UNIVERSITÀ DI FIRENZE

Dipartimento di Statistica, Informatica, Applicazioni “Giuseppe Parenti” (DISIA)

IMT School for Advancend Studies - Lucca

2ND LEVEL MASTER COURSE IN DATA SCIENCE AND STATISTICAL LEARNING (MD2SL)

Bayesian Approach for a Media Mix Modelling

Supervisor:

Prof. Fabio Schoen

Candidate:

Giuseppe carbonara

Academic Year 2021/2022

1 Introduction

The present paper illustrates the activity conducted during the project work experience at IQVIA Solutions s.r.l.

Specifically, the work is aimed to explore a Bayesian approach for media mix models which are used by advertisers to assess the effectiveness and capability of the adds' campaigns with the final scope of optimal budget allocation for return's maximization.

This work follows the example of Jin, Yuxue, et al. (2017) [JWS+17] and describes an implementation of a Bayesian method for media mix modeling with carryover and shape effects.

In general, media mix models aim to uncover the causal effect of paid media on a metric of interest, typically sales, and may include other data as prices, media spend in different channels, product distribution and availability, or also external factors as seasonality, macroeconomic patterns or even weather information; these models are usually based on weekly data.

Metrics such as return on advertising spend (ROAS) and optimal advertising budget allocations are derived from these models, starting from the assumption that these models provide valid causal results.

Media mix models have been around in various forms sine the '60s [Bor64]; however, it is not the only approach that can be used to answer questions around advertising effectiveness, other approaches include randomized experiments which represents the gold standard for answering causal questions by randomly splitting the population into a test group, where an action X is performed and a control group, where no action is taken.

The randomization controls for all other sources of variation, so that the only difference between the test and control groups is the action X. However, this approach can not be widely used because may need to run many experiments with many different conditions over time and this would require a huge amount of resources and is not always feasible [LR15], especially in this context.

If the "experimental" approach is not practicable one is therefore reliant on historical data, and a media mix model is a valuable solution.

Data used to fit Media mix models include usually:

- Response data, which are typically sales but can be other KPIs such as clients' acquisition.
- Media metrics in the different media channels, such as impressions, clicks, with media spend being the most common.
- Marketing metrics (control factors) such as price, promotion, product distribution.
- External control factors such as seasonality, weather and market competition.

The approach is usually parameterized with a general regression framework e.g.,

$$y_t = F(X_{t-L+1}, ...X_t, Z_{t-L+1}, ..., Z_t : \Phi)t = 1, ...T \tag{1}$$

where y_t is the sale (or any response variable) at the time t, $F(...)$ is the regression function, $X_t = \{x_{t,m}, m = 1, ..., M\}$ is a vector of ad channel variables at time t for a specific channel m, $Z_t = \{z_{t,c}, c = 1, ..., C\}$ is a vector of control variables at time t and Φ is the vector of parameters in the model. L indicates the length of lag effect that media and control variables have on the response variable.

The typical approach is to have the media channels variables enter in an additive way in the model, while the control variable are parameterized without a lag effect, so:

$$y_t = \sum_{m=1}^{M} \beta_m f_m(x_{t-L+1,m}, ..., x_{t,m}) + \lambda^T z_t + \epsilon_t \tag{2}$$

Where β_m is the channel-specific coefficient, λ is the vector of coefficient for the vector of control variables z_t and ϵ_t is the error term for the variation not explained by the variables.

Estimates from a model such as above are generally more trustworthy when:

- There are enough data needed to estimate all the parameters in the model
- There is useful variability in the advertising levels and control variables
- Model inputs vary independently
- The model accounts for all the important variables that might impact sales
- The model captures the causal relationship between variables

However, The data quantity available for the model is often limited and may consist of few years of national or regional weekly data.

The industry standard typically uses a weekly time period. This is because monthly data granularity is too long and daily level data has too much variation which leads to poor accuracy. Therefore, aggregate data at a weekly level is often the best practice for creating a media mix model.

2 Model description

In this paper is illustrated the implementation of a general model and estimation method that can be applied to any media mix model which uses a **carryover** and **saturation effect** to captures non-linearity in the impact of each single marketing channel on sales.

2.1 Carryover effect (Delayed Adstock)

Advertising carryover essentially states that the positive benefits from advertising, especially increased sales, are not perfectly in step with advertising movements but rather delayed and spread out over time so that changes may not be noticeable immediately or measurable right after the advertising strategy has gone into effect

It rely on the assumption that each new exposure to advertising builds awareness with time, and this awareness will be higher if there have been recent exposures, in a additive manner, and lower if there have not been. In the absence of further exposures the advertisement effect eventually decays to negligible levels.

To model the carryover effect of advertising, time series of each media channel spend need to be transformed through the adstock function:

$$adstock(x_{t-L+1,m}, ..., x_{t,m}; w_m, L) = \frac{\sum_{l=0}^{L-1} w_m(l) x_{t-l,m}}{\sum_{l=0}^{L-1} w_m(l)} \tag{3}$$

Where:

- m is media channel
- w_m non negative weight function
- L is the maximum duration assumed for a ad channel (generally in weeks)

A correct choose of L is critical for estimating the weights in the adstock transformation, if there is not a preliminary knowledge about the right value of L it can be set as large as the weight $w_m(l)$ for $l > L$ is close to 0 beyond such period.

There are different functions to parameterized the weight; in order to account for a possible lagged effect, hence to consider the delay in the peak of such effect, here I used a geometric decay function w_m:

$$w_m(l; \alpha_m, \theta_m) = \alpha_m^{(l-\theta_m)^2} \tag{4}$$

Where:

- $l = [0, ..., L-1]$
- $0 \leq \alpha_m \leq 1$
- $0 \leq \theta_m \leq L-1$, is the delay of the peak effect

Other functional shapes, such as the geometric adstock and negative binomial density function used in Hanssens et al. (2003), [HPS03] , can also be used.

Figure 1 for comparison shows the weight functions of delayed adstock (with a bell shape) and geometric adstock (which does not consider the delayed effect), for the same value of α_m .

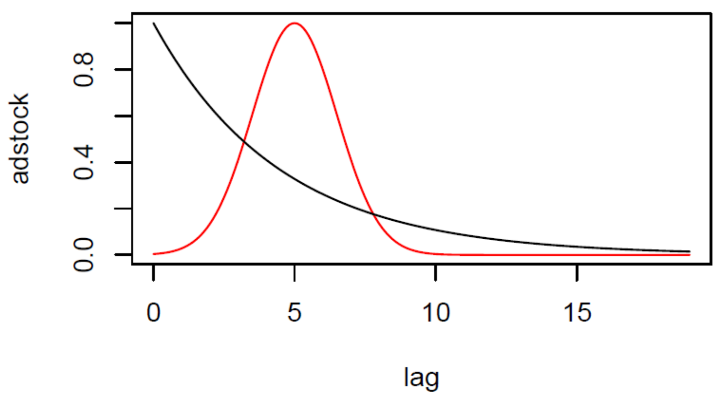

Figure 1: Example weight functions for geometric and delayed adstock. The black line is geometric adstock with $\alpha_m = 0{:}8$; the red line is delayed adstock with the same α_m

2.2 Saturation effect

Advertising saturation means that incremental amount of advertising causes a progressively lesser effect on demand increase. Indeed, even though increasing the amount of advertising would increase the percent of the audience reached, a linear increase in the advertising exposure does not have a similar linear effect on demand. Namely does exist a threshold level after which the effect of advertisement start to decline, and that can be captured by a saturation effect function.

$$\beta(1 - e^{(-\nu x)}) \tag{5}$$

In this case I used a simple exponential function with a β parameter for avoiding to bound the effect in the range [0,1]. However different function can be applied to shape the saturation effect, for example the BetaHill function [JWS+17].

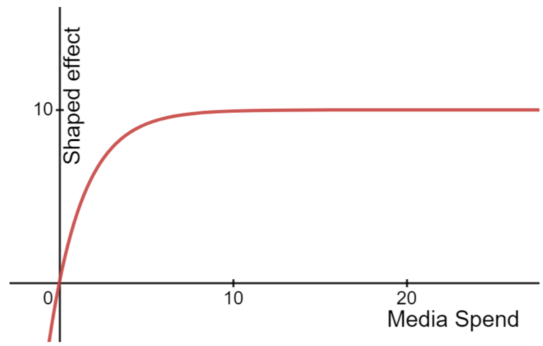

Figure 2: Example of saturation effect in eq. 5 with parameters $\beta = 10$ and $\nu = .5$

2.3 Complete model

The final model is a combination of the two aforementioned functions within the media mix model's framework, by applying the saturation function after the adstock transformation in an additive manner. This assumes that there is no synergy effect between media, which of course in some cases may not be true.

The response y_t (which in this case is sales at the week t) can be modeled by the following equation:

$$y_t = \tau + \sum_{m=1}^{M} \beta Saturation(x^*_{t,m}; \nu) + \sum_{c=1}^{C} \gamma_c z_{t,c} + \epsilon_t \tag{6}$$

Where:

- $x^*_{t,m} = adstock(x_{t-L+1,m}, ..., x_{t,m}; w_m, L)$ as shown in eq. 3
- β and ν are the parameters of the saturation effect in eq. 5
- τ is the baseline sales (intercept)
- γ_c is the effect of control variable z_c
- ϵ_t is white noise uncorrelated with other variables

here it is assumed that there is a linear relationship between the control variables and the response. Of course, can be considered also other forms of adstock and shape transformations.

3 Bayesian approach to Media Mix Model

Traditionally, a frequentist approach such as Ordinary Least Squares is used to fit a media mix model but, using a Bayesian approach (which was first introduced by Google in 2017 [JWS+17]), gives some benefits such as:

- **Incorporate prior knowledge:** they can come from various sources such as industry experience, previous media mix models of the same or similar advertisers or simply intuitions of marketing managers
- **Combine that prior knowledge with data:** a Bayesian approach enables to combine prior valuable knowledge with modeling insights from the MMM's data-driven strategy, augmenting such human knowledge.
- **Estimate the level of confidence:** a classic benefit of the Bayesian approach is the awareness about the level of confidence about parameters estimation. Hence, Parameters associated with channels that are estimated with high level of uncertainty may trigger marketing efforts to resolve this uncertainty (i.e. new tests or further data as control variables).

The frequentist approach finds the most likely value of the parameters by maximizing the likelihood (MLE), as in:

$$\hat{\Phi} = arg_{\Phi} max \mathcal{L}(y|X, Z, \Phi) \tag{7}$$

Where Φ represent the vector of parameters of the model in 6, X denote the media spend for all channel, Z all the control variables, and y the vector value of the response variable. Hence, $\mathcal{L}(y|X, Z, \Phi)$ is the log likelihood given the data and the parameters.

While the Bayesian approach, on the other hand, treats the model's parameters as random variables and is based on the posterior distribution of the parameters given the data and the prior distribution $\pi(\Phi)$:

$$p(\Phi|y, X) \propto \mathcal{L}(y|X, Z, \Phi)\pi(\Phi) \tag{8}$$

Random samples from the posterior distribution is used to make inference of the parameters and, as summary result, here can be used the mean or median of the posterior distribution and the credible interval of the parameters derived from Φ.

In this paper is used and described the Bayesian approach to estimate the model, such framework, as mentioned, allows to incorporate available knowledge into the model as prior distribution of the parameters. Such knowledge are important because the information available within a single media mix model data set is usually low compared to the number of parameters to be estimated.

Even though the Bayesian framework guarantees a number of advantages, it also implies some challenges which may arise when implementing it. firstly, the choose of appropriate likelihood function among several options which however normally is a normal distribution but can even change. Secondly, the priors to use over number of parameters, such as regression coefficients on the control measures and the parameters in the saturation and adstock functions. Some choices may be not appropriate and so may lead to bad convergence problems. Also the parameterization of the adstock and saturation functions may be challenging and can not be left to chance.

Another issue is the computational time for big data sets, indeed the sampling mechanism of the Bayesian algorithm may involve a longer computational time.

3.1 PyMC3 library

The sampling method used here is based on PyMC3 library (see [SWF16]), which is an open source Probabilistic Programming framework written in Python that uses Theano to compute gradients via automatic differentiation and allows model specification directly in Python code while making use of sampling algorithms such as the No-U-Turn Sampler (NUTS; Hoffman, 2014), a self-tuning variant of

Hamiltonian Monte Carlo (HMC; Duane, 1987). The syntax is quite close to the statistical notation, as concrete example see the script 2 for fitting the model with the relative priors and the likelihood expression an in 6.

Models in PyMC3 are centered around the Model class. It has references to all random variables (RVs) and computes the model logp and its gradients instantiated as part of a with. While every unobserved RV has the following calling signature: name (str), parameter keyword arguments.

Thus, a normal prior can be defined in a model context as "mu" in example 1. Observed RVs are defined just like unobserved RVs but require data to be passed into the observed keyword argument as shown in 1 for "obs".

Listing 1: PyMC script example

```
with pm.Model() as model:
    mu = pm.Normal("mu", mu=0, sigma=1)
    obs = pm.Normal("obs", mu=mu, sigma=1, observed=np.random.randn(100))

    idata = pm.sample(2000, tune=1500, return_inferencedata=True)
```

The main entry point to MCMC sampling algorithms is via the pm.sample() function. By default, this function tries to auto-assign the right samplers and auto-initialize if anything is passed.

By default the function apply a No U-Turn Sampler (NUTS) which is a Markov chain Monte Carlo method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence then can be used to approximate the distribution, or to compute an integral (such as an expected value).

4 Attribution metrics and budget optimization

Besides estimates of the parameters in the model, a modeler should be also interested in attribution metrics derived from the model, in particular the **Return on Ad Spend (ROAS)**, and to find the **optimal media mix** that maximizes the revenue under a budget constraint in the selected time period, which in turn is guided by ROAS.

4.1 ROAS

ROAS is the change in sales (or revenue) per unit spent on a single channel; it is calculated by setting spend of the channel to zero in the selected time period and comparing the predicted revenue against that of the current media spend. While, the predicted sales is the sum of the first three terms in equation 6 , excluding the noise term.

Of course, in this case we assume that the spend of one media channel can be changed without affecting the spend of any other channel.

Let $\hat{Y}_t$ denote the predicted sales, which depends on media X, control variables Z and parameters Φ. Suppose the historical spend of the mth channel is denoted by $x_{t,m}$ while $\tilde{x}_{t,m}$ is the spend of the same channel set to 0 for the selected time period (called change period), The predicted sales using all historical spend is denoted with $\hat{Y}_t^m(x_{t-L+1,m}, ..., x_{t,m}; \Phi)$, while the predicted sales using all the spend but the mth channel to evaluate (which is set to 0) is denoted by $\hat{Y}_t^m(\tilde{x}_{t-L+1,m}, ..., \tilde{x}_{t,m}; \Phi)$ ROAS on the media mth si calculated as:

$$ROAS_m = \frac{\sum_{t_0 \leq t \leq t_1+L-1} \hat{Y}_t^m(x_{t-L+1,m}, ..., x_{t,m}; \Phi) - \hat{Y}_t^m(\tilde{x}_{t-L+1,m}, ..., \tilde{x}_{t,m}; \Phi)}{\sum_{t_0 \leq t \leq t_1} x_{t,m}} \tag{9}$$

Parameter	Media 1	Media 2	Media 3
α	.6	.8	.8
θ	5	3	4
β	1.2	2	1.5
ν	1	2	2.8

Table 1: Generating parameters: Media specific parameters

Parameter	Media 1
L	13
τ	4
γ	-.5
ϵ	.3

Table 2: Generating parameters: Other variables

where $(t_0; t_1)$ is the change period. Since has been using a Bayesian framework, the posterior samples of Φ are plugged into 9 to obtain posterior samples of ROAS. As summary result can be used either the mean, or the median, and the credible interval of the posterior distributions of ROAS.

4.2 Budget Optimization

Suppose the total budget of all media in the change period $(t0; t1)$ is C, the optimal media mix for this period $X^o = \{x^o_{t,m}, t_0 \leq t \leq t_1, 1 \leq m \leq M\}$ is obtained by

$$maximise \sum_{t_0 \leq t \leq t_1 + L - 1} \hat{Y}(x_{t-L+1,m}, ..., x_{t,m}; 1 \leq m \leq M; \Phi) \tag{10}$$

$$subjet\ to \sum_{t_0 \leq t \leq t_1} \sum_{1 \leq m \leq M} x_{t,m} = C \tag{11}$$

Each posterior sample of Φ should be plugged into the objective function in 10, and get the optimal mix for the jth sample X^o_j. Again here the mean/median and the credible interval of the posterior distribution of the optimal mix can be used to summarize it.

5 Application to a Simulated Data Set

Here is described how has been estimated the media mix model in 6 on a synthetic data set by using the saturation (see 2.2) and carryover functions (see 2.1)

5.1 Data set description

The data set used to test the model has been synthetically generated, media variables were generated by adding white noise to a sinusoidal seasonality with 200 weeks as period and has been scaled in range [0-1], while prices (the unique control variable) have been generated as an ARIMA time series and the sales variable was generated using the saturation and carryover functions with fixed parameters. Table 1 and 2 contain the parameters used for the model in eq. 6 to generate sales as response variable.

The data set is formed by the weekly spend for three media channels for 200 weeks length, plus the price variable generated though ARIMA time series and the simulated sales. Figure 3 shows the charts of the data: subplot 'a' shows the raw media spend for the 200 weeks, while the subplot 'b' represent the media's spend already processed though the saturation and carryover functions using the fixed parameters reported in table 1.

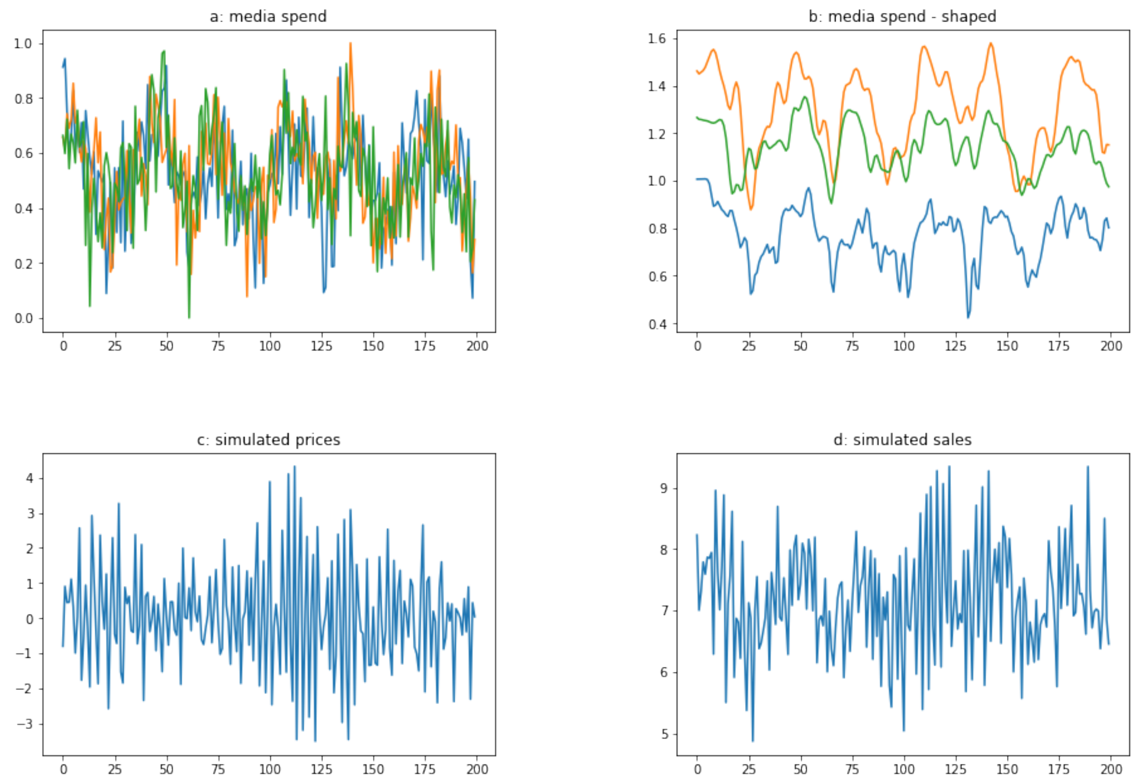

Figure 3: Sales, raw media spend, shaped media spend and simulated price in the data set.

As can been seen, the time series of the media spend already processes through the mentioned functions is more smooths and has less noise respect to the raw data.

5.2 Priors used and model fitting

Parameter	**Prior**	**Parameter**	**Prior**
α	Beta(2,2)	τ	Normal(0,5)
θ	Uniform(0,12)	λ	Normal(0,5)
β	Gamma(3,1)	ϵ	InverseGamma(3,.5)
ν	Gamma(3,2)		

Table 3: Priors on parameters.

The choose of the priors has a large influence on the posterior distributions especially in the cases where the size of data set is not that big. For this model alternatives priors have been tested to check the difference in the posterior distributions. Table 3 shows the priors used for the final model.

The value of β must be not negative, so a Gamma(2,2) distribution was chosen because we know that the value for the simulated data set is around 1.6 for all the media, similarly the prior of ν is a Gamma(1.5,2); however different alternatives of positive distribution (like as HalfNormal) may be tested for other data set.

The retention rate α is constrained on $[0; 1)$ and should have a prior that is defined on $[0; 1)$, such as a beta or uniform distribution, so was chosen a Beta(2,2). Similarly, the delay parameter θ should have a prior that is constrained on $[0; L-1]$ with $L = 13$, such as a uniform or scaled beta distribution, therefore a Uniform(0,12) was chosen.

Ad explained in section 3.1 the implementation of the model is quite similar to the statistical notation for the definition of the priors and the likelihood expression. As shown in script 2 for each parameter (i.e. alpha, theta or beta) is specified a prior distribution with relative internal parameters and the shape (number of variable interested by that prior), for example for the delay weight θ of the adstock function (carryover) is defined a Gamma(3,1) distribution and shape 3 (which corresponds to the number of media channels considered in the model).

The media spend series for each channel is shaped firstly through the carryover and then through the saturation function using the parameters sampled, afterward the sales are predicted ("y_hat" in script 2) using a normal distribution though the likelihood function as defined in eq. 6.

Listing 2: Fitting model script in PyMC3

```
import pymc3 as pm

with pm.Model() as m:

  #var,      dist, pm.name,             params,  shape

  alpha = pm.Beta('alpha',              2 , 2, shape=3  ) # retain rate in adstock (0 < alpha < 1)
  theta = pm.Uniform('theta',           0 , 12, shape=3 ) # delay weight in adstock(0 < theta < L-1)
  beta  = pm.Gamma('beta',              3 , 1,  shape=3 ) # features coefficient   (> 0)
  nu    = pm.Gamma('nu',                3 , 2,  shape=3 ) # parameter for  saturation(>0)
  tau   = pm.Normal('intercept',        0 , 5           ) # model intercept
  lamb  = pm.Normal('lamb'            , 0 , 5           ) # price coefficient
  noise = pm.InverseGamma('noise',      3 , .5          ) # variance about y


   M1 = [saturation(x, nu[0], beta[0]) for x in carryover(X_train[:,0],
                                             alpha[0], L, theta = theta[0])]
   M2 = [saturation(x, nu[1], beta[1]) for x in carryover(X_train[:,1],
                                             alpha[1], L, theta = theta[1])]
   M3 = [saturation(x, nu[2], beta[2]) for x in carryover(X_train[:,2],
                                             alpha[2], L, theta = theta[2])]

   y_hat = pm.Normal('y_hat',   mu= tau + M1+M2+M3 + (lamb * price_train),
                                           sigma=noise,    observed=y_train)

   trace1 = pm.sample()
```

5.3 Result on ROAS

For calculating the Return On Advertising Spend (ROAS) has been used the entire period of the data set, so I assumed the change period as 176 weeks, and the pre- and post-change periods are 12 weeks each, since L is assumed to be 13.

The most recent 12 weeks are the post-change period, and one year prior to it is the change-period. This is illustrated as example in Figures 4 and 5 respectively for media 1 and 2 using the described data set. The predicted sales are calculated using a randomly selected posterior sample of Φ.

In the figures the green solid line is the scaled historical spend on the medium, and the blue solid lines is the historical spend, while dashed green line is the predicted sales generated by setting the spend of the considered media channel in the change period to zero.

The total difference between the two time series of predicted sales (the blue solid and dashed green lines) is the change in revenue attributable to the spend for all media but the one analyzed.

Again here as summary result we can take the mean value and the standard deviation of the ROASs

Metrics	Media 1	Media 2	Media 3
Median ROAS	1.3631	2.0652	1.1933
standard deviation	.772	.8032	.8381

values for the uncertainty level.

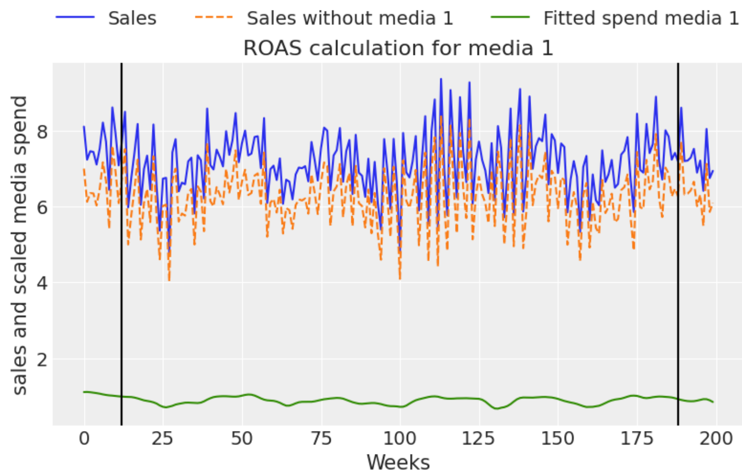

Figure 4: Illustration of ROAS calculation for one media channel 1.

Media 1 shows a ROAS value of 1.3631 with a standard deviation of 0.7727 this means that each unit spend in the first marketing media should give an average return of 1.3631 unit (in this case sales), however the relatively high standard deviation of the mean ROAS indicates also an high level of uncertainty.

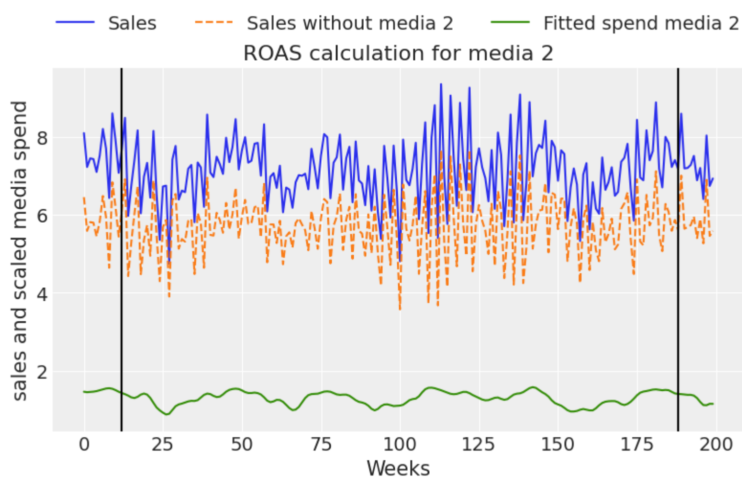

Figure 5: Illustration of ROAS calculation for one media channel 2.

Media 2 has the highest ROAS value of 2.0652 with a standard deviation of 0.8032; hence, for each unit spend in the second marketing media we have a calculated mean return of 1.8528 unit (of sales)

On the other hand, media 3 has the lowest ROAS value of 1.1933 but with a relatively high standard deviation of 0.8381; hence, for each unit spend in the second marketing media we should have an average return of 1.1933 unit (of sales).

5.4 Result on optimization

The budget optimization has been performed trough SciPy library. SciPy optimize provides functions for minimizing (or maximizing) objective functions, possibly subject to constraints. It includes solvers for nonlinear problems (with support for both local and global optimization algorithms), linear programming, constrained and nonlinear least-squares, root finding, and curve fitting.

Since we are dealing with a non-linear constrained optimization problem, with a non convex objective function, a constrain on non-negativity of the features (the expense can not be negative) and another constrain of equality for the maximum budget set (see 11).

The method chosen for the optimization has been the Sequential Least SQuares Programming optimizer (SLSQP), it is ideal for minimizing a function of several variables with any combination of bounds, equality and inequality constraints. The method wraps the SLSQP Optimization subroutine originally implemented by Dieter Kraft [Kra94].

The optimization has been performed for an interest period of 30 weeks (this is a critical choice but depends from marketing strategies) with a maximum budget of 50.000 unit (can be any currency) but scaled with the same scale used to initially scale the X features, so the spend in each marketing channel.

The features have been initialized randomly because, given the non-convexity of the objective function, the result may highly depend from the starting point of the algorithm.

Similarly to the calculation of ROAS metrics the predicted sales are calculated using a randomly selected posterior sample of Φ and, as summary result, the mean maximum value of the result (maximum sales) of the optimization about the spend for all 30 weeks.

With a budget of 50.000, optimal spend for the 30 Weeks period on Media 1 appear to be a total of 15577.63 (unit of currency) with a standard deviation of 12834.7, similarly the optimal spend for media 2 is the highest with a value of 21407.56 and Standard deviation of 12239.12; while media 3 had the lowest optimal spend of 14270.59 with a standard deviation of 14543.08. This is in line with the calculated ROAS which gave the media 2 as the most performing media followed by media 1 and media 3 as the less performing. However, here we have high values of standard deviation which indicates high uncertainly.

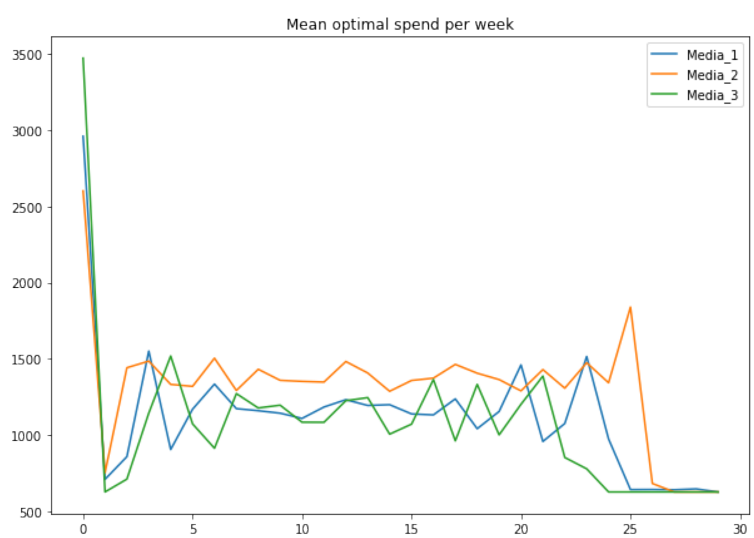

Figure 6: Mean optimal spend per week

Metrics	Media 1	Media 2	Media 3
Optimal Spend	15577.636	21407.565	14270.593
standard deviation	12834.746	12239.122	14543.08

5.5 Model Evaluation

To understand how well the model performs at recovering the true model parameters we can see the posterior distribution of model parameters and its mean value (in dashed red) versus the true model parameter (black lines) in figure 7.

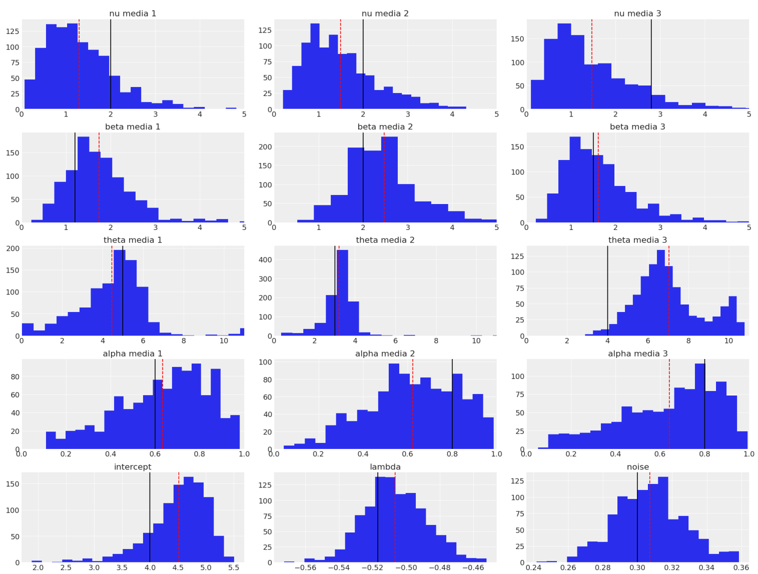

Figure 7: Density of posterior distributions of all parameters

Between the saturation parameters ν and β, β seems to be estimated better by the model and to have a slightly lower uncertainty respect to ν; in the specific, media 3 has a lower precision in the estimation of ν respect to the other media, while media 2 seems to have the best estimation both in terms of lower bias and uncertainty respect to ν.

The estimation of delay parameter θ has quite low bias for all media except the third, but an higher uncertainty in media 3, followed by media 1.

The estimation of α exhibits somewhat more bias in the second and third media with a similar uncertainty level for all media.

For what concern the other parameters, the intercept looks a bit more biased (considering the scale of the x axis in chart 7) while the and ϵ look quite similar and show lower bias and uncertainty

In order to evaluate the model performance, beside the mean and credibility interval of the parameters' posterior distributions around the real values, the data set has been also divided into training (80%) and test (20%) set, and has been compared the predicted sales calculated with the sampled posterior distribution of the parameters, with the real sales of the test set.

As can be seen by figure 8 which shows in black 500 time series of predicted sales generated with as many as set sampled parameters, and in red the actual sales for each week of the data set, the in-sample fit (RMSE/MAE) is pretty good: 0.369 the RMSE and 0.136 the MSE.

This was expected because the data were generated with the same functional form which was modelled. So, it basically shows how well the PyMC algorithm works, however it can be used as tools for a possible model selection ans so ti compare different models.

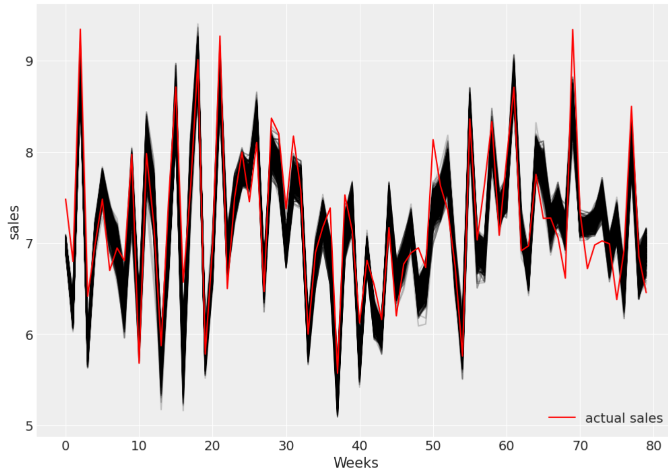

Figure 8: Fitted versus Actual sales

6 Conclusion

In this work has been shown one of the possible model specification for media mix models on a simulated data set using a Bayesian approach, as performed during the curricular intern experience in IQVIA Solutions s.r.l..

The media mix model described in this paper applies a media carryover and saturation effects and is shown how the model parameters can be estimated in a Bayesian framework. Furthermore, the attribution metric ROAS and the optimal budget allocation mix has also been calculated using the posterior samples of the parameters.

For future work the model and algorithm can be tested on real data sets with both media spend and multiple control variables, different alternative parametrization and priors can be experimented to select the setting more widely capable and performing.

References

[Bor64] Neil H Borden. The concept of the marketing mix. *Journal of advertising research*, 4(2):2–7, 1964.

[HPS03] Dominique M Hanssens, Leonard J Parsons, and Randall L Schultz. *Market response models: Econometric and time series analysis*, volume 2. Springer Science & Business Media, 2003.

[JWS+17] Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, and Jim Koehler. Bayesian methods for media mix modeling with carryover and shape effects. 2017.

[Kra94] Dieter Kraft. Algorithm 733: Tomp–fortran modules for optimal control calculations. *ACM Transactions on Mathematical Software (TOMS)*, 20(3):262–281, 1994.

[LR15] Randall A Lewis and Justin M Rao. The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, 130(4):1941–1973, 2015.

[SWF16] John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016.