Chapter - III
CHAPTER III

METHODOLOGY OF THE STUDY

The methodology of study consists of Research Design, Source of Data, Period of Study and Method of Data Analyses.

Research Design

The study has given a brief summary of trade in Goods and trade in services of BRICS countries. The relationship among trade in goods and trade in services are analyzed by applying the econometric techniques. Hence the present study is descriptive and as well as exploratory.

Sources of Data

Present study is based on secondary data. The data on trade in goods and trade in services for BRICS countries were collected from various sources, United Nation Conferences on Trade and Development (UNCTAD), World Bank, and various issues of RBI Bulletin.

Period of Study

Data on trade in goods and trade in services were collected from 24 years (from 1990 to 2013) and the structure data is Time Series Data.

Method of Data Analysis

The objective of the study were analysed by application of tolls such as Percentages, Analysis of Variance, Simple Linear Regression Model, Exponential Growth model, Johansen Cointegration, Chow and F tests.

Analysis of Variance (ANOVA)

To examine the difference between Trade in goods and Trade in services of BRICS countries, ANOVA was used. Professor R.A. Fisher was the first man to use the term ‘Variance’ and in fact, it was he who developed a very elaborate theory concerning ANOVA, explaining its usefulness in practical field. Later on Professor Snedecor and many others contributed to the development of this technique. ANOVA is essentially a
procedure for testing the difference among different groups of data for homogeneity. “The essence of ANOVA is that the total amount of variation in a set of data is broken down into two types, that amount which can be attributed to chance and that amount which can be attributed to specified causes.” There may be variation between samples and also within sample items. ANOVA consists in splitting the variance for analytical purposes. Hence, it is a method of analysing the variance to which a response is subject into its various components corresponding to various sources of variation.

Thus, through ANOVA technique one can, in general, investigate any number of factors which are hypothesized or said to influence the dependent variable. One may as well investigate the differences amongst various categories within each of these factors which may have a large number of possible values. If take only one factor and investigate the differences amongst its various categories having numerous possible values, are said to use one-way ANOVA and in case investigate two factors at the same time, then we use two-way ANOVA. In a two or more way the interaction between two independent variables affecting a dependent variable can as well be studied for better decisions.

The Basic Principle of ANOVA

The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples. In terms of variation within the given population, it is assumed that the values of \((X_{ij})\) differ from the mean of this population only because of random effects that is., there are influences on \((X_{ij})\) which are unexplainable, whereas in examining differences between populations assume that the difference between the mean of the \(j\)th population and the grand mean is attributable what is called a ‘specific factor’ or what is technically described as treatment effect. Thus while using ANOVA, assume that each of the samples is drawn from a normal population and that each of these populations has the same variance. Assume that all factors other than the one or more being tested are effectively controlled. This in other words means that we assume the absence of many factors that might affect our conclusions concerning the factor(s) to be studied.
In short have to make two estimates of population variance viz., one based on between samples variance and the other based on within samples variance. Then the said two estimates of population variance are compared with $F$-test, where in work out.

\[
F = \frac{\text{Estimate of population variance based on between samples variance}}{\text{Estimate of population variance based on within samples variance}}
\]

This value of $F$ is to be compared to the $F$-limit for given degrees of freedom. If the $F$ value work out is equal or exceeds the $F$-limit value may say that there are significant differences between the sample means (Kothari 2009)\(^1\).

**Simple Linear Regression Model**

To analyse impact of BRICS trade in goods and Trade in services on world trade in goods and services Simple Linear regression model was used. Regression is the determination of a statistical relationship between two or more variables. In simple regression, have only two variables, one variable (defined as independent) is the cause of the behaviour of another one (defined as dependent variable). Regression can only interpret what exists physically i.e., there must be a physical way in which independent variable $X$ can affect dependent variable $Y$. In other words, simple linear regression fits a straight line through the set of n points in such a way that makes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible. The slope of the fitted line is equal to the correlation between $y$ and $x$ corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass ($x, y$) of the data points.

\[
Y = \beta_0 + \beta_1x + U_t
\]

$Y =$ Trade in Goods  
$\beta_0 =$ Constant  
$\beta_1x =$ Trade in services  
$U_t =$ Error Term

---

Exponential Growth Model (Semi Log Model)

To analyses the growth pattern of exports and imports of trade in goods and trade in services of BRICS countries, exponential growth model was used. A quantity is growing exponentially if it increases by the same percent in each unit of time. This is called exponential growth.

Economists, businesspeople, and governments are often interested in finding out the rate of growth of certain economic variable, such as population, GNP, money supply, employment, productivity, and trade deficit. Here we want to find out the growth pattern of exports and imports trade in goods and services for BRICS countries. Let \( Y_t \) denote real growth pattern on services at time \( t \) and \( Y_0 \) the initial value of the growth pattern on service. The CAGRs were estimated by fitting an exponential then compound growth rate formula is

\[
Y_t = Y_0 (1+r)^t \quad \text{.................. (1)}
\]

Where \( r \) is the compound (i.e., over time) rate of growth of \( Y \). Taking the natural logarithm of equation (1) can write

\[
\ln Y_t = \beta_1 + \beta_2 t + u_t \quad \text{...............(2)}
\]

This model is like any other linear regression model in that the parameters \( \beta_1 \) and \( \beta_2 \) are linear. The only difference is that the regressand is logarithm of \( Y \) and the regressor is time, which will take values of 1, 2, 3, etc. This model called semi log model because only one variable (in this case regressand) appears in logarithmic form.

In this model the slope coefficient measures the constant proportional or relative change in \( y \) for a given absolute change in the value of the regressor. That is,

\[
\beta_2 = \text{relative change in regressand is divided by absolute change in regressor. If we multiply the relative change in } y \text{ in by } 100, \text{ then give the percentage change or growth rate, in } y \text{ for an absolute change in } x, \text{ the regressor. Compound Growth rate can be obtained by taking antilog of estimated regression coefficient, subtracting 1 from it.}
\]

For expressing the compound growth rate in percentage term is \([ (b_2 - 1) \times 100 \].

94
Estimation of Long run Relationship

To estimate long run relationship among trade in goods and trade in services of BRICS Countries study was used cointegration test developed by Johansen (1988), Johansen and Juselius (1990). For examining causality, the study was used Granger Causality Test. That has been employed in order examine Trade in Goods and Trade in services of BRICS Countries. Cointegration and Granger Causality Tests require a certain stochastic structure of the time series and for this stationarity test was performed to determine the order of integration for each time series by using Augmented Dickey Fuller Test (ADF) (1979) and Phillip Perron Test (PP) (1988).

Stationarity, Cointegration and Related Tests

Stationarity Tests

A brief outline why testing the time series data is important before applying various econometric test is outlined this section. Data properties are one of the major concerns in time series econometrics as most macroeconomic variables, in particular in level, contain non stationary process (Nelson and Plosser, 1998)\(^2\). The two central properties of many economic time series are non stationarity and time – volatility (Wei, 2006)\(^3\). Applying OLS regression on non stationary variables will end up with a spurious regression that is, the regression estimation will yield high R\(^2\), statistically significant coefficients, and low Durbin – Watson’d’ statistics (Gujarati, 1995)\(^4\). The non – random behaviour of the time serious data could undermine the usefulness of the standard econometrics methods, if it is applied directly without considering time series properties of the data (Russel, 1993)\(^5\). There are several versions of testing stationary or the presence of unit roots.

Stationary process is one whose statistical properties do not change over time. A stationary time series is one whose properties do not depend on the time at which the


series is observed. So time series with trends, or with seasonality, are not stationary. The trend and seasonality will affect the value of the time series at different times. On the other hand, a white noise series is stationary; it does not matter when you observe it. It should look much the same at any period of time. Non-stationary is a property common to many applied time series. This means that a variable has no clear tendency to return to a constant value or linear trend. It is generated by a non stationary process and follows stochastic trends. One major objectives of empirical research in economics is to test hypotheses and estimate relationships derived from economic theory, among other such aggregated variables. It can originate from various sources but the most important one is the unit root (Pfaff, 2006)\textsuperscript{6}. As to Harris (1996)\textsuperscript{7} a data series is said to be stationary if its error term has zero mean, constant variance, and the covariance between any two time periods depends only on the distance or lag between the two periods and not on the actual time at which it is computed.

If time series are non stationary it is not possible to use the classical statistical methods used in building and testing large simultaneous equation models, such as OLS to estimate their long run linear relationship because it would lead to spurious or nonsensical regression. The reason for the spurious or false results is that when applying the OLS model it is assumed or based on the assumption that the time series under investigation is assumed to be stationary.

Spurious regression is a regression shows significant results due to the presence of unit root in both variables, in other words it is situation in which there appears to be a statistically significant relationship between variables but the variables are unrelated. It may result when non stationary time series is regressed against one or more non stationary time series. Previously the difficulty of non stationary was not understood well. However, this is no longer the case because the technique of cointegration has been introduced according to which models containing non stationary stochastic variables can be constructed in such a way that the results are both statistically and economically


meaningful (Gujarati, 1995). The most successful way of curbing the problem of spurious regression is by checking for cointegration of the variables used in time series modeling. Hence, prior to the estimation of the long run models the time series properties of the variables concerned should be distinguished between stationary and non-stationary variables.

In this study applied two main methods: Augmented Dickey Fuller (ADF) and Phillip Perron (PP) tests respectively. This is due to the fact that the data generating process is not an AR (1) process. Therefore, the ADF is correctly specified in the higher order case (Engle and Granger, 1987). The ADF procedure attempts to retain the validity of the tests based on white noise errors in the regression model by ensuring that the errors are indeed white noise. On the other hand, Phillip Perron (PP) procedures correct the problem of serial correlation through a parametric correction to the standard statistic.

**Augmented Dickey Fuller (ADF) Test of Stationarity**

Dickey Fuller test assumed that the error term $U_t$ was uncorrelated. But in case the $U_t$ are correlated. Dickey and Fuller have developed a test known as Augmented Dickey–Fuller test (ADF) (Damodar Gujarati). The test use null hypothesis is that there exist unit root, meaning that existing series is non-stationary and alternative hypothesis is that there exist no unit root test, meaning that existing series is stationary.

The formal ADF test proceeds by estimating following specification.

$$\Delta y_t = \alpha_1 + \alpha_2 T + P y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-1+i} + U_t, \quad t = 1, \ldots, n$$

Where $Y_t$ is the variable under examination, $a_1$ is the intercept and $T$ is the time trend, $e_t$ is the error term assumed to be white noise. The parameter of interest here is $p$. The null hypothesis is $H_0: p = 0$ or the variable constitutes unit root. The critical value of this hypothesis test follows the ADF critical value (Mackinnon critical values) as the standard $t$ statistic is not applicable. The decision of accepting or rejecting the null hypothesis carried out by comparing the $t$ statistic of the estimated parameter with the critical value. The null hypothesis of non-stationary (unit root) is rejected if the $t$ statistic

---

8  Ibid. 4.
is larger (in absolute) than the critical value and then can be concluded that the series is stationary. In this case the time series is said to be integrated of order zero I (0).

Actually there are several objections against the use of the sequential testing procedure just described. First, a general size problem can result from sequential testing by the potential accumulation of wrong test decision for example, when the F-tests rejects although the nulls are true. Second, the use of standard normal distribution is misplaced as soon as the deterministic of the data generation process do not match the one of the test equation. To rely on the result of the F - test involves a risk: If the true data generating process does not contain a deterministic trend, its presence in the test equation introduces a downside bias in the estimate of 4B thereby making a rejection of the null of non stationarity more probable. In this case it is not clear whether the rejection of the null in the F-test really stems from the presence of a deterministic trend or whether the process is in fact stationary. Clearly, the F-test cannot solve the DF test’s problem of low power against a trend stationary alternative. Third, if the true value of 6B is close to but smaller than one, convergence to the asymptotic test distribution is slow. Thus, in small samples the DF distribution may be a better approximation (Banerjee and Newman 1993)\textsuperscript{10}. Therefore, sequential testing is not appropriate to solve the huge uncertainty surrounding the true nature of the data generating process in economic time series.

Philip Perron (PP) Test

The PP test differs from the ADF test in that it does not assume white noise residuals, but corrects the problem of serial correlation in the residuals. The test uses a non-parametric method to account for serial correction in the residuals. It does not augment the Dickey-Fuller test equation when accounting for serial correction but it instead adjusts the test (t) statistic to account for serial correction. The modified statistic of the PP test however follows the same distribution as the ADF statistic. As with the ADF test, in the PP test, the null hypothesis of a unit root (non stationary), i.e. p=0 is tested against the alternative hypothesis that no unit root is present/ stationarity p < 0. The calculated t-statistic of the estimated p is compared with its respective critical value.

If the absolute value of the calculated t statistics greater than the critical value, then the null hypothesis of a unit root (non-stationarity) is rejected and we concluded that the series is stationary. In this case the level of time series said to be integrated of order zero I

**Cointegration Test**

While the preceding test examines the presence of the unit root each individual variable, cointegration test will check the stationary of linear combination of those variables as a group. The concept of cointegration is particularly important in VAR analysis, since it is closely related to the existence and relevance of long run equilibrium relationship among non stationary variables being studied. The idea of long run equilibrium implies that two or more variables may wander away from each other in the short run but move together in the long run (Enders, 1995)\(^{11}\).

Cointegration means linear combination of two or more variables could be stationary (long run stable relationship) in spite of the existence of non-stationary in their individual time series. Two time series \(Y_t\) and \(X_t\) said to be cointegrated of order \((d, b)\), where \(d \geq b \leq 0\), if both time series are integrated of order \(d\) and there exists a linear combination of these two time series, say \(a_1Y_t + a_2X_t\), which is integrated of order \((d - b)\). The vector of the coefficient, which constitutes the linear combination the two series, it is called co integrating vector. In a bivariate (Two variable case) cointegration test, the null hypothesis that there is no cointegrating vector against the alternative hypothesis that there is one cointegrating vector. Those variables are integrated same order have been considered for Johansen Cointegration Test.

Furthermore, cointegration techniques are also powerful when modeling the behavior of data series that stochastic trends. This is because the existence of a cointegrating relationship between two or more non stationary variables counters the ordinary least squares (OLS) problem of spurious regression and regression estimates have been shown to be super consistent in this context. Engel and Granger (1987)\(^{12}\) demonstrated that if two variables are shown to be cointegrated, then the possibility of no causation in the Granger sense (not structural sense) is ruled out; and causality must exist at least in one direction , unidirectional or bi-directionally.

---


\(^{12}\) Ibid. 9.
The use of cointegration technique will allow capturing the equilibrium relationship between non stationary series within stationary model. It permitted the combination of long run and short run information in the same model and overcame the problem of losing information which could have occurred when attempting to address non stationary series through differencing (Adam, 1998). Cointegration technique made it possible to capture the information of non stationary series without sacrificing the statistical validity of the estimated equation. Two main tests for cointegration, namely Johansen cointegration test and Granger two step methods are usually used. The Engle Granger (1987) approach is inappropriate in the case of model with more than two variables. Johansen (1988) developed a maximum likelihood estimation procedure that allows one to test for the number of cointegrating relations.

Cointegration ranks are contingent upon the presence or absence of deterministic components in the dynamic model. The next question is to investigate whether all the variables in the model should enter into a long run equilibrium relationship. This can be done by testing linear restrictions on the long run coefficient after they have been normalized. The hypothesis of long run exclusion of each variable is tested using a likelihood ratio test. The Johansen tests are called the maximum eigenvalue test and the trace test. The Johansen tests are likelihood-ratio tests. There are two tests:

1. Maximum Eigenvalue Test
2. Trace Test.

For both test statistics, the initial Johansen test is a test of the null hypothesis of no cointegration against the alternative of cointegration. The tests differ in terms of the alternative hypothesis. If the test statistics exceeds the 95 per cent critical value then those coefficients are significant implying that the concerned variables should be present in the long run equilibrium relationship. The number of cointegrating relationships found will result in

---

14 Ibid.9
corresponding number of residual series and hence error correction terms (ECTs) to be used in the subsequent Vector Error Correction Model (Uddin and Habib 2009)\textsuperscript{16}.

**Granger Causality Test**

The causal relationship between trade in goods and services of BRICS to be analysed by employing the Granger Causality test. Regression analysis deals with the dependence of one variable on other variables, it does not necessarily imply causation. In other words, the existence of a relationship between variables does not prove causality or the direction of influence. But in regression involving time series data, the situation somewhat different because time does not run backward. That is, if event ‘A’ happens before event ‘B’ then it is possible that ‘A’ is causing ‘B’. However, it is not possible that ‘B’ is causing ‘A’. in other words, event in the past can cause events to happen today. Future events cannot. This is roughly the idea behind the so called Granger Causality Test (Gujarati)\textsuperscript{17}. Granger test is used to assess whether another variable lag makes or does not make a net significant incremental contribution to the movement of the dependent variable, once the own correlation of the dependent variable is taken in to account.

Testing for causality in the contest of stable VAR models involves whether some parameters of the model are jointly zero. The older approaches utilized techniques such as the Standard Granger (1969) or Sims (1972) methods which uses the non stationary data. It is shown, however that the use of non stationary data in causality tests can yield spurious causality results. Hence most causality studies since the late 1980s employ unitroot tests to examine the stationarity properties of variables perform Cointegration analysis mostly following the Johansen Procedure and formulate Vector Error Correction model inorder to capture both long run and short run sources of causality between the variables. An alternative method is that which presupposes the use of unit root and Cointegration tests but applying the standard Granger test rather than a VEC Model, in cases when no cointegration is found; Yoo (2006) has applied this method among others.


\textsuperscript{17} Ibid 4.
As unit root and cointegration tests are known to have low power and size properties in small samples (Cheung and Lai, 1993)\(^{18}\) there has been an increasing drift towards the use of methods which do not require that the variables be pre-tested for stationarity and cointegration. Such methods are the autoregressive distributed lag (ARDL) approach due to Pesaran and shin (1999) and the Dolado-Lutkepohl and Toda–Yamamoto methods, which involve a modified Wald test in an augmented vector autoregressive model (Dolado and Lutkepohl 1996)\(^{19}\). Hypothesis tests can be carried out with these methods irrespective of whether the variables involved are stationary or not and regardless of the existence of a co-integrating relationship among them.

In principle, the selection of causality test method should not affect the results as long as the time series properties of variables are accounted for appropriately. However, it has been found that asymptotically equivalent methods do not necessarily demonstrate similar properties in small samples of between 25 and 35 observations, which is the actual sample size in most causality studies. In this respect, (Pesaran and Shin 1999)\(^{20}\) have shown that the ARDL model is more rigorous in small samples than cointegration methods and Zapata and Rambaldi(1997) have demonstrated that the modified Wald tests employed by the Toda- Yamamoto and Dolado- Lutkepohl methods have lower power the Johansen based VEC approach in multivariate modes with simple size of 50 or less..

Given two variables \(X_t\) and \(Y_t\), \(X_t\) is said to granger-cause \(Y_t\), if lagged values of \(X_t\) help in the prediction of \(Y_t\) or if the coefficients on the lagged values of \(Y_t\), are statistically significant in the equation of \(Y_t\).

The test equation would thus take the form;

\[
y_t = \alpha + \sum_{j=0}^{k} \beta_j x_{t-j} + \sum_{i=0}^{k} y_{t-i} x_{t-i} + U_t
\]

The null hypothesis is that \(\beta_j = 0\) and it rejection that \(x_t\) can be said to Granger cause \(y_t\).

---


The Granger method (Granger 1988) seeks to determine how much of variable ‘Y’ can be explained by past values of ‘Y’ and whether adding lagged values of another variable ‘X’ can improve the explanation. Granger Causality test considers the amount of the current variable (e.g. Trade in goods) explained by past values of its self and then finds out whether adding lagged values of other variable (e.g. Trade in services) could improve the explanation. Because in most cases the direction of causality is not known, various tests have been suggested to identify the direction. The most popular well known test is “Granger Causality Test” proposed Granger (1969, 1988). Only cointegrated variables are allowed for running Granger Causality test. The null hypothesis is that there is no Granger Causality (Gujarati 1995). This test based on the concept of the Vector Auto regressive Model (VAR).

Before testing for causality, the presence of unit root in the time series is tested by using the Augmented Dicky Fuller Test (ADF) and Phillips Perron - Test. If time series found stationary the Granger Causality Test can be applied to test Causality between variables. Here regression equations to test the 2 null hypotheses are shown below:

1. Trade in goods does not Granger- Cause Trade in service

   The regression equation is,

   \[ TGS_t = c + TSS_{t-1} + Ut \]

2. Trade in service does not Granger- Cause Trade in goods

   The regression equation is,

   \[ TSS_t = c + TGS_{t-1} + Ut \]

**Vector Error Correction Model (VECM)**

As explained by Engle et al.(1987) an error correction model is a dynamical system with the characteristics that the deviation of the current state from its long run relationship will fed in to its short run dynamics. An error correction model (ECM) that corrects the error in another model. Error Correction Models (ECMs) are a category of multiple time series models that directly estimate the speed at which a dependent variable Y returns equilibrium after a change in an independent variable X. ECMs are a
theoretically driven approach useful for estimating both short term and long term effects of one time series on another. ECMs are useful models when dealing with cointegrated data, but can also be used with stationary data.

If counter action has been detected between series it can be concluded that there exist a long term equilibrium relationship between them and that is also sufficient condition to run VECM in order to evaluate the short run properties of the cointegrated series. In case of no cointegration VECM is no longer required and it is possible to directly proceed to Granger Causality tests to establish causal link between variables.

**Chow Break Point and F Test**

To find out structural break of trade in goods and trade in services of BRICS Countries after the formation of group the study was used Chow Break Point and F Tests. Identifying structural instability in models is of major concern to econometric practitioners. Chow Test examines whether parameters of one group of the data are equal to those of other groups. Most of the work has concentrated on detecting the presence of structural break(s) and estimating the location of the break(s).

There are two well-known problems with structural break estimation. The first one is the difficulty of differentiating data that is subject to a structural break (before and after which data shows stationary and trend stationary patterns) from data having a unit root. The second one is that although break locations in data can be estimated consistently, there is no efficiency condition for the limiting distribution of the estimates. Although consistency is a sufficient condition for the purpose of many empirical studies, efficiency could still be of interest if the aim is to obtain the smallest confidence intervals around the break dates. The stated reason behind these difficulties of estimating structural breaks is that the problem is nonstandard; a break date only appears under the alternative hypothesis, not under the null of no break. Perron (2005)\(^2\) empirical study makes a comprehensive review of both problems. It is very technical and seemingly there is a lack of resources summarizing the relevant literatures. To overcome this, Perron proposed allowing for a known or exogenous structural break in the Augmented Dickey-Fuller (ADF) test.

---

Literature Review of Structural Break

Sudden change can occur in time series data or cross sectional data, when there is a sudden change in the relationship being examined. A data can be found to be non-stationary if it has a unit root, or if it includes a structural break, before and after which data shows different patterns. As it is intricate play between unit roots and structural breaks (Perron 1989)\textsuperscript{22}. Most tests that attempt to distinguish between a unit root and a (trend) stationary process will favor the unit root model when the true process is subject to structural changes, but is otherwise (trend) stationary within regimes specified by the break dates. Most tests trying to assess whether a structural change is present will reject the null hypothesis of no structural change when the process has a unit root component but also constant model parameters. Accordingly, there is much literature on testing for a unit root under structural break(s). These tests also give break dates as a by-product, but they are not as efficient as the break estimators. The early influential empirical study of Perron (1989) tests null hypothesis of unit root under the assumption of known (exogenous, pre-tested) break date in both null and alternative hypotheses. Later Christiano (1992)\textsuperscript{23} criticizes Perron’s known date assumption as data mining. He argues that the data based procedures are typically used to determine the most likely location of the break that is. By pre-test examination of the data and this approach invalidates the distribution theory underlying conventional testing.

Zivot and Andrews (1992)\textsuperscript{24} and Perron (1997)\textsuperscript{25} proposed determining the break point endogenously from the data. However, these endogenous tests were criticized for their treatment of breaks under the null hypothesis. They do not allow for break(s) under the null hypothesis of unit root and derive their critical values accordingly. So they exclude the possibility that there may be a unit root process with a break. One way of


overcoming this problem would be taking log difference of the data, which made the series stationary and look for a break in the growth rate of the series. It would be wise to avoid data conversions that smooth the data; especially when the data is not long enough or includes outliers. It is because under these conditions break estimation tends to catch any kind of one time deviation in the data rather than finding a change in trend or in mean. In this case, these tests declare data as stationary with breaks. So it seems literature on this subject arrives at the approach of Lee and Strazicich (2003)\textsuperscript{26} employing minimum Lag range Multiplier (LM) tests.

One test allows for two-breaks in time series data, and the other allows one. While testing for a unit root, they both estimate break date(s) endogenously from the data and also allow break(s) both under the null and alternative hypotheses. By simulation exercises they show that their test outperforms existing ones. Besides, recently Glynn et al. (2007)\textsuperscript{27} analysed existing tests and mentioned the superiority of Lee and Strazicich’s test. Conversely, they also point out that instead of univariate models, common feature analysis of unit root with breaks has more potential, while indicating the development in this area is very limited. Hendry and Massmann (2007)\textsuperscript{28} applied to test unit roots under structural breaks, or directly to test for structural breaks, which rests on the principle that there is an appropriate combination of variables, having a break in common, that does not display the breaks any longer. But this very reason also prevents co-feature analysis from always being applicable. Applying them requires using more than one series, which are suspected to have common breaks. In order to deal with breaks in the growth rate of export, import and GDP, it requires using different type regressions and cannot be tested at a time with co-breaking analysis. Alternatively, each variable may have been tested for a break independently from the other.


Structural break tests can be divided into three categories. The Chow test is used within the first category. It tests whether the series has a break in the tested date. The tests in the second category look for the presence of a break in the series, which may exist at any time within the sample period. Some tests in this category also reveal the most possible break date as a by-product. The tests in the last category are in fact estimators, they first estimate the unknown date of the break, then test it. For any type of break, the date of the break, if it exists, is unknown so that it falls into the third category. But to understand the basics of the structural break estimators that are used to find unknown break dates and test them, it is better to start with the Chow Test. It is because unknown date estimators that use more complicated tests basically rest on the same principles as this test. Chow test looks for the following. Whether splitting data from the possible break point and estimating two generated sub samples separately by least square gives significantly better than using the whole sample at once; if the answer is yes, the null hypothesis of no break is rejected. The resulting statistics would be F-statistics, log likelihood ratio or the Wald statistic. However, as there can be more than one break in the data, the estimators can be further divided into two categories; single break estimators and multiple break estimators.

Actually it is theoretically proven that consistency for the break date estimates is satisfied for single break estimators even more than one break in the data exist (Bai and Perron, 1998)\textsuperscript{29}. This works by first finding one break in the data, and then splitting the data from there and searching for new breaks in the new samples. However, as there is no efficiency condition for any estimator, multiple break estimators are used to get more precise estimates to find smaller confidence intervals around the breaks, and also to increase the rate of convergence to the break dates. This increases efficiency in the estimation of parameter values subject to the structural change. Conversely, Multi-Equations Systems is used to get more precise estimates for any type of estimator.

Single Break Estimators

For the unknown break date, Quandt (1960)\textsuperscript{30} proposed likelihood ratio test statistics for an unknown change point, called Supremum (Max)-Test, while Andrews (1993)\textsuperscript{31} supplied analogous Wald and Lagrange Multiplier test statistics for it. Then he developed Exponential (LR, Wald and LM) and Average (LR, Wald and LM) tests. These tests are calculated by using individual Chow Statistics for each date of the data except from some trimmed portion from both ends of it. While the Supremum test is calculated for and finds the date that maximizes Chow Statistics, the most possible break point, the Average and Exponential tests use all the Chow statistic values and are only informative about existence of the break but not its date. The deficiency of the Supremum test is that it only has power if one break occurs under the alternative hypothesis. This means they do not show heterogeneity before and after the break, it is also a necessary condition for the Chow test. Heteroscedasticity and autocorrelation robust version of this test (also called Quandt Likelihood Ratio or Andrews- Quandt statistics, which is the estimator used most commonly in this literature) can be used, even though it still gives the most possible break date (it is so because of small sample properties). It also strongly suffers from large confidence intervals around the break date.

Single break model, Bai et al. (1998)\textsuperscript{32} use quasi likelihood estimation in a VAR setting and show that with common breaks across equations, the precision of the estimates increases with the number of equations in the system. However, their methodology obviously can only be carried out as long as equations are expected to show a break in the same time period. This could be the case when several variables are co-integrated. Besides, this test is designed for a single break and there could be more than one break date in the data, in which case these test exhibits non-monotonic power function (Vogelsang, 1997\textsuperscript{33}).


Multiple Break Estimators

Perron and Qu (2006) following the work of Bai and Perron first define minimum segment length (in proportion to the total data). Given this constraint, they search for the optimal partition of all possible segments of data to obtain global minimisers of the sum of squared residuals. By this way, they obtain the location of breaks, minimizing their objective function for any possible number of breaks. Then they sequentially test for whether an additional break date significantly reduces the sum of squared errors. Their methodology inherits both pure and partial structural change models. Though this method consistently identifies the break dates, this is due to when estimating a single break model in the presence of multiple breaks, the estimate of the break fraction will converge to one of the true break fractions, the one that is dominant in the sense that taking it into account allows the greatest reduction in the sum of squared residuals (in the case of two breaks that are equally dominant, the estimate will converge with probability half (½) to either break). This procedure states the fact that the method of estimation is based on the least-squares principle implies that, even if changes in the variance of error terms are allowed, provided they occur at the same dates as the breaks in the parameters of the regression, such changes are not exploited to increase the precision of the break date estimators. This is due to the fact that the least squares method imposes equal weights on all residuals allowing different weights, as needed when accounting for changes in variance, requires adopting a quasi-likelihood framework.

Finally, Perron and Qu (2007) bring a novel approach to structural change analyses which enable to find considerably small confidence intervals around the break dates. They use a multiple equation model. They first define the minimum segment length of the data that could be separated with breaks. Given this constraint, they then search for the optimal partition of all possible segments of data which the model fits, where the objective function being maximised is a quasi-likelihood one based on normal errors.

A series of data can often contain a structural break, due to a shock to the economy, that is formation of BRICS Group 2010. The F test (chow test) was applied to test the existence of endogenously determined structural break time in this date. Thus, the study signifies structural break with adopted Chow test of Perron (1989) structural break analysis model.

Further, study was analysed that a single regression line is not a good fit of the data due to the obvious structural break in 2010. Then analysis of three separate regression equations was done which are more efficient than Single regression lines.

Period 1 .......... From 1990 to 2013
Period 2 .......... From 1990 to 2009
Period 3 .......... From 2010 to 2013

This needs the Chow test, which is a variation of the F-test for a restriction expressed as:

\[ F = \frac{RSSR - (RSS_1 + RSS_2)}{K} \]

Where, RSSR = Residual sum of squares of the model on all data

\[ RSS_1 = \text{Residual sum of squares of the model before structural break time} \]

\[ RSS_2 = \text{Residual sum of squares of the model after structural break time} \]

\[ K = \text{Number of parameters estimated} \]

\[ N_1 = \text{number of observations before structural break time} \]

\[ N_2 = \text{number of observations after structural break time} \]