Dimitri Bouvier

# Analysis and Results Sections Paper

## 3. Analysis of results and discussions

My first attempt was to forecast sales volume daily, but that did not go well because the error metrics such as RMSE (the error metric I am optimizing the model for), and thus MAPE, were quite high and not accurate enough from a business standpoint even after running hyperparameters tuning. For a daily forecast, the average MAPE for both brands was around 30%. As a result, I decided to try a monthly forecast and was able to significantly reduce the RMSE and MAPE. Sandro's forecast has a MAPE of 11.5% and Maje's has a MAPE of 13.5%. More on that below.

Prophet computes these error rate values on unseen data by using forward-chaining cross validation. Prophet does not support k-fold cross validation, which is widely used in data science, because time series data violates the independence assumption on which k-fold cross validation is based. One cannot train a model using more recent data and then test the model by predicting older data. As a result, forward-chaining cross-validation (FCCV) is the type of cross validation I used to determine how well my model generalizes on unseen data.

In Prophet, FCCV works by starting with a training set of a specific size, known as the 'initial.' The model learns from this training set and then predicts future data points based on a value known as the horizon. The beginning of this time period is referred to as the 'cutoff.' Based on the predicted values y hat vs actual values y, the framework computes 7 error rates (mse, rmse, mae, mape, mdape, smape, and coverage). The distance between each 'cutoff,' that is, between each start of the prediction horizon, is referred to as the 'period.' The model learns from all available data up to each cutoff for each training fold and then predicts future values based on the horizon. Rafferty's [1] visual representation of Prophet's forward-chaining cross-validation is shown below.

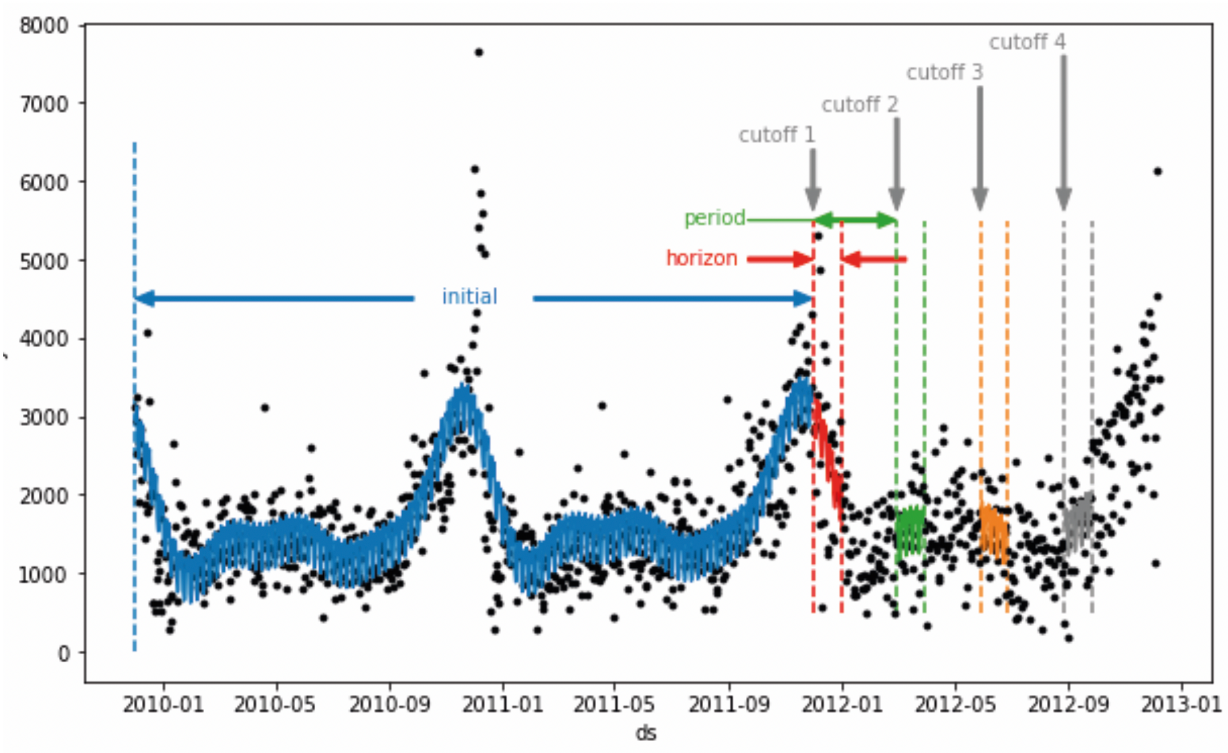



Figure 11.7 – Cross-validation terminology

The analyst provides input values for the horizon, initial, and period. The allocation and FP&A teams like to have a 6-month forecast to do all various types of planning, and this forecast is refreshed monthly. As a result, I set the 'initial' to 3 years (Prophet teams recommend three full cycles of seasonality, with the longest being a yearly seasonality in this case), the 'horizon' to 6 months, and the 'period' to 1 month. Then, in a pandas' data frame, I keep track of all folds' error rate computations, where each error metric is a column and the values of each fold's error metrics are rows. The average of each error rate is then computed.

The final crucial step in obtaining the best model is hyper parameter tuning. In my case scenario, two parameters are worth tuning: changepoint prior scale and seasonality prior scale. These prior scales influence the trend and seasonality components of the forecast. A lower prior scale value results in greater regularization of the component to which the prior scale is applied. See below for a visualization of such a technique by Rafferty [1]:

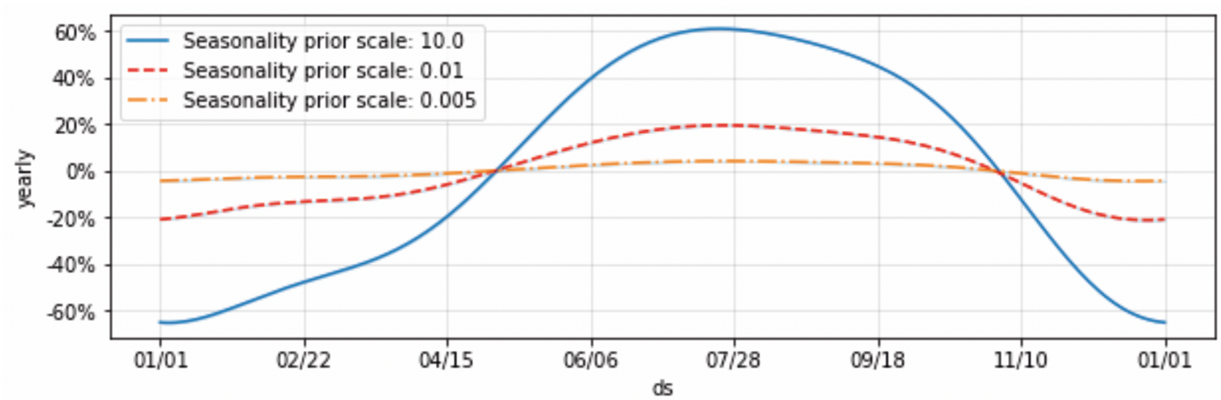



Figure 4.21 – Yearly seasonality with different prior scales

After many experiments and combinations of external variables (additional regressors in Prophet's jargon) fed to the model, I discovered that feeding the discount rate, University of Michigan's consumer sentiment (UMCSENT) [2], and unemployment rate (UNRATE) [3] to the model resulted in the best RMSE score and tightest confidence intervals for the trend component of the forecast. I am modeling 99% confidence intervals on all components of the forecast (trend, seasonality, and external regressors) by computing 300 Markov Chain Monte Carlo samples.

In terms of the academic portion of this project, I intend to present and analyze the model's results for one brand only, Sandro, in order to demonstrate competency in analyzing the results while not making this paper unnecessarily long. The format of the results is the same for both brands, only the numbers change

In [30]:
```
results.sort_values(by=['rmse'],ascending=True).head(20)
```

Out[30]:

| | changepoint_prior_scale | seasonality_prior_scale | horizon | mse | rmse | mae | mape | mdape | smape | coverage |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 0.010 | 10.00 | 180 days | 2.281960e+07 | 4776.986467 | 3361.459952 | 0.115173 | 0.098560 | 0.112675 | 0.765957 |
| 9 | 0.010 | 1.00 | 180 days | 2.300100e+07 | 4795.935598 | 3383.259807 | 0.115875 | 0.098199 | 0.113474 | 0.761229 |
| 10 | 0.010 | 0.10 | 180 days | 2.322178e+07 | 4818.898605 | 3394.065417 | 0.116086 | 0.099980 | 0.113516 | 0.761229 |
| 12 | 0.001 | 10.00 | 180 days | 2.509426e+07 | 5009.417109 | 3531.243158 | 0.121767 | 0.103829 | 0.118030 | 0.763593 |
| 13 | 0.001 | 1.00 | 180 days | 2.529582e+07 | 5029.495113 | 3551.617009 | 0.122295 | 0.101641 | 0.118715 | 0.754137 |
| 14 | 0.001 | 0.10 | 180 days | 2.550918e+07 | 5050.661429 | 3550.749633 | 0.122292 | 0.101068 | 0.118454 | 0.735225 |
| 11 | 0.010 | 0.01 | 180 days | 2.745957e+07 | 5240.187564 | 3685.031321 | 0.127109 | 0.112564 | 0.126322 | 0.815603 |
| 15 | 0.001 | 0.01 | 180 days | 2.988316e+07 | 5466.548800 | 3819.345441 | 0.132243 | 0.114738 | 0.130433 | 0.829787 |
| 4 | 0.100 | 10.00 | 180 days | 3.844872e+07 | 6200.703104 | 3958.192036 | 0.136267 | 0.110201 | 0.127799 | 0.695035 |
| 5 | 0.100 | 1.00 | 180 days | 3.856295e+07 | 6209.907176 | 3969.381875 | 0.136707 | 0.110333 | 0.128182 | 0.709220 |
| 6 | 0.100 | 0.10 | 180 days | 4.027029e+07 | 6345.887956 | 3996.814932 | 0.137256 | 0.110651 | 0.128243 | 0.709220 |
| 1 | 0.500 | 1.00 | 180 days | 4.492458e+07 | 6702.580430 | 4258.470947 | 0.147317 | 0.103974 | 0.137552 | 0.661939 |
| 0 | 0.500 | 10.00 | 180 days | 4.495998e+07 | 6705.220291 | 4264.901594 | 0.147616 | 0.105261 | 0.137911 | 0.659574 |
| 2 | 0.500 | 0.10 | 180 days | 4.825006e+07 | 6946.225974 | 4299.582764 | 0.148276 | 0.102442 | 0.137408 | 0.661939 |
| 7 | 0.100 | 0.01 | 180 days | 5.247835e+07 | 7244.194189 | 4592.565453 | 0.157809 | 0.113415 | 0.146779 | 0.756501 |
| 3 | 0.500 | 0.01 | 180 days | 1.147398e+08 | 10711.666368 | 5494.073631 | 0.186437 | 0.105062 | 0.158842 | 0.751773 |

The first results I'd like to present are the hyperparameter tuning results. The combination of changepoint prior scale = 0.01 (high regularization) and seasonality prior scale = 10.0 (no regularization) had the lowest RMSE, corresponding to an 11.51% MAPE, out of 16 combinations. These are the parameters I used in the final model.

Before diving into the model's results, I'd like to remind readers that I've hidden all data from 2020-01-01 to 2021-04-01 from the model so that it doesn't learn anything from this unusual time due to a black swan event. When this period is fed into the model, the confidence intervals and uncertainty around the trend grow significantly wider because the model starts to model covid’s impact as part of a new yearly seasonality pattern, which is obviously incorrect.

The full forecast results are listed below:

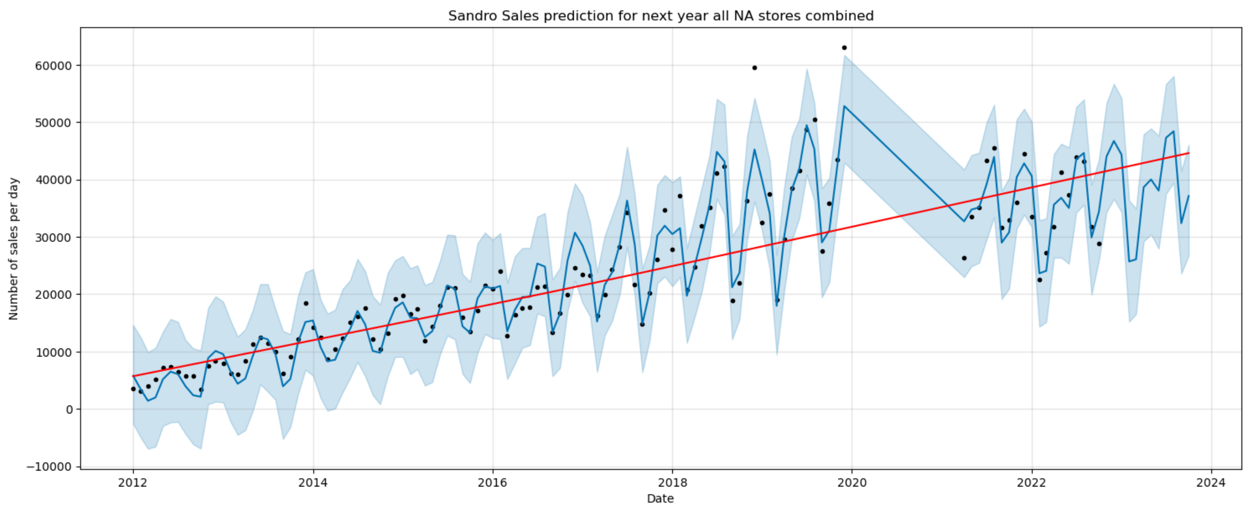



Figure 1 – Forecast and trend plot

Figure 1 is most likely Prophet's most important plot. This plot depicts the actual forecast as well as the trend component. As previously stated, the prior scale of the changepoint is low, implying that the trend component is subjected to a high level of regularization. As a result, the long-term trend is fairly rigid. According to the model, the trend component represents the time series "organic' trend that has been isolated from seasonality, holiday effects, and extra regressors. Each component's confidence interval is represented by the shaded areas. In this case, I set the confidence level to 99%, which explains why the shaded area is rather large. Reduced confidence tightens the shaded areas but decreases the likelihood that the actual components are within the shaded areas.

The model paints a positive picture: organic demand for Sandro's item is on the rise and shows no signs of losing steam, according to the most recent data points. It increased from about 7.5k units sold per month in early 2012 to about 40k units sold per month at the time of writing this paper. However, actual sales volume appears to be leveling off since pre-pandemic levels. This is due to other forecast components, which I will discuss below.

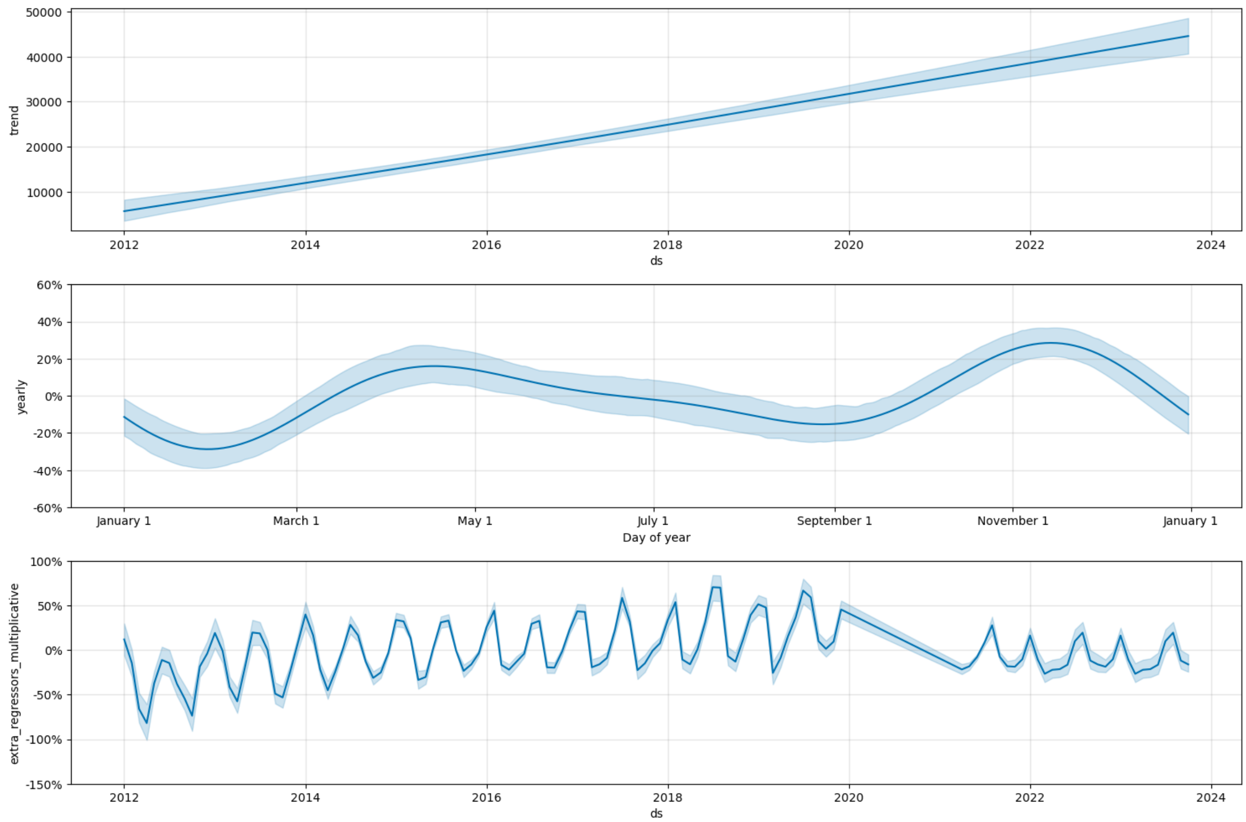



Figure 2 – Standalone trend, yearly seasonality, and combined effect extra regressors plots

The first subplot of Figure 2 has already been discussed. This is the same trendline as shown in Figure 1.

Figure 2's second subplot depicts the yearly seasonality. On average, the lowest sales volume of the year occurs in February, with sales volume typically 25% lower than the trend. Two real-world factors that I can think of that could explain such a phenomenon are that by that time of year, winter markdowns are usually over, and most people have already purchased their winter clothes, so there is nothing going on to boost sales volumes.

On the other hand, the peak of sales volume during a sales year typically occurs around November/December, with sales volume being 30% higher than the trend. Two real-world factors that come to mind are that during this time period, important retail calendar events such as Black Friday and Christmas occur, which significantly boosts sales volume, and that temperatures begin to cool in North America, prompting consumers to begin purchasing winter clothing.

Figure 2's third subplot depicts the combined effect of additional variables on the predicted variable. This plot is difficult to analyze and interpret without separating the additional variables, but I can note that we have been oscillating between 25% below the trend and 20% above it since pre-pandemic times. Let's take a closer look at it below.

| | regressor | regressor_mode | center | coef_lower | coef | coef_upper |
|---|---|---|---|---|---|---|
| 0 | cci | multiplicative | 85.836522 | 0.000689 | 0.003497 | 0.006415 |
| 1 | drate | multiplicative | 41.755254 | 0.012197 | 0.015755 | 0.019613 |
| 2 | un | multiplicative | 5.272174 | -0.119647 | -0.072349 | -0.027336 |

Figure 3 – Extra regressor coefficients

Figure’s 3 terminology:

Cci: University of Michigan’s consumer sentiment

Drate: Discount rate

Un: Unemployment rate

The linear coefficients of additional regressors are depicted in Figure 3. The 'coef' column represents the predicted variable's percent increase/decrease for a one unit increase in the additional regressor from its center. For example, if we look at the 'un' variable, we can see that an increase in one unit, in this case the unemployment rate, results in a 7.23% decrease in the predicted variable, in this case sales volume. In simpler words, this means that historically (since 2012), a 1% increase in unemployment in the US economy reduces sales volume by 7.23%. In plain language, a 1% increase in the US unemployment rate has historically resulted in a 2.73% to 11.9% decrease in sales volume with a 99% certainty, with 7.23% being the most likely sales decrease rate, according to the model.

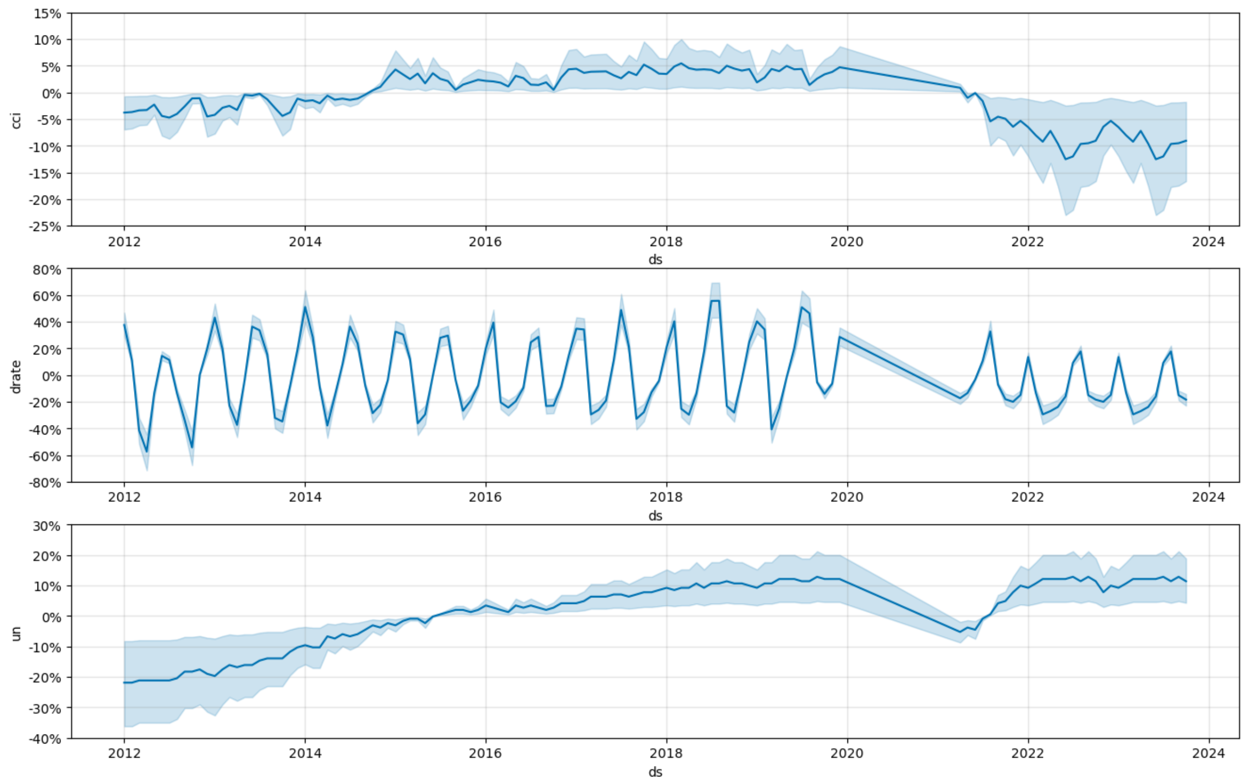


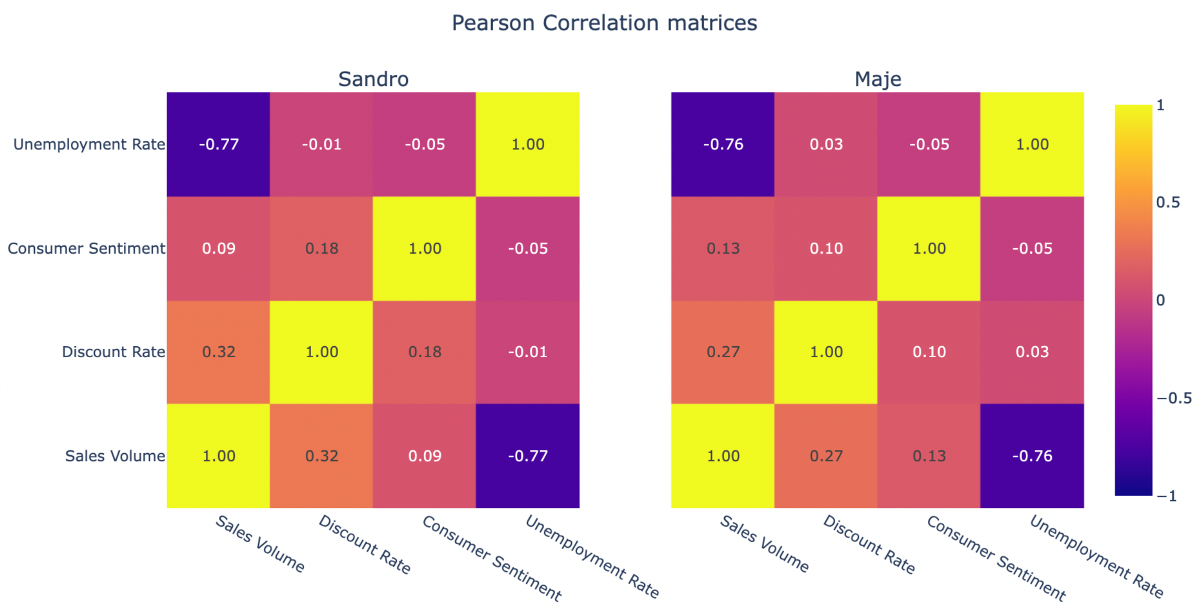

Figure 4 – Separated additional regressors impacts

The figure above depicts the historical standalone impact of each additional regressor on sales volume. Once again, the 99% confidence intervals are worth paying attention to. We can see that the discount rate has tighter confidence intervals than that of the unemployment rate (un) and consumer sentiment (cci), indicating that Prophet understands that the relationship between the discount rate and sales volume has less uncertainty.

All three variables have a significant impact on sales volume, with the discount rate being the only variable over which SMCP has control. I would like to point out that the unemployment rate has a -0.77 Pearson correlation with sales volume, whereas the discount rate has a positive Pearson correlation with sales volume of 0.32. (see appendix A). I hesitated to include the consumer sentiment index in this analysis because it has a lower correlation rate (0.09) with sales volume but also, according to Prophet, has a wide confidence interval. However, when performing FCCV and hyper parameter tuning, adding it to the mix of additional regressors still helped in achieving a lower RMSE and MAPE, so I decided to keep it.

Dimitri Bouvier

Appendix A. Correlation coefficients




References


1. Rafferty, Greg. “Forecasting Time Series Data with Facebook Prophet.” *Build, Improve, and Optimize Time Series Forecasting Models Using the Advanced Forecasting Tool*, 2021.
2. “University of Michigan: Consumer Sentiment.” *University of Michigan: Consumer Sentiment (UMCSENT) | FRED | St. Louis Fed*, 23 Nov. 2022, fred.stlouisfed.org/series/UMCSENT.
3. “Unemployment Rate.” *Unemployment Rate (UNRATE) | FRED | St. Louis Fed*, 4 Nov. 2022, fred.stlouisfed.org/series/UNRATE.