International Journal of Education and Science Research Review
ISSN 2348-6457
www.ijesrr.org June- 2017, Volume-4, Issue-3 Email- editor@ijesrr.org

A TIME SERIES ANALYSIS OF COTTON (*GOSSYPIUM HIRSUTUM)* IN TAMIL NADU STATE INDIA

MYLSAMY SUNDAR RAJAN
Department of Statistics
Government Arts College Udumalpet
Tamil Naidu

MALAYAPPA PALANIVEL
Department of Statistics
Government Arts College Coimbatore
Tamil Naidu

ABSTRACT
This study aims at presenting models for the forecasting time-series data of cotton (*Gossypium hirsutum)* Area, Production and Productivity of Tamil Nadu by using Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) models of time-series forecasting. The cotton data for (1970-2013) area, production and productivity were forecasted to five years staring from 2014-2018. Presence of trend in data was checked through time series stationary through auto correlation and partial auto correlation functions ARIMA (1,1,2), (2,1,2) and (1,1,3) model performed better among other models of ARIMA family for modeling as well as forecasting purpose.
KEYWORDS: Forecasting, ACF, PACF, Time series, ARIMA, Cotton.

INTRODUCTION
Cotton is an important cash crop and India cultivates the highest acreage in the world. It provides the basic raw material (cotton fibre) to cotton textile industry. Known also as 'White Gold' Cotton enjoys a predominant position amongst all cash in India. The major cotton growing states are Punjab, Haryana, Rajasthan, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Tamil Nadu and Karnataka. With an objective to improve the quality of cotton, enhance/ha productivity, enhance the income of cotton growers by reducing the cost of cultivation, to improve the processing facilities etc. Cotton is at present 2012-13 cultivated in the Tamil Nadu state in 1.28 lakh areas, with a production of 3.8 lakhs bales. India has the largest area under cotton cultivation in the world though she is the world's third largest producer of cotton after China and the USA. India, with an estimated annual production of 29 million bales during 2008-09(1 bale=17o kg), is the second largest cotton producers in the world.

OBJECTIVE OF THE STUDY
The main objective of this study is to develop an ARIMA model for forecasting of Cotton Crops area, production and productivity in the Tamil Nadu State, India.

METHODOLOGY
The study was conducted by using time series data of Cotton area, production and productivity during the years 1970-2013 of Tamil Nadu State, India. The data was collected from the various issues of Cotton Advisory Board, published by Government of India. Data was analyzed by using SPSS software. ARIMA model as proposed by Box and Jenkins (1970) was applied to forecast Cotton area, production and productivity for the period 1970-2013. The ARMA (Autoregressive Moving Average) method is used in this study is because the characteristic of each cascading is stationary (has a mean and constant variance also covariance lag that does not depend on where the calculation is done). The Box-Jenkins procedure is concerned with fitting a mixed ARIMA model to a given set of data. The results showed that the prediction is quite accurate using method of ARIMA in the study area, production and productivity.

TECHNIQUES OF PREDICTION USING ARIMA MODEL

ARIMA time series models traditionally expressed as ARIMA (p,d,q) combine as many as three types of processes, viz.

AR: p = order of the autoregressive part

I: d = degree of first differencing involved

MA: q = order of the moving average part

The form of the ARIMA (p,d,q) respectively,

- A p^{th} - order autoregressive model: AR(p), which has the general form: $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$

  Where, Y_t = Response (dependent) variable at time t

  $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ =Response variable at time

  Lags $t-1, t-2, \ldots, t-p$, respectively

  $\varphi_0, \varphi_1, \varphi_{2,,,} \varphi_p$ =Coefficient to be estimated

  ε_t = Error term at time t.
- A q^{th} - order moving average model: MA(q), which has the general form: $Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$

  Where, Y_t = Response (dependent) variable at time t

  μ = Constant mean of the process

  $\theta_1, \theta_2, \ldots, \theta_q$ = Coefficients to be estimated

  ε_t = Error term at time t

  $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \epsilon_{t-q}$ = Error in previous time

  Periods that is incorporated in the response Y_t
- Finally, the general form of ARIMA model of order (p,d,q) has the representation

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \mu + \epsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \varepsilon_t$$

Where Y_t is the cotton yield, ε_t's are independently and normally distributed with zero mean and constant variance δ^2 for $t = 1, 2, \ldots, n$ and d is the fraction differenced while interpreting AR and MA and φ_s and θ_s are the coefficients to be estimated.

TREND FITTING

For evaluating the adequacy of AR, MA and ARIMA processes, maximum value of coefficient of determination (R^2), Among the competitive models, best models are selected based on minimum value of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Normalized Bayesian Information Criteria (BIC) as suggested by Schwartz, and Q statistics have been used. Best fitted models are put under diagnostic checks through Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) of the residuals. The lower the values of above statistics, the better are the models which are given by,

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right]^{1/2} \qquad MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(Y_i - \hat{Y}_i)}{Y_i}\right|$$

$$MAE = \frac{\sum_{i=1}^{n}\left|X_i - \hat{X}_i\right|}{n} \qquad R^2 = \frac{\sum_{i=1}^{n}\left(\hat{X}_i - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

$$BIC(p,q) = In v^*(p,q) + (p,q)[In(n)/n]$$

Where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v^* is the estimate of white noise variance σ^2.

The significance level of individual coefficients is measured by Box-Jenkins Q-statistics were computed Fig1,

$$Q^* = n(n+2)\sum_{k=1}^{h}(n-k)^{-1}r_k^2$$

Where n is the number of residuals, h is the maximum lag and rk is the residuals autocorrelation at lag^k.

Box and Jenkins Methodology for time series modeling steps:

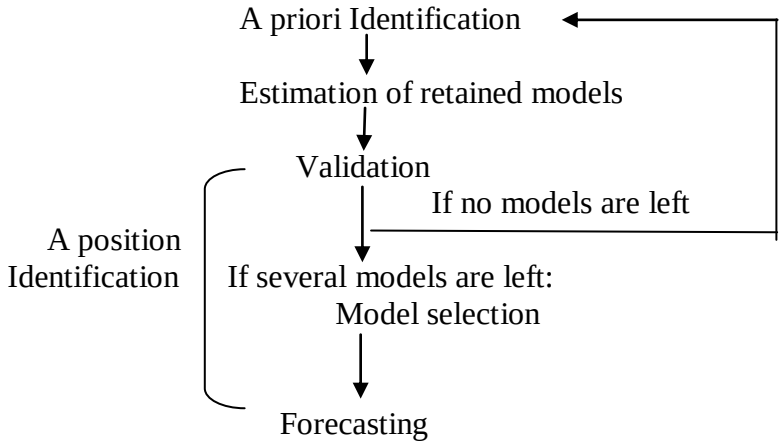

Fig 1. Schematic representation of the Box - Jenkins Methodology for time series modeling.

The Box-Jenkins Methodology is regarded as the most efficient forecasting technique, used in analysis and forecasting and is used widely and extensively specially for univariate time series modeling. The modeling of the time series was applied according to four stages namely:

Model identification: Specify the order of AR and MA components were determined.
Model estimation: Linear model coefficients were estimated.
Model validation: certain diagnostic methods were used to test the suitability of the estimated model.
Forecasting: If the model passes the validation step, then it can be used best model chosen for forecasting.
ARIMA methodology may be precisely visualized from Figure 1.

RESULT AND DISCUSSION

In the present study the data for Cotton cultivated area, production and productivity for the period 1970-2013 were used following the four stages of ARIMA model.

MODEL IDENTIFICATION

ARIMA model is designed after assessing which varies under forecasting as a stationary series. The stationary series is the set of values varied over time around a constant mean and constant variance. The results indicate that area , production and productivity of cotton crop are stationary at level. i.e. The series are integrated of zero I (0). So it is not needed to make these series stationary by taking difference. Similarly, the area, production and productivity series of cotton was stationary at level. In this stationary data the various Autocorrelations up to 12 lags were computed and the same along with their significance which is tested by Box-Ljung test are shown in Table 1 (a, b and c) with partial auto correlation shown Table 2. And also from the figures (Fig.2, 3 and 4) of ACF and PACF are the tentative ARIMA models are chosen has minimum normalized BIC (Bayesian Information Criterion).

MODEL ESTIMATION AND VERIFICATION

The model parameters estimated are presented in Table 3(a, b and c).The stationary is the data by ARIMA(1,1,2), (2,1,2) and (1,1,3) models were estimated using SPSS computer package and estimation of the models for the cotton area, production and productivity data as this model had statistically significant coefficient, the lowest Normalized BIC, Good R^2 and model fit statistics for Root Mean Square Error and Mean Absolute Percentage Error.

DIAGNOSTIC CHECKING

The model verification is concerned with checking the residuals of the model to see if they contained to improve the selected ARIMA. For this purpose, we used ACF and PACF of the plotted residuals of area, production and productivity were found to well fit models and are provided in Figs 5. In this proved that the selected ARIMA model was an appropriate model for forecasting cotton crops.

Table 1. Auto Correlations for Cotton Area, Production and Productivity

a) Area

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.820	.146	31.639	1	.000
2	.659	.144	52.536	2	.000
3	.606	.142	70.687	3	.000
4	.529	.141	84.854	4	.000
5	.439	.139	94.862	5	.000
6	.421	.137	104.313	6	.000
7	.378	.135	112.140	7	.000
8	.219	.133	114.825	8	.000
9	.093	.132	115.330	9	.000
10	.033	.130	115.394	10	.000
11	-.021	.128	115.420	11	.000
12	-.102	.126	116.081	12	.000

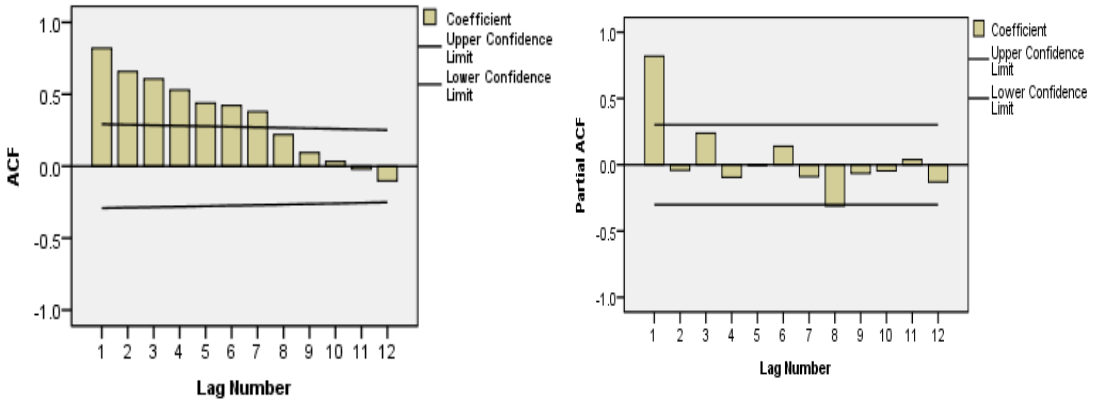

Fig 2. a) ACF and PACF of fitted Cotton Area

b) Production

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.534	.151	13.412	1	.000
2	.166	.189	14.739	2	.001
3	.152	.192	15.875	3	.001
4	.176	.195	17.437	4	.002
5	.187	.198	19.243	5	.002
6	.299	.202	24.012	6	.001
7	.251	.212	27.464	7	.000
8	.028	.219	27.509	8	.001
9	.019	.219	27.531	9	.001
10	.133	.219	28.592	10	.001
11	.056	.221	28.786	11	.002
12	-.020	.221	28.811	12	.004

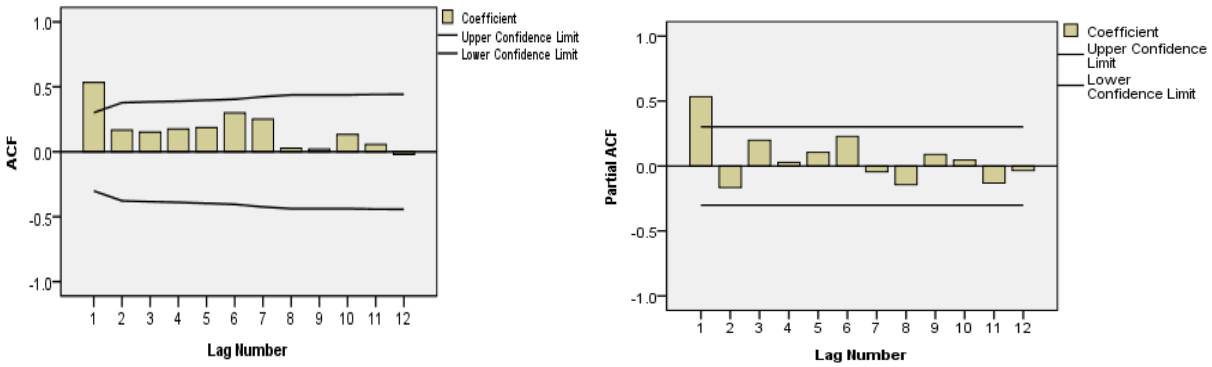

Fig 3. ACF and PACF of fitted Cotton Production

c) Productivity

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.901	.151	38.232	1	.000
2	.845	.244	72.606	2	.000
3	.743	.303	99.824	3	.000
4	.672	.342	122.659	4	.000
5	.564	.371	139.145	5	.000
6	.514	.390	153.243	6	.000
7	.422	.405	162.979	7	.000
8	.337	.415	169.366	8	.000
9	.265	.421	173.424	9	.000
10	.195	.425	175.697	10	.000
11	.130	.427	176.739	11	.000
12	.076	.428	177.100	12	.000

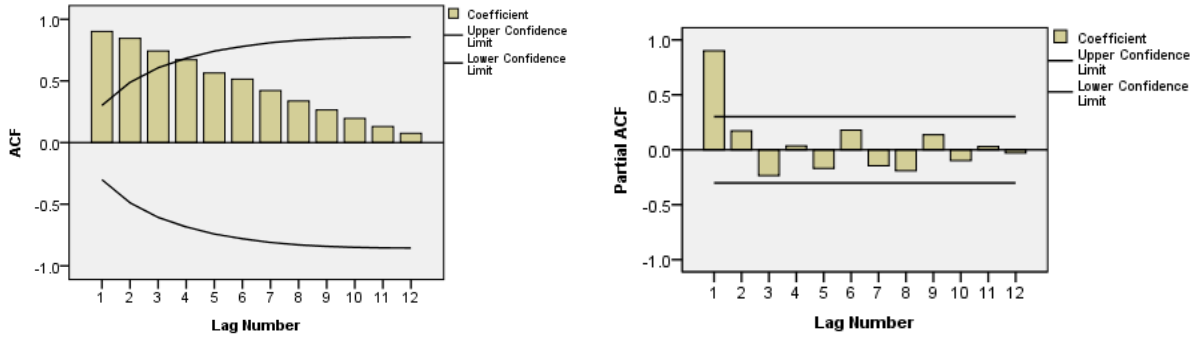

Fig 4. ACF and PACF of fitted Cotton Productivity

Table 2. Partial Auto Correlations for Cotton Area, Production and Productivity

Lag	Area		Production		Productivity	
	Partial Autocorrelation	Std. Error	Partial Autocorrelation	Std. Error	Partial Autocorrelation	Std. Error
1	.820	.151	.534	.151	.901	.151
2	-.042	.151	-.166	.151	.172	.151
3	.239	.151	.197	.151	-.235	.151
4	-.093	.151	.028	.151	.035	.151
5	-.003	.151	.105	.151	-.169	.151
6	.139	.151	.227	.151	.179	.151
7	-.090	.151	-.045	.151	-.146	.151
8	-.311	.151	-.144	.151	-.191	.151
9	-.065	.151	.089	.151	.136	.151
10	-.047	.151	.046	.151	-.098	.151
11	.038	.151	-.132	.151	.029	.151
12	-.130	.151	-.035	.151	-.028	.151

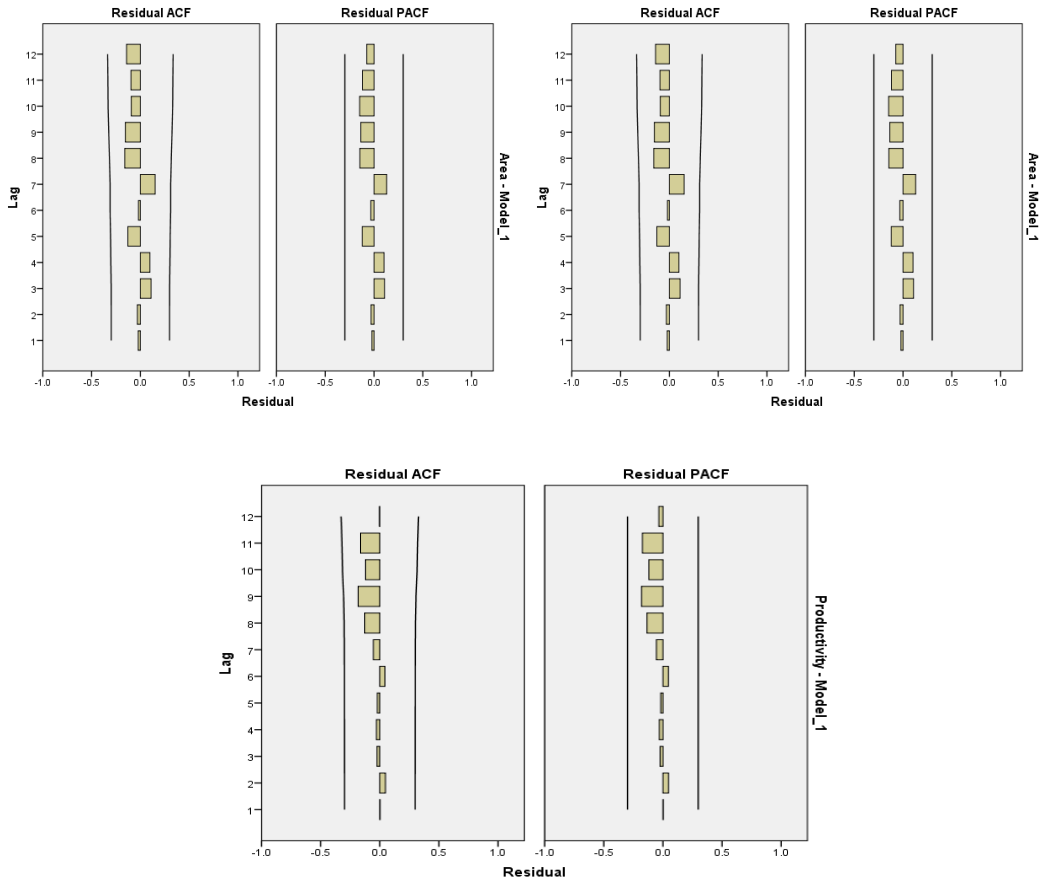

Fig 5. ACF and PACF of residuals of fitted ARIMA model for Cotton Area, Production and Productivity

FORECASTING COTTON AREA, PRODUCTION AND PRODUCTIVITY USING THE FITTED MODEL

The selecting the best models, now we are going to use these models to forecast different types of Cotton area, production and productivity in Tamil Nadu State. To forecast the following" Forecast Criteria" are considered which are shown in the Tables.

Table 3. Estimates of the fitted ARIMA model for Cotton Area, Production and Productivity

a) Area

	Forecasting Criteria for the best selected model					
Area	Selected Model	R-Square	RMSE	MAPE	MAE	Normalized BIC
	1,1,2	0.76	0.37	14.607	0.265	-1.566

b) Production

	Forecasting Criteria for the best selected model					
Production	Selected Model	R-Square	RMSE	MAPE	MAE	Normalized BIC
	2,1,2	0.42	0.91	17.931	0.677	.326

c) Productivity

	Forecasting Criteria for the best selected model					
Productivity	Selected Model	R-Square	RMSE	MAPE	MAE	Normalized BIC
	1,1,3	0.91	72.93	11.845	50.453	9.104

Table 3 (a,b and c) shows a goodness of the fit statistics for the first data set. R-Squared represents an estimated that is explained by the model. The maximum values indicated better fit. A value of 0.76, 0.42 and 0.91 means that model does an excellent cotton area, production and productivity of explaining the observed variations in the series. And here, best model (1,1,2), (2,1,2) and (1,1,3) ARIMA models can be combined to produce a great variety of models this selected model.

The ARIMA models developed in the present study are finally used to forecast the corresponding variables. The forecasting values thus obtained are the MAPE is 14.607, 17.931 and 11.845 which indicate that the accuracy level of forecasting is much higher this shows quite good level of forecasting accuracy.

Table 4. Fitted an ARIMA model for Cotton Area, Production and Productivity

a) Area

Summary statistic of the ARIMA model for Cotton Area					
Type		Estimate	SE	t	Sig.
	Constant	2.446	3.052	.802	.428
AR	Lag 1	.329	.270	1.219	.230
Difference		1			
MA	Lag 1	.581	7.568	.077	.939
	Lag 2	.418	3.108	.135	.894

b) Production

Summary statistic of the ARIMA model for Cotton Production					
Type		Estimate	SE	t	Sig.
	Constant	16.566	7.364	2.250	.031
AR	Lag 1	.023	.363	.064	.950
	Lag 2	-.093	.282	-.329	.744
Difference		1			
MA	Lag 1	.426	3.255	.131	.897
	Lag 2	.569	1.979	.287	.775

c) Productivity

Summary statistic of the ARIMA model for Cotton Productivity					
Type		Estimate	SE	t	Sig.
	Constant	-1543.426	501.583	-3.077	.004
AR	Lag 1	-.928	.134	-6.906	.000
Difference		1			
	Lag 1	-.418	26.143	-.016	.987
MA	Lag 2	.794	37.072	.021	.983
	Lag 3	.623	16.343	.038	.970

The model parameters estimated are presented in Table 4 (a, b and c).The stationary is the data by ARIMA (1,1,2), (2,1,2) and ARIMA(1,1,3) were found appropriate for area, production and productivity of cotton

respectively. The parameters can be said to be the selected estimated using SPSS computer package and estimation.

FORECASTING WITH ARIMA MODEL

The forecasting model ability of the fitted ARIMA model, the measures of the sample period forecast accuracy was computed. Table 5 shows the forecasted value for cotton area and production (with 95% confidence limit).

Table 5. Forecasted on Cotton Area, Production and Productivity based on best select model

Year	Area (Lakhs. ha)	Production (Lakh. Bales)	Productivity (Kg/ha)
2014	1.19	3.03	847
2015	0.96	3.51	880
2016	0.84	3.32	1011
2017	0.75	3.08	956
2018	0.67	2.90	1075

CONCLUSION

A time series mode accounts for patterns in the past movement of a variable and uses that information to predict its future movements. Time series data have become very popular to be intensively used in empirical research and econometricians have recently begun to pay very careful attention to such data. To select the best model for a particular time series that latest available model selection criteria are used.

This objective of this paper was to develop a model which could explain and predict changes in cotton area, production and productivity in the Tamil Nadu state. The prediction of future area, production and productivity of dates in Tamil Nadu state for a period of five years. Forecasting is done for the period of 2014 to 2018, Table 5 shows that the year's area, production and productivity of Tamil Nadu is showing a decreasing growth percentage could attain 1.19 lakhs/ha in 2014 with 0.67 lakhs/ha decreased year of 2018 and that the year's production could attain 3.03 lakhs bales in 2014 with 2.90 lakhs bales forecast is decreased for year of 2018. Further, the productivity 847 kg/ha for the year of 2014 with 1075 kg/ha forecasted is increased for the year of 2018. The finding of the study indicates that the years of area production and productivity in ARIMA model (1,1,2), (2,1,2) and (1,1,3) has been provided better results and upward trend for future.

[1]Tamil Nadu state lowest cotton producing states, CCI has not undertaken any commercial operations in the last 20 years. Tamil Nadu itself to increase the production, necessary steps to be taken to retain the existing area under cotton.

([1] The Financial Express,15 Dated.(2005).

REFERENCES

1. Box G.E.P. and Jenkins. G.M.(1970). Time series analysis: Forecasting and control. Holden day.,San Francisco.
2. Makridakis, Spyros G, Wheelwirght and Rob J.Hyndman.(1998). Forecasting: methods and applications,John Wiley & Sons, New York.
3. Bansal K.K,Sharma M.K.(2015): A Comparative Study of Reliability Analysis of a Non-Series Parallel Network, *International Journal of Education and Science Research ReviewVol.2(6)*