Time series analysis of CAR GDP and imports

Nicole Shum – STA137 final project – SQ2025

1. Introduction

1.1 The Central African Republic (CAR)

The Central African Republic (CAR) is a landlocked nation in Africa with a population of approximately 5 million as of 2024 [1]. Despite its rich natural resources, the country remains among the ten poorest in the world, holding the lowest GDP per capita globally as of 2017 [2]. After gaining independence from France in 1960 [3], the CAR has faced decades of political instability, conflict, and economic challenges. This study uses data spanning from the year of CAR's independence up to 2017 to examine the country's economic performance over time.

The goal of this analysis is to visualize and understand the economic trajectory of the CAR since its independence, with a focus on two key indicators: Gross Domestic Product (GDP) and imports. GDP measures the total economic output of a country and serves as a primary indicator of economic health and growth. Imports represent the value of goods and services brought into the country and can reflect both domestic demand and international economic integration. By analyzing trends in these two variables, we aim to assess past performance and forecast the CAR's future economic prospects.

1.2 Time series analysis

This study will utilize time series analysis to provide insight into this data. Time series analysis is a statistical approach used to analyze data points collected or recorded at successive time intervals, where observations are typically not independent but influenced by preceding values. The primary goal is to understand the underlying structure and patterns in the data, such as trends, cycles, and seasonality, and to use this information for forecasting future values. Two common components of time series models are autoregressive (AR) and moving average (MA) terms. AR models predict future values based on past observations, while MA models use past forecast errors to improve predictions. To help determine the appropriate model, we will use autocorrelation functions (ACF) and partial autocorrelation functions (PACF), which graphically display how data points in the series are correlated with their past values across different time lags.

1.3 Data & study overview

This study analyzes a dataset containing annual economic data for the Central African Republic from 1960 to 2017, consisting of 58 rows and 9 columns. The dataset includes key indicators such as GDP, economic growth, population, consumer price index (CPI), imports, and exports. All statistical tests in this study are conducted at a significance level of 0.1. The analysis proceeds in three main stages: first, exploring the data to assess whether transformations or differencing are necessary to achieve stationarity and homoscedasticity; second, selecting and fitting the most appropriate time series model using the Akaike Information Criterion (AIC) as the model selection criterion; and third, using the finalized models to generate forecasts and assess the future economic trajectory of the country.

2. Exploratory data analysis

2.1 Plots & summary statistics

To analyze time series data, we first need to assess the stationarity and homoscedasticity of the raw data to determine if any differencing or transformations need to be applied. Stationarity refers to a time series whose statistical properties, such as mean and variance, do not change over time. To evaluate this, we applied the Augmented Dickey-Fuller (ADF) test. A p-value below 0.1 in this test indicates that the series is likely stationary, indicating that the time series does not need to be differenced. Additionally, to assess whether the variance of the data is constant (homoscedastic), we performed Levene's test by dividing each series into two equal segments, representing the first and second halves, and testing for equal variance between them. A p-value in this test above 0.1 will indicate homoscedasticity. Stable variance supports the application of linear models without transformation.

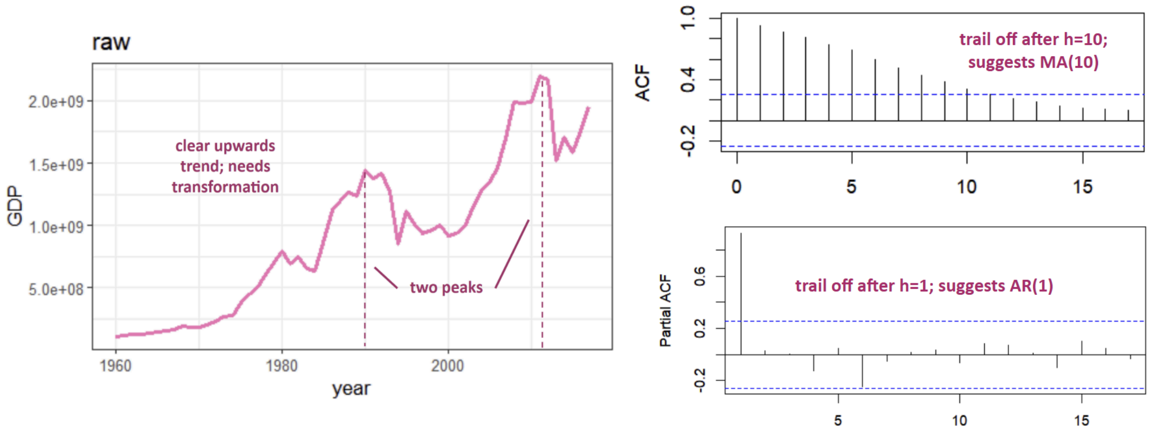

Figure 1. GDP data (raw) and ACF/PACF plots.

The GDP series (Figure 1) exhibits a clear upward trend. The ADF test produced a p-value of 0.09692, suggesting that while the series may be marginally stationary at $\alpha = 0.1$, it is on the cusp and may benefit from transformation or differencing to improve model performance. Levene's test resulted in a p-value of 0.4013, indicating stable variance across time. The data has a mean of 9.394×10^8 and a standard deviation of around 6.159×10^8.

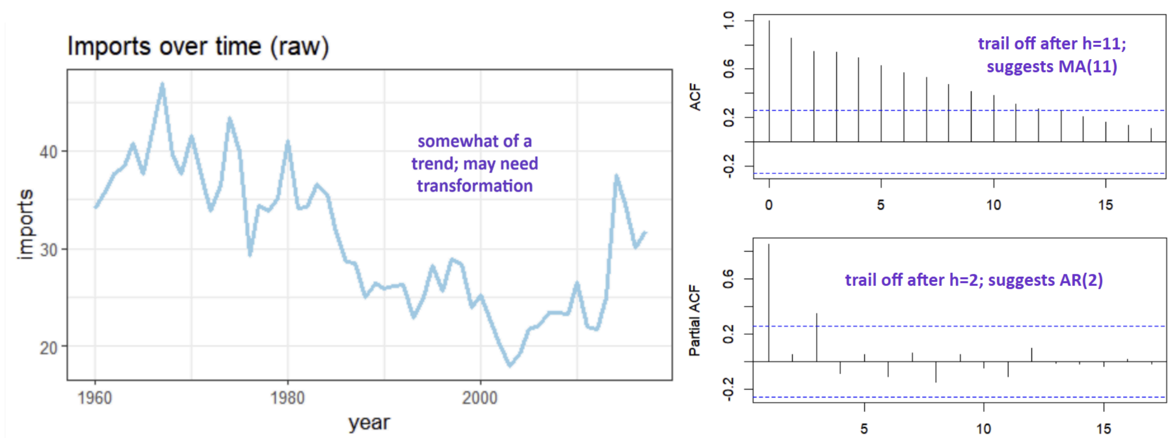

Figure 2. Imports data (raw) and ACF/PACF plots.

The imports data (Figure 2) shows a more complex pattern with a downward trend initially, followed by an upward trend in more recent years. The ADF test yielded a high p-value of 0.852, indicating strong non-stationarity and a clear need for differencing or transformation before modeling. Levene's test again indicated constant variance ($p = 0.5115$), suggesting that volatility is not changing substantially over time. The import data has a mean of 30.83 and a standard deviation of 7.13, reflecting smaller magnitude but meaningful variation in trade volume.

2.2 Transformations and differencing

Differencing is a common method used to remove trends and stabilize the mean of a time series, which helps to achieve stationarity, which is an essential condition for most time series models. First-order differencing subtracts each value from its immediate predecessor and is often sufficient for removing linear trends. Higher-order differencing, such as second-order differencing, can be used for more complex patterns but risks over-differencing, which may distort the underlying structure of the data and introduce unnecessary noise. Therefore, the goal is to apply the minimal level of differencing needed to achieve stationarity while preserving the integrity of the original signal.

For the GDP data, first-order differencing improved stationarity compared to the raw series. The ADF test p-value dropped to 0.04988, indicating the differenced series is now stationary at the 0.05 significance level. However, Levene's test yielded a p-value of 0.004, suggesting unequal variance between the first and second halves of the series. This result points to potential heteroskedasticity, especially as the series shows increased variability in later years, and indicates that a variance-stabilizing transformation may be needed before modeling.

In contrast, the differenced imports data (Figure 3) shows strong signs of improvement and requires no further transformation. The ADF test result is highly significant ($p < 0.01$), confirming that the series has become stationary after differencing. Levene's test returned a p-value of 0.1388, indicating stable variance across time. Visually, the series appears to have a constant mean and variability, suggesting that first-order differencing has sufficiently addressed both trend and variance issues.

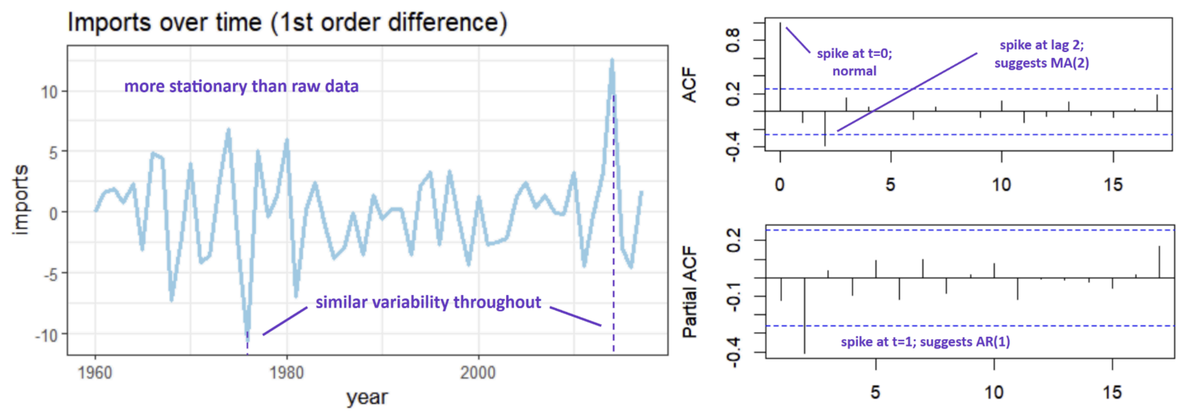

Figure 3. 1st order difference of imports data and ACF/PACF plots.

Transformations are often used in time series analysis to stabilize variance and improve the performance of statistical models. Two common approaches are the logarithmic transformation and the Box-Cox transformation. The log transformation is a special case of the Box-Cox method (with $\lambda = 0$) and is commonly used when the data are strictly positive and exhibit multiplicative or exponential patterns. The Box-Cox transformation, on the other hand, is more flexible—it identifies the optimal power parameter λ that best stabilizes variance and normalizes the data. While both methods aim to address heteroskedasticity and improve stationarity, the Box-Cox transformation can adapt to more complex variance structures but may require more interpretation.

Applying the Box-Cox transformation to the GDP data improved its stationarity and slightly stabilized the variance compared to differencing alone. The ADF test returned a p-value of 0.057, indicating marginal stationarity. Levene's test yielded a p-value of 0.02796, which, while still indicating unequal variance, shows improvement relative to the differenced data without transformation. This indicates that the transformed data may be sufficiently stationary and ready for modeling, but would be more optimal if the variance was more constant.

The log transformation produced more promising results (Figure 4). The ADF test yielded a p-value of 0.07012, indicating borderline stationarity at the 10% significance level. However, Levene's test resulted in a much more favorable p-value of 0.383, suggesting the variance is now statistically constant. Similar to the Box-Cox transformation, the ACF and PACF plots contain no significant autocorrelations beyond lag 0, indicating that the series is well-prepared for modeling with no evident AR or MA components. Given its simplicity and improved variance stabilization, the log transformation may be preferable for this dataset over the more complex Box-Cox approach.

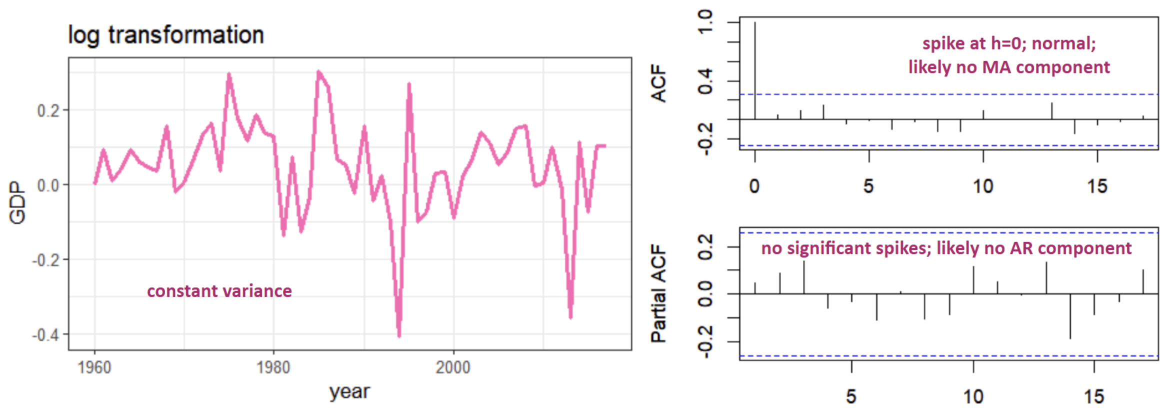

Figure 4. Log transformation of GDP data and ACF/PACF plots.

Based on the exploratory analysis, the final transformation applied to the GDP data was a log transformation followed by first-order differencing, which successfully stabilized both trend and variance. For the imports data, first-order differencing alone was sufficient to achieve stationarity and constant variance, making further transformation unnecessary. These transformations ensure both series meet the assumptions required for effective time series modeling.

3. Modeling

3.1 Model selection

In ARIMA models, the parameters p, d, and q represent the autoregressive (AR) order, the degree of differencing, and the moving average (MA) order, respectively. These parameters are typically selected by examining the autocorrelation (ACF) and partial autocorrelation (PACF) plots of a stationary version of the time series. To complement this visual approach, we also use the auto.arima() function in R, which selects the best-fitting model by minimizing the Akaike Information Criterion (AIC), a measure of model quality that balances goodness of fit with model complexity. Lower AIC values indicate better models, helping us avoid overfitting by penalizing unnecessary parameters.

For the GDP data, we used the log-transformed, first-order differenced series identified in Section 2.2 as the most stable version (Figure 4). The ACF plot showed only a spike at lag 0, and the PACF displayed no significant lags, indicating that neither AR nor MA terms were needed. With one level of differencing, the initial model was ARIMA(0, 1, 0), which produced an AIC of -62.05. However, the auto.arima() function suggested a better model with an AIC of -68.11. This improvement was achieved by including a drift term, which accounts for the non-zero mean and gradual upward trend in the differenced series. The final selected model for GDP is ARIMA(0, 1, 0) with drift, as it achieves the lowest AIC and best balance of simplicity and fit.

For the imports data, the first-order differenced series (Figure 3) showed a spike in the ACF at lag 2 and a trailing pattern thereafter, suggesting a second-order MA component. The PACF had a spike at lag 1, implying a first-order AR component, leading to an initial model of ARIMA(1, 1, 2) with an AIC of 309.49. However, the auto.arima() function produced a slightly lower AIC of 309.37 with an ARIMA(0, 1, 2) model. Through further testing, we found that removing the AR term not only improved the AIC but also simplified the model, reducing the risk of overfitting. Thus, the final model selected for the imports data is ARIMA(0, 1, 2).

3.2 Model performance

To evaluate how well our ARIMA models fit the data, we analyze the residuals of each model. Specifically, we examine residual plots, the autocorrelation function (ACF) of the residuals, and apply the Ljung-Box test. Our goal is for the residuals to resemble white noise—a random series with no autocorrelation, trend, or structure—indicating that the model has successfully captured all meaningful patterns in the data. The Ljung-Box test checks for autocorrelation in the residuals; using a significance level of 0.1 and a lag of 10, a p-value greater than 0.1 suggests that the residuals are indistinguishable from white noise and that the model is appropriate.

For the GDP data, the residual diagnostics provide strong evidence of good model fit. As shown in Figure 5, the residuals appear random with no discernible trend, and the histogram of residuals indicates approximate normality. The ACF plot of the residuals shows no significant spikes, suggesting minimal autocorrelation. Most importantly, the Ljung-Box test yields a p-value of 0.874, well above our threshold of 0.1, confirming that the residuals behave like white noise. These results support the conclusion that the final model—ARIMA(0, 1, 0) with drift—is a good fit for the log-differenced GDP data.

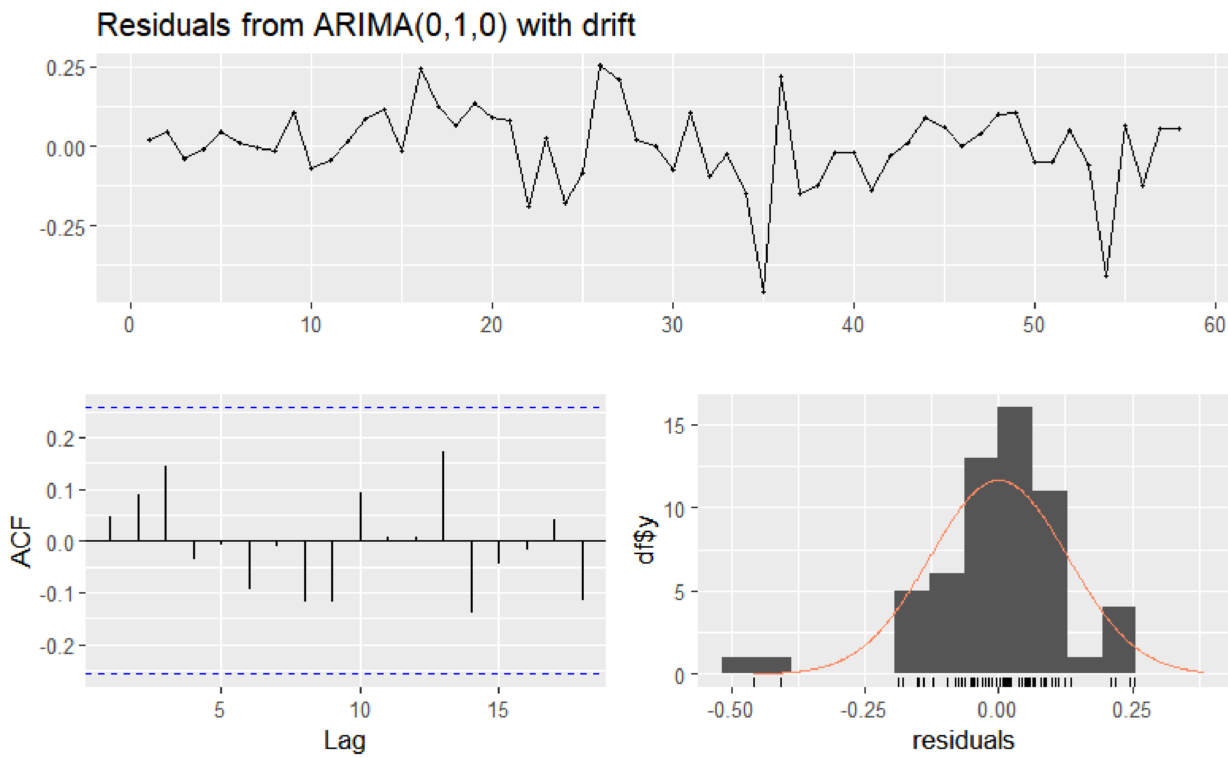

Figure 5. Plot (top), ACF plot (bottom left), and histogram (bottom right) of the residuals for final GDP model.

Similarly, the residuals for the imports data indicate a strong model fit. The residual plot (Figure 6) shows no visible structure or trend, and the histogram suggests a roughly normal distribution. The ACF of the residuals lacks significant spikes, again consistent with white noise behavior. The Ljung-Box test produces a p-value of 0.8357, further affirming that there is no meaningful autocorrelation remaining in the residuals. This confirms that the final model—ARIMA(0, 1, 2)—effectively captures the dynamics of the differenced imports data and is appropriate for forecasting.

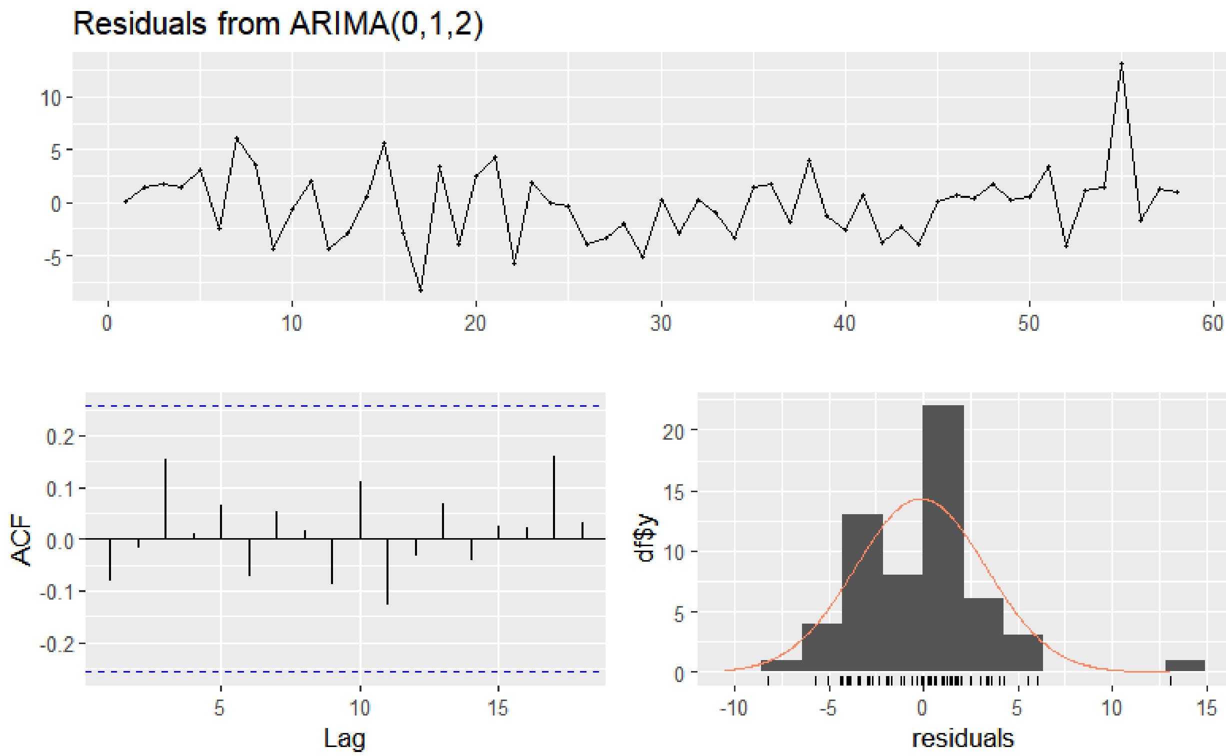

Figure 6. Plot (top), ACF plot (bottom left), and histogram (bottom right) of the residuals for final imports model.

3.3 Forecasts

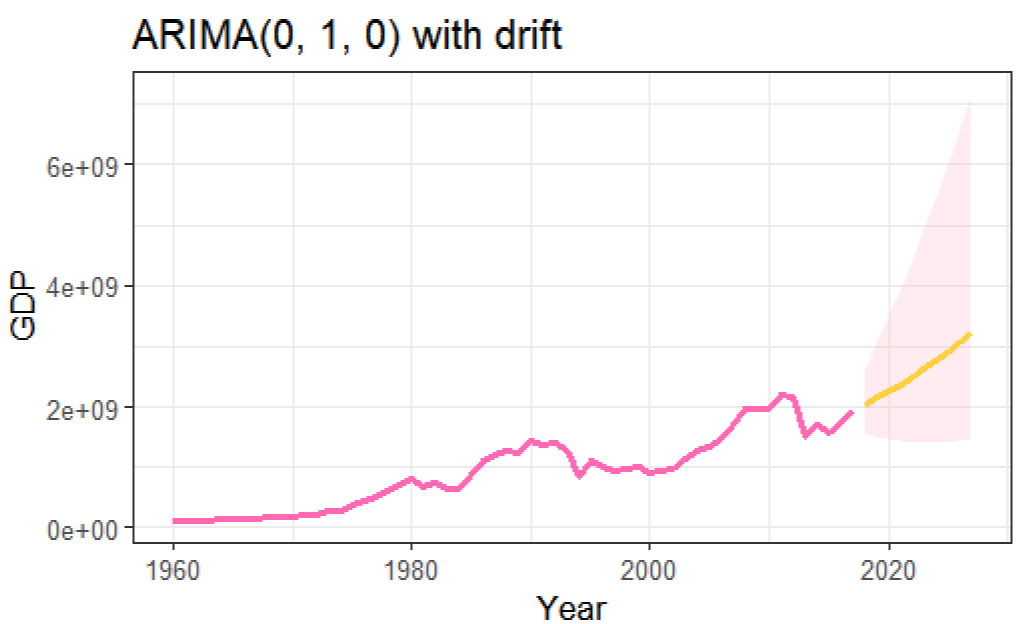

Figure 7. 10 year forecast for GDP.

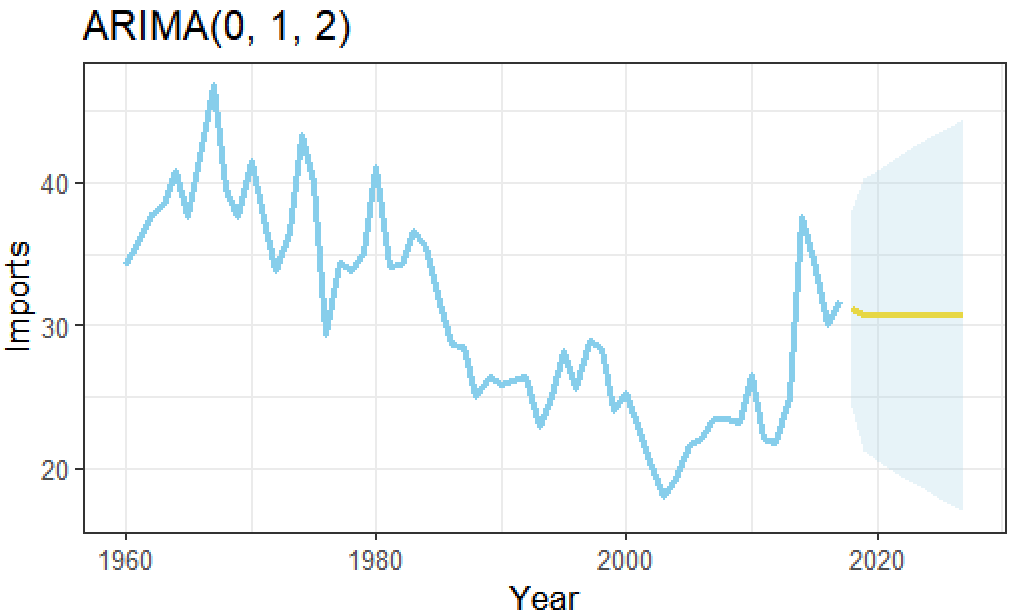

Figure 8. 10 year forecast for imports.

4. Discussion & conclusions

The GDP model, a random walk with drift applied to the log-differenced data, suggests that the Central African Republic's economic output follows a largely unpredictable path from year to year, but with a consistent positive average growth rate (Figure 7). This is a promising sign, especially given the CAR's status as one of the poorest countries in the world, as it indicates that the nation's economy is gradually expanding. The presence of drift in the model implies that there are structural factors that are contributing to long-term growth. While this upward trend does not guarantee stability, it provides an encouraging outlook for the country's economic future.

In contrast, the ARIMA(0, 1, 2) model for imports indicates that year-to-year changes in imports are neutral in overall direction but closely tied to the past two periods' irregularities or shocks (Figure 8). This pattern suggests a reactive import economy that is sensitive to recent economic conditions, reflecting a degree of ongoing instability. While imports can benefit a country by improving access to goods and lowering consumer prices, they can also suppress domestic industries and negatively affect GDP [4]. In the case of the CAR, the fact that imports show no consistent upward or downward trend may be beneficial. It implies that imports are not currently expanding in a way that might harm the country's fragile economic growth; however, they are not positively affecting it.

Overall, the economic outlook for the Central African Republic appears generally positive. The steady upward trend in GDP suggests that the country is making slow but meaningful progress in its economic development. While imports remain stable and responsive to short-term changes, their neutral trajectory indicates a need for continued economic stabilization. With continued effort, the CAR has the potential to continue economic growth in the years to come.

5. References

1. Mudge, Lewis (11 December 2018), "Central African Republic: Events of 2018", *World Report 2019: Rights Trends in Central African Republic*, archived from the original on 30 May 2019, retrieved 13 June 2024
2. World Economic Outlook Database, January 2018 Archived 3 May 2018 at the Wayback Machine, International Monetary Fund Archived 14 February 2006 at Archive-It. Database updated on 12 April 2017. Accessed on 21 April 2017.
3. Cherkaoui, Said El Mansour (1991). "Central African Republic". In Olson, James S. (ed.). *Historical Dictionary of European Imperialism*. Greenwood. p. 122. ISBN 0-313-26257-8.
4. Kramer, L. (2024, December 20). *How Importing and Exporting Impacts the Economy*. Investopedia. https://www.investopedia.com/articles/investing/100813/interesting-facts-about-imports-and-exports.asp#:~:text=A%20country's%20importing%20and%20exporting,on%20a%20country's%20exchange%20rate.