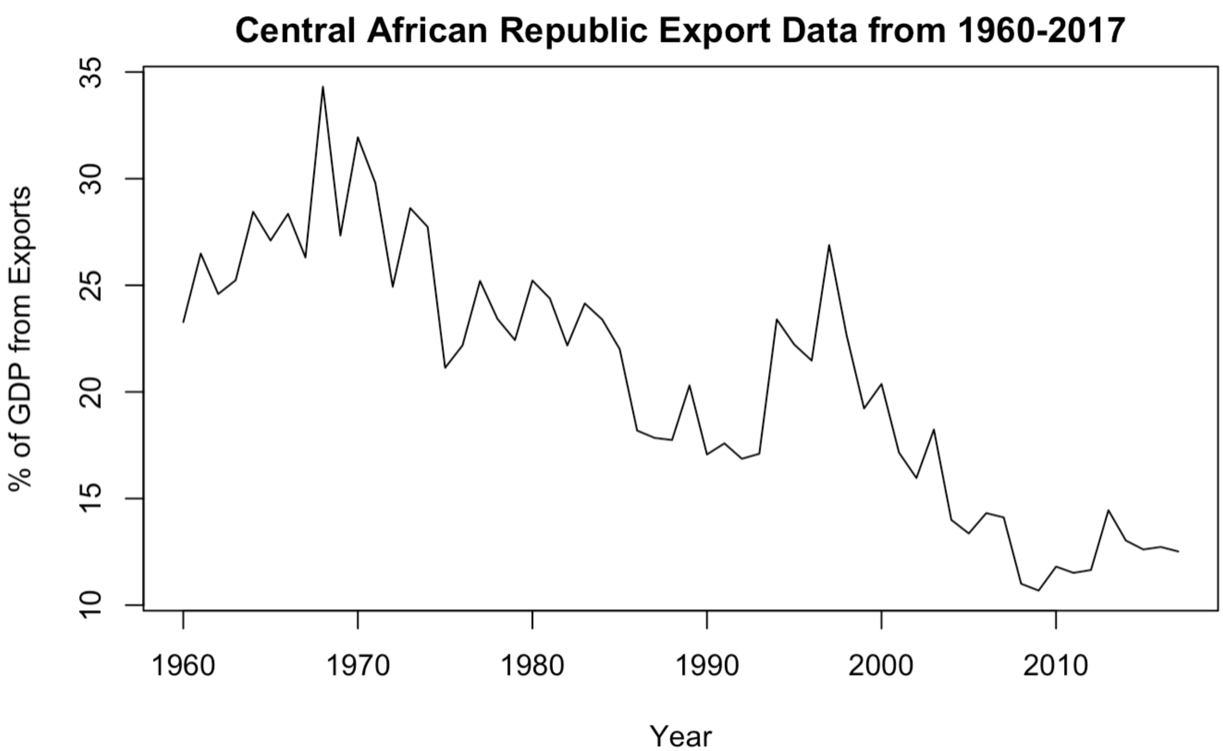
Jonathan Hui
STA 137 Section 001
Dr. Ding
6/4/2022

A Brief Analysis on the Central African Republic's Export Data

The Central African Republic (C.A.R.) is a landlocked country within the continent of Africa. The country was previously a French colony until gaining independence in 1960. Unfortunately, due to an ongoing civil war since 2012 as well as many other geopolitical and socioeconomical factors, the C.A.R. has one of the lowest GDP per capita in the world. As a matter of fact, the IMF's World Economic Outlook report in 2018 showed that the C.A.R. had the lowest GDP per capita in the world, in 2017. I have obtained data of the C.A.R.'s % of GDP that comes from exports, from the nation's founding in 1960 to the date of the IMF's reporting in 2017. The data can be seen as a time series below.



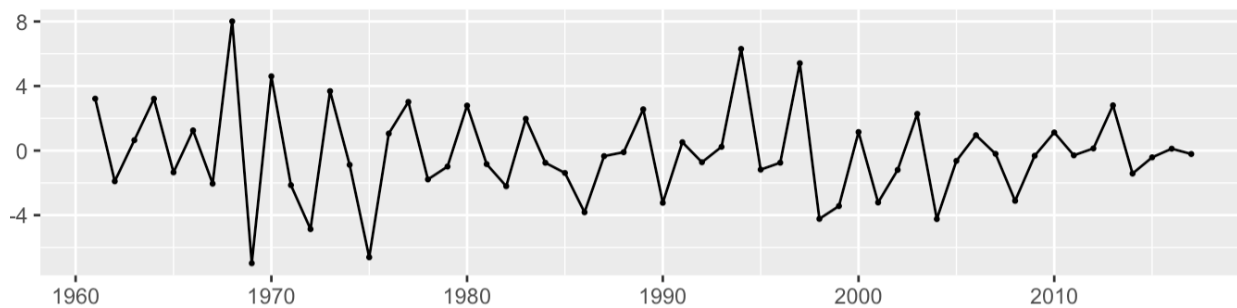
One of the fundamental goals in analyzing time series is forecasting. Since we are interested in predicting the C.A.R.'s export data in the years following 2017, we can use fit the data to a model. Then, we can use that model to predict the trajectory of the nation's percent of GDP that comes from exports. We observe that the data above shows a decreasing trend non-stationary time series during 1960 to 2017. However, there are 2 brief increases in the trend of percent of GDP from exports: between 1960-1970 and between 1993-1997.

Typically, statisticians and data scientists will use Box-Cox transformations to a time series that shows evidence of unstable variance. Because the graph above does not show evidence of unstable variance, I will not do a Box-Cox transform on the data. When I attempted to use it, there were no significant changes to trends in the data.

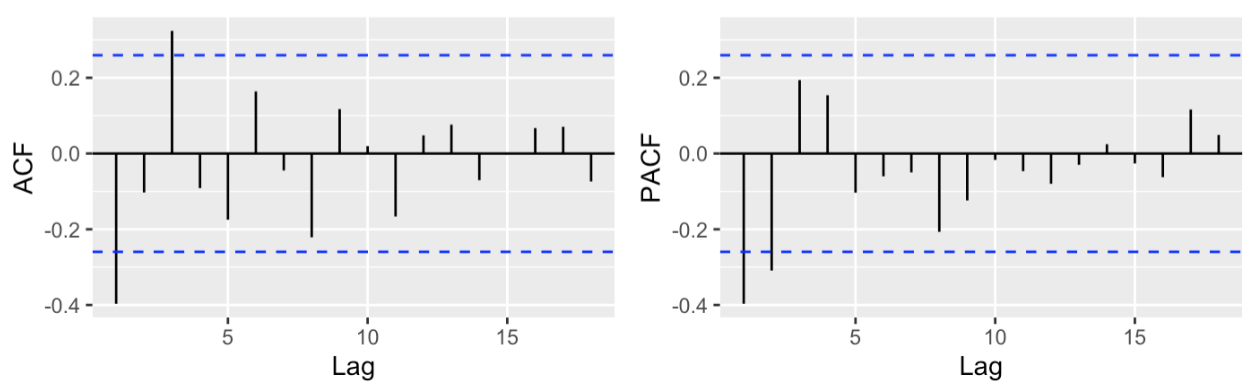
Since the data in the graph above is clearly non-stationary, we will transform the data until it is stationary. This prepares the data for the upcoming model-fitting steps. We take a first order difference of the data, such that:

$$\Delta x_t = x_t - x_{t-1}$$

Once doing so, we achieve data that looks approximately stationary, as graphed below.



With that data, we want to graph its ACF and PACF. In layman terms, this is a diagnostic step that helps us understand if there are relationships between our datapoints versus lag. Based on the ACF graph below, a moving average model for q = 1, MA(1), can be an appropriate fit for our data. Based on the PACF plot below, an autoregressive model for p = 2, AR(2), can be an appropriate fit for our data.



To find the best fit model, I analyzed various ARIMA models, a type of model that we can use to help us in forecasting. There are 3 different components of ARIMA models: p, d, and q. These all represent patterns in autoregression, degree of differencing, and moving average, respectively. Previously, I established that an AR model of p = 2 and an MA model of q = 1 are good fits for our data. When it comes to ARIMA model fitting, I tested many different ARIMA models to find the best fit model, for models near the ballpark of p=2, or q = 1, or both — and for various values of d. Given all of these models, I can find the best fit model by finding the model with the lowest AIC, and coefficients with a low p-value.

Using the sarima function of the astsa r package, I tested every combination of ARIMA model with parameters of $p = 0, 1,$ or 2, $q = 0, 1, 2,$ or 3, and $d = 0, 1$ or 2, except for ARIMA(0,0,0), with no constants. After doing this for all 35 combinations, I found that the ARIMA(2,1,0) model is the best fit for our data. Comparing AICs for all 36 models that I tested, it has the lowest AIC at 4.8164. Thus, when fitting our data to an ARIMA(2,1,0) model, our data follows the formula:

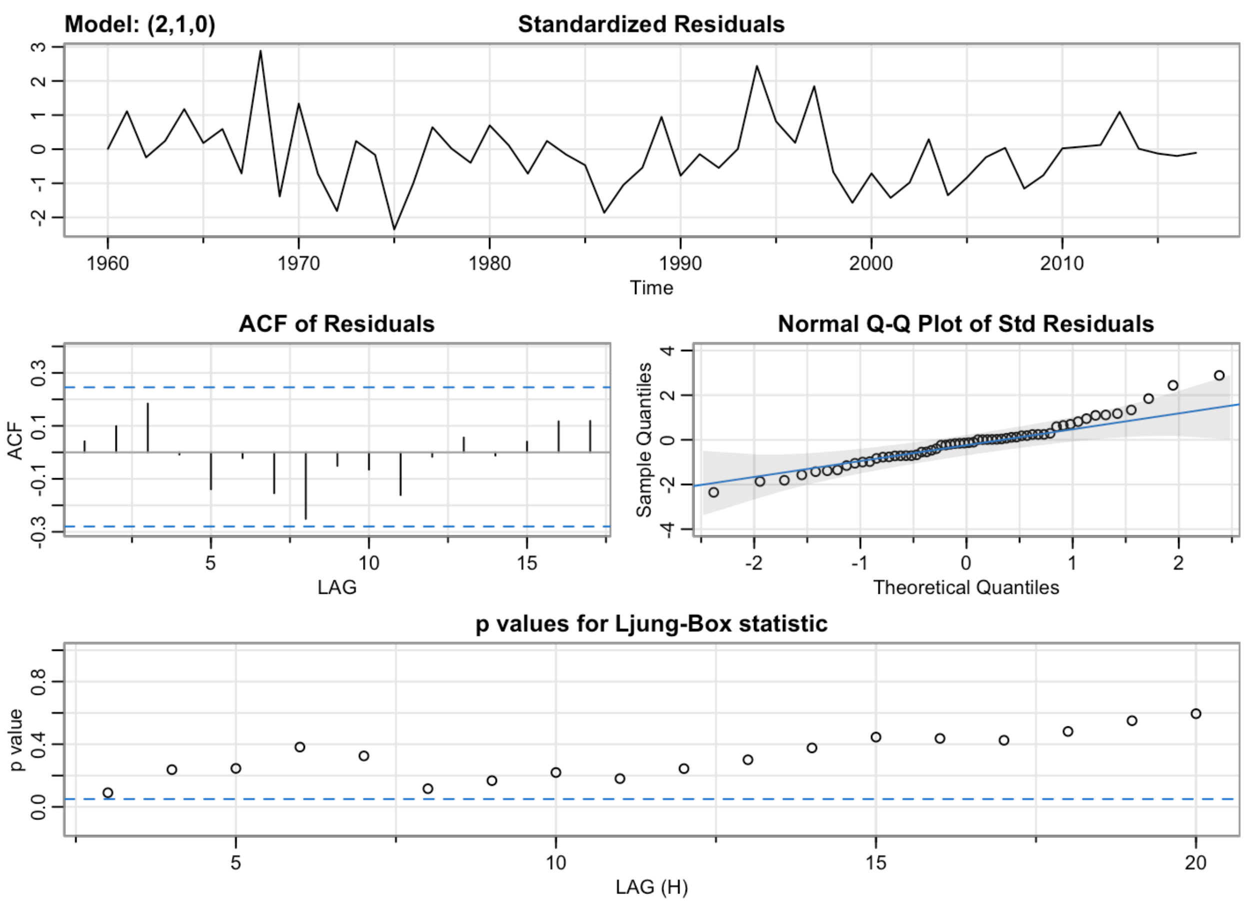
$$\boldsymbol{\Delta x_t = -0.5230\, x_{t-1} - 0.2897 x_{t-2} + w_t}$$

The p-values of the coefficients for the ARIMA(2,1,0) model are 0.0002 for $\Phi_1 = -0.5050$ and 0.0247 for $\Phi_2 = -0.2897$. These coefficients all have a p-value lower than 0.1 meaning that these coefficients are significant.

The next best model based on AIC is the ARIMA(0,1,1) model, with an AIC of 4.83678. The equation for when our data is fitted to this model is:

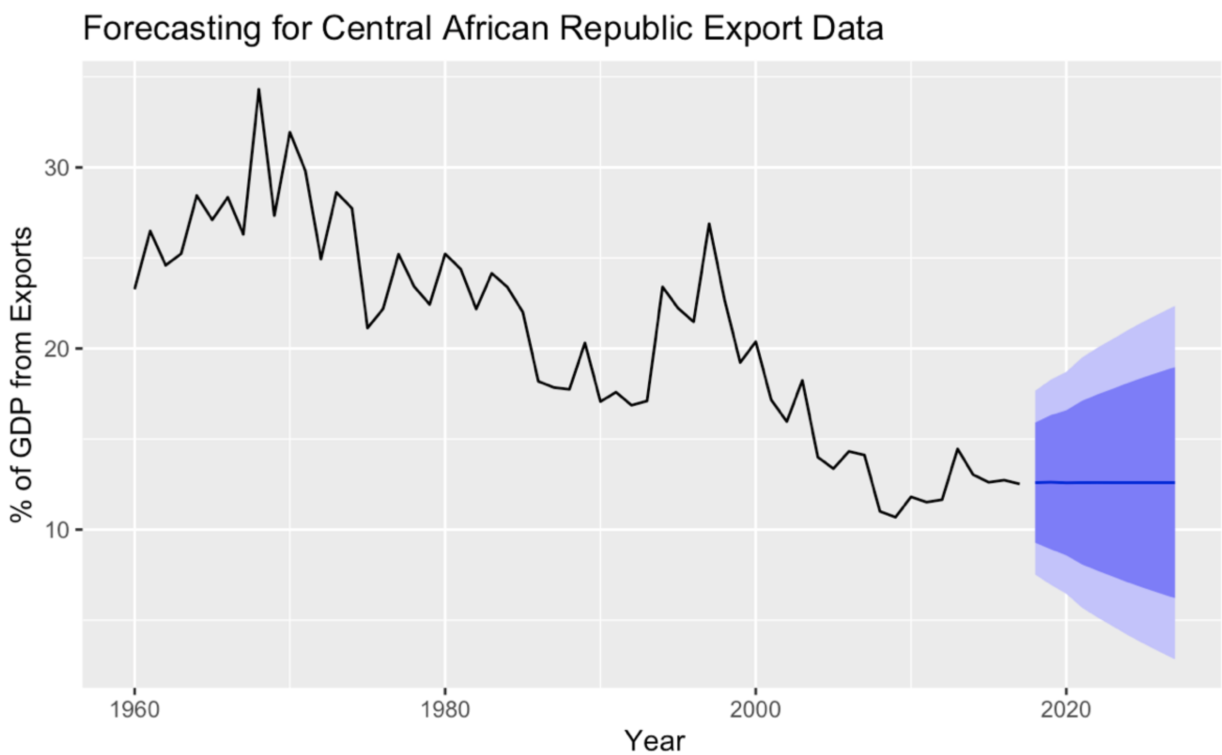
$$\boldsymbol{\Delta x_t = -0.4172 w_{t-1} + w_t}$$

In this model, the coefficient of the moving average term had a p-value of 0.0002. However, I still decided that the ARIMA(2,1,0) model was still a better fit for our data than the ARIMA(0,1,1) model. I proceed to run some diagnostic measures, as seen in the graphs below. This helps us to understand the nature of our residuals when fitting the ARIMA(2,1,0) model.



The standardized residual plot shows that there are no trends in the residuals, no outliers, and a very low change in variance across time. The plot that shows ACF of residuals shows no significant autocorrelation. The Q-Q plot shows that the residuals are normally distributed. The Ljung-Box graph shows that the residuals have non-significant p-values, showing that the residuals are white noise.

After fitting the data to an ARIMA(2,1,0) model, we are finally ready to proceed to forecast the percent of GDP, that the C.A.R. receives, from exports. This can be seen in the graph below.



We have found that the Central African Republic's % of the country's GDP that comes from exports can be modelled best by an ARIMA(2,1,0) model. This is essentially an AR(2) model with a degree of differencing variable of d = 1.

In conclusion, when we fit that data to the model and perform forecasting, we see that the model predicts a stagnating trend in percent of GDP from exports in the next 10 years. This data quantitatively elucidates the story of the C.A.R.'s economy, one characterized by the negative effects of post-colonization and civil war on that nation's economy.

Shortly following the timespan of our data, the COVID-19 pandemic greatly interrupted global supply chain, so it is likely that the true trend in percent of GDP that comes from exports for the C.A.R. decreased more than our model predicted from 2017-2022. We can only hope, for the sake of individuals and families living within the C.A.R., that better days await their nation's economy.