

Geo-level Bayesian Hierarchical Media Mix Modeling

Yunting Sun, Yueqing Wang, Yuxue Jin, David Chan, Jim Koehler

Google Inc.

1 Abstract

Media mix modeling is a statistical analysis on historical data to measure the return on investment (ROI) on advertising and other marketing activities. Current practice usually utilizes data aggregated at a national level, which often suffers from small sample size and insufficient variation in the media spend. When sub-national data is available, we propose a geo-level Bayesian hierarchical media mix model (GBHMMM), and demonstrate that the method generally provides estimates with tighter credible intervals compared to a model with national level data alone. This reduction in error is due to having more observations and useful variability in media spend, which can protect advertisers from unsound reallocation decisions. Under some weak conditions, the geo-level model can reduce ad targeting bias. When geo-level data is not available for all the media channels, the geo-level model estimates generally deteriorate as more media variables are imputed using the national level data.

2 Introduction

Media mix model (MMM) is an analytical approach that uses historical information, such as sales, marketing spend, price, macroeconomic variables, etc., to quantify the impact of various marketing activities on key performance indicators (KPIs, e.g., sales). Regression is used to infer causation from observational data. Although the gold standard for making causal statements is randomized experiments (Chan & Perry, 2017), regression is popular because experiments in advertising face many hurdles (Lewis & Rao, 2015). Advertisers are often unwilling to incur the technical and operational expense of running randomized experiments and a large number of experiments would be needed over sufficiently long time periods, to adequately capture ad shape and carryover effects (Tellis (2006)). Historical media mix data is usually aggregated weekly or sometimes monthly for 2-5 years (Chan and Perry (2017)), providing the possibility to model the media impact over a longer period of time than experiments. Jin, Wang, Sun, Chan and Koehler (2017) propose a Bayesian media mix model (BMMM) with ad shape and carryover effects for a single brand aggregated at the national level. The model suffers from small sample size and insufficient variation in media spend unless strong priors are used. To address these issues, Wang, Jin, Sun, Chan and Koehler (2017) pool data from brands in the same product category and pass the knowledge via informative priors to a single brand within the same category.

Another way to enhance the data available to the modeler is to use regional level information. The country of focus in an MMM can usually be partitioned into a set of geographic areas – which we

call “geos” – and data can often be gathered at that level. The geo level data has a larger effective sample size compared to the national level data as long as the geo level time series are not perfectly correlated with national level values. Also it is common that an advertiser may never decrease the national level media spend close to zero or always keep the spend of a media channel at a level too small compared to other channels. The marketing spend at the geo level generally has a wider range than that at the national level, which is critical to MMM as insufficient variation often leads to extrapolation issues. If we assume that the mechanism of media impact is similar across geos, we can take advantage of this additional level of variation using a geo-level model.

In section 3, we develop a Bayesian hierarchical model, pooling information across geos and incorporating prior knowledge, which could be based on industry experience. Section 3.1 describes the properties of the geo-level data. Section 3.2 introduces the notation and explores various variable transformations. Section 3.3 describes the geo-level model and how to estimate it. The comparison between the national-level and geo-level models is discussed in section 3.4, while the attribution metrics are introduced in section 3.5. We evaluate the GBHMMM via simulation in section 4 and apply the method to the historical geo-level media mix data of advertisers in the auto category in section 5.

In addition, ad targeting bias occurs when the advertiser targets the underlying base demand. Hence the media spend is highly correlated with the base demand. When control variables do not perfectly capture the base demand, biases are introduced and media attribution is incorrect (Chen et al., 2017). Section 4.3 presents simulation studies of the GBHMMM in the presence of ad targeting bias.

Another challenge to the geo-level model is that the geo-level data may not be available for all media channels. Section 4.4 evaluates the GBHMMM in the presence of imputed geo-level ad spend. All computations and figures were done within R (R Core Team, 2015).

3 Methodology

3.1 Geo-level data

The geo-level Bayesian hierarchical MMM (GBHMMM) begins with the identification of geos. It must be possible to serve ads according to these geos with reasonable accuracy and to track ad spend and the response metrics at the geo-level. One possible set of geos is to use cities, which is adopted as a geo-targeting unit by many advertising platforms. It is useful to aggregate cities into larger geos because 1) data at the city level can be sparse and noisy for small cities; 2) consumers could see an ad in one city but travel across city boundaries when they purchase, in which case the wrong city will get credit for the sale in the model, leading to inaccurate attribution.

Advertisers generally track KPIs at a very granular level and it is relatively straightforward to aggregate the KPIs to geos. It is easier to obtain geo-level ad spend data for certain types of media than others. Spot TV advertising is by definition aggregated at specific geographic areas where the TV ads are placed. National TV ad spend in each geo can be estimated from TV Gross Rating Points data at the geo-level and is available from comScore Inc.¹. Price and promotion

¹comScore is an American global media measurement and analytics company providing marketing data and analytics to enterprises, media and advertising agencies, and publishers., <https://www.comscore.com/>

are sometimes available at the geo level. Information Resources, Inc (IRI)² provides price and promotion data for consumer packaged goods (CPG) brands at the geo level and R.L.Polk & Company³ provides such data for auto brands. Digital advertising publishers generally have the capability to break down media exposure and spend by geo.

It may not always be possible to get data from an exhaustive set of geos that cover the entire business region. In that case, the sum of media spend or sales across geos would not be equal to the national level sums. The model we propose is generally still useful in such a case as long as the available geos represent the majority of sales and ad spend.

The geo-level variability in the data is crucial for the geo-level model to outperform a national-level model. In many ad platforms, geo-targeting techniques make it possible to vary ad spend across geo locations. Geo experiments (Vaver & Koehler, 2011) are often used to measure the ad effectiveness by modifying the media spend in randomly picked treatment geos and estimating the response relative to the control geos. Independent variation in marketing spend across geos from the experiments offers the possibility to improve MMM results by eliminating or reducing ad targeting bias and increasing the effective sample size. It is also less expensive for advertisers to change marketing spend in a few selected geos than nationally.

It is also important to normalize both the response and predictor variables by the target population size of a geo in order to build a model with similar media impact across geos. While control variables such as temperature, unemployment rate, etc. are independent of the target population size in each geo, KPI and marketing spend are generally positively correlated with it. In addition, a certain amount of media spend may reach the saturation point in one city but not in another city, simply because they differ in the target population size. Hence in order to build a model with shared ad shape effect, media spend should be standardized to the amount per capita. In some use cases, the census population size may not be a good proxy for the target population. For example, for an auto advertiser, sales may be a better proxy for the target population than the population size because a wealthy smaller town might have a larger target market for BMWs than a poor bigger town. The better we can control for such issues, the more plausible our assumption of a similar media impact mechanism across geos.

3.2 Notation

This section introduces the notation used in the rest of the paper. For geo $g, g = 1, \dots, G$ at time $t, t = 1, \dots, T$, we observe the geo-level response variable $Y_{t,g}$, media variables $X_{t,m,g}$ in the media channel $m = 1, \dots, M$ and control variables $Z_{t,c,g}, c = 1, \dots, C$. 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. Using media spend or exposure for modeling depends on the assumptions modelers want to make and the data available. For simplicity, we refer to media variables in terms of spend throughout the paper. The control variables include product price, promotions, and macroeconomic factors, such as unemployment rate, gasoline price, etc. The number of geos G is usually in tens and the number of time periods T is usually in hundreds depending on the aggregation level of the data.

Let the target population size for geo g be s_g . It varies across geos but is assumed to be the same across time during the observation period. Let Ω_x and Ω_z be the set of indices for media

²A third party data vendor, <https://www.iriworldwide.com>

³Polk is now IHS Automotive, a third party data vendor, <https://www.ihs.com/btp/polk.html>

variables and control variables that are subject to the target population size and Ω_x^c and Ω_z^c be their complements. We standardize the former variables to be the amount per capita using s_g (the reasons can be found in section 3.1). The standardized variables are obtained as follows,

$$\begin{aligned} y_{t,g} &= Y_{t,g}/s_g \\ x_{t,m,g} &= X_{t,m,g}/s_g, m \in \Omega_x; & x_{t,m,g} &= X_{t,m,g}, m \in \Omega_x^c \\ z_{t,c,g} &= Z_{t,c,g}/s_g, c \in \Omega_z; & z_{t,c,g} &= Z_{t,c,g}, c \in \Omega_z^c \end{aligned}$$

for $g = 1, \dots, G, t = 1, \dots, T$.

The response variables may be further transformed by a function $F_y(\cdot)$, for example, $F_y(v) = \log(v + 1)$, although this should be decided on a case-by-case basis since it implies a multiplicative relationship between variables. The ranges of media variables could vary significantly across channels. In order to reduce the search space of parameters and employ common priors across media channels, we recommend normalizing $x_{t,m,g}$ to be between zero and one for each of the M media channels. More details about the standardization of the variables can be found in Appendix 7.1.

It is widely believed that advertising exhibits a lag or carryover effect, i.e., a delayed response that occurs in subsequent time periods of advertising (Tellis, 2006). Several functional forms, such as geometric adstock, polynomial distributed lags (PDL), and delayed adstock, have been proposed to model the carryover effect (Jastram (1955), Jin et al. (2017), Palda (1965)). Also Tellis (2006) shows that advertising could have diminishing returns at high level of spend, referred to as shape effect. Non-linear transformation of media variables are often used to model the curvature of the response, including the logarithmic transformation, Hill function, and logistic growth function (Cain (2005), Jin et al. (2017), Little (1979)).

In this paper, we follow the notations in Jin et al. (2017) to model the carryover and shape effect of advertising through the (normalized) adstock function and the Hill function. The adstock function is defined as

$$\text{adstock}(x_{t-L+1}, \dots, x_t; w, L) = \frac{\sum_{l=0}^{L-1} w(l)x_{t-l}}{\sum_{l=0}^{L-1} w(l)},$$

where $\{x_t, t \geq 1\}$ is a media spend time series and $w(l)$, for $l \in \{0, \dots, L-1\}$ is a nonnegative weight function. The integer L is the maximum duration of carry effect. A commonly used weight function takes the form of geometric decay,

$$w^g(l; \alpha) = \alpha^l, \quad \text{for } l \in \{0, \dots, L-1\}, \quad (1)$$

where $\alpha \in (0, 1)$ is the retention rate of ad effect of the media.

This function assumes that the ad effect peaks at the same time period as the ad exposure, which might not be the case for media channels taking longer to build up effect. To model the delayed peak effect, Jin et al. (2017) introduces the delayed adstock function as

$$w^d(l; \alpha, \theta) = \alpha^{(l-\theta)^2}, \quad \text{for } l \in \{0, \dots, L-1\}, \quad (2)$$

where $\theta \in [0, L-1]$ is the delay of the peak effect and $\alpha \in (0, 1)$ is the retention rate.

To simplify the notation, we denote $x_t^*(\alpha, L) = \text{adstock}(x_{t-L+1}, \dots, x_t; w^g, L)$ as the geometric adstock transformed media spend (1) and $x_t^*(\alpha, \theta, L) = \text{adstock}(x_{t-L+1}, \dots, x_t; w^d, L)$ as the delayed adstock transformed media spend (2). For simplicity of illustration, we will use geometric adstock $x_t^*(\alpha, L)$ in the model specification hereafter. Other adstock functions can be used in a similar fashion.

We apply the Hill function after the adstock transformation in this paper. The Hill function is defined as,

$$\text{Hill}(x; \mathcal{K}, \mathcal{S}) = \frac{1}{1 + (x/\mathcal{K})^{-\mathcal{S}}}, \quad (3)$$

where $\mathcal{S} > 0$, $\mathcal{K} > 0$ and x is the adstock transformed media spend. \mathcal{K} is also referred to as EC_{50} , the half saturation point as $\text{Hill}(\mathcal{K}; \mathcal{K}, \mathcal{S}) = 1/2$ for any value of \mathcal{K} and \mathcal{S} . The Hill function goes to 1 as the media spend goes to infinity. The corresponding response curve is defined as $\beta \text{Hill}(x; \mathcal{K}, \mathcal{S})$ where β is the maximum ad effect achievable. If the true \mathcal{K} is far outside the range of observed historical media spend, the parameters $\mathcal{K}, \mathcal{S}, \beta$ are essentially unidentifiable (Jin et al., 2017). The range of geo-level media spend is generally wider than that at the national level and thus makes the estimation of these parameters more feasible.

3.3 Model specification and estimation

The GBHMMM is a Bayesian hierarchical model with non-linear transformation on some of the predictors. An overview of the Bayesian hierarchical model can be found in Gelman and Hill (2006) and Gelman and Pardoe (2006). The hierarchical linear model or mixed effect model assumes the data is drawn from a hierarchy of subpopulations with repeated measurements on each. Such a model controls for unobserved heterogeneity across subpopulations with random coefficients. The random coefficient has a different value for each subpopulation.

We assume for simplicity that there is no synergy effect between media, which may not be true in practice (Zhang & Vaver, 2017). The geo-level response is modeled as,

$$y_{t,g} = \tau_g + \sum_{m=1}^M \beta_{m,g} \text{Hill}(x_{t,m,g}^*(\alpha_m, L); \mathcal{K}_m, \mathcal{S}_m) + \sum_{c=1}^C \gamma_{c,g} z_{t,c,g} + \epsilon_{t,g}, \quad (4)$$

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

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

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

where each geo is a sample from the overall population and is allowed to deviate from the population level mechanism through the random effects $\tau_g, \beta_g = (\beta_{1,g}, \dots, \beta_{M,g})$ and $\gamma_g = (\gamma_{c,1}, \dots, \gamma_{C,g}), g = 1, \dots, G$. The geo-level variation is controlled by the standard deviations $\kappa, \eta = (\eta_1, \dots, \eta_M)$ and $\xi = (\xi_1, \dots, \xi_C)$, respectively. The parameters $\tau, \beta = (\beta_1, \dots, \beta_M), \gamma = (\gamma_1, \dots, \gamma_C)$ are the fixed coefficients or hyperparameters, representing the common mechanism of media impact at the total population level. Priors are needed for the hyperparameters τ, β, γ and standard deviations κ, η, ξ .

Furthermore, non-linear transformation is applied on media variables to capture carryover and shape effects and $\alpha = (\alpha_1, \dots, \alpha_M), \mathcal{K} = (\mathcal{K}_1, \dots, \mathcal{K}_M)$ and $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_M)$ are the corresponding transformation parameters. We restrict these parameters to be the same across geos but allow

them to vary across media. While we could allow these parameters to vary across geos, we worry that this would raise identifiability problems as described in Jin et al. (2017). Priors for parameters associated with the carryover and shape effects, such as $\alpha, \mathcal{K}, \mathcal{S}$, are needed. In this paper, the maximum duration of carryover effect L is fixed to be the same for all the media channels. L is predetermined by users and not estimated in the GBHMMM, although it would be possible to allow L to be a parameter estimated in the model as well.

We follow the general guidance of setting default priors as in Jin et al. (2017). Alternatively, category level priors (Wang et al., 2017) can be used as informative priors for the GBHMMM. If the data is strong enough, the posterior will be pulled towards the true parameters; otherwise, the posteriors will be close to the priors. If non-negativity is desired for β , we could put a gamma prior or a half-positive normal prior on β . In our experience, the use of an improper prior $\text{uniform}(0, \infty)$ on scale parameters $\sigma, \kappa, \eta, \xi$ could lead to slow convergence. The scale parameters are unlikely to exceed a low range, especially when modeling the response variable on the log scale. We recommend either having a bounded prior or a prior with much of its probability mass near zero, for example, half normal distribution as suggested in Gelman (2006).

A variety of Markov Chain Monte Carlo (MCMC) methods can be used to fit the model described above. For example, STAN (Gelman, Lee & Guo, 2015), which uses Hamiltonian Monte Carlo, offers a general implementation to fit Bayesian models. Customized algorithms, which take advantage of the specific model form, can also be developed to fit such models and be faster than STAN. In particular, the geo-level model is simply a hierarchical linear model when conditioning on the transformation parameters. Gibbs sampling is used to fit the geo-level model by alternating between sampling the transformation parameters and the hierarchical linear model parameters. When conditioning on the parameters of the hierarchical linear model, slice sampling (Neal, 2003) is used to draw samples from the full conditional distribution of the transformation parameters. When conditioning on the transformation parameters, Gibbs sampling is used to draw samples for parameters associated with the hierarchical linear model (Rossi, Allenby & McCulloch, 2005).

The hierarchical structure of the model may not be needed for all predictors. For example, some of the media coefficients and/or the control coefficients could be the same across geos, i.e. $\beta_{m,1} = \dots = \beta_{m,G} = \beta_m$ for some $m \in \{1, \dots, M\}$ and/or $\gamma_{c,1} = \dots = \gamma_{c,G} = \gamma_c$, for some $c \in \{1, \dots, C\}$, while the intercepts are allowed to vary across geos. Such models have much fewer parameters to estimate than a full GBHMMM (Equation (4)). To tell whether hierarchical structures are needed, we could use the Watanabe-Akaike information criterion (WAIC, (Watanabe, 2010)) and cross-validation to estimate pointwise out-of-sample prediction accuracy from the Bayesian models. The model with higher out-of-sample prediction accuracy would be selected.

The geo-level model of a single brand can be extended to multiple brands within a product category, and we call such extension a category-geo-level model. The response of brand b at geo g and time t is modeled as,

$$y_{t,g,b} = \tau_{g,b} + \sum_{m=1}^M \beta_{m,g,b} \text{Hill}(x_{t,m,g,b}^*(\alpha_m, L); \mathcal{K}_m, \mathcal{S}_m) + \sum_{c=1}^C \gamma_{c,g,b} z_{t,c,g,b} + \epsilon_{t,g,b}, \quad (5)$$

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

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

$$\tau_{g,b} \stackrel{iid}{\sim} \text{normal}(\tau, \kappa^2), \quad \epsilon_{t,g,b} \stackrel{iid}{\sim} \text{normal}(0, \sigma^2),$$

for $b = 1, \dots, B, g = 1, \dots, G, t = 1, \dots, T$. The model assumes shared carryover and shape effects across geos and brands, while random intercepts and coefficients are used to account for geo and brand level variation. The model will be illustrated with a real data example from the auto category in section 5.2.

3.4 Comparing GBHMMM with BMMM

When only national-level data is available, we would fit a Bayesian media mix model (BMMM, (Jin et al., 2017)). The BMMM uses media mix data aggregated at the national level and assumes

$$y_t = \tau + \sum_{m=1}^M \beta_m \text{Hill}(x_{t,m}^*(\alpha_m, L); \mathcal{K}_m, \mathcal{S}_m) + \sum_{c=1}^C \gamma_c z_{t,c} + \epsilon_t,$$

$$\epsilon_t \stackrel{iid}{\sim} \text{normal}(0, \sigma^2)$$

where y_t is the response variable, $x_{t,m}$ is the spend of media channel m , for $m = 1, \dots, M$ and $z_{t,c}$ is the control variable for $c = 1, \dots, C$ at the time t . $\tau, \boldsymbol{\beta} = (\beta_1, \dots, \beta_M), \boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_C)$ are the fixed coefficients. Priors are needed for $\tau, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathcal{K}, \mathcal{S}, \sigma$.

It makes sense to fit the GBHMMM using the geo-level data and then compare with the BMMM using the national level data aggregated from the geo level. In order to have a fair comparison, the Hill(\cdot) and the adstock(\cdot), the chosen priors, as well as the standardization and scaling of variables should be the same for the GBHMMM and BMMM.

As the geo level model pools data across geos to increase the effective sample size, it generally provides tighter credible intervals than the national level model when the geos are similar in terms of the media impact mechanism. If the GBHMMM yields wider credible intervals than the BMMM, it may indicate that there is considerable variation in the mechanism of media impact across geos and it makes more sense to conduct individual MMM for each geo, rather than building a joint model across geos.

3.5 Attribution metrics

In this section, we illustrate the methods to estimate the optimal media mix, average return on ad spend (ROAS), and marginal return on ad spend (mROAS) for the geo level model. The method for the national level model can be found in Jin et al. (2017). We will use the national ROAS and mROAS to evaluate the BMMM and GBHMMM in the following sections.

Let the predicted sales at the geo g and the time t be $\hat{Y}_{t,g}(\mathbf{X}_{t,g}; \Phi_g)^4$, where $\mathbf{X}_{t,g} = \{X_{s,m,g}, s \leq t, 1 \leq m \leq M\}^5$ is the time series of the media spend at geo g up to time t and Φ_g is the model parameter of geo g . Similar to the national level model (Jin et al., 2017), an optimal media mix \mathbf{X}_g^* can be derived by maximizing the posterior mean of the predicted sales in the change period given a constraint on the total media spend for geo g . This method generally leads to a different optimal mix for each geo. If a common optimal mix is desired for all the geos, the optimization

⁴The GBHMMM yields predicted sales per capita $\hat{y}_{t,g}$ and we have to multiply back the target population s_g to obtain the predicted sales in geo g , i.e., $s_g \hat{y}_{t,g}$.

⁵The GBHMMM use media spend per capita $x_{s,m,g}$ and we need to scale back to the original media spend $X_{s,m,g}$ for calculating attribution metrics.

should be applied at the national level given a constraint on the total national level media spend, i.e., maximizing the posterior mean of the predicted national sales in the change period with fixed proportion of total media spend allocated to each geo.

The ROAS is the average change in revenue per unit spend on a particular media channel. Following the definition of ROAS in Jin et al. (2017), the ROAS at the geo g for the media m given a model parameter Φ_g is defined as

$$\text{ROAS}_{m,g}(\Phi_g) = \frac{\sum_{T_0 \leq t \leq T_1 + L - 1} \left(\hat{Y}_{t,g}(\mathbf{X}_{t,g}^{1,m}; \Phi_g) - \hat{Y}_{t,g}(\mathbf{X}_{t,g}^{0,m}; \Phi_g) \right)}{\sum_{T_0 \leq t \leq T_1} X_{t,m,g}}$$

where $0 \leq T_0 < T_1 + L - 1 \leq T$ and $\mathbf{X}_{t,g}^{a,m}$ represents the media spend time series at geo g up to time t with the m -th media spend multiplied by a constant a during the period $[T_0, T_1]$, for example, $\mathbf{X}_{t,g}^{1,m}$ represents the observed media spend time series and $\mathbf{X}_{t,g}^{0,m}$ represents the media spend time series with m -th media channel turned off during $[T_0, T_1]$. Although the media spend is only changed during the period $[T_0, T_1]$, the impact on sales is calculated in the range $[T_0, T_1 + L - 1]$ to account for the carryover effect. One reasonable choice of $[T_0, T_1]$ is the most recent one year of the sample period as older data may not be representative of the current environment while a shorter period is not protected from seasonality and may make the estimator less stable.

The mROAS is the incremental change in revenue caused by an additional unit in media spend. Following Jin et al. (2017), the mROAS perturbed at a 1% multiplicative increment on media channel m for geo g is defined as

$$\text{mROAS}_{m,g}(\Phi_g) = \frac{\sum_{T_0 \leq t \leq T_1 + L - 1} \left(\hat{Y}_{t,g}(\mathbf{X}_{t,g}^{1.01,m}; \Phi_g) - \hat{Y}_{t,g}(\mathbf{X}_{t,g}^{1,m}; \Phi_g) \right)}{0.01 \times \sum_{T_0 \leq t \leq T_1} X_{t,m,g}}.$$

where $\mathbf{X}_{t,g}^{1.01,m}$ is the media spend time series with the m -th media spend multiplied by 1.01 during the period $[T_0, T_1]$.

The national ROAS and mROAS can be calculated similarly using the change of predicted sales and media spend aggregated at the national level, i.e., $\sum_{g=1}^G \hat{Y}_{t,g}(\mathbf{X}_{t,g}; \Phi_g)$ and $\sum_{g=1}^G X_{t,m,g}$. The national ROAS and mROAS for the media m given a model parameter $\Phi = (\Phi_1, \dots, \Phi_G)$ are simply a weighted average of geo-level values,

$$\text{ROAS}_m(\Phi) = \sum_{g=1}^G w_g \text{ROAS}_{m,g}(\Phi_g), \quad \text{mROAS}_m(\Phi) = \sum_{g=1}^G w_g \text{mROAS}_{m,g}(\Phi_g)$$

where $w_g = \left(\sum_{T_0 \leq t \leq T_1} X_{t,m,g} \right) / \left(\sum_{1 \leq g \leq G} \sum_{T_0 \leq t \leq T_1} X_{t,m,g} \right)$ is the proportion of media spend in geo g during the change 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.

The values of ROAS and mROAS depend on the model parameters and the fighting strategy of the media spend over time. The estimation of ROAS requires prediction at zero media spend. If there are few observations close to zero, the model may not predict well at zero spend and thus the ROAS won't be estimated accurately; the mROAS is not affected as much.

As the national-level model cannot estimate ROAS and mROAS for each geo, we will compare the BMMM and GBHMMM through the national ROAS and mROAS.

4 Results From Simulations

This section presents properties of the geo level model with simulated datasets generated from the same model class as in Equation (4). We compare the geo-level model with the national-level model in various simulation scenarios mimicking the challenges in reality. The comparison is based on the national ROAS and mROAS as well as the response curve.

4.1 Simulation configuration

We assume that there is only one media variable ($M = 1$), one control variable ($C = 1$), five geos ($G = 5$) and variables are aggregated weekly. The unobserved weekly seasonal base demand per capita h_t is simulated as a sinusoid over time for two years (104 weeks), with mean zero and standard deviation one, $h_t = 1.41 \cos(2\pi(t - 12)/52)$.

We then simulate the media spend per capita in the geo g to have a positive correlation ρ_g with the base demand, $x_{t,g} = u_g + \rho_g h_t + \sqrt{1 - \rho_g^2} v_{t,g}$, where u_g is a constant and $v_{t,g}$ is white noise independent of h_t and $vec(v) \sim \text{normal}(0, I_T \otimes I_G)$. As the media spend should be non-negative, we take the positive part of $x_{t,g}$ as the media spend.

The geo level sales per capita $y_{t,g}$ are simulated to depend on the base demand and the media spend,

$$y_{t,g} = \tau_g + \beta_g \text{Hill}(x_{t,g}^*(\alpha, L); \mathcal{K}, \mathcal{S}) + \gamma_g h_t + \epsilon_{t,g}, \quad (6)$$

$$\tau_g \stackrel{iid}{\sim} \text{normal}(\tau, \kappa^2), \quad \beta_g \stackrel{iid}{\sim} \text{normal}(\beta, \eta^2), \quad \gamma_g \stackrel{iid}{\sim} \text{normal}(\gamma, \xi^2), \quad \epsilon_{t,g} \stackrel{iid}{\sim} \text{normal}(0, \sigma^2)$$

where the parameters are summarized in Table 1. The patterns of unobserved base demand h_t as well as the carryover and shape effect are shared across geos.

We also simulate a control variable $z_{t,g}$ for each geo to be positively correlated with the underlying base demand h_t , $z_{t,g} = \rho_z h_t + \sqrt{1 - \rho_z^2} \omega_{t,g}$, where $\omega_{t,g}$ is white noise independent of h_t and $vec(\omega) \sim \text{normal}(0, I_T \otimes I_G)$. As h_t is unobserved, we would use $z_{t,g}$ instead of h_t to build models.

α	\mathcal{K}	\mathcal{S}	L
0.5	2	3	4
β	τ	γ	
2	4	2	
κ	η	ξ	σ
0.1	0.1	0.1	0.2

Table 1: Model parameters

Assuming that the geos are of equal size, the national level variables are simply the average of the geo-level variables⁶,

$$x_t = \frac{1}{G} \sum_{g=1}^G x_{t,g}, \quad z_t = \frac{1}{G} \sum_{g=1}^G z_{t,g}, \quad y_t = \frac{1}{G} \sum_{g=1}^G y_{t,g}.$$

⁶We assume that all the simulated variables are the amount per capita.

α	\mathcal{K}	\mathcal{S}
uniform(0, 1)	uniform(0, 5)	gamma(1.5, 0.5)
β	τ	γ
normal(0, 5 ²)	normal(0, 5 ²)	normal(0, 5 ²)

Table 2: Priors on the model parameters

When the geos have the same distribution of media spend, the true national-level response curve is approximately the same as the geo-level response curve $\beta\text{Hill}(x; \mathcal{K}, \mathcal{S})$. The two curves diverge when the geos have distinct average media spend⁷.

Figure 1 summarizes this bottom-up simulation process. In the following simulation studies, we fit GBHMMM using the geo level data and BMMM using the aggregated national level data for each of the 100 simulated datasets. As the media spend time series are different across simulated datasets, the true ROAS and mROAS are slightly different in each simulation as well. We place the same priors on the shared model parameters of GBHMMM and BMMM (Table 2). Both models are fit using a Gibbs sampler with 10,000 iterations⁸.

Posterior means of the response curves $\beta\text{Hill}(x; \mathcal{K}, \mathcal{S})$ and ROAS metrics (defined in section 3.5) are reported for each model on each simulated dataset. Let R_i be the true ROAS and \hat{R}_i be the model based posterior mean of the ROAS from the i -th simulated dataset, for $i = 1, \dots, 100$. The relative bias of the ROAS is defined as $\frac{1}{100} \sum_{i=1}^{100} \frac{\hat{R}_i - R_i}{R_i}$ and the mean squared error (MSE) of the ROAS is defined as $\frac{1}{100} \sum_{i=1}^{100} (\hat{R}_i - R_i)^2$.

⁷The national level response curve is approximately $\frac{1}{G} \sum_{g=1}^G \beta_g \text{Hill}(w_g x; \mathcal{K}, \mathcal{S})$, where w_g is the proportion of media spend allocated to geo g and x is the media spend per capita at the national level.

⁸We set the first half of the iterations as burn-in iterations.

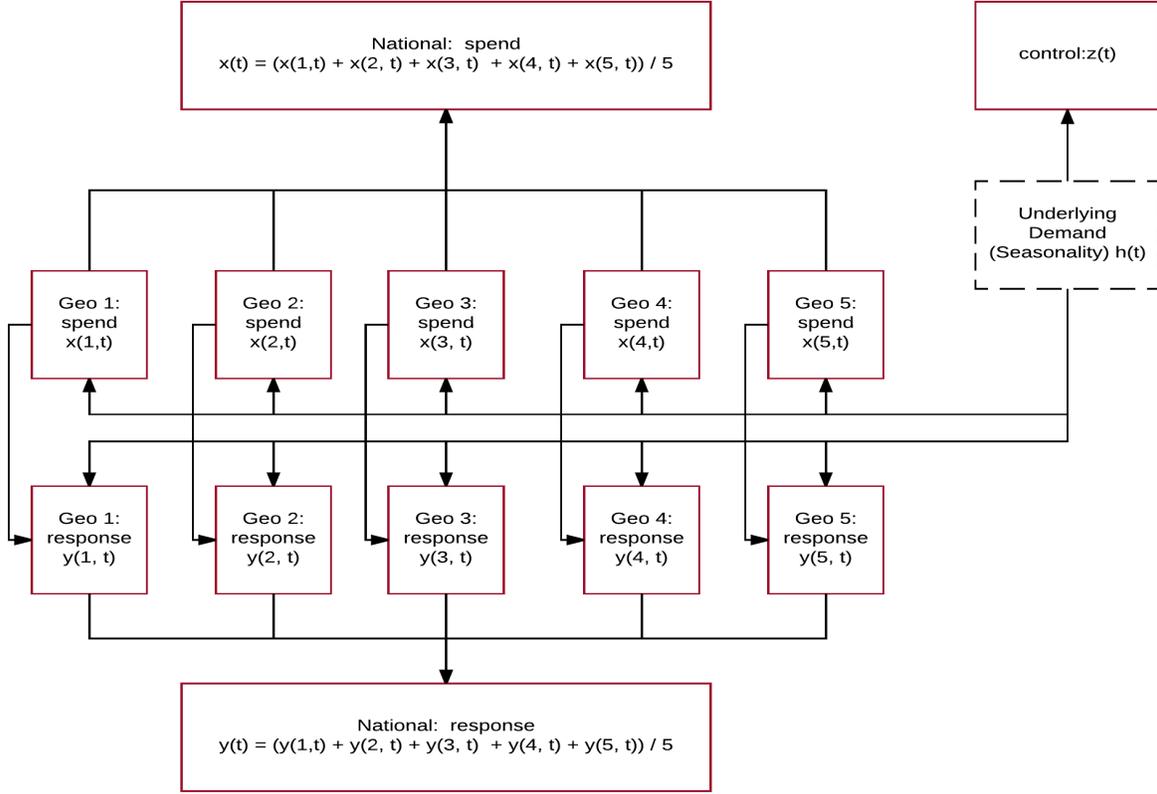


Figure 1: The diagram illustrates the process of simulating the geo and national level media mix data. Simulate the unobserved base demand h_t as a sinusoid shared across geos; Simulate the geo-level media spend positively correlated with the base demand; Simulate the geo-level response variable based on the media spend and the base demand; Simulate the control variable positively correlated with the base demand h_t . Simulated data in the dashed line box is unobserved and cannot be used in the model. ⁹

4.2 Baseline

In this section, we show that the geo-level model benefits from a larger effective sample size and more useful variation in the data, and thus has tighter credible intervals and yields less biased point estimates than the national level model. The improvement is more pronounced when the ranges of the media spend per capita significantly vary across geos. We set the correlation between the media spend and the base demand ρ_g to be 0.5 across all the geos. The control variable is simulated to perfectly capture base demand h_t , i.e., $z_{t,g} = h_t$, for $g = 1, \dots, G$.

⁹As noted in Chan and Perry (2017), the ratio of standard deviation of spend to standard deviation of dollar sales is a key indicator of how much information the data can contain about the impact of the media. The parameter values chosen for this simulation are more favorable to the modeler than is typical.

4.2.1 Simulation I: homogeneous spend distribution across geos

In this study, all the geos have the same distribution of media spend, with a mean spend of $u_g = 2$ and standard deviation one, for $g = 1, \dots, G$.

Pointwise posterior means of the response curves across simulated datasets are shown in Figure 2, plotted using the R package *boom*. It shows that the geo level data has wider range of media spend than the national level data. The BMMM performs much worse than the GBHMMM with more uncertainty and larger bias, especially in the region where few observations are available, i.e. in the beginning of the curve ($x < 1$) and the end of curve $x > 3$. In the middle of the curve $1 \leq x \leq 3$ where the curve is almost linear, the national-level model captures the shape of the curve but is positively biased. The bias in the response curve comes from the particular shape of the response curve and low variation in the media spend which mislead the national-level model to underestimate the slope \mathcal{S} and the intercept τ while overestimating the media coefficient β (Figure 3). The parameters are compensating for each other so that the model has better fit in regions with more observations.

Figure 4 shows the average and marginal ROAS estimates. As the national-level model overestimates the response curve, it overestimates the average ROAS. On the other hand, as the national-level model captures the slope of the curve in the region with majority of the observations, its marginal ROAS seems to be much less biased than the average ROAS. Nonetheless it still has much more uncertainty than the geo-level estimate. In general, the geo-level model has tighter credible interval and lower bias for the average ROAS than the national-level model as demonstrated by Table 3.

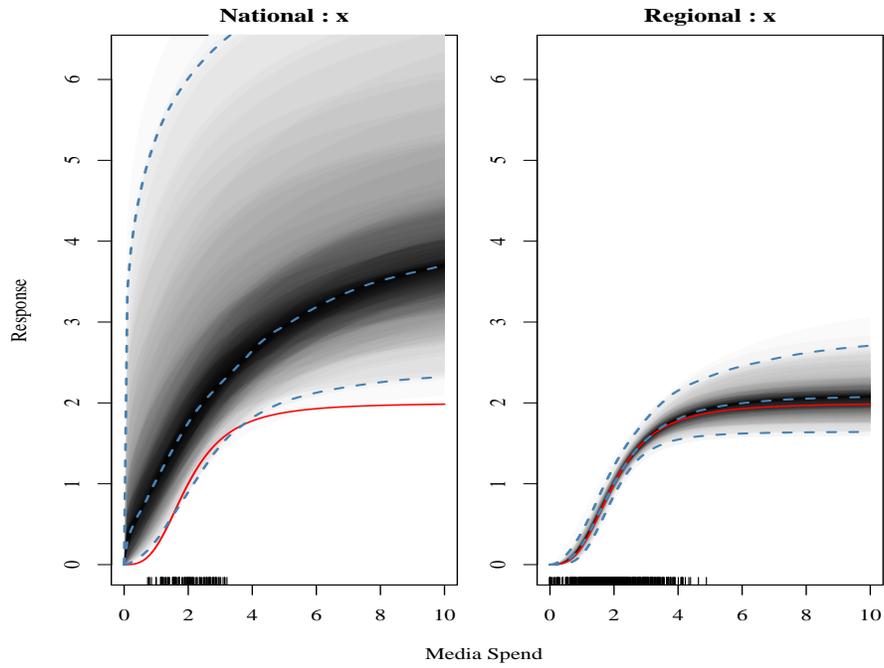
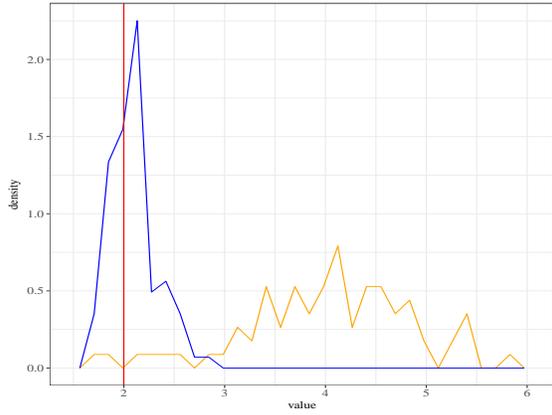
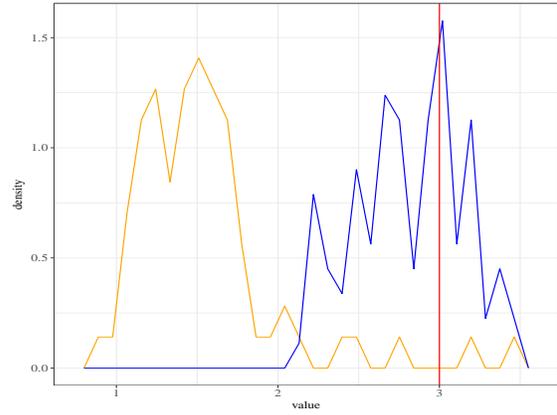


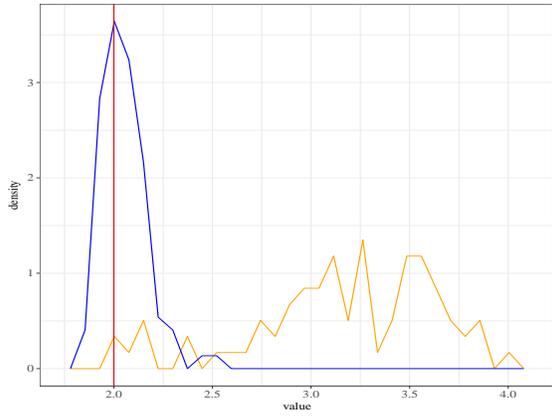
Figure 2: Simulation I: pointwise posterior means of response curves $\beta_{\text{Hill}}(x; \mathcal{K}, \mathcal{S})$ over 100 simulated datasets. The geo level media spends are simulated to have the same distribution. The darker the area, the denser the curves. The true response curve is in red. The tick marks along the X-axis shows the values of media spend used in the model.



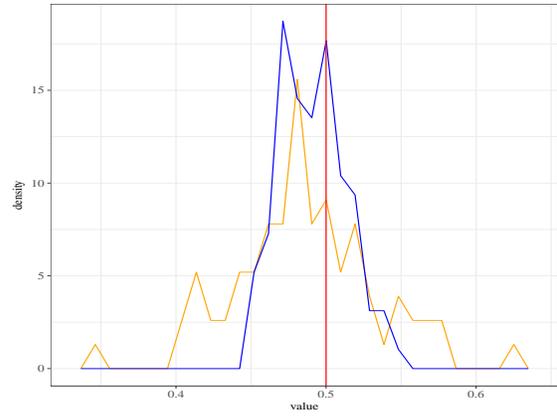
(a) media coefficient β



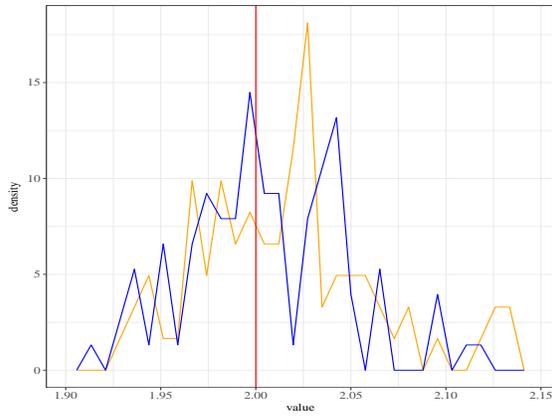
(b) slope \mathcal{S}



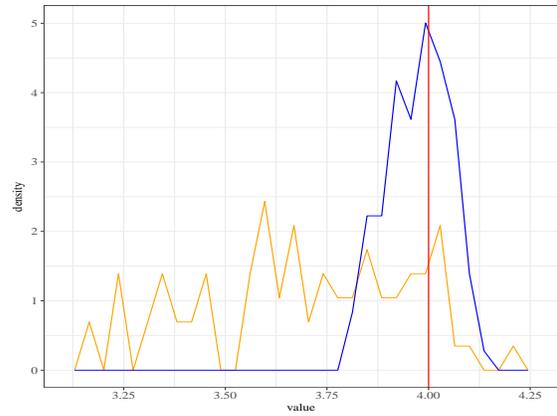
(c) EC_{50}, \mathcal{K}



(d) retention rate α



(e) control coefficient γ



(f) intercept τ

Figure 3: Density of posterior median of parameters over 100 simulated datasets (Simulation I). The national level model is represented in orange and the geo level model is represented in blue. The red vertical line indicates the true value.

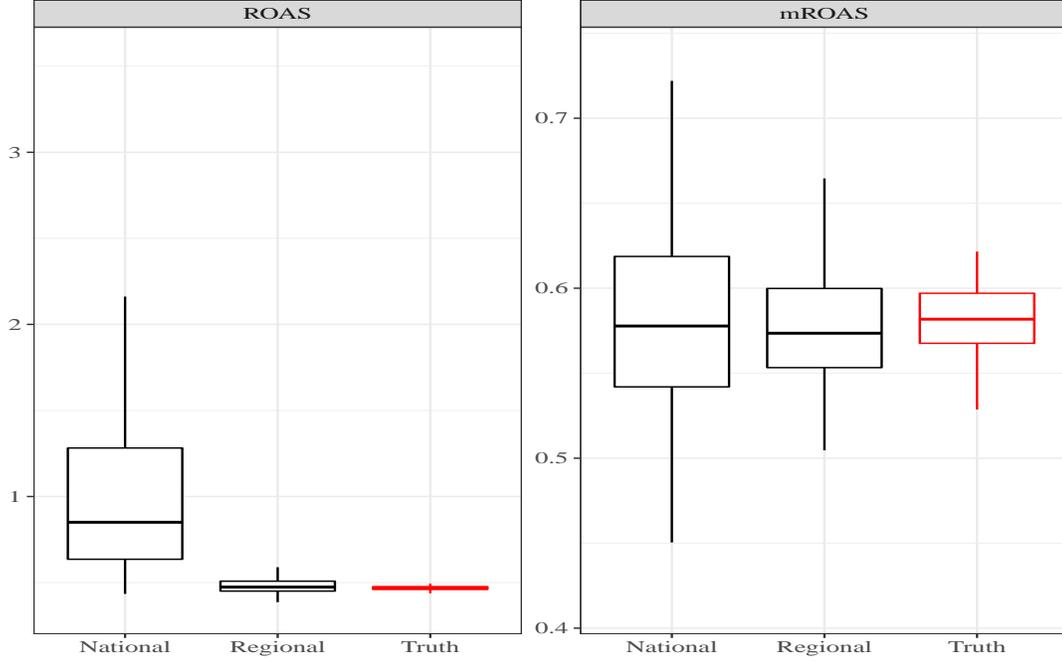


Figure 4: Simulation I: boxplot of the estimated national ROAS and mROAS over 100 simulated datasets. The geo level media spends are simulated to have the same distribution.

model	mean.bias	mse	relative.mean.bias	relative.mse
National	0.62	0.82	132.62%	3.81
Regional	0.01	0.00	2.06%	0.01

Table 3: Simulation I: Bias and MSE of the average ROAS estimates

4.2.2 Simulation II: heterogeneous spend distribution across geos

In this study, the mean media spend is different across geos, in particular, geo g has an average media spend $g + 1$ and standard deviation one for $g = 1, \dots, G$. Some geos have reached the saturation point of media spend while others have not.

Figure 5 shows that geo-level data has wider range of media spend than national-level data and the difference is more pronounced than that in Simulation I. As a result, the nationally aggregated data masks even more information available at the geo level. The national-level model is not able to recover the true response curve while the geo-level model can.

Comparisons of the average ROAS and marginal ROAS estimates can be found in Figure 6. The national-level model yields estimates with more bias and uncertainty for both average and marginal ROAS than the geo-level model. Table 4 demonstrates that the GBHMMM has lower bias and mean squared error for the average ROAS and the improvement over BMMM is more pronounced than the first simulation study. The relative bias of the ROAS estimate from the national-level model is more than 300 times of that from the geo-level model.

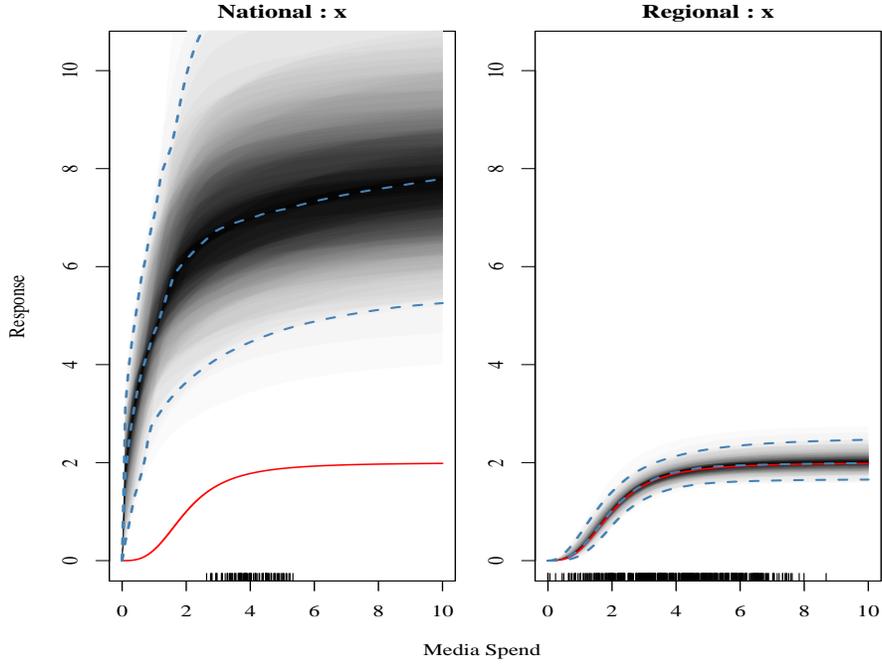


Figure 5: Simulation II: pointwise posterior means of response curves $\beta_{\text{Hill}}(x; \mathcal{K}, \mathcal{S})$ over 100 simulated datasets. The geo level media spends are simulated to have distinct distributions and thus explore different parts of the response curves. The darker the area, the denser the curves. The true response curve is in red. The tick marks along the X-axis shows the values of media spend used in the model.

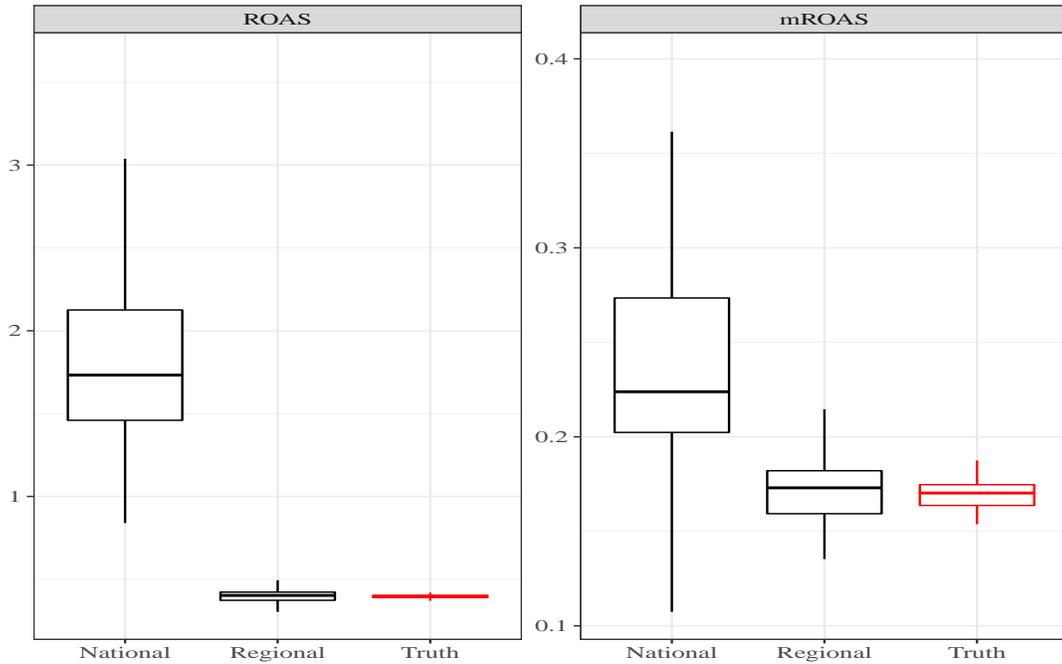


Figure 6: Simulation II: boxplot of the estimated national ROAS and mROAS over 100 simulated datasets. The geo level media spends are simulated to have distinct distributions and thus explore different parts of the response curves.

model	mean.bias	mse	relative.mean.bias	relative.mse
National	1.41	2.22	356.7%	14.18
Regional	0.01	0.00	1.35%	0.01

Table 4: Simulation II: Bias and MSE of the national ROAS estimates

These two simulation studies are ideal cases, free of omitted variables and model mis-specification which often exist in the real data. In the following sections, we will simulate data in more complex settings mimicking the challenges faced by modelers in practice.

4.3 Ad targeting bias

In this section, we demonstrate that with ad targeting bias, the geo-level model helps tighten credible intervals and can reduce bias in the presence of sufficient independent variation in media spend at the geo level.

Ad targeting bias occurs when the advertiser targets the underlying base demand, and hence the media spend is highly correlated with the base demand. When the control variable $z_{t,g}$ does not perfectly capture the base demand h_t , bias is introduced and the media attribution will be incorrect. The higher the correlation between the media spend and base demand, the more severe the bias. Chen et al. (2017) explores the use of search query data to control for ad targeting bias in paid search.

4.3.1 Simulation III: ad targeting bias

We set the correlation between the media spend and the base demand ρ_g to be 0.5 and the average media spend u_g to be 2 across all the geos. The control variable $z_{t,g}$ is simulated to be positively correlated with the base demand h_t with correlation $\text{cor}(z_{t,g}, h_t) = 0.8, g = 1, \dots, G$. The simulation setting is identical to the simulation I except that the control variable does not well capture the underlying base demand and thus introduces ad targeting bias.

As the national level media spend is an average over the geo level media spend, it generally has a higher correlation with the base demand h_t than the geo level data, as long as the geo level base demands are positively correlated (see the derivation in the Appendix 7.2). As a result, we expect the national level model to have worse performance than the geo level model when h_t is not perfectly captured by the control variable. In other words, the independent variation at the geo level can help reduce the correlation between the media spend and base demand and thus reduces the biases in the estimates.

Figure 7 shows that both BMMM and GBHMMM yield biased estimates of response curves because of ad targeting. However, the geo level model benefits from the lower correlation between the media spend and base demand and thus has lower bias for the national ROAS and mROAS as illustrated in Figure 8 and Table 5. The mean squared error of the ROAS estimates from the geo-level model is only one fifth of that from the national-level model.

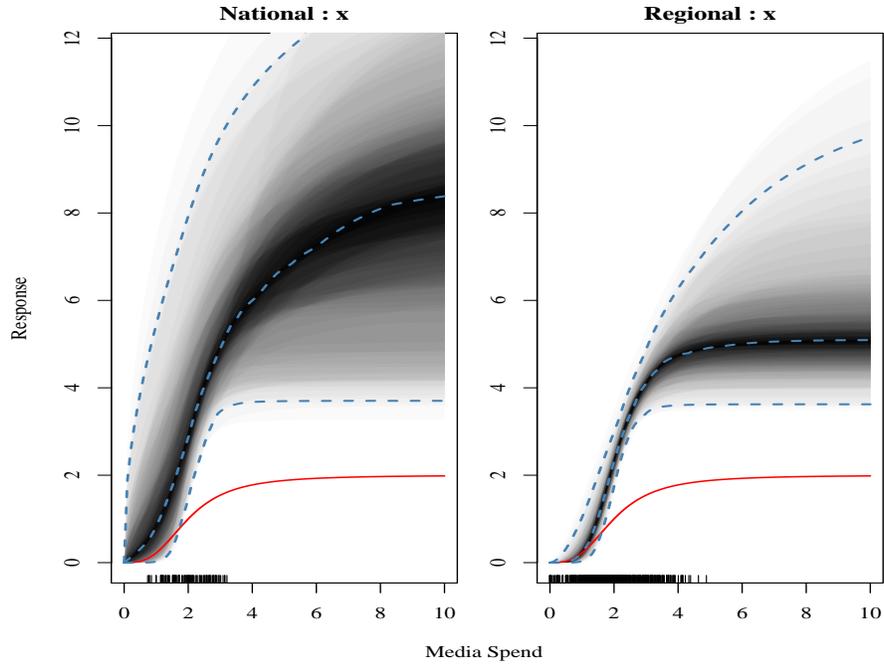


Figure 7: Simulation III: pointwise posterior means of response curves $\beta_{\text{Hill}}(x; \mathcal{K}, \mathcal{S})$ over 100 simulated datasets. In this study, the control variable does not fully capture the base demand and thus introduce ad targeting bias. The correlation between the media spend and the base demand is simulated to be the same across geos. The darker the area, the denser the curves. The true response curve is in red. The tick marks along the X-axis shows the values of media spend used in the model.

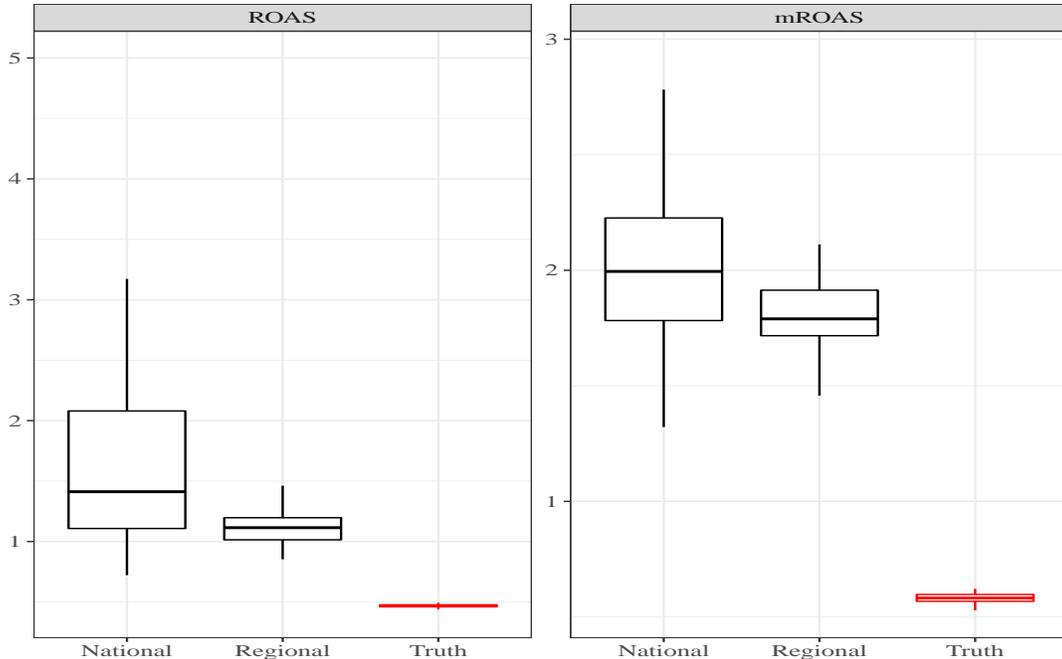


Figure 8: Simulation III: boxplot of the estimated national ROAS and mROAS over 100 simulated datasets. In this study, the control variable does not fully capture the base demand and thus introduce ad targeting bias. The correlation between the media spend and the base demand is simulated to be the same across geos.

model	mean.bias	mse	relative.mean.bias	relative.mse
National	1.23	2.22	263.03%	10.19
Regional	0.65	0.45	140.06%	2.08

Table 5: Simulation III: Bias and MSE of the national ROAS estimates

4.4 Imputed geo-level media spend

In this section, we explore the risk of using imputed geo-level media spend in the GBHMMM through simulation.

It is generally easier to obtain geo-level data for digital media than offline media such as TV or magazine, whose spend is usually aggregated at the national level. When a media variable is not available at the geo level, we have the option to impute the data. One naive imputation method is to approximate the geo-level media variable from its national level value, using the proportion of the population in the geo, relative to the population. There are other imputation methods, but most of them, like the naive one, do not introduce useful variability over time. For the purpose of this simulation study, we simply use the naive imputation approach to illustrate the potential issues with the imputed geo-level media spend.

We refer to the GBHMMM with the actual regional level spend as the full GBHMMM and the geo-level model with the media spend imputed by the national-level spend as the imputed

GBHMMM.

4.4.1 Simulation IV: imputed geo-level media spend

We simulate three media variables using the same model parameters as in Table 1, denoted `media.1`, `media.2`, `media.3`, each positively correlated with the base demand at the geo level. The correlation is set to be 0.5 for all three media channels. Conditioning on the base demand, the three media variables are independent over time. Other simulation settings are kept the same as in simulation I.

We fit the GBHMMM to each of the four datasets derived from a simulated dataset:

- Full: all media variables are observed at the geo level
- `Impute_1_vars`: only `media.1` is imputed.
- `Impute_2_vars`: both `media.1` and `media.2` are imputed.
- `Impute_3_vars`: all three media variables are imputed.

Geo-level response variables are observed in the four datasets despite that the geo-level media spend are partially observed in some. Figure 9 shows that the imputed GBHMMM generally yields larger biases and wider credible intervals for the response curves than the GBHMMM using the full geo-level media spend and the performance of the estimates deteriorates as more media variables are imputed. Similar to the national-level model in section 4.2, the positive biases in the estimated response curves and the ROASs of the imputed geo-level model stem from the low variation in the imputed geo-level media spend and the geometry of the Hill transformation. In this simulation, the geo-level time series of a media channel are positively correlated among themselves, so the national level media spend as a substitute has positive correlation with the geo-level spends. The situation could be even worse if some of the geo-level media spends are negatively correlated. Hence, we want to emphasize the importance of obtaining accurate geo level data for valid media mix inference.

On the other hand, this example confirms that the (partially) imputed geo-level model still yields useful information about the model parameters. In Figure 10 and Table 6, it seems that when only `media.1` is imputed, the ROAS estimates for `media.2` and `media.3` are slightly worse than those of the full geo-level model, but still look reasonable. It might be due to the fact that the correlation between the media variables and the base demand is only 0.5. If the correlations among the media variables increase, we would expect to see worse performance for `media.2` and `media.3`.

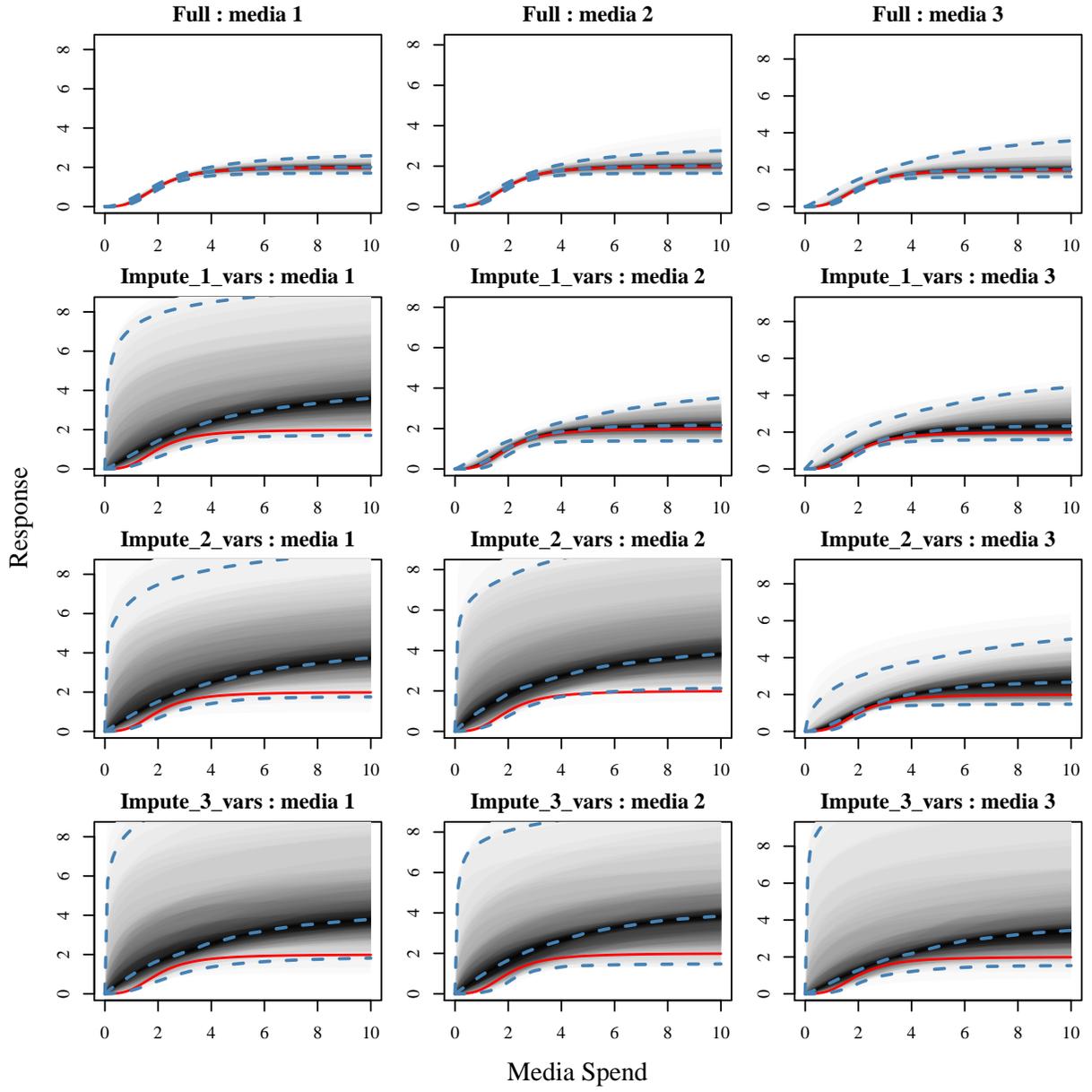


Figure 9: Simulation IV: estimated response curves over 100 simulated datasets. Each curve is the pointwise posterior mean of the curves given draws of the posterior samples of parameters for each simulated dataset. The columns are media channels and the rows are simulated datasets with increasing number of imputed media variables (from top to bottom).

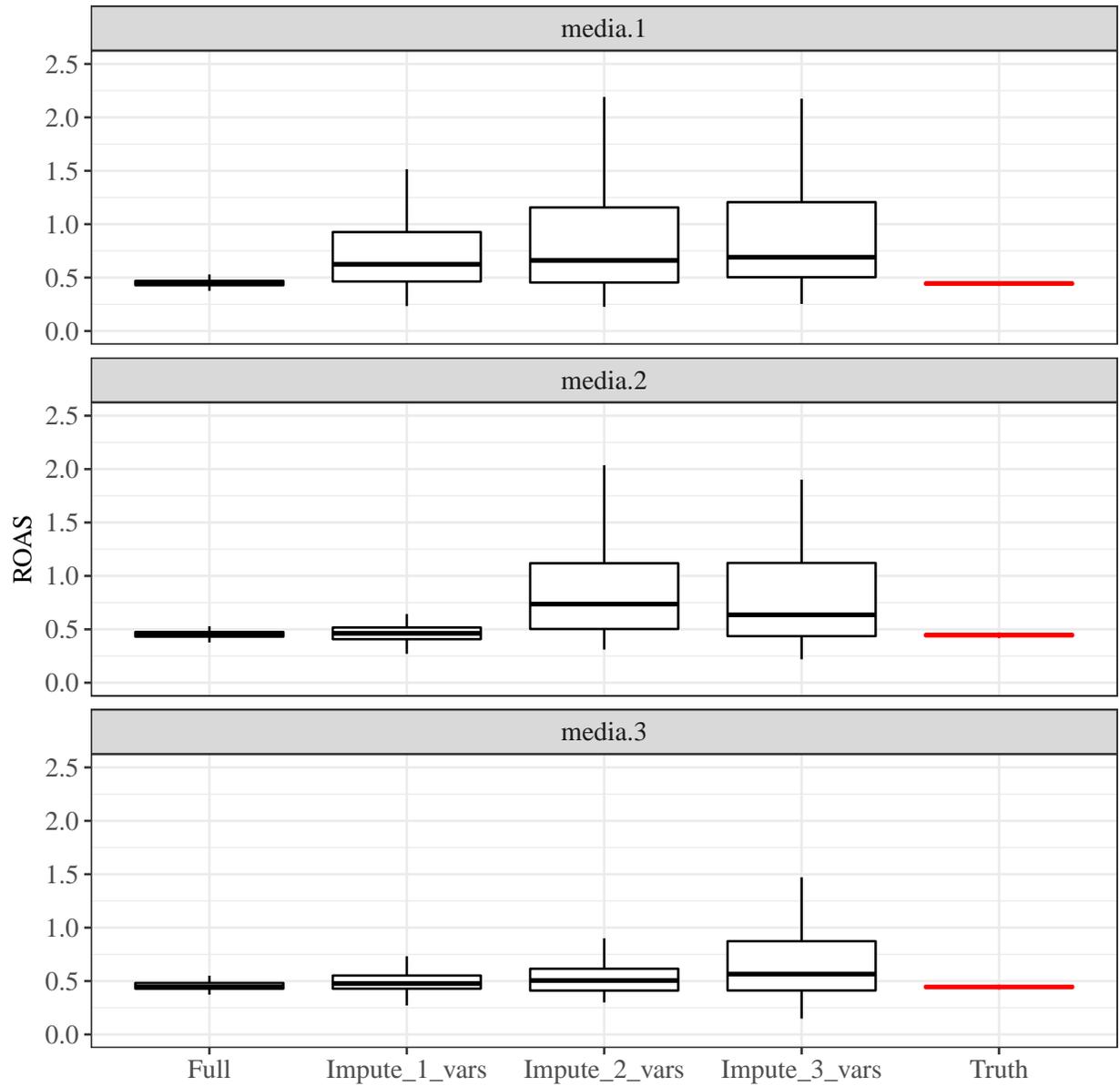


Figure 10: Simulation IV: boxplot of the estimated national average ROAS over 100 simulated datasets. It compares the model performance over simulated datasets with increasing number of imputed media variables.

media.var	model	mean.bias	mse	relative.mean.bias	relative.mse
media.1	Full	0.00	0.00	1.13%	0.01
media.1	Impute_1_vars	0.57	1.10	128.9%	5.57
media.1	Impute_2_vars	0.70	1.82	158.91%	9.47
media.1	Impute_3_vars	0.76	1.75	170.81%	8.85
media.2	Full	0.01	0.00	2.29%	0.01
media.2	Impute_1_vars	0.02	0.01	4.74%	0.04
media.2	Impute_2_vars	0.72	1.40	162.48%	7.08
media.2	Impute_3_vars	0.60	1.12	134.51%	5.62
media.3	Full	0.02	0.00	3.41%	0.02
media.3	Impute_1_vars	0.07	0.03	16.04%	0.15
media.3	Impute_2_vars	0.13	0.09	28.91%	0.46
media.3	Impute_3_vars	0.56	1.55	125.9%	7.73

Table 6: Simulation IV: Bias and MSE of the average ROAS estimates

5 Real Data Case Study

In this section, we illustrate the geo-level model with data of advertisers in the auto category and discuss various challenges.

5.1 Data

The dataset includes monthly media mix data for 12 auto brands at 18 DMAs¹⁰ in the United States from January, 2013 to December, 2014 (24 months). The number of new car registrations was sourced from R.L. Polk & Company. The TV ad spend was sourced from comScore, Inc. The Google search ad spend and search query volume were sourced from Google Inc. Other variables, including new car incentives, market price, tier two media spend¹¹, major and minor car model updates (redesigns) were provided by Neustar MarketShare, which consolidated data from sources such as Kantar Media, IRI, ITG, JD Power, and Rentrak. The brand names and DMA names are anonymized and we will refer to them as brand 1 to 12 and DMA 1 to 18 hereafter.

The 18 DMAs cover more than 50% of total new car registrations in the entire category. Advertising for the 12 auto brands in the US is dominated by TV which comprises more than 80% of tier one¹² nameplate marketing budgets. We calculate the monthly dollar sales by multiplying the number of car registrations with the average sales price of a brand at each DMA. Figure 11 shows the distribution of sales, TV ad spend, and search ad spend across 18 geos (the numbers are relative). These variables are positively correlated across geos because they are all confounded by the geo level target population size.

¹⁰Designated Market Area, geographic areas in the US in which local television viewing is measured by Nielsen.

¹¹Ad spend controlled by local dealerships

¹²Ad spend that comes from the national advertising budget of an automobile manufacturer

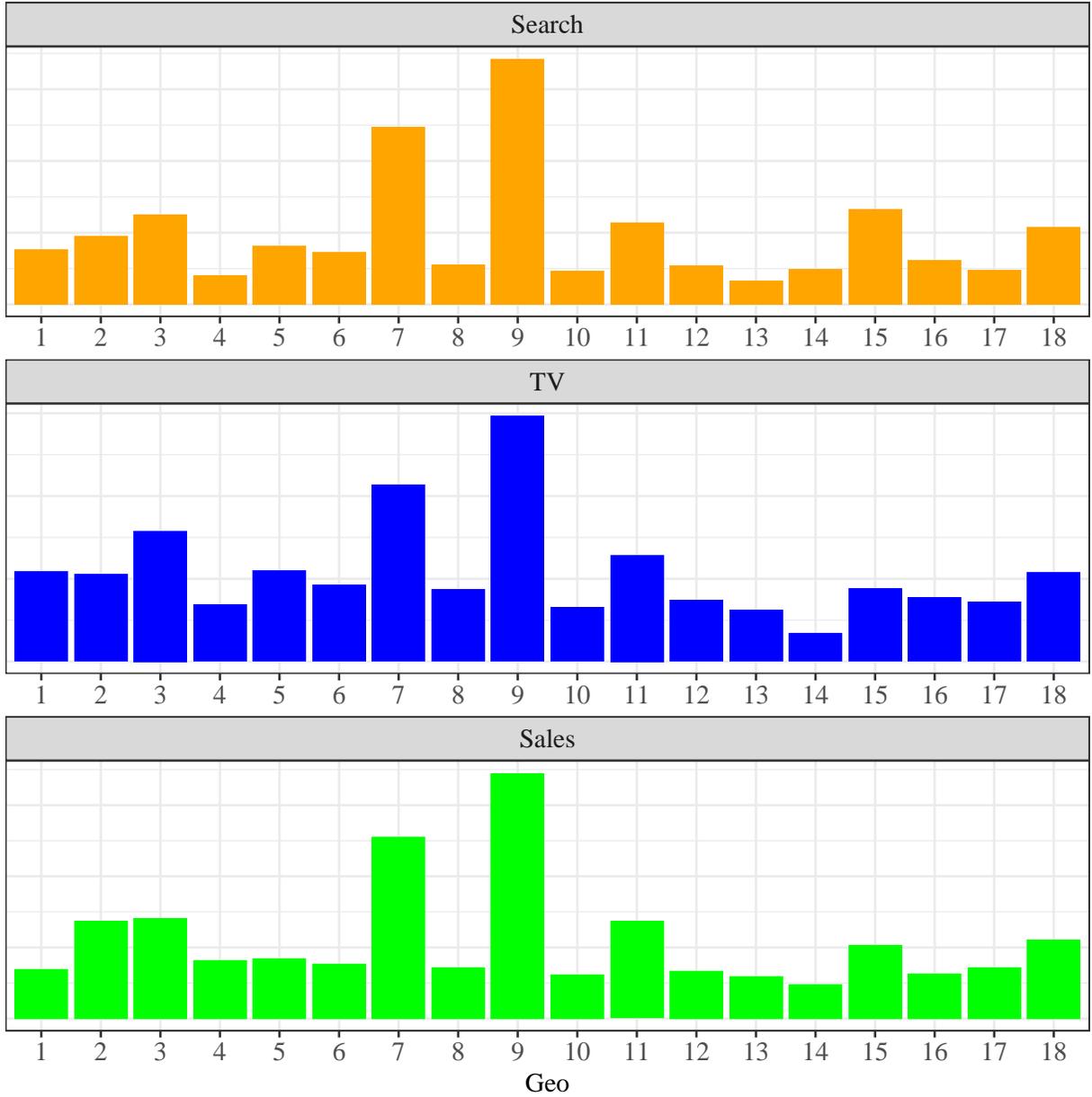


Figure 11: Proportion of total sales, TV ad spend and search ad spend across 18 geos of the auto category.

The target population size in each DMA is estimated by the average monthly sales during the entire sample period. The monthly sales and media spend in each DMA are standardized by the target population size to be the amount per capita. We log transformed dollar sales to be the response variable. The media spends are transformed to be between zero and one according to Appendix 7.1. There is little change in the market price over the two year period for the 12 brands and thus the price is not included in the model. Control variables are centered and scaled within each DMA.

The aggregated search query volume for the 12 brands in each DMA is included in the model to approximate the local seasonal demand for automobiles.

The time of consideration of purchasing a car could be quite long and some ad channels, especially TV, could have a delayed peak effect on sales. To model the carryover effect, we set the maximum duration of ad effect to be three months and employ the delayed adstock function defined in Equation (2). For media channels $m = 1, \dots, M$, we use a uniform(0, 1) prior on the carryover decay parameter α_m and a uniform(0, 3) prior on the carryover delay parameter θ_m . The shape effect is modeled by Hill transformation defined in Equation (3). We use a uniform(0, 1) prior on \mathcal{K}_m and a gamma(1.5, 0.5) prior on \mathcal{S}_m . A normal(0, 1) prior is placed for all the fixed effects (hyperparameters) τ, β, γ . A customized Gibbs sampler with 10,000 iterations is used to fit the model.

5.2 Geo level model of the auto category

We want to compare the geo-level and national-level models for a single brand. However, the national-level data for a single brand only have 24 observations (monthly over the two-year period). Hence we instead compare the category-level model using the national level data (see Wang et al. (2017)) with the category-geo-level model (see Section 3.3 Equation (5)) to understand the benefit of the geo-level data. The national-level media mix data for the entire category (12 brands) have 288 observations. The geo-level media mix data for the entire category have $288 \times 18 = 5184$ observations.

The category-geo-level model yields considerably tighter credible intervals than the category-level model due to larger sample size and wider range of media spend. In Figure 12, both models show that the TV effect peaks around one month after the ad exposure and decays gradually over the next two months, while the search effect peaks around one to two weeks. Figure 13 shows that the distribution of TV spend at the national level is extremely skewed with many observations at or close to zero while the search spend at the national level is approximately normally distributed, with fewer observations close to zero. TV seems to be closer to the saturation point at its maximum historical spend than search. Figure 14 shows the boxplot of the posterior samples of the average ROAS at the national level for 12 brands. The geo-level data reduces the estimation uncertainty in the ROAS estimates. The improvement varies across brands and channels, which could be a result of different media spend variation at the geo level.

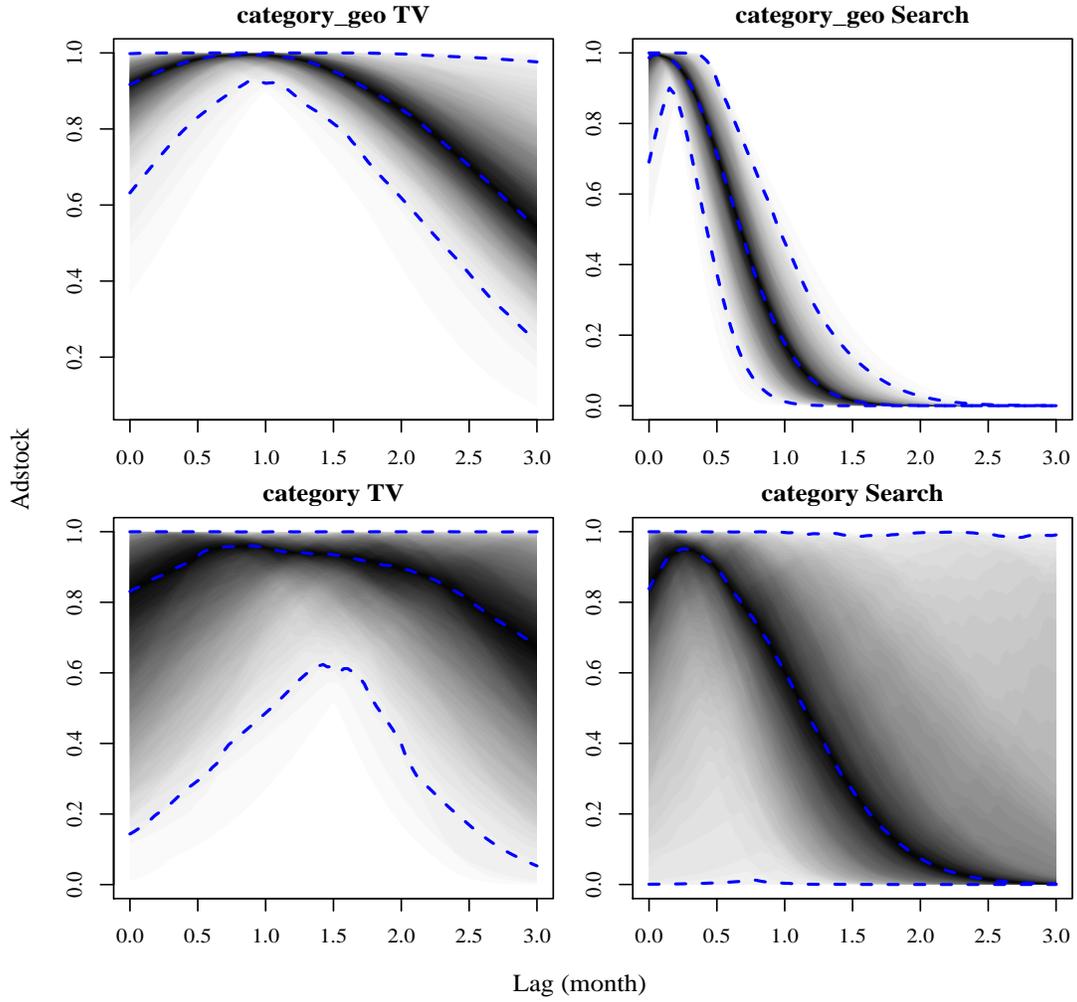


Figure 12: Posterior samples of the delayed adstock functions. The X-axis represents the number of lagged months and the Y-axis represents the delayed adstock function taking values at the X-axis.

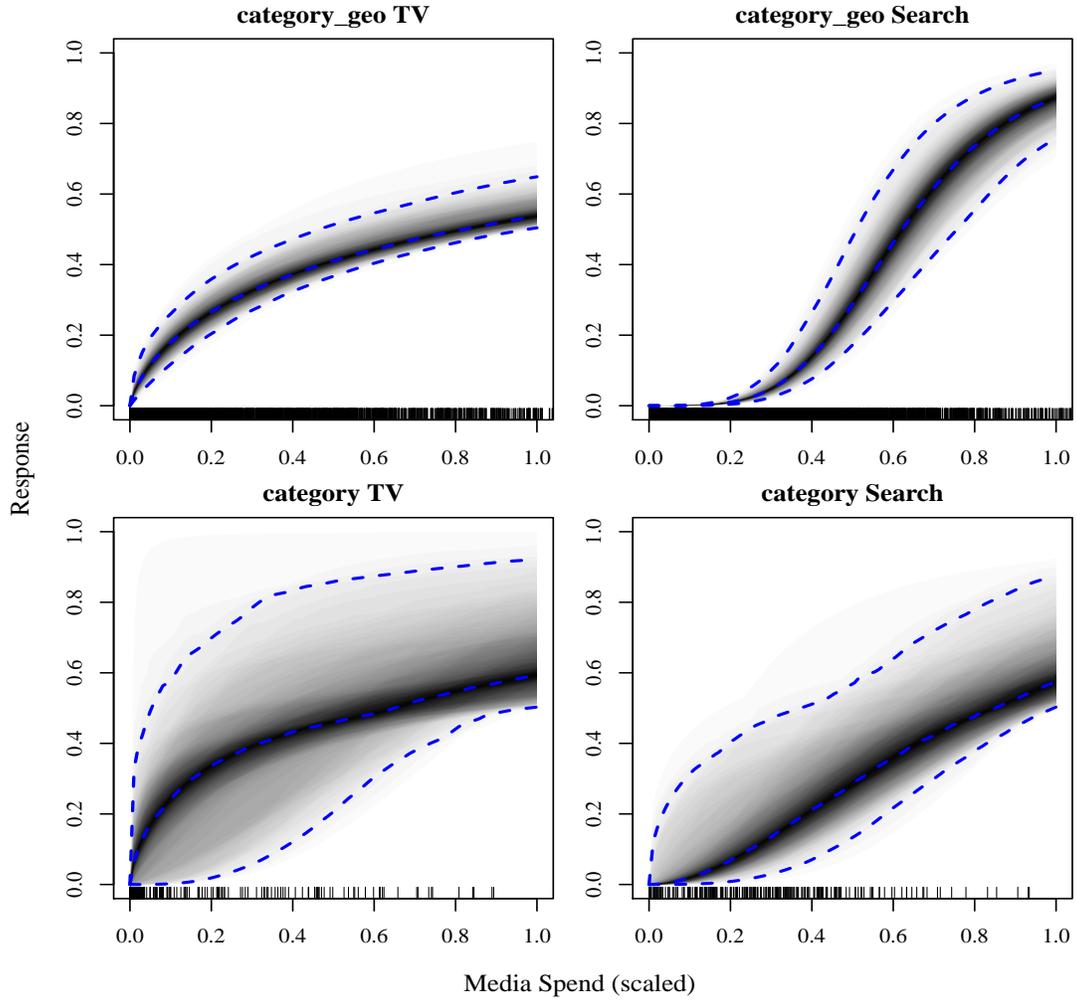


Figure 13: Posterior samples of the Hill functions. The X-axis is the scaled media variable ranging from zero to one and the Y-axis represents the Hill transformation taking values at the X-axis. The tick marks along the X-axis shows the values of scaled media spend.

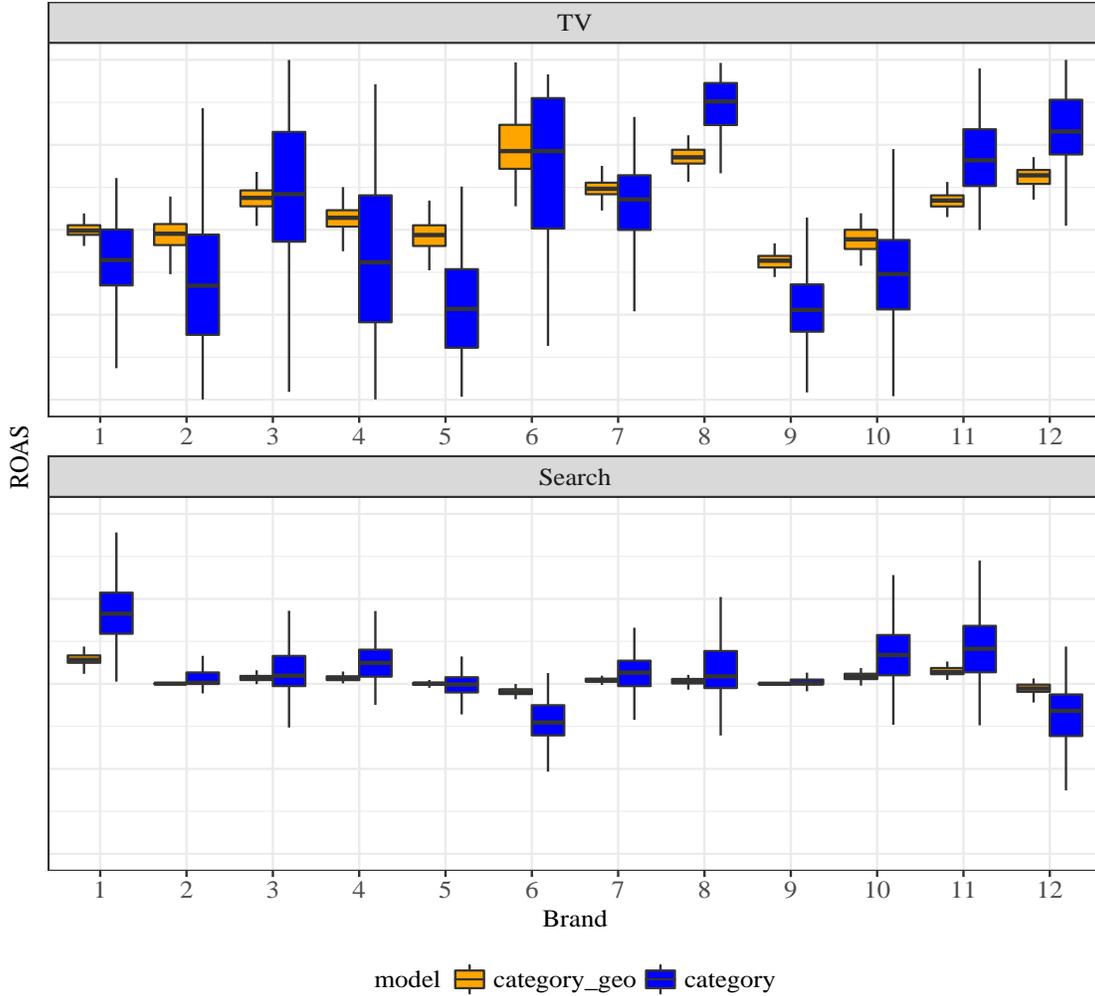


Figure 14: Boxplot of the brand ROAS estimates at the national level. The X-axis represents the 12 brands and Y-axis represents the estimated ROAS values from the category-level model (blue) and the category-geo-level model (orange).

6 Discussion

Media mix data are usually aggregated at a national level, which suffers from small sample size and insufficient variation with a large amount of model parameters. We develop a Bayesian hierarchical model incorporating regional variations to enhance media mix modeling. Simulations and a real example from the auto category demonstrate that the geo level Bayesian hierarchical model (GBHMMM) helps reduce uncertainty of estimators compared to the Bayesian media mix model (BMMM) using national level data alone, and thus yields more accurate attribution results. As the target population size varies across geos, standardization of the variables to be the amount per capita is necessary to employ the GBHMMM properly. The improvement over the BMMM is more pronounced when the level of media spend per capita varies significantly across geos. Although the geo-level model can't eliminate ad targeting bias, it generally tightens the credible intervals of the estimators and sometimes lowers the bias. Another common challenge is that regional level

spend data is not always available for all media channels. Using the national level media spend to impute the geo level spend generally leads to worse performance and we want to emphasize the importance of obtaining accurate geo-level data. However, the imputed geo-level model could still generate useful information for the media channels with complete geo-level data. Hence it is better to build a geo-level model when the geo-level data is available, or partially available, as long as the geos have similar media impact mechanism.

The model specification in this paper is not the only way to set up a geo-level model. Other functions to model shape and carryover effect could be used instead of the Hill and geometric adstock functions. The hierarchical structure of the model may not be necessary for all the predictors. Bayesian model selection techniques such as WAIC can be employed to choose a model.

The geo level model can be extended to help correct ad targeting bias in MMM, methods of which have been developed at the national level in Chen et al. (2017). The simulations in this paper were simple by design. Zhang and Vaver (2017) introduces a multi-stage simulator which models a wider variety of marketing situations and can be used to further evaluate the geo level model.

Acknowledgement

We want to thank Michael Perry, Stephanie Zhang, Zhe Chen, Aiyou Chen, Wiesner Vos, Bob Bell, Jon Vaver, Steve Scott, Penny Chu, Tony Fagan, Conor Sontag, Xiaojing Huang, Luis Gonzalez Perez and Shi Zhong for constructive feedback and support.

7 Appendix

7.1 Standardization of variables

We typically rescale a media variable to be relative to its minimum and maximum across times and geos,

$$x_{t,m,g} = \left(x_{t,m,g} - \min_{t,g}(x_{t,m,g}) \right) / \left(\max_{t,g}(x_{t,m,g}) - \min_{t,g}(x_{t,m,g}) \right).$$

The relative ordering of a media variable across geos are kept the same because the normalization function is the same across geos for each media channel. Note that this is not the only way to normalize media variables. Alternatively, if all the media variables are denominated in dollars, we could take the sum of total media spend in each time period, $x_{t,\cdot,g} = \sum_{m=1}^M x_{t,m,g}$ and rescale media variables relative to the range of total weekly ad spend or exposure across times and geos.

$$x_{t,m,g} = \left(x_{t,m,g} - \min_{t,g}(x_{t,\cdot,g}) \right) / \left(\max_{t,g}(x_{t,\cdot,g}) - \min_{t,g}(x_{t,\cdot,g}) \right).$$

We denote the transformation on the media variable of the m -th channel as $F_{x,m}(\cdot)$ for $m = 1, \dots, M$. No matter what transformations we apply to the response and media variables, we have to apply the inverse transformations to the model estimates afterwards.

We also recommend centering control variables within each geo and even scaling them. For example, the average household income may vary across geos, and as a result advertisers may set different baseline product prices in geos according to the income levels.

$$z_{t,c,g} = \zeta_{c,g} + \nu_{t,c,g}$$

where $\nu_{t,c,g}$ is white noise and $\zeta_{c,g}$ is the baseline product price in geo g . Without centering $z_{t,c,g}$, the Bayesian estimate of the coefficient of $z_{t,c,g}$ would be highly correlated with the intercept of geo g . Also when both household income and product price serve as control variables, their coefficients would be correlated. On the other hand, with centering, the model would focus on explaining the impact of change in control variables within each geo and leave the difference of price across geos to the intercept. The centering and scaling also facilitate the use of common priors on the coefficients of the control variables. However, these transformations alter the meaning of the coefficients, so modellers should decide on a case by case basis what transformation is needed for the control variables.

7.2 Derivation of the geo level correlation

Let the geo-level media spend be $x_{t,g}$ and the base demand be $h_{t,g}$,

$$x_{t,g} = u_g + \rho_g h_{t,g} + \sqrt{1 - \rho_g^2} \nu_{t,g}$$

where $\rho_g = \text{Cor}(x_{t,g}, h_{t,g})$ is the correlation between the media spend and the base demand. We assume that the geos are of equal size but the conclusion could be extended to the case that geos are of different sizes. As $u_g, g = 1, \dots, G$ are fixed, without loss of generality, we could set $u_g = 0$

for $g = 1, \dots, G$ and $E(h_{t,g}) = 0$. We first consider a simple case that the geo-level base demand time series are the same $h_{t,g} = h_t$, $\text{Var}(h_t) = 1$ and the correlations $\rho_g = \rho, g = 1, \dots, G$ are the same across geos.

As the geos are of equal size and variables are in the amount per capita, the national-level variables are simply averages of the geo-level variables, i.e.,

$$x_t = \sum_{g=1}^G x_{t,g} = \rho h_t + \sqrt{1 - \rho^2} \frac{1}{G} \sum_{g=1}^G v_{t,g}$$

where h_t is independent of the geo specific factor $v_{t,g}$. The correlation between the national-level media spend x_t and the base demand h_t is

$$\begin{aligned} \text{Cor}(x_t, h_t) &= \frac{\text{Cov}(x_t, h_t)}{\sqrt{\text{Var}(h_t)\text{Var}(x_t)}} \\ &= \frac{\rho \text{Var}(h_t)}{\sqrt{\text{Var}(h_t)(\rho^2 \text{Var}(h_t) + (1 - \rho^2) \frac{1}{G})}} \\ &= \frac{\rho}{\sqrt{\rho^2 + \frac{1 - \rho^2}{G \times \text{Var}(h_t)}}}. \end{aligned}$$

As $\text{Var}(h_t) = 1$, $\text{Cor}(x_t, h_t) = \frac{\rho}{\sqrt{\rho^2 + \frac{1 - \rho^2}{G}}} \geq \rho$, for $G \geq 2$, i.e., the national level correlation between the media spend and the base demand is higher than that at the geo level.

Now let's generalize to the case that $h_{t,g}$ are not the same but positively correlated, i.e., $\text{Var}(h_{t,g}) = 1, g = 1, \dots, G$ and $\text{Cor}(h_{t,m}, h_{t,l}) \geq 0, m \neq l$, at least one pair of geos have strictly positive correlation.

$$\begin{aligned} \text{Var}(h_t) &= \text{Var}\left(\frac{1}{G} \sum_{g=1}^G h_{t,g}\right) \\ &= \frac{\sum_{l,m} \text{Cov}(h_{t,l}, h_{t,m})}{G^2} \\ &> \frac{\sum_{g=1}^G \text{Var}(h_{t,g})}{G^2} \\ &= \frac{1}{G}. \end{aligned}$$

Hence $\text{Cor}(x_t, h_t) > \frac{\rho}{\sqrt{\rho^2 + (1 - \rho^2)}} = \rho$, i.e., the national level correlation between the media spend and the base demand is higher than that at the geo level.

References

- Cain, P. M. (2005). Modelling and forecasting brand share: a dynamic demand system approach. *International Journal of Research in Marketing*, 22(2), 203–220. doi:dx.doi.org/10.1016/j.ijresmar.2004.08.002
- Chan, D. & Perry, M. (2017). Challenges and opportunities in media mix modeling. *research.google.com*.
- Chen, A., Chan, D., Perry, M., Jin, Y., Sun, Y., Wang, Y. & Koehler, J. (2017). Bias correction for paid search in media mix modeling. *Forthcoming on <https://research.google.com>*.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515–534. doi:10.1214/06-BA117A
- Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge University Press.
- Gelman, A., Lee, D. & Guo, J. (2015). Stan: a probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavior Science*, 40(5), 530–543. doi:10.3102/1076998615606113
- Gelman, A. & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, 48(2), 241–251. doi:10.1198/004017005000000517
- Jastram, R. W. (1955). A treatment of distributed lags in the theory of advertising expenditure. *Journal of Marketing*, 20(1), 36–46. doi:10.2307/1248159
- Jin, Y., Wang, Y., Sun, Y., Chan, D. & Koehler, J. (2017). Bayesian methods for media mix modeling with carryover and shape effects. *research.google.com*.
- Lewis, R. A. & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics*, 130(4), 1941–1973. doi:10.1093/qje/qjv023
- Little, J. D. C. (1979). Aggregate advertising models: the state of the art. *Operations Research*, 27(4), 629–667.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3), 705–767. doi:10.1214/aos/1056562461
- Palda, K. S. (1965). The measurement of cumulative advertising effects. *The Journal of Business*, 38(2), 162–179. doi:10.1086/294759
- R Core Team. (2015). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rossi, P. E., Allenby, G. M. & McCulloch, R. (2005). *Bayesian statistics and marketing*. Chichester, England: Wiley-Interscience.
- Tellis, G. J. (2006). Modeling marketing mix. In R. Grover & M. Vriens (Eds.), *Handbook of marketing research: uses, misuses, and future advances* (pp. 506–522). Thousand Oaks, CA. doi:10.4135/9781412973380.n24
- Vaver, J. & Koehler, J. (2011). Measuring ad effectiveness using geo experiments. *research.google.com*. Retrieved from <https://research.google.com/pubs/pub38355.html>

- Wang, Y., Jin, Y., Sun, Y., Chan, D. & Koehler, J. (2017). A hierarchical bayesian approach to improve media mix models using category data. *research.google.com*.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
- Zhang, S. S. & Vaver, J. (2017). Introduction to the Aggregate Marketing System Simulator. *research.google.com*.