काल श्रृंखला आकड़ों में स्ट्रक्चरल ब्रेक के साथ कृषि मूल्य का पूर्वानुमान

Agriculture Price Forecasting with Structural Break in Time Series Data

RAJEEV RANJAN KUMAR

Doctor of Philosophy

in

Agricultural Statistics

ICAR-Indian Agricultural Statistics Research Institute
ICAR-Indian Agricultural Research Institute
New Delhi- 110012

2020

Agriculture Price Forecasting with Structural Break in Time Series Data

By

Rajeev Ranjan Kumar

Thesis submitted to the Faculty of Post-Graduate School,
ICAR-Indian Agricultural Research Institute, New Delhi,
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

AGRICULTURAL STATISTICS

ICAR-Indian Agricultural Statistics Research Institute
ICAR-Indian Agricultural Research Institute
New Delhi - 110012

Approved by:

Chairman: .....
(Dr. Girish Kumar Jha)

Co-chairman: .....
(Dr. Rajender Parsad)

Member: .....
(Dr. Achal Lama)

Member: .....
(Dr. Anshu Bharadwaj)

Member: .....
(Dr. Dwijesh Chandra Mishra)

Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
PUSA, New Delhi - 110012

Dr. Girish Kumar Jha
Principal Scientist (Agricultural Statistics)

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled "**Agriculture Price Forecasting with Structural Break in Time Series Data"** submitted in partial fulfilment of the requirement for the degree of **Doctor of Philosophy** in **Agricultural Statistics** of the **Post-Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi,** is a record of bonafide research carried out by **Mr. Rajeev Ranjan Kumar** under my guidance and supervision and no part of this dissertation has been submitted for any other degree or diploma.

All assistance and help received during the course of this investigation has been duly acknowledged.

Girish Kumar Jha

Date: 09/11/2020 ***(Girish Kumar Jha)***

Place: New Delhi Chairman, Advisory Committee

Dedicated to My Esteemed Mother

Late Rita Sinha

and

My Father Sh. Anandi Prasad Singh

Acknowledgements

This thesis came to fruition due to the knowledge gained over the entire period of my Ph.D. study at ICAR–Indian Agricultural Statistics Research Institute (IASRI), New Delhi during which, I have been in touch with a great number of people whose contribution in varied yet myriad ways has led to the research and making of the thesis which deserve special mention. It is a pleasure to convey my gratitude to all of them by way of my humble acknowledgements.

First and foremost, with reverence, I want to express my deepest sense of gratitude to ***Dr. Girish Kumar Jha****, Principal Scientist, Division of Agricultural Economics, ICAR-Indian Agricultural Research Institute, New Delhi and Chairman of my advisory committee for his initiative, benevolence, endurance, constructive criticism and constant monitoring during the period of my investigation and also in the preparation of this thesis. Above all and most needed, he provided me constant support and encouragement in various ways. I consider myself fortunate in having the privilege of being guided by him. I am really indebted to him.*

I am equally indebted to ***Dr. Rajender Parsad****, Director and Principal Scientist, ICAR-IASRI, New Delhi and Co-chairman of my Advisory Committee for his counsel, valuable suggestions and ready to help for accomplishment of my research work.*

I am extremely thankful to ***Dr. Achal Lama****, Scientist, Division of Forecasting and Agricultural Systems Modelling, ICAR-IASRI, New Delhi and member of my Advisory Committee for his useful suggestions and constant support during my research work.*

I am also highly grateful to ***Dr. Anshu Bharadwaj****, Principal Scientist, Computer Applications, ICAR-IASRI, New Delhi and member of my Advisory Committee for her useful suggestion and constant support during my thesis work.*

I am extremely thankful to ***Dr. Dwijesh Chandra Mishra****, Scientist, Centre for Agricultural Bioinformatics, ICAR-IASRI, New Delhi and member of my Advisory Committee for his useful suggestions and immense support during my investigation.*

I am also highly grateful to ***Dr. Bishal Gurung****, Scientist, Division of Forecasting and Agricultural Systems Modelling, ICAR-IASRI, New Delhi and ex-member of my Advisory Committee for his useful suggestion during my thesis work.*

I would like to convey my deep sense of gratitude and appreciation to ***Dr. Seema Jaggi****, Principal Scientist and Professor, Discipline of Agricultural Statistics, ICAR-IASRI for her encouragement for timely completion of course work. I am indebted to her for valuable advice and constructive criticism during preparation of my investigation.*

*I am extremely thankful to **Dr. K. N. Singh**, Principal Scientist and Head (A), Division of Forecasting and Agricultural Systems Modelling, ICAR-IASRI, New Delhi for his immense administrative support during my research work.*

*I am also extremely thankful to **Sanjeev Kumar Sir** and **Neeraj Budhlakoti**, Scientist, Centre for Agricultural Bioinformatics, ICAR-IASRI, New Delhi for their constant motivation and immense moral support during my research work.*

*I want to express my respect and gratitude to **my parents, my wife Ritu Rani, my elder brothers Dr. Chandan Kumar, Nivash Kumar and Ravi Ranjan, my bhabhiji, my younger brother Abhishek Ranjan and all the family members** for their sacrifices they did for me. These were the inspirations and moral boosting which gave me sufficient energy to complete this thesis in time.*

*I am thankful to my classmate and friends **Harun, Pramod, Sadikul, Pratish, Saurav and Sumeet** for their friendly approach and moral support throughout my entire Ph.D. tenure.*

*I do not have any words to express the help and guide given by my seniors specially, **Raju Sir**, **Prakesh Sir**, **Sunil Sir**, **Pradeep Sir, Vandita mam, Anandita mam** and my juniors **Rahul, Shyam, Pankaj, Nitin, Kuldeep, Gopal, Satpati, Amit, Asish, Akhilesh, Ronit, Yashin, Dipankar, Asit, Kapil, Jitendra, Novin, and all others** for their affectionate support and help.*

I am also thankful to all my seniors, juniors and friends in IARI for their moral support and inspirations.

*I take this opportunity to appreciate the help rendered by the staff of PG school and TAC. Special thanks are due to **Sanjeev ji**, **Gagan ji**, and **Sunil ji.***

I extend my thanks to the Director, the Dean and the Associate Dean, ICAR-IARI, New Delhi and the Director, ICAR-IASRI for their helpful attitude and cooperation, throughout the period of study.

*I am thankful to **ICAR** and **Ministry of Human Resource Development (MHRD)** for the financial assistance provided to me during the tenure of my study.*

Finally, I would like to thank everybody who was imperative to the successful realization of this thesis, as well as articulate my apology that I could not mention personally one by one.

Rajiv Ranjan Kumar

Date: 09/11/2020

ICAR-IASRI, New Delhi-110012 *(Rajeev Ranjan Kumar)*

Contents

Abbreviations

ACF- Autocorrelation Function

ADF- Augmented Dickey-Fuller

AGMARKNET- Agricultural Marketing Information Network

AIC- Akaike Information Criteria

ANN- Artificial Neural Network

AR- Autoregressive

ARCH- Autoregressive Conditional Heteroscedasticity

ARIMA- Autoregressive Integrated Moving Average

ARFIMA- Autoregressive Fractionally Integrated Moving Average

ARMA- Autoregressive Moving Average

BDS- Brock-Dechert-Scheinkman

BF- Basis Function

BIC- Bayesian Information Criteria

BP- Backpropagation

CDC- Correct Directional Change

DF- Dickey-Fuller

DM- Diebold-Mariano

ECM- Error Correction Model

ECT- Error Correction Term

ELM- Extreme Learning Machine

EML- Exact Maximum Likelihood

GARCH- Generalized Autoregressive Conditional Heteroscedasticity

GPH- Geweke and Porter-Hudak

ICSS- Iterative Cumulative Sum of Squares

KPSS- Kwiatkowski, Perron, Schmidt and Shin

LM- Lagrange Multiplier

MA- Moving Average

MAD- Mean Absolute Deviation

MAE- Mean Absolute Error

MAPE- Mean Absolute Percentage Error

MLP- Multi Layer Perceptron

MS-GARCH- Markov Switching GARCH

NHRDF- National Horticultural Research and Development Foundation

NN- Neural Network

OLS- Ordinary Least Squares

PACF- Partial Autocorrelation Function

PELT- Pruned Exact Linear Time

PP- Phillips-Perron

RBF- Radial Basis Function

RESET- Regression Equation Specification Error Test

RMSE- Root Means Square Error

R/S- Rescaled Range

SLFN- Single Layer Feedforward Neural Network

SVR- Support Vector Regression

TDNN- Time Delay Neural Network

VAR- Vector Autoregressive

VECM- Vector Error Correction Model

Measurement Units

mt- Metric Tonne

q- Quintal

₹- Indian Rupee

$- US dollar

CHAPTER I

INTRODUCTION

1.1 Background

Contribution of agriculture in total Gross Domestic Product (GDP) is relatively low in the developed countries i.e. (1%), but it raises to 15% for the developing countries like India (The World Bank, 2018). In India, around 60-70% of total population directly or indirectly depend on agriculture and allied sectors. The contribution of agriculture and its allied sectors to the Indian economy is inescapable. Given the importance of agricultural commodity prices on food security and its large negative economic and social effects, understanding the dynamics of agricultural commodity prices is vital. The prices of agricultural commodities are determined by a variety of factors, including adverse weather conditions, natural disasters, shifts in demand and supply (e.g., due to agricultural policy changes) *etc.* Such variables cannot be quantified by the same norm, and have different effects on various agricultural commodities in various wholesale markets, which further makes the forecasting of agricultural commodity prices extremely challenging.

Agricultural products are characterized by means of enormous price fluctuations. Agricultural commodities price volatility create a level of uncertainty which increases risks for farmers, traders, consumers and governments as they need greater human resource skills to manage markets in a volatile state. The effect of volatility in agricultural commodities prices become more severe in underdeveloped and developing countries like India where poor households may suffer food scarcity problems. There is emerging consensus that the global food system is becoming more vulnerable and susceptible to occurrences of extreme price volatility (FAO, 2011). After the food price crisis of 2007–2008, the level and volatility of staple food prices have increased by more than 50% (Brummer *et al.,* 2016). To mitigate the agricultural commodity price volatility, it is necessary to understand its causes, patterns and impacts on the farmers and consumers. Modelling agricultural commodity price volatility helps to forecast the absolute magnitude, in fact, the entire distribution of price changes. Such forecasts are extensively

utilized in risk management, derivative pricing and hedging, portfolio selection, among other financial activities.

1.2 Forecasting Agriculture Price

Modelling and forecasting of agricultural commodity prices are different from others non-farm commodity prices due to its seasonality of production, adverse weather conditions and biological nature of production. In the literature, there are mainly two basic approaches of forecasting viz. structural and time series models. The structural forecasting models proceed from the first principles of consumer and producer theory to identify the demand and supply schedules and the equilibrium prices resulting from their intersection. Due to computational and data demands of structural price forecasting, it is generally far exceed than what are routinely available in the developing countries. Contemporary parsimonious form of agriculture price forecasting relies mainly on time series modelling.

During the last few decades, so much effort has been devoted to the development and improvement of time series forecasting models. Initially, exponential smoothing techniques were used for the extrapolating of univariate time series data. After the publication of 'Time Series Analysis: Forecasting and Control' by Box and Jenkins (1970), the Autoregressive Integrated Moving Average (ARIMA) model became a most common and widely used method for forecasting. For the fitting of ARIMA model, they provided a systematic three-stage procedure viz. identification, estimation and diagnostics. When a stationary time series follows ARMA process, then its values depends on the past values and also the error term is related to its pasts. Due to this feature, ARIMA model is preferred for short term forecasting. The ARIMA model and its extensions have been heavily used in the agriculture price forecasting in last few decades. However, it can only capture linear patterns of time series, which is not always satisfactory for complex real world problems.

1.3 Structural Break

In time series analysis, factors such as major changes in technology like introduction of genetically modified crops as well as implementation of new economic policy *etc*. causes structural break(s) in the parameters of forecasting models. Structural break is being

commonly observed in many economic and agricultural time series data. In the statistical term, we can define structural break as a permanent change in the parameter vector of the model. A key question that arises in the context of forecasting is how future values of the time-series of interest might be affected by structural breaks. Generally, we consider the case where such breaks are exogenous, in the sense that they were determined by events outside the model under study. We also usually assume that such breaks were unanticipated given the historical data upto that point. Such structural breaks pose a formidable challenge to economic forecasting and it is the main source of forecast failure (Clements and Hendry, 1998).

Detection of structural breaks is a critical empirical activity, for the obvious reason that if such breaks are ignored then econometric relations are misspecified, from which numerous problems may flow. Detection of structural breaks in a series is now a researchable issue let that be single or multiple. For single break point, one can use centered cumulative sum of squares algorithm or the AtMost One Change Point (AMOC) algorithm. The algorithm for detection of multiple change points in a series was first given by Inclan and Tiao (1994) in the form of Iterated Cumulative Sum of Squares (ICSS). Another efficient algorithm for multiple change points detection has been developed as Pruned Exact Linear Time (PELT) by Killick *et al.* (2012).

1.4 Time Series Volatility Models

To model non-constant conditional variance of financial series, Engle (1982) proposed Autoregressive Conditional Heteroscedasticity (ARCH) model. The generalization of the ARCH model, which is popularly known as Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model was proposed by Bollerslev (1986). The following steps may be considered for the building of a volatility model:

- Specify a mean equation
- Use the residuals of the mean equation to test for ARCH effects
- Specify a volatility model if ARCH effects are statistically significant
- Check the fitted model carefully and refine it if necessary

In order to forecast price volatility, ARCH and GARCH models are extensively used due to its good performance in capturing the time-varying features of the data (Perales 2009, Lama *et al.* 2016). However, the standard GARCH model is intrinsically symmetric, and the forecasting results may be biased when skewed time series are considered (Franses and Dijk, 1996). To solve this problem, some nonlinear and asymmetric GARCH models are proposed for volatility forecasting, like gjrGARCH model by Glosten *et al.* (1993) and eGARCH model by Nelson (1991).

In all the above-mentioned models it is assumed that the GARCH process is stationary (white noise process for which the weak stationarity assumption holds) and only captures time-varying conditional variances that are not highly irregular. However, financial markets are occasionally hit by rather some extreme events that causes large disturbances. Events such as financial crises, flash crashes and market disruptions, *etc.* can impact the series and lead to sharp breaks in the unconditional variance and thus violate the assumptions for the GARCH process. Failure to accommodate this in the model may have serious effects on the forecasting abilities of the model. Upward bias in the degree of persistence in the estimated GARCH model can be caused by failure to accommodate structural breaks in the model, which was shown by Diebold (1986), Lamoureux and Lastrapes (1990), and Hillebrand (2005).

To solve this issue, regime switching models are employed to forecast price volatility. Cai (1994), Hamilton and Susmel (1994) and Gray (1996) introduced the regime switching process into the GARCH model, in order to consider potential structural breaks which is popularly known as Markov-Switching GARCH models. In the Markov-Switching GARCH models (MS-GARCH), parameters are changes over time according to a discrete latent (i.e., unobservable) variable. These models can adapt quickly to variations in the unconditional variance, which improves the volatility predictions (Ardia, 2008). Haas *et al.* (2004) did the further extension of the MS-GARCH model of Gray (1996). Markov-switching GARCH models allow for different GARCH behavior in each regime to capture the difference in the variance dynamics of low and high-volatility periods.

1.5 Long Memory Process

In the modelling of some agricultural price and macroeconomic data series, high persistence over relatively long periods of time and dependence structure across time play an important role. These data series are characterized by distinct, but non-periodic, cyclical patterns and are behavioral in such a way that current values are affected not only by immediate past values, but also by values from periods of time. There has been growing interest in capturing such a feature referred to as long memory. A time series has a long memory, whenever the dependence between apart events diminishes very slowly as the number of lags increases. The first attempt to parameterize the long memory character represented as memory parameter *d* was proposed by Granger and Joyeux (1980). In certain empirical contexts, Granger and Joyeux (1980) coupled the concept of long memory and fractional integration to provide theoretical explanation for the hyperbolic decay of the correlogram.

Consider a stochastic process, X_t is written in a simple generalization of a random walk,

$$X_t = (1 - L)^{-d}\epsilon_t$$

and the process $\epsilon_t \in I(0)$, i.e. an *i.i.d.* with zero mean and finite variance series. When $d = 0, X_t = \epsilon_t$ and thus X_t is weakly autocorrelated. For $d > 0$, X_t is strongly dependent and referred to as persistent, while when $d < 0$, X_t is negatively dependent and referred to as anti-persistent. The term $(1 - L)^{-d}$ can be defined by the binomial expansion and only exists for $-\frac{1}{2} < d < \frac{1}{2}$

$$(1 - L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)} L^j$$

where $\Gamma(\cdot)$ is the gamma function. The Autocorrelation Function (ACF) of X_t can be expressed by using Stirling's formula,

$$\rho_k \sim C k^{2d-1} as\ k \to \infty$$

where *C* is a positive constant. The summary of time series characteristics related to the long memory parameter *d* is given in Table 1.1.

Table 1.1: Characteristics of long memory time series

d	**Series**	**Memory**	**ACF**
$d \in (-0.5,0)$	Stationary fractional	Anti-persistant	Hyperbolic
$d = 0$	Stationary	Short memory	Exponential
$d \in (0, 0.5)$	Stationary fractional	Long memory	Hyperbolic
$d \in (0.5, 1)$	Non-stationary fractional	Long memory	Hyperbolic
$d = 1$	Unit root	Infinite memory	Linear

Recently, there has been a rise of interest in the possibility of confusing long memory and structural breaks. Perron (1989) showed that structural breaks and unit roots (*d=1*) are easily confused, for example a stationary process contaminated with structural breaks, tests of the null hypothesis of a unit root in the process are biased towards non rejection. This phenomenon has already been shown to apply in the context of the long memory as well. When a stationary short memory process is contaminated by structural breaks in level, the estimate of *d* is biased away from zero and the ACF exhibits a slow rate of decay (Granger and Hyung, 2004).

1.6 Artificial Intelligence (AI) Models

Artificial Intelligence is a branch of applied science that improves the performance of a machine or model by providing the power to mimic like human brain for solving complex problems. These AI models have better performance compared to the traditional models in case of nonlinear pattern due to the self-adaptive and data driven nature. A frequently used artificial intelligence model for forecasting agricultural price is the Artificial Neural Network (ANN). ANN contains three layers namely input layer, one or more hidden or middle layers and output layer. It is non-linear, nonparametric, data driven self-adaptive method with a few a priori assumption about the data series. ANN can learn from experience and have ability to infer about the unseen part of the population even if the sample data is noisy and incomplete. Hence, it is suitable for forecasting agricultural price series, which is inherently noisy and nonlinear in nature.

ANN is a powerful data-driven algorithm to capture the complex patterns present in the data series. However, a major challenge encountered by the ANN is the requirement of iterative tuning of model parameters and overfitting of the model. To overcome these limitations of ANN, Huang *et al.* (2006) proposed Extreme Learning Machine (ELM) model. ELM is based on single-hidden layer feedforward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. It has good generalization performance and fast learning speed compared to the ANN.

1.7 Hybrid Time Series Models

In order to achieve desired and accurate forecast, hybridization of existing forecasting models is an important technique. Hybridization is generally performed due to the lack of the comprehensive individual model in capturing various patterns in the data, concurrently. Hybrid time models combines the strength of each individual models. The main advantages of hybrid models are as follows:

- Improve forecast accuracy due to comprehensive pattern detection and modeling.
- Reduce the risk of using inappropriate model due to the combination of forecasts.
- Simplify the procedure of model selection due to the use of different components.

1.8 Co-integration Analysis

Granger (1981) shown that macroeconomic models containing nonstationary variables can be constructed in such a way that the results are both statistically sound and economically meaningful. Granger has achieved this breakthrough by introducing the concept of co-integrated variables. Co-integration is an econometric concept, which mimics the existence of a long-run equilibrium relationship among economic time series. If two or more series are themselves nonstationary, but a linear combination of them is stationary, then they are said to be co-integrated (Engle and Granger, 1987).

1.8.1 Engle-Granger Method for Testing Co-integration

Suppose that two variables Y_t and X_t are integrated of order 1 and we want to determine whether there exists an equilibrium relationship between these two variables. Engle and Granger (1987) proposed a four step procedure to determine if two $I(1)$ variables are co-integrated of order $CI(1,1)$.

Step-1: Pretest the variables for their order of integration

The first step in the analysis is to pretest each variable to determine its order of integration. For the co-integration analysis, each variable should be integrated of the same order. If both variables are stationary, it would not be necessary to proceed since standard time series methods apply to stationary variables and if the variables are integrated of different orders, it should be concluded that they are not co-integrated.

Step-2: Estimation of long-run equilibrium relationship

If, both variables Y_t and X_t are $I(1)$, then estimate the long-run equilibrium relationship in the form

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

If the variables are co-integrated, an ordinary least squares regression yields a super-consistent estimator of the co-integrating parameters β_0 and β_1.

Step-3: Estimate the Error Correction Model (ECM)

If the variables are co-integrated, the residuals from the equilibrium regression can be used to estimate the error correction model.

Step-4: Assess model adequacy

There are several procedures available that determine whether the estimated error-correction model is appropriate or not.

Disadvantages of Engle-Granger procedure

In this method it is possible to find that one regression indicates that selected variables are co-integrated, but when the order of variables get reversed it indicates no co-integration. This method has no systematic procedure for the separate estimation of the multiple co-integrating vectors.

1.8.2 Johansen Co-integration

In order to overcome the limitations of Engle-Granger method of co-integration, Johansen (1988) proposed test for presence of multiple co-integrating vectors, which is multivariate generalization of the Dickey-Fuller test. Johansen test of co-integration allows the researcher to test restricted version of the co-integrating vector(s) and speed of adjustment parameters. The first step of Johansen test involves the determination of co-integrating

rank, that is, the number of co-integrating relations. Johansen derived the maximum likelihood estimator using sequential tests for determining the number of cointegrating vectors. Two statistics, namely 'trace statistic' and 'maximum eigenvalue statistic' are used for testing the hypothesis. The trace statistic tests the null hypothesis of *r* cointegrating vectors against the alternative hypothesis of *K* (number of variables) co-integrating vectors. If $r = 0$, it means that there is no relationship among the variables. The maximum eigenvalue statistic, on the other hand, tests the null hypothesis of *r* co-integrating vectors against the alternative hypothesis of *(r + 1)* co-integrating vectors.

1.9 Motivation of the Study

High levels of volatility persistence may be spurious if there are structural breaks or regime shifts in the volatility process (Lamoureux and Lastrapes, 1990). This biased persistence of volatility indicates that the current information will still impose significant impacts on the conditional variance forecast for all horizons because of the very close to permanent influence on volatility. The standard GARCH models fail to account such biased persistence of volatility. Therefore, to obtain more robust estimates of conditional volatility, it would require a more general class of GARCH model that allows for structural breaks as part of the data generating process.

It is being increasingly evident from the time series literature that the presence of long memory in the data can be easily confused with structural break. Wrongly accounting for one when the other is present or acknowledging only one when both are present may lead to serious forecast failure. Given that it is often difficult to distinguish between the two, it is desirable to establish forecast methods that are robust to structural change and also appropriately account for long memory persistence.

The equilibrium relationship among nonstationary economic variables is often characteristic of structural breaks. Standard co-integration tests such as Engle-Granger test or Johansen co-integration test have low power under a co-integration relationship with structural breaks. Therefore, testing for co-integration in the presence of structural breaks is important for examining the equilibrium relationship among economic variables.

Keeping these research gaps in mind, following objectives have been formulated for the present investigation:

1. To develop and empirically evaluate GARCH type model for agriculture price forecasting in the presence of structural breaks
2. To assess the performance of long memory forecasting models with structural breaks for agriculture prices
3. To investigate co-integrated time series in the presence of structural break for agriculture prices

1.10 Orientation of the Thesis

Chapter 1 gives the brief idea about the structural breaks and its effects on the forecasting performance of the volatility and long memory models. The brief overview of co-integration, ANN and ELM models are also described in this chapter. The review of literature related to various time series models in the presence of structural break are discussed in the chapter 2. Chapter 3 describes the detail discussion of structural breaks and its detection techniques, volatility models, long memory models and co-integration with structural break, and proposed methodology. Overall results of proposed hybrid models and the inference of the obtained results are presented in the chapter 4. Chapter 5 summarizes the overall obtained results of the present study. Finally, the thesis is concluded with abstracts (english and hindi both) followed by appendix and bibliography.

CHAPTER II

REVIEW OF LITERATURE

Empirical evidence suggests that many macroeconomic and financial time series are subject to occasional structural breaks. Structural breaks in the parameters of forecasting models could arise from factors such as major changes in the economic policy, introduction of new variety, *etc.* Here, an effort is made to review popular time series volatility models, long memory models and co-integration analysis with structural breaks.

Box and Jenkins (1970) proposed a comprehensible, versatile, iterative three-stage cycle of time series model development, which includes (i) identification, (ii) estimation, and (iii) verification. This approach is popularly known as Box–Jenkins methodology.

Granger and Joyeux (1980) and Hosking (1981) introduced Autoregressive Fractionally Integrated Moving Average (ARFIMA) model. They examined long memory process through a parametric tool. They showed that ARFIMA model is good for capturing the long range dependency existing in the dataset by using differencing parameter (*d*) which is non-integer in contrary to the Box-Jenkins methodology which permits only integer value.

Booth *et al.* (1982) proposed a statistical tool for testing the presence of long memory in any time-series data called Hurst exponent (H). They used it to empirically investigate the behaviour of forex during the latest experiences with both flexible and fixed rates. More specifically, they tested the possible presence of long-term dependency in the exchange rate of the French franc, British pound, and German mark.

Engle (1982) developed the class of Autoregressive Conditional Heteroscedastic (ARCH) model in order to explain the dynamic fluctuations in conditional variance of a time series. Conditional variance was used as a deterministic function of past returns.

Geweke and Porter-Hudak (1983) proposed GPH estimator for the estimation of Long-memory parameter *d*. They developed an estimator of the parameter of ARFIMA model by regressing log periodogram on the deterministic regressor.

Helms *et al*. (1984) applied rescaled range (R/S) analysis for testing the presence of long memory. In addition, they showed empirically that their method is more efficient in forecasting the prices of soybean. Their hypothesis of testing was that the changes in price for chosen commodities futures contracts are not dependent on earlier price changes.

Bollerslev (1986) generalized ARCH model in which conditional variance is also a linear function of its own lags. This model was able to overcome the deficiencies of the previously proposed ARCH model.

Engle and Granger (1987) introduced the concept of co-integration for analysis of the long-run equilibrium relationship between economic variables.

Johansen (1988) proposed multivariate generalization of the Dickey-Fuller test for testing the presence of multiple co-integrating vectors. This test of co-integration permit the researcher to test the restricted version of the vector(s) and swiftness of adjustment parameters.

Lamoureux and Lastrapes (1990) demonstrated the effect of structural breaks in volatility process by introducing deterministic shifts in the variance and find marked reduction in the degree of volatility persistence relative to standard GARCH models.

Cheung (1993) discovered that in long run forecasting, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models and Autoregressive Fractionally Integrated Moving Average (ARFIMA) do not have much success rate. They used three different tests *i.e.* Modified Rescaled Range (MRR) test, Geweke-Porter-Hudak (GPH) test and two Lagrange Multiplier (LM) type tests for fractional integration. They reported that GPH and MRR tests were robust to autoregressive and heteroscedasticity components.

Allen (1994) observed that initially conventional methodologies of economics for price forecasting had major occupancies but later comparison and adoption of more accurate forecast methods have been started for time series observations. He pointed out the issue of little emphasis on the accuracy of models rather than explanation. He also commented that the principles underlying time series models like Autoregressive Conditional Heteroscedasticity (ARCH) model and its generalized form as the GARCH model assume

that there are periods of relative high and low volatility, though the underlying unconditional variance remains unchanged.

Clements and Hendry (1995) considered the implications for forecast accuracy of imposing unit roots and cointegrating restrictions in linear systems of $I(1)$ variables in levels, differences, and cointegrated combinations.

Robinson (1995) introduced an LM test to test nonstationarity as null hypothesis in the form of unit root. The Robinson's test was more robust than the other tests based on AR alternatives. Besides, the author has also shown that asymptotically the test is most powerful against a wide class of alternative hypothesis. At last, he showed that the test statistic is a chi-square distribution asymptotically.

Donaldson and Kamstra (1996) used ANN to combine time series forecasts of stock market volatility from the Japan, USA, Canada and the UK. They found that combining with nonlinear ANNs generally produces more accurate forecasts compare to the linear combination. They also highlighted the superiority of the ANN, because of its flexibility to account for potentially complex nonlinear relationships.

Gray (1996) proposed a generalized regime switching model in which first-order Markov process with state-dependent transition probabilities governs the switching between regimes.

Malliaris and Urrutia (1996) discovered that the agricultural futures contracts including Corn, Oats, Soybeans, Wheat, Soybean meal, and Soybean oil were correlated after examining their interrelationships. Similarly, Dawson and White (2002) applied the co-integration analysis to discover any sort of relationships among Barley, Coffee, Cocoa, Wheat, Sugar and futures, however they identified only one such relationship which was between Wheat and Barley.

Donaldson and Kamstra (1997) applied the GARCH-NN models to model of stock market volatility for New York, London, Tokyo and Toronto. They reported that the nonlinear model was semiparametric and was better than the other traditional models in both in-sample and out-of-sample performance.

Clements and Hendry (1998) argued that structural breaks pose a formidable challenge to economic forecasting and it as the main source of forecast failure.

Diebold and Inoue (2001) were the pioneers in studying the analysis of both structural break and long memory together and showed the way how one misinterprets structural break as the long memory. They have displayed analytically that how stochastic regime switching can be easily confused with the long memory, also asymptotically, so far as only a "small" magnitude of regime switching take place. They have reported that the long memory and regime switching are closely related to each other. In addition, they have given an evidence of improvement of their concept by considering numerous environments, like using a simple mixture model, Hamilton's (1989) Markov-switching model and Engle and Smith's (1999) stochastic permanent break model.

Fong and See (2002) employed a Markov Regime-Switching (MRS) approach for GARCH-dynamics, and abrupt changes in both mean and variance for modelling of the conditional volatility of crude-oil prices. Their result showed that the regime-switching model performed better than other non-switching models.

Gil-Alana (2002) has proposed a joint test which is an altered version of Robinson's test and it tests the fractional integration order and the necessity of structural break at the known point of time together. The test was applied to the yearly GDP of US. The results of the test revealed that the integration of order greater than 1 is required for raw time-series data and at the same time a slope dummy must be introduced in the regression model in consideration with the break due to World War II.

Zhang (2003) developed a hybrid model by combining both ARIMA and Neural Network (NN) model for forecasting a time series. He established that the recommended hybrid methodology has improvement over the unique strength of individual ARIMA and ANN models in both linear and nonlinear modelling.

Man (2003) discussed the usefulness of ARMA models having low order with fractionally differenced ARFIMA (0, d, 0) structure, $-0.5<d<0.5$, in the prediction of time series containing long memory. He argued that for short term prediction, a suitably adjusted ARMA (2, 2) model can yield competitive forecasts. Statistical evaluation showed that its

prediction error variance is at the most 0.6% higher at one step ahead, and at most 2.8% higher up to 10 steps ahead in comparison to the true model. The predictability of the model is also calculated and matched with that of the ARFIMA (0, d, 0). For empirical illustration, they forecasted the consumer price index of US and inflation rates for four different countries using both the adjusted ARMA (2, 2) and ARFIMA (0, d, 0). Using an out-of-sample prediction mean square errors as evaluating criterion, the empirical results suggested that the ARMA(2, 2) outperformed the ARFIMA (0, d, 0) in forecasts of up to 100 steps ahead.

Ai *et al.* (2006) fitted a partial equilibrium model using quarterly inventory and harvest data of Barley, Wheat, Corn, Oats and Soybeans collected from January, 1957 to September, 2002. They found that there was no such excessive co-movement as it was claimed and much of the co-movements were due to common trends in supply and demand factors.

Aiolfi and Timmermann (2006) developed a novel four-stage method for forecasting purpose. At first stage, models were sorted into clusters on the basis of their past performance. Then at second stage, forecasts were pooled within each cluster and at third stage the optimal forecast combination for these clusters were estimated and followed by shrinkage towards equal weights. This method worked well empirically for out-of-sample forecasting.

Yu *et al.* (2006) investigated the interrelationship among crude oil prices and vegetable oils including Soybean, Rapeseed, Sunflower, and Palm oil used for biodiesel production. In addition, they applied co-integration analysis to examine the interdependence among the vegetable oils only. They discovered only one cointegrating vector among the selected vegetable oils and crude oil. Besides, they also found that the crude oil price shocks did not have a substantial impact on altering vegetable oil prices.

Tsangari (2007) used the exchange rates for three countries relative to US dollar: Mexico (Peso), South Korean (Won) and the United Kingdom (British Pound) for developing an alternative methodology for combining different forecast models. The study revealed that neither ANN nor GARCH was able to capture volatility efficiently but their combination was found superior.

Clark and McCracken (2009) discussed situation when linear models are subject to structural change by giving logical, Monte Carlo and practical evidence on combining rolling and recursive forecasts. They used the trade-off between bias and variance for getting optimum configuration and concluded by their Monte Carlo experiments and many empirical examples that combination of the recursive scheme and the rolling scheme can often result improvements in forecast accuracy relative to forecasts made using individual scheme with a fixed window width.

Che and Wang (2010) developed a hybrid model called SVRARIMA model for forecasting purpose, which is nonlinear and nonstationary in nature. The SVR model of the hybrid model was used to capture the nonlinearity and then the ARIMA method was used for the estimation of the residual of the regression. They applied the method to predict decontrolled electricity data. The empirical results indicated that SVRARIMA model outperformed the other models.

Natanelov *et al.* (2011) applied Johansen co-integration technique for co-integration analysis between agricultural commodity future market price and crude oil future market price. The commodities selected were Cocoa, Coffee, Corn, Soybeans, Wheat and Rice prices. They found that parallel movement between Cocoa, Wheat and Gold pairs and crude oil exist for the last two decades, indicating strong linkages between these markets and crude oil.

Ghosh *et al.* (2011) applied AR-GARCH model in modeling and forecasting of the rainfall data. Their study comprised of finding the formulae and estimate the out-of-sample forecasts and its error variance up to three-steps ahead. The study showed that the Periodic autoregressive (PAR) model with AR-GARCH errors outperformed the Seasonal Autoregressive Integrated Moving Average (SARIMA) model with respect to modeling as well as forecasting performance.

Bildirici and Ersin (2012) analysed the family of regime switching GARCH neural network models, which allowed the generalization of Markov Switching type regime switching GARCH models to MS-GARCH-NN models by incorporating with neural network architectures with different dynamics. They also investigated the forecasting capabilities of these models.

Jha *et al*. (2012) applied multivariate co-integration approach to examine the relationship between energy and agricultural GDP of India from 1980 to 2005. They recognized the direct and robust relationship between both of them using the co-integration analysis.

Rosa and Vasciaveo (2012) tried to analyse the interactions among the prices of selected agricultural commodities in United States and Italy by applying the time series analysis method. The null hypothesis was that the volatility increase in the commodities prices is caused by the crude oil prices. For the empirical analysis, the data of commodity spot price of Wheat, Soybean and Corn in US and Italy and crude oil price were collected. The results suggested: i) there is an influence of the oil price on the selected commodities of the US markets; ii) there is co-integration between US and Italian commodities and thus the distinctive price condition; iii) no clear proof of causality between crude oil and Italian commodities, indicating that the volatility in oil is directly transmitted to the US market only whereas indirectly to the Italian one.

Jha (2013) examined the relationship between energy and agricultural production for some states of India. The study shown that states like Punjab and Haryana having high-productivity use more than seven-times energy as compared to the states like Odisha (4GJ/ha) having low-productivity. The paper also revealed that the energy-intensive inputs was used higher on marginal farms in compare to large farms.

Wang *et al*. (2013) shown that an ARFIMA process subjected to a break in mean and a shift in the long memory parameter can be approximately model by an Autoregressive (AR) model using an information criteria (AIC or Mallows' Cp) to choose the order of the approximated AR model. The perception of theoretical analysis is supported by Monte Carlo experiments, through which they have found that their proposed method gives a substantial improvement over the existing methodology in terms of better out-of-sample forecasting performance. They have also mentioned that the main reason for the success of the proposed forecasting methodology which is ignorance of inaccurate estimation of break dates.

Chaabane (2014) developed a hybrid model combining least square SVM and ARFIMA. He used the hybrid model to forecast the electricity market price. He came to the conclusion of better prediction accuracy of the hybrid model.

Jha and Sinha (2014) compared ANN and ARIMA model for agricultural price series forecasting. The suggested time-delay neural network (TDNN) model as a feed-forward neural network can handle nonlinearity feature of dataset very well. They studied and investigated the TDNN model on monthly wholesale price of oilseeds of India and concluded that this model performed better than the linear models.

Paul (2014) and Paul *et al*. (2014) applied ARFIMA model for the prediction of agricultural commodity price series. They demonstrated that the model performed well in terms of both variability explanation and prediction.

Bildirici and Ersin (2015) assessed a new set of nonlinear models by combining the forecasting abilities of multilayer perceptron (MLP) and radial basis function (RBF) neural networks with GARCH models and at last this model is augmented by LSTAR type nonlinear econometric models given by Terasvirta (1994). The LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF family models aim at modeling both the conditional mean processes and the conditional variance at a time with STAR type nonlinearity, which capture the regimes shifting with logistic transition functions. Therefore, at the first stage, they modelled crude oil prices with GARCH models with fractional integration and asymmetric power terms. Second stage was the introduction of LSTAR to the baseline models. At the third stage, the resulted LSTAR-LST-GARCH models were finally augmented with MLP and RBF neural networks.

Chen *et al.* (2015) developed a general GARCH method to predict volatility using those returns which are sampled at a higher frequency than the horizon of prediction. They called them High Frequency Data-Based Projection-Driven GARCH, or HYBRID-GARCH models. In addition, they examined the theoretical properties and statistical inference of the developed models. Their illustration showed that the developed models were superior in out-of-sample forecasting.

Paul *et al.* (2015a) combined AR and Fractionally Integrated GARCH (FIGARCH) and used them for forecasting lentil's spot price. They suggested through their results that the discussed model can be used as another approach for price series forecasting.

Paul *et al*. (2015b) applied ARFIMA-FIGARCH model for the modeling and then forecasting of volatility with long memory in India's agricultural commodities price series. They showed that there is a clear and strong evidence of occurrence of long memory in lentil's spot price and thus by considering the big role of long memory, the forecasting performance evaluated with less percentage error.

Salisu and Olako (2015) examined three ways to model oil price–US stock nexus. First, they employed the VARMA–AGARCH model introduced by McAleer *et al.* (2009) within the perspective of BEKK framework using Brent and West Texas Intermediate (WTI) and S&P stocks as proxies for oil market and US stock market, respectively. Secondly, they modified the model by including endogenously determined structural break. Third, they used the adopted model to calculate optimal portfolio hedge and weight ratios between oil price and US stocks taking data on the break date. Their empirical evidence suggested a significant return spillover from stock market to oil market and bidirectional shock spillover between the two markets.

Zhang and Goh (2015) developed a hybrid method based on decomposition and ensemble approach for crude oil price forecasting. They employed the ensemble empirical mode decomposition (EEMD) method for decomposition of raw series into many intrinsic mode functions (IMFs) and the residual term. After that, the least square support vector machine combined with the particle swarm optimization (LSSVM–PSO) method and the GARCH model were developed to model and forecast the nonlinear and time-varying components of those IMFs, respectively. At last, each of the forecasted component are added as the final forecasted results.

Lama *et al.* (2016) evaluated the forecasting performance of time-delay neural network and GARCH models for the volatility forecasting using monthly price series of edible oils in domestic and international markets. They also investigated the performance of the combined GARCH-ANN model.

Maciel *et al.* (2016) developed an evolving fuzzy-GARCH modelling technique to forecast asset returns of stock market. The technique aimed to account for volatility from GARCH approach and volatility clustering and identification of nonlinear time series from evolving fuzzy systems.

Chen *et al.* (2017) applied the two deep learning models that are the deep belief network and the recurrent neural network for modelling the dynamics of the crude oil price movement. They constructed a hybrid model by combining the ARMA model forecasts and the deep learning model forecasts.

Gopal and Ramasamy (2017) developed a hybrid model for the prediction of one-day future price for the stocks MSFT, Goldman Sachs, Apple, and JP Morgan. They used the Markov switching model together with radial basis function (RBF) network for prediction. Besides, they also forecasted the buying/selling strategy and this explored the risk of investment and the trading performance.

Jebli and Youssef (2017) examined the effect of co-integration and Granger causality on Tunisian economy. They examined short and long-run relationships among per capita carbon dioxide (CO_2) emissions, renewable and non-renewable energy consumption, real gross domestic product (GDP), agricultural value added (AVA) and trade openness ratio in Tunisia. The Johansen-Juselius test showed bidirectional causalities of long-run between all selected variables.

Kristjanpoller and Hernandez (2017) developed a hybrid neural network model with GARCH-type models, to forecast volatility of the returns from the prices of gold, silver and copper.

Kumar and Jha (2017) applied Johansen co-integration approach in order to examine the co-movement and causality between agricultural commodities price series and energy.

Zhang and Zhang (2017) developed a hybrid forecasting method based on exponential GARCH, the hidden Markov and least squares support vector machine (LSSVM), and the performance is compared with that of GARCH and other forecasting methods. Their results indicated that the proposed hybrid method outperformed others for crude oil price volatility in terms of forecasting accuracy criteria.

Zhao *et al.* (2017) developed an ensemble approach using deep learning. In this approach, bootstrap aggregation (bagging) was used to generate multiple data sets from single data set and then Stacked Denoising Autoencoders (SDAE) were applied on each generated for training and testing. SDAEs is a deep learning tool which was used here to model the complex and nonlinear relationships of crude oil price with its factors.

Chai *et al.* (2018) combined some forecast approaches that capture diverse fluctuation features present in crude oil series, like change points, trend decomposition, regime-switching, time-varying determinants, and the nonlinearity of model setting. At first stage, product partition model-K-means (PPM-KM) model was used to capture change points in the price sequence. Next stage, consisted a time-varying transition probability Markov regime switching (TVTP-MRS) model in order to capture the regime-switching characteristic. Then, Bayesian model averaging (BMA) was used to filtrate important determinants at each regime. Finally, the time series model TVP-STSM was used to decompose the sequence, capture the coefficients in "volatile upward" regime, and predict the oil price.

Johansen and Nielsen (2018) examined the nonstationary co-integration in a time series data. They revealed that likelihood function of the fractional co-integrated vector autoregressive (CVAR) model has some asymptotic properties.

Gaetano (2018) developed three combination schemes that consider structural breaks in the variance of a GARCH (1, 1) model. They were obtained by averaging forecasts of different estimation obtained on rolling on fixed window. The first scheme used the equal weights to average the individual forecasts; the second scheme simply allotted heavier weights to forecasts using latest information and the third trimmed a fixed fraction of the lowest and highest individual forecasts.

Khashei and Hajirahimi (2018) combined the Multilayer Perceptrons (MLPs) and the ARIMA methodology and developed many different hybrid models. They selected a time series comprising of both linear as well as nonlinear components for fitting the model. Their conclusion was that ANN-ARIMA model had superior performance in forecasting accuracy.

Choudhury *et al.* (2019) developed a hybrid model by combining EMD and ANN for price index of potato in India. Their empirical results suggested that the EMD-ANN model is the best methodology for price forecasting.

David *et al.* (2019) studied the co-integration between ethanol and some agricultural commodity price series. They also showed that how this relationship can affect the prediction ability and the efficiency of the co-integrated price series.

Garcia and Kristjanpoller (2019) examined the performance of a set of time series models such as ARIMA and GARCH, Artificial Neural Networks (ANN) as well as Fuzzy Inference Systems (FIS) and hybrid specifications of both. Genetic Algorithm was used to see the adaptability characteristic of these models in exogenous variables, configuration parameters and window size at the same time. They also examined the out-of-sample forecasting performance on the basis of Heteroskedasticity-adjusted Mean Squared Error (HMSE).

Lin *et al.* (2020) proposed a hybrid forecast model to forecast crude oil price on considering the long memory, heavy-tail distribution, asymmetric, nonlinear and non-stationary characteristics of crude oil price. They showed empirically that the proposed hybrid model WPD–EMD–ARMA–FIGARCH-M achieves significant effect during periods of extreme incidents.

Torre-Torres *et al.* (2020) used Markov-Switching GARCH (MS-GARCH) models in an active trading algorithm for corn and soybean future markets. Their results suggested that the Gaussian Markov-Switching GARCH model is the most appropriate to generate extra returns (from a passive investment strategy) in the corn market and the t-Student Markov-Switching GARCH is the best one for soybean trading.

CHAPTER III

MATERIALS AND METHODS

In this chapter, various time series models relevant to the present study are described in detail. The chapter starts with an overview of the data sets used in the study. In the subsequent sections, data diagnostic tests, structural break, GARCH type models, long memory models and co-integration analysis are described in detail. A short description of model evaluation techniques are given in the last section.

3.1 Data Sets

Under the three objectives of the present study, we considered different types of situation for the modelling and forecasting of agricultural price in the presence of structural break(s). Accordingly, we used three different data sets to accomplish three objectives. In the first objective, to evaluate the forecasting performance of the proposed time series volatility model, we used three different agricultural commodity price series viz. weekly Potato price of Delhi market, international monthly Groundnut oil price and international monthly Palm oil price. Potato price (₹/q) data of Delhi market are obtained from website of National Horticultural Research and Development Foundation (NHRDF) (http://nhrdf.org/en-us/). In the website of NHRDF, actually daily data are available, so we collected daily data for the period 1st January 2005 to 31st December 2019, and further converted it into weekly data. The monthly price ($/mt) series of international Groundnut oil and Palm oil are obtained from World Bank Pink Sheet (https://www.worldbank.org/en/research/ commodity-markets) for the period January 1980 to December 2019.

In the second objective, to assess the performance of long memory models in the presence of structural break(s), we used daily Mustard price (₹/q) data of Agra and Bharatpur market. Generally, long memory property are found in daily price data and it is very difficult to find this property in weekly or monthly agricultural price series. Daily Mustard price series of Agra and Bharatpur market are obtained from Agricultural

Marketing Information Network (AGMARKNET), (https://agmarknet.gov.in/) website, for the period 1st January 2016 to 31st January 2020.

For the investigation of co-integrated time series in the presence of structural break as in the final objective of the study, we have taken major four Potato markets of India viz. Agra, Bangalore, Delhi and Mumbai. The monthly Potato price data of these markets were collected from National Horticultural Research and Development Foundation (NHRDF) (http://nhrdf.org/en-us/) website for the period January 2005 to December 2019.

3.2 Data Diagnostic Tests

3.2.1 Stationarity and Unit Root Tests

A stationary time series is one whose properties do not depend on the time at which the series is observed. Therefore, time series with trends, or with seasonality, are not stationary *i.e.* the trends and seasonality will influence the time series value at different times. In other words, a stochastic process is stationary if the mean and variance are constant over time and covariance between two time points depends only on the distance of the lags between the two time periods and not on the actual time that the covariances are computed. The effect of an exogenous shock on a stationary series is relatively short-lived, meaning that we would expect spikes in the series to be followed by the series quickly reverting back to an equilibrium value. If the exogenous shocks were short-lived, we would expect the series to move up or down, but to revert quickly to a mean value. In literature, to test the stationarity of the data, unit root tests like Dickey-Fuller (DF) test, Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, KPSS test (Kwiatkowski, Perron, Schmidt, and Shin, 1992), *etc.* has been used. In this study, Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests have been used to test the stationarity of the data.

(A) Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) Test

Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) test are two most common tests for a unit root process. For a time series $X_t, (t = 1, \dots, T)$, an autoregressive process can be written as

$$X_t = \rho X_{t-1} + \epsilon_t \quad (3.2.1)$$

Now, rewrite the equation 3.2.1 by subtracting X_{t-1} from each side of the equation

$$\Delta X_t = \gamma X_{t-1} + \epsilon_t \quad (3.2.2)$$

where $\gamma = (\rho - 1)$. Now, testing $(\rho = 1)$ is the same as testing for $(\gamma = 0)$.

There are three versions of the Dickey-Fuller test, each with their own set of critical value *t*-test tables. The following regressions may also be used to test for a unit root:

$$\Delta X_t = \mu + \gamma X_{t-1} + \epsilon_t \quad (3.2.3)$$

$$\Delta X_t = \mu + \gamma X_{t-1} + \beta_1 t + \epsilon_t \quad (3.2.4)$$

The equation 3.2.2 is a pure random walk model, the equation 3.2.3 is a random walk with a drift, and the equation 3.2.4 is a random walk with a drift and a deterministic time trend.

Augmented Dickey-Fuller test is similar to the Dickey-Fuller test but allows for additional autoregressive terms, while taking those terms into consideration for unit root testing. The Dickey-Fuller test uses ordinary least squares (OLS) to generate γ, it assumes ϵ_t is independent and identically distributed, but if a higher order autoregressive process exists, autocorrelation will remain in the residuals and this violate OLS assumption, leading to poor estimates of γ. The ADF test allows for heterogeneity and serial correlation in the errors. The ADF test for the general p^{th} order autoregressive process is

$$\Delta X_t = \mu + \gamma X_{t-1} + \sum_{i=2}^{p} \zeta_i \Delta X_{t-i+1} + \epsilon_t \quad (3.2.5)$$

where $\gamma = -(1 - \sum_{i=1}^{p} \beta_i)$ and $\zeta_i = -\sum_{j=1}^{p} \beta_j$

In Augmented Dickey-Fuller test, the null hypothesis of unit root $(\gamma = 0)$ is tested against the alternative of stationarity.

(B) Phillips-Perron (PP) Test

Phillips and Perron (1988) introduced the non-parametric test for the null hypothesis of a unit root that explicitly allows for weak dependence and heterogeneity of the error

process, which is popularly known as Phillips-Perron test (PP test). They consider the following two test regressions:

$$X_t = \mu + \rho X_{t-1} + \epsilon_t \tag{3.2.6}$$

$$X_t = \mu + \rho X_{t-1} + \beta\left(t - \frac{1}{2}T\right) + \epsilon_t \tag{3.2.7}$$

For equation 3.2.6, Phillips and Perron (1988) defined following test statistics:

$$Z(\hat{\rho}) = T(\hat{\rho} - 1) - \hat{\lambda}/\bar{m}_{XX} \tag{3.2.8}$$

$$Z(\tau_{\hat{\rho}}) = \left(\frac{\hat{s}}{\hat{\sigma}_{Tl}}\right) t_{\hat{\rho}} - \hat{\lambda}'\hat{\sigma}_{Tl}/\bar{m}_{XX}^{\frac{1}{2}} \tag{3.2.9}$$

$$Z(\tau_{\hat{\mu}}) = \left(\frac{\hat{s}}{\hat{\sigma}_{Tl}}\right) t_{\hat{\mu}} - \hat{\lambda}'\hat{\sigma}_{Tl} m_X/\bar{m}_{XX}^{\frac{1}{2}} m_{XX}^{\frac{1}{2}} \tag{3.2.10}$$

$\bar{m}_{XX} = T^{-2}\sum(X_t - \bar{X})^2$, $m_{XX} = T^{-2}\sum X_t^2$, $m_X = T^{-3/2}\sum X_t$ and $\hat{\lambda} = \frac{1}{2}(\hat{\sigma}_{Tl}^2 - \hat{s}^2)$, where $\hat{s}^2$ is the residuals sample variance, $\hat{\lambda}' = \hat{\lambda}/\hat{\sigma}_{Tl}^2$ and $t_{\hat{\rho}}$, $t_{\hat{\mu}}$ are *t* ratios of $\hat{\rho}$ and $\hat{\mu}$, respectively. The long-run variance $\hat{\sigma}_{Tl}^2$ is estimated as

$$\hat{\sigma}_{Tl}^2 = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2 + 2T^{-1}\sum_{s=1}^{l} w_{sl}\sum_{t=s+1}^{T}\hat{\epsilon}_t\hat{\epsilon}_{t-s} \tag{3.2.11}$$

where $w_{sl} = 1 - s/(l+1)$.

Similarly, to test regression with a linear time trend included as in equation 3.2.7, the following test statistics are defined:

$$Z(\tilde{\rho}) = T(\tilde{\rho} - 1) - \tilde{\lambda}/M \tag{3.2.12}$$

$$Z(t_{\tilde{\rho}}) = \left(\frac{\tilde{s}}{\tilde{\sigma}_{Tl}}\right) t_{\tilde{\rho}} - \tilde{\lambda}'\tilde{\sigma}_{Tl}/M^{\frac{1}{2}} \tag{3.2.13}$$

$$Z(t_{\tilde{\mu}}) = \left(\frac{\tilde{s}}{\tilde{\sigma}_{Tl}}\right) t_{\tilde{\mu}} - \tilde{\lambda}'\tilde{\sigma}_{Tl} m_X/M^{\frac{1}{2}}(M + m_X^2)^{\frac{1}{2}} \tag{3.2.14}$$

$$Z(t_{\tilde{\beta}}) = \left(\frac{\tilde{s}}{\tilde{\sigma}_{Tl}}\right) t_{\tilde{\beta}} - \tilde{\lambda}'\tilde{\sigma}_{Tl}\left(\frac{1}{2}m_X - m_{tX}\right)/(M/12)^{\frac{1}{2}}\bar{m}_{XX}^{\frac{1}{2}} \tag{3.2.15}$$

where $m_{tX} = T^{-5/2}\sum tX_t$

The critical values of these Z statistics are similar to those of the Dickey-Fuller type tests.

3.2.2 Test of Linearity

Under the null hypothesis of linearity, the residuals should be independent for a correctly specified linear model. Any deviation from independence in the residuals indicates inadequacy of the model, including the linearity assumption. This is the basic idea behind various nonlinearity tests. To test the linearity of the time series data, nonparametric methods like Ljung–Box statistics of squared residuals, the bispectral test, Brock, Dechert, and Scheinkman (BDS) test, *etc.* and parametric methods like RESET test (Ramsey, 1969), F tests of Tsay (1989), *etc*. have been proposed. In this study, BDS test has been used to test the linearity of agricultural price data.

BDS Test

Brock, Dechert, and Scheinkman (1987) proposed a test statistic, commonly known as BDS test, to detect the independent and identically distributed (*i.i.d.*) assumption of a time series. The basic idea of the BDS test is to make use of a "correlation integral" popular in chaotic time series analysis. Given a *k*-dimensional time series X_t and observations $\{X_t\}_{t=1}^{T_k}$, define the correlation integral as

$$C_k(\delta) = \lim_{T_k \to \infty} \frac{2}{T_k(T_k-1)} \sum_{i<j} I_\delta(X_i, X_j) \tag{3.2.16}$$

where $I_\delta(X_i, X_j)$ is an indicator variable, which is equal to one if $\|X_i - X_j\| < \delta$, and zero otherwise, where $\|\cdot\|$ is the supnorm. The correlation integral measures the fraction of data pairs of $\{X_t\}$ that are within a distance of δ from each other. Consider next a time series x_t. Construct *k*-dimensional vectors $X_t^k = (x_t, x_{t+1}, \ldots, x_{t+k-1})'$, which are called *k* histories. The idea of the BDS test is to treat a *k* history as a point in the *k*-dimensional space. If $\{X_t\}_t^T$ are indeed *i.i.d.* random variables, then the *k*-histories $\{X_t\}_t^{T_k}$ should show no pattern in the *k*-dimensional space. Consequently, the correlation integrals should satisfy the relation $C_k(\delta) = [C_1(\delta)]^k$. Any departure from the prior relation suggests that x_t are not *i.i.d.*

Now define

$$C_l(\delta,T) = \frac{2}{T_k(T_k-1)}\sum_{i<j} I_\delta(X_i^*, X_j^*), l = 1, k, \quad (3.2.17)$$

where $T_l = T - l + 1$ and $X_i^* = x_i \;\; if \; l = 1$ and $X_i^* = X_i^k \;\; if \; l = k$.

Under the null hypothesis that $\{x_i\}$ are *i.i.d.* with a nondegenerated distribution function $F(\cdot)$, Brock, Dechert, and Scheinkman (1987) show that $C_k(\delta,T) \to [C_1(\delta)]^k$ with probability 1; as $T \to \infty$ for any fixed k and δ. Furthermore, the statistic $\sqrt{T}\{C_k(\delta,T) - [C_1(\delta,T)]^k\}$ is asymptotically distributed as normal with mean zero and variance:

$$\sigma_k^2(\delta) = 4\left(N^k + 2\sum_{j=1}^{k-1} N^{k-j}C^{2j} + (k-1)^2C^{2k} - k^2NC^{2k-2}\right) \quad (3.2.18)$$

where $C = \int [F(z+\delta) - F(z-\delta)]dF(z)$ and $N = \int [F(z+\delta) - F(z-\delta)]^2 dF(z)$. $C_1(\delta,T)$ is a consistent estimate of C, and N can be consistently estimated by

$$N(\delta,T) = \frac{6}{T_k(T_k-1)(T_k-2)}\sum_{t<s<u} I_\delta(x_t,x_s)I_\delta(x_s,x_u) \quad (3.2.19)$$

The BDS test statistic is then defined as

$$D_k(\delta,T) = \frac{\sqrt{T}\{C_k(\delta,T) - [C_1(\delta,T)]^k\}}{\sigma_k(\delta,T)} \quad (3.2.20)$$

where $\sigma_k(\delta,T)$ is obtained from $\sigma_k(\delta)$ when C and N are replaced by $C_1(\delta,T)$ and $N(\delta,T)$, respectively.

3.2.3 ARCH-LM Test

ARCH-LM test is used to test for the presence of heteroscedasticity in the squared residuals. Let $\epsilon_t = X_t - \alpha X_{t-1}$ be the residual series of an autoregressive process. To check for conditional heteroscedasticity, the squared residual series ϵ_t^2 is used, which is also known as the ARCH effects. Two tests are available for it. In the first test, apply the usual Ljung–Box statistics $Q(m)$ to the ϵ_t^2 series. The null hypothesis of the test statistic is that the first *m* lags of ACF of the ϵ_t^2 series are zero. The second test for conditional heteroscedasticity is the Lagrange multiplier test of Engle (1982). This test is equivalent to the usual *F* statistic for testing $\beta_i = 0, (i = 1, \dots m)$ in the linear regression

$$\epsilon_t^2 = \beta_0 + \beta_1\epsilon_{t-1}^2 + \cdots + \beta_m\epsilon_{t-m}^2 + \varepsilon_t, \;\; t = m+1, \dots, T \quad (3.2.21)$$

where ε_t denotes the error term, *m* is a pre-specified positive integer, and *T* is the sample size.

Let $SSR_0 = \sum_{t=m+1}^{T}(\epsilon_t^2 - \bar{\omega})^2$, where $\bar{\omega} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$ is the sample mean of ϵ_t^2, and $SSR_1 = \sum_{t=m+1}^{T}\hat{\varepsilon}_t^2$ where $\hat{\varepsilon}_t$ is the least squares residual of the prior linear regression. Then under H_0

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-2m-1)} \tag{3.2.22}$$

which follows an *F* distribution with degrees of freedom *m* and $T - 2m - 1$.

3.3 Detection of Structural Break(s)

A structural break(s) occurs when an entire series is no longer characterized by the same underlying process, and instead, there are two (or more) distinct subsamples of the observations, each of which is characterized by a different underlying process. These breaks can result either from an observed event, or from an unobserved combination of factors. The presence of structural break(s) has long been conjectured in the agricultural price data. The trigger of such structural breaks may involve with the implementation of new economic policy, introduction of new varieties, *etc*. Induced by these major factors, the "shocks" may trigger an abrupt shift in the mean or variance structure, which cause parameters inconsistency of the model. Ignoring such factors may lead to biased forecast. To test the structural break(s) in the time series data several test have been proposed, of which few tests are described in the following subsections:

3.3.1 Bai and Perron (2003) Test

Consider a system of linear regression equations, for a data series $X_t, t = 1, \dots, T,$ in which a set of segments determined by the locations of potential structural breaks, namely, $[T_1, \dots T_m]$ with *m* being the number of potential breaks:

$$X_t = \varphi_t'\beta + \phi_t'\delta_1' + \epsilon_t, \qquad t = 1, \dots, T_1$$

$$X_t = \varphi_t'\beta + \phi_t'\delta_2' + \epsilon_t, \qquad t = T_1 + 1, \dots, T_2$$

$$\vdots$$

$$X_t = \varphi_t'\beta + \phi_t'\delta_{m+1}' + \epsilon_t, \qquad t = T_m + 1, \dots, T$$

where X_t represents a segment of observations in each equation from the total series $X_t, t = 1, \dots, T$. φ_t' and ϕ_t' are two vectors of covariates, with dimensions of $p \times 1$ and $q \times 1$ respectively. The former has a dimension of $p \times 1$, indicating the start of one segment; while the latter is of $q \times 1$, $q = 1, \dots, T$, and it indicates the end of that segment. The coefficients β and $\delta_i', i = 1, \dots, m + 1$ are obtained by minimizing the sum of squared residuals $\sum_{i=1}^{m+1} \sum_{T_i}^{t=T_{i-1}+1} [X_t - \varphi_t'\beta - \phi_t'\delta_1']$. The estimates of the position of structural breaks are $[\bar{T}_1, \dots, \bar{T}_m]$, which makes the smallest sum of these minimized sum of squared residuals obtained in each segment.

Bai and Perron (2003) developed different methods to test each partition choice to find the breaks. One is through a Supremum F-test with the null hypothesis of no break versus the alternative of the presence of break with a number of a fixed positive finite integer. The number of breaks in this case needs to be identified in advance. Another approach is through a double maximum procedure, with the null hypothesis of no break and the alternative of an unspecified number of breaks. This approach consists of two tests, one is called the *U D max* test with all weights equal to unity, and the other is the *W D max* test with varying weights. The last approach is known as the sequential test, with the null hypothesis of *m* break(s) against *m+1* break(s).

3.3.2 Lagrange Multiplier Test of Andrews (1993)

Andrews (1993) proposed a structural break test based on the Lagrange Multiplier (LM) to locate a one-time unknown change point in non-linear parametric models. Consider an econometric model that fits to time series $X_t, t = 1, \dots, T$ with parameter vector $\boldsymbol{\varphi_t}$; define ω as the location of a potential break near the known events with $\omega \in (0,1)$; take $[\omega T]$, where $[\cdot]$ is the integer part operator, as the proportion of sample observations before the break occurs at the $[\omega T]^{\text{th}}$ observation. In this way, the model parameters before and after the break then become φ_1 for $t = 1, \dots, [\omega T]$ and φ_2 for $t = [\omega T] + 1, \dots, T$ respectively. Thus the null hypothesis of no structural break with alternative being the presence of such at $[\omega T]$ in the parameter are formulated as below:

$$H_0 = \boldsymbol{\varphi_t} = \boldsymbol{\varphi_0}$$

and

$$H_1: \boldsymbol{\varphi_t} = \begin{cases} \varphi_1(\omega), & for\ t = 1, \dots, [\omega T] \\ \varphi_2(\omega), & for\ t = [\omega T] + 1, \dots, T \end{cases}$$

In particular, for a normal linear regression model, if the location of structural break is known, the LM test that is constructed under the above hypotheses is equivalent to F test, which is also referred to as Chow test (Chow, 1960) in the literature.

Moreover, under the null hypothesis of no structural break, there is only one set of the parameter vector and it can be estimated via maximum likelihood; when there is one structural break, in other words, H_1 is true, and the location of it is known at $[\omega T]$ in a non-linear model, then the LM test statistic $LM(\omega)$ is calculated as below:

$$LM(\omega) = \frac{T}{\omega(1-\omega)} \overline{\boldsymbol{g}}_{\boldsymbol{1T}}(\widehat{\boldsymbol{\varphi}}, \omega)' \boldsymbol{S_T^{-1} D_T} \left(\boldsymbol{D_T' S_T^{-1} D_T} \right)^{-1} \boldsymbol{D_T' S_T^{-1}} \overline{\boldsymbol{g}}_{\boldsymbol{1T}}(\widehat{\boldsymbol{\varphi}}, \omega) \qquad (3.3.1)$$

where

$$\overline{\boldsymbol{g}}_{\boldsymbol{1T}}(\widehat{\boldsymbol{\varphi}}, \omega) = \frac{1}{T} \sum_{t=1}^{\omega T} g(X_t; \widehat{\boldsymbol{\varphi}}) \qquad (3.3.2)$$

$$\boldsymbol{S_T} = \frac{1}{T} \sum_{t=1}^{\omega T} [g(X_t; \widehat{\boldsymbol{\varphi}}) - \overline{\boldsymbol{g}}_{\boldsymbol{T}}(\widehat{\boldsymbol{\varphi}})] \, [g(X_t; \widehat{\boldsymbol{\varphi}}) - \overline{\boldsymbol{g}}_{\boldsymbol{T}}(\widehat{\boldsymbol{\varphi}})]' \qquad (3.3.3)$$

$$\boldsymbol{D_T} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial (X_t; \varphi)}{\partial \widehat{\varphi}'} \qquad (3.3.4)$$

In equation 3.3.2, $g(X_t; \widehat{\boldsymbol{\varphi}}) = \partial log f(X_t; \widehat{\boldsymbol{\varphi}}) / \partial \widehat{\boldsymbol{\varphi}}$ is the score in terms of the partial derivative of the log density with respect to the parameter vector $\widehat{\boldsymbol{\varphi}}$. $\boldsymbol{D_T}$ in equation 3.3.4 is the restricted estimator that is used to construct weight matrices for LM test statistics. The $LM(\omega)$ statistic asymptotically follows a chi-squared distribution with degree of freedom equals to the number of parameters in the model.

3.3.3 Iterative Cumulative Sum of Squares of Inclan and Tiao (1994)

Consider a series of uncorrelated random variables $X_t, t = 1, \dots, T$ with mean 0 and variance $\sigma_t^2, t = 1, \dots, T$. Define the following expression:

$$D_k = C_k / C_T - k/T\,, \ \ k = 0,1,2, \dots, T \qquad (3.3.5)$$

D_k is the centred and normalized cumulative sum of squares; $C_{k+1} = \sum_{i=1}^{k} X_t^2,\ k = 0,1,2,\dots,T$ is the cumulative sum of $k+1$ squares of the data observations. The underlying concept of this Iterated Cumulative Sums of Squares (ICSS) algorithm of Inclan and Tiao (1994) is to assume the variability of $\sigma_t^2, t = 1,\dots,T$ is made of constant σ_t^2 at different time periods over the whole sample period *T*. In other words, the variance stays constant for some time, until it takes up a new value at k^*; the variance then stays at this new value for some other time until another variance value occurs. It is then said that one structural break has occurred at k^*. Within this context, the construction of D_k in expression 3.3.5 will oscillate around zero until the occurrence of a structural break, where D_k varies away distinguishably from zero. Therefore, an *IT* test is developed to find the variation of D_k that is statistically significant, which takes the form as below:

$$IT = max\sqrt{T/2}\,|D_k| \tag{3.3.6}$$

where $\sqrt{T/2}$ is to standardize the distribution. Under a null hypothesis of no structural break against the alternative of presence of one break, when *IT* exceeds the critical value at a selected confidence level, one structural break is detected in the variance or volatility of this data series. In order to find multiple unknown breaks in the whole series, an iterative scheme is specifically designed to systematically search for change points by applying the *IT* test to sub-samples created consecutively after a possible change point is identified.

3.3.4 Pruned Exact Linear Time (PELT) Algorithm

The PELT algorithm (Killick *et al.*, 2012) uses a standard approach of detecting structural breaks through minimization of costs. In order to find multiple change points in the data, the PELT algorithm is first applied to the whole data series and iteratively and independently to each partition until no further structural breaks are detected. The main assumption of the PELT algorithm is that the numbers of structural breaks increases linearly with the increase of data set, *i.e.* the breaks are spread throughout the data and are not restricted to particular one portion of the data. It is based on the algorithm of Jackson, *et al.* (2005), which has aims to minimize:

$\sum_{i=1}^{m+1}[C(X_{\tau_{i-1}+1}, \dots, X_{\tau_i}) + \beta]$, where C denotes the cost function for the i^{th} segment and β is a penalty to guard against the over fitting.

In the PELT algorithm, optimal partitioning method of Jackson, *et al.* (2005) is modified by pruning. This algorithm combines the optimal partitioning and pruning to achieve exact and efficient computational cost. The optimal segmentation,

$$F(n) = \min_{\tau}\{\sum_{i=1}^{m+1}[C(X_{\tau_{i-1}+1}, \dots, X_{\tau_i}) + \beta]\} \quad (3.3.7)$$

Conditioning on the last break point, τ_m and calculating the optimal segmentation of the data up to that breakpoint gives,

$$F(n) = \min_{\tau_m}\left\{\min_{\tau/\tau_m}\sum_{i=1}^{m+1}[C(X_{\tau_{i-1}+1}, \dots, X_{\tau_i}) + \beta] + C(X_{\tau_m+1}, \dots, X_n)\right\} \quad (3.3.8)$$

This could equally be repeated for the second to last, third to last and so on, break points.

3.4 Volatility Models with Structural Break(s)

3.4.1 Auto-regressive Integrated Moving Average (ARIMA) Model

In some applications, the Auto-regressive (AR) or Moving Average (MA) models become cumbersome because one may need a higher order model with many parameters to adequately describe the dynamic structure of the data. To overcome this difficulty, the Auto-regressive Moving Average (ARMA) models are introduced. Basically, an ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization.

A stationary time series X_t follows an ARMA(1,1) process if it satisfies:

$$X_t - \phi X_{t-1} = \phi_0 + \epsilon_t - \theta_1\epsilon_{t-1} \quad (3.4.1)$$

where ϵ_t is a white noise series. The left-hand side of the equation 3.4.1 is the autoregressive component of the model and the right-hand side gives the moving average component.

If the ARMA model is extended by allowing the AR polynomial to have 1 as a characteristic root, the model will then become the well known Autoregressive

Integrated Moving Average (ARIMA) model. An ARIMA model is said to be unit root nonstationary because its AR polynomial has a unit root. In other words, a time series X_t is said to be an ARIMA(p, 1, q) process if the series $Z_t = X_t - X_{t-1} = (1 - B)X_t$ follows a stationary and invertible ARMA(p, q) model. The price series are generally believed to be nonstationary, but the return series, $r_t = ln(X_t) - ln(X_{t-1})$, is stationary. In this case, the log price series is unit root nonstationary and hence can be treated as an ARIMA process.

Box–Jenkins Methodology

The aim of this methodology is to find the most appropriate ARIMA(p, d, q) model and to use it for forecasting. It uses an iterative six-stage scheme:

(I) Identification of the differentiation order *d*

The first step for the fitting of ARIMA model is to check the stationarity of the series. ADF test and PP test are performed to see the presence of unit root in the data series.

(II) Identification of the order *p* and *q*

If data series is stationary then we plot Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Here ACF plot helps to decide the parameter q and PACF plot helps to decide the parameter p.

(iii) Estimation of the parameters

When we write our ARMA(p, q) or ARIMA(p, d, q) model after testing stationary of data series, we have different $\phi_i (i = 1,2, \dots, p)$, $\theta_j (j = 1,2, \dots, q)$ model's parameters are unknown . So our next work is to estimate these unknown parameters. There are different technique to estimate these parameters like Ordinary Least Square (OLS) technique, likelihood method, *etc.*

(IV) Validation (Residual diagnostic)

In the validation step, we examine the residuals. For a satisfactory model, the residuals should be white noise. The correlograms are evaluated, and portmanteau tests are used to assess if the residuals are near enough to white noise.

(V) Choice of a model

If many models pass the validation test, selection criteria may be used, the most common being the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The predictive properties of the models can be considered in complementing these criteria. Hence the theory of parsimony would lead one to select the simplest model, the one with the fewest parameters.

(VI) Forecasting

After considering all the above steps, we find a best suited model for our time series data and it can be used for forecasting purpose.

3.4.2 Generalized Auto-regressive Conditional Heteroscedasticity (GARCH) Model

Generalized Auto-regressive Conditional Heteroscedasticity (GARCH) models have been widely applied to model the volatility of agricultural commodity prices. Introduced by Bollerslev (1986), GARCH provides a more flexible lag structure of the ARCH (Engle, 1982) type models. For the return series r_t, the standard GARCH (p, q) can be written as:

$$r_t = \mu + \epsilon_t, \;\; \epsilon_t | I_{t-1} \sim N(0, \sigma_t^2) \tag{3.4.2}$$

$$\epsilon_t = z_t \sqrt{\sigma_t^2}, \quad z_t \sim N(0,1) \tag{3.4.3}$$

$$\sigma_t^2 = \omega + \sum_{t=1}^{q} \alpha_t \epsilon_{t-1}^2 + \sum_{t=1}^{p} \beta_t \sigma_{t-1}^2 \tag{3.4.4}$$

σ_t^2 is the conditional variance, with $\omega > 0$, $\alpha_t > 0$ and $\beta_t > 0$ to ensure the positivity of σ_t^2. z_t is an independent and identically distributed error term with zero mean and unit variance. When $p = 0$, then equation 3.4.4 becomes an ARCH (q) process. The volatility persistence is measured by the sum of α_t and β_t; the more it approaches unity the greater the persistence of shocks to the volatility. Moreover, GARCH (1,1) is recognised to perform quite sufficiently when forecasting volatility. (Hansen and Lunde, 2005). Consider the GARCH (1,1) process, the equation 3.4.4 can be rewrite as:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \alpha \sigma_{t-1}^2 - \alpha \sigma_{t-1}^2 \tag{3.4.5}$$

$$\sigma_t^2 = \omega + \alpha (\epsilon_{t-1}^2 - \sigma_{t-1}^2) + (\alpha + \beta) \sigma_{t-1}^2 \tag{3.4.6}$$

Let $\upsilon_{t-1} = \epsilon_{t-1}^2 - \sigma_{t-1}^2$ represent the shock, and $\lambda = \alpha + \beta$, then

$$\sigma_t^2 = \omega + \alpha\upsilon_{t-1} + \lambda\sigma_{t-1}^2 \tag{3.4.7}$$

Continue decomposing σ_{t-1}^2 in the form shown in equation 3.4.7 and then replace σ_{t-1}^2 in equation 3.4.7 with its new form in the following manner:

$$\sigma_{t-1}^2 = \omega(1 + \lambda + \lambda^2 + \lambda^3 + \cdots) + \alpha(\upsilon_{t-1} + \lambda\upsilon_{t-2} + \lambda^2\upsilon_{t-3} + \cdots) \tag{3.4.8}$$

$$\sigma_{t-1}^2 = \omega\frac{1-\lambda^t}{1-\lambda} + \alpha(\upsilon_{t-1} + \lambda\upsilon_{t-2} + \lambda^2\upsilon_{t-3} + \cdots) \tag{3.4.9}$$

In equation 3.4.9, $\omega\frac{1-\lambda^t}{1-\lambda} = \frac{\omega}{1-\lambda}$ is the unconditional variance when $\lambda < 1$. When $\lambda = 1$, the variance contains a unit root, thus the process has no unconditional variance and is defined as an integrated GARCH or I-GARCH process (Engle and Bollerslev, 1986). Furthermore, it can also be noticed from equation 3.4.9 that, the effect imposed by the shock on the conditional variance σ_t^2 relies on the degree of λ; that is to say, the larger the sum of α and β, the longer the shock lasts, *i.e.* the more persistent of the volatility. Therefore, the persistence of volatility for a certain shock is calculated by the sum of α and β. Lamoureux and Lastrapes (1990) pointed out that high levels of volatility persistence may be spurious if there are structural breaks or regime shifts in the volatility process. More explicitly, this biased persistence of volatility indicates that the current information will still impose significant impacts on the conditional variance forecast for all horizons because of the very close to permanent influence on volatility. Lamoureux and Lastrapes (1990) demonstrate this structural break in volatility process by introducing deterministic shifts in the variance and find that this results to a marked reduction in the degree of volatility persistence relative to standard GARCH models. They suggested that to obtain more robust estimates of conditional volatility would require a more general class of GARCH models that allows for regime shifts as part of the data generating process.

3.4.3 Markov Switching GARCH Model

Hamilton and Susmel (1994), Cai (1994), and Gray (1996) introduced the Markov Switching GARCH models. These models have many common features. First, it allows for the conditional variance process to switch stochastically between a finite number of

regimes. Second, the timing of regime switch is usually assumed to be governed by a first-order Markov process. The transition probability of the Markov process determines the probability that volatility will switch to another regime, and thus the expected duration of each regime. Transition probabilities may be constant or a time-varying function of exogenous variables.

For a return series r_t, consider a model:

$$r_{it} = \mu_{it} + \epsilon_{it} \tag{3.4.10}$$

$$\epsilon_{it} \mid \Phi_{t-1} \sim N(0, \sigma_{it}^2), \quad i = 0,1 \ states, \tag{3.4.11}$$

where μ_{it} and σ_{it}^2 are the conditional mean and conditional variance, respectively and both of which are allowed to switch between two regimes. Φ_{t-1} denote the information up to time $(t-1)$ and $\sigma_{i,t}^2 = var(\epsilon_t \mid s_t = i, \Phi_{t-1})$. Cai (1994) considers an ARCH(q) model:

$$\sigma_{i,t}^2 = \omega_0 + \omega_1 i + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2, \ i = 0,1 \tag{3.4.12}$$

Hamilton and Susmel (1994) proposed the switching regime ARCH (SWARCH(q)) model:

$$\sigma_{i,t}^2 = \lambda_i \eta_t = \lambda_i \left(c + \sum_{j=1}^{q} \alpha_j \varsigma_{t-j}^2\right), \quad i = 0,1 \tag{3.4.13}$$

The conditional variances of equation 3.4.12 have left shifts, but those of equation 3.4.13 have different scales. Cai (1994) and Hamilton and Susmel (1994) models are very special forms of regime switching conditional variances.

However it is not straightforward to expand the above two models to account for lagged conditional variances. It is observed that the conditional variance $\sigma_{i,t}^2$ depends on $\sigma_{i,t-1}^2$, it is determined not only by s_t but also by s_{t-1} due to the presence of $\sigma_{i,t-1}^2$. The dependence of $\sigma_{i,t-1}^2$ on $\sigma_{i,t-2}^2$ then implies that $\sigma_{i,t}^2$ must also be affected by the value of s_{t-2}, and so on. Accordingly, the conditional variance at time *t* is determined by the realization of $(s_t, s_{t-1}, \dots, s_1)$ which has 2^t possible values. This property of "path dependency" will result in a very complex model and make intractable estimation of the

model. Gray (1996) circumvents this problem by postulating that $\sigma_{i,t}^2$ depends on $\sigma_t^2 = IE(\epsilon_t^2|\Phi_{t-1})$, the sum of $\sigma_{i,t}^2$ weighted by the prediction probability IP$(s_t = i \mid \Phi_{t-1})$.

$$\sigma_{i,t}^2 = \omega_i + \sum_{j=1}^{q} \alpha_{i,j}\epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_{i,j}\,\sigma_{t-j}^2, \;\; i = 0,1 \tag{3.4.14}$$

$$\sigma_t^2 = \sigma_{0,t}^2 \text{IP}(s_t = 0 \mid \Phi_{t-1}) + \sigma_{1,t}^2 \text{IP}(s_t = 1 \mid \Phi_{t-1}) \tag{3.4.15}$$

It is important to note from equation 3.4.14 that $\sigma_{i,t}^2$ are no longer path dependent because both $\sigma_{0,t-j}^2$ and $\sigma_{1,t-j}^2$ have been used to from σ_{t-j}^2. So the conditional variance can be determined without taking into account all possible values of $(s_t, s_{t-1}, \dots, s_1)$. Compare to the Cai (1994) and Hamilton and Susmel (1994) models, the switching GARCH model of Gray (1996) does not impose any constraint on these parameters and allows all the GARCH parameters to switch.

Regime switching is assumed to be driven by a first-order Markov process with transition probability:

$$Pr[s_t = 0 \mid s_t = 0] = P$$

$$Pr[s_t = 1 \mid s_t = 0] = 1 - P$$

$$Pr[s_t = 1 \mid s_t = 1] = Q$$

$$Pr[s_t = 0 \mid s_t = 1] = 1 - Q$$

3.4.4 Artificial Neural Network (ANN)

Unlike traditional forecasting approaches, ANN are able to adapt nonlinearity and approximate complex relationships without extensive data or knowledge. ANN is loosely based on the structure of neurons in the brain. ANN is non-linear, nonparametric, data-driven and self-adaptive approaches as opposed to the model-based non-linear methods. ANN is capable of performing non-linear modeling without prior knowledge about the relationship between input and output data. One significant advantage of neural network models over other classes of nonlinear model is that ANN is a universal approximators which can approximate any continuous function with a desired accuracy. A neural network can be formed more accurately by using either long-term or short-term memory, depending on the retention time, into the structure of a static

network. For time series data forecasting we use some form of short-term memory to make neural network dynamic. One simple way of building short–term memory into the structure of a neural network is through the use of time delay, which can be implemented at the input layer of the neural network. Such architecture is a time-delay neural network (TDNN) that has been used for the present study. Figure 3.1 represents the general form of time-delay neural network.

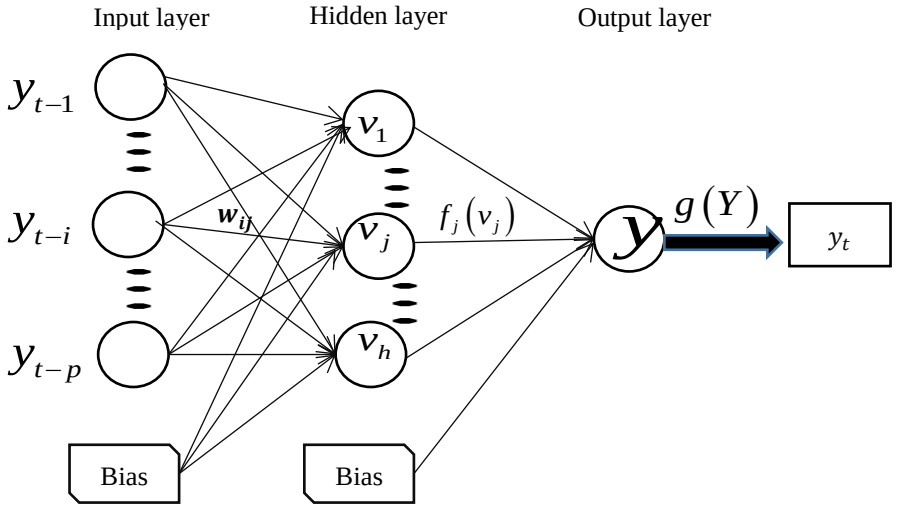


Figure 3.1: Time delay neural network (TDNN)

3.4.4.1 Architecture of Neural Network

Artificial Neural Network consists of a set of connected neurons (cells). The neurons receive information from either input cells or other neurons and perform some kind of transformation of the input and transmit the outcome to other neurons to output cells. The neural networks are built from layers of neurons connected so that one layer receives input from the preceding layer of neurons and passes the output on to the subsequent layer.

ANN consists of many neuron which are connected through weights. Accordingly different layers *i.e.* input, hidden, output layer and different nodes (input, hidden, output nodes) in every neural network has a different design. To design an ANN, we have to determine following things-

(1) Number of input, hidden and output layer

According to Zhang *et al.* (1998) there is no clear-cut guidelines to fix these components before construction of a neural network. By trial-error method we can construct a best neural network for current situation. A lot of research has been done in different field of neural network. It was found that for a univariate time series data to form a typical ANN, mainly one input, one output and one or at most two hidden layers are sufficient.

(II) Number of input, hidden, output node

The number of input node are equal to number of time lagged observation used to find out data pattern. If we use too few or too many time lagged observation then it may effect either learning or accuracy of neural network. So here we always use AR(p) to find out number of time lagged observations. Forecasting performance of a neural network always affected due to hidden nodes. To find a best network, the best way is trial-and-error approach. To avoid over fitting some thumb rule also given by different researcher like Lippmann (1987) used $2n + 1$ and Tang (1993) used n hidden nodes, where n is number of input node. The number of output node is equal to forecasting horizon. All layer's nodes are interconnected with each other from one layer to other layer respectively. Each node's output is a function of weighted sum of their inputs. Here we also use an activation function which transform input to output according to their relationship. When a i^{th}node n_i is connected with preceding layer's nodes $n_j, j = 1,2, \dots, (total\ nodes\ in\ this\ layer)$, than i^{th} node get some input value v_i *i.e.*

$$v_i = \sum_{j=1}^{total\ node} w_{ij} * output_j + b_i \qquad (3.4.16)$$

where w_{ij} is connection weight between i^{th} node and j^{th} node, b_i is bias for i^{th} node.

3.4.4.2 Activation function

There is some activation function or transfer function which convert net input value of a node to output for a particular node. Here mainly two activation function required one for hidden layer and another for output layer. Here in our research work we use sigmoid or logistic function for a hidden node n_i in hidden layer *i.e.* $f(v_i) = \frac{1}{1+e^{-v_i}}$ and identity function for output layer *i.e.*

$$\emptyset(o_i) = \begin{cases} 1, & o_i \geq 0 \\ 0, & otherwise \end{cases}$$

where o_i is net output for i^{th} output node

3.4.4.3 Training algorithm

The neural network training algorithm is used to improve forecasting accuracy. Training algorithm usually update the weights of a neural network. Due to updation of different weights we find a local optimum solution for a nonlinear problem. There are different methods which are used to find an optimum solution but no one guarantee to find out a global optimum solution for a given time series data. Most commonly method which is used for training is backpropagation (BP) algorithm. BP uses steepest descent method to find out weights for a neural network. It use supervised learning algorithm to find out more accurate weights. After finalize architecture of neural network it will used to predict the univariate time series.

So by using general expression for $\hat{y}_t$ at time t, a TDNN with single hidden layer is given by (Jha and Sinha, 2014) -

$$\hat{y}_t = g\left(\alpha_0 + \sum_{j=1}^{q} \alpha_j f\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right)\right)$$

where

$\hat{y}_t$ is the predicted value for y_t at time t

p input and q hidden nodes

i, j is i^{th} node of input layer, j^{th}node of hidden layer respectively

y_{t-i} ; $(i = 1,2, \ldots p)$ are network input nodes

β_{ij} $(j = 1, 2, \ . \ . \ ., q)$ refer the weight connection between neuron i^{th} and j^{th}

α_j Refer the weight between j^{th} neuron of hidden node and output node.

α_0 and β_{0j} are bias term for output layer and j^{th} hidden node

f and g are respectively hidden and output layer activation function, mainly logistic $f\left(v_j\right) = \frac{1}{1+e^{-v_j}}$ and g as identity function

3.4.5 Extreme Learning Machine (ELM)

Extreme learning machine (ELM) developed by Huang *et al.* (2006) is the state-of-art novel machine learning algorithm for Single Layer Feedforward Neural Network (SLFN). Consequently the ELM model has been widely used for the solution of

estimation problems in many different fields and is now gaining attention within the financial time series. The ELM model is easy to use and no parameters need to be tuned except the predefined network architecture, thus avoiding many complications faced by the gradient-based algorithms such as learning rate, learning epochs, and local minima. Importantly the ELM model has also been proven to be a faster algorithm compared with other conventional learning algorithms such as backpropagation (BP) or support vector machines (SVM). In the ELM approach most of the training is accomplished in time span of seconds or at least in minutes in large complex applications which are not easily achieved by using the traditional neural network models.

ELM was proposed for ‘‘generalized’’ single-hidden layer feedforward networks where the hidden layer need not be neuron alike (Huang *et al.*, 2006). The output function of ELM for generalized SLFN is

$$f_L(\boldsymbol{X}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i h_i(\boldsymbol{X}) = \boldsymbol{h}(\boldsymbol{X})\boldsymbol{\beta} \tag{3.4.17}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_L]^T$ is the output weight vector between the hidden layer of L nodes to the $m \geq 1$ output nodes, and $\boldsymbol{h}(\boldsymbol{X}) = [h_1(\boldsymbol{X}), \dots, h_L(\boldsymbol{X})]$ is ELM nonlinear feature mapping, e.g., the output (row) vector of the hidden layer with respect to the input $\boldsymbol{X}$.

$h_i(\boldsymbol{X})$ is the output of the i^{th} hidden node output. The output functions of hidden nodes may not be unique. Different output functions may be used in different hidden neurons.

In particular, in real applications $h_i(\boldsymbol{X})$ can be

$$h_i(\boldsymbol{X}) = G(\boldsymbol{a}_i,\ b_i, \boldsymbol{X}), \qquad \boldsymbol{a}_i \in \boldsymbol{R}^d,\ b_i \in \mathrm{R} \tag{3.4.18}$$

where $G(\boldsymbol{a}, b, \boldsymbol{X})$ (with hidden node parameters (a, b)) is a nonlinear piecewise continuous function satisfying ELM universal approximation capability theorems (Huang, *et al*., 2006).

Basically, ELM trains an SLFN in two main stages:
(1) random feature mapping and
(2) linear parameters solving.
In the first stage, ELM randomly initializes the hidden layer to map the input data into a feature space by some nonlinear mapping functions. The nonlinear mapping functions in

ELM can be any nonlinear piecewise continuous functions. In ELM, the hidden node parameters (a, b) are randomly generated (independent of the training data) according to any continuous probability distribution instead of being explicitly trained, leading to remarkable efficiency compared to traditional back propagation neural networks. In the second stage of ELM learning, the weights connecting the hidden layer and the output layer, denoted by β, are solved by minimizing the approximation error in the squared error sense:

$$\underset{\beta \in R^{L\times m}}{min} \ \|\boldsymbol{H}\beta - \boldsymbol{T}\|^2 \tag{3.4.19}$$

where

H is the hidden layer output matrix (randomized matrix):

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}(\boldsymbol{X}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{X}_N) \end{bmatrix} = \begin{bmatrix} h_1(\boldsymbol{X}_1) & \dots & h_L(\boldsymbol{X}_1) \\ \vdots & \vdots & \vdots \\ h_1(\boldsymbol{X}_N) & \cdots & h_L(\boldsymbol{X}_N) \end{bmatrix} \tag{3.4.20}$$

and **T** is the training data target matrix:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \tag{3.4.21}$$

where $\|\cdot\|$ denotes the Frobenius norm.

The optimal solution is given by:

$$\beta^* = \boldsymbol{H}^{\boldsymbol{\tau}}\boldsymbol{T} \tag{3.4.22}$$

where $\boldsymbol{H}^{\boldsymbol{\tau}}$ denotes the Moore–Penrose generalized inverse of matrix **H**.

Generalization performance:

ELM aims to reach better generalization performance by reaching both the smallest training error and the smallest norm of output weights:

$$minimize\text{: } \|\beta\|_p^{\sigma_1} + C\|H\beta - T\|_p^{\sigma_2} \tag{3.4.23}$$

where $\sigma_1 > 0,\ \sigma_2 > 0,\ \ p, q = 0, \frac{1}{2}, 1, 2, \dots, +\infty$

The first term in the objective function is a regularization term which controls the complexity of the learned model. Huang, *et al.* (2012) especially studied the stability and generalization performance of ELM with $\sigma_1 = \sigma_2 = p = q = 2$:

$$\underset{\beta \in R^{L\times m}}{min} \ \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N}\|e_i\|^2 \qquad s.t. \ \ \boldsymbol{h}(\boldsymbol{X}_i)\beta = \boldsymbol{t}_i^T - \boldsymbol{e}_i^T, \ \ i = 1, \dots, N. \quad (3.4.24)$$

By substituting the constraints of above equation into its objective function, we obtain the following equivalent unconstrained optimization problem:

$$\underset{\beta \in R^{L\times m}}{min} \ L_{ELM} = \ \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\|\boldsymbol{T} - \boldsymbol{H\beta}\|^2 \quad (3.4.25)$$

If **H** has more rows than columns ($N > L$), which is usually the case where the number of training patterns is larger than the number of the hidden neurons, we have the following closed form solution for β:

$$\boldsymbol{\beta}^* = \left(\boldsymbol{H}^T\boldsymbol{H} + \frac{\boldsymbol{I}}{C}\right)^{-1}\boldsymbol{H}^T\boldsymbol{T} \quad (3.2.26)$$

where **I** is an identity matrix of dimension L. If the number of training patterns is less than the number of hidden neurons ($N < L$), then $\boldsymbol{H}$ will have more columns than rows, so

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{H}^T + \frac{I}{C}\right)^{-1}\boldsymbol{T} \quad (3.2.27)$$

where, **I** is an identity matrix of dimension N.

3.4.6 Hybrid MS-GARCH – ELM Methodology

In order to propose the methodology for agricultural price volatility forecasting, we follow the following procedure:

Step-1

We apply the logarithmic transformation to stabilize the variance of the data. Logarithmic transformation is a special case of Box-Cox transformation when $\lambda = 0$. To make the series stationary, we take first difference as $lnX_t - lnX_{t-1}$, also known as return in the financial term. We divide the whole data in to training and testing sets. Last 12 observations are taken in the testing set and remaining data are used as training set.

Step-2

After logarithmic transformation and taking the first difference of the series, we find the ARMA structure of the differenced series based on the autocorrelation function, partial autocorrelation function and AIC value of the estimated model.

Step-3

Given the ARMA structure of the differenced series, we determine the residual series (actual observations minus predicted values) and square them to create a new variable 'esquare'. The 'esquare' series is taken as input variable for the Artificial Neural Network (ANN) and Extreme Learning Machine (ELM) model.

Step-4

We examine the heteroskedastic structure of the residuals. To test the ARCH effect in the residual series, we used Engle's Lagrange Multiplier test. If the ARCH effect are present in the residual series, then apply MS-GARCH model for the modelling and forecasting of conditional variance.

Step-5

We build ANN and ELM model by using 'esquare' series as an input variable. Since the 'esquare' is comparable to the conditional variance from the best GARCH model or MS-GARCH model. When, we design a robust ELM model, it is very important to consider important factors such as number of input nodes (which are the lags), number of nodes in the hidden layer, selection of best training algorithm and the activation function, and the number of output nodes. These factors are selected in such a way that the optimum model ensures the smallest generalization error and best performance assessment matrices.

Step-6

At last, we combine the forecast of the conditional variance from the fitted MS-GARCH as well as squared residuals from ELM using non parametric Nadaraya–Watson kernel weight function.

3.5 Long Memory Model and Structural Break

3.5.1 Defining Long Memory Process

Let $X_t(t = 0,1,2, \dots)$ be a stationary time series process and the autocorrelation function (ACF) of the time series with a time lag k of is given as

$$\rho_k = cov(X_t, X_{t-1})/var(X_t) \tag{3.5.1}$$

The series X_t is said to have short memory if the autocorrelation coefficient at lag k approaches to zero as k tends to infinity, *i.e.* $lim_{k\to\infty}\rho_k = 0$.

A standard formal definition of different types of linear dependence structures is given as follows:

Let X_t is a stationary process with autocovariance function $\gamma_X(k)$ and spectral density

$$f_X(\lambda) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty}\gamma_X(k)\exp(-ik\lambda) \quad (\lambda \in [-\pi, \pi]) \tag{3.5.2}$$

$$f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d} \tag{3.5.3}$$

where $L_f(\lambda) \geq 0$ is a symmetric function that is slowly varying at zero. Then X_t is said to exhibit (linear)

a) Long-range dependence if $d \in (0, \frac{1}{2})$
b) Intermediate dependence if $d = 0 \; and \; \lim_{\lambda\to 0} L_f(\lambda) = \infty$
c) Short-range dependence if $d = 0 \; and \; \lim_{\lambda\to 0} L_f(\lambda) = c_f \in (0, \infty)$
d) Antipersistence if $d \in (-\frac{1}{2}, 0)$

For a $X_t \sim I(1)$ series, the ACF declines linearly, and for a stationary $X_t \sim I(0)$ process, the ACF declines exponentially so that observations separated by a long time span may be regarded as independent. However, some empirically observed time series share neither of these characteristics, even though they are transformed to stationary by suitable differencing. These time series still exhibit a dependence between distant observations, and as the number of lags increases, dependence between apart events decreases very slowly. Such type of time series process is known as fractionally integrated process or long memory process. According to McLeod and Hipel (1978), a

process is said to possess a long memory if $\lim_{T\to\infty} \sum_{k=-T}^{T} |\rho_k|$ is non-finite. It is equivalent to stating that the spectral density of a long memory process becomes unbounded at low frequencies. For long memory processes, decaying of ACF occur at much slower rate (hyperbolic rate) which is consistent with $\rho_k \approx Ck^{2d-1}$, as k increases indefinitely, where *C* is a constant and d is the long memory parameter. The autocorrelation function of a long memory process exhibits persistency structure which is neither consistent with an *I(1)* process nor an *I(0)* process.

According to Beran (1995), a stationary long memory process has the following qualitative and quantitative properties:

(A) Qualitative properties

i. The observations of the time series data tends to stay either high level or low level for long period
ii. When we examine time plot of the series, then there seem to be cycles or local trends at short time periods. However, looking at the whole series, there is no apparent persisting trend or cycle.
iii. Overall, the series looks stationary.

(B) Quantitative properties:

i. The variance of the sample mean seems to decay to zero at a slower rate than $1/n$, in more appropriate proportional to $n^{-\alpha}$ for some $0 < \alpha < 1$.
ii. If the logarithm of periodogram $I(\lambda)$ is plotted against the logarithm of frequency, it seems to be scattered around a straight line having negative slope

3.5.2 Testing of Long Memory Parameter

One of the primary interests in the literature of time series with long memory properties is to estimate the unknown parameter *d* that describes the long memory properties or the low frequency behaviour of the spectral density function $f_X(\lambda)$. Two major classes of estimation methods are used for long memory testing *i.e.* parametric estimation and semiparametric estimation. For the parametric estimation, a complete parametric model such as Autoregressive Fractionally Integrated Moving Average (ARFIMA) model is

built that expresses autocovariance $\gamma_X(k)$ for all k, or the spectral density function $f_X(\lambda)$ for all λ, as a parameteric function of the parameters *d* and unknown scale factors. In contrast, the semi-parametric estimation is only interested in the memory parameter *d* and do not require the modelling of a complete set of the autocovariances. Parametric long memory estimation is computationally expensive and subject to misspecification. On the other hand, semi-parametric estimation consider *d* as the main parameter and avoid difficulties over the specifications of other parameters. Apart from parametric and semiparametric methods, some heuristic approach like R/S statistic (Hurst, 1951), visualizing ACF plot, variance plot, *etc.* and nonparametric methods like wavelet methodology (Jensen, 1999) are also used for the testing of long memory in the data series.

3.5.2.1 Rescaled Range Statistic (R/S statistic)

Hurst (1951) proposed the rescaled range statistic (R/S statistic) for detecting the presence of long-term memory. This descriptive statistics is defined as

$$R/S = \frac{1}{s_T}\left[\max_{1\le k\le T}\sum_{j=1}^{k}\left(X_j - \bar{X}\right) - \min_{1\le k\le T}\sum_{j=1}^{k}\left(X_j - \bar{X}\right)\right] \tag{3.5.4}$$

where s_T is the usual maximum likelihood standard deviation estimator,

$$s_T = \left[\frac{1}{T}\sum_{j=1}^{T}\left(X_j - \bar{X}\right)^2\right]^{\frac{1}{2}}. \tag{3.5.5}$$

This measure is always non-negative because the deviations from the sample mean $\bar{X}$ sum up to zero. Hence, the maximum of the partial sums will always be positive, and likewise the minimum will always be negative. Hurst (1951) showed that the probability limit

$$\operatorname*{plim}_{T\to\infty}\left\{T^{-H}\left(\frac{R/S}{s_t}\right)\right\} = const. \tag{3.5.6}$$

H is known as Hurst coefficient it is estimated as

$$\widehat{H} = \frac{\log(R/S)}{\log(T)} \tag{3.5.7}$$

A short-memory process is associated with a value of $H = \frac{1}{2}$, and estimated values greater than $\frac{1}{2}$ are taken as hindsight for long-memory behavior. Therefore, the differencing parameter *d* can be estimated as $\hat{d} = \widehat{H} - \frac{1}{2}$

3.5.2.2 GPH Estimate

An efficient method for estimating *d* was proposed by Geweke and Porter-Hudak (1983). They provided a semi-parametric estimator of *d* in the frequency domain. They consider as a data-generating process $(1 - L)^d X_t = \epsilon_t$ where $\epsilon_t \sim I(0)$. This process can be represented in the frequency domain

$$f_X(\lambda) = 1 - \exp(-i\lambda)|^{-2d} f_\epsilon(\lambda) \tag{3.5.8}$$

where $f_X(\lambda)$ and $f_\epsilon(\lambda)$ assign the spectral densities of X_t and ϵ_t, respectively. Equation (3.5.8) can be transformed to

$$log\{f_X(\lambda)\} = \left\{4sin^2\left(\frac{\lambda}{2}\right)\right\}^{-d} + log\{f_\epsilon(\lambda)\}, \tag{3.5.9}$$

$$log\{f_X(\lambda_j)\} = log\{f_\epsilon(0)\} - dlog\left\{4sin^2\frac{\lambda_j}{2}\right\} + log\left\{\frac{f_\epsilon(\lambda_j)}{f_\epsilon(0)}\right\} \tag{3.5.10}$$

The test regression is then a regression of the ordinates of the log spectral density on a trigonometric function of frequencies,

$$log\{I_X(\lambda_j)\} = \beta_1 + \beta_2 log\left\{4sin^2\frac{\lambda_j}{2}\right\} + v_j, \tag{3.5.11}$$

where $I_X(\lambda_j)$ be the periodogram of X_t with Fourier frequencies $\lambda_j = 2\pi_j/T$, $j = 1, \dots, m$. $v_j = log\left\{\frac{f_\epsilon(\lambda_j)}{f_\epsilon(0)}\right\}$ is an error term, *m* is the number of periodogram ordinates which will be used in regression and *T* is the number of observations. The error term is assumed to be *i.i.d.* with zero mean and variance $\frac{\pi}{6}$. The estimated order of fractional differencing is equal to $\hat{d} = -\hat{\beta}_2$. Its significance can be tested with either the usual *t* ratio distributed as Student *t* or one can set the residual variance equal $\frac{\pi}{6}$. A third possibility would be to choose m such that the estimated standard error of the regression is approximately equal to $\sqrt{\pi/6}$.

3.5.2.3 Local Whittle Estimation

Local Whittle (LW) estimator for testing long memory was first proposed by Kunsch (1987) and further developed by Robinson (1995). This estimator represents approximately a maximum likelihood estimation in the frequency domain, since for larger *T*

$$I_X(\lambda_j) \sim e^{f_X(\lambda_j)^{-1}} \tag{3.5.12}$$

The likelihood function is,

$$L\{I_X(\lambda_j), \dots, I_X(\lambda_m), \theta\} = \prod_{j=1}^{m} \frac{1}{f_\theta(\lambda_j)} e^{-I_X(\lambda_j) f_X(\lambda_j)^{-1}} \tag{3.5.13}$$

where $\theta = (C, d)$ is the parameter vector. The log-likelihood function becomes,

$$l(\theta) = \sum_{j=1}^{m} \left[-log f_\theta(\lambda_j) - \frac{I_X(\lambda_j)}{f_\theta(\lambda_j)} \right] \tag{3.5.14}$$

In the neighbourhood of zero frequency we obtain,

$$L(d, C) = \sum_{j=1}^{m} \left[logC - 2d\ log(\lambda_j) + \frac{I_X(\lambda_j)}{C{\lambda_j}^{-2d}} \right] \tag{3.5.15}$$

$$\frac{\partial l(C,d)}{\partial C} = \sum_{j=1}^{m} \left[\frac{1}{C} + \frac{I_X(\lambda_j)}{C{\lambda_j}^{-2d}} \right] \tag{3.5.16}$$

$$\hat{C} = m^{-1} \sum_{j=1}^{m} \left[\frac{I_X(\lambda_j)}{{\lambda_j}^{-2d}} \right] \tag{3.5.17}$$

Inserting $\hat{C}$ for C in equation 3.5.15 and by minimisation, the local Whittle estimator can be written as,

$$\hat{d}_{LW} = \arg min \left(log \left[m^{-1} \sum_{j=1}^{m} \left[\frac{I_X(\lambda_j)}{{\lambda_j}^{-2d}} \right] \right] - 2dm^{-1} \sum_{j=1}^{m} \log(\lambda_j) \right) \tag{3.5.18}$$

Robinson (1995) showed the Local Whittle estimator is consistent for $d \in (-0.5, 0.5)$. However, its consistency depends on the bandwidth *m*, which satisfy $\frac{1}{m} + \frac{m}{T} \to 0$ as $T \to \infty$. The LW estimator is more desirable due to its nice asymptotic properties, underlying mild assumptions and the likelihood interpretation. Robinson (1995 also showed that

$$\sqrt{m}(\hat{d}_{LW} - d) \rightarrow N(0, \frac{1}{4}) \quad (3.5.18)$$

3.5.2.4 The Exact Maximum Likelihood (EML)

Consider the following ARFIMA (p, d, q) process,

$$\Phi(L)(1-L)^d X_t = \Psi(L)\epsilon_t \quad (3.5.19)$$

where $\Phi(L)$ and $\Psi(L)$ are the polynomials

$$\Phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j \quad (3.5.20)$$

and

$$\Psi(L) = 1 + \sum_{j=1}^{q} \Psi_j L^j \quad (3.5.21)$$

involving autoregressive and moving average coefficients of order p and q respectively and ϵ_t is a white noise process. Now assume $X = (X_1, \dots, X_T)'$ follows a normal distribution with $X \sim N(0, \Sigma)$. The EML procedure allows for simultaneous estimation of both the long memory parameter and ARMA parameters. The maximum likelihood objective function is expressed as,

$$l_E(\Phi, \Psi, d; X) = -\frac{T}{2} log|\Sigma| - \frac{1}{2} X' \Sigma^{-1} X \quad (3.5.22)$$

As a result, the EML estimator of *d* can be derived as,

$$\hat{d}_{EML} = argmax \left[-\frac{T}{2} log|\Sigma| - \frac{1}{2} X' \Sigma^{-1} X \right] \quad (3.5.23)$$

3.5.3 Long Memory versus Structural Break

Understanding the difference between true and spurious long memory is of great importance for many agricultural price modelling. There is ambiguity over long memory and structural changes. The long memory process indicates constant unconditional volatility, while the structural change implies a dramatic shift in unconditional volatility and thus a structural break model is more plausible. Usually, time series with slowly decaying empirical autocorrelation functions are modeled as fractionally integrated processes. However, several authors point out that other data generating processes such as nonlinear time varying coefficient models, random level shift processes,

STOPBREAK models, and markov switching models can generate similar autocovariance features (Diebold and Inoue, 2001; Granger and Hyung, 2004). The existing literature on long memory and structural breaks suggests testing for long memory and structural breaks separately and then estimating a long memory model with breaks, after concluding for the existence of long memory and structural breaks.

Perron and Qu (2010) derived the properties of the periodogram of processes with short memory and structural break. They found that for low frequencies the effect of the shifts dominates the behavior of the spectral density and the implied value of *d* is one. For larger frequencies, on the other hand, the short memory component is dominant and the implied *d* is zero. These findings explain the sensitivity of semiparametric *d*-estimators with respect to the bandwidth choice. Therefore, Perron and Qu (2010) proposed a test statistic based on the difference between memory parameters estimated with different bandwidths. The same results on the spectral density of level shift processes are used by Qu (2011), who derives a score-type test that is based on the derivative of the local Whittle likelihood function to test true versus spurious long memory process. Simulation studies conducted by Qu (2011) and Leccadito *et al.* (2015) showed that against a wide range of alternatives the Qu test has the best power among the tests suggested so far. So, in this study to distinguish between true and spurious long memory process we used Qu test (Qu, 2011).

For the time series X_t , $(t = 1, \dots, T)$ with spectral density $f_X(\lambda)$ at frequency λ. Then the null hypothesis of the Qu test is as follows:

H_0: X_t is stationary with $f_X(\lambda) \cong G\lambda^{-2d}$ $as\ \lambda \to 0_+$ $with\ d \in (-1/2\,, 1/2)\ and\ G \in (0,\ \infty)$

The statistic is based on the local Whittle likelihood function written as

$$W = \sup_{r\in[\varepsilon,1]} \left(\sum_{j=1}^{m} \nu_j^2\right)^{-1/2} \left|\sum_{j=1}^{mr} \nu_j \left(\frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}} - 1\right)\right| \qquad (3.5.24)$$

where $\hat{d}$ is the local Whittle estimate of *d* using *m* frequency components and ε is a small trimming parameter.

3.5.4 Autoregressive Fractionally Integrated Moving Average (ARFIMA) Model

For modeling time series in presence of long memory, first of all Granger and Joyeux (1980) proposed Autoregressive Fractionally Integrated Moving Average (ARFIMA) model. Fractional integration is a generalization of integer integration. Normally a time series is assumed to be integrated with order zero or one but in the case of fractional integration method, the parameter may take any fractional value from 0 to 1. For example, an autoregressive moving average process integrated of order d *(ARFIMA*(*p, d, q*)) can be represented as

$$\phi(B)X_t = (1-B)^{-d}\theta(B)\epsilon_t \tag{3.5.25}$$

where ϵ_t is an *i.i.d.* random variable having zero mean and constant variance, B denotes the lag operator, $\phi(B)$ and $\theta(B)$ denote finite Autoregressive (AR) and Moving Average (MA) polynomials in the lag operator of order *p* and *q* respectively having roots outside the unit circle.

For parameter value $d = 0$ the process is stationary; for $0 < d < 0.5$ the process is said to have long memory. For the range $-0.5 < d < 0$, the sum of absolute values of its autocorrelations tends to a constant. In this case, the process exhibits negative dependency between distant observations and is therefore termed "anti-persistent" or to have "intermediate memory". It can be shown that the ACF of long memory processes declines hyperbolically instead of exponentially as would be the case for stable ARMA(p, q) models. The speed of the decay depends on the parameter value *d*. For instance, given a fractional white noise process ARFIMA(0, d, 0), Granger and Joyeux (1980) and Hosking (1981) have proved that the autocorrelations are given by

$$\rho_k = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} \tag{3.5.26}$$

3.5.6 Proposed Hybrid Methodology

To model the long memory property in the time series data, ARFIMA model has been widely used in the literature, and have very good performance compare to the ARIMA model. However, it is not able to capture the non-linear property of the data. On the other hand, to capture the nonlinear properties in the data, we generally used nonlinear

model like ANN, ELM, *etc*. ELM have good generalization power compare to the ANN and avoid over fitting problem. So, in this study we used ELM model to capture nonlinear patterns in the data. Both ARFIMA and ELM models have achieved successes in their own linear and nonlinear domains respectively. However, none of them is a universal model that is suitable for all conditions. Since, it is very difficult to know the characteristics of the data completely in a real world problem. Therefore, for the practical use, hybrid methodology that has both linear and nonlinear modeling capabilities can be a good strategy. Accordingly, we postulate that the time series data X_t can be decomposed into linear and nonlinear components.

$$X_t = L_t + N_t \tag{3.5.27}$$

where, X_t is the time series data under consideration, L_t is the linear autoregressive component, and N_t is the nonlinear component. To account the structural break present in the series, we estimate ARFIMA model with dummy variable. For this, first we detect the structural break in the data and accordingly put dummy variable as a regressor. Therefore, ARFIMA with dummy variable can take care of both long memory and structural break present in the data. Accordingly, to fit the linear component of the data series, we used ARFIMA model with dummy variable and obtained the residuals. To check the existence nonlinear component in the residuals, we employed BDS test. If there is the evidence of non-linearity then residuals are modelled using Extreme Learning Machine. Let e_t is the residual at time *t* from the ARFIMA model with dummy variable, then

$$e_t = X_t - \hat{L}_t \tag{3.5.28}$$

where, $\hat{L}_t$ is the forecast of ARFIMA model with dummy variable at time *t*. With *k* input nodes, ELM model for residuals will be

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-k}) + \epsilon_t \tag{3.5.29}$$

where, *f* is a nonlinear function and ϵ_t is a random error. Let forecast from the nonlinear component at time *t* be $\widehat{N}_t$, then the combined forecast will be

$$\hat{X}_t = \hat{L}_t + \widehat{N}_t \tag{3.5.30}$$

Hence, the proposed hybrid method would exploit the strength of both ARFIMA and ELM models.

3.6 Co-integration with Structural Break

3.6.1 Co-integration and Error Correction Models (ECM)

Granger (1981) introduced the idea of co-integration into the literature, and Engle and Granger (1987) published the general case in their seminal paper. The concept behind co-integration is to find a linear combination of two *I(d)*-variables that will produce a variable with a lower order integration. More formally, co-integration can be defined as "The components of the vector $\boldsymbol{X_t}$ are said to be cointegrated of order *d*, *b*, denoted $\boldsymbol{X_t} \sim CI(d, b)$ if (a) all components of $\boldsymbol{X_t}$ are *I(d)* and (b) a vector $\boldsymbol{\alpha}(\neq 0)$ exists so that $\epsilon_t = \boldsymbol{\alpha}'\boldsymbol{X_t} \sim I(d-b), b > 0$. The vector $\boldsymbol{\alpha}$ is called the co-integrating vector".

The great concern among researchers in this path-breaking development is mostly explained by the fact that stable, long-run relationships can now be identified among non-stationary variables. Consider the case of *d = 1, b =1*; *i.e.*, the components in the vector $\boldsymbol{X_t}$ are all integrated of order one, but if a linear combination $\boldsymbol{\alpha}$ of these exists, then the resultant series ϵ_t is integrated of order zero. Even though the individual series are non-stationary, the co-integrating vector binds them to each other. In the term of economics, deviations from a long- equilibrium direction may be possible, but these errors are characterized by a mean reversion to its stable long-run equilibrium.

To estimate the co-integrating vector $\boldsymbol{\alpha}$ and to model the dynamic behavior of *I(d)* variable, Engle and Granger (1987) proposed a two-step estimation technique. In the first step, a regression of the variables in the set of *I(1)* is run,

$$Y_t = \alpha_1 X_{t,1} + \alpha_2 X_{t,2} + \cdots + \alpha_k X_{t,K} + \epsilon_t \quad for\ t = 1, \dots, T \tag{3.6.1}$$

where ϵ_t is an error term. The estimated $(K+1)$ co-integrating vector $\hat{\boldsymbol{\alpha}}$ is given by $\hat{\boldsymbol{\alpha}} = (1, -\hat{\boldsymbol{\alpha}}^*)'$, where $\hat{\boldsymbol{\alpha}}^* = (\hat{\alpha}_1, \dots, \hat{\alpha}_K)'$. Hence, the co-integrating vector is normalized to the regressand. Engle and Granger (1987) have shown that the co-integrating vector can be consistently estimated in this static regression but with a finite sample bias of magnitude $O_p(T^{-1})$. Because the usual convergence rate in the *I(0)* case is only

$O_p(T^{-1/2})$, Stock (1987) termed the OLS estimation of the co-integrating vector as "superconsistent". Incidentally, although the co-integrating vector can be superconsistently estimated, Stock has shown that the limiting distribution is non-normal; hence, as in the case of spurious regressions, the typical *t* and *F* statistics are not applicable.

In the second step, an error-correction model (ECM) is specified. Here, the bivariate case in which two cointegrated variables Y_t and X_t, each *I(1)*, are considered. The general specification of an ECM is as follows:

$$\Delta Y_t = \psi_0 + \gamma_1 \hat{\epsilon}_{t-1} + \sum_{i=1}^{k} \psi_{1,i} \Delta X_{t-i} + \sum_{i=1}^{l} \psi_{2,i} \Delta Y_{t-i} + \varepsilon_{1,t} \quad (3.6.2)$$

$$\Delta X_t = \xi_0 + \gamma_2 \hat{\epsilon}_{t-1} + \sum_{i=1}^{k} \xi_{1,i} \Delta Y_{t-i} + \sum_{i=1}^{l} \xi_{2,i} \Delta X_{t-i} + \varepsilon_{2,t} \quad (3.6.3)$$

where $\hat{\epsilon}_t$ is the error from the static regression in equation 3.6.1, and $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are white noise processes. The error correction model in equation 3.6.2 states that changes in Y_t are explained by their own past value, lagged changes of X_t, and the error from the long-run equilibrium in the previous period. The value of the coefficient γ_1 determines the speed of adjustment and should be negative in sign. Otherwise the system would diverge from its long-run equilibrium path. Incidentally, one is not restricted to including the error from the previous period only. It can be any lagged value as equations 3.6.2 and 3.6.3 are still balanced as $\hat{\epsilon}_{t-1}$ is stationary and $\hat{\epsilon}_{t-k}$ with $k > 1$ is also stationary. It can be inferred from these equations and the static regression, Granger causality will occur in at least one direction in the case of two co-integrated *I(1)*-variables, *i.e.* at least one variable can help forecast the other.

3.6.2 Systems of Co-integrated Variables

Campbell and Perron (1991) have defined co-integration in a broader sense as follows: An $(K \times 1)$ vector of variables $\boldsymbol{Y_t}$ is said to be cointegrated if at least one non zero n-element vector $\boldsymbol{\beta_i}$ exists such that $\boldsymbol{\beta_i' Y_t}$ is trend stationary. $\boldsymbol{\beta_i}$ is called a co-integrating vector. If *r* such linearly independent vectors $\boldsymbol{\beta_i}(i = 1, \dots, r)$ exist, we say that $\boldsymbol{Y_t}$ is cointegrated with cointegrating rank *r*. We then define the $(K \times r)$ matrix of co-

integrating vectors $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \dots, \boldsymbol{\beta_r})$. The r elements of the vector $\boldsymbol{\beta' Y}_t$ are trend-stationary, and $\boldsymbol{\beta}$ is called the cointegrating matrix.

This definition is broader than the one by Engle and Granger in the sense that now it is no longer required that each individual series be integrated of the same order. For example, some or all series can be trend-stationary. If $\boldsymbol{Y_t}$ contains a trend-stationary variable, then it is trivially co-integrated and the co-integrating vector is the unit vector that selects the stationary variable. On the other hand, if all series are trend-stationary, then the system is again trivially co-integrated because any linear combination of trend-stationary variables yields a trend-stationary variable.

Let us consider the vector autoregression model of order *p*

$$\boldsymbol{Y}_t = \boldsymbol{\Pi}_1 \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{\Pi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\Phi D}_t + \boldsymbol{\varepsilon}_t \; for \; t = 1, \dots, T \tag{3.6.4}$$

where $\boldsymbol{Y}_t$ assigns the $(K \times 1)$ vector of series at period *t*, the matrices $\boldsymbol{\Pi}_i (i = 1, \dots, p)$ are the $(K \times K)$ coefficient matrices of the lagged endogenous variables, $\boldsymbol{\mu}$ is a $(K \times 1)$ vector of constants, and $\boldsymbol{D}_t$ is a vector of non-stochastic variables such as seasonal dummies or intervention dummies. The $(K \times 1)$ error term $\boldsymbol{\varepsilon}_t$ is assumed to be i.i.d. as $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$

From equation (3.6.4), two versions of a VECM can be delineated. In the first form, the levels of $\boldsymbol{Y}_t$ enter with lag $t - p$:

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma}_1 \Delta \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{Y}_{t-p+1} + \boldsymbol{\Pi Y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\Phi D}_t + \boldsymbol{\varepsilon}_t \; for \; t = 1, \dots, T \tag{3.6.5}$$

$$\boldsymbol{\Gamma}_i = -(\mathbf{I} - \boldsymbol{\Pi}_1 - \cdots - \boldsymbol{\Pi}_i) \;\; for \; i = 1, \dots, p-1, \tag{3.6.6a}$$

$$\boldsymbol{\Pi} = -\big(\mathbf{I} - \boldsymbol{\Pi}_1 - \cdots - \boldsymbol{\Pi}_p\big), \tag{3.6.6b}$$

where $\mathbf{I}$ is the $(K \times K)$ identity matrix. As can be seen from Equation (3.6.6a), the $\boldsymbol{\Gamma}_i (i = 1, \dots, p-1)$ matrices contain the cumulative long-run impacts; hence, this specification is termed the long-run form. Please note that the levels of $\boldsymbol{Y}_t$ enter with lag $t - p$.

The other VECM specification is of the form

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma}_1 \Delta \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{Y}_{t-p+1} + \boldsymbol{\Pi} \boldsymbol{Y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t \; for \; t = 1, \dots, T \quad (3.6.7a)$$

$$\boldsymbol{\Gamma}_i = -\left(\boldsymbol{\Pi}_{i+1} + \cdots + \boldsymbol{\Pi}_p\right) \; for \; i = 1, \dots, p-1, \quad (3.6.7b)$$

$$\boldsymbol{\Pi} = -\left(\mathbf{I} - \boldsymbol{\Pi}_1 - \cdots - \boldsymbol{\Pi}_p\right). \quad (3.6.7c)$$

The $\boldsymbol{\Pi}$ matrix is the same as in the first specification. However, the $\boldsymbol{\Gamma}_i$ matrices now differ in the sense that they measure transitory effects; hence, this form of the VECM is termed the transitory form. Furthermore, the levels of the components in $\boldsymbol{Y}_t$ enter lagged by one period. Incidentally, as will become evident, inferences drawn on $\boldsymbol{\Pi}$ will be the same regardless of which specification is chosen, and the explanatory power is the same.

As per assumption, the individual components of $\boldsymbol{Y}_t$ are at most *I(1)*-variables. Therefore, the left-hand side of the VECM is stationary. Besides lagged differences of $\boldsymbol{Y}_t$, the error-correction term $\boldsymbol{\Pi} \boldsymbol{Y}_{t-p}$ or, depending on the specification of the VECM, $\boldsymbol{\Pi} \boldsymbol{Y}_{t-1}$ appears. This term must be stationary, too; otherwise the VECM will not balance. The question now is, what kind of conditions must be given for the matrix $\boldsymbol{\Pi}$ such that the right hand side is stationary? Three cases must be considered,

(i) $\text{rank}(\boldsymbol{\Pi}) = K$

(ii) $\text{rank}(\boldsymbol{\Pi}) = 0$

(iii) $\text{rank}(\boldsymbol{\Pi}) = r < K$

where $rank()$ assigns the rank of a matrix. In the first case, all K linearly independent combinations must be stationary. This can only be the case if the deviations of $\boldsymbol{Y}_t$ around the deterministic components are stationary. Equations (3.6.5) and (3.6.7) represent a standard VAR-model in levels of $\boldsymbol{Y}_t$. In the second case, in which the rank of $\boldsymbol{\Pi}$ is zero, no linear combination exists to make $\boldsymbol{\Pi} \boldsymbol{Y}_t$ stationary except for the trivial solution. Hence, this case would correspond to a VAR-model in first differences. The interesting case is the third one, in which $0 < rank(\boldsymbol{\Pi}) = r < K$. Because the matrix does not have full rank, two $(K \times r)$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ exist such that $\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$. Hence, $\boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{Y}_{t-p}$ is stationary, and therefore the matrix-vector product $\boldsymbol{\beta}' \boldsymbol{Y}_{t-p}$ is stationary. The r linear independent columns of $\boldsymbol{\beta}$ are the co-integrating vectors, and the rank of $\boldsymbol{\Pi}$ is equal to

the co-integration rank of the system $\boldsymbol{Y}_t$. That is, each column represents one long-run relationship between the individual series of $\boldsymbol{Y}_t$. However, the parameters of the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are undefined because any non-singular matrix $\mathbf{H}$ would yield $\boldsymbol{\alpha}\mathbf{T}(\boldsymbol{\beta}\mathbf{H}^{-1})' = \boldsymbol{\Pi}$. It implies that only the co-integration space spanned by $\boldsymbol{\beta}$ can be determined. The obvious solution is to normalize one element of $\boldsymbol{\beta}$ to one. The elements of α determine the speed of adjustment to the long-run equilibrium. It is referred to as the loading or adjustment matrix.

Johansen (1988), Johansen (1991) and Johansen and Juselius (1990) developed maximum-likelihood estimators of these co-integration vectors for an autoregressive process as in Equations (3.6.4) through (3.6.7). Their approach uses canonical correlation analysis as a means to reduce the information content of *T* observations in the K-dimensional space to a lower-dimensional one of *r* co-integrating vectors. Hence, the canonical correlations determine the extent to which the multicollinearity in the data will allow such a smaller *r*-dimensional space. To do so, $2K$ auxiliary regressions are estimated by OLS: $\Delta\boldsymbol{Y}_t$ is regressed on lagged differences of $\boldsymbol{Y}_t$. The residuals are termed $\boldsymbol{R}_{0t}$. In the second set of auxiliary regressions, $\boldsymbol{Y}_{t-p}$ is regressed on the same set of regressors. Here, the residuals are assigned as $\boldsymbol{R}_{1t}$. The $2K$ residual series of these regressions are used to compute the product moment matrices as

$$\widehat{\boldsymbol{S}}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R}_{it}\boldsymbol{R}'_{jt} \text{ with } i,j = 0,1 \tag{3.6.8}$$

Johansen (1991) showed that the likelihood-ratio test statistic of the null hypothesis that there are at most *r* co-integrating vectors is

$$-2\ln(Q) = -T\sum_{i=r+1}^{n}\left(1-\hat{\lambda}_i\right), \tag{3.6.9}$$

where $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ are the $n-r$ smallest eigenvalues of the equation

$$\left|\boldsymbol{\lambda}\widehat{\boldsymbol{S}}_{11} - \widehat{\boldsymbol{S}}_{10}\widehat{\boldsymbol{S}}_{00}^{-1}\widehat{\boldsymbol{S}}_{01}\right| = 0 \tag{3.6.10}$$

For ease of computation, the $(K \times K)$ matrix $\widehat{\boldsymbol{S}}_{11}$ can be decomposed into the product of a non-singular $(K \times K)$ matrix $\mathbf{C}$ such that $\widehat{\boldsymbol{S}}_{11} = \mathbf{C}\mathbf{C}'$. Equation (3.6.10) would then accordingly be written as

$$\left|\boldsymbol{\lambda}\mathbf{I} - \mathbf{C}^{-1}\widehat{\boldsymbol{S}}_{10}\widehat{\boldsymbol{S}}_{00}^{-1}\widehat{\boldsymbol{S}}_{01}\mathbf{C}'^{-1}\right| = 0 \tag{3.6.11}$$

where **I** assigns the identity matrix and $|\cdot|$ denotes determinant.

Johansen (1988) has tabulated critical values for the test statistic in Equation (3.6.9) for various quantiles and up to five co-integration relations; *i.e.*, $r = 1, \dots, 5$. This statistic has been named the trace statistic.

Besides the trace statistic, Johansen and Juselius (1990) have suggested the maximal eigenvalue statistic defined as

$$-2\ln(Q; r|r+1) = -Tln(1 - \hat{\lambda}_{r+1}) \tag{3.6.12}$$

Once the cointegration rank r has been determined, the cointegrating vectors can be estimated as

$$\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{v}}_1, \dots, \widehat{\boldsymbol{v}}_r) \tag{3.6.13}$$

where $\widehat{\boldsymbol{v}}_i$ are given by $\widehat{\boldsymbol{v}}_i = \mathbf{C}'^{-1}\boldsymbol{e}_i$ and $\boldsymbol{e}_i$ are the eigenvectors to the corresponding eigenvalues in Equation (3.6.11). Equivalent to this are the first r eigenvectors of $\hat{\boldsymbol{\lambda}}$ in equation (3.6.10) if they are normalized such that $\widehat{\mathbf{V}}'\widehat{\boldsymbol{S}}_{11}\widehat{\mathbf{V}} = \mathbf{I}$ with $\widehat{\mathbf{V}} = (\widehat{\boldsymbol{v}}_1, \dots, \widehat{\boldsymbol{v}}_k)$.

The adjustment matrix $\boldsymbol{\alpha}$ is estimated as

$$\widehat{\boldsymbol{\alpha}} = -\widehat{\boldsymbol{S}}_{01}\widehat{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{S}}_{11}\widehat{\boldsymbol{\beta}}\right)^{-1} = -\widehat{\boldsymbol{S}}_{01}\widehat{\boldsymbol{\beta}} \tag{3.6.14}$$

The estimator for α is dependent on the choice of the optimizing **β**. The estimator for the matrix **Π** is given as

$$\widehat{\boldsymbol{\Pi}} == -\widehat{\boldsymbol{S}}_{01}\widehat{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{S}}_{11}\widehat{\boldsymbol{\beta}}\right)^{-1}\widehat{\boldsymbol{\beta}}' = -\widehat{\boldsymbol{S}}_{01}\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}' \tag{3.6.15}$$

Finally, the variance-covariance matrix of the K-dimensional error process $\boldsymbol{\varepsilon}_t$ is given as

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{S}}_{00} - \widehat{\boldsymbol{S}}_{01}\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{S}}_{10} = \widehat{\boldsymbol{S}}_{00} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\alpha}}' \tag{3.6.16}$$

3.6.3 VECM and Structural Break

The pitfalls of falsely concluding non-stationarity in the data can also be encountered in the case of VECM. The structural break may causes wrongly acceptance co-integration

relationship. Lutkepohl *et al.* (2004) proposed a procedure for estimating a VECM in which the structural break is a simple shift in the level of the process and the break date is estimated first. Next, the deterministic part, including the size of the shift, is estimated, and the data are adjusted accordingly. Finally, a Johansen-type test for determining the cointegration rank can be applied to these adjusted series.

Lutkepohl *et al.* (2004) assume that the $(K \times 1)$ vector process $\boldsymbol{Y}_t$ is generated by a constant, a linear trend, and level shift terms

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\delta} d_{t\tau} + \boldsymbol{X}_t \tag{3.6.17}$$

Where $d_{t\tau}$ is a dummy variable defined by $d_{t\tau} = 0$ for $t < \tau$ and $d_{t\tau} = 1$ for $t \geq \tau$. The shift assumes that the shift point τ is unknown and is expressed as a fixed fraction of the sample size,

$$\tau = [T\lambda] \text{ with } 0 < \underline{\lambda} \leq \lambda \leq \bar{\lambda} < 1 \tag{3.6.18}$$

where $\underline{\lambda}$ and $\bar{\lambda}$ define real numbers and [.] defines the integer part. The meaning of equation 3.6.18 is that the shift might occur neither at the very beginning nor at the very end of the sample. Furthermore, it is assumed that the process $\{\boldsymbol{X}_t\}$ can be represented as a VAR(p) and that the components are at most $I(1)$ and cointegrated with rank r.

The estimation of the break point is based on the regressions

$$\boldsymbol{Y}_t = \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \boldsymbol{\delta} d_{t\tau} + \boldsymbol{A}_1 \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\epsilon}_{t\tau} \text{ for } t = p+1, \ldots, T \tag{3.6.19}$$

where $\boldsymbol{A}_i$ with $i = 1, \ldots, p$ assign the $(K \times K)$ coefficient matrices and $\boldsymbol{\epsilon}_t$ is the spherical K-dimensional error process. It should be noted that other exogenous regressors, like seasonal dummy variables, can also be included in Equation (3.6.19).

The estimator for the break point $\hat{\tau}$ is then defined as

$$\hat{\tau} = \arg\min_{\tau \in \lambda} \det\left(\sum_{t=p+1}^{T} \hat{\epsilon}_{t\tau} \hat{\epsilon}'_{t\tau}\right) \tag{3.6.20}$$

where $\lambda = [T\underline{\lambda}, T\bar{\lambda}]$ and $\hat{\epsilon}_{t\tau}$ are the least-squares residuals of Equation (3.6.19). The integer count of the interval $\lambda = [T\underline{\lambda}, T\bar{\lambda}]$ determines how many regressions have to be run with the corresponding step dummy variables $\boldsymbol{\delta} d_{t\tau}$ and how many times the

determinant of the product moment matrices of $\hat{\epsilon}_{t\tau}$ have to be calculated. The minimal one is the one that selects the most likely break point.

Once the break point $\hat{\tau}$ is estimated, the data are adjusted according to

$$\widehat{\boldsymbol{X}}_t = \boldsymbol{Y}_t - \widehat{\boldsymbol{\mu}}_0 - \widehat{\boldsymbol{\mu}}_1 t - \widehat{\boldsymbol{\delta}} d_{t\hat{\tau}} \tag{3.6.21}$$

3.7 Evaluating Forecast Accuracy

1) Root Mean Squared Error (RMSE)

Mean squared error is the average of squared errors. For n test set it is calculated as-

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{3.7.1}$$

The RMSE is obtained by taking square root of MSE

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{3.7.2}$$

where y_t and $\hat{y}_t$ are the actual and predicted value respectively and n is the number of observations.

2) Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{3.7.3}$$

3) Correct Directional Change (CDC)

$$CDC = \frac{1}{n}\sum_{t=1}^{n} a_t \times 100 \tag{3.7.4}$$

where, $a_t = \begin{cases} 1, & if\ [y_{t+1} - y_t][\hat{y}_{t+1} - \hat{y}_t] \geq 0 \\ 0, & otherwise \end{cases}$

This statistic indicates the direction of change. Larger the value of CDC greater is the change.

4) Theil U- Statistic (U)

$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}}{\left[\sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t^2} + \sqrt{\frac{1}{n}\sum_{t=1}^{n} \hat{y}_t^2}\right]} \tag{3.7.5}$$

This statistic is scale independent. The value of U ranges between 0 and 1 and value close to 0 indicates efficient model.

CHAPTER IV

RESULTS AND DISCUSSION

This chapter contains the comprehensive findings and discussion of the research work carried out to accomplish three specific objectives of the study. The chapter is divided into three main sections as per the objectives of the current investigation. For better understanding, results of these investigations are presented in tabular and graphical form and explanation of each table and figure is discussed wherever required.

4.1 Results of Volatility Models with Structural Break

Unprecedented variability or volatility in the agricultural commodity prices creates much uncertainty and risk for all market participants, and makes both short-term and long-term planning difficult. Often volatility is assumed same as risk, but the fact is, risk deals only with negative price shocks while volatility takes care of both negative and positive shocks. To estimate the volatility huge numbers of non-linear parametric and non-parametric models have been developed. A major issue with the prediction of volatility is the presence of structural break in the data. In this section, first we assess some existing methods and then propose a hybrid method for agricultural price volatility forecasting in the presence of structural break by combining the Markov Switching GARCH (MS-GARCH) model with Extreme Learning Machine (ELM) model.

4.1.1 Data Description

For the present study, weekly Potato price (₹/q) of Delhi market and monthly international price ($/mt) of Groundnut oil and Palm oil has been used. The daily data of Potato price of Delhi market are obtained from National Horticultural Research and Development Foundation (NHRDF) (http://nhrdf.org/en-us/) website for the period 1st January 2005 to 31st December 2019, and then converted to weekly data. International monthly Groundnut oil and Palm oil price from January 1980 to December 2019, are obtained from World Bank Commodity price available at official website (https://www.worldbank.org/en/research/commodity-markets). In this study, we applied

the natural choice of logarithmic transformation to the data to stabilize the variance, and have taken first difference to the log transformed data to form a stationary return series. The data sets are divided into training and testing sets. Last 12 observations are taken as testing set whereas remaining data are used for the training of the model.

Table 4.1 summarizes the descriptive statistics for the weekly Potato price of Delhi market from Jan-2005 to Dec-2019, International monthly Groundnut oil and Palm oil price series from Jan-1980 to Dec-2019. The mean value of Potato price series of Delhi market is ₹808.7 per quintal whereas 1103.60 $/mt and 581.80 $/mt for International Groundnut and Palm oil price series respectively. The positive skewness indicates that all the three price series are distributed with an asymmetric tail extending towards positive values. Each series has excess kurtosis compared to the normal distribution. Moreover, neither of the price series is normally distributed as indicated by the values of Jarque-Bera normality test.

Table 4.1: Descriptive statistics of Delhi weekly Potato price (Jan-2005 to Dec-2019), International monthly Groundnut oil and Palm oil price (Jan-1980 to Dec-2019)

	Delhi Potato	**Int. Groundnut oil**	**Int. Palm oil**
Mean	808.70	1103.60	581.80
Std. Dev.	419.45	438.86	238.10
Minimum	225.20	445.00	197.00
Maximum	2669.00	2502.20	1377.20
CV (%)	51.86	39.78	40.92
Skewness	1.43	1.00	0.97
Kurtosis	5.61	3.93	3.75
Jarque-Bera test (p-value)	<0.01	<0.01	<0.01

Note: Potato price (₹/q), Groundnut oil and Palm oil price ($/mt)

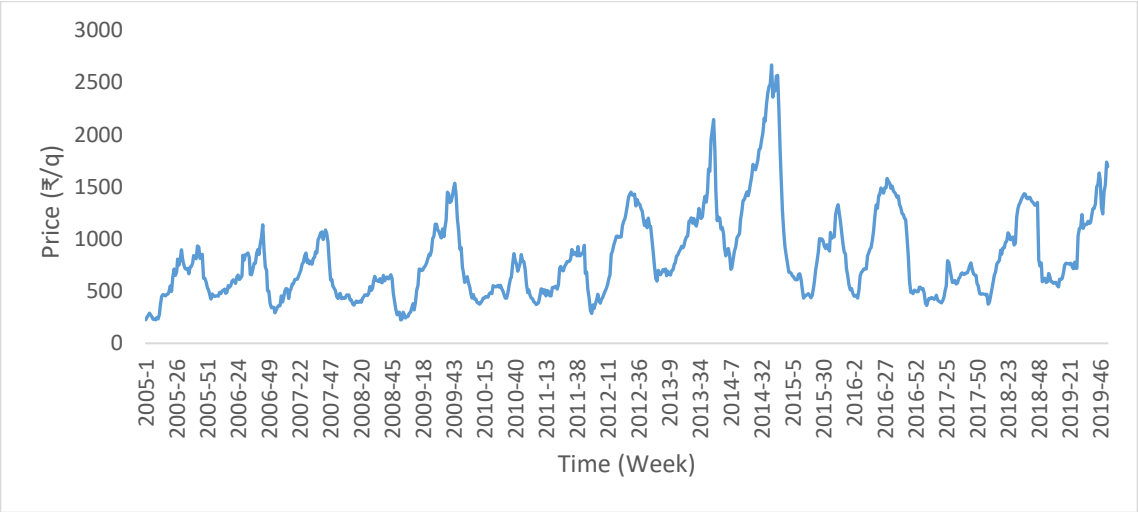


Figure 4.1a: Time plot of weekly Potato price of Delhi market (Jan-2005 to Dec-2019)

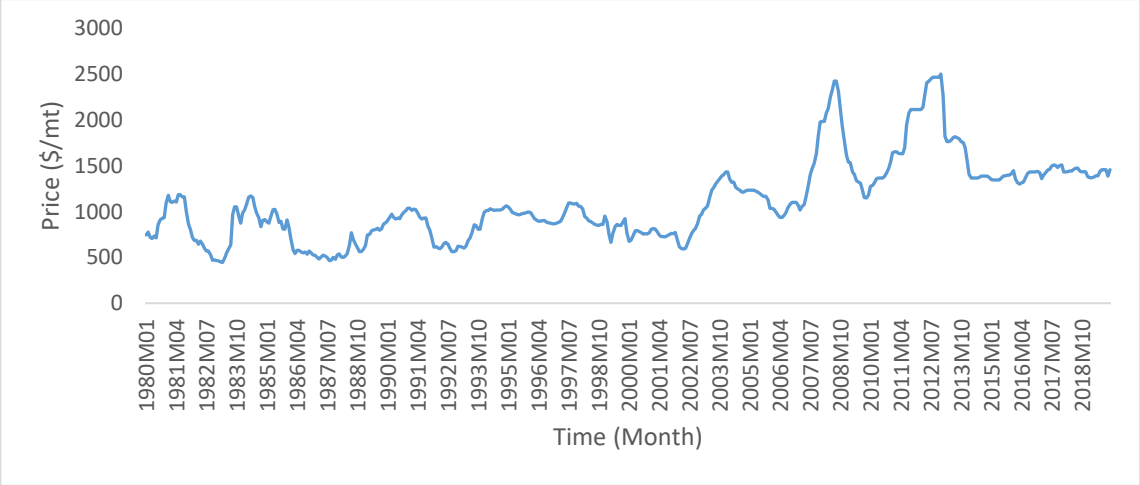


Figure 4.1b: Time plot of international monthly Groundnut oil price (Jan-1980 to Dec-2019)

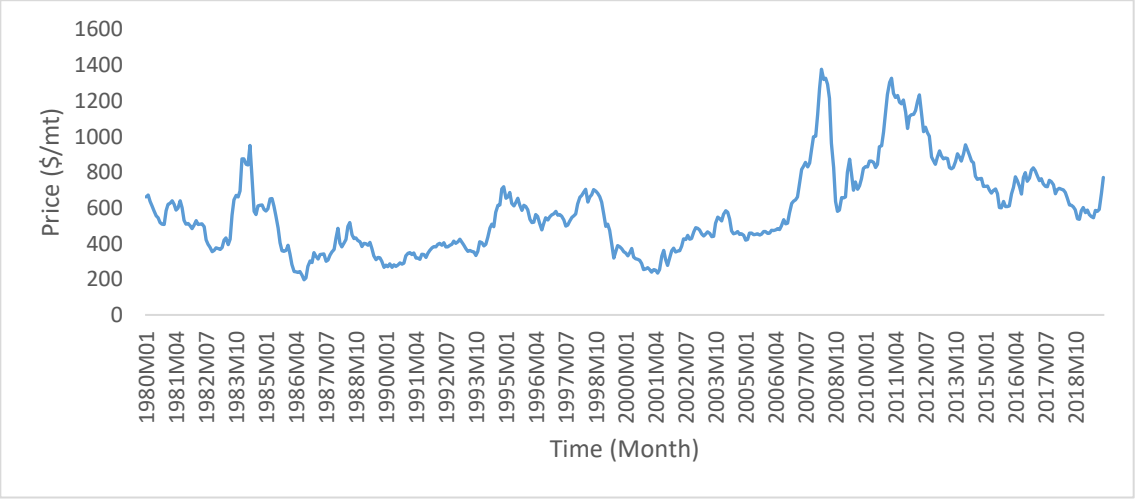


Figure 4.1c: Time plot of international monthly Palm oil price (Jan-1980 to Dec-2019)

4.1.2 Test for Stationarity, Linearity and Structural Break

To test the stationarity of the data series under consideration, we used Augmented-Dickey-Fuller (ADF) and Phillip-Perron (PP) tests, which provide the presence or absence of unit root in the data series. For both ADF and PP test, null hypothesis is presence of unit root in the data, which means data is non-stationary and alternative hypothesis being data series is generated from a stationary process. The results of both tests are reported in Table 4.2 for all the three price series. According to ADF and PP test results, all the series were non-significant at 5% level of significance *i.e.* all the series are non-stationary at level. However, on first difference, both the tests were found to be highly significant at 1% level of significance.

Table 4.2: Unit root tests

Price Series		**Augmented Dickey-Fuller Test**		**Phillip-Perron Test**	
		t-statistic	**p-value**	**t-statistic**	**p-value**
Delhi Potato	Level	-1.22	0.24	-4.42	0.20
	1st difference	-13.03	<0.001	-547.00	<0.001
International Groundnut oil	Level	-0.63	0.45	-0.74	0.52
	1st difference	-12.76	<0.001	-260.00	<0.001
International Palm oil	Level	-0.98	0.32	-1.79	0.42
	1st difference	-16.23	<0.001	-304.00	<0.001

We also employed Brock-Decher-Scheikman (BDS) test, to check the linearity characteristics of the price series. In this test, by using any linear econometric model, data is detrended first to remove linear structure and test whether the remaining residuals are independent or not, according to which we have to accept or reject the null hypothesis. If residuals are independent then we don't reject the null hypothesis which means data series is linear, and if residuals are not independent then we have to reject the null hypothesis which means data series is nonlinear. The detail results of the BDS test are presented in the Table 4.3. Here probability value is less than 0.01 for both the embedding dimensions, which shows that all three price series are nonlinear at 1% level of significance.

Table 4.3: Brock-Decher-Scheikman (BDS) test results

Series		Embedding dimension				Conclusion
		2		3		
	Epsilon	Statistic	p-value	Statistic	p-value	
Delhi Potato	0.5σ	112.96	< 0.001	174.76	< 0.001	Non Linear
	σ	69.31	< 0.001	80.17	< 0.001	
	1.5σ	53.71	< 0.001	55.05	< 0.001	
	2σ	48.30	< 0.001	46.53	< 0.001	
Int. Groundnut oil	0.5σ	145.41	< 0.001	250.02	< 0.001	Non Linear
	σ	71.84	< 0.001	86.20	< 0.001	
	1.5σ	48.83	< 0.001	50.48	< 0.001	
	2σ	40.62	< 0.001	39.16	< 0.001	
Int. Palm oil	0.5σ	140.49	< 0.001	235.33	< 0.001	Non Linear
	σ	67.16	< 0.001	80.31	< 0.001	
	1.5σ	48.01	< 0.001	49.66	< 0.001	
	2σ	42.12	< 0.001	40.81	< 0.001	

Recent studies have established that structural breaks could severely affect the results of volatility models. Structural break detection is the problem of determining the point at which a series of observations change its statistical properties. For identification of structural break in the variance, over years several algorithms have been proposed like binary segmentation, Iterated Cumulative Sum of Squares (ICSS) algorithm, Pruned Exact Linear Time (PELT) algorithm, *etc.* The PELT algorithm has been developed by Killick *et. al.* (2012), and it has several advantages over other algorithm. So, in this study, to check the structural break in the series PELT algorithm is used. The plots of return series along with break points are presented in the Figure 4.2. It demonstrates that all series have at least one structural break in their unconditional variance dynamics.

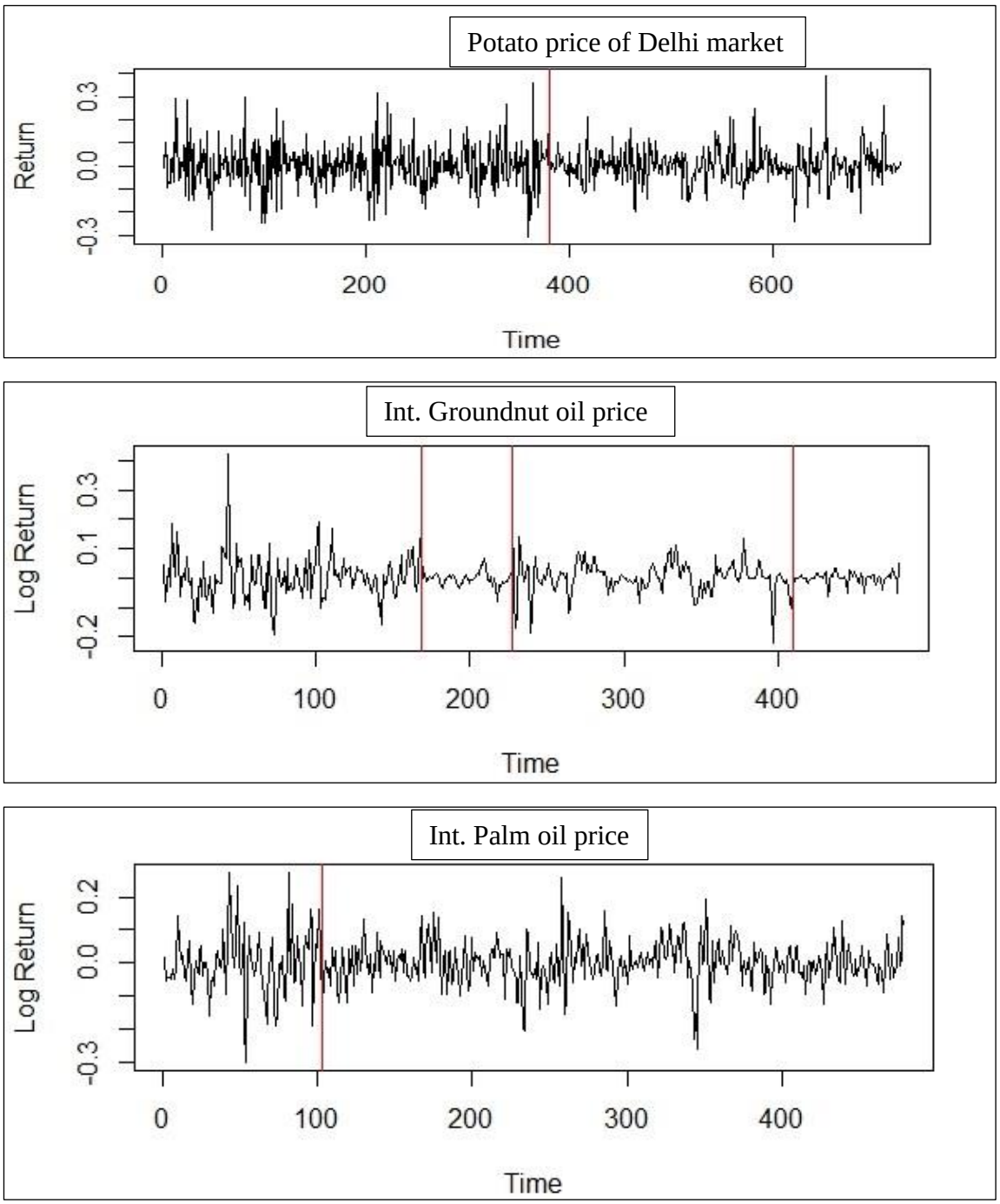


Figure 4.2: Test of structural breaks using PELT algorithm

4.1.3 Fitting of the ARIMA model

(a) Identification of the AR and MA order

The most important method to diagnose the time series properties of data is the correlogram. The correlogram is a graph that presents one of two statistics: the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). The

ACF measures the correlation between X_t and X_{t+k}, where k is the number of lead periods into the future. The PACF is the correlation between X_t and X_{t+k} after removing the effect of the intermediate observations. To understand the statistical properties, the ACF and PACF are estimated for each price return series. Figure 4.3 and Figure 4.4 shows the ACF and PACF for each price return series respectively. The autoregressive parameter (p) is determined by the PACF value and moving average parameter (q) is determined by the ACF value.

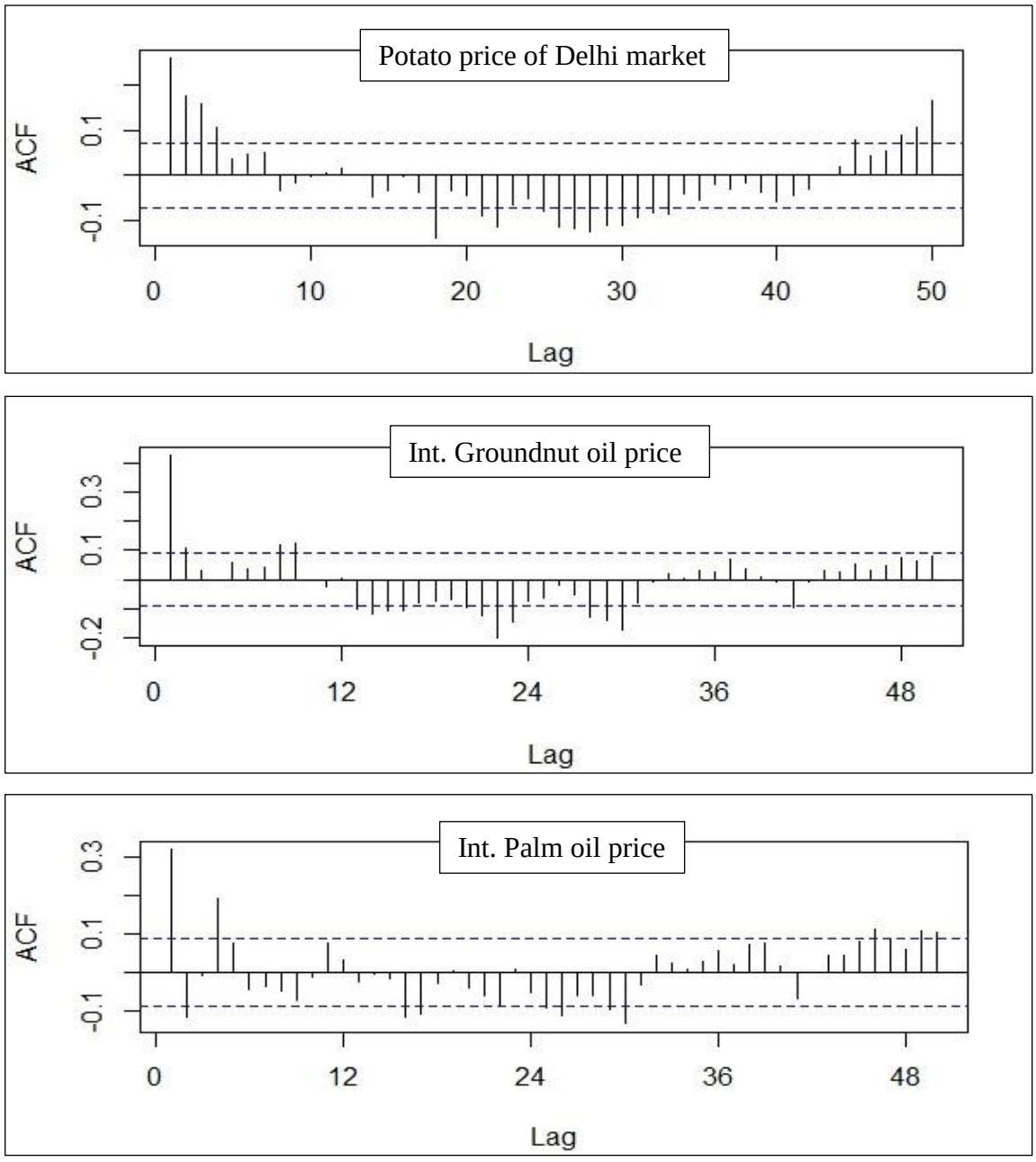


Figure 4.3: Autocorrelation Function (ACF) of return series

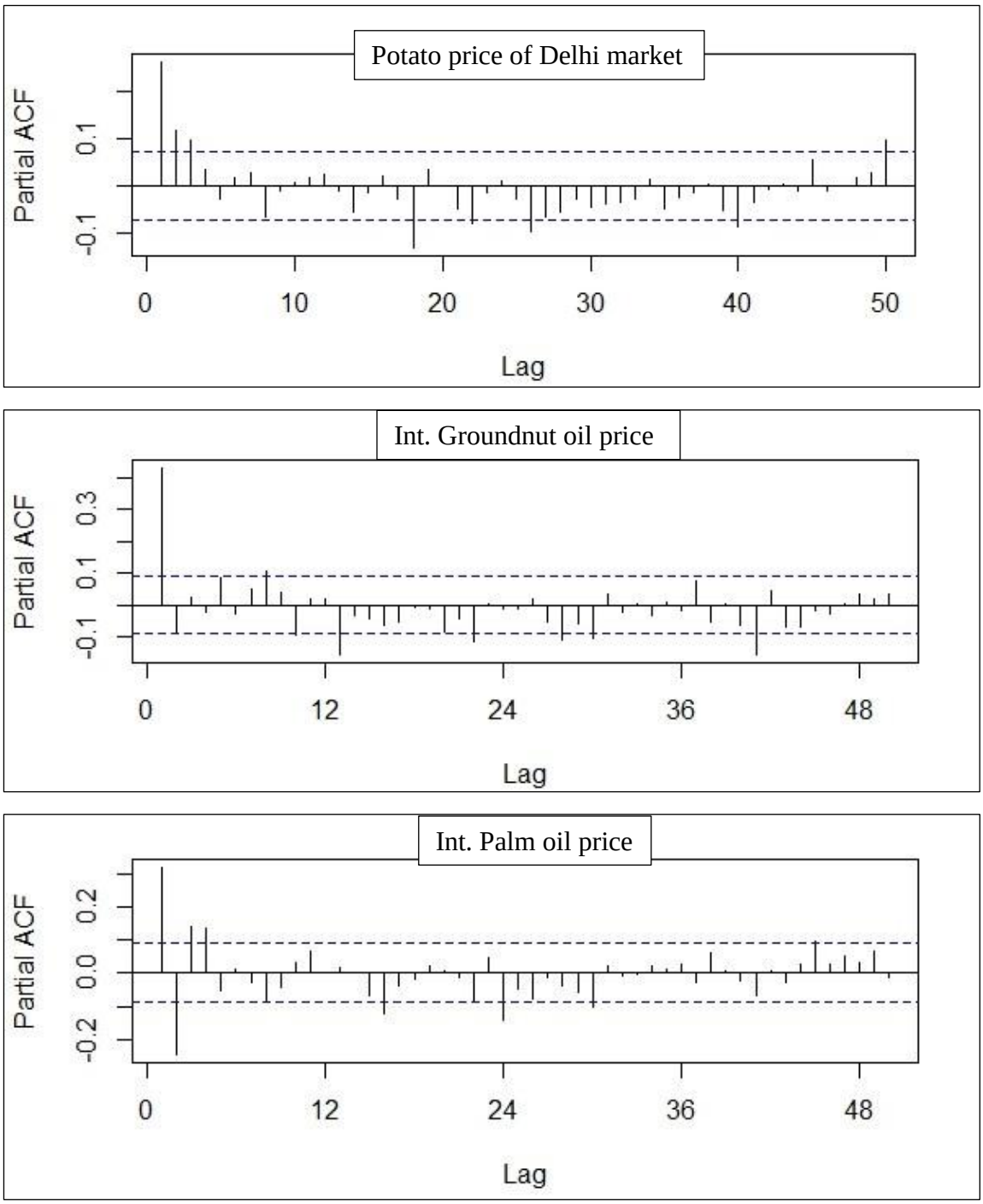


Figure 4.4: Partial Autocorrelation Function (PACF) of return series

(b) Estimation of the Parameters

From the visual inspection of the ACF and PACF in the Figures 4.3 and 4.4, it appears to be some low-order AR or MA processes are occurring in the data. However, we cannot have a confirmative structure of the model based on the correlograms. So, we

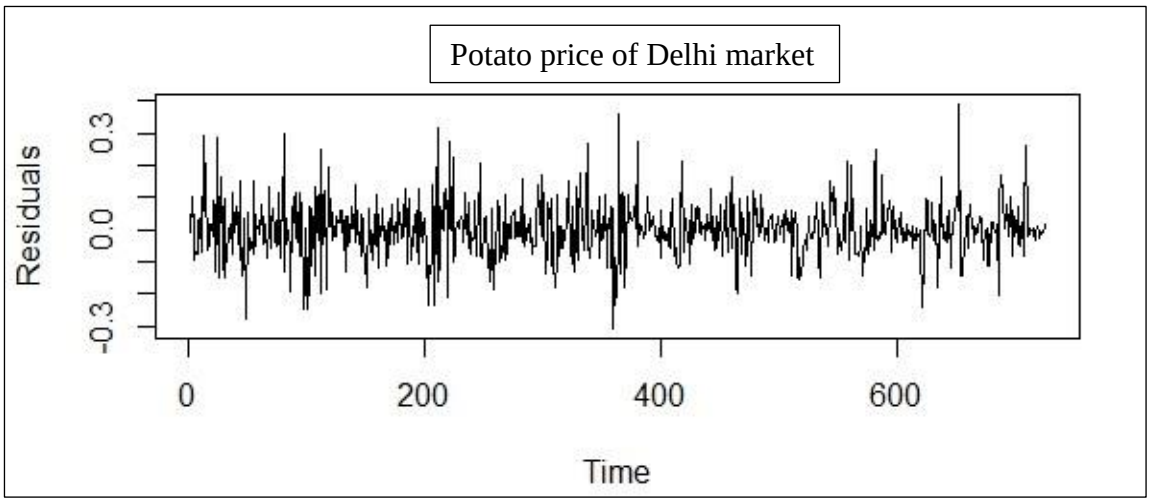
move to the estimation of the model. The parameters are estimated through maximum likelihood function such that an overall measure of errors is minimized or the likelihood function is maximized. We begin by proposing some simple models of each series, and then move to higher order model to assess model fit and adequacy. We have tried all possible combinations of *p* and *q* from order 0 to 3. Based on the minimum AIC value and parsimony of the model, the best model is selected. The details of the best-selected ARIMA model for each series are given in the Table 4.4. It is found that there is seasonal component in international Groundnut price series, but it is not present in the other two series.

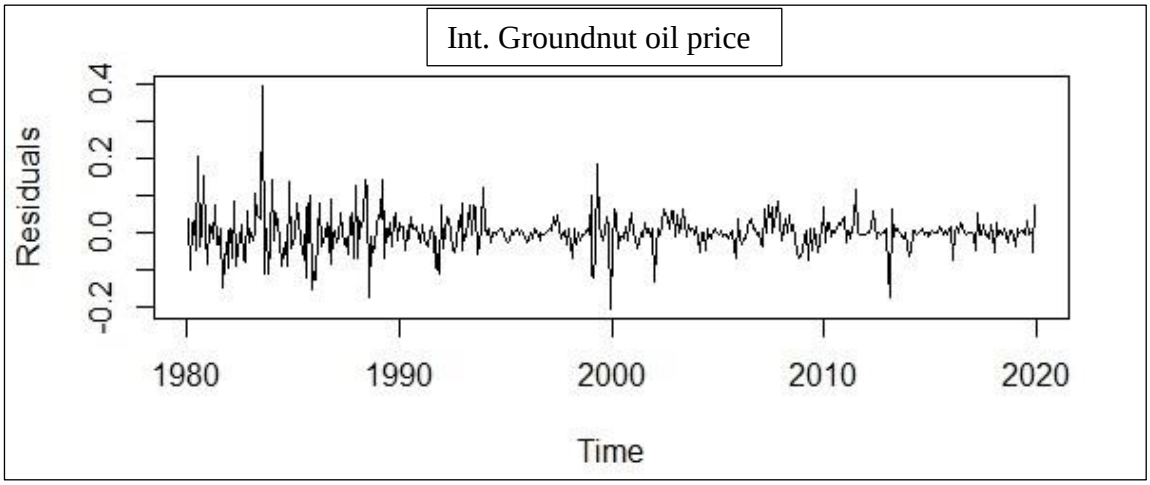
Table 4.4: Parameter estimates of the ARIMA model on the return series

Series	Model	Parameters					AIC
		AR-1	AR-2	MA-1	SMA-1	SMA-1	
Delhi Potato	ARIMA(1,0,1)	0.72 (0.07)	NA	-0.50 (0.08)	NA	NA	-474.03
Int. Groundnut oil	ARIMA(1,0,1) (1,0,1)[12]	0.25 (0.10)	NA	0.23 (0.10)	-0.04 (0.46)	0.11 (0.32)	-496.07
Int. Palm oil	ARIMA(2,0,1)	0.18 (0.11)	-0.18 (0.05)	0.23 (0.10)	NA	NA	-249.90

(c) Residual Diagnostic

The plot of the residuals and squared residuals are given in the Figure 4.5 and 4.6 respectively, for each price series. Before applying the test for ARCH effect to the residual series, we check 'goodness of fit' of the ARIMA model. For this, we inspect the autocorrelation function of the residuals. The ACF of the residuals in Figure 4.7 shows that there are some small spikes are present in the plots. To confirm the autocorrelation in the residuals, we employed Ljung-Box test. The details of the Ljung-Box test are presented in the Table 4.5. The null hypothesis of no correlation in the series of Ljung-Box test is not rejected for long lags, which indicates there is no autocorrelation in the residuals.





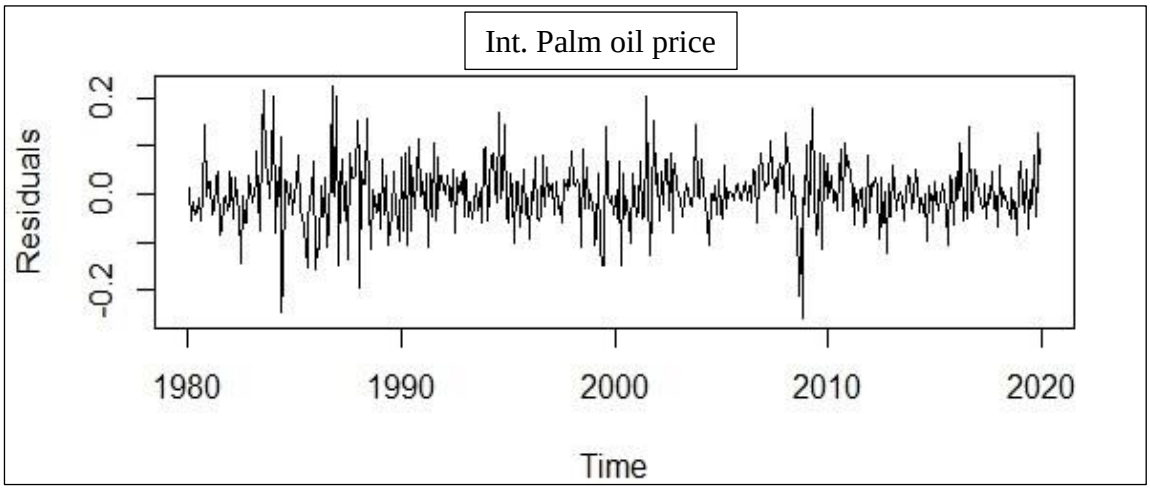


Figure 4.5: Time plot of residuals series

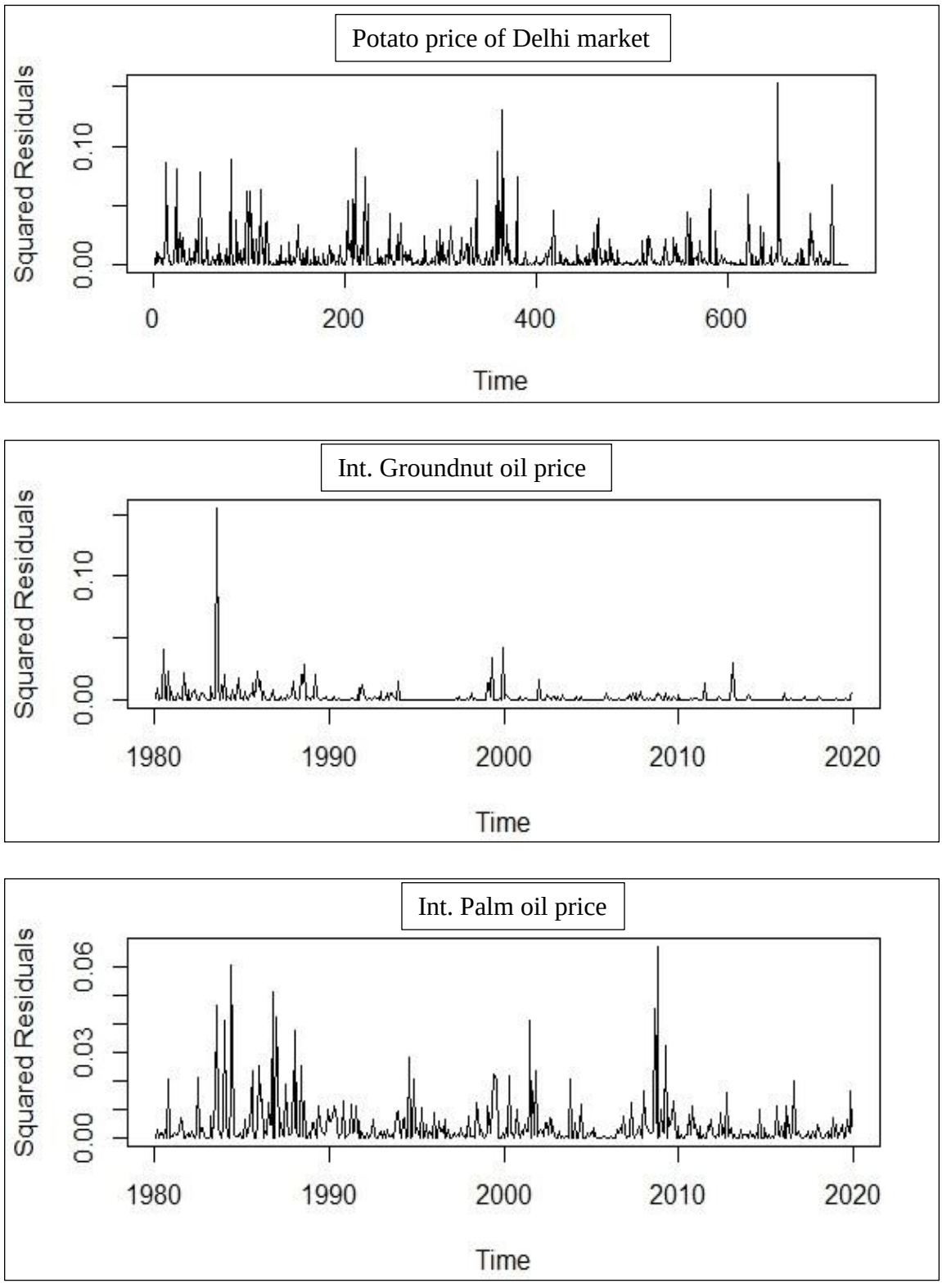


Figure 4.6: Time plot of squared residuals series

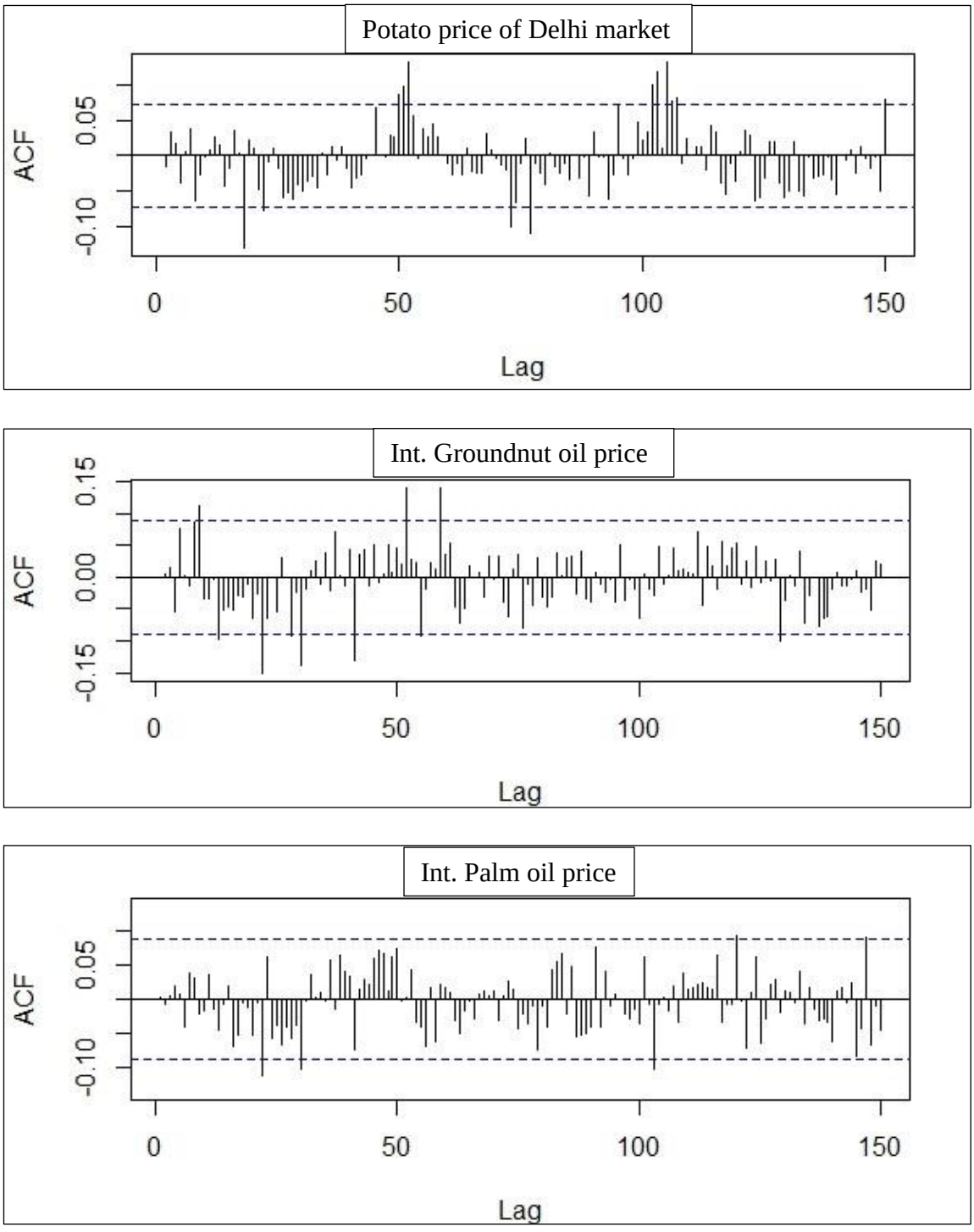


Figure 4.7: ACF plot of residuals of the fitted ARIMA model

However, results of Ljung-Box test show, residuals are statistically uncorrelated but they seem to be not identically distributed from visual inspection of the plot of the residuals in Figure 4.5 that is, the residuals are not independent and identically

distributed through time. There is a tendency that small (large) absolute values of the residual process are followed by other small (large) values of unpredictable sign, which is a common behaviour of GARCH processes. Granger and Andersen (1978) pointed out that some of the series modelled by Box and Jenkins (1970) exhibit autocorrelated squared residuals even though the residuals themselves doesn't seems to be autocorrelated over time. Therefore, they suggested that the ACF of the squared residuals could be useful in identifying nonlinear pattern. Bollerslev (1986) mentioned that the ACF and PACF of squared process are important in checking and identifying GARCH behaviour. Figure 4.8 and Figure 4.9 show the ACF and PACF of the squared residual series of the best fitted ARIMA model for each price series. It indicates, although the residuals seem to be uncorrelated, the squared residual series are autocorrelated. In other words, we can say variance of residual series is conditional on its past history *i.e.* the residual series may exhibit an ARCH effect. In this study, to check statistically ARCH effect in the process, we used Engle's Lagrange Multiplier test. The detail results of the Engle's Lagrange Multiplier test are given in the Table 4.6. The rejection of null hypothesis of 'No ARCH effect' at long lags provides strong evidence for the existence of ARCH effect in series under consideration.

Table 4.5: Ljung-Box test results for residuals

	Delhi Potato		Int. Groundnut oil		Int. Palm oil	
Order	Statistic	p-value	Statistic	p-value	Statistic	p-value
06	01.85	0.93	02.15	0.90	01.00	0.98
12	07.33	0.83	14.57	0.26	03.29	0.96
18	21.16	0.27	24.03	0.15	08.10	0.90
24	28.22	0.25	38.98	0.04	19.04	0.74

Table 4.6: Engle's Lagrange Multiplier test results for the ARCH effect

	Delhi Potato		Int. Groundnut oil		Int. Palm oil	
Order	Statistic	p-value	Statistic	p-value	Statistic	p-value
4	435.6	<0.001	1017	<0.001	284.1	<0.001
8	207.6	<0.001	482	<0.001	116.8	<0.001
12	131.5	<0.001	304	<0.001	72.8	<0.001
16	95.3	<0.001	201	<0.001	47.3	<0.001
20	74.0	<0.001	155	<0.001	37.1	<0.001
24	58.8	<0.001	115	<0.001	28.7	<0.001

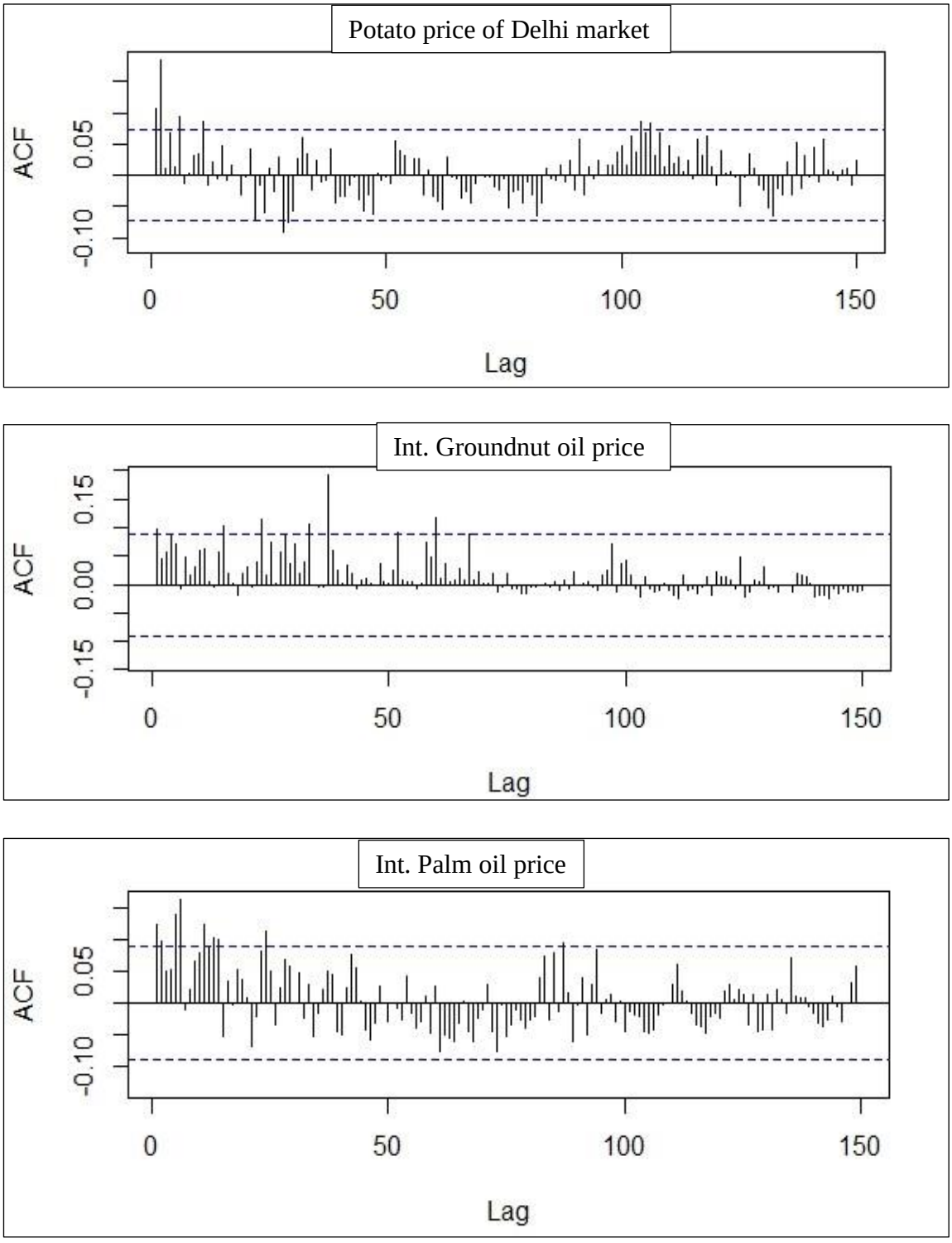


Figure 4.8: ACF plot of squared residuals of the fitted ARIMA model

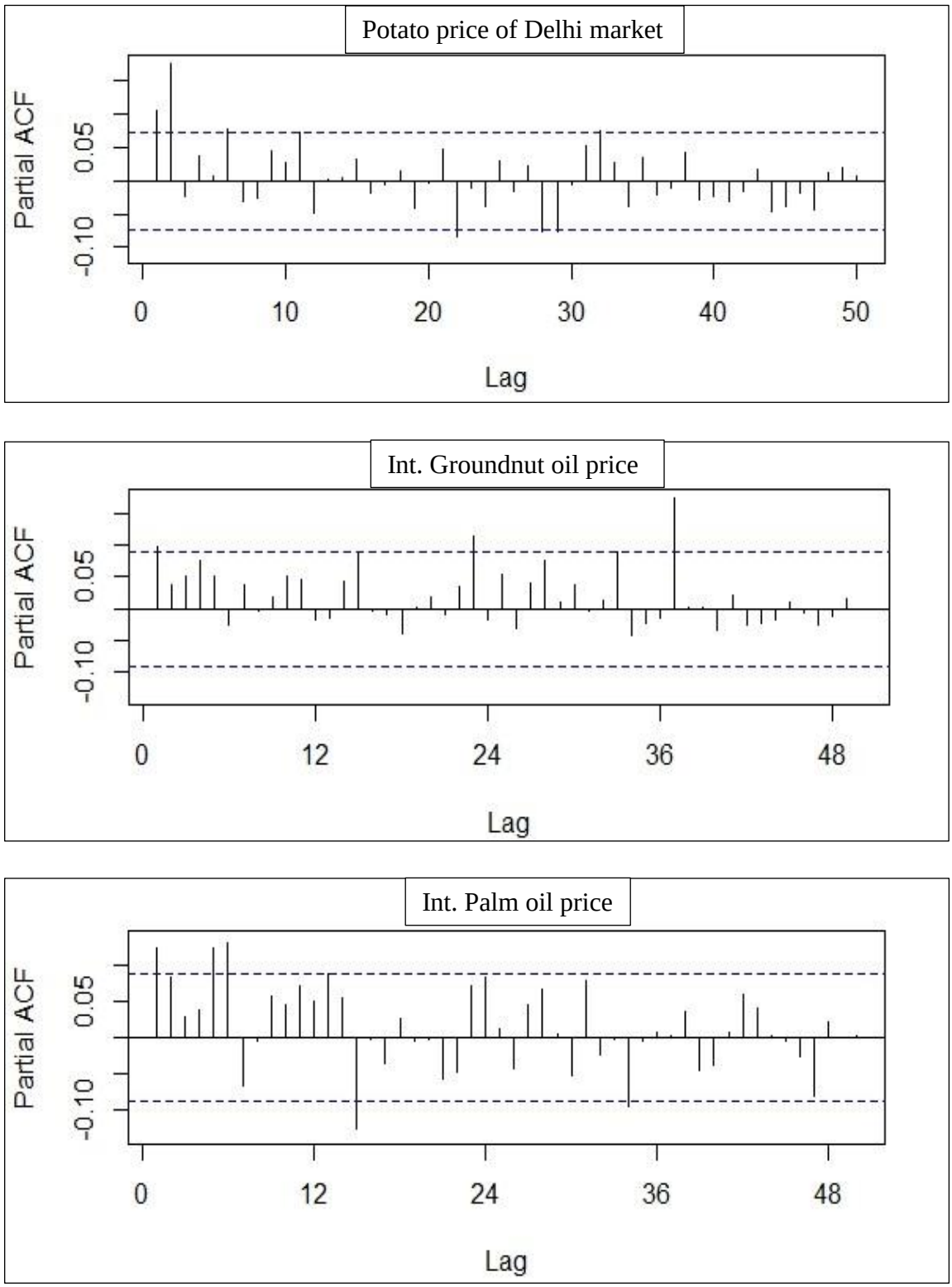


Figure 4.9: PACF plot of squared residuals of the fitted ARIMA model

4.1.4 Fitting of GARCH model

As mentioned earlier, the squared residuals from all three series exhibited heteroscedastic structure. Therefore, to model ARCH effect of the residual series, GARCH model is fitted to all the three residuals series. Under the GARCH(*p, q*) model structure, the conditional variance of the residuals at time *t*, depends on the squared residuals in the previous *q* time steps, and the conditional variance in the previous *p* time steps. Since GARCH models can be treated as ARMA process for squared residuals, the order of GARCH can be determined with the method for selecting the order of ARMA process, and model selection criteria, such as Akaike Information Criterion (AIC) can be used for selecting the appropriate model. In this study, GARCH models are estimated using maximum likelihood. The log-likelihood function is computed from the product of all conditional densities of the prediction errors. For all the three series, GARCH(1,1) model was identified to be the best model on the basis of in-sample performance. The estimates of the parameters of the GARCH (1,1) model along with their standard errors in brackets for individual series are given in Table 4.7.

Table 4.7: Parameter estimates of the best fitted GARCH model

Series	Model	Omega	Alpha1	Beta1
Delhi Potato price	GARCH(1,1)	0.02 (0.01)	0.21 (0.06)	0.53 (0.13)
Int. Groundnut oil	GARCH(1,1)	0.89 (0.35)	0.23 (0.06)	0.76 (0.05)
Int. Palm oil	GARCH(1,1)	0.22 (0.11)	0.10 (0.03)	0.85 (0.04)

4.1.5 Markov switching GARCH model

Plots of the residuals in Figure 4.5 clearly show that large peaks are more frequent at the starts, and at some places at the middle of the series. This indicates that the conditional variance is time varying according to a regime switching specification. In this study, we used two state Markov switching GARCH model. We tried all possible combination of three-variance model specification namely "sGARCH", "eGARCH", "gjrGARCH" and three distribution specification namely "norm", "std", "snorm" for each series. Based on

the minimum AIC value of the in-sample performance, we select the best combination of variance models and distributions (Table 4.8). The models are estimated through maximum likelihood estimation technique. The detail results of the parameter estimates are presented in the Table 4.9a, Table 4.9b, and Table 4.9c for the Potato price of Delhi market, International Groundnut oil and International Palm oil price series respectively. Parameter estimates show that the evolution of the volatility process is heterogeneous across two regimes.

Table 4.8: Best variance model and distribution specification of MS-GARCH model

Series	Variance Specification	Distribution Specification
Delhi Potato price	(sGARCH, eGARCH)	(norm, std)
Int. Groundnut oil	(sGARCH, gjrGARCH)	(norm, std)
Int. Palm oil	(sGARCH, gjrGARCH)	(norm, norm)

Table 4.9a: Parameter estimates of the MS-GARCH model for Potato price of Delhi market

Parameter	Estimate	Std. Error	p-value
Alpha0_1	0.01	0.01	0.04
Alpha1_1	0.11	0.08	0.02
Beta_1	0.31	0.18	<0.001
Alpha0_2	0.01	0.01	<0.001
Alpha1_2	0.10	0.01	0.02
Beta_2	0.81	0.07	0.03
P_1_1	0.78	0.06	<0.001
P_2_1	0.23	0.08	<0.001

Table 4.9b: Parameter estimates of the MS-GARCH model for international Groundnut oil price series

Parameter	Estimate	Std. Error	p-value
Alpha0_1	0.01	0.001	<0.001
Alpha1_1	0.02	0.01	0.03
Beta_1	0.97	0.01	0.02
Alpha0_2	0.21	0.02	<0.001
Alpha1_2	0.15	0.01	<0.001

Beta_2	0.81	0.01	0.05
P_1_1	0.77	0.02	<0.001
P_2_1	0.08	0.02	<0.001

Table 4.9c: Parameter estimates of the MS-GARCH model for international Palm oil price series

Parameter	Estimate	Std. Error	p-value
Alpha0_1	0.17	0.04	0.04
Alpha1_1	0.05	0.01	<0.001
Beta_1	0.91	0.01	<0.001
Alpha0_2	0.39	0.05	0.02
Alpha1_2	0.08	0.04	0.02
Beta_2	0.85	0.06	<0.001
P_1_1	0.95	0.11	<0.001
P_2_1	0.09	0.07	<0.001

4.1.6 Artificial Neural Network Training and Forecasting

With the advancement and evolution of the soft-computing technologies, artificial intelligence techniques have emerged in the price volatility forecasting to enhance the capacity in handling the significant randomness and non-stationarity in the data. Various Artificial Neural Network (ANN) methods have successfully been applied in the agricultural price volatility forecasting (Lama *et al.* 2016, Wu *at al.* 2017). When, we design a robust ANN based forecasting model, it is very important to consider pertinent factors such as number of input nodes, number of hidden layers, number of nodes per hidden layer, selection of best training algorithm and the activation function, and the number of output nodes. These factors are selected in such a way that the optimum model ensures the smallest generalization error and best performance assessment matrices. The number of hidden layers influences the complexity level of the input-output relationship. Using too many layers can cause overfitting, while too few leads to under-fitting. As neural networks, being a universal approximation can map any non-linear function with one hidden layer given sufficient number of nodes at hidden layer and adequate data points for training. Accordingly, in this study, the ANN with one hidden layer has been used.

There is no 'rule-of-thumb' to determine the number of input nodes, which are lagged observation of the same variable. In the literature, it is suggested that input nodes can be determined with the help of PACF. Therefore, we tried 2 to 10 input nodes for each series. The optimum input nodes for Potato price of Delhi market, international Groundnut oil, and international Palm oil price were 7, 4 and 6 respectively. The optimum number of nodes at hidden layer improves the out-of-the sample forecasting ability and avoids the over-fitting problem. If we use too few hidden nodes then it will lead to under-fitting, while too many nodes will lead to overfitting of the model. The number of hidden nodes is determined with help of experimentation. We varied the number of hidden nodes from 2 to 15 with basic cross validation method. For Potato price of Delhi market, International Groundnut oil, and International Palm oil price, the optimum number of hidden nodes were found 9, 8 and 8 respectively (Table 4.10). The number of replication is taken to be 50 for each series.

4.1.7 Extreme Learning Machine Training and Forecasting

Extreme Learning Machine (ELM) is based on the improvement of the conventional single layer feed forward neural network with excellent generalization performance (Xiong *et al.,* 2018). In ELM, input weights and biases are randomly generated, and its output weights are analytically calculated. The critical idea behind ELM is to transform difficult issues arising from nonlinear optimization, like the optimal determination for input weights, hidden layer biases, output weights, to a simple least square problem of deciding the optimal output weights. It means the users do not have to consider all the input weights and the hidden layer biases as long as the norm of weights is small enough, while the output weights are the only issue has to take care. This idea is completely different from the classic iterative learning techniques.

The basic architecture of the Extreme Learning Machine is almost similar to the ANN model. Therefore, similar to the ANN, for designing the ELM model three layers were used to build the architecture for forecasting agricultural price volatility. The number of input nodes is determined by the partial autocorrelation function. The number of output nodes is set as one, as the iterated strategy is used for implementing the multi-step-ahead forecasting. A logistic sigmoid function is selected as the activation function. As

discussed above, ELM randomly determines the input weights and hidden biases, which do not require any tuning in the training process. Thus, once these parameters are randomly generated in advance, the number of hidden nodes is determined in a trial-error fashion. We thereby construct fourteen ELM models with various numbers of hidden nodes (varying from 2 to 15). Each ELM model is trained repeatedly 50 times on the estimation samples, and then, the average MSE of each ELM is calculated on the testing set. The ELM that yields the smallest average MSE value is selected as the best model. By doing so, the optimal number of hidden nodes ELM for Potato price of Delhi market, International Groundnut oil, and International Palm oil price, are 8, 9 and 7 respectively.

Table 4.10: Neuronal arrangements for the ELM and the ANN models

Series	Model	Neuron structure
Delhi Potato	ANN	7 – 9 – 1
	ELM	7 – 8 – 1
Int. Groundnut oil	ANN	4 – 8 – 1
	ELM	5 – 9 – 1
Int. Palm oil	ANN	6 – 8 – 1
	ELM	6 – 7 – 1

4.1.8 Proposed MS-GARCH – ELM model

There are ample evidence that the parameters of time series models fitted to economic time series changes over time (Stock and Watson, 1996; Aiolfi and Timmermann 2006). According to Timmermann (2006), model instability as an important determinant of forecasting performance and a potential reason for hybrid models. Moreover, it becomes necessary to use hybrid model because individual model may not handle the inherent data patterns like non-stationarity and non-linearity simultaneously. In order to develop a hybrid model based on Extreme Learning machine (ELM) and Markov Switching GARCH (MS-GARCH) model for agricultural price volatility forecasting, we have employed parallel hybrid structure introduced by Bates and Granger (1969).

Motivated by the work of Lamoureux and Lastrapes (1990), in which he has shown the simple GARCH model tends to significantly overestimate volatility persistence in the

presence of structural breaks, Hamilton and Susmel (1994) proposed MS-GARCH model. In this model, the parameters switches from one regime to another according to an unobservable process that is assumed to follow a first-order Markov process. MS-GARCH model removes the bias of GARCH effects by allowing the coefficient on the unconditional volatility level to switch between regimes. The specification of MS-GARCH model allows periods of low and high unconditional volatility to be clearly identified, and it is robust against changes in market conditions and structural breaks. In view of the above advantages of MS-GARCH model in capturing volatility in the data under the assumption of structural break, we select it as one of the prominent candidate model in the hybridization process.

A frequently used machine-learning algorithm in the price volatility forecasting is the artificial neural network. ANN is a powerful and flexible data-driven algorithm designed to capture and represent non-linear and complex relationships present in the data. However, a big challenge encountered by the ANN is the need for iterative tuning of model parameters, slow learning rate and over fitting of the model. To overcome these problems, Huang (2006) proposed Extreme Learning Machine (ELM), which is based on single-hidden layer feedforward neural networks. ELM provides good generalization performance at very fast learning speed. Therefore, to capture the nonlinearity in the data with good generalization, we select ELM as another candidate model.

In this study, we adopt weighting scheme based on non-parametric smoothing techniques, in which for determining the optimum weights for the weighted average is based on kernels. In the kernel-based technique, the functional form of the weight function is not assumed to be known and it is non-linear. In addition, it is not assumed that the forecasting errors of the individual forecasting models are serially or cross uncorrelated. We make individual predictions of the squared residual series of ARMA process, both in the training and testing sets. Hence, we have in-sample and out-of-sample forecast value. The forecast of the conditional variance from the fitted MS-GARCH as well as squared residuals from ELM for each series were combined using Nadaraya–Watson kernel weight function. We used Gaussian kernel, which determines the shape of the kernel weights. The size of the weight is parameterized by the bandwidth, *h*. The bandwidth is chosen such that there is a trade-off between bias and

variance. The smaller bandwidth provide weights more concentrated around the forecast series. We employed cross validation and the leave-one-out method to estimate the bandwidth for each series. The forecasting ability of individual as well as hybrid models were assessed though Root mean squared error (RMSE), Theil U-statistic and Correct Directional Statistics (CDS). We take four individual model viz. GARCH, MS-GARCH, ANN and ELM along with the hybrid model for the evaluation of the performance. We evaluate the forecasting performance of the each model for both in-sample (training set) and out-of-sample (testing set). The detail results of in-sample evaluation and out-of-sample evaluation are presented in the Table 4.11 and Table 4.12 respectively for three price series. The results revealed that the both in-sample and out-of-sample, the proposed hybrid model outperform the individual models in all cases, since it has consistently lower RMSE and Theil-U and higher CDC value. The performance of the MS-GARCH model is better compare to the GARCH model, it may be due to MS-GARCH model capture the regime shift in the data. In the line of the earlier previous study of the forecasting agricultural price volatility as Lama *et al.* (2016), this study also confirm that the performance of the ANN model is better than the GARCH model. However, the overall performance of the ELM is better or at par compared to the ANN.

Table 4.11: In-sample forecasting performance of GARCH, ANN, ELM, MS-GARCH and Hybrid model

	GARCH	ANN	ELM	MS-GARCH	Hybrid
Delhi Potato price series					
RMSE	7.45	5.56	5.02	6.20	4.60
CDC	59.20	63.93	63.90	62.00	65.29
Theil-U	0.37	0.33	0.33	0.34	0.28
Int. Groundnut oil price					
RMSE	2.04	1.30	1.25	1.50	1.15
CDC	63.76	65.32	65.87	64.90	66.12
Theil-U	0.31	0.28	0.27	0.30	0.25
Int. Palm oil price					
RMSE	8.34	6.40	5.90	6.90	5.40
CDC	47.36	50.19	51.26	49.50	53.17
Theil-U	0.41	0.35	0.35	0.37	0.35

Note: RMSE value should be multiplied by 10^{-3}

Table 4.12: Out-of-sample forecasting performance of GARCH, ANN, ELM, MS-GARCH and Hybrid model

	GARCH	ANN	ELM	MS-GARCH	Hybrid
Delhi Potato price series					
RMSE	69.21	60.36	56.84	62.23	55.71
CDC	52.34	55.40	55.30	54.22	57.78
Theil-U	0.45	0.40	0.41	0.42	0.38
Int. Groundnut oil price					
RMSE	18.54	14.90	14.30	15.50	13.85
CDC	56.24	59.13	59.90	58.56	61.40
Theil-U	0.39	0.35	0.34	0.36	0.34
Int. Palm oil price					
RMSE	42.52	35.58	36.67	37.50	35.12
CDC	43.14	46.21	46.97	45.32	46.50
Theil-U	0.50	0.47	0.46	0.47	0.44

Note: RMSE value should be multiplied by 10^{-2}

4.2 Results of Long Memory models with structural break

4.2.1 Data Description and Stochastic Properties

For the present investigation, daily mustard price (₹/*q*) of Agra and Bharatpur market from 1st January 2016 to 31st January 2020, are collected from Agricultural Marketing Information Network (AGMARKNET), (https://agmarknet.gov.in/) website. The time plot of the mustard price series of Agra and Bharatpur markets are provided in Figure 4.10a and Figure 4.10b respectively. The descriptive statistics of the mustard price for both markets are reported in the Table 4.13. The average mustard price of Agra and Bharatpur markets are 3696 ₹/*q* and 3735 ₹/*q* respectively. There is not much difference in the average price of Agra and Bharatpur market, this can be due to the fact that both are production markets of mustard. Table 4.13 indicates, the value of Standard Deviation, Minimum, Maximum, Coefficient of variation, Skewness and Kurtosis are slightly higher in the Bharatpur market compare to the Agra market. Jarque-Bera test result shows that both series are not normally distributed. In this study, to stabilize the variance, we applied logarithmic transformation to the data.

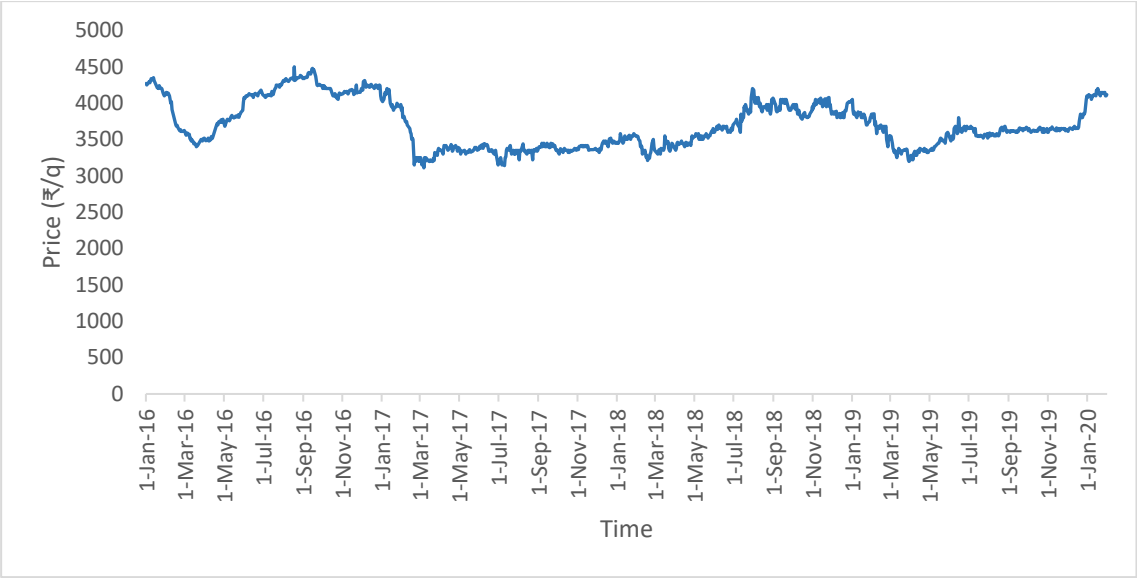


Figure 4.10a: Time plot of daily Mustard price of Agra market (01/01/2016 to 31/01/2020)

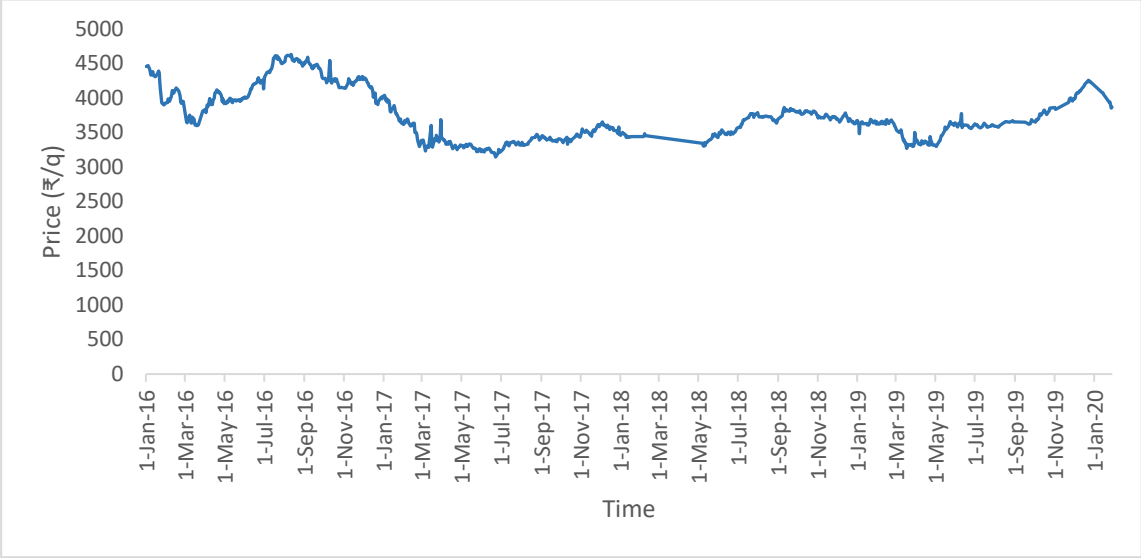


Figure 4.10b: Time plot of daily Mustard price of Bharatpur market (01/01/2016 to 31/01/2020)

Table 4.13: Descriptive statistics of Agra and Bharatpur market Mustard price series (01/01/2016 to 31/01/2020)

	Agra	**Bharatpur**
Mean *(₹/q)*	3696.00	3735.00
Std. Dev. *(₹/q)*	322.71	355.21
Minimum *(₹/q)*	3110.00	3143.00
Maximum *(₹/q)*	4500.00	4630.00
CV (%)	8.71	9.40

Skewness	0.46	0.69
Kurtosis	2.07	2.62
Jarque-Bera test (p value)	<0.001	<0.001

(*Source: Agricultural Marketing Information Network (AGMARKNET), https://agmarknet.gov.in/*)

The Autocorrelation Function (ACF) of the mustard price series of Agra and Bharatpur markets are shown in Figure 4.11a and Figure 4.11b respectively. Autocorrelation functions of both series are highly persistent over long lag and decaying very slowly towards zero (Hyperbolic rate), which indicates the possible presence of long memory properties in the data.

For selecting an appropriate technique for modelling and forecasting of the data, it is important to check whether the time series data under consideration is linear or not. If there is a solid evidence of nonlinearity in the dynamics of the data generating process, then in addition to linear models, nonlinear models should also be tried for forecasting of the data. To test the linearity of the series, we used Brock-Decher-Scheikman (BDS) test. The detail results of the BDS test are given in the Table 4.14, which indicates both price series are non-linear. In other words, if linear model is applied to mustard price series of both markets then some hidden structure is left unaccounted in the residuals of the fitted model. Thus, non-linear model can be more suitable for the forecasting of mustard price of both markets.

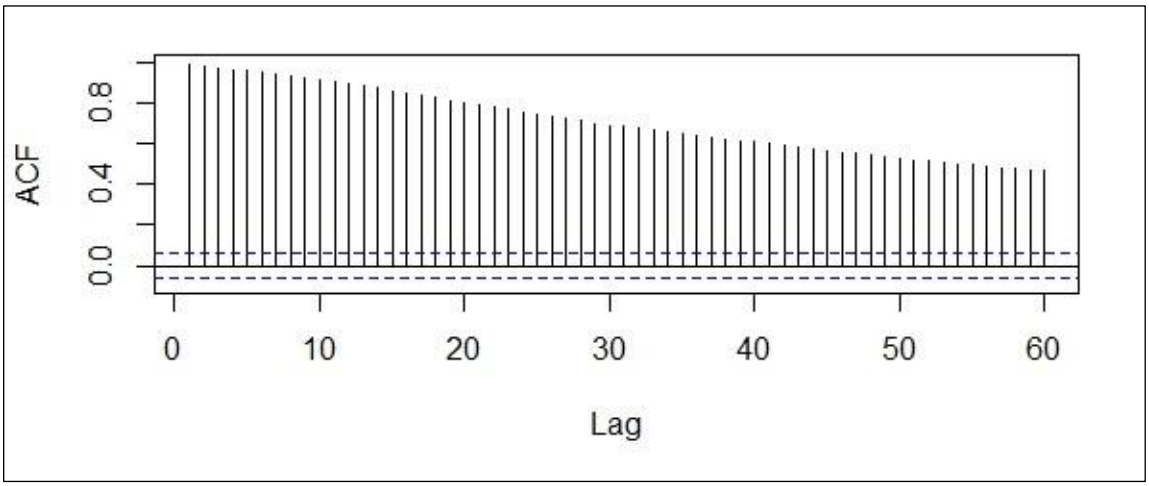


Figure 4.11a: ACF plot of Mustard price of Agra market

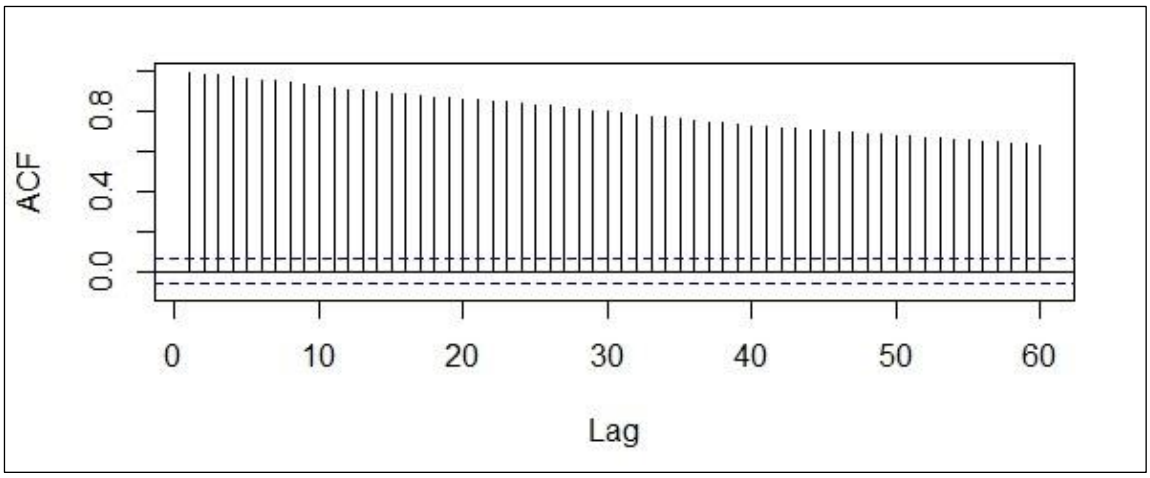


Figure 4.11b: ACF plot of Mustard price of Bharatpur market

Table 4.14: Brock-Decher-Scheikman (BDS) test results

Series		Embedding dimension				Conclusion
		2		3		
	Epsilon	Statistics	p-value	Statistics	p-value	
Agra	0.5σ	314.52	< 0.001	576.00	< 0.001	Non Linear
	σ	184.88	< 0.001	234.48	< 0.001	
	1.5σ	129.24	< 0.001	141.53	< 0.001	
	2σ	102.41	< 0.001	102.57	< 0.001	
Bharatpur	0.5σ	217.86	< 0.001	399.58	< 0.001	Non Linear
	σ	116.44	< 0.001	145.03	< 0.001	
	1.5σ	90.87	< 0.001	97.56	< 0.001	
	2σ	79.10	< 0.001	78.31	< 0.001	

4.2.2 Test for Long Memory and Structural Breaks

Long memory is a notable empirical feature of many financial time series data. Testing for long memory property is an important task since any evidence of long memory would support the use of long memory based models such as ARFIMA, ARFIMA-FIGARCH, *etc*. To test the long memory, we employ the semiparametric GPH (Geweke and Porter-Hudak, 1983) test to the mustard price of both markets. The obtained results are reported in Table 4.15. The GPH estimate of fractional integration parameter (d) are 0.14 and 0.16 for Agra and Bharatpur market respectively and are significant at 5% level of significance. To test the possible structural breaks in mustard price series in the mean,

we used PELT algorithm. Figure 4.12 shows the plot of the price series along with the break point detected through the PELT algorithm. It is found that there are three break points in the Agra market and two break points in the Bharatpur market (Table 4.15).

Table 4.15: Results of GPH long memory test and change point detection

	Agra	Bharatpur
d	0.14*	0.16*
Break points	08/02/17, 26/04/18, 02/03/19	06/02/17, 20/04/18

(* Significant at 5% level of significance)

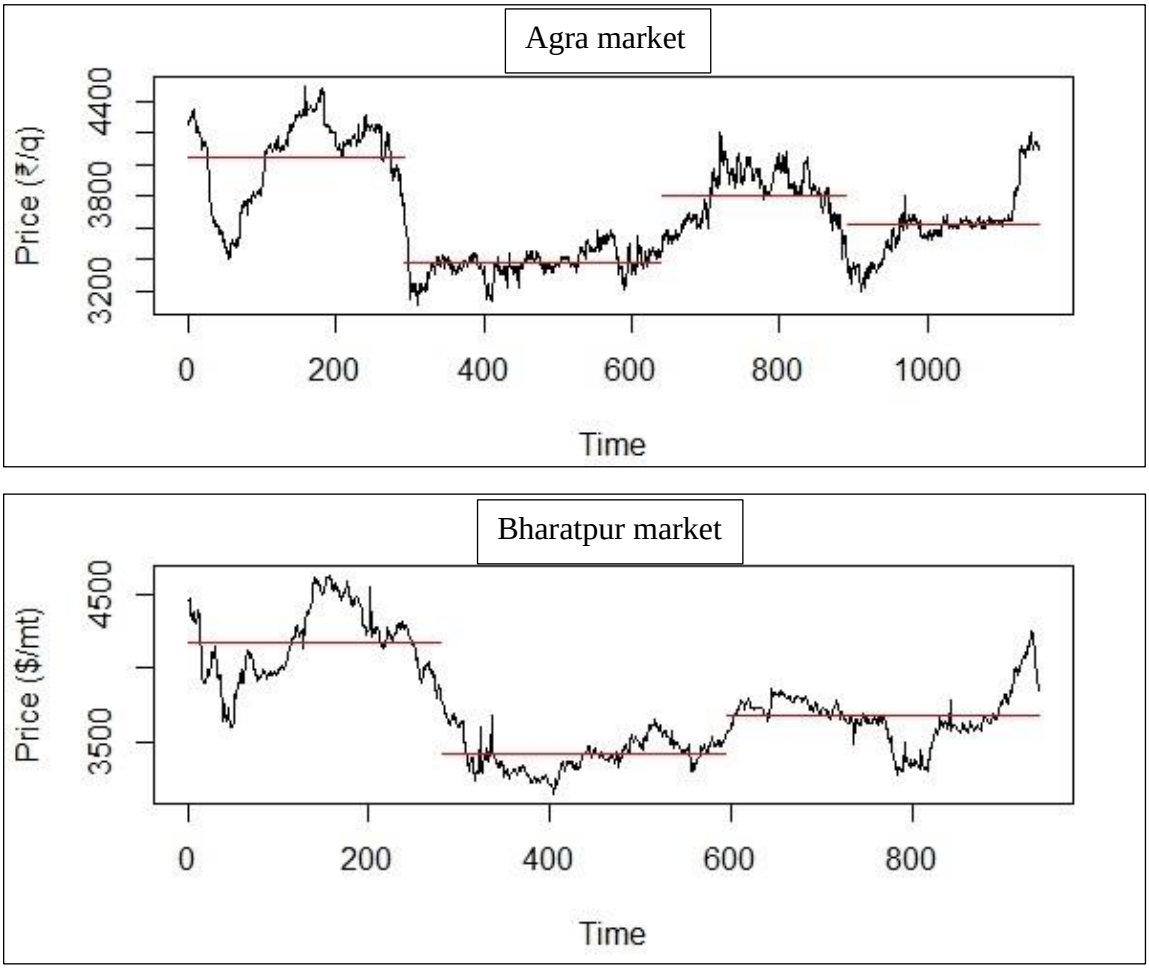


Figure 4.12: Plot of change point detection using PELT algorithm

Long memory and structural breaks are often confused. Several study shows that structural breaks can generate a spurious long memory component in the time series data (Granger and Hyung, 2004; Choi and Zivot, 2007). Therefore, we also investigated the possibility of possible spurious long memory in the data series. To test true long memory

against spurious long memory, we used Qu test (Qu, 2011). The test statistics of Qu test were 0.80 and 0.84 for Agra market and Bharatpur market respectively, and both are not significant at 5% level of significance. The results of Qu test confirms the presence of true long memory in both the series.

4.2.3 Proposed Hybrid Model based on ARFIMA with dummy variable and ELM

The proposed methodology for the modelling and forecasting long memory process with structural break is motivated by the fact that agricultural price data often contain both linear and non-linear pattern and no single model is capable to capture all patterns present in the data. A parsimonious technique to model long-term behaviour of a time series data is by means of an Autoregressive Fractionally Integrated Moving Average (ARFIMA) model. ARFIMA model is a parametric way of capturing long memory dynamics (Granger and Joyeux, 1980; Hosking, 1981). However, ARFIMA model can capture only the linear component present in the data series. To capture the non-linear pattern in the data series, we employed an artificial intelligence technique Extreme Learning Machine (ELM). To account the effect of structural break, ARFIMA model is estimated with dummy variable as a regressor. The proposed method exploits the strength and feature of ARFIMA model as well as Extreme Learning Machine. The hybrid model is constructed in a sequential manner, first ARFIMA model with dummy variable is applied to the original data series and then its residuals are modelled using extreme learning machine.

ARFIMA (p, d, q) is specified by the orders of the autoregressive (p) and the moving average (q) parts of the model, along with the non-integer order of differencing (d). The parameters are estimated through maximum likelihood function such that an overall measure of errors is minimized. We employed the 'trial and error' technique as one effort to minimize the risks of model misspecification. We adopted the Akaike's information criterion (AIC) to determine the appropriate model. The parameter estimates of the best-fitted model are presented in the Table 4.16 for both markets. We found ARFIMA (1, 0.11, 1) and ARFIMA (1, 0.13, 1) are the most adequate model for the Agra and Bharatpur market respectively, based on the AIC value and the concept of model parsimony.

Table 4.16: Parameter estimates of best-fitted ARFIMA model with dummy variable

Series	Parameter	Estimate	S.E	p-value	AIC
Agra	μ	0.01	0.01	0.54	3255.88
	AR – 1	0.15	0.09	0.08	
	MA – 1	-0.39	0.07	0.03	
	xreg	0.07	0.02	0.04	
	d	0.11	0.06	0.01	
Bharatpur	μ	-0.001	0.04	0.66	3066.07
	AR – 1	0.11	0.13	0.05	
	MA – 1	-0.36	0.17	0.03	
	xreg	0.06	0.03	0.03	
	d	0.13	0.10	0.02	

In the next step, residuals are extracted from the best-fitted ARFIMA model. Plots of the residuals obtained from the ARFIMA model are presented in the Figure 4.13. The ACF and PACF plots of the residuals are given in the Figure 4.14 and Figure 4.15 respectively. To assess the existence of non-linear pattern in the residuals, BDS test is applied. The detail results of BDS test on the residuals are given in the Table 4.17, indicate that null hypothesis of iid is rejected at 1% level of significance. The BDS test results suggest, non-linear patterns exist in the residuals of the selected ARFIMA model. Therefore, the obtained residuals are modelled through a non-linear Extreme Learning Machine (ELM) model. In order to find the best extreme learning machine architecture of the residuals, we tested all possible combinations of 1–7 input nodes and 2–10 hidden nodes and each combination was trained 50 times. The average mean square error of each ELM model is calculated on the testing set. The ELM that yields the smallest average MSE value is selected as the best model. Finally, the number of optimal input nodes and hidden nodes were found to be 5 and 7 respectively for Agra market. For the Bharatpur market, three input nodes and five hidden nodes were found to be optimal. Finally, the forecasted value of the ARFIMA with dummy variable and ELM model are summed to get the forecast of the hybrid model.

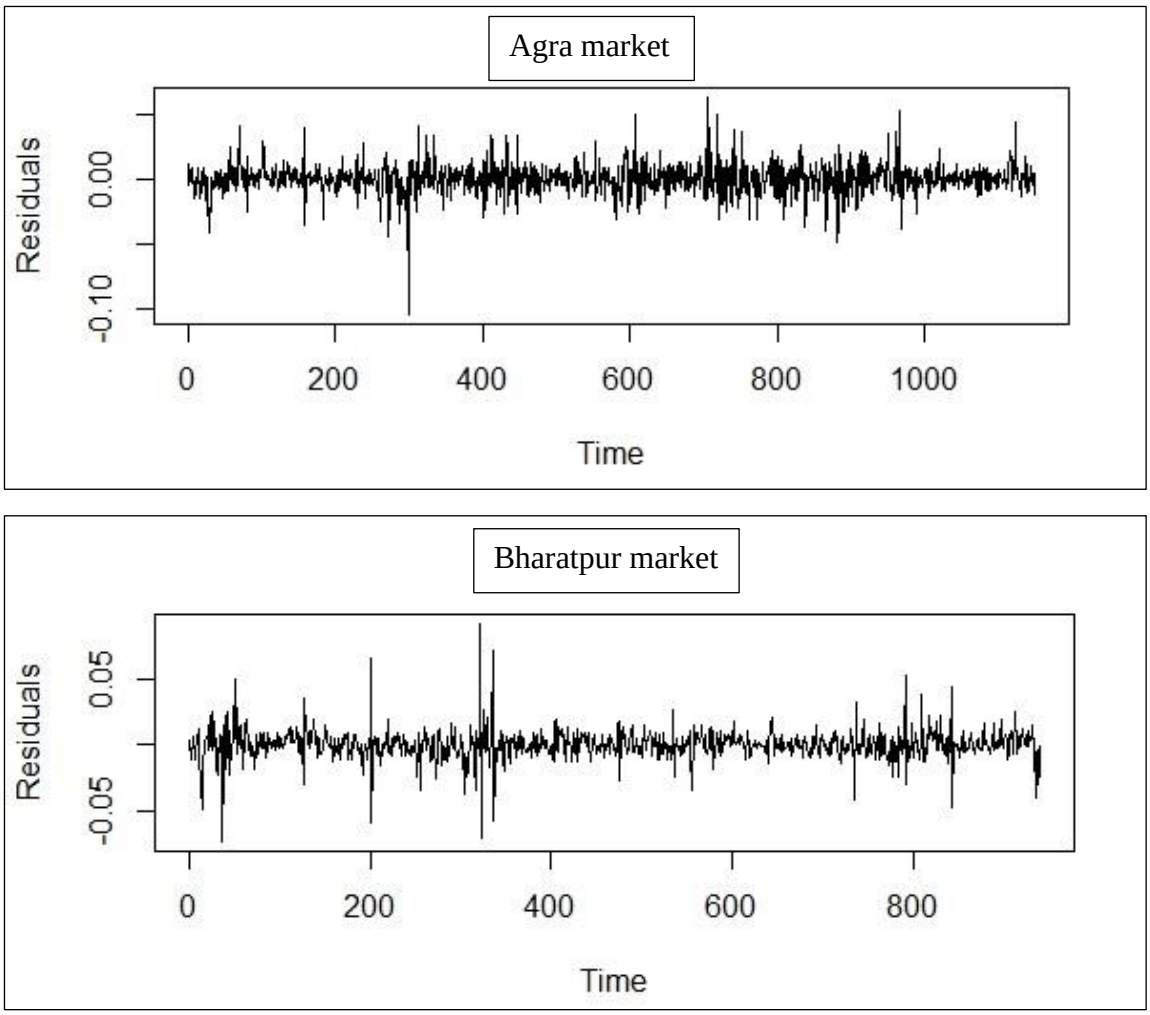


Figure 4.13 Plot of the residuals obtained from the fitted ARFIMA model with dummy variable

Table 4.17: Brock-Decher-Scheikman (BDS) test results on the residuals

Series		Embedding dimension				Conclusion
		2		3		
	Epsilon	Statistics	Probability	Statistics	Probability	
Agra	0.5σ	8.50	< 0.001	10.92	< 0.001	Non Linear
	σ	7.90	< 0.001	9.40	< 0.001	
	1.5σ	7.05	< 0.001	7.98	< 0.001	
	2σ	6.31	< 0.001	6.61	< 0.001	
Bharatpur	0.5σ	7.27	< 0.001	7.67	< 0.001	Non Linear
	σ	7.59	< 0.001	7.87	< 0.001	
	1.5σ	8.29	< 0.001	8.26	< 0.001	
	2σ	8.77	< 0.001	8.12	< 0.001	

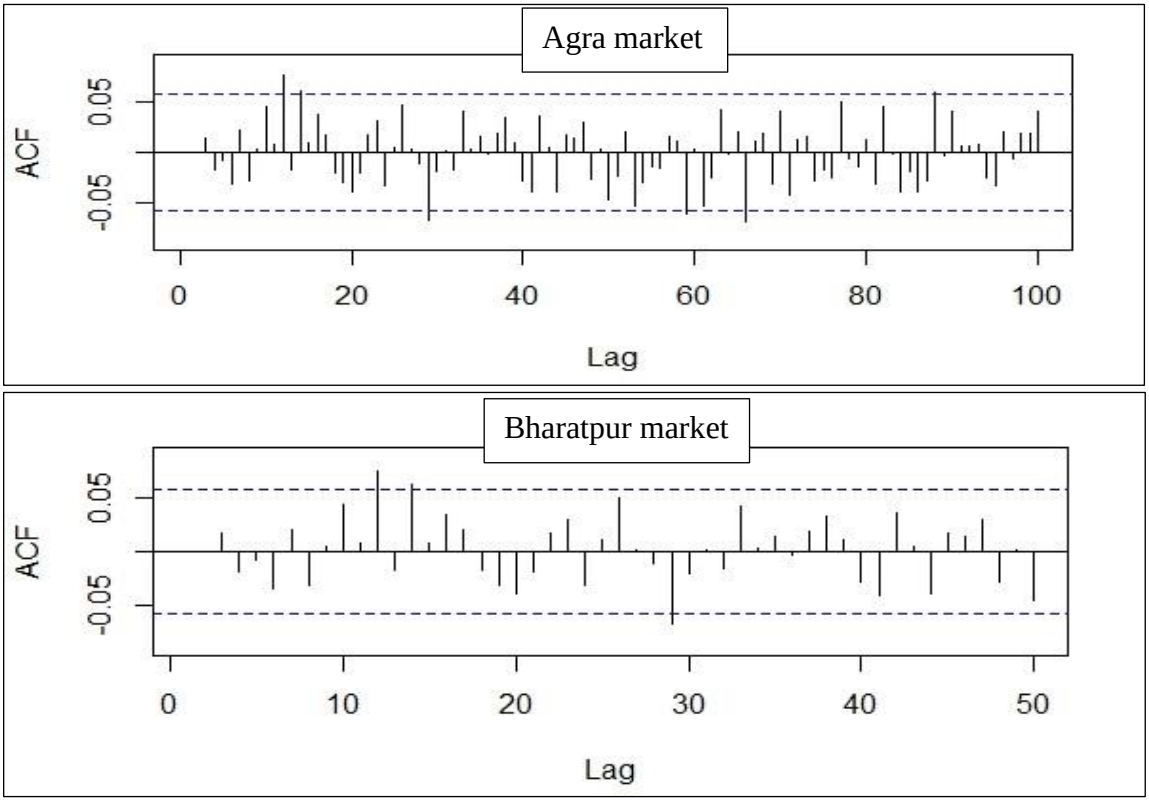


Figure 4.14 ACF plots of the residuals

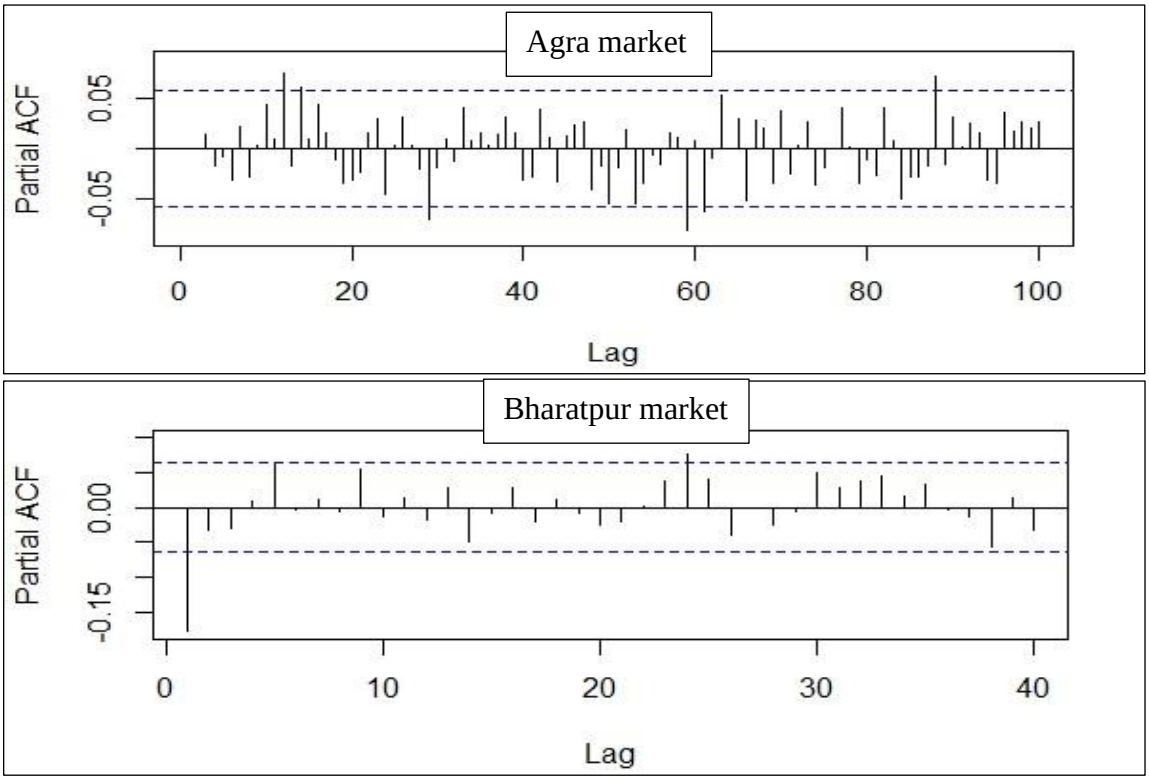


Figure 4.15 PACF plots of the residuals

To evaluate the forecasting performance of the proposed hybrid methodology, we carry out forecasting experiments using ARFIMA with dummy variable and a pure ELM model. In the process of fitting pure ELM to the original data set, we found 6 – 7 – 1 and 4 – 6 – 1 are the optimal architecture for the Agra and Bharatpur market respectively. Using the RMSE and MAE, in sample forecast evaluation results are reported in Table 4.18 and out of sample forecast evaluation are given in the Table 4.19. It shows that the proposed hybrid model outperforms all other competing techniques in terms of prediction accuracy. The forecasting performance of ELM model is better compare to the ARFIMA with dummy variable.

Table 4.18: In-sample forecasting performance of ARFIMA with dummy variable, ELM and Hybrid model

	ARFIMA model with dummy variable	ELM Model	Hybrid Model
Agra			
RMSE	1.02	0.42	0.28
MAE	0.93	0.44	0.19
Bharatpur			
RMSE	1.18	0.49	0.37
MAE	0.84	0.34	0.29

Note: All RMSE and MAE values should be multiplied by 10^{-3}

Table 4.19: Out-of-sample forecasting performance of ARFIMA with dummy variable, ELM and Hybrid model

	ARFIMA model with dummy variable	ELM Model	Hybrid Model
Agra			
RMSE	89.67	31.32	23.56
MAE	73.62	21.87	17.31
Bharatpur			
RMSE	96.65	45.81	36.23
MAE	90.43	41.89	30.74

Note: All RMSE and MAE values should be multiplied by 10^{-3}

4.3 Results of Co-integration with Structural Break

In this section, we investigated the spatial market integration across four major Potato markets viz. Agra, Bangalore, Delhi and Mumbai. Johansen's multivariate co-integration approach has been applied to identify the possible market integration. As we know, structural break can mislead the test results of co-integration. Accordingly, co-integration among these markets were investigated in the absence and presence of structural break.

4.3.1 Data sets and Descriptive Statistics

The data used in the empirical analysis comprises monthly wholesale prices (rupees per quintal, *₹/q*) of Potato of Agra, Bangalore, Delhi and Mumbai markets for the period January, 2005 to December, 2019. Identification of important markets has been based on arrival data of last five years as well as availability of the data. The wholesale prices of Potato series were obtained from National horticultural research and development foundation (NHRDF) (http://nhrdf.org/en-us/) website. The time plot of all the series are depicted in Figure 4.16, which clearly indicate the non-stationarity behaviour of a typical Potato price data.

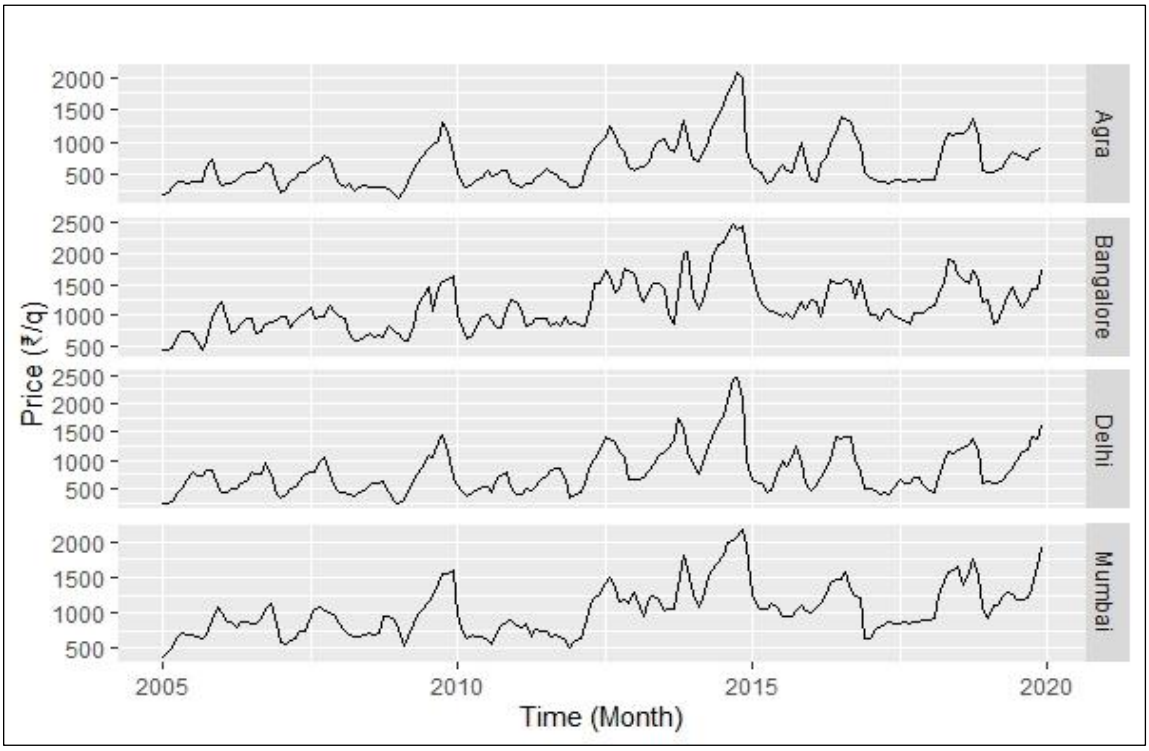


Figure 4.16: Time plot of monthly Potato wholesale prices of major markets

The basic descriptive statistics for all the four markets price series is presented in Table 4.20. The mean price of Potato for the period of January 2005 to December 2019 for the major four markets across India was the lowest at ₹643.43 per quintal in Agra market. It may be due to Agra is a production market and remaining three are consumption markets (Table 4.20). The highest average price was recorded at ₹1111.43 per quintal in Bangalore market. The minimum price was recorded in Agra market, at price of ₹147 per quintal whereas the maximum price was recorded in Bangalore market, at price of ₹2480 per quintal. Coefficient of variation (CV), a crude measure of volatility, indicates that Mumbai market has the lowest price volatility which is represented by 34.56% compared to 56.13% in Agra market, which has the highest price volatility. The price series of all markets appear to follow non-normal distribution and are leptokurtic in nature. Large fluctuations in the prices of a commodity may result in switching over of farmers to some other crops. The stable price level of Potato will provide incentives to the farmers to increase the production and adapt to new technology, which stabilizes the farm income.

Table 4.20: Descriptive statistics of monthly Potato wholesale prices of different markets (January 2005 to December 2019)

Market	Observation No.	Mean (₹/*q*)	Std. Dev. (₹/*q*)	Minimum (₹/*q*)	Maximum (₹/*q*)	CV (%)	Kurtosis	Jarque-Bera	p-value
Agra	180	643.43	361.19	147	2086	56.13	5.74	116.15	<0.001
Bangalore	180	1111.43	417.19	414	2480	37.53	4.24	43.15	<0.001
Delhi	180	783.98	406.67	245	2467	51.87	6.11	131.22	<0.001
Mumbai	180	996.49	344.44	394	2167	34.56	4.11	40.03	<0.001

4.3.2 Test for Stationarity

While investigating market integration, the first step is to check for the evidence of non-stationarity of data in order to confirm that co-integration approach is the appropriate method. In this study to check the non-stationarity of the price series, we used ADF test and PP test (Table 4.21). At level, we have found that all the major Potato price series were non-stationary according to the Augmented Dickey Fuller (ADF) and Philips-Perron (PP) test results as indicated in table 4.21. It reveals that series has time dependent statistical properties, which may be stochastic or deterministic. Augmented Dickey Fuller (ADF) and Philips-Perron (PP) test showed that the price series become

stationary when first differencing is done. It indicates that the price series were suitable for co-integration analysis.

Table 4.21: Unit root test using the Augmented Dickey-Fuller and Phillip-Perron tests

Series		Augmented Dickey-Fuller		Phillip-Perron Test	
		t-statistic	p-value	t-statistic	p-value
Agra	Level	-1.55	0.11	-1.24	0.19
	1st difference	-8.55	<0.001	-7.90	<0.001
Bangalore	Level	-0.63	0.44	-0.27	0.58
	1st difference	-5.60	<0.001	-12.23	<0.001
Delhi	Level	-1.72	0.08	-1.69	0.08
	1st difference	-8.00	<0.001	-6.76	<0.001
Mumbai	Level	-0.96	0.29	-0.57	0.46
	1st difference	-8.75	<0.001	-8.53	<0.001

4.3.3 Determination of Co-integration Rank

To check for co-integration among different Potato markets, a test for a suitable lag length to be included in the co-integration analysis was performed. Results of co-integration tests are quite sensitive to lag length included in the model. The number of lags is selected by applying Schwarz's information criterion (SIC). A Vector Autoregression (VAR) on the differenced series was conducted and lags length of the model with the least SIC values chosen as the appropriate lag length for the co-integration test. The interpretation of lag as, for example Agra and Bangalore markets, the pre estimation lag selection criteria indicates the average maximum 2 lag for the model. It indicates the maximum time for price to be transmitted from one Potato market (Agra) to the other (Bangalore) in the long run or to move into long run equilibrium is about two (2) month at most.

In this study, in order to identify a possible co-integration among selected major Potato markets, we have used Johansen multivariate co-integration technique. Johansen (1991) proposed a two-step method to first determine the lag length using either an information criterion or a likelihood ratio test and then to determine the co-integrating rank (r) using a likelihood ratio test, such as the λ max test or the trace test. The detail results of Johansen co-integration rank test without taking into consideration of structural break

are given in the Table 4.22. The rejection of first null hypothesis (co-integrating rank $r = 0$) at 1% level of significance confirms the presence of co-integration in the system. In the second test ($H_0: r \leq 1$) and third test ($H_0: r \leq 2$) also the test statistic is more than the critical value, which means more than two co-integrating rank is present in the system. The final test for $H_0: r \leq 3$ against $H_1: r > 3$ provides sufficient evidence for not rejecting the null hypothesis. Therefore, there are three co-integrating rank in the system, *i.e.* all the markets are co-integrated among themselves.

Table 4.22: Johansen co-integration rank test result without structural break

Null Hypothesis *Co-integrating rank (r)*	**Trace Statistic**	**Critical Value at**	
		5 %	**1 %**
$r = 0$	125.64	48.28	55.43
$r \leq 1$	72.12	31.52	37.22
$r \leq 2$	31.72	17.95	23.52
$r \leq 3$	6.71	8.18	11.65

However, researchers have been concerned with the effects that structural break may have on inference in models with co-integrated variables. Indeed, failure to detect and account the structural break can cause misspecification of the co-integrated system, which adversely affects inference procedure. If we account the structural break then the number of co-integrating rank (*r*) may change. Accordingly, in this study to test co-integration in the presence of structural break, we follow the procedure of Lutkepohl *et al.* (2004). In this approach, first the deterministic trend is estimated and then the data are adjusted accordingly, and finally Johansen test is applied in the adjusted series. The detail results of the Johansen co-integration rank test with structural break are presented in the Table 4.23. The test results indicate that after accounting the structural break, the test statistic changes but overall conclusion on the number of co-integrating rank (*r*) in the system remains same. In other words, we can say that all the major Potato markets under consideration are co-integrated among themselves even after accounting structural break.

Table 4.23: Johansen co-integration rank test result with structural break

Null Hypothesis *Co-integrating rank (r)*	**Trace Statistic**	**Critical Value at**	
		5 %	**1 %**
$r = 0$	143.43	45.20	51.60
$r \leq 1$	80.18	28.45	33.76
$r \leq 2$	43.43	15.83	19.85
$r \leq 3$	05.24	6.79	10.04

For the further study, we paired each market, resulting into 6 bi-variate systems viz. Agra – Bangalore, Agra – Delhi, Agra – Mumbai, Bangalore – Delhi, Bangalore – Mumbai and Delhi – Mumbai. Since the data series are integrated of the same order, co-integration techniques can be used to determine whether a stable long-run relationship exists between each pair. The results of Johansen's co-integration test for each pair markets are presented in Table 4.24 using the trace statistic and maximum eigen value statistic. The trace statistic and maximum eigen value statistic has resulted the same conclusion that all the six market pairs are co-integrated. In other words we can say that all the four selected Potato markets are well integrated and price signals are transferred from one market to the other to ensure efficiency. Thus, Johansen's co-integration test has shown that even though the selected Potato markets in India are geographically isolated and spatially segmented, they are well-connected in terms of prices of Potato, demonstrating that the selected Potato markets have long-run price linkage across them. This means that, selected major Potato market prices in India move closely together in the long run although in the short run they may drift apart. This also indicates that the Potato marketing is an open market of which the forces of demand and supply are the determinant of the various market prices hence ensuring high efficiencies between spatial markets.

Table 4.24: Bi-variate Johansen co-integration rank test results

	λ_{trace} **Statistic**	**Prob.**	λ_{max} **Statistic**	**Prob.**
Agra – Bangalore				
$H_0: r = 0 \; vs \; H_1: r \geq 1$	30.57	<0.001	30.14	<0.001
$H_0: r \leq 1 \; vs \; H_1: r \geq 2$	0.42	0.576	0.42	0.576

Agra – Delhi				
$H_0: r = 0 \; vs \; H_1: r \geq 1$	27.54	<0.001	24.90	<0.001
$H_0: r \leq 1 \; vs \; H_1: r \geq 2$	2.64	0.122	2.64	0.122
Agra – Mumbai				
$H_0: r = 0 \; vs \; H_1: r \geq 1$	22.17	<0.001	21.93	<0.001
$H_0: r \leq 1 \; vs \; H_1: r \geq 2$	0.24	0.682	0.24	0.682
Bangalore – Delhi				
$H_0: r = 0 \; vs \; H_1: r \geq 1$	42.53	<0.001	42.04	<0.001
$H_0: r \leq 1 \; vs \; H_1: r \geq 2$	0.49	0.547	0.49	0.547
Bangalore – Mumbai				
$H_0: r = 0 \; vs \; H_1: r \geq 1$	27.91	<0.001	27.48	<0.001
$H_0: r \leq 1 \; vs \; H_1: r \geq 2$	0.43	0.572	0.43	0.571
Delhi – Mumbai				
$H_0: r = 0 \; vs \; H_1: r \geq 1$	33.57	<0.001	33.33	<0.001
$H_0: r \leq 1 \; vs \; H_1: r \geq 2$	0.24	0.682	0.24	0.682

4.3.4 Test for Causality

Once co-integration between the markets is established it is of interest to analyze for causality of each co-integrating pair. Long run causality from the estimated Johansen VECM was analyzed through a likelihood ratio (LR) test by restricting the disequilibrium error term. The results of long-run causality are presented in Table 4.25, which shows that except Delhi → Agra and Delhi → Mumbai all the long-run causality tests are statistically significant.

Table 4.25: Estimate of error correction term from ECM for different agricultural market

Model	Regressors	Parameter estimated	t- test	P-value
Bangalore → Agra	ECT_{t-1}	-0.27	-3.33	<0.001
Agra → Bangalore	ECT_{t-1}	-0.26	-3.42	<0.001
Delhi → Agra	ECT_{t-1}	0.09	1.01	0.308
Agra → Delhi	ECT_{t-1}	-0.62	-5.15	<0.001
Mumbai → Agra	ECT_{t-1}	-0.29	-3.19	0.001
Agra → Mumbai	ECT_{t-1}	-0.15	-2.02	<0.001

Delhi → Bangalore	ECT_{t-1}	-0.15	-2.80	0.005
Bangalore → Delhi	ECT_{t-1}	-0.32	-4.87	<0.001
Mumbai → Bangalore	ECT_{t-1}	-0.24	-2.51	0.012
Bangalore → Mumbai	ECT_{t-1}	-0.19	-2.03	0.043
Mumbai → Delhi	ECT_{t-1}	-0.38	-5.11	<0.001
Delhi → Mumbai	ECT_{t-1}	-0.04	-0.87	0.382

Note: A→B = A causes B

In this study we found long-run bidirectional causality for the market pairs: Agra ↔ Bangalore, Agra ↔ Mumbai, Bangalore ↔ Delhi and Bangalore ↔ Mumbai, whereas for market pairs Agra → Delhi and Mumbai → Delhi have long-run unidirectional causality. To check short run causality, we applied Wald test. Table 4.26 represent results obtained from Wald test for different market pairs. According to the Wald test, there were short run unidirectional causalities between the market pairs: Bangalore → Agra, Agra → Mumbai, Mumbai → Bangalore and Delhi → Mumbai markets, meaning that a price change in the former market in each pair causes the price formation in the latter market in short run, whereas the price change in the latter market is not feed backed by the price change in the former market in each pair. There were short run bidirectional causalities for the market pairs: Agra ↔ Delhi and Bangalore ↔ Delhi.

Table 4.26: Short run causality by Wald Test

Model	Chi suquare test	P value
Bangalore → Agra	6.21	0.044
Agra → Bangalore	4.76	0.092
Delhi → Agra	59.08	<0.001
Agra → Delhi	10.20	0.006
Mumbai → Agra	1.86	0.394
Agra → Mumbai	22.22	<0.001
Delhi → Bangalore	5.59	0.061
Bangalore → Delhi	13.27	0.001
Mumbai → Bangalore	6.79	0.033
Bangalore → Mumbai	2.72	0.255
Mumbai → Delhi	4.24	0.115
Delhi → Mumbai	30.00	<0.001

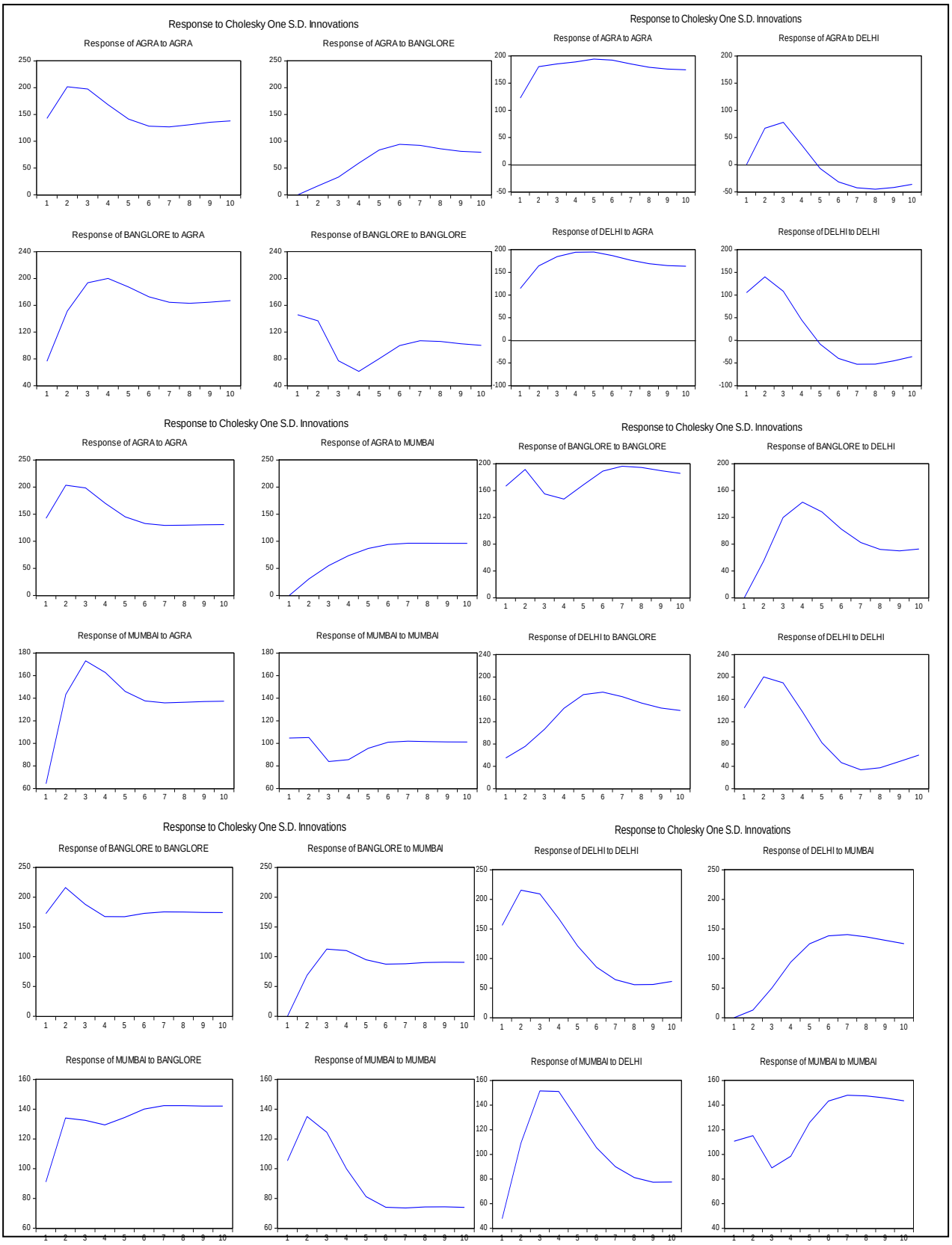


Figure 4.17: Impulse response function for different markets

The best way to interpret the implications of the models for patterns of price transmission, causality and adjustment are to consider the time paths of prices after exogenous shocks, *i.e.* impulse responses. The impulse response function traces the effect of one standard deviation or one unit shock to one of the variables on current and future values of all the endogenous variables in a system over various time horizons. Impulse responses identify the responsiveness of the dependent variable which is (endogenous variables) in the models when a shock is put to the error term. The results of impulse response functions, given in Figure 4.17, show how and to what extent a standard deviation shock in one of the Potato markets affects the current as well as future prices in all the integrated markets over a period of ten months.

CHAPTER V

SUMMARY

Agricultural commodities prices play a vital role in the affordability of food to consumers, as they directly affect their real income, especially among the small and marginal farming community, which spend a large proportion of their income on food. Predicting the prices of agricultural commodities accurately is very important for avoiding market risk, raising farm income and achieving macroeconomic control of the country. However, forecasting of agricultural commodities prices is a complex and difficult task due to its peculiar nature like seasonality, perishability, inelastic demand, etc.

Over the past several decades, lot of efforts have been devoted to develop statistical and machine learning time series models for agriculture price forecasting. However, factors such as major changes in technology like introduction of genetically modified crops, implementation of new economic policy, etc. causes structural break in the parameters of forecasting models. In the agricultural price series data, such structural break is being commonly observed. These structural breaks raise a serious challenge for agricultural commodities price forecasting and it is a key source of forecast failure. In order to improve the forecast accuracy, hybridization of statistical model that account for structural breaks with artificial intelligence model is a good idea, because hybrid model can capture various patterns in the data, concurrently along with breaks.

In view of the above, we studied three different contexts of structural break in agricultural time series data. In the first case, we considered time series volatility model and proposed a hybrid model based on Markov-Switching GARCH (MS-GARCH) and Extreme Learning Machine (ELM), for agricultural price volatility forecasting in the presence of structural break. Some agricultural price series contain long memory property i.e. dependence between apart events diminishes very slowly as the number of lags increases. Therefore, under second objective of the study, we developed a hybrid model based on ARFIMA with dummy variable and ELM in order to get accurate forecast of a long

memory process in the presence of structural break. Lastly, we also investigated co-integrated price series in the presence of structural break for major potato markets in India.

The thesis comprises of five chapters viz. introduction, review of literature, materials and methods, results and discussion, and summary followed by abstract (English and Hindi), appendix (which contains R source code to generate the results) and bibliography.

Chapter I includes brief idea about the characteristics of agricultural price data, challenges in the agriculture price forecasting, and effect of structural break on the volatility models, long memory models and co-integrated time series models. A short introduction of artificial neural network, extreme learning machine and hybrid time series models are also presented in this chapter. The chapter ends with motivation and scope of this study.

Chapter II covers the review of literature related to the structural breaks, time series volatility models, long memory models, co-integrated time series models, artificial intelligence models and hybrid time series models.

Chapter III provides detail descriptions of the data and methodologies used in this study. The data diagnostic tests like test for stationarity, test for linearity, test for heteroscedasticity, test for long memory and detection of structural breaks have been described. This chapter also provide the details about time series volatility models, long memory models, co-integration analysis along with the proposed models.

Chapter IV presents the results and discussion based on research work carried out to accomplish different objectives of the current study. The performance of the proposed hybrid MS-GARCH – ELM model for agricultural price volatility forecasting in the presence of structural break, is evaluated using three different agricultural commodity price series. The results of the study demonstrated the superiority of the proposed hybrid MS-GARCH – ELM model over individual models. Empirical results of long memory process in the presence of structural break show that the forecasting performance of the proposed hybrid model based on ARFIMA with dummy variable and ELM is better than the existing models. The results of overall co-integration test indicated that different potato markets in India are well integrated and have long-run price association across them, which revealed that the potato markets have high efficiency. In other words, we can

say that even though the selected potato markets in India are geographically isolated and spatially segmented, they are well connected in terms of prices of potato, demonstrating that the selected potato markets have long-run price linkage across them.

Following future prospective can be suggested:

- For the agricultural price volatility forecasting, the proposed model can be further improved by using deep learning algorithm like LSTM, etc.
- ARFIMA-FIGARCH with structural break can be studied for the long memory process
- Fractional co-integration with structural breaks can also be considered

ABSTRACT

Accurate price forecasting of agricultural commodities is very important for raising income of the farmers as well as for avoiding market risk. However, due to biological nature of production of agricultural commodities, forecasting of their prices become a challenging task. These challenges become more severe when structural breaks are present in the observed agricultural price series due to factors like major changes in technology, sudden changes in economic policy, *etc.* In this study, an effort has been made to account for the structural break along with the other complex patterns like non-stationarity, non-linearity, long memory and cointegration present in the agricultural price series.. Generally, single model may not be able to capture all complex patterns present in the data series concurrently. Therefore, to capture various complex patterns in the data along with structural break, hybridization of statistical model that account for structural break with artificial intelligence model has been done. Accordingly, for agricultural price volatility forecasting in the presence of structural break, a hybrid model based on Markov-Switching GARCH (MS-GARCH) and Extreme Learning Machine (ELM) is proposed. The performance of the proposed hybrid MS-GARCH–ELM model is evaluated on the weekly potato price of Delhi market, monthly international Groundnut oil and Palm oil price series, and it is found that the proposed model outperformed its counterparts. Empirical results of agricultural price series that contain long memory property with structural break show that the forecasting performance of the proposed hybrid model based on ARFIMA with dummy variable combined with ELM is better than the individual model. Further, the effect of structural break in the co-integrated system has also been evaluated. Accordingly, spatial market integration among major Potato markets in India are investigated in the absence and presence of structural break. The overall co-integration test results indicated that selected potato markets in India are well integrated and have long-run price association across them.

सार

किसानों की आय बढ़ाने और बाजार के जोखिम से बचने के लिए कृषि वस्तुओं के मूल्य का सही पूर्वानुमान बहुत महत्वपूर्ण है। हालांकि, कृषि वस्तुओं के उत्पादन की जैविक प्रकृति के कारण, इनकी कीमतों का पूर्वानुमान एक चुनौतीपूर्ण कार्य बन जाता है। ये चुनौतियां तब और गंभीर हो जाती हैं, जब प्रेक्षित आकड़ों में स्ट्रक्चरल ब्रेक मौजूद होते हैं। टेक्नोलॉजी में बड़े बदलाव, आर्थिक नीति में अचानक बदलाव आदि, टाइम सीरीज़ डेटा में स्ट्रक्चरल ब्रेक का कारण बनता है। इस तरह के स्ट्रक्चरल ब्रेक कृषि मूल्य श्रृंखला में अक्सर पाए जाते हैं और यह पूर्वानुमान विफलता का एक स्रोत हो सकता है। इस अध्ययन में, कृषि मूल्य श्रृंखला में मौजूद नॉन-स्टेसनरिटी और नॉन-लिनियरिटी जैसे अन्य जटिल पैटर्न के साथ स्ट्रक्चरल ब्रेक को ध्यान में रखते हुए अस्थिर मूल्य श्रृंखला, लंबी स्मृति गुण के साथ मूल्य श्रृंखला और सह-एकीकृत मूल्य श्रृंखला को मॉडल करने का प्रयास किया गया है। एक व्यक्तिगत मॉडल समवर्ती रूप से डेटा श्रृंखला में मौजूद सभी जटिल पैटर्न को कैप्चर करने में सक्षम नहीं होता है। इसलिए, स्ट्रक्चरल ब्रेक के साथ डेटा में विभिन्न जटिल पैटर्न को पकड़ने के लिए, सांख्यिकीय मॉडल और कृत्रिम बुद्धिमत्ता मॉडल का संकरण किया गया है। तदनुसार, कृषि मूल्य में अस्थिरता का पूर्वानुमान करने के लिए मार्कोव-स्विचिंग गार्च (एमएस-गार्च) मॉडल तथा ईक्स्ट्रिम लर्निंग मशीन (ईएलएम) पर आधारित एक संकर मॉडल का प्रस्ताव किया गया है। प्रस्तावित हाइब्रिड MS-GARCH-ELM मॉडल के प्रदर्शन का मूल्यांकन दिल्ली बाजार के आलू की कीमत, अंतरराष्ट्रीय मूंगफली तेल और पाम तेल की मूल्य श्रृंखला पर किया गया है, और यह पाया गया है कि प्रस्तावित मॉडल का पूर्वानुमान सभी व्यक्तिगत मॉडल से सटीक है। कृषि मूल्य श्रृंखला जिसमे लॉन्ग मिमोरी के गुण होते हैं, डमी चर के साथ ARFIMA और ELM पर आधारित प्रस्तावित हाइब्रिड मॉडल का प्रदर्शन अन्य मॉडल के तुलना मे बेहतर पाया गया है। इस अध्ययन में, सह-एकीकृत प्रणाली में स्ट्रक्चरल ब्रेक के प्रभाव का मूल्यांकन करने का भी प्रयास किया गया है। तदनुसार, भारत के प्रमुख आलू बाजारों के बीच स्ट्रक्चरल ब्रेक की अनुपस्थिति और उपस्थिति में स्थानिक बाजार एकीकरण की जांच की गयी है।

APPENDIX

R code to generate the results

```
# ipak function: install and load multiple R packages.

# check to see if packages are installed. Install them if they are not, then load them into the R session.

#if(!require(changepoint)){

#install.packages('changepoint')

#}

ipak <- function(pkg){

  new.pkg <- pkg[!(pkg %in% installed.packages()[ ,"Package"])]

  if (length(new.pkg))

    install.packages(new.pkg, dependencies = TRUE)

  sapply(pkg, require, character.only = TRUE)

}

# usage

packages <- c("tseries", "forecast", "rmgarch", "rugarch", "quantmod", "MSGARCH", "neuralnet", "changepoint", "aTSA", "nnfor",

          "moments","portes")

ipak(packages)

# Load multiple packages

Packages <- c("tseries", "forecast", "rmgarch","rugarch","quantmod", "MSGARCH", "neuralnet", "changepoint", "aTSA", "nnfor",

          "moments", "portes")

lapply(Packages, library, character.only = TRUE)

#########################################################################
#

# Import Data from copied file of Excel

data=read.table(file = "clipboard", sep = "\t", header = TRUE)

#data<-read.csv(choose.files())
```

```
data_ts<-ts(data)
# To make data into time series
data_ts<-ts(data, frequency = 12, start = c(1980, 1))
head(data_ts)
tail(data_ts)
plot(data_ts, xlab="Time (Month)", ylab="Price ($/mt)", col="black")
#chartSeries(data_ts,  theme = chartTheme("white"))
# For return series
data_log<-log(data_ts) # Natural Log of the data series
data_rt<- diff(data_log) # Difference of the log series
data_rt<- 100*data_rt
plot(data_rt)
data_rt<- periodReturn(data_ts)
data_df<- diff(data_ts)
plot(data_df)
# Seasonal Plot
#ggseasonplot(data_ts, year.labels=F, year.labels.left=F, ylab("Price (Rs/Qtl)"))
# Summary statistics of the data
length(data_ts) # No. of observations
summary(data_ts) # Summary Statistics
cv(data_ts) # Coefficient of variation
sd(data_ts) # Standard deviation
skewness(data_ts) # Skewness
kurtosis(data_ts) # Kurtosis
jarque.bera.test(data_ts) # Test for normality
bds.test(data_ts) # Test for linearity
# ADF test for stationarity
adf.test(data_ts)
```

```
adf.test(data_df)
adf.test(data_log)
adf.test(data_rt)
# PP test for Stationarity
pp.test(data_ts)
pp.test(data_df)
pp.test(data_log)
pp.test(data_rt)
# For Portmonteau test install package "portes"
BoxPierce(resid,lags = 6)
BoxPierce(resid,lags = 12)
BoxPierce(resid,lags = 18)
BoxPierce(resid,lags = 24)
# Auto correlation function
Acf(data_rt, lag.max = 50)
Acf(data_rt, lag.max=100)
# Partial auto correlation functions
Pacf(data_rt, lag.max = 50)
Pacf(data_rt,lag.max=40)
#########################################################
# Test of structural Break in Variance (PELT algorithm)
data_pelt<- as.numeric(data_rt)
soya.pelt <- cpt.var(data_pelt, method = "PELT", minseglen = 10)
plot(soya.pelt, xlab = "Time", ylab="Log Return")
logLik(soya.pelt)
print(soya.pelt)
#####################################################################
# To make ARMA structure
```

```
# Finding the order of ARIMA model
fit<- auto.arima(data_rt)
fit
accuracy(fit)
# Fitting of ARMA model
fit_arma<-arima(data_rt,order=c(2,0,1), seasonal = list(order=c(0,0,0),period=NA))
fit_arma<- Arima(data_rt, order = c(1, 0, 1),include.mean = FALSE,
                 seasonal = c(0, 0, 0))
#checkresiduals(fit_arma)
fit_arma
accuracy(fit_arma)
# To obtained residuals from ARMA Model
resid<-fit_arma$residuals # It is taken as input variable for GARCH and MSGARCH
plot(resid, ylab="Residuals")
resid_sqr<- (resid)^2 # It is taken as input variable for ANN and ELM
plot(resid_sqr, ylab="Squared Residuals")
# Obtaining the ACF plot of Squared Residuals
Acf(resid,lag.max=150)
residacf<-Acf(resid_sqr,lag.max=150)
# Obtaining the PACF plot of squared residual
residpacf<-Pacf(resid_sqr,lag.max=50)
# ARCH LM Test
arch.test(fit_arma)
# The ARCH Engle's test is constructed based on the fact that if the residuals
#(defined as e[t]) are heteroscedastic, the squared residuals (e^2[t]) are autocorrelated.
##########################################################################
##
#GARCH model using ruGarch package
# Specification of GARCH model
```

```
# Model Forecasting Performance of GARCH model
# Training set
data_trn<- ts(data_rt[1:467], frequency = 12, start = c(1980, 2))
data_trn_resid<- ts(resid[1:467], frequency = 12, start = c(1980, 2))
#data_trn_resid<- ts(resid[1:712])
data_trn_residsq<- ts(resid_sqr[1:467], frequency = 12, start = c(1980, 2))
#data_trn_residsq<- ts(resid_sqr[1:712])
# Test set
data_test<- ts(resid_sqr[468:479], frequency = 12, start = c(2019, 1))
tail(data_test)
#garch_spec<- ugarchspec()
#data_trn<- ts(data_rt[1:712])
#data_test<- ts(resid_sqr[712:724])
#garch_spec
garch_spec<-ugarchspec(variance.model = list(model = "sGARCH",
                                             garchOrder = c(1, 1),
                                  submodel = NULL, external.regressors = NULL,
                                  variance.targeting = FALSE),
                mean.model = list(armaOrder = c(1, 1), include.mean = FALSE,
                                  archm = FALSE,
                                  archpow = 1, arfima = FALSE,
                                  external.regressors = NULL, archex = FALSE),
                distribution.model = "norm", start.pars = list(), fixed.pars = list())
# Model Estimation
garch_fit<- ugarchfit(spec=garch_spec, data= data_trn)
garch_fit
#garch_spec
paste("Element in the @fit slot")
```

```
names(garch_fit@fit)
paste("Element in the @model slot")
names(garch_fit@model)
# Estimated conditional variance
est_var<- ts(garch_fit@fit$var)
est_var<-ts(est_var, frequency = 12, start = c(1980, 2))
plot(est_var, ylab="Variance")
lines(resid_sqr)
#plot(est_res2, ylab="Squared Res/ Variance", col=1)
#lines(est_var, col=2)
#est_var
# Estimated residuals
#est_res<- garch_fit@fit$residuals
#est_res<-ts(est_res, frequency = 12, start = c(1960, 2))
#plot(est_res, ylab="Residuals")
# Estimated squared residuals
#est_res2<- (garch_fit@fit$residuals)^2
#est_res2<-ts(est_res2, frequency = 12, start = c(1980, 2))

# Forecasting
fcast<- ugarchforecast(garch_fit, n.ahead=12)
fcast
# garch_fcast contains two slots @model and @forecast
paste("Element in the @forecast slot")
names(fcast@forecast)

fcast_sigma<- fcast@forecast$sigmaFor
fcast_sigma.sq<- (fcast_sigma)^2
```

```
plot(fcast_sigma, xlab="Time", ylab="Sigma", col="red")
fcast_garch<- ts(fcast_sigma.sq, frequency = 12, start = c(2019, 1))
plot(fcast_garch, xlab="Time", ylab="Sigma", col="red")
plot(resid_sqr, ylab="Squared Res/ Variance", col=1)
lines(est_var, col=2)
lines(fcast_garch, col=3)
#fcast_garch<- ts(fcast_sigma.sq)
# RMSE and MAPE
rmse_garch=sqrt(mean((data_test-fcast_garch)^2))
rmse_garch
mape_garch=(mean((abs(data_test-fcast_garch))/data_test))*100
mape_garch
plot(data_test)
lines(fcast_garch, col="green")
# Load the data
# Model specification
ms2.garch.n <- CreateSpec(variance.spec = list(model = "sGARCH"),
 distribution.spec = list(distribution = "norm"),
 switch.spec = list(K = 2))
# Maximum Likelihood Estimation of MS-GARCH model
fit.ml <- FitML(spec = ms2.garch.n, data = dem2gbp)
summary(fit.ml)
# MCMC Estimation of MS-GARCH model
set.seed(1234)
fit.mcmc <- FitMCMC(spec = ms2.garch.n, data = dem2gbp)
library(coda)
summary(fit.mcmc)
# Forecasting
```

```
pred <- predict(fit.ml, nahead = 5, do.return.draw = TRUE)

pred$vol

pred$draw[, 1:4]

# Value at Risk and Expected Shortfall

risk <- Risk(fit.ml, alpha = c(0.01, 0.05), nahead = 5)

risk$VaR

risk$ES

#To extract the in-sample conditional volatility

Volatility()

#Latent states

states()

#To simulate

simulate(fit.ml, nsim = 2, nahead = 4, nburn = 500)

#To compute the predictive density

PredPdf()

# the probability integral transform

PIT()

# Estimation of Long memory model with structural Break

# check to see if packages are installed. Install them if they are not, then load them into
the R session.

ipak <- function(pkg){

  new.pkg <- pkg[!(pkg %in% installed.packages()[ ,"Package"])]

  if (length(new.pkg))

    install.packages(new.pkg, dependencies = TRUE)

  sapply(pkg, require, character.only = TRUE)

}

# usage

packages <- c("tseries", "forecast", "LongMemoryTS", "rugarch", "changepoint",
"aTSA",
```

```
        "moments","portes")
ipak(packages)
# Load multiple packages
Packages <- c("tseries", "forecast", "LongMemoryTS", "rugarch", "changepoint",
"aTSA",
        "moments","portes")
lapply(Packages, library, character.only = TRUE)
#########################################################################
#
# Import Data from copied file of Excel
data=read.table(file = "clipboard", sep = "\t", header = TRUE)
#data<-read.csv(choose.files())
x=read.table(file = "clipboard", sep = "\t", header = TRUE)
x=ts(x)
# To make data into time series
data_ts<- ts(data)
head(data_ts)
tail(data_ts)
plot(data_ts, xlab="Time", ylab="Price (Rs/q)", col="green")
#chartSeries(data_ts,  theme = chartTheme("white"))
data_df<- diff(data_ts)
data_df_sq<- (data_df)^2
# For return series
data_log<-log(data_ts) # Natural Log of the data series
data_rt<- diff(data_log) # Difference of the log series
plot(data_rt)
data_rt_sq<- (data_rt)^2
# Seasonal Plot
#ggseasonplot(data_ts, year.labels=F, year.labels.left=F, ylab("Price (Rs/Qtl)"))
```

```
# Summary statistics of the data
length(data_ts) # No. of observations
summary(data_ts) # Summary Statistics
sd(data_ts) # Standard deviation
skewness(data_ts) # Skewness
kurtosis(data_ts) # Kurtosis
jarque.bera.test(data_ts) # Test for normality
bds.test(data_ts) # Test for linearity

# ADF test for stationarity
adf.test(data_ts)
adf.test(data_log)
# PP test for Stationarity
pp.test(data_ts)
pp.test(data_log)
# For Portmonteau test install package "portes"
BoxPierce(data_log,lags = 6)
BoxPierce(data_log,lags = 12)
BoxPierce(data_log,lags = 18)
BoxPierce(data_log,lags = 24)
#########################################################
# Auto correlation function
Acf(data_rt, lag.max = 40)
Acf(data_rt_sq, lag.max=40)
# Partial auto correlation functions
Pacf(data_rt, lag.max = 40)
Pacf(data_rt_sq,lag.max=40)
#################################################
```

```
# Long memory test

gph(data_rt, 60)

gph(data_rt_sq, 60)

ELW(data_rt, 75)

ELW(data_rt_sq, 75)

############################################################

# Two packages for test of structural break "changepoint" and "strucchange"

# Test of structural Break in Variance (PELT algorithm)

data_pelt<- as.numeric(data_ts)

cpt_mean<- cpt.mean(data_pelt, penalty = "None", method = "BinSeg", Q=3, test.stat =
"Normal", class = "TRUE", param.estimates = "TRUE")

plot(cpt_mean, xlab = "Time", ylab="Price ($/mt)")

logLik(cpt_mean)

print(soya.pelt)

cpt_variance<- cpt.var(data_pelt, penalty = "None", method = "PELT", Q=3, test.stat =
"Normal", class = "TRUE", param.estimates = "TRUE")

plot(cpt_variance, xlab = "Time", ylab="Price ($/mt)")


v1.man=cpt.mean(data_pelt,method='PELT',penalty='Manual',pen.value='2*log(1000)',
minseglen=200)

plot(v1.man, xlab = "Time", ylab="Price ($/mt)")


m1.cusum=cpt.var(data_pelt, method = "PELT")

plot(m1.cusum, xlab = "Time", ylab="Price ($/mt)")

soya.pelt <- cpt.mean(data_pelt, method = "AMOC")

plot(soya.pelt, xlab = "Time", ylab="Price ($/mt)")

logLik(soya.pelt)

print(soya.pelt)
```

```
############################################################
# Test for Long memory in the data   #### "LongMemoryTS"
library(fracdiff)
T<-1000
m<-floor(1+T^0.5)
m
d=0.4
series<-fracdiff.sim(n=T, d=d)$series
gph(data_rt_sq, 50)
ELW(data_rt_sq, 50)
elw
Qu.test(data_rt, 50)
############################################################

# Fitting of ARIMA model
aic_mat<-matrix(nrow=3,ncol=3)
rownames(aic_mat)<-paste("p",0:2,sep="=")
colnames(aic_mat)<-paste("q",0:2,sep="=")
for(i in 0:2){
  for(j in 0:2){
    fit_arima<-arima(datats,order=c(i,1,j), seasonal =
list(order=c(0,0,0),period=NA),include.mean=TRUE)
    aic_mat[(i+1),(j+1)]=fit_arima$aic
  }
}
fit_arima

#####################################################################
# Training set
```

```
data_trn<- ts(data_rt[1:1138])


# Test set
data_test<- ts(data_rt[1139:1150])
tail(data_test)
#garch_spec<- ugarchspec()
#garch_spec
##############################################################################
##
arfima_spec<- arfimaspec(mean.model = list(armaOrder = c(1, 1),
                                           arfima = TRUE, external.regressors = x))
fit_arfima<- arfimafit(arfima_spec, data_rt)
fit_arfima
fited_arfima<- ts(fit_arfima@fit$fitted.values)
#plot(data_rt, col=1)
#lines(fited, col=2)
write.csv(fited_arfima, file = "afima.csv")
# To obtained residuals from ARMA Model
resid<-ts(fit_arfima@fit$residuals) # It is taken as input variable for GARCH and
MSGARCH
plot(resid, ylab="Residuals")
resid_sqr<- (resid)^2 # It is taken as input variable for ANN and ELM
plot(resid_sqr, ylab="Squared Residuals")


# Obtaining the ACF plot of Squared Residuals
Acf(resid,lag.max=100)
Acf(resid_sqr,lag.max=50)
# Obtaining the PACF plot of squared residual
Pacf(resid,lag.max=100)
```

```
Pacf(resid_sqr,lag.max=50)
##############################################################################
##
figarch_spec<-ugarchspec(variance.model = list(model = "sGARCH",
                                               garchOrder = c(1, 1),
                                               submodel = NULL, external.regressors = NULL,
                                               variance.targeting = FALSE),
                         mean.model = list(armaOrder = c(1, 1), include.mean = FALSE,
                                           archm = FALSE,
                                           archpow = 1, arfima = TRUE,
                                           external.regressors = NULL, archex = FALSE),
                         distribution.model = "norm", start.pars = list(), fixed.pars = list())

# Model Estimation
figarch_fit<- ugarchfit(spec=figarch_spec, data= data_trn)
figarch_fit
fited_afigarch<- figarch_fit@fit$fitted.values
plot(data_rt, col=1)
#GARCH model using ruGarch package
# Specification of GARCH model
# Model Forecasting Performance of GARCH model
# Training set
data_trn<- ts(data_rt[1:707], frequency = 12, start = c(1960, 2))

# Test set
data_test<- ts(resid_sqr[708:719], frequency = 12, start = c(2019, 1))
tail(data_test)
#garch_spec<- ugarchspec()
#garch_spec
```

```
#########################################################################
#########
# Fitting of ARIMA model
aic_mat<-matrix(nrow=3,ncol=3)
rownames(aic_mat)<-paste("p",0:2,sep="=")
colnames(aic_mat)<-paste("q",0:2,sep="=")
for(i in 0:2){
  for(j in 0:2){
    fit_arima<-arima(data_rt,order=c(i,1,j), seasonal = list(order=c(0,0,0),period=NA))
    aic_mat[(i+1),(j+1)]=fit_arima$aic
  }
}
fit_arima
aic_mat
#############################################################
# To put these forecast in the context, lets disply with the last 20 obs. of estimation
est_var_t<- ts(c(tail(est_var, 12))) # To get last 12 obs
est_res2_t<- c(tail(est_res2, 20, rep(NA, 10))) # To get last 20 obs
garch_fcast_t<- c(rep(NA, 20),(fcast_sigma)^2 )

plot(est_res2_t, type = "1")
line(est_var_t, col="green")
line(garch_fcast_t, col="orange")
# Fitting of ANN model
data_trn.ann<- data_trn.elm

fit.ann<- mlp(data_trn.ann, m = frequency(data_trn.ann), hd = NULL, reps = 20, comb =
c(
   "mean"), lags = NULL, keep = NULL, difforder = NULL,
```

```
    outplot = c(TRUE), sel.lag = c(TRUE),
    allow.det.season = c(FALSE),
    hd.auto.type = c("cv"), hd.max = NULL)

names(fit.ann)
fcast.ann<- predict(fit.ann, h=12)
fcast_ann<- fcast.ann$mean
fcast_ann

# RMSE and MAPE
rmse_ann=sqrt(mean((data_test-fcast_ann)^2))
rmse_ann
mape_ann=(mean((abs(data_test-fcast_ann))/data_test))*100
mape_ann
plot(data_test)
lines(fcast_ann, col="green")
```

BIBLIOGRAPHY

Ai, C.R., Chatrath, A. and Song, F. (2006). On the co-movement of commodity prices. *American Journal of Agricultural Economics*, **88**, 574-588.

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, **135**, 31-53.

Allen, P.G. (1994). Economic forecasting in agriculture. *International Journal of Forecasting*, **10**, 81-135.

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**, 821-856.

Ardia, D. (2008). Financial risk management with bayesian estimation of GARCH models: Theory and applications. *Lecture Notes in Economics and Mathematical Systems*, Volume 612, Springer-Verlag, Berlin Heidelberg. https://www.springer.com/gp/book/9783540786566.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, **18**(1),1-22.

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **74**, 3-30.

Bates, J.M. and Granger, C.W.J. (1969). The combination of forecasts. *Operations Research Quarterly*, **20**, 451-468.

Beran, J. (1995). *Statistics for Long-Memory Processes*. Chapman and Hall Publishing Inc., New York.

Bildirici, M. and Ersin, O. (2012). Markov switching artificial neural networks and volatility modelling with an application to a Turkish stock index. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2118917.

Bildirici, M. and Ersin, O. (2015). Forecasting volatility in oil prices with a class of nonlinear volatility models: Smooth transition RBF and MLP neural networks augmented GARCH approach. *Petroleum Science*, **12**, 534-552.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307-327.

Booth, G.G., Kaen, F.R. and Koveos, P.E. (1982). R/S analyses of foreign exchange rates under two international monetary regime. *Journal of Monetary Economics*, **10**, 407-415.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Brock, W.A., Dechert, W.D. and Scheinkman, J.A. (1987). A test for independence based on the correlation dimension. Working paper, *Department of Economics. University of Wisconsin at Madison, University of Houston, and University of Chicago*.

Brummer, B., Korn, O., Chlubler, K. and Jaghdani, T. (2016). Volatility in oilseeds and vegetable oils markets: Drivers and spillovers. *Journal of Agricultural Economics*, **67**, 389-405.

Cai, J. (1994). A Markov Model of Switching-Regime ARCH. *Journal of Business and Economic Statistics*, **12**(3), 309-316.

Campbell, J.Y. and Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomic Annual 1991*, National Bureau of Economic Research, Cambridge, MA, 141-218.

Campbell, J.Y. and Shiller, R.J. (1987). Cointegration and tests of present value models. *Journal of Political Economy*, **95**, 1062-1088.

Chaabane, N. (2014). A hybrid ARFIMA and neural network model for electricity price prediction. *International Journal of Electrical Power and Energy Systems*, **55**, 187-194.

Chai, J., Xing, L.M., Zhou, X.Y., Zhang, Z.G. and Li, J.X. (2018). Forecasting the WTI crude oil price by a hybrid-refined method. *Energy Economics*, **71**, 114-127.

Che, J. and Wang, J. (2010). Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Conversion and Management*. **51**(10), 1911-1917.

Chen, X., Ghysels, E. and Wang, F. (2015). HYBRID-GARCH: A generic class of models for volatility predictions using high frequency data. *Statistica Sinica*, **25**, 759-786.

Chen, Y., He, K. and Tso, G.K.F. (2017). Forecasting crude oil prices: A deep learning based model. *Procedia Computer Science*, **122**, 300-307.

Cheung, Y.W. (1993). Tests for fractional integration: A Monte Carlo investigation. *Journal of Time Series Analysis*, **14**(4), 331-345.

Choi, K. and Zivot, E. (2007). Long memory and structural changes in the forward discount: An empirical investigation. *Journal of International Money and Finance*, **26**(3), 342-363.

Choudhury, K., Jha, G.K., Das, P. and Chaturvedi, K.K. (2019). Forecasting Potato price using ensemble artificial neural networks. *Indian Journal of Extension Education*, **55**(1), 73-77.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591-605.

Clark, T.E. and McCracken, M.W. (2009). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, **50**, 363-395.

Clements, M.P. and Hendry, D.F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127-146.

Clements, M.P. and Hendry, D.F. (1998). *Forecasting Economic Time Series*. Cambridge, Cambridge University Press.

David, S.A., Inácio, C.M.C. and Machado, A.J.T. (2019). Quantifying the predictability and efficiency of the cointegrated ethanol and agricultural commodities price series. *Applied Sciences*, **9**(24), 1-14.

Dawson, P.J. and White, B. (2002). Interdependencies between agricultural commodity futures prices on the LIFFE. *The Journals of Futures Markets*, **22**, 269-280.

Diebold, F. (1986). Modelling the persistence of conditional variances: A comment. *Econometric Reviews*, **5**(1), 51-56.

Diebold, F.X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, **105**, 131-159.

Donaldson, R.G. and Kamstra, M. (1996). Forecast combining with neural networks. *Journal of Forecasting*, **15**, 49-61.

Donaldson, R.G. and Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance*, **4**, 17-46.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, **50**, 987-1008.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, **55**, 251-276.

Engle, R.F. and Smith, A.D. (1999). Stochastic permanent breaks. *Review of Economics and Statistics*, **81**, 553-574.

Engle, R.F. and Bollerslev, T. (1986) Modelling the persistence of conditional variance. *Econometric Reviews*, **5**, 1-50.

FAO, UNCTAD, (2011). Price volatility in food and agricultural markets: Policy responses. *Food and Agricultural Organization*, https://doi.org/10.1596/27379

Fong, W.M. and See, K.H. (2002). A Markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, **24**, 71-95.

Franses, P.H. and Dijk, D.V. (1996). Forecasting stock market volatility using non-linear GARCH models. *Journal of Forecasting*, **15**(3), 229-235.

Gaetano, D.D. (2018). Forecasting with GARCH models under structural breaks: An approach based on combinations across estimation windows.

Communications in Statistics-Simulation and Computation, https://www.tandfonline.com/doi/abs/10.1080/03610918.2018.1520876.

Garcia, D. and Kristjanpoller, W. (2019). An adaptive forecasting approach for copper price volatility through hybrid and non-hybrid models. *Applied Soft Computing Journal*, **74**, 466-478.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221-238.

Ghosh, H., Paul, R.K. and Prajneshu. (2011). Nonlinear time series modelling and forecasting for periodic and ARCH effects. *Journal of Statistics Theory and Practice*, **4**(1), 27-44.

Gil-Alana, A.L. (2002). A joint test of fractional integration and structural break at a known period of time. *Journal of Time Series Analysis*, **25**, 691-700.

Glosten, L.R., Jagannathan, R. and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, **48**(5), 1779-1801.

Gopal, S. and Ramasamy, M. (2017). Hybrid multiple structural break model for stock price trend prediction. *The Spanish Review of Financial Economics*, **15**(2), 41-51.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121-130.

Granger, C.W.J. and Andersen, A. (1978). On the invertibility of time series models. *Stochastic Processes and their Applications*, **8**, 87-92.

Granger, C.W.J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, **11**(3), 399-421.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15-29.

Granger, C.W.J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2**(2), 111-120.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, **42**(1), 27-62.

Haas, M., Mittnik, S. and Paolella, M.S. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, **2**(4), 493-530.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357-384.

Hamilton, J.D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, **64**, 307-333.

Hansen, P.R. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1,1)? *Journal of Applied Econometrics*, **20**, 873-889.

Helms, B.P., Kaen, F.R. and Rosenman, R.E. (1984). Memory in commodity futures contracts. *The Journal of Futures Markets*, **10**, 559-567.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, **129**, 121-138.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, **168**, 165-176.

Huang, G.B., Zhou, H., Ding, X. and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **42**, 513-529.

Huang, G.B., Zhu Q.Y., and Siew, C.K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, **70**, 489-501.

Hurst, H.E. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Agricultural Engineers*, **116**, 770-799.

Inclan, C. and Tiao, G. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, **89**(427), 913-923.

Jackson, B., Sargle, J.D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L. and Tsai, T.T. (2005). An algorithm for

optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, **12**(2), 105-108.

Jebli, M.B. and Youssef, S.B. (2017). Renewable energy consumption and agriculture: Evidence for co-integration and Granger causality for Tunisia economy. *International Journal of Sustainable Development and World Ecology*, **24**(2), 149-158.

Jensen, M.J. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting*, **18**, 17-32.

Jha, G.K. (2013). Energy growth linkage and strategy for meeting the energy demand in Indian agriculture. *Agricultural Economics Research Review*, **26**, 119-127.

Jha, G.K. and Sinha, K. (2014). Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India. *Neural Computing and Applications*, **24**, 563-571.

Jha, G.K., Pal, S. and Singh, A. (2012). Changing energy-use pattern and demand projection for Indian agriculture, *Agricultural Economics Research Review*, **25**, 61-68.

Johansen, S. (1988). Statistical analysis of co-integration vectors. *Journal of Economic Dynamics and Control*, **12**, 231-254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**(6), 1551-1580.

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration with application to the demand for money. *Oxford Bulletin of Economics and Statistics*, **52**, 169-210.

Johansen, S. and Nielsen, M.O. (2018). Nonstationary co-integration in the fractionally co-integrated VAR model. *Journal of Time Series Analysis*, **40**(4), 1-25.

Khashei, M. and Hajirahimi, Z. (2018). A comparative study of series ARIMA/MLP hybrid models for stock price forecasting. *Communication in Statistics-Simulation and Computation*, **48**(9), 1-16.

Killick, R., Fearnhead, P. and Eckley, I.A. (2012). Optimal detection of change points with a linear computational cost. *Journal of the American Statistical Association*, **107**, 1590-1598.

Kristjanpoller, W. and Hernandez, E. (2017). Volatility of main metals forecasted by a hybrid ANN-GARCH model with regressors. *Expert Systems with Applications*, **84**, 290-300.

Kumar, R.R. and Jha, G.K. (2017). Examining the co-movement between energy and agricultural commodity prices in India. *Journal of the Indian Society of Agricultural Statistics*, **70**(3), 241-252.

Kunsch, H.R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika*, **74**, 517-524.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**, 159-178.

Lama, A., Jha, G.K., Gurung, B., Paul, R.K., Bharadwaj, A. and Parsad, R. (2016). A comparative study on time-delay neural network and GARCH models for forecasting agricultural commodity price volatility. *Journal of the Indian Society of Agricultural Statistics*, **70**(1), 7-18.

Lamoureux, C.G. and Lastrapes, W.D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economic Statistics*, **8**(2), 225-234.

Leccadito, A., Rachedi, O. and Urga, G. (2015). True versus spurious long memory: Some theoretical results and a Monte Carlo comparison. *Econometric Reviews*, **34**(4), 452–479.

Lin, L., Jiang, Y., Xiao, H. and Zhou, Z. (2020). Crude oil price forecasting based on a novel hybrid long memory GARCH-M and wavelet analysis model, *Physica A: Statistical Mechanics and its Applications*, https://www.sciencedirect.com/science/article/pii/S0378437119319697

Lippmann, R. (1987). An introduction to computing with neural nets, *IEEE ASSP Magazine*, **4**, 4-22.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, **65**(2), 297-303.

Lutkepohl, H., Saikkonen, P. and Trenkler, C. (2004). Testing for the cointegrating rank of a VAR with level shift at unknown time. *Econometrica*, **72**(2), 647-662.

Maciel, L., Gomide F. and Ballini, R. (2016). Evolving fuzzy-GARCH approach for financial volatility modeling and forecasting. *Computational Economics*, **48**, 379-398.

Malliaris, A.G. and Urrutia, J.L. (1996). Linkages between agricultural commodity futures contracts. *The Journals of futures markets*, **16**, 595-609.

Man, K.S. (2003). Long memory time series and short term forecasts. *International Journal of Forecasting*, **19**(3), 477-491.

McAleer, M., Hoti, S. and Chan, F. (2009). Structure and asymptotic theory for multivariate asymmetric conditional volatility. *Econometric Reviews*, **28**, 422-440.

McLeod, A.I. and Hipel K.W. (1978). Simulation procedures for Box-Jenkins Models. *Water Resources Research*, **14**(5), 969-975.

Mitra, D. and Paul, R.K. (2017). Hybrid time-series models for forecasting agricultural commodity prices, *Model Assisted Statistics and Applications*, **12**(3), 255-264

Natanelov, V., Alam, M., Mckenzie, A.M. and Huylenbroeck, G.V. (2011). Is there co-movement of agricultural commodities prices and crude oil? *Energy Economics*, **39**, 4971-4984.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, **59**(2), 347-370.

Paul, R.K. (2014). Forecasting wholesale price of pigeon pea using long memory time-series models. *Agricultural Economics Research Review*, **27**(2), 167-176.

Paul, R.K., Gurung, B. and Samanta, S. (2015a). Monte Carlo simulation for comparison of different estimators of long memory parameter: An application of ARFIMA model for forecasting commodity price. *Model Assisted Statistics and Applications*, **10**(2), 117-128.

Paul, R.K., Gurung, B. and Samanta, S. (2015b). Analyzing the effect of dual long memory process in forecasting agricultural prices in different markets of India. *International Journal of Empirical Finance*, **4**, 235-249.

Paul, R.K., Prajneshu and Ghosh, H. (2014). Development of out-of-sample forecast formulae for ARIMAX-GARCH model and their application. *Journal of Indian Society of Agricultural Statistics*, **68**(1), 85-92.

Perales, G.B. (2009). Price volatility forecasts for agricultural commodities: An application of volatility models, option implieds and composite approaches for futures prices of corn and wheat. *Journal of Management, Finance and Economics*, **3**(2), 40-59.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, **57**(6), 1361-1401.

Perron, P. and Qu, Z. (2010). Long memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics*, **28**, 275-290.

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R.* Springer Science and Business Media LLC.

Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, **75**, 335-346.

Qu, Z. (2011). A Test against Spurious Long Memory. *Journal of Business and Economic Statistics*, **29**(3), 423-438.

Ramsey, J.B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, Series B*, **31**(2), 350-371.

Robinson, P.M. (1995). Long-periodogram regression of time-series with long-range dependence. *The Annals of Statistics*, **23**, 1048-1072.

Rosa, F. and Vasciaveo, M. (2012). Agri-commodity price dynamics: The relationship between oil and agricultural market. *International Association of Agricultural Economists (IAAE)*, Triennial Conference, Foz do Iguacu, Brazil, August, 18-24.

Salisu, A.A. and Oloko, T.F. (2015). Modeling oil price–US stock nexus: A VARMA–BEKK–AGARCH approach. *Energy Economics*, **50**, 1-12.

Stock, J.H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, **55**, 1035-1056.

Stock, J.H. and Watson, M.W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11-30.

Tang, Z. and Fishwick, P.A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing*, **5**(4), 374-385.

Terasvirta T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of American Statistical Association*, **89**, 208-218.

The World Bank (2018). World development indicators: Structure of output. *World Development Indicators*, http://wdi.worldbank.org/table/4.2#.

Timmermann, A. (2006). Forecast combinations. In Graham Elliott, Clive W.J. Granger and Allan Timmermann (eds.). *Handbook of Economic Forecasting. Elsevier*, 135-196.

Torre-Torres, O.V.D., Montoya, D.A., Gracia, J.A. and Simonetti, J. (2020). Using Markov-switching models with Markov chain Monte Carlo inference methods in agricultural commodities trading. *Soft Computing*, https://link.springer.com/article/10.1007/s00500-019-04629-5

Tsangari, H. (2007). An alternative methodology for combining different forecast models. *Journal of Applied Statistics*, **34**, 403-421.

Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, **84**, 231-240.

Wang, C.S.H., Bauwens, L. and Hsiao, C. (2013). Forecasting a long memory process subject to structural breaks. *Journal of Econometrics*, **177**, 171-184.

Wu, H., Wu, H., Zhu, M., Chen, W. and Chen, W. (2017). A new method of large-scale short-term forecasting of agricultural commodity prices: Illustrated by the case of agricultural markets in Beijing. *Journal of Big Data*, **4**(1). https://link.springer.com/article/10.1186/s40537-016-0062-3

Xiong, T., Li, C. and Bao, Y. (2018). Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing*, **275**, 2831-2844.

Yu, T.H.E., Bessler, D.A. and Fuller, S. (2006). Cointegration and causality analysis of world vegetable oil and crude oil prices. *Agricultural and Applied Economics Association*, http://ideas.repec.org/p/ags/aaea06/21439.html.

Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**, 159-175.

Zhang, G.P., Zhang, B.E. and Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, **14**, 35-62.

Zhang, W. and Goh, A.T.C. (2015). Nonlinear structural modeling using multivariate adaptive regression splines. *Computers and Concrete*, **16**(4), 569-585.

Zhang, Y.J. and Zhang J.L. (2017). Volatility forecasting of crude oil market: A new hybrid method. *Journal of Forecasting*, **37**(8), 781-789.

Zhao, Y., Li, J. and Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, **66**, 9-16.