Central African Republic Exports Analysis

Diego Alatorre, Aria Hamidi
University of California, Davis - Statistics Department[1]
dialatorre@ucdavis.edu, hamidi@ucdavis.edu

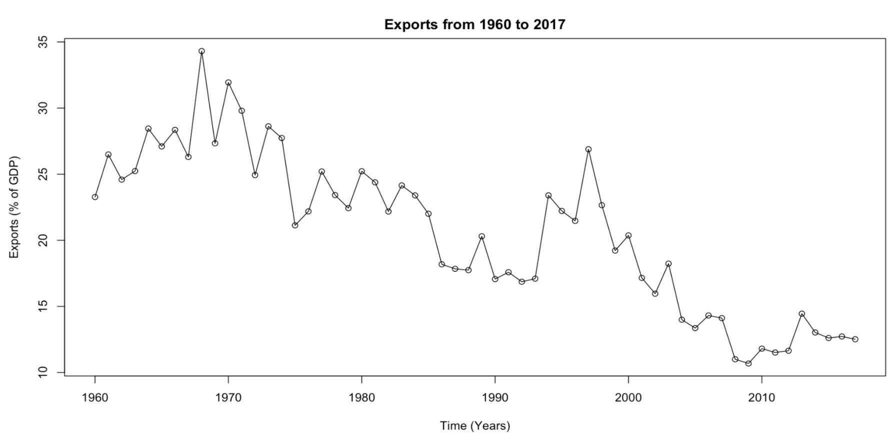
I. Introduction

The purpose of this project is to analyze the Central African Republic Exports time series in order to select an ideal model for forecasting purposes. For this analysis we used a time series data set that contains 58 observations and 9 variables (country, code, year, GDP, growth, CPI, imports, exports, and population) for the Central African Republic. However, we only focused on the exports variable, which contains information regarding exports as a percentage of GDP from 1960 to 2017.

The Central African Republic has faced export issues, it was exporting much more during the 1970s than what they currently are. According to Wilfried Kouame, this economic decline happens because of "the formation of political and armed groups that act as roving bandits, impeding anyone from producing and investing securely," and because the "political elites have been exploiting economic inequalities … to undermine social cohesion [and] capitalize on local grievances." Therefore, forecasting its exports for the next couple of years is important to see if it is expected to see further decline in exports or an increase in exports. This will allow us to see if the Central African Republic is becoming more or less reliant on exports and if its economy is improving or not.

II. Exploratory data analysis

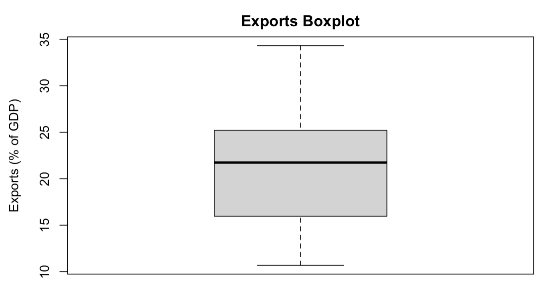
Before beginning to analyze the data, we explored the data visually and numerically. The first step was to plot the time series to check for any trends. The resulting plot is shown below.

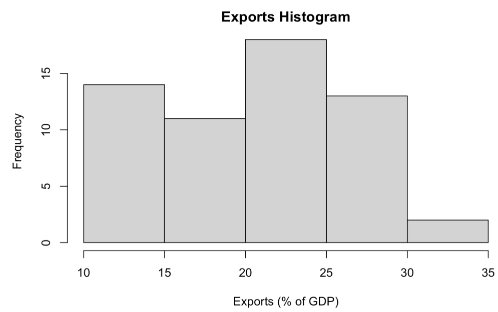


[1] Remark: Our report is 3 pages long without the figures. We just decided to include the plots throughout the project for clarity and readability purposes. Not all output is provided in the report. To see everything refer to the appendix which contains all our code and outputs.

From this plot, there appears to be a decreasing trend in exports over the years. This might be indicative that the time series is non-stationary. To verify this, we also plotted the ACF and PACF (provided in the appendix). The ACF shows an exponential decay which confirms that the time series is not stationary. Since it is non-stationary, we will have to do a transformation or do differencing to stabilize the variance and get stationary results.

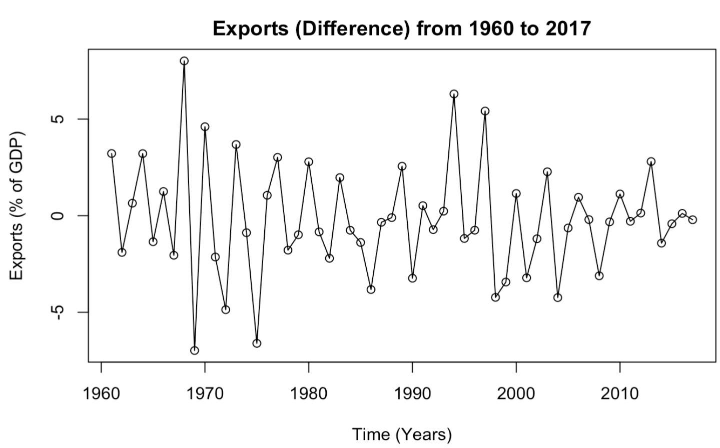
Before trying different transformations, we decided to further explore the data by computing summary statistics. From the summary statistics, we learned that the median for exports is 21.74% while the mean is 20.66%. In addition, we did not find any missing values and created the following boxplot and histogram that shows that there are no outliers in the data.



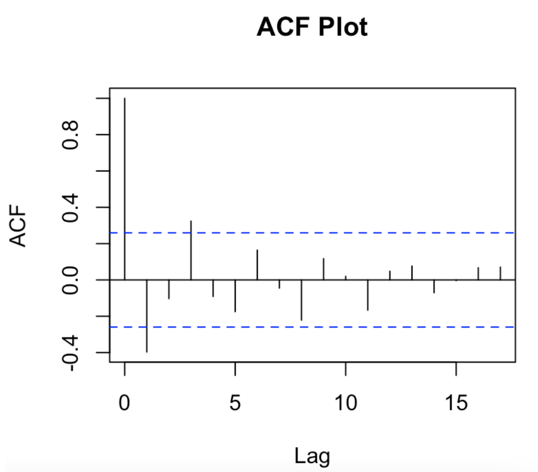


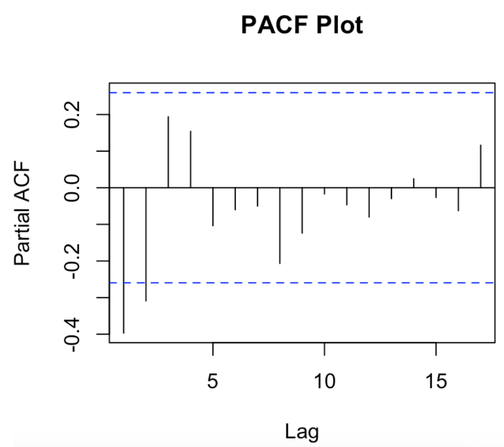
III. Model Selection

Prior to applying ARIMA, we tried some transformations on the data. We first tried a log transformation, but doing so did not yield the results we wanted. After applying the log transformation our ACF and PACF plots (shown in appendix) looked very similar to the ones without any transformations. The ACF still was exponentially decaying. Therefore, we decided to do first order differencing. Differencing proved to be a better alternative because as shown in the plot below, it stabilized the variance, removed the trend, and resolved the non-stationary issue.



Given that we are dealing with economic data, we decided to explore by doing first order differencing on the log-transformed data too. The resulting time series, ACF, and PACF plots were similar to the results of doing differencing without log transforming the data. Since the transformation was not helping much and in practice we prefer simplicity, we decided to use the model with first order differencing on the regular data (without log transformation). Doing so yielded the following ACF and PACF plots.





After doing differencing, we get improved ACF and PACF plots that follow what we expected. From the ACF plot we see that an ARIMA(0,1,3) model might be appropriate because it is zero after the third lag, while from the PACF it looks like an ARIMA(2,1,0) model would be appropriate because the lags are zero after lag 2.

Therefore we decided to fit both of these models. We first fitted the ARIMA(0,1,3) model to estimate the coefficients. As a result, we got the following model:

$$Z_t = -0.1999 - 0.4537w_{t-1} + 0.0922w_{t-2} + 0.2677w_{t-3} + w_t$$

For this model however, only -0.4537 and 0.2677 are significant because their p-values (0.0011 and 0.0533 respectively) are smaller than 0.1, so we reject the null hypothesis that the coefficient is equal to 0. The remaining terms have a p-value of 0.5 which is greater than 0.1 so they are not significant, we fail to reject the null hypothesis that the coefficient is equal to 0. After dropping the insignificant coefficients we get the following model:

$$Z_t = -0.4537w_{t-1} + 0.2677w_{t-3} + w_t$$

Next, we fitted the ARIMA(2,1,0) model and got the following model as a result:

$$Z_t = -0.2120 - 0.5230z_{t-1} - 0.3065z_{t-2} + w_t$$

In this case, only -0.5230 and - 0.3065 are significant because their p-values (0.0001 and 0.0173 respectively) are smaller than 0.1, so we reject the null hypothesis that the coefficient is equal to 0. The mean has a p-value (0.2546) greater than 0.1 so it is not significant, we fail to reject the null hypothesis that the coefficient is equal to 0. Therefore, we can drop the mean and now get that the actual ARIMA(2,1,0) model is:

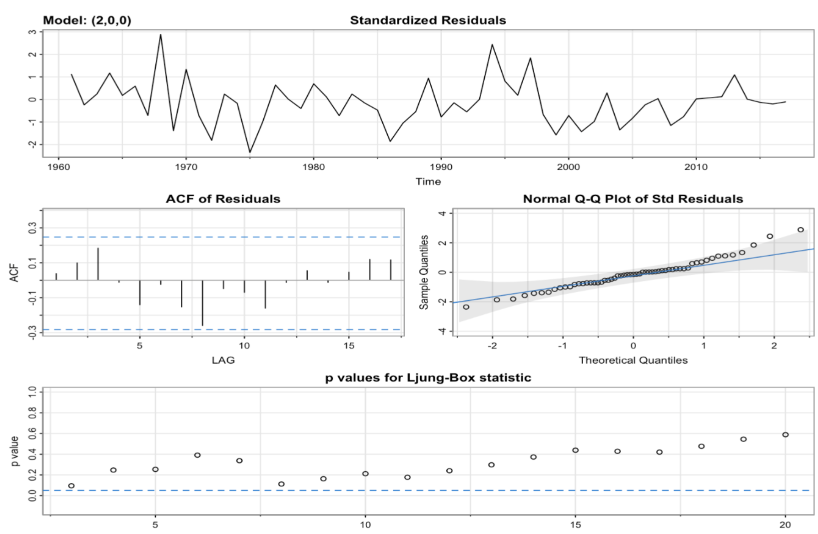
$$Z_t = -0.5230 z_{t-1} - 0.3065 z_{t-2} + w_t$$

To then figure out which model is better, we compared each model's AIC and BIC using the table below:

	ARIMA(2,1,0)	ARIMA(0,1,3)
AIC	4.829009	4.838539
BIC	4.972381	5.017754

Since as shown in the table above, the ARIMA(2,1,0) model has a smaller AIC and BIC, we chose to use the ARIMA(2,1,0) model. Additionally, AIC and BIC are similar, but in practice we prefer ARIMA(2,1,0) because it only has two unknown parameters instead of three, it is simpler.

IV. Residual Diagnostics and Model Interpretation



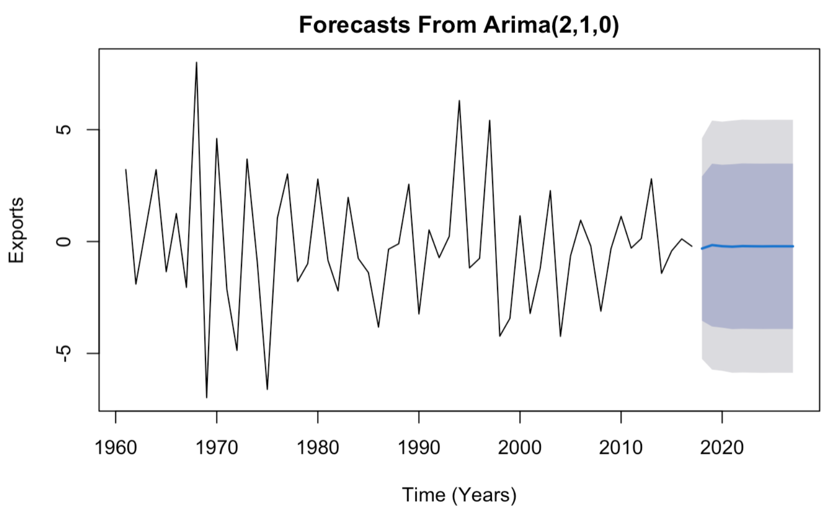
Once we selected the ARIMA(2,1,0) model, we conducted residual diagnostics and got the results shown above. From the Normal Q-Q plot we see that the residuals are normal as they follow a linear trend. In addition, from the ACF of the Residuals we see that all the lags are between the dashed lines which indicates that the residuals follow white noise. This is confirmed by the Ljung-Box test in which we see that all p-values are above the dashed line. This means that all p-values are greater than 0.1, which means that we fail to reject the null hypothesis that residuals follow white noise. Thus, we conclude that the residuals do follow white noise. These findings confirm that our model is appropriate and ready to be used for forecasting.

V. Forecast

Since we selected ARIMA(2,1,0) as the more reasonable model, we used it to forecast the exports for the next 10 years. We obtained the point forecasts and the corresponding confidence intervals for the next 10 years up until 2027. All forecasts are negative with most having values around -0.21 as shown below:

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
2018	−0.3133799	−3.536014	2.909254	−5.241973	4.615213
2019	−0.1592450	−3.796054	3.477564	−5.721264	5.402774
2020	−0.2084660	−3.846826	3.429894	−5.772858	5.355926
2021	−0.2299685	−3.913053	3.453116	−5.862760	5.402823
2022	−0.2036345	−3.896364	3.489095	−5.851177	5.443908
2023	−0.2108169	−3.903721	3.482087	−5.858626	5.436993
2024	−0.2151324	−3.909405	3.479140	−5.865034	5.434770
2025	−0.2106736	−3.905180	3.483833	−5.860933	5.439586
2026	−0.2116829	−3.906200	3.482835	−5.861960	5.438594
2027	−0.2125217	−3.907081	3.482037	−5.862862	5.437818

Moreover, for better understanding and visualization of the future trend of exports, we decided to plot the forecasts. The gray area represents the 95% confidence interval, while the blue area represents the 80% confidence interval.



As we can observe, there is an initial small decline. However, overall there is a more stabilized pattern in the forecasted exports for the next 10 years, from 2017 to 2027. This means that the Central African Republic should not expect any significant fluctuations in their exports for the upcoming years. From these forecasts, we see that the amount of exports is going to be more stable and predictable than in the past where they suffered significant increases and decreases. This is positive because the Central African Republic, which greatly depends on exports, will not suffer any important declines in exports. The reason for this stability is because the Central African Republic has important exporting partners. According to oec.world, they mainly export to China which has a strong economy. They also export to other economic powers such as the United Arab Emirates and France. In addition, according to the Macrotrends.net website, since 2017 actual exports did not have a lot of growth nor declines. The exports have stabilized which confirms that our forecasts were accurate and that our model performed well.

References

"Central African Republic (CAF) Exports, Imports, and Trade Partners." *OEC*, https://oec.world/en/profile/country/caf.

"Central African Republic Exports 1960-2023." *MacroTrends*, https://www.macrotrends.net/countries/CAF/central-african-republic/exports.

Kouame, Wilfried A. "How the Central African Republic Can Move from Fragility to Inclusive Growth." *World Bank Blogs*, https://blogs.worldbank.org/africacan/how-central-african-republic-can-move-fragility-inclusive-growth#:~:text=Weak%20governance%20and%20judicial%20services,1990%20to%20%24414%20in%202020.

Appendix

```r
# load in data
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
library(astsa)
```

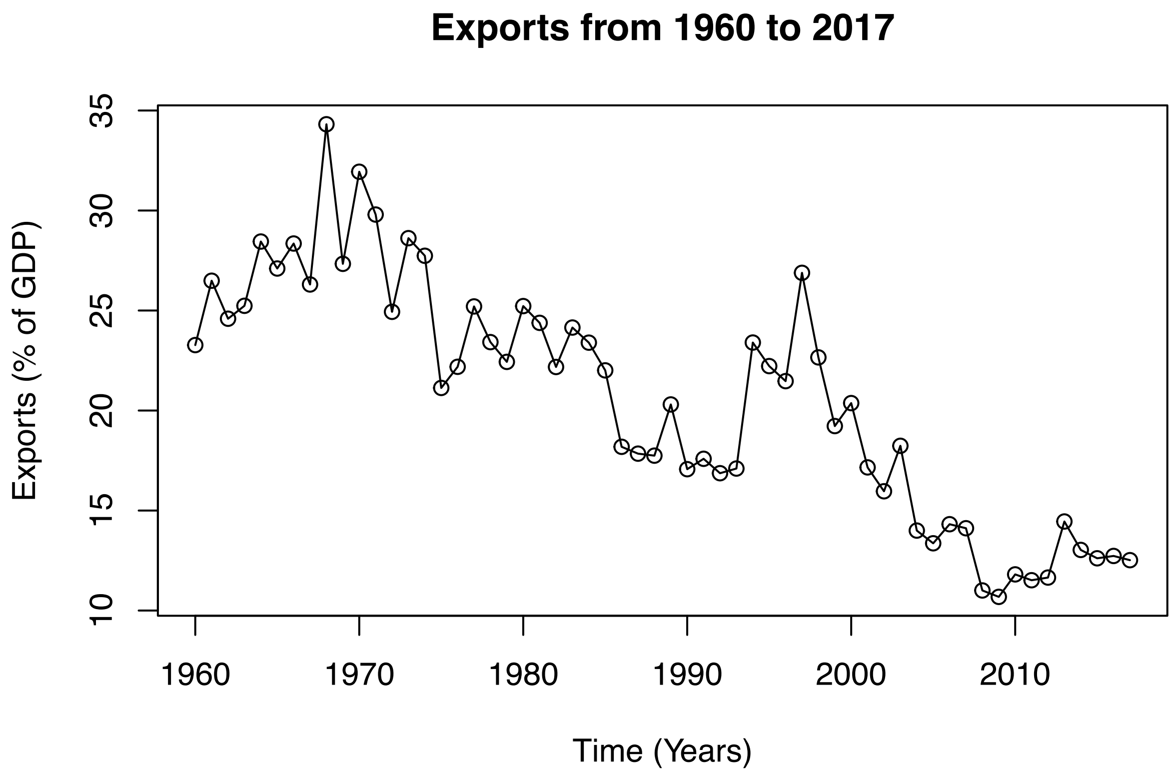
```
##
## Attaching package: 'astsa'

## The following object is masked from 'package:forecast':
##
##     gas
```

```r
load("/Users/diegoalatorre/Downloads/Sta 137/Project/finalproject.Rdata")
data  = finalPro_data[8] # only look at exports variable, drop the rest
head(data)
```

```
## # A tibble: 6 x 1
##   Exports
##     <dbl>
## 1    23.3
## 2    26.5
## 3    24.6
## 4    25.2
## 5    28.4
## 6    27.1
```

```r
# create time series data
ts_data = ts(data, start=1960, end=2017, frequency=1)
plot(ts_data, type = 'o',  main = "Exports from 1960 to 2017", ylab = "Exports (% of GDP)", xlab = "Tim
```



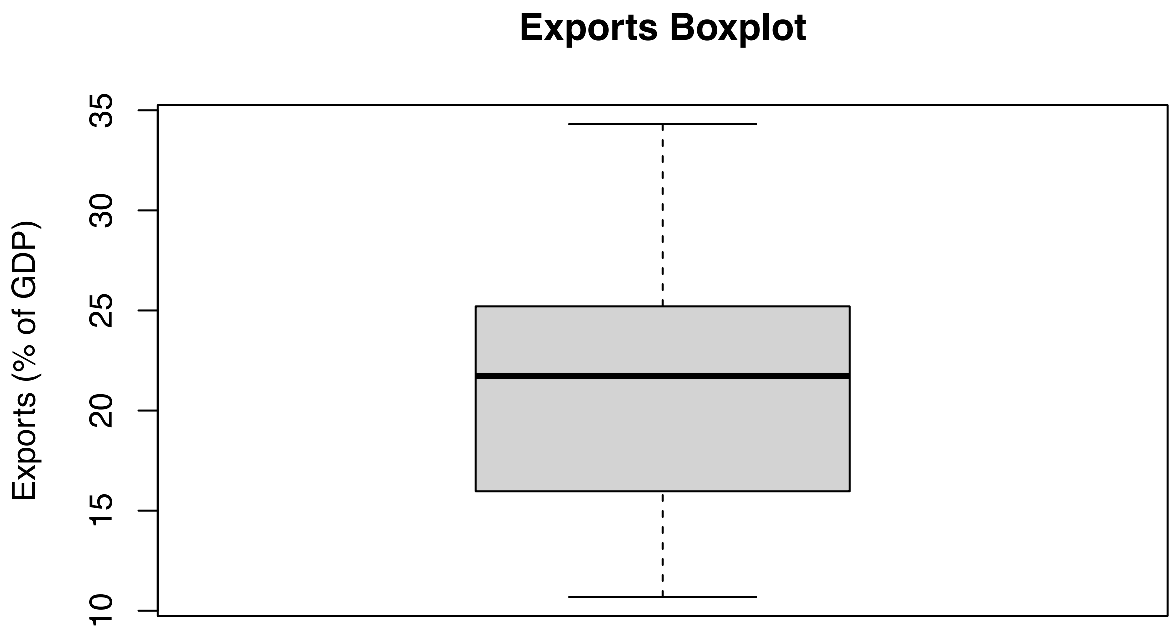

```
summary(data$Exports)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.68   16.19   21.74   20.66   25.14   34.31
```

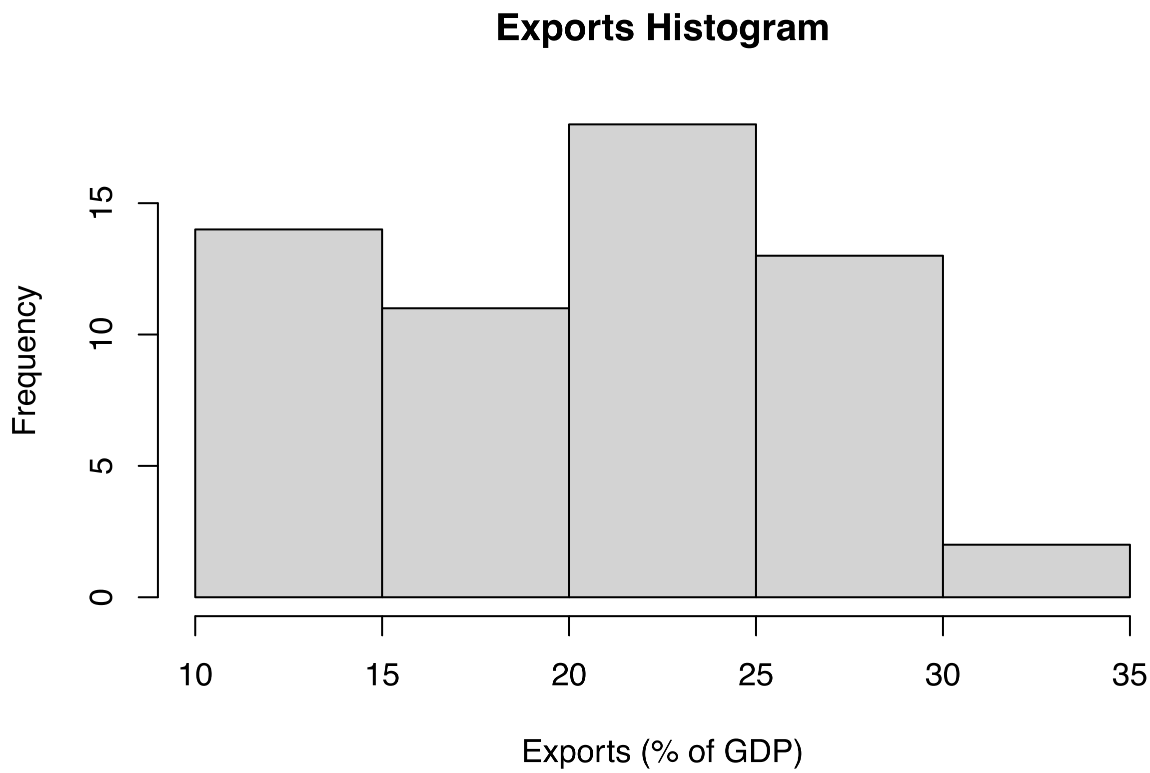
```
which(is.na(data)) # no missing value
```

```
## integer(0)
```

```
boxplot(data$Exports, main = "Exports Boxplot", ylab = "Exports (% of GDP)")
```

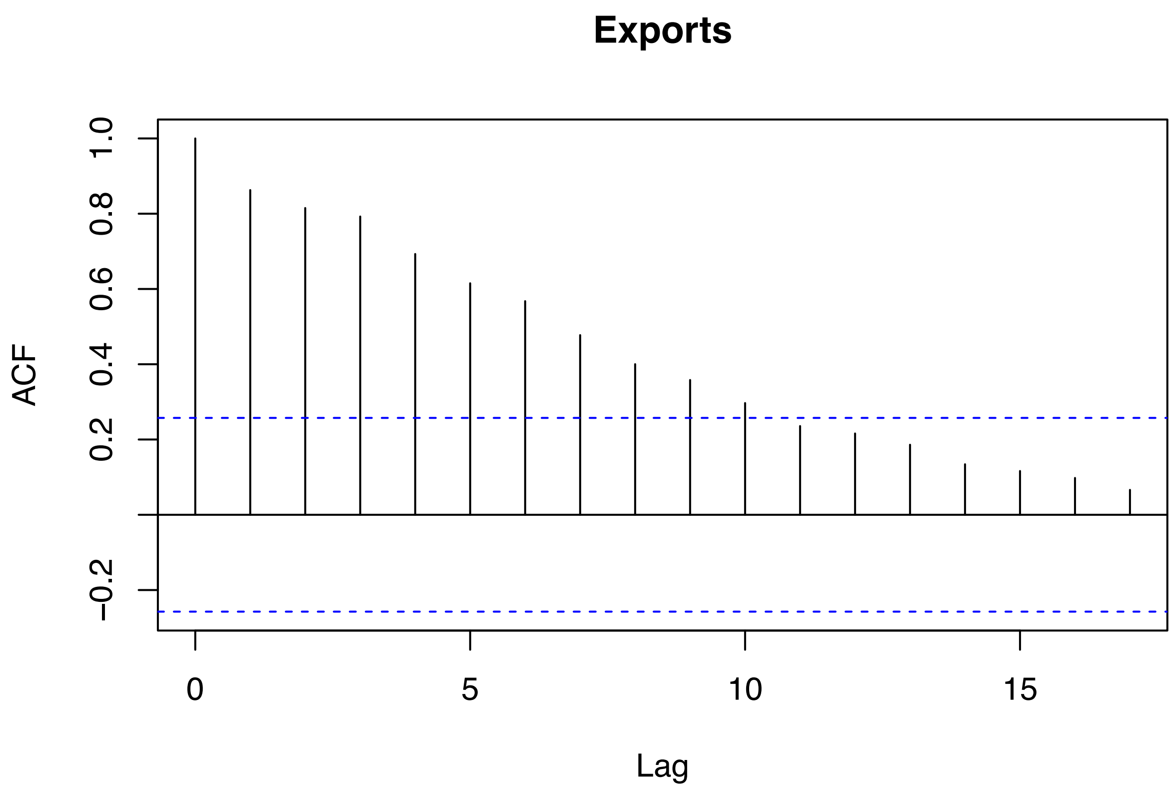



```r
hist(data$Exports, main = "Exports Histogram",xlab = "Exports (% of GDP)")
```

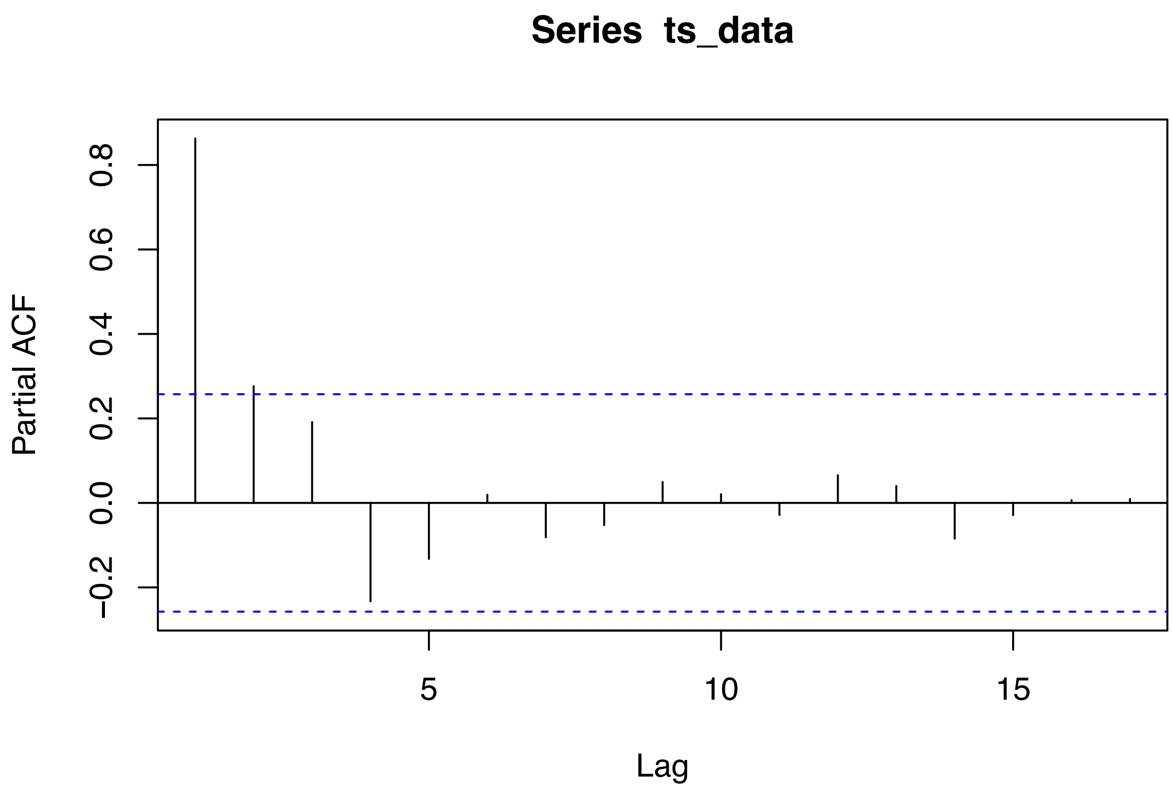


Without Transformations

```
# without transformation
acf(ts_data)
```

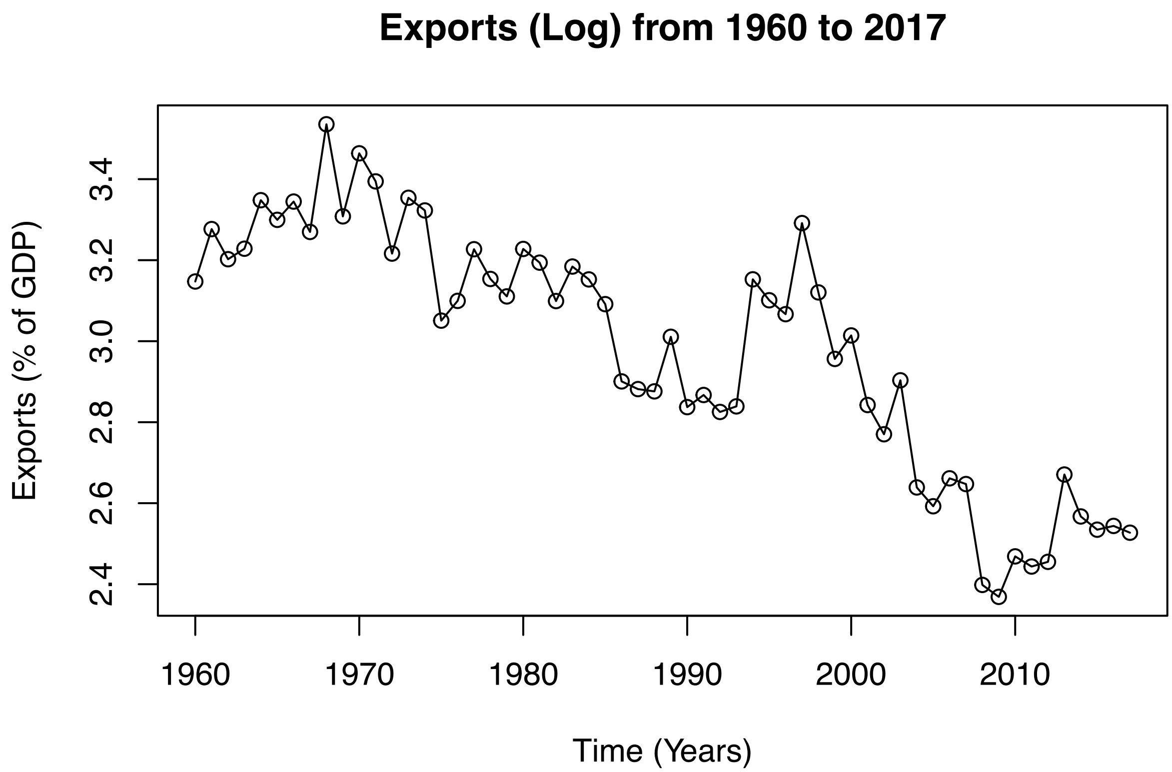



```
pacf(ts_data)
```

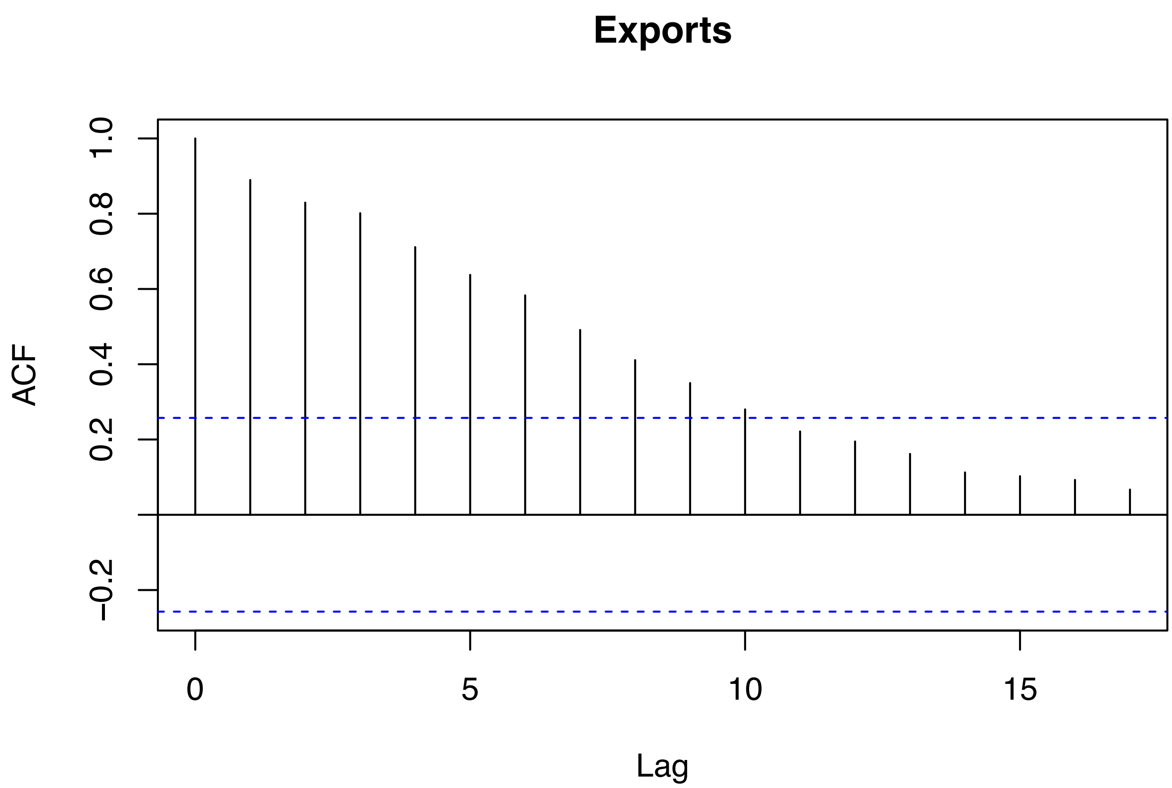


Log Transformation

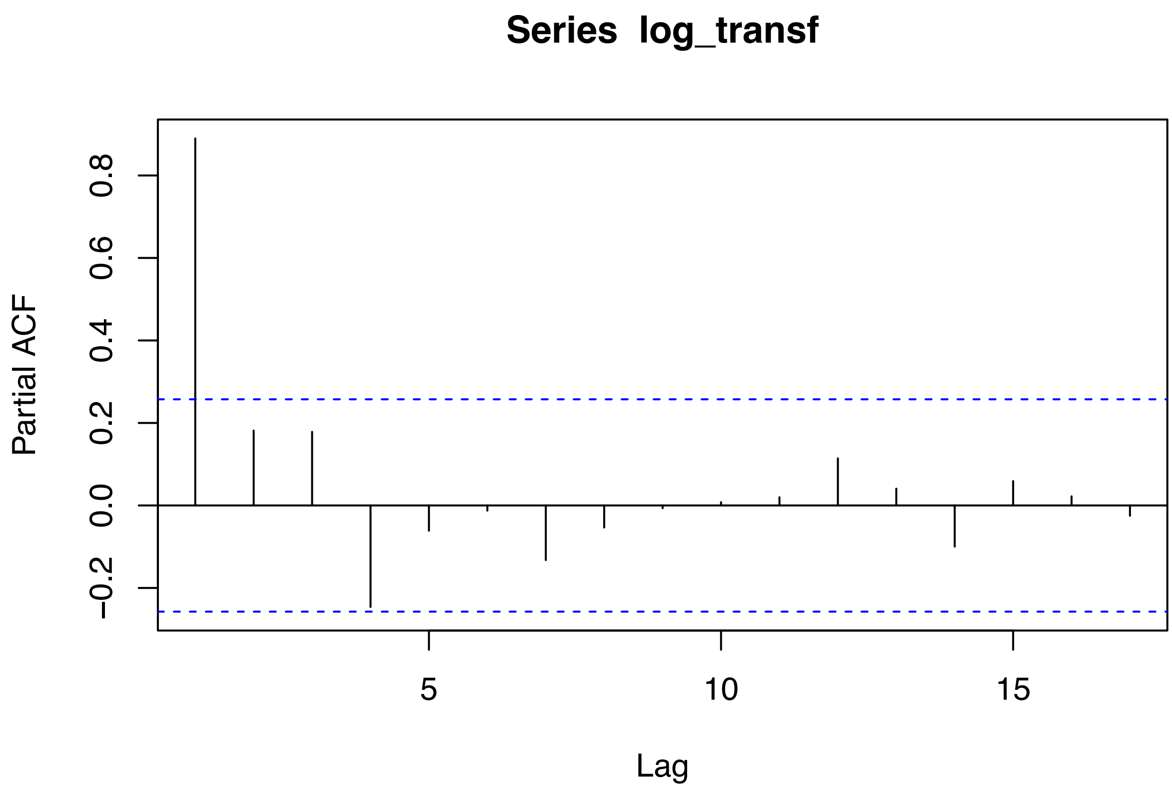
```
# try log transformation
log_transf = log(ts_data)
plot(log_transf, type = 'o',  main = "Exports (Log) from 1960 to 2017", ylab = "Exports (% of GDP)", xla
```




```
acf(log_transf)
```

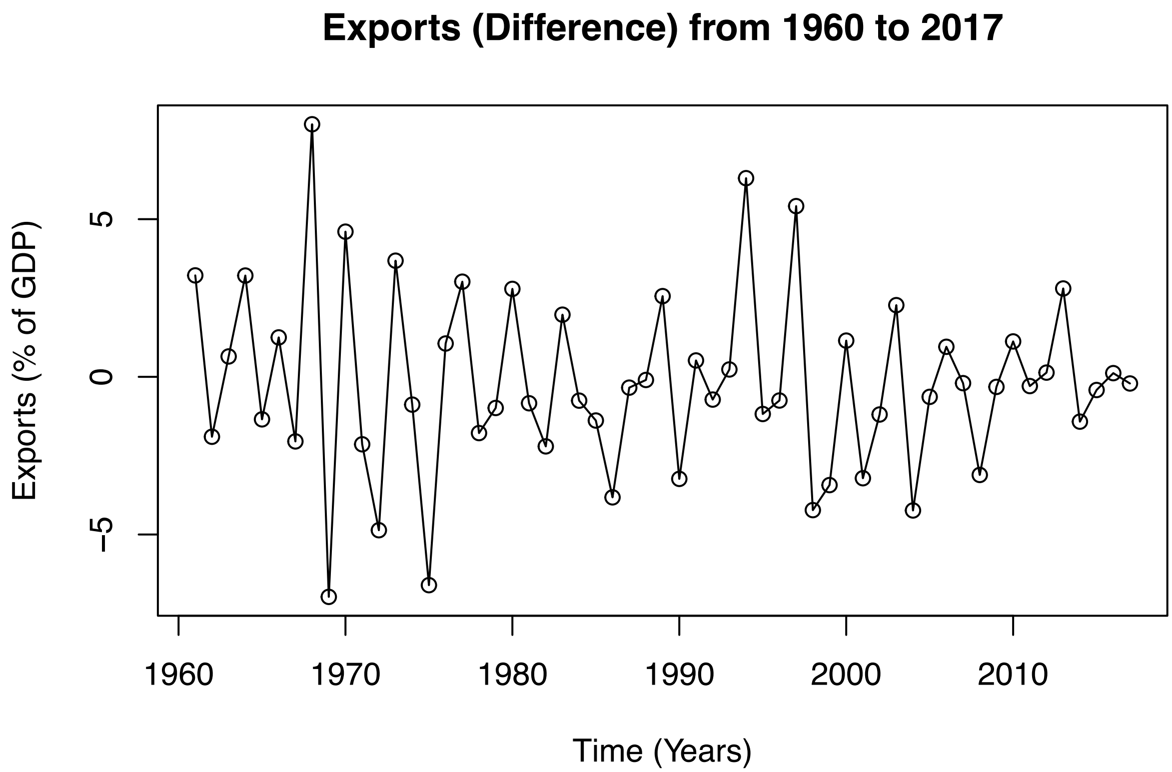



```
pacf(log_transf)
```

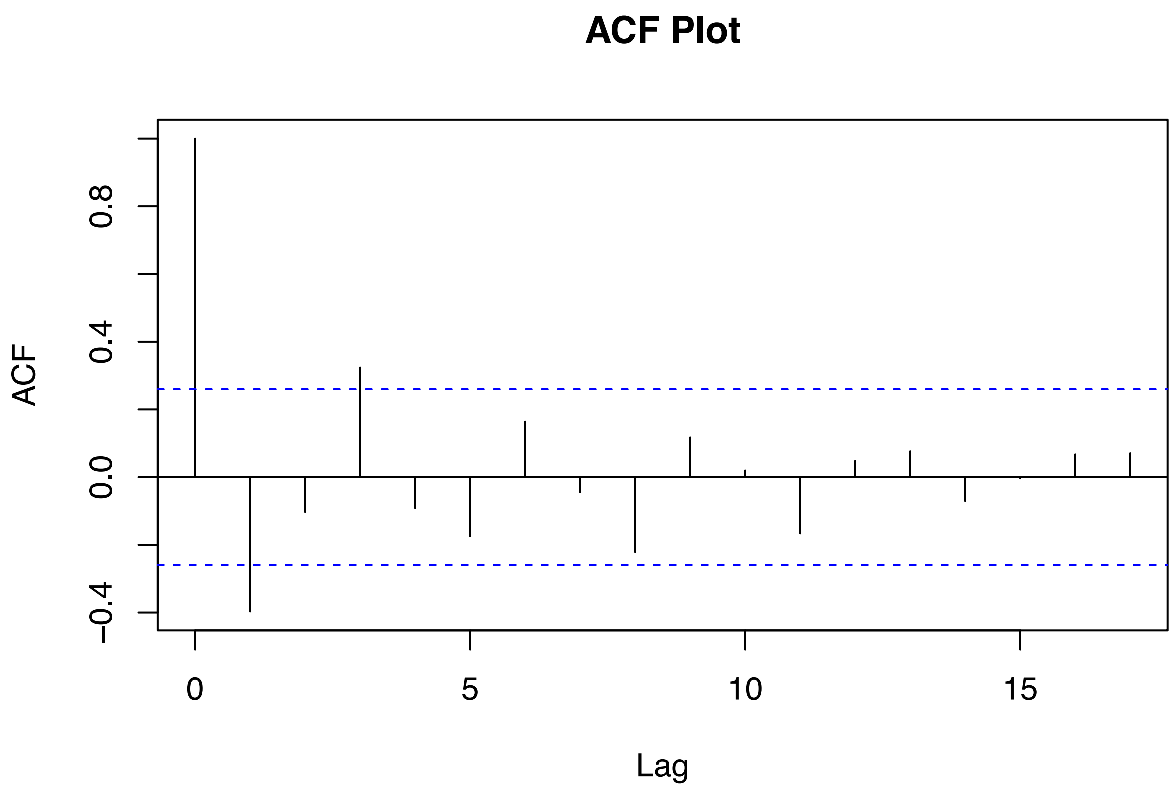


First Order Differencing

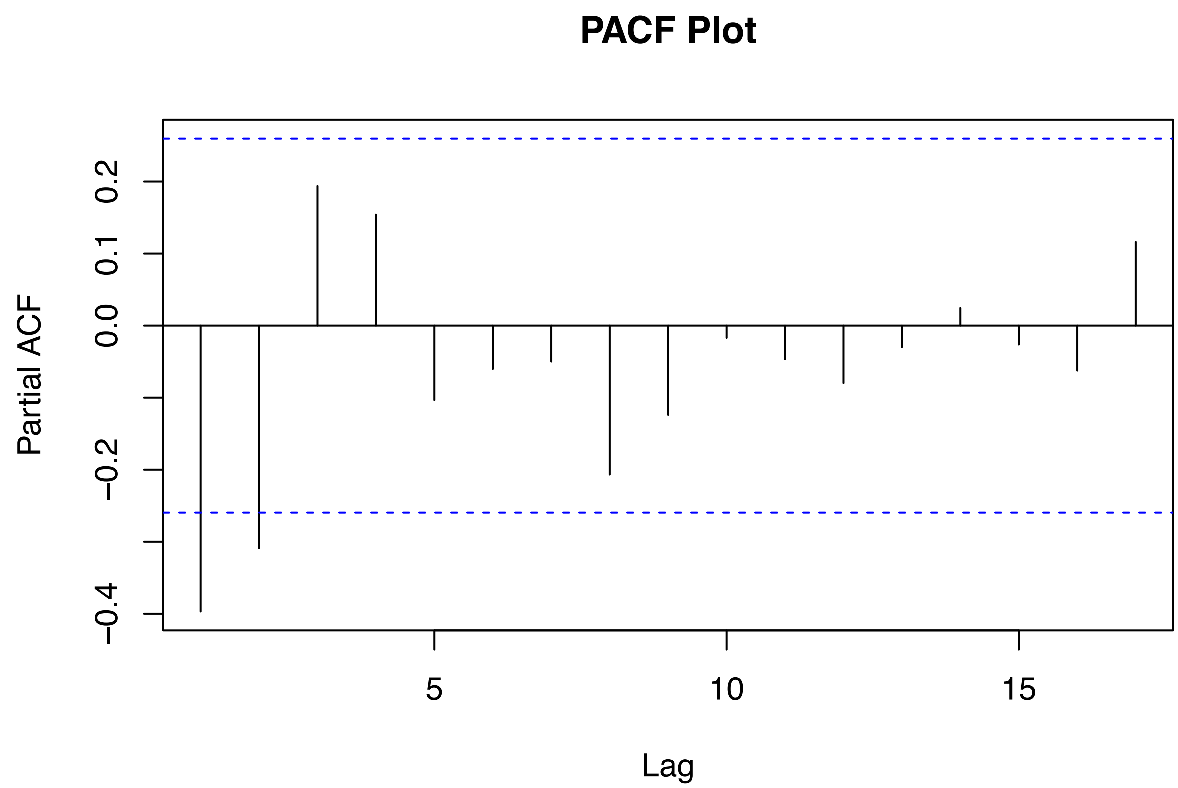
```
# try differencing
differ = diff(ts_data)
plot(differ, type = 'o',  main = "Exports (Difference) from 1960 to 2017",
     ylab = "Exports (% of GDP)", xlab = "Time (Years)")
```




```r
acf(differ, main = 'ACF Plot' )
```

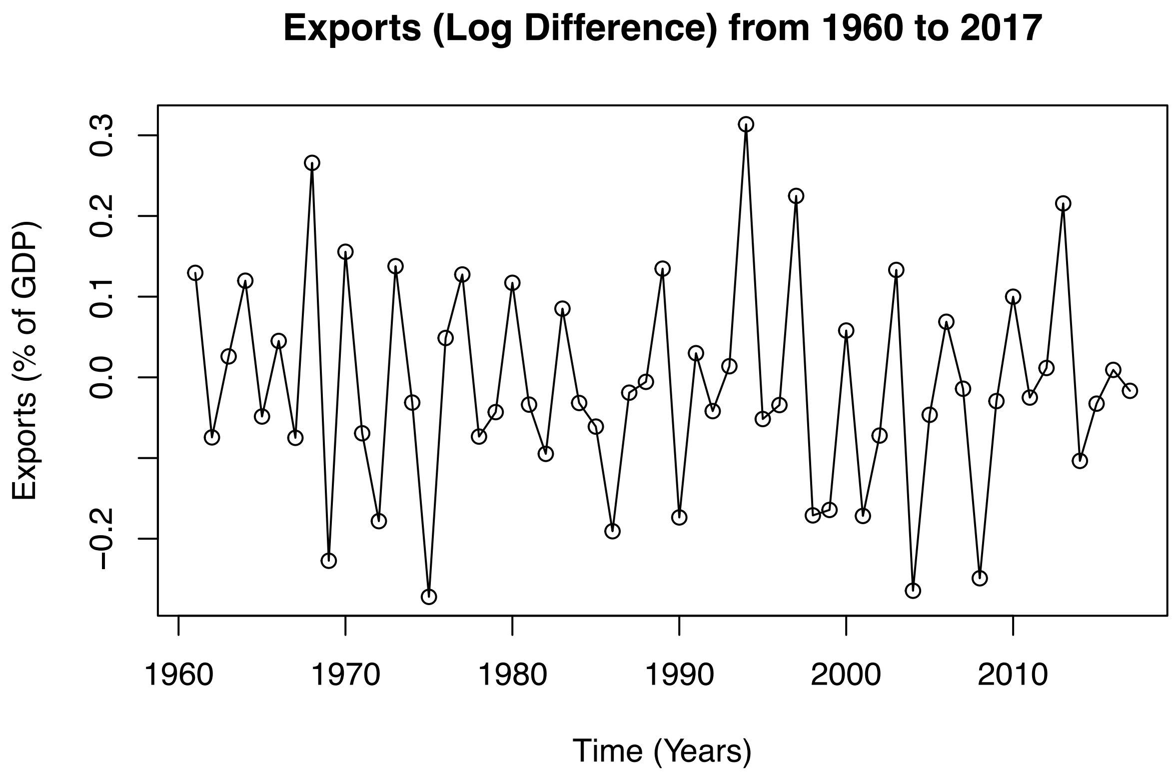



```r
pacf(differ, main = 'PACF Plot')
```

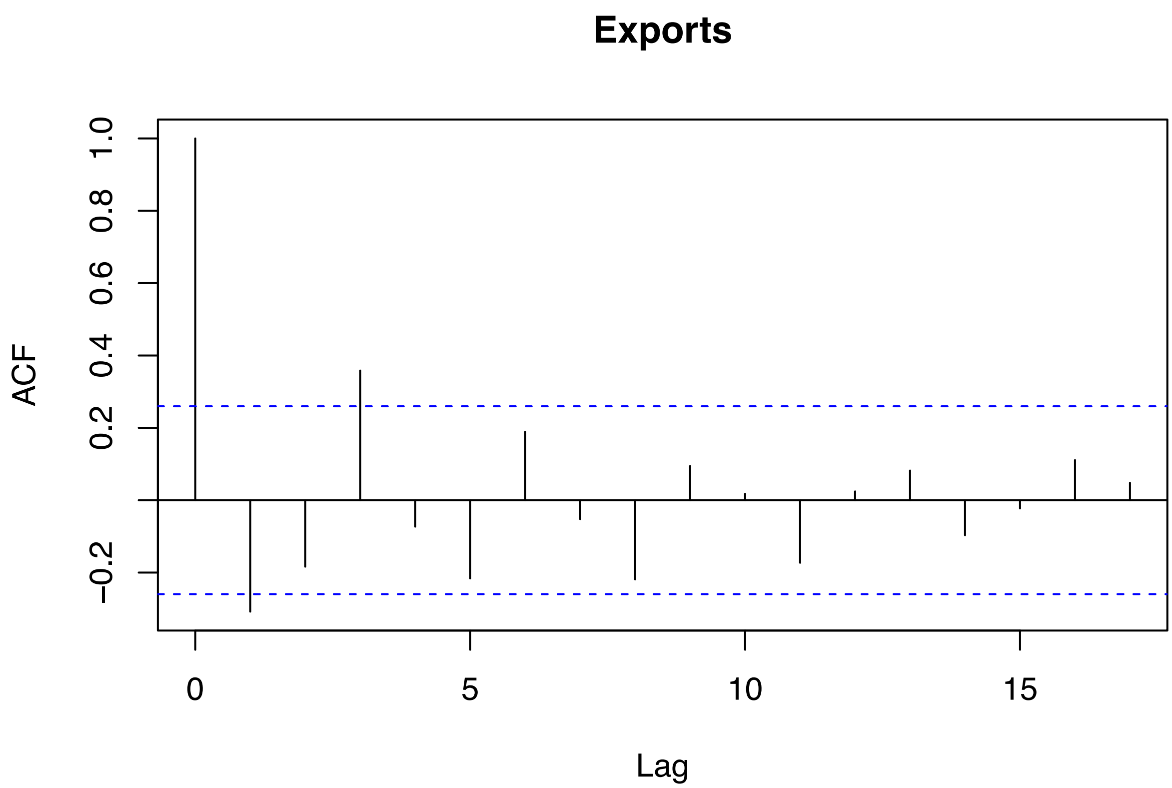


First order differencing on log transformed data

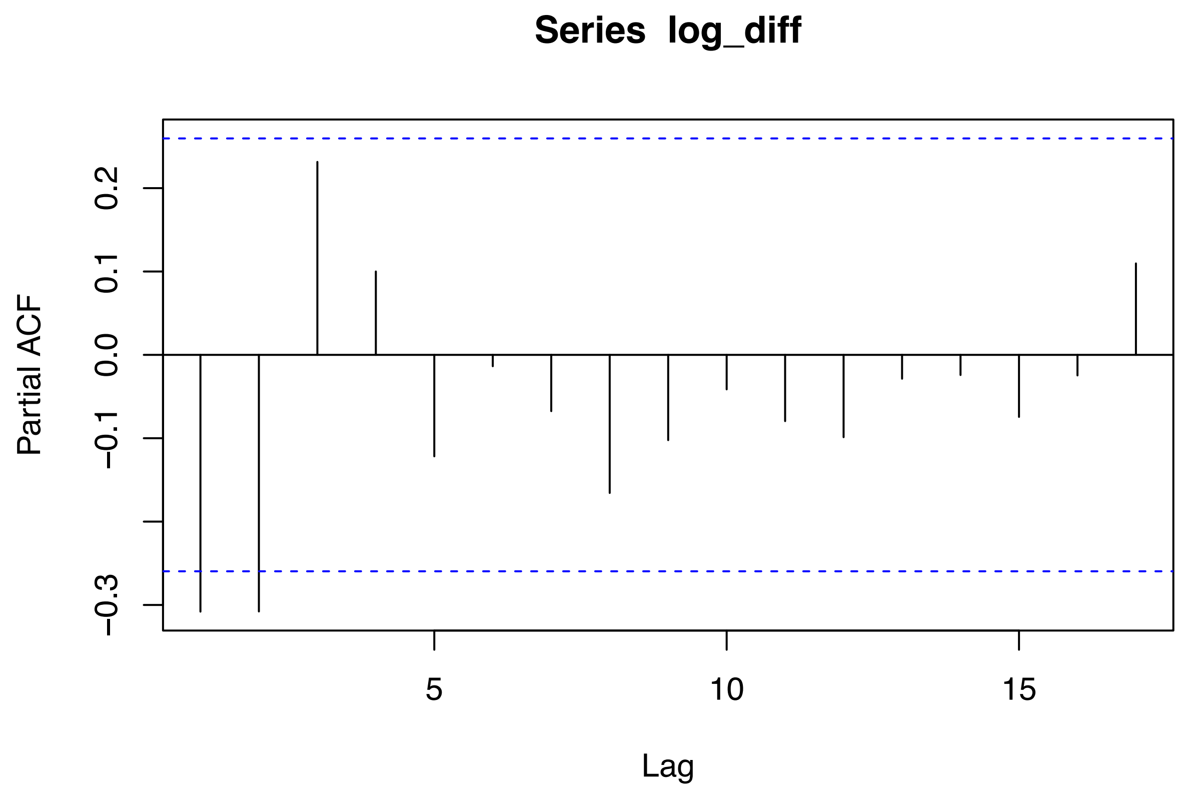
```
# log diff
log_diff = diff(log(ts_data))
plot(log_diff, type = 'o',  main = "Exports (Log Difference) from 1960 to 2017",
     ylab = "Exports (% of GDP)", xlab = "Time (Years)")
```


Exports (Log Difference) from 1960 to 2017
Exports (% of GDP)
−0.2
0.0
0.1
0.2
0.3
1960
1970
1980
1990
2000
2010
Time (Years)

```
acf(log_diff)
```

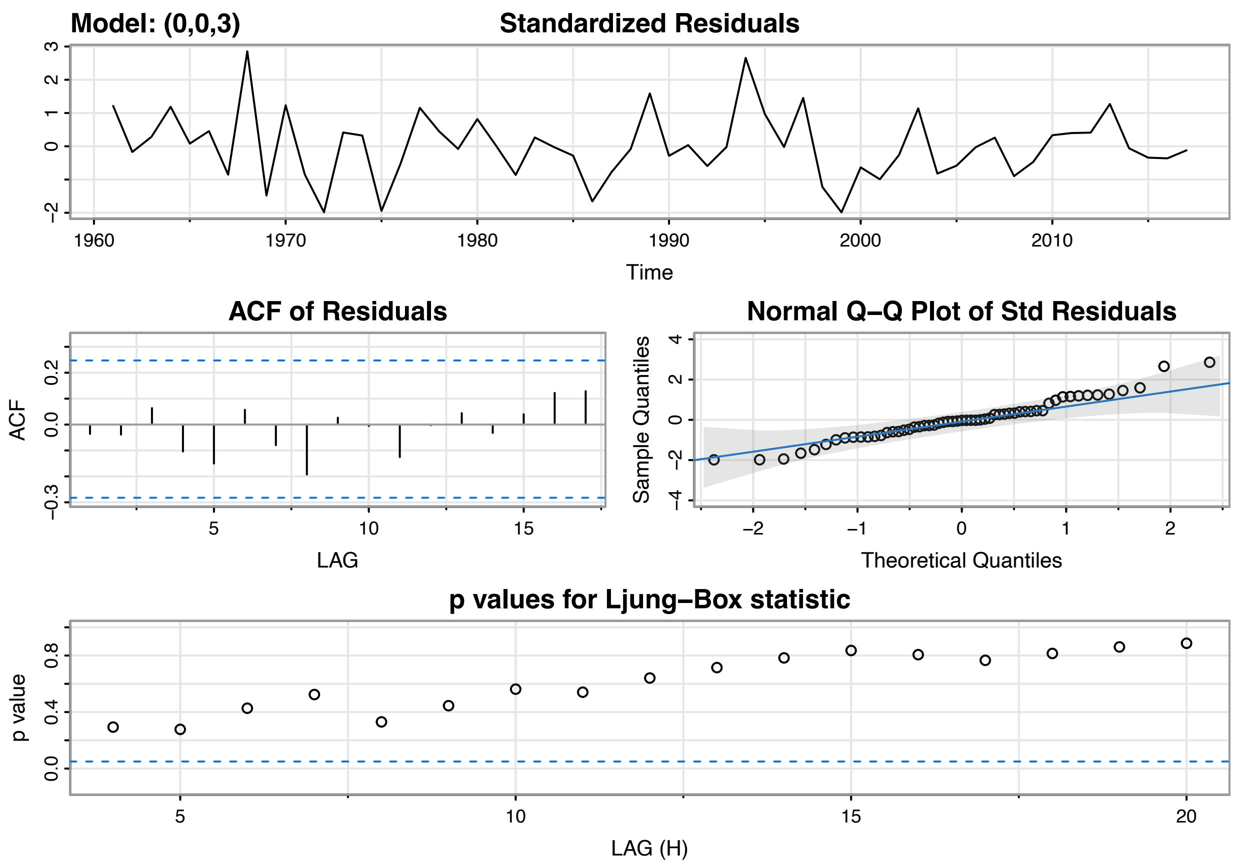



```
pacf(log_diff)
```




```
sarima(differ,0,0,3)
```

```
## initial  value 1.061848
## iter   2 value 0.925456
## iter   3 value 0.912523
## iter   4 value 0.910179
## iter   5 value 0.909894
## iter   6 value 0.909889
## iter   7 value 0.909843
## iter   8 value 0.909841
## iter   9 value 0.909841
## iter   9 value 0.909841
## iter   9 value 0.909841
## final  value 0.909841
## converged
## initial  value 0.912718
## iter   2 value 0.912652
## iter   3 value 0.912622
## iter   4 value 0.912620
## iter   5 value 0.912612
## iter   6 value 0.912612
## iter   7 value 0.912612
## iter   7 value 0.912612
## iter   7 value 0.912612
## final  value 0.912612
## converged
```

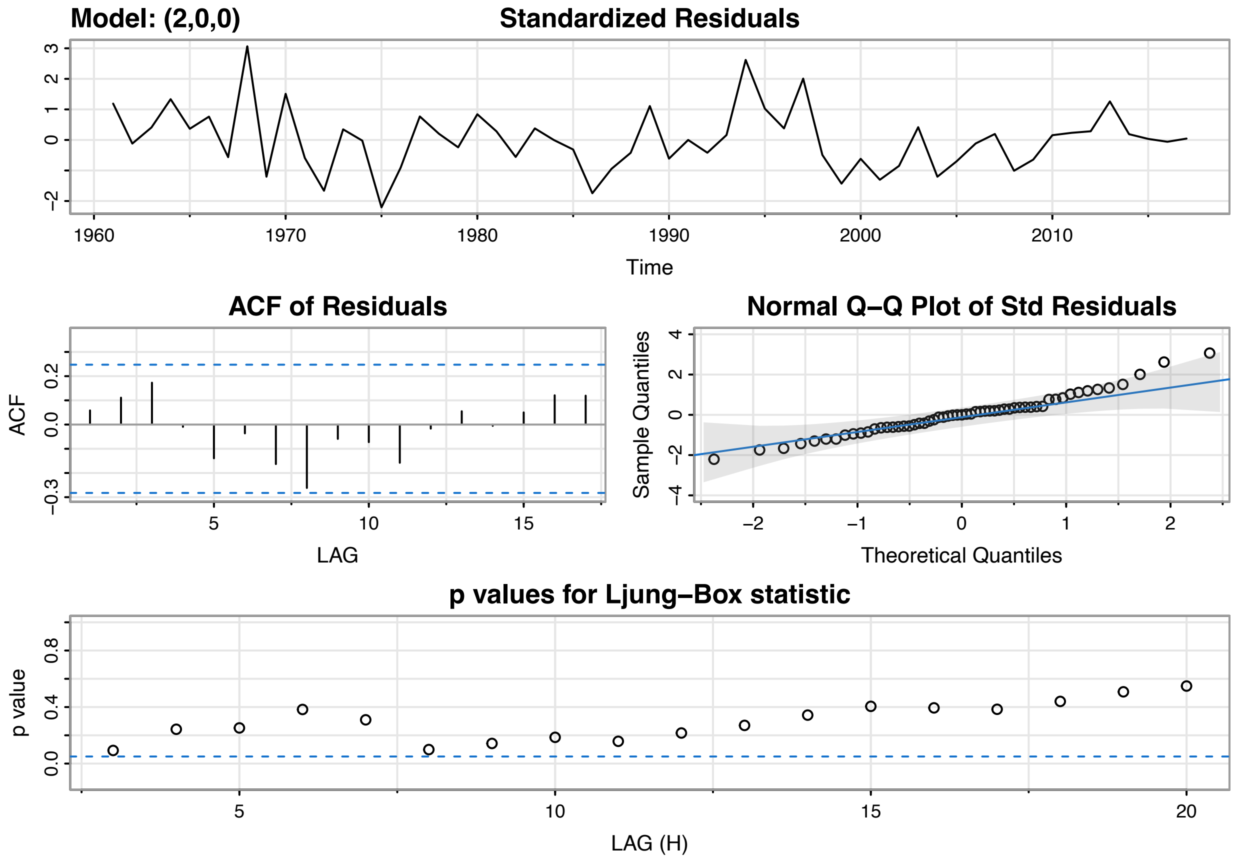



```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     xreg = xmean, include.mean = FALSE, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##           ma1     ma2     ma3    xmean
##       -0.4537  0.0922  0.2677  -0.1999
## s.e.   0.1319  0.1532  0.1354   0.2946
##
## sigma^2 estimated as 6.147:  log likelihood = -132.9,  aic = 275.8
##
## $degrees_of_freedom
## [1] 53
##
## $ttable
##       Estimate     SE t.value p.value
## ma1    -0.4537 0.1319 -3.4387  0.0011
## ma2     0.0922 0.1532  0.6018  0.5499
## ma3     0.2677 0.1354  1.9762  0.0533
## xmean  -0.1999 0.2946 -0.6787  0.5003
##
## $AIC
## [1] 4.838539
```

```
##
## $AICc
## [1] 4.852034
##
## $BIC
## [1] 5.017754
```

```r
sarima(differ,2,0,0)
```

```
## initial  value 1.064229
## iter   2 value 0.951368
## iter   3 value 0.935509
## iter   4 value 0.932371
## iter   5 value 0.929084
## iter   6 value 0.927239
## iter   7 value 0.927238
## iter   8 value 0.927238
## iter   8 value 0.927238
## final  value 0.927238
## converged
## initial  value 0.925441
## iter   2 value 0.925409
## iter   3 value 0.925393
## iter   4 value 0.925391
## iter   5 value 0.925391
## iter   5 value 0.925391
## iter   5 value 0.925391
## final  value 0.925391
## converged
```




```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     xreg = xmean, include.mean = FALSE, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##           ar1      ar2    xmean
##       -0.5230  -0.3065  -0.2120
## s.e.   0.1262   0.1248   0.1841
##
## sigma^2 estimated as 6.323:  log likelihood = -133.63,  aic = 275.25
##
## $degrees_of_freedom
## [1] 54
##
## $ttable
##       Estimate     SE t.value p.value
## ar1    -0.5230 0.1262 -4.1460  0.0001
## ar2    -0.3065 0.1248 -2.4563  0.0173
## xmean  -0.2120 0.1841 -1.1514  0.2546
##
## $AIC
## [1] 4.829009
##
```

```
## $AICc
## [1] 4.836953
##
## $BIC
## [1] 4.972381
```

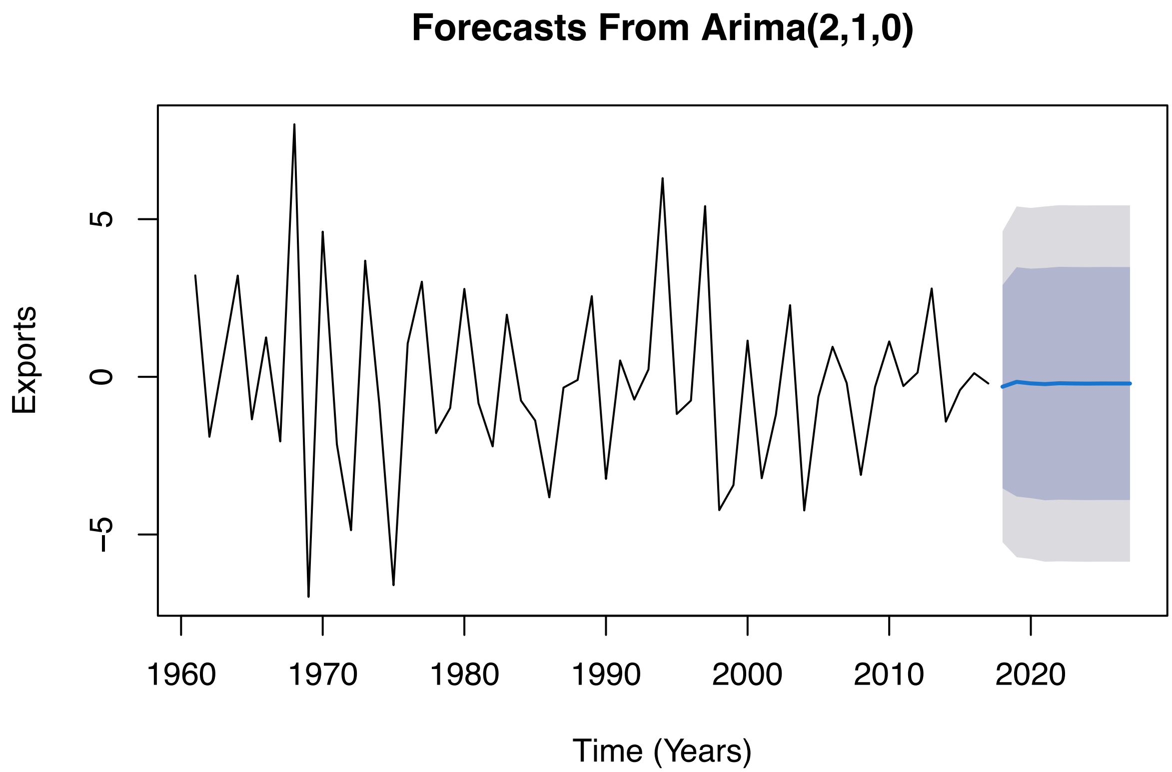
```r
ARMAtoMA(ar=c(-0.5230,-0.3065), ma=0,10)
```

```
##  [1] -0.523000000 -0.032971000  0.177543333 -0.082749552 -0.011139016
##  [6]  0.031188443 -0.012897447 -0.002813893  0.005424734 -0.001974677
```

```r
# forecast
l = forecast(arima(differ,order=c(2,0,0)), h = 10)
l
```

```
##      Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 2018     -0.3133799 -3.536014 2.909254 -5.241973 4.615213
## 2019     -0.1592450 -3.796054 3.477564 -5.721264 5.402774
## 2020     -0.2084660 -3.846826 3.429894 -5.772858 5.355926
## 2021     -0.2299685 -3.913053 3.453116 -5.862760 5.402823
## 2022     -0.2036345 -3.896364 3.489095 -5.851177 5.443908
## 2023     -0.2108169 -3.903721 3.482087 -5.858626 5.436993
## 2024     -0.2151324 -3.909405 3.479140 -5.865034 5.434770
## 2025     -0.2106736 -3.905180 3.483833 -5.860933 5.439586
## 2026     -0.2116829 -3.906200 3.482835 -5.861960 5.438594
## 2027     -0.2125217 -3.907081 3.482037 -5.862862 5.437818
```

```r
plot(l, ylab = "Exports", xlab = "Time (Years)", main = 'Forecasts From Arima(2,1,0)')
```


Forecasts From Arima(2,1,0)
Exports
5
0
−5
1960
1970
1980
1990
2000
2010
2020
Time (Years)