Bayesian Modeling for Measuring Effectiveness of Digital Marketing

YU ZHU and YANG SU

In today's digital marketing era, assessing the effectiveness of advertising campaigns across multiple media channels is crucial for optimizing strategic decisions and budgets. Traditional econometric models, while robust for immediate impact analysis, often fail to address the complex dynamics and non-linear effects prevalent in digital advertising. This study advances the field by integrating Bayesian statistical methods, which accommodate the inherent uncertainties of digital media data and allow for the incorporation of prior knowledge. By employing a Bayesian approach, our research analyzes both immediate and delayed impacts of advertising on application installations, using sophisticated weight functions like geometric decay and delayed impact to model temporal dynamics effectively. The study highlights the challenges of traditional models in capturing the extended influence of advertisements and proposes a refined method that enhances predictability and accuracy. Our findings provide actionable insights that can significantly enhance media planning and budgeting in digital marketing.

Additional Key Words and Phrases: Bayesian modeling, Digital marketing, Advertising effectiveness measurement

ACM Reference Format:

Yu Zhu and Yang Su. 2024. Bayesian Modeling for Measuring Effectiveness of Digital Marketing . In . ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the dynamic and digitally-driven marketing landscape, understanding the effectiveness of advertising campaigns has become a cornerstone for strategic decision-making and budget optimization. The vast proliferation of digital media channels has introduced complex dynamics into consumer behavior, necessitating sophisticated methodologies to measure and interpret the effects of advertising. Traditional models, while foundational, often fall short in addressing the intricacies of digital advertising, such as the interplay between different media types and the delayed effects of campaigns.

Historically, advertising effectiveness research has been dominated by econometric models, which provide a robust framework for assessing immediate impacts. However, as digital marketing evolves, there is a growing recognition of the limitations of these models in capturing the longitudinal and non-linear effects of advertising expenditures. Recent advancements in statistical modeling, particularly the integration of machine learning and Bayesian statistics, have begun to address these challenges, offering deeper insights and more accurate predictions.

The literature on advertising effectiveness highlights several critical issues, such as the attribution of sales or conversions to specific advertising activities and the estimation of carryover effects, where the influence of an ad extends beyond its initial exposure period. These issues are particularly salient in digital advertising, where consumer interactions with media are more frequent and varied than in traditional media environments. Addressing these complexities requires a nuanced approach that can adapt to the rapid changes in media consumption and the subtleties of consumer response patterns.

Our study contributes to this evolving field by employing a Bayesian modeling approach to rigorously analyze advertising effectiveness. This approach not only accommodates the inherent uncertainties and variabilities associated with media data but also allows for the incorporation of prior knowledge and expert judgment. Bayesian methods are particularly well-suited for modeling complex systems where traditional approaches may struggle with overfitting or underfitting, especially in scenarios with limited data but high variability.

By focusing on a real-world campaign across multiple media channels, our research seeks to uncover both the immediate and delayed impacts of advertising on consumer behavior—specifically, application installations. We utilize Bayesian inference to model these effects, employing sophisticated weight functions such as geometric decay and delayed impact to capture the temporal dynamics of advertising influence. This methodological choice enables us to provide a more comprehensive understanding of how advertising efforts translate into consumer actions over time.

Furthermore, the application of Bayesian techniques allows for robust sampling and estimation procedures, facilitating detailed posterior analyses that enhance the interpretability of the results. Through this study, we aim to deliver actionable insights that can guide marketers in optimizing their advertising strategies, ensuring that each campaign is not only effective but also economically efficient in a highly competitive digital marketplace.

The rest of the paper is organized as follows. In Section 2, we introduce the advertising process in the game marketing, and discuss the goals of the advertising modeling with specific challenges. In Section 3, we propose the Bayesian media advertising models including the carryover and shape effect in detail. In Section 4, we conduct empirical studies under the real-world ads campaign data and compare the performance of different priors and weight functions. Section 5 analyzes optimizing the future advertising budget allocations based on the fitted model.

2 BAYESIAN MEDIA ADVERTISING MODEL

We follow the digital marketing advertising framework to evaluate the effectiveness of the cost for each media channel relative to the business metrics of interest, i.e., the number of attributed installs for the game. Assume We have the advertising spend (cost) $x_{t,c}$ at time t with $t \in \{1, ..., T\}$ under the media channel c with $c \in \{1, ..., C\}$, where T is the total advertising days and C is the total number of media channels. Define $Y_{t,c}$ as the response variable, specifically in our context, the attributed installs of channel c at day t. Usually the media models would also include the non-media features as the control variables to help improve the model's goodness of fit and interpretability. However, due to the limited availability of the variables, we ignore those factors for a simpler structure.

2.1 Carryover Effect and Shape Effect

As proposed by [1], we model the carryover effect of media advertising with the *adstock* function, which is established as the normalized mixture of weight functions under the lag L, In comparison to the simple decay-effect function, where the effect only modeled by a single decay parameter, the weighted function depends on the lag $l \in \{0, ..., L-1\}$ so that the cumulative media effect can be the weighted average of cost in the current and previous $L-1$ days. The choice of L can be tricky as it depends on the data conditions. A relatively larger L, say $L = 13$, in common can achieve a good approximation. However, a large L will potentially lead to an over-smoothing model if the data is noisy with a short-term advertising period. We also assume the same L among different media channels.

$$adstock(X_{t-L+1,c}, ..., X_{t,c}; \omega_c, L) = \frac{\sum_{l=1}^{L-1} \omega_c(l) X_{t-l,c}}{\sum_{l=1}^{L-1} \omega_c(l)} \quad (1)$$

The weight functions can vary in different forms based on the characteristics of the advertising effects. For example, the media peak effect can appear in a very short term or extend to a longer time. This comes to the utilization of two types of weight functions: the *geometry decay* and the *delayed adstock*. The functions are as follows:

$$\textit{geometry decay:} \quad \omega_c(l;\alpha_c) = \alpha_c^{l}, \quad l = 0, ..., L-1 \tag{2}$$

$$\textit{delayed adstock:} \quad \omega_c(l;\alpha_c,\theta_c) = \alpha_c^{(l-\theta_c)^2}, \quad l = 0, ..., L-1 \tag{3}$$

where α_c is the retention rate of the $c-th$ channel and θ_c is the delay of the peak effect.

The shape effect quantifies advertising saturation. After transforming media costs using the adstock function, these adjusted values are further applied to a curvature function to emulate the diminishing returns of increased advertising. The Hill function, known for its flexibility, is utilized for this purpose:

$$Hill(X_{t,c};\kappa_c,S_c) = \frac{1}{1+(X_{t,c}/\kappa_c)^{-S_c}}, \qquad X_{t,c} \geq 0 \tag{4}$$

This function helps model how additional advertising investment yields progressively smaller increases in output, reflecting real-world market saturation dynamics.

We further combine those two functions as $Hill(X^*_{t,c};\kappa_c,S_c)$ where $X^*_{t,c}$ is the transformed $X_{t,c}$ with the adstock function.

2.2 Bayesian Modeling

We introduce a Bayesian framework to model the impact of advertising expenditure on the game installations, encompassing all considered media channels. The model is defined as follows:

$$y_{t,c} = \tau + \beta_c Hill(X^*_{t,c};\kappa_c,S_c) + \epsilon_{t,c} \tag{5}$$

where $X^*_{t,c} = adstock(X_{t-L+1,c}, ..., X_{t,c};\omega_c, L)$. τ denotes the base effect, capturing the intrinsic level of installs independent of advertising; β_c quantifies the effectiveness of advertising through the $c-th$ media channel; $\epsilon_{t,c}$ is the independent noise.

Utilizing a common baseline effect τ across all media channels simplifies the model by assuming a uniform base level of installs, irrespective of the channel. This assumption proves beneficial when the baseline level of engagement or activity does not vary significantly between channels, or when data limitations prevent a more detailed analysis. Moreover, a common τ facilitates easier isolation and interpretation of the individual effects of media cost on installs. With this setup, variations in Y can more directly be attributed to changes in X, not to intrinsic differences between the channels, which enhances the clarity and directness in interpreting results.

Even with a common baseline, allowing β_c to vary by channel acknowledges that each channel may contribute differently to the outcome. This flexibility captures the unique responsiveness of each channel to advertising spending, thus enabling precise adjustments in marketing strategies. Knowing the specific impact of β_c for each channel aids in allocating budgets more effectively, focusing resources on channels that yield the highest returns on investment.

The model structure offers significant benefits. With fewer parameters to estimate (a single τ instead of one per channel), the model may require less data to achieve stable and reliable estimates, thereby reducing the computational complexity and potential for overfitting. This structure also facilitates straightforward comparisons between channels on how effectively they convert advertising spend into installs, as each β_c reflects the efficiency of one channel independent of a channel-specific baseline. Moreover, although τ is fixed across channels, the model can still be expanded to include other effects such as interaction terms between channels or non-linear effects of advertising spend, providing a robust framework for more complex analyses if needed.

Parameter estimation is conducted through sampling the posterior distribution, defined as:

$$p(\Phi \mid y, X) \propto L(y \mid X, \Phi)\pi(\Phi) \tag{6}$$

Other than τ and β_c, Φ also includes the following parameters:

- σ^2: Variance of the noise, where $\epsilon_{t,c} \sim N(0, \sigma^2)$.
- $\kappa_c(> 0)$: Half-saturation point for the Hill function, ensuring $\text{Hill}(\kappa_c) = \frac{1}{2}$ for any κ_c and S_c.
- $S_c(> 0)$: Slope parameter, controlling the steepness of the response curve.
- $\alpha_c(\in (0, 1))$: Retention rate, indicating the percentage of advertising impact retained over time.
- $\theta_c(\in [0, L - 1])$: Delay of the peak effect.

The distributions for our model parameters are selected as below:

$$\begin{aligned}
\beta_c &\sim \text{HalfNormal}(0, 5) \\
\tau &\sim \text{Normal}(0, 5) \\
\sigma^2 &\sim \text{InvGamma}(0, 5) \\
\kappa_c &\sim \text{Beta}(2, 2) \\
S_c &\sim \text{Gamma}(3, 1) \\
\alpha_c &\sim \text{Beta}(3, 3) \\
\theta_c &\sim \text{Uniform}(0, L - 1)
\end{aligned}$$

The priors set up are based on the previous studies and sensitivity analysis to ensure appropriate prior robustness. For this first stage, we perform prior comparisons with the informative and non-informative priors due to the small sample size limitation. For example, we can try the informative prior for the retain rate under the expert knowledge that the digital marketing generally present a retain rate around 0.2. We also try a hierarchical form to facilitate learning from the data, but the current setting can already capture the underlying dynamics and relationships among the parameters effectively.

3 EMPIRICAL DATA ANALYSIS

We conduct a comprehensive analysis of a real-world advertising campaign that spanned five days across four media channels. Our dataset includes the number of installations recorded during the campaign and continues for an additional ten days after the campaign concludes, capturing the extended impact of the advertising efforts.

To ensure model stability and improve interpretability, we apply a logarithmic transformation to the response variable y, representing the number of installs. Similarly, the media expenditure variable X undergoes a logarithmic transformation. These transformations help to normalize the data, reducing skewness and dampening the influence of outliers. We also evaluate the effectiveness of the advertising by comparing two types of weight functions.

For the statistical inference, we use the Stan programming language, executing our Bayesian model with four chains, each running for 2000 iterations.

4 CASE STUDY

In this study, we focus on evaluating the effectiveness of an advertising campaign conducted from February 20, 2024, to March 10, 2024 (total 20 days). The campaign itself only last for around 5 days, and due to the retain effect, the positive number of installs are still recorded after the campaign closed. We utilize the number of installs as our response variable y, and the advertising cost is considered as our media variable X. To enhance the robustness of our model, we apply some transformations to both y and X (the true data is not public and the transformation detail is masked) in order to reduce skewness and mitigate the impact of extreme values thereby providing a more symmetrical distribution for statistical analysis. We also keep a lag of $L = 13$ days.

We fit the geometry and the delay functions respectively. Figure 1 and 2 visualize the density of posterior samples and posterior predicted log installs over time under the geometry function.

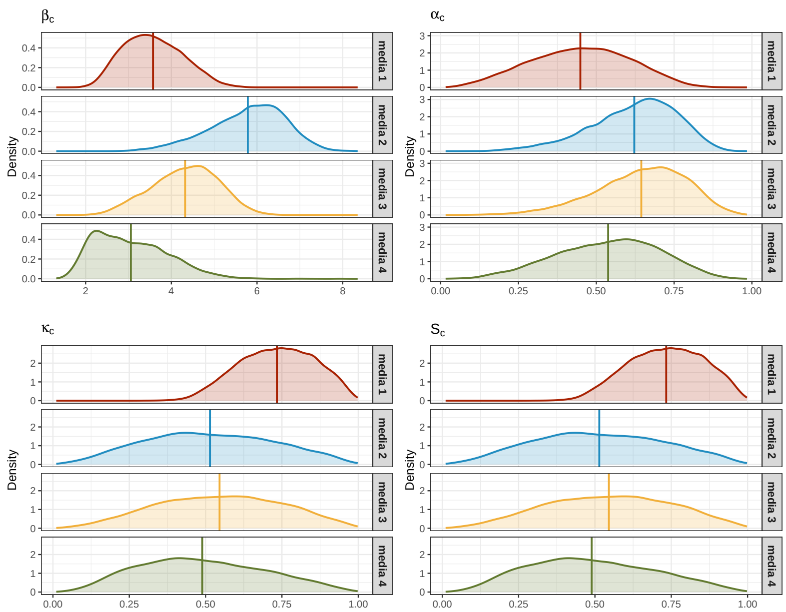

Fig. 1. The density plots of posterior samples under four media channels under the geometry function.

To illustrate the posterior samples of the model parameters, the overall regardless of the model chosen, the posterior means of the effectiveness of the media cost β_c consistently show the following order: media 2 > media 3 > media 1 > media 4. This observation suggests that increasing advertising investment in Media 2 could potentially yield a greater increase in installs compared to the other channels. Moreover, media 2 and 3 tend to present the larger retain rates comparing with media 1 and 4, indicating that investments in these channels are likely more effective in sustaining installs over time.

When comparing the performance across the four media channels, it's evident that each channel exhibits an immediate decline in predicted values from the beginning, except for the media 1. Notably, the number of installs appears to follow a cyclical pattern with a roughly seven-day window, characterized by an initial increase during the first two to three days followed by a subsequent decrease. The raw data are relatively noisy, likely due to the limited sample size. Despite

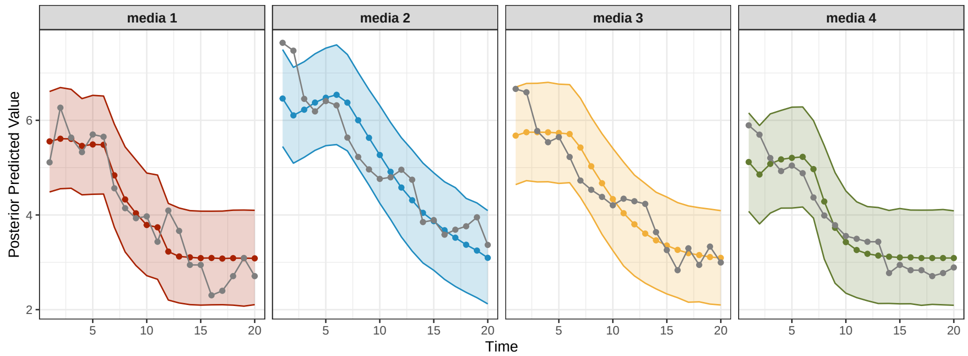

Fig. 2. The posterior predicted log installs over 20 days under four media channels under the geometry function.

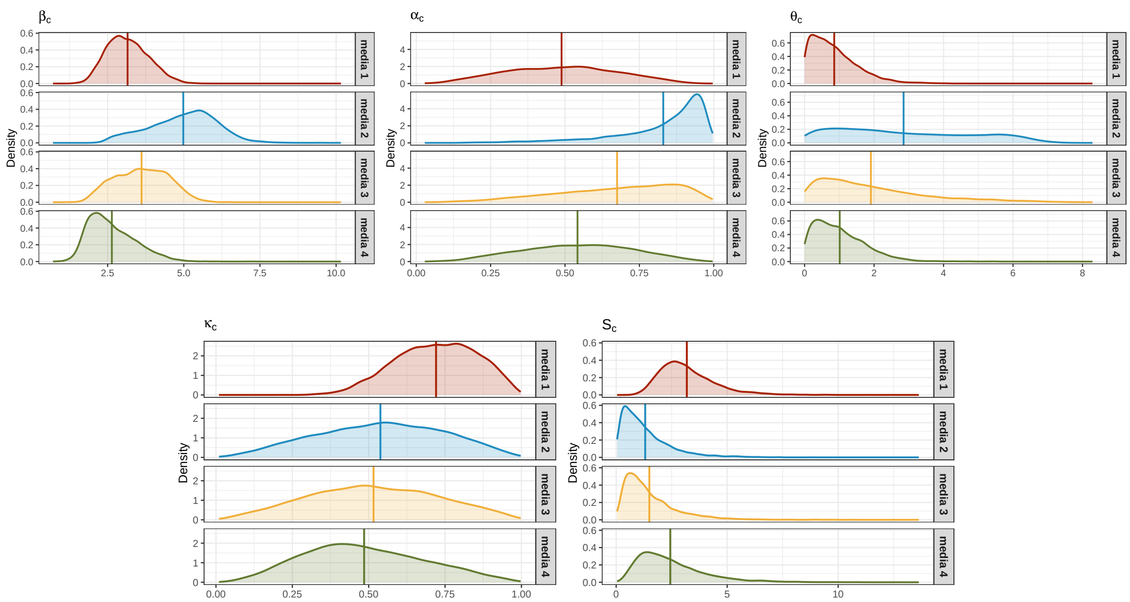

Fig. 3. The density plots of posterior samples under four media channels under the delay function.

this, our model successfully captures the general trend of installs over time across the various media channels. However, it tends to underfit the data, particularly at the beginning of the advertising campaign, where it often underestimates the number of installs.

As for the comparisons between the geometry decay and delay functions, we can observe that the geometric decay often results in a steeper initial drop. This is consistent with the nature of geometric decay, where effects diminish exponentially, causing a sharper decrease initially. Moreover, the stabilization seen in the geometric decay model typically occurs after a sharper decline, whereas the delay function might show more of a gradual tail-off. This reflects the different underlying assumptions: geometric decay assumes an immediate and rapid diminishing impact, while

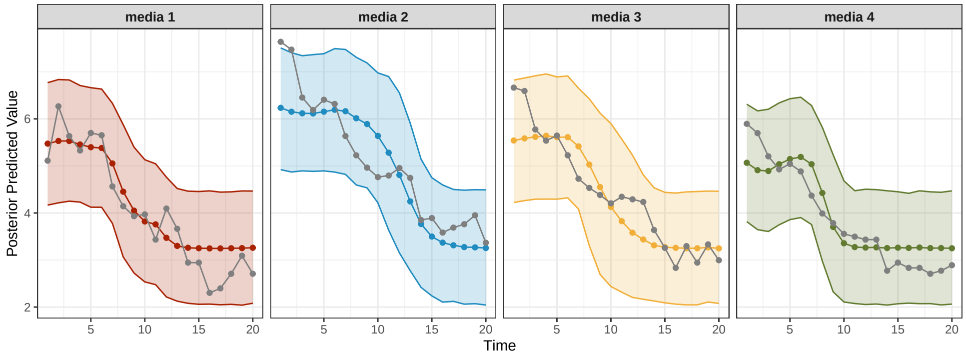

Fig. 4. The posterior predicted log installs over 20 days under four media channels under the delay function.

delay might incorporate more of the cumulative effects over time. From the uncertainty perspective, in geometric decay, uncertainty can be quite wide initially and narrow quickly as the effects stabilize. In contrast, the delay function might show a more gradual narrowing of uncertainty if the model incorporates a wider range of influencing factors over time.

5 ADVERTISING BUDGET OPTIMIZATION

In addition to fitting models based on historical advertising data, we are confronted with a further challenge: if a fixed total advertising budget is available in the future, how can this budget be optimally allocated across different regions and media channels to achieve the best possible advertising outcomes?

To further illustrate this issue, we aim to quantitatively define the 'optimal effect' as maximizing the total number of installations, which implies maximizing the installs garnered under a constrained budget, or equivalently, minimizing the Cost per Install (CPI).

Our optimization problem can be numerically formulated as follows: Suppose we have a utility function $f : X \to \mathbb{R}^+$ that we wish to minimize over a defined domain $X \subseteq \mathcal{X}$.

$$min \frac{\sum_t X_{ct}}{\sum_t f(X_{ct})} \quad s.t \quad \sum_c X_c^* \leq X_{total} \tag{7}$$

This scenario can be approached as a Bayesian decision problem, where the utility function is defined in terms of the expected outcomes (total installations or CPI) based on posterior distributions derived from our model.

Identifying the optimal allocation involves sophisticated optimization techniques because the utility function in question may exhibit complex dependencies on the allocation variables due to the dynamic and uncertain nature of advertising response. These dependencies necessitate the use of advanced statistical methods, potentially including stochastic optimization, to navigate the uncertainties and leverage the predictive power of the Bayesian framework.

6 CONCLUSION

In conclusion, our study provides a comprehensive framework for analyzing the effectiveness of the online advertising campaign with Bayesian models. We incorporate the carryover and shape effects into the model, and validate it under an empirical data set with extremely small sample size and noisy patterns.

It is worth noticing that the posterior mean and variance of the parameters can be sensitive to the choice of priors, especially under the small sample size. Moreover, our model assumes the same kernel, i.e, the weight function and number of lag for all the media channels, which might oversimplify the complex and unique pattern of each media. However, the trade-off between the model flexibility and the fitting needs to be further discussed. For future work, we propose to refine our model structures to account for the heterogeneity of media channels while keeping the model simple and interpretable. We also intend to enhance our analytical framework by incorporating a spatial-temporal dimension into our models, which allows us to examine the effects of geographic variables on our data, providing a more comprehensive understanding of location-based trends and patterns.

ACKNOWLEDGMENTS

We would like to thank Ning Zhang and Tian Zhou for useful comments and discussions.

REFERENCES

[1] Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, and Jim Koehler. 2017. *Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects.* Technical Report. Google Inc.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009