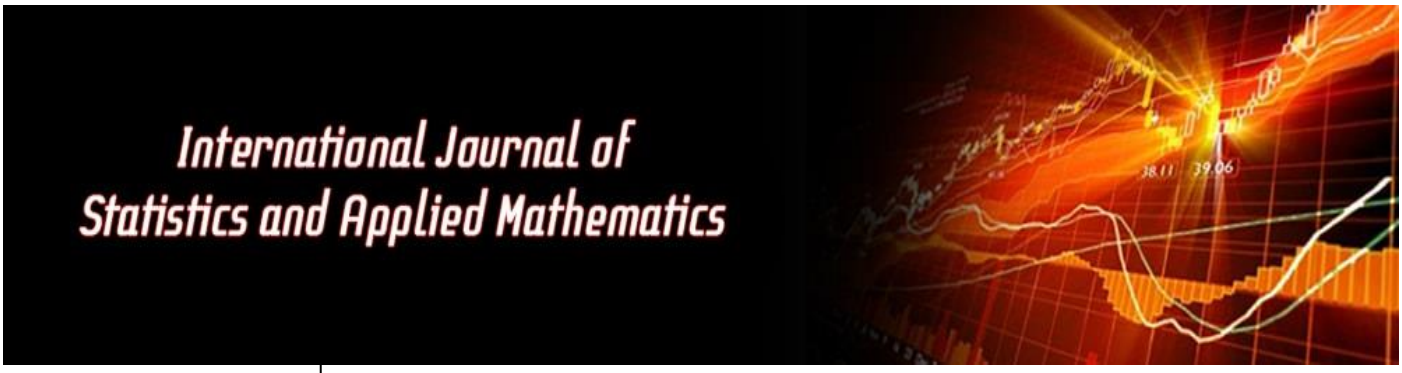

ISSN: 2456-1452
Maths 2023; 8(6): 157-164

https://www.mathsjournal.com
Received: 20-09-2023
Accepted: 21-10-2023

Siddharajsinh R Raj
M.sc Agricultural Statistics, Department of Statistics, BACA, Anand Agricultural University, Anand, Gujarat, India

Dr. AN Khokhar
Associate Professor, Department of Statistics, BACA, Anand Agricultural University, Anand, Gujarat, India

Sneh J Devra
Ph.D. Scholar, Depa Department of Statistics, Junagadh Agricultural University, Junagadh, Gujarat, India

Corresponding Author:
Siddharajsinh R Raj
M.sc Agricultural Statistics, Department of Statistics, BACA, Anand Agricultural University, Anand, Gujarat, India

Statistical models for forecasting area, production and productivity of banana in Gujarat: An empirical study

Siddharajsinh R Raj, Dr. AN Khokhar and Sneh J Devra

Abstract
The present study was carried out to estimate the trends of area, production and productivity of Banana of Gujarat. The time series data on area, production and productivity of Banana for the period 1996-97 to 2015-16 were collected from the Directorate of Horticulture, Gujarat state, Gandhinagar. The data from 1996-97 to 2012-13 have been used for model fitting and remaining for testing the forecast. Different polynomial models (linear, quadratic, and cubic) and time series models (ARIMA) were considered. The statistically most suited polynomial models were selected on the basis of adjusted R^2, significant regression coefficients, RMSE values, MAE values and assumptions of residuals (Shapiro-Wilk's test for normality and Run test for randomness). Appropriate ARIMA models were fitted after judging the time series data for stationarity based on graphically, auto-correlation function and partial auto-correlation function. The statistically model was selected on the basis of various goodness of fit criteria *viz.*, Akaike's information criteria (AIC), Bayesian information criteria (BIC), RMSE values, MAE values and assumptions of residuals (Shapiro-Wilks test for normality and Box-Ljung test for independence). The result showed that most of the cubic (third degree polynomial model) was found suitable for area, production and productivity of banana. For banana crop ARIMA (1,1,1), (2,1,0) and (1,1,0) suitable for area, production and productivity, respectively.

Keywords: Area, production, productivity

1. Introduction
India is currently producing about 283 million tons of horticulture produce and horticulture production has surpassed the food production in the country. It has proven beyond doubt that, productivity of horticulture crops is much higher compared to the productivity of food grains. The productivity of horticulture crops has increased by about 34% between 2004- 05 and 2014-15. The special thrust given to the sector, especially after the introduction of the Horticulture Mission for North East and Himalayan States (HMNEH) and the National Horticulture Mission (NHM) in the XI plan, has borne positive results. Given the increasing pressure on land, the focus of growth strategy has been on raising productivity by supporting high-density plantations, protected cultivation, micro-irrigation, and quality planting material, rejuvenation of several orchards and thrust on Post-Harvest Management and marketing of produce for better price realization.

India is the second largest producer of fruits and vegetables globally. Horticulture contributes about 30% of GDP in agriculture, using only 17% of land area. The area under horticulture increased 29% in 8 years, from 18.7 million ha in 2005-06 to 24.2 million ha in 2013-14 as more farmers are venturing into horticulture in their quest for diversification in agriculture. Horticulture production increased from 167 million tons in 2004-05 to 283 million tons in 2014-15 or 69% increase in 9 years.

India is a leader in producing fruits like Mango, Banana, Pomegranate, Sapota, Acid Lime and Aonla. Per capita availability of fruit to the Indian population is 189 gm/ person/ day and has been helping in supplementing nourishment.

Gujarat produces about 20.81 m. MT of horticulture from an area of 1.55 m.ha. accounting for 7.5% of total horticultural produce in the country. The major share of production is from vegetables (55.60%) and fruits (38.45 %).

In present investigation is undertaken to study fluctuation in productivity to arrive at a methodology that can precisely explain the fluctuation in area, production and productivity for banana in Gujarat state through different models *viz.*, Linear, Quadratic, Cubic, along with Autoregressive Integrated Moving Average (ARIMA).

2. Materials and Methodology

The time series data on area, production and productivity of banana crops for the period 1996-97 to 2015-16 were obtained from Directorate of Horticulture, Gujarat state, Gandhinagar. The data from 1996-97 to 2012-13 have been used for model fitting and remaining for testing the forecast**.**

In present investigation, the polynomial (linear, quadratic, cubic, and ARIMA) models were applied to study the trend analysis of area, production and productivity of banana. The details of various linear and non-linear models to be employed are given below in Table 1.

Table 1: List of linear and non-linear models

Model No.	Model	Name of the Model
I.	$Y=A+B*t$	Linear equation
II.	$Y=A+B*t+C*t^2$	Quadratic equation
III.	$Y=A+B*t+C*t^2+D*t^3$	Cubic equation

Where, Y is the area/ production/ productivity and X is the time points.

In the case of parametric (linear and non-linear) models, the model was selected if it fulfils the following characteristics.

1. The model should have significant F value.
2. The regression co-efficient in the model should be significant.
3. The residuals should be independently and normally distributed.

In the case of stochastic time-series (ARIMA) models, the model was selected if it fulfils the following characteristics.

1. It is parsimonious (uses the smallest number of co-efficient needed to explain the available data).
2. It is stationary (has AR co-efficient which satisfy some mathematical inequalities).
3. It is invertible (has MA co-efficient which satisfy some mathematical inequalities).

2.1 Fitting of polynomial models

A fundamental problem in statistics is to develop models based on a sample of observations and inferences are made using the model so developed. Over the last several decades, regression and time-series models play an important tool for statistical modeling and data analysis. Polynomial models *viz.*, regression (linear and non-linear) and time-series models provides information on relation between a response (dependent) variable and one or more predictor (independent) variables. Often, it is very difficult to select the most appropriate functional form just from looking at the data and sometimes there may not exist a suitable parametric form to express the functional form.

2.1.1 Linear Regression Approach (Rangaswamy, 2006) [7]

Regression analysis become one of the most widely used statistical tool for analysing functional relationships among the variables which is expressed in the form of an equation connecting the response or dependent variable Y (area, production and productivity) and time variable (t) as independent variable. The following model was fitted to original data

$$Y = a + bt \quad \ldots (3.1)$$

Where, a and b are the regression constant and regression coefficient.

2.1.2 Quadratic Regression Approach (Montgomery *et al.*, 2003) [5]

The fitted regression equation was as under

$$Y=a + bt + ct^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

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

2.1.3 Cubic Approach (Montgomery *et al.*, 2003) [5]

The model for the Cubic (Third degree polynomial) fitted to the data of each crop was as under

$$Y = a + bt + ct^2 + dt^3 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

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

2.2 Goodness of fit of the model

To test the goodness of fit of the fitted polynomial model, the co-efficient of determination R^2 defined as the proportion of total variation in the response variable (time) being explained by the fitted model is widely used and was calculated as under

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

and

$$\text{Adj. } R^2 = 1 - \frac{(n-1)(1-R^2)}{(n-k)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.6)$$

To test the overall significance of the model the F test was used.

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.7)$$

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.

$$t = \frac{b_i}{S.E.(b_i)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.8)$$

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

Where, bi is estimated i[th] coefficient, S.E. (bi) is the standard error of bi

In addition to the above criteria, two more reliability statistics Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were computed to measure the adequacy of the fitted model (*Liew et al.*, 2000). They can be computed as follows:

$$RMSE = \left[\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/n\right]^{1/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.9)$$

And

$$MAE = \sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|/n \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.10)$$

The fitted models which had lower values of these estimates were considered to be better.

As pointed out by Kvalseth (1985) [3], 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.

2.2.1 Test for the randomness of the residuals (Sidney and Castellan, 1988) [10]

The sample must be random to arrive at any conclusion about the population by using the information in the sample.

Let m be the number of elements of one kind (+ve sign residuals) and n be the number of elements of the other kind (-ve sign residuals) in sequence of N = m + n. To use one 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, determine the value of Z as under

$$\text{Mean} = \mu = \frac{2mn}{N} + 1 \text{ and standard deviation} = \sigma_r = \left[\frac{2mn(2mn-N)}{N^2(N-1)}\right]^{1/2} \quad (3.11)$$

$$Z = \frac{r - \mu_r}{\sigma_r} = \frac{r + h - \frac{2mn}{N} - 1}{\sqrt{\frac{2mn(2mn-N)}{N^2(N-1)}}} \quad \ldots\ldots\ldots\ldots\ldots (3.12)$$

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

Use normal table (Appendix A by Sidney and Castellan, 1988) [10] for testing Z value. The non-significant Z value indicates randomness of the residual.

2.2.2 Test for normality of the residuals (Shapiro-Wilk, 1965) [8]

The Shapiro–Wilk (1965) [8] statistic was used to test whether the residuals are normally distributed or not. The test is based on n residuals. These are arranged in non-decreasing sequence and is designated by $e_{(1)}$, $e_{(2)}$, $e_{(3)}$ … $e_{(n)}$. The following hypothesis is to be tested.

H_0: The residuals are normally distributed Vs H_1: These are not normally distributed.

The required test statistic W is defined as $W = S^2/b$ where

$$S^2 = \sum a(k)\left[e(n+1-k) - e(k)\right] \quad \ldots\ldots\ldots\ldots\ldots (3.13)$$

The parameter k takes the values

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

And

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

The values of co-efficients "a(k)" for different values of n and k are given in by Shapiro – Wilk (1965) [8]. When the calculated value of W is non-significant *i.e.* very close to unity, the null hypothesis regarding normality of residual was accepted.

2.3 Fitting of time–series model

In regression model it is usually assumed that the error terms are assumed to be uncorrelated. This implies that the various observations within a series are statistically independent. However, this assumption is rarely met in practice. Usually serial correlations in the observations often exist if the data are collected sequentially over time. *i.e.* each observation of the observed data series $\{Y_t\}$, which being a family of random variables $\{ Y_t , t \in T \}$, where T is the index set, T = { 0, ±1, ±2, … , } and apply standard time-series analysis technique to develop a model which will adequately represent the set of realizations and also their statistical relationship in a better way.

The statistical concept of correlation is to measure the relationships existing among the observations within the series. In these models, the values of correlations between the value of Y at time t (i.e., Y_t) and Y at earlier time periods (i.e., Y_{t-1}, Y_{t-2}, …) were examined. The algebraic forms of Autoregressive and Moving average processes are:

Autoregressive process

$$Z_t = C + \phi_1 Y_{t-1} + \varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots (A)$$

Moving average process

$$Z_t = C - \theta_1 a_{t-1} + \varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots (B)$$

Process (A) involving past (time – lagged) Y terms is called an autoregressive (Abbreviated as AR) process. The longest time lag associated with a Y term on the right hand side is called the AR order of the process. The equation (A) is thus an AR process of order one, abbreviated as AR (1). On the left hand side, Y_t represents the set of possible observations on a time sequenced random variables Y_1. The co-efficient ϕ_1 has a fixed numerical value which tells how Y_t is related to Y_{t-1}, C is a constant term related to the mean μ of the process. The constant term of an AR process is equal to the mean times the quantity one minus the sum of the AR co-efficients., i.e. for an AR (1) process $C = \mu(1 - \phi_1)$.

The variable a_t stands for a random shock element at the time point, t. Although Y_t is related to Y_{t-1}, the relationship is

not exact; it is probabilistic rather than deterministic. The random shock represents this probabilistic factor.

Now consider the process (B). The process with past (time – lagged) random shocks only are called moving average (abbreviated as MA) processes. The longest time lag associated with an error term (i.e. a_t) is called MA order of the process. The equation (B) is an MA process of order one, abbreviated as MA (1). C is a constant term related to the mean (μ) of the process. For a pure MA model C is equal to mean of the process, or, in symbol, $C=\mu$.The negative sign attached to θ_1 is merely a convention. The standard formula for calculating the auto-correlation co-efficient is

$$r_k = \frac{\sum_{t=1}^{n-k}(Y_t-\bar{Y})(Y_{t-k}-\bar{Y})}{\sum_{t=1}^{n}(Y_t-\bar{Y})^2} \qquad k = 1, 2, 3, \ldots, \ldots, \ldots$$

The above formula can be written more compactly since $\hat{Y}_t$ is defined as

$$Y_t - \bar{Y}$$

$$r_k = \frac{\sum_{t=1}^{n-k}\hat{Y}_t\hat{Y}_{t-k}}{\sum_{t=1}^{n}(\hat{Y}_t)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{(3.14)}$$

We use the symbol r_k for the estimated auto-correlation co-efficient among the observations separated by k time periods within a time series. After calculating estimated auto correlation co-efficient, we plot them graphically for different lags in an estimated auto correlation function (i.e. ACF).

The sample partial auto-correlation co-efficient can be estimated by using the following set of recursive equations.

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

$$\hat{\varphi}_{kk} = \frac{r_k - \sum \hat{\varphi}_{k-1,j} r_{k-j}}{1-\sum_{j=1}^{k-1}\hat{\varphi}_{k-1,j} r_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{(3.15)}$$

k = 2, 3, …………

Where,

$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j}$,k = 3, 4, …, … , ; j = 1,2,3, … , …, k – 1), $\hat{\phi}_{kk}$

is the estimate of the true partial auto correlation co-efficient ϕ_{kk} , r_k is the auto correlation co-efficient for k lags apart and $\hat{\phi}_{kj}$ is the estimate of partial auto-correlation co-efficient for k lags apart when the effect of j intervening lags has been removed.

2.3.1 Box – Jenkins Auto Regressive Integrated Moving Average (ARIMA) Models

Box – Jenkins time – series models written as ARIMA (p, d, q) was first popularized by Box and Jenkins (1976) [1]. This model amalgamates three types of processes, *viz.*, auto regressive of order ‘p’ differencing to make a series stationary of degree ‘d’ and moving average of order ‘q’. This method applies only to a stationary time series data. When the data is non-stationary which has to be brought into stationary by the method of differencing i.e. $W_t = Y_t - Y_{t-1}$. The series W_t is called the first differences of Y_t and the second difference of the series is $V_t = W_t - W_{t-1}$. In many cases first differencing is sufficient to bring about a stationary mean and second differencing is done in few cases only.

2.3.1.1 Test for Stationarity

The stationarity requirement ensures that one can obtain useful estimates of the mean, variance and ACF from a sample. If a process has a mean that is changing in each time period, one could not obtain useful estimates since only one observation available per time period. This necessitates testing any observed series of data for stationarity.

There are three ways to determine whether the above-mentioned stationarity requirement is met.

1. Examine the realization visually to see if either the mean or the variance appears to change over time.
2. Examine the estimated AR co-efficient to see if it satisfies the stationary condition. In case of AR (1) process the condition for stationary is that absolute value of ϕ_1 must be less than one, or symbolically, $|\phi_1|<1.$ In practice one don’t know ϕ_1, therefore, one apply the condition to $\hat{\phi}_1$ (i.e. estimate of ϕ_1) rather than ϕ_1.
3. For an MA (1) process the corresponding condition is that the absolute value of θ_1 must be less than one. Which is called the condition of invariability, or in symbols $|\theta_1|<1.$

Examine the estimated ACF to see if the auto-correlations move rapidly towards zero. In practice, “rapidly” means that the absolute t-values of the estimated auto-correlations should fall below roughly 1.6 by about lags 4 or 5. These numbers are only guidelines and are not absolute rules. If the ACF does not fall rapidly to zero, we should suspect a non-stationary mean and consider differencing of the data.

To find out the t-value of the estimated auto-correlation Barlett’s approximate expression for the standard error of the sampling distribution of r_k values can be used. The estimated standard error, designated as $S(r_k)$ is calculated using the following expression.

$$s(r_k) = \left(1+2\sum_{j=1}^{k-1} r_j^2\right)^{1/2} n^{-1/2} \quad \ldots\ldots\ldots\ldots\ldots \text{(3.16)}$$

The following null hypothesis is to be tested

$$H_0 : \rho_k = 0$$

for k = 1, 2, 3, … using the test statistics

$$t_{r_k} = \frac{r_k - \rho_k}{S(r_k)} \qquad k = 1, 2, 3,\ldots\ldots\ldots\ldots\ldots \text{(3.17)}$$

If the value of t comes out to be significant, we reject H_0 at the level of significance and conclude that $\rho \neq 0$.

2.3.1.2 Methodology related to ARIMA model

ARIMA modelling consists of three operational steps:

i. Identification ii. Estimations iii. Diagnostic checking

(i) Identification

At the identification stage, compare the estimated ACF and PACF's to find a match. Choose, as a tentative model, the ARMA process whose theoretical ACF and PACF best match the estimated ACF and PACF. In choosing a tentative model, one should keep in mind the principle of parsimony. The most important general characteristics of theoretical ACF and PACF of AR and MA models are

1. A stationary AR process has a theoretical ACF and decays or "damps out" toward zero. But it has theoretical PACF that cuts off sharply to zero after few spikes. The lag length of the last PACF spike equals the AR order (p) of the process.
2. A MA process has a theoretical ACF that cuts off to zero after a certain number of spikes. The lag length of the last ACF spike equals the MA order (q) of the process. The theoretical PACF decays or "dies out" toward zero.

The general characteristics of theoretical ACF and PACF of five common stationary process, *viz.* AR(1), AR(2), MA(1), MA(2) and ARMA(1,1) are summarized in the following table.

(ii) Estimation

Estimating the parameters for Box – Jenkins models is a quite complicated non-linear estimation problem. For this reason, the parameter estimation should be left to a high quality software program that fits Box-Jenkins models. The main approaches for fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation. Maximum likelihood estimation is generally the preferred techniques.

(iii) Diagnostic Checking

At the identification stage of the Box – Jenkins time series methodology, a tentative model based on the patterns of ACF and PACF can be selected. The parameters of such a model were estimated at the estimation stage. Now at the final stage of ARIMA model building, namely the diagnostic checking stage it is necessary to test the suitability of the selected model. For this purpose, the following goodness of fit statistics were calculated.

Akaike's Information Criterion (AIC)

Denoting by v^*, the estimate of white noise variance σ^2, obtained by fitting the corresponding ARIMA model, the AIC consists in computing the statistic,

$$AIC(p,q) = Ln\, v^*(p, q) + (2/n)(p+q), \quad \text{..... (3.18)}$$

Where, p and q are the order of AR and MA processes respectively and n is the number of observations in the time-series.

Bayesian Information Criterion (BIC)

This is computed as

$$BIC(p,q) = Ln\, v^*(p,q) + (p+q)[Ln(n)/n] \quad \text{..... (3.19)}$$

A modification to BIC is the Schwarz – BIC (Cromwell *et al*., 1994), given by $SC(p,q) = n * Ln\, v^*(p+q)+(p+q) Ln\, n$. The lower the values of these statistics, the better is the selected model. To test the independency assumption of the residuals, the Box – Ljung statistic (Q) (Cromwell *et al*., 1994) is utilized.

2.3.1.3 Test for independence of errors (Chi – squared test)

The Ljung and Box Chi-square can be used to test the residual auto-correlations are independent or not. The following null hypothesis about the correlations among the random shocks was to be tested

$$H_0 : \rho_1(a) = \rho_2(a) = ...= \rho_k(a) = 0$$

with the test statistic

$$Q = n(n+2)\sum_{i=1}^{k}(n-i)^{-1} r_i^2(\hat{a}) \quad \text{..... (3.20)}$$

Where, n is the number of observations used to estimate the model. The statistics Q approximately follows a Chi - squared distribution with (k - m) degrees of freedom where k is the number of lags and m is the number of parameters estimated in the ARIMA model. If the χ^2 (or Q value) is less than the tabulated χ^2, then the residual auto-correlations are not significantly different from zero. It means that residuals are independent.

To summarize the three stages of ARIMA model building, the parameters of the tentatively selected ARIMA model at the identification stage are estimated at the estimation stage and the adequacy of the chosen model is tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for representing the time-series observations under consideration.

2.4 Model selection criteria

In the case of parametric (linear and non-linear) models, the model was selected if it fulfils the following characteristics.

1. The model should have significant F value.
2. The regression co-efficients in the model should be significant.
3. The residuals should be independently and normally distributed.
4. In the case of stochastic time-series (ARIMA) models, the model was selected if it fulfils the following characteristics.
5. It is parsimonious (uses the smallest number of co-efficients needed to explain the available data).
6. It is stationary (has AR co-efficients which satisfy some mathematical inequalities).
7. It is invertible (has MA co-efficients which satisfy some mathematical inequalities).
8. The estimated co-efficients must be significant (absolute t – values about 2.0 or larger).
9. It should have statistically independent and normally distributed residuals.

3. Results and Discussion

3.1 Trend for area of banana crop

The data presented in Table 2 for area under cultivation of banana crop revealed that among different linear (first, second and third degree) models fitted. Among them the third degree

polynomial was found suitable to fit the trend in area under the banana crop.

$Y=28.730^{**} - 1.173^{**}t + 0.481^{**}t^2 - 0.016^{**}t^3$

As series was found non-stationary, the new variable Xt was constructed by taking differences of one (*i.e.* d=1) to make the series stationary. Among the models, those models having lower value of AIC and SBC are given in Table 3 ARIMA (1,1,1) model had comparatively lower value of AIC and SBC with significant AR (φ) coefficient and MA (θ) coefficient was significant.

By using cubic model of original data approach and ARIMA (1,1,1), Predicted values are given in Table 4

Table 1: Characteristics of fitted linear and non-linear models for area, production and productivity of banana crop for Gujarat

Aspects	Model	Regression constant	Partial regression co-efficient			Goodness of fit				
		a	b	c	d	Adj.R^2 (%)	S-W Test	Run Test (|Z|)	RMSE	MAE
Area	Linear	20.506**	2.899**	-	-	95.5	0.920	-1.936	2.979	2.297
	Quadratic	23.246**	2.034**	0.048	-	95.8	0.922	-2.499	2.796	2.321
	Cubic	28.730**	-1.173	0.481**	-0.016**	96.7	0.930	-2.499	2.390	2.027
Production	Linear	142.435	245.55**	-	-	93.7	0.928	-1.997	301.45	252.60
	Quadratic	722.396**	62.41	10.175**	-	96.8	0.921	-1.802	208.97	187.90
	Cubic	1361.577**	-311.38**	60.636**	-1.869**	98.8	0.928	-2.499	122.46	109.34
Productivity	Linear	27.070**	2.217**	-	-	89.8	0.956	-1.417	3.526	2.901
	Quadratic	31.773**	0.732	0.083*	-	91.8	0.915	-1.494	3.054	2.784
	Cubic	42.414**	-5.492**	0.923**	-0.031**	98.7	0.965	1.020	1.176	0.960

Table 2: Characteristics of fitted time series models for area, production and productivity of banana crop for Gujarat

Aspect	Model (p,d,q)	Constant	AR(Φ)		MA(θ)	AIC	BIC	S-W TEST	B-Q TEST	RMSE
			Φ1	Φ2						
Area	(1,1,1)	-35966.8	0.340*	-	0.119*	89.11	90.06	0.947	0.244	171.76
	(2,1,0)	-34875.1	0.185*	0.109	-	91.42	92.37	0.957	0.014	171.21
Production	(1,1,0)	8448.82	0.157*	-	-	217.18	219.50	0.876	0.106	227.56
	(2,1,0)	11407.44	0.089*	0.457*	-	216.90	219.22	0.895	0.026	209.14
Productivity	(1,1,0)	-116.51	0.262*	-	-	80.72	83.04	0.866	0.227	2.49
	(2,1,0)	-127.62	0.250*	0.072	-	81.69	84.01	0.877	0.664	2.56

Table 3: Testing of forecast values for remaining three year by using best fitted models cubic model and ARIMA (1,1,1) for area of banana

Year	Observed	Predicted			
		Cubic	Deviation	(1,1,1)	Deviation
2013-14	66.50	70.14	3.64	72.88	6.38
2014-15	67.01	70.34	3.33	76.79	9.78
2015-16	64.69	69.67	4.98	79.36	14.67

3.2 Trend for production of banana crop

The data presented in Table 2 for production under cultivation of banana crop revealed that among different linear (first, second and third degree) models fitted. Among them the third degree polynomial was found suitable to fit the trend in production under the banana crop.

$Y=1361.57^{**} - 311.38^{**}t + 60.636^{**}t^2 - 1.869^{**}t^3$

As series was found non-stationary, the new variable Xt was constructed by taking differences of one (*i.e.* d=1) to make the series stationary. Among the models, those models having lower value of AIC and SBC are given in Table 4.1.2 ARIMA (2,1,0) model had comparatively lower value of AIC and SBC with significant AR (φ) coefficient and MA (θ) coefficient was significant.

By using cubic model of original data approach and ARIMA (2,1,0), Predicted values are given in Table 5

Table 5: Testing of forecast values for remaining three year by using best fitted models cubic model and ARIMA (2,1,0) for production of banana

Year	Observed	Predicted			
		Cubic	Deviation	(2,1,0)	Deviation
2013-14	4225.49	4502.78	277.29	4892.37	666.88
2014-15	4324.35	4515.47	191.12	5298.22	973.87
2015-16	4185.52	4436.37	250.85	5711.57	1526.05

3.3 Trend for productivity of banana crop

The data presented in Table 2 for productivity under cultivation of banana crop revealed that among different linear (first, second and third degree) models fitted. Among them the third degree polynomial was found suitable to fit the trend in productivity under the banana crop.

$Y=42.414^{**} - 5.492^{**}t + 0.923^{**}t^2 - 0.031^{**}t^3$

As series was found non-stationary, the new variable Xt was constructed by taking differences of one (*i.e.* d=1) to make the series stationary. Among the models, those models having lower value of AIC and SBC are given in Table 3 ARIMA (1,1,0) model had comparatively lower value of AIC and SBC with significant AR (φ) coefficient and MA (θ) coefficient was significant.

By using cubic model of original data approach and ARIMA (1,1,0), Predicted values are given in Table 6.

Table 6: Testing of forecast values for remaining three year by using best fitted models cubic model and ARIMA (1,1,0) for productivity of banana

Year	Observed	Predicted			
		Cubic	Deviation	(1,1,0)	Deviation
2013-14	63.54	61.81	1.73	67.09	3.55
2014-15	64.53	58.64	5.89	70.47	5.94
2015-16	64.70	53.77	10.93	74.06	9.36

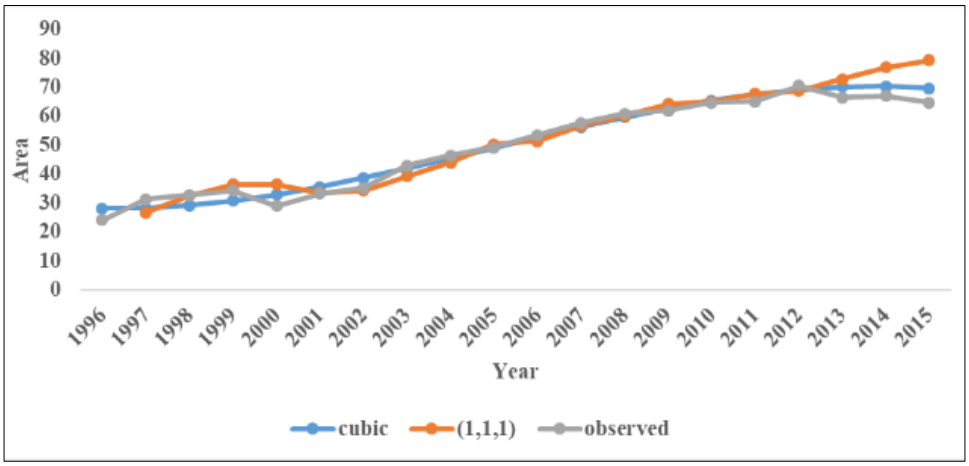

Fig 1: Trend in area of banana based on ARIMA (1,1,1) and cubic in Gujarat

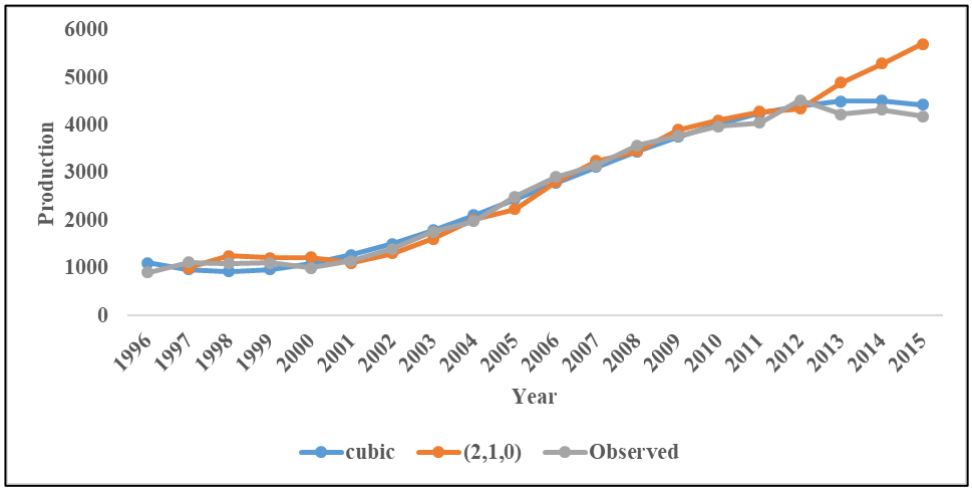

Fig 2: Trend in production of banana based on ARIMA (2,1,0) and cubic in Gujarat

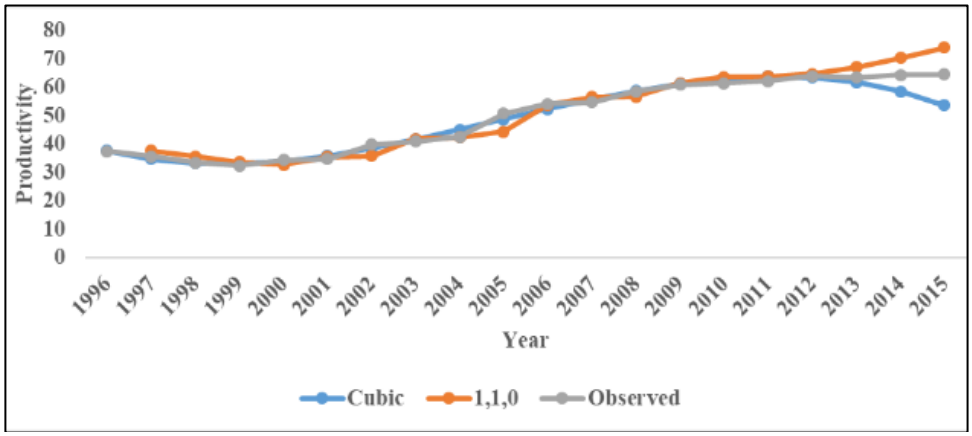

Fig 3: Trend in productivity of banana based on ARIMA (1,1,0) and cubic in Gujarat

4. Summary and Conclusion

The present study was carried out to estimate the trends of area, production and productivity of banana of Gujarat. The time series data on area, production and productivity of major fruit crops and total fruit crops for the period 1996-97 to 2015-16 were collected from the Directorate of Horticulture, Gujarat state, Gandhinagar.

For this purpose, different polynomial models (linear, quadratic, and cubic) and time series models (ARIMA) were considered. The statistically most suited polynomial models

were selected on the basis of adjusted R^2, significant regression coefficients, RMSE values, MAE values and assumptions of residuals (Shapiro-Wilk's test for normality and Run test for randomness). Appropriate ARIMA models were fitted after judging the time series data for stationarity based on graphically, auto-correlation function and partial auto-correlation function. The statistically model was selected on the basis of various goodness of fit criteria *viz.,* Akaike's information criteria (AIC), Bayesian information criteria (BIC), RMSE values, MAE values and assumptions of residuals (Shapiro-Wilks test for normality and Box-Ljung test for independence).

4.1 Fitting trends of area of Banana

Among the different polynomial models (linear, quadratic and cubic) fitted for area under the banana crop following third degree polynomial model was evolved as best fitted trend model

$$Y=28.730^{**} - 1.173^{**}X + 0.481^{**}X^2 - 0.016^{**}X^3$$

ARIMA (1,1,1) model fulfilled all statistical requirement for selected. The equation of this model is

$$Y = -35966.8 + 0.340^{*} y_{t-1} + 0.119^{*} y_{a-1}$$

4.2 Fitting trends of production of Banana

Among the different polynomial models (linear, quadratic and cubic) fitted for production under the banana crop following third degree polynomial model was evolved as best fitted trend model

$$Y=1361.57^{**} - 311.38^{**}X + 60.636^{**}X^2 - 1.869^{**}X^3$$

ARIMA (2,1,0) model fulfilled all statistical requirement for selected. The equation of this model is

$$Y = 11407.442 + 0.089^{*} y_{t-1} + 0.457\ y_{t-2}$$

4.3 Fitting trends of productivity of Banana

Among the different polynomial models (linear, quadratic and cubic) fitted for productivity under the banana crop following third degree polynomial model was evolved as best fitted trend model

$$Y=42.414^{**} - 5.492^{**}X + 0.923^{**}X^2 - 0.031^{**}X^3$$

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

$$Y = -116.513 + 0.262^{*} y_{t-1}$$

5. References

1. Box GEP, Jenkins GM, Reinsel GC. Autocorrelation function and spectrum of stationary processes and analysis of seasonal time series. In: Time Series Analysis: Forecasting and Control. 2nd ed. San Francisco: Holden-Day; c1976. p. 21-43.
2. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control. 2nd ed. Holden Day; c1976.
3. Kvalseth TO. Cautionary note about R2. The American Statistician. 1985;39(4):279-285.
4. Liew VKS, Shitan M, Hussain H. Time series modeling and forecasting of Sarawak Black Pepper Price. Munich Personal RePEc Archive; c2000. Paper No.-791. Available from: http://mpra.up.uni-muenchena.de/791/
5. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. John Wiley & Sons, Inc.; c2003. p. 221-258.
6. Pankratz A. Forecasting with univariate Box (Jenkins models). John Wiley & Sons, Inc; c1983.
7. Rangaswamy R. A text book of agricultural Statistics. New Age International (P) Limited Publishers; c2006. p. 158-210.
8. Shapiro SS, Wilk MB. An analysis of variance test for normality (Complete samples). Biometrika. 1965;52:591-611.
9. Shitan M, Liew KS. Time series modelling of Sarawak black pepper price. Unpublished paper; c2000.
10. Sidney Siegel, John Castellan NJr. An analysis of variances test for normality (Complete samples). Biometrica. 1988;52:591-611.