بسم الله الرحمن الرحيم

ASSESSING INDEPENDENCE AND CAUSALITY IN TIME SERIES

PhD Econometrics

Researcher: Waqar Muhammad Khan

Registration No. 21-SE/PhD (ET)/S12

Supervisor: Dr. Asad Zaman

Co-supervisor: Dr. Atiq ur Rehman

School of Economics

INTERNATIONAL INSTITUTE OF ISLAMIC ECONOMICS

INTERNATIONAL ISLAMIC UNIVERSITY, ISLAMABAD

ASSESSING INDEPENDENCE AND CAUSALITY IN TIME SERIES

Waqar Muhammad Khan,

Registration No. 21-SE/PhD (ET)/S12

Submitted in partial fulfillment of the requirements for the

Doctor of Philosophy in Econometrics

at International Institute of Islamic Economics

International Islamic University, Islamabad

Supervisor: Dr. Asad Zaman

Co-supervisor: Dr. Atiq ur Rehman

June 2018

1 - Causality
2 - Time series.
3 - Thesis.

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.

Waqar Muhammad Khan

DEDICATED

To my mother

APPROVAL SHEET

Assessing Independence and Causality in Time Series

by

Waqar Muhammad Khan
Reg. No: 21-SE/PhD(Et)/S12

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

Co-Supervisor: Dr. Atiq-ur-Rehman
Assistant Professor, PIDE Islamabad

Internal Examiner: Dr. Abdul Jabbar
Assistant Professor, IIIE, IIUI

External Examiner: Dr. Ejaz Ghani
Professor, PIDE Islamabad

22.10.18

External Examiner: Dr. Zahid Asghar
Associate Professor, Department of Statistics, Quaid-e-Azam University Islamabad

Dr. Hamid Hassan
Chairman School of Economics

Dr. Atiquzzafar
Director General

International Institute of Islamic Economics
International Islamic University, Islamabad.

Date of Viva-Voce Examination: 1st June, 2018

ACKNOWLEDGEMENTS

All praises are to Allah, Almighty, Peace and Blessing of Allah on his last Prophet Muhammad who taught the humanity to human beings.

All achievements by me are nothing but a blessing of Allah. I am most obliged to my parents, sisters, brothers, bhabi, Dr Atiq-ur-Rehman, Dr Asad Zaman and specially to my wife and children who made it possible for me to earn this degree.

At this occasion, I am thankful to all my teachers who taught me at any stage of my studies. I would acknowledge the teaching of Dr. Arshad Ali Bhatti and Dr. Faiz-ur-Raheem for their kind style of teaching. I must acknowledge cooperation and support of my teachers and seniors including Dr. Iftikhar Hussain Adil, Dr. Mudassar Rashid and Dr. Ehasn Satti. I would like to acknowledge many friends and colleagues who helped me during my PhD and in my early career and provided me excellent company to enjoy. I would thank all my class fellows during PhD including Mr. Khan Bahadar, Mr. Mehmood Hussan, Mr. Irfan Malik, Mr. Yasir Riaz, Mr. Shahid Akbar, Mr. Asad-Ul-Islam Khan and Mr. Abdul Waheed whose company I always enjoyed. I also acknowledge the cooperation of the staff of IIIE including Mr. Syed Niaz Ali Shah, Mr. Zareef and Mr. Tauqeer Ahmed. Many thanks are also due to my class fellows Mr. Asad-Ul-Islam for cooperation in preparing initial draft of this thesis. I am not able to find words to say thanks to my mother Ms. Khalida Bi, Bhai Gulfraz, and Nisar, Brother in law Muhammad Khan, Aftab Mehmood, Bhabis, sisters and my wife Ms. Saima Khan for their continuous cooperation, support, encouragement and lot more. Last but not the least I am thankful to my parents and wife for their prayers, sacrifices, physical contribution, patience and immortal love that encouraged me at every flash of study period.

Waqar Muhammad Khan

ABSTRACT
Assessing Independence and Causality in Time Series

Assessing independence of two series was, is and will be the most fundamental goal of econometric/economic practitioners. Most of the economic (especially macroeconomic) data are time dependent and economists are interested to validate different economic theories using sophisticated data analysis tools. At the end of 19th century Karl Pearson developed coefficient of correlation to assess correlation between two series, however Yule (1926) criticized the use of this coefficient for time series because of its autocorrelated behavior. Haugh (1976) was considered the first to propose a measure of correlation for time series. His idea of pre-whitening the series' first and then apply correlation coefficient became very famous. Since then different versions of Haugh test were developed. The latest version was developed by Rehman and Malik (2014) that is also based on the same idea, however, the pre-whitening process became more refined and modified.

For the last four decades, several tests of independence for time series are developed. Every test was developed on a particular property of underlying assumptions, and it works in its own domain and fails to work in other situations/domains. Researchers working in this area, presented few results to compare proposed test and one or two previously developed tests and concluded superiority of his/her proposed test. These studies include Hong (1996b), Duchesne and Roy (2003a) and so on. These comparisons are ad hoc in nature and no comprehensive study available in the literature to unfold the strengths and weaknesses of these tests.

We organized a comprehensive Monte Carlo simulation study to compare tests of independence and selected a broad data generating process in a common framework. For this purpose, standard stringency criteria of Zaman (1996) has been used. In presented study, we have selected eleven available tests of independence for time series. To use stringency criteria, tests of independence should maintain a stable size and then based on its powers we can decide about the appropriateness of the test. For checking the size stability of the tests, we have used twenty-one different specifications of stochastic part in our data generating process, similarly two specifications of deterministic part have been used for each stochastic part case. So, we have tested size stability at total forty-two different combinations of data generating process. Simulated critical values were used, as past studies suggested that asymptotic critical values are not appropriate. All eleven tests of independence have their sizes around nominal size of 5%. Following the stable size, power analysis has been carried out for the same combinations of the data generating process and results suggests that Atiq test performs well in small sample size in almost all cases. PhamRoy test remains on second position in small samples but in many situations, it supersedes Atiq test in medium and large sample sizes. Haugh test remains at third place in almost all cases of the simulation study however the difference between the shortcomings of Haugh and PhamRoy test are very large. The remaining tests have not shown any considerable performance. Kim Lee, LiHui and Bouhaddioui tests considered worst in all sample sizes and with and without deterministic part cases.

Another important contribution of this study is to compare three important techniques used to check the dependence of two or more-time series, these include cointegration, Granger causality and Tests of independence for time series. Using renowned Keynesian function of income and consumption, we applied these tests on real economic data of income and consumption of 100 countries from 1970 to 2014. The results depict that three selected tests of independence, i.e. Atiq, PhamRoy and Haugh tests have appreciable power gains and lower size distortions. Again, it is observed that Atiq test shows better empirical power gain with least size distortion. PhamRoy and Haugh tests also shows good performance, have good power gains and small size distortions. However, five cointegration tests which are considered relatively better by Khan (2017) and famous Granger causality test have shown very poor performance both in terms of real empirical size and power.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ACRONYMS

ACRONYMS	COMPLETE NAME
MSC	Maximum Shortcomings
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
HQC	Hannan–Quinn information criterion
DGP	Data Generating Process
AR	Autoregressive
ARMA	Autoregressive Moving average
IID	Independent and Identically Distributed
VARMA	Vector Autoregressive Moving average
CV	Cramer-von Mises
TECM	t Test of Cointegration in a Single Equation Error Correction Model
HAZA	Hansen's variation of the $\hat{Z}_\alpha$ test of Cointegration
HADF	Hansen's variation of the ADF test of Cointegration
POZA	Phillips and Ouliaris' $\hat{Z}_\alpha$ test of Cointegration
POPU	Phillips and Ouliaris' $\hat{P}_\mu$ test of Cointegration
PO	Point Optimal
AS	Alternate Space
MCSS	Monte Carlo Sample Size
LLH(H0)	Log Likelihood under Null Hypothesis
LLH(H1)	Log Likelihood under Alternative Hypothesis
MA	Moving Average
Haugh Test	Test of Independence proposed by Haugh in 1976
Koch and Yang Test	Test of independence proposed by Paul D. Koch and Shie-Shien Yang in 1986
Li and Hui Test	Test for cross correlation in Time series by W. K. Li and Y. V. Hui in 1994
Hong Test	Test of Independence proposed by Yongmiao Hong in 1996
Hallin and Saddi	Test of noncorrelation proposed by Marc Hallin and Abdessamad Saidi in 2001
PhamRoy Test	Test of noncorrelation proposed by Dinh Tuan Pham, Roch Roy and Lyne C´edra in 2003
Bouhaddioui Test	Test of noncorrelation proposed by Chafik Bouhaddioui and Jean-Marie Dufour in 2008
Bouhaddioui and Roy Test	Test of independence proposed by Chafik Bouhaddioui and Roch Roy in 2006
Kim and Lee Test	Test of independence proposed by Eunhee Kim and Sangyeol Lee in 2004
Mariano Test	A symbolic test of independence proposed by Mariano Matilla-Garcı´a, Jose´ Miguel Rodrı´guezb and Manuel Ruiz Marı´n in 2010
Atiq Test	The Modified R test proposed by Atiq ur Rehman and Muhammad Irfan Malik 2014

CHAPTER 1: INTRODUCTION

In this chapter we will briefly describe the problem of dependence for time series and its implications on Econometric literature. We will further discuss the significance of this study and describe which factors inspired us to execute such study. Main objectives will also be illustrated to show the strengths of this study.

1.1: Contextual of the Study

In many situations, the difficulty arises in assessing the independence of two time series. Evaluating the possible relationships between two or more variables is central in many Economic applications. Perceiving the particular interest in this context, is the hypothesis of independence or non-correlation between two observed series. Therefore, it is necessary to construct methods which are simple, powerful, and easily interpretable for testing this hypothesis.

Generally the engineering and economic series are time dependent and in modeling the systems, the problem often translate in describing the interrelationship existing between two time series. Mainly when someone's initial goal is of system identification, his/her first concern is checking the independence of the two series, after that, the second question is discovering whether one of the series may act as a leading indicator of the other.

In macroeconomic system, it may also happen that both variables in the system remain response ones, or we can say that causal relations may happens in both ways. Say we have two series X_t and Y_t, Our objective is to test the null hypothesis that these two series are uncorrelated at all leads and lags. If Gaussian assumptions are made (on the joint distribution of the two series), non-correlation of-course is equivalent to independence. Dependence is a symmetric property which means that two variables are interrelated; however in causality X_t variable causes Y_t, Y_t variable causes X_t or both cause each other.

It would be helpful to check the independence of two series first before going for further advanced level analysis techniques. The cross-correlation function between the two series is the basic tool used for this purpose, but especially when we are dealing with the macroeconomic time series (which are generally autocorrelated), the usual estimates of the lagged cross-correlation function between the two series can be difficult to interpret. The autocorrelation present in each of the series (even if the series have been detrended) can inflate the variance of cross-correlation estimates.

In 19^{th} century Karl Pearson familiarized the findings of Francis Galton as coefficient of correlation to analyze the two series. Pearson product moment correlation coefficient (PPMCC) is used to check the strength of relation between two variables; however the practice of this coefficient for time series was criticized by Yule (1926). As per his objections the nature of time series is different from the cross sectional data. Most of the time series have property of the serial dependence of the observations. Owing to this property, cross sectional and time series attained different distributions of several estimators. Serial dependence is an important source of spurious correlation among time series.

Notion of spurious correlation is revolutionary in Econometrics and scientists believed that usual tests of independence are not suitable for time series, as propensity of autocorrelation in time series may lead to spurious results. In literature, it is believed that Haugh (1976)was the first who gave the idea of test of independence for time series and after that Geweke (1981), Koch and Yang (1986), Hong (1996b) and many others proposed different test of independence for time series.

As we mentioned above during the last three decades many tests of independence for time series have been developed. Every one of these test is based on some philosophy and works in its own domain. There is need to compare these tests in a common frame work to judge their capabilities. In literature we did not found a single comprehensive study to decide the superlative test for independence in time series. This study has been conducted to actually fill this gap and provide Economists with a clear understanding of the test to be used to test the independence in time series. Furthermore this study also enhances the vision of whether we can use the advance level analysis for time series or some basic analysis give us more information about the relationship between two variables.

This study comprises of seven chapters, the first one contains some background of the proposed problem, inspiration motivation and significance of the study. In second chapter we briefly describe the literature, when and how these tests were formulated, and if there was a comparative study conducted in the literature, what were its findings. Third chapter encompasses the brief methodology used in our study. In fourth chapter we analyze the size comparison of tests of independence for time series. Power comparison of tests of independence for time series are debated in chapter five. Real data assessment conducted in our study is briefly explained in chapter six and in last chapter the inferences, recommendations and guidelines for further study are provided.

1.2: Inspiration

There are a number of tests developed by different researchers at times. These tests are used to check the independence of time series under different conditions. All these tests have some weaknesses; however some tests are stronger in one area and weaker in other. Each one of these tests is constituted on a different theme and a particular test should be strong in its own domain, however no particular study has been conducted yet to unfold the best test amongst those available tests in the literature which have least shortcomings. Very few studies (see (Duchesne & Roy, 2003b)) are conducted but no one jointly compared all these tests in a common framework to check the weak and strong points of these tests under the same conditions.

We have seen a number of articles in which more complicated techniques used for analysis of time series are applied. Without knowing the theoretical applicability, the researchers determine the relationship of one variable with the other on the basis of different advanced level time series analysis. There is need of a comprehensive theoretical analysis of basic and advanced level techniques and clear verdict about the applicability of these practices.

1.3: Worth of the Study

Assignment of assessing potential dynamic relationship between two time series is given to Econometricians, it is particularly crucial in economics and other relevant sciences. The point being that several economic phenomena (specifically macroeconomic) can be designated by time series. It indicates that it is relevant and customary to test the existence of a relationship, mainly dependence between series, and then to investigate the dynamic nature of it. When we have to deal with two time series, the question often arises of describing the interrelationships existing between them. In economics, for example, elucidation of causality relationships between time series may be very important in a prediction context. Before applying sophisticated methods for describing relationships between two time series, it is important to check whether they are independent (or serially uncorrelated in the non-Gaussian case) or not. If sufficiently powerful methods, which are simple both to apply and to interpret, were available for checking the independence of two time series, more sophisticated analyses as causality analysis might become redundant.

In distinction to the general relevance of investigating the dependence of two or more series, comparatively few attempts have been made to test this kind of dependence between two time series. However one needs a reliable measure of dependence for checking the two time series. Among all available tests, it is important to distinguish a single one which dominates others.

In regression the question of true determinants of regressand always have a vital position. To address this question, several automated model selection criterions such as AIC. BIC, HQC, forward and backward selection criterion, autometrics etc. have been established. The problem of identifying the relevant variables for regression is still complicated to sort out. Test of independence can be used to check whether the variable is relevant in explaining the variation in regression analysis or not, so it can also be used as a model selection criterion. Independence test will check whether the two variables have some relation (true relation) or they are independent, so in making this decision we also decide whether this variable is a true determinant of the regressand or not.

Granger causality and cointegration tests remain less favored by the econometricians because of their complexity and wavered behavior. But still decision about the causality direction is always a key point for economists; however at different times it is believed that econometricians are capable of getting results of their own choice from these tests. The comparative study of the Granger causality, cointegration and independence tests in the common frame work using real data set is helpful in this regard. As we have discussed above, our best independence test can act like a model selection criterion, this will help in initial scrutiny for the advanced level econometric analysis.

1.4: Objectives of the Study

There are two main objectives of our study;

1. The primary objective of this study is to compare the different tests of independence for time series. We will frame a simulation study in which a common DGP (data generating process) is used. This DGP will provide firm basis to compare all the tests in the literature. In the methodology section there is a brief description of how this simulation study proceeds. In this simulation study, we use the stringency criterion of Zaman (1996), as this criterion is never used to compare the test of independence. The analysis is based on the shortcomings of each test. Stringency is used to find the maximum shortcomings (MSC) of each test; the test which has least MSC declared to be the best test.

   Further, as we mentioned above, test of independence can be used as model selection criterion e.g. AIC, BIC and others, so by providing a superlative test in this study we suggest a better model selection criterion. Our primary object can be categorized in supplementary bifurcations as;

i) In macroeconomics, most time series are unit root or highly nonstationary and our DGP is primarily meant to coup up this issue. We introduced AR coefficients for two series in our DGP and analyze the behavior of each test of independence for different combinations of AR coefficients. This analysis provides the basis to predict the behavior of each test of independence under different stationarity conditions. In this regard, we first ensure the size stability of each test, and then evaluate each test on the basis of power comparison using MSC.

ii) Atiq-ur-Rehman and Zaman (2008) describe the prominence of deterministic part as "Performance of unit tests depends on several specification decisions prior to their application e.g., whether or not to include a deterministic trend". Similarly, Ahking (2002)identified the importance of model specification as "The inclusion or exclusion of the linear deterministic time trend should be justified formally in cointegration analysis. Otherwise, the results may be misleading". Accepting the importance of deterministic part, we are integrating it in DGP. Tests of independence compared on the basis (presence or absence) of deterministic part help to conclude the most powerful test.

2. After the verdict about the best test of independence for time series which has least maximum shortcomings, next major aim of this study is to equate the top best tests of independence with Granger causality and cointegration tests by an empirical study. The main intention to conduct this empirical study is to see the shortcomings of advanced level Econometric techniques to evaluate the time series and a basic technique of test for independence. Actually the Granger

causality is also based on the dependence of one variable on the other and the dependence is measured by the usual way of regressing one variable and its lags on another variable. We characterized our second primary objective in these three modules;

i) This study becomes the first extensive comparison of tests of independence for time series. Top best recognized tests are used for a comprehensive empirical analysis. For this purpose we use the Keynes' consumption function, which is a well-known and empirically proven phenomena. Income (Y) and consumption (C) of 100 countries are taken for this purpose. Size of the tests is calculated by rejecting the null of independence for income and consumption of same countries. Similarly power of the tests is determined by rejecting the null of independence for income and consumption of two different countries.

ii) In this empirical study, famous Granger causality test is also analyzed using same data set of income and consumption of 100 countries. Size and power of the test is calculated by means of the above mentioned technique.

iii) Using the results of a study conducted by Khan (2017), declared best tests of cointegration are used in our empirical analysis. The study includes almost all of the available tests in literate for cointegration, whether with null of cointegration or null of no cointegration. Size and power is calculated for five best tests declared in the Khan (2017) study.

CHAPTER 2: LITERATURE REVIEW

Although independence in time series is a very important issue, unfortunately little research effort has been devoted to the problem. As we mentioned in the previous chapter, numerous complicated and faltered model selection criterion are introduced, but very little effort is given to present a trustworthy test of independence. Several tests are presented by scientists from time to time, however a comprehensive study was never conducted to see the weaknesses and strengths of these tests.

2.1: Brief Overview

There is a unanimous agreement by all the econometricians in our knowledge that Haugh (1976)paper is the first attempt at developing a procedure based on residual cross-correlations for examining the independence of two stationary ARMA series. It is a two-stage procedure. At first place it fits univariate ARMA models to each of the series separately and then, examines the dependence of the series by means of the two resulting residual series. The author showed that under the null of independence between the original series, the residual series are asymptotically multivariate normal, with zero mean and covariance matrix as identity. This result is the basic principal to the definition of a portmanteau test which rejects the null hypothesis of non-correlation at given significance level.

The first critique on Haugh test presented by Pierce (1977), which detected low power against some alternatives appearing in the econometric context. His finding got strengthened by another extensive study conducted by Geweke (1981). These two studies provided incentive for other scientists for further development in this field.

In order to fill this gap in the literature, Koch and Yang (1986) proposed an alternative approach that accounts for such possible patterns. On the basis of a Monte Carlo study, Koch and Yang concluded that their test is preferable to Haugh against a wide range of alternatives. We have to keep in mind that this study is based on only Koch and Yang's test and Monte Carlo study have DGP which favors their own test. So these finding cannot be categorized as a comparative study, as these two tests were not compared in the common frame work.

Robinson (1991), Chan and Tran (1992) and Skaug and Tjøstheim (1993) offered nonparametric tests of independence, but none acquired great appreciation. Meanwhile a robustified version of Haugh statistic to outliers is described by Li and Hui (1994). Hong (1996b)statistic is a suitably standardized version of the sum of weighted squared residual cross-correlations, with weights defined by a kernel function. The test has a modified version of Haugh procedure for stationary infinite order autoregressive series. In this test, a finite-order autoregression is fitted to each series first and then the weighted squared residuals ore calculated using kernel function. With truncated uniform kernel, Haugh test can be a special case of Hong's test. Daniell kernel recognized best, as Hong shows that the power of his test maximizes using Daniell kernel under some particular local and fixed alternatives. Alike Haugh, Hong's test is also considered sensitive to outliers.

Furthermore, the usual tests which are based on cross-correlation function can be affected considerably by outliers.

Kim and Lee (2005) designed the independence test for two stationary infinite order AR processes. The empirical process method followed and produced the Cramer-von Mises type test statistics based on the least squares residuals. The behavior of the test showed that asymptotically the results are same as those tests which are based on true errors.

Hallin and Saidi (2005) develop a test statistic which was based on the definitions of Haugh (1976)and Himdi and Roy (1997), which states that under the null of non-correlation between the two series, there may be an arbitrary vector of residual cross-correlations which asymptotically follows a multivariate normal distribution. This test takes into account a possible pattern in the signs of cross-correlations at different lags. Actually Hallin and Saidi (2005) test is the multivariate generalization of the Koch and Yang (1986) procedure. Furthermore, El Himdi and Roy's technique was extended by Pham, Roy, and Cédras (2003) to partially nonstationary (cointegrated) VARMA series.

In first decade of 21st century, substantial amount of work was dedicated to testing the independence of two stationary time series, e.g. Bouhaddioui and Roy (2006), Duchesne and Roy (2003a), Eichler (2008), Hallin and Saidi (2005) and Hallin and Saidi (2007). Despite the fact that a large number of tests were introduced during this era, no one bothered to take into account the legitimacy of these tests using a comprehensive and exclusive comparison study. These tests are based purely on ad-hoc arrangements and authors checked the applicability of these tests in certain domain.

In literature we have seen that by prewhitening the two time series separately we obtain the innovation series, this will happen if we fit a parametric model to the individual series. For example Haugh (1976) used ARMA model to each time series. We discussed above that Himdi and Roy (1997) and Pham et al. (2003) are the multivariate extension of Haugh (1976) idea. However, theoretically the asymptotic null distribution of the test statistic is valid if the parametric models are correctly specified. To insure the correct model specifications, a long AR can be fitted to each time series. Under the null of no correlation, Bouhaddioui and Roy (2006) derived the asymptotic distribution of an arbitrary vector of residual cross-correlations. On the basis of these derived results they propose test statistics for the alternative of serial cross-correlation at a particular lag j or at a fixed number of lags j. Bouhaddioui and Roy extended the work of Hong for a multivariate case here.

Matilla-García, Rodríguez, and Marín (2010), proposed a test of independence for two time processes. They translated the problem into symbols using an easy symbolization technique and constructed a test of independence on the basis of entropy measure related with these symbols. This test is a novel nonparametric test for cross-independence and gives better result for a broad range of dependent alternatives. By the reason of model free test it does not require restrictive assumptions; secondly, it avoids the misspecification problem of many parametric tests which are based on the estimation of the true linear representation (say ARMA) of the data set.

A distinctive measure of association between two time series was proposed by Rehman and Malik (2014). This statistics is the correlation between recursive forecasts errors of autoregressive models fitted to both the series. The basic aim to develop this statistic is that its distribution is invariant to the change of IID assumption. The desired result are achieved by building a model which can approximate any of the dynamic structure. The obtained innovation series closely resembles to the true ones. He showed that when correlation between these innovation series is calculated, the results become more consistent.

Shao (2009)derived the test of independence for two long memory time series, which is the frequency domain analogues of Hong's test statistic. Eichler has recently reformulated the frequency domain. However, theoretically he developed procedure under stringent conditions. Shao's proposed statistics are derived under mild conditions to find the asymptotic standard normal distributions.

2.2: Tests to Be Compared

Although several tests are developed during the course of almost forty years. Very few comparative studies exists, according to our knowledge, which are conducted to assess the validity of these tests. Furthermore, the available comparative studies do not cover the entire literature but compare the tests of its own nature only. This shortcoming create breach in the literature and an exclusive study should be needed to fill this gap. A simulation base study to assess the strength and weaknesses of tests of independence for time series has been conducted. The underlying study covers all the tests of independence available in literature compared under the common DGP on several alternatives. Brief summary of independence tests for time series are discussed in this section.

2.2.1. Haugh (1976) Test

Haugh test was the first test which developed a procedure to check the independence of two stationary ARMA series based on residual cross-correlations. Haugh proposed a two stage procedure. First, he gets residuals from each series by fitting ARMA models to each of the series and then, finding cross-correlations between the two resulting residual series. Let $r_{uv}(j)$ be the residual cross-correlation at lag j:

$$r_{uv}(j) = \frac{\sum_{t=j+1}^{n} u_t v_{t-j}}{(\sum_{t=1}^{n} u_t^2 \sum_{t=1}^{n} v_t^2)^{1/2}} \qquad \text{For } |j| \leq n-1,$$

Where u_t, v_t, are the two residual series and n is the length of each time series t = 1, . . . , n.

The asymptotic distribution of a fixed number of residual cross-correlations is given by Haugh. He considered a portmanteau statistic which is based on the first M residual cross-correlations, where $M \leq n-1$ is a fixed integer. More precisely, he introduced the statistic

$$S_M = n \sum_{j=-M}^{M} r^2_{uv}(j)$$

Its asymptotic distribution is chi-square under the null hypothesis of independence, and the hypothesis is rejected for large values of the test statistic.

2.2.2. Koch and Yang (1986) Test

Koch and Yang introduced a modification of Haugh S_M statistic that absorbs potential pattern in the residual cross-correlation function. It also incorporates information about their magnitudes, in such a manner that the null of independence can be rejected when appropriate. This type of test should prove more powerful than the Haugh procedure for certain alternatives. The proposed test statistic is as follows;

$$r* = n \sum_{j=-M}^{M-i} \left[\sum_{l=0}^{i} r_{uv}(j+l) \right]^2$$

Where $i=0,1,2,\dots\dots,M-1$

$i=0$ is the special case which is Haugh statistic. To incorporate information about the covariation between estimated cross-correlation coefficients one lag apart, they need to include a term, $r_{uv}(k+1)$. This is achieved by the statistic in with $(i = 1)$, similarly for $i=2$, $r_{uv}(k+1)$ *and* $r_{uv}(k+2)$ terms can be integrated. In general, this process will employ cross-correlation coefficients up to *i lags*, as i is varied from zero to M-1.

Certainly, we can see from the above statistic that, if consecutive cross- correlations are on the same side of (say, greater than or less than) zero, this technique places more weight on successive coefficients. While the Haugh test overlooks this information.

2.2.3. Li and Hui (1994) Test

Li and Hui studied the "Robust residual cross correlation tests for lagged relations in time series". They proposed the following robust cross-correlation function:

$$\Upsilon_{uv}(j;\acute{\eta}) = \begin{cases} n^{-1} \sum_{t=j+1}^{n} \acute{\eta}(\frac{u_t}{\sigma_u}, v_{t-j}/\sigma_v), j \geq 0, \\ n^{-1} \sum_{t=-j+1}^{n} \acute{\eta}(\frac{u_{t+j}}{\sigma_u}, v_t/\sigma_v), j < 0, \end{cases}$$

Where u_t and v_t are the two innovation processes, with variances σ^2_u and σ^2_v respectively. When $\acute{\eta}(u, v) = uv$, we retrieve the usual sample cross-correlation function, and we write $\Upsilon_{uv}(j;\acute{\eta}) = r_{uv}(j)$. Similarly, a robust residual cross-correlation function is obtained by replacing $\{u_t\}$ and $\{v_t\}$ by the residual series $\hat{u}_t$ *and* $\hat{v}_t$. Li and Hui (1994) studied the asymptotic properties of a fixed length vector of robust residual cross-correlations.

A computationally simple approximation follows: r_i^* is approximately distributed as β_i $\chi^2 v_i$, here the degrees of freedom, v_i, are usually fraction.

2.2.4. Hong (1996c) Test

Hong incorporated all possible lags of the cross correlations for the innovation series. His proposed test is the generalization of Haugh statistic, as if we restrict this to the first lag only these two statistics become equallent. Hong's test statistic is a weighted sum of residual cross-correlations of the form;

$$Q_n = \frac{n\sum_{j=1-n}^{n-1} k^2\left(\frac{j}{m}\right) r^2_{uv}(j) - M_n(k)}{[2V_n(k)]^{1/2}}$$

Where $M_n(k) = \sum_{j=1-n}^{n-1}(1 - |j|/n)k^2\left(\frac{j}{m}\right)$ and

$V_n(k) = \sum_{j=2-n}^{n-2}(1 - |j|/n)(1 - (|j| + 1)/n)k^4\left(\frac{j}{m}\right)$

The 'k' is weighting kernel function. Using the truncated uniform kernel, Hong statistic Q_n resembles to a standardized version of Haugh statistic. However author believed that there are several kernels by which he can attain higher power than the truncated uniform kernel. In this article, he showed that the asymptotic power of his test maximized when he used the Daniell kernel under some local alternatives. Also, Hong (1996a) used parametric AR model to predict the residuals then Haugh ARMA modeling. Under the null of independence, the test statistic Q_n asymptotically follows the standard normal distribution. The test is one-sided and is rejected for the large values of Q_n.

2.2.5. Hallin and Saidi (2005) Test

Portmanteau type (like Haugh and El Himdi and Roy) tests have same weakness as these are based on unweighted sums of squared quantities. They completely ignored the patterns of signs that may provide strong evidence against the null hypothesis. Koch and Yang, is sensitive to such patterns. Hallin and Saidi have proposed a generalization of Koch and Yang procedure for VARMA series.

The test statistic Q_{HR}^{M} certainly is of the form$v_M v_M^T$, where

$$v_M := \sqrt{N} I_{2M+1} \otimes \left(R_{\hat{\eta}}^{(22)}(0) \otimes R_{\hat{\eta}}^{(11)}(0)\right)^{-1/2} r_{\hat{\eta}}^{(12)}(M)$$

Where v^Munder the null follows asymptotically N(0, I). This test statistic is built on the basis of the same idea as Koch and Yang's (1986) test. It also incorporates the patterns of signs as well as magnitude.

$$Q_{*i}^{M} := \sum_{k=1}^{(2M+1)d1d2-i} \left[\sum_{l=0}^{i} v_M(k+l)\right]^2, i = 0,1, \ldots, Md1d2 - 1,$$

Where $v_M(j)$ is the j[th] element of the vectorv_M. Note that, the test statistic become El Himdi and Roy's for i=0, while, for d1=d2=1, it matches with Koch and Yang's univariate test statistic.

2.2.6. Pham et al. (2003) Test

Pham et. al. presented the generalization of Haugh (1976) test for multivariate cointegrated case to assess the existence of relationship between two time series. The key contribution of authors is to show that when we are dealing with two uncorrelated cointegrating time series, the asymptotic distribution of the vector of cross-correlation matrices between the two residual series is same as an arbitrary vector of residual cross-correlation matrices. Derivation of asymptotic distribution for the test statistic, results of the article depicts that it follows chi-square distribution.

Pham et. al. offered two types of test statistics. In the first test statistic individual lags of cross-correlations are measured. Test statistic is given as;

$$QH(k) = (n-p)vecR_{\hat{a}}{}^{(12)}(k)[\rho_a{}^{(11)}(0)^{-1} \otimes \rho_a{}^{(22)}(0)^{-1}]vecR_{\hat{a}}{}^{(12)}(k)$$

Under the null of no correlation, *QH(k)* is asymptotically chi square distributed. The null hypothesis is rejected at a significance level α if$QH(k) > \chi^2{}_{d1d2,1-\alpha}$.

Now for the study of several lags at the same time *(say –M to M)*, another test statistic is given which rejects the null hypothesis of independence if for at least one lag the test statistic shows significance, i.e. $k \in \{-M,\dots,M\}, QH(k) > \chi^2{}_{d1d2,1-\alpha}$.

Test statistics for evaluating relationship at several lags at same time is given as under;

$$QH_M = (n-p)(r^{(12)}{}_{\hat{a}})'\left[I_{2M+1} \otimes \hat{\rho}_a{}^{(22)}(0) \otimes \hat{\rho}_a{}^{(11)}(0)\right]^{-1}(r^{(12)}{}_{\hat{a}}) = \sum_{k=-M}^{M} QH(k)$$

Hence QH_M is asymptotically distributed as $\chi^2{}_{(2M+1)d1d2}$.

2.2.7. Kim and Lee (2005)Test

Kim and Lee objected Haugh and other researchers in the way that "their procedure requires the dependence structure of given time series in a correct manner, and any model selection procedures require a step for diagnostics to set up a true model. However, the cross correlation method merely guarantees the uncorrelatedness of observations and does not ensure the independence. Moreover, the cross correlation just checks the linear relationship but cannot find a nonlinear dependence".

Therefore, an alternate of these procedures is presented, they use empirical process method devised by Hoeffding (1948) and Blum, Kiefer, and Rosenblatt (1961). Their proposed test statistics is of Cramer-von Mises (CV) statistics category, and can be used in different conditions. They gave attention to the case of two stationary infinite order AR processes for evaluating independence. So in these lines they used ARMA processes and developed a method which is based on residuals.

Test statistics for proposed test is given as;

$$B_{nk} = \begin{cases} (n-|k|)^{-1} \sum_{i=|k|+1}^{n} S^2{}_{nk}(\epsilon_{i-k}, \eta_i);\ k \geq 0 \\ (n-|k|)^{-1} \sum_{i=|k|+1}^{n} S^2{}_{nk}\left(\epsilon_i, \eta_{i-|k|}\right);\ k < 0 \end{cases}$$

Where

$$S_{nk}(x,y) = \begin{cases} (n-|k|)^{-1} \sum_{t=|k|+1}^{n} I(\epsilon_{t-k} \le x) I(\eta_t \le y) \\ (n-|k|)^{-2} \sum_{t=|k|+1}^{n} I(\epsilon_{t-k} \le x) \sum_{t=|k|+1}^{n} I(\eta_t \le y) \,; k \ge 0, \\ (n-|k|)^{-1} \sum_{t=|k|+1}^{n} I(\epsilon_t \le x) I(\eta_{t-|k|} \le y) \\ (n-|k|)^{-2} \sum_{t=|k|+1}^{n} I(\epsilon_t \le x) \sum_{t=|k|+1}^{n} I(\eta_{t-|k|} \le y) \,; k < 0, \end{cases}$$

This statistic can be used for testing the independence for single explicit lag k. However, several lags can be tested for independence in real situations, so testing for the single lag may be not enough for some cases and several lags *say* "*K*" may be needed, they proposed three kind of alternatives in this regard. First one is the summation of the test statistics, secondly they suggested the weighted summation and thirdly, maximum of the test statistics. Although they proposed all three but preferred the third one.

2.2.8. Bouhaddioui and Roy (2006) Test

Bouhaddioui and Roy proposed a semiparametric approach for testing the independence between two infinite-order VAR series, this test is an extension of Hong's univariate results. Like Hong they also apply AR model and find residuals and then calculate weighted sum of quadratic forms in the residual cross-correlation matrices at all possible lags. These weights depend on a kernel function and on a truncation parameter. They also used several Kernel functions like truncated uniform, Bartlett, Bartlett and Priestley, Parzen and Daniell.

In multivariate time series, the squared cross-correlation $r_{\hat{a}}^{(12)}(j)^2$ is replaced by a quadratic form in the vector $r_{\hat{a}}^{(12)}(j)^2 = vec\left(R_{\hat{a}}^{(12)}(j)^2\right)$.The test is based on the following sum of weighted quadratic forms at all possible lags

$$\mathcal{T}(\hat{\alpha},\hat{\Sigma}) = \sum_{j=1-N}^{N-1} k^2\left(\frac{j}{M}\right) Q_{\hat{a}}(j)$$

Where

$$Q_{\hat{a}}(j) = N r_{\hat{a}}^{(12)}(j)^T\left(R_{\hat{a}}^{(22)}(0)^{-1} \otimes R_{\hat{a}}^{(11)}(0)^{-1}\right) r_{\hat{a}}^{(12)}(j),$$

and$k(\cdot)$ is a kernel function. The test statistic is a standardized version of $\mathcal{T}(\hat{\alpha},\hat{\Sigma})$ given by

$$Q_N = \frac{\mathcal{T}(\hat{\alpha},\hat{\Sigma}) - m1m2S_N(k)}{\sqrt{2m1m2D_N(k)}}$$

where the truncation parameter $M = M(N) \to \infty$ and $M/_N \to 0$ $when$ $N \to \infty, and$

$$S_N(k) = \sum_{j=1-N}^{N-1} \left(\frac{N-|j|}{N}\right) k^2\left(\frac{j}{M}\right)$$

$$D_N(k) = \sum_{j=2-N}^{N-2} \left(\frac{N-|j|}{N}\right)\left(\frac{N-(|j|+1)}{N}\right) k^4\left(\frac{j}{M}\right)$$

Note that $S_N(k)$ and $D_N(k)$ are essentially the asymptotic mean and variance of $\mathcal{T}(\hat{\alpha},\hat{\Sigma})$ under H_0 respectively. If k is the truncated uniform kernel, apart from the standardization factors $S_N(k)$ and $D_N(k)$, Q_N corresponds to the multivariate version of Haugh statistic introduced in Himdi and Roy (1997) which is defined by

$$P_M = \sum_{j=-M}^{M} Q_{\hat{a}}(j)$$

In this case, M is a fixed integer which does not depend on the sample size N. The properties of PM in the $var(\infty)$ context are studied in Bouhaddioui and Roy (2006).

2.2.9. Bouhaddioui and Dufour (2008) Test

Bouhaddioui proposed test statistic based on cross correlation and partial cross correlations for testing independence between the residual series obtained by two nonstationary possibly cointegrated vector processes, in the general case where the processes have infinite-order autoregressive representations. The results revealed that, under the hypothesis of independence, residual cross-correlation and partial cross-correlation matrices follow the same asymptotic Gaussian distribution as the corresponding cross-correlation and partial cross-correlation matrices based on the true residuals.

Cross-correlation based test statistic presented as;

$$Q_{\hat{a}}(j) = Nvec\left(R_{\hat{a}}{}^{12}(j)\right)' (\hat{\rho}_2{}^{-1} \otimes \hat{\rho}_1{}^{-1})vec((R_{\hat{a}}{}^{12}(j))$$

Under the null of non-correlation, $Q_{\hat{a}}(j)$ is asymptotically distributed as $\chi^2{}_{m1m2}$. Thus, for a given significance level α, H_0 is rejected if $Q_{\hat{a}}(j) > \chi^2{}_{m1m2,1-\alpha}$.

The corresponding modified statistic $\tilde{Q}_{\hat{a},M}$ is defined by

$$\tilde{Q}_{\hat{a},M} = \sum_{j=-M}^{M} \tilde{Q}_{\hat{a}}(j) = \sum_{j=-M}^{M} \frac{N}{N-|j|} Q_{\hat{a}}(j)$$

2.2.10. Mariano (2010) Test

Non-parametric test for independence between two time series are presented by Mariano et. al. This test is based on symbolic dynamics and permutation entropy. In contrast to Haugh-type tests, it avoids autoregressive moving average (ARMA) pre-specification and kernel selection like Hong. The test is suitable for both linear and nonlinear processes.

To test the independence between *{Xt}* and *{Yt}*, following null hypothesis is considered:

$$Ho\colon \{X_t\}_{t\in I} \ and \ \{Y_t\}_{t\in I} are \ independent$$

Let we have two dimensional time series as; $Wt = (Xt, Yt)_{t\in I}$, here "*I*" belongs to the time domain.

Then relative frequencies for symbols can be calculated as;

For$\pi^x{}_i$ *and* $\pi^y{}_i$;

$$P(\pi^x{}_i) := P\pi^x{}_i = \frac{\{t\epsilon I | t \ is \ of \ \pi^x{}_i type\}}{T-m+1} = \frac{n_{\pi^x{}_i}}{T-m+1}$$

$$P(\pi^y{}_i) := P\pi^y{}_i = \frac{\{t\epsilon I | t \ is \ of \ \pi^y{}_i type\}}{T-m+1} = \frac{n_{\pi^y{}_i}}{T-m+1}$$

Relative frequency for η_{ij}can be obtained as;

$$P(\eta_{ij}) := P\eta_{ij} = \frac{\{t\epsilon I | t \ is \ of \ \eta_{ij} type\}}{T-m+1} = \frac{n_{\eta_{ij}}}{T-m+1}$$

Then under the null hypothesis of independence between series, the statistic

$$\Lambda(m) := -2\left[k ln(k) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} n_{\eta_{ij}} \ln\left(\frac{p_{\eta_{ij}}{}^{(0)}}{n_{\eta_{ij}}}\right)\right]$$

is asymptotically $\chi^2{}_{(m!-1)^2}$ distributed, where $p_{\eta_{ij}}{}^{(0)}$ refers to the probability of each symbol under the null.

Importantly, notice that $p_{\eta_{ij}}{}^{(0)}$ is unknown and hence fortunately, under the H_0 it has to be satisfied that

$$p_{\eta_{ij}}{}^{(0)} = p_{\pi^X{}_j\pi^Y{}_j}$$

In addition, note that

$$\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{k} = 1$$

And that

$$p_{\pi^X{}_j\pi^Y{}_j} = \frac{n_{\pi_i{}^X}}{k}\frac{n_{\pi_j{}^Y}}{k}$$

Where $k=T-m+1$.

2.2.11. Atiq (2014) Test

Test presented by Atiq et. al. claims that it is robust in all stationarity and deterministic conditions. For two time series of length T, let $T1$ be number smaller than T.

1. For$n \leq N$, estimate the autoregressive model$X_T = \hat{a}_{T1} + \hat{b}_{T1}X_{T-1} + e_T$using ordinary least square method.

2. Estimate $\hat{X}_{T1+1} = \hat{a}_{T1} + \hat{b}_{T1} X_{T1}$.

3. Now calculate $\hat{e}_{T1+1} = X_{T1+1} - \hat{X}_{T1+1}$.

4. Now repeat this process for remaining values of "T" as: T_{1+1}, T_{1+2}, ...T-1 to compute $e_{T1+2}, e_{T1+3}, \ldots, e_T$.

5. As calculated above for first series now calculate recursive residuals for second series and call them $u_{T1+1}, u_{T1+2}, u_{T1+3}, \ldots, u_T$.

6. The final test statistic is simply the correlation between these two residual series.

2.3: Cointegration Tests

In Econometric literature the question of non-stationarity and spurious results in time series open new doors for researchers. Granger and Newbold (1974)challenged the reliability of econometric analysis due to the non-stationary behavior of time series. However, Engel and Granger give a way out of this situation by introducing a new procedure called cointegration. Many tests to assess the cointegration between time series were introduced during the last thirty years. These tests are designed for capturing different features of time series, so these tests have their own domain of strength.

We have found many studies which are conducted to show the strengths and faintness of these tests, however a recent study conducted by Khan (2017)is considered most reliable and extensive. Twenty-four tests of cointegration were simultaneously assessed in a common frame work using standard stringency criteria and a detailed simulation study was carried out. We have chosen five best declared tests with null of no cointegration from above said simulation study. These tests includes;

1. t Test of Cointegration in a Single Equation Error Correction Model (TECM)
2. Hansen's variation of the $\hat{Z}_{\alpha}$ test of Cointegration (HAZA)
3. Hansen's variation of the ADF test of Cointegration (HADF)
4. Phillips and Ouliaris'$\hat{Z}_{\alpha}$ test of Cointegration (POZA)
5. Phillips and Ouliaris' $\hat{P}_{\mu}$ test of Cointegration (POPU)

Khan (2017)concluded his study with remarks that among the tests having null of no cointegration, "Phillips and Ouliaris $\hat{P}_{\mu}$ test was the leading performer as it was the most stringent test at all specifications of deterministic part. In addition to this test Choi Durbin-Hausman test, Phillips and Ouliaris' $\hat{Z}_{t}$ test, Phillips and Ouliaris' $\hat{Z}_{\alpha}$ test and the t-test of cointegration in a single equation Error Correction Model also performed overall better than the rest of the tests".

Here in literature review section of cointegration tests we are not presenting the mathematics for construction of the tests, as we mentioned a comprehensive study from where we have taken the tests of cointegration for real data analysis.

2.4: Granger Causality Test

Causality testing between variables is most important issue in economics, however econometricians believe that causality cannot be determined by correlation. So researchers performs analysis for some selected variables where all other possible causes are kept constant, although this type of testing leads to another problem that selecting a different set of variables in the analysis, previously established results are changed.

As philosophers and social scientists believed that there is no answer for this query but they introduced different techniques to capture the effect of one variable on another. Granger (1969)introduced the foreseeability as a measure to capture these effects which is called Granger causality testing, as Granger believes that only past can cause the present and future and secondly past must have some unique information for the present and future to establish the relationship between two variables.

We have two null hypothesis in case of assessing causality between two time series (*sayX and Y), i.e.*

$$H_{0a}: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_{0b}: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

Now we can test the Granger causality using these two regressions as;

$$Y_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i Y_{t-i} + \sum_{j=1}^{k} \beta_j X_{t-j} + \mu_t$$

$$X_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i Y_{t-i} + \sum_{j=1}^{k} \beta_j X_{t-j} + e_t$$

There may be four possible outcomes listed below with possible explanation;

1. If H_{0a} is rejected and H_{0b} is accepted we can say that *X Granger cause Y.*
2. If H_{0a} is accepted and H_{0b} is rejected we can say that *Y Granger cause X.*
3. If H_{0a} and H_{0b} both are accepted we can say that *no way Granger causality.*
4. If H_{0a} and H_{0b} both are rejected we can say that *both way Granger causality.*

We applied this simple Granger causality test to see the authority of Granger causality test on real data. In next section we discuss the methodology to conduct real data analysis in detail. Real data analysis for income consumption hypothesis is carried out in chapter six.

CHAPTER 3: METHODOLOGY

In this chapter we will briefly describe the methodology for evaluating the tests of independence for time series in a simulation study using stringency criteria and also compare the cointegration and Granger causality test in a real data empirical study. First we describe the data generating process for our simulation study, secondly size and power calculation steps will be presented. Construction of point optimal (PO) will be discussed in detail. Stringency criteria which is used in this study will be incorporated in next section and at the end of the study we will elaborate the methodology for real data comparison.

3.1: Data Generating Process

Selection of DGP for simulation study is very crucial in most comparative analysis. Different tests have different theoretical base so selecting one DGP can give an extra benefit to some test statistic. As we are concerned to compare all of the tests, which are used to test the independence of two time series. These tests can be compared in homogeneous frame work to conclude the superiority of a single test or show the strength and weaknesses of the tests under some alternative. In this regard, a simulation study should be needed; however we are proceeding our study by selecting a DGP which is feasible to compare all the tests.

In order to compare different tests of independence, data will be generated by following bivariate autoregressive process

$$\binom{x_t}{y_t} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \binom{x_{t-1}}{y_{t-1}} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \binom{1}{t} + \binom{\epsilon_{xt}}{\epsilon_{yt}} \quad where \binom{\epsilon_{xt}}{\epsilon_{yt}} \sim N\left(\binom{0}{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \text{....Eq (3.1)}$$

if $\rho = 0$ then X_t and Y_t are not correlated but for $0 < \rho \leq 1$ two series will be correlated.

The above process can also be written as

$$Zt = AZ_{t-1} + BD + \epsilon \qquad \epsilon \sim N(0, \Sigma) \qquad \text{..............Equation (3.2)}$$

$$where\ A = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \quad B = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}, \quad D = \binom{1}{t}, \epsilon = \binom{\epsilon_{xt}}{\epsilon_{yt}}\ and\ \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Under the Null hypothesis of Independence between X and Y (H_0: $\rho=0$) and under alternative hypotheses of dependence between X and Y (H_1: $0 < \rho \leq 1$).

Here in this analysis we assume the window size of 0.1; i.e. our alternate space becomes;

$AS = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$

All of these comparisons will be done by using different values of A, B and Σ. The analysis is primarily based on Σ matrix;

$$here\ \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} where\ 0 \leq \rho \leq 1$$

Which shows the strength of dependence between the two considered series and all other parameters in the model i.e. A and B are the nuisance parameters for the mutual correlation of two series. This data generating process produces two autoregressive series.

$$where\ A = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$

Here "A" reflects the autoregressive or stochastic part (strength of autocorrelation in the two series).

If $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, this implies two series have no autocorrelation. $б_1$, indicates the AR coefficient of the first series and $б_2$illustrates AR coefficient for the second series. These coefficients can vary from zero to one.

if $б_1$=0, it means the first series under consideration is strongly stationary,

Similarly when $0 < б_1 < 1$, we can say that the first series is "mild stationary";

And if $б_1$=1, then the first series becomes unit root.

Likewise explanation will apply on "$б_2$ ", when we talk about the second series.

Now we have second nuisance parameter "B";

$$B = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix},$$

First column presents drift and second column indicates linear trend.

$if\ B = \begin{pmatrix} a_1 & 0 \\ b_1 & 0 \end{pmatrix}$,it implies only drift and no linear trend,

Similarly$B = \begin{pmatrix} 0 & a_2 \\ 0 & b_2 \end{pmatrix}$, illustrates only linear trend and absence of drift;

And$B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, means no deterministic part i.e. absence of drift as well as linear trend.

If both "A and B" becomes null matrix, i.e. $\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix}\ where \begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, then we can say that series would be IID.

3.1.1: Selection of Stochastic Part

Autoregressive or stochastic part of our DGP can play an important role in assessing the tests of independence in this simulation study. As we all know that time series have autoregressive behavior and these tests are designed to attain higher precision in nonstationary or exactly unit root cases.

"A" matrix from equation (3.2) describes the AR coefficients of two series, and as debated earlier these coefficients can attain values from zero to one. Fluctuating the value of one parameter, keeping the other parameter constant, can provide several combination. Choice of gap between these coefficients depends on the number of combinations for analysis and loss of information. So an appropriate opening of 0.2 is selected and different combinations of AR coefficients are used to generate the data. There are twenty-one cases of different combinations of AR coefficients. We started with AR coefficient as zero which means strongly stationary and increase it by 0.2 in each case.

Table 3.1 given below shows all twenty-one cases in which different combinations of AR coefficients for first and second series are depicted.

Table 3.1: All possible Combinations of AR Coefficients

Case #	$б_1$	$б_2$	**Case #**	$б_1$	$б_2$	**Case #**	$б_1$	$б_2$
1	0	0	**8**	0.2	0.4	**15**	0.4	1
2	0	0.2	**9**	0.2	0.6	**16**	0.6	0.6
3	0	0.4	**10**	0.2	0.8	**17**	0.6	0.8
4	0	0.6	**11**	0.2	1	**18**	0.6	1
5	0	0.8	**12**	0.4	0.4	**19**	0.8	0.8
6	0	1	**13**	0.4	0.6	**20**	0.8	1
7	0.2	0.2	**14**	0.4	0.8	**21**	1	1

3.1.2: Selection of Deterministic Part

Mostly parametric tests of independence for time series are based on the maximum likelihood method and it takes several days to simulate a single test with different combinations of AR coefficients. In equation (3.1) deterministic part (matrix B) is the second nuisance parameter, which can influence the performance of any test. There may be more than two options of deterministic part, but we incorporated only two of them as presence or absence of deterministic part due to shortage of time and resources.

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} б_1 & 0 \\ 0 & б_2 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix} + \begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \ where \begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \text{....Equation (3.3)}$$

The above equation shows the presence of deterministic part with variable stochastic part. Similarly the equation given below shows the absence of deterministic part with variable stochastic parameter.

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} б_1 & 0 \\ 0 & б_2 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \ where \begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \text{..........Equation (3.3a)}$$

3.1.3: Selection of Sample Size and MCSS

Selection of sample size is very important in analysis of any sort of data. Usually in macroeconomic time series we have annual data sets and these series usually contain no more than forty or fifty observations. We have selected three different sample sizes for this simulation study; i.e. 30, 60 and 120; as small, medium and large respectively.

Against each sample size we have twenty-one combinations (see table 3.1) for stochastic part and two options for deterministic part (see equation (3.3) and (3.3a)), so we have total forty-two simulations for each sample size. Using three sample sizes we have to simulate hundred and twenty-six times to get a result against each alternative. Similarly we have ten alternatives for assessing the power of any test, therefore we need one thousand, two hundred and sixty simulations to get the power analysis for a single test and moreover one hundred and twenty-six simulations for size examination of each test.

Monte Carlo sample size has a problem, when we use small MCSS some distortions may exist but unfortunately it is not possible for us to make MCSS large. Simulation of every test consumes several days to produce a single result, in these circumstances slight increase in MCSS affect the required time for each simulation. In order to keep the precision and time in mind, we have fair enough MCSS of 1000, which gives better approximation.

3.2: Monte Carlo Simulation Design for critical values of Tests

For the decision of rejection or acceptance of null hypothesis using simulated critical value we first have to find the critical values which are based on simulation for a particular test. Following are the steps involved in finding the required simulated critical values;

1. Generating data for a fixed sample size using DGP under null hypothesis.
2. Tests statistic is calculated for this generated data.
3. Repeat this process for a fixed number of times say Monte Carlo sample size and save the results each time in array.
4. Now if we have one sided test say right tailed test, which most of the tests are, then 95^{th} percentile of this array is our simulated critical value and if two sided test than 2.5^{th} and 97.5^{th} percentiles of this array are our lower and upper simulated critical values respectively.

3.3: Monte Carlo Simulation Design for Size of Tests

When a true hypothesis is rejected this is an error which is denoted by "α" and it constitutes to the size of the test.

$$\alpha = P(rejecting\ \ H_0\ /\ H_0\ \ is\ true)$$

In this study, size of a test is calculated following these steps;

1. Generate data a fixed number of time under null hypothesis (say n=30 or 60 or 120).
2. Apply test statistic and calculate the value for this test statistic.
3. Now compare the value of calculated test statistic with simulated critical value obtained in previous section using DGP under null, if significant count.
4. Repeat this process for fixed Monte Carlo sample size, and percentage of this count is called size of the test.

3.4: Point optimal (Neyman Pearson) Test

Neyman-Pearson lemma is used in Statistics for driving point optimal test when we test a hypothesis against a simple hypothesis;

i.e.$H_0: \theta = \theta_0$ and $\quad H_1: \theta = \theta_1$,

Multivariate normal density for "P" dimensions is given as;

$$f(x_{1,}x_{2,}x_{3,}x_{4,} \dots \dots, x_p) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\{-1/2\,(x-\boldsymbol{\mu})'|\Sigma|^{-1}(x-\boldsymbol{\mu})\},$$

$where\ x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\ and\ \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ are the p variables and p mue (mean) respectively;

Similarly

$$\Sigma = \begin{bmatrix} \delta_1{}^2 & \delta_{12} & \cdots & \delta_{1p} \\ \delta_{21} & \delta_2{}^2 & \delta_{23} & \delta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{p1} & \delta_{p2} & \cdots & \delta_p{}^2 \end{bmatrix}$$ are the variance covariance matrix for p dimensions, with variances on diagonal and covariances are at off diagonal.

Now likelihood function for the above expression is given as;

$$LF = \prod_{j=1}^{n} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\{-1/2\,(x-\boldsymbol{\mu})'|\Sigma|^{-1}(x-\boldsymbol{\mu})\}$$

And log likelihood function is given as;

$$LLH = -np\log(2\pi) - \frac{n}{2}\log|\Sigma| - \{n/2\,(x-\boldsymbol{\mu})'|\Sigma|^{-1}(x-\boldsymbol{\mu})\}$$

Now when we generate data under null for bivariate case "Σ" becomes;

$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and when it is under alternative hypothesis it is $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

When we find log likelihood ratio then first term of LLH vanishes, as it is constant.

Test statistic for likelihood ratio test is given below as;

$$\Lambda(x) = \frac{L(x/\theta_0)}{L(x/\theta_1)} \leq \eta \qquad \text{.........................Equation (3.4)}$$

Where $P\left(\Lambda(x) \leq \eta/_{H_0}\right) = \alpha$ ....Equation (3.5)

The statistic is called likelihood-ratio test which rejects null hypothesis when equation (3.4) holds. It is said to be point optimal (PO) test when it has maximum power against a single alternate.

We follow the following steps to calculate the point optimal test (PO) using Neyman-Pearson Lemma.

1. Generate data under null of independence, i.e. two independent series X_t and Y_t which are generated using standard normal distribution N(0,1).
2. Now generate two series with dependence equal to "**ρ**", i.e. first generate two independent standard normal series X_t and Y_t and then generate dependent series as;

   $X^*_t = \rho * X_t + (1 - \rho^2) * Y_t$ and $Y^*_t = Y_t$. These two series are now correlated and its correlation equals to Rho (**ρ**).
3. Now we have to find log likelihood under null when data is generated under null; i.e.

   $LLH_0(H_0) = 0.5 * (X^2{}_t + Y^2{}_t)$.

   Similarly log likelihood under null, when data is generated under alternative, can be calculated as;

   $LLH_0(H_1) = 0.5 * (X^{*2}{}_t + Y^{*2}{}_t)$.
4. For finding log likelihood under alternative, when data is generated under null, we used the following expression;

   $LLH_1(H_0) = 0.5 * log(1 - \rho^2) - 0.5 * \{(X^2{}_t + Y^2{}_t - 2 * \rho * X_t * Y_t)/ \rho\}$

   In the same way we can find log likelihood under alternative when data is generated under alternative hypothesis;

   $$LLH_1(H_1) = 0.5 * log(1 - \rho^2) - 0.5 * \{(X^{*2}{}_t + Y^{*2}{}_t - 2 * \rho * X^*{}_t * Y^*{}_t)/ \rho\}$$
5. Now we are interested in calculating log likelihood ratio for data under null hypothesis, i.e.

$$L\left(X/_{\theta_0}\right) = \frac{LLH_0(H_0)}{LLH_1(H_0)}$$

Likewise we can find log likelihood ratio for data under alternate hypothesis as;

$$L\left({}^{X}/_{\theta_1}\right) = \frac{LLH_0(H_1)}{LLH_1(H_1)}$$

6. In the end we find the Neyman Pearson test statistic described in equation (3.4) which is the ratio of two log likelihood ratios.

3.5: Monte Carlo Simulation Design for Power of Tests

Power of the test means rejecting a hypothesis when alternative hypothesis is true.

$$Power = P(rejecting\ \ H_0\ /\ H_1\ \ is\ true)$$

In our analysis we compared eleven different tests used to test the independence of the time series. All the tests are compared in a common framework so we can assess the ability to reject the null hypothesis when it is wrong.

Powers of the tests are calculated following these phases;

1. Generate data a fixed number of time (say n=30, 60 or 120) under any point mentioned in alternate space (AS).
2. Apply test statistic and calculate the test statistic for that particular alternate point.
3. Now compare the value of test statistic result obtained in 2 with critical value obtained in section 3.2 under corresponding alternate point, if significant count.
4. Repeat this process for fixed MCSS, percentage of these counts are called power of the test at a specific alternate point.
5. Replication for every point given in alternate space gives us power of the test against all possible alternates.

Debate about the DGP conducted in section 3.1; powers of the test should be considered against each alternate so we have overall ten alternates revealed in alternate space (AS).

We have one hundred and twenty-six simulations for each point of alternate space so powers of each test against all alternatives will be calculated using one thousand two hundred and sixty simulations. These simulations give us powers of the test using three sample sizes; small (30), medium (60) and large (120).

3.5.1: Power Curves

Powers of each test plotted against the elements of alternate space will give us power curve for a specific test. We have eleven tests to compare so eleven power curves will be obtained.

3.5.2: Approximation of Power Envelope

Calculating powers using log likelihood ratio test statistic against every element of the alternate space yields powers for the point optimal (PO) test. By plotting these powers against the corresponding point of alternate space gives power envelop.

3.6: Stringency Criteria

Problem of comparing statistical tests on the basis of its performance was addressed be Zaman (1996). Size and power are basically two important features in evaluating the performance of different tests. Different statistical tests can be compared on the basis of its powers while their size should be equal. Stringency criteria usually works under the condition that nominal size should be achieved first and then power envelops and power curves decide about the stringency of the tests. Before going to the formal definition of the stringency we will first elaborate the term shortcomings in next subsection.

3.6.1: Maximum Shortcomings

Comparison of the tests will be on the basis of their shortcomings. These shortcomings of each test can be calculated following these lines;

1. Say we have $\Phi_i^{\ j}$as the power of i^{th} test at j^{th} alternative.
2. Similarly $\pi^j$is the power of point optimal test at j^{th} alternative.
3. Now we have $\Omega_i^{\ j} = \pi^j - \Phi_i^{\ j}$as the shortcomings of the i^{th} test at j^{th} alternative.
4. Repeat this process for all alternatives and find shortcomings against each alternative. As in our case we have eleven tests to compare against ten alternatives so *i=1, 2, ……, 11, and j=1, 2, ……., 10.*
5. Now for maximum shortcomings we have $\Theta_i = \max(\Omega_i^{\ j})\ \forall j = 1, 2, \dots\dots, 10$.
6. For most stringent test we have the following expression as;

$$Most\ stringent\ test = \min(\Theta_i)\ \ \forall\ i = 1,2,\dots,10$$

We present formal definition of stringency; "If test T_1 has a smaller shortcoming than T_2 on all alternate points (and both have nominal size too), then we will say that T_1 is more stringent than T_2. If a test T_1 has a smaller shortcoming than all other tests of the same level, we will say that T_1 is the most stringent test of level α."

3.7: Comparing Tests of Independence with Granger Causality and Cointegration Test using Real Data

Test of independence for time series can be used as a variable selection criteria for modeling; this will happen only when it is empirically proven that tests of independence have relatively greater power than other advanced econometric techniques. In this regard a real data analysis is designed to assess the statistical power of the tests of independence and some advanced econometric procedures which are practiced in econometric literature nowadays.

3.7.1: Keynesian Consumption Function

For real data analysis we first have to choose a well-established economic function, so we selected almost eighty years old economic function known as "Keynesian consumption function". It was presented JOHN MAYNARD (1936) in his famous book "*The General Theory of Employment, Interest and Money*". Keynes describes the relationship between income and consumption as;

$$C = f(Y_d)$$

Where Y_d is the disposable income.

This phenomena became very popular and acceptable in economical literature. According to Keynes consumption is a function of income. Acceptance of the function is primarily the basis of our real data analysis.

3.7.2: Real Data Analysis for Tests of Independence for Time series

In simulation study presented in chapter 4 and 5, eleven tests of independence for time series are contested, however two tests are considered best as Atiq test performed better in small sample size and PhamRoy test performed well in medium and large sample size. Furthermore Haugh test reflected as a mediocre test, as no other test performed close to it. Therefore in real data analysis we selected the top best and mediocre tests to study the behavior of tests of independence.

Null and alternate hypothesis are given as;

$$H_0: C_i \neq f(Y_j) or\ [consumption_i\ and\ Income_j\ are\ independent]$$

When $i=j$, consumption of country is the function of same country's income and this null is assumed to be false hypothesis according to the economic literature. Similarly when $i \neq j$, then this hypothesis is true as income of one country and consumption of another country cannot be logically dependent.

3.7.3: Real Data Analysis for Cointegration Tests

Recent simulation study conducted by Khan (2017), using same stringency criteria to check the reliability of cointegration tests has been used to select the top best tests. Five best tests which performed better in his study has been selected for real data analysis here. These includes;

1. *Hansen's variation of the ADF test of Cointegration (HADF)*
2. *Hansen's variation of the $\hat{Z}_\alpha$ test of Cointegration (HAZA)*
3. *Phillips and Ouliaris $\hat{P}_\mu$ test of Cointegration (POPU)*
4. *Phillips and Ouliaris '$\hat{Z}_\alpha$test of Cointegration (POZA)*
5. *t Test of Cointegration in a Single Equation Error Correction Model (TECM)*

Null hypothesis for these tests are as under;

$$H_0: There\ is\ no\ cointegration\ between\ consumtion_i\ and\ income_j$$

True for, $\forall i \neq j$.

3.7.4: Real Data Analysis for Granger Causality Test

Granger causality test claims to detect the correlation plus direction of the causality. However this claim is also challenged in this empirical study. The two null hypothesis for Granger causality are claimed here as;

$$H_0: Y_i does\ not\ Granger\ cause\ C_j$$

And

$$H'_0: C_i does\ not\ Granger\ cause\ Y_j$$

In first hypothesis it is claimed that income does not Granger cause consumption, if this null is rejected for $\forall i=j$, it is claimed as statistical power of the test.

3.7.5: Size Distortions and Power Gains

Size and power of each test is calculated using above mentioned standard theory. Empirical analysis is based on two main observations which are calculated as under;

$$size\ distortion = Empirical\ size\ of\ the\ test - nominal\ size\ (5\%)$$

$$power\ Gain = Empirical\ power\ of\ the\ test - Empirical\ size\ of\ the\ test$$

3.7.6: Data Source

In Keynesian consumption function we have two variables; i.e. Income Y and consumption C. we selected data for hundred countries from the time period of 1970 to 2014 from online WDI (2014) data bank. This comprises of 45 observations after cleaning and adjusting for the missing data. These observations are used in our empirical analysis.

CHAPTER 4: SIZE EVALUATION

As debated in the preceding chapter, tests of independence for time series are only comparable when these tests have nominal or equal size. In this chapter we calculate the size of all the tests included in our analysis. Simulated critical values are used instead of asymptotic ones. Why we do this? In the next section, it is discussed in detail. Data is generated under null hypothesis and three sample sizes; i.e. small (30), medium (60) and large (120). Monte Carlo sample size for investigation is 1000. In this simulation study nominal size is 5%; i.e. empirical size of each test calculated on the basis of simulated critical values of 5%. Our investigation is based on two nuisance parameters, first one being deterministic and second stochastic. Deterministic part incorporated as absence and presence of deterministic part; it has two components a drift and linear trend. In stochastic part we have different combinations of AR coefficients to generate the data. In table 3.1 these twenty-one cases of different combinations of AR coefficients are discussed in detail.

4.1: Why Simulated Critical Value?

In Econometrics it is widely believed that simulated critical values gives less size distortions than asymptotic critical values while performing a simulation study. Such phenomena holds because asymptotic critical values can be used with a specific sample size; i.e.Kim and Lee (2005) reveal that using asymptotic critical values, size distortions are very high especially in small samples. They used two sample sizes 100 and 300, small and high respectively. The results depict that using small sample size, all the tests have high size distortions but when they used large sample size, distortion minimizes.

Similarly El Himdi, Roy, and Duchesne (2003) also exposed the small sample distortions of asymptotic critical values. Using sample size of 100 and 200, they showed that size distortion of 200 sample size is far-off low relatively to sample size of 100.

Comparison for asymptotic critical values with a renowned bootstrap procedure was proposed by Bouhaddioui and Dufour (2008). In his study, both type critical values were used and the results suggested that in Monte Carlo study bootstrap procedure gives empirical size more close to the nominal size than asymptotic critical values. Shao (2009)also confirms that asymptotic critical values have higher size distortion than simulated critical values in Monte Carlo simulation study, especially in small sample.

The only comparative study of Duchesne and Roy (2003a) (which we have seen in the literature of comparison of tests of independence for time series), also showed that while comparing Haugh, Hong and Duchesne and Roy tests mostly empirical size is too higher than nominal size while using asymptotic critical values. Recent study of Robbins and Fisher (2015) also reported the oversize of the tests in a small simulation study using asymptotic critical values.

In section 3.2 we already discussed that simulation design is too much lengthy and takes one hundred and twenty six simulations to get the empirical size for a single test against three proposed sample sizes. We have number of examples presented in this section that why we are not using asymptotic critical values in this simulation study. In next two sections we present the empirical size of the tests of independence for time series using simulated critical values. These sections are distributed with respect to the presence of deterministic part and absence of it, whereas stochastic part varies in both sections. At the end of this section a brief summary for this chapter will be presented.

4.2: Size Comparison without Deterministic Part

We are going to discuss the comparison of size first with respect to change in stochastic part of our DGP, keeping the deterministic part absent. In our simulation study we fixed level of significance to 5%, therefore size of each test should be around 5%. In table 3.1 of methodology chapter we listed all possible combinations of stochastic part for two series. There are twenty-one cases enumerated; we discuss five important cases in this section and remaining graphs are incorporated in the appendix.

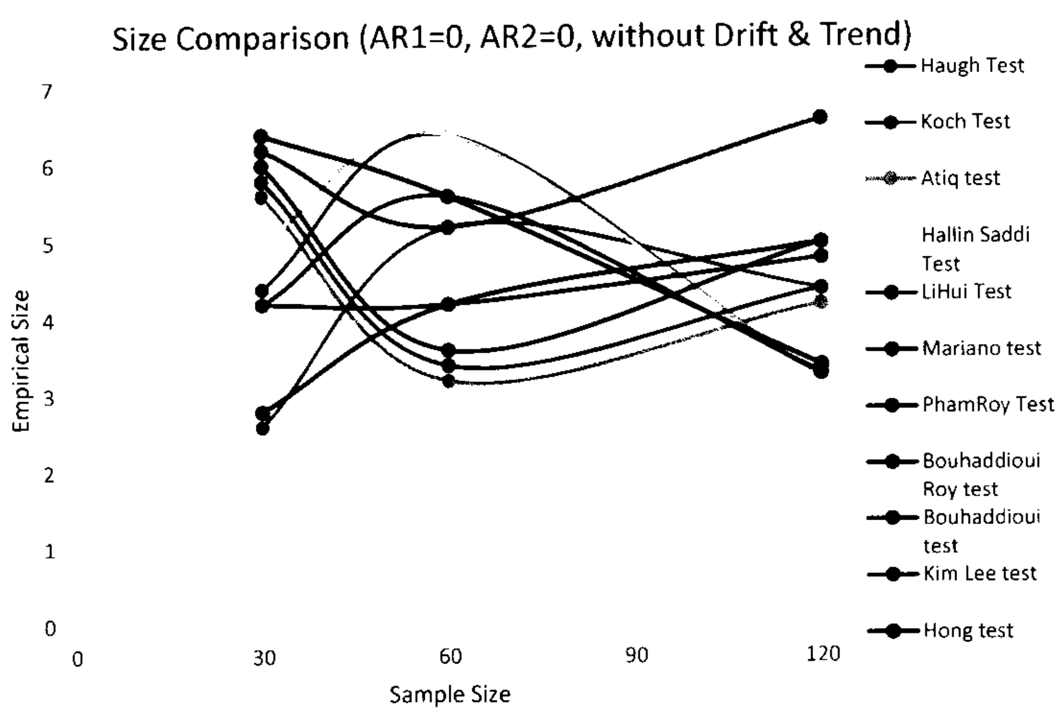
These five cases are listed below;

1. When both series are strongly stationary.
2. When one series is strongly stationary and second series mild stationary.
3. When one series is strongly stationary and second series is unit root.
4. When both series are mild stationary.
5. When both series are unit root.

In Figure 4.2.1, size of tests of independence are depicted for three sample sizes of 30, 60 and 120 when both series are strongly stationary i.e. case no. 1 in table 3.1, by assuming absence of deterministic part. Figure 4.2.1 clearly shows that all of eleven tests have stable size around nominal size of 5%. However the size of Haugh and Bouhaddioui and Roy tests are lesser than 3% at sample size of 30, but with increase in sample size i.e. 60 and then 120 their size get stabilized around nominal size.

Similarly Hong, Hallin Saddi, Koch and Bouhaddioui tests have slightly greater than 6%. These small deviations appears as MCSS is 1000. Most of the tests are within band of 4% to 6%, which is quite fine for this MCSS. Although we mentioned the points where our tests show size distortion but still we can say that size of the tests are appropriate and power comparison is legitimate.

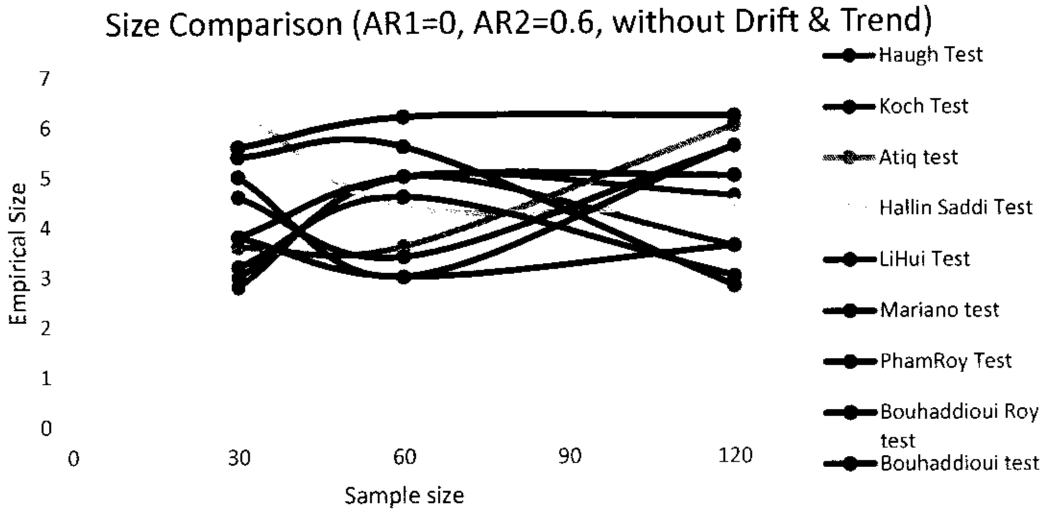
Figure 4.2. 1: When both series are strongly stationary



As was discussed earlier, we skip most of the cases from table 3.1; now we discuss fourth case of table 3.1, in figure 4.2.2, where the first series generated is strongly stationary (say AR coefficient = 0) and second series is mild stationary (AR coefficient = 0.6).

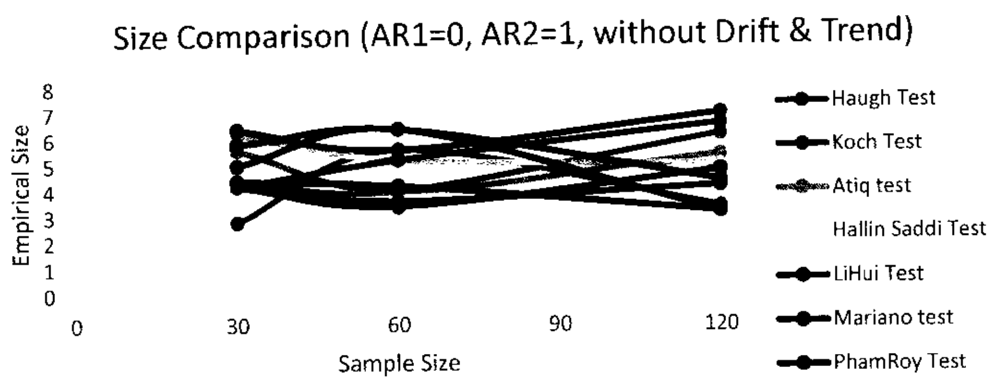
Slight departure does not reflect that size is distorted. Hong and Hallin Saddi tests have size slightly higher than 6% in few situations. Similarly Haugh and Bouhaddioui tests have size slightly less than 3% in couple of points. Overall size of the tests is fine.

***Figure 4.2. 2*: When one series is strongly stationary and second series mild stationary.**



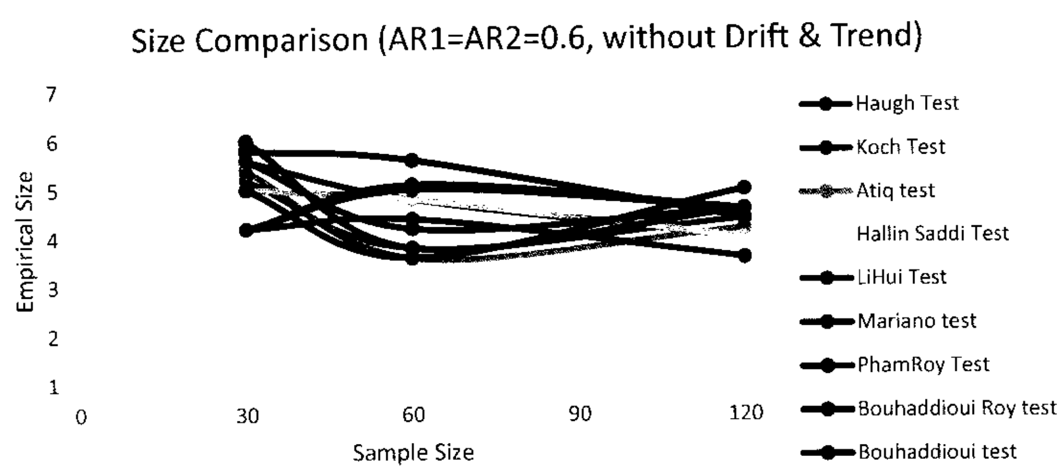
Combination of a stationary and a unit root series is discussed in the figure given below. Where Koch test have size less than 3% on one point and Haugh and Mariano tests have jumped slightly above 6% on couple of points. Where we haven't seen a single odd entry which is alarming. Size of the tests is nominal around 5% and slight departure is standard.

Figure 4.2. 3: **When one series is strongly stationary and second series is unit root.**



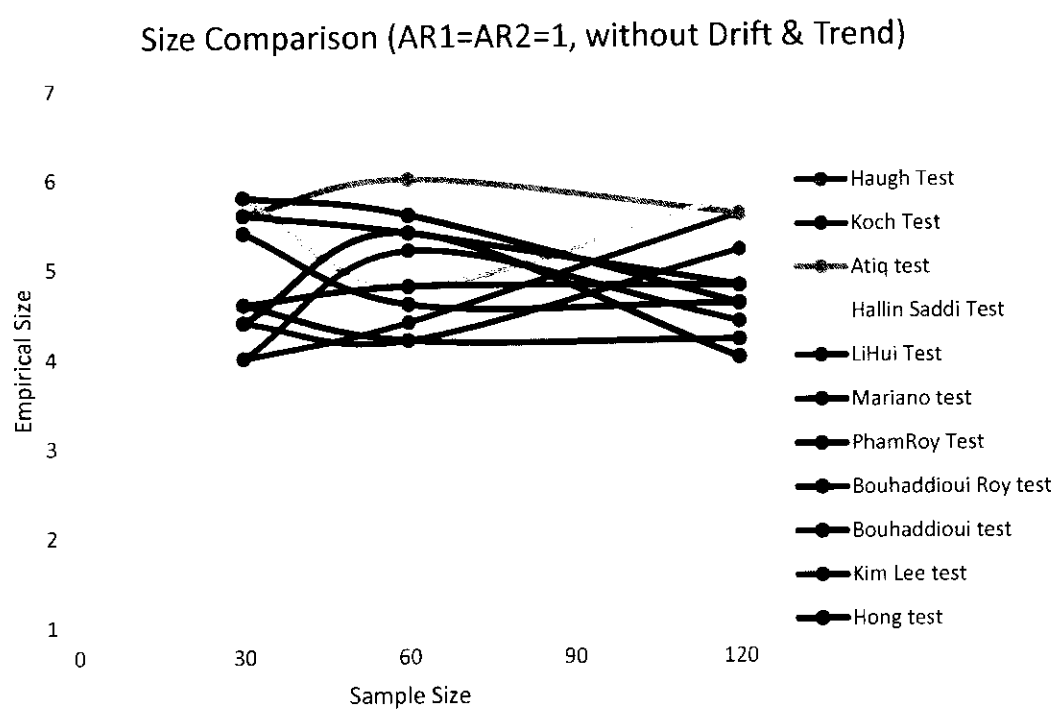
In figure 4.2.4 we have two mild stationary series (AR1=AR2=0.6) without deterministic parameters. There are eleven separate lines each represents the size of a particular test, each line consists of three points which are sample sizes 30, 60 and 120. Almost all the lines are within the band of 4 to 6, which means every test have nominal size of 5% using simulated critical value. Distortion in size is now more decreased when we are heading toward the unit root cases, as tests of independence for time series are more feasible for unit root series than stationary ones.

Figure 4.2. 4: **When both series are mild stationary.**



As we discussed above that tests of independence for time series are inevitable for unit root series and in figure 4.2.5, it is tangible that when both series are generated unit root then size distortions for the tests of independence are more negligible. Here all the lines are within the band of 4% to 6%. So the nominal size of all the tests is achieved.

***Figure 4.2. 5*: When both series are unit root.**

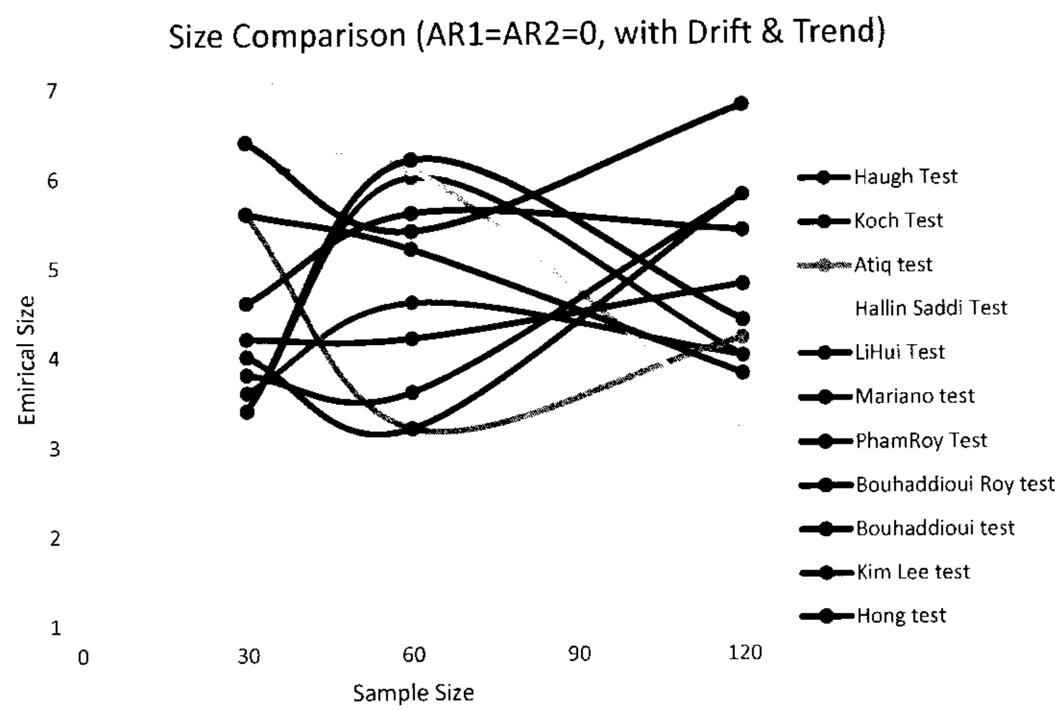


4.3: Size Comparison with Deterministic Part

Deterministic part of our DGP is the second imperative factor of this simulation study. In this section we are comparing the size of the tests of independence for time series in the presence of drift and linear trend. We have already discussed the results of change in stochastic part without incorporating drift and trend. Similar cases which are argued in the above section will be contended here with presence of drift and trend.

Figure 4.3.1 shows the size of the tests when data is generated with strong stationary conditions in the presence of drift and linear trend. Nominal size for each test is same as above (5%), and MCSS is 1000. Again the results are more or less equal to figure 4.2.1, where data is generated under same conditions without incorporating deterministic part. So we can say here that change of deterministic part may not have any noteworthy effect on the size of the tests. Not a single test have size below 3% however Hong, Hallin Saddi and Bouhaddioui and Roy tests have size between 6% to 7% at times. All the tests have nominal size of 5% with small distortions.

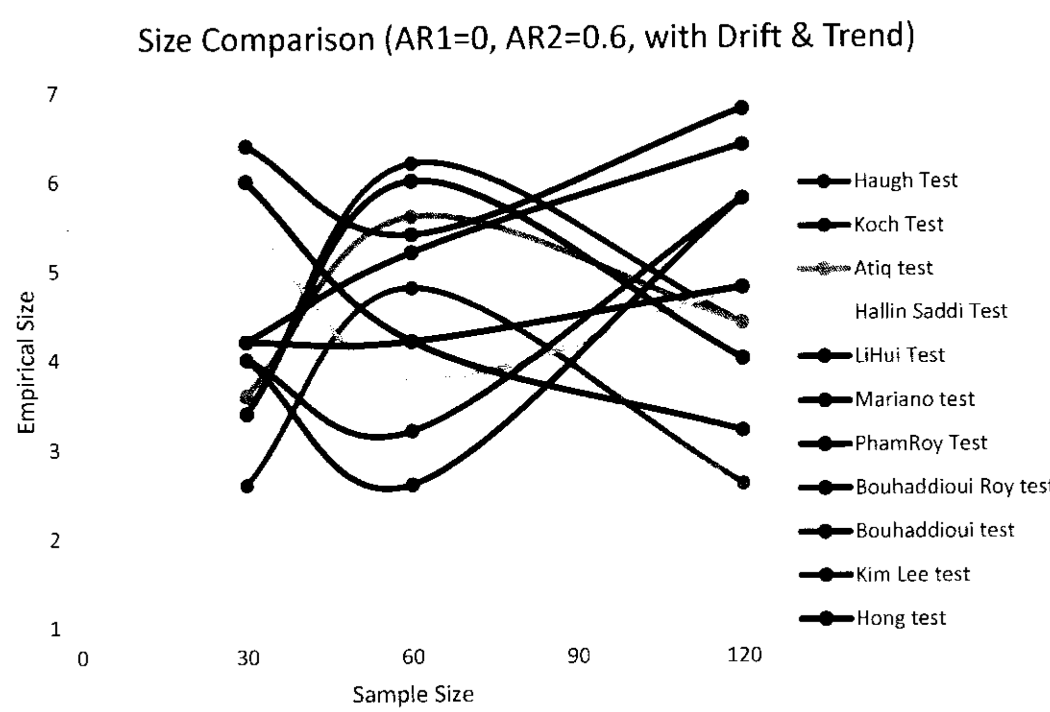
***Figure 4.3. 1*: When both series are strongly stationary.**



No consistent oversized or undersized test appears in our analysis, all the tests have slight departure from nominal size which is random. They appear in the band of 3% to 7%, which is quite fair.

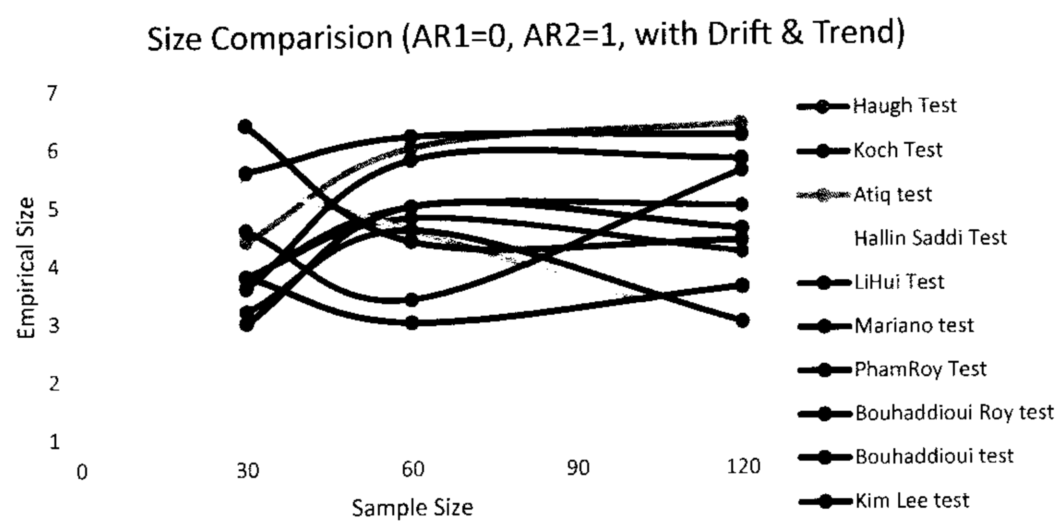
By comparing figure 4.2.2 with figure 4.3.2, where both have same stochastic part (AR1=0 and AR2= 0.6), the former figure does not have drift and trend however the later one contains deterministic part. Resemblance appears in both figures. Size of the tests is round about 5% when first series is stationary and second series is mild stationary, deterministic part does not situate any substantial effect here.

***Figure 4.3. 2*: When one series is strongly stationary and second series mild stationary.**



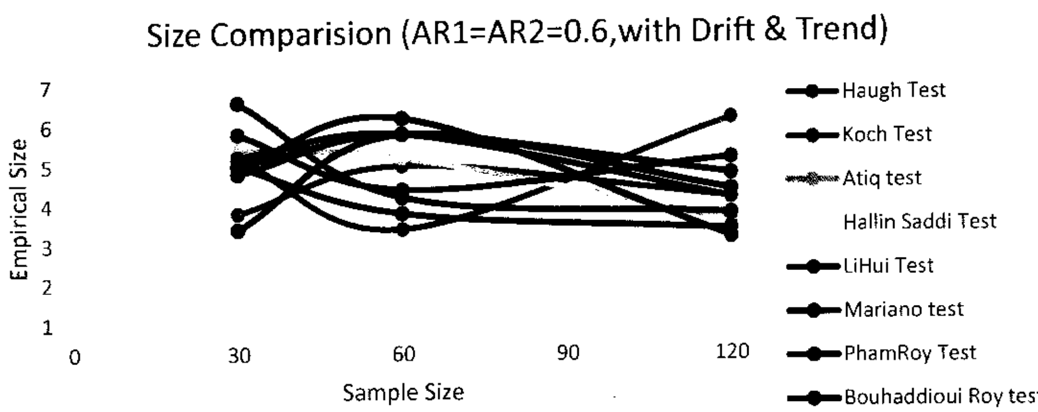
Graph available below is the size comparison of the tests when data is generated by the mixture of a stationary and a unit root series with drift and linear trend. Corresponding to this we have figure 4.2.3, in which we have same stochastic part but drift and trend is absent. Here, again size of all the tests is nearly 5% with small distortions which are negligible.

***Figure 4.3. 3*: When one series is strongly stationary and second series is unit root.**



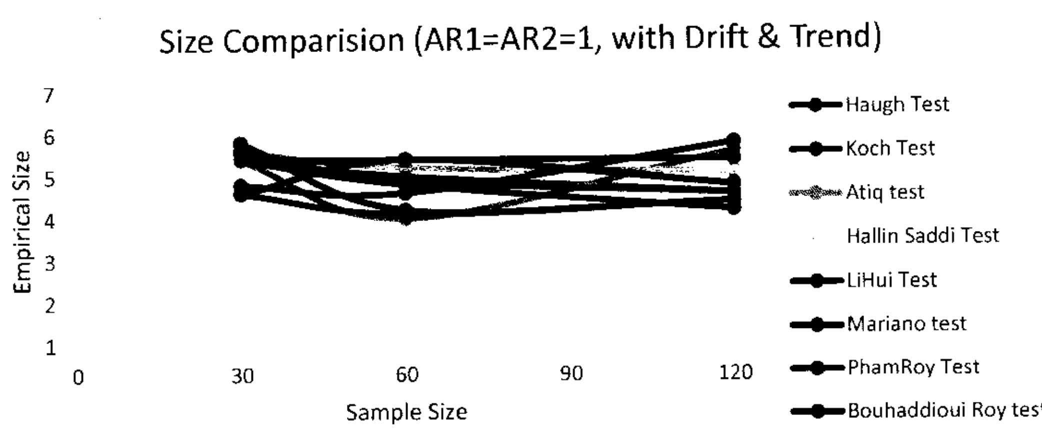
Now we have size of tests when two series are generated under mild stationary conditions with drift and trend in figure 4.3.4. As we discussed in the above section, when we move towards the unit root cases the distortion in size decreased, although this distortion is not so prominent. If we compare it with the same stochastic part and no deterministic part in figure 4.2.4, we notice that scenario seems not different, size of the tests is nearly same.

Figure 4.3. 4: **When both series are mild stationary.**



In figure 4.3.5, we have size of the tests when both series are unit root with drift and trend. Here we can see that distortion in size becomes more reduced and approaches to almost 5%. The same situation arises in figure 4.2.5, when we have same stochastic part but no drift and trend. Size of the tests is between 4% and 6%, and most of the times the size is close to 5%.

Figure 4.3. 5: **When both series are unit root.**



4.4: Conclusion

We have debated only five prominent cases here in this chapter for evaluation of the size stability and remaining graphs are assimilated in annexure. Results show that nominal size of 5% is attained in all twenty-one cases of variations in stochastic part. No clear pattern of a particular test is present in all these cases.

It is also evident that there is slight distortion in size of the tests when we are close to stationary condition but as we move towards the unit root cases the results are smoother. We do not claim that there is a clear pattern here but still it is evident from the results that when DGP is close to stationarity size distortions seem a bit high (not higher than ±2%), however when DGP is closer to unit root these distortions minimize.

Parallel to stochastic part our second nuisance parameter is deterministic part of DGP. Section 4.2 and 4.3 are bifurcated on the basis of this deterministic part. Results extracted from size comparison of deterministic part show that size stability is also same here. Size of the tests nearly 5% when drift and linear trend incorporated in the DGP.

Many specimens elaborated in section 4.1 that using simulated critical values size distortions become negligible as compared to the asymptotic critical values. We are running out of time and resources to show the results in both cases; as we mentioned above that empirical size calculation for a single test consumes one hundred and twenty-six calculations using all three sample sizes. Many of these tests are based on maximum likelihood function and take several hours in a single simulation. But still we achieved the target of stable empirical size which provides firm bases for the comparison of tests w.r.t. the power analysis.

As debated in section 3.5, using stringency criterion, tests of independence for time series are compared. For this purpose, size of the tests should be equal so shortcomings of the tests calculated by power of the tests should be compared. In the next chapter we can compare the power of the tests using stringency criterion.

CHAPPTER 5: POWER COMPARISION

It is determined in chapter 4 that size of all the tests compared in this study are around nominal size. Eleven tests of independence for time series are incorporated in this analysis. Size stability has been examined in Monte Carlo simulation study using simulated critical values, for three sample sizes i.e. small, medium and large. Few important cases are discussed comprehensively in preceding chapter while the other cases are merged in the appendix.

In section 3.5, it is discussed in detail that these tests are compared using stringency criteria, so we can assess the ability to reject the null hypothesis when alternate hypothesis is true. According to stringency criteria the empirical size of the tests should be approximately nominal and the decision of the best or worst test are taken on the basis of shortcomings of the tests. Calculation method for shortcomings are discussed in section 3.5.1.

In DGP we have two nuisance parameters; first one is deterministic in nature while the other is stochastic one. DGP is deeply discoursed in section 3.1. Here in power comparison, we use all twenty-one combinations of autoregressive coefficients listed in table 3.1. Significance of deterministic part is assessed with the presence and absence of drift and linear trend. Three sample sizes, i.e. small, medium and large are illustrated in the simulation study and Monte Carlo sample size of 1000 is selected for smoothing the results.

Next three sections are arranged as; section 5.1 discusses the power comparison in presence of stochastic part without incorporating the deterministic part, section 5.2 illustrates the comparison in the presence of deterministic part as well stochastic parameter fluctuations and finally section 5.3 gives a brief conclusion regarding the most powerful and worst performing tests.

5.1: Power Comparison without Deterministic Part

Firstly we are going to discuss the analysis which is on the basis of change in stochastic part keeping the deterministic part absent. As we have listed in table 3.1 different combinations of AR coefficients are used to generate the data. We have selected only six noticeable cases here in this chapter for debate and remaining cases are incorporated in appendix.

Debated cases in next two sections are listed below;

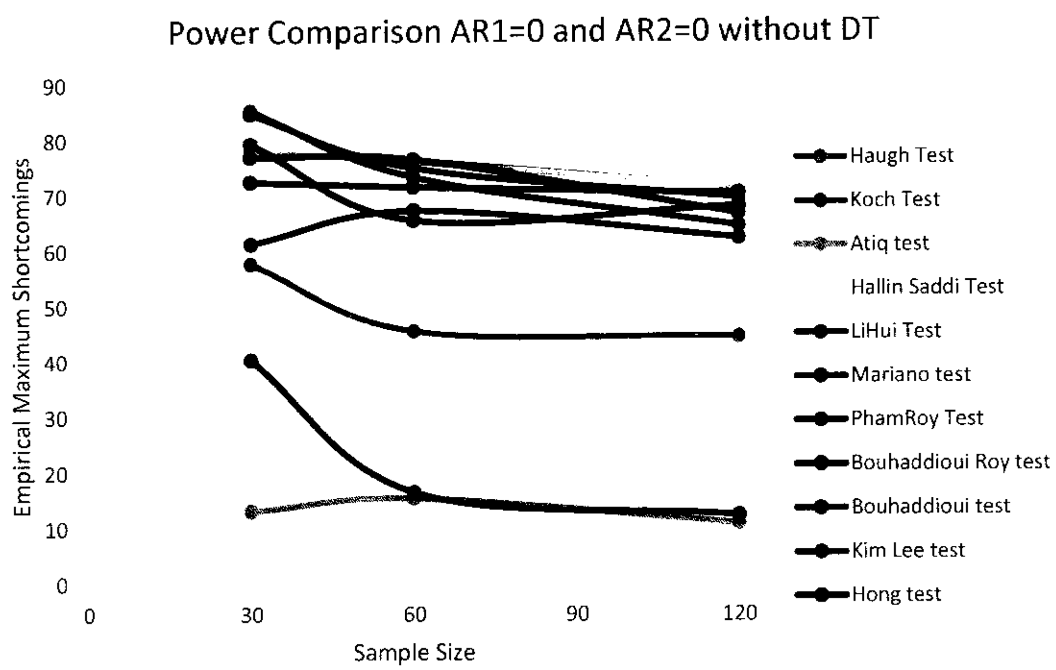
1. Both series generated under strong stationary conditions.
2. One series is strongly stationary and second is mild stationary.
3. One series is strongly stationary and second series is unit root.
4. Both series are mild stationary.
5. One series is mild stationary and second series is unit root.
6. Both series are unit root.

We start our discussion with the simplest case when both series are generated under strong stationarity conditions.

Figure 5.1.1 shows the behavior of the tests when data is generated under the condition of stationarity. We can see the maximum shortcoming curves for each test. We have three sample sizes 30, 60 and 120. Shortcomings are calculated by standard stringency procedure and the results suggest that at sample size 30, Atiq test comprehensively beat other contenders; as the shortcomings of this test are very small (below 15%); however MSC of PhamRoy test decrease drastically and at sample size 60 and 120 the shortcomings of PhamRoy and Atiq test almost coincides. Still Atiq test performs well comparatively.

Another important finding of this assessment is that maximum shortcomings have a decaying pattern as the sample size increases, although Atiq test does not depict a strong decaying pattern but PhamRoy test has a huge shift by decreasing maximum shortcomings when sample size is shifted from small to medium. Haugh test is considered the third best test in this comparison the gap between the shortcomings of PhamRoy test and Haugh test are very large. This gap shows that the first two tests outperforms the remaining ones by a fair margin.

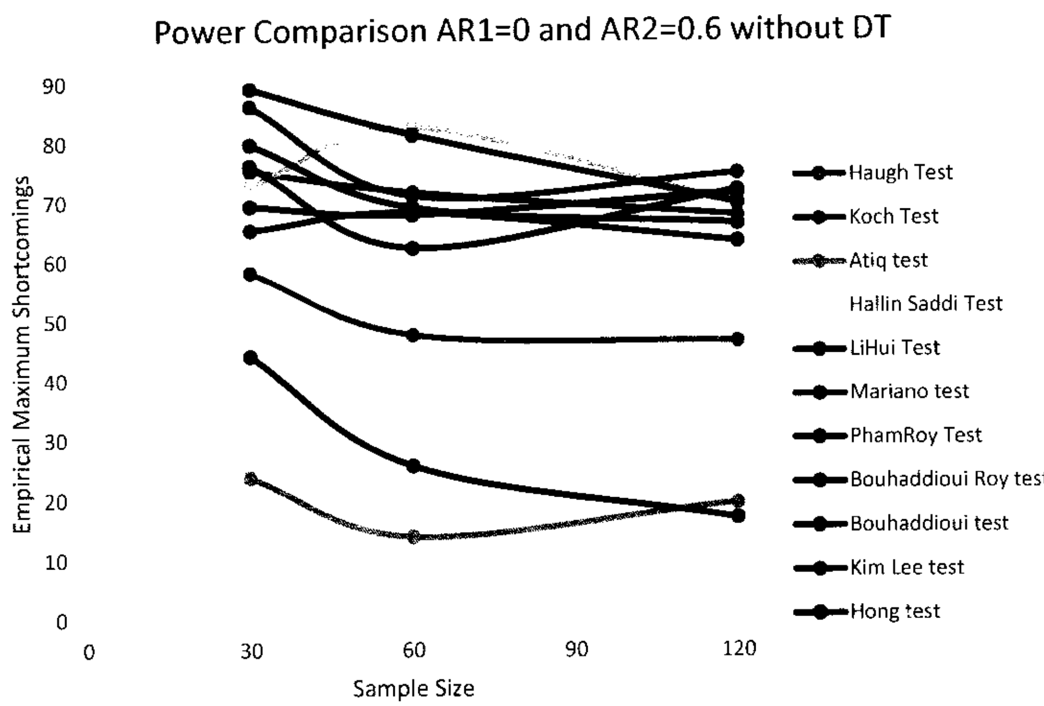
Figure 5.1. 1: Both series generated under strong stationary conditions.



At sample size of 30 Mariano test is the only one which is near Haugh test but its shortcomings are increased when sample size changes from small to medium. Although all other tests have decreasing pattern of MSC but their MSC does not decreased so effectively to prove them a good test. So all the remaining tests have huge shortcomings and they overlap each other in different sample sizes. Since their MSC decreased from 85% to about 65% but our best tests have very small MSC which are below 20%.

In figure 5.1.2 we have discussed the case when we have first series generated under strong stationary condition and second series under mild stationary, drift and trend are absent here. Atiq test performs well in the case of sample size 30 and 60, however PhamRoy test competes the test when sample size increases to 120, which means at large sample size both of these tests perform well but Atiq test performs well in small samples.

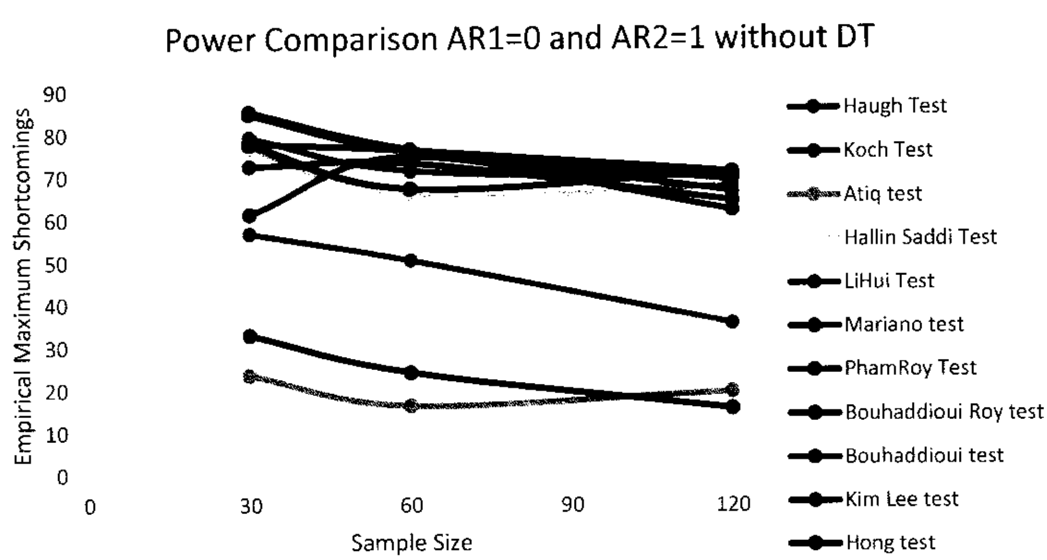
Figure 5.1. 2: One series is strongly stationary and second is mild stationary.



Another important behavior we have seen in these results is that MSC of Atiq test first decreased from 25% to just below 15% but again jumped up to nearly 20%; however PhamRoy test behavior is smooth as its MSC decreases as sample size increases. Again Haugh test is the third most powerful test, its MSC at large sample size improved a bit but at small sample sizes it performs similar as a stationary case. Mariano test which shows slight improvement in sample size 30 when DGP is stationary, its performance is devastating at medium and large sample sizes. Remaining tests perform poorly as their MSC are above 60%.

Results for DGP under which, one series is generated strongly stationary and the second series is generated unit root, are depicted in figure 5.1.3. Atiq test again performs well in small and medium samples and its MSC are 25% and just below 15% in small and medium sample sizes respectively. So we can say that MSC gain is about 10% but as was seen in figure 5.1.2, MSC of Atiq test increased when we move from medium to large sample size. However this behavior is not good but the shift is not big enough to consider it as a shocking result. PhamRoy test again performs well and has a smooth decaying effect as the sample size increases. In small and medium sample size MSC are above 20% but in large sample size it decreased to 15.8%. Haugh test has MSC above 50% in small and medium sample size but decreased to 35% in large samples and a steep decay can be seen in the figure. Bouhaddioui and LiHui tests have performed worst as they have MSC above 80% in small samples and they also perform badly in medium and small samples. Kim Lee, Hallin Saddi and Hong tests also have increasing MSC Patterns at times.

Figure 5.1. 3: One series is strongly stationary and second series is unit root.



When both series are generated mild stationary, the behavior of tests can be examined in figure 5.1.4. Atiq test performed well in all sample sizes and have a decaying effect in MSC. In small sample size MSC are slightly above 24% and decreased to about 14% when sample size is 60 but again it goes slightly above 16% when sample size increased to 120. The pattern of quadratic trend is somewhat persistent in MSC of Atiq test.

PhamRoy test have MSC of 40% in small sample but improved to about 20% in medium and large sample size. Haugh and Mariano test both started from 60% MSC in small sample but Mariano test increased MSC instead of decreasing, so Haugh test again appears to be the third most powerful test for testing independence in time series. Again LiHui, Bouhaddioui and Kim Lee tests performs worst as they have MSC almost 90%.

Figure 5.1. 4: Both series are mild stationary.

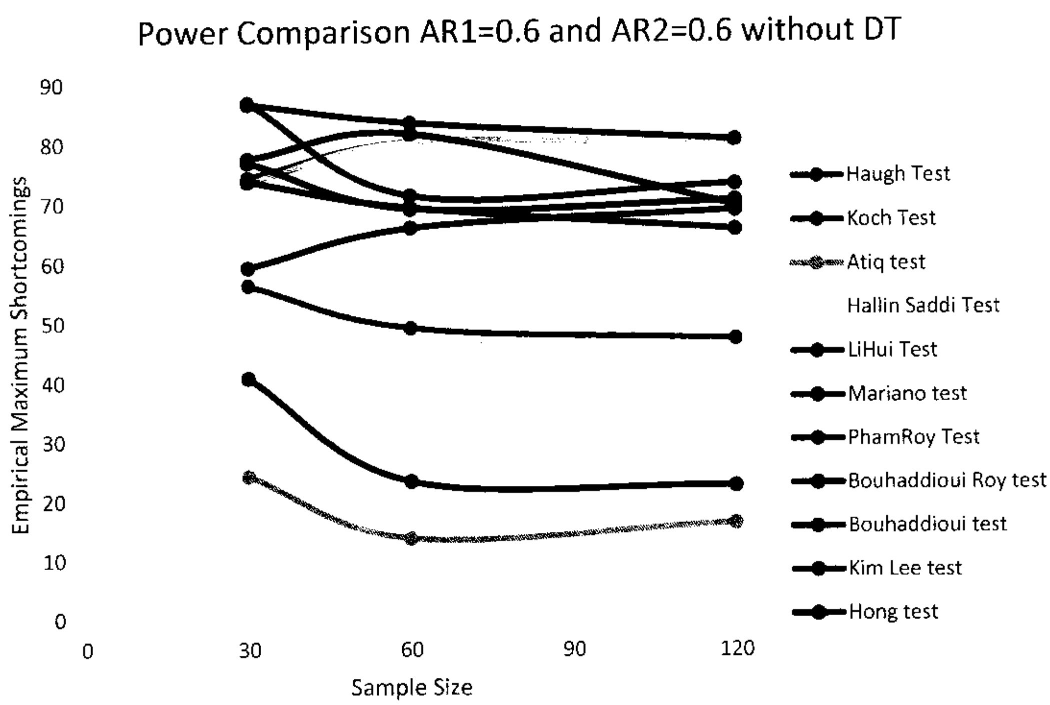


Figure 5.1.5 describes the behavior of incorporated tests of independence for time series when first series is generated mild stationary and second series is generated unit root. We can see that Atiq test dominates other tests in all three sample sizes. MSC at sample size 30 are just below 18% but it went up to above 20% in sample size 60 and then decreased again to 11%. Although Atiq test performs better in almost all cases and sample sizes but increase in MSC with increase in sample size is still not acceptable in literature, however this increase is not so significant. PhamRoy test have MSC of 38% when sample size 30 but increase in sample size reduced the MSC with a steady and smooth behavior. Haugh test again is on number three with MSC at 61%, 44% and 42% at respective sample sizes. LiHui test is considered worst in this case as its MSC are highest of all in all three sample sizes. Remaining tests similarly have MSC above 60% except Bouhaddioui test at sample size 120.

Figure 5.1. 5: One series is mild stationary and second series is unit root.

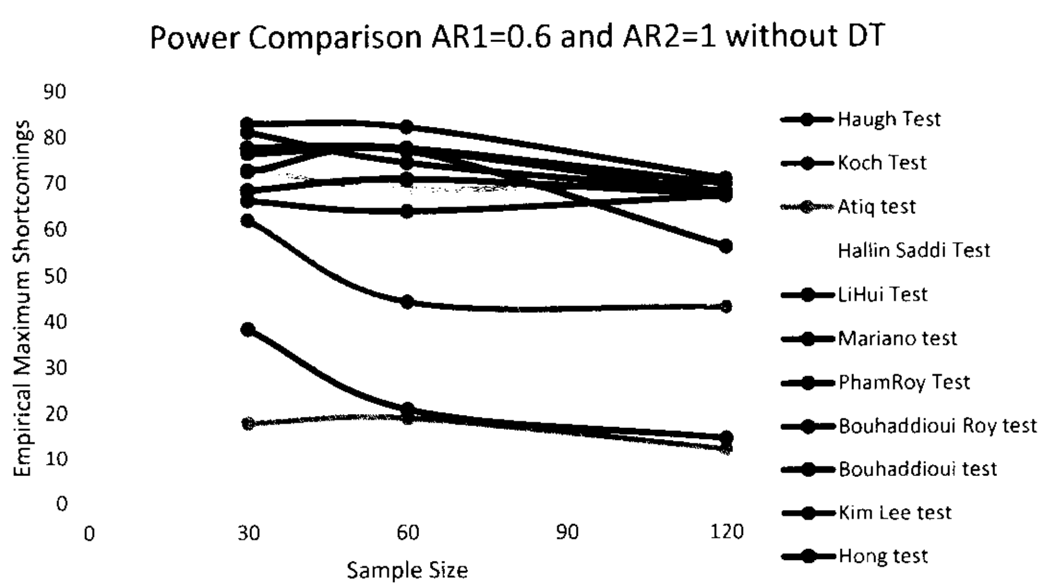
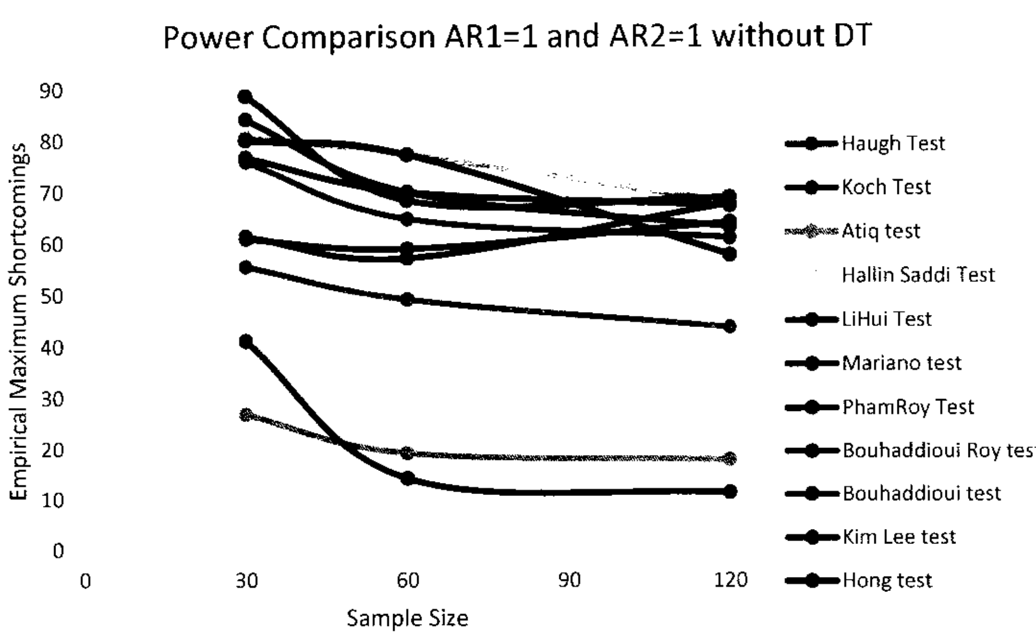


Figure 5.1.6 described the behavior of our tests of independence for time series when both series are generated exactly unit root. The results illustrate that Atiq test performed well in small sample and have decreasing pattern of MSC in sample size 60 and 120 but comparatively to other cases discussed above MSC are bit high. At sample size 30 MSC of Atiq test is 29% and falls when sample size is 60 at 19% and decreased 1% more when sample size=120.However PhamRoy test starts MSC of 40% in small sample but performed remarkably well as it reaches 14% in medium and at 11% when sample size is high. Here we can see that PhamRoy test even perform better than Atiq test. We noticed that remaining tests even performed worst in unit root case as compared to the other cases, as few tests achieved MSC almost 90% in small samples. Haugh test remains on third position; although MSC decreased when sample size increased but this decrease is not as sharp as we have seen in previous case.

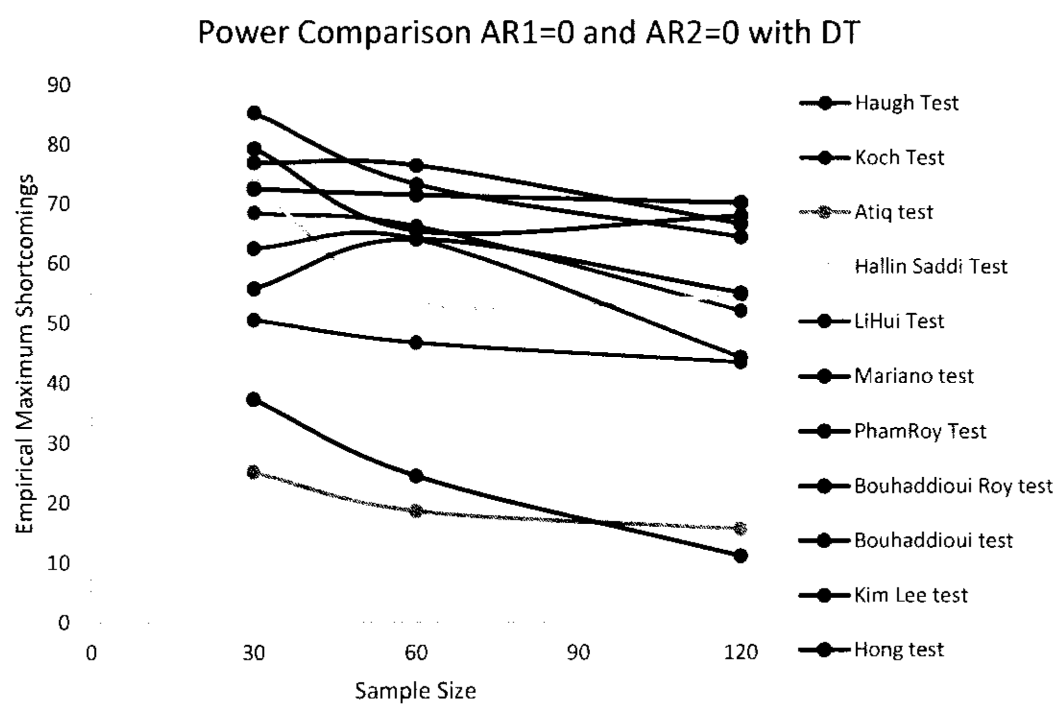
Figure 5.1. 6: Both series are unit root.



5.2: Power Comparison with Deterministic Part

Deterministic part of our study includes a constant or drift and a linear trend in DGP. In this section analysis is based on different autoregressive coefficient, i.e. change in stochastic part in the presence of deterministic part (inclusion of drift and linear trend in the DGP). Same six cases mentioned in the earlier section are discussed here and remaining cases are integrated in the appendix.

Figure 5.2. 1: Both series generated under strong stationary conditions



In figure 5.2.1 we have MSC of tests when both series generated under strong stationary conditions with drift and linear trend included in the DGP. Results show that Atiq test becomes dominant in small and medium sample size with MSC of 25% and 18% respectively but in large sample size PhamRoy test rules the rest with MSC of 11%.

Haugh test remained on third position in all sample sizes. Mariano, Hallin Saddi and LiHui tests are on number four on 30, 60 and 120 sample sizes respectively. Bouhaddioui Roy, Bouhaddioui, Kim Lee and Hong tests consistently performed poorly. All the tests have decreasing pattern of MSC except Mariano and Hallin Saddi.

Comparing with the same stochastic part i.e. strong stationary but in the absence of deterministic part we have seen that Atiq test dominates in small sample but comes quite close to PhamRoy test for medium and large sample sizes. Although MSC without deterministic part are below 15% and here these are above 25% but decreased to 15% in large sample size. Haugh test is on third position there. So we can say that remaining tests almost performs likewise when deterministic part is present.

Figure 5.2. 2: One series is strongly stationary and second is mild stationary.

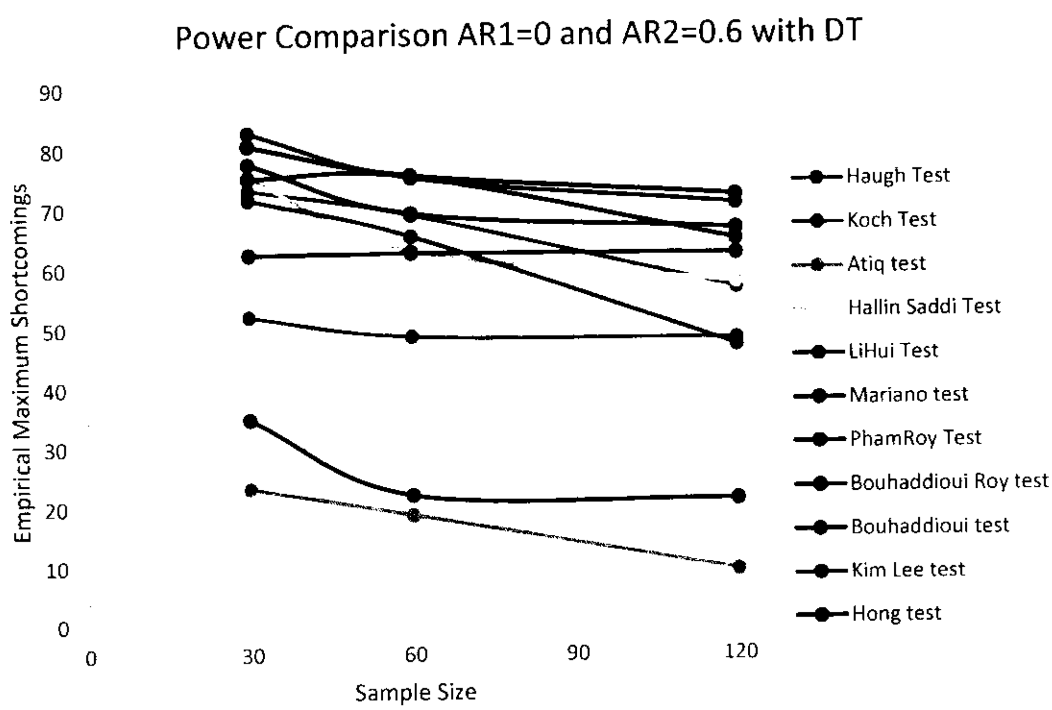
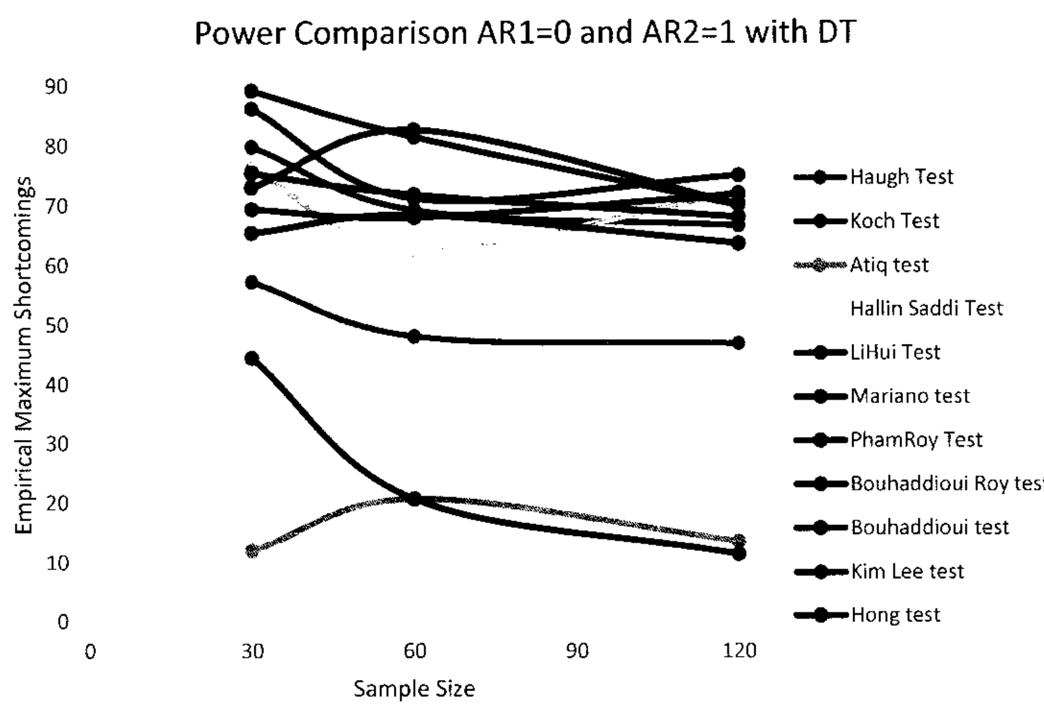


Figure 5.2.2 depicts the MSC of one strongly stationary and second mild stationary series with deterministic part. Atiq test outclasses the competitors in all sample sizes with MSC of 23%, 18% and 9% respectively in small, medium and large sample sizes. Similarly Pham Roy test rests on number two in all sample sizes, although MSC of this test decreased from 35% to 22% when sample size increased from 30 to 60 but it remained above 20% in large sample. Haugh test have lower MSC of 52% here and 48.6% in medium but it remains 48.2% in large sample size, so decline in MSC is not very quick. Here another interesting thing is that LiHui test becomes equal to Haugh test or slightly lower MSC in large sample size. Remaining results are almost same as in stationary case where Bouhaddioui Roy, Bouhaddioui, Kim Lee and Hong tests stayed on last positions.

Figure 5.2. 3: One series is strongly stationary and second series is unit root.

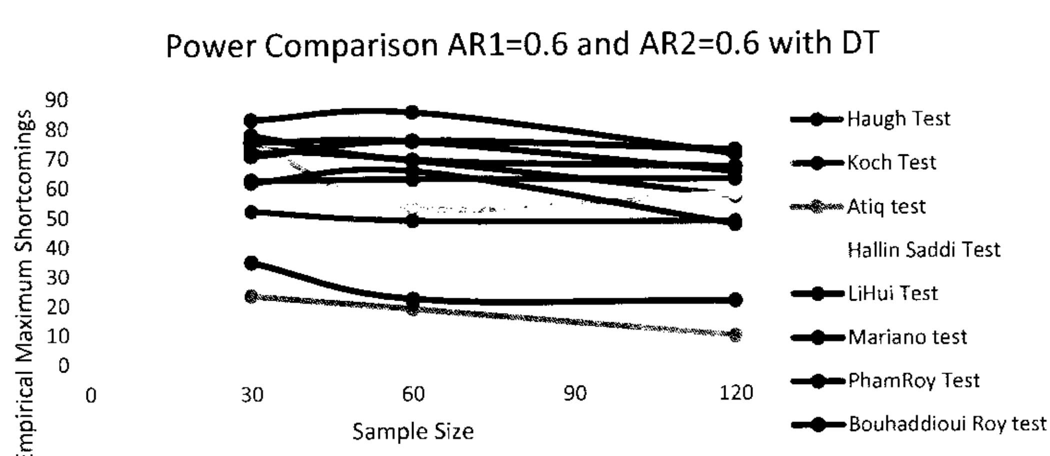


The same case without deterministic part shows that Atiq test has almost same MSC in small sample and decreases in medium sample size but increases MSC when sample size becomes large, so it performed well, similarly PhamRoy test also has same start but sharp decline in MSC takes it to first position.

In figure 5.2.3, we have MSC for DGP with a mixture of a strongly stationary and a unit root series including drift and trend. Atiq test have smallest MSC of 12% in small sample but went high to 20% in medium and again goes down to 13% in large sample size. We have also seen this increasing behavior of Atiq test in previous section. Our consistent test of PhamRoy again become dominant in medium and large sample sizes with MSC of 20% and 11% respectively.

Haugh test with same MSC of between 60% to 45% remains on third position. LiHui test performed worst in small samples and the remaining tests are not noticeable as there MSC are overlapping each other, so one test perform better than others in one sample size and worst in others.

Figure 5.2. 4: Both series are mild stationary.



Comparing figure 5.1.3 with 5.2.3 we have noticed that Atiq test behavior is similar as its MSC first decreased and then increased, however in magnitude MSC are comparatively low in DGP with deterministic part. PhamRoy test have comparatively higher MSC in small samples but it coincides in medium and large sample sizes when we compare with and without drift and trend cases.

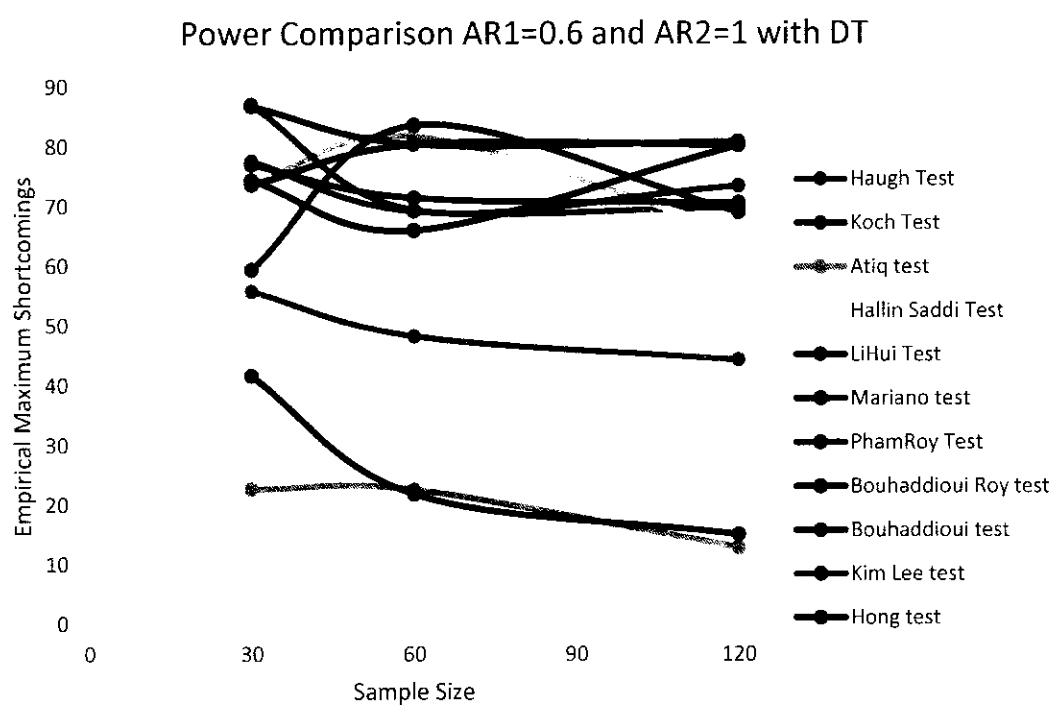
In case of two mild stationary time series with drift and linear trend Monte Carlo simulation results are illustrated in figure 5.2.4. Atiq test leads other contestants with a smooth decline in MSC, i.e. 23%, 18% and 9% respectively in small, medium and large sample sizes. Although PhamRoy test have MSC of 34% in small sample and reduced to 22% in medium sample but fail to ease in large sample size and remain in second position.

Haugh test could not present good performance here. Although it beats others in small and medium sample sizes but LiHui test performs better in large sample size. At one point we can see that Hallin Saddi test minimizes MSC when sample size increases from small to medium but it again creeps up. Kim Lee test considered worst in this case as MSC are very high in all three sample sizes. Remaining tests again have very abrupt behavior.

Relating figures 5.1.4 and 5.2.4, Atiq test performs well in the presence of deterministic part. In both cases Atiq test accomplishes well especially in small sample, however with deterministic part using medium and large sample size its MSC reduced with smooth and predictive pattern. Therefore it outclasses other tests in all three sample sizes here. PhamRoy test have lower MSC in case of deterministic part using small sample size but equal in medium and large sample sizes. Haugh test performs similarly in both cases with slight deviations in MSC.

With the combination of a mild stationary and a unit root series DGP, MSC of the tests are illustrated in figure 5.2.5. Empirical MSC show that at sample size 30 Atiq test performs comparatively better than PhamRoy and other tests with MSC of 22%, i.e. runner-up test which is PhamRoy test have MSC of 42%. However at sample size 60 and 120 both tests are almost at same position. MSC of Atiq test are 22% and about 13%, on the other hand PhamRoy test have 22% and 15%.

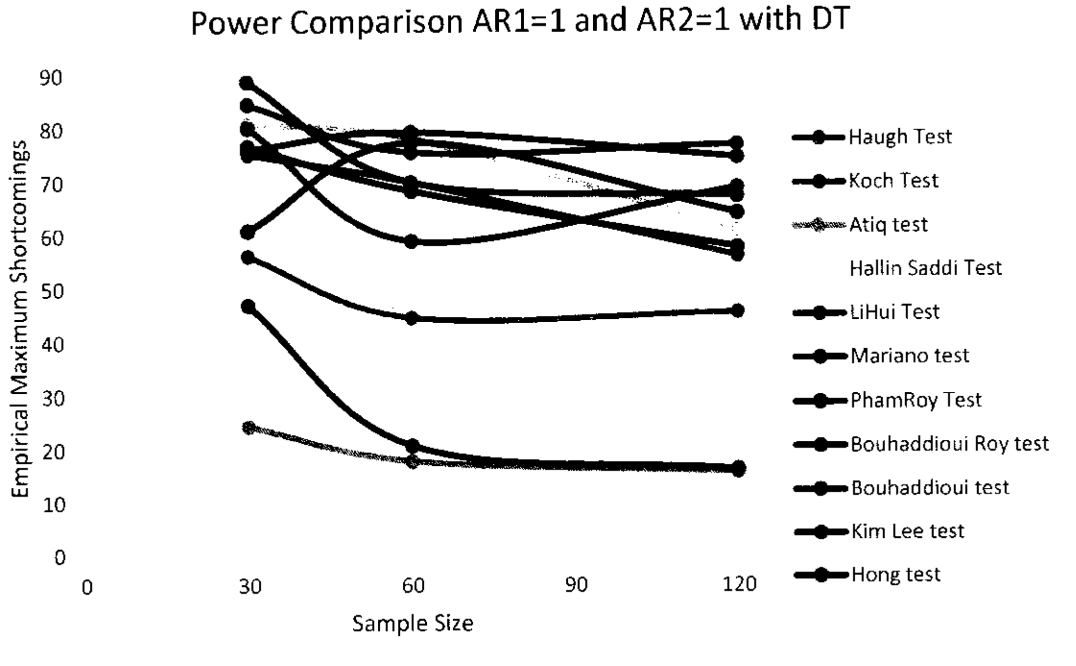
Figure 5.2. 5: One series is mild stationary and second series is unit root.



Haugh and Kim Lee tests have MSC of 56% and 59% respectively at sample size 30 but Kim Lee test MSC increases with increase of sample size, though Haugh remained on third position with 48% and 44% MSC. Hong with Daniel Kernel, Bouhaddioui and Kim Lee tests have MSC over 85%. Similarly other tests also have abrupt behavior and not a single test performs better.

Comparing figure 5.2.5 with the same case without drift and trend, we have noticed that Atiq test have relatively higher MSC of 22% in small and medium samples and 13% in large samples with deterministic part case however 17%, 18% and 12% respectively without deterministic part case. PhamRoy test also have 4% increase in MSC from without to with deterministic part case in small sample, however in medium and large samples MSC are almost same. Remaining tests have comparatively higher MSC without deterministic part case as they have largest discrepancy of 82% in previous section and 87% in this section.

Figure 5.2. 6: Both series are unit root.



When data is generated under strict unit root conditions with drift and linear trend, the MSC of compared tests are given in figure 5.2.6. Atiq test again performs well in small, medium and large sample sizes. In comparison to without deterministic part Atiq test have lower MSC in all sample size. PhamRoy test have MSC of 47%, 21% and 16% in three respective sample size which is comparatively higher than 41%, 14% and 11% with deterministic part case. Haugh test is consistently on third position in all sample sizes with almost same MSC as without deterministic part case.

Bouhaddioui test performs worst in presence and absence of deterministic part with the largest MSC of about 90%. Remaining tests perform in similar manner.

With obtained results we are in position to clinch that Atiq test performs well in small samples in all cases of stationary, mild nonstationary and unit root. However PhamRoy test performs well in large samples most of the times. Haugh test becomes the third most effective test of our analysis in nearly all sample sizes and DGP's.

5.3: Conclusion

Power comparison of eleven tests of independence for time series was conducted under DGP with two nuisance parameters. These parameters are of stochastic and deterministic nature, however stochastic part attain twenty-one different options while deterministic part have two possible outcomes entertained in this study. A Neyman Pearson (PO) test developed to get the power envelop and shortcomings for each test is calculated using this power curve. Now maximum of these shortcomings against each alternative is calculated and tests are compared on the basis of these maximum shortcomings (MSC).

Out of forty-two cases examined in this study only twelve important cases are discussed in this chapter and remaining are enclosed in the appendix. Almost all the cases depicted that there is a head to head rivalry in Atiq and PhamRoy tests. Summarizing the results we can conclude that Atiq test performs well in small samples, though for medium and large sample sizes Atiq and PhamRoy tests produce almost identical results. Another important finding is that these two tests performed likewise when deterministic trend is present or not.

Briefly discussing about the Atiq test, we have noticed on number of occasions when MSC of the test increased with the increase in sample size, however this phenomena is globally acceptable as with increase in sample size power of the test should be increased. In statistics, it is assumed that increment in sample size is the easiest technique to boost the statistical power of a test. It is harder to perceive the effects in relatively small samples and addressing the same issue, i.e. when DGP is same, larger sample size should give equal or higher power. In section 5.1, we have seen this happened in many cases, like when both series are strongly stationary Atiq test have MSC of 13% in small samples but increased to 16% when applied on medium samples. Again we discussed strongly stationary and mild stationary case where MSC of 14% in medium sample and increased to 19% in large sample. We discussed the case of strongly stationary and unit root series in figure 5.1.3, it is again evident that in medium sample MSC are 16% but in large sample it is 19%. In case of two mild stationary series it is again obvious that the test have increasing behavior with MSC of 13% in medium sample and 16% in large sample size. Case number 5, 9, 13 and 18 mentioned in table 3.1 are also showing the similar results, figures for these cases are enclosed in appendix.

Although Atiq test shows increasing behavior in many cases discussed in section 5.1 but this frequency reduced a lot when we moved towards the section 5.2, i.e. when deterministic part was added in the DGP. In most of the cases usual downward sloped line appears which shows reduction in MSC due to increase in sample size.

PhamRoy is the second most favorite test considered which have smaller MSC. Although in almost all cases it has MSC greater than Atiq test in small sample but it shows a smooth predictive behavior, i.e. as sample size increased MSC decreased. Due to this predictive behavior it seems more reliable test in my opinion when we have large data set, since we have seen on occasions that PhamRoy and Atiq tests have equal MSC when medium and large sample sizes are used.

Another important finding in this study is when we have DGP with no drift and linear trend then tests have greater MSC on overall sample sizes but with drift and trend these MSC reduced. In both stationary series case and without deterministic trend, MSC in small sample for Haugh and PhamRoy test are above 60% and 40% respectively however it reduced to 50% and below 40% when deterministic trend was included in the DGP. Similarly with one strongly stationary and a mild stationary case, when we move from without deterministic trend to deterministic trend case MSC are reduced by approximately 10% for PhamRoy and Haugh test and about 5% for Atiq test. These patterns can also be seen in other cases.

Mariano test performs better on number of occasions, both with deterministic part and without deterministic part. Especially this happens when series are closer to strong stationarity and particularly in small samples. In figure 5.1.1 it is seen that Mariano test has slightly lower discrepancy then Haugh in small sample (MSC of 61%), similarly in figure 5.1.4 the results depict that Mariano test has MSC of below 60% which is just above Haugh test. From table 3.1, in cases 2, 7, 9, 12, 19 and 21; Mariano test performs better as compared to the other tests except Atiq, PhamRoy and Haugh tests in small sample.

Kim Lee, LiHui and Bouhaddioui tests are considered worst in all sample sizes and with and without deterministic part cases. In figure 5.1.1, Kim Lee test have MSC of above 85% in small sample, in figure 5.1.2, it turns out to be 90%, MSC of about 87% in figure 5.1.4 and above 80% in figure 5.1.5. Similarly in figure 5.1.3 Bouhaddioui test have MSC of 85% in small sample and above 90% in figure 5.1.6. It is evident from figure 5.1.4 and 5.1.5 that LiHui test along with Kim Lee test perform worst in all sample sizes with MSC of above 80%.

It is concluded by the above debate that Atiq test performs well in small sample of all cases even with deterministic trend and without deterministic trend DGP. However the test have a shortcoming without deterministic trend case as MSC increased with increase of sample size in number of occasions, though this behavior is not so significant and vanished when we analyzed with deterministic trend DGP. PhamRoy test have smooth predictive behavior of MSC, although it has higher MSC in small sample but it performs very well in medium and large samples even better than Atiq test in some cases. Haugh test emerged as the third best test of the analysis, its MSC are way higher than Atiq and PhamRoy test but as we see that other tests have very rough behavior so comparatively, Haugh test perform well in all sample sizes and deterministic parts. At times Mariano test performs better in small sample when we are close to stationary conditions but fails in medium and large sample sizes. Kim Lee, LiHui and Bouhaddioui tests are considered worst in this analysis.

CHAPTER 6: REAL DATA COMPARISON

In general we can say that simulation is invented to deal with a real thing but really working with an artificial. Simulation is widely used in Statistics to verify the precision of any model or test. It is very important and acceptable in researchers because large data sets are not available for recalculating the models a large number of times (especially in Economic data sets). So simulation reveals the mimicking of the real empirical analysis of the data. Simulation approach is rather different then analyzing a real data set, i.e. an artificial data set is generated under controlled DGP, where the analysis is purely based on theoretical assumptions. As the simulation technique gives more flexibility and convenience but real data analysis is more reliable. These two methods are complementary for one another.

6.1: Agenda of the Real Data Analysis

In this chapter, we are interested to apply the test of independence on real economic data set. We move an artificial environment to the real situation for assessing the credibility of our tests. As mentioned above in the introductory words of this chapter that simulation based investigation and real data analysis are complementary for evaluating the integrity of the tests, so one important motive of this chapter is to validate the results of simulation study presented in chapter 4 and 5. However this is not the only objective of this chapter. As mentioned in section 1.3, test of independence for time series can be used as a powerful tool for model selection criteria, so no need for sophisticated and tedious methods to check the relation between the variables.

Granger causality and cointegration techniques are widely acceptable techniques for analyzing time series nowadays, even not knowing about the credibility of these techniques. These techniques come under critique time to time but people come up with a new version (slight maneuvering).

6.1.1: Tests of Independence for Time Series

Tests of independence are designed to measure the interdependence between two or more time series. Detailed discussion about the origin and types of these tests are carried out in chapter 2, so there is no need to further explain them here.

6.1.2: Cointegration Tests

If we have two or more series and all are integrated (in the time series integrated means unit root) but any linear combination of these series has stationary or in case of second order integrated has a lower order of integration then we can say that these series are cointegrated. The logic behind the theory is that discrepancy between observations of a series make it unit root so the adjacent linear combination that wash out the discrepancy between two unit root series are called cointegration factor and the process is called cointegration.

6.1.3: Granger Causality Test

If we have two time series and one series is predicted by its own passed values, another time series and its past values, and this prediction is better than the prediction of only on its own past values, we can say that this second series Granger cause first series.

Granger causality is based on two important principles:

1. "The cause happens prior to its effect.

2. The cause has unique information about the future values of its effect".

6.2: Size and Power Comparison of Tests of Independence

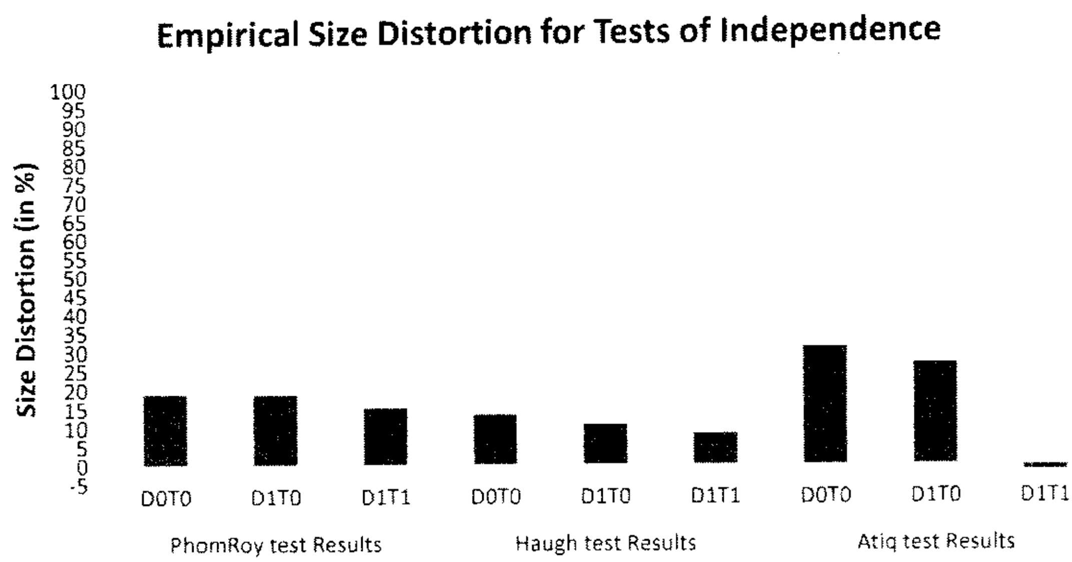
Tests of independence for time series are compared in previous sections on the basis of their size and power discrepancies using stringency criteria in a Monte Carlo simulation study. In this section we are evaluating the tests of independence using a real data analysis. In this regard a renowned economic phenomena of income consumption hypothesis is selected for analysis. Three prominent tests which emerged during the simulation based study are selected. These tests are top two and a mediocre test which performed relatively better than other tests, i.e. Atiq, PhamRoy and Haugh Tests. Income (Y) and consumption (C) series are selected for hundred countries and assume the income and consumption of same countries have a strong relationship, similarly income and consumption of different countries have logically no relationship between them.

Brief mathematics of the tests is described in section 2.3 and this is evident from the construction of these three tests that all of them are based on prewhitened residuals. These residuals can be attained by applying a linear (ARMA) model, so we have a stochastic parameter for this model which is calculated using Akaike Information criteria (AIC) and deterministic part of the model which has three possible outcomes, i.e. no drift and trend (D_0T_0), drift and no trend (D_1T_0) and both drift and trend (D_1T_1). Method to find size distortions and power gains are illustrated in section 3.7.5.

In figure 6.2.1, size distortions for three tests of independence for time series are presented. These three tests have further three deterministic part cases each. In PhamRoy test when we have no deterministic part case (D_0T_0), the size distortion is about 19%, but as we included the drift in the model i.e. (D_1T_0) this size distortion decreases to 18%, although this decrease is not so much significant but as we incorporated both drift and trend (D_1T_1) a nominal decrease of 3% occurred in size distortion and it becomes 15%. Haugh test also present the same decreasing pattern of size distortion when we move from no deterministic part to all deterministic parameters. Haugh test have relatively smaller distortions in size than PhamRoy test in all three cases. When no drift and trend (D_0T_0) included in the model Haugh test have 13% size distortions and 11% when we include drift only (D_1T_0), however it reduces to 8% when both drift and trend (D_1T_1) are incorporated in the model. In absence of deterministic part i.e. (D_0T_0), Atiq test have 31% size distortion, by economic theory we can say that GDP and consumption series are drift and trend in it, so Atiq test behavior is natural and we can say that this test is oversized here. Similarly when we incorporated drift only ($D1T_0$), these distortions reduced to 27% but it is still very high, however when both drift and trend included (D_1T_1), size distortions becomes -1.6%, it is rather surprising for a test which has 31% and 27% size distortions in previous cases and now it approaches to nominal size of 5%.

Concluding the discussion we can say that Haugh test have fairly close to nominal size and show good smooth behavior, PhamRoy test also performed better and size distortions are good enough to compare on the basis of power gains however Atiq test shows high size distortions in without drift and trend cases but done remarkably well in both drift and trend cases.

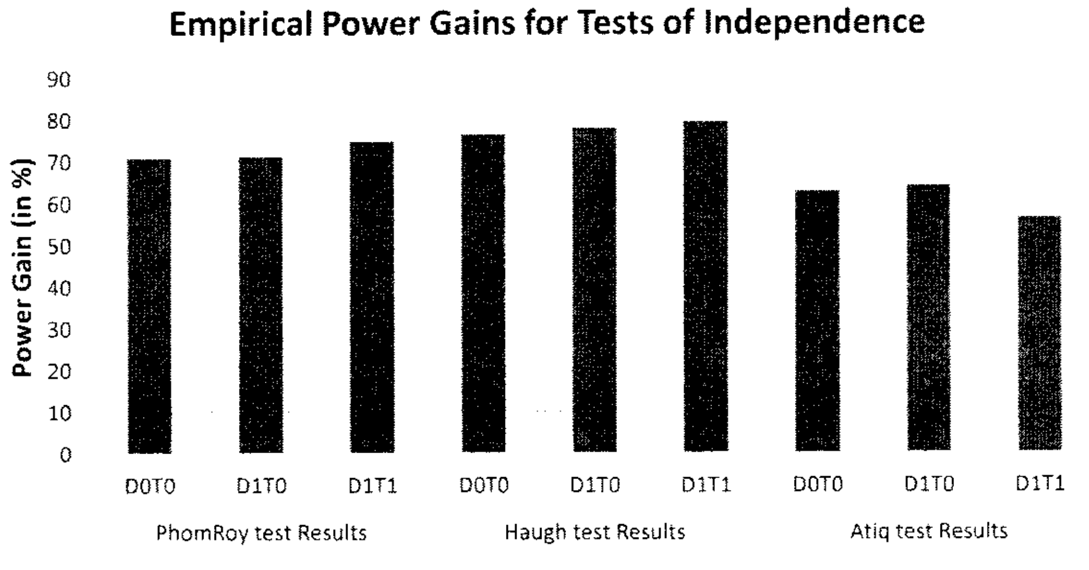
Figure 6.2. 1: **Size Distortions for Tests of Independence using real economic data**



In figure 6.2.2, we have power gains for all three tests with three possible deterministic cases each. PhamRoy test have 71% power gains in both without deterministic case (D_0T_0) and only drift case (D_1T_0). However power gain increased to 75% when drift and linear trend (D_1T_1) both are included in the model. Haugh test performed well here as it has 76% power gain in no deterministic part (D_0T_0) case and 78% in drift (D_1T_0) case only, however it jumps up to 80% when both drift and trend (D_1T_1) are included in the model. Atiq test has relatively low power gains in all three cases. In no deterministic case (D_0T_0) Atiq test has 63% power gain and by increment of 1% it becomes 64% in drift only (D_1T_0) case while it reduces to 56% in both drift and trend (D_1T_1) case. In above figure we have seen, when both drift and trend (D_1T_1) are included in the model size distortions for Atiq test are also very low, so if a test have small size distortions than its low power gains are also acceptable.

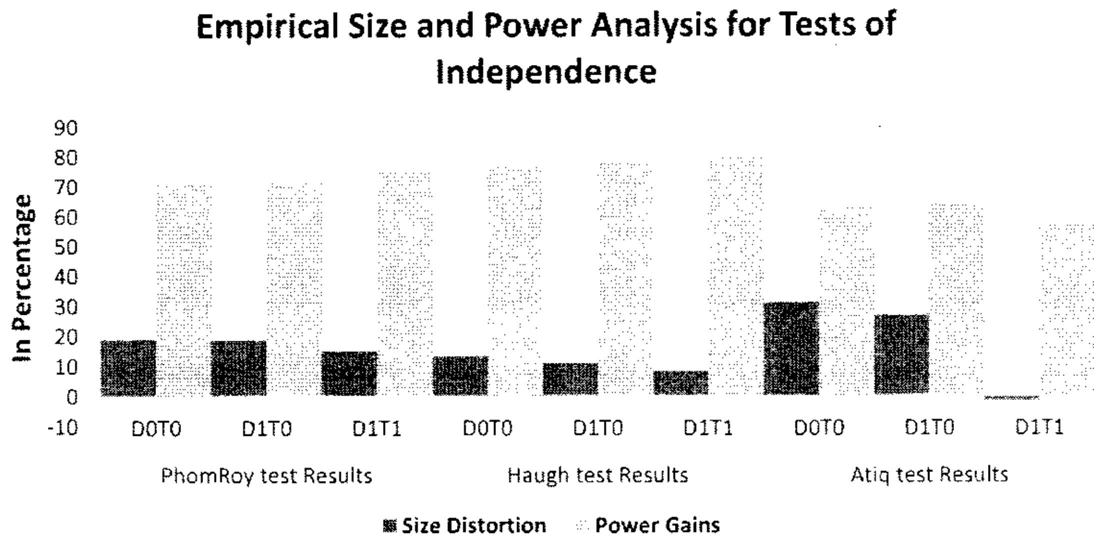
Concluding the debate we can say that Haugh test achieved a good reputation in all tests incorporated in this section of real data comparison. However PhamRoy test also performed better although its power gains are relatively lower than Haugh test but still it gives a consistent performance in all three cases. Atiq test has not achieved as good position in without drift and trend cases, however when we have incorporated both drift and trend (D_1T_1) it considered on first place as was rated at the first place in simulation study.

Figure 6.2. 2: Power Gains for Tests of Independence using real economic data



In figure 6.2.3, we have a comparison of size distortion and power gains for all three tests and all three deterministic part variations of each test. A test which has lower size distortions and higher power gains will be considered better than other. Haugh and PhamRoy tests have less than 20% size distortions and more than 70% power gains in all cases. However Atiq test have more than 30% size distortion in first two cases i.e. in no deterministic (D_0T_0) and drift only (D_1T_0) cases and less than 70% power gain in these respective cases but it has -1.6% size distortions and almost 56% power gain in both drift and trend (D_1T_1) case, although with almost no size distortion this power gain is good enough for the test. We can say that Atiq test performance is not consistent but it is very much consistent with the economic theory, as we have previously discussed the economic phenomena of the two concerned economic variables.

Figure 6.2. 3: Size Distortions and Power Gains for Tests of Independence using real economic data



At the end of this section we can say that Haugh test performed well in real data analysis as its size distortions are comparatively lower than PhamRoy test and its power gains are relatively higher than PhamRoy test in all cases.

6.3: Size and Power Comparison of Cointegration Tests

Simulation study of Khan (2017) concluded that five tests having null of no cointegration are relatively better than overall sixteen compared tests. Although in Monte Carlo simulation study these tests have very low power whether using asymptotic critical values or simulated ones, but in these comparison using standard stringency criteria these tests perform better than others.

In our real data analysis all five tests have three options for deterministic part i.e. no drift and trend (D_0T_0), drift only (D_1T_0) and both drift and trend (D_1T_1). Critical values for the rejection of null has been calculated using simulations and nominal size is assumed to be 5%. All five tests are applied on income and consumption data for hundred countries. Assuming the economic theory of income and consumption as true, the null hypothesis are defined as;

$$H_0: Income\ and\ consumption\ series\ are\ independent$$

We have hundred incomes and hundred consumptions, so when we are using the income and consumption of same countries i.e. $i=j$ we assume null is wrong using economic theory and when this hypothesis is rejected in such cases, percentage will be referred to as power of the test. Similarly when we have income of one country with consumption of another country i.e. $i \neq j$, it is not logically true that cointegration exists between them, so we assume the null as true and rejection of true null becomes the size of the test.

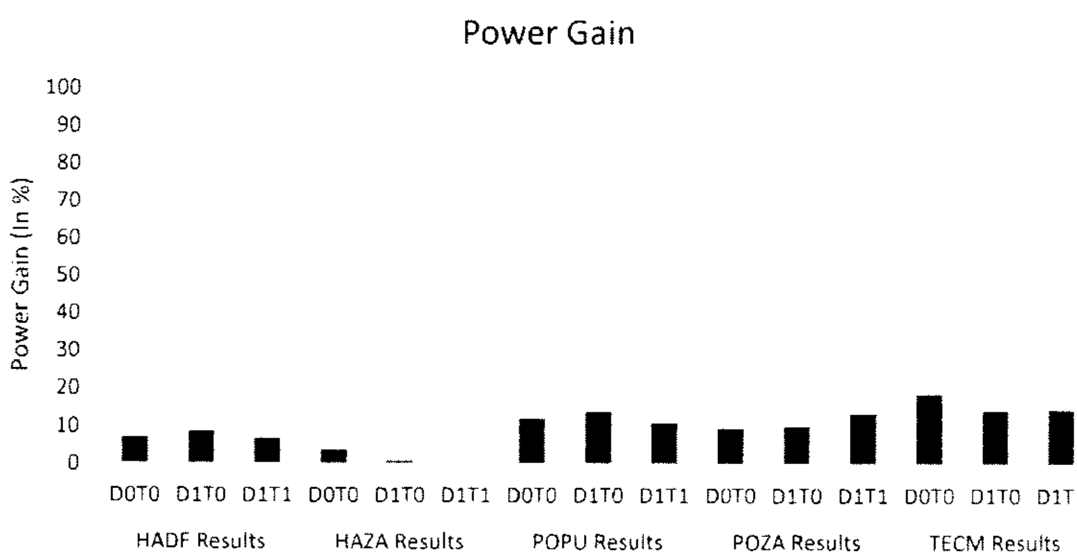
Results of tests are summarized in graphical presentation and tables are enclosed in the appendix. Size distortions and power gains of each test is reported with three deterministic parts for each. For simplification, we discuss size distortions and power gains separately in two different graphs and after that we present a joint graph for size distortion and power gains to conclude the tests which have smallest size distortions and maximum power gains.

In figure 6.3.1, we have size distortions for five cointegration tests included in our real data analysis. Graph depicts that most of the tests have minor negative size distortions. Only TECM test have a very small positive distortions. HAZA test with drift and trend (D_1T_1) case is considered to be the outlier in the analysis as it has almost 95% size distortions. Remember that these distortions are calculated using the nominal size of 5% and in above mentioned case we have actual size of the test using simulated critical value is almost 100%. It is considered to be the only shocking result in cointegration testing, although remaining tests have a very reasonable behavior, as almost all the remaining tests have nominal size.

Figure 6.3. 1: Size Distortions for Cointegration Tests using real economic data

As discussed above we have nominal size of 5% in all cointegration tests and size distortion are negligible in almost all cases except a single case mentioned above. In figure 6.3.2, we have power gains for cointegration tests which are considered most stringent in Monte Carlo simulation study conducted by Khan (2017). Although these power gains are not as high as we have seen in tests of independence in the above section but still we have power gains for five tests for three deterministic part cases for each. We have maximum power gains for t Test of Cointegration in a Single Equation Error Correction Model (TECM), however these gains are below 20%. TECM have power gain of just above 18% for no drift and trend (D_0T_0) case and 14.25% and 14.5% in only drift (D_1T_0) and both drift and trend (D_1T_1) cases respectively. Second cointegration test which is considered to have relatively high power gains is Phillips and Ouliaris $\hat{P}_\mu$test of cointegration (POPU), for three cases of no drift and trend (D_0T_0), only drift (D_1T_0) and both drift and trend (D_1T_1) cases it receives 12%, 14% and 11% power gains respectively. Similarly Phillips and Ouliaris'$\hat{Z}_\alpha$ test of Cointegration (POZA) is considered to be in the third place in real data analysis with power gains of 9%, 10% and 13.38% in three deterministic part cases of (D_0T_0), (D_1T_0) and (D_1T_1) respectively.

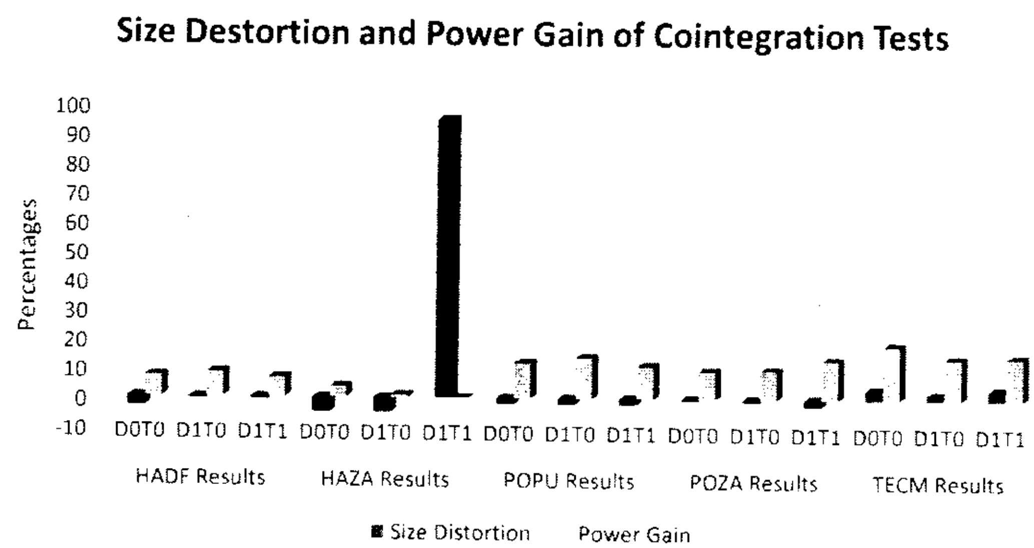
Figure 6.3. 2: Power Gains for Cointegration Tests using real economic data



Hansen's variation of the ADF test of Cointegration (HADF) remains on fourth position as it receives 7.25%, 8.7% and 6.9% power gains in three respective cases of deterministic part. Hansen's variation of the $\hat{Z}_{\alpha}$test of Cointegration (HAZA) performed poorly in all three cases of deterministic part and obtains 3.93%, 0.98% and 0% power gains in respective deterministic part cases.

In figure 6.3.3, we have both size distortions and power gains for cointegration tests. This figure depicts that cointegration tests have very low size distortions except a single case but besides size distortions these tests also have very low power gains, most appealing test considered in this real data analysis have very low power gains. Not a single test detects one out of five cases for cointegration. TECM is considered best test for cointegration in real data analysis of Keynesian hypothesis. Similarly HAZA remains on last in analysis of same hundred countries real data analysis.

Figure 6.3. 3: Size Distortions and Power Gains for Cointegration Tests using real economic data



6.4: Size and Power Comparison of Granger Causality Test

In this section we present the real data analysis for Granger causality test. We have data of income and consumption for hundred countries and analysis for these hundred countries is carried out in the way that we have null hypothesis as;

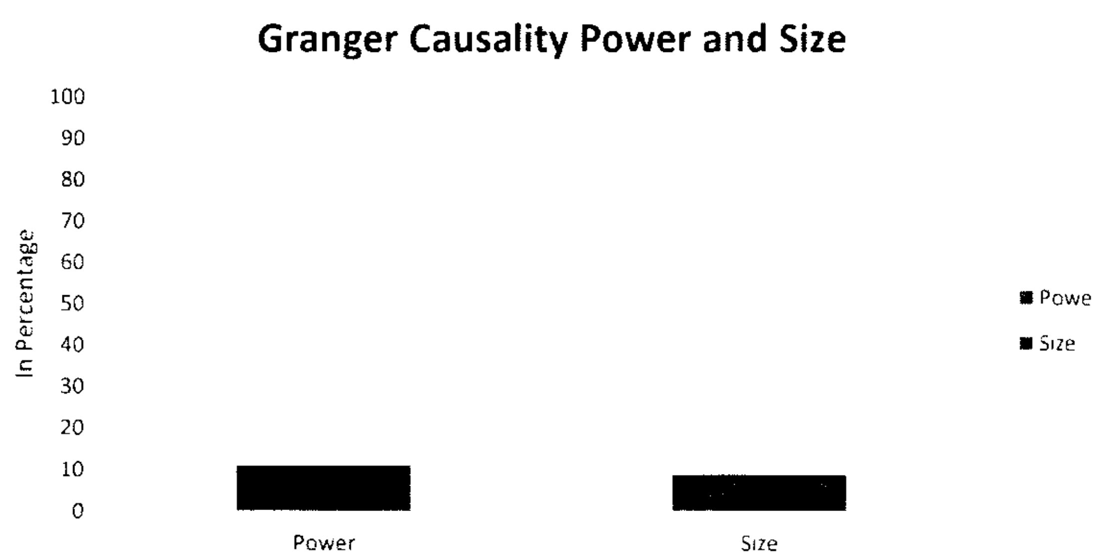
$$H_0: Y_i \text{ does not Granger Cause } C_j$$

According to standard economic theory income Granger causes consumption, so rejection of null hypothesis (*when i=j*), we consider it as power of the Granger causality test. Correct decision of Granger causality test referred as power of the test in figure 6.4.1. Actually we can check two hypothesis for Granger causality as income Granger cause consumption and consumption Granger cause income. If test concludes consumption Granger cause income this is wrong decision. Similarly when Granger causality test resulted to no causal relation or both way causality we can call it inconclusive, because according to Granger causality test causal variable should happens prior to its effect, so in our analysis both way causality should not happen and secondly causal variable have unique information about the future values of its effect variable, i.e. income should effect consumption so no causality case also eliminated from our analysis.

In figure 6.4.1, we have used two main bifurcations in our analysis, i.e. income and consumption of the same countries *(when i=j in null hypothesis)* and income and consumption of the different countries *(when i≠j in null hypothesis)*. In first case "Power" represents the cases when Granger causality gives verdict that income Granger cause consumption which is 11%, so it is considered as the power of the test. Similarly 89% of times Granger causality test fails to detect causal relation or in other words it is inconclusive. It means we have a strong causal relation according to economic theory that income Granger cause consumption but test fails to detect this 89% times.

In this classification we have two further categories, first one, when test concludes no way causality which shows that in the presence of strong economic relationship (income should cause consumption) test fails to detect it and second one, when test concludes both way causality, it is considered false as we know that consumption could not cause income. Using the assumption of Granger causality we can say that cause should happen prior to the effect so by the increase in income one can increase the consumption but the converse is not true.

Figure 6.4. 1: Size and Power for Granger causality Tests using real economic data



In second case we have opposite situation, we have income and consumption of different countries so logically income of one country and consumption of another country could not effect one another. Here we have 9900 cases and out of these 886 cases (8.95%) concludes that income causes consumption, as this null is true when income and consumption of same countries are taken but in this case it is not true, so we have rejected a true null and by statistical definition of size we name it as the size of the test.

6.5: Conclusion

Three different procedures heading towards the same inference, as testing independence between the variables, finding cointegrating relationship between variables or discovering which variable causes the other are the same conclusions. In real data analysis we adopted three different techniques which are used to get the same result in time series data analysis. Testing independence in time series is relatively simple and quick procedure comparative to other techniques under discussion, however results of the study show that tests of independence have performed comparatively better than other advanced techniques. In section 6.2 we have seen that tests of independence have moderate size distortions but its power gains are pretty high. Haugh test attains lowest size distortions as 13% and highest power gains as 80%. Similarly PhamRoy test also have low size distortions at below 20% and high power gains at about 75%. Atiq test also performed well and has low size distortions and high power gains.

Comparing tests of independence with cointegration tests, later performed poorly in all deterministic part cases. Although cointegration tests almost achieved nominal size of 5% except one case but power gains for these tests are very low as compared to the tests of independence. Highest power gain for cointegration test could not exceed 20%, which means these tests on the basis of sample provided, predict only one out of five correct cases and in other words these tests have four incorrect decisions out of five using real data.

Similarly Granger causality test also performed poorly in real data analysis. From consumption and income data for hundred countries using Keynesian consumption function Granger causality took correct decision about the causality direction of 11% countries. These countries includes Algeria, Cote d'Ivoire, Kuwait, Mauritania, Oman, Saudi Arabia, United States of America and Uruguay. It is also noted that Algeria's income not only Granger cause its own consumption but it also Granger cause twelve other countries consumption, similarly Saudi income Granger cause its own consumption as well as consumption of seventy-seven other countries, so we can conclude that Granger causality test could not detect true causality patterns.

Form the above discussion about the three testing procedures for assessing relationship between variables, we can say that tests of independence performed relatively better than the other two techniques. Although tests of independence need more advancements to reduce the size distortions but still these tests have higher power gains to detect the pattern of the data.

CHAPTER 7: CONCLUSION, RECOMMENDATIONS AND DIRECTIONS FOR FUTURE RESEARCH

In this chapter we try to summarize the findings of our simulation study about the tests of independence for time series and also about the real data analysis for three procedures i.e. tests of independence, cointegration tests and Granger causality test. These inferences support us to give some recommendations about the future econometric practices and also help us to show the way for forthcoming research in this field.

7.1: Conclusions

Firstly we conclude about the Monte Carlo simulation study for tests of independence for time series which comprised of chapter 4 and 5. In chapter 4 we illustrated size comparison for tests of independence, following a brief note about why we used simulated critical values and not asymptotic critical values for taking decision about the null hypothesis. In that discussion we gave examples that how much size was distorted in simulation study when asymptotic critical values were used. In the very next section we presented stability of size for eleven tests of independence for time series using simulated critical values. We saw that all the tests had nominal size of 5% using all the stochastic part cases and deterministic part cases. We checked stability of size for twenty-one stochastic part cases in which autoregressive coefficients were changed in DGP. That process was repeated in the presence of deterministic part as well as in the absence of it, so in general we have testified the size stability of about forty-two cases for each test.

In section 4.2 we had size comparison of tests for variation in stochastic part without having deterministic part in DGP. We have noticed that almost all the tests are between the bandwidth of 3% to 7%, it confirmed the stability of size for all tests of independence. Similarly in section 4.3 we had size stability for tests of independence with variation in stochastic part in the presence of deterministic part. We had drift or constant and a linear trend in deterministic part. The results depicted that all the tests are within the band of 3% to 7% which ensured the stability of size. As we used the stringency criteria for comparing the tests of independence so we had to stabilize the size of tests and then we were able to compare the tests on the basis of their powers. So chapter 4 concluded that size of tests are stable and we can compare the tests on the basis of their powers.

In chapter 5 we have compared eleven tests of independence for time series using stringency criteria. For comparison we used power analysis as we had stable size for all of these tests. Maximum shortcomings for each test calculated against three sample sizes 30, 60 and 120. DGP had same characteristics described in the size comparison section with above mentioned forty-two cases. A Neyman Pearson test was designed to get the locally maximum powers so we can get a power envelop for the comparison of all the tests. Actually this NP test acted as a benchmark here and powers of each test should be compared with this benchmark and maximum difference is noted as maximum shortcomings, now a test said to be most stringent if maximum shortcomings of a test are minimum, this is what we call stringency criteria.

The main findings of that power analysis was that Atiq and PhamRoy tests performed well in almost all cases. Atiq test ranked as number one as its MSC were minimum in most of cases however PhamRoy test also performed better and its MSC were close to Atiq test. Mostly for sample size 30 Atiq test performed far better than PhamRoy test. As we noticed that when both series were generated under strong stationarity conditions Atiq test had 13.2% MSC on sample size 30 while PhamRoy test had 40.4%. When we had first series as strong stationary and second one mild stationary, MSC became 23.8% for Atiq test and 44.2% for PhamRoy test. Similarly when we had first series as strong stationary and second series unit root then MSC were 23.6% for Atiq test and 33% for PhamRoy test. Deterministic trend was not included in above said all cases. When deterministic trend included in the DGP, for both series strongly stationary case, Atiq test had 25.2% MSC and PhamRoy test had 37.4%. Similarly when one series is strong stationary and second one is mild stationary, Atiq test had 23.2% and PhamRoy test had 34.8% MSC and when we had one series strong stationary and second series unit root, Atiq and PhamRoy tests had 12% and 44.4% MSC. Although Atiq test had clear superiority over PhamRoy test when we had small sample size as 30, however we haven't seen any vibrant pattern when we were changing DGP from Strong stationarity to unit root or DGP with deterministic trend to without it.

At sample size 60 and 120, both tests had mixed performance. Generally Atiq test became dominant but at times Atiq test had a habit of unorthodox behavior of increase in MSC with the increase of sample size, however PhamRoy test had a decreasing pattern of MSC with the increase of sample size. We have seen that when we had both series generated as strong stationary and deterministic trend is absent then on sample size 30 Atiq test had13.2% and PhamRoy test had 40.4% MSC but when we increased the sample size to 60 both tests had very minor difference of 1% in MSC so it became 15.5%

and 16.6%. Further when we reached large sample size as 120 it became 10.9% and 12.4% respectively. So we can say that for large sample size PhamRoy and Atiq test had almost similar performance.

For one strong stationary series and second mild stationary series, Atiq test performed better than PhamRoy test in small sample, however when we had sample size as 60 MSC of Atiq test were 14% and PhamRoy test were 25.8%, still Atiq test had good upper hand but when we moved towards the large sample size Atiq test had MSC of 19.7% and PhamRoy test had 17.2%. So PhamRoy test became superior to Atiq test in terms of MSC and also with respect to the theoretical behavior of the test as we know that with the increase of sample size power of the test should increase. Similar situation arrived when we had one series as strong stationary and second series as unit root. We had MSC of Atiq test as 16.4% and PhamRoy test had 24.2%, when sample size was 60, but it became 19.7% for Atiq test and 15.8% for PhamRoy test when we had sample size of 120. So PhamRoy test again became superior on large sample size. When we had both series generated as mild stationary without deterministic trend, Atiq test had MSC of 15% and 20.6% on sample size 60 and it became 21.1% and 14.2% respectively for sample size 120. For both series generated as unit root, MSC for Atiq and PhamRoy tests were 19.2% and 14.2% respectively for sample size 60 and 17.9% and 11.4% respectively for sample size 120.

As we perceived from the above discussion that in the absence of deterministic part the behavior of Atiq and PhamRoy tests confirmed the superiority of PhamRoy test in large sample sizes and also unconventional behavior of Atiq test. Now we can conclude it in the presence of deterministic part. So when we had both series generated strong stationary including deterministic part in DGP, then Atiq and PhamRoy test had MSC of 18.7% and 24.6% respectively for sample size 60 and 15.7% and 11.2% when sample size is 120. Likewise when we had one series generated as strong stationary and second series as mild stationary, here MSC on sample size 60 for Atiq and PhamRoy test were 16.7% and 20.1 % respectively and for sample size of 120 were 11.1% and 9.3%. When we had first series as strong stationary and second series as unit root, these numbers became 20.7% and 20.6% for 60 sample size and for 120 sample size it turned out to be 13.5% and 11.4% respectively. In the presence of deterministic part when both series were generated as mild stationary, at sample size 60 MSC of Atiq and PhamRoy test were 15.1% and 20.7% respectively while on sample size 120 it became 20.7% and 13.8% respectively.

When our DGP had first series as mild stationary and second series as unit root, MSC for above mentioned tests were 18.7% and 20.6% respectively on sample size 60 and similarly when sample size were 120 it became 11.7% and 14.2%. At the end of the discussion about these two tests we must mention the case where both series were generated as unit root here we had MSC of Atiq and PhamRoy tests were 17.9% and 20.8% respectively for sample size 60 and it became 15.9% and 16.4% when sample size were 120. Concluding the above discussion we can summarize that when we had small sample size like 30 or even 60, Atiq test had very low MSC. So we can say that Atiq test has higher powers or in other words Atiq test is more stringent in small sample sizes however when we had large sample size, say 120 or so, PhamRoy test is more stringent than Atiq test in most of the cases. Atiq test had a habit of unorthodox behavior, which make it notorious for large sample sizes, however for small sample sizes it performed very well.

In second phase we discuss about the mediocre position tests. Surprisingly we had a single test which emerged as a mediocre one. Haugh test which declared by consensus as the first test of testing independence for time series became third most stringent test in our analysis. Haugh test had MSC for small sample size 30 between 50% to just above 60% in almost all the cases and for medium and large sample sizes these figures are between just below 40% at times to 50%. However no one other test had MSC near Haugh test so the remaining tests considered to be the unacceptable as there MSC are very high.

Remaining tests performed worst and there MSC overlaps each other at times. As one test performed better under one DGP or at times using only one sample size and performed worst under the other, so the performance of remaining tests are not notable. Since Kim Lee, LiHui, Koch and Bouhaddioui tests considered worst in all sample sizes and with and without deterministic part cases. When we had both series generated strong stationary in the absence of deterministic part Kim Lee test had MSC of above 85% in small sample, similarly when we had first series stationary and second series mild stationary it turns out to be 90%, MSC of about 87% when both series were mild stationary and above 80% when one series is mild stationary and second series was unit root. Similarly in the absence of Deterministic part with one series strongly stationary and second series unit root, Bouhaddioui test had MSC of 85% in small sample and above 90% when both series were generated as unit root. It is evident that, when we generated both series as mild stationary and one series mild stationary and second series as unit root, LiHui test along with Kim Lee test performed worst in all sample sizes with MSC of above 80%. So concluding about this section we can say that these eight tests could not perform consistently as performed by Atiq, PhamRoy and Haugh tests.

At second stage we briefly conclude about the empirical or real data analysis in which we not only compared the tests of independence for time series but also compared two other well-known and frequently used econometric tools to analyze the time series. These techniques are cointegration and Granger causality tests. The results are described with reference to relevant graphs and figures in chapter 6. Eminent Keynesian hypothesis of income were taken for analyzing the data of hundred countries from 1970 to 2014. Top three tests of independence for time series were selected for real data analysis, top five cointegration tests for null of no cointegration were nominated from Khan (2017) and famous Granger causality tests is also incorporated in the analysis for comparison.

Khan (2017)conducted Monte Carlo simulation study using stringency criteria concluded that five tests for the null of no cointegration had relatively better power. In our empirical analysis these five tests represented the cointegration techniques. The results depicted that t Test of Cointegration in a Single Equation Error Correction Model (TECM) had attained highest power gain of just above 18% when no deterministic part included in the DGP, while power gain of about 14% in remaining two cases of only drift and both drift and linear trend cases. As we noted that size distortion of TECM are very small and negligible, as the highest size distortion is 3.5%.

Phillips and Ouliaris $\widehat{P}_{\mu}$test of cointegration (POPU) remained on second position in our analysis with highest power gain of 14% when only drift included in the analysis, similarly power gain of 12% when no deterministic part included in the DGP while its power gains decreased to 11.1% when both drift and trend were present in the DGP. Size distortion for POPU test were also minor as size for this test is smaller than the nominal size in all cases so the size distortions were negative. Results demonstrated that Phillips and Ouliaris'$\hat{Z}_{\alpha}$test of Cointegration (POZA) on third position with power gain of 13.38% when both drift and trend included in the DGP, while it turnedout to be 9.6% when DGP included no drift and trend and became 10% when only drift included in the DGP. Like TECM and POPU tests POZA test also had minor size distortions, so we can say that all these tests had nominal size of 5%.

Hansen's variation of the ADF test of Cointegration (HADF) considered to be on fourth position. Power gains for HADF were 7.25%, 8.7% and 6.9% for no deterministic part, only drift and both drift and trend cases respectively. Here once again size distortions were very low. Hansen's variation of the $\hat{Z}_{\alpha}$test of Cointegration (HAZA) remained on the last place with power gains of 3.93%, 0.98% and 0% power gains in above mentioned respective deterministic part cases. It is also evident that size distortion of about 95% when DGP had both drift and trend, so unanimously with both size distortion and power gains HAZA test considered to be the worst in this empirical analysis of cointegration technique.

Granger causality test was considered to be the reputed technique for assessing the dependence of one variable on another. In our analysis this time series analysis procedure not showed decent results. In chapter 6 we described briefly that according to economic theory change in income induced change in consumption and this is direct proportion, as income increases so consumption increases and when income decreases so consumption decreases. When we had data for the income and consumption of the same country the null hypothesis of "Income does not Granger cause Consumption" is false and the statistical definition for POWER says that rejecting the null when it is false. So the test showed only 11% power, as out of hundred countries data, null of eleven countries rejected. Similarly when income and consumption for different countries were considered, there is no logical relation between them so null of "Income does not Granger cause Consumption" is true here. The definition of size says rejecting a true null hypothesis, so the size of the test was 8.95%. We had considered 9900 of such cases and out of it 886 time test rejected the null hypothesis.

In our Monte Carlo simulation study, we had found two tests i.e. Atiq and PhamRoy tests which considered to be the best tests. Similarly we had a mediocre test i.e. Haugh test which performed better than other eight tests, so we selected top three tests for our empirical analysis section. Results of empirical study depicted that Haugh test had smallest size distortions amongst the competitors. It had 8.47% size distortion, when we had both drift and trend included for prewhitening the errors using AR model. Similarly it was 10.99% when only drift included in the model and 13.42% when AR model did not contain any deterministic part. Haugh test also had highest power gains of 79.53% when drift and trend included in the model, power gain of 78.01% when only drift included in the model while it became 76.58% when no deterministic part present in the model.

PhamRoy test again performed well and had least size distortions as 15.15% when both drift and trend included in the model, similarly it became 18.75% when only drift is present in the model while it turns out to be 19.01% when no drift and trend included in the model. Power gains are 70.99%, 71.25% and 74.85% in without deterministic part, only drift and both drift and trend cases respectively. Atiq test, here again in empirical analysis test showed again the weakness of inconsistency. Atiq test had size distortion of 31.14% when no deterministic part present in the model, it became 26.91% when drift included in the model while it turned out to be -1.62% when both drift and trend included in the model. Power gains were not as high as other two tests but still it performed well as it has power gains of 62.86% in no deterministic case, 64.09% in drift only case while power gain of 56.62% when both drift and trend included in the model.

Comparing these three econometric techniques together we can clearly conclude that tests of independence for time series are far better than cointegration and Granger causality tests in terms of size distortions and power gains. Although tests of independence especially Atiq test had greater size distortions in without drift and trend but still Atiq test is consistent with economic theory remaining tests also had reasonable size distortions and there power gains are far better than other two techniques.

7.2: Recommendations

Mostly data series of economic variables are time dependent. In time series data analysis when we are screening the variables on first sight we do not need to apply the sophisticated econometric techniques. This observations can be strengthen mutually by our Monte Carlo simulation study and empirical data analysis. Using empirical study we can recommend that tests of independence are more powerful tool while deciding about the interrelationship between the variables. Similarly we can say that tests of cointegration and Granger causality test are very weak in analyzing the real data. Results suggests that PhamRoy and Haugh tests are very consistent in assessing independence between economic variables while Atiq test have good power gain and reasonable size distortion but lack consistency at large samples as we have seen in Monte Carlo simulation study.

In Monte Carlo simulation study we have seen that for small sample sizes Atiq test always dominate its competitors while in medium and large sample sizes PhamRoy test have upper hand. We have also recommend that for small sample sizes researchers and practitioners can use Atiq test however we have not found any such evidence for including deterministic part or not. PhamRoy test will be recommended for medium and especially for large sample sizes.

7.3: Directions for Future Research

In this study we focus only on interrelationship between two variables while tests of independence are available which analyze more than two variables, so one can assess the independence using these multivariate tests. We use simulated critical values in our Monte Carlo simulation study and give many examples that how different researchers got size distortions using asymptotic critical values, similarly we give examples of stable size using asymptotic critical values. Unfortunately we have many tests which are based on maximum likelihood estimation technique and take several days to get a single result so we are not able to show these size distortions here as we have a huge power analysis and empirical analysis also. We have used ARMA model for obtaining residuals which are used for finding the test statistics for many nominated tests. While fitting ARMA model we have taken maximum AR and MA terms up-to 3 lags and BIC as model selection criterion. One can use different selection criteria and set maximum lag different to check the deviations in the results.

For empirical analysis we have selected three tests of independence. These tests are selected on the basis of their performance in Monte Carlo simulation study. We have to restrict our self to three tests because of shortage of time and resources, so can check the performance of all the tests of independence for time series using real economic data set. Although Granger causality test performed very poorly in our empirical study however Granger causality test have many modified versions which are available in literature. Most of these modifications of Granger causality test are on the basis of different tests of independence for time series. So one can use all these modified versions and can see if some modified version performed better or not.

REFERENCES

Ahking, F. W. (2002). Model mis-specification and Johansen's co-integration analysis: an application to the US money demand. *Journal of Macroeconomics, 24*(1), 51-66.

Atiq-ur-Rehman, A.-u.-R., & Zaman, A. (2008). Model specification, observational equivalence and performance of unit root tests.

Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The annals of mathematical statistics*, 485-498.

Bouhaddioui, C., & Dufour, J.-M. (2008). Tests for non-correlation of two infinite-order cointegrated vector autoregressive series. *Journal of Applied Probability and Statistics, 3*(1), 78-94.

Bouhaddioui, C., & Roy, R. (2006). A Generalized Portmanteau Test For Independence Of Two Infinite-Order Vector Autoregressive Series. *Journal of Time Series Analysis, 27*(4), 505-544.

byJOHN MAYNARD, K. (1936). The general theory of employment, interest, and money.

Chan, N. H., & Tran, L. T. (1992). Nonparametric tests for serial dependence. *Journal of Time Series Analysis, 13*(1), 19-28.

Duchesne, P., & Roy, R. (2003a). Robust tests for independence of two time series. *Statistica Sinica*, 827-852.

Duchesne, P., & Roy, R. (2003b). Robust tests for independence of two time series. *Statistica Sinica, 13*(3), 827-852.

Eichler, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis, 99*(5), 968-1009.

El Himdi, K., Roy, R., & Duchesne, P. (2003). Tests for non-correlation of two multivariate time series: a nonparametric approach. *Lecture Notes-Monograph Series*, 397-416.

Geweke, J. (1981). A comparison of tests of the independence of two covariance-stationary time series. *Journal of the American Statistical Association, 76*(374), 363-373.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics, 2*(2), 111-120.

Hallin, M., & Saidi, A. (2005). Testing Non-Correlation and Non-Causality between Multivariate ARMA Time Series. *Journal of Time Series Analysis, 26*(1), 83-105.

Hallin, M., & Saidi, A. (2007). Optimal tests of noncorrelation between multivariate time series. *Journal of the American Statistical Association, 102*(479), 938-951.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association, 71*(354), 378-385.

Himdi, K. E., & Roy, R. (1997). Tests for noncorrelation of two multivariate ARMA time series. *Canadian Journal of Statistics, 25*(2), 233-256.

Hoeffding, W. (1948). A non-parametric test of independence. *The annals of mathematical statistics*, 546-557.

Hong, Y. (1996a). Consistent testing for serial correlation of unknown form. *Econometrica: journal of the Econometric Society*, 837-864.

Hong, Y. (1996b). Testing for independence between two covariance stationary time series. *Biometrika*, 615-625.

Hong, Y. (1996c). Testing for independence between two covariance stationary time series. *Biometrika, 83*(3), 615-625.

Khan, A. u. I. (2017). *Theoretical and Empirical Comparisons of Cointegration Tests.* (PhD Econometrics), International Islamic University Islamabad, Islamabad.

Kim, E., & Lee, S. (2005). A test for independence of two stationary infinite order autoregressive processes. *Annals of the Institute of Statistical Mathematics, 57*(1), 105-127.

Koch, P. D., & Yang, S.-S. (1986). A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. *Journal of the American Statistical Association, 81*(394), 533-544.

Li, W., & Hui, Y. (1994). Robust residual cross correlation tests for lagged relations in time series. *Journal of Statistical Computation and Simulation, 49*(1-2), 103-109.

Matilla-García, M., Rodríguez, J. M., & Marín, M. R. (2010). A symbolic test for testing independence between time series. *Journal of Time Series Analysis, 31*(2), 76-85.

Pham, D. T., Roy, R., & Cédras, L. (2003). Tests for non-correlation of two cointegrated ARMA time series. *Journal of Time Series Analysis, 24*(5), 553-577.

Pierce, D. A. (1977). Relationships—and the lack thereof—between economic time series, with special reference to money and interest rates. *Journal of the American Statistical Association, 72*(357), 11-22.

Rehman, A.-u., & Malik, M. I. (2014). The modified R a robust measure of association for time series. *Electronic Journal of Applied Statistical Analysis, 7*(1), 1-13.

Robbins, M. W., & Fisher, T. J. (2015). Cross-Correlation Matrices for Tests of Independence and Causality Between Two Multivariate Time Series. *Journal of Business & Economic Statistics, 33*(4), 459-473.

Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *The Review of Economic Studies, 58*(3), 437-453.

Shao, X. (2009). A generalized portmanteau test for independence between two stationary time series. *Econometric Theory, 25*(01), 195-210.

Skaug, H. J., & Tjøstheim, D. (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika, 80*(3), 591-602.

WDI. (2014). World Bank. from data.worldbank.org/data-catalog/world-development-indicators

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between Time-Series?--a study in sampling and the nature of time-series. *Journal of the royal statistical society, 89*(1), 1-63.

Zaman, A. (1996). *Statistical foundations for econometric techniques*: Academic Press.

APPENDIX

Figure 8.1: Size for Tests of independence for Time Series with DT

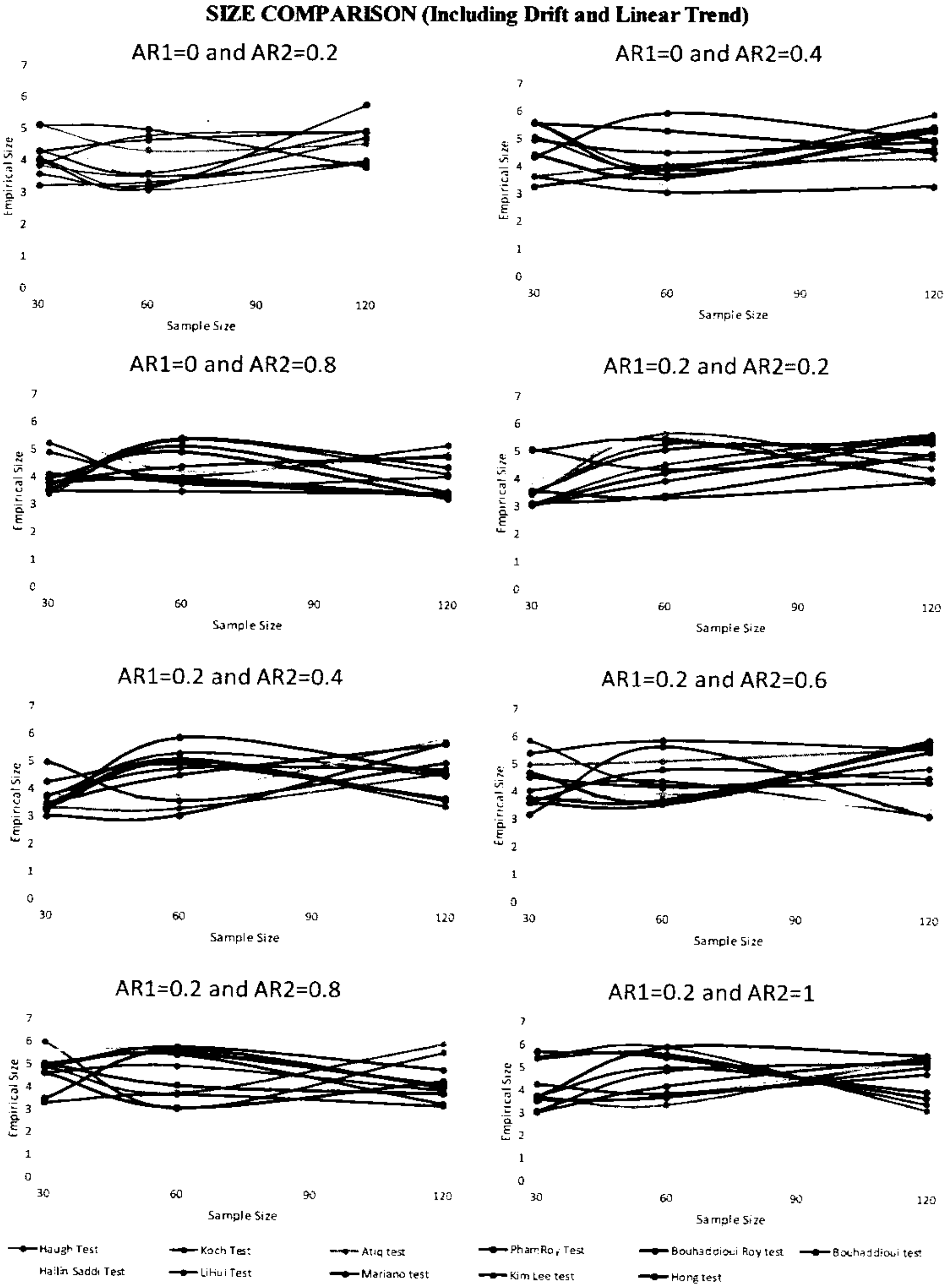


Figure 8.2: Size for Tests of independence for Time Series with DT (Continue)

SIZE COMPARISON (Including Drift and Linear Trend) (Continue)

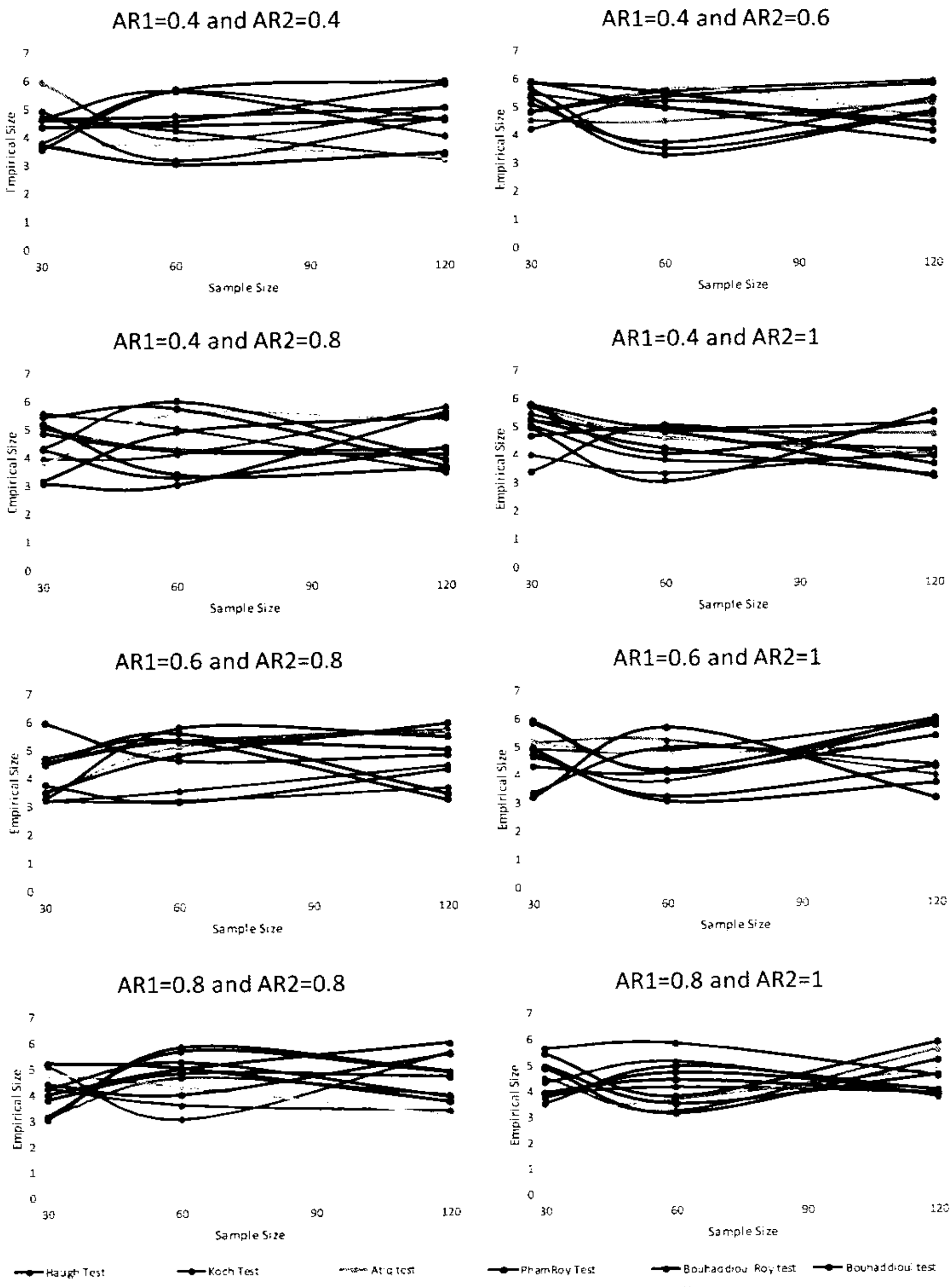


Figure 8.3: Size for Tests of independence for Time Series without DT

SIZE COMPARISON (Without Drift and Linear Trend)

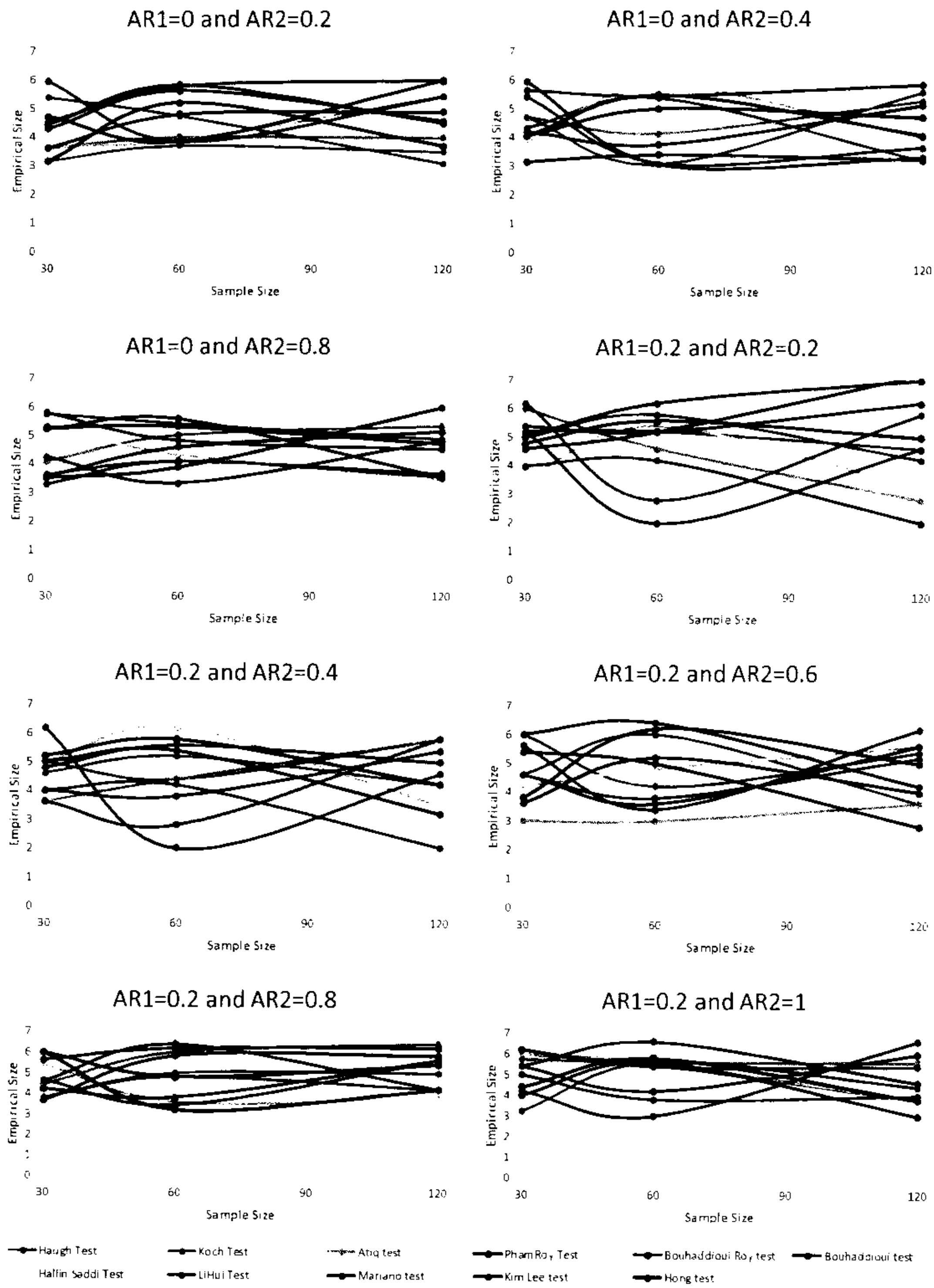


Figure 8.4: Size for Tests of independence for Time Series without DT (Continue)

SIZE COMPARISON (Without Drift and Linear Trend) (Continue)

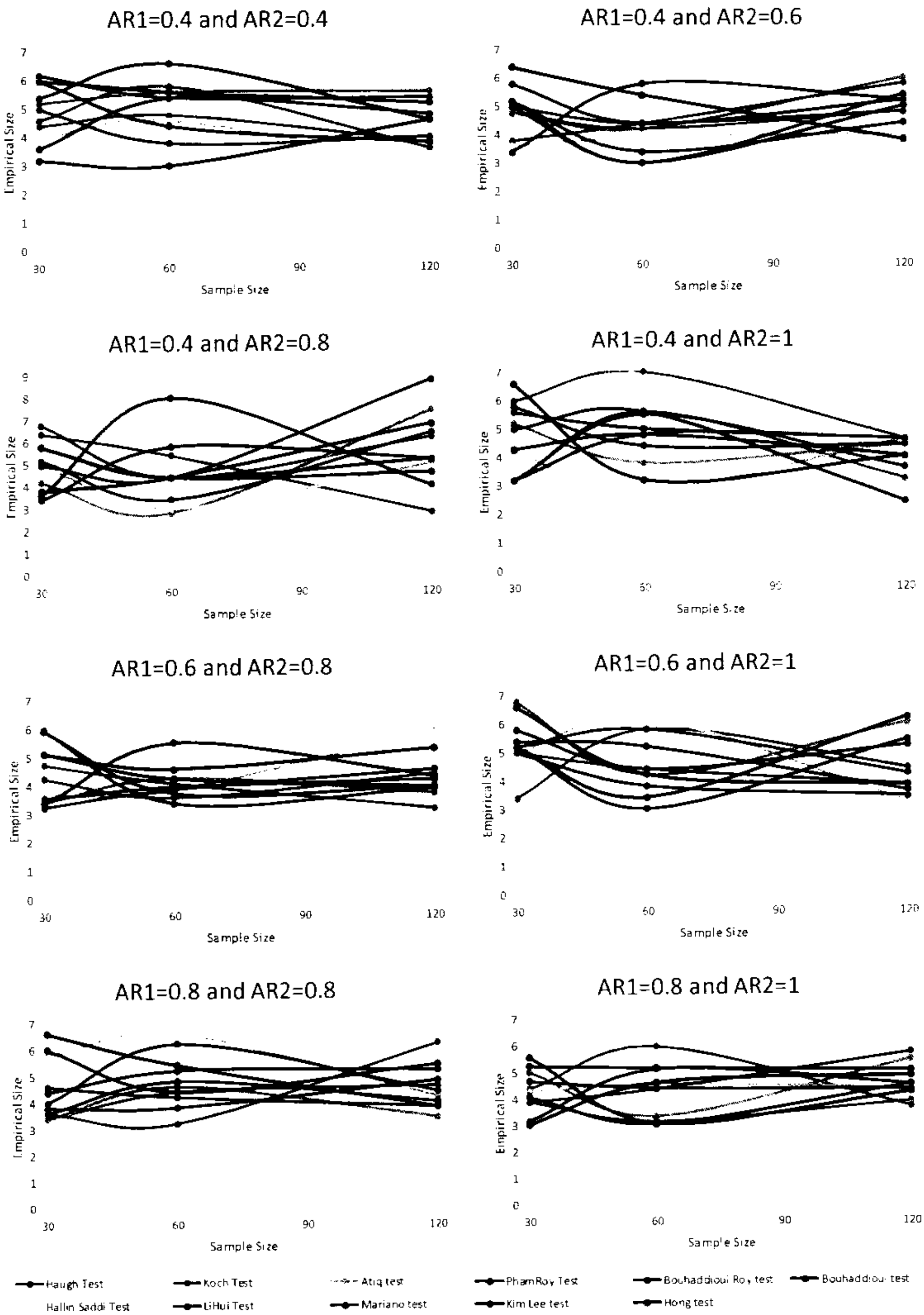


Figure 8.5: MSC for Tests of independence for Time Series with DT

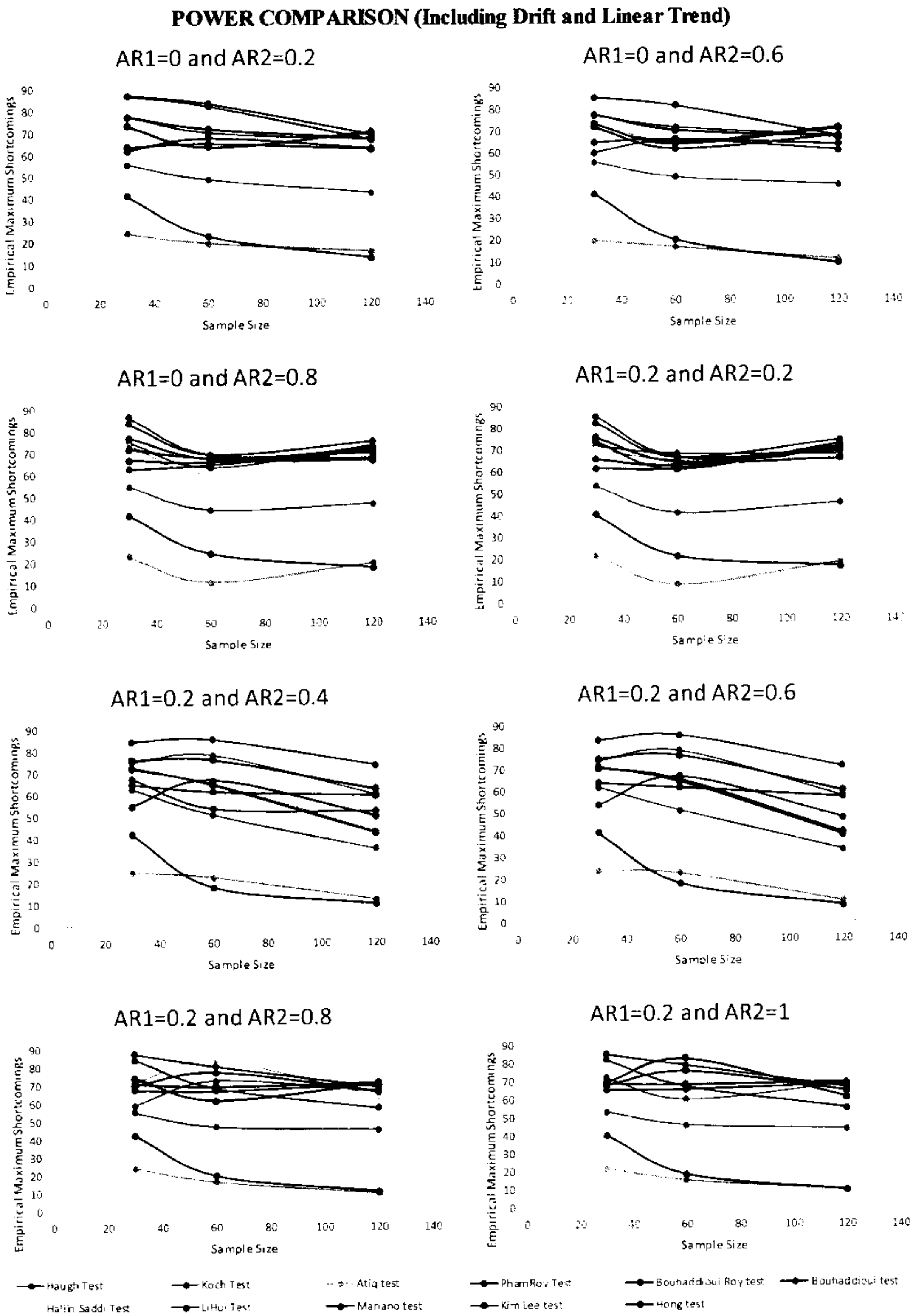


Figure 8.6: MSC for Tests of independence for Time Series with DT (Continue)

POWER COMPARISON (Including Drift and Linear Trend)

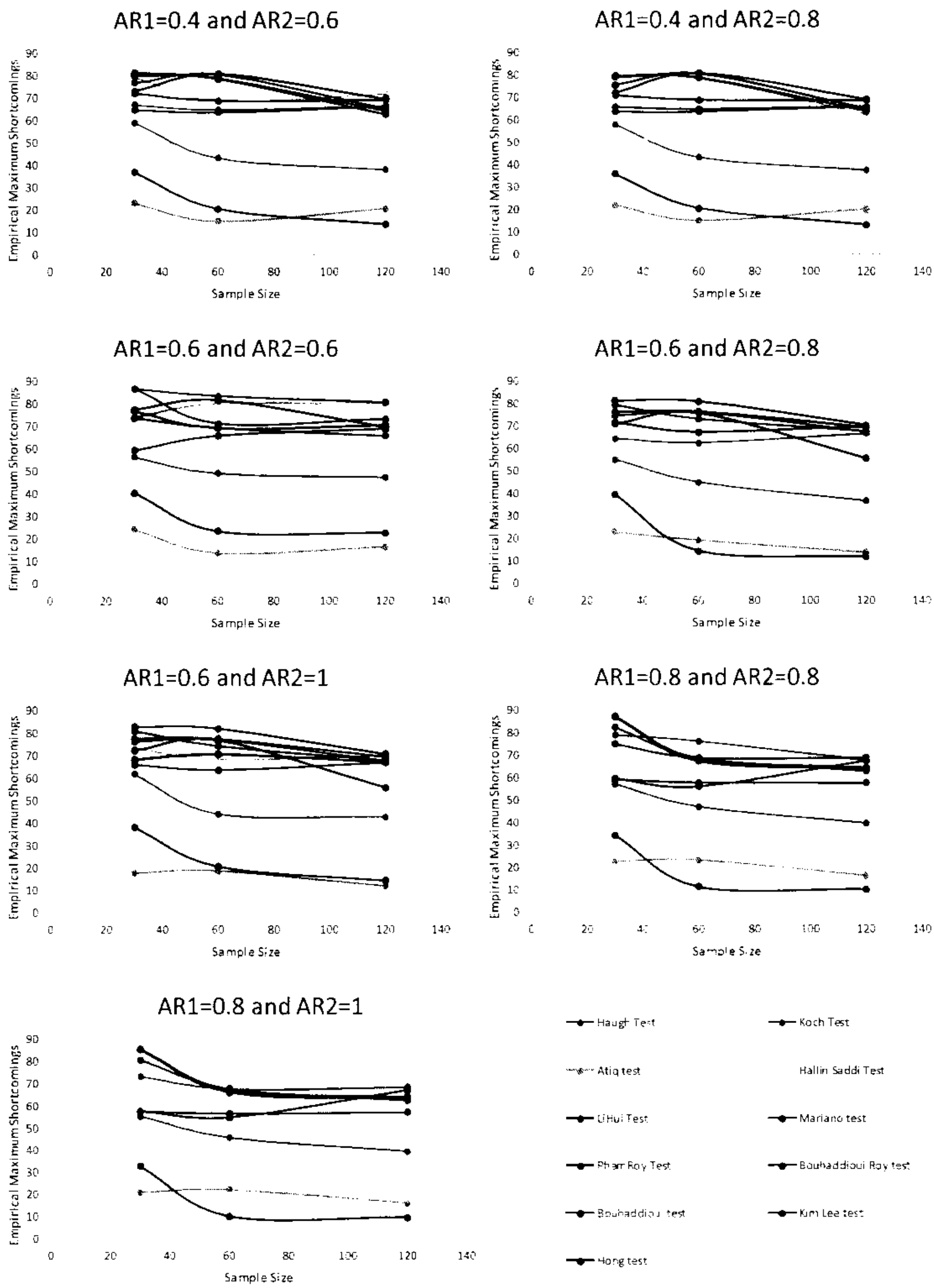


Figure 8.7: MSC for Tests of independence for Time Series without DT

POWER COMPARISON (Without Drift and Linear Trend)

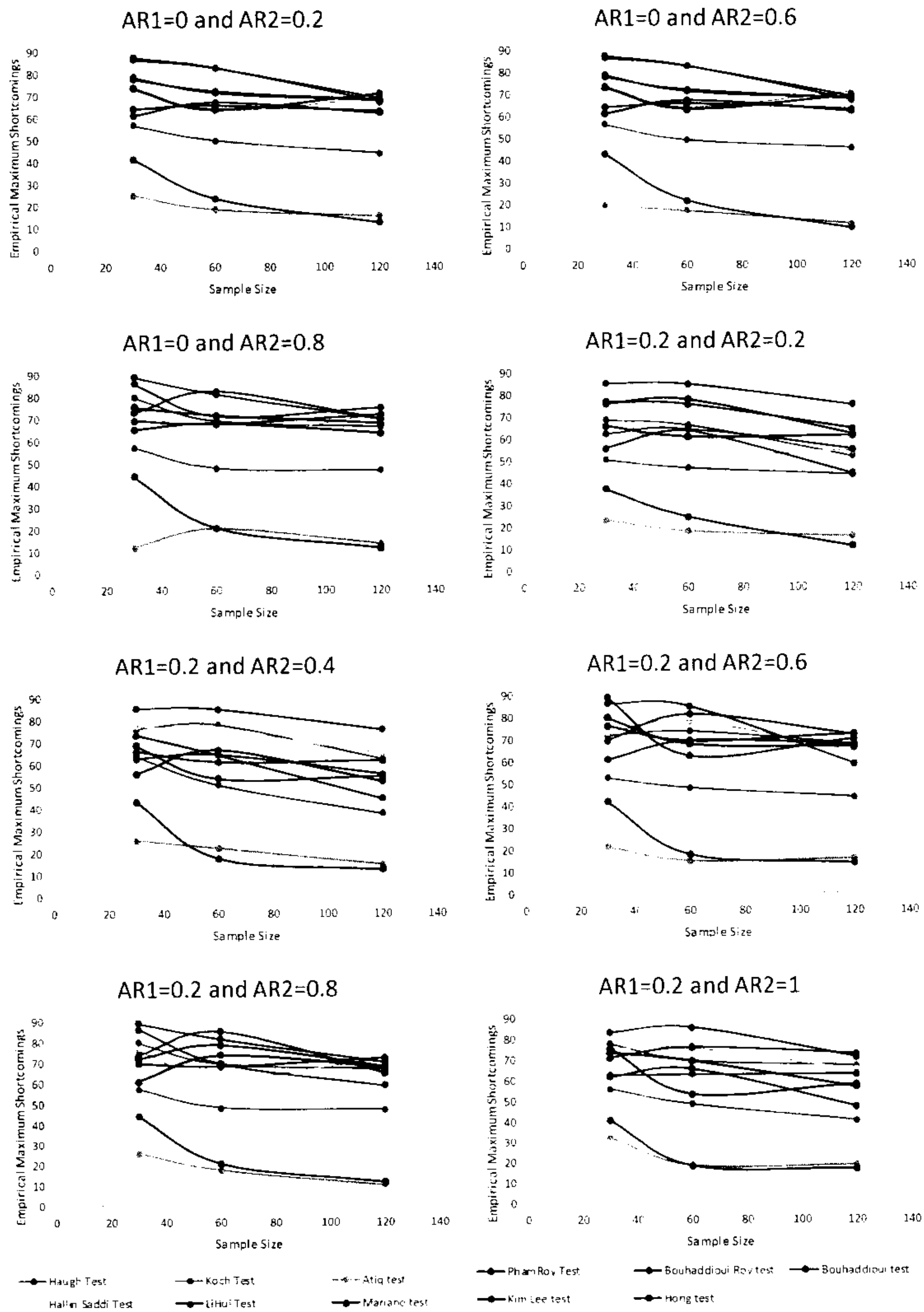


Figure 8.8: MSC for Tests of independence for Time Series without DT (Continue)

POWER COMPARISON (Without Drift and Linear Trend) (Continue)