Special Proceedings: ISBN #: 978-81-950383-5-0
26th Annual Conference, 26-28 February 2024; pp 51-60

Forecasting Models for the Production of Walnut in Jammu and Kashmir - A Comparative Study

Nishant Jasrotia[1], Manish Sharma[1], Anil Bhat[2], Sunali Mahajan[1] and Shavi Gupta[1]

[1]*Division of Statistics and Computer Science, SKUAST-Jammu*
[2]*Division of Agril. Economics and ABM, SKUAST-Jammu*

Received: 30 April 2024; Revised: 03 June 2024; Accepted: 06 June 2024

Abstract

Horticulture is an important sector that contributes towards the economic growth of our country. The UT of Jammu and Kashmir is the largest producer of walnut in India and provides important source of livelihood for many people. The study aims to forecast the production of walnut for Jammu and Kashmir using Time series models. Therefore, Holt linear exponential Smoothing and Autoregressive Integrated Moving Average (ARIMA) model have been applied and it shows that ARIMA(1,2,1) is appropriate model for forecasting on the basis of minimum value of information criterion and maximum value of coefficient of determination as compared to other models. Based on the forecast provided by the proposed model, there is a projected 56.69 percent increase in walnut production for the year 2035 with respect to 2022. This increase is contingent upon policymakers implementing policies aimed at boosting production.

Key words: ARIMA; Walnut production; Holt's linear exponential smoothing model.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

Jammu and Kashmir is northern most state of India. The Jammu and Kashmir holds important position in horticultural production. The contribution of horticulture in gross state domestic product (GSDP) is more than 9 percent as per Sharma *et al.* (2023). Horticulture is an important sector which provides employment to people. Growing dependence on service sector for employment has added a lot of burden on union territory so; horticulture development makes a strong case. Walnut productions are light demanding species and are drought tolerant. Drought tolerating nature of walnuts makes a special case for their cultivation in Jammu & Kashmir. Walnuts (genus Juglans) are plants in the family Juglandaceae. Walnut is believed to have originated in Iran and its surrounding areas and brought to Europe by Alexander the Great and from Europe it was brought to China. In India walnut was earlier confined to Jammu and Kashmir and in late century it was brought to Himachal Pradesh, Uttarakhand hills and expanding up to Darjeeling and Sikkim. Walnut is called by

Corresponding Author: Manish Sharma
Email: manshstat@gmail.com

different names in different parts of India. The most commonly used name is akhroot, while in Kashmir it is called dun. Walnuts became the viable horticulture industry in India since 1980s particularly in the valley of Kashmir (Pandey and Shukla, 2007). Jammu and Kashmir occupy almost 90 per cent share of walnut industry in India. According to provisional data from the National Horticulture Board, Jammu and Kashmir recorded a production of 206.43 thousand metric tons of walnuts, cultivated across 69.24 thousand hectares in the 2016-17 period. In contrast, the rest of India produced 21.8 thousand metric tons of walnuts, covering an area of 22.85 thousand hectares during the same period (Horticulture Statistics at a Glance 2017). Jammu and Kashmir have been declared as an Agri-Export Zone for walnuts as discussed by Shah *et al.* (2021). The major walnut production areas of Jammu and Kashmir are Anantnag, Kupwara, Kulgam, Budgam, Doda, Poonch, Kishtwar, Rajouri and Kathua. The demand of Kashmiri walnut is increasing rapidly which needs to bring more land under it and require a regular attention to this industry so that it can better flourish in the times to come. Forecasting is an important problem that spans many fields including business and industry, government, economics, environmental sciences, medicine, social science, politics, agriculture and finance. Forecasting problems are often classified as short-term and long-term. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, and months) into the future and long-term forecasting problems can extend beyond that by many years. Short term forecasts required for activities that range from operations management to budgeting and selecting new research and development projects. Long-term forecasts impact issues such as strategic planning. Most forecasting problems involve the use of time series data which is a time-oriented or chronological sequence of observations on a variable of interest. It is a sequential set of data points, measured typically over successive times. The measurements taken during an event in a time series are arranged in a proper chronological order. The time series in general supposed to be affected by four main components, which can be separated from the observed data. These components are: Trend, Cyclical, Seasonal and Irregular components. A time series model is linear or non-linear whether the variable of interest is forecasted using a linear or non-linear combination of past value of the variable. The linear time series models are designed to model the auto-covariance structure in the time series. Some of the forecasting models like Exponential smoothing, Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). The two popular sub groups of Linear time series models are the Autoregressive (AR) and Moving Average (MA) models. AR models combined with moving average (MA) model to form a general and useful class of time series models called the Autoregressive Moving Average (ARMA) models. Sharma *et al.* (2018) applied Box-Jenkins methodology to build Autoregressive Integrated Moving Average (ARIMA) model for monthly arrival of Rohu fish in Jammu region of J&K state among many models the best model obtained was ARMA (2, 2) on the basis of significance of model and parameters. Mahajan *et al.* (2020) applied ARIMA model for the production of Rice crop in India. The best model would be ARIMA(0,2,2) on the basis of minimum AIC and SBIC. Kumari *et al.* (2022) applied Exponential smoothing and ARIMA model for the area, production and productivity of total fruit crops in Gujarat.

2. Material and methods

This study used annual time series data of walnut production in MT from 1973 to 2022 (Directorate of Horticulture Kashmir). We have to find Instability Index to check if the time series exhibit any trend. Cuddy Della Valle Index (CDVI) method has been used as proposed Cuddy and Della Valle (1978) for measuring the instability in time series data. It is measured through CDVI = CV $*\sqrt{1-\bar{R}^2}$ where, CV is the coefficient of variation in percent, and $\bar{R}^2$ is the adjusted coefficient of determination. An appropriate modeling technique has been used for the forecasting of walnut production. Box and Jenkins methodology or ARIMA modeling has been introduced by Box and Jenkins (1976) is commonly used for forecasting purpose. It combines the Autoregressive Process (AR) and Moving Average Process (MA). The structure of ARIMA model is; ARIMA (*p, d, q*), where p and q are the order of the autoregressive and moving average process respectively while d is the order of differencing. The mathematical form of ARIMA (p, d, q) model is:

$$Y_t = C + (\Phi_1 Y_{t-1} + \ldots + \Phi_{t-p} Y_{t-p}) + \epsilon_t(-\theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q})$$

where, C is the constant, Y_t is the data on which the ARIMA model is to be applied, $\Phi_1, \ldots \Phi t - p$ are AR coefficients, $\theta_1, \ldots, \theta_q$ are MA coefficients and e_t is the random error. However, the AR model of order p is $Y_t = C + (\Phi_1 Y_{t-1} + \ldots + \Phi_{t-p} Y_{t-p}) + e_t$. Similarly, the MA structure of order q is $Y_t\theta = C + \epsilon_t - (\theta_1 \epsilon_(t-1) + \ldots + \theta_q \epsilon_(t-q))$.

The three main stages of Box-Jenkins forecasting model are used, first is identification of the model or specification of the model second is estimating the parameters and third is diagnostic checking of the residuals and forecasting. Another most commonly used univariate time series forecasting technique is the exponential smoothing (ES). In this technique, forecasts are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. Exponential smoothing method classified according to the type of component presented in the time series data. The current study exclusively employs a single exponential smoothing method, namely Holts linear trend exponential smoothing technique, utilizing time series data. Holt (2004) introduced an extension of simple exponential smoothing tailored to forecast data exhibiting a trend. This method entails a forecast equation and two smoothing equations.

Forecast equation $\hat{Y}_t = l_t + hb_t$

Level Equation $l_t = \alpha Y_t + (1-\alpha)(l_{t-1} + b_{t-1})$.

Trend Equation $b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}$.

where, Y_t, $\hat{Y}_t$ is observed and predicted value of series at time t, l_t and b_t are the estimate of level and trend of the series at time t. The α, β are the smoothing parameters for the level and the trend, $0 \leq \alpha, \beta \leq 1$.

Model selection is done on the basis of following measures:

- Akaike's information criterion (AIC) proposed by Akaike (1979) is a useful statistic for model identification and evaluation. It defined as $-2logL + 2n$, where, L is the likelihood function and n is the number of hyper parameters estimated from the model.

- Bayesian Information Criterion (BIC) also as known as Schwarz Criterion. Schwarz (1978) proposed the criterion from Bayesian likelihood maximization. And is defined as $SBIC = -2LogL + nLogT$ where, T is total number of observations.

- Coefficient of determination (R^2) Wright (1921) and calculated as $R^2 = 1 - RSS/TSS$. where, RSS is residual sum of square and TSS total sum of square Range of R^2 is 0 to +1.

- Mean Absolute Percentage Error: The mean absolute percentage error (MAPE) is one of the most popular measures of the forecast accuracy. It was used as the primary measure in the M-competition Makridakis *et al.* (1982). MAPE is defined as $MAPE = \frac{1}{N}\sum_{t=1}^{N} |\frac{A_t - F_t}{A_t}| * 100$ where, A_t is the actual value, F_t is the forecast value, N is the number of data points.

3. Results and discussion

The annual time series data of the walnut production in Jammu & Kashmir for the last fifty years are used in the forecasting model. The data has been segmented into two 25-year phases: 1972-1997 and 1998-2022, aiming to analyze trends and stability. Notably, the average production during the latter phase surpassed that of the former (refer to Table 1). A similar trend is evident in the standard deviation. Consequently, the instability index was higher during the initial phase (33.30 percent) compared to the second phase (12.49 percent), possibly attributed to government subsidies, support programs, and the use of high-quality seeds to promote walnut cultivation. The overall instability index of the walnut production was 32.62 percent clearly indicating the production instability. The increasing market demand for walnuts incentivized farmers in Jammu and Kashmir to ramp up production for greater profits. The Durbin-Watson test yielded a value of 0.257, indicating positive autocorrelation among the residuals, necessitating the utilization of time series models.

Table 1: Descriptive statistics of walnut production from 1973-1997 and 1998-2022

Period	Mean (MT)	Std. Dev. (MT)	Max.Value (MT)	Min.Value (MT)	Instability Index (%)	Durbin Watson test (Overall)
Phase I (1973-1997)	27954.20	18620.05	68880.00	10212.00	33.03	0.257
Phase II (1998-2022)	169281.00	75563.61	279422.00	74906.00	12.49	
Phase III (1973-2022)	98617.60	89786.98	279422.00	10212.00	32.62	

Both graphical and empirical methods have been employed for this investigation. The line chart presented in Figure (1) illustrates an upward trend in walnut production. Additionally, long-term patterns suggest that the data is non-stationary, with a mean production of 98,617.60 metric tons (MT) and a standard deviation of 88,884.50 MT.

Table 2: ADF test value of actual Series and second differenced series

Test Statistic	Actual Series		Second Differenced Series	
	Value	P-Value	Value	P-Value
ADF	1.05	0.99(NS)	-10.12	0.000**

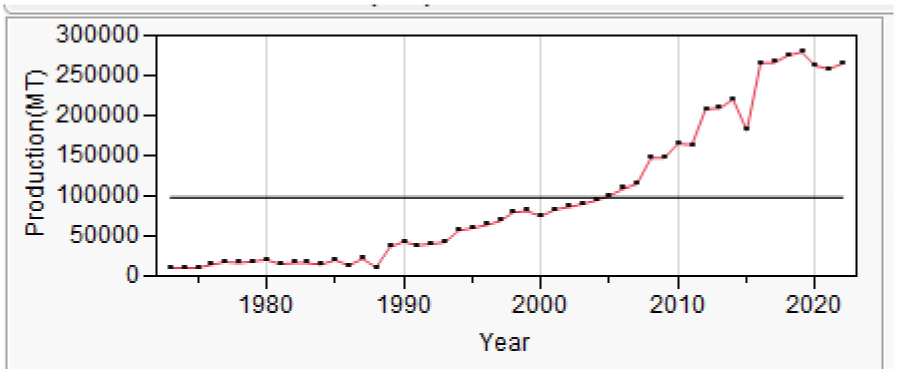


Figure 1: Trend of the annual production of walnut in Jammu and Kashmir

The Augmented Dickey-fuller (ADF) test value (see Table 2) is 1.05 and non-significant. It depicts non-stationarity of the data and p-values of Ljung-Box Q values are significant which means the residuals are dependent.

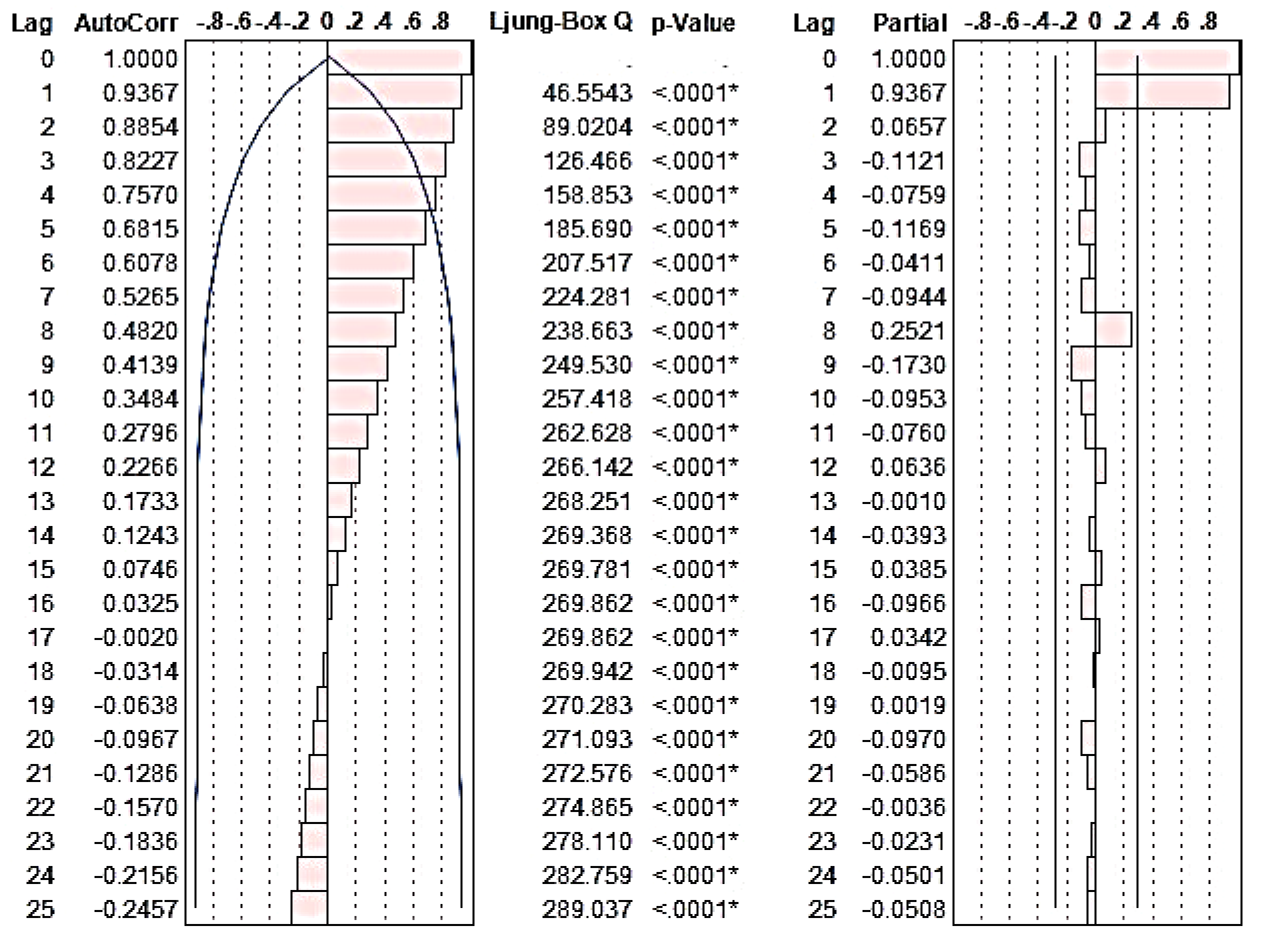


Figure 2: ACF and PACF plots of production of walnut in Jammu and Kashmir

The ACF and PACF plots in figure (2) show that the spikes are outside from the insignificant zone and fail to follow the assumption of randomness. Therefore, in order to met the stationary of the data first is to apply differencing method Mahajan *et al.* (2020).

The line chart, correlogram and ADF test have been used again after taking first and second order differencing until stationarity is achieved.

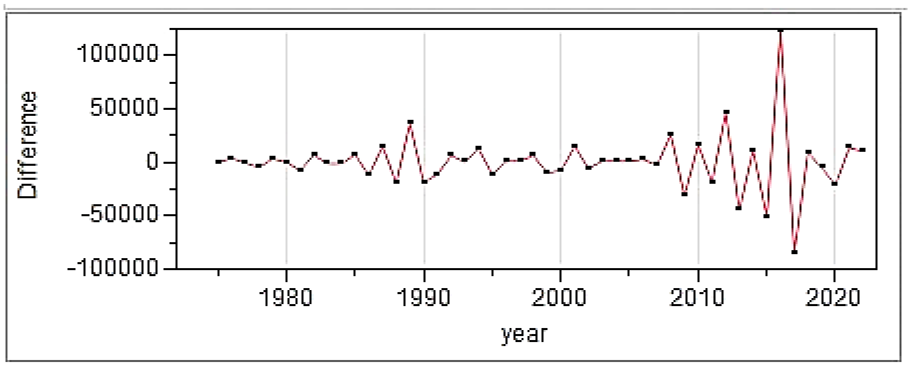
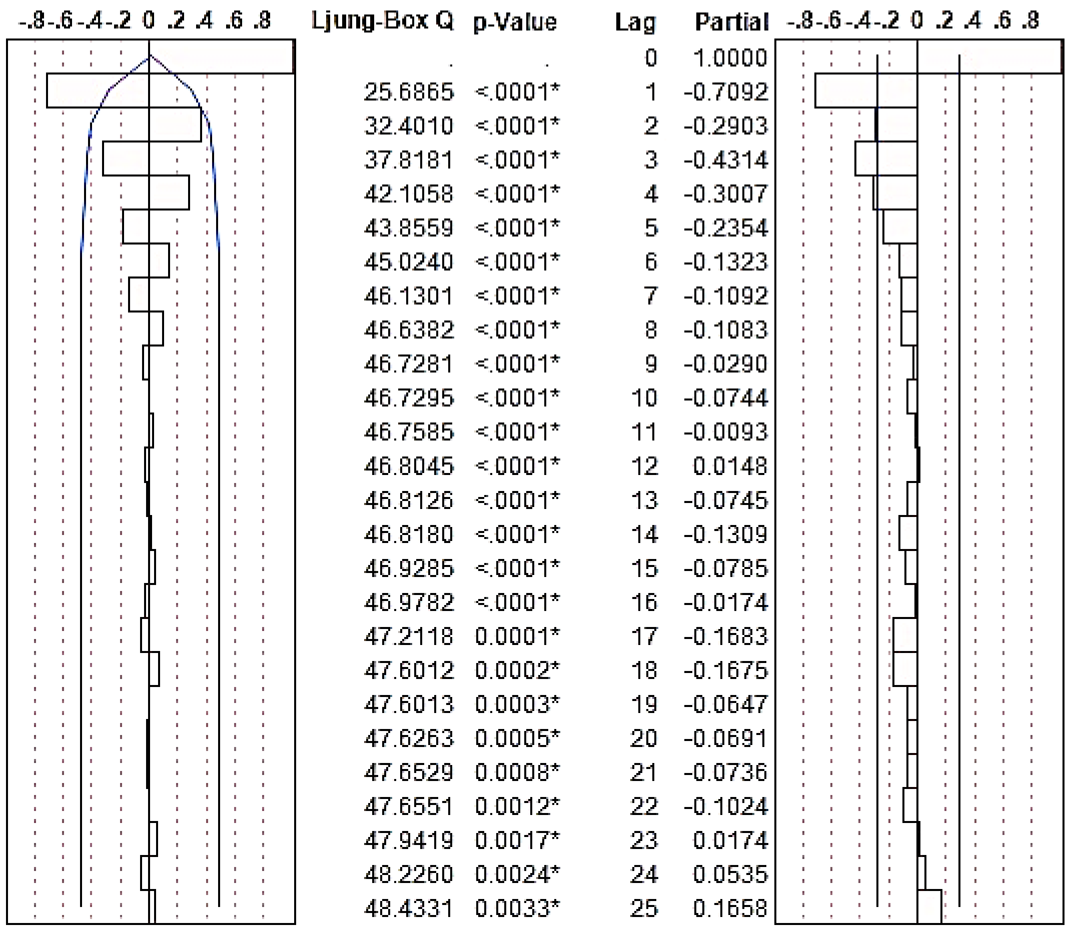


Figure 3: Trend of the data after taking differencing of order 2



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.7092	25.6865	<.0001*	1	-0.7092
2	0.3587	32.4010	<.0001*	2	-0.2903
3	-0.3187	37.8181	<.0001*	3	-0.4314
4	0.2804	42.1058	<.0001*	4	-0.3007
5	-0.1771	43.8559	<.0001*	5	-0.2354
6	0.1430	45.0240	<.0001*	6	-0.1323
7	-0.1375	46.1301	<.0001*	7	-0.1092
8	0.0920	46.6382	<.0001*	8	-0.1083
9	-0.0382	46.7281	<.0001*	9	-0.0290
10	0.0047	46.7295	<.0001*	10	-0.0744
11	0.0212	46.7585	<.0001*	11	-0.0093
12	-0.0263	46.8045	<.0001*	12	0.0148
13	-0.0108	46.8126	<.0001*	13	-0.0745
14	0.0087	46.8180	<.0001*	14	-0.1309
15	0.0390	46.9285	<.0001*	15	-0.0785
16	-0.0258	46.9782	<.0001*	16	-0.0174
17	-0.0549	47.2118	0.0001*	17	-0.1683
18	0.0698	47.6012	0.0002*	18	-0.1675
19	-0.0014	47.6013	0.0003*	19	-0.0647
20	-0.0171	47.6263	0.0005*	20	-0.0691
21	-0.0173	47.6529	0.0008*	21	-0.0736
22	0.0048	47.6551	0.0012*	22	-0.1024
23	0.0547	47.9419	0.0017*	23	0.0174
24	-0.0533	48.2260	0.0024*	24	0.0535
25	0.0446	48.4331	0.0033*	25	0.1658

Figure 4: Trend and correlogram of the data after taking differencing of order 2

The line chart in Figure (3) after taking the second order differencing shows that mean has no change and having constant variation. Thus, stationarity has been achieved through differencing. Moreover, ADF test value after second order differencing found to be significant which indicate that data are stationary. The order of (p) and (q), as depicted in Figure (4) through the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), is determined to be one. In the ACF, only one spike falls outside the range, while in the PACF, two spikes extend beyond the range. This suggests that the data has achieved stationarity. Different ARIMA models have been developed on the basis of Box-Jenkins methodology. Among them five best models have been proposed on the basis of minimum AIC (Akaike Information Criterion), SBIC (Schwartz Bayesian Information Criterion) and R2. The model ARIMA(1,2,1) have minimum AIC(1068.32), and SBIC(1073.94) values with R2 (0.97) with significant parameters found to be the best. The estimates of the parameters are shown in Table (3) having AR(1) as -0.40. MA(1) 1.00 are found to be significant respectively.

Table 3: Different models for annual production of rice in India

Models	R^2	AIC	SBIC	Significance of Parameters/models
ARIMA(1,2,2)	0.95	1094.00	1101.54	Non-Significant
ARIMA(1,2,1)	**0.97**	**1068.32**	**1173.94**	**Significant**
ARIMA(2,2,1)	0.97	1092.63	1100.19	Non-Significant
ARIMA(2,2,2)	0.97	1072.00	1082.01	Non-Significant
ARIMA(1,2,3)	0.94	1093.62	1100.01	Significant

Table 4: Parameters estimates of ARIMA(1,2,1) for production of walnut

Terms with orders	Estimates	Standard Error	t-Ratio	P-Value
AR(1)	-0.40	0.12	-3.10	0.00**
MA(1)	1.00	0.06	15.85	$< 00^{**}$
Intercept	202.30	108.03	1.87	0.06

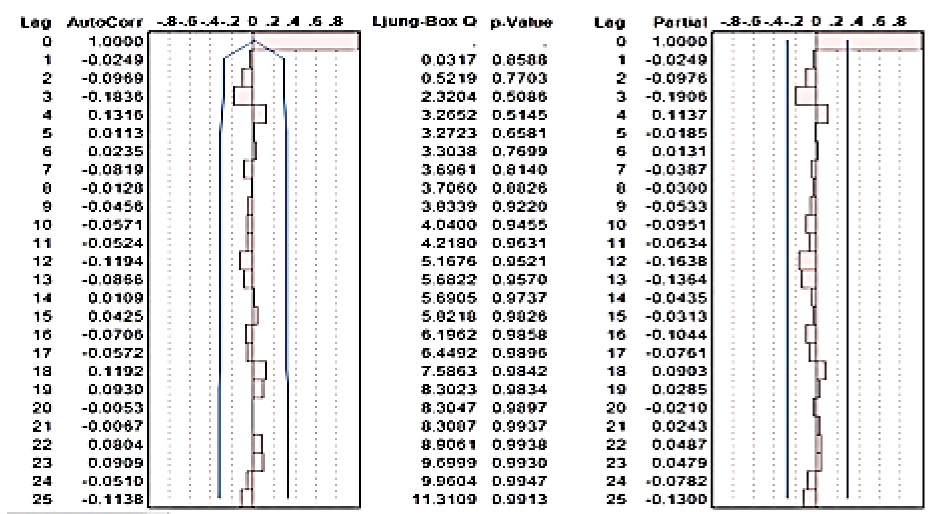


Figure 5: ACF and PACF plots of ARIMA(1,2,1)

The ACF and PACF plots in Figure (5) of the residuals indicate a good fit of the model, with p-values of the Ljung-Box Q test exceeding the significance level of 0.05. This implies that the residuals are independent, satisfying the assumption of randomness. Meanwhile, the (Holt) linear exponential smoothing model is employed to estimate walnut production. It exhibits an AIC of 1068.71 and SBIC of 1074.12, with an R-squared value of 0.96. Table (5) displays the parameter estimates of the Holt model. The Level smoothing weight is calculated to be 0.59 and the Trend Smoothing Weight is 0.14. The Level smoothing weight is found to be significant, while the Trend smoothing weight is not significant so this model is not considered.

Table 5: Parameter estimates of the linear (Holt) ES model for production of walnut

Term	**Estimate**	**Std Error**	**t-Ratio**	**Prob $> \|t\|$**
Level Smoothing Weight	0.59	0.13	4.29	$< .0001^{**}$
Trend Smoothing Weight	0.14	0.13	1.05	0.29

Comparison of performance of model done on the basis of AIC,SBIC, R2, and MAPE. Table (6) shows that the ARIMA(1,2,1) have maximum R2, with minimum Mean Absolute Percentage Error as compared to Linear Holt Exponential Smoothing model. So, on the basis of that ARIMA(1,2,1) model is best model for forecasting of walnut production of Jammu and Kashmir.

Table 6: Comparison of performance of different fitted models

Model	**Model Selection Measures**						
	AIC	**SBIC**	R^2	**MAPE %**	**Forecasted Value (2022)**	**Actual value (2022)**	**Difference**
ARIMA	1068.32	1073.94	0.97	13.10	274397.68	265423	8974.68
Holt Linear Smoothing Model	1068.71	1074.12	0.96	13.97	276750.55	265423	11327.55

Thus, the proposed model for estimating walnut production is ARIMA(1,2,1), specified as: $\bar{Y} = 202.30 + (-0.40)[Y_t + Y_{t-1} + Y_{t-2}] + 1.000[\epsilon_t - \epsilon_{t-1} - \epsilon_{t-2}]$. For validation, data are splitted into training(80 percent) and testing (20 percent) and as per the result of Table (7) the testing has minimum RMSE as compared to the training. Thus the model is best fitted.

Table 7: RMSE of training and testing for ARIMA (1,2,1)

Model	**RMSE**	
	Training	**Testing**
ARIMA (1,2,1)	115859.98	96998.59

The graphical representation of forecasting of ARIMA(1,2,1) model in Figure (6), shows there is upward trend in the production of walnut in Jammu and Kashmir.

Table 8: Forecasting values for walnut production in Jammu and Kashmir

year	Forecasted values (MT)	U-95	L-95	Percentage increase in production on the basis of 2022
2025	297927.24	339403.54	256450.95	12.24
2030	354356.67	417331.60	291381.74	33.50
2035	415908.24	494827.67	336988.81	56.69

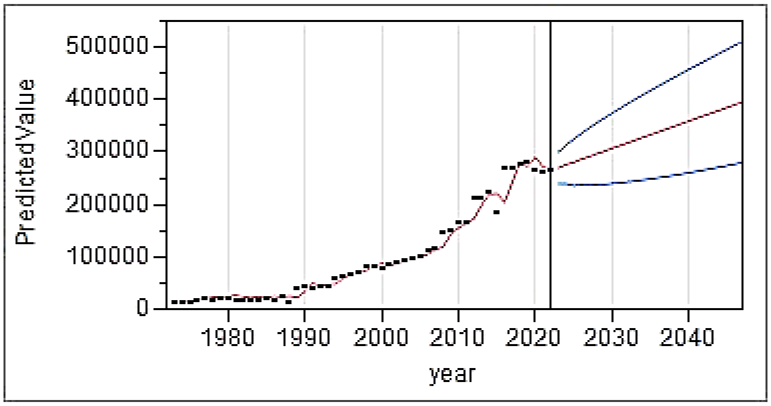


Figure 6: Forecasting graph of ARIMA (1,2,1) for annual production of Walnut

After conducting a forecast, the results indicate minimal variance between the actual and predicted values for the year 2022, affirming the accuracy of the model. The model's effectiveness is further validated by evaluating the lower and upper bounds of the forecasted values. Table (8) presents the projected walnut production for the years 2025, 2030, and 2035 based on this model. It illustrates a consistent upward trend in future walnut production. Specifically, the percentage increase in walnut production for 2025, 2030, and 2035 is 12.24%, 33.50%, and 56.69% respectively.

4. Conclusion

The current research aimed to forecast walnut production in Jammu and Kashmir, employing several time series models including linear Holt exponential smoothing and autoregressive integrated moving average (ARIMA). The findings suggest that the ARIMA model is the most appropriate for predicting walnut production. According to this model, the projected increase in walnut production for the year 2035 is 56.69%. This ARIMA model will be utilized for future walnut production forecasts, integrating up-to-date data.

Acknowledgements

I am indeed grateful to the Editors for their guidance and counsel. I am also thankful to the organizers of the Conference for giving me an opportunity to present my work.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, **66**, 237–242.

Box, G. E. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, volume 1.

Cuddy, J. D. and Della Valle, P. (1978). Measuring the instability of time series data. *Oxford Bulletin of Economics & Statistics*, **40**, 79–85.

Holt, C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages (onr memorandum no. 52). *International Journal of Forecasting*, **20**, 5–10.

Kumari, P., Parmar, D., Sathish Kumar, M., Mahera, A., and Lad, Y. (2022). Prediction of area, production and productivity of total fruit crops in Gujarat. *The Pharma Innovation*, **11**, 750–754.

Mahajan, S., Sharma, M., and Gupta, A. (2020). ARIMA modelling for forecasting of rice production: A case study of India. *Agricultural Science Digest-A Research Journal*, **40**, 404–407.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**, 111–153.

Pandey, G. and Shukla, S. (2007). The walnut industry in India - current status and future prospects. *International Journal of Fruit Science*, **6**, 67–75.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **1**, 461–464.

Shah, R. A., Bakshi, P., Sharma, N., Jasrotia, A., Itoo, H., Gupta, R., and Singh, A. (2021). Diversity assessment and selection of superior persian walnut (juglans regia l.) trees of seedling origin from north-western himalayan region. *Resources, Environment and Sustainability*, **3**, 1–14.

Sharma, M., Jasrotia, N., Kumar, B., Bhat, A., and Mahajan, S. (2018). Modeling of monthly arrival of rohu fish using ARIMA in Jammu region of J&K state. *Journal of Animal Research*, **8**, 259–262.

Sharma, M., Singh, I. J., and Gupta, S. (2023). Horticulture in Kashmir valley: opportunities and challenges. *Current Agriculture Research Journal*, **11**, 1057–1067.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, **20**, 557.