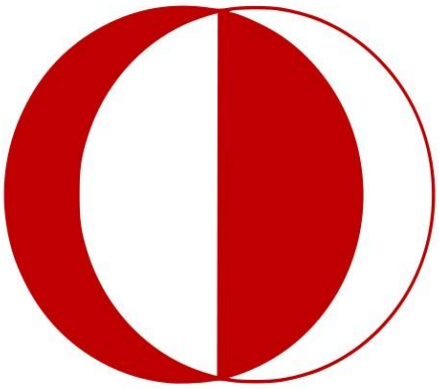


Middle East Technical University Department of Statistics

STAT 497 TIME SERIES PROJECT

TERM PAPER

Submitted to Assoc. Prof. CEYLAN TALU YOZGATLIGİL

2146181 Bennur Kaya

2146116 Merve Erşahin

CONTENT

INTRODUCTION

The Russian Federation is a Eurasian country established following the dissolution of the Soviets in 1991.

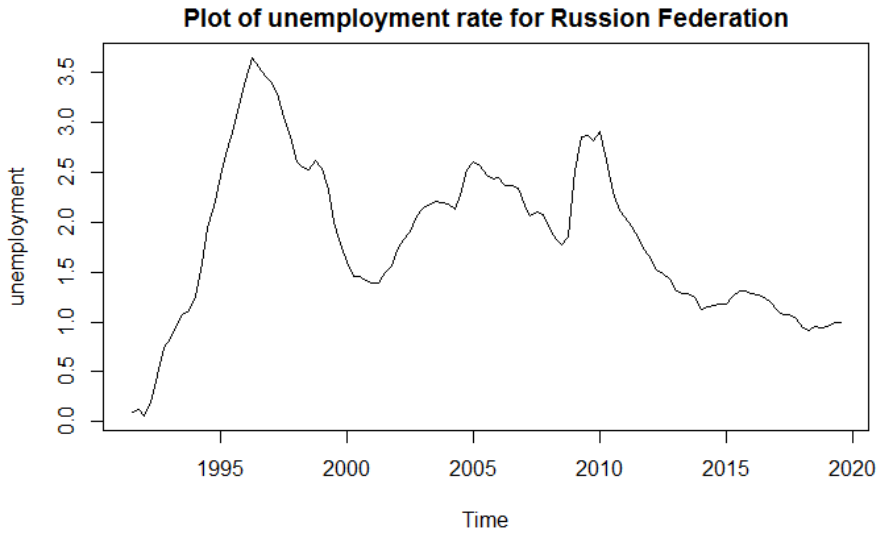
During the Soviet period, the administration was governed by communism, which established a social order based on the common ownership of the means of production. Following the dissolution of the Soviet Union, the Russian Federation experienced a process of transition to the principle of separation of powers (Executive, Legislative, Judicial). In the process, the most problematic situation for the country was the loss of earnings from trade. The gain from trade with the former USSR republics corresponded to more than half of RF foreign trade. For this reason, the RF made a lot of arrangements to recover its economy. This country, whose economy has been very unstable over the years, suffered an economic crisis in 1998. But it quickly recovered from the economic crisis with its income from oil in 1999.

This study aims to analyse Unemployment Rate for the Russian Federation between the years 1991 and 2019. The dataset 'Registered Unemployment Rate for the Russian Federation' used in the study is taken from foreign trade part of Federal Reserve Bank of St. Louis and it includes unemployment rate according to years.

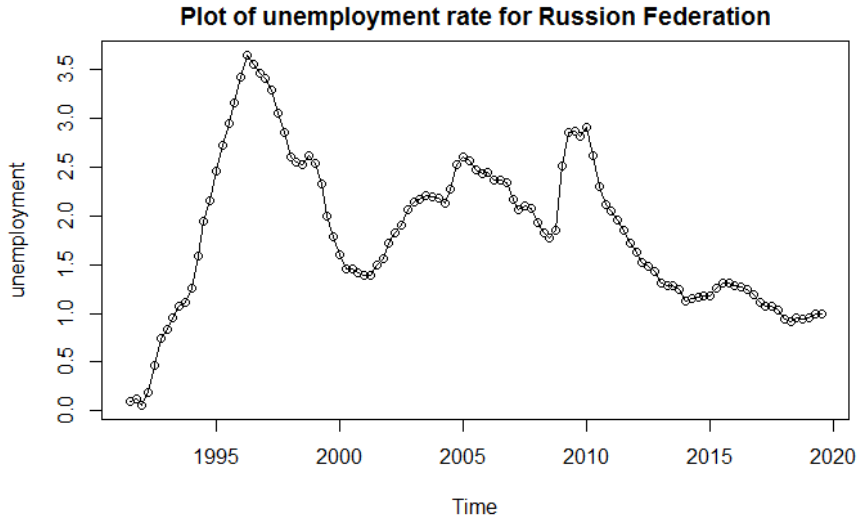
This study will be examined by making time series analysis of unemployment rate in Russian Federation between 1991 and 2019 years with R programming language. Observations of last 5 years will be forecasted since the frequency of data is quarterly.

2. METHODOLOGY

2.1. Plotting the time series



Graph1.Line plot of unemployment rate over the years

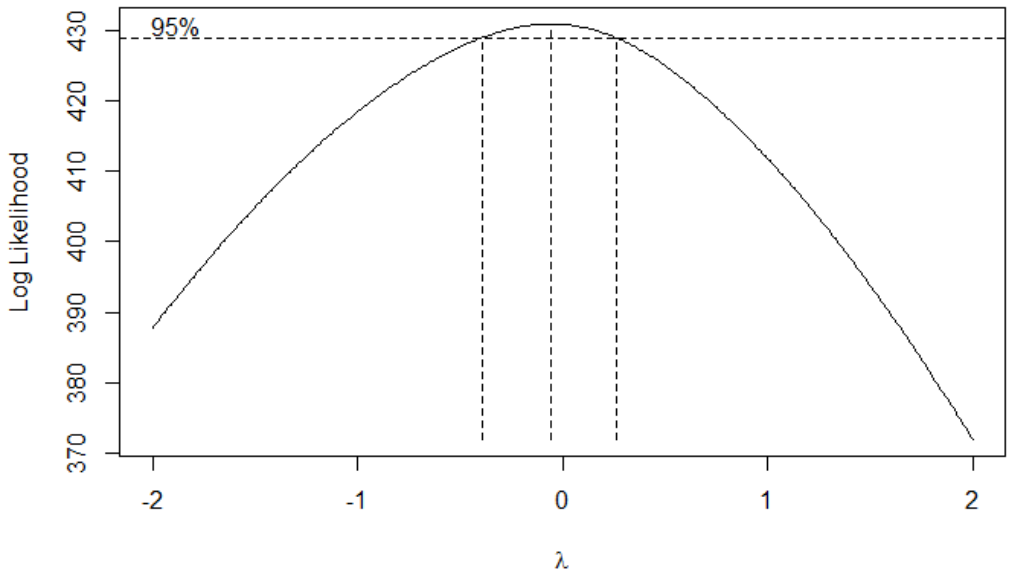


Graph2.Dotted line plot of unemployment rate over the years

The unemployment rate versus year plots of all observations shows that there seems a sharp increasing until 1996-1997 because of the some political reasons. Although there is a decline for unemployement after those years, there are fluctuations over the time for various reasons. After plotting all observations, last 5 observations of unemployment rates (2018-2019) are removed from data and kept as test data for forecasting.

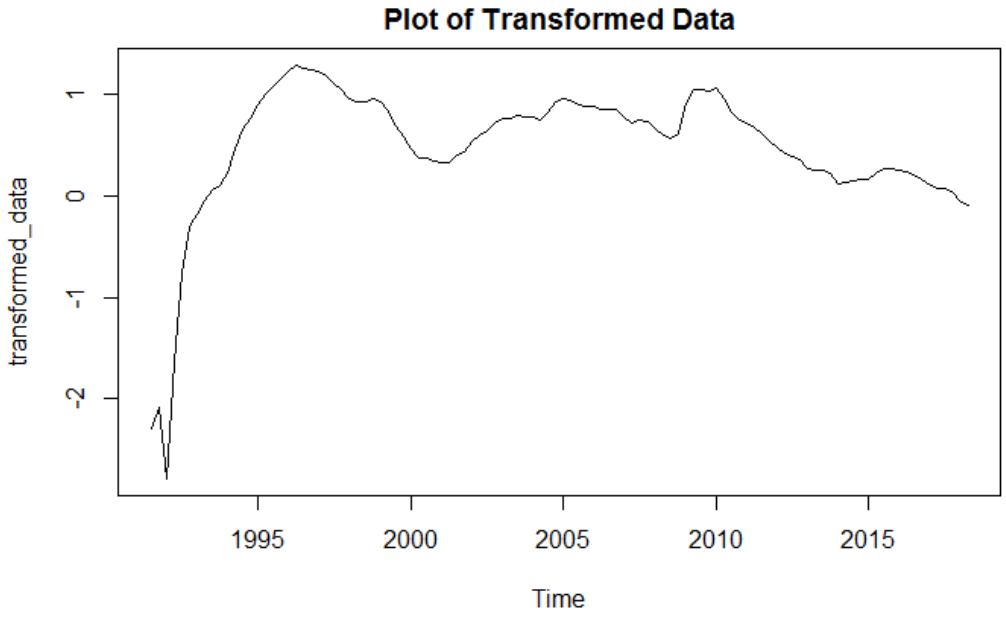
2.2 Box-Cox Transformation Analysis

Transformation may be needed to make a better and proper analysis. In order to see which type of transformation is required, plotting the lambda values of the series helps to determine the appropriate transformation.



Graph3. Graph of lambda by Ordinary Least Square Method

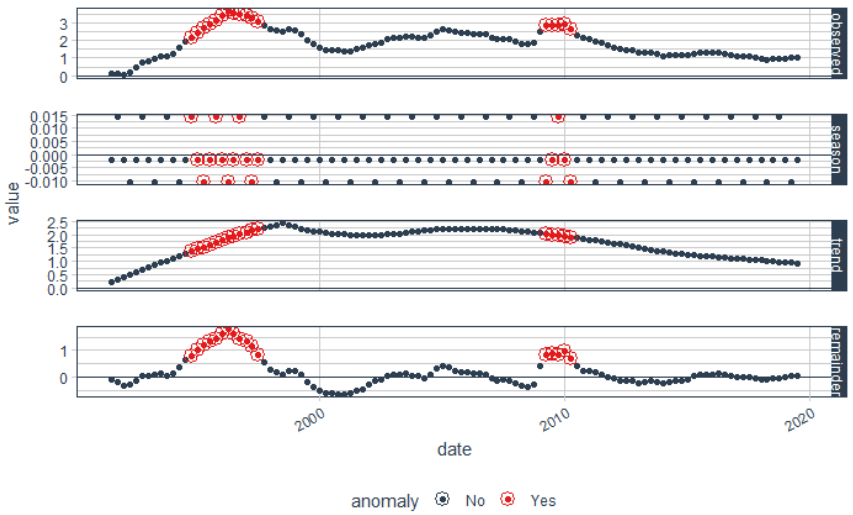
According to log likelihood versus lambda graphs by Ordinary Least Square method, since lambda is close to 0 and the exact lambda value is seen as 0.5, logarithmic transformation is needed.



Graph4. Plot of transformed unemployment rate

After transformation is applied to the train data, the plot shows that there is an steady increasing trend until 1995 and the data does not seem stationary.

2.3 Anomaly Detection



Graph 5. Anomaly Detection Plot

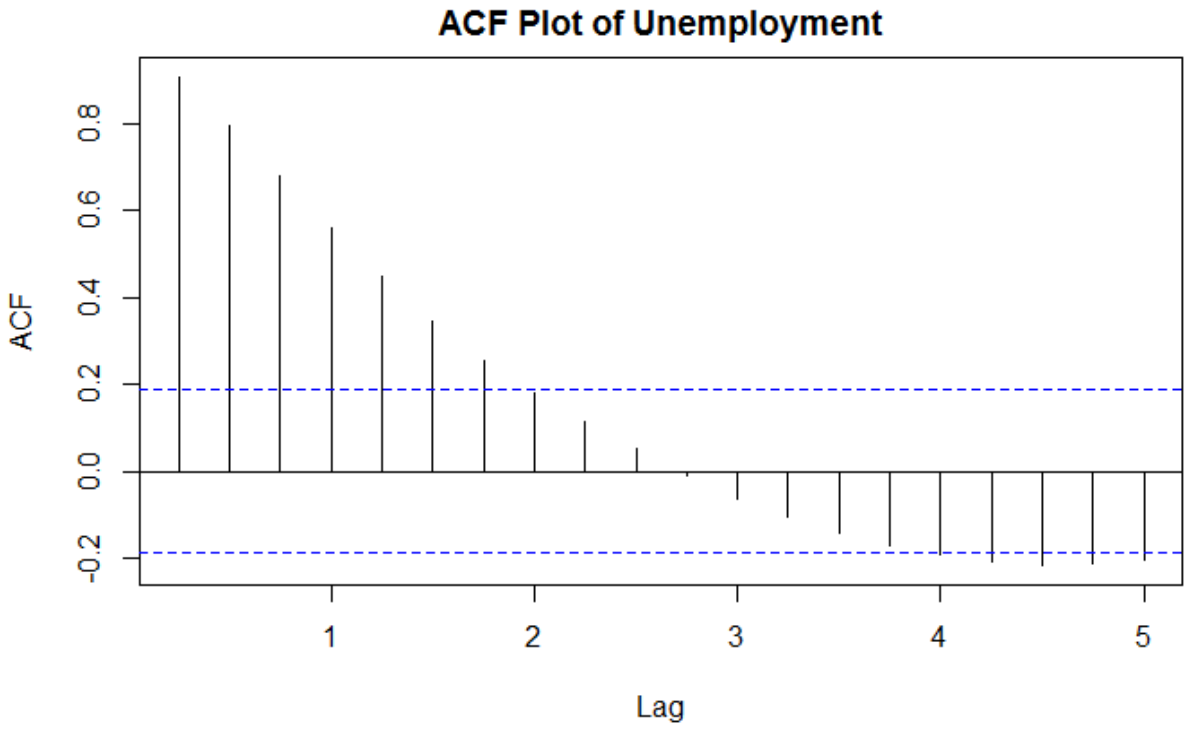
As it can be seen from the anomaly detection plot, there are some anomalies in the data. These outliers has been deleted.

2.4 ACF, PACF Plots, KPSS, PP and ADF Tests

In order to see whether the series is stationary and ready to analyze , ACF and PACF plots, and several tests are applied.

ACF Plot

ACF plot is the autocorrelation function plot. When ACF plot shows that the series is not stationary, there is no need to look at the PACF plot.



Graph 6.ACF plot of train data

According to the ACF plot, there is a slow decay in lags. Thus, the series is not stationary and seems to have unit root. There is no need to check PACF plot of the series.

After looking at the ACF plot and stating that the series is not stationary, unit root tests are applied.

KPSS TEST

KPSS test is applied in two levels in order to check whether the series is stationary or not and if it is not stationary , what the trend of the series is.

KPSS level test is used to state whether a series is stationary or not. The null hypothesis of KPSS level test is the series is stationary. That is, if the null hypothesis is rejected, it means the series is not stationary and there seems a trend. Therefore KPSS trend test is the next step for nonstationary series.

The result of the KPSS level test:

KPSS Level	Truncation lag parameter	p-value
0.2762	4	0.1

Table 1.KPSS test for level stationary

According to KPSS level test, the p-value 0.1 is greater than 0.05 significance level, the null hypothesis is failed to reject. That means the series is stationary. However, the series does not seem stationary according to ACF plot, clearly. So, other unit root tests are applied to be sure.

PP TEST

Hypothesis for PP test:

Null hypothesis: Non-stationarity exists

Alternative hypothesis: Non-stationarity does not exist.

In the PP test results, p- value (0.5727) is greater than 0.05, it indicates that fail to reject the null hypothesis. Therefore, the series is not stationary.

ADF TEST

Hypothesis for ADF test:

Null hypothesis: The process has unit root (non-stationary)

Alternative hypothesis: The process is stationary

Similarly, in the ADF test results, p- value (0.01) is smaller than 0.05, it indicates that reject the null hypothesis. Therefore, the series is not stationary so that differencing is needed.

HEGY TEST

HEGY test is applied to detect the type of unit root (i.e reason of stationarity). We have the following hypothesis for testing regular unit root and seasonal unit root respectively.

Regular Unit Root

Ho: The system has a regular unit root.

H1: The system does not contain any regular unit root.

Seasonal Unit Root

Ho: The system has a seasonal unit root.

H1: The system does not contain any seasonal unit root.

Stat.	p-value
tpi_1	0.10
tpi_2	0.01
Fpi_3:4	0.01
Fpi_2:4	NA
Fpi_1:4	NA

Table2.Hegy table

In this output, p-value of t_1 and the p-value of F_3:4 are used for testing regular and seosonal unit root, respectively.

The output shows that the system has only regular unit root because p value of t_1 is greater than α value.

To solve this problem, regular differencing is needed.

2.5 Trend Removing

After one differencing, KPSS test is necessary to check the current stationary of the series.

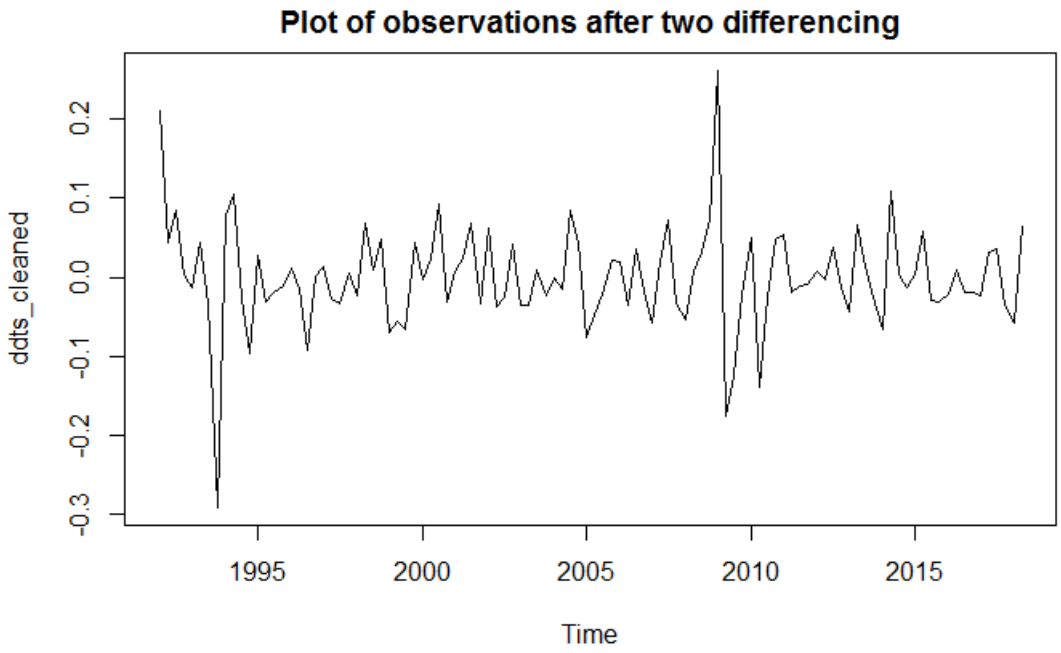
Firstly, KPSS level test is applied and the result is:

KPSS Level	Truncation lag parameter	p-value
0.72459	4	0.01

Table3. KPSS test for level stationarity after a differencing

KPSS test for level stationary indicates that the p values 0.01 is smaller than 0.05. This means that the series is still not stationary after one regular differencing.

HEGY test is conducted one more time and it is seen that regular unit root problem is not removed. So, one more regular differencing is taken.



Graph 7. The plot of the observations after two differencing

After two regular differencing, the graph of observations seems stationary in mean.

The mean of the differencing data is -0.000234 , so it is clearly seen that the mean is close to zero. Therefore, KPSS test is applied after two regular differencing to check the last situation.

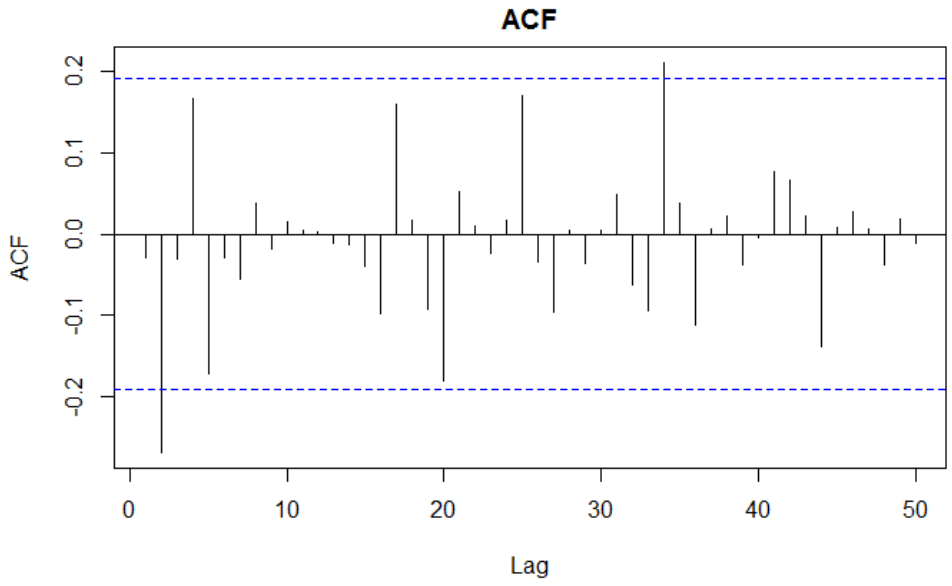
KPSS Level	**Truncation lag parameter**	**p-value**
0.038	4	0.1

Table4. KPSS test for level stationarity after two differencing

The null hypothesis means that the series is stationary and the p-value of test results is greater than 0.05. Thus, the null hypothesis is failed to reject. In other words, there is no unit root so the series is stationary anymore.

2.6 ACF - PACF

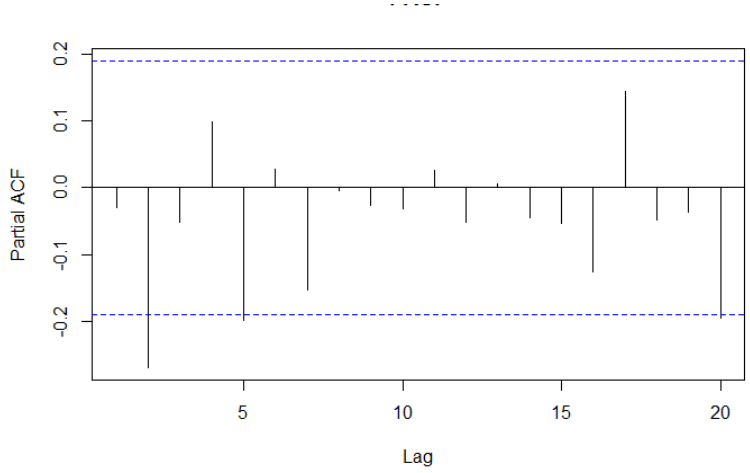
After checking test results and taking two difference, the series became stationary. However, it is needed to plot ACF and PACF graphs.



Graph 8.ACF plot of the series after two regular differencing

In ACF plot, most of the lags inside the white noise bands and there seems no slow decay.

PACF Plot of Differenced Series



Graph 9.PACF plot of the series after two regular differencing

Likewise the ACF plot, most of the lags inside the white noise bands and there seems no slow decay in PACF plot.

ESACF plot is not conducted since the data is quarterly.

2.7 MODEL IDENTIFICATION

By checking the ACF- PACF plots, the most appropriate model is ARIMA(2,2,2). The other model which is Arima(1,2,2)(1,0,0) is suggested by auto.arima function in R programming.

First fit is composed as ARIMA(2,2,2) which comes up with the values:

$\sigma 2$	0.00384
log likelihood	143.54
AIC	-279.07

Table5. Output of first fit

Second fit is composed by auto.arima function which is used to determine the most proper model for the series. According to the output, the most proper model is selected as Arima(1,2,2)(1,0,0) :

$\sigma 2$	0.00378
log likelihoodshs	146.68
sjjAIC	-282.76

Table6. Output of second fit

AIC informations criteria of two fits are compared and second fit is selected (Arima(1,2,2)(1,0,0)) most proper model for the series since its AIC information criteria is smaller than the first one.

2.8 MODEL SELECTION

After selecting the most proper model, maximum likelihood estimation is applied to the model and there seems not any convergence problem. Indeed, there should be more possible proper fits but for this series, there can be only one model. Therefore, comparison of maximum likelihood estimation of different fits cannot be applied.

2.9 DIAGNOSTIC CHECKING

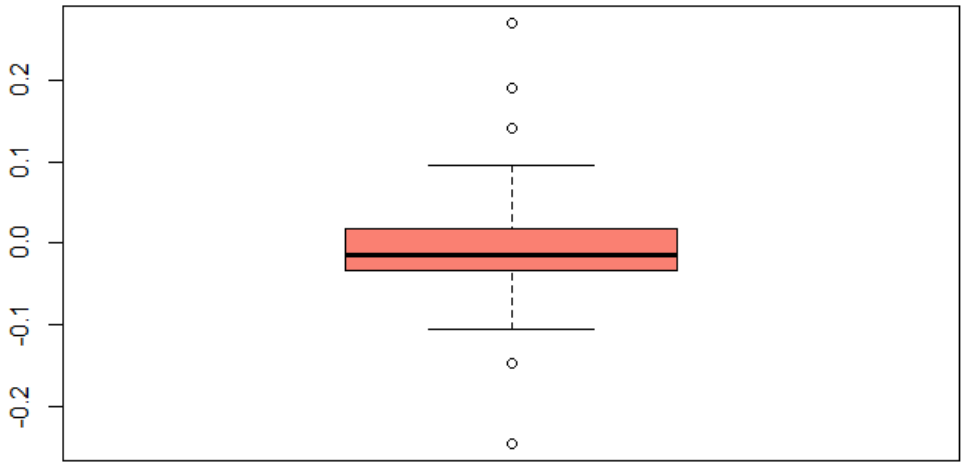
2.9.1. Formal Tests and Residual Checks

Firstly, residuals of selected fit are found. Later, their summary statistics values are calculated as:

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-0.245261	-0.033896	-0.013718	-0.005929	0.017775	0.270451

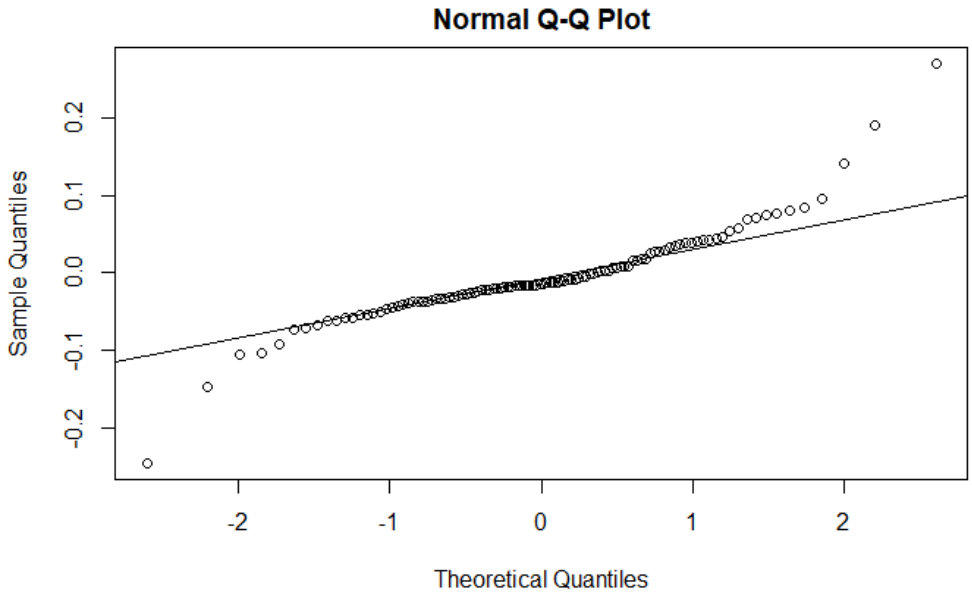
Table7. Summary statistics values of the residuals

After that, box plot of residuals are drawn to see how they are distributed and whether there are outliers or not.



Graph10. Box plot of residuals

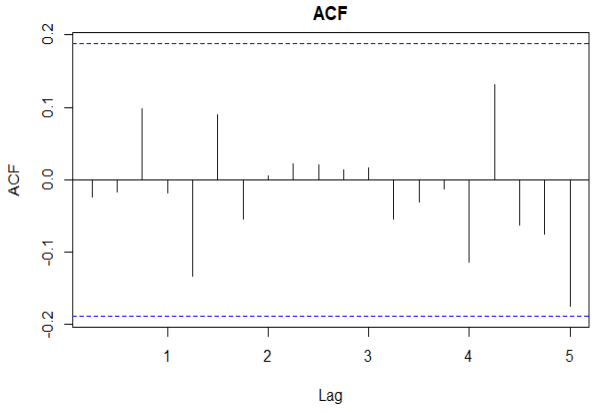
Box plot of residuals shows shat residuals seems normally distributed. However, there are outliers. Normality should be checked by conducting formal tests.



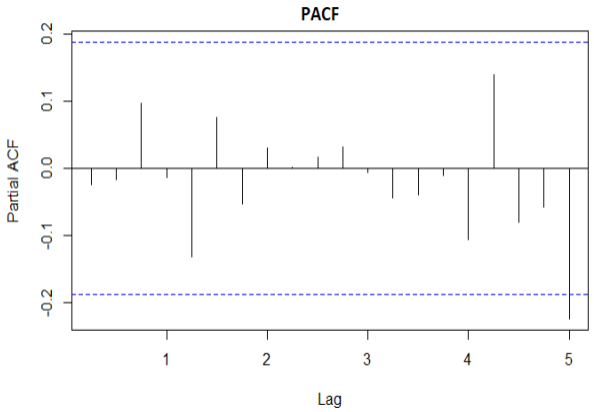
Graph11. Normal Q-Q Plot of residuals

According to the plot of residuals, there seems stationarity around mean 0. Also, there are several points which should be checked for outlier detection.

2.9.2 Autocorrelation Check



Graph12. ACF plot of residuals



Graph13. PACF plot of residuals

All lags are inside of the white noise bands for both ACF and PACF plots. Therefore, it seems there is no autocorrelation. After examining the plots, formal tests are applied.

All tests have the same hypothesis as
Ho:There is no serial autocorrelation in residuals.

Box-Ljung test

X-squared	degrees of freedom	p-value
5.2341	15	0.99

Table 8. Table of Box-Ljung test results

Box-Pierce test

X-squared	degrees of freedom	p-value
4.8632	15	0.9932

Table 9. Table of Box-Pierce test results

Breusch-Godfrey test

LM test	degrees of freedom	p-value
4.8632	15	0.9953

Table 10. Table of Breusch-Goldfrey test results

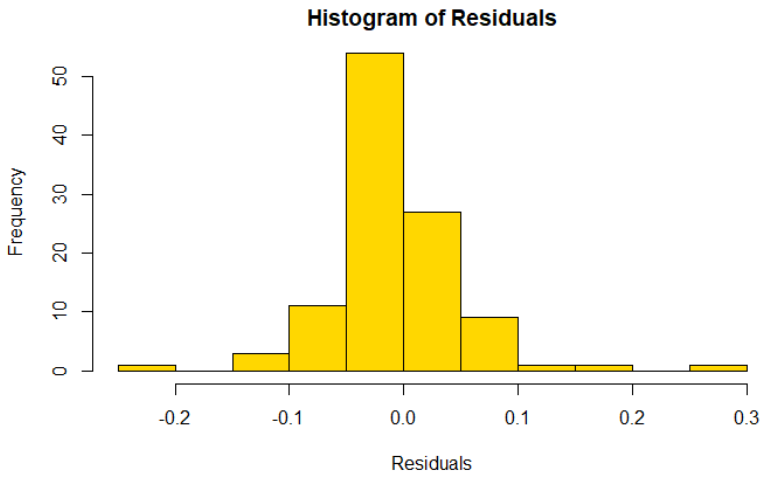
The result of formal tests give p-value is greater than significance level 0.05, so the null hypothesis is failed to reject. That means that there is no serial autocorrelation problem. Residuals are not correlated with each other.

2.9.3. Normality Checks

Three methods can be used in order to check the normality of residuals.

- Histogram of residuals
- QQ Plot of residuals
- Shapiro Wilk Test

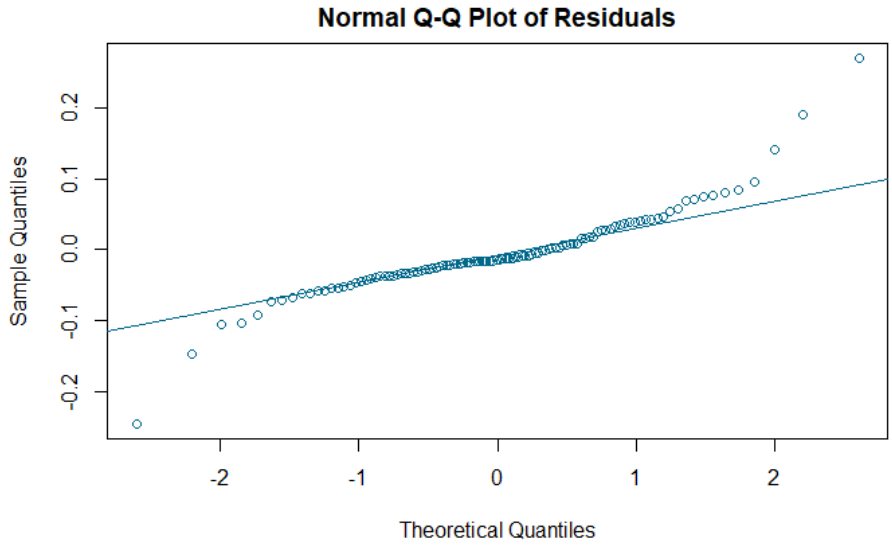
Histogram



Graph 14. Histogram of residuals

As it can be seen from histogram, residuals may be normally distributed. However, to be sure, common tests should be conducted.

Q-Q Plot



Graph 15. Normal Q-Q plot of residuals

The normal Q-Q Plot of residuals also indicates that residuals seems to follow normal distribution but common normality tests should be conducted. Moreover, there exists so many outliers.

Shapiro Test

Ho: Residuals are normally distributed.

Ha: Residuals are not normally distributed.

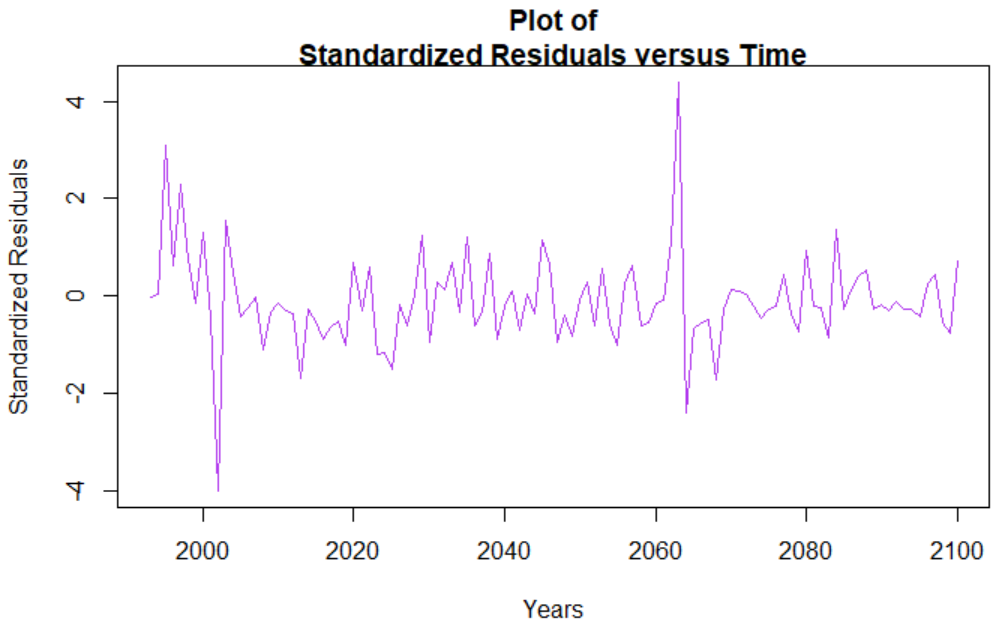
The results of Shapiro-Wilk Test Test:

W	**p-value**
0.88589	1.373e-07

Table12. Shapiro-Wilk test result

That is, the p- value of Shapiro-Wilk test is given as 1.373e-07 and it is smaller than significance level 0.05. Thus, the null hypothesis is rejected and it can be said that residuals are normally distributed.

As a result, by looking at the plot and tests, it can be said that residuals are not normally distributed.

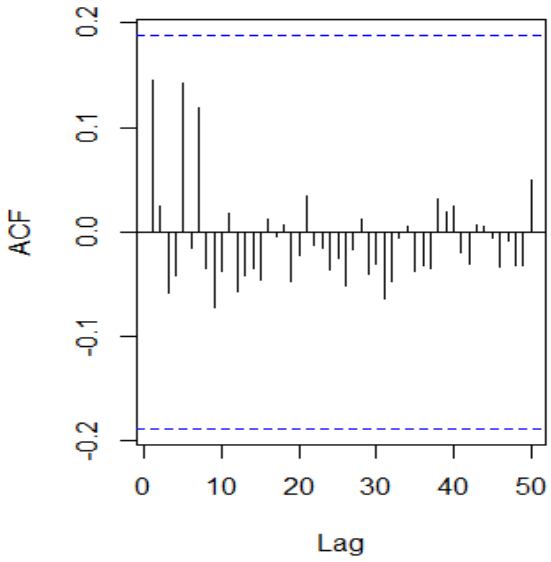


Graph16. Plot of standardized residuals versus time

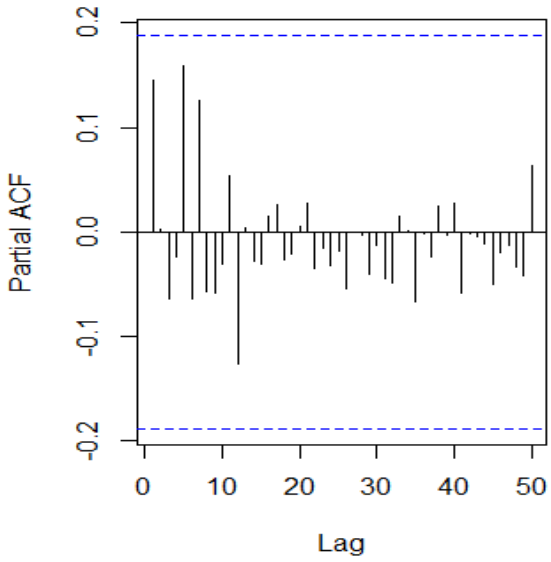
The standardized residuals vs time plot shows that there is stationarity around mean 0. However, there exists some outliers even though anomaly detection has been applied. After removing those several outlier points, this problem can be solved.

2.9.4. Heteroscedasticity Checks

Heteroscedasticity happens when the standard errors of a variable, monitored over a specific amount of time, are non-constant. In order to test heteroscedasticity , squared residuals are used.



Graph17. ACF Plot of Squared Residuals



Graph18. ACF Plot of Squared Residuals

Since all lags of squared residuals in ACF and PACF Plots are in white noise bands, there is no heteroscedasticity problem. Therefore, there exists homoscedasticity.

Furthermore, in order to be sure there is no heteroscedasticity problem, ARCH heteroscedasticity test for residuals is conducted.

The hypothesis for ARCH Test:

Ho: Residuals are homoscedastic.

Ha: Residuals are heteroscedastic.

After conducting the ARCH Test, p value equals to 0.368 which indicates that residuals are homoscedastic.

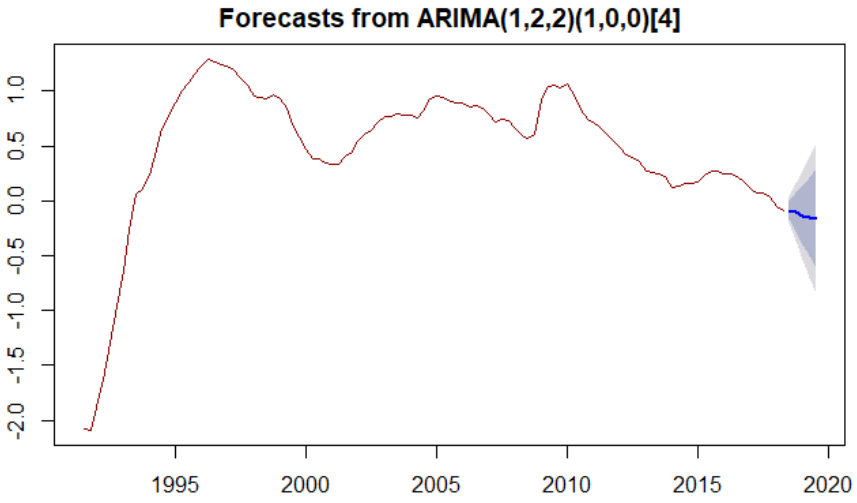
2.10 Forecasting

2.10.1 Minimum MSE Forecast

The prediction values of the Sarima model SARIMA(1,2,2)(1,0,0)4 is seen as the table.

Prediction Values				
-0.087406	-0.102034	-0.134421	-0.150393	-0.159496
Standard Errors				
0.061482	0.134709	0.207463	0.276880	0.352765

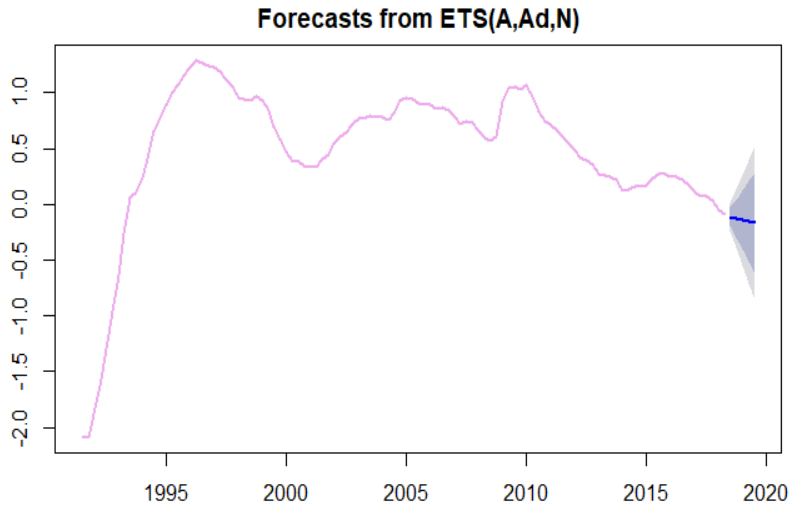
Table 13. Prediction values of ARIMA model



Graph 19. Plot of forecasts from ARIMA

By using forecast function, unemployment rate from 2018 to 2019 are forecasted.

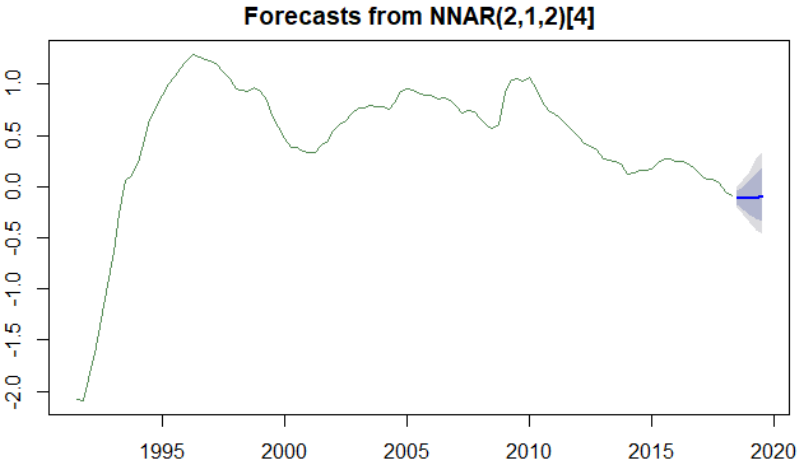
2.10.2 Exponential Smoothing for Deterministic Forecasting



Graph 20. Plot of forecasts from ETS model

ETS gives the additive error, additive trend and no seasonality.

2.10.3 Forecasting with Neural Network



Graph 21. Plot of forecasts from NNAR model

2.10.4 Forecast Accuracy Measures

In order to determine which model gives better forecast for the future values of the time series, forecast accuracy is used. By using MASE values for test set, we choose the best model.

	ARIMA	ETS	NNETAR
MASE	4.223554	4.27837	4.1855
MAPE	113.0566	114.5457	112.1086

Table14. MASE and MAPE values for the forecasts

According to MASE and MAPE values, NNETAR is chosen as the best technique.

2.11 Back Transforming

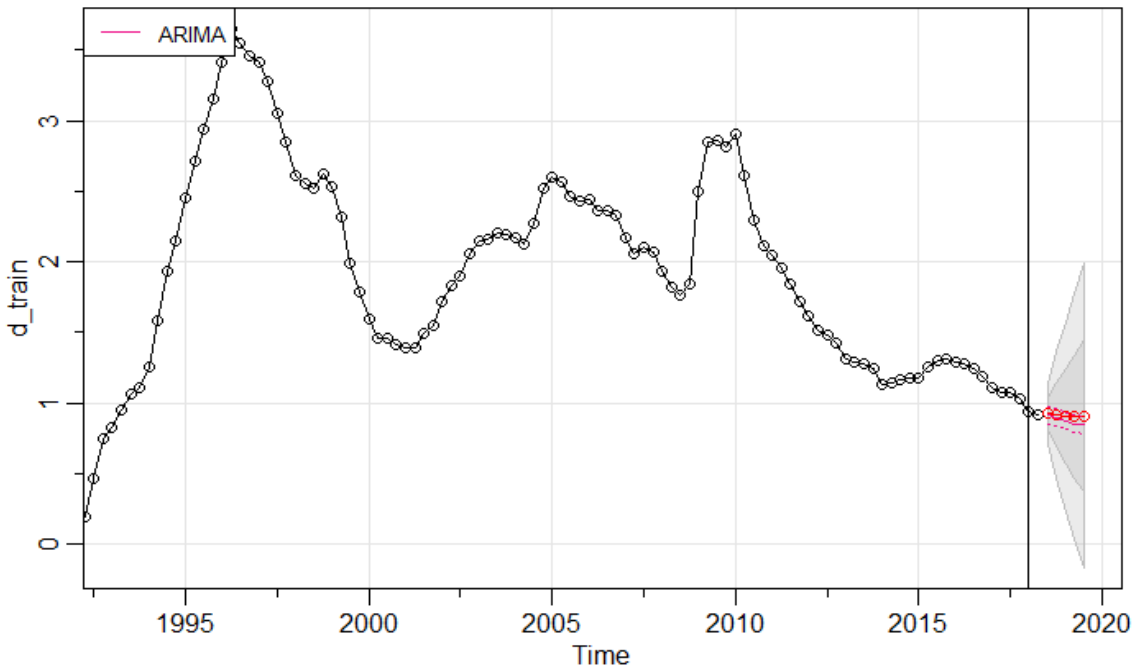
Backtransformed predictions	0.893496	0.8780222	0.874502	0.879287	0.890864
Test data	0.947664	0.9360405	0.953393	0.997214	0.998183

Table 15. Table of back transforming

The difference between back transformed predictions and test data is not so far from each other. That means, the NETAR model represents the data well.

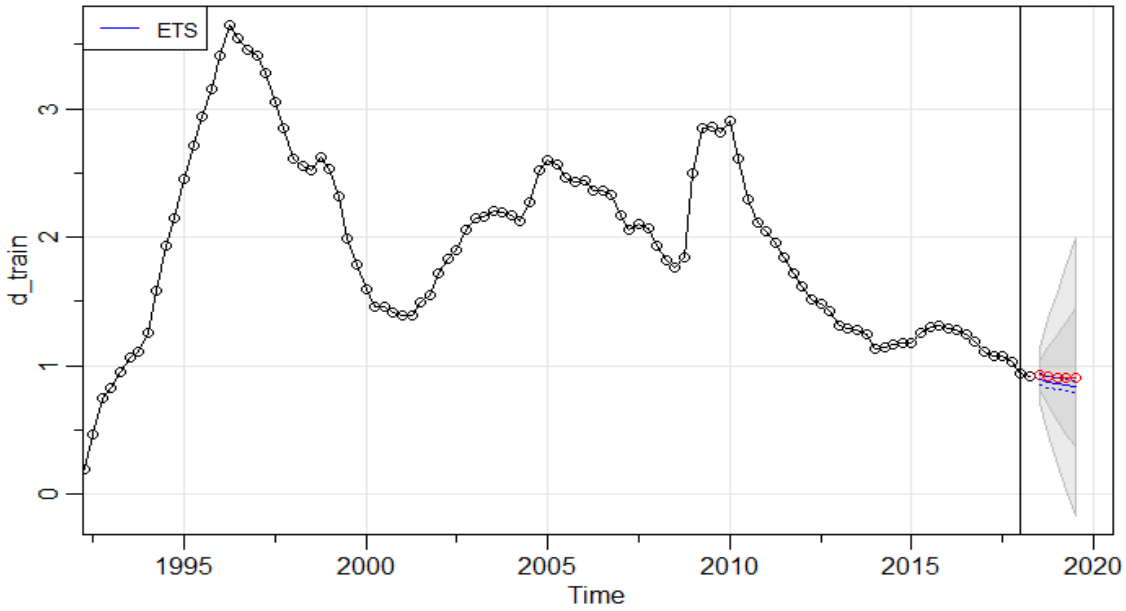
2.12 Plots of the Forecast Values

ARIMA



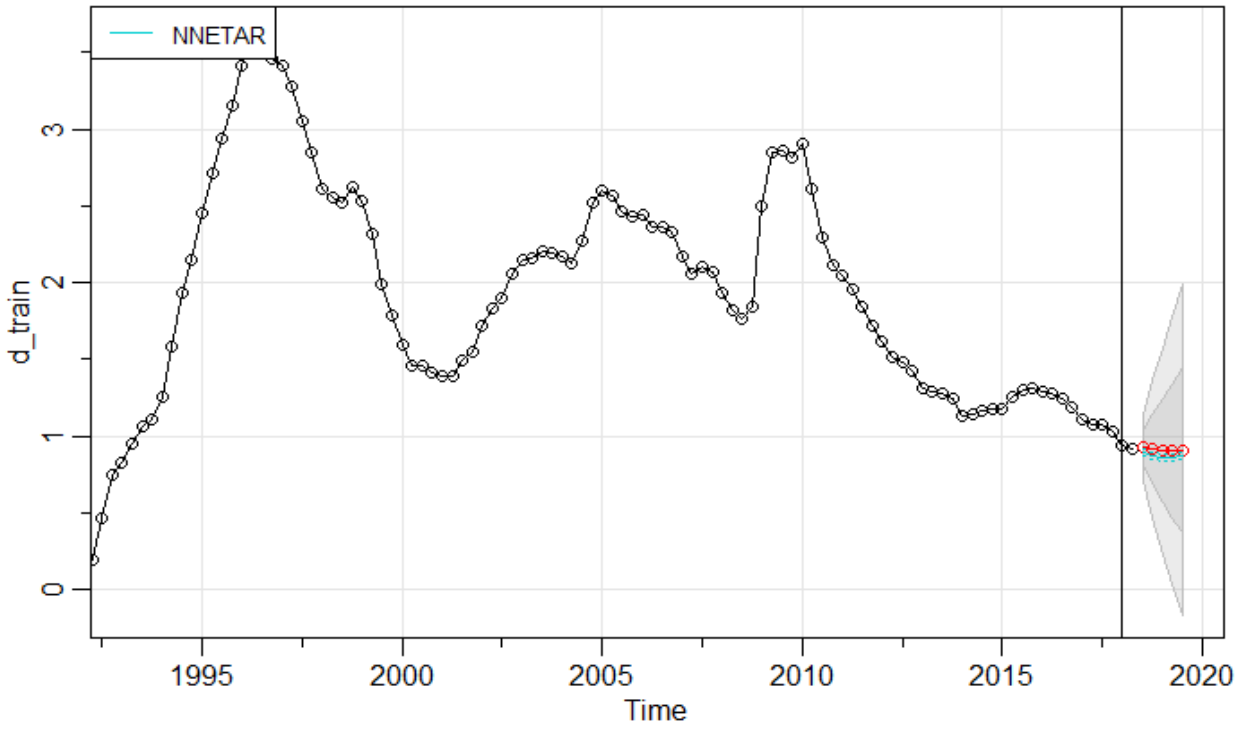
Graph 21. Plot of the data values and forecasts from ARIMA

ETS



Graph 22. Plot of the data values and forecasts from ETS

NNETAR



Graph 23. Plot of the data values and forecasts from NNETAR

All the prediction techniques are in the confidence region. However, NETAR seems to follow the structure in a best way.

3. Conclusion

Unemployment Rate for the Russian Federation data for 1991&2019 were used for analysis. The last five observations were eliminated as test data for the prediction process.

After examining the data, Box Cox transformation was applied for a better and proper analysis. After making anomally detection, a number of tests were carried out to check seasonality and stationarity. After two regular differencing, our data became stationary. Then, two models suggested by us and the R programming and these models were examined. Then, the appropriate model was chosen. Diagnostic check was done on that model. Finally, prediction values were obtained with three different forecasting methods and NNETAR was chosen as the best prediction method.

Since it was a newly established country in 1991, there has been a strict increase in unemployment until approximately 1996-1997. Although there has been a drop in unemployment rates after those years, there are fluctuations over time because of political reasons.

For further studies, variance models can be modelled by (G)ARCH method to get rid of heteroscedasticity problem.

REFERENCES

https://fred.stlouisfed.org/series/LMUNRRTTRUQ156S