Modelling and Forecasting of Area, Production and Productivity of Tomatoes in Haryana and India

Article · April 2022

DOI: 10.48165/IJEE.2022.58231

Parveen Kumar Nimbrayan
CCS Haryana Agricultural University

Muhammed Jaslam P K
University of Idaho

Aniket Vilas Chandanshive
Indian Agricultural Research Institute

Research Note

Indian Journal of Extension Education
Vol. 58, No. 2 (April–June), 2022, (205-208)

ISSN 0537-1996 ***(Print)***
ISSN 2454-552X ***(Online)***

Modelling and Forecasting of Area, Production and Productivity of Tomatoes in Haryana and India

Parveen Kumar Nimbrayan[1], P. K. Muhammed Jaslam[2]* and Aniket Chandanshive[3]

[1]Directorate of Human Resource Management, [2]Department of Math and Stat, CCS HAU, Hisar-125004, Haryana, India
[3]Officer Incharge, Tomato Improvement Scheme, MPKV Rahuri-413722, Maharashtra, India
*Corresponding author email id: jaslam.stat@hau.ac.in

ARTICLE INFO

Keywords: AIC, ARIMA, Linear Model, Nonlinear model and SBIC

http://doi.org/10.48165/IJEE.2022.58231

ABSTRACT

An effort was made to investigate the behaviour of the area, production and productivity of tomato crop in the Haryana and India. For the purpose modelling and forecasting, linear trend, exponential trend, quadratic trend, S-curve trend, ARIMA modelling techniques were used and analysed the available information from 1991 to 2018. The results show that there will not be a significant increase in tomato productivity in Haryana, but it will raise yield in India. The total production of tomatoes in Haryana will be 1029 thousand tons by 2024 and the current production (2018-19) is 643.55 thousand tons and an increase of 4043 tons can be achieved in 2024 in India. It is noteworthy that although the area under tomatoes will increase in the near future in Haryana, but productivity remains the same. Productivity in India may increase in the coming years, although the area under cultivation remains the same.

INTRODUCTION

Vegetables are one of the most important parts of the horticulture sector in both Haryana and India. Vegetables productivity in Haryana and India has increased over the last several years. Various factors have led to an increase in the area and production of vegetables such as an increase in per capita income, a rise in health awareness, a shift in farmers' growth of higher-value vegetables due to higher returns etc. Favourable income-elasticity of demand has also contributed in increasing trend in vegetable production worldwide. Vegetables are an essential aspect of Indian agriculture because of its short growing season, better yield, nutrient diversity, and economic viability. India is the world's second-largest producer of horticulture crops. Among vegetables, tomato (*Solanum lycopersicum*) is the important vegetable crops in India with 2nd rank in the production and area worldwide. Globally, China is largest producer accounts for 27.8 per cent followed by India (11.2%), respectively (Kumar et al., 2016; Harisha et al., 2019; Gupta et al., 2021). Haryana ranks 12th in tomato production with 753.72 thousand tonnes (Horticulture at a glance, 2018). In Haryana, tomato cultivated in open and protected cultivation both. Tomato production is higher in protected as compared to open cultivation (Kumar et al., 2017; Nimbrayan et al., 2018). The objective of this study is that of understanding the trend in production and productivity in the Haryana area compared to India, this analysis allows better policy decisions in terms of food security and allocation of land use too.

The ARIMA model is most used for forecasting time series used following the literature. The auto regressing integrated moving average (ARIMA) model is used to forecast using a univariate time series model. The Box and Jenkin model is another name for it (Box et al., 1976). The ARIMA model is one of the most prominent stochastic time series models for forecasting using observed data with little forecast error. In the literature, several statistical and economic forecasting models have been created that may be used to forecast a variety of topics, including agricultural output, marketing, demand, trade, and so on (Hanke & Wichern, 2008). Just to mention a few works; Verma et al., (2015); Kumar et al., (2019);

Received 09-02-2022; Accepted 15-02-2022

Naidu et al., (2018) etc. are working on several forecasting facets in agriculture in India. Forecasters should explore a variety of approaches and compare their performance across a random sample of series, according to Fildes & Lusk (1984). In agriculture, which is full of uncertainty, reliable and timely predictions give valuable and practical recommendations for successful, foresighted, and perceptive planning. Forecasting crop area, production, and productivity are critical parameters in our model for establishing a support policy choice on food security, optimal land use allocation, technical issues, and environmental challenges, among other things.

METHODOLOGY

The time series data of Haryana and India, period from 1990-91 to 2018-19 of tomato have been used to study the growth trends. The time series data have been taken from Department of Horticulture, Haryana, Department of Agriculture, Cooperation & Farmers Welfare, Government of India, Indian Institute of Vegetable Research, ICAR.

The study tried to fit univariate forecasting models such as linear trend, nonlinear trend and ARIMA models to predict vegetable production. Model diagnostic checking were doing through minimum of root mean squared error (RMSE), Akaike Information Criteria (AIC) and Schwarz Bayesian Information Criteria (SBIC) etc. Linear model and nonlinear model are presented in the next sub-section.

Linear trend is a basic function that is defined as a straight line that runs across multiple points on a time series graph and has a consistent pattern.

$$Y_n = c + bT_n$$

Where, c is the constant of production at base period and b is the coefficient of trend line direction. Method least squares can be applied to find these coefficients.

$$b = \frac{N\sum Y_n T_n - \sum Y_n \sum T_n}{N\sum T_n^2 - \sum (T_n)^2} \text{ and } c = \bar{T}_n - b\bar{Y}_n$$

There are several nonlinear trends, and this study uses three different trends:

Quadratic trend $Y_n = c + bT_n + rT^2_n$

Exponential trend $Y_n = cb^n$

S-curve or Logistic trend $Y_n = 1 / 1+e^{c+bTn}$

Non-linear equations can be solved using linearization, Newton Raphson methods etc. see Weisberg (2005).

The basic goal of fitting an Autoregressive Integrated Moving Average (ARIMA) model is to identify the time series' stochastic process and properly forecast future values. There are two types of stochastic processes: stationary and non-stationary. The first point to keep in mind is that most time series are non-stationary, while ARIMA models only consider stationary time series (Chatfield & Yar, 1988). Because ARIMA models only consider stationary time series, the first stage of the Box-Jenkins model involves extracting first order differences to convert non-stationary time series to stationary time series.

To measure the adequacy of the fitted model RMSE and AIC value are utilized and it can be computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum e_n^2}$$

$$\text{AIC} = 2 \ln(\text{RMSE}) + 2k / N$$

$$\text{SBIC} = \ln(N)\, k - 2 \ln(\hat{L})$$

Where, k is the number of estimated model parameters and $\hat{L}$ is the maximized value of the likelihood function of the model.

RESULTS AND DISCUSSION

The mean, maximum (Max.) and minimum (Min.) values, as well as other statistical features, are displayed in descriptive statistics. Table 1 shows that during the research period, there was a substantial variation in the minimum and maximum value of area, production, and productivity of tomato in Haryana and India (Kalia et al., 2021). The standard deviation value (Stdev.) for all variables is too high for tomato production, resulting in a volatile pattern. Even if the means are substantially different, the coefficient of variation (CV) is a helpful statistic for comparing the degree of variation from one data series to another. The standard deviation to mean ratio is used to compute it. In terms of area, production, and productivity, Haryana has a higher CV than India. Skewness is a metric for symmetry, or more specifically, the lack of it. If a distribution, or data collection, looks the same to the left and right of the centre point, it is said to be symmetric (Skewness value around zero). Kurtosis is a measure of how heavy-tailed or light-tailed the data are in comparison to a normal distribution.

The area data set in case of Haryana and India, production data set in case of Haryana and productivity data set in case of India found to be following platykurtic distribution. All the data set are more or less symmetrically distributed around the central value. The results of fitting different models to the data are compared in Table 2. The first model shows results for Haryana area; while the second model reports results for India. The model with based on RMSE, AIC and SBIC was selected and used to generate the forecast values.

Table 1. Descriptive statistics for tomato in Haryana and India from 1990-91 to 2018-19

	Min.	Max.	Mean	Stdev.	CV (%)	Skewness	Kurtosis
Haryana							
Area	5	35	16.93	9.772	57.729	.255	-1.514
Production	93	746	307.57	206.720	67.212	.952	-.490
Productivity	11	25	18.31	3.865	21.113	-.426	-.957
India							
Area	289	882	571.65	199.272	34.859	.269	-1.345
Production	4243	20708	11125.70	5569.049	50.056	.501	-1.358
Productivity	14	26	18.55	3.343	18.027	.925	-.058

Table 2. Comparison of different time series models based on selection criteria

Model	Area			Production			Productivity		
	RMSE	AIC	SBIC	RMSE	AIC	SBIC	RMSE	AIC	SBIC
Haryana									
Linear Trend	2375	15.684	15.779	84936	22.837	22.932	3.935	2.878	2.972
Quadratic trend	2325	15.623	15.764	53956	21.999	22.140	3.309	2.600	2.742
Exponential trend	3244	16.307	16.401	57505	22.057	22.152	3.962	2.891	2.986
S-curve trend	8086	18.134	18.228	186617	24.412	24.506	3.928	2.874	2.969
ARIMA	2297	15.618	15.712	59460	22.124	22.218	3.344	2.414	2.414
India									
Linear trend	67	8.568	8.663	1609	14.910	15.006	1.474	0.919	1.015
Quadratic trend	68	8.676	8.819	1269	14.506	14.649	0.959	0.132	0.274
Exponential trend	76	8.827	8.922	1332	14.533	14.628	1.331	0.714	0.810
S-curve trend	156	10.243	10.338	4664	17.038	17.134	2.975	2.323	2.418
ARIMA	54	7.985	7.985	1152	14.170	14.218	1.208	0.450	0.497

Table 3. Forecasted value of area, production, and productivity of tomato in Haryana and India

Year	Area (000'ha)	Lower limit@95%	Upper limit@95%	Production (tonnes/ha)	Lower limit@95%	Upper limit@95%	Productivity (Tonnes/ha)	Lower limit@95%	Upper limit@95%
Haryana									
2019-20	34.64	29.91	39.38	800	670491	928928	21.37	14.52	28.22
2020-21	34.58	29.20	39.96	854	719422	987997	21.37	11.68	31.06
2021-22	36.92	30.29	43.55	910	769459	1050060	21.37	9.51	33.23
2022-23	38.16	30.68	45.63	968	820572	1115150	21.37	7.67	35.07
2023-24	40.06	31.63	48.49	1029	872746	1183280	21.37	6.06	36.68
India									
2019-20	778	666.82	889.19	20145.70	17777.80	22513.50	26.58	24.27	28.90
2020-21	778	620.76	935.24	20923.20	17509.30	24337.10	27.51	25.10	29.92
2021-22	778	585.42	970.58	21730.80	17467.20	25994.40	28.48	25.95	31.01
2022-23	778	555.63	1000.37	22569.50	17548.00	27591.00	29.48	26.82	32.15
2023-24	778	529.38	1026.62	23440.60	17713.00	29168.30	30.53	27.71	33.34

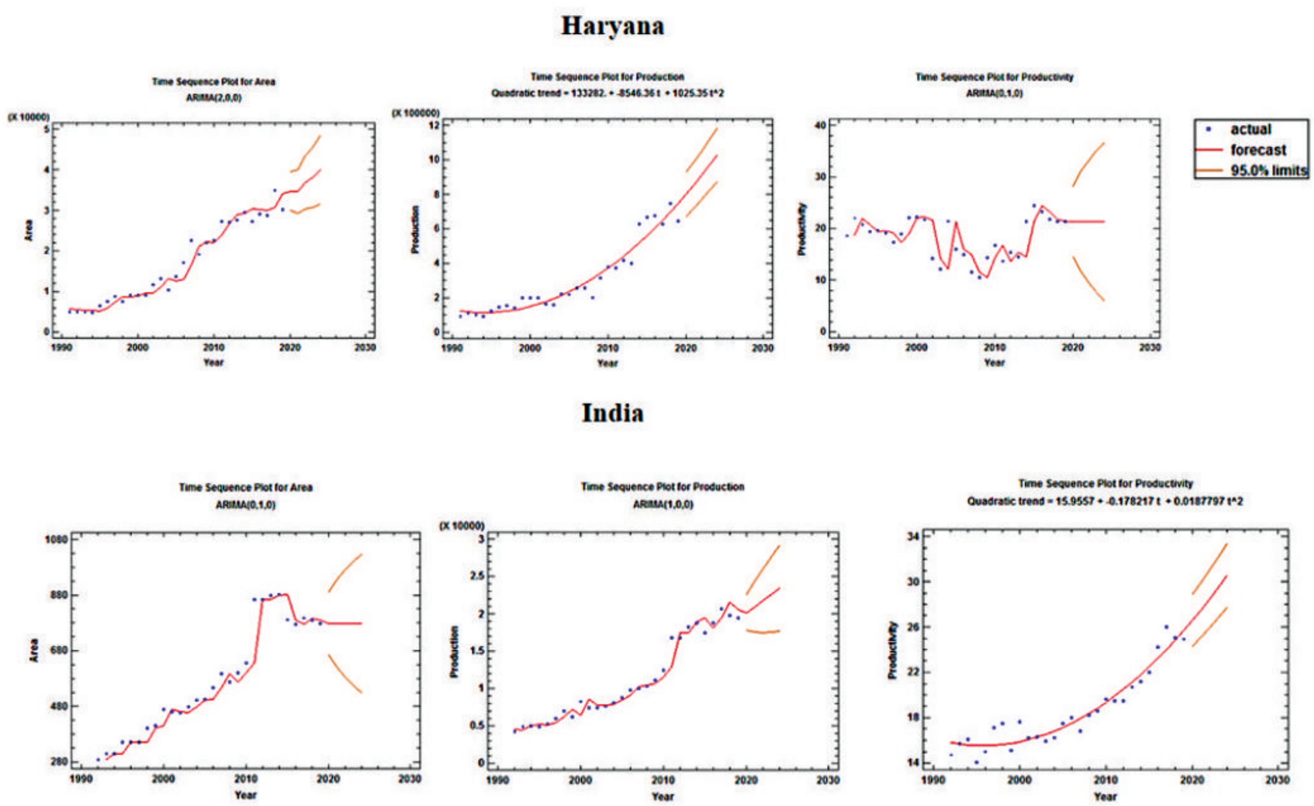

Figure 1. Time series forecast plots for tomato in Haryana and India

ARIMA model was shown to be better suited for forecasting Area [ARIMA (2, 0, 0)] and Productivity [ARIMA (0, 1, 0)] in Haryana. This approach posits that a parametric model linking the most current data value to prior data values and historical noise provides the best forecast for future data. In Haryana, however, tomato output follows a quadratic pattern (133282- 8546.36 t + 1025.35 t^2). This model assumes that the quadratic regression curve fitted to all prior data provides the best forecast for future data. In India, ARIMA model found more suitable to forecast Area [ARIMA (0, 1, 0)] and Production [ARIMA (1, 0, 0)] while Quadratic trend (15.9557 + -0.178217 t + 0.0187797 t^2) is best fit for time series data on tomato productivity in India. Table 3 displays the anticipated values from the different fitted models, as well as the predictions' 95.0 per cent prediction limitations. If the fitted model is acceptable for the data, these limits reveal where the real data value at a given future time is expected to be with 95.0 per cent confidence. Figure 1 shows several projected values and their patterns. It is interesting to note that in the case of in the case of tomato productivity in Haryana and Area under cultivation in India, the selected models provided a same forecast value for all the forecast years that means the forecast value depends solely on the recent available data. The rest of variables indicates increasing trends, although nonlinear in many cases.

CONCLUSION

As results shows, it is observed from results that the CV (%) found high in Haryana as compared to India with respect to Area, Production and productivity. In Haryana, ARIMA model found more suitable to forecast Area [ARIMA (2, 0, 0)] & Productivity [ARIMA (0, 1, 0)]. In India, ARIMA model found more suitable to forecast Area [ARIMA (0, 1, 0)] and Production [ARIMA (1, 0, 0)] while Quadratic trend (15.9557 + -0.178217 t+0.0187797 t^2) is best fit for time series data on tomato productivity in India. The results of the study also show that there will not be a significant increase in tomato crop productivity in Haryana, while a yield in India may upsurge significantly. There is more scope of increasing yield in Haryana using modern agricultural practices, high yielding varieties etc.

REFERENCES

Anonymous. (2018). Horticulture at a glance.

Box, G. E. P., & Jenkins, G. M. (1976) Time Series Analysis: Forecasting and Control. *Revised Edition, Holden Day, San Francisco*.

Chatfield, C. & Yar, M. (1988). Holt-Winters Forecasting: Some Practical Issues. *Journal of the Royal Statistical Society*, *(The Statistician), 37*, 129-140.

Fildes, R., & Lusk, E. J. (1984). The choice of a forecasting model. *Omega, 12*(5), 427-354.

Gupta, B. K., Dwivedi, S. V., Mishra, B. P., Mishra, D., Ojha, P. K., Verma, A. P., & Kalia, A. (2021). Adoption gap analysis in tomato cultivation in Banda District of Bundelkhand (U.P.). *Indian Journal of Extension Education*, *57*(4), 126-130.

Hanke, J. E., & Wichern, D. W. (2008). *Business Forecasting, Pearson Education, 9th edition*, New Delhi.

Harisha, N., Tulsiram, J., Meti, S. K., Chandargi, D. M., & Joshi, A. T. (2019). Extent of adoption of tomato cultivation practices among farmers under shade nets in Kolar District of Karnataka. *Indian Journal of Extension Education*, *55*(1), 28-33.

Kalia, A., Shukla, G., Mishra, G., Mishra, B. P., & Patel, R. R. (2021). Comparative trend analysis of mustard in Bundelkhand region, Uttar Pradesh and India. *Indian Journal of Extension Education*, *57*(1), 15-19.

Kumar, A., Deepankar, Jaslam, P. K. M., & Kumar, A. (2019). Wheat yield forecasting in Haryana: A time series approach. *Bulletin of Environment, Pharmacology and Life Sciences, 8*(3), 63-69.

Kumar, P., Chauhan, R. S., & Grover, R. K. (2016). Economics analysis of tomato cultivation under poly house and open field conditions in Haryana, India. *Journal of Applied and Natural Science, 8*(2), 846-848.

Kumar, P., Chauhan, R. S., Tanwar, N., & Grover, R. K. (2017). Status and constraints in vegetable cultivation under polyhouse in Haryana. *Advances in Bioresearch, 9*(2), 61-66.

Naidu, G. M., Reddy, B. R., & Murthy, B. R. (2018). Time series forecasting using Arima and neural network approaches. *International Journal of Agricultural and Statistical Sciences, 14*(1), 275-278.

Nimbrayan, P. K., Chauhan, R. S., Mehta, V. P., & Bhatia, J. K. (2018). A review on economic aspect of protected cultivation in India. *Research Trends in Horticulture Sciences*. Akinik Publications, New Delhi, pp 43–59.

Verma, U., Piepho, H. P., Hartung, K., Ogutu, J. O., & Goyal, A. (2015). Linear mixed modeling for mustard yield prediction in Haryana State (India). *Journal of Mathematics and Statistical Science,* pp 96-105.

Weisberg, S. (2005). *Applied Linear Regression, 3rd edition*. John Wiley & Sons, New York.