Karlsruher Institut für Technologie

Problemset 4 - Marketing Analytics

Institute of Information Systems and Marketing (IISM)

Julius Korch, Marco Schneider, Stefan Stumpf, Zhaotai Liu

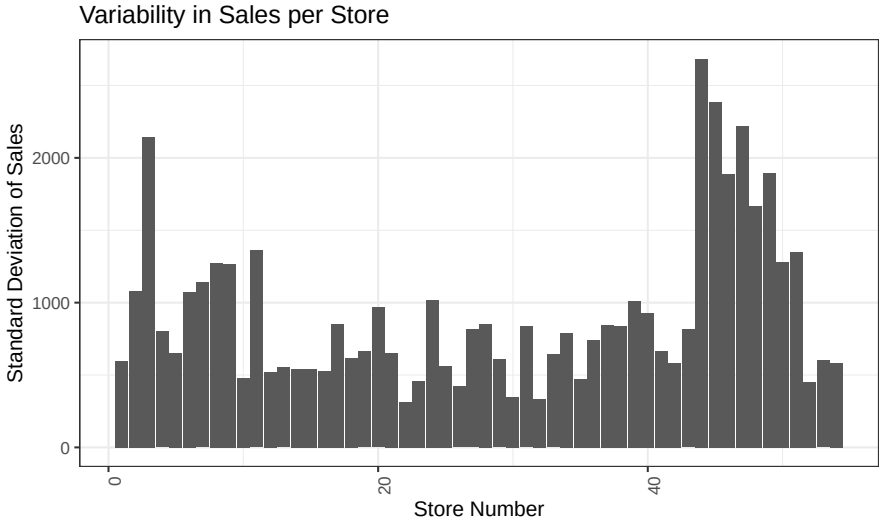
Last compiled on January 12, 2024

Contents

Task 1

Choice of Store

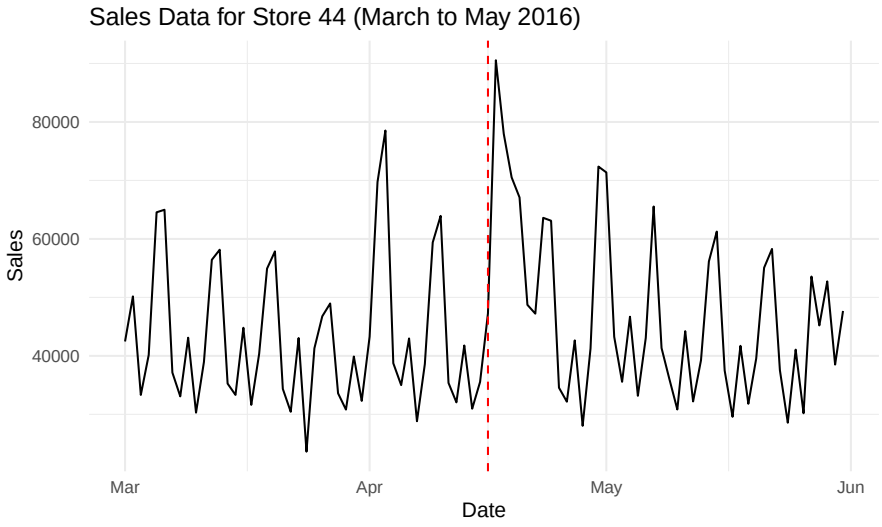
At the beginning of the task we explored the dataset to find out which store to choose. First we looked at the overall sales for each store and found that the store with ID 44 had the highest sales. Then we looked at the variability of sales per store. The graph also shows that ID 44 has the highest variability, suggesting that the store is a good choice for making predictions. Therefore, the store with ID 44 will be the store we look at for the rest of Task 1 and Task 2.



Preprocessing

After filtering the data for store 44, we also did one important preprocessing step. Because one data can have numerous different sales, we aggregated the sales of a day to have a total amount of sales per day. Furthermore, our data exploration showed that the sales data for each year on the 01.01.X and 25.12.X are missing. Because this missing data could result in noise when looking at seasonality, we added those days back into the data and imputed them with the yearly average of sales.

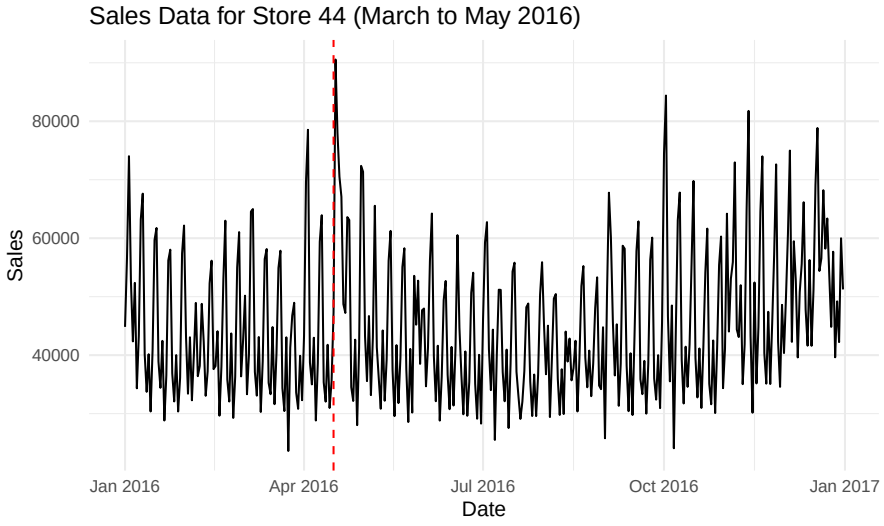
Analyzing the earthquake period



We also carried out an analysis of the earthquake period to determine the impact period of the earthquake

on the store. We looked at the data from two perspectives. First, we plotted the data from March to June to get an enlarged view of the sales in this short period. In the graph above, the date of the earthquake is marked by a red line. From the graph we can see that sales actually increased in the days following the earthquake. This suggests that 1. the shop was probably not directly affected by the earthquake because it was still open the day after, and 2. it's likely that other shops were affected in such a way that they couldn't open the day after, leading to the increase in sales for store 44.

To get a better feel for the data, we have also plotted the data for the whole year in the graph below. This allows us to look at the patterns to see if the month of the earthquake seems unusual. But looking at the graph, apart from the sharp increase in sales for the week of the earthquake, the pattern of sales returns to a normal state for the rest of the month. Moreover, there is no apparent lasting effect of the earthquake on the data. We therefore conclude that the impact pattern for the selected store only lasts for about one week after the earthquake.

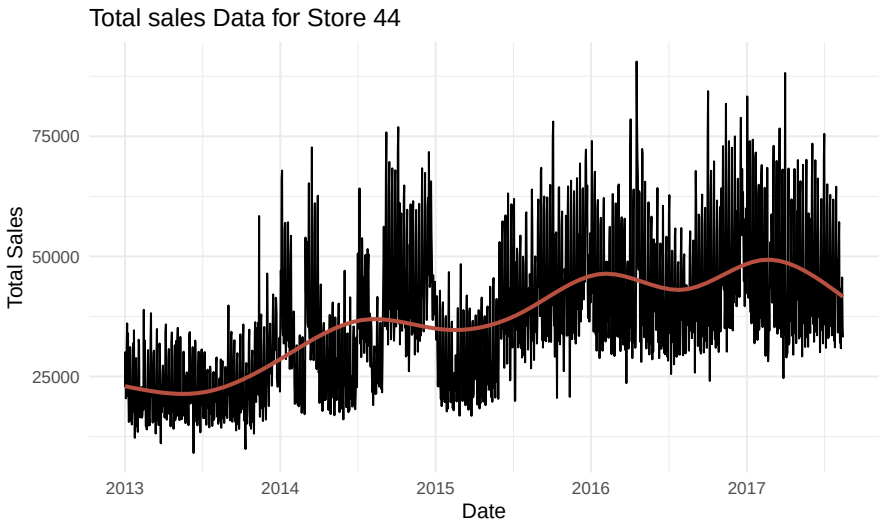


Analysis of the Stationarity of the Data

The graph below shows the aggregated sales for the entire dataset. Looking at the graph, we can see a clear upward trend in the data. The largest amount of sales occurred on 2016-04-17, one day after the earthquake. The second largest amount of sales occurred on 2017-04-01. We found out that there was a holiday in the neighbouring province (Cotopaxi) at that time. The province of our shop (number 44) did not have a holiday on that day. Therefore, we can speculate that people from the province of Cotopaxi decided to shop in store 44.

Going back to the trend that can be observed, it can bee seen that overall this is a upward trend which would indicate a deterministic trend. Such trends may occur due to various factors like market growth, expanding customer base, inflation etc. Furthermore, shocks like the earthquake seem to have no long lasting effect on the store's sales. As it only caused a temporary spike in the sales, which however lasted only about a week which can be seen in our analyses on the earthquake period Analyzing the earthquake period (seen above).

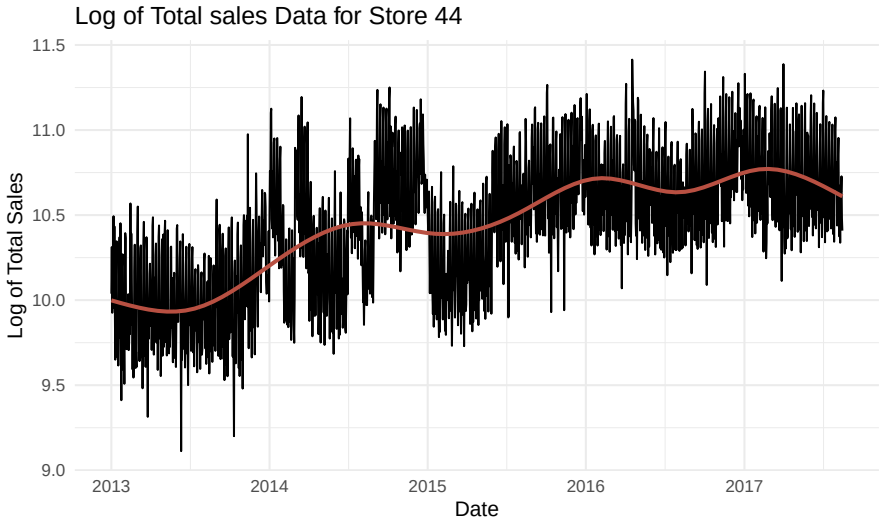
It can be seen that the amount of sales for the shop increases over time. This trend supports our conclusion that the data is not initially stationary. In the following sections we will process the data to make it stationary.



Logging the data

The sales data show non-constant variance over time, with higher sales on weekends and holidays. Sales often fluctuate over different time periods, such as during holiday seasons or special events such as Black Friday or Christmas, and decline during the off-season. In addition, external shocks such as natural disasters, political changes or technological disruptions can have a sudden impact on sales. For example, the earthquake caused a sudden spike in sales. Sales volumes exhibit non-constant variance, indicating the presence of heteroskedasticity in the data. To deal with this problem, we apply a logarithmic transformation to the data.

After logging, the data can be seen in the plot below. It can be seen that the trend is still clearly visible, suggesting further processing steps such as differencing, which we will apply in the next section.



Tests for Stationarity

Before we apply differencing, we also conduct tests to see if they support our assumption of the data not being stationary. In the lecture we learned about the augmented Dickey-Fuller test, which checks if the data is stationary:

```
##
##  Augmented Dickey-Fuller Test
##
## data:  log_sales_ts
```

```
## Dickey-Fuller = -5.723, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```

The application of the augmented Dickey-Fuller test reveals that the p-value (0.01) is lower than the significance level (0.05). Therefore, we reject H0 (the data contains a unit root and is not stationary) and choose H1 (the data is stationary). This indicates that the data is stationary, despite the plot suggesting otherwise. We assume that this is due to the augmented Dickey-Fuller test checking for a stochastic trend (also known as a unit root). If the test indicates that the data are stationary, this means that there is no stochastic trend. However, this does not necessarily mean that our data is trend free. Since we are assuming that our data is deterministic and we can observe a clear trend in the data, it is technically not stationary because the mean of the series changes over time. Stationarity requires that the statistical properties of the series, such as mean, variance and autocorrelation, remain constant over time. However, this is not the case if there is a deterministic trend. Therefore, we use the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to further examine the stationarity of our data:

```
##
##  KPSS Test for Level Stationarity
##
## data:  log_sales_ts
## KPSS Level = 12.048, Truncation lag parameter = 8, p-value = 0.01
```

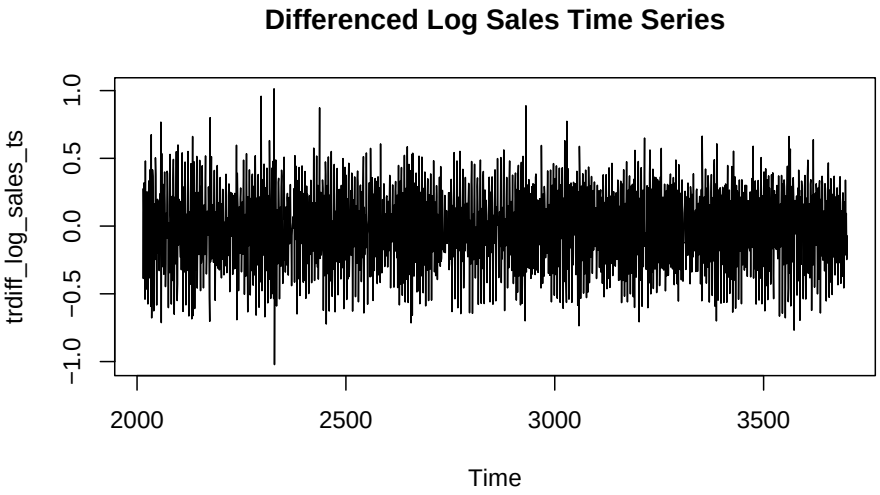
The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a unit root test that determines the stationarity of a given series around a deterministic trend. The null hypothesis of the KPSS test is that the series is stationary around a deterministic trend, while the alternative hypothesis is that it is not. After running the KPSS test on our data, we found that the p-value (0.01) is less than the significance level (0.05). Therefore, we reject the null hypothesis and conclude that the data is not stationary. This contradicts the result of the Augmented Dickey Fuller test. However, it is important to note that the ADF test tests for a stochastic trend, whereas the KPSS test tests for a deterministic trend. Therefore, we can conclude that our series contains a deterministic trend but no stochastic trend and is therefore not stationary. This supports our initial assumption when examining the data points of the time series.

Differencing to Remove the Trend

Because the data is non-stationary, we need to make it stationary and remove the deterministic trend. We do this by differencing the sales data with a lag of 1. After differencing, we run the KPSS test again to see if the data is now stationary around a deterministic trend. We do this because sometimes it is not enough to run the differencing once, but perhaps more often to properly remove a trend.

```
##
##  KPSS Test for Level Stationarity
##
## data:  trdiff_log_sales_ts
## KPSS Level = 0.0047545, Truncation lag parameter = 8, p-value = 0.1
```

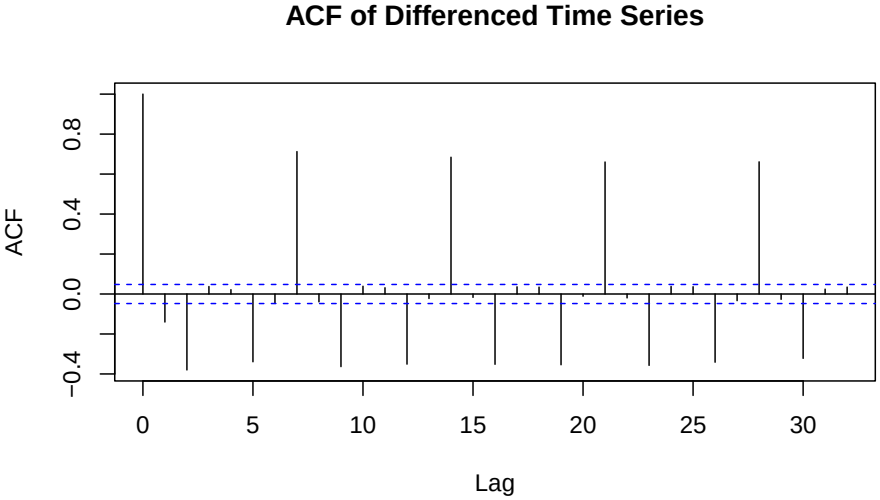
The p-value of the KPSS test performed on the differenced data is 0.1. This is greater than the 0.05 significance level. We can therefore reject H0 and conclude that the data are now stationary around a deterministic trend. In other words, the trend has been successfully eliminated. This can also be seen in the plot of the differenced data (see below).



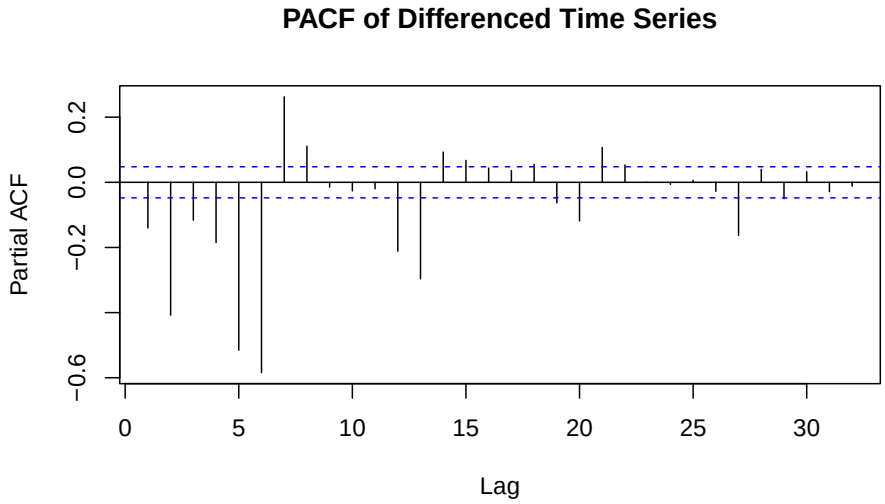
The data appears to move consistently around the zero point over time. Our next step is to analyse the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the differenced time series

ACF and PACF Plots of Differenced Time Series

After removing the trend from the data, the ACF and PACF plots indicate that there is still autocorrelation in the data, represented by the significant spikes in the plots.



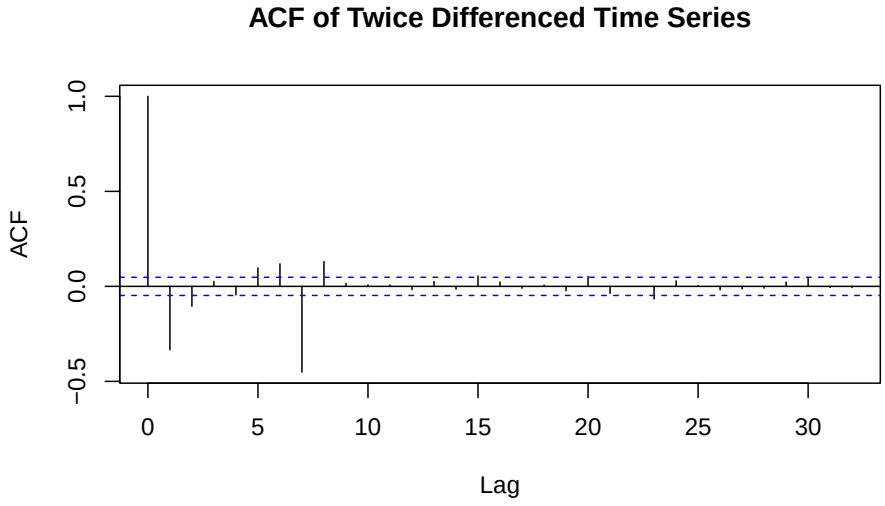
Looking at the ACF plot of the differenced time series, it is interesting to note that there is a significant autocorrelation every seven lags. In statistical terms, this means that sales values are significantly positively correlated with their values 7 lags later (or earlier). This is a strong indication that the sales value observed today is influenced in some predictable way by the value observed exactly 7 lags ago. As we analyse the sales data on a daily basis, this means that there is a weekly seasonality, as 7 lags is equal to 7 days.

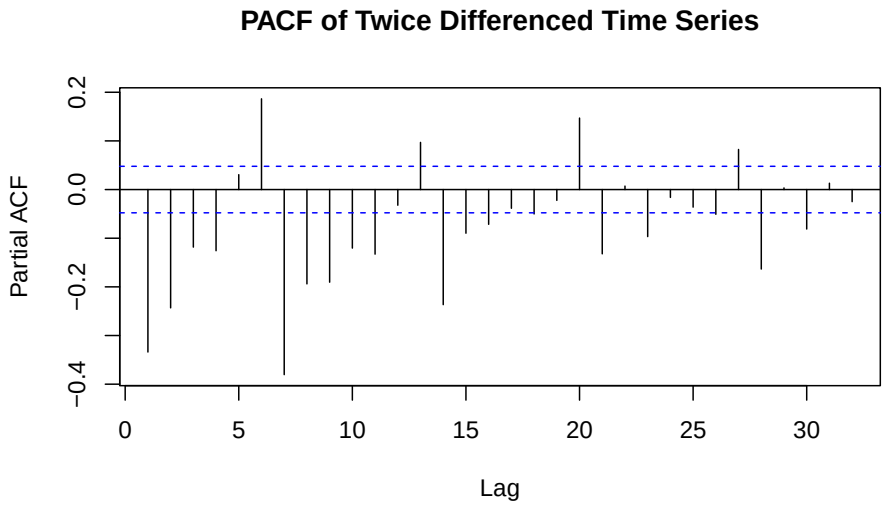


This is also supported by the PACF plot, which shows significant partial autocorrelation peaks at lags 7, 14 and 21. Unlike the ACF, which shows the overall correlation, the PACF isolates the direct relationship between an observation and its lagged version, controlling for relationships at shorter lags. Thus, a spike at lag 7 in the PACF indicates a direct correlation between an observation and its value 7 periods earlier that is not explained by correlations at lags 1 to 6. It also seems logical to us that there is a weekly seasonality in supermarket sales data due to consumer habits and the typical working week cycle. Shoppers often do their main grocery shopping at weekends, which coincides with leisure time and weekly work schedules, leading to a consistent weekly pattern in sales.

Differencing to Remove Seasonality

In order to remove the weekly seasonality from our time series, we decided to run the differencing again. This time we use lag = 7 because a lag is seven days and a week is exactly 7 days (also in our data). After differencing, the ACF and PACF plots are analysed again to determine the best parameters for the SARIMA model, which will be used in the next step to forecast the next 15 days of sales for store 44.





After differencing for weekly seasonality, we can see that the ACF and PACF plots have changed. In the ACF plot, there are only a few significant autocorrelation spikes left, which means that there is less observed correlation between lags. There are still some significant spikes in the PACF plot. Now the spikes start at lag 6 with a distance of 7 lags. This means that there is still a correlation between an observation and its logged version when controlling for the previous lags. However, if we look at the y-axis of the PACF plot, we can see that many of the significant spikes are close to the significance level compared to the spikes of the ACF plot before removing the weekly seasonality (the y-scale is larger in the ACF plots than in the PACF plots). We therefore conclude that there is no relevant seasonality left in the time series that we should take into account. We also checked for possible annual and monthly seasonality by setting lag.max to 366 (annual) and 31 (monthly) in the ACF and PACF plots. However, there were no relevant significant spikes in the plots. We therefore conclude that there is no annual or monthly seasonality in the data. This can be seen in the code.

Forecasting of Sales Data

Choosing Parameters Based on ACF/PACF Plots

Having removed the trend and seasonality from the time series, we now want to forecast the next 15 days of sales for Store 44. Due to the seasonality that is present in the original data, we use a SARIMA model to properly account for it. To find the best parameters for the SARIMA model, we first look at the ACF and PACF plots of the data (as shown above). To determine the seasonal coefficients of the SARIMA model, we look at the ACF and PACF plots of the data before removing the seasonality. To determine the non-seasonal coefficients, we look at the ACF and PACF plots of the data after removing the seasonality. The reason for this is that the twice differenced data no longer contain seasonality, which is important for the non-seasonal coefficients as they do not take these patterns into account.

Seasonal coefficients (P, D, Q) + seasonal period s:
We can see significant spikes in the ACF and PACF plots of the data before removing the seasonality at lag 7 and multiples of 7. Thus, we choose the seasonal period s equal to 7 (s=7) and the seasonal AR as well as seasonal MA coefficient equal to 1 (P=1, Q=1). For the reason that the significant spikes in the PACF are close to the significance level, we will also try P=0. P=3 will also be checked, since there are 3 significant spikes in the PACF plot (lag 7, 14 and 21). The seasonal differencing is already done in the previous step. Thus, we set D=1.
In conclusion, the seasonal coefficients are:
P=1, D=1, Q=1, s=7 or P=0, D=1, Q=1, s=7 or P=3, D=1, Q=1, s=7

Non-seasonal coefficients (p, d, q):
We can see significant spikes in the PACF after seasonal differencing until and including lag 4. Thus, we choose the non-seasonal AR coefficient equal to 4 (p=4). In the ACF plot after seasonal differencing we can see significant spikes until and including lag 2. Thus, we choose the non-seasonal MA coefficient equal to 2

(q=2). Since we performed non-seasonal differencing, we set d=1.
In conclusion, the non-seasonal coefficients are p=4, d=1, q=2.

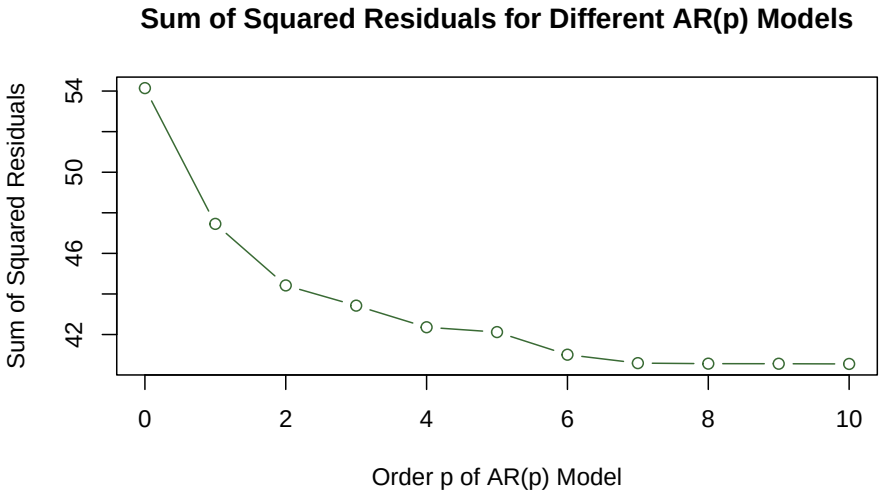
So, our initial models to try are:
SARIMA(4,1,2)(1,1,1)7, SARIMA(4,1,2)(0,1,1)7 and SARIMA(4,1,2)(3,1,1)7.

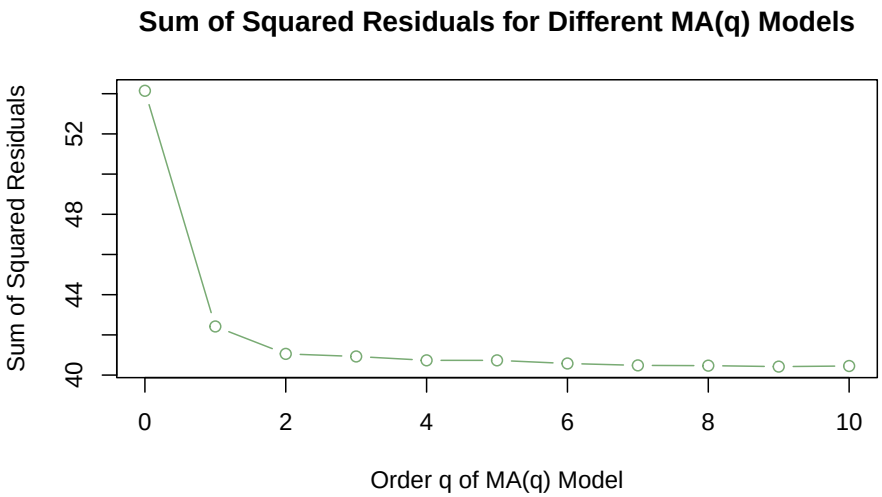
These coefficients have been deduced only by looking at the ACF and PACF plots. We want to further check these coefficients by including other criteria, such as the elbow criterion, by running repeated analyses with different coefficients and comparing their empirical performance (sum of squared residuals). We also check the AIC and RMSE/MAE of the different possible models and select the best model according to these criteria.

Recheck Parameters Based on Elbow Criteria

For this, we wrote functions which try out different parameters for the SARIMA model and plot the sum of squared residuals for the non-seasonal AR- and MA-coefficients with every parameter combination. We give the functions the following information which is always included:
d=1, D=1, Q=1, s=7
The reason for those preset parameters are that we performed non-seasonal and seasonal differencing, we identified a weekly seasonality which consistently occurs in the ACF plot after seasonal differencing.
We did the procedure three times to try it with P=1, P=0 and P=3. Due to space reasons, we only included the plots for P=1 in the document. The plots for P=0 and P=3 can be found in the code.

First the plots for P=1: Following the elbow method on the residuals, we would choose p=4 or p=2 as the best parameter for the non-seasonal AR part of the model.
Additionally by looking at the graphs below we would choose q=1 or q=2 as the best parameter for the non-seasonal MA part of the model.
Thus, we further analyse the following models: SARIMA(4,1,1)(1,1,1)7, SARIMA(4,1,2)(1,1,1)7, SARIMA(2,1,1)(1,1,1)7 and SARIMA(2,1,2)(1,1,1)7





For P=0: Following the elbow method would indicate here to choose p=4 or p=2 as the best parameter for the non-seasonal AR part of the model. It would also indicate to choose q=1 and q=2 as the best parameters for the non-seasonal MA part of the model.
Thus, we also further analyse the following models: SARIMA(4,1,1)(0,1,1)7, SARIMA(2,1,2)(0,1,1)7, SARIMA(4,1,2)(0,1,1)7 and SARIMA(2,1,1)(0,1,1)7

For P=3: Following the elbow method would indicate here to choose p=2 as the best parameter for the non-seasonal AR part of the model. It would also indicate here to choose q=1 and q=2 as the best parameters for the non-seasonal MA part of the model.
Thus, we also further analyse the following models: SARIMA(4,1,2)(3,1,1)7 (see ACF/PACF), SARIMA(2,1,1)(3,1,1)7 and SARIMA(2,1,2)(3,1,1)7

Disclaimer:
When training the selected models, we always received an error message for the SARIMA(2,1,1)(3,1,1)7 model. We could not fully identify the reason for this, but we suspect that the combination of parameters may not fit the time series and therefore the model cannot process the data properly. We therefore decided not to include this model in further analysis.

Select Model

Finally, we compare the AIC, RMSE/MAE and the distribution of the residuals of the different models to select the best model. In general, the distribution of the residuals should be normally distributed and the AIC and RMSE/MAE should be as low as possible. To keep the report short, we will only show the comparison of the AIC and RMSE/MAE of the models. The distribution plots as well as the ACF and PACF plots of the residuals can be seen in the code.

model	AIC	RMSE	MAE
SARIMA(4,1,1)(1,1,1)7	-1443.470	0.1552101	0.1045668
SARIMA(4,1,2)(1,1,1)7	-1443.470	0.1551451	0.1045362
SARIMA(2,1,1)(1,1,1)7	-1443.470	0.1551451	0.104609
SARIMA(2,1,2)(1,1,1)7	-1449.893	0.1548517	0.1046317
SARIMA(4,1,1)(0,1,1)7	-1431.256	0.1558858	0.1051186
SARIMA(2,1,2)(0,1,1)7	-1441.587	0.1553226	0.105138
SARIMA(4,1,2)(0,1,1)7	-1435.035	0.1556042	0.1051745
SARIMA(2,1,1)(0,1,1)7	-1433.308	0.1559736	0.1052392
SARIMA(2,1,2)(3,1,1)7	-1453.532	0.1546095	0.10477282
SARIMA(4,1,2)(3,1,1)7	-1449.431	0.1546508	0.1045284

By comparing the models based on the AIC, RMSE and MAE, we can see that all models are very close to each other. Thus, probably the results of the forecasting would be similar for all of the models. However,

the AIC is lowest for SARIMA(2,1,2)(3,1,1)7 what means the model explains the variability in the data with fewer parameters better than the other models. This model also has the lowest RMSE which indicates that this model can handle outliers slightly better than other models since the RMSE gives a relatively high weight to large errors because it squares the residuals before averaging. SARIMA(2,1,2)(3,1,1)7 has not the best MAE. This indicates that the model performs better at minimizing larger errors but might not be as accurate on average for all predictions.

Through analyzing the distribution of the residuals of the models (see plots in code), one can see for all models a normal distribution as it is supposed to be to ensure an appropriate capturing of the error structure without systematic bias. When analyzing the ACF and PACF plots of the residuals of the model, we can still detect some significant (partial) autocorrelation spikes. This indicates that the models are still not perfect and that there still might be some undetected patterns or structures left in the data. A further analysis is beyond the scope of this task and thus, will not be conducted.

Since all analyzed metrics (AIC, RMSE and MAE) are very close to each other, we decided to stick with SARIMA(2,1,2)(3,1,1)7 for forecasting the sales values of the next 15 days because this model showed the best AIC and RMSE results.

Forecast

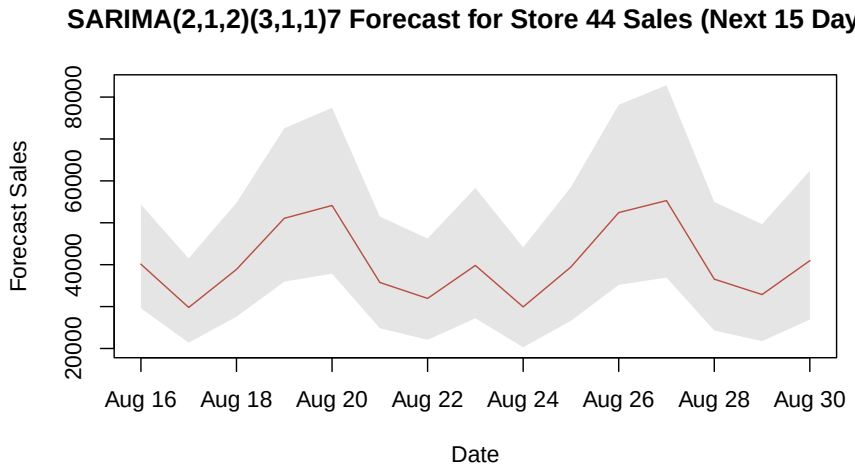
Finally, we forecast the next 15 days with the selected model. To do this, we had to make sure that we reversed the pre-processing steps that were applied to the training data. We then used the forecast package to forecast the next 15 days, which is stored in a table format that can be seen in the code. We then aggregated the sales to get a total of the predicted sales expected by the model.

```
## [1] "The aggregated mean forecast is:  608994"
```

```
## [1] "The aggregated lower forecast is:  476411"
```

```
## [1] "The aggregated upper forecast is:  778816"
```

We also plotted the predicted sales per day in a graph, which can be seen below.



Task 2

Data preparation

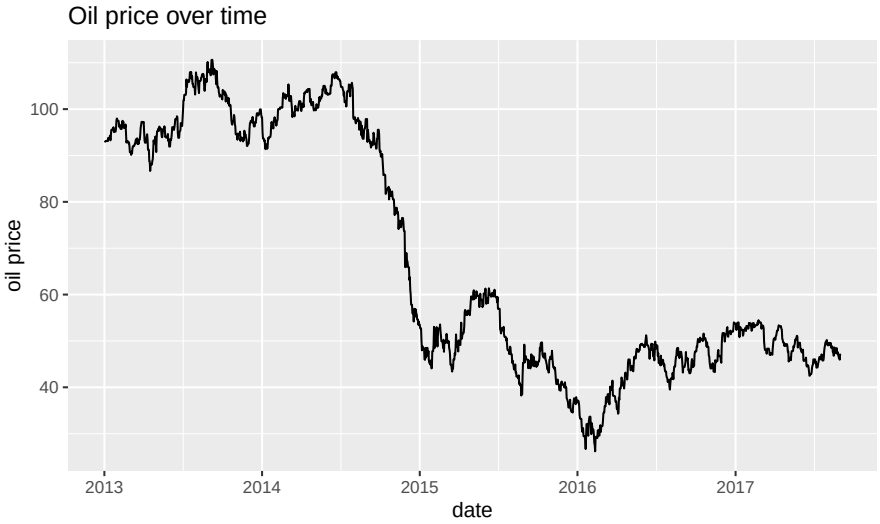
To analyse the interdependencies between store sales and two other factors, we chose the two factors 'oil price' and 'promotion sales'. We chose these two different factors because we thought that the oil price might be interesting as the Ecuadorian economy is very dependent on oil and an oil price shock could have a big impact on sales. The decision to also include promotions was made because we would expect that the number of promotional purchases could really influence total sales in the next few days, as people might tend to buy

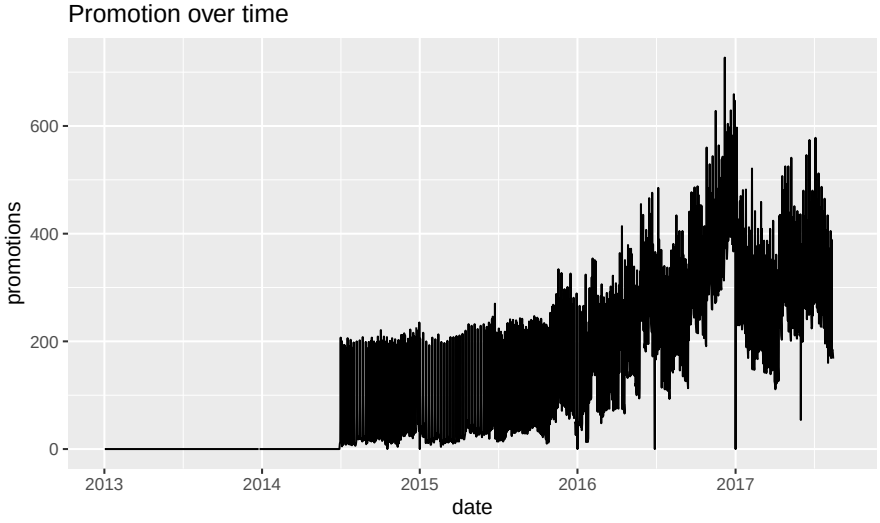
more stock when they get a good promotion and therefore need to shop less in the next few days. Another reason for choosing these two factors was that both values are on a numerical scale, which makes it easier to manipulate the data to simulate a shock.

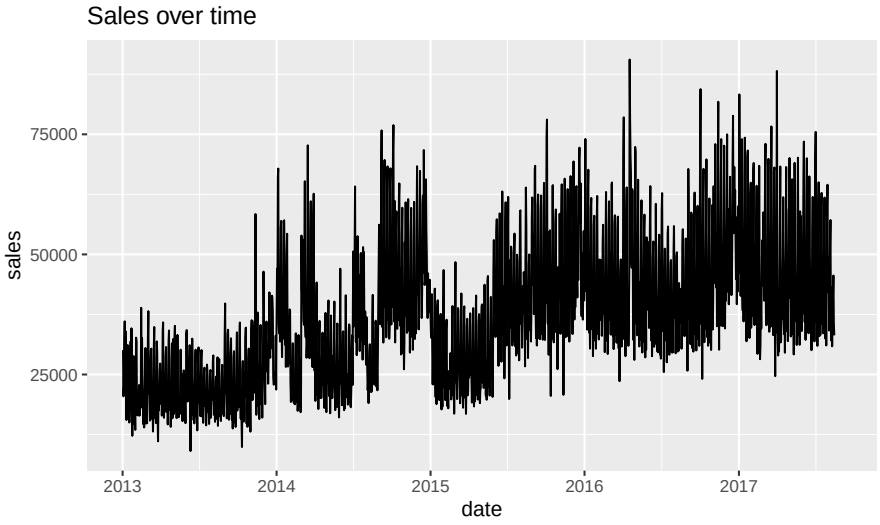
The first thing we did was to put all the information we needed (store 44 sales, promotions and oil price) into one data frame. Then we looked at the data to see if there was any missing data and if there was a trend in the data. The first thing we noticed was that there were no values for the oil price on weekends and also for some other dates in the data (possibly holidays). To fill in the missing values, we decided to simply use the oil price of the day before, as this is also the procedure of the stock markets, where prices are fixed over the weekend because the stock market is closed.

Regarding store sales, we have already noted that the store appears to be closed on the 25th of December and the 1st of January, so we have no values for total sales and promotions on these two dates. In order to be able to carry out our analysis, we decided to use the yearly average on these two dates for total sales & promotions.

We then looked at the graphs of these 3 time series to see if there were any trends or other things to look out for in the data.







The first thing we are able to see is that there are no promotions in the store until the 1st of July in 2014. So for this analysis we decided to only examine the period between 2014-07-01 and 2017-08-15.

For total sales, we already know from Task 1 that there is a trend in the data and a strong argument can be made for heteroskedasticity, so we decide to log total sales and perform differencing to remove heteroskedasticity and trend from the data.

For the promotions, we would also expect heteroskedasticity to be present because they are also sales data. Another argument for heteroscedasticity is that there are likely to be promotions on different days for different items, so we would expect the variance not to be constant over time. This is why we decided to log the promotions as well. Also, looking at the plot, there seems to be evidence of a trend, so we ran a KPSS test because this test was able to identify a trend in the sales data when the Dickey-Fuller test had a different result. Since we expect promotions to follow a similar pattern to sales (because promotions are somewhat dependent on overall sales), we ran the KPSS test directly.

The results of the test also support the assumption of a trend. Therefore, we also performed a differencing on the promotion data. To be able to use the results of the differencing & logging, we replaced every 0 in the dataset with a 0.1, which was only necessary on two dates (2014-10-18 & 2016-06-28).

For the oil price, we were not sure whether heteroskedasticity was present by looking at the plot. However, we decided to log the data anyway, as this allows us to account for heteroscedasticity if it is present, and even if it is not, this step does not harm our further analysis. To test whether there is a trend in the oil price, we performed an augmented Dickey-Fuller test.

This test result seems to support the hypothesis of a trend in the oil price data. We have therefore also performed a differencing on this time series.

Looking again at the plots of the three time series, we can see that the trend has been removed from the data and also because of the logging, heteroskedasticity should no longer be a problem in the data.

VAR model

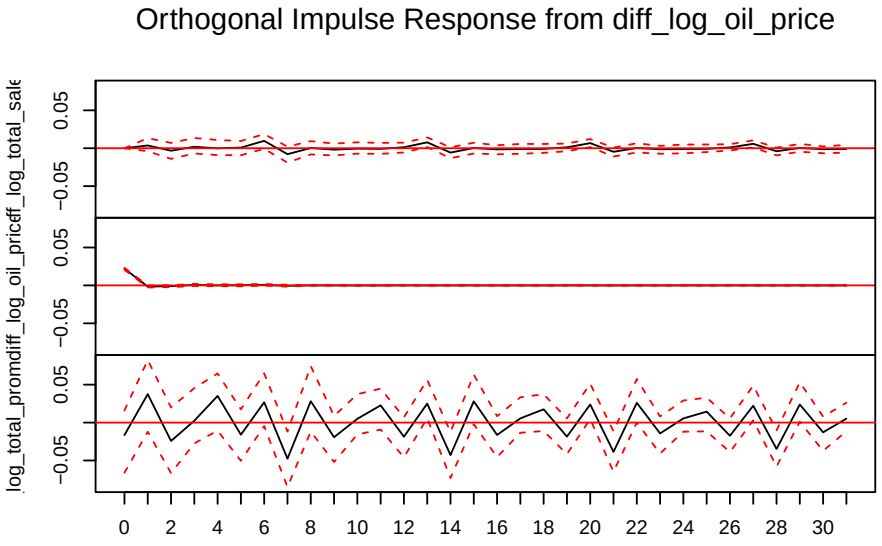
After accounting for trends and heteroskedasticity we analyzed the optimal number of lags to include in our VAR model. Therefore, we calculated the Bayesian Information Criterion (BIC) for the number of lags between 1 & 31 and then asked R to display the number of lags with the lowest BIC:

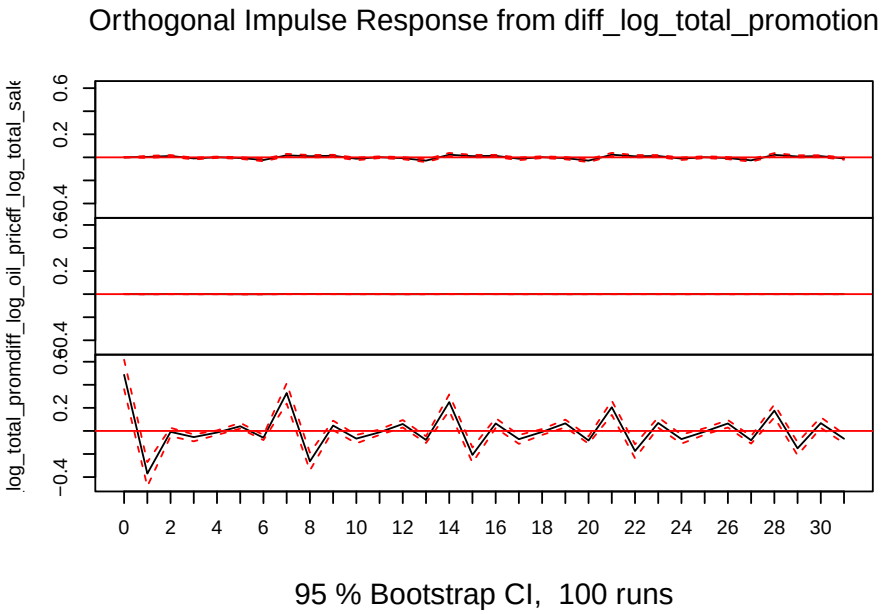
```
## [1] 7
```

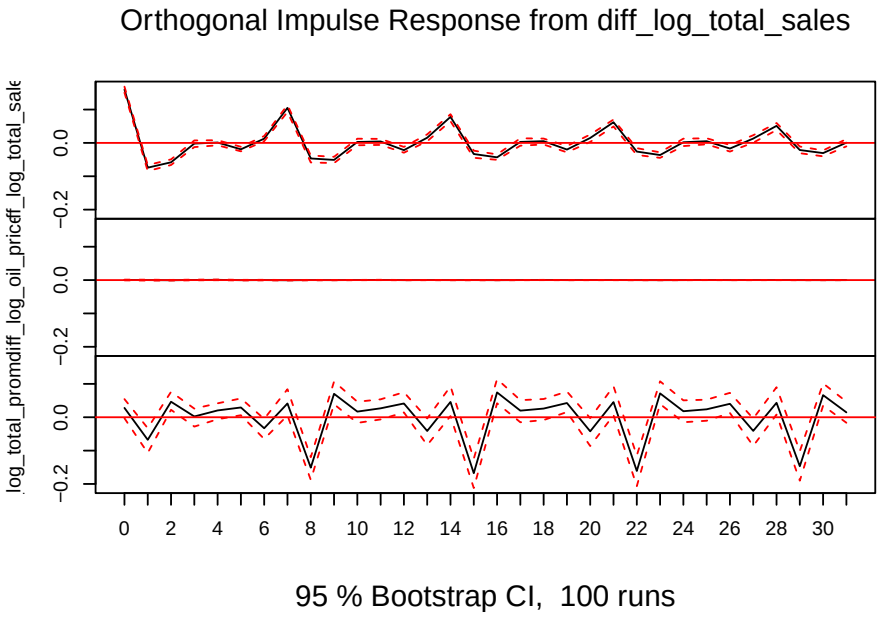
As we can see, the lowest BIC value was found when 7 lags were included. This makes sense as we have already seen in task 1 that there seems to be a weekly seasonality in the sales data. By including 7 lags (1 week) we are able to account for this seasonality in our model.

So now we run our model with the differenced & logged sales, promotion & oil price data by including 7 lags. The results of the model can be seen in the code.

For total sales we can see that the total sales of all 7 days before seem to have a significant influence, but also the oil price and promotions a few days before seem to be significant in predicting total sales. As the results of the VAR model are not intuitively interpretable, we have also plotted the impulse response functions of the VAR model to improve the interpretation:







Looking at the impulse response function, where the impulse was the oil price, we can see that an increase in the oil price does not seem to affect the total sales of our shop, as the response function for sales is not really different from zero. Furthermore, the confidence interval for the entire response function covers 0, which means that all the results are insignificant.

The effect of the oil price on the promotions is stronger, but again the confidence interval covers 0 for the whole function, which also means that the influence of the oil price on the promotions is not significant. Also, the effect of the oil price on the oil price itself does not seem to be present after the initial impulse, as the response of the oil price is basically zero after the first response.

When looking at the impulse response function where the promotions were the impulse we can also derive that the amounts of the promotions does not seem to influence the total sales since the whole response function is around 0. For the oil price we also cant see any influence of the promotions. The effect of an impulse in promotions on the promotions seems to be that after a lot of promotions on the next day there seem to be significantly less promotions and after seven days there are again a lot of promotions and the pattern starts again. This probably implies a weekly seasonality of the promotions.

For the impulse response function, where we have used sales as the impulse, we can infer that sales do not affect the oil price, as the function is again not different from 0. The effect of total sales on itself seems to be more or less the same as the effect of the promotion on itself. After the first impulse, the response function implies significantly lower sales the next day. Then again, the function is around zero until day 7 when there are significantly more sales and the pattern starts again. This also implies a weekly seasonality of sales. Looking at the effect of sales on promotions, we can see that a sales impulse also increases promotions significantly in the beginning, followed by significantly less promotion sales the next day. Again, the pattern repeats after 7 days, suggesting a weekly seasonality. It seems that promotions and total sales are closely linked, as the response function here shows a similar pattern. This also makes sense from a logical point of view, because if a supermarket has more sales, it seems intuitive that there will also be more sales of promoted items.

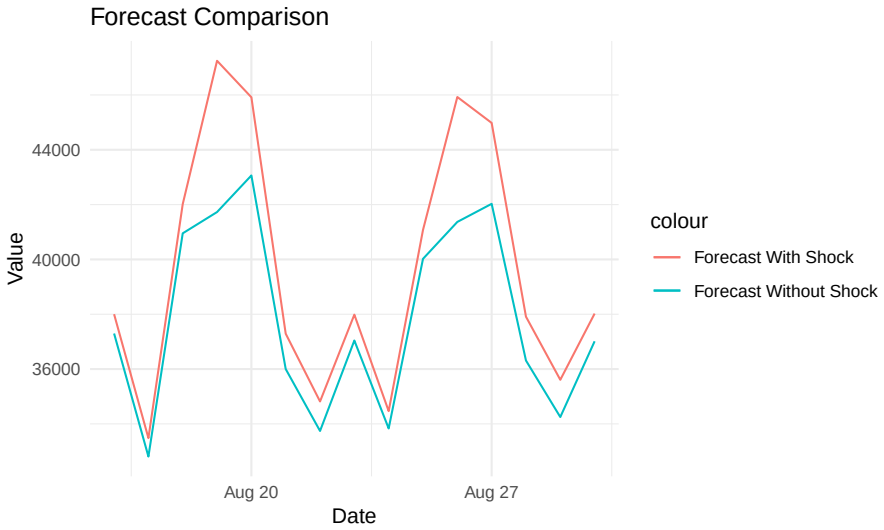
Oil price shock

We now want to simulate an oil price shock that results in a 25% increase in the oil price over the last three days of our data. A realistic reason for such an increase in the price of oil could be that it has been discovered that the oil reserves are smaller than expected. Another possible reason could be that the government has limited oil production to a certain amount, making the available oil more expensive.

From the impulse response functions, we assume that the oil price increase has little or no effect on total sales, since the response function is not really different from zero, with the confidence interval covering zero throughout the impulse response function. Logically, we would expect total sales to be lower after the shock because we would expect people to have less money to spend on groceries, which could lead to a decline in sales.

So we took our dataset and increased the oil price by 25% on the 15th, 14th and 13th of August 2017. Then, we did the differencing and logging of the oil price to get the values for the VAR model that includes the shock. After that, we calculated the forecast for the next 15 days for both models (with and without the shock). We can then compare the differences between the two models to see how sales were affected by the oil price shock.

Once we had calculated the forecast for the two different models, we reversed the differencing and logging step to obtain the forecast for the amount of sales in our store. We then created a new dataframe to compare the different results of the forecasts:



The plot shows that the oil price shock seems to increase the forecast for the next 15 days. The largest difference occurs on the 19th of August, when the forecast with the shock is about 5500 units higher than the forecast without the shock. After that, the differences between the two different forecasts become smaller again.

Overall, we would conclude that the shock did not make much of a difference to store orders, since on most days the forecast differs by only 1000-2000, which is less than 1% of total sales on those days. However, contrary to our initial assumption that the shock would reduce sales, it actually increased them. One reason for this could be that the people who shop in this store are part of the oil industry themselves, so they get more money when the oil price rises, which leads to more consumption. But without looking at revenue, such statements are just speculation, because it could be possible, for example, that total sales rise after an oil price shock, but the total revenue of the shops falls, so that people have simply become more efficient in their shopping. Another possible reason could be that the rise in the oil price made people think that the price of supermarket products would also rise soon, so they bought more stock in order to reduce their expenditure in the future (panic buying).