Regression, Bayesian and Machine Learning Methods for Media Mix Modeling - A Comparative Study

Author: Lukas Scharfe Supervisor: Prof. Dr. Antje Jahn Co-Supervisor: Prof. Dr. Arnim Malcherek

Hochschule Darmstadt, University of Applied Sciences, Fachbereich Mathematik und Naturwissenschaften & Informatik

h_da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES

Motivation and Research Objectives

"Half the money I spend on advertising is wasted; the trouble is I don't know which half" - John Wanamaker

Advertisers face a fundamental challenge: the efficient allocation of their advertising budget. Media Mix Models are used in marketing to **measure the effectiveness of advertising** and support marketers in making future budget allocation decisions. Key results in Media Mix Modeling (MMM) are metrics for marketing attribution such as Return on Advertising Spend (ROAS). These attribution metrics measure how advertising investments contribute to business-relevant Key Performance Indicators (KPIs) such as sales or brand awareness, and provide information for re-allocating media budgets. Nonetheless, modeling an advertising response is a challenge since advertising has lagged effects and diminishing returns [3, 1].

The aim of this thesis is a **comparative analysis of parametric, semi- and non-parametric regression methods**, which take the unique characteristics of advertising into account, and their evaluation regarding their applicability in Media Mix Modeling. The assessment of the examined methods is based on their prediction accuracy and their precision in determining attribution metrics. The regression methods are applied on both simulated data and a real-world data set.

Patterns of an Advertising Response

Carryover Effects

Advertising can have lagged effects and carry over through time. For example, one sees ads today and still remembers the next week. This can be modeled by applying a **geometric Adstock transformation** to the media spend [2]:

$$\text{Adstock}(x_{t-(L-1),m}, \dots, x_{t,m}; \alpha_m, L) = \frac{\sum_{l=0}^{L-1} \alpha_m^l \cdot x_{t-l,m}}{\sum_{l=0}^{L-1} \alpha_m^l}, \quad 0 < \alpha_m < 1,$$

where $x_{t,m}$ is the media spend of the m-th media channel, α_m the retention rate of the ad effect from one time unit to the next and L specifies the duration of the carryover effect.

Shape Effects

Additionally, advertising investments can have diminishing returns. A flexible parameterization for modeling concave and S-shaped response curves is given by the **Hill function**

$$\text{Hill}(x_{t,m}; K_m, S_m) = \frac{1}{1 + (x_{t,m}/K_m)^{-S_m}}, \quad x_{t,m} \geq 0,$$

where $S_m > 0$ is the slope and $K_m > 0$ the half saturation point of the m-th media channel [2]. Fixing $S_m = 1$ yield concave Hill functions.

A Parametric Media Mix Model

Combining carryover and shape effects, an advertising response y_t can be modeled by

$$y_t = \beta_0 + \sum_{m=1}^{M} \beta_m \cdot \text{Hill} \circ \text{Adstock}(x_{t-L+1,m}, \dots, x_{t,m}; \Phi) + \sum_{c=1}^{C} \gamma_c z_{t,c} + \epsilon_t. \quad (1)$$

The media spend $x_{t,m}$ of the m-th media channel acts via an Adstock and Hill transformation on the response variable, while the c-th control variable $z_{t,c}$ contributes linearly. β_m denotes the media-specific regression coefficient and γ_c the regression coefficient of the c-th control variable. Φ denotes the vector of transformation parameters (e.g. retention rates α_m) in the model.

Data

Simulated Data

The role of the simulated data is to evaluate the examined methods against artificial generated data where the *true underlying data structure is known*. This enables the comparison of estimated attribution metrics and model parameters with their respective true values.

Data generation of a single simulated data set:

- One data set contains a response variable y_t, 3 media spend variables $x_{t,m}$, $m \in \{1,2,3\}$, and 1 control variable z_t.
- The response variable is generated based on an additive media mix model, which is from the same model class as in Eq. (1). A geometric Adstock transformation with $L = 13$ and a concave Hill transformation ($S_m = 1, \forall m$) is used.
- The media spend variables $x_{t,m}$ are generated using a MMM data generator (R package dammmdatagen) and the control variable is generated as an $(1,0)$-ARMA time series.
- The media-specific parameters in the model (regression coefficient β_m, retention rate α_m and half saturation point K_m) are different for each media. For example, the parameters used in the model for Media 2 are $\beta_2 = 1.5, \alpha_2 = 0.3, K_2 = 0.4$.

In total, 500 simulated data sets are generated following the same model, but with a different random seed for generating the media spend variables and control variable. The estimation procedures of the examined methods are repeated across these 500 simulated data sets to eliminate randomness in a single simulated data set. Statements concerning the uncertainty and potential biases in the attribution metrics thus become more reliable. To study the impact of sample size on the method's outcomes, 500 data sets with normal sample size (4-year weekly history) and 500 data sets with large sample size (50-year weekly history) are generated.

Data (cont.)

Real-world Data

The role of the real-world data is to assess the methods on data with real-world complexities. The data set contains 3 years of weekly data. Ad Awareness, quantifying whether consumers can recall a brand's advertisement, serves as the response variable. Weekly media budgets of 3 media channels (TV, Print and Online Video) and 2 control variables, quantifying the COVID-19 impact and summer vacation impact on the response, are used for modeling.

Methods

Parametric regression methods

- **OLS regression with time-series cross-validation ("OLSReg")**: This method estimates a linear model using the transformed media spend variables and the control variables as predictors. It employs time-series cross-validation (TSCV) to estimate the media transformation parameters α_m and K_m.
- **Constrained regularized regression with time-series cross-validation ("ConstrRegulReg")**: This method works similar as the OLSReg method, but additionally applies $L2$-regularization and constraints the media regression coefficients β_m in the model to be positive.
- **Bayesian regression ("BayesianReg")**: The Bayesian approach is employed to estimate the model in Eq. (1) using Markov Chain Monte Carlo (MCMC) algorithms. This approach enables the integration of prior knowledge via the choice of informative prior distributions.

Semi- and non-parametric regression methods

- **Generalized Additive Models with shape constraints ("MonGAM")**: This method utilizes the adstocked media variables and control variables as predictors, estimating the non-linear relationship (shape effect) between the adstocked media spend and the response via monotonically increasing regression splines. TSCV is employed to estimate the retention rate parameters α_m.
- **XGBoost**: Similar to MonGAM, this method uses the adstocked media variables and control variables as predictors. TSCV is employed to optimize XGBoost's hyperparameters and the retention rates α_m.

Attribution metrics

The Contribution Share (CS) and Return on Ad Spend (ROAS) are attribution metrics that can be computed for each medium. CS measures the effect share of a medium on the response variable (e.g. sales) and ROAS quantifies the change in the response variable per Euro spent on the medium. In the additive model in Eq. (1), the ROAS of the m-th media channel can be computed by comparing the contribution of the m-th media channel to the response variable against the media spend $x_{t,m}$. This approach does not work for non-additive models (XGBoost). In this case, the concept of partial dependence is used to compute attribution metrics.

Results

Results across 500 simulated data sets

Bayesian regression achieves the best prediction accuracy (see Table 1) and shows the lowest uncertainty in ROAS estimates (see Figure 1). "BayesianRegInf", which employs informative priors for the media-specific model parameters (β, α and K), demonstrates enhanced prediction accuracy on normal-sized data sets compared to "BayesianRegNonInf", which utilizes non-informative priors. This shows that informative priors have an impact on prediction accuracy and model outcomes, especially on small data sets which are typically in MMM. The influence of informative priors diminishes when the sample size is large. On normal-sized data sets, informative priors additionally mitigate uncertainty in attribution metrics. "OLSReg" shows high uncertainty in attribution metrics and "ConstrRegulReg" additionally exhibits underestimated ROAS estimates. "MonGAM" shows low levels of uncertainty, but biased ROAS estimates. XGBoost exhibits poor prediction performance on normal-sized data sets and high variability in determining ROAS.

Method	Sample Size	
	Normal	Large
OLSReg	2.612	2.512
ConstrRegulReg	2.665	2.608
BayesianRegInf	**2.547**	2.504
BayesianRegNonInf	2.557	**2.503**
MonGAM	2.655	2.513
XGBoost	3.014	2.583

Table 1. Average prediction RMSE ($\cdot 10^{-1}$) across 500 simulated data sets with normal and large sample size.

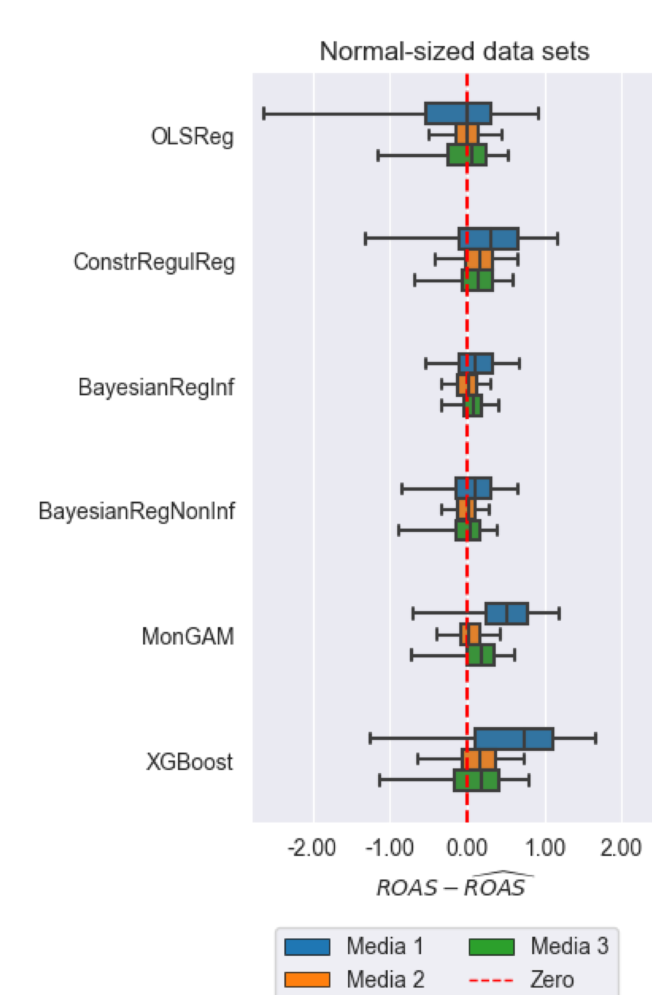

Figure 1. Distribution of the deviation between the true and estimated ROAS across 500 normal-sized data sets.

Results (cont.)

These findings are reflected in the estimated response curves for Media 1 in Figure 2. For all methods that apply regularization techniques ("ConstrRegulReg", "MonGAM" and XGBoost), the majority of response curves lie below the true curve. The media impact is consequently underestimated, which results in a biased ROAS. The estimated response curves of "BayesianRegInf" most closely resemble the true response curve.

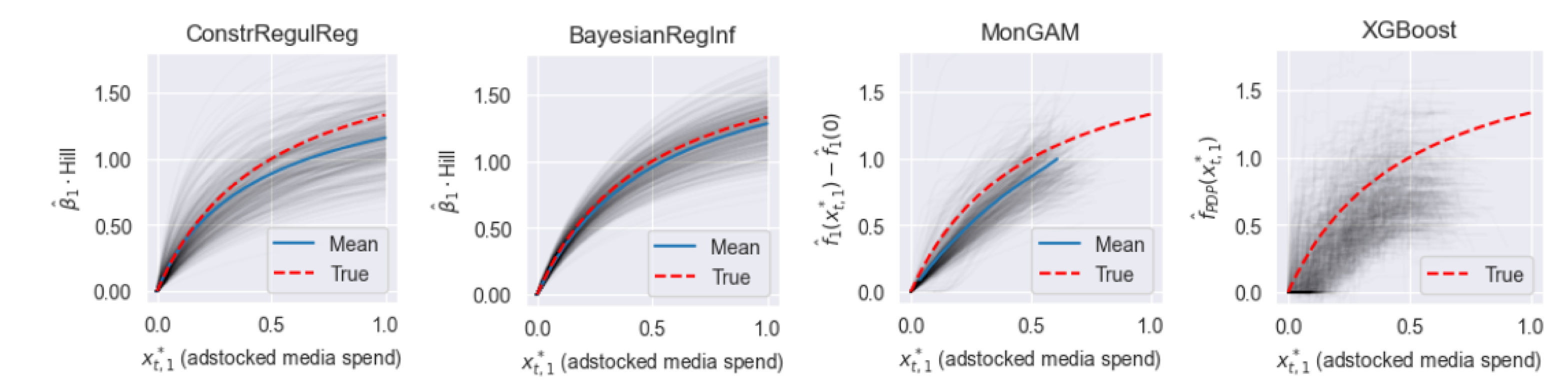

Figure 2. Response curves for Media 1 estimated by ConstrRegulReg, BayesianRegInf, MonGAM and XGBoost. Each grey line depicts a response curve of one simulated data set.

Results of the real-world data set

On the real-world data set, the parametric methods yield better prediction accuracy compared to MonGAM and XGBoost (see Table 2). The poor prediction accuracy of the semi- and non-parametric methods can be attributed to their limited extrapolation capabilities: given that the test set encompassed spending levels absent in the training set, the methods were forced to extrapolate beyond the training set. Thus, the more complex response curves estimated by MonGAM (e.g. a S-shaped response curve for Online Video) and the interactions revealed by XGBoost (e.g. a positive interactive influence between Print and Online Video) can neither be confirmed nor rejected.

The "ConstrRegulReg" method achieves the best prediction accuracy. The results from "ConstrRegulReg" show that TV has the highest impact on Ad Awareness among the media variables. Print follows with the second highest impact on Ad Awareness.

Bayesian regression best mitigates the uncertainty of the attribution metrics, but also stands out for differing results compared to other methods (e.g. the Contribution Share (CS) for Print in Figure 3). This difference primarily arises from the informative prior distributions chosen for the model parameters. In-depth analysis revealed that the data has not enough information content to alter the prior belief of the half saturation parameter K. Thus, the results rely on the prior knowledge incorporated into the model.

Method	RMSE ($\cdot 10^{-1}$)
OLSReg	7.543
ConstrRegulReg	**7.452**
BayesianReg	7.648
MonGam	9.233
XGBoost	8.151

Table 2. Prediction RMSE of the real-world data set.

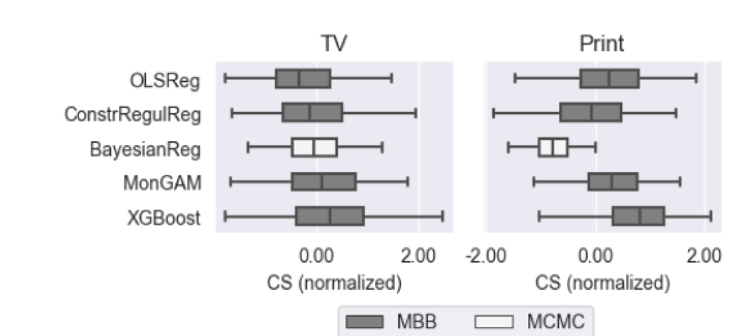

Figure 3. Exemplary uncertainty measurements of the Contribution Share (CS) for TV and Print. The CS values are normalized for data protection reasons.

Conclusion & Outlook

Bayesian regression methods emerged in this comparative study as the most promising approach in MMM for reducing uncertainty in the model, achieved through the integration of informative priors. On small data sets with less information content, the priors exert a significant influence on the model outcomes. It is therefore important to collect powerful MMM data sets and choose the prior distributions carefully.

Applying regularization appeared as a possibility to enhance prediction accuracy on real data, but in turn it can lead to biased attribution metrics due to an underestimated media effect.

Semi- and non-parametric methods emerged as an opportunity in MMM (e.g. for estimating more complex response curves). However, the media effect on the response can also be underestimated due to regularization. The suggested derivation of attribution metrics from XGBoost using partial dependence curves showed high uncertainty, which might be linked to overfitting issues and the instability of tree algorithms on small data sets. It needs further research if a more regularized ML model yields more precise results.

The extrapolation behavior of the examined semi- and non-parametric methods is a problem in MMM, suggesting the need for caution in making predictions. The potential of these methods warrants further exploration on real-world data, particularly in scenarios where extrapolation is not required.

References

[1] David X. Chan and Mike Perry. *Challenges and Opportunities in Media Mix Modeling*. 2017.

[2] Yuxue Jin et al. *Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects*. 2017.

[3] Gerard J. Tellis. "Modeling Marketing Mix". In: *The handbook of marketing research: uses, misuses, and future advances* (2006), pp. 506–522.