Unemployment Rate

Business, Economics and Financial Data

Roya Ghamari, Shabnam Sattar

January 2024

Introduction

Unemployment in Europe has been a persistent and multifaceted challenge, influenced by a combination of economic, social, and political factors. Despite the region's diverse economies and policy frameworks, the issue of unemployment has manifested differently across European countries, with some nations experiencing higher rates than others.

In the present work, we will analyze the unemployment rates in three European countries, each situated in distinct regions: North, Middle, and South.

Dataset

Main dataset

- **Unemployment by sex and age monthly data**
- Compiling agency: The statistical office of the European Union, Eurostat

Complementary dataset

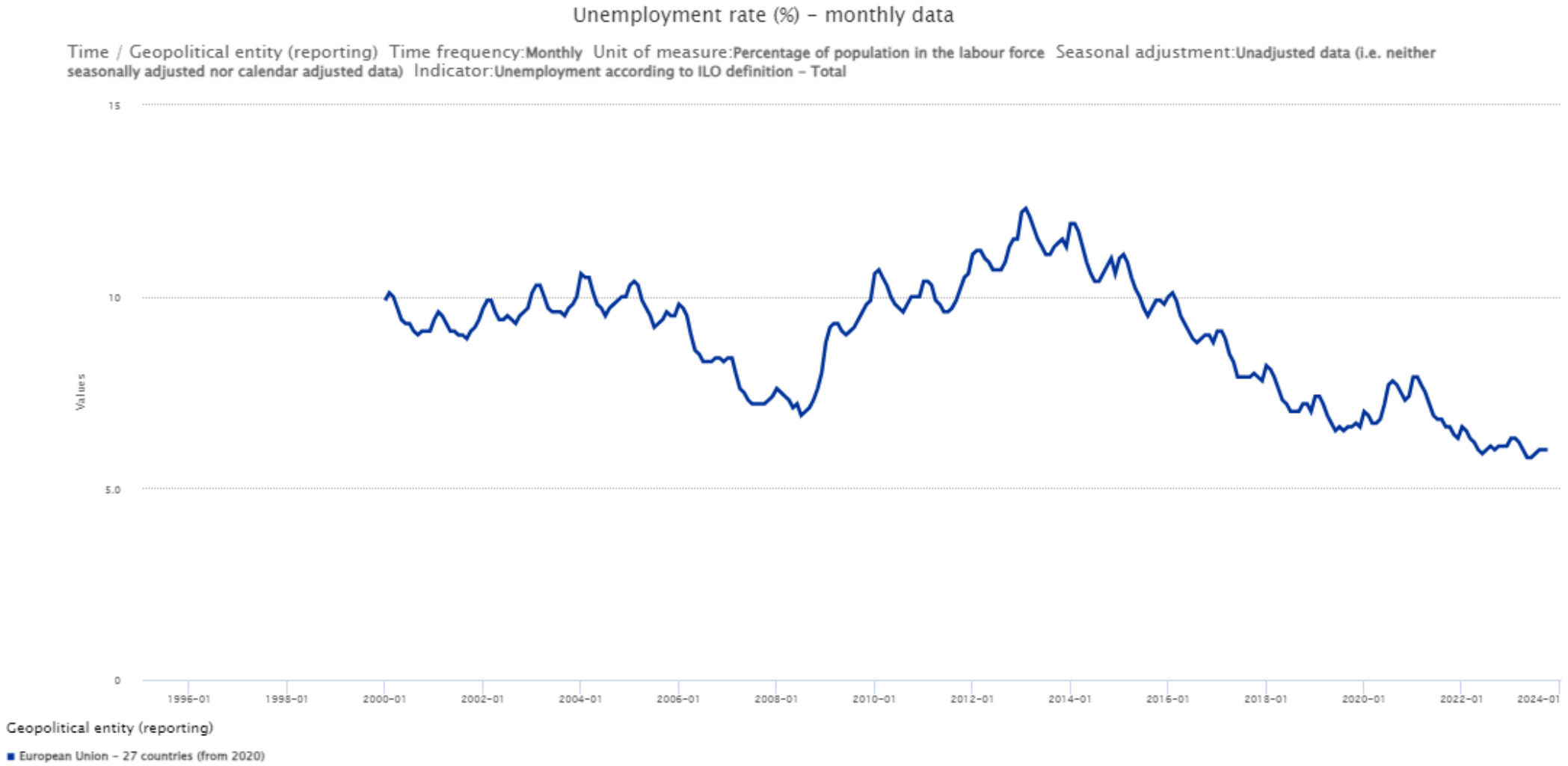
- **Gross Domestic Product (GDP) for Greece**
- **Consumer Price Index: All Items: Total for Greece**
- Compiling agency: Economic Research Division Federal Reserve Bank of St. Louis

- **Balance of payments by Greece**
- Compiling agency: The statistical office of the European Union, Eurostat

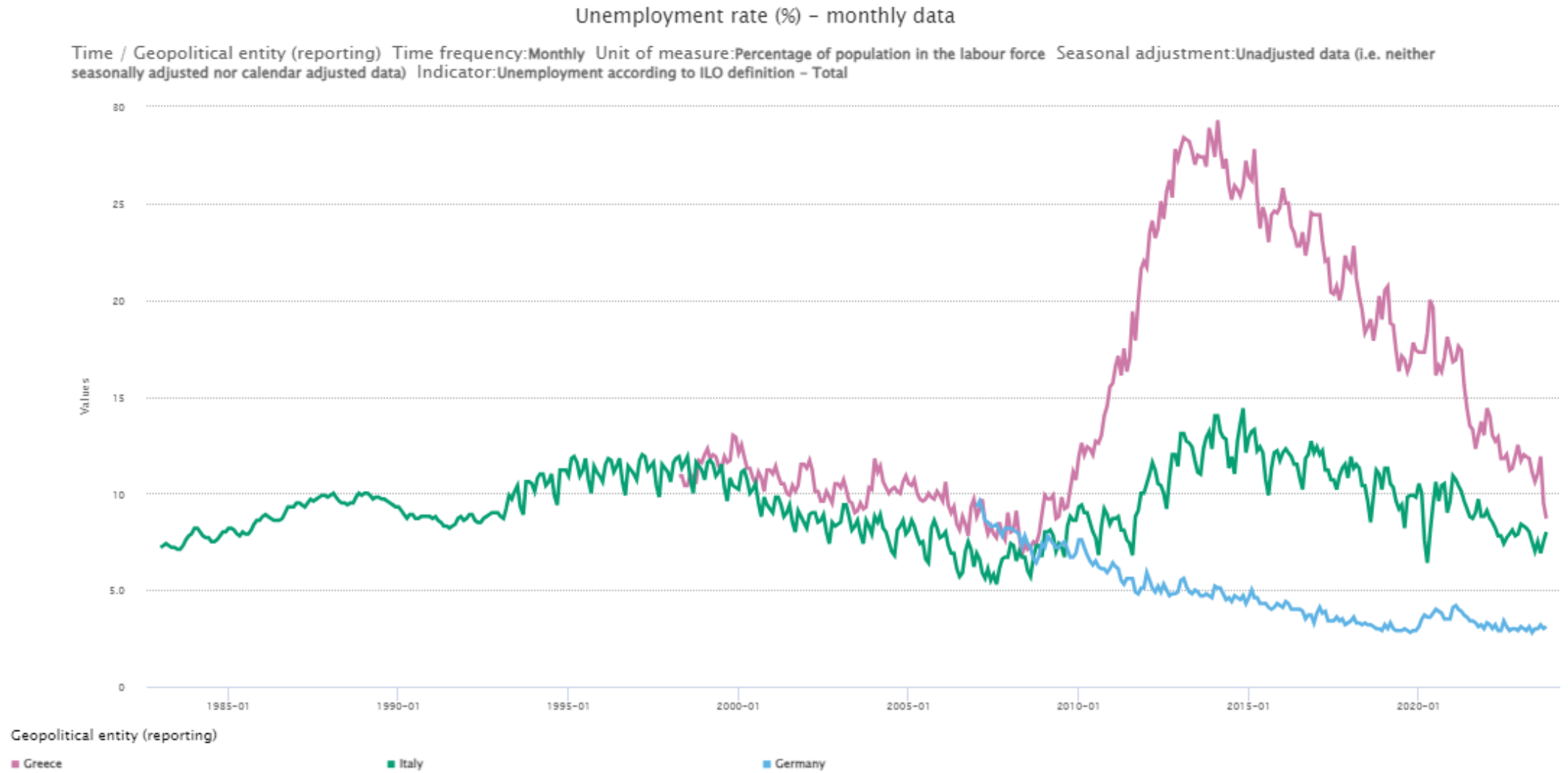
People between the ages of 15 and 74 who were not working in the past week, actively looked for a job in the last four weeks and are ready to start working either immediately or within two weeks are considered **unemployed**.

The unemployment rate is the number of people unemployed as a percentage of the labor force. The labor force is the total number of people employed and unemployed.

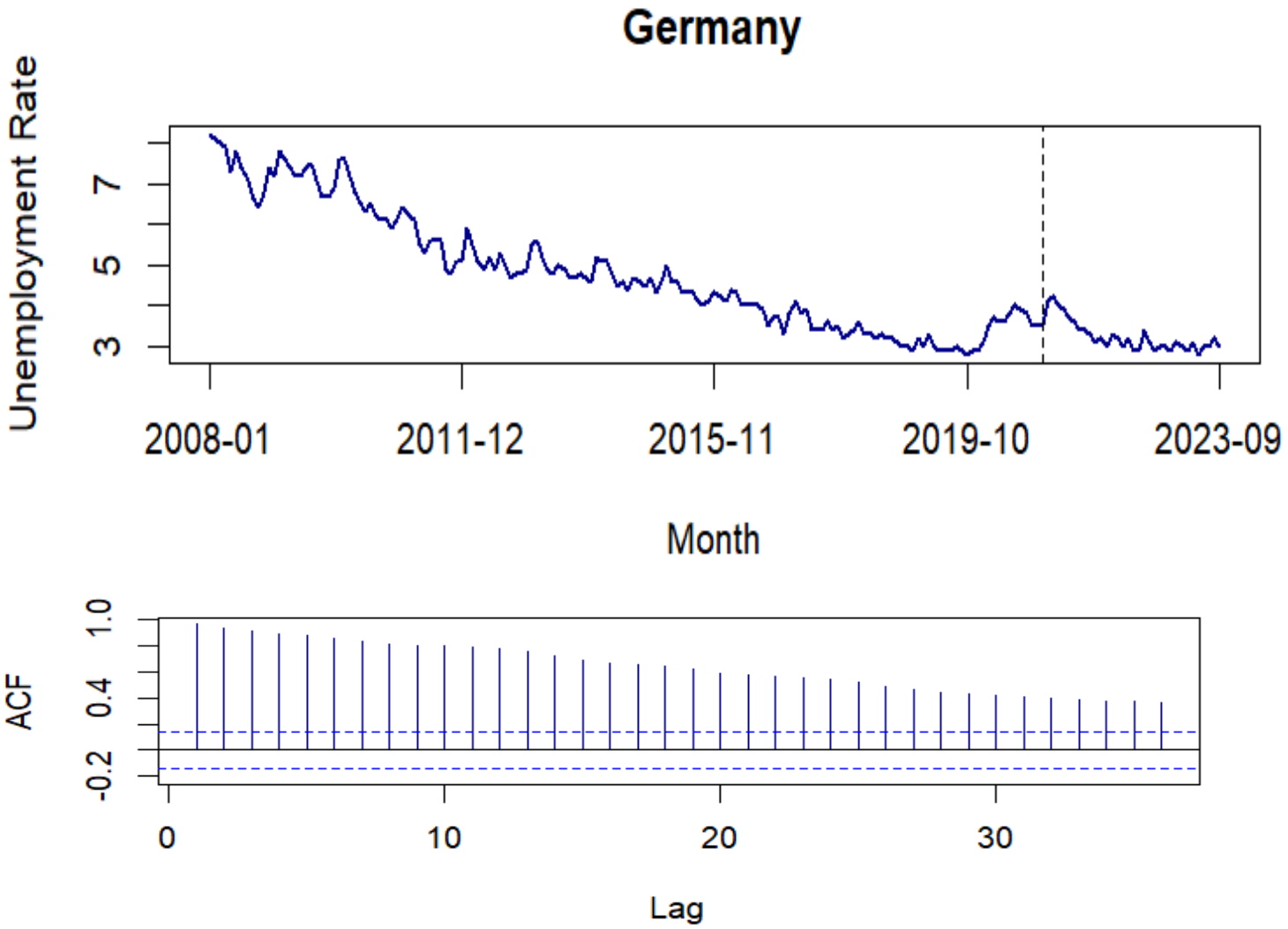
European Union Unemployment Rate



Germany, Italy and Greece Unemployment Rate



Germany



- Longterm decreasing trend

- Correlations are positive and significant(decreasing). This is a confirmation of existence of the trend

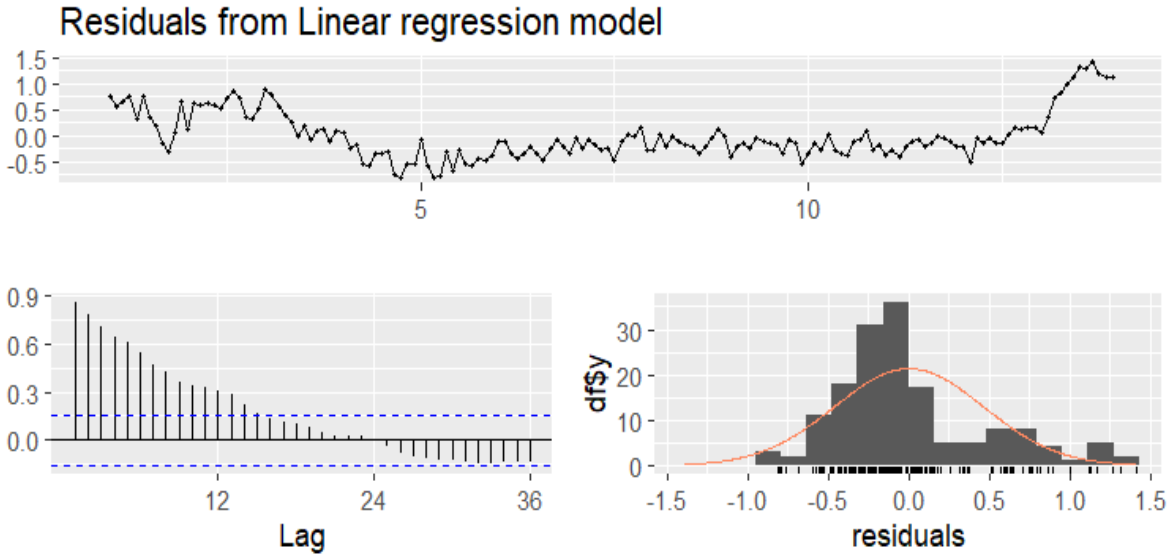
Modelling

The time-series is divided into two periods: the training set and the test set. The models are built using the training set. Once the models are defined, they are tested on the test set to assess their performance.

- Linear Regression model (trend + seasonality)
- Linear regression model (y ~ t + (t^2) + season)
- Holt Winters methods
- ARIMA
- Generalized additive model (y ~ s(t) + s)

Linear Regression (y ~ t + s)

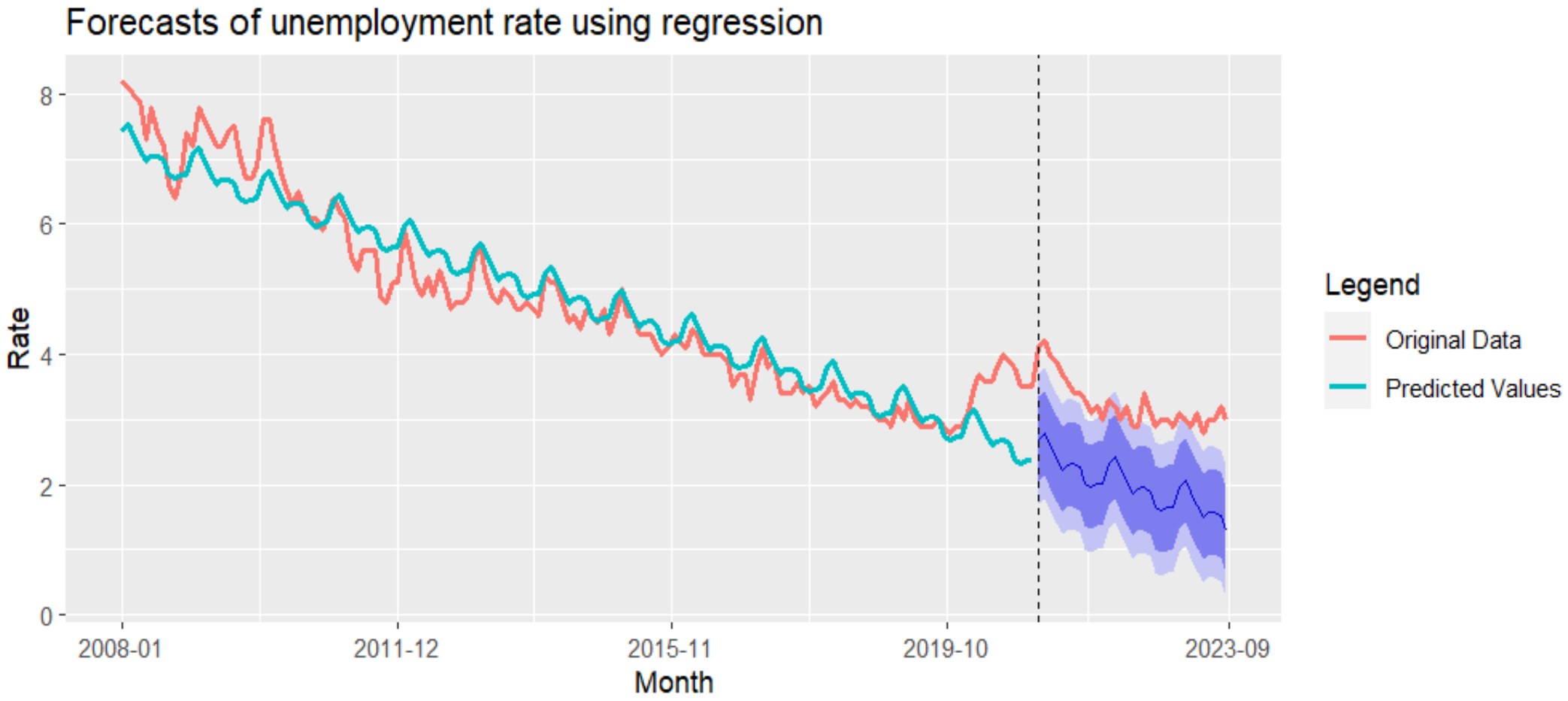
```
Call:
tslm(formula = GG2 ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8082 -0.2897 -0.1195  0.1522  1.4165

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.4622253  0.1476838  50.528   <2e-16 ***
trend       -0.0303571  0.0008595 -35.318   <2e-16 ***
season2      0.1303571  0.1890700   0.689   0.4916
season3     -0.0085165  0.1890759  -0.045   0.9641
season4     -0.1858516  0.1890857  -0.983   0.3273
season5     -0.3401099  0.1890993  -1.799   0.0742 .
season6     -0.2405220  0.1891169  -1.272   0.2055
season7     -0.1947802  0.1891384  -1.030   0.3048
season8     -0.2182692  0.1891638  -1.154   0.2505
season9     -0.4340659  0.1891931  -2.294   0.0232 *
season10    -0.4575549  0.1892263  -2.418   0.0169 *
season11    -0.3810440  0.1892634  -2.013   0.0460 *
season12    -0.3429945  0.1893043  -1.812   0.0721 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.482 on 143 degrees of freedom
Multiple R-squared:  0.9007,    Adjusted R-squared:  0.8923
F-statistic: 108.1 on 12 and 143 DF,  p-value: < 2.2e-16
```




```
        Durbin-Watson test

data:  m1
DW = 0.22501, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

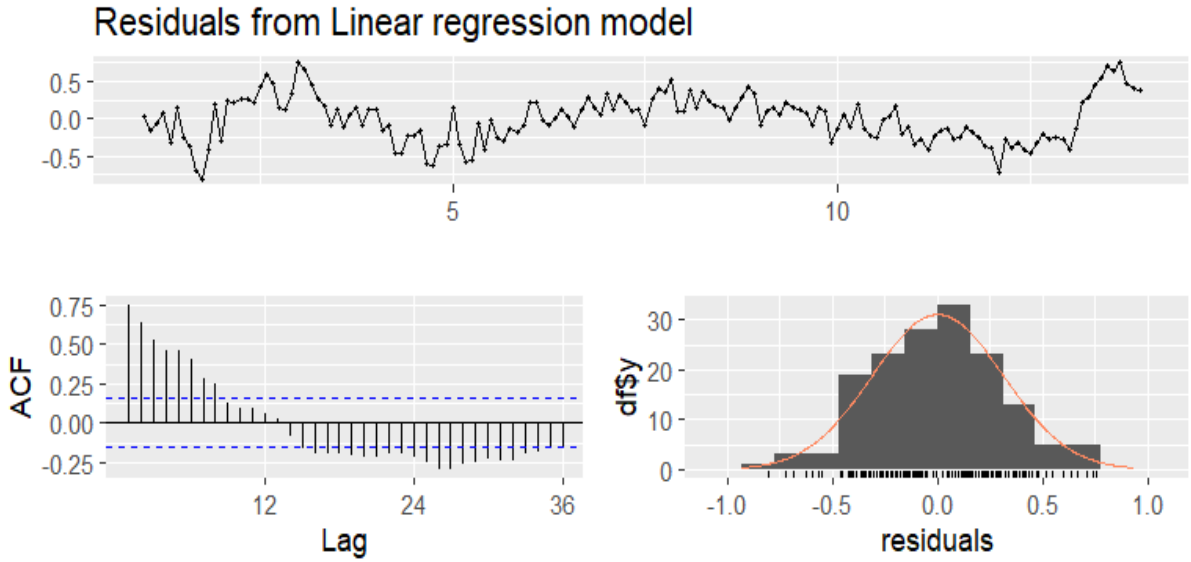
- **Motivation:** Simple and interpretable
- highly correlated residuals

Linear Regression (y ~ t + s)



Linear regression (y ~ t + (t^2) + season)

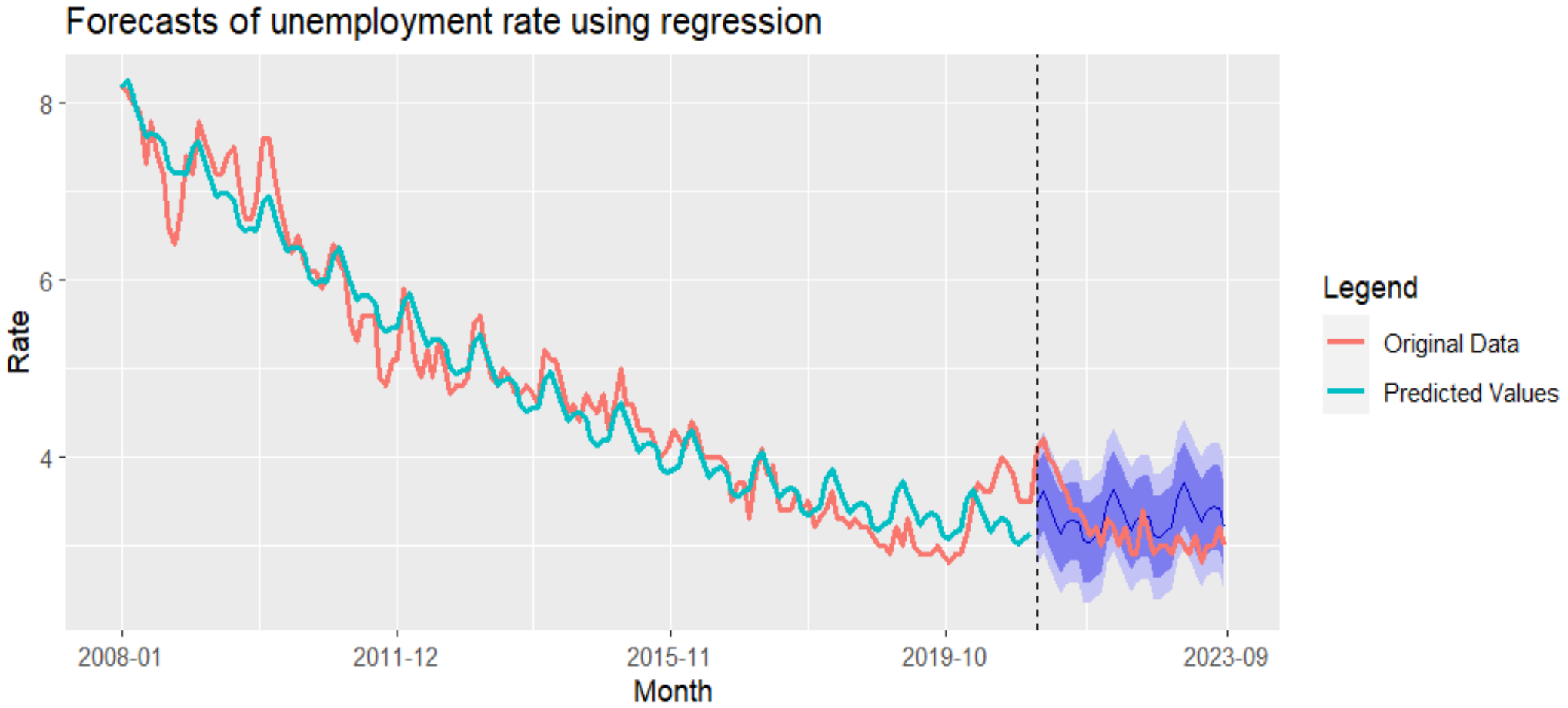
```
Call:
tslm(formula = GG2 ~ trend + I(trend^2) + season)

Residuals:
     Min       1Q   Median       3Q      Max
-0.80304 -0.23862  0.02173  0.20779  0.75175

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.238e+00  1.158e-01  71.133  < 2e-16 ***
trend       -5.994e-02  2.327e-03 -25.762  < 2e-16 ***
I(trend^2)   1.884e-04  1.435e-05  13.128  < 2e-16 ***
season2      1.322e-01  1.275e-01   1.037 0.301494
season3     -5.125e-03  1.275e-01  -0.040 0.968000
season4     -1.813e-01  1.275e-01  -1.422 0.157268
season5     -3.348e-01  1.275e-01  -2.625 0.009606 **
season6     -2.349e-01  1.276e-01  -1.841 0.067659 .
season7     -1.891e-01  1.276e-01  -1.483 0.140407
season8     -2.130e-01  1.276e-01  -1.669 0.097237 .
season9     -4.295e-01  1.276e-01  -3.366 0.000981 ***
season10    -4.542e-01  1.276e-01  -3.559 0.000508 ***
season11    -3.792e-01  1.277e-01  -2.970 0.003495 **
season12    -3.430e-01  1.277e-01  -2.686 0.008085 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3251 on 142 degrees of freedom
Multiple R-squared:  0.9551,    Adjusted R-squared:  0.951
F-statistic: 232.5 on 13 and 142 DF,  p-value: < 2.2e-16
```



Durbin-Watson test

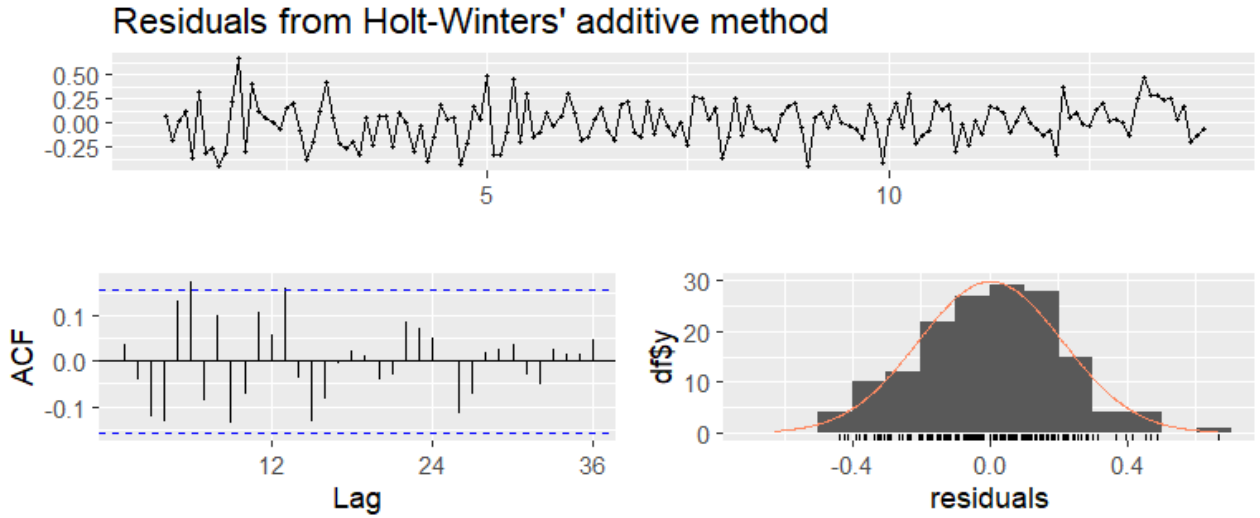
```
data:  m2
DW = 0.49365, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

Linear regression (y ~ t + (t^2) + season)



Holt-Winters method (additive)

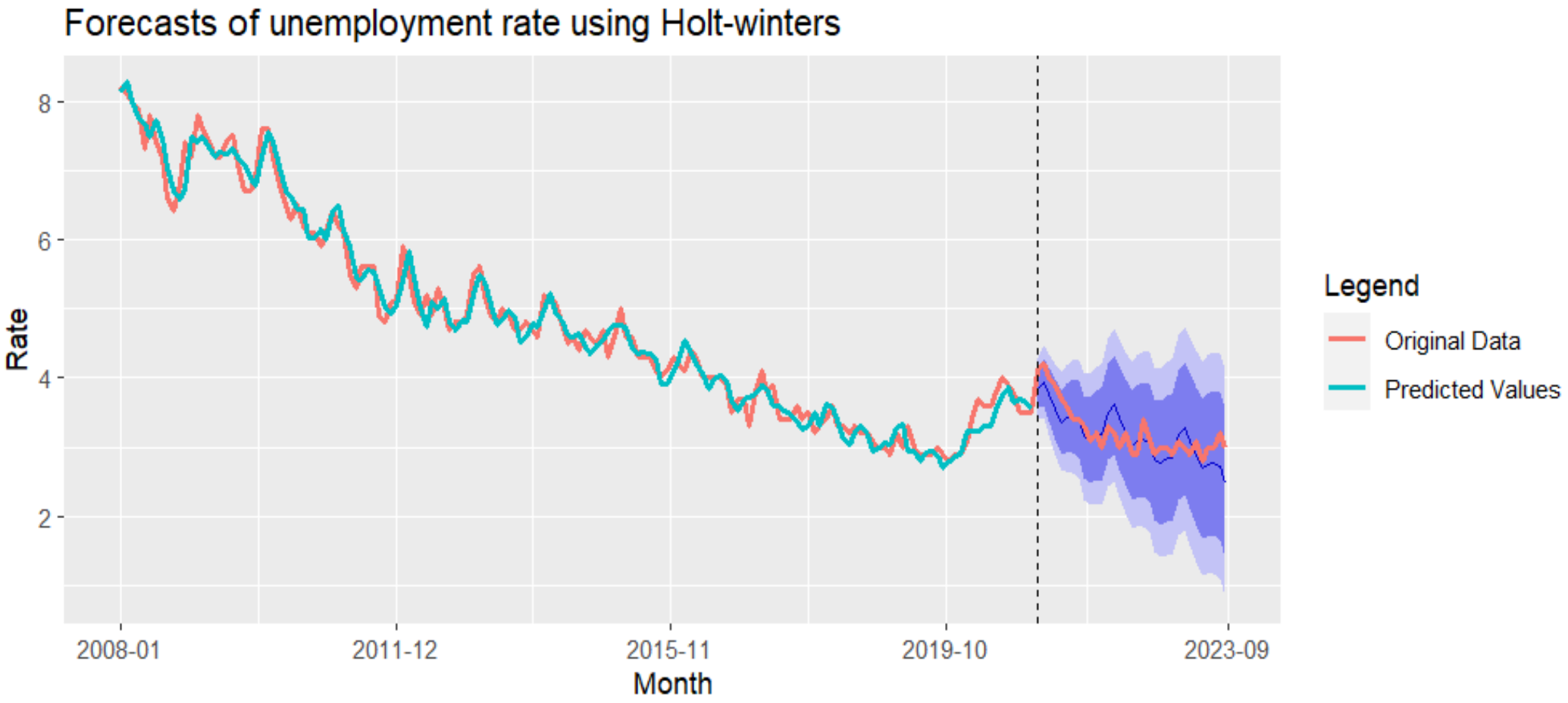
```
Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
 hw(y = GG2, h = 33, seasonal = "additive")

  Smoothing parameters:
    alpha = 0.649
    beta  = 1e-04
    gamma = 1e-04

  Initial states:
    l = 7.9179
    b = -0.0281
    s = -0.1075 -0.1486 -0.2423 -0.2256 -9e-04 0.0168
           -0.0409 -0.1265 0.0334 0.2227 0.3737 0.2458

  sigma:  0.2209

     AIC     AICc      BIC
333.7354 338.1702 385.5830

Error measures:
                     ME      RMSE       MAE        MPE     MAPE      MASE       ACF1
Training set 0.001090096 0.2092475 0.1679646 0.0699846 3.583593 0.3721062 0.03781422
```



Motivation:

- capture both the trend and seasonality
- Adaptability to Changing Patterns
- Forecasting Accuracy

Holt-Winters method (additive)

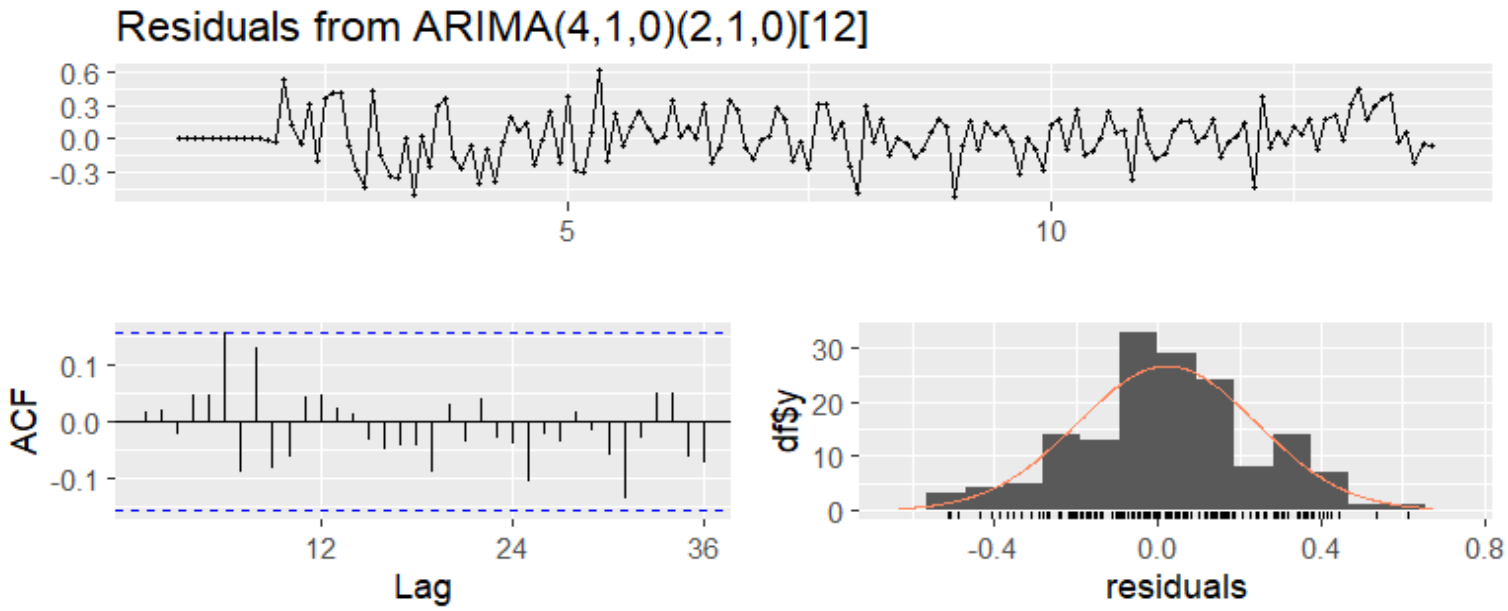


Seasonal ARIMA (4,1,0)(2,1,0)[12]

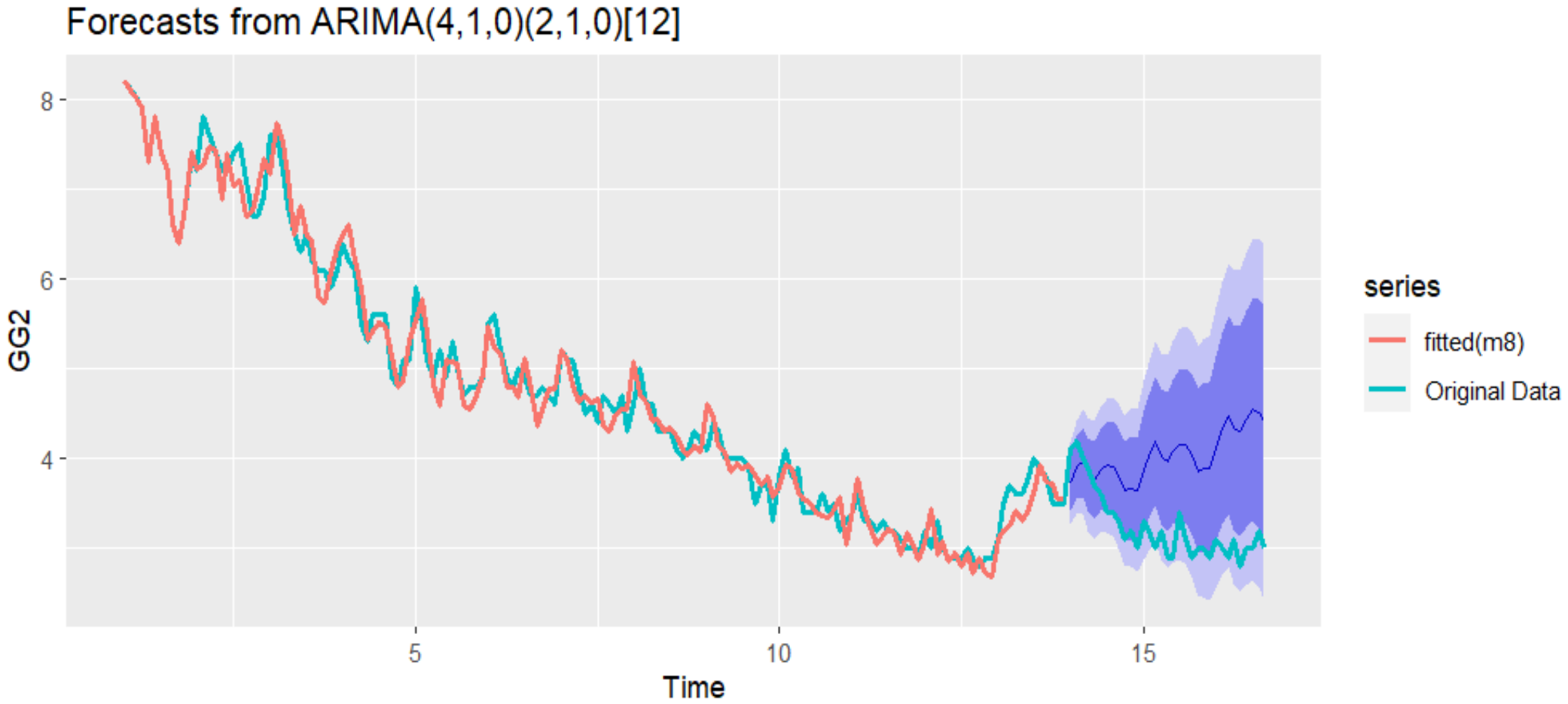
```
Series: GG2
ARIMA(4,1,0)(2,1,0)[12]

Coefficients:
          ar1      ar2      ar3      ar4     sar1     sar2
      -0.4372  -0.1927  -0.1452  -0.1783  -0.6137  -0.2930
s.e.   0.0838   0.0892   0.0895   0.0835   0.0920   0.0948

sigma^2 = 0.05449:  log likelihood = 5.44
AIC=3.12   AICc=3.95   BIC=23.86
```

Motivations:

- Robustness
- Interpretability
- No Assumptions about Data Distribution
- Ease of Use

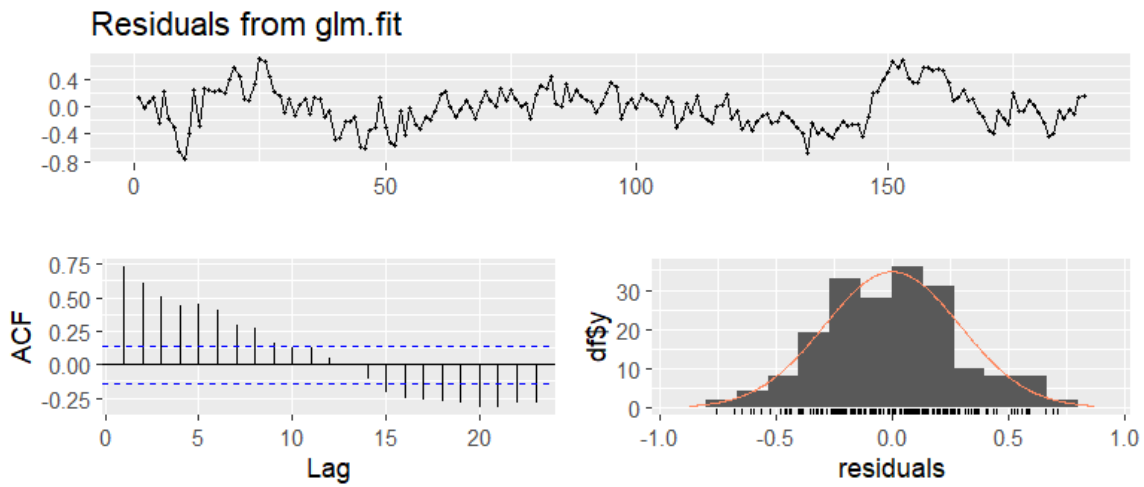


Seasonal ARIMA (4,1,0)(2,1,0)[12]



Generalized Additive Model (y ~ s(t) + s)

```
Call: gam(formula = G ~ s(t) + seas)
Deviance Residuals:
      Min        1Q    Median        3Q       Max
-0.759575 -0.205052  0.001231  0.181654  0.711276

(Dispersion Parameter for gaussian family taken to be 0.0918)

    Null Deviance: 410.4277 on 188 degrees of freedom
Residual Deviance: 15.8874 on 173.0001 degrees of freedom
AIC: 102.3529

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
           Df Sum Sq Mean Sq   F value    Pr(>F)
s(t)        1 345.54  345.54 3762.6616 < 2.2e-16 ***
seas       11   5.64    0.51    5.5856  1.23e-07 ***
Residuals 173  15.89    0.09
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
           Npar Df Npar F     Pr(F)
(Intercept)
s(t)             3  153.3 < 2.2e-16 ***
seas
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Motivation:

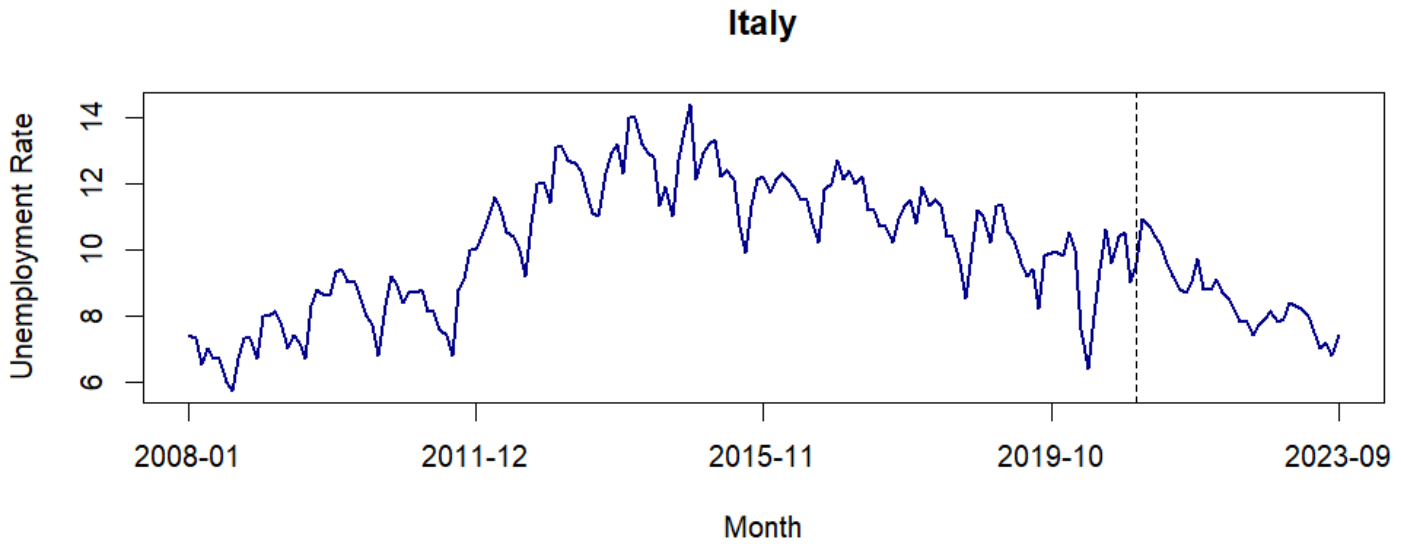
- Capturing Non-Linear Patterns

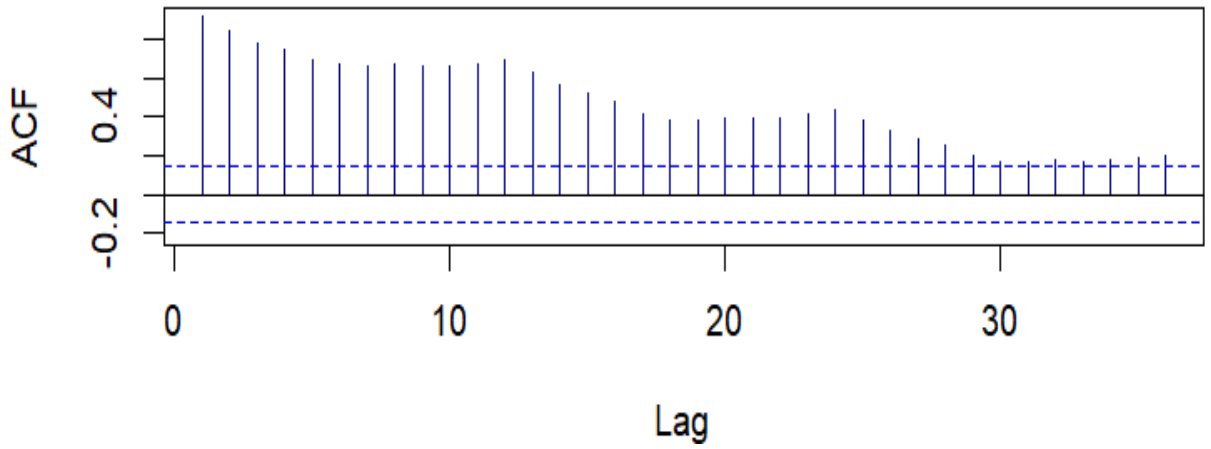
Generalized Additive Model (y ~ s(t) + s)

Comparing Models for Germany Via Metrics

Models	Train/Test	RMSE	MAPE	AIC	BIC
TSLM1	Training set	0.46	7.52	229.45	272.15
	Test set	1.25	38.31		
TSLM2	Training set	0.31	5.77	107.48	153.23
	Test set	0.38	10.12		
Holt-Winters method(additive)	Training set	0.21	3.58	333.73	385.58
	Test set	0.24	6.18		
SARIMA	Training set	0.21	3.64	3.12	23.86
	Test set	0.96	27.71		
GAM	Training set	0.29	5.54	102.35	147.74
	Test set	0.86	25.69		

Italy





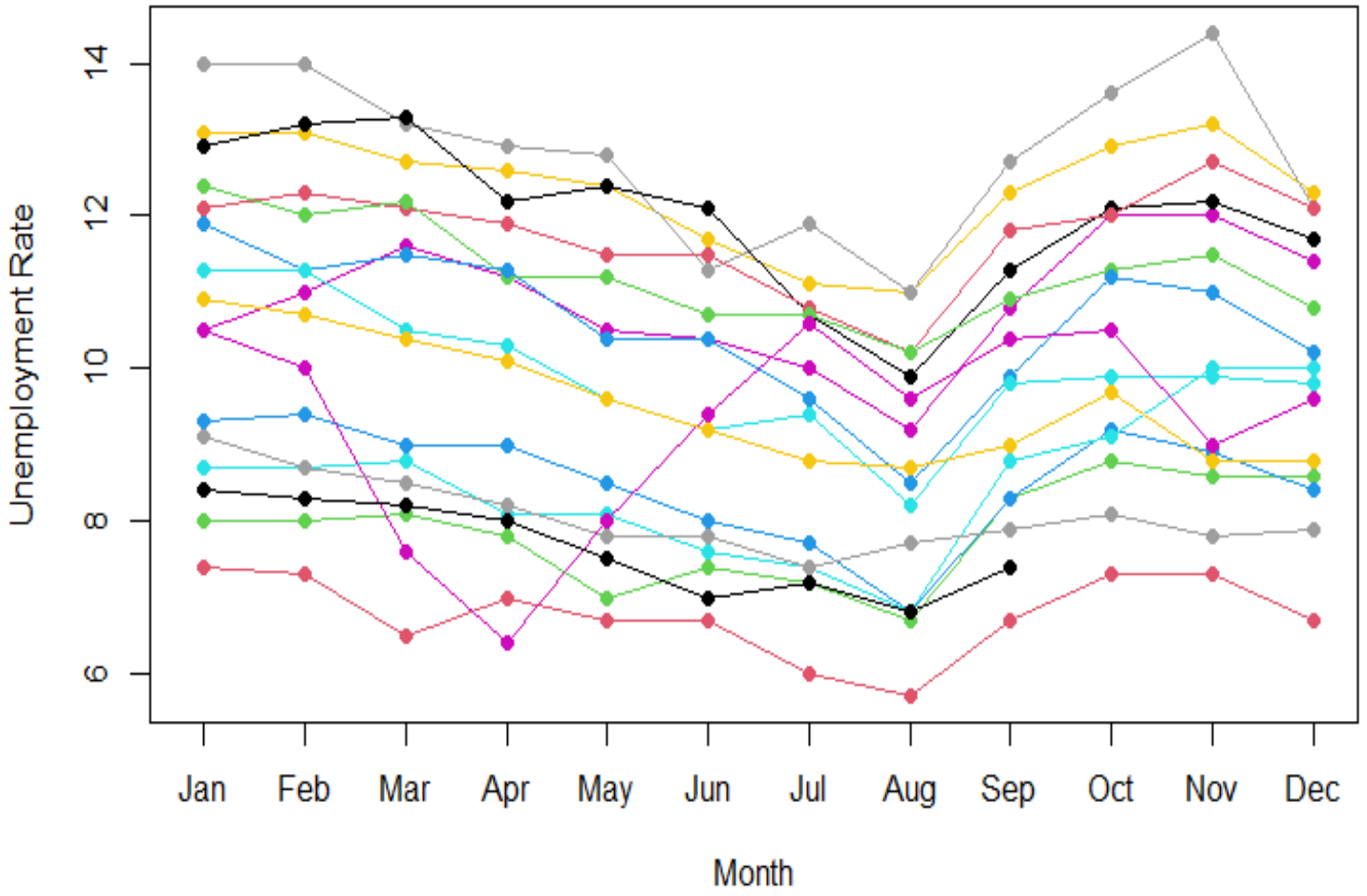
- Political Instability
- Banking Sector Challenges
- Before the COVID-19 pandemic, which began in 2019, Italy's unemployment rate was gradually declining

- Trend and Seasonality

Seasonal Plot

Lowest rate in August:

- Seasonal Employment
- Summer Hiring
- Youth Employment
- Government Employment Programs

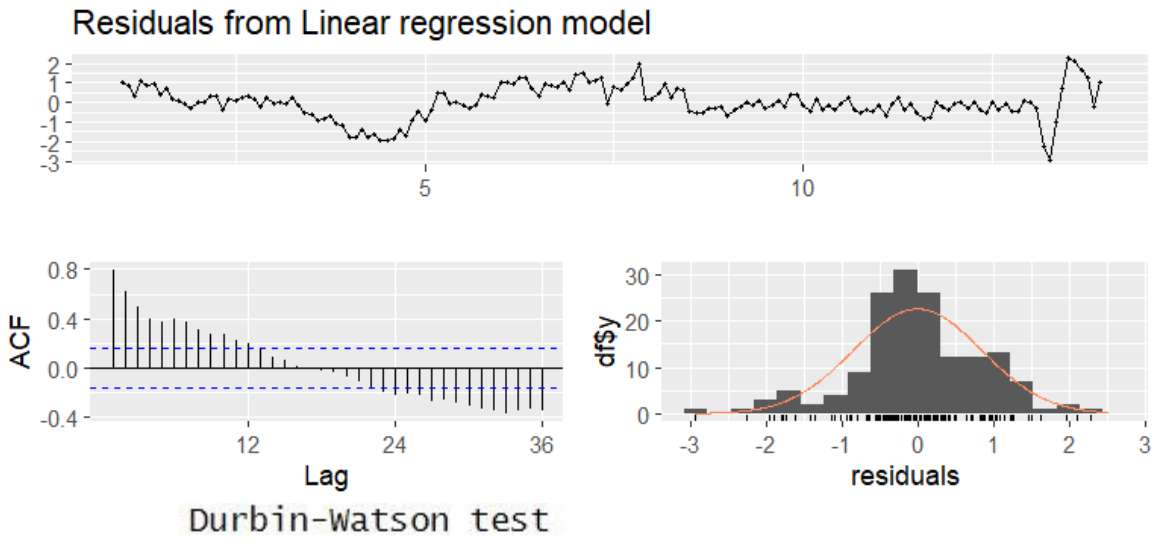


Modelling

- Linear regression model (y ~ t + (t^2) + season)
- Holt Winters methods
- ARIMA
- Generalized additive model (y ~ s(t) + s)

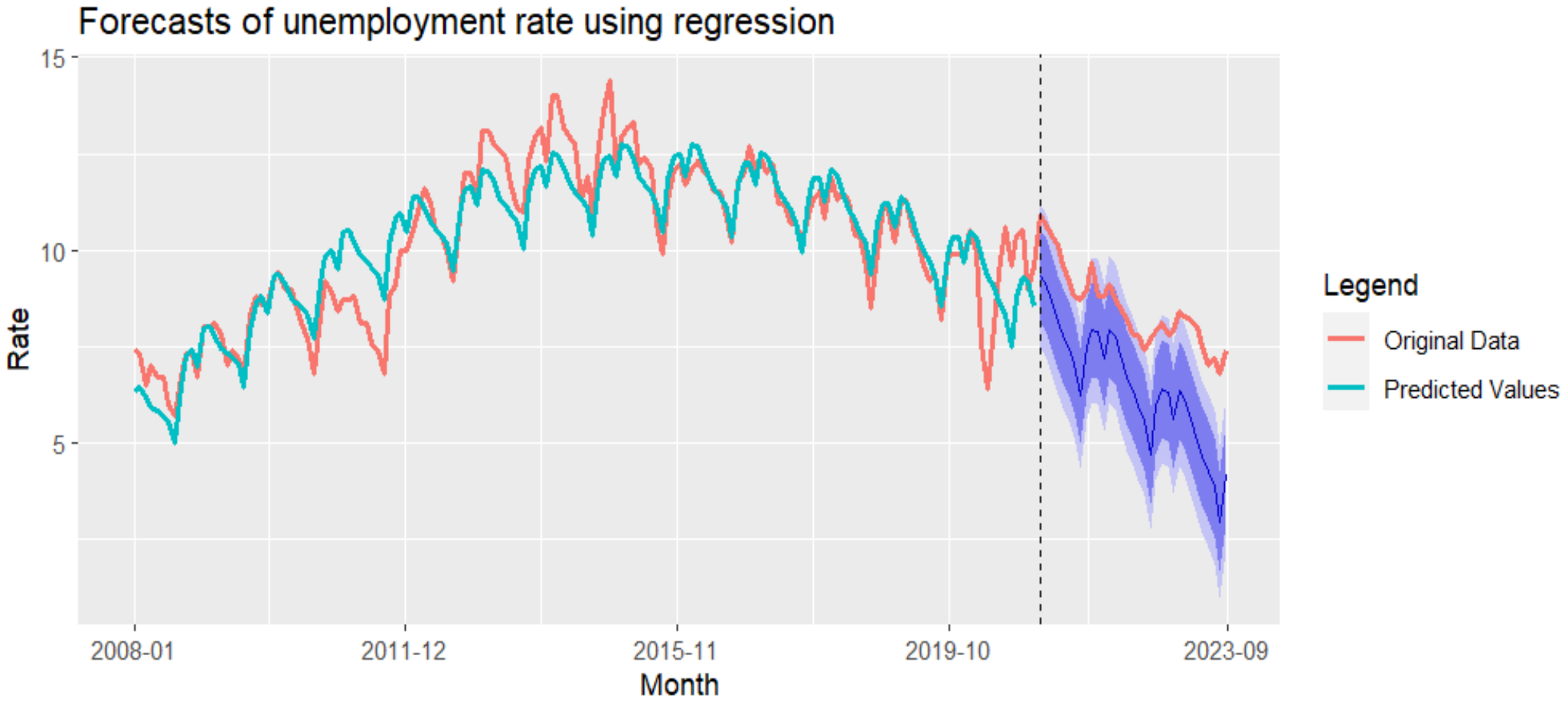
Linear regression (y ~ t + (t^2) + season)

```
Call:
tslm(formula = I1 ~ trend + I(trend^2) + season)

Residuals:
     Min       1Q   Median       3Q      Max
-2.93832 -0.39243 -0.02749  0.44546  2.28246

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.207608   0.314678  19.727  < 2e-16 ***
trend        0.144787   0.006322  22.902  < 2e-16 ***
I(trend^2)  -0.000796   0.000039 -20.410  < 2e-16 ***
season2     -0.066243   0.346509  -0.191 0.848663
season3     -0.438587   0.346520  -1.266 0.207697
season4     -0.863185   0.346539  -2.491 0.013895 *
season5     -1.101575   0.346564  -3.179 0.001817 **
season6     -1.330682   0.346597  -3.839 0.000185 ***
season7     -1.604350   0.346636  -4.628 8.23e-06 ***
season8     -2.337965   0.346682  -6.744 3.60e-10 ***
season9     -0.954603   0.346736  -2.753 0.006675 **
season10    -0.361958   0.346796  -1.044 0.298389
season11    -0.313874   0.346863  -0.905 0.367055
season12    -0.864198   0.346938  -2.491 0.013893 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8834 on 142 degrees of freedom
Multiple R-squared:  0.8217,    Adjusted R-squared:  0.8054
F-statistic: 50.34 on 13 and 142 DF,  p-value: < 2.2e-16
```



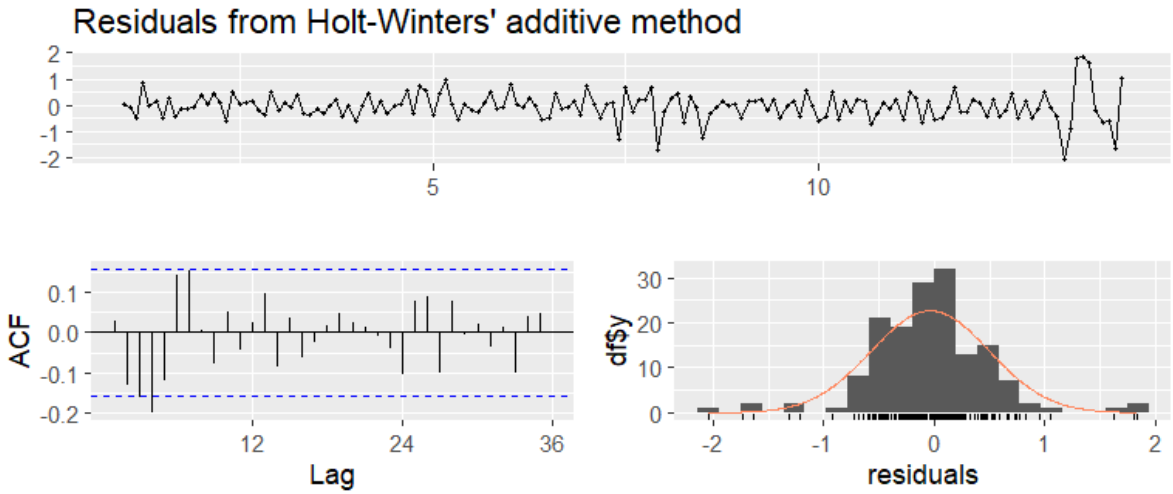

```
Durbin-Watson test

data:  n1
DW = 0.39254, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

Linear regression (y ~ t + (t^2) + season)

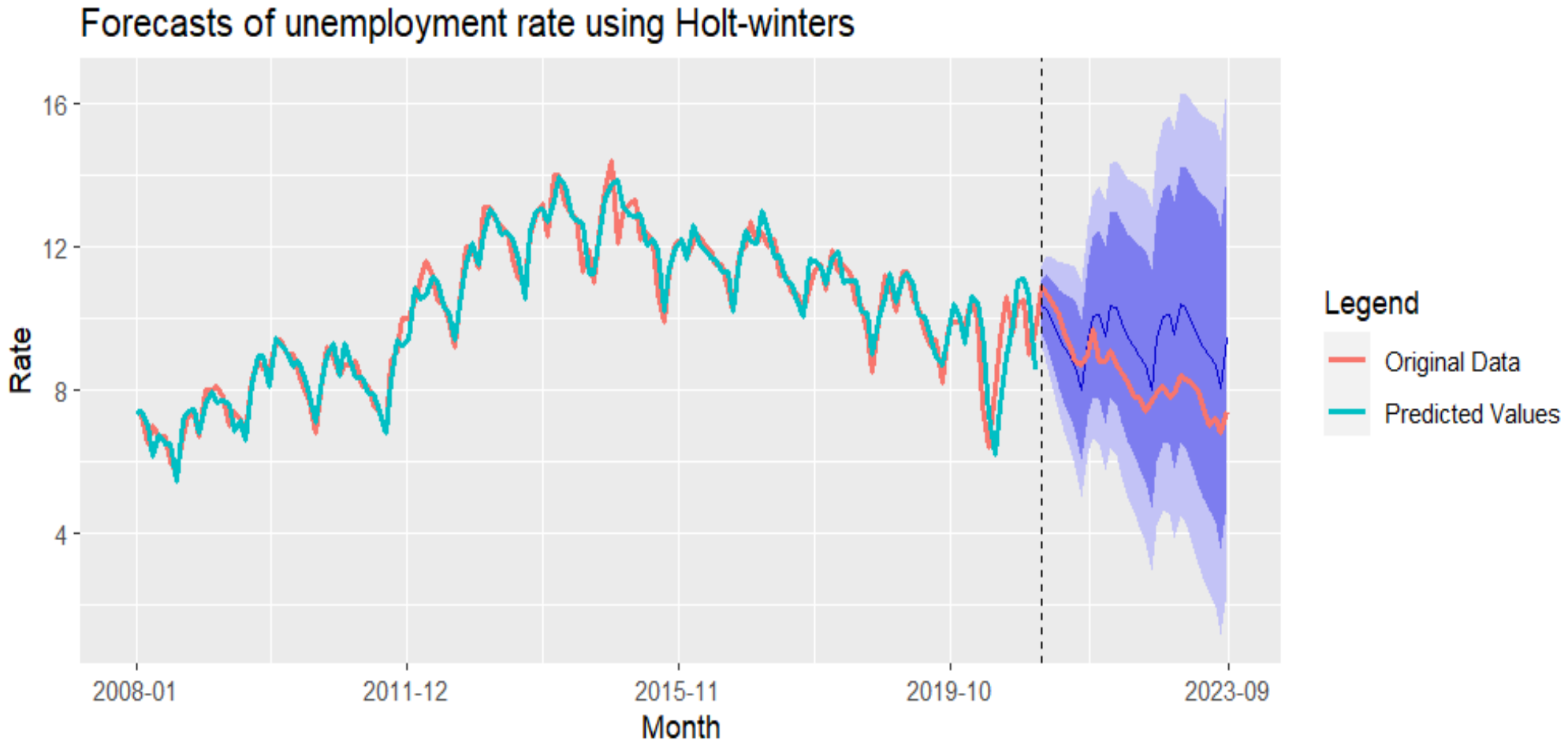


Holt-Winters method (additive)

```
Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
 hw(y = I1, h = 33, seasonal = "additive")

  Smoothing parameters:
    alpha = 0.9067
    beta  = 0.0121
    gamma = 1e-04

  Initial states:
    l = 6.4288
    b = 0.0701
    s = 0.0264 0.6134 0.5377 -0.0726 -1.4783 -0.8077
           -0.552 -0.2952 -0.038 0.4044 0.7959 0.866

  sigma:  0.5624

     AIC     AICc      BIC
625.3502 629.7850 677.1977

Error measures:
                     ME      RMSE       MAE        MPE     MAPE      MASE       ACF1
Training set -0.03624571 0.5328183 0.3777551 -0.4797393 3.878343 0.3938938 0.02852021
```

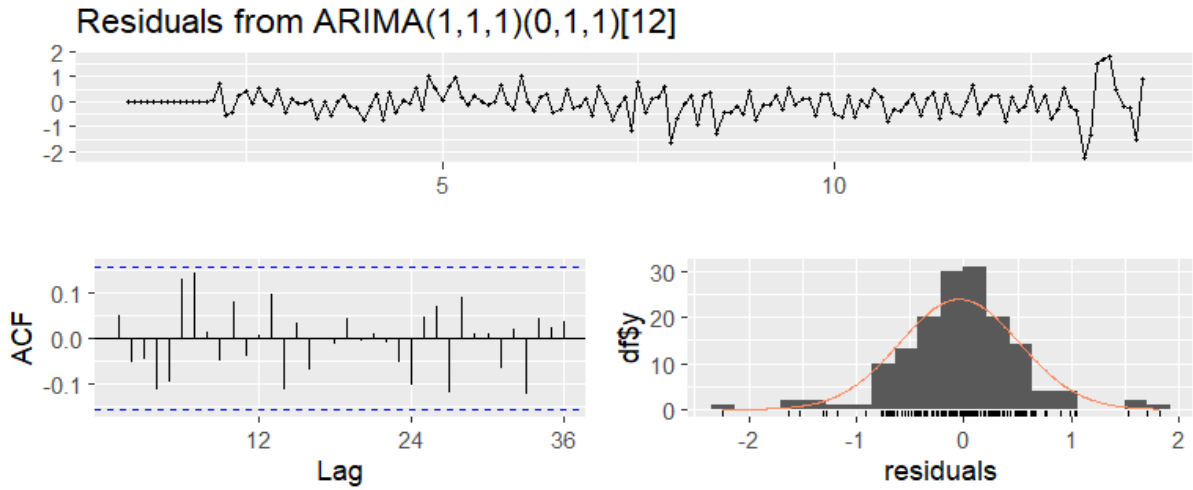


Holt-Winters method (additive)

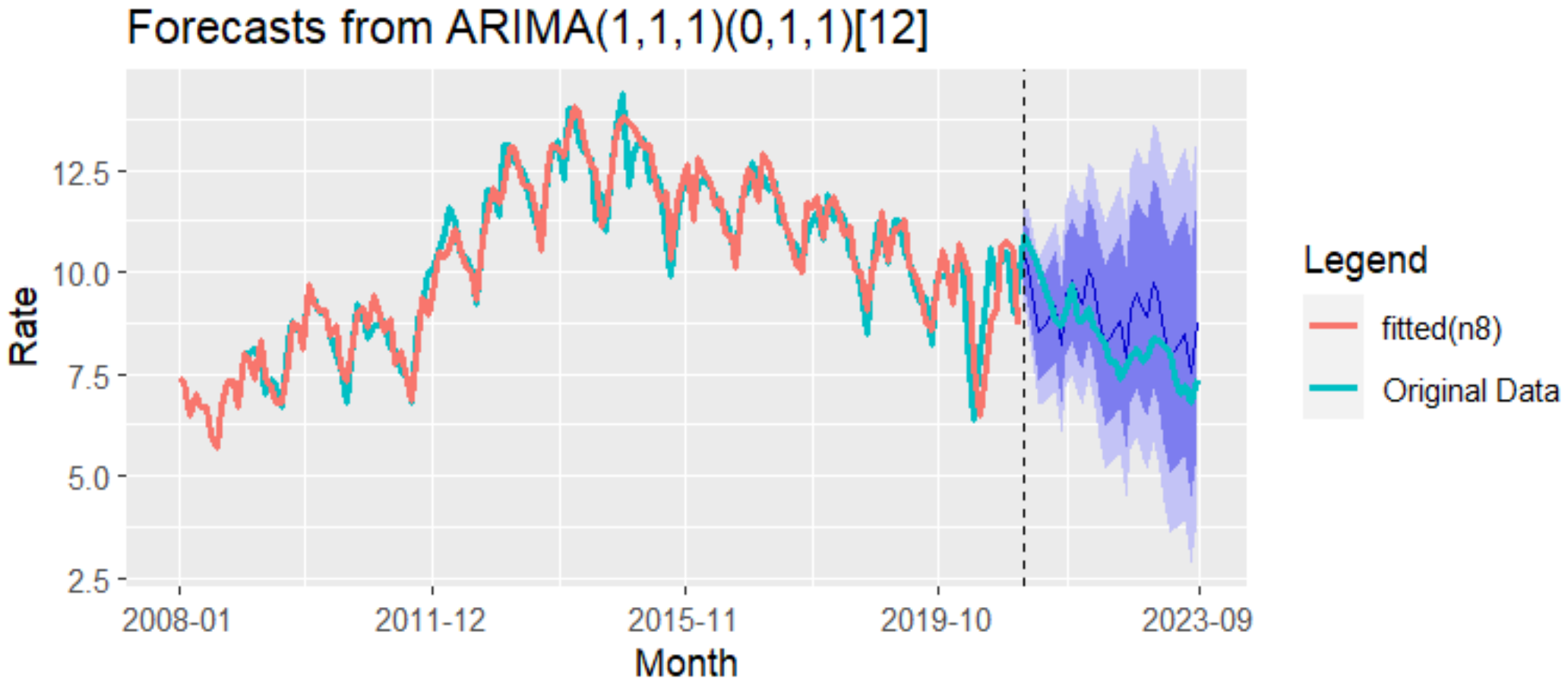


Seasonal ARIMA (1,1,1)(0,1,1)[12]

```
Series: I1
ARIMA(1,1,1)(0,1,1)[12]

Coefficients:
          ar1      ma1     sma1
      0.595  -0.7969  -0.6716
s.e.  0.130   0.0929   0.1193

sigma^2 = 0.3418:  log likelihood = -128.35
AIC=264.71   AICc=265   BIC=276.56

Training set error measures:
                     ME      RMSE       MAE         MPE     MAPE      MASE        ACF1
Training set -0.04285323 0.5538336 0.392712 -0.5206556 3.92337 0.4094897 0.04855059
```

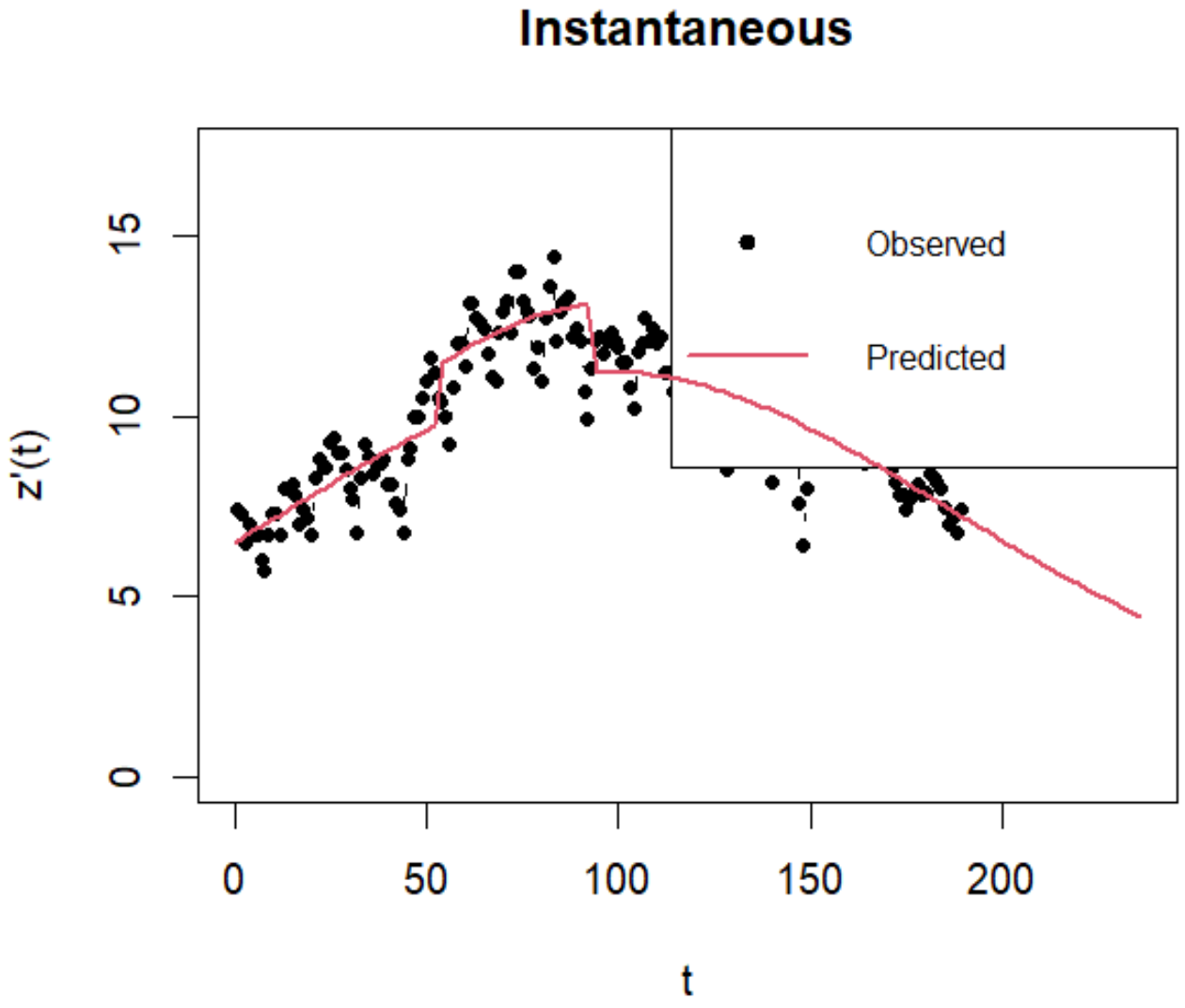


Seasonal ARIMA (1,1,1)(0,1,1)[12]



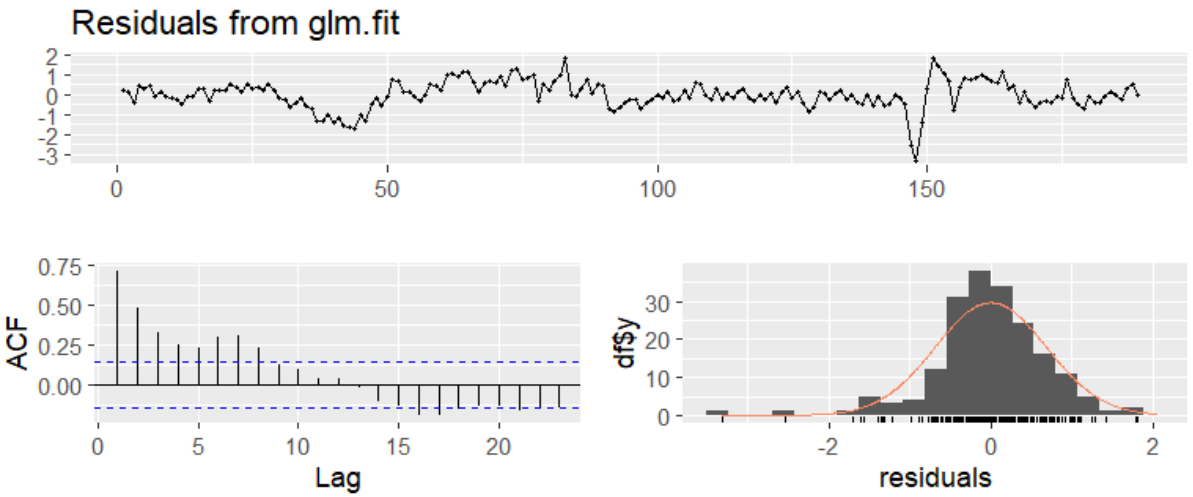
Generalized Bass Model

- **Motivation:** Check if there is shock after the 2008 economic crisis
- may not be good models because of their assumptions

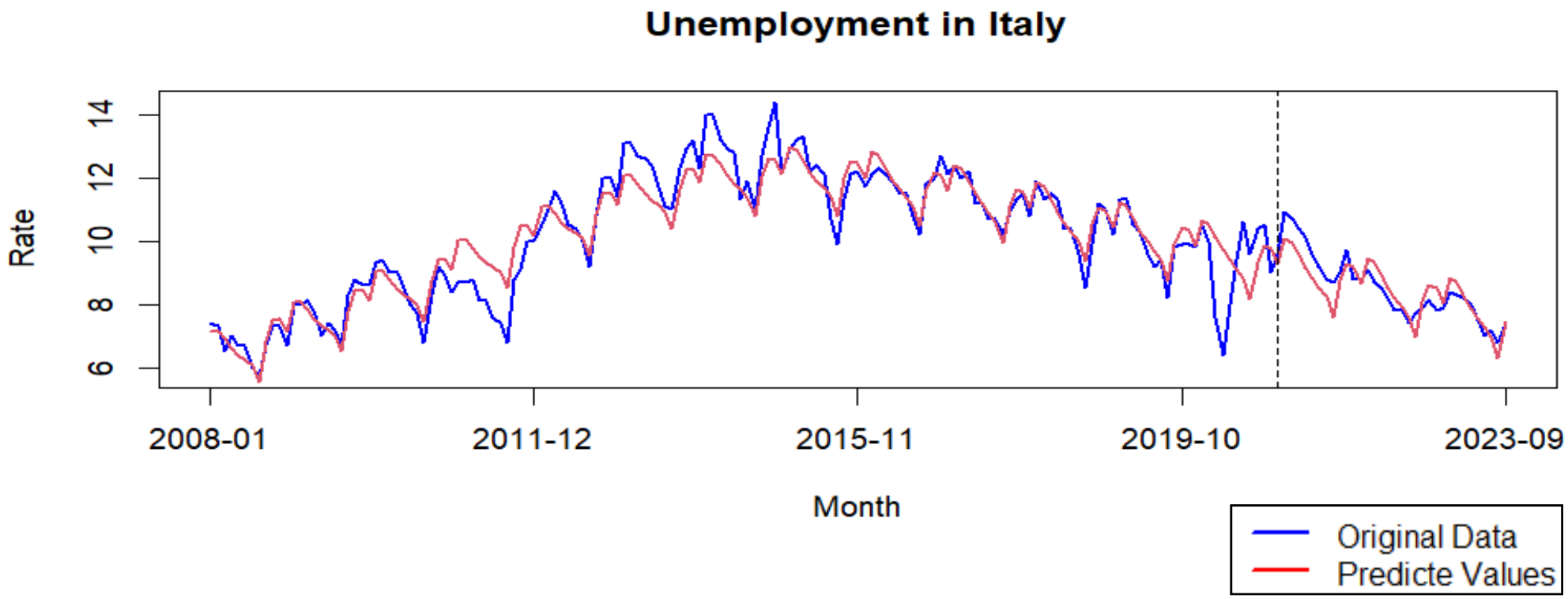


Generalized Additive Model (y ~ s(t) + s)

```
Call: gam(formula = I ~ s(t) + seas)
Deviance Residuals:
     Min       1Q   Median       3Q      Max
-3.30475 -0.35263 -0.02209  0.43941  1.81674

(Dispersion Parameter for gaussian family taken to be 0.51)

    Null Deviance: 735.2674 on 188 degrees of freedom
Residual Deviance: 88.2329 on 173.0001 degrees of freedom
AIC: 426.3847

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
           Df Sum Sq Mean Sq F value    Pr(>F)
s(t)        1  3.986  3.9855  7.8145  0.005769 **
seas       11 69.156  6.2869 12.3269 < 2.2e-16 ***
Residuals 173 88.233  0.5100
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
            Npar Df Npar F     Pr(F)
(Intercept)
s(t)              3 372.02 < 2.2e-16 ***
seas
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```



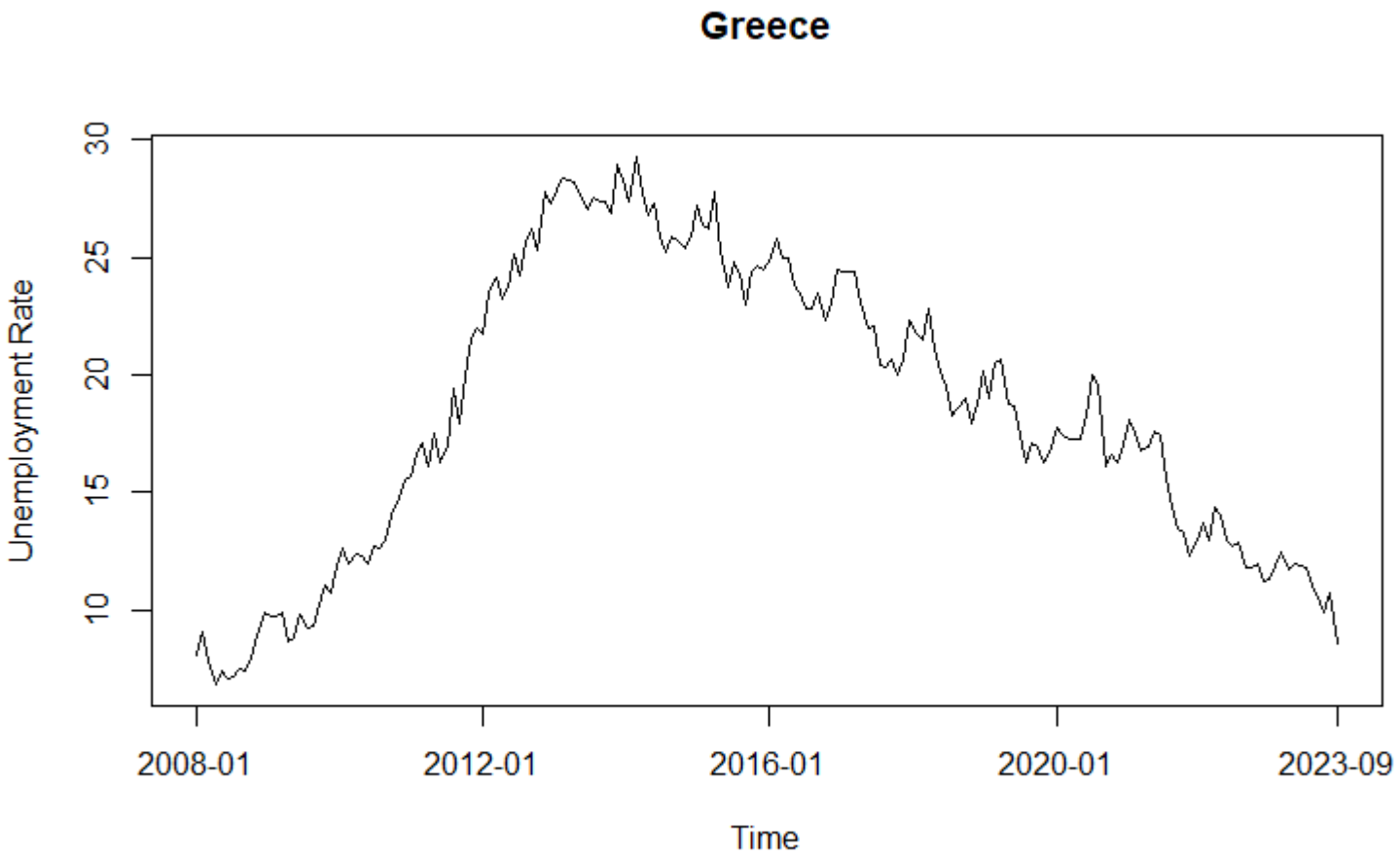
Generalized Additive Model (y ~ s(t) + s)



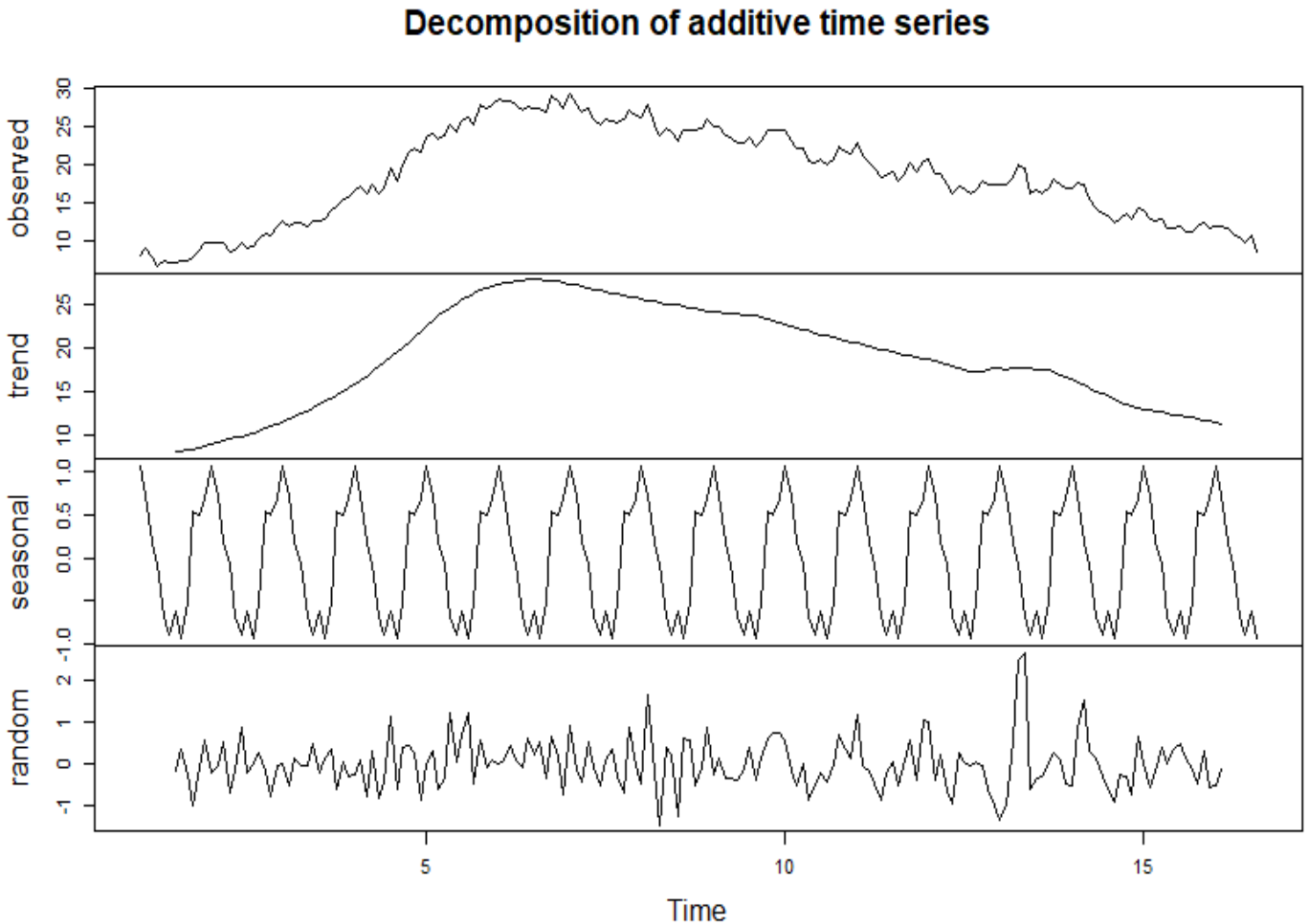
Comparing Models For Italy Via Metrics

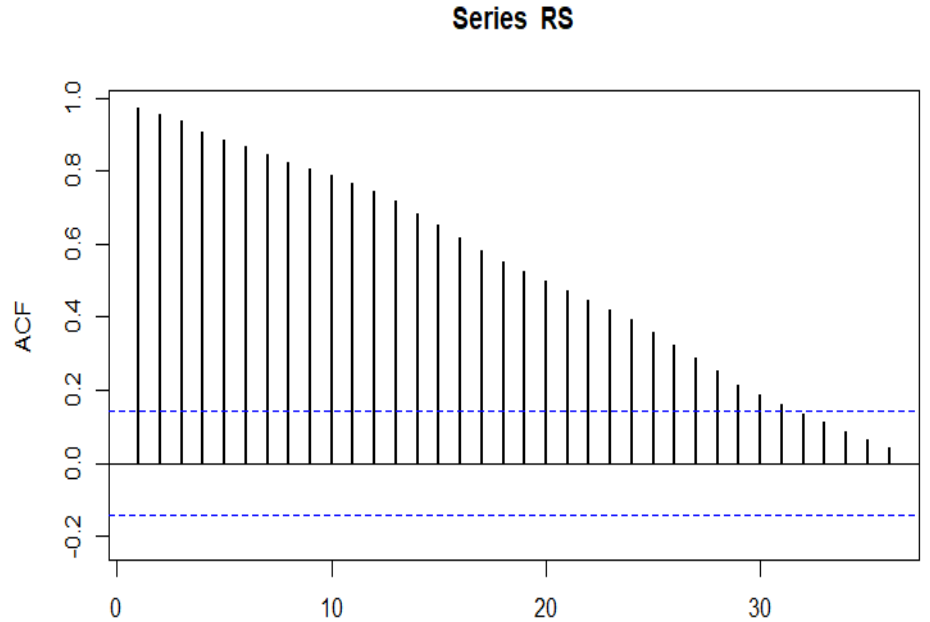
Models	Train/Test	RMSE	MAPE	AIC	BIC
TSLM	Train set	0.84	6.52	419.36	465.11
	Test set	2.14	24.68		
Holt-Winters method(additive)	Train set	0.53	3.88	777.46	832.57
	Test set	1.34	14.72		
SARIMA	Train set	0.55	3.92	264.71	276.56
	Test set	0.91	9.49		
GAM	Train set	0.68	5.29	426	471.76
	Test set	1.12	12.31		

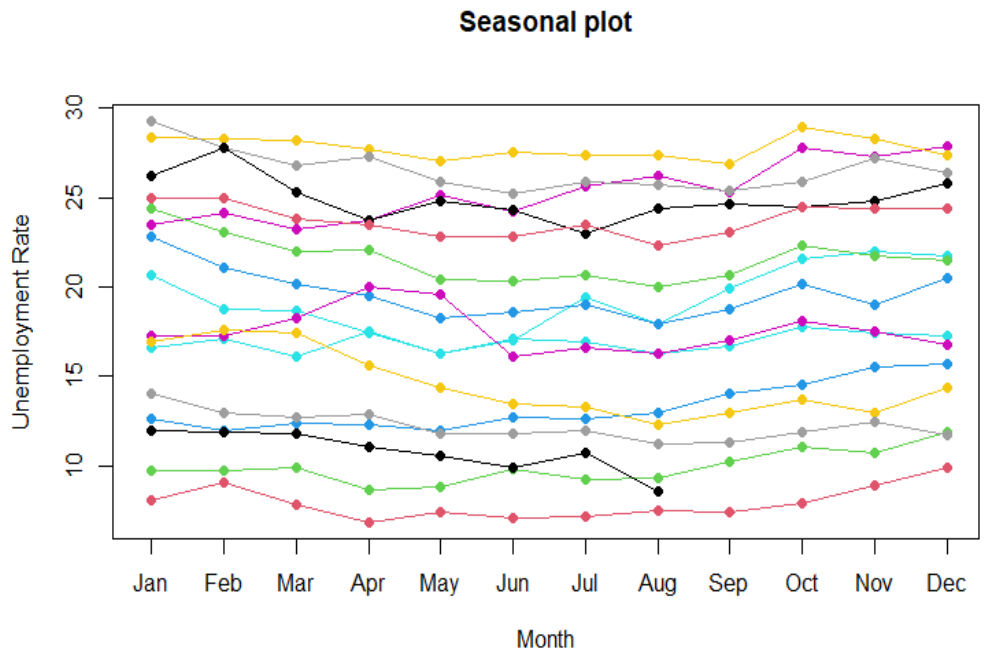
Greece



- Great Recession (2007-2009)
- Eurozone Crisis (2010-2012)
- Greece's Own Debt Crisis
- EU Funding and Support
- Structural Reforms


Decomposition of additive time series
observed
trend
seasonal
random
Time


Series RS
ACF


Seasonal plot
Unemployment Rate
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Modelling

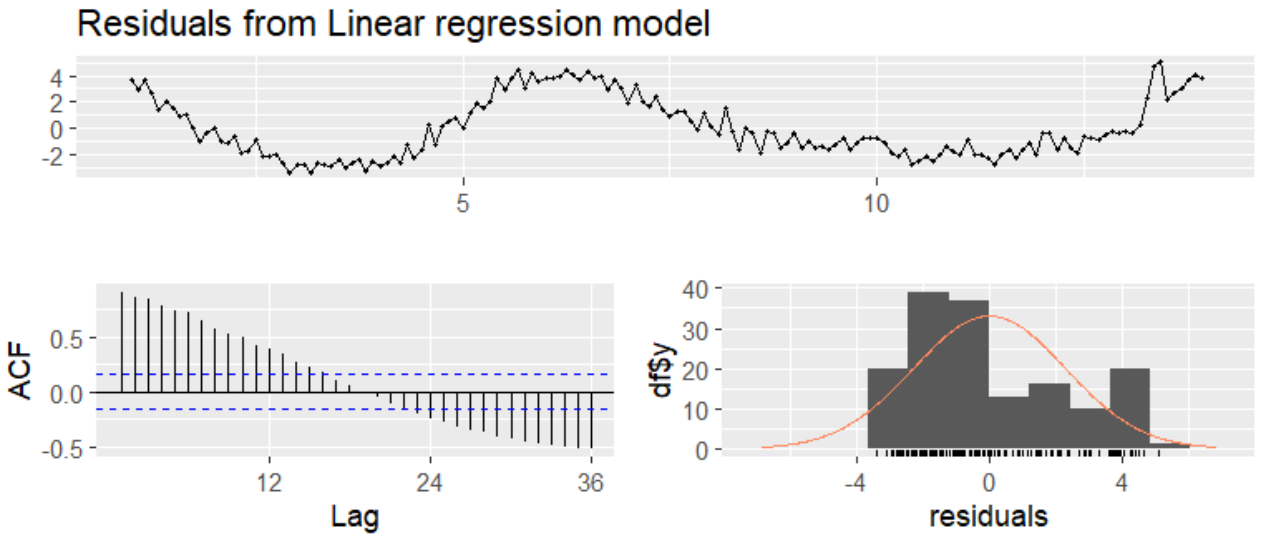
- Linear regression model (y ~ t + (t^2) + season)
- Smoothing Spline
- Diffusion models
- Holt Winters method (additive)
- Gradient Boosting

Linear regression (y ~ t + (t^2) + season)

```
Call:
tslm(formula = Grr1 ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4060 -1.8961 -0.5716  1.7366  5.0781

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.8499506  0.8481848   4.539 1.19e-05 ***
trend        0.5022978  0.0170406  29.477  < 2e-16 ***
I(trend^2)  -0.0028102  0.0001051 -26.734  < 2e-16 ***
season2      0.3877216  0.9339817   0.415   0.6787
season3      0.0426021  0.9340120   0.046   0.9637
season4     -0.6892048  0.9340617  -0.738   0.4618
season5     -0.9538529  0.9341305  -1.021   0.3089
season6     -1.4436498  0.9342181  -1.545   0.1245
season7     -1.6585957  0.9343242  -1.775   0.0780 .
season8     -1.3833057  0.9344489  -1.480   0.1410
season9     -1.6485493  0.9345924  -1.764   0.0799 .
season10    -1.2466340  0.9347549  -1.334   0.1845
season11    -0.1237138  0.9349367  -0.132   0.8949
season12    -0.1874809  0.9351386  -0.200   0.8414
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.381 on 142 degrees of freedom
Multiple R-squared:  0.8679,    Adjusted R-squared:  0.8558
F-statistic: 71.75 on 13 and 142 DF,  p-value: < 2.2e-16
```

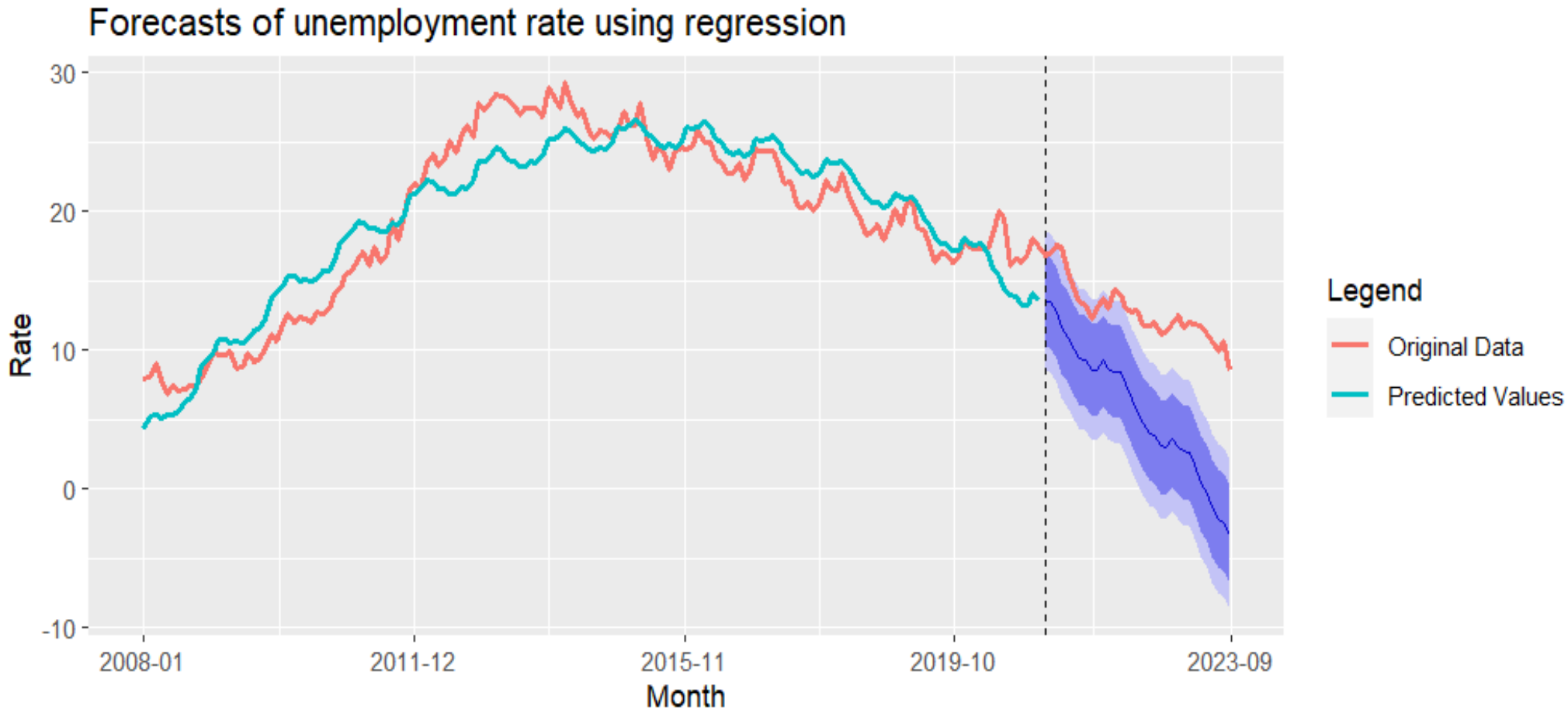
Motivations:

- Simple and interpretable



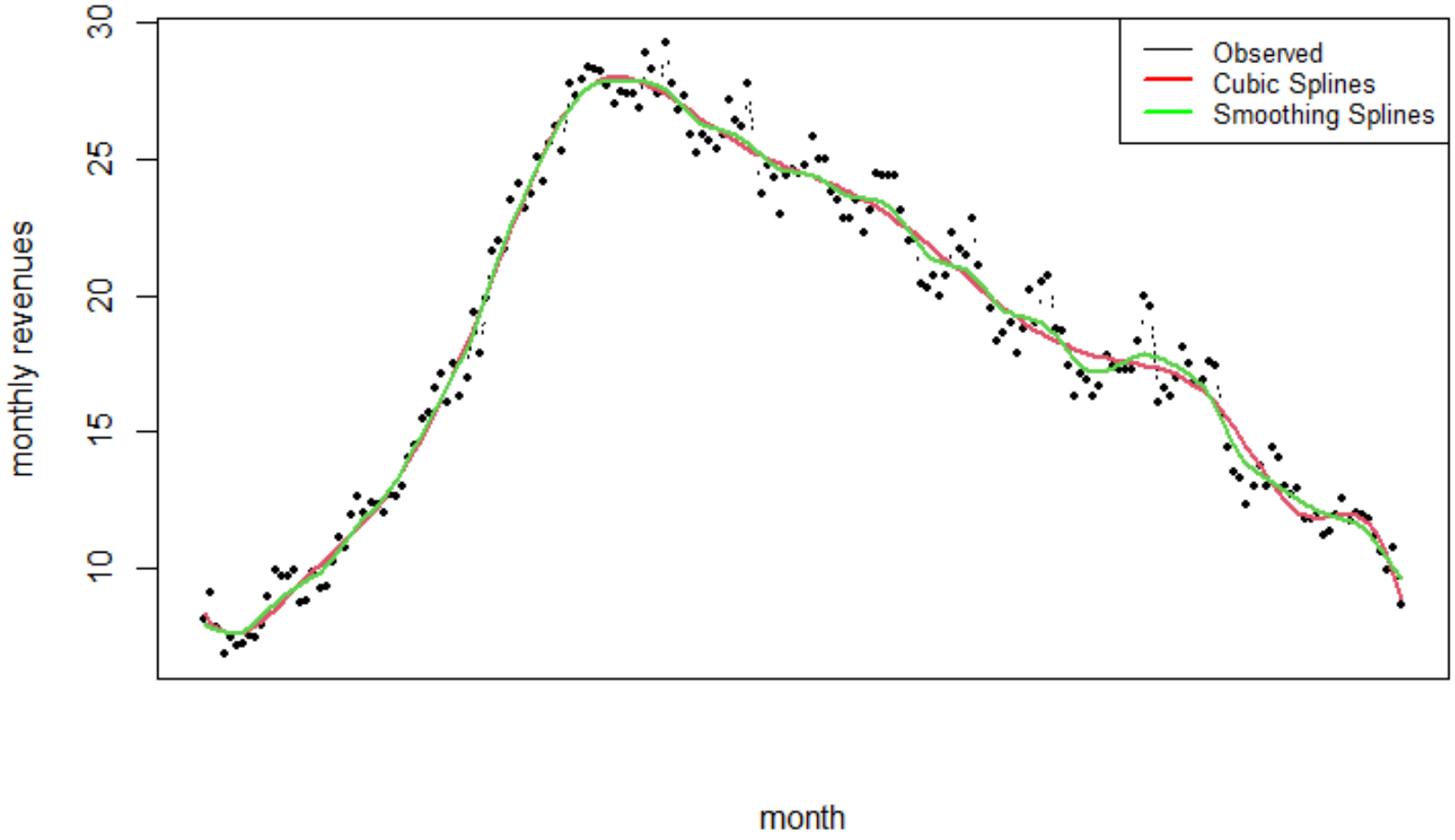

```
        Durbin-Watson test

data:  b2
DW = 0.14136, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

Linear regression (y ~ t + (t^2) + season)

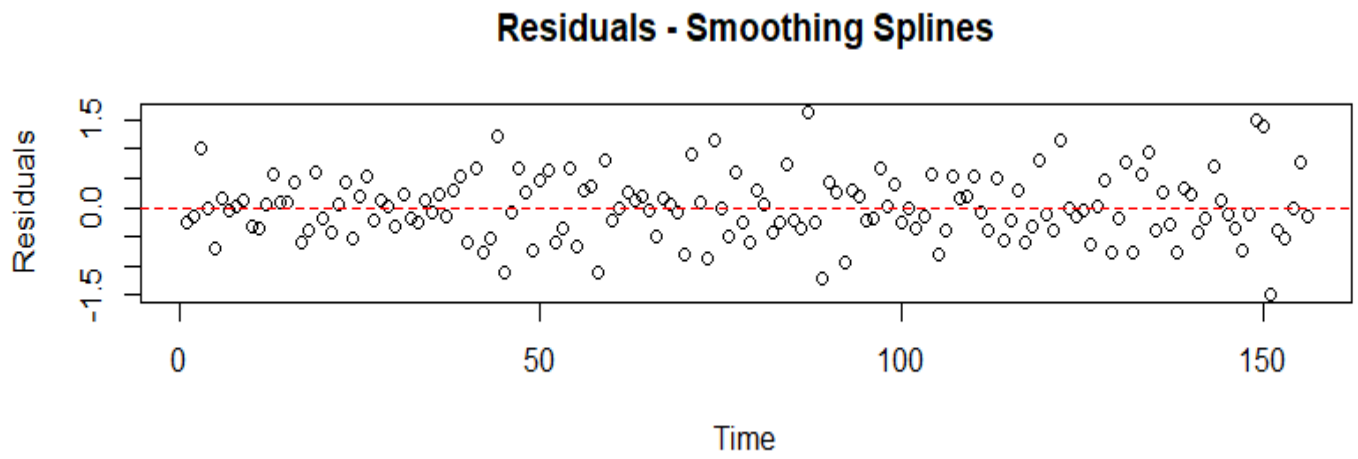


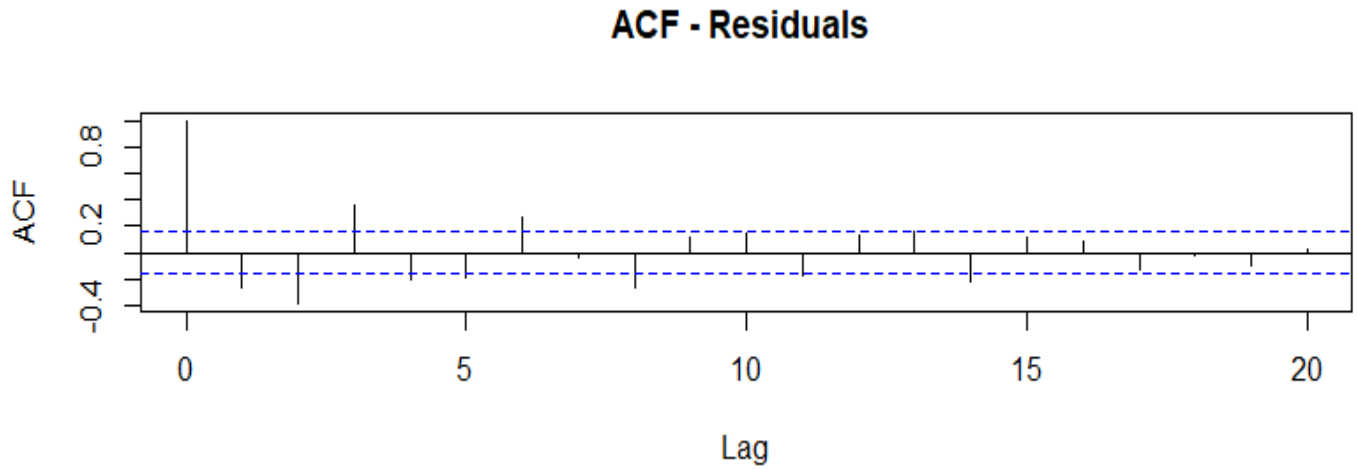
splines

- **Motivation:** flexible and can capture complex patterns with the ability to balance flexibility and overfitting.
- cubic splines
- Smoothing splines



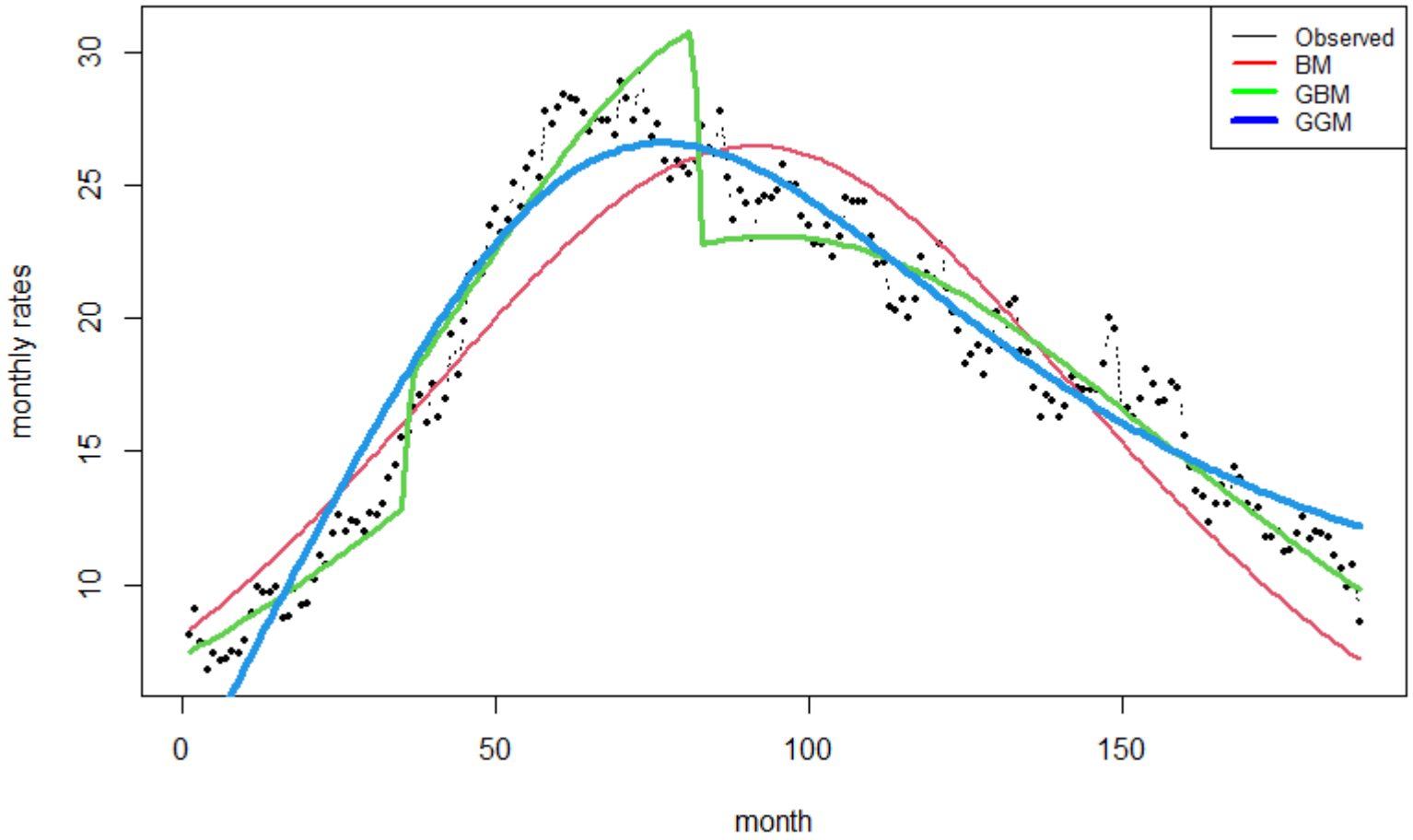
Smoothing Splines





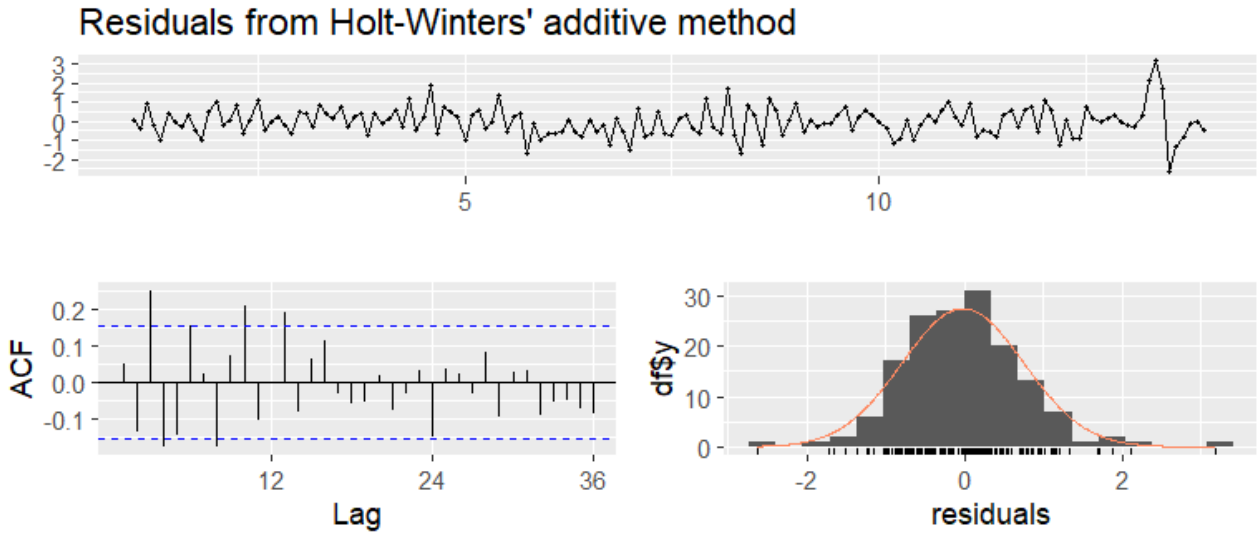
Diffusion Models

- **Motivation:** Check if there is shock after the 2008 economic crisis
- BM, GBM and GGM
- may not be good models because of their assumptions



Holt-Winters method (additive)

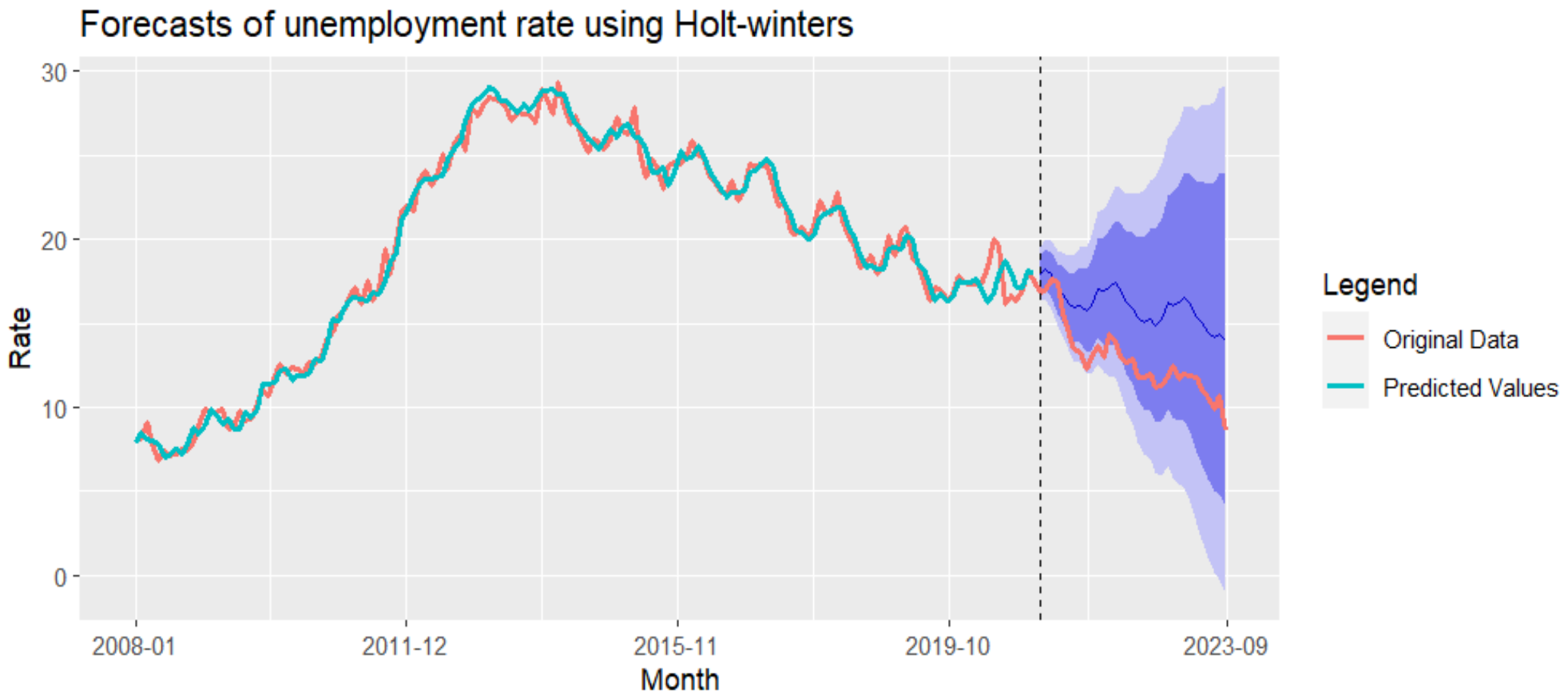
```
Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
 hw(y = Grr1, h = 33, seasonal = "additive")

  Smoothing parameters:
    alpha = 0.4342
    beta  = 0.0673
    gamma = 1e-04

  Initial states:
    l = 7.0654
    b = 0.1485
    s = 0.5028 0.5798 -0.5113 -0.9162 -0.6318 -0.8902
           -0.6725 -0.1957 0.0649 0.8199 1.1145 0.7358

  sigma:  0.8143

     AIC     AICc      BIC
740.8213 745.2561 792.6688

Error measures:
                     ME      RMSE       MAE          MPE     MAPE      MASE       ACF1
Training set -0.02093172 0.7714564 0.588549 -0.07968682 3.318622 0.2234996 0.04915017
```



Motivation:

- capture both the trend and seasonality
- Adaptability to Changing Patterns
- Forecasting Accuracy

Holt-Winters method (additive)



Adding Features

GDP

- Gross Domestic Product normalized for Greece, Index
- Monthly

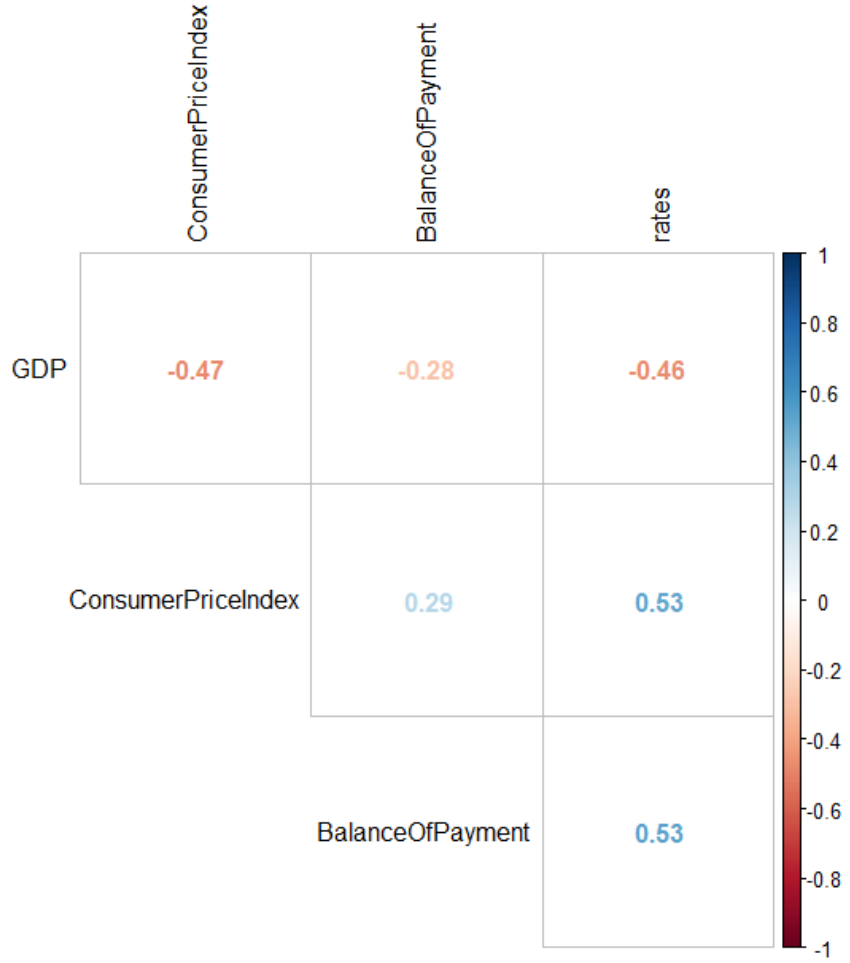
Consumer Price Index

- All Items, Total for Greece
- Index 2015=100,
- Monthly

Balance Of Payment

- Currency: Million euro
- Total economy
- Monthly
- Partner: Rest of the world

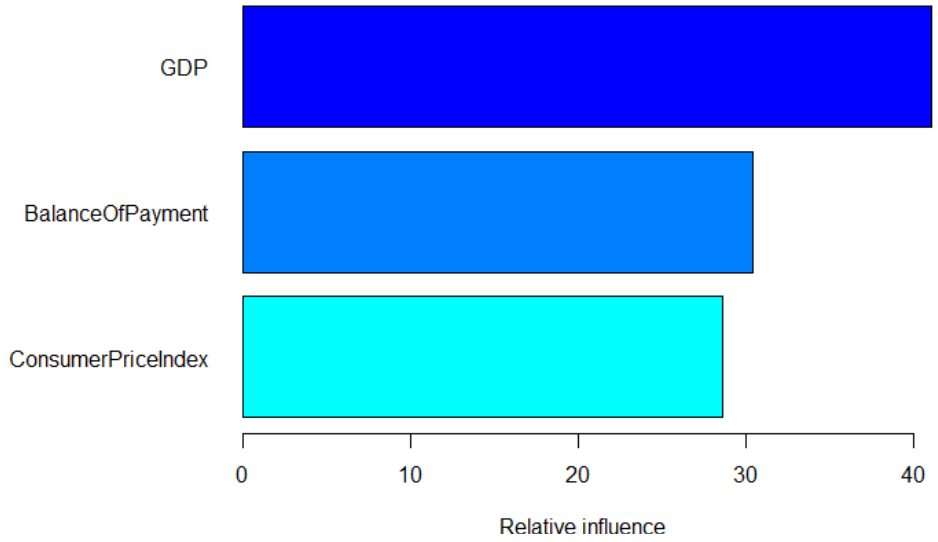
Correlation Matrix

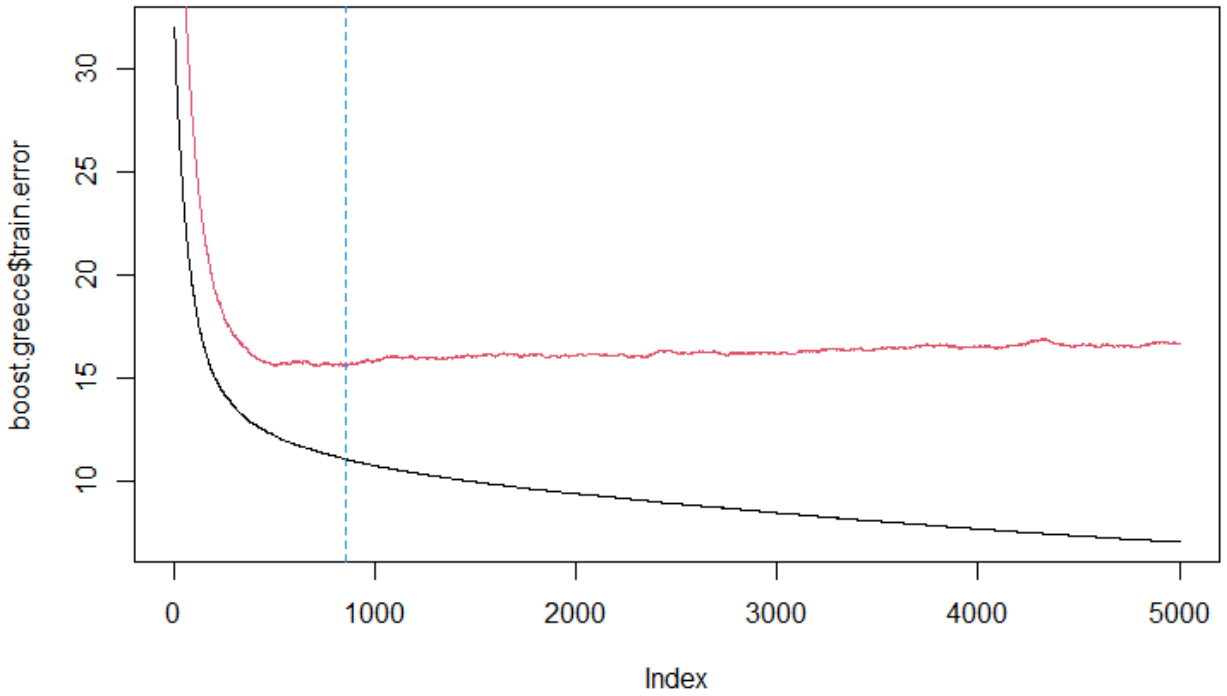


Response variable:
rates

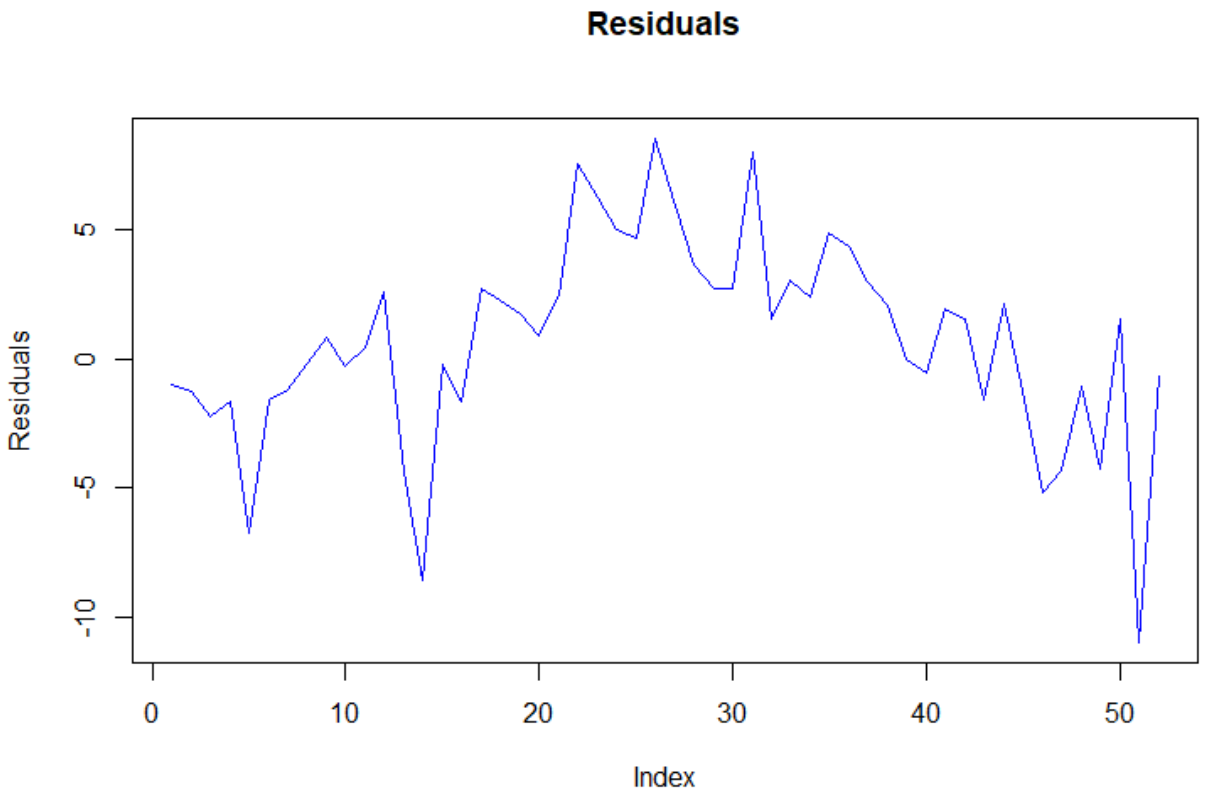
Gradient Boosting

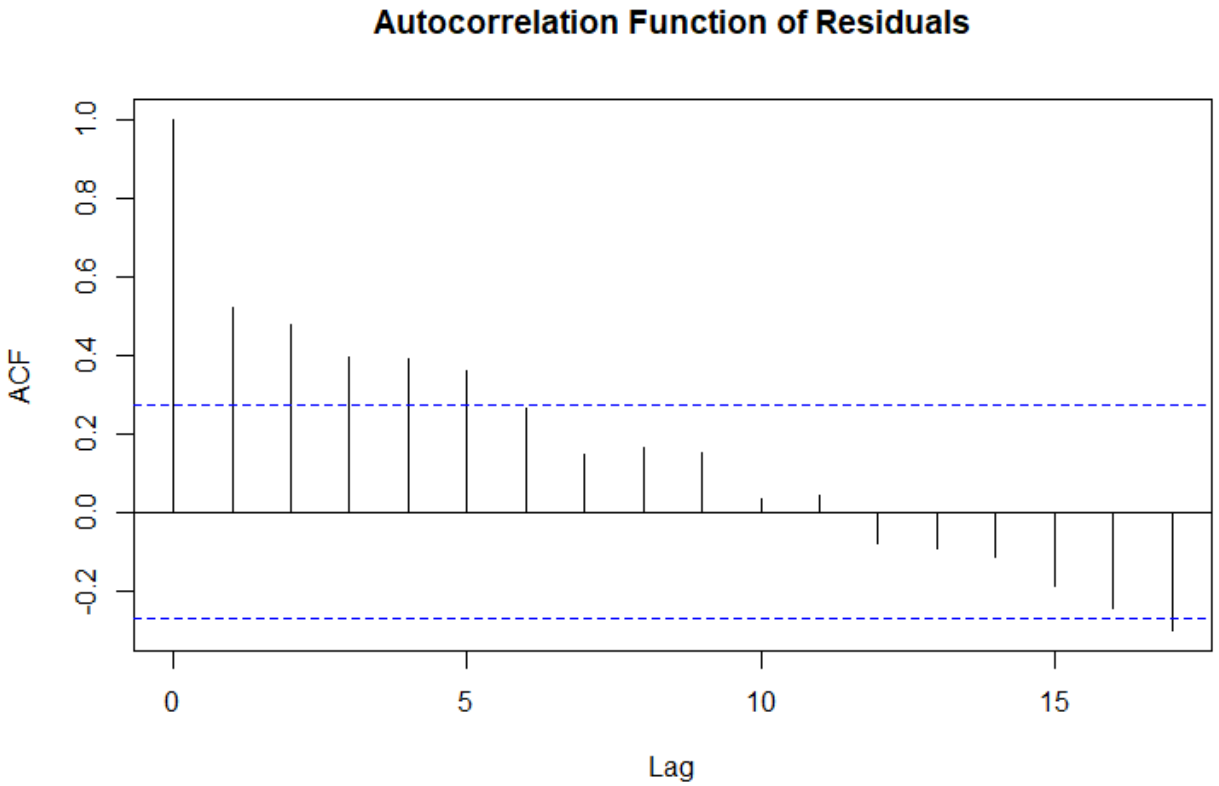
- **Motivations:** provide insights into feature importance, Combine multiple weaker models, non-linear relationships
- No cross validation
- rates ~ .-Greece, data=data.train, n.trees=5000, interaction.depth=1, shrinkage=0.01)
- minimum error: 15.56





Results


Residuals
Residuals
5
0
-5
-10
0
10
20
30
40
50
Index

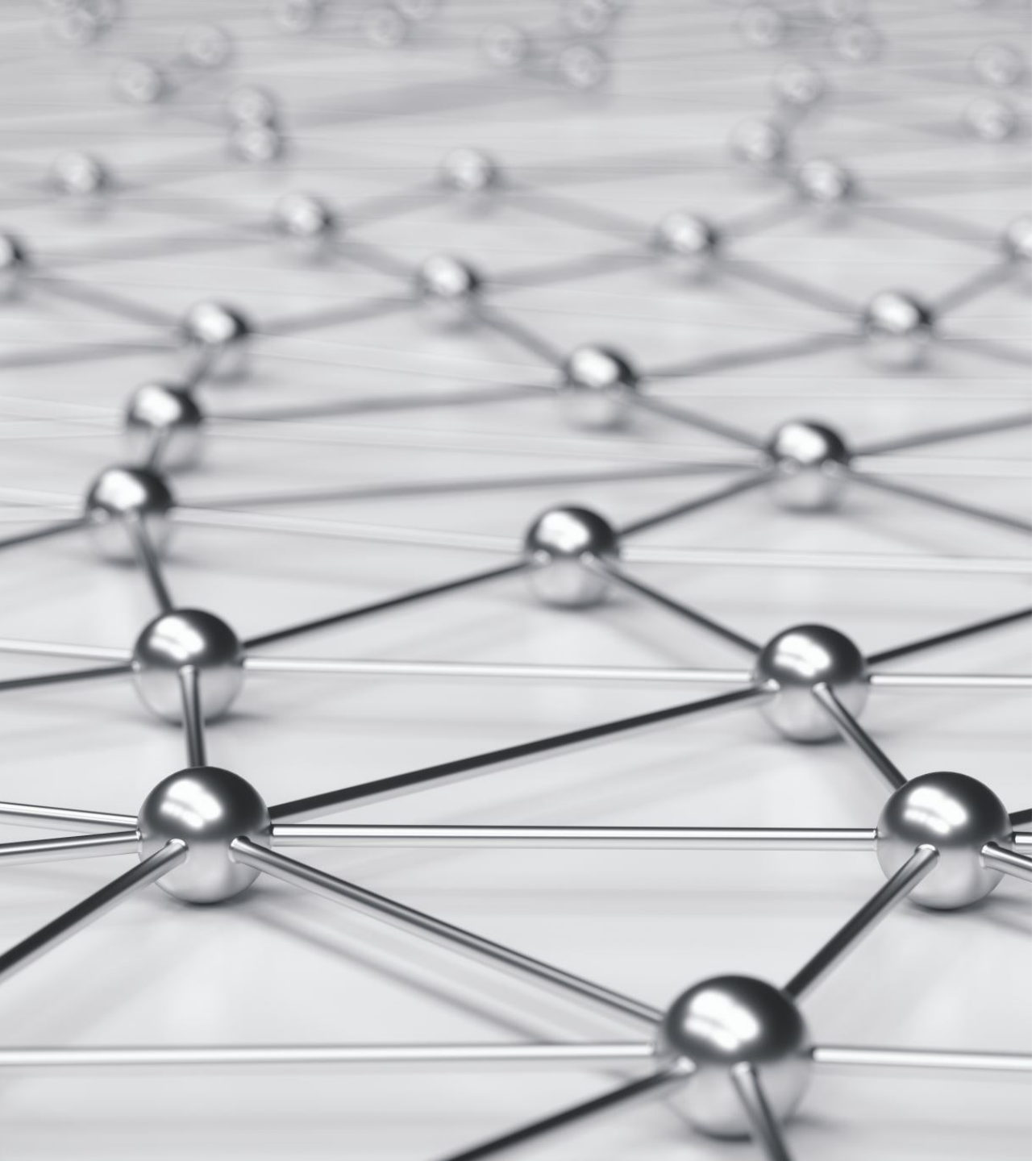

Autocorrelation Function of Residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

Comparing Models for Greece Via Metrics

Models		RMSE	MAPE	AIC	BIC
TSLM	Train set	2.27	11.31	728.73	774.47
	Test set	7.82	60.66		
Smoothing Splines	Data set	0.79	3.84	455.07	464.78
Holt-Winters method(additive)	Train set	0.77	3.32	740.82	792.66
	Test set	3.4	26.63		
Gradient Boosting	Train set	3.32	15.06	-	-
	Test set	3.94	17.8		

Conclusions and Future Works

- After 2008 we have the most variations in unemployment rates in Europe countries
- Holt-Winters methods for Germany and Greece and SARIMA for Italy seems to be the best models for forecasting
- The Greece time series is less regular and predictable
- The other predictors found for Greece seems to be relevant and effective
- Further analysis can exploit the usage of other predictors in the data



Thank you