

Chapter 1

Introduction

1.0 Introduction to the Project

The main emphasis of the study is on the importance of agriculture and agricultural related information for our country India. The importance and the various objectives to be covered in the study are also to be highlighted in this chapter.

1.1 Background

Agriculture, with its allied sectors, is unquestionably the largest livelihood provider in India, more so in the vast rural areas. It also contributes a significant figure to the Gross Domestic Product (GDP). Sustainable agriculture, in terms of food security, rural employment, and environmentally sustainable technologies such as soil conservation, sustainable natural resource management and biodiversity protection, are essential for holistic rural development. Indian agriculture and allied activities have witnessed a green revolution, a white revolution, a yellow revolution and a blue revolution. Today's agricultural marketing has to undergo a series of exchanges or transfers from one person to another before it reaches the consumer. There are three marketing functions involved in this, i.e., assembling, preparation for consumption and distribution. Selling on any agricultural produce depends on some couple of factors like the demand of the product at that time, availability of storage etc. The products may be sold directly in the market or it may be stored locally for the time being. Moreover, it may be sold as it is gathered from the field or it may be cleaned, graded and processed by the farmer or the merchant of the village. Sometime processing is done because consumers want it, or sometimes to conserve the quality of that product. The task of distribution system is to match the supply with the existing demand by whole selling and retailing in various points of different markets like primary, secondary or terminal markets. Most of the agricultural products in India are sold by farmers in the private sector to

moneylenders (to whom the farmer may be indebted) or to village traders. Products are sold in various ways. For example, it might be sold at a weekly village market in the farmer's village or in a neighbouring village. If these outlets are not available, then produce might be sold at irregularly held markets in a nearby village or town, or in the mandi.

So, the Central Sector Scheme project of Agricultural Marketing Information Network (AGMARKNET) was started in March of the year 2000 by the Ministry of Agriculture under Government of India. It targeted to network all the agricultural commodities wholesale markets throughout the nation and the State Agricultural Marketing Boards and Directorates. The project is technically supported by the National Informatics Centre (NIC). The aim is to collect, break down and distribute market information to the stakeholders. More than 3200 markets are mapped under the scheme and more than 2700 markets are giving information to AGMARKNET portal. More than 350 commodities and 2000 varieties are secured under the plan. Every unit is furnished with important IT hardware, alongside web network. "AGMARK," a convenient portal has been made to encourage gathering and dissipation of information at market level. The framework is currently web empowered. It additionally gives week by week information for real markets and important agribusiness products. Online Exchange Portals are connected to the AGMARKNET entryways to give current and future costs.”

1.1.1 Directorate of Marketing and Inspection

The Directorate of Marketing and Inspection (DMI), under Ministry of Agriculture & Farmers Welfare, was set up in the year 1935. The Agricultural Marketing Adviser to the Government of India (AMA) heads the Directorate.

Its prime aim is to implement agricultural policies and plans so that an integrated development of marketing information pertaining to agriculture can take place. By maintaining a close liaison with Central Government and State Government DMI also thrives to safeguard the interest of both farmers and consumers (source: <http://dmi.gov.in/About.aspx>).

1.1.2 National Informatics Centre (NIC)

Since its establishment in the year 1976, National Informatics Centre (NIC) has been the major builder applications pertaining to the domain of e-governance having a reach up to the grassroots levels. "NICNET", connects all the Ministries and Departments of the Central Government, links 36 of the State Governments or Union Territories, and spans about 688 district administrations across India. NIC aims to implement Information Communication Technology based applications to derive competitive advantage in social as well as public administration. "NIC also has under its belt various initiatives such as Government eProcurement System (GePNIC), Office Management Software (eOffice), Hospital Management System (eHospital), Government Financial Accounting Information System (eLekha)."

For the Agricultural Marketing Network Scheme, NIC had provided computer hardware, developed the software, provided training to market personnel towards the operation of the hardware and software systems and provided internet connectivity. It has also developed the integration between the software packages developed by the various states with AGMARKNET to bring about seamless uniformity in the database.

1.1.3 State Agricultural Marketing Boards

For AGMARKNET, the State Government/Marketing Boards gave the list of markets to be secured under the AGMARKNET project. The chosen markets were to give site to establishment including offices for computer installation, phone network and computer operating personnel.

Showcase Committees/Controlling specialists of AGMARKNET hub at market level were doled out to gather pertinent information and data, encourage it and transmit it to the State level and AGMARKNET entry. NIC had additionally

prepared suitable people from every hub in working on a computer and dealing with programming bundle.

At each market hub, there is a man appointed to gather information and transmit it. A motivation plot has been acquainted with reward information section administrators for keeping up execution guidelines consistently.

1.2 Significance of Study

“The Agricultural Marketing Information System Network (AGMARKNET) which is based on the “NICNET” targets at linking all the important Agricultural Produce Market Committees (APMC), State Marketing Boards and Directorate of Marketing and Inspection regional offices. So, that there can be seamless networking of information.”

The idea is to help AGMARKNET to enhance the data quality of the reporting prices. As data quality of an agricultural commodity is of utmost significance to the concerned party, particularly when it comes to issues such as inflation. Moreover, forecasting the future prices of an agricultural commodity can help the stakeholders to be better prepared for price rise or fall. Through timely awareness, predicted values can provide the much-needed leverage of time to adapt and perform accordingly. Data filtering, sorting and data gap filling can help us to draw meaningful inferences from the agricultural data and perform activities such as forecasting and diagnostic tool development using that data for distinct markets.

1.3 Objectives of the Study

The data belonging to both arrivals and prices is available on the AGMARKNET portal, though the reported data has many shortcomings when it comes to the quality. The daily price reporting data also shows irregularities because of lack of daily updated data. Also, the daily reporting makes the data vulnerable to mistakes which might be unintentional or deliberate in nature. The study aims to find out any gaps in the data reporting, and then to bridge those gaps in order to improve upon the data quality. The study will

primarily revolve around the agricultural commodity Paddy and will focus on the study of modal prices. After the analysis of data, the study aims to fulfil two major objectives:

- To develop a tool which can report irregularities in price for major Paddy markets.
- Predict the future prices of Paddy for markets and mandis.

1.4 Remarks

In this chapter, it was seen that how AGMARKNET is emphasizing the need to collect information and is facilitating it. Various agencies involved with the AGMARKNET project were also highlighted. The importance of the project and the key objectives of the study were highlighted in this chapter.

Chapter-2

Review of Literature

2.0 Introduction to Review of Literature

A lot of literature is available on agricultural marketing and its' management. It is neither desirable nor possible to survey the whole literature. Therefore, only those relevant works have been reviewed, which reveal some general idea and provide a rationale for the present study. Such review of literatures always helps the researcher in getting an overview of the problem under study. It, moreover, helps to identify areas where in-depth research has not been carried out. Such as identification of fresh areas facilitate the taking up of new and meaningful research work.

2.1 What Is Agricultural Marketing?

Agricultural marketing involves in its simplest form the buying and selling of agricultural produce. In olden days, when the village economy was more or less self-sufficient, the marketing of agricultural produce presented no difficulty, as the farmer sold his produce direct to the consumer on a cash or barter basis. Agricultural marketing consists of all the functions and services used in moving the commodities from the producer to the final consumer. It includes not only the physical movement to the place where the product is wanted but also putting it into the form and amount is desired and having it ready at the time it is wanted.

Marketing functions: In modern marketing, the agricultural produce has to undergo a series of transfers or exchanges from one hand to another before it finally reaches the consumer. This is achieved through three important marketing functions namely

- a. **Assembling (Concentration)** – Concentration pertains to the operations concerned with the assembly and transport of produce from the field to a common assembling

area or the market.

b **Preparation for consumption(processing)** – The produce may be sold, as obtained from the field, or may be cleaned, graded, processed and packed either by the farmer or village merchant before it is taken to the market. Some of the processing is necessary for the conservation of quality.

c. **Distribution (Dispersion)** – It involves the operations of whole selling and retailing as various points. By a series of indispensable adjustments and equalizing functions, it is the task of distribution system to match the available supplies with the existing demand.

(Source: indiaagronet.com).

2.2 Management Information system

MIS is the use of information technology, people, and business processes to record, store and process data to produce information that decision makers can use to make day to day decisions.

MIS is the acronym for **Management Information Systems**. In a nutshell, MIS is a collection of systems, hardware, procedures and people that all work together to process,

It also encourages spatial dispersion of items from country territories to urban zones and between urban markets by sending the representation of data in its true form from urban shoppers to rural producers in regards to quantities and varieties required.

Current interchanges innovations open up the conceivable outcomes for market data administrations to enhance data conveyance through SMS on mobile phones and the fast development of FM radio stations in many creating nations offers the possibilities of more localized information services. Radio projects like Kisanvani on All India Radio and Kisan Suvidha on versatile give farmers the genuinely necessary data in regards to farm and farm produce delivery.

Also, there has been an increase in penetration of internet over the last 16 years. This is evident from the fact in the year 2000, India had about 0.5% of its population using the internet whereas in July, 2016 34.8% of the total population of India uses internet, be it on desktop, laptop or mobile phones. As a result, there is rapid flow of information. AGMARKNET is one such portal giving agricultural related information of various varieties of crop to its stakeholders. The information needs to be complete and correct.

2.3 Data Reporting on AGMARKNET portal

Data on the AGMARKNET portal is primarily reported as product Arrivals and Prices. The arrival data shows the amount of agricultural produce received in tonnes. Arrival data is reported on yearly, monthly, weekly and even daily basis which is used to compile reports. One can easily see which state, district or market receives maximum arrivals of a particular agricultural commodity.

On the other hand, the price data is reporting in the unit of Rs/Quintal. There are three types of prices reported on the portal they are:

- Maximum Price

The maximum prices paid for an agricultural commodity during several transactions held in a day in a market is termed at maximum price.

- Minimum Price

The minimum price paid for an agricultural commodity during several transactions held in a day in a market is termed at minimum price.

- Modal Price

The modal price is the price at which most of the commodity was sold in a market during the day.

Along with these three prices one can also get the details of state name, district name, group to which agricultural commodity belongs, variety and grade.

Thus, the information being disseminated through the AGMARKNET portal is quite comprehensive and vital to all the stakeholders of agriculture.

2.4 Overview of Paddy Production

According to Agricultural Market Information System (AMIS) and FAO, Paddy is a Kharif crop and it is sown in mid-June-mid-July and reaped in November-December. It develops well in cool, damp atmosphere and ages in a warm, dry atmosphere. The cool winters and the hot summers are helpful for a decent product. A cloudless sky having brilliant daylight amid maturing and gathering periods will improve quality Paddy. Winter precipitation is perfect.

The climatic conditions suitable for growing Paddy are given below:

- Temperature: The month to month normal temperature ought to go in the vicinity of 16°C and 27°C amid the time of sowing. Paddy grain does not develop completely if the temperature falls underneath 15°C.
- Rainfall: Paddy develops better in those ranges where precipitation happens in winter. Regions having 100 to 200 cm yearly normal rainfall develop Paddy. Irrigation serves the best if rainfall falls underneath 200 cm.

2.4.1 Paddy Statistics

“Directorate of Economics and Statistics reports that in the previous 10 years India's Paddy generation has expanded at a CAGR of 1.43 percent. However in an unlikely trend, area under cultivation of paddy has not grown significantly in last 10 years. India is the global leader in paddy cultivation with 43.2 million hectares of land under paddy cultivation. It is shortly followed by China. According to library of congress country studies, India produces 20% of the world Paddy. As per AGMARKNET database (for Domestic Price) and FAO database (at International Costs) the local costs have been lower than that of International costs since April 2014 onwards. Commission at Agricultural Costs and Prices (CACP) announces the Minimum Support Price (MSP) for Paddy and this figure has increased by a whopping 120.9 percent over the past 10 years.

2.4.2 Trade Policy

- Export Policy: Under the policy of the Government of India, the export of Paddy is free.
- Import Policy: Import of Paddy for human utilization is allowed through State Trading Enterprises but import of seeds is confined.”

2.4.3 Major Paddy Producing States (2015-16)

The Major Paddy producing states for the year 2015-16 according to Directorate of Economics and Statistics, Ministry of Agriculture are:

Table 2.1: Major Paddy producing states
 (Source: Analysis using R on AGMARKNET data)

State Name	Arrivals (Tonnes)
Punjab	11566774.99
Chattisgarh	6183534.44
Haryana	3758286.51
Uttar Pradesh	3716519.61
Madhya Pradesh	1986428.31
Uttrakhand	1809145.82
Karnataka	1558452.00
West Bengal	1363543.35
Orissa	1282701.86

Thus, it is seen that when it comes to agriculture then the quality of data reported becomes of prime importance. Data reporting of arrivals and prices on a daily basis becomes a cumbersome task often involving errors. Thus, the entire data reporting process keeps on running without any checks and controls.

2.5 Statistical Process Control

Statistical Process Control (SPC) is an industry-standard methodology for measuring and controlling quality during the manufacturing process. Quality data in the form of Product or Process measurements are obtained in real-time during manufacturing. This data is then plotted on a graph with pre-determined control limits. **Control limits** are determined by the capability of the process, whereas **specification limits** are determined by the client's needs.

Statistical process control (SPC) procedures can help you monitor process behaviour. Arguably the most successful SPC tool is the control chart, originally developed by Walter Shewhart in the early 1920s. A control chart helps you record data and lets you see when an unusual event, e.g., a very high or low observation compared with “typical” process performance, occurs.

Control charts attempt to distinguish between two types of process variation:

- Common cause variation, which is intrinsic to the process and will always be present.
- Special cause variation, which stems from external sources and indicates that the process is out of statistical control.

Various tests can help determine when an out-of-control event has occurred. However, as more tests are employed, the probability of a false alarm also increases.

Over time, other process-monitoring tools have been developed, including:

- Cumulative Sum (CUSUM) charts: the ordinate of each plotted point represents the algebraic sum of the previous ordinate and the most recent deviations from the target.
- Exponentially Weighted Moving Average (EWMA) charts: each chart point represents the weighted average of current and all previous subgroup values, giving more weight to recent process history and decreasing weights for older data.

More recently, others have advocated integrating SPC with Engineering Process Control (EPC) tools, which regularly change process inputs to improve performance.

2.6 Time Series Forecasting

Time series forecasting is the use of a model to predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

Ramasubramanian V. of Indian Agricultural Statistics Research Institute(IASR) with expertise in Agricultural Economics, Aquaculture, Artificial Neural Network has discussed various time series models for agricultural forecasting in

Forecasting Techniques in Agriculture” namely:

- Exponential Smoothing Models
- Auto Regressive Integrated Moving Average (ARIMA) Models

So, from the above references it is seen that there is evidence that forecasting of time series has been done in agriculture sector to predict the future crop yields.

2.6.1 Exponential Smoothing Models

Exponential smoothing weights past observations with exponentially decreasing weights to forecast future values. For any time period t , the smoothed value S_t is found by computing where S_i stands for smoothed observation and Y stands for the original value.

$$S_t = \alpha Y_{t-1} + (1-\alpha) * S_{t-1} \quad , \quad 0 < \alpha \leq 1, \quad t \geq 3.$$

Alpha (α) is known as the smoothing constant. The optimized value of alpha can be chosen using the value of Mean Absolute Percentage Error (MAPE) i.e. choose the value of alpha for which the value of MAPE is least.

Another form of Exponential Smoothing method is ETS (Error-Trend-Seasonality) exponential smoothing method. The ETS modelling framework was developed in 2002 IJF paper (with Hyndman, Koehler, Snyder and Grose), and in 2008 Springer book (with Koehler, Ord and Snyder). Exponential smoothing methods were originally classified by ‘Pegels’ (1969). This was later extended by Gardner (1985), modified by Hyndman et al. (2002), and extended again by Taylor (2003), giving a total of fifteen methods seen in the following table.

Trend Component	Seasonal Component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A _d (Additive damped)	A _d ,N	A _d ,A	A _d ,M
M (Multiplicative)	M,N	M,A	M,M
M _d (Multiplicative damped)	M _d ,N	M _d ,A	M _d ,M

Figure 2.2: The fifteen exponential smoothing methods

(Source: Automatic Time Series Forecasting: The forecast Package for R)

According to Rob J. Hyndman and Yeasmin Khandakar, “the cells (N, N) describes the simple exponential smoothing (or SES) method, cell (A, N) describes Holt’s linear method, and cell (Ad, N) describes the damped trend method. The additive Holt-Winters’ method is given by cell (A, A) and the multiplicative Holt-Winters’ method is given by cell (A, M). Equations for method (A, A), the Holt-Winters’ additive method.”

$$\begin{aligned}
 \text{Level:} \quad & \ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\
 \text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \\
 \text{Seasonal:} \quad & s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \\
 \text{Forecast:} \quad & \hat{y}_{t+h|t} = \ell_t + b_t h + s_{t-m+h_m^+}
 \end{aligned}$$

Figure 2.3: Equations in ETS Models

(Source: Automatic Time Series Forecasting: The forecast Package for R)

Also, according to Rob J. Hyndman and Yeasmin Khandakar, in the above equation “where m is the length of seasonality, t represents the level of the series, b_t denotes the growth, s_t is the seasonal component, $\hat{y}_{t+h|t}$ is the forecast for h periods ahead. Some interesting special cases can be obtained by setting the smoothing parameters to extreme values. For example, if $\alpha = 0$, the level is constant over time; if $\beta = 0$, the slope is constant over time; and if $\gamma = 0$, the seasonal pattern is constant over time.

2.6.2 Making Time Series Stationary

To check whether a time series is stationary or not a test known as Breakpoint Unit Root Test is used. This test has a null hypothesis that *Ho: Series has a unit root.*

If a series has unit root then it is a non-stationary series. To check the hypothesis, p-value is seen i.e. if p-value is less than 0.05 then hypothesis is rejected and time series is stationary, whereas if p-value is greater than 0.05 then hypothesis is accepted and the time series is non-stationary.

One approach to make a time series stationary is to find the differences between the data recorded. This is known as differencing. Changes, for example, logarithms can balance out the variations of a time series. Differencing can help balance out the mean of a time series.

- Random walk model

The differenced series is formed by subtracting the next observation from the previous one, and can be written as

$$y_t' = y_t - y_{t-1}$$

The differenced series will have only $t-1$ values since it is not possible to calculate a difference for the first observation.

2.7 Conclusion of the Literature

Through the review, it has been brought to light that extensive literature is available online regarding marketing information. Also, how AGMARK reports prices and arrivals to deliver timely information was also noted. There are evidences regarding the applicability of statistical process control in non-manufacturing sector and how time series forecasting has been done on the agricultural sector as well. Now, the methodology involved in the study shall be discussed.

Chapter-3

Methodology

3.0 Introduction to Methodology

In organizations, **analytics** enables professionals to convert extensive data and statistical and quantitative analysis into powerful insights that can drive efficient decisions. Therefore with **analytics** organizations can now base their decisions and strategies on data rather than on gut feelings. After the data is gathered and assessed, the sources are utilized to demonstrate a speculation or bolster a thought. A person can bring out small details to form greater assumptions about the material by using critical thinking skills effectively.

The research in this project focusses on secondary data obtained from the AGMARKNET.GOV.IN portal. Through this analysis, the study aims to bring to forward the gaps in order to develop an effective diagnostic tool and a forecasting model.

3.1 Project Phases

3.1.1 Understanding Phase

Step 1: To comprehend the broader purpose(s) of the project.

The project is targeted at developing a diagnostic tool for a particular commodity in order to improve the quality of the data reported in future. Developing a diagnostic tool will help farmers and citizens of the country to have correct information regarding the crop produce as it will help address the issues such as modal prices reported as zeros, modal prices reported as averages of maximum and irregularities in data reporting. Thereby reducing the error in reporting.

The second objective of the project is to forecast the prices of the chosen commodity. This activity will help farmers and government to understand the prices in the coming months of 2017.

Step 2: To understand the data reporting fields of AGMARKNET portal.

The AGMARKNET portal reports Arrivals (in tonnes) and Prices (in Rs/Quintal) along with crop commodities' state, district, variety, grade and date of reporting.

3.1.2 Defining Phase

Step 1: To choose a commodity for analysis.

The commodity chosen for the project was Paddy.

Step 2: To define a time frame of the data reporting period for that commodity.

The time frame for the data so that the diagnostic tool can be developed was 01 January, 2016 to 31 December, 2016. Whereas, the time period for data forecasting for the year 2017 was chosen to be 01 January, 2017 to 31 December, 2017.

Step 3: To identify the key data term(s) to be worked upon for analysis from amongst maximum price, modal price and minimum price.

The data reporting on the AGMARKNET portal consists of maximum price, minimum price and modal price. Out of them modal prices were chosen for the analysis part as this is the price at which most of the commodity is transacted in a market.

3.1.3 Data Preparation and Analysis Phase

Step 1: Tools chosen for Analysis

The software chosen for development of diagnostic tool was Microsoft Excel 2013, where the software chosen for time series forecasting was 'R', an open source programming language.

Step 2: To find the state, district and market with maximum arrivals within the reporting period.

The state with maximum arrivals for Paddy was found out. Then within that state the district and market with maximum arrivals for Paddy were found.

Step3: Check for the reporting frequency of the data fields (modal price and date).

The plot of prices reported per month was done for the whole reporting period to check the consistency in the data.

Step 4: Clean the data for better analysis.

The data gaps were identified and they were filled using the average method in order to have a consistently reported past data which can be worked upon easily to develop diagnostic tool and forecast model.

Step 5: Analyse the data to develop diagnostic tool and forecast model in the next phase.

The data was analysed for data gaps, zeros and reporting of averages of maximum prices and minimum prices in place of modal prices.

3.1.4 Developing Phase

Step 1: To develop a diagnostic tool incorporating statistical process control charts.

To develop a diagnostic tool, control chart was used. Various reports pertaining to control charts were studied and the apt chart was chosen according to the best data fit. The chart helped us to get the upper and lower control limits.

Step 2: Develop a time series forecast model for the modal price of the chosen commodity.

To forecast the time series data into 2017, Error, Trend, Seasonal (ETS) forecasting method was chosen.

3.1.5 Testing Phase

Step 1: Test the diagnostic tool on other markets reporting the same commodity. The diagnostic tool thus developed for one market was applied on other markets as well by using their respective control limits to check the applicability of the tool across markets for the same commodity.

Step 2: Check the accuracy of the forecast model.

To check the accuracy of the forecast various parameters like Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were used.

3.1.6 Findings and Recommendations Phase

After extensive analysis, one will be able to bring out some findings and recommendations through which the AGMARKNET portal can benefit. The limitations of the diagnostic tool and forecasting model are also highlighted.

3.2 Concluding Remarks

The approach towards the study of the project and the project methodology was discussed in this chapter. Now in the next chapter the data analysis, findings, limitations and recommendations regarding the study shall be discussed.

Chapter-4

Data Analysis

4.0 Introduction to Data Analysis

The commodity chosen for analysis in the understanding phase was Paddy. The commodity data for Paddy obtained from the AGMARKNET portal was usually in.xls or .csv format and various operations such as data filtering, data sorting and data cleaning was done in order to draw and make sense out of the compiled data. Since, the project had two major objectives the entire analytics process was classified into two major headings.

- Developing a Diagnostic Tool
- Developing a Forecasting Model

4.1 Development of a Diagnostic Tool

The development of diagnostic tool was a multi phased process. The phases were as follows:

4.1.1 Choosing a Tool to Analyse Data

Since, the data obtained from AGMARKNET was in excel format, so, Microsoft Excel 2013 was chosen to analyse the data and develop a diagnostic tool. MS Excel is a dynamic tool with lots of formulae, which is chosen to build a custom diagnostic tool. MS Excel is applied in various situations.

4.1.2 Selection of Market for Analysis

The arrival data for Paddy for the past one year i.e. from 01 January,2016 to 31 December,2016 was obtained for all the states from the AGMARKNET portal. The arrivals were for each state and the percentage share of total arrivals for each state was found. It was clear to see that Punjab was the leader in Paddy arrivals with 20.9% share of the total Paddy arrivals. Hence, the data of the state Punjab was analysed to find the district and market showcasing maximum arrivals in tonnes. The district that gave maximum arrivals was Faridkot with a share of 12.67% and in that district Jaitu market showed the highest percentage of arrivals of about 14.45%. So, it was decided to develop the diagnostic tool for Jaitu market of Faridkot district in Punjab. However, it was found that the Jaitu market showed all its arrival in the month of April 2016 and the data reporting of prices during the same period was also not up to the mark. So, it was decided to do analysis for the next major state in terms of Paddy arrivals viz. Chattisgarh. Hence the state of analysis was shifted from Punjab to Chattisgarh, which had 15.52% of the total Paddy arrivals during the year 2016. The Rajnandgaon district (11.93% of Paddy arrivals) ranked one in terms of Paddy arrival and within Rajnandgaon district, the main market was seen to have 73.33% of the total district arrivals. The arrivals were seen throughout the year making the market appropriate for further analysis.

Table 4.1: Arrival Data for Paddy (State wise)
(Source: Own Analysis on AGMARKNET Data)

State Name	Arrivals (Tonnes)
Punjab	11566774.99
Chattisgarh	6183534.44
Haryana	3758286.51
Uttar Pradesh	3716519.61
Madhya Pradesh	1986428.31
Uttrakhand	1809145.82
Karnataka	1558452.00
West Bengal	1363543.35
Orissa	1282701.86

Table 4.2: Punjab Arrival of Paddy (District and variety wise)
(Source: Own Analysis on AGMARKNET Data)

Row Labels	Count of Price Date
Rajpura	210
1121	131
Basmati 1509	15
D.B.	19
Other	12
Paddy	33
Amritsar	187
1121	113
Basmati 1509	37
Other	26
Paddy	11
Khanna	152
1121	1
Basumathi	113
Paddy	38
Kotkapura	129
1121	95
Basmati 1509	6
Basumathi	3
Paddy	25

Table 4.3:No. of Arrivals of Paddy (Variety wise)
(Source: Own Analysis on AGMARKNET Data)

Variety	No. of Arrivals
Paddy fine	4381
Paddy	3851
Paddy Coarse	2801
Other	2411
Common	2179
Paddy Medium	1178
Swarna Masuri (New)	1169

Table 4.4: Arrival Data for Paddy for Chattisgarh
(District and Market wise)

(Source: Own Analysis on AGMARKNET Data)

District	No. of Arrivals	Percentage of Total Arrivals
Raipur	5528	20.74%
Rajnandgaon	3237	7.82%
Dhamtari	2936	6.10%
Bastar	2457	3.69%
Mahasamund	2266	5.67%
Bilaspur	2257	8.81%
North Bastar	1329	1.24%
Durg	1271	1.24%
Janjgir	1054	12.46%
Raigarh	917	7.84%
Surguja	759	20.84%
Jashpur	410	1.05%
Kawardha	395	0.05%
Korba	170	1.73%
Kanker	159	0.38%
Koria	111	0.03%
Bijapur	95	0.05%
Dantewada	72	0.26%
Narayanpur	31	0.01%

Market	Percentage of Total Arrival	No. of Arrivals
Rajnandgaon	22.85%	1592
Kheragarh	19.82%	386
Gandai	15.78%	250
Dongargarh	1.94%	117
Dongargaon	24.92%	230
Bazar Atriya	2.01%	386
Bandhabazar	12.59%	229
Ambagarh Chowki	0.08%	19

4.1.3 Checking for Modal Price Reporting Frequency

Along with arrivals it became necessary to see the frequency with which prices were reported for Rajnandgaon market in Chattisgarh. We found the price reporting to be good throughout the year as it reported for 249 days out of 366 days.

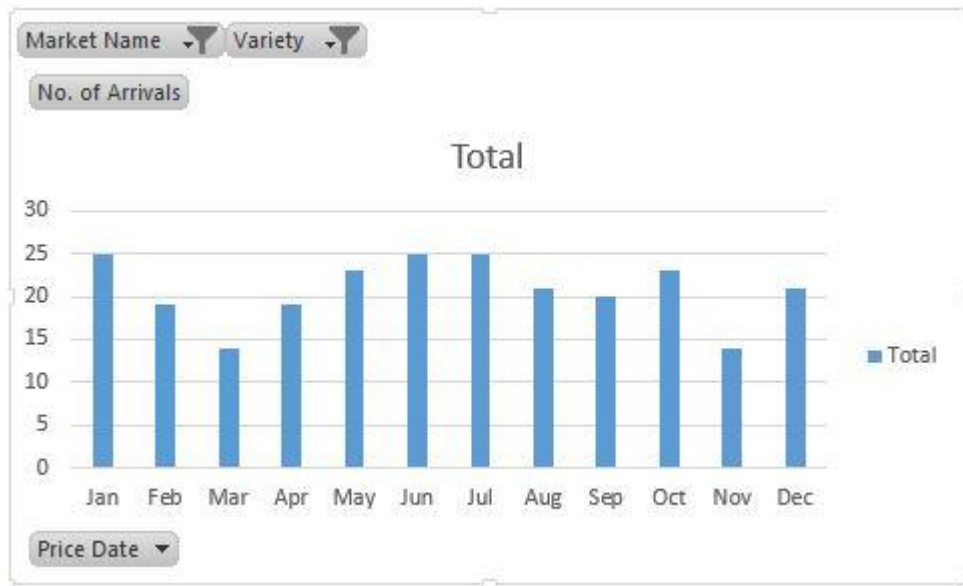


Figure 4.5: Modal Price Reporting frequency
(Source: Own Analysis on AGMARKNET Data)

4.1.4 Data Cleaning of Modal Prices

The price reporting sheet for Rajnandgaon market was analysed on the following points

- Modal Price is Average

In the 249 entries, it was checked that how many entries had modal price equal to average of maximum and minimum price by using the “IF”

statement in MS Excel. It was found that there were 2 entries in all where modal price was actually average.

- Modal Price is Zero

The reported entries were checked for where the modal price is reported to be zero. No entry was found to be reporting modal price zero. The entries having modal price as average of maximum and minimum prices were deleted. 247 entries were now reported after deletion and by using VLOOKUP command PIVOT TABLE the daily data gaps were filled by the average modal price of that month. Thereby, having a continuous series.

4.1.5 Developing the control limit and control chart for Diagnostic tool

Once the data was properly cleaned it was necessary to develop control limits. For this statistical process control charts were referred. Since, the modal prices were of variable nature, measured on a continuous scale of daily dates and the data set had a subgroup size of one (as each day’s price data cannot be combined with the next day’s data), so, the Control Chart was chosen to get the upper control limit and lower control limit.

Table 4.6:- Diagnostic Tool in Excel

(Source: Own Analysis on AGMARKNET Data)

Arrival Date	Modal Price	Mean	LCL	UCL	Out of Bound
01-Jan-16	1394.4	1541.733	1231.223	1852.242	
02-Jan-16	1394.4	1541.733	1231.223	1852.242	
03-Jan-16	1394.4	1541.733	1231.223	1852.242	
04-Jan-16	1360	1541.733	1231.223	1852.242	
05-Jan-16	1350	1541.733	1231.223	1852.242	
06-Jan-16	1391	1541.733	1231.223	1852.242	
07-Jan-16	1350	1541.733	1231.223	1852.242	
08-Jan-16	1391	1541.733	1231.223	1852.242	
09-Jan-16	1391	1541.733	1231.223	1852.242	
10-Jan-16	1394.4	1541.733	1231.223	1852.242	
11-Jan-16	1390	1541.733	1231.223	1852.242	
12-Jan-16	1390	1541.733	1231.223	1852.242	
13-Jan-16	1394.4	1541.733	1231.223	1852.242	
14-Jan-16	1400	1541.733	1231.223	1852.242	
15-Jan-16	1380	1541.733	1231.223	1852.242	
16-Jan-16	1405	1541.733	1231.223	1852.242	

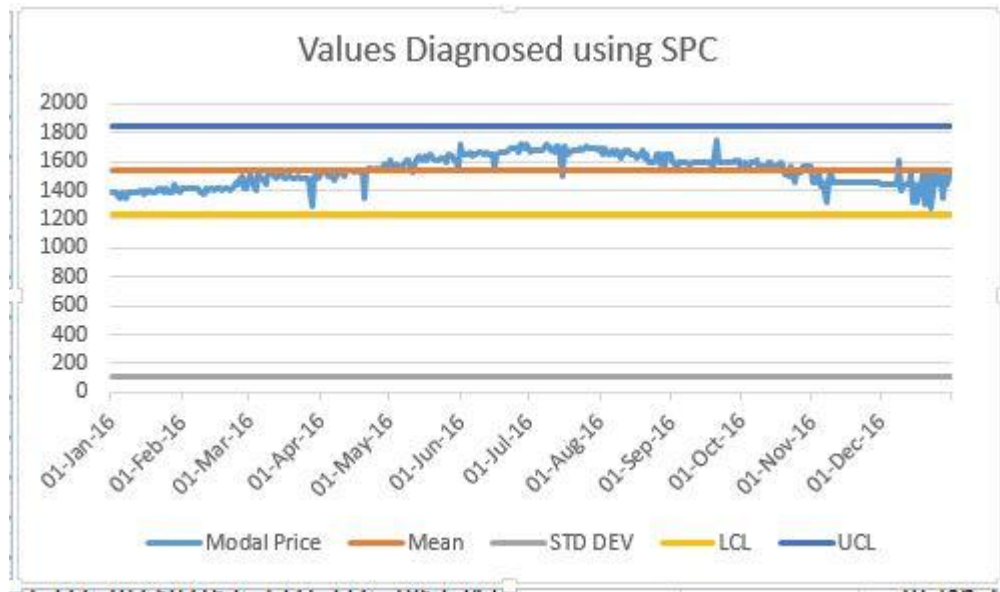


Figure 4.7: Plot of Diagnostic Tool in excel
(Source: Own Analysis on AGMARKNET Data)

4.1.6 Developing the Interface of the diagnostic tool

The diagnostic tool was a standalone excel sheet where users can enter the data for modal price to check whether the price values were within the range or not. The excel sheet also consisted of dynamic Control Charts, reflecting new data value entered dynamically. The tool prompted for a warning message "Out of bound" if the date did not belong to 2016 or if the modal price was out of the UCL and LCL.

Table 4.8: Diagnostic Tool in Excel (Showing out of bound Data)

(Source: Own Analysis on AGMARKNET Data)

Arrival Date	Modal Price	Mean	LCL	UCL	Out of Bound
01-Jan-16	1394.4	1543.509	1226.258	1860.759	
02-Jan-16	1394.4	1543.509	1226.258	1860.759	
03-Jan-16	1394.4	1543.509	1226.258	1860.759	
04-Jan-16	1360	1543.509	1226.258	1860.759	
05-Jan-16	1350	1543.509	1226.258	1860.759	
06-Jan-16	1391	1543.509	1226.258	1860.759	
07-Jan-16	2000	1543.509	1226.258	1860.759	OUT OF BOUND
08-Jan-16	1391	1543.509	1226.258	1860.759	
09-Jan-16	1391	1543.509	1226.258	1860.759	
10-Jan-16	1394.4	1543.509	1226.258	1860.759	
11-Jan-16	1390	1543.509	1226.258	1860.759	
12-Jan-16	1390	1543.509	1226.258	1860.759	
13-Jan-16	1394.4	1543.509	1226.258	1860.759	
14-Jan-16	1400	1543.509	1226.258	1860.759	
15-Jan-16	1380	1543.509	1226.258	1860.759	
16-Jan-16	1405	1543.509	1226.258	1860.759	

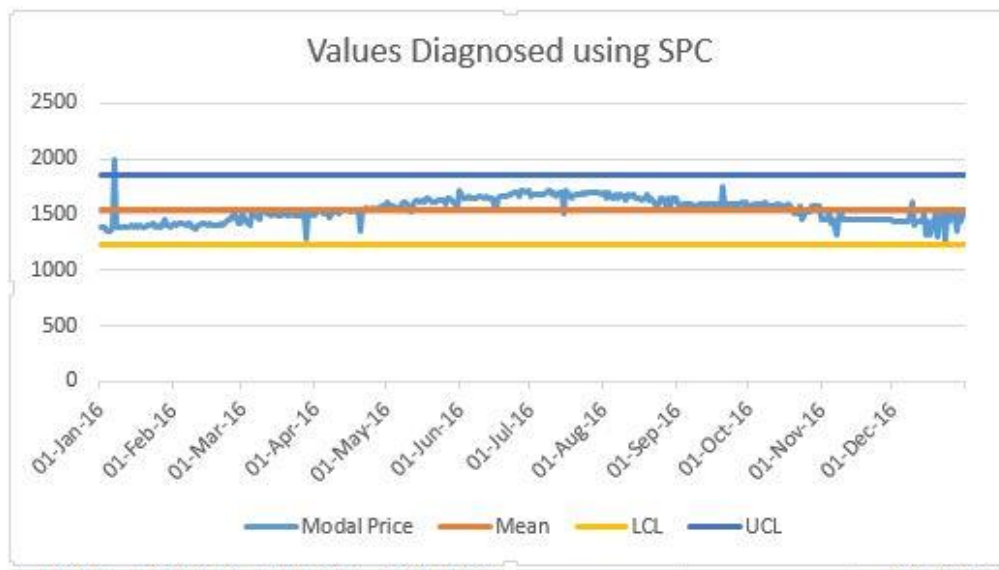


Figure 4.9: Plot of Diagnostic Tool in excel (Showing out of bound data)

(Source: Own Analysis on AGMARKNET Data)

The activities performed for one year data of 2016 of Rajnandgaon Market to check the applicability of the developed diagnostic tool is performed in following steps as:

- Drawing Inferences from the Arrivals data
- Checking for modal price Reporting frequency
- Cleaning of data for Modal Prices
- Developing a control limit & Control chart
- Developing the logic & Interface4

4.2 Forecasting the Modal Prices of Paddy

In order to forecast the Paddy modal prices for the year 2017 based on the modal prices of the previous years following steps were followed:

4.2.1 Choosing the Tools of Analyses

To develop a better forecast model two tools were used

- Microsoft Excel 2013: For sorting, filtering and cleaning data
- R (Language) To help choose a forecast model

4.2.2 Drawing Inferences from the Arrivals Data

It was aimed to develop forecast model for one of the top Paddy producing markets of Chattisgarh, and for that it became necessary to look into the arrivals data of Chattisgarh's markets.

But after analysing the data of Surajpur Market (leading market in terms of arrivals) belonging to the leading district Surguja in terms of Paddy arrivals it was found that there were a large data gaps because of price reporting gaps.

Table 4.10: Arrivals of paddy Variety wise and marketwise
(Source: Own Analysis on AGMARKNET Data)

Variety	No. of Arrivals
Paddy fine	4381
Paddy	3851
Paddy Coarse	2801
Other	2411
Common	2179
Paddy Medium	1178
Swarna Masuri (New)	1169

Market	No. of Arrivals
Rajnandgaon	1772
1001	251
Paddy Medium	250
Swarna Masuri (New)	250
MTU-1010	250
Other	249
Paddy fine	249
HMT	245
I.R. 64	28

Paddy fine is the leading variety of Paddy with maximum number of arrivals as shown in the figure at 4381 arrivals. Also, in our chosen market, Rajnandgaon the arrival frequency is fine as well as the number of arrivals standing at 249. So, the choice of this particular market and this variety is apt for analysis.

4.2.3 Data Cleaning of Modal Prices

To forecast the data, it was necessary to fill the data gaps and improve the quality of data. For this it was seen that how many number of times (count) the Paddy-Paddy fine variety reported is month i.e. in days. To bridge the data gaps, the months which had more than 15 days of reporting of modal prices had their remaining non-reported days filled with the averages of that month. Whereas, the months having data reported days' count greater than 10 but less than or equal to 15 were replaced by the average of the modal price values present of that month and the average of annual modal price. Lastly, if the count of the modal price

reported days was less than 10, then, the missing values were replaced by annual average.

4.2.4 Choosing a Forecast Model and Technique

R is the perfect programming language for forecasting when a time series is involved. STLF framework was used which utilises different forecasting model and smoothen the data through a single algorithm.

First of all, it was checked that was the series stationary or not. Using the Dickey-Fuller test, the p-value came out to be 0.480 which means that the null hypothesis “Series has a unit root” was not rejected, this acceptance means that the time series is **non-stationary** in nature.

```
> adf.test(y1)
Augmented Dickey-Fuller Test
data: y1
Dickey-Fuller = -2.232, Lag order = 11, p-value = 0.480
alternative hypothesis: stationary
```

Figure 4.11: Dickey Fuller Test for Stationary

(Source: Analysis using R on AGMARKNET data)

So, in order to make the time series stationary first level difference was taken. We again ran the unit root test on it based on Dickey-Fuller test to get the p-value. The p-value this time was found to be less than 0.01, hence the null hypothesis was rejected, meaning that it does not have a unit root and thus is **stationary** in nature.

```
> adf.test(y2, k=0)

Augmented Dickey-Fuller Test

data: y2
Dickey-Fuller = -65.934, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary

warning message:
In adf.test(y2, k = 0) : p-value smaller than printed p-value
> |
```

Figure 4.12: Dickey Fuller Test for Difference
(Source: Analysis using R on AGMARKNET data)

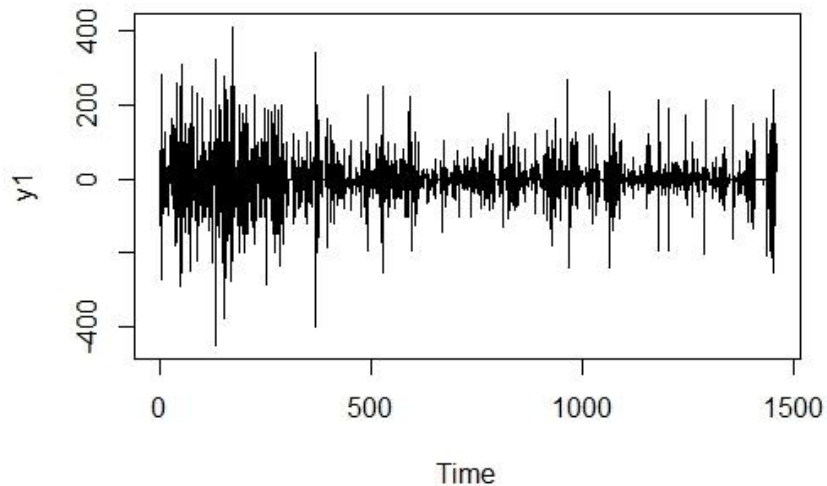


Figure 4.13: Stationary Series Plot
(Source: Analysis using R on AGMARKNET data)

Stationary series was found at the first level of difference. The mean and variance are time independent.

R has a several time series packages which are helpful in analysis and helps users to find the best fit model out of the given models. The ETS (Error-Trend-Seasonality) Exponential Smoothing of forecasting gives the best fit model according to the select criteria and justifies the chosen model statistically.

Due to the limitation of the ETS model. ETS cannot handle frequency over 24 which limit us to use STL function which is a combination of STF () (Seasonal and Trend decomposition using Loess) + F () (Forecasting) functions.

R program is developed using Time Series and Forecast packages that includes-

- Reading Excel files from AGMARKNET.
- Converting data into Time Series.
- Dividing data into Training and Test Data.
- Checking the data for seasonality (Dickey-Fuller Test).
- Checking the data for Correlation (ACF and PACF function).
- Removal of Seasonality Error and Trend in data.
- Applying the ETS modal and forecasting future values
- Transfer of Forecasted Data into Excel sheet for further evaluation

```

1 # Time Series Forecasting of paddy of year 2016-2017
2 # install packages
3 library(tseries)
4 library(forecast)
5
6 #Reading Data from Excel file
7 mydata<- read.csv("C:/Users/SHAKTI/Desktop/SHIVESH PROJECT/s4.csv")
8 attach(mydata)
9
10 # Descriptive statistics and plotting the data
11 summary(mydata)
12 plot.ts(mydata$PRICE)
13 Summary(mydata)
14
15 #Converting the data into Training data and Test data
16 a1<-ts(mydata$PRICE[c(1:1095)], frequency=365) #Training Data
17 a2<-ts(mydata$PRICE[c(1095:1461)], frequency=365) #Test Data
18 a3<-ts(mydata$PRICE[c(1:1461)],frequency=365)
19
20 ##Testing the Data for seasonality and correlation
21 # ACF and PACF (Auto Correlation Function and Partial Auto Correlation Function)
22 acf(a1)
23 pacf(a1)
24
25 #Applying the ETS(Error,Trend,seasonal) in Testing data
26 #ets(y, model="ZZZ", damped=NULL, alpha=NULL, beta=NULL,gamma=NULL)
27 summary(a1.ets)
28 plot(a1.ets)
29
30 #Transfer the data to excel file
31 write.table(a1.ets,file="forecated10.csv", append=TRUE,sep="," ,col.names=FALSE,row.names=FALSE)
32
33 #Applying model for 2017 data
34 f1.ets<-stlf(a3,method="ets",h=100)
35 summary(f1.ets)
36 plot(f1.ets)
37
38 #Transfer the data to excel file
39 write.table(f1.ets,file="forecated11.csv", append=TRUE,sep="," ,col.names=FALSE,row.names=FALSE)

```

Figure 4.14R program for forecasting

(Source: Analysis using R on AGMARKNET data)

The ETS exponential smoothing method was first employed on the period 01 January 2013 to 31 December 2015 to forecast the modal prices for the month January to December 2016. Since the data for 2016 was already present it was easy to check the model's accuracy by finding out various errors.

In order to find appropriate values of Error, Trend and Seasonality the model was chosen as “ZZZ” which mean automatic selection of Error, Trend and Seasonality for best fit of model. To further optimize the model the frequency is set as 365, for no. of days in a year.

We split our data into Training data (data from 2013 to 2015) and Testing data (data of 2016) to compare the forecasted values of 2016 with forecasted values of 2016.

After running the function, it was found that model gives “M N N” where M defines multiplicative error in data and no trend and seasonality. Value of parameters are-

Alpha=0.1785

Beta=0

Gamma=0

```
Forecast method: STL + ETS(M,N,N)
Model Information:
ETS(M,N,N)
Call:
ets(y = x, model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)
Smoothing parameters:
alpha = 0.1785
Initial states:
l = 1454.2366
sigma: 0.0348
AIC      AICC      BIC
16206.53 16206.55 16221.53
Error measures:
           ME      RMSE      MAE      MPE      MAPE      MASE
Training set -0.364504 50.57707 37.73693 -0.1375963 2.621538 0.2765407
           ACF1
Training set 0.005195634
Forecasts:
Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
4.0000 1296.252 1234.526 1357.979 1201.8500 1390.655
4.0027 1298.917 1236.213 1361.620 1203.0199 1394.813
4.0055 1307.184 1243.519 1370.850 1209.8159 1404.553
4.0082 1272.219 1207.605 1336.833 1173.4010 1371.037
4.0110 1332.459 1266.911 1398.007 1232.2124 1432.706
4.0137 1457.460 1390.991 1523.929 1355.8042 1559.116
4.0164 1344.867 1277.489 1412.244 1241.8211 1447.912
4.0192 1360.293 1292.019 1428.568 1255.8766 1464.710
```

Figure 4.15R Summary of forecasted Modal Price for2016
(Source: Analysis using R on AGMARKNET data)

Forecasts from STL + ETS(M,N,N)

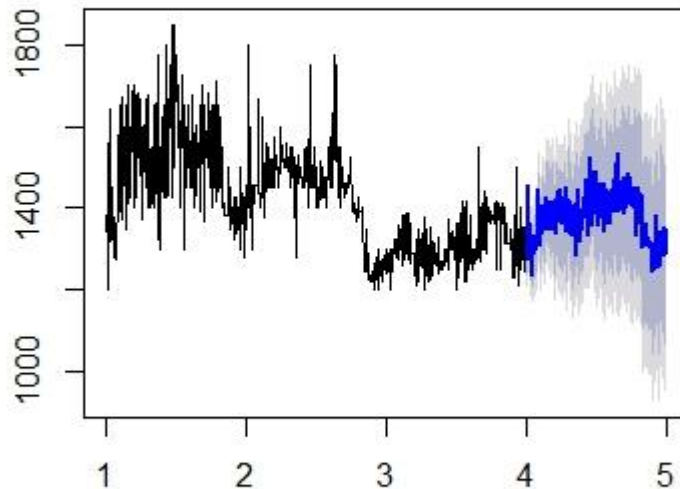


Figure 4.16R Plot of forecasted Modal Price for 2016
(**Source:** Analysis using R on AGMARKNET data)

The Blue line shows the forecasted data for 2016 and grey area is 80% and 90% confidence limit of the forecast.

4.2.5 Forecast Errors for the Months of year 2016

The program returns daily forecast of modal price for 366 days of 2016 with the 80% and 95% confidence limit in a excel file, then mean absolute error and mean absolute error were calculated. Graph between actual modal price and forecasted modal price with 95% confidence limit is plotted using excel.

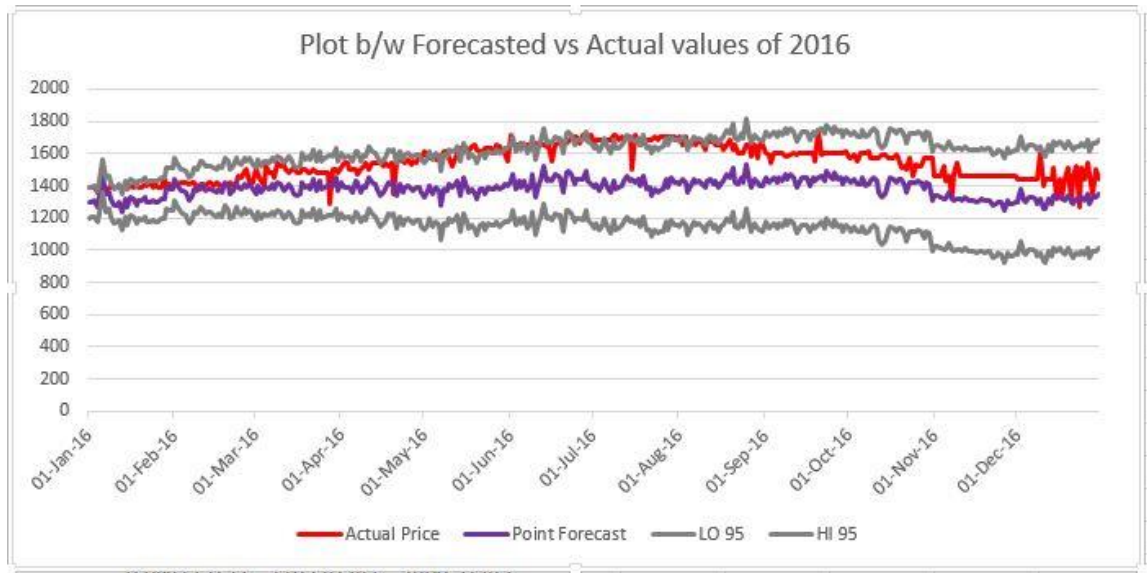


Figure 4.17Plot between Actual and Forecasted Values with Confidence limit for 2016

(Source: Analysis on Forecast R data and AGMARKNET data)

Table 4.18Actual and Forecasted Values with Confidence limit for 2016

(Source: Analysis on Forecast R data and AGMARKNET data)

Arrival Date	Actual Price	Point Forecast	Percentage Error in Forecast	LO 95	HI 95
01-Jan-16	1394.4	1296.252298	7.04%	1201.85003	1390.65456
02-Jan-16	1394.4	1298.916624	6.85%	1203.01985	1394.8134
03-Jan-16	1394.4	1307.184299	6.25%	1209.8159	1404.5527
04-Jan-16	1360	1272.219198	6.45%	1173.40103	1371.03737
05-Jan-16	1350	1332.459465	1.30%	1232.21243	1432.7065
06-Jan-16	1391	1457.460067	4.78%	1355.8042	1559.11593
07-Jan-16	1350	1344.866595	0.38%	1241.8211	1447.91209
08-Jan-16	1391	1360.29328	2.21%	1255.8766	1464.70996
09-Jan-16	1391	1317.761962	5.27%	1211.99182	1423.53211
10-Jan-16	1394.4	1275.324632	8.54%	1168.21807	1382.43119
11-Jan-16	1390	1280.060123	7.91%	1171.63357	1388.48668
12-Jan-16	1390	1296.759264	6.71%	1187.02854	1406.48999

The Forecast accuracy using the program is found out to be 89.84%. The forecasted value were close to the actual values of modal price, and hence the data from 01 January 2013 to 31 December 2016 was used to forecast the values of various months of 2017 using the same model constraints obtained earlier.

4.3 Findings

The Forecasting Program was applied to the whole data from 2013 to 2016 to get forecasts for the months of January, February and March 2017.

After running the function, it was found that model gives “M N N” where M defines multiplicative error in data and no trend and seasonality. Value of parameters are-

```

Forecast method: STL + ETS(M,N,N)
Model Information:
ETS(M,N,N)
Call:
ets(y = x, model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)
Smoothing parameters:
alpha = 0.1823
Initial states:
l = 1448.0924
sigma: 0.0315
AIC      AICC     BIC
21814.15 21814.17 21830.01
Error measures:
           ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.4693326 46.29023 33.24152 -0.06771418 2.279626 0.1975816
           ACF1
Training set -0.009737437
Forecasts:
Point Forecast  Lo 80  Hi 80  Lo 95  Hi 95
5.0027  1527.826 1464.260 1591.392 1430.610 1625.042
5.0055  1529.405 1464.790 1594.020 1430.586 1628.224
5.0082  1506.417 1440.770 1572.063 1406.019 1606.815
5.0110  1512.442 1445.780 1579.105 1410.491 1614.394
5.0137  1609.596 1541.933 1677.259 1506.114 1713.078
5.0164  1502.707 1434.057 1571.356 1397.716 1607.697
5.0192  1556.400 1486.778 1626.022 1449.923 1662.877
5.0219  1522.728 1452.147 1593.308 1414.784 1630.671
5.0247  1495.798 1424.271 1567.325 1386.407 1605.189
Alpha=0.1823          Beta=0          Gamma=0

```

Figure 4.19R Summary of forecasted Modal Price for2017

(Source: Analysis using R on AGMARKNET data)

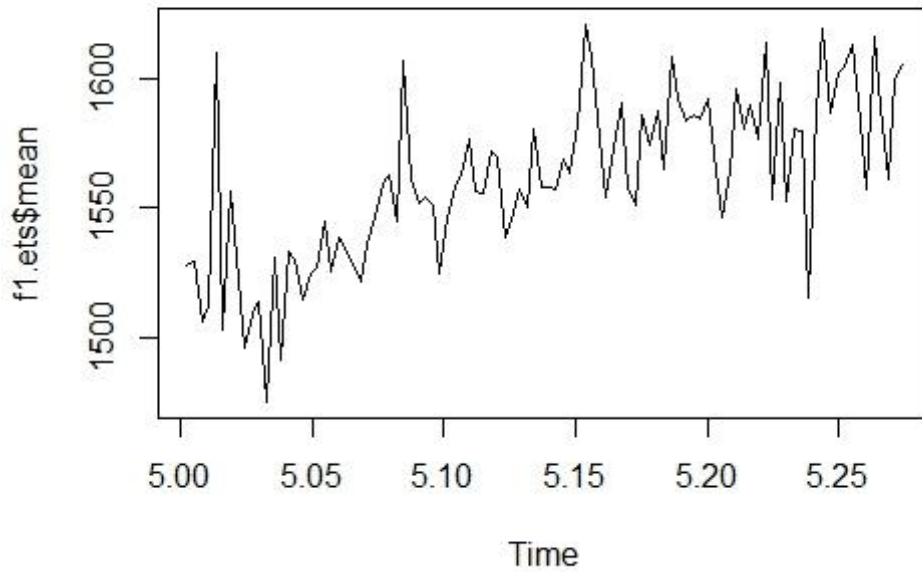


Figure 4.20R Plot of forecasted price for 2017
 (Source: Analysis using R on AGMARKNET data)

Forecasts from STL + ETS(M,N,N)

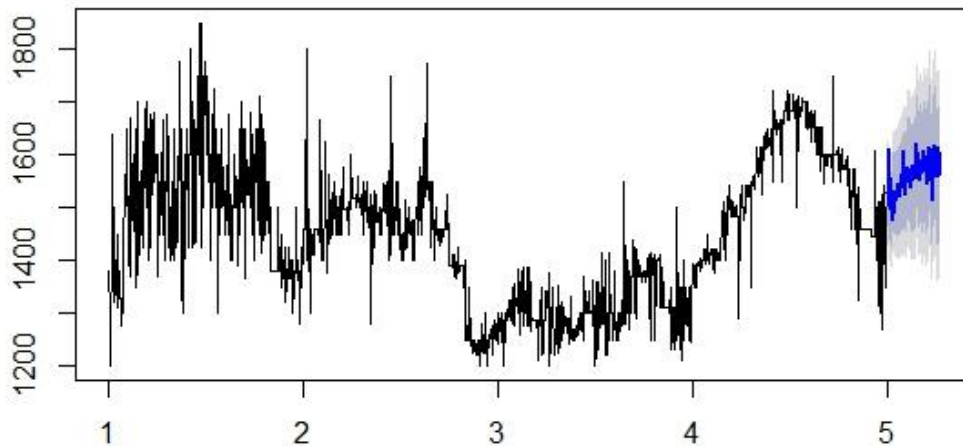


Figure 4.21R Plot of previous data and forecasted price from 2013 to 2017
 (Source: Analysis using R on AGMARKNET data)

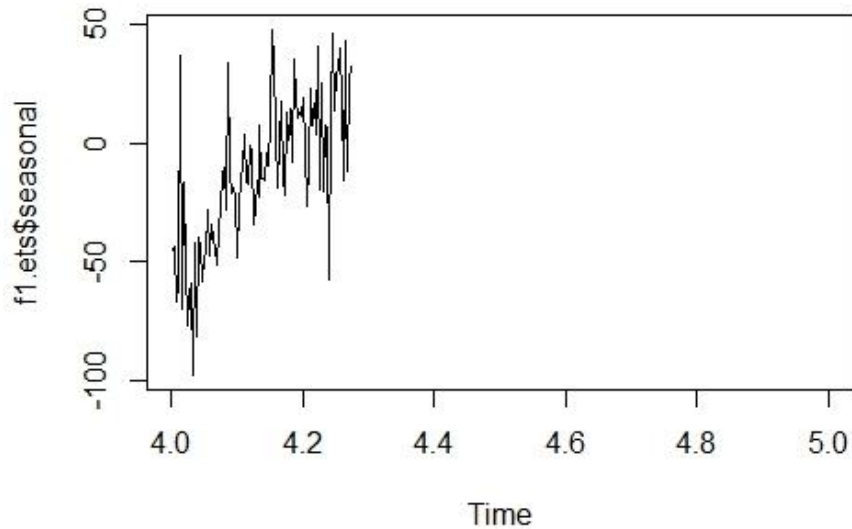


Figure 4.22 R Plot of seasonality

(Source: Analysis using R on AGMARKNET data)

The program finds out seasonality in first quarter of 2016 which was automatically get rectified using the ETS.

The MAPE for first three months of 2017 was 3.38% whereas the MAE was Rs-51.41/Quintal. The Market was Rajnandgaon and Paddy variety Paddy Fine

Date	Actual	point forecast	percentage error			LO 95	HI 95
01-01-2017	1549.8	1527.825971	1.42%	21.97	MAE	1430.61	1625.042
02-01-2017	1625	1529.405024	5.88%	95.59	51.41824	1430.586	1628.224
03-01-2017	1511	1506.41694	0.30%	4.58		1406.019	1606.815
04-01-2017	1545	1512.442309	2.11%	32.56	MAPE	1410.491	1614.394
05-01-2017	1340	1609.59606	20.12%	269.60	3.39%	1506.114	1713.078
06-01-2017	1511	1502.706598	0.55%	8.29		1397.716	1607.697
07-01-2017	1549.8	1556.399888	0.43%	6.60		1449.923	1662.877
08-01-2017	1549.8	1522.727593	1.75%	27.07		1414.784	1630.671
09-01-2017	1620	1495.797786	7.67%	124.20		1386.407	1605.189
10-01-2017	1621	1508.929472	6.91%	112.07		1398.111	1619.748
11-01-2017	1597	1513.835402	5.21%	83.16		1401.607	1626.064
12-01-2017	1550	1475.086304	4.83%	74.91		1361.465	1588.708
13-01-2017	1607	1530.917489	4.73%	76.08		1415.92	1645.915
14-01-2017	1600	1491.252713	6.80%	108.75		1374.896	1607.609

Figure 4.23 Actual and Forecasted Values with Confidence limit for 2017

(Source: Analysis on Forecast R data and AGMARKNET data)

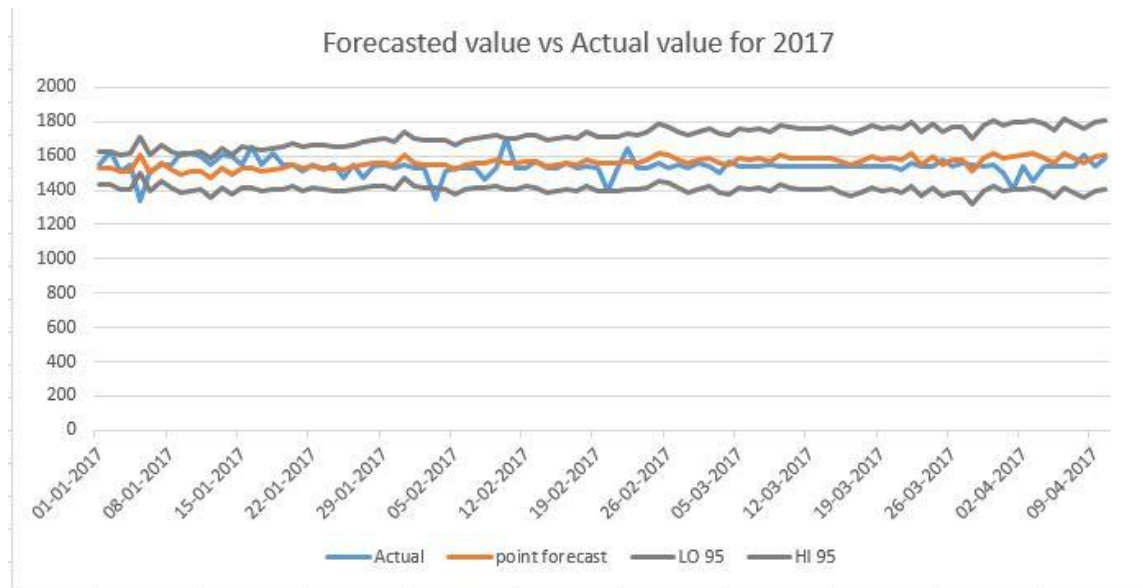


Figure 4.24 Plot between Actual and Forecasted Values with Confidence limit for 2016
 (Source: Analysis on Forecast R data and AGMARKNET data)

Month	MAE	MAPE
January 2017	52.36	3.41%
February 2017	48.82	3.25%
March 2017	52.49	3.46%
January –March 2017	51.41	3.39%

Table 4.25: MAE and MAPE values for 2017
 (Source: Own Analysis on AGMARKNET Data)

4.4 Limitations

- The program will not be able to take factors like quality of Paddy, demand in market, weather conditions as parameters for forecasting which are the main factors for selling price in the market.
- The analysis was done only on modal price and not on maximum and minimum prices.
- The data used for the development of the diagnostic tool was only of one year i.e. of 2016. To develop a better tool the data horizon can be widened to include previous year data.
- The diagnostic tool needs to be developed for each market and each variety separately which can make the task cumbersome.
- The data used for forecasting the modal prices of 2017 was taken from 01 January 2012 to 31 December 2016, giving a forecast horizon of only five years which can be made large.
- The forecast was done only for one market and one specific variety of Paddy. Thus, it can only be applied to that market and that variety of Paddy.

4.5 Concluding Remarks and Recommendations

After the data analysis, the diagnostic tool was developed for major Soybean markets (in terms of arrivals) of Chattisgarh. Also, the time series forecasting for Paddy fine variety of Paddy for the chosen market (Rajnandgaon-Chattisgarh) was done. To conclude the data analysis of the project, here are some recommendations.

- The diagnostic tool developed as part of the project work can be accommodated alongside the data entry reporting portal of Soybean prices i.e. http://agmarknet.nic.in/market_online/. Integration with the portal will give the checks and raise alerts then and there itself, making the data entry process effective and thereby improving the data quality of reporting.
- The price data reporting frequency should be monitored daily as it can help provide the actual data. This actual data can in turn help in developing robust diagnostic tools and accurate forecast models. For example, Punjab-the leading Paddy market for 2016 had a poor reporting frequency and the market of analysis was shifted.
- The time series forecast in the case of this study was done for a specific variety of Paddy and for a specific market but instead the forecast model needs to be developed for all varieties of Soybean across various markets.
- The variety terms of Paddy on AGMARKNET portal like “Others” need to be classified as well to maintain effective tracking on the Paddy arrivals.
- In the findings of the study it was seen that the data reported historically pertaining to an agricultural commodity can be edited and changed. The editing of already reported data can disturb and hamper the entire data analysis of the project and thus, the changing of the data should be done in a restrictive manner to ensure integrity of reported data.

Chapter 5

Conclusion of the Study

The project began by understanding the agencies associated with AGMARKNET. Also, the importance of the study and objectives of the study were highlighted which gave a direction to develop a well-planned methodological approach to the project study.

Further on in the project, the review of literature had highlighted the importance of the information particularly when it comes to the aspect of agricultural marketing information. The review of literature had also brought to light the need manage and improve the quality of data being reported. Various definitions of marketing information were obtained. The working of AGMARKNET, an agricultural commodity data reporting portal which can help in bringing insightful information to the farmers, government and consumers was understood. It was also seen how the data is being reported on this portal. The agricultural commodity's data worked on in this project was that of Paddy. It was also necessary to understand some facts related to Paddy like major Paddy producing states, temperature and rainfall conditions and trade policies, and the same was done in this study.

The study had also highlighted how statistical process control can be used in non-manufacturing sector to continuously monitor the data being reported. In this project, it was also emphasized that which statistical process control charts can be used to have mechanism of check on the prices being reported on the AGMARKNET portal for a particular commodity such as Paddy. The equations of upper control limit, lower control limit and control limit were highlighted for process control chart. It was also discussed that how the control limits can be improved by entering new arrival data for each month so that marginally neglected prices can be accommodated. The diagnostic tool developed on the platform of Microsoft Excel was on the basis of statistical process control charts for one major Paddy arrival market of Chattisgarh.

To check whether the approach adopted for developing the diagnostic tool was scalable to other markets or not, the same approach was applied to other four markets of Chattisgarh. It was seen that more than 90% of the modal prices reported for all the five markets was within the limits developed by using statistical process control chart. The diagnostic tool also took care of the modal prices that were being reported as average of maximum and minimum prices and reported a warning whenever modal prices was average.

Also, the study of this project emphasized the importance of time series forecasting particularly when it comes to the prices of agricultural commodities. Through the review of the available literature it was seen that what were the different models and techniques to analyse the time series. With a well-defined mechanism for data gap filling and after applying data filtering and data sorting in Microsoft Excel 2013, R was used to develop a suitable model to forecast the modal prices for Soybean for the year 2017. Using the ETS forecasting gave a particular model, this model was applied to a subset of the dataset and was checked for errors. The error percentage came out to be less than 4%. Then the ETS model was applied on the entire data set to get the forecasted values for the year 2017. Till March 2017, the mean absolute error percentage was 3.39%. The forecasting of time series was done for one particular market of Chattisgarh and for one specific variety of Paddy.

Thus, we see that by developing a diagnostic tool for various markets can help in improving the quality of data being reported as it will create a mechanism of checks and balances. The need for developing these kinds of diagnostic tools across all markets and for all varieties of agricultural commodities was emphasized in the study. Whereas, the forecasted modal prices will assure the farmers, government agencies and consumers to get a prior information regarding the agricultural commodities. It was also highlighted that with a wider range of data set the forecasted model thus developed can be more accurate.

Bibliography

- Biswas, A. (2004, JULY 13). Domain-b. Retrieved from Domain-b: http://www.domain-b.com/economy/agriculture/20040713_marketing.html
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 2-3.
- InfinityQS International, I. (2014, January). A Practical Guide to Selecting the Right Control Charts. Fairfax, Virginia, United States of America.
- S. A. S. (2008). Agricultural Marketing in India. In A. S. S., *Agricultural Marketing In India* (pp. 1-8). Oxford & Ibh.
- Statistical Process Control (SPC). (n.d.). Retrieved from www.moresteam.com: <https://www.moresteam.com/toolbox/statistical-process-control-spc.cfm>
- United Nations Report (1995). *The State of Food and Agriculture*. Rome: Food and Agriculture Organization.
- Thomsen, F. L. (1951). *Agricultural Marketing*.
- www.internetlivestats.com. (2017, March). Retrieved from internetlivestats
- Gareth James, Daniela Witten, Trevor Hastie Robert Tibshirani; *An Introduction to Statistical Learning with Applications in R*.
- <https://stats.stackexchange.com/questions/63250/forecasting-function-r-holt-winters-hw-approach>.
- <https://www.rstudio.com/wp-content/uploads/2016/11/ggplot2-cheatsheet-2.1.pdf>, Data visualization with ggplot2 RStudio Cheat Sheet.
- Rob Hyndman [aut, cre, cph], Mitchell O'Hara-Wild [aut], Christoph Bergmeir [aut], Slava Razbash [aut], Earo Wang [aut] *Forecasting Functions for Time Series and Linear Models*.
- <https://stats.stackexchange.com/questions/72243/holts-linear-and-holt-winters-in-r>