HAMS Data Science Challenge

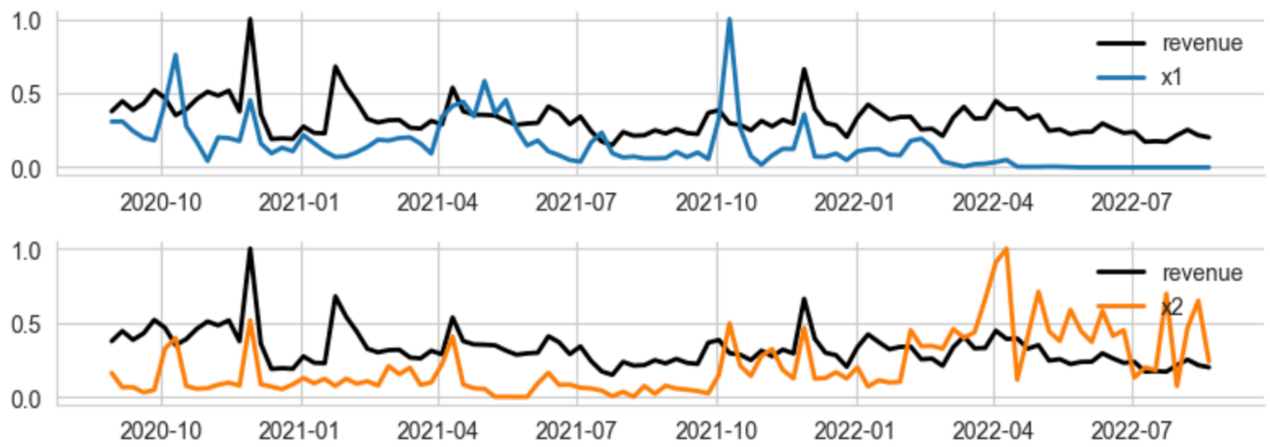
1. How do you model spend carry over?

We model the carryover effect using a **delayed adstock** transformation with the weights of the adstock function are given by:

$$w_m^d(l; \alpha_m, \theta_m) = \alpha_m^{(l-\theta_m)^2}, \quad l = 0, \ldots, L-1, \quad 0 < \alpha < 1, \quad 0 \leq \theta_m \leq L-1$$

where α_m is the retention rate of the ad effect in the m-th media channel (m=1…7) from one period to the next, θ_m is the delay of the peak effect and l=1,…,L-1, denotes the week numbers over which we model the carryover effect. We choose a rather large L (=13) to be the maximum lag period.

Motivation: A comparison of the maximum-absolute-scaled (MaxAbsScaled) media variables time series with the MaxAbsScaled revenue seems to indicate, for certain media variables (eg, channels 1,2,5), there is a delayed response to revenue, i.e, a peak in the media spend is followed after some time with a peak in the revenue. While, there are other variables which do not show any prominent delayed response and could have been modelled using a geometric adstock function, for simplicity we use the delayed adstock function for all the media variables.



Caption: Comparison of MaxAbsScaled revenue and spend channels 1 and 2; peaks in the media spends seem to be followed by peaks in revenue, suggesting a possible delayed response to the marketing spend.

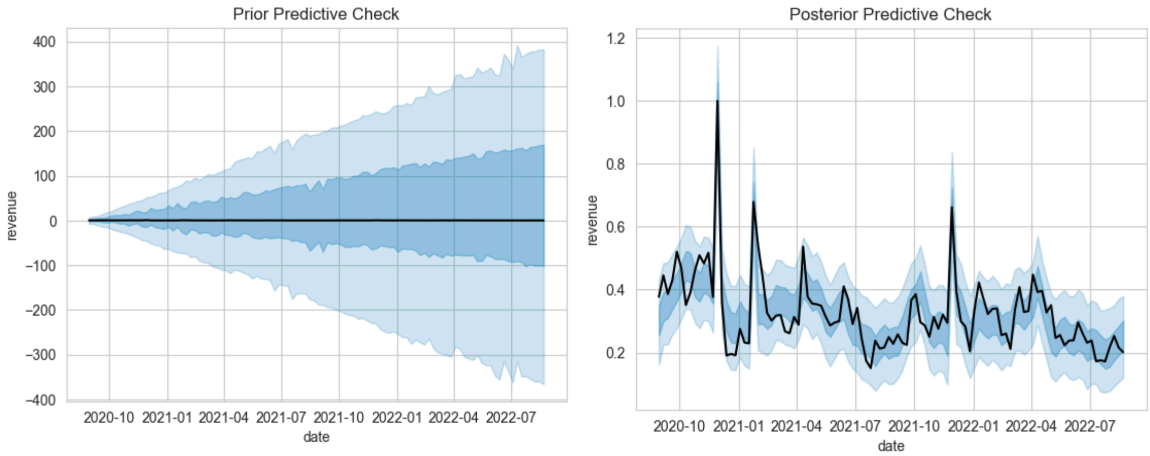
2. Explain your choice of prior inputs to the model?

Parameter	Prior	Reasons for Choice
τ	$\mathrm{Normal}(0, 2)$	
α	$\mathrm{Beta}(1, 3)$	$\alpha \in [0, 1)$; Other choices: Uniform()
θ	$\mathrm{Uniform}(0, 12)$	$\theta \in [0, L)$ and here $L = 13$; Other choices: Beta()
L	13	Without prior information, L can be set to a very large number, as an approximation to infinity, so the weights $w_m(l) \sim 0$ for $l > L$
β	$\mathrm{HalfNormal}(0, 1)$	$\beta_m > 0$, since media effect is assumed non-negative
λ	$\mathrm{Gamma}(3, 1)$	Following Jin et al, Google Inc. (2017) which says a common prior is a gamma distribution with a positive mode.
γ	$\mathrm{Normal}(0, 2)$	
σ	$\mathrm{HalfNormal}(0, 2)$	

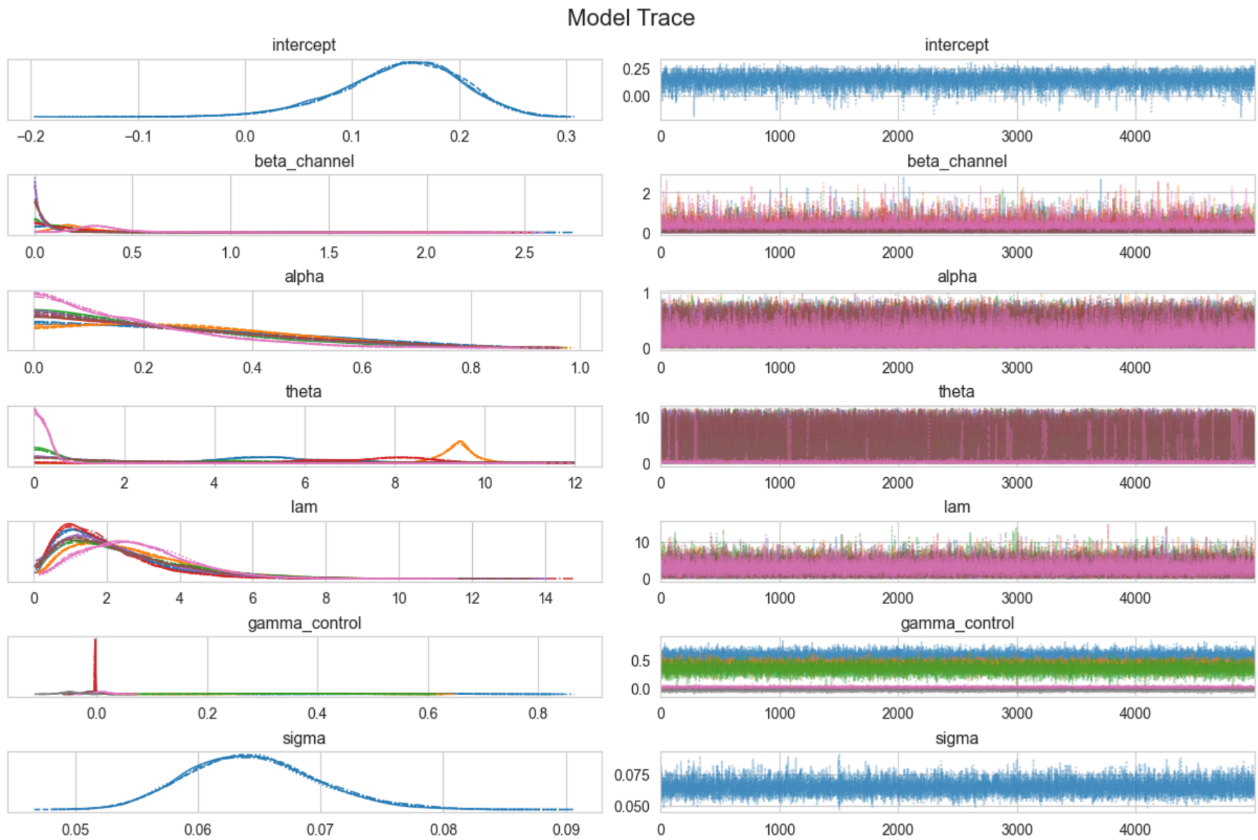
For some parameters, eg, α and λ, we tried out a few different priors, eg Uniform(), Beta(a,b) with a few different choices of (a,b), within the scope of our limited testing time, these were the prior choices that gave stable sampling results, and hence, we report them here.

3. How are your model results based on prior sampling vs. posterior sampling?

Here is a comparison of the prior and the posterior probability distributions on our target variable, the MaxAbsScaled revenue. The light and dark blue bands represent the 50 and 94% credible intervals, while the black solid line in the right hand plot represents actual MaxAbsScaled revenue provided to us.



4. How good is your model performing? How you do measure it?



We calculated a crude Bayesian Information Criteria (BIC) for our model, which was ~ 15. In the time I could spend on this case study, I tried out only one model, because of which I could not compare BICs against different models. For eg, Section 8 in Jin et al, 2017 compares 4 different models (which differ on the functional forms of the adstock and the shape transformations) using the BIC.

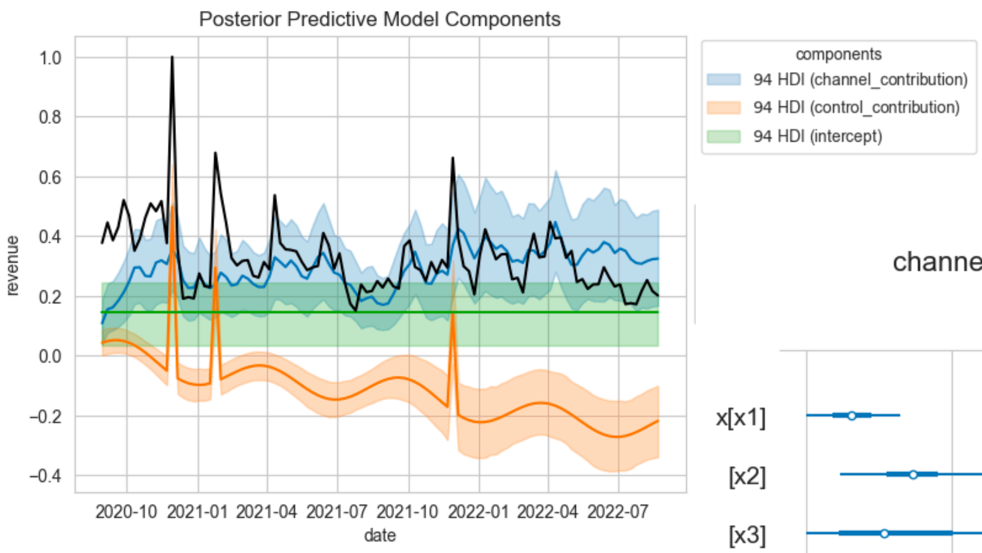
A more qualitative way to diagnose our model would be to check it's convergence using the model trace. In that regard, all our MCMC chains seem to have converged well (rhat=1), providing smooth posteriors on all our parameters [See Model Trace plot above].

5. What are your main insights in terms of channel performance/ effects?

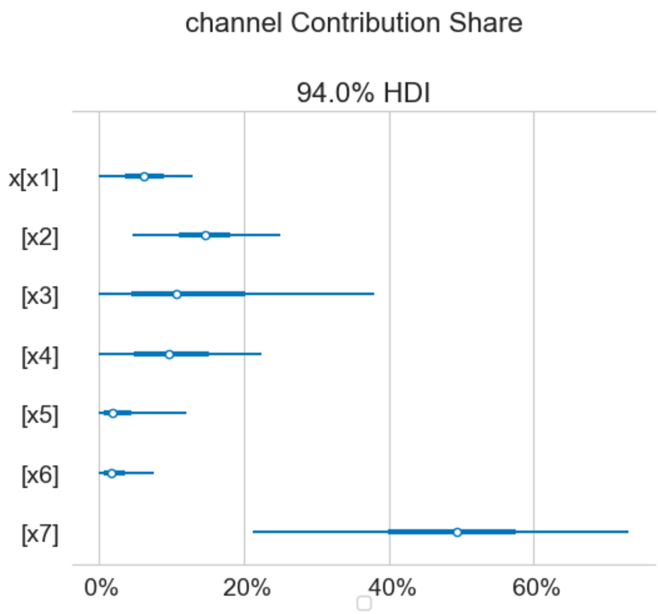
From the posterior probability distributions of $\{\alpha_m, \theta_m, \beta_m, \lambda_m\}$, it seems:

- **Retention:** All the channels have a retention rates between 0.15 - 0.3. Channel 1 has the highest mean retention rate (but highest standard deviation also), while channel 7 has the least.
- **Delay:** All channel **except** channel 7, rule out a delay = 0 at the 96% credible interval, which seem to indicate that we were correct in using the delayed adstock function. Channel 2 has the highest (~9 weeks). Channel 7, on the other hand, seems to have no delay (almost immediate ad exposure) and could have been safely modelled using a geometric adstock.
- **Shape:** Channel 1,2 and 7 are the only channels for which the shape parameters are confidently predicted to peak away from 0, which seems to suggest that at some point media spends in these channels do tend to saturate and then show diminishing returns. From the posteriors, that effect seems to be most prominent for channel 7.

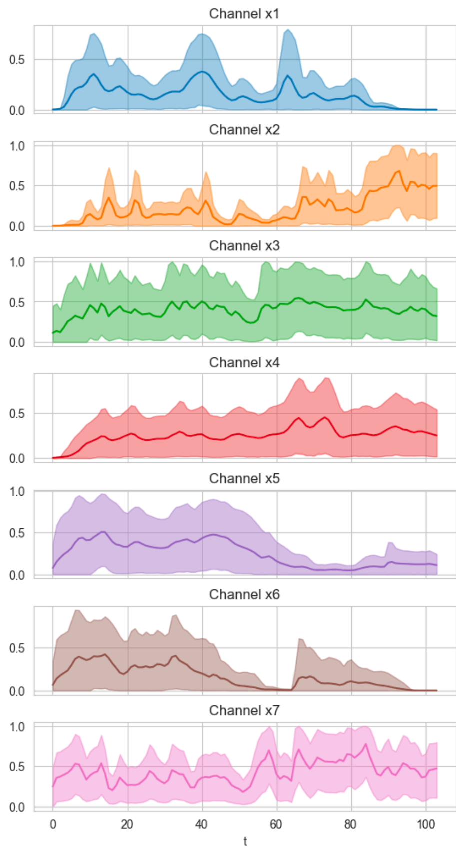
We could calculate the contributions of the different components of the final predicted revenue, i.e., contributions from the base revenue, seasonality/exception events and media spends. It would seem that the majority of the contribution comes from the media spends, even sometimes despite negative effects of seasonality.



Finally, if we isolate the contribution from the media variables, we could predict the relative contribution of each channel to the revenue. Media spend in channel 7 contributes the largest towards the revenue.



We also find that some channels have a more consistent contribution to the revenue than other variables, eg, channels 3,4 and 7. The others contribute in bursts. For these channels, it might have made sense to apply the carryover transformation **after** the shape transformation.



Caption: relative contributions of each channel to the revenue over time

6. (Bonus) Can you derive ROI (return on investment) estimates per channel? What is the best channel in terms of ROI?

Calculating the Return on Ad Spend (ROAS) metric for each channel, we find the change in the revenue per Euro spend on the channel. We approximate the calculation shown in Jin et al 2017, by diving the contribution to the revenue of each channel, divided by the total cost in that channel.

We find that channel 2 has the highest ROAS, followed by channels 1 and 7 respectively.