

Estimated trend in production of cotton for Ahmedabad district of Gujarat state by statistical models

N. J. RANKJA¹ AND S.M.UPADHYAY²

Department of Agricultural Statistics, College of Agriculture
Junagadh Agricultural University, Junagadh-362 001 (Gujarat)

Received : 17 July, 2017

ABSTRACT

An investigation was carried out using the polynomial (linear, quadratic and cubic) models were fitted on original data as well as three, four and five year moving averages data. The Autoregressive Integrated Moving Average (ARIMA) models were fitted to original time series data after checking the stationary condition, to arrive at a methodology that can precisely explain the fluctuation in production for cotton crop in Ahmedabad district of Gujarat for the period 1960-61 to 2014-15 (55 years) to compare different models. The error percentage for the selected model was also carried out to test the prediction power. The data from year 1960-61 to 2006-07 were used for model fitting and remaining year for testing the forecast. In polynomial models, the most suitable model was selected on the basis of R^2 , significant regression coefficient, root mean square error, mean absolute error, normality (Shapiro and Wilk test) and randomness of residual's (Run test) distribution. The different ARIMA models (p, d, q) were judged on the basis of autocorrelation function (ACF) and partial autocorrelation function (PACF) at various lags. Among different fitted ARIMA models, the final models were selected on the basis of significant autoregressive and moving average term, Akaike's Information Criterion (AIC), Schwartz-Bayesian Criterion (SBC) and normality (Shapiro-Wilk test) and randomness of residual's (Run test) distribution. Suitable model for cotton production was in Ahmedabad district was ARIMA (1, 1, 0) with the lowest error percentage of prediction was during 2014-15 with 2.602% and highest during 2010-11 with 42.036%.

Key words: Shapiro and Wilk test, run test, polynomial model, ARIMA, autocorrelation function.

Cotton is one of the oldest and the most important commercial and fiber crop of the world. Cotton is referred to as "King of Fibres" and also known as "White Gold". It is heat loving and sun loving (heliophyte) plant. A daily minimum temperature of 16°C is required for germination and 21°C to 27°C for proper vegetative growth. It can tolerate temperature as high as 43°C, but does not do well if the temperature falls below 21°C. During fruiting phase, the day temperature ranging from 27°C to 30°C and cool nights are needed. Abundant sunshine during the crop growth period particularly the period of boll maturation and harvesting is essential to obtain a good quality produce. Successfully grown in areas receiving an average annual rainfall ranging from 500mm, of which 175-200 mm should be received during crop growth period. Cotton- Quality parameters-and yield attributes are

categorised into (1) Colour of cotton fibre is white with few exceptions like desi cotton which have reddish or yellowish tinge. White coloured cotton which is shiny is considered as superior cotton. Cotton is grown in the nine major states in three different zones. Punjab, Haryana and Rajasthan in north zone; Maharashtra, Gujarat and Madhya Pradesh in central zone and Andhra Pradesh, Karnataka and Tamil Nadu in the south zone are the major cotton growing states. Central zone accounts for about 60 per cent of all cotton, and where only 16 per cent is irrigated. About 74 per cent of the total area and production of cotton in the country are contributed by four states of Gujarat, Andhra Pradesh, Maharashtra and Punjab. Cotton is mainly a rainfed crop in Gujarat. About 54 per cent of the state's production comes from Surendranagar, Rajkot, Vadodara, Ahmedabad and Sabarkantha districts. Other important producers include Mehsana, Bhavnagar, Bharuch, Kheda, Surat, Amreli and

¹Assistant Professor, ²Professor and Head

Panchmahals districts. The main long and medium stapled varieties include V-797, Digvijay, Sanjay, Kalyan, Dholleras, Deviraj, Wagad and Varalaxmi etc. Gujarat state accounts 17.10 lakh hectare area in 1960 and increases up to 30.6 lakh hectare area during year 2014-2015. Production of cotton crop increases 36.78 lakh bales to 125 lakh bale from year 1960 to 2014-2015. Productivity of cotton crop increases from 86 kg/hectare to 707 kg/hectare from year 1960 to 2014-2015. (Anonymus, 2015b). The observation on these time series variables are varying in some pattern due to the existence of artificial and natural forces. But it is, indeed, a hard truth that policy makers, while formulating policies have to venture into the future (predicted) values of these time series variables for formulating the strategic all round developmental plan for zone or a state or a country. In literature, a large number of univariate time series models are available viz., linear, polynomial, exponential, logistic etc. The model developed by Box and Jenkins during the seventies, popularly known as Autoregressive Integrated Moving Average (ARIMA) model, is often found to be superior among all above mentioned models in case of the univariate time series variables which are found correlated with their lag variables. The advantage of ARIMA modelling over the other univariate time series model is that besides displaying the intrinsic behaviour (generating process) in the time series variables itself, it produces the smallest mean squared forecast error variances. But as a mark of caution, it is noted that ARIMA models are useful only for forecasting the near future values destined to appear, usually, within the five years beyond the last datum observed on the time scale. Thus, in respect of given time series, the predicted values corresponding to the future time point considered in data set, cannot be relied upon. ARIMA model is an extrapolation method for forecasting and like any other such method, it requires only the historical time series data on the variables under forecasting. Among the extrapolation methods, this is one of the most sophisticated method, as it incorporates the future of all such methods, does not require the investigator to choose initial values of any variables and values of the various parameters a priori. It is robust to handle any data pattern. As one would expect this is quite a difficult model to develop and apply as it involves transformation of the variable, identification of the model, estimation through nonlinear method, verification of the model and derivation of the forecasts (Gupta, 1993).

The strength of ARIMA models lies in their ability to reveal complex structures of temporal interdependence in time series. It has also been shown that ARIMA models are highly efficient in short term forecasting (Fatimah and Gaffar, 1986).

Therefore, present investigation was undertaken to study the time series trend using data of cotton crop production Ahmedabad district by different statistical model. The data were obtained for Ahmedabad district during *kharif* session of 1960-60 to 2014-15 with following objectives.

- To study fitting of Polynomial models for cotton crop in Ahmedabad district.
- To study fitting of ARIMA model for cotton crop in Ahmedabad district.
- To compare different models according to their ability of prediction.

MATERIALS AND METHODS

Considering the production of cotton crop for the period 1960-61 to 2014-15 the time series data were obtained from Directorate of Agriculture, Gujarat state, Gandhinagar and WHO Cell, Department of Agricultural Economics, College of Agriculture, J.A.U. The data from 1960-61 to 2006-07 were used for model fitting. Data from 2007-2008 to 2014-15 were used for testing of forecast.

Fitting of Polynomial models

(1) Linear Regression Approach (Rangaswamy, 2006)
The following proposed model was fitted to original data as well as moving average data as:

$$Y = a + bt \quad \dots\dots\dots (1)$$

Where, response or dependent variable Y was production, time variable (t) as independent variable and a, b are the regression constant and regression coefficient, respectively.

(2) Quadratic Regression Approach (Montgomery *et al.*, 2003)

The fitted regression equation for quadratic regression model for each of proposed districts was given as

$$Y = a + bt + ct^2 \quad \dots\dots\dots (2)$$

The unknown parameter viz. a, b and c were estimated by using least square method.

(3) Cubic Regression Approach (Montgomery *et al.*, 2003)

As for consideration of the third degree polynomial model fitted to the data of Ahmedabad district was

as under

$$Y = a + bt + ct^2 + dt^3 \dots\dots\dots (3)$$

The constant a and coefficient b , c and d were estimated using least square method.

(4) Goodness of fit of the models (Montgomery *et al.*, 2003)

The most common and widely used test statistics for the goodness of fit of the fitted polynomial models is coefficient of determination R^2 and it was calculated as under

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad \text{and}$$

$$\text{Adj. } R^2 = 1 - \frac{(n-1)(1-R^2)}{(n-k)} \dots\dots\dots (4)$$

Where R^2 indicates the amount of variation in dependent variable accounted due to the model. For a given data set, it is well-known that R^2 is a non decreasing function of the number of parameters used. To overcome this weakness, an adjusted R^2 is proposed, since we interested in overall significance of the model. The F test was applied and the formula is given below.

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \dots\dots (5)$$

Which follows F distribution with $[k, (n-k-1)]$ degrees of freedom.

Where, n = number of observations

k = number of independent variables

The individual regression coefficients were tested using the t test under the null hypothesis, $\beta = 0$

$$t = \frac{b_i}{\text{S.E.}(b_i)} \dots\dots (6)$$

t value with $(n-k-1)$ degrees of freedom

Where b_i is estimated i^{th} coefficient and S.E. (b_i) is the standard error of b_i

For adequacy of fitted model, in addition to the above criteria, two more reliability statistics Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were computed (Liew *et al.*, 2000). Both were computed by using the following formulae.

$$\text{RMSE} = \left[\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / n \right]^{1/2} \dots\dots\dots (7)$$

$$\text{and MAE} = \sum_{i=1}^n |Y_i - \hat{Y}_i| / n \dots\dots\dots (8)$$

From the fitted models which had lower values of these estimates were considered better one.

According to Kvalseth (1985), before taking any final decision about the appropriateness of the fitted model, it is paramount importance to investigate the basic assumptions regarding the error term, viz., randomness and normality.

(5) Test for the randomness of the residuals (Sidney and Castellan, 1988)

For the randomness of residuals i.e. it is independently distributed or in other word sample must be random to arrive at any conclusion about the population by using the information in the sample.

Suppose m be the number of positive sign residuals, considering one group of element and n be the number of negative sign residuals, being second group of element in sequence of $N=m+n$. To use sample run test, first observe the m and n events in which they occurred and determined the value of r (i.e. no. of runs). If m or n is larger than 20, Z value needs to calculate as under.

$$\text{Mean} = \mu = \frac{2mn}{N} + 1 \text{ and standard deviation} =$$

$$\sigma_r = \left[\frac{2mn(2mn-N)}{N^2(N-1)} \right]^{1/2} \dots\dots(9)$$

$$Z = \frac{r - \mu_r}{\sigma_r} = \frac{r + h - \frac{2mn}{N} - 1}{\sqrt{\frac{2mn(2mn-N)}{N^2(N-1)}}} \dots\dots(10)$$

Where $h = +0.5$ if $r < \frac{2mn}{N} + 1$ and

$$h = -0.5 \text{ if } r > \frac{2mn}{N} + 1$$

Use normal table for testing Z value. The not significant Z value indicates randomness of the residual.

(6) Test for normality of the residual (Shapiro – Wilk, 1965)

To test for normality of the residual, The Shapiro - Wilk statistic is best suited among different test statistic to test the normality of residuals for time series data of above mentioned polynomial model and also in ARIMA model.

The required test statistics W was defined as

$$W = \frac{S^2}{b} \quad \text{Where} \quad S^2 = \sum a(k) [e(n+1-k) - e(k)] \dots (11)$$

The parameter k takes the value

$$k = \begin{cases} 1, 2, 3, 4, \dots, n/2 & \text{when } n \text{ is even} \\ 1, 2, 3, 4, \dots, (n-1)/2 & \text{when } n \text{ is odd} \end{cases}$$

$$\text{and } b = \sum_{i=1}^n (e_i - \bar{e})^2$$

The values of coefficients "a(k)" for different values of n and k are given (Shapiro - Wilk, 1965). When the calculated value of W is not-significant i.e. very close to unity, the null hypothesis regarding normality of residual was accepted.

Autoregressive (AR) and Moving Average (MA) models (Pankratz, 1983)

Talking on regression model, in ideal condition the parameters β 's are assumed to be constant over the time. In the forecasting models the errors ε_t 's within time period ($t = 1, 2, 3, \dots, n$) are assumed to be uncorrelated i.e. the observations Y_t 's are uncorrelated and thus it is an independent and identically distributed sequence. However, this assumption is rarely happen in practice and general condition. Usually serial correlations in the observations often exist in time-series data, which is indication of stationarity of time series data. The statistical concept of autocorrelation was used to measure the relationships between the value of Z at time t (i.e., Y_t) and Y at earlier time periods (i.e., $Y_{t-1}, Y_{t-2}, \dots$). The algebraic forms of Autoregressive (AR) and Moving average (MA) processes are:

(1) Autoregressive (AR) process :

This model is in the same form as the well-known simple linear regression model in which Y_t (Z_t) is the dependent variable and Y_{t-1} is the explanatory variable.

$$Z_t = C + \phi_1 Y_{t-1} + a_t \dots (12)$$

Where Z_t = time sequenced random variable

C = constant term related to mean (μ) such that

$$C = \mu(1 - \phi_1)$$

ϕ_1 = relationship of Y_t with Y_{t-1}

a_t = a random shock element at time t

A process involving past (time-lagged) Y terms is called an autoregressive (abbreviated as AR) process and the longest time lag associated with Y term is called the AR order of the process. The above equation is thus an AR process of order one, abbreviated as AR (1). Here it suffices to note that an AR (1) model implies that, conditional on the past return Y_{t-1} .

(2) Moving average (MA) process :

There are several ways to introduce MA models. One approach is to treat the model as a simple extension of white noise series. Another approach is to treat the model as an infinite-order AR model with some parameter constraints. We adopt the second approach. The general equation for moving average (MA) model was given as,

$$Z_t = c - \theta_1 a_{t-1} + a_t \dots (13)$$

Where C = constant term related to mean μ and

θ = relation of a_t with a_{t-1}

The process with past (time-lagged) random shocks only are called moving average (abbreviated as MA) processes and the longest time lag associated with an error term (i.e. a_t) is called order of MA. The above equation is an MA process of order one, abbreviated as MA (1). For a pure MA model $C = \mu$. The negative sign attached to θ_1 is merely a convention. It makes no difference whether to use negative or positive sign. The formula for calculating autocorrelation coefficient is :

$$\hat{\phi}_{11} = r_1$$

$$\hat{\phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j} \quad \text{Where } k = 1, 2, 3 \dots (14)$$

$$\text{and } \hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j}$$

$$(k = 3, 4, 5, \dots; j = 1, 2, 3, \dots, k-1)$$

$\hat{\phi}_{kk}$ = the estimate of the true partial autocorrelation coefficient t ϕ_{kk}

r_k = autocorrelation coefficient k lag apart

$\hat{\phi}_{kj}$ = estimate of partial autocorrelation of coefficient
t for lags apart when the effect of the
intervening lags has been removed

As long as data series is stationary the following set of recursive equation gives fairly good estimates of the partial autocorrelation

Fitting of Box-Jenkins ARIMA Models :

The AR or MA models discussed in the previous sections become cumbersome because one may need a high-order model with many parameters to adequately describe the dynamic structure of the data. To overcome this difficulty, the autoregressive moving-average (ARMA) models was introduced, most widely known as Box-Jenkins time-series models i.e. ARIMA (p, d, q) is known as "Univariate Box-Jenkins technique" (Box and Jenkins, 1976). An ARIMA model is an algebraic statement telling how observations on a variable are statistically related to past observation. To know whether data follow purely AR process or purely MA process or an ARMA process or an ARIMA process, BJ methodology comes in handy to answering preceding question.

This model includes three types of process, viz., Autoregressive of order p; differencing to make a series stationary of degree d and moving average of order q. This method applied only to a stationary time series data. When the data is non-stationary then it has to be brought into stationary by the method of differencing.

(1) Test for Stationary

The foundation of time series analysis is stationarity. A time series $\{Y_t\}$ is said to be *strictly stationary* if the joint distribution of $(Y_{t_1}, \dots, Y_{t_k})$ is identical to that of $(Y_{t_1+k}, \dots, Y_{t_k+k})$ for all t, where k is an arbitrary positive integer and $(t_1, \dots, t_k)$ is a collection of k positive integers. The Stationarity requirement ensures that one can obtain useful estimates of the mean, variance and ACF from a sample. The Stationarity condition of a series was tested by examining the:

1. Change of mean and variance over time.
2. By the coefficients of AR and MA process i.e. in case of AR (1) and MA (1) process it should be $|\phi_1| < 1$ and $|\theta| < 1$.
3. Estimated ACF values which should be tails-off towards zero rapidly.

The significance of autocorrelation was tested by t-test. The standard error of autocorrelation (Bartlett, 1946) was calculated as under

$$s(r_k) = \left(1 + 2 \sum_{j=1}^{k-1} r_j^2 \right)^{1/2} n^{-1/2} \dots\dots\dots(15)$$

$$t_{r_k} = \frac{r_k - \rho_k}{S(r_k)} \dots\dots\dots(16)$$

k=1, 2, 3...

The significant value of "t" indicates the presence of autocorrelation. The process of time series modelling involves transformation of data in order to achieve Stationarity, followed by identification of appropriate models, estimation of parameters, validation of models and finally for prediction. The complete description of these process and steps of time series modelling is clearly explained below.

RESULTS AND DISCUSSION

According to objective of the present research work, polynomial (Linear, Quadratic and Cubic) and ARIMA models were subjected to fitted on cotton production of Ahmedabad district of the Gujarat. The data were recorded for the year 1960-61 to 2014-15 from the published reports by the Directorate of Agriculture, Gujarat State, Gandhinagar and WHO cell, Department of Agricultural Economics J.A.U., Junagadh. The data from 1960-61 to 2006-07 were used for model fitting and remaining data from 2007-08 to 2014-15 were used for testing the forecast value of best fitted model among different selected model.

The different selected polynomial models (Linear, Quadratic and Cubic) were fitted to the original data as well as also on three years, four years and five years moving average data while Autoregressive Integrated Moving Average (ARIMA) models were fitted to the original data for cotton production of Ahmedabad districts. In fitting of Univariate Box-Jenkins (UBJ) Autoregressive Integrated Moving Average (ARIMA) models, the autocorrelation up to 12 lags were worked out for production of the Ahmedabad district. If the spikes did not sharply tails-off towards zero and if the visual inspection of the realization indicates that the mean, variance and autocorrelation were not constant over time then the series was considered as non-

Table 1 Fitted polynomial models for cotton production in Ahmedabad district

Model	Moving Average	Regression				Adj. R ²	RMSE	MAE	S-W Test	Run test(Z)
		constant	Regression coefficients							
		a	b	c	d					
Linear	Original	1586.960**	2.211	-	-	-2.0	647.264	471.996	0.953	2.653
	3 year	1568.424**	1.914	-	-	-2.1	484.179	362.200	0.974	3.581**
	4 year	1577.679**	0.664	-	-	-2.3	417.908	320.933	0.980	4.098**
	5 year	1494.223**	-5.725	-	-	-0.98	376.741	302.060	0.986	4.825**
Quadratic	Original	2372.833**	-94.015**	2.005**	-	22.7**	557.000	382.842	0.926	1.472
	3 year	2302.287**	-91.771**	2.003**	-	37.5**	374.393	280.774	0.955	5.029**
	4 year	2246.536**	-85.679**	1.929**	-	42.0*	310.780	243.182	0.966	5.025**
	5 year	2265.588**	-92.747*	2.141	-	58.6**	264.922	219.808	0.974	4.583**
Cubic	Original	2048.168**	-16.800	-1.971	0.055	23.9**	546.430	386.964	0.939*	0.587
	3 year	1931.638**	-0.080	-2.892	0.071*	42.0**	359.753	265.119	0.960	3.605**
	4 year	1923.715**	-5.039	-2.541	0.066*	47.5**	291.983	228.796	0.958	3.445**
	5 year	1922.868**	-7.965	2.417	0.066**	65.2**	247.005	202.079	0.966	3.618**

* Significant at 5% level, ** Significant at 1% level

Table 2 Fitted ARIMA models for cotton production in Ahmedabad district

ARIMA	AIC	SBC	AR(ϕ)	MA(θ)	CONS	RMSE	SW-TEST	BLQ-TEST
(1,1,0)	352.545	262.408	-0.303*	-	2.893	674.834	0.957	20.946
(1,1,1)	349.265	260.591	0.331	0.857**	4.412	625.967	0.944*	11.554
(2,1,0)	353.292	264.607	-0.350*, -0.141	-	5.700	675.591	0.946*	20.065
(3,1,0)	352.172	265.105	-0.391*, -0.251, -0.313*	-	9.492	648.438	0.943*	10.890

* Significant at 5% level, ** Significant at 1% level

stationary. Therefore, the new variable X_t was constructed by taking difference of one (*i.e.* $d = 1$) to make the series stationary.

Fitting of trend on cotton production in Ahmedabad district:

(1) *Fitting of polynomial models:*

The results of fitted polynomial models are given in Table 1. The results indicated that all coefficients of determination ($\text{Adj. } R^2$) were significant in quadratic and cubic models and it was maximum (65.20) in cubic regression of five year moving average approach. All coefficients of determination ($\text{Adj. } R^2$) were negative in linear model for all data approaches. The regression constant values were found significant in all polynomial models of all approaches. None of linear regression coefficients were significant in linear and third degree polynomial models of all data approaches. The quadratic regression coefficients were significant in all approaches of second degree polynomial models. All linear regression coefficients of quadratic and cubic model were negative were as it was positive for linear model (except for five year moving data approach). In original data approach the value of $\text{Adj. } R^2$ were increased by 44.70 per cent in quadratic regression as compared to linear regression while, in case of cubic regression, it was 1.20 per cent over the quadratic model. By taking five years moving averages R^2 value was observed to be increased by 1.92, 30.90 and 41.30 percent in case of first, second and third degree polynomial models, respectively over original data.

(2) *Fitting of ARIMA models:*

As series was found non stationary, the new variable X_t was constructed by taking differences of one (*i.e.* $d=1$) to make the series stationary.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) of various order of X_t were computed to identify the value of p and q . The ACF (γ_k) of the transformed variable was damping towards zero with cut off second and third's spike and the PACF (ϕ_{kk}) also cut-off at first lags. This suggested that the algebraic family of ARIMA on $p=2$ and 3 $d=1$ and $q=1$ can be used. The different models among the different value of p and q were fitted. Among the models, those models having lower value of AIC and SBC are given in Table 2. ARIMA (1, 1, 0) model had comparatively lower value of AIC and SBC with significant AR (ϕ) coefficient. The assumptions of residuals *i.e.* normality and independence of residuals were tested by Shapiro- Wilk test and Box-Ljung (Q) test indicated that all ARIMA models did not satisfied the assumption of normality (except ARMA(1,1,0)) while all ARMA model satisfied the assumption of independent residuals.

Thus, for the production of cotton in Ahmedabad district none of the polynomial models were found satisfactory to explain the trend but ARIMA (1, 1, 0) was found satisfactory to explain the trend of cotton production in Ahmedabad district. Balanagammal *et al.* (2000) and found ARIMA (1,1,0) model which fulfilled all diagnostic criteria for prediction of cotton production in Tamilnadu Thus these results are similar to present investigation.

Table 3 Testing of forecast values for remaining eight years by using selected models *i.e.* ARIMA (1, 1, 0) of cotton production in Ahmedabad district.

Years	Observed Values	ARIMA (1,1,0)	
		Predicted values	Error Per cent
2007-08	2569	2466.281	3.998
2008-09	2682	2547.689	5.007
2009-10	3613	2744.557	24.036
2010-11	5920	3395.542	42.036
2011-12	3916	5223.151	-33.593
2012-13	5706	4703.492	17.573
2013-14	6123	5194.492	15.164
2014-15	6244	6081.493	2.602

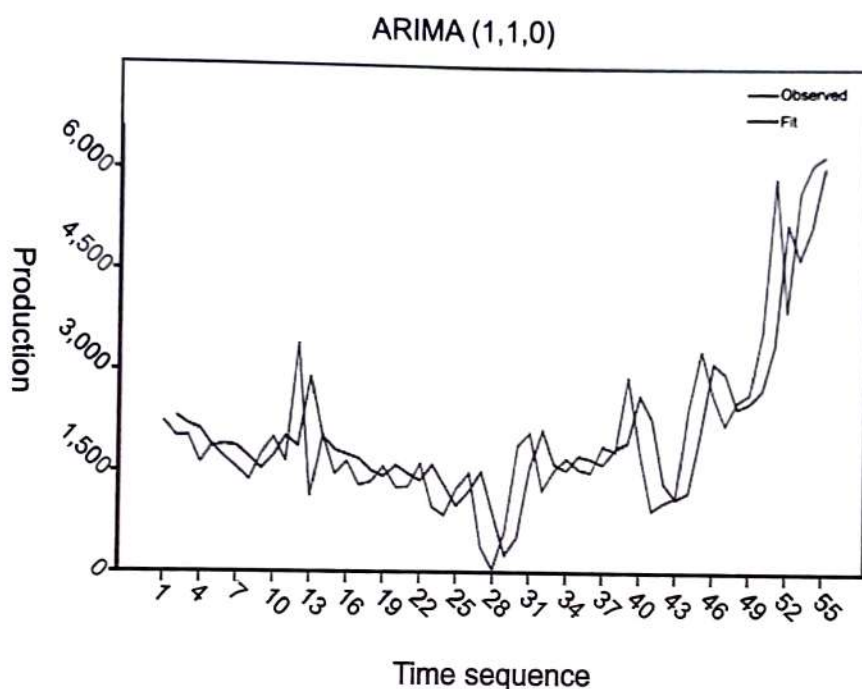


Fig. 1. Trend in cotton production based on ARIMA (1, 1, 0) models in Ahmedabad district

By using ARIMA (1, 1, 0), Predicted values and their error percentage are given in Table.3. The ARIMA (1, 1, 0) showed that during 2009-10 and 2010-11 it was over estimated and during 2011-12, it was underestimated while rest of year ARIMA (1, 1, 0) had good prediction for cotton production in Ahmedabad district.

CONCLUSION

The quadratic model on original data series from polynomial models was found suitable for fitting trend on production of Ahmedabad district. The production showed significant negative linear and positive quadratic trend. The fitted quadratic model for the production is

$$\hat{Y}_t = 2372.833 - 94.015t + 2.005t^2 \quad (\text{Adj. } R^2 = 22.7\%)$$

ARIMA (1, 1, 0) model fulfilled all the statistical requirement for being selected. The equation is

$$\hat{Y}_t = 2.892 - 0.303Y_{t-1} + a_t$$

The lowest error percentage of prediction was during 2014-15 with 2.602% and highest during 2010-11 with 42.036%.

It is suggested that the model based on original data with reasonably good R^2 can be used for future

prediction while model based on moving averages can be used to predict average trend value. ARIMA models would be suitable only for short term prediction and hence need to be updated every year for achieving valid forecast.

REFERENCES

- Anonymous, 2015 a The Pocket book on Agricultural Statistics 2015, Directorate of Economics and Statistics, New Delhi.
- Anonymous, 2015 b Current Cotton Scenario - The Cotton Corporation of India, Ltd. <http://cotcorp.gov.in/current-cotton>. assessed on 17 November, 2014.
- Anwar, S., Javed, M. S. and Mahmood, K. 2010. Prospect of cotton export from Pakistan up to the year 2010. *Pakistan Journal of Agricultural Science*. **33** : 996.
- Chaudhari, D.J., Chaudhari C.J., and Tingre, A.S. 2014. Forecasting cotton (*Gossypium* spp) production in India. *Journal of Cotton Research* **28** (2) 342-345
- Debnath, M. K., Bera, K. and Mishra, P. 2013. Forecasting area, production and yield of cotton

- in India using ARIMA Model. *Journal of Space Science & Technology*, 2(1): 16-20
- Gomes, E.M., Celidonio, O.L.M., Ferrerira, D.L., Alves, L.G., Guides, J.P. and Anzenello, M.J. 2013. Forecasting model for the production and consumption of cotton fiber versus polyester. *Brazilian Journal of Agricultural Economics*, 21(1):49-62.
- *Kvalseth, T.O. 1985. Cautionary note about R^2 . *The American Statistician*, 39(4):279-285.
- Liew, V. K.S., Shitan, M. and Hussain, H. 2000. Time series modeling and forecasting of Sarawak Black Peeper Price. Munich Personal RePEc Archive. Paper No.-791, posted November 13, 2006. Online at <http://mpra.up.uni-muenchen.de/791/>.
- Montgomery, D. C., Peck, E. A. and Vining, G. G. 2003. In: Introduction to Linear Regression Analysis. John Wiley & sons, Inc.:221-258.
- Rajan, S. and Palanivel. 2014. Time series model to forecast production of cotton from India: An Application of Arima Model. *AE International Journal of Multi Disciplinary Research*, 2(1): 203-212.
- *Shapiro, S. S. and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52:591-611.
- Sidney, S. and Castellan, Jr. N. J. 1988. An analysis of variances test for normality (complete samples). *Biometrika*, 52: 591-611.
- Zhong, W.Z., Xiu-Lu, L., Shu-Lin, W. and Yong-Zen, L. 2007. Prediction of time series of main yield characters for cotton varieties in the yellow river valley region using ARIMA models. *Journal of China National Knowledge Infrastructure*, 4(2):133-152