Bayesian Hierarchical Media Mix Model Incorporating Reach and Frequency Data

Yingxiang Zhang, Mike Wurm, Alexander Wakim, Eddie Li, Ying Liu

Google LLC.

July 2023

Abstract

Reach and frequency (R&F) is a core lever in the execution of ad campaigns, but it is not widely captured in the marketing mix models (MMMs) being fitted today due to the unavailability of accurate R&F metrics for some traditional media channels. Current practice usually uses impressions aggregated at regional level as inputs for MMMs, which does not take into account the fact that individuals can be exposed to an advertisement multiple times, and that the impact of an advertisement on an individual can change based on the number of times they are exposed. To address this limitation, we propose a R&F MMM which is an extension to Geo-level Bayesian Hierarchical Media Mix Modeling (GBHMMM) and is applicable when R&F data is available for at least one media channel. By incorporating R&F into MMM models, the new methodology is shown to produce more accurate estimates of the impact of marketing on business outcomes, and helps users optimize their campaign execution based on optimal frequency recommendations.

1 Introduction

Media mix model (MMM) is a statistical technique that helps businesses understand the impact of their marketing mix on sales. To fit MMMs, businesses need to collect aggregate historical time series data on their marketing efforts and sales. This includes data on media channel exposures, other marketing activities, control variables such as weather and seasonality, and sales data. Once the data is collected, statistical models can be used to estimate the return on advertising spend (ROAS) of each marketing channel and to optimize the advertising budget allocations.

MMMs typically assume that there is a time delay between exposure to an ad and a consumer taking action, and that there is ad saturation and diminishing returns at high levels of spend. These effects are nontrivial to estimate by maximizing the log likelihood or equivalently minimizing the residual sum of squares. Jin, Wang, Sun, Chan and Koehler (Jin et al. [2017]) propose a Bayesian media mix model (BMMM) with carryover and shape effects for a single brand aggregated at the national level. The BMMM applies the Bayesian framework and Markov Chain Monte Carlo (MCMC) to estimate model parameters and sample posterior distributions, which incorporates known knowledge into model as prior information. However, The BMMM utilizes data aggregated at a national level, which often suffers from small sample size and insufficient variation in the media spend. To address these issues, Geo-level Bayesian Hierarchical Media Mix Modeling (GBHMMM) (Sun et al. [2017]) is developed, which uses regional level data and pools information across regions to increase the effective sample size. It is also demonstrated that GBHMMM generally provides estimates with tighter credible intervals compared to a model with national level data alone.

In MMMs, advertising spend or number of impressions are generally used as media exposure variables. Impressions can be decomposed into reach and frequency (R&F). Reach is the number of unique people who are exposed to an advertisement at least once, while frequency is the average number of times a person is exposed to an advertisement. Reach and frequency are important factors in the execution of ad campaigns, particularly digital ad campaigns. However, traditional MMMs lack methods tailored to digital media's complexity such as accommodating ad reach and frequency. Additionally, Reach and frequency are more informative than impressions alone. For example, 100 impressions could represent one person seeing the same ad 100 times, or 100 people seeing the ad once each. These two scenarios could lead to very different outcomes, which cannot be captured by MMM if it only takes impressions as input.

Moreover, setting the right frequency for a media channel is crucial for advertisers, as it can have a significant impact on the effectiveness of their media campaigns. However, traditional MMM does not provide a clear answer to the question of what the right frequency is for a particular media channel. There is a copious literature on ad-response to timing and frequency. For example, Sahni [2015] combines field experiments and econometric modeling to measure the causal effect of temporal spacing between ad exposures in an actual market setting, and demonstrates the importance of the number of times s a banner ad is displayed on consumers' choices made during a session. Krugman [1972] argues that three exposures are typically enough to move a consumer from curiosity, to recognition, to decision. He also notes that additional exposures beyond three may not have a significant impact on the consumer's decision-making process. Burton et al. [2017] indicates that consumers exposed to an advertisement 10 or more times had greater purchase intentions than consumers with less exposure.

In this paper we propose a R&F MMM that takes into account both reach and frequency data. R&F MMM is an extension to GBHMMM (Sun et al. [2017]) and is applicable when R&F data is available for at least one media channel. R&F MMM provides a data-driven way to investigate the optimal frequency that maximizes the return on ad spend. The benefits of the R&F MMM over GBHMMM are twofold: First, it helps users optimize their campaign execution based on the optimal frequency recommendations. Second, the new methodology can produce more accurate estimates of the impact of marketing on business outcomes, which we demonstrate through a simulation study.

The remainder of this paper is organized as follows. In Section 2, we briefly review the GBHMMM. In Section 3, we introduce the R&F MMM. Section 3.1 describes the model specification of the R&F MMM. Section 3.2 compares the GBHMMM and the R&F MMM. Section 3.3 discusses the two attribution metrics, ROAS and mROAS, in detail. Section 3.4 introduces optimal frequency, one of the main usages of the R&F MMM. In Section 4, we evaluate the R&F MMM through simulation studies. We conclude this paper with a brief discussion in Section 5.

2 GBHMMM

We begin by outlining the Geo-level Bayesian Hierarchical Media Mix Modeling, building upon the model form introduced by Sun et al. [2017],

$$y_{t,g} = \tau_g + \sum_{m=1}^{M} \beta_{m,g} Adstock(Hill^*(x^*_{t,m,g}, K_m, S_m), \alpha_m, L) + \sum_{c=1}^{C} \gamma_{c,g} z_{t,c,g} + \epsilon_{t,g}. \tag{1}$$

For geo g at time t, we observe the geo-level response variable $y_{t,g}$, media variables $x_{t,m,g}$ for the media channel $m = 1, ..., M$ and control variables $z_{t,c,g}, c = 1, ..., C$. The time-series of media variable is denoted by $x^*_{t,m,g} = \{x_{s,m,g}, s \leq t\}$. The response variable is usually a KPI (e.g. revenue, online inquiries, etc.). The media variables could be advertising spend or number of impressions delivered. The control variables could include product price, promotions, and macroeconomic factors, such as unemployment rate, gasoline price, etc. The sales and media variables can be scaled by the geo population or target market size. Any control variable that roughly scales with population or market size can also be adjusted to a "per capita" scale. The model parameters follow a Bayesian hierarchical structure where each geo is a sample from the overall population and is allowed to deviate from the population level.

$$\beta_{m,g} \overset{\text{iid}}{\sim} normal(\beta_m, \eta_m^2), m = 1, ..., M,$$

$$\gamma_{c,g} \overset{\text{iid}}{\sim} normal(\gamma_c, \xi_c^2), c = 1, ..., C,$$

$$\tau_g \overset{\text{iid}}{\sim} normal(\tau, \kappa^2),\ \epsilon_{t,g} \overset{\text{iid}}{\sim} normal(0, \sigma^2),$$

where τ, β_m and γ_c are the fixed coefficients or hyperparameters, representing the common mechanism of media impact at the total population level. The geo-level variation is controlled by the standard deviations κ, η_m and ξ_c. Priors are needed for the hyperparameters τ, β_m, γ_c and standard deviations κ, η_m, ξ_c

The shape and carryover effect of advertising is modeled through the Hill function and the geometric Adstock function respectively. The Hill function and the geometric Adstock function are defined as

$$Hill(x; K, S) = \frac{1}{1 + (x/K)^{-S}}, \tag{2}$$

$$Adstock(x_0, ..., x_t; \alpha, L) = \frac{\sum_{l=0}^{L} \alpha^l x_{t-l}}{\sum_{l=0}^{L} \alpha^l}, \tag{3}$$

where $\alpha \in (0,1)$ is the retention rate of the ad effect of the media. The integer L is the maximum duration of carry effect. To simplify the notation, we denote $Hill^*(x^*_{t,m,g}, K_m, S_m) = \{Hill(x_{s,m,g}; K_m, S_m), s \leq t\}$ as the vectorized hill function (3). Hill function is applied before the Adstock transformation to capture the diminishing return of media spend with parameters $K > 0$ and $S > 0$. K is also refered to as EC_{50}, the half saturation point as $Hill(K; K, S) = 1/2$ for any value of K and S. The Hill function goes to 1 as the media spend goes to infinity.

3 R&F MMM

R&F MMM models the response variable directly as a two-dimensional function of R&F instead of a one-dimensional function of impressions. In other words, we are shifting from modeling a response curve (impressions vs response) to a response surface (R&F vs response). This approach is appealing in theory, because it would directly answer many of the fundamental questions advertisers have: if I increase my reach or frequency, how will my sales change? What is the optimal frequency for maximizing my return on ads spend?

3.1 Model Specification

For notation purposes, we describe weekly average frequency as $f_{t,m,g}$ and weekly reach [1] as $r_{t,m,g}$. Both metrics are in-period, meaning that they are not affected by the previous or following week. For example, if the same individual is exposed to the ad in two consecutive weeks, they will be counted as one reach for each week. Let $r^*_{t,m,g} = \{r_{s,m,g}, s \leq t\}$ and $f^*_{t,m,g} = \{f_{s,m,g}, s \leq t\}$ denote the time-series of reach and frequency. We model the R&F MMM using the following generic equation:

$$y_{t,g} = \tau_g + \sum_{m=1}^{M} \beta_{m,g} Adstock(r^*_{t,m,g} Hill^*(f^*_{t,m,g}, K_m, S_m), \alpha_m, L) + \sum_{c=1}^{C} \gamma_{c,g} z_{t,c,g} + \epsilon_{t,g}, \quad (4)$$

$$\begin{aligned}
\beta_{m,g} &\overset{\text{iid}}{\sim} normal(\beta_m, \eta_m^2), m = 1, ..., M, \\
\gamma_{c,g} &\overset{\text{iid}}{\sim} normal(\gamma_c, \xi_c^2), c = 1, ..., C, \\
\tau_g &\overset{\text{iid}}{\sim} normal(\tau, \kappa^2),\ \epsilon_{t,g} \overset{\text{iid}}{\sim} normal(0, \sigma^2).
\end{aligned}$$

R&F MMM modifies the GBHMMM by first applying the Hill function (2) to the weekly average frequency $f_{t,m,g}$ to adjust for saturation and thresholding effects. Next, the adjusted frequency is multiplied by reach to obtain an adjusted impression count, which is in turn fed into the Adstock function (3) to capture lagged effects of media exposure over time. The model specification forms the shape or curvature (an "S" curve) of sales response to frequency, which reflects the fact that higher frequency can help reinforces brand recall, leading to increased incremental sales, and that excessive frequency may result in ad fatigue and diminishing returns. The shape effect is controlled by the parameters K_m, S_m as seen in Figure 1.

[1] The reach can be scaled by geo population or market size, depending on whether the response variable is adjusted to a "per capita" scale.

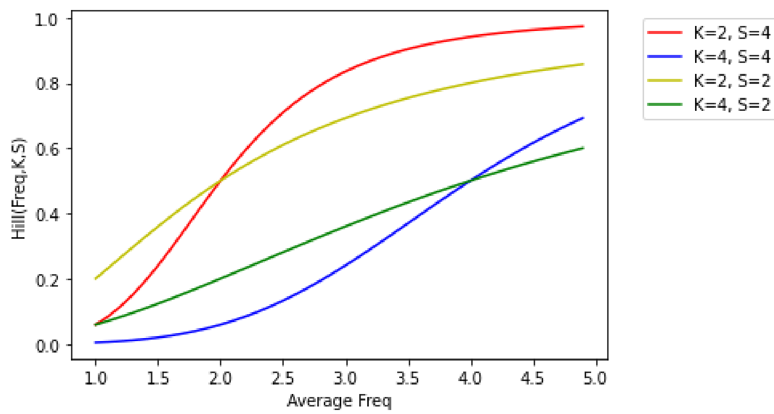


Figure 1: Shape effect of frequency

On the other hand, reach is assumed to have a linear relationship with sales response holding frequency fixed. Reach is contingent on the definition of the target audience, which could be a combination of different groups, each with its own responsiveness to advertising. When using a linear relationship with sales response, we are implicitly assuming that reach across different audiences increases or decreases proportionally. However, it is possible that as the advertiser increases reach, reach on the most desirable audiences does not increase as much as on the less “valuable” audiences, which will lead to diminishing returns. We restrict the effect to be linear instead of using a non-linear transformation like hill function to avoid model overparameterization and non-identifiability of $\beta Hill$ (Jin et al. [2017]). The poor identifiability makes it challenging to estimate the parameters well with any statistical method, and could lead to non-convergence in Bayesian framework. Therefore, we make a trade-off between parsimony and complexity.

3.2 Comparing GBHMMM with R&F MMM

When only impression or cost data is available for each media channel, a Geo-level Bayesian Hierarchical Media Mix Modeling (GBHMMM, Sun et al. [2017]) is typically fit. However, GBHMMM does not account for the fact that individuals can be exposed to an advertisement multiple times, and that the impact of an advertisement on an individual can change based on the number of times they are exposed. To address this limitation, we propose a new approach, R&F MMM, which can be used when R&F data is available for at least one media channel. The side-by-side comparison between R&F MMM and GBHMMM is shown below.

R&F MMM :

$$y_{t,g} = \tau_g + \sum_{m=1}^{M} \beta_{m,g} Adstock(r^*_{t,m,g} Hill^*(f^*_{t,m,g}, K_m, S_m), \alpha_m, L) + \sum_{c=1}^{C} \gamma_{c,g} z_{t,c,g} + \epsilon_{t,g};$$

GBHMMM :

$$y_{t,g} = \tau_g + \sum_{m=1}^{M} \beta_{m,g} Adstock(Hill^*(x^*_{t,m,g}, K_m, S_m), \alpha_m, L) + \sum_{c=1}^{C} \gamma_{c,g} z_{t,c,g} + \epsilon_{t,g}.$$

The R&F MMM and GBHMMM models have the same number of parameters, and they both follow the same Bayesian hierarchical structure which pools information across geos and incorporates prior knowledge. An overview of Bayesian hierarchical models can be found in Gelman and Hill [2006] and Gelman and Pardoe [2006]. Furthermore, R&F MMM is also applicable to data where some channels have only impression or cost data, while the other channels have reach and frequency data. For the channels that only have impression or cost data, R&F MMM yields the same formula as GBHMMM if we replace $f_{t,m,g}$ with $x_{t,m,g}$ and replace $r_{t,m,g}$ with 1. For the channels that have reach and frequency data, R&F MMM in Equation (4) can be fitted and help users optimize their campaign execution based on the optimal frequency recommendations.

3.3 ROAS and mROAS

Before estimating the optimal frequency in the R&F MMM, we first define the attribution metrics, in particular the Return on Ad Spend (ROAS) and the marginal ROAS (mROAS). In this section, we illustrate the methods to estimate ROAS and mROAS for the R&F MMM. The method for the GBHMMM can be found in Sun et al. [2017].

ROAS is the change in revenue per dollar spent on the media; it is usually estimated by setting spend of the media to zero in the selected time period and comparing the predicted revenue against that of the current media spend. For consistency, we use the same notations as described in Sun et al. [2017]. Let the predicted sales at geo g and time t be $\hat{Y}_{t,g}(R_{t,g}, F_{t,g}, Z_{t,g}; \Phi_g)$, where $R_{t,g} = \{r^*_{t,m,g}, 1 \le m \le M\}$ and $F_{t,g} = \{f^*_{t,m,g}, 1 \le m \le M\}$ are the time series of reach and frequency at geo g up to time t, $Z_{t,g} = \{z_{t,c,g}, 1 \le c \le C\}$ is the control variables at geo g and time t, and $\Phi_g = \{\tau_g, \beta_{m,g}, K_m, S_m, \alpha_m, L, \gamma_{c,g}, 1 \le m \le M, 1 \le c \le C\}$ is the model parameters of geo g. To estimate the predicted sales when changing the spend of media m in the selected time period, we denote $R^{a,m}_{t,g}$ and $F^{a,m}_{t,g}$ as the time series of reach and frequency at geo g up to time t, with the m-th media reach and frequency multiplied by a constant a during the period $[T_0, T_1]$, for example, $R^{1,m}_{t,g}$ represents the observed reach time series and $R^{0,m}_{t,g}$ represents the reach time series with m-th media channel turned off during $[T_0, T_1]$. Let the media spend for the mth media channel at time t and geo g be $C_{t,m,g}$. The ROAS at geo g for media m is defined as

$$ROAS_{m,g} = \frac{\sum_{T_0 \le t \le T_1 + L} (\hat{Y}_{t,g}(R^{1,m}_{t,g}, F^{1,m}_{t,g}, Z_{t,g}; \Phi_g) - \hat{Y}_{t,g}(R^{0,m}_{t,g}, F^{1,m}_{t,g}, Z_{t,g}; \Phi_g))}{\sum_{T_0 \le t \le T_1} C_{t,m,g}}, \quad (5)$$

Although the media variables are only changed during the period $[T_0, T_1]$, the impact on sales is calculated in the range $[T_0, T_1 + L]$ to account for the carry over effect. By substituting $\hat{Y}_{t,g}$ with equation $(4)^2$, the ROAS can be re-written as

$$ROAS_{m,g} = \frac{\sum_{T_0 \le t \le T_1} \beta_{m,g} r_{t,m,g} Hill(f_{t,m,g}, K_m, S_m)}{\sum_{T_0 \le t \le T_1} C_{t,m,g}}. \tag{6}$$

Note that the numerator does not include the model parameters $\{\tau_g, \alpha_m, L, \gamma_{c,g}\}$, as they are canceled out during the derivation. As the model in (4) is additive in media effect, $ROAS_{m,g}$ does not depend on other media channels except the m-th channel or on control variables.

mROAS for the m-th medium is the additional revenue generated by one-unit increase in spend, usually from the current spent level. When estimating mROAS for R&F MMM, a small increase in spend must somehow be translated to a small increase in Reach and/or Frequency, which could be done in multiple ways. mROAS by reach perturbed at a 1% multiplicative increment on both reach and media spend at geo g, while holding frequency fixed, is defined as

$$mROAS^{Reach}_{m,g}(\Phi_g) = \frac{\sum_{T_0 \le t \le T_1 + L} (\hat{Y}_{t,g}(R^{1.01,m}_{t,g}, F^{1,m}_{t,g}; \Phi_g) - \hat{Y}_{t,g}(R^{1,m}_{t,g}, F^{1,m}_{t,g}; \Phi_g))}{0.01 \times \sum_{T_0 \le t \le T_1} C_{t,m,g}}, \tag{7}$$

mROAS by frequency perturbed at a 1% multiplicative increment on both average frequency and media spend at geo g, while holding reach fixed, is defined as

$$mROAS^{Freq}_{m,g}(\Phi_g) = \frac{\sum_{T_0 \le t \le T_1 + L} (\hat{Y}_{t,g}(R^{1,m}_{t,g}, F^{1.01,m}_{t,g}; \Phi_g) - \hat{Y}_{t,g}(R^{1,m}_{t,g}, F^{1,m}_{t,g}; \Phi_g))}{0.01 \times \sum_{T_0 \le t \le T_1} C_{t,m,g}}. \tag{8}$$

where $R^{1.01,m}_{t,g}$ and $F^{1.01,m}_{t,g}$ are the R&F time series with the m-th media R&F at geo g multiplied by 1.01 during the period $[T_0, T_1]$. The national ROAS and mROAS for the media m given its spend during the period $[T_0, T_1]$ are simply a weighted average of geo-level values,

$$ROAS_m = \sum_{g=1}^{G} w_{m,g} ROAS_{m,g} = \sum_{g=1}^{G} \sum_{t=T_0}^{T_1} \frac{\beta_{m,g} r_{t,m,g} Hill(f_{t,m,g}, K_m, S_m)}{C_m}, \tag{9}$$

$$mROAS^{Reach}_m = \sum_{g=1}^{G} w_{m,g} mROAS^{Reach}_{m,g}, \tag{10}$$

$$mROAS^{Freq}_m = \sum_{g=1}^{G} w_{m,g} mROAS^{Freq}_{m,g} \tag{11}$$

[2] For simplicity, we assume that the response variable in equation (4) is dollar sales, and that no transformation nor population scaling has been applied. If a transformation of sales is used as the response variable, a corresponding inverse transformation should be applied to get the predicted sales.

Where $w_{m,g} = (\sum_{T_0 \le t \le T_1} C_{t,m,g}) / (\sum_{1 \le g \le G} \sum_{T_0 \le t \le T_1} C_{t,m,g})$ is the proportion of media spend in geo g during the change period $[T_0, T_1]$, and $C_m = \sum_{1 \le g \le G} \sum_{T_0 \le t \le T_1} C_{t,m,g}$ is the total media spend over all geos during the period $[T_0, T_1]$. By plugging in each of the draws from the joint posterior distribution of model parameters, we obtain posterior samples of ROAS and mROAS. The calculation can be done for each geo as well as nationally.

3.4 Optimal Frequency

One of the main usages of R&F MMM is to find the optimal frequency for a channel in the selected time period. Optimal frequency refers to the ideal average number of times an advertisement should be shown to the target audience for each geo and time to achieve the maximum ROAS. As it is impracticable for advertisers to execute different frequencies for each geo and time, we imposed a restriction that the same average frequency is applied in all geos and times, denoted by f_m as below

$$f_{t,m,g} = f_m \forall t, m, g,$$

This restriction serves two purposes: first, to obtain a target frequency strategy that is more practical to implement, and second, to make the optimization routine more tractable. It is then intuitive to use the national ROAS (9) as the objective function, such that the optimal frequency would maximize the national ROAS for the m-th media given its spend during $[T_0, T_1]$. The national ROAS equation (9) involves two set of variables, $r_{t,m,g}$ and $f_{t,m,g}$, and each one can be expressed as a function of the other, as in

$$r_{t,m,g} = \frac{x_{t,m,g}}{f_{t,m,g}},$$

where $x_{t,m,g}$ is the total number of impressions for m-th media at geo g and time t. Assume that the total media spend, C_m, and the number of impressions, $x_{t,m,g}$, do not vary with the average frequency, $f_{t,m,g}$, and that the same average frequency, f_m, is applied in all geos and times, the national ROAS equation (9) then only involves one variable, f_m, for a given set of parameter $\beta_{m,g}, K_m, S_m$

$$ROAS_m = \sum_{g=1}^{G} \sum_{t=T_0}^{T_1} \frac{x_{t,m,g} \beta_{m,g} Hill(f_m, K_m, S_m)}{C_m f_m}. \tag{12}$$

In the Bayesian framework, with multiple posterior samples of the parameters, there are two approaches to obtain the optimal frequency. In the first approach, we make the objective function to be the expected ROAS across all posterior samples of $\beta_{m,g}, K_m, S_m$, and the optimal frequency that maximizes the expected ROAS is obtained by

$$
\begin{aligned}
f_m^{optimal} &= \underset{f_m}{\arg\max}\ \frac{1}{J}\sum_{j=1}^{J} ROAS_m^j \\
&= \underset{f_m}{\arg\max}\ \frac{1}{J}\sum_{j=1}^{J}\sum_{g=1}^{G}\sum_{t=T_0}^{T_1} \frac{x_{t,m,g}\beta_{m,g}^j Hill(f_m, K_m^j, S_m^j)}{C_m f_m} \\
&= \underset{f_m}{\arg\max}\ \frac{1}{J}\sum_{j=1}^{J}\sum_{g=1}^{G}\sum_{t=T_0}^{T_1} \frac{x_{t,m,g}\beta_{m,g}^j Hill(f_m, K_m^j, S_m^j)}{f_m}.
\end{aligned} \tag{13}
$$

where $\beta_{m,g}^j, K_m^j, S_m^j$ are the jth sample of the parameters, $ROAS_m^j$ is the estimated national ROAS for the jth posterior sample, and there are J posterior samples in total. It is noteworthy that equation (13) does not have an analytical solution and must be solved numerically.

In the second approach, we plug each posterior sample of $\beta_{m,g}, K_m, S_m$ into the ROAS equation (12) to solve the optimal frequency that maximizes the ROAS for each posterior sample. For example, for the jth sample, the optimal frequency denoted by $f_m^{optimal,j}$ is calculated as

$$
\begin{aligned}
f_m^{optimal,j} &= \underset{f_m}{\arg\max}\ ROAS_m^j \\
&= \underset{f_m}{\arg\max}\ \sum_{g=1}^{G}\sum_{t=T_0}^{T_1} \frac{x_{t,m,g}\beta_{m,g}^j Hill(f_m, K_m^j, S_m^j)}{C_m f_m} \\
&= K_m^j (S_m^j - 1)^{1/S_m^j}.
\end{aligned} \tag{14}
$$

which is the point of intersection between the tangent line through the origin and the hill function as illustrated in Figure 2. By plugging in each of the draws from the joint posterior distribution of model parameters, we obtain the posterior samples of the optimal frequency $\{f_m^{optimal,j}, 1 \leq j \leq J\}$. The reader can also use either the mean, or the median, or the credible interval of the posterior distribution of optimal frequency as a summary result. An auxiliary benefit of getting the posterior distribution of optimal frequency is that it incorporates the uncertainty and informs the user how much he should trust the model in guiding the optimal frequency.

The first approach provides a point estimate of the optimal frequency that maximizes the expected ROAS across all posterior samples, while the second approach shows the posterior distribution of the optimal frequency that maximizes the ROAS for each posterior sample. Both approaches are illustrated on a simulated data set in Section 4.3. Although these two approaches may theoretically produce significantly different results, our simulation studies did not observe significant divergence between the results of these two approaches.

Optimal frequency is also a consideration in media mix optimization, which allocates a total budget across all media to maximize total incremental sales. One key property of optimal frequency is that it yields the highest media-level ROAS for a given media spend. In other

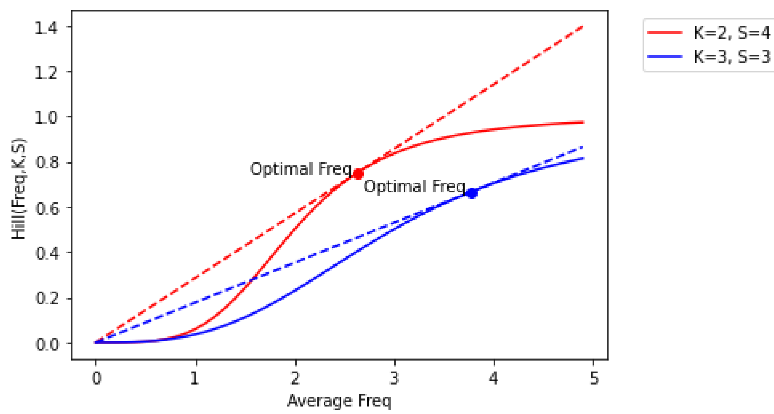


Figure 2: Optimal Frequency

words, the optimal frequency yields the steepest response curves [3], which are then used by media mix optimization to find the highest incremental sales across all media given a total budget. Depending on whether the media mix optimization is to maximize the average of total incremental sales across all posterior samples, or the total incremental sales for each posterior sample, the optimal frequency under the first or the second approach would be used to construct the response curves for the media.

4 Simulation

In this section, we illustrate some of the key benefits of R&F MMM through simulation studies. The simulation is designed to mimic natural consumer-level behavior while also providing ground truth for evaluating R&F MMM and GBHMMM estimation accuracy. Rather than generating MMM aggregate-level data directly, the simulation generates data at the user-level. The user-level data is then aggregated into weekly R&F and sales time series data. We also demonstrate that the optimal frequency estimated from R&F MMM is aligned with the ground truth of the simulation studies.

4.1 Data simulation

To systematically generate the aggregate time-series of reach, frequency, impression, and sales, we begin the simulation at the user level. Specifically, we simulate the individual-level reach events and the number of impressions if being reached, respectively.

Let $r_{i,t,m}$ denote the event that individual i is reached by the ad of media m at week t. It is assumed to be a Bernoulli random variable with probability $p_{t,m}$. This means that each individual has a probability of $p_{t,m}$ of being reached by the ad of media m at week t. If the individual is reached by the ad, the number of ad impressions of media m at week t, denoted by $x_{i,t,m}|r_{i,t,m} = 1$, is assumed to be a zero-truncated Poisson (ZTP) distribution with parameter $\lambda_{t,m}$, which represents the average rate or intensity of ad exposure of channel

[3]Response curves show the estimated relationship between media-level spend and media-level incremental sales. The steeper the response curves, the higher the incremental sales given a media spend.

m at week t if being reached by the ad. If the individual is not reached by the ad, the number of ad impressions of media m at week t, denoted by $x_{i,t,m}|r_{i,t,m} = 0$, is zero. To account for the fact that different media channels have different levels of audience exposure, $\lambda_{t,m}$ is assumed to be an Exponential distribution with rate λ_m, and $p_{t,m}$ is assumed to be a Beta distribution with parameter A_m and B_m

$$\begin{aligned}
&r_{i,t,m} \sim Bernoulli(p_{t,m}),\ p_{t,m} \sim Beta(A_m, B_m),\\
&(x_{i,t,m}|r_{i,t,m} = 1) \sim ZTP(\lambda_{t,m}),\ \lambda_{t,m} \sim Expenential(\lambda_m),\\
&(x_{i,t,m}|r_{i,t,m} = 0) = 0.
\end{aligned}$$

The shape and adstock effect of ad impressions is also captured at individual level. For the number of impressions $x_{i,t,m}$, the immediate convertiness at week t, denoted by $Conv(x_{i,t,m}, t)$, is calculated using a hill function with parameters K_m and S_m. The lagged convertiness at week $t+n$, denoted by $Conv(x_{i,t,m}, t+n)$, is calculated by a geometric decay function with a decay rate of a_m.

$$\begin{aligned}
Conv(x_{i,t,m}, t) &= \frac{1}{1 + (x_{i,t,m}/K_m)^{-S_m}},\\
Conv(x_{i,t,m}, t+n) &= a_m^n Conv(x_{i,t,m}, t).
\end{aligned}$$

The incremental sales for individual i at week T, denoted by $IncSales_{i,T}$, is then calculated as a linear function of the accumulated convertiness for each media channel. Figure 3 illustrates the process of simulating ad impressions and incremental sales at individual level.

$$IncSales_{i,T} = \sum_{m=1}^{M} \beta_m \sum_{t=1}^{T} Conv(I_{i,t,m}, T).$$

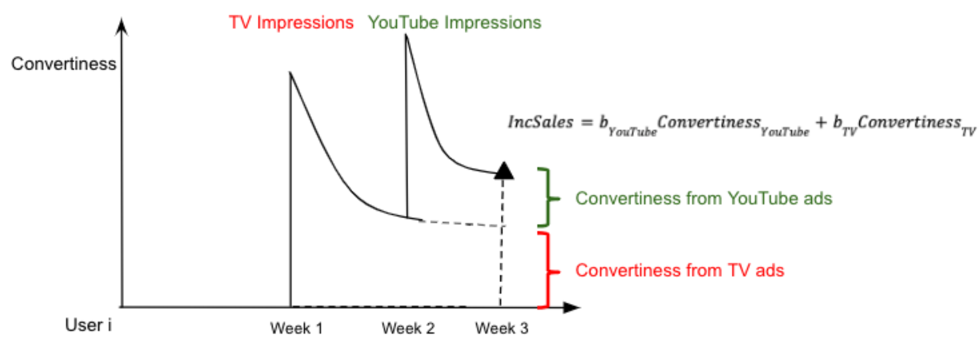


Figure 3: Data Simulation at individual level

Geo-level time-series of media data is then obtained by aggregating individual-level values across all users, with each user randomly assigned to one geo. Let $R_{t,m,g}$ denote the aggregated reach, $F_{t,m,g}$ the aggregated frequency, $I_{t,m,g}$ the total impression, CPM_m the cost per impression, $C_{t,m,g}$ the total media spend, and $IS_{t,m,g}$ the total incremental sales. The calculation is shown below.

$$
\begin{aligned}
R_{t,m,g} &= \sum_{i \in g} \mathbb{1}_{x_{i,t,m}>0}, \\
I_{t,m,g} &= \sum_{i \in g} x_{i,t,m}, \\
F_{t,m,g} &= R_{t,m,g}/I_{t,m,g}, \\
C_{t,m,g} &= I_{t,m,g} CPM_m, \\
IS_{t,g} &= \sum_{i \in g} IncSales_{i,t}.
\end{aligned}
$$

From this simulation scenario, we simulated two media channels and 10000 users who were randomly assigned to three geos with the specifications listed in Table 1. The final data set contains 4 years of weekly data, including sales, media variables and one control variable (weather). The control variable (weather), denoted by $W_{t,g}$, is generated as a Gaussian process with mean W_g to reflect the average temperature of each geo. The sales, denoted by $S_{t,g}$, is the sum of the weekly baseline sales, BS_s, the incremental sales, $IS_{t,g}$, the effect of control variable, $\beta_w W_{t,g}$, and white noise that is assumed to be uncorrelated with the other variables. Figure 4 shows the time series plots of the sales and the media variables for one of the geos.

(a) Media specific parameters

	Media 1	Media 2
A_m	5	5
B_m	4	6
λ_m	4	2.5
K_m	2	2.5
S_m	2	2.5
a_m	0.5	0.7
β_m	2	3
CPM_m	0.2	0.2

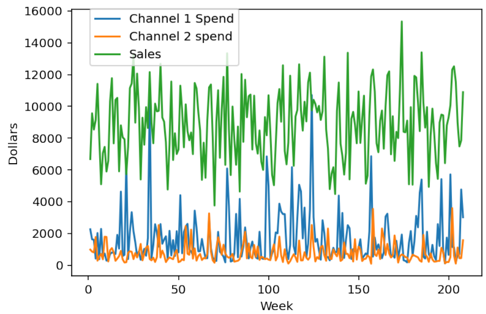
(b) Other variables

	Geo 1	Geo 2	Geo 3
W_g	10	20	15
BS_g	4000	3000	2000

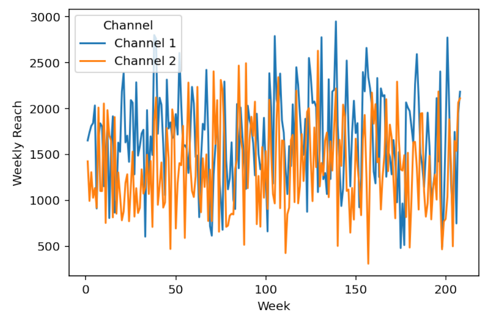
Table 1: Generating parameters

4.2 Comparison of R&F MMM and GBHMMM

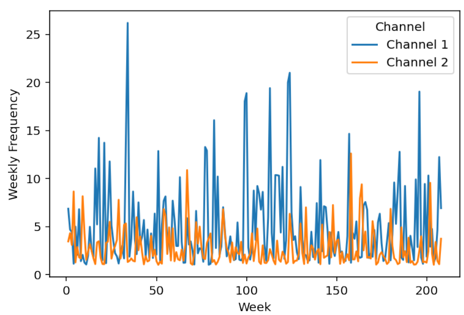
The data simulated from user level provides a fair comparison between GBHMMM and R&F MMM in terms of model performance. Although both GBHMMM and R&F MMM work with aggregated data, which is less granular than user-level data and lacks the ability to capture the nuances of individual user responses, the differences in data aggregation could lead to different levels of information loss, which would impact model performance. In the Bayesian framework, with multiple posterior samples of the parameters, model performance is assessed by plugging in each of the draws from the joint posterior distribution of model



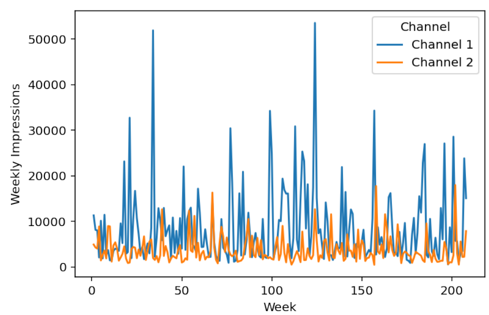
(a) Sales and media spends



(b) Reach



(c) Average Frequency



(d) Number of impressions

Figure 4: Simulated data series for geo 1

parameters to obtain the posterior samples of predicted sales. Then, R-squared is calculated based on the posterior mean of predicted sales and the actual sales.

Table 2 summarizes the R-squared values of R&F MMM and GBHMMM based on the geo-level and national-level data, which is further split into training and testing sets for a comprehensive comparison. R&F MMM captures the actual audience exposure to advertising by considering unique reach and average frequency, rather than simply counting impressions. Impressions, on the other hand, count every instance an ad is served, including potential duplication or multiple exposure of the same individual. R&F MMM offers a more accurate representation of the actual audience reached by marketing efforts.

(a) R&F MMM model fit

Sample	R^2 (Geo)	R^2 (National)
Train	0.993028	0.998032
Test	0.991758	0.998267
Total	0.992819	0.994207

(b) GBHMMM model fit

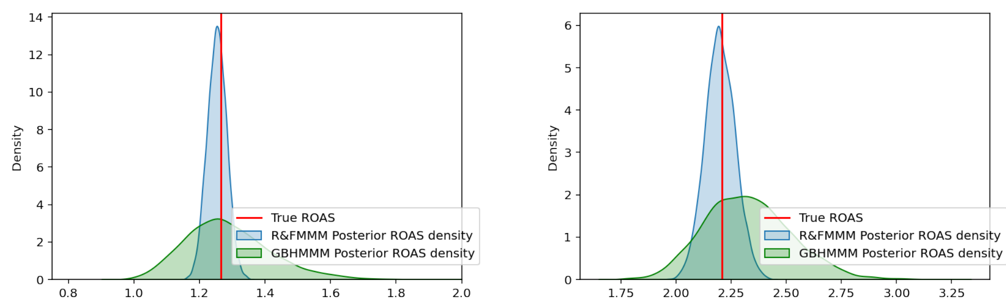
Sample	R^2 (Geo)	R^2 (National)
Train	0.980722	0.989061
Test	0.973629	0.992957
Total	0.979555	0.962520

Table 2: Model performance comparison

4.3 Comparison of R&F MMM and ground truth

Another benefit of using simulated data is that it provides a benchmark, or ”ground truth,” against which to validate R&F MMM. This is because simulated data is generated using a known process, the results can be compared to the “ground truth” to see if there is bias in the model estimates. In this case, the true ROAS was calculated by dividing the total incremental sales by the total media spending. The true optimal frequency was revealed by plotting the weekly frequency for each geo against the ROAS of the media spend for that week and geo, including the lagged effect. This allowed us to see what is the optimal weekly frequency that maximizes the ROAS.

As described in Section 3.3, we obtain the posterior samples of ROAS by plugging the posterior samples of Φ into (9). For comparison, we also obtain the posterior samples of ROAS for GBHMMM. In figure 5, the blue curves are the posterior density of ROAS for R&F MMM, the green curves are the posterior density of ROAS for GBHMMM, and the red lines are the true ROAS of all time. ROAS is estimated reasonably well with the true ROAS falling within the 90% credible interval. Moreover, R&F MMM shows a tighter credit interval for the ROAS estimation than GBHMMM.

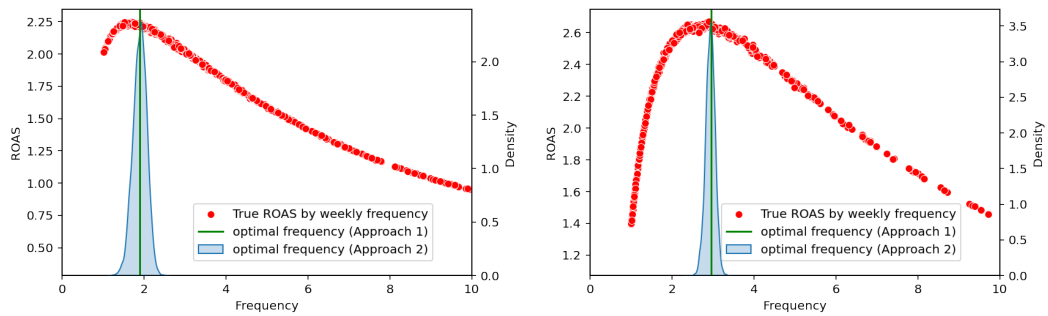


(a) Media channel 1

(b) Media channel 2

Figure 5: Posterior ROAS vs ground truth

As discussed in section 3.4, there are two approaches to obtain the optimal frequency. The first approach provides a point estimate of the optimal frequency that maximizes the expected ROAS, while the second approach provides the posterior distribution of the optimal frequency that maximizes the ROAS for each posterior sample. In figure 6, the green line represents the point estimate of the optimal frequency based on the first approach, the blue curves are the posterior distribution of the optimal frequency according to the second approach, and the red dots are the weekly frequency for each geo against the true ROAS of the media spend for that week and geo, including the lagged effect. The figure shows that the optimal frequencies obtained from both approaches align with the weekly frequency that provides the highest true ROAS.



(a) Media channel 1 (b) Media channel 2

Figure 6: Posterior optimal frequency vs ground truth

5 Conclusion

Reach and frequency (R&F) is a core lever in the execution of ad campaigns. However, it is not widely captured in the marketing mix models being fitted today. We propose a R&F MMM which is an extension to GBHMMM (Sun et al. [2017]) and is applicable when R&F data is available for at least one media channel. R&F MMM is to model the response variable directly as a two-dimensional function of R&F instead of a one-dimensional function of impressions or spends. The model specification forms the shape or curvature (an “S” curve) of sales response to frequency, while assuming a linear relationship between sales response and reach. R&F MMM also uses a Bayesian hierarchical model structure incorporating regional variations to enhance media mix modeling.

By incorporating R&F into MMM models, the new methodology is shown to produce accurate estimates of the impact of marketing on business outcomes, and helps users optimize their campaign execution based on the optimal frequency recommendations. The simulation study provides evidence that R&F MMM is a promising approach for estimating the optimal frequency of advertising. The posterior samples of optimal frequency align with the ”true” optimal frequency which provides the highest ROAS based on the simulation study. R&F MMM and GBHMMM are compared in terms of model performance using data simulated from user level. Although both models work with aggregated data, which is less granular than user-level data, R&F MMM captures the actual audience exposure to advertising by considering unique reach and average frequency. This results in a more accurate representation of the actual audience reached by marketing efforts, and specific insights for media which R&F are relevant.

Acknowledgment

We would like to thank Hitesh Wadhwani and Lisa Jakobovits for their helpful comments and suggestions, Joseph Kelly and Brenda Price for their great discussion and feedback, Harikesh Nair for his guidance and support throughout the research, David Chan, Susanna Makela and Mike Hankin for their previous work on R&F MMM.

References

Jennifer Lee Burton, Jan Gollins, Linda E. McNeely, and Danielle M. Walls. Revisiting the relationship between ad frequency and purchase intentions. *Journal of Advertising Research*, 2017.

Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models (1st ed.)*. Cambridge University Press, 2006. ISBN 052168689X.

Andrew Gelman and Iain Pardoe. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. 2006.

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

Herbert E. Krugman. Why three exposures may be enough. *Journal of Advertising Research*, 1972.

Navdeep S. Sahni. Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quantitative Marketing and Economics*, 2015.

Yunting Sun, Yueqing Wang, Yuxue Jin, David Chan, and Jim Koehler. Geo-level bayesian hierarchical media mix modeling. 2017.