Available online at https://doi.org/10.5281/zenodo.7193147 Research Article

DIUAJ Daha International University Academic Journal

Time Series Analysis of Rice Prices using Box-Jenkins ARIMA Methodology: Case Study Hargeisa-Somaliland

Abdishakur Ismail Adam[1]

Received, 1 September 2022 *Accepted, 3 October 2022* *Available Online, 13 October 2022*

1 Department of Statistics, Ondokuz Mayıs University, Turkey,* cabsha1994@gmail.com

ABSTRACT

The international prices of agricultural commodities have been increasing considerably. This upward trend, which may cause a new food crisis, has attracted the attention of the world. Several explanations for these movements in prices have been provided by analysts, researchers, and development institutions. The main purpose of this study was to determine and get forecasts of rice prices in Hargeisa, Somaliland by using Box-Jenkins ARIMA modeling. Rice prices in Hargeisa were examined in order to identify if it is stationary or not. In order to check if it is stationary, we have used time series plot, correlograms and done Augmented Dickey-Fuller test. The results revealed that the data is non-stationary. We have used some approaches such as taking differences to make the data stationary. After getting it stationary we have determined some time series Box-Jenkins models as candidates. After that the determined models were compared with respect to the model accuracy criteria such as Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). Then we have found the best model fitted well to the data set. After doing diagnostic checking we have calculated the forecasts. And all of the results of whole analysis were presented. The outcome of this study can aid both Somaliland government and policy makers in making optimal production decisions and in managing overall price risks

Keywords: Agricultural Commodities, Trend, Box-Jenkins ARIMA modeling, Stationary, Hargeisa

1. INTRODUCTION

Global food prices have recently become an increasingly important international concern with the occurrence of the food price crisis of 2007-2008. The Food and Agriculture Organization (FAO) food price index, which measures the international prices of meat, dairy, cereals, oils and fats, and sugar, climbed from 127 in 2006 to 159 in 2007 to 200 in[1].

Continuing a decade-long increase, global food prices rose 2.7 percent in 2012, reaching levels not seen since the 1960s and 1970s. However, still well below the price spike of 1974. Between 2000 and 2012, the World Bank global food price index increased 104.5 percent, at an average annual rate of 6.5 percent. Food price volatility has increased dramatically since 2006. European Union, (2010). According to the United Nations, Food and Agriculture Organization the standard deviation—or measurement of variation from the average—for food prices between 1990 and 1999 was 7.7 index points, but it increased to 22.4 index points in the 2000–12 period [2].

WFP (2015) reported that Africa is facing its worst food crisis in years, as El Nino rages across the continent [3]. New analysis from *Mail & Guardian Africa* collating data from the UN, the Famine Early Warning System Network (FEWS Net) and various news agencies reveals that more than 40 million people in Africa are facing food insecurity and starvation [4].

The continent needs at least $4.5 billion for emergency relief, but just a fraction of that has been raised so far, even as analysis from Oxfam shows that an early response is far cheaper than a late one. Also, the report states that The Horn of Africa particularly Ethiopia and much of southern Africa is in bad straits and the weather is not the only factor at play. A country's ability to cope depends partly on its public finances and ability to mobilize resources; for some weakening currencies make food imports more expensive, and conflict is making it difficult to move supplies around [4].

As Somaliland is facing an upward pressure on food prices in the country with deteriorating terms of trade and a larger food import bill to pay. Somaliland is totally dependent on imports of sugar, rice, spaghetti, wheat flavor and vegetable oil, Somaliland does not produce rice and does not have the climatic conditions to grow wheat. Because of converging trends including water scarcity, land degradation, lack of on-farm technological innovation, population growth, and climate change [5]. Barrett, Christopher (1996), Much food supply exists in Somaliland and this resulted shock and surprise Food price rates in the country[6]. Hargeisa is the second largest city of Somaliland also it's the second most populous city of whole the country and still there is a high food insecurity and high food price rates. As Ministry of National Planning and Development reported in 2011 The people most likely to be negatively affected are urban poor, disabled people, elderly people, children and women, the unemployed and IDPs [5].

Time series analysis is a statistical method which is used forecasting for the future based on the events of past and present. It comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data[7].

In this report, we applied the principles of Box-Jenkins methodology to secondary data comprised monthly Food prices (particularly Rice price) from November 2013 to April 2016, Hargeisa, Somaliland, which was obtained from the Ministry of National Planning and Development (MoNPD) especially Department of Statistics.

The study seeks to model, validate and forecast the Hargeisa monthly food prices and highlight the trend analysis followed by Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) to forecast Hargeisa Food Prices for the next year Apr, 2017. The findings of the study could serve as a guide for a review and help assess the current interventions to curb the high food price rates of Somaliland.

2. MATERIALS AND METHODS

2.1. Box-Jenkins Methodology

Box-Jenkins methodology (named after the statisticians George Box and Gwilym Jenkins) is a statistical procedure that is used to model time series data by using autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models [8]. The Box - Jenkins Analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models. The model is generally referred to as an ARIMA (p, d, q) model where p, d and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. The study used secondary monthly data, collected from MoNPD, Hargeisa, Somaliland which covered the period of 4 years, from 2013 to 2017. The data were modelled using ARIMA models. The ARIMA models help to fit datasets that have time series structure to describe the trend of food prices and forecast points ahead. It also provides a forecast interval and it is based on a proven model [9].

2.2. The ARIMA Model

The basic processes of the Box–Jenkins ARIMA (*p*, *d*, *q*) model include the autoregressive process, the integrated process, and the moving average process. A dataset Y_t follows ARIMA model if the d^{th} differences $\nabla^d Y_t$ follows a stationary ARMA model. The parameters that help build the ARIMA model are three; *p*, which determines the AR order; *d*, denotes the number of differencing required before stationarity, and MA order is given by *q* [10].

Hence, ARIMA (*p*, *d*, *q*) is represented in a general form according to Tebbs J.M [10] as;

$$\emptyset(B)(1-B)^d Y_t = \theta(B)e_t \tag{1}$$

Where, the AR and MA characteristic operators are

$$\emptyset(B) = 1 - \emptyset_1(B) - \emptyset_2(B)^2 - \cdots - \emptyset_p(B)^p \tag{2}$$

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B)^2 - \cdots - \theta_p(B)^p \tag{3}$$

And

$$(1-B)^d Y_t = \nabla^d Y_t \tag{4}$$

Where, $\emptyset$ is the autoregressive parameter to be estimated; θ is the moving average parameter to be estimated; ∇, the difference operator; B, the backward shift operator; e_t random process having zero mean and variance. Box and Jenkins [8] proposed the estimation of parameters of ARIMA model, and their approach involves the steps: identification of ARIMA model, model parameter estimation, and model diagnostics [10].

2.3. Unit Root Test

In order to make inferences in time series analysis, it is necessary to determine whether the time series is stationary or not. This study we used the Augmented Dickey Fuller (ADF) test for assessing stationarity of the dataset [11]. The test assumes that y_t follows a randomness in the time series data:

$$Y_t = \rho Y_{t-1} + e_t \tag{5}$$

Where, ρ, the characteristic root of an AR polynomial and e_t is white noise with mean zero and variance σ^2 [10]. The ADF test helps to test the null hypothesis of non-stationarity in the data. This results in the following ADF unit root test: H_0: $\delta = 0$ (non-stationary) versus H_1: $\delta < 0$ (stationary) [12].

2.4. Identification of ARIMA Model

There are techniques under ARIMA model identification which estimate the p, q and d values. The autocorrelation function (ACF) and partial autocorrelation function (PACF) help to determine the p, q and d values. The theoretical PACF of ARIMA (p, q, d) process usually show non-zero PACF at first p lags, with remaining lags having zero PACF. The first q lags also report non-zero ACF and the remaining lags having zero ACF for the theoretical ACF. We determine q and p by the total frequency of the significant lags which are not zero for ACF and PACF respectively. If the values of p, d, and q are inaccurately selected, models derived can be inadequate, hence cannot be used for predictions [10].

2.5. Estimation of Model Parameters, Model Diagnostic and Validation

If the ARIMA model is identified, then we can estimate the parameters of the model using EViews, Model estimation is followed by model selection, and it is done by considering minimum values of Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) [10].

$$AIC = -2\ln(\bar{L}) + 2h \qquad (6)$$

and $$BIC = -2\ln(\bar{L}) + \ln(n)h \qquad (7)$$

where, L is the likelihood value of the likelihood function, h and n are number of parameters to be estimated and number of residuals respectively. For any two competing models, the model with the minimum AIC or BIC will be selected as a better one. After fitting the model, we will be checked whether the model is appropriate or not, or we investigated how the model fits the data. Therefore, the Normality plot and ACF and PACF plots of Residuals are convenient graphical techniques for model diagnostic we used. To validate the model selected, the dataset was modelled using a training set which comprised of data from Nov 2013 to Feb 2017 and validated using a testing sample from Mar 2017 to Dec 2017. The validation measures included mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) will be used [13].

3. RESULTS

3.1 Data Handling

The data for this study is a Hargeisa monthly rice prices (Nov 2013 – Dec 2017) from Ministry of National Planning and Development, Hargeisa, Somaliland. The time series plot below in Figure 1 shows Hargeisa monthly rice prices from November 2013 to December 2017.

The time series plot of our data shows that there is a deterministic trend which provides that the time series is not stationary.

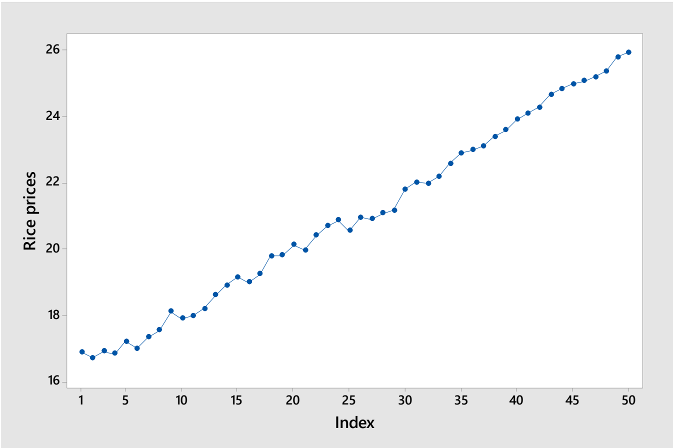

Figure 1. Time series plot of Hargeisa monthly rice prices, Hargeisa, Somaliland.

3.2 Testing Stationarity

It is clear there was an increasing trend in our original data which clarifies that the series is non-Stationary. In addition to the graphical approach for testing stationarity, it is also necessary to test with statistical testing. One of these approaches is the Unit Root Test which has been widely used in recent years. Adedia D, Nanga, S, et al (2018).

The Augmented Dickey Fuller (ADF) test were applied for assessing stationarity of the dataset. The results of no differenced data series for ADF tests (constant, constant & trend and none) revealed that the data series in non-stationary since the absolute of the ADF test statistic is less than ADF critical values 1%,5% and 10%, however ADF tests confirmed stationarity after the data series was differenced of the first order as can be seen in Table 1 and Since the data series is non-Stationary we should make the data series stationary before estimation of the best model fit using first difference method.

Table 1: The results of the Unit Root test (ADF)

ADF TEST	**Differencing Order**	**ADF 1% critical value**	**ADF 5% critical value**	**ADF 10% critical value**	**ADF test statistic**	**Decision**
No constant and no trend	0	-2.6162	-1.9481	-1.6123	-1.2089	Accept H_o
Constant and no trend	0	-3.5811	-2.9266	-2.6014	1.1879	Accept H_o
Constant and trend	0	-4.1705	-3.5107	-3.1855	-2.1075	Accept H_o
No constant and no trend	1	-2.6198	-1.9486	-1.612	-6.2385	Reject H_o
Constant and no trend	1	-3.5811	-2.9266	-2.6014	-7.7084	Reject H_o
Constant and trend	1	-4.1705	-3.5107	-3.1855	-7.8762	Reject H_o

We recognized below Figure 2 clearly that First-differenced time series is *stationary* (constant mean and approximately constant variance) and now since the data series is stationary, we can study its behavior, calculate the best fit model the data series using the Box-Jenkins Methodology (ARIMA Modelling).

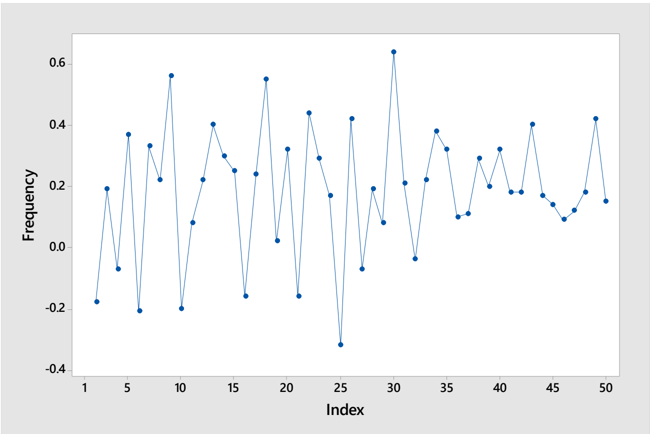

Figure 2: Time series plot of differenced Hargeisa rice prices.

3.3 Model Identification

The output in **Table 1** shows that after the first difference the dataset became stationary. By using the spikes in the ACF and the PACF plot of the differenced data of the first order, we suggest both the q and p values. Figure 2 shows the ACF and PACF plots. The ACF plot has spikes (significant lags) at lags 1, lag 3 and 4, which is the moving average (MA)part to the model and the PACF plot has spikes for lags 1and 3 which shows the autoregressive part (AR) to the model.

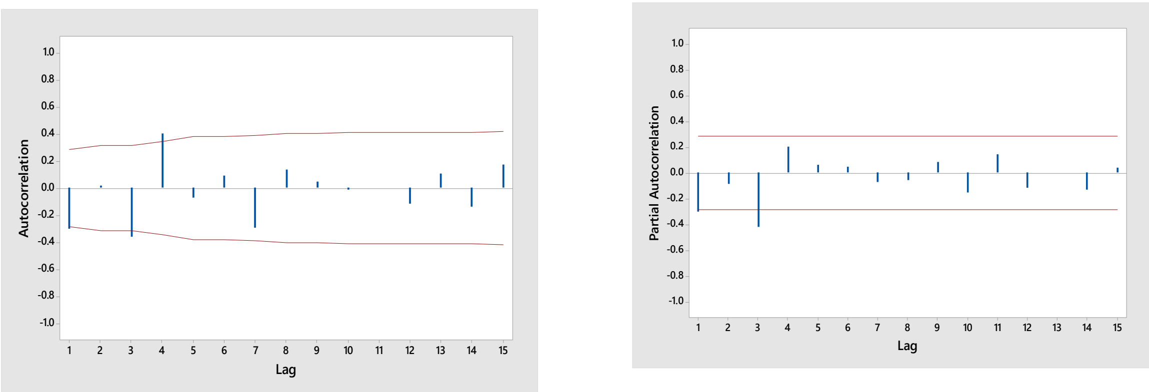

Figure 3: ACF and PACF of first differenced Hargeisa rice prices

Therefore, models were tentatively suggested based on the combination of the significant spikes in both the ACF and PACF plots (Figure 2), and through Box-Jenkins approach the best model was selected as the best. ARIMA (1, 1, 1* (Row 1)) in Table 2 is the best model because it is the model with the least AIC and BIC values. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994).

Table 2. The fitted ARIMA (p, 1, q) models

No	**Models**	**AIC**	**BIC**
*1**	*ARIMA (1,1,1) **	*-0.8556**	*-0.72631**
2	ARIMA (1,1,3)	-0.53888	-0.3234
3	ARIMA (1,1,4)	-0.25738	0.001188
4	ARIMA (3,1,1)	-0.36045	-0.14051
5	ARIMA (3,1,3)	-0.57886	-0.27095
6	ARIMA (3,1,4)	-0.41534	-0.06345

3.4 Model Estimation

Once we indicated the best fitted model for the data series, we estimated the parameters of chosen model ARIMA (1, 1, 1) as Table 3 below shows

Table 3: Model parameters estimates of ARIMA (1, 1, 1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.191119	0.014141	13.51519	0.0000
AR(1)	0.378571	0.069416	5.453648	0.0000
MA(1)	-1.34053	0.193358	-6.93289	0.0000

Therefore, our model will be as following:

$$\hat{y}_t = 0.1911 + 0.3786y_{t-1} - 1.3405\varepsilon_{t-1} \quad (8)$$

3.5 Model Diagnostic

After fitting the model, we should check whether the selected model is appropriate or not, or in other words we study how the model fits our data. Therefore, the *normal probability plot*, ACF and PACF plots of residuals of the selected model is a good convenient graphical technique for model diagnostic.

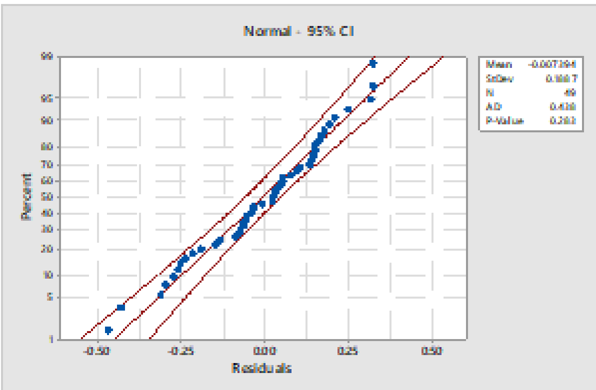

Figure 4: Normal Probability Plot for ARIMA (1, 1, 1) model

According the above figure 4 it indicates that the normal distribution provides an adequate fit for this model since p-value of 0.283 > α = 0.05 which confirmed the normality of the residuals.

The Ljung-Box test also showed that the ARIMA (1, 1, 1) was adequate with p-value = 0.104 > α = 5% and could be used to forecast Hargeisa monthly food prices.

Figure 5 shows the ACF and PACF graphs of residuals obtained from ARIMA (1, 1, 1) model. As we can examine that there is no significant lags and we can conclude that the residuals are white noise (independent and identically distributed) so ARIMA (1, 1, 1) model is the best model for prediction.

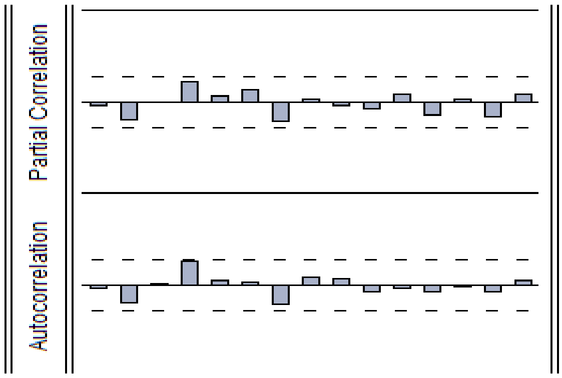

Figure 5: ACF and PACF plots of Residuals for ARIMA (1, 1, 1) model

3.6 Model Validation

The dataset was partitioned as training and testing sample. The training sample contains about 80 % (Nov-2013 to Feb-2017) portion of the dataset for modeling the data. The sample for testing the validity of the model (test sample) contains the remaining portion, 20 % (Mar-2017 to Dec 2017) of the dataset. Since there was an obvious deterministic trend we can also use trend analysis for forecasting, but we determined that ARIMA (1,1,1) is the best model , so we compare both of them based on the estimates of MSE, MAE and MAPE so the model with least estimated is the best model which fit the data series. It is demonstrated in **Table 4** that the ARIMA (1,1,1) has a good predictive ability than Trend Analysis since its estimates is smaller than the other and this confirms that ARIMA (1,1,1) is the best model that can be used for prediction.

Table 4: Model Validation

Model Fit Indexes	ARIMA (1,1,1)	Trend Analysis
MSE	0.1109	0.1437
MAE	0.3126	0.3667
MAPE	1.2439	1.463

Finally **Figure 6** below shows the predicted Hargeisa rice prices which lie within the 95% confidence intervals. The lower confidence limit (LCL) and upper confidence limits (UCL) are also indicated. Hargeisa rice prices are expected to be increase over the time period after December 2017.

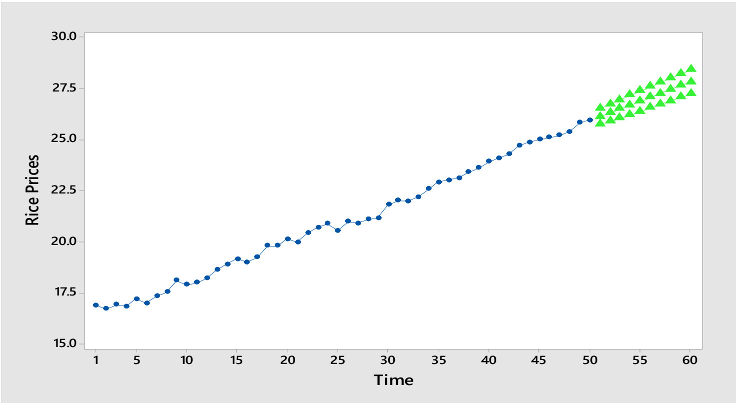

Figure 6: 10 months forecasts of Hargeisa rice price with their lower and upper confidence limits**.**

Our study supports the report by World Food Programme (WFP), that there is a high food prices in all Somaliland regions and specially Hargeisa since all rice are imported from other countries [14].

4. CONCLUSION

This study applied the Box-Jenkins methodology to model the Hargeisa rice prices from Nov 2013 up to Dec 2017 recorded at Ministry of National Planning and Development Hargeisa, Somaliland. The time series modeling was employed by first assessing the time plot, ACF and the PACF of the series. The time plot showed increasing trend in rice prices Nov 2013 to Dec 2017. Finally, the appropriate model ARIMA (1, 1, 1) was used to forecast 2018 (10 months) for the Hargeisa rice price. The model adequacy and validation have also shown that ARIMA (1,1,1) is the most appropriate in predicting the rice prices and it was used to forecast data from

Jan 2018 to Dec 2018. The forecast values fell within the required 95% confidence interval highlighting the adequacy of the fitted model. The results of the forecasting showed that the Hargeisa rice price rates were steadily increasing and this is due to imports all rice from abroad, For the above findings, there is need to integrate possible actions, efforts, programs and policies from the government, international agencies and local NGOs into Food security plans to achieve a maximum reduction of Food prices in Hargeisa city and whole of the country.

5. ACKNOWLEDGEMENTS

The author would like to thank the Somaliland Ministry of National Planning and Development (MoNPD) for allowing access to the data we have used.

REFERENCES

[1] F. A. O. U. Nations, *OECD-FAO Agricultural Outlook 2018-2027*. Food & Agriculture Organization of the UN (FAO), 2018.

[2] S. Wenzlau, "Global food prices continue to rise," *Worldwatch Institute*, vol. 11, no. 4, p. 2013, 2013.

[3] L. Hove and C. Kambanje, "Lessons from the El Nino–induced 2015/16 drought in the Southern Africa region," in *Current Directions in Water Scarcity Research*, vol. 2: Elsevier, 2019, pp. 33-54.

[4] M. Christine Mungai, Guardian Africa and FEWS Net. "Which African countries are battling the biggest food shortages?" world economic forum. https://www.weforum.org/agenda/2016/02/which-african-countries-are-battling-the-biggest-food-shortages/ (accessed.

[5] S. Government, "Somaliland Food & Water Security strategy," Goverment, Hargeisa, Somaliland, Goverment report 2011, vol. DMS.

[6] C. B. Barrett and D. Maxwell, *Food aid after fifty years: Recasting its role*. Routledge, 2007.

[7] S. Sinharay, "An overview of statistics in education," *Elsevier*, pp. 1-11, 2010. [Online]. Available: https://doi.org/10.1016/B978-0-08-044894-7.01719-X.

[8] G. E. P. a. J. Box, G.M, "Time Series Models for Forecasting and Control," no. Holden-Day San Francisco, 1979.

[9] C. Chatfield, " The Analysis of Time Series: An Introduction.," *Chapman & Hall*, no. London, 2004.

[10] J. M. Tebbs, "STAT 520 Forecasting and Time Series. ," *University of South Carolina. Department of Statistics*, vol. 80-306, 2013.

[11] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?," *Journal of econometrics*, vol. 54, no. 1-3, pp. 159-178, 1992.

[12] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American statistical association*, vol. 74, no. 366a, pp. 427-431, 1979.

[13] A. Bartczak, R. Glazik, and S. Tyszkowski, "The application of Box–Cox transformation to determine the Standardised Precipitation Index (SPI), the Standardised Discharge Index (SDI) and to identify drought events: Case study in Eastern Kujawy (Central Poland)," *Journal of Water and Land Development,* vol. 22, no. 1, pp. 3-15, 2014.

[14] F. N. WFP, Nasa Fldas "Persistent drought in Somaliland leads to major food security crisis," no. Mogadishu, Somalia., 2016.