IJTA
© Serials Publications
National Academy of Agricultural Science (NAAS)
Rating : 3. 03

# Forecasting Cultivated Area and Production of Vegetables in India

*Prema Borkar**

***Abstract:*** *The paper describes an empirical study of forecasting time series data of area and production of vegetables in India. Yearly vegetable area and production data for the period of 1960-1961 to 2012-2013 of India were analyzed by time-series methods. Autocorrelation and partial autocorrelation functions were calculated for the data. The Box Jenkins ARIMA methodology has been used for forecasting. The diagnostic checking has shown that ARIMA (1, 0, 1) and ARIMA (1, 1, 0) is appropriate for vegetable area and production. The forecasts from 2013-2014 to 2024-2025 are calculated based on the selected model. The forecasting power of autoregressive integrated moving average model was used to forecast vegetable area and production for twelve leading years. This projection is important for the policy makers to foresee the future requirements of vegetable production, import and/or export and adopt appropriate measures in this regard.*

***Key words:*** *ACF - autocorrelation function, ARIMA - autoregressive integrated moving average, PACF - partial autocorrelation function, Vegetables, trends.*

India's diverse climate ensures availability of all varieties of fresh fruits & vegetables. It ranks second in fruits and vegetables production in the world, after China. As per National Horticulture Database published by National Horticulture Board, during 2012-13 India produced 162.19 million metric tons of vegetables. The area under cultivation of vegetables stood at 9.21 million hectares. India is the largest producer of ginger and okra amongst vegetables and ranks second in production of potatoes, onions, cauliflower, brinjal, cabbages etc. The vast production base offers India tremendous opportunities for export. During 2014-15, India exported vegetables worth Rs. 4702.78 crores. Onions, Okra, Bitter Gourd, Green Chillies, Mushrooms and Potatoes contribute largely to the vegetable export basket. The major destinations for Indian vegetables are UAE, Bangladesh, Malaysia, UK, Netherland, Pakistan, Saudi Arabia, Sri Lanka and Nepal.

Though India's share in the global market is still nearly 1 per cent only, there is increasing acceptance of horticulture produce from the country. This has occurred due to concurrent developments in the areas of state-of-the-art cold chain infrastructure and quality assurance measures. Apart from large investment pumped in by the private sector, public sector has also taken initiatives and with APEDA's assistance several Centers for Perishable Cargoes and integrated post harvest handling facilities have been set up in the country. Capacity building initiatives at the farmers, processors and exporters' levels has also contributed towards this effort.

Forecasts have traditionally been made using structural econometric models. Concentration have been given on the univariate time series models known as auto regressing integrated moving average (ARIMA) models, which are primarily due to world of Box and Jenkins (1970). These models have been extensively used in practice for forecasting economic time series, inventory and sales modeling (Brown, 1959; Holt *et al.*, 1960) and are generalization of the exponentially weighted

* Assistant Professor, Gokhale Institute of Politics and Economics (Deemed to be a University), Pune - 411004. Maharashtra. India.

moving average process. Several methods for identifying special cases of ARIMA models have been suggested by Box-Jenkins and others. Makridakis *et al.* (1982), and Meese and Geweke (1982) have discussed the methods of identifying univariate models. Among others Jenkins and Watts (1968), Yule (1926, 1927), Bartlett (1964), Quenouille (1949), Ljune and Bos (1978) and Pindyck and Tubinfeld (1981) have also emphasized the use of ARIMA models.

In this study, these models were applied to forecast the area and production of vegetables in India. This would enable to predict expected vegetable area and production for the years from 2014 onwards. Such an exercise would enable the policy makers to foresee ahead of time the future requirements for vegetable production, import and/ or export of vegetables thereby enabling them to take appropriate measures in this regard.

## MATERIAL AND METHODS

Time Series data was used for the Study. The data were obtained from the website of Indian Horticulture Database from 1960-61 to 2012-13. Box and Jenkin (1976) linear time series model was applied on the data. This model is commonly known as Autoregressive Integrated Moving Average Model (ARIMA Model).

One of time series models which is popular and mostly used in ARIMA model. ARIMA (p, d, q) model is a mixture of Autoregressive (AR) model which shows that there is a relation of a value in the present (Z) and value in the past ($Z_{t-k}$), added by random value and Moving average (MA) model which shows that there is a relation between a value in the present ($Z_t$) and residuals in the past

$$(Z_{t-k}\ k = 1,2...)$$

with a non-stationary data pattern and d differencing order. The form of ARIMA (p,d,q) is:

$$\Phi_p(B)(1-B)^d Z_t = \theta_q(B) a_t$$

where, p is AR model order, q is MA model order, d is differencing order and:

$$\Phi_p(B) = (1 - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_p B^p)$$
$$\Phi_p(B) = (1 - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_q B^q)$$

Generalization of ARIMA model for a seasonal patter data, which is written as:

$$ARIM\ (p, d, q)\ (P, D, Q)^2 =$$
$$\Phi_p(B)\ \Phi_p(B^2)\ (1-B)^d\ (1-B^s)(1-B^s)^D Z_t$$
$$= \theta(B)^{\Theta}\ Q\ (B^s)a$$
$$t$$

where, s is seasonal period.

$$\Phi_p(B^s) = (1 - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_p B^p)$$
$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)$$

**Model Identification:** To determine whether the series is stationary or not we considered the graph of ACF. If a graph of ACF cuts of fairly quickly or dies down fairly quickly, then the time series value should be considered stationary. Model for non-seasonal series are called Autoregressive Integrated Moving average model, denoted by ARIMA (p,d,q). Here p indicates the order of the autoregressive part, d indicates the amount average of difference and q indicates the order of the moving average part. If the original series is stationary, d = 0 and the ARIMA models reduce to the ARMA models.

The difference linear operator ()), denoted by:

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1-B)Y_t$$

The stationary series:

$$W_t = \Delta^d Y_t = (1-B)^d\ Y_t = \mu + \theta_q(B)\varepsilon_t$$
$$or\ \Phi_p(B)W_t = \mu + \theta_q(B)\ \varepsilon_t$$

**Model estimation and checking:** Estimate the parameters for a tentative model has been selected. The derived model must be checked for adequacy by considering the properties of the residuals whether the residuals from an ARIMA model is normal and randomly distribution. An overall check of the model adequacy is provided by Ljung-Box Q statistics. The test statistics Q is given in equation below:

$$\underset{k}{Q_m} = n(n+2)\sum_{k-1}^{nr_e^2} \chi^2_{m-r}$$

where, $r_k$ (e) = the residual autocorrelation at lag K.

n = the number of residuals

m = the number of time lags includes in the test.

If the p-value associated with the Q Statistics is small (p-value < "), the model is considered inadequate. The analysts should consider a new or modified model and continue the analysis until a satisfactory model has been determined.

## RESULTS AND DISCUSSION

The maximum cultivated area of vegetables was 7.59 thousand hectares in 2013-13 and was minimum 2.78 thousand hectares in 1960-61. For production, the maximum production of vegetables in India was obtained in 2012-2013 year (105.80 thousand tons) and minimum in 1950-1951 year (18.47 thousand tons).

Last 53 years data of cultivation area and production of vegetables in India was used for modeling purpose. In model specification, we looked at the plots of auto-correlation function (ACF) for vegetable cultivated areas (Fig. 1) and production figures (Fig. 2). Also, partial auto correlation function (PACF) for vegetable cultivated area (Fig. 3) and production (Fig. 4). Auto correlation function indicated the order of the auto regression compounds "q" of the model while the partial correlation function gave an indication for the parameter p. The ACF and PACF of the residuals (Fig. 5 and 6) also indicate 'good fit' of the model.

The time series plot (Fig. 7 and 8) of cultivation areas and production showed an increasing trend. ACF of both series showed non-stationary as ACF did not fall as quickly as the log K increased. To check the further stationary, second difference of the original series for cultivation area and production was taken. The auto correlation formulation of second series and correlogram shows some more stationary than that of the first different. The corellogram of the auto correlation function of first difference series showed that the auto correlation function falls finally after lag 1 for area and production, hence the respective values of the parameter "q" decided to be 1.

PAC function of the first differenced series of the cultivation area and production was used to determine parameter "p". Thus, we chose "p" to be 1 and 0 for cultivation area and production respectively which gave good results consequently,

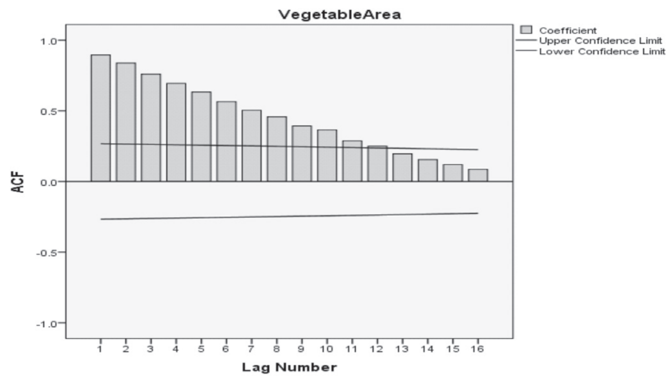


**Figure 1: Autocorrelation function of Vegetable cultivated area**

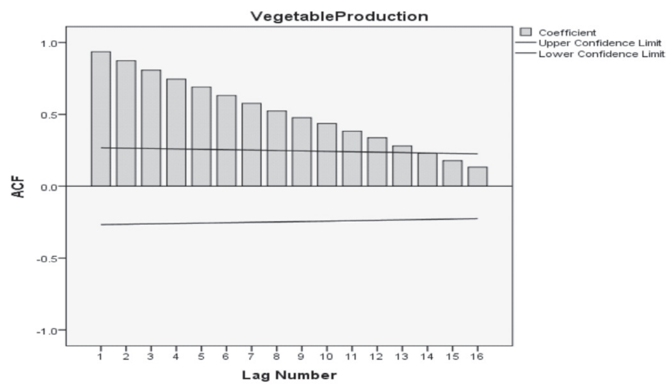


**Figure 2: Autocorrelation function of Vegetable production**

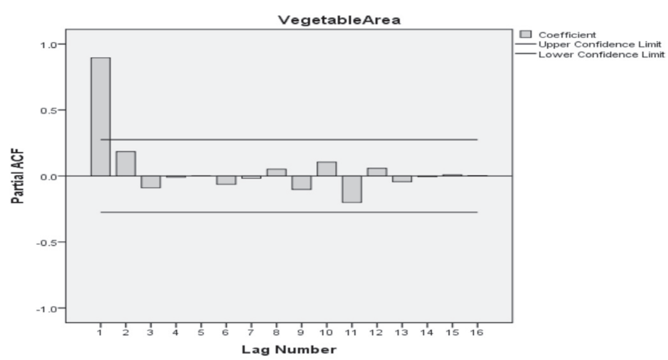


**Figure 3: Partial autocorrelation function for Vegetable cultivated area**

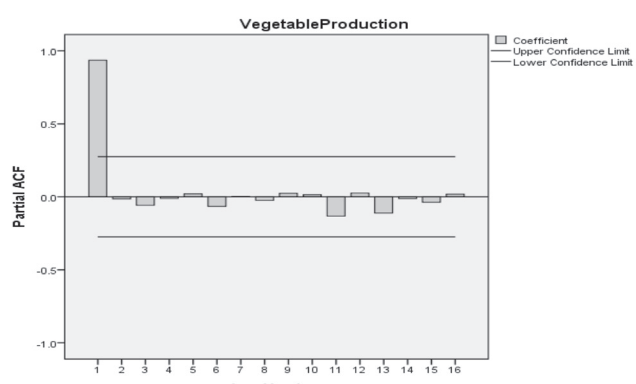


**Figure 4: Partial autocorrelation function for Vegetable production**

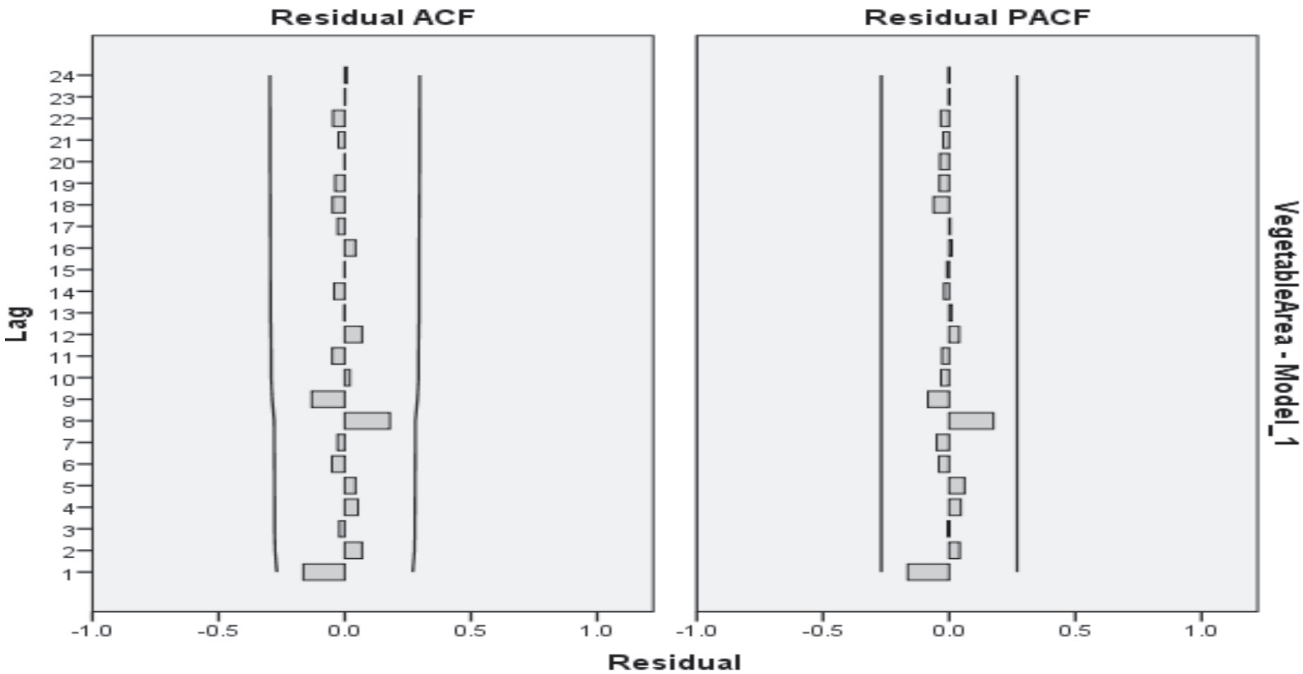


**Figure 5: ACF and PACF of residuals of fitted Vegetable cultivated area**

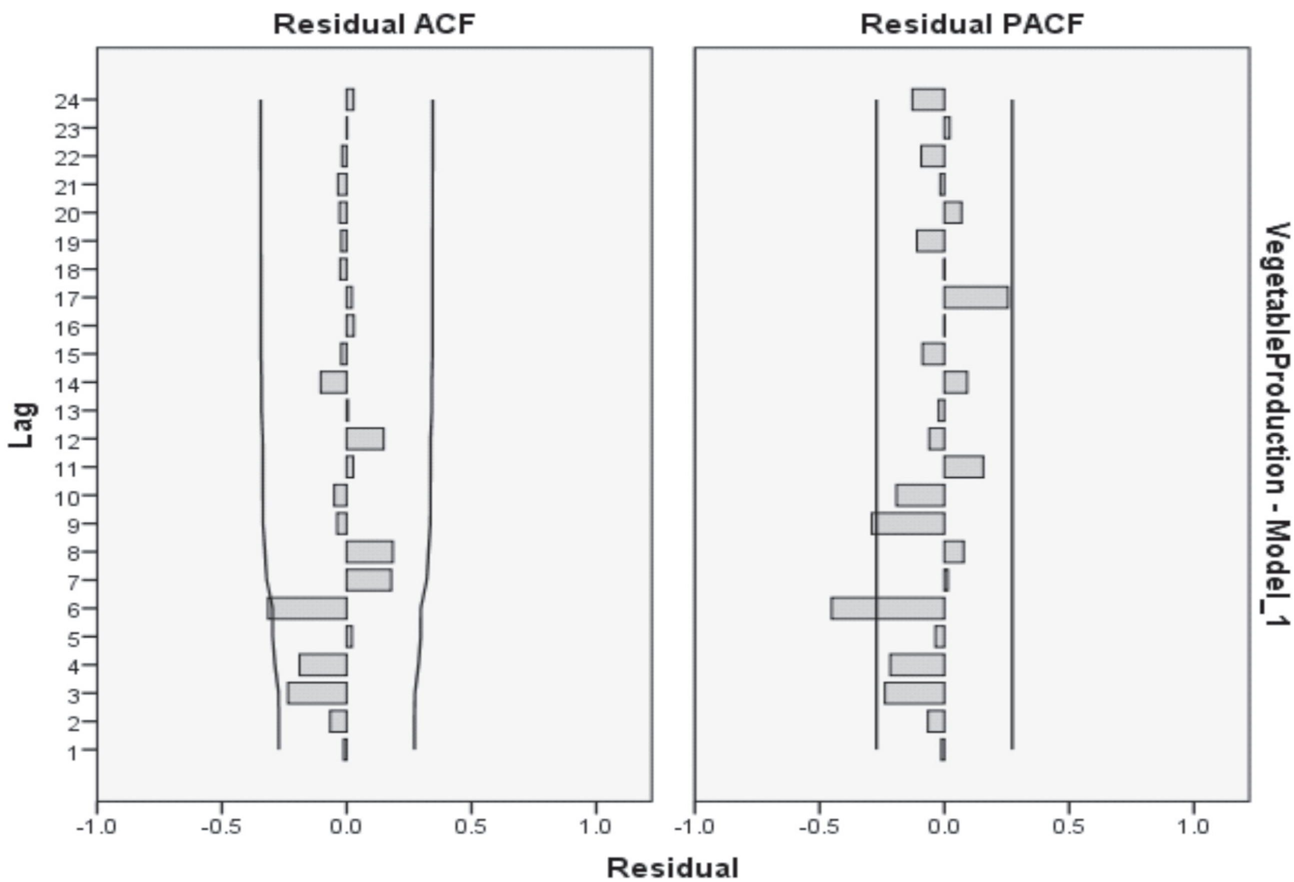


**Figure 6: ACF and PACF of residuals of fitted Vegetable production**

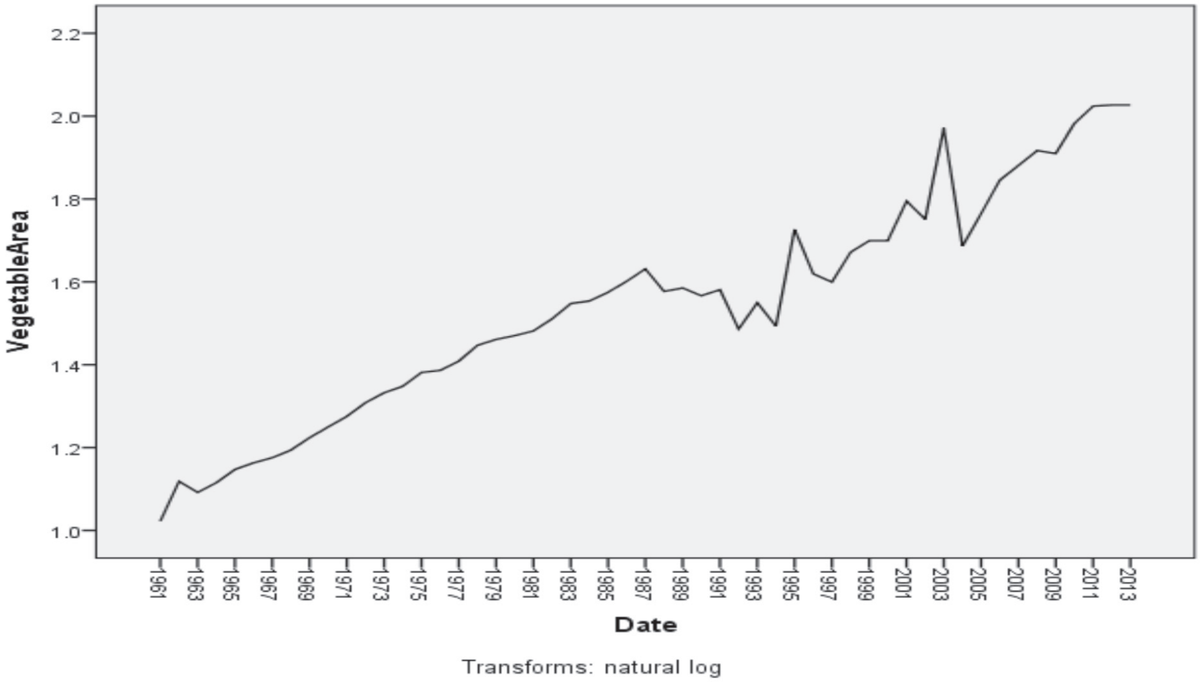


**Figure 7: Time series plot for Vegetable cultivated area**

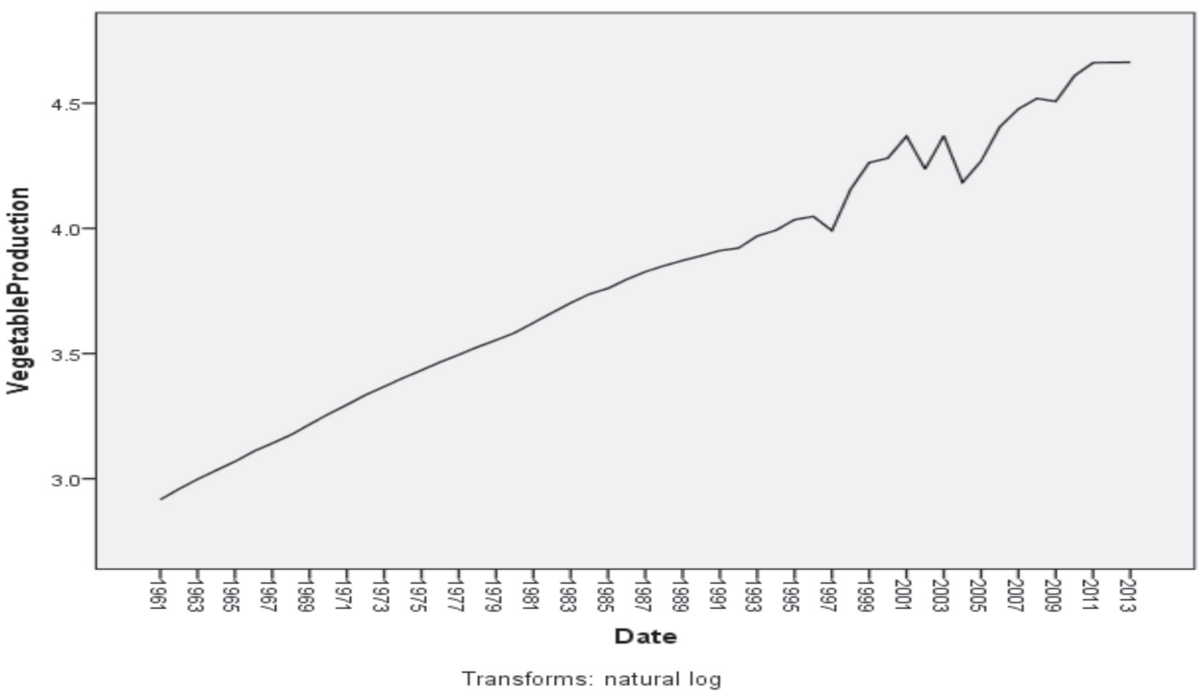


**Figure 8: Time series plot of Vegetable production**

**Table 1**
**Estimate of area parameters**

| *Type* | *Coefficients* | *S.D.* | *t-ratio* |
|---|---|---|---|
| Constant | 5.233 | 6.354 | 0.824 |
| AR Lag 1 | 0.994 | 0.031 | 32.402 |
| MA Lag 1 | 0.385 | 0.164 | 2.348 |

**Table 2**
**Estimate of production parameters**

| *Type* | *Coefficients* | *S.D.* | *t-ratio* |
|---|---|---|---|
| Constant | 0.034 | 0.005 | 6.520 |
| AR Lag 1 | 0.359 | 0.132 | 2.714 |
| MA Lag 1 | 0.563 | 0.145 | 2.318 |

**Table 3**
**Forecasts for Vegetable area and production (2013-14 to 2024-2025)**

| *Years* | *Area (Thousand Hectares)* | *Production (Thousand Tonnes)* |
|---|---|---|
| 2013-2014 | 7.53 ± 0.78 | 111.07 ± 11.68 |
| 2014-2015 | 7.52 ± 0.91 | 114.45 ± 14.40 |
| 2015-2016 | 7.51 ± 1.02 | 118.65 ± 17.94 |
| 2016-2017 | 7.49 ± 1.13 | 122.73 ± 20.98 |
| 2017-2018 | 7.48 ± 1.22 | 127.05 ± 24.11 |
| 2018-2019 | 7.47 ± 1.30 | 131.49 ± 27.19 |
| 2019-2020 | 7.46 ± 1.38 | 136.10 ± 30.33 |
| 2020-2021 | 7.44 ± 1.46 | 140.86 ± 33.54 |
| 2021-2022 | 7.43 ± 1.52 | 145.79 ± 36.82 |
| 2022-2023 | 7.42 ± 1.59 | 150.90 ± 40.19 |
| 2023-2024 | 7.41 ± 1.65 | 156.18 ± 43.68 |
| 2024-2025 | 7.39 ± 1.71 | 161.65 ± 47.29 |

the respective value of p,d,q were determined for ARIMA, that is ARIMA (1,0,1) and ARIMA (1,1,0).

### Model estimation

ARIMA (1,0,1) and (1,1,0) model were estimated using S-plus statistical difference and estimation of the models for the rice area and production are given in Table 1 and 2.

Goodness of fit of the model given in as the diagnostic check of the estimated model.

**Residual analysis:** The time series plot of the residual cultivated area and production data showed scattered trend, therefore, models were fitted properly by residual analysis.

For normality test, Shapiro-wilk test was used. The test was significant and assumption of normality was accepted. Since the series fitted shows normality, the model is a good fit.

Forecast of area and production. ARIMA (1,0,1) and ARIMA (1,1,0) were taken for 12 years ahead and forecasts for rice area and production are given in Table 3 at 95% confidence interval values.

For 2013-14, a forecast of vegetable area was 7.53 thousand ha with lower and upper limit of 0.78 thousand hectare. Vegetable area forecast for the year 2024-2025 was 7.39 thousand ha with lower and upper limit of 1.71 thousand hectare. Forecast of vegetable production showed an increasing trend. For 2013-14, vegetable production was about 111.07 thousand tons with lower and upper limit of 11.68 thousand tons. The maize production forecast for the year 2024-2025 is about 161.65 thousand tons.

## CONCLUSION

In our study, the developed model for vegetable area and production was found to be ARIMA (1,0,1) and (1,1,0). The forecasts of vegetable area and production, lower control limits (LCL) and upper control limits (UCL) are presented in Table 3. The validity of the forecasted values can be checked when the data for the lead periods become available. ARIMA model being stochastic in nature, it could be successfully used for modeling as well as forecasting the vegetable area and production of India. The model demonstrated a good performance in terms of explaining variability and predicting power. The supply projection of an agricultural commodity especially vegetable plays a vital role in the adjustment of supply to demand. These projections help the government to make policies with regards to relative price structure, production and consumption patterns and also, to establish relationship with other countries of the world.

### *References*


Bartlett, M.S., (1964), "On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series". *J. Roy. Stat. Soc.,* B 8: 27–41.

Box, G.E.P., and G. M. Jenkins. (1970), Time series analysis: forecasting and Control, Holden Day, San Francisco, CA.

Box, G.E.P. and G.M. Jenkins, (1976), Time Series Analysis: Forecasting and Control. Rev. Ed. San Francisco. Holden-Day.

Brockwell, P.J., and Davis, R. A. (1996), Introduction to time series and Forecasting, Springer.

Brown, R.G., (1959), Statistical Forecasting for Inventory Control. New York, McGraw-Hill.

Iqbal, N., K. Bakhsh, A. Maqbool and A. S. Ahmad. (2005), Use of the ARIMA Model for forecasting Wheat area and production in Pakistan". Journal of Agriculture & Social Sciences., 2: 120-122.

Jenkins, G. M. and D.G. Watts, (1968), Spectral Analysis and its Application, Day, San Francisco, California, USA.

Ljunge, G.M. and G.E.P. Box, (1978), "On a Measure of Lack of Fit in Time Series Models". Biometrika, 65: 67–72.

Kendall, M. G., and A. Stuart. (1966), The advanced theory of statistics. Vol. 3. Design and Analysis and Time-Series. Charles Griffin & Co. Ltd., London, United Kingdom.

Holt, C.C., F. Modigliani, J.F. Muth and H.A. Simon, (1960), Planning, Production, Inventores, and Work Force. Prentice Hall, Englewood Cliffs, NJ, USA.

Muhammad, F., M. S. Javed and M. Bashir. (1992), Forecasting Sugarcane Production in Pakistan using ARIMA Models, Pak. J. Agri. Sci., 9(1): 31-36.

Makridakis, S, A. Anderson, R. Filds, M. Hibon, R. lewandowski, J. Newton, E. Parzen and R. Winkler, (1982), The Accuracy of Extrapolation (time series) methods: Results of a Forecasting Competition, Journal of Forecasting Competition. J. Forecasting, 1: 111–53.

Meese, R. and J. Geweke, (1982), A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series. Unpublished Manuscript, University of California, Berkeley, CA, USA.

Prindycke, R.S. and D.L. Rubinfeld, (1981), Econometric Models and Economic Forecasts, 2nd Ed. New York, McGraw-Hill.

Quenouille, M.H., (1949), "Approximate Tests of Correlation in Time- Series. J. Roy. Stat. Soc., B11: 68–84.

Saeed, N., A. Saeed, M. Zakria, T. M. Bajwa. (2000), Forecasting of Wheat Production in Pakistan using ARIMA models, International Journal of Agriculture and Biology., 4: 352-353.

Yule, G.U., (1926), "Why Do We Sometimes Get Nonsence-corrleations Between Times Series. A study in Sampling and the Nature of Series". J. Roy. Stat. Soc., 89: 1–69.

Yule, G.U., (1927), "On a method of Investigation Periodicities in Disturbed Series, With Specia; Reference To Wolfer's Sunspot Number". Phill. Trans., A 226: 267–98.