Central African Republic Exports

Jalen Lee, Eric Sun, Allison Peng

It is essential to look at country exports and the underlying patterns found within the economic data. The goal of our project is to analyze the Central African Republic (CAR) Exports during the time period 1960 to 2017. Since export data is factored into the Gross Domestic Product, we can use the exports to better understand the underlying health of the economy. In addition, analyzing the exports of a country provides essential information that can better understand policy making, political sentiments, and internal/external conflicts during a specific time frame.

1. Exploratory Data Analysis

Even though we are only interested in Exports, we want to look at the structure of the data to better understand the general trends and variables. In the beginning of the time series plot, the 1960's saw a general rise in exports, potentially due to the country gaining its independence from France. Around 1975, exports dropped, and during this time the prime minister, Bokassa, demonstrated authoritarian rule such as proclaiming himself president for life and later the emperor of the nation. Exports fluctuated after Bokassa was ousted, but, the moment he came back to power in 1986 we saw another tanking in exports. After three years in 1989, there was again an increase when Bokassa disappeared for good when he became imprisoned for life. With the return of democracy, there was a general rise in exports in 1993 when Ange-Felix Patasse became elected, ending military rule the same year. This increase peaked in 1997 when France removed all its troops. Then we see a large decrease in exports as the nation plunged into conflict which led to a civil war. There are constant decreases after this point with more and more violence. (https://www.bbc.com/news/world-africa-13150044) Furthermore, we see in Figure 2 that there are 9 variables in the CAR dataset, and we are given 58 observations.

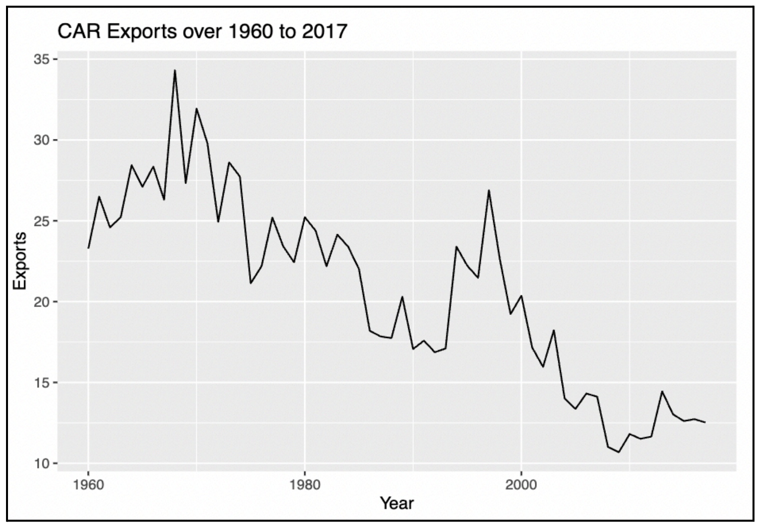


Figure 1: Line graph of exports from CAR dataset

	X	Country	Code	Year	GDP	Growth	CPI	Imports	Exports	Population
	<int>	<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	1	Central African Republic	CAF	1960	112155599	NA	NA	34.18181	23.27272	1503508
2	2	Central African Republic	CAF	1961	123134584	4.9535538	NA	35.76159	26.49007	1529227
3	3	Central African Republic	CAF	1962	124482749	-3.7138002	NA	37.70491	24.59017	1556661
4	4	Central African Republic	CAF	1963	129379098	-0.7070108	NA	38.48581	25.23659	1585763
5	5	Central African Republic	CAF	1964	142025069	2.0803246	NA	40.80459	28.44827	1616516
6	6	Central African Republic	CAF	1965	150574816	0.9475787	NA	37.66938	27.10027	1648833

6 rows | 1-10 of 11 columns

```
## [1] "X"          "Country"    "Code"       "Year"       "GDP"
## [6] "Growth"     "CPI"        "Imports"    "Exports"    "Population"
```

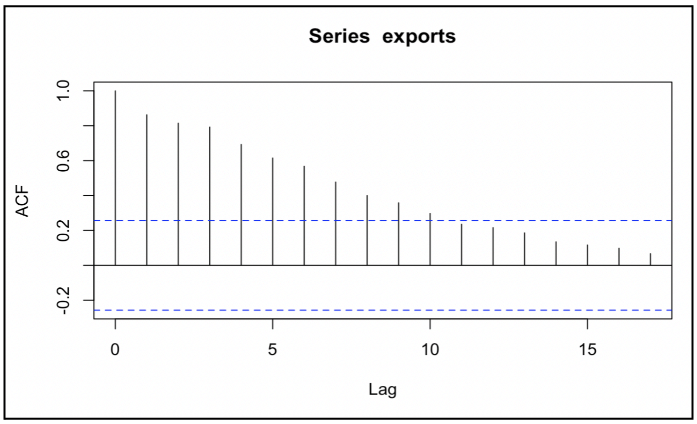
Figure 2: Data Structure of the Central African Republic (CAR) Dataset.

a) Stationary:

In order to fit the data to a time series model, we need to know if it is stationary or not. Since we focused on stationary models in the class, we will need to transform the data if it does not follow the stationary rules. In order to be stationary, the data needs to follow these requirements:

1) $E(x_t) = \mu$, or the mean is constant over time
2) $\gamma(h) = cov(x_{t+h}, x_t)$ only depends on h

In order to check if our data is stationary, we will plot the ACF and PACF plot using the Exports variable. We will also run a hypothesis test to get an objective result.



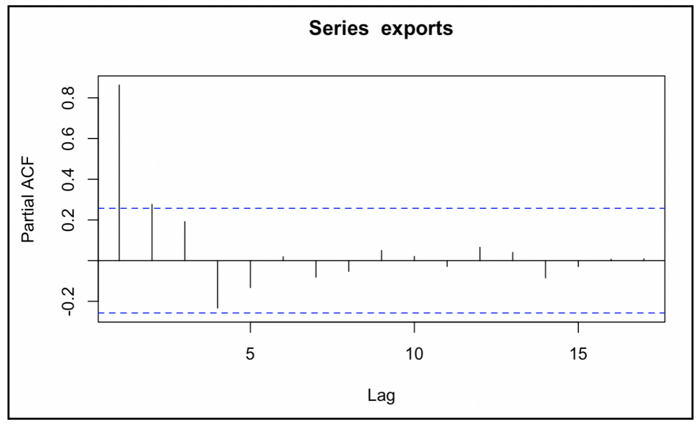


Figure 3: Line graph of the Exports variable and corresponding ACF (left) and PACF (right) plots

KPSS Test for Level Stationarity, $\alpha = 0.05$

Ho: The data is stationary

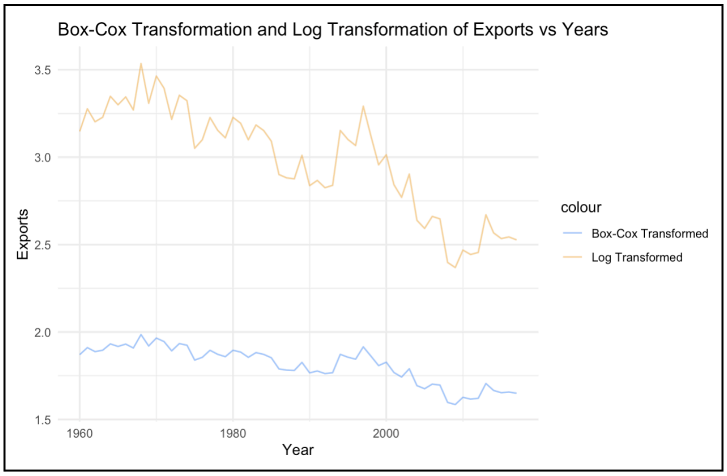
Ha: The data is not stationary

```
Original Data
KPSS Test for Level Stationarity
data:  exports
KPSS Level = 1.2824, Truncation lag parameter = 3,
p-value = 0.01
```

Figure 4: Results to KPSS test for original data

The autocorrelation function (ACF) shows how the correlation between two values changes as the time between them changes. For stationary data, the ACF plot is exponentially decreasing as lag increases. Figure 3 shows the ACF plot, and the ACF value is slowly decreasing as the lag increases, thus it is difficult to determine that it is stationary. The partial autocorrelation function (PACF) shows how the correlation between two values changes as the time between them changes, but removes the effect of the previous lags. We ran the KPSS test to determine if the data is stationary or not. Looking at Figure 4, we see that the p-value of the test is 0.01, we reject the null hypothesis using alpha = 0.05. Thus, the test says that the variable exports is non-stationary. Our next step is to transform the data in order to stabilize the variance for model fitting.

2. Transformations




```
Box-Cox Transformation
KPSS Test for Level Stationarity
data:  transformed_exports
KPSS Level = 1.2639, Truncation lag
parameter = 3, p-value = 0.01


Log Transformation
KPSS Test for Level Stationarity
data:  log(exports)
KPSS Level = 1.276, Truncation lag
parameter = 3, p-value = 0.01
```

Figure 5: Transformation graph (left) and KPSS test results (right)

We attempted to transform the data using a log-transformation and a Box-Cox transformation. In Figure 5 it shows the two transformations led to different results and that the Box-Cox displayed greater stabilization. However, after conducting the KPSS test for stationarity, both test results stated that the resulting data is non-stationary. Thus, we discuss additional methods to reduce trends.

3. Differencing/Testing Stationary

Since the transformations did not lead to a stationary dataset, we now attempt to remove any trends from the data through differencing and a moving average. The definitions are provided below:

1) **Differencing**: allows us to remove linear trends or seasonality in the data by computing the change between consecutive time points:

   a) $\nabla x_t = x_t - x_{t-1}$

2) **Simple Moving Average**: an estimation of the trend for the data. When we remove this estimation, it serves as a way to detrend the time-series and view it as stationary, not dependent on time. For the sake of making forecasts, we prioritize the latest dates available over the starting ones to predict next export values. Including too many years seemed to take away too much information as the importance of the years in a local time frame gets diminished. In this case, we define the difference of 2-MA as:

   b) $x_t - (x_t - x_{t-1})/2$

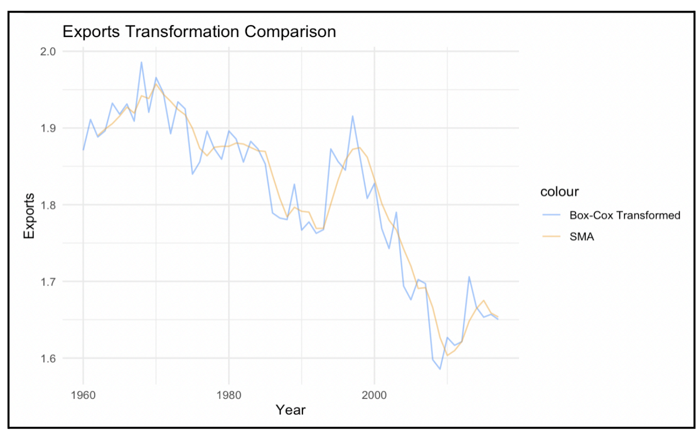
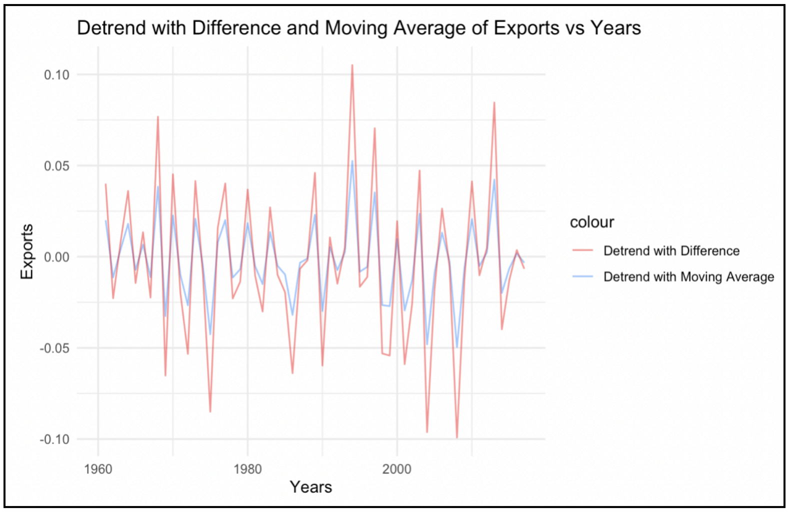


Figure 6: Comparison of Box-Cox transformation with Moving Average detrending




```
Difference
KPSS Test for Level Stationarity

data:  diff(transformed_exports,
1)
KPSS Level = 0.1058, Truncation
lag parameter = 3, p-value = 0.1

Moving Average:
KPSS Test for Level Stationarity

data:  detrend
KPSS Level = 0.1058, Truncation
lag parameter = 3, p-value = 0.1
```

Figure 7: Detrend with First-Order Difference and Simple Moving Average (left) and KPSS tests

The two detrending techniques above incorporated only the Box-Cox transformation before proceeding with the KPSS evaluation. Utilizing a significance level of 0.05, the KPSS tests conclude that the detrending methodologies of 1st Order Differencing and MA with p-values of 0.10 both fail to reject the null. In addition, it can be observed that the MA has a more stable variation compared to first order differencing, suggesting more promising modeling results. We were able to conclude that both methods are stationary series, however we decided to continue to model selection with the moving average detrending.

4. ACF/PACF for Detrended Data and Model Selection

a) <u>ACF/PACF of Detrended Data with Box-Cox Transformation</u>

For ACF/PACF analysis, we decided to derive potential models using the combination of simple moving average detrending and Box-Cox transformation. The ACF and PACF charts are shown below. We are able to choose the best model for the data by looking at the cutoff points for each graph.

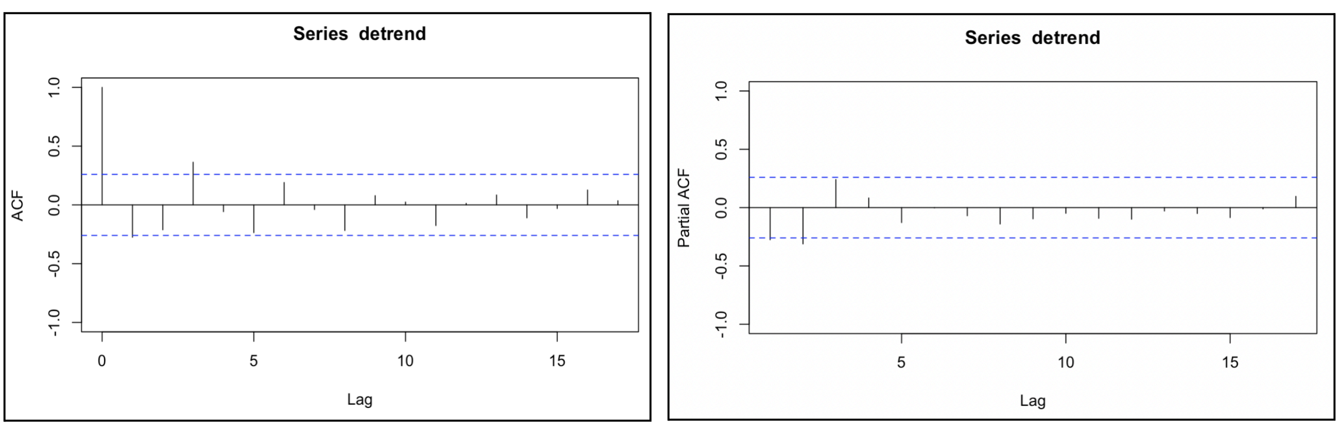


Figure 8: ACF and PACF of detrended data

As seen in Figure 8 the cutoff lag for PACF seems to cut on the second lag, however, the third lag is very close to the cutoff line. From this, a few models we could test are AR(2) and AR(3), or autoregression functions. The ACF cutoff is close at lag 3, to be safe we will model up to MA(4), moving average function. More information about the AR and MA functions are in the following section.

b) Model Selection

Autoregressive model, **AR(p)**: Using previous data outputs in this case exports, as predictors, since previous years may affect future years. w_t represents the residual of the model prediction. The variable p represents the number of phi coefficients in the model. This is an **AR(1) model:**

c) $x_t = \phi x_{t-1} + w_t$

Moving Average model, **MA(q)**: Using previous data residuals, or the difference between predicted outputs and the actual outputs, as predictors for future years. The variable q represents the number of theta coefficients in the model. This is a **MA(1) model:**

d) $x_t = \Theta w_{t-1} + w_t$

ARMA model, ARMA(p,q): ϕ and Θ serves as a coefficient, scaling the importance of the previous years output and residual. With a greater p and q, there will be more of these coefficients. An **ARMA** model is the combination of the two, here is an ARMA(1,1) example:

e) $x_t = \phi x_{t-1} + \Theta w_{t-1} + w_t$

Model metrics, AIC/BIC: AIC and BIC, information criterions, used to determine the model performance, with a lower score indicating a better result. It could be understood as the model performance penalized by the model complexity, including too many variables, which leads to overfitting potentially causing bad predictions.

f) **AIC** $= 2p - 2ln(\hat{L})$ where p are the number of predictors and $\hat{L}$ being the MLE (maximum likelihood estimate) which optimizes the model parameters

g) **BIC** $= pln(n) - 2ln(\hat{L})$ where n represents sample size

In Table 1, we calculated the AIC and BIC for several models. We used the detrended data with box-cox transformation and fit several AR, MA and ARMA models as shown below. Our goal is to find the most negative AIC and BIC measurement and fit the best model to the data. When looking at the potential models from the ACF/PACF, we were interested in AR(2) and AR(3) and MA(4). The MA(4) model has a significantly lower AIC and BIC compared to the AR models. We can see that even though AR(3) has a higher AIC than AR(2), the BIC is higher for AR(2). These results are very close, thus we will end up picking the simpler model, AR(2). This prevents the model from getting too complex and avoids overfitting. We also looked at models ARMA(2,4) and ARMA(3,4) but they both have a significantly lower BIC, thus we will not consider them for our model.

AIC/BIC	p = 0	p = 1	p = 2	p = 3
q = 0	AIC BIC	-275.0384 -268.9092	-278.7482 -270.576	-280.037 -269.8217
q = 1	-277.0533 -270.9241	-275.0562 -266.884	-278.4439 -268.2287	-278.1956 -265.9373
q = 2	-275.0754 -266.9031	-274.7874 -264.5722	-279.5555 -267.2972	-277.6373 -263.3359
q = 3	-278.0504 -267.8352	-276.2664 -264.0081	-277.6414 -263.34	-280.0133 -263.6689
q = 4	-276.5816 -264.3233	-278.4689 -264.1676	-277.793 -261.4485	-278.148 -259.7606

Table 1: AIC/BIC Metrics for multiple models, green highlight is selected model

c) Final Model

Chosen Model: AR(2)

We ran the summary function to see the coefficients of the chosen model and evaluated the p-values to determine if the coefficients were significant. The models and p-values are displayed below:

$$\textbf{Model: } x_t = -.359908x_{t-1} - .310981x_{t-2} - 0.003608455 + w_t$$

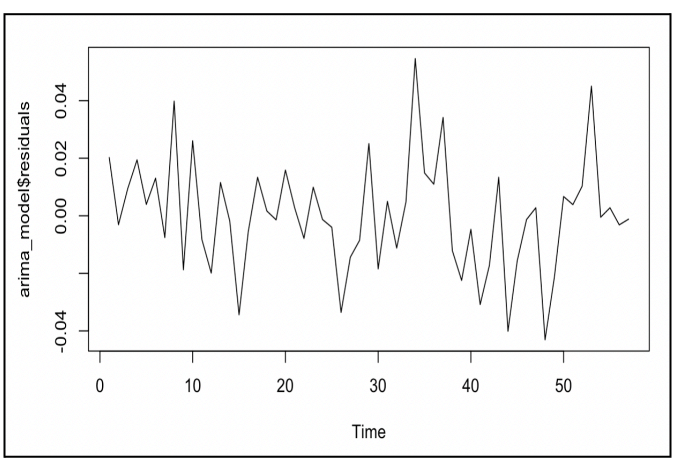
P-value: $\phi_1 = 0.0042$, $\phi_2 = 0.0126$

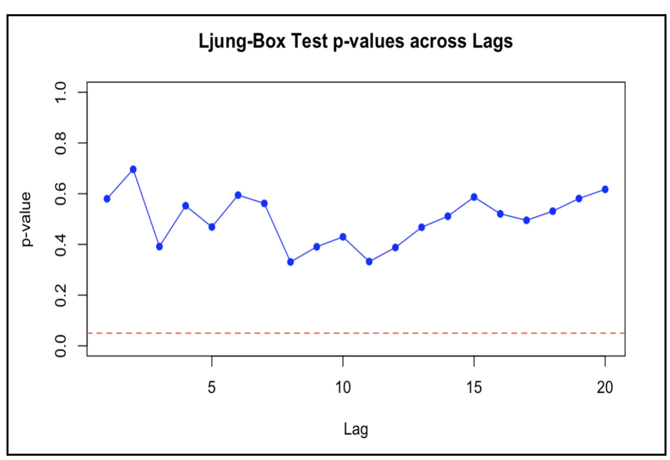
Since we are using a significance level of alpha = 0.05, we see that both coefficients are significant, and we will keep both of them in the model. Now, we will evaluate the model's performance even further by looking at the residuals.

5. Model Residual Diagnostics

To further evaluate the model's performance in fitting the data, we look at the residuals. If the residuals are normally distributed and behave like white noise, we can confidently say the model is a good fit for the data.

Residuals for the AR(2)





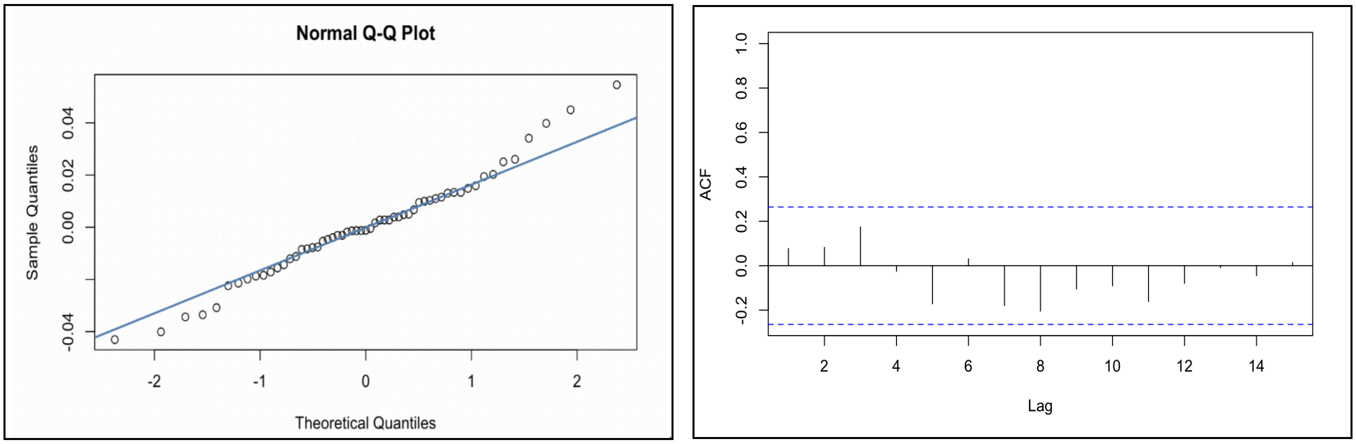


Figure 9: Residual Diagnostics, Residual plot (upper left), Ljung-box test p-value (upper right), Quartile-Quartile plot (bottom left), ACF plot (bottom right)

The model appears to fit the data well, as the residuals seem to be white noise, albeit a few of them potentially being outliers from the Q-Q plot. The ACFs of the residuals are all statistically significant (all confined within the blue dotted lines) and the Ljung-Box test displays every p-value having a value greater than the chosen significance level of 0.05, indicating that we fail to reject the null hypothesis that the residuals are white noise. As our residuals look good, we conclude that the AR(2) model fits well to this data compared to other alternatives, hence is used for forecasting future values.

6. Future Predictions

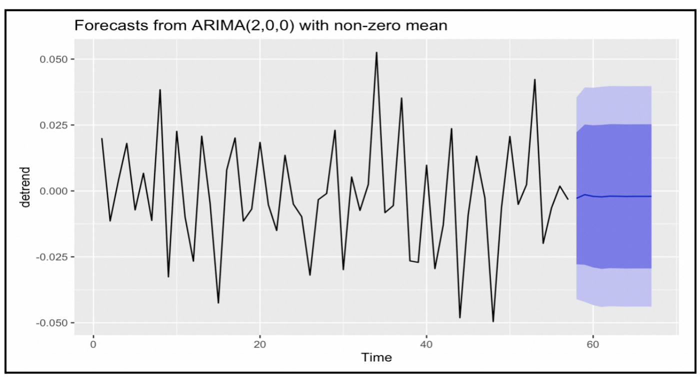


Figure 10: Predicting future export values using AR(2)

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
58	-0.002805610	-0.02781198	0.02220076	-0.04104956	0.03543834
59	-0.001414069	-0.02799701	0.02516888	-0.04206918	0.03924104
60	-0.002070691	-0.02901601	0.02487463	-0.04328000	0.03913862
61	-0.002259961	-0.02955412	0.02503420	-0.04400278	0.03948286
62	-0.001990647	-0.02928570	0.02530440	-0.04373482	0.03975353
63	-0.002029822	-0.02935359	0.02529395	-0.04381792	0.03975828
64	-0.002098155	-0.02942686	0.02523055	-0.04389380	0.03969749
65	-0.002061516	-0.02939092	0.02526789	-0.04385824	0.03973520
66	-0.002053807	-0.02938418	0.02527656	-0.04385200	0.03974439
67	-0.002067806	-0.02939819	0.02526258	-0.04386602	0.03973041

Figure 11: Confidence intervals for 10 future years

Using the selected AR(2) model with the detrended and transformed data, we now can predict future values. We used the forecast function in R to predict the exports for 10 additional years. Each row represents a prediction of exports using the AR(2) model, starting from 2018. The forecasted data does look monotone compared to the original data, thus there are possible issues with the model performance. However, since we transformed and detrended the data, it is difficult to interpret the values.

7. Limitations/Discussion

The process of fitting a model to the exports data was a multi-step process that involved trial and error. Initially, we did try to compare the first-order difference detrending and the simple moving average detrending. After finding the optimal model for first-order difference, we got an AIC of -202.0509 and a BIC of -193.8787. This is significantly higher than the AIC and BIC for the SMA detrending, thus we proceeded to use SMA detrending. When using the final model to predict the values, the predicted values did not follow similar patterns to the original data. This could be due to a small sample size as we would need a larger set of observations to train the model on. It is also difficult to interpret the data because we transformed the data leading to a change in units. Future work could include trying different models to fit the data and running the forecast function on those models.

8. Conclusion

The Central African Republic (CAR) Exports time series displayed a non-stationary trend with fluctuating variances at different points in time. As forecasting is a powerful tool for groups to capitalize on economic opportunities and mitigate risks, the analysis we conducted revolved around the idea of selecting the best model for forecasting future CAR Exports. In the analysis, transforming the data with the Box-Cox and Simple Moving Average procedures appeared more optimal for stabilizing the variance and satisfying stationary conditions compared to other methods (i.e first order differencing). After these transformations, the ACF/PACF analysis displayed multiple potential candidates for selection, and ultimately we decided that ARIMA(2,0,0) or AR(2) fit the best due to its low BIC value and relative model simplicity. The validity of the AR(2) model was further verified through producing residual diagnostic plots, which displayed the residuals of the model conforming to the standards of a white noise process.

Contributions:

Allison: Document Formatting, Analysis and Interpretation, Model Selection
Jalen: Coding, Model Selection, EDA, Detrending, Transformation
Eric: ACF/PACF Analysis, Residual Diagnostics, Conclusion