# A Comparative Analysis of Panel Unit Root Tests under a Single Framework

**PhD Econometrics**


| | |
|---|---|
| **Researcher:** | **Shahid Akbar** |
| **Registration No.** | **26-SE/PhD (ET)/F13** |
| **Supervisor:** | **Dr. Abdul Rashid** |
| | **Associate Professor, IIU Islamabad.** |
| **Co-supervisor:** | **Dr. Atiq-ur-Rehman** |
| | **Associate Professor, AJK University, Muzaffarabad.** |

SCHOOL OF ECONOMICS

INTERNATIONAL INSTITUTE OF ISLAMIC ECONOMICS (IIIE)

INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD (IIUI), PAKISTAN.

I Panel Analysis
ii Time-Series Analysis
iii Simulated Critical value

# A Comparative Analysis of Panel Unit Root Tests under a Single Framework

Shahid Akbar

Registration No. 26-SE/PhD (ET)/F13

Submitted in partial fulfillment of the requirements for the

Doctor of Philosophy in Econometrics

at International Institute of Islamic Economics (IIIE)

International Islamic University Islamabad (IIUI)

Supervisor: Dr. Abdul Rashid

Co-supervisor: Dr. Atiq-ur-Rehman

January 2020

# DECLARATION

I hereby declare that this thesis, neither as a whole nor as a part thereof, has been copied out from any source. It is further declared that I have carried out this research by myself and have completed this thesis on the basis of my personal efforts under the guidance and help of my supervisor. If any part of this thesis is proven to be copied out or earlier submitted, I shall stand by the consequences. No portion of work presented in this thesis has been submitted in support of any application for any other degree or qualification in International Islamic University or any other university or institute of learning.

Shahid Akbar

January 2020.

## *DEDICATION*

***To***

***My Parents***

***&***

***My Family***

# APPROVAL SHEET

A Comparative Analysis of Panel Unit Root Tests under a Single Framework

by

**Shahid Akbar**

**Reg. No: 26-SE/PhD(Et)/F13**

Accepted by the International Institute of Islamic Economics (IIIE), International Islamic University Islamabad (IIUI), as partial fulfillment of the requirements for the award of degree of DOCTOR OF PHILOSOPHY IN ECONOMETRICS.

Supervisor

Dr. Abdul Rashid
Associate Professor, IIIE, IIUI.

Co-supervisor

Dr. Atiq-ur-Rehman
Associate Professor, AJK University, Muzaffarabad.

Internal Examiner

Dr. Arshad Ali Bhatti
Assistant Professor, IIIE, IIUI.

External Examiners

Dr. Asad Zaman
Ex-Vice Chancellor, Pakistan Institute of Development Economics (PIDE), Islamabad.

Dr. Zahid Asghar
Director, School of Economics, QAU, Islamabad.

Dr. Hamid Hassan
In-charge,
School of Economics

Dr. Atiquzzafar Khan
Director General

**International Institute of Islamic Economics (IIIE)**

**International Islamic University Islamabad (IIUI)**

Date of Viva-Voce Examination: **January 1, 2020**

# Acknowledgement

Starting with the name of Almighty Allah, Who is most merciful and beneficiary upon His creatures. I am very much thankful to Almighty Allah to give me knowledge and understanding to enhance my skills in the field of Econometrics.

I am thankful to my loving, obligated and respected parents; especially to my late mother, father, sister, brothers, wife and all of my other family members who give me courage and support me at every stage of my study. I am also thankful to my Supervisor Dr. Abdul Rashid, Co-Supervisor Dr. Atiq-ur-Rehman and Dr. Asad Zaman to teach me and guide me in a good manner.

Here I would acknowledge the ability and enhance teaching and research methods of my teachers Dr. Asad Zaman, Dr. Abdul Rashid, Dr. Mumtaz Ahmad, Dr. Iftikhar Hussain Adil and Dr. Atiq-ur-Rehaman. In this regard I am especially thankful to my respected and honorable teachers Dr. Asad Zaman, Dr. Abdul Rashid, Dr. Atiq-Ur-Rehaman, Dr. Mumtaz Ahmad and all faculty members of IIIE who encouraged and guided me whenever I need them. I am also very much thankful to my valuable, caring and loving friends and class fellows including Asad-ul-Islam, Mehmood-ul-Hussan, Abdul Waheed, Khan Bahadar, Ali Raza, Muhammad Irfan Malik, Yasir Riaz, Gulfam Haider, Waqar Muhammad Khan and Abdul Ghafar Shah. I also acknowledge the cooperation of administration staff of IIIE, especially, Syed Niaz Ali Shah for his support during my stay at IIIE.

Again I would acknowledge that today I am here just because of prayers of my father, mother, sister and brothers Jamshed Akbar and Sher Akbar, who always encouraged, financed and supported me at every stage of my study period.

**Shahid Akbar**

بسم الله الرحمن الرحيم

# Abstract

This study compare twenty four panel unit root tests belonging to the null hypothesis of unit root and null hypothesis of stationarity on the basis of size and power properties using Monte Carlo simulations. Eighteen tests having the null hypothesis of panel unit root and six tests having the null hypothesis of stationary have been compared using stringency criterion discussed by Zaman (1996) to make comparison under a unified framework. For this comparison, first size of all (unit root and stationary) tests have been stabilized around nominal size of 5% by using simulated critical value instead of asymptotic critical values. The critical values are computed by Monte Carlo simulations assuming different level of the cross section in the panel and different time series length. After equalizing size of all panel unit root tests, power comparison of tests have carried out for both categories of tests for two specification of deterministic parts: with intercept term and both with intercept and trend terms. A standard bench mark on the basis of maximum shortcomings is made to identify best, mediocre and worst tests before making comparison for fixed cross section units with varying level of time series and vice versa. It is observed that De Wachter, Harris and Tzavalis (DWH); Im, Pesaran, and Shin (IPS); Levin, Lu, and Chu (LLC); and Westerlund (WT) tests having the null hypothesis of panel unit root are found to be best at small, medium, and large samples. The second category tests having the null hypothesis of stationary have concluded Hadri (HD) and Hadri and Larsson (HL) tests as the best performing tests as compared to other stationary tests in the both specification of deterministic cases. Empirical evaluation of best performing panel unit root tests have been carried out using purchasing power parity hypothesis on the basis of bootstrap method for sixteen OECD countries. The results of empirical study justify simulation study results for the best performing tests on the basis of empirical power.

Keywords: Simulated Critical Value, Stringency Criterion, Maximum Shortcomings, Convergence Pattern, Divergence Pattern, Bootstrap Method.

JEL Classification: C12, C15, C23.

# Table of Contents

# List of Tables

# List of Figures

# List of Acronyms

| Acronyms | Complete Name |
|---|---|
| MP | Moon and Perron (2004) Test |
| PS | Phillips and Sul (2003) Test |
| BNG | Bai and Ng (2004) Test |
| LLC | Levin, Lin and Chu (2002) Test |
| BU | Breitung (1999) Test |
| IPS | Im Pesaran-Shin (2003) Test |
| MW | Maddala and Wu (1999) Test |
| CHO | Choi (2001) Test |
| CH | Chang (2002) Test |
| OS | Oh and So (2004) Test |
| CIPS | Cross Sectional Im, Pesaran, and Shin (2007) Test |
| CIPS_star | Modified Cross Sectional Im, Pesaran, and Shin (2007) Test |
| DWH | De Wachter, Harris and Tzavalis (2007) Test |
| RMA | Recursive Mean Adjustment (2009) Test |
| LW | Lee and Wu (2012) Test |
| DH | Demetrescu and Hanck (2012) Test |
| WT | Westerlund (2014) Test |
| WL | Westerlund and Larsson (2012) Test |
| HD | Hadri (2000) Test |
| HL | Hadri and Larsson (2005) Test |

| | |
|---|---|
| HLM | Harris, Leybourne and McCabe (2005) Test |
| SS | Shin and Snell (2006) Test |
| KK | Hadri and Kurozumi (2009) Test |
| DHT | Demetrescu, Hassler and Tarcolea (2010) Test |
| OLS | Ordinary Least Squares |
| MCSS | Monte Carlo Sample Size |
| ADF | Augmented Dickey Fuller |
| GLS | Generalized Least Square |
| DGP | Data Generating Process |
| AR | Autoregressive |
| | |
| | |
| | |
| | |
| | |

# Chapter 01

## Introduction

### 1.1. Brief Introduction

In the last three decades, unit root techniques have been utilized in the empirical literature to observe whether a process has an infinite long memory (implying that process will be permanently affected by an innovation) or limited memory (implying that process will be temporarily affected by an innovation). From economic point of view, these long- and short-run memory properties make the existence of a unit root an interesting question. In stationarity models, shocks (such as policy interventions) have a temporary effect, whereas in the unit root models, shocks have persistence effects that remain forever.

In the time series literature, there exists a large number of unit root tests developed in different studies, Dickey and Fuller (1979), Phillips and Perron (1988), Kwiatkowski et al. (1992), Ng and Perron (2001), Elliott, Rothenberg, and Stock (1996), Dickey and Pantula (1987), to identify an infinite long-memory or short-memory process. These tests have been developed based on different assumptions and structures for time series data. These tests get reasonable power if the prior specification decisions are correct, however, they do not have much power when there is mismatch in the data generating process and testing model. This problem of making specification decisions have been solved through the induction of adequate autocorrelation, trend, moving average, and heteroscedastic terms in the test model to get higher power for the tests under investigation (see, Dickey and Fuller (1981), Elliott, Rothenberg, and Stock (1996)). Despite all these, the performance of time series unit root tests is under investigation because every decision has a specific probability of error. One of these errors in the decision-making is the large sample size needed to make a reliable inference. In time series, such huge sample seems implausible showing the lack of information which results in low power gain of time series unit root tests to reject the null hypothesis of interest.

In order to increase power of unit root tests, researchers have introduced panel tests and bivariate unit root tests in their studies, these studies include: Levin, Lin, and Chu (2002), Im, Pesaran, and Shin (2003), Phillips and Sul (2003), Bai and Ng (2004), Hadri (2000), Hadri and Larsson (2005) etc. A two dimension unit root tests, panel unit root (PUR) tests, have an advantage over time series unit root tests by exploiting the information of cross section units along with time series units to have a superior power gain over time series unit root tests. In time series, if $Y_{it}$ and $X_{it}$ are two independent nonstationary random vectors, and a regression is performed from $Y_{it}$ to $X_{it}$ with respect to time series for a certain cross section unit resulting a lack of cointegrating relation between $Y_{it}$ and $X_{it}$, Results of this regression is characterized as spurious with nondegenerative limiting distribution of the regression coefficient, Barreira and Rodrigues (2005). Panel data tackle this problem when testing for unit root and cointegration by adding the cross section unit and increase the number of observations and power of the tests.

There are many PUR tests proposed in the literature to describe dissimilar features of panel data but these tests are complicated due to heterogeneity, unbalanced panels, and cross sectional dependence as compared to time series. Hadri (2000), Choi (2001), Levin, Lin, and Chu (2002), Im, Pesaran, and Shin (2003), Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004), Hadri and Larsson (2005), Pesaran (2007) etc. introduced their tests based on different properties to test for PUR. These tests are categorized into two groups. First group of tests are called first generation PUR tests, which are designed for independent and uncorrelated cross section properties. Hadri (2000), Choi (2001), Levin, Lin, and Chu (2002) and Im, Pesaran, and Shin (2003) are the first generation PUR tests. On the other hand, the second generation PUR tests have been developed for the correlated cross sections properties. The unit root tests of Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004), Hadri and Larsson (2005), and Pesaran (2007) belong to the second generation PUR tests.

Levin, Lin, and Chu laid the foundation of PUR tests in 1993 by introducing their test but their test was published in 2002. Levin, Lin, and Chu (2002) derived the limiting distribution for the test statistic which follow normal distribution as compared to non-standard limiting distribution of time series unit root tests. Maddala and Wu (1999) developed their PUR test, with the basic idea of Fisher (1934), using the p-value from the ADF regressions provided that cross-sectional independence is assumed. Breitung (1999) suggested a pooled PUR test which does not need bias correction factors. Breitung (1999) used an appropriate variable transformation to get test statistic without bias correction. Choi (2006) test was originally developed in year 2001 but published in 2006, introduced a PUR test using a two way error component model to model the cross-dependence. This two way error component model imposes same pair wise covarinces across the different cross section units. Chang (2002) introduced a nonlinear instrumental variable based approach to deal with cross sectional dependence. Im, Pesaran and Shin (2003) suggested a panel test based on the mean of individual unit root statistic for the dynamic heterogeneous panels.

Similarly, Phillips and Sul (2003) proposed PUR test using orthogonalization procedure which eliminates the common factors in the residual one factor model. Moon and Perron (2004) developed their PUR test based on de-factored observations by estimating the factor loadings using the principal component method. Bai and Ng (2004) proposed their test by allowing the possibility of unit roots in the common factors. Oh and So (2004) introduced a robust test based on the sum of sign test statistics in the dynamic heterogeneous panels across the groups. Pesaran (2007) announced his PUR test which uses individual ADF regressions with cross section averages to incorporate error cross section dependence.

De Wachter et al. (2007) introduced PUR tests based on generalized and instrumental variable estimators. Sul (2009) presented a recursive mean adjustment (RMA) procedural based PUR test under the cross sectional dependence. Lee and Wu (2012) introduced their test by accommodating smooth structural changes in deterministic components and cross sectional dependence among variables. Demetrescu and Hanck (2012) carried their research and developed an instrumental based error volatility PUR test, this test is constructed on the basis of robust to nonstationary of error volatility. Another robust test for heteroscedastic and cross correlated errors for the null of PUR has been introduced by Westerlund (2014). Westerlund (2014) elaborated this panel test statistic by assuming heteroscedasticity in the deterministic part of the model. Westerlund and Larsson (2012) used a random autoregressive coefficient across cross section units to propose a heterogeneous PUR test.

Besides PUR tests for the null of unit root there are few tests developed to test the null hypothesis of stationarity. Hadri (2000) laid the foundation of PUR tests under the null of stationarity and proposed a residual-based Lagranger Multiplier (LM) test to test the null hypothesis of stationarity of an individual series against alternative of PUR. Hadri and Larsson (2005) further enhanced Hadri (2000) test and introduced the test for finite time dimension in the panel data. Harris et al. (2005) developed a robust stationary test in the occurrence of both serial and cross sectional dependence across the panels. Snell and Shin (2006) accelerated their research in the null hypothesis of stationarity testing and developed a panel-based mean group test in the presence of serial correlation across time and heterogeneity across cross section units. Hadri and Kurozumi (2009) introduced a test which is based on common factor of the error term of the model to model cross section dependence. Demetrescu et al. (2010) suggested a new test for the null of stationarity which allows persistance correlation of the cross units in the form of an unbounded norm of the panel long run correlation matrix. However, the number of developing PUR tests are growing fast with enhanced assumptions and mathematical structure to tackle more and more problems which exist in panel data.

All of the above tests (i.e. tests for the null hypothesis of unit root and stationarity) rely on different assumptions specified for the independent and dependent cross sections and time series, and cross correlated and uncorrelated errors. Many comparative studies, which are based on Monte Carlo methods featuring different behavior of the panel data, have been conducted to evaluate the performance (i.e. comparison of size and power properties) of the PUR tests. These studies include, Maddala and Wu (1999), Hlouskova and Wagner (2006), Gutierrez (2006), Pesaran (2007) etc. However, all these comparative studies have taken a small number of PUR tests and many tests remained out of the comparison making the comparison meaningless. Also, every test is based on a particular set of assumptions and performs well in Monte Carlo experiments when the data generating process matches the assumptions for that test. There is no evidence that which of the test is most robust to the distributional assumptions and give reasonable performance even in case of violation of assumptions. Moreover, most of the comparative studies have carried out using asymptotic calculation which lose the validity in finite samples. A critical value adjustment is needed for small sample to make a reliable inference of comparative tests under investigation.

All these comparative studies have not achieved any inclusive inference showing the week and strong areas of these tests. For example, one author shows a specific PUR test to be better regarding size and power properties as compared to the other tests. While second author demonstrates a second study and shows that an another different test is having better size and power properties as compared to the size and power of all other tests under investigation.

This study, therefore, evaluate the performance of two categories of PUR tests: tests having the null hypothesis of unit root and tests having the null hypothesis of stationarity under a single frame work of stringency. Stringency criterion has been discussed by Zaman (1996) which compares the tests by considering the whole set of alternatives.

## 1.2. Motivations

There are four main problems in the previous comparative studies of PUR tests. However, we try to take these into account so that an extensive overview of the performance of PUR tests could be given.

In the existing literature, asymptotic critical value has been utilized to find the empirical size of tests by keeping in view nominal size equal to 5%, which causes over rejection problem. This over size problem creates misleading conclusions about panel data under investigation. All of the existing comparative studies have used asymptotic critical values to make comparison. This have caused over rejection problem for a certain test but that test is observed as powerful test according to power evaluation creating wrong conclusion about that test. All of such studies, Maddala and Wu (1999), Hlouskova and Wagner (2006), and Gutierrez (2006), have reported conclusive results. It is required to adjust the size of the tests before making comparison to make comparison meaningful. This study tackles this over and under sized problems and provides a simulated based critical value to stabilize the size of all the tests.

In previous comparative studies, a small number of PUR tests have been considered to evaluate the size and power properties of tests. Many tests remain out of the comparison, making the comparison meaningless. Every PUR test is based on particular set of assumptions and performs well into Monte Carlo simulation method when the data generating process matches the assumptions. There is no idea that which of the test is most robust to the distributional assumptions and give reasonable performance even in case of violation of assumptions. To the best of our knowledge, so for, Hlouskova and Wagner

(2006) have contacted an extensive study in which they have compared few PUR and stationarity tests through large scale simulation study but they get mixture results. This study takes a large number of major PUR tests to make an extensive comparison of the tests for different combination of time series and cross section dimensions.

All the existing comparative studies take a few alternatives hypothesis to make comparisons, our study take the whole set of alternatives for a similar category of tests (tests with null hypothesis of unit root and tests with null hypothesis of stationary) to properly evaluate the power performance of PUR tests at each alternative.

No unified framework has been used to compare PUR tests in the existing comparative studies. All of these simulation studies, Maddala and Wu (1999), Hlouskova and Wagner (2006), and Gutierrez (2006), produce different results because these studies do not follow a specific standard framework to compare PUR tests. Our study compares all PUR tests under a single framework to make comparison meaningful. In this study, this framework is known as stringency criterion to compare a similar category of tests.

## 1.3. Research Gaps

In the literature, numerous comparative studies have used asymptotic critical value to evaluate size and power performance of PUR tests but this cause over rejection problem for the null hypothesis of interest. This study uses simulated critical value instead of asymptotic critical value to overcome size distortion problem and stabilize the size of all PUR tests so that power comparison of PUR tests makes scene. Also, the unified framework (stringency criterion) which is used in this study based on stabilized size to make power comparison meaningful.

Previous comparative studies, Maddala and Wu (1999), Hlouskova and Wagner (2006), and Gutierrez (2006), have taken few alternatives to analyze the power of tests having the null hypothesis of unit root and null hypothesis of stationary. In order to tackle this issue, this study considers all set of alternatives in a whole sample space for the tests under consideration to evaluate the power of PUR tests comprehensively for different specification of deterministic terms.

The main gap in the literature is that no single study has compared a large number of tests under a single framework. However, this study has tackled this problem and have compared twenty four PUR tests both having the null hypothesis of PUR and stationary under a single frame work which provides us to have a comprehensive look of size and power properties of all PUR tests.

## 1.4. Objectives of the Study

1. The first objective of our study is to investigate the least size distortive, largest size distortive, and stable size PUR tests having the null hypothesis of unit root and stationary from nominal size of 5% under the usage of asymptotic critical values.
2. As mentioned earlier that stringency criterion uses a stabilized size of the tests to be compared in a single framework. The second main objective of the study is to stabilize the size of all PUR tests by finding simulated critical values of all tests under consideration. These simulated critical values are calculated for various specification (i.e. with all combination of cross section and time series dimensions in the presence of deterministic terms) under $H_0$.
3. After stabilizing the size, the third main objective of the study is to find power at various specification under $H_A$ and identify the best performer test for the null

hypothesis of unit root in panel data for various specifications under stringency criterion. The secondary objective of the study is to identify mediocre and worst performer tests for the null hypothesis of unit root for various specifications (with all combination of cross section and time series dimensions in the presence of deterministic terms) under stringency criterion. For this purpose, a point optimal test for null hypothesis of PUR under the DGP-A (Data Generating Process) in methodology section has been discussed which is based on Neyman Pearson Lemma (Neyman and Pearson (1992)) for stringency criterion.

4. The fourth main objective of the study is to identify the best performer test having the null hypothesis of stationarity for varying level of time series and cross section dimensions in the presence of deterministic terms. The secondary objective is to identify mediocre and worst performer tests for the null hypothesis of stationarity in panel data by examining power of the tests under stringency criterion for all combination of cross section and time series dimensions. A point optimal test for the null of stationarity is given in the methodology section under the DGP-B to determine the best, mediocre and worst test for each cross section with smallest to largest sample sizes.
5. The fifth main objective of our study is to apply best performing PUR tests on monthly real data from OECD countries based on purchasing power parity hypothesis and empirically evaluate the powers of these tests using bootstrap method.

## 1.5. Significance of the Study

This study helps researchers to use simulated critical values rather than asymptotic critical values to avoid over rejection problem for certain tests in the real data. If a test faces over rejection problem then it provides a wrong decision about the behavior of integrated series.

Further, this study guides practitioner to evaluate the power performance of all tests under a whole set of alternative rather than for a few alternatives to make power evaluation clearly at each specific alternative point of all tests with all combination of specifications belong to a similar category (i.e. tests having the null hypothesis of unit root and test having the null hypothesis of stationary).

Point optimal tests help researchers how to use these tests for comparative purposes to avoid misleading and biased results that is observed in the past comparative studies. These tests also help researchers to understand the importance of such a large number of tests for comparison purpose which enables practitioners to use a PUR test in real data which has better power as compared to the other tests for small, medium and large time series and cross section dimensions corresponding to different specification of deterministic terms under a unified framework. Also, practitioners will also avoid to use a test of PUR which has worst performance under stringency criterion for the null hypothesis of unit root and stationarity for panel data.

This study will help researchers in the field of applied economics and finance to apply best performer tests and avoid applying worst performer ones for different level of time series and cross section dimensions to investigate the integrated series in panel data. In this study, purchasing power parity example has been taken and discussed for the best performer tests by using bootstrap method to empirically investigate their powers in finite and large time and cross section units. This will help practitioners to use the same methodology for other real panel data.

## 1.6. Structure of the Thesis

This study contains eight chapters. The first chapter includes the brief introduction of our study, research motivation, research gap, study objectives and study significance. The second chapter is organized to discuss the background of the PUR tests proposed in the literature in the first section, while second section discusses revision of the comparative studies of PUR tests. A methodological layout is discussed in Chapter 3, which describes two type of data generating processes, point optimal tests, power curves and envelopes, and stringency criterion to carry out comparative study.

In Chapter 4, size distortion of both types of PUR tests (PUR and panel stationary tests) have been discussed under the asymptotic critical value. In Chapter 5, a detailed study of controlling the size under a simulated critical value is described for the null hypothesis of unit root and stationary. Sixth chapter of the study makes comparison of the PUR and stationary tests under the stringency criterion to analyze the effect of time series over cross section units and vice versa. Seventh chapter of our study empirically evaluates the power performance of best performing PUR tests for OECD countries. Last chapter of this study is organized to discuss overall conclusions and recommendations. The directions for future research are also mentioned at the end of final chapter (Chapter 8).

# Chapter 02

# Literature Review

## 2.1. The Priors Panel Unit Root Tests

Testing for unit root in the field of economics and finance for panel data has become a regular and mandatory practice among researchers after the appearance of the paper of Levin, Lin, and Chu (1993). Although, they have published their PUR based paper after ten years. It is generally accepted argument that unit root tests based on time series data, like Dickey-Fuller (DF), Phillips-Perron (PP), Augmented Dicky-Fuller (ADF), Kwiatkowski, Phillips, Schmidt and Shin (KPSS), Ng and Perron (NP), Generalized Least Square transformed Dickey-Fuller (DFGLS), and Dickey–Pantula (DP) tests lack power in distinguishing the null of unit root from alternatives of stationary in small samples. In this aspect, information related to various countries or individuals are added to increase the power of unit root tests.

## 2.2. Review of Panel Unit Root Tests

In the first section of revision of PUR tests, first and second generation of PUR tests have been discussed under the null hypothesis of PUR and under the alternative hypothesis of stationary. In the second and last part of the literature review, the first and second generation of panel stationary tests have been deliberated with their assumptions and mathematical structure.

### 2.2.1. Panel Unit Root Tests with Null Hypothesis of Unit Root

#### 2.2.1.1. Moon and Perron (MP) Test

Moon and Perron (2004) proposed their test for cross-correlated panels using the factor model approach to model the cross sectional units. They argues that tests with assumption of cross sectional independence, like LLC test, face size distortion problem if the common factors are present in the panel. They incorporated cross sectional dependence problem by eliminating the common components of the series through proper transformation. This test considers the following dynamic model.

$$y_{it} = X_{it}'\delta_i + x_{it}\ , \quad x_{it} = \rho x_{it-1} + u_{it}\ , \ u_{it} = \beta_i' f_t + e_{it} \tag{2.1}$$

where $X_{it}$ denotes the deterministic terms, $x_{it}$ represents an autoregressive process and $u_{it}$ follows a factor model with a common factor component $f_t$ and with $\beta_i$ factor loading coefficients. Moon and Perron (2004) used projection matrix $Q_\beta$ to remove the cross section dependence in equation (2.1) and compute the unbiased estimator as follows,

$$\rho_{pool}^{+} = \frac{tr\left(Y_{-1}Q_\beta Y'\right) - NT\lambda_e^N}{tr\left(Y_{-1}Q_\beta Y_{-1}'\right)} \tag{2.2}$$

where $Y_{-1}$ denotes the lagged matrix, $tr(.)$ and $\lambda_e^N$ shows the trace operator and one-sided long-run variance of the cross sectional average of $e_{it}$, respectively. They used principal components of $\hat{e}'\hat{e} = (Y - \hat{\rho}_{pool}Y_{-1})'(Y - \hat{\rho}_{pool}Y_{-1})$ to compute projection matrix $Q_\beta$ and factor loadings coefficient $\beta_i$, where a pooled AR ( $\hat{\rho}_{pool}$ ) is the ordinary least square estimate. On the basis of (2.2), Moon and Perron (2004) proposed the following two modified test statistics:

$$t_a^* = \frac{\sqrt{N}T\left(\hat{\rho}_{pool}^{+}-1\right)}{\sqrt{2\hat{\phi}_e^4 / \hat{\omega}_e^4}} \tag{2.3}$$

And

$$t_b^* = \sqrt{N}T\left(\hat{\rho}_{pool}^{+}-1\right)\sqrt{\frac{1}{NT^2}tr\left(Y_{-1}Q_{\beta}Y_{-1}^{'}\right)\frac{\hat{\omega}_e^2}{\hat{\phi}_e^4}} \tag{2.4}$$

where $\hat{\rho}_{pool}^{+}$ is the biased corrected pooled AR estimate of equation (2.2), $\hat{\omega}_e^2$ is the cross sectional averages of long run variance of $\hat{e}_{it}$ and $\hat{\phi}_e^4$ denotes the cross sectional average of $\hat{\omega}_{e,i}^4$.

Moon and Perron (2004) found that both of the statistics have converge to normal distribution under null hypothesis of unit root while have a divergent behavior under alternative hypothesis of stationarity. In fact, the models of MP and LLC tests become similar with common AR root if cross section dependence unit removed from Moon and Perron (2004). In other words, both of the models with pooled estimators have same convergence rate by removing the cross section dependence in Moon and Perron's (2004) model.

#### 2.2.1.2. Phillips and Sul (PS) Test

Phillips and Sul (2003) suggested a PUR test similar to the Moon and Perron (2004) test procedure but differs in that only one factor is allowed in equation (2.1) which is supposed to be independently distributed over time. Phillips and Sul (2003) computed projection matrix $Q_{\beta}$ by a moment based method instead of principal components method by eliminating the common factors which creates the main difference between Moon and Perron (2004) and their own test. They proposed the following test statistic:

$$G_{OLS}^{++} = \frac{1}{\sqrt{N}\sigma_{\varsigma}} \sum_{i=1}^{N-1} \left[ \frac{(\hat{\rho}_i^{+} - 1)}{\hat{\sigma}_{\hat{\rho}_i^{+}}} - \mu_{\varsigma} \right] \tag{2.5}$$

where $\hat{\rho}_i^{+}$ and $\hat{\sigma}_{\hat{\rho}_i^{+}}$ denote de-factored data's cross sectional autoregressive estimates and its standard errors, while $\mu_{\varsigma}$ and $\hat{\sigma}_{\hat{\rho}_i^{+}}$ show the asymptotic mean and variance, respectively. Further, Phillips and Sul (2003) calculated their statistic asymptotic distribution and showed that it converge to normal distribution as the time and cross section dimension goes towards infinity.

#### 2.2.1.3. Bai and Ng (BNG) Test

Bai and Ng (2004) suggested a different methodology as compared to the MP and PS PUR tests. They considered a factor structure of large dimensional panels to understand the nature of non-stationary in the data. Bai and Ng (2004) started with the following model to develop their test

$$y_{it} = X_{it}^{'}\delta_t + \lambda_i^{'}F_t + e_{it} \tag{2.6}$$

$$F_t = F_{t-1} + f_t$$

and $$e_{it} = \rho_i e_{it-1} + \varepsilon_{it}$$

where $X_{it}$ denotes the deterministic components, $F_t$ and $\lambda_i$ show a vector of common factors and associated factor loadings, respectively. In other words, series $y_{it}$ is generated as the combination of deterministic components, common component, and the idiosyncratic component. Moreover, $e_{it}$ and $f_t$ are generated from moving average process. Bai and Ng (2004) adopted equations (2.1) and (2.6) with only one factor to test the null hypothesis of unit root. Hence, they have the following model in the presence of both of the non-stationary components (i.e. idiosyncratic components and common factor).

$$y_{it} = \bar{\alpha}_{i0} + \lambda_i^{'} \sum_{t=1}^{T} f_t + \sum_{t=1}^{T} \varepsilon_{it}.$$

Both of these components, idiosyncratic components and common factor, are separately tested for unit root. Also, under the alternative hypothesis, equations (2.1) and (2.6) show very marginal differences. They used principal component method of the matrix of observed data to compute common factors ($\Delta f_t$) and associated factor loadings ($\beta$) consistent estimates. Thus, estimated residuals are calculated as $\Delta \hat{\varepsilon}_{it} = \Delta y_{it} - \hat{\lambda}_i^{'} \Delta \hat{f}$, and

$$\hat{f}_t = \sum_{s=2}^{T} \Delta f_s \,,\ \hat{\varepsilon}_{it} = \sum_{s=2}^{T} \Delta \hat{\varepsilon}_{is} \,,\ i = 1, 2, \ldots, N.$$

Thus indicates a straight and simple way to test the null hypothesis of unit root for the idiosyncratic component and common factor separately. If only one factor is present with factor component, Bai and Ng (2004) suggested to use simple ADF test. However, if the number of factors are more than one then a modified test statistics of Stock and Watson (1988) have been suggested by Bai and Ng (2004). If the idiosyncratic components are to be treated, then Bai and Ng (2004) used a meta-analysis procedure which is basically developed by Maddala and Wu (1999). This procedure combines p-values from any unit root test statistic. Bai and Ng (2004) computed ADF test for each cross section and combined p-values against each ADF test statistic for each idiosyncratic component. The test statistic is given by,

$$P_{\hat{e}} = \frac{-2\sum_{i=1}^{N} \log p_{\hat{e}}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1) \tag{2.7}$$

where $p_{\hat{e}}$ denotes the p-value of the ADF test on the estimated residuals $\hat{e}_{it}$ in equation (2.6). The test statistic in equation (2.7) converges to standard normal distribution if errors are assumed as independent.

**2.2.1.4. Levin, Lin, and Chu (LLC) Test**

Levin, Lin, and Chu published their work in 2002 which was initiated in 1993, in their study, they proposed a pooled based panel test statistic to test the null hypothesis of unit root for each individual time series against the alternative hypothesis of stationary for each time series. This test is considered to be the first one in the PUR literature as well under the assumption of cross sectional independence. Similarly, error variance and serial correlation across individuals are permitted to vary freely to develop their test. They used the following model to develop their test statistic.

$$\Delta y_{it} = \alpha_i y_{i,t-1} + \sum_{j=1}^{p_i} \beta_{ij} \Delta y_{i,t-j} + X'_{it}\delta + \varepsilon_{it} \tag{2.8}$$

To test $H_0 : \alpha_i = 0$ against homogenous alternative $H_1 : \alpha_i < 0$, where $\alpha_i = \rho_i - 1$ and $p_i$, to vary across cross sections. $X_{it}$ shows the deterministic terms involved in equation (2.8). LLC test has the following procedure.

First, regressing both $\Delta y_{it}$ and $y_{it-1}$ on $\Delta y_{i,t-j}$ (for j = 1,2,...,pi) and $X_{it}$ to get $(\hat{\beta},\hat{\delta})$ and $(\tilde{\beta},\tilde{\delta})$. Secondly, by considering

$$\Delta \bar{y}_{it} = \Delta y_{it} - \sum_{j=1}^{p_i} \hat{\beta}_{ij} \Delta y_{i,t-j} - X'_{it}\hat{\delta} \text{ and } \bar{y}_{i,t-1} = y_{i,t-1} - \sum_{j=1}^{p_i} \tilde{\beta}_{ij} \Delta y_{i,t-j} - X'_{it}\tilde{\delta}$$

Auxiliary equations have been used to remove autocorrelation and deterministic terms. Thirdly, proxy variables $\Delta\tilde{y}_{it}$ and $\tilde{y}_{i,t-1}$ have been obtained by dividing $\Delta\bar{y}_{it}$ and $\bar{y}_{i,t-1}$ by standard error $S_i$ of the ADF equation (2.8). Lastly, a pooled proxy equation $\Delta\tilde{y}_{it} = \alpha\tilde{y}_{i,t-1} + \eta_{it}$ is used to estimate $\hat{\alpha}$. In the absence of deterministic terms the test statistic $t_\alpha$ converges to standard normal distribution while in the presence of deterministic terms (i.e. with drift term only and, with drift and trend term both) this test statistic diverges to minus infinity. In the case of divergence, a normalized has been achieved towards normal distribution by reentering and normalizing the $t_\alpha$. Thus, LLC test statistic is given by

$$t_\alpha^{*} = \frac{t_\alpha - (N\tilde{T})\hat{S}_N \hat{\sigma}_{\tilde{\varepsilon}}^{-2} se(\hat{\alpha}) \mu_{m\tilde{T}}{*}}{\sigma_{m\tilde{T}}{*}} \rightarrow N(0,1)$$

where, $t_\alpha = \hat{\alpha} / se(\hat{\alpha})$, $\tilde{T} = T - (\sum_i p_i / N) - 1$, $\hat{\sigma}_{\tilde{\varepsilon}}^2$ and $\hat{S}_N$ denote the error term estimated variance and mean ratios of long run to short run standard deviation in equation (2.8) for each individual, respectively. A kernel based technique has been used to find the long run variance. The adjustment terms of mean ($\mu_{m\tilde{T}}{*}$) and variance ($\sigma_{m\tilde{T}}{*}$) have been used from

LLC tables for different deterministic terms. Also, this adjusted test statistic converges to standard normal distribution in the present of deterministic terms.

### 2.2.1.5. Breitung (BU) Test

Breitung (1999) developed a pooled PUR test like LLC test using model (2.8) and assumption of independence. This test does not require bias correction factors as compared to LLC test which is achieved through appropriate variable transformation. Like LLC test, BU test also has three cases to develop PUR test. In the first case, if there is no deterministic term in the model then no bias correction is required to develop the test statistic. If only drift term is present then by subtracting the initial value of the data bias correction factors can be removed. In these both cases, the test statistics follow standard normal distribution by avoiding the correction factors which were used in LLC test. However, if both of the deterministic terms are present in the model equation then a different method is adopted as compared to first two cases. In this case, a complex correction factors are applied after removing the serial correlation in the regression. In short, BU test is different in two distinct ways than LLC. First, proxy variables are obtained by removing only AR terms. That is,

$$\Delta\tilde{y}_{it} = \left( \frac{\Delta y_{it} - \sum_{j=1}^{p_i} \hat{\beta}_{ij} \Delta y_{i,t-j}}{s_i} \right), \quad \tilde{y}_{i,t-1} = \left( \frac{y_{i,t-1} - \sum_{j=1}^{p_i} \tilde{\beta}_{ij} \Delta y_{i,t-j}}{s_i} \right)$$

Second, both of the proxy variables have been transformed and detrended.

$$\left( \Delta y_{it}^{*} = \sqrt{\frac{(T-t)}{(T-t+1)}} \left[ \Delta\tilde{y}_{it} - \frac{1}{T-t} (\Delta\tilde{y}_{it+1} + \ldots\ldots + \Delta\tilde{y}_{iT}) \right] \right)$$

and $$y_{it}^{*} = \tilde{y}_{it} - \tilde{y}_{i1} - \frac{t-1}{T-1} (\tilde{y}_{iT} - \tilde{y}_{i1})$$

Then estimated value of $\alpha$ is obtained from the following regression.

$$\Delta y_{it}^{*} = \alpha y_{i,t-1}^{*} + v_{it}$$

According to Breitung (1999), $\alpha^*$ converges to standard normal distribution asymptotically. The number of lags specification, which is used in cross section ADF regression, and exogenous variables are required in Breitung (1999) test method to distinguish it from LLC test. Also, BU test does not need to compute kernel density estimator as LLC does.

**2.2.1.6. Im, Pesaran, and Shin (IPS) Test**

Im, Pesaran, and Shin (2003) suggested a PUR test for dynamic heterogeneous panel based on average of individual unit root test of the ADF regression. This test is developed under the assumption of cross sectional dependence and can be used if drift only or both drift and trend terms are present in the model equation under investigation. Im, Pesaran, and Shin (2003) have used an identical lag lengths and a balanced panel is required to apply this test statistic. They used the model presented in equation (2.8) to test the hypothesis $H_0 : \alpha_i = 0$ against $H_A : \alpha_i < 0$. The average of the *t*-statistics from equation (2.8) of individual ADF regressions is defined as by Im, Pesaran, and Shin (2003),

$$\tilde{t}bar = \frac{1}{N}\sum_{i=1}^{N}\tilde{t}_{iT}$$

where $\tilde{t}_{iT} = \dfrac{\Delta y_i' M_\tau y_{i,-1}}{\tilde{\sigma}_{iT}\left(y_{i,-1}' M_\tau y_{i,-1}\right)^{1/2}}$, $\tilde{\sigma}_{iT}^2 = \dfrac{\Delta y_i' M_\tau \Delta y_i}{(T-1)}$, $M_\tau = I_T - \tau_T\left(\tau_T'\tau_T\right)^{-1}\tau_T'$, and $\tau = (1,1,\ldots,1)'$. This test statistic is simply called t-bar test in the literature and is abbreviated as the IPS test. According to Lyapunov central limit theorem, this test follows a standard normal distribution under the null hypothesis for a fixed T when T>5 in the case of drift term only and, T>6 in the presence of both of the drift and trend terms, and is defined as.

$$Z_{\tilde{t}bar} = \frac{\sqrt{N}\left\{\tilde{t}bar - N^{-1}\sum_{i=1}^{N} E\left(\tilde{t}_{T_i}\right)\right\}}{\sqrt{N^{-1}\sum_{i=1}^{N} Var\left(\tilde{t}_{T_i}\right)}} \overset{N}{\Rightarrow} N(0,1)$$

Here $E\left(\tilde{t}_{T_i}\right)$ and $Var\left(\tilde{t}_{T_i}\right)$ denotes the mean and variance of $\tilde{t}bar$ test statistic, respectively. The values of these moments are given in the Im, Pesaran, and Shin (2003) paper for different level of time series and cross section dimensions which have been obtained from Monte Carlo Simulations. Im, Pesaran, and Shin (2003) conducted a simulation study and found that IPS test is performing better as compared to LLC test if the number of lag order to be selected is large from the ADF regressions.

**2.2.1.7. Fisher Type Tests**

Maddala and Wu (1999) and Choi (2001) used the Fisher (1934) results of combining the p-values to develop their PUR tests.

**2.2.1.7.1. Maddala and Wu (MW) Test**

This test statistic is dated back to Fisher (1934) and is used to test the significance of independent tests of a hypothesis from independent cross sectional results. MW PUR test is developed by combining the observed significance level (p-value) of the individual test statistic of the ADF regression using the assumption of cross sectional independence. Using equation (2.8) as a base model to test $H_0 : \alpha_i = 0$ against $H_A : \alpha_i < 0$, p-value of each cross section is obtained and PUR test is defined as.

$$P = -2\sum_{i=1}^{N} \ln p_i \rightarrow \chi^2_{2N}$$

where 'pᵢ' represent the observed significance level of the ADF regression for each cross section. This test statistic can be applied for any type of PUR problem fulfilling the condition of cross sectional independence. Also, MW test statistic does not need balanced and identical lag length for each cross section equation.

### 2.2.1.7.2. Choi (CHO) Test

CHO test developed using the same regression model and same methodology as MW test did but for sufficiently large cross section unit and with a degenerate limit distribution. Choi (2001) suggested the following panel statistic to test $H_0 : \alpha_i = 0$ against $H_A : \alpha_i < 0$ with $E(-2\ln p_i) = 2$ and $Var(-2\ln p_i) = 4$.

$$Z = \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (-2\ln p_i - 2) \rightarrow N(0,1)$$

This test statistic is corresponded to the standardized form of the average of cross sectional univariate p-values. According to Lindeberg-Levy central limit theorem, this test statistic converges to standard normal distribution under the null hypothesis of unit root and assumption of cross sectional independence of p-values as both time series and cross section units approaches towards infinite.

### 2.2.1.8. Chang (CH) Test

CH test introduced by Chang in 2002 to model the cross sectional dependency of panel data using a non-linear instrumental variable estimation method. Cross sectional dependency structure is utilized among the residuals of the cross units. Also, CH test allows both balance and unbalance panels under the null hypothesis of PUR under investigation. The idea is to estimate the ADF regression using a nonlinear instrumental variable estimation method and lagged levels of $y_{i,t-1}$'s as instruments nonlinear transformation for each cross section. A nonlinear transformation of lagged values of dependent variable is used as instruments to develop CH test. Chang (2002) used the following model to introduce CH test.

$$y_{it} = X_{it}'\delta_i + \rho y_{i,t-1} + \sum_{k=1}^{p_i} \alpha_{i,k} \Delta y_{i,t-k} + \varepsilon_{it} \tag{2.9}$$

where $\alpha^i(L)u_{it} = \varepsilon_{it}$ is an AR(pi) process. An average of the individual test statistic from equation (2.9) has been taken to test the null hypothesis $H_0 : \rho = 1$ for all $i$ vs $H_1 : |\rho| < 1$ for some $i$. Chang (2002) introduced the following average t-ratio test statistic

$$S_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_i, \quad \text{where } Z_i = \frac{\hat{\rho}_i - 1}{s(\hat{\rho}_i)}$$

where $\hat{\rho}_i$ is defined as $\rho_i - 1 = B_{T_i}^{-1} A_{T_i}$ and , $s(\hat{\rho}_i)$ is the standard error of the IV-estimator $\hat{\rho}_i$ given by $s(\hat{\rho}_i)^2 = \hat{\sigma}_i^2 B_{T_i}^{-2} C_{T_i}$,

$$C_{T_i} = \sum_{t=1}^{T_i} F(y_{i,t-1})^2 - \sum_{t=1}^{T_i} F(y_{i,t-1}) x_{it}' \left( \sum_{t=1}^{T_i} x_{it} x_{it}' \right)^{-1} \sum_{t=1}^{T_i} x_{it} F(y_{i,t-1})$$

$$A_{T_i} = \sum_{t=1}^{T_i} F(y_{i,t-1}) \varepsilon_{it} - \sum_{t=1}^{T_i} F(y_{i,t-1}) x_{it}' \left( \sum_{t=1}^{T_i} x_{it} x_{it}' \right)^{-1} \sum_{t=1}^{T_i} x_{it} \varepsilon_{it},$$

$$B_{T_i} = \sum_{t=1}^{T_i} F(y_{i,t-1}) y_{i,t-1} - \sum_{t=1}^{T_i} F(y_{i,t-1}) x_{it}' \left( \sum_{t=1}^{T_i} x_{it} x_{it}' \right)^{-1} \sum_{t=1}^{T_i} x_{it} y_{i,t-1},$$

$$x_{it}' = \left( \Delta y_{i,t-1}, \ldots\ldots, \Delta y_{i,t-p_i} \right),$$

And $\hat{\sigma}_i^2 = T_i^{-1} \sum_{t=1}^{T_i} \hat{\varepsilon}_{it}^2$,

$\hat{\varepsilon}_{it}$ denotes the fitted residuals from the base regression and F, i.e. $F(y_{i,t-1})$ is called the instrument generating function (IGF) which is to be correlated with $y_{i,t-1}$. According to

Chang (2002), CH test statistic follows standard normal distribution asymptotically for time series diemension only. In the presence of the determinstic terms, a proper demeaned and detrended method is suggested before applying CH test statistic. If model equation has only, a drift term, then the demeaned series is defined on the basis of equation (2.9) and $z_{it} = \mu_i + y_{it}$ as

$$y_{it}^{\mu} = \alpha_i y_{i,t-1}^{\mu} + \sum_{k=1}^{p_i} \alpha_{i,k} \Delta y_{i,t-k}^{\mu} + e_{it},$$

where $$y_{it}^{\mu} = z_{it} - \frac{1}{t-1} \sum_{k=1}^{t-1} z_{ik},$$

$$y_{i,t-1}^{\mu} = z_{i,t-1} - \frac{1}{t-1} \sum_{k=1}^{t-1} z_{ik},$$

$$\Delta y_{i,t-k}^{\mu} = \Delta z_{i,t-k}, \qquad k = 1, 2, \dots\dots, p_i$$

If a linear trend model is taken into account, then the following detrended procedure is utilized for $z_{it} = \mu_i + \delta_i t + y_{it}$ as

$$y_{it}^{\tau} = \alpha_i y_{i,t-1}^{\tau} + \sum_{k=1}^{p_i} \alpha_{i,k} \Delta y_{i,t-k}^{\tau} + e_{it},$$

where $$y_{it}^{\tau} = z_{it} + \frac{2}{t-1} \sum_{k=1}^{t-1} z_{ik} - \frac{6}{t(t-1)} \sum_{k=1}^{t-1} k z_{ik} - \frac{1}{T} z_{iT_i},$$

$$y_{i,t-1}^{\tau} = z_{i,t-1} - \frac{2}{t-1} \sum_{k=1}^{t-1} z_{ik} - \frac{6}{t(t-1)} \sum_{k=1}^{t-1} k z_{ik},$$

$$\Delta y^{\tau}_{i,t-k} = \Delta z_{i,t-k} - \frac{1}{T} z_{iT_i}, \qquad k = 1, 2, \ldots\ldots p_i$$

### 2.2.1.9. Oh and So (OS) Test

Oh and So (2004) published their work in the PUR literature and suggested a robust PUR test grounded on the sum of sign-type test statistic across the groups in the dynamic heterogeneous panels. OS test is considered to valid under weaker assumptions than OLS based tests on the error terms. According to Oh and So (2004), in the literature, error term is treated under normal assumption which cause lack of power problem in the tests under investigation. In order to tackle this problem, Oh and SO (2004) developed their test on the basis of non-normal assumption of the error term and have taken heavy-tailed errors. This test also applicable both for balance and unbalance panels. Oh and So (2004) used the following model to introduce OS test

$$y_{it} = X'_{it}\delta_i + u_{it}$$

$$u_{it} = \rho_i u_{i,t-1} + e_{it}, \; i = 1, \ldots, N, \qquad t = 1, \ldots\ldots, T_i$$

And using hypothesis: $H_0 : \rho_i = 1$ vs $H_1 : |\rho_i| < 1$, for all $i$.

Using this model, OS test statistic is defined as follows

$$PS = \sum_{i=1}^{N} S^{(i)}_{T_i} \text{ , here } PS_p = \sum_{i=1}^{N}\sum_{t=1}^{T_i} sign\left(\Delta y_{it} - \sum_{j=1}^{p-1}\tilde{\alpha}_{ij}\Delta y_{i,t-1}\right) sign\left(y_{i,t-1} - \hat{\mu}_{i,t-1}\right)$$

where $\hat{\mu}_{it}$ is the recursive median of $y_{i1}, \ldots\ldots, y_{it}$ and $\tilde{\alpha}_{ij}$ is an estimator of $\alpha_{ij}$ based on $y_{i1}, \ldots\ldots, y_{iT_i}$. Under the null hypothesis, this test statistic converges to binomial distribution as $(PS + n)/2 \sim BIN(n, 0.5)$ and as n tends to infinite then $PS/\sqrt{n} \Rightarrow N(0,1)$. Oh and So

(2004) compared this test with IPS test and concluded that their test (i.e. OS test) performs remarkably well then IPS test in terms of size and power properties in the presence of heavy tailed non-normal errors.

### 2.2.1.10. Pesaran Tests

Pesaran (2007) further contributed in PUR tests literature and developed a PUR test to test unit roots in dynamic panel. Pesaran's (2007) test takes the standard ADF regressions for the individual series with averages of current and lag cross sections to deal with cross section dependence instead of orthogonalization procedure, which have been used in the previous PUR tests, in the panel data before the PUR tests are utilized on the transformed data. Pesaran (2007) developed his tests by utilizing the following heterogeneous model:

$$y_{it} = (1-\phi_i)\mu_i + \phi_i y_{i,t-1} + u_{it}, \quad i = 1,\ldots\ldots N; t = 1,\ldots\ldots T$$

$$u_{it} = \gamma_i f_t + \varepsilon_{it},$$

and a convenient way to form both of these equations as.

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i f + \varepsilon_{it},$$

where $\alpha_i = (1-\phi_i)\mu_i,\ \beta_i = -(1-\phi_i),\ \Delta y_{it} = y_{it} - y_{i,t-1}$

Now, to test the null hypothesis $H_0 : \beta_i = 1$ for all $i$ vs $H_A : \beta_i < 1$, Pesaran (2007) proxied the common factor $f_t$ with mean of cross sectional series $y_{it}$. Pesaran (2007) used the following cross section augmented Dickey Fuller (CADF) regression model to develop his test statistic.

$$\Delta y_{it} = X_{it}^{'}\delta_i + by_{i,t-1} + c_i \bar{y}_{t-1} + d_i \Delta\bar{y}_t + e_{it}$$

Pesaran (2007) proposed two test statistics for the cross sectional augmented Dicky Fuller (CADF) regression.

i- The first test statistic is named as CIPS and is defined as follows.

$$CIPS(N,T)=N^{-1}\sum_{i=1}^{N}t_i(N,T)$$

This shows that CIPS test is simply the cross-sectional average of the individual CADF t-statistics, which is defined as follows.

$$t_i(N,T)=\frac{\Delta y_i'\bar{M}_i y_{i,-1}'}{\bar{\sigma}_i\left(y_{i,-1}'\bar{M}_i y_{i,-1}\right)^{1/2}}$$

where $\bar{\sigma}_i^2=(T-6)^{-1}\Delta y_i'\bar{M}_{i,w}\Delta y_i$, $\bar{M}_i=I_T-\bar{W}_i\left(\bar{W}_i'\bar{W}_i\right)^{-1}\bar{W}_i'$, $M_{i,w}=I_T-G_i\left(G_i'G_i\right)^{-1}G_i'$ and $G_i=\left(y_{i,-1},\bar{W}_i\right)$, $\bar{W}_i=\left(\Delta y_{i,-1},\Delta\bar{y},\Delta\bar{y}_{-1},\tau_T,\bar{y}_{-1}\right)$, $\tau=(1,1,......,1)'$.

ii- The second test statistics is named as CIPS_star, which is the truncated PUR test of CIPS defines earlier and is defined as.

$$CIPS\_star(N,T)=N^{-1}\sum_{i=1}^{N}t_i^*(N,T)$$

where,

$$t_i^*(N,T)=t_i(N,T), \quad \text{if } -K_1<t_i(N,T)<K_2$$

$$t_i^*(N,T)=-K_1, \quad \text{if } t_i(N,T)\le -K_1$$

$$t_i^*(N,T)=K_2, \quad \text{if } t_i(N,T)\ge K_2$$

$K_1$ and $K_2$ are sufficiently large positive constants, so that $\Pr\left[-K_1<t_i(N,T)<K_2\right]$ is also sufficiently large. As, we are conducting study for two deterministic parts of the model i.e. model with intercept and no trend case-I, and model with an intercept and trend case-II. Hence,

For case-I: $K_1=6.19, \quad K_2=2.61$

For case-II: $K_1 = 6.42, \quad K_2 = 1.70$

In simulation part of Pesaran paper, he showed that both of these tests perform better from size and power properties.

### 2.2.1.11. De Wachter, Harris and Tzavalis (DWH) Test

DWH test is developed to inspect the effect of time dimension and serial correlation of the residuals in the panel data model under the additional assumption of homogenous nuisance parameters across time and cross section dimensions. De Wachter et al. (2007) developed an instrumental variable estimator based on a PUR test using the following model.

$$y_{i,t} = (\eta_i - z_{i,0})(1-\rho) + \rho y_{i,t-1} + u_{i,t}, \; u_{i,t} = v_{i,t} + \theta v_{i,t-1}, \quad i = 1,2,\ldots.N, t = 1,2,\ldots\ldots T.$$

where $u_{i,t}$ error terms with zero mean are dependent process. Moreover, $u_{i,t}$ shows a homogenous moving average process with $v_{i,t}$ innovations as independently and identically distributed having mean zero and variance constant. DWH test statistics is defined as follows.

$$\tau_2 = \frac{T\sqrt{N}}{\sqrt{2}}(\hat{\rho}_{IV} - 1) \xrightarrow{L} N(0,1)$$

where $\hat{\rho}_{IV} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T-p-1} y_{i,t} y_{i,t+p} \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T-p-1} y_{i,t} y_{i,t+p+1} \right)$ is instrumental variable estimator.

In their study, they concluded that near the unit root against an extensive range of alternatives DWH performs better with very high power when there is positive or zero correlation between error terms. Here, the simplest condition of no correlation is taken.

### 2.2.1.12. Sul (RMA) Test

Recursive Mean Adjustment (RMA) PUR tests are developed by Sul in 2009 to model the cross sectional dependence by applying recursive mean adjustment procedure. Bai and Ng (2004) test estimates the common factor using principal component method, however, Sul (2009) test is based on sequential procedure to handle factor structure estimation even in small time series and cross section length. Sul (2009) used the following model to test the hypothesis $H_0 : \rho_i = 1$ for all $i$ against $H_A : \rho_i < 1$ for all $i$.

Constant (M1): $$y_{it} = a_i + x_{it},\ x_{it} = \rho x_{it-1} + \varepsilon_{it} \tag{2.10}$$

Linear Trend (M2): $$y_{it} = a_i + b_i t + x_{it},\ x_{it} = \rho x_{it-1} + \varepsilon_{it} \tag{2.11}$$

Under the null hypothesis, Sul (2009) defined the following sequential regression estimations procedure corresponding to regression (2.10) and (2.11).

$$y_{it} - \hat{\rho} y_{it-1} = a_i + \sum_{j=1}^{p} \phi_{ij} \Delta y_{it-j} + e_{it}, \tag{2.10a}$$

$$y_{it} - \hat{\rho} y_{it-1} = a_i + \beta_i t + \sum_{j=1}^{p} \phi_{ij} \Delta y_{it-j} + e_{it} \tag{2.10b}$$

Further, Sul (2009) has introduced covariance matrix by supposing the mentioned entities from the core variables, that is, $y_t = (y'_{1t}, ...., y'_{Nt})'$, $c_{t-1} = (c'_{1t}, ...., c'_{Nt})$, $e_t = (e'_{1t}, ...., e'_{Nt})'$, and letting $c_{it-1} = \sum_{s=1}^{t} y_{is} \Big/ (t-1)$. Now covariance matrix is defined as $\hat{\Sigma} = \Lambda\Lambda'$, where $y_t^{+} = \Lambda y_t$, and $c_{t-1}^{+} = \Lambda c_{t-1}$. Hence,

$$y_{it}^{+} - c_{it-1}^{+} = \rho (y_{it-1}^{+} - c_{it-1}^{+}) + \sum_{j=1}^{p} \phi_{ij} \Delta y_{it-j}^{+} + e_{it}^{+} \tag{2.10c}$$

$$y_{it}^{+} - 2c_{it-1}^{+} = \beta_i + \rho(y_{it-1}^{+} - 2c_{it-1}^{+}) + \sum_{j=1}^{p} \phi_{ij} \Delta y_{it-j}^{+} + e_{it}^{+} \tag{2.10d}$$

where $y_{it}^{+}$, $c_{it-1}^{+}$, and $e_{it}^{+}$ represent the ith components of $y_t$, $c_{t-1}$, and $e_t$, respectively. Then, from equation (2.10c) and (2.10d), Sul (2009) proposed Panel Feasible Generalized Least Square combined with Recursive Mean Adjustment (PFGLS-RMA) test for point estimates $\hat{\rho}^{rc}$ and $\hat{\rho}^{rr}$ as follwos.

$$t^{rc} = \frac{\hat{\rho}^{rc}}{\sqrt{V(\hat{\rho}^{rc})}} \text{ and } t^{rr} = \frac{\hat{\rho}^{rr}}{\sqrt{V(\hat{\rho}^{rr})}}$$

Sul (2009) concluded that RMA test is under sized whether T and N is small or large, but this gains power by utilizing recursive mean adjustment procedure over BN test.

**2.2.1.13. Lee and Wu (LW) Test**

LW test is proposed by Lee and Wu (2012), which handles changes in the smooth structure of deterministic terms accompanied with cross sectional dependence among variables. The proposed test statistic is constructed from the breaks and cross-sectional dependence ADF (hence called BCADF) regression procedure introduced by Pesaran (2007) to incorporate a single frequency component Fourier function which is used to identify multiple structure break. They used the following model to develop their test

$$\Delta y_{it} = \beta_i \left( y_{i,t-1} - \alpha_i' d_{t-1} \right) + \alpha_i' \Delta d_t + \gamma_i f_t + \varepsilon_{it},$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $d_t = \left(1, \sin(2\pi kt/T), \cos(2\pi kt/T), t\right)'$ is a $4\times1$ deterministic common effect, $\Delta d_t = \left(0, \Delta\sin(2\pi kt/T), \Delta\cos(2\pi kt/T), 1\right)'$, $\beta_i = -(1-\rho_i)$, and $\alpha_i = \left(\mu_i, \alpha_{i,1}, \alpha_{i,2}, \varsigma_i\right)'$. $f_t$ is a common factor and $\gamma_i$ is the associated factor loading.

Frequency component, k, reflecting the number of homogenous cycles in the sample period. The unit-root hypothesis, $\rho_i = 1$ for above model can be expressed as:

$$H_0 : \beta_i = 0 \;\; \forall i, \text{ against } H_1 : \beta_i < 0, \;\; i = 1,2,....,N; \;\; \beta_i = 0, \; i = N_1+1, N_1+2,....,N.$$

To carry on testing procedure the following ADF regression has been performed for above model.

$$\Delta y_{it} = c_{i,0} + c_{i,1}\Delta \sin(2\pi kt/T) + c_{i,2}\Delta\cos(2\pi kt/T) + c_{i,3}\bar{y}_{t-1} + c_{i,4}\Delta\bar{y}_t + b_i y_{i,t-1} + e_{it}$$

and suggested the following test statistic, by following the procedure of Pesaran (2007), based on individual t-statistics of the above regression

$$LW = BCIPS^*(N,T) = \frac{1}{N}\sum_{i=1}^{N} t^*_i(N,T)$$

where

$$t_i^*(N,T) = t_i(N,T), \quad \text{if } -M_1 < t_i(N,T) < M_2$$

$$t_i^*(N,T) = -M_1, \quad \text{if } t_i(N,T) \le -M_1$$

$$t_i^*(N,T) = M_2, \quad \text{if } t_i(N,T) \ge M_2$$

$M_1$ and $M_2$ are sufficiently large positive constants, so that $\Pr\left[-M_1 < t_i(N,T) < M_2\right]$ is also sufficiently large. Values of $M_1$ and $M_2$ have been obtained from simulation results and are corresponding to CIPS_star test of Pesaran (2007). Also,

$$t_i(N,T) = \frac{\Delta y_i' \bar{M}_z y_{i,-1}}{\hat{\sigma}_i \left(y_{i,-1}' M_z y_{i,-1}\right)^{1/2}}, \quad \hat{\sigma}_i^2 = (T-6)^{-1}\,\Delta y_i' M_{i,z} \Delta y_i, \quad M_z = I_t - Z(Z'Z)^{-1} Z',$$

$M_{i,z} = I_T - G_i(G_i'G_i)^{-1} G_i'$ and $G_i = (Z, y_{i,-1})$, $Z = (\Delta\bar{y}, \tau, \Upsilon_1, \Upsilon_2, \bar{y}_{-1})$, $\tau = (1,1,......,1)'$,

$\Upsilon_1 = (\Delta\sin(2\pi k1/T),......,\Delta\sin(2\pi kT/T))'$ and $\Upsilon_2 = (\Delta\cos(2\pi k1/T),........,\Delta\cos(2\pi kT/T))'$.

### 2.2.1.14. Demetrescu and Hanck (DH) Test

Demetrescu and Hanck (2012) contributed in the PUR tests and introduced an instrumental variable based and robust to nonstationary error volatility PUR test. Demetrescu and Hanck (2012) used the following model to test $H_0: \rho_i = 1$ vs $H_1: |\rho_i| < 1$.

$$y_{it} = \mu_i + x_{i,t} \text{ with } x_{i,t} = \rho_i x_{i,t-1} + u_{i,t}$$

where $x_{i,0}$ are fixed and $u_{i,t} = \sum_{j=1}^{p} a_{i,j} u_{i,t-1} + \varepsilon_{i,t}$. Majority of instrumental variable based PUR tests use orthogonalizaton nonlinear instrumental variable procedure to utilize PUR test but DH test avoids to do so and considers the robust panel standard errors which is based on pooled instrumental variable estimator. The instrumental pooled estimator and its robust standard error is defined as

$$\hat{\rho} = 1 + \frac{\sum_i \sum_t \operatorname{sgn}(\tilde{y}_{i,t-1}) \bar{\varepsilon}_{i,t}}{\sum_i \sum_t |\tilde{y}_{i,t-1}|} \text{ and}$$

$$\sigma_{\hat{\rho}} = \left( \frac{\sqrt{\sum_i \sum_t \operatorname{sgn}(\tilde{\mathbf{y}}_{t-1})' \bar{\boldsymbol{\varepsilon}}_t \bar{\boldsymbol{\varepsilon}}_t' \operatorname{sgn}(\tilde{\mathbf{y}}_{t-1})}}{\sum_i \sum_t |\tilde{y}_{i,t-1}|} \right), \text{ respectively.}$$

Accordingly, DH test on the basis of above estimator and its standard error is defined as

$$t_{IV} = \frac{\hat{\rho} - 1}{\sigma_{\hat{\rho}}}$$

$$t_{IV} = \frac{\sum_i \sum_t \operatorname{sgn}(\tilde{y}_{i,t-1}) \bar{\varepsilon}_{i,t}}{\sqrt{\sum_t \operatorname{sgn}(\tilde{\mathbf{y}}_{t-1})' \bar{\boldsymbol{\varepsilon}}_t \bar{\boldsymbol{\varepsilon}}_t' \operatorname{sgn}(\tilde{\mathbf{y}}_{t-1})}}$$

where $\tilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{t-1}\sum_{j=1}^{t-1} y_{i,j-1}$ is the recursive demeaned series and $\bar{\varepsilon}_{i,t} = \Delta y_{i,t-j} - \sum_{j=1}^{p} \hat{a}_{i,j} \Delta y_{i,t-j}$. Demetrescu and Hanck (2012) dervied approximated asymptotic distribution of their test statistic and found convergence towards standard normal probabailty distribution.

### 2.2.1.15. Westerlund (WT) Test

Westerlund (2014) developed a robust panel test in the presence of uncorrelated or cross-correlated and unconditionally heteroscedastic errors. In the time series literature, this serial and heteroscedastic problem has been addressed through the estimation of nonparametric and resampling. Westerlund (2014) assumed heteroscedasticity in the deterministic and stochastic terms to develop his PUR test statistic. In this test statistic, no prior evidence is required about the structure of the heteroscedasticity. Westerlund (2014) used the following model to develop his test statistic to test $H_0 : \rho_i = 1$ vs $H_1 : |\rho_i| < 1$

$$y_{it} = \alpha_i + u_{it} ,$$

$$u_{it} = \rho_i u_{it-1} + v_{it} ,$$

$$\phi_i(L) v_{it} = e_{it} ,\ \ e_{it} = \Theta_i' F_t + \in_{it} ,\ \ \in_{it} = \sigma_{mi} \varepsilon_{it}$$

where $\phi_i(L) = 1 - \sum_{k=1}^{p_i} \phi_{ki} L^k$ and $F_t$ shows the common factor vector with r-dimension. Moreover, $\varepsilon_{it}$ is independently and identically distributed across both dimensions having mean zero and variance one. Also, $F_t$ is supposed as independently and identically distributed but only across time series level with $E(F_t)=0$ and $E(F_tF_t')=\Sigma_t$. However, both of terms are mutually independent across time and cross section dimensions. The following procedure is used to develop robust PUR test.

First, residuals denoted by $\Delta\hat{R}_{i,t}$ are calculated from the base model by regressing $\Delta y_{i,t}$ on $\Delta y_{i,t-1},\dots,\Delta y_{i,t-p_i}$. Second, a principal components method is used to get $\hat{F}_t$ and $\hat{\Theta}_i$. Third, variance is calculated as $\hat{w}_{m,t}=1/\hat{\sigma}_{m,t}$, where $\hat{\sigma}_{m,t}=\sum_{i=N_{m-1}+1}^{N_m}\hat{\epsilon}_{i,t}^2/(N_m-N_{m-1})$. Lastly, the desired test statistics is computed as

$$\tau=\frac{\left(\sum_{i=1}^{N}\sum_{t=p_i+2}^{T}\Delta\hat{R}_{wi,t}\hat{R}_{wi,t-1}\right)^2}{\sum_{i=1}^{N}\sum_{t=p_i+2}^{T}\hat{R}_{wi,t-1}^2}$$

where $\Delta\hat{R}_{wi,t}=\hat{w}_{mt}\,\hat{\epsilon}_{it}$ and above test statistic is asymptotically distributed as Chi-square distribution. In simulation study, Westerlund (2014) concluded that WT has smaller size distortion and have better as compared to DH test of Demetrescu and Hanck (2012).

### 2.2.1.16. Westerlund and Larsson (WL) Test

Westerlund and Larsson (2012) suggested a heterogeneous PUR test by adopting AR coefficient to be random across the cross sections. Here, AR term is independently and identically distributed with mean zero and variance one to test the joint null hypothesis of mean and variance related to AR coefficient. According to Westerlund and Larsson (2012), a similar situation of joint restrictive null and alternative hypothesis is not adopted in the previous PUR tests. Further, they addressed that using random AR, parameter of interest, leads to efficiency and reduces the number of parameters to be estimated. They used the following model to develop their test statistic.

$$y_{it} = d_{it} + z_{it},$$

$$z_{it} = \rho_i z_{i,t-1} + u_{it},$$

$$\phi(L)u_{it} = \xi_{it},$$

where $\rho_i \sim iid(\mu_\rho, \sigma_\rho^2)$, the following procedure has been adopted by Westerlund and Larsson (2012) to compute WL test statistic. First, define $w_{it}$ using base regression of $z_{it}$ as

$$w_{it} = \phi_i(L)(y_{it} - \alpha_i^0 - \beta_i^0(t-p)),$$

$$w_{it} = y_{it} - \Phi_i^{'}\mathbf{y}_{it} - \alpha_i - \beta_i^0\phi_i(L)(t-p),$$

then $$\Delta w_{it} = \phi_i(L)(\Delta y_{it} - \beta_i^0) = \Delta y_{it} - \Phi_i^{'}\Delta\mathbf{y}_{it} - \beta_i$$

where $\mathbf{y}_{it} = (y_{it-1}, \ldots\ldots, y_{it-p})^{'}$, $\Phi_i = (\phi_{1i}, \ldots\ldots, \phi_{pi})^{'}$. After some mathematical manipulation of all the given entities the following difference error term has been obtained,

$$\Delta e_{it} = (\rho_i - 1)e_{i,t-1} + \varepsilon_{it}, \quad \hat{w}_{it} = y_{it} - \hat{\Phi}_i^{'}\mathbf{y}_{it} - \hat{\alpha}_i$$

where $\hat{e}_{it} = \frac{1}{\sigma_{\xi_i}}\hat{w}_{it}$, $\varepsilon_{it} = \frac{1}{\sigma_{\xi_i}}\xi_{it}$, $\hat{\sigma}_{\xi_i}^2 = \frac{1}{T-p-1}\sum_{t=p+2}^{T}(\Delta\hat{w}_{it})^2$ and $\hat{k} = \frac{1}{N(T-p-1)}\sum_{i=1}^{N}\sum_{t=p+2}^{T}(\Delta\hat{e}_{it})^4$.

WL test statistic, also called LM type test, based on maximum likelihood estimation is defined as follows:

$$LM_{\mu,\sigma^2} = \frac{A_{NT}^2}{B_{NT}} + \frac{12}{5(\hat{k}-1)} \frac{C_{NT}^2}{D_{NT}},$$

where $$A_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta\hat{e}_{it}\hat{e}_{i,t-1},$$

$$B_{NT} = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \hat{e}_{i,t-1}^2,$$

$$C_{NT} = \frac{1}{\sqrt{N}T^{3/2}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \left((\Delta\hat{e}_{it})^2 - 1\right)\hat{e}_{i,t-1}^2,$$

$$D_{NT} = \frac{1}{\sqrt{N}T^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (\Delta\hat{e}_{it})^2 \hat{e}_{i,t-1}^4.$$

### 2.2.2. Panel Unit Root Tests with Null Hypothesis of Stationarity

In this section of literature review, a brief review of almost all (six tests) PUR tests having the null hypothesis of stationarity has been discussed. 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. Also, all of them are right tail tests according to their hypothesis of interest and referred as Langrange multiplier (LM) statistics. These tests are elaborated below in the ascending order according the year of publication.

#### 2.2.2.1. Hadri (HD) Test

Hadri (2000) proposed a residual-based Lagrange multiplier (LM), which is a panel extension of Kwiatkowski et al. (1992) test (KPSS 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. Hadri (2000) used the following model to develop his test,

$$y_{it} = r_{it} + \varepsilon_{it} \tag{2.12}$$

and

$$y_{it} = r_{it} + \beta_i t + \varepsilon_{it} \tag{2.13}$$

$$r_{it} = r_{i,t-1} + u_{it} \tag{2.14}$$

where $t = 1,......,T$ and $i = 1,........,N$, $\varepsilon_{it}$ and $u_{it}$ are mutually and independent normals and iid across cross sections and over time with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$, $E(u_{it}) = 0$, and $E(u_{it}^2) = \sigma_u^2$. Given the residuals $\varepsilon_{it}$ from the individual regressions, the LM statistic, labelled as *HD* henceforth is defined for the $H_0 : \lambda = \frac{\sigma_u^2}{\sigma_\varepsilon^2} = 0$ against $H_A : \lambda > 0$, as

$$HD = \frac{\sqrt{N}\left(LM - \xi\right)}{\zeta}$$

where

$$LM = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{S_{it}^2}{\hat{\sigma}_{\varepsilon i}^2},$$

$$\hat{\sigma}_{\varepsilon i}^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2}{NT}, \quad S_{it} = \sum_{s=1}^{t} \hat{\varepsilon}_{is}$$

LM is basically the average of KPSS test in the panel structure. $S_{it}$ are the cumulative sums of residuals and $\hat{\sigma}_{\varepsilon i}^2$ is defined as the variance of $\varepsilon_{it}$. $\xi$ and $\zeta$ denote the mean and standard deviation taking different values in the presence of drift term only and, drift and trend terms both. If there is only drift term in the model then $\xi = 1/6$ and $\zeta^2 = 1/45$, but if both drift and trend terms are present then $\xi = 1/15$ and $\zeta^2 = 11/6300$. HD test is asymptotically converges to standard normal distribution. Hadri (2000) concluded that the empirical size of the test statistic is very close to nominal size (i.e. 5%) when both time series and cross section dimensions are large enough.

### 2.2.2.2. Hadri and Larsson (HL) Test

Hadri and Larsson (2005) extended the Hadri (2000) test for the fixed T which expands the finite sample performance of test statistic under the assumption of cross sectional independence. The main difference between Hadri (2000) and Hadri and Larsson (2005) test is derivation of exact finite sample mean and variance. Hadri and Larsson (2005) computed the exact means and variance for model (2.12) and model (2.13). This test has the same mathematical procedure as Hadri (2000) does have. Their test statistic is abbreviated as, *HL* and defined as,

$$HL_i = \frac{\sum_{t=1}^{T} S_{iT}^2}{T^2 \hat{\sigma}_{\varepsilon i}^2}$$

While its asymptotic distribution converges to standard normal but with different exact finite sample mean and variance, which makes the main difference from the Hadri (2000) test, is

$$Z_{HL} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{HL_i - E(HL_i)}{\sqrt{\mathrm{var}(HL_i)}} \right) \Rightarrow N(0,1)$$

With $E(HL_i) = (T+1)/6T$, $\text{var}(HL_i) = (T^2+1)/20T^2 - ((T+1)/6T)^2$ for model (12) and with $E(HL_i) = (T+2)/15T$, $\text{var}(HL_i) = (T+2)(13T+23)/2100T^3 - ((T+2)/15T)^2$ for model (2.13). Hadri and Larsson (2005) have shown from their simulations results that these moments are $E(HL_i)$=1/6 and $\text{var}(HL_i)$=1/20 in the presence of drift term only in the given model. Similarly, if both of the deterministic terms are present in the model then these moments take value of $E(HL_i)$=1/15 and $\text{var}(HL_i)$=13/2100 according to their simulations, in both asymptotic results $T \to \infty$. HL test is also applicable both for balance and unbalance panels. Hadri and Larsson (2005) discussed through simulation results that with asymptotic T assumption leads to size distortion and can only be controlled if T is kept fixed. Also, HL test has better power as compared to KPSS test of time series for the same null hypothesis of interest.

#### 2.2.2.3. Harris, Leybourne and McCabe (HLM) Test

Harris et al. (2005) suggested a new nonparametric panel stationarity test which is robust in the presence of serial dependence and cross-sectional dependence across the panel. HLM test avoid to fit the individual series separately and treat the short time series. This test statistic is simple to use and uses standard normal critical values to separate acceptance and rejection region. It can be applied for approximate factor model as well. Harris et al. (2005) used the following model to develop their test statistic.

$$y_{it} = \beta_i' x_{it} + z_{i,t}, \; i = 1,\ldots\ldots,N, \;\; t = 1,\ldots\ldots,T$$

$$z_{it} = \phi_i z_{it-1} + \xi_{it}$$

where disturbance term has mean zero and, $\xi_{it}$ and $\xi_{jt}$ may be correlated. They were interested to test the null hypothesis of joint stationarity against the alternative of unit root, that is

$$H_0 : |\phi_i| < 1 \text{ against } H_A : |\phi_i| = 1$$

They assumed that $\xi_t = (\xi_{1,t}, \ldots\ldots, \xi_{N,t})'$ is a linear stationary process to permit cross sectional correlation in the form of either contemporaneous or cross sectional in the series of panel. An OLS residuals of above regression is obtained as $\hat{z}_{i,t} = y_{it} - \hat{\beta}_i' x_{it}$ then these residuals are standardized as $\tilde{z}_{it} = \hat{\xi}_{it} / s_i$, here $s_i$ denotes the sample standard deviation. Then the test statistic is obtained as

$$\hat{S} = \frac{\tilde{C}_k}{\hat{w}\{\tilde{a}_{kt}\}},$$

where $\tilde{C}_k = \sum_{i=1}^{N} \tilde{C}_{ik}$, $\tilde{C}_{ik} = T^{-1/2} \sum_{t=k+1}^{T} \tilde{z}_{it} \tilde{z}_{i,t-k}$, and $\hat{w}^2\{\tilde{a}_{kt}\} = \hat{\gamma}_0\{a_{kt}\} + 2\sum_{j=1}^{l}\left(1 - \frac{j}{l+1}\right)\hat{\gamma}_j\{a_{kt}\}$, is the long-run variance estimator where $\hat{\gamma}_j\{a_{kt}\} = T^{-1} \sum_{t=j+k+1}^{T} a_{kt} a_{k,t-j}$, $a_{kt} = \sum_{i=1}^{N} \tilde{z}_{it} \tilde{z}_{i,t-k}$.

According to asymptotic theory, this test statistic converges to standard normal distribution under the null hypothesis of joint stationarity while under the alternative hypothesis of unit root this test statistic diverges to positive infinity. This test statistic is abbreviated as HLM in this study to evaluate its size and power properties.

#### 2.2.2.4. Shin and Snell (SS) Test

Shin and Snell (2006) proposed a panel-based mean group test for the null of stationarity against the alternative of unit roots in the occurrence of both serial correlation across time periods and heterogeneity across cross-section units. Using models given in equation (12) and (13) and under the null hypothesis $H_0 : \sigma^2_{v_1} = ... = \sigma^2_{v_N} = 0$ against $H_A : \sigma^2_{v_1} > 0$ with the same estimation method that we have observed for HD and HL tests, they developed the following test statistic.

$$SS = \tau_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\eta_{iT} - \mu) \times \frac{1}{w^{\delta} \hat{w}_N^{1-\delta}}, \quad 0 \le \delta \le 1$$

where

$$\hat{w}_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\eta_{iT} - \overline{\eta}_T)^2,$$

$$\overline{\eta}_T = \frac{1}{N} \sum_{i=1}^{N} \eta_{iT},$$

$$\eta_{iT} = \frac{T^{-1} \sum_{t=1}^{T} S_{iT}}{\hat{\sigma}^2_{\varepsilon i}}, \quad \mu = \frac{1}{6} \text{ and } S_{iT} = \sum_{s=1}^{t} \hat{\varepsilon}_{is}.$$

Shin and Snell (2006) have taken $\delta = 0.5$ in their study. This test statistic converges to standard normal distribution under null hypothesis while diverges to positive infinity when both N and T tend towards infinity under joint asymptotic theory. Shin and Snell (2006) have concluded that their test statistic performs very well when time series dimension is small when errors are uncorrelated. But situation becomes worse with over size problem when errors are correlated and time series level is small.

### 2.2.2.5. Hadri and Kurozumi (KK) Test

Hadri and Kurozumi (2009) introduced a simple PUR test for the null hypothesis of stationarity in heterogeneous panel data with cross-sectional dependence in the form of a common factor in the error term of the model. This common factor is estimated by Hadri and Kurozumi (2009) by using the procedure of Pesaran (2007). Their test statistic is same as the KPSS test is in univariate time series literature to test the null hypothesis of stationaity against alternatve hypothesis of unit root but with augumented cross sectional averag. Further, Hadri and Kurozumi (2009) test uses the same model that has been used in HD and HL test but with common factor errors. They used the following model to present their test statistic.

$$y_{it} = z_t'\delta_i + r_{it} + u_{it}, \quad r_{it} = r_{it-1} + v_{it}, \quad u_{it} = f_t\gamma_i + \varepsilon_{it}$$

where $f_t$ denotes the unobserved common factor. The estimation test procedure of this test is same that have been developed and adopted for HD and HL test but to eliminate the effect of common factor Hadri and Kurozumi (2009) used Pesaran (2007) method of cross-sectional average of the model.

$$\bar{y}_t = z_t'\bar{\delta} + \bar{r}_t + f_t\bar{\gamma} + \bar{\varepsilon}_t,$$

By assuming $\bar{\gamma} \neq 0$, we have $$f_t = \frac{1}{\bar{\gamma}}\left(\bar{y}_t - z_t'\bar{\delta} - \bar{r}_t - \bar{\varepsilon}_t\right)$$

Hence $$y_{it} = z_t'\tilde{\delta}_i + \tilde{\gamma}_i\bar{y}_t + \epsilon_{it}$$

and introduced their test statistic.

$$KK = Z_A = \frac{\sqrt{N}\left(\overline{ST} - \xi\right)}{\varsigma}$$

to test $$H_0 : \rho = \frac{\sigma_v^2}{\sigma_\varepsilon^2} = 0 \text{ vs } H_A : \rho \geq 0$$

where $\overline{ST} = \frac{1}{N}\sum_{i=1}^{N} ST_i$, $ST_i = \frac{1}{\sigma_\varepsilon^2 T^2}\sum_{t=1}^{T}\left(S_{it}^{u}\right)^2$, $S_{it}^{u} = \sum_{s=1}^{t} \hat{\epsilon}_{is}$, $\tilde{\delta}_i = \delta_i - \tilde{\gamma}_i\overline{\delta}$, $\tilde{\gamma}_i = \gamma_i / \overline{\gamma}$ and $\epsilon_{it} = r_{it} - \tilde{\gamma}_i \overline{r}_t + \varepsilon_{it} - \tilde{\gamma}_i \overline{\varepsilon}_t$. $\xi$ and $\varsigma$ denote the mean and standard deviation. If there is constant in the model then $\xi = 1/6$ and $\varsigma^2 = 1/45$, and if both constant and time trend both are present in the model then $\xi = 1/15$ and $\varsigma^2 = 11/6300$. The limiting distribution of this test statistic is same as have been defined for HD and HL tests. Here after, this test is denoted as KK test in this simulation study.

#### 2.2.2.6. Demetrescu, Hassler and Tarcolea (DHT) Test

Demetrescu et al. (2010) proposed their test for large cross sectional dimension in the presence of cross unit correlation. They proposed an unbounded norm of the long-run correlation matrix of the panel is used to allow for persistent cross correlation. They used equation (2.12) and (2.13) as models to develop their test statistic, they used the following demean or de-trend procedure recursively for estimation.

$$y_{i,t}^{\mu} = y_{i,t} - \frac{1}{t}\sum_{i=1}^{t} y_{i,j}$$

$$y_{i,t}^{\tau} = y_{i,t} + \frac{2}{t}\sum_{i=1}^{t} y_{i,j} - \frac{6}{t(t+1)}\sum_{j=1}^{t} j y_{i,j}$$

and proposed test statistics, $\tilde{k}^{\mu} = \frac{1}{NT^2 \hat{\rho}^{\mu}} \sum_{t=1}^{T} S_t' S_t - \frac{1}{6} \frac{1-\hat{\rho}^{\mu}}{\hat{\rho}^{\mu}}$ and

$$\tilde{k}^{\tau} = \frac{1}{NT^2 \hat{\rho}^{\tau}} \sum_{t=1}^{T} S_t' S_t - \frac{1}{6} \frac{1-\hat{\rho}^{\tau}}{\hat{\rho}^{\tau}}$$

for equation (2.12) and equation (2.13) respectively.

where $$\hat{\rho}^{\mu} = \max\left(N^{-0.5}, 1-q^{\mu}\right) + 0.2\sqrt{\frac{2}{N-1}}\left(1-\max\left(N^{-0.5}, 1-q^{\mu}\right)\right)$$

$$\hat{\rho}^{\tau} = \max\left(N^{-0.5}, 1-q^{\tau}\right) + 0.2\sqrt{\frac{2}{N-1}}\left(1-\max\left(N^{-0.5}, 1-q^{\tau}\right)\right)$$

$$q^{\mu} = \frac{1}{(N-1)T} \sum_{i=1}^{N} \left(S^{\mu}_{i,T} - \bar{S}^{\mu}_{T}\right)^2, \quad q^{\tau} = \frac{1}{(N-1)T} \sum_{i=1}^{N} \left(S^{\tau}_{i,T} - \bar{S}^{\tau}_{T}\right)^2$$

$S_t$ Constructed from data demeaned or de-trended series. Here, $\hat{\rho}$ denotes the consistent estimator. In this study, this test statistic is abbreviated as DHT test.

## 2.3. Comparison Studies of Panel Unit Root Tests

Many authors developed PUR tests for two main types. One type of PUR tests have been developed to test the null hypothesis of PUR while the second type of tests have been developed for the null hypothesis of stationarity. Second type of tests are also called residual based PUR tests. Authors developed their own PUR test relying on different features of the underlying panel data. These tests are then compared and evaluated on the basis of their size and power properties. Many comparative studies have been carried out to judge the performance of panel tests but these comparison create no ultimate answer that which test is performing better.

Maddala and Wu (1999) conducted a simulation study, in which, they compared three type of PUR tests using two types of experiments. They used a single alternative to compare the power of three PUR tests. In the first experiment, a DGP for the general set up of the null of unit root and alternative of stationary has been used to compare tests. In the second experiment, the null hypothesis remains the same but in alternative a mixture of unit root and stationary has been taken to analyze size and power properties of tests under investigation. For both of the experiments, a serial correlation of errors has also been considered.

For the first experiment, both in case of constant only and with constant and trend, Fisher test (MW) with less size distortion and better power is identified as better performer as compared to IPS and LLC tests. In the second experiment, Maddala and Wu (1999) showed that IPS test is slightly more powerful than Fisher (MW) and LLC test when there is no cross sectional correlation in the errors. At this stage, Fisher test has better power as compared to LLC test. All tests perform well in case of no serial correlation for all samples. But, when errors are correlated for different samples (or cross sectional units) then all the tests (MW, IPS, and LLC) perform very poorly. However, the results suggested that MW test less effected in the presence of cross sectional correlation of error terms as compared to the IPS and LLC tests. They also showed through their simulations results that size distortion with MW test is small when T (time dimension) is large but N (cross sectional dimension) is not very large as compared to IPS and LLC tests. But for medium values of T and large N, the size distortion of the MW test becomes savior. Maddala and Wu (1999) results show that MW test performs much better for the cases when there is a mixture of stationary and nonstationary series in the group as alternative hypothesis. Overall, Maddala and Wu (1999) reported that MW test stands best among the other two tests (IPS and LLC).

Gutierrez (2006) evaluated the performance of four cross sectional correlated panels tests through simulation study. Gutierrez (2006) compared Choi's (2006), Bai and Ng's (2004), Moon and Perron's (2004), and Phillips and Sul's (2003) tests using the size and power properties of these tests. Gutierrez (2006) compared these tests, for small sample sizes using different values of cross section (N) and time (T) dimension and for different number of factors, under a common structure rather than utilizing different methods or procedures to identify idiosyncratic and common components.

Gutierrez (2006) investigated the power comparision of these tests under a single alternative. Gutierrez (2006) concluded results based on by varying the common and idiosyncratic components in the DGP. When equal importance is given to common and idiosyncratic components in the DGP then Gutierrez (2006) noticed that for large T and N all the four tests have good stable size except Choi's (2006) test. However, for small values of N a strong size distortion observed. In case of cross sectional dependence in the DGP, Moon and Perron's (2004) and Bai and Ng's (2004) tests suffer from oversized problems, this distortion become more savior for small N. Size and power of Moon and Perron's (2004) test for different values of T and N and model specification performs well as compared to other tests. Gutierrez (2006) simulation results show that Bai and Ng's (2004) test based on pooled GLS procedure gains much better power as compared to the pooled OLS based procedure. Choi's (2006) test is highly oversized when there is cross section unit heterogeneity under the common factors influences. Lastly, Gutierrez (2006) observed that in the presence of deterministic term in the included process, all the tests have a lack of power. The study of Gutierrez (2006) reported a mixture results for all the situations under consideration.

Hlouskova and Wagner (2006) carried out an extensive and comprehensive study in the PUR literature in which they compared PUR tests both having the null of unit root and null of stationarity using a single DGP and asymptotic critical values through simulations. Their study is the only one study in which, both categories of the tests have been compared. Hlouskova and Wagner (2006) used four PUR tests having the null of unit root, Levin, Lin, and Chu (2002), Breitung (1999), Im, Pesaran, and Shin (2003), Maddala and Wu (1999), and two PUR tests having the null of stationarity, Hadri's (2000) and Hadri and Larsson's (2005) tests to carry on their simulation study.

Hlouskova and Wagner (2006) performed simulation study by addressing four aspects. First, they evaluated the performance of the tests by varying time series and cross section dimensions. Second, for a moving average roots having values (0.2, 0.4, 0.8, 0.9, 0.95, 0.99) tending to 1 has been taken to assess the performance of the tests. Third, a first order autoregressive coefficient having values of (0.7, 0.8, 0.9, 0.95, 0.99) has been taken to evaluate the power performance of PUR tests having the null hypothesis of unit root, and AR(1) coefficient having values of (0, 0.1, 0.2, 0.3, 0.4, 0.5) is used to analyze the size performance of panel tests having the null hypothesis of stationary. Lastly, with two form of deterministic terms (with intercept, and with intercept and trend) the investigation of panel tests have been made. They observed that Levin, Lin, and Chu's (2002) and Breitung's (1999) tests gain better powers as comapred to other tests having the null of unit root when there is intercept under stationary series. Moreover, with short and uncorrelated panels Levin, Lin, and Chu's (2002) test has substaintianl improvement over all other tests.

Maddala and Wu's (1999) test with varaition of panels does not show better performance. They observed that Levin, Lin, and Chu's (2002) and Breitung's (1999) tests have the minumum size distirtions over all the situations. Hlouskova and Wagner (2006) noticed that the tests of Maddala and Wu's (1999) and Im, Pesaran, and Shin's (2003) tests perform very poorly. Their size and power are not much poor as compared to other tests having the null of unit root. Besides, Hlouskova and Wagner (2006) observed that Hadri's (2000) and Hadri and Larsson's (2005) tests having the null of stationarity perform very poorly. Hadri (2000)'s and Hadri and Larsson's (2005) tests tend to reject statioanrity most of the times even for highly stationary series.

Pesaran (2007) conducted a simulation study based on Cross-sectionally Augmented Dickey Fuller (CADF) regressions with the first difference of each cross section and cross section averages of lags. Pesaran (2007) used DGP with cross section dependence and fixed effects for dynamic panels to compare CIPS, MW, CHO tests and truncated version of t-bar test denoted by (CIPS_star) using CADF regressions. Pesaran (2007) observed that all tests based on CADF methodology show satisfactory power and size both for short and long T and N. However, CIPS and CHO tests perform better as compared to MW and CIPS_star tests in the case of with serial correlation, but moderate, and without serial correlation. In the case of linear trend in the model, power of these tests crucially depends upon time dimension and cross section dimension and power get rises when T is greater than 30. Overall, Pesaran (2007) concluded that tests based on CADF regressions perform better as compared to standard regressions on which these tests have been developed for cross sectional dependence panels.

Gengenbach (2009) used DGPs of Bai and Ng (2004) based PUR test to conduct comparison among CIPS, BNG, RMA, MP, and BU PUR tests. Their results are based on asymptotic critical values and they have used Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to select the leg length. All tests have serious size distortions for almost all cases when time dimension is small and in the presence of serial correlation. CIPS and BNG tests are more powerful than the rests. In the case of single factor model, CIPS test has better size and power properties for large N and T. However, in the presence of two or more factors leads to serious size distortion and worse power.

Hoang and McNown (2006) investigated the size and power properties of IPS and MW tests. They have taken AR (1) DGP to compare both of the tests using two different estimation methods. Hoang and McNown (2006) performed their simulation study for IPS and MW tests using least square and weighted symmetric estimation methods. They concluded that IPS and MW tests have more power and stable size by using weighted symmetric method as compared to the DF and ADF based estimation method. IPS test based on weighted symmetric method for balance panel performs better as compared to MW estimated using the same method for different time dimensions and cross sectional dimensions. In the case of unbalanced panel, MW test performs better when T and N is small.

## 2.4. Concluding Remarks

All PUR tests introduced in the literature follow different models and assumptions and a comparative study is carried out to show the superiority of the newly introduced test using size and power properties. In the first section of this chapter, PUR tests (to be used in our study) having the null hypothesis of unit root and stationary have been discussed with their models, assumptions and mathematical structure. PUR tests having the null hypothesis of unit root are either residuals based, maximum likelihood based, p-value based or instrumental variable based according to their structure. However, all PUR tests having the null hypothesis of stationary are based on the residuals. In the second section of this chapter, previous comparative studies have been discussed with DGPs and results. According to existing literature, the following gaps have been identified in the comparative studies.

First, previous comparative studies have used asymptotic critical values to investigate the size and power properties of PUR tests. This causes size instability as a result produce less or over power gain of PUR tests. In this study, simulated critical value is used to make stabilize the size and make power comparison of PUR tests meaningful.

Second, previous comparative studies take only few PUR tests to evaluate the size and power properties. For example, Hlouskova and Wagner (2006) have compared six PUR tests out of which four tests are having the null hypothesis of unit root while two of the tests having the null hypothesis of stationary. While this study compares twenty four PUR tests in which eighteen tests belong to the null hypothesis of unit root while six tests belong to the null hypothesis of stationary.

Third, all comparison studies of the PUR tests have been made for one, two or very few alternatives of parameter of interest. In order to make further contribution in the literature, this study compares all the tests for all the alternatives in a whole space under investigation for both type of tests (PUR and stationary tests).

Lastly, the review of the previous comparative studies reveal no single significance technique to keep a bench mark for comparison of the PUR tests for both categories of the tests. A unified bench mark is needed to make comparison extensively and meaningful for both categories of the tests. In this study this technique is called as stringency criterion.

# Chapter 03

## Research Methodology

### 3.1. Methodology Description

To carry out the comparative analysis of PUR tests, this study investigates twenty four PUR tests both under the null hypothesis of unit root (eighteen tests) and null hypothesis of stationarity (six tests). This study makes comparison on the basis of size and power properties of each test under consideration. In this study a time dimension of T=10, 25, 50, 100 and cross sectional dimension of N=2, 4, 8, 16, 32 are considered for Monte Carlo Simulation experiments.

Since, both categories of tests are taken i.e. tests having the null hypothesis of unit root and tests under the null hypothesis of stationarity, for this purpose two Data Generating Processes (DGPs) are considered for this simulation study. On the basis of these DGPs, simulated critical values are calculated.

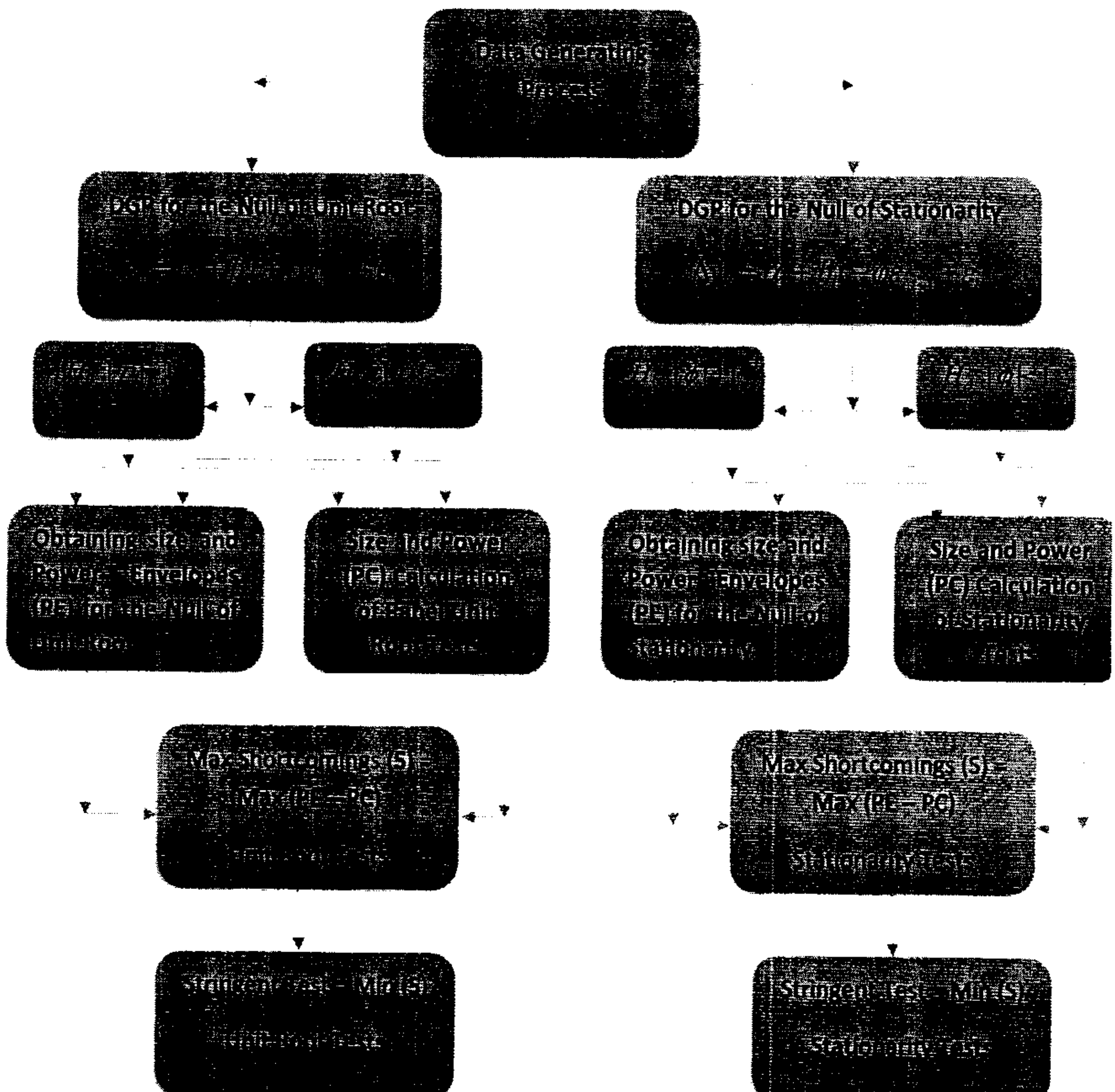
A detailed Monte Carlo simulation designs is explained to find power curves of PUR tests and power envelops of point optimal tests.

Stringency criterion, which is the main theme of our study, is used to compare all the tests under consideration. A stringency criterion takes the whole alternate space for the study under consideration. This part of our methodology explains how to find the most stringent tests for the given DGPs (i.e. one for the null hypothesis of unit root and the other for null hypothesis of stationarity).

## 3.2. Monte Carlo Simulation Design

In Monte Carlo simulation design of methodology section, DGPs, size, powers, power curves, power envelopes, maximum shortcomings (MSC) and most stringent tests have been discussed both for the null hypothesis of PUR and stationary. Graphically, general framework for this study is represented as follows.

**Figure 3.1: General Framework for the Simulation Study**



### 3.3. Specification of Data Generating Processes

In this section, two types of data generating processes are discussed, first type of DGP belongs to the null hypothesis of PUR versus alternative hypothesis of panel stationary while the other one belongs to the null hypothesis of panel stationary versus alternative hypothesis of PUR.

#### 3.3.1. DGP-A: Null Hypothesis of Panel Unit Root with Deterministic Term Specifications

Let $y_{it}$ denotes the panel series, where $i \in N$ and $t \in T$, panel series is generated using the following mathematical model.

$$y_{it} = \alpha_i + \beta_i t + \rho y_{i,t-1} + e_{it}$$

where $\alpha_i \sim N(0,1)$, $\beta_i \sim N(0,1)$, $e_{it} \sim N(0, \delta_{e_i})$ and error variance is defined as $\delta_{e_i} \sim U[0.5, 1.5]$ to generate a complete heterogeneous model. In this model, under the null hypothesis of unit root $\rho = 1$ and $0 \leq \rho < 1$ for the alternative hypothesis of stationarity. Here, this study takes $\rho = \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0\}$.

#### 3.3.2. DGP-B: Null Hypothesis of Stationarity with Deterministic Term Specifications

This study considers the following DGP for the null hypothesis of stationarity,

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

where $\alpha_i \sim U[0,10]$, $\beta_i \sim U[0,2]$, $\varepsilon_{it} \sim N\left(0,\delta_{\varepsilon_i}\right)$ and error variance is defined as $\delta_{\varepsilon_i} \sim U[0.5,1.5]$. This DGP has been used in Hwang and Schmidt (1993) for the null hypothesis of stationarity in time series literature. According to Hwang and Schmidt (1993), under the null hypothesis of stationarity $\varphi = 0.99999 \cong 1$ and under alterative hypothesis of unit root $0 \leq \varphi < 1$. In this study, $\varphi = \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0\}$ under alternative hypothesis.

## 3.4. Specification of Cross Section Length and Time Series Length

This research choose five different cross section units and categorize them into three different groups to proceed this study. These cross section units are 2, 4, 8, 16, and 32 in which first two represent small cross section units, cross section number third and fourth shows medium cross section unit and last cross section unit indicate the largest one.

Similarly, four time series length are taken to evaluate the performance of PUR tests, these time series level are 10, 25, 50, and 100. As similar to the categorization of cross section length into three types, a time series length of 10 indicate small time series, 25 and 50 are assigned as medium time series length while 100 is allotted as large time series length in this study.

## 3.5. Finding Simulated Critical Value of a Test

To find the simulated critical value of a given test (for both categories), the following simulation procedure is adopted:

First, data are generated according the given DGP under the null hypothesis. Secondly, given test statistic is calculated for the given DGP (i.e. DGP under the null hypothesis).

Thirdly, DGP and test statistics under the null hypothesis is repeated a fixed number of times which is called Monte Carlo Sample Size and given test statistic under consideration is sorted and saved in a column matrix. Lastly, simulated critical value is calculated depending upon nature (left tail test, right tail test or two tail test) of the test statistic. If the given test statistic is right tail test then $95^{th}$ percentiles of the column matrix is the desired simulated critical value. For the left tail test, $5^{th}$ percentiles is the required simulated critical value in the column matrix. While $2.5^{th}$ and $97.5^{th}$ percentiles are the appropriate simulated critical values in the column matrix for the given two tail test.

## 3.6. Finding Size of the Test

In this study, a test size is calculated using the following procedure:

First, data are generated a fixed number of time and cross section dimension under the null hypothesis. The number of time series and cross section dimensions have considered are 10, 25, 50, 100 and 2, 4, 8, 16, 32, respectively. Second, test statistic is calculated for the given time series and cross section dimensions and the number of significant value is counted if test statistic value is greater than simulated critical value. A repetition of this process is carried for fixed Monte Carlo sample size of 10,000 and size of test is obtained as

$$Size\ of\ Test = \frac{significant\ count}{MCSS} \times 100$$

## 3.7. Panel Log-Likelihood Ratio (Point Optimal) Tests

This section explains two point optimal tests under the null hypothesis of PUR and null hypothesis of stationarity. This study uses panel point optimal tests by using log likelihood ratio concept introduced by Neyman Pearson Lemma (Neyman and Pearson (1992)) to get most powerful test statistic at each alternative in entire sample space under consideration.

### 3.7.1. Point Optimal Test under the Null Hypothesis of Unit Root

Monn, Perron, and Phillips (2007) developed a PUR point optimal test for the null hypothesis of unit root. They follow the following model to introduce their test statistic,

$$Y_{it} = \pi G' + \rho Y_{i,t-1} + U_{it}$$

where $U_{it} \sim N(0, \sigma_i^2)$, $G = (G_0, G_1)$, $G_0 = (1, \ldots\ldots, 1)$, $G_1 = (1, 2, \ldots\ldots, T)$, and $\pi = (\alpha_i, \beta_i)$.

To test the hypothesis,

$$H_0 : \rho = 1 \text{ vs } H_A : |\rho| < 1$$

Monn, Perron, and Phillips (2007) find the following log likelihood function,

$$L(\rho) = -\frac{1}{2}[vec((Y_{it} - \rho Y_{it-1})' - G\pi')]' \Delta_\rho' (\Sigma^{-1} \otimes I_{T+1}) \Delta_\rho [vec((Y_{it} - \rho Y_{it-1})' - G\pi')],$$

where, $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots\ldots\ldots, \sigma_n^2)$,
$$\Delta_{\rho_{(T+1)\times(T+1)}} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & -\rho & 1 & 0 \\ 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}$$

According to Neyman Pearson Lemma (Neyman and Pearson (1992)), a point optimal test is the ratio of two log likelihood functions (i.e. log likelihood function under null to log likelihood function under alternative) hence Monn, Perron, and Phillips (2007) panel point optimal test statistics is as follows

$$PO = -2\left[L(\rho) - L(1)\right].$$

### 3.7.2. Point Optimal Test under the Null Hypothesis of Stationarity

This section introduce a point optimal test for the null hypothesis of panel stationary against alternative hypothesis of PUR using log likelihood ratio concept for DGP-B.

By definition of Neyman Pearson Lemma (Neyman and Pearson (1992)),

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

where log likelihood for DGP-B is,

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

$$L(1) = -\frac{NT}{2}\log(2\pi) - \frac{NT}{2}\log\left(\delta_{\varepsilon_i}^2\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it} - y_{i,t-1}\right)^2$$

Hence, Panel Point Optimal test for null hypothesis of stationarity is,

$$PPO = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{i,t} - y_{i,t-1}\right)^2 + \frac{1}{2}\log\left(1 + NT(1-\varphi)^2\right) + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{\left(y_{i,t} - y_{i,t-1}\right)^2}{\left(\delta_{\varepsilon_i}^2\left(1 + NT(1-\varphi)^2\right)\right)}$$

## 3.8. Finding Power Curve of the Tests

The following Monte Carlo simulation design is adopted to calculate and plot the power curves of the PUR tests for both the categories.

First, data are generated using DGP-A and DGP-B under a point alternative hypothesis and calculate test statistics for the generated data one by one. Here in this study a point alternative hypothesis for both type of tests lies between "0" and "0.9" (i.e. $\rho = \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0\}$ and $\varphi = \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0\}$ ).

Secondly, using the simulated critical value of each test statistic a decision of rejecting or not rejecting a null hypothesis is made.

Thirdly, a fixed number of Monte Carlo Sample Size is used to have a repetition of the above two steps and rejection of the null hypothesis is counted.

Fourthly, power of each test statistic at a specific alternative is calculated i.e.

$$Power\ of\ the\ Test = \frac{proportion\ of\ rejections\ out\ of\ MCSS\ at\ a\ specfic\ alternative}{MCSS}$$

Fifthly, a repetitions of the above four steps is made for each test statistic under a specific alternative to calculate the power.

Lastly, for all specific alternative, power of the tests is plotted to get power curve of the tests.

## 3.9. Attaining Power Envelop of the Point Optimal Test

Power envelop of the point optimal tests is obtained using the same simulation design as discussed for obtaining the power curve of the PUR tests for both categories.

First, data are generated using DGP-A and DGB-B under a point alternative and calculate the point optimal test (i.e. point optimal test for the null hypothesis of PUR and null hypothesis of stationary).

Secondly, using the simulated critical value a decision of rejecting or not rejecting the null hypothesis of point optimal tests is made.

Thirdly, a repetition of the above two steps using a fixed number of Monte Carlo Sample Size (MCSS) is made and rejection of the null hypothesis is counted for the given point optimal test under consideration.

Fourthly, power of point optimal test statistic is calculated for each point alternatives.

Lastly, power envelops of the point optimal tests (i.e. point optimal test for the null hypothesis of unit root and null hypothesis of stationary) are plotted against all the specified alternatives.

## 3.10. Finding Most Stringent Test

Stringency criterion is a newer one in the literature which use a point optimal test to compare different test-statistics to have a most stringent test among different test-statistics of the same field of study. A stringency criterion relates to the power curves of the tests and power envelop of the point optimal test under consideration.

**Shortcomings of the Test:** First to calculate the shortcomings of each test, the difference between the power envelope and power curve of each test for a specific alternative is taken. All shortcomings should be positive that is $S_j^m = PE_j - PT_j^m$ , $S_j^m \geq 0$ where $m = 1,2,.......,k$ and $j = 1,2,........,l$ denotes the total number of tests and total number of alternative, and $S_j^m$ denotes the shortcomings of test "$m$" at a specific alternative "$j$". $PE_j$ denotes the power envelop and $PT_j^m$ denotes the power of test "$m$" at a specific alternative "$j$". The calculation of shortcomings for all the point alternative hypothesis $j = 1, 2, \ldots\ldots\ldots, l$ for all the tests under consideration are made.

**Maximum Shortcomings (MSC) of The Test:** In the second step, the MSC of each test is calculated i.e. $\Pi = \max_{j=1,2,\ldots,l} \left(S_j^m\right)$

**Minimum of the MSC:** In third step, the minimum of MSC among all tests are obtained, i.e. $\Pi_{\min} = \min_{m=1,2,\ldots\ldots,k} (\Pi)$

**Stringent Test:** The test having minimum of the MSC i.e. $\Pi_{\min}$ is the most stringent test.

## 3.11. Bootstrap Empirical Power Evaluation

An empirical power investigation of best performing PUR tests has been taken place by using economic data to validate the results based on artificially data generating process. A purchasing power parity hypothesis has been taken by using real exchange rate data to evaluate the empirical power performance of best performing PUR tests by applying bootstrap method. For this purpose, bootstrap critical values have been calculated for different combination of time and cross section units to decide the acceptance and rejection regions of best performing PUR tests. This empirical experiment has been applied on monthly data of sixteen OECD countries by decomposing these sixteen OECD countries into four data sets.

# Chapter 04

## Empirical Size Evaluation of Panel Unit Root Tests

In this chapter, size evaluation of PUR tests, both having the null hypothesis of PUR and stationary, have been explained using asymptotic critical values for time series level of 10, 25, 50, and 100 corresponding to cross section unit of 2, 4, 8, 16, and 32. A Monte Carlo simulation size of 10,000 has been taken to analyze the size of PUR tests under the nominal size of 5% in two cases, first, in the presence of only drift term, second, in the presence of both drift and trend terms. In the first section, size evaluation of PUR tests having the null hypothesis of unit root based on asymptotic critical value have been discussed with drift term only while in the second section the empirical size analysis of PUR tests having the null hypothesis of unit root have been analyzed in the presence of both drift and trend terms. Similarly, size evaluation based on asymptotic critical value of PUR tests having the null hypothesis of stationary have been explained in the presence of both of the deterministic terms in the last two sections of this chapter.

### 4.1. Empirical Size Evaluation of Panel Unit Root Tests, Intercept Case

In this section, asymptotic critical value has been used for different combination of time series and cross section dimensions to investigate the size of eighteen tests having the null hypothesis of PUR when data are generated with intercept term only. Figure 4.1 to Figure 4.5 represent the empirical size evaluation of these eighteen tests in which x-axis shows time series dimension while y-axis represents empirical size.

**Figure 4.1: Empirical Size of Tests with Null of Unit Root (Intercept Case), N=2**

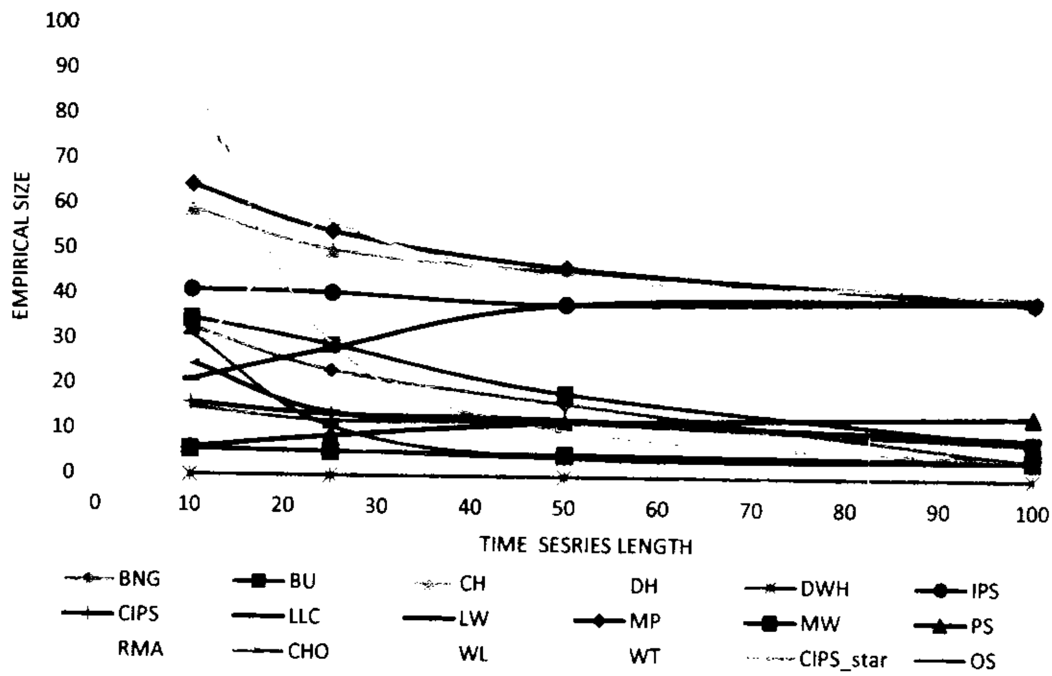



Figure 4.1 shows that majority of the tests converge to nominal size of 5% as the sample size gets large. Also, the p-value based MW and CHO tests have stable size equal to nominal size of 5% at each stage of time series level which means that only two tests among eighteen tests have empirical size equal to nominal size of 5% at each level of time series dimension. Majority of the tests have convergence behavior towards nominal size of 5% as the time series dimension approaches from small to large. While, LW, PS and WL tests are detected with divergence pattern as the time series level increases from small to large. Among these three over size distortion tests, WL test achieves 100% empirical size at large time series dimension while PS test has gained empirical size around nominal size of 5% at small (i.e. T=10) and medium time series (i.e. T=25), however, as the time series progress this test loss its stable size gain and become over size test. Further, DH and DWH tests have 0% of empirical size at each time series dimension and are identified as under

sized PUR tests having the null hypothesis of unit root. Moreover, BNG, BU, CIPS, LLC, CIPS_star, and OS tests with convergence behavior are identified as less size distortion tests with minimum empirical size around 5% and maximum empirical size of 34% as compared to other PUR tests having the null hypothesis of unit root.

**Figure 4.2: Empirical Size of Tests with Null of Unit Root (Intercept Case), N=4**

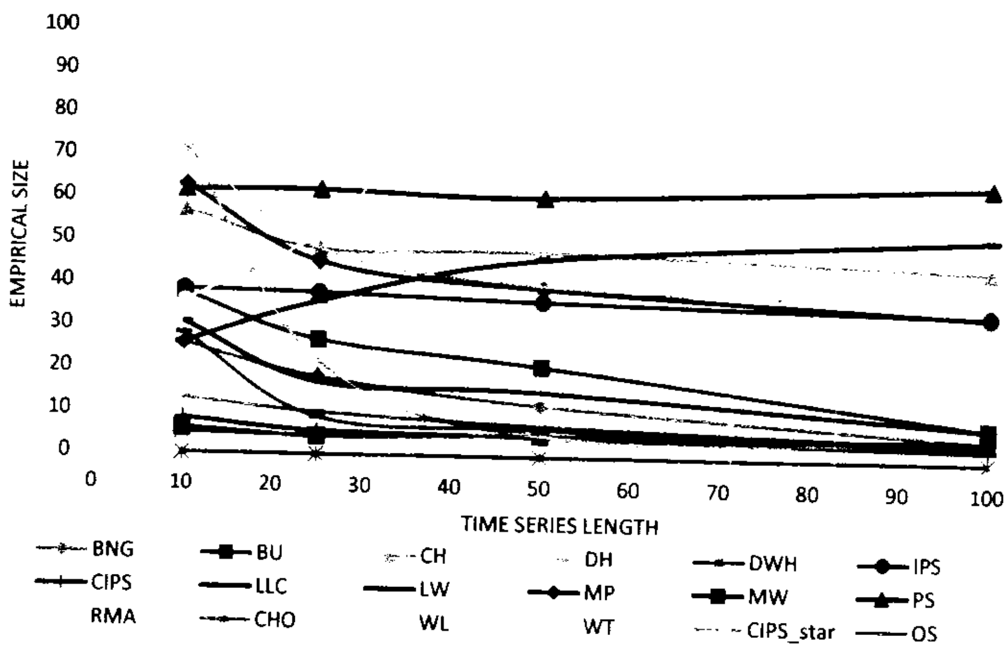
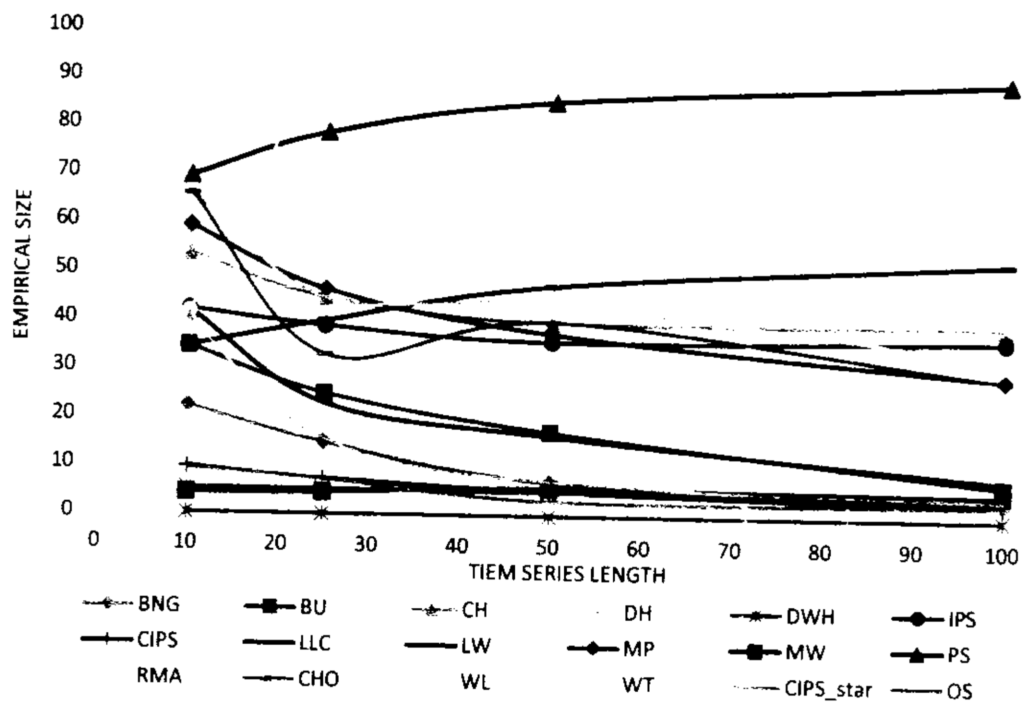



Figure 4.2 demonstrates the empirical size pattern of eighteen PUR tests having the null hypothesis of unit root when the number of cross sections are 4 over time series level of 10, 25, 50, and 100, and data are generated with intercept term only. It is clear from Figure 4.2 that majority of the tests have convergence behavior of empirical size as the time series dimension increases from small to large and these tests have achieved empirical size equal to nominal size of 5% at large time series dimension (i.e. T=100) while at small and medium time series level their empirical size are not stable. Moreover, CIPS, MW, and CHO tests have empirical size equal to nominal size of 5% at each time series level. Also,

CIPS_star test is very little over sized at time series 10 and 25 while at T=50 and T=100 its size is stable around nominal size of 5%. Similarly, OS test is stable at medium and large time series but at small time series (i.e. T=10) its size is 28% and is unstable. However, LW, PS, and WL tests have a divergence pattern and at each stage of time series level their empirical size remains unstable. It is also observed from Figure 4.2 that DH and DWH tests have empirical size equal to 0% and are detected as only two under size tests.

**Figure 4.3: Empirical Size of Tests with Null of Unit Root (Intercept Case), N=8**




Similarly, Figure 4.3 presents the empirical size performance of PUR tests having the null hypothesis of unit root when the number of cross sections are 8 over small, medium, and large time series level. It is observed from this figure that majority of the tests are unstable at small and medium time series level while at large time series dimension (i.e. T=100) the empirical size of these tests are become stable. Also, these tests have convergence pattern towards nominal size of 5% as the time series length increases. Moreover, it is analyzed that DH and DWH tests are detected as under size tests at each level of time series length and gets 0% empirical size at each stage level of time series. A stable empirical size of p-value based MW and CHO tests have been observed at small, medium and large time series while CIPS and CIPS_star tests have a stable empirical size at medium and large sample sizes while at small time series their empirical size is a little over size from nominal size of 5%. Moreover, there are three tests (LW, PS, and WL) with divergence behavior and unstable empirical size as the time series level increases from small to large. Also, WL test achieves 100% size distortion as the time series gets larger.

**Figure 4.4: Empirical Size of Tests with Null of Unit Root (Intercept Case), N=16**

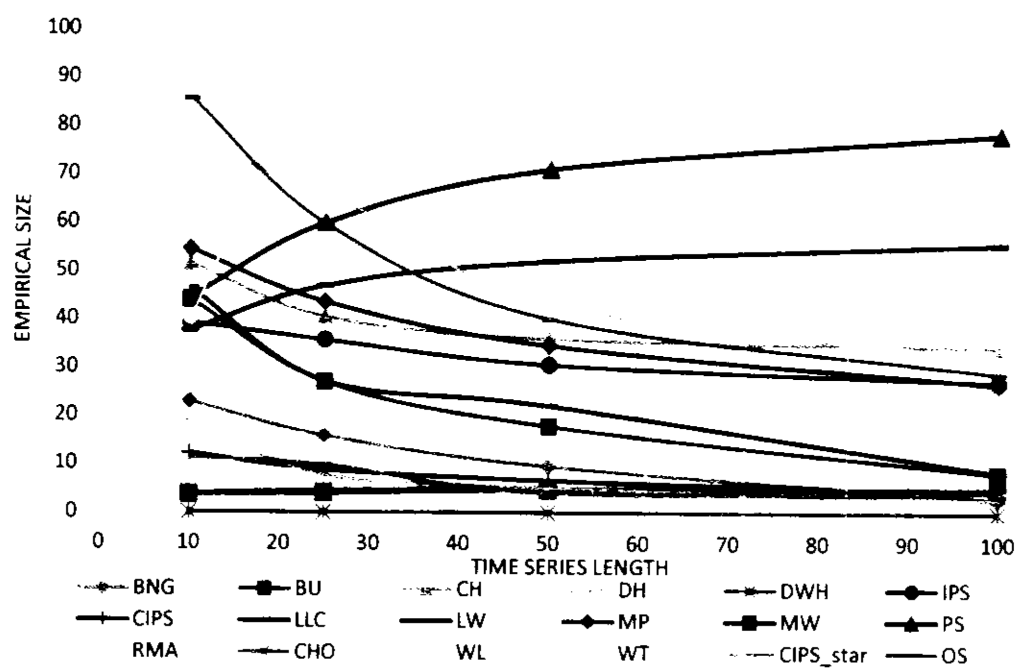



Figure 4.4 shows the behavior of PUR tests having the null hypothesis of unit root when the number of cross sections are 16. Again, a convergence and unstable empirical size pattern of majority tests have been observed as the time dimension increases from small to

large. At small time series majority of the tests are unstable and are oversized while at large sample their empirical size becomes stable around nominal size of 5%. MW and CHO tests are the only two tests out of eighteen tests which have stable empirical size at each level of time series while CIPS, WT and CIPS_star tests are also stable at time series 25, 50, and 100 but at T=10 their empirical size is oversized. Moreover, LW, PS, and WL tests with divergence pattern over increasing time series are classified as oversized tests. Among these three tests, WL test has gained much size distortion as compared to other two oversized tests. Similarly, DH and DWH tests with zero percent empirical size at all time series level are the only two under sized tests among eighteen tests at N=16.

**Figure 4.5: Empirical Size of Tests with Null of Unit Root (Intercept Case), N=32**

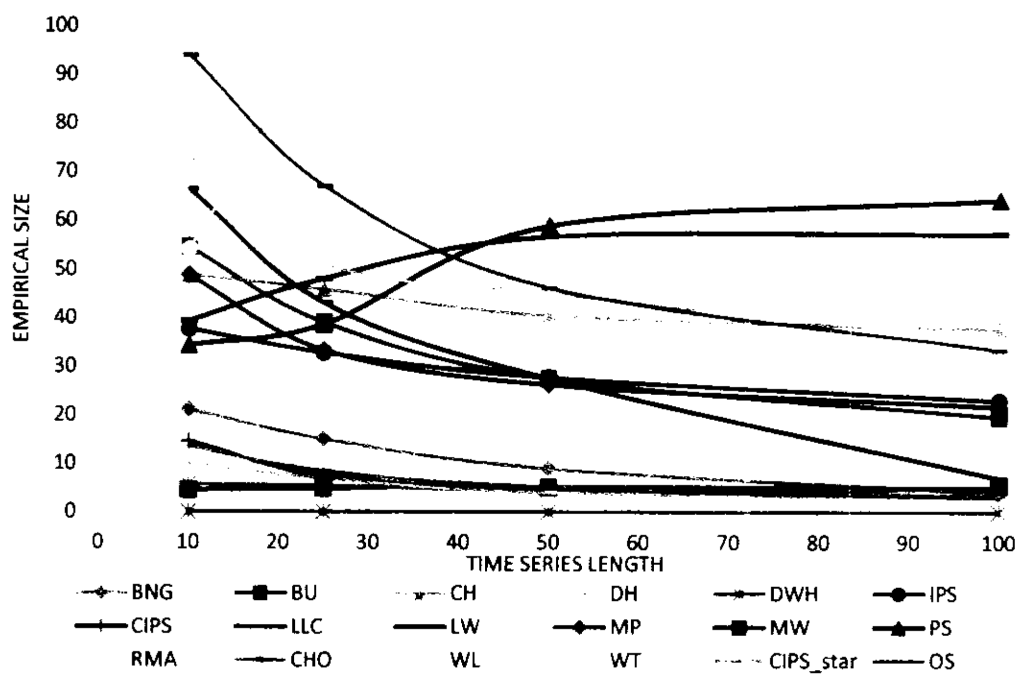



Figure 4.5 explains the empirical size performance of PUR tests having the null hypothesis of unit root when number of cross sections are large (i.e. N=32) and data are generated with intercept term only. The same results are observed from Figure 4.5 for the PUR tests

having the null hypothesis of unit root as have been elaborated earlier when the number of cross sections are less than 32. Again, majority of these tests have convergence behavior as the time series level approaches from small to large with unstable empirical size at small and medium time series while at large time series their empirical size become stable around nominal size of 5%. MW and CHO tests have stable size whether time series level is small, medium or large as these tests are developed based on p-value. Similarly, CIPS, WT, and CIPS_star tests become stable from empirical size point of view at medium and large sample sizes but at small time series these tests are a little oversized. Moreover, PS, LW, and WL tests have also unstable size with divergence behavior, as the time length gets larger. While, DH and DWH tests having 0% empirical size at each level of time series are identified only two under sized tests among eighteen PUR tests when the number of cross sections are 32.

Overall, Figure 4.1 to Figure 4.5 concludes that MW and CHO tests have stable empirical size around nominal size of 5% whether time length is small, medium or large corresponding to each level of cross section unit when data is generated with intercept term only. Similarly, CIPS and CIPS_star tests also get stable size at medium and large time dimensions but at small time series level these two tests are a little oversized whether cross section is small, medium or large. Also, majority of the tests show convergence behavior over time series and cross section length while there are three tests having divergence behavior over time and cross section units. Moreover, DH and DWH are the only two tests with 0% empirical size and are the only two tests identified as under size.

## 4.2. Empirical Size Evaluation of Panel Unit Root Tests, Intercept and Trend Case

In this section the empirical size behavior of eighteen PUR tests having the null hypothesis of unit root are investigated using asymptotic critical value for different combinations of time series and cross section length when data are generated in the presence of both of the deterministic terms. Figure 4.6 to Figure 4.10 shows the analysis of empirical size of these eighteen tests in which x-axis shows time series dimension while y-axis represents empirical size.

**Figure 4.6: Empirical Size of Tests with Null of Unit Root (Intercept and Trend Case), N=2**

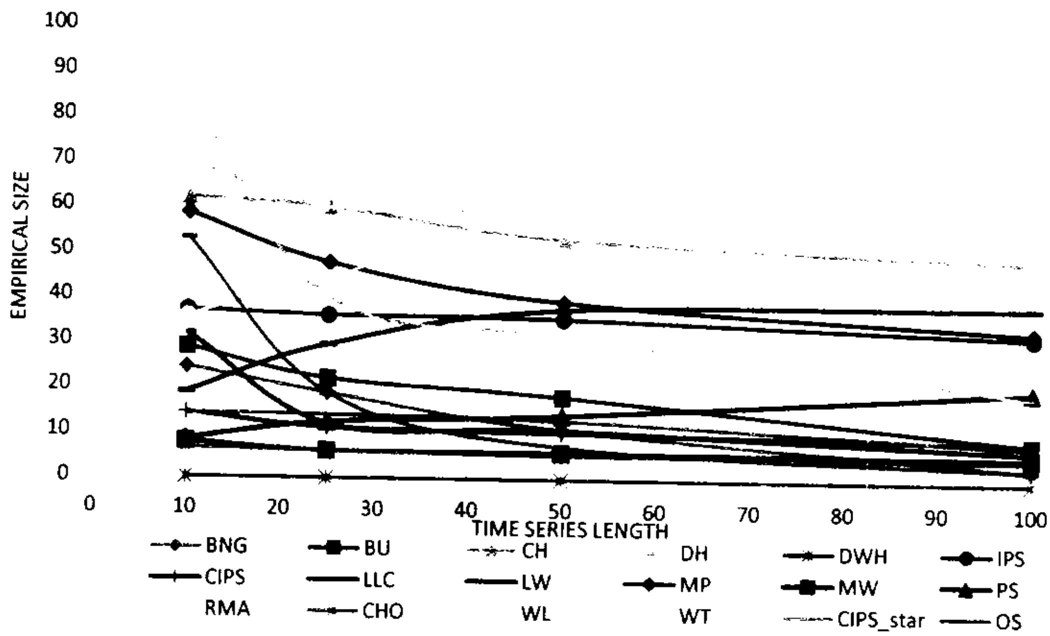
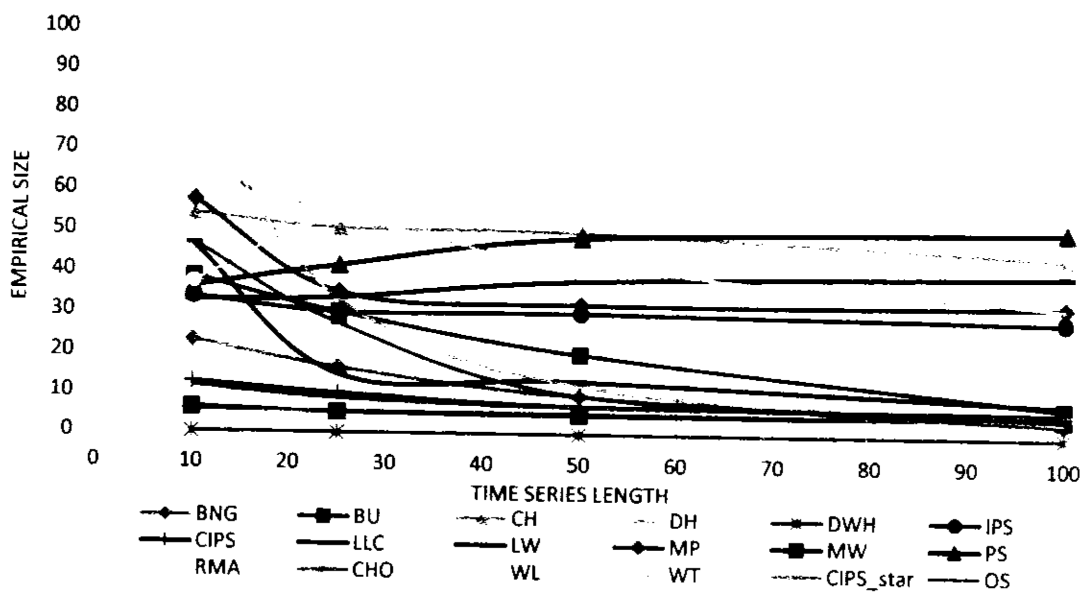



Figure 4.6 demonstrates the empirical size analysis of PUR tests using asymptotic critical value when data is generated with intercept and trend terms for the number of cross section units 2 over the small, medium, and large time series dimensions. This figure shows a

convergence pattern with unstable size at small and medium time dimensions for majority of the tests under investigation. At large time series length these tests become stable from empirical size point of view with a little fluctuations from the nominal size of 5%. However, only three tests (i.e. LW, PS, and WL) with unstable empirical size have divergence pattern as the time dimension gets larger. WT test has the maximum empirical size distortion of 83% at T=10 while WL test has 71% empirical size distortion at very large time series dimension 100. Only MW and CHO tests have stable size at each level of time length with a little fluctuation around nominal size of 5% and are identified as stable tests from asymptotic critical value point of view. Also, CIPS and CIPS_star tests with a little oversized pattern at small time series length have also stable size around nominal size of 5% at large time series length. Moreover, DH and DWH tests with zero percent empirical size are the only two under size tests from nominal size of 5% when N=2.

**Figure 4.7: Empirical Size of Tests with Null of Unit Root (Intercept and Trend Case), N=4**



When the number of cross section units are 4 and data are generated by taking both of the deterministic terms then Figure 4.7 shows that PUR tests having the null hypothesis show the same results as have been observed in the case of intercept for the same cross section unit.

At small time series level (i.e. T=10), WT test has the maximum empirical size of 78% while at large time length (i.e. T=100) this test has the highest empirical size of 78.32% as compared to other tests. According to Figure 4.7, majority of the tests converge towards nominal size of 5% as the time series length gets larger. Only, MW and CHO tests have stable size around nominal size of 5% at each level of time series length among eighteen pane unit root tests having the null hypothesis of unit root as both of these tests have been developed on the basis of p-value. Also, CIPS and CIPS_star tests have stable size at T=50 and T=100 while both of them are a little oversized from nominal size of 5% at T=10 and T=25. Moreover, LW, WL, and WT tests with divergence behavior are identified as the most size distortive tests among all eighteen-PUR tests. While, DH and DWH tests with zero empirical size at each time series length are identified as the only two under size tests from the nominal size of 5%.

These results of Figure 4.7 are parallel to Figure 4.2 when the numbers of cross sections are 4 and data are generated in the presence of both of the deterministic terms with varying time series length of 10, 25, 50, and 100. In both figures, MW and CHO tests are identified as stable tests around nominal size of 5% by using asymptotic critical value.

**Figure 4.8: Empirical Size of Tests with Null of Unit Root (Intercept and Trend Case), N=8**

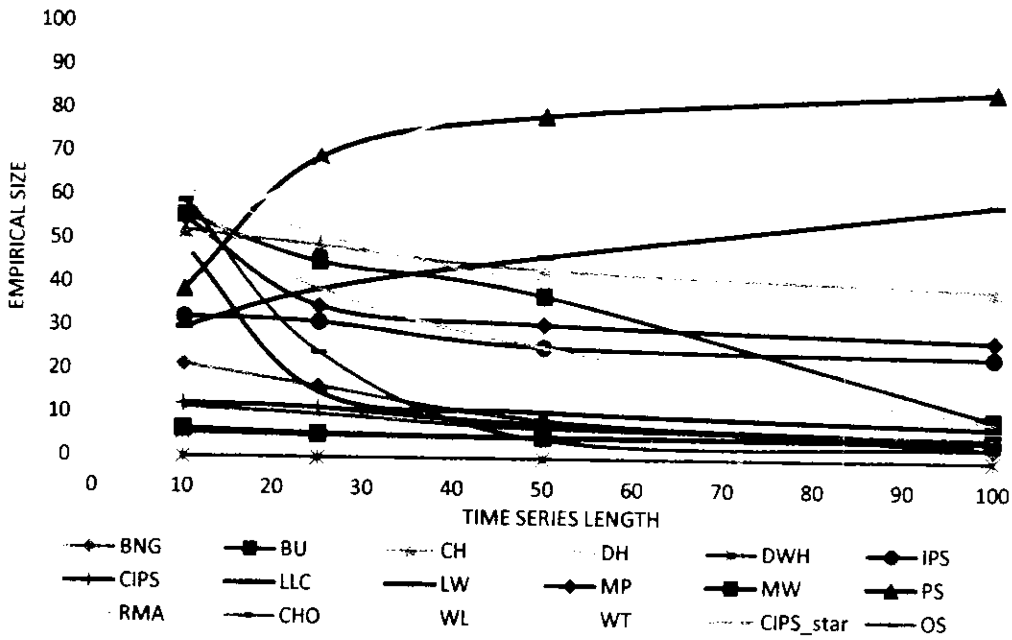
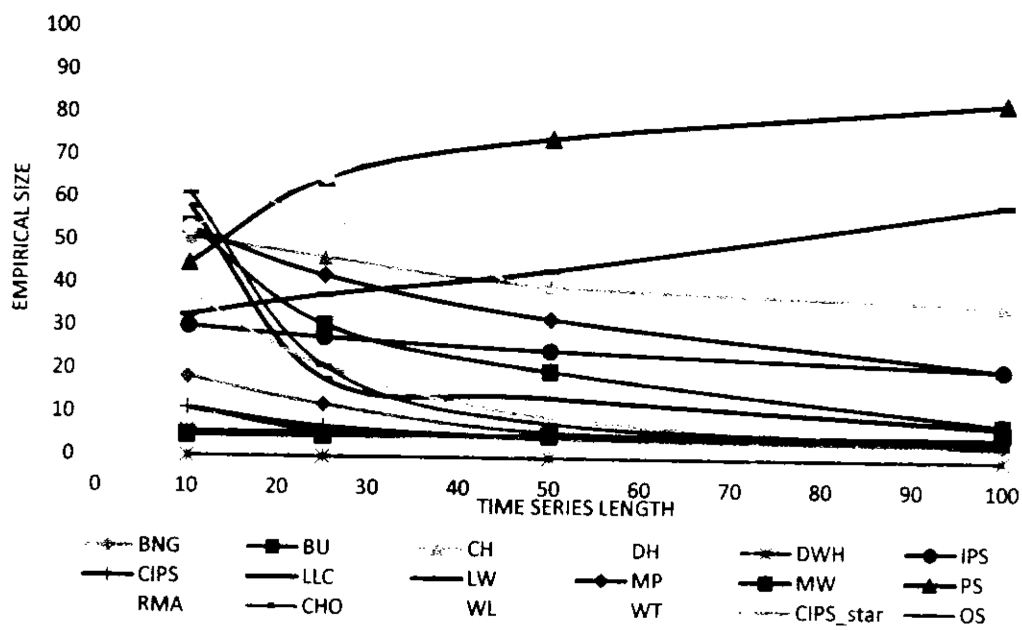



Figure 4.8 shows the empirical size pattern of PUR tests having the null hypothesis of unit root when the number of cross sections is 8 when data are generated in the presence of both of the deterministic terms. From Figure 4.8, a convergence and unstable empirical size pattern of majority tests have been observed as the time dimension increases from small to large. At small time series majority of the tests are unstable and are oversized while at large sample their empirical size becomes stable around nominal size of 5%.

According to Figure 4.8, MW and CHO tests are the only two tests among eighteen tests which have stable empirical size at each level of time series. Similarly, CIPS and CIPS_star tests are also stable at time series 25, 50, and 100 but at T=10 their empirical size is oversized. Moreover, LW, PS, and WL tests with divergence pattern over increasing time series are classified as oversized tests. Among these three tests, WL test has gained much size distortion as compared to other two oversized tests. However, DH and DWH tests with zero percent empirical size at each level of time series are the only two under sized tests among eighteen tests at N=8. These results of Figure 4.8 are similar to the results of Figure 4.3 when data are generated with intercept term only.

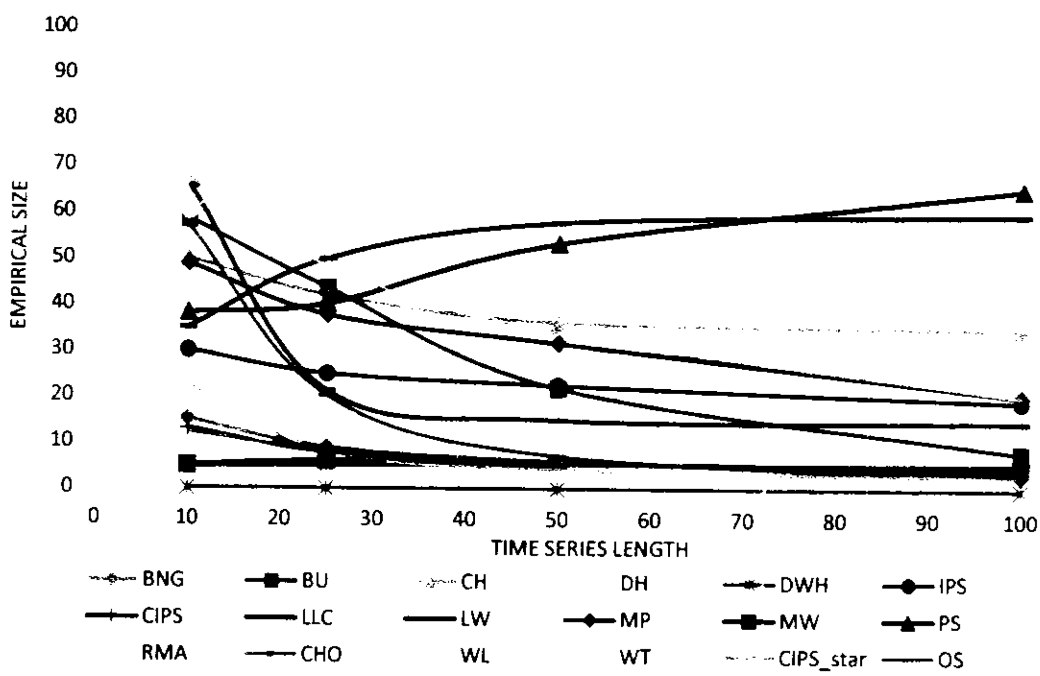
**Figure 4.9: Empirical Size of Tests with Null of Unit Root (Intercept and Trend Case), N=16**




When the numbers of cross sections are 16 then Figure 4.9 shows a very similar picture as has been observed for the previous cross section units when data are generated in the presence of both of the deterministic terms. A convergence pattern of empirical size towards nominal size of 5% has been observed for fifteen PUR tests having the null hypothesis of unit root. Among these tests, the maximum and minimum empirical size has been observed as 64% corresponding to RMA test at T=10 and 2.99% corresponding to WT test at T=100. Also, majority of these tests become stable at large time series level.

Similarly, LW, PS, and WL tests with divergence behavior towards 100% empirical size are identified as most size distortive size tests as the time series level gets larger. Only, two tests (i.e. MW and CHO) have stable empirical size around nominal size of 5% at each time series length while CIPS and CIPS_star tests are a little oversized at small cross section but both of them become stable as the time dimension increases. However, there are two tests having 0% empirical size and are identified as only two under sized tests from nominal size of 5%. Figure 4.9 results are very similar to results of Figure 4.4 when data is generated in the presence of intercept term only at N=16.

**Figure 4.10: Empirical Size of Tests with Null of Unit Root (Intercept and Trend Case), N=32**



In the last, Figure 4.10 presents the empirical size behavior of PUR tests having the null hypothesis of unit root when the number of cross sections are 32. It is obvious from Figure 4.10 that all tests behave very similarly as has been concluded from the previous figures when the number of cross sections are less than 32.

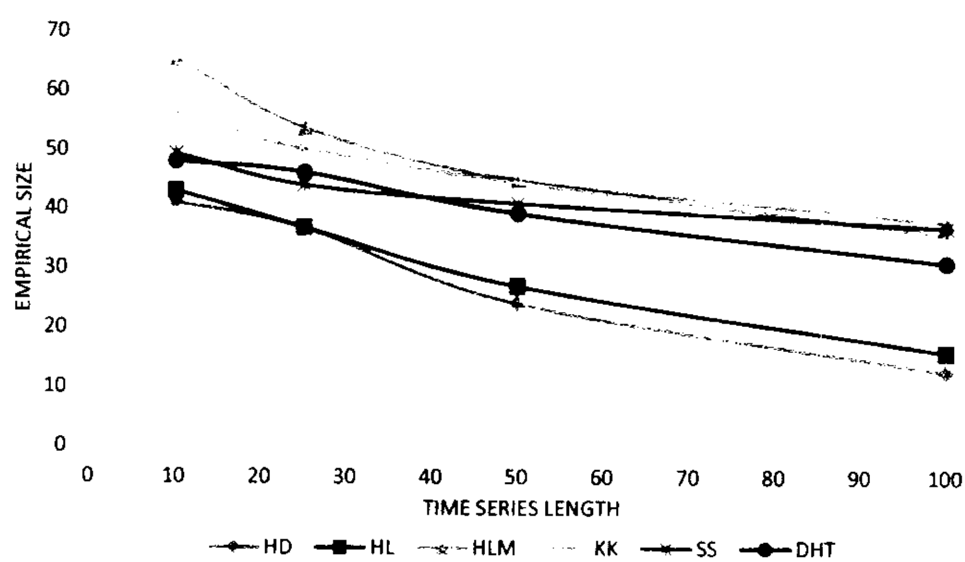
Again, a convergence empirical size pattern towards nominal size of 5% for majority tests have been observed as the time dimension increases from small to large. At small time series majority of the tests are unstable and are oversized while at large sample their empirical size becomes stable around nominal size of 5%. MW and CHO tests are the only two tests out of eighteen tests which have stable empirical size at each level of time series while CIPS, WT and CIPS_star tests are also stable at time series 25, 50, and 100 but at T=10 their empirical size is a little oversized from nominal size of 5%. Moreover, LW, PS, and WL tests with divergence and unstable empirical size pattern over increasing time series are classified as oversized tests from nominal size of 5%. Among these three tests, WL test has gained much size distortion as compared to other two oversized tests. Similarly, DH and DWH tests with zero percent empirical size at each level of time series are the only two under sized tests among eighteen tests when the number of cross sections are 32.

Overall, Figure 4.6 to Figure 4.10 concludes a convergence picture of empirical size towards nominal size of 5% of majority of the tests when data are generated in the presence of both of the deterministic terms. At each level of cross section units over fixed time series level MW and CHO tests are identified as more stable tests while CIPS and CIPS_star tests are only stable at medium and large time series length but they are a little oversized from nominal size of 5% at T=10. Moreover, LW, PS, and WL tests corresponding to each cross section unit with fixed time series length are classified as divergent tests from empirical size point of view. However, DH and DWH tests are the only two tests having 0% empirical size at each level of time series dimension corresponding to cross section units 2, 4, 8, 16, and 32. Overall, both of these two tests are detected as under sized tests.

## 4.3. Empirical Size Evaluation of Panel Stationarity Tests, Intercept Case

In this section, the empirical size pattern has been evaluated using asymptotic critical value of six PUR tests (i.e. HD, HL, HLM, KK, SS, and DHT) having the null hypothesis of stationary when data are generated in the presence of intercept term only. Figure 4.11 to Figure 4.15 explains the analysis of empirical size of these tests in which x-axis represents time series dimension while y-axis shows empirical size.

**Figure 4.11: Empirical Size of Tests with Null of Stationary (Intercept Case), N=2**




In Figure 4.11, empirical size evaluation of stationary tests have been shown when the numbers of cross section units are 2 and data are generated with an intercept term only. It is observed that at small time series length all stationary tests have empirical size in between 40% and 70% in which HD and HL tests have minimum size distortion as compared to other tests from nominal size of 5%. As the time series length increases from small to large then all stationary tests have convergence behavior with respect to their empirical size. HD and HL tests have less size distortion as compared to other stationary tests at each level of time series dimension while HLM test has the most empirical size distortion at each level of time series length. Also, HD and HL tests converge to towards nominal size of 5% as the time dimension gets larger. At T=100, their (HD and HL tests) empirical size become 12% and 15% respectively which is not too far from nominal size of 5% when N=2. However, from Figure 4.11, all other stationary tests (i.e. HLM, KK, SS, and DHT) have achieved empirical size in between 30% to 40% at large time dimension (i.e. T=100) when the number of cross sections are 2, thus these tests are more size distortive ones.

**Figure 4.12: Empirical Size of Tests with Null of Stationary (Intercept Case), N=4**

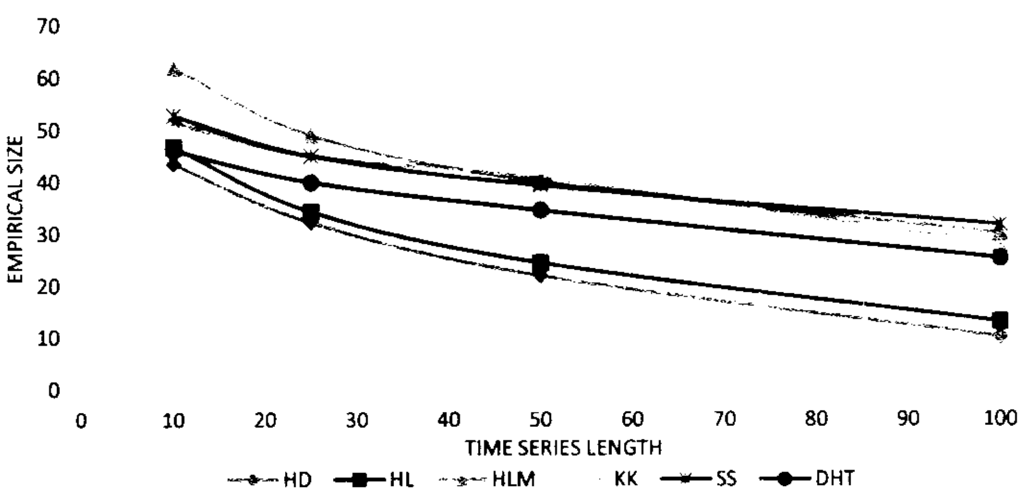
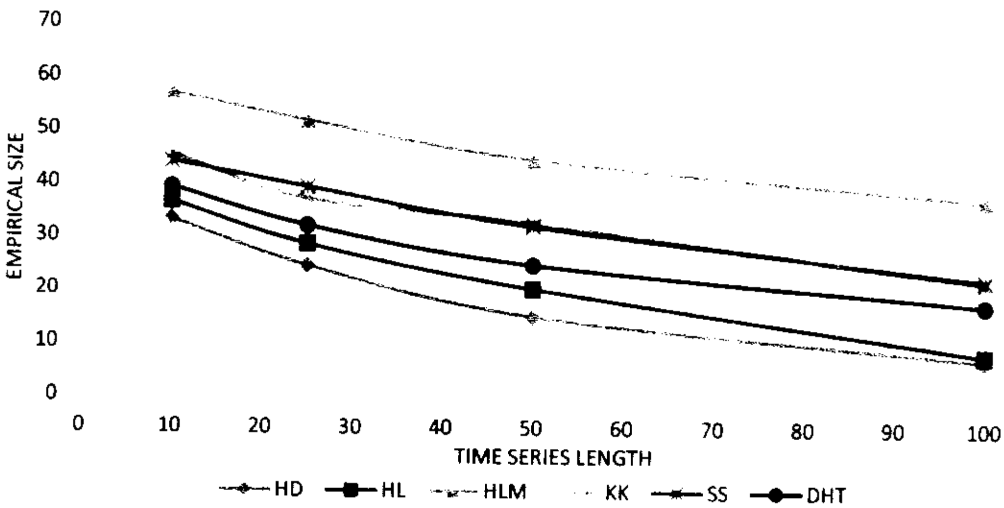


Figure 4.12 represents the empirical size performance of stationary tests when the number of cross sections is four and data are generated in the presence of intercept term only. A convergence pattern of empirical size is obtained by all stationary tests as the time dimension varies from small to large with minimum value of 10% of HD test corresponding to large time dimension (i.e. T=100) while a maximum value of 61% of HLM test corresponding to T=10. According to Figure 4.12, HD and HL tests gain minimum empirical size whether time series level is small, medium or large and both of these tests perform similar. Also, their empirical size is very close to the size of nominal size of 5% as compared to other stationary tests. While, HLM test has the highest empirical size at each time dimension and this test is identified as the more size distortive test among the all stationary tests. From Figure 4.12, it is noticed that as the number of cross section units gets larger then empirical size of each stationary test drops towards nominal size of 5%.

When the number of cross section units are 8 then Figure 4.13 explains the empirical size behavior of stationary tests over time series 10, 25, 50, and 100 when data are generated with intercept term only. All stationary tests show a convergence pattern as the time series length increases from small to large. At T=10, the maximum (i.e. 57%) and minimum (6.97%) empirical size has been observed corresponding to HMLM and HD tests. Similarly, when the time series length is 25 then HLM test with 51% and HD test with 24% empirical size are identified as the most and less size distortive tests from nominal size of 5%.

Further, as the time dimension reaches to 50 and 100 then HD test with 15% and 6% empirical size has been detected as less size distortive test while HLM test with 44% and 36% is assigned as more size distortive test from nominal size of 5%.

**Figure 4.13: Empirical Size of Tests with Null of Stationary (Intercept Case), N=8**




Overall, as the time length moves from small to large then HD and HL tests having minimum value of their empirical size at each time series are identified as the less size distortive tests while HLM test is classified as more size distortive test. Also, HD and HL tests gets their empirical size equal to nominal size of 5% when the time length is large. While, KK and SS tests have very close empirical size at each stage of time dimension just like HD and HL tests.

Moreover, from Figure 4.13, it is concluded that as the number of cross sections are increases then empirical size of each stationary test decreases towards nominal size of 5%.

**Figure 4.14: Empirical Size of Tests with Null of Stationary (Intercept Case), N=16**

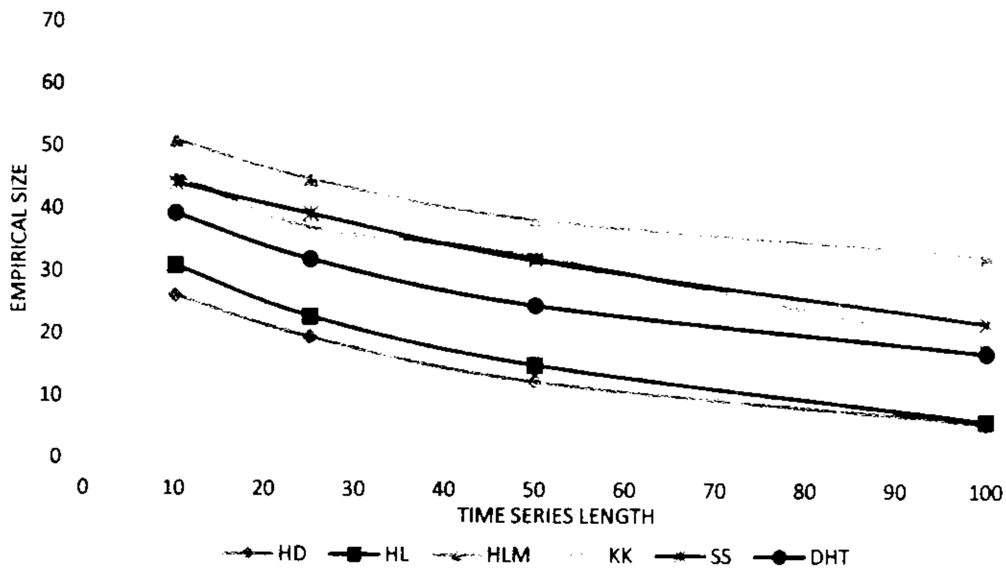



Figure 4.14 investigates the empirical size of stationary tests obtained by using asymptotic critical value for different time series length when the number of cross section unit is 16. Figure 4.14 shows that all stationary tests have declining pattern as the time series length increases from small to large.

At small time series dimension (i.e. T=10) Figure 4.14 shows HLM test with highest empirical size, thus indicating high size distortion from nominal size of 5%, while HD test with empirical size of 26% is recognized as less size distortive test. Moreover, as the time dimension increases from 10 to 25, 50, and 100 then again HLM and HD tests have maximum and minimum empirical size at stage of time dimension, respectively.

Figure 4.14 shows that HD and HL tests have less empirical size distortion from nominal size of 5% at each time series length and at large time dimension their empirical size become equal to nominal size of 5%. Similarly, KK, SS, and DHT tests with convergence picture have very close empirical size at each time series length.

**Figure 4.15: Empirical Size of Tests with Null of Stationary (Intercept Case), N=32**

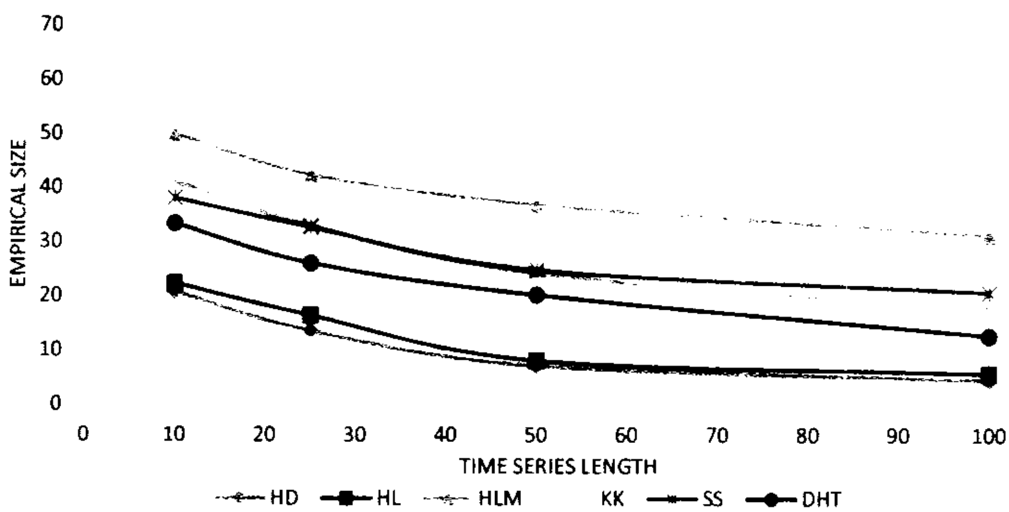



Figure 4.15 analyzes the asymptotic critical value based empirical size of six stationary tests over time series length small, medium, and large when the number of cross sections are large (i.e. 32) and data are generated with intercept term only. Again, a convergent picture towards nominal size of 5% has been observed for all stationary tests as the time series dimension increases from small to large.

Figure 4.15 explains that HLM test has highest empirical size at each time series length and is the more size distortive stationary test while HD and HL tests with minimum empirical size are identified as less size distortive tests from nominal size of 5%. Also, at large time series length both of these tests become equal around nominal size of 5%. Overall, Figure 4.15 concludes an unstable size behavior of all stationary tests at all time series dimension.

Figure 4.11 to Figure 4.15 concludes that all six panel stationary tests are unstable at small and medium time series length but only HD and HL tests become stable at large time and cross section units from nominal size of 5% when data are generated with intercept term only. Similarly, HLM test with unstable empirical size pattern for all combination of time series and cross section units is identified more size distortive test when data are generated with intercept term only. Also, a decreasing size distortive picture has been observed for all panel stationary tests as both time series and cross section units increases. At very large time series and cross section units our analysis indicates that the empirical size will eventually become equal to nominal size of 5% for almost all tests when data are generated with intercept term only.

## 4.4. Empirical Size Evaluation of Panel Stationarity Tests, Intercept and Trend Case

In the last section of this chapter, the empirical size of all PUR tests having the null hypothesis of stationary have been discussed when data are generated with intercept and trend terms for different number of cross section units (i.e. N=2, 4, 8, 16, and 32) having the fixed time series levels (i.e. T=10, 25, 50, and 100). The empirical size of these tests can be analyzed from Figure 4.16 to Figure 4.20.

Figure 4.16 indicates the empirical size pattern of panel stationary tests for the number of cross section units 2 over time series length of 10, 25, 50, and 100 when data are generated both with intercept and trend terms.

**Figure 4.16: Empirical Size of Tests with Null of Stationary (Intercept and Trend Case), N=2**

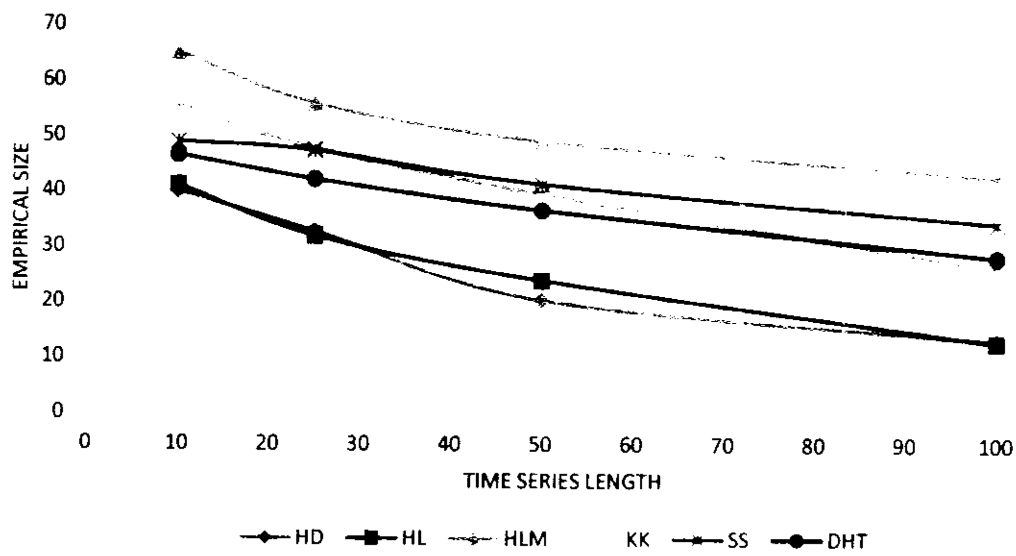



Figure 4.16 shows an unstable (from nominal size of 5%) and convergence picture for all panel stationary tests as the time series length increases from small to large. Among these six panel stationary tests, only HD and HL tests gain less size distortion at small to large time series dimensions as compared to other stationary tests. While HLM test with more empirical size distortion from nominal size of 5% is identified as more unstable test among all stationary tests.

Moreover, KK, SS, and DHT tests with approximately same unstable and convergence pattern get empirical size in between 26% to 34% when time dimension is large as compared to their empirical size in between 46% to 55% at small time dimension. This means that all these three tests gain less size distortion at large time series level as compared to their empirical size at small time series length. Results of Figure 4.16 are similar to that of Figure 4.11 when data are generated with intercept term only.

**Figure 4.17: Empirical Size of Tests with Null of Stationary (Intercept and Trend Case), N=4**

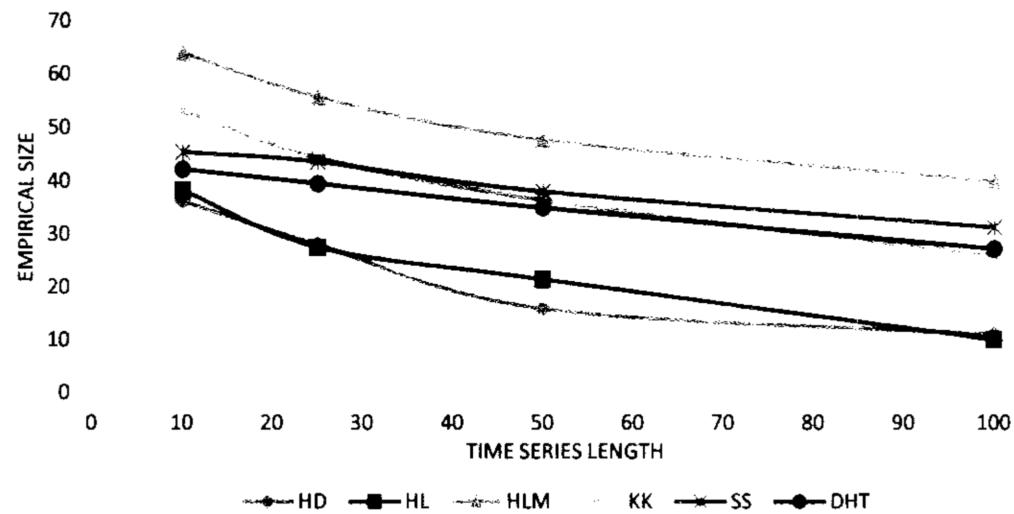
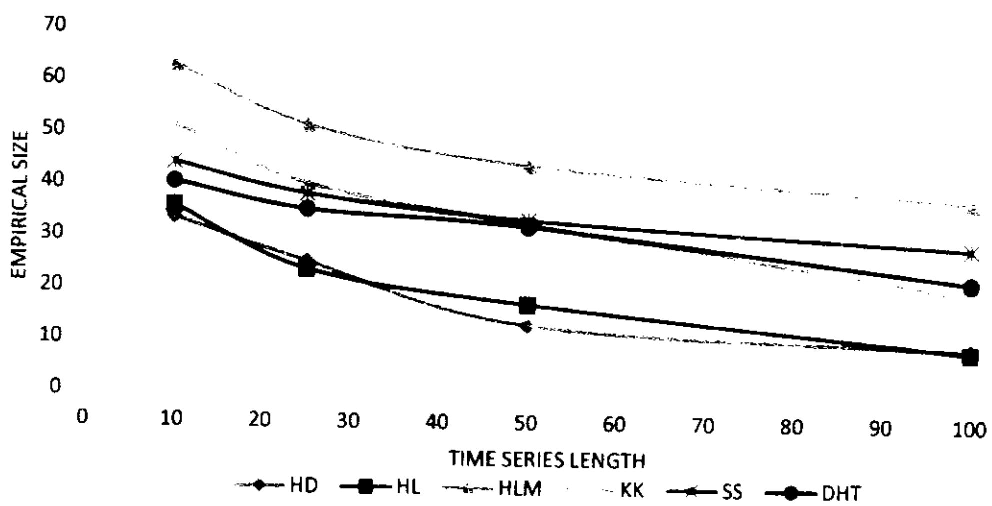



Figure 4.17 shows the empirical size performance of panel stationary tests when the number of cross sections are 4. Again, all tests show a convergence (towards nominal size of 5%) and unstable (from nominal size of 5%) pattern as the time dimension progress from small to large at N=4 when data are generated in the presence of both of the deterministic terms.

At small to large time series dimension, only HD and HL tests have maintained less size distortion while HLM test has achieved more size distortion from nominal size of 5% among all panel stationary tests. However, KK, SS, and DHT tests also achieve unstable empirical size at small to large time series but their behavior remains the same over different time dimension. Figure 4.17 shows a more decreasing value of empirical size of all tests as compared to Figure 4.16 when the number of cross section units is 2. Also, the results of Figure 4.17 are the similar to Figure 4.12 when data are generated with intercept term only.

**Figure 4.18: Empirical Size of Tests with Null of Stationary (Intercept and Trend Case), N=8**




When the number of cross section units are 8 then a very similar picture has been obtained from Figure 4.18 as has been observed from Figure 4.16 and Figure 4.17 when the number of cross section units are less than 8 and data are generated with intercept and trend terms. First, a convergence and unstable pattern of all panel stationary has been obtained as the time series length progress from small to large. At small time series level all tests have empirical size in between 33% to 62% which decreases to 8% to 36% at large time dimension when the number of cross sections are 8. Second, HD and HL tests remains less size distortive as compared to HLM, KK, SS, and DHT tests when time series length approaches from small to large. At T=100, both of these two tests get empirical size very close to nominal size of 5% which will eventually become equal to 5% at very large time series level. Third, HLM test with highest level of size distortion is identified as more size distortive test among the panel stationary tests at each level of time series length.

Lastly, KK, SS, and DHT tests with very similar pattern of their unstable empirical size over small to large cross units are ranked in between HD, HL, and HLM tests. Further, the empirical size of panel stationary tests have been decreased at N=8 as compared to their empirical gain at N=4 and N=2. Also, Figure 4.18 indicates a very similar sketch as it has been observed from Figure 4.13 when data are generated in the presence of an intercept term only.

**Figure 4.19: Empirical Size of Tests with Null of Stationary (Intercept and Trend Case), N=16**

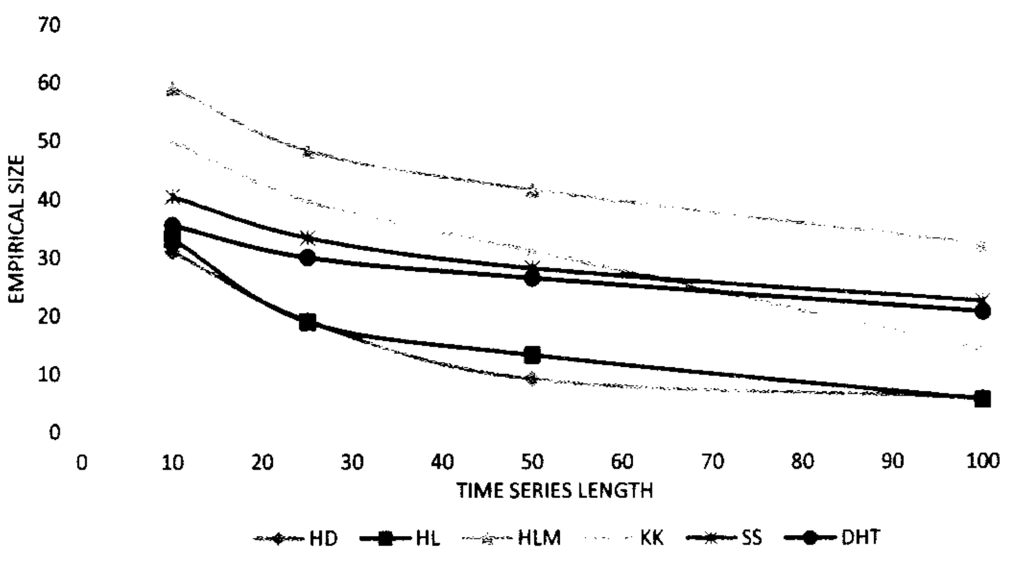


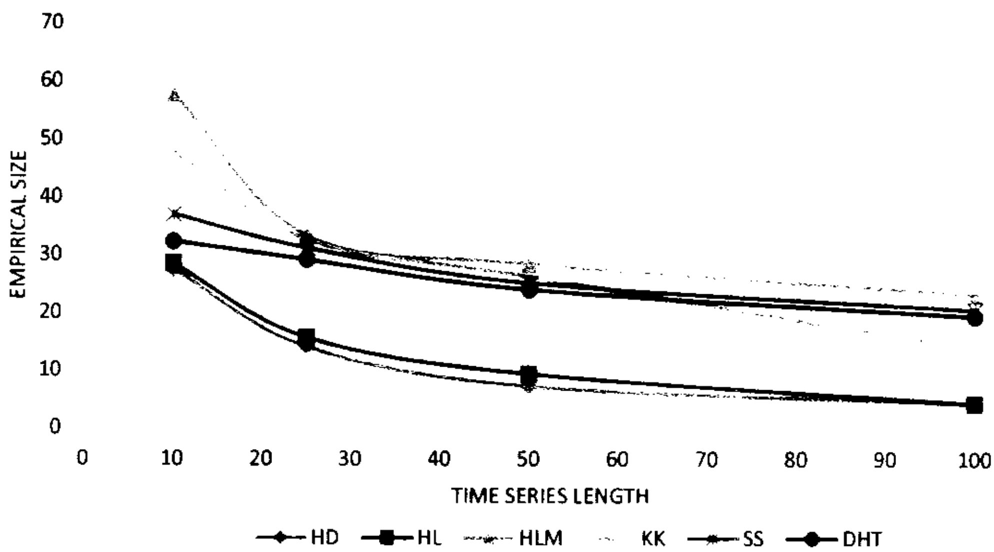

Figure 4.19 investigates the empirical size of stationary tests when the number of cross section units is 16 over small to large time series. Again, a convergence and unstable empirical size picture has been observed over small to large time dimension for all panel stationary tests.

At T=10, the maximum empirical size of 60% is gained by HLM test while a minimum of 30% by HD test among all six tests. Similar results have been observed in the case of HLM and HD tests among all panel stationary tests as the time progress from medium to large. Among these six panel stationary tests, HD and HL tests have less size distortion from nominal size of 5% at each level of time dimension. Also, both of these tests achieve their empirical size equal to nominal size of 5% as the time dimension increases to 100. While, HLM test with unstable and more size distortion from nominal size of 5% is concluded as the most size distortive test over time series length 10 to 100.

Moreover, KK, SS, and DHT tests with very similar unstable empirical size pattern lie in between less and more size distortive tests from nominal size of 5%. Figure 4.19 indicates a very similar picture as has been concluded from Figure 4.14 when data are generated with intercept term only. Also, the empirical size of each panel stationary test has been dropped as compared to their empirical size from Figure 4.16 to Figure 4.18.

When the number of cross section units are 32 then Figure 4.20 explains the empirical size behavior of stationary tests when data are generated with intercept and trend terms. All panel stationary tests maintain a convergence and unstable empirical size pattern as the time dimension moves from small to large. Moreover, the empirical size of HD and HL tests become stable around nominal size of 5% at medium and large time dimensions. Also both of these tests take less size distortion at each time series length which fluctuate around 28% to 5% as compared to all other stationary tests. While, HLM test with maximum empirical size at each time series level is considered as more size distortive test among KK, SS, and DHT tests. However, HLM test has become very close to empirical size of KK, SS, and DHT tests as the number of cross section units become large as compared to its performance of empirical size for number of cross section less than 32.

**Figure 4.20: Empirical Size of Tests with Null of Stationary (Intercept and Trend Case), N=32**




The findings of Figure 4.20 are similar to the results of Figure 4.15 when the data are generated in the presence of intercept term only. Also, from Figure 4.16 to Figure 4.20, a decreasing empirical size of each test has been seen as the number of cross section and time units gets larger. At very large cross section and time series units all these tests will get approximate empirical size towards nominal size of 5%.

## 4.5. Concluding Remarks

In the first (Section 4.1) and second section (Section 4.2) of this chapter it is concluded that majority of PUR tests having the null hypothesis of unit root have unstable empirical size from nominal size of 5% for different combination of time series and cross section lengths when asymptotic critical value is used to calculate size. Majority of the PUR tests (BNG, BU, IPS, LLC, CIPS, CIPS_star, CH, WT, MP, OS, RMA) have convergence pattern towards nominal size of 5% while only three tests (LW, PS, and WL) are detected with divergence behavior from nominal size of 5% as the number of cross section units varies over fixed time series level of 10, 25, 50, and 100.

Further, MW and CHO tests are recognized as more stable tests whether time and cross section dimensions ae small, medium or large. While CIPS and CIPS_star tests are only stable at medium and large time series length but they are a little oversized from nominal size of 5% at T=10. However, DH and DWH tests are the only two tests having 0% empirical size at each level of fixed time series dimension corresponding to cross section units 2, 4, 8, 16, and 32. Overall, both of these two tests are detected as under sized tests.

Similarly, the last two sections (Section 4.3 and Section 4.4) have evaluated the empirical size performance of six PUR tests (i.e. HD, HL, HLM, KK, SS, and DHT) having the null hypothesis of stationary when asymptotic critical value is used to calculate size. It is concluded that all tests have unstable empirical size from nominal size of 5% when the time dimension is small, medium or large. At large time series dimension HD and HL test become equal to nominal size of 5% when the number of cross sections are greater than 4. Also, both of these two tests have less size distortion as compared to other stationary tests at each level of time series for all cross section units. Similarly, HLM test has the more size distortion from nominal size of 5% at each combination of time series and cross section lengths. While, KK, SS, and DHT tests also have unstable empirical size from nominal size of 5% but all these three tests lie in between category of less and more size distortive tests at each combination of time series and cross section length. Hence, it is necessary to stabilize size before making any comparison of PUR tests.

# Chapter 05

## Size Analysis of Panel Unit Root Tests

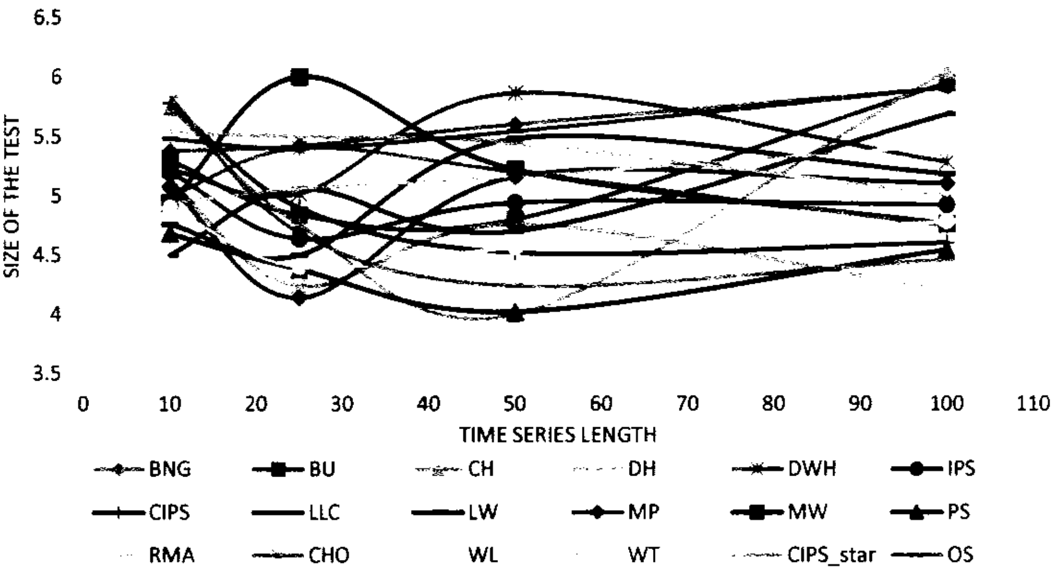
### 5.1. Stabilizing Size using Simulated Critical Value

This chapter focuses on size of both type of PUR tests (i.e. null hypothesis of unit root and null hypothesis of stationary). In previous chapter, it is concluded that asymptotic critical value causes unstable size for both categories of PUR tests. In this chapter, simulated critical value are used to equalize type-I error at nominal level of 5% for all PUR tests under consideration. This size stabilization of tests are made for different level of time series and cross sectional dimension for model with deterministic parts. Tests to be compared on the basis of size are abbreviated as; MP, PS, BNG, CH, CHO, LLC, BU, IPS, MW, CIPS_star, CIPS, DWH, RMA, LW, DH, WT, OS, and WL. All these tests belong to the null hypothesis of PUR. The null hypothesis of stationary tests is abbreviated as HD, HL, HLM, SS, KK, and DH. In the first part of this chapter, the size of tests having the null hypothesis of PUR are evaluated using simulated critical value. Lastly, this chapter shit light on the second category of PUR tests having the null hypothesis of stationary under simulated critical value. This stud considers T=10, 25, 50, 100 and N= 2, 4, 8, 16, 32 for time series and cross section dimensions respectively to analyze size of both categories of tests using Monte Carlo Simulation Size (MCSS) 10,000.

### 5.2. Stabilized Size with the Null Hypothesis of Panel Unit Root (Intercept Case)

In this section, size of eighteen PUR tests having the null hypothesis of unit root based on simulated critical value are analyzed. Figure 5.1 shows the stabilized size behavior of PUR tests for N=2 and intercept case only. In this graph, x-axis represents the time dimension and y-axis shows the size of the tests. It is clear that all tests size fluctuate around 5% of nominal size as the time dimension progress from 10 to 100, showing the stability of the tests based on simulated critical value. This fluctuation varies from 4% to 6% for almost all tests.

**Figure 5.21: Size of the Tests with Null Hypothesis of Unit Root and Intercept Case only, N=2**




When number of cross sections increases from 2 to 4 then the same picture is obtained as has been described for N=2 as the sample size gets larger from smallest (N=10) to largest (N=100). Figure 5.2 shows a smooth fluctuation around 4% to almost 6% which is very close to nominal size of 5%.

Figure 5.3 clearly portrays that all tests are size stabilized at nominal size of 5% when simulated critical values are used for N=8 as the time series gets larger from 10 to 100. WL has the smallest size of 4.18% while LW has the largest size of 6% when simulated critical value is used to calculate nominal size of 5%.

**Figure 5.22: Size of the Tests with Null Hypothesis of Unit Root and Intercept Case only, N=4**

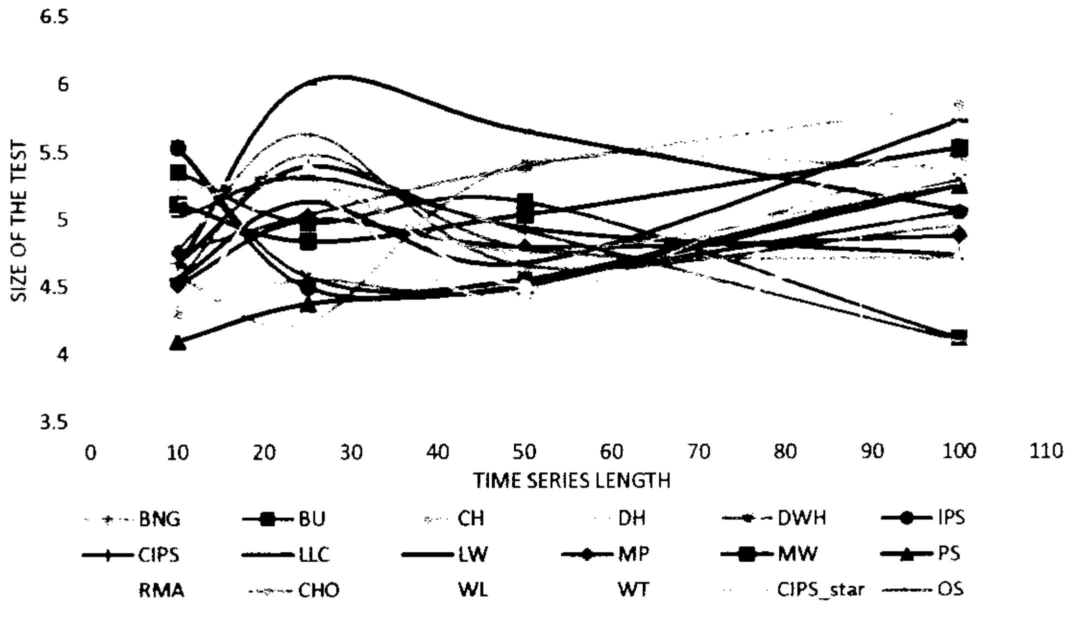



**Figure 5.23: Size of the Tests with Null Hypothesis of Unit Root and Intercept Case only, N=8**

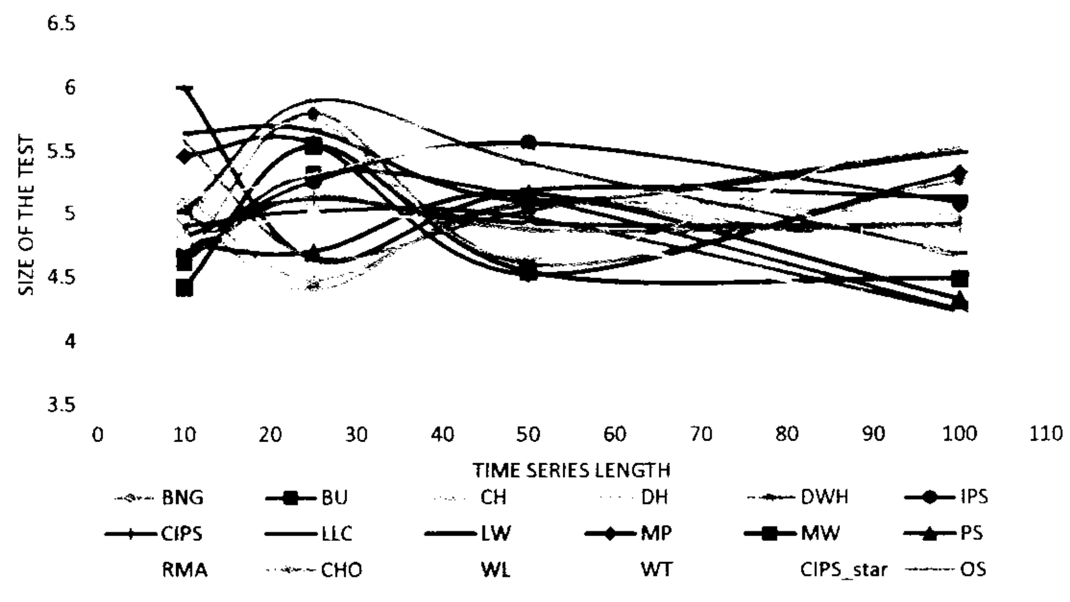


Figure 5.4 and Figure 5.5 concludes that all tests have a stable size around nominal size of 5% for N=16 and N=32 as time dimension varies from 10 to 100 when simulated critical value is calculated for different combinations of time series.

**Figure 5.24: Size of the Tests with Null Hypothesis of Unit Root and Intercept Case only, N=16**

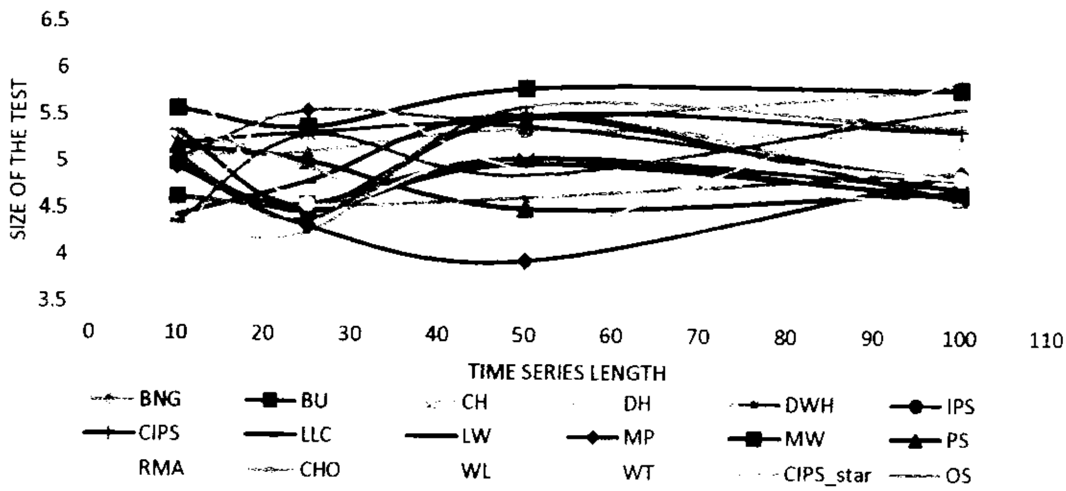
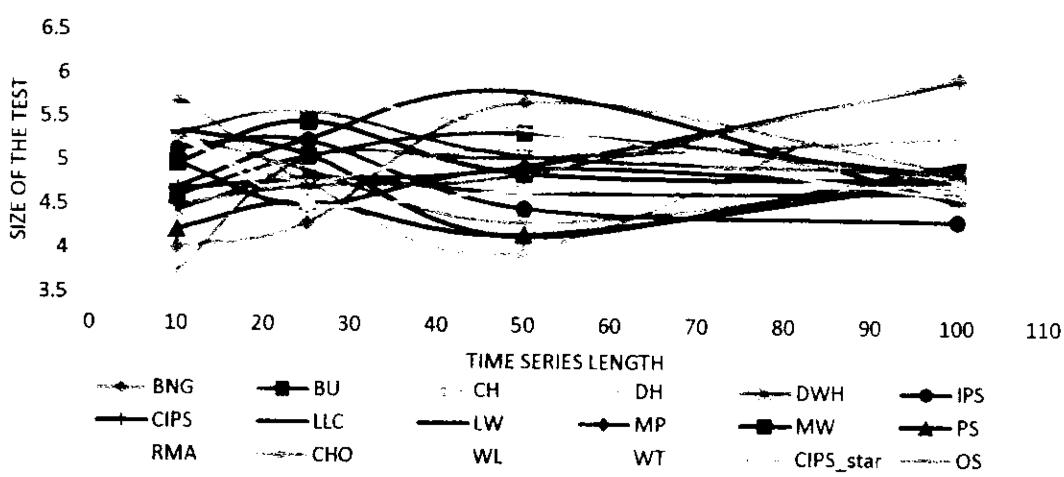



**Figure 5.25: Size of the Tests with Null Hypothesis of Unit Root and Intercept Case only, N=32**

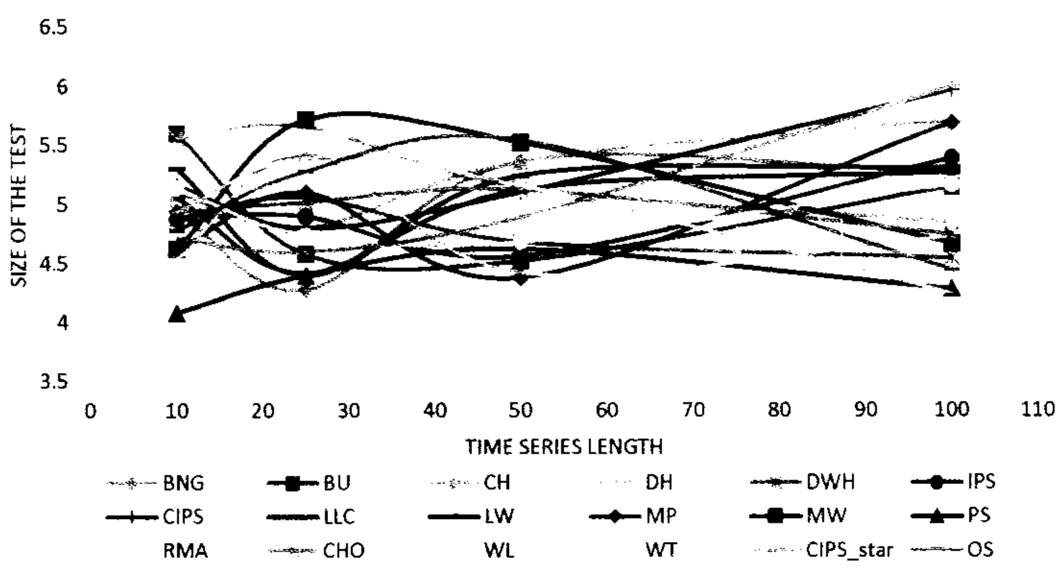


Clearly, Figure 5.2 to Figure 5.5 shows the size stability behavior of the PUR tests having the null hypothesis of unit root for N=4, 8, 16, and 32 and intercept case only. Like Figure 5.1, x-axis and y-axis represents the time dimension and size of the tests respectively. All these figures present a clear view of stabilizing size for all the cross section and time dimensions. In these figures, a stabilize size ranges from 4% to 6% by using a simulated critical value.

## 5.3. Stabilized Size with the Null Hypothesis of Panel Unit Root (Intercept and Trend Case)

Figure 5.6 to Figure 5.10 explains the size of the PUR tests having the null hypothesis of unit root, when data are generated using both of the deterministic terms, using simulated critical value for cross sections N=2, 4, 8, 16, and 32 and time series dimension T=10, 25, 50, and 100.

**Figure 5.26: Size of the Tests with Null Hypothesis of Unit Root, Intercept and Trend Case (N=2)**



According to Figure 5.6, when N=2, a stable size of around nominal size of 5% has been observed for different level time series dimensions when data are generated both with intercept and trend terms. Also, a stabilized size of tests with fluctuation around 4% to 6% are observed for different level of time and cross section dimensions.

**Figure 5.27: Size of the Tests with Null Hypothesis of Unit Root, Intercept and Trend Case (N=4)**

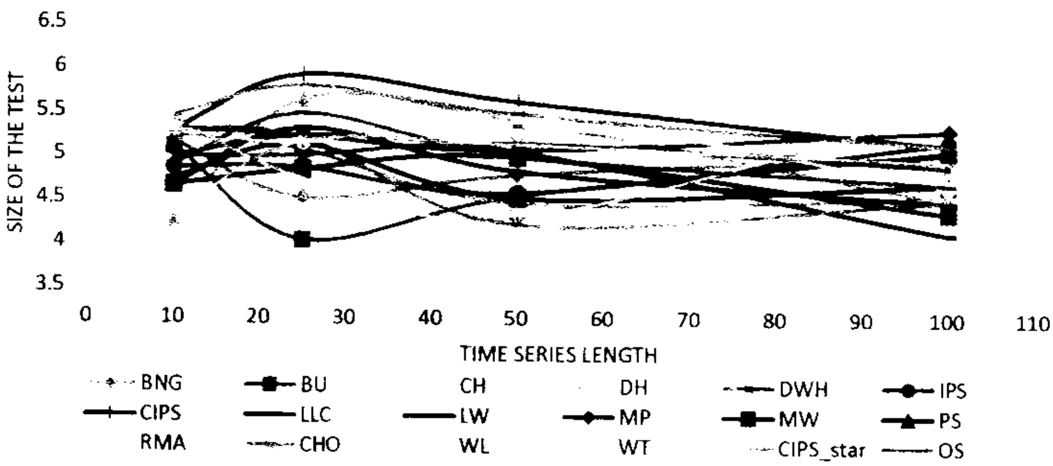
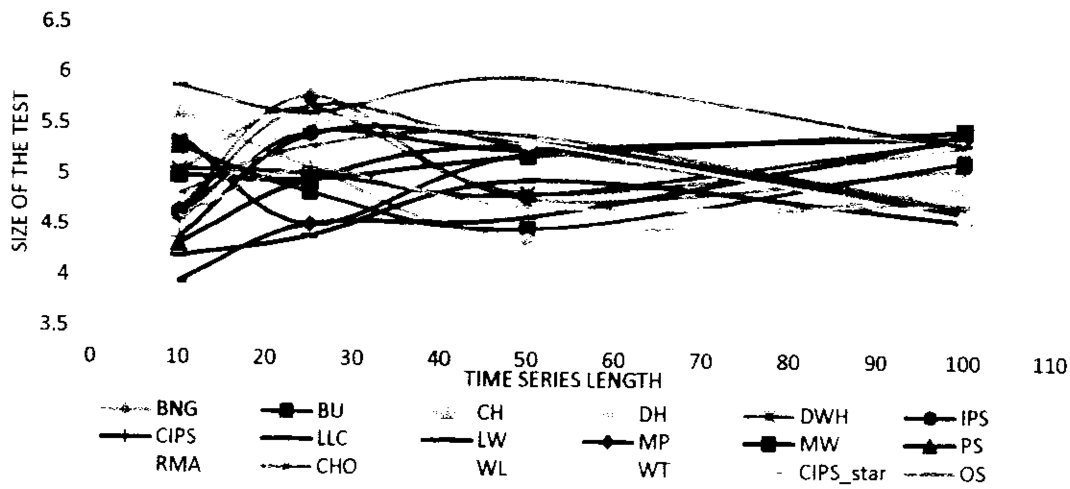



Figure 5.7 describes, when data are generated in the presence of both (intercept and trend terms) of deterministic terms, and simulated crucial value is calculated for different level of time series and cross section unit 4 for all PUR tests having the null hypothesis of unit root then size of PUR tests are approximately equal to nominal size of 5%. It can be observed that at time series 25, BU test has the smallest size of 4.04% and at the same time series level CIPS_star test has the largest size of 5.8% among all other time series levels when N=4.

**Figure 5.28: Size of the Tests with Null Hypothesis of Unit Root, Intercept and Trend Case (N=8)**




It is concluded from Figure 5.8 that all tests have an approximate size of nominal size of 5% for all sample sizes when size is calculated for N=8 using simulated critical value in the presence of both of the deterministic terms in the DGP. A minimal size of 3.95% for LLC test and a maximum size of 5.96% for OS test has been observed at time series level of 10 and 50, respectively. But both of them are not too far from nominal size of 5%.

Similarly, Figure 5.9 clearly shows the same type of variation of minimum size of 4% and maximum size of 6% around nominal size of 5%, when data are generated with intercept and trend terms for N=16.

Lastly, when data are generated for N=32, all PUR tests have an approximate fluctuation equal to nominal size of 5% as sample size varies from smallest (i.e. T=10) to largest (i.e. T=100). This result is presented in Figure 5.10.

**Figure 5.29: Size of the Tests with Null Hypothesis of Unit Root, Intercept and Trend Case (N=16)**

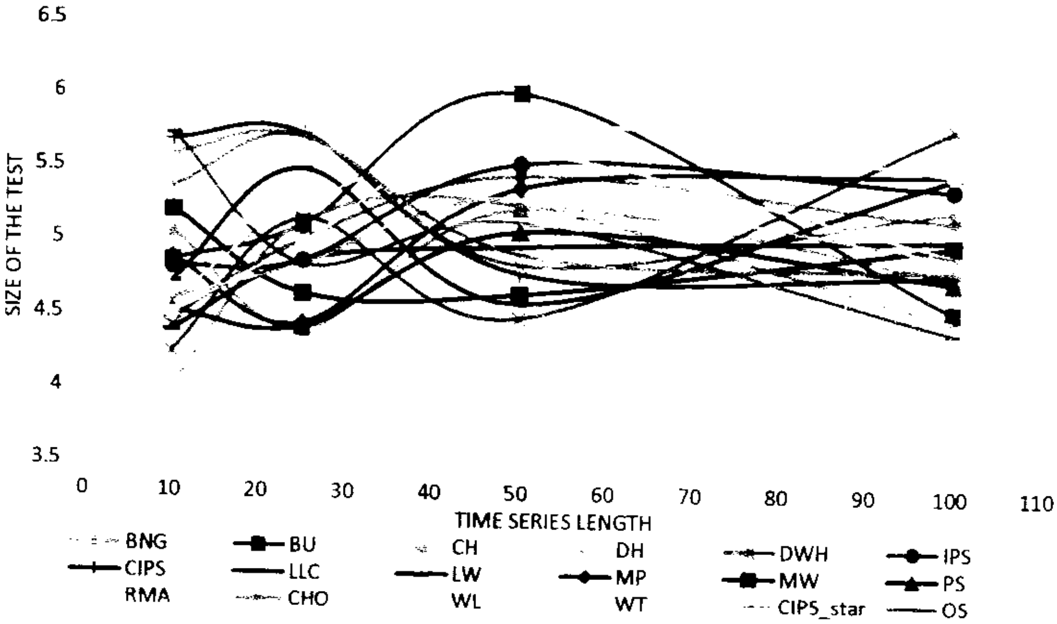
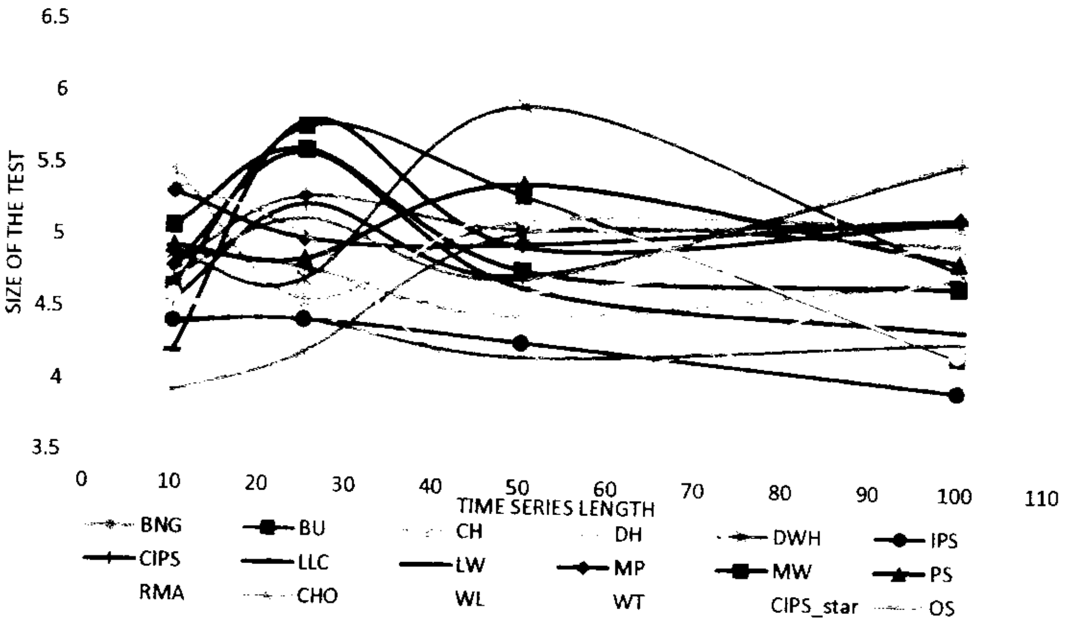



**Figure 5.30: Size of the Tests with Null Hypothesis of Unit Root, Intercept and Trend Case (N=32)**



Overall, the results for both cases (i.e. i- intercept case, ii-intercept and trend case) show that size of all tests for different cross section units and time series level remain stable around nominal size of 5%.

## 5.4. Stabilized Size with the Null Hypothesis of Stationarity (Intercept Case)

This section investigates size performance of PUR tests having the null hypothesis of stationarity when data are generated with intercept term only. Figure 5.11 to Figure 5.15 are given to analyze size of the six stationary tests for different cross section units with varying sample size from 10 to 100 using simulated critical values for each type of time series and cross section unit combination. It is clear that size of all tests (i.e. HD, HL, HLM, KK, SS, and DHT) varies from 4% to 6% (approximately) as the time series progress from small to larger, which suggests that size of all tests are equal to nominal size of 5%.

**Figure 5.31: Size of the Tests with Null Hypothesis of Stationarity, Intercept Case (N=2)**

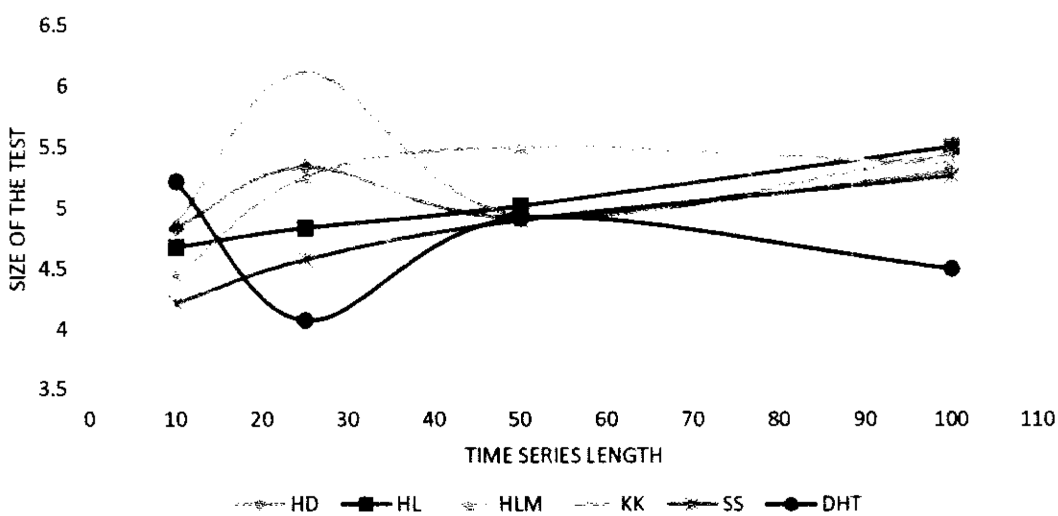
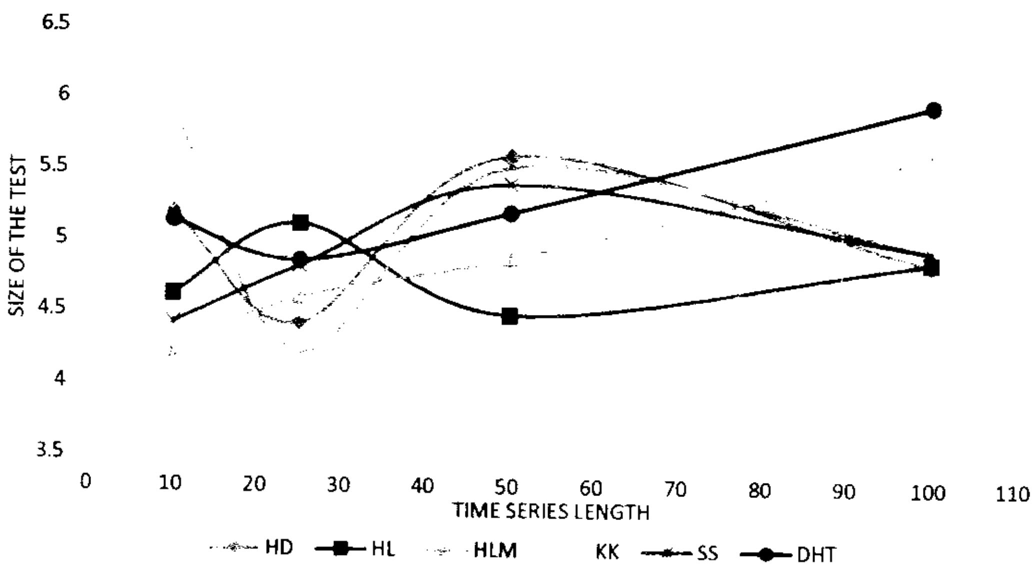


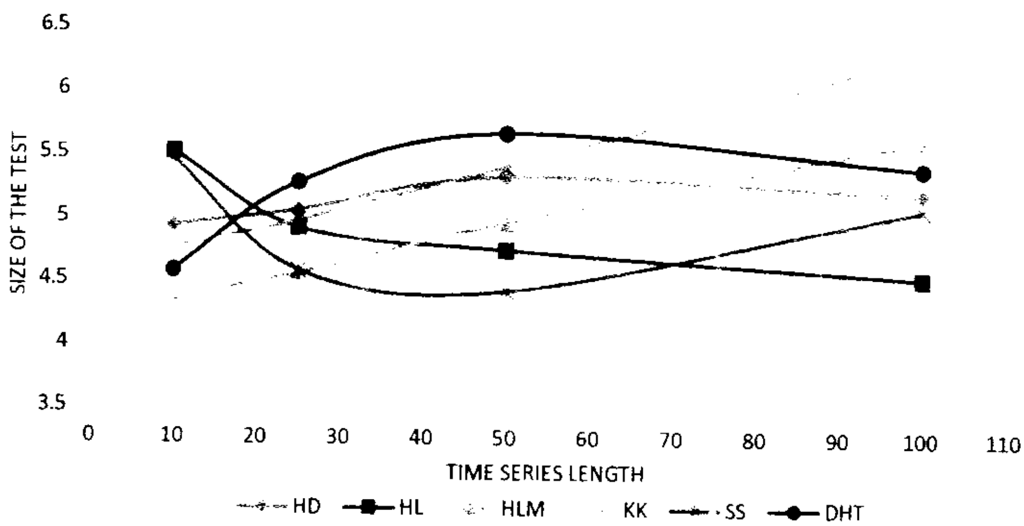
Figure 5.11 displays a graphical representation of HD, HL, HLM, KK, SS, and DHT tests have a size around nominal size of 5% for almost all sample sizes when N=2 and data are generated with intercept term only. Here, DHT test has the minimal size of 4.08% at sample size 25 and at the same sample size KK test has the maximum size of 6.12% among all sample sizes when N=2.

**Figure 5.32: Size of the Tests with Null Hypothesis of Stationarity, Intercept Case (N=4)**




At N=4 for sample sizes 10, 25, 50, and 100, Figure 5.12 shows that all panel stationary tests fluctuates around 4% to 6% when data are generated with intercept term only. HLM has the least size of 4.22% at sample size 10 while DHT test has greatest size of 5.96% at sample size 100 among all sample sizes.

**Figure 5.33: Size of the Tests with Null Hypothesis of Stationarity, Intercept Case (N=8)**




When number of cross sections is 8 then Figure 5.13 displays a similar picture as have been observed in the last two figures. Figure 5.13 indicates that HLM test has the minimal size of 4.36% at T=10 and KK test has the maximum size of 6.26% at T=100. Overall, Figure 5.13 demonstrate an approximate fluctuation of 4% to 6% at nominal size of 5% which represents a stabilized size to compare panel stationary tests using stringency criterion.

Figure 5.14 and Figure 5.15 concludes that HD, HL, HLM, KK, SS, and DHT tests have a stabilized size equal to nominal size of 5% when data are generated with intercept term only for N=16 and N=32. According to Figure 5.14, HL test has minimum size of 4.3% and KK test has maximum size of 6.36% around benchmark size of 5%. Similarly, Figure 5.15 shows that the least size of 4.1% has been taken by DHT test while KK test has the greater size of 6.12% at N=32.

Figure 5.34: Size of the Tests with Null Hypothesis of Stationarity, Intercept Case (N=16)

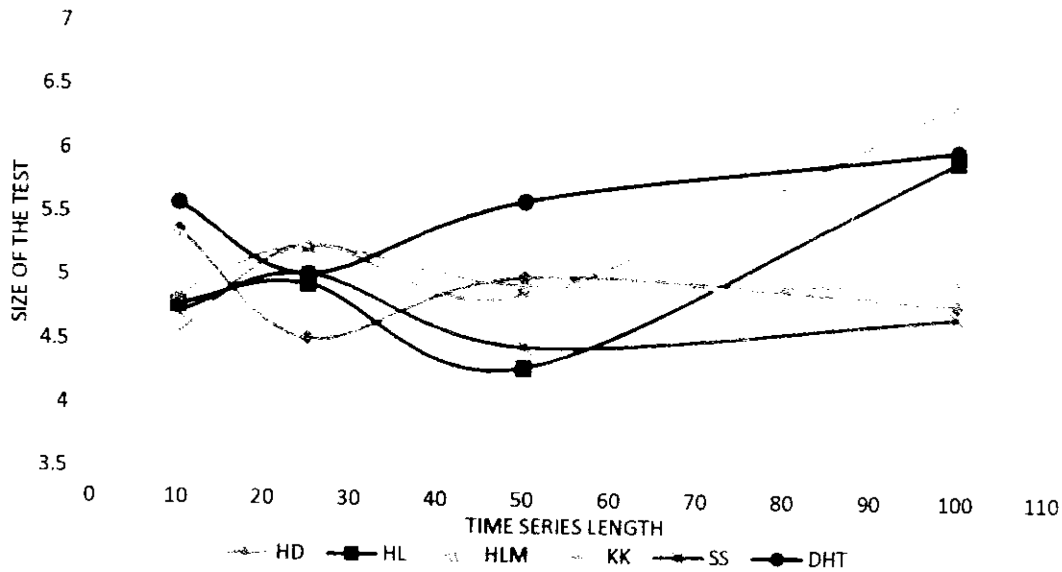
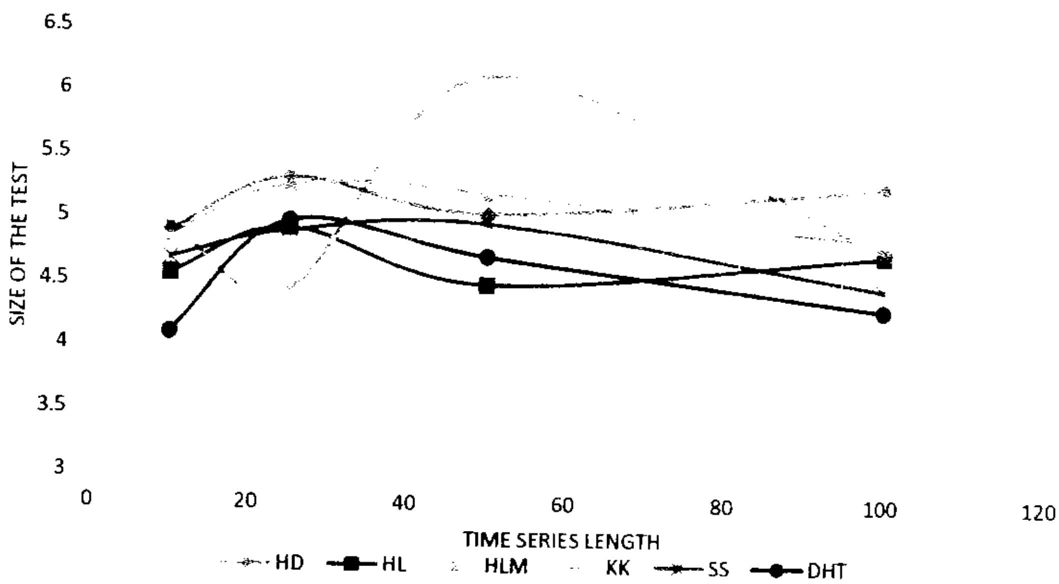



Figure 5.35: Size of the Tests with Null Hypothesis of Stationarity, Intercept Case (N=32)

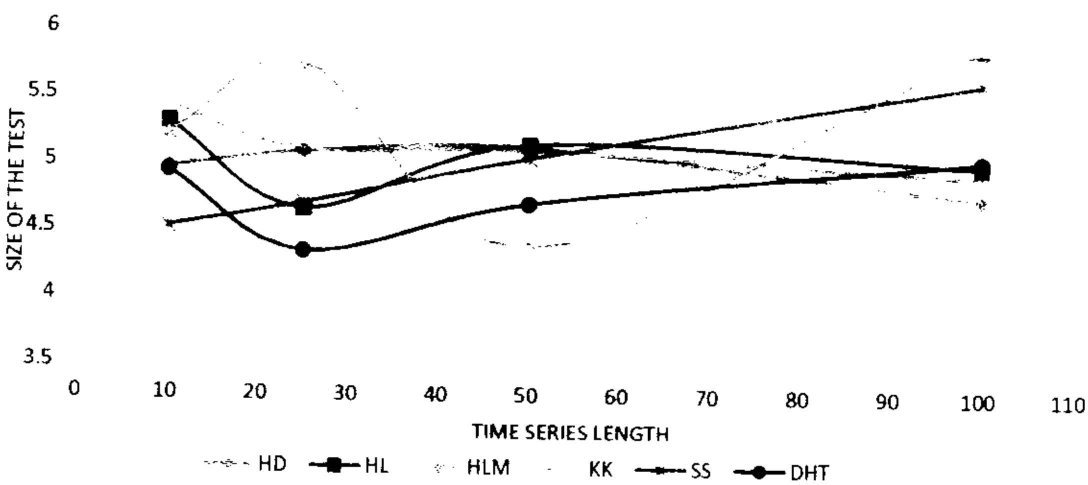


It is investigated that all tests having the null hypothesis of panel stationary fluctuate their size around nominal size of 5% for small, medium and large sample size with different level of cross section units.

## 5.5. Stabilized Size with the Null Hypothesis of Stationarity (Intercept and Trend Case)

This section examines size of panel stationary tests when data are generated in the presence of both of the deterministic terms using simulated critical values for different combination of time series and cross sections. Figure 5.16 to Figure 5.20 explains that all stationary tests approximate their size in between 4% to 6% which clarify that all these tests have size equal to nominal size of 5% at each combination of N and T.

**Figure 5.36: Size of the Tests with Null Hypothesis of Stationarity, Intercept and Trend Case (N=2)**




From Figure 5.16 for N=2, It can be determined that HD, HL, HLM, KK, SS, and DHT tests have stable size of around 5% (nominal size) for sample sizes 10, 25, 50, and 100 when data are generated with intercept and trend term.

**Figure 5.37: Size of the Tests with Null Hypothesis of Stationarity, Intercept and Trend Case (N=4)**

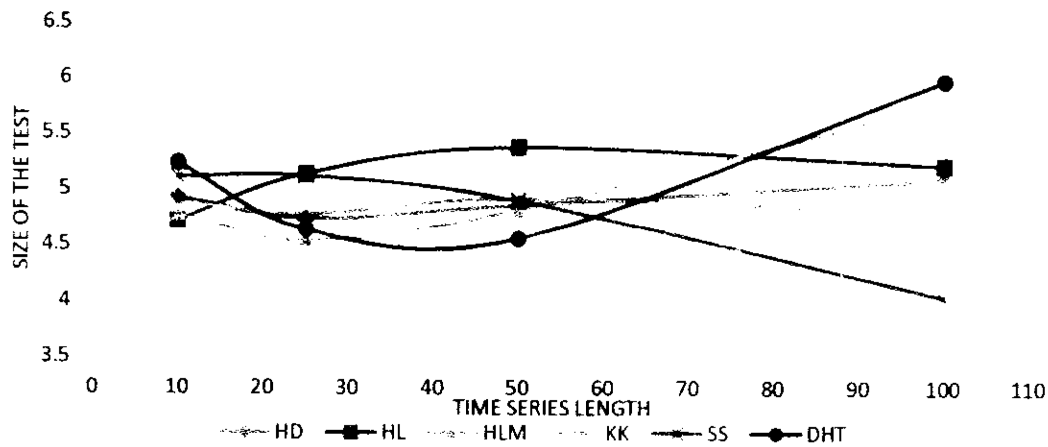



Figure 5.17 shows a very clear view that size of all panel stationary tests based on simulated critical value fluctuate between 4% to 6% when data are generated in the occurrence of both of the deterministic terms for N=4. Here, 4.04% and 5.98% are the lowest and largest size corresponding to SS and DHT tests at the same time level. Overall, all test have stabilized size to nominal size of 5% at N=4.

**Figure 5.38: Size of the Tests with Null Hypothesis of Stationarity, Intercept and Trend Case (N=8)**

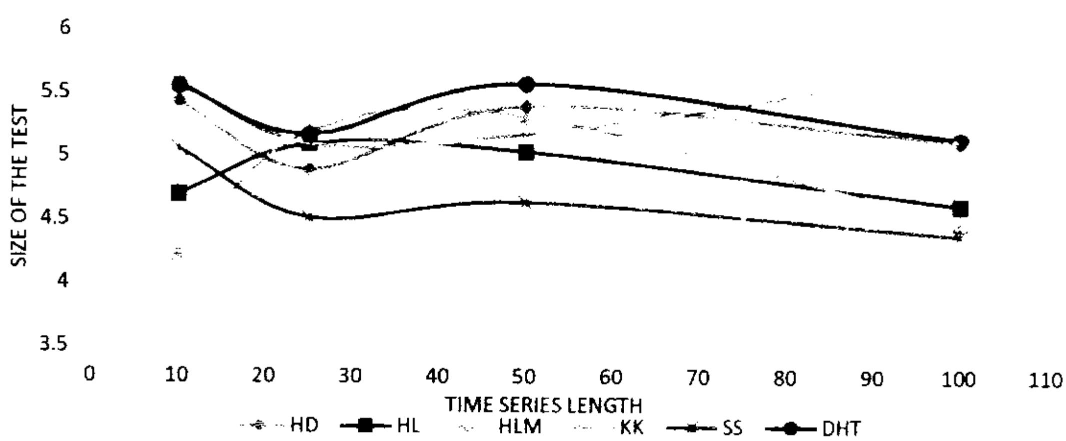


Figure 5.18 clearly portrays that all stationary tests have a stable size around nominal size of 5% when data are generated with intercept and trend terms by using simulated critical value. More ever, Figure 5.18 concludes that HLM and KK tests have minimum and maximum size at sample size 10 and 100 respectively and both of them are very close to usual nominal size of 5%.

**Figure 5.39: Size of the Tests with Null Hypothesis of Stationarity, Intercept and Trend Case (N=16)**

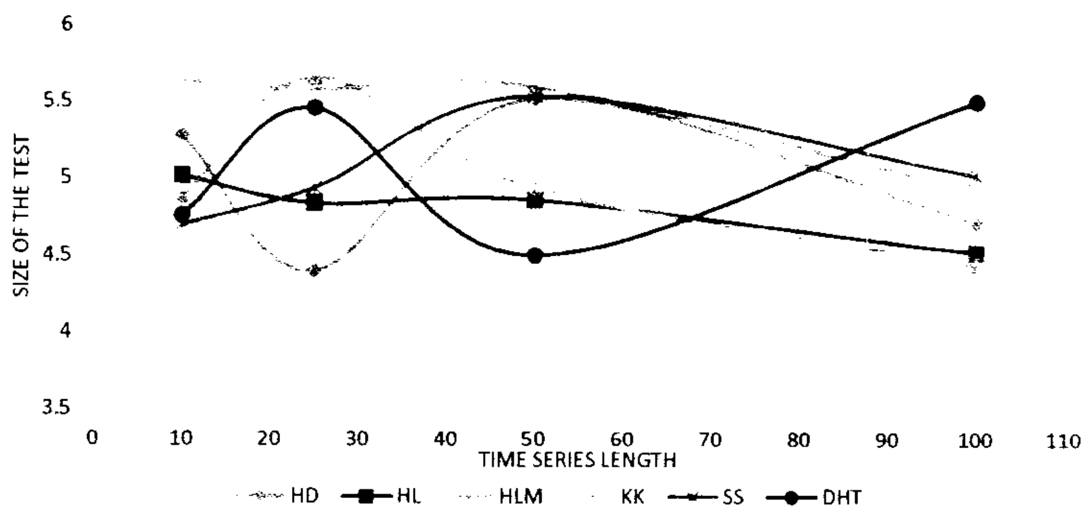
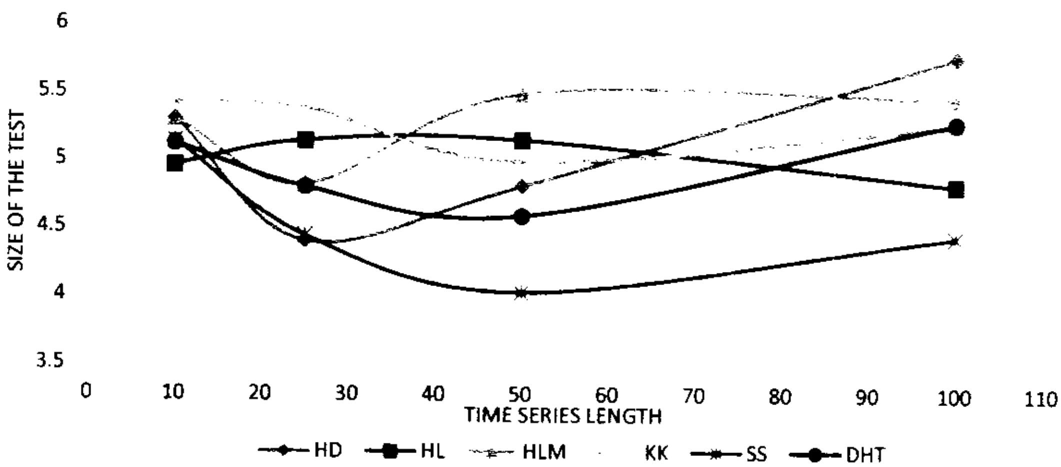



Figure 5.19 and Figure 5.20 indicate a similar graphical presentation as described by previous pictures when data are generated in the occurrence of both of the deterministic terms using simulated critical value when number of cross sections are 16 and 32, respectively. Both of the figures conclude that all panel stationary tests have an approximate stable size around nominal size of 5%.

**Figure 5.40: Size of the Tests with Null Hypothesis of Stationarity, Intercept and Trend Case (N=32)**




It is observed from Figure 5.16 to Figure 5.20 that HD, HL, HLM, KK, SS, and DHT tests have stabilized size around nominal size of 5% when data are generated in the presence of both of deterministic terms for varying number of number of time series and different cross section units.

Overall summary for both type of deterministic cases show that all tests having the null hypothesis of panel stationary have stabilize size approximately equal to nominal size of 5%.

## 5.6. Concluding Remarks

Overall results reveal that tests for the null hypothesis of unit root and stationary have stabilized size around nominal size of 5% when simulated critical value is used instead of asymptotic critical value. This evidence is observed in the presence of both type of deterministic models (i.e. with intercept, and with intercept and trend). Therefore, size of all PUR tests are stable around nominal size of 5% and according to stringency criterion both categories of PUR tests under simulated critical value can now have a meaningful comparison.

# Chapter 06

## Power Properties of Panel Unit Root Tests

### 6.1. Introduction

In Chapter 05, size of the PUR tests are calculated using simulated critical value to stabilize size of all tests at a nominal level size of 5%. In this chapter, power of PUR tests for the null hypothesis of unit root and stationarity are calculated using the same simulated critical value obtained from Chapter 05 for different level of time series and cross section dimensions along with both of deterministic parts. MP, PS, BNG, CH, CHO, LLC, BU, IPS, MW, CIPS_star, CIPS, DWH, RMA, LW, DH, WT, OS, and WL tests having the null hypothesis of PUR, and HD, HL, HLM, SS, KK, and DHT tests having the null hypothesis of panel stationarity are compared using stringency criterion for, both of the deterministic parts, different time series (T=10, 25, 50, 100) and cross section (N= 2, 4, 8, 16, 32) dimensions and using Monet Carlo Sample Size (MCSS) of 10,000. A long-run variance is calculated for majority of the tests for both of the categories. And an automatic bandwidth of Bartlett Kernel is used to find the long run variance.

Power curves are calculated for all the tests having the null hypothesis of PUR and power envelope obtained using the point optimal test having the null hypothesis of PUR. A maximum difference between powers envelop and power curve is obtained to find the MSC of each test to identify most stringent (best) test having the null hypothesis of PUR, for both deterministic models and, for different time series and cross section dimensions. A similar procedure is adopted for the null hypothesis of panel stationary tests to obtain most stringent test, for both deterministic models and, for different time series and cross section dimensions.

## 6.2. Tests Ranking

All tests are categorized into three types by analyzing the power performance of all tests using MSC. First, a test is considered to be best test if its MSC lie in between 0 and 10%. Second, a mediocre test satisfies the criteria if its MSC is in between 10% to 50%. Third, a worst test is defined to have MSC above 50%. In our analysis best tests and mediocre tests are shown by double and single stars in the tables.

The first part of this chapter analyzes the power of both type of PUR tests (i.e. tests having the null hypothesis of unit root and tests having the null hypothesis of stationary) when the time series dimensions varies while cross section dimensions remains fixed for both cases of deterministic parts. The second part of this chapter evaluates power of both types of PUR tests in the both cases of deterministic terms when time series dimension is fixed while cross section unit remains varying. In the last section, we discuss overall summary of the power comparison for both categories of tests.

## 6.3. Effect of Time Series Length on Power of Panel Unit Root Tests

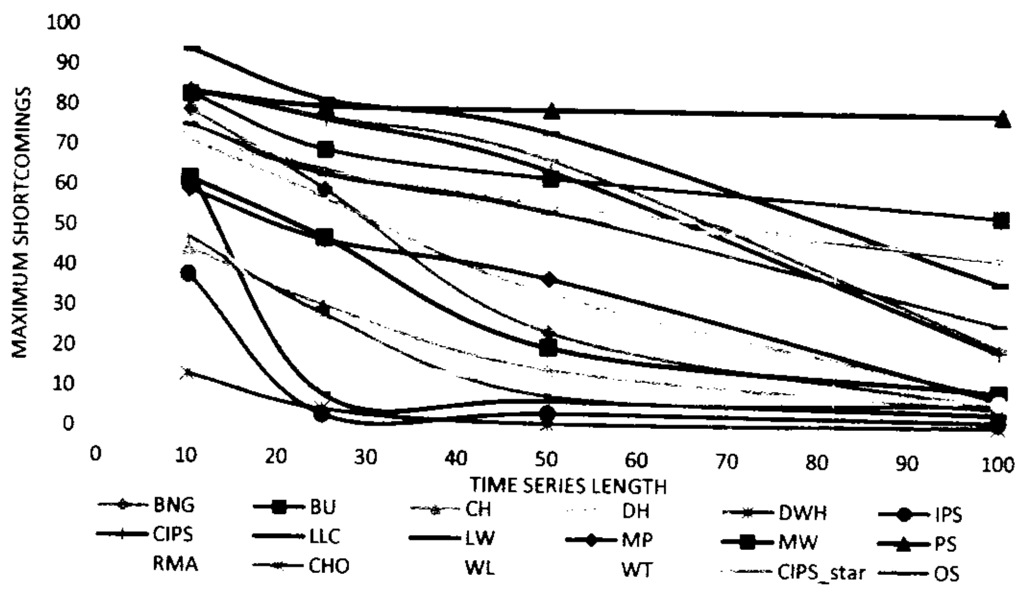
In this section, PUR and stationary tests are examined keeping cross section unit fixed while time dimension varies for the test models with drift term only and, with both drift and trend cases. First part of this section evaluates the power behavior of PUR tests for both cases (i.e. with drift only and, with both drift and trend) and the last part analyzes power of stationary tests in the presence of both of the deterministic terms.

### 6.3.1. Effect of Time Series on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Unit Root (Intercept Case only)

This section of the study evaluates power behavior of PUR tests by using MSC with null hypothesis of unit root for the specification of drift term only in the DGP and model of tests used.

In Figure 6.1 to Figure 6.5, it is observed that all tests follow a downward pattern for MSC corresponding to increasing pattern of sample size T=10, 25, 50, 100 and for different cross sections N=2, 4, 8, 16, 32 when data are generated with drift deterministic term only along with same specification for the model of the tests. The criteria for a best, mediocre and worst performer test has already defined, Figure 6.1 to Figure 6.5 include all these three type of tests. In all figures x-axis shows time series dimensions and y-axis represents MSC for each cross section unit.

**Figure 6.1: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept only, N=2**



At N=2 and T=10, Figure 6.1 and Table 6.1 (Panel-A) show that no single test fulfills the best test criteria but all tests show a convergence pattern as the time series length increases. At a small time series and cross section dimension all the tests have MSC greater than 30% excluding DWH test with 13% MSC. At the same cross section and time series dimension (i.e. T=10 and N=2), the number of worse performing tests are more as compared to mediocre tests, CH, DWH, IPS, and CHO tests are mediocre tests having MSC greater than 10% but less than 50%, while, BNG, BU, DH, CIPS, LLC, LW, MP, MW, PS, RMA, WL, WT, CIPS_star, and OS tests are the worst performing according to their MSC of greater than 50%. Hence, no single test is classified as best at small time series dimension (i.e. T=10).

At N=2 and T=25, an improvement has been observed for majority of the tests as compared to last time dimension. Analysis of Figure 6.1 shows that DWH, IPS, and LLC are best performing tests corresponding to their attained MSC as compared to their performance for time dimension 10. Among these three best tests, LLC test has attained a remarkable gain in its power and has ranked as best at T=25 from worst performing test previously at T=10. Similarly, at T=25, IPS test has also achieved a good power and has managed to classify as best test with MSC 3% which was 37% at time series length 10. While, CH, MP, MW, CHO, and WL are mediocre performing tests and all other tests are worst performing tests at the same level of time series and cross section dimension (i.e. T=25 and N=2). In these worst performing tests BNG, BU, DH, CIPS, LW, RMA, WT, CIPS_star, and OS tests have attained a significant declining gain in the pattern of their MSC as compared to their MSC of last sample size (i.e. T=10) while PS test shows a little improvement in its pattern.

Further increasing sample size (SS) to 50 and 100 at N=2, it is observed that a few more tests fulfill the criteria of best tests and those tests which performing poor at small sample sizes become better at large sample sizes according to their MSC. CHO test has become best test at sample size T=50 which was having MSC greater than 10% at small sample size of T=10 and T=25. In the mediocre test category for T=50, BNG and DH tests have shown a rapid decrease in their MSC among four other mediocre tests at the same time series level, as both of these tests placed in worst performing category for time series level 10 and 25. All other tests (i.e. BU, CIPS, LW, PS, RMA, WT, CIPS_star, and OS) have the same status of worst performing tests as have been observed for last two sample sizes, but showing a little improvement in their pattern.

At larger and final sample size of T= 100 for our simulation study when N=2, results of Figure 6.1 and Table 6.1 (Panel A) reveals improvement in the PUR tests having MSC greater than 10% for small sample sizes. BNG, CH, DH, MP, MW, and WL tests have further accelerated their positions at larger sample size of T=100 and named themselves as best performing tests besides other best tests for the previously small and medium sample sizes. In the mediocre tests category, CIPS, LW, RMA, WT, OS, and CIPS_star tests have further improved their performances among the previously mediocre ranked tests for sample sizes 10, 25, and 50. For all sample sizes at N=2, BU and PS tests show no improvement and remain the worst performing tests. Figure 6.1 indicates that majority of the best performing tests have attained their maximum power almost equal (very close) to the power of point optimal test as the sample size increases from 50 to 100. It is also observed that best, mediocre and one worst performing test tests have a downward pattern (convergence behavior) as sample size increases from T=10 to T=100 while PS (i.e. worst performing) test has no such a downfall pattern. Further, it is analyzed that as the sample size increases more tests become in the category of best performing tests and, the number of mediocre and worst performing tests decreases.

**Figure 6.2: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept only, N=4**

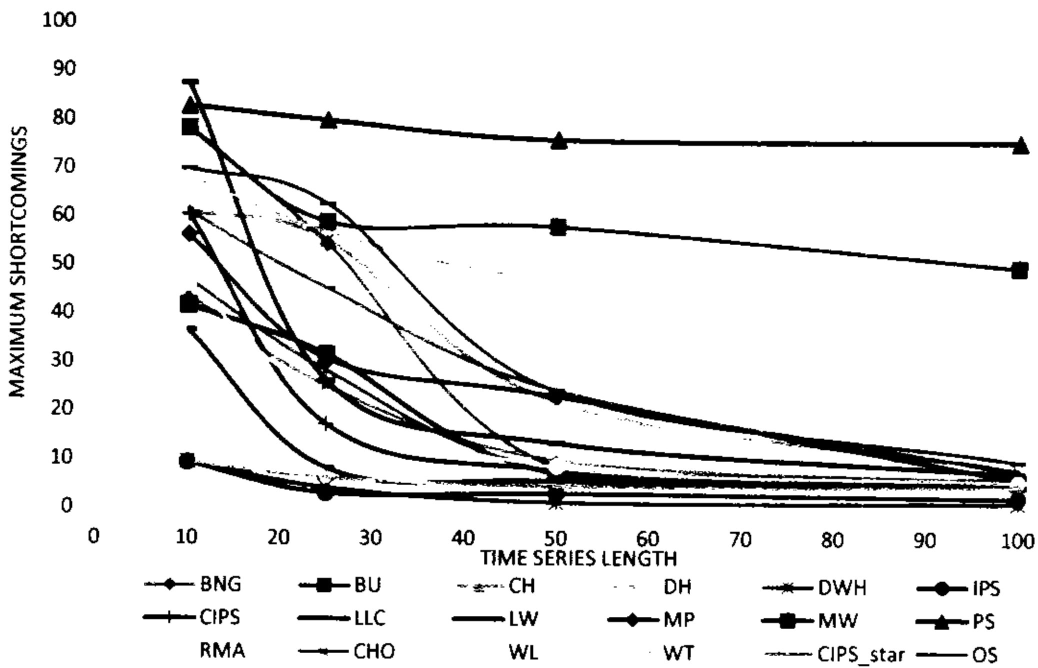



Figure 6.2 and Panel B of Table 6.1 show the results for N=4 with different sample sizes. Figure 6.2 reveals that DWH, IPS and WT are the best performing tests for sample size 10 having MSC of 4%, 3%, and 9%, respectively. At the same sample size, CH, LLC, MW, CHO, and WL are mediocre tests according to their attained MSC. The MSC of all other tests (i.e. BNG, BU, DH, CIPS, LW, MP, PS, RMA, CIPS_star, and OS) are greater than 50% and these are the worst performing tests at sample size 10.

It can be seen that at N=4 three tests ranked as best performing tests as compared to N=2, in which no single test is best performing, for the sample size 10. If we further observe

Figure 6.2 for the sample size 25, it is clear that more tests satisfy the best performing tests criteria. One of the mediocre tests (i.e. LLC) at sample size 10 has much improved its performance and become best performing test besides DWH, IPS and WT for sample size 25. At the same time series level, Figure 6.2 and Table 6.1 (Panel B) show that CH, CIPS, LW, MP, and CIPS_star tests have attained a downfall pattern with MSC of less than 50%, hence ranked as mediocre tests, as compared to their MSC at sample size 10. However, BNG, BU, DH, PS, RMA, and OS tests remain the worst performing tests for T=25 like for T=10. It is clearly seen from Figure 6.2 that one more test become part of the best performing tests category as the sample size increases from 10 to 25 while a few tests remain as worst performing tests for both of the samples. A similar analysis is seen for the mediocre tests for sample size 25 as compared to sample size 10.

A further larger sample size of 50 shows a very different picture as compared to previous sample sizes (i.e. T=10 and T=25) for N=4. At sample size 50 a more declining pattern (convergence behavior) of MSC is analyzed for majority of the tests (i.e. ten tests). Out of these ten best performing tests, DWH and IPS tests have power very close to the power of point optimal test at T=50. Also, among these best performing tests (BNG, CH, DWH, IPS, CIPS, LLC, MW, CHO, WL, WT), BNG test has remarkably achieved higher power and ranked as best performing tests from the previously assigned position of worst performing test for time series level of 10 and 25. Moreover, DH and RMA tests corresponding to their MSC remain mediocre tests in which DH test has improved its position from worst to mediocre test as compared to previous sample sizes of 10 and 25. However, BU and PS tests remain in the same status (i.e. worst performing tests) as have been observed for sample sizes 10 and 25.

When sample size reaches to 100 for N=4, the number of best performing tests increases from ten to fifteen (i.e. BNG, CH, DWH, DH, IPS, CIPS, LLC, LW, MP, MW, CHO, WL, WT, CIPS_star, and OS) as compared to best performing tests for time dimension 50. BU test has ranked into midcore performing test from worst performing test according to its percentage MSC. However, only PS test remains worst performing test for N=4 and for all sample sizes. All the tests show a good convergence behavior towards zero at large sample size as compared to previous sample sizes of 10, 25, and 50, excluding BU, RMA, and PS tests.

It is observed from Figure 6.2 that as the time dimension increases from small to large, the best performing tests achieve maximum power which is very close to the power of point optimal test. It is also noticed from Figure 6.1 and Figure 6.2 that at large time series level (T=100) the number of best performing tests increases to fifteen from ten. Similarly, at the same level of time series, both of the figures reveal that the number of mediocre and worst performing tests also decreases. Overall, Figure 6.2 shows that DWH, IPS, and WT tests perform well at small, medium and large time series dimensions. Moreover, LLC test also performs better at T=10 among mediocre performing tests while it is identified as best performing test for the remaining time series. However, majority of the other tests are assigned as best performing tests only when time series level is greater than 25.

**Table 6.1: MSC for Null Hypothesis of Unit Root in the Presence of Intercept Case, at N=2 and N=4**

| | Panel A: N=2 | | | | Panel B: N=4 | | | |
|---|---|---|---|---|---|---|---|---|
| Test/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| BNG | 79 | 59.42 | 24.06* | 6.5** | 78.14 | 54.34 | 8.54** | 6.1** |
| BU | 82.9 | 69.3 | 62.5 | 53.48 | 78.22 | 58.84 | 57.84 | 49.5* |
| CH | 44.58* | 30.5* | 14.9* | 5.7** | 43.74* | 25.08* | 9.9** | 5.51** |
| DH | 71.42 | 57.22 | 33.78* | 9.08** | 61.08 | 55.02 | 21.78* | 8.08** |
| DWH | 13.28* | 4.52** | 1.3** | 1.2** | 9.48** | 4.18** | 1.14** | 1.03** |
| IPS | 37.92* | 3.24** | 3.84** | 2.5** | 9.32** | 3.02** | 2.98** | 2.08** |
| CIPS | 83.78 | 76.8 | 64.18 | 19.92* | 60.46 | 17.14* | 7.92** | 5.6** |
| LLC | 63.03 | 8.38** | 7.02** | 4.5** | 36.62* | 8.2** | 5.58** | 5.29** |
| LW | 94 | 81.64 | 73.72 | 36.9* | 87.5 | 26.08* | 13.3* | 7.1** |
| MP | 59.14 | 46.8* | 37.38* | 8.14** | 56.16 | 30.62* | 22.93* | 5.92** |
| MW | 61.92 | 47.28* | 20.28* | 9.9** | 41.68* | 31.54* | 7.1** | 5.7** |
| PS | 83.54 | 80.06 | 79.48 | 78.82 | 82.78 | 79.8 | 75.74 | 75.2 |
| RMA | 75.1 | 64.28 | 53.62 | 42.86* | 68.23 | 56.86 | 46.78* | 40.67* |
| CHO | 47.36* | 28.34* | 8.28** | 6.5** | 47* | 28.27* | 7.9** | 5.3** |
| WL | 52.42 | 22.84* | 11.1* | 7.7** | 46.12* | 22.48* | 8.4** | 5.5** |
| WT | 72 | 63.92 | 56.04 | 42.98* | 9.42** | 6.24** | 4.5** | 4.4** |
| CIPS_star | 83.48 | 77.3 | 66.98 | 21.06* | 60.88 | 45.21* | 24.21* | 7.5** |
| OS | 75.3 | 63.04 | 54.1 | 26.6* | 69.88 | 62.58 | 23.9* | 9.6** |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.3: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept only, N=8**

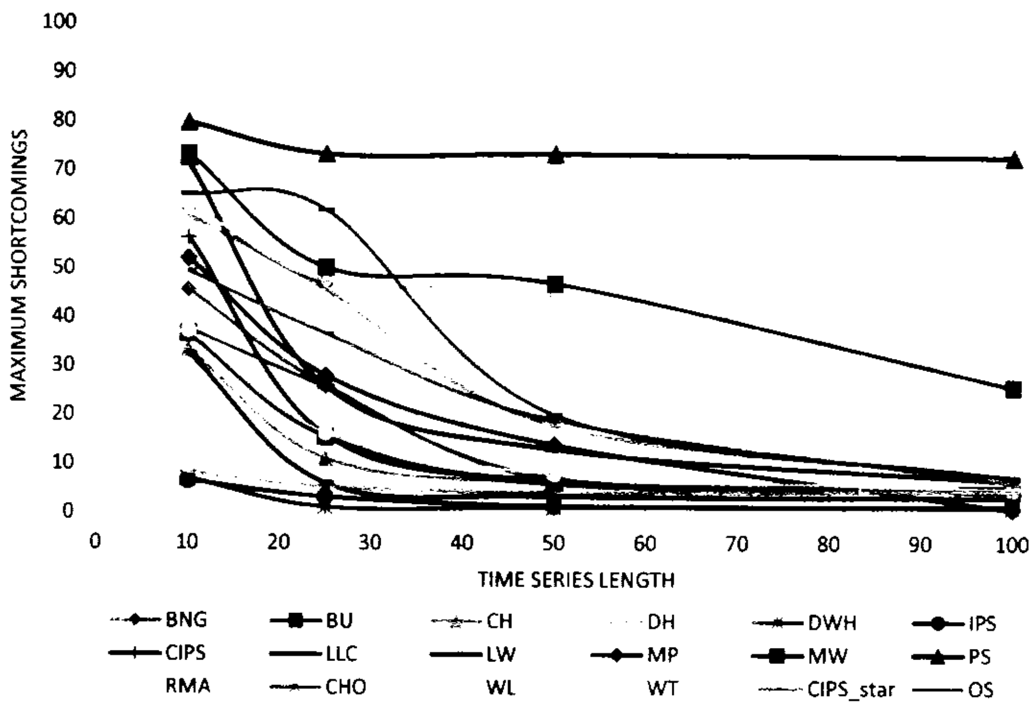


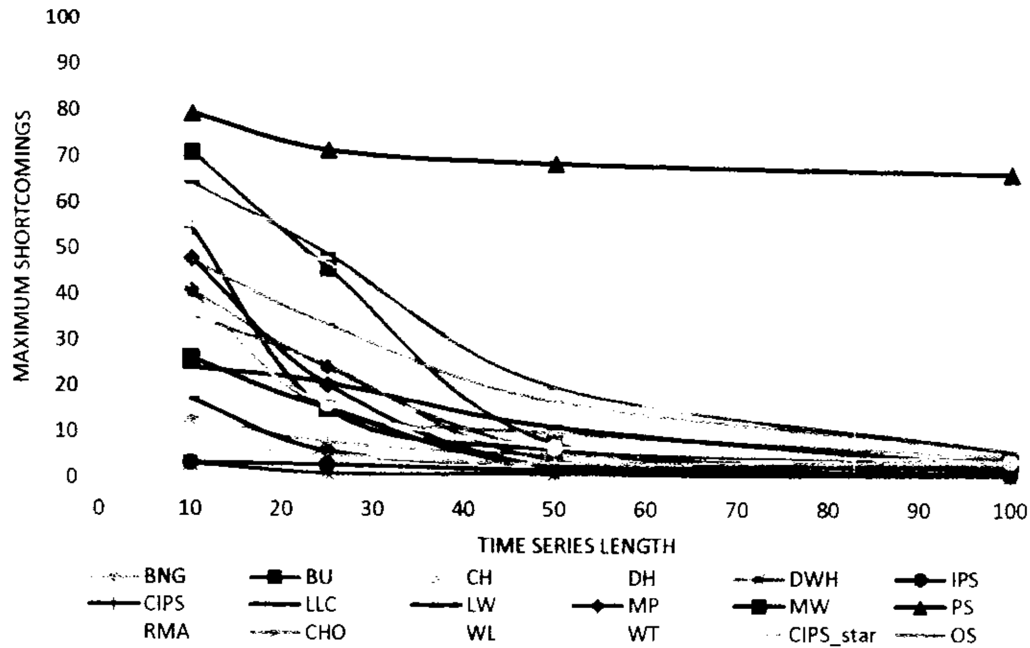

Figure 6.3 and Table 6.2 (Panel A) show a convergence pattern for majority of the tests as the sample size increases from 10 to 100 for N=8. Staring from sample size 10, Figure 6.3 reveals that DWH, IPS, and WT as best performing tests at that sample size. While, BNG, CH, LLC, MW, CHO, WL, and CIPS_star tests have ranked into mediocre tests according to their achieved MSC. However, all other eight (i.e. BU, DH, CIPS, LW, MP, PS, RMA, and OS) tests have categorized as worst performing tests at sample size 10.

At sample size 25, LLC test has categorized as best performing test besides DWH, IPS, and WT tests according to its behavior of MSC. Moreover, LLC test has attained a huge downfall in its pattern as compared to its MSC at last sample size of 10. Six of the worst performing tests (i.e. BU, DH, CIPS, LW, MP, and RMA) for sample size 10 have attained good pattern for sample size 25 and assigned as mediocre tests for that sample size. Lastly, two (i.e. PS and OS) tests also perform very poorly at sample size 25 as we have observed their same status at sample size 10.

For sample size 50 and 100, a very similar pattern can be observed for most of the tests besides MP and OS tests. These two tests have MSC less than 10% for time series dimension 100 as compared to their MSC of less than 50% at time series level 50. At both sample sizes, PS test has MSC almost with a constant pattern and ranked as worst performing tests. At sample size 50, ten tests (i.e. BNG, CH, DWH, IPS, CIPS, LLC, MW, CHO, WL, and WT) while at sample size 100 fifteen tests (i.e. BNG, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, CHO, WL, WT, CIPS_star, and OS) have ranked as best performing tests according to their MSC which shows that the number of best performing tests increases as the sample size increases. Moreover, BU, DH, LW, MP, RMA, CIPS_star, and OS tests have assigned as mediocre tests with respect to their MSC at T=50. Further, as the time series level increases to 100 then the number of mediocre tests decreases and only three tests (BU, PS, and RMA) are ranked as mediocre tests.

It is observed from Figure 6.3 that majority of the PUR tests having the null hypothesis of unit root are ranked as best performing test at large sample sizes when the number of cross sections are 8.

**Figure 6.4: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept only, N=16**




A further improvement in the declining pattern (convergence behavior) of PUR tests can be observed in Figure 6.4 and also from Table 6.2 (Panel B) for N=16. At sample size of 10, DWH, IPS and WT tests have MSC less than 10% and categorized as best performing tests. At the same time series level, BNG, CH, DH, LLC, LW, MP, MW, CHO, WL, and CIPS_star tests have MSC in between assigned benchmark of MSC for the mediocre tests and all of these ten tests have ranked as mediocre tests. However, all other tests (i.e. BU, CIPS, PS, RMA, and OS) have worst performance with MSC greater than 50% at sample size 10.

As sample size increases from 10 to 25, more tests make their presence in the best performing tests category. These include, CH and LLC tests with MSC lie in the benchmark of MSC considered for a best performing test. Figure 6.4 also indicates that powers of DWH and IPS tests have almost become equal to the power of a point optimal test as the sample size increases. While BU, CIPS, RMA, and OS tests have shown improvement as well and have ranked as mediocre tests at time series level 25 as compared to their MSC for the last sample size. At sample 25, only PS test has MSC greater than 50% and ranked as worst performing tests.

At large sample size of 50 and 100, thirteen (i.e. BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, MP, MW, CHO, WL, and WT) and sixteen (i.e. BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, CHO, WL, WT, CIPS_star, and OS) tests have concluded as best performing tests, respectively. Further, LW, RMA, CIPS_star, and OS tests have MSC less than 50% but greater than 10% while only one test (i.e. RMA test) has a MSC of less than 50% but greater than 10% have categorized as mediocre tests for sample size 50 and 100, respectively. At sample size 50, PS test has worst performance having MSC of 68%. While increasing sample size to 100, the same test (PS) remains the worst performing test. Excluding the worst performance of PS test, overall analysis of the pattern in Figure 6.4 and Table 6.2 (Panel B) conclude that as sample size increases MSC of majority of the tests decreases and at last reaches to almost zero showing the good performance of these tests.

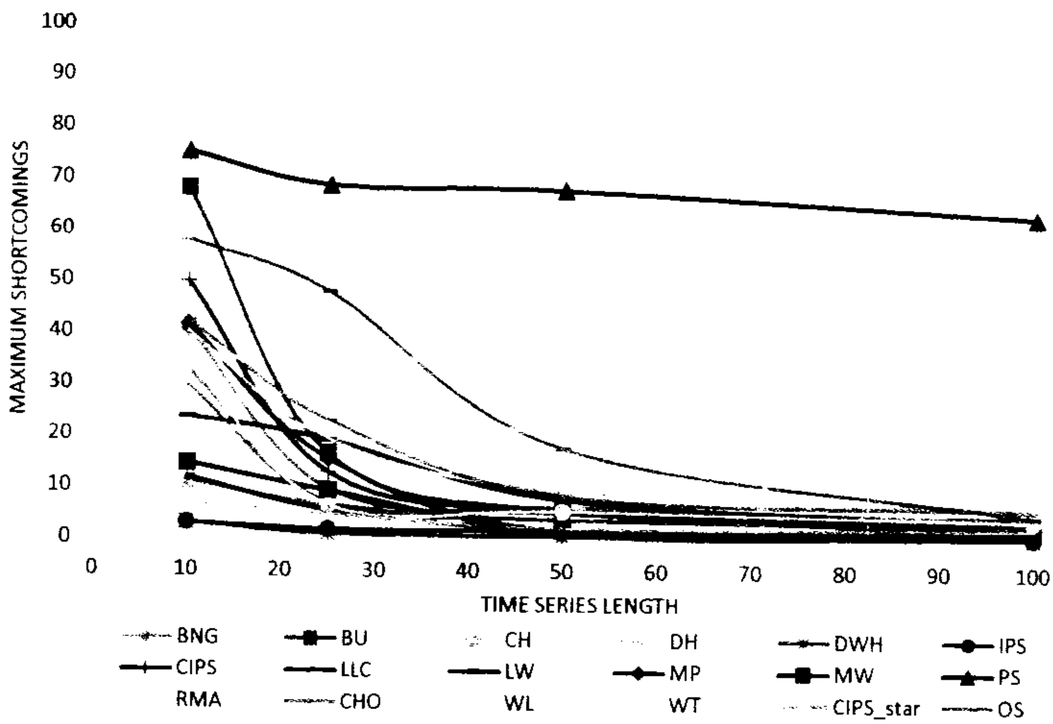
Figure 6.4 indicates that majority of the tests performs well and are ranked best performing tests at large time series levels. Also, PS test is the only worst performing test but with further increase of time series (i.e. greater than 100) this test may get good power and will eventually ranked into mediocre class.

**Table 6.2: MSC for Null Hypothesis of Unit Root in the Presence of Intercept Case, at N=8 and N=16**

| | Panel A: N=8 | | | | Panel A: N=16 | | | |
|---|---|---|---|---|---|---|---|---|
| Test/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| BNG | 45.64* | 26* | 7.04** | 5.2** | 40.93* | 24.26* | 6.22** | 2.1** |
| BU | 73.3 | 50* | 46.8* | 25.64* | 71.07 | 45.79* | 7.64** | 2.72** |
| CH | 33.46* | 11.08* | 6.1** | 5.08** | 13.44* | 7.9** | 5.2** | 4.3** |
| DH | 60.76 | 45.5* | 17.79* | 7.34** | 40.34* | 15.18* | 9.9** | 5.01** |
| DWH | 6.96** | 1.11** | 1.02** | 1.08** | 3.22** | 1.16** | 1.04** | 0.8** |
| IPS | 6.54** | 3.19** | 1.39** | 1.27** | 3.4** | 2.97** | 1.9** | 1.2** |
| CIPS | 56.34 | 15.9* | 6.4** | 4.1** | 54.42 | 14.96* | 6.38** | 4.1** |
| LLC | 32.62* | 6.03** | 3.3** | 3.1** | 17.3* | 5.97** | 3.2** | 2.4** |
| LW | 71.32 | 25.74* | 12.9* | 6.4** | 24.18* | 20.96* | 11.16* | 3.02** |
| MP | 52.04 | 27.99* | 13.88* | 0.6** | 47.89* | 20.48* | 2.46** | 1.18** |
| MW | 36.64* | 15.52* | 5.9** | 5.1** | 26.3* | 15.16* | 2.9** | 4.2** |
| PS | 79.68 | 73.2 | 73.08 | 72.26 | 79.46 | 71.5 | 68.5 | 66.3 |
| RMA | 63.28 | 47.12* | 43.84* | 19.06* | 55 | 46.82* | 22.62* | 11.04* |
| CHO | 37.62* | 25.56* | 6.6** | 4.7** | 35.54* | 24.67* | 4.01** | 3.13** |
| WL | 37.26* | 16.23* | 7.9** | 4.8** | 36.82* | 16.14* | 6.91** | 3.97** |
| WT | 8.06** | 5.33** | 4.1** | 4.38** | 7.84** | 5.2** | 3.3** | 3.81** |
| CIPS_star | 49.4* | 36.58* | 18.9* | 6.2** | 47.44* | 33.56* | 16.76* | 6.1** |
| OS | 65.23 | 61.84 | 20.01* | 7.2** | 64.42 | 48.88* | 19.94* | 6.1** |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.5: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept only, N=32**




At N=32, which is the last level of cross sections in our simulation study, for different time series levels (i.e. from small to large) a clear decline pattern (convergence behavior) is shown in Figure 6.5 for almost all tests. In Table 6.3 and from Figure 6.5 it is observed that only DWH, IPS, and WT tests have MSC less than 10% and have categorized as best performing tests for sample size 10. At the same time series dimension, BNG, CH, DH, CIPS, LLC, LW, MP, MW, RMA, CHO, WL, and CIPS_star have ranked as mediocre tests according to their MSC performances. According to worst performing tests criteria, BU, PS and OS have assigned as worst performing tests.

As the sample increases from 10 to 25, Figure 6.5 shows that all other tests have either ranked as best or mediocre performing tests according to their MSC criteria except PS test which is worst performing test. The MSC of BU test for sample size 25 is declined very fast and has ranked as mediocre test along with OS test as compared to their previous MSC for sample size 10. Now at sample size 25, there are ten tests (i.e. BNG, CH, DH, DWH, IPS, LLC, MW, CHO, WL, and WT) as compared to three tests for sample size 10 which has MSC less than 10% and are categorized best performing tests for that sample size. Moreover, most of the tests have gained higher power and have showed a decreased pattern in their MSC.

At time series level of 50, Figure 6.5 shows a good and smooth downward fall for almost all tests excluding PS test. At the same time series level, sixteen (i.e. BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, RMA, CHO, WL, WT, and CIPS_star) tests have adjusted themselves in best performing tests category. Only OS test has categorized as mediocre test according to its MSC of 17%. However, only one (i.e. PS) test has ranked into worst performing test showing a constant behavior over small time series level to time series level 50. At T=50 for N=32, the number of best performing tests have been increased as compared to the number of best performing tests for time series level 10 and 25. Also, the number of mediocre tests also decreased as the time series increases from 10 to 50.

Moving to time dimension 100 for cross section unit 32, Figure 6.5 further reveals that OS test has classified as best performing test corresponding to its MSC among BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, RMA, CHO, WL, WT, and CIPS_star best performing tests for the previous sample size 50. Among these seventeen best performing tests, majority of the tests have gained maximum power and have attained MSC equal to almost zero. Moreover, PS test has categorized as worst performing test with respect to its MSC of 62% at sample size 100.

Generally, analysis of all PUR tests from Figure 6.1 to Figure 6.5 and Table 6.1 to Table 6.3 have concluded the effect of time series on the power of PUR tests for different cross section units. Only, DWH, IPS, and LLC tests are ordered as best performing tests according to their MSC from small time series unit (i.e. T=10) to large time series level (i.e. T=100) when test equation and DGP has drift term only. Also, WT test has also shown such performance but at small time and cross section level (T=10 and N=2) this test has not achieved high power as compared to DWH and IPS tests. However, as the sample increase from smallest to largest the number of best performing tests also increases corresponding to each cross section unit. At the last sample size, the number of best performing tests reaches to seventeen at high level of cross section unit (i.e. 32). However, only PS test has not showed any improvement and has categorized as worst performing test at each time series level corresponding to all cross section units.

**Table 6.3: MSC for Null Hypothesis of Unit Root in the Presence of Intercept Case, at N=32**

| N=32 | | | | |
|---|---|---|---|---|
| Test/TL | 10 | 25 | 50 | 100 |
| BNG | 40.72* | 9.86** | 5.2** | 4.9** |
| BU | 68.18 | 16.64* | 5.68** | 2.4** |
| CH | 11.18* | 6.8** | 4.01** | 4.17** |
| DH | 32.88* | 7.1** | 6.4** | 4.5** |
| DWH | 3.2** | 1.1** | 0.81** | 0.87** |
| IPS | 3.16** | 2.01** | 1.66** | 0.88** |
| CIPS | 49.96* | 13.02* | 5.71** | 4.01** |
| LLC | 12* | 5.8** | 1.96** | 1.45** |
| LW | 23.85* | 19.28* | 7.9** | 3.6** |
| MP | 41.66* | 15.58* | 1.54** | 1.3** |
| MW | 14.66* | 9.5** | 2.05** | 2.01** |
| PS | 75.16 | 68.78 | 67.88 | 62.86 |
| RMA | 40.5* | 19.42* | 9.42** | 6.6** |
| CHO | 29.84* | 7.1** | 4.12** | 3.25** |
| WL | 33.92* | 6.6** | 5.7** | 4.15** |
| WT | 7** | 4.9** | 2.4** | 2.5** |
| CIPS_star | 42.38* | 23.13* | 8.63** | 6.2** |
| OS | 58.04* | 48.08* | 17.78* | 5.06** |

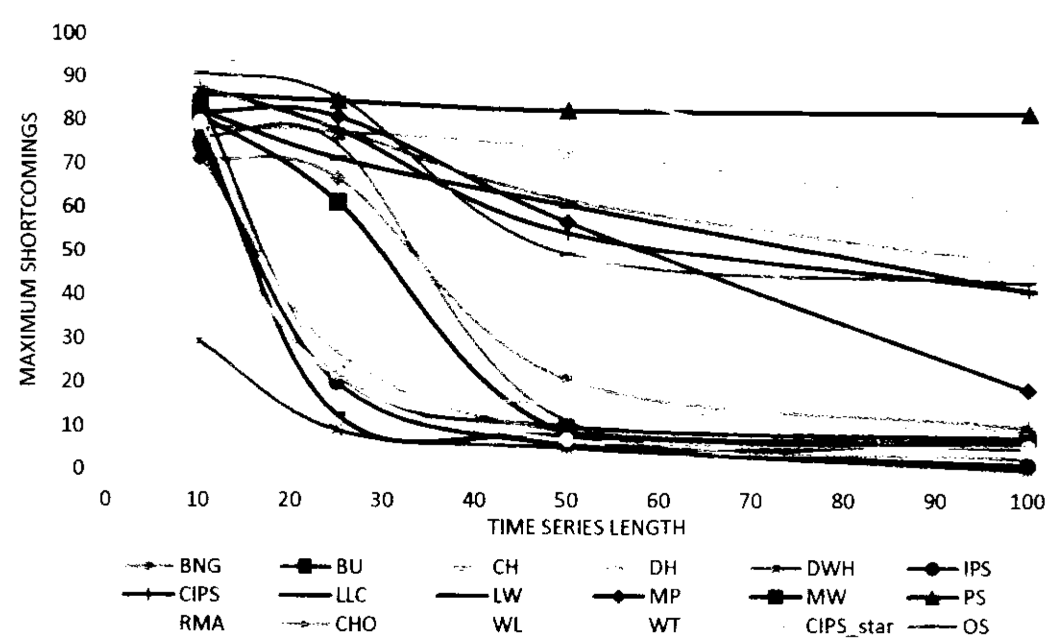
Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

### 6.3.2. Effect of Time Series on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Unit Root (Intercept and Trend Case)

This section analyzes the power performance of PUR tests having the null hypothesis of unit root when specification of drift and trend terms of PUR test equation matches with DGP.

In the presence of both of the deterministic terms in the data generating process for fixed cross section unit, that is, N=2, 4, 8, 16, and 32 for varying time series dimensions T=10, 25, 50, and 100 corresponding to each cross section unit, Figures 6.6 to Figures 6.10 shows a decreasing pattern for almost all tests having the null hypothesis of unit root. In each graph, x-axis and y-axis represent time series and MSC for different cross sections.

**Figure 6.6: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend, N=2**



At N=2, Figure 6.6 and Table 6.4 (Panel A) show a declining pattern (convergence behavior) for almost all of the tests as the sample size varies from smaller to larger. At a small sample size T=10, Figure 6.6 reveals that all tests are worst performing having MSC greater than 50%, excluding DWH test which has MSC 29% and stands the only mediocre test among all worst performing tests. In other words, at T=10, no single test is considered to be best performing according to attained MSC. Increasing sample size from 10 to 25, DWH test has MSC less than 10% and is the only best performing test for that sample size. Moreover, at T=25, BU, DH, IPS, LLC, WL, and WT have categorized as mediocre tests according to their MSC as compared to their performances for sample size 10. Figure 6.6 shows that at sample size 25 all other tests (i.e. BNG, CH, CIPS, LW, MP, MW, PS, RMA, CHO, CIPS_star, and OS) are worst performing as there is much gap between power curve and power envelope points.

At larger sample size of 50 for N=2, BU, DH, IPS, LLC, MW, and WL tests have classified into best performing tests beside DWH test. Among these best performing tests, MW test has maintained a high power and ranked as best performing test as compared to its status as worst performing test for the last two sample sizes (i.e. T=10 and T=25). While, BNG, CHO and OS tests are categorized as mediocre tests as compared to their MSC performance at sample size 10 and 25. However, with the same status for T=10 and T=25, CH, CIPS, LW, MP, PS, RMA, and CIPS_star tests have ranked as worst performing tests for time series level 50 as well. Among these worst performing tests for time series 50, CH, PS, and RMA tests have shown very little amount of improvement in their powers that have been observed for time series 25.

At a much larger sample size of 100, Figure 6.6 demonstrates a much decreasing pattern for almost all tests and ten (i.e. BNG, BU, DH, DWH, IPS, LLC, MW, CHO, WL, and WT) of them are ranked as best performing tests. While, approximately zero MSC of DWH and IPS tests show that both of these tests have almost equal power against power of point optimal test at sample size 100. At this sample size CIPS, LW, MP, CIPS_star, and OS tests have been considered as mediocre performing tests with respect to their MSC which have been ordered as worst performing tests for the previous sample sizes. While, CH, PS, and RMA have classified as worst performing tests with respect to their MSC at time series level 100. However among these worst performing tests, PS test shows no pattern of MSC and retains a constant pattern for all sample sizes (i.e. T=10, 25, 50, 100) indicating its worst performing behavior. Overall, Figure 6.6 indicates a convergence behavior according to MSC as the time series level moves from small to large.

**Figure 6.7: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend, N=4**

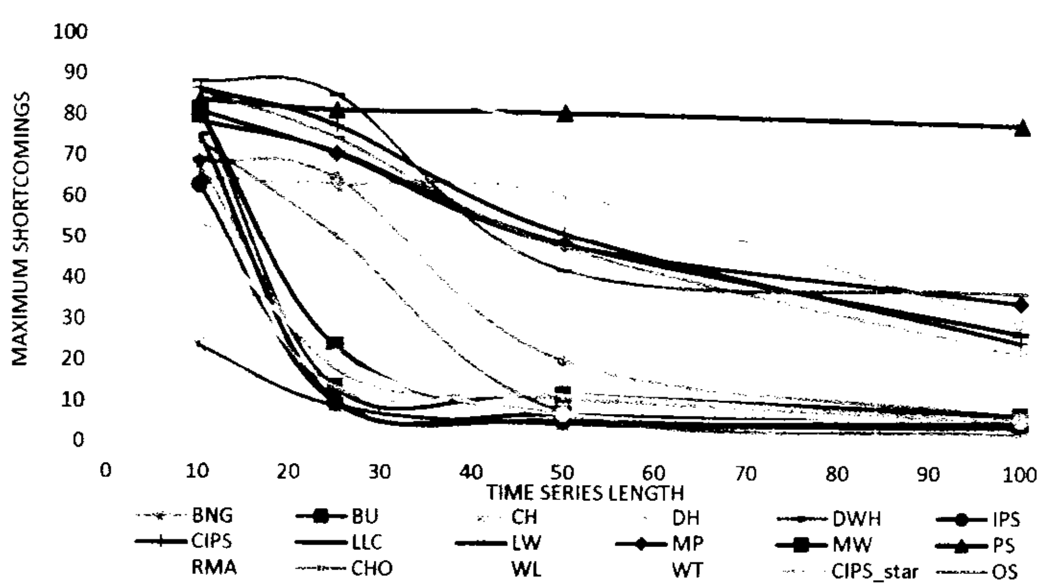


Figure 6.7 shows a convergence pattern for almost all tests as the sample sizes increases from 10 to 100 excluding PS test when N=4. This declining pattern is also mentioned in Table 6.4 (Panel B) corresponding to each sample size. In the given figure, DWH is showing a 23% MSC, which is the only one mediocre test at sample size 10. All other seventeen tests (i.e. BNG, BU, CH, DH, IPS, CIPS, LLC, LW, MP, MW, PS, RMA, CHO, WL, WT, CIPS_star, and OS) are worst performing tests with respect to their MSC for time series 10. However, the MSC of DWH test is less as compared to the MSC of this test at T=10 and N=2. Similarly, all other tests have the same situations for N=2 and T=10, even though all of them are worst performing tests.

Figure 6.7 also indicates that corresponding to sample size 25 for N=4, DWH, IPS, and LLC tests have MSC less than 10% and have classified as best performing tests. Here, IPS and LLC tests have remarkably gained a good power and ranked as best performing tests from the previously assigned worst performing tests for sample size 10. While, BU, DH, MW, CHO, WL, and WT tests are ordered as mediocre tests corresponding to their MSC. However, according to benchmark assigned for the worst performer tests, BNG, CH, CIPS, LW, MP, PS, RMA, CIPS_star, and OS tests get not less than 50% MSC and classified as worst performing tests. Moreover, at T=25 and N=2, only DWH test was categorized as best performing test while at the same time series level but N=4, the number of best performing tests increases to three (i.e. DWH, IPS, and LLC).

An analysis of Figure 6.7 for time series 50 corresponding to N=4 and Table 6.4 (Panel B) shows a further improvement in the performance of tests for all categories of tests. MW, CHO, and WL tests display a maximum shortcoming less than 10% alongside best performer tests (i.e. DWH, IPS, and LLC) for the previous sample sizes. While, BNG, BU, DH, LW, MP, CIPS_star, and OS tests have managed a decreased MSC of 20%, 12%, 11%, 48.9%, 48.7%, 47% and 42% at sample size 50 respectively as compared to their MSC for previous sample sizes and have ordered as mediocre test. However, CH, CIPS, PS, and RMA tests have graded as worst performing tests according to their assigned MSC for sample size 50.

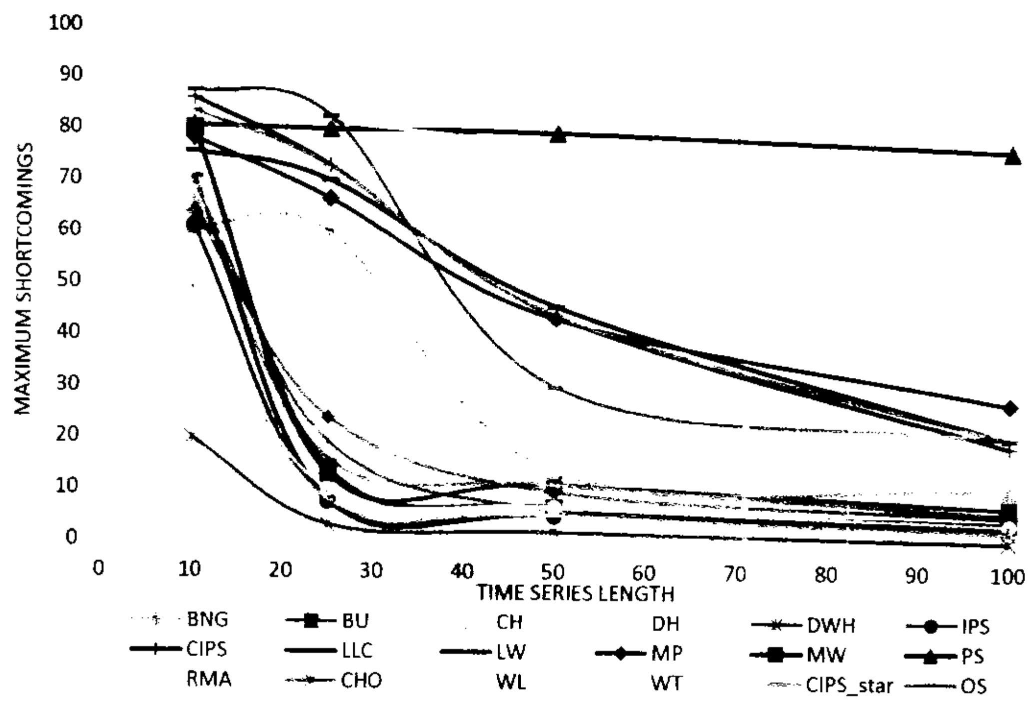
At a much higher sample size of 100, Figure 6.7 reveals that BNG, BU, DH, and WT tests induce themselves in the best performer tests category with a MSC of 6%, 7%, 5%, 5.6% respectively among all other best performer tests (i.e. DWH, IPS, LLC, MW, CHO, and WL) for the previous sample size 50. Two tests have further make their place in the mediocre tests category. These tests are CH and CIPS tests along with LW, MP, CIPS_star, and OS tests in the mediocre tests category. Moreover, PS and RMA tests are observed to be the worst performing tests for sample size 100. Although, RMA test shows a slight increasing pattern at sample size 100 as compared to its performance last three time series levels but PS test shows a very little improvement as the time series progress. Overall, Figure 6.7 and Table 6.4 (Panel B) determine that majority of the tests are either ranked into worst or mediocre tests at small sample size, however, as the sample size increases from small to large then almost of these PUR tests are categorized into best performing tests with respect to their attained MSC. While, only two tests (i.e. PS, and RMA) are ordered as worst performing tests for all sample sizes with respect to their MSC at N=4. Also, the performance of all tests at each level of time series for number of cross section 4 are more better as compared to N=2.

**Table 6.4: MSC for Null Hypothesis of Unit Root in the presence of Intercept and Trend Case, at N=2 and N=4**

| | Panel A: N=2 | | | | Panel B: N=4 | | | |
|---|---|---|---|---|---|---|---|---|
| Test/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| BNG | 71.3 | 66.94 | 20.96* | 10.01** | 68.92 | 64.7 | 20.2* | 6.88** |
| BU | 84.16 | 22.18* | 10.06** | 7.44** | 81.38 | 13.58* | 12.01* | 7.07** |
| CH | 78.64 | 77.48 | 73.26 | 59.06 | 64.78 | 63.42 | 61.18 | 29.81* |
| DH | 71.72 | 27.16* | 9.16** | 6.9** | 67.82 | 17.96* | 11.01* | 5.9** |
| DWH | 29.46* | 9.28** | 5.02** | 0.02** | 23.82* | 9.06** | 4.74** | 4.01** |
| IPS | 74.82 | 20.18* | 5.9** | 1.1** | 63.12 | 9.81** | 5.41** | 5.12** |
| CIPS | 87.54 | 77.96 | 54.32 | 41.26* | 86.46 | 77.6 | 51.36 | 24.94* |
| LLC | 77.8 | 12.82* | 8.02** | 5.01** | 75.28 | 10.42* | 7.02** | 5.01** |
| LW | 82.22 | 71.4 | 60.68 | 41.08* | 78.52 | 71.09 | 48.98* | 27.08* |
| MP | 81.96 | 80.98 | 56.84 | 18.44* | 80.92 | 70.78 | 48.7* | 34.6* |
| MW | 81.24 | 61.4 | 8.64** | 6.7** | 81.06 | 23.74* | 7.09** | 5.21** |
| PS | 86.03 | 84.62 | 82.38 | 81.77 | 83.72 | 81.42 | 80.9 | 78.18 |
| RMA | 93.89 | 91.9 | 86.46 | 65.46 | 93.87 | 90.18 | 77.76 | 61.26 |
| CHO | 76.28 | 74.6 | 11.3** | 2.58** | 73.74 | 50.7 | 7.88** | 2.3** |
| WL | 79.64 | 22.44* | 7.1** | 5.5** | 52.66 | 21.02* | 7.06** | 5.9** |
| WT | 61.34 | 21.6* | 15.16* | 8.86** | 56.16 | 13.56* | 11.9* | 5.63** |
| CIPS_star | 87.61 | 78.19 | 61.98 | 47.2* | 85.88 | 74.53 | 47.91* | 22.13* |
| OS | 90.98 | 85.4 | 49.46* | 43.08* | 88.34 | 85.08 | 42.28* | 37.06* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.8: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend, N=8**




At N=8 for different time series, Figure 6.8 and Table 6.5 (Panel A) categorize best, mediocre and worst performer tests according to their MSC. At sample size 10, no single test is categorized as best performing test while only DWH and WL tests fulfill the criteria of mediocre test at that sample size. Moreover, all other tests (i.e. BNG, BU, CH, DH, IPS, CIPS, LLC, LW, MP, MW, PS, RMA, CHO, WT, CIPS_star, and OS) have ranked into worst performing tests at sample size 10. While, the number of mediocre tests has increased to two at N=8 and T=10 as compared to N=4 and T=10.

Increasing time series 10 to 25, Figure 6.8 reveals that DWH, IPS, LLC, WL, and WT tests are best performing tests according to their MSC. Among these best performer tests, IPS, LLC, WL, and WT tests have obtained higher power thus attaining a lower MSC for T=25 as compared to their performances at T=10. Also, BU, DH, IPS, LLC, MW, CHO, WL, and WT tests improve their positions very fast from worst at T=10 to best performer at T=25. In the mediocre tests category for T=25, five tests (i.e. BNG, BU, DH, MW, and CHO) have fulfilled their criteria as compared to their worst performances for sample size 10. However, eight tests (i.e. CH, CIPS, LW, MP, PS, RMA, CIPS_star, and OS) are not able to get good position at sample size 25 and have ranked as worst performing tests with the same status that have been received for sample size 10. Moreover, all eighteen test have improved their MSC for N=8 and T=25 as compared to their behavior of MSC for N=4 and T=25.

Furthermore, Figure 6.8 shows that BNG, MW, and CHO tests have obtained MSC less than 10% and counted in the best performer test category for sample size 50 along with DWH, IPS, LLC, WL, and WT tests at sample size 25. At the same time series level, CH, CIPS, LW, MP, CIPS_star, and OS tests fulfill mediocre test criteria with 13%, 43.8%, 45%, 43%, 43.8%, and 30% MSC, respectively, besides BU and DH tests. While, only two tests (i.e. PS and RMA) have not achieved their positions at T=50 and have classified as worst performing tests. Overall, results indicate an over improvement of PUR tests at N=8 and time series dimension of 50 as compared to results of N=4 for the same level of time series.

At a larger sample size 100, Figure 6.8 and Table 6.5 (Panel A) demonstrate that BU, CH, and DH tests are ranked as best performer tests besides other eight tests (i.e. BNG, DWH, IPS, LLC, MW, WL, CHO, and WT) for the previous sample sizes. Moreover, the number of mediocre tests decreases as the sample size increases from 50 to 100. At time series dimensions 50 and 100, CIPS, LW, MP, CIPS_star, and OS tests have graded as mediocre tests. While, only RMA test is ranked as mediocre at T=100 which is worst performing test at T=50. Analysis of Figure 6.8 further shows that PS test has remained with the same status as this test has ranked as worst performing test for previous sample sizes (i.e. 10, 25, and 50).

Inclusively, Figure 6.8 and Table 6.5 (Panel A) indicate that at small time series level none of the tests performs well from best performer tests point of view but as the sample size increases from medium to large then the number of best performing tests also increases and at last sample size the number of best performing tests becomes eleven. Also, as the time series level moves from 10 to 100 then the number of worst performing tests become decreasing in numbers. Lastly, it is also observed that only one test (i.e. PS) remains in the worst performer test category at small, medium and large time series dimensions. Also, the performance of PUR tests have been improved and more tests have categorized as best performer tests at N=8 with all time series levels as compared to the results of N=4 with varying time series level.

**Figure 6.9: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend, N=16**

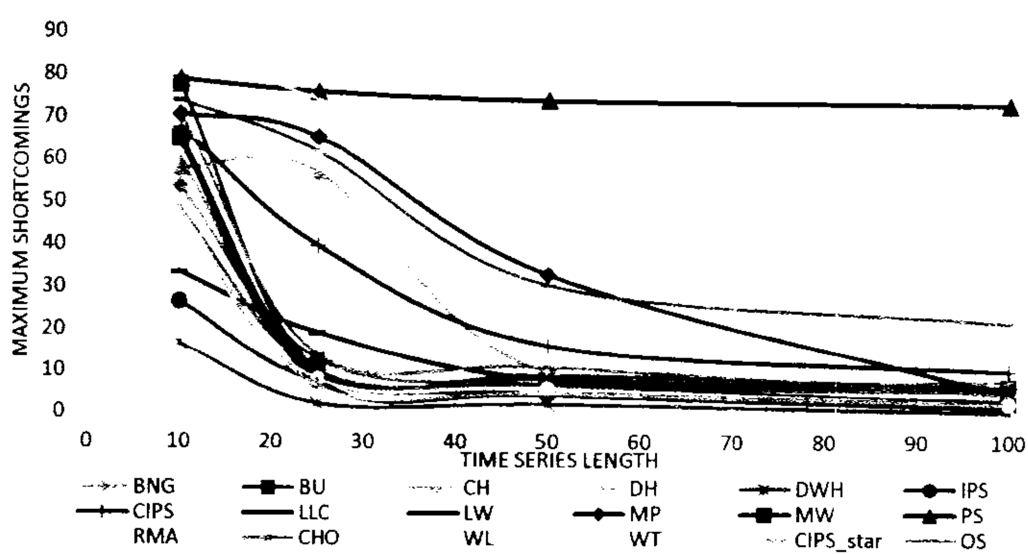


Figure 6.9 and Table 6.5 (Panel B) describe the declining phenomena for N=16 with varying time series levels. Starting from small time series 10, Figure 6.9 shows that no single test lies in the best performing test category while at the same time series level DWH, IPS, LW, CHO, WL, and WT tests have been categorized as mediocre tests. Moreover, twelve tests (i.e. BNG, BU, CH, DH, CIPS, LLC, MP, MW, PS, RMA, CIPS_star, and OS) classify into worst performing tests with respect to their MSC at time series dimension 10. But, with an increase time series level from 10 to 25, it is observed that nine tests are declared as best performing tests according to their attained MSC. These tests are BU, DH, DWH, IPS, CIPS, LLC, MW, CHO, WL, and WT with MSC 9.92%, 8.4%, 2.02%, 7.51%, 7.6%, 9.6%, 9.9%, 7.9%, and 6.3%, respectively. At the same time series level, BNG, CIPS, LW, and CIPS_star are assigned as mediocre tests. Moreover, six tests (i.e. CH, MP, PS, RMA, and OS) of the remaining tests have graded into worst performing tests with respect to their assigned MSC.

Further, increasing time dimension to 50 for N=16, Figure 6.9 demonstrates that besides BU, DH, DWH, IPS, LLC, MW, CHO, WL, and WT tests; BNG, CH, LW, and CIPS_star tests have also ordered into best performing tests with respect to their MSC. However, CIPS, MP, and OS tests have ranked into mediocre test while PS and RMA tests with no improvement in their positions are classified as worst performing tests at sample size 50. At the same time series level but cross section unit 16, the number of best performing tests are ranked as thirteen tests while at the same time series level with N=8 the number of best performing tests were eight.

When the time series dimension reaches to 100 corresponding to number of cross sections 16, then Figure 6.9 displays very similar results as we have observed for time series 50. At

T=100, the number of best performing tests becomes fifteen (i.e. BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, CHO, WL, WT, and CIPS_star) corresponding to their attained MSC. Moreover, there are two tests (i.e. RMA and OS) which fulfill mediocre tests criteria with 47% and 21% MSC respectively but only PS test with respect to its MSC graded into worst performing test at T=100.

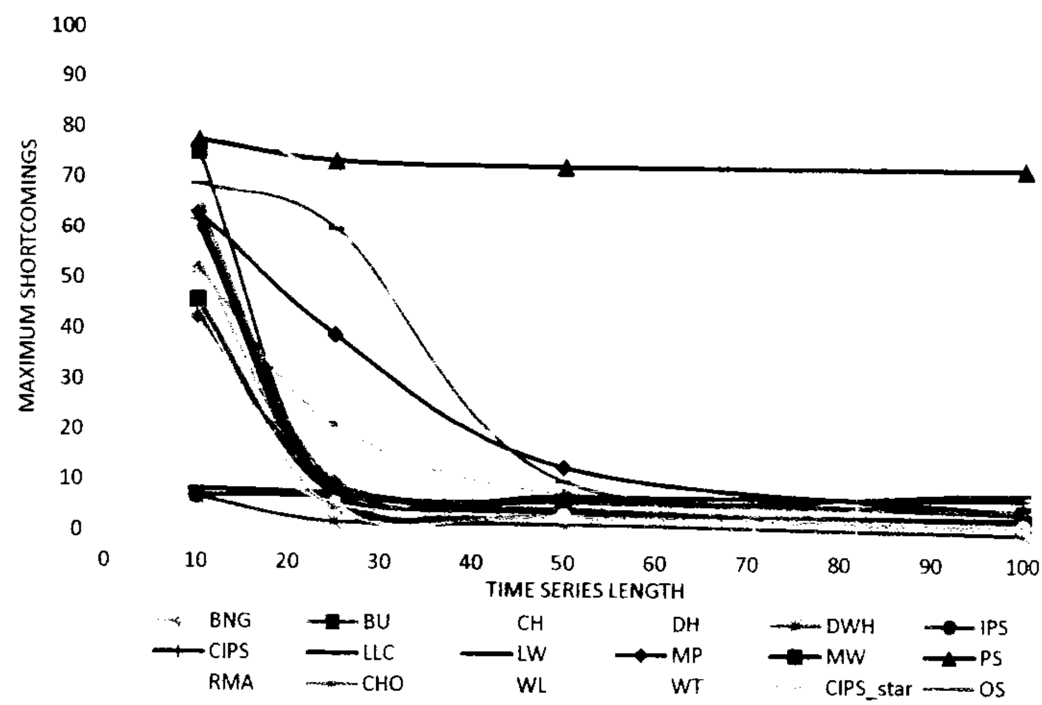
Overall results of Figure 6.9 shows that at small time (T=10) series dimension no test has ordered into best performer tests category, however as the time series level increases from 10 to 25, 50, and 100 then the number of best performing test become nine, twelve, and fifteen, respectively. Also, Figure 6.9 demonstrates that the number of mediocre and worst performing decreases as the sample size increases. We have noticed that PS test is the only worst performing test at all sample sizes. Finally, RMA and OS tests have maintained good pattern and has ranked as mediocre test at time dimension 100, while for time series 10, 25, and 50, RMA test is categorized as worst performing test. Generally, Figure 6.9 demonstrates an improvement of power of all PUR tests for each time series level when number of cross sections are 16 over the number of cross sections 2, 4, and 8 with small to large time series dimensions. At each cross section level and with increasing time series level, the number of best performing tests increases while the number of worst performing tests decreases and at last reaches to one or two tests. This shows a great gain in power of PUR tests as the number of cross section and time series unit increase.

**Table 6.5: MSC for Null Hypothesis of Unit Root in the presence of Intercept and Trend Case, at N=8 and N=16**

| | Panel A: N=8 | | | | Panel B: N=16 | | | |
|---|---|---|---|---|---|---|---|---|
| Test/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| BNG | 63.98 | 24* | 9.6** | 5.01** | 53.76 | 12.82* | 6.7** | 4.97** |
| BU | 80.06 | 12.98* | 11.08* | 6.67** | 77.74 | 9.92** | 8.89** | 6.21** |
| CH | 60.82 | 60.06 | 13.52* | 10.01** | 58 | 56.98 | 9.52** | 7.55** |
| DH | 66.64 | 15.98* | 11.01* | 5.5** | 60.38 | 8.44** | 8.17** | 6.01** |
| DWH | 19.86* | 3.18** | 1.81** | 0.07** | 16.6* | 2.02** | 1.86** | 0.06** |
| IPS | 61.22 | 7.7** | 5.09** | 3.5** | 26.64* | 7.51** | 4.42** | 2.99** |
| CIPS | 85.96 | 72.88 | 43.78* | 18.42* | 67.62 | 39.74* | 16.02* | 9.88** |
| LLC | 70.7 | 8.09** | 5.05** | 2.88** | 64.24 | 7.6** | 4.22** | 1.21** |
| LW | 75.64 | 70.02 | 45.58* | 20.06* | 33.56* | 19.2* | 7.9** | 5.9** |
| MP | 78.2 | 66.46 | 43.18* | 26.86* | 70.74 | 65.46 | 32.94* | 4.4** |
| MW | 79.9 | 14.1* | 6.6** | 3.83** | 65.1 | 9.61** | 6.22** | 2.48** |
| PS | 80.64 | 79.88 | 79.09 | 75.64 | 78.9 | 76.02 | 74.06 | 73.24 |
| RMA | 89.56 | 81.4 | 75.38 | 50.12 | 87.64 | 74.22 | 60.26 | 47.04* |
| CHO | 69.7 | 19.3* | 6.16** | 2.02** | 48.66* | 9.92** | 6.02** | 2.01** |
| WL | 47.96* | 10.02** | 6.6** | 3.5** | 43.96* | 7.9** | 5.7** | 2.23** |
| WT | 50.3 | 8.06** | 5.05** | 2.1** | 47.7* | 6.33** | 4.45** | 1.92** |
| CIPS_star | 83.42 | 72.66 | 43.83* | 20.02* | 70.78 | 13.98* | 10.2* | 4.5** |
| OS | 87.5 | 82.49 | 30.02* | 20.08* | 74.06 | 62.01 | 30.46* | 21.72* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.10: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend, N=32**




At a very last cross section unit 32 Figure 6.10 demonstrates a very similar picture as previously demonstrated for cross section unit 2, 4, 8, and 16. All the PUR tests have convergence behavior as the time series varies from small to large. Table 6.6 exhibits that as sample size varies from small time series level 10 to highest time series level 100 the number of best performer tests increases and decreasing count of worst performer tests have observed. It is observed that when sample size is 10 then only three tests (i.e. DWH, IPS, and LW) are considered to be the best performing tests according to given criteria. In these three best tests, DWH test has the least MSC as compared to MSC of IPS and LW tests. At the same level of time series, BNG, MW, CHO, WL, and WT tests with respect to their MSC are categorized as mediocre tests. However, ten tests (i.e. BU, CH, DH, CIPS, LLC, MP, PS, RMA, CIPS_star, and OS) have been ranked as worst performing tests according to their attained MSC for time series level 10. At the same time series level but with different cross section units, it is observed that number of best and mediocre performer tests are more when N=32.

As the sample size increases from 10 to 25 the number of best performer tests increases, besides the other three best performer tests for sample size 10. BNG, BU, DH, CIPS, LLC, MW, CHO, WL, WT, and CIPS_star tests are also graded into best performing test category at time series level 25. Surprisingly, BU, DH, CIPS, LLC, and CIPS_star have attained higher power and have ordered as best performing tests from worst performing at time series 25. At the same time series level, only two tests (i.e. CH and MP) corresponding to their MSC are assigned as mediocre tests. However, PS, RMA, and OS tests have earned the same status according to their MSC, as have been defined for sample size 10, with little improvement in their powers.

For time series 50 and 100 the number of best performer tests getting larger and larger. CH and OS tests have MSC less than 10% and so the number of best performer tests become fifteen (BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, LW, MW, CHO, WL, WT, CIPS_star, and OS) at T=50 as compared to number of best performer tests for previous time series levels. Figure 6.10 shows that only MP test has sustained its position in mediocre test category while PS and RMA tests remain the worst performing tests for sample size 50.

At sample size 100, MP test has MSC less than 10% and the number of best performer tests become sixteen (BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, MP, MW, CHO, WL, WT, LW, CIPS_star, and OS). Similarly, Figure 6.10 and Table 6.6 also indicate very similar decreasing phenomena for mediocre tests as have been observed for other cross section dimensions. Figure 6.10 also indicates that RMA and PS tests have been classified as mediocre and worst performing tests with 20% and 71% MSC, respectively. It is observed that RMA test has gained a good power at large time series unit as compared to its position for small and medium time dimensions.

Figure 6.10 and Table 6.6 summarize that as the time dimension increases from smallest to largest then the number of best performing tests also increases while the number of mediocre and worst performing tests decreases. First, at T=10 the number of best performing tests are three but as the time series level increases to 25, 50, and 100 then number of best tests become as thirteen, fifteen, and sixteen, respectively. Second, at last sample size (i.e. T=100), fourteen tests have gained an ultimate power of 100% which indicates that high cross section unit and time series dimension effect the power of PUR tests. However, PS test is having worst performance whether time dimension is small, medium or large while RMA has managed its position from worst to mediocre as the time dimension reaches to 100.

Generally, Figure 6.6 to Figure 6.10 indicate a convergence pattern for almost all tests excluding PS test with almost constant pattern of its MSC as the time dimensions moves from smallest to largest level corresponding to each cross section unit when data are generated both with drift and trend terms. The least number of MSC at each higher combination of time series and cross section units indicate a higher power of PUR tests.

**Table 6.6: MSC for Null Hypothesis of Unit Root in the presence of Intercept and Trend case, at N=32**

| Test/TL | 10 | 25 | 50 | 100 |
|---|---|---|---|---|
| BNG | 42.48* | 9.84** | 6.06** | 4.6** |
| BU | 75.26 | 7.02** | 6.6** | 4.1** |
| CH | 52.26 | 21.02* | 7.7** | 4.69** |
| DH | 53.34 | 7.95** | 6.02** | 4.5** |
| DWH | 6.82** | 1.91** | 1.58** | 0.03** |
| IPS | 7.08** | 6.9** | 3.4** | 1.99** |
| CIPS | 61.88 | 8.02** | 7.03** | 6.81** |
| LLC | 60.34 | 7.18** | 3.85** | 1.3** |
| LW | 8.38** | 7.55** | 3.9** | 2.9** |
| MP | 63.24 | 38.92* | 12.86* | 4.16** |
| MW | 45.82* | 7.38** | 4.38** | 2.4** |
| PS | 77.58 | 73.44 | 72.28 | 71.72 |
| RMA | 84.52 | 70.49 | 60.02 | 20.08* |
| CHO | 45.94* | 7.51** | 4.7** | 1.99** |
| WL | 22.74* | 5.55** | 3.31** | 1.7** |
| WT | 44.24* | 4.6** | 3.01** | 1.06** |
| CIPS_star | 64.5 | 9.81** | 6.07** | 4.46** |
| OS | 68.92 | 60.02 | 10.12* | 7.86** |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

### 6.3.3. Effect of Time Series on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Stationary (Intercept Case)

This part of the study discusses the simulation based performance of PUR tests having the null hypothesis of stationarity. The power analysis based on MSC are mentioned from Figure 6.11 to Figure 6.15 for HD, HL, HLM, KK, SS, and DHT and when data are generated with drift term only. A Monte Carlo Simulation size of 10,000 is carried out for all given tests for the fixed number of cross sections 2, 4, 8, 16, and 32 with the varying time series levels of 10, 25, 50, and 100. In each figure x-axis and y-axis show time series size and MSC, respectively. Graphical MSC is also mentioned in Table 6.7 to Table 6.9 for different level of cross section and time series levels when data are generated with only deterministic term of intercept. Figure 6.11 to Figure 6.15 and Table 6.7 to Table 6.9 indicate a poor performance of majority of the PUR tests having the null hypothesis of panel stationary when test equation and DGP have drift term only which justify the bad performance of PUR tests having the null hypothesis in the existing literature.

**Figure 6.11: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept, N=2**

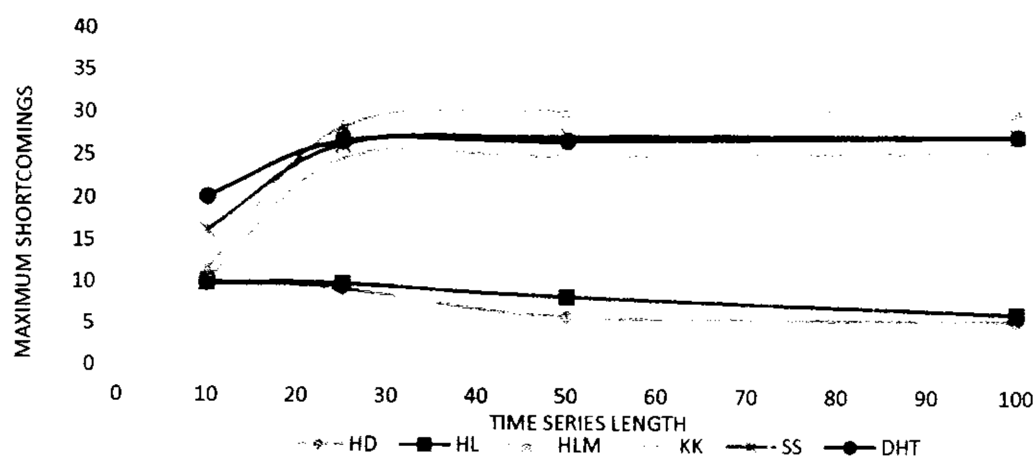


Figure 6.11 and Table 6.7 (Panel A) display results of different time series level (T=10, 25, 50, and 100) for cross section unit 2. Only two tests (HD and HL) have convergence pattern of their MSC when sample size varies from 10 to 100. At T=10, all tests have MSC under 21% but as the time series level increases four of the six tests show a divergence behavior excluding HD and HL tests. The reason of least MSC for four (HLM, KK, SS, and DHT) of the tests at N=2 and T=10 as compared to other time series level is because of their very close power pattern behavior with point optimal test (see Table A.1 (Panel A) in Appendix-A). However, as the time series level increases the power of point optimal test increasing fast as compared to power of stationarity tests resulting a huge a gap between them.

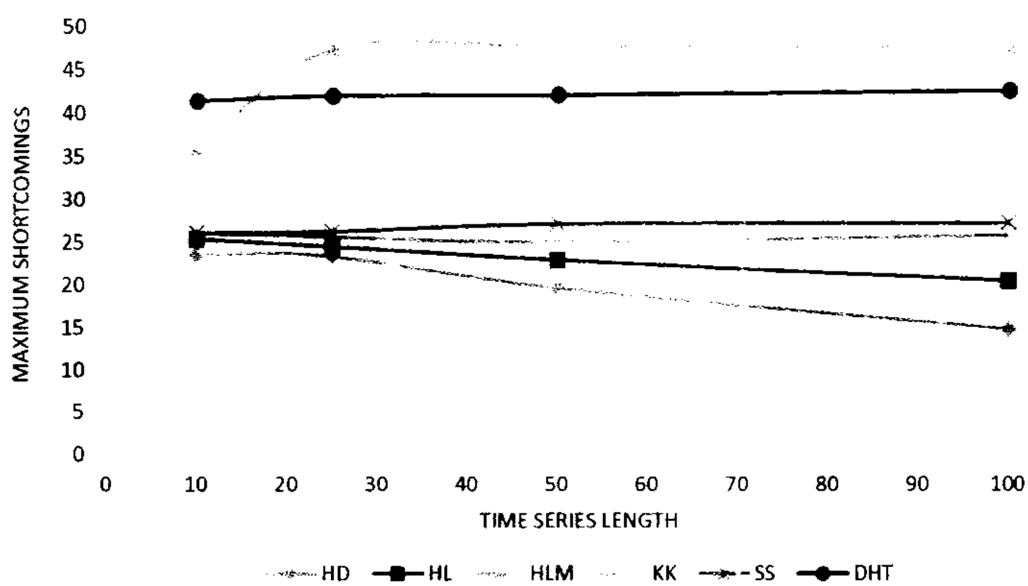
At all sample sizes, only HD and HL tests with highest MSC of 9.89% and 10% at sample size 10, and lowest MSC of 5.62% and 6.46% at sample size 100 have ranked as best performer tests as the time dimension varies from 10 to 100. Also, both of these tests have lowest MSC at each level of time series level as compared to other tests which indicates a very close power pattern of these two tests to the power of point optimal test. Moreover, HD test has lower MSC at each level of time series as compared to its best category counterpart (i.e. HL test) thus indicating its good performance in the best performer category. While, all other four tests according to their MSC are assigned as mediocre tests as the time series level moves from small (10) to large (100) with approximately constant divergence behavior. Figure 6.11 indicates HLM test with highest MSC over other five tests at time series 25, 50, and 100 showing bad performance of this test. Moreover, KK, SS, and DHT tests with very close pattern of their MSC with each other are also not far away from the pattern of HLM test as the sample size varies.

**Table 6.7: MSC for the Null Hypothesis of Stationarity in the presence of Intercept, N=2 and N=4**

| | Panel A: N=2 | | | | Panel B: N=4 | | | |
|---|---|---|---|---|---|---|---|---|
| Tests/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| HD | 9.86** | 9.28** | 6.02** | 5.62** | 23.58* | 23.44* | 19.72* | 14.84* |
| HL | 10** | 9.9** | 8.36** | 6.46** | 25.44* | 24.56* | 23.02* | 20.58* |
| HLM | 11.72* | 28.42* | 30.51* | 30.52* | 36* | 47.54* | 47.96* | 47.99* |
| KK | 10.3* | 24.72* | 25.06* | 25.66* | 26.12* | 25.68* | 25.16* | 25.94* |
| SS | 16.32* | 26.4* | 27.38* | 27.56* | 26.17* | 26.32* | 27.26* | 27.46* |
| DHT | 20.24* | 26.75* | 26.82* | 27.5* | 41.46* | 42.08* | 42.26* | 42.88* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

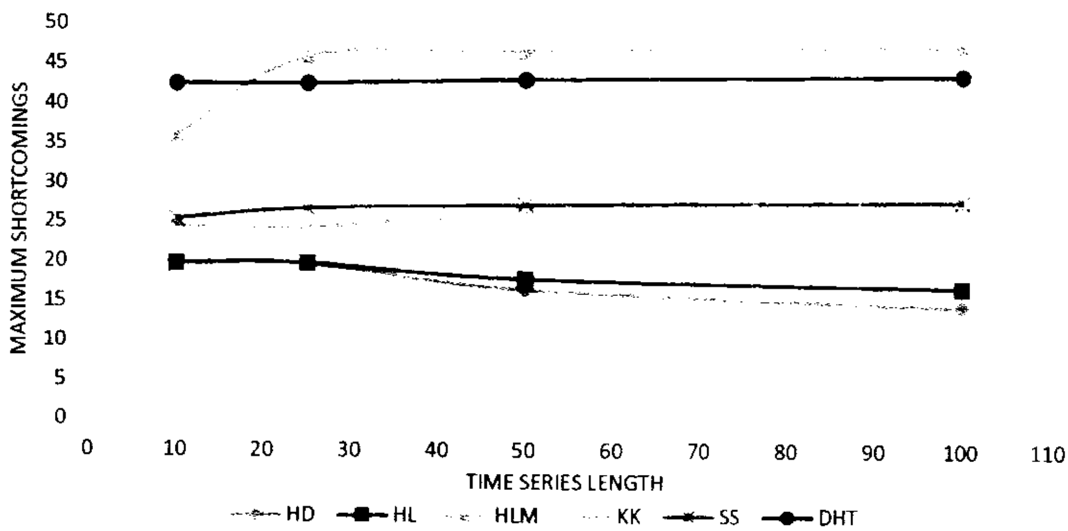
**Figure 6.12: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept, N=4**



At N=4 with different time series level Figure 6.12 and Table 6.7 (Panel B) demonstrate a very different behavior as compared to N=2 showing a remarkable power gain of point optimal test as compared to power of stationarity tests. This indicates a high MSC at each level of time series at N=4 as compared to the results of N=2. All tests have very similar pattern of their MSC as have been observed at N=2 as the time series level goes from 10 to 100. Again, HD and HL tests with respect to their convergence pattern of MSC have stood as better performer tests among all mediocre performer tests. Moreover, HD test with lowest MSC of 23.6%, 23.4%, 19.7%, and 14.8% as compared to MSC of 25.4%, 24.6%, 23%, and 20.6% of HL test over time dimension 10, 25, 50, and 100 respectively is considered as better in the mediocre category.

However, KK and SS tests with constant pattern over time dimension with MSC fluctuation in between 25% and 28% are categorized as mediocre tests beside with HLM and DHT tests with their power curve having large distance from power envelope of point optimal test. Also, HLM, SS, and DHT tests have divergence pattern as the time dimension increases from 10 to 100. Overall, no test have ordered as best or worst performer test according to assigned MSC at each level of time series level. Also, at very large time dimension, power of HD and HL tests will eventually become equal to power of point optimal test thus will have MSC equal to zero.

**Figure 6.13: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept, N=8**




Further, as the number of cross section increases from 4 to 8 a similar picture has been shown in Figure 6.13 and Table 6.8 (Panel A) as has been inspected at N=4 over time series 10, 25, 50, and 100 when data are generated with drift term only. Here, HD test with MSC 19.84%, 19.78%, 16.46%, and 14.38% for time series level 10, 25, 50, and 100 respectively are categorized as better performing tests in the mediocre tests category as compared to all other mediocre tests. This is observed from Figure 6.13 as HD test has lowest graph representation of MSC at each time series level with convergence pattern.

Similarly, HL test has the same phenomenon with MSC 19.84%, 19.82%, 17.82%, and 16.74% for time series dimension 10, 25, 50, and 100 respectively but it has more MSC at each level of time series as compared to HD test and ranked as the second better mediocre performing test among the five mediocre tests with convergence behavior. However, KK and SS tests with approximately same constant divergence pattern are graded as third and fourth number of mediocre tests having MSC of 24.42%, 24.38%, 25.94%, and 26.02%; and 25.44%, 26.75%, 27.24%, and 27.68% respectively over time series length 10, 25, 50, and 100. While, DHT test with MSC of 42.54%, 42.56%, 42.94%, and 43.34% at time dimension 10, 25, 50, and 100 are ranked as fifth mediocre test among all other mediocre tests. Lastly, HLM test with divergence pattern and highest MSC at each time series level is ordered as the last mediocre test. Also, the MSC of HD and HL tests at N=8 are less at each time series level as compared to N=2 while all other stationarity tests have the same performance for both type of cross section levels. Generally, according to Figure 6.13, HD and HL tests are ranked as better performing mediocre tests among all mediocre tests with the convergence graph as the sample size varies from small to large. Also, both of these tests will eventually get zero MSC at very high cross section and time series level. Moreover, none of the tests have fulfilled best performer test criteria.

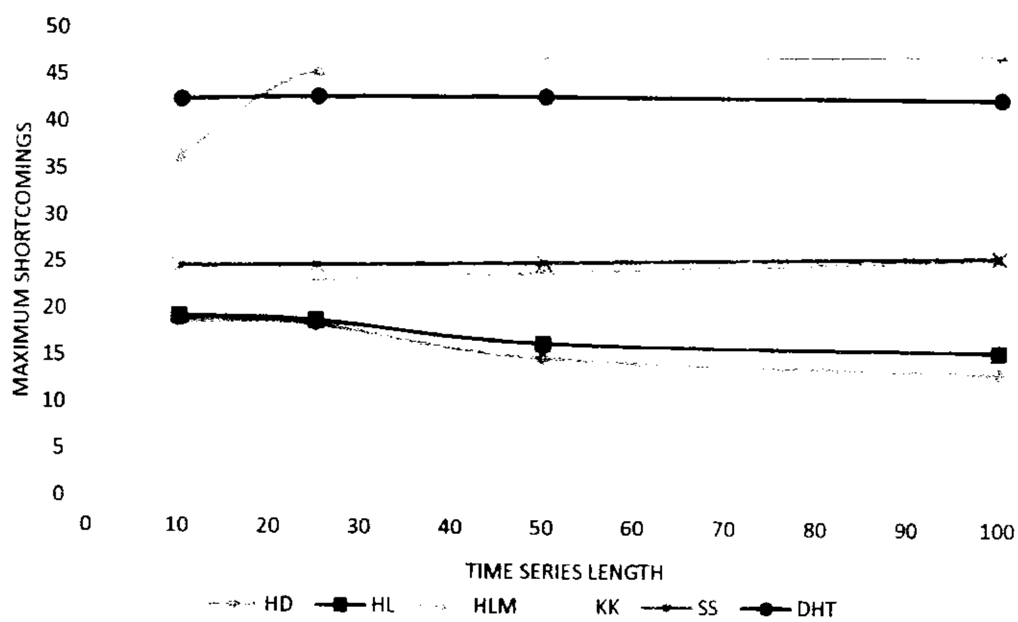
**Table 6.8: MSC for the Null Hypothesis of Stationarity in the presence of Intercept, N=8 and N=16**

| | Panel A: N=8 | | | | Panel B: N=16 | | | |
|---|---|---|---|---|---|---|---|---|
| Tests/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| HD | 19.84* | 19.78* | 16.46* | 14.38* | 18.74* | 18.5* | 14.94* | 13.54* |
| HL | 19.84* | 19.82* | 17.82* | 16.74* | 19.28* | 18.93* | 16.56* | 15.86* |
| HLM | 36.04* | 45.88* | 46.57* | 47.03* | 36.6* | 45.74* | 46.76* | 47.8* |
| KK | 24.42* | 24.38* | 25.94* | 26.02* | 22.54* | 23.18* | 24.16* | 25.78* |
| SS | 25.44* | 26.75* | 27.24* | 27.68* | 24.63* | 24.86* | 25.22* | 26.08* |
| DHT | 42.54* | 42.56* | 42.94* | 43.34* | 42.52* | 42.9* | 42.98* | 43* |

Note: "*" indicates Mediocre Test.

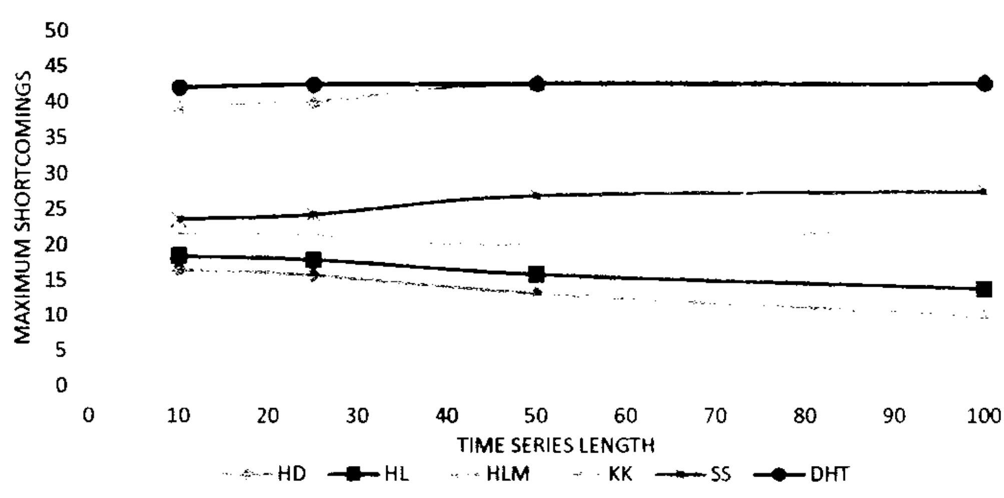
Figure 6.14 demonstrates the behavior of all six stationarity tests when the numbers of cross sections are 16 with varying time series level 10, 25, 50, and 100. Almost all tests have the same type of behavior but with little improvement in their power as the number of cross section units move from 8 to 16. At time series level 10, 25, 50, and 100, HD test with MSC 18.7%, 18.5%, 14.9%, and 13.5% respectively and convergence pattern is labeled as better performer mediocre test in the class of mediocre tests. Also, HL test with MSC 19.3%, 18.9%, 16.6%, and 15.9% at time series level 10. 25, 50, and 100 respectively and convergence behavior is ranked as second better performing mediocre test among mediocre tests. Moreover, the MSC of both of these tests are less at N=16 as compared to their MSC at N=8 with convergence picture, these results can also be observed from Table 6.8 (Panel A) and (Panel B).

**Figure 6.14: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept, N=16**



However, HLM test with divergence behavior of MSC of 36.6%, 45.7%, 46.8%, and 47.8% at time series level 10, 25, 50, and 100 respectively is ranked as bad performing 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 MSC 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 the second bad performing test beside HLM test in the category of mediocre tests. The MSC of both of these tests fluctuates in between 40% to 48% as the time series level increases from 10 to 100, except MSC of 36% of HLM test at T=10. These results for N=16 indicates HD and HL tests as the better performing tests in the category of mediocre tests according to their assigned convergent MSC and are preferred to apply at each time series level as compared to other mediocre tests. Also, at very large time series and cross section level both of these tests will have their power curve equal to power envelope.

**Figure 6.15: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept, N=32**



Increasing number of cross section to last number of cross section 32, Figure 6.15 and Table 6.9 analyze the performance of all six tests with varying time series level when data generating process and test equation both have drift term only. Again, all of the stationarity tests have MSC under 50% but over 10% for almost all tests indicating that these tests are categorized as mediocre tests. In other words, no test is assigned as worst performing test and also all the tests have not taken the position of best performing tests excluding only HD test with MSC of 9% at T=10.

At N=32, HD test over varying time series level of 10, 25, 50, and 100 corresponding to MSC 16.52%, 15.72%, 13.2%, and 9.82% is classified as better performing test among all mediocre class tests. However, at time series 100, HD test with its convergence pattern is ranked as best performing test having MSC 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. Further, HL test has with convergence pattern of its MSC of 18.48%, 17.94%, 15.86%, and 13.74% over time series level of 10, 25, 50, and 100 respectively is categorized as second better performing test among mediocre tests category apart HD 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 have higher power thus will be included in best performing category.

Moreover, HLM test with MSC 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 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 MSC

42.06%, 42.52%, 42.62%, and 42.73% over time series level of 10, 25, 50, and 100 respectively. Moreover, DHT test shows a very similar constant behavioral picture that has been analyzed for HLM test.

Finally, KK test with little improvement in their MSC, which fluctuates in between 20% to 22%, is categorized as third mediocre test. Similarly, SS test with divergence pattern of its MSC is also remained in the mediocre tests class like its performance for the previous cross section units.

**Table 6.9: MSC for the Null Hypothesis of Stationarity in the presence of Intercept, N=32**

| Tests/TL | 10 | 25 | 50 | 100 |
|---|---|---|---|---|
| HD | 16.52* | 15.72* | 13.2* | 9.82** |
| HL | 18.48* | 17.94* | 15.86* | 13.74* |
| HLM | 39.44* | 40.14* | 42.76* | 43.56* |
| KK | 21.66* | 21.54* | 20.08* | 22.68* |
| SS | 23.64* | 24.3* | 26.95* | 27.58* |
| DHT | 42.06* | 42.52* | 42.62* | 42.73* |

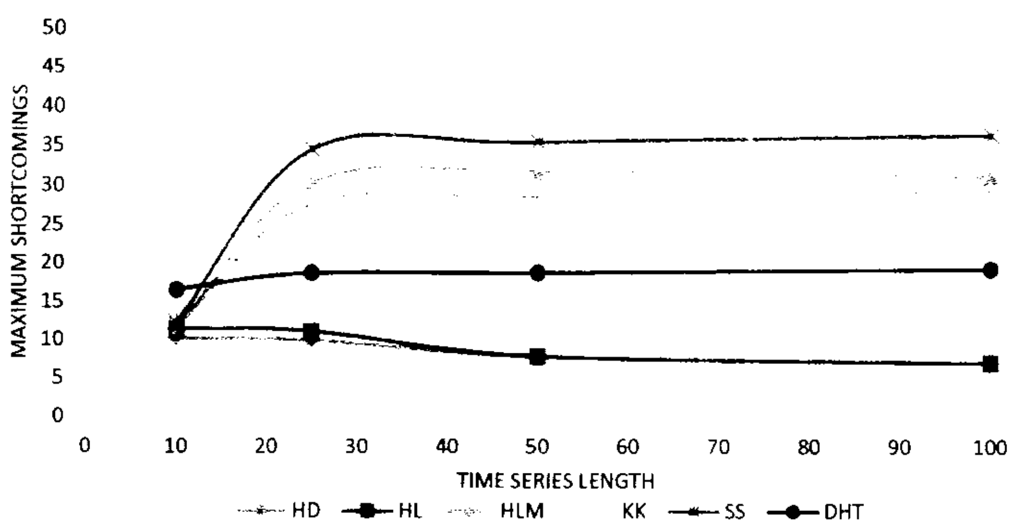
Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

Overall, this analysis for N=32 indicates a very similar picture that has been analyzed for previous cross section units. It is observed from Figure 6.11 to Figure 6.15 that all stationarity tests lie in the class of mediocre tests majority of the time when data generating process and test equation have drift term only, excluding the performance of HD and HL tests at N=2. Also, HD and HL tests have maintained their convergence behavior and have ranked as better performing tests in the category of all mediocre tests. While, other four (i.e. HL, KK, SS, and DHT) tests with divergence behavior remain in the same category, that is, in mediocre tests.

### 6.3.4. Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Stationary (Intercept and Trend Case)

This section investigates the behavior of PUR tests having the null hypothesis of stationarity when data are generated with intercept and trend using Monte Carlo simulation size 10,000 with time series varying 10, 25, 50 and 100 for fixed cross section units of 2, 4, 8, 16 and 32. The performance of all six tests and for all combination of time series and cross section unit are patriated from Figure 6.16 to Figure 6.20. In each figure, x-axis displays sample sizes and y-axis shows MSC. Table 6.10 to Table 6.12 also points out MSC performances of stationary tests for the varying time series level and fixed cross section unit when test equations and DGP have both drift and trend terms.

**Figure 6.16: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend, N=2**



In the presence of both deterministic terms Figure 6.16 and Table 6.10 (Panel A) show the performance of all six tests as the sample size varies from 10 to 100 for a single cross section unit 2 when data are generated with both of deterministic terms. At first sample size T=10, all six (HD, HL, HLM, SS, KK, and DHT) tests are identified as mediocre with respect to their MSC, however, as the time dimension increases only HD and HL tests attain their positions in best performing tests while other tests remain in the mediocre tests category corresponding to their MSC.

Figure 6.16 indicates HD test as the most stringent test as it has minimum value of the MSC, that is, 10.24%, 9.96%, 7.84%, and 6.96% corresponding to sample size 10, 25, 50, and 100 respectively as compared to MSC of all other stationarity tests at the same time series level. It can be noted that at small time series level (T=10), HD test is classified as mediocre test, however, at T=25, 50, and 100 it is ranked as best performer test with its closer power curve to power envelope (see Table A.11 (Panel A and Panel B) in Appendix-A). While, HL test with its MSC 11.44%, 11.2%, 7.86%, and 7% at T=10, 25, 50, and 100 respectively is more near or approximately with same pattern to HD test is classified as second better performing test in the mediocre tests category at time series level of 10 and 25. At T=50 and 100 this test is graded into best performing tests class where its MSC is approximately equal to MSC of HD test.

However, HLM test with its MSC of 11.7% at small time series level is supposed to be very close to best performing tests criteria, but as the time dimension increases to 25, 50, and 100 it achieves corresponding MSC of 30.46%, 31.82%, and 31.24% respectively showing its divergence behavior in the mediocre tests class. Moreover, a divergence pattern has also been analyzed for KK, SS, and DHT tests as the time dimension increases from small to large. At T=10, all these three tests (KK, SS, and DHT) show a MSC of 11.78%, 12.29%, and 16.56% respectively and have the least values of MSC as compared to their values for other time series levels.

According to MSC performance, DHT test is ranked as the third better performing test in the mediocre tests class with MSC 16.56%, 18.725, 18.74%, and 19.18% for time series level of 10, 25, 50, and 100 respectively. Further, KK test is ranked as the fourth better performing test in the mediocre tests class. While, SS test is classified as the bad performing test among mediocre tests category with increasing pattern of MSC of 12.29%, 34.6%, 35.54%, and 36.58%.

**Table 6.10: MSC for the Null Hypothesis of Stationarity in the presence of Intercept and Trend, N=2 and N=4**

| | Panel A: N=2 | | | | Panel B: N=4 | | | |
|---|---|---|---|---|---|---|---|---|
| Tests/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| HD | 10.24* | 9.96** | 7.84** | 6.96** | 22.4* | 21.56* | 21.26* | 20.56* |
| HL | 11.44* | 11.2* | 7.86** | 7** | 22.84* | 22.78* | 21.34* | 21.01* |
| HLM | 9.7** | 30.46* | 31.82* | 31.24* | 30.5* | 31.42* | 33.66* | 34.22* |
| KK | 11.78* | 27.68* | 28.46* | 29.22* | 27.76* | 27.6* | 26.5* | 26.58* |
| SS | 12.29* | 34.6* | 35.54* | 36.58* | 25.24* | 25.34* | 26.34* | 27.06* |
| DHT | 16.56* | 18.72* | 18.74* | 19.18* | 27.18* | 27.82* | 24.76* | 23.2* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

Generally, these results for N=2 suggest HD and HL tests as the better performing tests among stationarity tests. At very large time series dimension the MSC of both these tests will become closer to zero MSC indicating their powers equal to the power of point optimal test. However, due to minimum value of MSC of HD test at each level of time series length it is considered to be most stringent test as compared to all other stationarity tests.

**Figure 6.17: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend, N=4**

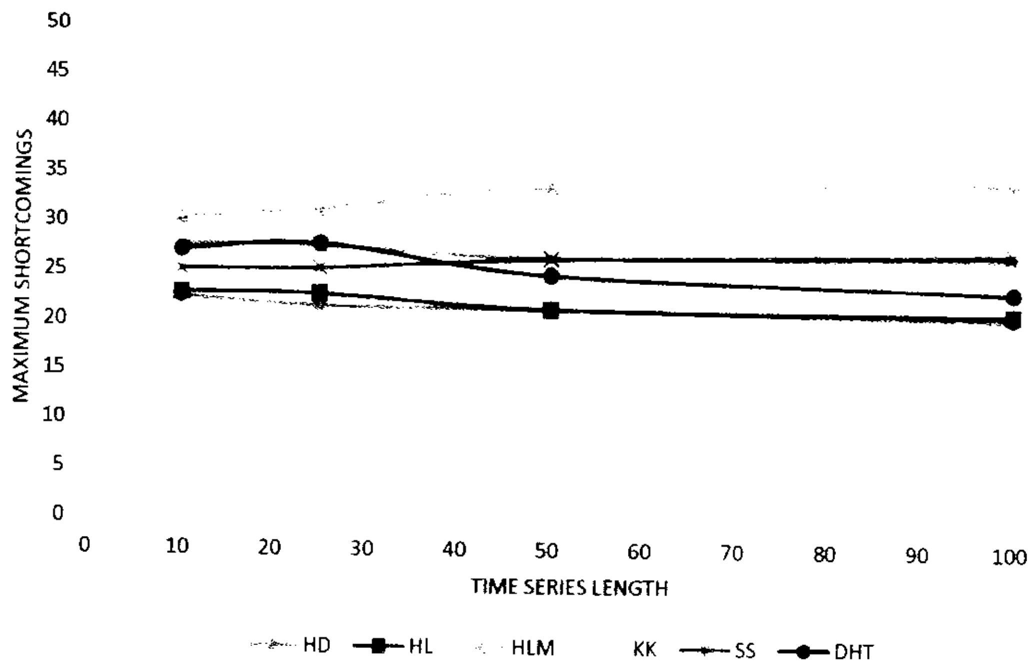



Figure 6.17 displays the MSC picture of stationarity tests at N=4 for T=10, 25, 50, and 100 when data are generated with drift and trend terms. At T=10, all stationarity tests MSC fluctuates in between 22% to 28% indicating the approximate same power behavior of these tests with margin difference and ranked as mediocre tests. Similarly, a very same picture can be seen for time series 25, 50, and 100 for almost all tests with mediocre positions.

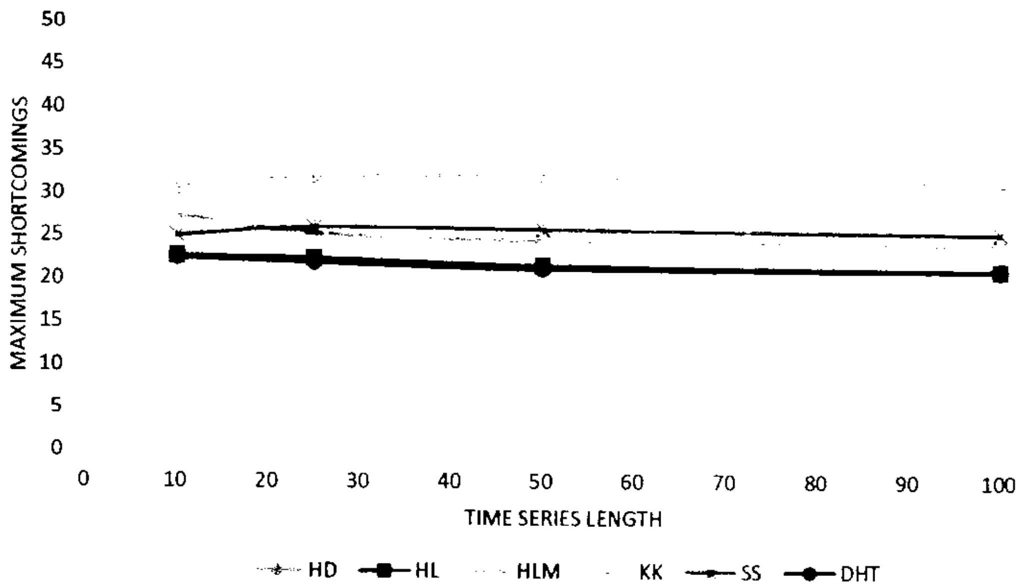
HD test and HL tests have lowest MSC among other mediocre performer tests at each stage of time series length. However, as HD test has minimum values (i.e. 22.4%, 21.56%, 21.26%, and 20.56%) of shortcomings than HL test's shortcomings (i.e. 22.8%, 22.8%, 21.3%, and 21%) at time dimension length of 10, 25, 50, and 100 respectively thus HD test is considered as most stringent test. Also, these results for HD and HL show a very slow convergence pattern of MSC over time dimension.

Moreover, KK and DHT tests with a slow decreasing pattern of MSC of 27.76%, 27.6%, 26.5%, 26.58%, and, 27.2%, 27.85, 24.8%, 23.2% respectively have ranked as mediocre tests at each level of time dimension. Both of these tests show a very similar pattern but at large time dimension DHT test has less MSC than KK test. SS test shows an increasing pattern and has lowest MSC value of 25.24% at sample size 10 while highest value of MSC 27.06% corresponding to time dimension 100.

However, with MSC of 30.5%, 31.42%, 33.66%, and 34.22% for time series level of 10, 25, 50, and 100 respectively HLM test with its slow divergence pattern is ranked as the bad performing test among mediocre tests at each time series level.

Table 6.10 (Panel B) and Figure 6.17 indicate HD and HL tests as better performing tests being in the category of mediocre performing tests with convergence pattern. However, HD test with lowest MSC value among all tests at each level of time series level is labeled as most stringent test.

**Figure 6.18: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend, N=8**




A more detail analysis of stationarity tests have been evaluated for N=8 in Figure 6.18 and Table 6.11 (Panel A) indicating a similar situations as have been observed in N=4. At T=10, all tests have classified into mediocre tests category with MSC of 22.38%, 22.7%, 30.8%, 27.44%, 25.03%, and 22.64% corresponding to HD, HL, HLM, KK, SS, and DHT tests respectively. However, as the time series level reaches to 25 then HD, HL, KK, and DHT tests with decreasing MSC of 21.92%, 22.44%, 25.32%, and 21.56% respectively remain in the mediocre tests class. While, HLM and SS tests with increasing MSC of 31.9% and 25.96% also persist to lie in mediocre class. At T=50, again all these tests continue to remain in the mediocre tests category with further decreasing pattern of HD, HL, KK, SS, and DHT tests having MSC of 21.14%, 21.58%, 31.98%, 24.42%, 25.7%, and 21.22% respectively while with increasing pattern of HLM test with MSC of 31.98%. Moreover, at T=100, all these tests continue to be the part of mediocre tests class having the lowest of the MSC value 20.86% against HD test while highest of the MSC value of 31.18% against HLM test.

**Table 6.11: MSC for the Null Hypothesis of Stationarity in the presence of Intercept and Trend, N=8 and N=16**

| | Panel A: N=8 | | | | Panel B: N=16 | | | |
|---|---|---|---|---|---|---|---|---|
| Tests/TL | 10 | 25 | 50 | 100 | 10 | 25 | 50 | 100 |
| HD | 22.38* | 21.92* | 21.14* | 20.86* | 20.3* | 18.68* | 17.02* | 16.06* |
| HL | 22.7* | 22.44* | 21.58* | 20.94* | 20.52* | 19.82* | 18.18* | 17.74* |
| HLM | 30.78* | 31.9* | 31.98* | 31.18* | 31.3* | 30.78* | 31.96* | 32.96* |
| KK | 27.44* | 25.32* | 24.42* | 23.88* | 23.56* | 21.6* | 21.3* | 19.88* |
| SS | 25.03* | 25.96* | 25.7* | 25.16* | 25.04* | 23.56* | 22.8* | 21.24* |
| DHT | 22.64* | 21.96* | 21.22* | 21* | 21.46* | 18.86* | 17.56* | 16.98* |

Note: "*" indicates Mediocre Test.

Figure 6.18 demonstrates HD test with convergent MSC of 22.38%, 21.92%, 21.14%, and 20.86% for time series level of 10, 25, 50, and 100 respectively as the better mediocre performing test among all mediocre tests category. Also, HD test is considered to be most stringent test as it has minimum value of MSC at each time series level than other tests. Apart from HD test, HL test also has a convergence behavior with MSC of 22.7%, 22.44%, 21.58%, and 20.94% corresponding to time series length of 10, 25, 50, and 100 respectively and has ranked as mediocre test for each time series level. Similarly, KK and DHT tests have also maintained a decreasing pattern of their MSC of 27.44%, 25.32%, 24.42%, 23.18% and 22.64%, 21.96%, 21.22%, 21% respectively and have ranked as mediocre tests at all sample sizes.

However, HLM test with increasing pattern and SS test with approximately constant pattern of their MSC have also made place in mediocre tests whether sample size is small, medium or large but only HLM test has named as bad mediocre performing test in the mediocre tests category having maximum value of MSC at each time series level. While SS test is considered to be the second bad performing mediocre test beside HLM test as the sample size varies from 10 to 100.

Figure 6.18 concludes that HD and HL tests are the better performing test among mediocre tests category, however, HD test is detected as most stringent test having minimum value of shortcomings among all other stationarity tests at time series dimension 10, 25, 50, and 100. While, HLM test with its divergent behavior over time dimension is identified as bad performing test among mediocre tests category.

Further, as the number of cross section increases to 16 then a convergence pattern for majority of the tests have been identified in Figure 6.19 and Table 6.11 (Panel B) as the time series progress from small to large. Figure 6.19 indicates a better performance of stationarity tests at N=16 as compared to their performances for N=2, 4, and 8. The graphical representation of HD, HL, KK, SS, and DHT tests demonstrate a convergence behavior while HLM test shows a divergence pattern over time according to their attained MSC.

**Figure 6.19: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend, N=16**

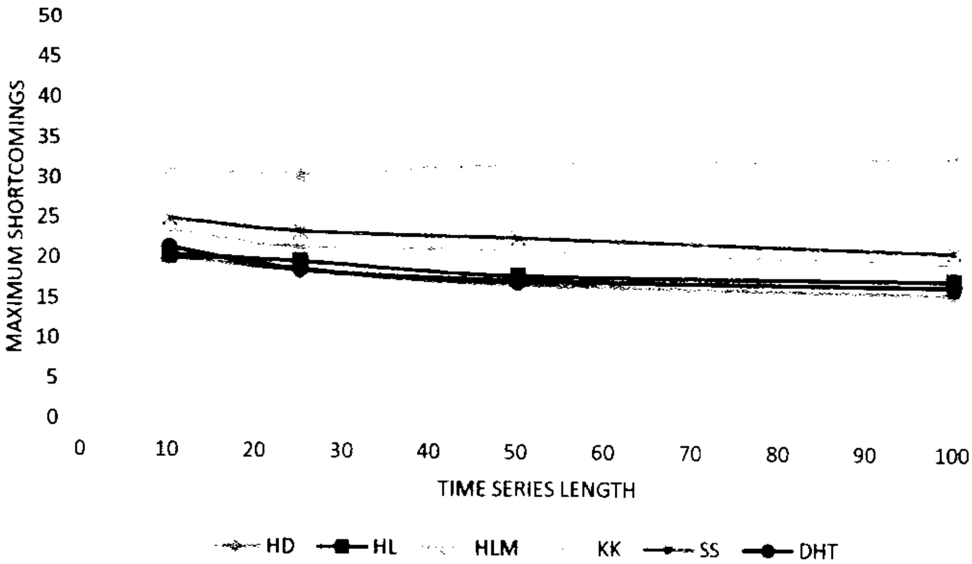


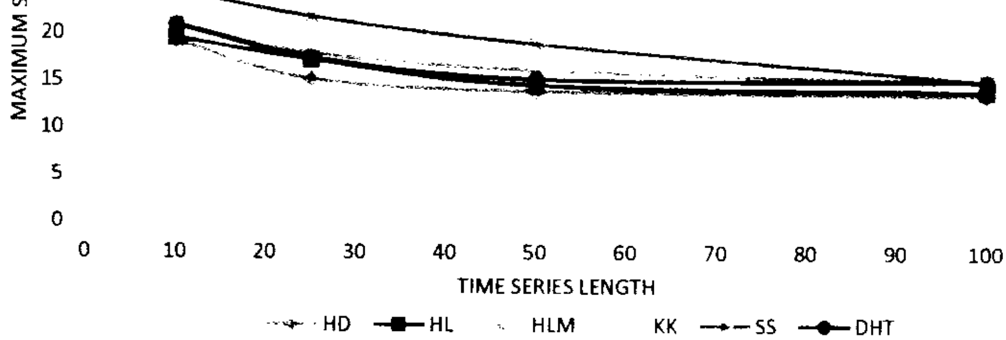

Figure 6.19 identifies HD test as the first better performer test among mediocre tests category having MSC of 20.3%, 18.68%, 17.02%, and 16.06% as the time series increases from 10 to 100. Similarly, HL and DHT tests corresponding to MSC of 20.52%, 19.82%, 18.18%, 17.74% and 21.46%, 18.86%, 17.56%, 16.98% respectively have also detected as better performing tests among mediocre tests over time series level of 10, 25, 50, and 100. However, due to highest values of these tests over the values of HD test at each sample size, HD test is ranked as most stringent test.

Similarly, KK test with attained MSC of 23.56%, 21.6%, 21.3%, and 19.88% at time series level 10, 25, 50, and 100 respectively is identified as fourth mediocre test in the mediocre tests class. Further, this test has improved its power behavior as compared to its power for the last cross section units.

Also, SS test has a decreasing pattern of MSC of 25.04% to 21.24% as the time length increases from small (10) to large (100) respectively indicating its second last position from the graphical representation in the mediocre tests category. Moreover, this test has switched its pattern from divergence to convergence pattern as compared to its performance for the last cross section units. However, HLM test with as usual divergence behavior of its MSC remains in the bed performing position in the mediocre tests class.

Figure 6.19 concludes that HD test is the most stringent test as it has minimum value of its MSC at each level of time series while HLM test as the bad performing test with divergence pattern over time series level 10 to 100 when the number of cross section units are 16. Also, all of the tests have improved their MSC behavior as compared to their MSC for previous cross section units.

**Figure 6.20: MSC Assessments of Tests having Null of Stationarity with Intercept and Trend, N=32**



At very last cross section unit of 32, Figure 6.20 and Table 6.12 display results of all six stationarity tests in which almost all tests have further attained a convergence pattern according to their MSC as the time series varies from small to large when specification of test equation model and DGP is same (i.e. case of intercept and trend terms). Here, MSC values fluctuate in between 13% and 32% for all sample sizes.

HD test with MSC of 19.26%, 15.22%, 13.96%, and 13.62% with respect to time dimension 10, 25, 50, and 100 respectively is ranked as the most stringent tests among all mediocre tests. Moreover, HD test has better performance corresponding to MSC at N=32 as compared to its MSC for previous sample sizes.

Similarly. HL test is detected as the second better performer test having MSC of 19.56%, 17.35%, 14.58%, and 14.04% at time series level 10, 25, 50, and 100 respectively. Also, HL test has attained a good convergence behavior at N=32 as compared to its pattern for MSC less than N=32 but remains in the mediocre test category. While, two other mediocre tests (i.e. KK and DHT) have also very close pattern to HD and HL tests at each level of time series dimension. Both of these tests show an overlap pattern corresponding to each level of time at N=32, however, at N=16 and less both of these tests have very different behavior with respect to their MSC.

Further, SS test also behaves very similarly like KK and DHT tests in the mediocre tests class but with higher MSC at each time dimension. SS test has MSC of 24.56%, 21.9%, 19%, and 15.08% over sample size 10, 25, 50, and 100 respectively and has stood as the fifth better performing test and second bad performing test among mediocre tests class. However, the decreasing MSC performance of this test at N=32 is better than the previous cross section units.

Lastly, HLM test has also shown a decreasing behavior with respect to time dimension indicating its switching position from divergent class to convergence class category. This test has MSC of 31.98%, 31.1%, 30.76%, and 30.2% at sample size 10, 25, 50, and 100 respectively and is ranked as the last mediocre performing test. This result shows that HLM test will eventually get more decreasing pattern and at last will achieve zero MSC at very large time series and cross section level. Figure 6.20 concludes a very similar results from stringent test (HD test) point of view as has been observed for the previous number of cross section units when DGP and test equation have same specification.

**Table 6.12: MSC for the Null Hypothesis of Stationarity in the presence of Intercept and Trend, N=32**

| Tests/TL | 10 | 25 | 50 | 100 |
|---|---|---|---|---|
| HD | 19.26* | 15.22* | 13.96* | 13.62* |
| HL | 19.56* | 17.3* | 14.58* | 14.04* |
| HLM | 31.98* | 31.1* | 30.76* | 30.2* |
| KK | 21* | 18* | 16.2* | 15.06* |
| SS | 24.56* | 21.9* | 19* | 15.08* |
| DHT | 20.94* | 17.5* | 15.22* | 15.06* |

Note: "*" indicates Mediocre Test.

## 6.4. Effect of Cross Section Length on Maximum Shortcomings of Panel Unit Root Tests

This section analyzes best, mediocre and worst performing PUR and stationary tests by keeping time series dimension fixed while cross section unit varies for the test models having drift only and, with both drift and trend cases. In the first part of this section, the power behavior of PUR tests for both cases (i.e. with drift only and, with both drift and trend) have been analyzed and the second and last part of this section analyzes power of stationary tests for the same specification cases (i.e. with drift only and, with both drift and trend). Here, number of cross section units are divided into three category, these are small (N=2, 4), medium (N=8, 16), and large (N=32). While, we have already categorized time dimension as, small (T=10), medium (T=25 and T=50), and large (T=100).

### 6.4.1. Effect of Cross Section Length on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Unit Root (Intercept Case only)

This section assesses the performance of power property of PUR tests with null hypothesis of unit root for the specification of drift term only in the DGP and model of tests used for fixed time series level (i.e. T=10, 25, 50, 100) and with varying cross section unit (i.e. N=2, 4, 8, 16, 32).

Figure 6.21 to Figure 6.24 and Table 6.13 to Table 6.15, investigate the effect of cross section units on the performance of PUR tests having the null hypothesis of unit root when data are generated with drift deterministic term only along with same specification for the model of the tests. In all of these figures, x-axis shows cross section dimensions and y-axis represents MSC for each time series level.

**Figure 6.21: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept Term only, T=10**

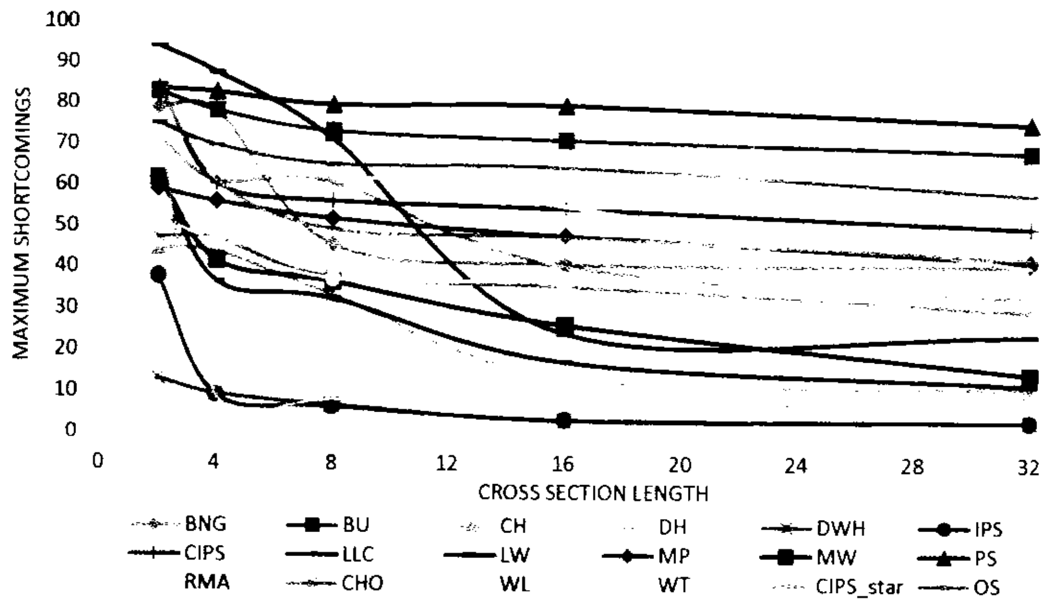



Figure 6.21 and Table 6.13 (Panel A) show the behavior of PUR tests having the null hypothesis of unit root when data are generated with drift term only for fixed T=10 with varying number of cross sections to analyze the effect of cross sections. Figure 6.21 shows a convergence pattern for almost all PUR tests. When number of cross sections are two, then all of the tests either take maximum shortcoming against mediocre class or worst performer class showing not a very significant performance of power of all tests at small cross section. Four (CH, DWH, IPS, CHO) of the tests have ranked into mediocre performing tests while fourteen (BNG, BU, DH, CIPS, LLC, LW, MP, MW, PS, RMA, WL, WT, CIPS_star, OS) tests have been ordered into worst performing tests category with respect to their attained MSC at cross section unit 2 for T=10.

As the number of cross section increases from 2 to 4 then DWH, IPS, and WT tests fulfill the best performing tests criteria with 9.48%, 9.32%, and 9.42% MSC respectively and have ranked as best performing tests. While, BNG, BU, DH, CIPS, LW, MP, PS, RMA, CIPS_star, and OS tests have again categorized into worst performing tests as all these tests have been in the same status when N=2. But, a few test (i.e. CH, LLC, MW, CHO, and WL) have managed to rank into mediocre tests form worst performing test category. These results indicate an increasing power of PUR tests having less MSC at N=4 as compared to N=2. However, as the number of cross section units reaches to 8 then almost all the test have the same status that have been observed for N=4, except BNG and CIPS_star tests which have gained power and ranked into mediocre test class with MSC 45% and 49% respectively.

At small time series (T=10) when the number of cross section are medium (N=16) then few tests have managed to place their previous positions and have either ordered into best or mediocre performing classes. DH, LW, and MP tests is ranked into mediocre tests category with MSC 40%, 24%, and 47% respectively. However, the number of best performing tests (DWH, IPS, and WT) remain the same but their MSC have been decreased.

When the number of cross section becomes large (N=32) for small time series dimension (T=10) then CIPS and RMA tests have gained a reasonable power and have ranked into mediocre test class, however, all other tests have remained in the same status that have been observed when N=16. Generally, all of the tests have maintained a convergence behavior as the number of cross varies from small to large when the time series dimension is small (T=10). As the number of cross section unit changes from small to large then decreasing pattern of MSC affect the behavior of PUR tests from ranking point of view. However, BU, PS, and OS tests with MSC over 50% for each level of cross section unit have remained the worst performing tests. Overall, DWH, IPS, and WT tests have graded as best performing test for N=4, 8, 16, 32 with respect to their MSC.

**Table 6.13: MSC for Null Hypothesis of Unit Root in the Presence of Intercept Case, at T=10 and T=25**

| | Panel A: T=10 | | | | | Panel B: T=25 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Test/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| BNG | 79 | 78.14 | 45.64* | 40.93* | 40.72* | 59.42 | 54.34 | 26* | 24.26* | 9.86** |
| BU | 82.9 | 78.22 | 73.3 | 71.07 | 68.18 | 69.3 | 58.84 | 50.1 | 45.79* | 16.64* |
| CH | 44.58* | 43.74* | 33.46* | 13.44* | 11.18* | 30.5* | 25.08* | 11.08* | 7.9** | 6.8** |
| DH | 71.42 | 61.08 | 60.76 | 40.34* | 32.88* | 57.22 | 55.02 | 45.5* | 15.18* | 7.1** |
| DWH | 13.28* | 9.48** | 6.96** | 3.22** | 3.2** | 4.52** | 4.18** | 1.11** | 1.16** | 1.1** |
| IPS | 37.92* | 9.32** | 6.54** | 3.4** | 3.16** | 3.24** | 3.02** | 3.19** | 2.97** | 2.01** |
| CIPS | 83.78 | 60.46 | 56.34 | 54.42 | 49.96* | 76.8 | 17.14* | 15.9* | 14.96* | 13.02* |
| LLC | 63.03 | 36.62* | 32.62* | 17.3* | 12* | 8.38** | 8.2** | 6.03** | 5.97** | 5.8** |
| LW | 94 | 87.5 | 71.32 | 24.18* | 23.85* | 81.64 | 26.08* | 25.74* | 20.96* | 19.28* |
| MP | 59.14 | 56.16 | 52.04 | 47.89* | 41.66* | 46.8* | 30.62* | 27.99* | 20.48* | 15.58* |
| MW | 61.92 | 41.68* | 36.64* | 26.3* | 14.66* | 47.28* | 31.54* | 15.52* | 15.16* | 9.5** |
| PS | 83.54 | 82.78 | 79.68 | 79.46 | 75.16 | 80.06 | 79.8 | 73.2 | 71.5 | 68.78 |
| RMA | 75.1 | 68.23 | 63.28 | 55 | 40.5* | 64.28 | 56.86 | 47.12* | 46.82* | 19.42* |
| CHO | 47.36* | 47* | 37.62* | 35.54* | 29.84* | 28.34* | 28.27* | 25.56* | 24.67* | 7.1** |
| WL | 52.42 | 46.12* | 37.26* | 36.82* | 33.92* | 22.84* | 22.48* | 16.23* | 16.14* | 6.6** |
| WT | 72 | 9.42** | 8.06** | 7.84** | 7** | 63.92 | 6.24** | 5.33** | 5.2** | 4.9** |
| CIPS_star | 83.48 | 60.88 | 49.4* | 47.44* | 42.38* | 77.3 | 45.21* | 36.58* | 33.56* | 23.13* |
| OS | 75.3 | 69.88 | 65.23 | 64.42 | 58.04 | 63.04 | 62.58 | 61.84 | 48.88* | 48.08* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.22: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept Term only, T=25**

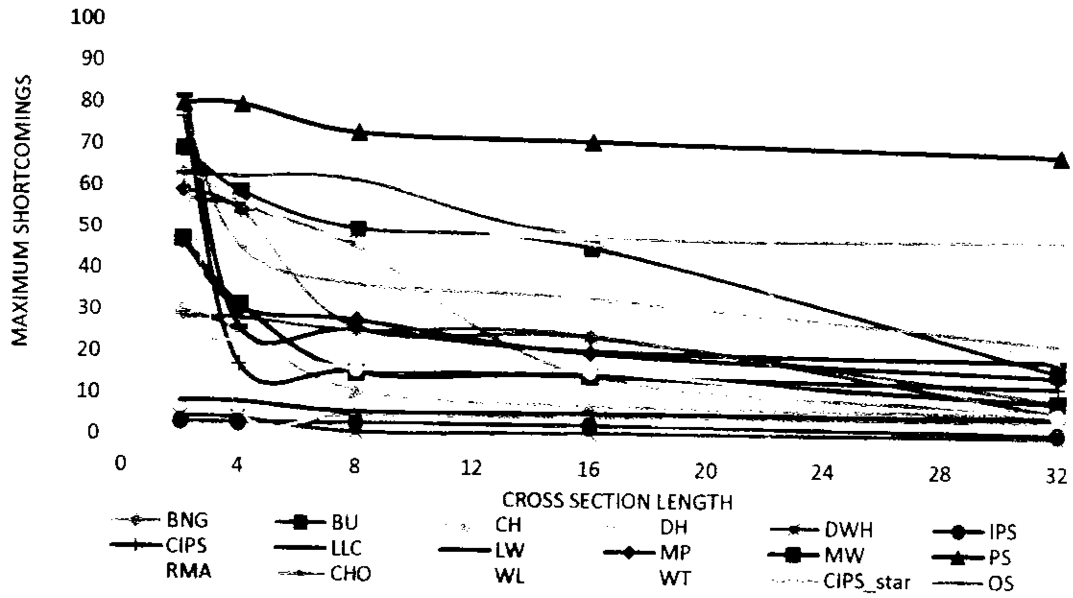



Figure 6.22 and Table 6.13 (Panel B) show the MSC behavior of eighteen PUR tests for medium time series (i.e. T=25) with different cross section units when DGP and test equation take only drift term. A convergence behavior is observed for almost all tests from Figure 6.22 as the cross section unit increases from small to large. At small cross section length (N=2), ten (BNG, BU, DH, CIPS, LW, PS, RMA, WT, CIPS_star, OS) of the PUR tests have shown to be worst performer tests according to their MSC while five (CH, MP, MW, CHO, WL) and three (DWH, IPS, LLC) of the tests have stood mediocre and best performing tests with respect to their MSC respectively. Also, the performance of PUR tests at T=25 and N=2 have been significantly better as compared to their performance at T=10 with same cross section unit.

As the number of cross section increases to another small number 4, then CIPS, LW, and CIPS_star tests having 17%, 26% and 45% MSC are ranked as mediocre performer tests from the previously assigned worst performing tests for N=2. Similarly, WT test with MSC 6.24% from previously assigned MSC of 63% is placed into best performing tests beside DWH, IPS and LLC best performer tests. While, BNG, BU, DH, PS, RMA, and OS tests remain in the worst performer tests having MSC above 50%.

As the number of cross section increases to 8 (i.e. medium) then at medium time series length BNG, BU, DH, and RMA with 26%, 50%, 45%, 47% respectively have gained power and adjusted their places in mediocre group. While, only PS and OS tests with MSC 73% and 61% remain in the worst performer tests. However, the number of best performer tests remains the same that have been observed in the previous cross section unit. At N=16, only OS test has finally adjusted its place into mediocre performing tests while CH has placed into best performing tests as compared to its performance for the last cross section unit. However, only PS test remains in the worst performing test with MSC 71%.

Finally, at cross section unit 32, BNG, DH, MW, CHO, and WL tests with MSC 9.86%, 7.1%, 9.5%, 7.1%, and 6.6% have graded into best performing test category. While, all other tests have remained in the same status as have observed for N=16. Also, PS test with respect to its MSC has been the worst performing test as the cross section level varies from small to large. Moreover, it is observed that DWH, IPS, LLC, and WT tests are the best performer tests as the cross section varies from small to large.

**Figure 6.23: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept Term only, T=50**

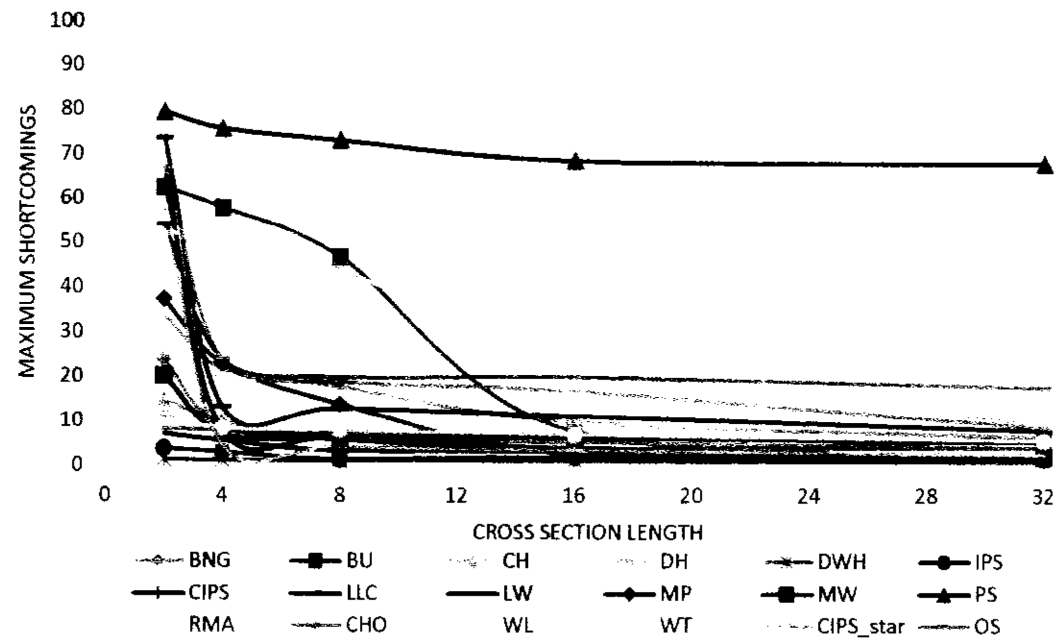



Figure 6.23 and Table 6.14 (Panel A) explain the performance of PUR tests for fixed T=50 with varying cross section units N=2, 4, 8, 16, 32 when data are generated with drift term only. Figure 6.23 demonstrates a convergence pattern for almost all PUR tests which shows an increasing power of these tests as the time series and cross section unit increases. Starting form N=2, Figure 6.23 demonstrates that four (DWH, IPS, LLC, and CHO) of the tests are ranked as best performing tests according to their MSC of 1.3%, 3.84%, 7.02%, and 8.28% respectively. While six tests (BNG, CH, DH, MP, MW, and WL) have assigned as mediocre performer tests according to assigned benchmark. Unfortunately, BU, CIPS, LW, PS, RMA, WT, CIPS_star, and OS tests with 62%, 64%, 73%, 79%, 53%, 56%, 66%, and 54% MSC respectively are classified as worst performer tests.

However, as the number of cross section increases to 4 then majority of the tests show a further improvement and ranked into either mediocre or best performer category. Among best performer tests for N=2, BNG, CH, CIPS, MW, WL, and WT tests have also adjusted their position in the best performing class with 8.54%, 9.9%, 7.92%, 7.1%, 8.4%, and 4.5% respectively. While, in the mediocre performing tests, LW, RMA, CIPS_star, and OS tests with 13.3%, 46.7%, 24.2%, and 23.9% MSC respectively have placed their positions in mediocre group. Further, as the number of cross section increases to 8 then BU test with 46.8% MSC is ordered into mediocre performing category beside DH, LW, MP, RMA, CIPS_star, and OS tests. Moreover, the number of best performer tests remain the same as that have been observed for N=4. At N=8, PS test has defined as worst performing test with 73% MSC.

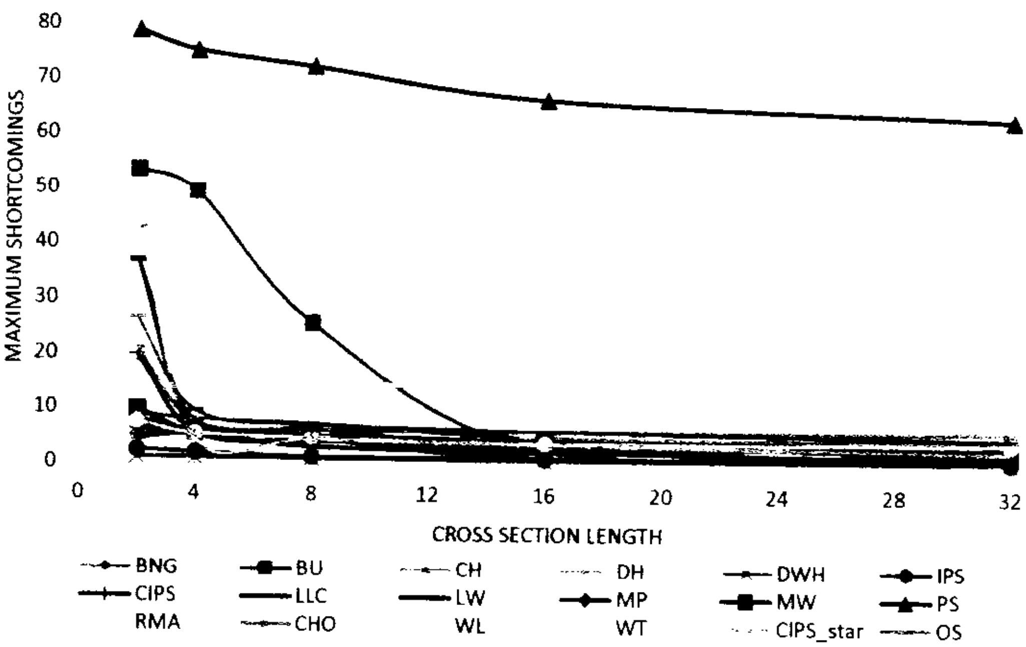
Now, if the number of cross section units increases to 16 and 32, then only PS test remains the worst performer test at both cross section units. While, BU, DH, and MP tests have made their places in best performing tests from mediocre category at N=16 and N=32. Moreover, LW, RMA, CIPS_star, and OS tests with 11.16%, 22.62%, 16.76%, and 19.94% respectively are identified as mediocre tests for number of cross section 16 as well. However, LW, RMA, and CIPS_star tests with MSC 7.9%, 9.42%, and 8.63% respectively have managed to take place in best performer tests category while OS test with no improvement is ranked as in mediocre at N=32. Generally, Figure 6.23 shows a convergence pattern towards zero for almost all tests, excluding PS test, as the number of cross section increases. Finally, at very large cross section unit the number of best performer tests reaches to seventeen.

**Table 6.14: MSC for Null Hypothesis of Unit Root in the Presence of Intercept Case, at T=50 and T=100**

| | Panel A: T=50 | | | | | Panel B: T=100 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Test/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| BNG | 24.06* | 8.54** | 7.04** | 6.22** | 5.2** | 6.5** | 6.1** | 5.2** | 2.1** | 4.9** |
| BU | 62.5 | 57.84 | 46.8* | 7.64** | 5.68** | 53.48 | 49.5* | 25.64* | 2.72** | 2.4** |
| CH | 14.9* | 9.9** | 6.1** | 5.2** | 4.01** | 5.7** | 5.51** | 5.08** | 4.3** | 4.17** |
| DH | 33.78* | 21.78* | 17.79* | 9.9** | 6.4** | 9.08** | 8.08** | 7.34** | 5.01** | 4.5** |
| DWH | 1.3** | 1.14** | 1.02** | 1.04** | 0.81** | 1.2** | 1.03** | 1.01** | 0.8** | 0.87** |
| IPS | 3.84** | 2.98** | 1.39** | 1.9** | 1.66** | 2.5** | 2.08** | 1.27** | 1.2** | 0.88** |
| CIPS | 64.18 | 7.92** | 6.4** | 6.38** | 5.71** | 19.92* | 5.6** | 4.1* | 4.1** | 4.01** |
| LLC | 7.02** | 5.58** | 3.3** | 3.2** | 1.96** | 4.5** | 5.29** | 3.1** | 2.4** | 1.45** |
| LW | 73.72 | 13.3* | 12.9* | 11.16* | 7.9** | 36.9* | 7.1** | 6.4** | 3.02** | 3.6** |
| MP | 37.38* | 22.93* | 13.88* | 2.46** | 1.54** | 8.14** | 5.92** | 0.6** | 1.18** | 1.3** |
| MW | 20.28* | 7.1** | 5.9** | 2.9** | 2.05** | 9.9** | 5.7** | 5.1** | 4.2** | 2.01** |
| PS | 79.48 | 75.74 | 73.08 | 68.5 | 67.88 | 78.82 | 75.2 | 72.26 | 66.3 | 62.86 |
| RMA | 53.62 | 46.78* | 43.84* | 22.62* | 9.42** | 42.86* | 40.67* | 19.06* | 11.04* | 6.6** |
| CHO | 8.28** | 7.9** | 6.6** | 4.01** | 4.12** | 6.5** | 5.3** | 4.7** | 3.13** | 3.25** |
| WL | 11.1* | 8.4** | 7.9** | 6.91** | 5.7** | 7.7** | 5.5** | 4.8** | 3.97** | 4.15** |
| WT | 56.04 | 4.5** | 4.1** | 3.3** | 2.4** | 42.98* | 4.4** | 4.38** | 3.81** | 2.5** |
| CIPS_star | 66.98 | 24.21* | 18.9* | 16.76* | 8.63** | 21.06* | 7.5** | 6.2** | 6.1** | 6.2** |
| OS | 54.1 | 23.9* | 20.01* | 19.94* | 17.78* | 26.6* | 9.6** | 7.2** | 6.1** | 5.06** |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.24: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept Term only, T=100**




At large time series level (T=100) for different cross section units (N=2, 4, 8, 16, 32), Figure 6.24 and Table 6.14 (Panel B) explain the MSC behavior of PUR tests having the null hypothesis of unit root when data generating process and test equation both have same specification in deterministic term (i.e. only drift term).

It is observed from Figure 6.24 that majority of the tests with high power and attained MSC have ranked as best performing tests at N=2. BNG, CH, DH, DWH, IPS, LLC, MP, MW, CHO, and WL tests are classified as best performer tests with MSC 6.5%, 5.7%, 9.08%, 1.2%, 2.5%, 4.5%, 8.14%, 9.9%, 6.5%, and 7.7% respectively. While, six (CIPS, LW, RMA, WT, CIPS_star, and OS) of the test have managed to take place in mediocre test class and only two (i.e. BU and PS) tests have graded into worst performing tests according to their MSC. Also, the number of best performer tests have been increased at N=2 and T=100 as compared to number of best performer tests at T=50 with same cross unit.

As the number of cross section units increases to 4 then five other tests become the part of best performing tests, these tests are CIPS, LW, WT, CIPS_star, and OS beside BNG, CH, DH, DWH, IPS, LLC, MP, MW, CHO, and WL best performing tests at N=2. While one test (i.e. BU) is managed to take its position in mediocre category alongside RMA at N=4. However, due to its bad performance PS test has ranked as worst performing test with its MSC 75%. This result shows that the number of best performer tests have increased at T=100 and N=4 as matched to the number of best performing tests at T=50 and N=4.

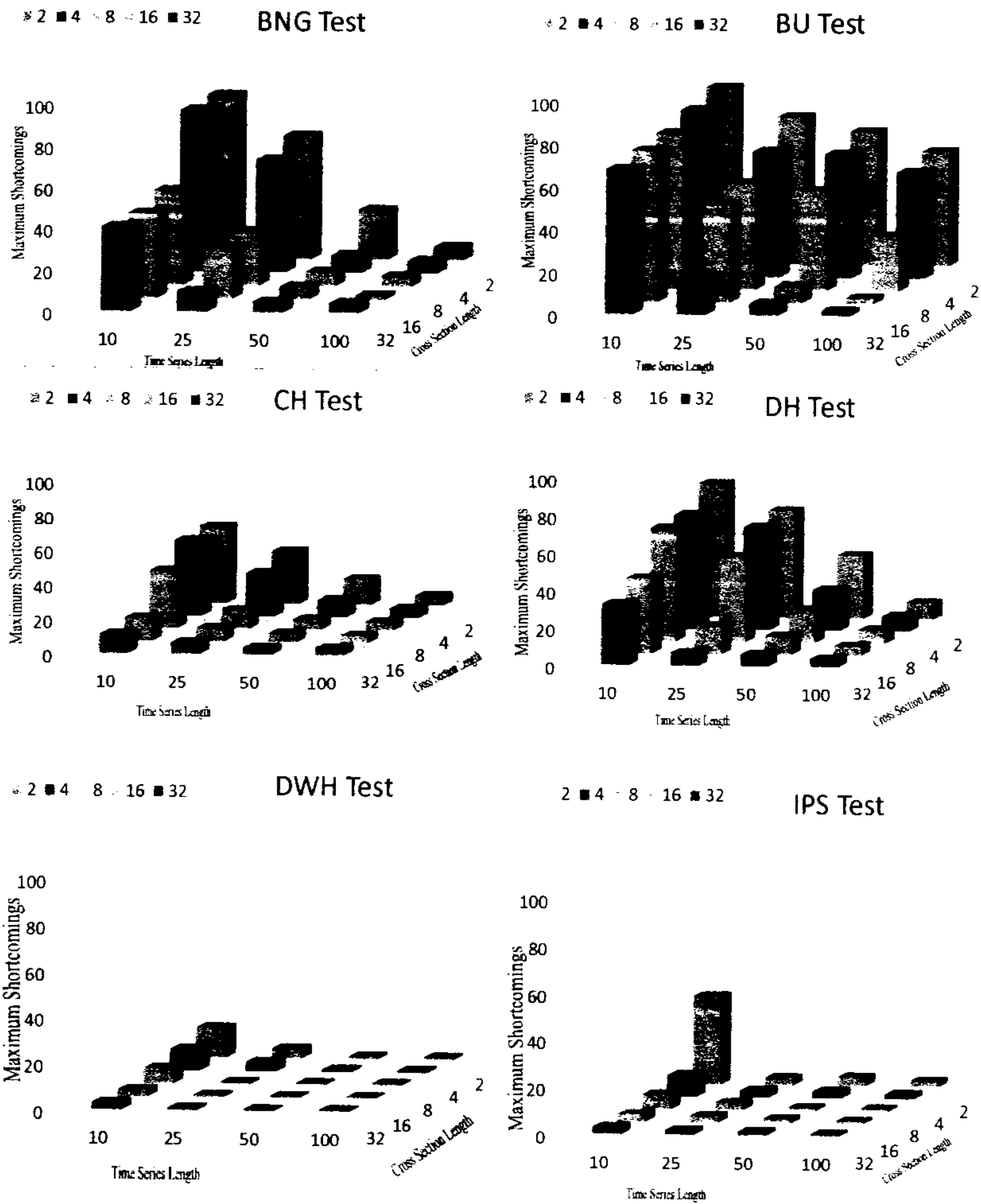
At N=8, Figure 6.24 indicates the same status for all PUR tests that have been observed at N=4. Fifteen (BNG, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, CHO, WL, WT, CIPS_star, and OS) and two (BU and RMA) tests have remained in the best and mediocre performing tests category with further decreased in their MSC while PS test remains the worst performer.

However, as the number of cross section units increases to 16 then BU test also become part of the best performer tests at T=100. While, all other tests have remained in the same status with decreasing pattern in their MSC. At N=32, RMA test is identified as best performer test with 6.6% MSC along with sixteen other best performing tests. PS test with MSC 62% is unable to make its place in mediocre class and has ranked as worst performer test. While, no single test is identified as mediocre one at N=32.

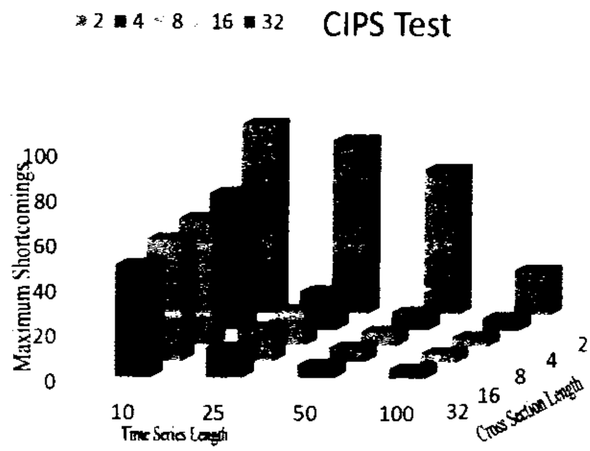
Generally, Figure 6.24 shows a convergence pattern towards zero for almost all tests, excluding PS test, as the number of cross section increases. At very large cross section unit the number of best performer tests reaches to seventeen while PS stands as worst performer one according to its MSC behavior.

Figure 6.21 to Figure 6.24 concludes a very similar picture for the effect of cross section unit over time series that have been observed from Figure 6.1 to Figure 6.5 for all PUR tests when data are generated with drift term only. The effect of cross section unit over time series length shows a convergence behavior for all PUR tests. Also, all these tests have been individually shown in Figure 6.25 corresponding to their MSC for each combination of cross section time series dimensions. This figure indicates best performer tests with their bars downward towards zero while worst performing tests with bar upwards towards 100% at each combination of time series and cross section units which further clarify their positions in our simulation study.

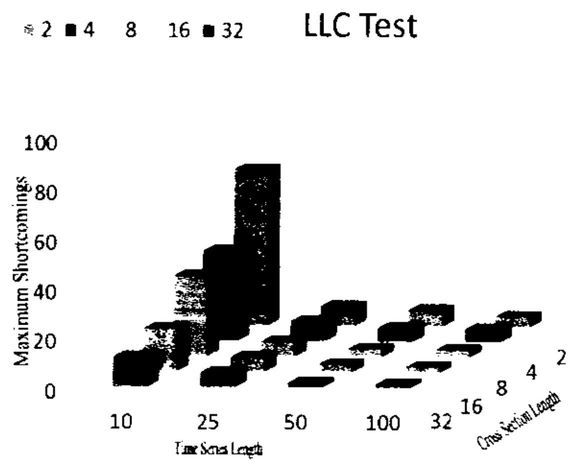
**Figure 6.25: MSC of PUR Tests with Drift Term only**



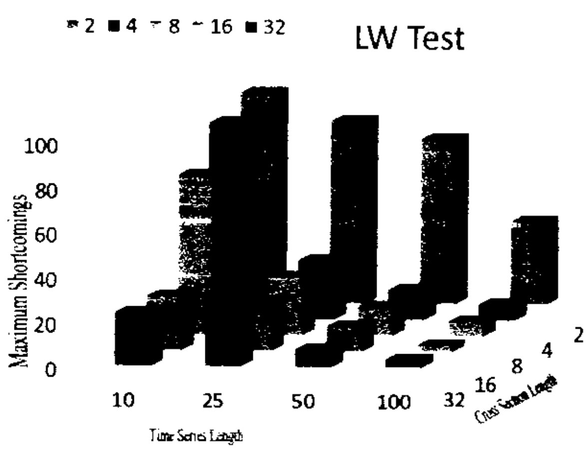
Continue.......



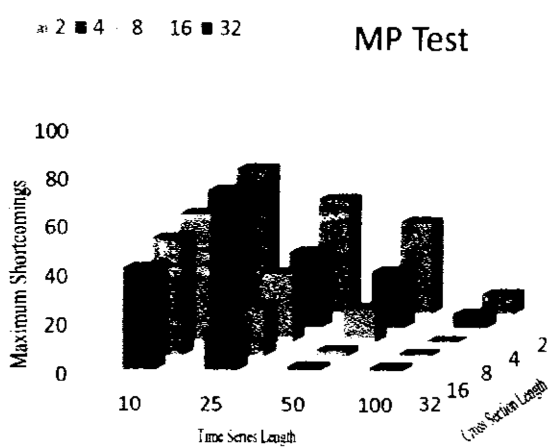
2 4 8 16 32
CIPS Test
Maximum Shortcomings
0
20
40
60
80
100
10
25
50
100
32
16
8
4
2
Time Series Length
Cross Section Length

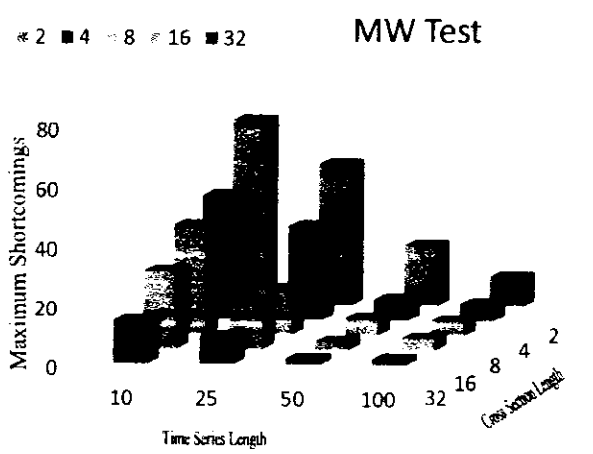
2 4 8 16 32
LLC Test
Maximum Shortcomings
0
20
40
60
80
100
10
25
50
100
32
16
8
4
2
Time Series Length
Cross Section Length

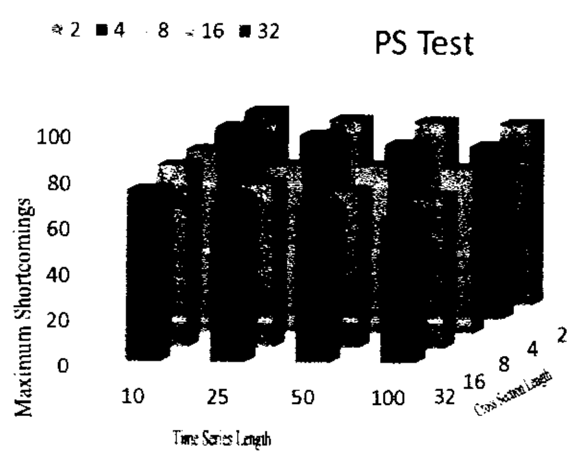
2 4 8 16 32
LW Test
Maximum Shortcomings
0
20
40
60
80
100
10
25
50
100
32
16
8
4
2
Time Series Length
Cross Section Length

2 4 8 16 32
MP Test
Maximum Shortcomings
0
20
40
60
80
100
10
25
50
100
32
16
8
4
2
Time Series Length
Cross Section Length

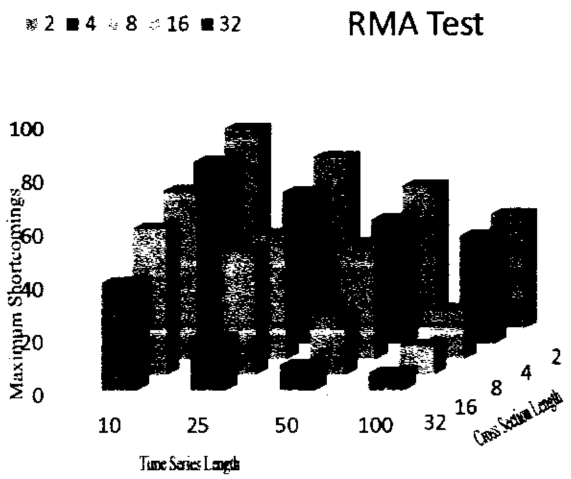



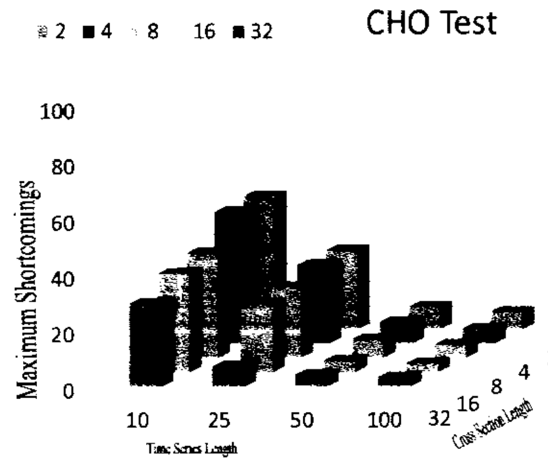
2 4 8 16 32
MW Test
Maximum Shortcomings
0
20
40
60
80
10
25
50
100
32
16
8
4
2
Time Series Length
Cross Section Length

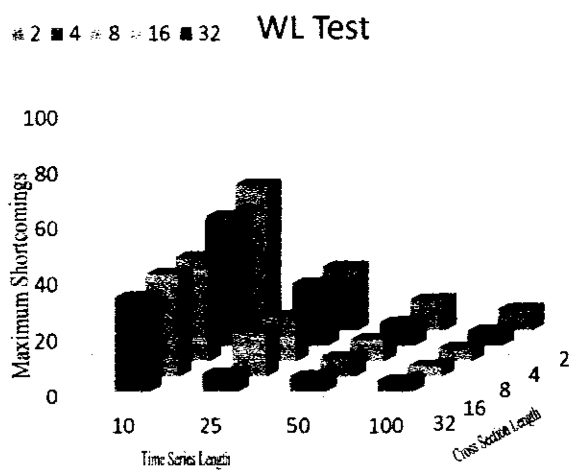



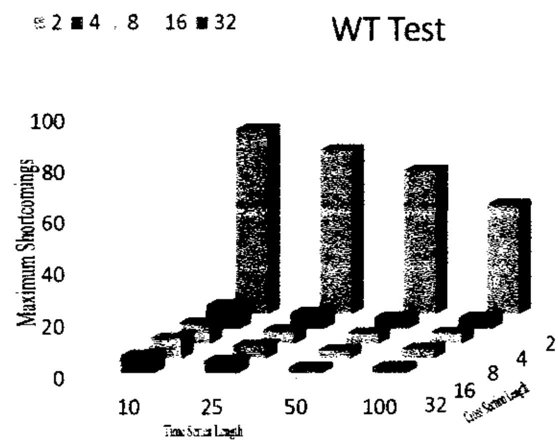
2 4 8 16 32
PS Test
Maximum Shortcomings
0
20
40
60
80
100
10
25
50
100
32
16
8
4
2
Time Series Length
Cross Section Length

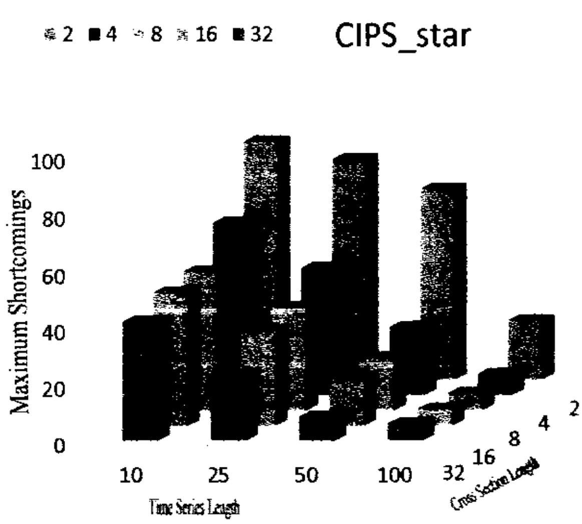
Continue.......

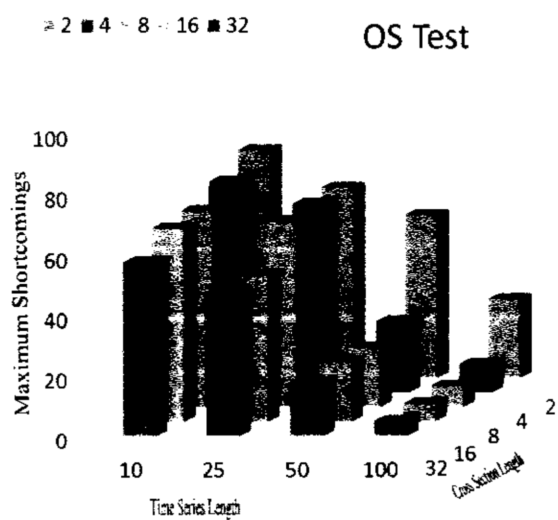

2 4 8 16 32
RMA Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
CHO Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
WL Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
WT Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
CIPS_star
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
OS Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

### 6.4.2. Effect of Cross Section Length on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Unit Root (Intercept and Trend Case)

This section of the power analysis of PUR tests investigate the performance of PUR tests with null hypothesis of unit root when DGP and test equation both have drift and trend terms to check the effect of cross section units over time series length.

Figure 6.26 to Figure 6.29 and Table 6.15 to Table 6.16, examine the effect of cross section units over different time series level to evaluate the performance of PUR tests when specification of DGP and model of the test is same. In these mentioned figures, x-axis shows cross section dimensions and y-axis represents MSC for each time series level.

**Figure 6.26: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend Term, T=10**

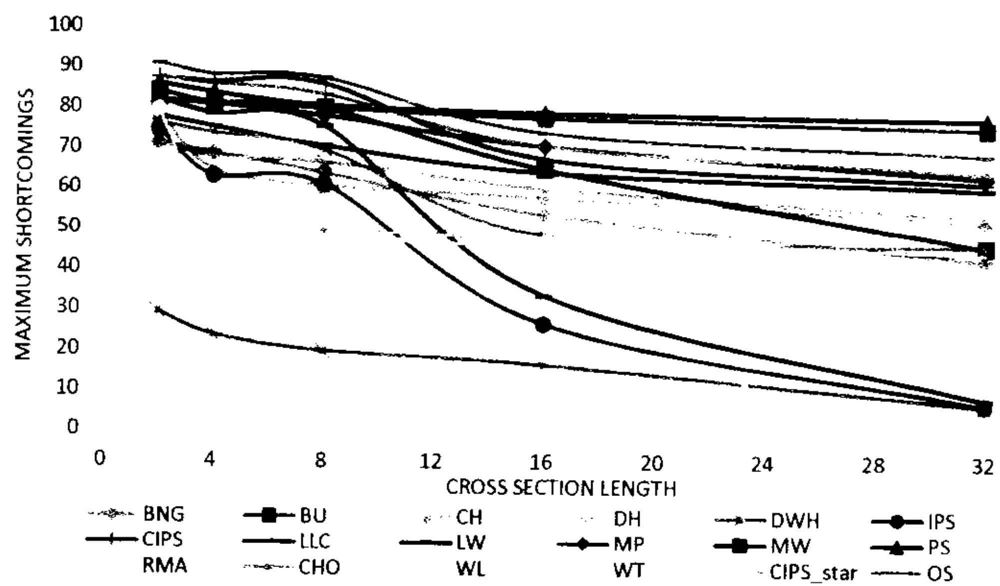


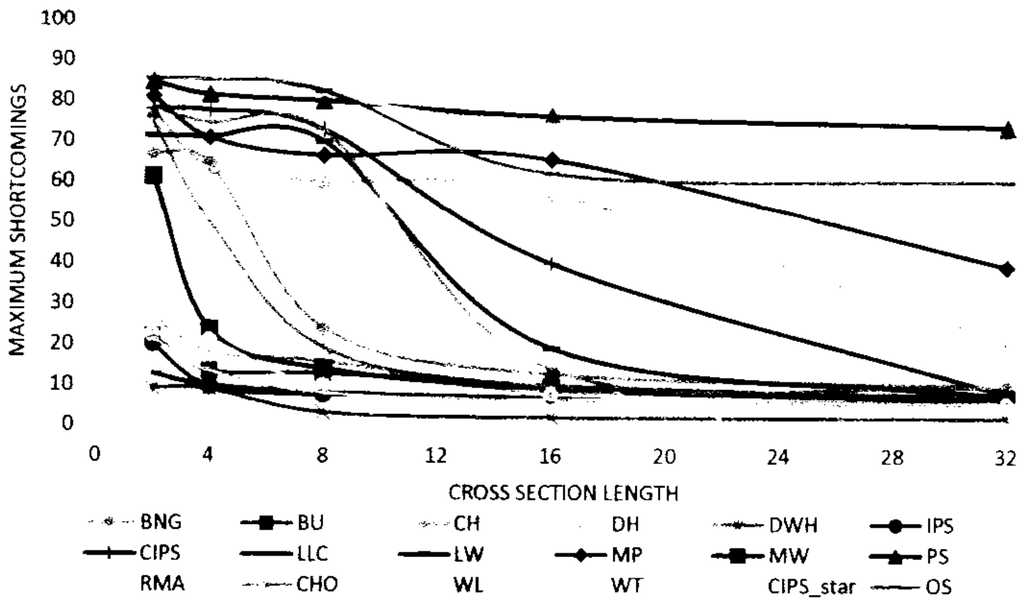
Figure 6.26 and Table 6.15 (Panel A) evaluate the MSC behavior of all tests when the time series level is 10. All of the PUR tests have convergence pattern as the number of cross section units moves from small to large. Starting from N=2, DWH test with MSC 29% is stood as the only single mediocre test while all other tests with MSC over 50% are classified as worst performer tests. As the number of cross sections increases to 4 then all tests have the same status that have assigned previously for N=2 even with decreased MSC. Again, DWH with 23% MSC is identified as mediocre test while no test has adjusted place into best performing tests. Further, as the number of cross section are 8, WL and WT tests are ranked as mediocre tests beside DWH test while all other tests remain in the same status. At medium cross section unit 16 with small time series length 10, IPS, LW, and CHO tests with MSC 26%, 33%, and 48% are graded as mediocre tests beside DWH and WL and WT tests. While, BNG, BU, CH, DH, CIPS, LLC, MP, MW, PS, RMA, CIPS_star, and OS tests with no improvement from rank point of view have ordered as worst performer ones. However, all tests have maintained a decreasing pattern of their MSC. As the number of cross section gets larger (i.e. N=32) then at small time series length DWH, IPS, and LW tests against 6%, 7%, and 8% MSC respectively are classified as best performer ones. Somehow, BNG and MW tests with improved MSC of 42% and 45% respectively have categorized as mediocre tests. While, BU, CH, DH, CIPS, LLC, MP, PS, RMA, CIPS_star, and OS tests with MSC over 50% are identified as worst performer tests. Overall, all tests have shown a downfall pattern as the number of cross section units increases at small time series level. As the number of cross section units gets larger and at last reaches to 32 then maximum number of best performing tests are observed as three while the number of mediocre tests are detected as five.

**Table 6.15: MSC for Null Hypothesis of Unit Root in the Presence of Intercept and Trend Case, at T=10 and T=25**

| | Panel A: T=10 | | | | | Panel B: T=25 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Test/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| BNG | 71.3 | 68.92 | 63.98 | 53.76 | 42.48* | 66.94 | 64.7 | 24* | 12.82* | 9.84** |
| BU | 84.16 | 81.38 | 80.06 | 77.74 | 75.26 | 22.18* | 13.58* | 12.98* | 9.92** | 7.02** |
| CH | 78.64 | 64.78 | 60.82 | 58 | 52.26 | 77.48 | 63.42 | 60.06 | 56.98 | 21.02* |
| DH | 71.72 | 67.82 | 66.64 | 60.38 | 53.34 | 27.16* | 17.96* | 15.98* | 8.44** | 7.95** |
| DWH | 29.46* | 23.82* | 19.86* | 16.6* | 6.82** | 1.78** | 9.06** | 3.18** | 2.02** | 1.91** |
| IPS | 74.82 | 63.12 | 61.22 | 26.64* | 7.08** | 20.18* | 9.81** | 7.7** | 7.51** | 6.9** |
| CIPS | 87.54 | 86.46 | 85.96 | 67.62 | 61.88 | 77.96 | 77.6 | 72.88 | 39.74* | 8.02** |
| LLC | 77.8 | 75.28 | 70.7 | 64.24 | 60.34 | 12.82* | 10.02** | 8.09** | 7.6** | 7.18** |
| LW | 82.22 | 78.52 | 75.64 | 33.56* | 8.38** | 71.4 | 71.09 | 70.02 | 19.2* | 7.55** |
| MP | 81.96 | 80.92 | 78.2 | 70.74 | 63.24 | 80.98 | 70.78 | 66.46 | 65.46 | 38.92* |
| MW | 81.24 | 81.06 | 79.9 | 65.1 | 45.82* | 61.4 | 23.74* | 14.1* | 9.61** | 7.38** |
| PS | 86.03 | 83.72 | 80.64 | 78.9 | 77.58 | 84.62 | 81.42 | 79.88 | 76.02 | 73.44 |
| RMA | 93.89 | 93.87 | 89.56 | 87.64 | 84.52 | 91.9 | 90.18 | 81.4 | 74.22 | 70.49 |
| CHO | 76.28 | 73.74 | 69.7 | 48.66* | 45.94* | 74.6 | 50.07* | 19.3* | 9.92** | 7.51** |
| WL | 79.64 | 52.66 | 47.96* | 43.96* | 22.74* | 22.44* | 21.02* | 10.02** | 7.9** | 5.55** |
| WT | 61.34 | 56.16 | 50.3 | 47.7* | 44.24* | 21.6* | 13.56* | 8.06** | 6.33** | 4.6** |
| CIPS_star | 87.61 | 85.88 | 83.42 | 70.78 | 64.5 | 78.19 | 74.53 | 72.66 | 13.98* | 9.81** |
| OS | 90.98 | 88.34 | 87.5 | 74.06 | 68.92 | 85.4 | 85.08 | 82.49 | 62.01 | 60.02 |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.27: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend Term, T=25**




At time series length 25 with varying cross section units, Figure 6.27 and Table 6.15 (Panel B) represent the convergence behavior of all PUR tests in the presence of both of the deterministic terms. At small cross section unit 2, DWH is stood as the only best performer test with 1.78% MSC. At the same cross section level, BU, DH, IPS, LLC, WL, and WT with 22%, 27%, 20%, 12%, 22%, and 21% respectively are ordered as mediocre tests showing an improvement over N=2 and T=10. While, BNG, CH, CIPS, MP, MW, PS, RMA, CHO, CIPS_star, and OS tests with MSC over 50% have assigned as worst performer tests.

At second small cross section level (i.e. N=4) with medium time series length (i.e. T=25), three of the tests (DWH, IPS, and LLC) are classified as best performer tests with respect to their assigned MSC. Similarly, MW and CHO tests with 23% and 50% MSC are ordered as mediocre tests besides BU, SH, WL, and WT tests. While, BNG, CH, CIPS, LW, MP, PS, RMA, CIPS_star, and OS tests with MSC over 50% have remained in worst performing category.

Further, as the number of cross section units goes to 8 (i.e. medium cross section unit), WL and WT tests against 10% and 8% have managed to make place in best performing tests. Moreover, BNG test is the only single test that has managed its position from worst to mediocre category according to its attained MSC. However, CH, CIPS, LW, MP, PS, RMA, CIPS_star, and OS tests have remained in the same position with MSC over 50%. A further improvement in more tests have been observed as the number of cross section reaches to 16 from 8. Among best performer tests, BU, DH, MW, and CHO tests have also placed their positions with 9.9%, 8.4%, 9.6%, and 9.92% MSC respectively. Similarly, CIPS, LW, and CIPS_star tests are graded into mediocre tests group with MSC 39%, 19%, and 13% respectively. While, CH, MP, PS, RMA, and OS tests have not changed their position with respect to MSC and have maintained in worst performer position.

At large cross section unit 32 with medium time series dimension, Figure 6.27 displays a further improvement in the best performer tests having MSC below 10%. At N=32, BNG, CIPS, LW, and CIPS_star tests have managed to adjust their places in the best performing tests from their previously assigned mediocre performer positions at medium cross section unit 16. While, CH and MP tests with MSC 21% and 38% are ranked into mediocre tests from the worst performing positions. However, PS, RMA, and OS tests have continued to be in the worst performing category according to their MSC.

Overall, Figure 6.27 explains convergence pattern of all tests as the number of cross section units increases from small to large for the same level of time series. Most importantly, the performance of PUR tests from MSC and ranking point of view have been improved at each cross section unit for T=25 (medium) as compared to their performance at T=10 (small) with same cross section units.

**Figure 6.28: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend Term, T=50**

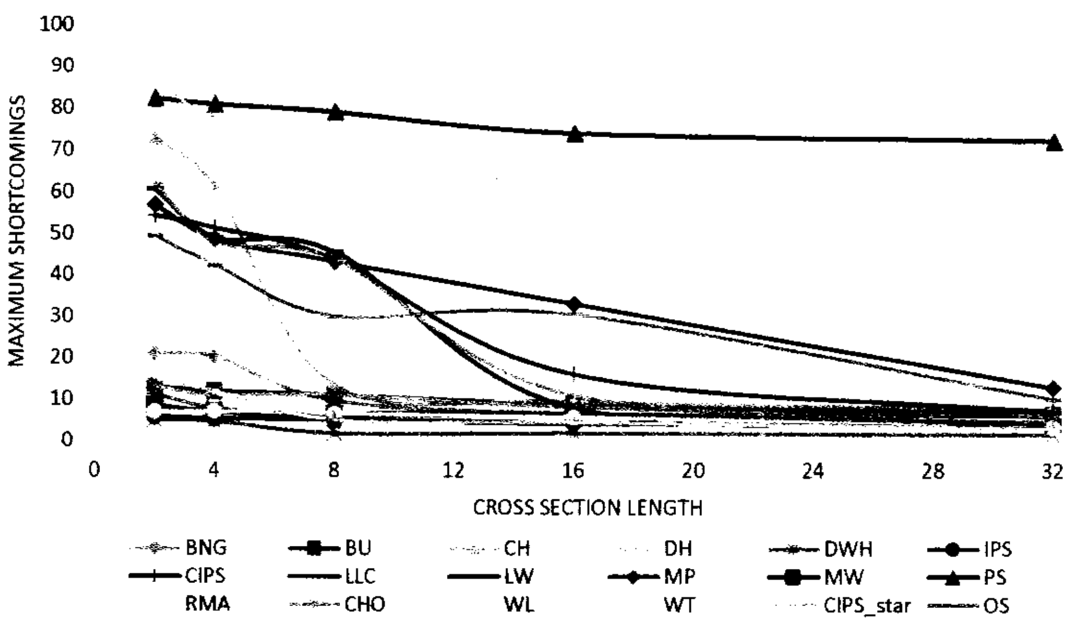



Figure 6.28 and Table 6.16 (Panel A) demonstrate a declining phenomenon of PUR tests as the number of cross section increases from small to large at time series level 50. Starting from small cross section 2, six (DH, DWH, IPS, LLC, MW, and WL) of the PUR tests are classified as best performer tests according to their attained MSC. Similarly, five (BNG, BU, CHO, WT, and OS) tests are ranked into mediocre tests with 20%, 11%, 15%, and 49% MSC respectively. While, CH, CIPS, LW, MP, PS, RMA, and CIPS_star tests have left with worst performer ones according to the assigned benchmark criteria for worst performing tests. Now, as the number of cross section increases to 4 then only CHO test with 7.8% MSC has managed to include in best performer tests. Similarly, LW, MP, and CIPS_star tests with 48.9%, 48.7%, and 47.1% MSC respectively are ranked into mediocre

performer category. However, CH, CIPS, PS, and RMA tests have remained in its previously assigned position of worst performing tests for N=2 with a decreasing pattern in their MSC.

Further, increasing cross section unit from 4 (small) to 8 (medium) for medium time series level of 50, two other tests (BNG and WT) with less than 10% MSC have ordered into best performing tests alongside six previously assigned best performing tests. Moreover, CH and CIPS tests with MSC 13% and 43% respectively are categorized into mediocre performing tests. While, with no progress in their assigned ranked, PS and RMA have placed into worst performing tests. At cross section length 16 for T=50, BU, CH, DH, and LW tests with MSC 8.89%, 9.52%, 8.17% and 7.9% have fulfilled best performing tests criteria and included in best performer tests. However, all other tests with their assigned behavior of MSC have remained in the same status at N=16 as have observed for N=8, with a decreasing pattern in their MSC. Finally, at large cross section unit of 32, CIPS, CIPS_star, and OS tests have assigned as best performing tests with respect to their MSC besides BNG, BU, CH, DH, DWH, IPS, LLC, LW, MW, CHO, WL, and WT tests. While, MP is defined as the only mediocre performing test with MSC 12%. Lastly, PS and RMA tests have failed to include in mediocre or best performer tests category and have graded as worst performer test even at N=32.

**Table 6.16: MSC for Null Hypothesis of Unit Root in the Presence of Intercept and Trend Case, at T=50 and T=100**

| | Panel A: T=50 | | | | | Panel B: T=100 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Test/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| BNG | 20.96* | 20.2* | 9.6** | 6.7** | 6.06** | 10.01** | 6.88** | 5.01** | 4.97** | 4.6** |
| BU | 13.06* | 12.01* | 11.08* | 8.89** | 6.6** | 7.44** | 7.07** | 6.67** | 6.21** | 4.1** |
| CH | 73.26 | 61.18 | 13.52* | 9.52** | 7.7** | 59.06 | 29.81* | 10.1** | 7.55** | 4.69** |
| DH | 12.96* | 11.01* | 11.01* | 8.17** | 6.02** | 6.9** | 5.9** | 5.5** | 4.01** | 4.5** |
| DWH | 5.02** | 4.74** | 1.81** | 1.86** | 1.58** | 4.02** | 4.01** | 0.07** | 0.06** | 0.03** |
| IPS | 5.9*8 | 5.41** | 5.09** | 4.42** | 3.4** | 5.81** | 5.12** | 3.5** | 2.99** | 1.99** |
| CIPS | 54.32 | 51.36 | 43.78* | 16.02* | 7.03** | 41.26* | 24.94* | 18.42* | 9.88** | 6.81** |
| LLC | 8.02** | 7.02** | 5.05** | 4.22** | 3.85** | 5.01** | 5.01** | 2.88** | 1.21** | 1.3** |
| LW | 60.68 | 48.98* | 45.58* | 7.9** | 3.9** | 41.08* | 27.08* | 20.06* | 5.9** | 2.9** |
| MP | 56.84 | 48.7* | 43.18* | 32.94* | 12.86* | 38.44* | 34.6* | 26.86* | 4.4** | 4.16** |
| MW | 8.64** | 7.09** | 6.6** | 6.22** | 4.38** | 6.7** | 5.21** | 3.83** | 2.48** | 2.4** |
| PS | 82.38 | 80.9 | 79.09 | 74.06 | 72.28 | 81.77 | 78.18 | 75.64 | 73.24 | 71.72 |
| RMA | 86.46 | 77.76 | 75.38 | 60.26 | 60.02 | 65.46 | 61.26 | 50.12 | 47.04* | 20.08* |
| CHO | 11.3* | 7.88** | 6.16** | 6.02** | 4.7** | 2.58** | 2.3** | 2.02** | 2.01** | 1.99** |
| WL | 7.1** | 7.06** | 6.6** | 5.7** | 3.31** | 5.5** | 4.9** | 3.5** | 2.23** | 1.7** |
| WT | 15.16* | 11.9** | 5.05** | 4.45** | 3.01** | 8.86** | 5.63** | 2.1** | 1.92** | 1.06** |
| CIPS_star | 61.98 | 47.91* | 43.83* | 10.9* | 6.07** | 47.2* | 22.13* | 20.02* | 4.5** | 4.46** |
| OS | 49.46* | 42.28* | 30.02* | 30.46* | 10.12* | 43.08* | 37.06* | 20.08* | 21.72* | 7.86** |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

**Figure 6.29: MSC Assessments of Tests having Null Hypothesis of PUR with Intercept and Trend Term, T=100**

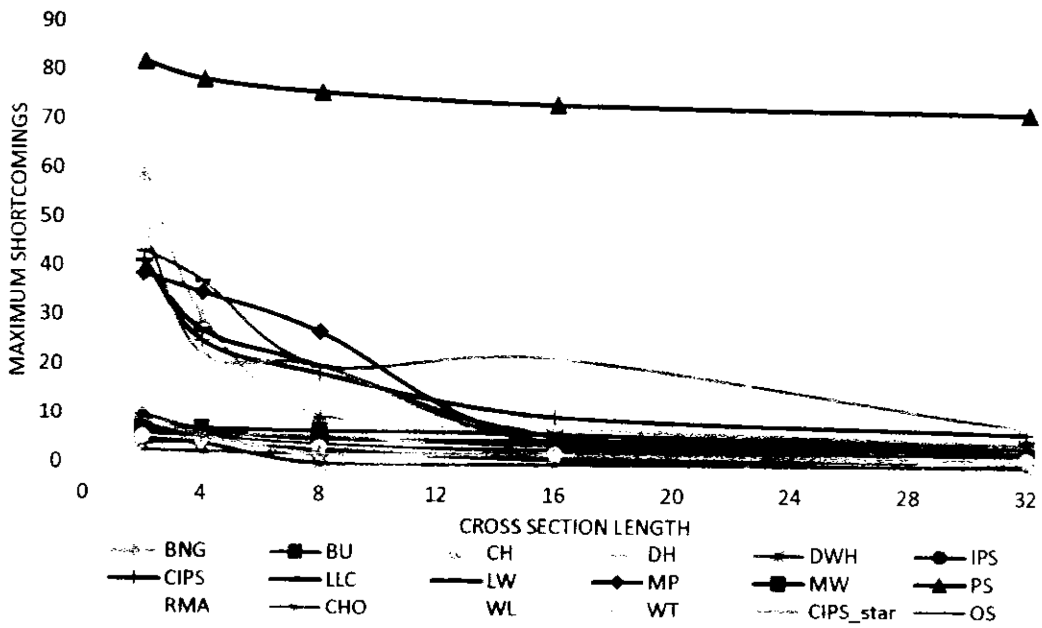



Figure 6.29 explains the performance of PUR tests in the presence of both of the deterministic terms (i.e. drift and trend) for number of time series length 100 with varying number of cross section units. Figure 6.29 and Table 6.16 (Panel B) investigate that all PUR tests have convergence pattern towards zero as the number of cross section units gets larger. At N=2 with large time series level of 100, majority of the PUR tests show a least number of MSC as compared to previous performances of these tests at small cross section units but with small and medium time series length. At this cross section unit ten tests have qualified as best performing tests, these tests are; BNG, BU, DH, DWH, IPS, LLC, MW, CHO, WL, and WT corresponding to MSC 10%, 7.44%, 6.9%, 4.02%, 5.81%, 5.01%, 6.7%, 2.58%, 5.5%, and 8.86%, respectively. Whereas, CIPS, LW, MP, CIPS_star, and OS tests having MSC 41.26%, 41.08%, 38.44%, 47.2%, and 43.08%, respectively are identified as mediocre performer tests. However, only three (CH, PS, and RMA) tests with

MSC over 50% are unable to adjust their places in best or mediocre classes and ranked as worst performing tests.

Similarly, as the number of cross section units reaches to another small cross section unit 4, for large time series dimension all tests become better from their MSC point of view. Here, the number of best performing test remain the same while the number of mediocre performing tests have increased from five to six (CH, CIPS, LW, MP, CIPS_star, and OS) as the number of cross section moves from 2 to 4. For the same cross section unit 4, the number of worst performer tests have also decreased to two (PS and RMA) from three as compared to previous cross section unit. However, as the number of cross section further increases to 8 then one more test (i.e. CH) is added into best performing test category with MSC 10% besides BNG, BU, DH, DWH, IPS, LLC, MW, CHO, WL, and WT tests. Similarly, with one elimination from mediocre tests while one new addition into mediocre tests from worst performing tests, the number of mediocre tests reaches to six (CIPS, LW, MP, RMA, CIPS_star, and OS) with MSC 18.42%, 20.06%, 26.86%, 50%, 20.02%, and 20.08% respectively. PS test with MSC 75% remains the only worst performer test at N=8.

At N=16 with large time series level 100, four more tests have obtained their MSC less than 10% and are ranked as best performer tests. These tests are CIPS, LW, MP, and CIPS_star with MSC 9.88%, 5.9%, 4.4%, and 4.5%, respectively. Similarly, the number of mediocre tests has decreased to two (RMA and OS) while only PS test is detected as worst performing test with MSC 71% at N=16 when number of time series dimension is large (T=100). Finally, as the cross section level reaches to 32, OS test is also ranked as best performing test with MSC 7.86% besides BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, LW, MP, MW, CHO, WL, WT, and CIPS_star tests. Moreover, RMA and PS tests with

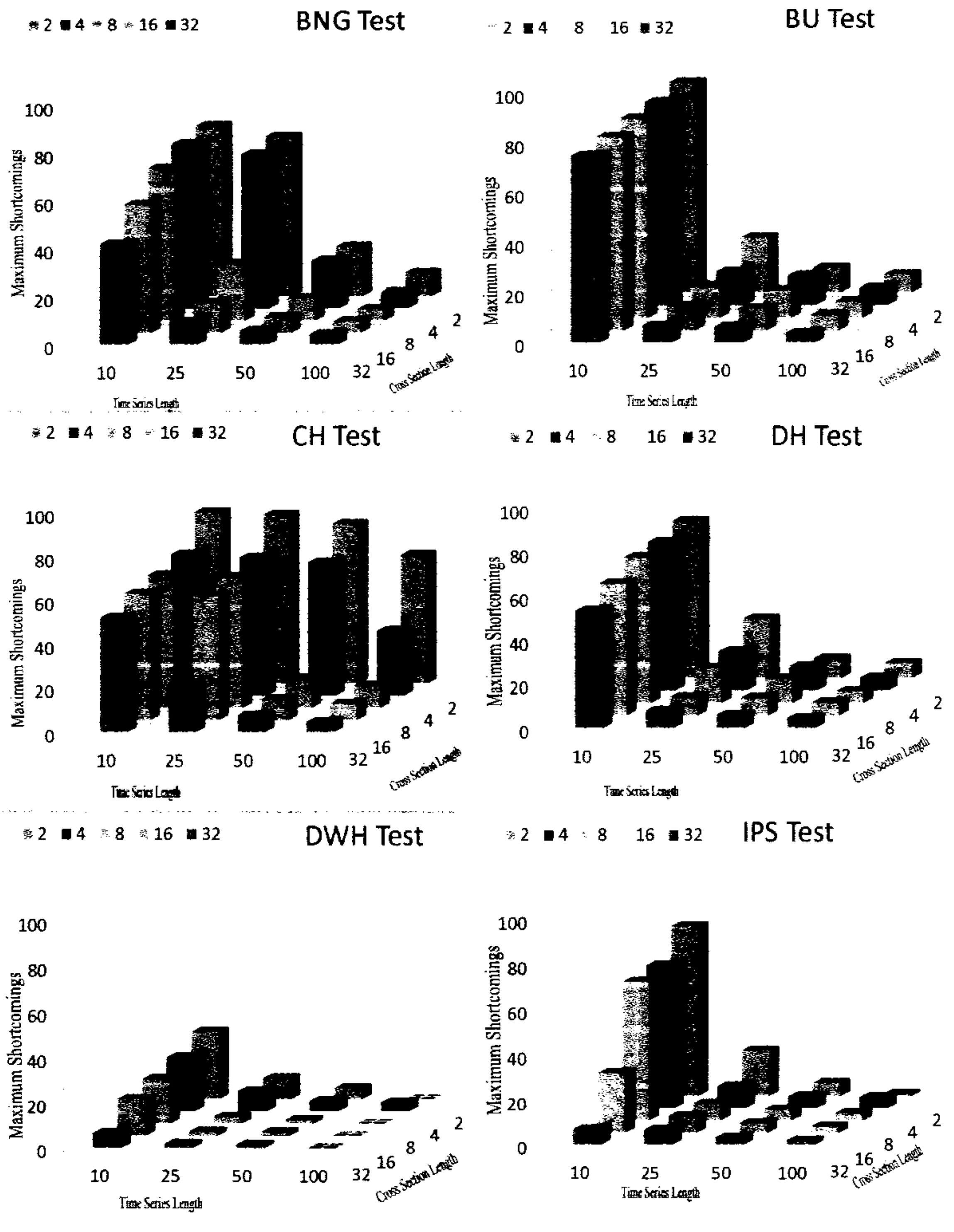
MSC 20% and 71% remain in the mediocre and worst performing tests category. Overall, results of Figure 6.29 shows a better performance of almost all PUR tests as the number of cross section increases and at very large cross section units these all tests will eventually show a more smooth convergence behavior towards zero and will be classified as best performing tests.

Figure 6.26 to Figure 6.29 and Table 6.15 to Table 6.16 demonstrate a very good performance of almost all PUR tests showing a convergence pattern as the number of cross sections moves from small to medium and to large with fixed time series dimensions defined as small, medium and large when data are generated in the presence of both of the deterministic terms. Overall, these figures have identified sixteen tests as the best performer tests at large cross section and time series length. Also, PS test is identified as worst performer test at each level of cross section and time series length but at very large cross section and time series dimensions this test will eventually become the part of either mediocre or best performing class. All these tests have been individually picturized in Figure 6.30 corresponding to their MSC for each combination of cross section and time series dimensions. This figure shows best performer tests with their three dimension bars downward towards zero while worst performing tests with bar upwards towards 100% at each combination of time series and cross section units which further clarify their positions in our simulation study.

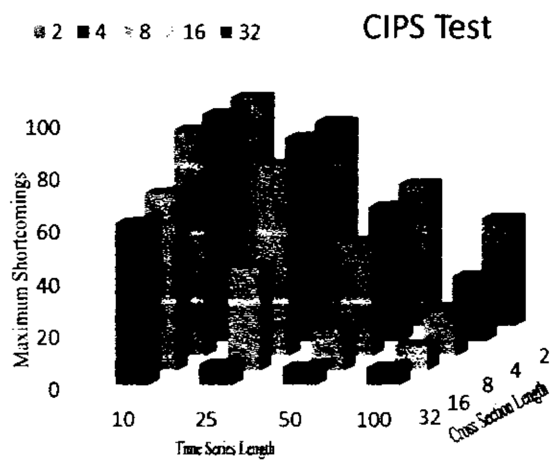
Overall, our findings indicate a lot number of best performing tests having the null hypothesis of unit root at large time series and cross section dimensions, hence it is needed to tighten the best tests criteria and select a few best tests which performs well in small to medium time and cross section units. Further, a benchmark criteria is developed by taking

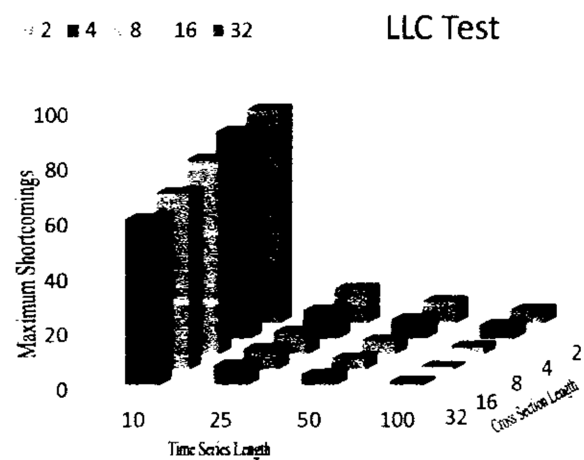
N=4 (small) and N=8 (medium) with T=25 (medium) to pick the best performer tests with respect to their less MSC as compare to other tests. Overall, according to the assigned benchmark criteria Figure 6.1 to Figure 6.10, Figure 6.21 to Figure 6.28, Figure 6.25 and Figure 6.30 suggest, DWH, IPS, LLC, and WT tests as best performing tests in the small, medium, and large time series and cross section dimensions in both situations of deterministic terms. Altogether, if the number of observations is 50 or 100 then these four tests are most preferable over other tests at the same number of observations. Similarly, overall PS test according to its attained MSC is identified as worst performer test whether panel dimension is small, medium or large (see Appendix B, Table B.1 to Table B.6).

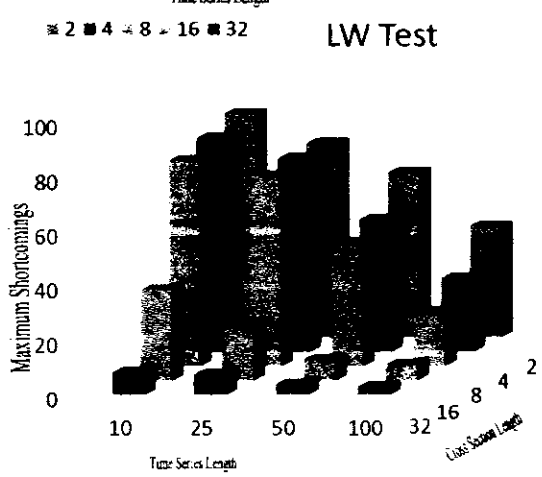
**Figure 6.30: MSC of PUR Tests with Drift and Trend Terms**

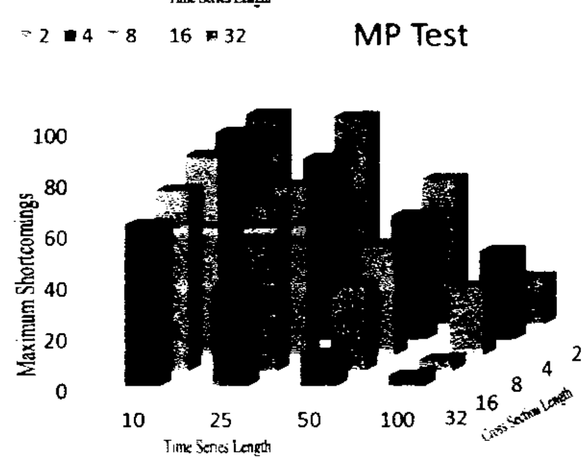


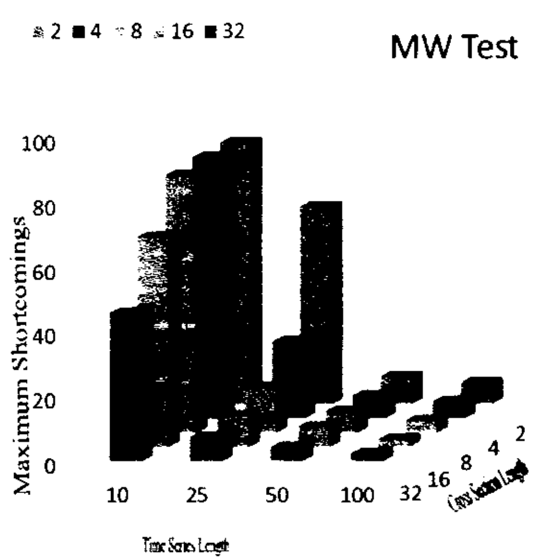
Continue…….

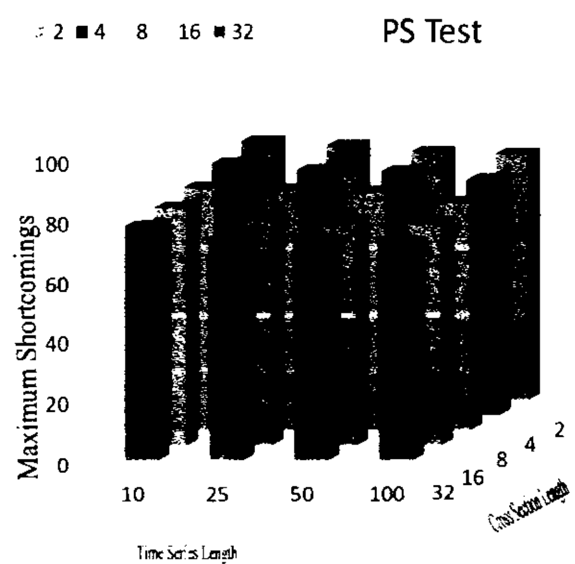

CIPS Test
2 4 8 16 32
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


LLC Test
2 4 8 16 32
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

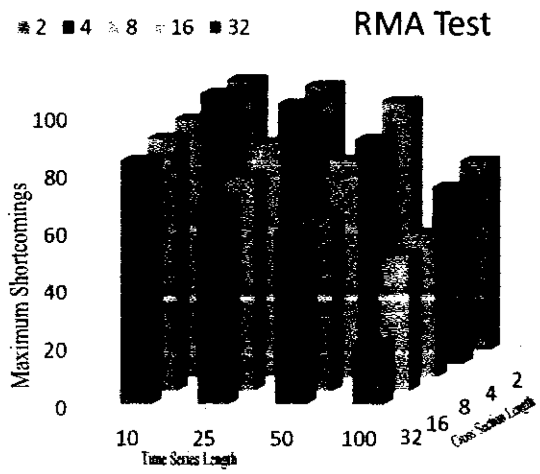

LW Test
2 4 8 16 32
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

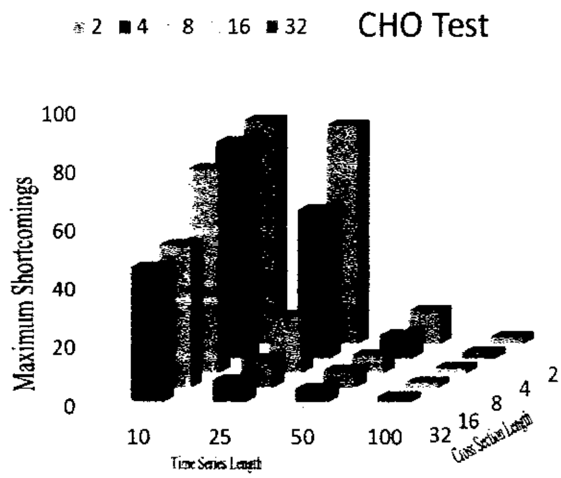

MP Test
2 4 8 16 32
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

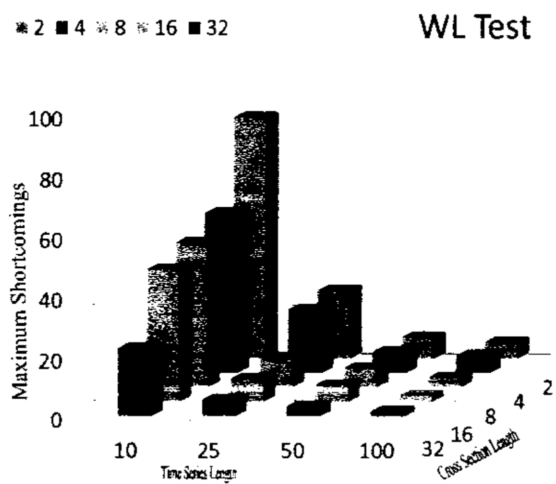

MW Test
2 4 8 16 32
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

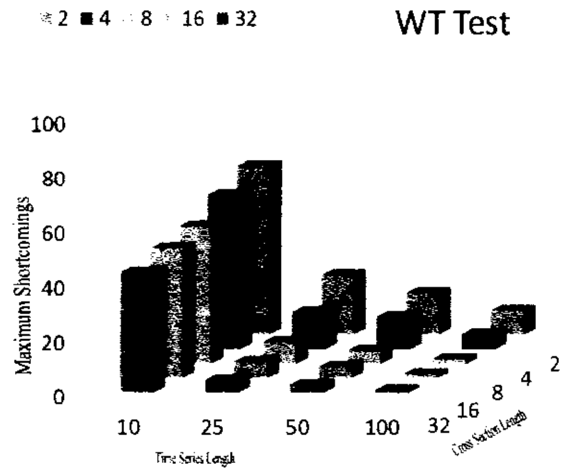

PS Test
2 4 8 16 32
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

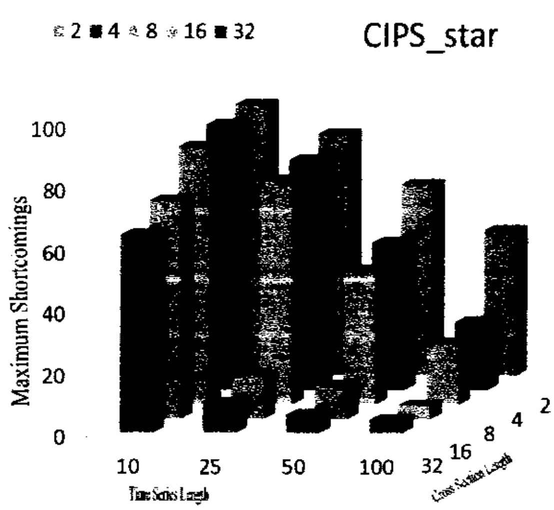
Continue.......

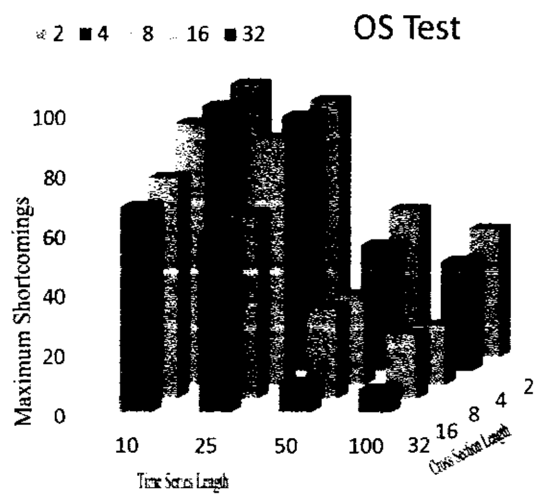

2 4 8 16 32
RMA Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
CHO Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
WL Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
WT Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
CIPS_star
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length


2 4 8 16 32
OS Test
Maximum Shortcomings
0 20 40 60 80 100
10 25 50 100
Time Series Length
32 16 8 4 2
Cross Section Length

### 6.4.3. Effect of Cross Section Length on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Stationarity (Intercept Case)

In this part of power analysis a comprehensive comparative evaluation has been explained for the tests having the null hypothesis of panel stationarity by keeping the varying effect of cross section units (i.e. N=2, 4, 8, 16, and 32) for fixed time series dimension (i.e. T=10, 25, 50, and 100) using Monte Carlo Simulation size of 10,000 when data generating process and test equation have same specification (i.e. with drift term only).

In graphical representation from Figure 6.31 to Figure 6.34, x-axis shows the number of cross units for each level of time series dimension while y-axis displays MSC attained by stationarity tests.

**Figure 6.31: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept Term, T=10**

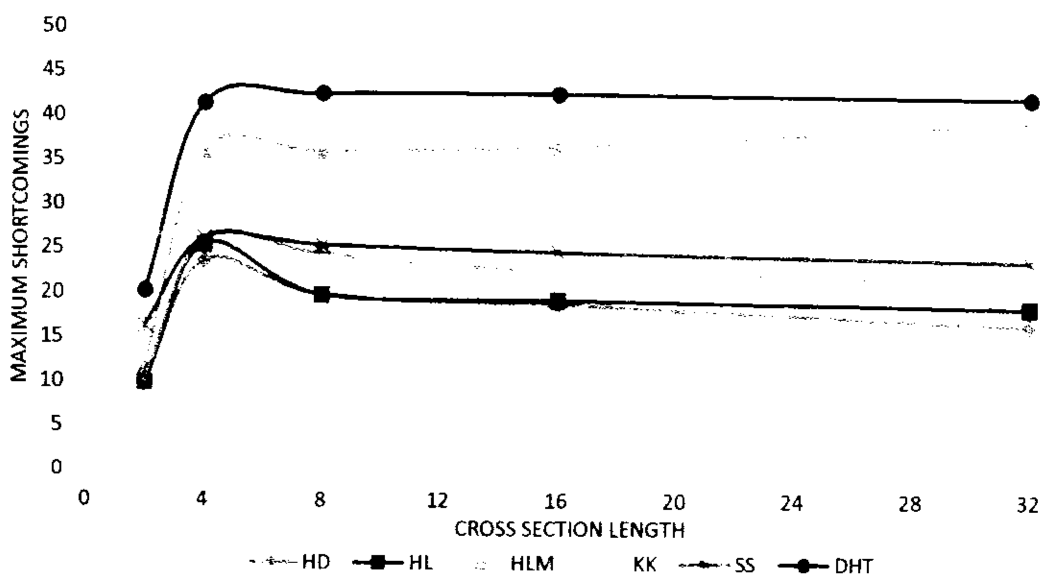


Figure 6.31 demonstrates the MSC behavior of panel stationarity tests to analyze the effect of cross section length on time series length 10 when data are generated with drift term only. At N=2, HD and HL tests with MSC 9.86% and 10% are ranked as best performing tests, however, KK test with 10.3% MSC is ordered into mediocre test class. While, HLM, SS, and DHT tests corresponding to MSC 11.72%, 16.32%, and 20.24% respectively but more than MSC of HD and HL tests have classified into mediocre tests category. This result indicates that HD test has the lowest value and DHT test has the highest value of MSC corresponding to number of cross sections 2.

At N=4, the number of MSC for all tests are more as compared to MSC for N=2 with smallest MSC of 23.58% against HD test while highest MSC of 41.46% in favor of DHT test in the mediocre tests category. Similarly, HL, HLM, KK, and SS tests have also continued to be endured in the mediocre tests with 25.44%, 36%, 26.12%, and 26.17% respectively. However, as the number of cross section units reaches to 8 then HD, HL, KK, and SS tests with decreasing MSC of 19.84%, 19.84%, 24.42%, and 25.44% respectively have also persisted to be in the mediocre tests category. Again, HD test has the smallest value at this cross section unit while DHT test with highest value of 42.54% of its MSC has divergence pattern aside HLM test.

As the number of cross section increases to 16 then all tests have the same status of being in the mediocre tests category, however, the MSC for the convergent pattern stationarity tests have further decreased. Again, HD test with MSC of 18.74% has maintained to remain in the better performing among the mediocre tests category. Moreover, with MSC of 19.28%, HL test has remained in the position of second better performing test in the mediocre tests class. While, KK and SS tests with approximately same MSC are ranked as third and fourth in the mediocre tests at N=16. However, HLM and DHT tests corresponding to MSC of 36.6% and 42.52% respectively have persisted to be in same positions of bad performer tests among the mediocre tests.

**Table 6.17: MSC for Null Hypothesis of Stationarity in the Presence of Intercept Case, at T=10 and T=25**

| | Panel A: T=10 | | | | | Panel B: T=25 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Tests/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| HD | 9.86** | 23.58* | 19.84* | 18.74* | 16.52* | 9.28** | 23.44* | 19.78* | 18.5* | 15.72* |
| HL | 10** | 25.44* | 19.84* | 19.28* | 18.48* | 9.9** | 24.56* | 19.82* | 18.93* | 17.94* |
| HLM | 11.7*2 | 36* | 36.04* | 36.6* | 39.44* | 28.42* | 47.54* | 45.88* | 45.74* | 40.14* |
| KK | 10.3* | 26.12* | 24.42* | 22.54* | 21.66* | 24.72* | 25.68* | 24.38* | 23.18* | 21.54* |
| SS | 16.32* | 26.17* | 25.44* | 24.63* | 23.64* | 26.4* | 26.32* | 26.75* | 24.86* | 24.3* |
| DHT | 20.24* | 41.46* | 42.54* | 42.52* | 42.06* | 26.75* | 42.08* | 42.56* | 42.9* | 42.52* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

Finally, at N=32 the convergent behavioral stationarity tests have further decreasing gain in their MSC while HLM and DHT tests remain in the same status as have been observed for N=16. First, HD and HL tests have kept to place in first two positions among the mediocre tests with MSC 16.52% and 18.48% respectively. Similarly, KK, SS, HLM, and DHT tests with MSC of 21.66%, 23.64%, 39.445, and 42.06% also take place the same place as have been assigned for N=16.

Overall, HD, HL, KK, and SS tests show a decreasing pattern of MSC as the number of cross section moves from 2 to 32 and majority of these tests remain in the mediocre tests category, excluding HD and HL tests which are identified as best performing tests at N=2. While among these tests, HD test with its minimum value of MSC at each level of cross section unit is identified as most stringent test. However, only HLM and DHT tests have detected as bad performing tests among the mediocre tests class with high MSC over 30% against number of cross 4 to 32, excluding HLM test with MSC 11.72% and 20.24% respectively. Also, in Figure 6.31 the lowest number of MSC at N=2 shows a low power behavior of stationarity tests and point optimal test which have also been observed for the effect of time series on number of cross units (see Table A.1 (Panel A) in Appendix-A).

**Figure 6.32: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept Term, T=25**

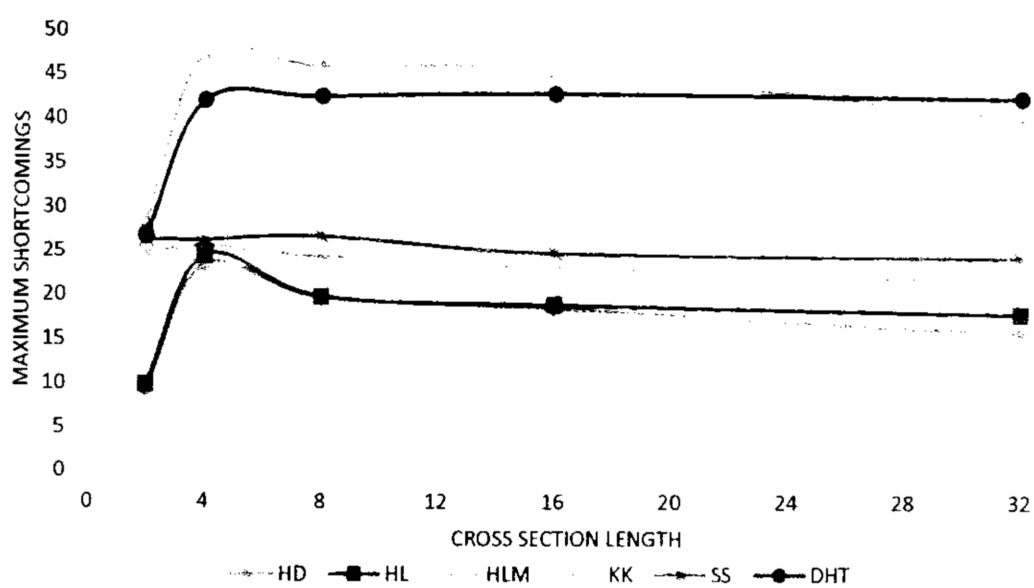



Figure 6.32 and Panel B of Table 6.17 determine the MSC performance of stationarity tests at time series length of 25 over the number of cross section units 2, 4, 8, 16, and 32. At N=2, only two tests (HD and HL) are diagnosed as best performing tests with respect to MSC of 9.28% and 9.9% respectively. At the same cross section unit, HLM, KK, SS, and

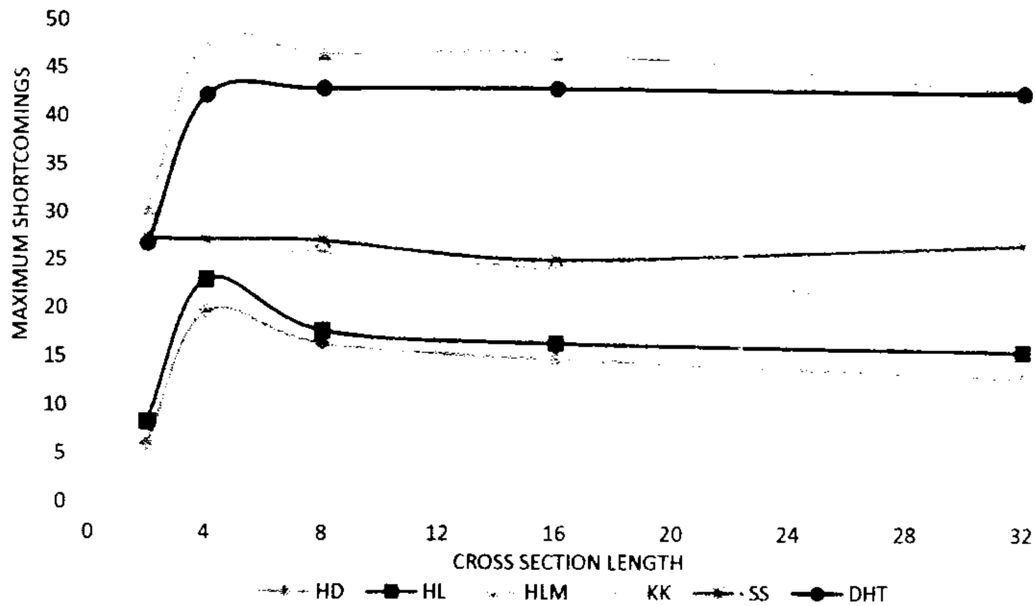
DHT tests with MSC 28.42%, 24.72%, 26.4%, and 26.75% respectively are graded as mediocre tests.

As the number of cross section increases to 4 then HD and HL tests switch into mediocre tests category with MSC 23.44% and 24.56% respectively. Similarly, KK and SS tests have kept their positions in the mediocre tests category but with little increase in their MSC. However, DHT and HLM tests are classified as the bad performing tests in the mediocre tests class with MSC 42.08% and 47.54%. Further, increasing cross section unit to 8, HD and HL tests with MSC 19.78% and 19.82% respectively have gained a less value of MSC as compared to their previous cross section counterpart. Also, other tests have also kept their positions in the mediocre tests category with respect to MSC 24.38%, 26.75%, 42.56% and 45.88% corresponding to KK, SS, DHT, and HLM tests respectively in which last two tests are assigned with the same previous status according to cross section unit.

At N=16, all six test remain in the same type of class with minimum value of MSC against HD test as 18.5% while maximum value of MSC of 45.74% against HLM test. Lastly, when the number of cross sections are 32 then again all tests with the same status remain in mediocre tests class in which a MSC of 15.72%, 17.94%, 40.14%, 21.54%, 24.3%, and 42.52% against HD, HL, HLM, KK, SS, and DHT tests have been observed. These results indicate a better performance of HD and HL tests while bad performance of DHT and HLM tests over all cross section units. However, HD test stands as the most stringent tests with a little improvement over HL test at all cross section units. Also, a convergence pattern has been attained by HD, HL, HLM, KK, and SS tests over increasing number of cross section units, excluding N=2. At N=2, the low value of MSC indicate that stationarity tests and point optimal test have very low and close power at each alternative. However, other than

N=2 all cross section units show a good power behavior of stationarity tests and more powerful performance of point optimal test over stationarity tests.

**Figure 6.33: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept Term, T=50**




At T=50, a further investigation of stationarity tests have been analyzed in Figure 6.33 and Panel A of Table 6.18 with intercept term only. Like previous results, a similar picture has been observed for N=50 over small to large cross section units. First, at N=2, HD and HL tests with MSC of 6.025 and 8.36% are classified as best performing tests while HLM, KK, SS, and DHT tests against MSC of 30.51%, 25.06%, 27.385, and 26.82% respectively are placed into category of mediocre tests.

Second, as the cross section unit increases to 4 then HD and HL tests switch their previous best performing positions into mediocre performing tests with 19.72% and 23.02% of MSC respectively. However, HLM, KK, SS, and DHT tests have remained in the mediocre tests class with MSC of 47.96%, 25.16%, 27.26%, and 42.26% respectively indicating a more increase of MSC for HLM and DHT tests thus defined as bad performing tests in the mediocre class. Also, HD test has the least value while HLM test with highest value of their MSC at N=2 and N=4.

Third, with further increase of cross section unit to 8, Figure 6.33 demonstrates a same situations for all test as have been observed for N=4 with HD and HL tests as the better performing tests regarding their MSC while HLM and DHT tests with MSC of 46.57% and 42.94% remains in bad performing category in the mediocre tests category. Fourth, with N=16, HD and HL tests corresponding to MSC of 14.94% and 16.56% have again stood as the better performer tests in the mediocre tests class. Similarly, KK and SS tests with approximately same value (i.e. 24.16% and 25.22%) of MSC remain in their previous assigned positions in the mediocre tests class. While, as usual, HLM and DHT tests having MSC far away from other mediocre tests MSC have continued to be in the last positions.

Lastly, with number of cross sections 32, KK test has further decreased its MSC from 24.16% to 20.08% indicating that, this test, eventually gets MSC very close to zero as the size of cross section units gets larger. Similarly, HD and HL tests with further improvement in their powers and with assigned MSC of 13.2% and 15.86% have again managed their top positions in the mediocre tests category. Moreover, HLM and DHT tests have also identified as the bad performing tests in the mediocre test class with MSC over 40%. These results show good performance of HD and HL tests with convergence behavior as the number of cross sections increases from 2 to 32 for T=50, however, between these two better performing tests in the class of mediocre tests HD test has identified as the most stringent test. While, HLM and DHT tests with bad performance in the mediocre tests category have detected as bad performing tests over number of cross sections 2 to 32. However, at N=2, the power of stationarity tests and point optimal test are very low at each alternative thus indicating minimum value of MSC, at N=2, as compared to MSC of stationarity and point optimal tests at other cross section units.

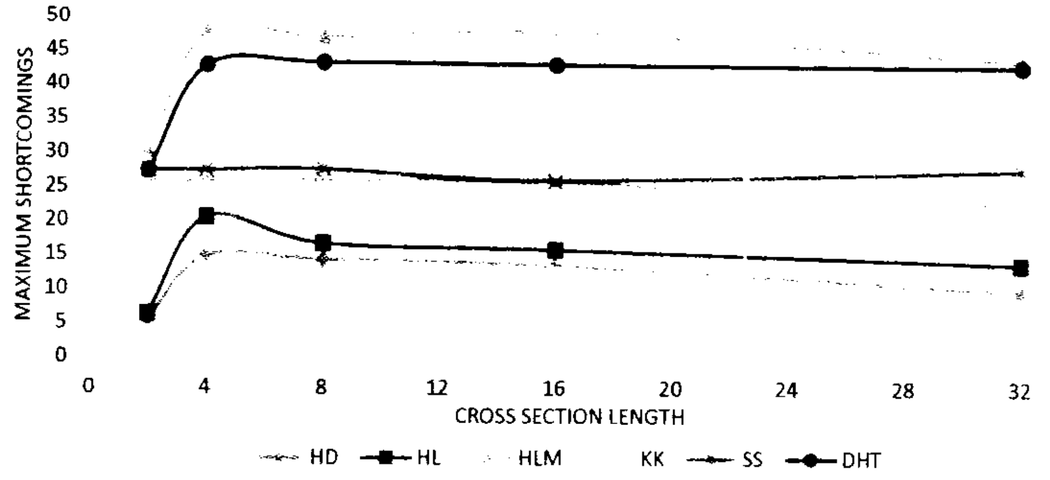
**Table 6.18: MSC for Null Hypothesis of Stationarity in the Presence of Intercept Case, at T=50 and T=100**

| | Panel A: T=50 | | | | | Panel B: T=100 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Tests/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| HD | 6.02** | 19.72* | 16.46* | 14.94* | 13.2* | 5.62** | 14.84* | 14.38* | 13.54* | 9.82** |
| HL | 8.36** | 23.02* | 17.82* | 16.56* | 15.86* | 6.46** | 20.58* | 16.74* | 15.86* | 13.74* |
| HLM | 30.51* | 47.96* | 46.57* | 46.76* | 42.76* | 30.52* | 47.99* | 47.03* | 47.8* | 43.56* |
| KK | 25.06* | 25.16* | 25.94* | 24.16* | 20.08* | 25.66* | 25.94* | 26.02* | 25.78* | 22.68* |
| SS | 27.38* | 27.26* | 27.24* | 25.22* | 26.95* | 27.56* | 27.46* | 27.68* | 26.08* | 27.58* |
| DHT | 26.82* | 42.26* | 42.94* | 42.98* | 42.62* | 27.5* | 42.88* | 43.34* | 43* | 42.73* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

When the time series increases to 100 then the performance of stationarity tests regarding to their gained MSC under 50% are explained in Figure 6.34 and Table 6.18 (Panel B) when data are generated in the intercept term only. At N=2, all stationarity tests have minimum value of 5.62%, 6.46%, 30.52%, 25.66%, 27.56%, and 27.5% of their MSC corresponding to HD, HL, HLM, KK, SS, and DHT tests respectively as compared to MSC of these tests over other cross section units. Only, HD and HL tests with MSC of 5.625% and 6.46% are assigned as best performer tests among four (HLM, KK, SS, and DHT) mediocre tests at N=2.

**Figure 6.34: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept Term, T=100**




At N=4, the MSC of all tests have high value as compared to N=2, however, HLM, KK, SS, and DHT tests have remained in the mediocre tests class while HD and HL tests with MSC of 14.84% and 16.74% have switched their positions form best performing test class to mediocre class beside four other tests. Further, as the number of cross section units increases to 8 then all stationarity tests have the same status as have been explained at N=4 with approximately same MSC value of HD, HLM, KK, SS, and DHT tests while a little decrease in the value of HL test has been observed. Moreover, at N=16, HD and HL test with MSC of 13.54% and 15.86% have classified as the better performing tests among HLM, KK, SS, and DHT tests having values of their MSC 47.8%, 25.78%, 26.08%, and 43% in the mediocre tests category. At large cross section unit of 32, only HD test with value of MSC of 9.82% has managed to take its position again in best performing test class. While HL test with more close to best performing test category with respect to MSC value of 13.7% is remained in the mediocre class besides HLM, KK, SS, and DHT tests having MSC value of 43.6%, 22.7%, 27.6%, and 42.7% respectively.

Overall, the results of Figure 6.34 indicate HD test as the most stringent test having a convergence pattern from N=4 to N=32 and has assigned as best performer test twice as compared to other mediocre tests. HL test is ranked as the second better performing test in the class of mediocre tests with convergence behavior over N=4 to N=32. While, KK and SS tests with approximately same pattern over N=2 to N=32 is assigned as average performing tests in the mediocre tests class. However, HLM and DHT tests are ranked as the bad performing tests in the mediocre test class with MSC over 40% at N=4 to N=32. Finally, the minimum value of MSC of all tests at N=2 indicate a low power of stationarity tests and point optimal test as compared to power of these tests at other cross section units and T=100. Generally, HD test is assigned as most stringent test at each level of time series length (i.e. t=10, 25, 50, and 100) with varying number of cross section units (i.e. N=2, 4, 8, 16, and 32) when DGP and test equation have the same specification (i.e. drift term only). While, HLM and DHT tests are categorized as bad performer tests among mediocre tests at each level of time series and cross section units.

**Figure 6.35: MSC of Panel Stationary Tests with Drift only**

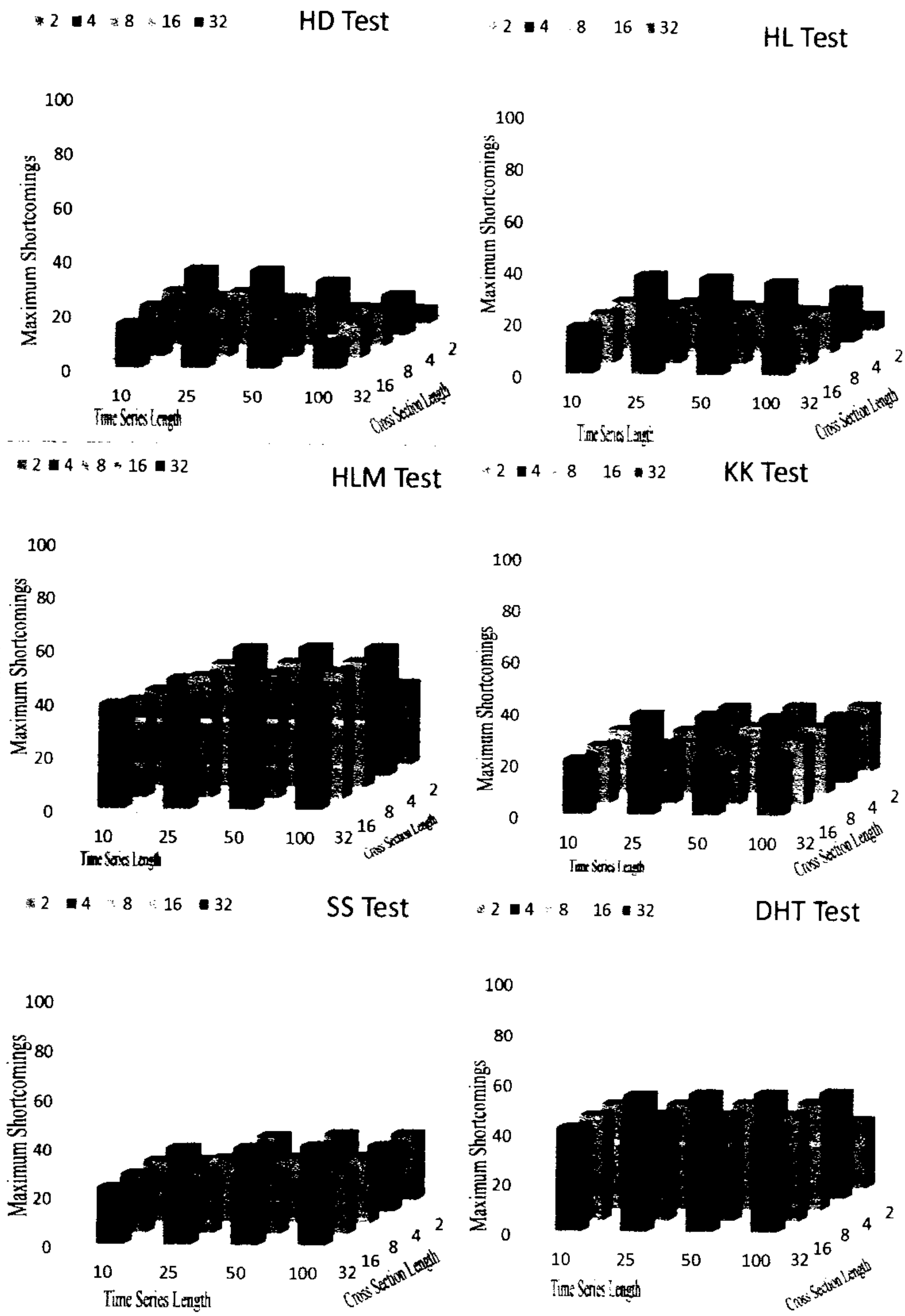


Figure 6.31 to Figure 6.34 conclude a very similar picture for the effect of cross section unit over time series that have been observed from Figure 6.11 to Figure 6.15 for all panel stationarity tests when data are generated with drift term only. The effect of cross section unit over time series length shows a convergence behavior for all panel stationarity tests. Also, all these tests have been individually shown in Figure 6.35 corresponding to their MSC for each combination of cross section time series dimensions. This figure indicates best performer tests (i.e. HD and HL) with their bars downward towards zero at each combination of time series and cross section units which further clarify their positions in our simulation study.

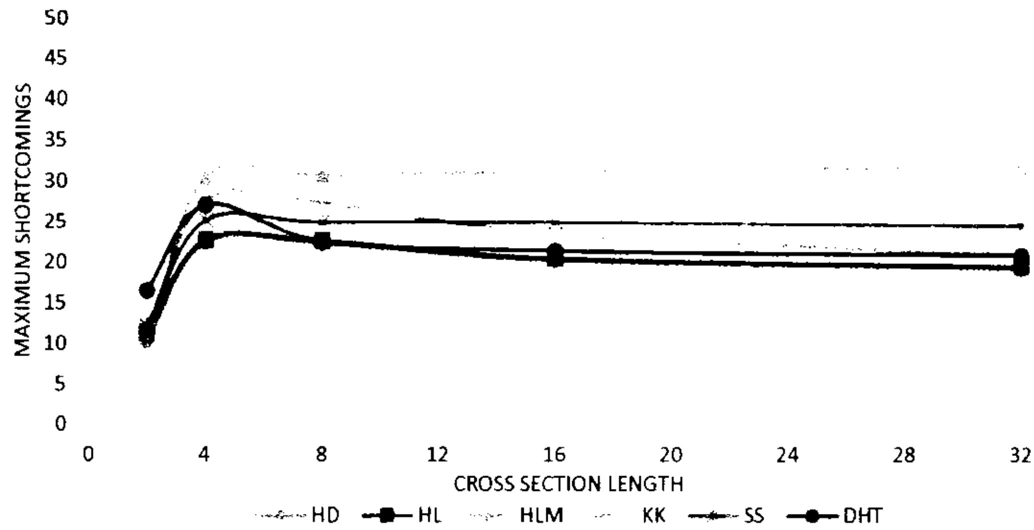
### 6.4.4. Effect of Cross Section Length on Maximum Shortcomings Evaluation of the Tests having the Null Hypothesis of Stationarity (Intercept and Trend Case)

This section demonstrates the performance of stationarity tests on the basis of MSC to see the effect of cross section length over fixed time series length to identify better and bad performer tests in the mediocre tests category when DGP and test equation have same specification of the deterministic term, here in this case both drift and trend terms, using a Monte Carlo size of 10,000.

Figure 6.36 to Figure 6.39 and Table 6.19 to Table 6.20 explain behavior of these tests. In these figures, x-axis indicates cross section length while y-axis displays MSC of stationarity tests.

Figure 6.36 and Table 6.19 (Panel A) display the results of stationarity tests at T=10 with varying number of cross section units in the case of both intercept and trend terms. When the number of cross section units are 2 then all tests have MSC in between 9% to 17% in which HLM is ranked as best performer test with 9.7% MSC while DHT test with MSC of 16.56% is classified as bad performer in the mediocre class. However, HD and HL tests having MSC of 10.24% and 11.44%, which are very close to best tests criteria, are graded as mediocre tests beside KK and SS tests with MSC below 15% at N=2.

**Figure 6.36: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend Terms, T=10**




HD and HL tests have remained in the mediocre test class with 22.4% and 22.84% value of MSC but have ordered as better tests in the mediocre class when the number of cross section are 4. While, HLM has lost its previous best performing position at N=2 and has ranked as mediocre test with MSC of 30.5% at N=4. Similarly, KK, SS and DHT tests with high value of their MSC, as compared to N=2, have kept their same positions as the number of cross section units reaches to 4.

Further, as the number of cross units increases to 8 then Figure 6.36 indicates that all tests remain in their previous positions with HD and HL tests as better performing tests in the mediocre class with 22.38% and 22.7% MSC, respectively. While, HLM, KK, SS, and DHT tests have managed to sustain their positions in the mediocre test class with 30.78%, 27.44%, 25.03%, and 22.64% MSC, respectively. At N=16 and 32, HD test has continued to lie in top position in the mediocre tests category with MSC of 20.3% and 19.26%, respectively. While HL test with MSC of 20.52% and 19.56% against N=16 and N=32 is identified as the second better performing test in the mediocre tests category. Moreover, KK and DHT tests with their MSC of 23.56%, 21% and 21.46%, 2.94% against N=16 and N=32 respectively have sustained to get closer to the MSC of HD and HL tests. However, HLM test with 31.3% and 31.98% MSC in favor of N=16 and N=32 is continued to remain in bad performer test in the mediocre tests category.

**Table 6.19: MSC for Null Hypothesis of Stationarity in the Presence of Intercept and Trend Case, at T=10 and T=25**

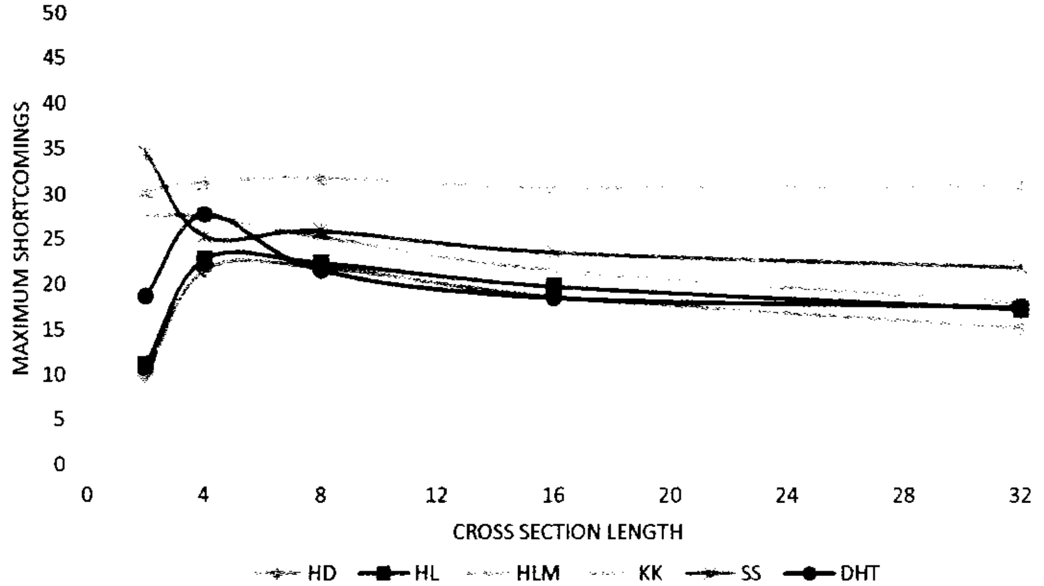
| | Panel A: T=10 | | | | | Panel B: T=25 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Tests/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| HD | 10.24* | 22.4* | 22.38* | 20.3* | 19.26* | 9.96** | 21.56* | 21.92* | 18.68* | 15.22* |
| HL | 11.44* | 22.84* | 22.7* | 20.52* | 19.56* | 11.2* | 22.78* | 22.44* | 19.82* | 17.3* |
| HLM | 9.7** | 30.5* | 30.78* | 31.3* | 31.98* | 30.46* | 31.42* | 31.9* | 30.78* | 31.1* |
| KK | 11.78* | 27.76* | 27.44* | 23.56* | 21* | 27.68* | 27.6* | 25.32* | 21.6* | 18* |
| SS | 12.29* | 25.24* | 25.03* | 25.04* | 24.56* | 34.6* | 25.34* | 25.96* | 23.56* | 21.9* |
| DHT | 16.56* | 27.18* | 22.64* | 21.46* | 20.94* | 18.72* | 27.82* | 21.56* | 18.56* | 17.5* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

Generally, results of Figure 6.36 concludes that HD test with its convergence pattern and minimum value of the MSC at N=4 to N=32, excluding its MSC at N=2, is identified as better performing test. Similarly, HL test with very close to MSC of HD test at each level of cross section is identified as second performer better test among HLM, KK, SS, and DHT mediocre tests category. While, HLM test is categorized as bad performer test at each level of cross section unit in the mediocre class. Again, the minimum value of MSC of all stationarity tests at N=2 show low and close power of stationarity and point optimal tests at each alternative at T=10 (see Table A.11 (Panel A) in Appendix-A).

Figure 6.37 and Table 6.19 (Panel B) show the behavior of MSC of panel stationarity tests against small, medium and large cross section units at fixed time series level of 25 when data are generated in the presence of both of deterministic terms.

**Figure 6.37: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend Terms, T=25**



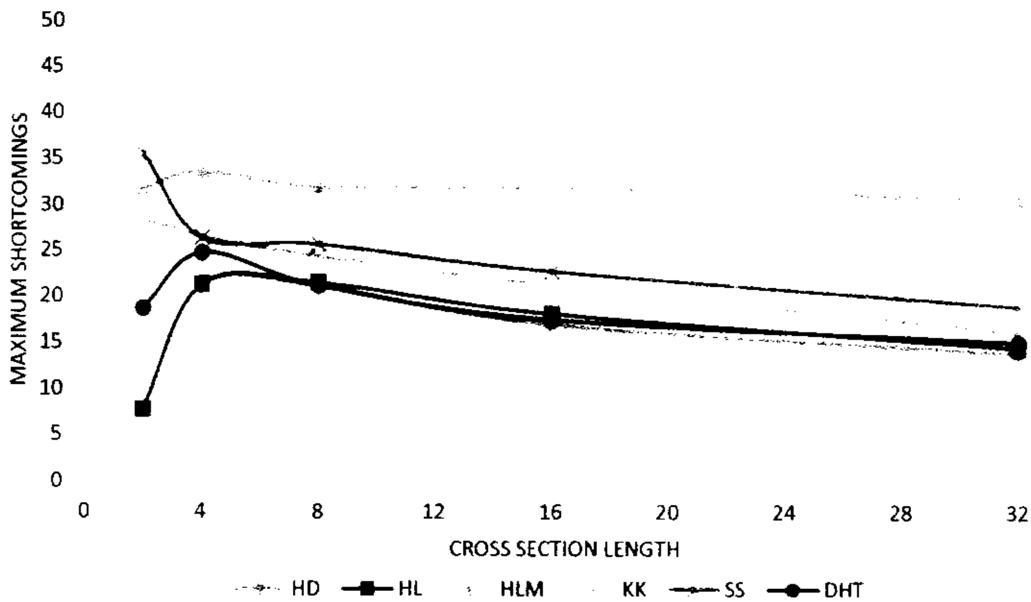

At N=2, Figure 6.37 shows that HD test is classified as best performer test with 9.96% MSC while HL test with 11.2% MSC is detected as second better performing test. While, HLM, KK, SS, and DHT tests having corresponding MSC of 30.46%, 27.68%, 34.6%, and 18.72% are graded into mediocre tests category in which SS test is diagnosed as bad performing test. As the number of cross section increases to 4 then HD test is unable to have a further decrease in its shortcomings and is ranked as mediocre test with MSC

21.56%. Similarly, HL and DHT tests with their increased MSC of 22.78% and 27.82% have managed its previous position. While, only SS test has decreasing MSC of 25.34% as compared to its previous MSC of 34.6%. However, HLM and KK tests with their constant behavior of MSC for N=4 and N=2 are again classified as mediocre tests.

Further, as the number of cross section increases to 8 then HD, HL, HLM, KK, and SS tests with approximately same MSC as have been observed at N=4 have kept their last positions in the mediocre tests category. However, DHT test has remarkably gained its power thus having minimum value of MSC at N=8 as compared to all other stationarity tests and has ranked as better performer test at that cross section unit. At number of cross section unit 16, all tests have gained a decreasing pattern according to their MSC and all of these tests remain in the mediocre test class. Out of these mediocre tests, DHT with 18.56% and HLM with 30.78% MSC are categorized as the better performing and bad performer tests in mediocre tests category. When the number of cross section units are 32 then HD test is classified as better performing test with MSC of 15.22% while HLM test with 31.1% MSC is graded as bad performing test in the mediocre tests category. While all other test with further decreasing pattern also remain in the same status.

Overall, Figure 6.37 demonstrates that HD, HL, KK, SS, and DHT tests have convergence pattern over cross section unit 4 to 32, excluding their MSC at N=2. However, HD test corresponding to its MSC at each level of cross section unit is identified as better performer test. Moreover, HL test with very parallel pattern of its MSC to HD test is also classified as the second better performing test as the number of cross increases from 2 to 32. A very similar picture is observed for DHT test but after N=4. Finally, due to its constant behavior regarding MSC HLM test is identified as bad performer test in the mediocre tests category.

**Figure 6.38: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend Terms, T=50**




A more detail review of PUR tests having the null hypothesis of stationarity has been analyzed in Figure 6.38 and Table 6.20 (Panel A) for T=50 over number of cross section unit 2, 4, 8, 16, and 32 when both drift and trend terms are present in the DGP and test equation. This figure shows a declining pattern for all most all tests over small to large cross section unit. At N=2, only HD and HL test have graded as best performing tests with MSC of 7.84% and 7.86%. While, HLM, KK, SS, and DHT tests corresponding to MSC of 31.82%, 28.46%, 35.54%, and 18.74% are ranked as mediocre tests. However, as the number of cross section increases from 2 to 4 then HD and HL tests lose their previous positions and have ranked as mediocre tests with MSC of 21.26% and 21.34%. While, all other four tests remain in the same category (i.e. mediocre tests) corresponding to their attained MSC. At N=4, HD and HLM tests with 21.26% and 33.66% MSC are classified as better and bad performer tests in the mediocre tests category.

Further, as the number of cross section moves to 8 then all stationarity tests continue to sustain their previous positions with little decreased in their MSC. Again, HD test with 21.14% and HLM test with 31.98% of MSC are labeled as better and bad performing tests in the mediocre test class. Similarly, with decreasing pattern all tests have kept their previous positions at N=16 with 17.02% minimum value of MSC of HD test and 31.86% maximum value of HLM test.

Finally, when the number of cross section increases to 32 then HD, HL, KK, SS, and DHT tests have attained their MSC less than 20%, excluding the bad performer test HLM with MSC 30.76%. Again, HD test with MSC of 13.96% is referred as better performing test in the class of mediocre tests.

Overall, Figure 6.38 indicates that HD test with convergence pattern after N=2 is identified as the most stringent test by attaining the minimum value of its MSC at each level of cross section unit when T=50. Similarly, HL, KK, SS, and DHT tests have also maintained their convergence pattern after N=2. Out of these mediocre performer tests, HL and DHT tests have very close pattern to pattern of HD test. However, HLM test is ranked as bad performer test in the mediocre test class at each level of cross section unit at T=50.

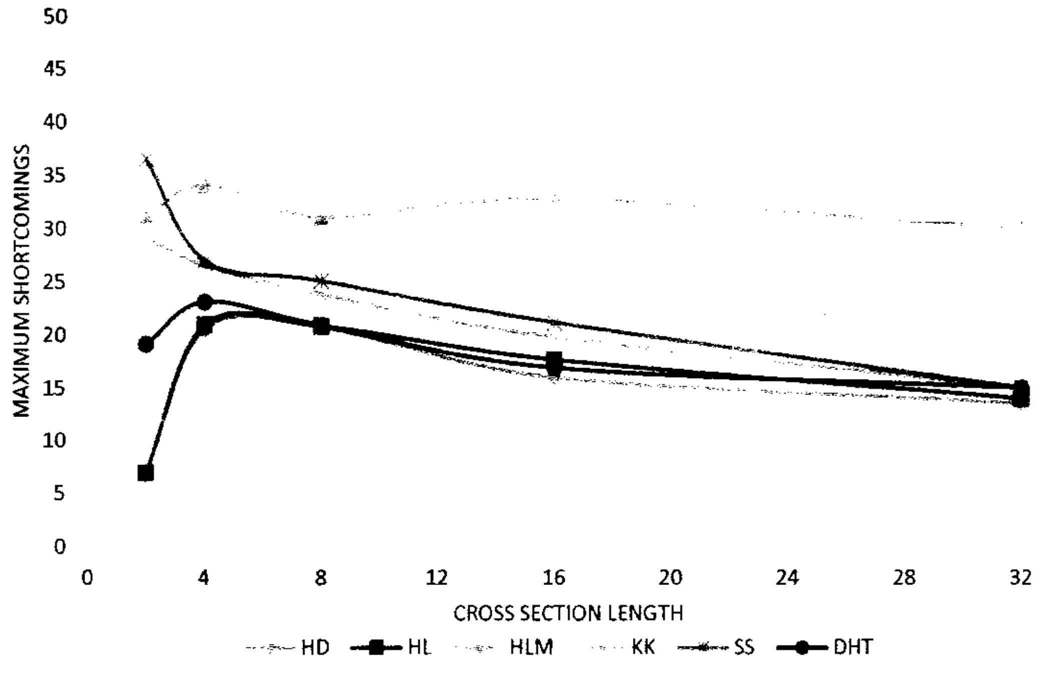
**Table 6.20: MSC for Null Hypothesis of Stationarity in the Presence of Intercept and Trend Case, at T=50 and T=100**

| | Panel A: T=50 | | | | | Panel B: T=100 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Tests/CL | 2 | 4 | 8 | 16 | 32 | 2 | 4 | 8 | 16 | 32 |
| HD | 7.84** | 21.26* | 21.14* | 17.02* | 13.96* | 6.96** | 20.56* | 20.86* | 16.06* | 13.62* |
| HL | 7.86** | 21.34* | 21.58* | 18.18* | 14.58* | 7** | 21.01* | 20.94* | 17.74* | 14.04* |
| HLM | 31.82* | 33.66* | 31.98* | 31.96* | 30.76* | 31.24* | 34.22* | 31.18* | 32.96* | 30.2* |
| KK | 28.46* | 26.5* | 24.42* | 21.3* | 16.2* | 29.22* | 26.58* | 23.88* | 19.88* | 15.06* |
| SS | 35.54* | 26.34* | 25.7* | 22.8* | 19* | 36.58* | 27.06* | 25.16* | 21.24* | 15.08* |
| DHT | 18.74* | 24.76* | 21.22* | 17.56* | 15.22* | 19.18* | 23.2* | 21* | 16.98* | 15.06* |

Note: "**" and "*" indicate Best and Mediocre Tests, respectively.

At very last time dimension of 100 of this study, Figure 6.39 and Table 6.20 (Panel B) display the performance of stationarity tests at varying size of cross sections. At N=2, two tests (i.e. HD and HL) have ranked as best performer tests corresponding to MSC of 6.96% and 7% while other four tests (HLM, KK, SS, and DHT) with respect MSC of 31.24%, 29.22%, 36.58%, and 19.18% are classified as mediocre performing tests. However, as the number of cross section increases to 4 then HD and HL tests switch their position into mediocre tests class with MSC 20.56% and 21.01%, respectively. Similarly, previously assigned mediocre tests at N=2 with upward and downward fluctuations have remained in the same came category with minimum value of shortcomings 23.2% against DHT test and maximum value of shortcomings of 34.22% against HLM test.

**Figure 6.39: MSC Assessments of Tests having Null Hypothesis of Stationarity with Intercept and Trend Terms, T=100**




Further, at N=8, all of the stationarity tests (i.e. HD, HL, HLM, KK, SS, and DHT) corresponding to MSC of 20.86%, 20.94%, 31.18%, 23.88%, 25.16%, and 21% respectively have gained a decreasing pattern at N=8 as compared to their previous values for N=4 but have kept their places in mediocre tests category. When the number of cross section units are 16 then a similar picture is detected as has been analyzed at N=8 for almost all tests in which HD and HLM tests with MSC of 16.06% and 32.96% are identified as better and bad performer tests among the mediocre tests class. Finally, at N=32, these tests have remained in the mediocre class with convergence behavior of their MSC. A very close value of MSC of HD, HL, KK, SS, and DHT tests have been observed at N=32, in which KK SS, and DHT tests with same values have an overlapping picture. Unfortunately, HLM test according to MSC of 30.2% is classified as bad performing test for N=32 as well.

Figure 6.39 indicates a convergence pattern for all six tests as the number of cross section increases from 2 with HD test having the minimum value of MSC at each level of cross section unit, hence, identified as most stringent test in the mediocre tests class. At small cross section units the value of HD and HL tests are small over other tests, however, as the number of cross section moves to medium and large cross section units these tests get their MSC value very close to HD and HL tests at T=100. Which shows that all these tests will eventually have zero MSC as both cross section dimension and time dimension gets larger. However, among mediocre tests category, HLM test remains the bad performing test having maximum value of MSC at each cross section level at T=100.

**Figure 6.40: MSC of Panel Stationary Tests with Drift and Trend**

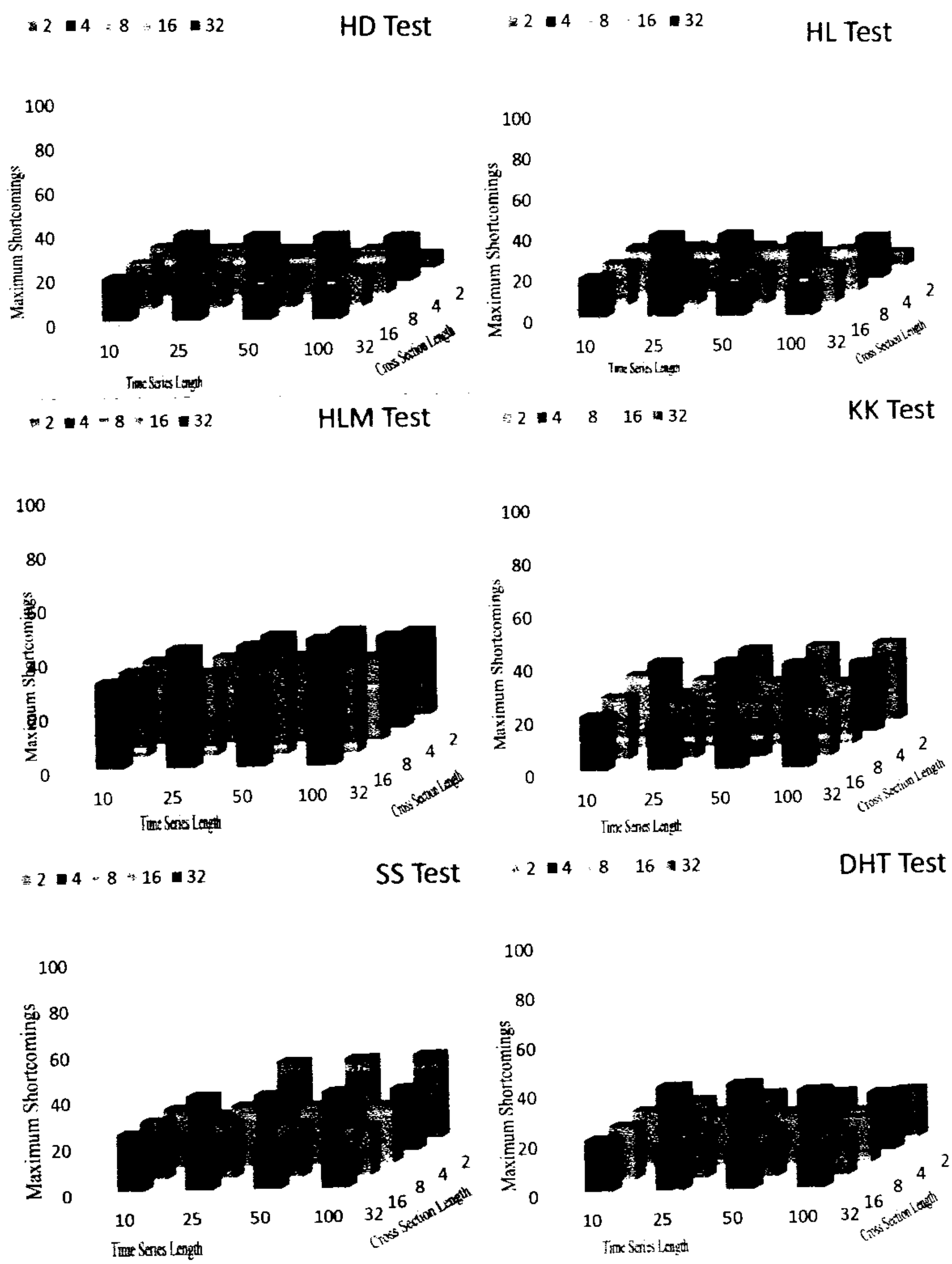


Figure 6.36 to Figure 6.39 conclude a very similar picture for the effect of cross section unit over time series that have been observed from Figure 6.16 to Figure 6.20 for all panel stationarity tests when data are generated with drift and trend terms. The effect of cross section unit over time series length shows a convergence behavior for all panel stationarity tests. Also, all these tests have been individually shown in Figure 6.40 corresponding to their MSC for each combination of cross section and time series dimensions. This figure indicates best performer tests (i.e. HD and HL) with their bars downward towards zero at each combination of time series and cross section units which further clarify their positions in this simulation study.

## 6.5. Concluding Remarks

It is evident from Figure 6.1 to Figure 6.5, Figure 6.25 and Table 6.1 to Table 6.3 that when data are generated with only an intercept term for PUR test with null hypothesis of unit root to observe the time series effect over cross section units then approximately all tests have minimal difference from power envelope as time series progress from smallest to largest for each fixed cross section unit. Also, the number of best performer tests increases as time series and cross section dimension grows and at larger time series for all cross sections majority of the tests become better performer according to their less and decreasing pattern of MSC. When cross section dimension is very small (i.e. N=2) then all tests have convergence pattern but none of them fulfill best test criteria corresponding to time series 10, however, for the same cross section but with varying time series level of more than 10 these PUR tests become in more decreasing pattern and become the part of best performing tests category. The number of worst and mediocre performer tests decreases while number of best performer tests increases as sample size and cross section units increases. At sample size 25 and 50, on average there are five (5) and eleven (11) best performer tests respectively for all cross section units as compared to 4 best performer tests for time series 10. Lastly at T=100, there are on average (i.e. N=8) fifteen (15) best performing tests and are the most number of best performing tests for that sample size as compared to previous sample sizes. As most of tests are performing well at sample size 50 and 100 so it is recommended to take those better tests which stands best in lower sample size. A standard benchmark of N=4 and N=8 against T=25 are taken to further tighten the best performer tests criteria and find the best performing tests in the overlapped picture shown in each graph. We have selected N=4 and N=8 corresponding to T=25 because these combination of cross section and time series are not too small and large in our finite sample comparative study. This concludes DWH, IPS, LLC, and WT tests as most suitable best performing tests for small and large sample sizes having MSC less than 10% at N=4 and N=8 against T=25 when data are generated with intercept but without trend, excluding PS test with worst performance. Overall, PS test with respect to its MSC is categorized as worst performing test at each combination of cross section and time series level.

It is also observed from Figure 6.6 to Figure 6.10, Figure 6.30 and Table 6.4 to Table 6.6 that when data are generated in the presence of both of deterministic terms for the null hypothesis of PUR tests to investigate the effect of time series on the performance of PUR tests then results are very similar as noticed in the case of with intercept but without trend model. Again, as the time series increases and cross section dimensions moves from small to large then a very large number of PUR tests having the null hypothesis of unit root are classified as best performing tests and the number of mediocre or worst performing tests remains either one or two. Hence, a benchmark of N=4 and N=8 against T=25 has been taken into account to find the most suitable best performing tests which can be used for small as well as for large samples. Hence, our analysis suggests DWH, IPS, LLC and WT tests as best performing tests for all combination of varying time series level and fixed cross section units according to assigned benchmark. There is only one test (i.e. PS) which performs very poorly according to its MSC for small to medium and then large panels. Therefore, besides this one worst performing test, all tests are preferable to be used at large time and cross section units according to simulation results but among these tests DWH, IPS, LLC, and WT tests are the most appropriate because it can be used for short and long sample sizes and cross section units.

In the third part of comparative study, tests having the null hypothesis of stationarity have been compared to categorize them as best, mediocre and worst performer tests to observe the effect of time series on the MSC performance of panel stationary tests. When data are generated with an intercept but without trend then it is observed from Figure 6.11 to Figure 6.15, Figure 6.35, Table 6.7 to Table 6.9 and Table A.1 to Table A.10 (Appendix-A) that overall HD and HL tests show an approximate declining phenomena for their MSC for all cross section units as the sample size varies from small to large. Both of these tests are ranked as better performing tests at sample size 10, 25, 50, and 100 when number of cross sections are 2. Moreover, HLM, KK, SS, and DHT tests with slowly divergence behavior, are the worst performing tests being in the category of mediocre tests according to their observed MSC as the sample size increases from small to large. Therefore, if the data are generated with intercept but without trend then HD and HL tests are most preferable tests over other four tests, as both of these tests have minimum discrepancy from power envelope for all cross section units as the sample size increases. Moreover, HD test is identified as most stringent test for most of the cross section and time series levels having minimum value of MSC as compared to other stationarity tests.

If the data are generated with intercept and trend for the null hypothesis of panel stationary then two of the stationary tests perform better as compared to other four tests as the sample size increases for different level of cross section units to analyze the effect of time series on the MSC performance. It is evident from Figure 6.16 to Figure 6.20, Figure 6.40, Table 6.10 to Table 6.12 and Table A.11 to Table A.20 (Appendix-A) that HD and HL tests with convergence pattern have less distance, especially at large sample size their distance, from power envelope for all cross section units and are preferable over other stationary tests when data are generated in the presence of both of deterministic terms. Also, three tests (KK, SS, and DHT) also show a decreasing pattern as the sample size gets larger but overall results indicate that DHT test is more close to HD and HL tests as the number of cross section and time series level increases. However, HLM test with a constant behavior of its MSC over all time series for each cross section unit is identified as bad performing test among the mediocre tests class.

This study has concluded the performance of both type (PUR and stationary tests) of PUR tests to observe the effect of time series over cross section unit on the performance of PUR tests having the null hypothesis of unit root and it is found that DWH, IPS, LLC, WT, HD, and HL tests are identified as best performing tests from the rest of other tests with respect to their MSC in both cases of deterministic parts. Out of these five best performing tests, first four belongs to null hypothesis of PUR while the last two belong to null hypothesis of panel stationarity.

A similar results have been observed to see the effect of cross section units on the performance of PUR tests having the null hypothesis of unit root in both cases of deterministic terms. In the presence of drift term only, PUR tests having the null hypothesis of unit root have evaluated through Figure 6.21 to Figure 6.25 and Table 6.13 to Table 6.14. Overall analysis of these figures have concluded that almost all PUR tests gets MSC equal to zero with overlapped pattern as the sample size increases from 25 to 100 and it is needed to find the best performing tests which performs well in small as well as in large time series and cross section combinations. Hence, a similar benchmark criteria has been adopted as has been taken previously to observe the effect of time series on the performance of PUR tests in the first section. Again, DWH, IPS, LLC, and WT tests are detected as the best performing tests at each cross section unit (i.e. from small to large) over time series level of 10, 25, 50, and 100 according to the assigned benchmark of N=4 and N=8 against T=25. It is observed that the number of mediocre and worst performing gets smaller as the time and cross section unit increases from small to large. Again, PS test is identified as worst performing test with MSC over 50% at all cross section level for fixed level of time series.

In the presence of both of deterministic terms (drift and trend terms) a similar situation has been concluded from Figure 6.26 to Figure 6.30 and Table 6.15 to Table 6.16 as has been observed when only drift term is present in the test equation and DGP to see the effect of cross section units on the performance of PUR tests having the null hypothesis of unit root. Overall, an overlapping pattern has been observed for almost all tests at large time series and cross section dimensions for best performing tests. Again, DWH, IPS, LLC, and WT tests are assigned as best performing tests according to their MSC behavior over small to large cross section units using the assigned benchmark criteria mentioned earlier. While, PS test is ranked as worst performing test at each level of varying cross section unit over fixed time series of 10, 25, 50, and 100.

Similarly, the effect of cross section units on the performance of PUR tests having the null hypothesis of stationary, from Figure 6.31 to Figure 6.35, Table 6.17 to Table 6.18 and Table A.1 to Table A.10 (Appendix-A) have the same remedy as have been concluded for the effect of time series dimension over level of cross section units. In the case of drift term for DGP and test equation under consideration, HD and HL tests with convergence pattern of their MSC over varying level of cross section units are detected as best performing tests among four (HLM, KK, SS, and DHT) tests. Between these two best performing tests, HD test with minimum value of its MSC over varying level of cross section units is identified as most stringent test. However, HLM test is assigned as bad performing test being in the category of mediocre tests.

In the presence of both of deterministic terms in the DGP and test equation, the effect of cross section units on the performance of PUR tests having the null hypothesis of stationary from Figure 6.36 to Figure 6.40, Table 6.19 to Table 6.20 and Table A.11 to Table A.20 (Appendix-A) show that HD and HL tests with convergence pattern are identified as best performing tests having minimum value of their MSC as compare to other four stationary tests. However, HD test with minimum value of its MSC at each level of cross section unit is detected as most stringent test. HLM test with its high MSC value at each level of cross section unit for time dimension is classified as bad performer test in the category of mediocre tests.

Generally, it is also analyzed that among both type of PUR tests (tests having null hypothesis of unit root and null hypothesis of stationary), power curve of tests having the null hypothesis of unit root are observed as more close to power envelope of point optimal test as compared to power curve of panel stationary tests to power envelope of point optimal test. Further, the results of this study concludes that at large sample sizes power curve of majority of PUR tests having null hypothesis of unit root overlap the power envelope of point optimal test, however, panel stationary tests do not have such a situation. Hence, PUR tests with null hypothesis of unit root are more preferable to apply then panel stationary tests. These results of poor performance of panel stationary tests are also parallel to the results of existing literature, Hlouskova and Wagner (2006), about panel stationary tests.

# Chapter 07

## Bootstrap Empirical Power Evaluation of Panel Unit Root Tests

In last chapter (i.e. chapter 06) we have studied the performance of PUR tests on the basis of size and power properties of these tests using simulated critical values and have concluded best, mediocre, and worst performing tests with respect to their assigned MSC. In this chapter we empirically evaluate the power behavior of best performing tests having the null hypothesis of unit root by using the Purchasing Power Parity (PPP) hypothesis. PPP hypothesis has been discussed in the literature by using time series and PUR tests, the idea is, if the real exchange rate is found to be unit root then this represent the violation of PPP hypothesis. If, it is found to be stationary then it is taken as strong evidence in favor of long run PPP. According to literature, some studies found evidence in favor of validity of PPP while other studies have failed to find the validity of PPP. According to existing literature the real exchange rate is constructed as follows:

$$\ln RER_{it} = \ln NER_{it} + \ln CPI_t^f - \ln CPI_{it}^d \qquad (7.1)$$

Here, $\ln NER_{it}$ shows the log of nominal exchange rate of domestic currency of country $i$ at time t against one unit of foreign currency (US dollar) and, $\ln CPI_t^f$ and $\ln CPI_{it}^d$ denote the log of foreign (US) and domestic consumer price index, respectively. Here, the best performer tests have been utilized on real exchange rate by using the bootstrap method. Hence, bootstrap critical values are calculated and used to carry out empirical performance of best performing tests.

This bootstrap simulation study is carried out using four different data sets corresponding to four cross section units from OECD countries and four time series dimensions. A monthly data is taken from January 2010 to April 2018 for Belgium, Canada, Finland, France, Germany, Ireland, Italy, the Netherland, Spain, Denmark, Iceland, Norway, Sweden, Switzerland, the United Kingdom, and the United States from IFS database. First data set include two (United Kingdom and United States) OECD countries corresponding to time series level of 10, 25, 50, and 100. Second data set takes four (France, Germany, the United Kingdom, and the United States) OECD countries with same level of time series. Third data set consists of eight countries including France, Germany, Italy, the Netherland, Spain, Denmark, the United Kingdom, and the United States with same time series levels. Similarly, fourth data set consist of sixteen countries including Belgium, Canada, Finland, France, Germany, Ireland, Italy, the Netherland, Spain, Denmark, Iceland, Norway, Sweden, Switzerland, the United Kingdom, and the United States for time dimension 10, 25, 50, and 100.

In the first section the bootstrap empirical power of best performing PUR tests have been analyzed when only intercept case is present in the DGP and test equation. In the second the bootstrap empirical powers have been explained in the presence of both of the deterministic terms in the DGP and test equation. In the last part of this chapter the summary of this chapter has been given.

## 7.1. Bootstrap Empirical Powers with Intercept Term, at N=2

Table 7.1 (Panel A) shows the bootstrap empirical power performance of four best performer PUR tests (i.e. DWH, IPS, LLC, and WT) for time series length of 10, 25, 50, and 100 when the number of cross sections are 2. At first data set, DWH test has bootstrap empirical power 67.12% at T=10 which is the highest as compared to empirical power of 59.98%, 47.96%, and 54.06% corresponding to IPS, LLC, and WT tests, respectively. As the time dimension increases to 25 then DWH, IPS, LLC, and WT tests have bootstrap empirical power 95.36%, 89.1%, 88.98%, and 54.06%, respectively. Here, DWH test again has maximum power as compared to other three PUR best performing tests while at the same time series level WT test has the least bootstrap empirical power of 67.78%. Moreover, at T=50 and 100, DWH, IPS, and LLC tests have achieved their maximum bootstrap empirical power of 100% while WT test has the least empirical power of 83.98% and 95%. These results are very similar to conclusion of Chapter 06, that is, as the time dimension increases the power of PUR tests also increases.

**Table 7.1: Bootstrap Power Evaluation with Intercept Case, at N=2 and N=4**

| Panel A: N=2 | | | | | Panel B: N=4 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TL/Tests | DWH | IPS | LLC | WT | TL/Tests | DWH | IPS | LLC | WT |
| 100 | 100 | 100 | 100 | 95 | 100 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 83.98 | 50 | 100 | 100 | 100 | 98.48 |
| 25 | 95.36 | 89.1 | 88.98 | 67.78 | 25 | 99.8 | 96.71 | 94.4 | 75.48 |
| 10 | 67.12 | 59.98 | 47.96 | 54.06 | 10 | 76.4 | 66.22 | 58.46 | 54.34 |

## 7.2. Bootstrap Empirical Powers with Intercept Term, at N=4

Panel B of Table 7.1 prsents a very similar results of best performing PUR tests that we have observed at N=2 for time series length of 10, 25, 50, and 100 when number of cross sections are 4. In the second data set, when the time length is 10 then DWH, IPS, LLC, and

WT tests have attained their bootstrap empirical power of 76.4%, 66.22%, 58.46%, and 54.34%, respectively. Clearly, at T=10, DWH and WT tests have the highest and smallest empirical power as compared to other two PUR best performing tests. As the time dimension increases from 10 to 25 then empirical power of all best performing PUR tests increases. DWH, IPS, LLC, and WT tests with 99.8%, 96.71%, 94.45%, and 75.48% empirical powers are ranked as 1st, 2nd, 3rd, and 4th tests, respectively. Even though, there is not much gape in between their attained empirical powers at N=25. As the time series level reaches to 50 and 100 then DWH, IPS, LLC, and WT tests have maximum power of 100%, although WT test has 98.48% bootstrap empirical power at T=50 which it is almost equal to 100%. Table 7.1 (Panel B) results are corresponding to results of simulation results of chapter 06 in which we have concluded that as the time and cross section dimension increases the power of PUR tests also increases.

## 7.3. Bootstrap Empirical Powers with Intercept Term, at N=8

Table 7.2 (Panel A) demonstrates the bootstrap empirical power of best performing PUR tests when the number of cross section units is 8 (data set 3rd) with varying time series length of 10, 25, 50, and 100. At T=10, it is observed that all tests have empirical power greater than 50%. DWH test with bootstrap empirical power of 82.78% has the highest empirical power while IPS and LLC tests with bootstrap empirical power of 78.84% and 72.9% respectively are the 2nd and 3rd number tests but not with much power difference. However, WT test has the least empirical power of 59.8% among all four best performing tests at N=10. Further, as the time series level is 25 then empirical power of all four best performing tests also increases. DWH, IPS, and LLC tests have almost 100% bootstrap empirical power while WT test has the least bootstrap empirical power of 78.04% and is the least one among all four best performing PUR tests. As the time dimension increases to 50 and 100 then almost all tests have their maximum power of 100% including WT test. Table 7.2 (Panel A) explains a similar conclusion as we have drawn from the results of simulations study of Chapter 06 which demonstrates highest power achievement of almost all tests as the number of cross section and time series lengths increases.

**Table 7.2: Bootstrap Power Evaluation with Intercept Case, at N=8 and N=16**

| Panel A: N=8 | | | | | Panel B: N=16 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TL/Tests | DWH | IPS | LLC | WT | TL/Tests | DWH | IPS | LLC | WT |
| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 98.84 | 50 | 100 | 100 | 100 | 100 |
| 25 | 100 | 100 | 99.9 | 78.04 | 25 | 100 | 100 | 100 | 99.34 |
| 10 | 82.78 | 78.84 | 72.9 | 59.8 | 10 | 100 | 100 | 100 | 73 |

## 7.4. Bootstrap Empirical Powers with Intercept Term, at N=16

When the number of cross section unit is 16 (4th data set) then Table 7.2 (Panel B) explains that almost all tests have attained maximum bootstrap empirical power of 100% as the time dimension increases from 10 to 25 and then 100. Even though, WT test has 73% bootstrap empirical power at T=10 but as the time series level increases to 25 then this test attains its maximum bootstrap empirical power of almost 100% along with DWH, IPS, and LLC tests. Table 7.2 (Panel B) demonstrates that at very large cross section unit with each level of time series length the empirical power of all tests reaches to their maximum power of 100%.

Overall, Table 7.1 and Table 7.2 indicates that as the number of cross section and time series length increases then bootstrap empirical power of best performing tests also increases and at last reaches to their maximum attained bootstrap empirical power of 100% in the specification of only intercept term in the specified test's model. These results are very similar to the results of simulation results of Chapter 06.

## 7.5. Bootstrap Empirical Powers with Intercept and Trend Terms, at N=2

Table 7.3 (Panel A) analyzes the bootstrap empirical power performance of best performing PUR tests when the number of cross section are two (first data set) in the presence of specification of intercept and trend terms both in DGP and test equation of the model. At T=10, it is observed that all PUR tests have their bootstrap empirical power greater than 50% with very little difference between these empirical powers. At small time series the empirical powers 74.38%, 65.3%, 59.78%, and 54.1% have been attained by DWH, IPS, LLC, and WT tests, respectively. As the time dimension increases to 25 then DWH and IPS tests attained maximum empirical power of 100% while LLC with 94.76% (almost 100%) empirical power are more near to empirical power of DWH and LLC tests. However, WT has taken 71.8% empirical power at T=25 and is the least power among four best performer tests. At T=50 and 100, DWH, IPS, and LLC tests have empirical power equal to 100%. While WT test has bootstrap empirical power of 88% and 100% at T=50 and T=100, respectively. These results are in favor of our simulation study results that we have carried out in Chapter 06 which shows that power of PUR tests increases as the time series length increases.

**Table 7.3: Bootstrap Power Evaluation with Intercept and Trend Case, at N=2 and N=4**

| Panel A: N=2 | | | | | Panel B: N=4 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TL/Tests | DWH | IPS | LLC | WT | TL/Tests | DWH | IPS | LLC | WT |
| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 88 | 50 | 100 | 100 | 100 | 99.48 |
| 25 | 100 | 100 | 94.76 | 71.8 | 25 | 100 | 100 | 100 | 77.48 |
| 10 | 74.38 | 65.3 | 59.78 | 54.1 | 10 | 94.38 | 88 | 74.4 | 58.34 |

## 7.6. Bootstrap Empirical Powers with Intercept and Trend Terms, at N=4

A very similar picture has been displayed for almost all tests in Table 7.3 (Panel B) when the number of cross sections is four (second data set). At small time series (T=10), DWH, IPS, LLC, and WT tests have taken bootstrap empirical power of 94.38%, 88%, 74.4%, and 58.34%, respectively. However, as the time dimension increases to 25 then DWH, IPS, and LLC tests have attained their maximum bootstrap empirical power of 100% while WT test has taken bootstrap empirical power equal to 77.48%. Further, at T=50 and T=100, almost all test have attained their highest bootstrap empirical power equal to 100% which has been also observed from simulations results in Chapter 06 that as the time series and cross section length increases the power of PUR tests also increases.

## 7.7. Bootstrap Empirical Powers with Intercept and Trend Terms, at N=8

Table 7.4 (Panel A) investigates the bootstrap empirical power when the number of cross sections are eight (third data set). At T=10, DWH test has the highest empirical power of

100% while LLC, IPS, and WT tests have attained their empirical power of 98.2%, 93.86%, and 72.8%, respectively. As the time length increases to 25, then almost all tests have taken their empirical power equal to 100% except WT test with 98% empirical power almost equal to 100%. At T=50 and T=100, all best performing tests have bootstrap empirical power equal to 100%. Table 7.4 (Panel A) concludes that as the time and cross section length increases then bootstrap empirical power of PUR tests also increases, these results are very similar to our simulation results of Chapter 06.

**Table 7.4: Bootstrap Power Evaluation with Intercept and Trend Case, at N=8 and N=16**

| Panel A: N=8 | | | | | Panel B: N=16 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TL/Tests | DWH | IPS | LLC | WT | TL/Tests | DWH | IPS | LLC | WT |
| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 100 | 50 | 100 | 100 | 100 | 100 |
| 25 | 100 | 100 | 100 | 98 | 25 | 100 | 100 | 100 | 100 |
| 10 | 100 | 98.2 | 93.86 | 72.8 | 10 | 100 | 100 | 100 | 79.01 |

## 7.8. Bootstrap Empirical Powers with Intercept and Trend Terms, at N=16

At very large cross section unit of sixteen countries (fourth data set) in this study, Table 7.4 (Panel B) demonstrates bootstrap empirical power of best performing PUR tests when DGP and test equation have both intercept and trend terms. At small time series length (T=10), only WT test has empirical power (i.e. 79.01%) less than 100% while all other tests (DWH, IPS, and LLC) have bootstrap empirical power equal to 100%. At time series length 25, 50, and 100, all best performing PUR tests have attained their bootstrap empirical power of 100%. These results, at large cross section and time series length, are very similar to our simulation results of Chapter 06 that as the number of cross section and time length increases the power of PUR tests also increases

Table 7.3 and Table 7.4 show that as the number of cross section and time length increases then bootstrap empirical power of tests also increases when data are generated with intercept and trend terms.

## 7.9. Concluding Remarks

In this chapter, bootstrap empirical power evaluation of four best performing PUR tests (DWH, IPS, LLC, and WT) have been carried out using four type of data sets under the specification of both type of the deterministic terms. The empirical power evaluation of these tests have been demonstrated using purchasing power parity hypothesis by utilizing bootstrap method. In the case of intercept term only, Table 7.1 and Table 7.2 conclude that DWH test has least empirical power of 67.1% at T=10 and at high time series level this empirical power is observed as 100% when the number of cross section units are 2.

Similarly, IPS and LLC tests have the same phenomenon with least bootstrap empirical power of 60% and 47.96% when T=10 and highest empirical power of 100% when T=100 at N=2, respectively. However, WT test with the least empirical power of 54.06% at T=10 and highest empirical power of 95% at T=100 has the least empirical power at every time series length as compared to other three tests when N=2.

As the number of cross section units increases then each PUR test gains much power at each stage of increasing time series level. At large cross section unit (N=16) of this study, DWH, IPS, and LLC tests have gained 100% empirical power even at small time series level of 10. However, WT test has 67% empirical power at T=10 but as the time series level increases to 50 and 100 then power of this test statistic is observed 100% at N=16. These results present similar picture as has been observed in Chapter 06 of our simulation study which explains 100% power for almost all tests as the number of time series and cross section units increases if DGP and test equation has intercept term only.

Similarly, Table 7.3 and Table 7.4 indicate empirical power evaluation of PUR tests in the presence of both of the deterministic term in the DGP and test equation. These results display the same picture as have been observed in case of intercept term only. At small time series (T=10) and cross section unit (N=2), DWH, IPS, LLC, and WT tests have the least empirical power 74.38%, 61.3%, 49.78%, and 54.1%, respectively. However, as the time dimension increase to 50 and 100 then empirical power of DWH, IPS, and LLC tests have been observed as 100% while WT test is observed with bootstrap empirical power of 88% at T=50 and 100% at T=100.

At very large cross section all these PUR tests achieve maximum bootstrap empirical power of 100% from small to large time series level. All these results are parallel to results of simulation study of Chapter 06 that as the number of cross section and time series length increases the power of PUR tests also increases in the presence of both of the deterministic terms.

# Chapter 08

## Conclusions and Recommendations

This chapter summarizes and discusses the overall results of the study for the both (i.e. tests having the null hypothesis of PUR and null hypothesis of stationary) categories of PUR tests to mention some recommendation for the future research in the field of econometrics.

### 8.1. Key Conclusions

In Chapter 04, empirical size of twenty four tests have investigated using asymptotic critical value for different combination of time series and cross section units. Out of these twenty-four tests, eighteen belongs (i.e. MP, PS, BNG, CH, CHO, LLC, BU, IPS, MW, CIPS, DWH, RMA, LW, DH, CIPS_star, WT, OS, and WL) to the null hypothesis of unit root and the remaining tests belong to the null hypothesis of stationary. It is concluded from PUR tests having the null hypothesis of unit root that sixteen tests have unstable size at small time dimension while at very large time dimension majority of these tests become stable. Only two tests (MW and CHO), based on p-value, have stable size around nominal size of 5% at all combination of cross section and time series length.

Similarly, all tests (HD, HL, HLM, KK, SS, and DHT) having the null hypothesis of stationary have unstable and convergence pattern of their empirical size as the time dimension increases from small to large with each level of cross section unit. At small, medium, and large time series dimensions, four of the tests have unstable size from nominal size of 5% while only HD and HL tests become stable at large time series and cross section units. Overall, an unstable empirical size has been observed for almost all tests when asymptotic critical value is used to calculate the size of the tests.

In order to get a stable size, in Chapter 05 size of twenty-four PUR tests (i.e. tests having the null hypothesis of unit root and null hypothesis of stationary) have been assessed by using simulated critical value. In which eighteen tests (i.e. MP, PS, BNG, CH, CHO, LLC, BU, IPS, MW, CIPS_star, CIPS, DWH, RMA, LW, DH, WT, OS, and WL) belong to null hypothesis of PUR while six tests (i.e. HD, HL, HLM, KK, SS, and DHT) are from the category of null hypothesis of stationary. A time dimension of 10, 25, 50, and 100 for the cross section level of 2, 4, 8, 16, and 32 have used to evaluate the size performance of both type of tests for two different situations.

In first situation all PUR tests have analyzed by using DGP only with intercept term, and in the second situation DGP with intercept and trend terms have been utilized to examined size stability for all possible combination of time series and cross section level. First, the size of eighteen PUR tests having the null hypothesis of unit root has been analyzed with intercept term only. It is concluded that size of all eighteen tests are stable around nominal size of 5% when simulated critical value is used. This stabilized size varies from 4% to 6% for almost all tests either time series level is small, medium or large for all cross section units. Second, tests having the null hypothesis of PUR have analyzed for small to large sample sizes corresponding to each cross section level when data are generated in the occurrence of both of the deterministic terms. It is concluded that all tests have a stable size approximately equal to nominal size of 5% for all combination of time series and cross section units.

Similarly, six PUR tests having the null hypothesis of stationary are examined and the same results are obtained as described for the PUR tests having the null hypothesis of unit root. In the first section, size of stationary tests has been analyzed when data are generated in the presence of intercept term only. It is concluded that all six stationary tests for different level of time series (i.e. small, medium and large) and cross section units (small, medium and large) have an approximate stabilized size equal to nominal size of 5%.

Similarly, the results of stationary tests when data are generated both with intercept and trend terms have the same remedy as have been explained with drift term only. At sample size 10, 25, 50, and 100 for different cross section units 2, 4, 8, 16, and 32, we have observed a stabilized size which varies from 4% to 6% and are approximately equal to nominal size of 5%. Overall, these results conclude that a stabilized size equal to nominal size of 5% corresponding to PUR and panel stationary tests to avoid from the problem of over and under rejection which do exist if asymptotic critical value is used. In short, we have seen that size of both type of PUR tests have been controlled by using simulated critical value.

Chapter 06 of this study has analyzed Monte Carlo power comparison of both type of PUR tests under the newly introduced criterion i.e. stringency criterion discussed by Zaman (1996). According to stringency criterion a stabilized size is needed, which is achieved in Chapter 05, to compare PUR tests for the null hypothesis of PUR and null hypothesis of stationary, a Neyman Perason (NP) test for both type of tests have been discussed. Power comparison for both type of tests has carried out for cross section units 2, 4, 8, 16, and 32 with different time dimension 10, 25, 50, and 100 using DGPs with intercept term only and with both intercept and trend terms. At first, we have made a standard to categorize tests according to their MSC performance. A test is named as best performing test if it has MSC in between 0% to 10%. Similarly, test having MSC greater than 10% but less than or equal to 50% are assigned as mediocre performing test. Lastly, a test is known to be a worst performing test if its MSC are greater than 50%.

A time series effect over the cross section units of PUR tests having the null hypothesis of unit root have been analyzed with respect to MSC when data are generated with only intercept term for N=2, 4, 8, 16, and 32 with time series T=10, 25, 50, and 100. When data are generated with only intercept term for PUR test with the null hypothesis of unit root then approximately all tests have minimal difference from power envelope at large time series and cross section unit. When cross section dimension is very small (i.e. N=2 and N=4) then all tests have convergence pattern as the time series approaches from 10 to 100, excluding PS test is identified as the worst performing test for all time and cross section dimensions. As the time series and cross section length become large, the number of worst performer and mediocre performer tests decreases while the number of best performer tests increases.

The findings of this study indicate a lot number of best performer PUR tests (i.e. seventeen) at large time series (T=100) and cross section (N=32) dimensions, hence there is need to select such a cross section and time series combination which is most favorable to detect a few best performer tests having MSC less than 10%. So, a time series length of 25 and cross section length of 4 and 8 have been kept as standard to further tight the criteria of best performer test and picked the most favorable best performing tests from over all possible combination of time series and cross section units. Also, the combination of N=4, 8 and T=25 are not very small and very large in our study, since this study is based on finite sample sizes, to evaluate the performance of panel tests in the best tests class. According to obtained results, there are four tests which satisfy this benchmark criteria, these tests are; DWH, IPS, LLC, and WT tests if the effect of time series over cross section units is evaluated. In other words, if the number of observations are 50 or 100 then these four tests are better to use as compared to other PUR tests in the case of deterministic drift term only. Most importantly, PS test is identified as worst performer test at all level of time series and cross section combinations.

When data are generated in the presence of both of deterministic terms for the null hypothesis of PUR tests to see the effect of time series length over cross section units then results are very similar as noticed in the case of with intercept but without trend model. As sample size increases from 10 to 100 then the number of best performer tests also increases while the number of mediocre and worst performer tests decreases at each level of cross section units. At large number of time series (T=100) and cross section (N=32) lengths the number of best performer tests reaches to sixteen, hence a benchmark of T=25 against N=4 and N=8 have been selected to further find most favorable best performer tests. Overall analysis according to assigned benchmark suggests DWH, IPS, LLC, and WT tests as best performing tests for all combination of varying time series level and fixed cross section units. There is only one test (PS) which performs poorly according to its MSC for all level of time series and cross section dimensions.

A similar results have obtained over the cross section effect on time series levels in the cases when both of deterministic terms are taken into account corresponding to PUR test having the null hypothesis of unit root. In the presence of drift term only, PUR tests having the null hypothesis of unit root have been evaluated and it is found that the number of best performing tests increases while the number of mediocre and worst performer tests decreases as the cross section lengths varies from small to large for the large time dimension of 100. At large combination of cross section and time series dimension seventeen tests (i.e. BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, MP, MW, CHO, WL, LW, WT, OS, CIPS_star, and RMA) have categorized as best performer tests according to their MSC of less than 10%. A benchmark of N=4 against T=25 and N=8 against T=25 have been taken to further tighten the best tests criteria and find the most favorable best performing tests. This benchmark is taken from moderate number of cross section and time series units which are not very small and not too large. According to this criteria, DWH, IPS, LLC, and WT tests have been identified as the best performing tests over other best performing tests at small, medium and large cross section as well as time series units. However, PS test with MSC over 50% has been ranked as worst performing test at each level of cross section unit for all time series length.

In the presence of both of deterministic terms (i.e. drift and trend terms) for PUR tests having the null hypothesis of unit root a similar situation has been concluded as has been observed when only drift term is present in the test equation and DGP to observe the effect of cross section units over fixed level of time series. It is observed that as the number of cross section and time series units vary from small to large then the number of worst performer and mediocre performer tests shrink while the number of best performer tests increases. These findings show a lot of number of PUR tests lying in the category of best performing tests as the time series and cross section length reaches to 100 and 32 respectively the number of best performing tests at that combination are sixteen (i.e. BNG, BU, CH, DH, DWH, IPS, CIPS, LLC, MP, MW, CHO, WL, LW, WT, OS, and CIPS_star).

Again, a benchmark of N=4 against T=25 and N=8 against T=25 has been taken to find the best performer tests which are most favorable at small sample sizes as well. Our results have concluded that DWH, IPS, LLC, and WT tests are assigned as best performing tests according to their MSC behavior over the assigned benchmark combination of N=4 and N=8 corresponding to T=25. However, our analysis also indicate PS test as worst performing test at each level of cross section and time series length.

These results of tests having the null hypothesis of PUR have revealed that effect of time series and effect of cross section units over PUR tests have the same results in both cases of deterministic terms (i.e. with drift term only, and with drift and trend terms). In both cases, DWH, IPS, LLC, and WT tests are identified as best performing tests according to assigned benchmark criteria of T=25 against N=4 and N=8 among best performing tests. Also, the results show that majority (i.e. excluding one or two worst performer tests) of the PUR tests having the null hypothesis of unit root are ranked as best performing tests at large time series (T=100) and cross section length (N=32). However, at all possible combinations PS test is ordered as worst performing test with almost no improvement in its pattern.

This study has also evaluated the performance of PUR tests having the null hypothesis of stationarity with respect to MSC for different combination of time series and cross section units. In this part of comparative study the effect of time series has been evaluated for the tests having the null hypothesis of stationarity to categorize best, mediocre and worst performer tests for different combinations of time series and cross section units. When data are generated with an intercept but without a trend then it is observed that only HD and HL tests show an approximate declining phenomena for their MSC for all cross section units as the sample size varies from small to large. Both of these tests are ranked as better performing tests at sample size 10, 25, 50, and 100 when number of cross sections are 2. Even though, HD and HL tests are ranked as best performing test corresponding to N=2, however, both of these have classified as better mediocre tests in the mediocre tests category with respect to their MSC. Moreover, HLM, KK, SS, and DHT tests with slowly divergence behavior, are the poor performing tests being in the category of mediocre tests according to their observed MSC as the sample size increases from small to large. Therefore, if the data are generated with intercept but without trend then HD and HL tests are most preferable tests over other four tests, as both of these tests have minimum discrepancy from power envelope for all cross section units as the time dimension increases. Also, HD test is identified as most stringent test for most of the cross section and time series levels having minimum value of MSC as compared to other stationarity tests.

If the data are generated both with intercept and trend terms for the PUR tests having the null hypothesis of stationary to observe the effect of time series then two (i.e. HD and HL) of the stationary tests perform better as compared to other four (i.e. HLM, KK, SS, and DHT) tests as the sample size increases for different level of cross section units. It is also observed that HD and HL tests with convergence pattern have less distance, especially at large sample size, from power envelope for all cross section units and are preferable over other stationary tests when data are generated in the presence of both of deterministic terms. Also, three tests (KK, SS, and DHT) show a decreasing pattern as the sample size gets larger but overall results indicate that DHT test is more close to HD and HL tests as the number of cross section and time series level increases. However, HLM test with a constant behavior of its MSC over all time series corresponding to each cross section unit is identified as bad performing test among the mediocre tests class. Among the tests having the null hypothesis of stationary no single test has identified as worst performing test at all combination of time series and cross section units. Also, no single test has fulfilled best test criteria corresponding to MSC performance excluding HD and HL tests only at N=2.

Similarly, the effect of cross section units over fixed level of time series for PUR tests having the null hypothesis of stationary have the same remedy as has been concluded in the presence of the effect of time series dimension for panel stationary tests. When DGP and test equation both have drift term only then HD and HL tests with convergence pattern of their MSC over varying level of cross section units are detected as better performing tests among four (HLM, KK, SS, and DHT) other stationary tests. Out of these two best performers, HD test with minimum value of its MSC over varying level of cross section units is identified as most stringent test. Also, as the number of cross section increases then the pattern of KK, SS, and DHT tests gain convergence pattern and at very large number of cross section and time series length these four tests will eventually lie in the best tests category beside HD and HL tests having their MSC very close to zero. However, HLM test with approximately constant behavior of its MSC over cross section units is assigned as bad performing test being in the category of mediocre tests.

When both of the deterministic terms (i.e. drift and trend) are taken into account in the DGP and test equation, the effect of cross section units on the performance of PUR tests having the null hypothesis of stationary with respect to MSC concludes that HD and HL tests with convergence pattern are identified as best performing tests in the category of mediocre tests having minimum value of their MSC as compare to other four mediocre stationary tests (HLM, SS, KK, and DHT). The MSC of HD and HL tests at N=2 for all time series level are less than 10% and are identified as best performing tests at that cross section unit for each level of time series but as the number of cross section increases these tests lose their position in the best performing test class and have identified as mediocre performer tests with each level of time series. Moreover, these tests are ranked as better performing tests being in the class of mediocre tests among HLM, SS, KK, and DHT tests. However, HD test with minimum value of its MSC at each level of cross section unit is detected as most stringent test. A convergence pattern has been observed for KK, SS, and DHT tests as the number of cross section and time series dimension increases. While, HLM test with its high MSC value at each level of cross section unit over the time series level is classified as bad performer test in the category of mediocre tests.

These results for the PUR tests having the null hypothesis of stationary conclude that the effect of time series and cross section units over performance of panel stationarity tests are same with respect to MSC at each combination of time series and cross section units in both of the deterministic cases. These findings show HD and HL tests as the better performing tests with convergence pattern corresponding of their MSC if the effect of time series or cross section units are taken into account, however, HD test with minimum value of its MSC at each combination of cross section and time series unit is identified as most stringent test. While, HLM test with its position in mediocre tests category has remained the bad performer test if the effect of time series or cross section units are analyzed.

Overall results of both type of tests (i.e. PUR and stationary tests) indicate that the effect of time series and cross section units over the MSC performance are same in both cases of deterministic terms. These findings have suggested DWH, IPS, LLC, WT, HD, and HL tests as best performing tests from the rest of other tests with respect to their MSC in the presence of both cases of deterministic terms if the effect of either time series or cross section units are taken into account. Out of these six best performing tests, first four belongs to null hypothesis of PUR while the last two belong to null hypothesis of panel stationarity. More importantly, due to high power of PUR tests having the null hypothesis of unit root over panel stationarity tests of this study findings justify the results of existing literature about the poor performance of panel stationarity tests. Hence, PUR tests having the null hypothesis of unit root are more preferable to use then panel stationarity tests in real world applications.

An empirical evaluation of best performing tests having null hypothesis of unit root have been discussed using constructed real exchange rate data from nominal exchange rate and consumer price index of the domestic country and foreign country data by applying bootstrap method in the last chapter of this study. A monthly data is taken from January 2010 to April 2018 for Belgium, Finland, France, Germany, Ireland, Italy, the Netherland, Spain, Denmark, Iceland, Norway, Sweden, Switzerland, the United Kingdom, and the United States from IFS database. It is observed that all best performing tests have increasing pattern of their empirical power as the time series level progress from small (T=10) to large (T=100) corresponding to each cross section unit (i.e. N=2, 4, 8, and 16). At small time series and cross section level all tests have empirical power greater than 50% but as the time dimension increases to 100 then three of the tests achieves 100 percent power. However, as the number of cross section reaches to 16 then at T=50 and T=100 almost all test have 100% empirical power whether data are generated with intercept term only or both intercept and trend terms. These empirical results justify our results of simulation study of Chapter 06.

## 8.2. Recommendations

We have observed from the findings of this study that simulated critical value gets a stabilized size for all PUR tests either number of cross sections and time series are small, medium or large. We investigate a stabilized size when data are generated in the presence of only intercept term and then in the occurrence of both of the intercept and trend terms, and we observed that in both situations size of all PUR tests (i.e. tests having the null hypothesis of unit root and null hypothesis of stationary) are stable around nominal size of 5%. If a researcher uses asymptotic critical value instead of simulated critical value then a greater number of size distortions can be observed for majority of PUR tests. So, we recommend avoiding over rejection problem for both type of tests (i.e. tests having the null hypothesis of unit root and null hypothesis of stationary), use simulated critical value rather than to use asymptotic critical value for any combination of cross section and time series levels.

From power evaluation point of view, our simulation study recommends four PUR tests having the null hypothesis of unit root to be used in applied research for different level of time series and cross section combinations in the presence of both of the specification of deterministic terms. These tests are DWH, IPS, LLC, and WT having MSC less than 10% whether number of cross section and time series is small or large. In simple words, if the number of observations is 50 or 100 then these four tests are better to apply as compare to other tests. However, if the number of observation are greater than 100 then almost all tests are recommend to apply, excluding the PS test with its worst performance.

Likewise, in the PUR tests having the null hypothesis of stationary, HD and HL tests are recommended to practice if time series and cross section dimensions are small, medium or large in both cases of specification of deterministic terms. Also, our findings recommend HLM test as the worst performing test and is avoided to use in practice.

Overall, it is also concluded from the results that PUR tests having the null hypothesis of unit root performed better as compared to PUR tests having the null hypothesis of stationary. Hence, it is recommended that tests having null hypothesis of unit root may be given priority over tests having null hypothesis of stationary.

## 8.3. Directions for Future Research

In this comparative simulation study, we have studied the performance of PUR and stationary tests using two different DGPs for each category of tests to find the most stringent tests which needs point optimal tests under stringency criterion. For future research study, a single DGP can be used to develop a point optimal test for both categories of tests and, to evaluate the size and power performance of PUR tests under stringency criterion. We have taken simple DGP with heterogeneous panel and with no cross sectional dependence situation, in future a research can be conducted to compare PUR tests using a cross sectional dependence DGP.

Also, in our study, we have observed that at small time series and cross section dimensions there is a huge gap between point optimal tests and PUR tests. So, there is a need to develop a PUR test having appropriate maximum power at small time series and cross section length to minimize the gap.

# References

Bai, J., and Ng, S. (2004). "A PANIC Attack on Unit Roots and Cointegration." Econometrica **72**: 1127-1177.

Barreira, A. P., and Rodrigues, P. M. (2005). "Unit Root Tests for Panel Data: A Survey and an Application." Estudos II, 665-685.

Breitung, J. (1999). "The Local Power of Some Unit Root Tests for Panel Data", Discussion Papers, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.

Chang, Y. (2002). "Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency." Journal of Econometrics **110**(2): 261-292.

Choi, I. (2001). "Unit Root Tests for Panel Data." Journal of International Money and Finance **20**(2): 249-272.

Choi, I. (2006). "Combination Unit Root Tests for Cross-Sectionally Correlated Panels." Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter CB Phillips. Cambridge University Press, Chapt **11**: 311-333.

De Wachter, S., Harris, R.D. and Tzavalis, E., (2007). "Panel Data Unit Roots Tests: The Role of Serial Correlation and The Time Dimension." Journal of Statistical Planning and Inference **137**(1): 230-244.

Demetrescu, M., and Hanck, C. (2012). "A Simple Nonstationary-Volatility Robust Panel Unit Root Test." Economics Letters **117**(1): 10-13.

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.

Dickey, D.A. and Fuller, W.A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," Journal of the American Statistical Association, 74, 427–431.

Dickey, D. A. and Fuller, W. A. (1981). "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica, 49, 1057–1072.

Dickey, D.A. and Pantula, S.G. (1987). "Determining the Order of Differencing in Autoregressive Processes," Journal of Business and Economic Statistics 15, 455–461.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). "Efficient Tests for an Autoregressive Unit Root," Econometrica, 64, 813–836.

Fisher, R. A. (1934). "Statistical Methods for Research Workers." Oliver & Boyd, Edinburgh, 4th Edition.

Gengenbach, C., Palm, F.C., and Urbain, J-P. (2009). "Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling." Econometric Reviews **29**(2): 111-145.

Gutierrez, L. (2006). "Panel Unit-root Tests for Cross-sectionally Correlated Panels: A Monte Carlo Comparison." Oxford Bulletin of Economics and Statistics **68**(4): 519-540.

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.

Hoang, N. T., and McNown, R.F. (2006). "Panel Data Unit Roots Tests using various Estimation Methods." University of Colorado Bulletin, **6**, 33-66.

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

Im, K. S., Pesaran, M.H., and Shin, Y. (2003). "Testing for Unit Roots in Heterogeneous Panels." Journal of Econometrics **115**(1): 53-74.

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.

Lee, C., and Wu, J-L. (2012). "A Panel Unit-Root Test with Smooth Breaks and Cross-Sectional Dependence." Working Paper, Institute of Economics, National Sun Yat-sen University, Kaohsiung, Taiwan.

Levin, A., and Lin, C.F. (1993). "Unit Root Tests in Panel Data: New Results." UC San Diego Working Paper 93-56.

Levin, A., Lin, C.F., and Chu, C-S.J. (2002). "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties." Journal of Econometrics **108**(1): 1-24.

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.

Moon, H. R., and Perron, B. (2004). "Testing for a Unit Root in Panels with Dynamic Factors." Journal of Econometrics **122**(1): 81-126.

Moon, H. R., Perron, B., and Phillips, P.C.B. (2007). "Incidental Trends and the Power of Panel Unit Root Tests." Journal of Econometrics **141**(2): 416-459.

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

Ng, S. and Perron, P. (2001). "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," Econometrica, 69(6), 1519-1554.

Oh, Y. J., and So, B.S. (2004). "Robust Tests for Unit Roots in Heterogeneous Panels." Economics Letters **84**(1): 35-41.

Pesaran, M. H. (2007). "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence." Journal of Applied Econometrics **22**(2): 265-312.

Phillips, P.C.B. and Perron, P. (1988). "Testing for a Unit Root in Time Series Regression," Biometrika, 75, 335–346.

Phillips, P. C. B., and Sul, D. (2003). "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence." The Econometrics Journal **6**(1): 217-259.

Stock, J. H. and M. W. Watson (1988). "Testing for common trends." Journal of the American Statistical Association **83**(404): 1097-1107.

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

Sul, D. (2009). "Panel Unit Root Tests under Cross-Section Dependence with Recursive Mean Adjustment." Economics Letters **105**(1): 123-126.

Westerlund, J. (2014). "Heteroscedasticity Robust Panel Unit Root Tests." Journal of Busines & Economic Statistics **32**(1): 112-135.

Westerlund, J., and Larsson, R. (2012). "Testing for a Unit Root in a Random Coefficient Panel Data Model." Journal of Econometrics **167**(1): 254-273.

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

# Appendix

## Appendix-A

**Table A.1: Power Analysis of Stationarity Tests with Drift but without Trend, at N=2, T=10 and T=25**

| Panel A: T=10 and N=2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 27.68 | 27.54 | 25.82 | 27.24 | 22.32 | 17.3 | 37.54 |
| 0.1 | 24.18 | 23.91 | 23.1 | 24.74 | 17.52 | 15.14 | 32.54 |
| 0.2 | 22.12 | 21.58 | 20.48 | 21.1 | 14.56 | 13.54 | 30.88 |
| 0.3 | 19.9 | 17 | 18.44 | 17.22 | 12.64 | 12.18 | 25.12 |
| 0.4 | 17.68 | 14.94 | 15.16 | 14.26 | 11.54 | 10.24 | 23.52 |
| 0.5 | 14.52 | 13.12 | 13.7 | 11.64 | 10.98 | 9.18 | 20.98 |
| 0.6 | 12.7 | 11.1 | 11.96 | 10.64 | 9.9 | 8.64 | 15.94 |
| 0.7 | 10.72 | 9.74 | 9.36 | 9.9 | 8.24 | 7.5 | 12.96 |
| 0.8 | 9.72 | 8.26 | 8.24 | 8.02 | 7.22 | 6.78 | 10.84 |
| 0.9 | 7.14 | 7.22 | 7.24 | 7.76 | 6.76 | 5.34 | 9.7 |
| 1 | 4.84 | 4.68 | 4.44 | 4.88 | 4.22 | 5.22 | 5.08 |
| **Panel B: T=25 and N=2** | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 80.76 | 79.92 | 64.42 | 67.7 | 71.42 | 63.04 | 89.68 |
| 0.1 | 73.98 | 71.88 | 53.7 | 57.06 | 55.38 | 55.03 | 81.78 |
| 0.2 | 64.26 | 63.96 | 43.96 | 50.4 | 50.08 | 45.82 | 72.38 |
| 0.3 | 52.66 | 52.84 | 38.98 | 43.78 | 44.78 | 35.6 | 61.22 |
| 0.4 | 39.64 | 40.06 | 31.36 | 36.58 | 36.32 | 26.92 | 48.92 |
| 0.5 | 30.46 | 29.98 | 24.36 | 28.48 | 29.4 | 21 | 38.98 |
| 0.6 | 20.78 | 21.98 | 19.9 | 22.72 | 21.78 | 15.6 | 29.82 |
| 0.7 | 14.22 | 15.26 | 13.96 | 18.74 | 15.16 | 10.88 | 20.06 |
| 0.8 | 11.14 | 11.14 | 10.56 | 12.12 | 11.82 | 9.26 | 16.98 |
| 0.9 | 8.92 | 7.3 | 6.8 | 8.64 | 8.76 | 6.22 | 12.98 |
| 1 | 5.34 | 4.84 | 5.26 | 6.12 | 4.58 | 4.08 | 4.4 |

**Table A.2: Power Analysis of Stationarity Tests with Drift but without Trend, at N=2, T=50 and T=100**

| Panel A: T=50 and N=2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 99.64 | 99.32 | 79.66 | 95.64 | 91.14 | 89.42 | 100 |
| 0.1 | 98.74 | 97.62 | 73.76 | 90.76 | 84.94 | 83.46 | 100 |
| 0.2 | 96.4 | 93.52 | 66.99 | 80.74 | 77.78 | 75.58 | 97.34 |
| 0.3 | 89.22 | 85.78 | 59.98 | 69.04 | 62.94 | 65.08 | 90.32 |
| 0.4 | 77.74 | 72.04 | 48.99 | 54.44 | 52.98 | 52.68 | 79.5 |
| 0.5 | 61.44 | 57.08 | 35.64 | 39.3 | 42.72 | 39.48 | 63.92 |
| 0.6 | 44.9 | 40.8 | 25.88 | 29 | 28.72 | 28.34 | 49.16 |
| 0.7 | 27.8 | 25.6 | 16.3 | 18.84 | 19.78 | 19.7 | 32.3 |
| 0.8 | 16.8 | 15.7 | 12.64 | 12.66 | 12.16 | 12.64 | 20.72 |
| 0.9 | 8.96 | 8.38 | 8.6 | 8.64 | 8.02 | 8.06 | 14.98 |
| 1 | 4.92 | 5.02 | 5.5 | 4.94 | 4.9 | 4.92 | 4.88 |
| Panel B: T=100 and N=2 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 97.64 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 99.98 | 87.92 | 100 | 100 | 97.8 | 100 |
| 0.2 | 99.5 | 99.88 | 80.92 | 97.88 | 99.96 | 91.64 | 100 |
| 0.3 | 98.08 | 98.76 | 73.12 | 91.48 | 92.46 | 83.64 | 99.64 |
| 0.4 | 96.02 | 95.36 | 66.64 | 80.44 | 79.76 | 70.22 | 96.86 |
| 0.5 | 88.56 | 85.14 | 59.9 | 64.76 | 67.94 | 62.92 | 90.42 |
| 0.6 | 70.34 | 65.88 | 43.68 | 45.1 | 49.46 | 44.96 | 70.56 |
| 0.7 | 48.86 | 43.1 | 33.2 | 29.68 | 22 | 29.76 | 49.56 |
| 0.8 | 27.08 | 25.3 | 25.98 | 16.64 | 17.42 | 14.98 | 27.74 |
| 0.9 | 10.86 | 10.08 | 12.42 | 9.9 | 10.68 | 9.22 | 16.48 |
| 1 | 5.3 | 5.5 | 5.36 | 5.44 | 5.26 | 4.5 | 5.76 |

**Table A.3: Power Analysis of Stationarity Tests with Drift but without Trend, at N=4, T=10 and T=25**

| Panel A: T=10 and N=4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 51.92 | 52.46 | 39.5 | 49.38 | 49.33 | 34.04 | 75.5 |
| 0.1 | 40.48 | 39.54 | 36.5 | 39.66 | 46.62 | 28.48 | 61.54 |
| 0.2 | 34.7 | 31.98 | 33.14 | 32.2 | 40.4 | 23.18 | 54.38 |
| 0.3 | 28.06 | 22.08 | 27.04 | 29.36 | 35.64 | 19.66 | 47.52 |
| 0.4 | 23.38 | 18.26 | 23.54 | 24.16 | 31.58 | 16.16 | 37.6 |
| 0.5 | 19.42 | 14.8 | 19.36 | 20.82 | 15.6 | 12.6 | 29.68 |
| 0.6 | 16.16 | 12.48 | 16.12 | 18 | 13.14 | 10.76 | 21.24 |
| 0.7 | 12.34 | 10.16 | 13.1 | 13.86 | 11.68 | 9.3 | 15.9 |
| 0.8 | 10.22 | 8.38 | 10.42 | 10.88 | 9.38 | 8.06 | 13.84 |
| 0.9 | 7.94 | 7.18 | 7.86 | 7.96 | 7.9 | 7.44 | 11.99 |
| 1 | 5.2 | 4.62 | 4.22 | 5.88 | 4.42 | 5.14 | 4.8 |
| Panel B: T=25 and N=4 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 95.12 | 94.98 | 62.96 | 97.02 | 89.64 | 78.84 | 100 |
| 0.1 | 90.14 | 90.58 | 57.84 | 93.26 | 83.76 | 71.78 | 99.06 |
| 0.2 | 82.54 | 82.06 | 50.2 | 84.98 | 76.72 | 63.18 | 97.08 |
| 0.3 | 71.82 | 70.82 | 45.96 | 73.24 | 66.54 | 51.68 | 92.86 |
| 0.4 | 60.95 | 58.94 | 35.5 | 62.04 | 57.96 | 40.96 | 83.04 |
| 0.5 | 44.02 | 42.9 | 27.76 | 41.78 | 47.32 | 31.1 | 67.46 |
| 0.6 | 27.82 | 30.96 | 21.74 | 30.06 | 34.66 | 23.04 | 48.64 |
| 0.7 | 18.5 | 19.7 | 14.98 | 21.22 | 26.92 | 16.04 | 33 |
| 0.8 | 11.9 | 14.06 | 11.91 | 13.48 | 19.5 | 10.48 | 24.84 |
| 0.9 | 7.44 | 8.7 | 8.76 | 9.98 | 10.64 | 7.54 | 14.38 |
| 1 | 4.42 | 5.12 | 4.6 | 4.22 | 4.82 | 4.86 | 4.44 |

**Table A.4: Power Analysis of Stationarity Tests with Drift but without Trend, at N=4, T=50 and T=100**

| Panel A: T=50 and N=4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 85.16 | 100 | 100 | 97.38 | 100 |
| 0.1 | 99.98 | 99.9 | 78.16 | 100 | 98.16 | 94.7 | 100 |
| 0.2 | 99.78 | 98.98 | 69.02 | 99.48 | 93.46 | 89.58 | 100 |
| 0.3 | 98.4 | 95.28 | 60.2 | 96.64 | 85.22 | 80.34 | 99.92 |
| 0.4 | 92.18 | 86.98 | 53.28 | 89.9 | 72.34 | 67.1 | 98.52 |
| 0.5 | 79.68 | 72.54 | 46.26 | 74.68 | 66.96 | 52.22 | 94.22 |
| 0.6 | 61.16 | 58.56 | 33.04 | 55.08 | 53.92 | 37.98 | 80.24 |
| 0.7 | 36.72 | 33.42 | 24.26 | 36.72 | 38.84 | 24.92 | 56.44 |
| 0.8 | 21.2 | 19.78 | 19.6 | 21.12 | 25.16 | 15.52 | 31.1 |
| 0.9 | 11.44 | 9.76 | 13.6 | 13.82 | 14.76 | 9.02 | 17.86 |
| 1 | 5.6 | 4.48 | 4.88 | 5.52 | 5.4 | 5.2 | 5.4 |
| Panel B: T=100 and N=4 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 98.38 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 90.38 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 83.98 | 100 | 100 | 97.62 | 100 |
| 0.3 | 100 | 99.82 | 76.14 | 99.98 | 100 | 94.04 | 100 |
| 0.4 | 99.82 | 98.76 | 68.98 | 98.86 | 99.7 | 85.32 | 100 |
| 0.5 | 97.96 | 92.18 | 59.64 | 93.12 | 88.34 | 70.32 | 99.88 |
| 0.6 | 87.38 | 79.06 | 49.81 | 76.84 | 70.91 | 55.58 | 97.8 |
| 0.7 | 69.04 | 63.3 | 37.23 | 57.94 | 56.42 | 41 | 83.88 |
| 0.8 | 38.66 | 37.44 | 28.02 | 29.4 | 33.74 | 20.74 | 52.3 |
| 0.9 | 14.74 | 12.54 | 16.14 | 17.24 | 17.94 | 15.32 | 25.7 |
| 1 | 4.84 | 4.86 | 5.6 | 4.94 | 4.94 | 5.96 | 4.72 |

**Table A.5: Power Analysis of Stationarity Tests with Drift but without Trend, at N=8, T=10 and T=25**

| Panel A: T=10 and N=8 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 72.74 | 74.86 | 58.04 | 75.98 | 67.95 | 54.46 | 92.58 |
| 0.1 | 69.64 | 68.56 | 52.36 | 63.98 | 62.96 | 45.96 | 88.4 |
| 0.2 | 62.24 | 63.78 | 46.78 | 56.92 | 57.72 | 36.5 | 79.04 |
| 0.3 | 57.16 | 58.72 | 40.66 | 47.96 | 50.66 | 30.14 | 69.72 |
| 0.4 | 51.06 | 51.1 | 34.94 | 39.96 | 44.24 | 23.1 | 62.08 |
| 0.5 | 43.96 | 44.66 | 27.36 | 27.04 | 35.92 | 18.26 | 49.46 |
| 0.6 | 35.4 | 35.68 | 22.2 | 21.24 | 27.84 | 14 | 38.72 |
| 0.7 | 24.36 | 23.38 | 17.28 | 16.08 | 23.12 | 12.12 | 29.64 |
| 0.8 | 16.88 | 18.94 | 12.02 | 13.4 | 17.22 | 9.16 | 17.62 |
| 0.9 | 9.04 | 8.8 | 8.92 | 10.56 | 11.86 | 8.62 | 13.98 |
| 1 | 4.94 | 5.52 | 4.36 | 4.76 | 5.46 | 4.58 | 4.96 |

| Panel B: T=25 and N=8 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 76.6 | 100 | 99.3 | 96.84 | 100 |
| 0.1 | 98.44 | 97.9 | 71.04 | 100 | 97.02 | 93.72 | 100 |
| 0.2 | 95.56 | 93.24 | 64.28 | 98.78 | 92.92 | 87.42 | 99.98 |
| 0.3 | 91.16 | 82.72 | 58.26 | 94.92 | 83.18 | 78.26 | 99.24 |
| 0.4 | 82.52 | 78.72 | 49.96 | 84.56 | 70.94 | 64.9 | 95.84 |
| 0.5 | 69.92 | 71.96 | 43.99 | 68.28 | 64.94 | 47.14 | 89.7 |
| 0.6 | 54.04 | 53.88 | 36.2 | 49.32 | 46.95 | 36.2 | 73.7 |
| 0.7 | 32.4 | 32.96 | 24.84 | 31.46 | 33.96 | 23.44 | 51.12 |
| 0.8 | 18.88 | 17.8 | 15.1 | 20.08 | 23.3 | 14.3 | 29.1 |
| 0.9 | 11.64 | 10 | 10.91 | 12.38 | 12.28 | 10.56 | 18.92 |
| 1 | 5.06 | 4.92 | 4.56 | 4.98 | 4.58 | 5.28 | 5.4 |

**Table A.6: Power Analysis of Stationarity Tests with Drift but without Trend, at N=8, T=50 and T=100**

| Panel A: T=50 and N=8 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 94.58 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 88.04 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 83.58 | 100 | 100 | 98.64 | 100 |
| 0.3 | 100 | 100 | 78.04 | 99.94 | 96.42 | 95.72 | 100 |
| 0.4 | 99.76 | 98.74 | 70.96 | 99.02 | 89.42 | 88.22 | 100 |
| 0.5 | 96.86 | 93.04 | 59.38 | 94.56 | 73.08 | 74.7 | 99.44 |
| 0.6 | 84.4 | 77.6 | 47.99 | 80.32 | 69.08 | 51.62 | 94.56 |
| 0.7 | 59.56 | 58.36 | 33.48 | 50.08 | 48.78 | 34.42 | 76.02 |
| 0.8 | 33.3 | 29.9 | 20.68 | 29.68 | 32.12 | 20.8 | 47.72 |
| 0.9 | 14.36 | 13.12 | 15.66 | 15.52 | 16.24 | 13.96 | 23.62 |
| 1 | 5.32 | 4.74 | 4.96 | 5.4 | 4.42 | 5.66 | 5.34 |
| Panel B: T=100 and N=8 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 99.82 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 93.82 | 100 | 100 | 99.9 | 100 |
| 0.3 | 100 | 100 | 85.22 | 100 | 100 | 99.28 | 100 |
| 0.4 | 100 | 100 | 78.56 | 100 | 100 | 96.94 | 100 |
| 0.5 | 99.94 | 99.32 | 69.88 | 100 | 95.18 | 88.18 | 100 |
| 0.6 | 98.66 | 94 | 59.66 | 96.54 | 79.08 | 70.54 | 99.78 |
| 0.7 | 85.92 | 79.24 | 48.95 | 77.9 | 68.94 | 53.94 | 95.98 |
| 0.8 | 63.42 | 57.56 | 34.04 | 46.86 | 45.2 | 29.54 | 72.88 |
| 0.9 | 22.34 | 21.86 | 19.46 | 19 | 19.96 | 15.98 | 36.72 |
| 1 | 5.18 | 4.52 | 5.6 | 6.26 | 5.06 | 5.38 | 4.42 |

**Table A.7: Power Analysis of Stationarity Tests with Drift but without Trend, at N=16, T=10 and T=25**

| Panel A: T=10 and N=16 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 92.06 | 91.04 | 79.5 | 99.58 | 99.99 | 78.22 | 100 |
| 0.1 | 87.06 | 84.32 | 71.3 | 92.06 | 95.46 | 71.46 | 100 |
| 0.2 | 81.88 | 81.88 | 65.86 | 82.58 | 88.58 | 63.08 | 98.64 |
| 0.3 | 76.98 | 76.92 | 58.98 | 73.04 | 77.91 | 53.06 | 95.58 |
| 0.4 | 69.98 | 68.96 | 52.96 | 65.94 | 68.94 | 47.91 | 88.24 |
| 0.5 | 55.84 | 55.74 | 43.52 | 52.92 | 49.95 | 34.18 | 74.58 |
| 0.6 | 41.44 | 39.48 | 36.08 | 37.96 | 33.98 | 22.8 | 58.44 |
| 0.7 | 28.98 | 26.4 | 30.42 | 24.32 | 24.78 | 16.26 | 43.98 |
| 0.8 | 23.38 | 21.82 | 23.92 | 17.82 | 19.54 | 10.74 | 30.98 |
| 0.9 | 15.78 | 14.58 | 14.32 | 13.22 | 13.92 | 11.6 | 20.91 |
| 1 | 5.36 | 4.78 | 4.86 | 4.56 | 4.72 | 5.58 | 4.64 |

| Panel B: T=25, N=16 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 97.5 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 93 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 84.82 | 100 | 100 | 100 | 100 |
| 0.3 | 99.8 | 99.86 | 75.46 | 100 | 98.74 | 97.36 | 100 |
| 0.4 | 97.74 | 98.9 | 66.3 | 99.24 | 94.08 | 91.08 | 99.9 |
| 0.5 | 88.38 | 93.52 | 57.18 | 94.88 | 82.68 | 78.12 | 99.44 |
| 0.6 | 76.92 | 77 | 49.68 | 77.92 | 70.56 | 60.7 | 95.42 |
| 0.7 | 58.93 | 57.93 | 38.82 | 53.68 | 53.86 | 33.96 | 76.86 |
| 0.8 | 33.66 | 29.6 | 27.08 | 28.72 | 32.34 | 24.18 | 47.2 |
| 0.9 | 17.46 | 14.46 | 17.49 | 16.82 | 15.96 | 12.1 | 27.76 |
| 1 | 4.52 | 4.94 | 5.24 | 5.24 | 5.02 | 5.02 | 100 |

**Table A.8: Power Analysis of Stationarity Tests with Drift but without Trend, at N=16, T=50 and T=100**

| Panel A: T=50 and N=16 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 99.22 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 94.14 | 100 | 100 | 99.96 | 100 |
| 0.4 | 100 | 100 | 87.76 | 100 | 100 | 99.16 | 100 |
| 0.5 | 99.98 | 99.76 | 81.32 | 99.92 | 99.32 | 95.58 | 100 |
| 0.6 | 98.48 | 96.76 | 73 | 97.62 | 86.79 | 83.72 | 99.9 |
| 0.7 | 86.68 | 80.62 | 50.42 | 82.64 | 71.98 | 60.5 | 97.18 |
| 0.8 | 58.76 | 57.28 | 35.94 | 49.54 | 48.48 | 30.72 | 73.7 |
| 0.9 | 20.4 | 23.4 | 19.12 | 19.1 | 19.76 | 16.3 | 30.44 |
| 1 | 5 | 4.3 | 4.94 | 4.86 | 4.46 | 5.6 | 4.54 |
| Panel B: T=100 and N=16 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 91.16 | 100 | 100 | 99.98 | 100 |
| 0.5 | 100 | 100 | 87.12 | 100 | 100 | 99.12 | 100 |
| 0.6 | 100 | 99.98 | 80.84 | 99.96 | 98.12 | 93.9 | 100 |
| 0.7 | 98.96 | 96.36 | 65.74 | 97.66 | 87.7 | 74.92 | 99.98 |
| 0.8 | 82.28 | 79.96 | 48.02 | 70.04 | 69.74 | 52.82 | 95.82 |
| 0.9 | 34.68 | 33.84 | 20.93 | 30.82 | 26.18 | 29.8 | 47.26 |
| 1 | 4.8 | 5.92 | 4.96 | 6.36 | 4.7 | 6 | 5.34 |

**Table A.9: Power Analysis of Stationarity Tests with Drift but without Trend, at N=32, T=10 and T=25**

| Panel A: T=10 and N=32 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 98.14 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 99.7 | 91.88 | 100 | 100 | 96.76 | 100 |
| 0.2 | 95.6 | 95.58 | 82.42 | 100 | 100 | 93.24 | 100 |
| 0.3 | 90.98 | 88.58 | 71.2 | 96.58 | 95.9 | 87.58 | 100 |
| 0.4 | 84.91 | 80.64 | 62.54 | 87.14 | 90.96 | 78.6 | 98.64 |
| 0.5 | 76.94 | 74.98 | 54.02 | 73.42 | 76.64 | 58.22 | 93.46 |
| 0.6 | 66.62 | 63.8 | 43.82 | 59.92 | 57.94 | 39.52 | 81.58 |
| 0.7 | 55.97 | 52.92 | 35.16 | 41.96 | 40.3 | 29.1 | 59.66 |
| 0.8 | 30.68 | 28.46 | 29.56 | 29.52 | 28.54 | 22.9 | 39.62 |
| 0.9 | 18.86 | 15.14 | 17.86 | 17.04 | 17.84 | 13.94 | 27.9 |
| 1 | 4.9 | 4.56 | 4.86 | 4.92 | 4.68 | 4.1 | 4.88 |
| Panel B: T=25 and N=32 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 94.66 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 91.68 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 86.8 | 100 | 100 | 99.62 | 100 |
| 0.5 | 99.52 | 99.82 | 78.14 | 99.86 | 98.28 | 97.12 | 100 |
| 0.6 | 94.36 | 93.54 | 66.8 | 97.64 | 89.06 | 82.76 | 99.66 |
| 0.7 | 78.96 | 80.94 | 54.54 | 81.42 | 70.38 | 54.5 | 94.68 |
| 0.8 | 56.96 | 52.02 | 38.96 | 48.42 | 45.92 | 27.44 | 69.96 |
| 0.9 | 23.1 | 20.98 | 22.28 | 23.3 | 20.4 | 17.9 | 34.32 |
| 1 | 5.32 | 4.92 | 5.26 | 4.44 | 4.9 | 4.98 | 5.02 |

**Table A.10: Power Analysis of Stationarity Tests with Drift but without Trend, at N=32, T=50 and T=100**

| Panel A: T=50 and N=32 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 94.32 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 100 | 87.4 | 100 | 99.84 | 99.98 | 100 |
| 0.6 | 100 | 100 | 78.42 | 100 | 97.82 | 98.3 | 100 |
| 0.7 | 98.92 | 97.72 | 63.76 | 98.94 | 83.46 | 86.74 | 99.86 |
| 0.8 | 78.94 | 76.28 | 49.38 | 72.06 | 65.19 | 49.52 | 92.14 |
| 0.9 | 32.32 | 30.38 | 28.24 | 34.02 | 24.02 | 22.72 | 42.82 |
| 1 | 5.04 | 4.48 | 5.2 | 6.12 | 4.96 | 4.7 | 5.4 |
| Panel B: T=100 and N=32 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 99.98 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 100 | 92.18 | 100 | 100 | 100 | 100 |
| 0.6 | 100 | 100 | 81.68 | 100 | 99.6 | 99.86 | 100 |
| 0.7 | 100 | 99.92 | 69.16 | 99.96 | 92.6 | 96.5 | 100 |
| 0.8 | 98.5 | 93.82 | 56.14 | 85.5 | 72.12 | 69.34 | 99.7 |
| 0.9 | 60.88 | 56.96 | 37.98 | 48.02 | 46.76 | 27.97 | 70.7 |
| 1 | 5.26 | 4.72 | 4.82 | 4.76 | 4.46 | 4.3 | 4.84 |

**Table A.11: Power Analysis of Stationarity Tests with Drift and Trend, at N=2, T=10 and T=25**

| Panel A: T=10 and N=2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 25.5 | 24.3 | 24.04 | 23.96 | 23.45 | 19.18 | 35.74 |
| 0.1 | 23.2 | 22.68 | 23.06 | 22.38 | 21.99 | 18.08 | 31.48 |
| 0.2 | 21.5 | 20.62 | 21.9 | 19.84 | 19.13 | 16.54 | 29.18 |
| 0.3 | 19.12 | 17.36 | 18.6 | 17.84 | 17.12 | 14.86 | 24.66 |
| 0.4 | 17.36 | 15.5 | 16.54 | 15.88 | 14.27 | 13.54 | 22.7 |
| 0.5 | 15.32 | 13.22 | 14.08 | 13.9 | 12.95 | 12.62 | 20.24 |
| 0.6 | 13.8 | 11.14 | 12.22 | 11.78 | 10.74 | 10.76 | 16.18 |
| 0.7 | 10.76 | 9.76 | 10.26 | 9.32 | 8.91 | 9.14 | 13.02 |
| 0.8 | 8.72 | 8.66 | 9.82 | 7.9 | 7.39 | 7.24 | 11.86 |
| 0.9 | 7.01 | 6.42 | 7.74 | 6.76 | 6.71 | 6.88 | 9.62 |
| 1 | 4.96 | 5.3 | 5.22 | 5.42 | 4.52 | 4.94 | 4.9 |

| Panel B: T=25 and N=2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 81.76 | 78.1 | 58.58 | 69.62 | 57.06 | 74 | 89.04 |
| 0.1 | 74.69 | 73.06 | 53.88 | 56.58 | 49.66 | 65.54 | 84.26 |
| 0.2 | 64.42 | 67.6 | 47.22 | 46.24 | 42.4 | 56.94 | 73.68 |
| 0.3 | 52.98 | 56.28 | 41.82 | 38.5 | 36.76 | 47.7 | 62.6 |
| 0.4 | 44.52 | 45.44 | 35.76 | 29.2 | 31 | 37.36 | 52.6 |
| 0.5 | 30.08 | 35.7 | 30.12 | 22.24 | 27.18 | 28.62 | 37.06 |
| 0.6 | 22.58 | 29.88 | 25.8 | 16.62 | 23.62 | 20.18 | 32.54 |
| 0.7 | 16.78 | 18.5 | 19.36 | 12.92 | 16.72 | 14 | 18.64 |
| 0.8 | 11 | 12.54 | 12.32 | 8.72 | 11.98 | 9.98 | 14.5 |
| 0.9 | 7.54 | 7.66 | 8.78 | 7.16 | 8.92 | 7.3 | 11.6 |
| 1 | 5.08 | 4.66 | 5.74 | 5.1 | 4.7 | 4.94 | 4.78 |

**Table A.12: Power Analysis of Stationarity Tests with Drift and Trend, at N=2, T=50 and T=100**

| Panel A: T=50 and N=2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 99.4 | 99.54 | 74.91 | 93.4 | 78.44 | 99.08 | 100 |
| 0.1 | 97.62 | 98.98 | 68.92 | 88.38 | 70.1 | 96.98 | 100 |
| 0.2 | 93.2 | 95.46 | 64.92 | 77.08 | 63.66 | 92.76 | 96.74 |
| 0.3 | 85.12 | 90.2 | 59.98 | 63.92 | 55.14 | 83.98 | 90.68 |
| 0.4 | 73.84 | 78.06 | 48.92 | 51.62 | 48.58 | 62.66 | 80.08 |
| 0.5 | 58.08 | 61.9 | 39.14 | 38.52 | 41.9 | 47.18 | 65.92 |
| 0.6 | 41.64 | 44.12 | 32.2 | 26.6 | 35.56 | 33.48 | 47.12 |
| 0.7 | 28.08 | 26.42 | 27.26 | 16.56 | 25.74 | 22.46 | 34.28 |
| 0.8 | 16.38 | 15.18 | 18.74 | 11.96 | 14.8 | 14.64 | 18.38 |
| 0.9 | 9.66 | 10.04 | 10.46 | 7.56 | 10.46 | 8.44 | 13.08 |
| 1 | 5.12 | 5.14 | 4.38 | 5.1 | 5.04 | 4.7 | 5.18 |
| Panel B: T=100 and N=2 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 92.6 | 99.9 | 92.52 | 100 | 100 |
| 0.1 | 100 | 100 | 86.16 | 99.46 | 89.28 | 100 | 100 |
| 0.2 | 99.82 | 99.98 | 81.12 | 97.4 | 83.48 | 99.98 | 100 |
| 0.3 | 99.02 | 99.08 | 75.02 | 93.06 | 75.4 | 99.42 | 99.78 |
| 0.4 | 94.62 | 98.08 | 68.02 | 81.36 | 63.28 | 89.8 | 98.14 |
| 0.5 | 82.92 | 89.66 | 58.64 | 64.14 | 53.3 | 79.98 | 89.88 |
| 0.6 | 69.06 | 68.56 | 44.92 | 46.34 | 41.48 | 58.96 | 75.56 |
| 0.7 | 49.01 | 48.6 | 35.6 | 27.66 | 27.36 | 33.08 | 52.26 |
| 0.8 | 26.54 | 27.3 | 23.64 | 16.64 | 16.36 | 21.9 | 29.82 |
| 0.9 | 13.26 | 12.7 | 14.8 | 8.58 | 10.56 | 12.14 | 17.1 |
| 1 | 4.76 | 5 | 5.88 | 4.92 | 5.62 | 5.04 | 4.46 |

**Table A.13: Power Analysis of Stationarity Tests with Drift and Trend, at N=4, T=10 and T=25**

| Panel A: T=10 and N=4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 49.46 | 46.9 | 38.84 | 44.48 | 44.1 | 43.34 | 69.34 |
| 0.1 | 42.88 | 42.44 | 35.8 | 37.52 | 41.89 | 38.1 | 65.28 |
| 0.2 | 37.76 | 37.01 | 32.82 | 30.6 | 37.57 | 33.76 | 54.46 |
| 0.3 | 32.66 | 33.58 | 27.28 | 25.2 | 32.4 | 28.2 | 48.7 |
| 0.4 | 27.98 | 28.06 | 24.56 | 20.64 | 27.9 | 22.94 | 36.74 |
| 0.5 | 22.22 | 23.44 | 21.68 | 15.64 | 22.39 | 17.86 | 26.8 |
| 0.6 | 17.38 | 18.5 | 18.84 | 12.7 | 17.21 | 14.4 | 22.06 |
| 0.7 | 12.9 | 13.96 | 14.46 | 9.68 | 12.89 | 11.22 | 16.88 |
| 0.8 | 9.4 | 10.04 | 10.72 | 7.54 | 9.09 | 9.24 | 11.72 |
| 0.9 | 7.78 | 8.32 | 7.96 | 6.22 | 7.76 | 6.56 | 9.3 |
| 1 | 4.94 | 4.72 | 4.78 | 4.9 | 5.11 | 5.24 | 5.62 |
| Panel B: T=25 and N=4 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 94.78 | 99.9 | 76.24 | 96.18 | 92.86 | 92.9 | 100 |
| 0.1 | 89.24 | 94.9 | 71.66 | 92.34 | 86.32 | 85.28 | 100 |
| 0.2 | 80.66 | 89.6 | 67.96 | 83.7 | 79.64 | 77.08 | 97.98 |
| 0.3 | 73.26 | 75.84 | 62.42 | 71.16 | 71.76 | 67.24 | 93.3 |
| 0.4 | 65.26 | 64.88 | 55.24 | 59.06 | 61.32 | 58.84 | 86.66 |
| 0.5 | 49.44 | 48.22 | 44.16 | 44.02 | 50.28 | 44.38 | 71 |
| 0.6 | 32.44 | 34.34 | 32.32 | 28.76 | 38.74 | 31.14 | 53.16 |
| 0.7 | 19.16 | 17.04 | 22.46 | 19.58 | 26.68 | 19.92 | 35.84 |
| 0.8 | 12 | 11.64 | 14.38 | 12.5 | 10.36 | 11.98 | 22.46 |
| 0.9 | 8.1 | 7.9 | 8.82 | 7.62 | 7.2 | 7.98 | 13.76 |
| 1 | 4.74 | 5.14 | 4.54 | 4.78 | 5.12 | 4.64 | 4.78 |

**Table A.14: Power Analysis of Stationarity Tests with Drift and Trend, at N=4, T=50 and T=100**

| Panel A: T=50 and N=4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 94.66 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 89.44 | 99.94 | 98.28 | 100 | 100 |
| 0.2 | 100 | 100 | 82.54 | 99.4 | 93.84 | 99.9 | 100 |
| 0.3 | 94.44 | 98.04 | 76.64 | 97.62 | 86.52 | 98.7 | 100 |
| 0.4 | 86.82 | 93.76 | 68.8 | 90.3 | 79.42 | 94.28 | 98.96 |
| 0.5 | 74 | 81.14 | 60.82 | 75.76 | 69.92 | 83.16 | 94.48 |
| 0.6 | 58.94 | 61.02 | 50.98 | 57.78 | 53.08 | 55.42 | 79.42 |
| 0.7 | 38.3 | 38.22 | 36.24 | 33.06 | 36.66 | 34.8 | 59.56 |
| 0.8 | 23.02 | 22.3 | 21.9 | 20.22 | 21.92 | 21.44 | 32.5 |
| 0.9 | 9.52 | 11.34 | 12.84 | 10.8 | 8.54 | 10.92 | 17.5 |
| 1 | 4.86 | 5.38 | 4.82 | 4.92 | 4.9 | 4.56 | 5.52 |
| Panel B: T=100 and N=4 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 97.88 | 100 | 100 | 100 | 100 |
| 0.4 | 98.76 | 99.8 | 90.92 | 99.56 | 97.54 | 99.96 | 100 |
| 0.5 | 93.6 | 97.8 | 82.62 | 96.26 | 91.62 | 99.42 | 100 |
| 0.6 | 78.56 | 87.28 | 70.16 | 84.82 | 78.84 | 82.94 | 98.4 |
| 0.7 | 65.96 | 65.51 | 52.3 | 59.94 | 59.46 | 63.32 | 86.52 |
| 0.8 | 36.26 | 34.42 | 31.28 | 32.72 | 28.76 | 32.4 | 53.56 |
| 0.9 | 13.32 | 14.44 | 16.56 | 14.26 | 14.66 | 18.34 | 21.16 |
| 1 | 5.1 | 5.22 | 5.78 | 4.8 | 4.04 | 5.98 | 5.02 |

**Table A.15: Power Analysis of Stationarity Tests with Drift and Trend, at N=8, T=10 and T=25**

| Panel A: T=10 and N=8 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 75.8 | 73.42 | 69.2 | 74.44 | 69.99 | 77.74 | 94.62 |
| 0.1 | 67.56 | 67.32 | 63.82 | 63.9 | 64.91 | 70.12 | 89.94 |
| 0.2 | 59.98 | 58.96 | 55.9 | 54.52 | 56.79 | 62.32 | 81.66 |
| 0.3 | 53.7 | 51.91 | 42.6 | 45.94 | 49.93 | 50.74 | 73.38 |
| 0.4 | 44.58 | 42.34 | 33.12 | 32.86 | 37.22 | 40.58 | 59.94 |
| 0.5 | 36.14 | 34.7 | 25.94 | 23.62 | 29.65 | 31.6 | 48.06 |
| 0.6 | 27.34 | 27.3 | 22.08 | 17.94 | 21.1 | 22.9 | 35.18 |
| 0.7 | 17.06 | 16.9 | 18.76 | 14.46 | 16.91 | 17.9 | 22.86 |
| 0.8 | 9.34 | 11.54 | 11.38 | 9.82 | 11.01 | 10.96 | 16.3 |
| 0.9 | 7.76 | 8.74 | 8.04 | 7.24 | 6.95 | 7.94 | 10.8 |
| 1 | 5.44 | 4.7 | 4.24 | 5.58 | 5.06 | 5.56 | 4.76 |
| Panel B: T=25 and N=8 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 98.3 | 100 | 98.38 | 100 | 100 |
| 0.1 | 100 | 100 | 90.04 | 100 | 95.38 | 100 | 100 |
| 0.2 | 98.14 | 98.02 | 82.6 | 98.62 | 89.36 | 99.7 | 100 |
| 0.3 | 93.64 | 93.28 | 74.34 | 95.32 | 81.72 | 98.22 | 100 |
| 0.4 | 83.74 | 81.96 | 66.54 | 85.16 | 72.48 | 91.98 | 97.26 |
| 0.5 | 67.96 | 68.42 | 57.98 | 64.56 | 63.92 | 78.98 | 89.88 |
| 0.6 | 49.94 | 49.18 | 42 | 49.4 | 49.16 | 50.06 | 71.62 |
| 0.7 | 31.3 | 30.14 | 32.08 | 31.44 | 29.28 | 30.98 | 49.4 |
| 0.8 | 19.52 | 17.18 | 19.42 | 17.32 | 17.16 | 18.78 | 27.4 |
| 0.9 | 10.08 | 9.52 | 12.06 | 9.7 | 11.58 | 10.34 | 18.14 |
| 1 | 4.9 | 5.1 | 5.2 | 5.1 | 4.52 | 5.18 | 5.46 |

**Table A.16: Power Analysis of Stationarity Tests with Drift and Trend, at N=8, T=50 and T=100**

| Panel A: T=50 and N=8 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 97.6 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 92.46 | 100 | 100 | 100 | 100 |
| 0.4 | 98.96 | 99.4 | 85.18 | 99.54 | 96.26 | 99.98 | 100 |
| 0.5 | 93.36 | 95.76 | 76.02 | 96.06 | 88.22 | 98.92 | 100 |
| 0.6 | 78.28 | 82.66 | 63.92 | 82.52 | 78.54 | 92.28 | 95.9 |
| 0.7 | 58 | 57.56 | 47.56 | 54.72 | 53.44 | 71.2 | 79.14 |
| 0.8 | 29.14 | 31.08 | 28.04 | 31.22 | 28.62 | 27.16 | 48.38 |
| 0.9 | 13.26 | 13.02 | 13.5 | 13.58 | 12.16 | 13.28 | 24.78 |
| 1 | 5.4 | 5.04 | 5.32 | 5.18 | 4.64 | 5.58 | 4.96 |
| Panel B: T=100 and N=8 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 100 | 100 | 99.86 | 100 | 100 |
| 0.5 | 99.66 | 100 | 96.86 | 100 | 97.16 | 97.1 | 100 |
| 0.6 | 94.78 | 98.84 | 82.56 | 98.68 | 87.24 | 89.86 | 99.92 |
| 0.7 | 76.9 | 86.24 | 65.92 | 85.9 | 72.94 | 76.3 | 97.1 |
| 0.8 | 53.06 | 52.98 | 46.34 | 50.04 | 48.76 | 52.92 | 73.92 |
| 0.9 | 24 | 20.18 | 23.24 | 21.32 | 20.68 | 17.5 | 34.12 |
| 1 | 5.12 | 4.62 | 4.48 | 5.68 | 4.38 | 5.14 | 4.6 |

**Table A.17: Power Analysis of Stationarity Tests with Drift and Trend, at N=16, T=10 and T=25**

| Panel A: T=10 and N=16 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 96.8 | 93.5 | 88.24 | 96.4 | 91.84 | 95.32 | 100 |
| 0.1 | 90.8 | 87.2 | 82.22 | 91.28 | 85.64 | 91.12 | 100 |
| 0.2 | 85.18 | 81.12 | 74 | 82.24 | 79.68 | 84.72 | 98.82 |
| 0.3 | 78.76 | 75.12 | 66.6 | 72.08 | 72.13 | 75.54 | 95.64 |
| 0.4 | 69.66 | 69.5 | 56.36 | 65.4 | 62.93 | 66.2 | 87.66 |
| 0.5 | 56.68 | 58.68 | 45.88 | 54.84 | 51.94 | 55.9 | 76.98 |
| 0.6 | 40.5 | 44.78 | 32.34 | 42.16 | 43.41 | 40.52 | 60.28 |
| 0.7 | 27.6 | 27.56 | 28.62 | 26.82 | 27.21 | 26.84 | 38.6 |
| 0.8 | 17.98 | 16.36 | 17.1 | 17.22 | 16.77 | 16.4 | 22.66 |
| 0.9 | 10.9 | 10.36 | 10.56 | 8.72 | 8.5 | 8.3 | 13.48 |
| 1 | 5.28 | 5.02 | 4.88 | 5.64 | 4.7 | 4.76 | 5.62 |
| Panel B: T=25 and N=16 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 95.48 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 87.5 | 99.94 | 98 | 100 | 100 |
| 0.4 | 98.54 | 97.9 | 78.6 | 99.1 | 94.12 | 99.8 | 100 |
| 0.5 | 91.98 | 88.92 | 70.22 | 94.1 | 83.76 | 97.6 | 100 |
| 0.6 | 78.22 | 75.92 | 64.94 | 78.58 | 75.5 | 87.12 | 95.72 |
| 0.7 | 59.84 | 59.24 | 50.14 | 56.92 | 54.96 | 59.96 | 78.52 |
| 0.8 | 31.16 | 29.58 | 35.54 | 28.2 | 28.54 | 36.66 | 49.4 |
| 0.9 | 12.62 | 12.34 | 15.46 | 13.9 | 10.36 | 16.66 | 20.44 |
| 1 | 4.4 | 4.84 | 5.66 | 5.58 | 4.94 | 5.46 | 4.86 |

**Table A.18: Power Analysis of Stationarity Tests with Drift and Trend, at N=16, T=50 and T=100**

| Panel A: T=50 and N=16 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 94.8 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 100 | 86.76 | 100 | 97.06 | 100 | 100 |
| 0.6 | 96.84 | 98.28 | 77.84 | 98.46 | 86.22 | 99.78 | 100 |
| 0.7 | 80.76 | 84.42 | 65.78 | 86.54 | 74.94 | 93.34 | 97.74 |
| 0.8 | 58.1 | 56.94 | 45.88 | 53.82 | 52.94 | 57.56 | 75.12 |
| 0.9 | 19.48 | 18.1 | 17.46 | 20.62 | 18.76 | 25.02 | 29.18 |
| 1 | 5.52 | 4.86 | 4.94 | 5.6 | 5.54 | 4.5 | 3.96 |

| Panel B: T=100 and N=16 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 100 | 98.4 | 100 | 100 | 100 | 100 |
| 0.6 | 100 | 100 | 88.12 | 100 | 98.34 | 100 | 100 |
| 0.7 | 96.42 | 98.96 | 76.92 | 98.72 | 89.32 | 99.94 | 100 |
| 0.8 | 79.96 | 78.98 | 63.06 | 80.1 | 74.78 | 79.08 | 96.02 |
| 0.9 | 38.4 | 33.56 | 29.44 | 31.42 | 30.08 | 34.32 | 51.3 |
| 1 | 4.72 | 4.52 | 4.44 | 4.92 | 5.02 | 5.5 | 4.74 |

**Table A.19: Power Analysis of Stationarity Tests with Drift and Trend, at N=32, T=10 and T=25**

| Panel A: T=10 and N=32 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 99.56 | 100 | 100 | 100 | 100 |
| 0.2 | 98.6 | 95.2 | 94.56 | 98.7 | 97.52 | 98.9 | 100 |
| 0.3 | 96.48 | 89.12 | 87.68 | 94.52 | 92.48 | 95.9 | 100 |
| 0.4 | 89.16 | 83.74 | 74.18 | 83.52 | 84.9 | 88.44 | 98.52 |
| 0.5 | 75.08 | 74.92 | 65.24 | 73.34 | 70.12 | 73.9 | 94.34 |
| 0.6 | 63.78 | 63.46 | 49.88 | 65.86 | 57.3 | 61.38 | 81.86 |
| 0.7 | 39.66 | 39.2 | 33.56 | 48.42 | 37.91 | 37.82 | 58.76 |
| 0.8 | 22 | 21 | 22.08 | 30.96 | 23.88 | 21.18 | 35.6 |
| 0.9 | 13.6 | 14.4 | 15.46 | 15.58 | 14.9 | 12.08 | 20.68 |
| 1 | 5.3 | 4.96 | 5.3 | 5.42 | 5.14 | 5.12 | 4.96 |
| Panel B: T=25 and N=32 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 96.9 | 100 | 100 | 100 | 100 |
| 0.5 | 99.72 | 99.5 | 87.18 | 99.98 | 98.28 | 100 | 100 |
| 0.6 | 96.9 | 94.34 | 74.02 | 97.38 | 91.92 | 99 | 100 |
| 0.7 | 80.04 | 77.96 | 64.16 | 83.92 | 73.74 | 88.22 | 95.26 |
| 0.8 | 57.28 | 53.98 | 48.24 | 52.56 | 48.66 | 53.06 | 70.56 |
| 0.9 | 20.58 | 17.34 | 16.04 | 18.68 | 16.76 | 18.7 | 27.9 |
| 1 | 4.4 | 5.14 | 4.82 | 5.38 | 4.44 | 4.8 | 5.12 |

**Table A.20: Power Analysis of Stationarity Tests with Drift and Trend, at N=32, T=50 and T=100**

| Panel A: T=50 and N=32 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 100 | 98.68 | 100 | 100 | 100 | 100 |
| 0.6 | 100 | 100 | 92.82 | 100 | 97.46 | 100 | 100 |
| 0.7 | 97.92 | 99.3 | 81.64 | 98.92 | 82.38 | 99.78 | 99.88 |
| 0.8 | 79.74 | 81.64 | 62.76 | 81.7 | 74.52 | 78.3 | 93.52 |
| 0.9 | 31.3 | 30.68 | 29.04 | 29.06 | 29.8 | 33.28 | 45.26 |
| 1 | 4.8 | 5.14 | 5.48 | 4.98 | 4.02 | 4.58 | 4.3 |
| Panel B: T=100 and N=32 | | | | | | | |
| $\varphi$ | HD | HL | HLM | KK | SS | DHT | PPO Test |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.6 | 100 | 100 | 98.68 | 100 | 100 | 100 | 100 |
| 0.7 | 100 | 100 | 83.82 | 99.98 | 95.08 | 100 | 100 |
| 0.8 | 94.76 | 98.38 | 69.58 | 98.48 | 86.1 | 89.56 | 99.78 |
| 0.9 | 54.3 | 53.88 | 38 | 52.86 | 52.84 | 52.86 | 67.92 |
| 1 | 5.74 | 4.8 | 5.44 | 5.24 | 4.42 | 5.26 | 4.6 |

# Appendix-B

**Figure B.1: MSC Assessments of PUR Tests, Excluding Worst Performer, with Intercept only, N=4**

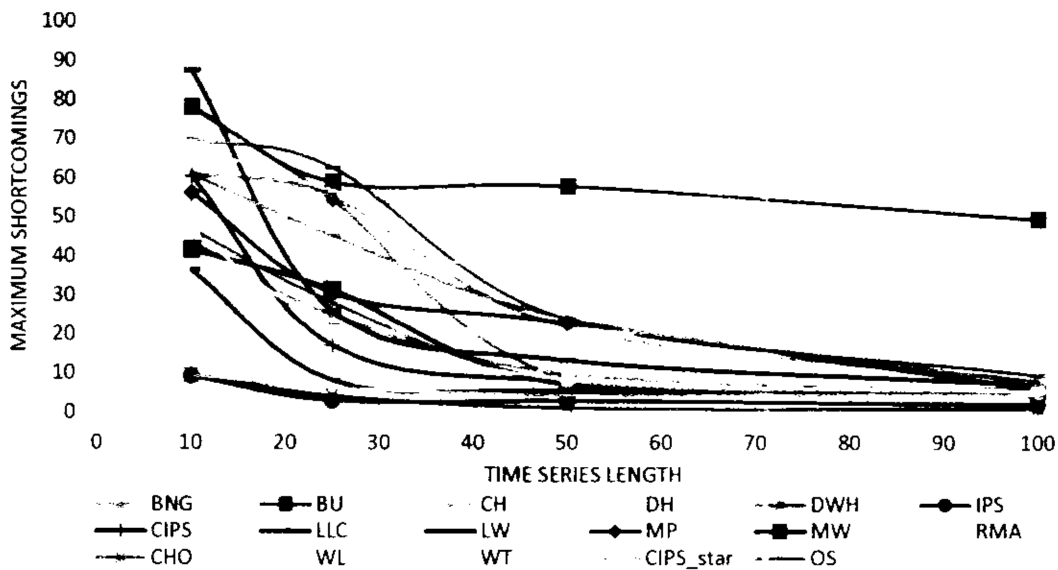



**Figure B.2: MSC Assessments of PUR Tests, Excluding Worst Performer, with Intercept only, N=8**

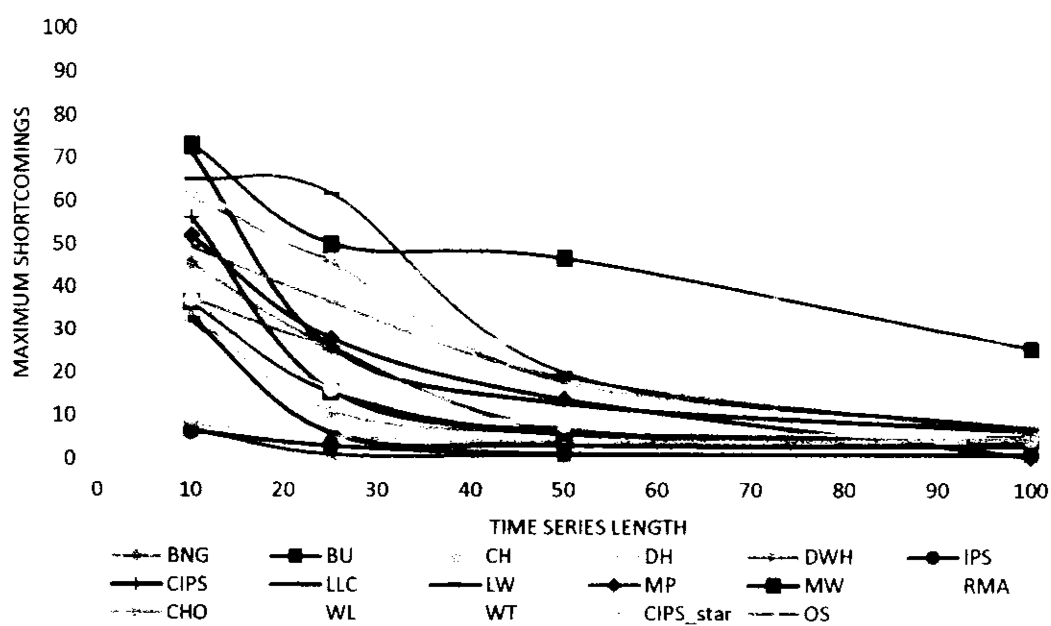


**Figure B.3: MSC Assessments of PUR Tests, Excluding Worst Performer, with Intercept and Trend, N=4**

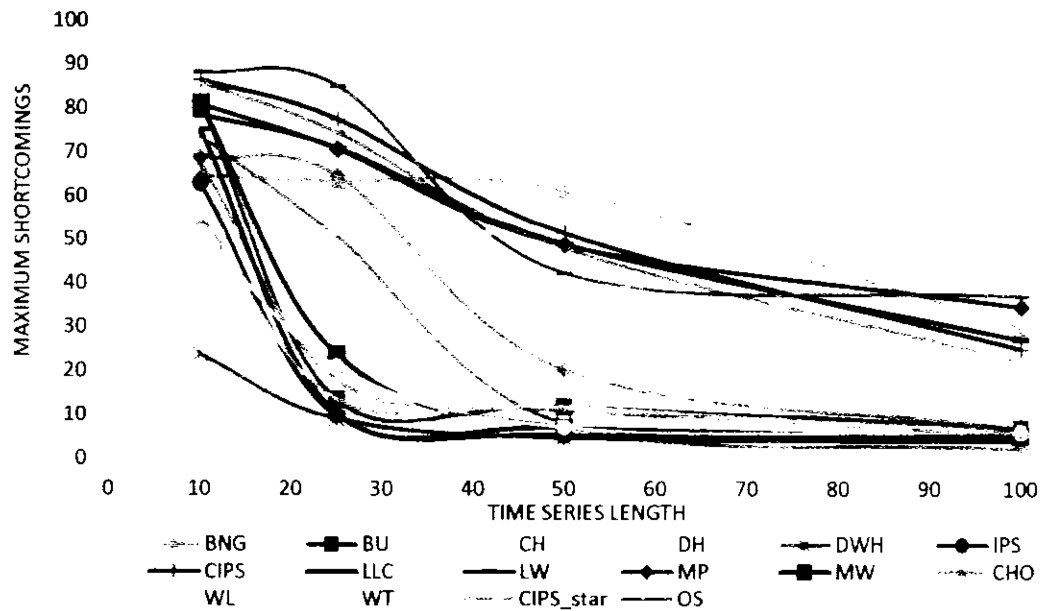



**Figure B.4: MSC Assessments of PUR Tests, Excluding Worst Performer, with Intercept and Trend, N=8**

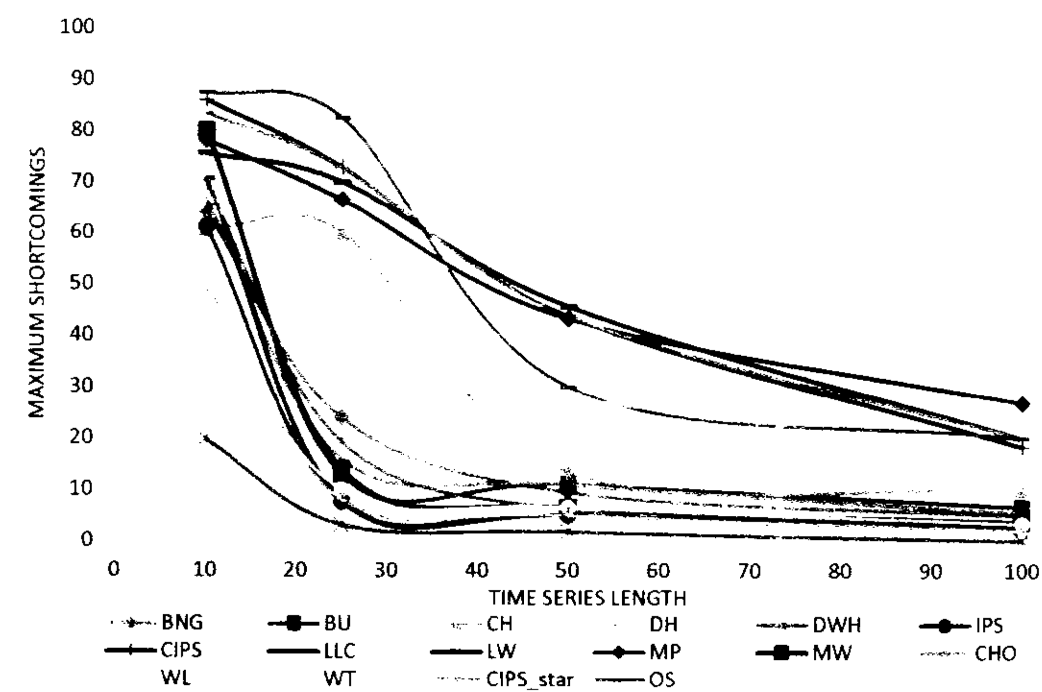


**Figure B.5: MSC Assessments of PUR Tests, Excluding Worst Performer, with Intercept Term only, T=25**

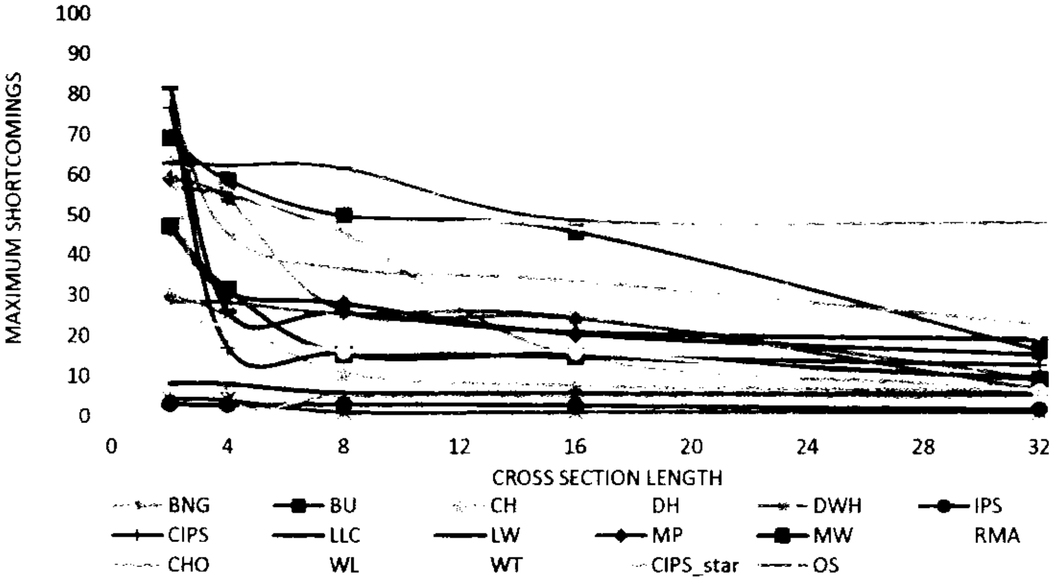



**Figure B.6: MSC Assessments of PUR Tests, Excluding Worst Performer, with Intercept and Trend Term, T=25**