Macroeconomic Forecasting of French Economy using Machine Learning Approach- Configuration Manual

M.Sc. Research Project
Data Analytics

Manu Mishra
Student ID: x18127347

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Manu Mishra
Student ID:	x18127347
Programme:	M.Sc.-Data Analytics
Year:	2019
Module:	Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	12/12/2019
Project Title:	Macroeconomic Forecasting of French Economy using Machine Learning Approach-Configuration Manual
Word Count:	XXX
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	12th Deceember 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

1 Introduction **2**

2 Specification of Experimental Environment **2**
2.1 System Specifications 2
2.2 Technical Specifications 3
2.3 Data Sources 4

3 Related Work **5**

4 Model Implementation and Evaluation **7**
4.1 Vector Autoregressive Model 7
4.2 Long Short-Term Memory 18

Configuration Manual

Manu Mishra
x18127347

1 Introduction

Configuration Manual provides an extensive description of the environment setup for the Research Project. This includes the system configuration, details of programming language used, description of libraries and packages used.

This document also provides discussion of results of different tests being used in the research. Some relevant information, graphs and output metrics, which were a part of implementation and the relevant work, are also included in this report. The reason for including such information in the Configuration Manual is that these details were not included in the main Research Report but are worth considering for relevance and discussion.

Therefore, these details are a part of this manual. The discussions in the Research Report which have relevant information mentioned in the Configuration Manual have been properly referenced for the ease of understanding.

2 Specification of Experimental Environment

2.1 System Specifications

The research has been carried out on the local machine which has the following configuration:

- Operating System: Windows 10, 64 bit
- Processor: Intel® Core™ i5, 8th Generation, 1.80 GHz
- Installed Memory (RAM): 12 GB

2.2 Technical Specifications

Following software environments and programming languages are used in the research:

1. R [1]

- Version 3.5.2
- RStudio 1.1.463 [2] is used as an Integrated Development Environment (IDE) for R.
- The following packages in R are used in the research project:
  - vars [3]
  - urca [4]
  - tseries [5]
  - forecast [6]

2. Python [7]

- Version 3.7.4
- Jupyter Notebook 6.0.2 on Anaconda 1.9.7 is used as a platform for Python.
- The following packages in Python are used in research project:
  - Pandas 0.25.3 [8]
  - Numpy 1.17.3 [9]
  - Scikit-learn 0.21.3 [10]
  - Keras 1.0.8 [11]
  - Matplotlib 3.1.1 [12]
  - Tensorflow 1.13.1 [13]

[1] https://www.r-project.org/
[2] https://rstudio.com/
[3] https://cran.r-project.org/web/packages/vars/vars.pdf
[4] https://cran.r-project.org/web/packages/urca/urca.pdf
[5] https://cran.r-project.org/web/packages/tseries/tseries.pdf
[6] https://cran.r-project.org/web/packages/forecast/forecast.pdf
[7] https://www.python.org
[8] https://pandas.pydata.org/
[9] https://numpy.org/
[10] https://scikit-learn.org/stable/index.html
[11] https://keras.io/
[12] https://matplotlib.org/
[13] https://www.tensorflow.org/install/

2.3 Data Sources

1. Unemployment Rate of France[14] : Monthly time series data for unemployment rate for France is downloaded from Statistical Data Warehouse - European Central Bank for 36 years, starting from January 1983 to December 2018. The series level information is given in below figure:

Series Level Information

Series Key	STS.M.FR.S.UNEH.RTT000.4.000
Title	Unemployment rate
Title Complement	France - Standardised unemployment, Rate, Total (all ages), Total (male & female); unspecified; Eurostat; Seasonally adjusted, not working day adjusted, percentage of civilian workforce
Unit	Percent
Dataset	STS : Short-Term Statistics Metadata page (Series and Dataset Level Information)
+ More attributes	

Figure 1: Data Source Information-Unemployment Rate

2. Inflation Rate of France[15] : Monthly time series data for inflation rate for France is downloaded from Statistical Data Warehouse - European Central Bank for 23 years, from January 1996 to December 2018. The series level information is provided below:

Series Level Information

Series Key	ICP.M.FR.N.000000.4.INX
Title	HICP - Overall index
Title Complement	France - HICP - Overall index, Monthly Index, Eurostat, Neither seasonally nor working day adjusted
Unit	2015 = 100
Dataset	ICP : Indices of Consumer prices Metadata page (Series and Dataset Level Information)
+ More attributes	

Figure 2: Data Source Information-Inflation Rate

[14]https://sdw.ecb.europa.eu/quickview.do?SERIES$_KEY = 132.STS.M.FR.S.UNEH.RTT000.4.000$

[15]http://sdw.ecb.europa.eu/quickview.do;jsessionid=4D1C7FE561D95E47067811B20DF8CD8B?SERIES$_KEY = 122.ICP.M.FR.N.000000.4.INX$

3 Related Work

Unemployment Rate in France and Germany:

France has comparatively higher unemployment in youth, as compared to Germany(Guillaume et al.; 2019). This is represented in figure 1.

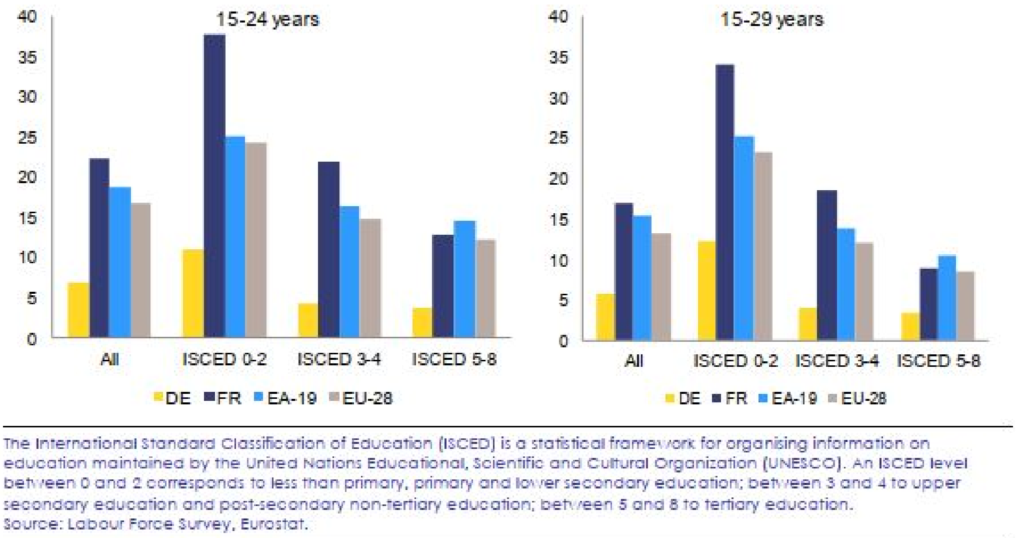

Figure 3: Youth Unemployment Rate by age and education(Guillaume et al.; 2019)

Relevance of Phillips Curve in present times:

It can be seen that there has been a symbolic decrease in the slope of Phillips Curve during 1980 to mid-1990 and since then, has remained stable around 0.5. Referring to the coefficient of inflation in figure 3, it maintained a value less than 1 during the mid-1990s and then began to descend to limit around the value of zero during the current times. Therefore, it can be said that there has been a shift from influence of unemployment rate over variations in inflation.

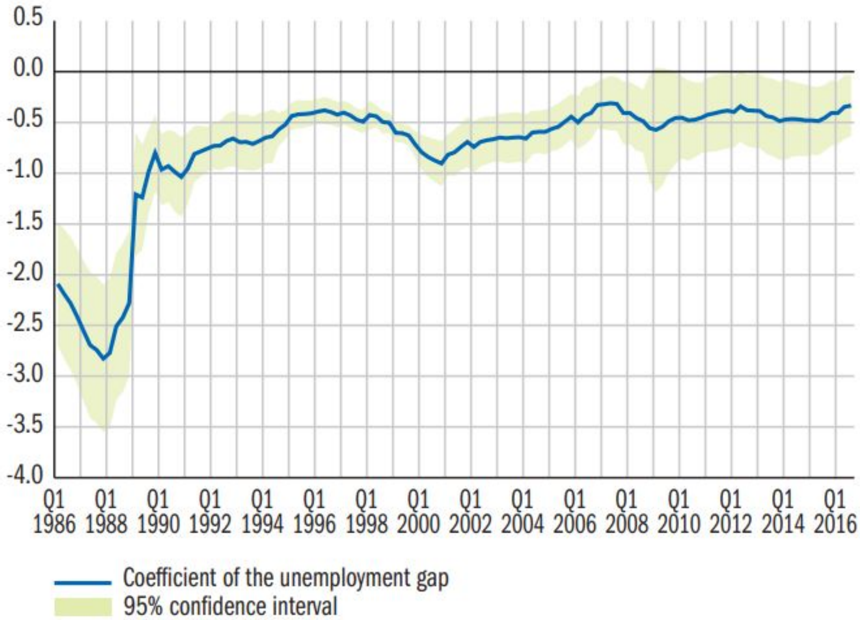

Figure 4: Estimated coefficient of the unemployment gap in G7 countries(Berson et al.; 2018)

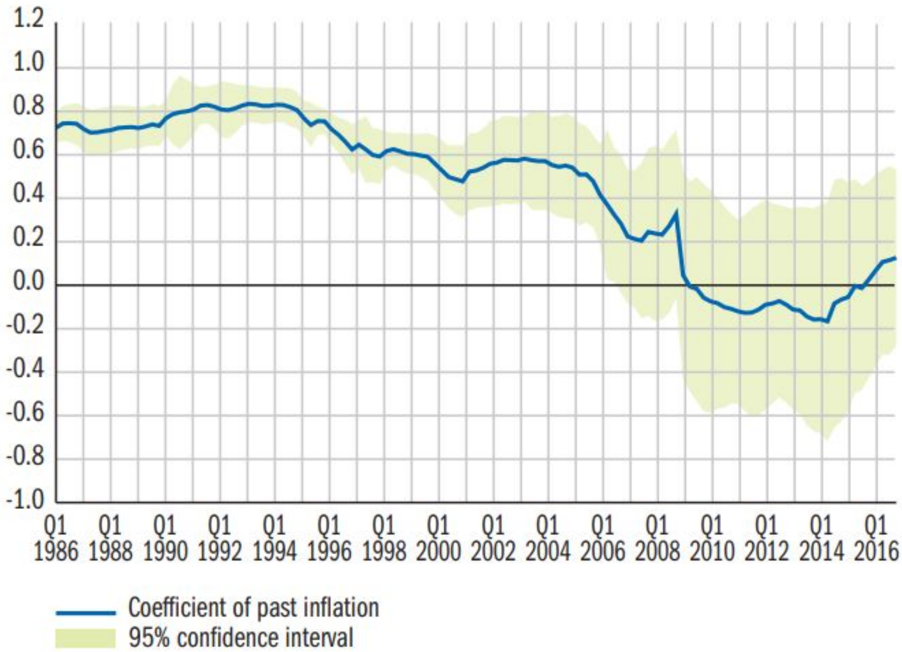

Figure 5: Estimated coefficients of past inflation in G7 countries(Berson et al.; 2018)

4 Model Implementation and Evaluation

4.1 Vector Autoregressive Model

- **Converting Time Series into time-series object**

  The time series of unemployment rate is converted into time-series object to ease the implement time-series analysis.

````
Converting the time series into time-series (ts) object
```{r}
UNRATE <- ts(france$UNRATE,
             start = c(1996, 1),
             end = c(2018, 12),
             frequency = 12)

autoplot(UNRATE)+
  ggtitle("Unemployment rate for France")+
  xlab("Year")+
  ylab("Percentage")
```
````

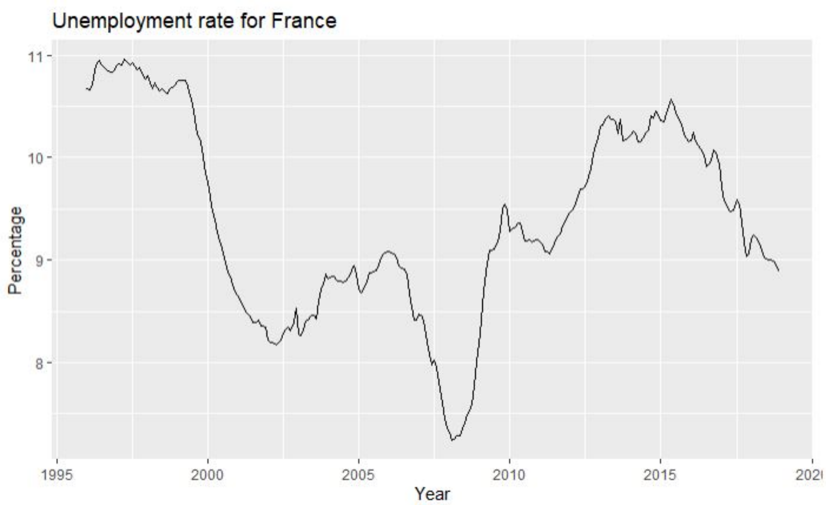

Figure 6: Converting unemployment rate to time-series object

The time series of inflation rate is converted into time-series object to ease the implement time-series analysis.

~~~
HICP <- ts(france$HICP,
           start = c(1996, 1),
           end = c(2018, 12),
           frequency = 12)

autoplot(HICP)+
  ggtitle("Inflation rate for France")+
  xlab("Year")+
  ylab("Percentage")

```
~~~

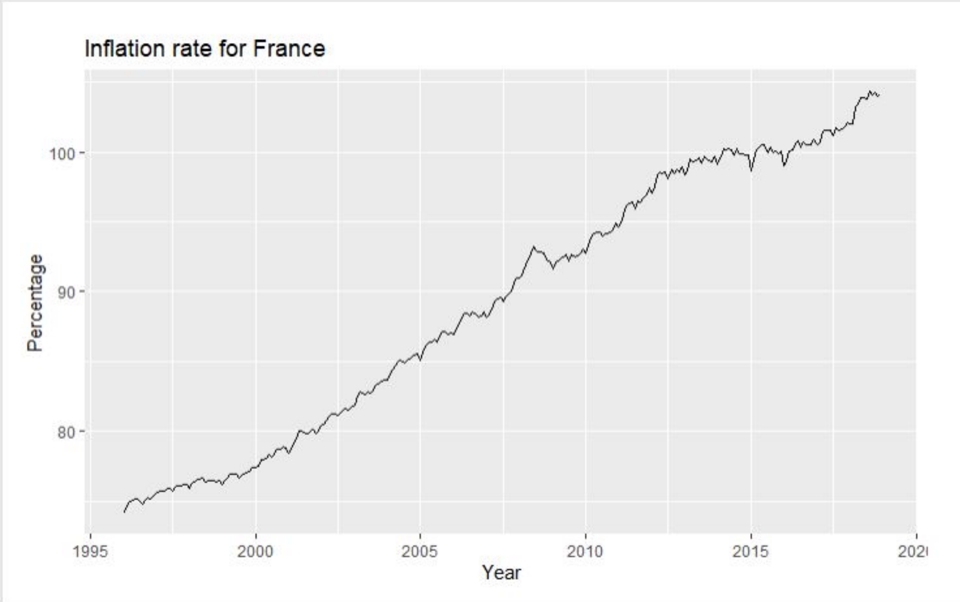

Figure 7: Converting inflation rate to time-series object

- **Check for seasonality**

  As per figure 8 and 9, both the seasonal plots confirms the absence of seasonal component in unemployment rate series.

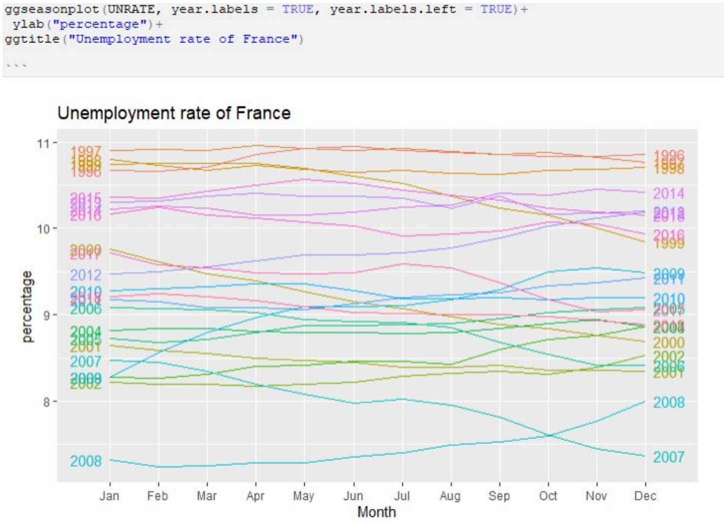

Figure 8: Seasonal plot for unemployment rate

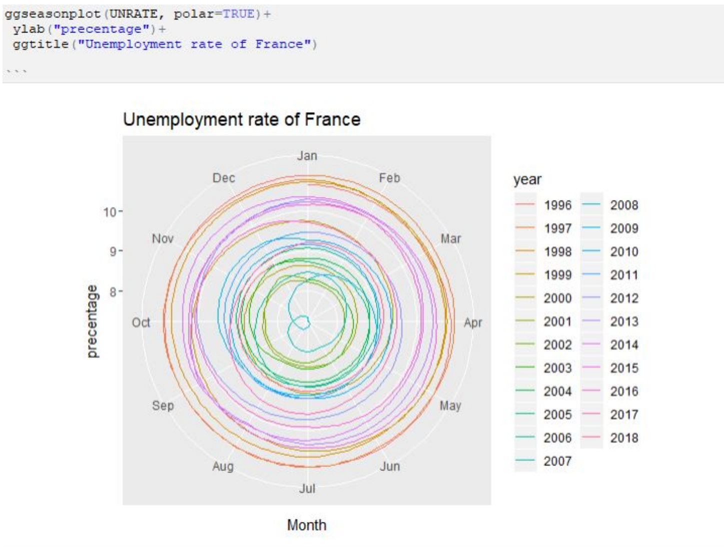

Figure 9: Seasonal plot for unemployment rate

Both the seasonal plots for series of inflation rate confirms the absence of seasonal component.

```
ggseasonplot(HICP, year.labels = TRUE, year.labels.left = TRUE)+
 ylab("percentage")+
ggtitle("Inflation rate of France")
```

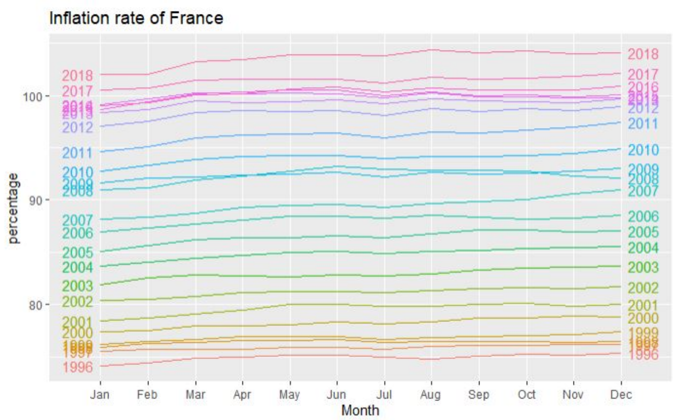

Figure 10: Seasonal plot for inflation rate

```
ggseasonplot(HICP, polar=TRUE)+
 ylab("precentage")+
 ggtitle("Inflation rate of France")
```

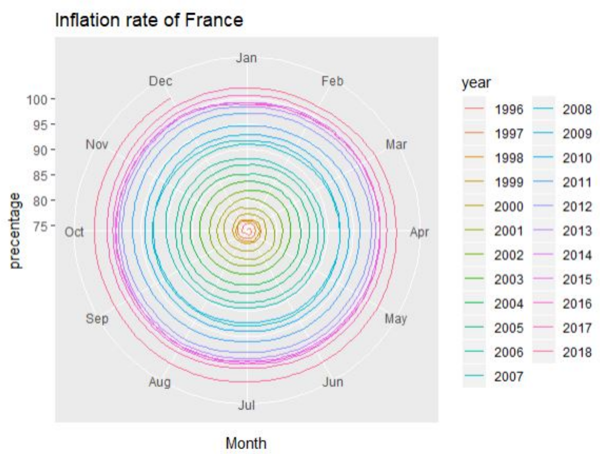

Figure 11: Seasonal plot for inflation rate

- **Check for lag length**: Lag length is selected as 2 because it has the minimum value for the information criteria.

```
VARselect(france, lag.max = 5, type = "both")

VARselect(france, lag.max = 5, type="both")$selection

```
```

```
$selection
AIC(n)  HQ(n)  SC(n) FPE(n) 
     2      2      2      2 

$criteria
                   1             2             3
AIC(n) -7.5047159410 -7.8349043647 -7.8206400909
HQ(n)  -7.4620210399 -7.7708620131 -7.7352502887
SC(n)  -7.3983803301 -7.6754009483 -7.6079688690
FPE(n)  0.0005504846  0.0003956859  0.0004013785
                   4             5
AIC(n) -7.8044772733 -7.7944225079
HQ(n)  -7.6977400206 -7.6663378047
SC(n)  -7.5386382459 -7.4754156751
FPE(n)  0.0004079319  0.0004120744

AIC(n)  HQ(n)  SC(n) FPE(n) 
     2      2      2      2 
```

Figure 12: Lag length selection
```

- **Check for stationarity**: Augmented Dickey Fuller (ADF) test (Dickey and Fuller; 1981)is used to check stationarity of both the time series. The hypothesis for the test are:
  - Null Hypothesis : There is presence of a unit root.
  - Alternate Hypothesis : There is no unit root and staionarity exists.

  p-value for unemployment rate is 0.2641, which is greater than significance level of 0.005. Therefore the series is non-stationary.

  Similarly, p-value for inflation rate is 0.6085, which is greater than the significance level of 0.05. This confirms that the series is non-stationary.

````
```{r}
adf.test(france[,"UNRATE"])
```
````

```
	Augmented Dickey-Fuller Test

data:  france[, "UNRATE"]
Dickey-Fuller = -2.7406, Lag order = 6, p-value =
0.2641
alternative hypothesis: stationary
```

The p-value for UNRATE is 0.2641, which is greater than the significance level of 5% or 0.005

````
```{r}
adf.test(france[,"HICP"])
```
````

```
	Augmented Dickey-Fuller Test

data:  france[, "HICP"]
Dickey-Fuller = -1.9231, Lag order = 6, p-value =
0.6085
alternative hypothesis: stationary
```

The p-value for HICP is 0.6085, which is greater than the significance level of 5% or 0.05

Figure 13: ADF test on both the time series

- **Creating Differenced Series**:

```{r}
adf.test(diff(france[,"UNRATE"]))
```

```
        Augmented Dickey-Fuller Test

data:  diff(france[, "UNRATE"])
Dickey-Fuller = -3.555, Lag order = 6, p-value =
0.03789
alternative hypothesis: stationary
```

The p-value is 0.03789, which is less than 0.05. The series is now stationary.

```{r}
adf.test(diff(france[,"HICP"]))
```

```
p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  diff(france[, "HICP"])
Dickey-Fuller = -6.6377, Lag order = 6, p-value =
0.01
alternative hypothesis: stationary
```

The p-value is less than 0.01, which is less than significance level of 0.05.

Figure 14: ADF test on both the differenced time series

The p-value of unemployment rate is 0.03789, which is less than 0.05, which confirms that the series is stationary after first differencing. For inflation rate, p-value is less than 0.01, which is less than significance level of 0.05. Therefore, the inflation rate is stationary after first differencing.

The plots for differenced series for unemployment rate and inflation rate is shown below:

```
unrate_diff = diff(france[,"UNRATE"])
hicp_diff = diff(france[,"HICP"])

autoplot(unrate_diff)+
  ggtitle("Unemployment rate for France - First order of differencing")+
  xlab("Year")+
  ylab("Percentage")
```

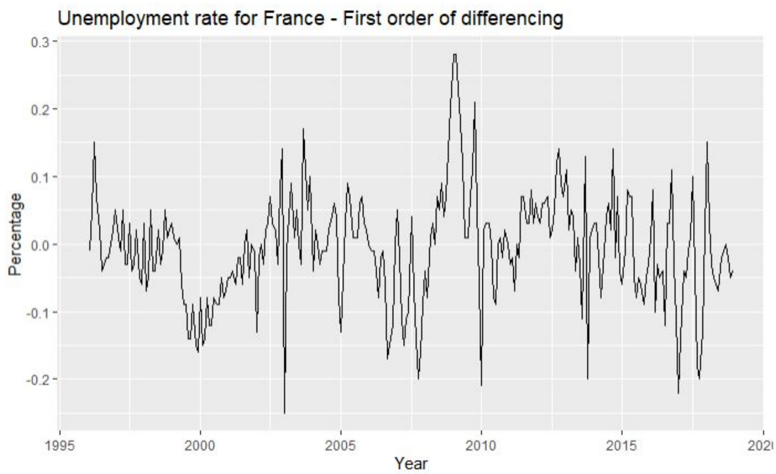


Figure 15: First Differencing - Unemployment Rate

```
autoplot(hicp_diff)+
  ggtitle("Inflation rate for France - First order of differencing")+
  xlab("Year")+
  ylab("Percentage")
```

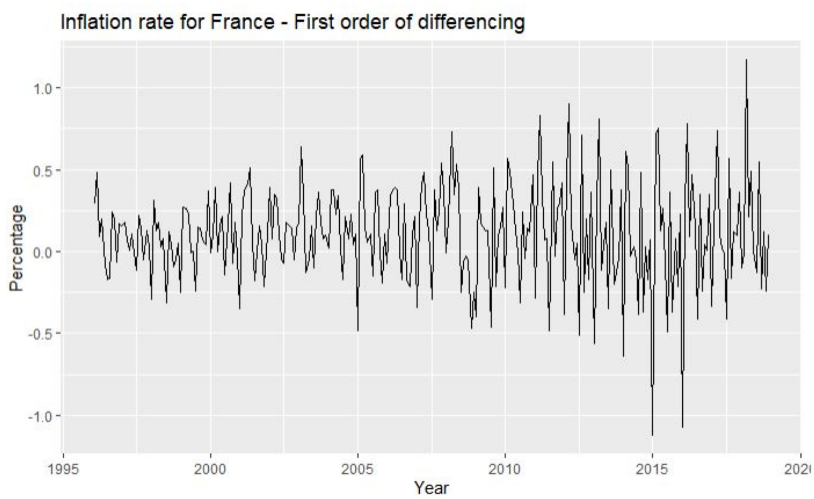


Figure 16: First Differencing - Inflation Rate

- **Lag length selection for differenced series**:

  Lag length is selected as 1 because it has the minimum value for the information criteria.

```
VARselect(france_diff, lag.max = 5, type = "both")
```

```
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     1      1      1      1

$criteria
                   1             2             3
AIC(n) -7.7766374834 -7.7578368044 -7.7308933842
HQ(n)  -7.7338235714 -7.6936159364 -7.6452655602
SC(n)  -7.6700175735 -7.5979069395 -7.5176535644
FPE(n)  0.0004194219  0.0004273863  0.0004390669
                   4             5
AIC(n) -7.7149946993 -7.763296149
HQ(n)  -7.6079599193 -7.634854413
SC(n)  -7.4484449246 -7.443436419
FPE(n)  0.0004461181  0.000425103
```

Figure 17: Lag length selection on differenced series

- **Check for Cointegration**:

  For first hypothesis, test statistic is less than value at significance level of 5%. For second hypothesis, test statistic is less than value at significance level of 5%. This confirms the absence of cointegration.

```
summary(ca.jo(france, type = "trace", ecdet = "trend",
              K=2, spec = "transitory" ))
```

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend in cointegration

Eigenvalues (lambda):
[1] 0.05183064 0.02344972 0.00000000

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  6.50 10.49 12.25 16.26
r = 0  | 21.08 22.76 25.32 30.45

Eigenvectors, normalised to first column:
(These are the cointegration relations)

           UNRATE.l1    HICP.l1  trend.l1
UNRATE.l1  1.0000000 1.00000000 1.0000000
HICP.l1   -2.0544242 0.31710810 -0.9364279
trend.l1   0.2379305 -0.03698642 0.2378768

Weights W:
(This is the loading matrix)

            UNRATE.l1      HICP.l1      trend.l1
UNRATE.d -0.006332252 -0.001552822  9.410437e-18
HICP.d    0.005453655 -0.043770407 -2.377358e-17
```

Figure 18: Output for Cointegration Test

- **Check for Causality**: The output of Granger Causality Test (Granger and Newbold; 1974) on both the variables shows that both variables do not explain the variation in each other.

```
grangertest(hicp_diff~unrate_diff, order=2, data=france_diff)
```
```
Granger causality test

Model 1: hicp_diff ~ Lags(hicp_diff, 1:2) + Lags(unrate_diff, 1:2)
Model 2: hicp_diff ~ Lags(hicp_diff, 1:2)
  Res.Df Df      F Pr(>F)
1    268
2    270 -2 1.128 0.3252
```

Figure 19: Output for Causality Test of unemployment granger causes inflation

```
grangertest(unrate_diff~hicp_diff, order=2, data=france_diff)
```
```
Granger causality test

Model 1: unrate_diff ~ Lags(unrate_diff, 1:2) + Lags(hicp_diff, 1:2)
Model 2: unrate_diff ~ Lags(unrate_diff, 1:2)
  Res.Df Df      F Pr(>F)
1    268
2    270 -2 0.4574 0.6334
```

Figure 20: Output for Causality Test of inflation granger causes unemployment

- **Check for Autocorrelation in both the time series**: Check for individual autocorrelation tests on both the differenced series shows weak correlations.

```
acf(unrate_diff, lag.max=5, plot=FALSE)
```
```

Autocorrelations of series ‘unrate_diff’, by lag

0.0000 0.0833 0.1667 0.2500 0.3333 0.4167
 1.000  0.592  0.393  0.256  0.203  0.285
```

Figure 21: Autocorrelation Test on differenced unemployment rate

```
acf(hicp_diff, lag.max=5, plot=FALSE)
```
```

Autocorrelations of series ‘hicp_diff’, by lag

0.0000 0.0833 0.1667 0.2500 0.3333 0.4167
 1.000 -0.093 -0.016 -0.002 -0.081 -0.142
```

Figure 22: Autocorrelation Test on differenced inflation rate

## 4.2 Long Short-Term Memory

- **LSTM forecast of Time Series - Normal Split of Train and Test**: The below given figures shows the forecast plot by LSTM network on normal data split into train and test. Figure 23 for unemployment rate and figure 24 for inflation rate represents values of the variables on Y-axis and the length of the series on X-axis. In the forecast plot, the blue curve represents the actual values while the red curve represents the forecasted values.

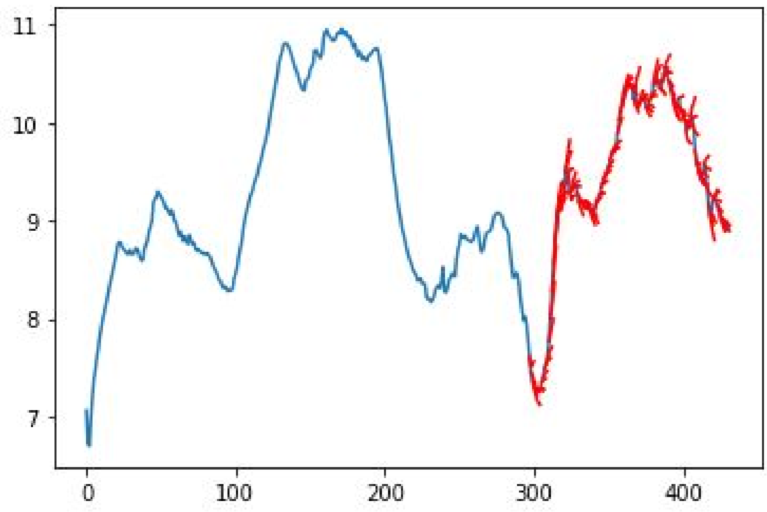


Figure 23: Unemployment Rate forecast plot for LSTM - Normal Split

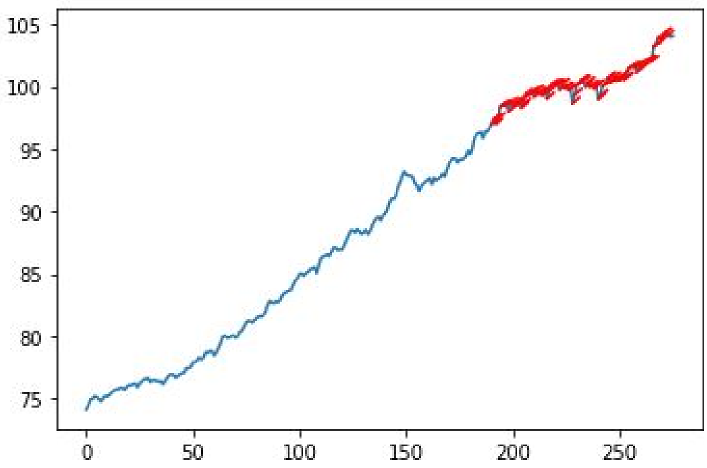


Figure 24: Inflation Rate forecast plot for LSTM - Normal Split

- **LSTM forecast of Time Series - Cross Validation**: The below given figures (Figure 25 for Unemployment Rate and Figure 26 for Inflation Rate)shows the forecast plot by LSTM network on the basis of cross-validation in time-series, where the blue curve represents actual values and red curve represents forecast values. Y-axis shows the value of variables and X-axis represents the length of series. As per (Hyndman; 2016), the train set data always appear before the train set data, in terms of their appearance in the original series.

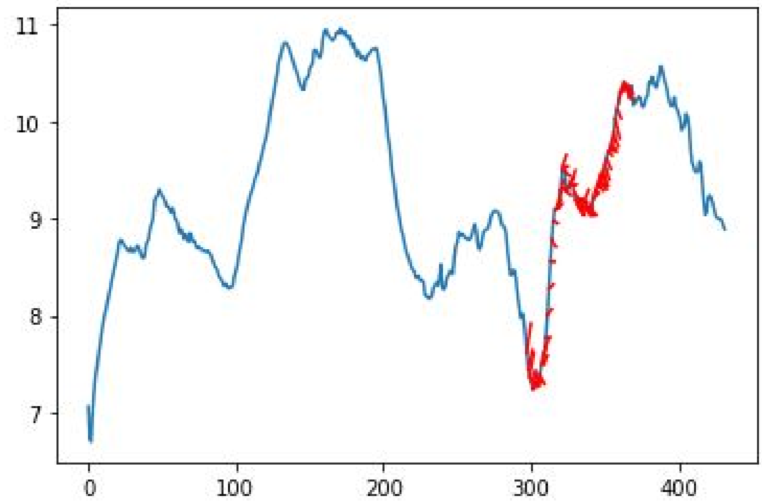


Figure 25: Unemployment Rate forecast plot for LSTM - Cross Validation

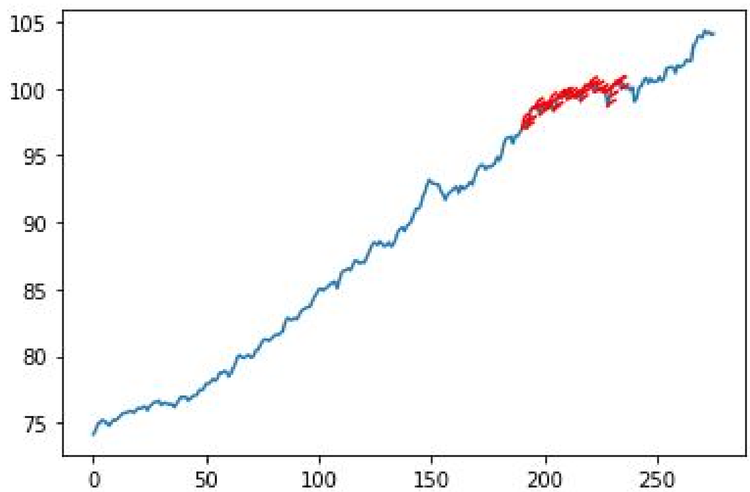


Figure 26: Inflation Rate forecast plot for LSTM - Cross Validation

# References

Berson, C., Charsonville, L. d., Diev, P., Faubert, V., Ferrara, L., Nefussi, S. G., Kalantzis, Y., Lalliard, A., Matheron, J., Mogliani, M. and et al. (2018). *Does the Phillips curve still exist?*, Vol. 56.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* **49**(4): 1057.

Granger, C. J. and Newbold, P. (1974). Spurious regression in econometrics, *Journal of Econometrics* **2**: 111–120.

Guillaume, C., de Castro Fernandez Francisco, Duran, L. J., Granelli, L., Hallet, M., Jaubertie, A., Rodriguez, C. M., Ognyanova, D., Balazs, P., Tsalinski, T. and et al. (2019). *Cruising at different speeds: similarities and divergences between the German and the French economies.*

Hyndman, R. J. (2016). Cross-validation for time series.
**URL:** *https://robjhyndman.com/hyndsight/tscv/*