Journal of Statistical and Econometric Methods, vol.5, no.4, 2016, 63-91
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2016

Autoregressive Distributed Lag (ARDL) cointegration technique: application and interpretation

Emeka Nkoro[1] and Aham Kelvin Uko[2]

Abstract

Economic analysis suggests that there is a long run relationship between variables under consideration as stipulated by theory. This means that the long run relationship properties are intact. In other words, the means and variances are constant and not depending on time. However, most empirical researches have shown that the constancy of the means and variances are not satisfied in analyzing time series variables. In the event of resolving this problem most cointegration techniques are wrongly applied, estimated, and interpreted. One of these techniques is the Autoregressive Distributed Lag (ARDL) cointegration technique or bound cointegration technique. Hence, this study reviews the issues surrounding the way cointegration techniques are applied, estimated and interpreted within the context of ARDL cointegration framework. The study shows that the adoption of the

[1] Department of Economics, University of Port Harcourt, Port Harcourt, Nigeria.
E-mail: nkoro23@yahoo.co.uk
[2] Department of Economics, University of Port Harcourt, Port Harcourt, Nigeria.

Article Info: *Received* : September 2, 2016. *Revised* : October 4, 2016.
Published online : December 1, 2016.

ARDL cointegration technique does not require pretests for unit roots unlike other techniques. Consequently, ARDL cointegration technique is preferable when dealing with variables that are integrated of different order, I(0), I(1) or combination of the both and, robust when there is a single long run relationship between the underlying variables in a small sample size. The long run relationship of the underlying variables is detected through the F-statistic (Wald test). In this approach, long run relationship of the series is said to be established when the F-statistic exceeds the critical value band. The major advantage of this approach lies in its identification of the cointegrating vectors where there are multiple cointegrating vectors. However, this technique will crash in the presence of integrated stochastic trend of I(2). To forestall effort in futility, it may be advisable to test for unit roots, though not as a necessary condition. Based on forecast and policy stance, there is need to explore the necessary conditions that give rise to ARDL cointegration technique in order to avoid its wrongful application, estimation, and interpretation. If the conditions are not followed, it may lead to model misspecification and inconsistent and unrealistic estimates with its implication on forecast and policy. However, this paper cannot claim to have treated the underlying issues in their greatest details, but have endeavoured to provide sufficient insight into the issues surrounding ARDL cointegration technique to young practitioners to enable them to properly apply, estimate, and interpret; in addition to following discussions of the issues in some more advanced texts.

Keywords: Cointegration; Unit Roots; the Autoregressive Distributed Lag Cointegration technique; Error Correction Mechanism

JEL Words: C5; C51; C52

1 Introduction

Theoretically, economic analysis suggests that there is a long run relationship between variables under consideration. Oftentimes, econometricians/researchers have ignored the inherent dynamic features of most time series in the process of analyzing time series and formulating traditional regression models. It was assumed that the underlying time series were stationary or at least stationary around a deterministic trend and as well exhibited a long run relationship. Hence, it was normal to formulate an econometric model in the conventional way assuming that the means and variances of the variables were constant and not depending on time. Thus, the estimated models were used to analyze theories formulated at abstract level and, to forecast, evaluate and stimulate policies.

Recent development in econometrics have however, revealed that often times, most time series are not stationary as was conventionally thought. Therefore, different time series may not display the same features. Hence, it is possible to see some time series that display the feature of diverging away from their mean over time while others may converge to their mean over time. Time series that diverge away from their mean over time are said to be non-stationary. Therefore, the classical estimation of variables with this relationship most times gives misleading inferences or spurious regression.

To overcome this problem of non-stationarity and prior restrictions on the lag structure of a model, econometric analysis of time series data has increasingly moved towards the issue of cointegration. The reason being that, cointegration is a powerful way of detecting the presence of steady state equilibrium between variables. Cointegration has become an over-riding requirement for any economic model using non-stationary time series data. If the variables do not cointegrate, then we have the problems of spurious regression and the results therein become almost meaningless. On the other hand, if the variables do cointegrate then we have cointegration.

In applied econometrics, the Granger (1981) and, Engle and Granger (1987), Autoregressive Distributed Lag(ARDL) cointegration technique or bound test of cointegration(Pesaran and Shin 1999 and Pesaran et al. 2001) and, Johansen and Juselius(1990) cointegration techniques have become the solution to determining the long run relationship between series that are non-stationary, as well as reparameterizing them to the Error Correction Model (ECM). The reparameterized result gives the short-run dynamics and long run relationship of the underlying variables. However, given the versatility of cointegration technique in estimating relationship between non-stationary variables and reconciling the short run dynamics with long run equilibrium, most researchers still adopt the conventional way of estimation even when it is glaring to test for cointegration among the variables under consideration. That is most of the researchers are not conversant with the conditions that necessitate the application of cointegration test and the interpretation of the results therein, hence, presenting misleading inferences.
With this background, the objective of this paper is to examine the conditions that necessitate the application of the Autoregressive Distributed Lag (ARDL) cointegration or bound test of cointegration technique and its interpretation. Accordingly, this paper is divided into five sections. Section one, which is the introduction. Section two, examines the concept of stationarity, section three focuses on various unit roots tests, section four deals on ARDL cointegration approach, section five focuses on summary and conclusions.

2 Stationary and Non-Stationary Series Concept

A non-stationary time series is a stochastic process with unit roots or structural breaks. However, unit roots are major sources of non-stationarity. The presence of a unit root implies that a time series under consideration is non-stationary while the absence of it entails that a time series is stationary. This depicts that unit root is one of the sources of non-stationarity. A non-stationary stochastic process could be Trend Stationary (deterministic) Process (TSP) or Difference Stationary Process (DSP). A time series is said to be trend stationary process if the trend is completely predictable and not variable, whereas if it is not predictable, we call it difference or integrated stochastic trend or difference stationary process. In the case of deterministic trend, the divergence from the initial value (represents non-stationary mean) is purely random and they die out quickly. They do not contribute or affect the long run development of the time series. However, in the case of integrated stochastic trend, the random component (U_t) or divergence affects the long run development of the series. Utilizing time series with these features in any meaningful empirical analysis, the series must be purged of this trend. This is referred to as *detrending* of the series. This could be carried out in two ways, depending on whether the series is a difference stationary process or deterministic stationary process. If a series is DSP, it means it has a unit root; hence, the differencing of such series is stationary. Therefore, the solution to the non-stationary series is to difference the series. Also, if a series is TSP, it means it exhibits a deterministic trend, while a trend stationary variable with non-constant mean may be I(0) after removal of a deterministic trend. That is, regressing such series on time(t) and the residuals from this regression will be stationary($Y_t = \beta t + U_t$). Hence, cointegration cannot be seen as a means to an end but restricted. It should be made clear that if a time series is TSP, but treated as DSP, this is called *over-differencing*. On the other hand, if a time series is DSP, but treated as TSP; this is referred to as *under-differencing*. The implications of these types of specification error can be serious, depending on how the serial correlation properties of the resulting error terms are handled. However, it has been observed that most time series are DSP rather than TSP. Therefore, when such non-stationary time series (DSP) are used in estimation of an econometric model, the Ordinary Least Square (OLS) traditional diagnostic statistics for evaluation of the validity of the model estimates such as, coefficient of determination (R^2), Fisher's Ratio(F-

Statistic), Durbin-Watson(DW-Stat), t-statistic etc. become highly misleading and unreliable in terms of forecast and policy. In such series, the mean, variance, covariance and autocorrelation functions change overtime and affect the long run development of the series. The presence of unit root in these series leads to the violation of assumptions of constant means and variances of OLS. However, this review dwells on Difference Stationary Process rather than Trend Stationary Process since most time series are Difference Stationary Process.

As demonstrated above, many time series variables are stationary only after differencing. Hence, using differenced variables for regressions imply loss of relevant long run properties or information of the equilibrium relationship between the variables under consideration. This means that we have to devise a way of retaining the relevant long run information of the variables. Cointegration makes it possible to retrieve the relevant long run information of the relationship between the considered variables that had been lost on differencing. That is, it integrates short run dynamics with long run equilibrium. This is the basis for obtaining realistic estimates of a model, which is the driver of a meaningful forecast and policy implementation. Cointegration is a preferred step for modeling empirically meaningful relationships of DSP. Cointegration is concerned with the analysis of long run relations between integrated variables and reparameterizing the relationship between the considered variables into an Error Correction Model (ECM). Under the conventional Granger (1981) and, Engle and Granger (1987) cointegration analysis is not applicable in cases of variables that are integrated of different orders (i.e, series-A is I(1) and series-B is I(0)) while in Johansen and Juselius(1990), and ARDL cointegration procedure it is applicable. The ARDL cointegration technique is used in determining the long run relationship between series with different order of integration (Pesaran and Shin, 1999, and Pesaran et al. 2001). The reparameterized result gives the short-run dynamics and long run relationship of the considered variables.

Although ARDL cointegration technique does not require pre-testing for unit roots, to avoid ARDL model crash in the presence of integrated stochastic trend of I(2), we are of the view the unit root test should be carried out to know the number of unit roots in the series under consideration. This is presented in the next section.

3 Unit Root Stochastic Process

Given a Random Walk Model (RWM);

$$Y_t = \rho Y_{t-1} + U_t \tag{3.1}$$

$$-1 \leq \rho \leq 1$$

In the above RWM without drift, If $\rho = 1$, we are faced with unit root problem, that is, a situation of non-stationarity. In this case the variance of Y_t is not stationary.

However, If $/\rho/ \leq 1$, that is if the absolute value of ρ is less than one, then the series, Y_t is said to be stationary. Given this, U_t is said to be white noise and distributed normally with zero mean and unit variance. Hence, it follows that $E(Y_t) = 0$ and $Var(Y_t) = 1/(1-\rho^2)$.

A stochastic process Y_t is assumed to have a unit root problem if its first difference, $(Y_t\text{-}Y_{t-1})$ is stationary. In practice, the presence of unit root shows that the time series under consideration is non-stationary unless the reverse is the case. On the other hand, a series with unit root have no tendency to return to long-run deterministic path and the variance of the series is time dependent. A series with unit root suffers permanent effects from random shocks, thus, follow a random walk. That is, using (dependent and independent) time series that contain unit root in regression analysis, the classical results of the regression may be misleading. However, I(1) variables that exhibit a random walk without drift may have a mean that is constant over time, expected value of zero and, with trending variance; hence making the series with unit root to have the tendency to return to long-run path after removing deterministic trend. This reemphasized that; cointegration cannot be seen as a means to an end, but restricted. However, this paper focuses on series with unit root, I(1) (no constant mean and variance) that have no tendency of returning to the long-run path.

There are various methods of testing unit roots. They include; Durbin-Watson (DW) test, Dickey-Fuller test(1979)(DF), Augmented Dickey-Fuller(1981)(ADF) test, Philip-Perron(1988) (PP) test, among others. It is of the view that before pursuing formal tests to plot the time series under consideration, to determine the likely features of the series and; run the classical regression. If the series is trending upwards it shows that the mean of the series has been changing with time. In the case of the classical regression, if Durbin– Watson statistics is very low and a high R^2 (Granger–Newbold, 1974), this perhaps reveals that the series is not stationary. This is the initial step for a more formal test of stationarity. The most popular strategy for testing the stationarity property of a single time series involves using the Dickey Fuller or Augmented Dickey Fuller test respectively. The choice of the right tests depends on the set up of the problem which is of interest to the practitioner. It is difficult to follow the latest advances or to understand the problems between employing various tests. This should not be understood as a motive for not performing other types of unit root tests. Comparing different results from different test methods is a good way of testing the sensitivity of your conclusions. Once you understand how these tests work, and their limitations, you will understand when to use any test. The advantage is that it enables us to understand the meaning and purpose of any test. However, when a test result is inconclusive, the usual way is to continue the analysis with a warning note, or simply assume one of the alternatives. Thus, the unit roots test is basically required to ascertain the number of times a variable/series has to be differenced to achieve stationarity. From this comes the definition of integration: A variable Y, is said to be integrated of order d, I(d)] if it attained stationarity after differencing d times(Engle and Granger, 1987).

3.1 The Durbin-Watson Test

This test is a simple but unreliable test for unit root. To understand how this test works, recollect that the DW-value is calculated as $DW = 2(1-\hat{\rho})$(Harvey, 1981), where $\rho = \hat{\rho}$ is the estimated first order autocorrelation. Thus, if Y_t is a random walk, ρ will equal unity and the DW value is zero. Under the null that Y_t is a random walk, the DW statistic calculated from the first order autocorrelation of the series $Y_t = Y_{t-1} + V_t$, will approach one. The DW value approaches 0 under the null of a random walk. A DW value significantly different from zero rejects the hypothesis that Y_t is a random walk and I(1), in favor of the alternative that Y_t is not I(1), and perhaps I(0). The test is limited by the assumption that Yt is a random walk variable. This test is not good for integrated variables in general. The critical value at the 5% level for the maintained hypothesis of I(1) versus I(0) is 0.17. A higher value rejects I(1))(Bo Sjö, 2008).

3.2 Dickey-Fuller (DF) (1979) Test for Unit Roots

Assume that Y_t is random walk process, $Y_t = Y_{t-1} + \mu_t$, then the regression model becomes $Y_t = \rho Y_{t-1} + \mu_t$. Subtract Y_{t-1} from both sides of the equation,

$$Y_t - Y_{t-1} = \alpha_1 Y_{t-1} - Y_{t-1} + u_t \tag{3.2}$$

$$\Delta Y_t = (\alpha - 1) Y_{t-1} + u_t \tag{3.3}$$

$$\Delta Y_t = (\alpha - 1) Y_{t-1} + \alpha_2 T + u_t \tag{3.4}$$

Where $\alpha - 1 = p_1$, Δ is change in Y_t or first difference operator and t is the trend factor. u_t is a white nose residual.

$$\Delta Y_t = p_1 Y_{t-1} + u_t \tag{3.5}$$

With a drift we have;

$$\Delta Y_t = \alpha_0 + p_1 Y_{t-1} + u_t \tag{3.6}$$

In practice, we test the hypothesis that $p=0$. *If* $p=0$, *"α"* in equation 3.2 will be equal to 1, meaning that we have a unit root. Therefore, the series under consideration is non-stationary. In the case where $p \geq 0$, that is, the time series is stationary with zero mean and in the case of 3.4, the series, Y_t is stationary around a deterministic trend. If $p \geq 1$, it means that the underlying variable will be explosive.

However, conducting the DF test as in (3.3) or (3.4), it is assumed that U_t is uncorrelated. But in the case the error terms (U_t) are correlated, the Augmented Dickey-Fuller (ADF) is resorted to, since it adjusts the DF test to take care of possible autocorrelation in the error terms (U_t), by adding the lagged difference term of the dependent variable, ΔY_t.

3.3 The Augmented Dickey-Fuller (ADF) (1981) tests for Unit Root

$$\text{Restrictive ADF Model: } \Delta Y_t = p_1 Y_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta Y_{t-i} + u_t \tag{3.7}$$

$$\text{Restrictive ADF Model: } \Delta Y_t = p_1 Y_{t-1} + \alpha_2 T + \sum_{i=1}^{k} \alpha_i \Delta Y_{t-i} + u_t \tag{3.8}$$

$$\text{General ADF Model: } \Delta Y_t = \alpha_0 + p_1 Y_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta Y_{t-i} + u_t \tag{3.9}$$

$$\text{General ADF Model: } \Delta Y_t = \alpha_0 + p_1 Y_{t-1} + \alpha_2 T + \sum_{i=1}^{k} \alpha_i \Delta Y_{t-i} + u_t \tag{3.10}$$

u_t is a pure white noise error term and $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$, $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$, etc. The number of lagged difference terms to be included is often determined empirically, the reason being to include enough terms so that the error term in (3.5) and (3.6) are serially uncorrelated. k is the lagged values of ΔY, to control for higher-order correlation assuming that the series follow an AP(p). In ADF *$p=0$ is* *still tested and follow the same asymptotic distribution as DF statistic.* H_0*:* $p_1 = 0(p_1 \sim I(1))$*, against* $H_a : p_1 < 0(p_1 \sim I(0))$*.*

In practice, an DF or ADF value with less than its critical value shows that the underlying series is non-stationary. Contrarily, when an DF or ADF value that is greater than its critical value shows that the underlying series is stationary. However, the null hypothsis cannot be rejected about non-stationarity based on ADF test, since its power is not strong as such. This decision can be verified using other related tests, such as Kwiatkowski-Phillips-Schmidt-Shin (1992)(KPSS) or Philips-Perron (PP) test. PP test has the same null hypothesis as ADF, and its asymptotic distribution is the same as the ADF test statistic. But in the case of KPSS test, the null hypothesis is different; it assumes stationarity of the variable of interest. The results from ADF test differ from KPSS as KPSS does not provide a p-value, showing different critical values instead. In this case, the test statistic value is compared with the critical value on desired significance level. If the test statistic is higher than the critical value, we reject the null hypothesis and when test statistic is lower than the critical value, we cannot reject the null hypothesis. However, when there is a conflicting of the tests, it all depends on the researchers aim and objective. *In general, the null hypothesis for ADF reads that the series is non-stationary while KPSS reads that the series is stationary. For the treatment of serial correlation, PP reads that there is serial correlation (non-parametric) while ADF reads that there is serial correlation (parametric).*

The test can also be performed on variables in first differences as a test for I(2). Under the null, $\hat{p}_1$ will be negatively biased in a limited sample, thus, unless y_t is explosive. A significant positive value implies an explosive process, which can be a very difficult alternative hypothesis to handle. Conversely, When testing for I(2) or differencing twice, a trend term is not a possible alternative. The two interesting models here are the ones with and without a constant term. Furthermore, lag length in the augmentation can also be assumed to be shorter.

However, it is a good strategy to start with the model containing both a constant and a trend (3.10), because this model is the least restricted. If a unit root is rejected here, due to a significant p_1, there is no need to continue testing. If $p_1 = 0$ cannot be rejected, the improved efficiency in a model without a time trend might be better. There is also the lag length in the augmentation to consider (Bo Sjö, 2008).

A substantial weakness of the original Dickey-Fuller test(equation 3.3) as earlier stated is that it does not take account of possible autocorrelation in the error process U_t. If μ_t is autocorrelated (that is, it is not white noise) then the OLS estimates of the equations and, of its variants are inefficient. Therefore the simple solution is to apply ADF by using the difference lagged dependent variable as explanatory variables to take care of the autocorrelation.

The choice of the number of lags (p) to be included in the unit root test is based on the significant lag of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots of the correlogram and partial correlogram. The value of p is taken to be the number of lags at which the ACF cuts of or the number of lags of the PACF that are significantly difference from zero. By rule of thumb, we compute ACF up to one-third to one-quarter of the length of the time series. The ACF and PACF show different lags that are correlated and compared with the confidence bounds, mostly at 95 percent level. This will lead to AR process in cognizance of the properties of the residual(Uko and Nkoro, 2012). The characteristic of a time series has a far reaching implication for economic and policy formulation and implementation. When a series has a unit root *($p_1 = 0$),* any shock to the data series is long lasting. Hence, there will be a cumulative divergence from the mean/trend of the series. The instability exhibited by this series will tend to render any policy formulated and implemented on the basis of a model estimated using such data series inefficient. This is because what drives any policy formulation and implementation is the clear assumption of the stability of the series.

However, the Augmented Dickey-Fuller(ADF) test is considered superior because of its popularity and wide application. The ADF test adjusts the DF test to take care of possible autocorrelation in the error terms by adding the lagged difference term of the dependent variable. In the case of PP test it also take cares of the autocorrelation in the error term and, its asymptotic distribution is the same as the ADF test statistic. However, ADF is commonly used because of its easy applicability.

4 Cointegration Test

Modeling time series in order to keep their long-run information intact can be done through cointegration. Granger (1981) and, Engle and Granger(1987) were the first to formalize the idea of cointegration, providing tests and estimation procedure to evaluate the existence of long-run relationship between set of variables within a dynamic specification framework. Cointegration test examines how time series, which though may be individually non-stationary and drift extensively away from equilibrium can be paired such that the workings of equilibrium forces will ensure they do not drift too far apart. That is, cointegration involves a certain stationary linear combination of variables which are individually non-stationary but integrated to an order, I(d). Cointegration is an econometric concept that mimics the existence of a long-run equilibrium among underlying economic time series that converges over time. Thus, cointegration establishes a stronger statistical and economic basis for empirical error correction model, which brings together short and long-run information in modeling variables. Testing for cointegration is a necessary step to establish if a model empirically exhibits meaningful long run relationships. If it failed to establish the cointegration among underlying variables, it becomes imperative to continue to work with variables in differences instead. However, the long run information will be missing. There are several tests of cointegration, other than Engle and Granger(1987) procedure, among them is; Autoregressive Distributed Lag cointegration technique or bound cointegration testing technique. This becomes the focal point of this paper.

4.1 Autoregressive Distributed Lag Model (ARDL) Approach to Cointegration Testing or Bound Cointegration Testing Approach

When one cointegrating vector exists, Johansen and Juselius(1990) cointegration procedure cannot be applied. Hence, it become imperative to explore Pesaran and Shin (1995) and Pesaran et al (1996b) proposed Autoregressive Distributed Lag (ARDL) approach to cointegration or bound procedure for a long-run relationship, irrespective of whether the underlying variables are I(0), I(1) or a combination of both. In such situation, the application of ARDL approach to cointegration will give realistic and efficient estimates. Unlike the Johansen and Juselius(1990) cointegration procedure, Autoregressive Distributed Lag (ARDL) approach to cointegration helps in identifying the cointegrating vector(s). That is, each of the underlying variables stands as a single long run relationship equation. If one cointegrating vector (i.e the underlying equation) is identified, the ARDL model of the cointegrating vector is reparameterized into ECM. The reparameterized result gives short-run dynamics (i.e. traditional ARDL) and long run relationship of the variables of a single model. The re-parameterization is possible because the ARDL is a dynamic single model equation and of the same form with the ECM. Distributed lag Model simply means the inclusion of unrestricted lag of the regressors in a regression function.

This cointegration testing procedure specifically helps us to know whether the underlying variables in the model are cointegrated or not, given the endogenous variable. However, when there are multiple cointegrating vectors ARDL Approach to cointegration cannot be applied. Hence, Johansen and Juselius(1990) approach becomes the alternative. The next sections expose the requirement for using this approach and its application.

The ARDL$(p,q_1,q_2......q_k)$ model specification is given as follows;

$$\Phi(L,p)y_t = \sum_{i=1}^{k} \beta_i \left(L, q_i\right) x_{it} + \delta w_t + u_t \tag{4.1}$$

where

$$\Phi(L,p) = 1- \Phi_1 L - \Phi_2 L^2 -\ldots - \Phi_p L^p$$

$$\beta(L,q) = 1- \beta_1 L - \beta_2 L^2 -\ldots - \beta_q L^q, \qquad for\ i=1,2,3\ldots\ldots k,\ \boldsymbol{u}_t \sim iid(0;\delta^2).$$

L is a lag operator such that $L^0 y_t = X_t$, $L^1 y_t = y_{t-1}$, and w_t is a s x1 vector of deterministic variables such as the intercept term, time trends, seasonal dummies, or exogenous variables with the fixed lags. $P=0,1,2\ldots,m,\ q=0,1,2\ldots.,m,\ i=1,2\ldots.,k:$ namely a total of $(m+1)^{k+1}$ different ARDL models. The maximum lag order, m, is chosen by the user. Sample period, $t = m+1, m+2\ldots.,n.$

OR

The ADRL(p,q) model specification:

$$\boldsymbol{\Phi}(L)y_t = \boldsymbol{\varphi} + \boldsymbol{\theta}(L)x_t + u_t, \tag{4.2}$$

with

$$\boldsymbol{\Phi}(L) = 1 - \boldsymbol{\Phi}_1 L - \ldots - \boldsymbol{\Phi}_p L^p,$$

$$\boldsymbol{\theta}(L) = \boldsymbol{\beta}_0 - \boldsymbol{\beta}_1 L - \ldots - \boldsymbol{\beta}_q L^q.$$

Hence, the general ARDL$(p,q_1,q_2......q_k)$ model;

$$\Phi(L)y_t = \boldsymbol{\varphi} + \boldsymbol{\theta}_1(L)x_{1t} + \boldsymbol{\theta}_2(L)x_{2t} + \boldsymbol{\theta}_k(L)x_{kt} + \mu_t \tag{4.3}$$

Using the lag operator L applied to each component of a vector, $L^k y = y_{t-k}$, is convenient to define the lag polynomial $\Phi(L,p)$ and the vector polynomial $\beta(L,q)$. As long as it can be assumed that the error term u_t is a white noise process, or more generally, is stationary and independent of $x_t, x_{t-1}, \ldots$ *and* $y_t, y_{t-1}, \ldots$, the ARDL models can be estimated consistently by ordinary least squares.

4.2 Requirements for the Application of Autoregressive Distributed Lag Model (ARDL) Approach to Cointegration Testing

- Irrespective of whether the underlying variables are I(0) or I(1) or a combination of both, ARDL technique can be applied. This helps to avoid the pretesting problems associated with standard cointegration analysis which requires the classification of the variables into I(0) and I(1). This means that the bound cointegration testing procedure does not require the pre-testing of the variables included in the model for unit roots and is robust when there is a single long run relationship between the underlying variables,
- If the F-statistics (Wald test) establishes that there is a single long run relationship and the sample data size is small or finite, the ARDL error correction representation becomes relatively more efficient.
- If the F-statistics (Wald test) establishes that there are multiple long-run relations, ARDL approach cannot be applied. Hence, an alternative approach like Johansen and Juselius (1990) can be applied. That is, if the various single expression/equation of the underlying individual variable as dependent variable shows a feedback effect(multiple long run relationships) between the variables, then a multivariate procedure need to be employed.
- If the trace or Maximal eigenvalue or the F-statistics establishes that there is a single long-run relationship, ARDL approach can be applied rather than applying Johansen and Juselius approach.

To determine whether the above requirements are met or not see section 4.3.

4.3 Advantages of ARDL Approach

- Since each of the underlying variables stands as a single equation, endogeneity is less of a problem in the ARDL technique because it is free of residual correlation (i.e. all variables are assumed endogenous). Also, it enable us analyze the reference model.

- When there is a single long run relationship, the ARDL procedure can distinguish between dependent and explanatory variables. That is, the ARDL approach assumes that only a single reduced form equation relationship exists between the dependent variable and the exogenous variables (Pesaran, Smith, and Shin, 2001).
- The major advantage of this approach lies in its identification of the cointegrating vectors where there are multiple cointegrating vectors.
- The Error Correction Model (ECM) can be derived from ARDL model through a simple linear transformation, which integrates short run adjustments with long run equilibrium without losing long run information. The associated ECM model takes a sufficient number of lags to capture the data generating process in general to specific modeling frameworks.

4.4 The steps of the ARDL Cointegration Approach

This sub-section explores how one determines whether the above requirements are met.

Step 1: Determination of the Existence of the Long Run Relationship of the Variables

At the first stage the existence of the long-run relation between the variables under investigation is tested by computing the Bound F-statistic (bound test for cointegration) in order to establish a long run relationship among the variables. This bound F-statistic is carried out on each of the variables as they stand as endogenous variable while others are assumed as exogenous variables.

In practice, testing the relationship between the forcing variable(s) in the ARDL model leads to hypothesis testing of the long-run relationship among the underlying variables. In doing this, current values of the underlying variable(s) are excluded from ARDL model approach to Cointegration.

This approach is illustrated by using an ARDL (p,q) regression with an I(d) regressor,

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \theta_0 x_t + \theta_1 x_{t-1} \ldots + q_1 x_{t-p} + u_{1t} \tag{4.4}$$

or

$$x_t = \Phi_2 x_{t-1} + \ldots + \Phi_p x_{t-p} + \theta_0 y_t + \theta_1 y_{t-1} \ldots + q_1 y_{t-p} + u_{2t} \tag{4.5}$$

$$t = 1,2,\ldots T \qquad \mu_t \sim iid(0, \delta^2).$$

For convenience the deterministic regressors such as constant and linear time trend are not included. Where Φ, θ_0 and θ_1 are unknown parameters, and $x_{t(}$ or y_t) is an I(d) process generated by;

$$x_t = x_{t-1} + \mathcal{E}_t;$$

or

$$y_t = y_{t-1} + \mathcal{E}_t;$$

u_t and $\mathcal{E}_t$ are uncorrelated for all lags such that x_t (or y_t) is strictly exogenous with respect to $u_{t..}$ $\mathcal{E}_t$ is a general linear stationary process.

(Cointegration/stability Condition) /Φ/ <1, so that the model is dynamically stable. This assumption is similar to the stationarity condition for an AR(1) process and implies that there exists a stable long-run relationship between $y_t(x_t)$ and x_t (y_t). If Φ =1, then there would be no long-run relationship. *In practice, this can also be denoted as follows:*

The ARDL ($p,q_1,q_2......q_k$) model approach to Cointegration testing;

$$\Delta X_t = \delta_{0i} + \sum_{i=1}^{k} \alpha_i \Delta X_{t-1} + \sum_{i=1}^{k} \alpha_2 \Delta Y_{t-i} + \delta_1 X_{t-1} + \delta_2 Y_{t-1} + v_{1t} \tag{4.6}$$

$$\Delta Y_t = \delta_{0i} + \sum_{i=1}^{k} \alpha_i \Delta Y_{t-1} + \sum_{i=1}^{k} \alpha_2 \Delta X_{t-i} + \delta_1 Y_{t-1} + \delta_2 X_{t-1} + v_{1t} \tag{4.7}$$

k is the ARDL model maximum lag order and chosen by the user. The F-statistic is carried out on the joint null hypothesis that the coefficients of the lagged variables ($\delta_1 X_{t-1}$ $\delta_1 Y_{t-1}$ *or* $\delta_1 Y_{t-1}$ $\delta_1 X_{t-1}$) are zero. ($\delta_1 - \delta_2$) correspond to the long-run relationship, while ($\alpha_1 - \alpha_2$) represent the short-run dynamics of the model.

The hypothesis that the coefficients of the lag level variables are zero is to be tested.

The null of non-existence of the long-run relationship is defined by;

Ho: $\boldsymbol{\delta_1 = \delta_2 = 0}$ (null, i.e. the long run relationship does not exist)

H1: $\boldsymbol{\delta_1 \neq \delta_2 \neq 0}$ (Alternative, i.e. the long run relationship exists)

This is tested in each of the models as specified by the number of variables.

This can also be denoted as follows:

$$F_X(X_1 \mid Y_1, \ldots . Y_k) \tag{4.8}$$

$$F_y(Y_1 \mid X_1, \ldots . X_k) \tag{4.9}$$

The hypothesis is tested by means of the F- statistic (Wald test) in equation 4.8 and 4.9, respectively. The distribution of this F-statistics is non-standard, irrespective of whether the variables in the system are I(0) or I(1). The critical values of the F-statistics for different number of variables (K), and whether the ARDL model contains an intercept and/or trend are available in Pesaran and Pesaran (1996a), and Pesaran et al. (2001). They give two sets of critical values. One set assuming that all the variables are I(0)(i.e. lower critical bound which assumes all the variables are I(0), meaning that there is no cointegration among the underlying variables) and another assuming that all the variables in the ARDL model are I(1)(i.e. upper critical bound which assumes all the variables are I(1), meaning that there is cointegration among the underlying variables). For each application, there is a band covering all the possible classifications of the variables into I(0) and I(1). However, according to Narayan (2005), the existing critical values in Pesaran et al. (2001) cannot be applied for small sample sizes as they are based on large sample sizes. Hence, Narayan (2005) provides a set of critical values for small sample sizes, ranging from 30 to 80 observations. The critical values are *2.496 - 3.346, 2.962 – 3.910, and 4.068 – 5.250 at 90%, 95%, and 99%, respectively.*

If the relevant computed F-statistic for the joint significance of the level variables in each of the equations(4.6 and 4.9), δ_1, and δ_2 falls outside this band, a conclusive decision can be made, without the need to know whether the underlying variables are I(0) or I(1), or fractionally integrated. That is, when the computed F-statistic is greater than the upper bound critical value, then the H_0 is rejected (the variables are cointegrated). If the F-statistic is below the lower bound critical value, then the H_0 cannot be rejected (there is no cointegration among the variables). If long run (or multiple long-run relationships) relationships exist in both equations (4.8 and 4.9) the ARDL approach cannot be applied, hence, Johansen and Juselius (1990) approach becomes the alternative.

If the computed statistic falls within(between the lower and upper bound) the critical value band, the result of the inference is inconclusive and depends on whether the underlying variables are I(0) or I(1). It is at this stage in the analysis that the investigator may have to carry out unit root tests on the variables (Pesaran and Pesaran, 1996a). Also, if the variables are I(2), the computed F-statistics of the bounds test are rendered invalid because they are based on the assumption that the variables are 1(0) or 1(1) or mutually cointegrated (Chigusiwa et al., 2011). However, to forestall an effort in futility, it may be advisable to first perform unit roots, though not as a necessary condition in order to ensure that none of the variables is I(2) or beyond, before carrying out the bound F-test.

Step 2: Choosing the Appropriate Lag Length for the ARDL Model/ Estimation of the Long Run Estimates of the Selected ARDL Model

If a long run relationship exists between the underlying variables, while the hypothesis of no long run relations between the variables in the other equations cannot be rejected, then ARDL approach to cointegration can be applied. The issue of finding the appropriate lag length for each of the underlying variables in the ARDL model is very important because we want to have Gaussian error terms (i.e. standard normal error terms that do not suffer from non-normality, autocorrelation, heteroskedasticity etc.). In order to select the appropriate model of the long run underlying equation, it is necessary to determine the optimum lag length(k) by using proper model order selection criteria such as; the Akaike Information Criterion(AIC), Schwarz Bayesian Criterion (SBC) or Hannan-Quinn Criterion(HQC).

The values of AIC, SBC and LP for model 4.3 are given by;

$$AICp = -n/2(1+\log 2\pi)-n/2\log\delta^2-P$$

$$SBCp = \log(\delta^2) +(\log n/n)P$$

$$HQC =\log \delta +(2\log\log n/n)P$$

$$LRp,p = n(\log[\textstyle\sum p]-\log[\hat{\ }\textstyle\sum p])$$

Where δ^2 is Maximum Likelihood(ML) estimator of the variance of the regression disturbances, $\hat{\ }\sum p$ is the estimated sum of squared residuals, and n is the number of estimated parameters, $p=0,1,2\ldots\ldots P$, where P is the optimum order of the model selected.

The ARDL model should be estimated given the variables in their levels (non-differenced data) form. The lags of the variables should be alternated, model re-estimated and compared. Model selection criteria- The model with the smallest AIC, SBC estimates or small standard errors and high R^2 performs relatively better. The estimates from the best performed become the long run coefficients. This is appropriate to embark on if it is satisfied that there is long-run relationship between the underlying variables in order to avoid spurious regression.

The long-run coefficients for y_t(or x_t) to a unit change in x_t(or y_t) are estimated by;

$$\hat{\theta}_i = \frac{\hat{\beta}_i(1,\hat{q}_i)}{\hat{\phi}(1,\hat{p})} = \frac{\hat{\beta}_{i0}+\hat{\beta}_{i1}+\ldots+\hat{\beta}_{iq}}{1-\hat{\phi}_1-\hat{\phi}_2-\ldots\hat{\phi}_p} \qquad i = 1, 2\ldots$$

Where $\hat{p}$ and $\hat{q}_i$, i =1, 2,. . .k are the selected(estimated values of p and q, i =1, 2. . ,k

Similarly, the long-run coefficients associated with the deterministic/exogenous variables with fixed lags are estimated by;

$$\hat{\psi} = \hat{\delta}(\hat{p},\hat{q}_1,\hat{q}_2,\ldots\hat{q}_k)$$

$$1 - \hat{\phi}_1 - \hat{\phi}_2 - \ldots \hat{\phi}_p$$

Where $\hat{\delta}(\hat{p}, \hat{q}_1, \hat{q}_2, \ldots \hat{q}_k)$ denote the OLS estimate of δ in (equation 4.1) for the selected ARDL model.

In practice, this can also be denoted as follows:

The selected ARDL(k) model long run equation;

$$Y_t = \delta_0 + \sum_{i=1}^{k} \alpha_1 X_{1t} + \sum_{i=1}^{k} \alpha_2 X_{2t} + \sum_{i=1}^{k} \alpha_3 X_{3t} + \sum_{i=1}^{k} \alpha_n X_{nt} + v_{1t} \qquad (4.10)$$

Xs ($X_{1t}, X_{2t}, X_{3t}, \ldots\ldots\ldots\ldots X_{nt}$) are the explanatory or the long run forcing variables, k is the number of optimum lag order.

The best performed model provides the estimates of the associated Error Correction Model (ECM).

Step 3: Reparameterization of ARDL Model into Error Correction Model

As we said earlier, when non-stationary variables are regressed in a model we may get results that are spurious. One way of resolving this is to difference the data (since most data exhibit DSP) in order to achieve stationarity of the variables. In this case, the estimates of the parameters from the regression model may be correct and the spurious equation problem resolved. However, the regression equation only gives us the short-run relationship between the variables. It does not give any information about the long run behaviour of the parameters in the model. This constitutes a problem since researchers are mainly interested in long-run relationships between the variables under consideration, and in order to resolve this, the concept of cointegration and the ECM becomes imperative. With the specification of ECM, we now have both long-run and short-run information incorporated.

The unrestricted error correction model associated with the ARDL$(\hat{p}, \hat{q}_1, \hat{q}_2, \ldots \hat{q}_k)$ model can be obtain by rewriting equation 4.1 in terms of the lagged levels and the first differences of $y_t, x_{1t}, x_{2t}, \ldots x_{kt}$ and w_t. First note that;

$$y_t = \Delta y_t + y_{t-1}$$

$$y_{t-1} = y_t - \sum_{j=1}^{s-1} \Delta y_{i-j} \qquad s = 1,2, . . p$$

and similarly,

$$w_t = \Delta w_t + w_{t-1}$$

$$x_t = \Delta x_t + x_{t-1}$$

$$x_{1t-s} = yi_{t-1} - \sum_{j=1}^{s-1} \Delta x_{it-j} , \qquad s = 1,2, . . q_i$$

Substituting these relations into 4.1 we have;

$$\Delta y_t = - \phi(1,\hat{}p)EC_{t-I} + \sum_{i=1}^{k} \beta_{i0}\Delta x_{it} + \delta\Delta w_t - \sum_{j=1}^{p-1} \phi_j \Delta_{t-j} - \sum_{i=1}^{k} \sum_{j=1}^{\hat{}q-1} \beta_{ij}\Delta x_{i,1-j} + \mu_t \quad (4.11)$$

EC_t is the error correction term defined by;

$$EC_t = \varepsilon_t = y_t - \sum_{i=1}^{k} \hat{}\theta_i x_{it} - \psi' w_t$$

The term EC_t as the speed of adjustment parameter or feedback effect is derived as the error term from the cointegration models (4.6 and 4.7) whose coefficients are obtained by normalizing the equation on X_t (4.6) and Y_t (4.7) respectively. The EC_t shows how much of the disequilibrium is being corrected, that is, the extent to which any disequilibrium in the previous period is being adjusted in y_t. A positive coefficient indicates a divergence, while a negative coefficient indicates convergence. If the estimate of EC_t = 1, then 100% of the adjustment takes place within the period, or the adjustment is instantaneous and full, if the estimate of EC_t = 0.5, then 50% of the adjustment takes place each period/year. EC_t = 0, shows that there is no adjustment, and to claim that there is a long-run relationship does not make sense any more.

Recall that $\phi(1,\hat{}p) = 1 - \hat{}\phi_1 - \hat{}\phi_2 - . . . \hat{}\phi_p$ measures the quantitative importance of the error correction term. The remaining coefficients $\hat{}\phi_j$ and β_{ij}, relate to the short-run dynamics of the model's convergence to equilibrium. EC_t is the residuals that are obtained from the estimated cointegration model of equations 4.6 and 4.7.

The ARDL models and its associated ECM can be estimated by the OLS method.

5 Summary and Conclusion

Given the deficiencies associated with standard Johansen and Juselius(1990) cointegration procedure, it becomes imperative to explore Pesaran and Shin (1999) and Pesaran et al (1996b) proposed Autoregressive Distributed Lag (ARDL) approach to cointegration or bound procedure for a long-run relationship. Some of the deficiencies include: identifying the cointegrating vector(s) where there are multiple cointegrating relations; applicability when one cointegrating vector of different order exists. Based on this, this study reviewed Autoregressive Distributed Lag (ARDL) Approach to cointegration testing in terms of its application, estimation and interpretation. Given this, the following findings were made:

- ARDL cointegration technique is adopted irrespective of whether the underlying variables are I(0), I(1) or a combination of both, and cannot be applied when the underlying variables are integrated of order I(2). However, to avoid crashing of the ARDL technique and, effort in futility, it is advisable to tests for unit roots since variables that are integration of order I(2) leads to the crashing of the technique.
- If the trace or Maximal eigenvalue or the F-statistics establishes that there exists a single long-run relation among the variables (i.e underlying variables), ARDL approach can be applied rather than applying Johansen and Juselius approach. The ARDL technique provides a unified framework for testing and estimating of cointegration relations in the context of a single equation.

- If the F-statistics (Wald test) establishes that there is a single long run relationship and the sample data size is small (n≤ 30) or finite, the ARDL error correction representation becomes relatively more efficient.
- The ARDL model is reparameterized into ECM when there is one cointegrating vector among the underlying variables. The reparameterized result gives the short-run dynamics and long run relationship of the underlying variables.
- When there are multiple long-run relationships, ARDL approach cannot be applied. Hence, an alternative approach like Johansen and Juselius (1990) becomes more appropriate.

This review is an important starting point for future practitioners, as well as a more reliable research. ARDL cointegration technique is one of the greatest discoveries of the 20th century solution to the analysis of series with one cointegrating vector and, it does not require pretesting of unit root. Therefore, there is need to explore the necessary conditions that give rise to ARDL cointegration technique in order to avoid its wrongful application, estimation, and interpretation which may in turn lead to model misspecification and unrealistic estimates. However, this paper cannot claim to have treated the underlying issues in their greatest details, but have endeavoured to provide sufficient insight into the issues surrounding Autoregressive Distributed Lag (ARDL) cointegration technique to young practitioners to enable them apply the technique, estimate the problem therein, and interpret the result thereafter. Also, to enable them follow discussions of the issues in some more advanced texts.

References

[1] D. Asteriou, and S.G. Hall, *Applied Econometrics: A Modern Approach*, PALGRAVE MACMILLAN, New York, 2007.

[2] J.E. Davidson, H. David, F. Hendry, F. Srba, and S. Yeo, Econometric Modeling of the Aggregate Time Series Relationship Between Consumers' Expenditure and Income in the United Kingdom, *Economic Journal*, **88**, (1978), 661–692.

[3] M. Bahmani-Oskooee, and T.J.A. Brooks, New Criteria for Selecting the Optimum Lags in Johansen's Cointegration Technique, *Applied Economics*, **35**, (2003), 875-880.

[4] Bo Sjö, Testing for Unit Roots and Cointegration, *Memo*, (2008).

[5] G. Box and G. Jenkins, Time Series Analysis, Forecasting and Control. *San Francisco: Holden-Day*, (1970).

[6] P.T. Brandt, and J.T. Williams, *Multiple Time Series Models: Quantitative Applications in the Social Sciences*, Sage Publications Ltd, London, 2006.

[7] L. Chigusiwa, S. Bindu, V. Mudavanhu, L. Muchabaiwa and D. Muzambani, Export-Led Growth Hypothesis in Zimbabwe: Does Export Composition Matter? *International Journal of Economic Resources*, (**2**), (2011), 111-129.

[8] D. Dickey and W. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, **74**, (1979), 427-431.

[9] D. Dickey and W. Fuller, Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, **49**, (1981), 1057-1072.

[10] W. Enders, *Applied Econometric Time Series*, *2nd Ed*, John Wiley & Sons Inc, New York, 2004.

[11] R. Engle, and G. Granger, Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica*, **55**, (1987), 251-276.

[12] J. Geweke, R. Meese and W. Dent, Comparing Alternative Tests of Causality in Temporal System, *Journal of Econometrics*, **77**, (1983), 161-194.

[13] C.W.J.Granger, Investigating Causal Relations by Econometric Models and Cross Spectral Methods, Econometrica, **37**, (1969), 428-438.

[14] C.W.J. Granger, Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics,* **28**, (1981), 121-130.

[15] C.W.J. Granger, Cointegrated Variables and Error-Correcting Models, *UCSD Discussion, Paper 83-13*, (1983).

[16] C.W.J. Granger, Some Recent Developments in a Concept of Causality, *Journal of Econometrics*, **39**, (1988), 199-211.

[17] C.W.J. Granger, and J. Lin, Causality in the Long-run. *Econometric Theory,* **11**, (1995), 530-536.

[18] C.W.J. Granger, and P. Newbold, Spurious Regressions in Econometrics, *Journal of Econometrics,* **26**, (1974), 1045-1066.

[19] D. Hendry and G. Mizon, Serial Correlation as a Convenient Simplification not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England. *Economic Journal,* **88**, (1978), 349-363.

[20] D. Hendry and J.F. Richard, The Econometric Analysis of Economic Time Series, *International Statistical Review,* **51**, (1983), 111-163.

[21] M. Iyoha and O.T Ekanem, *Introduction to Econometrics*. Mareh Publishers, Benin City, Nigeria, 2002.

[22] S. Johansen and K. Juselius, Hypothesis Testing for Cointegration Vectors with an Application to the Demand for Money in Denmark and Finland. *Working Paper No. 88-05, University of Copenhagen,* (1988).

[23] S. Johansen and K. Juselius, Maximum Likelihood Estimation and Inference on Cointegration-With Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics,* **52**(2), (1990), 169-210.

[24] D. Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root, *Journal of Econometrics*, (1992), *15978*.

[25] G.S. Maddalas, *Introduction to Econometrics*, *2nd Ed*, Englewood Cliffs, Prentice Hall, 1992.

[26] P.K Narayan, The Saving and Investment Nexus for China: Evidence from Cointegration Tests, *Applied Economics*, **37**, (2005), 1979–1990.

[27] C.R. Nelson and C.I. Plosser, Trend and Random Walks in Macroeconomics Time Series. *Journal of Monetary Economics*, **10**(2), (1982), 139-162.

[28] M.H. Pesaran and B. Pesaran, *Microfit 4.0*, Cambridge London, Windows Version, CAMFIT DATA LIMITED, 1996.

[29] M.H. Pesaran, R.J. Smith, and Y. Shin, Testing for the Existence of a long run Relationship, *DAE Working paper No.9622*, Department of Applied Economics, University of Cambridge, (1996b).

[30] M.H. Pesaran, R.J. Smith and Y. Shin, Bounds Testing Approaches to the Analysis of Level Relationships, *Journal of Applied Econometrics*, **16**, (2001), 289-326.

[31] M.H. Pesaran and Y. Shin, An Autoregressive Distributed Lag Modeling Approach to Cointegration Analysis, In: Strom, S., Holly, A., Diamond, P. (Eds.), *Centennial Volume of Rangar Frisch*, Cambridge University Press, Cambridge, (1999).

[32] P. Phillips and P. Perron, Testing for a Unit Root in Time Series Regression. *Bimetrika,* **75**, (1988), 335-346.

[33] J.D. Sargan, Wages and Prices in the United Kingdom: A Study in Econometric Methodology. *In Econometric Analysis for National Economic Planning* (R.E. Hart, G. Mills, J.K. Whittaker, eds.), Butterworths, London, (1964), 25–54.

[34] H.R. Seddighi, K.A. Lawler and A.V. Katos, *Econometrics: A Practical Approach,* Routledge, London EC4P4EE, 2000.

[35] C. Sims, Money, Income and Causality, *American Economic Review*, **62**, (1972), 540-552.

[36] A.K. Uko and E. Nkoro, Inflation Forecast with ARIMA, Vector Autoregressive and Error Correction Models in Nigeria, *EJEFAS*, Issue 50, July, (2012).