Code

CONSUMER HISTORY: Predicting USA Gros Domestic Product (GDP)

Author: Thiago Tanure Andozia

GO TO TOP

In this document, you will find time series models to predict USA Gros Domestic Product (GDP).

```
The time series will be generated by the column "GDP".
Also, we will try to extrapolate with the data from the Unemployment column.
```

Important:

```
We will create the time series ending in late 2019.
```

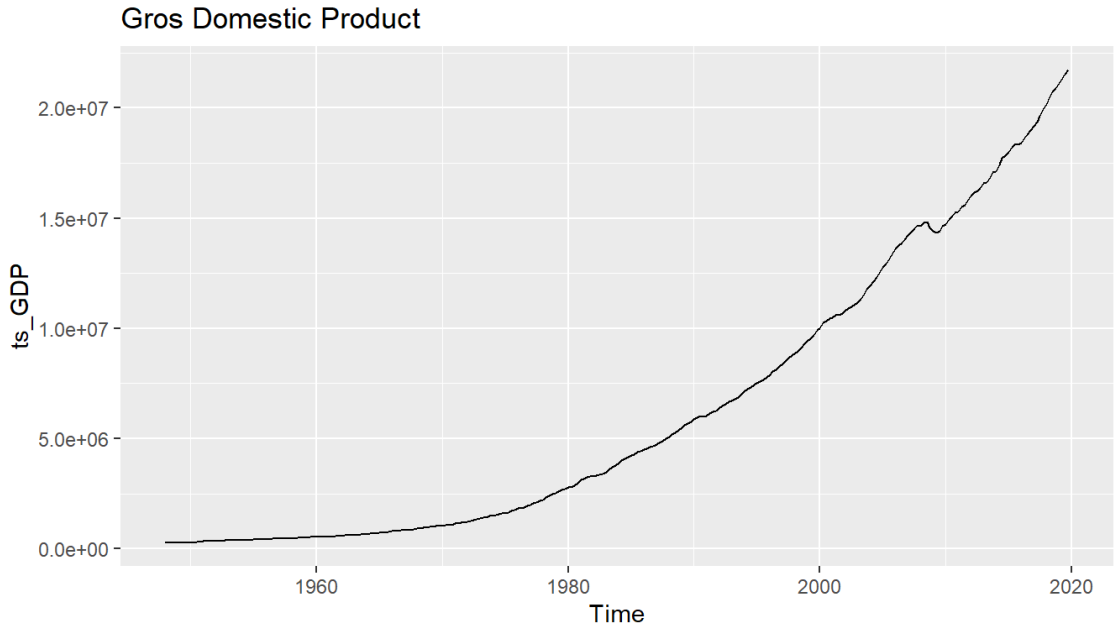
What we will have:

```
The extrapolation of the expectation of GPD by it-self and with a scenario with unemployment as a predictor.
```

Code

```
The first time serie and some informations about it.
```

Code



Code

```
## [1] "Mean: 6279515.81"
```

Code

```
## [1] "Median: 3851380"
```

Code

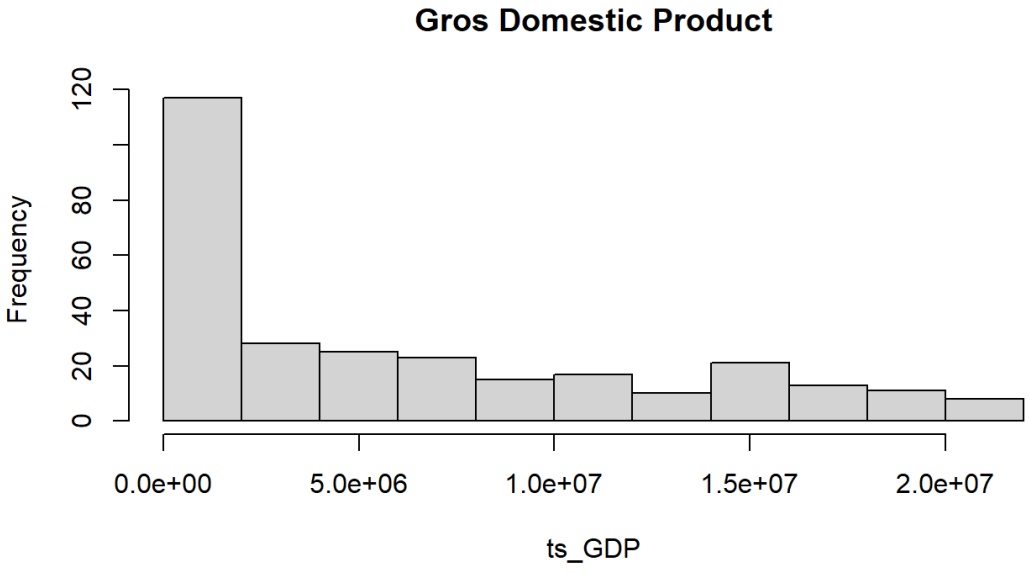
```
## [1] "length: 288"
```

Code

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   265742   789765  3851380  6279516 10692458 21729124
```

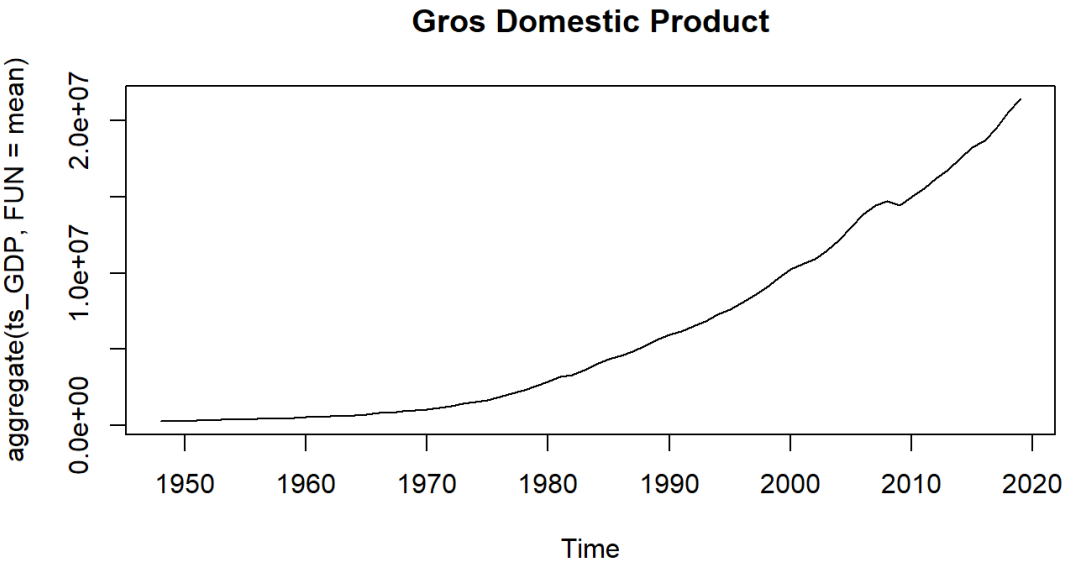
Graphics

Code



Code

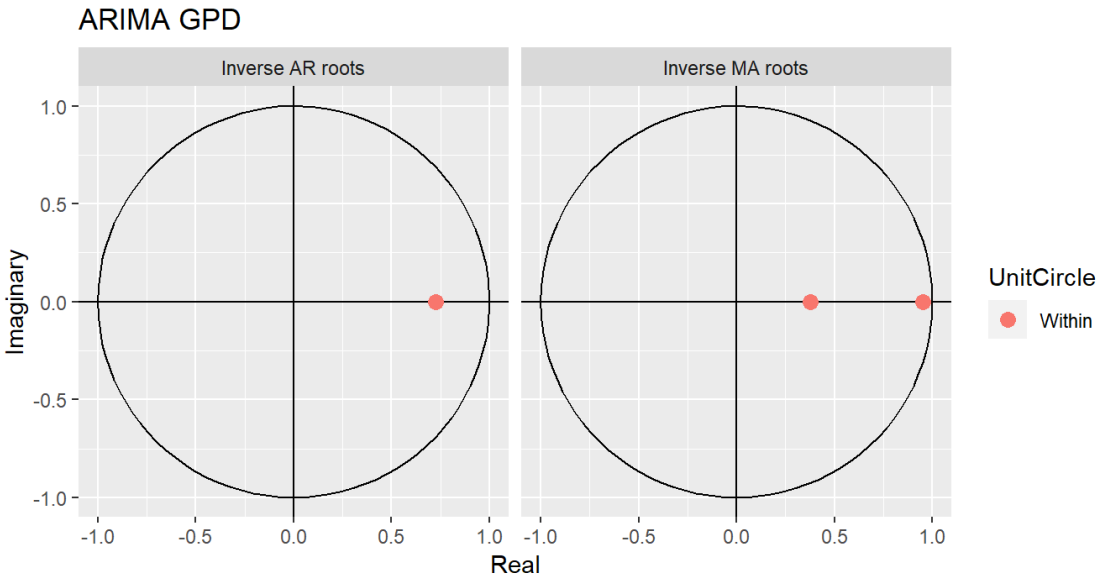
Code



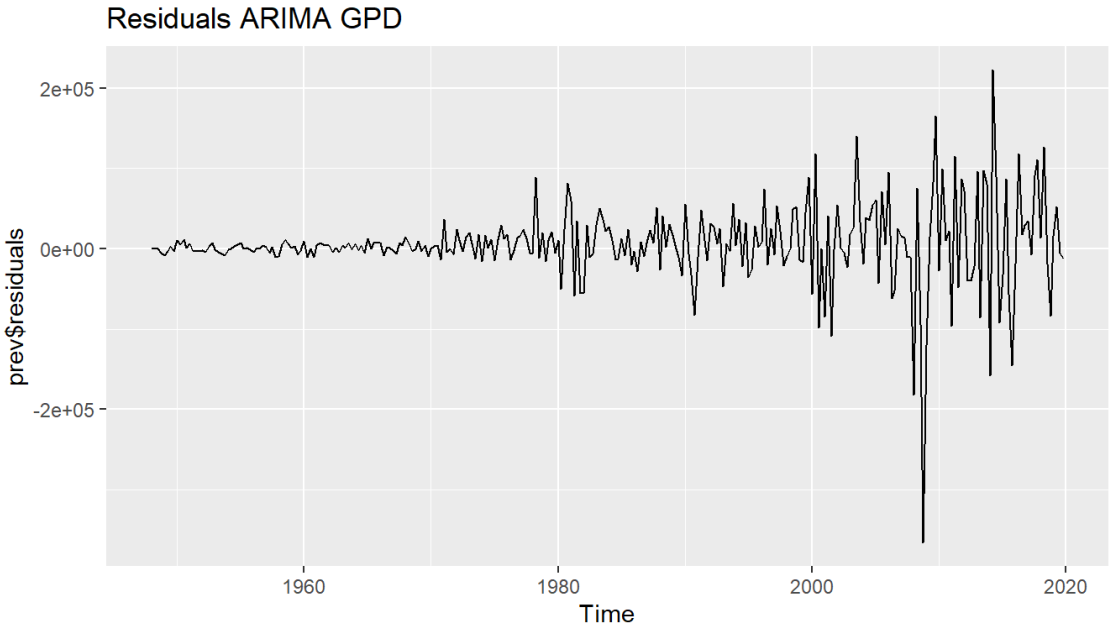
Note: FUN is the aggregation function. Softened data that is very seasonal.

PREDICTIVE AUTO ARIMA

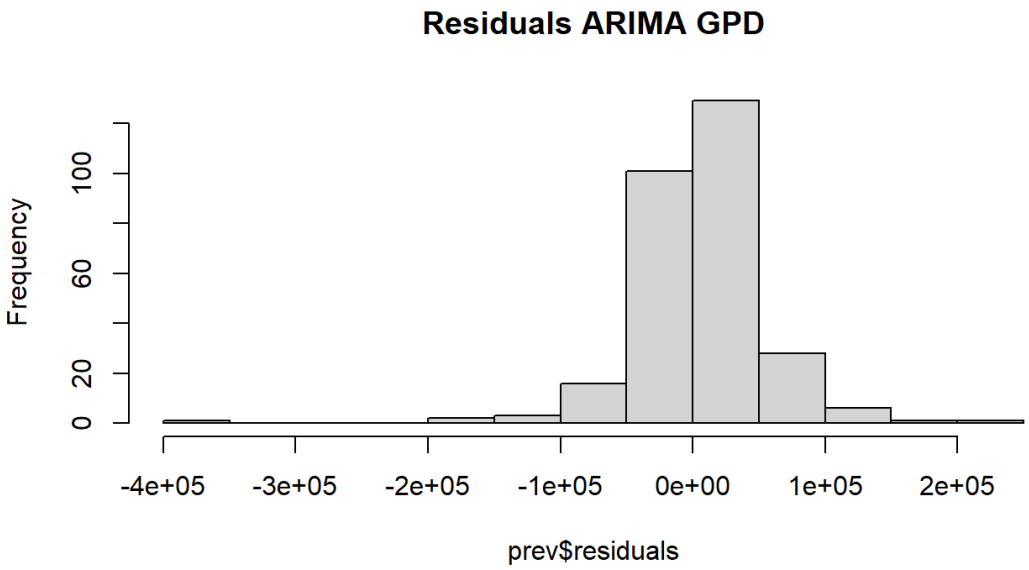
Code


ARIMA GPD
Inverse AR roots
Inverse MA roots
Imaginary
Real
UnitCircle
Within

Code

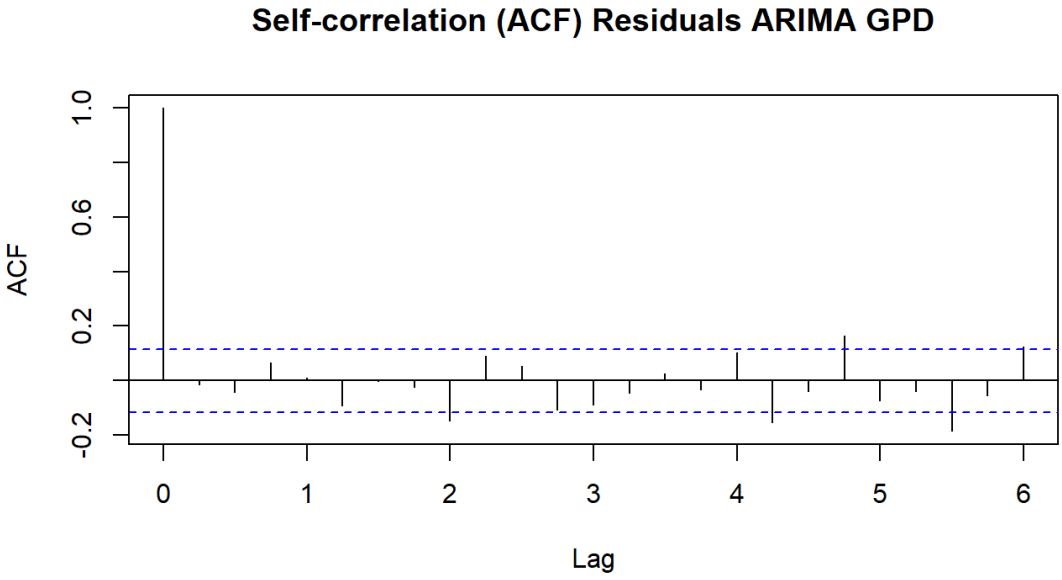

Residuals ARIMA GPD
prev$residuals
Time

Code

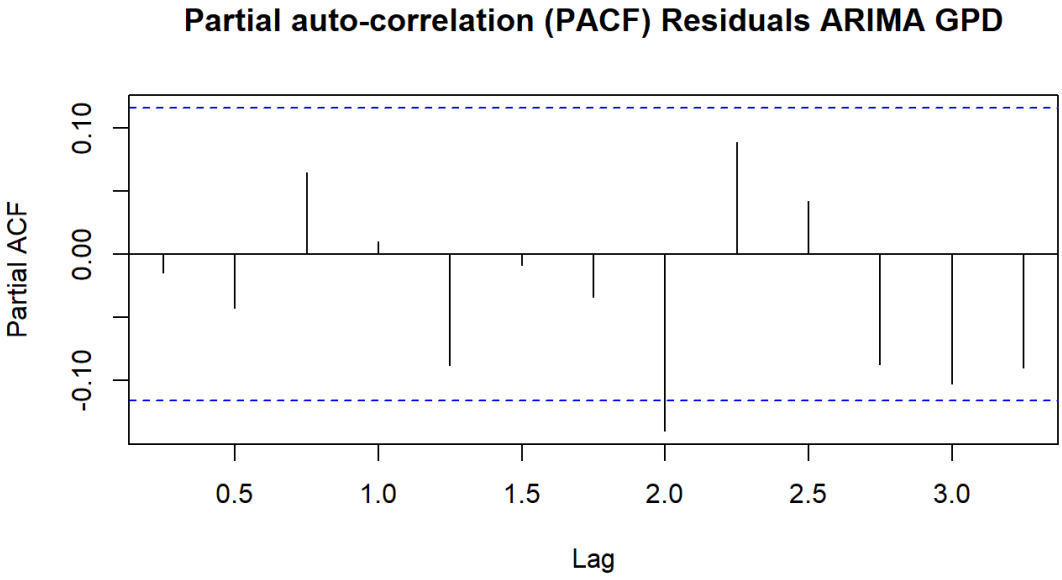

Residuals ARIMA GPD
Frequency
prev$residuals

Code

Code



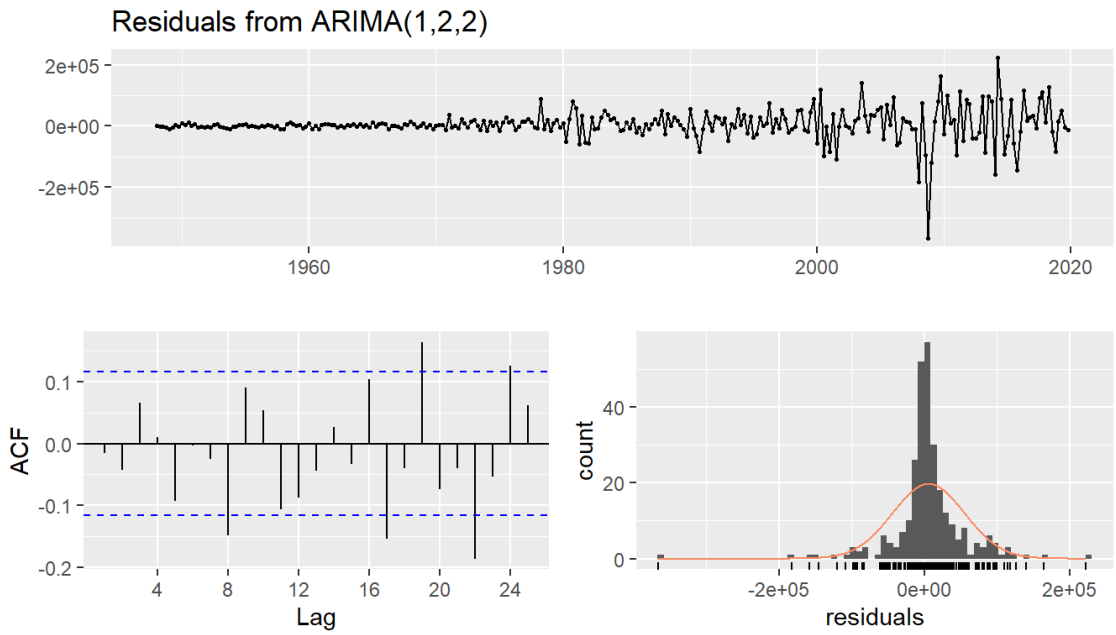
Code



As we can see, just white noise.

Here we can see another way to analyze residuals:

Code



ARIMA (1, 2, 2) means that this has:
- 1 auto-regressive parameter;
- 2 differentiations from the original series;
- 2 parameters of moving averages.'

To see if are normally distributed:

Code

```
##
##  Shapiro-Wilk normality test
##
## data:  prev$residuals
## W = 0.82394, p-value < 2.2e-16
```

```
The data does not come from a normal distribution
```

STATIONARITY & DIFFERENTIATIONS

STATIONARITY

Code

```
##
## ########################################
## # KPSS Unit Root / Cointegration Test #
## ########################################
##
## The value of the test statistic is: 4.4638
```

```
This series has a tendency. Probably not stationary.
The test is greater than 0.05. It shows us that this series is not stationary.
```

DIFFERENTIATIONS

It transforms from non-stationary to stationary

Function to know how many differentiations are needed

Code

```
## [1] "It is necessary to do 2 differentiations"
```

Code

```
##
## ###############################################
## # KPSS Unit Root / Cointegration Test #
## ###############################################
##
## The value of the test statistic is: 3.5597
```

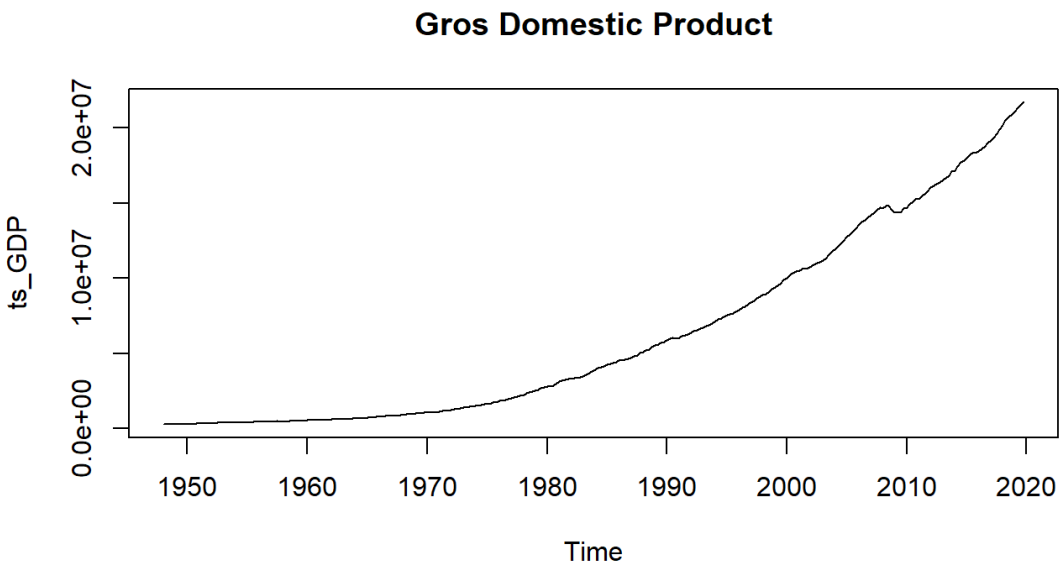
Code

```
##
## ###############################################
## # KPSS Unit Root / Cointegration Test #
## ###############################################
##
## The value of the test statistic is: 0.0127
```

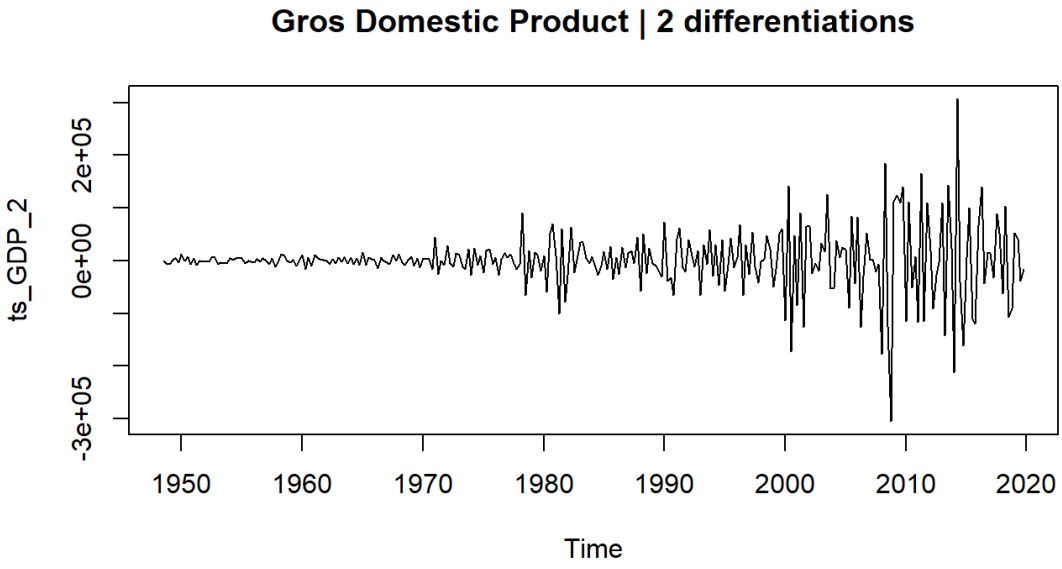
After 2 differentiation processes, we managed to transform it into a stationary one.

Visually analyzing both (original and after 2 differentiations)

Code

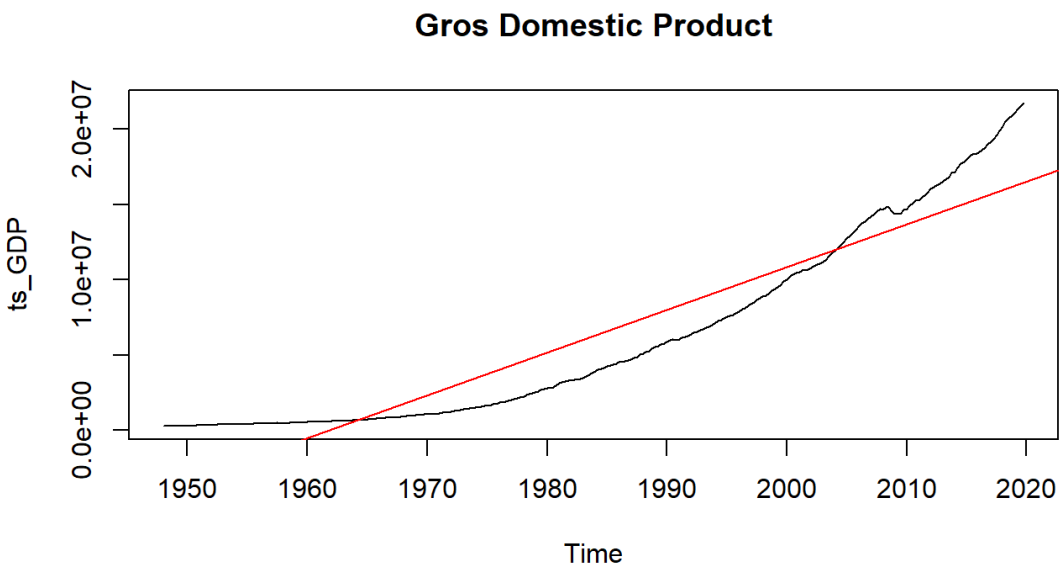


Code

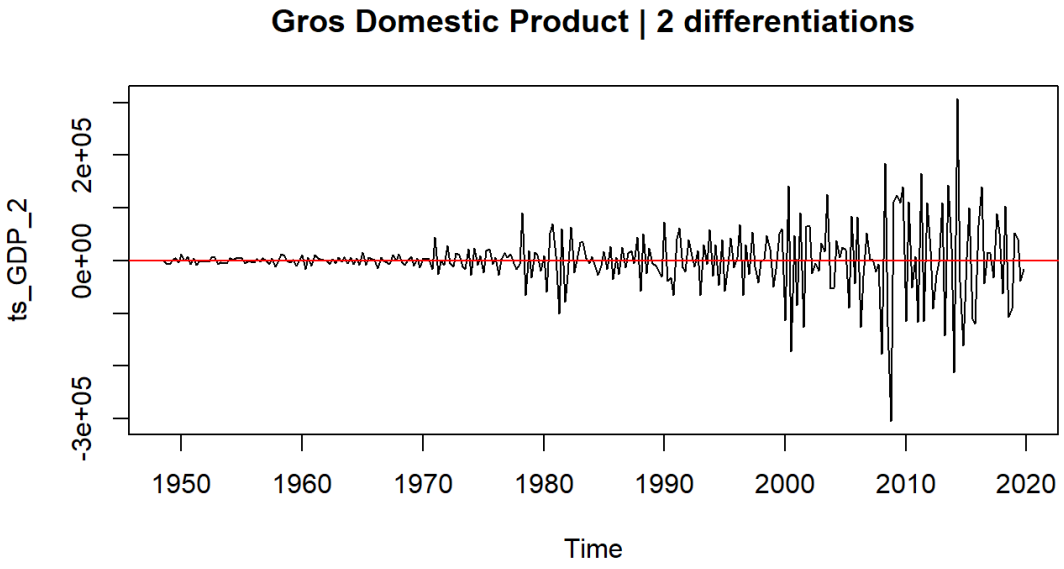


SEASONNALITY & TREND

Code



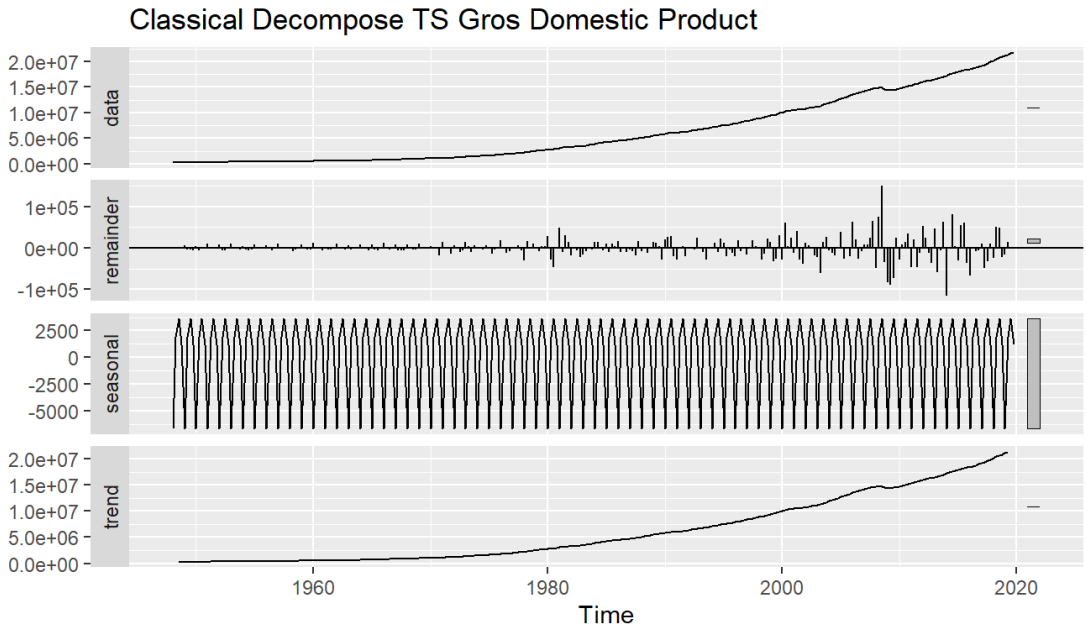
Code



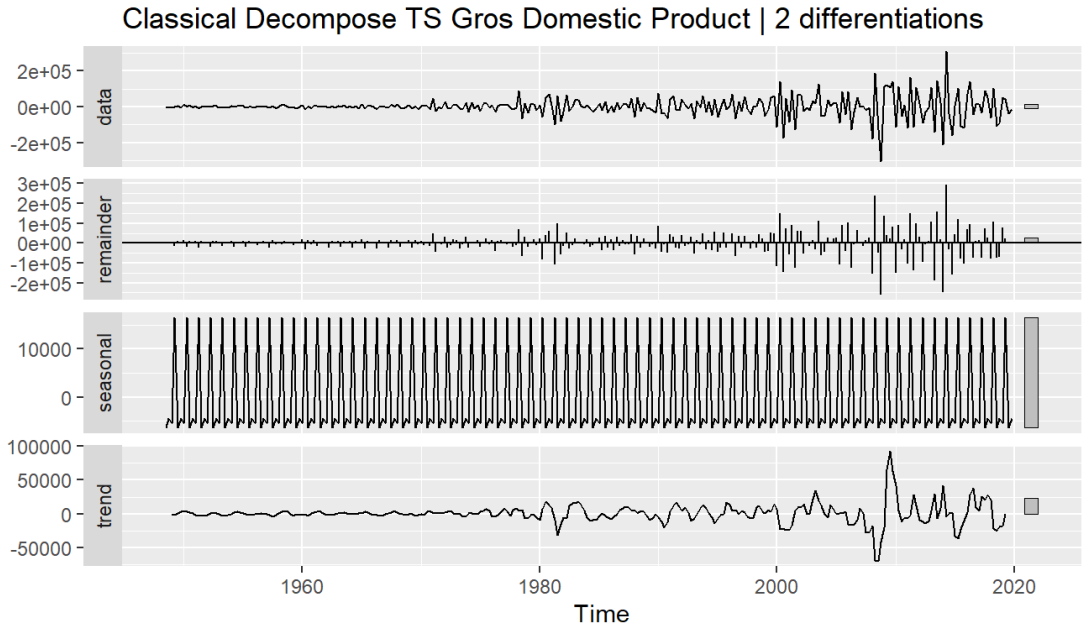
DECOMPOSION & BOXCOX TRANSF.

CLASSICAL DECOMPOSION

Code

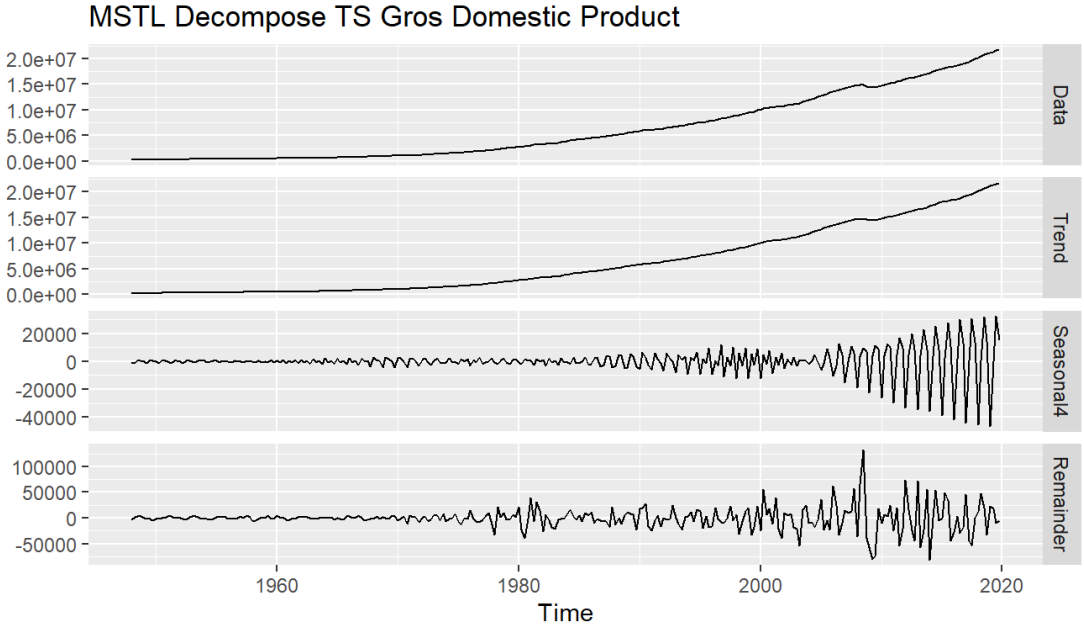


Code

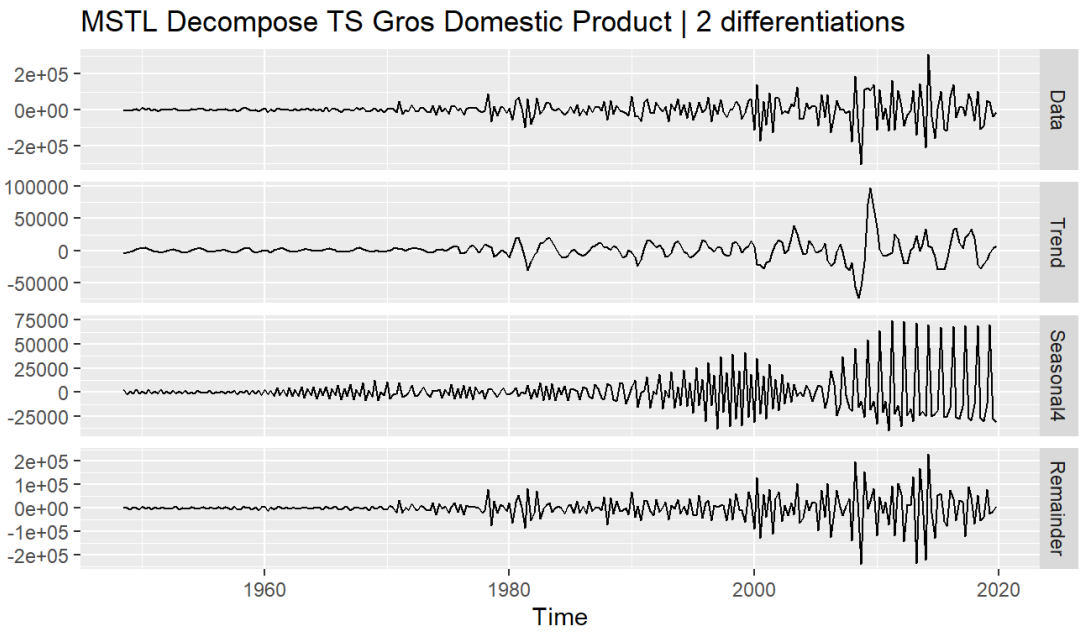


MSTL DECOMPOSION

Code



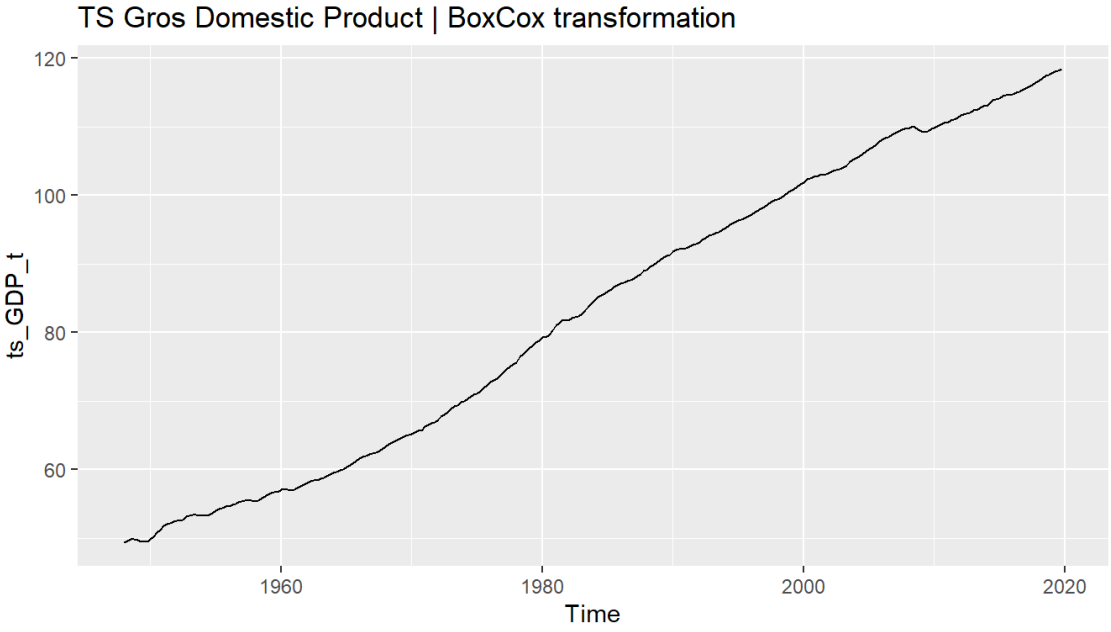
Code



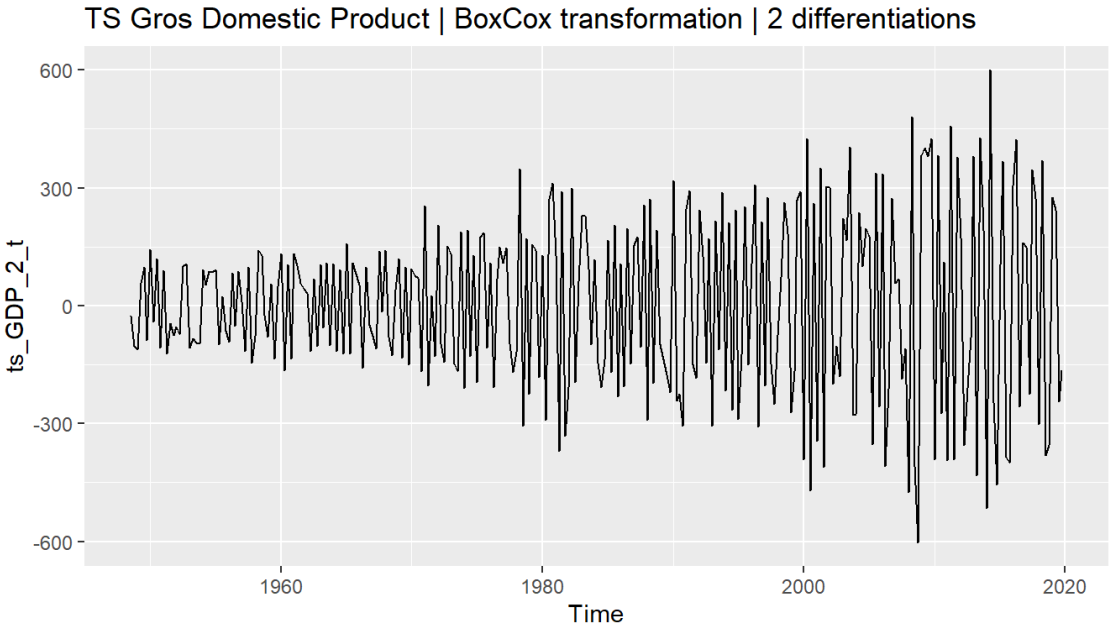
BOXCOX TRANSFORMATION

We will generate automatic logarithmic labda

Code

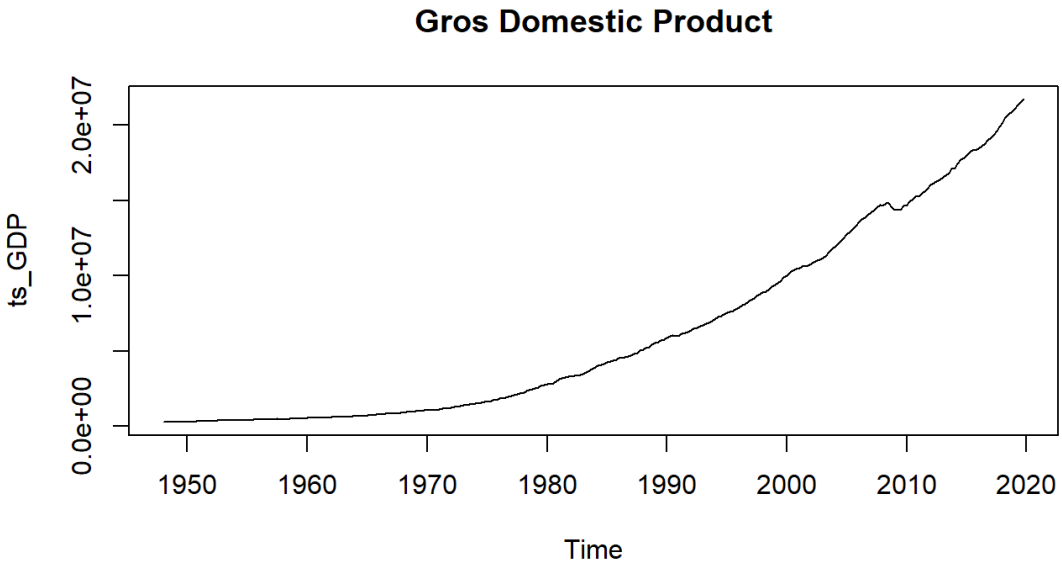


Code

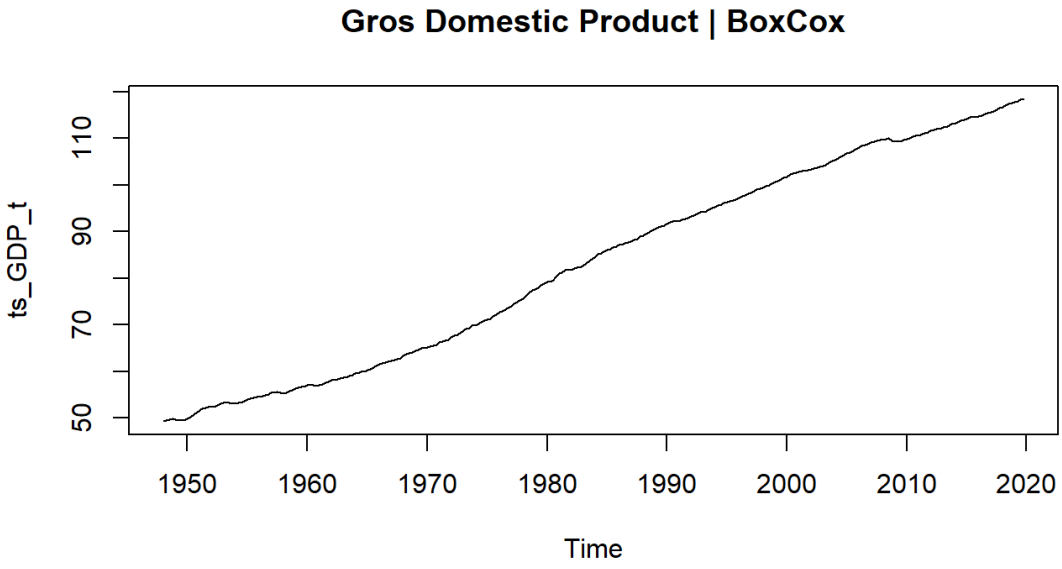


PLOTS TO COMPARE

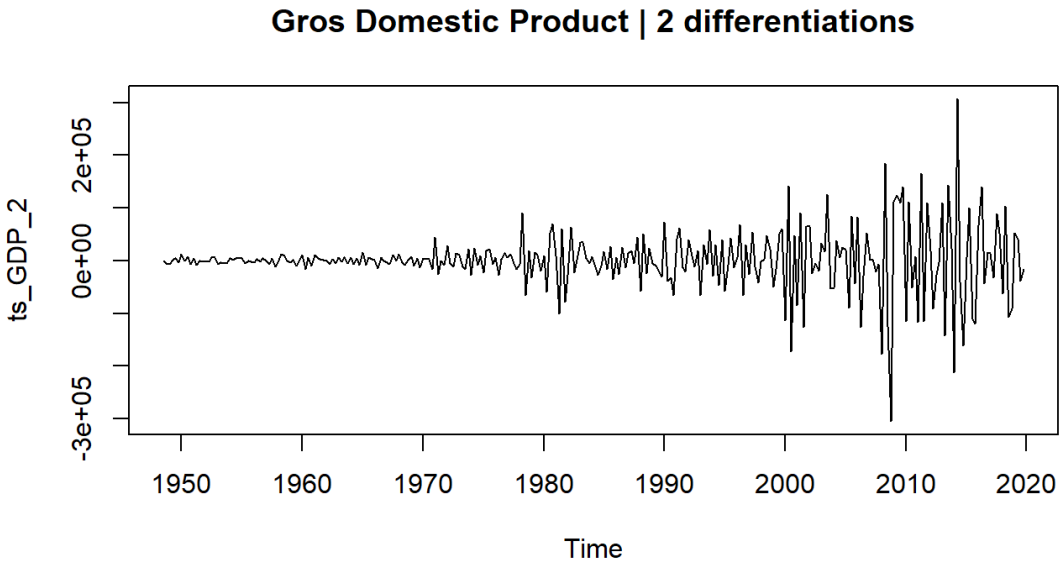
Code



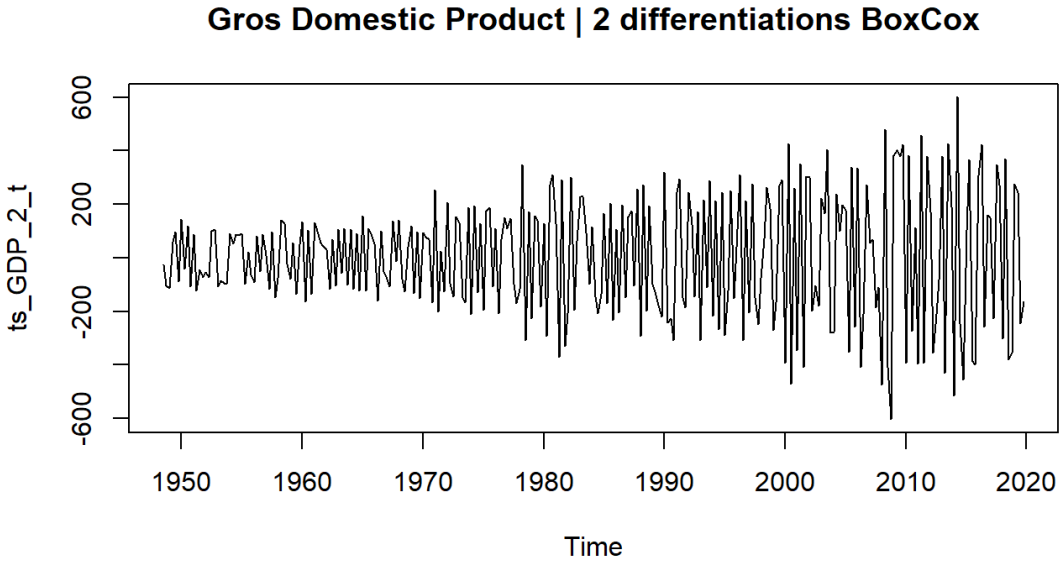
Code


Gros Domestic Product | BoxCox
ts_GDP_t
50
70
90
110
1950
1960
1970
1980
1990
2000
2010
2020
Time

Code

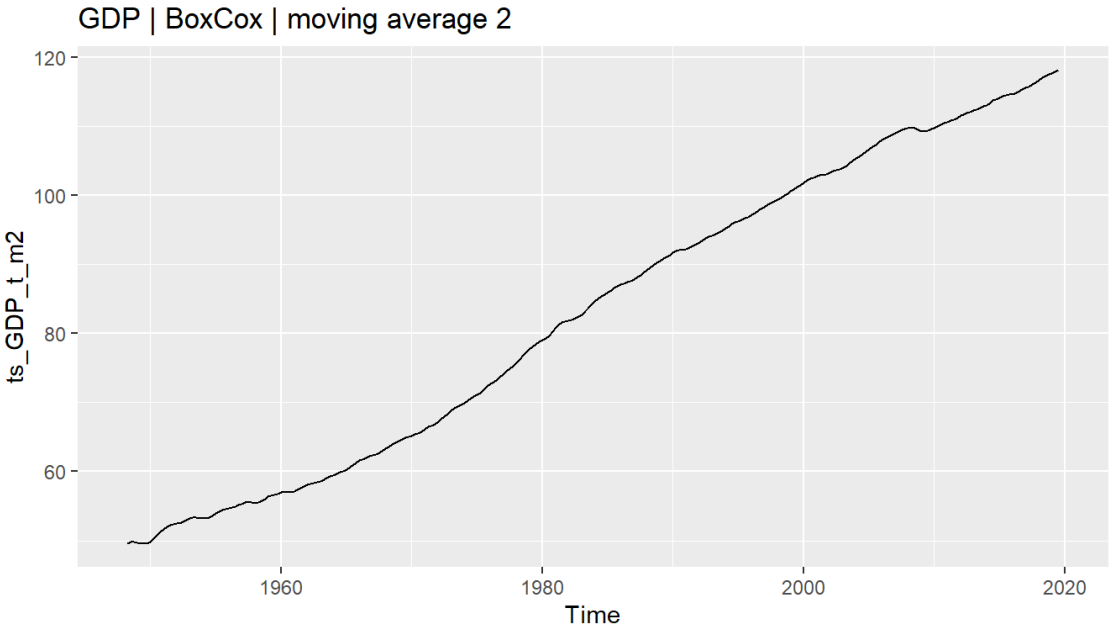

Gros Domestic Product | 2 differentiations
ts_GDP_2
-3e+05
0e+00
2e+05
1950
1960
1970
1980
1990
2000
2010
2020
Time

Code

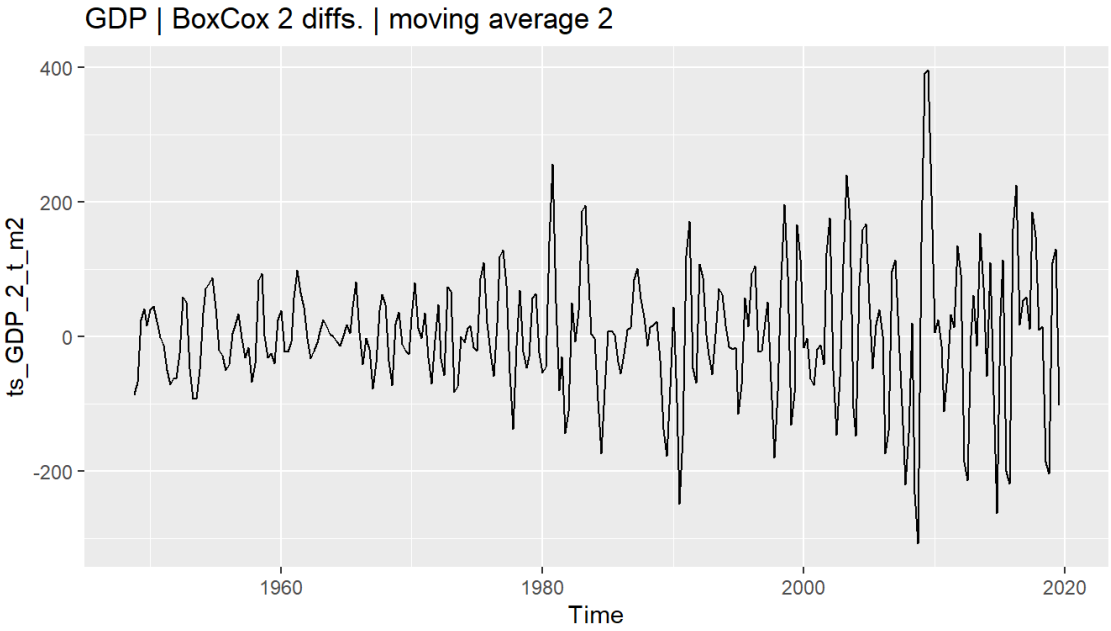

Gros Domestic Product | 2 differentiations BoxCox
ts_GDP_2_t
-600
-200
200
600
1950
1960
1970
1980
1990
2000
2010
2020
Time

MOVING AVERAGE & DATA CLEANING

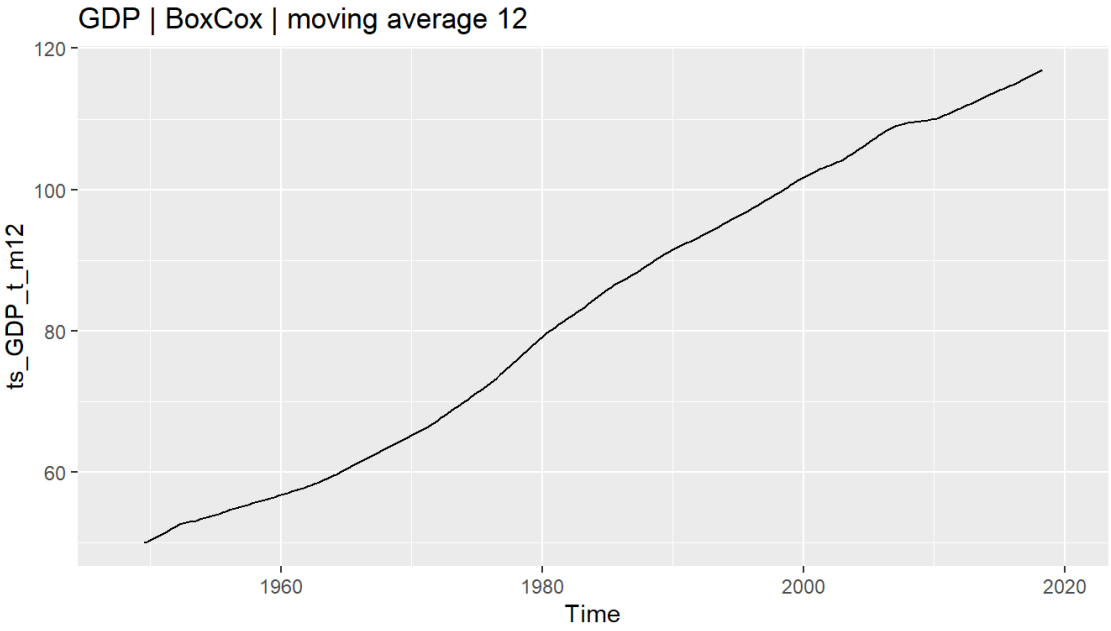
Code



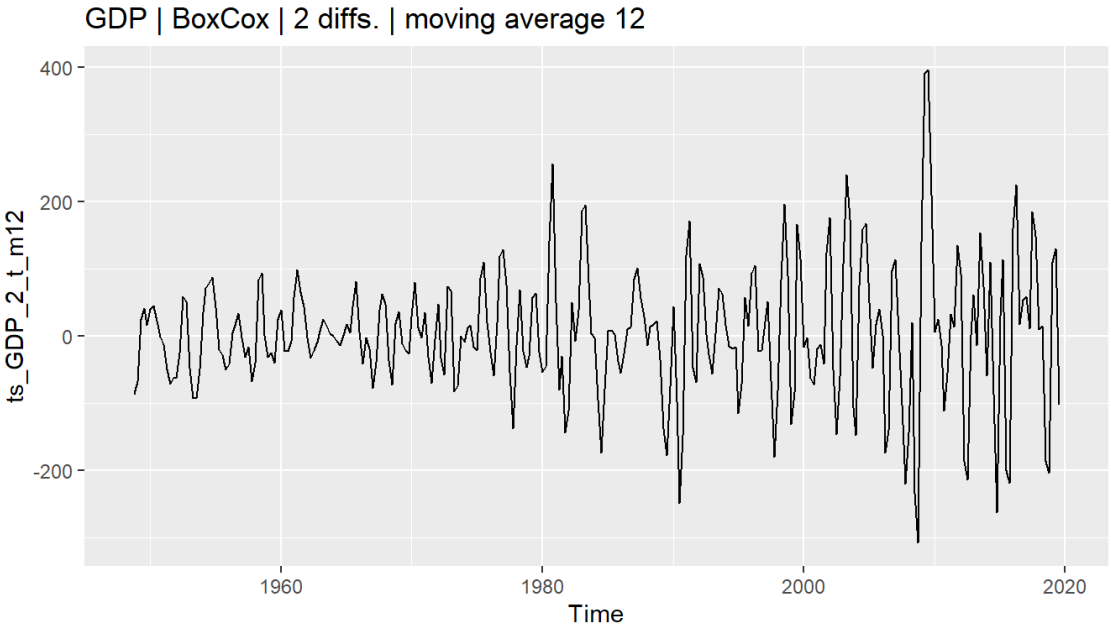
Code



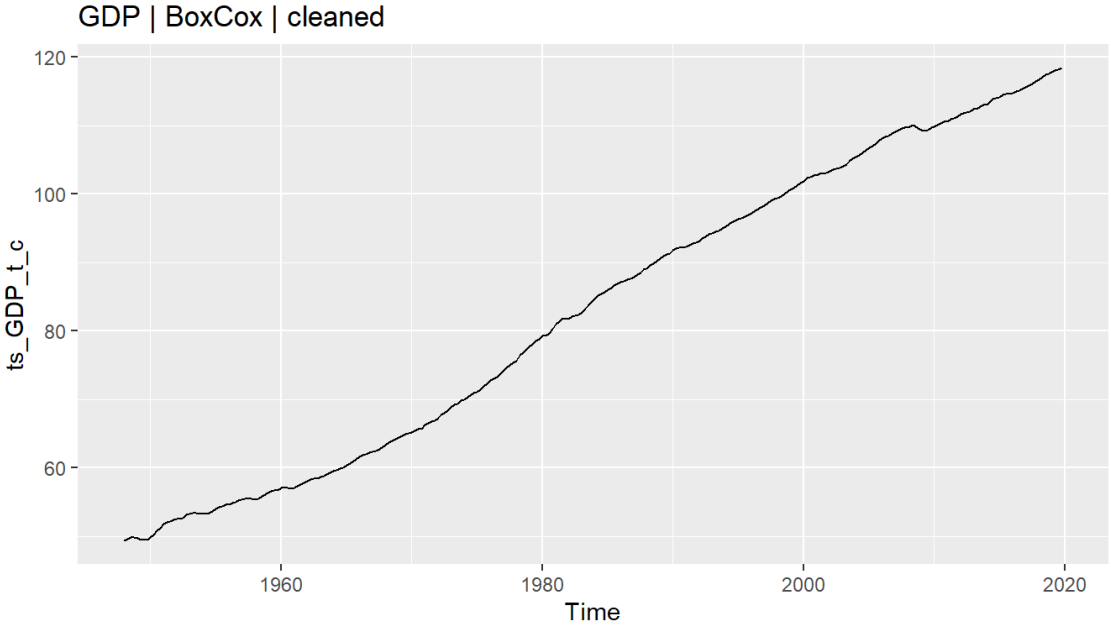
Code


GDP | BoxCox | moving average 12
ts_GDP_t_m12
120
100
80
60
1960
1980
2000
2020
Time

Code


GDP | BoxCox | 2 diffs. | moving average 12
ts_GDP_2_t_m12
400
200
0
-200
1960
1980
2000
2020
Time

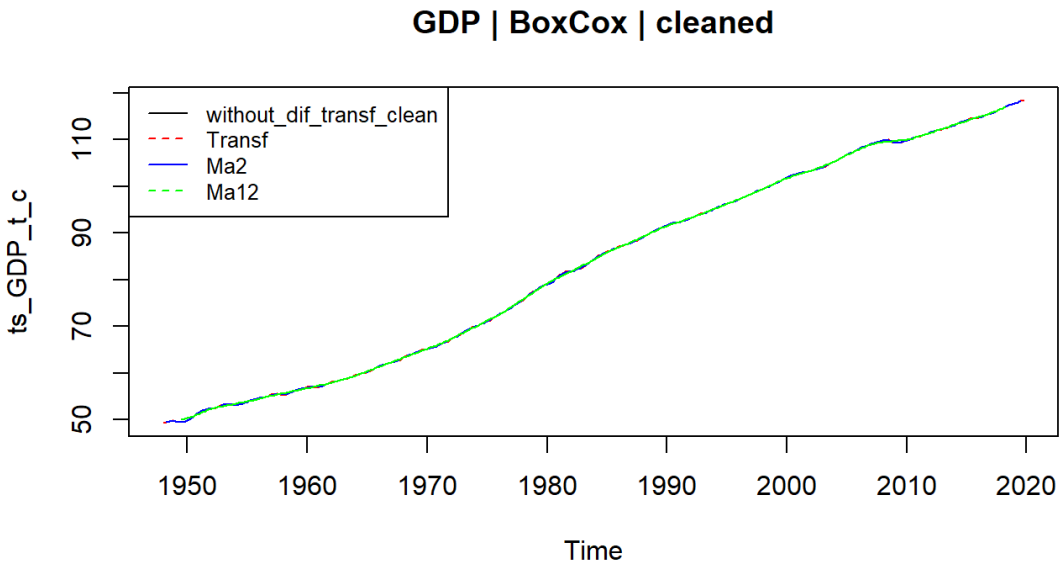
Code


GDP | BoxCox | cleaned
ts_GDP_t_c
120
100
80
60
1960
1980
2000
2020
Time

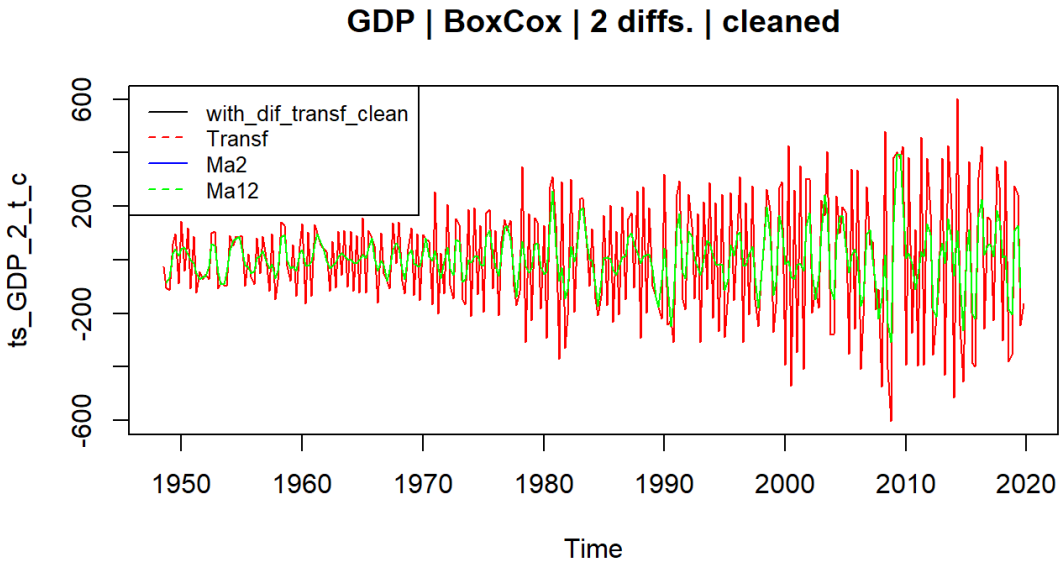
Code

COMPARING

Code



Code

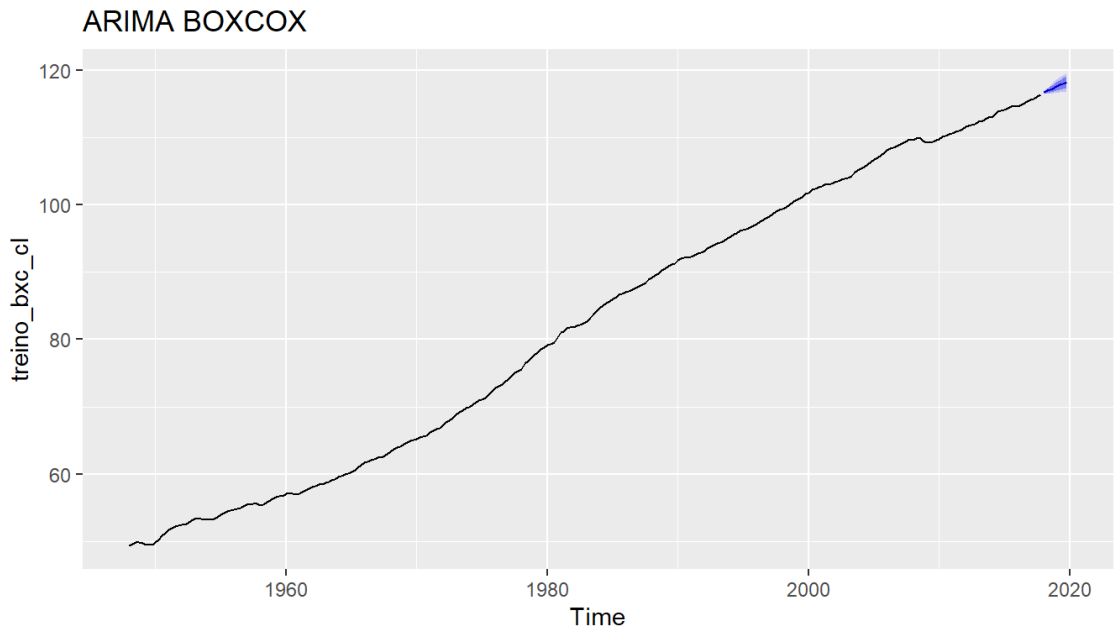


PREDICTIONS

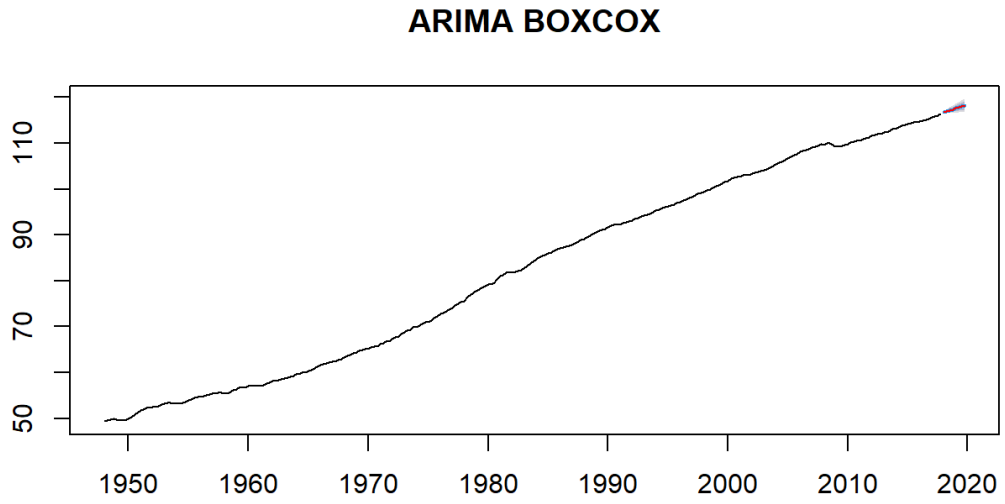
MODEL 1

TIME SERIES AFTER BOXCOX AND CLEANING PROCESS

Code



Code



ACCURACY

Code

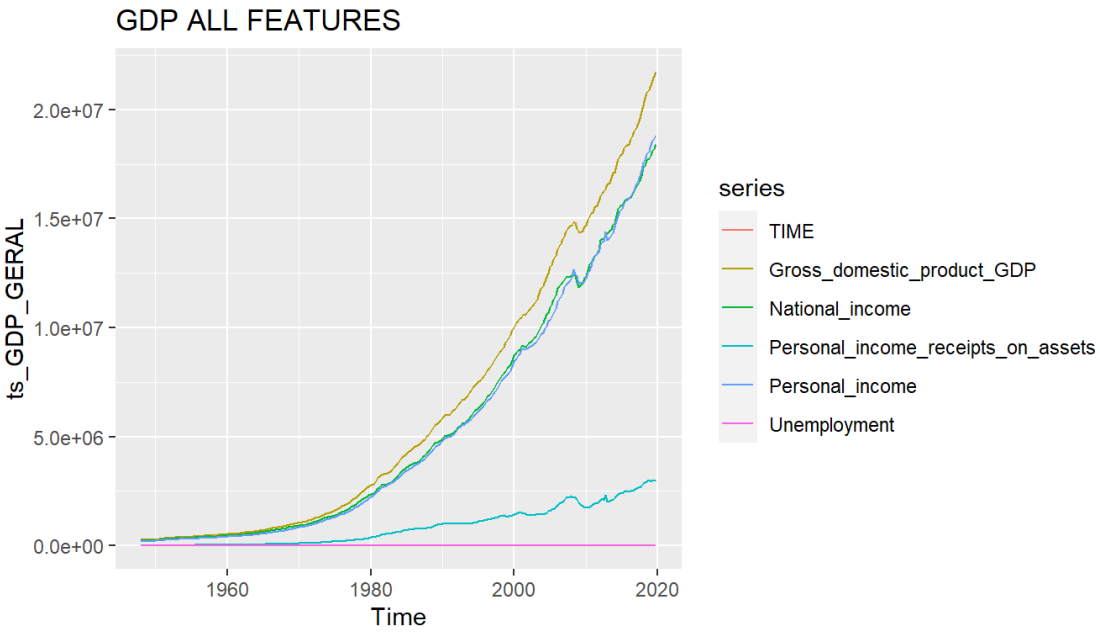
```
##                  ME      RMSE       MAE        MPE      MAPE       ACF1
## Test set -0.1384159 0.1494275 0.1384159 -0.1176566 0.1176566 0.04204106
##          Theil's U
## Test set 0.7293204
```

```
The Mean Absolute Percent Error was 12%
```

MODEL 2

REGRESSION WITH 1 COLUMN + TREND

Code



Code

The range of the unemployment variable from 1948 to 2019 will be used in xreg when creating the model.
Then we will extract the period from 2017 to 2019 (8 in total) to use forecast function

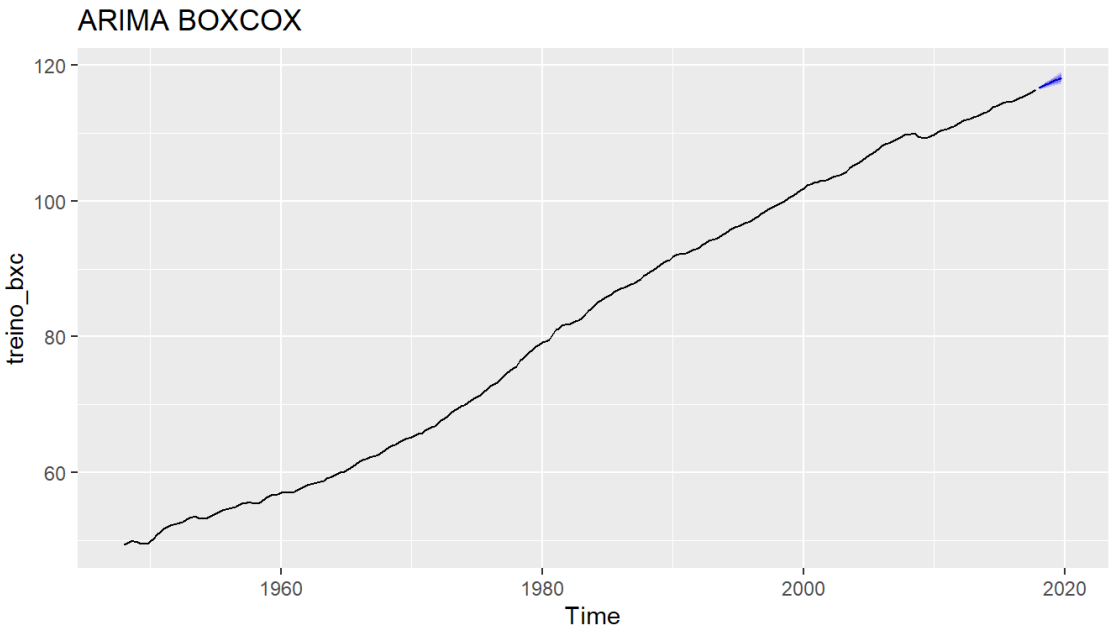
Code

PREDICT

Code

```
##         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2018 Q1       116.6549 116.5159 116.7939 116.4424 116.8675
## 2018 Q2       116.8942 116.6878 117.1006 116.5785 117.2099
## 2018 Q3       117.1277 116.8604 117.3950 116.7189 117.5365
## 2018 Q4       117.3309 117.0070 117.6547 116.8356 117.8262
## 2019 Q1       117.5089 117.1270 117.8908 116.9248 118.0930
## 2019 Q2       117.7752 117.3369 118.2136 117.1048 118.4457
## 2019 Q3       117.9713 117.4755 118.4670 117.2131 118.7294
## 2019 Q4       118.2052 117.6528 118.7576 117.3603 119.0500
```

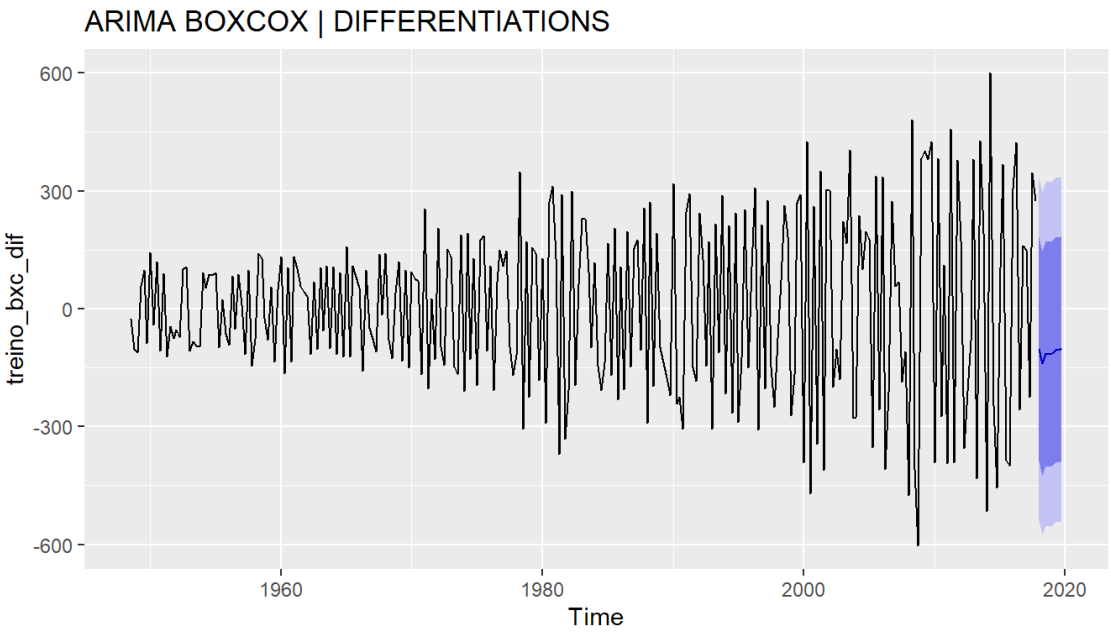
Code



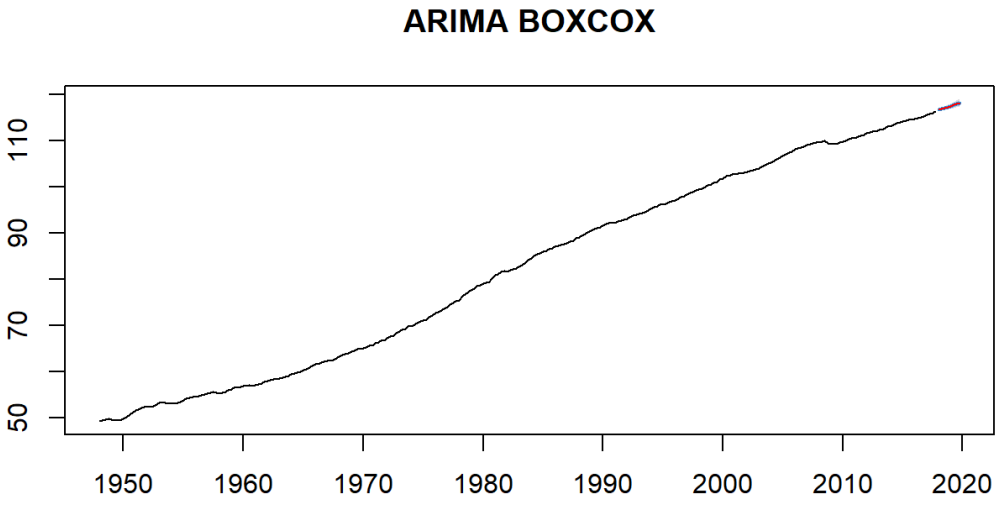
Code

```
##         Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 2018 Q1      -103.1729 -388.0749 181.7291 -538.8929 332.5470
## 2018 Q2      -139.8941 -424.8057 145.0174 -575.6288 295.8405
## 2018 Q3      -114.1542 -400.3811 172.0727 -551.9005 323.5921
## 2018 Q4      -115.8717 -402.2513 170.5079 -553.8515 322.1081
## 2019 Q1      -113.8635 -400.3846 172.6576 -552.0597 324.3327
## 2019 Q2      -105.6083 -392.1757 180.9592 -543.8754 332.6589
## 2019 Q3      -104.3169 -390.9084 182.2745 -542.6207 333.9869
## 2019 Q4      -100.6827 -387.2842 185.9187 -539.0018 337.6363
```

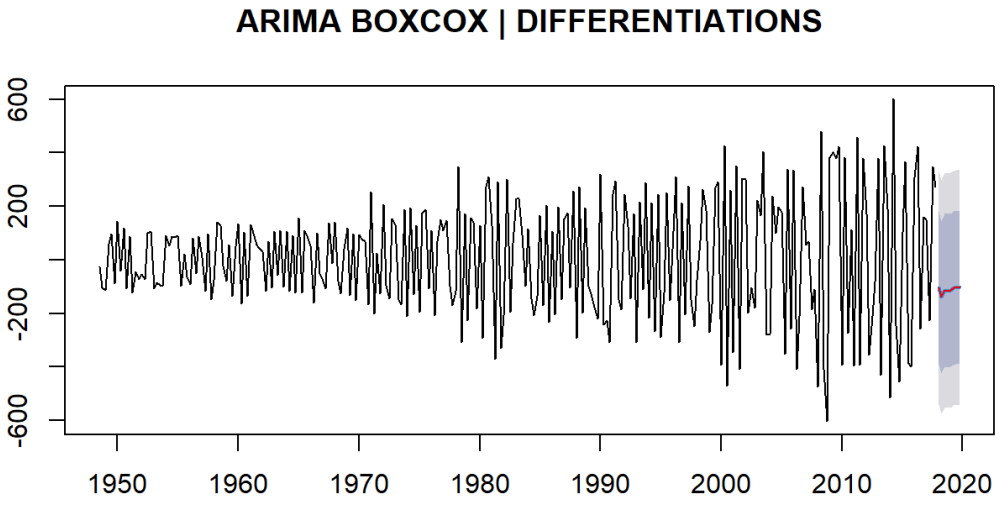
Code


ARIMA BOXCOX | DIFFERENTIATIONS
treino_bxc_dif
600
300
0
-300
-600
1960
1980
2000
2020
Time

Code


ARIMA BOXCOX
50
70
90
110
1950
1960
1970
1980
1990
2000
2010
2020

Code


ARIMA BOXCOX | DIFFERENTIATIONS
600
200
-200
-600
1950
1960
1970
1980
1990
2000
2010
2020

ACCURACY

Code

```
##                 ME      RMSE      MAE        MPE      MAPE       ACF1 Theil's U
## Test set -0.214725 0.2222662 0.214725 -0.1827175 0.1827175 0.06874263  1.057604
```

The Mean Absolute Percent Error was 18%

Code

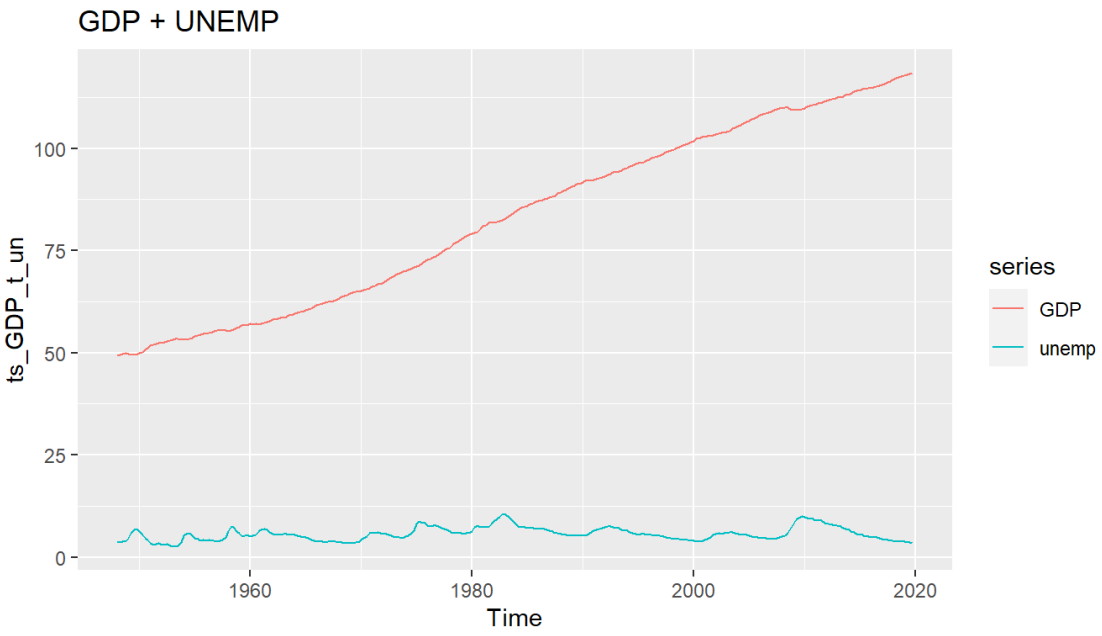
```
##                 ME     RMSE     MAE      MPE    MAPE       ACF1 Theil's U
## Test set -42.30404 300.9591 269.594 25.69617 233.415 -0.2784379   18.1507
```

The model with the series after 2 differentiations was not efficient.

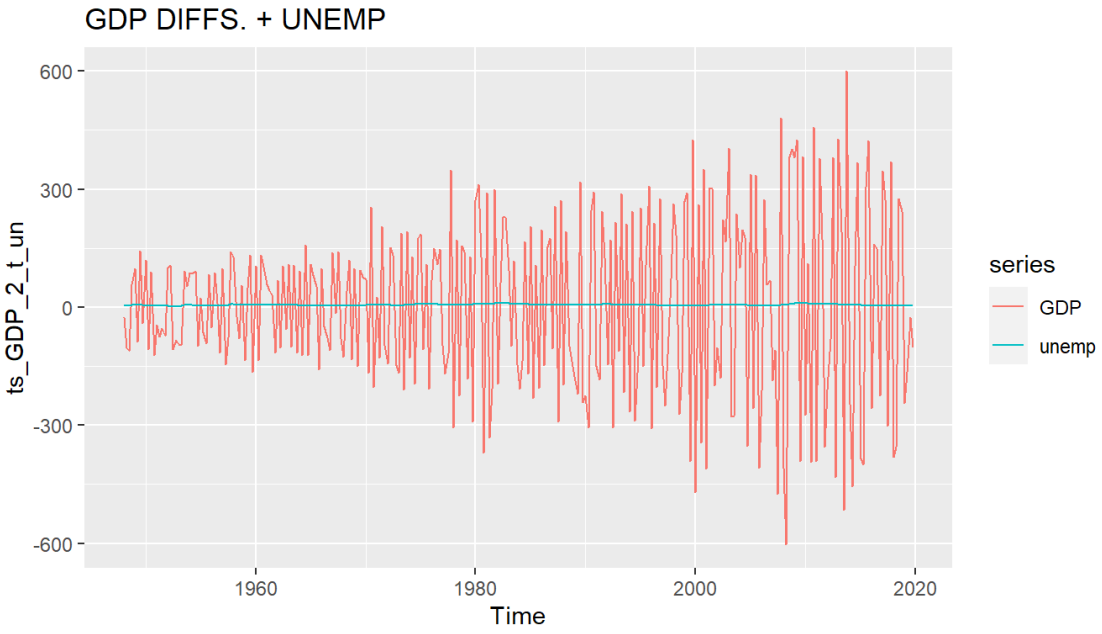
MODEL 3

ARIMA WITH 1 COLUMN + TREND

Code



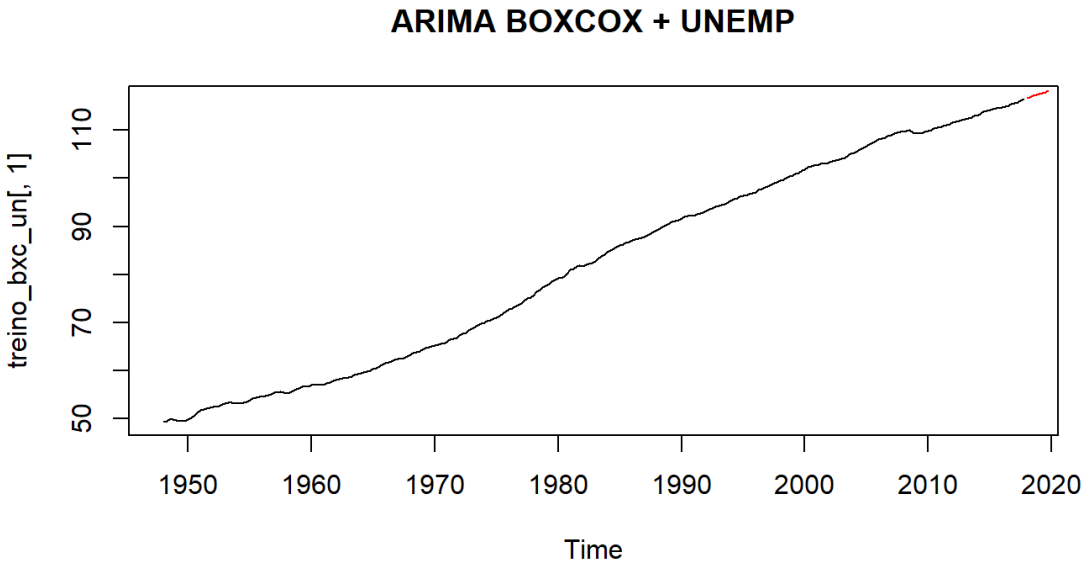
Code



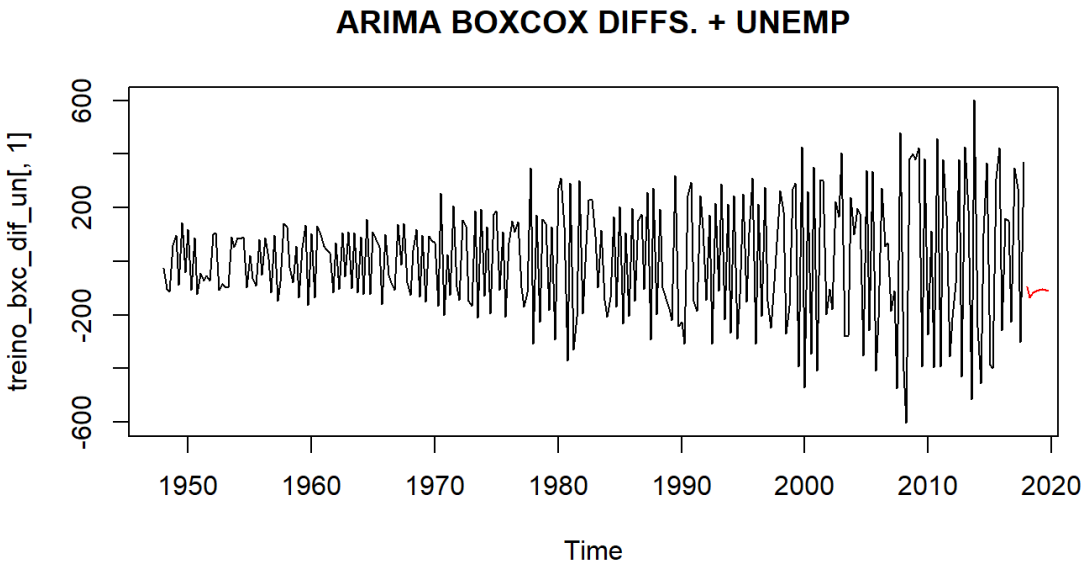
Code

PREDICT

Code



Code



ACCURACY

Code

```
##                ME      RMSE      MAE       MPE      MAPE       ACF1 Theil's U
## Test set 0.214725 0.2222662 0.214725 0.1823607 0.1823607 0.06874263 0.9323222
```

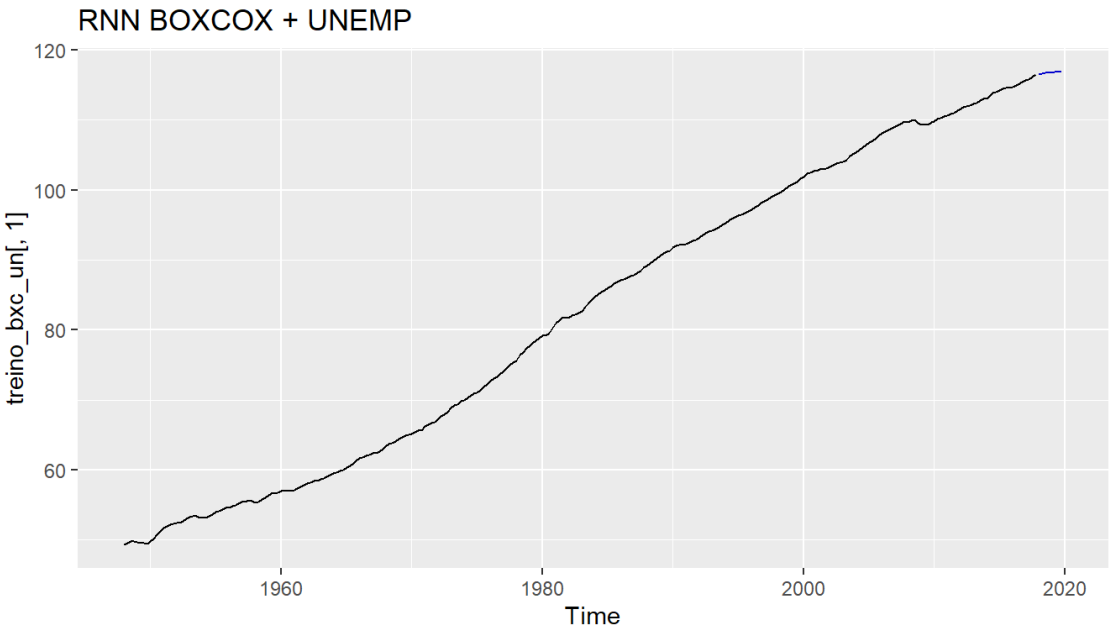
Code

```
##                ME     RMSE     MAE      MPE     MAPE      ACF1 Theil's U
## Test set 16.19244 234.4263 192.607 23.73014 105.9712 0.1462585 0.4740075
```

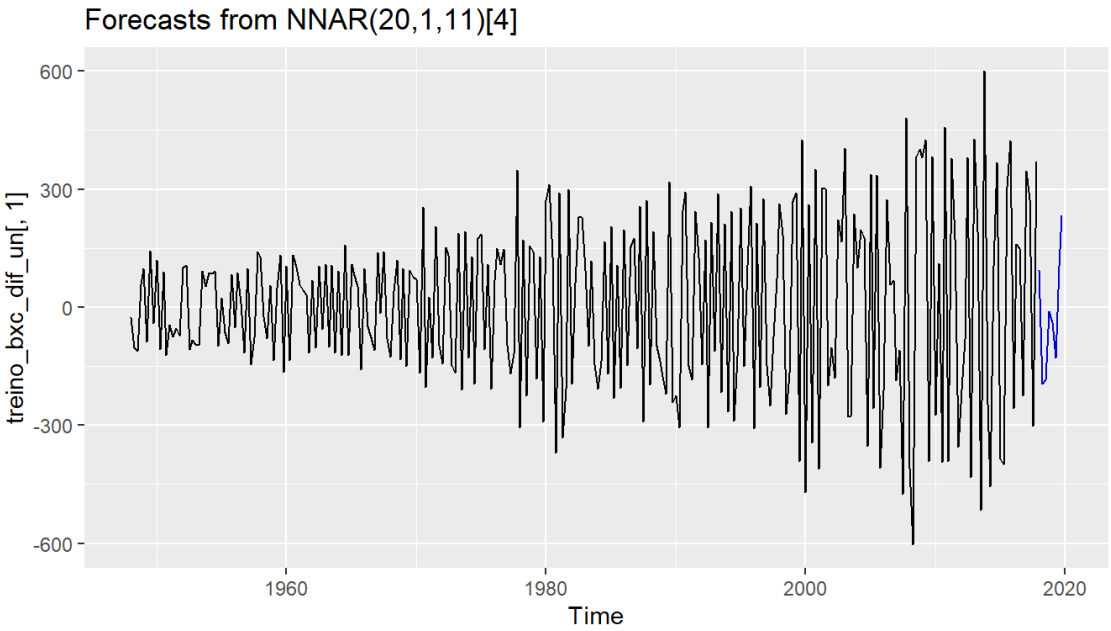
MODEL 4

RECURRENT NEURAL NETWORK (RNN)

Code



Code



ACCURACY

Code

```
##                ME     RMSE       MAE       MPE      MAPE      ACF1 Theil's U
## Test set 0.8996696 1.001543 0.8996696 0.7630121 0.7630121 0.5943856  4.208843
```

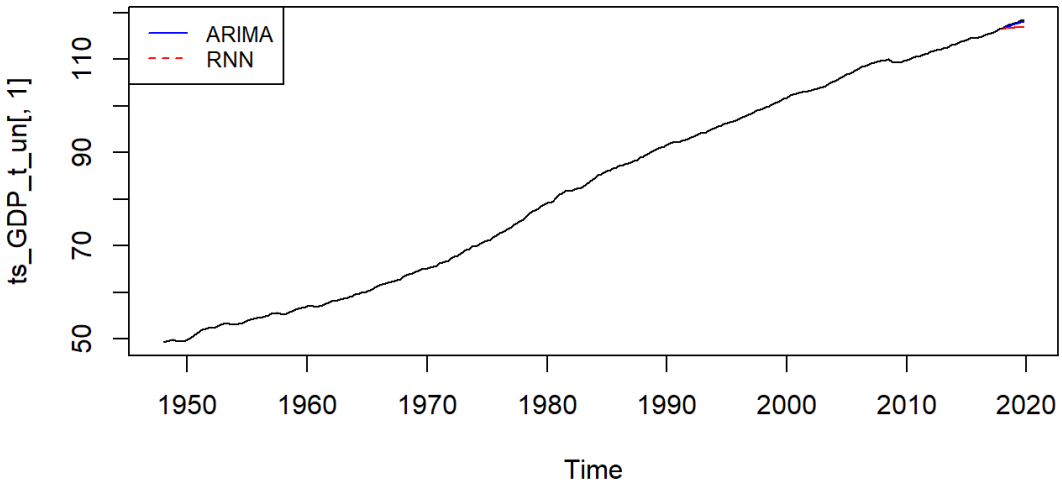
Code

```
##                ME     RMSE     MAE      MPE     MAPE      ACF1 Theil's U
## Test set -78.39287 295.7262 256.973 171.9638 171.9638 0.2009339  3.189306
```

COMPARING ARIMA AND RNN

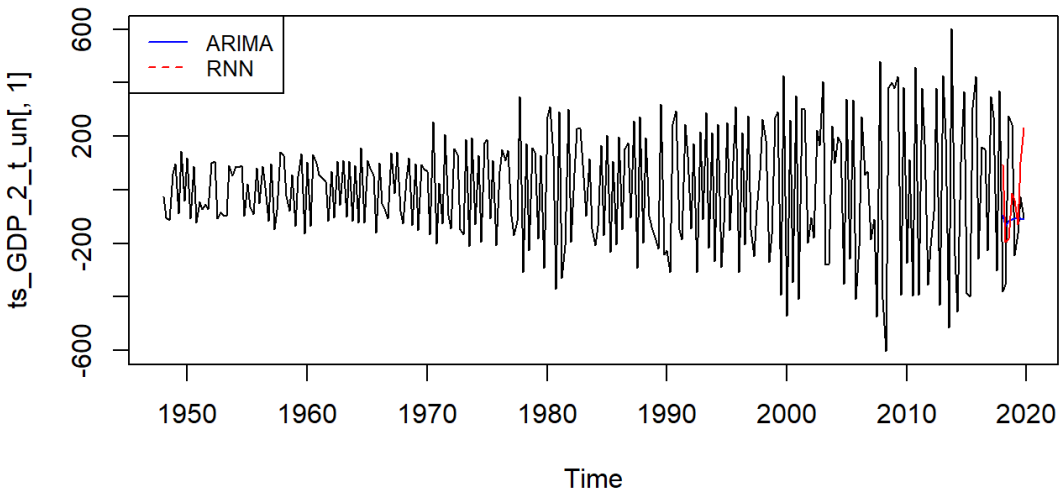
Code

CPMPARING MODELS



Code

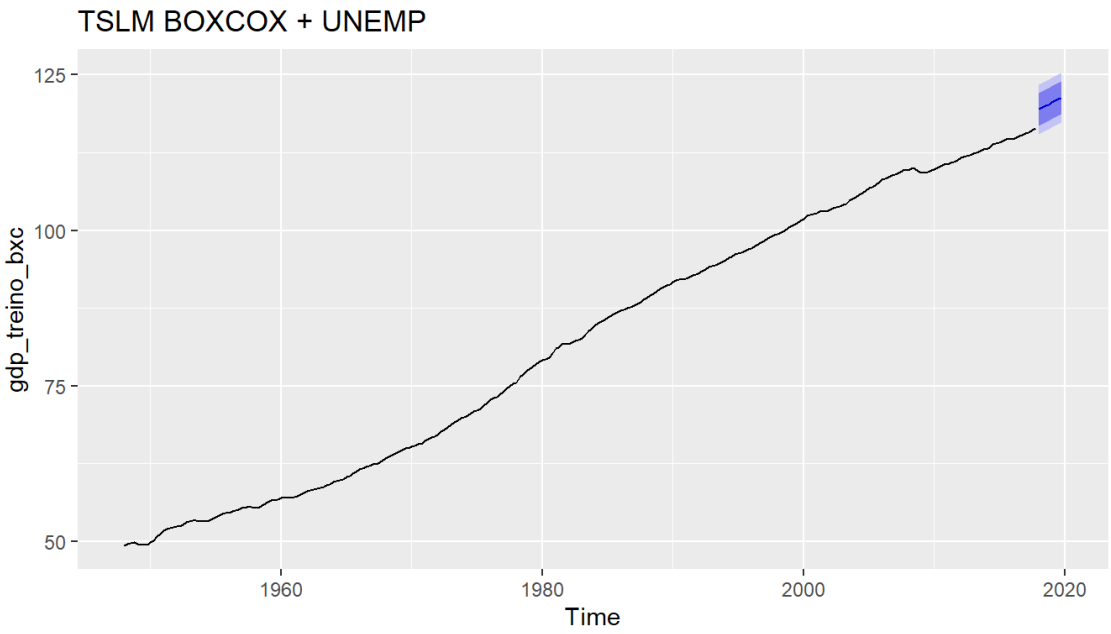
COMPARING MODELS DIFFS.



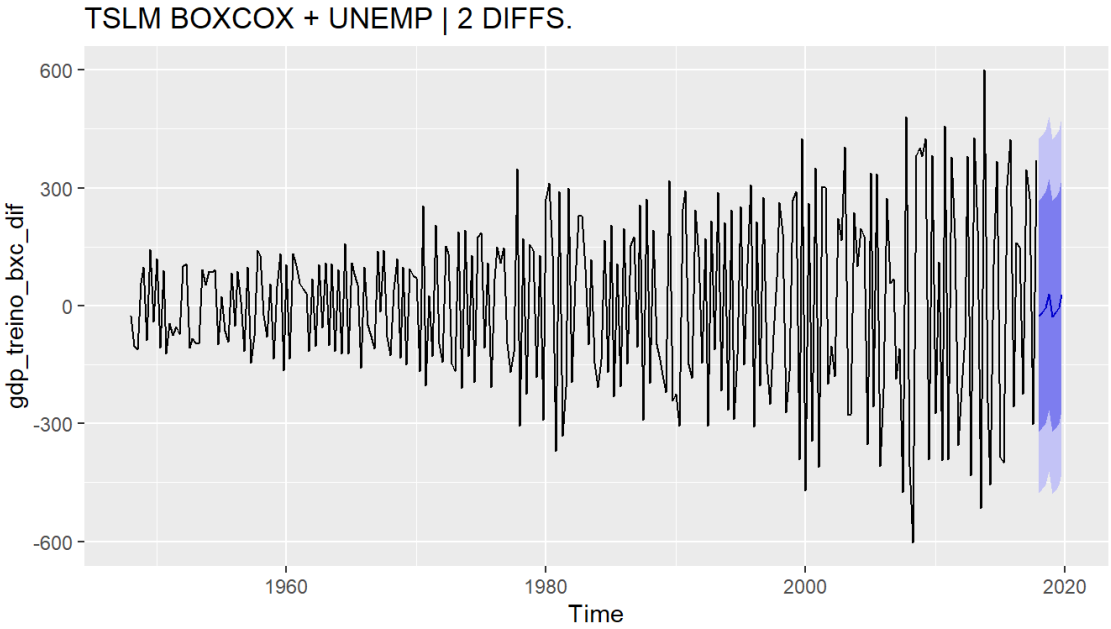
MODEL 5

REGRESSION FOR TIME SERIES (TSLM)

Code



Code



ACCURACY

Code

```
##                ME    RMSE      MAE       MPE     MAPE     ACF1 Theil's U
## Test set -2.763548 2.76512 2.763548 -2.348766 2.348766 0.624693   10.9208
```

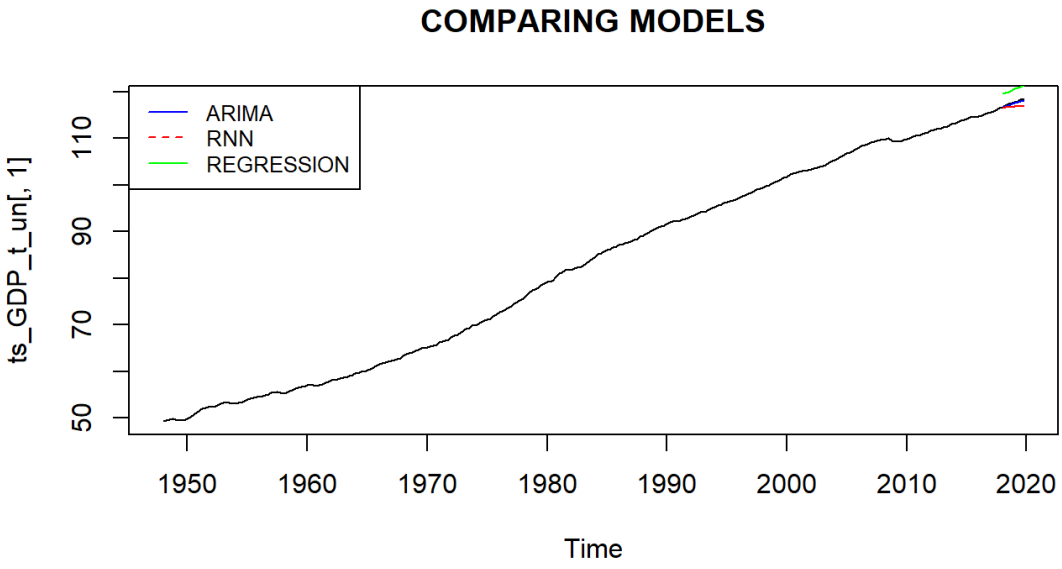
Code

```
##                ME     RMSE      MAE      MPE     MAPE      ACF1 Theil's U
## Test set -89.28551 237.5145 212.8591 95.24695 95.24695 0.1239704  1.298814
```

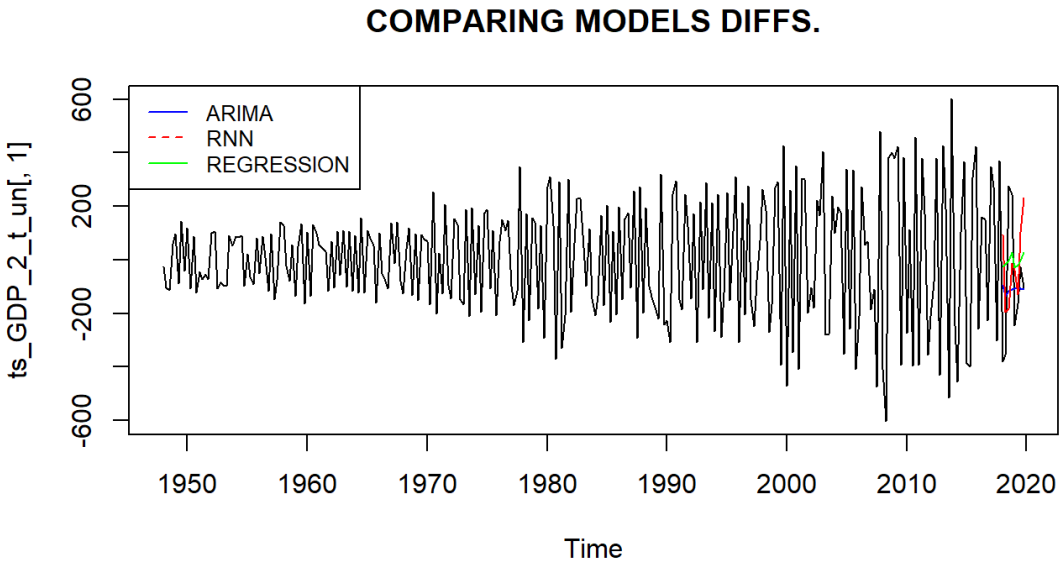
COMPARING 3 BEST MODELS

ARIMA, RNN AND TSLM

Code



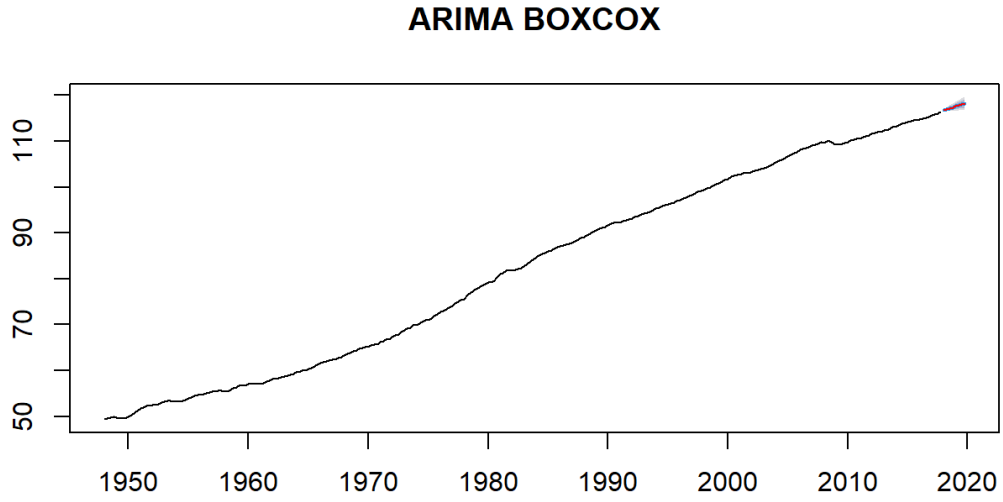
Code



CONCLUSION

The best model was the first one (MODEL 1). it is a Forecast, applied in ARIMA (2,2,1), built with a set of data after BoxCox transformation and process of cleaning outliers. Its Mean Absolute Percent Error was 12%. Below is a model's summary:

Code



ACCURACY

Code

```
##                    ME      RMSE       MAE        MPE      MAPE       ACF1
## Test set -0.1384159 0.1494275 0.1384159 -0.1176566 0.1176566 0.04204106
##          Theil's U
## Test set 0.7293204
```

Code