Journal of Agricultural Engineering and Food Technology
p-ISSN: 2350-0085; e-ISSN: 2350-0263; Volume 4, Issue 1; January-March, 2017 pp. 22-26

http://www.krishisanskriti.org/Publication.html

Application of ARIMA Model for Forecasting Pulses Productivity in India

Prema Borkar[1] and V.M. Bodade[2]

[1]Gokhale Institute of Politics and Economics (Deemed to be a University), Pune - 411004. Maharashtra. India
[2]Department of Agricultural Economics & Statistics, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola – 444101. Maharashtra, India

Abstract—*Forecasting of any issues, events or variables requires an in-depth understanding of the underlying factors affecting it. Such is the case for forecasting annual productivity of pulse crops. India's ubiquitous position as the leading producer, the foremost consumer and the largest importer of pulses is besmirched by abysmally mediocre policy intervention and equally unimpressive agricultural research budgets. Pulses in India recorded less than 40 per cent growth in production in the past 40 years while its per capita availability declined from 60 grams a day in the 1950 to 39.4 grams a day in 2011. Pulses productivity, in the context of India, extensively depends upon numerous factors namely: good rainfall, timely use of appropriate fertilizer and pesticides, favourable climate and environments etc. Currently, even as production has stabilized at 18.5 million tones, our consumption is hovering at 22 million tones, which necessitates yearly pulse imports of around 3.5-4 million tones. Therefore, forecasting productivity of pulse crop is indispensible, as large chunk of people depends on agriculture for their livelihood.*

Various uni-variate and multi-variate time series techniques can be applied for forecasting such variables. In this paper, ARIMA model has been applied to forecast annual productivity of selected pulse crops. For empirical analysis a set of different has been considered, contingent upon availability of required data. Applying annual data from 1950-51 to 2014-15, forecasted values has been obtained for another 5 years since 2016. The evaluation of forecasting of pulses production has been carried out with root mean squares prediction error (RMSPE), mean absolute prediction error (MAPE) and relative mean absolute prediction error (RMAPE). The residuals of the fitted models were used for the diagnostic checking. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of grain storage, import and/or export and adopt appropriate measures in this regard.

Keywords*: ARIMA, Forecasting, Production, Pulses.*

1. INTRODUCTION

Pulses are an important source of protein, high in fibre content and provide ample quantity of vitamins and minerals. Keeping in view large benefits of pulses for human health, the United Nations has proclaimed 2016 as the International Year of Pulses. Thus, due attention is required to enhance the production of pulses not only to meet the dietary requirement of protein but also to raise the awareness about pulses for achieving nutritional, food security and environmental sustainability. Pulses are important component to sustain the agriculture production as the pulse crops possess wide adaptability to fit into various cropping systems, improve the soil fertility being leguminous in nature and physical health of soil while making soil more porous due to tap root system.

Despite India being the largest producer (18.5 million tons) and processor of pulses in the world also imports around 3.5 million tons annually on an average to meet its ever increasing consumption needs of around 22.0 million tons. According to Indian Institute of Pulses Research's Vision document, India's population is expected to touch 1.68 billion by 2030 and the pulse requirement for the year 2030 is projected at 32 million tons with anticipated required annual growth rate of 4.2 per cent. Thus, there is need to increase production and productivity of pulses in the country by more intensive interventions.

This paper applies Autoregressive Integrated Moving Average (ARIMA) forecasting model, the most popular and widely used forecasting models for uni-variate time series data. Although it is applied across various functional areas, it's application is very limited in agriculture, mainly due to unavailability of required data and also due to the fact that agricultural product depends typically on monsoon and other factors, which the model failed to incorporate. In this context, it is worth mentioning, few applications of ARIMA model for forecasting agriculture product. Applying ARIMA model Hossian *et al.* (2006) forecasted three different varieties of pulse prices namely motor, mash and mung in Bangladesh with monthly data from Jan 1998 to Dec 2000; Wankhade *et al.* (2010) forecasted pigeon pea production in India with annual data from 1950-1951 to 2007-2008; Mandal (2005) forecasted sugarcane production in India; Iqbal *et al.* (2005) forecasted area and production of wheat in Pakistan; Khin *et al.* (2008) forecasted natural rubber price in world market; Shukla and Jharkharia (2011) forecasted wholesale vegetable market in Ahmedabad; Assis *et al.* (2010) forecasted cocoa bean prices in Malaysia along with other competing models; Nochai and Nochai (2006) forecasted palm oil prices in

Thailand; Masuda and Goldsmith (2009) forecasted world Soybean productions; Cooray (2006) forecasted Sri Lanka's monthly total production of tea and paddy beyond Sept 1988 using monthly data from January 1988 to September 2004. With these exceptions, there is paucity of studies regarding applications of ARIMA model for forecasting agricultural products.

In lieu of this, the paper applies ARIMA model for forecasting. The model not only apprehends its own past information but also current and past information of error term and thereby considers all sorts of information surrounded with the uni-variate time series, while forecasting. Although the model is widely used for forecasting any given stationary time series, it is quite robust to handle any data pattern. The application of the model involves certain steps such as identifying, fitting, estimating and forecasting the interested variable. It specifies and identifies the AR and MA process of an integrated series with 0 or 1 order and then forecasts. The details of the methodology, empirical analysis, and details of data, empirical estimation and analysis of results are discussed in the different sections of this paper

2. MATERIALS AND METHODS

The existing study applies Box-Jenkins (1970) forecasting model popularly known as ARIMA model. The ARIMA is an extrapolation method, which requires historical time series data of underlying variable. The model in specific and general forms may be expressed as follows. Let Yt is a discrete time series variable which takes different values over a period of time. The corresponding AR (p) model of Yt series, which is the generalizations of autoregressive model, can be expressed as:

$$AR(p)_t \; Y_t = \phi_o + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots.. + \phi_p Y_{t-p} + \varepsilon_t \quad \ldots\ldots\ldots..(1)$$

Where, Y_t is the response variables at time t,

Y_{t-1}, Y_{t-2}……Y_{t-p} is the respective variables at different time with lags;

ϕ_o, ϕ_1,….. ϕ_p are the coefficients; and ε_t is the error factor.

Similarly, the MA (q) model which is again the generalizations of moving average model may be specified as:

$$MA\ (q) : Y_t = \mu_t + \varepsilon_t + \delta_1 \varepsilon_{t-1} + \ldots. + \delta_q \varepsilon_{t-q} + vt \quad \ldots\ldots\ldots..(2)$$

Where, μ-i is the constant mean of the series;

$\delta_1 \ldots \delta_q$ is the coefficients of the estimated error term;

ε_t is the error term.

Combining both the model is called as ARIMA models, which has general form as:

$$Y_t = \phi_o + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots.. + \phi_p Y_{t-p} + \varepsilon_t + \delta_1 \varepsilon_{t-1} + \ldots.. + \delta_q \varepsilon_{t-q} + vt$$

If Y_t is stationary at level or I(0) or at first difference I(1) determines the order of integration, which is called as ARIMA model. To identify the order of p and q the ACF and PACF is applied. The details of the estimation and forecasting process are discussed below.

Data: To fit an ARIMA model requires a sufficiently large data set. The paper considers annual secondary data on 9 different pulses for forecasting. Data is collected from the website of Indian Institute of Pulses Research e-pulses data book. The period of study for chickpea and pigeonpea is from 1950-51 to 2014-15, whereas for urdbean and mungbean it is 1965-66 to 2014-15, for lentil, peas and lathyrus it is 1970-71 to 2014-15 and for mothbean and kulthi it is 1976-77 to 2014-15 depending upon the availability of data.

3. BOX-JENKINS ARIMA FORECASTING MODEL:

ARIMA forecasting model is applied for large stationary data and involved four different but interrelated steps. These steps and estimated results are discussed below.

Identification: The first step of applying Box-Jenkins forecasting model is to identify the appropriate order of ARIMA (p, d, q) model. Identification of ARIMA model implies selection of order of AR(p), MA(q) and I(d). The order of d is estimated through I(1) or I(0) process of unit root stationarity tests. The model specification and selection of order p and q involved plotting of autocorrelations (ACF) and partial autocorrelations functions (PACF) or correlogram of variables at different lag length. The autocorrelation functions specify the order of moving average process, q and partial autocorrelations select autoregressive of order p. The ACF shows autocorrelation coefficients at different lag length with 95 per cent confidence interval whether they are statistically different from zero or not. For example, if up to certain lag, say 6, the autocorrelation coefficients lies outside the 95 per cent confidence bound, then it selects the order of q as 6. Similarly order p is selected from PACF. The significance level of individual coefficients is measured by Box-Pierce Q statistics and for all the coefficients jointly together by Ljung-Box statistics. The Box-Pierce Q statistics is defined as

$$Q = \sum^{m} \hat{\rho} k^2 \sim \chi^2_m ;$$

and Ljung Box (LB) Statistics is defined

$$\text{as } LB = n(n+2)\sum_{k=1}^{m} \frac{\hat{\rho} k^2}{n-k} \sim \chi^2_m \text{ where n= sample size and ,}$$

m is lag length. Due to limited space, the results of ACF and PACF are not reported, but can be obtained from the author upon request. Thus, in the process it selects the order of p and q.

Estimation of the model: Once the order of p, d, and q are identified, next step is to specify appropriate regression model and estimate it. With the help of SPSS software various order of ARIMA model has been estimated to arrive at the optimal model. For example, if a series is identified as ARIMA (2, 1, 1) it means the series is stationary at first difference and follows AR (2) and MA (1) process. The regression model is estimated with simple ordinary least squares methods. Once the model is estimated, significance of each coefficient are tested applying 't' test and jointly together by 'F' test. The adjusted R^2 provides whether the model is a good model or not as does in case of multiple linear regression models. As cited above, in their respective studies, the most suitable ARIMA (1,1,1) model was selected by Wankhede *et al.* (2010), ARIMA (2,1,0) by Mandal (2005); ARIMA (1,1,1) and ARIMA (2,1,2) for area and productions of wheat by Iqbal; multiplicative ARIMA $(3,1,3)\times(2,0,2)_{12}$ for both motor and mash prices and multiplicative ARIMA $(3,1,2)\times(3,0,2)_{12}$ mung prices by Hossain (2006).

Diagnostic checking: Now the question may arise, how do we know whether the identified model is appropriate or not? One simple way to answer is diagnostic checking on residual term obtained from ARIMA model applying the same ACF and PACF functions. Obtain ACF and PACF of residual term up to certain lags of the estimated ARIMA model and then check whether the coefficients are statistically significant or not with Box-Pierce Q and Ljung-Box LB statistics, respectively. If the result obtains from the model is purely random, then estimated ARIMA model is correct or else we have to look for alternative specification of the model. Similarly, diagnostic checking can also be done through Adjusted R2, minimum of Akaike Information Criteria (AIC) and Schwarz Bayesian Criteria (SBC) and lowest mean absolute percent error (MAPE). The paper reports the Adjusted R2, minimum of AIC and MAPE values to obtain the optimal ARIMA model. Table II reports the estimated results.

Forecasting: Once the three previous steps of ARIMA model is over, then we can obtained forecasted values by estimating appropriate model, which are free from problems. The forecasted values obtained from ARIMA model are reported in Table II. The forecasted values are reported for a maximum 5 years as too much long term forecasting might not be appropriate. For example, Chickpea crop follows ARIMA (0, 1, 1) model, with Adjusted R^2 being equal to 0.249, the forecasted values for 2016 is 8.24 million tonnes, for 2017 it is 8.32 million tonnes, for 2018 it is 8.40 million tonnes and so on. Similarly, Pigeonpea follows ARIMA (0, 1, 1) model. The forecasted values for 2016 and onwards are 2.85 million tonnes, 2.87 million tonnes and so on and so forth.

4. CONCLUSION

ARIMA model offers a good technique for predicting the magnitude of any variable. Its strength lies in the fact that the method is suitable for any time series with any pattern of change and it does not require the forecaster to choose a priori the value of any parameter. Its limitations include its requirement of a long time series. As the model requires large data points, considering the availability of required annual data, 9 different pulses data has been selected. Annual data from 1951, 1966, 1971 and 1977 onwards to 2015 as the case may be have been used. All the necessary steps of ARIMA model have been applied systematically for forecasting 5 periods ahead from 2016 onwards. Among these items, urdbean provides lowest MAPE value, whereas mungbean provides lowest AIC values. Similarly, highest MAPE is obtained for peas and highest AIC value is for chickpea. Now the question may arise is since agricultural productivity depend upon many factors such as rainfall, irrigation facility, monsoon, climate, soil, fertilizer etc., forecasted values might be more accurate only with ceteris paribus assumption. However, generally all the factors do not go well every time and in right direction; therefore reliability of these forecasted values might be questionable. In this context one need to rethink about other forecasting model, which could incorporate more information for forecasting the agricultural products. This could be one of the limitations of the paper.

REFERENCES

[1] Assis, K., A. Amran, Y. Remali and H. Affendy, 2010. A comparison of univariate time series methods for forecasting cocoa prices. Trends Agric Econ., 3: 207-215.

[2] Box, G. and G. Jenkins, 1970. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

[3] Hossain, M. Z., Q. A. Samad and M. Z. Ali, 2006. ARIMA model and forecasting with three types of pulse prices in Bangladesh: A case study. Int. J. Social Econo., 33: 344-353.

[4] Mandal, B. N., 2005. Forecasting Sugarcane Production in India with ARIMA Model. Inter Stat, October, 2005.

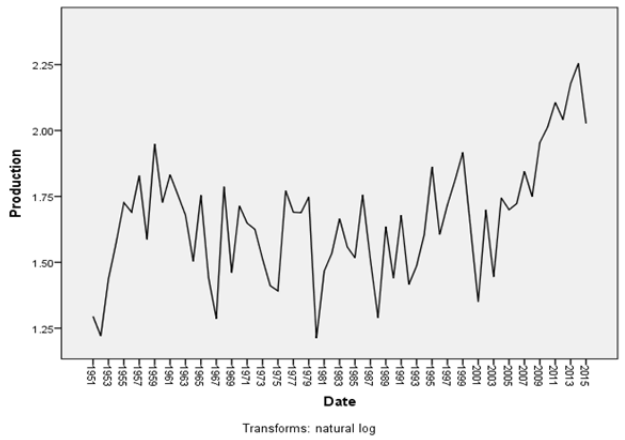
[5] Shukla, M. and S. Jharkharia, 2011. Applicability of ARIMA models in wholesale vegetable market: An investigation. Proceedings of the 2011 International Conference on Industrial Engineering and Operations Managenment. Kuala Lampur, Malaysis, January 22-24, 2011.

Table 1: Descriptive statistics for various pulse crops

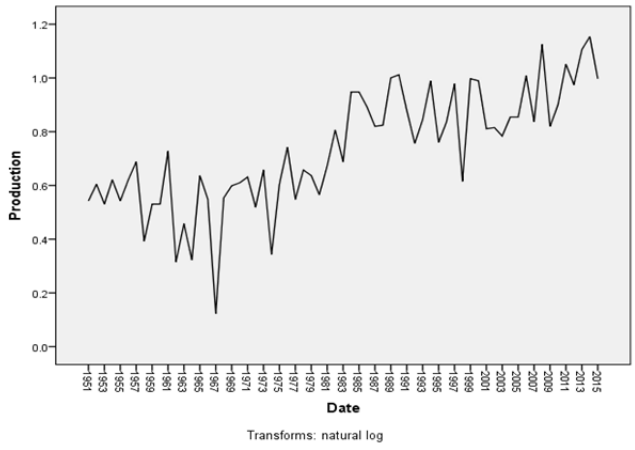
Variables	Mean	Max	Min.	Std. Dev.	Sk.	Kurtosis	C.V.	Obs.
Chickpea	5.39	9.53	3.36	1.27	0.98	1.33	23.54	65
Pigeonpea	2.13	3.17	1.13	0.46	0.23	-0.54	21.34	65
Urdbean	1.16	1.90	0.52	0.40	-0.23	-1.12	34.37	50
Mungbean	1.04	1.80	0.40	0.34	0.15	-0.52	32.76	50
Lentil	0.74	1.13	0.30	0.25	-0.15	-1.26	34.40	45
Peas	0.55	0.84	0.15	0.18	-0.24	-0.78	32.11	45
Mothbean	0.30	0.83	0.04	0.19	0.83	0.99	61.50	39
Lathyrus	0.49	0.81	0.24	0.14	0.70	0.69	27.94	45
Kulthi	0.44	0.76	0.21	0.19	0.25	-1.58	44.07	39

Table 2: Forecast values with ARIMA model

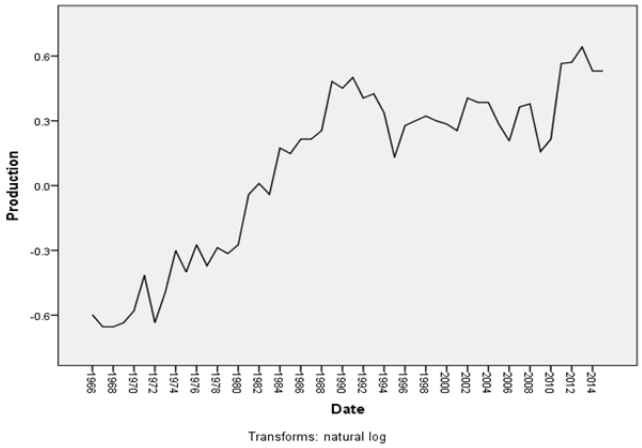
Variables	ARIMA (p,d,q) Model	Adj R^2	MAPE	AIC Values	Forecast Values (million tones)				
					2016	2017	2018	2019	2020
Chickpea	0,1,1	0.249	15.122	0.974	8.24	8.32	8.40	8.46	8.52
Pigeonpea	0,1,1	0.428	10.948	0.862	2.85	2.87	2.89	2.91	2.92
Urdbean	0,1,0	0.016	8.980	0.373	1.75	1.80	1.86	1.92	1.97
Mungbean	0,1,1	0.404	18.231	0.341	1.51	1.54	1.58	1.61	1.65
Lentil	0,1,1	0.248	9.896	0.805	1.18	1.22	1.25	1.29	1.33
Peas	0,1,0	0.016	21.018	0.936	0.88	0.92	0.97	1.01	1.06
Mothbean	1,0,1	0.038	11.274	0.833	0.24	0.25	0.26	0.27	0.27
Lathyrus	1,0,1	0.296	18.793	0.488	0.41	0.42	0.42	0.43	0.43
Kulthi	1,0,1	0.848	13.632	0.838	0.23	0.24	0.25	0.26	0.27



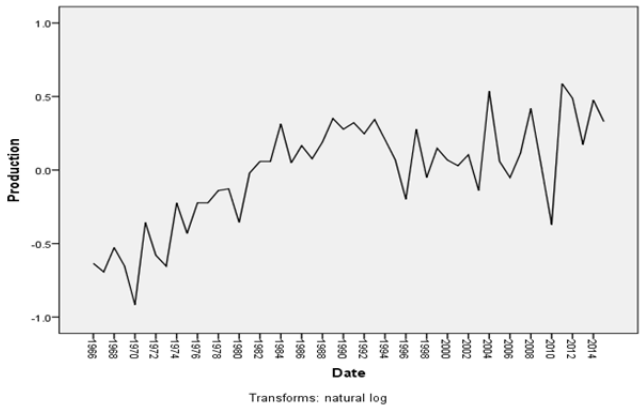
Chickpea



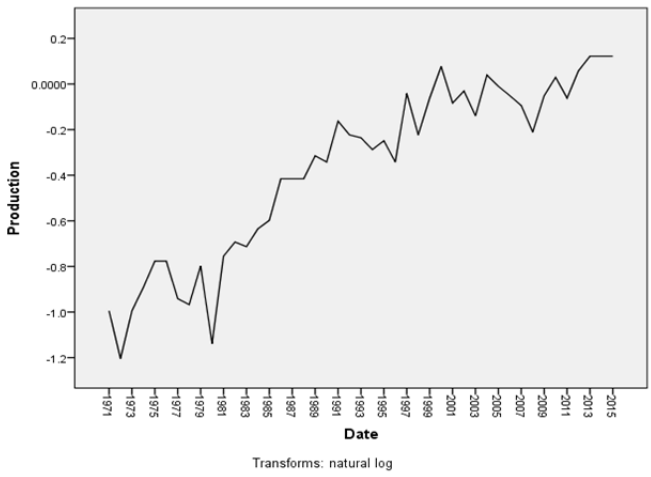
Pigeonpea



Urdbean



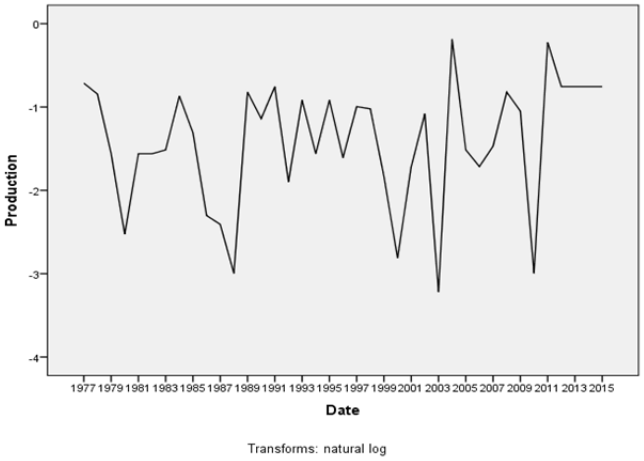
Mungbean



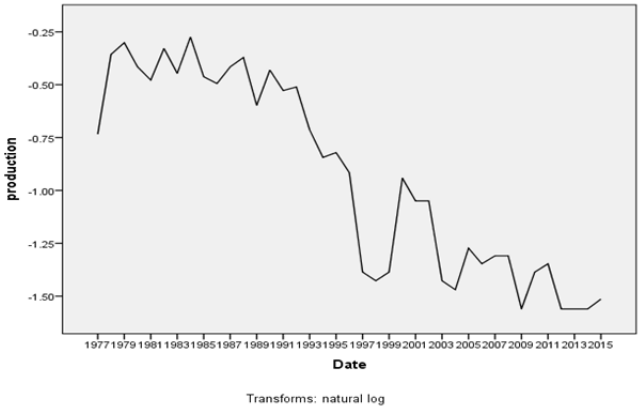
Lentil

Peas

Lathyrus



Mothbean



Kulthi

Fig. 1: The time series plot of different pulses in India