Forecasting

Individual Assignment – Eduardo Razo

IÉSEG SCHOOL OF MANAGEMENT

May 1st, 2020

Forecasting

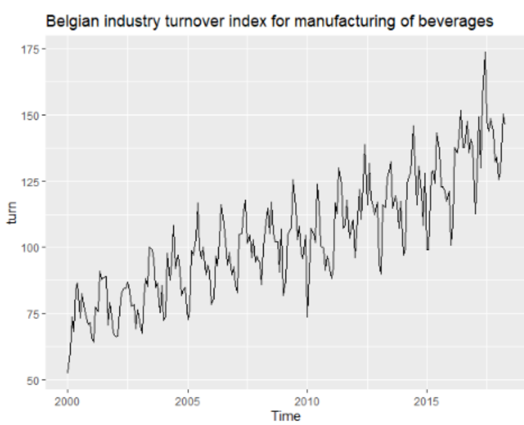
Exercise 1

The purpose of this document is to describe the application of forecasting models and the corresponding evaluation of them. For this forecasting task, we will be using the data set **Turnover** (provided in the class materials) which contains the Belgian industry turnover index for manufacturing of beverages from January 2000 to January 2018.

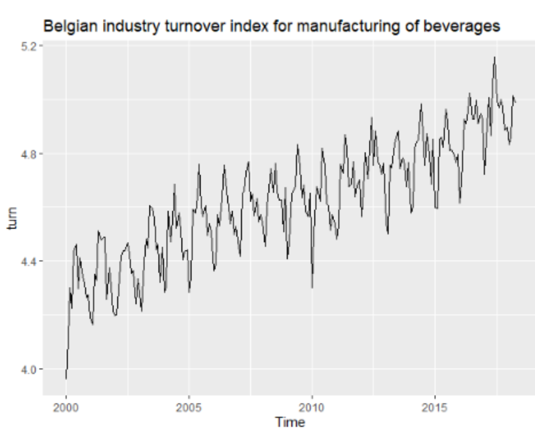
The forecasting process will consist of three main phases such as exploratory analysis, fitting the models and evaluating them.

Exploratory analysis

By applying **autoplot ()** on the timeseries information, I obtained the following graph:

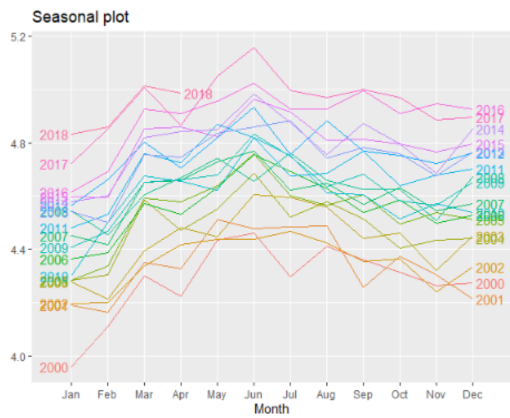


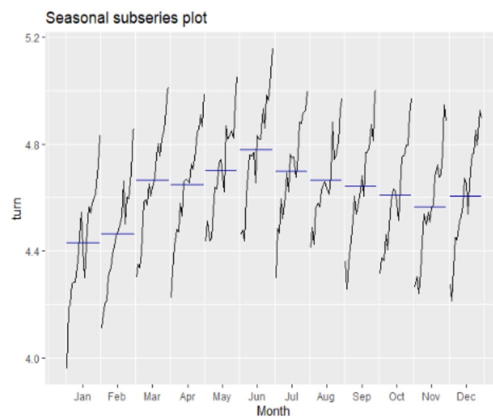
It is possible to observe that there is no stationarity, instead, there is a clear increasing trend and a seasonal pattern as well. The seasonality variation increases slightly as the level of the series increases, therefore it is a good idea log transforming the time series data, so I proceeded to this transformation using the function **log** (). In the following graph, we observe the behavior of the data after the log transformation.



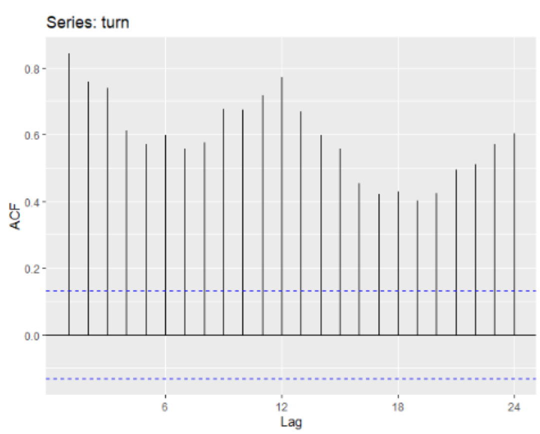
Now, the time series variations look more compact and the changes can be noticed mainly after year 2010.

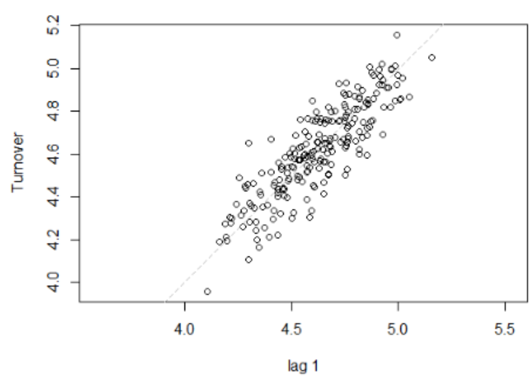
In the following graphs, we can see the seasonality component in a more detailed way in each of the years.



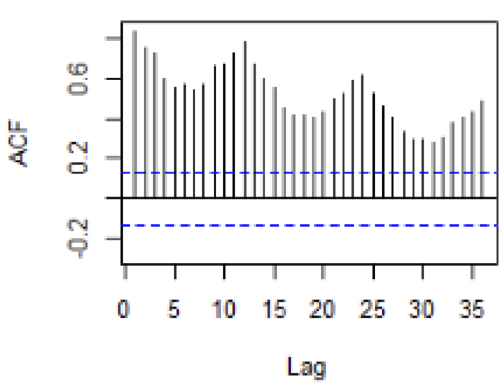


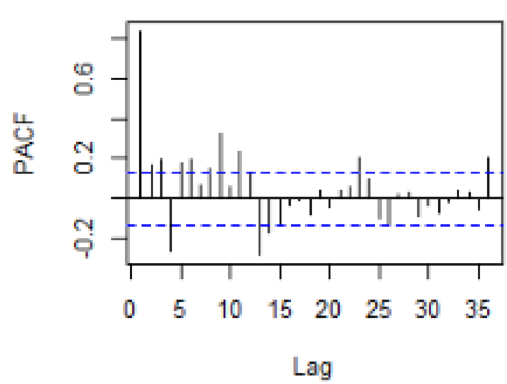
In addition, if we explore autocorrelation (AFC), the graphs below show that there is a strong autocorrelation and also a strong component of seasonality and that the greatest autocorrelation values occur at lags 1, 2, 3, 11 and 12, which are above of 0.70. From this information, we can start wondering about the lack of white noise.





As complement, if we look a partial autocorrelation (PACF), we confirm the absence of white noise.



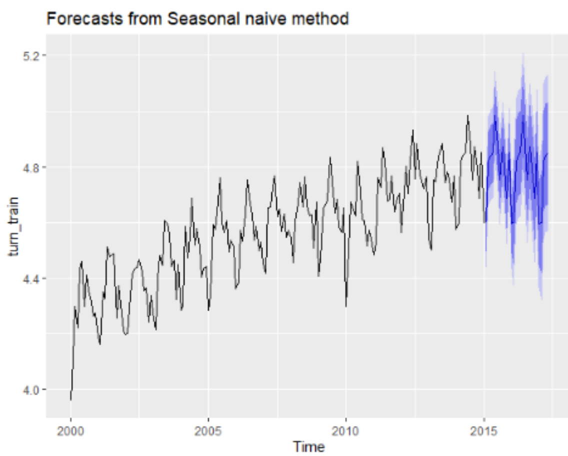


Fitting the models

Seasonal Naïve Method

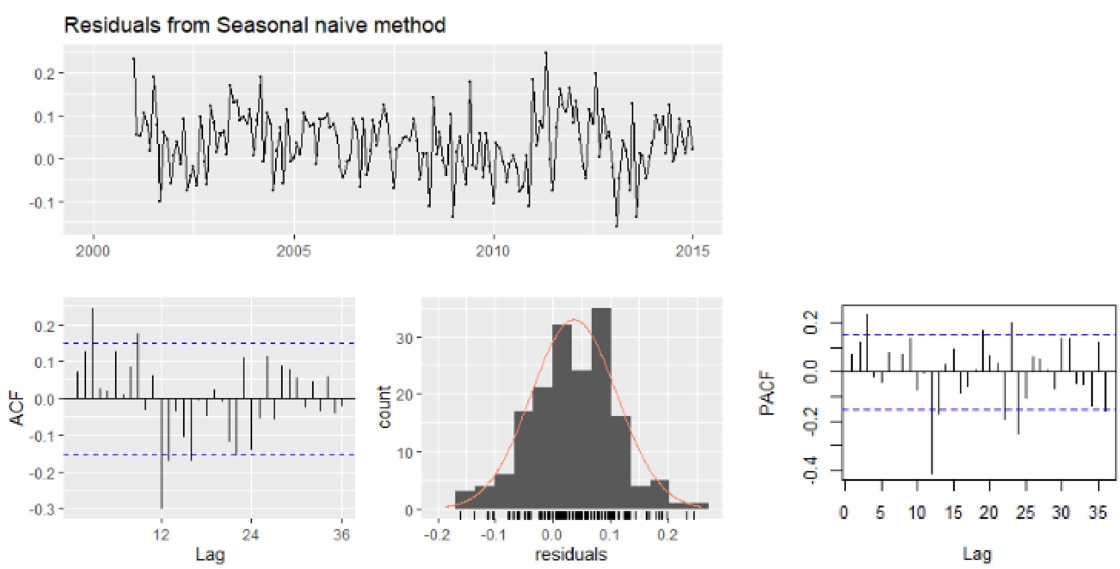
As initial step, the data was first divided into train and test. The training set contains information from January 2000 to December 2015 and the testing set from January 2016 to December 2020. To the extent, in the next models we will also use this data partitioning for evaluation purposes.

The seasonal naïve model was fitted on the training set and the evaluation on both training and testing set. The results are described as follow:



Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.



From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

lags	statistic	df	p-value
1	0.8878892	1	3.460499e-01
2	3.6799844	2	1.588187e-01
3	14.1728357	3	2.679088e-03
4	14.2689768	4	6.484273e-03
5	14.3171250	5	1.371566e-02
6	17.2018379	6	8.569394e-03
7	17.2142973	7	1.606540e-02
8	18.4833344	8	1.788084e-02
9	24.0830521	9	4.171920e-03
10	24.3138741	10	6.809463e-03
11	24.9674384	11	9.217258e-03
12	41.7416869	12	3.680779e-05
13	47.2925279	13	8.622786e-06
14	47.5691767	14	1.548802e-05
15	49.6478084	15	1.374510e-05
16	55.2569609	16	3.226485e-06

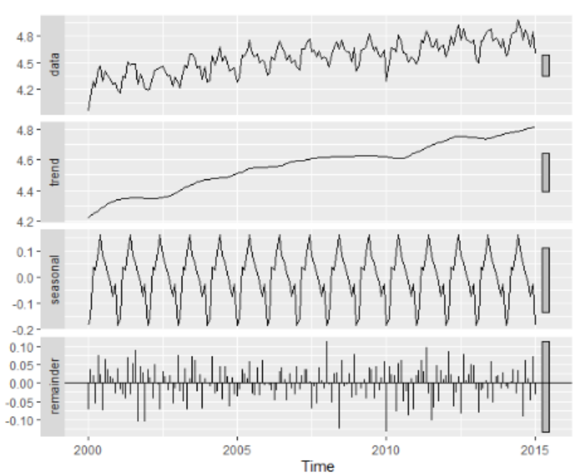
Box-Ljung test
p-value = 2.255e-06

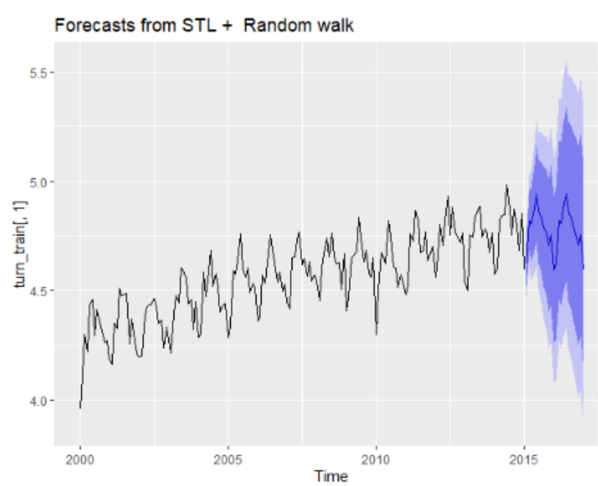
In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.03627381	0.08243054	0.06575223	0.7853386	1.435412	1.000000	0.07184435
Test set	0.11837251	0.13861018	0.11837251	2.4074041	2.407404	1.800281	-0.04209979

Seasonal Naïve Method with STL decomposition

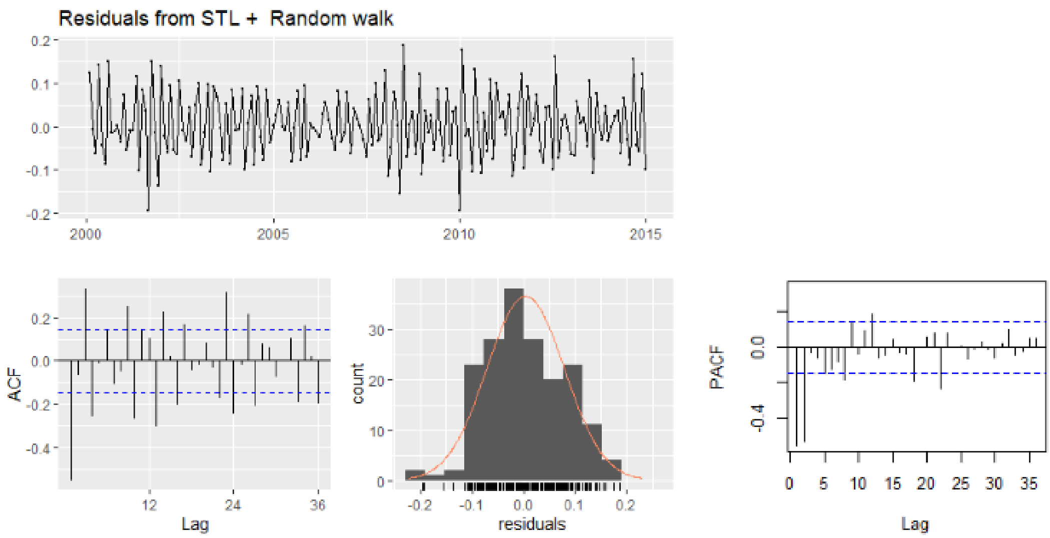
The seasonal naïve model with STL decomposition was fitted on the training set and the evaluation on both training and testing set. The results are described as follow:





Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.



From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

```
lags statistic df       p-value
   1  55.53198  1 9.192647e-14
   2  56.35137  2 5.799805e-13
   3  76.78226  3 1.110223e-16
   4  89.21729  4 0.000000e+00
   5  89.24162  5 0.000000e+00
   6  93.10876  6 0.000000e+00
   7  95.39161  7 0.000000e+00
   8  95.92264  8 0.000000e+00
   9 108.15089  9 0.000000e+00
  10 122.13113 10 0.000000e+00
  11 126.29209 11 0.000000e+00
  12 128.27859 12 0.000000e+00
  13 146.47921 13 0.000000e+00
  14 156.28129 14 0.000000e+00
  15 156.34456 15 0.000000e+00
  16 164.62055 16 0.000000e+00
```

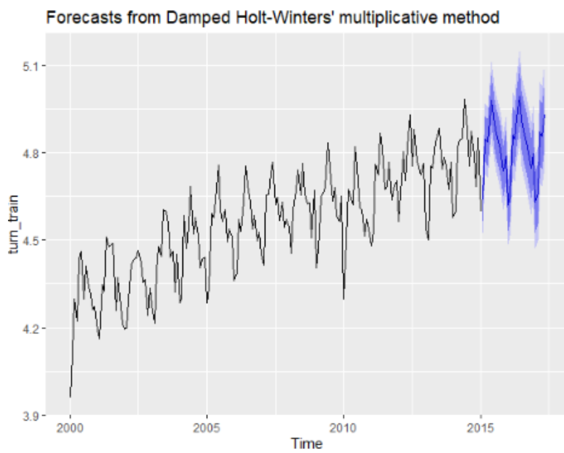
Box-Ljung test
p-value = 2.2e-16

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

```
                     ME       RMSE        MAE        MPE     MAPE      MASE       ACF1
Training set 0.00354455 0.07526845 0.06132697 0.06645455 1.347016 0.9326978 -0.5508409
Test set     0.11545703 0.12967191 0.11545703 2.35172863 2.351729 1.7559408  0.5052186
```

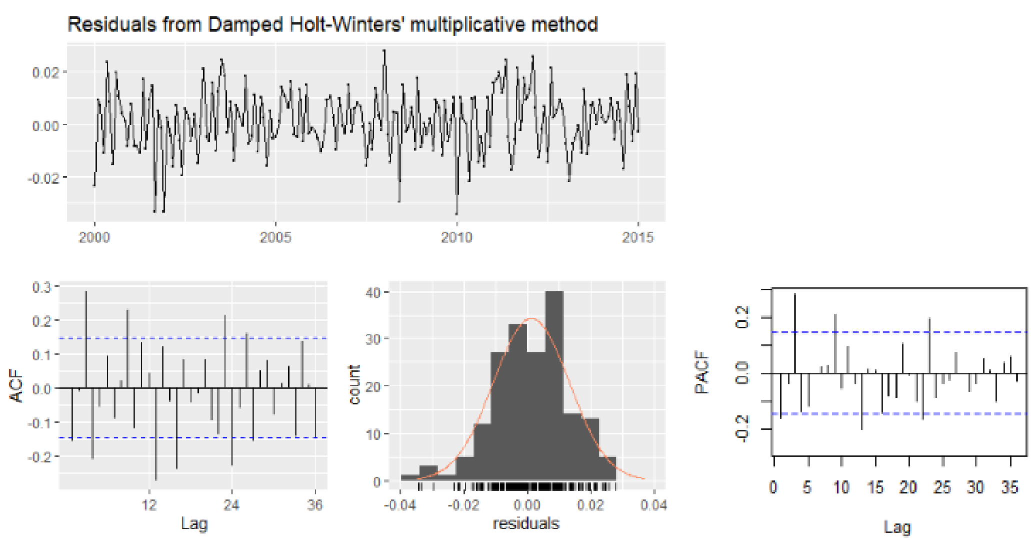
Holt-Winter´s method

The Holt-Winter´s model was fitted on the training set and the evaluation on both training and testing set. The results are described as follow:



Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.



From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

```
lags  statistic df p-value
   1   4.541334  0      NA
   2   4.561136  0      NA
   3  19.494034  0      NA
   4  27.655164  0      NA
   5  28.262061  0      NA
   6  29.868133  0      NA
   7  31.462326  0      NA
   8  31.542708  0      NA
   9  41.606681  0      NA
  10  44.468485  0      NA
  11  47.953728  0      NA
  12  48.332548  0      NA
  13  62.911756  0      NA
  14  65.672873  0      NA
  15  66.010087  0      NA
  16  77.705577  0      NA
  17  79.051955  0       0
  18  79.430357  1       0
```

Box-Ljung test
p-value = 1.781e-12

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

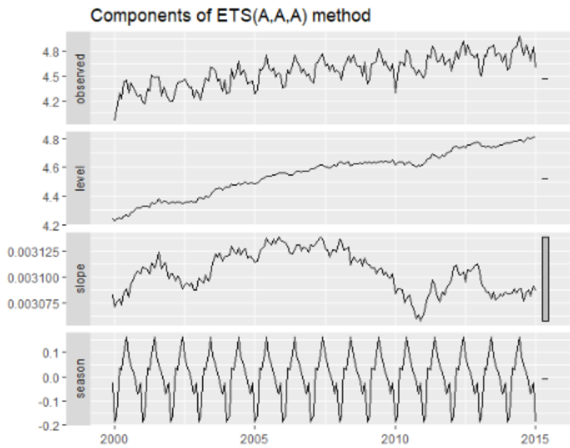
```
                      ME       RMSE        MAE       MPE     MAPE      MASE       ACF1
Training set 0.005276321 0.05419074 0.04378702 0.1004128 0.961077 0.6659397 -0.1563936
Test set     0.084502790 0.10487706 0.08520471 1.7156995 1.730911 1.2958451  0.3564325
```

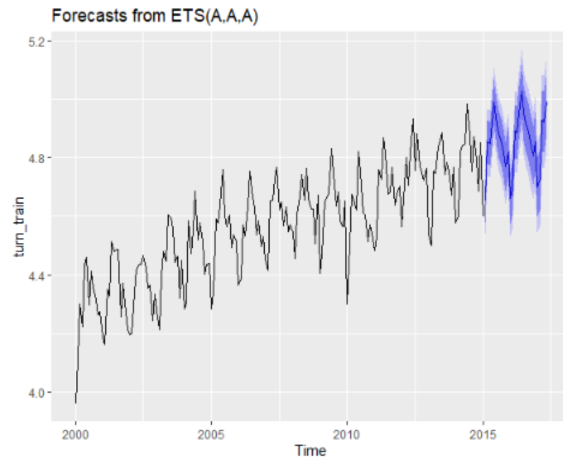
ETS method

In this case, before applying the automated ETS to find the optimal model to fit the data, some models were selected manually. First, from the 18 available models, those that do not incorporate the seasonality and trend components were excluded, also those ones that only consider a linear approach. Taking this into account the list of models used consists of:

- **ETS1 (A, N, A)**
- **ETS2 (A, N, Ad)**
- **ETS3 (A, A, A)**
- **ETS4 (A, A, Ad)**
- **ETS5 (M, N, A)**
- **ETS6 (M, N, Ad)**
- **ETS7 (M, A, A)**
- **ETS8 (M, A, Ad)**

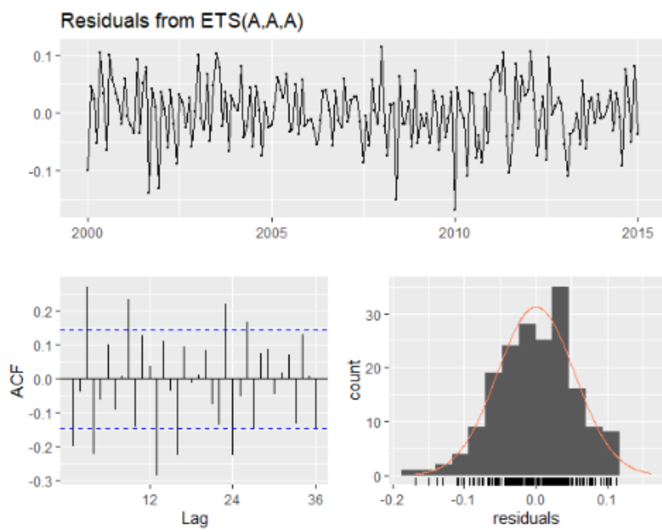
Regarding these models, the results obtained in terms of residuals and accuracy did not perform very well and for more detail I suggest referring to the code. For communication purposes, in this report I decided to include only the output of the model selected by the automated ETS method, which is **ETS(A, A, A)**. Results are described as follow:

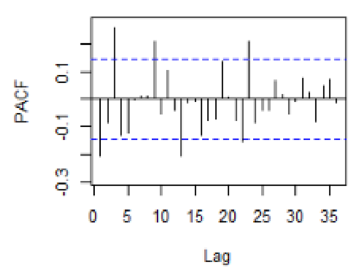




Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.





From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

lags	statistic	df	p-value
1	7.424094	0	NA
2	7.727553	0	NA
3	21.491831	0	NA
4	30.645384	0	NA
5	31.379941	0	NA
6	33.325735	0	NA
7	34.996703	0	NA
8	35.010545	0	NA
9	45.632593	0	NA
10	49.577054	0	NA
11	52.681872	0	NA
12	52.934049	0	NA
13	68.902725	0	NA
14	71.403098	0	NA
15	71.639475	0	NA
16	81.809391	0	0
17	83.541741	1	0
18	83.566955	2	0

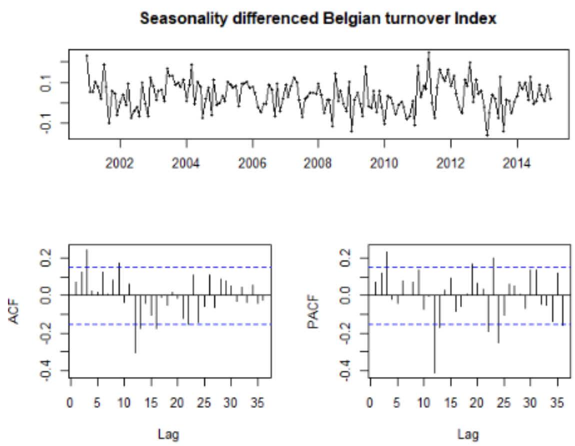
Box-Ljung test
p-value = 4.08e-13

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.0002207855	0.05397451	0.04376293	-0.004760354	0.9625227	0.6655732	-0.2008599
Test set	0.0353226473	0.06584905	0.05157446	0.710721192	1.0498163	0.7843758	0.1936066

Seasonal ARIMA method

In this case as well, before applying the auto ARIMA to find the optimal model to fit the data, some models were selected manually. As first step, in order to add a component of stationarity and help to stabilize the mean, I computed the first difference or variation in the observations and the output was plotted.

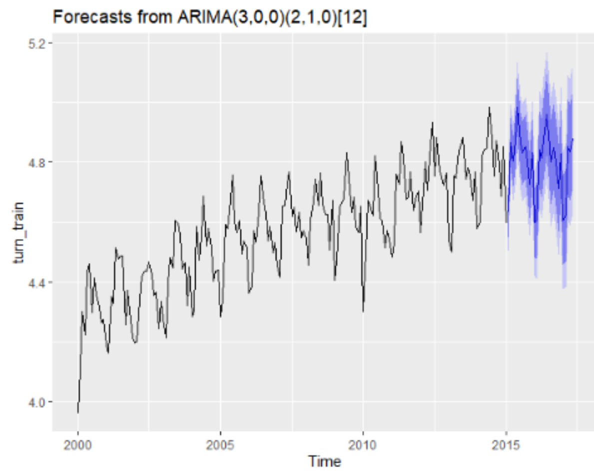


Based on the previous results, we can still identify some significant peaks in the PACF graph, so in order to address this situation, some models were suggested manually adjusting the elements in the trending and seasonal components (P, D, Q). Taking this into consideration the list of initial ARIMA models used consists of:

- **ARIMA (1, 0, 1) (2, 1, 1)$_{12}$**
- **ARIMA (3, 0, 1) (0, 1, 1)$_{12}$**
- **ARIMA (3, 0, 1) (3, 1, 1)$_{12}$**
- **ARIMA (3, 1, 1) (3, 1, 1)$_{12}$**
- **ARIMA (3, 0, 3) (3, 1, 1)$_{12}$**
- **ARIMA (3, 0, 3) (3, 1, 3)$_{12}$**

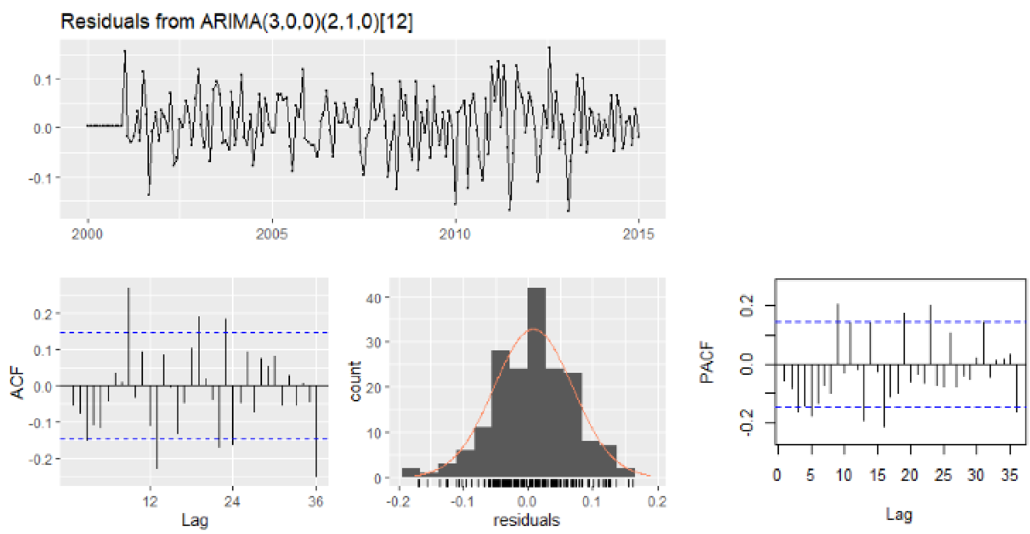
Regarding these models, the results obtained in terms of residuals and accuracy did not perform very well and for more detail I suggest referring to the code. For communication purposes, in this report I decided to include only the output of the model selected by the auto ARIMA method, which is **ARIMA (3, 0, 0) (2, 1, 0)$_{12}$**. Results are described as follow:

```
Coefficients:
         ar1     ar2     ar3     sar1     sar2
      0.1057  0.2602  0.5117  -0.5750  -0.3534
s.e.  0.0684  0.0658  0.0715   0.0784   0.0778

sigma^2 estimated as 0.004156:  log likelihood=222.81
AIC=-433.63   AICc=-433.11   BIC=-414.85
```



Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.



From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

lags	statistic	df	p-value
1	0.5924811	0	NA
2	1.6675886	0	NA
3	6.0217380	0	NA
4	8.2339109	0	NA
5	10.8201391	0	0.000000e+00
6	11.1672517	1	8.325386e-04
7	11.3826858	2	3.375058e-03
8	11.4013687	3	9.742200e-03
9	25.4193765	4	4.142668e-05
10	25.6283485	5	1.053320e-04
11	27.3856680	6	1.225931e-04
12	29.8546095	7	1.009709e-04
13	40.2576285	8	2.868411e-06
14	41.6967635	9	3.733648e-06
15	41.7045160	10	8.462953e-06
16	45.2373287	11	4.407341e-06

Box-Ljung test
p-value = 3.612e-07

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

```
                     ME       RMSE        MAE        MPE      MAPE      MASE        ACF1
Training set 0.007990924 0.0613624 0.0477721 0.1668333 1.042377 0.7265472 -0.05674253
Test set     0.110706430 0.1262979 0.1107064 2.2517306 2.251731 1.6836908  0.07429014
```

Choosing a model

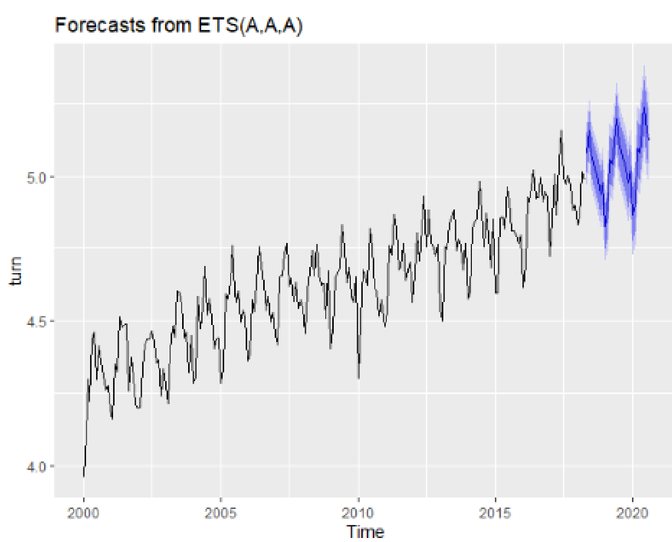
After running all the models previously described and considering the evaluation metric, it is possible to conclude the following:

- After analyzing residuals graphically and using Ljung Box, all the models failed this test. Despite all models failed the residual test, **ETS (A, A, A)** model has the best performance in terms of **RMSE**, therefore, we are in position of selecting this model as final.

As support and summary of performance, the chart below describes the metrics used in the evaluation phase.

	SNAIVE	SNAIVE STL	Holt-Winters'	Auto ETS	Auto ARIMA
P-value	2.255e-06	2.2e-16	1.781e-12	4.08e-13	3.612e-07
RMSE	0.138610	0.129671	0.104877	0.065849	0.126297
MAE	0.118372	.115457	0.085204	0.051574	0.110706
AIC	-	-	-	-81.69056	-433.63
AICc	-	-	-	-77.93596	-433.11

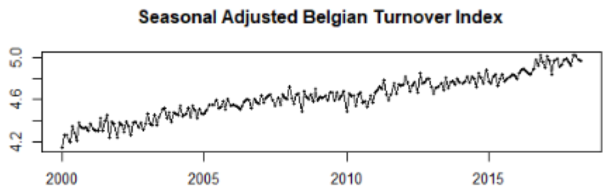
Application of the best forecasting model over the complete data.

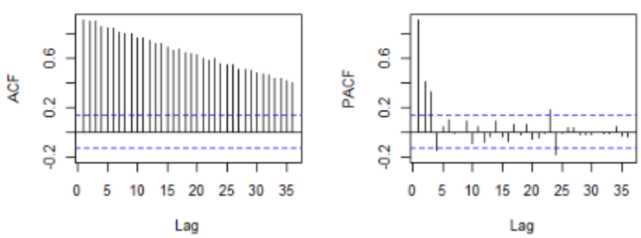


ARIMA vs ETS

In this section of the report, the purpose is to compare the performance of ARIMA model vs ETS model on seasonally adjusted time series.

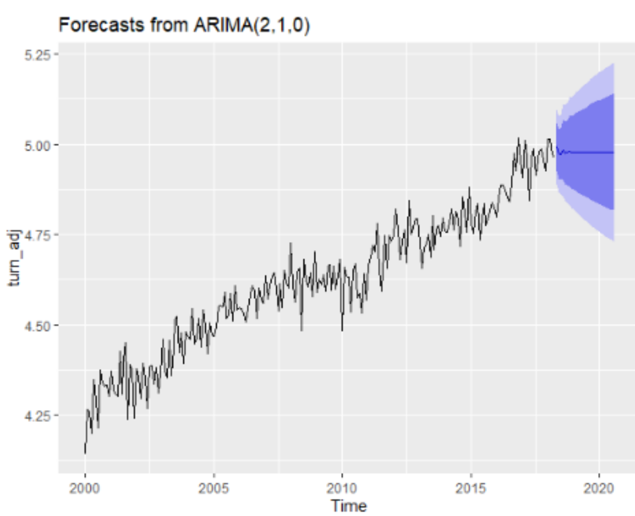
In order to eliminate the effect of seasonality on the data, functions **stl()** and **seasadj()** were applied and from there, we can analyze how data looks like.





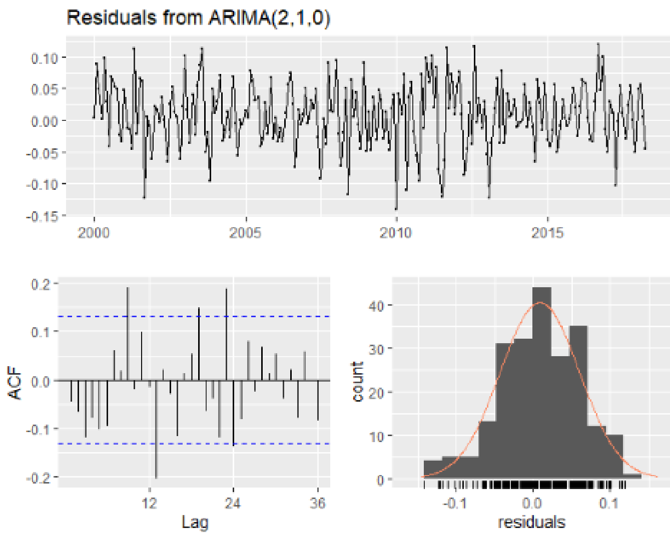
Next, non-seasonal auto ARIMA model was used and it came up with **ARIMA (2, 1, 0)** as the best model and with these parameters and forecast.

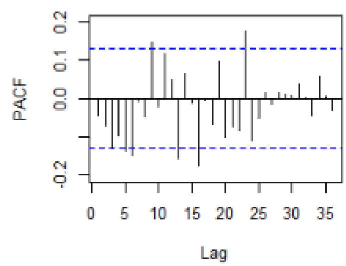
```
ARIMA(2,1,0)

Coefficients:
          ar1      ar2
      -0.8070  -0.5435
s.e.   0.0571   0.0569

sigma^2 estimated as 0.002738:  log likelihood=335.84
AIC=-665.69   AICc=-665.58   BIC=-655.52
```



Evaluation of non-seasonal ARIMA model

These charts show the behavior of the residuals and the accuracy of the model.





From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

lags	statistic	df	p-value
1	0.4815998	0	NA
2	1.5059979	0	0.000000e+00
3	4.7593430	1	2.913967e-02
4	6.1484866	2	4.622459e-02
5	8.5853136	3	3.534401e-02
6	10.6525037	4	3.075988e-02
7	11.4703519	5	4.281189e-02
8	11.5377046	6	7.311327e-02
9	20.0982879	7	5.361250e-03
10	20.1918865	8	9.633954e-03
11	22.4693496	9	7.504726e-03
12	22.5260429	10	1.263790e-02
13	32.3928748	11	6.594886e-04
14	32.4977424	12	1.158893e-03
15	32.7018377	13	1.888595e-03
16	36.0437197	14	1.027832e-03
17	36.0839559	15	1.719159e-03
18	36.7719495	16	2.257687e-03

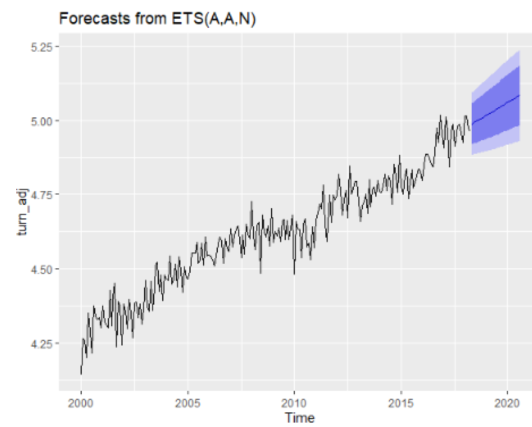
Box-Ljung test
p-value = 4.829e-05

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.008588409	0.05197118	0.04155871	0.1781071	0.9008767	0.6259177

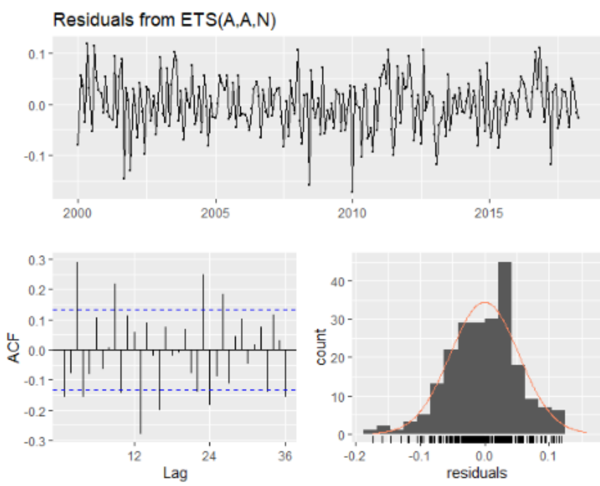
Finally, auto ETS model was used and it came up with **ETS (A, A, N)** as the best model and with these parameters and forecast.

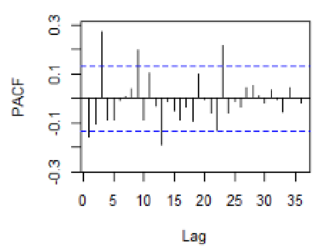
```
ETS(A,A,N)

Call:
 ets(y = turn_adj)

  Smoothing parameters:
    alpha = 0.2049
    beta  = 1e-04

  Initial states:
    l = 4.2178
    b = 0.0036

  sigma:  0.0534

     AIC      AICc       BIC
-96.45939 -96.17902 -79.49125
```



Evaluation of non-seasonal ETS model

These charts show the behavior of the residuals and the accuracy of the model.





From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

lags	statistic	df	p-value
1	5.454673	0	NA
2	6.789921	0	NA
3	25.881173	0	NA
4	31.564378	0	0.000000e+00
5	33.100404	1	8.752052e-09
6	35.685243	2	1.782575e-08
7	36.663846	3	5.420021e-08
8	36.678985	4	2.097513e-07
9	47.756639	5	3.982289e-09
10	52.608317	6	1.406886e-09
11	55.536925	7	1.166861e-09
12	56.343829	8	2.417684e-09
13	74.651932	9	1.851186e-12
14	76.479028	10	2.450262e-12
15	76.596446	11	6.689871e-12
16	86.289403	12	2.570166e-13
17	87.648283	13	3.929079e-13
18	87.751982	14	1.007083e-12

Box-Ljung test
p-value = 1.41e-14

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

```
                        ME       RMSE        MAE         MPE      MAPE      MASE       ACF1
Training set -0.0004733927 0.05293096 0.04232005 -0.02050358 0.9214371 0.6373843 -0.1563935
```

In order to compare both models easily, the chart below shoes the results of each method.

	ARIMA	ETS
P-value	4.829e-05	1.41e-14
RMSE	0.051971	0.052930
MAE	0.041558	0.042320
AIC	-665.69	-96.4593
AICc	-665.58	-96.1790

As conclusion for this exercise, even though the models used failed the residuals test and based on the accuracy of them, we can select non-seasonal **ARIMA (2, 1, 0)** as the final model since it is the one with better RMSE.

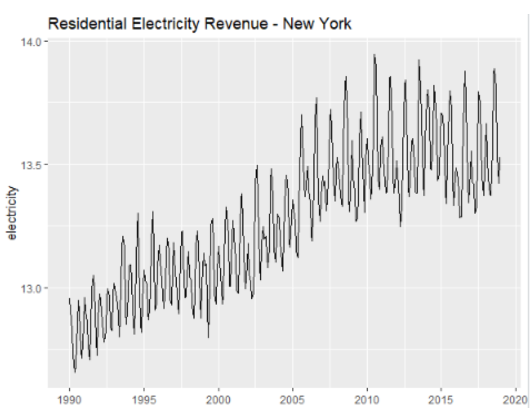
Exercise 2

For this exercise, the data analyzed and forecasted, corresponds to the monthly revenue of residential electricity consumption in New York, USA. The dataset contains information from January 1990 to December 2018.

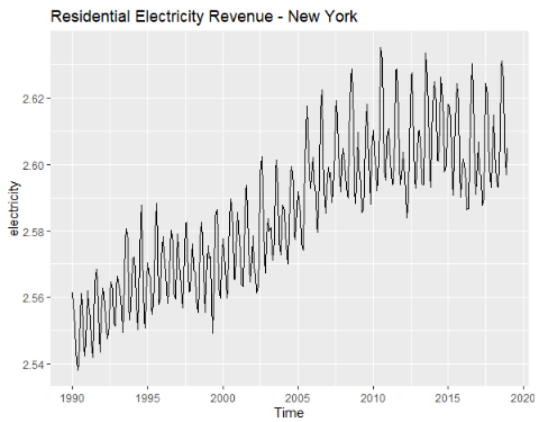
The forecasting process will consist of three main phases such as exploratory analysis, fitting the models and evaluating them.

Exploratory analysis

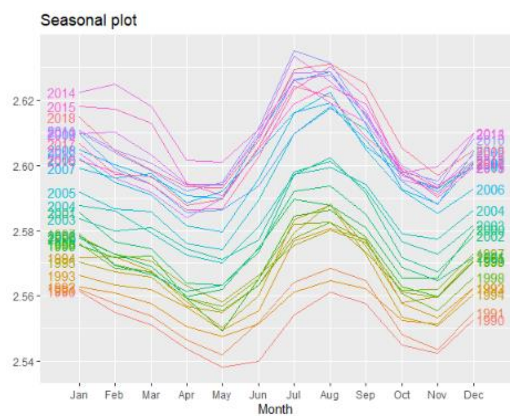
By applying **autoplot ()** on the timeseries information, I obtained the following graph:

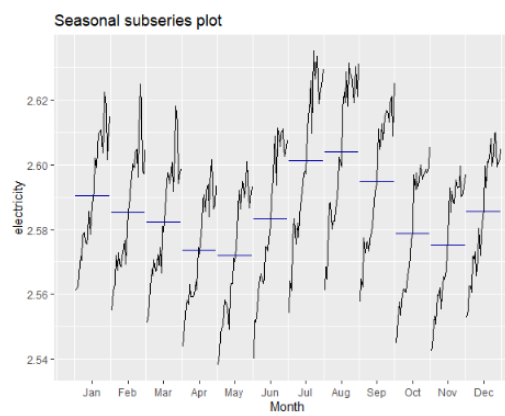


It is possible to observe that there is no stationarity, instead, there is a clear increasing trend and a strong seasonal pattern as well. The seasonality variation increases slightly as the level of the series increases, therefore it is a good idea log transforming the time series data, so I proceeded to this transformation using the function **log** (). In the following graph, we observe the behavior of the data after the log transformation.

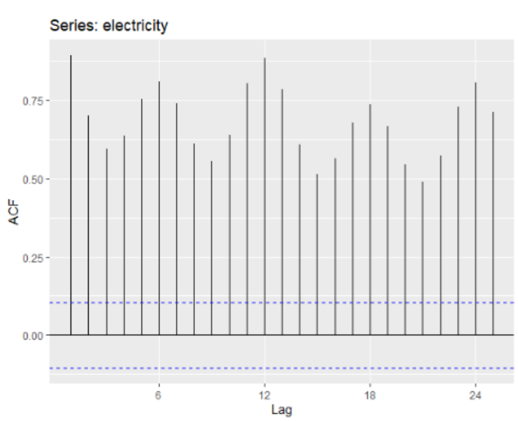


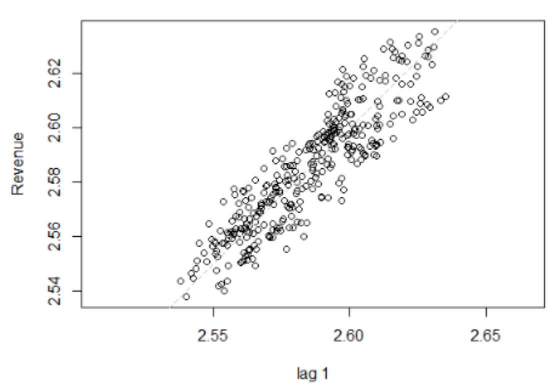
In the following graphs, we can see the seasonality component in a more detailed way in each of the years.



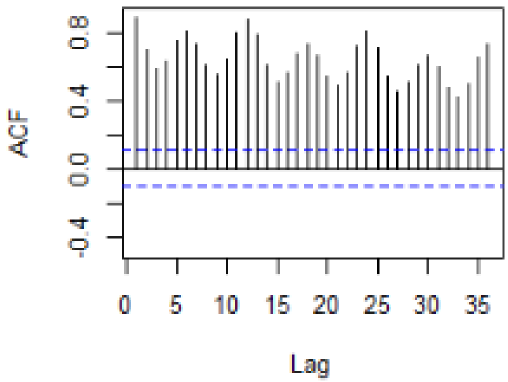


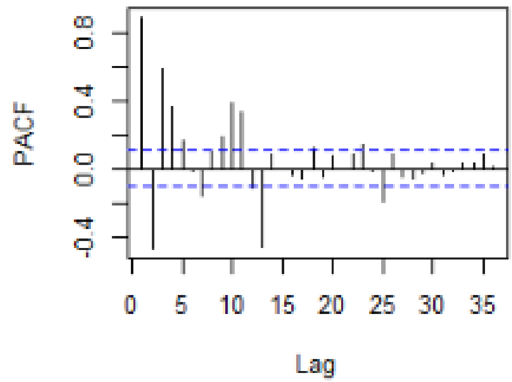
In addition, if we explore autocorrelation (AFC), the graphs below show that there is a strong autocorrelation and also a strong component of seasonality and that the greatest autocorrelation values occur at lags 1, 12, and 24, which are above of 0.75. From this information, we can start wondering about the lack of white noise.





As complement, if we look a partial autocorrelation (PACF), we confirm the absence of white noise.



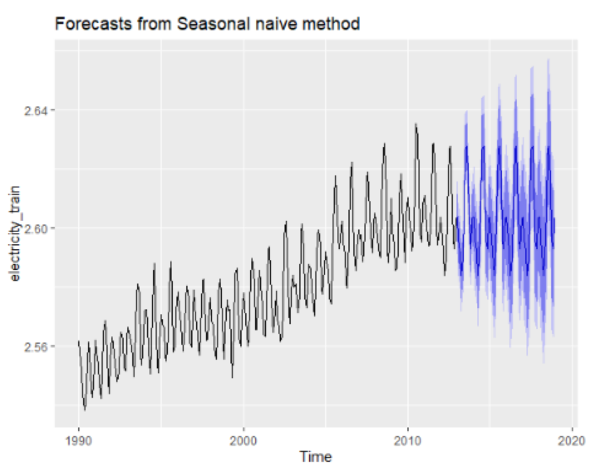


Fitting the models

Seasonal Naïve Method

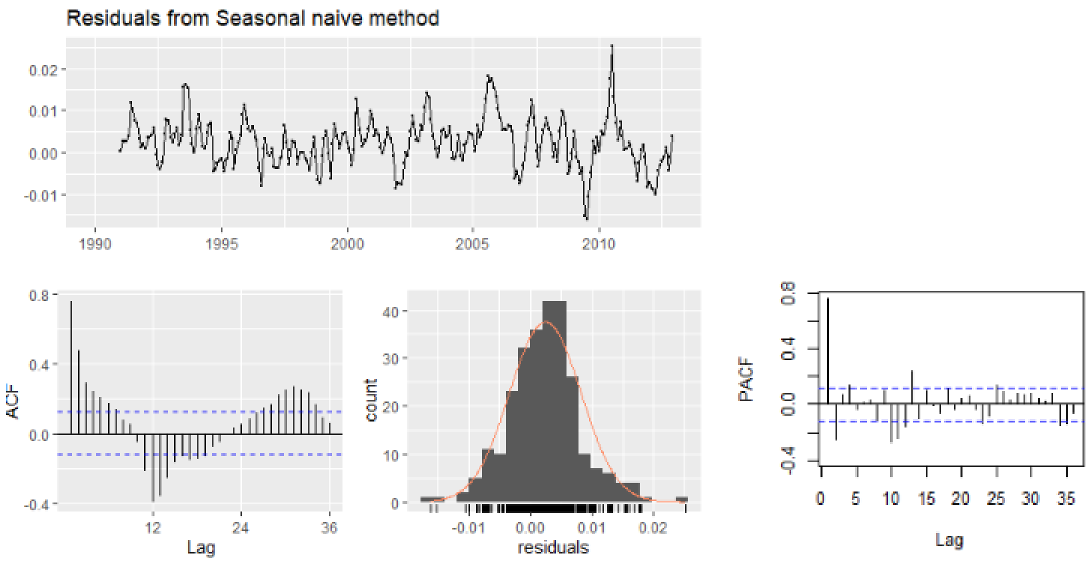
As initial step, the data was first divided into train and test. The training set contains information from January 1990 to December 2012 and the testing set from January 2013 to December 2018. To the extent, in the next models we will also use this data partitioning for evaluation purposes.

The seasonal naïve model was fitted on the training set and the evaluation on both training and testing set. The results are described as follow:



Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.



From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

```
lags statistic df p-value
   1  155.1978  1       0
   2  216.2457  2       0
   3  239.2047  3       0
   4  254.4382  4       0
   5  265.9913  5       0
   6  273.9637  6       0
   7  279.3312  7       0
   8  280.8490  8       0
   9  281.5753  9       0
  10  282.1896 10       0
  11  295.5291 11       0
  12  338.6355 12       0
  13  374.4971 13       0
  14  393.5351 14       0
  15  401.5417 15       0
  16  406.4043 16       0
```

Box-Ljung test
p-value = 2.2e-16

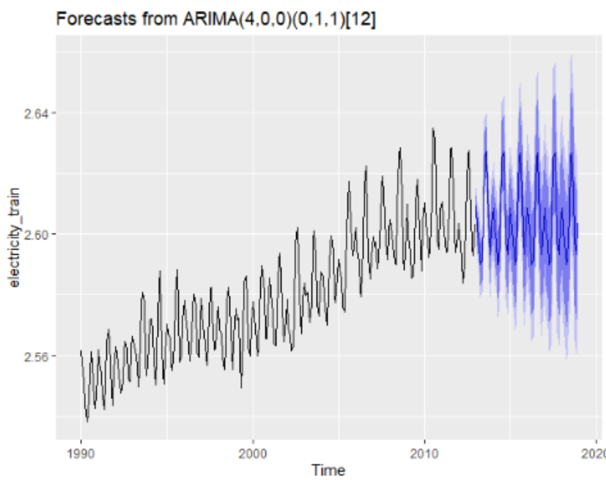
In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

```
                      ME        RMSE         MAE        MPE      MAPE     MASE      ACF1
Training set 0.002388141 0.006255172 0.004816424 0.09235956 0.1862804 1.000000 0.7623913
Test set     0.004124382 0.008510145 0.005996752 0.15801827 0.2297283 1.245063 0.7583867
```

Seasonal ARIMA method

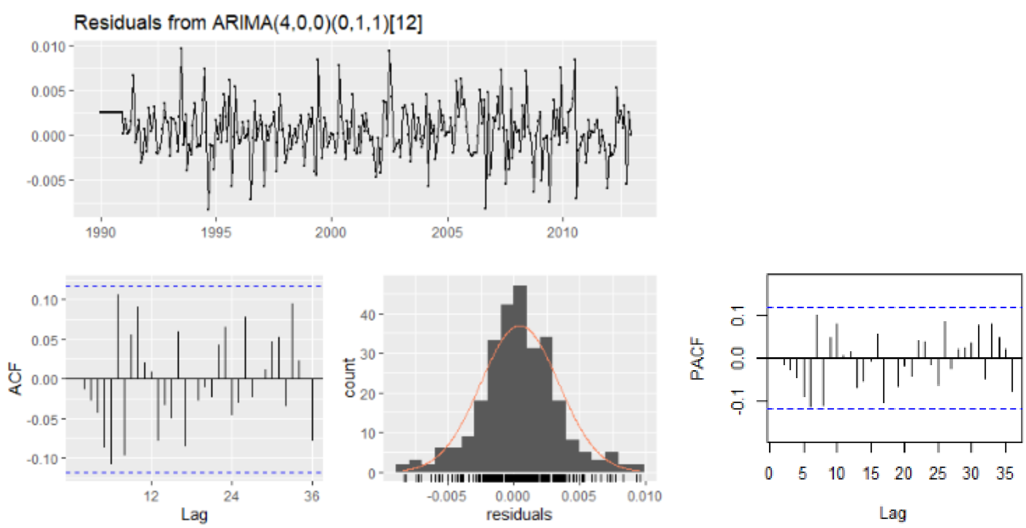
In this case, the auto ARIMA method was used and we ended up with **ARIMA (4, 0, 0) (0, 1, 1)$_{12}$** as the best model. Results are described as follow:

```
Coefficients:
         ar1      ar2      ar3     ar4     sma1
      1.0268  -0.2070  -0.0979  0.2430  -0.7406
s.e.  0.0607   0.0867   0.0877  0.0603   0.0493

sigma^2 estimated as 9.528e-06:  log likelihood=1152.93
AIC=-2293.87   AICc=-2293.54   BIC=-2272.41
```



Evaluation of the model

These charts show the behavior of the residuals and the accuracy of the model.



From the residuals visualizations above and the Ljung Box tests performed for this model, we can reject the null hypothesis of white noise since the parameters in the tests (p-values) are less than .05 which means that they are significant. that means there is still something in the residuals that is not been captured by the model in the data generating process and we still can learn more from the information.

```
lags     statistic df    p-value
   1 3.663396e-04  1 0.9847294
   2 5.625374e-02  2 0.9722650
   3 2.749327e-01  3 0.9646719
   4 8.004660e-01  4 0.9383856
   5 2.954789e+00  5 0.7069560
   6 6.304049e+00  6 0.3900062
   7 9.504385e+00  7 0.2184413
   8 1.223956e+01  8 0.1408311
   9 1.310221e+01  9 0.1580358
  10 1.547808e+01 10 0.1155797
  11 1.559768e+01 11 0.1567361
  12 1.562099e+01 12 0.2092187
  13 1.746350e+01 13 0.1789735
```

Box-Ljung test
p-value = .4241

In terms of accuracy, we obtained the following results. These will be used further as reference for model selection purposes.

```
                       ME        RMSE         MAE        MPE       MAPE      MASE         ACF1
Training set 0.0004294394 0.002990185 0.002230584 0.01663733 0.08638859 0.4631204 -0.001145859
Test set     0.0022019930 0.006503662 0.004668379 0.08407158 0.17877116 0.9692625  0.740668214
```

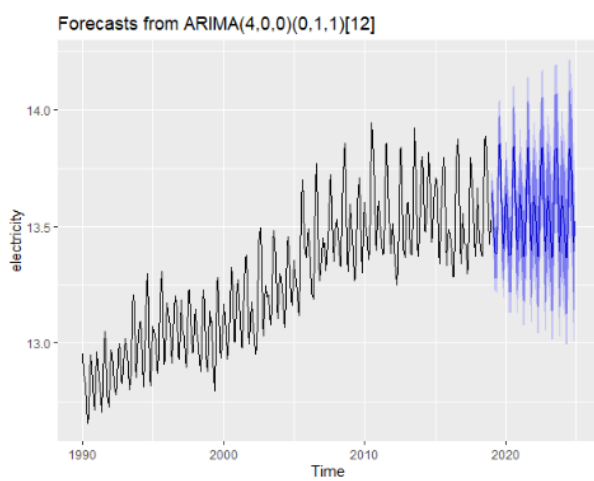
Choosing a model

To compare both models easily, the chart below shoes the results of each method.

	SNAIVE	ARIMA
P-value	2.2e-16	.4241
RMSE	.008510	.006503
MAE	.005996	.004668
AIC	-	-2293.87
AICc	-	-2293.54

As conclusion for this exercise, even though the models used failed the residuals test and based on the accuracy of them, we can select seasonal **ARIMA (4, 0, 0) (0, 1, 1)$_{12}$** as the final model since it is the one with better RMSE.

Finally, we compute the forecast on the whole data and obtain the final forecast.



Reference of the data:

https://www.eia.gov/electricity/data.php#elecenv