CHAPTER 4

RESEARCH METHODOLOGY

This chapter deals with various research methods adopted in the study. The description and sources of data, population and the sample for the study along with a detailed description of various statistical and econometric tools used have been presented in this chapter.

4.1 SOURCES OF DATA

The study has been made to analyze the market behavior, Price Discovery and volatility spillover of Indian commodity futures market and respective spot markets. The impact of selected macroeconomic variables on commodity futures market has also been a analysed. The daily commodity futures and spot data has been collected from the website of MCX. The data of macroeconomic variables have been collected from the website of RBI.

4.2 POPULATION OF THE STUDY

The study focuses on the price discovery and volatility spillover process of Indian Commodity Market. As on March 31st 2017, there are 6 national level commodity exchanges and 4 regional level commodity exchanges and about 40 commodities are allowed to trade in these exchanges.

4.3 SAMPLING METHOD

The sampling method used in the study is Purposive sampling. The purpose of the study is to analyse the price discovery process and volatility spillover of Indian Commodity Market. Since Multi Commodity Exchange has more than 80 % of the market share, this exchange has been chosen for the study to represent the Indian commodity market. The thirteen actively traded commodities which are traded on all days in the Multi commodity exchange from the sectors Bullion, Base Metals, Energy and Agricultural segment have been taken up in the study.

4.4 SAMPLE SIZE

In the study thirteen actively traded commodities which are traded on all days in Multi commodity exchange have been included. Some commodities which are no longer trading in MCX have been omitted. Out of the thirteen commodities two are from Bullion Segment (Gold and Silver); two from Energy Segment (Crude Oil and Natural Gas); five from Base Metals Segment (Copper, Aluminium, Zinc, Lead, Nickel) and four are agricultural commodities (Cardamom, Mentha Oil, Crude Palm Oil and Cotton).

4.5 PERIOD OF THE STUDY

The period of the study is from April 2007 to March 2017. However, data period varies across the commodity futures due to late introduction of futures contract in the MCX. The list of sample commodities and the respective data periods have been presented in the table no:

Table no: 4.1 List Of Sample Commodities And The Respective Data Period And Spot Market.

SL.NO	COMMODITY	DATA PERIOD	SPOT MARKET
Bullion			
1.	Gold	01/04/2007-31/03/2017	Ahmedabad
2	Silver	01/04/2007-31/03/2017	Ahmedabad
Energy			
3.	Crude Oil WTI	01/04/2007-31/03/2017	Mumbai
4.	Natural Gas	01/04/2007-31/03/2017	Hazira
Base Metals			
5.	Copper	01/04/2007-31/03/2017	Mumbai
6.	Aluminium	01/04/2007-31/03/2017	Mumbai
7.	Zinc	01/04/2007-31/03/2017	Mumbai
8	Lead	01/04/2007-31/03/2017	Mumbai
9	Nickel	01/04/2007-31/03/2017	Mumbai
Agricultural Commodities			
10	Cardamom	01/04/2007-31/03/2017	Vandanmedu
11	Mentha Oil	01/04/2007-31/03/2017	Chandausi
12	Crude Palm Oil	07/06/2008-31/03/2017	Kandla
13	Cotton	07/10/2011-31/03/2017	Rajkot

4.6 DATA DESCRIPTION

The research uses data which are secondary in nature. The selected macroeconomic indicators are collected from the website of RBI. The daily spot and futures data are collected from the website of MCX. The study uses logarithmic transformation of the variables to minimize the heteroscedastic nature.

4.7 HYPOTHESES OF THE STUDY

The following hypotheses have been framed based on the objectives of the study.

H_0^1: There is no impact of selected macro economic variables on selected commodity futures

H_0^2: The future and spot prices of selected commodities are not stationary.

H_0^3: There is no cointegration between future and spot prices of selected commodities.

H_0^4: There is no long run causality between future and spot prices of selected Commodities.

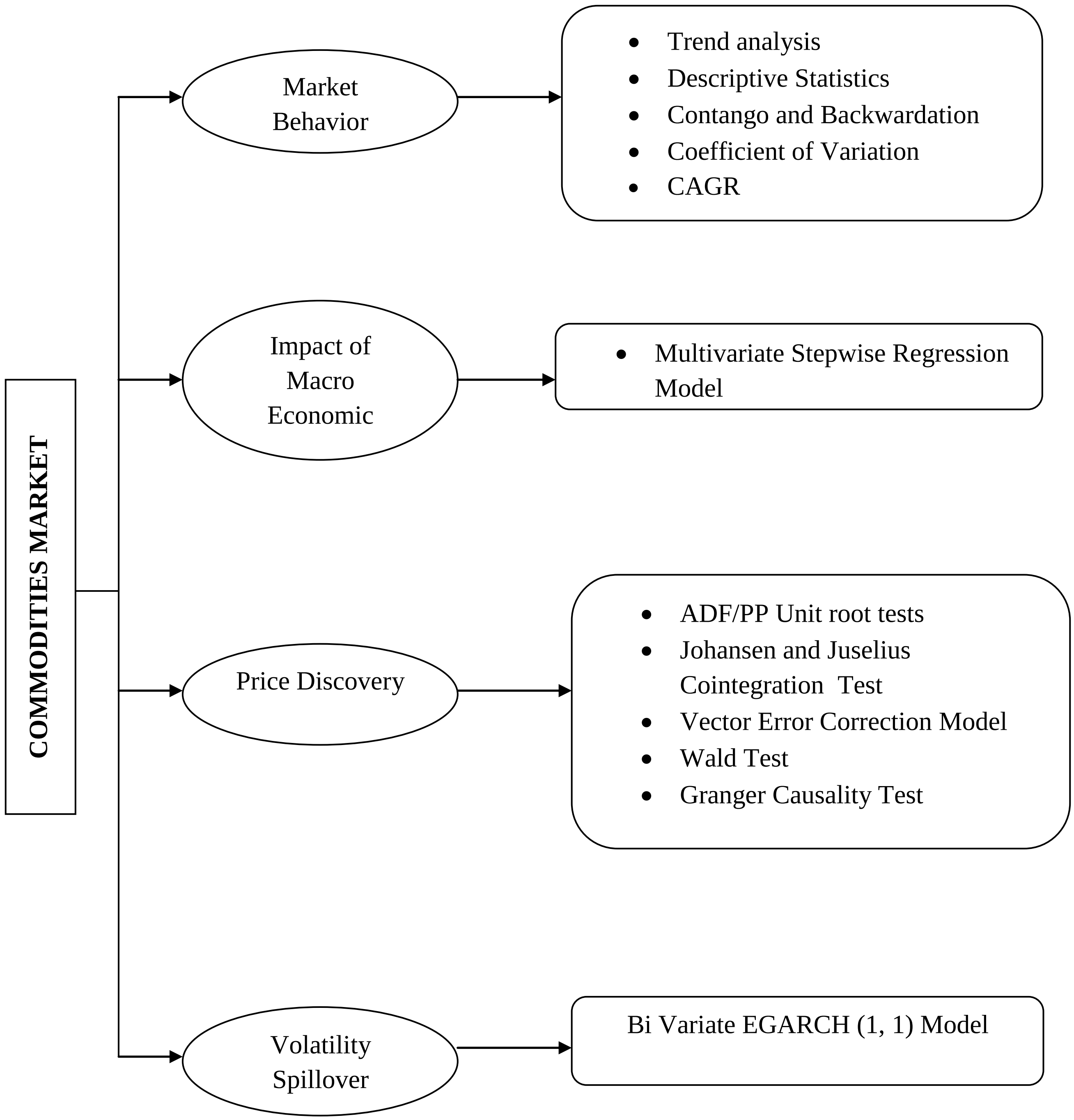
H_0^5: There is no short run causality between future and spot prices of selected Commodities.

H_0^6: There is no lead lag relationship between future and spot prices of selected commodities

H_0^7: There is no volatility spillover between future and spot prices of selected commodities.

4.8 RESEARCH DESIGN

The research is largely descriptive in nature. The market behavior of commodity spot and futures market is analysed using descriptive statistics, Contango/Backwardation, CAGR and coefficient of variation. The impact of Macroeconomic indicators on commodity market is analysed using Multivariate step wise regression to avoid spurious regression. The research study aims to find out the price discovery of Indian commodity market using various econometric tools like ADF and PP tests, Johansen and Juselius Cointegration test, VECM, Wald test and Granger causality tests. The volatility spill over between spot and future market has been analysed using Bi variate EGARCH Model. The research framework used for the study is presented in the following diagram.



4.9 TOOLS USED FOR THE STUDY

4.9.1TREND ANALYSIS

Trend analysis is a method of technical analysis where past data of the series are used to predict the future movement of the series. It is used to identify the general pattern of movement of the series and how the series move in relation to other.

4.9.2 DESCRIPTIVE STATISTICS

Descriptive statistics, as the name suggests involves in describing the basic features of the data. It mainly involves in summarizing and simplifying the data for the purpose of meaningful analysis. It involves various measures which help to bring out the characteristics of the numerical data and systematic reporting thereof, which will help in further discussions, analysis and interpretations. The various measures which are used in descriptive statistics are measures of central tendency, dispersion, skewness and kurtosis which are used to highlight the hidden characteristics of the data.

4.9.3 CONTANGO/BACKWARDATION

Contango and Backwardation are the terms which are used incommodity markets. Contango is a situation where future prices are traded at a premium than spot price, where as backwardation is the reverse. The Contango situation in a commodity market reveals that future prices of the particular commodity is anticipated to be more than spot price in future. The backwardation is a situation where future prices are anticipated to be less expensive than spot price.

4.8.4 COEFFICIENT OF VARIATION

The coefficient of variation, which is denoted as CV is a popular measure of relative dispersion. It is expressed in terms of standard deviation as a percent of mean. That is

$$CV=[Standard\ Deviation/Mean]*100 \qquad .....(1)$$

4.9.5 COMPOUNDED ANNUAL GROWTH RATE

The compounded annual growth rate popularly known as CAGR is the average growth rate of an investment over a period of time longer than a year. As the market is highly volatile, year on year growth rate may interpret the result wrongly. The advantage of CAGR comes in smoothening out the investment return so that it is able to understand the growth rate more clearly. The CAGR is calculated in the following manner:

$$CAGR=\left(\frac{Begining\ Value}{Ending\ Value}\right)\wedge\ (1/\text{Number of Years})\ -1 \qquad \ldots.\ (2)$$

4.9.6 MULTIVARIATE STEPWISE REGRESSION

Multivariate stepwise regression analysis has been used in the study as the independent variables have a strong association between them. When independent variables are closely related to each other, the resulting regression will have multicollinearity, which will result in spurious regression. Hence multivariate stepwise regression have been used in the study, which is essentially a multiple regression a number of times, each time removing weakest correlated variable. At the end, we are left with variables which describe the distribution best.

4.9.7 STATIONARITY /UNIT ROOT TESTS

A time series is said to be stationary when it's mean, variance and co variance at any point of time remains the same that is it is time invariant. Such a time series will tend to return to its mean which is called mean reversion. A most popular test to measure Stationarity is unit root test. Unit root tests are a descriptive tool which is used to classify a series into stationary or not stationary. If it is possible to classify the series as integrated, stationary or deterministic trend stationary, it is possible to study the long run and short run effects of the model and to set up a model where statistical inference will be meaningful. The way of transforming a non stationary series to a stationary one is called differencing. By applying unit root tests and understanding the order of integration, suitable econometric models can be applied to make inferences.

One of the basic test to test the unit root for understanding order of integration is Augmented Dickey Fuller Test. The necessary equation for estimating the Stationarity of time series using Augmented Dickey Fuller test is mentioned below:

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t, \quad \text{…… (3)}$$

where, c_t is the deterministic index of time index t and $\Delta x_t = x_t - x_{t-1}$ is the differenced series of x_t. In practice, c_t can be zero or a constant. The t-ratio of $\hat{\beta} - 1$,

$$\text{ADF-test} = \frac{\hat{\beta} - 1}{\text{std}(\hat{\beta})}, \quad \text{……. (4)}$$

Another popular unit root test which is used widely is Philips Perron unit root test which incorporates an auto correction to DF procedure to allow for auto correlated residuals. The primary difference between PP test and ADF test is that the former ignores serial correlation while the later uses a parametric auto regression to approximate the structure of errors.

4.9.8 JOHANSEN AND JUSELIUS TEST OF COINTEGRATION AND VECTOR ERROR CORRECTION MODEL

Johansen test of cointegration is one of the popular tools for checking cointegration for two time series. One of the pre conditions before doing cointegration test is that the variables should be integrated at the same order. Thus before applying cointegration technique, unit root tests are done in the series to check whether they are integrated in the same order. A n-dimensional time series are co integrated if some linear combination $\beta_1 y_{1t} + \ldots .. \beta_n y_{nt}$ of the component variables are stationary. The cointegration is different from the common equilibrium in which balance of forces creates a stable long term relationship. Cointegrating relationship exhibit instability at levels, but they exhibit a mean reverting spreads that move around same stochastic trend. Thus two integrated variables share a common stochastic trend such that their linear combinations are stationary, then the variables are said to be cointegrated.

This study employs Johansen and Juselius Cointegration test to analyse the cointegration between Commodity Future and Spot market. The procedure uses two tests to determine the number of cointegrating vectors: Maximum Eigen test and Trace test. The Trace and maximum eigen value can be computed as follows:

$$J_{trace} = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i)$$

$$J_{max} = -T \ln(1 - \hat{\lambda}_{r+1}) \quad \text{....... (5)}$$

Here, T is the Sample size, and $\hat{\lambda}_i$ is the i[th] canonical correlation. The trace statistics tests null hypothesis of r cointegrating vectors against alternative hypothesis of n cointegrating vectors. The maximum eigen value test the null hypothesis of r cointegrating vectors against the alternative hypothesis of r+1 cointegrating vectors. The critical values for maximum eigen value and trace statistics are compared with the test statistics where, if the test statistics are found to be greater than the critical values, the null hypothesis rejected.

If cointegration is detected between the series, it is concluded that there is long run equilibrium relationship between the variables, so that Vector Error Correction Model is applied in order to find out the short run properties of the time series. In case, there is no cointegration detected, directly Granger causality tests are done to establish the causal linkages between the variables. The regression equation form for VECM is as follows:

$$\Delta S = c1 + \sum_{k=1}^{n} \beta_{li} \ \Delta S_{t-k} + \sum_{k=1}^{n} \alpha_{2i} \ \Delta F_{t-k} + \rho_1 ECT_{t-1} + u_{2t} \quad \text{.....(6).}$$

$$\Delta F = c2 + \sum_{k=1}^{n} \alpha_{li} \ \Delta F_{t-k} + \sum_{k=1}^{n} \beta_{2i} \ \Delta S_{t-k} + \rho_2 ECT_{t-1} + u_{1t} \quad \text{.......(7)}$$

In VECM the cointegration Rank shows number of cointegrating vectors. A negative significant coefficient of Error correction term (ECT in above equation) indicates that any short term fluctuations between the independent variables and dependent variables will give rise to long run relationship between the variables. The error correction term shows the direction of causal relation and reveals the speed of discrepancy from equilibrium is corrected or minimized.

4.9.9 GRANGER CAUSALITY TEST

A general specification for bivariate (X, Y) can be expressed as below:

$$x_t = \gamma_0 + \sum_{k=1}^{M} \delta_k y_{t-k} + \sum_{l=1}^{N} \gamma_l x_{t-l} + v_t \quad \text{….. (8)}$$

$$y_t = \beta_0 + \sum_{k=1}^{M} \beta_k y_{t-k} + \sum_{l=1}^{N} \alpha_l x_{t-l} + u_t \quad \text{….. (9)}$$

where x_t and y_t are two variables at time '*t*'; *k* and *l* are the number of lags; β_0 and γ_0 is deterministic; u_t and v_t is an error term; α and β are coefficients on the lagged *x* and *y* values, and δ_k and γ_l are coefficients on lagged *y* and *x* values respectively. In this test, two null hypothesis have been tested, First one being, X does not granger cause Y, Second, Y does not granger cause X. If one can reject one hypothesis, but fail to reject the other one, it can be concluded that there is a unidirectional causality relationship between the variables. If one is able to reject both null hypotheses, there is a bi directional causality relationship between the variables. If it is not able to reject both hypotheses, there is no causality between the variables.

4.9.10 BIVARIATE EGARCH (1, 1) MODEL

The volatility spillover mechanism investigates how news from one market affects the volatility of other market. GARCH family models are used to model the volatility of the markets. Bollerslev's (1986) generalized auto regressive conditional heteroscedastic (GARCH) model cannot be applied as one of the assumption in that model is that, any news, whether good or bad have same consequence on volatility of the markets. Another drawback of GARCH model is that the coefficient of all the variables needs to be positive to ensure that conditional variance (measure of risk) is never negative. To overcome these weakness of GARCH model, Nelson (1991) developed an exponential GARCH (EGARCH) model which captures the asymmetric impact on volatilities and to avoid the imposing of non negativity restrictions on the coefficients. The following EGARCH model has been used in the study:

$$\ln(\sigma^2_{x,t}) = \omega_x + \alpha \left| \frac{\varepsilon t-1}{\sigma t-1} - \sqrt{\frac{2}{\pi}} \right| + \gamma_x \ln \frac{\varepsilon t-1}{\sigma t-1} + \beta_x \ln(\sigma^2{}_{2,t-1}) + \Psi_x \ln(\varepsilon^2_{y,t-1}) \quad \text{....(8)}$$

$$\ln(\sigma^2_{y,t}) = \omega_y + \alpha \left| \frac{\varepsilon t-1}{\sigma t-1} - \sqrt{\frac{2}{\pi}} \right| + \gamma_y \ln \frac{\varepsilon t-1}{\sigma t-1} + \beta_y \ln(\sigma^2{}_{2,t-1}) + \Psi_y \ln(\varepsilon^2_{x,t-1}) \quad \text{.....(9)}$$

The unrelated residuals ε_{xt} and ε_{yt} are obtained from equations (6) and (7).”This two step approach, the first step for VECM and second step for bivariate EGARCH model is asymptotically equivalent to joint estimation for the VECM and EGARCH models”(Greene, 1997).