Original Research Article

GROWTH RATE, TREND IN AREA, PRODUCTION AND PRODUCTIVITY AND ALSO FORECASTING TECHNIQUE ARIMA MODEL FOR PRODUCTION OF COTTON CROP IN BALLARI DISTRICT OF KARNATAKA

ABSTRACT

Cotton is an important principal commercial fiber crop. It is one of the most leading and important cash crops in Indian economy. In the present study an attempt has been made using secondary data for forty six years (from 1970 to 2016) to understand the growth rates in area, production and productivity of cotton crop in Ballari district of Karnataka. The results revealed that area, production and productivity of cotton crop marked a significant increase in growth rate during the study period. In case of Ballari district, increasing growth rate in area (2.78%), production (3.26%) and productivity (2.99%) was observed. To understand the trends in area, production and productivity of cotton crop in Ballari district of Karnataka. It was observed that among different polynomial models fitted for area, production and productivity of cotton crop, For Ballari district, we observe that quartic model was found to be the best fit for area under cotton with RMSE of 12.848 and adjR^2value of 0.813.For production quartic model was the best fit with RMSE of 18.653 and adjR^2value of 0.869 while for productivity cubic model was the best fit with RMSE of 50.548 and adjR^2value of 0.7.And also to understand the forecast in production of cotton crop in Ballari district of Karnataka. The results revealed that, forecast was made for the production of cotton crop. Forecasting was carried out using ARIMA based on RMSE and MAPE values for production of cotton crop in Ballari district of Karnataka. Forecasting have been done for next four years were carried out using a short term forecasting technique ARIMA for production of cotton crop in Ballari district of Karnataka. It was found that ARIMA (2, 1, 2) model was appropriate for Ballari district.

KEYWORDS: Cotton, growth rate, trend, polynomial models, forecasting, ARIMA, RMSE and MAPE.

Comment [M1]: This sentence seems incomplete.

Comment [M2]: Is it a continuation of the previous sentence or a new sentence?

Comment [M3]: This sentence is wrong.

INTRODUCTION

Cotton is the most essential natural fiber crop in the world for textile produce, accounting for about 50 per cent of all fibers used in the textile industry. It is more important than the various synthetic fibers, and it is grown all over the world in about 80 countries. It is unique among agricultural crops, because it is the main natural fiber crop, which provides edible oil and seed by-products for livestock feed. Further, it also provides income for hundreds of millions of people. It is one of the agro-industrial crops which are produced in both developing and developed countries. Cotton fibers are used in clothing and household furnishings. It has played an important role since the industrial revolution of the 17th century. Currently, it is an important cash crop especially for a number of developing countries at local and national levels (Gudeta and Egziabher, 2019). India is primarily an agriculture based country and its economy largely depends on agriculture. Agricultural growth is necessary not only for attaining high overall growth but also for accelerating the poverty reduction in a developing country like India (Pavitra, et al., 2018).India cultivates the highest acreage under cotton in the world. It provides the basic raw material (cotton fiber) to the cotton textile industry (Rajan and Palanivel, 2018). It is the leading textile fiber in the world accounting for 35 per cent of the world fiber use. Cotton was first cultivated about 7,000 years ago, by the inhabitants of the Indus Valley Civilization. This civilization covered a huge swath of the north-western part of the Indian sub-continent, comprising today's parts of eastern Pakistan and north-western India (Mayilsami and Selvaraj, 2016). Cotton has been traditionally known as the backbone of nonfood crops of agricultural economy of India (Sharma, 2015). About 25 per cent of our country's Gross Domestic Product (GDP) comes from agricultural sector. Nearly 75 per cent of the country's population lives in villages and depends on agriculture (Parmar, et al., 2016). Cotton is an important principal commercial fiber crop. It is also known as 'White gold' or the "King of Fibers" due to its importance in agricultural as well as industrial economy throughout the world. Cotton is one of the leading and important cash crops in Indian economy (Mohammad, et al., 2018). Cotton provides gainful employment to several million people in cultivation, trade, processing, manufacturing and marketing, *etc.* It serves vast handloom sector of the country. Apart from its use in textiles, cotton is also used as surgical lint and for various domestic purposes. Other plant parts of cotton are utilized in the manufacture of industrial products like paper, card board, blotting paper, *etc.* Cotton seeds have recently assumed greater importance as a source of edible

Comment [M4]: has

oil. Thus, cotton plays a vital role in the Indian economy by contributing to human utilization in a number of ways. Cotton cultivation provides livelihood to millions of farmers. Their economic welfare heavily relies on cotton productivity. Cotton yarn, fabric, garment, etc. are dependent on the availability of cotton. Export of raw cotton and cotton products can earn foreign exchange for India. But India is not able to come up with satisfactory performance in cotton productivity in spite of having varieties of cotton and availability of technology. Its productivity is far below the major cotton producing countries. Demand for cotton textile is rising progressively all over the world. Being the first largest cotton producing area, India has an immense opportunity in the growth of the cotton textile industry and strengthening of Indian economy in the coming years (Mal and Pandey, 2013). India is the only country in the world growing all the four cultivated species of cotton, *viz.*,*G. hirustum*, *G. arboretum*, *G. herbaceaum* and *G. barbadense*. The maximum area has been covered by the hybrids (Samuel et al., 2013). India is unique among the major cotton growing countries because of the broad range of agro-climatic and soil conditions which permit cultivation of all varieties and staple lengths of cotton. Major Cotton producing countries are India, China, USA, Pakistan, Brazil, Australia, Uzbekistan, Turkey, Turkmenistan and Burkina (Rajan and Palanivel, 2017).In the recent period, cotton is gaining momentum in non-traditional areas such as Odisha, West Bengal and Tripura. India accounts for approximately 25 per cent of worlds total cotton area and 18 per cent of global cotton production (Kulkarni et al., 2017). India ranks first with respect to area and production and eighth rank with respect to productivity of cotton. Cotton in India occupies an area of 118.81 lakh hectares with a production of 345.82 lakh bales and productivity of 495 Kg/ha. Cotton is cultivated in a majority of the states in the country. The ten major cotton producing states of India are Gujarat, Maharashtra, Telangana, Karnataka, Andhra Pradesh, Haryana, Madhya Pradesh, Rajasthan, Punjab and Tamil Nadu and accounts for more than 95 per cent of the area under cotton. In Karnataka area under cotton is around 7.5 lakh hectares which is 7 per cent of country's area. The production of crop is 28 lakh bales (around 4 per cent of country's production) while productivity is 653 kg/ha. The main cotton growing districts in Karnataka are Dharwad, Ballari and Raichur. The Government of India has launched " Technology Mission on Cotton" in February 2000 with an objective of improving the production and productivity of cotton through development of high yielding varieties; enhance the income of the cotton growers by reducing cost of cultivation, appropriate transfer of technology and better farm management practices, cultivation of *Bt*-cotton hybrids

Comment [M5]: *Gossypium* for the first time etc. Area estimation and forecasting of production are essential procedures supporting in policy decisions with respect to production, land use allocation, food security, environmental issues price structures as well as consumption of cotton in the country. Increased global demand for cotton should induce higher production in the next decade. With these backgrounds, it is necessary to know the extent of cotton production in future with available resources. Various approaches have been used for forecasting such agricultural systems. Borkar Prema *et al*. in their empirical study showed that ARIMA (2, 1, 1) is the appropriate model for forecasting the production of cotton in India. The study of Debnath *et al*. revealed that area, production and yield of cotton in India would increase from 2016-17 to 2020-21. Similar studies have been conducted by Payyamozhil, S. *et al*. and Sundar Rajan *et al*. for forecasting cotton production in India, the analysis revealed that ARIMA (0, 1, 0) is the best model for forecasting cotton production. The present study has been undertaken with an objective to forecast the production of cotton in India in future using Box- Jenkins ARIMA model. Among the stochastic time series models ARIMA types are very powerful and popular as they can successfully describe the observed data and can make forecast with minimum forecast error. These types of models are very difficult to identify and estimate.

Comment [M6]: When?

Comment [M7]: More explanation

METHODOLOGY

In the present study, major cotton growing district of Karnataka *viz.*, Ballari are selected. In order to study the growth, trend and forecasting by using ARIMA model of cotton, secondary data pertaining to the area, production and productivity of cotton crop for the period of 46 years (from 1970-71 to 2015-16) was obtained from the Directorate of Economics and Statistics, Bengaluru.

Comment [M8]: What?

STATISTICAL TOOLS AND MODELS EMPLOYED

The Linear and compound growth rates for area, production and productivity were estimated from the time series data for the period from 1970-71 to 2015-16 collected from the Directorate of Economics and Statistics, Bengaluru. The following analytical tools were used to estimate the growth rates by using the method of ordinary least squares. To understand the trend in area, production and productivity of cotton, an appropriate polynomial models were fitted by using the method of ordinary least squares. Cubic and Quartic models have been used in the present study (Singh and Supriya, 2017). And to fit the forecasting model ARIMA for production of cotton crop in Ballari district of Karnataka.

LINEAR GROWTH FUNCTION

Linear growth function is given by, $Y_t = a + b_t + e_t$

Where, t is the time in years, Y_t is the trend value, a and b are constants and e_t is error term.

The linear growth rate is calculated by the formula,

$$Linear\ growth\ rate\ (LGR\ \%) = \frac{b}{\bar{y}}\ x100$$

COMPOUND GROWTH FUNCTION

Compound growth function is given by, $Y_t = ab^t$

Taking log on both sides, we have, $Log\ Y_t = loga + t\ log\ b$

Where, t is the time in years, Y_t is the characteristic and a and b are parameters (Ramakrishna and Bhave, 2017).

The compound growth rate (CGR %) is calculated by using the formula,

$$\text{CGR (\%)} = (\text{antilog b} - 1) \text{ x } 100$$

The significance of the growth rates can be tested by applying student's t- test as

$$\text{t} = \frac{\text{r}}{\text{SE (r)}} \sim \text{t}_{(\text{n}-2)} \text{ df}$$

Where, r is the growth rate, n is the total number of years under study and $SE(r)$ is the standard error of growth rate.

CO-EFFICIENT OF VARIATION (CV):

The co-efficient of variation (CV) has been worked out to find out the variation in the area, production and productivity of cotton crop in selected districts of Karnataka, over the years.

$$\text{CV} = \frac{\text{Standard deviation}}{\text{Mean}} \text{x } 100$$

Cubic function:

The cubic model is given by the equation, $Y_t = a + bt + ct^2 + dt^3 + e_t$

Where, Y_t is the dependent variable i.e., area, production and productivity, t is the independent variable, time in years, e_t is error term a, b, c and d are parameters.

Quartic function:

The quartic model is given by the equation, $Y_t = a + bt + ct^2 + dt^3 + et^4 + e_t$

Where, Y_t is the dependent variable i.e., area, production and productivity, t is the independent variable, time in years, e_t is error term and a, b, c, d and e are parameters.

Model Adequacy Checking:

Adequacy of a model indicates the suitability of the model to explain the inherent nature of the collected information. The assumptions made in a linear regression model are, linear dependence of "Y" on regressors, independence and identical distribution (normal) of errors with zero mean. Gross violations of the assumptions may yield an unstable model in the sense that a different sample could lead to a totally different model with opposite conclusion. We cannot detect departures from the underlying assumptions by examination of the summary statistics such as "t" or "F" statistics or R^2. These are "Global" model properties and as such they do not ensure model adequacy. Hence,diagnostic methods, primarily based on study of the model residuals. The diagnostics checks of randomness and normality of residuals ascertains the independence and distribution assumption of data. The coefficient of determination (R^2) is a test statistic that will give information about the appropriateness of a model. R^2 value is the proportion of variability in a data set that is accounted for by the statistical model. It provides a measure of how well the assumed model explains the variability in dependent variable.

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

Where, ESS is error sum of squares, RSS is regression sum of squares, TSS is total sum of squares. Computed R^2 value lies between zero and one. If R^2 value is closer to 1 indicates that the model fits the data. Adjusted R^2 and Root Mean Square Error (RMSE) are also used for the checking of the fit of model.

Adjusted R^2

The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The adjusted R-squared can be negative, but it's usually not. It is always lower than the R-squared.

$$Adjusted\ R^2 = \frac{RSS/df}{TSS/df}$$

Where, RSS is regression sum of squares, TSS is total sum of squares and df is the respective degrees of freedom.

Testing for significance of regression coefficient

Significance of regression coefficient is tested using F-test statistic.

Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) (also called the root mean square deviation, RMSD) is used to assess the amount of variation that the model is unable to capture in the data. The RMSE is obtained as the square root of the mean squared error hence considered as the model prediction capability and is obtained as,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}{n}}$$

Where, Y_t= observed value, $\hat{Y}_t$= predicted value and n= number of observation.

To Forecast the Production of Cotton Crop in selected Districts of Karnataka

Forecasting using Auto Regressive Integrated Moving Average (ARIMA) Model

The Box-Jenkins procedure is concerned with fitting a mixed Auto Regressive Integrated Moving Average (ARIMA) model to a given set of data. The main objective in fitting this ARIMA model is to identify the stochastic process of the time series and predict the future values accurately. These methods have also been useful in many types of situation which involve the building of models for discrete time series and dynamic systems. Before using ARIMA model for forecasting the data should be checked for autocorrelation, which can be done by using Durbin Watson test. A value below 1.5 and a value above 2.5 indicate the presence of autocorrelation in data. Auto Regressive (AR) models were first introduced by Yule in 1926. These were consequently supplemented by Slutsky who in 1937 presented Moving Average (MA) schemes. Wold (1938), combined both AR and MA schemes and showed that ARMA processes can be used to model all stationary time series as long as the appropriate order of p, the number of AR terms, and q, the number of MA terms stands. Before discussing the ARIMA model building, some basic concepts of linear time series analysis, such as ~~stationarity~~stationary, seasonality and a short reference to the most classical common types of time series forecasting process are discussed.

a. Stationarity and non-~~stationarity~~stationary

A Time series is said to be stationary if its underlying generating process is based on a constant mean and constant variance with its autocorrelation function (ACF) essentially constant through time. Otherwise it is called non-stationary. A statistical test for stationarity has been proposed by Dickey and Fuller (1979).$\Delta Y_{t-1} = \gamma Y_{t-1} + \varepsilon_t$

Where, $\gamma = \Phi - 1$,Then, null hypothesis of $H_0: \gamma = 0$ against the alternative hypothesis $H_1: \gamma < 0$. Acceptance of null hypothesis indicates that the series is stationary. Usually, differencing is applied until the ACF shows an interpretable pattern with only a few significant autocorrelations.

b. Seasonality

In addition to trend, which has now been provided for, stationary series quite commonly display seasonal behavior where a certain basic pattern tends to be repeated at regular seasonal intervals. The seasonal pattern may additionally display constant change over the time as well. Just as regular differencing was applied to the overall trending series, seasonal differencing (SD) is applied to seasonal non-stationarity as well as autoregressive and moving average tools are available with the overall series, so too, they are available for seasonal phenomena using seasonal autoregressive parameters (SAR) and seasonal moving average parameters (SMA).

c. Autocorrelation Function (ACF)

The most important tools for the study of dependence is the sample autocorrelation function. The correlation coefficient between any two random variables X, Y, which measures the strength of linear dependence between X, Y, always takes values between -1 and 1. If ~~stationarity~~stationary is assumed and autocorrelation function ρ_k for a set of lags K = 1, 2..., is estimated by simply computing the sample correlation coefficient between the pairs, k units apart in time. The correlation coefficient between Y_t and Y_{t-k} is called the lag-k autocorrelation or serial correlation coefficient of Y_t and it is denoted by symbol ρ_k, under the assumption of weak ~~stationarity~~stationary, define as:

$$\rho_k = \frac{\sum_{t-k+1}^{r}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t-1}^{r}(Y_t - \bar{Y})} = \frac{\gamma_k}{\gamma_0}; for\ k = 1,2, \dots where, \gamma_k = cov(Y_t, Y_{t-k})$$

It ranges from -1 to +1. Box and Jenkins has suggested that maximum number of useful ρ_k are roughly $N/4$where N is the number of period upon which information on $Y_t$is available.

Comment [M9]: Un clear

d. Partial Autocorrelation Function (PACF)

The correlation coefficient between two random variables Y_t and Y_{t-k} after removing the impact of the intervening $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$ is called (PACF) at lag k and denoted by ϕ_{kk}

$$\phi_{00} = 1, \phi_{11} = P_1$$

$$\phi_{kk} = \frac{P_k - \sum_{j-1}^{k-1} \phi_{k-1} P_{k-j}}{1 - \sum_{j-1}^{k-1} \phi_{k-1} P_j}, k = 2,3, \ldots \ldots where\ \phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-1}$$

e. Autocorrelation function (ACF) and partial autocorrelation function (PACF)

Theoretical ACFs and PACFs (Autocorrelations versus lags) are available for the various models chosen and for various values of orders of autoregressive and moving average components i.e. p and q. Thus comparing the correlograms (plot of sample ACFs versus lags) obtained from the given time series data with these theoretical ACF/PACFs, we find a reasonably good match and tentatively select one or more ARIMA models.

White Noise (WN)

Avery important case of stationary process is called white noise. For a white noise series, all the ACFs are zero or close to zero .If $\{r_t\}$ is normally distributed with zero mean and variance σ^2 and no autocorrelation, then it is said to be Gaussian white noise.

Autoregressive process (AR)

A stochastic model that can be extremely useful in the representation of certain practically occurring series is the autoregressive model. In this model, the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a errorε_t.

A model written in the form$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + \epsilon_t$is called autoregressive model of order p and abbreviated as AR (p), where $\phi$is autoregressive coefficient and ε_t is white noise. In general, a variable r_t is said to be autoregressive of order p [AR (p)], if it is a function of its p past values and can be represented as:

$$r_t = \sum_{i-1}^{P} \phi_i r_{t-i} + \epsilon_t$$

Moving Average process (MA)

A second type of Box-Jenkins model is called a "moving average" model. Although these models look very similar to the AR model, the concept behind them is quite different. Moving average parameters relate what happens in period t only to the random errors that occurred in past time periods. A series $\{r_t\}$ is called moving average of order q and abbreviated as MA (q), expressed in following form of equation:

$$r_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \cdots - \theta_q\epsilon_{t-q}$$

Where, θ is moving average coefficient and ε_t is white noise

The above equation can be written as,

$$r_t = \epsilon_t - \sum_{i-1}^{q} \theta_i\epsilon_{t-i}$$

Autoregressive Moving Average process (ARMA):

An autoregressive moving average is expressed in the form:

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + .. + \phi_p r_{t-p} + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - .. - \theta_q\epsilon_{t-q}$$

A stationary solution to above mentioned equation exists if and only if all the roots of the AR characteristic equation $\phi(x)=0$ are outside the unit circle. For invariability, the roots of $\theta(x)=0$ lie outside the unit circle, Where ε_t is a sequence of uncorrelated variables, also referred to as a white noise process, and ($\phi_1,...,\phi_p,\theta_1,...,\theta_q$) are unknown constants or parameters. The above equation can be written as:

$$(1 - \phi_1 B^1 - \phi_2 B^2 - .. - \phi_p B^p) r_t = (1 - \theta_1 B^1 - \theta_2 B^2 - \cdots - \theta_q B^q)\epsilon_t$$

Where B is the backshift operator, that is $B\ (X_t) = X_{t-1}$ and

$$\phi(B) = (1 - \phi_1 B^1 - \phi_2 B^2 - .. - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B^1 - \theta_2 B^2 - .. - \theta_p B^p)$$

Autoregressive Integrated Moving Average process (ARIMA)

ARIMA is one of the most traditional methods of non-stationary time series analysis. In contrast to the regression models, the ARIMA model allows r_t to be explained by its past, or lagged values and stochastic error terms. These models are often referred to as "mixed models". Although this makes forecasting method, more complicated, but the structure may indeed simulate the series better and produce a more accurate forecast. Pure models imply that the structure consists only of AR or MA parameters - not both. The models developed by this approach are usually called ARIMA models because they use a combination of autoregressive (AR), integration (I) - referring to the reverse process of differencing to produce the forecast, and moving average (MA) operations. An ARIMA model is usually stated as ARIMA (p, d, q). An autoregressive integrated moving average is expressed in the form:

If $w_t = \nabla^d r_t = (1 - B)^d r_t$ then

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{p-1} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_p \epsilon_{t-p}$$

If $\{W_t\}$ follows the ARMA (p, q) model, and $\{r_t\}$ is an ARIMA (p, d, q) process. For practical purposes, we can take is usually d = 1 or 2 at most. Above equation is also written as:

$$\phi(B) w^t = \theta_0 + \theta(B)\epsilon_t$$

Where, $\phi(B)$ is a stationary autoregressive operator, $\theta(B)$ is a stationary moving average operator, and ϵ_t is white noise and θ_0 is a constant. In the case of the pattern of seasonal time series ARIMA model is written as follows:

$$\phi(B)\Phi(B)\nabla^d \nabla_s^D r_t = \theta(B)\Theta(B)\epsilon_t$$

Where,

$w_t = \nabla^d \nabla_s^D r_t$, $\nabla^d = (1 - B)^d$ is number of regular differences and $\nabla_s^D = (1 - B^s)^D$ is number of seasonal differences.

Seasonal ARIMA model is denoted by (p, d, q) (P, D, Q), where p denotes the number of autoregressive terms, q, number of moving average terms and d, number of times a series must be differenced to induce stationarity. P, number of seasonal autoregressive components, Q, number of seasonal moving average terms and D denotes the number of seasonal differences required to induce stationarity.

The main stages in setting up a Box-Jenkins forecasting model are described below:

1. Identification

The foremost step in the process of modeling is to check for the stationarity of the series, as the estimation procedures are available only for stationary series. If the original series is non-

Comment [M10]: Reference?

stationary, then first of all it should be made stationary. Stationarity is achieved by differencing the data for required number of times they could be obtained by looking for significant autocorrelation and partial autocorrelation coefficients. Say, if second order auto correlation coefficient is significant, then an AR (2), or MA (2) or ARMA (2) model could be tried to start with. This is not a hard and fast rule, as sample autocorrelation coefficients are poor estimates of population autocorrelation coefficients. Still they can be used as initial values while the final models are achieved after going through the stages repeatedly. Stationarity can be analyzed graphically using a ACF plot. A slow decay over the period indicates non-stationarity. A sudden change in lags of ACF plot shows the data has become stationary. Further, if the sequence graph of data is stationary over mean or variance we say as stationarity is achieved. Order of AR i.e. p and MA i.e., q is obtained by the examination of PACF and ACF plots respectively. Number of lagged values outside the limit is the order of the model.

Comment [M11]: ?

2. Estimation of parameters

The model is said be good fit for the data if the Ljung Q statistics is non-significant. At the estimation stage, the coefficients of the identified models are estimated. Generally, method of least squares is used to estimate the parameters which utilizes the concept of minimizing the sum of squares due to residuals. At the estimation phase, Stationarity and invariability are checked for the coefficient obtained at the same time checking is also done in order to know, whether the model fit the data satisfactorily or not?. If the model is significant and estimates of the parameters are non-significant suitable transformation can be done depending on the data. Outliers can be detected and removed in order to get an appropriate model.

Outlier in Time Series: Time series observations may sometimes be affected by unusual events, disturbances, or errors that create spurious effects in the series and result in extraordinary patterns in the observations that are not in according with most observations in the time series. Such unusual observations may be referred to as outliers. They may be the result of unusual external events such as strikes, sudden political or economic changes, sudden changes in a physical system, and so on, or simply due to recording or gross errors in measurement. The presence of such outliers in a time series can have substantial effects on the behavior of sample autocorrelations, partial autocorrelations, estimates of ARMA model parameters, forecasting, and can even affect the specification of the model.

Types of Outliers

- Additive Outlier (AO): An outlier that affects a single observation. For example, a data coding error might be identified as an additive outlier.
- Level shift (LS): An outlier that shifts all observations by a constant, starting at a particular series point. A level shift could result from a change in policy.
- Innovational Outlier (IO): An outlier that acts as an addition to the noise term at a particular series point. For stationary series, an innovational outlier affects several observations. For non-stationary series, it may affect every observation starting at a particular series point.
- Local Trend (LT): An outlier that starts a local pattern at a particular series point.
- Transient: An outlier whose impact decays exponentially to 0.
- Seasonal additive: An outlier that affects a particular observation and all subsequent observations separated from it by one or more seasonal periods. All such observations are affected equally. A seasonal additive outlier might occur in beginning of a certain year, if sales are higher every January.
- Additive patch: A group of two or more consecutive additive outliers. Selecting this outlier type results in the detection of individual additive outliers in addition to patches of them.

Detection of outliers:

These outliers are detected one by one using SPSS 20 software. The outliers are removed until the parameter estimates are significant. In most of the cases 10 per cent of the more influential observations are deleted to obtain a significant parameter estimate. The importance of the coefficients is measured by their statistical significance. Each estimated coefficient has a sampling distribution with a certain standard error that is to be estimated. Most ARIMA estimation routine automatically tests the hypothesis that the true coefficient is zero. If the coefficients are highly correlated the estimates are of poor quality. To check the closeness of the fit, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and some more were calculated.

3. Diagnostic checking

Having chosen a particular ARIMA model and having estimated its parameters, the next step is to check whether the chosen model fits the data reasonably well, as it is possible that another ARIMA model might do the job well. Here selection of model will be done by criteria like R-square and MAPE (Mean Absolute Percent Error).

Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

$$M = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where, A_t is the actual value and F_t is the forecast value. The difference between A_t and is divided by the actual value A_t again. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points n. multiplying by 100 makes it a percentage error.

Forecasting accuracy checking

Among the best fitted ARIMA and exponential smoothing technique a best model is used for forecasting based on the accuracy of the testing. The accuracy is checked using two measures namely RMSE and MAPE. A major part of the data used for model fitting is called as training set and a smaller portion (usually 10%) of data used for checking forecasting accuracy is called as testing set.

RESULTS AND DISCUSSION

Ballari district

Growth rates for area, production and productivity of cotton crop for Ballari district during the study period 1970-71 to 2015-16 are shown in Table 1. It was observed that average area under cotton was 70.46 thousand hectares with a coefficient of variation of 44.74 per cent. The linear and compound growth rates during the study period were 2.48 and 2.78per cent per annum respectively. The area under cotton in Ballari district exhibited a positive significant trend according to linear and compound growth rates. The average production of cotton during the study period 1970-71 to 2015-16 was 64.52 thousand bales with a coefficient of variation of

84.59 per cent. The linear and compound growth rates recorded for the study period were 4.04 and 3.26 per cent per annum respectively. The production of cotton in Ballari district exhibited a positive significant trend. Regarding productivity of cotton in Ballari district, the average productivity of cotton during the study period 1970-71 to 2015-16 was 232.59 kg/ha, with a coefficient of variation of 48.57per cent. The linear and compound growth rates during the study period were 2.96 and 2.99 per cent respectively.The productivity of cotton also had exhibited a positive significant trend during the study period in Ballari district. Hence, the growth rates for area, production and productivity were significantly increasing during the study period for Ballari district. Similar growth rates were observed for area, production and productivity of cotton for Coastal Andhra region, Telangana region and Andhra Pradesh as a whole (Panasa, 2014)

Table 1: Growth rates for area, production, productivity of Cotton crop in Ballari district

Ballari district	Average	CV (%)	LGR (%)	CGR (%)
Area	70.46	44.74	2.48**	2.78**
Production	64.52	84.59	4.04**	3.26**
Productivity	232.59	48.57	2.96**	2.99**

**** Significance at 1% level**

Table 2: Parameter estimates of fitted models for Area, Production and Productivity of Cotton in Ballari district

Ballari District	**Model**	**Parameter estimates**					R^2	**Adj** R^2	**RMSE**	**P Values**
		a	**b**	**c**	**d**	**e**				
Area	**Quartic**	151.99**	-16.97**	1.517**	-0.05**	0.00**	0.83	0.81	12.85	<0.001
Production	**Quartic**	51.66**	-9.45**	1.04**	-0.04**	0.00**	0.88	0.87	18.65	<0.001
Productivity	**Cubic**	51.50**	19.26**	-0.91**	0.02**		0.79	0.78	50.55	<0.001

**** Significant at 1% level, * Significant at 5% level, NS-Non-significant**

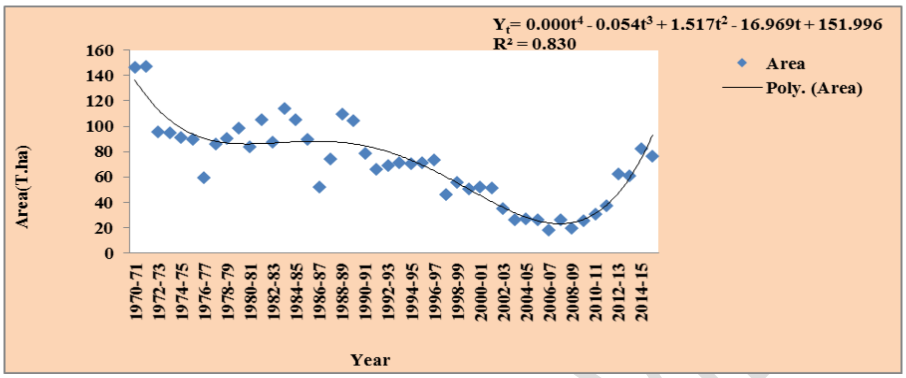

Fig.1a: Best fitted Quartic model for Area under Cotton in Ballari district

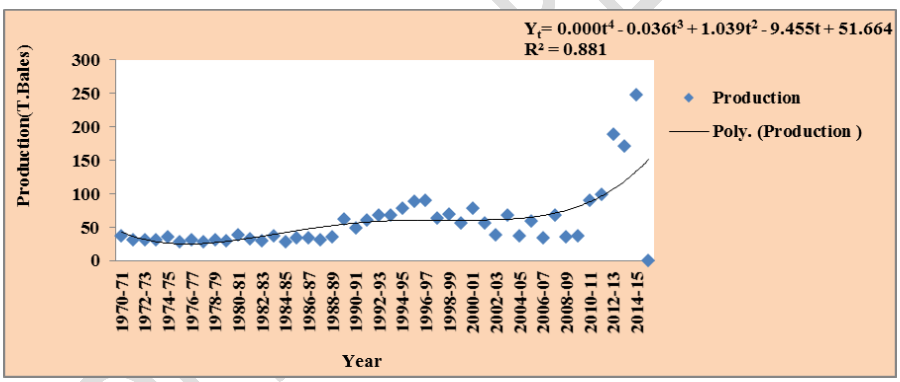

Fig.1b: Best fitted Quartic model for Production of Cotton in Ballari district

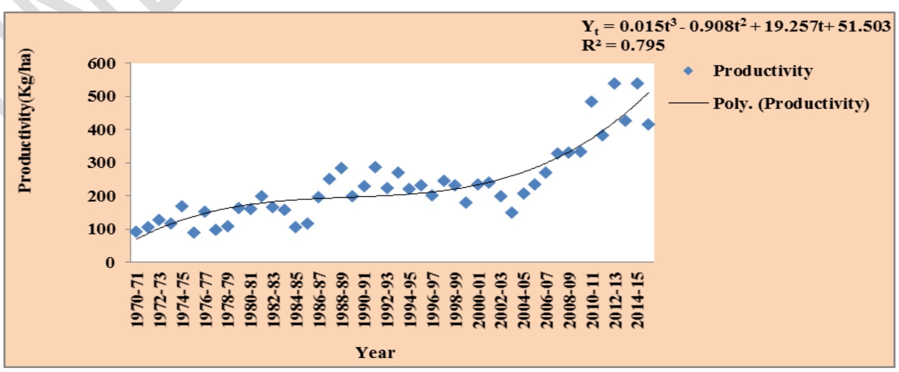

Fig.1c: Best fitted Cubic model for Productivity of Cotton in Ballari district

Table 3: Detected outliers of the ARIMA (2, 1, 2) model

Year	Type of outlier	Estimate	SE	t statistic	p-value
1997	Local Trend	12.055**	1.252	-9.626	0.000
2011	Local Trend	37.967**	1.252	30.328	0.000

**** Significant at 1 % level**

Table 4: Estimate of the ARIMA Model parameter for production of cotton in Ballari district

Transformation	Parameters	Lag	Estimate	SE	t statistic	p-value
No Transformation	**Constant**		-743.30**	105.204	-7.065	0.000
	AR	Lag 1	-1.538**	0.119	-12.87	0.000
		Lag2	-0.719**	0.142	-5.06	0.000
	Difference		1			
	MA	Lag 1	-0.199^{NS}	32.612	-0.006	0.995
		Lag2	0.801^{NS}	26.184	0.031	0.976

**** Significant at 1 % level, NS-Non-significant**

Table 5: Model fit statistics and Ljung-Box Q statistics for production of Cotton in Ballari district

Model Fit statistics			Ljung-Box Q			Number of Outliers detected
R^2	RMSE	MAPE	Statistic	DF	p-value	
0.977	9.117	13.660	19.859^{NS}	14	0.135	2

NS-Non-significant

Table 6: Forecasted values for production of cotton crop in Ballari district

Year	Forecast Value for production('000t)	95% Confidence interval	
		Lower	Upper
2017	318.73	283.81	353.64
2018	355.68	316.68	394.68
2019	391.04	335.31	446.77
2020	424.51	350.81	498.20

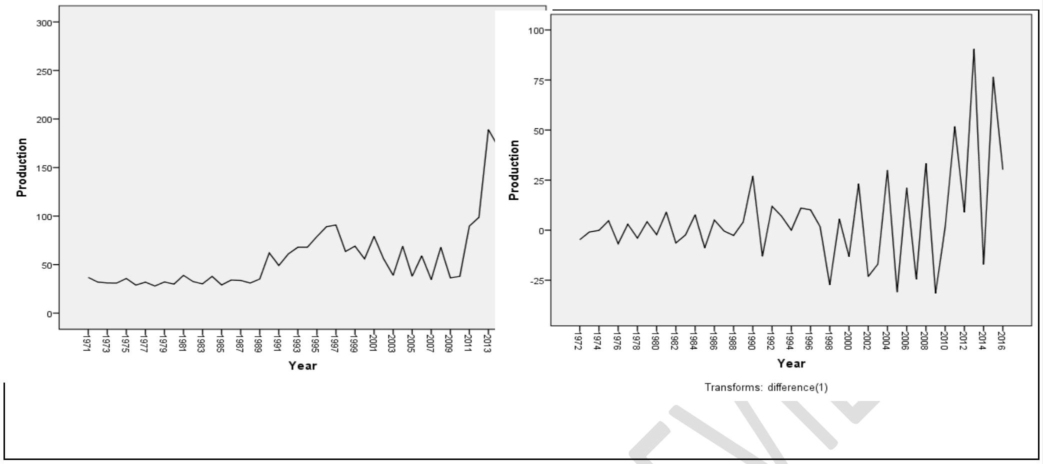

Fig.2: Time plot for production of cotton in Ballari district

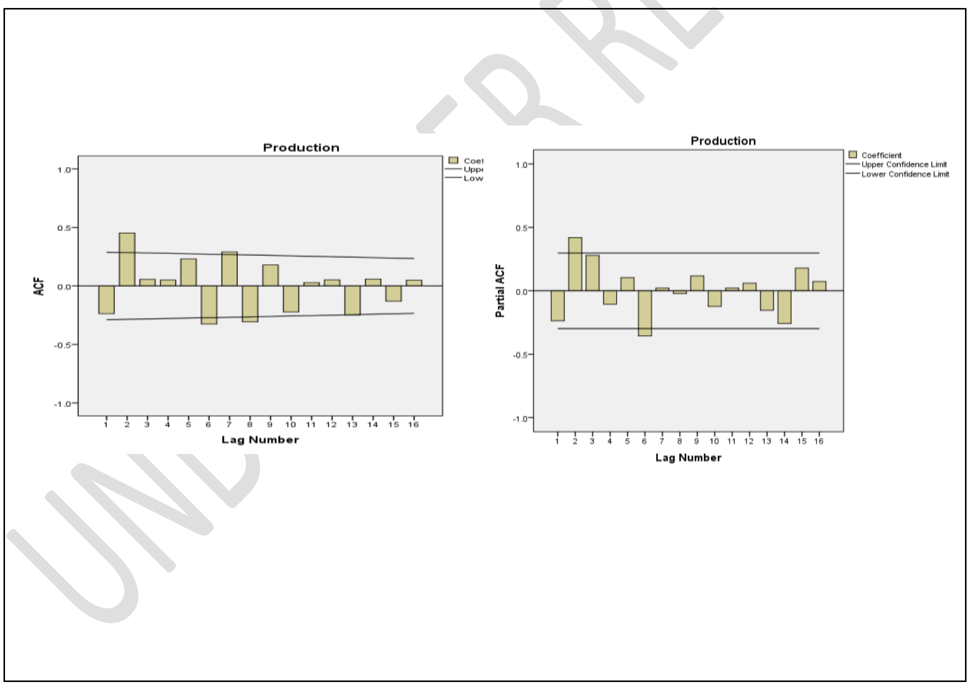

Fig. 3: ACF AND PACF of cotton in Ballari district

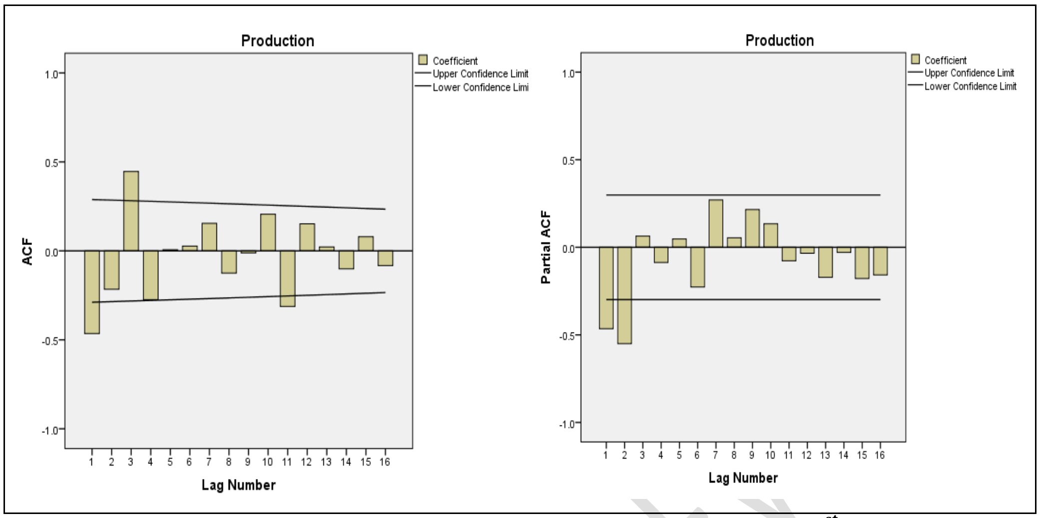

Fig.4: Autocorrelations and Partial autocorrelations at different lags of 1st differenced time series for production of cotton in Ballari district

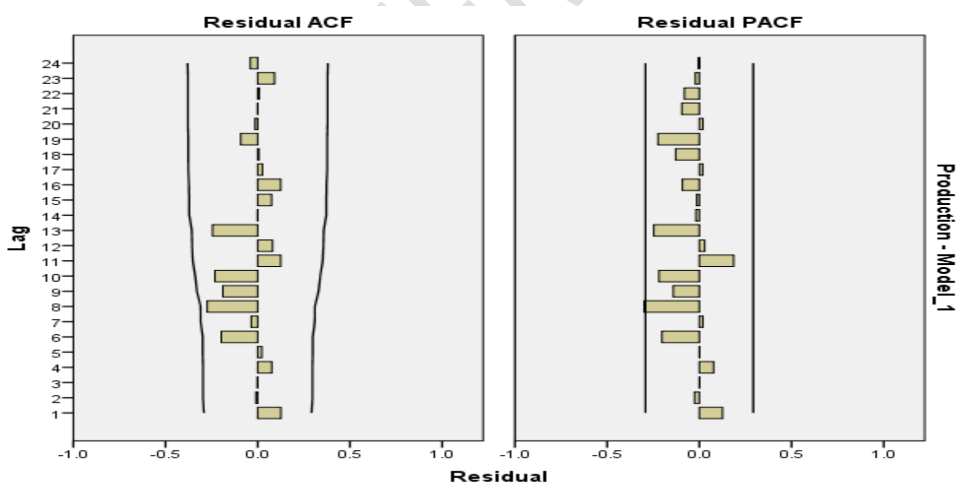

Fig. 5: Residual autocorrelation and partial autocorrelations for production of cotton in Ballari district

Different polynomial models were fitted for the time series data on area, production and productivity of cotton crop in selected districts of Karnataka. A model was selected as a best fit for the data when the coefficient of the higher order polynomial was non-significant. In the present study, four growth models *viz.*, linear, quadratic, cubic and quartic have been fitted to the time series data on area, production and productivity of the cotton crop in selected districts of Karnataka.

Polynomial models for area, production and productivity of cotton crop in Ballari district

Based on the results of p-value, adj R^2 value and RMSE value presented in the Table2 best polynomial model was selected. For Ballari district, we observe that quartic model was found to be the best fit for area (Fig.1a), under cotton with an RMSE of 12.848 and adjR^2 value of 0.813. On the other hand, for production (Fig.1b), quartic model was the best fit with an RMSE of 18.653 and adjR^2value of 0.869while for productivity (Fig.1c), cubic model was the best fit with an RMSE of 50.548 and adj R^2value of 0.781. Thus, for the overall study period from 1970-71 to 2015-16, for Ballari district it was observed that quartic regression model was the best fit with an increasing trend for area and production of cotton. While, for productivity of cotton, cubic model was found to be the best fitted model for Ballari district. Further, area, production and productivity showed increasing trends.

Time series models for forecasting the production of cotton crop in Ballari district

ARIMA model

SPSS 20.0 statistical package was used to fit ARIMA model.The first step in the analysis was to plot the given data. Fig. 2shows the plot of production of cotton crop from 1970-71 to 2015-16 for Ballari district. An examination of Fig.3 revealed a positive trend over time which indicates the non-stationary nature of the series. This was confirmed, through the Auto Correlation Function (ACF) Partial Autocorrelation Function (PACF).ACF of the time series in Fig.4 shows a slow linear decay of the autocorrelation coefficients. Fig.4 represents the PACF plot which showed significance at lags 1. This indicates the non-stationarity of the series. To make the series stationary, it was first differenced after which the data attained stationarity as shown in Fig.4.

Identification of the model

ARIMA (2, 1, 2) model was fitted best based on the Autocorrelation function and Partial autocorrelation function of the differenced series as shown in Fig.4. as all the lagged values remained within the limit for both ACF and PACF plots. The outliers were detected and removed to get a significant model. Based on R^2, RMSE and MAPE values, it was observed that ARIMA (2, 1, 2) was found to be the best fit after eliminating three significant outliers. Three significant outliers were detected and removed as depicted in Table3. The estimates of the parameters are given in Table2. The adequacy of the model was also appraised based on the values of Ljung-Box Q statistics as shown in Table 4 which are found to be non-significant. The R^2, RMSE and MAPE values for ARIMA (2, 1, 2) model are given in Table 5.Residual analysis was carried out to check the adequacy of the model. The residuals of ACF and PACF were obtained from the tentatively identified model. All the lags were found to be non-significant as shown in Fig.6. Hence, it was inferred that ARIMA (2, 1, 2) model was adequate for forecasting production in Ballari district.

Forecasting accuracy and forecasting

The forecasting adequacy was checked using the RMSE and MAPE values. The predicted values using ARIMA with the model fit statistics like RMSE and MAPE values are given in Table 4.15. ARIMA (2, 1, 2) model was the best fit with least values of RMSE (9.117) and MAPE (13.660) as given in Table 4.15.Forecasting was done for the next four years using the ARIMA (2, 1, 2) model. The forecasted values are shown in Table 6. ARIMA (2,1,2) was selected as a model for forecasting in ARIMA technique after analyzing the ACF and PACF plots given in Fig.4. Thus, ARIMA (2, 1, 2) model was observed be the best fit with a R^2 value of 97.70 per cent. Hamjah (2014).have also obtained ARIMA (2, 1, 2) model for rice production in Bangladesh.

CONCLUSION

A prudent attempt has been made in the present study to understand the growth rate, trend in area, production and productivity and also forecasting technique ARIMA for production of cotton crop in Ballari district of Karnataka., The data was obtained from Directorate of Economics and Statistics, Bengaluru for the period from 1970-71 to 2015-16. The results revealed that area, production and productivity of cotton crop marked a significant increase in growth rate during the study period. In case of Ballari district, increasing growth rate in area, production and productivity was observed. The trend equations were fitted to the area, production and productivity of cotton crop and best fitted model was chosen for the purpose of future prediction. It is clear from the analysis that there is an increasing trend in area, production and productivity of cotton crop for Ballari. Based on adjR^2 and RMSE values it is evident that the quartic model was sought to be best fit for area under cotton for productivity, cubic model is found to be the best fit while for production, quartic model is said to be the best fit for Ballari district. Forecasting is also done using the best fitted ARIMA model. Due to autocorrelation in the data, time series forecasting model such as ARIMA were adopted. Forecasting have been done for next four years were carried out using a short term forecasting technique ARIMA for production of cotton crop. ARIMA (2, 1, 2) was the best model for forecasting production of cotton in Ballari district.

REFERENCES

Gudeta, B. and Egziabher, A.G.,2019, Cotton production potential areas, production trends, research status, gaps and future directions of cotton improvement in Ethiopia.*Greener J. Agric. Sci*., 9(2): 163-170.

Kulkarni, K.P., Jadhav, M.C. and Sharief, Z., 2017, Trend in cotton production in Nanded District of Maharashtra. *Trends in Biosciences.*, 10(32):6846-6848.

Mal, M. and Pandey, A., 2013, Instability and relative growth trend analysis of area, production and productivity of cotton crop in India.*Int. J. Agric. Econ. and Management.*, 3 (2):35-42.

Mayilsami, K. and Selvaraj, A., 2016. Growth of cotton cultivation: A study in Tamil Nadu. *Imperial J. Interdisciplinary Res.*, 2(11): 1-5.

Mohammad, D., Shiyani, R.L. and Ardeshna, N.J., 2018, Growth dimensions of long staple cotton area, production and yield in Gujarat, India.*Int. J. Curr. Microbiol. App. Sci.*, 5(7): 2993-3005**.**

Parmar, R.S., .Rajarathinam, A., Patel, H.K. and Patel, K.V., 2016,Statistical modeling on area, production and productivity of cotton *(Gossypium spp.)* crop for Ahmedabad Region of Gujarat State. *J. pure and app. microbiol.*,10(1):751-759.

Pavithra, N.L., Ashalatha, K.V., Megha, J., Manjunath, G.R. and Hanabar, S., 2018, Growth in area, production and productivity of food grains in Karnataka State, India.*Int. J. Curr. Microbiol. App. Sci .*, 7(8): 2532-2535.

Rajan, S.M. and Palanivel, M., 2017, Forecasting and growth model of cotton in Tamil Nadu State.*Asian J. Agric. Extension, Econ. &Sociology., 17(1): 1-5.*

Rajan,S.M. and Palanivel, M., 2018, Application of regression models for area, production and productivity growth trends of cotton crop in India.*Int. J. Stat. Distributions* &*Applications*. 4(1): 1-5.

Ramakrishna, G. and Bhave, M.H.V., 2017, Temporal variations in area, production and productivity of turmeric crop in India, *Int. J. Res. in Business Management.,* 5(7): 83-90.

Samuel, J., Basavaraja, H., Puspanjali and Rejani, R., 2013, Trends in area, production and productivity of cotton across the major states in India, *Int. J. Humanities, Arts, Med.* & *Sci.*,1(2): 97-102.

Sharma, A., 2015, Growth and variability in area, production and yield of cotton crop.*Int. J. Agric. Innovations* & *Res.,* 4 (3): 2319-1473.

Panasa,V., 2014,A study of temporal variations in area, production and productivity of cotton crop in three regions of Andhra Pradesh.*M.Sc. (Agri.) Thesis (Unpub.)*,*ANGRAU*Rajendranagar,Hyderabad.

Singh,M. and Supriya, K., 2017,Growth rate and trend analysis of wheat crop in Uttar Pradesh, India**.***Int. J. Curr. Microbiol. App. Sci.*,*6*(7): 2295-2301.Wali, V.B., Devendra, B. andLokesh H., 2017,Forecasting of Area and Production of Cotton in India: An Application of ARIMA Model. *Int. J. Pure App. Biosci.*, 5(5)**:** 341-347.

Wold,H.1938, A study in the analysis of stationary time series. Stockholm: Almqvist and Wiksell.

Yule, G.U., 1926, Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series. *Journal of the Royal Statistical Society.*, 89: 1–64.

Ali, S., Badar, N. and Fatima, H., 2017, Forecasting Production and Yield of Sugarcane and Cotton Crops of Pakistan for 2013-2030. *Sarhad Journal of Agriculture.*,31(1): 1-10.

Debnath, M. K., Bera, K. and Mishra., 2015, Forecasting area, production and yield of cotton in India using ARIMA Model. *Res & Rev: J. Space Sci. & Tech.*, 2(1): 16-20.

Dickey, D.A.and Fuller, W. A., 1979, Distribution of estimators for autoregressive time series with a unit root. *J. American Statistical Association.*, 74: 427-431.

Hamjah, M. A., 2014, Rice production forecasting in Bangladesh- An application of Box-Jenkins ARIMA model. *Mathematical Theory Modelling.*,4(4): 1-11.

Iqbal, N., Bakhsh, K., Maqbool, A. and Ahamad, A. S., 2005, Use of the ARIMA model for forecasting wheat area and production in Pakistan. *J.Agric. Soc. Sci.*, 1(2): 120-12.