Haensel AMS Job Application Task: Mixed Marketing Model

Applicant: **Philipp von Mengersen**

June 20, 2024

1 Model

In the provided dataset we can find sales and channel spend timeseries data for 2 years. In order to understand, how the marketing actions impact the sales company X's online shop, we have to associate the channel spend with the sales series. Therefore I created two different models:

- The first model serves as our baseline model, It is a simple multivariate bayesian regression model, that only considers the direct impact of the weekly ad-spendings on the sales, without any additional effects.
- The second model (called mmm) incorporates some additional features, that are known to have an impact on the sales: (i) The adstock effect, which models the delayed impact of the ads on the sales. (ii) The trend, seasonality and the blackfriday, which are known to have an impact on the sales.

The basis is an additive model, that means I identify multiple components of the observed variable (here: sales) and then add these together. In the following I will explain the components of the model in more detail:

1.1 Media Contributions

The media contributions follow this schema:

$$\text{sales} \sim \mathcal{N}\left(\sum_{i=1}^{N} \beta_i \cdot \text{channel\_spend}_i, \sigma\right)$$

with $\beta_i \sim \text{TruncatedNormal}(\mu_i, 0.1, 0, 0.5)$[1] and $\sigma \sim \text{Uniform}(0, 0.5)$.

1.2 Adstock Effect

The adstock effect with the α and θ parameters is calculated using weights that decay according to both parameters. The weights for period k are given by:

$$w_k = \alpha^{(k-\theta)^2}$$

These weights are normalized by dividing each by the sum of all weights to ensure they sum to 1:

$$\tilde{w}_k = \frac{w_k}{\sum_{j=0}^{L-1} w_j} = \frac{\alpha^{(k-\theta)^2}}{\sum_{j=0}^{L-1} \alpha^{(j-\theta)^2}}$$

The adstocked spend at time t is then obtained by convolving the advertising spend Spend_t with these normalized weights:

$$\text{Adstock}_t = \sum_{k=0}^{L-1} \tilde{w}_k \cdot \text{Spend}_{t-k}$$

This formula captures the weighted effect of advertising spend over time, incorporating both decay and a shift controlled by the parameters α and θ.

[1] the last two parameter of truncated normal are the lower and higher bound

1.3 Trend + Season + Blackfriday

1.3.1 Trend

To model the trend over time, we employ a logistic curve, which is a commonly used function for representing (negative) growth processes that start slowly, accelerate, and then decelerate as they approach an asymptote. The logistic curve is defined as follows:[2]

$$f(t) = \frac{capacity}{1 + \exp(-k \cdot (t - t_0))}$$

1.3.2 Seasonality

To model seasonality in time series data, we use a cyclic component that captures the periodic fluctuations inherent in the data. This component can be expressed as a combination of sine and cosine functions, which allows for the representation of seasonal patterns. The cyclic component is defined as follows: (with period = 52)

$$g(t) = a_0 + a_1 \cos\left(\frac{2\pi t}{\text{period}}\right) + b_1 \sin\left(\frac{2\pi t}{\text{period}}\right)$$

1.3.3 Black Friday

The contribution of Black Friday to the overall analysis is modeled by multiplying this binary indicator by a coefficient and the corresponding data value at each time point. Mathematically, the contribution BF_Contrib_t at time t is given by: $\text{BF\_Contrib}_t = \text{BlackFriday}(t) \cdot \beta_{\text{BF}} \cdot (\text{trend}_t + \text{cycle}_t)$ and the indicator function is defined as:

$$\text{BlackFriday}(t) = \begin{cases} 1 & \text{if } t \mod \text{period} = \text{offset} \\ 0 & \text{otherwise} \end{cases}$$

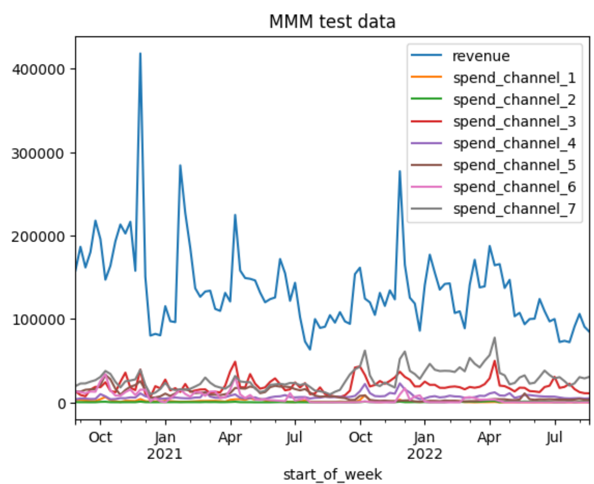

Figure 1: Original Data

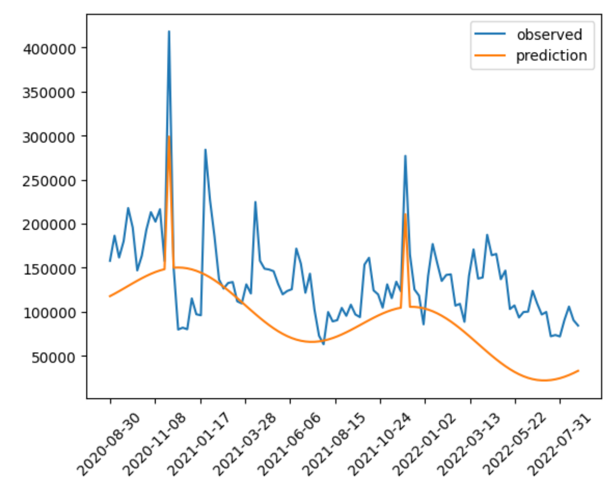

Figure 2: Trend + Season + Black Friday

1.4 Prediction Analysis

In the Prediction Analysis Figures, we can see the prior and posterior predictions of the baseline and the mmm model. For the priors, I decided on the following values. Since no label for the spend channel was given, I assumed the same initial prior contribution, lag and decay for all channels and wrapped in the same distributions. The prior and posterior predictions are shown in the figures Prior Predictions and Posterior Predictions. All the values have been choosen by thinking of a plausible range for the parameters. This was an experimental process, in which I tried to find meaningful values, that represent some real aspect, rather than forceful pushing them to fit the data.

[2]for more details regarding the individual parameters, please refer to the notebook

Variable	Description	Distribution and Parameters
L	describes period length of adstock impact	constant values = 5
$\alpha_{channel}$	Adstock decay parameter for each channel	$\alpha_{channel} \sim \mathcal{U}(0.3, 0.9)$
$\theta_{channel}$	Adstock delay parameter for each channel	$\theta_{channel} \sim \text{DiscreteUniform}(0, L-1)$
$\text{weights}_{channel}$	Weights for adstock effect	$\text{weights}_{channel} = \alpha_{channel}^{(\text{range}(L)-\theta_{channel})^2}$
μ_{channel}	Mean contribution coefficient for each channel	inital value is constant = 0.05 for all channels
$\text{coefficient}_{channel}$	Contribution coefficient for each channel	$\text{coefficient}_{channel} \sim \text{TruncatedNormal}(\mu_{\text{channel}}, 0.1, 0, 0.5)$
capacity_coef	Coefficient for logistic growth capacity	$\text{capacity\_coef} \sim \mathcal{N}(1.2, 0.3)$
growth_rate	Growth rate for logistic trend	$\text{growth\_rate} \sim \text{TruncatedNormal}(-0.02, 0.1, -0.15, 0.0)$
cos_coef	Coefficient for cosine component	$\text{cos\_coef} \sim \mathcal{N}(-1, 0.3)$
sin_coef	Coefficient for sine component	$\text{sin\_coef} \sim \mathcal{N}(2, 0.3)$
year_coeff	Scaling coefficient for yearly component	$\text{year\_coeff} \sim \mathcal{N}(0.1, 0.05)$
black_friday_coef	Coefficient for Black Friday effect	$\text{black\_friday\_coef} \sim \mathcal{N}(1, 0.2)$
σ	Standard deviation of sales (observed variable)	$\sigma \sim \mathcal{U}(0, 0.5)$

Table 1: Summary of Model Variables and Their Distributions

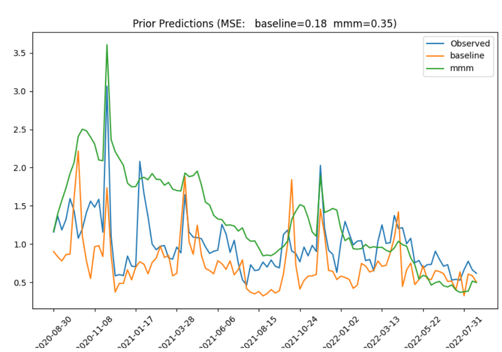

Figure 3: Prior Predictions

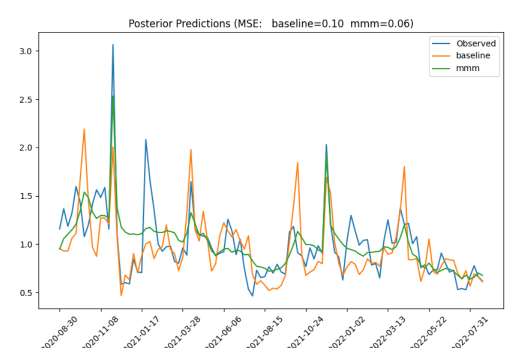

Figure 4: Posterior Predictions

1.5 Model Performance

As we can see in the table Model Performance, the mmm model performs worse in the prior, but becomes better in posterior (regarding the mean squared error). Also by comparing the loo likelihoods with the arviz compare function, we can see that the mmm model is better than the baseline model. (The resuls presented here can slightly differ from the one on the notebook)

Model	MSE
baseline prior	0.20
baseline posterior	0.31
mmm prior	0.10
mmm posterior	**0.06**

Table 2: Model Performance

	rank	elpd_loo	p_loo	elpd_diff	Weight	SE
mmm	0	-7.86	11.00	0.00	0.88	10.52
baseline	1	-34.47	10.28	26.60	0.12	13.33

Table 3: LOO Arviz Comparison (posterior)

2 Channel Spend Analysis

Now that we have a model that fits the data, we can analize the impact of the individual channels on the sales. On the one side we can see how much each channel contributes to the sales, and on the other side we can see the ROI of each channel. The ROI is given by the ratio of the sales to the channel spend.

We can see that channel 2 has an outstanding ROI. Since the model contains some bias, we probably should not take the ROI values too seriously, but we can see that the channel 2 has a much higher ROI than the other channels. In the second table, we can observe the other channels without channel 2 with a better scaling. Interesting here is, that channel 7 has the highest contribution, but the ROI is the lowest.

Now we could argue, that we should invest more in channel 2, since it has the highest ROI, but we should also consider the contribution of the channels. Another interesting measure, would be the channel_spend_coefficients, since they show how much the sales increase for a 1 unit increase in the channel spend. Since there are also other effects we have to consider, like saturation, channel independent phenomena that drive the sales, etc. we can not immediately take the ROI as the only measure for the channel spend.

In addition to that, we should also take more advantage of our prior knowledge about the channels, so we get more precise results.

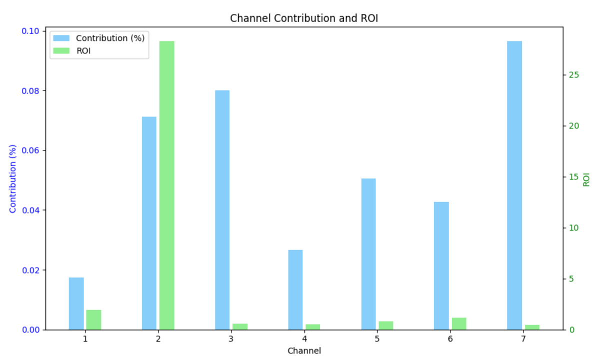

Figure 5: Channel Spend Analysis

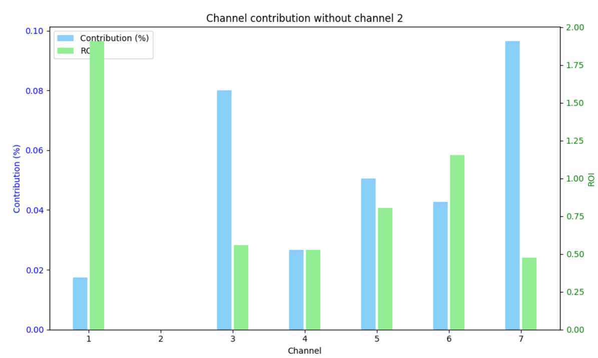

Figure 6: Channel Spend Analysis (Without Channel 2)

3 Discussion

As we can see in the Prediction Analysis Section, the model does not properly fit the data and still many adaption are to be made. Part of this work, was to show which basic effects can be considered in a model and how to implement them in a bayesian framework.

With ongoing improvement of the model, it might be possible to generate more accurate predictions and to better understand the impact of the marketing actions on the sales.