02417 Time Series Analysis

Assignment 1

Anna Raabo, s203691
Frederik Skou Fertin, s203679
Jacob Sterup Skaarup, s203693
Thomas Reenberg Trosborg, s203658

2024-02-23

Contents

1 Plot Data

1.1 Creation of time variable x

The data used in this assignment includes the total number of registered vehicles, $\boldsymbol{Y}$, for each month, $\boldsymbol{x}$, over a time period. The training data runs from January 2018 to November 2022, whereas the test data runs from December 2022 to November 2023. The time series data, with $x_0 = 2018$, is shown in fig. 1.

1.2 Description of time series

The series has a general, linearly increasing trend. The trend seems to have a seasonal variation with a steeper slope during the first half of each year compared to the second halves of the years where the increase stagnates. The latest year, 2022, seems to introduces a new, changed trend, which has the same seasonal variation in slope, but the general slope is flatter and an actual decrease in number of registered vehicles is seen during the second half of 2022.

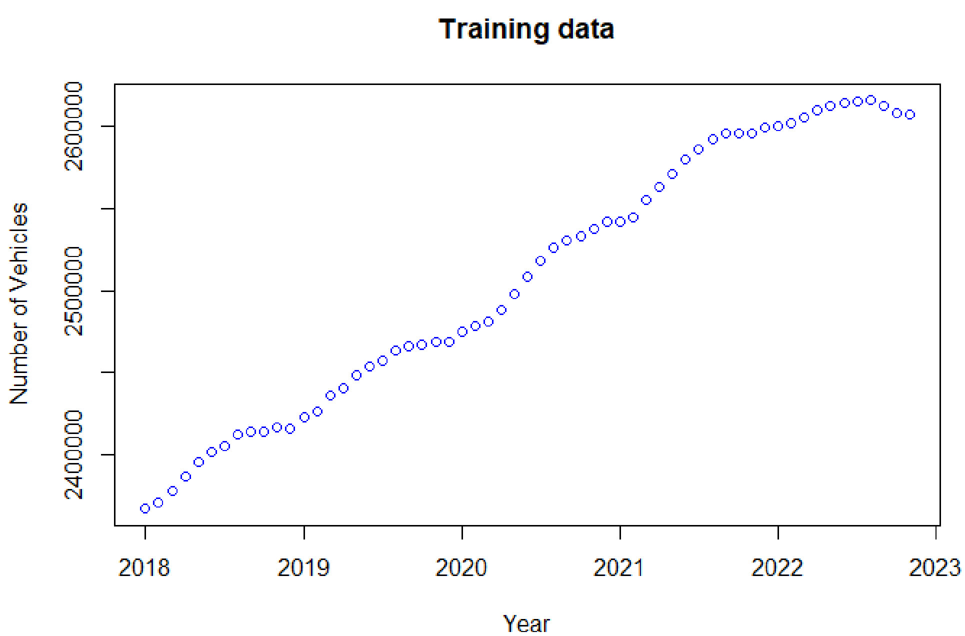


Figure 1: The time series of training data showing total number of vehicles in Denmark in each month.

2 Ordinary Least Squared model

2.1 Estimation of β_0 and β_1

A global linear model is introduced to try and model the number of vehicles in Denmark and takes the form of eq. (1). Here Y_t is the number of vehicles at time t, β_0 and β_1 are the model parameters to be determined, x_t the date (with monthly granularity), and finally the error ϵ_t is assumed to be a normally distributed random variable with mean $\mu = 0$, a fixed variance σ^2, and are mutually uncorrelated.

$$Y_t = \beta_0 + \beta_1 \cdot x_t + \epsilon_t \tag{1}$$

Based on the training set the best estimate of the parameters can be found by minimizing the squared errors of the model when fitted to the training data. These parameters can be determined deterministically using eq. (2) from theorem 3.1 [1].

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{x}^T\boldsymbol{x}\right)^{-1}\boldsymbol{x}^T\boldsymbol{Y}, \quad \hat{\boldsymbol{\theta}} = [\ \beta_0 \ \ \beta_1\ ]^T \tag{2}$$

2.2 Values of $\hat{\beta}_0$ and $\hat{\beta}_1$ and estimation of $\hat{\sigma}_{\hat{\beta}_0}$ and $\hat{\sigma}_{\hat{\beta}_1}$

The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ is found by applying the training data into eq. (2). The parameters are then found as:

$$\hat{\beta}_0 = -109\,499\,933 \quad \hat{\beta}_1 = 55\,438 \tag{3}$$

The intercept is at year 0, explaining why the model shows $\hat{\beta}_0$ = -109 million vehicles. According to the model, the number of vehicles increases by 55 437 vehicles per year. The standard errors, $\hat{\sigma}_{\hat{\beta}_0}$ and $\hat{\sigma}_{\hat{\beta}_1}$, are found by first determining the estimate of residual variance using eq. (4), and then the variance of parameters using eq. (5) with the training data based on theorem 3.2 from [1]. In eq. (4) the error between the fitted model and training data is ϵ, the number of observations in the training set n, and the number of parameters in the model p.

$$\hat{\sigma}^2 = \frac{\left(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)}{n-p} \tag{4}$$

$$\hat{Var}\left[\boldsymbol{\theta}\right] = \hat{\sigma}^2\left(\boldsymbol{x}^T\boldsymbol{x}\right)^{-1} \tag{5}$$

The value of the standard errors are shown in equation (6) and is simply the square root of the variance determined in eq. (5).

$$\hat{\sigma}_{\hat{\beta}_0} = 1{,}904{,}918 \quad \hat{\sigma}_{\hat{\beta}_1} = 942.83 \tag{6}$$

2.3 Forecast for the next 12 months

With the model parameters estimated it is possible to forecast the number of vehicles in Denmark for the following 12 months. The forecast is produced by using equation eq. (7) where $\hat{\boldsymbol{\theta}} = [\hat{\beta}_0\ \hat{\beta}_1]^T$.

$$\hat{Y}_t = \boldsymbol{x}_t\hat{\boldsymbol{\theta}} \tag{7}$$

The forecast is the best estimate for what will happen in the future based on the past as described by the model. However, there is uncertainty on the forecasts and this can be described by the prediction interval which is a confidence interval on the forecast. This is computed using eq. (8) taken from equation 3.61 in [1]. A 95% prediction interval is found by setting $\alpha = 0.05$, using the estimate of the variance of the errors found in eq. (4), the forecast of eq. (7), the training data dates x, and future dates x_t.

$$CI = \hat{Y}_t \pm t_{\alpha/2}(n-p)\cdot\hat{\sigma}\sqrt{1+\boldsymbol{x}_t^T\left(\boldsymbol{x}^T\boldsymbol{x}\right)^{-1}\boldsymbol{x}_t} \tag{8}$$

Here, $t_{\alpha/2}$ denotes the $1-\alpha/2$ quantile of the t-distribution. The forecast for 12 months ahead together with the prediction interval is reported in table 1.

Table 1: Forecast of the number of vehicles in Denmark 12 months ahead together with the 95% prediction interval boundaries.

Date	Number of vehicles	Lower prediction boundary	Upper prediction boundary
01-12-2022	2,646,082	2,624,807	2,667,358
01-01-2023	2,650,702	2,629,391	2,672,013
01-02-2023	2,655,322	2,633,975	2,676,670
01-03-2023	2,659,942	2,638,557	2,681,327
01-04-2023	2,664,562	2,643,138	2,685,986
01-05-2023	2,669,182	2,647,718	2,690,645
01-06-2023	2,673,801	2,652,297	2,695,306
01-07-2023	2,678,421	2,656,875	2,699,968
01-08-2023	2,683,041	2,661,451	2,704,631
01-09-2023	2,687,661	2,666,027	2,709,294
01-10-2023	2,692,281	2,670,602	2,713,959
01-11-2023	2,696,900	2,675,176	2,718,625

2.4 Plot of fitted model with training data and forecast values

Plotting the training data, the fitted model, and forecast with prediction interval results in fig. 2.

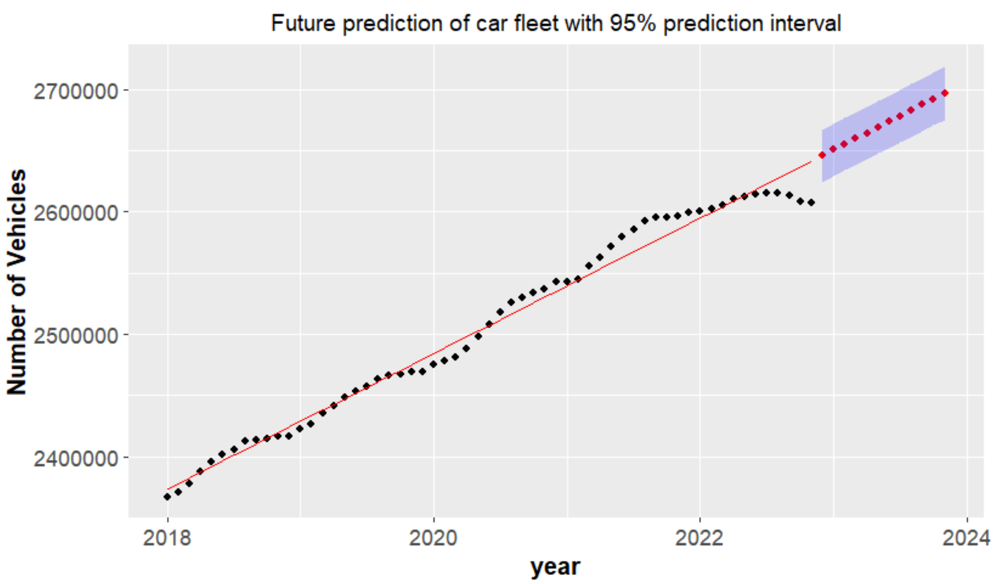


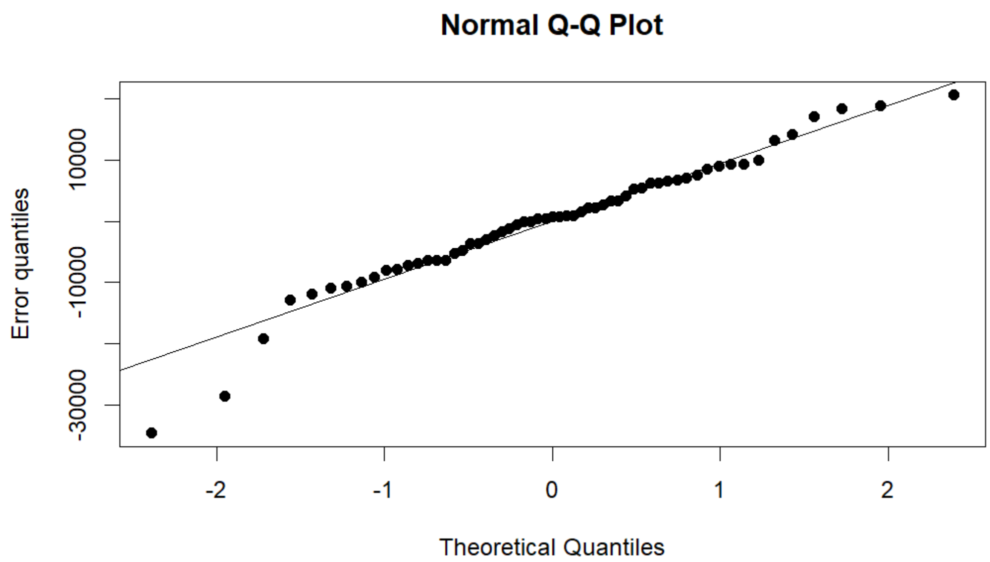
Figure 2: OLS forecast of the number of vehicles in Denmark. Training data in black, fitted model as the red line, predicted values in red, and 95% prediction interval in blue.

2.5 Comment on forecast

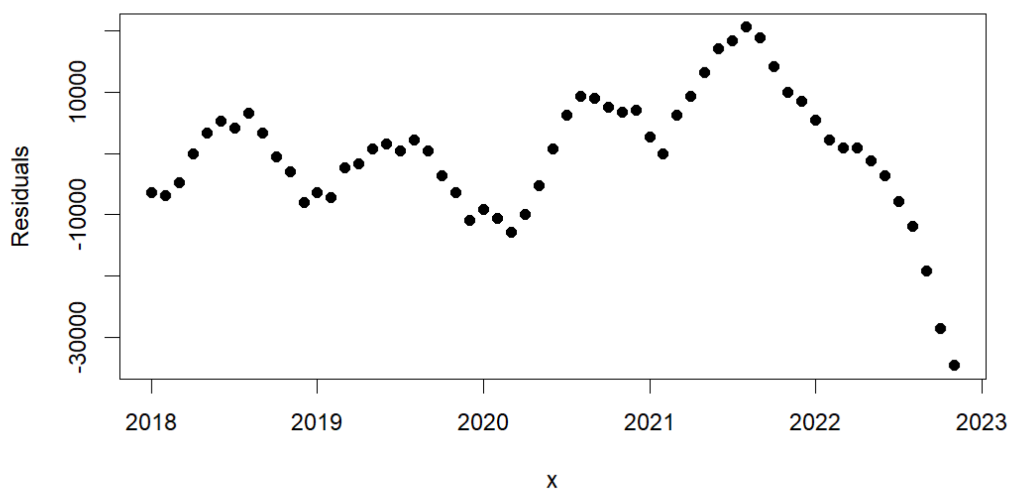
In fig. 2 the fitted model follows the historical data reasonably well as the number of vehicles has been rising almost linearly in the training period. However, near the end of the training period the number of vehicles in Denmark stops growing, which has little effect on the model parameters due to the few observations. As a result the model forecasts a continued rise in the number of vehicles in Denmark, while it is likely that a plateau or at least less growth will occur during the following years. Had the model put greater emphasis on more recent data the slope would be flatter and the forecast would be lower which is expected to fit better with the recent data. As we do not yet evaluate the model against the test data it is difficult to estimate quantitatively how well the model is performing in its forecast 12 months into the future. It is however very unlikely, that the first months of forecast is predicting a confidence interval, which includes the actual future values, as it would require a significant jump in registered vehicles. The historical data shows no sign of the fast system dynamics that would be required for this to happen.

2.6 Examine the residuals

When defining the model it was assumed that the errors were normally distributed and mutually uncorrelated. Whether these assumptions were reasonable can now be checked. In fig. 3a a Q-Q plot of the historical residuals is shown to check if the residuals can be assumed to be normally distributed. In the lower end a few points deviate significantly, but the vast majority of the residuals lie on the expected line. It is therefore expected that the assumption of normal distribution is reasonable. In fig. 3b the residuals are plotted against time to check whether the errors are uncorrelated. It is clear that there is a pattern in the residuals and they do not appear to be uncorrelated. If a previous error was positive, the following is also likely to be positive and vice versa. The residual typically peaks during the summer and dips during the winter which could indicate a seasonal trend that is not considered in this simple model. With one of the key assumptions about the model residuals violated, it is likely that a different and more complex model will perform better.



(a) Q-Q plot of the residuals of the model on training data.



(b) Residuals of the model on training data

Figure 3: Check of whether the model error assumptions hold true.

3 WLS - local linear model

3.1 The variance-covariance matrix

The local linear trend model introduces a forgetting factor, λ, which is used to assume variances for historical data. The variance-covariance matrix for the local model with $\lambda = 0.9$ is a diagonal matrix with assumed variances of each observation in the known data. The variance is assumed to increase as the time series data points become more distant from the current time step and assumed less important for the current trends. The variance of each time step is defined by $1/\lambda^{N-j}$ from time step $j = \{1, .., N\}$ with j=N being the current time step. This assumption differs from OLS where the variance of each observation is assumed to be the same, which results in an OLS variance-covariance diagonal matrix of 1's. A snippet of the WLS variance-covariance matrix is shown below.

$$\Sigma_{WLS} = \begin{array}{c c} & \begin{array}{ccccc} & N-3 & N-2 & N-1 & N \end{array} \\ \begin{array}{c} \\ N-3 \\ N-2 \\ N-1 \\ N \end{array} & \begin{pmatrix} .. & .. & .. & .. & .. \\ .. & 1.37 & 0 & 0 & 0 \\ .. & 0 & 1.23 & 0 & 0 \\ .. & 0 & 0 & 1.11 & 0 \\ .. & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{9}$$

3.2 Visualizing the λ-weights

The λ-weights follow a similar pattern, but inversely computed to the variance-covariance diagonal. As the weights are computed by $w = \lambda^{N-j}$, the weights increase for the data as the data gets closer to N, the latest observation in the time series. The weights are shown in fig. 4. At the most recent time step the weight is 1, while the oldest observation has a weight of just 0.0022.

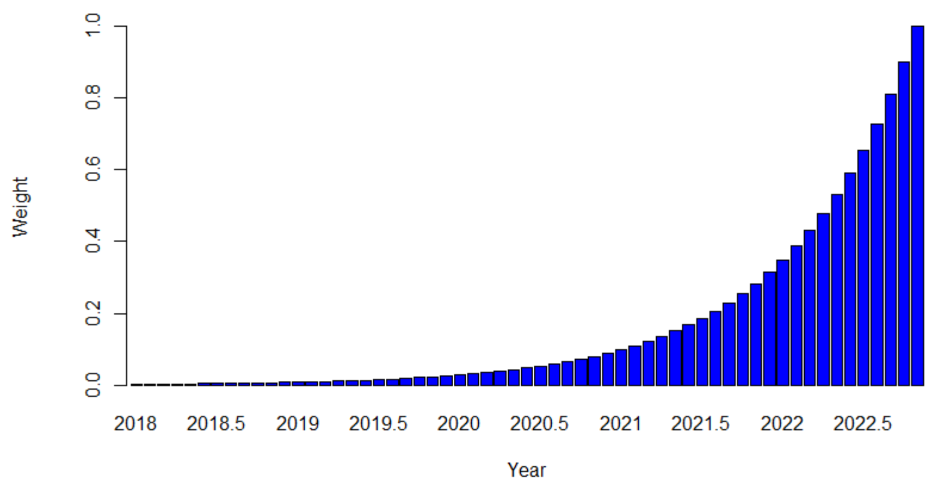


Figure 4: The weights of data points in the training data versus time. Weights are for the WLS model with $\lambda = 0.9$.

3.3 $\boldsymbol{T}$, the sum of weights in the WLS.

The sum of the weights - with $\lambda = 0.9$ - is here denoted by the symbol, $\boldsymbol{T}$. The equivalent, corresponding sum of weights for the OLS is $N = 59$, which is the number of data points in the training set. For the current data set $\boldsymbol{T}$ is calculated as following:

$$\boldsymbol{T} = \sum_{j=1}^{N} \lambda^{N-j} = 9.98 \tag{10}$$

The value $\boldsymbol{T}$ can be used in eq. (4) to help compute the expected variance of the model and thus also help compute confidence intervals for forecasts as in eq. (8).

3.4 Estimation of β_0 and β_1

Estimating the parameters of the local linear model follows nearly the same approach as the unweighted in eq. (2) but now includes the variance-covariance matrix to weight the observations. Including this results in eq. (11).

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\right)^{-1} \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Y} \tag{11}$$

The estimated values of the intercept $\hat{\beta}_0$, and the slope $\hat{\beta}_1$ are determined to be:

$$\hat{\beta}_0 = -84,089,780.95 \quad \hat{\beta}_1 = 42,868.56 \tag{12}$$

$\hat{\beta}_0$ and $\hat{\beta}_1$ of the WLS model with $\lambda = 0.9$ differ from the corresponding model parameters from the OLS, which estimated a steeper slope and a lower intercept than the WLS. This difference in model estimations stem from the fact that the WLS weighs the observations of 2022 higher where a decrease in the upwards trend can be seen. The WLS model predicts that 42.9 thousand more vehicles will be present on the Danish roads each year, and the negative intercept shows clearly that the model should not be used to try and describe time periods many years back as the model is clearly not valid when extrapolating further backwards. As the WLS model prioritizes recent trends more than the OLS, this is even more true for the WLS than for the OLS.

3.5 Forecasting 12 months from December 2022 to November 2023

Four WLS models have been constructed with different λ-values with the purpose of forecasting 12 months into the future. The WLS models have λ-values of 0.9, 0.8, 0.7, and 0.6. The estimated β-values can be seen in the following table 2.

Table 2: Estimated β-values of the four WLS models.

WLS model λ	$\hat{\beta}_0$	$\hat{\beta}_1$
0.9	-84.1e6	42,869
0.8	-37.8e6	19,965
0.7	-2.9e6	2,745
0.6	22.9e6	-10,025

The WLS models have been used to forecast the number of vehicles for each month from December 2022 to November 2023. Fig. 5 shows the estimated models in solid lines and the forecasted values for the next 12 months as dots in each model's respective color. The black dots are the historical data.

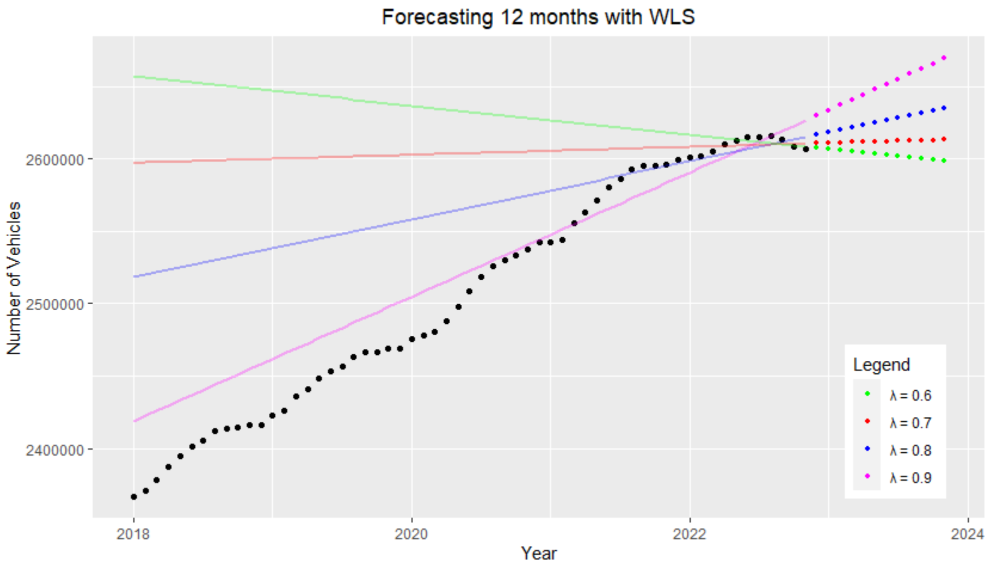


Figure 5: WLS models with varying λ-values. The lines shows the models' predictions of historical data, whereas the points show the forecasted values for the next 12 months.

3.6 Discussing the forecasts for different values of λ

The four WLS models vary significantly in the expected future trend of the number of vehicles on the Danish roads. The models with lower values of λ give relatively lower weights to old data, which gives the result of the model with $\lambda = 0.6$ actually expecting a negative future trend in number of vehicles, as the model gives large weight to the very recent negative trend in the training data. The other models, naturally, add a larger weight on older data, which results in the three other models predicting a positive slope in spite of the newest data showing a negative trend in the data. The slopes of the models thus does correspond to what we would expect when choosing the different λ-values.

3.7 Choosing the optimal model for forecasting

The model chosen for forecasting should depend on the amount of months ahead in time one is trying to predict. As discussed in section 3.6 the weights of old data is relatively lower, with a smaller value of λ, and relatively larger with a larger value of λ. Thus, for longer predictions, a higher value of λ is likely desirable, since the general trends in the historical data can be expected to be of significance. In this case, where a forecast of 12 months is needed, it should be expected of the model to depend on a larger amount of older data, than e.g. a forecast of only 1 month. Since there is a general positive trend for the training set, it would not be likely to follow the negative slope, shown for $\lambda = 0.6$, when predicting 12 months ahead, but it could make sense for the 1 month prediction horizon since the short term trend is likely to dominate the long term trends. Based on fig. 5, it seems the models with $\lambda = 0.7$ and 0.8 have a better fit with the data from the last 12 months and they are likely most suited for a 12-month forecast.

4 Iterative update and optimal λ

4.1 Provide $\boldsymbol{L}$ and $\boldsymbol{f}(0)$

When preparing the data for iterative updates of the model parameters, the time index variable is shifted to be $t = 0$ for the current time, and then the previous month is indexed by $t = -1$. Initially we assume that we only have access to the first observation in the training data, January 2018, which will be indexed by 0 for the first iteration of the WLS model updates.

On page 53 in [1] $\boldsymbol{f}(j)$ and $\boldsymbol{L}$ for the linear trend model are introduced and are written as in eq. (13) and eq. (14).

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{13}$$

$$\boldsymbol{f}(j) = \begin{bmatrix} 1 \\ j \end{bmatrix} \Rightarrow \boldsymbol{f}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{14}$$

4.2 Provide $\boldsymbol{F}_1$ and $\boldsymbol{h}_1$

$\boldsymbol{F}_N$ is defined by $\boldsymbol{F}_N = \boldsymbol{x}_N^T \Sigma^{-1} \boldsymbol{x}_N$ with N being the number of observations available in the training data. For $\boldsymbol{F}_1$ we only include the first observation.

$$\boldsymbol{F}_1 = \boldsymbol{x}_1^T \boldsymbol{x}_1 = \boldsymbol{f}(0)\boldsymbol{f}^T(0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{15}$$

$\boldsymbol{h}_N$ is defined by $\boldsymbol{h}_N = \boldsymbol{x}_N^T \Sigma^{-1} Y_N$ with N, similarly, being the number of observations available in the training data.

$$\boldsymbol{h}_1 = \boldsymbol{x}_1^T Y_1 = \boldsymbol{f}(0) Y_1 = \begin{bmatrix} 2,367,154 \\ 0 \end{bmatrix} \tag{16}$$

4.3 Update $\boldsymbol{F}_N$ and $\boldsymbol{h}_N$ recursively with $\lambda = 0.9$ and provide $\boldsymbol{F}_{10}$ and $\boldsymbol{h}_{10}$

$\boldsymbol{F}_N$ and $\boldsymbol{h}_N$ are now updated recursively by iterating through eqs. (17)-(18).

$$\boldsymbol{F}_{N+1} = \boldsymbol{F}_N + \lambda^N \boldsymbol{f}(-N)\boldsymbol{f}^T(-N) \tag{17}$$

$$\boldsymbol{h}_{N+1} = \lambda \boldsymbol{L}^{-1} \boldsymbol{h}_N + \boldsymbol{f}(+) Y_{N+1} \tag{18}$$

After 10 iterations the model can be described as:

$$\boldsymbol{F}_{10} = \begin{bmatrix} 6.51 & -23.75 \\ -23.75 & 137.46 \end{bmatrix} \tag{19}$$

$$\boldsymbol{h}_{10} = \begin{bmatrix} 15\,628\,557 \\ -56\,709\,556 \end{bmatrix} \tag{20}$$

The model after 10 iterations is further illustrated in fig. 6.

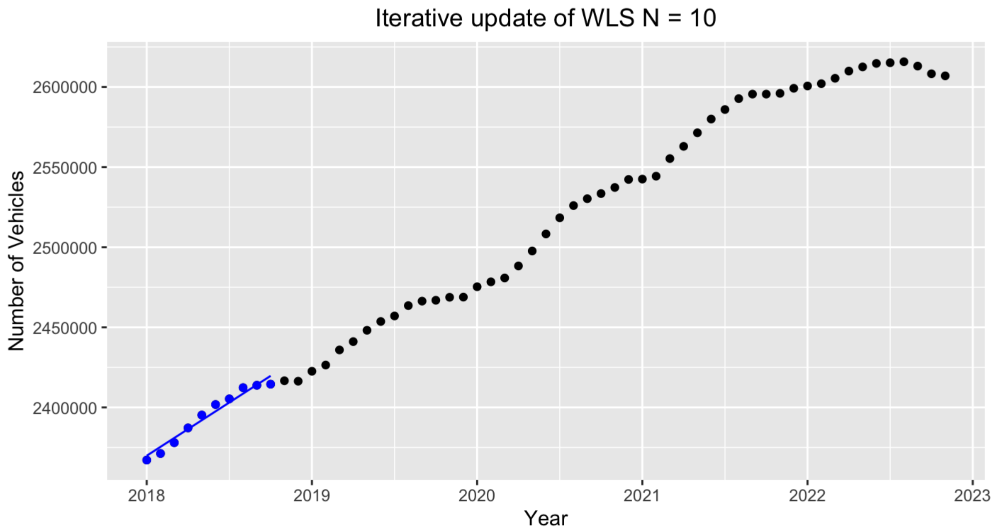


Figure 6: Resulting model after 10 update steps with $\lambda = 0.9$.

4.4 Update model up to $\boldsymbol{F}_{59}$ and $\boldsymbol{h}_{59}$

Again, using eqs. (17)-(18) the model is updated recursively to $N = 59$. At each iteration, predictions are calculated 1, 6, and 12 months into the future. This is done by estimating the model parameters $\hat{\boldsymbol{\theta}}_N$ based on $\boldsymbol{F}_N$ and $\boldsymbol{h}_N$ and subsequently, using $\hat{\boldsymbol{\theta}}_N$ to calculate predictions:

$$\hat{\boldsymbol{\theta}}_N = \boldsymbol{F}_N^{-1}\boldsymbol{h}_N \tag{21}$$

$$\hat{Y}_{N+\ell|N} = \boldsymbol{f}^T(\ell)\hat{\boldsymbol{\theta}}_N \tag{22}$$

Here, ℓ denotes how far into the future is being predicted and thus, $\hat{Y}_{N+\ell|N}$ denotes the prediction of the value of Y after $N + \ell$ steps when using the model at step N. Therefore, at each iteration, eq. (22) is used with $\ell = 1, 6, 12$ to calculate the desired predictions. The resulting predictions are plotted with the training data in fig. 7.

4.5 Plot 1-, 6-, and 12-month predictions

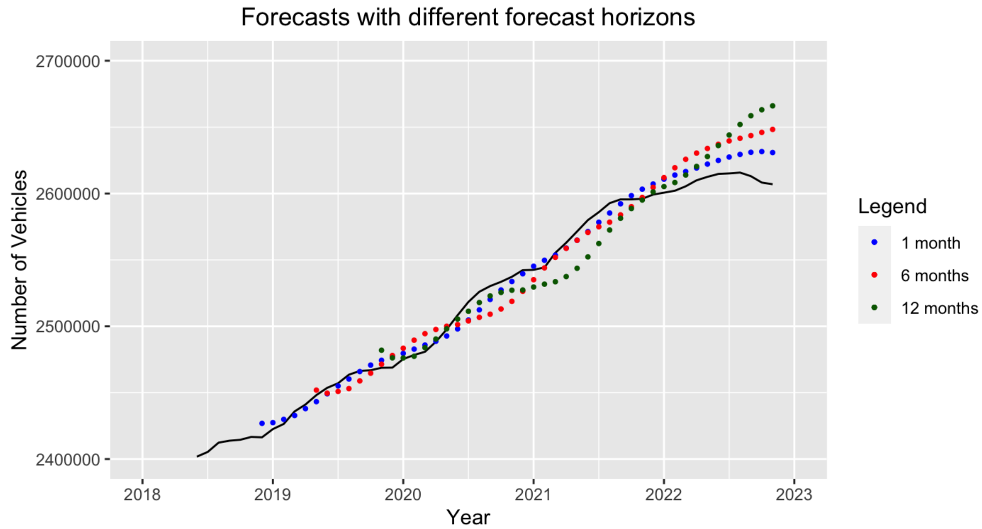


Figure 7: Model predictions for $N \in [11; 59]$ for $\ell = 1, 6, 12$ with $\lambda = 0.9$.

Since predictions are started at $N = 11$ the first three predictions are $\hat{Y}_{12|11}$, $\hat{Y}_{17|11}$, and $\hat{Y}_{23|11}$ which explains the difference in starting year for the three series of predictions. Naturally, as the 12-month predictions do not know the trends in the data at times $t \in [N : N + 12]$ it appears in the plot as if the model reacts slowly to the changes in slope. This is most evident from mid 2020 to mid 2021 and at the end of the period.

4.6 Iterative predictions with different values of λ

The iterative predictions performed with eqs. (17)-(18) and (21)-(22) are now calculated for forgetting factors between 0.55 and 0.95 with a 0.01 granularity. Subsequently, the root-mean-square of the prediction error (RMSE) is calculated in the following way:

$$RMSE = \sqrt{\frac{1}{N - (11 + \ell)} \sum_{t=11}^{N-\ell} (Y_{t+\ell} - \hat{Y}_{t+\ell|t})^2} \tag{23}$$

Given the different values of ℓ, the 1-, 6-, and 12-month predictions have different lengths (i.e., 48, 43, and 37, respectively). Therefore, the number of predictions from which the RMSE is calculated varies with ℓ leading to correspondingly different variances.

The resulting RMSE for each ℓ as a function of forgetting factor λ are plotted in fig. 8.

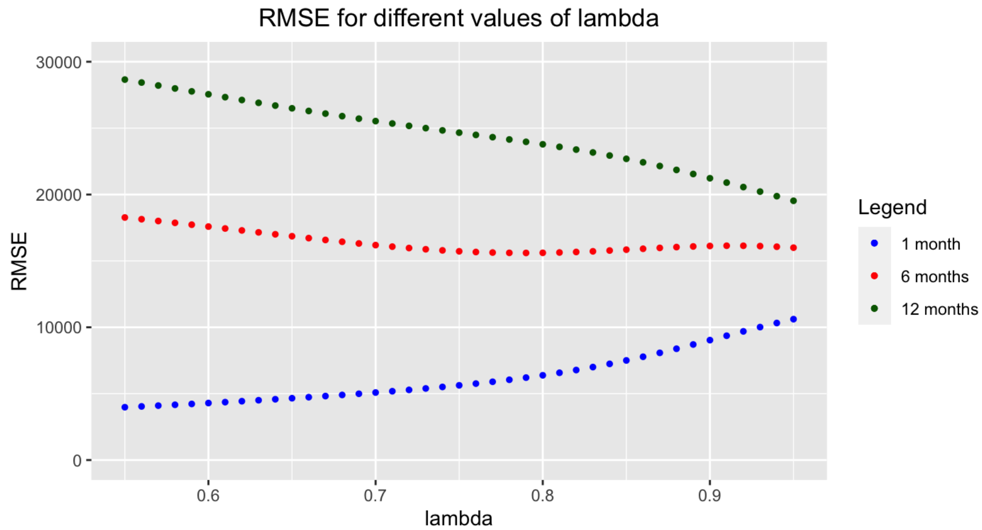


Figure 8: RMSE of prediction error as a function of forgetting factor λ.

The optimal forgetting factors for each prediction horizon are shown in tab. 3. The optimal factor is defined as the model which has the lowest RMSE during the iterative predictions on the training data.

Table 3: Optimal forgetting factors λ for different prediction horizons ℓ. Task 4.7, 4.8 & 4.9.

$\boldsymbol{\ell}$ [months]	1	6	12
λ	0.55	0.79	0.95

4.10 Minimum value of λ

If $\lambda = 0.5$ is inserted into eq. (10) the sum of the weights become 2. Since the variance of the residuals is estimated using eq. (24) no meaningful estimate of the variance can be made when $T \leq 2$ when estimating two parameters as here. In other word, the variance goes to infinity for T going to p.

$$\hat{\sigma}^2 = \frac{\left(\hat{Y} - \boldsymbol{x}_N \hat{\boldsymbol{\theta}}_N\right)^T \boldsymbol{\Sigma}^{-1} \left(\hat{Y} - \boldsymbol{x}_N \hat{\boldsymbol{\theta}}_N\right)}{T - p} \tag{24}$$

4.11 Naïve persistence model

The naïve persistence predictor simply predicts the data point at time $t+1$ to be equal to the value at time t:

$$\hat{Y}_{t+1|t} = Y_t \tag{25}$$

Due to its simplicity it serves as a great benchmark model to evaluate the performance of other models. A comparison of the 1-month ahead with its optimal $\lambda = 0.55$ and persistence model is shown in fig. 9. Naturally, as the general trend is increasing the persistence model generally has negative residuals whereas the 1-month predictor has both positive and negative residuals. When it comes to RMSE, the persistence model achieves an RMSE of 5330, while the 1-month ahead model performs better with an RMSE of 3981.

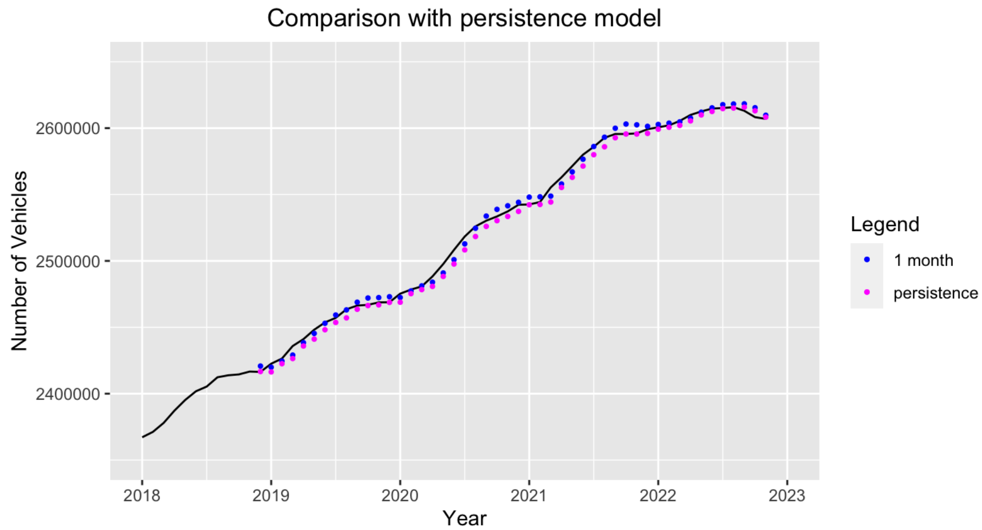


Figure 9: Predictions for local linear model with ℓ = 1 month and $\lambda = 0.55$ compared to persistence model.

4.12 Predicting into test data

For each forecast horizon the optimal λ is taken from table 3 and used in a local linear trend model with the given forecast horizon. The models are evaluated and plotted in fig. 10 with the training data as a solid line, the forecasts as orange points, and the realizations as black points.

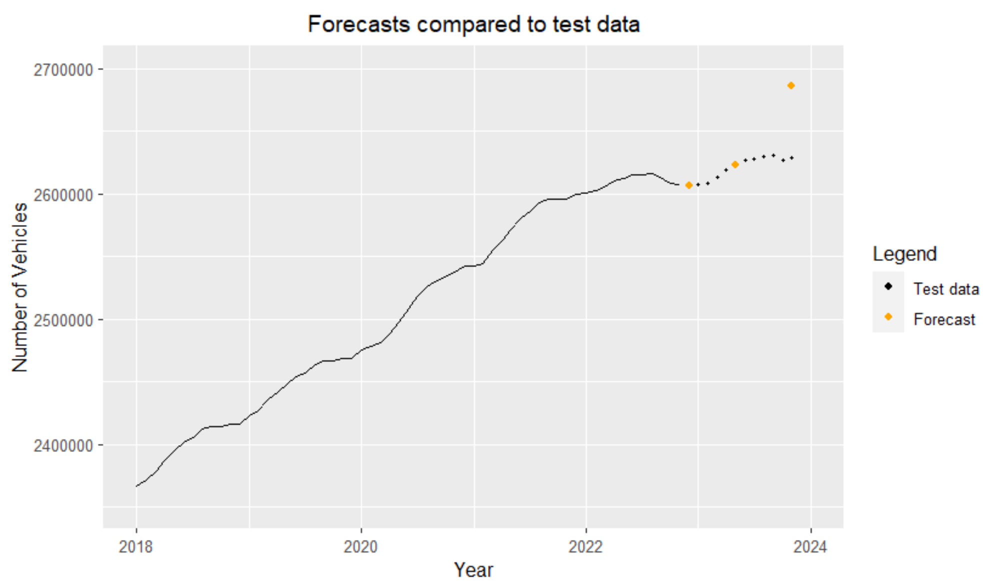


Figure 10: Predictions into test data for local linear model with ℓ = 1,6,12 months using optimal forgetting factors λ for each ℓ.

4.13 Prediction performance

In fig. 10 the forecasts for 1- and 6-months ahead fits well with the realization of the test data, while for 12-months ahead a large error is observed. In table 4 the errors at each step are reported. When considering that the total number of vehicles is around 2.6 million, the relative errors on 1- and 6-months ahead forecasts indicates near perfect forecasts, while the error on the 12-months ahead forecast is noticeable with a relative error of $\approx 2\%$.

Table 4: Prediction error on test data.

Forecast horizon [months]	1	6	12
Absolute error	1,294	869	57,278
Relative error	0.05%	0.03%	2.18%

The T-values for each λ-value and consequently each prediction horizon model is $T_{\lambda=0.95} = 19.0$, $T_{\lambda=0.79} = 4.8$, and $T_{\lambda=0.55} = 2.2$ using eq. (10). As mentioned, the T-value roughly corresponds to the value N in the global model, which denotes the number of data points trained on. For the three different prediction horizon WLS models, the T-value helps describe what trend of the training data can be expected to influence the local linear trend model. For the 1-month prediction the 2 most recent time steps define the trend, which shows in fig. 10 where the prediction is more or less a direct extrapolation of the two most recent data points. The model trend for the 12-month prediction can likewise also be somewhat interpolated by looking at the last 19 values of the training set, while the 6-month prediction horizon model trend cannot be fully described by the last 5 points, which highlights the fact that T should not be interpreted as a direct translation of N to local trend models since they do also give some weight to older data.

References

[1] H. Madsen, *Time Series Analysis*. Boca Raton, Florida: Chapman & Hall / CRC Press, 2007.