CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

Research methodology is the systematic process dealing with identification of problem, collection of facts or data, analyzing these data and reaching a certain conclusion either in the form of solutions towards the problem concerned or certain generalization for some theoretical formulation. Present chapter provides the detailed framework of research methodology used for carrying out the present research work. Efforts have been made to match the characteristics of research objectives and research methods. It also includes the details of the methodology used in the present research and justification for the use of the particular methodology in this study. As the study purports to use a high degree of quantitative modeling along with the primary research to understand and analyze the first-hand experience. The methodology has been carefully drafted so as to ensure a judicious mix of quantitative interpretations and intuitive insights. The basis of selection of appropriate methodology was the outcome of the previous research work and scope of its implementation in the Indian context. Appropriate adjustments have been made depending upon the requirements and references are given from different chapters to avoid duplication.

3.2 Research design

The concept of research design is explained by Kerlinger [177] in a very lucid manner as, “research design is the plan, structure, and strategy of investigation conceived so as to obtain the answer to the research question and control variance.” The research design is a master plan outlines the possible conclusions that could be drawn from specified methods and procedures, data collected and analysis for the research. Therefore, the above-mentioned explanation clearly envisages the importance of the research design and its impact on the final outcome. Many researchers emphases the need to clearly chart a research design [1, 17]. The most common research designs used by the researchers are exploratory, descriptive and causal. In the present study, the exploratory and descriptive study has been used as a purpose of the study to obtain and analyze the data. Hence, an outline is presented here that details a scheme from the formulation of hypothesis and its operational implementations to the analysis of data and final synthesis.

3.3 Outline of the study
3.3.1 Conceptualization of the problem

The idea for the present research work dates back to an age old concept of commodity derivatives. Though the basic objective and theme of commodity derivatives may not have changed over a period of time, the mechanism and practices have certainly undergone a huge change. The effectiveness of commodity derivatives dwells upon the performance of three basic functions namely, a) Risk (volatility) management of derivative market b) Price discovery, c) Hedging effectiveness.

Volatility in the prices of commodities or price risk is one of the most significant risks for traders, manufacturers, and consumers. Volatility is characterized by its stylized facts namely volatility persistence, mean reversion and leverage effect. The accuracy in forecasting of any model depends on how effectively these basic features of the volatility have been modeled. Thus, before modeling the volatility, it is imperative to get the facts straight regarding the basic features of volatility. A number of research studies have aptly identified the presence of stylized facts of volatility in the equity derivatives market and identify it as an important tool to manage this risk. The general approach used in literature is the simplistic technique, like historical volatility methods, and OLS methods without realizing its limitation. But, time and again, the academic research has proposed to use certain time series models which are entirely the domain of sophisticated investors. In the context of Indian commodity market, there is a dearth of literature which deals with time series model for volatility characteristics. Similarly, the derivative securities have also been seen with suspicion and, time and again, the doubts are raised on their utility for price discovery and hedging.

This study focuses on the commodity derivative securities in the Indian context. To completely understand the efficiency of the commodity derivative market, the study first explores the presence of stylized fact of volatility and the impact of the flow of information on the volatility dynamics by using GARCH family models. Then, Vector Autoregressive models (VAR) and time series sophisticated techniques are used to study market efficiency and unbiasedness of commodity derivative market with the effectiveness in price discovery and volatility spillover function. Further, a comparative analysis is done to compare the appropriateness of static and dynamic hedge models and then tests their efficiency in reducing price variance.

Since the formal commodity derivatives have been introduced in Indian market in the year 2003, there is little evidence available as regards the efficiency of these securities. Though some
studies have been conducted in the past [3, 95, 192, 193, 172, 261, 292] but they are constrained by either the sample size or the period of study or the coverage of some relevant issues. Most of the past studies are concentrated on the regional commodity exchanges.

Keeping this in view, the present study is undertaken with the objectives mentioned in Table 3.1. It seeks to extend the scope of previous studies and improve upon them in terms of methodological relevance and longer time period.

3.3.2 Identification of issues and their classification

The present study is dealing with the following issues of Indian commodity market which have been identified by extensive literature review and discussions with the academic researchers and industry professionals.

1) Study of stylized facts of volatility like, persistence, mean variance, leverage effect in Indian commodity derivative market.
2) The impact of the flow of information which is measured by exogenous variables i.e. open interest and volume on the dynamics of commodity price volatility. The Samuelson hypothesis (time to maturity) is also examined in Indian commodity market.
3) Examining the market efficiency of commodity futures market in performing the functions of Price discovery, volatility spillover, and hedging.
4) Understanding the users’ perspective about the derivative market activity, usage, price discovery and hedging effectiveness.

3.3.3 Objectives

Consequent upon the identification and classification of relevant issues, the detailed objectives have been formulated. These objectives are also amenable to the classification, quantitative or qualitative, as enumerated in Table 3.1

Table 3.1: Classification of objectives
<table>
<thead>
<tr>
<th>S. No.</th>
<th>Objectives</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>To analyze the stylized facts of volatility in Indian commodity futures such as persistence, mean reversion, volatility half-life and leverage effect.</td>
<td>Quantitative</td>
</tr>
</tbody>
</table>
| 2.     | To measure the impact of the flow of information on volatility dynamics of Indian commodity derivative market. This would be based on the following parameters namely;  
  - Impact of volume and open interest on volatility dynamics  
  - Impact of Samuelson hypothesis (time to maturity) on the volatility dynamics.                                                                 | Quantitative |
| 3.     | To measure the market efficiency and unbiasedness of commodity derivatives market and effectiveness in performing price discovery and volatility spillover functions.                                    | Quantitative |
| 4.     | To find out optimal hedge ratio by using static and dynamic models and evaluate the hedging effectiveness (reduction of variance) of various hedging models (static and dynamic).                                                     | Quantitative |
| 5.     | To gauge the perception of broker members about the efficiency of commodity derivative in India and to identify the purpose for which derivative securities are being used in the Indian context.         | Quantitative as well as qualitative |
3.3.4 Hypotheses

To address the objectives as mentioned above, some of the key hypotheses to be tested are:

**Hypothesis- I**

a) Null hypothesis (H₀): There is no significant impact of volume on volatility dynamics.
b) Null hypothesis (H₀): There is no significant impact of open interest on volatility dynamics
c) Null hypothesis (H₀): There is no significant impact of time to maturity on volatility dynamics

**Hypothesis- II**

a) Null hypothesis (H₀): There is no significant difference between the efficiency level of price discovery in commodity futures market and commodity spot market

**Hypothesis- III**

a) Null hypothesis (H₀): There is no significant difference between the volatility spillover from futures to spot or spot to futures.

3.3.5 Period of study

The time period has been taken differently for the agricultural and non-agricultural commodities. The study covers the total period of over 10 years from March 2004 to December 2014. To achieve the defined objectives of the present study, the data is required for the commodities futures and is obtainable from different commodity exchanges. As data is a critical input in the present study, it has to be ensured that it is collected from the source that has a transparent system of recording the prices, liquidity, reliability and long history. Two commodity exchanges qualify on these parameters namely, Multi-commodity Exchange (MCX) and National Commodity Derivative Exchange (NCDEX). MCX is ranked sixth among the global commodity bourses in terms of the number of futures contracts traded and also its futures contracts on Gold, Silver, Copper, Crude Oil are ranked among the top 20 global futures contracts in their respective segments [2]. NCDEX is recognized for high trading volume in agricultural commodities futures. Both the exchanges can be credited with bringing in specific systems in Indian commodity market and introduced the fully automated screen based trading system. Data for agricultural and non-agricultural commodities have been taken from National commodity and derivative exchange (NCDEX) and Multi commodity exchange (MCX) respectively. MCX and NCDEX have highest trading volume for non-agricultural commodities and agricultural commodities respectively and account for 97% of market share in terms of total turnover.
MCX and NCDEX both have started trading from the year 2003 and therefore the data is available for a period of over 10 years, which is appropriate to capture the different faces of volatility. Keeping this in view, the data has been taken from March 2004 to December 2014 for all the sample commodities that make it around 2500 trading days, over 10 years. Spot prices of the selected commodities are also taken from the respective exchanges. In the case of agricultural commodities, spot prices of the local market that is Disa for castor seed and Jodhpur for guar seed which is also the production center, are reported in Table 3.2. For the precious metal, the spot price of Ahmedabad market is used. The spot price of Mumbai is taken for industrial metal and for energy commodities i.e. crude oil and natural gas spot prices of Mumbai and Hazira is chosen for analysis. The data of all the sample commodities are not available from the March 2004 because they got listed at different point of time. Details of data period and spot market related to selected commodities are given in Table 3.2.

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Futures Market</th>
<th>Data Period</th>
<th>Spot Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castor Seed</td>
<td>NCDEX</td>
<td>March 2004 to Nov 2014</td>
<td>Disa</td>
</tr>
<tr>
<td>Guar Seed</td>
<td>NCDEX</td>
<td>March 2004 to Nov 2014</td>
<td>Jodhpur</td>
</tr>
<tr>
<td>Copper</td>
<td>MCX</td>
<td>March 2005 to Dec 2014</td>
<td>Mumbai</td>
</tr>
<tr>
<td>Nickel</td>
<td>MCX</td>
<td>March 2005 to Dec 2014</td>
<td>Mumbai</td>
</tr>
<tr>
<td>Gold</td>
<td>MCX</td>
<td>July 2007 to Dec 2014</td>
<td>Ahmadabad</td>
</tr>
<tr>
<td>Silver</td>
<td>MCX</td>
<td>June 2005 to Dec 2014</td>
<td>Ahmadabad</td>
</tr>
<tr>
<td>Crude Oil</td>
<td>MCX</td>
<td>March 2005 to Dec 2014</td>
<td>Mumbai</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>MCX</td>
<td>March 2005 to Nov 2014</td>
<td>Hazira</td>
</tr>
</tbody>
</table>

The selection of this time period ensures the following: a) uniformity of time period for sample commodities and b) recording of commodity futures prices in a transparent setup and liquid market. Further, this time period coincides with some of the major developments in the Indian commodity market which include, a) more active role of FMC after getting more statutory power from the government. b) Increase in the number of commodities futures c) Improvement in market
liquidity, d) acceptability and better understanding of the free regime of commodities in Indian market, e) beginning of an era where government emphasized upon the long- term vision for Indian commodity market, and d) more media attention led to nationwide awareness.

The data is obtained from the official website of MCX and NCDEX. For analysis both the market prices of futures contracts and spot market is considered. Data has been divided into two set i.e. the near month contracts and next to near month futures series. We have done this segregation on rolling basis i.e. data of next month contract is taken when the near month contract approaches maturity. Data on MCX Indices are taken from official website of MCX from June 2005 to December 2014.

3.3.6 Sample commodities and Indices

The major focus of this study is to examine the efficiency of commodity derivative market in performing the function of price discovery and hedging. Hence the universe has been defined as the individual commodity futures and commodity futures Indices. Currently, the derivative securities are traded through futures contracts in India for five Indices and 113 commodities. The futures contracts on MCX and NCDEX Indices were introduced in the year 2005 and 2010 respectively wherein futures contracts on commodities have been introduced in different phases. Initially, the futures contracts on the limited number of commodities were allowed to trade but nowadays more and commodities are bringing for futures trading.

For achieving the defined objectives of the present study, the sample consisting of 8 commodities, two belonging to the agricultural sector-castor seed and guar seed, two are industrial metal- copper and nickel, two are precious metal- gold and silver, and the rest two belonging to energy sector- crude oil and natural gas have been taken. The selected commodity futures are representative of their sector and their trading volume is relatively high from the year 2010 to 2014.

The MCX Indices namely MCX-COMDEX, MCX-ENERGY, MCX-METAL, and MCX-AGRI, are also used to study the defined objectives as they are representatives of the commodities derivatives. It is necessary to replicate the test used for individual commodities futures with these Indices to confirm that the results of individual commodities are not commodity specific rather can be generalized.

Data of futures contracts is divided into two parts: 1) near month futures data, and 2) next to near month futures data. For objectives 1 to 3, data of near month futures with respective spot data
has been taken for analysis. Market information has a greater impact on near month contracts than on contracts for farther-out delivery due to the smaller elasticity of supply and demand for shorter runs. For determining the hedging efficiency of Indian commodity derivative market which is the fourth objective of the study data of next to near month contract is also used.

3.3.7 Data description

The present study includes the analysis on the basis of both primary as well as secondary data. The primary data was collected through a survey conducted among the brokers, the key participants in the derivative market. The details of the survey methodology are covered in chapter 8.

3.3.7.1 The secondary data

The data for sample commodities and Indices have been collected from MCX and NCDEX, both for spot as well as futures contracts, because more than 97% of the trading in derivative securities takes place at these exchanges. Also, for the reasons mentioned in the section 3.3.5, all the inter-day data have been collected from official websites of MCX and NCDEX. The data collected include a) spot and futures prices for sample commodities (open, high, low, close and intraday), b) spot and futures prices for Indices namely, MCXCOMDEX, MCXAGRI, MCXMETAL, MCXENERGY c) Trading volume, open interest data for the derivative market.

3.3.8 Statistical tools

This study is based on extensive use of statistical tools. The spot and futures prices were obtained in .csv (Comma separated Values) and .dbf (Data Base File) format and they were initially processed using Microsoft Excel and Microsoft Access. Thereafter, statistical computations were carried out in SPSS 13 and EVIEWS 9. The graphical representation has been done using the graphical tools available in Microsoft Excel.

3.3.9 Methods of data analysis

There are a number of methods that have been employed for the investigation and analysis in the present study. They are classified into four sections keeping in view the objectives.

3.3.9.1 Methods used for studying the stylized facts of volatility

Over a period of time, a number of models have appeared in the literature that have attempted to capture the stylized features of volatility. On the basis of literature survey,
AutoRegressive Conditional Heteroscedasticity (ARCH) family models (as given in Table 3.3) have been chosen for inclusion in the current research study. The basic framework of ARCH family models works on the heteroscedasticity of the volatility of time series.

Table 3.3: Models used for studying volatility dynamics

<table>
<thead>
<tr>
<th>Stylized facts of volatility</th>
<th>Models used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistence, Mean reversion, Half-life and Leverage</td>
<td>ARMA (1,1) - GARCH(1,1) and ARMA(1,1) - EGARCH (1,1)</td>
</tr>
</tbody>
</table>

ARCH family models work better than the conventional models with the time series data as it does not base on the assumption of homoscedasticity of the variance of the error terms [330]. However, it is quite unlikely for the time series data to have variance of error constant over time and hence it exhibits heteroscedasticity. Heteroscedasticity means ‘changing variance’ and Autoregressive means ‘regressing on itself’. ARCH class of models have been developed keeping in view the heteroscedasticity and volatility clustering or volatility pooling. ARCH family models like GARCH and EGARCH models are specifically designed to model and forecast conditional variances. The variance of the dependent variable is modeled as a function of past values of the dependent variable and independent, or exogenous variables. Before explaining the ARCH family models, it would be pertinent to point out the idea of conditional as well as unconditional mean and variance. The conditional mean would be expressed as $E \left( y_{t+1} / \Omega_t \right)$, i.e. expected value of $y$ at the time t+1 given that all information up to and including time $t$ ($\Omega_t$) is available. In contrast to this, the unconditional expectation of $y$ is the expected value of $y$ without any reference to time. The key point to be noted here is that the term conditional implies explicit dependence on a past sequence of observations while unconditional is more concerned with the long term behavior of a time series and assumes no explicit knowledge of past. The distinction between conditional and unconditional variance is same as that of the conditional and unconditional mean. Similarly, the conditional variance of $u_t$ may be denoted as $\sigma_t^2$, which is written as:

$$
\sigma_t^2 = var\left( u_t / u_{t-1}, u_{t-2}, .... \right)
$$
\[ E \left[ (u_t - E(u_t))^2 / u_{t-1}, u_{t-2}, \ldots \right] \] (3.1)

Where it is assumed that \( E(u_t) = 0 \), therefore

\[ \sigma_t^2 = E \left[ (u_t^2 / u_{t-1}, u_{t-2}, \ldots) \right] \] (3.2)

This means that the conditional variance of normally distributed random variable \( u_t \) is equal to the conditional expected value of the square of \( u_t \). Thus in ARCH model, the autocorrelation in volatility is modeled by allowing the conditional variance of error term, \( \sigma_t^2 \) to depend upon the immediately preceding value of squared error.

\[ \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 \] (3.3)

This model is known as ARCH (1) as the conditional variance depends on only squared error. It could be extended to general case where the error variance depends on \( q \) lags of squared errors, also known as ARCH (q) model:

\[ \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2 \] (3.4)

The conditional variance in the literature is denoted by \( h_t \), so that the full model may be written as:

\[ h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2 \] (3.5)

Since \( h_t \) is a conditional variance, its value must always be positive. The values on the right hand side of the equation (3.5) contain all squared of lagged errors and hence they would always lead to positive outcome. However, in case one or more coefficients take the negative value, then the fitted value from the model for conditional variance may be negative and would become meaningless. Hence there is a non-negativity condition that needs to be attached with ARCH (q) specification, such that \( \alpha_i \geq 0 \), \( \forall \ i=1, 2, 3, \ldots q \) i.e. all the coefficient would be required to be non-negative. This is sufficient but not the necessary condition for the non-negativity of conditional variance.

**Limitations of ARCH (q) Models**

Despite being more robust than the traditional volatility estimates, ARCH models suffer from a number of limitations that include:
1. Deciding the number of lags necessary to obtain the conditional variance. Generally, likelihood ratio test is used in the absence of any other better approach.

2. Though Engle [97] uses ARCH (q) specification, the value of q might be very large to capture all the dependence in the conditional variance which makes the model unwieldy.

3. Non-negativity constraints on the coefficients might be violated.

3.3.9.1.1 Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model

To overcome the limitations mentioned above in ARCH (q) model, GARCH model as a natural extension of the ARCH (q) model is developed by Bollerslev [44]. According to Brooks [54], this model is more parsimonious than ARCH model and avoids overfitting. Further, this model is less likely to breach the non-negativity constraint. Pandey [245] has documented the significant explanatory power of GARCH models in the context of Indian stock market. The equation (3.6) is termed as GARCH (1, 1) model and it expresses the conditional volatility as a function of three parameters: a) Long-term average value ($\phi$); b) Information about volatility during the previous period ($u_{t-1}^2$); and c) Fitting variance from the model during the previous period ($\sigma_{t-1}^2$).

$$\sigma_t^2 = \phi + \beta \sigma_{t-1}^2 + \alpha_1 u_{t-1}^2$$  \hspace{1cm} (3.6)

In the financial context, the specifications of GARCH model is often interpreted as predicted value of current period’s variance ($\sigma_t^2$) by forming a weighted average of a long term average (the constant), the forecasted variance from last period (the GARCH term), and information about volatility observed in the previous period (the ARCH term) [100]. This model is also consistent with the volatility clustering often seen in financial returns data, where large changes in returns are likely to be followed by further large changes. GARCH model with heavy tailed innovation is efficient in forecasting the downside risk of returns whereas the estimates from GARCH models with normal innovations underestimate the potential down side risk [224].

The (1, 1) in GARCH (1, 1) refers to the presence of a first-order autoregressive GARCH term (the first term in parentheses) and a first-order moving average ARCH term (the second term in parentheses). An ordinary ARCH model is a special case of a GARCH specification in which there is no lagged forecast variance in the conditional variance equation i.e., a GARCH (0, 1).

This model is extended to a GARCH (p, q), where current conditional variance is parameterized to depend on upon q lags of the squared error and p lags of the conditional variance.
\[ \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \ldots + \beta_p \sigma_{t-p}^2 \]  
(3.7)

\[ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \]  
(3.8)

The unconditional variance can be calculated under GARCH specification as given below:

\[ Var(u_t) = \frac{\alpha_0}{1-(\alpha_1+\beta)} \]  
(3.9)

The unconditional variance of \( u_t \) is constant and the abovementioned specification holds good as long as \( \alpha_1 + \beta < 1 \). It is not defined for \( \alpha_1 + \beta \geq 1 \) and is termed as ‘non-stationarity in variance’ and when \( \alpha_1 + \beta = 1 \), it would be known as a ‘unit root in variance’ or ‘Integrated GARCH or IGARCH.

### 3.3.9.1.2 Estimation of ARCH/ GARCH models

ARCH/ GARCH models are not of the linear form and hence they require a different estimation procedure as compared to the one used for the estimation of parameters in the conditional mean equation where ordinary least square method is used that minimizes the residual sum of square. In the case of conditional variance, another technique, i.e. maximum likelihood, needs to be used that can be employed to find parameter values for both linear and non-linear models. This methodology works by finding the most likely values of the parameters given the actual data. The steps given in Exhibit 3.1 explain the estimation procedure for ARCH or GARCH model.

<table>
<thead>
<tr>
<th>Exhibit 3.1: Estimating an ARCH or GARCH model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Specify the appropriate equations for the mean and the variance- e.g. an ARMA (1,1)-GARCH(1,1) model</td>
</tr>
</tbody>
</table>

\[ Mean \ equation \quad R_t = \alpha_0 + \gamma_1 R_{t-1} + \gamma_2 u_{t-1} + u_t \]  
(3.10)
Variance Equation  \( \sigma_t^2 = \emptyset + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \)  \( \text{(3.11)} \)

2. Specify the log-likelihood function (LLF) to maximize under disturbance assumption of normality/student’s t/ Generalized Error Distribution (GED).

   \[
   l_t = -\frac{T}{2} \log(2\pi) - \frac{1}{\sigma_t^2} \sum_{t=1}^{T} (y_t - \mu - \phi y_{t-1})^2
   \]  (Normality assumption)  \( \text{(3.12)} \)

   \[
   l_t = -\frac{1}{2} \log \left( \frac{\pi(v-2)}{\Gamma(\frac{v}{2})} \frac{\left(\frac{v}{2}\right)^2}{\Gamma\left(\frac{v+1}{2}\right)^2} \right) - \frac{1}{2} \log \sigma_t^2 - \frac{(v+1)}{2} \log \left( \frac{(y_t - X_t' \theta)^2}{\sigma_t^2(v-2)} \right)
   \]  (student’s t distribution)  \( \text{(3.13)} \)

   \[
   l_t = -\frac{1}{2} \log \left( \frac{\Gamma(\frac{v}{2})^3}{\Gamma(\phi' \phi)^2} \right) - \frac{1}{2} \log \sigma_t^2 - \left( \frac{\Gamma(\frac{v}{2})(y_t - X_t' \theta)^2}{\sigma_t^2 \Gamma(\phi')^2} \right)
   \]  (GED assumption)  \( \text{(3.14)} \)

Where the tail parameter \( r > 0 \). The GED is a normal distribution if \( r = 2 \) and fat tailed if \( r > 2 \).

3. The computer will maximize the function (software EVIEWS, RATS, etc.) and generate the parameter values that maximize the LLF and will construct their standard errors.

This can be done by using the numerical procedure. All these methods work on the basic principle of ‘hit and trial’ to find out the values for parameters that maximize the log likelihood function. Even the software like EVIEWS, SAS, AUSS, TSP, Matlab, RATS and many others follow this principle and start with the initial guess of parameter values followed by the iterative process until program determines an optimum. The key success factor in implementation this methodology is the good initial guess of parameter values otherwise the log likelihood function will be maximized with respect of local optima. Given a distributional assumption, ARCH models are typically estimated by the methods of maximum likelihood. Maximizing the LLF involves jointly minimizing the value of error distribution that also implies minimizing the error variance. Brooks [54] has presented a simple summary of this procedure as given in the Exhibit 3.2.
Exhibit 3.2: Using Maximum Likelihood (ML) estimation in practice

1. Set up LLF.
2. Use regression to get initial estimates for the mean parameters.
3. Choose some initial guesses for the conditional variance parameters.
4. Specify a convergence criterion either by any specific criterion or by value.

There is a possibility that the true variance process may be different from the one specified using above mentioned procedure. In such a case, one may make use of the diagnostic tests. According to Engle [100], the simplest one is to construct the series of \{u_t\}, which is supposed to have constant mean and variance if the model is correctly specified. For example, Ljung Box test with 15 lagged autocorrelation can be used to examine the significance of autocorrelation in the return series of sample commodities.

Brook [54] has described this test to be extremely useful as a general test of linear dependence in time series. The LB- Q statistics is defined by Ljung and Box [211] as follows:

\[
Q = T(T + 2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{T-k} = x_m^2
\]  \hspace{1cm} (3.15)

Where T is sample size, m is minimum lag length such that m= 1,2,3,.....,k and \(\hat{\rho}_k\) is autocorrelation coefficient. Q-statistic is asymptotically distributed as a \(x_m^2\) under null hypothesis that all m autocorrelation coefficients are zero.

3.3.9.1.3 Exponential GARCH (EGARCH) model

As GARCH model incorporates squared value of error term, so fails to explain the leverage effect in the model. Nelson [238] introduces EGARCH (known as Exponential GARCH) model where conditional variance is constrained to be non-negative by assuming the logarithm of \(\sigma_t^2\) to be a function of the past \(u_t\)'s. \(\sigma_t^2\) depends upon the both size and sign of the lagged residual. The conditional mean and variance equations may be specified as follows:

\[
Mean \ equation \quad R_t = \alpha_0 + \gamma_1 R_{t-1} + \gamma_2 \varepsilon_{t-1} + \varepsilon_t
\]  \hspace{1cm} (3.16)

\[
Variance \ Equation \quad \ln(\sigma_t^2) = \omega + \varphi \left[ \frac{\varepsilon_{t-1}}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2)
\]  \hspace{1cm} (3.17)
As in the case of a typical GARCH model, \( \omega \) can simply be made a function of time to accommodate the effect of any non-trading periods or forecastable events [148]. This specification enables \( \sigma_t^2 \) to respond asymmetrically to rises and falls in the error term and has several advantages over GARCH model: a) \( \sigma_t^2 \) will always be positive even if the parameters are negative as the log specification is used in this model, thus avoiding the need to impose the non-negativity constraints; b) This model accounts for the asymmetries because \( \gamma \) will be negative in case the relationship between volatility and return is negative.

### 3.3.9.2 Methods used for finding the impact of volume, open interest and time to maturity on volatility dynamics

The present study investigates the significance of volume, open interest and time to maturity on the volatility dynamics (persistence and leverage) of commodity futures prices by using the augmented ARMA (1, 1)-EGARCH (1, 1) models with GED distribution of error terms. Equation (3.17) is extended to allow for the inclusion of exogenous or predetermined regressors in the conditional variance equation. The methodology used here has been taken from Lamoureux and Lastrapes [199] and Watanabe [318] which later replicated in context of Indian commodity market by Pati and Rajib [248].

To investigate the extent to which trading volume and open interest explain the persistence and leverage effect of volatility, we extend the EGARCH model [equation (3.17)] including volume and open interest as an exogenous variable in the conditional variance as given below:

\[
\text{ARMA (1, 1) – EGARCH (1, 1) conditional variance equation with current volume}
\]

\[
\ln(\sigma_t^2) = \omega + \varphi \left[ \frac{|\epsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\sqrt{\pi}} \right] + \gamma \frac{\epsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 V_t
\]

\[\text{(3.18)}\]

\[
\text{ARCH (1, 1) - EGARCH (1, 1) conditional variance equation with current open interest}
\]

\[
\ln(\sigma_t^2) = \omega + \varphi \left[ \frac{|\epsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\sqrt{\pi}} \right] + \gamma \frac{\epsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 \text{OI}_t
\]

\[\text{(3.19)}\]
Lamoureux and Lastrapes [199] and Najand and Yung [234] explained the simultaneity bias arises due to the inclusion of contemporaneous volume and open interest because they will not be an exogenous variable. Volume and open interest may be endogenous to this system in equation (3.18) and (3.19) respectively. To overcome this simultaneity problem, lagged volume and open interest is used in the conditional variance equations (3.20) and (3.21) respectively.

ARMA (1, 1) – EGARCH (1, 1) conditional variance equation with lagged volume

\[
\ln(\sigma_t^2) = \omega + \phi \left[ \frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\pi} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 V_{t-1} \tag{3.20}
\]

ARMA (1, 1) – EGARCH (1, 1) conditional variance equation with lagged open interest

\[
\ln(\sigma_t^2) = \omega + \phi \left[ \frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\pi} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 OI_{t-1} \tag{3.21}
\]

In the Equation (3.22) and (3.23) open interest is also added as an exogenous variable along with trading volume to understand whether open interest data can explain any newer dimension to this study.

ARMA (1, 1) - EGARCH (1, 1) conditional variance equation with current volume and open interest together.

\[
\ln(\sigma_t^2) = \omega + \phi \left[ \frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\pi} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 V_t + \delta_2 OI_t \tag{3.22}
\]

ARMA (1, 1) - EGARCH (1, 1) conditional variance equation with lagged volume and open interest together:

\[
\ln(\sigma_t^2) = \omega + \phi \left[ \frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\pi} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 V_{t-1} + \delta_2 OI_{t-1} \tag{3.23}
\]
Equation (3.24) explains the conditional variance equation with time to maturity as exogenous variable. Time-to-maturity variable is calculated by the number of calendars days’ remaining until the delivery of the futures contract.

\[ \text{ARMA (1,1)-EGARCH (1,1) conditional variance equation with time to maturity} \]

\[ \ln(\sigma_t^2) = \omega + \varphi \left( \frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \frac{\sqrt{2}}{\pi} \right) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2) + \delta_1 TTM \] (3.24)

This methodology cannot apply to commodity Indices because it requires the data for open interest and trading volume. Since the required data is not available for commodity Indices, the analysis is limited to sample eight commodities futures.

3.3.9.3 Methods used for testing the futures market efficiency, price discovery, and volatility spillover

Functioning of every financial market is based on the assumption of the efficient futures market. Further, an efficient futures market is considered to perform well the price discovery function and stabilize the volatility distribution. The following techniques are used to test market efficiency, price discovery and spillover functions in Indian commodity market.

3.3.9.3.1 Unit root test

Many economic and financial time series exhibit trending behavior or non-stationarity in the mean. An important initial econometric task with the financial data is to determine its stationarity. Many financial time series reflect a more complicated dynamic structure, a basic autoregressive unit root test fails to capture that. Said and Dickey [276] augmented the basic autoregressive unit root test to accommodate general ARMA(p, q) models with unknown orders and their test is referred as the Augmented Dickey-Fuller (ADF) test. The presence of stochastic trend (non-stationarity) is tested by using Augmented Dickey-Fuller test. Any time series is generally thought of as being generated by the stochastic or random process. A stochastic process is said to be stationary if its mean and variance are constant over time and value of covariance between two time periods depends only on the distance or lag between two time periods and not on the actual time at which covariance is computed.

The following are the implications of presence of unit root (hence, random walk behavior):

1) The permanent component in the fluctuation of the variable is highly volatile.
2) The random shock in the economy will result in a permanent increase in the level of series.

3) Since the price series in level form are I(1) time series, the standard hypothesis tests are inapplicable. The problem can be overcome by testing the cointegration relations.

Augmented Dickey-Fuller Regression is explained by the following equation.

\[ \Delta X_t = \rho_0 + \rho X_{t-1} + \sum_{i=1}^{n} \delta_i \Delta X_{t-i} + \varepsilon_t \]  \hspace{1cm} (3.25)

Where \( X_t \) = log price series, \( \rho_0 \) = a constant or drift, \( \Delta \) = first difference operator, \( \varepsilon_t \) = a pure white noise error term, \( i = 1 \) to \( n \) is number of lagged difference terms which is determined empirically to remove any autocorrelation in error term \( \varepsilon_t \). The null hypothesis is to test that \( \rho = 0 \).

If \( \rho = 0 \), then \( \alpha = 1 \), that is, the series has a unit root.

**Phillip-Perron (PP) unit root Test**

Phillip Perron test [250] builds on Dicky- Fuller test and address the issue of presence of high autocorrelation in the process generating data for \( X_t \). The presence of unit root is tested by regression (3.26) for the null hypothesis (\( H_0: \rho = 0 \)).

\[ \Delta X_t = \rho X_{t-1} + \varepsilon_t \]  \hspace{1cm} (3.26)

The main difference between ADF test and PP test is ADF addresses the issue of unit root by introducing lags of \( \Delta X_t \) as regressors in the test equation whereas the Phillips–Perron test makes a non-parametric correction to the t-test statistic. PP test is robust with respect to unspecified autocorrelation and heteroscedasticity in the disturbance process of the test equation.

In the present study, both ADF and PP test are used to examine the presence of Unit root in the time series.

**3.3.9.3.2 Johansen cointegration test**

The extensive review of literature shows that Johansen full information multivariate cointegrating procedure [163, 164] is widely used to perform the cointegration analysis. A vector time series \( Y_t (n \times 1) \) is said to be cointegrated if each of the \( n \) series is of the same degree of integration i.e. I(1). Johansen cointegration test is performed by using \( k^{th} \) order vector error correction model (VECM) which is represented as follows:

\[ \Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \tau_i \Delta Y_{t-1} + \vartheta + \varepsilon_t \]  \hspace{1cm} (3.27)

\[ \varepsilon_t / \Omega_{t-1} \sim (0, \sigma^2_t) \]
Where $Y_t$ is $(n \times 1)$ vector examine the cointegration, $\Delta Y_t = Y_t - Y_{t-1}$. $\partial$ represents the deterministic term or trend, and $\tau_t$ and $\Pi$ represent the coefficient matrix. The matrix $\Pi$ can be decomposed as $\Pi = \alpha \beta$, where matrix $\alpha$ is the speed of adjustment and matrix $\beta$, contains the cointegrating vectors ($\beta = 1-a-b$). The lag length of $k$ is selected on minimum value of Schwarz information criterion (SIC) such that $\epsilon_t$ is a multivariate normal white noise process with mean zero and finite covariance matrix. The existence of cointegration between endogenous variable is tested by examining the rank of coefficient matrix $\Pi$. If the rank of the matrix $\Pi$ is zero, no cointegration exists between the variables. If $\Pi$ is the full rank ($n$ variables) matrix then variables in vector $Y_t$ are stationary. If the rank lies between zero and $p$, cointegration exists between the variables under investigation. Two likelihood ratio tests namely Trace statistics and Max-Eigen value are used to test the long-run relationship [164]. Johansen and Juselius [164] and Cheung and Lai [68] argued that the trace test may lack power relative to the maximal eigenvalue test, and viewed that the trace test shows more robustness than the maximal eigenvalue test.

(1) Trace Statistics: The null hypothesis of at most $r$ cointegrating vectors against a general alternative hypothesis of more than $r$ cointegrating vectors is tested by trace statistics:

$$ (\lambda_{\text{trace}}) = -T \sum_{i=r+1}^{n} \ln(1 - \lambda_i) $$

(3.28)

Where $T$ is the number of observations and $\lambda_i$ is the largest squared canonical correlations. Null hypothesis for $\lambda_{\text{trace}}$ is the number of cointegration relationship is less than or equal to $r$ against the alternate hypothesis which assume the number of cointegration relationship is more than $r$.

(2) Maximum eigenvalue Statistics: The null hypothesis of $r$ cointegrating vector against the alternative of $r+1$ is tested by Maximum eigenvalue statistic:

$$ \lambda_{\text{max}}(r, r+1) = -T(\ln(1 - \lambda_{r+1}^1)) $$

(3.29)

3.3.9.3.3 Restrictions on cointegrating vectors

If the spot and futures prices are cointegrated, market efficiency and unbiasedness can be defined by a stationary linear combination of the two (3.30).

$$ \epsilon_t = S_t - a - bF_t $$

(3.30)
The present study tested the market efficiency and unbiasedness of the sample commodities by imposing a restriction on the parameters in equation (3.30). Market efficiency can be tested by imposing restriction of \( b \) as \( b = 1 \) and unbiasedness can be tested by imposing restrictions on both \( a \) and \( b \) as \( a = 0 \) and \( b = 1 \).

The present study has used the standard likelihood-ratio (LR) test proposed by Johansen and Juselius [164] to test the restriction on parameters. The standard likelihood ratio can be expressed as:

\[
L_r = T \sum_{i=1}^{r} \ln \left( \frac{1-\lambda_i^*}{1-\lambda_i} \right)
\]  

(3.31)

Where \( \lambda^* \) is the eigenvalues under the null hypothesis (restricted model) and \( \lambda \) is the eigenvalues under unrestricted models. The test statistics follows \( x^2 \) distribution with degree of freedom equal to number of restrictions imposed.

### 3.3.9.3.4 Vector Error Correction Model (VECM)

If the spot and futures prices are cointegrated, then causality must exist at least in one direction [139]. Karmarkar [173] argued that VECM captures the dynamic correlations and causalities between the cointegrated return series. VECM model identifies whether two variables move one after another or contemporaneously. When they move contemporaneously, one provides no information of other. If spot price \( (S_t) \) causes futures \( (F_t) \) prices, then changes in \( F_t \) should follow changes in \( S_t \). VECM model is better explained as VAR model in first difference augmented by the error-correction terms, \( \lambda_1 Z_{t-1} \) and \( \lambda_2 Z_{t-1} \). Equation (3.32) and (3.33) explain the specifications of the model.

\[
R_{st} = \mu_s + \lambda_1 Z_{t-1} + \sum_{i=1}^{k} \alpha_{s,i} R_{s,t-i} + \sum_{j=1}^{l} \beta_{s,j} R_{f,t-j} + \epsilon_{s,i} \quad (3.32)
\]

\[
R_{ft} = \mu_f + \lambda_2 Z_{t-1} + \sum_{i=1}^{k} \alpha_{f,i} R_{f,t-i} + \sum_{j=1}^{l} \beta_{f,j} R_{s,t-j} + \epsilon_{f,i} \quad (3.33)
\]

Where \( Z_{t-1} \) i.e. \( (S_{t-1} - F_{t-1}) \) is the Error Correction term (ECT). The magnitude of error correction terms \( \lambda_1 Z_{t-1} \) and \( \lambda_2 Z_{t-1} \) represent the speed of adjustment of returns towards long run equilibrium. The terms \( \alpha_s, \alpha_f, \beta_s, \beta_f \) are the short run parameters which measure short run integration.

### 3.3.9.3.5 Weak Exogeneity test
Even if the spot and futures price series show the long-term equilibrium relationship, they may still drift away from the relationship from time to time due to some transitory shocks. Weak exogeneity test can be used to test whether the equilibrium relationship can be restored quickly. If the price does not react to a shock in the long run relationship, it is said to be weakly exogenous.

If a cointegrating relationship exists then the coefficient matrix $\Pi$ in Equation (3.27) can be decomposed as $\Pi = \alpha \beta$, where $\beta$ is the cointegrating vector and $\alpha$ is a loading vector that measures the average speed of convergence towards the long-run equilibrium. The larger the value of $\alpha$, the faster the two price series converge to equilibrium [145]. If a price is weakly exogenous then the corresponding element of $\alpha$ will be zero. A likelihood-ratio statistic can be used to test the null hypothesis that $\alpha_i = 0$, for $i = 1, 2, \ldots, n$. The statistic has a chi-square distribution with one degree of freedom under the null hypothesis.

3.3.9.3.6 BEKK-GARCH (1, 1) model

Extensive literature review shows that GARCH type models have been used more frequently in studying the volatility spillover between markets [183, 16, 75]. The seminal work of Lin et al [209] used multivariate GARCH model to study the volatility spillover in the US and Japanese foreign exchange markets. The present study used GARCH (1, 1) - BEKK model [16] to study the volatility spillover between futures and spot prices of commodities traded in India. BEKK formulation reveals the existence of transmission of volatility from one market to another market [99]. One important feature of BEKK specification among others (diagonal specification, and constant correlation representation) is that it builds in sufficient generality allowing the conditional variances and covariance of the financial markets to influence each other. Besides, it has few parameters and ensures positive definiteness of the conditional covariance matrix which is a requirement needed to quadratic non-negative estimated conditional variance [99].

The covariance matrix of the unrestricted BEKK model is as follows:

$$H_t = C'C + a' \varepsilon_{t-1} + b'H_{t-1}b$$

(3.34)

Where $H_t$ represents the conditional covariance matrix and $C'C$ represents the upper triangular matrices. BEKK- GARCH (1, 1) is having a unique property of positive definite matrix for
conditional covariance. Covariance matrix \((H_t)\) will be positive definite as far as \(C'C\) is positive definite.

The present study used Berndt-Hall- Hall-Hausman [30] algorithm for estimating the parameters for BEKK – GARCH \((1, 1)\). This algorithm uses the first derivatives of the quasi-maximum likelihood (QML) with respect to the number of parameters that are contained in GARCH model.

The BEKK\((p, q)\) of variance of error term \(H_t\) is as follows:

\[
\begin{bmatrix}
h_{11,t} & h_{12,t} \\
h_{21,t} & h_{22,t}
\end{bmatrix} = \begin{bmatrix} c_{11,t} & c_{12,t} \\ c_{21,t} & c_{22,t} \end{bmatrix} \begin{bmatrix} c_{11,t} & c_{12,t} \\ c_{21,t} & c_{22,t} \end{bmatrix}^t + \begin{bmatrix} a_{11,t} & a_{12,t} \\ a_{21,t} & a_{22,t} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1} \varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11,t} & b_{12,t} \\ b_{21,t} & b_{22,t} \end{bmatrix} \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} \begin{bmatrix} b_{11,t} & b_{12,t} \\ b_{21,t} & b_{22,t} \end{bmatrix}
\]

\((3.35)\)

Where ARCH coefficient \((a_{ij})\) is the coefficient of square of residual \(e_t e_j\), and the GARCH coefficient \((b_{ij})\) represents the variance covariance term \(h_{ij}\).

Elements of \(a_{ij}\) represent the innovation from one market to another market. When \(i=j\) then element of \(a_{ij}\) represented as \(a_{11}\) and \(a_{22}\) which quantifies the effect of its own lagged value of squared residual term \((\varepsilon_{i,t-1}^2)\) on the present volatility of the spot market. Similarly, the coefficient \(a_{22}\) represent the effect of its own lagged value of squared residual term on the present volatility of the future market.

When \(i\neq j\), the element of \(a_{ij}\) represented as \(a_{21}\) and \(a_{12}\). Coefficient \(a_{12}\) measures the degree of innovation from spot to futures market and the coefficient \(a_{21}\) from futures to spot market. The element \(h_{11}\) represents the variance of change rate of spot market prices and \(h_{21}\) represents the covariance of the change rate of futures returns and spot returns. Volatility spillover from spot market to futures market is examined by testing whether the coefficients \(a_{12}\) and \(b_{12}\) are statistically significantly different from zero. Similarly to investigate the volatility spillovers effect from futures market to spot market, we should test \(a_{21}\) and \(b_{21}\) are significantly different from zero. If there is no volatility spillovers effect between futures and spot market, the non-diagonal elements of matrices \(a\) and \(b\) are not statistically significantly different from zero.
A program is written in Eviews software to estimate the parameters of the GARCH model under BEKK parameterization. Berndt, Hall, Hall and Hausman (BHHH) are used to estimate the variance-covariance matrix directly.

### 3.3.9.4 Methods used for testing the hedging effectiveness in commodity futures

There are a large number of models that can be used for measuring the hedging effectiveness. All these models can be classified into the following categories: a) Static hedging models, b) Dynamic hedging models. Dynamic hedging models clearly dominate as it incorporates the conditional information of time series data.

Further, the latest work in Indian context documented in Kumar and Pandey [188] shows the significant explanatory power of Vector Autoregressive Model (VAR). Therefore on the basis of literature survey, the following models (Table 3.4) have been shortlisted for inclusion in the current research study.

#### Table 3.4: Models used for measuring hedging effectiveness

<table>
<thead>
<tr>
<th>Static hedging methods</th>
<th>Dynamic hedging methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinary Least Square Regression, Vector Auto regression, Vector Error Correction Model</td>
<td>Vector Error Correction Model-Generalized AutoRegressive Conditional Heteroscedastic Model</td>
</tr>
</tbody>
</table>

#### (OLS) regression

To determine how many contract one should purchase to minimize the risk exposure, a conventional approach of OLS technique is used. The model of OLS regression is a linear regression of change in spot prices on the change in futures prices.

\[
R_s = \alpha_0 + \beta R_f + \varepsilon_t \quad (3.36)
\]

Where \(R_s\) and \(R_f\) are the return value of spot prices and the futures prices respectively, \(\beta\) is the slope coefficient of OLS regression which is the estimate of the optimal hedge ratio \(h\), also known as Minimum Variance hedge ratio [232].

\[
\beta = \frac{\sigma_{s,f}}{\sigma_f^2} \quad (3.37)
\]
Where $\sigma_{sf}$ is the covariance between spot and futures prices return and $\sigma_f^2$ is the variance of the futures price return.

The OLS technique is quite robust and simple to use. However, for the valid and efficient results, certain assumptions must be satisfied. If any of these assumptions is violated, the error term in the regression will be heteroscedastic. Under the first assumption, $\varepsilon_s$ and $\varepsilon_f$ are jointly normally distributed. Under the second assumption, both the spot and futures prices follow a pure random walk. The OLS uses unconditional sample moments instead of conditional sample moments [246]. OLS also assumes a constant relationship between spot and futures price, which has been proved to be wrong by many researchers [46].

### 3.3.9.4.2 Vector Autoregressive Method (VAR)

A simple regression method does not deal with the presence of autocorrelation among residuals. To overcome this shortcoming of OLS, bivariate vector autoregressive (VAR) model has been used. The optimal lag length is decided by minimum Akaike’s Information Criteria (AIC) and Schwarz’s information criteria (SIC).

The specifications of VAR model are presented as follow:

$$R_{st} = \alpha_s + \sum_{i=1}^{m} \beta_{si} R_{st-i} + \sum_{j=1}^{n} \gamma_{fj} R_{ft-j} + \varepsilon_{st} \quad (3.38)$$

$$R_f = \alpha_f + \sum_{i=1}^{m} \delta_{si} R_{st-1} + \sum_{j=1}^{n} \mu_{fj} R_{ft-j} + \varepsilon_{ft} \quad (3.39)$$

Where $\alpha_s$ and $\alpha_f$ are intercept term of spot and futures return equations respectively, $\beta_{si}, \gamma_{fj}, \delta_{si}, \mu_{fj}$ are parameters in VAR for $i=1,2,3,-----m,$ and $j=1,2,3,-----n,$ $R_{st-i}$ and $R_{ft-j}$ are lagged spot and futures returns, $\varepsilon_{st}$ and $\varepsilon_{ft}$ are the error terms of spot and futures return equations.

Residual series generated by equation (3.38) and (3.39) is used to calculate the hedge ratio by equation (3.40)

Minimum variance hedge ratio is calculated as follows:

$$(h) = \frac{\sigma_{sf}}{\sigma_f^2} \quad (3.40)$$

Where $\operatorname{Var} (\varepsilon_{st}) = \sigma_s^2$, $\operatorname{Var} (\varepsilon_{ft}) = \sigma_f^2$, and $\operatorname{Cov} (\varepsilon_{st}, \varepsilon_{ft}) = \sigma_{sf}$

### 3.3.9.4.3 Vector Error Correction Model (VECM)
Kroner and Sultan [187] proposed Vector Error Correction Model to estimate the hedge ratio. VECM model allows the existence of long-run equilibrium error correction in prices in the conditional mean equations. Equations 3.32 and 3.33 mentioned above are used to estimate the error term from the spot and futures series which are non-stationary and integrated of order one. The variance of error terms is used to estimate hedge ratio in equation (3.41).

The minimum variance hedge ratio is calculated as:

\[ h = \frac{\sigma_{s,f}}{\sigma_t} \]  

(3.41)

Where \( \text{Var}(\varepsilon_{st}) = \sigma_s \), \( \text{Var}(\varepsilon_{ft}) = \sigma_f \), \( \text{Cov}(\varepsilon_{st}, \varepsilon_{ft}) = \sigma_{sf} \)

3.3.9.4.4 Vector Error Correction Model – Multivariate Generalized AutoRegressive Conditional Heteroscedasticity (VECM - MGARCH)

The common feature of time series data is the presence of ARCH effect. The presence of ARCH effect is found in both the spot and futures return series which call for the estimation of dynamic hedge ratio. Seminal dynamic time series model VECH- GARCH model proposed by Bollerslev et al. [46] was difficult to follow because of complicated estimations and inference procedure of conditional variance and covariance matrix. To circumvent the estimations of a large number of parameters, Constant Correlation (CC)-MGARCH model is proposed by Bollerslev. We have considered the constant conditional correlation (CCC) in VECM- MGARCH for the calculation of time-varying hedge ratio.

VECM- MGARCH model is based on the assumption of constant conditional correlation between the observed residuals. VECM model (explained above) has used to estimates the mean equation. The error terms obtained from the VECM model is modeled for variance equation in CCC- GARCH model and covariance is calculated by using the equations (3.42), (3.43) and (3.44).

\[ h_{ss,t} = \omega_s + \sum_{i=1}^m \alpha_i \varepsilon_{s,t-i}^2 + \sum_{j=1}^n \beta_j h_{ss,t-j} \]  

(3.42)

\[ h_{ff,t} = \omega_f + \sum_{i=1}^m \alpha_i \varepsilon_{f,t-i}^2 + \sum_{j=1}^n \beta_j h_{ff,t-j} \]  

(3.43)

\[ h_{sf,t} = \rho(h_{ss,t} * h_{ff,t})^{1/2} \]  

(3.44)
Where $h_{ss,t}$ is the conditional spot variance, $h_{ff,t} = conditional$ $futures$ $variance$, $h_{sf,t}$ is covariance of futures and spot, $\rho = constant$ $conditional$ $correlation$, $\omega_s, \omega_f, \alpha_i, \beta_j$ are greater than zero, and $\alpha_i + \beta_j < 1$ for all the value of $i = 1, 2, 3 \ldots m$ and $j = 1, 2, 3, \ldots, n$.

Maximum likelihood estimation is used to estimate the parameters and optimal hedge ratio is calculated by using equation (3.45)

$$\text{Hedge ratio} = \frac{h_{ss,t}}{h_{ff,t}}$$

(3.45)

3.4 Concluding observations

The present chapter presents the methodology being used in the present study. The details are given with regard to the choice of sample, sample period, data collection, database, statistical tools and methods for investigation and analysis. The methodology is chosen on the basis of sound logic, the outcome of review of literature and its practicality. Despite the fact that some of the other methodologies might have proved to be more robust, the selection is done keeping in mind the availability of data from the Indian market. Based on the outline presented in this chapter, the empirical analysis is carried out in the forthcoming chapters.

CHAPTER-4

STYLIZED FACTS OF COMMODITY FUTURES VOLATILITY

4.1 Introduction

The immense importance of understanding the volatility and its patterns in the financial asset is evident from the fact that they have been used as critical input in portfolio selection, asset allocation, asset pricing, portfolio diversification and risk management. A number of researches in reputed journals all over the world covering the volatility explained it as a risk resulting from any
deviation in a given value from its mean \([107]\). People investing in the capital market desire to
gauge this risk (volatility) in order to maximize their benefit. Various models have been developed
from the time immemorial to cater to the risk minimization need of investors. The effectiveness of
any volatility model is determined by the accuracy with which it can forecast volatility. There are
some of the stylized facts about volatility that should be considered while designing an efficient
model (which has minimum error term). Stylized facts are a characterization of some empirical
statistical regularities which can be attained by analyzing the financial data \([61]\).

Examining volatility and its stylized facts help to understand the connection between
information transmission and volatility explicitly, since any change in the rate of information
arrival to the market will change the volatility \([77]\). Several features of financial time series and
stylized facts like volatility persistence and leverage have been well researched and accepted in
the developed market \([13, 142]\), but there is hardly any study which has examined the stylized
facts of volatility in Indian commodity market.

4.1.1 Stylized facts of volatility

Beginning from the seminal work of Mandelbrot \([218]\), the important features of the
volatility found in the literature are discussed here under:

1) Persistence or Volatility clustering: The seminal work of Mandelbrot \([218]\) reported that the
large changes in the price of assets are often followed by other large changes while the small
changes are often followed by small changes confirming persistence. This feature is referred
as volatility clustering. Persistence, as explained by Taylor \([301]\) is a phenomenon in which
market experience periods of high volatility and periods of low volatility, wherein, high
volatility leads to high dispersion of returns and vice-versa. The practical application of
 persistence is that the innovations today will influence the expectation of volatility many
periods in future. This characteristic was reaffirmed by numerous other studies \([19, 39, 45, 70, 104, 257, 281]\). Poterba and summers \([257]\) explained volatility as the weakly serially
correlated which implies that shocks to volatility do not persist.

 Poon and Granger \([254]\) provide a review of research issues concerning volatility, emphasizing
various aspects of volatility including persistence of volatility through time. ARCH imposes
an autoregressive structure on conditional variance, allowing volatility shocks to persist over
time. Persistence feature of volatility which documents the return of like magnitude to cluster
over time can be well explained by the non-normality and non-stability of empirical asset
The presence of ARCH effect is explained by the hypothesis of mixture of distribution (MDH) which assume the rate of daily arrival of information is stochastic mixing variable. The presence of MDH hypothesis in Indian commodity market is empirically tested in chapter 5.

Figure 4.1 shows the daily return of MCX- COMDEX Indices for the period ranging from January 1, 2005 to December 31, 2015. These returns are expressed in percentage terms and are continuously compounded. A visual inspection of this diagram makes it clear that there is volatility clustering.
an intermittent period of high volatility and any reasonable statistical test would infer that
returns are not independently and identically distributed over time.

2) **Volatility half-life:** A further quantitative measure of persistence in a volatility model is the
“half-life” of volatility. Half-life of volatility is defined as the time taken for the volatility to
move halfway back towards its unconditional mean following a deviation from it. The half-life is a better explanation of the persistence and provides a quantified approach in terms of
exact days required for volatility to come back to its average.

Zivot and Wang [329] proposed a formula to calculate the half-life as given below:

\[
L_{\text{half}} = \frac{\ln(1/2)}{\ln(\alpha_1 + \beta)}
\]

Where \(\ln\) is Natural logarithm, \(\alpha_1\) is ARCH term, and \(\beta\) is GARCH term

3) **Mean reversion:** The mean reversion is explained as a stationary series, where mean and
variance are constant and finite over time. Many researchers evidenced the existence of mean
reversion in developed and emerging countries by using a number of statistical techniques
namely, long-horizon regressions, variance ratio test, parametric contrarian investment
strategy etc.

Mean reversion in volatility advocates that there is always a normal level of volatility to which
the volatility will eventually return which implies that the current information has no long-
term impact on volatility. Mean reversion in time series data examines the nature of effect
shocks might have on volatility. The effect on volatility could be of permanent nature or
transitory nature. Transitory effect means that the variations in stock price do not have a
permanent effect on the volatility. Transitory effect of shocks on prices is confirmed if the
prices are stationary over the time. Hence, the mean reverting behavior of time series is
established. In other words, in the presence of mean reversion, prices will tend to return to its
mean value or follow the trend over the long run. Past trading strategies or price movement
could be used for forecasting future changes in the prices and to earn unusual returns in case
of transitory effect. However, if it is found that stock prices are non-stationary, then shocks will have a permanent effect, implying that stock prices will attain a new equilibrium and future returns cannot be predicted based on historical movements in stock prices [235].

4) Leverage effect: Black [38] introduced the concept of leverage effect in the context of financial time series data and later supported by many researchers [74, 101, 134, 238, 281]. Leverage effect confers a negative correlation between changes in stock price and changes in volatility. Under leverage effect, volatility tends to be more after a negative shock as compared to the positive shock of similar extent [74, 255]. deBondt and Thaler [87], in his study of US stock market, observed that past losing stocks over the previous 3-5 years significantly outperform past winning stocks over 3-5 years holding period.

In this chapter, the empirical analysis of stylized features of commodity futures is examined for both selected individual commodities and MCX commodity Indices. Section 4.2 covers the analysis for the presence of stylized facts of volatility in individual commodity futures. To further assess that the results are not limited to some selected individual commodities, the tests are replicated with MCX commodity Indices in section 4.3. Finally, the concluding observations are given at the end of the chapter.

4.2 Analysis of stylized facts of volatility

Natural logarithms of price series have been considered as the most consistent measure of variation of price changes in the past. Hence, for present analysis the price series has been converted in the return series as follows:

$$R_F = \ln \left( \frac{P_{FT}}{P_{FT-1}} \right)$$

(4.1)

Where $P_{FT}$ and $P_{FT-1}$ are current and previous day daily closing futures price of selected commodities respectively. $R_F$ is the return series.

Table 4.1 presents the descriptive statistics of individual commodities futures return. Results show that mean return of individual commodities varies from -0.0141 to 0.142. The highest return is of silver while the lowest mean return is of natural gas, which is negative. A perusal of standard deviation of all commodities suggests that the futures returns of castor seed are the most volatile. A measure of skewness and kurtosis suggest that all the commodities are positively skewed except guar seed, copper, and nickel which are negatively skewed. Return series of castor seed is more
skewed and highest peaked as compared to other which indicate that there is a great percentage of small deviations from the mean return and even a greater percentage of extremely large deviations from mean return. Most of the investors perceived such kind of behavior as increasing risk. Descriptive analysis of commodity futures return also exhibits the presence of several features of financial time series viz., fat-tailed distribution and skewness in Indian commodity futures. According to Nattenburg [237], financial asset returns exhibit non-normal skewness and kurtosis. Volatility skewness is a consequence of empirical violations of the normality assumption. These findings are also supported by the work of Corrado and Su [79], Clark [76] and Blattberg and Gonedes [40]. Adding to this, Ghysels et al. [130] contended that volatility clustering and fat tails of asset returns are intimately related.

| Table: 4.1 Descriptive statistics of near month futures return series |
|---------------------------|-----------------|----------------------|----------------|-----------------|----------------|-----------------|-----------------|
|                          | Castor Seed     | Guar Seed            | Copper          | Nickel          | Gold            | Silver          | Crude Oil        | Natural Gas     |
| Mean                     | 0.108           | 0.128                | 0.115           | 0.02            | 0.127           | 0.142           | 0.036            | -0.014          |
| Median                   | 0.023           | 0.062                | 0.06            | 0               | 0.064           | 0.105           | 0.078            | -0.071          |
| Std. Dev                 | 3.855           | 2.295                | 2.503           | 2.797           | 1.638           | 2.688           | 1.77             | 2.59            |
| Skewness                 | 34.602          | -7.26                | -2.198          | -0.992          | 2.525           | 0.061           | 0.086            | 0.514           |
| Jarque-Bera              | 1.97E+08        | 3660253              | 225288.8        | 1114.61         | 57510.41        | 13507.6         | 1592.28          | 1642.7          |

Standardized residual diagnostic

| LB-Q² (return)           | 0.036**         | 0.004*              | 0.05*           | 0.012*          | 0.042**         | -0.032*         | 0.072**         | 0.009*          |

** and * represent 1% and 5% level of significance respectively

Ljung-Box (LB-Q) statistics is also calculated to test the serial correlation in the returns. All value of LB- Q² statistics of returns are statistically significant indicating the presence of serial correlation. Significant Jarque-Bera statistics indicate the non-normal behavior of unconditional distributions of futures return [104, 296].

4.2.1 Stationarity of data
Regression analysis or modeling of non-stationary variables leads to potentially misleading inferences about the estimated parameters and the degree of association \([89, 90]\). Therefore, the order of integration or stationarity of price series must be determined. Near-month futures prices and their return series are tested for the stationarity by using the Augmented Dickey-Fuller unit root test (ADF) and Philip Perron test (PP).

Table 4.2 shows the result of ADF and PP test for the near month futures and the return series (log of first difference). Based on Schwarz information criteria (SIC), the optimal lag length chosen for ADF test is 2 for price series and 1 for return series. The results of both ADF and PP test indicate that the futures price series contains a single unit root, implying the fact that the futures price series is non-stationary. However, both tests statistics reject the hypothesis of a unit root at 1% level of significance in return series, implying the fact that the return series are stationary. Stationary of the return series also exhibits that these markets do not exhibit characteristics of random walk and as such are not efficient in the weak form.

Table: 4.2 Results of unit root test for near-month futures and spot prices

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Augmented Dickey-Fuller test</th>
<th>Phillips-Perron test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price series</td>
<td>Return</td>
</tr>
<tr>
<td><strong>Agricultural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castor Seed</td>
<td>near month futures</td>
<td>-0.7811</td>
</tr>
<tr>
<td></td>
<td>spot</td>
<td>-0.6912</td>
</tr>
<tr>
<td>Guar Seed</td>
<td>near month futures</td>
<td>8.1101</td>
</tr>
<tr>
<td></td>
<td>spot</td>
<td>7.8197</td>
</tr>
<tr>
<td><strong>Industrial Metal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>near month futures</td>
<td>-0.0563</td>
</tr>
<tr>
<td></td>
<td>spot</td>
<td>-0.3469</td>
</tr>
<tr>
<td>Nickel</td>
<td>near month futures</td>
<td>-1.3213</td>
</tr>
</tbody>
</table>
Precious Metal

<table>
<thead>
<tr>
<th></th>
<th>near month futures</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold</td>
<td>-0.6235</td>
<td>-35.953**</td>
<td>-0.5589</td>
<td>-32.603**</td>
</tr>
<tr>
<td></td>
<td>-0.5187</td>
<td>-35.685**</td>
<td>-0.497</td>
<td>-28.906**</td>
</tr>
<tr>
<td>Silver</td>
<td>-1.2538</td>
<td>-31.774**</td>
<td>-1.3001</td>
<td>-32.011**</td>
</tr>
<tr>
<td></td>
<td>-1.2393</td>
<td>-31.663**</td>
<td>-1.2867</td>
<td>-30.023**</td>
</tr>
</tbody>
</table>

Energy

<table>
<thead>
<tr>
<th></th>
<th>near month futures</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude Oil</td>
<td>-1.2079</td>
<td>-49.690**</td>
<td>-1.2080</td>
<td>-47.806**</td>
</tr>
<tr>
<td></td>
<td>-1.4185</td>
<td>-54.370**</td>
<td>-.0987</td>
<td>-45.962**</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>-2.0166</td>
<td>-45.856**</td>
<td>-1.4501</td>
<td>-42.905**</td>
</tr>
<tr>
<td></td>
<td>-1.9561</td>
<td>-49.842**</td>
<td>-.2198</td>
<td>-45.891**</td>
</tr>
</tbody>
</table>

** represents significance at 1 % level

### 4.2.2 Model description

The initial analysis of the commodity futures return series is showing the basic characteristics of financial time series facts like skewness, excess kurtosis and non-normal distribution and stationarity test is showing the stationarity of return series. Significant heteroscedasticity in return series as indicated by LB-Q statistics can be removed by using GARCH model. Literature supports the using GARCH specifications to characterize the presence of serial correlation [199, 248]. Thus the present study uses GARCH framework to examine the presence of stylized facts of volatility in Indian commodity futures. The first step in GARCH modeling is to accommodate the sufficient lagged value of autoregressive (AR) and moving average (MA) which removed any predictability associated with them [244, 101]. Akaike information criterion (AIC) and Bayesian information criterion (BIC) has been used to specify the mean equation with ARMA (1, 1) and variance equation as GARCH (1, 1). Generalized error distribution (GED) of the error term is used for the estimation of GARCH model because GED-GARCH model disposes of the leptokurtosis, fat-tailed behavior of financial time series [126].

Considering distributional assumption, maximum-likelihood method is used to estimate GARCH models. The estimates are generated by using ARMA (1, 1) – GARCH (1, 1) model equations (3.10) and (3.11) which are explained in detail in chapter 3 on Research Methodology. \( \alpha_1 \) and \( \beta \) are the ARCH and GARCH terms respectively (equation 3.11) which can be defined as
the coefficients measure the impact of recent news and old news on volatility respectively. Sum of $\alpha_1$ and $\beta$ measures the persistence of volatility. Large value of GARCH coefficient indicates the high volatility persistence while the large value of ARCH coefficient signifies less persistence. As GARCH model incorporates squared value of error term, so fails to explain the leverage effect in the model. Thus for the further examination of the asymmetricity of volatility in individual commodities EGARCH model proposed by Nelson [238] is used.

4.2.2.1 Persistence, half-life and mean reversion in commodity futures

Table 4.3 presents the estimated coefficients of ARMA (1, 1) - GARCH (1, 1) model without any variance regressors in the conditional variance equation. In mean equation the coefficient $\gamma_1$ and $\gamma_2$ are statistically significant at 1% for the copper, nickel, gold, and silver which signifies that conditional mean depends on its previous value and previous error term.

The results show that ARCH term, $\alpha_1$ (the coefficient on the lagged squared residual term) and GARCH term, $\beta$ (the coefficient of the lagged conditional variance) of the conditional variance equation are statistically significant for all the eight commodities. Persistence of the volatility is measured by the summation of $\alpha_1$ and $\beta$ which is quite high (greater than .98) for all the commodities. Persistence in volatility is the result of market reaction to the news [197, 11]. Thurman [306], Williams and Wright [319] and Karali and Thurman [171] concluded in their research that the large physical commodities inventory play a role of shock absorber which implied that if the commodity inventory size is large in the market, information is absorbed with greater speed as compared to small size commodity inventory. Sum value of ARCH and GARCH is less than 1 for all the commodity futures which indicate that return volatility will not move indefinitely upwards or downwards, thus, confirming mean reversion behavior in case of all commodity futures. The speed of change in the inventory of a commodity may be the cause of varying length of commodity futures volatility. Control of government in commodity futures and automatic forces of demand and supply are important factors which bring the return volatility back to equilibrium [131].

Results from Table 4.3 show that half-life time for sample commodities range from 22 to 107 days. Highest half-life is for energy commodities (107 and 70 days) and the least for industrial products (11 days and 35 days).
Table 4.3: Estimations of ARMA (1, 1) – GARCH (1, 1) Model

<table>
<thead>
<tr>
<th></th>
<th>Agricultural</th>
<th>Industrial Metal</th>
<th>Precious Metal</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Castor Seed</td>
<td>Guar Seed</td>
<td>Copper</td>
<td>Nickel</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.0297</td>
<td>0.05633</td>
<td>0.524**</td>
<td>0.04971*</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.2037</td>
<td>0.2818*</td>
<td>0.2917*</td>
<td>-0.6172**</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.2896</td>
<td>-0.3230**</td>
<td>-0.2741**</td>
<td>0.5954**</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.0368</td>
<td>0.1012*</td>
<td>0.06901</td>
<td>0.4522**</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0809</td>
<td>0.0908*</td>
<td>0.0181*</td>
<td>0.2078**</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9031</td>
<td>0.8911*</td>
<td>0.9623*</td>
<td>0.7609**</td>
</tr>
<tr>
<td>Persistence</td>
<td>0.984</td>
<td>0.982</td>
<td>0.981</td>
<td>0.969</td>
</tr>
<tr>
<td>Half Life</td>
<td>43 days</td>
<td>38 days</td>
<td>35 days</td>
<td>22 days</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>44 days</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47 days</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>107 days</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70 days</td>
</tr>
</tbody>
</table>

Standardized residual diagnostic

<table>
<thead>
<tr>
<th></th>
<th>ARCH-LM(24)</th>
<th>LB - Q(24)</th>
<th>LB - Q²(24)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0011</td>
<td>21.971</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>0.0398</td>
<td>22.4</td>
<td>0.7604</td>
</tr>
<tr>
<td></td>
<td>0.0621</td>
<td>19.908</td>
<td>1.608</td>
</tr>
<tr>
<td></td>
<td>3.237</td>
<td>5.42</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.0952</td>
<td>14.319</td>
<td>2.086</td>
</tr>
<tr>
<td></td>
<td>0.964</td>
<td>24.631</td>
<td>24.273</td>
</tr>
<tr>
<td></td>
<td>1.458*</td>
<td>25.382</td>
<td>36.45</td>
</tr>
<tr>
<td></td>
<td>1.4217*</td>
<td>15.22</td>
<td>35.638</td>
</tr>
</tbody>
</table>

** and * represent 1% and 5% level of significance respectively

LB-Q return, LB-Q² return has insignificant statistical value for all the commodities whereas ARCH-LM has significant statistical value for energy commodities futures. Insignificant statistical
value implies the acceptance of the null hypothesis of no serial correlation for all the commodities futures which indicates that GARCH model is perfect fit. Literature shows that informational efficiency of market impacts the persistence and asymmetricity of volatility. Hence, inclusion of these exogenous factors representing the flow of information in the model may affect the volatility persistence and asymmetricity. In chapter 5, a detailed analysis has been conducted to study the impact of flow of information on volatility dynamics.

4.2.2.2 Leverage effect in individual commodity futures

Assumptions of GARCH model restrict the model behavior towards bad or good news. In the present study, ARMA (1, 1) – EGARCH (1, 1) model is used to further study this asymmetric behavior of volatility. Table 4.4 reports the estimates of the coefficients of ARMA (1, 1) – EGARCH (1, 1) model. The asymmetric coefficient (γ) corresponding to $\frac{\varepsilon_t}{\sqrt{\sigma_t^2}}$ is negatively statistically significant for all the sample commodities which confirms the leverage impact and implies that variance tends to fall when return innovations are negative i.e. negative news have a greater impact on volatility then the positive news of same magnitude. Since β, coefficient of volatility persistence is positive and statistical significant for all the commodity futures which indicate that the current period volatility is highly impacted by the previous period’s volatility. According to Lien and Yang [208] prices in commodity market is more sensitive to negative noises than the positive one that leads to higher volatility in market. Thus, such activity in the market creates shaky position for the producers of the commodity. They can secure their profits by entering into commodity futures, options, short sales and warrants market.

The diagnostic test of Ljung – Box (Q) test and ARCH-LM test have been applied to check the robustness of the EGARCH (1, 1) model. The values of LB- Q and LB- Q² statistics accept the null hypothesis of no serial correlation for all commodity futures and indicate that ARMA (1, 1)- EGARCH (1, 1) model is well fitted and capture the leverage effect of innovations to volatility. ARCH –LM statistics confirm the presence no serial correlation among the residuals.
Table 4.4: Estimations of ARMA (1, 1) – EGARCH (1, 1) Model

<table>
<thead>
<tr>
<th></th>
<th>Agricultural</th>
<th>Industrial Metal</th>
<th>Precious Metal</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Castor Seed</td>
<td>Guar Seed</td>
<td>Copper</td>
<td>Nickel</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.0183</td>
<td>0.0266</td>
<td>-0.0021</td>
<td>-0.235**</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0178</td>
<td>0.999**</td>
<td>-0.555**</td>
<td>0.6067</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0418</td>
<td>-0.997**</td>
<td>0.5487*</td>
<td>-1.004**</td>
</tr>
<tr>
<td>$\omega$</td>
<td>-0.0247**</td>
<td>-0.065**</td>
<td>2.5401*</td>
<td>-0.053**</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0.074**</td>
<td>0.1174**</td>
<td>0.0201*</td>
<td>0.6067</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.0518**</td>
<td>-0.0624**</td>
<td>-0.0132**</td>
<td>-0.172*</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9687**</td>
<td>0.9830**</td>
<td>0.998**</td>
<td>0.8605</td>
</tr>
</tbody>
</table>

Standardized residual diagnostic

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>1.1294</th>
<th>1.357</th>
<th>1.57</th>
<th>1.587</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH-LM(24)</td>
<td>0.0011</td>
<td>0.047</td>
<td>0.6912</td>
<td>1.1294</td>
<td>1.357</td>
<td>1.57</td>
</tr>
<tr>
<td>LB-Q(24)</td>
<td>12.003</td>
<td>20.44</td>
<td>20.95</td>
<td>4.078</td>
<td>14.67</td>
<td>22.69</td>
</tr>
<tr>
<td>LB-Q^2(24)</td>
<td>0.418</td>
<td>0.321</td>
<td>0.885</td>
<td>0.0667</td>
<td>2.489</td>
<td>33.98</td>
</tr>
</tbody>
</table>

** and * represent 1% and 5% level of significance respectively

4.3 Stylized facts of volatility in MCX Indices

GARCH test has been applied on the MCX Indices to further test presence of stylized facts are not limited to sample individual commodities only. Natural logarithms of MCX futures and spot price series have been calculated by using the equation (4.1). Table 4.5 reports the descriptive
analysis for futures and spot price series for all the four MCX Indices namely, MCX COMDEX, MCX AGRI, MCX ENERGY, MCX METAL. All the futures series are represented by COMDEXF, AGRIF, ENERGYF and METALF whereas all the spot series are represented by COMDEXS, AGRIS, ENERGYS and METALS. Both, mean and standard deviation of futures and spot price series of every Index is approximately same. However, the kurtosis and skewness (absolute value) of the futures price series are greater than that of spot price series which indicates that the futures market appears more volatile than the spot market. These series are not normally distributed as value of kurtosis is less than 3 and value of skewness is not equal to 0 in any of them. Jaruque-Bera statistics provides evidence for the non-normality of the price series. There is better risk return trade-off displayed in the futures market as low standard deviation is observed in case of futures as compared to spot market. Investors have the opportunity to earn same amount of return in the futures market with less risk

Table: 4.5 Descriptive statistics of MCX Indices

<table>
<thead>
<tr>
<th></th>
<th>COMDEXF</th>
<th>COMDEXS</th>
<th>AGRIF</th>
<th>AGRIS</th>
<th>METALF</th>
<th>METALS</th>
<th>ENERGYF</th>
<th>ENERGYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.9153</td>
<td>7.9106</td>
<td>7.613</td>
<td>7.6770</td>
<td>8.0825</td>
<td>8.0747</td>
<td>7.9464</td>
<td>7.9296</td>
</tr>
<tr>
<td>Median</td>
<td>7.8971</td>
<td>7.8853</td>
<td>7.641</td>
<td>7.652</td>
<td>8.034</td>
<td>8.0315</td>
<td>7.9140</td>
<td>7.9034</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.2693</td>
<td>0.2828</td>
<td>0.266</td>
<td>0.335</td>
<td>0.347</td>
<td>0.3485</td>
<td>0.2339</td>
<td>0.2437</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.0725</td>
<td>-0.0309</td>
<td>0.171</td>
<td>0.152</td>
<td>-0.208</td>
<td>-0.2010</td>
<td>0.0955</td>
<td>-0.0261</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.7663</td>
<td>1.7628</td>
<td>2.022</td>
<td>1.600</td>
<td>2.098</td>
<td>2.0665</td>
<td>2.5295</td>
<td>2.7856</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>169.52</td>
<td>168.57</td>
<td>117.96</td>
<td>225.43</td>
<td>108.44</td>
<td>113.49</td>
<td>28.332</td>
<td>5.3512</td>
</tr>
</tbody>
</table>

** and * represent 1% and 5% level of significance respectively

The descriptive statistics of Indices futures are similar with the individual commodities futures. The findings of Figlewski [107] evidenced that equities and many other securities show
fatter tails since the log-normal diffusion model is inconsistent are reconfirmed by the descriptive analysis of the MCX Indices.

Augmented Dickey-Fuller test has been employed to test the stationarity of spot and future price series of MCX Indices futures. The results of Augmented Dickey-Fuller test are further examined by Phillips–Perron test. Results of both the test are displayed in Table 4.6. The results indicate that spot and the futures price series of commodity market are non-stationary at the level form and become stationary at the first order level. The findings of Augmented Dickey-Fuller test are supported by the Philips- Perron test.

Table 4.6: Results of unit root test for MCX Indices

<table>
<thead>
<tr>
<th>MCX Indices</th>
<th>Variables</th>
<th>Augmented Dickey-Fuller</th>
<th>Phillips-Perron test</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCXCOMDEX</td>
<td>COMDEXF</td>
<td>-0.407387</td>
<td>-63.429**</td>
</tr>
<tr>
<td></td>
<td>COMDEXS</td>
<td>-0.456252</td>
<td>-54.656**</td>
</tr>
<tr>
<td>MCXAGRI</td>
<td>AGRIF</td>
<td>-0.645416</td>
<td>-49.752**</td>
</tr>
<tr>
<td></td>
<td>AGRIS</td>
<td>0.019553</td>
<td>-48.211**</td>
</tr>
<tr>
<td>MCXMETAL</td>
<td>METALF</td>
<td>0.201905</td>
<td>-57.377**</td>
</tr>
<tr>
<td></td>
<td>METALS</td>
<td>0.052952</td>
<td>-53.150**</td>
</tr>
<tr>
<td>MCXENERGY</td>
<td>ENERGYF</td>
<td>-1.838784</td>
<td>-43.439**</td>
</tr>
<tr>
<td></td>
<td>ENERGYS</td>
<td>-1.968213</td>
<td>-44.124**</td>
</tr>
</tbody>
</table>

Note: ** indicates significance at 1% level. Optimal lag length is determined by SIC (Schwarz information criterion) for Augmented Dickey-Fuller test.

The results of the GARCH (1, 1) model with error term following GED is shown in Table 4.7. Results of Table 4.7 help in drawing inferences about the basic characteristics of Indian commodity futures market. The results from the GARCH model indicate that the METALF has
highest persistence value (0.9995) as compared to returns of other futures series. AGRIF has registered lowest persistence with persistence value of 0.752454. COMDEXF and ENERGYF have shown persistence value of 0.973126 and 0.997777 respectively. Lowest persistence value in AGRIF shows that since AGRI is a government regulated market, any innovation entering the market has short-lived impact due to government interventions. ENERGYF and METALF are found to be most volatile markets as these markets are mostly dependent on the foreign markets. If foreign innovation enters the market it has long lasting impact. The results are in line with Daal et al. [82] which states that volatility tends to persist in developed countries but it seems to be even more persistent in emerging countries. All the four return series of MCX Indices show sign of the Mean Reversion. The values for each index is less than one which indicates that return volatility will not move indefinitely upwards or downward. Eventually, the return volatility will come down to a mean level.

Volatility persistence can be better expressed by way of volatility half-life. Results show that METALF takes the longest time to return back to its unconditional mean. The impact of volatility lasts for 1571 days (approx.) on the METALF returns while the shortest impact in term of days, i.e. 2 days, has been registered in case of AGRIF returns. For COMDEXF, the impact of volatility lasts for 25 days and for ENERGYF, it lasts for 309 days. The reason for such a long memory in METAL futures could be the commodities included in this index. All the industrial used metals are constituents of METAL Index and get affected by the indigenous news and foreign news both. The impact of these news remains for a longer time. As Indian economy is agriculture driven and regulated by the government so the impact of previous volatility does not last for more than 2 days for AGRI futures index. As COMDEX is an average of all the other MCX Indices, previous periods’ volatility impact is averaged out which is evident from the volatility half-life value of 25 days.

Table 4.7: Estimations of GARCH (1, 1) Model for MCX Indices

<table>
<thead>
<tr>
<th></th>
<th>COMDEXF</th>
<th>AGRIF</th>
<th>METALF</th>
<th>ENERGYF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>0.056784</td>
<td>0.205986</td>
<td>0.032896</td>
<td>0.010266</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.121397</td>
<td>0.151634</td>
<td>0.09513</td>
<td>0.039989</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.851729</td>
<td>0.60082</td>
<td>0.904429</td>
<td>0.957776</td>
</tr>
</tbody>
</table>
Volatility persistence | 0.973126 | 0.752454 | 0.999559 | 0.99777
Mean reversion | 0.973126<1 | 0.752454<1 | 0.999559<1 | 0.99777<1
Volatility half-life | 25 days | 2 days | 1571 days | 309 days

Standardized residual diagnostic

<table>
<thead>
<tr>
<th></th>
<th>LB-Q(24)</th>
<th>LB-Q²(24)</th>
<th>ARCH-LM (24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatility persistence</td>
<td>19.018</td>
<td>0.6563</td>
<td>0.026471</td>
</tr>
<tr>
<td>Mean reversion</td>
<td>20.400</td>
<td>1.4385</td>
<td>0.058622</td>
</tr>
<tr>
<td>Volatility half-life</td>
<td>23.181</td>
<td>3.9106</td>
<td>0.16001</td>
</tr>
</tbody>
</table>

** and * represent 1% and 5% level of significance respectively

Ljung-box test and ARCH-LM test were conducted to test whether the model is a better fit with the data. Statistics of Ljung-box test accepts the null hypothesis of no serial correlation as the critical Q value from the Chi-square table exceeds the computed value of Q-statistics for all commodity Indices except for ENERGYF. This diagnostic test confirms that the GARCH (1, 1) is the best-fitted model for all the commodity Indices except ENERGYF.

**Leverage Effect:** It is realized by the researchers at large that negative innovations are likely to cause more volatility as compared to positive innovations of the same magnitude [173, 137]. EGARCH propounded by Nelson [238], gives conditional volatility model to address the problem of leverage effect.

Table 4.8: Estimations of EGARCH (1, 1) model for MCX Indices

<table>
<thead>
<tr>
<th></th>
<th>COMDEXF</th>
<th>AGRIF</th>
<th>METALF</th>
<th>ENERGYF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatility persistence</td>
<td>-0.167652*</td>
<td>0.486339*</td>
<td>0.098452*</td>
<td>-0.068935*</td>
</tr>
<tr>
<td>Mean reversion</td>
<td>0.274737*</td>
<td>-0.091439*</td>
<td>0.164669*</td>
<td>0.109565*</td>
</tr>
<tr>
<td>Volatility half-life</td>
<td>-0.076530*</td>
<td>0.058781*</td>
<td>0.030190*</td>
<td>-0.015614*</td>
</tr>
<tr>
<td>Leverage Effect</td>
<td>0.951566*</td>
<td>-0.369271*</td>
<td>0.952155*</td>
<td>0.985004*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LB-Q(24)</th>
<th>LB-Q²(24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatility persistence</td>
<td>22.799</td>
<td>3.1278</td>
</tr>
<tr>
<td>Mean reversion</td>
<td>24.207</td>
<td>23.105</td>
</tr>
<tr>
<td>Volatility half-life</td>
<td>25.747</td>
<td>7.2231</td>
</tr>
</tbody>
</table>

** and * represent 1% and 5% level of significance respectively
The results of EGARCH statistics to determine the leverage Effect have been depicted in Table 4.8. An analysis of $\gamma$ values for all four Indices under study show the negative and significant values for COMDEXF (-0.010447) and ENERGYF (-0.013382). This provides evidence for negative correlation between volatility and changes in the prices whereas the $\gamma$ value for AGRIF and METALF signals a positive relationship between volatility and change in the prices. Negative correlation implies that the negative innovations have more impact on the volatility as compared to positive news of the same magnitude. To check the robustness of the EGARCH (1,1) model, Ljung-Box(Q) test and ARCH-LM test is conducted. The value of Q and Q$^2$ statistics accepts the null hypothesis of no serial correlation for all commodity futures except Energy Futures. Thus the results show that the bivariate EGARCH model is well-fitted model with all commodity Indices except Energy futures. A study conducted by the Debasish and Dey [86] in the Indian Black Pepper commodity futures market found strong support for the presence of ARCH and GARCH effects.

### 4.4 Concluding observations

There are a number of stylized facts that have been uncovered in the present chapter about the Indian commodity derivative market; namely volatility persistence, volatility half-life, mean reversion, leverage effect and the presence of fat-tailed distribution of risky asset returns. The findings suggest that derivative market in India is also having a high persistence which results in high half-life. Strict delivery rules and availability of high inventory may be one of the reasons for such a high persistence in these commodity futures. The presence of mean reversion also confirms that volatility ultimately comes back to its mean level because of presence of arbitrage. Market volatilities form clusters and volatilities are not very persistent in India, contrary to experiences in many countries [22].

EGARCH model confirms the presence of leverage effect in agricultural and energy commodities. Existence of high persistence and leverage may be due of the presence of speculation and illegal traders or hoarders. It is suggested to strengthen the institutional mechanism in India that would bring the commodities under government's scanner and gather the evidence of
speculation and spot out the illegal traders or hoarders rather than imposing a ban on specific commodities futures trading. The findings are by and large similar for individual commodities as well as the overall market being represented by MCX Indices. Through the present research an effort has been done to complement and update previous results as well as to extend some of them into either new or different directions.