JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3591-3599
https://publishoa.com
ISSN: 1309-3452

Design and Development of Stochastic Modelling for Musa paradisiaca Linn Production in India

T. Jai Sankar and **P.Pushpa**
Department of Statistics, Bharathidasan University, Tiruchirappalli, Tamil Nadu, INDIA
tjaisankar@gmail.comand pushbkr@gmail.com

Received 2022 March 25; **Revised** 2022 April 28; **Accepted** 2022 May 15.

Abstract

This study aims at design and development of stochastic modelling for Musa paradisiaca Linn(Banana) production in India during the years from 1961 to 2019. In India,plantation of banana can take place from February to May in South India, and from July to August in North India. And, in South India, except during the summer. The study considers Autoregressive (AR), Moving Average (MA) and ARIMA processes to select the appropriate ARIMA model for Musa paradisiaca Linnproduction in India. Based on ARIMA (p,d,q) and its components Autocorrelation Function (ACF), Partial Autocorrelation Function(PACF), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Normalized BIC and Box-Ljung Q statistics estimated, ARIMA (0,1,1) was selected. Based on the chosen model, it could be predicted that Musa paradisiaca Linnproduction would increase to 36.09 million tons in 2025 from 30.46million tons in 2019 in India.

Keywords: ARIMA, BIC, Forecasting, MAPE, Banana Production, RMSE.

Introduction

Bananas are a major fruit crop in tropical and subtropical regions. Annually, 14.2 million tonnes of banana is produced in India, i.e. more than any other country in the world. In India, banana fruit is the second most important fruit after the mango. Tamil Nadu, Andhra Pradesh, Maharashtra, and Karnataka contribute to a significant percentage of banana production in India (Figure 1). Andhra Pradesh is the largest banana-producing state in India, with a share of 16.27%. Gujarat and Maharashtra are the second and third largest producers of bananas in India. Both states have a share of 14% and 13% respectively. Jalgaon in Maharashtra is called Banana city, as it contributes near to two-thirds of Maharashtra's banana production. Convenient, inexpensive, and delicious-bananas are one of the most popular fruits in the world. From the health benefits (Figure 2) of heart and weight loss, to beauty care and digestive support, there are dozens of reasons we should incorporate all varieties of this fruit into our daily diet.

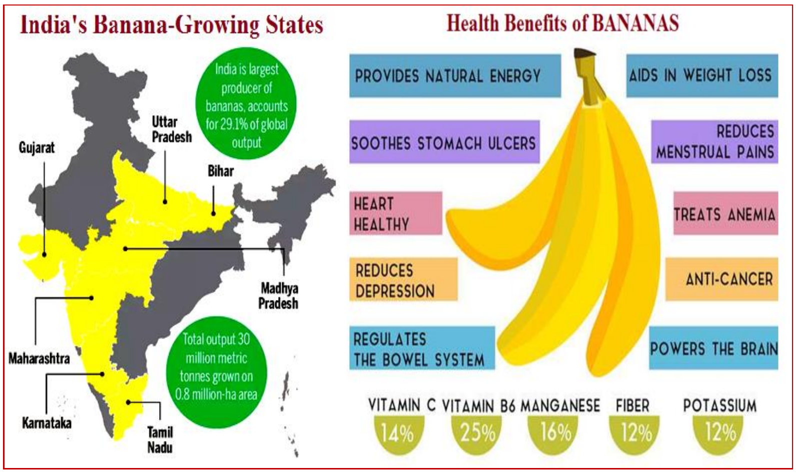

Figure 1. India's Banana-Growing States and Health Benefits

A tropical crop banana needs warm, humid, and rainy climate. Temperature ranging from 10 to 40°C and the relative humidity 80% and above is suitable for growing bananas. The banana plant is sensitive towards dew and frost, and cannot bear arid conditions. Strong dry winds can affect the growth of the banana plant, its yield and, fruit quality. Soil for bananas should be rich loamy soil with good drainage, adequate fertility, moisture, and plenty of organic matter. The suitable pH range is 6 to 8. Nutritionally deficient, saline solid and very sandy soil is not suitable for bananacultivation.The nutritional, minerals and vitamins values of bananas per 100 grams are given in Figure 2:

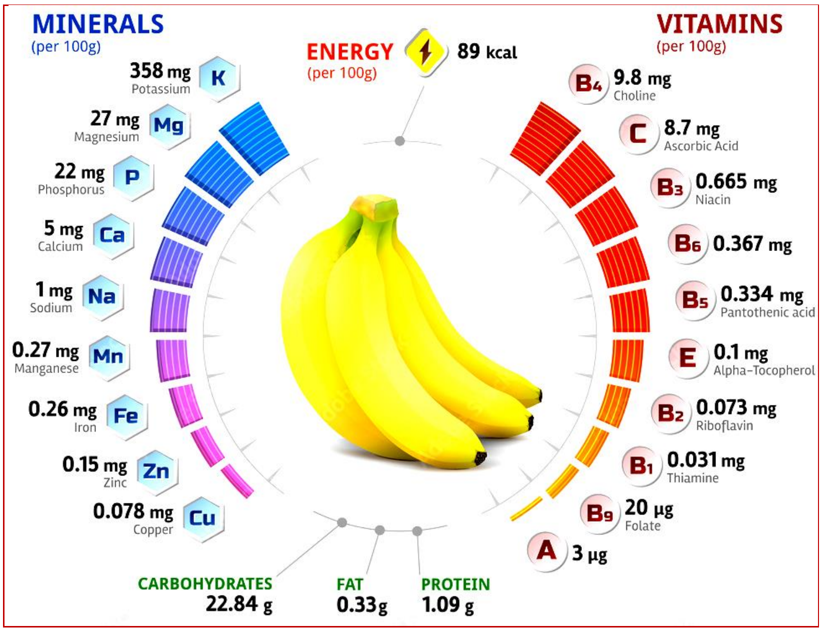

Figure 2. Nutritional, Minerals and Vitamins of Bananas

Material and Methods

As the aim of the study was to design and development of stochastic modelling for Musa paradisiaca Linnproduction in India, various forecasting techniques were considered for use. ARIMA model, introduced by Box and Jenkins (1976), was frequently applied for discovering the pattern and predicting the future values of the time series data. Box and Pierce (1970) measured the distribution of residual autocorrelations in ARIMA. Akaike (1970) found the stationary time series by an AR (p), where p is finite and bounded by the same integer. Moving Average (MA) models were applied by Slutzky (1973). Md.Moyazzem Hossain andFaruq Abdulla (2016) applied ARIMA (0,2,1) model for yearly potato production in Bangladesh for the period from 1971 to 2013. Borkar et al. (2016) found that ARIMA (2,1,1) is the appropriate model for forecasting the production of cotton in India. Kour et al., (2017) applied ARIMA model for forecasting of productivity of pearl millet of Gujarat for the period from 1960-61 to 2011-12 and validated ARIMA (0,1,1) model performs quite satisfactorily as the RMAPE value is less than 6 percent. Hemavathi and Prabakaran (2018) found rice production data during 1990-2015 and applied ARIMA (0,1,1) model up to 2020. BholaNath et al. (2018) discovered ARIMA (1,1,0) model for wheat production in India for the period from 1949-50 to 2016-17 and forecasted up to 2026-27.

Stochastic time-series ARIMA models were widely used in time series data which are having the characteristics (Alan Pankratz, 1983) of parsimonious, stationary, invertible, significant estimated coefficients and statistically independent and normally distributed residuals. ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated to fit the ARIMA models.

Autoregressive process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$;

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

The general form of ARIMA model of order (p, d, q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where Y_t is bananaproduction, ε_t 's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; d is the fraction differenced while interpreting AR and MA and ϕs and θs are coefficients to be estimated.

Trend Fitting : The Box-Ljung Q statistics was used to transform the non-stationary data into stationarity data and also to check the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and BIC were used. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics were computed as below:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right]^{1/2} ; \; MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right| \text{and BIC(p,q) = ln v*(p,q)+(p+q) [ln(n) / n]}$$

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.

$$Q = \frac{n(n+2)\sum_{i=1}^{k} rk^2}{(n-k)}$$

where n is the number of residuals and rk is the residuals autocorrelation at lag k.

In this study, the data on banana (Musaparadisiaca Linn) production in India were collected from the Agricultural Statistics at a Glance 2020*, Directorate of Economics and Statistics, Department of Agriculture, Government of India for the period from 1961 to 2019 and were used to fit the ARIMA model to predict the future production.

Table 1. Actual BananaProduction (million tons) in India

Year	Production	Year	Production	Year	Production	Year	Production
1961	2.26	1976	3.76	1991	7.85	2006	21.00
1962	2.43	1977	4.25	1992	8.52	2007	23.82
1963	2.60	1978	4.56	1993	9.95	2008	26.22
1964	2.68	1979	4.27	1994	10.69	2009	26.47
1965	3.27	1980	4.35	1995	10.18	2010	29.78
1966	3.41	1981	4.58	1996	10.30	2011	28.46
1967	3.20	1982	4.22	1997	13.34	2012	26.51
1968	3.13	1983	4.65	1998	15.10	2013	27.58
1969	3.17	1984	5.25	1999	16.81	2014	29.72
1970	2.90	1985	5.39	2000	14.14	2015	29.22

1971	3.37	1986	5.71	2001	14.21	2016	29.14
1972	3.19	1987	5.92	2002	13.30	2017	30.48
1973	3.17	1988	5.99	2003	13.86	2018	30.81
1974	3.27	1989	6.41	2004	16.74	2019	30.46
1975	3.41	1990	7.15	2005	18.89		

*Source: Directorate of Economics and Statistics, Department of Agriculture, Government of India

Results and Discussion

In this study, the data for banana production in India is collected from the period 1961 to 2019 is given in Table 1. To fit an Autoregressive model, Autoregressive process for any variable involves four steps: identification, estimation, diagnostic and forecasting. ARIMA (p,d,q) model is fitted to check stationarity through examining the graph or time plot of the data. Figure 3 reveals that the data is non-stationary. The autocorrelation and partial autocorrelation coefficients of various orders of Y_t are computed Table 2. The graphs of ACF and PACF are given in Figure 4.

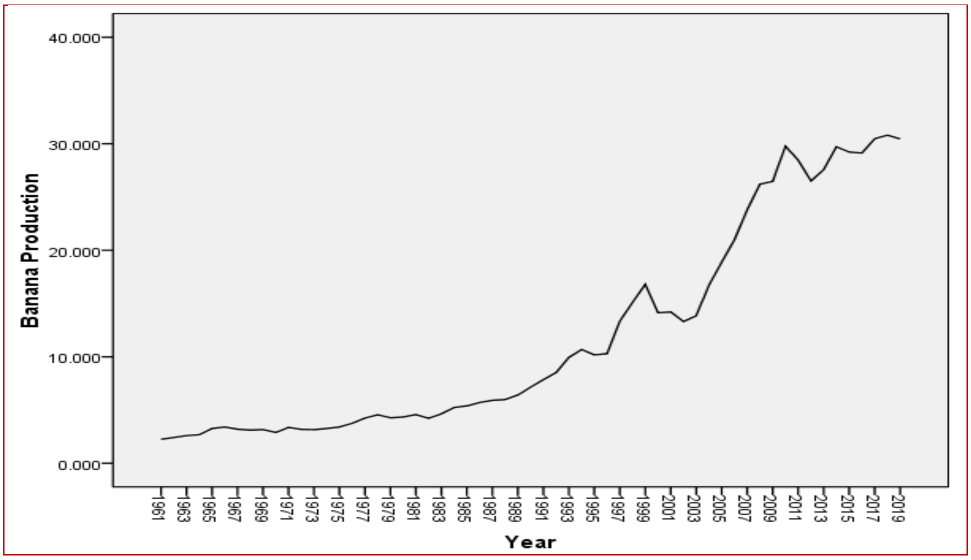

Figure 3 - Time plot of Banana Production

Table 2 - ACF and PACF of Banana Production

Lag	**AC**	**Std. Error^a**	**Box-Ljung Statistic**	**PAC**	**Std. Error**	**Lag**	**AC**	**Std. Error^a**	**Box-Ljung Statistic**	**PAC**	**Std. Error**
	Value	**Df**	**Sig.^{b}**	**Value**	**Df**		**Value**	**Df**	**Sig.^{b}**	**Value**	**Df**
1	0.954	0.127	56.503	0.954	0.130	**17**	0.114	0.108	390.543	0.062	0.130
2	0.902	0.126	107.902	-0.093	0.130	**18**	0.083	0.107	391.146	-0.030	0.130
3	0.849	0.125	154.246	-0.035	0.130	**19**	0.047	0.105	391.349	-0.090	0.130
4	0.798	0.124	195.886	-0.009	0.130	**20**	0.011	0.104	391.361	-0.041	0.130

5	0.745	0.122	232.831	-0.050	0.130	**21**	-0.035	0.103	391.478	-0.153	0.130
6	0.690	0.121	265.119	-0.048	0.130	**22**	-0.079	0.101	392.081	-0.035	0.130
7	0.638	0.120	293.328	0.014	0.130	**23**	-0.116	0.100	393.429	0.021	0.130
8	0.588	0.119	317.691	-0.034	0.130	**24**	-0.144	0.099	395.551	0.051	0.130
9	0.529	0.118	337.830	-0.122	0.130	**25**	-0.172	0.097	398.667	-0.078	0.130
10	0.461	0.117	353.468	-0.127	0.130	**26**	-0.201	0.096	403.064	-0.052	0.130
11	0.402	0.115	365.600	0.062	0.130	**27**	-0.228	0.094	408.900	0.007	0.130
12	0.339	0.114	374.425	-0.100	0.130	**28**	-0.251	0.093	416.209	0.006	0.130
13	0.281	0.113	380.625	0.012	0.130	**29**	-0.272	0.091	425.062	0.013	0.130
14	0.230	0.112	384.851	0.035	0.130	**30**	-0.289	0.090	435.446	0.050	0.130
15	0.185	0.111	387.644	0.015	0.130	**31**	-0.304	0.088	447.355	-0.047	0.130
16	0.146	0.109	389.425	0.012	0.130	**32**	-0.317	0.087	460.771	-0.032	0.130

[a] The underlying process assumed is independence (white noise).

[b] Based on the asymptotic chi-square approximation.

The models and corresponding BIC values are given in Table 3. The value of normalized BIC is 0.385 and R Squared value is 0.988. So the most suitable model for banana production is ARIMA(0,1,1) as this model has the lowest BIC value.

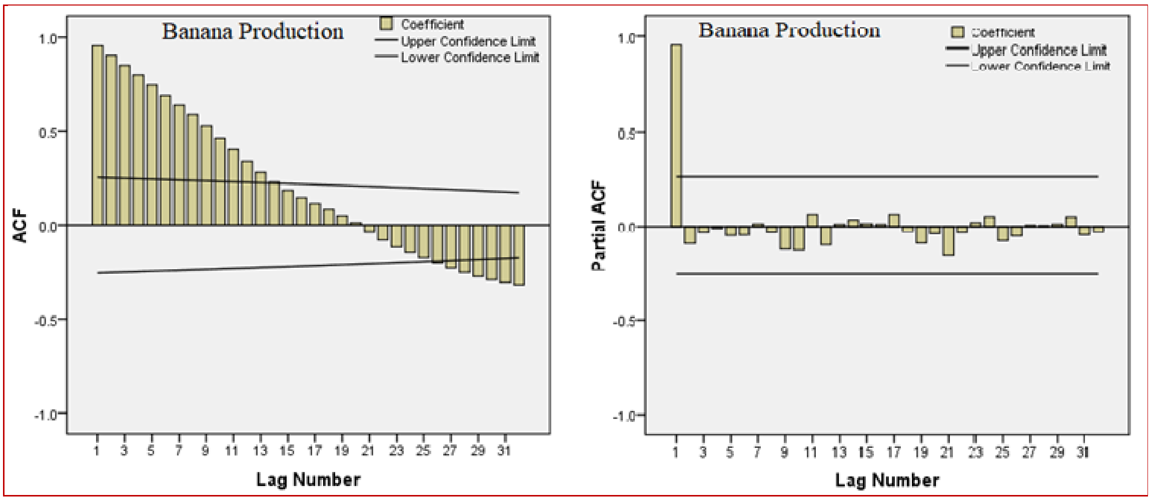

Figure 4 ACF and PACF of differenced data

Table - 3 BIC values of ARIMA(p,d,q)

ARIMA (p,d,q)	BIC Values
0,1,1	**0.385**
0,1,2	0.47
1,1,0	0.392
1,1,1	0.471
1,1,2	0.486
2,1,0	0.469
2,1,1	0.557
2,1,2	0.58
3,1,0	0.557
3,1,1	0.646
3,1,2	0.665

Model Estimation: Model parameters were estimated and reported in Table 4and Table 5. The model verification is concerned with checking the residuals of the model to improve on the chosen ARIMA (p,d,q). This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders, up to 32 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 6. This proves that the selected ARIMA model is an appropriate model.

Table 4 - Estimated ARIMA Model of Banana Production

	Estimate	SE	T	Sig.
Constant	-30.379	20.510	-1.481	0.144
MA 1	-0.214	0.133	-1.615	0.112

The ACF and PACF of the residuals are given in Figure 5. It also indicates 'good fit' of the model. So the fitted ARIMA model for the banana production data is

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$Y_t = -30.379 + 0.214\varepsilon_{t-1} + \varepsilon_t$$

Table 5 - Estimated ARIMA Model Fit Statistics

ARIMA (p,d,q)	Stationary R^2	R^2	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC
0,1,1	**0.094**	**0.988**	**1.091**	**5.928**	**24.819**	**0.699**	**3.509**	**0.385**
0,1,2	0.098	0.988	1.100	5.935	24.647	0.697	3.484	0.470
1,1,0	0.088	0.988	1.095	5.943	24.747	0.703	3.499	0.392
1,1,1	0.096	0.988	1.100	5.922	24.828	0.698	3.510	0.471
1,1,2	0.160	0.989	1.071	6.257	23.302	0.690	3.294	0.486
2,1,0	0.098	0.988	1.099	5.922	24.082	0.695	3.405	0.469
2,1,1	0.099	0.988	1.109	5.928	24.325	0.695	3.439	0.557

2,1,2	0.156	0.989	1.084	6.384	23.369	0.693	3.304	0.580
3,1,0	0.098	0.988	1.109	5.928	24.253	0.695	3.429	0.557
3,1,1	0.099	0.988	1.120	5.914	23.783	0.694	3.362	0.646
3,1,2	0.161	0.989	1.091	6.411	23.389	0.696	3.307	0.665

Table 6 - Residual of ACF and PACF of Banana Production

Lag	**ACF**		**PACF**		**Lag**	**ACF**		**PACF**	
	Mean	**SE**	**Mean**	**SE**		**Mean**	**SE**	**Mean**	**SE**
1	-0.009	0.131	-0.009	0.131	**17**	0.056	0.150	0.051	0.131
2	-0.065	0.131	-0.065	0.131	**18**	-0.071	0.150	-0.098	0.131
3	-0.006	0.132	-0.008	0.131	**19**	-0.047	0.151	-0.060	0.131
4	-0.011	0.132	-0.016	0.131	**20**	0.043	0.151	-0.022	0.131
5	-0.245	0.132	-0.247	0.131	**21**	0.009	0.151	0.106	0.131
6	0.013	0.139	0.005	0.131	**22**	-0.107	0.151	-0.185	0.131
7	0.032	0.140	-0.001	0.131	**23**	0.002	0.153	-0.010	0.131
8	-0.168	0.140	-0.186	0.131	**24**	0.018	0.153	-0.025	0.131
9	0.088	0.143	0.088	0.131	**25**	-0.079	0.153	-0.064	0.131
10	-0.043	0.144	-0.139	0.131	**26**	-0.026	0.153	-0.016	0.131
11	0.086	0.144	0.102	0.131	**27**	0.005	0.153	-0.124	0.131
12	-0.032	0.145	-0.042	0.131	**28**	-0.016	0.153	-0.051	0.131
13	0.070	0.145	-0.010	0.131	**29**	-0.027	0.153	-0.041	0.131
14	-0.154	0.146	-0.119	0.131	**30**	0.059	0.153	-0.013	0.131
15	-0.029	0.149	-0.073	0.131	**31**	-0.031	0.154	-0.005	0.131
16	0.099	0.149	0.113	0.131	**32**	0.019	0.154	-0.103	0.131

Forecasting: Forecasted value of banana production (quantity in million tons) for the year 2020 through 2025 respectively given by 31.16, 32.12, 33.09, 34.07, 35.07 and 36.09 are given in Table 7. To assess the forecasting ability of the fitted ARIMA (p,d,q) model, important measures of the sample period forecasts' accuracy were computed. This measure indicates that the forecasting inaccuracy is low. Figure 6 shows that the actual and forecasted value of banana production data with 95% confidence limits.

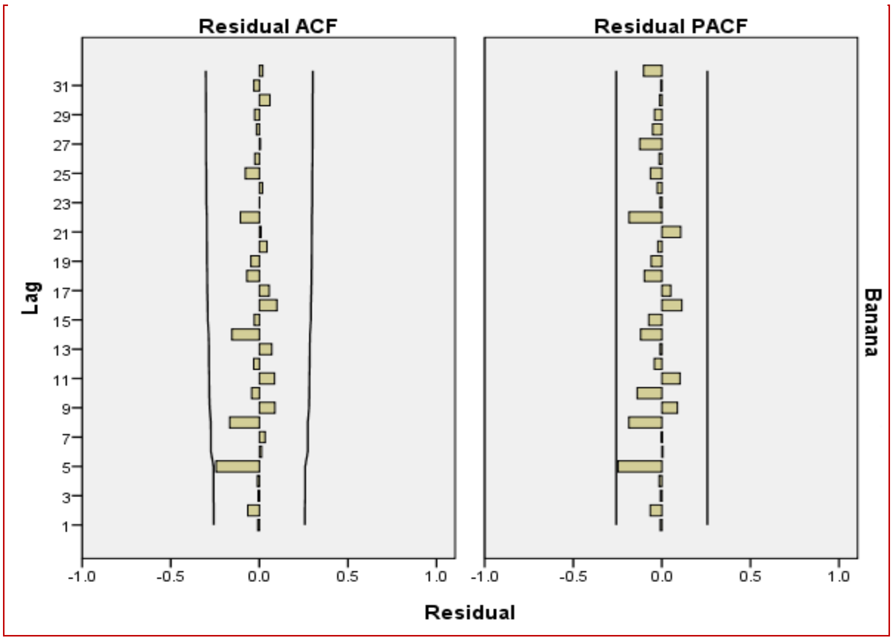

Figure 5 - Residuals of ACF and PACF

Table 7 - Forecast of Banana Production

Year	Predicted	LCL	UCL
2020	31.16	28.97	33.35
2021	32.11	28.67	35.56
2022	33.09	28.74	37.43
2023	34.07	28.98	39.17
2024	35.07	29.33	40.82
2025	36.09	29.76	42.42

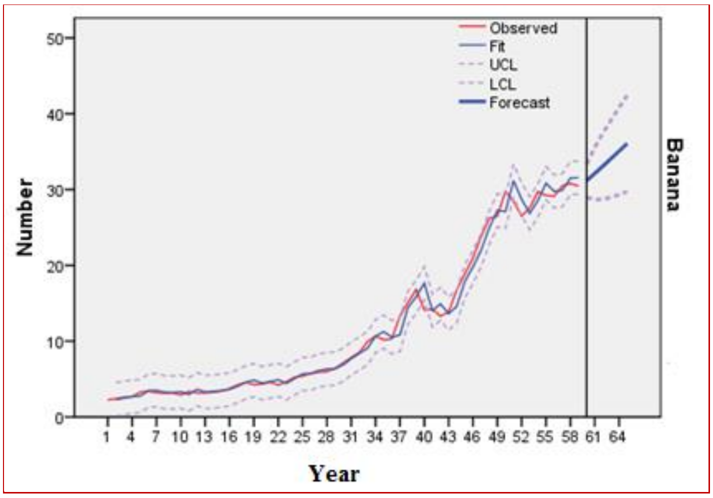

Figure 6 - Actual and Estimate of Banana Production

Conclusion

The most appropriate ARIMA model for banana production forecasting of data was found to be ARIMA (0,1,1). From the time series data, it can be found that forecasted production would increase to 36.09million tons in 2025 from 30.46 million tons in 2019 in India for using time series data from 1961 to 2019 on banana production, this study provides an evidence on future banana production in the country, which can be considered for future policy making and formulating strategies for augmenting and sustaining banana production in India.

References

1. Agricultural Statistics at a Glance (2020), Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India. https://desagri.gov.in/ document-report/agricultural-statistics-at-a-glance-2020.

2. Akaike, H. (1983). Statistical Predictor Identification. Annals of Institute of Statistical Mathematics, **22:** 203-270. https://link.springer.com/article/10.1007/BF02506337.

3. Alan Pankratz, (1983). Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wiley & Sons, New York. https://onlinelibrary.wiley.com/doi/book/ 10.1002/9780470316566.

4. Bhola Nath, Dhakre, D.S. and Debasis Bhattacharya, (2018). Forecasting wheat production in India: An ARIMA modelling approach. Journal of Pharmacognosy and Phytochemistry, **8(1)**:2158-2165.https://www.researchgate.net/publication/331471229.

5. Borkar, Prema and Tayade, P.M., (2016). Forecasting of Cotton Production in India Using ARIMA Model. International Journal of Research in Economics and Social Sciences, **6(5):** 1-7. https://euroasiapub.org/wp-content/uploads/2016/10/1ESSMay-3546-1.pdf.

6. Box, G.E.P. and Jenkins, G.M., (1976). Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day, California, USA. https://link.springer.com/chapter/10.1057/9781137291264_6.

7. Box, G.E.P. and Pierce, D.A., (1970). Distribution of Residual Autocorrelations in ARIMA Models. J. American Stat. Assoc.,**65**: 1509-1526. https://www.jstor.org/stable/2284333?seq=1.

8. Hemavathi, M. and Prabakaran, K., (2018). ARIMA Model for Forecasting of Area, Production and Productivity of Rice and Its Growth Status in Thanjavur District of Tamil Nadu, India. Int. J. Curr. Microbiol. App. Sci., **7(2)**: 149-156.https://doi.org/10.20546/ijcmas.2018.702.019.

9. Kour Satvinder, Pradhan, Ranjit Kumar Paul, U.K. and Vaishnav, P.R., (2017). Forecasting of Pearl Millet Productivity in Gujarat Under Time SeriesFramework. Economic Affairs, **62(1)**: 121-127. https://www.researchgate.net/publication/319122005.

10. Md. Moyazzem Hossain andFaruq Abdulla, (2016). Forecasting Potato Production in Bangladesh by ARIMA Model. Journal of Advanced Statistics, **1(4)**:191-198.https://www.academia.edu/30930688.

11. Slutzky, E., (1973). The summation of random causes as the source of cyclic processes. Econometrica, **5**: 105-146. https://www.scribd.com/document/388512380.