Forecasting Exports of the Central African Republic

STA 137 Time Series Project

Filip Wilhelm Sjostrand

2023-03-18

Contents

1 Introduction

Giles-Vernick, O'Toole, and Hoogstraten (2023) describe the Central African Republic as an agricultural state with abundant natural resources and excellent export potential that has reached a paradigm where diamonds stand for a majority of the exports. The authors claim poor governance and corruption have led to economic hardship since the late 80s, with inertia in maintaining the status quo. Yapatake Kossele, Ndjakwa Tonga, Anning, and Ngaba Mbai-Akem (2022) express this as a "resource curse"—a phenomenon resulting from abundant resources and poor governance, causing economic misfortune. They stress the importance of diamonds to the CAR's economy despite a decline since independence and argue for a stable political environment and an accountable diamond sector to improve the economy.

Given the background, we should consider some implications of the CAR's condition in our time series analysis. As mentioned, diamonds account for a significant portion of CAR's exports, which may make the country's export revenue vulnerable to fluctuations in the diamond market. Any shocks to the global demand or price of diamonds could significantly impact CAR's export earnings. Further, the CAR's economic development highly depends on the country's political and economic stability. Any instability, conflict, or civil unrest could disrupt the country's economic activities, including exports. Although diamonds are CAR's primary export, the country has great potential to export other agricultural commodities. Hence, the performance of the agricultural sector may also influence the country's export performance.

2 Method

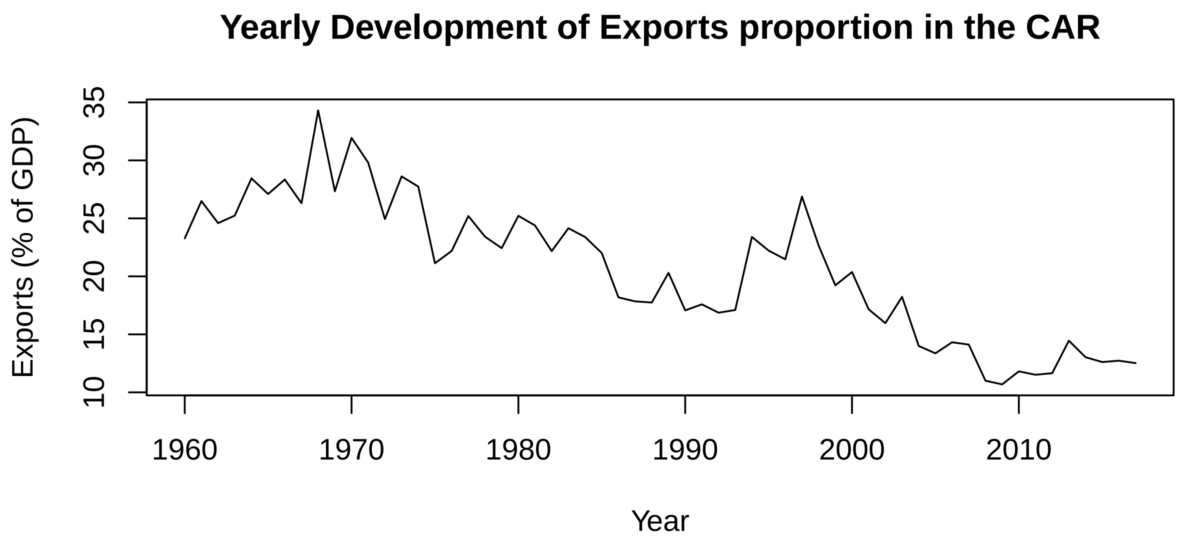
2.1 Data Descriptives

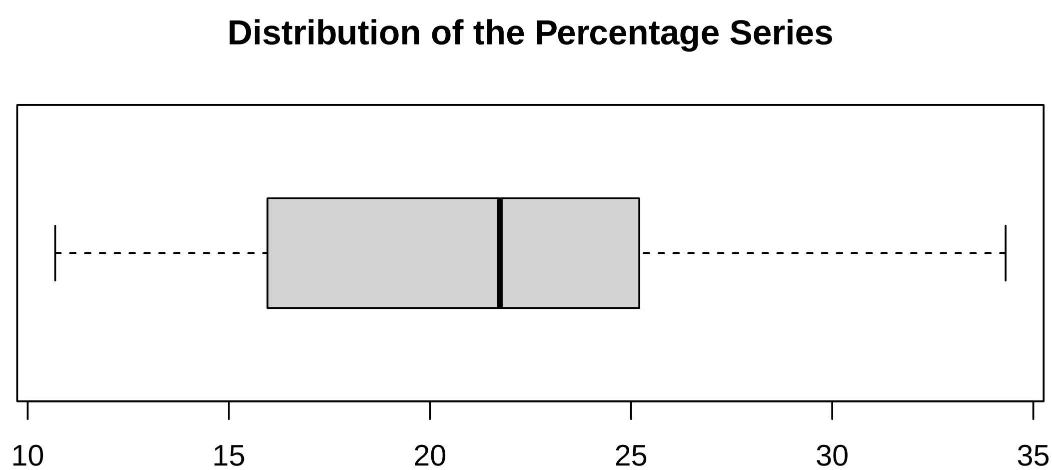
The first insight we make regards the property of the data. The export values are a percentage of the country's GDP (USD) (The World Bank, 2023). The report approaches the data by working with the percentage and monetary values. Regarding monetary values, operating with nominal values imposes an issue: what about the purchasing power? Since the nominal values do not account for inflation, we can get misguiding results. For example, exports can increase in nominal values but decrease in real terms if inflation is strong enough. Thus conversion is needed from nominal to real values. However, we face an obstacle: a lot of missing data for the CPI (our deflator). That raises the question of imputation.

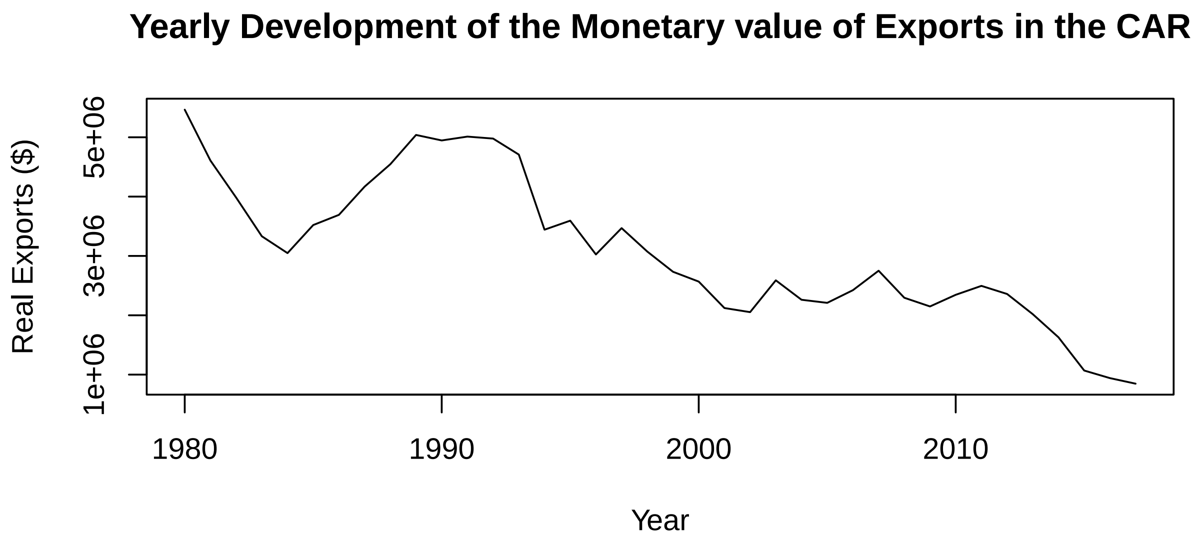
First, we must consider why the data may be missing. According to Harrell (2015), there are three types of missing data: missing completely at random (MCAR), missing at random (MAR), and informative missing (IM). He says MCAR occurs when data is missing for reasons unrelated to the subject's qualities or responses. Further, the author writes that MAR occurs when the probability of missing data depends on the observed data, not the unobserved data. Additionally, Harrell claims that IM occurs when unobserved data that may be related to the result of interest determines the likelihood of missing data. In the case of missing CPI data for the CAR, the missingness could, for instance, be MCAR if there is an issue with the instruments that record the CPI, MAR if the probability of CPI data being missing depends only on the governmental maturity and not on the actual inflation, and IM if the inflation is very high and therefore considered not necessary to record. The author states that multiple imputations are helpful if the missing data is MCAR or MAR; if the data is IM, we risk biasing whatever estimations we generate.

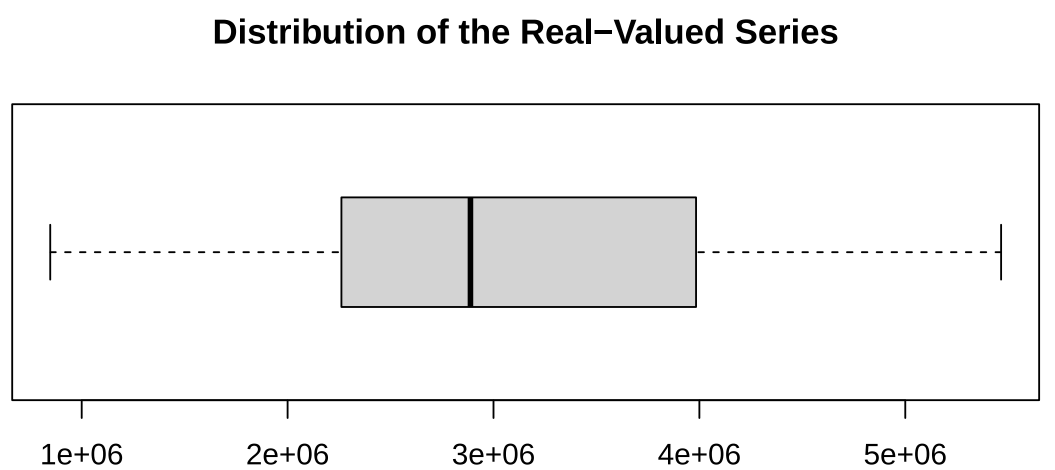
Three plausible reasons for the missing CPI data from 1960-1979 include a lack of resources, political instability, or limited data collection infrastructure. Therefore, it is difficult to conclude that the data is MCAR or MAR. In contrast, the missing CPI data from 2016-2017 is likely subject to a random event since the IMF had recently signed an Extended Credit Facility agreement to stabilize the economy and promote growth, implying rigorous CPI controls. Therefore, assuming that the data can be imputed for these years is safe. Consequently, we drop 1960-1979 and impute 2016-2017 CPI data. We recommend using Kalman smoothing to impute data with trend and seasonality (Moritz & Bartz-Beielstein, 2017).

In our plots, we can observe similarities. First, we find an apparent negative trend in both graphs. Such a trend implies that we are working with non-stationary data. We shall begin removing the trend using differencing. Further, we find in the boxplots that no outliers, although some skewness in each case, could negatively affect our analysis.



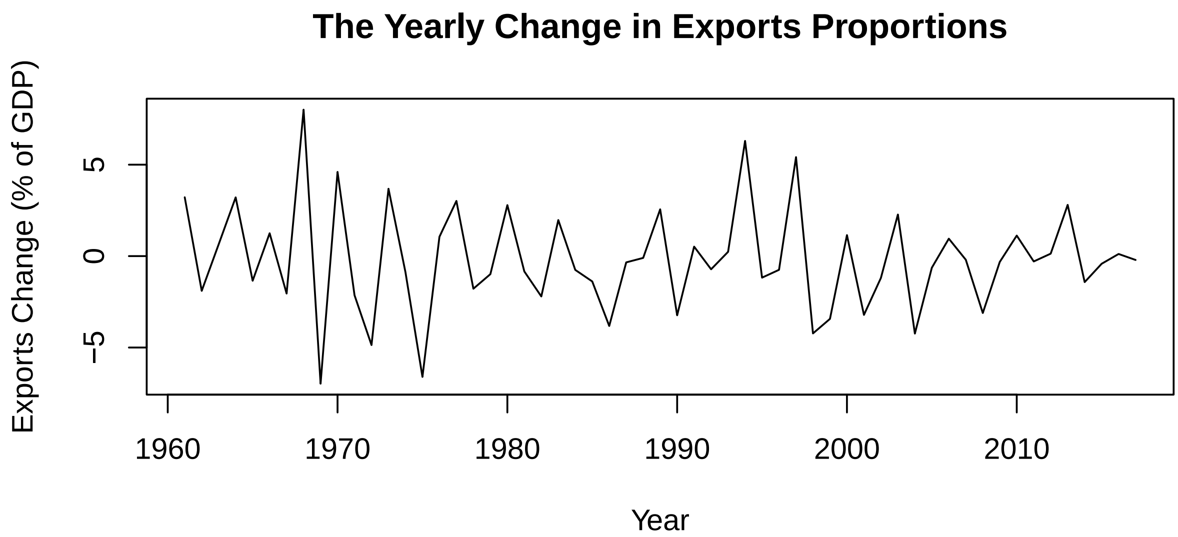


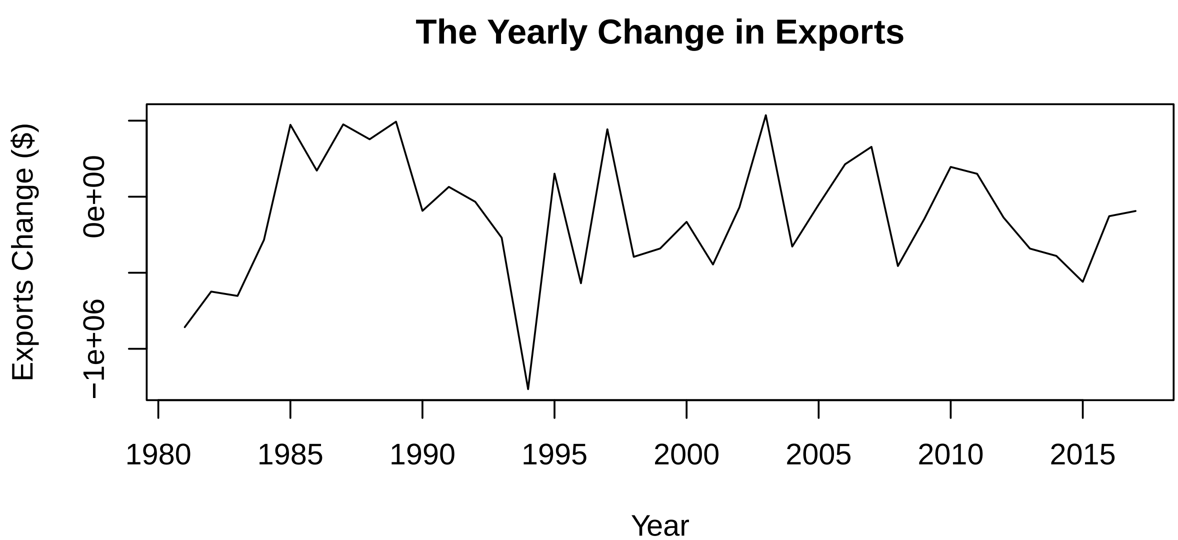


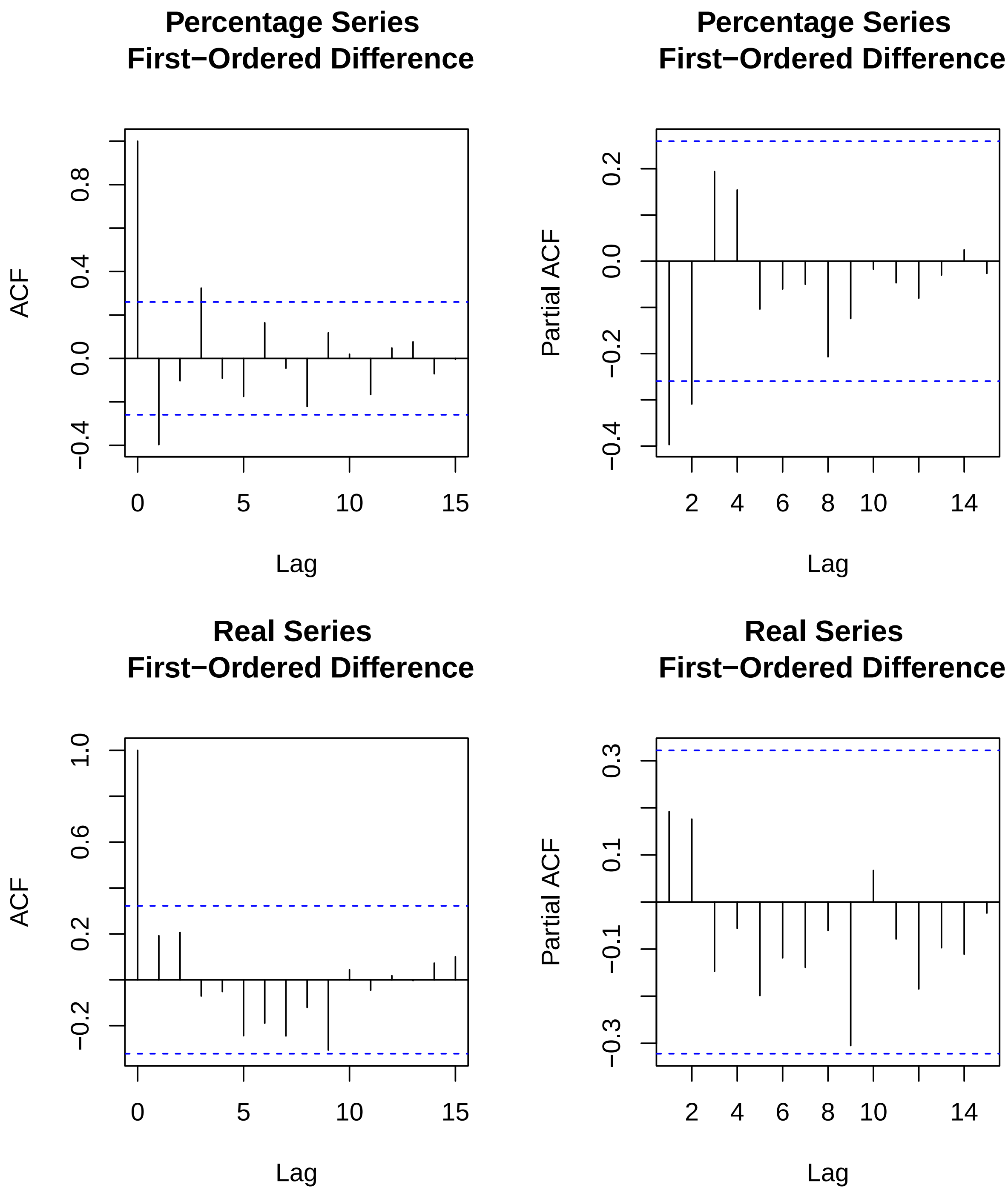


2.2 Transformations

First, we can observe that first-order differencing has removed any trend for both series. According to Nau (n.d.), a higher order of differencing is unnecessary if the autocorrelations are all small and patternless. For the percentage, the data is patternless but has a rather non-suggestive ACF and a PACF $= 0$ for $h > 2$, hence AR(2). For the real values, we have more of a pattern appearing but also no information from ACF/PACF. We might expect a second-order difference for the real values to provide better insight, which could also reduce the spikes in the time series plot.

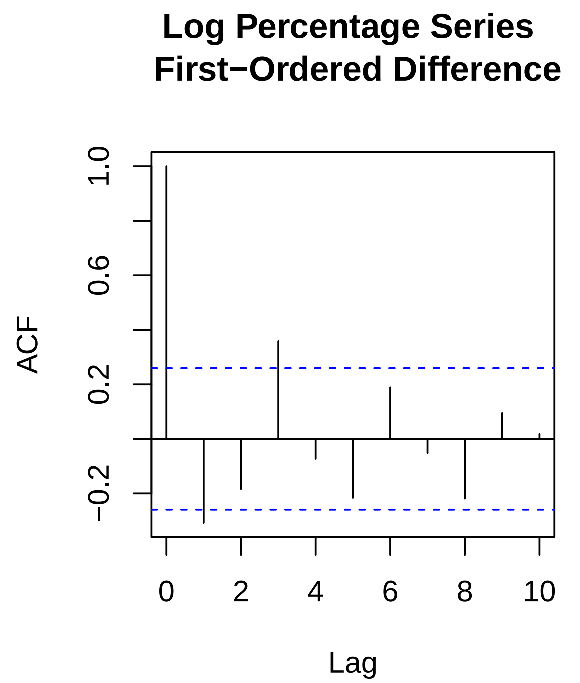

The Yearly Change in Exports Proportions
Exports Change (% of GDP)
5
0
−5
1960
1970
1980
1990
2000
2010
Year

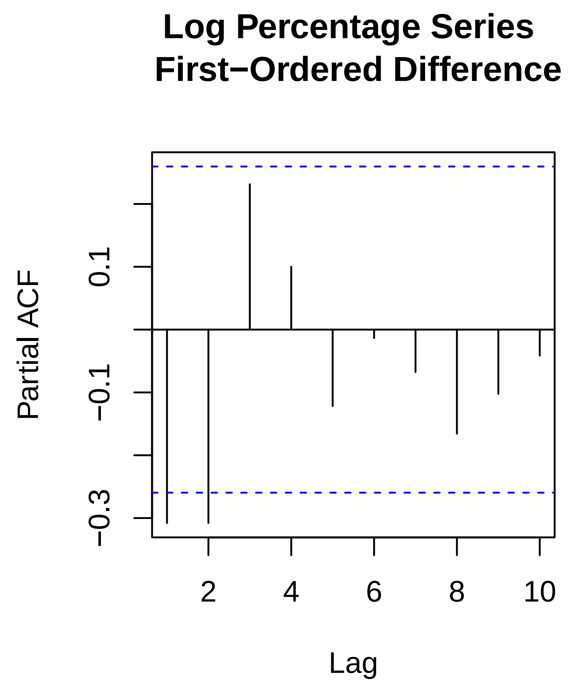

The Yearly Change in Exports
Exports Change ($)
0e+00
−1e+06
1980
1985
1990
1995
2000
2005
2010
2015
Year

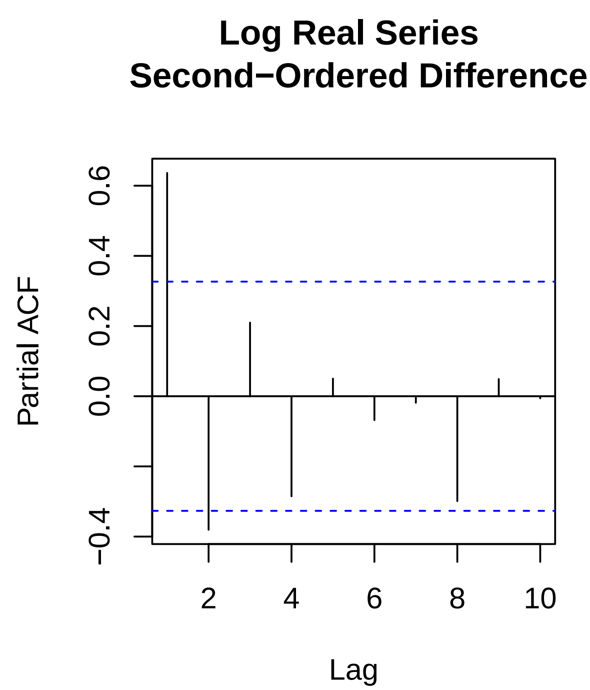


From the literature review, we find that, in general, Box-Cox transformations tend not to improve the analysis of macroeconomic data (Nelson & Granger, 1979) (Proietti & Lütkepohl, 2013). However, it appears as if power transformations are the most common method used (Cohen, 2001). Following Shumway & Stoffer (2017), we know that when data behaves as $x_t = (1 + p_t)x_{t-1}$, where p_t is a small percentage, then $\nabla log(x_t) \approx p_t$ is known as the growth rate. They claim that growth rates are often reported instead of actual values when reporting GNP and similar economic indicators. Therefore, the log transformation is a highly suitable option for our macroeconomic data.

The log transformation has provided some insight. It appears as if the percentage series is consistent with the AR(2) suggestion. Given that the real series was still not hinting at anything after the log transform, we decided to take the log transform of the second order difference and found that we have ACF $= 0$ for $h > 1$ and PACF $= 0$ for $h > 2$, i.e., MA(1) and (AR2).







2.3 diagnostics

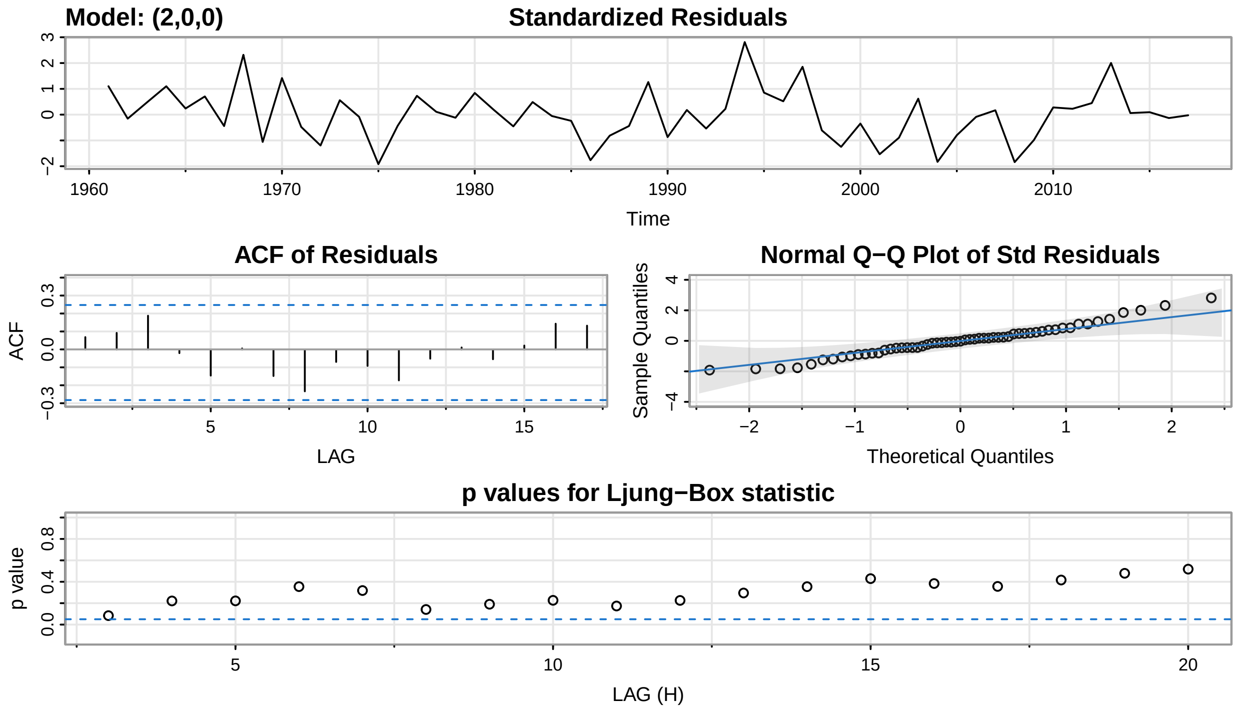
In each case we are testing the following hypothesis at $\alpha = 0.1$:

H_0 : The data are independently distributed

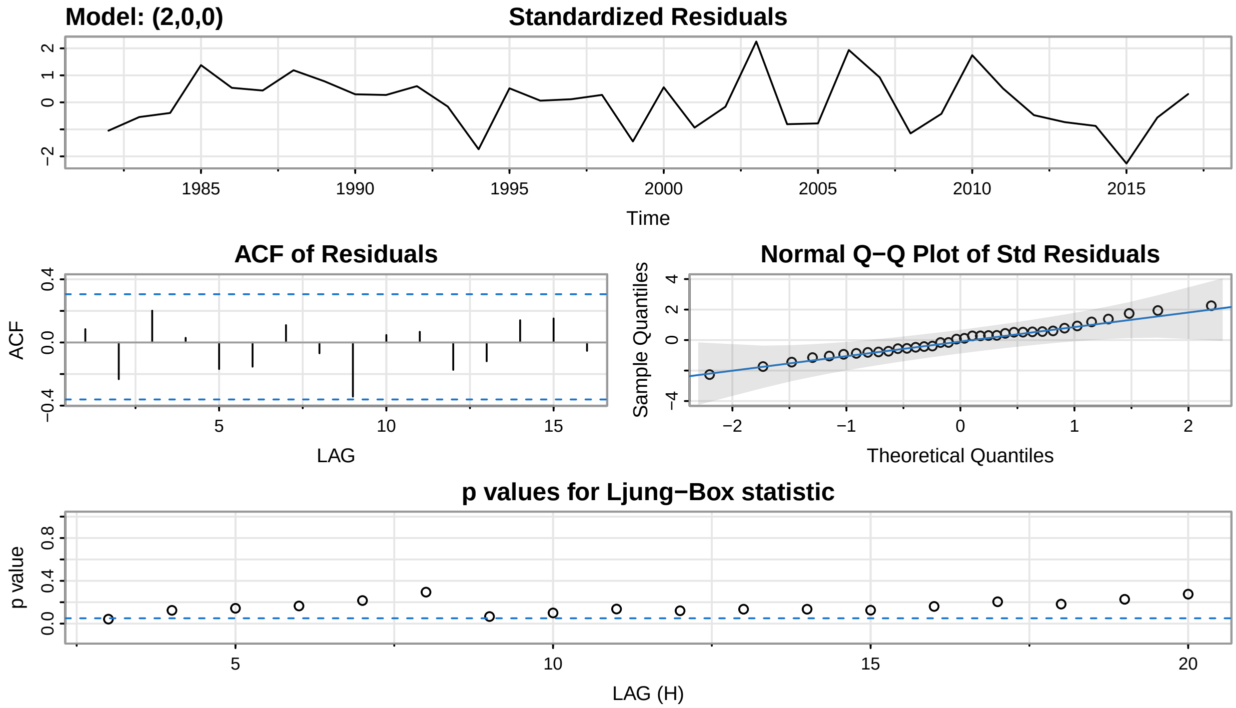
H_a : The data are not independently distributed; they exhibit serial correlation. (Wikipedia, 2023)

From each model, except possibly for the first residual of the real AR2 model, we fail to reject the null hypothesis. Thus, we conclude that each contender model's residuals are white noise. The residuals are approximately normally distributed, with no significant ACF in each case.

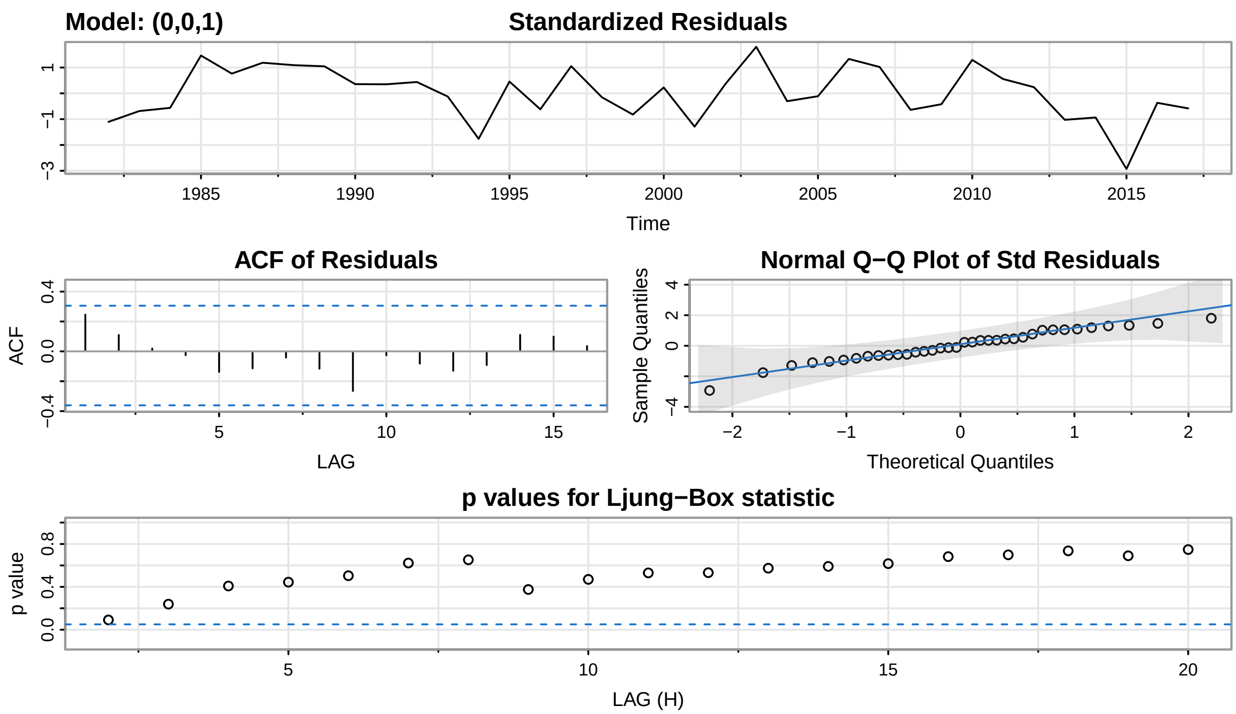
AR(2): Log Percentage



AR(2): Log Real



MA(1): Log Real



2.4 Parameter Estimation

In each case we are testing the following hypothesis at $\alpha = 0.1$:

H_0 : The parameter is zero

H_a : The parameter is non-zero

For the mean of the percentage series, we fail to reject the null. Thus, there is insufficient evidence that the mean is different from zero. In all other cases, we find sufficient evidence that each parameter is significant.

Table 1: p-values for parameters

	AR(2): Percentage	AR(2): Real	MA(1): Real
Coefficient 1	0.0023	0.0000	0.0000
Coefficient 2	0.0182	0.0211	NA
Mean	0.2027	0.0782	0.0299

2.5 Model Choice.

From the table below, we can derive that the AR(2) model based on the percentage series provides us with the lowest AIC and BIC values. It is the best model for our data.

Table 2: Selection Criterion

	AR(2): Percentage	AR(2): Real	MA(1): Real
AIC	-1.323425	-0.6800969	-0.8709509
BIC	-1.180053	-0.5041503	-0.7389910

3 Result & Conclusion

Our model looks like the following:

$$\begin{aligned} \text{Raw data:} \quad & x_t = \text{Export as percentage of total GDP} \\ \text{Transformation:} \quad & y_t = \log(x_t) - \log(x_{t-1}) \\ \text{Auto Regressive:} \quad & y_t = -0.4035 y_{t-1} - 0.3037 y_{t-2} + w_t \end{aligned}$$

and we find our forecast below. For each year, we observe that the confidence intervals are zero-containing and, therefore, insignificant. Hence, it is not different from zero.

Now, what does it imply? We can expect exports to explain the same proportion of GDP as the percentage data shows. This constant progression is reasonable from one perspective. As explained earlier, the country suffers from the resource curse, causing it to reach a plateau. This plateau has significantly affected the data, causing it to predict a few further changes.

On the other hand, Kouame (2022) explains that the CAR is abundant in its resources. He argues that we could expect immense growth with carefully designed and challenging reforms in the right area of the country's governance apparatus. Such analysis causes our prediction to be unstable. Therefore, the data tells the story of a double-edged sword. On the one hand, without further adjustment to the country's governance, we could expect few changes to its export capabilities, following the trend of that past data. On the other, with the right reformations of the country, we would see significant increases in the export sector. Thus, the forecast comes with great uncertainty and should be used short-term, given that any disruption to the paradigm will likely cause it to increase significantly—albeit unknown when.

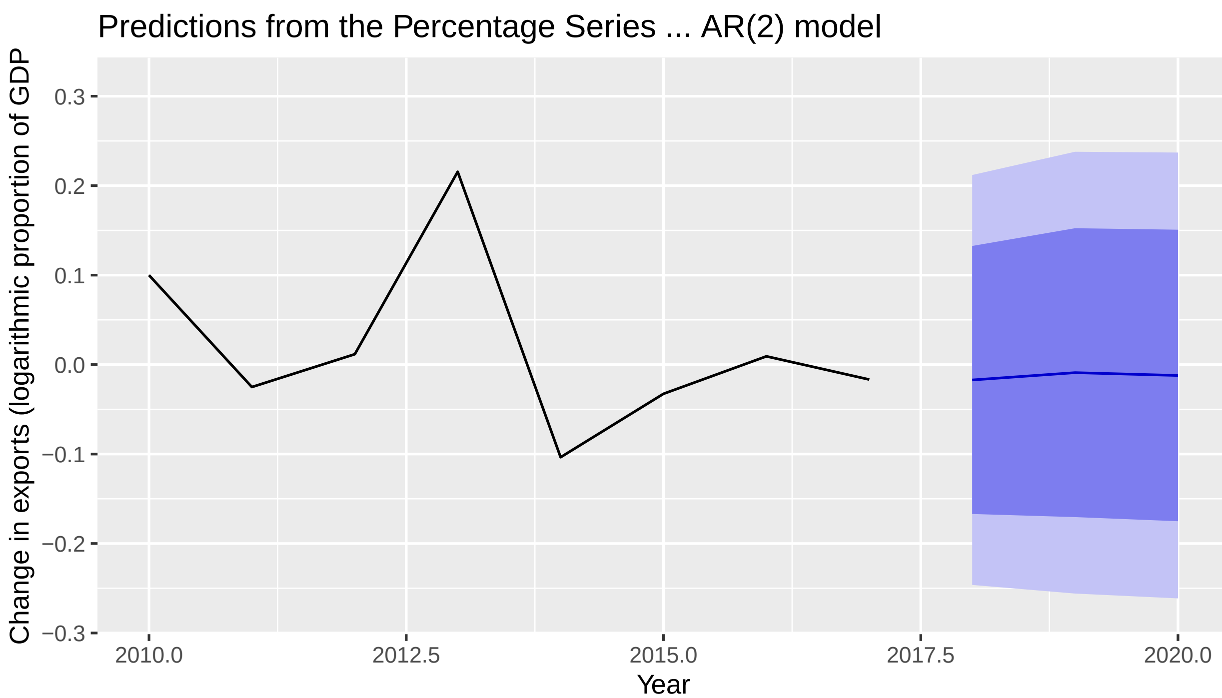


Table 3: Predictions and Confidence Intervals

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018	-0.0171864	-0.1669772	0.1326044	-0.2462717	0.2118989
2019	-0.0090087	-0.1704515	0.1524341	-0.2559142	0.2378968
2020	-0.0121503	-0.1750694	0.1507688	-0.2613135	0.2370129
2021	-0.0134037	-0.1785988	0.1517915	-0.2660479	0.2392405
2022	-0.0119331	-0.1771830	0.1533168	-0.2646610	0.2407949
2023	-0.0121386	-0.1775243	0.1532470	-0.2650741	0.2407969
2024	-0.0125085	-0.1779426	0.1529256	-0.2655181	0.2405011
2025	-0.0122965	-0.1777312	0.1531381	-0.2653071	0.2407140
2026	-0.0122679	-0.1777087	0.1531728	-0.2652877	0.2407518
2027	-0.0123447	-0.1777859	0.1530966	-0.2653653	0.2406759

4 Sources

Cohen, D. S. (2001). Linear Data Transformations Used in Economics, SSRN Scholarly Paper, 295624, Available Online: https://papers.ssrn.com/abstract=295624 [Accessed 13 March 2023]

Giles-Vernick, T. L., O'Toole, T. E. & Hoogstraten, J. S. F. van. (2023). Central African Republic, in Encyclopedia Britannica, Available Online: https://www.britannica.com/place/Central-African-Republic [Accessed 12 March 2023]

Harrell, Jr., Frank E. (2015). Regression Modeling Strategies. With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis., Springer Mathematics and Statistics EBooks 2015 English/International, 2nd ed. 2015., [e-book] Springer International Publishing, Available Through: ePublications https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat02271a&AN=atoz.ebs6971956e&site=eds-live&scope=site

International Monetary Fund (2015). IMF Executive Board Approves US$11.8 Million Disbursement Under the Rapid Credit Facility for the Central African Republic [Press release]. 15 September. Available at: https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr15417

Kouame. (2022). How the Central African Republic Can Move from Fragility to Inclusive Growth, The World Bank, Available Online: https://blogs.worldbank.org/africacan/how-central-african-republic-can-move-fragility-inclusive-growth [Accessed 12 March 2023]

Moritz, S. & Bartz-Beielstein, T. (2017). ImputeTS: Time Series Missing Value Imputation in R., R Journal, [e-journal] vol. 9, no. 1, pp.207–218, Available Through: Academic Search Complete https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=a9h&AN=123900019&site=eds-live&scope=site

Nau, bob. (n.d.). Identifying the Order of Differencing in ARIMA Models, Available Online: https://people.duke.edu/~rnau/411arim2.htm [Accessed 14 March 2023]

Nelson, H. L. & Granger, C. W. J. (1979). Experience with Using the Box-Cox Transformation When Forecasting Economic Time Series, Journal of Econometrics, [e-journal] vol. 10, no. 1, pp.57–69, Available Online: https://www.sciencedirect.com/science/article/pii/0304407679900642

Proietti, T. & Lütkepohl, H. (2013). Does the Box–Cox Transformation Help in Forecasting Macroeconomic Time Series?, International Journal of Forecasting, [e-journal] vol. 29, no. 1, pp.88–99, Available Through: ScienceDirect https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0169207012000830&site=eds-live&scope=site

Shumway, R. H. & Stoffer, D. S. (2017). Time Series Analysis and Its Applications With R Examples, 4th edn, Springer

The World Bank. (2023). Exports of Goods and Services (% of GDP) - Central African Republic | Data, Available Online: https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?locations=CF [Accessed 14 March 2023]

Yapatake Kossele, T. P., Ndjakwa Tonga, M., Anning, L. & Ngaba Mbai-Akem, M. G. (2022). Rich and Poor: How Political Governance Crises Have Affected the Relationship between Diamond Exports and Economic Growth in the Central African Republic, Politics & Policy, [e-journal] vol. 50, no. 2, pp.441–464, Available Online: https://doi.org/10.1111/polp.12459 [Accessed 14 March 2023]

Wikipedia. (2023). Ljung–Box Test, in Wikipedia, Available Online: https://en.wikipedia.org/w/index.php?title=Ljung%E2%80%93Box_test&oldid=1131351701 [Accessed 15 March 2023]

5 Code

```r
# Libraries -----------
library(astsa)
library(dplyr)
library(forecast)
library(ggplot2)
library(knitr)
library(imputeTS)

# Chunk settings ----------
opts_chunk$set(
  echo = FALSE,
  message = FALSE,
  error = FALSE,
  fig.height = 4,
```

```
  fig.width = 7,
  warning = FALSE
  )

# Data ----------
load("/Users/filipsjostrand/Documents/UC Davis/Courses/STA 137/Project/finalproject.Rdata")

percentage <- ts(finalPro_data$Exports, frequency=1, start=c(1960))

imp <- na_kalman(finalPro_data %>% filter(Year > 1979))
real_exports <- imp %>%
  mutate(nominal_exports = Exports/100 * GDP) %>%
  transmute(real_exports = nominal_exports/CPI)
real <- ts(real_exports, frequency = 1, start = 1980)


# Initial plots ----------
par(mfrow=c(2,1))
ts.plot(percentage,
        xlab = "Year",
        ylab = "Exports (% of GDP)",
        main = "Yearly Development of Exports proportion in the CAR"
        )
boxplot(percentage,
        main = "Distribution of the Percentage Series",
        horizontal=TRUE
        )

ts.plot(real,
        xlab = "Year",
        ylab = "Real Exports ($)",
        main = "Yearly Development of the Monetary value of Exports in the CAR"
        )
boxplot(real,
        main = "Distribution of the Real-Valued Series",
        horizontal=TRUE
        )

# Differencing & ACF/PACF ----------
percentage_diff <- diff(percentage, 1)
real_diff <- diff(real, 1)

par(mfrow = c(2,1))
ts.plot(percentage_diff,
        xlab = "Year",
        ylab = "Exports Change (% of GDP)",
        main = "The Yearly Change in Exports Proportions"
        )
ts.plot(real_diff,
        xlab = "Year",
        ylab = "Exports Change ($)",
        main = "The Yearly Change in Exports"
        )
```

```r
par(mfrow = c(1,2))
acf(percentage_diff,
    lag.max = 15,
    main = paste("Percentage Series", "\n", "First-Ordered Difference")
    )
pacf(percentage_diff,
     lag.max = 15,
     main = paste("Percentage Series", "\n", "First-Ordered Difference")
     )

acf(real_diff,
    lag.max = 15,
    main = paste("Real Series", "\n", "First-Ordered Difference")
    )
pacf(real_diff,
     lag.max = 15,
     main = paste("Real Series", "\n", "First-Ordered Difference")
     )

# Log transform ----------
percentage_log <- diff(log(percentage))
real_log <- diff(log(real), 2)

par(mfrow=c(1,2))
acf(percentage_log,
    lag.max = 10,
    main = paste("Log Percentage Series", "\n", "First-Ordered Difference")
    )
pacf(percentage_log,
     lag.max = 10,
     main = paste("Log Percentage Series", "\n", "First-Ordered Difference")
     )
acf(real_log,
    lag.max = 10,
    main = paste("Log Real Series", "\n", "Second-Ordered Difference")
    )
pacf(real_log,
     lag.max = 10,
     main = paste("Log Real Series", "\n", "Second-Ordered Difference")
     )

# Diagnostics ----------
diagnostic_percentage_AR2 <- sarima(percentage_log, 2,0,0)

diagnostic_real_AR2 <- sarima(real_log, 2,0,0)

diagnostic_real_MA1 <- sarima(real_log, 0,0,1)

# p-value table
pp2 <- diagnostic_percentage_AR2$ttable[,4]
pr2 <- diagnostic_real_AR2$ttable[,4]
pr1 <- diagnostic_real_MA1$ttable[,4]
pr1 <- append(pr1, NA)
```

```r
tmp_row  <- pr1[2]
pr1[2]  <- pr1[3]
pr1[3]  <- tmp_row

df <- data.frame(pp2, pr2, pr1) %>%
  rename(`AR(2): Percentage` = pp2, `AR(2): Real` = pr2, `MA(1): Real` = pr1)
rownames(df) <- c("Coefficient 1", "Coefficient 2", "Mean")

kable(df, caption = "p-values for parameters")

# Decision criteria ----------
df1 <- data.frame(diagnostic_percentage_AR2$AIC) %>%
  rbind(diagnostic_percentage_AR2$BIC)
df2 <- data.frame(diagnostic_real_AR2$AIC) %>%
  rbind(diagnostic_real_AR2$BIC)
df3 <- data.frame(diagnostic_real_MA1$AIC) %>%
  rbind(diagnostic_real_MA1$BIC)
df <- df1 %>% cbind(df2) %>% cbind(df3)
colnames(df) <- c("AR(2): Percentage", "AR(2): Real", "MA(1): Real")
rownames(df) <- c("AIC", "BIC")

kable(df, caption = "Selection Criterion")

# Model coefficients ----------
ar1 <- round(diagnostic_percentage_AR2$fit$coef[[1]], 4)
ar2 <- round(diagnostic_percentage_AR2$fit$coef[[2]], 4)

# Prediction ----------
model <- ar.ols(percentage_log, order = 2, demean = FALSE, intercept = TRUE)
pred <- forecast(model)
autoplot(pred, xlim = c(2010, 2020)) +
  labs(title = "Predictions from the Percentage Series - AR(2) model") +
  xlab("Year") +
  ylab("Change in exports (logarithmic proportion of GDP")
kable(pred, caption = "Predictions and Confidence Intervals")
```