Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2582-2845.7973

ISSN: 2582 – 2845

Ind. J. Pure App. Biosci. (2020) *8*(2), 64-68

Research Article

Indian Journal of Pure & Applied Biosciences

Peer-Reviewed, Refereed, Open Access Journal

Validation of Arima Model on Production of Papaya in India

B. Ramana Murthy[1*], G. Mohan Naidu[2], C. Tamilselvi[3] and N. Priyanka Evangilin[4]

[1]Assistant Professor, [2]Professor and Head, [3&4]M.Sc., Agricultural Statistics

Department of Statistics and Computer Applications,

Acharya N.G. Ranga Agricultural University, S.V. Agricultural College, Tirupati

*Corresponding Author E-mail: ramanastats@gmail.com

Received: 29.01.2020 | Revised: 26.02.2020 | Accepted: 2.03.2020

ABSTRACT

The present study was carried out to validate the Auto Regressive Integrated Moving Average (ARIMA) model on production of papaya in India. In this study, we used secondary data for the period from 1950-51 to 2017-18. The ARIMA (p, d, q) models were fitted for the period from 1950-51 to 2009-10 and the remaining periods i.e. 2010-2018 were used for validation of the ARIMA model. The validity of the models were tested by using standard statistical criterias (i.e., R^2, RMSE and MAPE). Among all the ARIMA (p, d, q) models, ARIMA (3, 1, 4) was found to be best fitted model for validation and forecast the future values. The forecasted values of papaya production from 2010-11 to 2017-18 were approximately similar to original values.

Key words: *Papaya, ARIMA, R^2, RMSE and MAPE.*

INTRODUCTION

The Papaya (*Carica papaya* L.) which belongs to member of the Caricaceae family and is the most economically important species in the family. Papaya is the third most cultivated tropical crop world-wide. Brazil and India are the largest producers of papaya although Mexico is the main exporter. Among common fruits, papaya is ranked first on nutritional scores for the percentage of vitamin A, vitamin C, potassium, folate, niacin, thiamine, riboflavin, iron and calcium, and fiber. Moreover, fruits, stems, leaves and roots of papaya are used in a wide range of medical applications and papain production. In India total cropped area under papaya is 138 thousand ha and production of 5989 thousand MT which **is** highest in the world. The average productivity of papaya in the country is 43.3 MT per ha. India contributes about 42.6% of world's papaya production. Only 0.08% of the total domestic production is exported and rest (99.92%) all is consumed within the country. Papaya is mostly cultivated in the states of Andhra Pradesh, Karnataka, Gujarat, Orissa, West Bengal, Assam, Kerala, Madhya Pradesh and Maharashtra. Andhra Pradesh is the leading producer of papaya with production of 1687.82 thousand MT. Three notable varieties are Pusa Delicious, Pusa Dwarf, Pusa Nanha.

Cite this article: Ramana Murthy, B., Mohan Naidu, G., Tamilselvi, C., & Evangilin, N.P. (2020). Validation of Arima Model on Production of Papaya in India, *Ind. J. Pure App. Biosci. 8*(2), 64-68. doi: http://dx.doi.org/10.18782/2582-2845.7973

Mahesh and Jain (2013) studied to determine the area, production and productivity of papaya in Raipur district and Chhattisgarh state by using Compound Growth Rates and they were found that the area, production of papaya would be tremendously increasing and productivity has showed the fluctuations during their study period.

Debnath et al. (2013) were studied forecasting the cultivated area and production of cotton in India using Autoregressive Integrated Moving Average (ARIMA) model by using the data from 1950-51 to 2010-11. They revealed that ARIMA (0, 1, 0) ARIMA (1, 1, 4) and
ARIMA (0, 1, 1) are the best fitted models for forecasting of cotton area, production and yield in India respectively.

Prabakaran et al. (2014) analyzed the Pulses Area and Production in India during the period from 1950-51 to 2011-12 by using ARIMA model and he found that ARIMA (1, 1, 0) and ARIMA (2, 1, 1) models were best fitted to forecast Pulses Area and Production in India.

Ramana Murthy et al. (2018) studied the appropriate ARIMA model for forecasting sunflower Production in India and he found that ARIMA (4, 1, 4) model was found to be a best fitted model to forecast Sunflower Production in India.

The objective of the present study is to fit and validate different ARIMA (p, d, q) models for the time series data on Production of Papaya in India.

MATERIALS AND METHODS

The data of the study for a period of 60 years (1950-51 to 2009-10) and the validation period from 2010-11 to 2012 - 18 pertaining to the production of papaya were collected from the source of indiastat.com. In order to examine and fitted various ARIMA (p, d, q) models on Production of Papaya in India by using SPSS 20 version.

2.1 Auto Regressive Integrated Moving Average (ARIMA)

The ARIMA methodology is also called as Box-Jenkins methodology (Box and Jenkins 1976). The Box-Jenkins procedure is concerned with fitting a mixed ARIMA model to a given set of data. The main objective in fitting ARIMA model is to identify the stochastic process of the time series and predict the future values accurately. This method shave also been useful in many types of situations which involve the building of models for discrete time series and dynamic systems. However the optimal forecast of future values of a time series are determined by the stochastic model for that series. A stochastic process is either stationary or non-stationary. The first thing to note is that most time series are non-stationary and the ARIMA models refer only to a stationary time series. Since the ARIMA models refer only to a stationary time series the first stage of Box-Jenkins model is for reducing non-stationary series to a stationary series by taking the differences.

The ARIMA (p, d, q) process is given by

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots\ldots\ldots + \phi_p y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \ldots\ldots - \theta_q \varepsilon_{t-q} \quad (1)$$

Where y_t and ε_t are the actual value and random error at time period t, respectively. $\phi_i (i = 1,2,3,\ldots\ldots, p)$ and θ_j (j=1, 2, 3,…..,q) are model parameters. p and q are integers and often referred to as orders of the model. Random errors ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

The main stages in setting up a Box-Jenkins forecasting model are as follows:

1. Identification 2. Estimating the parameters 3. Diagnostic checking 4. Forecasting

RESULTS AND DISCUSSION

In the present study, the data for Papaya Production for the period of 60 years (1950-51 to 2009-10) were used for the study.

3.1 Model Identification.

Among several methods studies the goodness of fitted models were examined by highest R^2 value, lowest RMSE (Residual Mean Square Error) and MAPE (Mean Absolute percentage Error) values. Based on these criterions ARIMA (3, 1, 4) is found to be the best fitted model for study period. The Coefficient of determination (R^2), Mean Absolute Percentage Error (MAPE) and Residual Mean Square Error (RMSE) are given by

$$R^2 = 1 - \left[\sum_{t=1}^{n} (y_t - \hat{y}_t)^2 / \sum_{t=1}^{n} (y_t - \bar{y})^2 \right] \quad (2)$$

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} \left(y_t - \hat{y}_t \right)^2}{n}} \quad (4)$$

Where y_t is the actual observation for time period't' and $\hat{y}_t$ is the predicted value for the same period and $\bar{y}$ is the overall sample mean of observations.

The different ARIMA (p, d, q) models were tested and the models which has Maximum R^2 and Minimum MAPE and RMSE were chosen. The models and the corresponding values are shown in table (1)

Table 1:

Papaya-Production	Models	R^2	RMSE	MAPE
ARIMA(p, d, q)	(1, 1, 2)	.900	314.060	24.471
	(2, 1, 3)	.895	325.444	22.101
	(3, 1, 3)	.926	278.818	22.108
	(3, 1, 4)	**.928**	**277.538**	**22.072**
	(4, 2, 4)	.916	304.557	24.328
	(4, 3, 4)	.886	366.145	30.235
	(1, 0. 2)	.874	351.093	28.792

3.2 Model Estimation and Verification

The parameters of the model were estimated by using SPSS 20 package. The ARIMA (3, 1, 4) model is best fitted model for forecasting future values (i.e validation period) of Production of Papaya. The model verification (or) diagnosed by the Ljung-Box Q statistic. The Ljung-Box Q statistic is to check the overall adequacy of the model. The test statistic Q is given by

$$Q_n = nr(nr+2) \sum_{i=1}^{n} \frac{r_i^2(e)}{nr - l} \quad (5)$$

Where $r_l(e)$ is the residual autocorrelation at lag l, nr is the number of residual, n is the number of time lags included in the test for model to be adequate, p-value associated with Q statistics should be large $(p - value > \alpha)$. The results of estimation are reported in Tables 2.

Table 2: Estimates of fitted ARIMA (3, 1, 4) model for Papaya Production

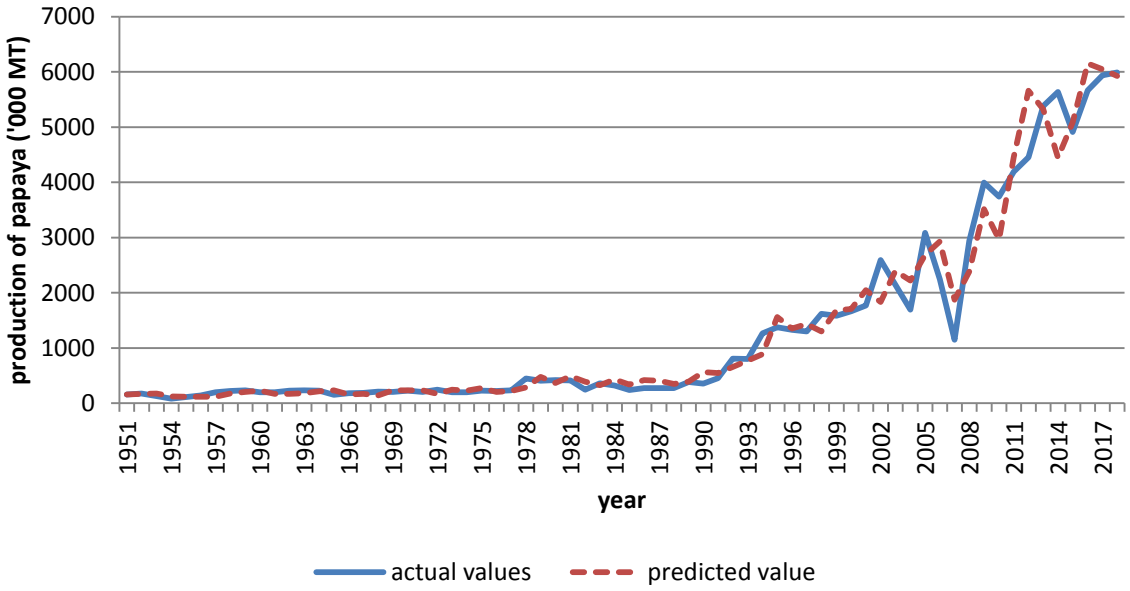
Model fit Statistics			Ljung-Box Q (18)	
R-Square	RMSE	MAPE	Statistic	p-value
0.928	277.538	22.072	10.102	0.521

3.3 Validation of ARIMA (3, 1, 4) model for the period 2010-2011 to 2017-18:

Table 3:

Year	Actual value of Production	Predicted values by ARIMA (3, 1, 4)
2010-11	4195.5	4461.0
2011-12	4457.0	5658.1
2012-13	5381.3	5321.4
2013-14	5639.3	4449.8
2014-15	4913.0	5087.2
2015-16	5667.0	6151.2
2016-17	5940.0	6051.2
2017-18	5989.0	5929.7

ARIMA (3, 1, 4) Model



CONCLUSION

The study revealed that among all ARIMA (p, d, q) models, the ARIMA (3, 1, 4) model was found to be the best fitted model for forecasting of future values of production of papaya in India. Based on p-value of Ljung-Box Q statistic, the model ARIMA (3, 1, 4) was good. The forecasted values of validation period were approximately closed to be actual values. Time series analysis and forecasting is an active research area over the last few decades. The accuracy of time series forecasting is fundamental to many decision processes and hence the research for improving the effectiveness of forecasting models. With the efforts of Box and Jenkins (1970), the ARIMA (p, d, q) models has become one of the most popular methods in the forecasting research and practice.

REFERENCES

Box, G. E. P., & Jenkin, G. M. (1976). Time series of analysis, Forecasting and

Control. Sam Franscico, Holden Day, California, USA.

Maske, M., & Jain, B.C. (2013). Compound Growth Rates area, production and productivity of papaya in Raipur district and Chhattisgarh state. *International Journal of Agriculture Environment & Biotechnolog*, *6*(1), 139-143.

Debnath, M.K., Bera, K., & Mishra, P. (2013). Forecasting Area, Production and Yield of Cotton in India using ARIMA Model. *Journal of Space Science & Technology*. *2*(1), 16-20.

Prabakaran, K., Nadhiya, P., Bharathi, S., & Isaivani, M. (2014). Forecasting of Pulses area and production in India – An ARIMA Approach. *Indian Streams Research Journal*. *4*(3), 1-8.

Ramana Murthy, B., Nafeez Umar, SK., & Hari Babu, O. (2018). ARIMA Models for forecasting Sunflower Production in India. *International Journal of Pure & Applied Bioscience*. *6*(6), 1121-1126.

Ramana Murthy, B., & Hari Babu, O. (2018). A Statistical Trend Analysis of Mango area, production and productivity in Andhra Pradesh. *International Journal of agricultural and statistical sciences*, *1*(14), 337-342.

Ramana Murthy, B., Mohan Naidu, G., Ravindra Reddy, B., & Nafeez Umar, SK. (2018). Forecasting Groundnut area, production and productivity of India using ARIMA model. *International Journal of Agricultural Statistical Sciences 14*(1), 153-156.

www.indiastat.com.