STATISTICAL ANALYSIS ON PRICE SERIES OF ARECA NUT

Harsha S[*1], Ismail B[1] and Sinchana N[1]
[1]Department of PG Studies and Research in Statistics Mangalore University, Mangalagangothri, Konaje, INDIA

ABSTRACT: Trade in areca nut is always a gamble and farmers are the worst hit. The price advantages are not transferred to farmers and the traders are making hay under such circumstances. Lack of market information and intelligence has pushed the farmers to the state of price takers. In order to protect the farmers and save those from distress sale of areca nut and to guide those in making proper decisions with regard to marketing we decided to carry out the analysis on price patterns of Areca nut. This study intended to carry out a detailed statistical analysis on Areca nut price with respect to two varieties namely, "cqca" and "New variety" related to Puttur areca nut market in Dakshina Kannada, Karnataka. We use some of the sophisticated statistical techniques like descriptive statistics, Transition probabilities for price rise and fall, Cluster analysis to identify and describe the price patterns. Further we adopted polynomial fitting and ARIMA class models to forecast the price series.

Keywords: Transition probability, Clustering, Polynomial fitting, ARIMA Model.

1.INTRODUCTION

Areca nut or betel nut (*Areca catechu L.*,) is an important cash crop in the Western Ghats, Eastern Ghats, East and North Eastern regions of India. Areca plant is a tall-stemmed erect palm, reaching varied height depending upon the environmental conditions. Palms attaining a height of 30 meters are not uncommon which poses problems for harvesting of nuts. Areca nut is an important component of the religious, social and cultural celebrations and economic life of people in India. Areca nut is also used in Ayurvedic and veterinary medicines. The habit of chewing areca nut is typical of the Indian subcontinent and its neighborhood. Although, production of areca nut is localized in a few states, the commercial product is widely distributed all over the country. About 20 per cent of total areca production in the country is consumed as ripe fruit.

Importance of Areca nut and Marketing

India is traditionally an Areca growing country; the India-Pakistan partition in 1947 led to India losing nearly 50 per cent of areca nut area to Pakistan. In the early fifties, the internal demand for areca nut went up and had to be made good by way of imports. The government of India decided to encourage large scale areca nut cultivation in India and institutional finance was provided through co-operatives and scheduled banks with an intention of harnessing the potential of areca nut cultivation as well as to avoid foreign exchange drain. Consequently, the production steadily increased and import was stopped. The emergence of new products such as *pan masala* and *gutkha* further gave a fillip to demand for areca nut and has resulted in remunerative prices for farmers leading to rapid expansion in area, not only in traditional belts, but also to non-traditional plains of Karnataka. Areca nut is a notified commodity in about 32 regulated markets of Karnataka. There are more than 15 co-operative marketing societies handling areca nut in this state. The co-operative societies have been fairly successful in their functioning and about 30 per cent of the marketable surplus in the state being handled by them. All these co-operatives are functioning as the agencies of CAMPCO (Cocoa and Areca nut Marketing and Processing Co-operative) Ltd., Mangalore.

As production of Areca nut have been increasing over the years, study on the price pattern of areca nut is much necessary to bridge gap between market and farmer. Thus we decided to carry out the analysis on price patterns of Areca nut in Puttur.

About Data

The data set which we used for analysis is collected from "agmarknet.gov.in" site with respect to the region Puttur, Karnataka. Our Study period includes weekly data from January 2012 to December 2016. Price series is collected with respect to two varieties namely, cqca and new variety. In order to easily and effectively carry out this work, we used R (3.4.4) software in each stage of the analysis.
In the next section we present brief review related to the analysis of the price series of the areca nut which are carried out in the past years. Section 3 gives clear picture about use of different Statistical methods with necessary references.
In the Section 4, the various results found throughout the analysis were arranged systematically. The major findings of the study and the conclusions notes were given in the last section.

2. LITERATURE REVIEW

Jose,C.T et al.(2013) forecasted minimum, maximum and average areca nut (*Areca catechu* L.) prices in the major Areca nut markets of the Assam as well as Meghalaya based on the monthly price data. Monthly minimum, maximum, and average market price data of areca nut (in Rs./quintal) for the period May-2003 to March-2012 (for Assam) and February-2003 to March-2012 (for Meghalaya) were used. Box-Jenkins autoregressive integrated moving average (ARIMA) methodology was adopted for developing the models. An interrupted time-series model was also applied to resolve the problem of intervention point (October-2011) for Meghalaya price data. The proposed models were ARIMA (1, 0, 1), ARIMA (1, 1, 1), ARIMA (0, 1, 1) (for Assam market price data series) and, log ARIMA (0, 1, 1), log ARIMA (1, 0, 1) with linear trend and a man-made intervention (Oct-2011) and log ARIMA (0, 1, 1) with linear trend and a man-made intervention (Oct-2011) (for Meghalaya market price data series) for minimum, maximum, and average monthly price series, respectively. Rethinam and

Sivaraman (2001) studied the growth in area, production and productivity of areca nut in Karnataka State with respect to price and non-price factors. He also estimated the costs and returns in the cultivation of areca nut and documented the constraints in production of areca nut in the study area. Surendra (1997) conducted a study on growth trends in area, production and productivity of Banana in Assam. Time series data on an area, production and productivity were collected for the period 1980-81 to 1999-2000 from Directorate of Economics and Statistics, Government of Assam and were analyzed using Linear, Quadratic and exponential growth models. The findings of the study revealed a significant growth rates for area (3.13%) and production (3.13%) and a non-significant growth rate for productivity. Jose and Jayashekar (2009) in his study on growth trends in area, production and productivity of areca nut in India used time series data on area, production and productivity of areca nut for a period of 30 years, employed compound growth rate and linear regression to analysis the data. The findings revealed that during last 15 years due to favorable price, area has increased by more than two times and production by three times.

3. Methodology

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group. The basic objective in cluster analysis is to discover natural groupings of the items (or variables). It is the most important unsupervised learning problem. It deals with finding structure in a collection of unlabeled data. In turn, we must first develop a quantitative scale on which to measure the association (similarity) between objects. Unsupervised learning is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data (a classification or categorization is not included in the observations). Similarity measures most efforts to produce a rather simple group structure from a complex data set require a measure of "closeness," or "similarity." There is often a great deal of subjectivity involved in the choice of a similarity measure. Important considerations include the nature of the variables (discrete, continuous, binary), scales of measurement (nominal, ordinal, interval, ratio), and subject matter knowledge. When items (units or cases) are clustered, proximity is usually indicated by some sort of distance. By contrast, variables are usually grouped on the basis of correlation coefficients or like measures of association (Johnson and Wichern).

In any market the rise and fall of price of a commodity is commonly observed. Suppose s_t is assumed to be a *n*-state, first order Markov process, taking the values 1………*n* with transition probability matrix.

$$P=\left\{\{p_{ij}\}\right\} \quad i, j = 1,2.......n;$$

where $p_{ij}=p_r\,[s_t = j|s_{t-1} = i]$ *with* $\sum_{i=1}^{n} p_{ij}=1$ *for all i …….(1)*

Based on equation (1), two state transition probabilities (symbolically 1 and 2) based on first order markov chain probabilities are defined. Here 1- denote rise and 2- denote fall.
Then the following probabilities are very appropriate to check the nature of price series.
p_{11} denotes the probability of price rise followed by price rise
p_{22} denotes the probability of price fall followed by price fall
p_{12} denotes the probability of price rise followed by price fall
p_{21} denotes the probability of price fall followed by price rise

Modeling Time series
The main purpose of building time series model is to know the past, to understand the present and forecast the future. Thus it is required to fit an appropriate model. We can build time series model with or without converting the original series into stationary.
Modeling Time Series with Trend (Non stationary series)
The observed time series may be generated by a regression type model, in which certain functions of time are taken as independent variables. For example, a series with trend is shown below:

$$x_t = m_t + E_t \quad …….. (2)$$

The function m_t may be linear or non-linear in time *t*.
Specifically , m_t may be a polynomial in *t* of the type:

$$m_t = \beta_0+\beta_1 t +\beta_2 t^2+\beta_3 t^3+………+\beta_p t^p \quad ……… (3)$$

To estimate beta coefficients in the above equation (3) one can use method of least square.

Method of Least Squares
The order *p* of the polynomial is a *model selection* problem. That is, we fit polynomials of degrees 1, 2,.. . . . for m_t and choose the one for which the least squares method provides a close fit for the data. Let the estimates of the coefficients be denoted by $\hat{\beta}_i$, *i* = 0, 1, 2, . . . *p*. The estimate of the trend function is then

$$m_t = \hat{\beta}_0+\hat{\beta}_1 t+\hat{\beta}_2 t^2+……….+\hat{\beta}_p t^p \quad ……… (4)$$

Now compute the residuals $\hat{E}_t=x_t-\hat{m}_t$ t=1, 2, ……. N and test for stationarity of { $\hat{E}_t$}.

Outline of the Approach:
- Select the first *nfit* observations as the Fitting/Training portion.
- Select the next *L=nfore* observations as the Holdout/Test portion.
- Fit MLR models to the Fitting portion of the data.
- Assess the fits. Compute in-sample model selection criteria.
- Use the fitted models to forecast the response time series for the Holdout portion, using coefficients from the fitted models and values of independent variables from the hold-out portion.
- Use forecast evaluation criteria as out-of-sample model selection criteria.

• Use the in-sample and out-of-sample criteria to select the best model(s).
• Use the best model(s) to forecast the future responses.

Fitting ARIMA class of models (Stationary series)

If the stochastic (or error) component provides a sequence of independent and identically distributed random variables, then we can analyze it by classical statistical tools. On the other hand, if the resulting error sequence is stationary, we can further analyze it by using the relevant tools in the area of discrete time stationary stochastic processes. Thus we will examine its temporal dependence structure and identify an appropriate linear time series model. The class of models is called Box-Jenkins models. The construction of these models consists following steps (Chatfield, 1996).

Step1: Plot the time profile and observe whether there are any time series components like trend and seasonality. One can use sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) to graphically access the time series component.

Step2: Remove time series components in order to make time series to be stationary. This can be achieved through making appropriate transformations like log, square root, inverse, or differencing the time series under study.

Step3: Select appropriate model and its order from ARIMA class model. This class contains MA, AR, ARMA and ARIMA models.

1. **Autoregressive Moving Average (ARMA) model:**

   Let $\{\varepsilon_t, t \geq 1\}$ is sequence of identically independent random variables with E(Ɛt)=0,V(Ɛt)= ϭ2. The time series {Xt, t>=1} is said to follow ARMA process of order(p,q) then it has the representation

   $$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + .. + \beta_p x_{t-p} + \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} .. - \alpha_q \varepsilon_{t-q}$$

   where $\beta_1 \ldots \beta_p$ are the parameters of AR process and $\alpha_1 \ldots \alpha_q$ are the parameters of MA process.

2. **Autoregressive Integrated moving average (ARIMA)process:**

   Let $\{\varepsilon_t, t\varepsilon I\}$ denotes the observed time series. Let this series becomes stationary after d differences denote Zt=$\nabla^d(x_t)$.Where ∇ =(1-B) and B is the backward shift operator .If Zt follows ARMA(p,q) then x_t is said to follow ARIMA(p,d,q). One can obtain initial values for p and q by observing the nature of ACF and PACF.

   - ACF of MA(q) process shows a clear cut of after lag q where as its PACF is exponentially decay to zero.
   - PACF of AR(p) process shows a clear cut of after lag p where as its ACF is exponentially decay to zero.
   - In ARMA(p,q) both ACF and PACF are exponentially decays to zero.

Step4: Estimation of parameter:

In literature several methods were discussed for parameter estimation such as,

- Method of lest square
- Yule-Walker method
- Maximum likelihood method

Step5: Diagnostic checking:

After parameter estimation, we have to assess model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the $\{\varepsilon_t\}$ is a white noise sequence. That is et's are uncorrelated random shocks with zero mean and constant variance. In classical time series set up it is common to assume that the white noise sequence $\{\varepsilon_t\}$ is iid Gaussian.

To check these assumptions on the residuals we have carried out Ljung and Box test (for uncorrelation) and Shapiro-Wilk test (for normality).The null hypothesis of Ljung–Box test may be defined as,

H_0: $\rho_1 = \rho_2 = \ldots\ldots\ldots\ldots = \rho_m = 0$

H_1: at least one $\rho_i \neq 0$

The test statistic is:

$$Q = n(n+2)\sum_{i=1}^{m} \frac{\hat{\rho}_i^2}{n-k}$$

where *n* is the sample size

To check the normality one can use Shapiro-Wilk test. The null hypothesis of Shapiro-Wilk test is defined as,

H_0: Residual follows normal

H_1: Residual do not follows normal

The test statistic is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Where $x_{(i)}$ is ith order statistic, i.e., the i^{th} smallest number in the sample, $\bar{x}$ is the sample mean and a_i is the Shapiro-Wilk constant.

If p value is less than the level of significance then we reject H_0 in both the test procedures.

Forecast Performance Measures

There are several forecast performance measures available in the literature to check the accuracy of the fitted model .The forecast performance measures used in this study is The Mean Absolute Error (MAE) and The Mean Percentage Error (MPE). Suppose y_t is the actual value, f_t is the forecasted value, then $e_t = y_t - f_t$ is the forecast error and *n* is the size of the test set [8]. Then the mean absolute error and mean absolute percentage error is defined as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{e_i}{y_i}\right|$$

4. Analysis and Discussions

The data set which we used for analysis is collected from "agmarknet.gov.in" site to the region Puttur, Karnataka. Our study period includes weekly data from January 2012 to December 2016. Price series is collected with respect to two varieties namely, "cqca" and "New variety". Statistical analysis of any data begins from a look into descriptive statistics. Table 1 represents the descriptive statistics for cqca and New variety

Table 1: Represents descriptive statistics for cqca and New variety

Variable	cqca	New variety
Maximum	15200	28000
Minimum	4500	7100
Mean	9548.859	16178.8
Median	9500	14500
Kurtosis	-0.4632	-0.8845
Skewness	-0.1396	0.3962

From above table it clearly states that Kurtosis of two varieties is negative, it indicates that price pattern of these two varieties are flatter than normal. Skewness of cqca variety is negative thus we can say that the distribution of weekly price pattern of cqca variety is negatively skewed and Skewness of New variety is positive thus we can say that the distribution of weekly price pattern of New variety is positively skewed.

A boxplot summarize the descriptive statistics graphically. The spacing between the different parts of the box helps indicate the degree of dispersion (spread) and skewness in the data, and identify outliers. Fig1 represent the box plot for two varieties.

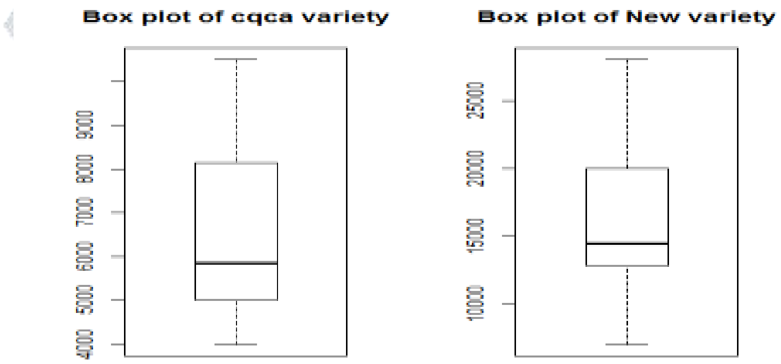

Fig 1: Represents Box plots for cqca and New variety

Above box plots clearly indicate that there is no outliers in data and "cqca" is positively skewed and "New variety" is negatively skewed.

Cluster Analysis

It is worth to know about similarity between weekly prices of a particular variety over the years. To check this cluster analysis is carried out for the price series of the year2016 by Ward's method of hierarchical clustering.

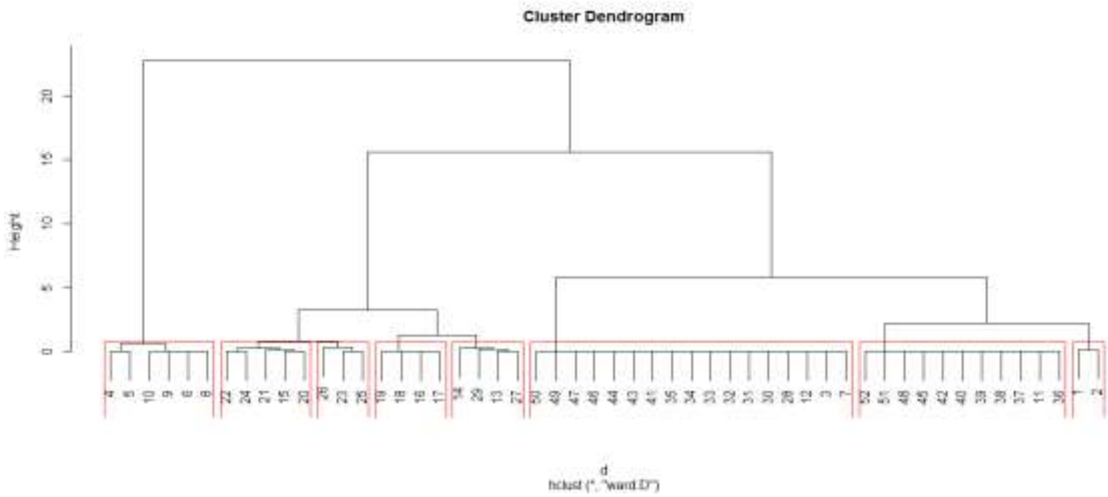

Fig 2: Represents Dendrogram of the price of the cqca variety in the year 2016

Table 2: Classification of weeks into different clusters

Number of cluster	Weeks in cluster
1	4,5,10,9,6,8
2	22,24,21,15,20
3	26,23,25
4	19,18,16,17
5	14,29,13,27
6	50,49,47,46,44,43,41,35,34,33,32,31,30,28,12,3,7
7	52,51,48,45,42,40,39,38,37,11,36
8	1,2

By looking at Fig 2 and Table 2 and we clearly says that week 1,2 are having similarity and forming smallest cluster and 50,49,47,46,44,43,41,35,34,33,32,31,30,28,12,3,7 weeks forming largest cluster.
Similarly, the dendrogram and classification table for new variety is presented below.

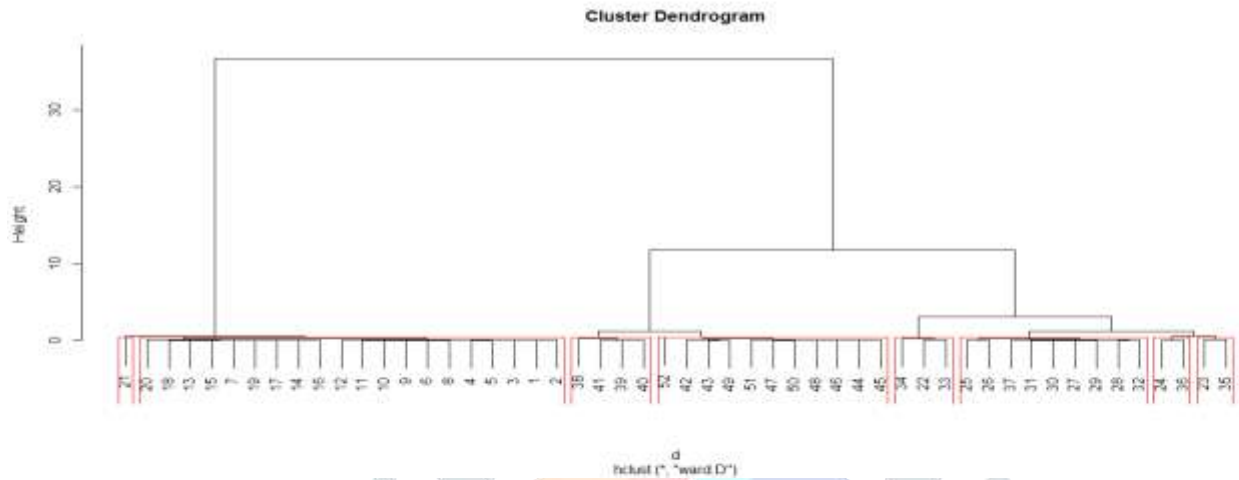

Fig 3: Represents Dendrogram of the price of the New variety in the year 2016

Table 3: Classification of weeks into different cluster

Number of cluster	Weeks in cluster
1	21
2	20,18,13,15,7,19,17,14,16,12,11,10 9,6,8,4,5,3,1,2
3	38,41,39,40
4	52,42,43,49,51,47,50,48,46,44,45
5	34,22,33
6	25,26,37,31,30,27,28,29,28,32
7	24,36
8	23,35

By looking at above Fig 3 and Table 3, we clearly says that week 21 forming separate cluster which is also forming as a smallest cluster and 20,18,13,15,7,19,17,14,16,12,11,10,9,6,8,4,5,3,1,2 weeks forming largest cluster.

Transition probabilities for price rise and fall:

In any market the rise and fall of price of a commodity is commonly observed. Then the following probabilities are very appropriate to check the nature of price series. Here 1- denote rise and 2- denote fall. Following Table 4 gives the transition probabilities matrix for price rise and fall for "cqca" variety

Table 4: Represents the transition probabilities matrix for price rise and fall for cqca variety.

	State1	State2
State1	0.9241	0.1785
State2	0.0759	0.8215

The probability that the state1 will be followed by another week of state1 is $p11$ = 0.9241and the state1 will persist, on average, for 9 weeks. However, the probability that the state2 will be followed by another week of state2 is $p22$ = 0.8215. This episode will typically persist for 8 weeks. Similarly, the probability that the state 1 will be followed by state2 phase is represented by *p12* = 0.1785 and the average persistence of this state is only 2 weeks. The probability that the state2 will be followed by another week of state1 *p21* = 0.0759 and the average duration of this state is 1 week.

Next, Table 5 gives the Transition probabilities matrix for price rise and fall for "New variety"

Table 5: Represents the Transition probabilities matrix for price rise and fall for New variety.

	State1	State2
State1	0.7517	0.2345
State2	0.2483	0.7655

The probability that the state1 will be followed by another week of state1 is *p11* = 0.7517 and the state1 will persist, on average, for 7 weeks. However, the probability that the state2 will be followed by another week of state2 is *p22 = 0.7655*. This episode will typically persist for 8 weeks. Similarly, the probability that the state 1 will be followed by state2 phase is represented by *p12 = 0.2483* and the average persistence of this state is only 2 weeks. The probability that the state2 will be followed by another week of state1 *p21 = 0.7655* and the average duration of this state is 8 weeks.

Fitting Time series model

The main purpose of building time series model is to know the past, to understand the present and forecast the future. Thus it is required to fit an appropriate model. We can build time series model with or without converting the original series into stationary. We use polynomial fitting method (Non-stationary data) and ARIMA (stationary data) class models. Time profile of price series of cqca and New variety is represented in following Fig 4.

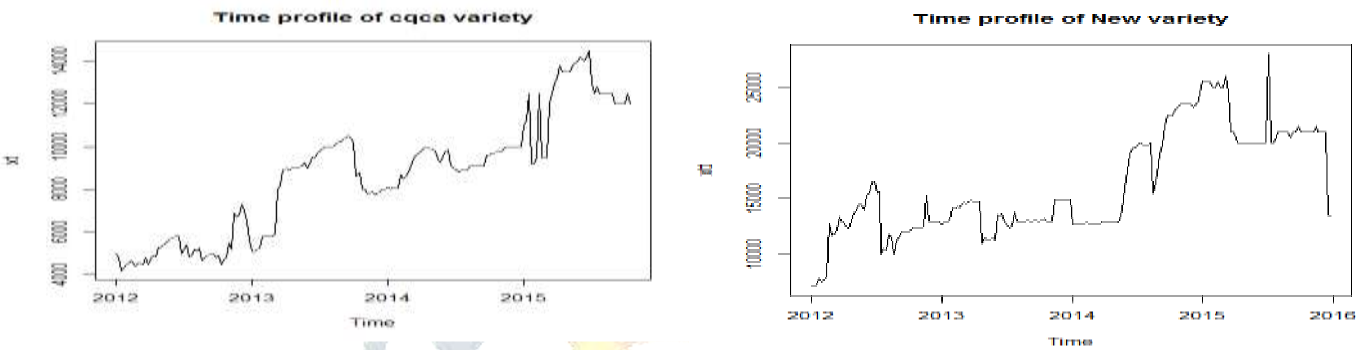

Fig 4: Represents Time profile of cqca and new variety

Next we plot the ACF of price series of cqca and New variety

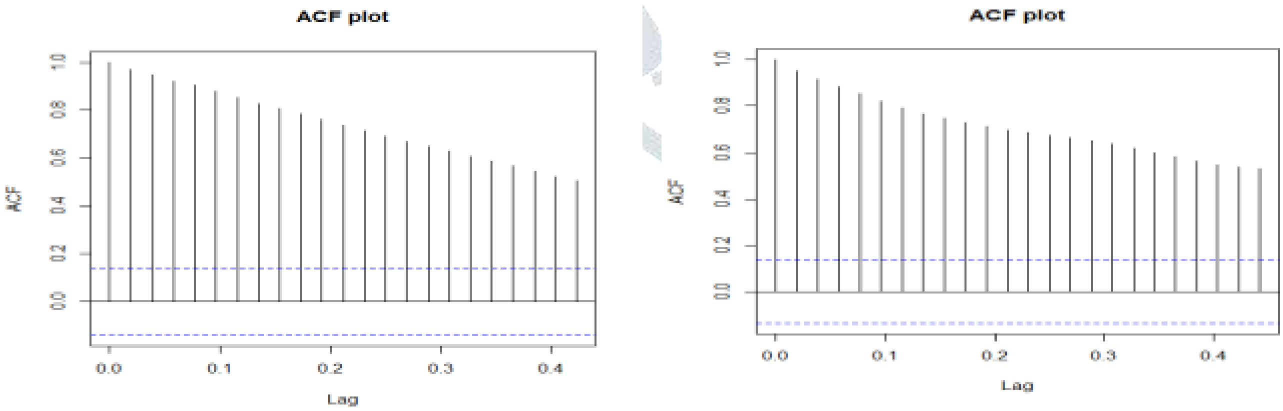

Fig 5: Represents ACF plots of cqca and new variety

From the Time profile and ACF plots it is clear that the trend component is present in the series under study. Theoretically to test for trend in a time series, we used Mann-Kendall (MK) test (Kendall and Ord, 1990). The computed value for Mann-Kendall test statistic is turn out to be 0.605 and 0.572 with very high significant p value (less than 2.22e-16) for the two series under study. Thus we can conclude that there is trend in the series.

Now, we divided the data under study into Training set and Testing set in order to give comparison to the forecast performance of the fitted models.

Polynomial fitting

The observed time series with trend can be represented as shown below

$x_t = m_t + E_t$ where $m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \ldots\ldots\ldots + \beta_p t^p$ ……. (5)

To estimate beta coefficients in the above equation, one can use method of least square. Here we fitted m_t using (5) up to 3rd degree. Following Fig 6 represent the fitted values with linear, quadratic and cubic polynomial for "cqca".

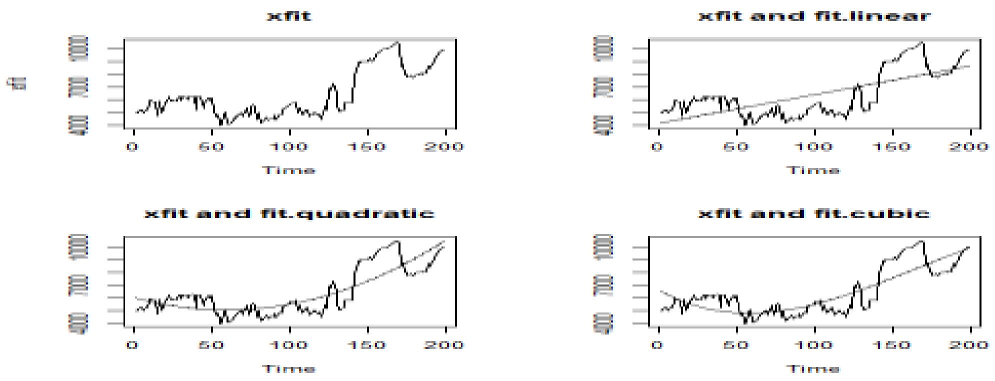

Fig 6: Represents fitted values with linear, quadratic and cubic polynomial for cqca variety

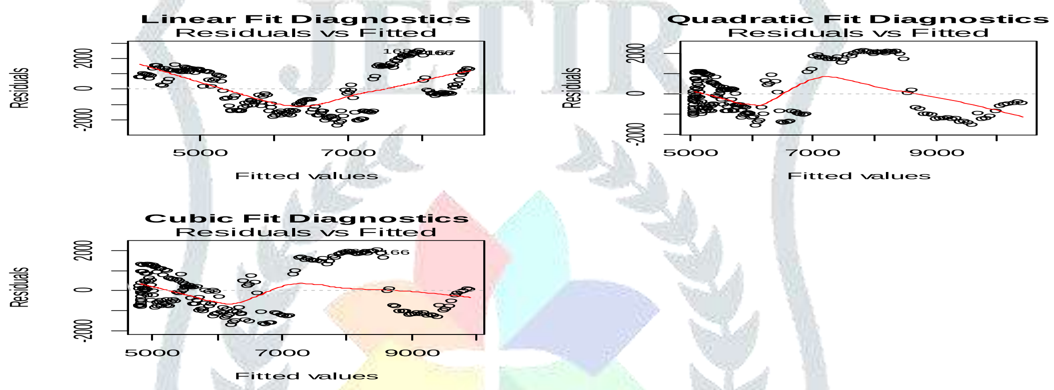

Fig 7: Represents the plot of Residual vs Fitted values of cqca variety

Table 6: Represents AIC values cqca variety

Order	AIC
1. Linear	15.40409
2.Quadratic	14.97976
3. Cubic	14.92581

By observing Residual vs Fitted plot and AIC values one can assess the goodness of fit. It can be observed that quadratic and cubic polynomials are competing models. However cubic polynomial has lower AIC value and hence it can be taken as a better model to fit trend in the price of "cqca" variety.

Now we try to fit a suitable polynomial to trend for the price series of "New variety". Fig 8 represents the fitted values with linear, quadratic and cubic polynomial for "New variety".

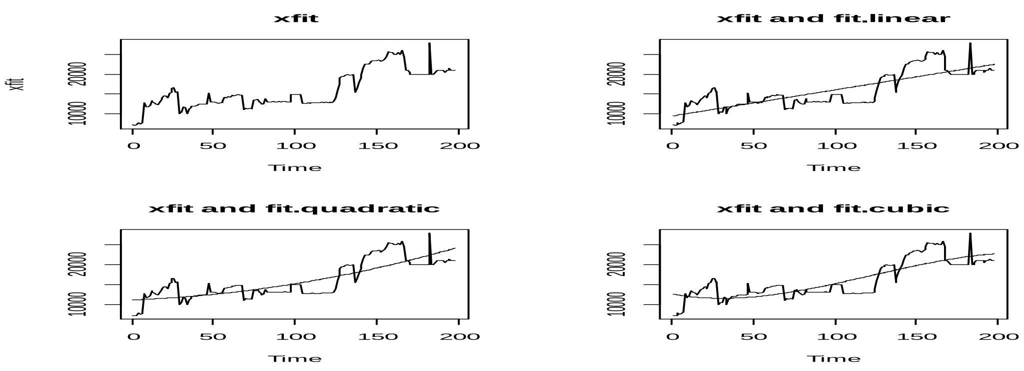

Fig 8: Represents fitted values with linear, quadratic and cubic polynomial for New variety

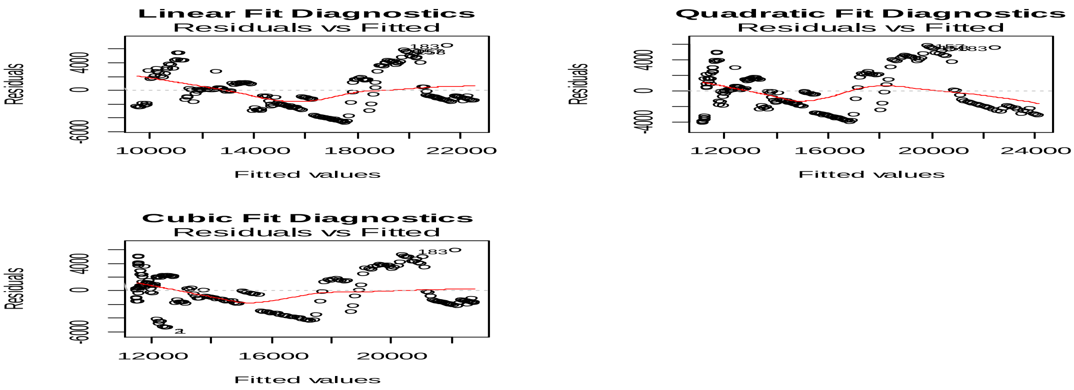

Fig 9: Represents the plot of Residual vs Fitted values for New variety

Table 7: Represents AIC values New variety".

Order	AIC
1. Linear	16.85786
2.Quadratic	16.79578
3. Cubic	16.76704

It can be observed that all the 3 fitted polynomials are competing models. However cubic polynomial has lower AIC value and hence it can be taken as a better model to fit trend in the price of "New variety". Thus for both varieties, one can use cubic polynomial to forecast the price series.

Fitting stationary model

In the time series analysis one of the common approach of fitting suitable models by converting the non-stationary series into stationary series. ARIMA class models are the best examples for this type of approach. So we converted time series under study into stationary by taking 1st difference. Using Augmented Dickey–Fuller (ADF) test, we confirm the stationarity of series. In order to identify the orders for parameters of ARIMA class model with respect to "cqca" variety, we plotted ACF and PACF plots as follows.

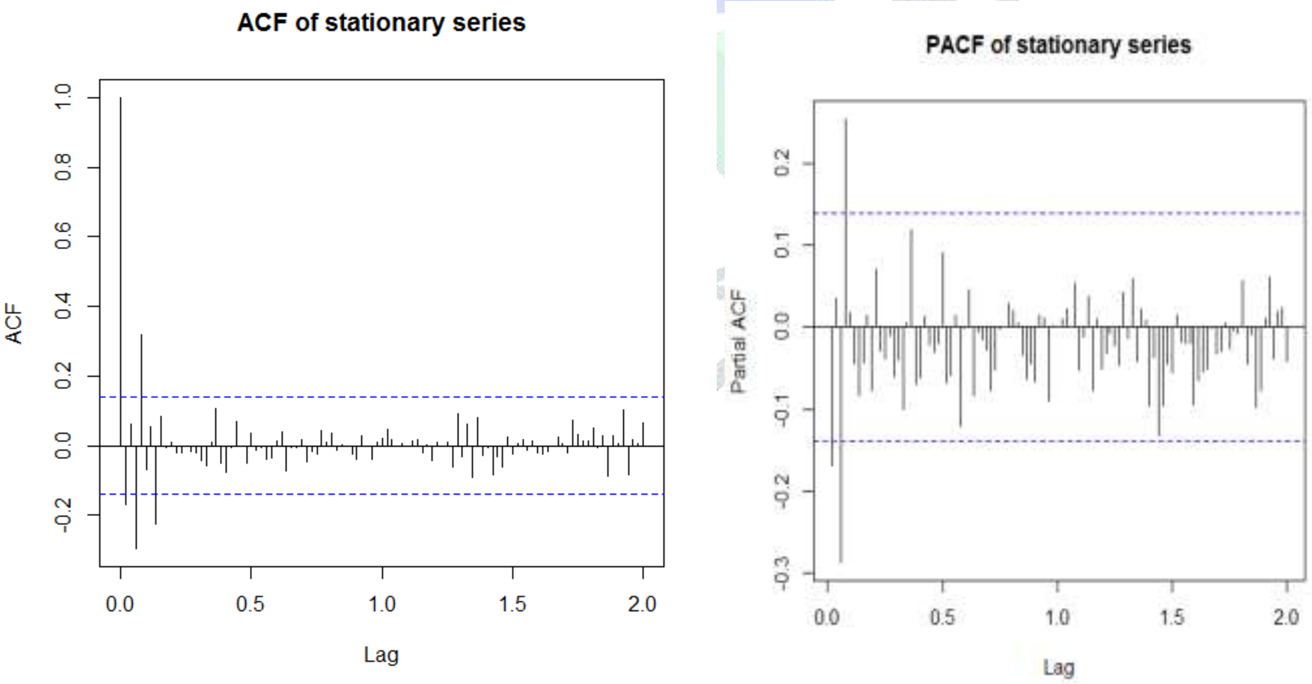

Fig 10: Represents ACF and PACF plots for stationary series of "cqca" variety

By observing ACF and PACF plots (Fig 10),we get the values p=3 and q=4 with d=1 as the initial guesses for p and q. Thus several combinations are possible. One can choose the best model by conducting analysis on residuals and by looking at model information criteria like AIC. The basic assumptions on residuals are

1) Residual series is uncorrelated
2) Residual series is normally distributed

Table8 represents some of the combinations with their assessing criteria's.

Table 8: Represents combinations with their assessing criteria's "cqca" variety.

Order	AIC	Wilks p value (for normality)	Ljung-Box p value(for un correlation)
1,1,0	3051.32	<2.2e-16	0.9411
2,1,0	3053.07	2.215e-15	0.8959
3,1,0	3038.08	2.215e-15	0.3074

By looking at Table 8, we can observe that only uncorrelation assumption on residuals is satisfied. Further the values of AIC for 3 different models were not differed much. Thus all the 3 models reported were competing models. Since ARIMA(1,1,0) has lower parameters, we decided that model is the best to predict price series of "cqca". Following Fig11 represents the summary of residual series. The estimates of ARIMA(1,1,0) model with respect to "cqca" variety are given below (Table 9).

Table 9: Represents the estimates of ARIMA(1,1,0) model with respect to "cqca" variety

	Estimate	SE	t.value	p.value
ar1	-0.1702	0.0702	-2.4233	0.0163
Constant	36.1030	33.5226	1.0770	0.2828

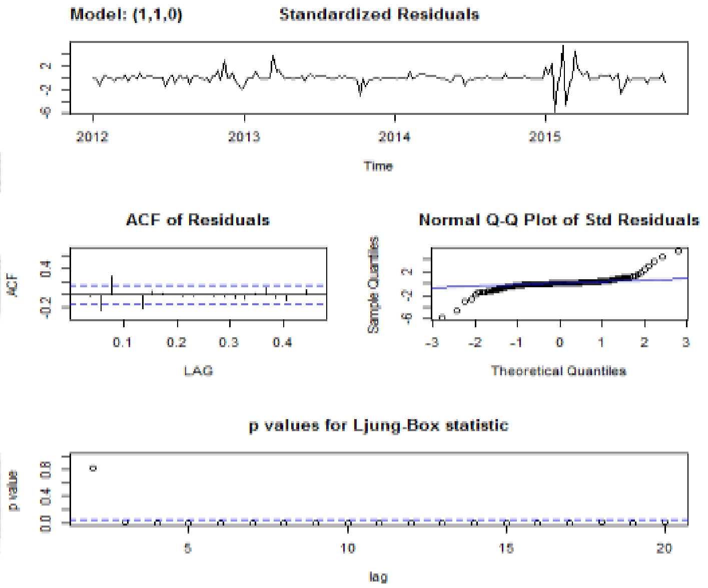

Fig 11: Represents the summary of residual series of cqca variety

Similarly, to identify the orders for parameters of ARIMA class model with respect to "New variety", we plotted ACF and PACF plots as follows.

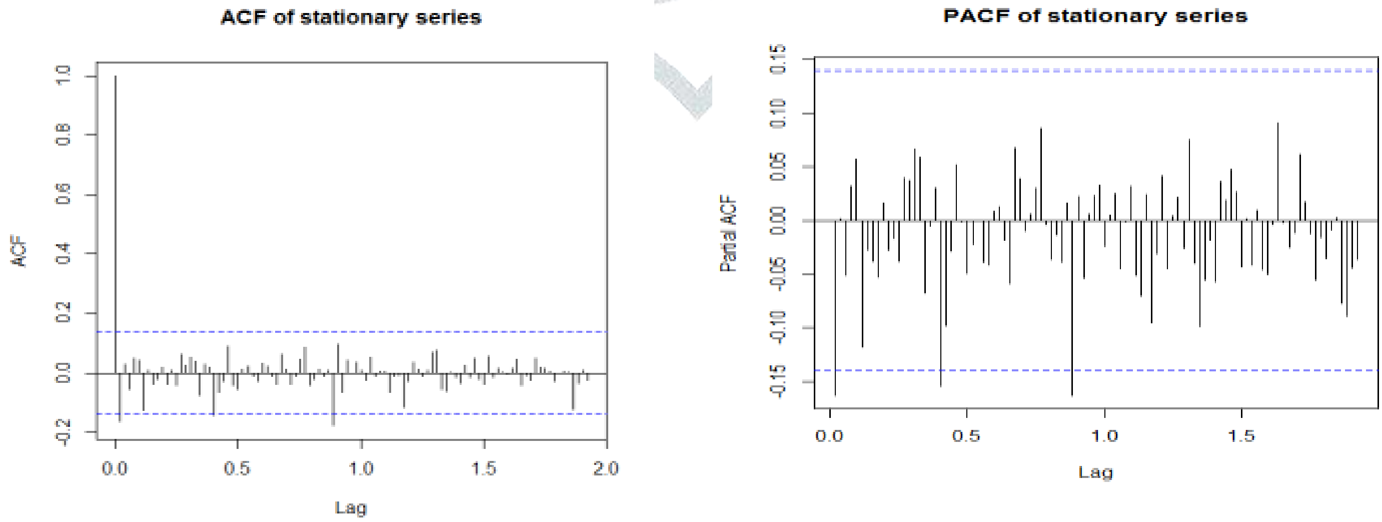

Fig 12: Represents ACF and PACF plots for stationary series of New variety

By observing ACF and PACF plots, we have obtained values for p=1and q=1 with d=1 as the initial guesses for p and q. Thus several combinations are possible. One can choose the best model by conducting analysis on residuals and by looking at model information criteria like AIC. Following table represents some of the combinations with their assessing criteria's.

Table 10: Represents combinations with their assessing criteria's.

Order	AIC	Wilks p value (for normality)	Ljung-Box p value (for uncorrelation)
0,1,1	3361.4	< 2.2e-16	0.9659
1,1,0	3363.42	< 2.2e-16	0.9879
1,1,0	3361.81	< 2.2e-16	0.8827

By looking at Table 10 we can observe that only uncorrelation assumption on residuals is satisfied. Further the values of AIC for 3 different models were not differed much. Thus all the 3 models reported were competing models. Since ARIMA(0,1,1) has lower parameters, we decided that model is the best to predict price series of "New variety". Following Fig 13 represents the summary of residual series of the fitted model.

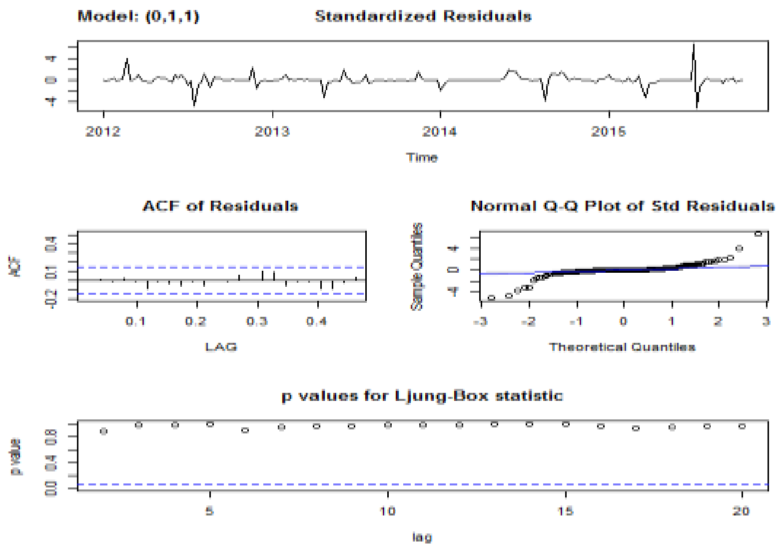

Fig 13: Represents summary of residual series of New variety

The estimates of ARIMA(0,1,1) model with respect to "New variety" are given below.

Table 11: Represents estimates of ARIMA(0,1,1) model with respect to New variety

	Estimate	SE	t.value	p.value
ma1	-0.2339	0.0687	-3.4046	0.0008
Constant	70.3479	66.0708	1.0647	0.2883

Forecast

By using fitted models (cubic polynomial and ARIMA(1,1,0)) for "cqca" variety, we forecasted weekly price for the period from Oct-25-2016 to Dec-31-2016. Further we compare the forecasted values with actual values (testing set).

Table 12: Represents comparison of the forecasted values with actual values (testing set) of cqca variety

Actual value	9950 9900 9800 9400 9300 9550 9800 9850 9150 9000
Forecasted price from Cubic Polynomial	9946.161 9990.233 10034.013 10077.491 10120.659 10163.510 10206.034 10248.223 10290.068 10331.562
Forecasted price from ARIMA (1,1,0)	12127.33 12147.91 12186.66 12222.31 12258.49 12294.58 12330.68 12366.79 12402.89 12438.99

Table 13: Represents comparison forecast accuracy measures of cqca variety

	Polynomial	ARIMA
MAE	571.5631	286.9602
MAPE	6.123471	2.339786

By comparing the forecast accuracy measures, MAE and MAPE, we found that ARIMA (1,1,0) has minimum value. Thus we can conclude that ARIMA (1,1,0) is better model to forecast the weekly price of "cqca" variety. Following Fig 14 represent forecasted values of price of "cqca" variety for the period from Oct-25-2016 to Dec-31-2016.

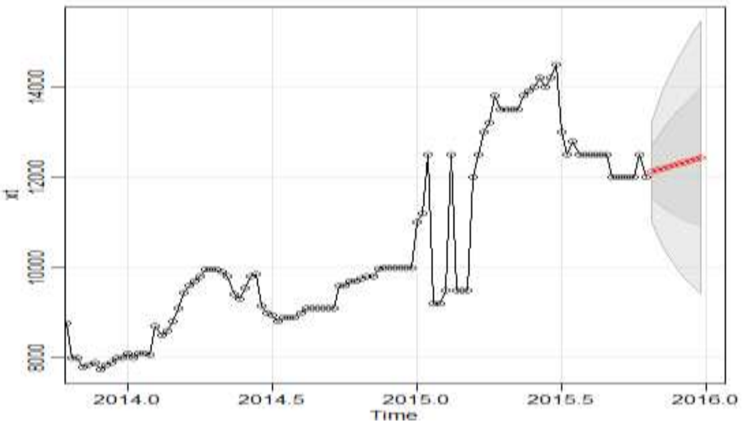

Fig 14: Represents Forecast graph of price cqca variety

Similarly by using fitted models (cubic polynomial and ARIMA(0,1,1,)) for "New variety", we forecasted weekly price for the period from Oct-25-2016 to Dec-31-2016. Further we compare the forecasted values with actual values (testing set)

Table 14: Represents comparison of the forecasted values with actual values (testing set) of New variety

Actual value	21000 21000 21000 21500 21000 21000 21000 21000 13500 13500
Polynomial forecasted value	22790.03 22817.02 22842.27 22865.76 22887.46 22907.35 22925.40 22941.61 22955.93 22968.36 10331.562
ARIMA forecasted value	21096.67 21167.02 21237.36 21307.71 21378.06 21448.41 21518.76 21589.10 21659.45 21729.80

Table15: Represents comparison forecast accuracy measures of New variety

	Polynomial	ARIMA
MAE	3340.119	1901.692
MAPE	20.89663	13.38933

By comparing the forecast accuracy measures, MAE and MAPE, we found that ARIMA (0,1,1,) has minimum value. Thus we can conclude that ARIMA (0,1,1) is better model to forecast the weekly price of "cqca" variety. Following Fig 15 represent forecasted values of price of "New variety" for the period from Oct-25-2016 to Dec-31-2016.

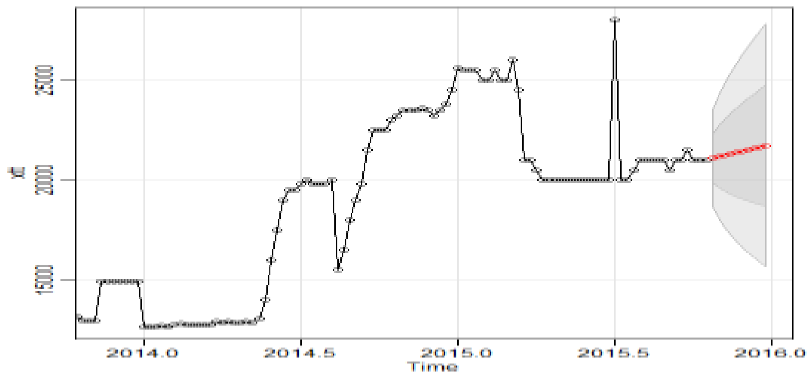

Fig 15: Represents Forecast graph of price New variety

5. CONCLUSION

- The data set which we used for analysis is collected from "agmarknet.gov.in" site to the region Puttur, Karnataka. This Study includes weekly data from January 2012 to December 2016. Price series is collected with respect to two varieties namely, "cqca" and "New variety".
- The coefficient of skewness calculated shown that the distribution of price pattern of "cqca" variety is negative skewed whereas distribution of price pattern of "New variety" positively skewed.
- Further from calculated coefficient of kurtosis, we can notice that, the distribution of price pattern of "cqca" variety and "New variety" are flatter than normal.
- Box plot clearly shows that there are no outliers in price data under study and it is negatively skewed with respect to "cqca" variety and positively skewed with respect to "New variety".
- In any market the rise and fall of price of a commodity is commonly observed. The probabilities in this matrix are very appropriate to check the nature of price series.

We observe that price fall followed by another price fall has highest transition probability 0.9241 related to "cqca" variety. Which means that on an average price fall followed by another price fall will continue upto 9 weeks. Whereas the corresponding probability related to "New variety" turned-out to be 0.7925. which means on on an average price fall followed by another price fall will continue upto 8 weeks.

- The largest cluster formed with respect to cqca variety has size=17 where as the largest cluster formed with respect to New variety has size=20.
- The smallest cluster formed with respect to cqca variety has size=2 whereas the smallest cluster formed with respect to New variety has size=1

By looking all these measures we can conclude that the price pattern of two varieties under study does not much differ.

One of the main objectives of this project is to build time series models to predict the future price and to check the accuracy of fitted model. Thus we divided the entire time series into two parts namely, Training set and Testing set. We adopted polynomial fitting method and ARIMA class models to forecast the price series. Further we used the forecast evaluation measures (MAE and MAPE) to check the forecast accuracy of fitted model. It is found that ARIMA model (ARIMA(1,1,0) for "cqca" and ARIMA(0,1,1) for "New variety") has minimum forecast errors. Thus one can use this model to predict the future price of Areca nut in Puttur.

References

[1] Chatfield, C., 1996, The Analysis of Time series An Introduction, Chapman and Hall.

[2] Jose, C. T. and Jayashekhar, S., 2009, Growth trends in area, production and productivity of Arecanut in India. *Agric. Situ. Ind*. **65 (3)**, pp.135-140.

[3] Kendall M.G. and Ord J.K.,1990,Time series, 3rd Ed., Edward Arnold

[4] Preeti, N. S., 1998, An economic analysis of arecanut cultivation and its impact on farm economy in Shivamogga district. *M. Sc. (Agri) Thesis*, Univ. Agric. Sci., Bangalooru, Karnataka (India).

[5] Rethinam, P. and Sivaraman, K., 2001, Arecanut (Areca catechu L) present status and future strategies. *Indian J. Arecanut Spice and Medicinal Pla*nts, **3(1)**, pp. 35-50.

[6] Shil, S., Acharya, G.C., Jose, C.T., Muralidharan, K., Sit, A.K.and George, V. T., 2013, Forecasting of arecanut market price in north eastern India: ARIMA modelling approach, Journal of Plantation Crops, **41(3)**, pp. 330-337.

[7] Surendra B. N., 1997, An analysis of arecanut production in Karnataka. *Agric. Situ. Ind*. **53(10)**, pp. 705-707.

[8] Johnson, R. A. and Wichern, D.W., 2007, Applied Multivariate Statistical Analysis, Prentice Hall, **U.S.A**.