PROJECT OF R-PROGRAMMING

PROJECT TOPIC:-

Production forecasting of Rice

SUBMITTED BY

Name -Amit kumar

Shanti Business School

Batch - PGDM-2015-17

Enrollment no - 1011517118

UNDER GUIDANCE OF:-

Prof. Amit Saraswat

ARIMA MODEL TO TIME SERIES DATA FOR FORECASTING USING R

ABSTRACT

Data on rice area and production for the period 1995-96 to 2014-15 were analyzed by time series method. Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) were calculated for the data. Appropriate Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) model was fitted. Validity of the model was tested using standard statistical techniques. ARIMA (0, 1, 1) and ARIMA (0, 1, 1) model were used to forecast area and production in Andhra Padesh, India for five leading years. The model also showed rice production forecast for the year 2019 to be about 11565.4 million tonnes with upper and lower limits 20979.01 and 2151.795 million tonnes respectively.

INTRODUCTION

There are several methods of forecasting. Among them is the exponential smoothing method which does not take the correlations between successive time series values into consideration. ARIMA model can make a better prediction since it takes into account the correlation between data. This aspect therefore assumes that the irregular component of time series can take non-zero autocorrelation. The ARIMA model is usually defined FOR STATIONARY time series. A stationary time series is the one whose statistical properties such as mean, variance, autocorrelation etc are all constant over time. If the time series is not stationary, then it has to be transformed to a stationary time series using any appropriate transformation technique, for instance, differencing, that is determine the differencing order d which makes it stationary. Differencing removes the trend component of a time series leaving the irregular component(Kasyoki, 2015).

A time series is a sequential set of data points, measured typically over successive times. It is mathematically defined as a set of vectors x (t),t = 0,1,2,... where t represents the time elapsed [21, 23, 31]. The variable) x (t is treated as a random variable. The measurements taken during an event in a time series are arranged in a proper chronological order. A time series is a collection of data recorded over a period of time, weekly, monthly, quarterly, or yearly. An analysis of history, a time series can be used by management to make current decisions and plans based on long-term forecasting. One usually assumes that past patterns will continue into the future. Long-term forecasts extend more than 1 year into the future; 5, 10, 15, and 20 year projections are common. Long-range predictions are essential to allow sufficient time for various departments to develop plans for future development(Agrawal, Adhikari, & Agrawal, 2013).

Forecasting is essential for decision making, unless insurance or hedging is selected to deal with the future (Armstrong, 1988). The growing importance of the forecasting function within companies is reflected in an increased level of commitment in terms of money, hiring of operational researchers and statisticians, and purchasing computer software (Wheelwright and Clarke, 1976; Pan et al., 1977; Fields and Hastings, 1994). Makridak is et al. (1983) note several factors which have caused the importance of forecasting within an organization to increase in recent years:-

- The increasing complexity of organizations (e.g. number of submarkets served and products offered) and their environments (e.g. changes in technology and demand structures) has made it more difficult for decision makers to take all the factors relating to the future development of the organization into account;
- organizations have moved towards more systematic decision making that involves explicit justifications for individual actions, and formalized forecasting is one way in which actions can be supported; and
- The further development of forecasting methods and their practical application has enabled not only forecasting experts but also managers (decision makers) to understand and use these techniques. With particular reference to the last point, it is

LITERATURE REVIEW

Multiple forecasts for autoregressive-integrated moving-average (ARIMA) models are useful in many areas such as economics and business forecasting. These methods were based on higher-order Bonferroni and product-type inequalities. Cheung, S. H., Wu, K. H. and Chan, W. S. (1998) found that the 'exact' method was computationally far more efficient. The 'exact' method is the evaluation of multivariate normal probabilities to the approximation method. Furthermore, the exact method can be applied to all ARIMA models while the approximation methods were limited to only a subset of ARIMA models(Khin et al., 2008).

ARIMA model

The autoregressive-integrated-moving average (ARIMA) model is discussed in detail in Box and Jenkins (1994) and O'Donovan (1983). Briefly, this technique is a univariate approach which is built on the premise that knowledge of past values of a time series is sufficient to make forecasts of the variable in question. There are two types of basic Box-Jenkins models: autoregressive (AR) models and moving-average (MA) models. The AR and MA models may also be combined to form ARMA models(Khin et al., 2008). These models are written as follows:

AR models: Xt = A1Xt-1 + ……………….ApXt-p + Et
Where Xt is directly related to one or more past series values.

MA models: Xt = - (B1Et-1 + ……………….BqEt-q) + Et
Where Xt is related to one or more past random errors.

ARMA models: Xt = (A1Xt-1 + ……………….ApXt-p) - (B1Et-1 + ……………BqEt-q) + Et
Where Xt is related to both past series values and past random errors.

The Ais are called autoregressive parameters and Bis are also called moving-average parameters. The subscripts on the A's and B's are called the orders of the parameters. In an AR model **p** is the order of the model, and in an MA model **q** is the order of the model. The order of an ARMA model is expressed in terms of both **p** and **q**. In terms of the original series such models are called integrated models and are denoted by Auto Regressive, Integrated, Moving-Average (ARIMA) models.

Box and Jenkins (1994) set four steps for this approach: model identification, parameter estimation, diagnostic checking and forecasting. The identification step involves the comparison of estimated autocorrelation and partial autocorrelation functions of known ARIMA processes. Given a class of ARIMA models from the first step, their parameter values can be estimated from the historical series using nonlinear least squares. Diagnostic checks are then applied to determine any possible inadequacies in the model, and the process is repeated if any found. Finally, having arrived at an adequate model, "optimal" forecasts are generated by recursive calculation.

In time series analysis literature (Box et al. 1994), an intervention event is an input series that indicates the presence or absence of an event. An intervention event causes a time series process to deviate from its expected evolutionary pattern. It is assumed that the intervention event occurs at a specific time, has a known duration, and is of a particular type. The time of the intervention is when the event begins to cause deviation. The duration of the intervention is how long the event causes deviation. The type of the intervention is how the event's influence changes over time. The intervention response is how the intervention causes deviation.

Description of Various Forecast Performance Measures

Now we shall discuss about the commonly used performance measures and their important properties. In each of the forthcoming definitions, t y is the actual value, t f is the forecasted value, $e_t = y_t - f_t$ is the forecast error and n is the size of the test set. Also, is the test mean and is the test variance(Winklhofer, Diamantopoulos, & Witt, 1996).

The Mean Forecast Error (MFE)

The properties of MFE are:

- It is a measure of the average deviation of forecasted values from actual ones.
- It shows the direction of error and thus also termed as the Forecast Bias.
- In MFE, the effects of positive and negative errors cancel out and there is no way to know their exact amount.
- A zero MFE does not mean that forecasts are perfect, i.e. contain no error; rather it only indicates that forecasts are on proper target.
- MFE does not panelize extreme errors.
- It depends on the scale of measurement and also affected by data transformations.
- For a good forecast, i.e. to have a minimum bias, it is desirable that the MFE is as close to zero as possible.

The Mean Absolute Error (MAE)

Important features are:

- This measure represents the percentage of average absolute error occurred.
- It is independent of the scale of measurement, but affected by data transformation.
- It does not show the direction of error.
- MAPE does not panelize extreme deviations.
- In this measure, opposite signed errors do not offset each other.

The Mean Absolute Percentage Error (MAPE)

- This measure represents the percentage of average absolute error occurred.
- It is independent of the scale of measurement, but affected by data transformation.
- It does not show the direction of error.
- MAPE does not panelize extreme deviations.
- In this measure, opposite signed errors do not offset each other.

The Mean Percentage Error (MPE)

- MPE represents the percentage of average error occurred, while forecasting.
- It has similar properties as MAPE, except,
- It shows the direction of error occurred.
- Opposite signed errors affect each other and cancel out.
- Thus like MFE, by obtaining a value of MPE close to zero, we cannot conclude that the corresponding model performed very well.
- It is desirable that for a good forecast the obtained MPE should be small.

The Mean Squared Error (MSE)

- It is a measure of average squared deviation of forecasted values.
- As here the opposite signed errors do not offset one another, MSE gives an overall idea of the error occurred during forecasting.
- It panelizes extreme errors occurred while forecasting.
- MSE emphasizes the fact that the total forecast error is in fact much affected by large individual errors, i.e. large errors are much expensive than small errors.
- MSE does not provide any idea about the direction of overall error.
- MSE is sensitive to the change of scale and data transformations.
- Although MSE is a good measure of overall forecast error, but it is not as intuitive and easily interpretable as the other measures discussed before.

The Sum of Squared Error (SSE)

- It measures the total squared deviation of forecasted observations, from the actual values.
- The properties of SSE are same as those of MSE.

The Signed Mean Squared Error (SMSE)

- It is same as MSE, except that here the original sign is kept for each individual squared error.
- SMSE panelizes extreme errors, occurred while forecasting.
- Unlike MSE, SMSE also shows the direction of the overall error.
- In calculation of SMSE, positive and negative errors offset each other.
- Like MSE, SMSE is also sensitive to the change of scale and data transformations.

The Root Mean Squared Error (RMSE)

- RMSE is nothing but the square root of calculated MSE.
- All the properties of MSE hold for RMSE as well.

The Normalized Mean Squared Error (NMSE)

- NMSE normalizes the obtained MSE after dividing it by the test variance.
- It is a balanced error measure and is very effective in judging forecast accuracy of a model.
- The smaller the NMSE value, the better forecast.
- Other properties of NMSE are same as those of MSE.

The Theil's U-statistics

- ❖ It is a normalized measure of total forecast error.
- ❖ $0 \leq U \leq 1$;$U = 0$ means a perfect fit.
- ❖ This measure is affected by change of scale and data transformations.
- ❖ For assessing good forecast accuracy, it is desirable that the U-statistic is close to zero.

We have discussed ten important measures for judging forecast accuracy of a fitted model. Each of these measures has some unique properties, different from others. In experiments, it is better to consider more than one performance criteria. This will help to obtain a reasonable knowledge about the amount, magnitude and direction of overall forecast error. For this reason, time series analysts usually use more than one measure for judgment.

Quantitative forecasting models can be grouped into two categories: the time series models and causal methods. Time series analysis tries to determine a model that explains the historical demand data and allows extrapolation into the future to provide a forecast in the belief that the demand data represent experience that is repeated over time.

This category includes naïve method, moving average, trend curve analysis, exponential smoothing, and the autoregressive integrated moving averages (ARIMA) models. These techniques are appropriate when we can describe the general patterns or tendencies, without regard to the factors affecting the variable to be forecast. Easy to develop and implement, times series models are preferred for their having been used in many applications such as: Economic Forecasting, Sales Forecasting, Budgetary Analysis, Stock Market Analysis, Process and Quality Control and Inventory Studies. Although in many forecasting applications the ARIMA model has been successfully used to predict seasonal time series, it suffers from limitation because of its linear form. However, due to its linearity, ARIMA is not always suitable for complex real-world problems.

STATISTICAL TECHNIQUES FOR TIME SERIES FORECASTING

Various organizations/ employees in India and abroad have done modeling using supported time series data exploitation. The various methodologies viz. statistic decomposition models, Exponential smoothing models, ARIMA models and their variations like seasonal ARIMA models, vector ARIMA models using variable time series, ARMAX models i.e. ARIMA with instructive variables etc. has been used. Many studies have taken place within the analysis of pattern and distribution of rainfall in numerous regions of the globe. Totally different time series methods with different objectives are used to investigate rain information in numerous literatures.

Stringer (1972) reported that a minimum of thirty five quasi-periods with over one year long are discovered in records of pressure, temperature, precipitation, and extreme climatic conditions over the earth the earth of the world surface. A very common quasi-periodic oscillation is the quasi-biennial oscillation (QBO), during which the environmental condition events recur each two to two.5 years. Win Stanley (1973a, b) reported that monsoon rains from Africa to India decreased by over five hundredth from 1957 to 1970 and expected that the long run monsoon seasonal rain, averaged over five to ten years is probably going to decrease to a minimum around 2030. Laban

(1986) uses time series supported ARIMA and Spectral Analysis of areal annual rain of two same regions in East Africa and counseled ARMA(3,1) because the best appropriate region indices of relative wetness/dryness and dominant quasi-periodic fluctuation around 2.2-2.8 years,3-3.7 years,5-6 years and 10-13 years.

Fig. shows a framework for organizational forecasting practice developed by integrating the (largely complementary) perspectives of Levenbach and Cleary (1981, 1982, and 1984) and Armstrong et al. (1987). The framework distinguishes between three different sets of issues, relating to design, selection/specification and evaluation. Design issues comprise the purpose and type of forecast required, the resources committed to forecasting, the characteristics of forecast preparers and users and the data sources used. Selection specification issues are concerned with forecasting techniques and address questions of familiarity with and selection and usage of alternative forecasting methods. Finally, evaluation issues focus on the outcomes of forecasting activity as reflected in the presentation and review of forecasts, the evaluation of forecast performance and the forces adversely affecting forecast accuracy. It should be noted that, as indicated by the two-way arrows, the three sets of issues are interlinked in that each can have implications for the others; for example, the adoption of a particular forecasting technique (a specification issue) will have implications for forecast accuracy (an evaluation issue) which, in turn, may lead to adjustments in, say, the data inputs used to develop the forecast (a design issue).provides a logical and orderly representation of organizational forecasting practice and establishes a clear overview of the latter; thus, it is a convenient 'navigation guide' for discussing the diverse findings of the empirical studies summarized earlier in Table 1 and for presenting future research directions(Winklhofer et al., 1996).

Design issues

❖ Purpose/use of forecast

Several empirical studies focused on why businesses produce forecasts and the use they make of the latter. In White's (1986) survey, 64% of respondents regarded the purpose of a sales forecast as a goal setting device-a statement of desired performance; only 30% wanted to derive a true assessment of the market potential. This finding was independent of firm size. However, smaller firms used sales forecasts more often for personnel planning while for larger firms sales quota setting and purchasing planning were frequent uses.

❖ Forecast level

The level for which forecasts are prepared (e.g. product item vs company forecast) seems to have been neglected, as only a handful of studies dealt with this issue. Small's (1980) investigation showed that the majority of firms produced sales estimates for more than one level of product/ market detail and also found a relationship between technique usage and forecast level.More specifically, firms using judgemental forecasting techniques (e.g. jury of executive opinion and sales force composite) were more likely to produce forecasts for geographic market areas, while firms employing regression and time series analysis were more likely to generate forecasts on the basis of product line/class of service. A further study of retailing firms by Peterson (1993) found that the level of forecast preparation could also be related to the size of the company; larger retailers were more inclined to develop industry forecasts and estimates by customer type and geographic area than smaller retailers, although both types of firms developed company and product forecasts.

❖ Time horizon and frequency of forecast Preparation

Several empirical investigations examined how far into the future firms prepared forecasts and the frequency of forecast preparation. In Pan etal.'s (1977) study the most popular short-term sales forecasting horizon was one month, while sales forecasts prepared for 1 year and 5 years ahead reflected typical long-term sales forecasting horizons. Cerullo and Avila (1975), White (1986) and Peterson (1990) also found that the Majority of firms prepared sales forecasts on a yearly basis. Naylor (1981) reported that firms which used econometric models developed forecasts for up to 7.7 years ahead on average; some companies employed them to generate forecasts as far as 25 years ahead. Investigating the time horizon of a firm's forecasts in relation to company characteristics, Small (1980, p. 21) discovered that "factors such as the industry of a firm, its market orientation and the forecasting role in which a technique is used have a significant impact on the time horizon over which a technique is used to forecast sales". In this context, McHugh and Sparkes (1983) found that firms operating in highly competitive markets put more emphasis on short-term than on long-term forecasts, and that subsidiaries prepared forecasts more frequently than independent firms.

❖ Forecast preparers

A large body of literature has focused on the organizational structure of forecasting, and the background and knowledge of the individuals involved in forecast preparation.

❖ Resources committed to forecasting

• Money and personnel

Cerullo and Avila's (1975) results indicated a weak commitment among Fortune 500 firms towards sales forecasting, as the forecasting function was generally not formally organized within these companies; two-thirds of respondents had no full-time employees working in sales forecasting, 24% had less than five employees and only 9% had more than five employees. In contrast, Wheelwright and Clarke (1976) (who surveyed large US firms with high involvement and concern for forecasting) observed a significant level of financial commitment (e.g. the resources typically committed by a company with annual sales up to $500 million comprised one or more specialist forecasting staff and between $10 000 and $50000 forecasting budget).

• Computers

Some 20 years ago, the majority (86%) of UK companies who utilised econometric forecasting models already worked with computers (Simister and Turner, 1973).

❖ Forecast users

In contrast to the large amount of empirical research on forecast preparers, relatively little is known about forecast users. However, findings on forecast purpose/use provide at least some insight, since they highlight the functional areas (e.g. production and marketing) in which the forecasts are applied. It should also be borne in mind that forecast preparers may also themselves be the principal forecast users as shown, for example, in the case study by Fields and Hastings (1994). Peterson (1993) found that top management, marketing, finance and accounting executives were the major users of forecasts, while five key user groups were identified in the study of Rothe (1978): production planning and operations management, sales and marketing management, finance and accounting, top corporate management and personnel.

❖ Data sources

A number of studies have concentrated on the external and internal information sources utilized for forecast preparers in companies, with only half of Mentzer and Cox's (1984a) sample having received formal training (see also Cerullo and Avila, 1975). On the other hand, Davidson (1987, p. 19) reported that his sample of forecasters regarded college courses in "quantitative methods, computer literacy, production/management, statistics, forecasting and market research" as "most important". Surveys that focused on forecasting courses offered at universities (Hanke, 1984, 1989; Kress, 1988; Hanke and Weigand, 1994) agreed that business schools emphasise different techniques (more quantitative than qualitative) than those commonly used in the business world. Furthermore, training in data collection, monitoring and evaluation of forecasts seemed to be rather neglected and a cause for concern given that "the intended audience of most forecasting courses is future managers/decision makers and not forecasters" (Kress, 1988, p. 28), and managers/decision makers are likely to spend a considerable time on such activities (Fildes and Hastings, 1994). Makridakis et al. (1983, pp. 805-806) stated that lack of formal training is often overrated, in that "the emphasis should not be on increasing the forecaster's knowledge of sophisticated methods, since doing so does not necessarily lead to improved performance. Perhaps the training should consider such issues as how to select a time horizon, how to choose the length of a time period, how judgment can be incorporated into a quantitative forecast, how large changes in the environment can be monitored, and the level of aggregation to be forecast".

Selection/specification issues

❖ Familiarity with forecasting techniques

A large body of literature (e.g. Wheelwright and Clarke, 1976; Fildes and Lusk, 1984; Mentzer and Cox, 1984a; Sparkes and McHugh, 1984; Sanders, 1992; Fildes and Hastings, 1994; Sanders and Manrodt, 1994) focused on companies' familiarity with forecasting techniques. However, the above studies vary with regard to the number and types of forecasting methods examined and the populations surveyed (e.g. analysts' vs senior managers).

❖ Criteria for technique selection

Wheelwright and Clarke (1976) identified four factors (cost, user's technical ability, problem specific characteristics and desired forecasting method characteristics) which were equally important for the selection of a forecasting technique in their sample. In Mahmoud et al.'s (1988) study, on the other hand, respondents rated accuracy as the most important criterion, followed by ease of use, data requirements, time horizon, and data pattern, number of items to be forecast and availability of software.

❖ Usage of alternative forecasting methods

Drury (1990), who compared surveys from Canada (Small, 1980; Drury, 1986), the UK (Simister and Turner, 1973; Sparkes and McHugh, 1984) and the USA (Dalrymple, 1975; Mentzer and Cox, 1984a), observed that in the late 1970s and early 1980s there were no differences across the three countries regarding the use of subjective and objective forecasting techniques. Surveys carried out in the mid 1980s (Mentzer and Cox, 1984a; Sparkes and McHugh, 1984; Drury, 1986) revealed an increase in the adoption of objective techniques in all three countries, but with different adoption rates in these countries (the UK having the highest rate and the US the lowest).

Evaluation issues

❖ Forecast presentation to management

Miller's (1985) case study emphasised how crucial the employment of graphical output was when presenting forecasts to decision makers. Graphical displays were "[t]he most important step toward the overall satisfaction of needs" of the organisation (Miller, 1985, p. 74). Surprisingly, none of the other empirical investigations raised the issue of how forecasts were presented to management.

❖ Forecast review and use of subjective judgement

Davidson (1987) examined whether interdepartmental meetings were used as a means of reviewing forecasts. Half of his respondent companies held such meetings, but some firms did not invite the forecast preparers! Peterson (1989, 1990) investigated whether different parties were involved in the forecast review process for different forecasting techniques, and found that the kinds of individuals involved in reviewing forecasts depended both upon the type of firm and the forecasting technique. For example, in terms of sales force forecasts, finance managers were highly involved in reviewing in consumer goods firms but to a lesser extent in industrial goods firms. The converse was true regarding the revision of expert opinion forecasts, i.e. finance managers were more involved in industrial goods firms than in consumer goods firms.

❖ Standards for forecast evaluation

Studies investigating the criteria used for evaluating forecasts agreed that accuracy was the most important factor, followed by ease of use, ease of interpretation, credibility and cost (Carbone and Armstrong, 1982; Mentzer and Cox, 1984a; Martin and Witt, 1988). Inaccurate forecasts led, in Sanders' (1992) sample, to inventory/ production and scheduling problems, wrong pricing decisions, customer service failures, etc. However, the accuracy aspect seemed to be more important for academics than for practitioners, the latter putting more emphasis on ease of interpretation, cost and time (Carbone and Armstrong, 1982). Speed, or timeliness of a forecast, also tended to be an important evaluation criterion for industrial goods producers but not for consumer goods producers in Herbig et al.'s (1994) sample (see also Small, 1980). Lastly, Martin and Witt (1988) reported that, with the extension of the forecasting horizon, the speed with which the forecast became available lost importance for their respondents.

❖ Forecast performance

From the empirical evidence, it is clear that maintaining records of forecast accuracy on a regular basis is not a universal practice. While Dalrymple (1975, 1987) reported that four out of five firms kept such records, in Rothe's (1978) sample some 40% did not have an objective accuracy figure. Similarly, more than a third of Drury's (1990) respondents did not have systems and procedures for analyzing forecast errors. One explanation for this may well be the lack of a mechanism for keeping records on the various types of forecasts, as shown in the case study company by Capon et al. (1975); this, in itself, may cause forecasts to be perceived as inadequate (Winklhofer et al., 1996).

OBJECTIVES

To forecast the Rice production of Andhra Pradesh year wise. To show the use of ARIMA models in forecasting.

CODES

```
Food<-read.csv ("C:/Users/WELCOME/Desktop/time.csv")
Foodtimeseries <- ts(Food, frequency=1, start=c(1995), end = c(2014))
View(Foodtimeseries)
plot.ts(Foodtimeseries)
class(Food)
frequency(Foodtimeseries)
summary(Foodtimeseries)

acf(Foodtimeseriesdiff1, lag.max=20) # plot a correlogram
acf(Foodtimeseriesdiff1, lag.max=20, plot=FALSE) # get the autocorrelation values

pacf(Foodtimeseriesdiff1, lag.max=20) # plot a partial correlogram
pacf(Foodtimeseriesdiff1, lag.max=20, plot=FALSE)

Foodtimeseriesarima <- arima(Foodtimeseries, order=c(1,0,0)) # fit an ARIMA(1,0,1) model
Foodtimeseriesarima
Foodtimeseriesarima <- arima(Foodtimeseries, order=c(2,0,0)) # fit an ARIMA(2,0,0) model
Foodtimeseriesarima
Foodtimeseriesarima <- arima(Foodtimeseries, order=c(0,0,1)) # fit an ARIMA(0,0,1) model
Foodtimeseriesarima
Foodtimeseriesarima <- arima(Foodtimeseries, order=c(1,0,3)) # fit an ARIMA(1,0,3) model
Foodtimeseriesarima
Foodtimeseriesarima <- arima(Foodtimeseries, order=c(1,0,1)) # fit an ARIMA(1,0,1) model
Foodtimeseriesarima
Foodtimeseriesarima <- arima(Foodtimeseries, order=c(2,0,3)) # fit an ARIMA(2,0,3) model
```

```
Foodtimeseriesarima
Foodtimeseriesarima <- arima(Foodtimeseries, order=c(0,1,0)) # fit an ARIMA(0,1,1) model
Foodtimeseriesarima
library(timeDate)
library(forecast)
Foodtimeseriesforecasts <- forecast.Arima(Foodtimeseriesarima, h=5)
Foodtimeseriesforecasts
plot.forecast(Foodtimeseriesforecasts)

acf(Foodtimeseriesforecasts$residuals, lag.max=10)
Box.test(Foodtimeseriesforecasts$residuals, lag=10, type="Ljung-Box")

plot.ts(Foodtimeseriesforecasts$residuals) # make time plot of forecast errors
hist(Foodtimeseriesforecasts$residuals) # make a histogram
```

RESULTS AND DISCUSSION

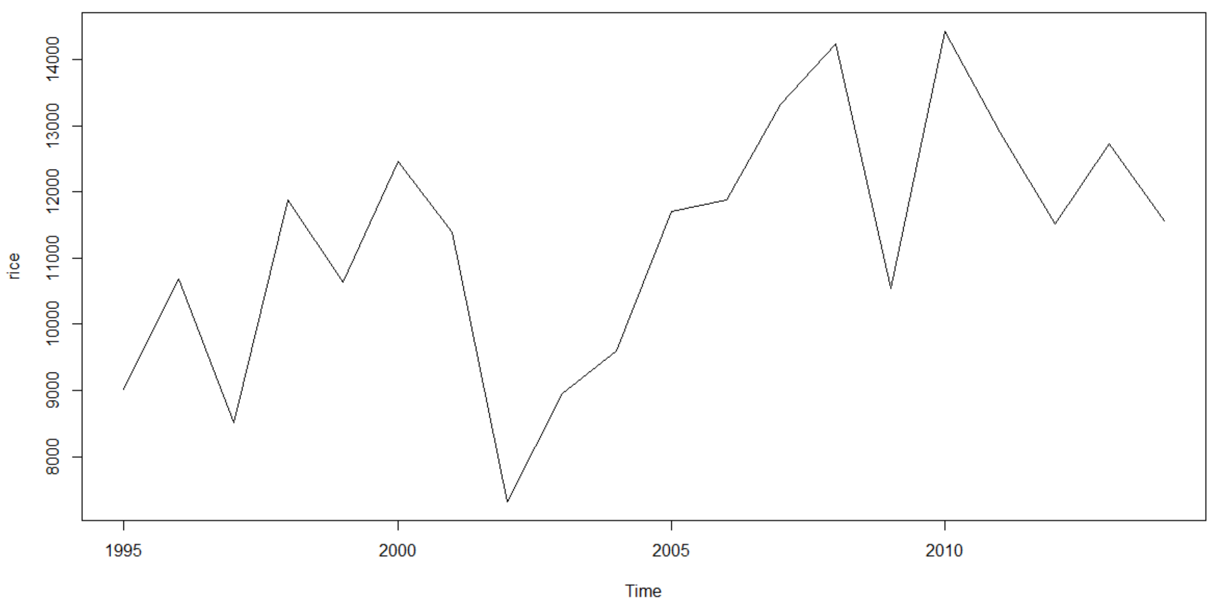
In the present study the data for rice cultivated production for the period 1995-96 to 2014-15 were used following the four stages of ARIMA model.

Plotting Time Series

Once you have read a time series into R, the next step is usually to make a plot of the time series data, which you can do with the plot.ts() function in R.

For example, to plot the time series of the production of rice of 20 successive rice production in A ndhra Pradesh, we type:

```
> Food<-read.csv ("C:/Users/WELCOME/Desktop/time.csv")
> Foodtimeseries <- ts(Food, frequency=1, start=c(1995), end = c(2014))
> View(Foodtimeseries)
> plot.ts(Foodtimeseries)
```



We can see from this time series that there seems to be seasonal variation in the Million tonnes Rice production per year: there is a peak every 2 to 3 year. Again, it seems that this time series could probably be described using an additive model, as the year wise fluctuations are roughly constant in size over time and do not seem to depend on the level of the time series, and the random fluctuations also seem to be roughly constant in size over time.

```
> Foodtimeseries
Time Series:
Start = 1995
End = 2014
Frequency = 1
         rice
 [1,]  9014.2
 [2,] 10686.0
 [3,]  8510.0
 [4,] 11878.0
 [5,] 10637.8
 [6,] 12458.0
 [7,] 11389.8
 [8,]  7327.0
 [9,]  8953.0
[10,]  9601.0
[11,] 11704.0
[12,] 11872.0
[13,] 13324.0
[14,] 14241.0
[15,] 10538.0
[16,] 14418.0
[17,] 12895.0
[18,] 11510.0
[19,] 12724.7
[20,] 11565.4


> class(Food)
[1] "data.frame"
> frequency(Foodtimeseries)
[1] 1
> summary(Foodtimeseries)
      rice
 Min.   : 7327
 1st Qu.:10304
 Median :11538
 Mean   :11262
 3rd Qu.:12525
 Max.   :14418
```

ARIMA MODEL

If your time series is stationary, or if you have transformed it to a stationary time series by differencing d times, the next step is to select the appropriate ARIMA model, which means finding the values of most appropriate values of p and q for an ARIMA (p, d, q) model. To do this, you usually need to examine the correlogram and partial correlogram of the stationary time series.

To plot a correlogram and partial correlogram, we can use the "acf()" and "pacf()" functions in R, respectively. To get the actual values of the autocorrelations and partial autocorrelations, we set "plot=FALSE" in the "acf()" and "pacf()" functions.

Plot ACF and PACF to identify potential AR and MA model

Now, let us create autocorrelation factor (ACF) and partial autocorrelation factor (PACF) plots to identify patterns in the above data which is stationary on both mean and variance. The idea is to identify presence of AR and MA components in the residuals. The following is the R code to produce ACF and PACF plots.

Example of the Rice production of Andhra Pradesh

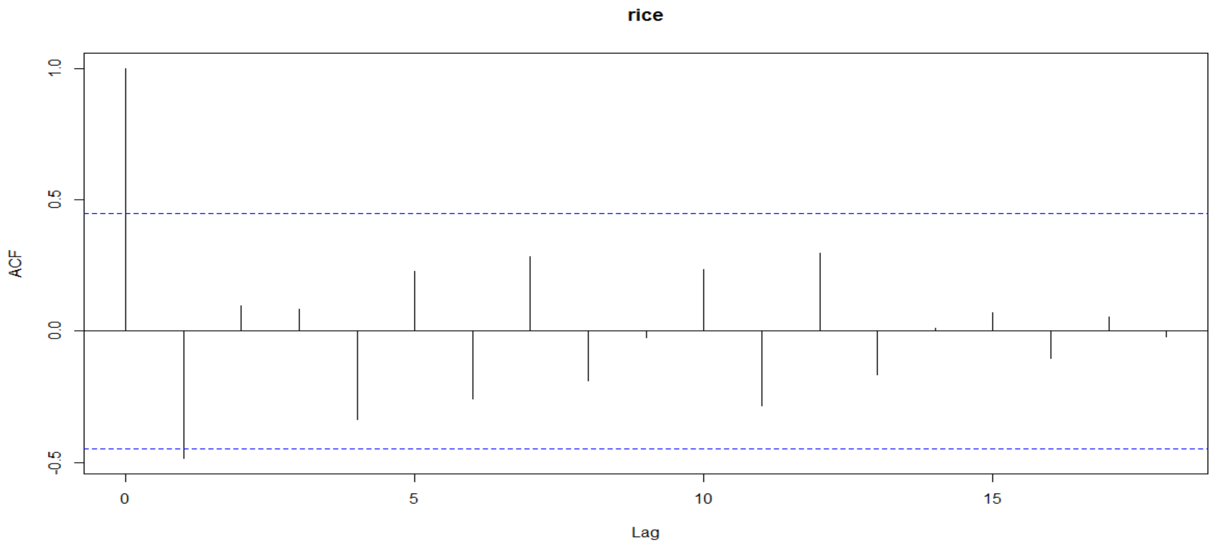
For example, to plot the correlogram for lags 1-20 of the once differenced time series of the Production of Rice in Andhra Pradesh, and to get the values of the autocorrelations, we type:

ACF

```
> acf(Foodtimeseriesdiff1, lag.max=20) # plot a correlogram
> acf(Foodtimeseriesdiff1, lag.max=20, plot=FALSE) # get the autocorrelation v
alues

Autocorrelations of series ‘Foodtimeseriesdiff1’, by lag

     0      1      2     3      4     5      6     7      8      9    10
 1.000 -0.483   0.098 0.084 -0.336 0.229 -0.257 0.285 -0.188 -0.025 0.234
  11     12        13     14         15        16         17       18
-0.165   0.012    0.071  -0.103     0.053    -0.023      -0.284   0.298
```



We see from the correlogram that the autocorrelation at lag 1 (-0.483) exceeds the significance bounds, but all other autocorrelations between lags 1-20 do not exceed the significance bounds.

PACF

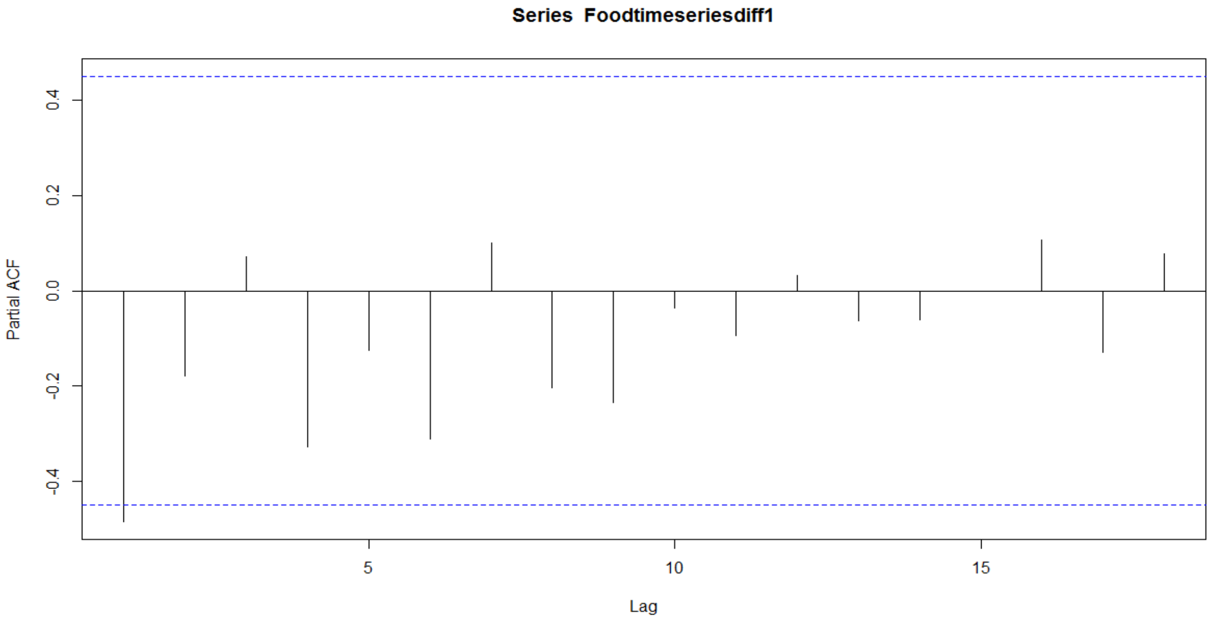
To plot the partial correlogram for lags 1-20 for the once differenced time series of the Production of Rice in Andhra Pradesh (year wise), and get the values of the partial autocorrelations, we use the "pacf()" function, by typing:

```
> pacf(Foodtimeseriesdiff1, lag.max=20) # plot a partial correlogram
> pacf(Foodtimeseriesdiff1, lag.max=20, plot=FALSE)
```

Partial autocorrelations of series 'Foodtimeseriesdiff1', by lag

1	2	3	4	5	6	7	8	9	10	11
-0.483	-0.178	0.072	-0.327	-0.124	-0.310	0.101	-0.202	-0.234	-0.036	-0.093

12	13	14	15	16	17	18
-0.061	0.000	0.107	-0.129	0.079	0.032	-0.062



The partial correlogram shows that the partial autocorrelations at lags 1, 2 exceed the significance bounds, are negative, and are slowly decreasing in magnitude with increasing lag (lag 1: -0.483, lag 2: -0.178). The partial autocorrelations tail off to zero after lag 2(i, e. 0.072).

Since the correlogram is zero after lag 1, and the partial correlogram tails off to zero after lag 2, this means that the following ARMA (autoregressive moving average) models are possible for the time series of first differences:

- ❖ An ARMA(3,0) model, that is, an autoregressive model of order p=3, since the partial autocorrelogram is zero after lag 2, and the autocorrelogram tails off to zero (although perhaps too abruptly for this model to be appropriate)
- ❖ An ARMA(0,1) model, that is, a moving average model of order q=1, since the autocorrelogram is zero after lag 1 and the partial autocorrelogram tails off to zero.

- ❖ An ARMA(p,q) model, that is, a mixed model with p and q greater than 0, since the autocorrelogram and partial correlogram tail off to zero (although the correlogram probably tails off to zero too abruptly for this model to be appropriate)

We use the principle of parsimony to decide which model is best: that is, we assume that the model with the fewest parameters is best. The ARMA(3,0) model has 3 parameters, the ARMA(0,1) model has 1 parameter, and the ARMA(p,q) model has at least 2 parameters. Therefore, the ARMA(0,1) model is taken as the best model.

An ARMA(0,1) model is a moving average model of order 1, or MA(1) model. This model can be written as: X_t - mu = Z_t - (theta * Z_t-1), where X_t is the stationary time series we are studying (the first differenced series of Rice production), mu is the mean of time series X_t, Z_t is white noise with mean zero and constant variance, and theta is a parameter that can be estimated.

Forecasting Using an ARIMA Model

- Once you have selected the best candidate ARIMA (p, d, q) model for your time series data, you can estimate the parameters of that ARIMA model, and use that as a predictive model for making forecasts for future values of your time series.
- You can estimate the parameters of an ARIMA (p, d, q) model using the "arima()" function in R.

```
> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(1,0,0))
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(1, 0, 0))

Coefficients:
         ar1   intercept
      0.3257 11217.5246
s.e.  0.2129    565.0063

sigma^2 estimated as 3026412:  log likelihood = -177.66,  aic = 361.33

> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(2,0,0))
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(2, 0, 0))

Coefficients:
         ar1     ar2     intercept
      0.2505    0.2271   11200.9446
s.e.  0.2161    0.2169     681.3241

sigma^2 estimated as 2856109:  log likelihood = -177.14,  aic = 362.27
> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(0,0,1))
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(0, 0, 1))

Coefficients:
         ma1   intercept
      0.2254  11236.8911
s.e.  0.1882    482.6728

sigma^2 estimated as 3151186:  log likelihood = -178.04,  aic = 362.08
```

```
> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(1,0,3))
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(1, 0, 3))

Coefficients:
          ar1     ma1     ma2     ma3    intercept
      -0.3706   0.6948  0.6400  0.6761   11242.2159
s.e.   0.3232   0.2673  0.3717  0.3836     711.8063

sigma^2 estimated as 2273946:  log likelihood = -175.68,  aic = 363.36

> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(1,0,1))
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(1, 0, 1))

Coefficients:
         ar1      ma1   intercept
      0.6395  -0.3377   11194.665
s.e.  0.3486   0.3990     666.101

sigma^2 estimated as 2931854:  log likelihood = -177.37,  aic = 362.74

> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(2,0,3))
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(2, 0, 3))

Coefficients:
         ar1      ar2     ma1     ma2     ma3    intercept
      0.1529  -0.3683  0.0540  0.9464  0.2600   11280.2440
s.e.  0.9125   0.3271 0.9832 0.2991 0.7911       584.4509

sigma^2 estimated as 2070344:  log likelihood = -175.53,  aic = 365.05

> Foodtimeseriesarima <- arima(Foodtimeseries, order=c(0,1,1)) # fit an ARIMA(
0,1,1) model
> Foodtimeseriesarima

Call:
arima(x = Foodtimeseries, order = c(0, 1, 1))
```

```
Coefficients:
          ma1
      -0.6383
s.e.   0.2429

sigma^2 estimated as 3327416:  log likelihood = -169.89,  aic = 343.78
```

- Finally, considering all graphical and formal test, it is clear that our fitted ARIMA (0, 1,1) mo del is the best selected model for forecasting Rice productions in the Andhra Pradesh.

- As mentioned above, if we are fitting an ARIMA (0, 1, 1) model to our time series, it means we are fitting an an ARMA (0, 1) model to the time series of first differences. An ARMA (0, 1) model can be written X_t - mu = Z_t - (theta * Z_t-1), where theta is a parameter to be estimated. From the output of the "arima()" R function (above), the estimated value of theta (given as 'ma1' in the R output) is -0.6383 in the case of the ARIMA(0,1,1) model fitted to the time series of Rice production of Andhra Pradesh.

Specifying the confidence level for prediction intervals

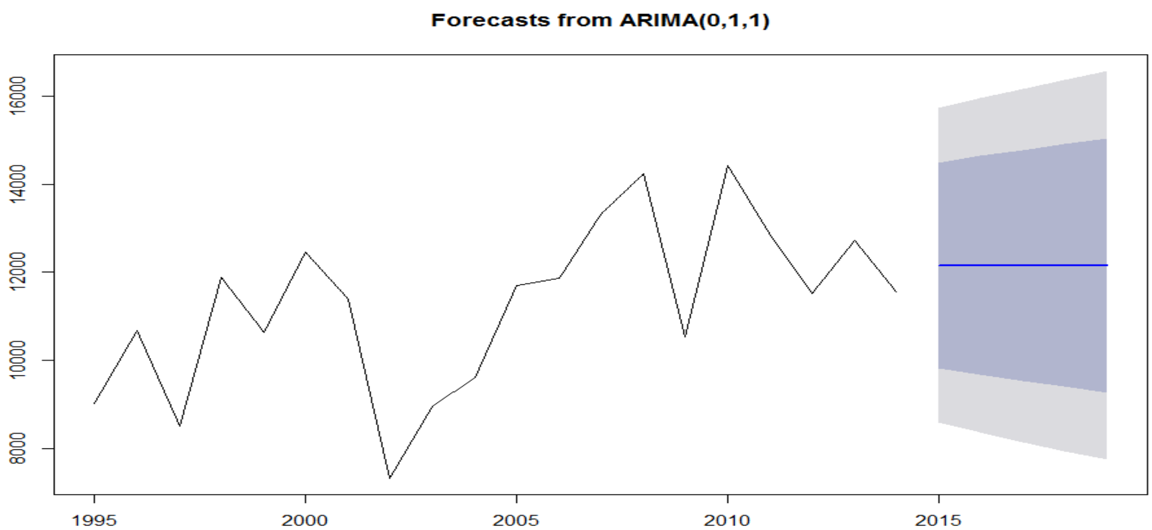
- You can specify the confidence level for prediction intervals in forecast. Arima() by using the "level" argument. For example, to get a 99.5% prediction interval, we would type "forecast. Arima (Foodtimeseriesarima, h=5)
- We can then use the ARIMA model to make forecasts for future values of the time series, using the "forecast. Arima()" function in the "forecast" R package. For example, to forecast the Rice Production of the next five years, we type:

```
> library(timeDate)
> library(forecast)
> Foodtimeseriesforecasts <- forecast.Arima(Foodtimeseriesarima, h=5)
> Foodtimeseriesforecasts
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2015        11565.4 8812.699 14318.10 7355.508 15775.29
2016        11565.4 7672.494 15458.31 5611.713 17519.09
2017        11565.4 6797.583 16333.22 4273.653 18857.15
2018        11565.4 6059.999 17070.80 3145.616 19985.18
2019        11565.4 5410.174 17720.63 2151.795 20979.01
```

- The original time series for the Rice production in Andhra Pradesh 20st years of production. The forecast.Arima() function gives us a forecast of the Rice Production of the next five years in Andhra Pradesh (Year 2015-2019), as well as 80% and 95% prediction intervals for those predictions. The Production of Rice of the 20th years (the last observed value in our time series), and the ARIMA model gives the forecasted Rice production of the next five years as 11565.4 million tonnes.

- We can plot the observed Rice production for the first 20 years, as well as the productions that would be predicted for these 20years and for the next 5 years using our ARIMA (0,1,1) model, by typing:

```
> plot.forecast(Foodtimeseriesforecasts)
```



- As in the case of exponential smoothing models, it is a good idea to investigate whether the forecast errors of an ARIMA model are normally distributed with mean zero and constant variance, and whether there are correlations between successive forecast errors.

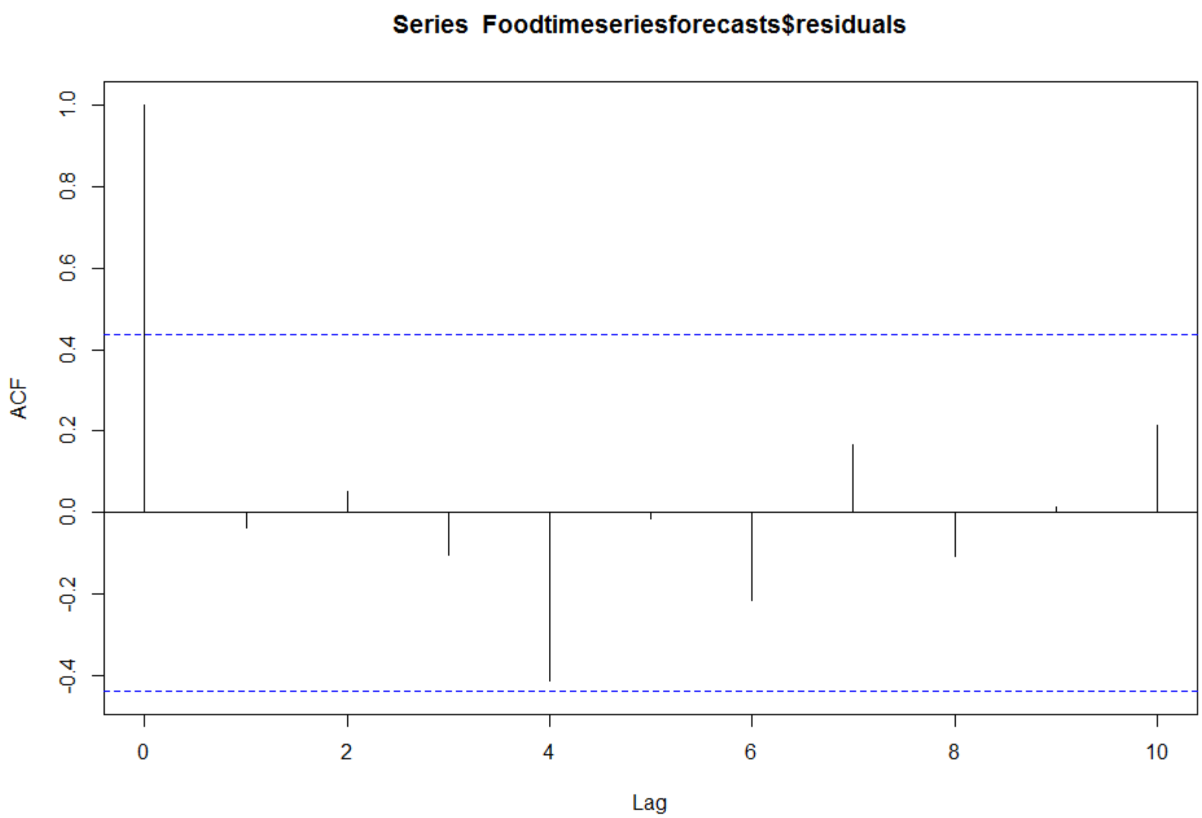
❖ For example, we can make a correlogram of the forecast errors for our ARIMA(0,1,1) model for the Rice production in Andhra Pradesh, and perform the Ljung-Box test for lags 1-10, by typing:

Plot ACF and PACF for residuals of ARIMA

```
> acf(Foodtimeseriesforecasts$residuals, lag.max=10)
> Box.test(Foodtimeseriesforecasts$residuals, lag=10, type="Ljung-Box")

        Box-Ljung test

data:  Foodtimeseriesforecasts$residuals
X-squared = 9.9709, df = 10, p-value = 0.4431
```



❖ Since the correlogram shows that none of the sample autocorrelations for lags 1-10 exceed the significance bounds, and the p-value for the Ljung-Box test is 0.4431, we can conclude that there is very little evidence for non-zero autocorrelations in the forecast errors at lags 1-10.

- To investigate whether the forecast errors are normally distributed with mean zero and constant variance, we can make a time plot and histogram (with overlaid normal curve) of the forecast errors.

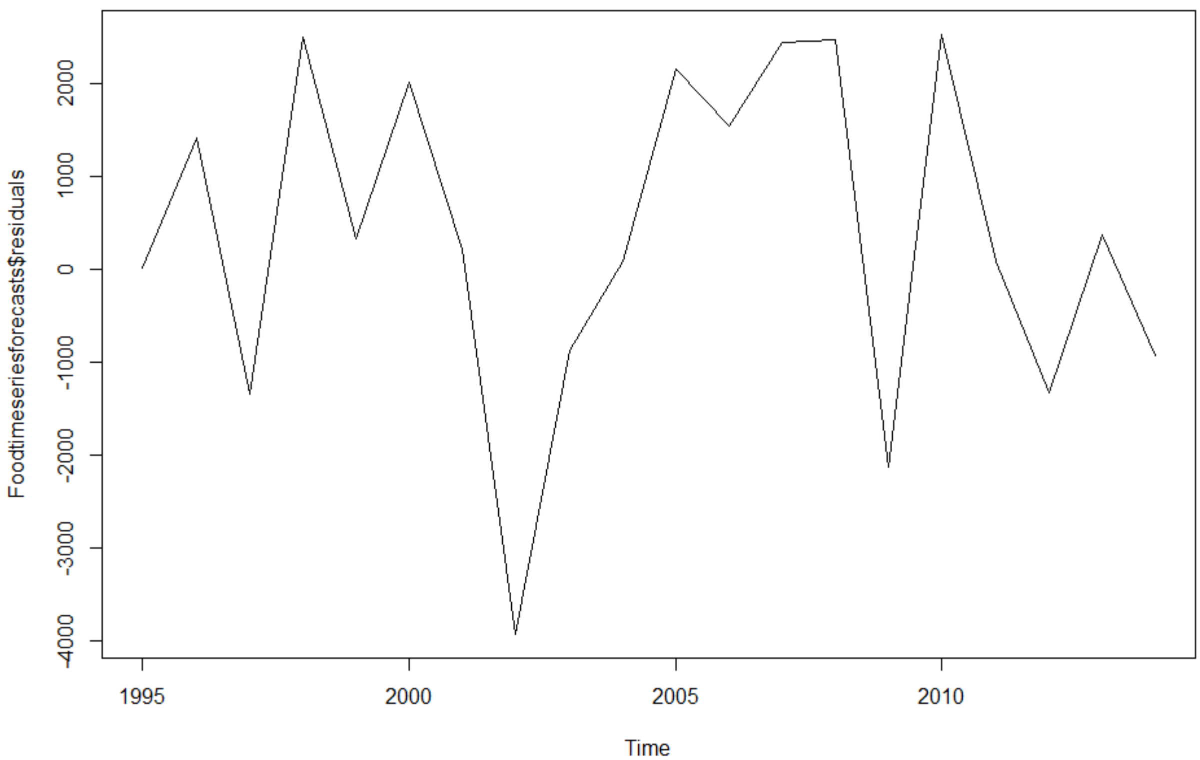
Here, it test overall randomness on the basis of lag,

H0: The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

Ha: The data are not independently distributed; they exhibit serial correlation.

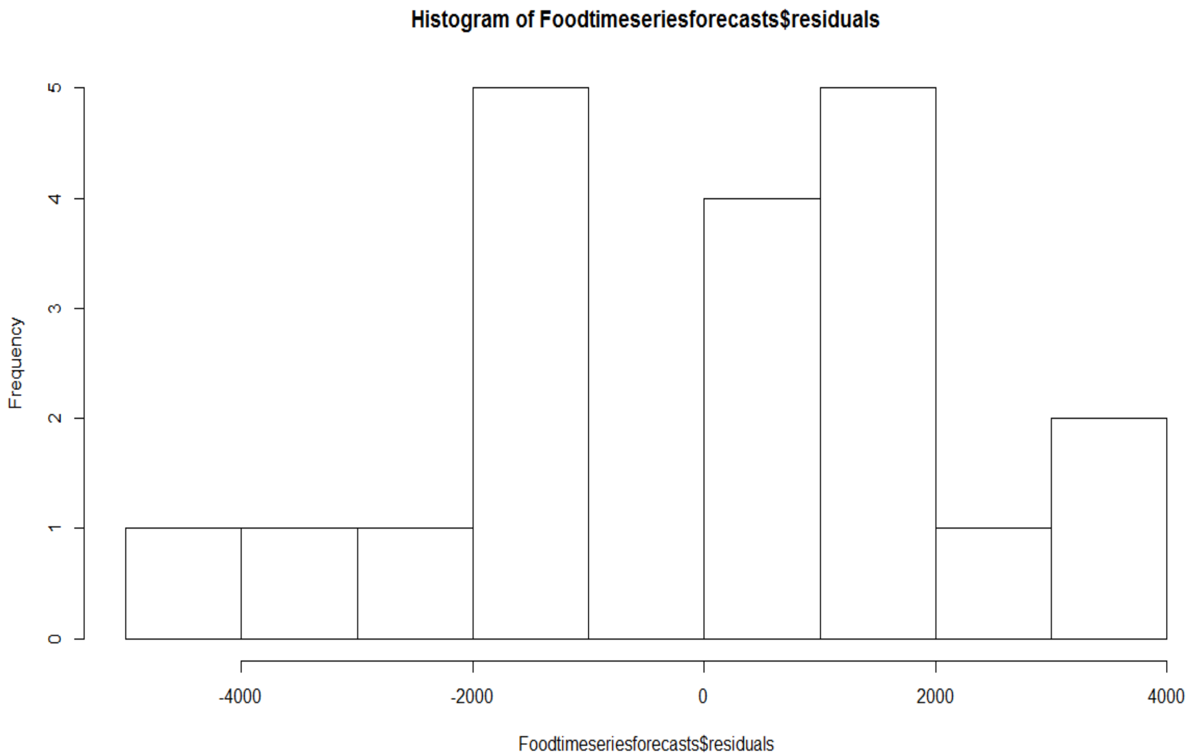
So here, as the p value is more than 0.05, so I can say that data are independently distributed.

```
> plot.ts(Foodtimeseriesforecasts$residuals) # make time plot of forecast err
ors
```



In case of arima(0,1,1) all the value has improved so it becomes favorable. Value of x square is almost same in both the situation.

```
> hist(Foodtimeseriesforecasts$residuals) # make a histogram
```



- The time plot of the in-sample forecast errors shows that the variance of the forecast errors seems to be roughly constant over time (though perhaps there is slightly higher variance for the second half of the time series). The histogram of the time series shows that the forecast errors are roughly normally distributed and the mean seems to be close to zero. Therefore, it is plausible that the forecast errors are normally distributed with mean zero and constant variance.

- Since successive forecast errors do not seem to be correlated, and the forecast errors seem to be normally distributed with mean zero and constant variance, the ARIMA(0,1,1) does seem to provide an adequate predictive model for the ages at death of English kings.

CONCLUSION AND DISCUSSION

In the study ARIMA (0, 1, 1) and ARIMA (0, 1, 1) were developed models for rice cultivated areas and production of rice respectively. From the forecast available by using the developed model it can be seen that forecasted rice cultivated production were to increase in the next five years. The validity of the forecasted value can be checked when the data for the lead periods become available. The model can be used by researchers for forecasting rice cultivated production in Andhra Pradesh, India. However data need to be updated from time to time with incorporation of current values.

REFERENCE

Agrawal, R. K. R. A. K., Adhikari, R., & Agrawal, R. K. R. A. K. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *arXiv Preprint arXiv:1302.6613*, *1302.6613*, 1–68.

Kasyoki, A. (2015). Simple Steps for Fitting Arima Model to Time Series Data for Forecasting Using R, *4*(3), 2013–2016.

Khin, A. A. A., F.C, E. C., Shamsudin, M. N., Mohamed, Z. A., Chong, E., Aye Aye Khin, B., … Chiew Fook Chong, E. (2008). Natural rubber price forecasting in the world market. *Agriculture Sustainability through Participative Global Extension*, 1–13. Retrieved from http://www.fachmannhunter.com/home/wp-content/uploads/2011/05/aye-eye-khin-66.pdf

Winklhofer, H., Diamantopoulos, a, & Witt, S. (1996). Forecasting practice: A review of the empirical literature and an agenda for future research. *International Journal of Forecasting*, *12*(2), 193–221. http://doi.org/http://dx.doi.org/10.1016/0169-2070(95)00647-8

Khin, A. A. A., F.C, E. C., Shamsudin, M. N., Mohamed, Z. A., Chong, E., Aye Aye Khin, B., … Chiew Fook Chong, E. (2008). Natural rubber price forecasting in the world market. *Agriculture Sustainability through Participative Global Extension*, 1–13. Retrieved from http://www.fachmannhunter.com/home/wp-content/uploads/2011/05/aye-eye-khin-66.pdf

Agrawal, R. K. R. A. K., Adhikari, R., & Agrawal, R. K. R. A. K. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *arXiv Preprint arXiv:1302.6613*, *1302.6613*, 1–68.

Kasyoki, A. (2015). Simple Steps for Fitting Arima Model to Time Series Data for Forecasting Using R, *4*(3), 2013–2016.

Khin, A. A. A., F.C, E. C., Shamsudin, M. N., Mohamed, Z. A., Chong, E., Aye Aye Khin, B., … Chiew Fook Chong, E. (2008). Natural rubber price forecasting in the world market. *Agriculture Sustainability through Participative Global Extension*, 1–13. Retrieved from http://www.fachmannhunter.com/home/wp-content/uploads/2011/05/aye-eye-khin-66.pdf

Winklhofer, H., Diamantopoulos, a, & Witt, S. (1996). Forecasting practice: A review of the empirical literature and an agenda for future research. *International Journal of Forecasting*, *12*(2), 193–221. http://doi.org/http://dx.doi.org/10.1016/0169-2070(95)00647-8