3.3 Artificial Neural Network

Artificial Neural Networks (ANN) are nonlinear autoregression models, biomimetically inspired by the neurons in the biological brain. An ANN consists of a number of artificial neurons that can pass signals of varying strength to each other (see Fig. 2). If the combined incoming signals are strong enough, the neuron becomes activated and the signal travels to other neurons connected to it.

ANN's have to be trained from examples, and cannot be explicitly programmed. That is why this model is often applied to problems where the solution is difficult to express in a traditional computer programming language.

3.4 Vector Auto Regression

The Vector Auto Regression (VAR) model is a multivariate generalization of the AR model. The VAR-model allows the inclusion of time series which are expected to be linearly interdependent of each other. Each variable has its own equation containing its own lagged values and those of the other variables in the model and is therefore explained by its own history and that of the other variables.
An example of a VAR-model based on 3 variables $Y_{1,t}$, $Y_{2,t}$ and $Y_{3,t}$ is shown in (3).

$$\begin{aligned}\hat{Y}_{1,t} &= c_1 + \sum_{i=1}^{p}(\alpha_{1,i}Y_{1,t-i} + \beta_{1,i}Y_{1,t-i} + \gamma_{1,i}Y_{1,t-i}) + \epsilon_{1,t}\\ \hat{Y}_{2,t} &= c_2 + \sum_{i=1}^{p}(\alpha_{2,i}Y_{2,t-i} + \beta_{2,i}Y_{2,t-i} + \gamma_{2,i}Y_{2,t-i}) + \epsilon_{2,t}\\ \hat{Y}_{3,t} &= c_3 + \sum_{i=1}^{p}(\alpha_{3,i}Y_{3,t-i} + \beta_{3,i}Y_{3,t-i} + \gamma_{3,i}Y_{3,t-i}) + \epsilon_{3,t}\end{aligned} \quad (3)$$

In [3] and [4] economic variables were used as extra variables. Economic variables reflect the state of the economy and it is assumed that this influences potential customers in their decision whether or not to purchase a new car. For this research several combinations of economic time series were considered to be included in the model. Based on their performance, Job Vacancies Index and Car Prices Index were selected as explanatory variables in this model.

3.5 Theta

The theta method [7] has caught interest in academic circles and among forecast practitioners due to its remarkable good performance for monthly series at the M3-forecasting competition [8]. The original description of this univariate model is rather involved. It is based on decomposition of the time series through second order differences into so-called Theta-lines to capture long-term behavior and short-term features separately. Hyndman and Bilah [9] however, found that for a large training set the Theta method is equivalent to simple exponential smoothing with drift.

3.6 Naive Seasonal

Finally, as a benchmark to evaluate the performance of the prediction models, the average monthly figures are used as a naive forecast. In an activity as future prediction it is recommendable to check if a sophisticated model is indeed an improvement with respect to simpler methods because, as said before, sometimes simple models perform better than difficult ones. Several common performance metrics (see Section 5.1) for the results of the models described above are calculated and compared to those of the naive seasonal model.

4 Results

The data were split into a training set and a test set. For the training set we used the monthly data from January 2007 to April 2016. This time series contains 112 data points. The above explained forecasting methods were applied to the training set to estimate the parameters of these models. After this, the trained models were used to create forecasts for the test set, the 12 months period from May 2016 to April 2017.
The results were calculated using the statistical open source language R. The univariate Holt-Winters ETS, ARIMA, NNET and Theta models have been estimated using the forecast package [10] while the VAR model was established using vars package functions [11].

4.1 Check of the Model Output Properties

To see whether the predictive models could be improved a few checks have been conducted on its residuals.

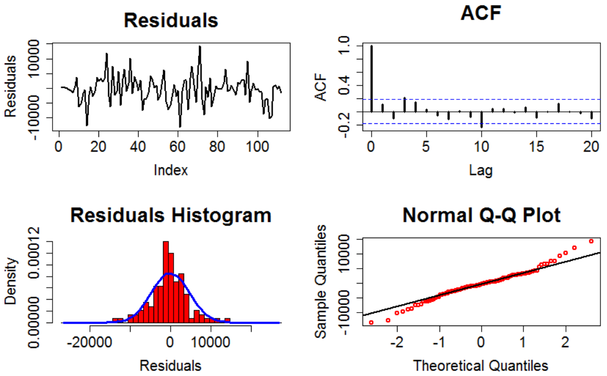

Figure 3: The residuals of the Holt-Winters model have more or less a constant variance, no significant autocorrelations and approximately a Normal distribution with mean zero.

In the top left of Fig.3 the in-sample residuals of the Holt-Winters model forecast are displayed. One can visually establish that the variance is more or less constant over time. Furthermore, in the top right of Fig. 3 one can see that no autocorrelations at lags 1-20 of the in-sample forecast errors greatly exceed the significance bounds at the dotted lines. This indicates that there is little evidence of non-zero autocorrelations at lags 1-20. In the lower part of the figure a histogram of the residuals with overlaid normal curve and a QQ-plot is displayed. From these pictures it seems plausible that the forecast errors are normally distributed with mean zero.

We may conclude that the Holt-Winters model fits the new car registrations appropriately and that it provides a forecast that probably cannot be improved.

The same analysis as above has been conducted with respect to the other models.

For establishing the ARIMA model, the Box-Jenkins methodology [12] was applied which lead to an ARIMA(2,1,0)(1,0,0)$_{12}$ model i.e. a differenced second order autoregressive model with a first order seasonal component of 12 months. The in-sample residuals and its properties are displayed in Fig.4.

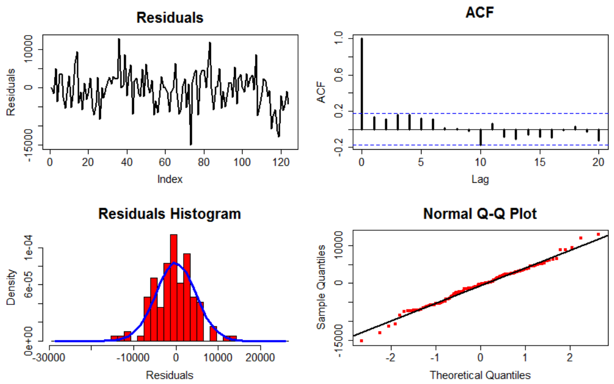

Figure 4: The ARIMA model residuals have more or less a constant variance, no autocorrelations and approximately a Normal distribution with mean zero.

A neural autoregressive network was estimated with 4 hidden nodes, 7 time lags and a seasonal component. See Fig.5 for its residual properties. Again the conclusion can be drawn that the model fits the data quite well and probably cannot be further improved without changing the model itself.

The VAR model used in this project contains 2 explanatory economic variables, namely Job Vacancies Index (JVI) and Car Prices Index (CPI). The rationale behind this choice is that demand for new cars tends to raise when more people acquire a (better) job and/or when car prices are low.

Several conditions are necessary to be fulfilled when creating a VAR model. First, it was checked whether these three time series are stationary. Using the Augmented Dickey Fuller test it was found that differencing was required in the case of JVI and CPI to acquire stationarity.

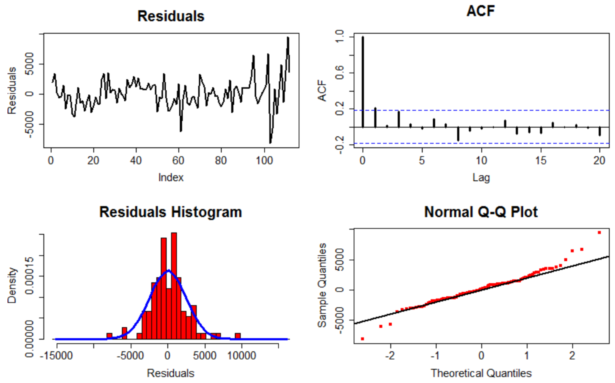

Figure 5: In-sample residuals plots of the ANN model.

Secondly, the autocorrelations of each of the time series were checked. See Fig.6 for the autocorrelation plots of the new car registrations. It shows that 3 lags seem appropriate for the AR part of the model.

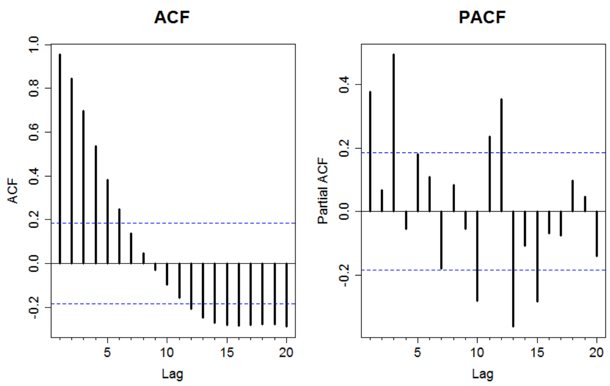

Figure 6: Plots of the autocorrelations and partial autocorrelations of the deseasonalized time series of the number of new car registrations.

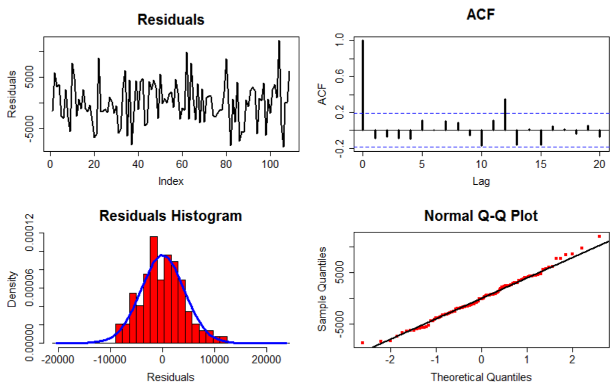

Figure 7: The residuals of the VAR model have more or less a constant variance, no autocorrelations and approximately a Normal distribution with mean zero.

This choice was confirmed by Akaike's Information Criterion. Finally, the model was estimated using the training data and its residuals were analyzed (see Fig.7). From the figure the residual properties are deemed satisfactory.

Finally, the Theta model was estimated. As with the other models the residuals were checked graphically (see Fig. 8) and it was concluded the model cannot be further improved.

In short, the residuals of all 5 estimated models have the same desirable properties (zero mean, constant variance, Normally distributed).

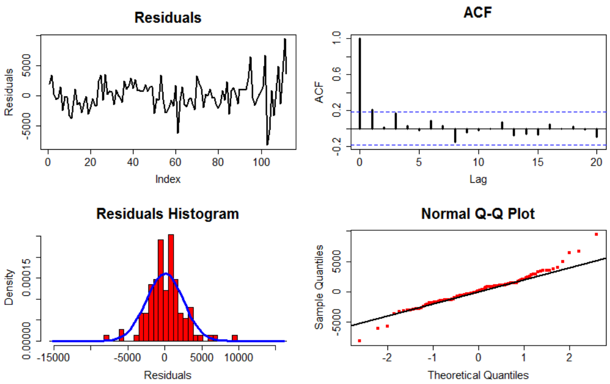

Figure 8: Several displays of the residuals of the Theta model to show its properties. The residuals seem more or less to have the necessary properties.

4.2 Establishment of the Ensemble

In [5] it was recommended to average the results of at least 5 different forecasting models which do not correlate positively. In such a case forecasting errors tend to single each other out which would result in a more accurate forecast. The correlations between the model forecasts were calculated and displayed in Fig.9. It can be seen that the outcomes of some models correlate exceptionally well.

Figure 9: Some of the forecast results of the models under consideration are highly correlated. Therefore, it was decided to remove the ARIMA and VAR models from the Ensemble forecast.

An explanation for this could be that ARIMA and the Holt-Winters models are both linear functions of lagged values of the time series and since both models are optimized on the data they have probably assigned comparable weights to them. The high correlation between the VAR model and Theta model is the result of their property that after a few time steps the forecasts tend to converge to the seasonal pattern.

Heeding the recommendation above, it was decided to leave out the least performing two of these models (i.e. ARIMA and VAR) and combine the remaining three models into an ensemble forecast.

The real data of the test period, the forecasts of all the models and the Ensemble based on three models have been plotted together in Fig.10 for visual inspection.

5 Performance Evaluation

In this section the model forecasts and the forecast of the combined models at the test data period are compared with the real outcomes. This gives an impression about the performance capability of these models if they were to be used for real forecasts.

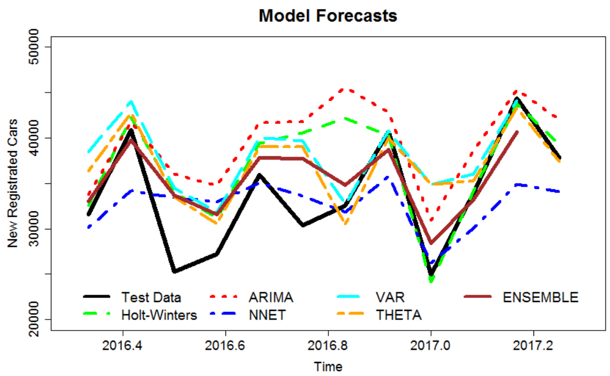

Figure 10: Forecast results of 5 models on the 12 months test set. The thick black line represents the real outcome of the new car registrations time series and the brown line is the Ensemble forecast.

5.1 Forecasting Performance

The models' performances are compared with each other using several standard evaluation metrics which are based on the forecast errors $e_i = y_i - \hat{y}_i$.

The Mean Error

$$ME = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t) \tag{10}$$

The Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{11}$$

The Mean Absolute Prediction Error

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{|y_t|}100\% \qquad (12)$$

The Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \qquad (13)$$

The Maximum Absolute Error

$$MXAE = \max_t |y_t - \hat{y}_t| \qquad (14)$$

Obviously, the lower the outcomes of these metrics the better the performance of the corresponding model. The metrics are applied to each model and are listed in Table 1 for easy comparison.

Forecasting Performance Metrics Calculated over a 12 Months Test Set					
Model \ Metric	ME	MAE	MAPE	RMSE	MXAE
ETS	-3188	3580	11.7%	5079	10200
ARIMA*	-5790	5790	19.1%	7018	12994
NNET	*1104*	4215	12.6%	5041	9469
VAR*	-4463	4517	15.6%	5785	9841
THETA	-3080	3796	13.1%	4975	9989
NAIVE*	-3706	3722	12.9%	4991	9555
ENSEMBLE	-1955	*3348*	*11.1%*	*4113*	*8391*

Table I: Comparison of the forecasting performances of the 5 models individually, the Naive seasonal model and the Ensemble. Models indicated with an "*" are not included in the Ensemble. The best performance figures are printed in italics.

6 Conclusion

In this paper 5 common forecasting models have been applied on a training set of 112 months of new car registrations. The results of a 12 months ahead forecast on a test set were evaluated. It turned out that the Exponential Smoothing outperformed the other individual models.

Interestingly, the naive seasonal model performed better than ARIMA and VAR, which happen to be the most sophisticated models in this study.

Furthermore, it was shown that the Ensemble forecast based on the ETS, NNET and Theta models performs slightly better than each model individually which is in confirmation of [5]. This shows that even with only a few available models, one can improve forecast accuracy.

Acknowledgements:
New car registration data was kindly provided by the Netherlands Vehicle Authority (Dutch: Rijksdienst voor het Wegverkeer (RDW)).
Special thanks goes to Aernout van Enter and Rudolf Roelfszema for their help and interesting discussions.

References:

[1] B. Brühl, M. Hülsmann, C.M. Friedrich, D. Reith, A sales forecast model for the German automobile market based on time series analysis and data mining methods, *ICDM 2009: Advances in Data Mining. Applications and Theoretical Aspects*, pp. 146-160.

[2] M. Hülsmann, D. Borscheid, C.M. Friedrich, D. Reith, General sales forecast models for automobile markets based on time series analysis and data mining techniques, *Transactions on Machine Learning and Data Mining*, (5)2, 2012, pp. 65-86.

[3] A. Sa-ngasoonsong, Multi-step sales forecasting in automotive industry based on structural relationship identification, *Int. J. Production Economics* (140), 2012, pp. 875-887.

[4] D. Fantazzini, et al., Forecasting German car sales using Google data and multivariate models, *Int. J. Production Economics*, 170A, pp. 97-135, 2015.

[5] J. Scott Armstrong, *Combining Forecasts Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, 2001.

[6] C.C. Holt, Forecasting trends and seasonals by exponentially weighted averages, Carnegie Institute of Technology, Pittsburgh ONR memorandum, no. 52. 1957.

[7] V. Assimakopoulos, K. Nikolopoulos, The theta model: a decomposition approach to forecasting, *Int. J. of Forecasting*, (16), 2000, pp. 521–530.

[8] S. Makridakis, M. Hibon, The M-3 Competition: results, conclusions, and implications, *Int. J. of Forecasting*, (16), 2000, pp. 451–476.

[9] R.J. Hyndman, B. Billah, Unmasking the Theta method, *Int. J. of Forecasting*, (19)2, 2003, pp. 287-290.

[10] R.J. Hyndman, Y. Khandakar, Automatic time series forecasting: the forecast package for R, *J. of Stat. Software*, (26)3, 2008, pp. 1–22.

[11] B. Pfaff, Package Vars, 2007. http://cran.r-project.org/web/packages/vars/index.html.

[12] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.