Research Methodology
3.1 Introduction

According to Clifford Woody, “research comprises defining and redefining problems, formulating hypothesis or suggesting solutions, collecting, organising and evaluating data, making deductions and reaching conclusions, and at last carefully testing the conclusions to determine whether they fit the formulating hypothesis.” (Kothari, 2006) D.Slesinger and M. Stephenson define research as “the manipulation of things, concepts or symbols for the purpose of generalising to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art”. (The Encyclopedia of Social Sciences, 1930). “The research process can be presented in the form of a model which usually starts with a broad area of interest, the initial problem that the researcher wishes to study.” (W, 2000).

The Oxford dictionary defines research as "investigation into and study of materials and sources in order to establish facts and reach new conclusions" and methodology as "a system of methods used in a particular area of study or activity". (The Advanced Learner's Dictionary of Current English, 1952). In literal sense, research methodology can be defined as systematic interrogation of data for extract information. It can be understood as a science of studying how research is done scientifically. It broadly defines series of steps for the undertaken research. Hence, after defining and understanding the research arena, the next step is to specify the question which needs to be rationally studied. This involves formulation of hypothesis or rather formulation of assumptions for the data. Once the hypothesis of the study is successfully formulated, the researcher contemplates the question of interest. A thorough review of problem under study helps to understand the nature of data required for the study. Data gathered can be primary or secondary in nature. A primary data set is collected by the researchers from the sample or population of study by itself. On the other hand, if the researcher wants to use data collected from other data sources by someone else, the data set is called secondary data. The data once collected from appropriate source is analysed in variety of ways.
The present research is an attempt to understand the impact of changes in the major macro-economic variables on the prices of the commodities traded on Indian Commodities Exchanges. In other words, it traces the reaction of the commodities prices to the macro economic variables. It will further analyse if it is the spot market which absorbs these shocks first and leads the future market in reacting to the shocks generated by the macro economic variables or the case stands vice versa where future market leads the spot. The study is likely to add to the knowledge of hedgers and speculators who are associated with the exchange based trading of the commodities. Moreover, an understanding of the changes in the prices of the commodities trading with respect to macro-economic variables can be of immense help for policy formulators as well. The proposed study is an attempt to understand the dynamics of prices of the commodities market.

The various types of research methods have been developed as a consequence of differing requirements of information, availability of time and need patterns. The various methods that are available today are outcome of modifications and improvements that have been done on the traditional methods over the passage of time. As a result, the new methods are relevant and befitting to the arenas of research. However, the foundation need of every methodology is the same- to derive information from the data. It is very important to use correct methodology for any research. The usage of right methods can yield fruitful results while inappropriate research methodology can prove fatal and lead to wastage of resources. Therefore, it is of utmost importance to outline the statement of problem and the research objectives at the beginning of the research.

The further section of the chapter discuss the statement of problem, objectives of study and the research questions and then specifies the methodology of the research. It also provides information regarding commodities under study and the period of research.


3.2 Statement of problem

The primary objective of this research is to analyse the volatility in the commodities market in India and to study the association between volatility of prices of commodity under study and volatility in macro-economic variables. The exchange based trading of the commodities has given a new face to the pricing of commodities in India. It has depicted commodities as an investment tool in India giving huge powers to speculators in trading. The high volume of speculative trade has made commodities prices on exchange platform more linked to the economic factors in India rather than to demand and supply forces. (Robles, Torero, & Braun, 2009)(Ghosh, 2009)(Kabra, 2007)(Susan, 2003)(Naik & Jain, 2002). The study is an attempt to understand the movement of commodity prices with respect to changes in macroeconomic variables in India. The Commodities Market in India today stands as an important segment in the economy providing numerous benefits to producers, investors, manufacturers and consumers. The basket of benefits includes price risk management, price discovery, diversification of portfolio risk, etc. Therefore there is a need to study how the macro economic factors impact the volatility in the commodities market keeping in mind that the portion of speculative trade in commodities is surging day by day in India and the market is dealing with various sensitive issues regarding goods of consumption.

3.3 Objectives of Study

The current study was proposed with the following objectives:

(i) To understand the structural breaks in the prices of commodities under study and to reason out these breaks

(ii) To study the lead lag relationship between spot and future market of commodities under study and to further know the existence of arbitrage opportunity (if any) and understand the efficiency of markets
(iii) To understand the impact of SENSEX (considered as a barometer of performance of stock market in India) on spot and future prices of the commodities under study

(iv) To understand the impact of US Dollar to Indian Rupee exchange rate (considered as a barometer of performance of foreign exchange market) on spot and future prices of the commodities under study

(v) To understand the impact of Whole Sale Price Index Announcements (considered as a barometer of inflation in Indian economy) on spot and future prices of the commodities under study

3.4 Research Questions

The central theme of this research is to understand the volatility in the Indian commodity market with a special reference to lead lag relationship between spot and futures market. The lead lag relationship between spot and future market of a commodity is given a special attention to understand if it is the spot market or the future market which reacts to the new shock introduced by the macro economic variables first. This is an important issue for the retail investors as they can look for existence of arbitrage opportunity between two markets with this information. On the other hand, this information is also important for policy makers as this can help them understand the efficiency in the market.

Taking these factors into consideration, the key questions of the research are:

(i) Are there any structural breaks in the prices of commodities under study? If yes, then what lead to these structural breaks in the data?

(ii) Is there any lead lag relationship between spot and future prices of commodities under study? Has this relationship changed with occurrence of structural breaks?
(iii) Is there any lead lag relationship between SENSEX (considered as barometer of performance of equity market in India) and spot prices of commodity under study? Has this relationship changed with occurrence of structural breaks?

(iv) Is there any lead lag relationship between US Dollar and Indian Rupee exchange rate (considered as barometer of performance of foreign exchange market in India) and spot prices of commodity under study? Has this relationship changed with occurrence of structural breaks?

(v) Is there any lead lag relationship between Whole Sale Price Index Announcements (considered as barometer of inflation in India) and spot prices of commodity under study? Do WPI Announcements show any unusual changes in prices of commodities under study? Has this relationship changed with occurrence of structural breaks?

(vi) Is there any lead lag relationship between SENSEX (considered as barometer of performance of equity market in India) and future prices of commodity under study? Has this relationship changed with occurrence of structural breaks?

(vii) Is there any lead lag relationship between US Dollar and Indian Rupee exchange rate (considered as barometer of performance of foreign exchange market in India) and future prices of commodity under study? Has this relationship changed with occurrence of structural breaks?

(viii) Is there any lead lag relationship between Whole Sale Price Index Announcements (considered as barometer of inflation in India) and future prices of commodity under study? Do WPI Announcements show any unusual changes in prices of commodities under study? Has this relationship changed with occurrence of structural breaks?
3.5 Commodities under Study

The proposed study has restricted its scope to three commodities. These commodities are:

(i) Copper

(ii) Silver

(iii) Crude Oil

The importance of these three commodities in Indian Economy is given below.

3.5.1 Copper: Copper is the most important and popular nonferrous metal used in various kinds for multiple uses (Cutler, 2010). The metal is best known for its conductive and antibacterial properties. The metal is also used for making various alloys like brass which are further put to various uses. The demand for the metal is highest in the electronic industry wherein the metal is used in form of cables, winding wires, electronic chips, etc. The metal is also put to use in construction industry for plumbing, valves and other fitting components. The metal forms important component for various vehicles. As per an estimate by ICSG (International Copper Study Group) most cars contain an average of 20 kg copper and luxury & hybrid vehicles contain about 45 kg copper (Government of India Ministry of Mines, 2011). Since the metal is put to various economic uses, the metals are supposed to be an important indicator for the economy (Saefong, 2013). The booms in the economies are usually characterized by rise in copper demand and the recessionary phase is supported by poor demand for the metal. The metal is widely known as ‘Dr. Copper’ because of its ability to predict the health of the economy. Keeping in mind the importance of the metal in the economic development, the present study is an attempt to understand the causality of volatility in the prices of metal with the help of various economic factors. The figures depicting the relationship between copper consumption and GDP growth and IIP growth depict interesting relationship between the variable. With exceptional to 2008,
since 2006-07 to 2010-11\(^1\) the consumption of copper has shown positive relationship with GDP growth and IIP growth. The year of 2008 was the year of global recession and as a result the government interfered in the working of the economy and released certain bailout packages. Such packages involved starting of certain infrastructure projects by Government of India which had important impact on the demand for the metal. These packages could not have a very improving impact on the IIP growth or GDP growth. Hence, the property of metal as 'Dr. Copper' might not be justified in case of India in 2008. This also adds on to the significance of the proposed study. The case is similar with the IIP statistics. The relationship between IIP and consumption of copper is well illustrated in fig. 3.3. Hence, the property of metal as 'Dr. Copper' can be justified in case of India. This also adds on to the significance of the proposed study.

\[\text{Fig.3.1: Copper Consumption by End Use}\]

\(^1\) Data after 2010-11 on consumption of copper is not available
Fig. 3.2: Consumption of Copper (in Tonnes) and GDP of India


Fig. 3.3: Consumption of Copper (in Tonnes) and IIP Growth (%)

3.5.2 Silver: Silver is one of the oldest metals known to mankind. The metal occupies an important status in the Indian history, tradition and economy making it the most exchange based traded metal in India. Silver has a trading volume of 7.16 lakh tonne which was approximately 28% of total value of commodities traded on MCX (Forward Market Commission, 2013). The major components of silver demand are Industrial use (54%), Photography (15%), Jewellery and Silverware (26%) and Coins (5%) (Aggrawal & Bhargava, 2011). The metal was chosen for analysis because the metal is most traded on the Indian Commodity Exchanges but has a week existing literature. There are very few studies which have been done on Silver prices in India in spite of the fact that it is the most traded precious metal in the country.

3.5.3 Crude Oil: Crude Oil can be rightly called the bloodline of an economy. Any disturbances in crude oil prices produce a direct and indirect impact on various arms of economy. The fuel prices have the power to influence the inflation level in an economy. It is one of those commodities which drive the economy. Oil is the largest traded commodity in the world both in value terms and volume. The crude oil prices are of prime importance to Indian economy. Any increase in the crude oil prices increases the economic growth by increasing the 'real resource' prices in the economy which is also called 'revenue effect'. However, the effect does not stop here. The increase in the value of trade leads to increase in growth of liquidity. But when the oil prices begin to fall, a wave of panic spreads in the economy. The oil prices are indicator of global sentiments towards the economic future of the economy. The logic is straight. A fast growing economy will need more oil while a sluggish economy needs less oil. Since the crude oil price are determined by demand and supply, they indicate the opinion of global experts towards our economy. The crude oil prices are an important input to determine if the economy is heading towards boom or recession. However, it is to be noted that sudden crash or boom in the prices may not be only by demand or supply
but also by sudden political, economic or geographical event. A further analysis of such events will be done in subsequent chapter.

3.6 Data Collection

The data requirement for the proposed study is as follows:

(i) **Spot and Future Prices of Copper, Silver and Crude Oil:** The daily closing spot and future prices of commodity under study were collected from the official website of Multi Commodity Exchange (MCX). MCX is the tenth largest exchange in the world in terms of volume of trade (Acworth, 2013). The exchange has market share of about 87% and has the highest volume of trade in India. (Forward Market Commission, 2013)

(ii) **Daily Closing Values of Sensex:** The study uses daily closing value of Sensex as an indicator of performance of stock market in India. Stock Market and commodity market have emerged as important arm of Indian Capital market where performance of one can influence the performance of the other. (Creti, Joëts, & Mignon, 2012)(Silvennoinen & Thorp, 2010)(Dwyer, Gardner, & Williams, 2011)(Vivian & Wohar, 2012)(Zapata, Detre, & Hanabuchi, 2012). The proposed study will analyse how the performance of stock market has influenced the volatility in commodities market in India.

(iii) **US- Dollar Exchange Rate:** The US Dollar is considered as "world's reserve currency". The countries all over the world hold US Dollar denominated assets as they are considered as "best way to hold foreign exchange"(U.S. Department of State)(Conerly, 2013). Dollar is known as an international currency for its international acceptance. Any fluctuations in dollar prices can influence the cost of exports and imports of any economy which in turn can impact the balance of payments of the economy. Moreover, India is world's number 3 crude importer, importing 3.86 million barrels-per-day of crude oil in 2013(Reuters, 2014). The country is among top ten importer of copper and a
net exporter of refined copper (Jolly, 2013)(InfodriveIndia.com, 2013). India is the one of the leading importer of silver.(Rudarakanchana, 2014). Since the metals under study are either exported or imported, the exchange value of global currency will stand as a strong factor influencing the volatility of prices of commodities under study. The study uses daily closing exchange value of US Dollar to Indian rupee as an indicator of foreign exchange market.

(iv) **Whole Sale Index Announcements:** The investment decisions are greatly influence by inflation related announcements. This is more relevant for investments in precious metals where people invest with a tendency to beat inflation. The proposed study is an attempt to understand if WPI announcements influence the price volatility in commodities under study. Inflation logically shares a direct relationship with crude oil (Hooker, 2002)(Cunado & Gracia, 2005). Investors make investments in precious metals to beat inflation (Conover, Jensen, Johnson, & Mercer, 2009)(Barnhart, 1989). Keeping these factors in mind, the influence of WPI announcements will be studied on price volatilities of commodities under study. The WPI announcements were released on weekly bases till January, 2012 (every Friday). From thereon, the government of India has decided to discontinue the weekly WPI announcements and come up with monthly announcements (14th of every month and in case of holiday, the data is released on next business day)(Press Trust of India, 2012). The effect of WPI announcement will be captured with the help of dummy variable which will take a value ‘1’ when there is a WPI announcement and take the value ‘0’ otherwise.

### 3.7 Period of Study

The exchange based commodities trading was established in India in 2003 with the establishment of 4 national commodity exchanges. However, the spot price data for various commodities was maintained by the exchange only from May, 2005. Hence
the period of the study is from 1 May, 2005 to 31 September, 2013. The researcher has taken next six months for the analysis of data. The application of models and test required a thorough study. Keeping in mind the nature of models and complexity of data, the expansion of time period of study was not feasible. Therefore, the study has been restricted to 31 September, 2013.

3.8 Research Methodology

The further section discusses the methods/ models/ techniques, etc. which the study proposes to use on the data to extract the required information for the raw data.

3.8.1 Structural Breaks in Data

An unexpected shift in the time series leads to structural break in the data. This can lead to change in the slope of the data which can lead to forecasting errors and increase the unreliability of the model. The existing literature indicates that the commendable work of Chow (1960) and Quandt (1992) focuses on test of one structural break in the data. A further development of literature led to development of test for estimating multiple structural changes at unknown break dates. The various prominent test proposed were by Andrews (1993), Andrews and Ploberger (1994) and Bai and Perron (1998). The study uses the test proposed by Bai and Perron (1998) as it is more applicable to our study.

The Bai–Perron (1998) test is based upon "information criterion in the context of a sequential procedure, and allows one to find the number of breaks implied by the data, as well as estimating the timing of the breaks and the parameters of the processes between breaks." It provides for multiple breaks at unknown dates. Also, the test is simple to apply in contrast to other test in the existing literature.

Bai and Perron have proposed two methods to test I+1 breaks given I breaks. The first is a sequential procedure which is based on each new null hypothesis on every previous break date generated. The second procedure is also called the sequential
test under the global null which is based on the break dates that are global minimisers of the sum of squared residuals. The current research follows the second approach.

Consider the following $m$-break model:

$$y_t = x_t \beta + z_t \delta_1 + \mu_t, t = 1, 2, \ldots, T_1$$

$$y_t = x_t \beta + z_t \delta_2 + \mu_t, t = T_1 + 1, 2, \ldots, T_2$$

$$\vdots$$

$$y_t = x_t \beta + z_t \delta_{m+1} + \mu_t, t = T_m + 1, 2, \ldots, T,$$

... Equation 1

Where,

$y_t =$ Value of dependent variable at time $t$

$x_t =$ $p \times 1$ vector of covariates with corresponding vector of coefficients $\beta$

$z_t =$ $q \times 1$ vector of covariates with corresponding vector of coefficients $\delta_j$ ($j = 1, 2, \ldots, m+1$)

$\mu_t =$ Disturbance or error component at time $t$

$T_1, \ldots, T_m =$ Unknown Breakpoint dates

The each derived set of break dates will divide the data into a separate segment where each fragment will have a different model associated with it. Therefore the derived segments will be separated treated for the derived objectives. The data set of the commodity spot and future prices will be subject to the structural test and each derived division will be separately treated.

(For details refer, (Bai & Perron, 1998))
3.8.2 Measures of Central Tendency

The study uses the data collected from various sources in its original form. The series are then subjected to initial analysis with the help of basic statistical tools before proceeding with econometric modelling. An initial assessment of mean and standard deviation of the series is done to comment on the volatility of prices of commodity under study. Standard deviation is often used as a measure of the risk associated with price-fluctuations of a given asset. Standard deviation is a statistical measurement that sheds light on historical volatility. The average volatility in the prices of commodities under study is an important determinant while undertaking any investment decision. The study of basic statistical measures will be coupled with graphical study. Other statistical measures under study are skewness and kurtosis. Skewness measure will help us understand the asymmetry from the normal distribution in a set of statistical data while Kurtosis measures the degree to which a distribution is more or less peaked than a normal distribution. The disturbance of the data from the condition of normality will be further confirmed with the help of JarqueBera test which is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution. The calculation of the test statistics is defined in equation 2:

\[ JB = \frac{n}{6} (S^2 + \frac{1}{4}(K - 3)^2) \]

... Equation 2

*Where,*

\[ JB = \text{Jarque- Bera Test Statistics} \]

\[ N = \text{Number of observations} \]

\[ S = \text{Skewness of the data} \]

\[ K = \text{Kurtosis of the data} \]
JB statistic asymptotically has a “chi-squared distribution with two degrees of freedom, so the statistic can be used to test the hypothesis that the data are from a normal distribution. The null hypothesis is a joint hypothesis of the skewness being zero and the excess kurtosis being zero. Samples from a normal distribution have an expected skewness of 0 and an expected excess kurtosis of 0. As the definition of JB shows, any deviation from this increases the JB statistic and indicates that the data is not normally distributed.” Gupta (1985)

3.8.3 Test for Stationary of Data

After understanding the condition of normality of data, the study will proceed to understand if the data under study is stationary or not. A stationary process is the one whose joint probability distribution does not change when shifted in time. Consequently, parameters such as the mean and variance, if they are present, also do not change over time and do not follow any trends.

A Time Series is stationary if has the following conditions:

(i) Constant $\mu$ (mean) for all $t$.

(ii) Constant $\sigma$ (variance) for all $t$.

(iii) The autocovariance function between $X_{t_1}$ and $X_{t_2}$ only depends on the interval $t_1$ and $t_2$.

In other words, a stationary time series $(X_t)$ must have three features: finite variation, constant first moment, and that the second moment only depends on $(X_{t_1},X_{t_2})$ and not depends on $t_1$ and $t_2$.

The modern econometrics describes various tests to check for the stationary of the data. However, these tests have some or the other drawback. Glynn, Perera, & Verma (2007) has done an extensive review of the recent developments in testing of the unit root hypotheses in the presence of structural change and provided empirical evidences. The current study uses Augmented Dickey Fuller Test for the analysis.
The Augmented Dickey Fuller Test (1984) is an improvement over the original Dickey Fuller Test (1979). The Dickey Fuller test stated that a simple autoregressive model (where the variable under study is regressed on its previous value) is indicated by:

\[ y_t = \alpha y_{t-1} + \mu_t \]  

... Equation 3

Where:

- \( y_t \) = Variable under study
- \( t = \) time index,
- \( \alpha = \) coefficient,
- \( \mu_t = \) error term.

If \( \alpha = 1 \), a unit root is present in the data meaning that the series of the variable under study \( y_t \) is non-stationary in this case. The ADF test removes all the structural effects (autocorrelation) in the time series and then tests using the same procedure. The testing procedure for ADF is defined in equation 3:

\[ \Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p+1} + \varepsilon_t \]  

... Equation 4

Where,

- \( \alpha = \) constant,
- \( \beta = \) coefficient on a time trend and
- \( p = \) the lag order of the autoregressive process.
Imposing the constraints $\alpha = 0$ and $\beta = 0$ corresponds to modelling a random walk and using the constraint $\beta = 0$ corresponds to modelling a random walk with a drift. By including lags of the order $p$ the ADF formulation allows for higher-order autoregressive processes.

The hypothesis under study while applications of the test are:

$H_0 =$ Series has a unit root (Series is non-stationary)

$H_1 =$ Series does not have a unit root (Series is stationary)

The null hypothesis will be rejected if the calculated $t$-statistic is more than the critical values and the null hypothesis will be accepted if the calculated $t$-statistic is less than the critical values. In case the null hypothesis is rejected and the data is found to be non-stationary, the first difference of the data is calculated. The first difference of a time series is its change between period’s $t$ and $t-1$. In other words,

$$\Delta Y_t = Y_t - Y_{t-1}$$

.... Equation 5

The difference of the series will be calculated till it becomes stationary.

3.8.4 Test for Linear Property of Data

One of the important assumptions with most of the models is the linear property of the parameters or rather the regression coefficients with the response variable. This is an important restriction of the variables under study. An extensive review of literature of the studies conducted in the proposed area did not take this property of data into consideration while framing the research methodology. This fact was specifically studied in the Indian context. The researcher proposes to test this property of data at the initial level of the study. The study uses two tests to check for linear property of data:

(i) BDS Test

(ii) Variance Ratio Test
3.8.4.1 BDS Test:

The BDS test statistics uses the concept of correlation integration which measures how frequently, the temporary patterns repeat themselves in the data. For a time series, correlation integration can be estimated by:

\[ C_{m,e} = \frac{2}{T_m(T_m - 1)} \sum_{m \leq s < t \leq T} I(x_t^m, x_s^m; \epsilon) \]

... Equation 6

Where,

\[ x_t^m = (x_t, x_{t-1}, \ldots, x_{t-m+1}) \] is the m-history of a time series \( x_t \) for \( t = 1, 2, \ldots, T \).

\( T_m = T - m + 1 \) and

\[ I(x_t^m, x_s^m; \epsilon) \] is an indicator function which is one if \( |x_t^m - x_s^m| < \epsilon \) and zero otherwise.

The correlation integral stated above estimated the probability that any two m-dimensional points are within a distance of \( \epsilon \) of each other. The joint probability is given by

\[ \Pr (|x_t - x_s| < \epsilon, |x_{t-1} - x_{s-1}| < \epsilon, \ldots, |x_{t-m+1} - x_{s-m+1}| < \epsilon) \]

... Equation 7

Now if \( x_t \) are iid, the limiting case will be defined by:

\[ C_{1,e}^m = \Pr (|x_t - x_s| < \epsilon)^m \]

... Equation 8

Brock, Dechert, Scheinkman, & LeBaron (1996) defined BDS statistics as:
Where,

\[ S_{m,e} \text{ is the standard deviation of } \sqrt{T} (C_{m,e} - C_{1,e}^m) \]

The BDS statistics converges to \( N (0, 1) \) under moderate regularity conditions as:

\[ V_{m,e} \xrightarrow{d} N (0, 1) \]

\[ \text{... Equation 10} \]

### 3.8.4.2 Variance Ratio Test:

The study proposes to use the variance ratio test proposed by (Lo & Mackinlay, 1988) to check the non-linearity of data. The test is based on the concept that the return of any series \( x_t \) follows pure random walk in a manner that:

\[ x_t = \emptyset + x_{t-1} + \varepsilon_t \]

\[ \text{... Equation 11} \]

Where,

\[ \emptyset = \text{drift parameter and,} \]

\[ \varepsilon_t = \text{error component which is Independent and Identically Distributed (i.i.d.)} \]

In such case, the variance of the \( k \)-differences will grow proportionally with the difference \( k \). Therefore, to check for random walk, we can compare the variance of one period with that of \( k \)-period returns using the formula:
\[
VR(k) = \left( \frac{1}{n} \right) \frac{VAR(x_{t+n} - x_t)}{VAR(x_{t+1} - x_t)}
\]

...Equation 12

Lo and MacKinlay (1988) proved that if \( x_t \) is IID, then under the null hypothesis \( VR(k) = 1 \). In that case, the test statistics \( M_1 \) is given by:

\[
M_1(x; k) = (VR(x; k) - 1) \left[ \frac{2(2k - 1)(k - 1)}{3KT} \right]^{-\frac{1}{2}}
\]

...Equation 13

The test statistics will follow normal distribution. With the passage of time, researcher realized that the financial time series have the property of conditional heteroskedasticity. Therefore Lo and Mackinlay, (1989) provided a new test statistics \( M_2 \) which is defined as:

\[
M_2(x; k) = (VR(x; k) - 1) \left[ \sum_{j=1}^{k-1} \left[ \frac{(2k - j)^2}{K} \delta_j \right] \right]^{\frac{1}{2}}
\]

...Equation 14

(see Hoque, Kim, & Pyun, 2007).

3.8.5 Test for Cointegration:

The researcher first used the Johansen Cointegration to find the long term association in the variables under study. The test is used to find the association between the linear variables. But since the variables under study were found to be nonlinear, the proposed test provided spurious results and indicated that there is no association between any variables under study. Therefore, the researcher then used the nonlinear test for cointegration which is defined in the subsequent section.
3.8.6 Johansen Test for Cointegration

The test for cointegration helps to know if the data share a long term relationship or not. If the data under study does not share a long term relationship then it is not appropriate to further conduct any other modelling on the data set. The long term association of the variables under study will be checked with the help of Johansen Cointegration Test. This test was first proposed by Johansen (1988). The concept considered Vector Autoregression of the variable (VAR) with $p$-dimensions and integrated of order $d$.

$$A_k(L)X_t = \mu_0 + \Psi D_t + \varepsilon_t$$

... Equation 15

Where,

$D= \text{Dummy Variable in the study}$

The method is applicable to the variables of the same order only. The VAR levels are now transformed to Vector Error Correction Model as:

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{k-1} \Delta X_{t-k+1} + \mu_0 + \Psi D_t + \varepsilon_t$$

... Equation 16

Where,

$\Gamma_t$ and $\Pi$ are matrixes of variables

The VECM can be written in a more compatible way as:

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-1} + \mu_0 + \Psi D_t + \varepsilon_t$$

... Equation 17
The number of cointegration equation is represented by $\Pi$ matrix. The matrix is filled with zero in case there is no cointegration.

(For details refer (Sjö, 2008))

When the data was check for Johansen Cointegration, there was no cointegration which could be found between variables under study. This meant that the researcher could not proceed with analysis on data. The researcher here would like to bring to the notice that the Johansen Cointegration Test assumes linear combinations of the parameters. But, the prior tests for linear property have indicated that the data is nonlinear. Therefore, there was a need to use a nonlinear test for cointegration analysis rather than the traditional Johansen Cointegration Test. Hence, at this stage, the researcher would like to emphasise on the check for linearity of data when the application of model demands a linear relationship. A lot of studies in the existing literature have no mention about this property of data.

### 3.8.4.2 Non-linear Cointegration

The study uses the popular Rank Test for Non-linear Cointegration to check the nonlinear cointegration in the data. The test was developed by Breitung (2001). He defined a new series of rank of variables under study than the variables in raw form.

For example, consider two series, $x_t$ and $y_t$. Breitung (2001) defined two rank series and named them as $R_T(x_t)$ and $R_T(y_t)$ where,

$$R_T(x_t) = \text{Rank of } x_t \text{ among } x_1, \ldots, x_T$$

And

$$R_T(y_t) = \text{Rank of } y_t \text{ among } y_1, \ldots, y_T$$

And $x_t$ and $y_t$ are integrated of order one $I(1)$ and the two series and nonlinearly related as
\[ y_t = f(x_t) + \mu_t \]

\[ \ldots \text{Equation 18} \]

Breitung (2001) further adds on “the sequence of ranks is invariant to a monotonic transformation of the data.” The distance measure between the ranks of two series is given by

\[ K_T = T^{-1} \sup_{t} |d_t| \]

\[ \ldots \text{Equation 19} \]

And

\[ \xi_T = T^{-3} \sum_{t=1}^{T} d_t^2 \]

\[ \ldots \text{Equation 20} \]

Where

\[ d_t = R_T(y_t) - R_T(x_t) \]

\[ \ldots \text{Equation 21} \]

The rank transformation is also capable of limiting the effect of extreme values and this leads to “distribution of a test statistic based on ranks tends to be more concentrated around its mean” Breitung (2001). Further, the two series \( f(x_t) \) and \( g(y_t) \) are correlated and converge to Brownian motions \( W_1(r) \) and \( W_2(r) \) with a correlation coefficient:

\[ \rho = E[W_1(1)W_2(1)] \]

\[ \ldots \text{Equation 22} \]

Breitung (2001) uses Monte Carlo simulations to calculate the relationship between two parameters as \( f(x_t) \) and \( g(y_t) \) are not observable and hence cannot be directly used to calculate \( \rho \).
The series of models are used as per the objectives of the research. The application and analysis of the output of models is discussed in subsequent section.

(for details, refer Breitung (2001))
References


