Research Article

Indian Journal of Extension Education
Vol. 58, No. 4 (October–December), 2022, (71-75)

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

Forecasting of Vegetable Production in Haryana by Ordinary Least Square Method and ARIMA Model

Vikash[1]* Satyveer Singh Meena[2] and R. K. Verma[3]

[1]Ph.D. Scholar, [2]Assistant Professor, Institute of Agri Business Management, [3]Professor & Head, Department of Extension Education Swami Keshwanand Rajasthan Agricultural University, Bikaner-334006, Rajasthan, India
*Corresponding author email id: vikash.js0705@hau.ac.in

ARTICLE INFO

Keywords: Vegetables, Ordinary least square method, ARIMA, Time series data

http://doi.org/10.48165/IJEE.2022.58415

ABSTRACT

India has the world's second-largest production of fruit and vegetables. In India, a variety of agro-climate zones with unique seasons allow for the cultivation of a diverse range of vegetables. From this study, it was attempted to estimate the production of vegetables in the upcoming year through means of ordinary least squares (OLS) method and ARIMA (Autoregressive Integrated Moving Average) model using secondary data. This study is based on time series data of vegetables from 1990-91 to 2020-21 in open field condition, which was taken from National Horticulture Board (NHB) website, while the area and production reports on cultivation of vegetables from the Horticulture Department, Haryana state of India. In the current context, it was fetched out that the forecasted value of the vegetables production for the year 2021-22 will be 7540814.31 tonnes in open field condition. Further, for predicting the area and production of vegetables in Haryana, ARIMA (1, 2, 1) was fitted after experimenting with various lags of the moving average and autoregressive procedures.

INTRODUCTION

India has the world's second-largest production of fruit and vegetables next to China. In India, a variety of agro-climate zones with unique seasons allow for the cultivation of a diverse range of vegetables. Vegetables are the best sources of vitamins, dietary fibre, phytochemicals, and minerals. Vegetables with a shorter lifespan and increased yield have given farmers larger financial benefits (Kumar et al., 2021). The primary challenges cited by homestead vegetable growers included a lack of water resources, a high prevalence of pests and diseases, price volatility, high labour costs, high cost of cultivation, labour shortage, lack of timely access to inputs, and no guarantee of premium prices for organic products (Chandran, 2020).

The technique of forecasting allows for future predictions to be made using data from the past and present as well as trend analysis. The projection of crop output is a crucial factor in determining support policies for concerns like food security, efficient land use allocation, technology, and the environment. Verma et al., (2015); Kumar et al., (2016, 2017-a&-b & 2019) and many other studies have been done to improve forecasting utilising various pre-harvest forecasting methodologies, and they encourage forecasters to take a variety of methods into consideration and compare their performance across a variety of series. The current study attempted to predict Haryana's vegetable output in this context, providing the general public, academics, and decision-makers with long-term information on state-wide vegetable production. According to the study of Prakash et al., (2022), to make informed decisions about sweet potato production, marketing, and consumption, sweet potato producers, governmental organisations, and other stakeholders in the sweet potato value chain will benefit greatly from the anticipated price of the crop. Nimbarayn et al., (2022) conducted a study on modelling and forecasting of area, production and productivity of tomatoes. The results revealed that Haryana's tomato productivity won't grow significantly, but India's yield will increase. In the current study,

Received 21-06-2022; Accepted 20-09-2022

an attempt was made to estimate the production of vegetables in the upcoming year through means of ordinary least squares (OLS) method using secondary data and forecasting was done by using the SPSS 26 statistical tool.

METHODOLOGY

This study was based on secondary data (time series data of vegetables from 1990-91 to 2020-21 in open field condition), which was taken from the National Horticulture Board (NHB) website, whereas area and production reports on cultivation of vegetables was obtained from Horticulture Department, Haryana state of India. Ordinary least square (OLS) method was used to forecast the production of vegetables in the upcoming year by using the SPSS 26 statistical tool. OLS is the most prominent method used in estimating the relationship in econometric models. It can be understood from simple regression model, which establishes the connection between two variables, where one is dependent and other is independent variable, related with a linear function (Koutsoyiannis, 2015). It is assumed that the variables are connected with the simplest mathematical form that is explained by the following equation

$$Y = a + b X$$

Where, Y = Production, X = Area Here, Y is a dependent variable and X is an independent variable, the parameters are a and b. Our main motive was to get the value of a and b which can be done through simple curve estimation in regression through SPSS 26 (Statistical Package for Social Science).

Box and Jenkins are credited with popularising the univariate ARIMA technique, and models created using this method are regarded to as univariate Box-Jenkins (UBJ) models. Identification, parameter estimation, diagnostic checking, and forecasting are the first three phases in univariate ARIMA modelling (Goyal et al., 2021). The ARIMA (p, d, q) model's basic functional form is:

$$\phi p(B)\ \Delta^d\ y_t = c + \theta_q\ (B) a_t$$

where, y = Agricultural Export, B = Lag operator, a = Error term (Y-yˆ, where yˆ is the estimated value of Y), t = time subscript, ϕp (B) = non-seasonal AR i.e. the autoregressive operator, represented as a polynomial in the back-shift operator, D^d = non-seasonal difference

θ_q (B) = non-seasonal MA i.e. the moving-average operator, represented as a polynomial in the back-shift operator, ϕ's and θ's are the parameters to be estimated.

RESULTS AND DISCUSSION

Forecasting the Production of Vegetables for Year 2021-22

The analysis of data was done through the SPSS 26, where the Ordinary Least Square estimate to forecast the production for the year 2021-22. In the equation for defining the production of vegetables (forecasted), it was assumed that there is a relation between variables with the simplest possible mathematical form. Here in this Function, researchers are interested to see the effect of area on vegetable production by keeping another factor as constant. As it is well known that the production of vegetables does not depend on a single factor than many factors but in this study, the researchers were interested to see the effect of area on vegetable production. By accessing the data of area and production in SPSS 26, the list of model summary and parameter estimates was taken out.

Table 1 shows the model summary and parameter estimates of the linear function. The R square value was 0.937 which is significant at 0.05 level of significance. The value of constant is -197384.526 (Normally it should be positive, but here a turn up with a –ve (negative) sign, so the –ve (negative) part of the production function, since a negative does not make sense in economics as per suggested by Koutsoyiannis, 2015 in his book the Theory of Econometrics. By getting the values of parameters when the calculation was done then the forecasted value of the vegetables production for the year 2021-22 will be 7540814.31 tonnes.

Regression line estimation

The Table 2 shows that the compound function is explaining better other than linear, cubic and quadratic to the overall production function where the value of R Square 0.976 is very significant at

Table 1. Model Summary and Parameter Estimates of Linear Function
Dependent Variable: production

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant (a)	b1
Linear	.937	417.941	1	28	.000	-197384.526	244780.993

Source: Researcher's computation from secondary data

Table 2. Model Summary and Parameter Estimates of Different Functions
Dependent Variable: production

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.937	417.941	1	28	.000	-197384.526	244780.993		
Quadratic	.967	392.502	2	27	.000	732219.221	70480.291	5622.603	
Cubic	.967	252.112	3	26	.000	702887.026	80983.467	4789.302	17.920
Compound	.976	1162.577	1	28	.000	864762.431	1.081		

Source: Researcher's computation from secondary data

0.00 (significance level). It can be interpreted that the compound function explaining the 97 per cent variability in the production which is the best explanation for the vegetable production function (Figure 1). Through SPSS 26 different plots are drawn for obtaining a line of fit in Figure 1 which depicted the linear (Li), quadratic (Qu), cubic (Cu) and compound (Co) fit of lines. The best fit line is represented by the compound function as shown in Figure 2. The best fit line of compound function representing the data of production which lies close to the lines are best described by the compound function line because the line is closer to the data values of vegetable production in different years. As the result obtained through production estimation it was found that the forecasted value of the vegetables production for the year 2021-22 will be 7540814.31 tonnes. This value was calculated through the coefficient values as shown in Table 3 which comes from SPSS 26 Table and the ordinary least square method equation number 1.

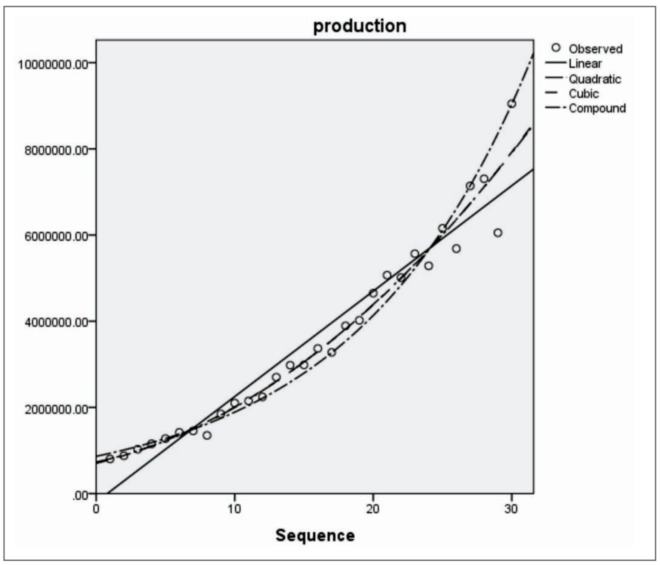

Figure 1. Regression line of diff. functions (Li, Qu, Cu, Co)

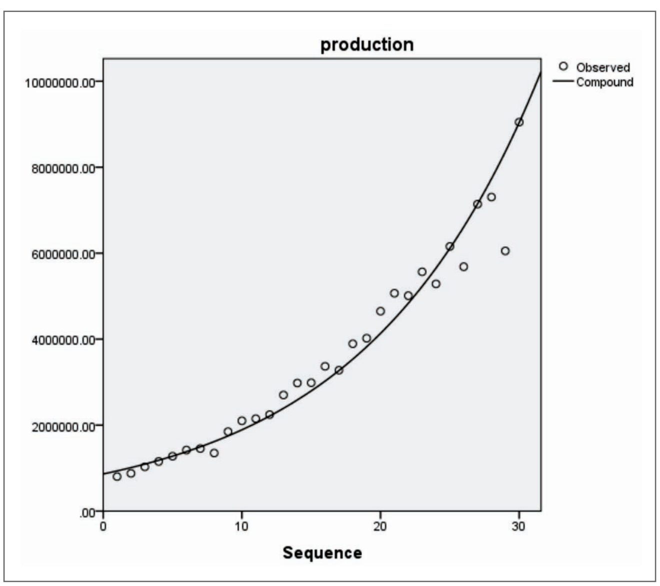

Figure 2. Regression line of compound function

Forecasting by using ARIMA Model

The ARIMA technique will be used to simulate India's exports of agricultural products. On the base of ACF and PACF analyses, the values of the parameters p, d, and q are determined. The area and production data of vegetables was discovered to be non-stationary, and order one differencing was all that was required to produce a suitable stationary series. In Table 3 and 4, the predicted ACF are displayed.

For predicting the area and output of vegetables in Haryana, ARIMA (1,2,1) was fitted after experimenting with various lags of the moving average and autoregressive processes. Past observations are averaged in the ARIMA model, although weights for more recent observations are bigger than weights for previous observations. The Marquardt algorithm (1963) has been used to minimise the sum of squared residuals. For this sort of data, a

Table 3. Autocorrelations: Area of Vegetables

Lag	Auto correlation	Std. error (a)	Box-Ljung Statistic Value	df	Sig (b)
1	.904	.180	27.861	1	.000
2	.820	.291	51.601	2	.000
3	.729	.358	71.026	3	.000
4	.631	.403	86.126	4	.000
5	.547	.434	97.905	5	.000
6	.453	.456	106.320	6	.000
7	.366	.470	112.043	7	.000
8	.268	.479	115.238	8	.000
9	.180	.484	116.740	9	.000
10	.089	.486	117.124	10	.000
11	-.004	.487	117.124	11	.000
12	-.084	.487	117.502	12	.000
13	-.166	.487	119.068	13	.000
14	-.223	.489	122.070	14	.000
15	-.289	.492	127.395	15	.000
16	-.337	.498	135.131	16	.000

a. The underlying process assumed is MA with the order equal to the lag number minus one. The Bartlett approximation is used.
b. Based on the asymptotic chi-square approximation.

Table 4. Autocorrelations: Production of Vegetables

Lag	Auto correlation	Std. error (a)	Box-Ljung Statistic Value	df	Sig (b)
1	.876	.180	26.148	1	.000
2	.793	.286	48.339	2	.000
3	.699	.350	66.188	3	.000
4	.602	.392	79.926	4	.000
5	.527	.421	90.861	5	.000
6	.438	.442	98.703	6	.000
7	.358	.456	104.159	7	.000
8	.252	.465	106.979	8	.000
9	.177	.469	108.435	9	.000
10	.099	.471	108.908	10	.000
11	.021	.472	108.930	11	.000
12	-.049	.472	109.061	12	.000
13	-.126	.472	109.966	13	.000
14	-.176	.473	111.822	14	.000
15	-.227	.475	115.125	15	.000
16	-.275	.479	120.278	16	.000

a. The underlying process assumed is MA with the order equal to the lag number minus one. The Bartlett approximation is used., b. Based on the asymptotic chi-square approximation.

Table 5. Parameter Estimates of ARIMA Model

			Estimates	Std. error	t	Sig.
Area	Natural Log	Constant	-0.004	0.001	-2.788	0.010
		MA Lag 1	1.000	32.015	0.31	0.975
Production	Natural Log	Constant	-0.002	0.001	-1.305	0.203
		AR Lag 1	0.975	1.208	0.807	0.427

Table 6. Diagnostic Checking of Residuals Autocorrelation: Area and Production of Vegetables

Model	No. of Predictors	Model Fit Statistics					
		R-Squared	RMSE	MAPE	Normalized BIC	Ljung-Box Statistics	Sig.
Area (1,2,1)	0	0.969	294106.20	20.68	20.48	19.49	0.244
Production (1,2,1)	0	0.937	564591.00	30.36	26.83	11.83	0.755

Table 7. Forecasted Value of Area and Production of Vegetables in Haryana

Model		2021-22	2022-23	2023-24	2024-25	2025-26
Area-Model_1	Forecast	427242	447120	443544	449366	448265
	UCL	493680	526777	541364	561289	574132
	LCL	367856	377004	359830	355339	344640
Production-Model_2	Forecast	7661519	9190527	8854857	9743716	9845482
	UCL	9427497	11501512	11702035	13206035	13841102
	LCL	6159987	7252040	6569575	7021375	6796828

prediction emphasising the most recent observations appears more logical than a forecast highlighting all previous observations equally. The criteria to estimate the AR and MA coefficients in the model were determined using log likelihood, Akaike's Information Criterion, AIC (1969), Schwarz's Bayesian Criterion, SBC (1978), and residual variance. Table 5 provides parameter estimates for the fitted models.

To determine if random shocks were white noise, Ljung & Box (1978) recommended using the residual ACF, the associated "t" tests, and the Chi-squared test (Table 6). In Haryana, the ARIMA model might be used to estimate and forecast the area and output of vegetables. Vegetable's area and production in Haryana have been determined to be trending significantly increased. For predicting the area and production of vegetables in Haryana, the degree of accuracy attained using ARIMA (1, 2, 1) was judged to be adequate, and residuals were considered to be white noise. Five-steps ahead (out-of-model development period i.e. 2021-22, 2022-23, 2023-24, 2024-25 and 2025-26) forecasted values of area and production of vegetables in Haryana are shown in Table 7.

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Goyal et al., (2021) in their study found that the moving average and autoregressive procedures were tested with various lags, and ARIMA (0,1,1) was fitted to estimate agricultural export in India. It was discovered that the estimated agricultural export values for the years 2016–17 to 2018–19 were relatively close to the actual values, with percent deviations between the estimated and observed figures ranging from –2 to –4, and forecasted values for the three years ahead—2019–20, 2020–21, and 2021–22—lying within confidence limits based on ARIMA models.

CONCLUSION

The OLS estimation revealed that the forecasted value of the vegetable production for the year 2021-22 will be 7540814.31 tonnes. The best fit line of compound function representing the data of production which lies close to the lines are best described by the compound function line because the line is closer to the data values of vegetable production in different years. The compound function is explaining better other than linear, cubic and quadratic to the overall production function where the value of R Square 0.976 was very significant at 0.00 (significance level). It can be interpreted that the compound function explaining the 97 per cent variability in the production which is the best explanation for the vegetable production function. Further, for predicting the area and production of vegetables in Haryana, ARIMA (1, 2, 1) was fitted after experimenting with various lags of the moving average and autoregressive procedures. Five-steps ahead (out-of-model development period i.e. 2021-22, 2022-23, 2023-24, 2024-25 and 2025-26) forecasted values of area and production of vegetables in Haryana.

REFERENCES

Albright, L.D. (2002). Controlling Greenhouse Environments. *Acta Horticulture, 578*, 47-54.

Anonymous. (2017). Review of annual administrative report of horticulture department, Haryana for the year 2017-18. Retrieved from http://hortharyana.gov.in/sites/default/files/Administative%20report%20today%202017-18.pdf

Anonymous. (2018). Area and production of horticulture crops: All India. Retrieved from http://nhb.gov.in/statistics/State_Level/2018-19%20(3rd%20Adv.Est_.)%20-%20Website.pdf.

Anonymous. (2018). Horticultural statistics at a glance 2018. Retrieved from http://nhb.gov.in/statistics/Publication/Horticulture%20Statistics%20at%20a%20Glance-2018.pdf.

Anonymous. (2020). Detail of area and production of fruit, vegetable and flower since 1992-93 to 2020-21. Retrieved from http://info.hortharyana.gov.in/hortharyana/SA.xlsx

Anonymous. (2020). Statistical data. Horticulture department, Government of Haryana. Retrieved from http://hortharyana.gov.in/en/statistical-data.

Chandran, V., & Podikunju, B. (2020). Constraints experienced by homestead vegetable growers in Kollam district. *Indian Journal of Extension Education*, *57*(1), 32–37.

Goyal, M., Agarwal, S., Goyal, S., & Agarwal, N. (2021). Forecasting of Indian agricultural export using ARIMA model. *Asian Journal of Research in Business Economics and Management*,*11*, 11-15.

Koutsoyiannis, A. (2015). Theory of econometrics: An introductory exposition of econometrics methods. New York: PALGRAVE

Kumar, M., Battan, K. R., & Sheoran, O. P. (2019). Preharvest forecast model for rice yield using principal component regression based on biometrical character with R software. *International Journal of Agricultural and Statistical Sciences*, *15*(1), 323-326.

Kumar, M., Jaslam, P. M., Kumar, S., & Dhillon, A. (2021). Forecast and error analysis of vegetable production in Haryana by various modeling techniques. *Journal of Applied and Natural Science*, *13*(3), 907-912.

Kumar, M., Paul, R. K., & Singh, B. K. (2016). Estimating area, production and productivity of cotton crop in Haryana state. *Journal of Cotton Research and Development*, *30*(2), 317-323.

Kumar, M., Rajendra. & Hasija, R. C. (2017b). Arima modelling and forecasting of cotton productivity in India. *Environment & Ecology 35*(1A), 224-228, January-March 2017.

Kumar, M., Raman, R. K., & Kumar, S. (2017a). Forecasting of soybean yield in India through ARIMA model. *International Journal of Pure and Applied Bioscience*, *5*(5), 1538-1546.

Nimbrayan, P. K., Jaslam, P. M., & Chandanshive, A. (2022). Modelling and forecasting of area, production and productivity of tomatoes in Haryana and India. *Indian Journal of Extension Education*, *58*(2), 205-208.

Prakash, P., Jaganathan, D., Immanuel, S., Lama, A., Sreekumar, J., & Sivakumar, P. S. (2022). Forecasting of Sweet Potato (Ipomoea batatas L.) Prices in India. *Indian Journal of Extension Education*, *58*(2), 15–20.

Verma, U., Piepho, H. P., Hartung, K., Oggutu, J. O., & Goyal, A. (2015). Linear mixed modeling for mustard yield prediction in Haryana state (India). *Journal of Mathematics and Statistical Science*, *1*(3), 96-105.