Forecasting – demonstration

Created by Elias Mayer ~ 25.08.2022

as a course submission for a ‘Forecasting’ module.

Contents

Executive summary

Data sources:

R library: Mcomp – N1355

R library: Mcomp – 70 quarterly time series of the M3 competition with IDs in [701, 1400] considering end numbers of 2 (702, 712, …).

Process:

In the first step, an explanatory data analysis was conducted on the data set: N1355.

From this analysis, matching models for forecasting were chosen through a variety of measurements, including AIC values, time-series cross validation and test forecasts.

The arima forecast performed the best out of these three methods and was estimated as an: (0,11)(0,1,1)[4] model. The ets model was estimated as an MMA model. The non-linear regression was constructed with knots to better fit the data but lead to inaccurate results due to over-fitting.

Through batch processes, automatic model fits to the time series N0702, N0712, N0722, … , N1392 with different model types was conducted.

A batch process for different automated exponential smoothing models was conducted. Also, a batch process for different automated arima models and a batch process for the Multiple Aggregation Prediction Algorithm (MAPA).

Afterwards, a dynamic approach, which uses cross validation of the different automated methods to choose the best fitting model type and conduct a forecast, was implemented.

All Batch processes lead to acceptable results. The dynamic approach and the automated arima model performed the best in a comparison of their MAPE (Mean absolute percentage error) scores.

In the last chapter, a hybrid (combined) forecast model on the basis of: arima, ets and naïve forecasting models, while also dynamically able to switch to MAPA for better results, was implemented and compared with the previous methodologies. This approach was not successful and underperformed the dynamic model.

A mix of methods (hybrid, MAPA, dynamic) used in dependence of their sector performance was estimated to be most suitable for the time series.

General Exploratory Analysis

Trend, Season and Remainders

The first step is the general exploratory analysis of the data, which enables to find suitable forecasting models. Specific characteristics of the data will be closer examined in the preparation chapters of the three models.

A simple plot of the data and its main data descriptive data is necessary to get an overview of the distribution of the time series and its seasonal, trend components (if applicable) and data labels.

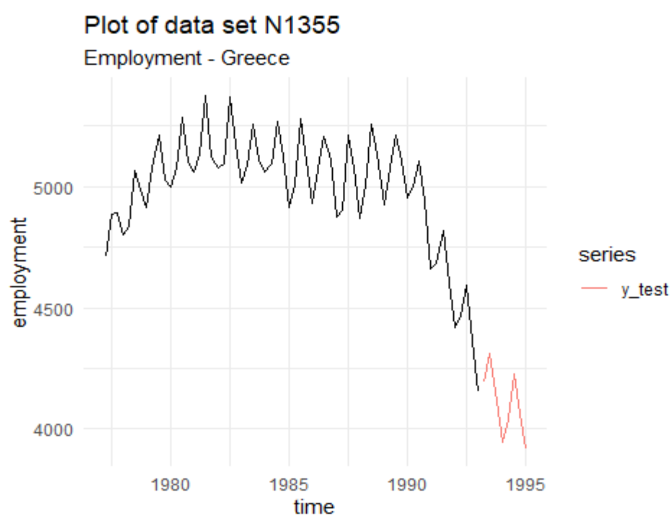


Figure 1 - Plot of dataset (partitioned in training and test data)

The examination of the data shows the data is: A quarterly time series. It is in the category of demographic data. It shows the employment levels in Greece through the dates: 1977 - Q2 to 1995 Q1. The data is labeled with "N1355" and has the type: demographic data. The data shows the development of employment in Greece through the quarters. The Pattern of the spikes shows a clear seasonal pattern in the data. A first increasing and then decreasing trend is visible too. Through a proceeding called "Decomposition" the single components of a time series can be separated and more clearly inspected. Which decomposition is suitable depends on the data (there are additive and multiplicative decompositions). The remainder component indicates which information is left in the data and should therefore have a random appeal.

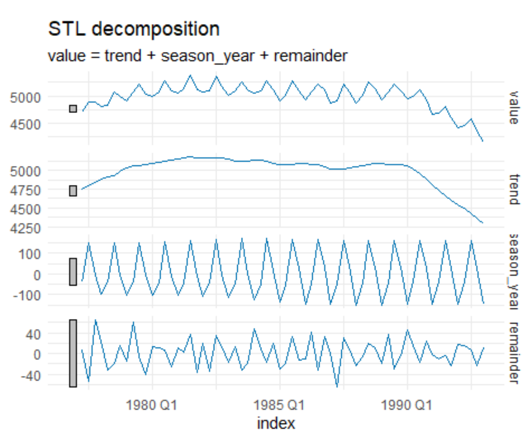


Figure 2 - separated components of the time series

The data without the seasonal component, for example, can be seen in Figure 3, which splits in the components below.

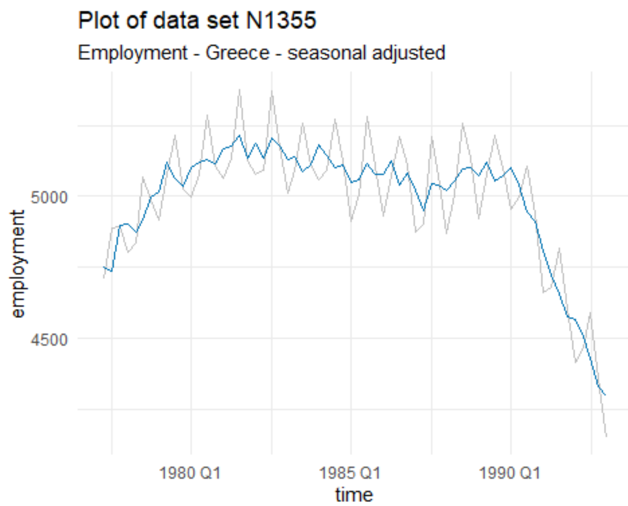


Figure 3

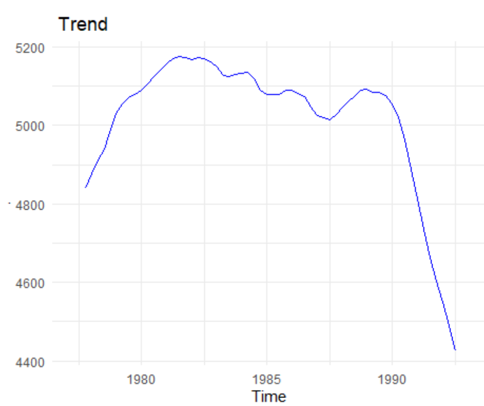


Figure 5

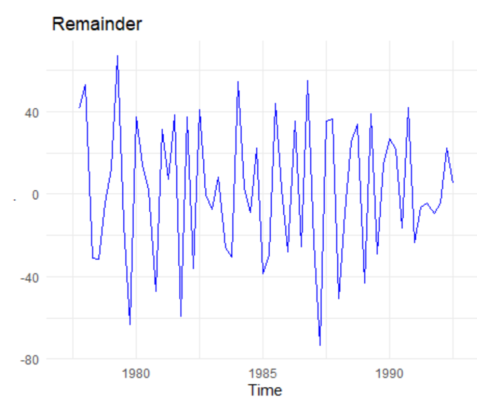
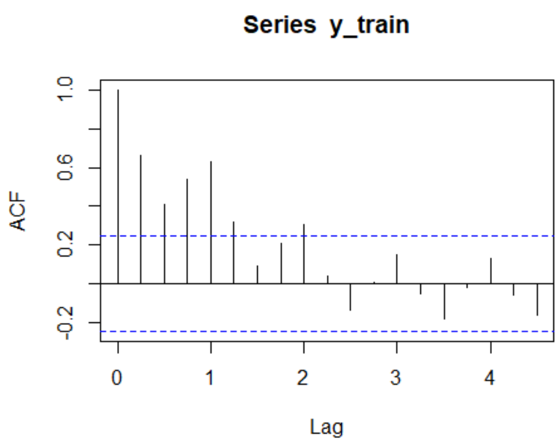
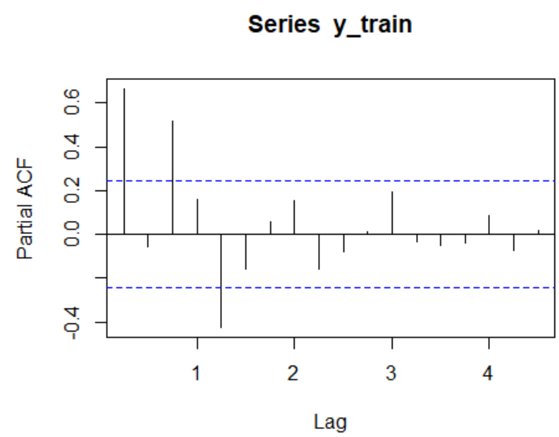


Figure 4

Autocorrelation and Stationarity

Autocorrealation measures the linear relationship between past values in a time series. These past values are called "lagged values". There can be different lag lengths. For example, variable y at t has a lagged value of y at t-1. This concept is important for regression models in time series and ARIMA models, which will be constructed in the following parts of the report. If there is no autocorrelation present in a time series, it is white noise, which means no values at different times are correlated. This is highly problematic if a meaningful model should be deprived from the data. A test if autocorrelation is present in the data source is therefore required. An ACF plot (correlogram) and PACF plot are solid ways to visualize and test for autocorrelation.





The blue line indicates the +- 1.96 standard error interval. This means in practice that if the lines cross this threshold, a significant autocorrelation is visible, therefore the data is not white noise.

```
Box-Ljung test
data: N1355
X-squared = 29.601, df = 1, p-value = 5.307e-08
```

A formal test for autocorrelation can be conducted additional to lower the chance of false-positive results. One of these tests is the "portmanteau test". The Test results in a tiny probability value, which means the test detected enough statistical evidence to reject the null hypothesis of the test. Therefore, dependence can be assumed.

Exponential Smoothing modeling, analysis and forecasting

Preparation

For successful exponential smoothing models, it is important to define if seasonal and trend components are additional or multiplicative. To clarify if a model should be additive or multiplicative, Pegel's classification was used in this report.

The trend component can be most likely classified under "dumped trend" an alternative possibility would be that the trend component is exponential.

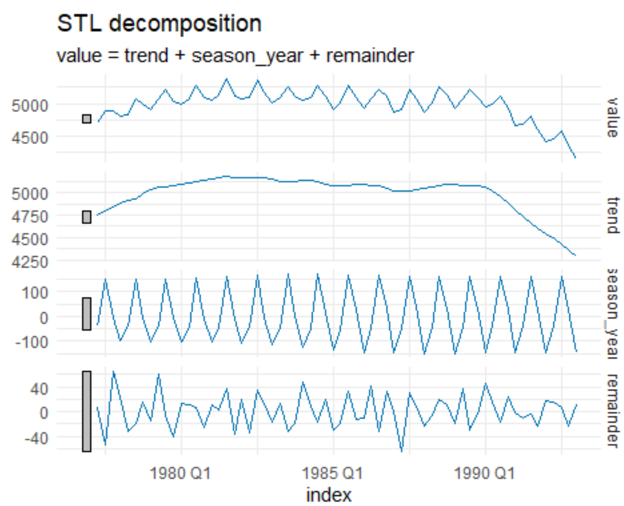


Figure 6

The seasonality seems to be stable over the time, but it is not completely clear if it is slightly decreasing at the beginning and the end of the distribution (the edges).

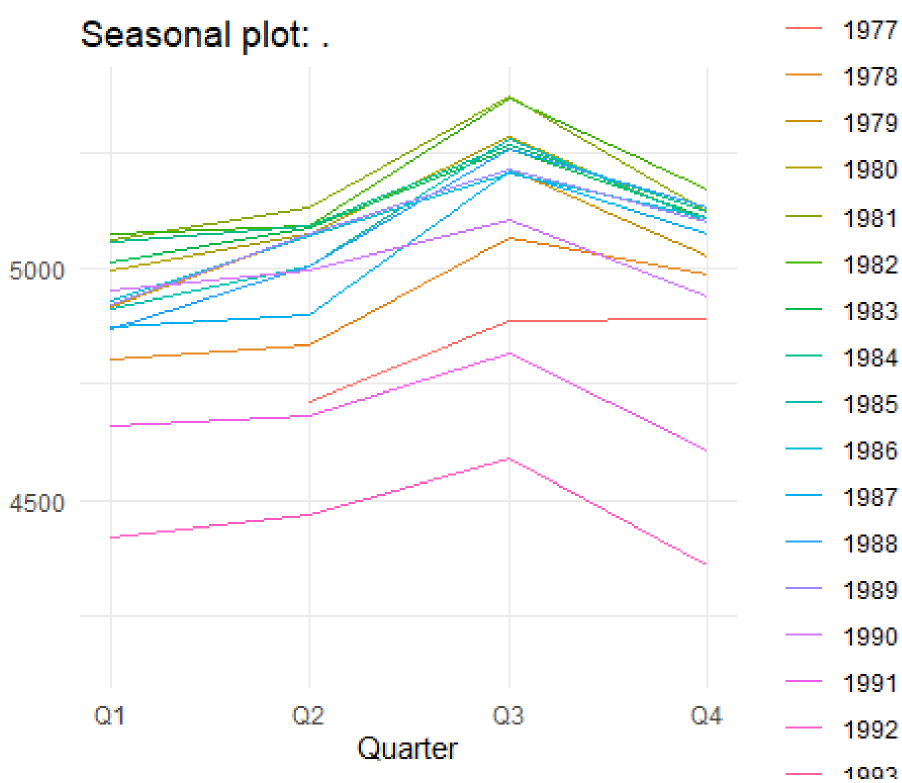
The assumption is that the seasonality should be additive due to its relatively stable height in peaks. In the following, these elements will be examined more closely.

A multiplicative version of the seasonality will be tested for the exponential smoothing model against a model fitted with additive seasonality.

Another way of insightful visualization is through a seasonal plot.

Underlying patterns, and especially pattern changes in the seasonal data, can be made more clear through the usage of seasonal plots. (Hyndman, 2018, p.2.4)

In the data, the seasonality seems stable to a degree. In the most years there is a decrease visible, starting at the beginning of the second quarter, which peeks at the beginning of the third quarter and declines afterwards. There are a few exceptions to this visible, e.g., in 1977, which doesn't decrease from the peak. Overall, the seasonality seems to be steady, which supports the hypothesis that the seasonality is additive but doesn't completely prove a lack of multiplicity.



With the gained knowledge about the data set, it is possible to formulate potential matching exponential smoothing models. The model has most likely an additive seasonality, damped trend and multiplicative error (due to only positive values).

Additionally, some other combinations will be tested to validate that the chosen model is resulting in the best possible estimation.

One model evaluated will be a model with multiplicity season and error and a damped additive trend. This model was chosen due to the small variation in the seasonal components towards the edges, which might indicate multiplicative seasonality. This model is known as the Holt-Winter's Exponential smoothing model with damped trend (MAdM). Another model will be constructed with multiplicity of errors, damped additive trend and additive seasonality (MAdA). A third less common model will be proposed. The model has multiplicity of errors, multiplicity of the trend component and an additive seasonality (MMA).

This method is called multiplicative trend exponential smoothing and is far less common than the Holt-Winter's Exponential smoothing model with damped trend. (Taylor, 2003, p.4)

AIC, AICc and BIC comparison

After the evaluation of different properties of the three exponential smoothing models, the one best fitting will be chosen and a final forecast implemented. The powerful tool in selecting an appropriate model is the AIC. AICc and BIC.

Model selection by these criteria shows which model is appropriate for forecasting a given time series. (Hyndman, 2018, p.8.6). Akaike's Information Criterion, in short AIC, penalizes for complexity (number of parameters) and measures the fit of the model. The smallest value for AIC is, in many cases, the most accurate model. (Hyndman, 2018, p.7.5) In the Table below, the details of the three fitted (matched to test data) models are displayed:

```
          AIC     AICc      BIC     BICc
 MMA 715.2094 718.5428 734.6394 741.5708

           AIC     AICc      BIC     BICc
 MAdA 715.8789 720.0298 737.4677 746.0994

           AIC     AICc     BIC     BICc
 MAdM 716.8052 720.9562 738.394 747.0257


Model estimated: ETS(MMA)
Sample size: 64
Number of estimated parameters: 9
Number of provided parameters: 1
Number of degrees of freedom: 55
Information criteria:
      AIC     AICc      BIC     BICc
 715.2094 718.5428 734.6394 741.5708


 Model estimated: ETS(MAdA)
 Sample size: 64
 Number of estimated parameters: 10
 Number of provided parameters: 1
 Number of degrees of freedom: 54
 Information criteria:
      AIC     AICc      BIC     BICc
 715.8789 720.0298 737.4677 746.0994
```

In the AIC, AICc and BIC the model MMA and MAdA have smaller values than the MAdM model, which leads to rejection of the MAdM Model. The MMA model has on parameter less and the MAdA model got penalized for having one parameter more. Therefore, the models get examined further.

Residual Exploration

To test if a model has captured all necessary information, the analysis of residuals can be conducted through plots.

For the model: MMA

The QQ plot shows that the residuals mostly follow a normal distribution.

The ACF plot is indicating there is one value which is significant at lag 7. This is neglect able due to the fact that it is a single occurrence (in a .95 confidence interval). The graph indicates that no autocorrelation of residuals is present.

It is necessary to double check this assumption with a formal test for autocorrelation called "Ljung-Box portmanteau test" on the residuals. The Test results in a probability value of 0.3681, which means the test did not detect enough statistical evidence to reject the null hypothesis of the test. Therefore, dependence can not be assumed. This is a satisfying result and shows that our models residuals don't contain significant patterns of information.

The plot of the residuals shows a somewhat even distribution of the residual means around zero, except for an outlier around ~ 1991. In conclusion, the model seems suitable and able to capture a fast enough amount of information from the underling series.

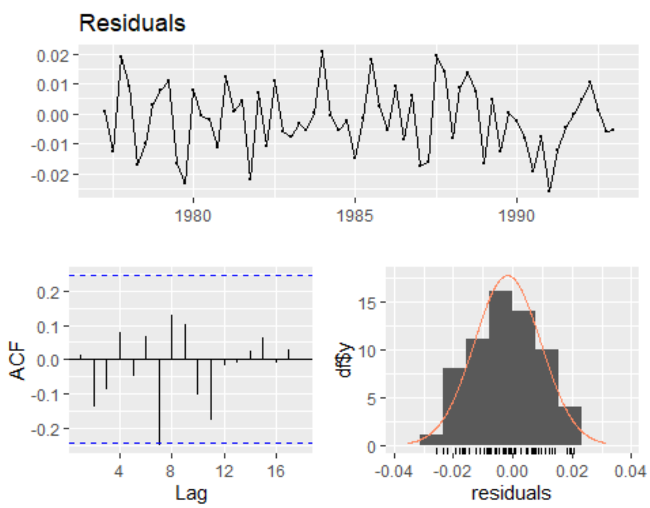


Figure 8

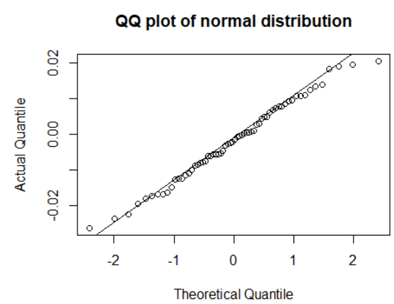


Figure 7

For the model: MAdA:

The QQ plot shows that the residuals mostly follow a normal distribution. In the tails, the residuals more clearly diverts from the norm compared to model one. Which can have implications for the prediction intervals.

The ACF plot is indicating there are no significant values and therefore no information left.

The Test results in a probability value of 0.4773, which means the test did not detect enough statistical evidence to reject the null hypothesis of the test. Therefore, dependence can not be assumed.

The variance in the residuals plot is slightly lower than in the graph of the previous model.

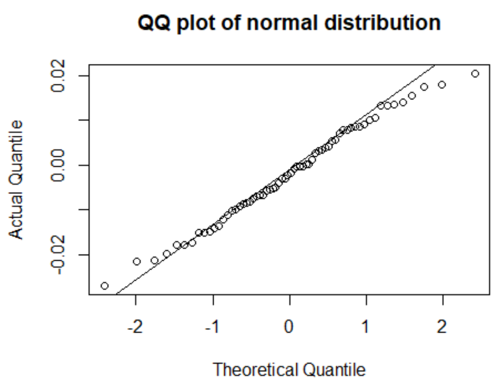


Figure 10

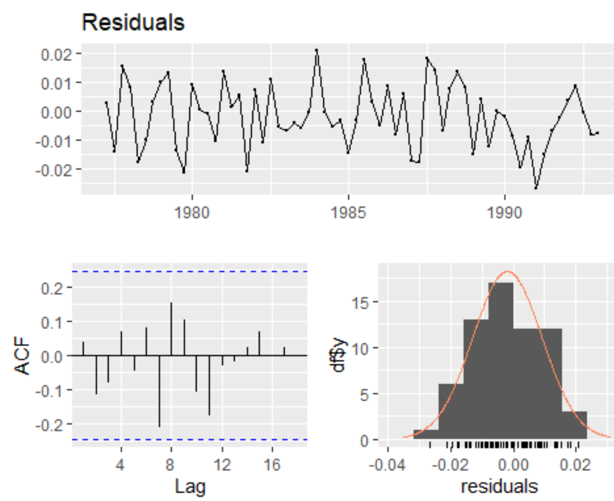


Figure 9

The analysis of the residuals of both fitted models doesn't clearly favor one model over another, so a third step in the model selection is conducted.

Model decision

Another method apart from the comparison of the AIC, AICc and BIC is the validation through the conduction of forecasts on a partition of the training data. The model with the best results is then chosen and used to do a forecast on all training data. In the case of the two models left, the best result in this frame can be defined as the model with the lowest MAPE value.

In the preparation for this step, we partition our training data in two parts. The training data from our training set is the training data without the 8 last quarters. Consequently, the last 8 quarters of observations in the training data set are the test set for this experiment.

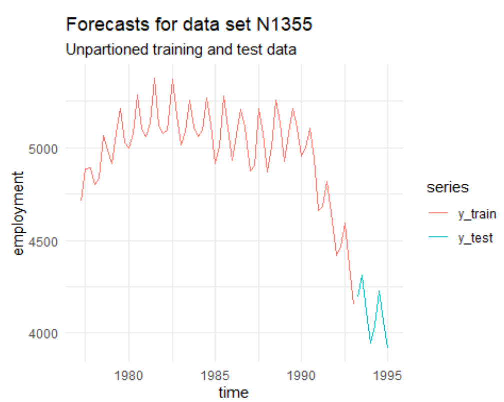


Figure 12

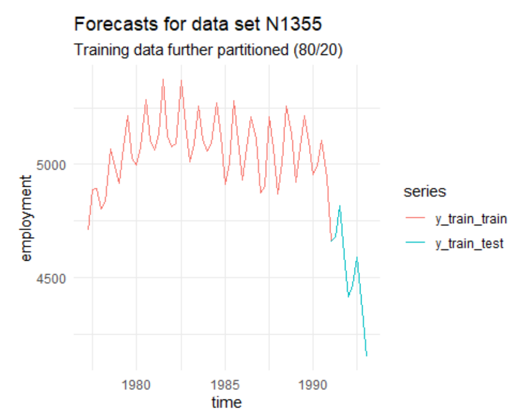


Figure 11

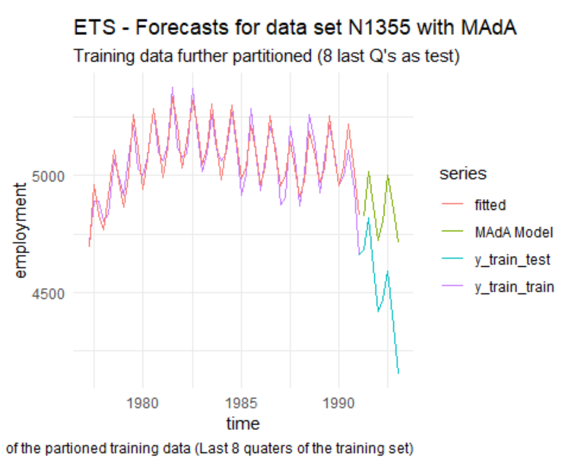


Figure 14

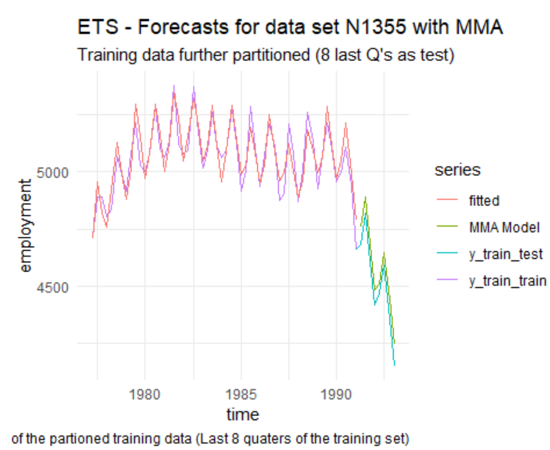


Figure 13

The direct comparison of the MAPE of both models clearly favors the MMA approach. This is also visible in the two given plots of the forecast on the partitioned training set. Therefore, we continue a final implementation with the MMA model.

```
1.581318 MAPE MMA Model

7.622148 MAPE MAdA Model
```

ARIMA modeling, analysis and forecasting

Preparation

The first step in creating robust "Autoregressive integrated moving average models" is to integrate the data (make it stationary).

The data source has a strong trend and seasonality and is therefore not stationary. A stationary time series is visually close to a white noise signal (it has plotted no recognizable visual pattern or systematic).

ARIMA models assume stationarity of the data, which must therefore be created in our data set, before we can continue with the model selection and construction. Through a process called "Differencing" (which computes the difference between following observations) it is possible to stabilize the mean and trend. Through the usage of a logarithmic transformation, it is possible to stabilize the variance (the variability from the mean).

A statistical test, the "Unit root test" shows which level of differencing is required. The test is showing that two differencing steps are required for the data, one for the normal component and one for the seasonal component.

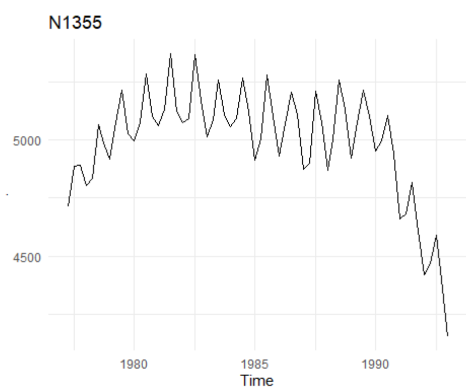
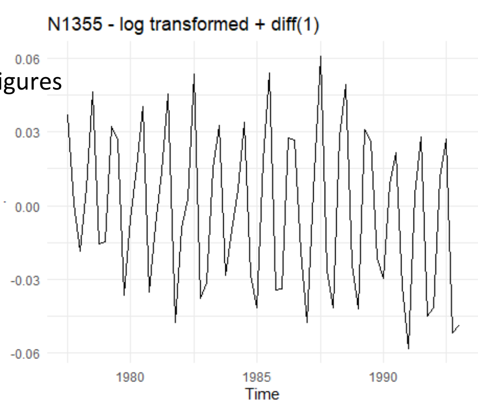


Figure 15 - N1355 Employment data

Figures



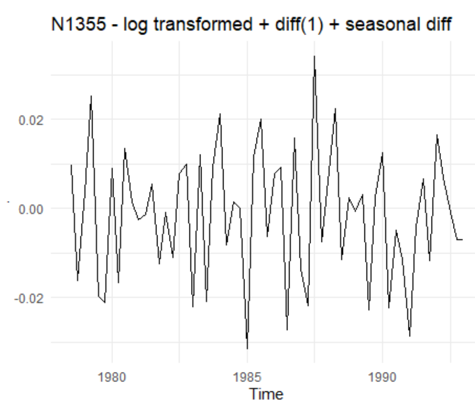


Figure 16 - N1355 after first differencing *Figure 17 - N1355 after second differencing*

Figure 17 - N1355 after second differencing is showing no visual pattern. The data seems random distributed and is therefore useable for the fitting of an ARIMA model.

The examination of the ACF and PACF plots of data helps in determining a suitable ARIMA model. A matching fit could be an ARIMA (4,1,0) $(0,1,0)_4$ model. The reasoning behind this will be very briefly explained.

The last significant spike in the PACF plot can show the number of required auto regressive values if the ACF plot shows a sinodel pattern, as in the graph below. This is due to the fact that these values show significant autocorrelation.

Another fitting model could be a simple seasonal moving average model.

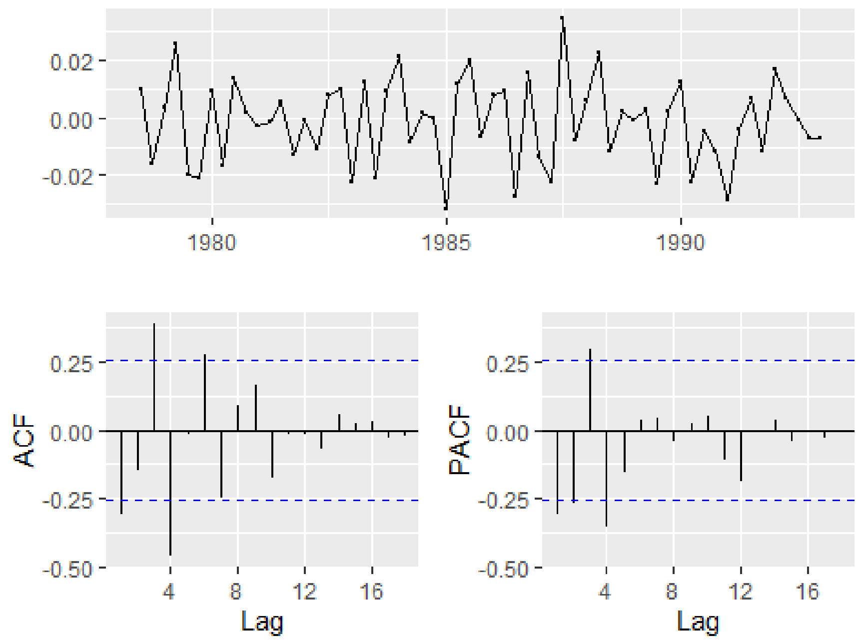


Figure 18 – ACF / PCAF Plots

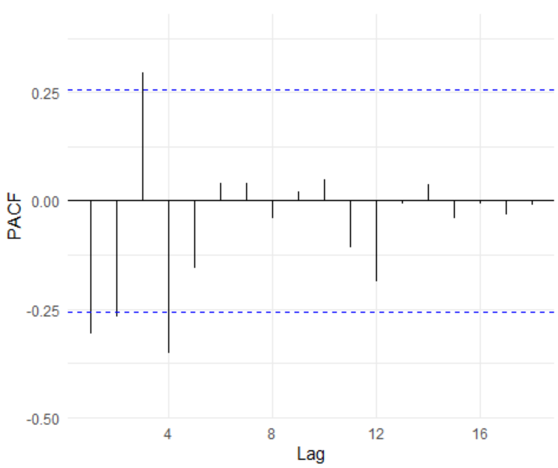


Figure 19 - Detailed view of the PACF plot

Therefore, the moving average seasonal model (ARIMA (0,1,1) $(0,1,1)_4$) will be tested against the previously defined model (auto regression model).

AIC, AICc and BIC comparison

Below, the fitted models are directly compared in terms of AIC, AICc and BIC. The comparison shows that the first model has a lower AIC value but higher AICc and BIC values.

```
Model: ARIMA(4,1,0)(0,1,0)4  - seasonal autoregressive model

Coefficients:
          ar1      ar2     ar3      ar4
      -0.1914  -0.1667  0.2007  -0.3360
s.e.   0.1215   0.1217  0.1200   0.1221

AIC=662.04   AICc=663.17   BIC=672.42
```

```
Model: ARIMA(0,1,1)(0,1,1)4 – seasonal moving average model

Coefficients:
          ma1     sma1
      -0.2333  -0.4674
s.e.   0.1219   0.1392

AIC=662.16   AICc=662.6   BIC=668.39
```

The comparison doesn't lead to an obvious result. In the next step, an exploratory analysis of the residuals will take place in both models.

Residual Exploration

An examination of the residuals is necessary to determine the suitability of the models in capturing the information from the data source. If the residuals of the fitted models show recognizable patterns, this indicates that not all informations from the data was captured.

Residuals of model one:

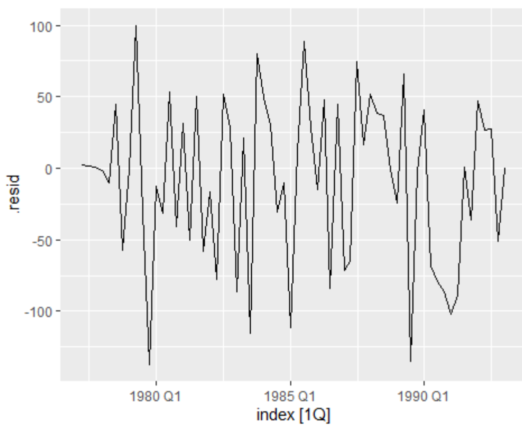


Figure 21

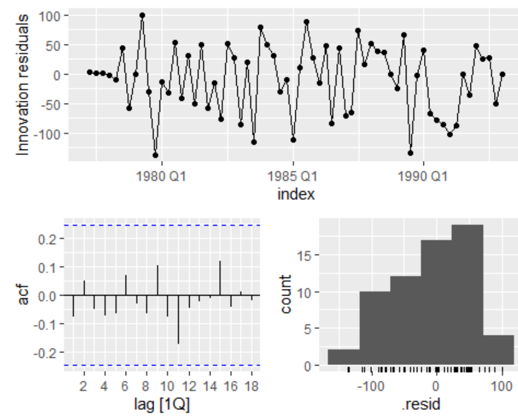


Figure 20

Residuals of model two:

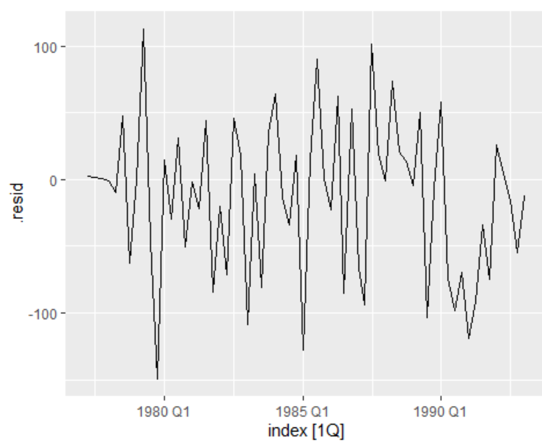


Figure 23

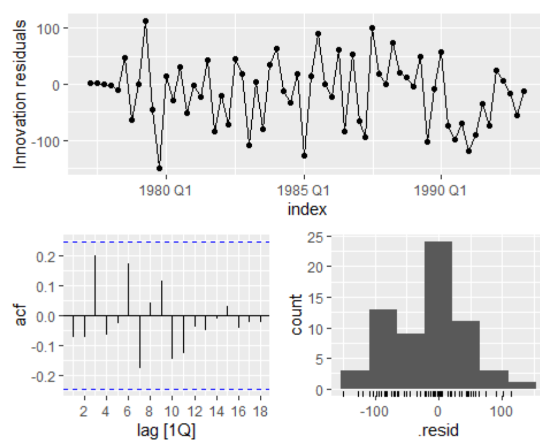


Figure 22

The residuals are not distinguishable from white noise in both models. Also, the Jung-Box Test Results of both models imply independence of residuals (p = 0.9945156 for model 1 and p = 0.3918196 for model 2). The normal distribution from model one is strongly skewed, which has implications for its prediction interval.

Cross Validation ARIMA

Another test will be conducted to make a final model decision. To find out which ARIMA model fits better, an implementation of time-series cross-validation on partitioned training data is conducted. This means there will be several forecasts conducted, from different origins, and their average performance will be compared. Performance refers, in this context, to the MAPE value of the models. The model specifications are listed below.

Results:

Average MAPE in the Model: `ARIMA (4,1,0)(0,1,0)[4] =` **2.217433**

Average MAPE in the Model: `ARIMA (0,1,1)(0,1,1)[4] =` **2.080882**

The cross-validation results in a lower average MAPE score for the seasonal moving average model:

ARIMA (0,1,1) $(0,1,1)_4$

Therefore, this model gets forecasted. Afterwards, its performance is measured against the real test data.

Regression modeling, analysis and forecasting

Exploratory Modeling

In the first step, a linear regression model is fitted to the data. Predictor variables of the regression model can be used through quarterly dummy variables (to capture seasonality) and a dummy variable for the trend.

The summary of the model shows that the second season has no significant predictive value in the model also, residuals show that information from the underlying data has not been captured well. The clear bow shape resembles the given trend of the data.

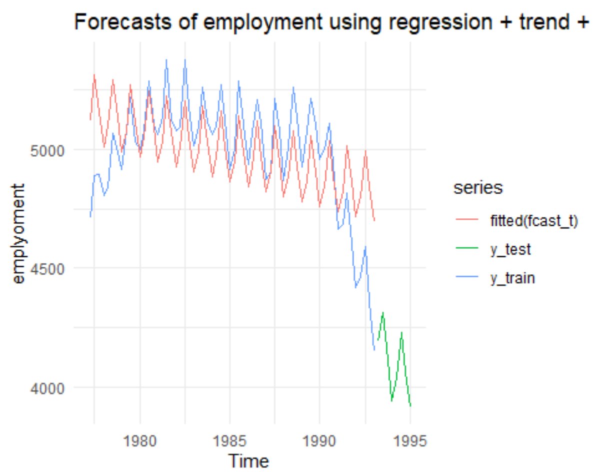


Figure 24 - Linear Regression with Season and Trend

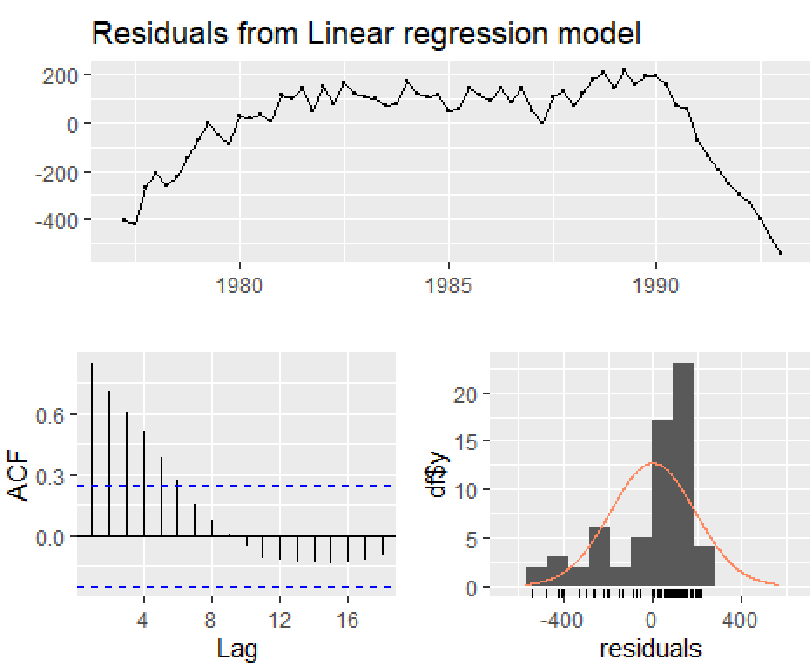


Figure 25 - Residual Analysis

Through the usage of subjective knots which determine changes in the historical data, we can fit the model more closely to the data. This method can lead to over-fitting. Due to the fact that the data is employment, economical events correlated with data, setting knots on peaks is somewhat acceptable. The identified data knots are the years 1980 (~ stop increasing movement) and 1990 (~ start decreasing). An examination of the residuals shows that trend was captured to a degree by the model. The ACF plot shows no significant values. Therefore, the model is seen as suitable and will be used in the forecasting process.

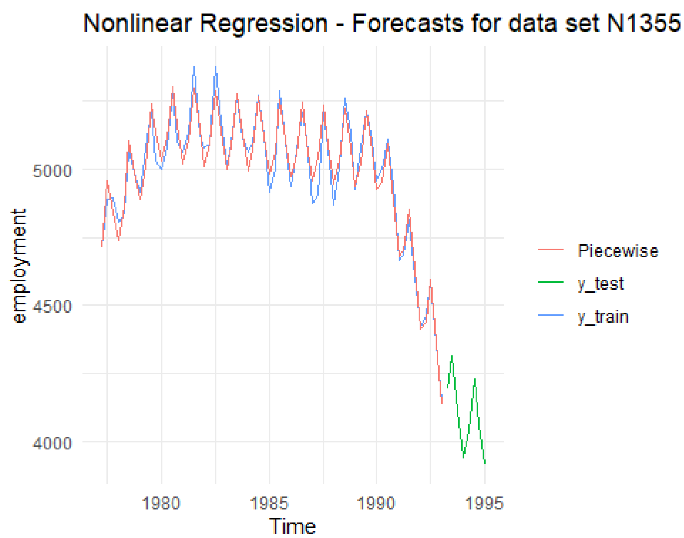


Figure 26 - Linear Regression with Knots

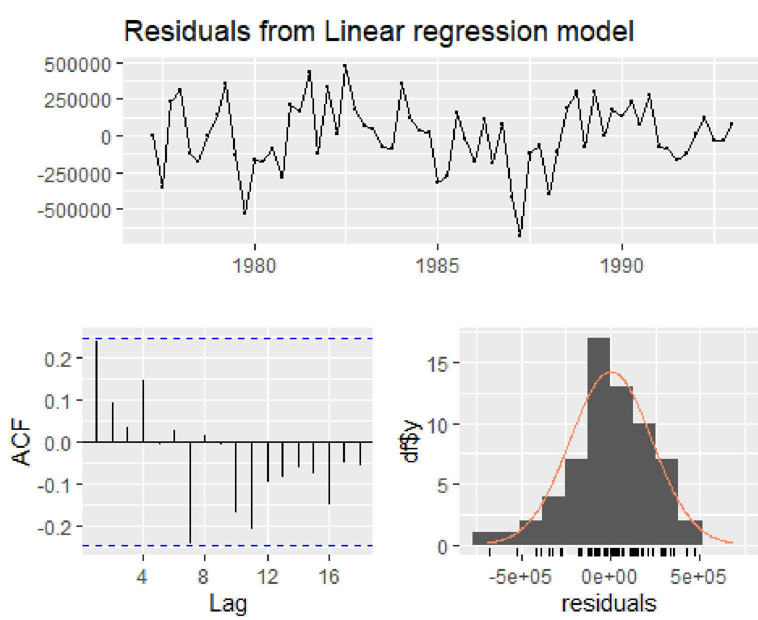
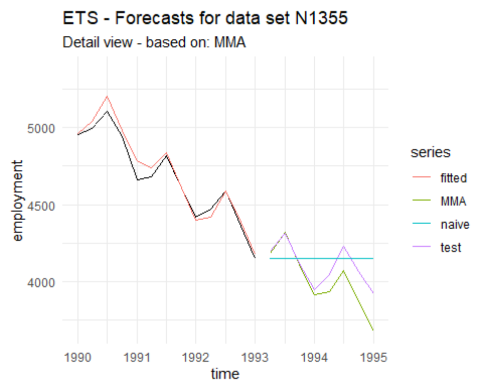


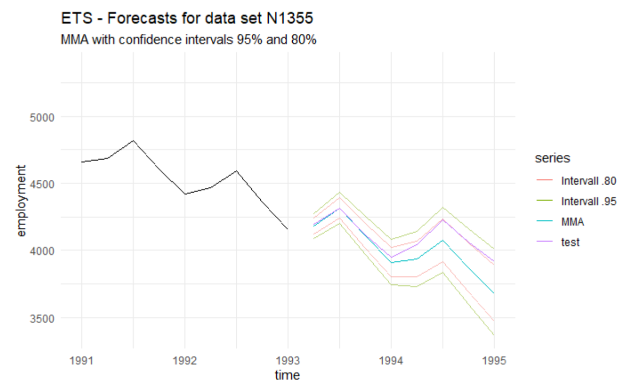
Figure 27

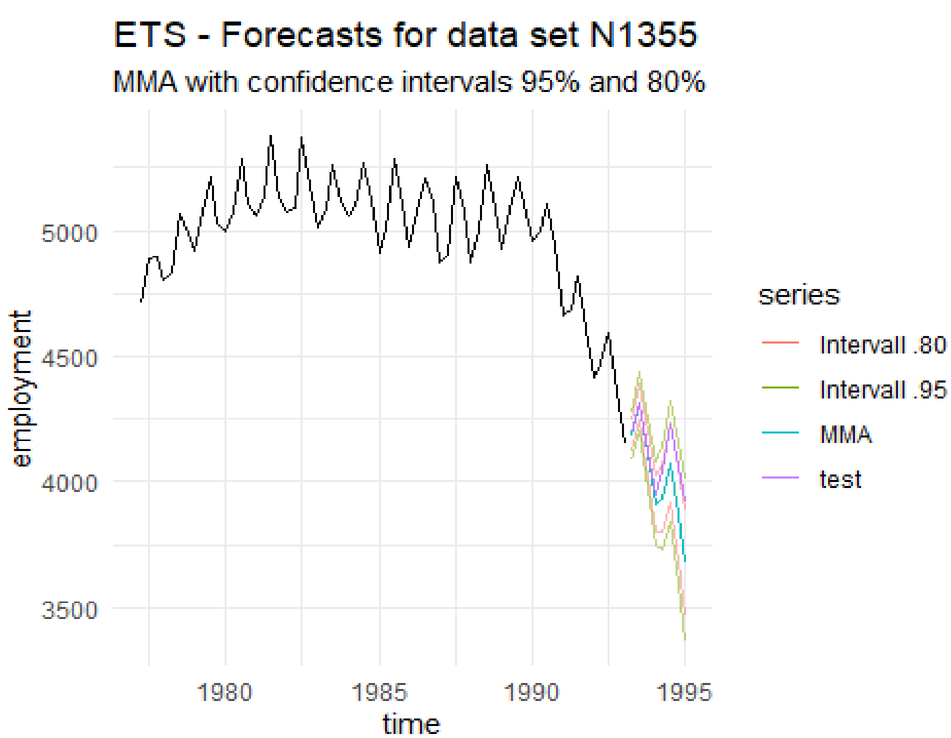
Final Forecasts

Forecast ETS – Exponential Smoothing Model

The final **MAPE** value for the ETS model is **2,35** (2.359635). The test value leaves the inner prediction interval of the model slightly in parts.

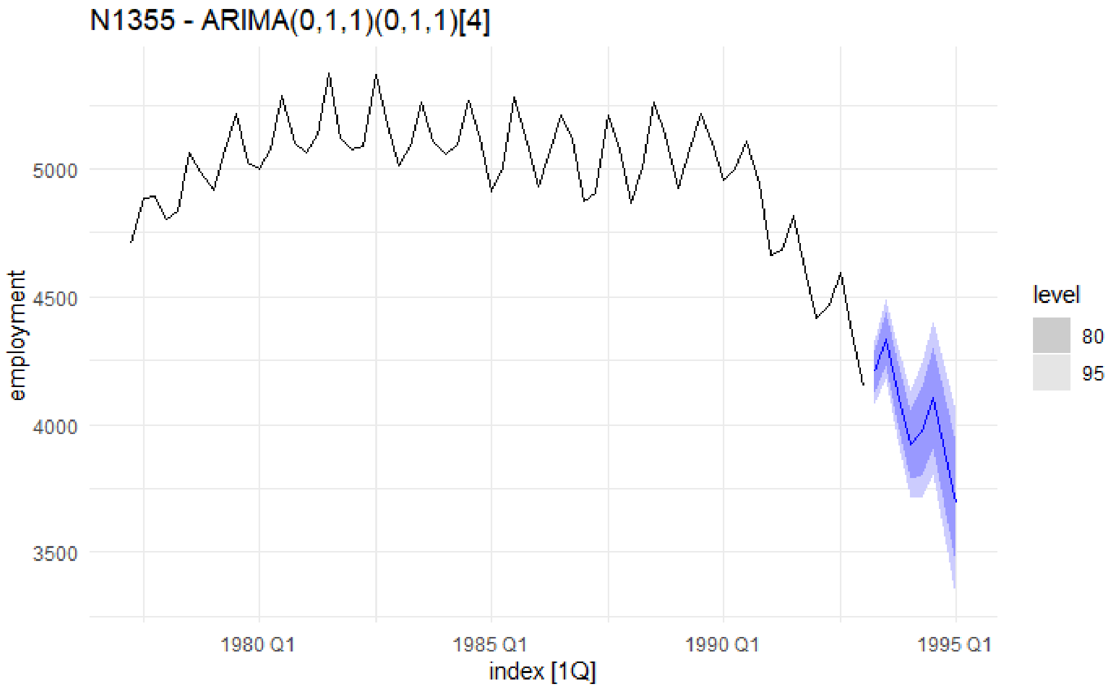






Forecast ARIMA – Auto-Regressive Integrated Moving Average

The final **MAPE** value for the ARIMA model is **2,03** (2.038594). This performance is in an acceptable area but could be further improved through batch testing.



Forecast NLR - non-linear regression with knots

The final **MAPE** value for the non-linear regression model with knots is **44.14** (44.14482). This performance is in the unacceptable area and strongly indicates an over-fitting of the model.

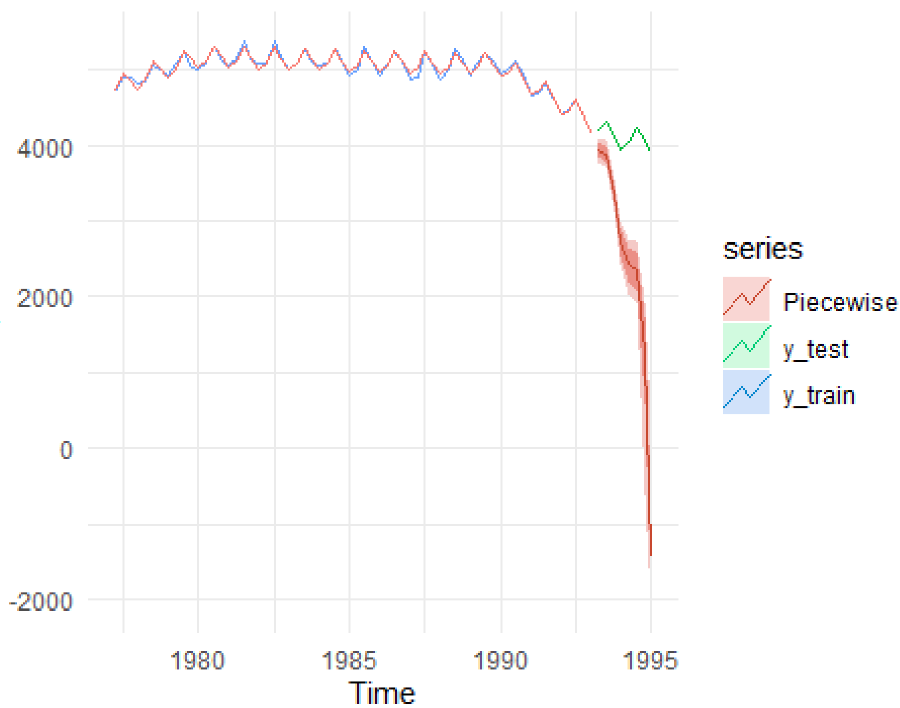


Figure 28

Evaluation of the final manual models

MAPE values per models:

- Non-linear regression with knots – model: **44.14**
- ARIMA model: **2,03**
- ETS model: **2,35**

It is clear that ARIMA performs the best, followed closely by the ETS model. The non-linear regression model performed unusable with the data.

Batch forecasting

Automatic exponential smoothing forecasting

In this part of the report, automated batch processes will be applied to a series of quarterly data. The Data has the different types: demographic, finance, industry, macro, micro. This report will evaluate how the automated models perform, which version was implemented for which type and trend (of data), how the models perform over different starting points (cross validation), and how they perform compared to three bench mark forecasts: naïve, seasonal naïve and mean. In total, the automated processes compute the results for 70 datasets on the basis of training data, which is cross checked with the matching test data set. data will be presented in averaged or sum formats to transport information efficient. The entire tables without sum or average functions are shown in the Appendix.

The data set analyzed contains the following data tables:

N0702, N0712, N0722, … , N1392. The data set is composed of 5 demographic time series, 8 financial time series, 9 industry time series, 33 macro data time series and 15 micro data time series. All data is quarterly data. 12% of the series show a steady or positive trend while the rest show a decline.

The error measures chosen for the evaluation of the automated batch process with exponential smoothing models are: MAPE, MASE and the Cross Validated average of MAPE. These error measurements were chosen for a variety of reasons:

MAPE is the average of absolute percentage error. It has the disadvantage that in dealing with zero or close to zero values; it produces undefined values. This disadvantage doesn't apply to our data set. MAPE is used extensively as a measure of forecast accuracy, also due to its easy interpretation as percentages. (Kim and Kim, 2016, pp.669-700)

MASE gives the forecast error by using naïve as a reference method and therefore doesn’t generate infinite or undefined values. (Kim and Kim, 2016, pp.669-700) It is also a great indicator for comparative performance and is therefore included.

The third metric is the average MAPE created through cross validation of the trainings data. Cross validations were also conducted with the benchmark methods: naïve, snaive and mean.

Table 1 shows the result of the batch process and shows the average performances in each sector (type), and further divided, for each model. The green highlighted models are the in average best performing ones per type (It doesn't state the absolute best performing model). Yellow marks show mediocre models and counts. While red marks show problematic models.

Row Labels	Average of MAPE_ETS	Average of MASE	Average of CV_MAPE_ETS	Count of model
DEMOGRAPHIC	**9.32843286**	**1.138200573**	**8.351535495**	**5**
ETS(A,A,A)	2.003229138	0.356719315	2.654596876	1
ETS(A,N,N)	2.604898813	1.977699148	1.360859358	1
ETS(M,A,M)	1.98541739	2.031793561	0.666229592	1
ETS(M,Ad,N)	0.82518534	0.101039463	8.98273301	1
ETS(M,N,N)	39.22343362	1.223751377	28.09325864	1
FINANCE	**15.72186597**	**1.321513841**	**11.02982143**	**8**
ETS(A,N,A)	12.59528174	1.986863471	8.505208547	1
ETS(A,N,N)	18.46180664	1.420172809	8.072774659	3
ETS(M,Ad,N)	30.91315574	1.233884479	34.09112861	1
ETS(M,N,N)	8.960356791	1.030281451	7.141303425	3
INDUSTRY	**8.718280954**	**0.709843069**	**18.45255114**	**9**
ETS(A,A,A)	27.7233083	0.374499261	95.56290992	1
ETS(A,Ad,N)	1.630689241	0.794347217	3.4497286	1
ETS(M,A,M)	3.851247209	0.749014884	7.104222699	1
ETS(M,Ad,M)	6.090784216	1.07860401	8.910777486	1
ETS(M,N,M)	6.396218956	0.368150479	12.16270978	2
ETS(M,N,N)	8.792020568	0.885273763	8.906633992	3
MACRO	**6.177063503**	**1.217931438**	**5.654819826**	**33**
ETS(A,A,A)	1.109309883	0.276831679	12.17089185	1
ETS(A,A,N)	3.871680852	1.021271075	2.227437162	11
ETS(A,N,A)	6.660484748	2.855548791	7.72035915	2
ETS(A,N,N)	9.026334704	0.7004758	12.72251638	1
ETS(M,A,A)	1.207096465	0.261772124	17.52250836	1
ETS(M,A,M)	4.178238574	1.027328621	3.631617326	2
ETS(M,A,N)	7.448985644	1.459318996	5.038959532	7
ETS(M,Ad,M)	0.982141197	0.405683307	2.620209869	1
ETS(M,Ad,N)	2.742609219	0.532493143	5.904430973	3
ETS(M,N,N)	16.72038754	1.933631371	10.34528911	4
MICRO	**16.32748171**	**1.529675301**	**19.01985446**	**15**
ETS(A,A,A)	7.127480552	0.804872455	6.861722543	1
ETS(A,N,A)	9.952963211	0.657216653	17.21313655	1
ETS(M,A,M)	9.762811408	1.204592685	5.87475801	2
ETS(M,A,N)	24.20998996	1.801189218	15.25367174	2
ETS(M,Ad,M)	9.247297077	1.168387419	7.215742726	1
ETS(M,N,A)	5.667360141	1.695008804	4.04262075	1
ETS(M,N,M)	27.81436216	2.196629529	12.19947213	2
ETS(M,N,N)	17.86855952	1.642964264	36.66175812	5

Table 1

From the results in the table, it becomes visible that in each sector (type) a different exponential smoothing model is in average, resulting in lower MAPE values. Interesting is the difference in model variety per type.

One area is finance, another area is Micro. The data above shows averages and is therefore heavily tinted by outliers and unsuitable forecasts. Macro and demographics are the best captured types in average.

Another interesting aspect is to see in which types (sectors) the benchmark methods outperform the automatic ets models (in average MAPE). The results in table Table 2 show that, on average, exponential smoothing was outperforming other models in each category. The least differences to the second-best performer are in the category: demographic, industry and finance.

Row Labels	**Average of CV_MAPE_ETS**	**Average of CV_MAPE_NAIVE**	**Average of CV_MAPE_MEAN**	**Row Labels**
DEMOGRAPHIC	8.351535495	9.965747577	25.73032282	DEMOGRAPHIC
FINANCE	11.02982143	13.21513429	21.23222222	FINANCE
INDUSTRY	18.45255114	16.90460107	57.94482376	INDUSTRY
MACRO	5.654819826	8.114440594	21.15500454	MACRO
MICRO	19.01985446	23.63077596	27.61216304	MICRO

Table 2

The last point is which model performed the best on average across all types and trends, etc. and how often the models were applied to the data source. Important in the examination of this is to remember that this is averages contain outlier forecasts which heavily tint the resulting scores.

Row Labels	**Average of MAPE_ETS**	**Average of CV_MAPE_ETS**	**Count of model**
ETS(M,A,A)	1.207096465	17.52250836	1
ETS(A,Ad,N)	1.630689241	3.4497286	1
ETS(A,A,N)	3.871680852	2.227437162	11
ETS(M,Ad,M)	5.440074163	6.248910027	3
ETS(M,A,M)	5.619794094	4.46386716	6
ETS(M,N,A)	5.667360141	4.04262075	1
ETS(M,Ad,N)	7.993233747	12.15743091	5
ETS(A,N,A)	8.967303613	10.28976585	4
ETS(A,A,A)	9.490831968	29.3125303	4
ETS(M,A,N)	11.17365327	7.308895578	9
ETS(A,N,N)	13.40333069	7.660339942	5
ETS(M,N,N)	15.54405709	18.80793862	16
ETS(M,N,M)	17.10529056	12.18109095	4

Table 3

Automatic ARIMA forecasting

The same process described in the exponential smoothing batch render process was conducted.

The comparison results to of ARIMA to ets are:

Sectors / Types:

• **Demographic** > **ARIMA AVG 5.67** - ETS AVG 9.32
• **Finance** > **ARIMA AVG 15.08** - ETS AVG 15.72
• **Industry** > ARIMA AVG 10.85 - **ETS AVG 8.71**
• **Macro** > ARIMA AVG 6.714 – **ETS AVG 6.177**
• **Micro** > **ARIMA AVG 15.114** – ETS AVG 16.32

The results show that in dependence of industry type, either ARIMA of ETS performs better.

For the complete tables, see Appendix.

Seasonality or cyclic	Count
1	35
2	8
3	1
4	10
10	1
11	2
14	1
16	2
18	1
20	1
21	2
26	1
29	1
42	1
45	2
53	1

Table 4 - models per seasonality count

DEMOGRAPHIC	Average of MAPE_ARIMA	Average of MASE_ARIMA	Count
DEMOGRAPHIC	**5.67395105**	**1.039261477**	**5**
0 0 1	4.254144319	3.239114847	1
1 0 0 0 1 0 4	3.042107479	0.552630076	1
1 1 0	10.27610775	0.435597409	2
1 1 0 0 1 0 4	0.521287949	0.533367644	1
FINANCE	**15.08574544**	**1.279147643**	**8**
0 1 0	8.555666355	1.027022079	4
0 1 1	31.82764401	1.531723295	1
1 0 1 1 1 0 4	10.88676664	1.721328123	1
1 1 0	30.27407325	1.211388173	1
2 1 2 0 0 1 4	13.47481417	1.660653237	1
INDUSTRY	**10.85420744**	**0.777384937**	**9**
0 1 0	4.081325957	0.982657203	2
0 1 0 0 1 1 4	5.722303877	1.036557057	1
0 1 0 1 0 0 4	12.83559899	0.310733106	1
0 1 0 1 0 1 4	27.00699432	0.366271377	1
1 0 0 0 1 2 4	1.121844304	0.367582766	1
1 0 0 1 0 1 4	6.795534079	0.62141306	1
2 0 0 0 1 1 4	4.966839119	0.877180298	1
2 0 1	31.07610038	1.451412363	1
MACRO	**6.714628354**	**1.18504635**	**33**
0 0 0 0 1 0 4	3.25632072	1.46594985	2
0 1 0	7.166639757	1.357554097	6
0 1 0 0 1 1 4	0.488016962	0.20042548	1
0 1 0 1 0 1 4	36.65957437	4.78631173	1
0 1 1 0 0 1 4	8.229503567	1.132526485	2
0 1 2 0 0 1 4	4.601313753	0.804152292	1
0 1 2 0 0 2 4	1.072123598	0.215892363	1
0 2 1	6.330553616	1.753307814	2
1 0 0 0 1 0 4	6.562912632	1.605768602	1
1 0 0 1 1 0 4	1.619169802	0.291201741	2
1 1 0	4.508946909	1.0968828	4
1 1 0 0 0 1 4	16.50587148	2.503587282	1
1 1 0 1 1 0 4	2.159662999	0.516023769	1
1 1 1	27.18029195	1.988851633	1
1 2 1	1.42287911	0.382399859	1
2 0 0	1.475476078	0.588866484	1
2 0 0 0 1 0 4	2.273137713	0.493143693	1
2 1 0	12.22629248	1.418436923	1
2 1 2 1 1 0 4	2.911522746	0.738998163	1
3 1 0	5.081786735	0.694055808	1
3 1 1	1.055151476	0.350787911	1
MICRO	**15.11455367**	**1.42578594**	**15**
0 0 0 0 0 1 4	7.433158184	1.291409267	1
0 0 0 0 1 1 4	12.99867314	0.96644351	1
0 0 1 0 1 0 4	48.68886329	3.222406122	1
0 1 0	13.66943421	1.014467442	2
0 1 1 0 1 0 4	5.273664833	0.79895486	1
0 1 3 0 1 1 4	10.88642271	1.283827897	1
1 0 0 0 1 1 4	9.116758738	1.462709716	3
1 0 1	37.99973591	1.738369337	1
1 0 1 1 0 0 4	8.256078229	0.671993195	1
1 1 1	23.43926032	3.399618734	1
2 0 0	7.190912122	0.936386298	1
2 0 0 2 1 0 4	9.862391651	0.660315851	1

Table 5

Multiple Aggregation Prediction Algorithm (MAPA)

A summarized description of the MAPA approach is to disassemble a time series into multiple frequencies, forecast them, and combine the forecasts. The technical details are beyond this report.

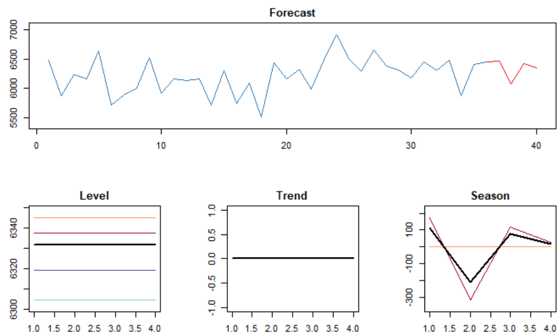


Table 6 - simple mapa (only example) output for ts 712

Table 7 - 9 MAPA example forecasts from the time series

Sectors / Types:

• **Demographic** > **ARIMA AVG 5.67** - ETS AVG 9.32 – MAPA AVG 12.86
• **Finance** > **ARIMA AVG 15.08** - ETS AVG 15.72 - MAPA AVG 16.32
• **Industry** > ARIMA AVG 10.85 - ETS AVG 8.71 - **MAPA AVG 8.62**
• **Macro** > ARIMA AVG 6.714 – **ETS AVG 6.177** - **MAPA AVG 6.12**
• **Micro** > **ARIMA AVG 15.114** – ETS AVG 16.32 - MAPA AVG 15.83

The comparison table shows MAPA had good average results in the sectors industry and Macro. For more detailed results, see the table in the appendix.

Dynamic model selection

To forecast always with the most suitable model, cross valuation of ARIMA, ETS and MAPA models will be evaluated. From all three models, cross validations will be conducted per time series, and the average MSE (Mean-square-error) results compared. The model with the better performance in terms of MSE value will be chosen and used for the final forecast, which will be evaluated as MAPE to enable comparison with previous model performances. In the picture below, the workflow is displayed in Table 8.

The windows for the time series cross validations were chosen as:

Length (tim series) – (2*forcasting horizon) → length (time series)

This keeps computation time in check and evaluates our models over ~ 16 months of the training data.

Row Labels	Average of MAPE_model
DEMOGRAPHIC	**9.25337008**
arima	1.479505989
ets	20.91416622
FINANCE	**15.67973184**
arima	21.87429325
ets	16.12370478
mapa	8.597224538
INDUSTRY	**8.038567746**
arima	3.709087298
ets	4.513144793
mapa	15.62545198
MACRO	**6.102449849**
arima	6.318785921
ets	6.882451211
mapa	1.872316398
MICRO	**14.93308031**
arima	19.05944431
ets	10.7623234
mapa	13.95698046

Table 9

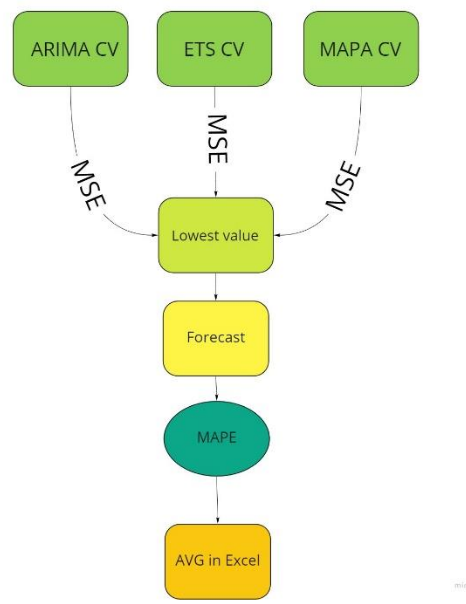


Table 8 - Workflow dynamic model

Table 9 shows that the averages per sector are favoring the dynamic method. For the model per table, please see the whole table in the Appendix. A second realization from the dynamic model is that the MSE as decision criteria might be suboptimal.

Sectors / Types:

• **Demographic** > **ARIMA AVG 5.67** - ETS AVG 9.32 – MAPA AVG 12.86 – dynamic AVG 9.25
• **Finance** > **ARIMA AVG 15.08** - ETS AVG 15.72 - MAPA AVG 16.32 – dynamic AVG 15.67
• **Industry** > ARIMA AVG 10.85 - ETS AVG 8.71 - MAPA AVG 8.62 – **dynamic AVG 8.03**
• **Macro** > ARIMA AVG 6.714 – ETS AVG 6.177 - MAPA AVG 6.12 – **dynamic AVG 6.10**
• **Micro** > ARIMA AVG 15.114 – ETS AVG 16.32 - MAPA AVG 15.83 – **dynamic AVG 14.93**

The below Table shows the method of the distribution per type (pink marked the 2 most commonly used models).

Row Labels	Count of choosen_model
DEMOGRAPHIC	**5**
arima	3
ets	2
FINANCE	**8**
arima	2
ets	4
mapa	2
INDUSTRY	**9**
arima	2
ets	4
mapa	3
MACRO	**33**
arima	19
ets	11
mapa	3
MICRO	**15**
arima	6
ets	5
mapa	4
Grand Total	**70**

Table 10 - dynamic model selection results

Combination strategy

For the creation of a suitable combination strategy, results from previous tests are taken into consideration.

In the first attempt, a comparison between the cross validation MSE of a hybrid forecast (arima, ets, naive) model and a MAPA model is conducted. And the forecast is made by the model with the lower MSE.

The automatic weighting of the hybrid model was set on equal because of insufficient knowledge about all the data series. If Knowledge is limited, a basic weighting is recommended. (Armstrong, 2001, p.417)

This led to comparable poor results (the dynamic model has a lower MAPE in all areas). For more details, please see the table in the appendix.

As the second attempts another hybrid model was constructed. This model widens its used models from arima, ets and naïve to theta. A theta model is a combination model which decomposes a series into short- and long-term components. (Assimakopoulos and Nikolopoulos, 2000, p.521) This attempt performed significantly better but still underperformed in some sectors (marked red in Table 11).

Row Labels	Average of hybrid_MAPE
DEMOGRAPHIC	10.10759014
FINANCE	16.3703608
INDUSTRY	8.649479827
MACRO	6.109349352
MICRO	14.24743203

Table 11

The final approach is to have a dynamic combination of models per sector. In dependence of the sector, the in average best performing model class is chosen. The average can be distorted in this case because of outliers, but is as a rough measurement usable. This dynamic mixed approach leads to good results.

Sectors / Types:

• **Demographic** > **ARIMA AVG 5.67** - ETS AVG 9.32 – MAPA AVG 12.86 – dynamic AVG 9.25 – hybridT 10.10
• **Finance** > **ARIMA AVG 15.08** - ETS AVG 15.72 - MAPA AVG 16.32 – dynamic AVG 15.67 – hybridT 16.37
• **Industry** > ARIMA AVG 10.85 - ETS AVG 8.71 - MAPA AVG 8.62 – **dynamic AVG 8.03** **–** hybridT 8.64
• **Macro** > ARIMA AVG 6.714 – ETS AVG 6.177 - MAPA AVG 6.12 – **dynamic AVG 6.10** **–** **hybridT 6.10**
• **Micro** > ARIMA AVG 15.114 – ETS AVG 16.32 - MAPA AVG 15.83 – dynamic AVG 14.93 – **hybridT 14.24**

Conclusion

The report shows that not one technique led to an optimal solution. Experimentation and the combination of different methods and tools are required. This has implications for the conduction of forecasts. The report shows that in dependence of factors like sector, the forecast methods have a high variability in performance. The examination and systematic categorization of all these factors in depth would exceed the scope of this report.

The managerial consequence of this is that forecasting can only be automated to a degree. For sensible data and if highly accurate forecasts are required skilled workforce has to be used. The underling data has to be explored in depth, or has to labeled and pre examined to make automated forecasting a suitable solution.

This doesn’t mean that the Batch forecasting solution, especially when combined with dynamic methods, can not give good estimates and reliable forecasts. In most of the data series, the forecast quality was acceptable.

Bibliography

Armstrong, J.S., 2001.
Combining Forecasts. Springer US, pp. 417-439.

Assimakopoulos, V. and Nikolopoulos, K., 2000.
The theta model: a decomposition approach to forecasting.
International Journal of Forecasting, 16(4), pp. 521-530.

Hyndman, R.J.A., George, 2018.
Forecasting: Principles and Practice. 3 ed.: OTexts.

Kim, S. and Kim, H., 2016.
A new metric of absolute percentage error for intermittent demand forecasts.
International Journal of Forecasting, 32(3), pp. 669-679.

Taylor, J.W., 2003.
Exponential smoothing with a damped multiplicative trend.
International Journal of Forecasting, 19(4), pp. 715-725.