

# STA 137 Final Project

Ryan Chiang, Manik Sethi, Muhammad Laiq

June 2025

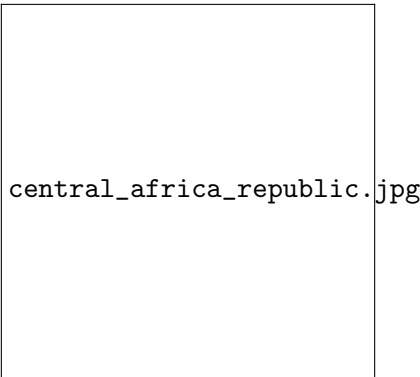


Figure 1: Central African Lumber

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Explanatory Data Analysis</b>	<b>2</b>
<b>3</b>	<b>Model Selection</b>	<b>3</b>
3.1	Transformation . . . . .	3
3.1.1	GDP . . . . .	3
3.1.2	Import Model . . . . .	4
3.2	ACF/PACF . . . . .	5
3.2.1	GDP . . . . .	5
3.2.2	Import Model . . . . .	5

3.3	Candidate Models . . . . .	6
3.3.1	GDP . . . . .	6
3.3.2	Import Model . . . . .	8
3.4	Final Model . . . . .	8
3.4.1	GDP . . . . .	8
3.4.2	Import Model . . . . .	8

## 1 Introduction

Agriculture is a vital part of the Central Africa Republic's (CAR) economy, occupying nearly four-fifths of the national work force (Ref 1). Goods such as crops, timber and diamonds are essential, and the hallmark of CAR's economy. However, it is rare that these goods stay within the republic itself. Many of these items make their way out of the country as exports, which in turn provide capital to keep the country running. Studying the trends of GDP and Exports over time allows us to be informed when deciding whether or not the Central African Republic should agree to certain trade deals. By knowing how dependent CAR is on the wealth of other countries, a delicate balance between domestic independence and international exploitation can be found. Time series analysis will offer insights into year-long trends of the economy.

## 2 Explanatory Data Analysis

For our explanatory data analysis, we will start by looking at our GDP Data. Plotting our data with 'year' on the x-axis (Fig. 2a), we are able to eye-ball the general trend and seasonality. There seem to be two bumps, each separated by approximately twenty years. To verify that this "seasonality" really does exist, we plot the moving average with a lag of 5 and see the two bumps still exist (Fig. 2c). To check if our GDP has a stable variance overtime, we calculate the standard deviation over five years and overlay the graph on top of our regular time series (Fig. 2e). This plot reveals to us that our standard deviation hovers around \$150 million, but quickly jumps up to values of \$250 million during periods of intense growth or recession.

Moving on to our Imports Data, we start off by creating a simple time series of Imports (in millions) by time. (Fig. 2b). It's difficult to discern a trend, as the number of imports goes down from 1960 to 2000, but then starts to rebound up. Whether this is a part of a season whose cycle is longer than the years recorded is

unclear, and will require some research into current CAR economic policies to say for sure. What is clear though, is the volatility of our data (Fig. 2f). Compared to our GDP data, the imports seem a lot more noisy.

Looking at these plots reveal the presence or absence of trends and seasonality in GDP and Imports data. Using this, we are able to make informed decisions on transformations needed to make our data meet the assumptions used for commonly-used time series models.

## **3 Model Selection**

### **3.1 Transformation**

#### **3.1.1 GDP**

The raw GDP time series exhibits a clear upward trend over time, suggesting non-stationarity. This is important because ARIMA models assume constant mean and variance over time. We transform the data, to prevent any model misfit which may produce misleading inferences and poor forecasts. To stabilize the variance and better interpret changes over time, we applied a log transformation to the GDP values:

$$Y_t = \log(\text{GDP}_t)$$

After transforming the data, we applied first differencing to remove the non-stationary trend component. The resulting series  $Z_t$  appears to fluctuate around a constant mean and shows more stable variance, indicating that it may now be stationary.

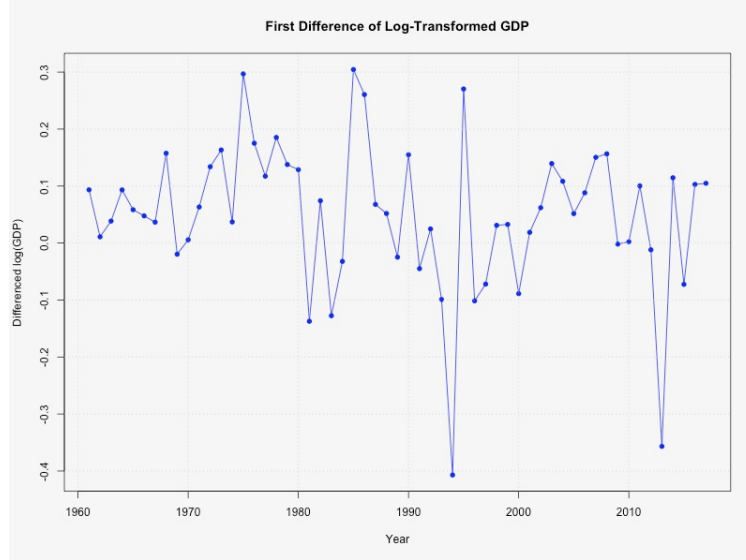


Figure 2: Transformation of GDP series

### 3.1.2 Import Model

The raw import time series appeared volatile and showed signs of non-constant variance with no clear deterministic trend. This is important to note as certain models such as ARIMA assume constant mean and variance. In order to stabilize our variance, we applied a Box-Cox transformation where we estimated the optimal lambda as  $\lambda = .439$ . After this transformation, we performed an Augmented Dickey-Fuller test to test if this transformation helped with our non-constant variance. However, a p-value of .852 would suggest that the series is still non-stationary. Therefore, we applied first differencing to the Box-Cox transformed series such that

$$Z_t = Y_t - Y_{t-1}$$

After differencing, we ran another Augmented Dickey-Fuller test onto our differenced model, yielding us a p-value of .01, suggesting that our transformed and differenced series is appropriate for time series modeling.

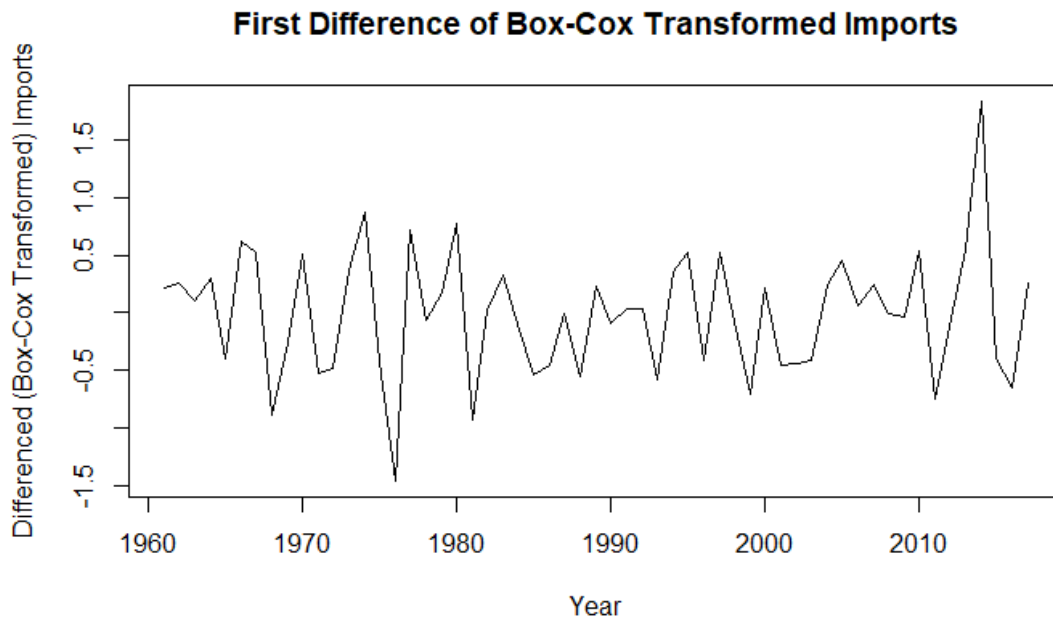


Figure 3: Differenced box-cox transformed import model

## 3.2 ACF/PACF

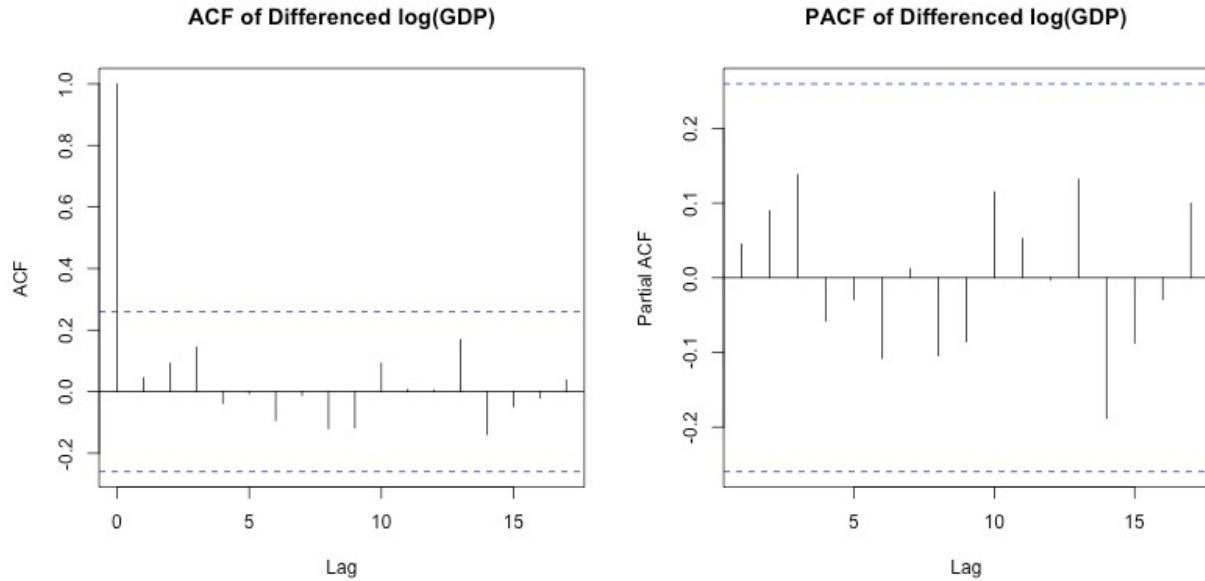
### 3.2.1 GDP

To determine potential models for ARIMA, we examine the ACF and and PACF of the transformed series.

- **ACF:** The ACF plot of the differenced series shows a a spike at lag 1, then followed by a gradual decay. This behavior is similar to a Moving Average (MA) process.
- **PACF:** The PACF shows no strong cut-off pattern, with no lags showing strong significance beyond the noise band. This suggests a lack of strong Autoregressive (AR) behavior, indicating the AR component might not be necessary.

### 3.2.2 Import Model

To determine potential models for ARIMA, we examine the ACF and PACF of the transformed and differenced model for imports.

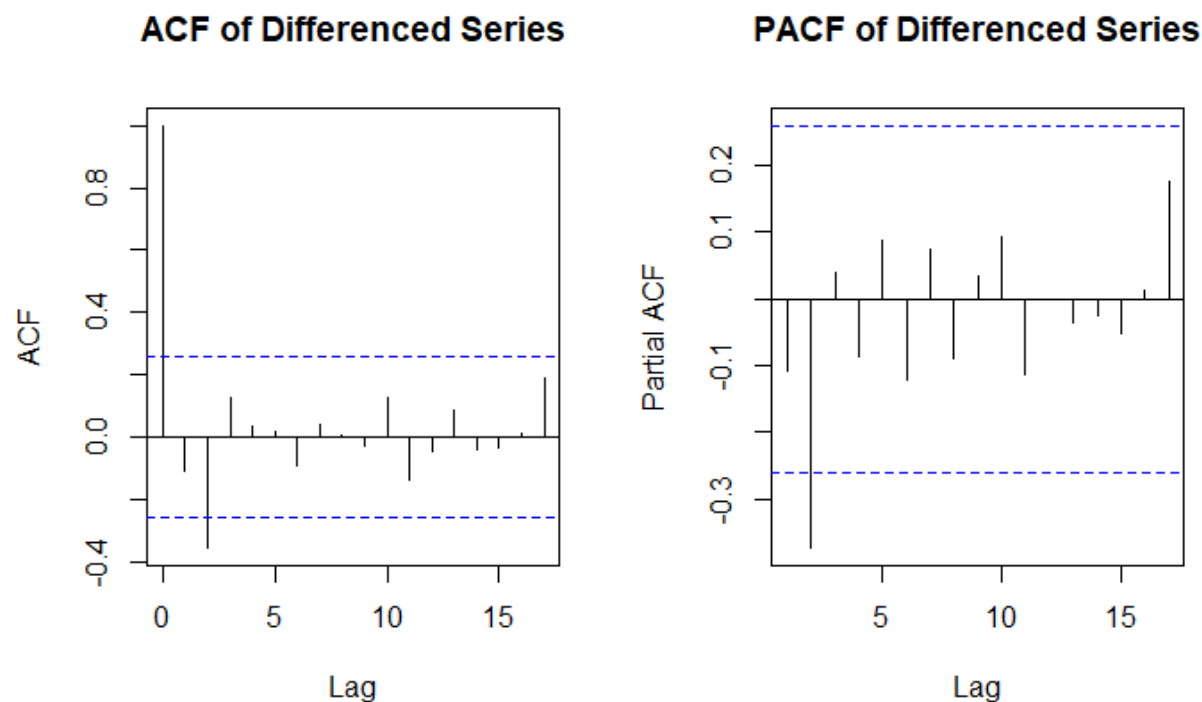


- **ACF:** The ACF plot of the differenced and transformed series shows a significant spike at lag 2 followed by a mostly gradual decay. This behavior suggests that a Moving Average (MA) process could be appropriate.
- **PACF:** The PACF shows that lag 2 is significant, suggesting that an autoregressive part of our model can be appropriate.

### 3.3 Candidate Models

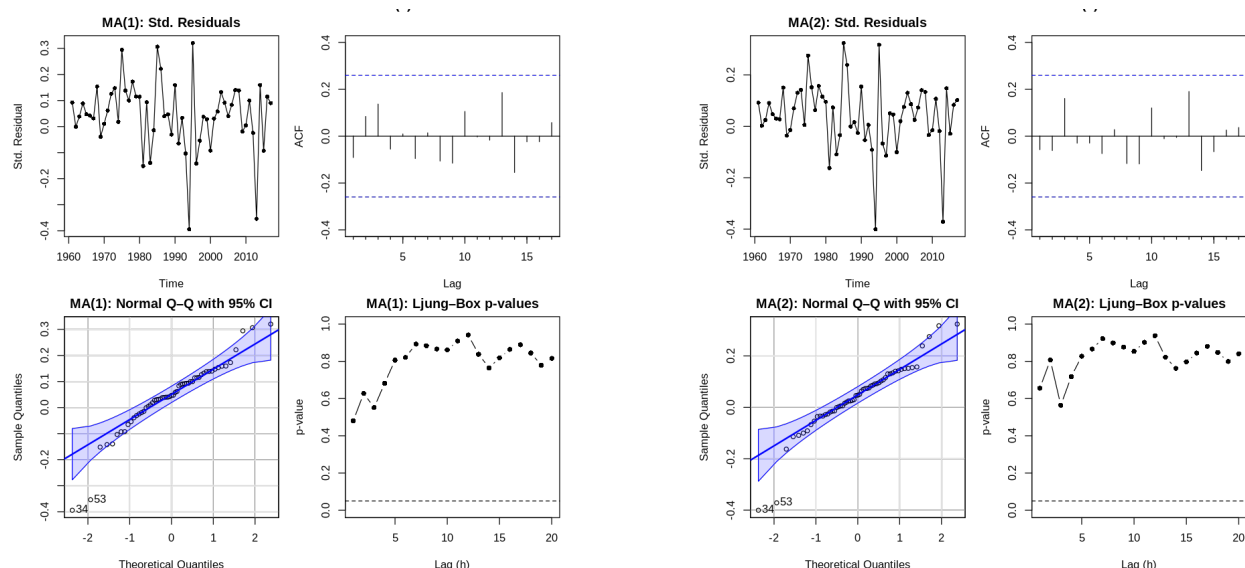
#### 3.3.1 GDP

Our Candidate Models were MA(1), MA(2), ARMA(1,1), and ARIMA(0,1,1). Evaluating the AIC, AICc, and BIC of each of the models returned the following results: As outlined in Figure 3, the model with the lowest AIC, AICc, and BIC is ARMA(1,1). This is a good indicator that our final model for GDP should be ARMA(1,1). However, we will look at residual plots to double check and look for trends in our QQ-plot that might indicate non-normal residuals.



Name	AIC	AICc	BIC	LB10	LB20
MA(1)	-61.21917	-60.99694	-57.13306	0.8618982	0.8159858
MA(2)	-60.48276	-60.02993	-54.35360	0.8532975	0.8406041
ARMA(1,1)	-63.39231	-62.93948	-57.26316	0.8807589	0.9000593
ARIMA(0,1,1)	-62.17999	-61.95776	-58.12928	0.8785951	0.8811332

Figure 4: Comparison Statistics



### 3.3.2 Import Model

The candidate models that we chose were ARIMA(1,1,2), ARIMA(0,1,2), and ARIMA (1,1,0). Evaluating our AIC, AICc, and BIC values for each of the models we find that ARIMA(1,1,2) has the lowest AIC and AICc values, however that ARIMA(0,1,2) has the lowest BIC value. We find that our differenced MA model of ARIMA(0,1,2) has a similar performance of ARIMA(1,1,2). This would suggest that either ARIMA(0,1,2) or ARIMA(1,1,2) would be the best fit. However, we will need to look at residual plots to check for trends and QQ plot to find any possibility of non-normal residuals.

Name	AIC	AICc	BIC
ARIMA(1,1,2)	214.73	215.51	222.83
ARIMA(0,1,2)	216.42	216.88	222.5
ARIMA(1,1,0)	246.43	246.66	250.48

Figure 6: Enter Caption

## 3.4 Final Model

### 3.4.1 GDP

Thus, based upon our Candidate Models, our selected Final Model is ARMA(1,1). With the lowest AIC of -63.4 , AICc of -62.9, and BIC of -57.3, the model showed slight issues with the Ljung-Box p-value. In contrast, the MA processes showed p-values that are more easily identifiable as white noise. However, our MA models have larger AIC, AICc, and BIC, with MA(1) having the lower value between the two MA values. Thus, a strong alternative to the ARMA(1,1) model would be the MA(1) model. All models used log-differenced GDP series.

### 3.4.2 Import Model

Based off of our candidate models, our selected final model is ARIMA(0,1,1) for our import data, with an AIC value of 216.42, AICc value of 216.88, and a BIC value of 222.5. Although this model has a higher AIC and AICc value than ARIMA(1,1,2), the simplicity of the model will help reduce problems with overfitting. Our Ljung-Box p-value of .22 also indicates that our model is a good fit and that our residuals are comprised of mostly white noise.