

Production and Consumption of Corn in Ghana: Forecasting Using ARIMA Models

Nasiru Suleman and Solomon Sarpong

Department of Statistics, Faculty of Mathematical Sciences, University for Development Studies, P.O. Box 24, Navrongo, Ghana, West Africa

Abstract: This study describes an empirical study of modeling and forecasting production and consumption of corn in Ghana using ARIMA models. The study revealed that ARIMA (2, 1, 1) and ARIMA (1, 1, 0) were the appropriate models for forecasting production and consumption respectively. Our forecast showed an increasing pattern in consumption and production of corn. Despite the increase in production, government has to still invest money into corn production, motivate corn farmers, and implement good policies for better land tenure systems for corn cultivation to ensure that production always exceeds consumption to avoid importation of corn into the country. This is imperative because importation could lead to high prices of corn and increase inflation rate, hence affecting the economy of the country.

Keywords: ARIMA, consumption, corn, Ghana, production

INTRODUCTION

Corn also known as maize (*Zea mays* L.) is an important staple food for more than 1.2 billion people in Sub-Saharan Africa (SSA) and Latin America. The consumption of corn is more than 116 million tons with Africa consuming 30% and SSA, 21%. The cereal is cultivated all over the world in a range of agro-ecological environment with worldwide production of 785 million tons (IITA, 2009). According to the Millennium Development Authority, corn is the number one crop cultivated in terms of area and accounts for 50-60% of total cereal production in Ghana. The cereal is cultivated along with rice and millet in most parts of the country especially Northern Ghana. The three cereals contribute significantly to agricultural Gross Domestic Product (GDP) and the economy of the country.

In Ghana, corn is a staple food of great socio-economic importance. The cereal is used in preparing local dishes and drinks. The consumption of corn sometimes outstrips the production, making it a threat to the country's economy. It is obvious that when consumption exceeds production, the country has to import in order to fill this gap in production. This would result in high prices of corn and thus increase inflation rate in the country. It is therefore necessary to monitor the pattern of consumption and production in order to curb such challenges. To do this, it is imperative to forecast future values of consumption and production.

Forecasts have been made using parametric univariate time series models, known as Autoregressive Integrated Moving Average (ARIMA) model popularized by Box and Jenkins (1976). These approaches have been employed extensively for forecasting economics time series, inventory and sales modeling (Brown, 1959). Ljung and Box (1978) and Pindyck and Rubinfeld (1981) have also discussed the use of univariate time series in forecasting. Rachana *et al.* (2010), used ARIMA models to forecast pigeon pea production in India. Badmus and Ariyo (2011), forecasted 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 Eatraz (2008), 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 methodology have been used extensively by a number of researchers to forecast demands in terms of internal consumption, imports and exports to adopt appropriate solutions (Muhammed *et al.*, 1992; Shabur and Haque, 1993; Sohail *et al.*, 1994).

Thus in this study, we modeled and forecasted the consumption and production of corn in Ghana using ARIMA methodology. This would help predict future values of consumption and production in the country.

MATERIALS AND METHODS

This study was carried out on the basis of corn consumption and production data from the period 1960 to 2010 collected from secondary source (Index Mundi, 2011). The data were model using Autoregressive Integrated Moving Average (ARIMA) stochastic model popularized by Box and Jenkins (1976). 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 process can be defined as:

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

where,

- y_t = Represents production or consumption
- μ = Mean of $\Delta^d y_t$

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

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

- ϕ_i = The i^{th} autoregressive parameter
- θ_i = The i^{th} moving average parameter

p, q and d denote the autoregressive, moving average and differenced order parameter of the process, respectively. Δ and B denote the difference and lag operators, respectively.

The estimation of the model consists of three steps, namely: identification, estimation of parameters and diagnostic checking.

Identification step: Identification step involves the use of the techniques to determine the values of p,q and d. The values are determined by using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). For any ARIMA (p, d, q) process, the theoretical PACF has non-zero partial autocorrelations at lags 1, 2, ..., p and has zero partial autocorrelations at all lags, while the theoretical ACF has non zero autocorrelation at lags 1, 2, ..., q and zero autocorrelations at all lags. The non-zero lags of the sample PACF and ACF are tentatively accepted as the p and q parameters. For a non stationary series the data is differenced to make the series stationary. The number of times the series is differenced determines the order of d. Thus, for a stationary data d = 0 and ARIMA (p, d, q) can be written as ARMA (p, q).

Estimation of parameters: The second step is the estimation of the model parameters for the tentative models that have been selected.

Diagnostic checking: The estimated model must be checked to verify if it adequately represents the series. The best model was selected based on the minimum values of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Normalized Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and highest value of R-square. Diagnostic checks are performed on the residuals to see if they are randomly and normally distributed. Here, the Kolmogorov-Smirnov test for normality was used. An overall check of the model adequacy was made using the modified Box-Pierce Q statistics. The test statistics is given by:

$$Q_m = n(n + 2) \sum_{k=1}^m (n - k)^{-1} r_k^2 \approx \chi_{m-r}^2$$

where,

- r_k^2 = The residuals autocorrelation at lag k
- n = The number of residuals
- m = The number of time lags included in the test

when the p-value associated with the Q is large the model is considered adequate, else the whole estimation process has to start again in order to get the most adequate model. Here all the tests were performed at the 5% level of significance.

RESULTS AND DISCUSSION

The maximum consumption of corn was 1650 thousand metric tons which occurred in the year 2010 and the minimum was 176 thousand metric tons which occurred in the year 1964. The maximum and the minimum production were 1676 thousand metric tons and 172 thousand metric tons respectively. The maximum production was achieved in the year 2010 and the minimum in the year 1983. In addition, the average production and consumption were 681 thousand metric tons and 695 thousand metric tons respectively.

The time series plot (Fig. 1) of consumption and production showed that the series was not stationary. The data was differenced to make it stationary. The first difference was enough to make the data stationary. As shown in Fig. 2, the differenced series fluctuates about the zero point indicating constant mean and variance which affirms that the series is stationary.

The rapid decay in the autocorrelation function and the partial autocorrelation function of the differenced series for consumption and production (Fig. 3 and 4) indicates that the series was stationary.

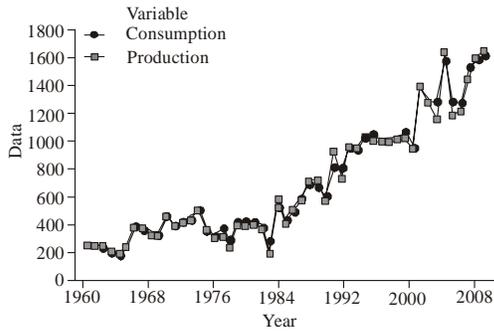


Fig. 1: Time series plot of consumption and production

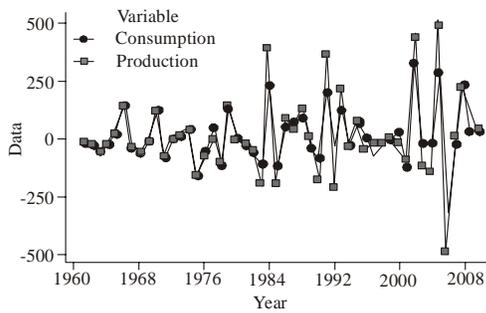
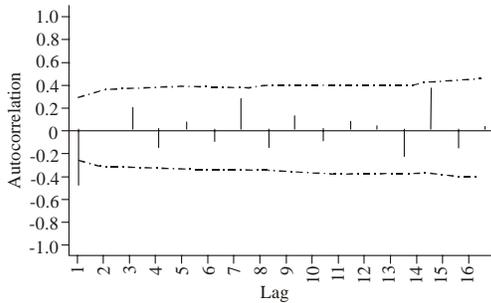
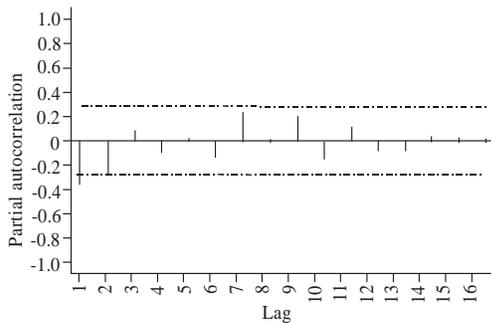


Fig. 2: Time series plot of the differenced series for consumption and production

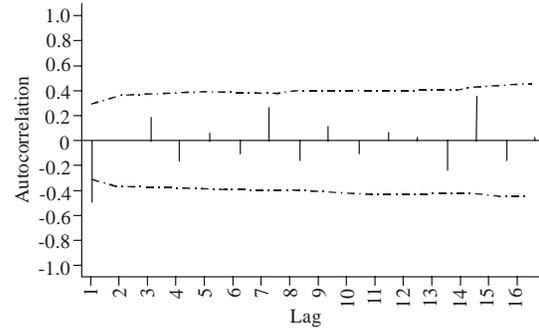


(a) ACF for consumption

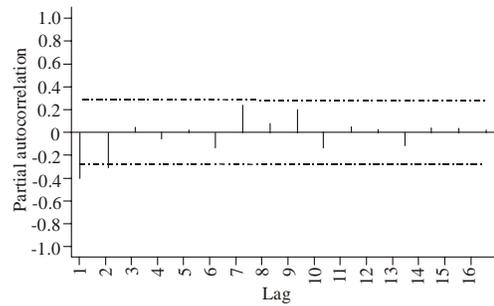


(b) PACF for consumption

Fig. 3: ACF and PACF of the differenced series for consumption



(a) ACF for production



(b) PACF for production

Fig. 4: ACF and PACF of the differenced series for production

To identify the orders of the ARIMA (p,d,q) model for production and consumption, the autocorrelation functions and partial autocorrelation functions were examined. The Augmented Dickey-Fuller test for the first differenced series shows that both the consumption and the production series were integrated of order one. For production, the autocorrelation function and the partial correlation function were examined and different models for production were fitted using different significant values of p and q. ARIMA (2, 1, 1) was selected as the best model for production based on the minimum values of RMSE, MAPE, BIC, AIC and highest R-square value as shown in Table 1. From Table 2, the estimates of the parameters of the ARIMA (2, 1, 1) model were highly significant.

Also, the ACF and PACF of the differenced series for consumption were examined and different models for consumption were fitted using different significant values of p and q. ARIMA (1, 1, 0) was identified as the best model for consumption based on the set selection criteria as shown in Table 3. As shown in Table 4, the estimate of the parameter of ARIMA (1, 1, 0) model was highly significant. The residuals for these models were diagnosed to see if they were normally distributed using the Kolmogorov-Smirnov normality test. The Kolmogorov-Smirnov statistics for the residuals of ARIMA (2, 1, 1)

Table 1: Suggested models for production

Model	R-square	RMSE	MAPE	BIC	AIC
ARIMA (0, 1, 1)	0.848	31.629	0.123	9.540	9.653
ARIMA (1, 1, 0)	0.861	32.810	0.148	9.617	9.154
ARIMA (1, 1, 1)	0.861	31.159	1.233	9.570	9.217
ARIMA (2, 1, 1)	0.861	31.143	0.268	9.570	9.195
ARIMA (2, 1, 1)*	0.871*	30.701*	0.036*	9.453*	9.051*

*: Means best based on that selection criterion

Table 2: ARIMA (2, 1, 1) model parameter estimates

Type	Coefficient	Standard error	t-value	p-value
AR 1	0.4788	0.136	3.52	0.001
AR 2	0.5189	0.1279	4.06	0.000
MA 1	0.9658	0.1143	8.45	0.000

Table 3: Suggested models for consumption

Model	R-square	RMSE	MAPE	BIC	AIC
ARIMA (0, 1, 1)	0.909	26.919	1.922	8.999	7.623
ARIMA (1, 1, 0)*	0.910*	21.629*	1.064*	8.992*	7.350*
ARIMA (1, 1, 1)	0.910	32.749	2.06	9.110	7.463

*: Means best based on that selection criterion

Table 4: ARIMA (1, 1, 0) model parameter estimates

Type	Coefficient	Standard error	t-value	p-value
AR 1	0.2888	0.1370	2.11	0.040

was 0.113 with an observed significance level of 0.106 indicating that the residuals for the model was normally distributed at the 5% (0.05) significance level. Also, for ARIMA (1, 1, 0) the Kolmogorov-Smirnov statistics was 0.132 with an observed significance level of 0.124 indicating that the residuals for this model was normally distributed at the 5% significance level. In addition, the modified Box-Pierce statistic was used to check the overall model adequacy for production and consumption. For production, the modified Box-Pierce statistic for lag 48 is 49.2 at 45 degrees of freedom which has observed significance level 0.307 indicating that it is insignificant at 5% significance level, hence the fit is good. Also, for consumption the modified Box-Pierce statistic for lag 48 is 37.2 at 47 degrees of freedom which has the observed significance level 0.523 indicating that it is insignificant at the 5% significance level. Hence, the diagnostic test indicates that ARIMA (2, 1, 1) and ARIMA (1, 1, 0) models were appropriate for production and consumption respectively.

Finally, ten years ahead forecast was made for production and consumption using ARIMA (2, 1, 1) and ARIMA (1, 1, 0), respectively. Table 5, shows the forecast for production and consumption at the 95% confidence limit. From Table 5, the forecast for the year 2011-2012 for consumption was 1635.56 thousand metric tons with a 95% confidence limit of (1413.57, 1857.55) thousand metric tons. In the same year the forecast for production was 1702.12 thousand metric tons with a 95% confidence limit of (1407.12, 1997.12) thousand metric tons. For the year 2020-2021, the forecast for consumption was 1685.97 thousand metric tons and that of production was

Table 5: Forecast for consumption and production of corn in thousand metric tons

Year	Consumption	Production
2011-2012	1635.56±221.99	1702.12±295.00
2012-2013	1639.73±272.40	1743.69±331.56
2013-2014	1648.87±324.51	1777.14±406.77
2014-2015	1655.77±366.93	1814.73±453.92
2015-2016	1657.80±405.43	1850.09±508.34
2016-2017	1663.79±440.56	1886.53±555.59
2017-2018	1668.12±472.94	1922.32±603.71
2018-2019	1670.62±503.33	1958.37±649.50
2019-2020	1673.32±531.00	1994.20±695.01
2020-2021	1685.97±559.28	2030.07±739.57

2030.07 thousand metric tons with 95% confidence limit of (1126.69, 2245.25) thousand metric tons and (1290.5, 2769.63) thousand metric tons respectively.

CONCLUSION

We employed the Box-Jenkins approach to model and forecast consumption and production of corn in Ghana. Our forecast showed an increasing pattern in consumption and production. Despite the increase in production, government has to still invest money into corn production, motivate corn farmers, and implement good policies for better land tenure systems for corn cultivation to ensure that production always exceeds consumption to avoid importation of corn into the country. This is imperative because importation could lead to high prices of corn and increase inflation rate, hence affecting the economy of the country.

REFERENCES

- Badmus, M.A. and O.S. Ariyo, 2011. Forecasting cultivated areas and production of maize in Nigeria using ARIMA model. *Asian J. Agric. Sci.*, 3(3): 171-176.
- Box, G.E.P. and G.M. Jenkins, 1976. *Time series Analysis, Forecasting and Control*. San Francisco, Holden Day, California, USA.
- Brown, R.G., 1959. *Statistical Forecasting for Inventory Control*. McGraw Hill Book Co., Inc., NY, USA.
- Falak, S. and A. Eatzaz, 2008. Forecasting Wheat production in Pakistan. *Lahore J. Econ.*, 3(1): 57-85.
- International Institute of Tropical Agriculture (IITA), 2009. *Maize Production, Consumption and Harvesting*. Retrieved from: www.iita.org, (Accessed on: 9th September, 2011).
- Ljung, G.M. and G.E.P. Box, 1978. On a measure of lack of fit in time series models. *Biometrika*, 65: 67-72.
- Muhammed, F., M. Siddique, M. Bashir and S. Ahamed, 1992. Forecasting rice production in Pakistan Using ARIMA models. *J. Animal Plant Sci.*, 2: 27-31.
- Pindyck, R.S. and D.L. Rubinfeld, 1981. *Econometric Models and Economic Forecasts*. McGraw Hill Book Co. Inc., NY, USA.

- Rachana, W., M. Suvarna and G. Sonal, 2010. Use of ARIMA models for forecasting pigeon pea production in India. *Int. Rev. Bus. Financ.*, 2(1): 97-107.
- Shabur, S.A. and M.E. Haque, 1993. Analysis of rice in Mymensing town market pattern and forecasting. *Bangladesh J. Agric. Econ.*, 16: 130-133.
- Sohail, A., A. Sarwar and M. Kamran, 1994. Forecasting total Food grains in Pakistan. *J. Eng. Appl. Sci.*, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, 13: 140-146.