Time Series Analysis
Assignment 1

Authors

Daniel Gonzalvez Alfert- s240404
Spiliopoulos Charalampos - s222948
Maria Kokali - s232486
Marios-Dimitrios Lianos - s233558

March 4, 2024

Contents

1 Plot Data

1.1 Plot the training data versus the time variable x

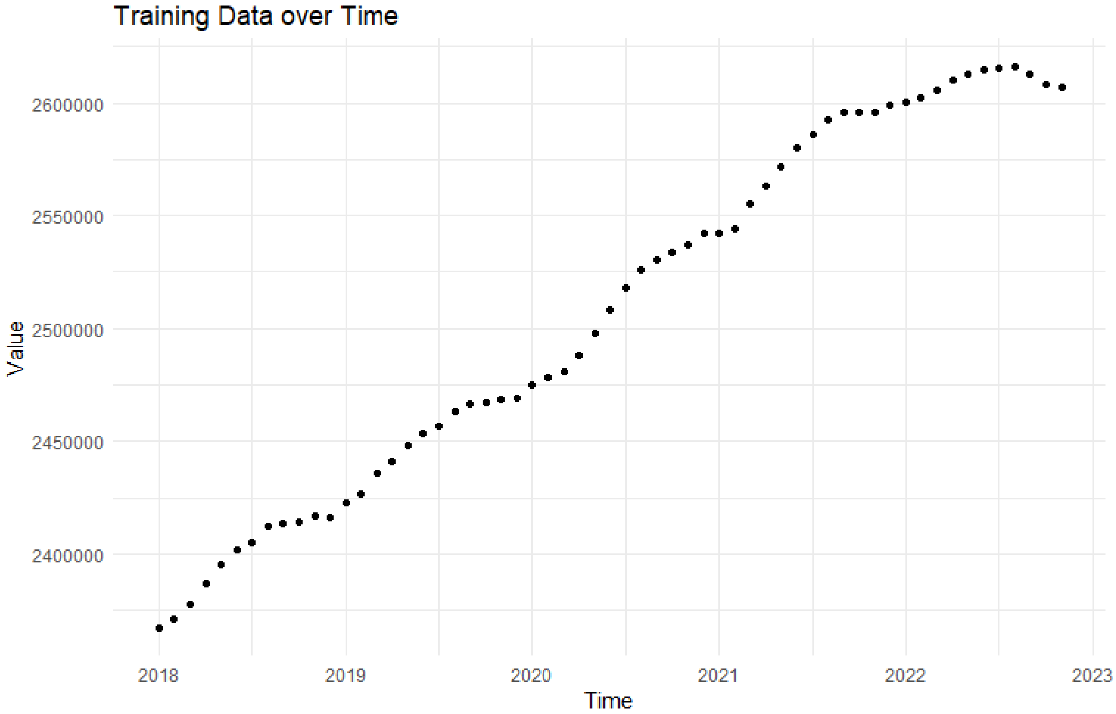

Figure 1: Training data versus time.

1.2 Describe the time series in your own words

A time series is a sequence of data points collected or recorded at successive points in time, usually at equally spaced intervals. In other words, it's a set of observations on the values that a variable takes at different times.

Specifically, in this particular case, we encounter a times series that represents the number of motor driven vehicles in Denmark each month starting from January 2018. The graph shows a steady increase in value over the years, starting from just above 2,400,000 in 2018, then reaching slightly below 2,600,000 in 2022 and after 2022, the increase slows down, and the values plateau as it approaches 2023. Overall, there is a clear upward trend in the data points.

2 Ordinary Least Squares

2.1 Estimation the parameters β_0 and β_1 using the training data (OLS model)

The Ordinary Least Squares estimation of the parameter β is given by $\hat{\beta}$ =$(X^TX)^{-1}X^Ty$, where X is the "design matrix" and y is the vector with observations.

The expression $\hat{\beta}$ =$(X^TX)^{-1}X^Ty$ comes from the fact that the OLS method attempts to minimize the expression $S(\beta) = (y - X\beta)(y - X\beta)^T$, which quantifies the error of the linear model. Therefore:

$$\nabla_\beta S(\beta) = -2X^T(y - X\beta) = 0 \iff X^TX\beta = X^Ty$$

If the matrix X^TX has a full rank, we can compute the expression $(X^TX)^{-1}X^Ty$ and obtain the OLS estimator $\hat{\beta}$. In this case we have added an intercept value β_0. This can be easily framed into the linear model by using a design matrix X in which the first column are all ones. Using such matrix X' in the linear model without intercept $y = X'\beta$ is equivalent to the model $y = \beta_0 + X'\beta_1$ that we desired in the first place. In this fashion we have calculated the estimators of β_0 and β_1.

2.2 $\hat{\beta}_0$, $\hat{\beta}_1$ and their estimated standard errors $\hat{\sigma}_{\hat{\beta}_0}$ and $\hat{\sigma}_{\hat{\beta}_1}$

$\hat{\beta}_0$	-109499932.75
$\hat{\beta}_1$	55437.78

Table 1: Estimations of β_0 and β_1

In order to obtain the standard errors, we just have to calculate the square root of the elements of the matrix $\hat{Var}[\beta] = \hat{\sigma^2_{OLS}}(X^TX)^{-1}$.

$\hat{\sigma}_{\hat{\beta}_0}$	1921556
$\hat{\sigma}_{\hat{\beta}_1}$	951.0687

Table 2: Estimations of $\hat{\sigma}_{\hat{\beta}_0}$and $\hat{\sigma}_{\hat{\beta}_1}$

2.3 Make a forecast for the next 12 months

First, we calculate the predicted values ($\hat{y}$), given by the formula $\hat{y} = X\hat{\beta}$, based on the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$. In Figure 2, we can see the predicted values represented by the red line.

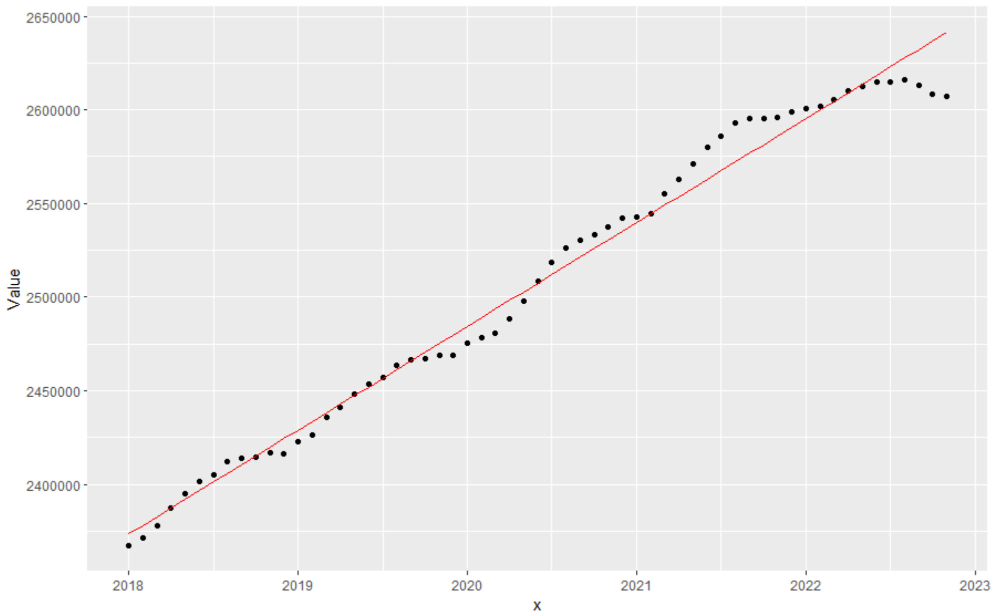

Figure 2: Predicted values $\hat{y}$

Then we continue by computing the residuals: e $= y$ - $\hat{y}$. The sum of squared residuals (RSS) is calculated as: $\text{RSS}_{\text{OLS}} = \mathbf{e}_{\text{OLS}}^T \mathbf{e}_{\text{OLS}}$.

The estimate of the variance (σ^2) is: $\hat{\sigma}^2_{\text{OLS}} = \frac{\text{RSS}_{\text{OLS}}}{n-p}$, where n is the number of observations and p is the number of parameters.

$R\hat{S}S$	6126129359
$\hat{\sigma^2}$	107475954

Table 3: Estimations of RSS and σ^2.

The residual standard error is the square root of the estimated variance: $\text{RSE}_{\text{OLS}} = \sqrt{\hat{\sigma}^2_{\text{OLS}}}$.

In Table 4, we have the variance-covariance matrix (V) of the OLS estimator is: $V_{\text{OLS}} = \hat{\sigma}^2_{\text{OLS}} \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$.The variances of the parameters are the values in the diagonal. Therefore, the desired estimate of standard errors is the square root of the respective elements of the diagonal of the variance-covariance matrix.

Variance and Covariance.	
3692375739353	-1827530914.1
-1827530914	904531.7

Table 4: Variance-covariance matrix.

RSE_{OLS}	1921555.5520	951.0687

Table 5: Standard Errors

In order to make a forecast for the next 12 months, we identify the last recorded time point in our dataset and generate future time points for the next 12 months, ensuring these are spaced one month apart. To account for the uncertainty in these predictions, we calculate the prediction intervals.

Eventually, in Table 6, we have the prediction intervals for the forecasted values.

Time	$\hat{y}$	**Lower**	**Upper**
2022-12-01	2646082	2625068	2667097
2023-01-01	2650702	2629653	2671751
2023-02-01	2655322	2634237	2676407
2023-03-01	2659942	2638820	2681064
2023-04-01	2664562	2643401	2685722
2023-05-01	2669182	2647982	2690382
2023-06-01	2673801	2652561	2695042
2023-07-01	2678421	2657139	2699703
2023-08-01	2683041	2661717	2704365
2023-09-01	2687661	2666293	2709029
2023-10-01	2692281	2670868	2713693
2023-11-01	2696900	2675443	2718358

Table 6: Forecasted Values with Prediction Intervals

Setting a confidence of 95%, having 59 observation, 2 parameters and using the following formula, we derive intervals shown in Figure 3.

$$\hat{Y}_t \pm t_{\frac{\alpha}{2}}(n-p) \cdot \hat{\sigma} \cdot \sqrt{1 + \mathbf{X}_t^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_t}$$

2.4 Plot of the forecast for the next 12 months

In Figure 3, we have the plot of the fitted model together with the training data, the forecasted values and the prediction intervals of the forecasted values.

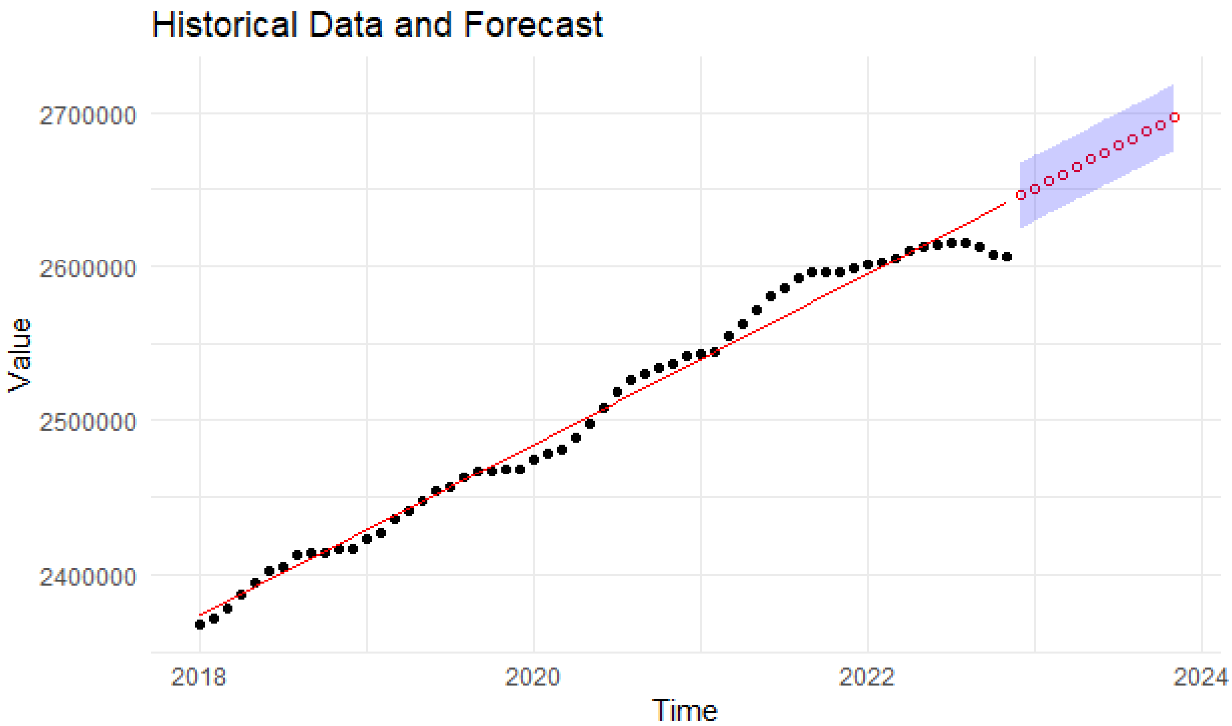

Figure 3: OLS 12-month forecast with prediction intervals

2.5 Comment on the Forecast Model

In general, the smooth extension of the trend in the forecast indicates a good model fit. The model closely matches past data without apparent overfitting and there is no presence of significant outliers.

More specifically, it appears that the model predicts the observations with relative accuracy (low bias) and precision (low variance) for the first 2 years, however, as time passes, we can see how the observed values diverge. This becomes clear when we try to forecast the next 12 months. The last observed values appear to be in a decreasing trend that is not very similar to the predicted values. Therefore, we could probably obtain a better forecast by changing to a local trend model.

2.6 Residuals of the Model

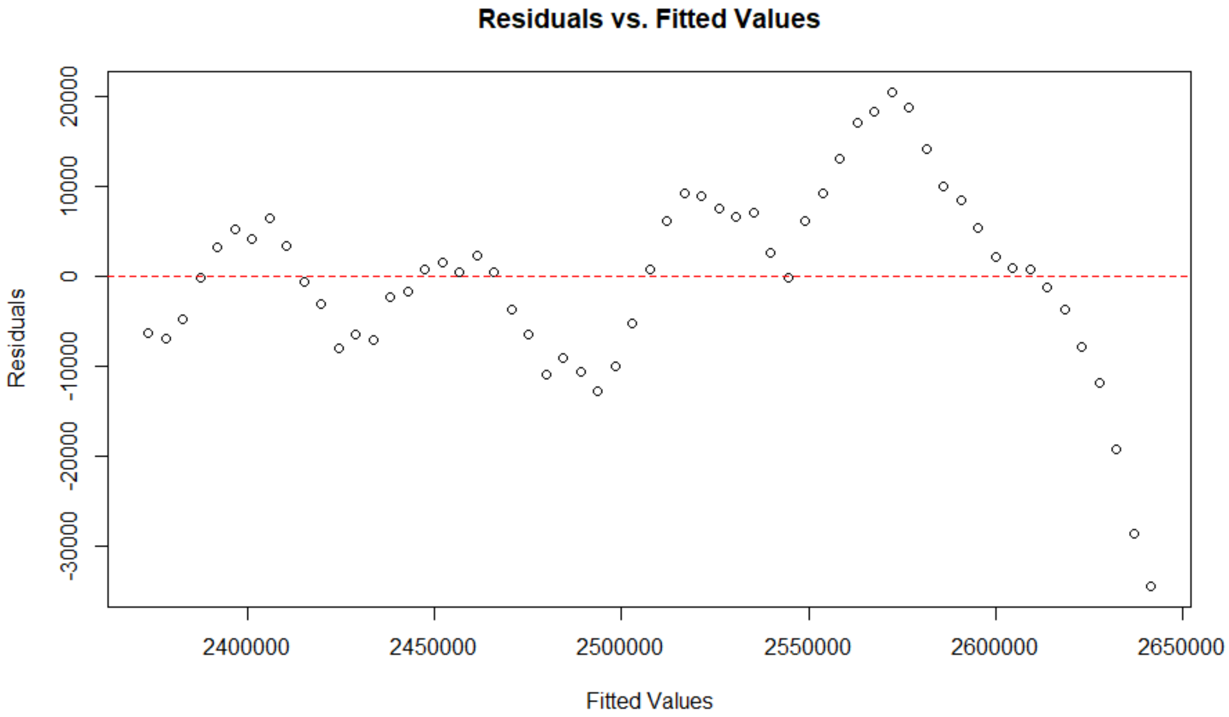

Figure 4: Residuals vs Fitted values

The linear regression model follows 4 general assumptions:

- *Linearity:* The relationship between X and y is linear. This assumption, according to the plot, appears to hold.
- *Independency:* The different observations of the random variable X are independent. In this case they obviously are.
- *Normality:* For any fixed value of X, y appears to distribute around a normal distribution. According to the QQ-plot and histogram, this assumption appears to hold with the exception of a few outliers. Furthermore, the histogram shows a roughly bell-shaped distribution but is not perfectly symmetrical, suggesting a slight departure from normality.
- *Homoscedasticity:* The variance of the residual is the same for any fixed value of X. This assumption does not appear to hold as the variance follows an increasing trend as time goes on. Heteroscedasticity can be an indicator that the model is not capturing all the predictive information and that errors have non-constant variance.

However, even with the issues mentioned, the residuals appear to be centered around zero, which is a good sign.

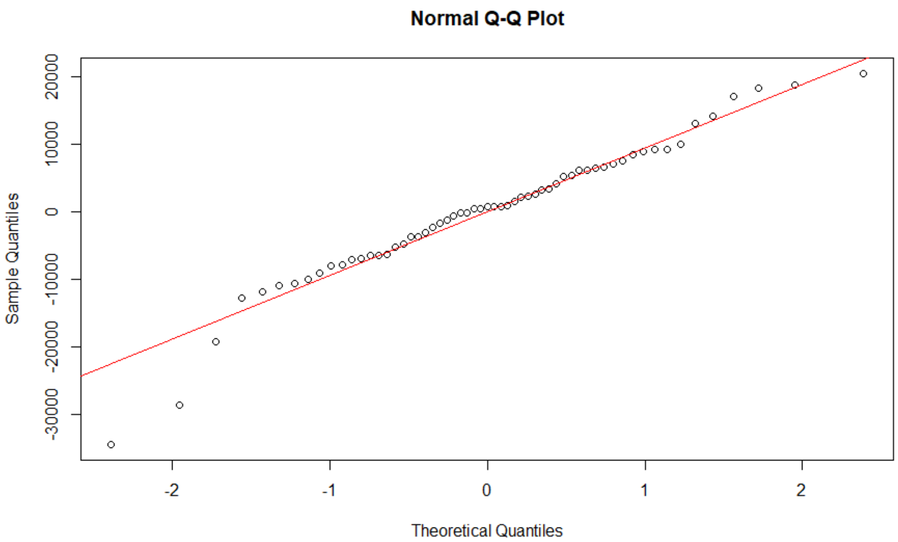

Figure 5: QQ plot

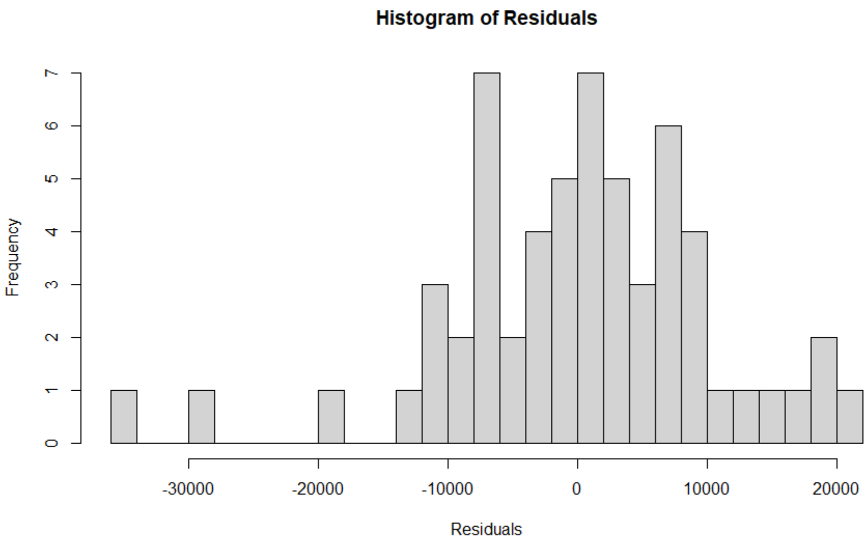

Figure 6: Histogram of Residuals

3 Weighted Least Squares - Local Linear Model

3.1 Describe the variance-covariance matrix for the local model and compare it to the variance-covariance matrix of the corresponding global model

For local models, the variance-covariance matrix is typically more complex due to the weighting applied to observations. In the case of a local linear model, weights are assigned to data points based on their proximity to the point of interest, with points closer to the target having more influence on the model fit at that specific location. This weighting scheme is designed to capture local trends and variations in the data.

Since we assumed homoscedasticity in OLS, the $n \times n$ (in this case $n = 59$) variance-covariance matrix Σ is just the n dimensional identity matrix. This is due to the fact that the random variable representing each residual is assumed to be independent from each other (and therefore covariance 0) and to have variance 1.

In contrast, the variance-covariance matrix for global models like Ordinary Least Squares (OLS) is simpler and assumes homoscedasticity, meaning the error variance is constant across observations. The OLS model does not use weights, and every observation contributes equally to the model, making it well-suited for scenarios where the underlying relationship is stable across the entire dataset. Σ is modified in the case of WLS since we want to assign a lesser weight to those observations that exhibit a greater variance. Therefore, the resulting variance-covariance matrix is:

$$\Sigma_{WLS} = \begin{bmatrix} \frac{1}{\lambda^n} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\lambda^{n-1}} & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \frac{1}{\lambda^1} & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} =$$

λ is a parameter in the range $0 < \lambda < 1$ that controls how fast the weight of the observations decrease.

In our case, we have set $\lambda = 0.9$, so the resulting matrix would be:

$$\Sigma_{WLS} = \begin{bmatrix} 500.83 & 0 & \cdots & 0 & 0 \\ 0 & 450.75 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1.11 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

The form of variance-covariance matrix is such because in order to implement WLS we will try to minimize the expression $(y - X\beta)^T \Sigma^{-1} (y - X\beta)$. In Σ^{-1} the elements of the diagonal are increasing in size, generating the desired effect of reduction in weight/importance of initial observations.

3.2 Plot the λ-weights vs. time. Which time-point has the highest weight?

We can clearly see in the graph that the maximum weight is assigned to the last observation: x_{59}.

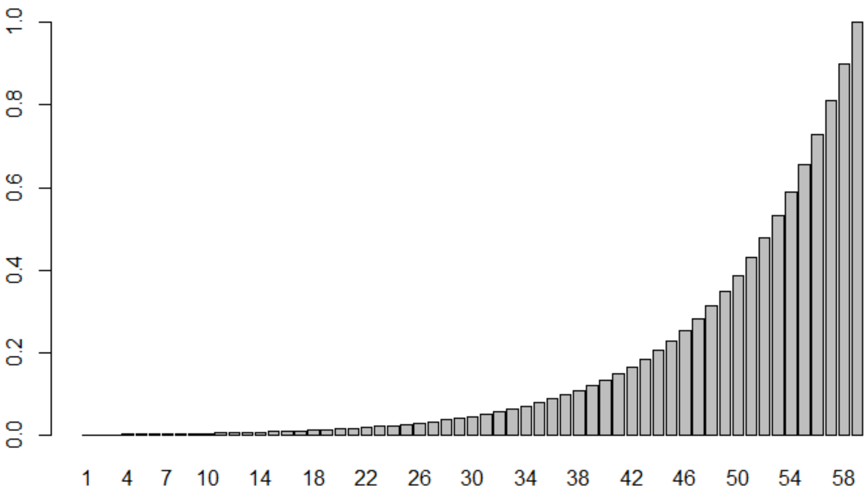

Figure 7: λ-weights vs. time

3.3 Calculate the sum of all the λ-weights. What would be the corresponding sum of weights in an OLS model?

The sum of all the weights is $\lambda_{sum} = 9.980033$. λ parametrizes the weight that every observation has. In the OLS model the weight equals to 1, so the sum of weights would be the number of the dataset observations.

3.4 Estimate and present $\hat{\beta}_0$ and $\hat{\beta}_1$ corresponding to the WLS model with $\lambda = 0.9$

$\hat{\beta}_0$	-840897815
$\hat{\beta}_1$	42868.56

Table 7: Estimations of β_0 and β_1 for λ=0.9

3.5 Make a forecast for the next 12 months with $\lambda = 0.9$. Plot the training data, the predicted values - one series for each λ), and the forecasted values for the next 12 months, also one series for each λ).

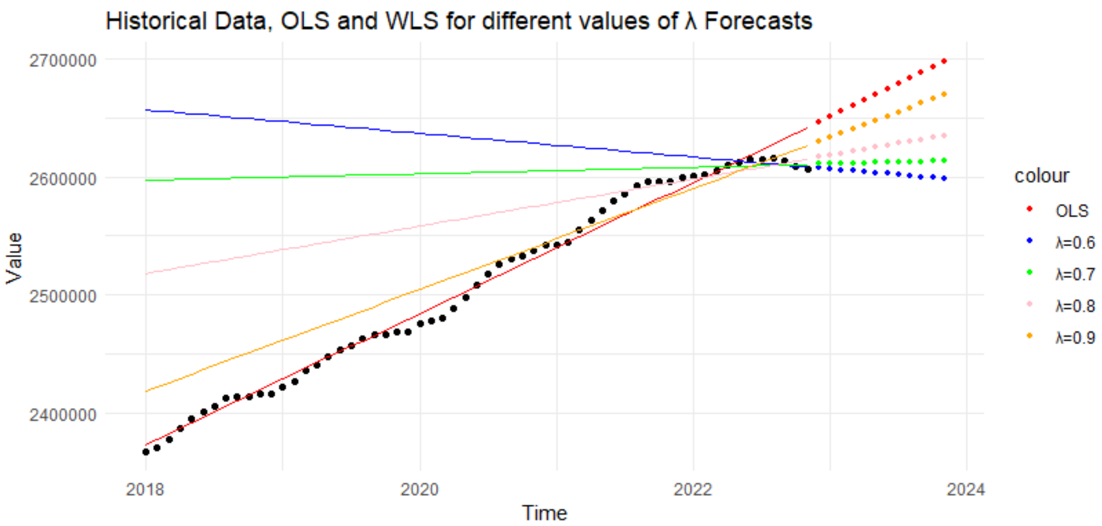

Figure 8: WLS 12-month forecast with prediction intervals, λ=0.9

3.6 Comment on the forecasts

We can see from how we define the expression that we minimize when implementing WLS, that in the end the initial observations will have a reduced weight, finilizing with the lowest one: λ^n. Therefore, the smaller the parameter λ, the faster and greater the reduction of weight is.

We can appreciate said phenomenon in the plots for each λ. The smaller the λ, the less the model adheres to initial observations and the more relative focus is put in latter observations.

One consequence of this strategy is the increase of the prediction intervals, due to the fact that by reducing the weight of certain observations, making them really close to 0 and therefore losing information initially available in the OLS model. In practicality, this implies an approximation to a model with less observations.

3.7 Which model would you use for decision making - or how would you decide which one to use knowing only the results generated until here?

In general it is preferable to maintain as much information as possible. It is always possible for new observations to demonstrate the inefficacy of any given model, so it seems unreasonable to sacrifice the certainty of new predictions and risk overfitting to the last observations just because there seems to be a trend.

On the other hand, dismissing said trend could lead to old and potentially obsolete observations introducing information to the model that is no longer useful, obscuring more recent trends.

All in all, the ideal choice of lambda would depend on the future horizon we would want to predict. For a time point closer to the present we would choose a small lambda around 0.6 while for time points further into the future a higher value should be a better choice. If we had to choose one lambda value though that would be $\lambda = 0.8$), giving new observations more importance but not generating a reduction in weight that is too drastic.

4 Iterative update and optimal λ

4.1 Provide L and f(0) for the model.

We define f for this model as:

$$f(0) = \begin{bmatrix} 1 \\ j \end{bmatrix}$$

Therefore $f(0)$ is:

$$f(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

On the other hand the matrix L such that $f(j+1) = Lf(j)$ is:

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

4.2 Provide F_1 and h_1.

$$F_1 = \lambda^0 f(0) f^T(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$h_1 = \lambda^0 f(0) y_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} 2367154 = \begin{bmatrix} 2367154 \\ 0 \end{bmatrix}$$

4.3 Using $\lambda = 0.9$ update F_N and h_N recursively and provide F_{10} and h_{10}.

We will use the following formulas:

$$\begin{aligned} F_{n+1} &= F_n + \lambda^n f(-n) f^T(-n) \\ h_{n+1} &= \lambda L^{-1} h_n + f(0) y_{n+1} \\ \hat{\beta}_{n+1} &= F_{n+1}^{-1} h_{n+1} \end{aligned}$$

To obtain the respective F_{10} and h_{10}:

$$F_{10} = \begin{bmatrix} 6.513216 & -23.7511 \\ -23751096 & 137.4602 \end{bmatrix} h_{10} = \begin{bmatrix} 15628557 \\ -56709556 \end{bmatrix}$$

4.4 Update the model recursively up to F_{59} and h_{59}, while also calculating predictions at each step. You should calculate predictions for 1 month ahead, 6 months ahead, and 12 months ahead.

In order to produce the predictions 1, 6 and 12 months ahead, we use the formula, for l steps ahead:

$$\hat{y}_{n+l|n} = f(l)^T \hat{\beta}_n$$

We use this formula to produce the following predictions for each $n = 11 \dots 59$ and $l = 1, 3, 12$.

n	1 step	6 step	12 step
12	2426890	–	–
13	2427428	–	–
14	2429919	–	–
15	2432803	–	–
16	2437984	–	–
17	2443241	2451920	–
18	2449177	2449614	–
19	2455069	2450941	–
20	2460285	2453081	–
21	2465919	2458863	–
22	2470773	2464670	–
23	2474389	2471409	2481956
24	2477449	2477977	2476236
25	2479605	2483477	2476167
26	2482791	2489541	2477413
27	2485879	2494449	2483918
28	2488730	2497613	2490386
29	2492700	2500049	2498088
30	2498014	2501303	2505466
31	2504716	2504059	2511308
32	2512377	2506724	2517888
33	2520207	2509107	2522861
34	2527366	2513040	2525481
35	2533719	2518779	2527170
36	2539550	2526331	2527340

Table 8: Forecasted Values with Prediction Intervals

n	1 step	6 step	12 step
37	2545234	2535086	2529580
38	2549699	2543977	2531737
39	2553530	2551898	2533559
40	2558913	2558703	2537448
41	2564818	2564791	2543696
42	2571338	2570668	2552269
43	2578385	2574947	2562337
44	2585288	2578416	2572500
45	2592253	2583918	2581336
46	2598402	2590085	2588684
47	2603262	2597027	2595081
48	2607152	2604618	2601190
49	2610806	2611990	2605244
50	2613906	2619410	2608279
51	2616546	2625759	2613924
52	2619211	2630450	2620407
53	2622146	2633915	2627855
54	2624914	2637106	2636099
55	2627458	2639621	2644032
56	2629453	2641585	2651998
57	2631036	2643623	2658587
58	2631587	2646038	2663076
59	2630848	2648272	2666031

Table 9: Predictions from local model for $n = 12 \dots 59$ for 1, 6 and 12 months.

4.5 Plot the resulting 1-month, 6-month, and 12-month prediction together with the training data.

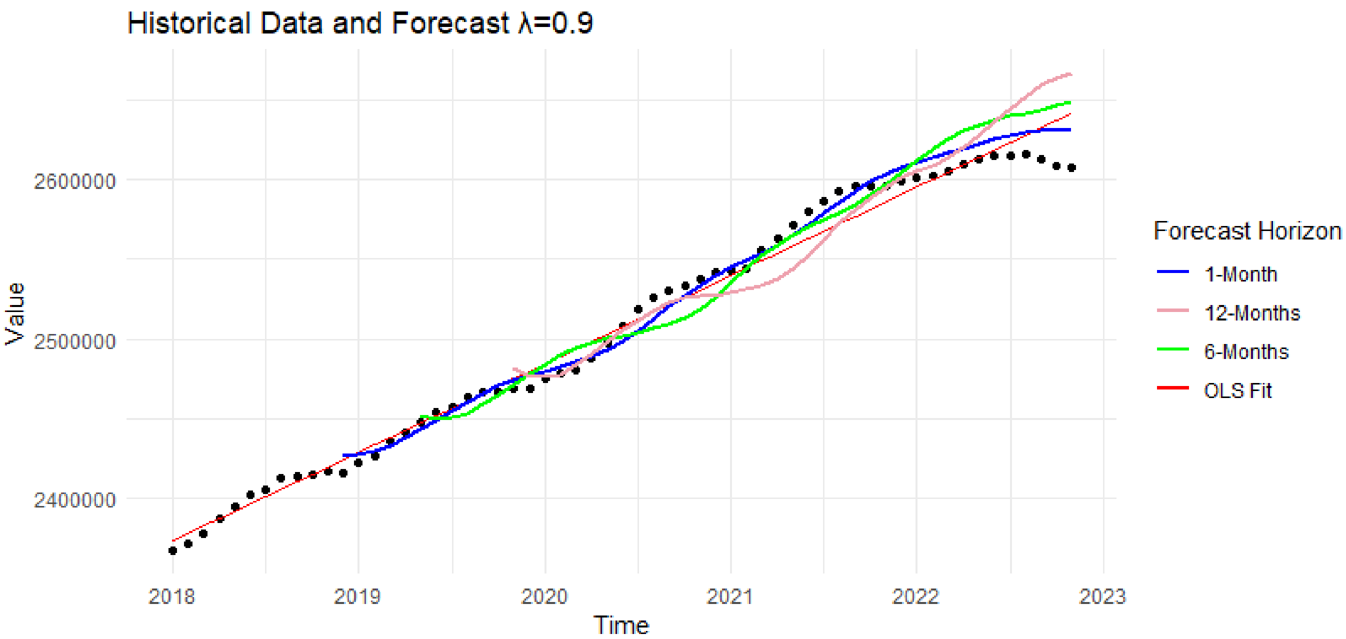

Figure 9: 1-Month, 6-Months, 12-Months predictions for $\lambda = 0.9$

The black dots represent the actual historical data points. They show an increasing trend over time, which the forecasts and OLS fit are attempting to model.
The 1-month forecast closely tracks the actual historical data and the OLS fit, which suggests that the model for Îť=0.9 is quite responsive to recent changes in the data. This close fit indicates that the model may be useful for short-term forecasting, as it adjusts quickly to the most recent data trends.

The 6-month forecast starts to deviate from the OLS fit but still follows the general trend of the data. The divergence suggests that as the forecasting horizon increases, the model begins to weigh recent observations and the inherent variability differently. There may be a slight lag in response to the most recent trends, likely due to the cumulative effect of weighting past observations.

The 12-month forecast deviates further from the OLS fit compared to the 1-month and 6-month forecasts. This model seems to flatten out towards the end of the forecast period, which could indicate that the influence of the more recent data is becoming diluted as the forecast extends further into the future. It may not capture potential cyclic patterns or longer-term trends as effectively due to the high value of Îť emphasizing recent data.

The three forecast models diverge from each other as the forecasting horizon increases, suggesting that the choice of Îť impacts longer-term predictions more significantly.

4.6 Repeat the iterative predictions for $\lambda = 0.55, 0.56, 0.57, ..., 0.95$, and calculate the root-mean-square of the prediction errors for each forecast horizon (1 month, 6 months, and 12 months) and for each value of λ. Plot the root-mean-square of the prediction errors versus λ for both 1-month, 6-month, and 12-month predictions.

The graphs depict the Root Mean Square Error (RMSE) for different forecast horizons (1-month, 6-months, and 12-months) as a function of the parameter λ in a time series forecasting model. Here are some observations:

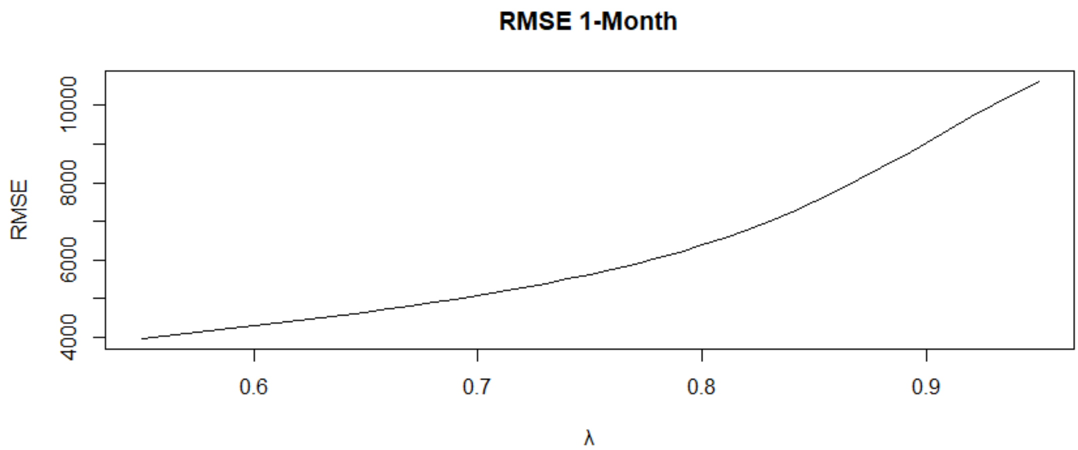

Figure 10: $RMSE$ for 1-Month vs λ

The RMSE appears to increase steadily as λ increases. This suggests that for short-term predictions, a lower λ, which puts more weight on the most recent observations, may lead to better prediction accuracy.

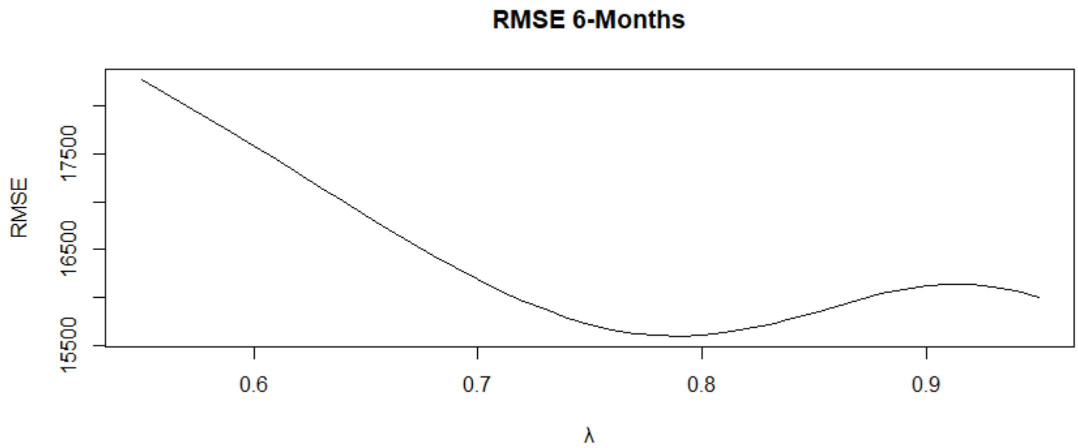

Figure 11: $RMSE$ for 6-Months vs λ

The RMSE decreases as λ increases from 0.55 to approximately 0.7 and then starts to increase slightly before flattening out. This indicates that there is an optimal value of λ around 0.7 for 6-month predictions, balancing the contribution of recent and older data.

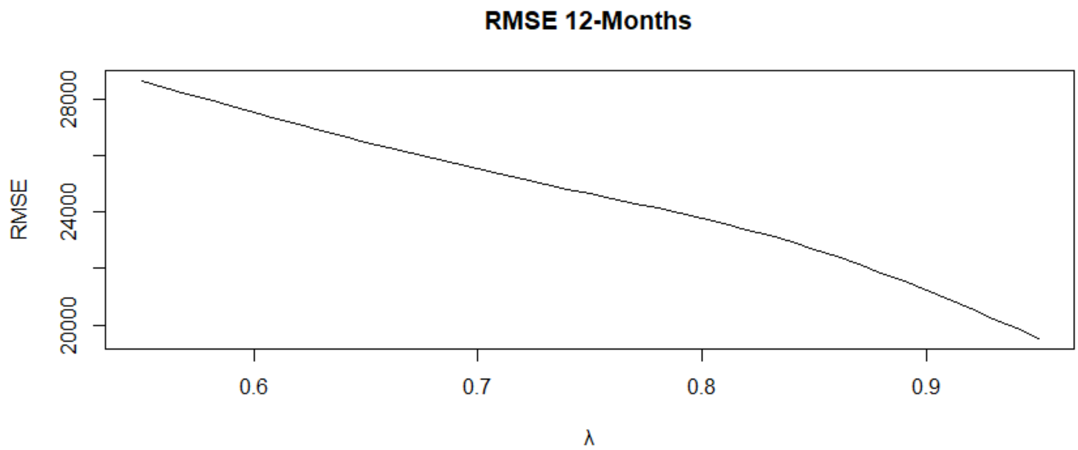

Figure 12: $RMSE$ for 12-Months vs λ

The RMSE decreases more consistently as λ increases. This trend suggests that for long-term predictions, a higher λ (giving more uniform weight to older observations) may result in better forecasting performance.

4.7 Which value of λ is optimal for predicting 1-month ahead?

$\lambda_{optimal} = 0.55$ with $RMSE = 3980.50$

4.8 Which value of λ is optimal for predicting 6-months ahead?

$\lambda_{optimal} = 0.79$ with $RMSE = 15600.56$

4.9 Which value of λ is optimal for predicting 12 months ahead?

$\lambda_{optimal} = 0.95$ with $RMSE = 19528.34$

4.10 It would be problematic to make λ as small as 0.5. Why is that?

When λ is set as low as 0.5 in a weighted least squares model, the sum of the weights for the observations decreases significantly, especially for older data points. This reduction in the effective number of observations, due to the diminishing weights, can lead to a scenario where the sum of the weights is comparable to or even less than the number of parameters to be estimated in the model. In such cases, the model risks becoming underdetermined, meaning there isn't enough information to provide reliable estimates for the parameters. Essentially, the model could suffer from overfitting to recent data while losing the ability to generalize well across the entire data set. This balance between the sum of weights and the number of parameters is crucial for ensuring the model's stability and its capacity to capture the underlying data structure accurately.

4.11 Compare the 1-month-ahead prediction with $\lambda = 0.55$ to the naive persistence model. Which one is better?

Here, we must compare the 1-month ahead prediction's RMSE with that of the naive persistence model.
The $RMSE$ for the 1-month ahead prediction model with $\lambda = 0.55$ is:

$$RMSE_{1-month} = 3980.50$$

The $RMSE$ for the naive persistence model is:

$$RMSE_{naive} = 5386.69$$

It can be observed that the naive model, which suggests that the next step prediction is equal to the previous one, performs worse than the Local Trend model for the 1-month prediction based on their RMSE, which is expected.

4.12 Now choose the best forecasts at time t = 59 of Y_t for time t = 60, t = 65, and t = 71, and plot them together with the training data and the test data.

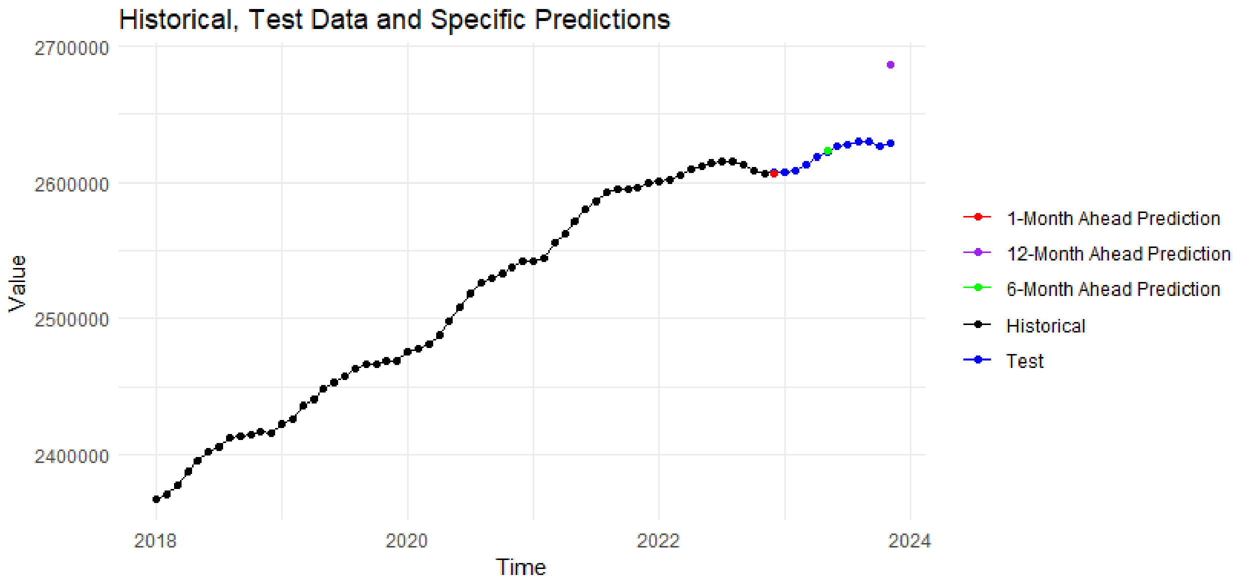

Figure 13

Based on our previous calculations, the models for the predictions were: for the 1-month ahead the model with $\lambda = 0.55$, for the 6-months ahead $\lambda = 0.79$ and for the 12-months ahead the model with $\lambda = 0.95$.

4.13 Comment on the model's prediction performance for both 1-month, 6-month, and 12-month ahead.

The 1-month ahead prediction appears to follow the trajectory of the historical data closely. Given its proximity to the last known data point, it suggests that the model expects little change in the short term, which aligns with the generally incremental nature of the time series.

The 6-month ahead forecast (green dot) aligns closely with the subsequent test data points, indicating that the model has successfully captured the underlying trend of the series over this medium-term horizon. It reflects a good balance between recent trends and historical behavior, making it a potentially reliable forecast for this specific time frame.

The 12-month ahead prediction is higher than the other predictions and the test data points. It suggests a significant increase, which could indicate an anticipation of a trend expected to raise the series value or the incapability of our model to successfully make forecasts that far in the future.

List of Figures

List of Tables