

International Journal of Statistics and Applied Mathematics



ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2026; 11(5): 147-155
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 20-03-2026
Accepted: 22-04-2026

Lokesh Kumar
Department of Physical Sciences,
MGCGV, Chitrakoot, Satna,
Madhya Pradesh, India

SS Guatam
Department of Physical Sciences,
MGCGV, Chitrakoot, Satna,
Madhya Pradesh, India

Aman Karun Ekka
Department of Physical Sciences,
MGCGV, Chitrakoot, Satna,
Madhya Pradesh, India

Instability analysis and forecasting of potato production dynamics in Madhya Pradesh

Lokesh Kumar, SS Guatam and Aman Karun Ekka

DOI: <https://www.doi.org/10.22271/maths.2026.v11.i5b.2328>

Abstract

Potato is a major commercial and staple food crop, which plays a vital role in agricultural development and food security in Madhya Pradesh. The current study examines the variability and predicts the future trends of potato cultivation in terms of area, production and productivity. The variability and growth trends were studied using secondary time-series data for the study period. The instability in the series was determined by the Coefficient of Variation (CV) which gives a measure of variability. The findings suggest that potato production in the state has exhibited considerable variability, especially in production and productivity, which is influenced by climatic variability, the use of inputs, and market factors. Although the trend in area and production is upwards, the presence of fluctuations indicates a degree of instability. To make projections, the Autoregressive Integrated Moving Average (ARIMA) model was used to forecast values until 2030. It fitted well to the past data and produced accurate predictions for area, production and productivity. The forecasts show an increasing trend in area and production, with a stable trend in productivity, which implies a steady improvement in potato cultivation. In conclusion, the study reveals that while potato production in Madhya Pradesh is growing, the presence of instability indicates a need for better technology, risk management and policy measures to maintain a stable and sustainable growth in the coming years.

Keywords: Instability analysis, potato, coefficient of variation (cv), Arima model, time series, forecasting, area, production, productivity, agricultural, Madhya Pradesh.

Introduction

Potato (*Solanum tuberosum* L.) is a globally significant food crop, belonging to Solanaceae family. It is the third most important food crop following rice and wheat and is essential for global food and nutritional security (Sood *et al*, 2022) ^[12]. It is a highly productive crop, and generates more edible energy and protein per unit area and time than many cereals, making it a key component of sustainable farming systems (FAO 2008) ^[7]. Potato is a herbaceous, annual, tuber crop with the tubers used for human consumption. Potato is propagated vegetatively using seed tubers, not botanical seeds, and thus genetic homogeneity is maintained and plants establish quickly. Potato is tetraploid ($2n = 4x = 48$) in nature. Potatoes are best grown in cooler climates and are a popular winter (rabi) crop in subtropical areas such as India. (Buckseth *et al.*, 2022) ^[4] Potato is native to the Andean region of South America, in the modern countries of Peru and Bolivia, and was domesticated almost 8,000-10,000 years ago (Li *et al.*, 2024) ^[10]. Since being introduced to other regions, it has become a staple food crop in many regions because of its versatility and productivity. Potato is now grown in over 150 countries under various climatic conditions. (FAO 2008) ^[7].

Potato is a good food nutritional wise as it provides carbohydrates, high-quality protein (having all the essential amino acids), vitamins (particularly vitamin C) and minerals (potassium and iron). It has a higher yield of dry matter, edible protein, and minerals per unit area than other food crops, making it a good choice for nutritional security Choudhary (2023) ^[6]. Special cultivation techniques such as using good quality seed tubers, spacing, irrigation, and integrated nutrient and pest management greatly contribute to improve potato yield. Recent developments such as better seed production technologies and climate-smart potato varieties

Corresponding Author:
Lokesh Kumar
Department of Physical Sciences,
MGCGV, Chitrakoot, Satna,
Madhya Pradesh, India

have also helped in enhancing yield and sustainability of potato production. (Buckseth *et al.*, 2022) ^[4].

Globally, potato is grown on an area of around 17-18 million hectares with production over 370 million tonnes, highlighting its significance and demand. China, India, Russia and US are the leading potato-producing nations. The growing demand for processed potato products and a shift in food habits have also contributed to increased global potato production. (Sood *et al.*, 2022) ^[12].

Potato in India is the second-largest crop (after China), with a production of more than 50-60 million tonnes from 2.1-2.2 million hectares. Potato is an important crop in Indian agriculture for its productivity and profitability. It is mainly cultivated in the Indo-Gangetic plains and has significant contributions from states such as Uttar Pradesh, West Bengal and Bihar. (FAOSTAT 2023). Potato cultivation has become increasingly important in recent years for the state of Madhya Pradesh because of the availability of suitable agro-climatic conditions and the expansion of irrigation. The area and production of potato in the state has grown over the years, with production reaching around 3925.65 thousand tonnes in recent years. Growth in potato area and production in the state has been phenomenal with more than 500% and 1000% increase in area and production over time, respectively. (Mishra 2025) But even though there is an increasing trend in area, production, and productivity, there are significant fluctuations in potato cultivation, which are attributed to climatic conditions, price, pest and disease outbreaks, and fluctuations in inputs. This results in production instability that impacts farmers' livelihood and the sustainability of agriculture. Hence, trend analysis, instability and forecasting of future production trends should be considered for planning and policy making. With this backdrop, the current research attempts to study the trends in area, production, and productivity of potato, instability, and sustainability using relevant indices and forecasting models to gain insights into the long-term trends in potato cultivation in the state of Madhya Pradesh.

2. Methodology

2.1 Instability

For measuring the instability in area, production and yield the index given by Cuddy and Della (1978) and used by Kumar *et al.*, (2022) ^[9]:

$$CV_t = (CV) \times \sqrt{1 - R^2} \quad \text{where, } C.V. = \frac{\sigma}{\bar{X}} \times 100$$

Where

σ = Standard Deviation

$\bar{X}$ = Mean

R^2 = coefficient of determination of the linear trend model of the variable concerned.

CV_t = CV around trend

The coefficient of variation (CV) is a popular measure of instability. But in the presence of a trend, the CV is not an appropriate measure of variability as it ignores the trend effect (Hasan *et al.*, 2008).

To address this issue, the approach suggested by Cuddy and Della Valle (1978) is commonly used, which modifies the CV by removing the trend effect, and thus gives a better measure of instability (CV_t).

However, the Cuddy–Della Valle method assumes a linear trend which may not be suitable for long-term time series data, where non-linear trends are often present. Aware of this issue, the current study uses non-linear trend models to capture the true nature of the trend. The coefficient of determination (R^2) of the best-fitted model is used for calculating the adjusted coefficient of variation (CV_t) of various series.

For the sake of convenience, the time period is also divided into two parts: Period I (before liberalization) and Period II (after liberalization) in order to compare the degree of instability in these two periods.

For the instability analysis divided the whole study period in two equal periods

- Period I 1960 to 1990
- Period II 1991 to 2022
- Whole period 1960 to 2022

2.2 Box- Jenkins model

The advent of the Box-Jenkins method in the 1970s marked the beginning of a new era for time series forecasting, and the use of this method has been made easier with the availability of computer packages. This method is based on the idea that the value of a time series at any given time depends on its history and random shocks. The Autoregressive Moving Average (ARMA) model is a popular model for time series analysis and forecasting. This model is a mixture of two models: the autoregressive (AR) model that accounts for the current value as a function of its own past values, and the moving average (MA) model that accounts for the effect of past random shocks. This combination of the two is known as an ARMA model.

Autoregressive Model: Autoregressive Integrated Moving Average (ARIMA) is a popular statistical model for time series analysis and forecasting. The word "integrated" means that the original time series has been differenced to stationarise it or remove any trends. If the series is already stationary and doesn't need to be differenced, the model is an ARMA (Autoregressive Moving Average) model. The AR(p) in the model stands for autoregressive model of order p, which means the current value of the series is explained by the p previous values. That is, the AR(p) model assumes that the current value of the series is a linear combination of the past values and an error term.

$$X_t = c + \sum_{i=1}^p \alpha_i X_{t-i} + \mu_t$$

Here, the coefficients are the parameters to estimate, c is a constant, and it is the random error (white noise) term. The constant term is sometimes omitted from the model.

Moving Average model: The notation MA (q) refers to the moving average model of order q:

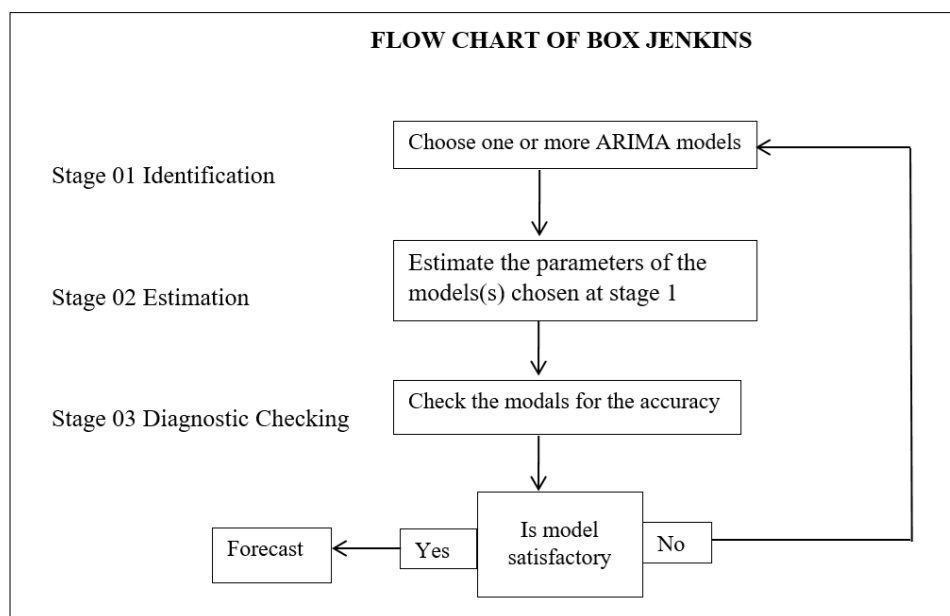
$$X_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

In this context, $\theta_1, \dots, \theta_q$ represent the model parameters, μ denotes the expected value of X_t (often believed to be 0), and the term signifies an error component.

Stationarity:

The joint distribution function defining the process must be time invariant, with $F(z_t, \dots, z_{t+k}) = F(z_{t+s}, \dots, z_{t+s+k})$ for s and all k , according to Box and Jenkins (1976), Anderson (1976) [2], Judge *et al.* (1982) [8], Chatfield (1984) [5], and Pankratz (1983) [11]. The covariance, variance, and mean are all constant due to this strong stationarity condition. The

majority of economic data increase with time, which causes the economic time series' mean to fluctuate and produce a nonstationary time series. A non-stationary time series in this situation needs to be transformed into a stationary one. Differencing or transformation can yield the same result, and in most cases, we don't need to vary more than once to demonstrate stationarity.

**Identification Techniques:**

The statistical features of a time series, such as its mean, variance, autocorrelation function (ACF) and partial autocorrelation function (PACF), are determined from the data to gain insight into its structure. The ACF and PACF, specifically, are used to determine the amount of correlation between observations and to suggest a suitable model in the model identification stage. But the model is only finalised after estimation and diagnostic analysis. Generally, if the PACF cuts off at lag p and the ACF tapers off, an $AR(p)$ is suggested. Then, if the ACF shows a sharp cut-off at lag q and the PACF dies down, a moving average model of order q [$MA(q)$] is appropriate. If the ACF and PACF both taper off, an $ARMA(p, q)$ model is appropriate. If the ACF does not taper off, it suggests the series is non-stationary, and it is necessary to transform it (e.g. through differencing) to make it stationary.

Estimation:

After determining suitable values of p and q , the parameters of the autoregressive (AR) and moving average (MA) components need to be estimated. In the Box-Jenkins approach, this is usually done by the maximum likelihood method or non-linear least squares, which can be difficult. The AR models are relatively simple as they are linear in parameters and lagged values, and the parameters can be estimated via least squares. But models containing MA terms are more complex due to the unobserved nature of the errors. This means that the model cannot be estimated directly using multiple regression, and requires non-linear estimation techniques. While the forecasts of ARIMA models are non-linear in the parameters, they are linear in past observations.

Model Formulation

The whole period under consideration (1960-2022) has been divided into two parts

- The model formulation period (1960-2018).
- Model Validation period (2019-2022)
- Forecasting period up to 2030

Checking for model adequacy:

The optimal Box- Jenkins model is chosen based on maximum R^2 , minimal root mean square error (RMSE), minimum mean absolute percentage error (MAPE), and minimum Akaike information production among the competing models (AIC). Any model that meets the majority of the above criteria is chosen.

R-squared:

A measure of how much of the standard deviation in the series is explained by the model. When the series is stationary, this metric is most useful. Positive values indicate that the current model is superior to the baseline model.

$$R^2 = \frac{\sum_{i=1}^n (\hat{X}_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Root Mean Square Error (RMSE):

The square root of mean square error. A measure of how much a dependent series varies from its model-predicted level, expressed in the same units as the dependent series.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}}$$

Mean Absolute Percentage Error (MAPE):

A metric for how far a dependent series deviates from the level predicted by the model. It is unaffected by the units used, hence it can be used to compare series with different units.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right|}{n} \times 100$$

Mean absolute error (MAE):

Measures how much the series varies from its model-predicted level. MAE is reported in the original series units.

$$MAE = \frac{\sum_{i=1}^n |X_i - \hat{X}_i|}{n}$$

Akaike's Information Creation (AIC): It is an estimator of prediction error and thereby relative quality of statistical model for a given set of data. AIC provide a means for model selection. AIC mathematically defined as, $AIC = -2\log(\text{maximum likelihood}) + 2k$

Where, $k = p+q+1$ (if model include intercept) otherwise $k=p+q$. (Kumar *et al.*, 2019)

Diagnostic checking:

This stage allows you to see if the model you've picked is a good fit for the data. The residuals calculated from the chosen model can be checked to see if they are white noise; if they are, the fit can be approved; if not, the operation must be resumed, making the Box Jenkins Methodology an iterative process (Mishra *et al.*, 2013). The study estimates ARIMA equations for all parameters using data from 1960 to 2020 and forecasts up to 2026.

Forecasting:

ARIMA's success in predicting is one of the reasons for its popularity. The fundamental Box-Jenkins approach for forecasting the values of a time series is as follows.

- First, check for stationarity. The autocorrelation function (ACF) and partial autocorrelation (PACF) can be computed, or a normal root analysis can be used.
- If the time series isn't stationary, divide it by one or more to attain stationarity.
- The stationary time series' ACF and PACF are then calculated to determine if the series is fully autoregressive, solely moving average, or a combination of the two.

d) Finally, the preliminary model is estimated.

- e) This model's residuals are investigated to see if they are white noise. If they are, the tentative model is most likely a decent approximation of the stochastic process at hand. If they aren't, the procedure is restarted from the beginning. As a result, the Jenkins approach in the box is iterative. The final model used can be used to forecast.

ACF and PACF graphs of the residuals are used to perform diagnostic tests on the fitted models. If the residuals are white noise, the model is used solely for predicting.

3. Result and Discussion

The information gathered from secondary sources was evaluated and inferred in line with the analysis's objective. The chapter that follows this section discusses the study's findings and conclusions.

3.1 Instability of Potato in Madhya Pradesh

The instability in area, production, and Productivity of potato in Madhya Pradesh across all three periods was computed and presented in Table & Fig. 3.1 The coefficient of variation around the trend for potato during the period 1960 to 2022 was also reported in the same table. The overall instability analysis, revealed that potato production exhibited the highest, instability (33.23), followed by area (21.66) and productivity (14.50). These finding indicated a relatively lower risk in the presence of decadal variations in instability across area, production, and productivity. In term of area under cultivation, the lowest, instability was observed during period I (6.61). Similarly, potato production, recorded comparatively lower instability in period I (15.03), while productivity also exhibited minimum variability during the same period (12.46). Indicating a phase of relative stability. In contrast, highest instability in area was observed, during period III (21.66). Period III (33.23) and period II (24.71) had the highest variability in production and also period III (14.50) and II (14.28) had higher instability in productivity. The instability may have been varied due to the lack of widespread dissemination and use of better technologies. In general, the findings indicated that potato cultivation in Madhya Pradesh was characterized by relatively low degree of risk, yet the variability was still observed over the decades.

Table 3.1: Instability in Area, Production, and Productivity of potato in Madhya Pradesh

Statistics	Period 1	Period 2	Whole period 3
Madhya Pradesh			
Area (000'ha)			
R ²	0.95	0.90	0.94
CV	31.39	58.87	88.82
CV _t	6.61	17.76	21.66
Production (000'MT ha)			
R ²	0.85	0.90	0.92
CV	39.45	81.64	123.96
CV _t	15.03	24.71	33.23
Yield (MT/ha)			
R ²	0.27	0.73	0.80
CV	14.61	27.57	32.70
CV _t	12.46	14.28	14.50

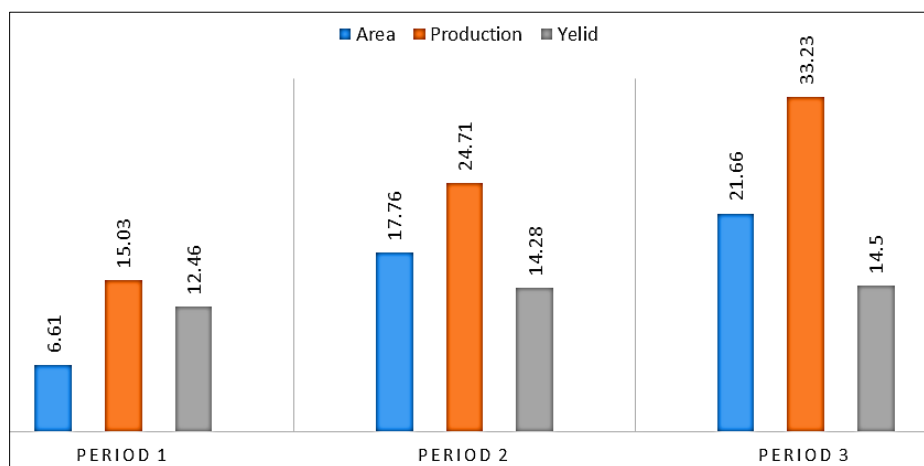


Fig 3.1: Instability in Area, Production and Productivity of potato in Madhya Pradesh

3.2 Modelling and forecasting

Following the analysis of the trend patterns for each of the time series, the next task was to develop future forecasts using the Box-Jenkins approach. The time period 1960-2018 was used to fit the model and 2019-2022 was used for validating the model. The best models were selected based on goodness-of-fit criteria and the selected models were used for prediction.

Analysis of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the potato series in Madhya Pradesh showed non-stationarity. To correct this, the potato series was differenced at lag-1, which resulted in a series with a constant mean and variance, suitable for ARIMA modelling.

The ACF and PACF plots of the differenced series also provided indications of the parameters. For potato area, the parameters were probably $p = 1$ and $q = 3$, while for potato production and potato productivity they were probably $p = 1$, $d = 1$ and $q = 5$. This provided a base for the estimation of several models, such as ARIMA (1,1,3), ARIMA (1,1,4) and ARIMA (1,1,5). The model was selected based on performance measures (RMSE, MAE, MAPE, AIC, and R^2)

and the diagnostics related to residuals. The findings suggested that ARIMA (1,1,3), ARIMA (1,1,4) and ARIMA (1,1,5) were the most suitable models for the series.

During the validation period (2019-2022), the estimated potato area in Madhya Pradesh was 161.12 thousand hectares, which was similar to the observed value of 172.21 thousand hectares, suggesting good model performance. Using the model, we projected a slow increase in potato area to 196.09 thousand hectares in 2030. Likewise, the potato production for the validation period was estimated to be 3740.6 thousand tonnes, which was close to the observed production of 3954.91 thousand tonnes, showing satisfactory accuracy. And the projected production could reach 4547.47 thousand tonnes. For productivity, the model estimated 21.56 tonnes per hectare during 2019-2022, whereas the actual productivity was 22.97 tonnes per hectare, suggesting good model accuracy. Projections also indicate that productivity may increase to around 23.59 tonnes per hectare. In general, the study reveals a steady expansion in potato area, production and productivity in Madhya Pradesh, along with good model fit and reliable forecasts.

Table 3.2.1: Different ARIMA Model for area, production and productivity under potato in Madhya Pradesh

Particular	Area	Production	Productivity
ARIMA Model	ARIMA (1,1,3)	ARIMA (1,1,5)	ARIMA (1,1,3)
R^2	0.964946	0.96848	0.824584
RMSE	8.86	200.25	1.877
MAPE	11.238	25.591	11.375
MAE	4.824	134.65	1.447
AIC	458.55	846.21	275.84

Table 3.2.2: Validation of predicated value of potato Area, Production, Productivity for Madhya Pradesh

Year	Variable	Observed	Predication
			ARIMA
2019	Area	151	148.52
	Production	3457.3	3391.9
	productivity	22.9	22.61
2020	Area	156.39	152.55
	Production	3566.94	3440.3
	productivity	22.81	21.65
2021	Area	158.43	158.14
	Production	3587.9	3607
	productivity	22.65	22.84
2022	Area	172.21	161.12
	Production	3954.91	3740.6
	productivity	22.97	21.56

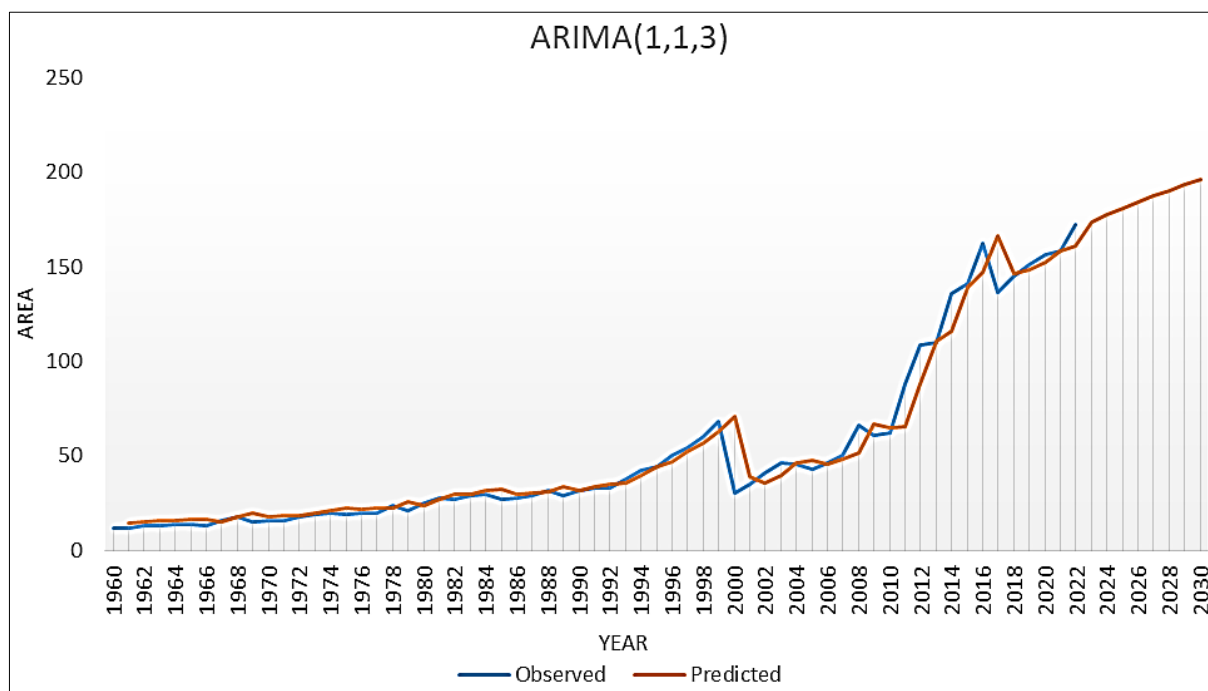


Fig 3.2.1: Observed and forecasting of area under potato in Madhya Pradesh

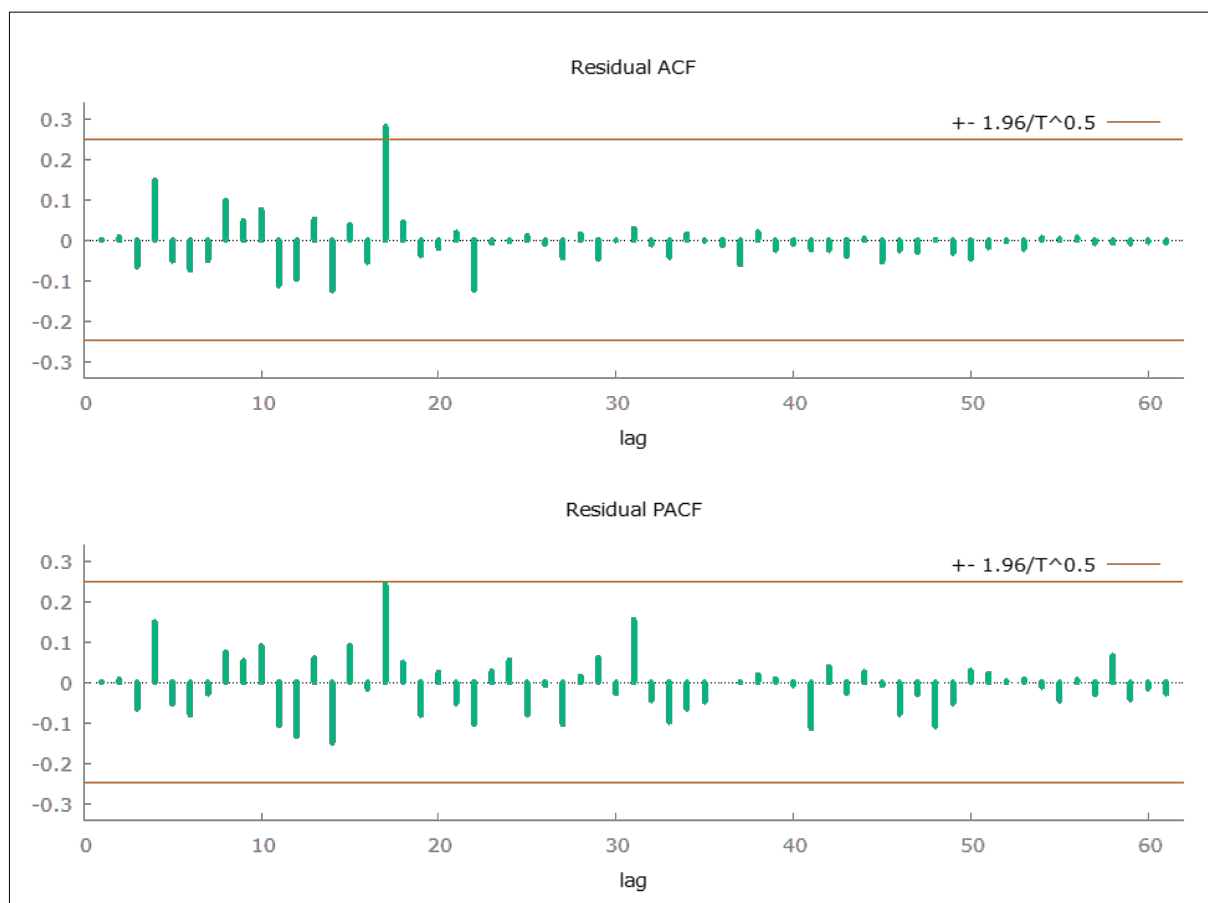


Fig 3.2.2: ACF and PACF graph of residuals for the best fitted models of area under potato in Madhya Pradesh

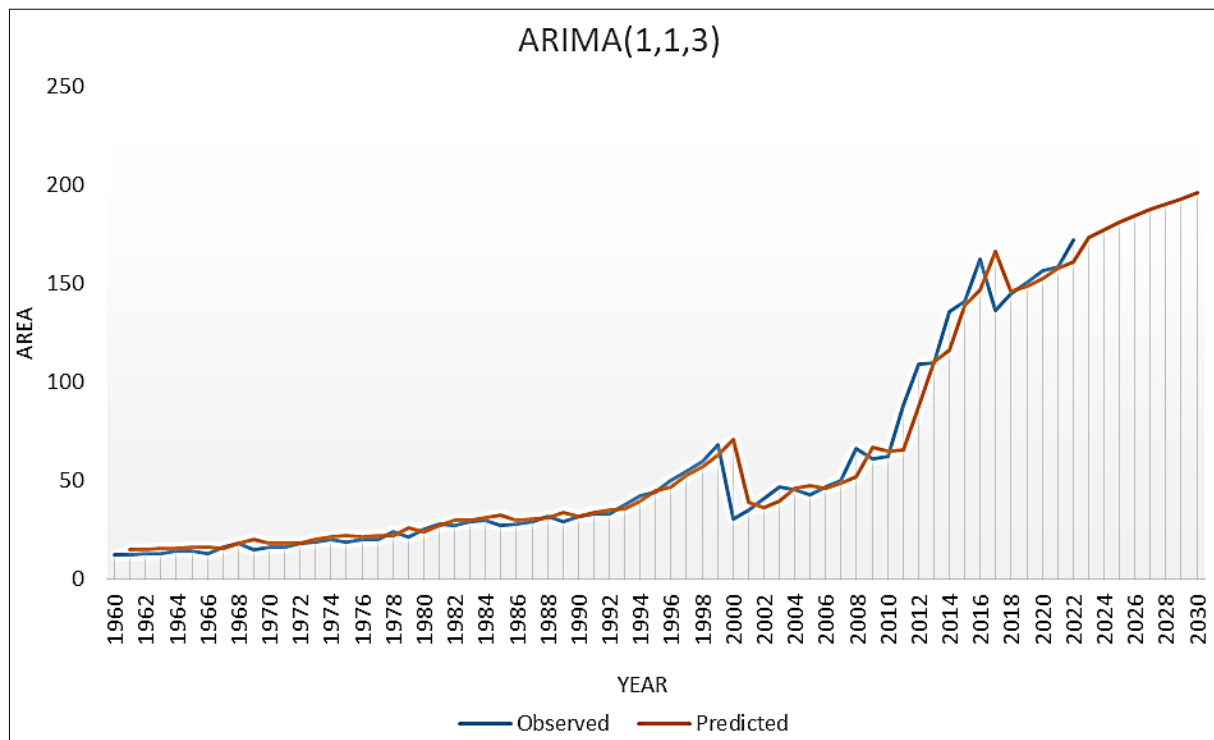


Fig 3.2.3: Observed and forecasting of Production under potato in Madhya Pradesh

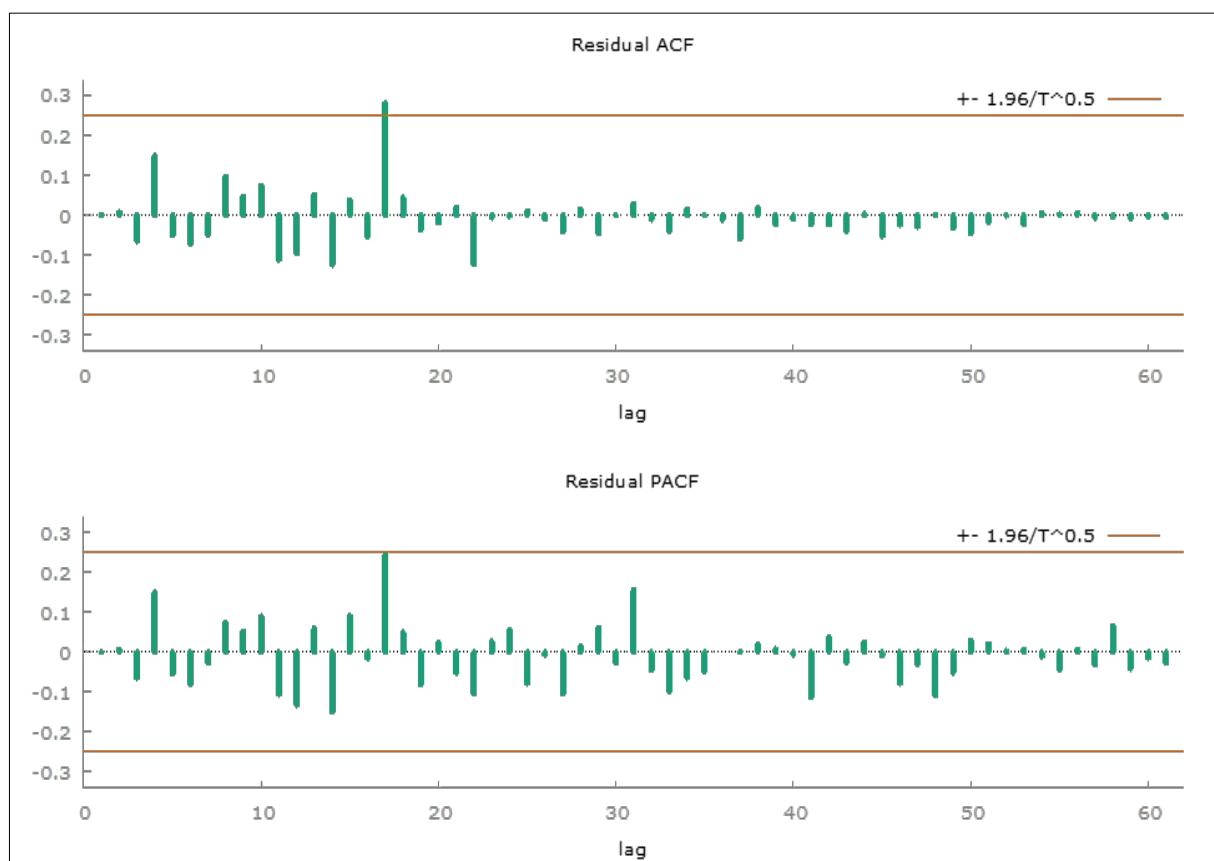


Fig 3.2.4: ACF and PACF graph of residuals for the best fitted models of Production under potato in Madhya Pradesh

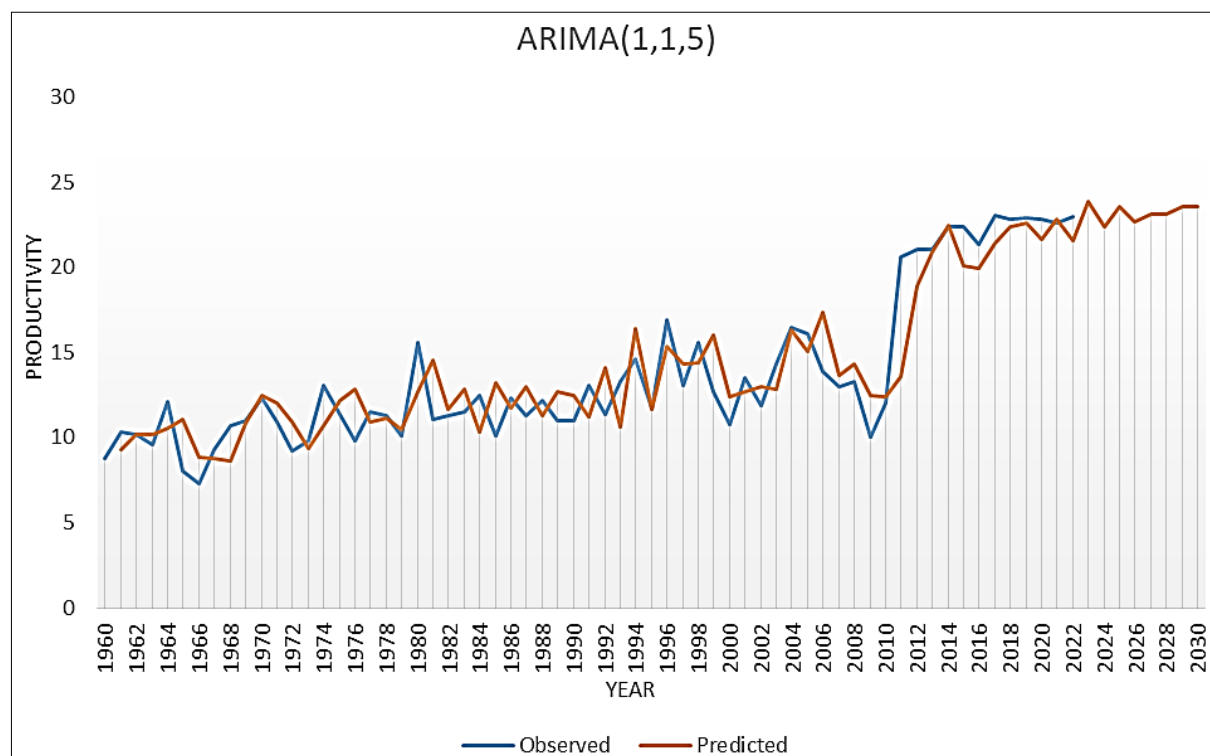


Fig 3.2.4: Observed and forecasting of Productivity under potato in Madhya Pradesh

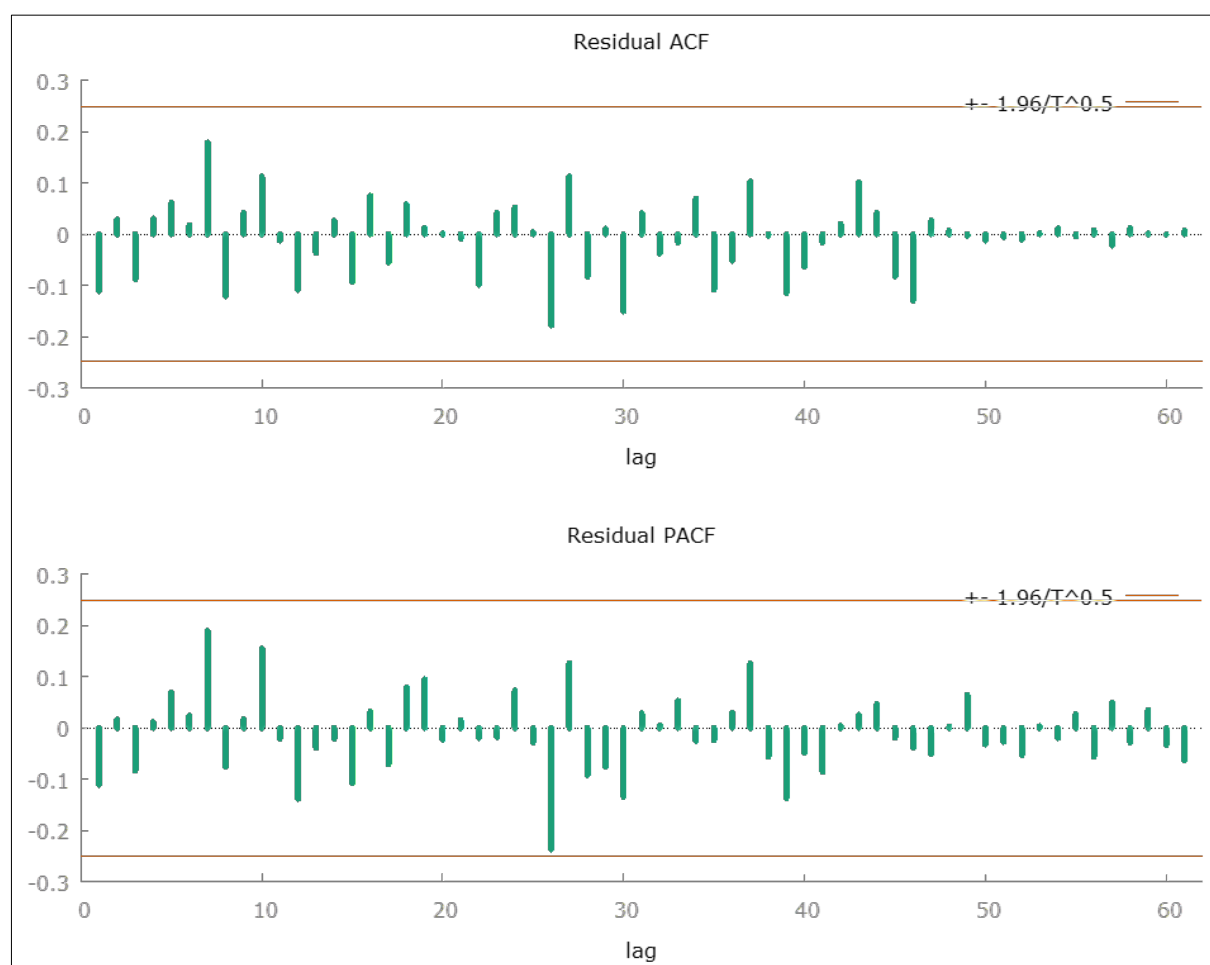


Fig 3.2.5: ACF and PACF graph of residuals for the best fitted models of Productivity under potato in Madhya Pradesh

References

1. Aashish, Mishra P, Singh U, Srivastava AB, Ray S, Badhe V, *et al.* Forecasting and yield sustainability insights into potato production in Madhya Pradesh, India. Agricultural Science Digest. 2025. doi:10.18805/ag.D-6282.
2. Anderson TW. The statistical analysis of time series. New York: Wiley; 1976.

3. Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Holden-Day; 1976.
4. Buckseth T, Tiwari JK, Singh RK, Kumar V, Sharma AK, Dalamu D, *et al.* Advances in innovative seed potato production systems in India. *Frontiers in Agronomy*. 2022;4:956667. doi:10.3389/fagro.2022.956667.
5. Chatfield C. The analysis of time series: An introduction. London: Chapman & Hall; 1984.
6. Choudhary B. Potential and status of potato production in Madhya Pradesh. In: [Book Title]. p. 83-89. ISBN: 978-93-91057-61-9.
7. Food and Agriculture Organization of the United Nations. Workshop to commemorate the International Year of the Potato: Proceedings of the workshop held at FAO, Bangkok. 2008. ISBN: 978-92-106005-6.
8. Judge GG, Griffiths WE, Hill RC, Lee TC. Introduction to the theory and practice of econometrics. New York: Wiley; 1982.
9. Kumar L, Mishra P, Singh RB, Sayyed M, Rathi D, Srivastava AB. Forecasting of area, production and productivity of wheat crop in Madhya Pradesh. *Indian Journal of Economics and Development*. 2022;18:1-10. doi:10.35716/IJED/22215.
10. Li L, Zhu T, Wen L, Zhang T, Ren M. Biofortification of potato nutrition. *Journal of Advanced Research*. 2024;75:23-34. doi:10.1016/j.jare.2024.10.033.
11. Pankratz A. Forecasting with univariate Box-Jenkins models: Concepts and cases. New York: Wiley; 1983.
12. Sood S, Bhardwaj V, Kumar V, Das R, Gupta VK, Mangal V, *et al.* Genetic progress in 50 years of potato breeding in India: Where do we stand? *European Journal of Agronomy*. 2022;141:126620. doi:10.1016/j.eja.2022.126620.