BOOTSTRAPPING THE AUTOREGRESSIVE-DISTRIBUTED LAG TEST FOR COINTEGRATION

SAM CHUNG YAN

UNIVERSITI SAINS MALAYSIA

2016

BOOTSTRAPPING THE AUTOREGRESSIVE-DISTRIBUTED LAG TEST FOR COINTEGRATION

by

SAM CHUNG YAN

Thesis submitted in fulfillment of the requirements

for the degree of

Master of Social Science

January 2016

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors Dr. Goh Soo Khoon from Centre for Policy Research & International Studies (CenPRIS) and Dr. Sek Siok Kun from School of Mathematical Sciences. Dr. Goh’s patience, understanding, care and support have greatly enriched my graduate studies’ experience. She always puts me at the first place and finds many ways to support and help me. She also employs me as her Research Assistant under the Universiti Sains Malaysia research grant to lessen my financial and living burdens. I appreciate her guidance and support in my thesis writing. In addition, Dr. Goh has helped me tremendously to communicate easily with Professor Dr. Robert McNown when discussing various research problems via email.

I would like to thank Professor Dr. Robert McNown, from University of Colorado at Boulder, Boulder, USA from whom I have learnt a great deal during his stay in CenPRIS USM under the Fulbright Senior Scholar program. He has continued to supervise and guide me even after his visit in USM for which I am most grateful. His expertise and vast knowledge has helped me greatly throughout this thesis. His patience in staying connected with me to discuss the problems is instrumental in helping me to complete my thesis.

A very special thanks goes to Dr. Sek, who without her invitation to this research project I would not have the opportunity to further my study. I must also acknowledge Dr. Tan Yao Sua, who is from CenPRIS, for giving me a chance to work as his Research Assistant where I learnt some analyzing skills from the projects undertaken. Dr. Tan is also a caring and friendly person and I am happy to work under him. Next, I would like to acknowledge Mr. Syed, administration staff from School of Mathematical Sciences, for his patience and kindness in allowing me to complete my data collection and results smoothly from the School’s laboratory. Besides, I also would like to thank the Dean of Mathematical School, Professor Dr. Ahmad Izani Md. Ismail.

In conclusion, I recognize that this research would not have been possible without the financial aid from the research grants from Universiti Sains Malaysia, the MyMaster scholarship from Kementerian Pengajian Tinggi, and the facilities provided by CenPRIS and School of Mathematical Sciences. I express my gratitude to these agencies.

TABLE OF CONTENTS

		PAGE
Acknowledgements		ii
Table of Contents		iii
List of Tables		vi
List of Diagrams		vii
Abstrak		viii
Abstract		x

Chapter 1: Introduction

1.1	Motivation of the Study	1
1.2	Study Objectives	5
1.3	Scope of Study	6
1.4	Significance of Study	7
1.5	Outline of the Thesis Chapters	7

Chapter 2: The ARDL Bounds Test and the Literature Review on Bootstrapping Cointegration

2.1	Overview	9
2.2	Pesaran et al. (2001) ARDL Bounds Testing Approach	9
2.3	Advantages of Bounds Testing Approach	11
	2.3.1 Model Deals with Mixed Integration Order Variables	12
	2.3.2 Single-Equation Model	13
	2.3.3 Super Consistent Properties Underlying ARDL Framework	13
2.4	Assumption 3 of Bounds Testing Approach	15
2.5	The Possibility of Inconclusive Inference	18
2.6	The Existence of Degenerate Case	19
2.7	Literature Review on Bootstrapping Cointegration	22

Chapter 3: Methodology

3.1	Overview	25
3.2	Monte Carlo Simulation	26

3.3 Data-generating process (DGP) 27
3.4 Statistical Models 31
3.4.1 Autoregressive-Distributed Lag (ARDL) 31
3.4.2 Test Statistics 32
3.5 Bootstrap Methods 33
3.5.1 Recursive Bootstrap Algorithm 36
3.6 Size and Power 42
3.6.1 Size and Power Analysis 43
3.7 Simulation Organization 46

Chapter 4: Results and Discussions
4.1 Overview 52
4.2 The DGP Setting 53
4.3 Size and Power Analyses Using Bootstrap Designs A & B 57
4.3.1 Asymptotic Test Analysis 58
4.3.2 Analyses Using Bootstrap Designs A & B 61
4.4 Size and Power Analyses Using Bootstrap Design C 62
4.5 Analyses with Different Combinations of Sample Size and Correlation 65

Chapter 5: Empirical Application: Estimating Saving-Investment Correlations
5.1 Overview 72
5.2 Saving-Investment Correlation 73
5.3 Empirical Model and Methodology 75
5.4 Empirical Results and Discussions 76

Chapter 6: Conclusion
6.1 Overview 90
6.2 Summary of the Motivation, Objectives and Findings of the Study 90
6.3 Limitations and Future Study 93
6.4 Conclusion 94

References

Appendices

LIST OF TABLES

		PAGE
Table 1	Summary of the parameter restriction in the bootstrap procedures for designs	41
Table 2	Parameter combinations used in the DGP simulation	54
Table 3	Cointegration and violation assumption status	57
Table 4	Size and power of asymptotic test (n = 50, ρ = 0.5)	59
Table 5	Size and power of asymptotic test and bootstrap Design C (n = 50, ρ = 0.5)	63
Table 6	Size and power of asymptotic test and bootstrap Design C (n = 50, ρ = 0.0)	66
Table 7	Size and power of asymptotic test and bootstrap Design C (n = 100, ρ = 0.5)	68
Table 8	Size and power of asymptotic test and bootstrap Design C (n = 100, ρ = 0.0)	70
Table 9	Unit root test for each country	78
Table 10	ARDL bounds test estimation using *IR* as dependent variable	80
Table 11	ARDL bounds test estimation using *SR* as dependent variable	85
Table 12	Summary of cointegration status tested at 5% significant level	89

LIST OF DIAGRAMS

PAGE

Diagram 1	Simulation structure and flowchart	47
Diagram 2	Flowchart of bootstrapping	48

BOOTSTRAP UJIAN AUTOREGRESSIVE-DISTRIBUTED LAG UNTUK KOINTEGRASI

ABSTRAK

Objektif tesis ini adalah untuk mengkaji prestasi ujian kointegrasi*: Autoregressive-Distributed Lag (ARDL) Bounds Test* yang dikembangkan oleh Pesaran et al. (2001). Pendekatan ini menjadi popular dan banyak digunakan dalam dua dekad atas kelebihan super konsistent estimasi dan menangani masalah pemboleh ubah bebas yang berintegrasi campur. Namun, ARDL Bounds Test sentiasa disalahgunakan dalam situasi yang tidak konsisten dengan andaian dalam rangka kerja tersebut. Pendekatan ini menganggap tiada kesan maklum balas di tahap dari pemboleh ubah bersandar ke pemboleh ubah bebas. Ini bermakna, salah satu pemboleh ubah mestilah sebagai *weakly exogenous*. Estimasi yang terlibat dalam beberapa kemungkinan pemboleh ubah endogenous yang digunakan sebagai regressors akan memberi keputusan yang keliru dan berat sebelah. Walau bagaimanapun, bukti dari simulasi menunjukkan prestasi pendekatan *Bounds Test* tersebut tidak dipengaruhi oleh andaian masalah *endogeneity*. Dalam tesis ini, kami mencadangkan satu ujian kointegrasi ARDL baru yang bergantung atas prosedur bootstrap. Ia dapat ditunjukkan bahawa jika memperkenalkan prosedur bootstrap dengan betul, sebahagian kelemahan daripada pendekatan ini dapat diatasi dan menambah baik dengan berdasarkan sifat-sifat saiz dan kuasa. Tambahan pula, ia menghapuskan kemungkinan inferensi yang tidak yakin daripada *ARDL Bounds Test*. Selain itu, inferensi berdasarkan signifikan ujian F dan tunggal ujian t sahaja adalah tidak mencukupi untuk mengelakkan kes *degenerate*. Dengan melakukan ujian tambahan daripada langkah yang dicadangkan atas pemboleh ubah bebas, kita dapat gambaran yang lebih baik untuk

menyimpulkan status kointegrasi. Empirikal yang berkait dengan bootstrap ARDL test ditunjukkan daripada anggaran korelasi simpanan-pelaburan dalam tesis ini.

BOOTSTRAPPING THE AUTOREGRESSIVE-DISTRIBUTED LAG TEST FOR COINTEGRATION

ABSTRACT

The objective of this thesis is to examine the performances of a cointegration test: Autoregressive Distributed Lag (ARDL) bounds test approach developed by Pesaran et al. (2001). This approach gained popularity and is widely used for over two decades due to its advantages of super consistent estimation and dealing with mixed integration order regressors. Nevertheless, the ARDL bounds test is often applied in environments that are inconsistent with the assumptions underlying that framework. This approach assumes that there is no feedback at level from the dependent variable to the regressors. That is, all variables except one must be weakly exogenous. Estimation involving several plausibly endogenous variables as regressors will give biased and misleading results. However, through simulation evidence our results show that the performance of the bounds test approach is not affected by this endogeneity problem. In this thesis, we propose a new ARDL cointegration test that relies on the bootstrap procedure. It is shown that by introducing a proper bootstrap procedures, some weaknesses underlying the approach are improved based on size and power properties. In addition, it eliminates the possibility of inconclusive inferences from bounds testing. Besides that, inference based solely on the significance of F test and single t test is insufficient to avoid degenerate case. Conducting an additional testing on the lagged independent variable comes from the proposed method to provide a better insight in concluding the status of cointegration. The empirical relevance of the bootstrap ARDL test is demonstrated by an estimation of saving-investment correlations.

CHAPTER 1

INTRODUCTION

1.1 Motivation of the Study

In the mid 80s, the emergence of the econometric methodology of cointegration attracted considerable attention from economists and researchers. They developed many cointegration tests and approaches. Enger and Granger (1987) first proposed a two-step residual-based procedure with the cointegrating regression estimated by ordinary least squares (OLS) under the null hypothesis of no cointegration. This approach is the first method in testing the existence of cointegration between several variables after Granger introduced the concept of cointegration and presented it at a conference organized in Gainesville, Florida in 1980. Another procedure named fully modified OLS (FMOLS) was developed by Phillips and Hansen (1990) later. In the following year, Engle and Yoo (1991) proposed a 'third step' to the standard Enger-Granger procedure to overcome the flaws in the previous two-step procedure method. Latter work has come out with a system-based reduced rank regression procedure proposed by Søren Johansen which is a procedure in estimating several integrated series in system form to test the cointegration relationship within the system (see Johansen, 1991, 1995). With the rise of cointegration analysis, other approaches such as procedures based on stochastic common system trends by Stock and Watson (1988), variable addition approach by Park (1990) and the residual-based procedure test under null hypothesis of cointegration by Shin (1994) have been considered.

In the beginning of 2000s, Pesaran et al. (2001) (PSS henceforth), developed a new approach for the cointegration test, namely, autoregressive-distributed lag (ARDL) bounds testing in dealing with mixed integration orders of the time series. This approach gained popularity due to its advantages and excellent performances over other cointegration testing approaches. This claim was mentioned by Shahbaz et al. (2013) and Satti et al. (2014). Although this approach is widely used in empirical economic studies, many researchers apply it in environments which are against the assumptions underlying the bounds testing framework. For example, the bounds testing approach assumes that there is no feedback in the levels from the dependent variable to the explanatory variables; in other words, there are no feedback effects from the dependent variable to the explanatory variables in the long-run. PSS states that in a set of variables, the assumption made in bounds testing restricts in that there exists at most one conditional level relationship between the dependent variable and explanatory variables. This indicates that, there is at most one endogenous variable and the rest are exogenous variables (or forcing variables). It is necessary to presume one of the variables as an endogenous variable and the rest as forcing variables. In applications, many applicants apply this approach by letting each of the variables to be the dependent variable in a sequence of regressions on the others. This implicitly allows two or more variables to be (weakly) endogenous and this already violates the assumptions underlying the test statistics presented in PSS.

In this study, the performances of PSS ARDL bounds testing based on size and power properties under various environments will be investigated. Environment designs involved in the study include variables that are designed as endogenous, cases which are against the assumption made in the approach by allowing multilateral feedbacks among variables. In addition, the bootstrap method will be applied to observe if any improvements can be made on the approach. Enormous literature exist on the use of the bootstrap method and it is no longer new. Many studies have shown that bootstrap methods efficiently give less distortion and more precise inferences of the test. Empirical studies with bootstrapping include unit root test, cointegrating regressions, and Granger causality (see Chang and Park, 2003; Li and Maddala, 1997; Ko, 2011).

Besides, in determining the existence of the long-run relationship between the dependent variable and regressors, PSS presented a pair of tests which are the F test that tests the joint significance of the coefficients of lagged level variables and t test on the individual coefficient of lagged level of the dependent variable. We can conclude that the dependent variable and the regressors are cointegrated if and only if both tests individually reject the null hypotheses. However, this is only valid on condition that the dependent variable must be stationary at integration order of one. One degenerate case might happen even though the F and t tests are significant, given that the dependent variable is not $I(1)$. As pointed out by PSS, degenerate cases are situations where either only the lagged dependent variable or lagged independent variable(s) is showing significance in error correction term. Degenerate cases are not cointegration by its incomplete structure of error correction term to adjust the system back to equilibrium. PSS discussed the possibility of degenerate cases by rejecting the null hypothesis with the significance of the F statistic. By assuming the dependent variable in $I(1)$ it can rule out one of the degenerate cases. With the additional test we can determine the cointegration status clearly and make a valid conclusion. Thus, ignoring the integration order of the dependent variable will mislead one to conclude that there is a cointegration, when in fact, there is none. In order to estimate the integration order of the dependent variable, the common method is through the unit root testing method. Nevertheless, it is well-known that unit root testing has low power in determining the existence of unit root especially for a small sample size. This study proposes an alternative method of avoiding the possibility of degenerate cases happening, that is, through an additional test in testing the existence of lagged level of the regressors instead of conducting a unit root test to ensure the dependent variable as $I(1)$. The type of test depends on the number of the regressor. If there is one regressor, the t test is used; if there is more than one regressor, the F test is used. The proposed test critical values are generated through the bootstrap method.

One imperfect feature of the bounds testing approach is, it may give us an inconclusive inference if the estimated F or t statistic falls inside the bounds of the critical values. Knowledge of the integration order of the variables need to be known before drawing a conclusive inference. This study tackles this problem and eliminates the possibility of inconclusive inferences through bootstrap as the generated critical values are based on the exact integration properties of the series. To examine the tests and obtain the findings, a Monte Carlo simulation is developed. Results will be obtained from the simulation.

After the investigation on the performances of both asymptotic and bootstrap ARDL tests, one empirical application will be demonstrated using the proposed method of ARDL testing approach. Estimation on correlation between savings and investment will be used as our empirical application. According to Feldstein and Horioka (1980), the correlation between savings and investment can be treated as an indicator of international capital mobility. The greater the correlation, lower is the capital mobility of one country, conversely, a smaller correlation indicates a greater capital mobility. But this idea is controversial as many findings showed that the correlation is high for most of the developed countries as seen by the findings of Feldstein and Horioka themselves. Besides demonstrating the use of the new proposed method, I also show the possibility of the occurrence of degenerate cases. Misleading conclusions can be made if in applying the bounds test approach is not handled well.

1.2 Study Objectives

The primary objective of this study was to investigate how the endogeneity problem affects the performances of the PSS's ARDL bounds testing approach and how well the bootstrap method improves the ARDL test. The specific objectives were:

1. To examine how well the bounds testing approach performs under a range of environments.

2. To examine how well the bootstrap method improves the ARDL test.

3. To demonstrate that the bootstrap ARDL test on lagged level of independent variable(s) has proper size and power properties.

4. To eliminate the possibility of inconclusive inferences through the bootstrap ARDL test.

5. To demonstrate the possibility of the occurrence of degenerate cases.

1.3 Scope of Study

This study used a simulation method in generating data to conduct the analysis. Simulation is developed by writing a program using the statistical software package EViews 8. Those data are generated according to several data-generating processes (DGPs) setting.

Methodologies involved in this study included Monte Carlo experiments, recursive bootstrap method, and OLS estimation method in ARDL model. The simulation named "Simulation_bootstrap" was developed and all the methodologies mentioned were included in the simulation. The program was used to conduct the analyses. The Monte Carlo experiment created $N = 2000$ replications under 16 different environments in testing the ARDL test performances based on size and power properties.

With each n_i replication in the experiment, there were $B = 999$ bootstrap replications. Asymptotic F statistic and t statistic on individual lagged level of the dependent variable were obtained as well as bootstrap statistics including the bootstrap test statistic on lagged level of the independent variable through estimation. Testing was carried out by letting each variable be the dependent variable in a sequence of regressions and naming them as Equation Y if variable Y is treated as the dependent variable and Equation X if X is treated as the dependent variable. Distributions of these estimated statistics were formed. Size and power properties of the statistics were studied from the distributions.

1.4 Significance of Study

The study of this thesis has provided several contributions and its significance covers the performances of the ARDL bounds test under the violation of endogeneity assumption, the performance of bootstrap method, and the existence of degenerate cases. The specific contributions are listed as follows:

1. This study shows that the performances of the ARDL bounds test are not affected by the assumption of endogeneity.

2. The bootstrap method if applied properly, helps to improve the ARDL test estimation.

3. The bootstrap ARDL test eliminates the inconclusive inferences underlying the bounds test.

4. This study demonstrates the occurrence of degenerate cases.

5. This study proposes an effective alternative testing to avoid the occurrence of the degenerate cases.

1.5 Outline of the Thesis Chapters

The study is organized as follows: Section 2, discusses the pros and cons of PSS ARDL bounds testing methodology, endogeneity problems and literature reviews. Next, Section 3 introduces the methodologies used in the study including the DGP setup, recursive bootstrap method, size and power properties and Monte Carlo experiment. Simulation's structure flow is discussed in this section as well. Then, Section 4 presents the empirical findings and discussions. Section 5 discusses an application in estimating the Saving-Investment correlation to demonstrate the use of the bootstrap ARDL test. This section also shows the possibility of the occurrence of degenerate cases. Lastly, Section 6 concludes the findings of this study.

CHAPTER 2

THE ARDL BOUNDS TEST AND THE LITERATURE REVIEW ON BOOTSTRAPPING COINTEGRATION

2.1 Overview

This chapter discusses PSS ARDL bounds testing approach and the literature review on bootstrapping cointegration. Section 2.2 discusses the use of the bounds testing approach and how it is used in determining an existence of cointegration. Section 2.3, discusses the advantages underlying bounds testing. The assumptions and limitations of the test are discussed in Section 2.4. This section elaborates on how researchers conducted the bounds test without considering the endogeneity assumption and its potential problems. Section 2.5 discusses the possibility of inconclusive inference. Next, Section 2.6 discusses the existence of degenerate cases and the possibility of making a misleading conclusion on cointegration. Section 2.7 discusses the literature reviews on bootstrapping cointegration. Lastly, discussion on the proposed alternative method in applying the bounds test through the bootstrap procedure is shown in Section 2.8.

2.2 Pesaran et al. (2001) ARDL Bounds Testing Approach

Before the 1980s, many economists were using linear regression on de-trended non-stationary time series data in analysis. But Granger and Newbold (1974) showed that this is a dangerous way in analyzing the data and it might produce a spurious correlation. Later, Granger (1981) first introduced the concept of cointegration which attracted attention from economists and researchers. Many researches on cointegration are carried out and a number of approaches were developed in dealing with cointegration. This includes the well-known PSS ARDL bounds testing approach.

ARDL bounds testing approach is inspired by the works of Banerjee et al. (1998). They were the first in proposing cointegration testing with a single-equation framework, based on the early contributions of Banerjee et al. (1986) and Kremers et al. (1992). In 2001, PSS proposed the bounds procedure based on the standard Wald or F and t statistics to test the significance of the lagged levels of the variables in a univariate equilibrium correction mechanism. The table of bounds testing critical values based on the non-standard asymptotic distribution under the null hypothesis of absence level relationship, with covering all the possible classifications of regressors which are purely $I(0)$, purely $I(1)$ or mutually cointegrated is provided. If one computed the Wald or F statistic in testing the joint significance of the variables with one lagged period level coefficients which is greater than the upper bound of the critical values, then the null of no cointegration is rejected, without knowing whether the regressors are purely $I(0)$, $I(1)$ or mutually cointegrated; if it is less than the lower bound critical values, then, the null is not rejected. If the computed F statistic falls between the bounds, inference is inconclusive. Determination of the order of the variables is needed to make the inferences. If the F test rejects the null, then proceed to the test of existence effect of the lagged dependent variable. If the computed t statistic of the lagged dependent variable is greater than the upper bound critical value as well, this confirms that there exists a level relationship between y_t and $\mathbf{x}_t$, with a condition that the dependent variable is in $I(1)$. Once the existence of the level relationship between y_t and $\mathbf{x}_t$ is determined, a ARDL model is built with orders $(p,p,\ldots,p)$ or ARDL$(p,p,\ldots,p)$, or a model is built specifically with a different order with ARDL$(p,p_1,\ldots,p_k)$ to determine the long-run and short-run relationship.

Narayan (2005) employed the PSS bounds testing approach in his research and argued that the provided critical values are not applicable in his research with a small sample size. PSS in generating the critical values based on asymptotic distribution was using 500 and 1000 observations with 20,000 and 40,000 replications, respectively. He showed that the upper bound critical value at 5% significance level for 31 observations with 4 regressors is 4.13 while the corresponding critical value for 1,000 observations is 3.49. Thus, PSS critical values are not accurate when it is used on a small sample size test. Hence, he generated a new set of critical values based on a small sample size from observations of 30 to 80. Narayan's contribution corrected biases that may be caused by the bounds testing approach in the test of a small sample size. He made the bounds testing approach more comprehensive.

2.3 Advantages of PSS Bounds Testing Approach

The emergence of the PSS bounds testing approach has attracted interest from researchers whose studies are related to the test of cointegration. The bounds testing approach gained its popularity over its outperforming performances. PSS bounds testing approach has some advantages over the other cointegration testing approaches, i.e., it can be used for regressors with a mixture of $I(0)$ and $I(1)$, involving only a single-equation setup and is easy to implement by its simple design, and the model can be set up with different variables with different lag lengths (general-to-specific framework). Moreover, the test provides better results with a small sample size compared to conventional cointegration approaches (Haug, 2002).

2.3.1 Model Deals with Mixed Integration Order Variables

In cointegration tests such as the Engle-Granger two-step residuals based cointegrating regression test, Phillips-Ouliaris two residual-based test, Stock and Watson's approach based on stochastic common system trends etc, it is assumed that all the tested variables must be non-stationary and integrated with the same order with at least order one. This restriction ensures that all the variables need to go through a preliminary testing to make sure the interest variables are having unit root and the same degree of integration. As for the conventional unit root testing methods, there is the augmented Dickey-Fuller (ADF) and Phillip and Perron (PP) unit root tests. In the test of cointegration, one of the crucial issues is to see if there are mixed degrees of integration. This is common in macroeconomics research and if this happens, the above mentioned cointegration testing approaches are inapplicable. PSS saw the weaknesses in having these cointegrating approaches and developed the bounds testing that is able to determine the existence of a level relationship between y_t and $\mathbf{x}_t$ regardless of whether the regressors are either purely $I(0)$, $I(1)$ or mixed orders. Inferences can be made by only using the computed Wald or F statistic and t statistic. Sets of critical values are provided by PSS in carrying out the hypothesis testing. However, this approach is not appropriate for multicointegration or variables which have a higher order than unity. Bounds test critical values are generated based on regressors with $I(0)$ and $I(1)$ only. Estimations that involve regressors with higher order than unity will lead to spurious results. Thus, a unit root test needs to be carried out to ensure that all the regressors' integration orders are not higher than unity.

2.3.2 Single-Equation Model

In contrast to the system cointegration tests, for example, the Banerjee approach and bounds testing approach only involve a single-equation setup in studying the level relationship between the variables by assuming only one variable as endogenous and the others as forcing variables (or weakly exogenous). A single-equation model is simpler and easier to apply compared to the system equation which involves many cross-sections and complicated estimations. Banerjee et al. (1998) were the first to propose a single-equation model in cointegration testing, but it is only valid for regressors which are $I(1)$. In addition, PSS pointed out that the bound test model is also allowed for differential lag lengths for each of the lagged variables involved. The flexible structure of the model will not affect the asymptotic properties. The specification of the lag length of the model gives a more accurate estimation in explaining the relationship of the study variables.

2.3.3 Super Consistent Properties Underlying ARDL Framework

The bounds testing approach adopts a traditional ARDL model or unrestricted ECM model as its framework. Pesaran and Shin (1999) proved that estimation using the estimator ordinary least square (OLS) in the ARDL model framework estimates the short-run parameters as consistent and for long-run is super consistent even in a small sample size. Besides that, Engle et al. (1983) also proved that, regressors with weak exogeneity for the parameters of interest are sufficient for OLS to provide asymptotically efficient estimates of the parameters in the conditional ARDL model. Under assumptions, $\mathbf{x}_t$ can be treated as weakly exogenous to the parameters of the estimation equation (see PSS Assumption 4, page 293-294). Therefore, estimation using the OLS method is consistent. As is known, most of the statistical approach estimations will become statistically less powerful when dealing with small sample size data.

In addition, Kremers et al. (1992) showed that in testing for cointegration, the ECM framework estimation can generate a more powerful test than the regression framework by avoiding the invalid common factor restriction. Compared to the dynamic model underlying the ECM framework, the model with the regression framework characterized by the Dickey-Fuller test implicitly imposed a common factor restriction which ignores potentially valuable information. Problems will still remain even if it includes additional variables, additional lags of variables, constant term seasonal dummies or a more sophisticated cointegrating vector.

Underlying the bounds testing approach is a more statistical powerful ARDL framework which gives more accurate inferences in the test of a small sample size. This is one of the main advantages over the use of other cointegrating approaches, especially for empirical analysis in economic studies in which the data involved are usually small. A more convenient and statistical powerful approach introduced by PSS has given researchers a pleasant experience in conducting cointegration testing. These are the reasons why this approach gained popularity. Authors such as Calza and Zaghini (2010), Fielding (2003), Garg and Dua (2014), and Ibarra (2011) employed the PSS ARDL bounds testing to carry out cointegration tests.

2.4 Assumption 3 of Bounds Testing Approach

PSS (2001) used 5 assumptions in developing bounds testing. Some researchers may miss out one of the crucial assumptions i.e. Assumption 3 which is spelled out by Pesaran et al. (2001), page 293.

Let us consider a (k+1)-VAR model of order p:

$$\mathbf{\Phi}(L)(\mathbf{z}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} t) = \boldsymbol{\varepsilon}_t, t = 1, 2, \ldots \tag{1}$$

where L is the lag operator, $\{\mathbf{z}_t\}_{t=1}^{\infty}$ is a (k+1) random process and can be partitioned into $(y_t, \mathbf{x}_t)'$, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are unknown (k+1)-vectors of intercept and trend coefficients, the $(k+1, k+1)$ matrix lag polynomial $\mathbf{\Phi}(L)$ is equal to $\mathbf{I}_{k+1} - \sum_{i=1}^{p} \mathbf{\Phi}_i (L)^i$, where $\mathbf{I}_{k+1}$ an identity matrix or order k+1, with $\{\Phi_i\}_{i=1}^{p}$ $(k+1, k+1)$ matrices of unknown coefficients. By considering PSS Assumptions 1 and 2, i.e. Assumption 1: *the roots of* $\left|\mathbf{I}_{k+1} - \sum_{i=1}^{p} \mathbf{\Phi}_i z^i\right| = 0$ *are either outside the unit circle* $|z| = 1$ *or satisfy* $z = 1$ and Assumption 2: *the vector error process* $\{\boldsymbol{\varepsilon}_t\}_{t-1}^{\infty}$ *is* $IN(\mathbf{0}, \mathbf{\Omega}), \mathbf{\Omega}$ *positive definite*, the VAR(p) in Equation (1) is derived into a system of conditional ECM as follows:

$$\Delta y_t = c_0 + c_1 t + \pi_{yy} y_{t-1} + \boldsymbol{\pi}_{yx} \mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i' \Delta \mathbf{z}_{t-1} + \boldsymbol{\omega}' \Delta \mathbf{x}_t + u_t, \ t = 1, 2, \ldots \tag{2}$$

$$\Delta \mathbf{x}_t = \mathbf{c}_0 + \mathbf{c}_1 t + \boldsymbol{\pi}_{xy}\mathbf{y}_{t-1} + \boldsymbol{\pi}_{xx}\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}'_{xi}\Delta \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_{xt},\ t = 1, 2, \ldots \tag{3}$$

where π matrix denotes the long-run multiplier, $\boldsymbol{\psi}'_i$ and $\boldsymbol{\Gamma}'_{xi}$ are the short-run multiplier, Δ is the difference operator, $\boldsymbol{\omega}'$ is the coefficient in respect to $\Delta \mathbf{x}_t$, u_t and ε_{xt} are the errors. PSS (2001) states Assumption 3 as: the *k*-vector $\boldsymbol{\pi}_{xy} = \mathbf{0}$, i.e. there is no feedback from the level of y_t in the system of conditional unrestricted ECM of $\mathbf{x}_t$, but it does not impose similar restrictions on the short-run multipliers in the equations for $\mathbf{x}_t$. Under Assumption 3, the Equation (3) becomes

$$\Delta \mathbf{x}_t = \mathbf{c}_0 + \mathbf{c}_1 t + \boldsymbol{\pi}_{xx}\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}'_{xi}\Delta \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_{xt},\ t = 1, 2, \ldots$$

This assumption restricts vectors $\mathbf{x}_t$ as *forcing variables* by process $\{\mathbf{x}_t\}_{t=1}^{\infty}$ long-run forcing for $\{y_t\}_{t=1}^{\infty}$. The reason behind this assumption is, irrespective of the level of integration of the process $\{\mathbf{x}_t\}_{t=1}^{\infty}$, it restricts consideration to cases in which there exists at most one conditional level relationship between y_t and $\mathbf{x}_t$. Under Assumption 3, the conditional ECM (2) now becomes

$$\Delta y_t = c_0 + c_1 t + \pi_{yy} y_{t-1} + \boldsymbol{\pi}_{yx.x}\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\psi}'_i \Delta \mathbf{z}_{t-1} + \boldsymbol{\omega}'\Delta \mathbf{x}_t + u_t.$$ [1]

[1] Due to its complexity of notations involved in the derivation of the model, the details of the coefficient $\boldsymbol{\pi}_{yx.x}$ are not discuss here. For further information, please refer to Pesaran et al. (2001), page 293.

This is the crucial assumption that supports the PSS methodology. By incorporating Assumption 3 together with other assumptions, one can detect cointegration irrespective of the level of integration of the regressors.

The PSS framework assumes weak exogeneity of the regressors and it is necessary a priori or presuming that one variable is endogenous and the rest are exogenous. These regressors are not receiving any effects from the dependent variable in the long-run, but it does not preclude that there are level relationships among the regressors or endogenous variable being Granger causal to the regressors. Since the asymptotic distributions presented by PSS builds on Assumption 3, its violation of this assumption may lead to invalid results. Unfortunately, this important assumption is always being ignored in empirical studies, for instance, Shahbaz et al. (2013), Satti et al. (2014), Blotch et al. (2015), and Baharumshah et al. (2009). They violated this particular assumption when applying the bounds testing approach. For example, Shahbaz et al. did not presume one variable as endogenous and set the others as exogenous. They had tested 4 different equations using variables: economic growth, terrorism, capital stock, and trade openess by subjecting each of these variables as the dependent variable and regressing on the rest. This implicitly allows these variables to be endogenous. Violation of the assumption is shown through its findings. Through the significance of the F (Wald) statistics, they concluded the there was the presence of 3 cointegrating vectors and validated the existence of a long-run relationship between economic growth, terrorism, capital stock, and trade openness in Pakistan over the period of 1973 to 2010. The findings show that these variables are endogenous and the direction of the feedbacks in level are moving among each other. The same issue happened to Blotch et al., Satti et al. and Baharumshah et al. as well.

It is widespread to assume all the variables as endogenous; in economics, all the economics terms are plausibly endogenous and giving effects to each other in level. It is not known whether if the violation of Assumption 3 happens, the estimation in the bounds testing is still consistent and reliable. Performances of the test under the above situation will be studied based on the size and power properties.

2.5 The Possibility of Inconclusive Inference

Although the bounds testing approach is superior in inferences of cointegration among the series, it has a flaw in that it can possibly come out as an inconclusive result. PSS provided sets of two polar asymptotic critical values i.e. $I(0)$ and $I(1)$. If the computed F statistic falls outside the bounds, conclusive inference can be made without knowing the integration order of the underlying regressors. On one hand, if the F statistic falls outside the lower bound $I(0)$, this indicates that there is no cointegration or there is an absence of level relationship between the dependent variable and the regressors by not rejecting the null hypothesis. On the other hand, if it falls outside the upper bound $I(1)$, this indicates that there is cointegration or existence of level relationship between the dependent variable and the regressors by rejecting the null hypothesis. However, if the F statistic falls between the bounds, it indicates inconclusive inference. If this case happens, knowledge of the order of the integration of the underlying regressors is required before a conclusive inference can be drawn. The possibility of indeterminacy is disturbing to the applicants.

In this study, proposing the indeterminacy of the status of cointegration can be eliminated through the bootstrap procedure. The bootstrap procedure generates critical values that are specific to the integration properties of each specific data. Thus, exact critical values of those specific to the integration properties in testing cointegration are generated and the possibility of an indeterminate test outcome is eliminated.

2.6 The Existence of Degenerate Case

In the test of cointegration by using the ARDL bounds testing approach, PSS pointed out a special case named degenerate case. It may happen even if the null hypothesis of no cointegration is rejected under the F test. PSS introduced the test in the absence of any level relationship between y_t and $\mathbf{x}_t$ by defining the null hypothesis $H_0 = H_0^{\pi_{yy}} \cap H_0^{\pi_{yx.x}}$, where $H_0^{\pi_{yy}} : \pi_{yy} = 0,\ H_0^{\pi_{yx.x}} : \boldsymbol{\pi}_{yx.x} = \mathbf{0}'$, whereas the alternative hypothesis is defined as $H_1 = H_1^{\pi_{yy}} \cup H_1^{\pi_{yx.x}}$ and it covers not only $\pi_{yy} \neq 0,\ \boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$ but also permits $\pi_{yy} \neq 0,\ \boldsymbol{\pi}_{yx.x} = \mathbf{0}'$ or $\pi_{yy} = 0,\ \boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$. Cases when $\pi_{yy} \neq 0,\ \boldsymbol{\pi}_{yx.x} = \mathbf{0}'$ and $\pi_{yy} = 0,\ \boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$ PSS named them as the degenerate level relationship between y_t and $\mathbf{x}_t$. The former case we call it as degenerate case #1 and the latter is named as degenerate case #2. Degenerate case #1 is defined as the situation where only the lagged dependent variable is significant but not for lagged independent variable(s). Degenerate case #2 is defined as only significant for lagged independent variable(s) but not for lagged dependent variable. Cointegration occurs only when $\pi_{yy} \neq 0$ and $\boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$, degenerate cases are not considered as cointegration.

In testing for the existence of a level relationship between y_t and $\mathbf{x}_t$, the F test has to be jointly significant. Significance of the test singly is insufficient, as it is only rejecting the null of $H_0 = H_0^{\pi_{yy}} \cap H_0^{\pi_{yx.x}}$ and favouring $H_1 = H_1^{\pi_{yy}} \cup H_1^{\pi_{yx.x}}$. Remember that the alternative hypothesis does contain the possibility of a degenerate level relationship. The t test null hypothesis is set as $H_0 = \pi_{yy} = 0$ against $H_1 = \pi_{yy} \neq 0$. PSS explained that we can conclude that y_t and $\mathbf{x}_t$ are cointegrated and that a level relationship does exist if and only if the F test is significant and in addition, the lagged level dependent variable t test shows significance or shows $\pi_{yy} \neq 0$ as well. However, this is valid only if y_t is stationary at integration order one because when y_t is in $I(1)$, in the event of a possibility of a degenerate case #1, $\pi_{yy} \neq 0, \ \boldsymbol{\pi}_{yx.x} = \mathbf{0}'$, will be eliminated. This is because degenerate case #1 is the case same as the Dickey-Fuller equation, suggesting that the variable is in $I(0)$. Hence, rejecting the null of the F test while favouring the alternative hypothesis remains as cases of $\pi_{yy} = 0, \ \boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$ and $\pi_{yy} \neq 0, \ \boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$ can happen. If y_t is $I(0)$ and $\mathbf{x}_t$ is $I(1)$, there can be no cointegration even though both F and t statistics are significant.

To confirm the status of a level relationship between y_t and $\mathbf{x}_t$, both the F test on joint significance of lagged level variables and the t test on individual lagged level dependent variable are needed. Failure to conduct the t test in testing the significance of the lagged level of the dependent variable may lead to a degenerate case. Papers by Alhassan and Fiador (2014), Garg and Dua (2014), Bloch et al. (2015), Shahbaz et al. (2013), and Getnet et al. (2005) show that neglect to carry out the t test in their investigations. Degenerate cases can possibly happen if it is found that the F statistic is significant. Moreover, the information of the integration order of y_t is important. Even if this information is ignored, the chances of the degenerate case #1 occurring still remains even if both the F and t tests are found significant, if y_t is $I(0)$. Jiang and Nieh (2012) in their study were concerned about the endogeneity problem to make sure that there is only one unique long-run relationship by chosing one variable as the dependent variable while the others are weakly exogenous. However, they failed to conduct the t test and the dependent variable was tested as $I(0)$. Morley (2006) in his study using the conventional ADF test found that the variables of per capital GDP from three countries were in $I(1)$ but for immigrations were in either $I(0)$ or $I(0)/I(1)$ borderline for immigrations. He had tested the models by taking these two variables as the dependent variable. However, it was incorrect for him to proceed with the test for the equation using the immigration variable as the dependent variable and regressing on per capital GDP. In the long-run relationship robustness testing, Muscatelli and Spinelli (2000) merely conducted the F test in the PSS bounds test to confirm that the variables are cointegrated and a long-run demand for money relationship exists. Ibarra (2011) did not show the results of the unit root test of variables and only mentioned their integration order are not more than unity. This is dangerous as degenerate case #1 might happen if the dependent variable is $I(0)$ resulting in misleading inferences. From the literature review, Goh and McNown (2015) is the only paper that has found a degenerate case in applying the PSS ARDL bounds testing in their study. They examined whether the Malaysian interest rate is cointegrated with that of the US during the recent managed float exchange rate regime. They found that although the F and t tests produced significant statistics, the t statistic for the lagged independent variable (i.e. US interest rate) was insignificant. Hence, this suggests that the lagged level of the dependent variable is the source for the overall statistical significance of the lagged level, suggesting that the dependent variable is actually stationary.

Therefore, information on the integration order of the variables is important for the bounds testing approach to be used appropriately. Besides, not only is the integration order of the regressors needs to be tested, it is necessary to ensure that y_t has to be $I(1)$ so that the degenerate case can be avoided and a valid conclusion can be made.

2.7 Literature Review on Bootstrapping Cointegration

Bootstrap methods on cointegration tests are commonly used in recent years. Authors such as Harris and Judge (1998) used the bootstrap on the Johansen cointegration test in a small sample size testing, Palm et al. (2010) used the sieve bootstrap on the single-equation conditional error-correction model, Seo (2006) conducted residual-based bootstrap testing for the null of no cointegration in a threshold vector error correction model, and Trenkler (2009) analyzed the vector error correction model of cointegration test with prior adjustment for deterministic terms using the bootstrap etc. They confirmed that bootstrap methods do help in improving the size properties of the cointegration tests. Swensen (2006) explained the use of recursive bootstrap on the Johansen cointegration test theoretically. He also showed that the limiting distribution of the bootstrap version of the Johansen test is the same as the asymptotic version Johansen test. Moreover, the size of the test can be improved by bootstrap in a small sample size with the evidence of simulation runs. The same investigation was conducted by Chang et al. (2006) and they developed some of the theoretical aspects for bootstrapping cointegrating regressions. They showed that the OLS estimator relying on the regressions in bootstrap procedure is consistent and asymptotically valid. The OLS estimator with size refinement properties produce an asymptotic unbiased test.

Based on the findings and theories from the literature, the bootstrap method is used in this study. Besides that, the finding of Palm et al. (2010) are supportive of my research as its framework of single equation conditional ECM is similar to the ARDL model. However, Palm et al.'s model is not concerned about the endogeneity problem underlying the bounds test. Hence, this study aims to fill in the literature gap. Furthermore, bootstrap on ARDL bounds testing can do more than merely improving the size and power of the test. In this study, an alternative method is proposed to the PSS bounds testing approach.

As mentioned, in carrying out the ARDL bounds testing the integration order of y_t needs to be *I*(1) and this can be determined by a unit root test. Nevertheless, there is a problem in the unit root test too. It is notorious for unit root tests to have a low-power estimation and the use of the existing unit root tests in identifying the integration order of one series is highly questionable (see Rudra et al., 2013, page 916). In theory, it is common for one variable to be predicted as stationary, but the tests suggest otherwise. For instance, the inflation and interest rate. In the article by Muscatelli and Spinelli (2000) in the test of the integration order of the study variables, by using the ADF test, it was found that the tested variable *INF* (rate of inflation) was *I*(0). However, they argued and treated the variables as *I*(1) by citing studies on the Fisher effect which showed that inflation and interest rates exhibit a common stochastic trend during policy regimes when they display trends. This shows how unreliable the unit root test is.

Besides by assumption, consideration of the degenerate case #1 i.e. only significant of the lagged dependent variable in the error correction term can be ruled out as long as y_t is tested as $I(1)$ and a conclusion can be made on the status of the level relationship between y_t and $\mathbf{x}_t$ with the F test and t test. The alternative method is proposed by adding an additional testing on $\boldsymbol{\pi}_{yx.x}$ together with the existing F and t tests to draw the conclusion. This alternative method has no need for a low-power unit root test to be conducted in determining the integration order of y_t, rather, to conduct an extra testing to determine the significance of $\boldsymbol{\pi}_{yx.x}$. The null hypothesis of the test is defined as $H_0 : \boldsymbol{\pi}_{yx.x} = \mathbf{0}'$ against $H_1 : \boldsymbol{\pi}_{yx.x} \neq \mathbf{0}'$. The proposed critical values of the test are generated through the bootstrap approach. This additional test on the lagged level of explanatory variables is either through the F test or t test depending on the number of regressors.