Asian Journal of Economics, Business and Accounting
2(3): 1-8, 2017; Article no.AJEBA.32414

SCIENCEDOMAIN *international*
www.sciencedomain.org

ARIMA Modeling to Forecast Pulses Production in Kenya

N. Mwangi Esther[1*] and N. Wangui Magdaline[2]

[1]Department of Economics, Karatina University, P.O.Box 1957-10101, Karatina, Kenya.
[2]Department of Economics, Moi University, P.O.Box 3900-30100, Eldoret, Kenya.

Authors' contributions

This work was carried out in collaboration between both authors. Author NME designed the study, collected data, analyzed and wrote the first draft of the manuscript. Author NWM managed the theoretical framework and the literature review in conjunction with author NME. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJEBA/2017/32414
Editor(s):
(1) Chen Zhan-Ming, Associate Professor, Department of Energy Economics, School of Economics, Renmin University of China, China.
Reviewers:
(1) Fateh Benatia, Mohamed Khider University of Biskra, Algeria.
(2) K. P. Vishwajith, Bidhan Chandra Krishi Viswavidyalaya, West Bengal, India.
(3) Razana Alwee, Universiti Teknologi Malaysia, Malaysia.
(4) Taofeek Aremu Kasali, Moshood Abiola Polytechnic, Abeokuta, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/18627

Original Research Article

Received 24th February 2017
Accepted 4th April 2017
Published 13th April 2017

ABSTRACT

Pulses are vital staple foods in Kenya and are ranked second after cereals. This study focuses on forecasting production of pulses in Kenya using Autoregressive Integrated Moving Average (ARIMA) model. Time series data on production collected from Food and Agriculture Organization of the United Nations (FAO) statistical yearbooks for the period 1961 to 2012 was used to model. The study found that ARIMA (1,1,2) was the appropriate model to forecast pulses production in Kenya. Based on this model the point forecasts showed that 25437.53 tons of pulses would be produced in 2020, 25342.27 tons in 2025 and 25357.44 tons in 2030 holding all things constant. The government should formulate and implement effective policies to promote pulses production in order to meet their consumption demand for the growing population and enhance food security in Kenya.

Keywords: ARIMA; forecast; Kenya; production; pulses.

**Corresponding author: E-mail: emwangi@karu.ac.ke;*

1. INTRODUCTION

Pulses are grain legumes belonging to the family leguminosae and mainly include dry beans, dry peas, chickpeas, cowpeas, peanut and lentils. They are essential food crops due to their nutritional value and soil nitrogen fixation benefits. Pulses are an important source of plant proteins, fiber, various vitamins and essential amino acids [1]. They contain 20-25 percent proteins and are rich in amino acids lysine and methionine. They contribute 7.5 per cent of total protein intake in developing countries and 2.5 per cent in developed countries. In Kenya, pulses are ranked second most important staple food after cereals and contribute to about 20 per cent of per capital protein intake [2]. Furthermore, pulses are gaining recognition globally as a healthy diet. The year 2016 was declared the International year of pulses (IYP) by the 68th UN general assembly which aimed at creating public awareness of the nutritional and environmental benefits of pulses towards global sustainable food production and food security [1]. Pulse-based proteins consumption is increasing overtime both in developed and developing countries not only because of their nutritional value but also because they are relatively cheaper compared to animal sources of proteins [3]. Like other leguminous plants, pulses play a significant role in nitrogen fixation in the soil thus improves soil fertility. The IYP 2016 thus aimed at educating the public through collaboration with government and non-government organizations on the importance of practicing crop rotation.

Pulses play an important role in Kenya's economy in that their production and value addition chains has the potential to contribute to development as an economic activity and a source of livelihood. Though Kenya has recently been ranked as a middle income economy, food insecurity is still a major problem. Pulses can solve the problem of protein deficiency due to their relatively high protein content and they command relatively lower market price compared to animal sources of proteins. However, World pulses production shows a declining trend from the 1970s to the late 2001 [4]. Moreover, [2] show that production of pulses does not receive much attention in terms of investment resources compared to cereals despite the important role they play. Globally, pulses are allocated less land resources that is about 80 million hectares and do not receive much policy attention. This global total area harvested for pulses is far below the area harvested for all cereals which is about 700 million hectares on average. Based on the projected rate of population growth, world consumption of pulses is expected to increase by 23% from the levels in 2008 by 2030 and much of the growth in pulses consumption is expected to be in Asia and Africa [5]. Pulses production therefore needs to be increased to meet the projected global demand for pulses consumption. The area harvested need be scaled up by 9.5 million hectares by 2030 over and above the global pulse area harvested in 2008 [2].

In Kenya pulses are often intercropped with cereals and treated as secondary crops relative to cereals and other cash crops. Moreover, pulses production is done in a traditional way by small holder farmers resulting to low yields. Pulses production is highly affected by the area cultivated, climatic conditions, the amount of key inputs such as seeds and fertilizer and government policies. Consumption of pulses outweighs their production in Kenya causing Kenya to be a net importer of pulses despite her potential to produce enough for domestic consumption and for export. Low production and low stock levels causes price spikes escalating the severity of food insecurity. The seasonal nature of pulses production result to high market volatility. Farmers who are also the consumers lack information about the future prospects of pulses production output and their future market prices while making their farm decisions. As the population continues to grow there is need to plan on how to meet the growing pulses consumption demand. Forecasting pulse production in Kenya is therefore imperative to provide information on the quantity that is likely to be produced in the future. This information is of importance to the government, farmers, consumers and middlemen in production planning against any potential risk to ensure consumption demand is met and also plan for possible opportunity in case of excess production.

ARIMA modeling has widely been used to forecast production using time series data. [6] carried out a study on forecasting pigeon pea pulse production in India using ARIMA model. In their study, ARIMA (1,1,1) model was selected as the best model for forecasting Pigeon pea pulse production. Using the model, pigeon pea production was expected to increase from 2.49479 million tons in 2008-09 to 2.73452 million tons for the year 2014-2015 with lower and upper limits of 2.05787 million tones and 3.41116 million tons respectively. [7] carried out a study on forecasting cultivated areas and production of maize in Nigeria using time series data for the period 1970-2005. ARIMA (1,1,1) and ARIMA (2,1,2) models were selected to forecast maize area cultivated and maize production respectively. [8] used data for the period 1968-2011 to forecast lentil pulse production in Bangladesh. They applied box-Jenkins ARIMA approach to model and forecast. Their study established that ARIMA (0,1,9) was the appropriate model to forecast lentil production in Bangladesh. Using this model, forecasts revealed that production of lentil in Bangladesh would decrease from 89625 metric tons in 2011-2012 to 58601 metric tons for the year 2015-2016. [9] forecasted wheat production in Pakistan using Box-Jenkins (1976) linear time series models. Using wheat production data for 1902-2005, ARIMA (1,2,2) was chosen as the best forecasting model for forecasting wheat production of Pakistan. The forecast showed that wheat in Pakistan would rise to 26623.5 thousand tons in 2020 and would double in 2060 as compared to the amount produced in 2010. [10] forecasted production of pulses in India using ARIMA model and Generalized Autoregressive Conditional Heteroscedastic (GARCH) models. Their study revealed that none of the models proved to be superior to the other in modeling and forecasting production of pulses in India. Based on the efficiency of ARIMA modeling in production forecasting, this study used the methodology to model and forecast pulses production in Kenya.

2. MATERIALS AND METHODS

The study aimed at forecasting production of pulses in Kenya by 2030. The study was carried out using time series data for the period 1961 to 2012. Secondary data on pulses production was collected from Food and Agriculture Organization Statistical Database (FAOSTAT). ARIMA (p,d,q) model also known as the Box-Jenkins approach was employed. An ARIMA (p,d,q) model is a combination of Autoregressive (AR), a random value also called the integration part (d) and a Moving Average (MA). AR shows that there is a relationship between present and past values. MA shows that the present value has something to do with the past residuals. The ARIMA model can be expressed as;

$$\phi(B)(\Delta^d y_t - \mu) = \theta(B)\varepsilon_t$$

Where, y_t = pulse production in tons

μ = Mean of $\Delta^d y_t$

$\phi(B) = 1 - \phi_1 B - ...... - \phi_P B^P$

$\theta(B) = 1 - \theta_1 B - ...... - \theta_q B^q$

ϕ_i = The ith autoregressive parameter

θ_i = The ith moving average parameter

p denotes the number of time lags, the integration part *d* is the degree of differencing i.e. the number of times the series has to be differenced before it becomes stationary and *q* the number of moving average terms. The ARIMA modeling involves the following steps.

2.1 Model Identification

In the identification stage the data is tested for stationary and the augmented dickey fuller test was applied. If the data is stationary, *d*=0 and the ARIMA (p,d,q) becomes ARMA (p, q). On the other hand if the data is not stationary, it is transformed by differencing in order to obtain a stationary series. The number of times the series is differenced determines the order of d. The AR and MA signatures are determined using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. A theoretical AR model of order *p* have an ACF that decays exponentially and a PACF that cuts off at lag *p* while a theoretical MA model of order q consist of a PACF that decays exponentially and an ACF that cuts off at lag q. The tentative time series model is specified on the basis of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) likelihood criteria developed by [11]. The model with the minimum AIC and BIC values is selected as the model that fits the data best.

2.2 Estimation of Parameters and Diagnostic Checking

Parameters of tentative best fit ARIMA model are estimated. The model is then checked for adequacy by testing whether the parameter estimates are significant. The residuals from the estimated model are generated and tested whether they resemble a white noise series (uncorrelated and have zero mean). Autocorrelation is tested by computing the Ljung-Box statistic test developed by [12]. ACF and PACF plots of the residuals are also used to detect the autocorrelation in the residuals [13].

If the parameter estimates are insignificant and the residual are not a white noise the entire process of model identification, parameter estimation and diagnostic checking is repeated until the appropriate model is attained.

2.3 Forecasting

After selection of an appropriate model, future values of the time series are forecasted and the confidence intervals for the forecasts generated. Reliability of forecasted values based on selected model is checked by computing Root Mean Square Error (RMSE) or Mean Absolute Error (MAE).

3. RESULTS AND DISCUSSION

The maximum production of pulses in Kenya was 374000 tons in 1983 while the lowest production was 10952 tons in 1997. On average the production of pulses in Kenya from 1961 to 2012 was 169285.4 tons. However a significant decrease in production was realized since 1996 onwards as illustrated in Figure 1. [2] reports that pulses get relatively lower land and other resources allocation globally compared to cereals. The same case applies to Kenya where the pulses area cultivated is much less compared to cereals, coffee and tea. There is low investment in pulses production and much investment and policy attention is directed to the main cash crops. Moreover, land degradation, climate change and lack of superior agricultural technology are responsible for the decrease in pulses production in Kenya. The government should therefore consider formulating and implementing policies that will promote pulses production such as increasing pulses area cultivated by cultivating marginalized land, use of irrigation, subsidizing pulses certified seeds and fertilizer among others.

The augmented dickey fuller test for pulses production data was done to test whether the data was stationary. The results (Table 1) showed that the data was not stationary since the p-value is greater than 0.05 at 5% level of significance so we fail to reject the null hypothesis which states that the data is not stationary. The data was differenced to make it stationary. After the first differencing, the augmented dickey fuller test for the differenced data gave the results shown in Table 2. The first differencing was sufficient to make the data stationary since the p-value was less than 0.05 at 5% level of significance hence production is integrated of order one, d=1. Figure 2 shows a plot of the differenced pulses production data against time.

Table 1. Test for stationary of production data

Dickey-Fuller	Lag order	p-value
-2.3282	3	0.4424

Table 2. Test for stationary of differenced data

Dickey-Fuller	Lag order	p-value
-5.0684	3	0.01

ACF and PACF plots for the original production data were done (Fig. 3). The results implied that production data was not stationary since the ACF die off slowly.

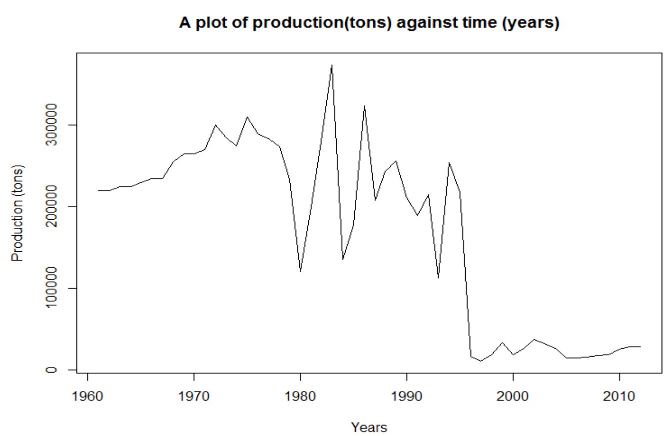

Fig. 1. Pulses production in Kenya (1961-2012)

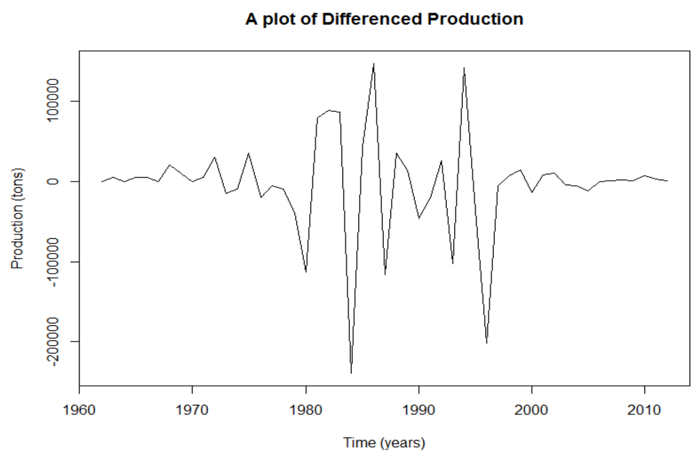

Fig. 2. Differenced pulses production in Kenya (1961-2012)

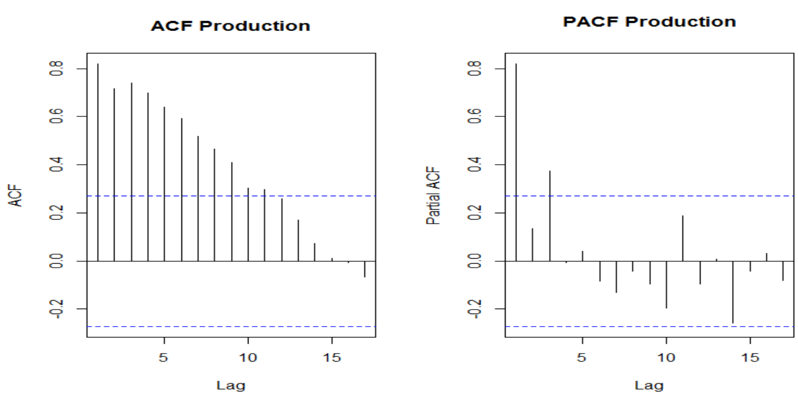

Fig. 3. ACF and PACF plots for pulses production

Since there were some spikes in the plots outside the insignificant zone (dotted horizontal blue lines) there existed some information to be extracted by AR and MA models. In order to obtain the order of AR and MA, the Plots of ACF and PACF for the differenced production data was done and the results are shown in Fig. 4.

The results implied that the model had an AR of order 1 as indicated by the PACF plot and an MA of order 2 as indicated by an ACF plot. The next step was to identify the best fit ARIMA model to be used in forecasting pulses production. The auto ARIMA function in R which identifies the best fit ARIMA model based on the minimum values of AIC and BIC values was applied. The results showed that the most appropriate ARIMA model to be used to forecast pulses production in Kenya is ARIMA (1,1,2) (Table 3). This was in line with the order of differencing the data (order 1) and the results of ACF and PACF Plots of the differenced data (Fig. 4).

Table 3. ARIMA fit and model coefficient estimates

ARIMA (1,1,2)

Parameter	Coefficient estimate	Standard error
AR1	-0.6925	0.1545
MA1	0.3854	0.3209
MA2	-0.5976	0.1905
Log likelihood=629.23		
AIC=1266.45	AICc=1267.32	BIC=1274.18

A diagnostic checking was done by plotting ACF and PACF of the residuals of the best fit ARIMA model, that is ARIMA (1,1,2) to check for any autocorrelation in the residuals. Ljung-Box test was also performed to check for independence of the residuals. The ACF and PACF plot of the

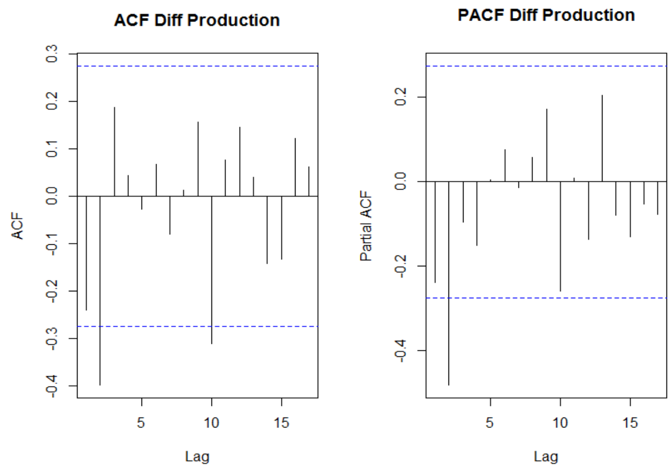

Fig. 4. ACF and PACF plots for differenced pulses production data

Table 4. ARIMA (1,1,2) validity summary

ME	RMSE	MAE	MPE	MAPE	MASE
-8061.059	53575.16	30951.67	-53.26992	64.14746	0.8546268

residuals of the ARIMA (1,1,2) model (Fig. 5) verified that the residuals are random and represent a white noise since there are no spikes outside the insignificant zone for both ACF and PACF plots. Ljung-Box test results (Table 5) has p-value being greater than 0.05 at 5% level of significance indicating that the residuals are independent. Hence, ARIMA (1,1,2) model was appropriate and working well.

Table 5. Test for independence

x-squared	df	p-value
0.072615	1	0.7876

The model for forecasting pulses production can be expressed as;

$$\hat{y}_t = -0.6925\hat{y}_{t-1} + 0.3854\hat{e}_{t-1} - 0.5976\hat{e}_{t-2}$$

Where $\hat{y}$ is the forecasted pulses production in time *t*, $\hat{y}_{t-1}$ is the forecasted pulses production in time *t-1*, $\hat{e}_{t-1}$ is the residual in time *t-1* and $\hat{e}_{t-2}$ is the residual in time *t-2.*

Finally, the ARIMA (1,1,2) model was used to forecast pulse production in Kenya by 2030.

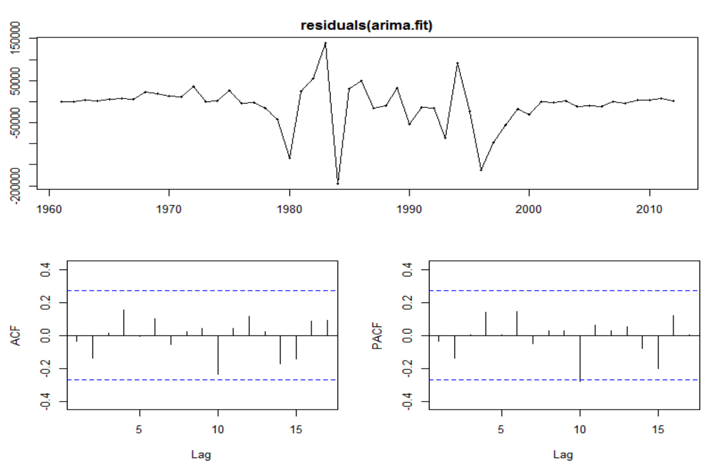

Fig. 5. Residual plot for ARIMA (1,1,2)

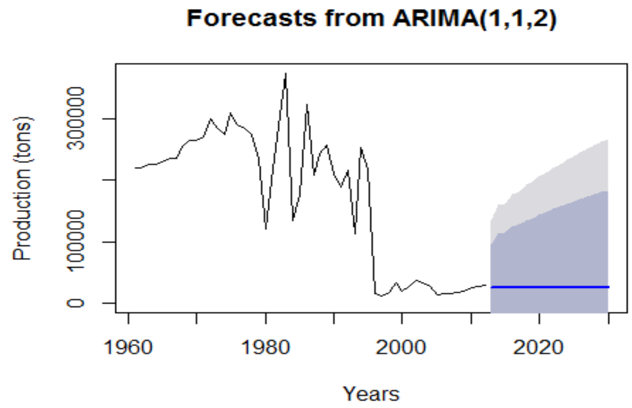

Fig. 6. Actual and forecasted pulses production in Kenya

We found that 25437.53 tons of pulses would be produced in 2020, 25342.27 tons in 2025 and 25357.44 tons in 2030 holding all things constant. The forecasted values with an 80% and 95% prediction intervals are plotted in Fig. 6. Pulses production in Kenya would be decreasing compared to the actual 28500 tons produced in 2012. These results are in line with [2] findings that pulses production does not receive much attention globally.

4. CONCLUSION

The study used Box-Jenkins ARIMA modeling to forecast production of pulses in Kenya. The results showed a decreasing trend in the production of pulses despite their nutritional and environmental benefits. Due to the increasing trend in population growth coupled with the decreasing trend in production, it is evident that there will not be enough pulses to feed the growing population in Kenya by 2030. The government will have to depend on imported pulses and this might lead to increased pulse prices. To curb this, the government should increase investment in pulses production, provide subsided certified seeds and fertilizer to the farmers and educate all stakeholders on the benefits associated with pulses production, consumption and value chain addition. Moreover the government should employ land tenure systems that would promote pulses production in Kenya.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Food and Agriculture Organization of the United Nations. Biannual report on global food markets. 2016;0251-1959. Available:www.fao.org/
2. Akibode S, Maredia M. Global and regional trends in production, trade and consumption of food legume crops; 2011. Available:http://impact.cgiar.org
3. USDA-Economic Research Service (ERS). Vegetables and melons outlook. 2011;23-26.
   Available:http://www.ers.usda.gov/briefing/drybeans/PDFs/DBnOutlook
   (Accessed March 2011)
4. Ben Belhassen B. Pulses: Past trends and future prospects; 2006.
   Available:http://www.fao.org/es/esc/common/ecg/98/en/PulsesStudy.pdf
   (Accessed March 2011)
5. Clancey B. World Pulse Outlook: Report to the Saskatchewan Pulse Growers. Stat Publishing; 2009.
6. 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.
7. Badmus MA, Ariyo OS. Forecasting cultivated areas and production of maize in Nigeria using ARIMA model. Asian Journal of Agricultural Sciences. 2011;3(3):171-176.
8. Farhat R, Azizul B, Ashutosh R, Monir H. Forecasting of lentil pulse production: An econometric analysis. Australian Journal of

Basic and Applied Sciences. 2013; 7(2):819-824.
9. Amin A, Amanullah M, Akbar A. Time series modeling for forecasting wheat production of Pakistan. The Journal of Animal & Plant Sciences. 2014;24(5):1444-1451.
10. Vishwajith KP, Dhekale BS, Sahu PK, Mishra P, Noman MD. Time series modeling and forecasting of pulses production in India. Journal of Crop and Weed. 2014;10(2):147-154.
11. Akaike H. Information theory and an extension of the maximum likelihood principle. Proc.2nd inter. Symposium on Information Theory. 1973; 267-281.
12. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978;65:297-303.
13. Elivli S, Uzgoren N, Savas M. Control charts for autocorrelated colemanite data. Journal of Scientific and Industrial Research. 2009;68:11-17.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/18627