Journal of Agriculture, Biology and Applied Statistics
Vol. 2, No. 1, 2023, pp. 25-44

URL : www.arfjournals.com
https://DOI:10.47509/JABAS.2023.v02i01.03

Modeling and Forecasting of Lentil in India

K. P. Vishwajith[1], P. K. Sahu, Aditya Bhooshan Srivastava[2] and Rajani Gautam[3]
[1]Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur- 741 252, Nadia, West Bengal, India
[2]Department of Agricultural Economics, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya-224229 (Uttar Pradesh)
[3]Awedhesh Pratap Singh University, Rewa (M.P.), India
[*]Corresponding email: vishwajithkp@gmail.com

To cite this article
K. P. Vishwajith, P. K. Sahu, Aditya Bhooshan Srivastava & Rajani Gautam (2023). Modeling and Forecasting of Lentil in India. Journal of Agriculture, Biology and Applied Statistics. Vol. 2, No. 1, pp. 25-44. *https://DOI:10.47509/JABAS.2023.v02i01.03*

Abstract: In this study an attempt has been made to apply the autoregressive integrated moving average (ARIMA) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model approach to investigate the trend in lentil area, production and productivity in Bihar, Madhya Pradesh, Uttar Pradesh, West Bengal, and India. Yearly data from 1970 to 2009 were used for forecasting up to 2020.In comparison, we get that in area the GARCH model is the best compared to ARIMA for forecasting based on the maximum value of R2 and minimum value of MAPE, MPE, RMSE, MAE, AIC, BIC.Whereas, univariate ARIMA model outperformed in case of production and productivity. Furthermore, according to the trend analysis analysis signifies that production of lentil in many state has shown decreasing trend in recent past which is a major concern towards food and nutritional security. Also from the forecasting value it can be said that area, production in lentil would improve in future with a major concern in productivity front.Moreover, this study will also help make good policies with respect to the production scenario of lentil.

Keywords: Area, ARIMA, GARCH,trend, modelling, forecasting, production and productivity.

Introduction

India, with a large population of poor and malnourished people, has long encouraged a cereal-based diet based on subsidised staples like rice and wheat. Dietary habits, on the other hand, are changing today. Policymakers, researchers, and health advocates are exploring for strategies to combat malnutrition, not only hunger, in the country. Pulses (the dried, edible seeds of legumes) are becoming more popular as the focus shifts from calorie consumption to nutrients. Hunger can be classified into three types: calorie insufficiency, protein deficiency, and micronutrient deficiency.

Article History Received : 29 January 2023; Revised : 24 February 2023; Accepted : 05 March 2023; Published : 22 March 2023

Pulses are one of the most significant food crops in the world and India, where they contribute the most to financial gains. Chickpeas, pigeon peas, moong beans, black beans, lentils, peas, and a variety of other beans are among the major pulses farmed (Mishra et al., 2021). Most countries' pulses are produced in India, which accounts for 25% of global production. It also consumes the most pulses, accounting for 27% of global consumption. In 2018, total pulse production was 92.28 million tonnes (FAO, 2018), with dry beans accounting for 32.98 percent, chickpeas for 18.63 percent, peas for 13.53 percent, cowpeas for 7.83 percent, lentils for 6.86 percent, and pigeon peas for 6.45 percent. (Srivastava *et al.,* 2010).

Although India is the world's largest producer of pulses (23020 tonnes in 2019), local production is insufficient to meet domestic demand, and the country must import 3 million to 5 million tonnes of pulses each year (15 percent of global imports), making it the world's top pulse importer (Suresh and Reddy, 2016). After, arhar and gram, lentil is the third most important pulse crop.

Lentil is recognised as a valuable pulse crop in India. It is known to be the most nutritive of the pulses due to high protein content. It is grown as a winter crop and the sowing time extends from October to December. Since it is a short duration crop, it becomes ready for harvest in about three months. The crop is harvested from February to April depending upon the time of sowing. During 2010-11, lentil was grown on an area of 1597.40 thousand hectare with production of 943.80 thousand tones. It is mainly cultivated in Uttar Pradesh, Bihar, Madhya Pradesh, West Bengal and Rajasthan which together contribute more than 80 percent area and production of this crop. Uttar Pradesh is the key state with 36.68 percent of all India area and 43.54 percent of production. Next in the array are Bihar (14.93%), Madhya Pradesh (36.96%) and which produced around 22.74 percent and 18.80 percent of country's total lentil. West Bengal and Rajasthan are minor players in lentil cultivation but their yield was as high as 929 kg/ha and 972.50 kg/hectare respectively during 2010-11. Under 700 kg per hectare was the lentil production in Uttar Pradesh. Even though it ranks second in terms of output, Madhya Pradesh only generates 301.3 kg per hectare. When it comes to yield, lentils are second only to chickpeas in India. Box-Jenkins Modeling and predicting time series have been aided by ARIMA. Multiple harvests have been predicted using ARIMA. Sahu's (2006) findings aid in the forecasting of crop yields for potatoes, mustard, and wheat. The growing of onions in India is the subject of Mishra et al. Vishwajith et al. (2016) forecasted sugarcane production in India's major growing regions, and Rahman et al. (2013) predicted lentil pulse production in Bangladesh.

The linearity and homoscedaticity assumptions of the ARIMA model, which is frequently employed in time series data modeling, have been called into question. As a result, they looked for substitutes. Generalized autoregressive conditional heteroscedastic (GARCH) models are used in time series analysis. Paul et al. (2009) used the Box-Jenkins Auto Regressive integrated moving average (ARIMA) technique and the GARCH nonlinear time series model, including its estimate processes, to assess the risky export data from

India's spice industry. According to research by Yaziz *et al.* (2011), the GARCH expansion model outperformed the ARIMA model at predicting future crude oil prices. The ARIMA and GARCH models were used to analyze and forecast Indian pulse production by Vishwajith *et al.* (2014). However, the aforementioned study and others like it have mostly addressed modeling utilizing time series data of a specific phenomenon, but all crops depend on production parameters such as rainfall, temperature, relative humidity, fertiliser, etc. Research into production forecasting using ARIMAx is scarce in India. The best ARIMA, GARCH, and ARIMAx models in the competition forecasted gram production through 2020 [Vishwajith *et al.,* 2016]. Researchers made an effort to incorporate manufacturing variables in their model. In order to analyze historical data and estimate gram output in India's significant agricultural states, this study examined the ARIMA, GARCH, and ARIMAx models.

Materials and Methods

The main approaches to the research problem with their methodologies are discussed here:

Source of data

The data gathered is entirely secondary. The data on lentil production from 1970 to 2012 was collected from Directorate of Economics and Statistics.

Trend models

The model can be described as a means of presenting a process/system. The statistical model generally traces the path of the process along with its statistical properties and implications. In the present topic, we are interested in studying the path and nature of the series under our preview through different models, which are briefly given in table 1.

Table 1: Different trend models

Model	*Form*
Linear Model	$Y_t = b_0 + (b_1 t)$
Quadratic Model	$Y_t = b_0 + (b_1 t) + (b_2 t^2)$
Compound Model	$Y_t = b_0(b^t)$ or $\ln(Y_t) = \ln(b_0) + \ln(b_1)$
Cubic Model	$Y_t = b_0 + (b_1 t) + (b_2 t^2) + (b_2 t^3)$
Exponential Model	$Y_t = b_0 e^{(b1t)}$ or, $\ln(Y_t) = \ln(b_0) + (b_1 t)$
Logarithmic Model	$Y_t = b_0 + b_1 \ln(t)$
Growth Model	$\ln(Y_t) = b_0 + b_1 t$ Yt

Where, Yt is the value of the series at time t and b_0, b_1, b_2, b_3 are the parameters.

ARIMA models stand for AutoregressiveIntegrated Moving Average models. An ARIMA model is in-fact a combination of AR, MA models withintegration.

Autoregressive model (AR) : The notation AR(*p*) refers to the autoregressive model of order *p*. The AR (*p*) model is written as

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

where, 1, 2, ..., *p* are the parameters of the model, *c* is a constant and μ*t* is white noise i.e. *μt* ~ *WN* $(0, \sigma^2)$. Sometimes the constant term is omittedfor simplicity.

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

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

where, the θ1, ..., θ*q* are the parameters of themodel, μ is the expectation of *Xt* (often assumed to equal 0) and the t is the error term.

ARMA model : A time series {Xt} is an ARMA(p, q) if {Xt} is stationary and if for every *t*,

$$\boldsymbol{X_t - \emptyset_1 X_{t-1} - \cdots - \emptyset_p X_{t-p} = z_t + \theta_1 Z_{t-1} + \cdots + \theta_q z_{t-q}}$$

Where,{Zt} ~ WN(0, σ^2) and the polynomials$(1 - \emptyset_1 Z - \cdots - \emptyset_p Z^p)$ and$(1 - \emptyset_1 Z + \cdots + \emptyset_p Z^p)$ have no common factors.

ARIMA Model: A time series $\{X_t\}$ is an ARIMA (p,d,q) if $Y_t = (1\text{-}B)^d X_t$ is a casual ARIMA (p, q) process. This mean $\{X_t\}$ satisfies $\emptyset(B)X_t \equiv \emptyset(B)(1-B)^d X_t = \theta(B)Z_t$.

Where, $\{Z_t\}$ ~ WN $(0, \sigma^2)$ (z) and θ(z) are polynomials of degree *p* and *q* respectively and ϕ(*z*) = 0 for |*Z*| ≤ 1. The polynomial $\phi^*(Z)$ has a zero of order *d* at *z* = 1. The process {Xt} is stationary if and only if *d* = 0 and in that case it reduces to ARMA (*p*, *q*) process.

GARCH (p,q) Model : GARCH stands for Generalized Autoregressive Conditional Heteroscedasticity.

Generalized : It is developed by Bollerslev (1986) as a generalization of Engle's original ARCH volatility modelling technique.

Autoregressive : It describes a feedback mechanism that incorporates past observations into the present.

Conditional : It implies a dependence on the observations of the immediate past.

Heteroscedasticity: Loosely speaking, we can think of heteroscedasticity as time-varying variance.

To formally define GARCH, let $\varepsilon_1, \varepsilon_2, ..., \varepsilon_T$ be the time series observations denoting the errors and let Ft be the set of ε t up to time T, including εt for t d ≤ 0. $h_t = \alpha_1 + \alpha_1 \varepsilon^2_{t-1} + ... + \theta_q \varepsilon^2_{t-q} + \beta_1 h_{t-1} + ... \beta_p h_{t-p}$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

ARCH (q) : The ARCH model is a special case of a GARCH specification in which, there is no GARCH terms in the conditional variance equation. Thus ARCH (q) = GARCH(0, q). The process ht is an Autoregressive Conditional Heteroscedastic process of order q or ARCH(q), if ht is given by

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2$$

Where, $q > 0$ and $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1,\ldots\ldots.q$. Again, the condition $\alpha_0 > 0$ and $\alpha_i \geq 0$ are needed to guarantee that the conditional variance $h_t > 0$. To carry out the process of parameter estimation consider the simplest model which is the GARCH (0,1) model, where h_t is given by

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

If Ω is an open interval and if L(θ) is differentiable and assumes a maximum on θ, then MLE will be a solution of the

$$\text{Equation } \frac{\partial L(\theta)}{\partial \theta} = 0$$

GARCH (1,1) : The most widely used GARCH (p, q) model for GARCH (1,1) takes the form of $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, where α_0 is constant term; $\alpha_1 \epsilon_{t-1}^2$ is ARCH term reflects the volatility from the previous period, measured as the lag of the squared residual from the mean equation and β_1 h_{t-1} is the GARCH term, it is the last periods forecast variance.

$$Y_t(1 - \sum_{s=1}^{P} \alpha_s L^s) = \mu + \sum_{i=1}^{q} \beta_s L^s x_t + (1 + \sum_{s=1}^{P} \gamma_s L^s) e_t$$

Where, L is the usual lag operator(L^s $y_t = y_{t-s}$ 'L^s $x_t = x_{t-s}$, etc),μ∈R, $\alpha_s \in R, \beta_s \in R^k$ and $\gamma_s \in R$ are parameters, et's errors and p, q and r are natural numbers specified in advance. R^2 max/min Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE) determine the best-fitted ARIMA, GARCH, and ARIMAx models. All three groups choose the model that best meets the criteria. The Ljung-Box-test, ACF, and PACF residual plots are used to diagnose best-fit models again. White-noise models have a stringent cutoff. Predictions were produced through 2020 using the best fitted ARIMA, GARCH, and ARIMAx AIC =2k-2 ln(L)

BIC = –2*ln(L) +k*ln(n)

$$ME=\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{X}_i\right)$$

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \hat{X}_i}{X_i}\right) * 100$$

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

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

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

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

where, $X, \bar{X}, \hat{X}$ are the values of the *i* th observation, mean and estimated value of the *i*th observation of the variable X and k is the number of parameters in the statistical model and *L* is the maximized value of the likelihood function for the estimated model

Result and Discussion

Trends in area, production and productivity of lentil in major states of India.

Knowing the overall performance of the each series, path of movement was traced through different parametric trend models. Among the significant competitive models, the best model was selected based on maximum value of R^2 and minimum value of RMSE. The following section presents the result of this exercise in table 2.

Area under lentil cultivation in India has followed significant linear trend while production and productivity follows the quadratic trend thereby indicating the two points of infections. The production scenario of lentil in all the states are followed non-linear trends, mostly polynomial. Area, production and productivity of lentil in Uttar Pradesh; production and productivity in case of Bihar and productivity in case of Madhya Pradesh have followed significant quadratic trend while area in case of Bihar; area and production in case of Madhya Pradesh have followed cubic trend. Positive coefficient of linearity in area under lentil for India clearly indicates that area has increased continuously throughout the study period, but production and productivity has increased continuously up to 2005 and shows slight decreasing trend thereafter (Fig. 2 & 3) same has been realized through nature of b_1 and b_2 coefficients of quadratic model. This has happened mainly because of decreasing trend of area, production and productivity of major lentil growing states of nation i.e., Uttar Pradesh, Bihar and Madhya Pradesh during recent years under study as evident by the nature of respective coefficients. This clearly a major concern towards food and nutritional security of the Indian people. One must think for resisting these tendencies so as to keep lentil production at steady state. Area and production of lentil in West Bengal has followed quadratic trend while productivity has increased linearly. West Bengal had reported maximum area and production of lentil in middle of seventies (Fig. 1&2), started declining thereafter and a slight improvement has been noticed in recent years under study, same has been revealed by the nature of b_1 and b_2 coefficient of the area and production. The area and production of lentil in case of Rajasthan has increased exponentially during the study period. But the productivity have undergone three point of fluctuations, negative value for cubic time coefficients indicates that productivity is declining in recent past hence a necessary steps needs to be taken to arrest this decreasing trend in productivity.

Table 2: Trends in area, production and productivity of lentil in major states of India.

Parameter	*Model*	R^2	*RMSE*	*Constant*	b_1	b_2	b_3
			Uttar Pradesh				
Area	**Quadratic**	0.962	27.513	106.426	26.001	-0.354	
Production	**Quadratic**	0.938	33.394	14.882	21.735	-0.263	
Productivity	**Quadratic**	0.783	53.071	448.747	16.265	-0.189	
			Bihar				
Area	**Cubic**	0.710	5.870	134.950	5.046	-0.212	0.003
Production	**Quadratic**	0.788	13.313	59.078	5.650	-0.088	
Productivity	**Quadratic**	0.807	59.249	422.400	26.720	-0.419	
			Madhya Pradesh				
Area	**Cubic**	0.917	28.378	353.765	-17.582	1.277	-0.018
Production	**Cubic**	0.944	13.295	143.999	-10.802	0.870	-0.014
Productivity	**Quadratic**	0.597	30.609	340.612	7.744	-0.116	
			West Bengal				
Area	**Quadratic**	0.793	11.820	139.273	-5.140	0.081	
Production	**Quadratic**	0.215	9.125	60.621	-1.450	0.028	
Productivity	**Linear**	0.726	67.407	424.919	9.505		
			Rajasthan				
Area	**Exponential**	0.319	6.669	12.525	0.017		
Production	**Exponential**	0.502	8.413	7.748	0.031		
Productivity	**Cubic**	0.739	87.307	799.143	-48.281	3.881	-0.067
			India				
Area	**Linear**	0.950	51.084	772.024	19.312		
Production	**Quadratic**	0.941	54.217	257.221	29.059	-0.257	
Productivity	**Quadratic**	0.844	35.229	382.070	17.665	-0.272	

Thus from the test of outliers, randomness test and trend analysis of area, production and productivity of lentil in major states of India the following important features has emerged out:

1. Analysis of data for different series rejected the presence of outliers in production and productive series for all the states under study, while area under lentil in Bihar and Rajasthan show the presence of outliers.
2. Productivity of the lentil in all the states under study has changed randomly.
3. Except for productivity series of West Bengal and area of whole India, all other data series exhibit non-linear trend.
4. All the states under study have shown decreasing trend for area under lentil during recent period under study except for Bihar and West Bengal.
5. Excepting West Bengal, all other states are showing decreasing trend in lentil productivity during recent periods under study.

Modeling and forecasting of area under lentil

Area, output, and productivity stationarity tests for lentils throughout Indian states are summarized in Table 3. The area under the lentil data failed both the KPSS and the ADF tests, indicating that the stationary data hypothesis can be rejected. To reach stationary values, all investigated series needed to be differentiated at the first-order. After ensuring that each series is stationary, multiple ARIMA models are tested before the best one is chosen based on error and R2 values. Table 4 displays the best GARCH model for each

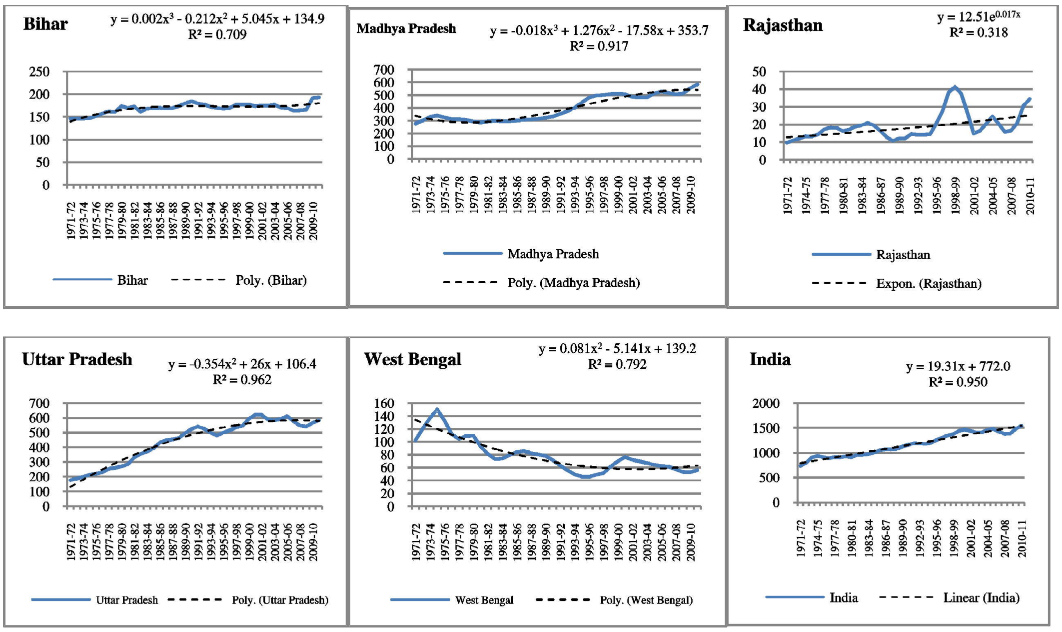

Figure 1: Observed and expected trends of area under lentil in major states of India

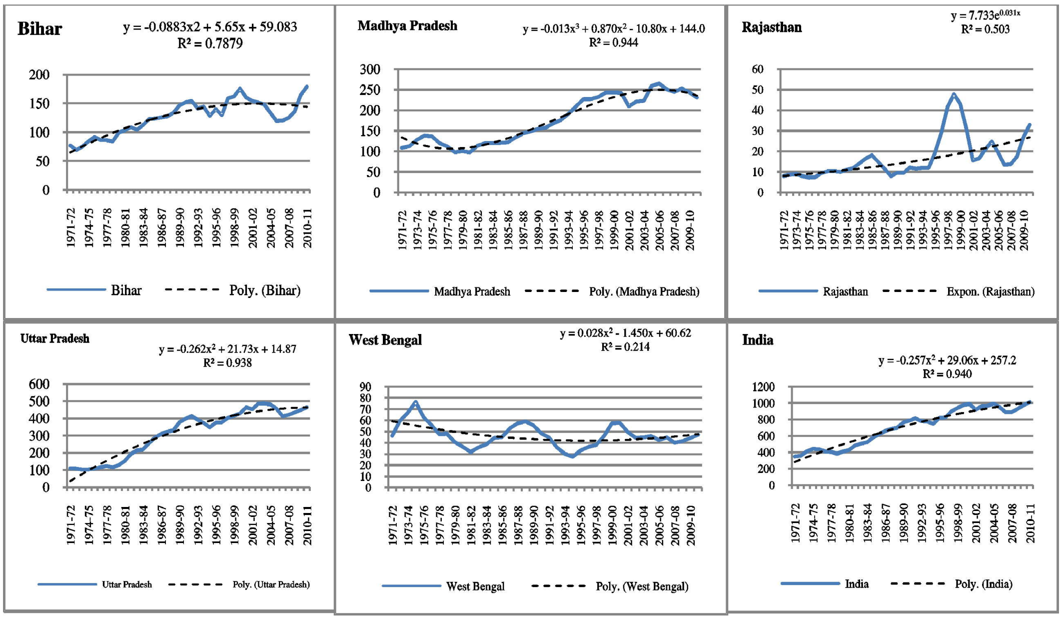

Figure 2: Observed and expected trends of lentil production in major states of India

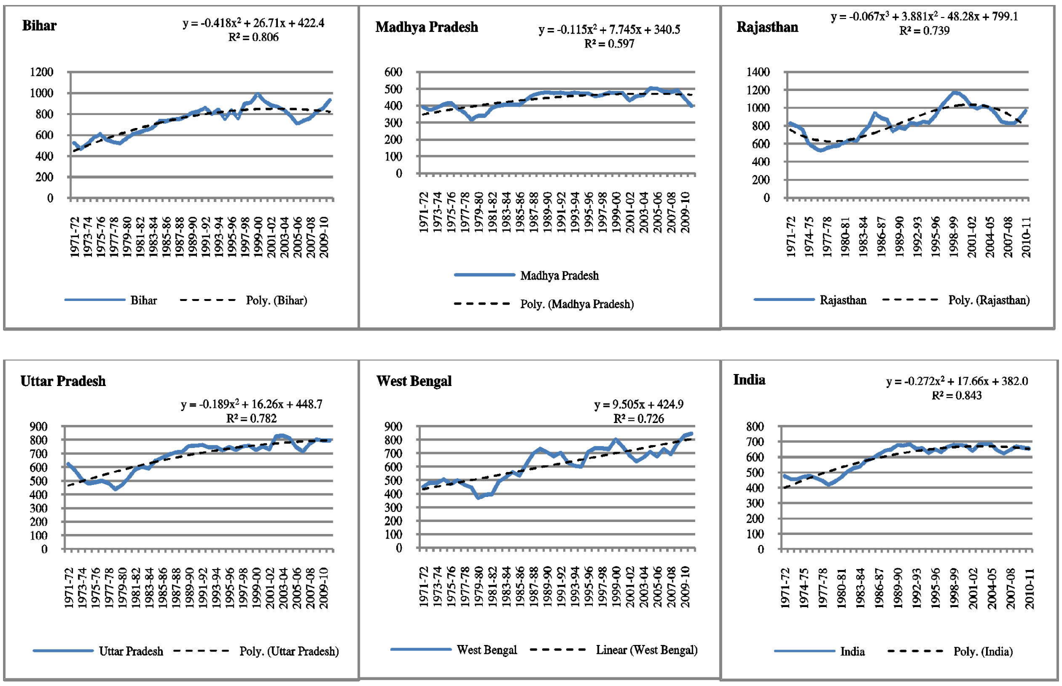

Figure 3: Observed and expected trends of lentil productivity in major states of India

dataset after numerous models were fitted. Models are validated using the Ljung-Box residual test.According to Table 4, the best ARIMA model for lentil production in West Bengal and Rajasthan is ARIMA(0,1,2), while the best models for Uttar Pradesh, Bihar, Madhya Pradesh, and India are ARIMA(2,1,2), ARIMA(1,1,2), and ARIMA(4,1,3). However, GARCH (1) is the best fit for the lentil area in every Indian state except West Bengal, where GARCH(2) is the best fit. Strong autocorrelation in the residuals for the best fitted model in ARIMA and GARCH is rejected by Ljung-Box residual tests.

ARIMA is superior to GARCH for predicting lentil production in Bihar and West Bengal. Table 4 shows that GARCH outperformed ARIMA in every other country except for Japan. Consequently, the residuals of selected models do not exhibit any correlation, as demonstrated by the ACF and PACF residual graphs (Fig. 4). The lentil harvest in 2020 is forecasted using these models. (Fig. 5). The model's predicted and observed values coincide, as depicted by the graph. All states except Rajasthan, where GARCH(1) models failed to capture the unexpected gain in area during 2011, had their chosen models validated using data from the prior three years (Table 5). Area underlentil is expected to increase steadily in Uttar Pradesh, Bihar, and India, reaching 688.08 thousand hectares, 230.40 thousand hectares, and 1670.16 thousand hectares respectively by 2020. The remaining major states for lentil production are projected to see a decline in lentil acreage (Fig. 5).

Table 3: Test of stationarity of area, production and productivity of lentil in India.

State	*ADF Value*	*P Value*	*Conclusion*	*KPSS Value*	*P Value*	*Conclusion*
			Area			
Uttar Pradesh	0.051	0.990	Non Stationary	1.847	0.010	Non Stationary
Bihar	-1.592	0.733	Non Stationary	0.934	0.010	Non Stationary
Madhya Pradesh	-2.110	0.530	Non Stationary	1.725	0.010	Non Stationary
West Bengal	-1.921	0.604	Non Stationary	1.308	0.010	Non Stationary
Rajasthan	-1.988	0.578	Non Stationary	0.672	0.016	Non Stationary
India	-1.781	0.659	Non Stationary	1.913	0.010	Non Stationary
			Production			
Uttar Pradesh	-1.187	0.892	Non Stationary	1.875	0.010	Non Stationary
Bihar	-1.926	0.602	Non Stationary	1.475	0.010	Non Stationary
Madhya Pradesh	-2.508	0.374	Non Stationary	1.867	0.010	Non Stationary
West Bengal	-3.781	0.042	Stationary	0.328	0.100	Stationary
Rajasthan	-1.893	0.615	Non Stationary	0.812	0.010	Non Stationary
India	-1.581	0.738	Non Stationary	1.950	0.010	Non Stationary
			Yield			
Uttar Pradesh	-2.878	0.230	Non Stationary	1.681	0.010	Non Stationary
Bihar	-1.346	0.829	Non Stationary	1.519	0.010	Non Stationary
Madhya Pradesh	-1.110	0.908	Non Stationary	1.414	0.010	Non Stationary
West Bengal	-2.207	0.492	Non Stationary	1.580	0.010	Non Stationary
Rajasthan	-2.870	0.233	Non Stationary	1.391	0.010	Non Stationary
India	-0.887	0.942	Non Stationary	1.612	0.010	NonStationary

Table 4: Best fitted ARIMA and GARCH models for area under lentil in India

State	*Model*	*Model Selection Criteria*								*Ljung-Box test for residuals*	
		AIC	*BIC*	*ME*	*RMSE*	*MAE*	*MPE*	*MAPE*	*R2*	χ^2	*P Value*
Uttar Pradesh	ARIMA(2,1,2)	379.910	389.740	-0.160	32.258	24.517	0.048	5.820	0.956	7.010	0.725
	GARCH(1)*	303.053	311.107	-0.413	12.820	10.490	2.333	0.216	0.991	1.259	0.262
Bihar	ARIMA(1,1,2)*	206.330	214.390	-0.028	3.446	2.473	-0.009	1.453	0.894	2.969	0.982
	GARCH(1)	208.691	216.745	-0.001	3.681	2.596	1.528	-0.004	0.825	0.304	0.581
Madhya Pradesh	ARIMA(4,1,3)	360.760	375.500	-1.656	21.935	16.687	-0.863	4.296	0.948	2.086	0.996
	GARCH(1)*	317.469	325.657	2.016	14.731	11.413	2.978	0.351	0.975	2.183	0.140
West Bengal	ARIMA(0,1,2)*	229.640	236.090	-0.131	4.763	3.084	-0.037	3.677	0.968	6.092	0.712
	GARCH(2)	314.742	324.723	1.850	14.269	9.278	11.048	-1.388	0.735	3.454	0.616
Rajasthan	ARIMA(0,1,2)	192.460	199.120	-0.006	2.559	1.887	0.382	9.207	0.911	0.395	0.530
	GARCH(1)*	166.007	174.062	-0.109	2.445	1.789	8.799	-1.566	0.839	3.403	0.101
India	ARIMA(1,1,3)	427.590	437.410	0.577	61.854	50.119	-0.259	4.752	0.929	7.355	0.692
	GARCH(1)*	323.298	331.353	0.994	17.179	12.632	1.108	0.072	0.994	0.966	0.326

Note: * indicates the best model and used further for forecasting purpose.

Table 5: Validation and forecasting of area (000' hectare) under lentil in India on the basis of selected best model

States	*Model*	*2010*		*2011*		*2012*		*2016*	*2018*	*2020*
		Observed	*Predicted*	*Observed*	*Predicted*	*Observed*	*Predicted*	*Predicted*	*Predicted*	*Predicted*
Uttar Pradesh	GARCH(1)	592.41	575.74	586.00	586.97	573.00	598.21	643.15	665.61	688.08
Bihar	ARIMA(1,1,2)	171.07	169.90	238.55	205.04	168.45	221.09	223.68	226.97	230.46
Madhya Pradesh	GARCH(1)	541.30	517.13	590.50	520.83	620.50	524.57	540.08	548.12	556.34
West Bengal	ARIMA(0,1,2)	51.70	49.78	57.45	50.19	59.30	49.14	44.91	42.80	40.68
Rajasthan	GARCH(1)	27.05	24.82	44.07	25.52	31.90	26.19	28.68	29.83	30.93
India	GARCH(1)	1479.81	1467.51	1597.44	1488.94	1562.36	1510.02	1591.63	1631.20	1670.16

Modeling and forecasting of lentil production

Stationarity tests for lentil production show that all data series are nonstationary except West Bengal (Table 3), but when plotted in ACF and PACF graphs, lentil production in West Bengal also shows a nonstationary pattern. First-order differencing makes all states data series stationary. After stationarity, multiple ARIMA models are attempted for each series. The significant model that meets the maximum criteria of minimal value AIC, BIC, ME, RMSE, MAE, MPE, MAPE, and maximum R2 is selected as the best ARIMA model and displayed in table 6. According to the table, the optimum ARIMA model for lentil production is ARIMA(1,1,2) for Bihar and West Bengal, ARIMA(2,1,3), ARIMA(2,1,4), ARIMA(0,1,2), and ARIMA(1,1,3) for Uttar Pradesh, Madhya Pradesh, Rajasthan, and India. Using the same criterion, the best GARCH model for lentil production in all states in India is GARCH(1). As with other crops, ARIMAx first models and forecasts up to 2020

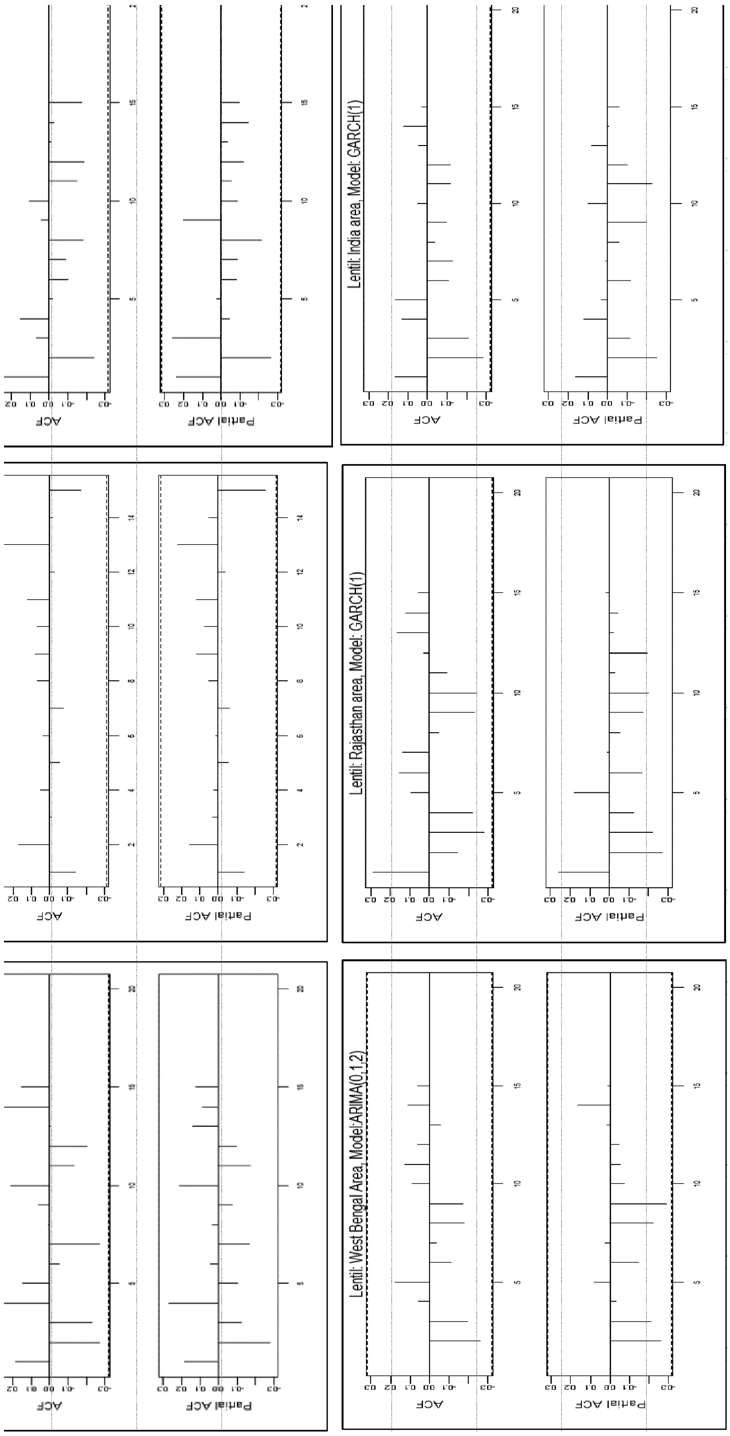

Figure 4: ACF and PACF graphs of residuals for the best fitted models of area under lentil in India

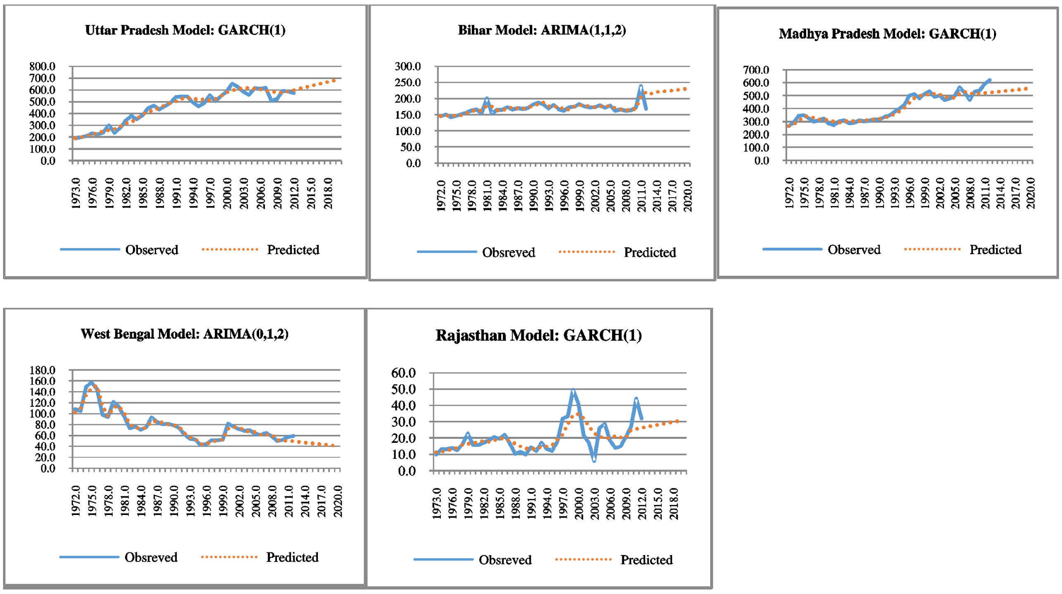

Figure 5: Observed and forecasted area ('000 ha) under lentil cultivation using best selected model in India

using ARIMA algorithm all independent factors that substantially affect lentil productivity. These anticipated values become ARIMAx model independent variables. Best ARIMAx model was chosen similarly. From table 6, ARIMAx(1,1,2) for Bihar, West Bengal, and India; ARIMAx(2,1,3), ARIMAx(3,1,4), and ARIMAx(3,1,4) for Uttar Pradesh, Madhya Pradesh, and Rajasthan are the best ARIMAx models for lentil production. Ljung–Box residuals test rejects strong auto correlation in the residuals of the best fitted ARIMA, GARCH, and ARIMAx model (table 6).

The best ARIMA, GARCH, and ARIMAx models were chosen using error and R2 criteria. From table 6, ARIMAx model meets highest criterion for lentil production in Bihar and Rajasthan, but ARIMA model outperforms GARCH and ARIMAx in all other states in India. ACF and PACF residual graphs (Fig. 6) show that the residuals of selected models are free from significant correlations and may be used to anticipate lentil output up to 2020. (Fig. 7). Figure 7 shows that throughout model construction, observed and projected values are near in all stages. The selected models are evaluated for accuracy using last three years data and found that the actual and predicted values are in range (Table 7) for all states and India. From the anticipated numbers, lentil output in Bihar and India will rise constantly in 2020 compared to 2012, while other states would increase lentil production moderately. 2020 lentil output would rise to 1223.29 thousand tonnes from 1058.67 in 2012.

Table 6: Best fitted ARIMA, GARCH and ARIMAx models for production of lentil in India

State	*Model*	*Model Selection Criteria*								*Ljung-Box test for residuals*	
		AIC	*BIC*	*ME*	*RMSE*	*MAE*	*MPE*	*MAPE*	R^2	χ^2	*P Value*
Uttar Pradesh	ARIMA(2,1,3)*	321.320	332.790	0.126	14.128	11.394	-0.003	4.009	0.989	6.947	0.730
	GARCH(1)	342.607	350.795	-0.980	19.530	15.264	5.702	-1.009	0.979	0.190	0.663
	ARIMAx(2,1,3)	322.330	335.430	0.152	13.956	11.285	0.017	3.983	0.990	3.205	0.976
Bihar	ARIMA(1,1,2)	272.670	280.860	0.147	7.677	5.909	-0.002	4.720	0.927	3.002	0.981
	GARCH(1)	286.510	294.698	0.084	9.293	7.559	6.166	-0.356	0.887	0.197	0.658
	ARIMAx(1,1,2)*	279.120	290.590	0.117	7.189	5.819	-0.027	4.690	0.936	7.735	0.655
Madhya Pradesh	ARIMA(2,1,4)*	246.160	259.050	-0.008	5.053	3.625	-0.009	2.045	0.992	2.664	0.988
	GARCH(1)	372.679	380.997	2.102	26.033	20.784	12.632	-1.202	0.804	2.793	0.095
	ARIMAx(3,1,4)	282.100	298.480	0.117	7.003	5.124	0.060	3.076	0.985	3.580	0.964
West Bengal	ARIMA(1,1,2)*	218.680	226.740	-0.037	4.003	3.101	-0.109	6.686	0.871	7.693	0.659
	GARCH(1)	248.387	256.575	0.091	5.918	4.700	10.116	-1.615	0.685	0.879	0.349
	ARIMAx(1,1,2)	227.390	237.210	-0.038	4.062	3.115	-0.097	6.749	0.866	8.469	0.583
Rajasthan	ARIMA(0,1,2)	193.910	200.460	0.004	2.711	1.871	0.140	10.720	0.935	8.362	0.594
	GARCH(1)	200.150	208.338	0.618	4.651	2.896	16.070	-2.444	0.760	0.977	0.393
	ARIMAx(2,1,2)*	190.060	202.940	-0.003	2.437	1.711	0.221	10.573	0.946	8.409	0.589
India	ARIMA(1,1,3)*	360.290	370.110	0.628	23.875	18.228	-0.049	2.665	0.988	2.478	0.991
	GARCH(1)	383.185	391.373	0.395	32.836	26.191	3.906	-0.173	0.977	1.854	0.173
	ARIMAx(1,1,2)	361.930	374.820	-0.185	23.641	18.465	0.010	2.670	0.988	0.736	0.865

Note: * indicates the best model and used further for forecasting purpose.

Table 7: Validation and forecasting of lentil production (000' tones) in India on the basis of selected best model

State	*Model*	*2010*		*2011*		*2012*		*2016*	*2018*	*2020*
		Observed	*Predicted*	*Observed*	*Predicted*	*Observed*	*Predicted*	*Predicted*	*Predicted*	*Predicted*
Uttar Pradesh	ARIMA (2,1,3)	475.94	464.94	411.00	455.16	505.00	469.38	488.69	505.72	522.75
Bihar	ARIMAx (1,1,2)	150.51	145.19	214.69	199.74	171.61	186.41	198.45	204.87	213.51
Madhya Pradesh	ARIMA (2,1,4)	284.40	236.59	177.90	227.14	230.00	220.27	227.82	238.77	243.14
West Bengal	ARIMA (1,1,2)	47.10	40.18	53.37	41.57	41.19	40.53	41.36	41.78	42.21
Rajasthan	ARIMAx (2,1,2)	24.20	24.06	38.45	29.41	35.93	34.52	38.73	42.52	44.88
India	ARIMA (1,1,3)	1031.62	991.79	943.81	1028.08	1058.67	1045.65	1148.17	1186.53	1223.29

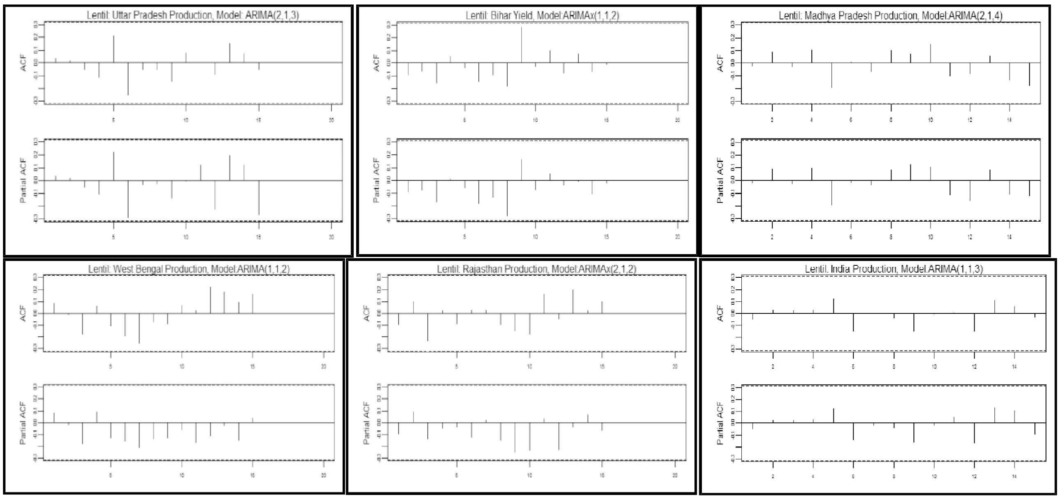

Figure 6: ACF and PACF graphs of residuals for the best fitted models of lentil production in India

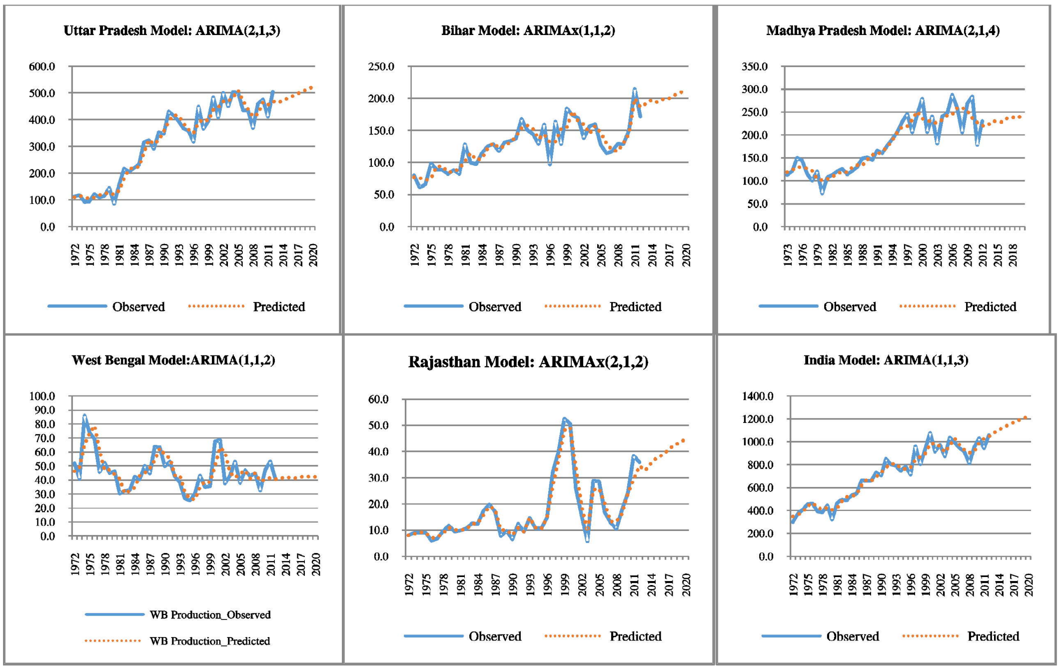

Figure 7: Observed and forecasted lentil production ('000 tonnes) using best selected model in India

Modeling and forecasting of lentil productivity

The ADF and KPSS stationarity tests for lentil production show that all states' lentil productivity is nonstationary (Table 3). First-order differencing stabilised it. After stationary, we proceeded similarly to production and chose the best ARIMA, GARCH, and ARIMAx models for all states under consideration, as shown in table 8. From table 8, ARIMA(1,1,4) is best for lentil productivity in Bihar and Madhya Pradesh, whereas ARIMA(1,1,0), ARIMA(1,1,2), ARIMA(2,1,2), and ARIMA(2,1,4) are best for Uttar Pradesh, West Bengal, and Rajasthan. GARCH fits lentil productivity well in all states in India (1). For all states' lentil productivity series, the best ARIMAx model is also chosen among competitive models. According to table 8, ARIMAx(1,1,2) for Bihar and Rajasthan, ARIMAx(2,1,4) for Uttar Pradesh and India, ARIMAx(0,1,2) for Madhya Pradesh, and ARIMAx(3,1,2) for West Bengal are the best-fitted ARIMAx models for lentil productivity in their states. The residuals of the best selected ARIMA, GARCH, and ARIMAx models were tested using Ljung–Box (Table 8), and all showed no significant auto correlation.

Table 8: Best fitted ARIMA, GARCH and ARIMAx models for productivity of lentil in India

State	*Model*	*Model Selection Criteria*								*Ljung-Box test for residuals*	
		AIC	*BIC*	*ME*	*RMSE*	*MAE*	*MPE*	*MAPE*	R^2	χ^2	*P Value*
Uttar Pradesh	ARIMA(1,1,0)*	313.370	318.200	0.298	17.017	13.693	0.017	2.129	0.976	3.767	0.957
	GARCH(1)	386.194	394.382	-6.733	34.591	27.020	4.353	-1.494	0.907	2.060	0.151
	ARIMAx(2,1,4)	358.960	373.700	0.277	21.003	16.937	0.012	2.600	0.966	6.596	0.763
Bihar	ARIMA(1,1,4)*	342.360	353.640	0.288	21.100	15.823	0.045	2.113	0.972	5.929	0.821
	GARCH(1)	408.114	416.302	-0.965	48.290	38.218	5.335	-0.710	0.868	0.039	0.844
	ARIMAx(1,1,2)	394.120	405.580	0.626	32.555	27.351	0.041	3.816	0.942	7.049	0.721
Madhya Pradesh	ARIMA(1,1,4)*	285.260	296.340	-0.126	11.243	8.009	-0.019	1.907	0.943	1.881	1.000
	GARCH(1)	338.704	346.892	-1.225	18.485	13.063	3.201	-0.554	0.858	1.424	0.233
	ARIMAx(0,1,2)	340.620	348.940	0.345	17.098	12.976	0.084	3.065	0.890	6.312	0.788
West Bengal	ARIMA(1,1,2)	388.780	396.960	0.254	36.877	31.130	-0.108	5.085	0.916	9.122	0.521
	GARCH(1)	402.430	415.740	0.427	35.688	28.772	-0.144	4.694	0.925	1.871	0.171
	ARIMAx(3,1,2) *	370.293	378.347	-2.178	31.503	26.694	4.509	-0.734	0.927	6.362	0.784
Rajasthan	ARIMA(2,1,2)*	366.570	376.240	0.157	30.592	25.339	0.003	3.059	0.966	9.775	0.460
	GARCH(1)	395.276	403.331	-7.394	44.166	37.472	4.638	-1.177	0.929	0.176	0.675
	ARIMAx(1,1,2)	386.190	397.470	-0.727	38.718	29.769	-0.140	3.745	0.952	9.901	0.449
India	ARIMA(2,1,4)	332.060	345.160	0.565	16.045	13.638	0.109	2.328	0.970	0.036	0.850
	GARCH(1)	349.862	358.050	-0.133	21.188	17.159	2.983	-0.175	0.945	1.118	0.290
	ARIMAx(2,1,4)*	334.390	352.110	-0.044	14.239	11.550	0.011	1.970	0.976	2.809	0.833

Note: * indicates the best model and used further for forecasting purpose.

The best ARIMA, GARCH, and ARIMAx models have least AIC, BIC, ME, RMSE, MAE, MPE, MAPE, and greatest R2 values. ARIMAx outperforms ARIMA and GARCH for lentil productivity in West Bengal and India. ARIMA outperforms GARCH and ARIMAx in Uttar Pradesh, Bihar, Madhya Pradesh, and Rajasthan. These selected models are again diagnostically checked using ACF and PACF residual graphs (Fig 8) and confirmed to be independent and used to estimate lentil productivity up to 2020. (Fig. 9). The chart shows that during model construction in all stages, real and projected values are quite near.The top models are evaluated using current three-year data (Table 9) and found to be close to real values for Bihar and India. However, models have missed unexpected lentil productivity shifts in all remaining states over one or more validation years. The 2020 lentil productivity projection for West Bengal is 982.68 kg per hectare, up from 694.62 kg in 2012. All states except Rajasthan will boost their output somewhat (Figure 9 and Table 9). Rajasthan may lose its lentil production potential in the future. India's lentil production is expected to rise to 721.90 kg/ha in 2020, however this is far below New Zealand's 2666.70 kg/ha and China's 2238.80 kg/ha.

Table 9: Validation and forecasting of lentil productivity (kg per hectare) in India on the basis of selected best model

State	*Model*	*2010*		*2011*		*2012*		*2016*	*2018*	*2020*
		Observed	*Predicted*	*Observed*	*Predicted*	*Observed*	*Predicted*	*Predicted*	*Predicted*	*Predicted*
Uttar Pradesh	ARIMA(1,1,0)	803.40	766.33	701.37	817.71	881.33	829.43	859.76	873.26	886.69
Bihar	ARIMA(1,1,4)	879.82	847.88	899.98	908.74	1018.76	957.60	1023.92	1044.20	1065.13
Madhya Pradesh	ARIMA(1,1,4)	525.40	450.37	301.27	428.59	370.67	421.25	422.38	423.39	424.61
West Bengal	ARIMAx(3,1,2)	911.06	793.65	929.01	897.27	694.62	808.65	858.46	959.05	982.68
Rajasthan	ARIMA(2,1,2)	894.66	872.97	872.45	927.98	1126.33	960.73	966.14	977.93	992.20
India	ARIMAx(2,1,4)	697.13	664.19	590.82	612.91	677.61	645.64	665.82	695.47	721.90

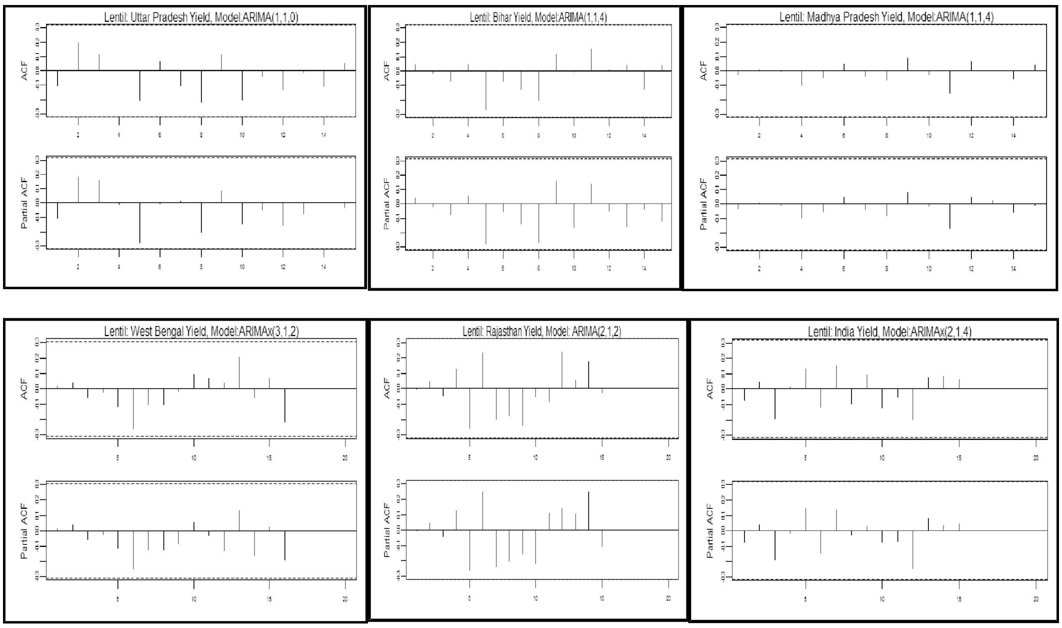

Figure 8: ACF and PACF graphs of residuals for the best fitted models of lentil productivity in India

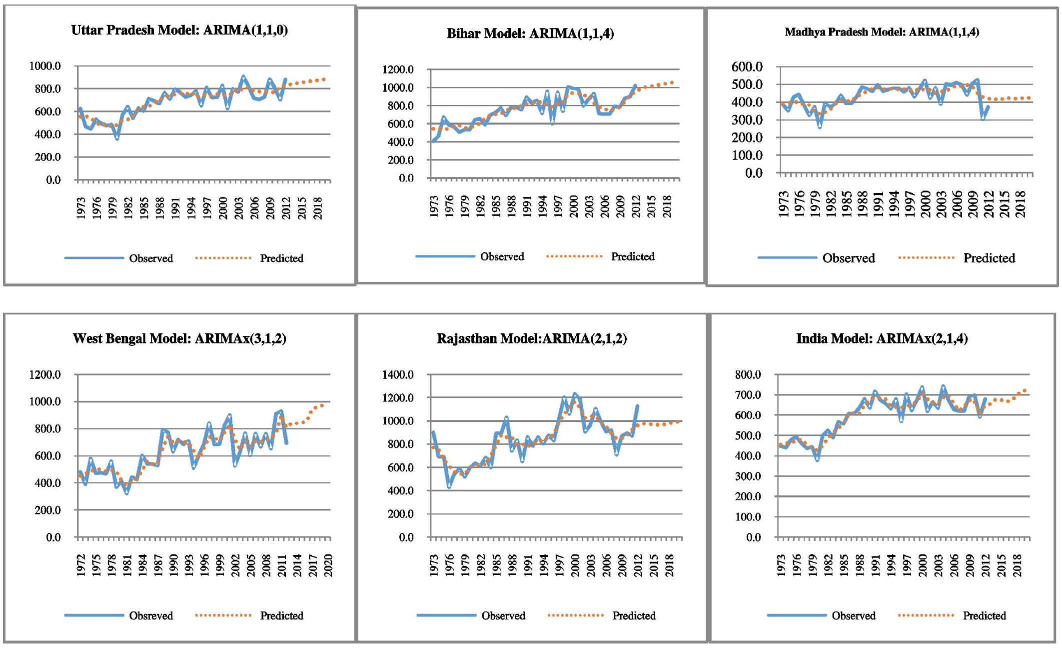

Figure 9: Observed and forecasted lentil productivity (kg per hectare) using best selected model in India

Thus, from the study of modeling and forecasting of area, production and productivity of lentil in major growing states and whole India following findings emerged out:

1. The data series of area, production and productivity of lentil for selected states and India are stationary and hence first order differencing is done to achieve stationarity.
2. For area under lentil in Bihar and West Bengal AIMRA model is best fitted model where as in all the other states including whole India GARCH model outperformed the ARIMA models.
3. In maximum cases of lentil production and productivity series under study univariate ARIMA outperformed ARIMAx and GARCH.
4. From the forecasted values it can be said that area under lentil in Uttar Pradesh, Bihar and whole India would increase whereas in other states it would decrease in future.
5. In case of production series of lentil, it can be noted that production would increase in 2020 as compared to 2012 in all the states under study in future.
6. Lentil productivity of India would increase marginally to 721.90 in 2020, but this increase in productivity is no match with present productivity of New Zealand (2666.70 kg/ha) and China (2238.80 Kg/ha) and so on.

Conclusion

The study reveals that there has been considerable expansion in area, production and productivity of lentil in all the states under study including whole India Except for Madhya Pradesh, in all other major states average productivity of lentil is above the national average productivity of 592.52 kg/ha. Hence there is need for increasing productivity in minor states to augment the production of lentil.Productivity of lentil in all the states under study has changed randomly,may be due to lack of stable varieties. Major concern in the production scenario of lentil production is that the major contributing states like Uttar Pradesh and Bihar including whole India showing decreasing trend in lentil production during recent period under study.

The data series of area, production and productivity of lentil for selected states and India are non-stationaryand hence first order differencing is done to achieve stationarity. In maximum cases GARCH model found to be best for modeling and forecasting area under lentil, whereas univariate ARIMA model outperformed incase of production and productivity. From the forecasted values it can be said that area under lentil in Uttar Pradesh, Bihar and whole India would increase whereas in other states it would decrease in future. Lentil productivity of India would increase marginally to 721.90 in 2020, but this increase in productivity is no match with present productivity of New Zealand (2666.70 kg/ha) and China (2238.80 Kg/ha) and so on.

Reference

FAO. (2018) Food Outlook. Rome: FAO.

Mishra. P., C. Sarkar, K. P. Vishwajith, B. S. Dhekale and P. K. Sahu (2013). Instability and forecasting using ARIMA model in area, production and productivity of onion in India. *J Crop Weed*., **9**, 96-101.

Mishra, P., Yonar, A. Yonar, H. Kumari, B., Abotaleb, M., Das, S.S., & Patil, S.G. (2021). State of the art in total pulse production in major state of India using ARIMA technique. *Current Research in food science*.800-806 *https://doi.org/10.1016/j.crfs.2021.10.009*

Paul, R.K, Prajneshu and Himadri Ghosh (2009). GARCH Nonlinear Time series analysis for Modeling and Forecasting of India's Volatile Spices Export Data. *Journal of the Indian Society of Agricultural Statistics,* **63(2)**, 123-131

Rahman, N.M.F., M.A. Aziz, M.M. Rahman and N. Mohammad (2013). Modeling on Grass Pea and Mung Bean Pulse production in Bangladesh using ARIMA model. *IOSR Journal of Agriculture and Veterinary Science,* **6(1)**, 20-31.

Sahu, P.K. (2006). Forecasting yield behaviour of potato, mustard, rice and wheat under irrigation. *J. Veg. Sci.*, **12**, 81-99.

Srivastava, S.K., Sivaramne, N. & Mathur, V.C. (2010) Diagnosis of pulses performance of India. *Agricultural Economics Research Review*, 23:137–148.

Suresh, A. & Reddy, A.A. (2016). Total factor productivity of major pulse crops in India: implications for technology policy and nutritional security. *Agricultural Economics Research Review*, 29:87–98

Vishwajith, K.P., B.S. Dhekale, P.K. Sahu, P. Mishra and Md. Noman (2014). Time series modelling and forecasting of pulses production in India. *J. Crop Weed*., **10**, 147-154.

Vishwajith, K.P., P.K. Sahu, B.S. Dhekale, Md. Noman and P Mishra (2016). Comparison of ARIMA, ARIMAx and GARCH model for forecasting the gram production scenario of India. *RASHI*, **1(1)**, 24-38

Yaziz, S.R., M.H. Ahmad, L.C. Nian & N. Muhammad (2011). A comparative study on Box-Jenkins and Garch models in forecasting crude oil prices. *Journal of Applied Sciences,* **11(7)**, 1129-1135.