Chapter 3 Research Methodology

3. Introduction to Research Methodology

This chapter will discuss various tests and techniques used to study the hypothesis that are formed for understanding of the research problems. The proposed study is an applied research in which research techniques and procedure are applied to the collection of data about the problem so that information gathered can be used meaningfully to understand the futures market. The time series analyses are utilized to analyze the long term and causal relationship between the variables. The study will be taking daily data for the span of almost 11 years from 2005 to 2014 depending on the availability of information for different commodities. The study uses various statistical softwares and programmes like EViews 8, Gauss, Microsoft excel and nonlinear causality code.

3.1 Definition of Problem

Agriculture commodity prices have seen extreme volatility in past and continuous effort is made to keep the price volatility in check and to explore the possible reasons leading to the wide fluctuations in their prices. The present study aims to analyse the volatility in agriculture commodity prices in presence of structural breaks (if any). Further to study the association between the volatility of agriculture commodity prices and volatility in macro-economic variables.

Agriculture price volatility apart from being attributed to fundamental demand-supply variation is also influenced by the varieties of factors like volatility in oil prices which serves as input cost, interest rates in the economy, foreign currency exchange rates, inflation, etc. (Rezitis & Sassi (2013), Abbott et al. (2008), Trostle (2008), Wright (2009)). Major structural change in agriculture commodities market in recent past has been brought by technological upgradation offering online trading. This has resulted into better dissemination of information to the interested parties (Irwin, 2012). The primary objective of the study is to understand the impact of the movement in macro-economic variables on agriculture prices. This will benefit the numerous groups that hold interest in the commodity market either for hedging purpose or with an aim to trade and make profit from the movements thereof.
3.2 Research Questions

Formulation of research questions is a vital step. The present study aims to analyse the volatility in the agriculture commodity prices and the impact of volatility in macro-economic factors on the agriculture prices. The lead-lag relationship between spot and future market of the commodity helps to unearth which market captures the information first and transmits to the other market. Also helps to evaluate which of the two markets i.e. spot or future reacts faster to the shocks introduced by movement in macro-economic variables. The testing of lead-lag relation between variables has been the most coveted technique in study of market efficiency. It helps investors as they plan to take advantage of arbitrage opportunities in the market. The research would also prove useful to the academia, investors, hedgers and policy makers to understand what causes the structural shift in the prices of agriculture commodities and lead-lag between them. This will help them to understand market efficiency and understand the usefulness of the variable from the point of view of making future forecast for the movement in prices.

The major research questions dealt with by the present study are

i. Can structural breaks be identified in the price series of agriculture commodities viz. Soy Bean, Chana and Turmeric? If yes, what are the possible reasons for the occurrence of these breaks?

ii. Is there presence of lead-lag relationship between spot and futures price of agriculture commodities viz. Soy Bean, Chana and Turmeric? Has this relationship undergone any change with occurrence of structural breaks in the prices?

iii. Is there any lead-lag relationship between spot price of agriculture commodity and Sensex? Has this relationship undergone any change with occurrence of structural breaks in the prices?

iv. Is there any lead-lag relationship between spot price of agriculture commodity and Forex (USD-INR Rate)? Has this relationship undergone any change with occurrence of structural breaks in the prices?
v. Is there any lead-lag relationship between spot price of agriculture commodity and Crude? Has this relationship undergone any change with occurrence of structural breaks in the prices?

vi. Is there any lead-lag relationship between Future price of agriculture commodity and Sensex? Has this relationship undergone any change with occurrence of structural breaks in the prices?

vii. Is there any lead-lag relationship between Future price of agriculture commodity and Forex (USD-INR Rate)? Has this relationship undergone any change with occurrence of structural breaks in the prices?

viii. Is there any lead-lag relationship between Future price of agriculture commodity and Crude? Has the relationship undergone any change with occurrence of structural breaks in the prices?

3.3 Data collection and Variables

The data for spot and future price of agriculture commodities is collected from National Commodity & Derivatives Exchange Limited (NCDEX). NCDEX is an online commodity exchange headquartered in Mumbai, India. It was established in the year 2003 and offers large varieties of agriculture derivatives contracts. In India, NCDEX is one of the most important commodity exchanges for agricultural commodities, hence selected for empirically testing of the causal association between Spot prices and Future prices. The chart below (refer to figure-3.1) gives the percentage of share of commodities trade on NCDEX.
The study proposes to explore the lead lag associationship between the spot and future prices of following agriculture commodities due to their importance in the consumption pattern in India as well as the interest of investor community that lends them volume and liquidity in the market. The spot and near month future contract price data is availed from NCDEX.

i. **Chana** – Chana/Chick Pea is a very sort after legume for its protein and nutrients and is an important part of the pulses basket. It is also referred to as Bengal gram and primarily comes in 2 varieties one being “Desi Chana” and other “the Kabuli Chana”. India is the world’s largest producer of Chana with production of approximately 8.8 million tonnes in 2012-2013 (Chana Market outlook). India contributes over 75% of total world production of the commodity, followed by Pakistan, Turkey and Iran. Despite of hefty production, India has been net importer of the Chana seeds due to high domestic consumption. Being a
part of essential commodities, Government of India had banned the export of Pulses from India to meet domestic demand. Also to support the growth in production, Government fixes minimum support price every year. It is a rabi crop with sowing season being October-December and harvest season being February to April. The chart given in Figure 3-2 below depicts the quantity of production, export and import of Chana from 2005-2013 in lakhs of metric tons.

**Figure 3-2 Production, Export and Import of Chana from 2005-2013**

Source: Agricorp.nic.in, (Indian Pulses and Grains Association)

The minimum support price (MSP) in case of Chana has grown at a compounded annual growth rate of approximately 9% over 2005-2015. The chart given in figure 3-3 depicts the MSP (in Rupees per quintal of Chana) over 2005-2015.
Source: Agricorp.nic.in, (Indian Pulses and Grains Association)

### Table 3-1 Contract Specifications for Chana on NCDEX

<table>
<thead>
<tr>
<th>Contract Specifications</th>
<th>Chana Future Contract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticker</td>
<td>CHARJDDEL</td>
</tr>
<tr>
<td>Trading hours</td>
<td>Monday to Friday 10 am to 5 pm</td>
</tr>
<tr>
<td>Contract months</td>
<td>As decided by the exchange</td>
</tr>
<tr>
<td>Contract Size</td>
<td>10 MT</td>
</tr>
<tr>
<td>Tick Size</td>
<td>Re. 1</td>
</tr>
<tr>
<td>Daily Price Limits</td>
<td>DPL (+/-) 4%</td>
</tr>
<tr>
<td>Expiry/Due date of Contract</td>
<td>20&lt;sup&gt;th&lt;/sup&gt; day of the delivery month</td>
</tr>
<tr>
<td>Delivery Center</td>
<td>Delhi</td>
</tr>
<tr>
<td>Delivery logic</td>
<td>Compulsory delivery of outstanding position on expiry of contract</td>
</tr>
</tbody>
</table>
ii. **Soy Bean** - Soy Bean is also referred to as Golden Bean and is classified as Oilseed on NCDEX. It is one of the important sources of protein diet and is also majorly used to produce Soya Oil (vegetable oil) and soya meal. India ranks 5th in the world after USA, China, Brazil and Argentina in terms of production of the Soy Bean seeds. It is a kharif crop sowed in the month of June –July and is harvested in the month of Oct-Dec (Source: NCDEX, ICEX India). It has multiple uses from human dietary product to production of Industrial products like paints, soaps, plastic, etc. and also serves as cattle feed, poultry feed and feed for aqua animals (ICEX INDIA). India is increasing its Soy Bean produce and the production of over 100 lakh tonnes of Soy Bean was made in 2012-2013. (InditradeIndia.com, agricoop.nic.in). India imports almost 55% of edible oil requirement and hence Soy Bean has good trading business. India exports Soy Bean meal and with increase in demand for Soy bean both for the meal and biofuel production, soy bean prices get subjected to forex risk and global supply-demand fluctuations. Majority of Soy Bean production is done in the states of Madhya Pradesh, Maharashtra and Rajasthan.

The minimum support price (MSP) in case of Soy Bean has grown at a compounded annual growth rate of approximately 8.4% over 1992-2013. The chart given in figure 3-5 depicts the MSP (in Rupees per quintal (100 Kgs) of Soy Bean) over 1992-2013.

**Figure 3-4 Production of Soy Bean oilseeds from 2005-2014**

![Soybean Production Chart](image-url)
Minimum support price depicted in the graph above is for black variety of Soy Bean in India.

### Table 3-2 Contract Specification for Soy Bean on NCDEX

<table>
<thead>
<tr>
<th>Contract Specification</th>
<th>Soy Bean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticker</td>
<td>SYBEANIDR</td>
</tr>
<tr>
<td>Trading hours</td>
<td>Monday to Friday 10 am to 5 pm</td>
</tr>
<tr>
<td>Contract months</td>
<td>As decided by the exchange</td>
</tr>
<tr>
<td>Contract Size</td>
<td>10 MT</td>
</tr>
<tr>
<td>Tick Size</td>
<td>Re. 1</td>
</tr>
<tr>
<td>Daily Price Limits</td>
<td>DPL (+/-) 4%</td>
</tr>
<tr>
<td>Expiry/Due date of Contract</td>
<td>20th day of the delivery month</td>
</tr>
<tr>
<td>Delivery Center</td>
<td>Indore</td>
</tr>
<tr>
<td>Delivery logic</td>
<td>Seller’s option exercised during tender period</td>
</tr>
</tbody>
</table>
iii. **Turmeric** – India is the largest producer and exporter of turmeric in the world. It is also termed as Curcuma longa or Haldi and belongs to the Zingiberaceae category (NCDEX). Turmeric falls under the “Spices” basket and it is famous for ample amount of active ingredient called curcumin, which is used as anti-cancer medicine and has antibiotic properties. India produces nearly 80-90% of world’s total production and is the largest exporter and consumer of turmeric. It is a kharif crop and takes nearly 7-8 months to be ready before harvest. It is sowed in the months of May-August and is harvested in December-March. This golden spice has diversified uses. Many consume it in its raw form or use as a dry spice for coloring and flavoring the food. Its antibiotic properties have led to huge demand from pharmaceutical sector and Ayurveda. In addition to this it has many industrial uses as colorant, formulations of dyes for textiles, in production cosmetics, oils and ointments. The Oleoresin extracted from turmeric is used in many formulations like brine pickles, gelatins etc. Major turmeric producing states in India are Andhra Pradesh, Karnataka, Orissa, Tamil Nadu, West Bengal, Maharashtra and Kerala (Spices Board of India).

Figure 3-6 shows the data for production, area under cultivation and yield for turmeric over 1997-2014. The yield for turmeric has been very volatile largely dependent on southwest monsoon.

**Figure 3-6 Area under Cultivation, Production and Yield for Turmeric**

![Figure 3-6 Area under Cultivation, Production and Yield for Turmeric](source: Horticulture and Spice Board of India, (Commoditiescontrol.com))
India is the world’s largest exporter for turmeric followed by Pakistan, Bangladesh, Sri Lanka, Taiwan, China etc.

**Figure 3-7 Export of turmeric from India (quantity and value)**

Source: Spice Board of India, Turmeric Crop Survey and Analysis by (Dsouza, R., Murali, N., Shah, K., 2015) of Nirmal Bang

**Table 3-3 Contract Specification For Turmeric at NCDEX**

<table>
<thead>
<tr>
<th>Contract Specification</th>
<th>Turmeric Future Contract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticker</td>
<td>TMCFGRNZM</td>
</tr>
<tr>
<td>Trading hours</td>
<td>Mondays through Fridays: 10:00 a. m. to 5:00 p.m.</td>
</tr>
<tr>
<td>Contract months</td>
<td>As decided by the exchange</td>
</tr>
<tr>
<td>Contract Size</td>
<td>5 MT</td>
</tr>
<tr>
<td>Tick Size</td>
<td>Rs. 2</td>
</tr>
<tr>
<td>Daily Price Limits (DPL)</td>
<td>DPL will be (+/-) 2%. If the price touches (+/-) 2%, trading will continue with 2% limit for the 15 minutes period from the time 2% limit was reached. Thereafter, price limit would be extended by another (+)/ (-) 2 %. No trade would be permitted during the day beyond the price limit of (+)/(-)4% from the previous day’s closing price.</td>
</tr>
</tbody>
</table>
3.3.2 Macroeconomic Variables under Study

i. **USD-INR exchange rate** – USD is the currency of United States of America and is most sort after currency for carrying out international trades. Some of the international researchers have studied the predictive power of USD exchange rate for future commodity prices. (Harri, Nalley, & Hudson, 2009), (Chen, Rogoff, & Rossi, 2008). USD is the world’s reserve currency and any fluctuation in USD-INR has a significant impact on the current and capital account deficit of the country. India being one of the major exporters of many of the agriculture commodities, USD-INR is expected to have relationship with the commodity prices. Also as India imports majority of its crude oil, any major fluctuation in the currency is expected to have significant impact on agriculture commodities as crude and its derivatives are used as input. Exchange rate is amongst the most important macro-economic factor and hence linkage with agriculture commodities is worth a study ( (Bradshaw & Orden, 1990), (Frank & Garcia, 2010), (Nazlioglou & Soytas, Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis, 2012)).

ii. **S&P BSE-SENSEX** – Bombay stock exchange (BSE) is the leading stock exchange of India. Established in 1875, BSE is amongst the worlds’ top twenty stock exchanges with market capitalization of Rs 95.72 trillion as of April 2015. Standard and Poors’ BSE-Sensex also called as Sensex, is the BSE’s benchmark stock market index for Indian equity market and serves as a barometer of investment sentiment in India. Many researchers have studied the linkage between commodity prices and other asset markets ((Rossi, 2012), (Chakrabarty & Sarkar, 2010),(Mensi, Beljid, Boubaker, & Managi, 2013), (Kaur & Rao, 2010), (Choudhary, Nair, & Purohit, 2015)). Information flow and
inter-linkage between agriculture commodities and stock market have been found relevant for the study as

- Commodity serves as inputs for many companies traded on stock exchange thus impacting the cost structures and hence the profitability of the companies

- Commodity markets are still less developed in India and if relationship and causality is established then it will prove to be significant in forecasting the future trends (Chakrabarty & Sarkar, 2010)

- volatility transmission between commodity prices and stock index would help to hedge effectively and help portfolio managers in building an optimal portfolio and forecasting future stock return volatility (Mensi, Beljid, Boubaker, & Managi, 2013)

- Commodity futures market has been used as an alternative investment class and would be used by speculators and further help in financialization of commodity markets in India

iii. **Crude** – Crude oil is the largest imported commodity (in value terms) in India and is the major cause of trade deficit of the country. Crude price serves as an input to agriculture sector in the form of petrochemicals, diesel, petrol, etc. and also is an ancillary cost for transportation of agriculture produce to the markets. Since India satisfies most of its crude requirement through importing from OPEC and others, the movement in crude price has become a sensitive issue for the economic growth. Many countries; in wake of satisfying the demand for fuel; have diverted agriculture commodities like corn, soybeans, rapeseed, etc. for the production of biofuel. The debate on “Food vs. Fuel” has attracted the attention of many researchers. The increasing demand for fuel is likely to become an issue with depleting oil reserves and ever increasing world population. Diversion of agriculture produce to manufacture biofuel has increasingly led to the demand for those commodities that help to produce fuel and in turn increase in their prices. Farmers have shifted the cultivable land to produce more of these profitable commodities ignoring the demand for food. The present study attempts to explore the interlinkage between crude oil and agriculture commodity market. Inflation has always been a political issue and government remains concerned primarily with reference to inflation in agriculture commodities which are
items of mass consumption. In order to control inflation in agriculture commodities, Government often and on delists many futures contracts in agriculture commodities. Thus exploring whether crude price movement is a major reason for the movement in commodity prices will be interesting (Saghaian (2010), Rezitis (2015), (Ghaith, 2011), (Zhang & Reed, 2008), (Choudhary N. , 2014). The present study uses crude oil-Brent FOB price in USD per Barrel from Thomson-Reuters database.

3.4 Period of Study

In the proposed analysis, researcher will be taking daily data for the span of almost 10 years from 2005 to 2014 depending on the availability of information for different commodities. The future market in India was reactivated in 2003 after the formation of national commodity exchanges. To avoid the impact of thin trading during early introduction phase; the daily data for spot and future price of agriculture commodities is collected from 2005 and onwards.

3.5 Research Methodology

3.5.1 Finding of Structural Breaks in Data

Time-series is many times subjected to unexpected shift due to multiple factors leading to structural break in the data. If structural breaks are not analysed in a long-term study, this could possibly lead to forecast error as the break leads to change in the slope of the model. Political, economic and social factors can cause the structural change in time series which can in turn change the relationship among variables over time (ÖNEL, 2013). Earlier works of some researchers like (Quandt, 1958) and (Chow, 1960) have explored tests for structural change for a known single break date. Further many tests were developed to study the unknown breaks in time series like those suggested by (Andrews, 1993), (Andrews & Ploberger, 1994), (Liu, 1997), (Bai & Perron, 1998) amongst many others. The current study applies test proposed by (Bai & Perron, 1998) to find unknown structural breaks in the agriculture commodities spot and future price series. It would help to understand multiple breaks at unknown dates. With these breaks; the data will be divided in to sub-periods and a model would be estimated for each sub-period. This will help to assess the efficiency of the markets with reference to each break period identified.
**Bai-Perron test for Structural breaks**

A structural break appears on account of unexpected shift in time series data. A structural break means that the values of the parameters of the linear regression model do not remain the same through the entire time period. This may be caused due to external forces or major policy changes or many other causes. (Gujarati, 2004). Bai and Perron have proposed many models for identification of unknown structural breaks for time series estimated using linear least squares method. They are as follows-

1. \(L+1\) breaks vs. \(L\) sequentially determined breaks
2. Tests of \(1\) to \(M\) globally determined break
3. \(L+1\) vs. \(L\) globally determined breaks

The first method determines the optimal number of breaks based on the sequential methodology. Researcher will employ Bai and Perron tests of \(L+1\) vs. \(L\) sequentially determined breaks. Here the researcher would test the single added breakpoint that most reduces the sum-of-squares (Sequential \(L+1\) breaks vs. \(L\)). This is a sequential test of \(l\) versus \(l+1\) structural change:

\[
\sup_{T_{l}} (l+1/l) = \{S_T(\hat{T}_l, \ldots, \hat{T}_1) - \min \inf S_T(\hat{T}_l, \ldots, t, \hat{T}_1, \ldots, \hat{T}_1)/\sigma^2 \} \quad \text{Eq. 1}
\]

Where,

\[
\Lambda_{i,r} = \{t: \hat{T}_l - 1 + (\hat{T}_l - \hat{T}_1)i \eta \leq t \leq \hat{T}_l + (\hat{T}_l - \hat{T}_1)i \eta\}.
\]

\[
\beta_T(\hat{T}_1, \ldots, \hat{T}_l, \ldots, \hat{T}_1, \ldots, \hat{T}_1) \text{ is the sum of squared residuals resulting from the least squares estimation from each m-partition (T_1, \ldots, \ldots, T_m) and } \sigma^2 \text{ is a consistent estimator of } \sigma^2 \text{ under the null hypothesis.}
\]

(Refer to (Bai & Perron, 1998) for further details)

The procedure to test breaks in time series variable, at first considers full sample and perform test of parameter constancy with unknown break dates. The null hypothesis of the test is constancy i.e. no break in the series vs. alternative hypothesis one break. The rejection of null; will determine the break date and thereafter the sample is divided into two samples. After this in each sub-sample; single unknown break point tests are performed. These test
will be seen as conducting the null hypothesis test of “l=1” vs. alternative hypothesis of “l+1=2 break”. This process will add a break point whenever the null hypothesis is rejected. This process is repeated till the level of maximum breaks specified in the test.

Based on the derived break dates, the spot and future price of agriculture commodities will be divided in to non-overlapping phases and each phase will be analysed separately. The study also employs graphical representation of structural breaks in the price series.

### 3.5.2 Descriptive Statistics

The study proceeds to calculate the descriptive statistics involving measures of central tendency i.e. mean, mode, median, along with measures of dispersion (standard deviation) and measures of asymmetry (skewness and Kurtosis). The average of agriculture spot and future prices will be studied together with the standard deviation and minimum and maximum statistic to understand the volatility in the data series during each phase. Standard deviation is the most used measure of dispersion as it less affected by fluctuations in sampling (Jin & Kim, 2012). The measures of asymmetry i.e. skewness helps to understand the formation of the series, shape of the curve, and deviation of the series from normal distribution. Further Kurtosis helps to understand whether the data series is peaked or flat as compared to the normal distribution. Generally a bell shaped curve is considered to have normal distribution.

It is well known that knowing the shape of the distribution curve is critical to the use of statistical methods in research analysis. Most of the methods make specific assumptions about the nature of the distribution curve. The study will further use Jarque-Bera test and its probability value to analyze whether the time series under study deviate from normality. The test compares the sample values with the value of skewness and kurtosis matching normal distribution.

The Jarque Bera test has a joint null hypothesis that the sample has zero skewness and zero excess kurtosis. If null hypothesis is accepted, it means the data series exhibit normal property. However the rejection of null hypothesis in favour of alternative hypothesis; would indicate that the data is not distributed normally.
3.5.3 Test of Stationarity

Once the structural breaks are determined in the price series, the sample is divided into phases. Each of the variables under each phase is tested for stationarity using Augmented Dickey Fuller test as specified by (Dickey & Fuller, 1979). As the study deals with time series data, to avoid spurious results before formulating a model, the data is tested for stationarity. It helps to determine the order of integration of time series variables. An I (0) series is stationary in level, whereas, an I (1) series is random walk and achieves stationarity with differencing. The ADF test is an improvement of the standard dickey fuller test and has the existence of a unit root as the null hypothesis. Refer to the simple auto regressive process as in equation 2, giving standard dickey fuller test.

\[ y_t = \rho y_{t-1} + x_t \delta + \varepsilon_t \quad \ldots \text{eq. 2} \]

Where,

\[ x_t = \text{optional exogenous regressors (consisting of constant or constant and trend)} \]
\[ \rho \text{ and } \delta \text{ are parameters to be estimated} \]
\[ \varepsilon_t = \text{white noise} \]

After subtracting \( y_{t-1} \) from both the sides of equation 2, we get

\[ \Delta y_t = \alpha y_{t-1} + x_t \delta + \varepsilon_t \quad \ldots \text{eq. 3} \]

Where \( \alpha = \rho - 1 \).

The null and alternative hypotheses may be written as,

\[ H_0: \alpha = 0 \]
\[ H_1: \alpha < 0 \quad \ldots \text{eq.4} \]

This is evaluated using the conventional t-ratio for \( \alpha: t_\alpha = \hat{\alpha} / (se(\hat{\alpha})) \ldots \text{Eq.5} \)

Where, \( \hat{\alpha} \) is the estimate of , and \( se(\hat{\alpha}) \) is the coefficient standard error.
The ADF test constructs a parametric correction for higher-order correlation by assuming that
the series follows an AR (1) process and adding lagged difference terms of the dependent
variable to the right-hand side of the test regression:

\[ \Delta y_t = \alpha y_{t-1} + x_t \delta + \beta_1 \Delta y_{t-1} + \beta_1 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \epsilon_t \]  

Eq. 6

This augmented specification in equation 6 is then used to test the hypothesis specified in
equation 4 using t-ratio as calculated in equation 5.

3.5.4 Testing of Linear Property of Time Series

The testing of the linear property of data is an essential requirement prior to estimation of any
of the regression model. Many international studies have focused on this issue and have
documented the empirical evidence of existence of non-linear dependence in the time series
data (to quote few such studies; (Scheinkman & LeBaron, 1989), (Hsieh, 1991), (Qi, 1999),
(Abhyankar, Copeland, & Wong, 1995), (Soni T. K., 2014), (Lim & Hooy, 2012), (Silvapulle
& Choi, "Testing for Linear and Nonlinear Granger Causality in stock Price-Volume
Relation: Korean Evidence", 1999)). Non-linear property of data is also reported in case of
agriculture commodities (Soni T. K., 2014), (Soni T., 2013). However empirical
investigations on agriculture commodity prices in many developing economies including
India have largely ignored the testing of the property of data and have formulated linear
models considering the data series to be linear (for example (Chauhan, Singh, & Arora,
2013), (Chhajed & Mehta, 2013), (Dasgupta, 2004) (Gupta & Sunitha, 2013)). The testing of
the property of data holds paramount importance to specify correct empirical model. Further,
the existence of nonlinearity in return series is considered as evidence against the efficient
market theory. Efficient market theory states “data as random walk” or more strictly “a
martingale” i.e. market discounts all available information in prices and no party can predict
the market.

3.5.4.1 Variance Ratio Test

To address this issue; the present study aims to employ Variance ratio test as proposed by (Lo
& MacKinlay, 1988). Variance Ratio test check whether differences in the series are
uncorrelated, or they follow a random walk/ martingale property. The test is based on the
concept that the return of any series \( X_t \) follows pure random walk in a manner
\[ X_t = \phi + X_{t-1} + \epsilon_t \] \quad \text{Eq. 7}

Where, \( \phi \) = Arbitrary drift parameter
\[ \epsilon_t = \text{Error terms are independently and identically distributed} \]

The null hypothesis for variance ratio test hints towards nonlinear property of data.

\[ H_0: \text{The series is a random walk (non-linear)} \]

\[ H_1: \text{The series is linear (hence rejection of nonlinear property and weak form EMH)} \]

The variance of K-differences will grow proportionally with difference K. Therefore the check the null of random walk; we can compare the variance of one period with that of K-period returns using the following formula

\[ VR(k) = \left( \frac{1/nVAR(X_{t+n}-X_t)}{VAR(X_{t+n}-X_t)} \right) \quad \text{Eq. 8} \]

The Variance ratio test proved that if \( X_t \) is Independently and Identically Distributed (IID), then under null hypothesis \( VR(K)=1 \). In that case the test statistic \( M_1 \) is given as

\[ M_1(x,k) = (VR(x,k) - 1) \left[ \frac{2(2k-1)(k-1)}{3KT} \right]^{-\frac{1}{2}} \quad \text{Eq. 9} \]

For further details refer to (Lo & MacKinlay, 1988).

### 3.5.4.2 BDS Test

BDS is “a portmanteau test for time based dependence in a series that helps in testing possible deviations from independence including linear, non-linear dependence or Chaos”. The test was developed by (Brock, Davis, Scheinkman, & LeBaron, 1996) to test the independence of series based on correlation integration.

The correlation integral at embedding dimension \( m \) is shown by following equation

\[ C_{m,n}(\epsilon) = \frac{2}{(n-m+1)(n-m)} \sum_{S=1}^{n-m+1} \sum_{T=S+1}^{n-m+1} \prod_{j=0}^{m-1} I_e (X_{S+j}, X_{T+j}) \quad \text{Eq. 10} \]
Where

\[ I_\varepsilon(x,y) = \begin{cases} 1 & \text{if } |x - y| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \]

\[ C_{m,n} = \text{correlation integral} \]

These sample estimates of the probabilities can be used to construct a test statistic for independence

\[ b_{m,n}(\varepsilon) = c_{m,n}(\varepsilon) - c_{1, n-m}(\varepsilon)^m \] … Eq. 11

Under the assumption of independence, this statistic is expected to be close to zero. In fact, it is shown in (Brock, Davis, Scheinkman, & LeBaron, 1996) that

\[ (\sqrt{n - m + 1}) \frac{b_{m,n}(\varepsilon)}{\sigma_{m,n}(\varepsilon)} \to N(0,1) \] … Eq. 12

For further details on the test refer to (Brock, Davis, Scheinkman, & LeBaron, 1996)
3.5.5 Cointegration Analysis

Cointegration analysis depicts that the two time series move together in the long run. The cointegrating associationship between two time series data variables suggest that the variables are bound by some relationship in the long-run. The may deviate from the relationship in the short run; however their association would return eventually in long run. The cointegrating associationship is normally checked in level form of the data; however it is proved that if the series are co-integrated in levels they will be also be co-integrated in log level form (Hendry & Juselius, 2000).

The study proposes to employ Johansen Cointegration as well as Rank cointegration test to study the relationship of macro-economic variables with spot and future prices of agriculture commodities.

3.5.5.1 Johansen Cointegration Test

Johansen Cointegration test was proposed by Johansen (1988). It is a linear test that helps understand whether long-run relationship exists between variables. The test is applied on raw price series i.e. data in levels form. However, when time series exhibit nonlinear property, the test may produce spurious results of no cointegration between variables; when in reality there may be cointegrating associationship between them.

Johansen co-integration methodology takes its starting point in the Vector Auto regression (VAR) of $P$ given as

$$ y_t = \mu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + B x_t + \varepsilon_t \ldots \ldots \text{Eq. 13} $$

Where

$y_t$ is an $K$- vector of variables that are integrated of order one $I(1)$ and

$x_t$ is $d$-vector of deterministic variables

$\varepsilon_t$ is an nx1 vector of innovations.

We can further rewrite the VAR as


\[
\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + B x_t + \epsilon_t
\]

\[\text{...........Eq. 14}\]

Where,

\[
\Pi = \sum_{i=1}^{p} A_i - I, \quad \Gamma_i = - \sum_{j=1}^{i} A_j
\]

\[\text{...........Eq. 15}\]

Here “r” is the number of cointegrating relations called ranks and each column.

(Refer to Johansen, (1988) for further details)

As already discussed, Johansen test assumes linear combinations between parameters which may give suprious results of no cointegration in case the data exhibit nonlinear property. The study further proposes to use nonlinear cointegration test to unearth the true cointegrating relationship between variables under consideration.

3.5.5.2 Rank Cointegration Test

Rank test developed by (Breitung, 2001) would be deployed to check for non-linear cointegration between variables. (Breitung, 2001) proposed cointegration test based on rank transformation on the time series. A new series would be defined giving the rank of variables to find cointegration among them.

Consider two variables \( S_t \) and \( F_t \) in raw form i.e. integrated of order one (I (1)). (Breitung, 2001) defined the rank series as

\[
R_T(S_t) = \text{Rank of } S_t \text{ amongst } S_1, S_2, \ldots, S_T R_T(F_t)
\]

\[
= \text{Rank of } F_t \text{ amongst } F_1, F_2, \ldots, F_T
\]

The non-linearity equation between the time series is given as

\[
F_t = F(S_t) + \mu_t \quad \text{........Eq. 16}
\]
The distance measures between the two series are given by

$$ K_T = T^{-1} \sup_t |d_t| \quad \text{Eq. 17} $$

and

$$ \xi_T = T^{-3} \sum_{t=1}^{T} d_t^2 \quad \text{Eq. 18} $$

Where,

$$ d_t = R_T(F_i) - R_T(S_t) \quad \text{Eq. 19} $$

The null hypothesis of the test is $H_0 = \text{No nonlinear cointegration between } S_t \text{ and } F_t$ is rejected if the test statistic are too small. For critical values refer to page 334 of the paper (Breitung, 2001). The test uses Monte Carlo simulations to calculate associationship between two variables.

### 3.5.6 Non-linear Causality Test

The cointegrating relationship between variables suggests that their might lead-lag relationship between them. The Agriculture commodity prices exhibit non-linear tendency and hence the study proposes to explore the lead-lag relation using the non-parametric granger causality developed by (Diks & Panchenko, 2006). The causality study would help in understanding the role of price discovery played by futures market vis-à-vis Spot market and also the lead-lag relation with the macro-economic variables with agriculture prices under study.

The study is conducted using log returns of the data series as specified below;
Log returns = LN \( \frac{P_{t-1}}{P_t} \) * 100 \( \ldots \) Eq. 20

Diks & Panchenko (2006) suggests an improvement over the test proposed by Hiemstra & Jones (1994) which was found to over reject the null hypothesis of Granger non-causality. Causality is said to exist between variables if past and current values of one variable contains some information on the future value of another variable and vice versa.

The Diks & Panchenko (2006) test can be summarized with an example as- Let us consider 2 stationary time series \( X_t \) and \( Y_t \). The null hypothesis for the test is

\[ H_0 = X_t \text{ does not granger cause } Y_t. \]

For a stationary bivariate time series process \( \{X_t, Y_t\} \), the causality can be denoted as

\[ (Y_{t+1}, \ldots, Y_{t+k}) | (\mathcal{F}_{X_t}, \mathcal{F}_{Y_t}) \sim (Y_{t+1}, \ldots, Y_{t+k}) | \mathcal{F}_{Y_t} \] \( \ldots \) Eq. 21

Assuming that there are delayed vectors

\[ X_{t-1} = (X_{t-1} \ldots X_{t-1}) \text{ and } Y_{t-1} = (Y_{t-1} \ldots Y_{t-1}) \]

Where,

\[ l_x, l_y > 1 \]

\( \ldots \ldots \) Eq. 22

The null hypothesis assumes no causality between variables i.e. the past values of one variable has no information on another variable under study. It is similar to testing for conditional independence of \( Y_t \) on \( X_{t-1} \ldots, X_{t-lx} \), given \( Y_{t-1} \ldots, Y_{t-ly} \) as
\[ H_0: Y_t \left| (X_{t-1}^{lx}, Y_{t-1}^{lx}) \right. \sim Y_t \right| Y_{t-1} \quad \text{Eq. 23} \]

Where;

\[ X_{t-1}^{lx} = (X_{t-1}^{lx} \ldots X_{t-1}) \quad \text{and} \quad Y_{t-1}^{ly} = (Y_{t-1}^{ly} \ldots Y_{t-1}) \]

Null hypothesis is stating about the invariant distribution of \((l_x, l_y, +1)\) - dimensional vector \(W_t = (X_{t-1}^{lx}, Y_{t-1}^{ly}, Z_t)\), where \(Z_t = Y_t\). The joint probability density function after dropping of the time variable for ease notation, and its marginal; must satisfy the following equation

\[
\frac{f_{x,y,z}(x,y,z)}{f_y(y)} = \frac{f_{x,y}(x,y)}{f_y(y)} = \frac{f_{y,z}(z)}{f_y(y)} \quad \text{Eq. 24}
\]

The equation 24; explicitly states that \(x\) and \(y\) are independent conditionally on \(Y = y\), for each fixed value of \(y\). The revised formulation of null hypothesis implies as below

\[
q = E[f_{x,y,Z}(X,Y,Z)f_y(Y) - f_{x,Y}(X,Y)f_{y,Z}(Y,Z)] \quad \text{Eq. 25}
\]

Further the test statistic in a simplified manner is given as;

\[
T_n(c) = \frac{(n-1)}{n(n-2)} \sum_i \left( \hat{f}_{x,y,Z}(X_i, Y_i, Z_i) \hat{f}_y(Y_i) - \hat{f}_{x,Y}(X_i, Y_i) \hat{f}_{y,Z}(Y_i, Z_i) \right) \quad \text{Eq. 26}
\]

The test statistic \(T_n\) will satisfy the following equation based on the bandwidth selected \(E_n\)

\[
\sqrt{n} \left( \frac{T_n(c_n) - q}{S_n} \right) \rightarrow N(0,1) \quad \text{Eq. 27}
\]

For further details and clarification on bandwidth for relevant size of data; refer to the paper by Diks & Panchenko (2006).
3.6 Statistical and Other Softwares

The study uses Microsoft excel 2010 to cleanse the data, exhibit graphical representation in the form of charts and graphs and tabular presentation of data and empirical results. Eviews 8, a statistical software is used to carry out different tests. Gauss program and nonlinear causality code is further used to derive empirical results. Now the researcher will proceed to analyze the results/outcomes of the research in subsequent chapter.


