Causal Impact analysis on Tesla stock after Elon Musk’s acquisition of Twitter

Josip Ćurković

Zagreb School of Economics and Management

Zagreb, Croatia

[jcurkovi@student.zsem.hr](mailto:fpoljak@student.zsem.hr)

1. INTRODUCTION

In this paper, we will explore how Elon Musk’s acquisition of Twitter impacted Tesla stock, using the Causal Impact package in python. The goal is to get an insight into how this purchase affected Tesla as well as looking into how the stock would have most likely behaved had that purchase not happened. Firstly we will be creating an artificial time series dataset, where one time series will be the response time series which is affected by the intervention and the second will be the control time series which is not impacted. Afterwards we will apply the package on a real dataset with multiple control time series.

1. DATA SETS USED

The packages used in this seminar paper include as previously mentioned CausalImpact, for the causal impact analysis, yfinance for importing datasets from Yahoo finance directly into Jupyter lab. Pandas to be able to work with data frames and finally Seaborn and Matplotlib to have the ability to construct graphs. Using the yfinance package we downloaded the stock data for multiple companies from July of 2022 until March of 2023 with a one day interval. These companies include our target Tesla as well as BMW, Volkswagen, Hyundai Motor Co. and General Motors.

1. CAUSAL IMPACT

Causal Impact is an algorithm made by Google that is used to create a Bayesian structural time series model which is based on one or multiple control groups in order to estimate how an intervention will affect a series as well as how the series would have behaved had there been no intervention with given current data.

1. DATA MODELLING

i. ARTIFICIAL DATA SET

Now that we’ve gone over what exactly Causal Impact is, we will apply it firstly to our artificial dataset. In order to do so we will first need to create the control time series and the response time series. To do this we will first need to create an autoregressive moving average process and generate a sample from it with any size of our liking, in our case here we used n = 1000. From there we get our series X where as our y series is just a function of X with an intercept which is taken from a normal distribution. Finally once we have our data frane, we need to set the impact to occur at a specific time and have a set increase or decrease, in our case here the impact is set at +20 units after the 750th datapoint in our time series. In order to be able to use the CausalImpact package we need to create a data frame with our X and y time series and add date indexes for X and y to truly be time series. Once we have done all of this we can plot the data to get a better insight into what is happening with our data.Chart, line chart

Description automatically generated

Figure 1 Plot of X, y and the point of impact

Another necessity in order to use the CausalImpact package is to specify the pre-period and post-period. The pre-period is the period before the intervention happened, so that means that here we need to set it to be from the beginning of our dataset until the point in time right before the impact. So in our case since the point of impact is at the 750th datapoint, we need our pre-period to last until the 749th datapoint. The post-period is the period after the intervention, so in our case the post-period will be from the 750th point until the end. Now once we have done all of this we can finally use the CausalImpact package, to do so we need to use the command CausalImpact(…) and in the parenthesis specify the data frame and the pre and post-periods. We then plot the causal impact and get the following graphs.

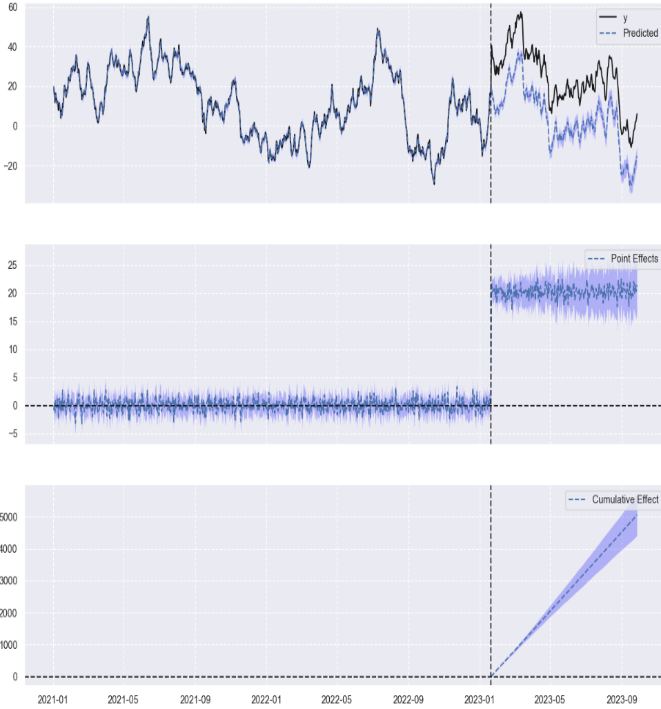


Figure 2 CausalImpact output for artificial dataset

Here we can see that at the point of impact our y and predicted y deviate from each other by those 20 units which we have specified at the beginning. This is what the goal of causal impact is, namely to show how an intervention will affect a time series at the point of intervention and in the future. Finally to conclude this part we will print the summary of the causal impact

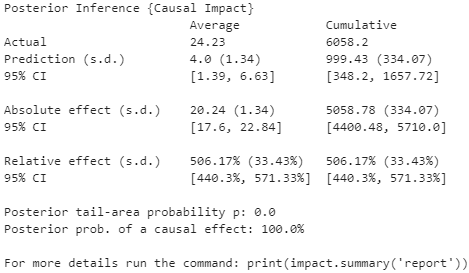


Figure 3 Causal Impact summary for artificial data set

From here we see the average and cumulative daily values compared to what was predicted by the model in the post-period. So in our case the average was 24.23 where our prediction was 4.0. This makes sense since our impact is in the post-period and if we subtract 4 from 24.23 we roughly get the value of our impact with some small deviation. The absolute effect shows us the average daily and cumulative difference between what our model predicted and the actual values in the post-period. The average difference was 20.24 and the 95% confidence interval was [17.6, 22.84]. And finally the relative effect just turns the absolute effect into a percentage.

ii. TESLA STOCK DATASET

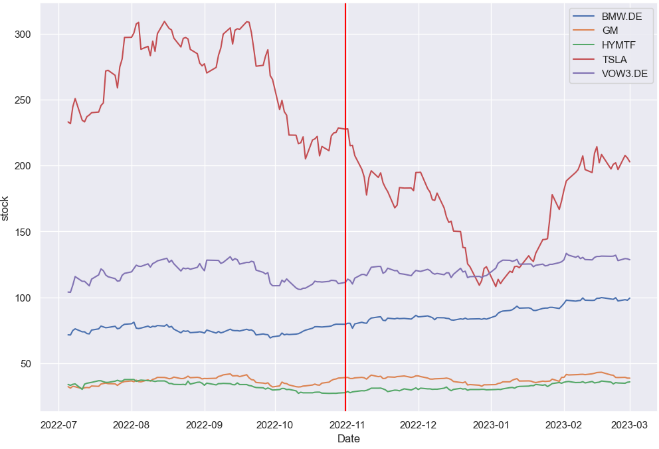
Now we will be taking a look at a real world data set, where our goal is to see whether Elon Musk’s acquisition of Twitter had an impact on the Tesla stock. To start we will firstly be importing our needed packages, so namely yfinance, pandas, matplotlib, seaborn and CausalImpact. Firstly we will need to specify the training start and end period as well as the treatment start and end period. Here we took the training period to be from the 5th of July 2022 until the 26th of October 2022 and the treatment period to be from the 27th of October 2022, which is the day he purchased Twitter, until the 1st of March 2023. Afterwards we need to get stock data for the period from the beginning of the training period to the end of the treatment period with an interval of 1 day. To do this we need to use the yfinance package and choose companies which are in the same or similar industry as Tesla. In this project I chose BMW, Volkswagen, Hyundai and General Motors. Now in order to continue we will need to do some data cleaning in the sense that from the dataset that yfinance has given us we need to firstly select only the adjusted close values and afterwards drop all of the NaN values. After we’ve done this we will use Seaborn to plot the stock data including the time at which the intervention happened.

Figure 4 Plot of the stock prices for over the whole period

Now we will need to check whether the stocks we have selected are good choices for our model. In order to do so we will be computing the correlation matrix for the training period only. To start we first have to create a new data frame which is equal to our old data frame in everything apart from it being filtered to only show the training period, so from the 5th of July 2022 to the 26th of October 2022. The reason we’re only using the training period is because we’re only interested in the correlation before the intervention as we know what happened before but we’re interested in how the stock will behave afterwards. With our new data set we just need to use the .corr() function and get our correlation matrix. Here we’re presented with the following matrix.

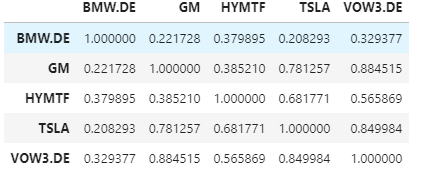


Figure 5 Correlation matrix

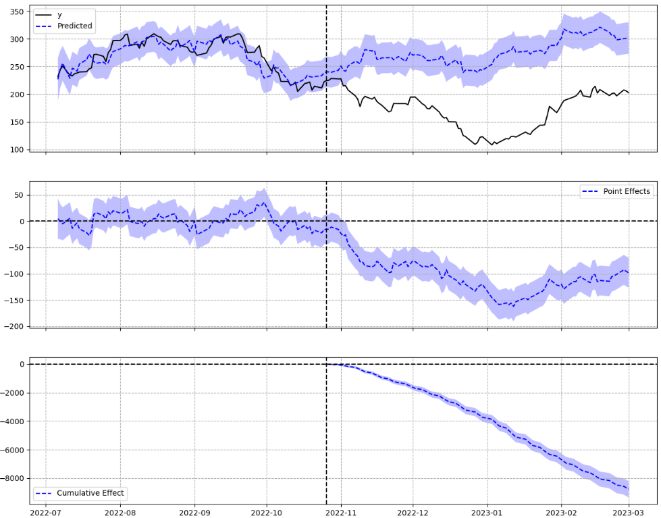
In this correlation matrix we’re only looking at the correlation between Tesla and the other companies. As we can see Tesla is fairly correlated with every company apart from BMW. Because of this we will now be removing BMW from our original data set which we will be using for the causal impact analysis. From here on we need to specify the pre-period and post-period. The pre-period is the period from the training start to training end, and the post period from the treatment start to the treatment end. Or in other words our pre-period is from the 5th of July 2022 until the 27th of October 2022, and out post-period is from the 28th of October 2022until the 1st of March 2023. Once this is done we can use the CausalImpact(…) function, which needs the data frame as well as the pre and post-periods as inputs. After we run the function we can use the .plot function to plot the causal impact.

Figure 6 Plot of the causal impact

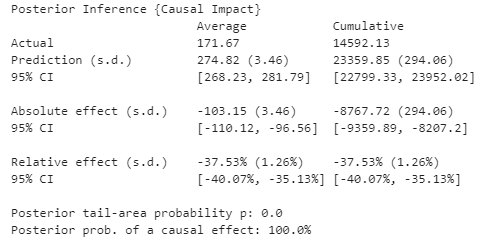
From these graphs we can see how at the intervention point the predicted and actual values start to deviate, namely the predicted values continue following the same trend as before, whereas the actual values start plummeting. Finally we use the .summary() function to see how substantial the changes in the predicted and actual values are.

Figure 7 Summary of the causal impact

Here we see what the average and cumulative daily predicted and actual values are. The average prediction was higher than the actual average by 103.15 units, this is an average decrease in 37.53%. Furthermore our posterior tail-area probability is equal to 0.0 which means that the causal impact is statistically significant, or in other words that the purchase of Twitter had a negative impact on the stock of Tesla.

1. Conclusion

To finish up, we have seen that the Causal Impact package is a very powerful tool in analyzing how a time series will behave after an intervention. We have also seen how Elon Musk’s purchase of Twitter and the rash decision he has done on it following the purchase affected the Tesla stock severely. Namely that actual average is less than the predicted average by more than 100 units, or in other words that it dropped by nearly 40%

# REFERENCES

(2020, 6 3). Retrieved from https://www.youtube.com/watch?v=86IEGJI1BlU

Amy. (2022, 9 10). *Medium*. Retrieved from Time Series Causal Impact Analysis in Python: https://medium.com/grabngoinfo/time-series-causal-impact-analysis-in-python-63eacb1df5cc

Brodersen, K. (2016, 12 13). Retrieved from https://www.youtube.com/watch?v=GTgZfCltMm8&t=206s

*Seer Interactive*. (2021, 10 5). Retrieved from What Is a Causal Impact Analysis and Why Should You Care?: https://www.seerinteractive.com/insights/what-is-a-causal-impact-analysis-and-why-should-you-care

Senouci, J. (n.d.). *nbviewer*. Retrieved from Causal Impact: https://nbviewer.org/github/jamalsenouci/causalimpact/blob/master/GettingStarted.ipynb