

FORECASTING OF POTATO PRODUCTION IN HIMACHAL PRADESH USING ARIMA MODEL

Sapna Bhardwaj^{1*} and Banti Kumar²

ABSTRACT: This study focuses on the application of the ARIMA time series model to forecast the future potato production in Himachal Pradesh based on past data from 1991-92 to 2019-20. The stationarity of the data has been checked using the augmented dickey Fuller test which showed that the data was non-stationary. Therefore, differencing up to the first order has been done to make it stationary. Based on various model selection criteria, the best-fitted model was ARIMA (0, 1, 1) which has been used for the forecast of the next five years i.e., from 2020-2024. From the forecasted figure, potato production will go from 29612.36 tons to 22543.08 tons in 2024 showing an increasing trend.

KEYWORDS: ARIMA, Stationarity, Forecasting, Time Series data, Potato production

INTRODUCTION

Potato plays a vital role in the agricultural economy. India is the 2nd largest producer of potato in the world after China, accounting for ~12% of global potato production (Choudhary *et al.*, 2021). It is economically important not only for its contribution to the livelihood of farmers but also for its dominance in the agricultural consumption basket of households (Paul *et al.*, 2016). China is the biggest producer of potatoes worldwide, with about 1/3rd of the world's potatoes produced in China and India. According to FAO estimates, in 2020, over 359 million tons (mt) of potatoes were produced worldwide, an increase from a production volume of 333.6 mt in 2010. Potato (*Solanum tuberosum*) is the 3rd top most consumed crop worldwide after rice and wheat (International Potato Center, 2013).

In Indian agricultural GDP, the contribution of potatoes from the unit area of cultivable land is about 3.7 times higher than rice and 5.4 times higher than wheat

(Indian Horticulture, 2019). Some popular potato varieties have been developed by ICAR-Central Potato Research Institute for farmers which cover almost 95 % of the total area under potatoes. India produced approximately 45.87 mt of potato annually during the triennium 2014-17 and contributed ~57,512 crores annually to the Gross Value Added (GVA) at current prices (Indian Horticulture, 2019). The economy of Himachal Pradesh is largely dependent on agriculture. About 12.73 % of the total gross state domestic product (GSDP) is generated through agricultural and allied sectors (Economics Survey of Himachal Pradesh, 2019-20). So, it is an essential sector for the sustained growth of the state economy. Although rice, wheat, and maize are major agricultural crops grown in the state yet, the potato is also one of the important commercial crops. Out of total vegetable production in the country, the major share goes to potatoes (28.9 %). Out of the total 17.22 lakh tons production of vegetables in the state, potato accounts for 1.87 lakh

¹Department of Agricultural Economics, Extension Education and Rural Sociology, College of Agriculture, CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur, India

²Department of Physical Sciences and languages, College of Basic Sciences, CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur, H.P. India

*Corresponding author; email: s3vsapna@gmail.com

tons i.e. 10.85 % which is the largest among all other vegetables (Economics Survey of Himachal Pradesh, 2019-20).

In this research paper, forecasting of potato production has been done for the next five years (2020 to 2024) for Himachal Pradesh. In recent times, forecasting methods have been increasingly applied by researchers. Some popular forecasting analysis methods are exponential smoothing methods, the Box-Jenkins ARIMA model, the Grey forecasting model, and others (Şahinli, 2020). Time series analysis for forecasting may be regarded as a useful practice to develop an appropriate econometric model for estimating future behavior based on previously observed behaviour (Ray and Bhattacharyya, 2020).

ARIMA (p, d, q) models are created by first taking the differences of the data from d degree for the stabilization process and then adding the ARMA (p, q) model. Hossain and Faruq (2016) used a time series model to find out the patterns in the past and utilize that information to forecast the future values of potato production in Bangladesh. They selected Box-Jenkins ARIMA (0,2,1) model for forecasting the potato production for the whole of Bangladesh. Jubair *et al.* (2019) carried out horizontal variability of some soil properties in Wasit governorate Iraq by using time series analysis. Mishra *et al.* (2021) did an estimation of fish production in India using ARIMA, Holt's Linear, BATS, and TBATS Models. Jamuna *et al.* (2021) analyzed the price of potatoes through the time series analysis using a price forecasting system that is implemented based on ARIMA Model. The potato price at one main market for each state was examined using the ARIMA model based on monthly average data obtained from Karnataka, Madhya Pradesh, Gujarat, Rajasthan, and Maharashtra. Kumar

and Baishya (2020) worked on forecasting of potato prices in India through the ARIMA model. Kumar *et al.* (2019) also used time series modeling to forecast the oilseed production in India.

ARIMA is also called as Box-Jenkins model because it was proposed by George Box and Gwilym Jenkins in 1976. Since, the ARIMA model employs non-zero autocorrelation between the subsequent values of the time series data, it was chosen for predicting in the current study. The present work has been taken with the following objectives:

- (i) To investigate the past trend in potato production in Himachal Pradesh.
- (ii) To develop an ARIMA model for potato production in Himachal Pradesh.
- (iii) To forecast the future potato production for the next 5 years i.e. from 2019 to 2020.

MATERIALS AND METHODS

Time series data on potato production in Himachal Pradesh for the last 30 years, i.e., from 1991-92 to 2019-20 were used for the present study. Data were procured from Statistical Yearbooks of Himachal Pradesh published by the Government of H.P.

Autoregressive Integrated Moving Average Model:

Three elements make up an ARIMA model: Moving Average (MA), Integration (I), and Autoregression (AR). The variable of interest is regressed on its own lagged values to create the AR model. I display how many times the series has been made stationary by differencing. A stochastic term's current value as well as several previous values are linearly reliant on the output variable that MA represents. A basic ARIMA model is expressed as (p, d, q), where p signifies the

order of the moving average process, d the number of times the series has undergone differencing to make it stationary, and q the order of the autoregressive process.

A time series must be stationary in order to be used for forecasting. In the event that the series is not stationary, it must be made stationary by taking the differencing up to the appropriate order. A time series generating process's statistical characteristic that doesn't change over time is called stationarity. If the mean, variance, autocorrelation, etc. remain constant across time, a series is said to be stationary. If the conditions listed below are true for every t , a time series Y_t is stationary.

$$E(Y_t) = m < \infty, V(Y_t) = E(Y_t - m)^2 = g_0 < \infty \text{ and } Cov(Y_t, Y_{t-k}) = E\{(Y_t - m)(Y_{t-k} - m)\} = g_k$$

To test whether a series is stationary or not, one may apply the unit root test. The unit root test checks whether the given time series has a unit root or not. Augmented Dickey-Fuller (1979) test is prominently used which assumes that the error term in the model is correlated.

Correlogram: It is a visualisation of a time series' time lag against partial and autocorrelation correlations. The underlying behaviour of a time series is examined using the patterns displayed in a correlogram. The basic connection between X_t and, let's say, X_{t+h} is shown by autocorrelation; this correlation between series has a lag. The autocorrelation function is expressed mathematically as:

$$\rho = \frac{Cov(y_t, y_{t-p \text{ or } q})}{var(y_t)}$$

Partial autocorrelation is the relationship between two observations after eliminating the linear relationship of every observation that comes between those two observations. Mathematically, it is given as:

$$\varphi(p) = \frac{\rho_p - \sum_{j=1}^{p-1} (\varphi_{p-2,j} - \varphi_{pp}\varphi_{p-2,p-j})\rho_{p-j}}{1 - \sum_{j=1}^{p-1} (\varphi_{p-2,j} - \varphi_{pp}\varphi_{p-2,p-j})\rho_j}, p \geq 3$$

The autocorrelation plot in the correlogram helps to decide the order of MA (q) whereas the partial autocorrelation plot helps to identify the order of AR (p).

Ljung-Box Q statistic: It is used to assess whether the observations are independent and arbitrary over time. It is a way to test for the *absence* of serial autocorrelation, up to a specified lag k . The test determines whether or not the autocorrelations for the errors or residuals are non-zero. The null hypothesis of the Ljung-Box Test is that H_0 : the model *does not* show a lack of fit. The alternative hypothesis is that H_1 : the model *does* show a lack of fit. A significant p-value in this test rejects the null hypothesis that the time series *data are not* autocorrelated. Ljung-Box Q statistic has an asymptotic chi-square distribution.

$$Q(p) = T(T+2) \sum_{i=1}^p \frac{\rho_i^2}{T-1} \sim \chi_p^2$$

A Box Jenkins methodology can be applied in three main steps viz., Identification, estimation and model checking, and forecasting.

Identification: To determine if the provided time series is stationary or not, model identification is used. If a stationary model is discovered, forecasting can be done using it. If it isn't, the series must be differentiated in order to make it stationary. An ARIMA model is then given for the differences series after the original series has been substituted by a series of differencing.

Estimation and Model Checking: Diagnostic testing, which is a stage in ARIMA modelling, entails looking at autocorrelation and partial autocorrelation functions for the estimation of residual behaviour. This aids

in determining the situations in which the model is inappropriate. At this point, the estimated models are tested to see if they meet the requirements for single-variable time series.

Forecasting: In this step, the estimated model is used to predict the future values of the time series.

RESULTS AND DISCUSSION

Figure 1 shows the behavior of the data on the annual production of potatoes. The scenario of potato production has shown an overall increasing trend except for steep low production in 2010.

Augmented Dickey-Fuller (ADF) test for checking stationarity of the data: ADF test is the most widely used test for determining the stationarity of data. The ADF test has the hypothesis that H_0 : the non-stationarity of the data against H_1 : the data are stationary. The critical value for the rejection of the null hypothesis of the unit root test was not significant as the p-value is (>0.05) at 5 % level of significance. The results showed that the data regarding potato production were non-stationary and could be used for forecasting purposes.

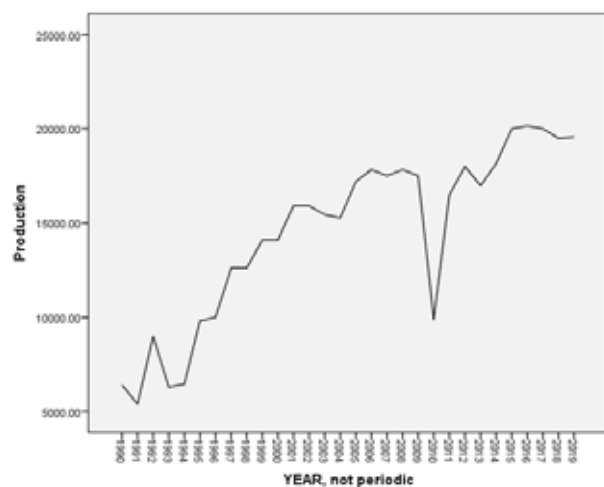


Fig. 1: Behaviour of the potato production in Himachal Pradesh for the period from 1999 to 2019

Since Table 1 reveals that the data were non-stationary, therefore, differencing up to order one is to be done to make it stationary.

It can be seen from above figure 2 that the mean and variance appeared to be normalized which indicates that data are now made stationary. ADF test statistics value is -5.65 and is highly significant which shows that the model is good and can be used further for forecasting purposes (Table 2).

The next step is to find out the values of p and q . For this, the autocorrelation and partial autocorrelation coefficients of various orders of X_t are computed. The ACF and PACF (Fig. 3, 4) show that the order of p is zero and q can be at the most one.

Table 1: Unit root test for checking the stationarity of the data

	Test - statistics	Significant value
ADF Test at 5 % level of significance	-1.79	0.38

Table 2: Unit root test for checking the stationarity of the data after the first difference

	test - statistics	Significant value
ADF Test at 5 % level of significance	-5.65	0.00

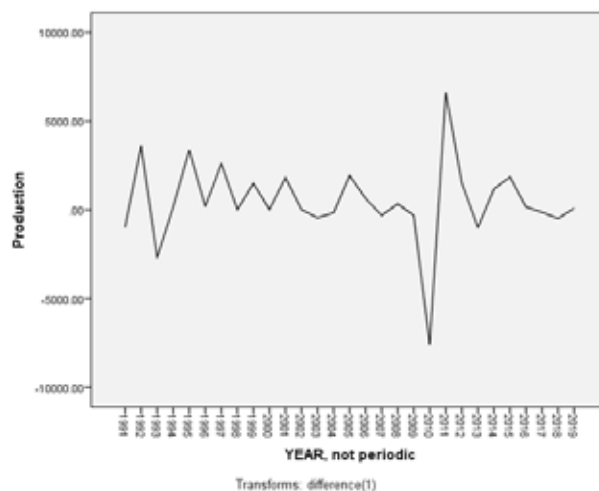


Fig. 2: Time-series plot of stationary Potato production data

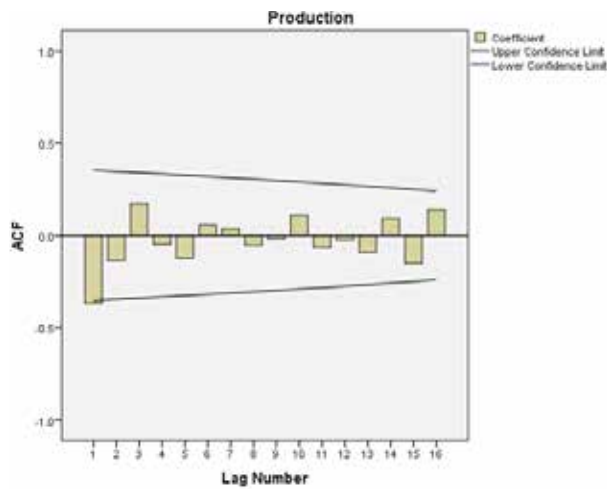


Fig. 3: Sample autocorrelation function

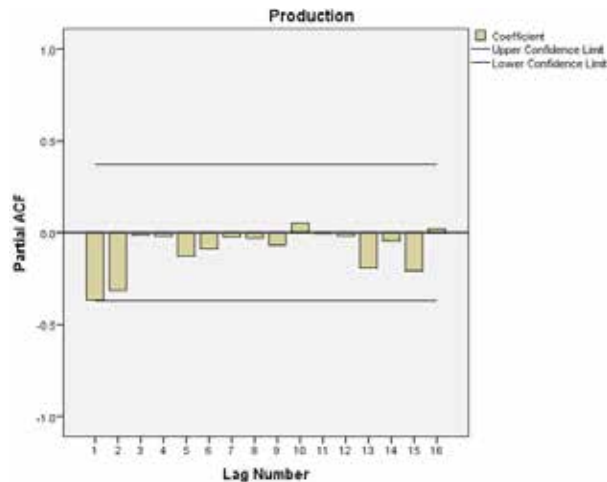


Fig. 4: Sample partial autocorrelation function

After trying all the possible ARIMA models, we chose the model ARIMA (0, 1, 1) and depicted it based on certain model selection criteria as:

ARIMA (p, d, q)	Stationary R-squared	Normalized BIC	MAPE
(0, 1, 1)	0.20	15.55	12.29

Table 3: Estimation of parameters

ARIMA Model Parameters	Estimate	SE	t	Significant
Constant	482.70	190.36	2.53	0.01
MA	0.53	0.16	3.21	0.00

Ljung-Box Q-statistics = 11.13 (p value = 0.85)

Model Estimation and Verification

Table 3 depicts the results of the estimation of parameters. The value of Ljung-Box Q-statistics shows that model does not show a lack of fit. MA estimates are highly significant at 1 per cent level which indicates that model is best fitted and good for forecasting.

The ACF and PACF of the residuals (Fig. 5) also indicate a 'good fit of the model and indicate that there is no correlation between sample errors.

Residuals of models are needed to be checked to see if they have any systematic pattern which will lead to the selection of a much better ARIMA model. This is done by examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 25 lags are computed and the same along with their significance

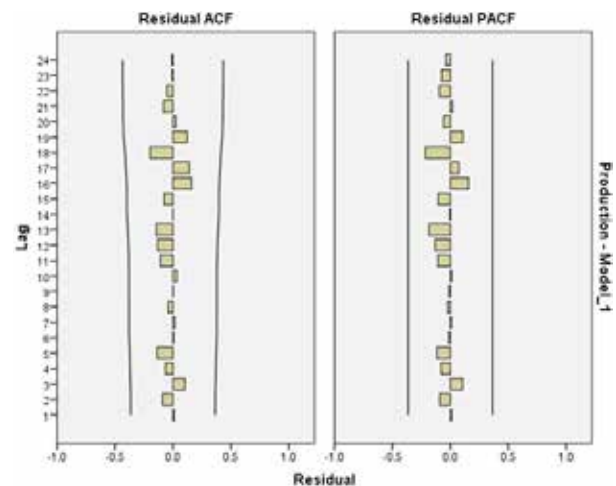


Fig. 5: ACF and PACF of residuals of fitted ARIMA (0, 1, 1) model.

Table 4: Forecasted values for the next five years

Year	Estimated Production	LCL	UCL
2020	20612.26	16263.18	24961.34
2021	21094.97	16301.95	25887.99
2022	21577.67	16378.48	26776.86
2023	22060.38	16484.53	27636.23
2024	22543.08	16614.45	28471.71

are tested by the Box-Ljung test. As the results indicate, none of these correlations is significantly different from zero at 5% level. This proves that the selected ARIMA model is appropriate.

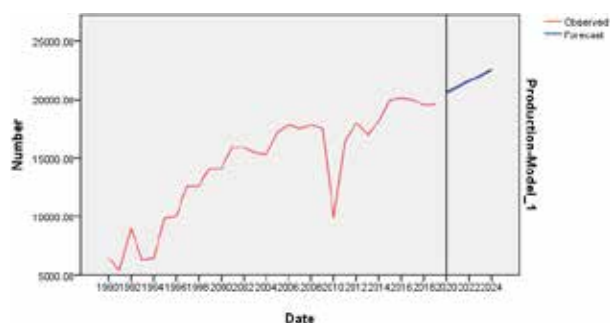
Forecasting with ARIMA model

Table 4 showed that for 2020, a forecast of maize area was 20612.26 mt with the lower and upper limit of 16263.18 and 24961.34 mt respectively. The forecasting of potato production in Himachal Pradesh in 2024 would be 22543.08 mt with the lower and upper limit of 16614.45 and 28471.71 mt respectively..

It can be seen from the fig.6 that potato production will increase in the upcoming five years according to the forecasted model.

CONCLUSION

In this study, ARIMA (0, 1, 1) model was developed and the validity of the forecasted values can be checked when the data for the lead periods become available. The correctness of the anticipated values may be verified once the data for the lead periods are available since an ARIMA (0, 1, 1) model was used in this study. Analysis of the performance of potato production from preceding 1991-92 to 2019-20 gives us the ARIMA (0, 1, 1) model which helps us in forecasting the future values of potato production. ARIMA (0, 1, 1) was chosen from different model parameters as it provides the best model which satisfies all the criteria of

**Fig 6: Forecasted graph for 2020 -2024**

fit statistics. For forecasting potato production in Himachal Pradesh, the model can be preferred by the researchers but it should be updated time and again with the inclusion of contemporary data.

CONFLICT OF INTEREST

The authors confirm that this manuscript has no conflict of interest.

LITERATURE CITED

- Anonymous (2019-20) Economic Survey of Himachal Pradesh, Department of Economics and Statistics, Himachal Pradesh: 1-172p
- Anonymous (2019-20) Statistical Year Book of Himachal Pradesh, Department of Economics and Statistics, Himachal Pradesh: 1-27p
- Anonymous (2019-20) Indian Horticulture Economic Survey of India, Ministry of Agriculture and Farmers Welfare, India: 1-78p
- Hossain Md and Abdulla Faruq (2016) Forecasting Potato Production in Bangladesh by ARIMA Model. *Journal of Advanced Statistics* 1(4):198-91
- International Potato Center (2013) Potato facts and figures <https://cipotato.org/potato/potato-facts-and-figures/>, 1st September 2022
- Jamuna C J, Patil Chetana and Kumar R Ashok (2021) Forecasting the Price of Potato Using Timer Series ARIMA Model. Proceedings of International Conference on Communication and Computational Technologies, Algorithms for Intelligent Sytems, Kumar S et al (eds.). Springer, Singapore pp.493-518
- Jubair Amel Radhi, Abboud Nada Farouq and Hamad Ibrahim Abdul Ghafour (2019) Horizontal Variability of Some Soil Properties in Wasit

- Governorate by Using Time Series Analysis. *Indian Journal of Ecology* **46**(3):565-58
- Kumar Banti, Monika, Bhanot Rajeev Prakash and Sharma Manish Kumar (2019) Forecasting of Oilseed Crop Production through Time Series Modeling. *Think India Journal* **22**(37):326-21
- Kumar and Baishya M (2020) Forecasting of Potato Prices in India: An Application of ARIMA Model. *Economic Affairs* **65**(4): 479-73
- Mishra P, Padmanaban K and Meenakshee D (2017) Modeling and forecasting of black pepper production in India. *Indian Journal of Ecology* **44**(4): 745-41
- Mishra Pradeep, Ray Soumik and Khatib (2021) Estimation of Fish Production in India using ARIMA, Holt's Linear, BATS, and TBATS Models. *Indian Journal of Ecology* **48**(5): 1261-54
- Ray S and Bhattacharyya B (2020) Time series modeling and forecasting on pulses production behavior of India. *Indian Journal of Ecology* **47**(4): 1149-40
- Sahinli MA (2020) Forecasting with Holt-Winters and ARIMA Methods: A Case Study. *American Journal of Potato Research* **97**: 346-36
-
- MS Received: November 29, 2022; Accepted: December 19, 2022