AJAS / SPAS

African Journal of Applied Statistics
Vol. 5 (2), 2018, pages 503 – 517.
DOI: http://dx.doi.org/10.16929/ajas/503.227

ISSN 2316-0861

Bayesian inference approach in modeling and forecasting maize production in Rwanda

Denise Uwamariya[1] **and Denis Ndanguza**[1,*]

[1] Department of Mathematics, College of Science and Technology,
University of Rwanda, P.O. Box 3900 Kigali, Rwanda.

Received : September 10, 2018. Accepted : October 15, 2018. Published : October 25, 2018

Abstract. Rwanda is the country whose economy relies on agriculture. Therefore, forecast in agriculture sector is very important in Rwanda for future plan. In our study, secondary annual data from the agricultural ministry (MINAGRI), spanning from 1960 to 2014 have been used. In the analysis, appropriate model is selected based on the appearance of ACF and PACF of the transformed data. In addition to that, we use the fitted model to provide a four year forecasts of maize production from 2015 to 2018. Through Box–Jenkins methodology, the appropriate model is ARIMA (1,2,1) and fit the data at 91%. From the results and forecast, it is seen that the production of maize in Rwanda will have an increasing trend in the future. To strengthen the model, we also use the MCMC algorithm as an alternative method in parameters estimation. Diagnostics prove the chains' convergence which is the sign of an accurate model.

Key words: maize; time series model; Box–Jenkins methodology; forecast; MCMC method.
AMS 2010 Mathematics Subject Classification Objects : 62-07; 62H12; 65C05;65C40.

Presented by Prof Gane Samb Lo, UGB and AUST,
Senegal and Nigeria
Member of the Editorial Board.

* Corresponding Author: dndanguza@gmail.com
Denise Uwamariya : uwadeny85@gmail.com

Résumé (French) Le Rwanda est le pays dont l'économie repose sur l'agriculture. Par conséquent, les prévisions dans le secteur agricole sont très importantes son futur. Dans notre étude, les données annuelles secondaires du ministère de l'agriculure (MINAGRI), couvrant la période allant de 1960 à 2014, ont été utilisées. Dans notre analyse, un modèle approprié est sélectionné en fonction des allures des graphes ACF et PACF des données transformées. En plus de cela, nous utilisons le modèle ajusté pour fournir prévisions sur quatre ans de la production de maïs de 2015 à 2018. À travers la méthodologie Box - Jenkins, le modèle approprié est ARIMA(1,2,1) et s'ajuste aux données à 91%. D'après les résultats et les prévisions, il est établi que la production de maïs au Rwanda aura une tendance à la hausse à l'avenir. Pour renforcer le modèle, nous utilisons également l'algorithme MCMC comme méthode alternative pour l'estimation des paramètres. Les diagnostics prouvent la convergence des chaînes, ce qui montre l'efficacité de notre modèle.

1. Introduction

Maize also called corn is the third largest planted crop after wheat and rice all over the world and is one of the oldest human domesticated plants, in 7000 years ago maize was found in central Mexico and maize has been transformed into a better source by native Americans (Abdolreza, 2006; Ranum et al., 2014). All over the world in many countries after some years, maize has been spread out and become an major food crop grown in different zones and farming systems, and food consumption.

In Rwanda, maize was introduced in 1960 and has been considered as a priority stable crop by the government of Rwanda within the context of the national crop intensification program. Furthermore, Rwanda is exporting the maize production with high export market potential to Burundi and Eastern DRC.
Actually, time series is statistical methodology that can be applied to forecast the future values of the given series based on the current and past values of that series. Moreover, time series analysis can be applied in different field to solve any related problem which may arise. For Instance, time series is most applied in the field of economics to understand or to forecast the future values of one price per trading day or one price for crop per season. For different applications of time series refer to (Shumway and Stoffer, 2010; Brockwell and Davis, 2006; Miller and Hickman, 1973; Hamilton, 1994; Box et al., 2015; McCleary et al., 1980; Ahmed and Cook, 1979) among others.

Forecasting has long been in existence and continues to receive extensive attention in the literature. Different authors such as (M.A.Sarika and Chattopadhyay, 2011; Irfan et al., 2011; Boken, 2000) among others provide the definition of forecasting in the literature based on the environment in which it is applied. Forecasting has been evolving over the years and saw many methods being established and some being developed.

Researchers have used methods of forecasting with time series data in different areas. For instance, M.A.Sarika and Chattopadhyay (2011) used ARIMA (Autoregressive Integreted Moving Average model) to model and forecast time series data of pigeonpia production in India from 1969 to 2007. In their study they used Root mean square, Akaike Information Criterion and Bayesian Criterion to identify the best model. As result they found that ARIMA (2,1,0) model was the best among other models of ARIMA family, for modeling as well as forecasting purpose (M.A.Sarika and Chattopadhyay, 2011).
Irfan et al. (2011), conducted a research where their main objective was to develop a suitable model and then forecast the yield rice in the four provinces of Pakistan during the period from 1947 to 1948 and from 2008 to 2009 using ARCH family models. They only discussed the ARMA methodology rather than the ARCH family models. As a result, on the basis of two criteria AIC (Akaike information criteria) and SIC (Schwartz information criteria), the GARCH model was selected for all provinces and the forecasting was also perfect (Irfan et al., 2011).

Amin et al. (2014) developed a quite number of time series models and suggested the best to forecast the future value of the wheat production of Pakistan in the coming years. Secondary data from 1902 to 2005 have been used for analysis. They have fitted different time series models to the data using JMP and Statgraphics statistical software for analysis. Based on different criteria and model adequacy, ARIMA(1, 2, 2) was selected to be the best (Amin et al., 2014).

Boken (2000) used time series techniques on the past yield data to forecast the future data for wheat yield estimation applied weather data over the growing season. In his paper, a few relevant techniques are tested to model the average spring wheat yield series for Saskatchewan, Canada. Using spring wheat series ($1975-1993, 1975-1994$ and $1975-1995$) the series were forecasted for $1994, 1995$ and 1996, respectively. More discussion have been made to provide the improvement of the forecasting by dividing the heterogeneous cropping region of Saskatchewan into rather yield-based homogeneous region by using spatial analysis tool of geographic information system software (Boken, 2000).

Based on the fact that maize is the main agriculture crop in Rwanda which plays a remarkable role in food security for both urban and rural population, an appropriate model to be used by decision makers for forecasting and planning is crucial. Therefore, the aim is to build such time series model to be used in forecasting maize production in Rwanda. Moreover, we refer to the Box-Jenkins methodology to identify, to estimate parameters and to forecast future production. Secondary data to be used are collected from Ministry of Agriculture and Animal resources. For model identification, ACF and PACF plots for available series of maize production in Rwanda are used to choose the appropriate model after data transformation (Adhikari and Agrawal, 2013). Finally, the identified unknown parameters are used to forecast the future values of maize production which may help the Rwandan decision makers to keep or increase its production in the future. To strengthen the accuracy of the model, we use the Markov chain

Monte carlo method in parameters estimation and test the convergence diagnostics.

The Bayesian inference philosophy is to regard the model parameters as random variables while the object of interest is to focus on the posterior distribution of the parameters given the data. In addition, the unobservable parameters are handled probabilistically, while the observed data are considered deterministically (Ndanguza, 2015).
The posterior density $p(\theta|y)$, where θ is the unknown parameter to be estimated, y the measurements or observations is defined through Bayes' formula as the normalised product of the prior density and the likelihood. The Bayes' formula is defined as follows (Ndanguza, 2015)

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{1}$$

where $p(y|\theta)$ is the likelihood function, $p(\theta)$ is the prior distribution and $p(y)$ is the normalizing constant. Normally, the prior distribution encloses all knowledge about the parameter values vacant to the researcher before the consideration of data. Usually, researchers use non-informative parameters so that the amount of prior knowledge incorporated in the analysis is small (O'Neill, 2002). The normalizing constant is defined as

$$p(y) = \int_{\theta} p(y|\theta)p(\theta)d\theta. \tag{2}$$

When models are complex or multidimensional, it is challenging to compute the posterior distribution since solving analytically (2) is tough (Gelman et al., 1996). If this occurs, one of the ways to handle it is to use Markov chain Monte Carlo. MCMC method is a general method established to draw values of θ from approximate distributions and then correcting those draws to better approximate target posterior distribution, $p(\theta|y)$ (Ndanguza, 2015).
Normally, the posterior distribution is used while computing many statistics like mean, moments and quantiles (Robert and Casella, 2004). To illustrate, given the function $p(\theta)$, the posterior expectation of $p(\theta)$ is

$$\mathbb{E}[p(\theta|y)] = \frac{\int p(\theta)p(y|\theta)p(\theta)d\theta}{\int p(y|\theta)p(\theta)d\theta}. \tag{3}$$

The MCMC method has been widely used by (Geyer, 1992; Brooks, 1998; Ndanguza, 2015) among many others.
The rest of the paper is as follows: Section 2, is the Box-Jenkins methodology, followed by Section 3, which is the estimation of parameters using MCMC method and convergence testing. The last Section is the conclusion.

2. Box–Jenkins methodology

2.1. Model Selection

We select an appropriate Box and Jerkins model to forecast the Rwanda maize production. The first step in developing a Box and Jerkins model is to determine if the series is stationary and this is done by using time plot of the available series. To capture the main features of the graph, we check in particular whether there is a trend, a seasonal component, any apparent sharp changes in behavior, any outlying observations. To achieve stationarity of series, difference transformation have been used to remove the trend. Once stationarity has been addressed, the next step is to identify the order of the Autoregressive (AR) and Moving Average (MA) terms corresponding to order of p and q. The primary tools for doing this is ACF and PACF plot.

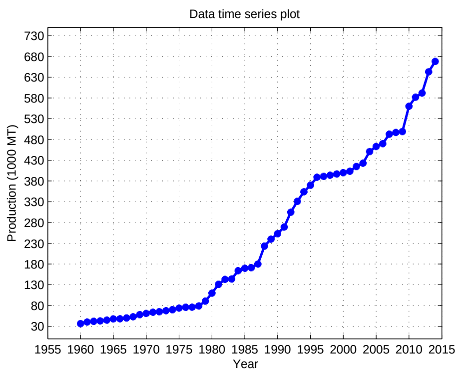


Fig. 1: Time plot of original data of Rwanda maize production.

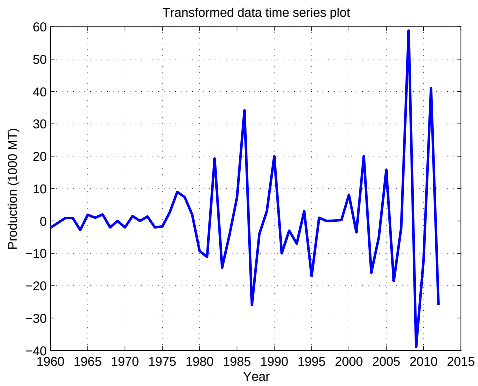


Fig. 2: Time plot of transformed data of Rwanda maize production.

Figure 1 shows that data are not stationary in mean and in variance. The series shows an increasing trend. Therefore, to make the series of data stationary, difference of order two have been carried out and series became stationary with respect to mean and variance as shown by Figure 2.
To identify the order of the model, ACF and PACF plots are used. The order of AR is 1 since there is one significant spike compared to others which are out of the boundaries on the ACF plot. The same procedure is followed to identify the order of MA and the Partial Auto correlation has one significant spike indicating that the order of Autoregressive in the model is 1.
We therefore suggest two combined models namely AR (1) and MA (1). Thus, since the order of difference is 2, the potential model for the transformed data is ARIMA (1, 2, 1) and the model is suggested to be

$$y_t = \mu + \rho y_{t-1} + \psi \epsilon_{t-1} + \epsilon_t, \qquad (4)$$

where μ is the constant, ρ the AR coefficient at lag 1, ψ, the MA coefficient at lag 1, ϵ_{t-1}, the forecast error that was made at period $t-1$ and ϵ_t, errors at time t. Parameters to be estimated are μ, ρ and ψ.

2.2. Estimation of parameters using Least square method

Parameters described in Equation (4) are estimated using data captured in Figure 1 by least-squares fit. This consists of minimizing the likelihood function which is the residual sum of squares (RSS).

$$RSS = \sum_{i=1}^{n} (Y_i - f(Y_i, \theta))^2,$$

where Y_i are data, $f(Y_i, \theta)$ stands for the model and θ the parameters to be estimated.
Estimated parameters are explained in Table 1.

Table 1: LSQ estimates of fitted Model

Parameters	Estimate	Standard Error	t. value	sig.
Constant (μ)	3.88	.001	-.734	.047
AR lag 1 (ρ)	1.03	.150	1.951	.037
Difference	2			
MA lag1 (ψ)	-0.106	.308	3.211	.002

Based on the estimated coefficients and constant for AR and MA terms found in Table 1, the actual fitted model is written as follows

$$\widehat{Y}_t = 3.88 + 1.03 y_{t-1} - 0.106 \varepsilon_{t-1} + \epsilon_t \tag{5}$$

2.3. Model checking

From Table 1, the SE (standard error) shows that there is no big difference between estimates and true values meaning that Equation (5) can be used. In addition to this, the table shows the tests value for coefficients estimates which are t-test and significance values in column 4 and 5 respectively. By using significance values, we can say that all coefficients are significant since their significance values are less than 0.05 p–value. Now, this model can be used to predict the quantity of maize that will be produced in the future. Usually, before using the model to predict, it is important to test whether the model is adequate or not. To test this, we use forecast errors by plotting its time plot, histogram and ACF and finally we use the multiple plot of actual values, predicted values, upper and lower limit.

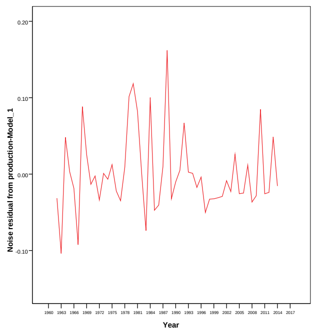


Fig. 3: Time plot of forecast error of Rwanda maize production.

The plot given in Figure 3 helps to investigate the behavior of errors and indicates whether the forecast errors have a regular pattern or not. From this plot, the forecast errors fluctuate around approximately constant mean with roughly constant variance and there is no obvious pattern in their curve.

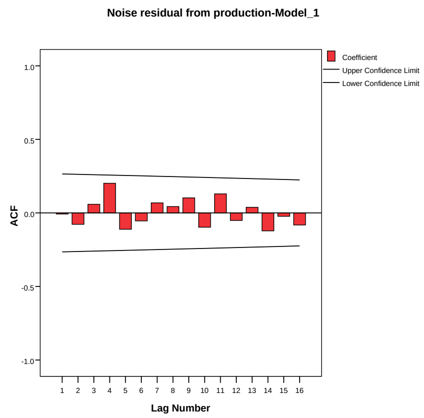


Fig. 4: Autocorrelation function of forecast errors of Rwanda maize production.

From the Auto Correlogram Function of forecast errors shown in Figure 4, all sample autocorrelation lie within confidence limits, so there is no evidence against the null hypothesis of zero auto correlation at lags 1 to 16.

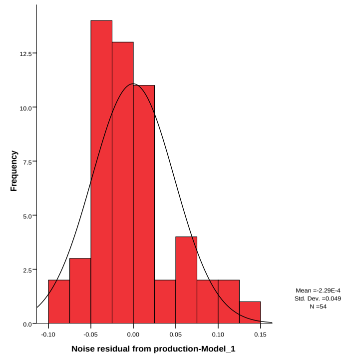


Fig. 5: Histogram of errors of Rwanda maize production.

From the histogram of errors shown in the Figure 5, we can easily see that errors are approximately normally distributed with mean -0.00029 and standard deviation 0.049.
Even if the model fits the generated data, the coefficient of determination is computed to ascertain the level as below

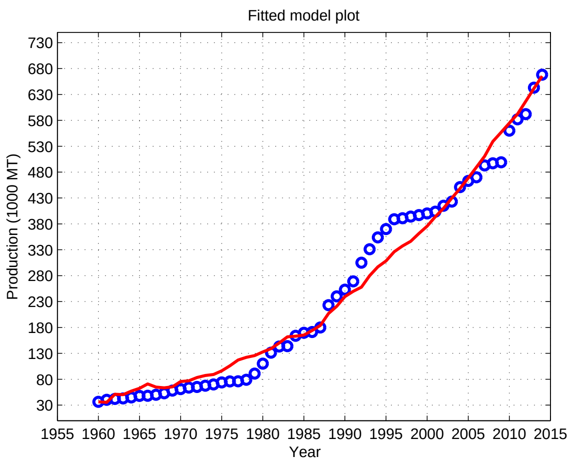
$$R^2 = 1 - \frac{\sum(Y_i - f(Y_i, \hat{\theta}))^2}{\sum(Y_i - \bar{Y}_i)^2},$$

where $\hat{\theta}$ and $\bar{Y}_i$ are estimated parameters and mean of data respectively.

Table 2: Model Statistics

Model Fits statistics	Ljung Box Q(18)			
R square	statistics	DF	Sig.	Number of outliers
.910	15.847	16	.046	0

The fitness of the model is captured in Figure 6 and it is clear that the model fits



the data.

Fig. 6: The estimated model vs real data for fitness checking.

2.4. Forecasting

From the table 2, showing model statistics, it is seen that R^2 for the fitted model is 0.910 and this means that our model explains the data at 91% and it can be said that the fitted model is good and can be used to predict maize production in Rwanda.

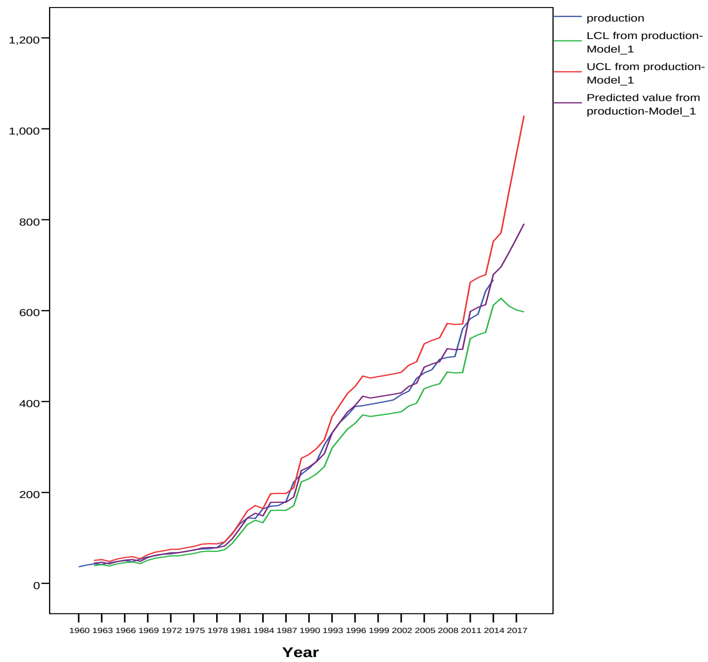


Fig. 7: Multiple plot of Rwanda maize production.

From the Figure 7, we can easily see that the period in which data are available, means, from 1960 up to 2014, the prediction intervals for the 1 step ahead forecasts are quite narrow. After this period, the forecasts are increasingly uncertain. This is due to the fact that the prediction intervals become wide as the time increases. From all those information, the $ARIMA(1, 2, 1)$ model seems to be adequate and can be used to predict for a short period of time.

Table 3: Predicted Values, Upper and Lower confidence interval for Maize Production in Rwanda from 2015 to 2018

Model		2015	2016	2017	2018
production model 1	forecast	696.23	726.52	758.23	791.15
	UCL	771.05	858.55	943.84	1028.92
	LCL	627.01	610.37	601.51	597.28

Table 3 reveals that maize production will increase as time increase with 696.23 (1000 MT) in the 2015 and 791.15 (1000 MT) in the 2018. Since our results show the reality on the ground, we go a little bite far to test the model by including bias in it and produce the chain of posteriors. Chains will be produced by the MCMC algorithms.

3. Markov chain Monte Carlo method

Markov chain Monte Carlo (MCMC) methods are numerical methods for computing multidimensional integrals of the above form by using Monte Carlo. The objective is to draw samples $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)}$ from the posterior distribution $p(\boldsymbol{\theta} \,|\, \mathbf{y}_1, \dots, \mathbf{y}_M)$ and approximate the expectation as the sample average

$$\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}) \,|\, \mathbf{y}_1, \dots, \mathbf{y}_M] \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}(\boldsymbol{\theta}^{(i)}). \tag{6}$$

The construction of the Markov chain depends upon choosing a density $q(.,.)$, considered as proposal density, with objective to suggest the next possible element of the chain. A suggested sample is either accepted to be an element of the chain or not and this is performed by the Metropolis-Hastings algorithm (Ndanguza, 2015). The objective function (likelihood function) $p(\theta|y)$ is computed as sum of squares of residuals and therefore, used in MCMC method (Ndanguza, 2015). We estimate the posteriors and results are found in Table 4.

Table 4: MCMC results for parameters estimation

Parameters	Initial value	Mean Posteriors	Median Posteriors	Standard deviation
μ	3.88	3.888	3.7442	0.26401
ρ	1.04	1.0319	1.0526	0.0008
ψ	-0.15	-0.15869	-0.17221	0.0152

After the computation of chain of posteriors, it is necessary to assess through some diagnostics whether the Markov chain has converged to its stationary distribution (Brooks and Roberts, 1998). We use the plot method for an MCMC object to produce a trace plot and a density plot for each parameter.

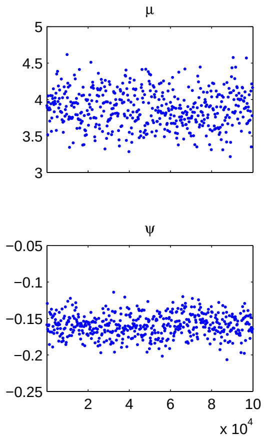


Fig. 8: Trace plot representing MCMC chains of parameters which displays the values the parameters have taken during the runtime of the chain.

Figure 8 shows that there is an almost perfect mixing because samples move from one region to another in 1 step. This is one of the most used ways to conclude that the stationarity has achieved. One can simply visualize the state of the chain through "time" (iterations, sometimes called generations). It is illustrative to plot each posterior as a function of iteration number to obtain a time series plot (Brooks and Roberts, 1998).
Figure 9 visualizes the marginal posterior distributions or some times called marginal density plot. To smooth the distribution it is recommended to use a kernel density estimate of the posterior (Gelman et al., 1996).

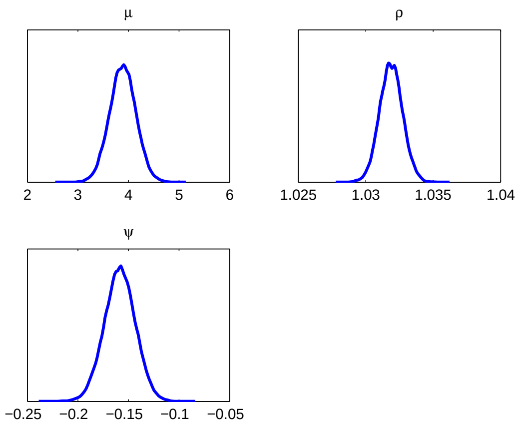


Fig. 9: Kernel density plots of the parameters (μ, ρ and ψ).

Generally, the trace-plot of the values of the parameter in the chain is changed into the (smoothed) histogram. We also compute the marginal densities which are an average a parameter takes with all other marginalized parameters. From Figure 9, all the distributions are gaussian which is a sign of a positive test of stationarity and convergence.

3.1. Pairs of posteriors

The purpose is to check whether there is a strong correlation among the scattered ones. We have a collection of points displaying the data in two variables. One variable is determining the x-axis and the other variable value determining the y-axis. We also call his kind of plot a scatter chart, scattergram, scatter diagram, or scatter graph. We sketch the two plots in Figure 10, where the scatter of two by two parameters is shown. The purpose is to check whether there is a strong correlation among the scattered ones.

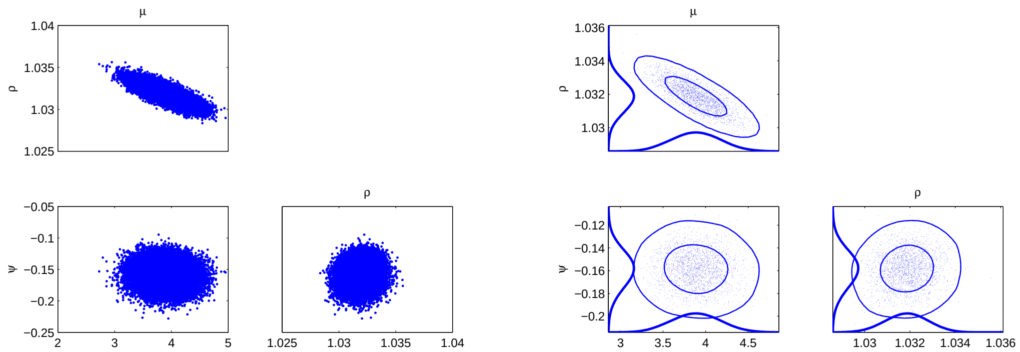


Fig. 10: 2D scatter plots of estimated parameters.

Since the poor mixing corresponds with a high correlation, there is a need of larger sample size for a suitable comparison of variance. From Figure 10, parameters are not correlated each other. As long as this happens, we conclude that there is a good mixing.

3.2. Posterior histogram plots

We plot the density of data in form of histograms captured in Figure 11 and we estimate the probability density function of the underlying parameter.

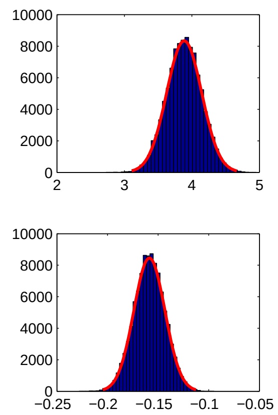


Fig. 11: Histograms plot for the estimated model parameters.

We normalize to 1 the total area of a histogram used for probability density. The judgement is based on the length of the intervals on the x-axis whether they are all 1. If yes, the histogram is equivalent to a relative frequency plot. Figure 11 highlights that all the parameters' distributions are almost gaussian, which is the evidence of good mixing.

3.3. Diagnostic using autocorrelation plot

Usually, the expectation is that the n^{th} lag autocorrelation has to be smaller as n increases (our 2^{nd} and 40^{th} draws have to be less correlated than our 2^{nd} and 5^{th} draws) (Ndanguza, 2015). This is justified by Figure 12.

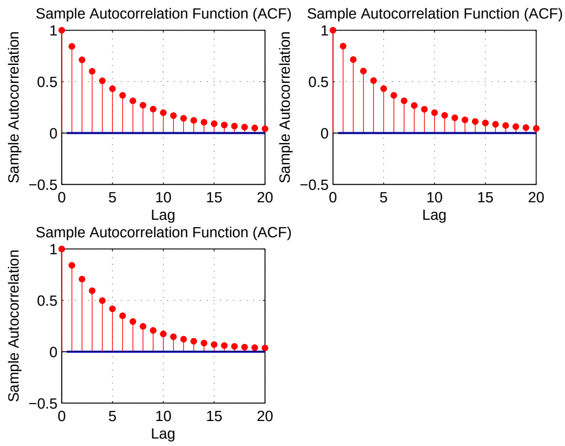


Fig. 12: The autocorrelation functions of estimated parameters with 20 lags.

From Figure 12, it is visible that the autocorrelation coefficients located on x-axis downfall near zero. After reaching zero, they stabilize as soon as the number of lags located on y-axis increases. This behaviour ascertains a good mixing and also proves the stationarity of sampled chains.

4. Conclusion

This paper aims at providing a time series model and forecasting the maize production of Rwanda. Using the Box–Jenkins methodology, we found that the appropriate model is
$ARIMA(1, 2, 1)$ with fitted equation of $\widehat{Y}_t = 3.88 + 1.03y_{t-1} - 0.106\varepsilon_{t-1} + \epsilon_t$. The model has been assessed and diagnosed using the MCMC method. The model forecast proves that Rwanda maize production will increase over time in the coming years. Furthermore, the estimates got by least square method are in agreement with those obtained by MCMC method. We can definitely argue that the model is accurate and can be applied by any organization in decision making. The model fits the data at a high percentage of coefficient of determination of 91% which is the sign of model accuracy.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Abdolreza, A. (2006). Maize international market profile. *Background Paper for the Competitive Commercial Agriculture in Sub-sahara African (CCAA) Study*.

Adhikari, R. and Agrawal, R. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.

Ahmed, M. S. and Cook, A. R. (1979). *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*. Number 722.

Amin, M., Amanullah, M., Akbar, A., et al. (2014). Time series modeling for forecasting wheat production of Pakistan. *The Journal of Animal & Plant Sciences*, 24(5):1444–1451.

Boken, V. K. (2000). Forecasting spring wheat yield using time series analysis. *Agronomy Journal*, 92(6):1047–1053.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brockwell, P. J. and Davis, R. A. (2006). *Introduction to time series and forecasting*. Springer Science & Business Media.

Brooks, S. (1998). Markov chain Monte Carlo method and its application. *Journal of the royal statistical society: series D (the Statistician)*, 47(1):69–100.

Brooks, S. P. and Roberts, G. O. (1998). Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing*, 8(4):319–335.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1996). *Bayesian Data Analysis*. Chapman and Hall, London, 2nd edition.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483.

Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.

Irfan, M., Irfan, M., and Tahir, M. (2011). Modeling the province wise yield of rice crop in Pakistan using GARCH model. *International Journal of Science and Technology*, 1(6).

M.A.Sarika, I. and Chattopadhyay, C. (2011). Modelling and forecasting of pigeonpea (cajanus cajan) production using autoregressive integrated moving average methodology. *Indian Journal of Agricultural Sciences*, 81(6):520–3.

McCleary, R., Hay, R. A., Meidinger, E. E., and McDowall, D. (1980). *Applied time series analysis for the social sciences*. Sage Publications Beverly Hills, CA.

Miller, R. B. and Hickman, J. C. (1973). Time series analysis and forecasting. *Transactions of the Society of Actuaries*, 25:267–302.

Ndanguza, R. D. (2015). Bayesian analysis of SEIR epidemic models. *Acta Universitatis Lappeenrantaensis*.

O'Neill, P. D. (2002). A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Mathematical biosciences*, 180(1):103–114.

Ranum, P., Peña-Rosas, J. P., and Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1):105–112.

Robert, P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York, USA: Springer-Verlag.

Shumway, R. H. and Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.