# Assignment 1

## 2025-02-21

**Group members: Keerthi (s243933), Katarina (s243906), Hubert (s243896), German (s243660)**
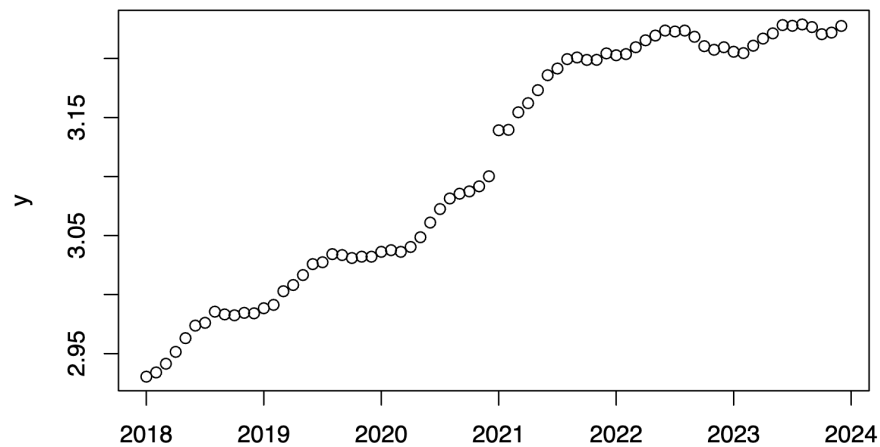
## 1 Plot data



Figure 1: The plot of observations.

**1.1.**

**1.2.** There seems to be a general positive trend of number of vehicles in Denmark over the years. However, there is also a seasonal pattern within each year where it increases for approximately the first half of a year and then decreases. We also noticed that there is a jump at the start of 2021, and also after that the trend seems to have gotten flatter but the seasonal pattern still exists.

## 2 Linear trend model

**2.1** The pictures are shown in Appendix.

## 3 OLS - global linear trend model

**3.1** We can estimate the parameters by solving the normal equation $\theta = (X^T X)^{-1} X^T y$.

**3.2** The parameters are represented in Table 1 and the estimated mean in Figure 2.

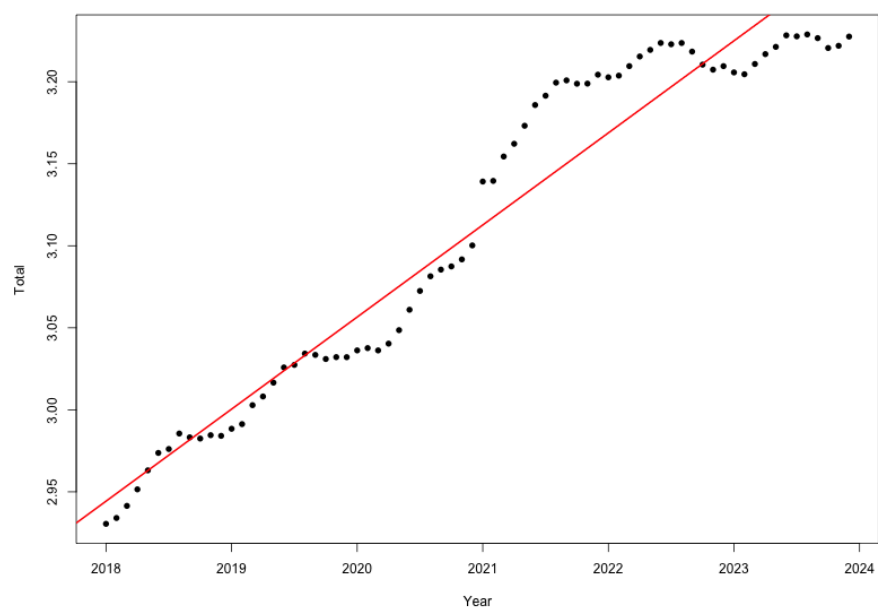**3.3** The forecast results are presented in Table 2.

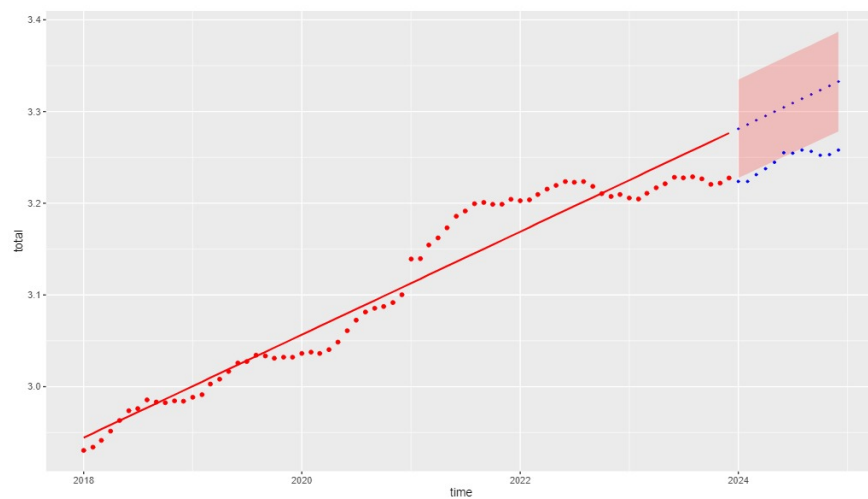Figure 2: Observations and OLS Estimated Mean



Figure 3: Fitted model together with the training data and the forecasted values.

| Parameter | Estimate $(\hat{\theta})$ | Standard Error $(\hat{\sigma}_{\hat{\theta}})$ |
|---|---|---|
| $\hat{\theta}_1$ | -110.36 | 3.59 |
| $\hat{\theta}_2$ | 0.056 | 0.0018 |

Table 1: Parameter Estimates and Standard Errors

Table 2: Forecast Results

|    | fit | lwr | upr |
|---|---|---|---|
| 73 | 3.281154 | 3.227579 | 3.334728 |
| 74 | 3.285832 | 3.232198 | 3.339467 |
| 75 | 3.290511 | 3.236815 | 3.344208 |
| 76 | 3.295190 | 3.241430 | 3.348950 |
| 77 | 3.299869 | 3.246044 | 3.353693 |
| 78 | 3.304547 | 3.250656 | 3.358439 |
| 79 | 3.309226 | 3.255267 | 3.363185 |
| 80 | 3.313905 | 3.259876 | 3.367934 |
| 81 | 3.318583 | 3.264483 | 3.372683 |
| 82 | 3.323262 | 3.269090 | 3.377435 |
| 83 | 3.327941 | 3.273694 | 3.382188 |
| 84 | 3.332620 | 3.278297 | 3.386942 |

**3.4** The plot is represented in Figure 3.

**3.5** The forecast appears to be limited as it overshoots the actual values with a slope that doesn't align with the patterns, especially during the test period. The inaccuracy is further demonstrated by most forecasted values falling outside the prediction interval. All in all, this is a basic forecast, following the general trend of the data, but not seasonal shifts.

**3.6** The residuals show confusing patterns, staying close to the reference line only in the middle while significantly deviating at both the beginning and end of the plot. The model assumptions seem to be violated as the errors aren't properly distributed around zero, with negative residuals indicating the forecast systematically overshoots actual values. This systematic error pattern suggests fundamental issues with the model specification rather than random noise, confirming the model assumptions are not fulfilled.

## 4 WLS - local linear trend model

**4.1** The variance-covariance matrix for the local model consists of the inverse of observation weights in the diagonal and zeros otherwise. $N$ is equal to the number of observation, which is equal to 72.

$$\Sigma_{WLS} = \begin{pmatrix} \frac{1}{\lambda^{N-1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda^{N-2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda^0} \end{pmatrix}$$

Conversely, the variance-covariance matrix for the global model contains 1 in the diagonal and zeros otherwise.

$$\Sigma_{OLS} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

3

**4.2** The highest weight is for the latest time point and it is equal to $\lambda^0 = 1$. The further in the past an observation point is, the more does its weight decrease. The weight distribution is visualized in Figure 4.
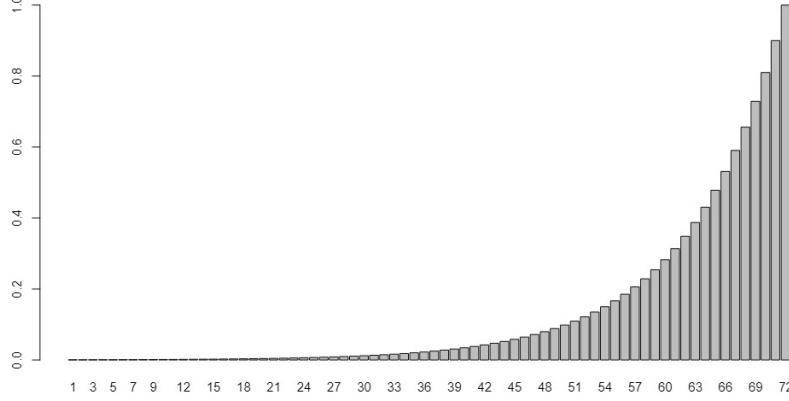


Figure 4: The weights of each observation in the train dataset.

**4.3** The sum of weights for the local model is $\sum_{n=1}^{N} \lambda^{n-1} = 9.994925$. The sum of weights for the global model is equal to the number of observations, 72.

**4.4** According to the local model with $\lambda = 0.9$, the parameters are equal to $\theta_1 = -52.4828617$ and $\theta_2 = 0.0275299$.

**4.5** Figure 5 shows the predictions made by OLS and WLS in red and blue respectively, along with the true predictions shown in black. From this, it can be seen that the WLS model prioritizes recent data points, creating a more gradual slope that adapts to current trends. While OLS treats all data equally and provides a generalized fit, WLS gives higher weight to recent observations. We would choose WLS when new predictions are crucial, and OLS when analyzing from a long-term perspective.

## 5 Recursive estimation and optimization of $\lambda$

**5.1** The pictures are shown in Appendix.

**5.2**

$$\hat{\theta}_1 = \begin{pmatrix} 7.196 \times 10^{-7} & 0.001452 \end{pmatrix} \hat{\theta}_2 = \begin{pmatrix} 9.775 \times 10^{-9} & 0.001453 \end{pmatrix} \hat{\theta}_3 = \begin{pmatrix} -3.696 \times 10^{-6} & 0.001455 \end{pmatrix}$$

The matrix calculations in RLS seem intuitive as they represent an iterative updating process. $R_t$ stores information about input correlations and $h_t$ stores the correlation between inputs and outputs. This allows us to efficiently update parameter estimates (without having to recompute using all historical data).

**5.3** The estimates of $\hat{\theta}_N$ and the OLS estimates are very far apart, with a difference of 104.81 for the first parameter. This is caused by the initial value of $R_0$ being too low. Surprisingly, reducing $R_0$ to 0.001 made the difference even worse. A larger $R_0$ would cause the estimates to be more influenced by the initial value, helping the RLS algorithm to converge faster to the OLS estimates.
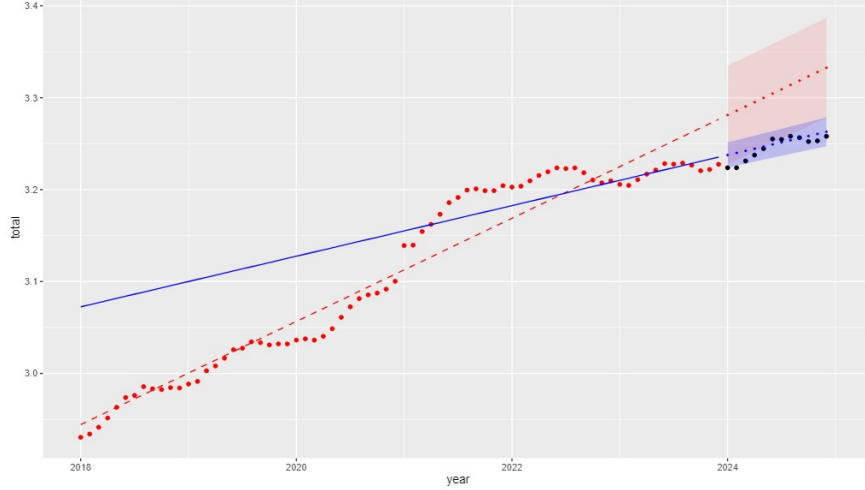
4

Figure 5: The observations as well as the predictions for the OLS & WLS models, including prediction intervals.

**5.4** The parameters are represented in Figure 6. The estimates of $\theta_1$ and $\theta_2$ show better stability with higher $\lambda$ values. Higher $\lambda$ places more emphasis on past observations. This results in smoother and more stable parameter estimates over time.

**5.5** The plot in Figure 7 shows the one-step-ahead residuals for $\lambda = 0.7$ in blue and $\lambda = 0.99$ in red. The residuals represent the difference between predicted and actual values at each time step. For $\lambda = 0.99$, the residuals are stable, showing that the model is more resistant to any recent fluctuations because of the higher weight given to past observations. In contrast, for $\lambda = 0.7$, the residuals diverge a lot more, suggesting that the model reacts stronger to recent data. This clearly shows how the choice of $\lambda$ affects the stability and adaptability of the recursive estimation process.

**5.6** The relationship between the forgetting factor $\lambda$ and the Root Mean Square Error (RMSE) across different forecast horizons ($k = 1, \ldots, 12$) is shown in Figure 8. A clear trend can be noticed where RMSE decreases as $\lambda$ approaches 1, meaning that a higher $\lambda$ results in lower prediction errors. However, for lower values of $\lambda$, the RMSE is much higher, which could mean that too much "forgetting" doesn't gives optimal predictions. These results would then suggest that the optimal choice of $\lambda$ lies within the range of 0.8 to 0.99, where RMSE is minimized. Moreover, since the RMSE patterns are the same across all horizons, I would use a single optimal $\lambda$ rather than letting it depend on the horizon.

**5.7** For $\lambda = 0.99$, the forecasts are represented in Figure 9, with the red points being OLS predictions, blue being WLS, and black being RLS.

**5.8** Time adaptive models are very useful for tracking changes of a dynamic system. However, they also come with a set of challenges that should be carefully considered or managed. One concern is to balance adaptation speed with stability. If $\lambda$ is too small, the model relies on recent data which makes is very sensitive to fluctuations and noise. On the other hand, if $\lambda$ is too high, our older data will have a stronger imact which then makes the model unresponsive to new trends. In other words, our model overfits or underfits and selecting the right $\lambda$ value is crucial for good generalization.

Time dependent data also introduces autocorrelation which makes cross-validation less effective. This is because splitting data randomly leads to data leakage where future data influences past predictions. Instead,
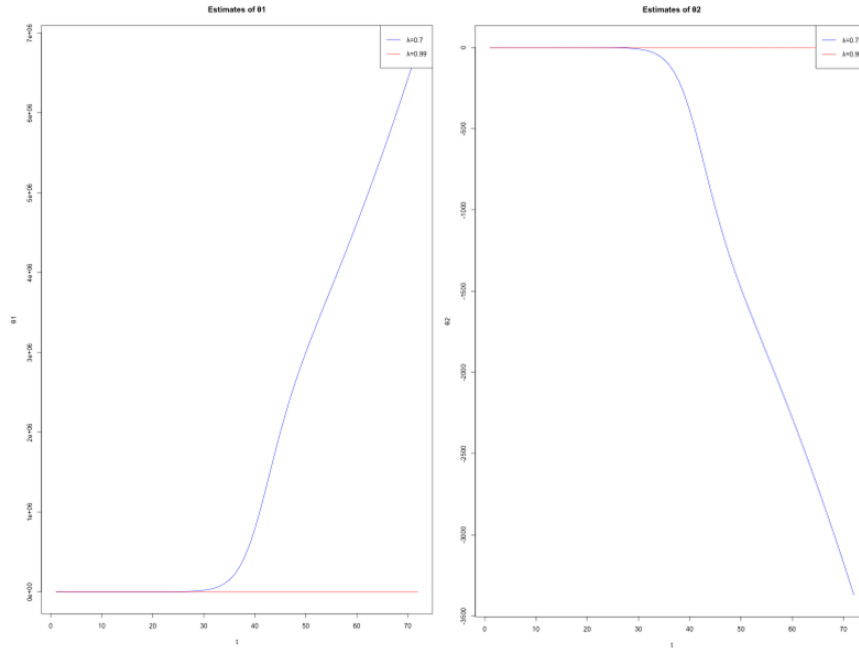
Figure 6: Parameter estimates for RLS with forgetting, with theta 1 on the left, and theta 2 on the right.
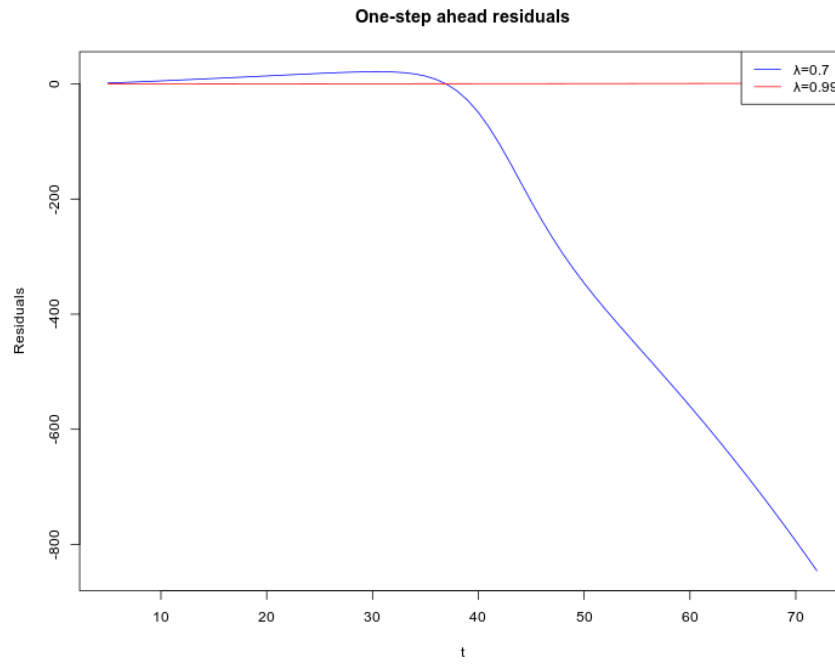


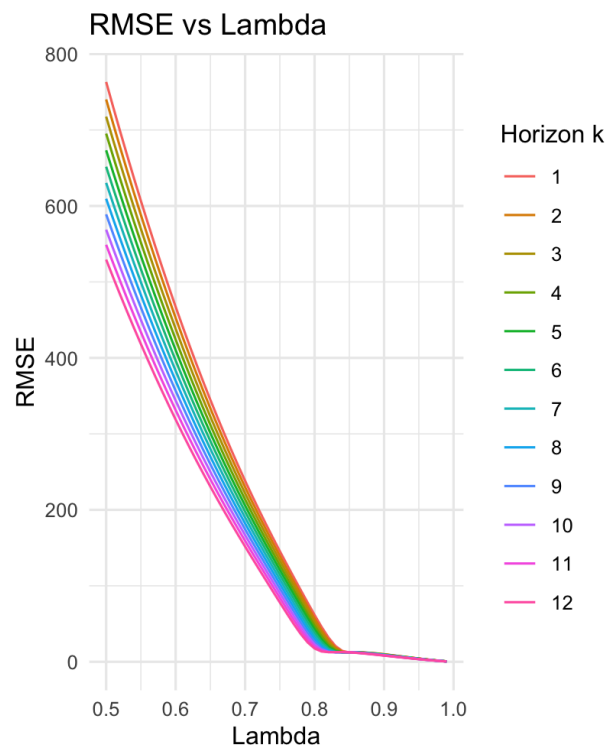Figure 7: Residuals for two different values of lambda.

Figure 8: RMSE vs Lambda



Figure 9: Forecasts for OLS, WLS, and RLS.

7

time dependent data requires sequential train-test splits such as rolling windows that make for a more realistic evaluation of the model performance.

Recursive estimation helps with challenges with test sets for time dependent data by allowing the model to update its predictions recursively as new data comes. So, instead of relying on a fixed set, the model refines its parameters and learns any new trends.

An alternative technique for adaptive estimation is the long short-term memory network that is designed to capture temporal dependencies and adapt over time.

## Appendix



Figure 10: s243660



Figure 11: s243896



Figure 12: s243933

**2.1**

**5.1**

Katarina, s243906

(2.1)   GLM → $y_T = \theta_1 + \theta_2 \cdot x_t + \varepsilon_t$, $\varepsilon_t \sim N(0,\sigma^2)$, $t = 1, \ldots N$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \implies Y = X\theta + \varepsilon$$

for Jan, Feb, Mar of 2018:  $x_t = 2018 + \dfrac{t-1}{12}$

→  $x_1 = 2018$   $x_2 = 2018 + \frac{1}{12} = 2018.083$   $x_3 = 2018.166$

from the csv data file: $Y_1 = 2930483$, $Y_2 = 2934044$, $Y_3 = 2941422$

$$\begin{bmatrix} 2930483 \\ 2934044 \\ 2941422 \end{bmatrix} = \begin{bmatrix} 1 & 2018 \\ 1 & 2018.083 \\ 1 & 2018.167 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

Figure 13: s243906

$\hat{\theta}_t = \ell_t^{-1} h_t$      $\ell_t = \ell_{t-1} + x_t x_t^T$      $h_t = h_{t-1} + x_t y_t$

$\ell_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$   $\theta_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$   $X_{[0,1]} = \begin{bmatrix} 1 & 2018 \\ 1 & 2018.083 \end{bmatrix}$   $Y_{[0,1]} = \begin{bmatrix} 2,930483 \\ 2,934044 \end{bmatrix}$

1) $\hat{\theta}_1 = \ell_1^{-1} h_1$

$= \begin{bmatrix} 7,196093\,e^{-07} \\ 0,00145272 \end{bmatrix}$

$h_1 = h_0 + x_1 y_1$

$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018 \end{bmatrix} \cdot 2,930483$

$= \begin{bmatrix} 2,930483 \\ 5913,314694 \end{bmatrix}$

$h_0 = \theta_0 \cdot \ell_0$

$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ell_0$

$\ell_1 = \ell_0 + x_1 x_1^T$

$= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018 \end{bmatrix} [1 \; 2018]$

$= \begin{bmatrix} 1.1 & 2018 \\ 2018 & 4072274.1 \end{bmatrix}$

2) $\hat{\theta}_2 = \ell_2^{-1} h_2$

$= \begin{bmatrix} 9,750073\,e^{-07} \\ 0,00145307 \end{bmatrix}$

$h_2 = h_1 + x_2 y_2$

$= \begin{bmatrix} 2,930483 \\ 5913,314694 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018.083 \end{bmatrix} \cdot 2,934044$

$= \begin{bmatrix} 5,864527 \\ 11874,86 \end{bmatrix}$

$\ell_2 = \ell_1 + x_2 x_2^T$

$= \ell_1 + \begin{bmatrix} 1 \\ 2018.083 \end{bmatrix} [1 \; 2018.083]$

$= \begin{bmatrix} 2.1 & 4036.083 \\ 4036.083 & 8144914.44 \end{bmatrix}$

Figure 14: s243660

9

⑤

$$\hat{Q}_t = R_t^{-1} h_t \qquad R_t = R_{t+1} + x_t x_t^T \qquad h_t = h_{t+1} + x_t y_t$$

$$R_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, \quad Q_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x_{1:2} = \begin{bmatrix} 1 & 2018,000 \\ 1 & 2018,083 \end{bmatrix} \quad y_{1:2} = \begin{bmatrix} 2,930483 \\ 2,934044 \end{bmatrix}$$

1° $\hat{Q}_1 = R_1^{-1} \cdot h_1 = \begin{bmatrix} 7,196093e^{-07} \\ 0,001452172 \end{bmatrix}$

$h_1 = h_0 + x_1 y_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018 \end{bmatrix} \cdot 2,930483 = \begin{bmatrix} 2,930483 \\ 5113,714694 \end{bmatrix}$

$h_0 = Q_0 \cdot R_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot R_0$

$R_1 = R_0 + x_1 x_1^T = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018 \end{bmatrix} [1 \ 2018] =$

$= \begin{bmatrix} 1,1 & 2018 \\ 2018 & 4072324,1 \end{bmatrix}$

2° $\hat{Q}_2 = R_2^{-1} \cdot h_2 = \begin{bmatrix} 9,775073e^{-07} \\ 0,001443024 \end{bmatrix}$

$h_2 = h_1 + x_2 y_2 = \begin{bmatrix} 2,930483 \\ 5913,714694 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018,083 \end{bmatrix} \cdot 2,934044 =$

$= \begin{bmatrix} 5,864527 \\ 11831,86 \end{bmatrix}$

$R_2 = R_1 + x_1 \cdot x_2^T = R_1 + \begin{bmatrix} 1 \\ 2018,083 \end{bmatrix} [1 \ 2018,083] =$

$= \begin{bmatrix} 2,1 & 4036,083 \\ 4036,083 & 8144984,44 \end{bmatrix}$

Figure 15: s243896

5.1 $R_t = R_{t-1} + x_t x_t^T \qquad \hat{Q}_t = R_t^{-1} h_t$ where $h_t = h_{t-1} + x_t y_t$

Given, $R_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$

then,

$R_1 = R_0 + x_1 x_1^T$

$= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018.000 \end{bmatrix} [1 \ 2018.000]$

$= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 & 2018.000 \\ 2018.000 & 2018^2 \end{bmatrix}$

$= \begin{bmatrix} 1.1 & 2018 \\ 2018 & 2018^2 \end{bmatrix}$

$R_2 = R_1 + x_2 x_2^T$

$= \begin{bmatrix} 1.1 & 2018 \\ 2018 & 2018^2 \end{bmatrix} + \begin{bmatrix} 1 \\ 2018.083 \end{bmatrix} [1 \ 2018.083]$

$= \begin{bmatrix} 1.1 & 2018 \\ 2018 & 2018^2 \end{bmatrix} + \begin{bmatrix} 1 & 2018.083 \\ 2018.083 & 2018.083^2 \end{bmatrix}$

$= \begin{bmatrix} 2.1 & 4036.083 \\ 4036.083 & 8144982.995 \end{bmatrix}$

Figure 16: s243933

Katarina, s243906

$R_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, \qquad \theta_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

RLS algorithm:

$\hat{\theta}_t = \hat{\theta}_{t-1} + R_t^{-1} X_t \left[ Y_t - X_t^T \hat{\theta}_{t-1} \right]$

$R_t = R_{t-1} + X_t X_t^T$

FIRST ITERATION: $\quad X_1 = \begin{bmatrix} 1 \\ x_1 = 2018 \end{bmatrix}$

$R_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 \\ x_1 \end{bmatrix} \begin{bmatrix} 1 & x_1 \end{bmatrix} = \begin{bmatrix} 0.1+1 & 0+x_1 \\ 0+x_1 & 0.1+x_1^2 \end{bmatrix} = \begin{bmatrix} 1.1 & 2018 \\ 2018 & 2018^2+0.1 \end{bmatrix}$

$\hat{\theta}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + R_1^{-1} X_1 \left[ Y_1 - X_1^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right] = R_1^{-1} X_1 Y_1$

SECOND ITERATION: $\quad X_2 = \begin{bmatrix} 1 \\ x_2 = 2018.083 \end{bmatrix}$

$R_2 = R_1 + X_2 X_2^T = \begin{bmatrix} 1.1 & 2018 \\ 2018 & 2018^2+0.1 \end{bmatrix} + \begin{bmatrix} 1 \\ x_2 \end{bmatrix} \begin{bmatrix} 1 & x_2 \end{bmatrix} = \begin{bmatrix} 2.1 & 4036.083 \\ 4036.083 & 2018^2+2018.083^2+0.1 \end{bmatrix}$

$\hat{\theta}_2 = R_1^{-1} X_1 Y_1 + R_2^{-1} X_2 \left[ Y_2 - X_2^T [R_1^{-1} X_1 Y_1] \right]$

Figure 17: s243906

11