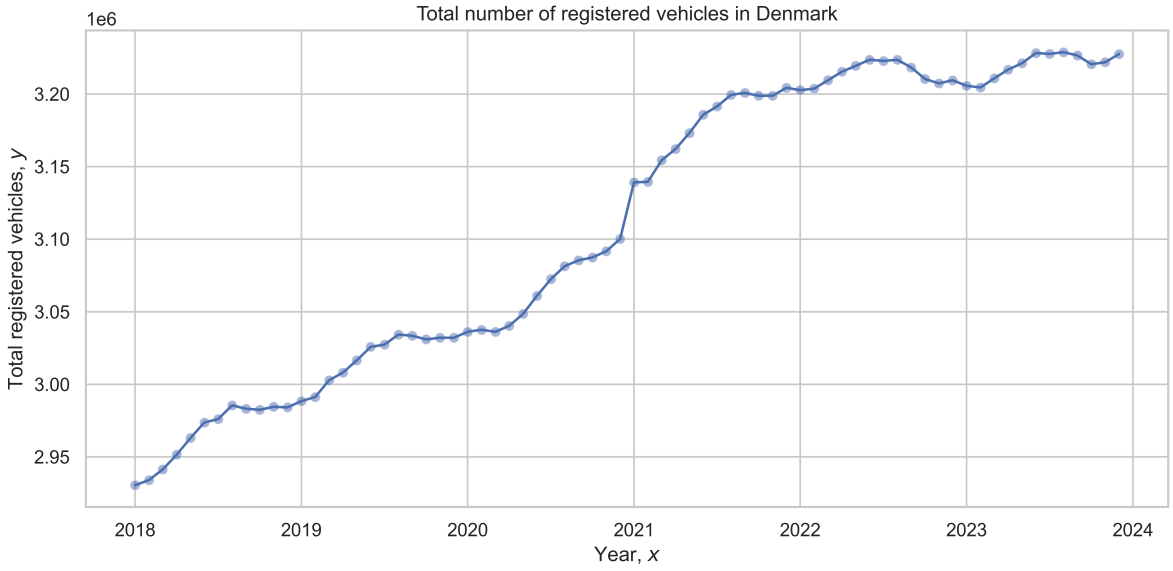
A Assignment 1

A.1 Plot Data

A.1.1 Construct time variable

A.1.2 Plotting observations



A.2 Linear Trend Model

We are given the General Linear Model (GLM) in sloppy notation:

$$Y_t = \theta_i + \theta_2 \cdot x_t + \epsilon_t \tag{1}$$

A.2.1 Matrix Form

We rewrite Eq. 1 as a matrix:

$$\underline{\mathbf{y}} = \underline{\underline{\mathbf{X}}}\,\underline{\boldsymbol{\theta}} + \underline{\boldsymbol{\epsilon}} \tag{2}$$

Where $\underline{\underline{\mathbf{X}}}$ denotes the *design matrix*, $\underline{\boldsymbol{\theta}}$ the *parameter vector*, and $\underline{\boldsymbol{\epsilon}}$ represents a stochastic noise term, $\underline{\boldsymbol{\epsilon}} \sim \mathcal{N}(0, \sigma^2)$.

In our case, the *design matrix* is constructed with an intercept θ_1 and a single trend parameter θ_2:

$$\underline{\underline{\mathbf{X}}} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \tag{3}$$

Using the first 3 data points of the given data, we can construct the following model:

$$
\begin{aligned}
\underline{\mathbf{y}} &= \underline{\underline{\mathbf{X}}}\underline{\boldsymbol{\theta}} + \underline{\boldsymbol{\epsilon}} \\
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} &= \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix} \\
\begin{bmatrix} 2930483 \\ 2934044 \\ 2941422 \end{bmatrix} &= \begin{bmatrix} 1 & 2018.000 \\ 1 & 2018.083 \\ 1 & 2018.167 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}
\end{aligned}
\tag{4}
$$

A.2.2 Parameter Estimation

We can estimate the parameters $\underline{\boldsymbol{\theta}}$ by deriving the *normal equations*:

$$
(2) \quad \Rightarrow \quad \underline{\underline{\mathbf{X}}}^T\underline{\mathbf{y}} = \underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\underline{\hat{\boldsymbol{\theta}}} \quad \Rightarrow \quad \underline{\hat{\boldsymbol{\theta}}} = \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\mathbf{y}} \tag{5}
$$

Where $\underline{\underline{\mathbf{X}}}$ has assumed to be invertible.

We additionally consider the standard errors of the parameter estimates, $\underline{\hat{\boldsymbol{\theta}}}$.

Under the assumption that the observations are described by Eq. 1, that is, the data is drawn from a *Simple Linear Model* overlaid with a stochastic (i.i.d) noise term, we find that the residuals of the mdoel are exactly the noise term, $\underline{\boldsymbol{\epsilon}}$.

$$
\underline{\boldsymbol{\epsilon}} = \underline{\mathbf{y}} - \underline{\hat{\mathbf{y}}} \tag{6}
$$

Where

$$
\underline{\hat{\mathbf{y}}} = \underline{\underline{\mathbf{X}}}\underline{\hat{\boldsymbol{\theta}}} \tag{7}
$$

For the benefit of the reader, we reproduce the relationship between the residuals and the covariance matrix, $\underline{\underline{\boldsymbol{\Sigma}}}$ [1, p. 36-37]:

From Eq. 5, we have:

$$
\begin{aligned}
\mathbb{E}\left[\underline{\hat{\boldsymbol{\theta}}}\right] &= \mathbb{E}\left[\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\mathbf{y}}\right] \\
&= \mathbb{E}\left[\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\left(\underline{\underline{\mathbf{X}}}\underline{\boldsymbol{\theta}} + \underline{\boldsymbol{\epsilon}}\right)\right] && (2) \\
&= \mathbb{E}\left[\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)\underline{\boldsymbol{\theta}}\right] && \underline{\boldsymbol{\epsilon}} \sim \underline{\mathcal{N}}(0, \sigma^2) \\
&= \underline{\boldsymbol{\theta}}
\end{aligned}
\tag{8}
$$

We consider the following:

$$
\begin{aligned}
\underline{\hat{\boldsymbol{\theta}}} - \mathbb{E}\left[\underline{\hat{\boldsymbol{\theta}}}\right] &= \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\mathbf{y}} - \underline{\boldsymbol{\theta}} \\
&= \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\left(\underline{\underline{\mathbf{X}}}\underline{\boldsymbol{\theta}} + \underline{\boldsymbol{\epsilon}}\right) - \underline{\boldsymbol{\theta}} \\
&= \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\boldsymbol{\epsilon}}
\end{aligned}
\tag{9}
$$

We can now evaluate the covariance of the predicted parameters, $\underline{\hat{\boldsymbol{\theta}}}$:

$$
\begin{aligned}
\operatorname{Cov}\left[\hat{\underline{\theta}}\right] &= \mathbb{E}\left[\left(\hat{\underline{\theta}} - \mathbb{E}\left[\hat{\underline{\theta}}\right]\right)\left(\hat{\underline{\theta}} - \mathbb{E}\left[\hat{\underline{\theta}}\right]\right)^{\mathsf{T}}\right] \\
&= \mathbb{E}\left[\left(\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\epsilon}\right)\left(\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\epsilon}\right)^{\mathsf{T}}\right] \\
&= \mathbb{E}\left[\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\underline{\epsilon}\underline{\epsilon}^T\underline{\underline{\mathbf{X}}}\left(\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\right)^{\mathsf{T}}\right] \\
&= \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\mathbb{E}[\underline{\epsilon}\underline{\epsilon}^T]\underline{\underline{\mathbf{X}}}\left(\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{\mathsf{T}}\right)^{-1} \\
&= \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\ \operatorname{Var}[\underline{\epsilon}\underline{\epsilon}^T]\underline{\underline{\mathbf{X}}}\left(\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{\mathsf{T}}\right)^{-1} \\
&= \left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}\underline{\underline{\mathbf{X}}}^T\sigma^2\underline{\underline{\mathbf{X}}}\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1} \qquad\qquad \underline{\epsilon} \sim \mathcal{N}(0, \sigma^2) \\
&= \sigma^2\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}
\end{aligned}
\tag{10}
$$

Where we understand the variances of the predicted variables to be the diagonal elements of the covariance matrix:

$$
\operatorname{Var}\left[\hat{\underline{\theta}}\right] = \operatorname{diag}\left(\operatorname{Cov}\left[\hat{\underline{\theta}}\right]\right) \tag{11}
$$

Which gives the following standard errors for the parameter estimates:

$$
\sigma_{\hat{\theta}_i} = \sqrt{\operatorname{Var}\left[\hat{\theta}_i\right]} = \sqrt{\sigma^2\left(\underline{\underline{\mathbf{X}}}^T\underline{\underline{\mathbf{X}}}\right)^{-1}_i} \qquad\qquad i \in \{1, 2\} \tag{12}
$$

Computing these for the entire training dataset gives:

$$
\begin{aligned}
\hat{\theta_1} &= (-110360 \pm 60) \times 10^3 \\
\hat{\theta_2} &= (3593.6 \pm 1.8) \times 10^3
\end{aligned}
\tag{13}
$$

Using these predicted parameters, we are able to produce an estimate of the vehicle registrations for the training dataset using the General Linear Model as seen in Figure 1.

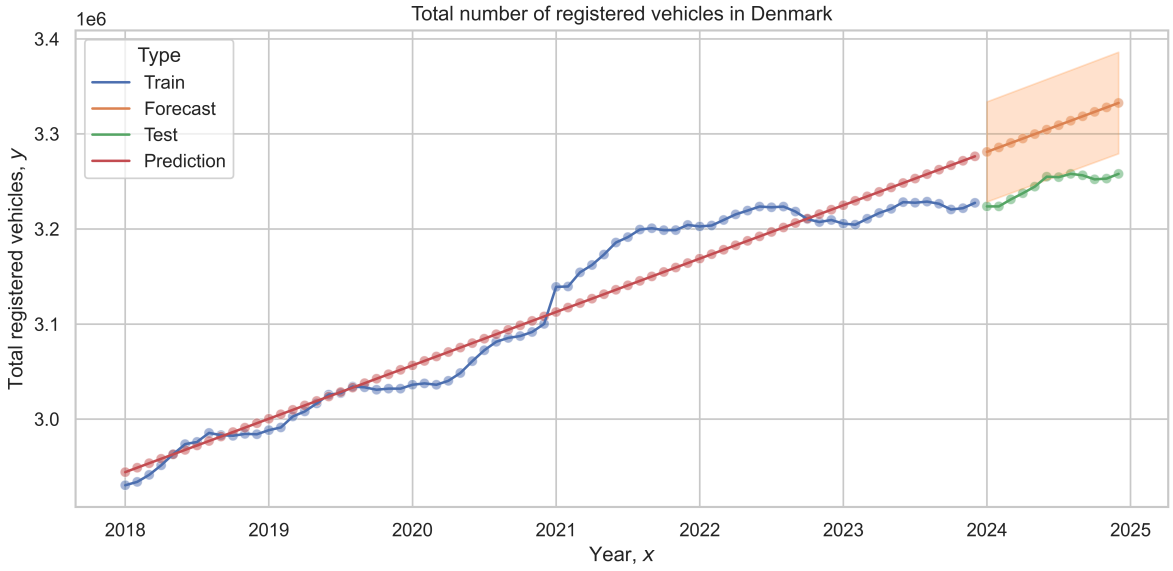


Figure 1: Estimation of vehicle registrations using the General Linear Model.

A.2.3 Prediction

We now wish to predict future vehicle registrations using our simple model. From Eq. 7, we see that doing so simply requires the construction of an appropriate design matrix, $\hat{\underline{\underline{\mathbf{X}}}}$. This may be constructed simply as:

$$\hat{\underline{\underline{\mathbf{X}}}}_{i} = \begin{bmatrix} 1 & 2024 + \frac{i-1}{12} \end{bmatrix} \qquad i \in \{1, 2, ..., 12\} \tag{14}$$

Where i indexes the rows of the design matrix.

Carrying out the forecasting as described by Eq. 7 along with appropriate estimation of the confidence interval of our prediction, we obtain Table 1. It should be noted that the estimation of the confidence interval valid only in the case where the observations are described by the General Linear Model.

Time	Total Vehicles Registered	95% Confidence Interval, lower bound	95% Confidence Interval, upper bound
2024.000	3281153	3228504	3333803
2024.083	3285832	3233123	3338540
2024.167	3290511	3237741	3343280
2024.250	3295189	3242358	3348021
2024.333	3299868	3246973	3352764
2024.417	3304547	3251586	3357508
2024.500	3309225	3256198	3362253
2024.583	3313904	3260808	3367000
2024.667	3318583	3265417	3371749
2024.750	3323262	3270024	3376499
2024.833	3327940	3274630	3381250
2024.917	3332619	3279235	3386003

Table 1: 12 month forecast of vehicle registrations in Denmark.

A.2.4 Plot of Forecast

We can now present the forecasted data in Table 1 along with the training, test, and prediction data sets

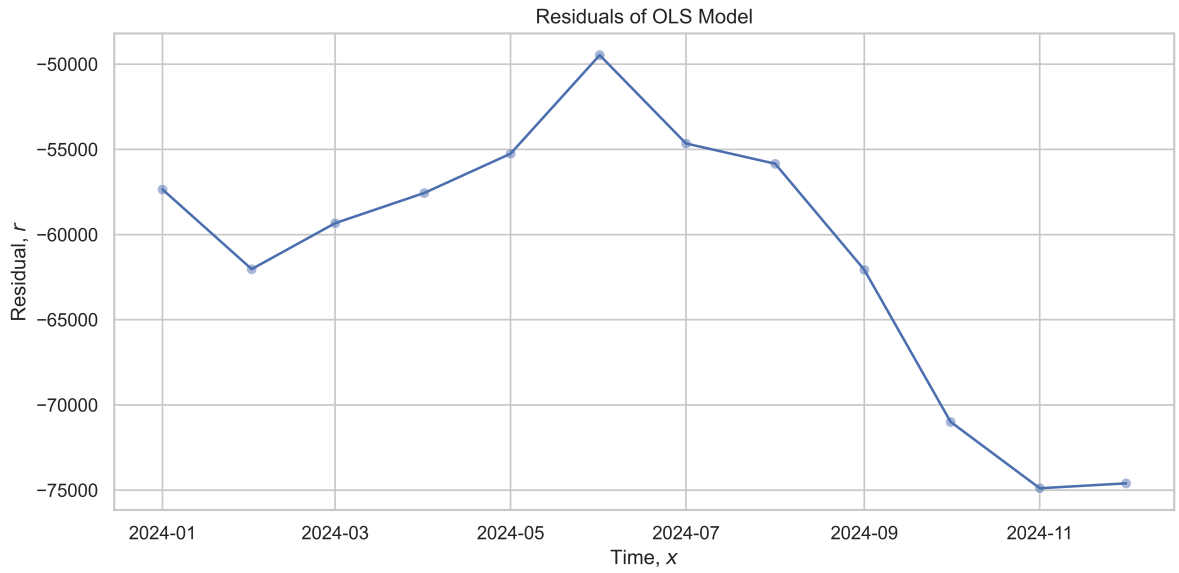


Figure 2: 12 month forecast of vehicle registrations in Denmark.

A.2.5 Commentary on Forecast

We find that the prediction is a relatively poor match against the test data, from which we understand that a Simple Linear Model is likely not an appropriate model for the data.

In particular, we observe significant local deviations from the model, which can be understood by considering the modelling domain. It is reasonable to assume that the number of registered vehicles will be influenced by market conditions, such as government subsidies, registration fees, and taxation schemes. Additionally we would expect supply chain disruptions to have strongly influence the contemporary pricing and availability of vehicles.

A better model would likely incorporate these factors. A model that incorporates locality without apriori domain knowledge could be a *Weighted Least Squares* model with local weights.

A.2.6 Residual Analysis

In order to substantiate Section A.2.5, we perform a residual analysis on the prediction and forecasting data.

Recalling the assumptions of our model, we expect the residuals to be normally distributed around zero:

$$\underline{\mathbf{r}} = \underline{\mathbf{y}} - \underline{\hat{\mathbf{y}}} \sim \mathcal{N}(0, \sigma^2) \tag{15}$$

It is readily apparent in Figure 3 that the residuals are not normally distributed, nor do they average to zero:

$$\mathbb{E}[\underline{\mathbf{r}}] = -8739 \tag{16}$$

As such, we conclude that the model is not appropriate to describe the data.

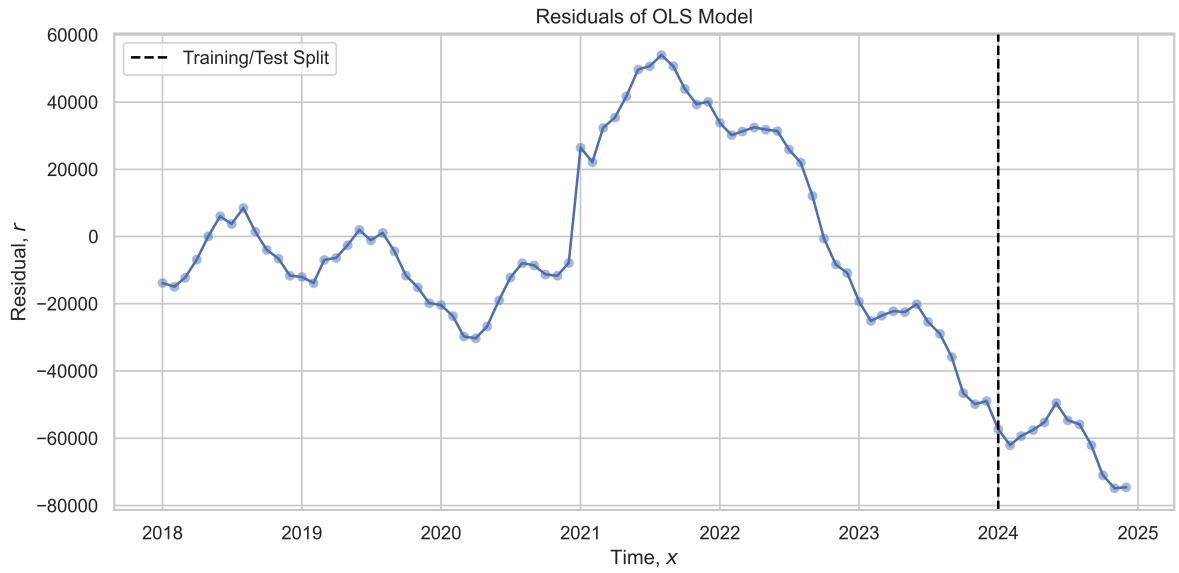


Figure 3: Residual analysis on prediction and forecasting of total vehicle registrations in Denmark. Dashed black line delineates the transition from prediction to forecasting.

Bibliography

[1] H. Madsen, *Time Series analysis*. Chapman & Hall/CRC, 2008.