Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2020; 9(1): 1193-1196
Received: 19-11-2019
Accepted: 21-12-2019

Mamata
Department of Agricultural Statistics, Visva Bharati, Sriniketan, West Bengal, India

Alisha Mittal
Department of Mathematics and Statistics, CCSHAU, Hisar, Haryana, India

Deepa Bharti,
Department of Mathematics and Statistics, CCSHAU, Hisar, Haryana, India

Corresponding Author:
Alisha Mittal
Department of Mathematics and Statistics, CCSHAU, Hisar, Haryana, India

Forecasting of Barly production in India using ARIMA Model

Mamata, Alisha Mittal and Deepa Bharti

Abstract
Forecasting is an important tool to estimate the area, production and productivity of any crop in near future. The method chosen depends on the purpose and importance of the forecasts as well as the cost and efficiency of the alternative forecasting methods. Keeping in view the importance of the subject matter, a study on yield trends of barley crop in India has been undertaken to see the forecasting performance of the developed ARIMA models for barley yield prediction. ARIMA models are built for the data related to barley yields in India. The crop yield data of the past three/four decades have been used for the model building and the forecast values are obtained for the year 2013-14, 2014-15, 2015-16, 2016-17 and 2017-18. After experimenting with different lags of the moving average and autoregressive processes; ARIMA (1,1,0) have been fitted for barley yield forecasting purpose in India. The overall results indicates that the percent relative deviations of the forecast yields from the observed yields are within acceptable limits and favours the use of ARIMA models to get short-term forecast estimates.

Keywords: Autocorrelation, partial autocorrelation, stationarity, ljung-box, univariate

Introduction

Barley (*Hordeum vulgare* L.) is one of the most important cereal crops in the world after rice, wheat and maize. Barley is a rabi cereal crop from the grass family *Poaceae*. Global production stands around 160 million tons. In the world, Europe is the most leading continent growing Barley followed by Asia. In India, Uttar Pradesh, Rajasthan, Madhya Pradesh, Haryana, Punjab and Himachal Pradesh are the major producers of barley crop. In India, barley production fluctuated substantially in recent years, it tended to decrease through 1969 - 2018 period ending at 1.781 MMT in 2018. In 2019-20 barley production is forecast at a record 1.95 MMT on reported higher planting. Traditionally, India produced six-row varieties of barley, which are mostly for food and feed use and unsuitable for malting.

Forecasts have been made using parametric univariate time series models, known as Autoregressive Integrated Moving Average (ARIMA) model popularized by *Box and Jenkins* (1976) [4]. These approaches have been employed extensively for forecasting economics time series, inventory and sales modeling (*Brown*, 1959) [5]. *Ljung and Box* (1978) [7] and *Pindyck and Rubinfeld* (1981) [8] have also discussed the use of univariate time series in forecasting. *Rachana et al.* (2010) [9] used ARIMA models to forecast pigeon pea production in India. *Badmus and Ariyo* (2011) [1] forecasted the area of cultivation and production of maize in Nigeria using ARIMA model. They estimated ARIMA (1, 1, 1) and ARIMA (2, 1, 2) for cultivation area and production respectively. *Falak and Eatzaz* (2008) [6] analyzed future prospects of wheat production in Pakistan. They obtained the parameters of their forecasting model using Cobb-Douglas production function for wheat, while future values of various inputs are obtained as dynamic forecasts on the basis of separate ARIMA estimates for each input and for each Province. The ARIMA technique have been used extensively by a number of researchers to forecast demands in terms of internal consumption, imports and exports to adopt appropriate solutions, *Sohail et al.*, (1994) [10]. Balanagammal *et al.* (2000) [2] and Boken (2000) [3] have used time series analysis for crop yield forecasting. Verma *et al.* (2009) have worked on ARIMA yield modeling for different crops in Germany.

Methodology

The barley production data of India for the period 1980-81 to 2017-18 collected from India stat website. This data used to develop the forecasting model.

The ARMA models are generalization of the simple AR model that uses three tools for modeling series correlation in the disturbance. The model can also be checked for adequacy by doing a chi-square test, known as the Ljung-Box Q statistic, on the autocorrelations of the residuals. The test statistic is:

$$Q = (N - d) \sum_{k=1}^{m} {r_k}^2$$

which is approximately distributed as a chi-square variate with "*k-p-q*" degree of
freedom. In this equation
N = length of the time series.
k = First k autocorrelation being checked.
m = Maximum no. of lags checked.
r_k = Sample autocorrelation function of the k^{th} residual term.
d = Degree of differencing to obtain a stationary series.
If the calculated value of Q is larger than the chi-square for k-p-q degree of freedom, the model should have been considered inadequate. It is possible that two or more models have been judge to be approximate, yet none of the models may be an exact fit for the data. In this case, the principle of parsimony should prevail, and simpler model should have chosen.
The data were model using Autoregressive Integrated Moving Average (ARIMA) model as proposed by *Box and Jenkins* (1976) [4]. An ARIMA (p,d,q) model is a combination of Autoregressive (AR) which shows that there is a relationship between present and past values, a random value and a Moving Average (MA) model which shows that the present value has something to do with the past residuals. The ARIMA model denoted as ARIMA (p, d, q) has the general form given by;

$$Y_t = \emptyset_1 X_t + \emptyset_2 X_{t-2} + \ldots + \emptyset_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots \theta_q e_t - q$$

Where, Y_t is the dependent variable at time t
X_{t-1}, X_{t-2}, ..., X_{t-p} are response variables at time lags t-1, t-2, ..., t-p
$Ø_1$, $Ø_2$, ..., $Ø_p$ are coefficients of past variables
e_{t-1}, e_{t-2}, ..., e_{t-q} are past errors
and $Θ_1$, $Θ_2$, ..., $Θ_q$ are coefficients of past errors

The above equation simply means that any given series X_t can be modeled as a combination of past errors e_t or past values X_t or both. Four steps are to be followed when analyzing data using ARIMA model. Firstly, the original series X_t is to be transformed in order for it to become stationary in its mean and variance. Stationarity condition is achieved when the series becomes constant in its mean and variance. Secondly, is the specification of order p and q; this is done by selecting the order that has the least values of log-likelihood, AIC, SBC and Hannan-Quin. Thirdly, is the estimation of the parameters $Ø_1$, $Ø_2$, ..., $Ø_p$ and/or $Θ_1$, $Θ_2$,..., $Θ_q$ using non-linear optimization procedure which will minimize sum of square roots. Finally the seasonal series is modelled practically and the of the order of the models specified. This stage includes carrying out diagnostic checks that show random residuals after which the model can be adopted for purposes of forecasting. The data was subjected to first and second log-differencing in order to attain the stationarity condition necessary for ARIMA. Stationarity transformations also involved plotting time series ACF plots and a review of descriptive summary statistics. Suitability of the model is achieved through Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (SBC), Log-likelihood Estimation and Hannan-Quinn Information Criterion. Diagnostic checks are carried out with the aid of tests for normality of residuals.
In this study, we are concerned with the long time-series crop yield data and the emphasis is to forecast a future value on the basis of previous time-series observations. In the standard regression analysis, the various observations within a single series are assumed to be statistically independent. However, with most time-series data, this assumption may not hold true. Therefore, the standard regression analysis is generally not adequate for forecasting time series data as the observations in the series may not be statistically independent. The Box-Jenkins (1976) [4]. methodology is a powerful tool for time-series analysis, when the time-sequenced observations in a data series may be statistically dependent or related to each other. In accordance with the objective formulated, 'a study on yield trends of barley crop in India' has been undertaken to see the forecasting performance of the developed ARIMA models.

Results and Discussions

The models have been fitted/tested using the barley production data of the period 1980-81 to 2017-18 for the country India. The models have been validated for the post-sample period i.e. 2013-14, 2014-15, 2015-16, 2016-17 and 2017-18, not included in the development of the models. The orders of AR and MA polynomials i.e. values for p and q were determined from the autocorrelation functions and partial autocorrelation functions of the stationary series. Almost all the autocorrelations upto lag 16 significantly different from zero in Table 1 confirm non-stationarity. However, the pacfs showed the presence of one significant spike at lag one, indicating that the series may have autoregressive component of order one (Figure 1).

Table 1: Autocorreations of barley production for India.

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.484	.166	8.442	1	.004
2	.480	.164	17.035	2	.000
3	.412	.161	23.571	3	.000
4	.245	.158	25.952	4	.000
5	.248	.156	28.486	5	.000
6	.151	.153	29.466	6	.000
7	.034	.150	29.517	7	.000
8	.010	.147	29.521	8	.000
9	.022	.144	29.545	9	.001
10	-.018	.141	29.562	10	.001
11	.028	.138	29.601	11	.002
12	-.056	.135	29.777	12	.003
13	-.126	.132	30.693	13	.004
14	.041	.128	30.798	14	.006

15	-.122	.125	31.759	15	.007
16	-.105	.121	32.514	16	.009

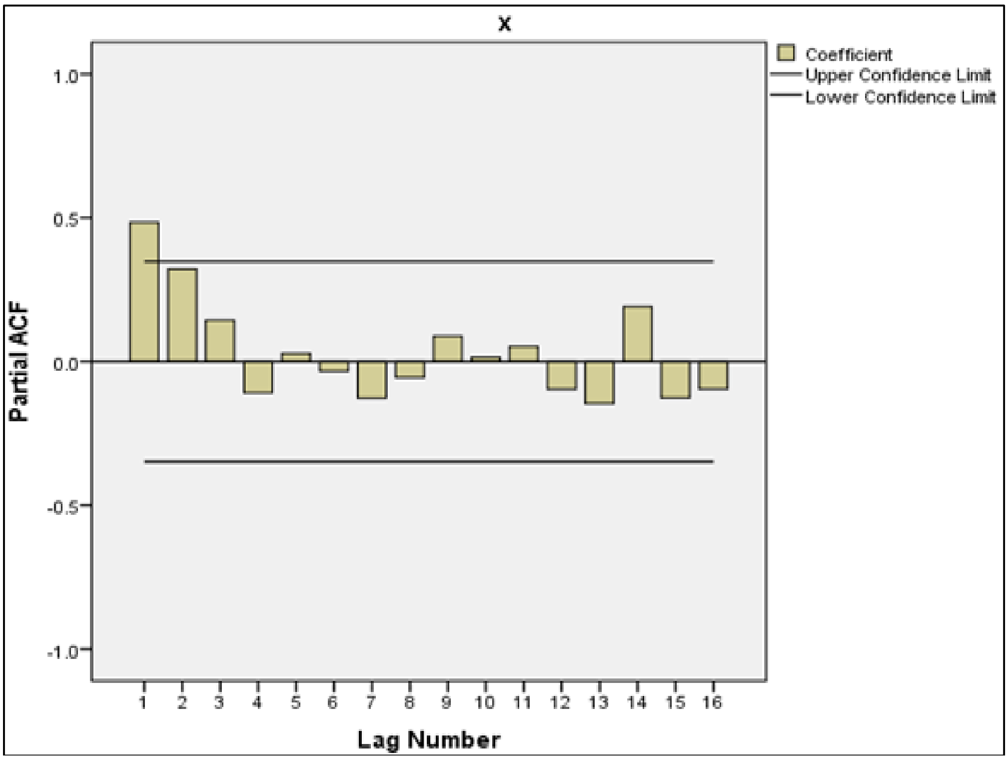

Fig 1: Partial autocorrelations of barley production of India.

The models ARIMA (1,0,0), ARIMA (1,0,1) and ARIMA (1,1,0) were tentatively selected in the identification stage. After experimenting with different lags of the moving average and autoregressive processes, ARIMA (1,1,0) model was fitted for estimating the barley production for India.

Table 2: Tentative ARIMA models for barley production of India.

Models		Estimate	Standard error	t-value	Approx. prob
ARIMA(1,0,0)	AR(1)	0.65	0.13	4.85	<0.01
ARIMA(1,0,1)	AR(1) MA(1)	0.93, 0.47	0.07, 0.21	13.21, 2.25	<0.01, 0.03
ARIMA(1,1,0)	AR(1)	-0.55	0.15	-3.61	<0.01

Table 3: Selection of ARIMA model

Models	RMSE	MAPE	BIC
ARIMA(1,0,0)	217.50	10.60	10.98
ARIMA(1,0,1)	204.99	10.14	10.96
ARIMA(1,1,0)	184.40	8.90	10.65

Table 4: Diagnostic checking of residual autocorrelations of the selected ARIMA model.

Model	Ljung-box Q statistic		
	Statistic	df	Sig.
ARIMA (1,1,0)	15.35	17	0.57

The diagnostic check involved testing whether the residuals from the estimated equations are white noise. All chi-Squared statistic(s) in this concern calculated using the Ljung-Box (1978) [7] formula showed that the residual acfs were not significantly different from zero as shown in Table 4.

Finally, a comparison between ARIMA based estimates and real time barley production was made in terms of percent relative deviation (RD %).

Table 5: RD%=100×(observed yield-estimated yield)/ observed yield.

Model	Year	Observed yield	Estimated yield	Percent relative deviation
ARIMA(1,1,0)	2013-14	1831	1655.10	9.61
	2014-15	1613	1684.46	-4.43
	2015-16	1437	1644.48	-14.44
	2016-17	1747	1642.58	5.98
	2017-18	1781	1619.76	9.05

Conclusion

The Autoregressive integrated moving average (ARIMA) model is considered to be one of the best model when the data consists if at least 50 observations. Barley has been an important commodity of the country. The present study attempts at modelling and forecasting of barly production in India was done using Autoregressive integrated moving average (ARIMA) model. Autocorrelation function (ACF) and partial autocorrelation function (PACF) functions were estimated, which led to the identification and construction of

ARIMA model (1,1,0). The overall results indicates that the percent relative deviations of the forecast yields from the observed yields are within acceptable limits and favours the use of ARIMA models to get short-term forecast estimates.

Bibliography

1. Badmus MA, Ariyo OS. Forecasting cultivated areas and production of maize in Nigeria using ARIMA model. Asian J Agric. Sci. 2011; 3(3):171-176.
2. Balanagammal D, Ranganathan CR, Sundaresan R. Forecasting of agriculture scenario Tamil Nadu – A time series analysis. Journal of Indian Society of Agricultural Statistics. 2000; 53(3):273-286.
3. Boken VK. Forecasting spring wheat yield using time series analysis: A case Study for the Canadian Prairies. Agronomy Journal. 2000; 92:1047-1053.
4. Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Holden Day, San Francisco. 1976; 575.
5. Brown RG. Statistical Forecasting for Inventory Control. McGraw Hill Book Co., Inc., NY, USA, 1959.
6. Falak S, Eatzaz A. Forecasting Wheat production in Pakistan. Lahore J Econ. 2008; 3(1):57-85.
7. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978; 65:67-72.
8. Pindyck RS, Rubinfeld DL. Econometric Models and Economic Forecasts. McGraw Hill Book Co. Inc., NY, USA, 1981.
9. Rachana W, Suvarna M, Sonal G. Use of ARIMA models for forecasting pigeon pea production in India. Int. Rev. Bus. Financ. 2010; 2(1):97-107.
10. Sohail A, Sarwar A, Kamran M. Forecasting total food grains in Pakistan. J Engi. Appl. Sci. 1994; 13:140-146.
11. Verma U, Koehler W, Goyal M. A study on yield trends of different crops using ARIMA analysis. Environment and Ecology. 2012; 30(4A):1459-1463.