

# A Hierarchical Bayesian Approach to Improve Media Mix Models Using Category Data

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

Google Inc.

7th April 2017

## Abstract

One of the major problems in developing media mix models is that the data that is generally available to the modeler lacks sufficient quantity and information content to reliably estimate the parameters in a model of even moderate complexity. Pooling data from different brands within the same product category provides more observations and greater variability in media spend patterns. We either directly use the results from a hierarchical Bayesian model built on the category dataset, or pass the information learned from the category model to a brand-specific media mix model via informative priors within a Bayesian framework, depending on the data sharing restriction across brands. We demonstrate using both simulation and real case studies that our category analysis can improve parameter estimation and reduce uncertainty of model prediction and extrapolation.

## 1 Motivation

Media Mix Models (MMMs) are widely used as the basis for understanding the effect different media types have on Key Performance Indices (KPIs, e.g. sales), as well as for optimizing media mix for maximal KPI. Modelers face several challenges when developing MMMs (Chan & Perry, 2017; Jin, Wang, Sun, Chan, & Koehler, 2017; Lewis & Rao, 2015; Quandt, 1964). One of the most critical is the lack of data with sufficient information to adequately estimate a model with the many parameters required to account for the all media types modern advertisers utilize.

Media mix studies are typically based on nationally aggregated weekly data over 2 to 5 years or monthly data over 4 to 6 years resulting in about 50 to 250 observations (Dekimpe & Hanssens, 2000). Even if data is available for a longer duration, e.g., more than 10 years, it is not desirable as the market dynamics could have shifted drastically during that time. The above restrictions of data availability and relevancy leaves a very modestly sized dataset for developing an MMM, given that the number of media types involved can be as many as 20 or more, and that media effects tend to be significantly smaller than some non-media factors, such as price or retailer promotion (e.g., retailer feature or special display of the products).

Moreover, advertisers often align their media spending with the underlying seasonality of their products' or brands' historically established promotion cycles. Advertisers also tend to resist significantly varying their spend from historic patterns due to planning inertia, lack of quantitative

knowledge on media’s true effects, and an aversion to risk of mis-spending. Such advertising behaviors often leads to highly correlated observations of media spend that vary within a relatively small range, making it almost impossible to distinguish each media’s impact on KPIs, manifesting as large model estimation and extrapolation uncertainties. Chan and Perry (2017) provides a discussion on the current challenges of MMMs.

One approach to address the problem of data sparsity is to inject variability through randomized experiments (Blake, Nosko, & Tadelis, 2015; Lewis & Rao, 2015), in particular geo experiments (Vaver & Koehler, 2011). They involve assigning non-overlapping geographic regions to either a control or treatment group, modifying advertising level on a certain media channel in the treatment region, and measuring the KPI against the unchanged control regions. Given the complex structures of media spending, however, not all advertisers can afford or would want to run experiments across all of their adopted media channels for sufficiently long time periods.

If experiments are expensive and difficult to scale, can we collect observational data with sufficient natural variability to measure the effectiveness of ads? Several research works have provided possible solutions, including observing “a natural experiment” over special events (Stephens-Davidowitz, Varian, & Smith, 2017), or using data collected at a finer geographic granularity (Sun, Wang, Jin, Chan, & Koehler, 2017).

In this paper, we propose pooling datasets from multiple brands within a product category to form a joint dataset with more independent variation and a wider range of media spend and potentially control factors, e.g., product price. We adopt the Bayesian framework described in Jin et al. (2017) and further use a hierarchical Bayesian model to represent the category-brand relationship (Gelman & Hill, 2006). In the media mix literature, there have been many different model frameworks proposed, for example, see Chan and Perry (2017), Little (1979), Quandt (1964). We think the benefits of pooling multiple brand datasets, and passing information learned from a category to a brand, is useful regardless of the specific model family used to describe the data.

Our method assumes media effects on sales are similar across the brands (in a manner discussed in more detail in Section 3) within the category. Therefore, media similarity should be used as a guideline for conducting category analysis on a group of entities, assessed from data and subject knowledge. If a group of entities have substantially diverse media responses, our method would not be appropriate. Instead, a subset of these entities among which media responses are similar can be considered. We discuss two examples of pooling multiple brands within the same Consumer Packaged Goods (CPG) category, i.e. shampoo (Section 6) and soda (Section 7). Such pooled datasets can be hard to acquire for individual advertisers due to cost, but may be accessible for some third-party MMM vendors.

When the assumption of similar media response within the category is reasonable, we hope to reduce the high correlation between various media variables (sometimes including non-media variables) that often exist within a single brand, by observing them along with the response variable (e.g., sales) over multiple brands in the category. We expect more variation in the data, because different brands do not always share the same targeting preferences and historical patterns in their media expenditure.

Category analysis also provides an increase in the number of observations available for the model. In the case studies included in later sections, the category datasets have more than ten times the number of observations than a single brand. An increase in sample size and improved variability gives us with a better chance at understanding the effects of media on sales.

One can also leverage the pooled datasets for a straightforward approximation of the underlying demand of a product category, in terms of seasonality and long-term trend, using the total sales of all brands within the category. It is unlikely these brands share the same promotional activities, and thus their total sales represent total demand for a category of products, rather than the brand’s individual traits.

One more advantage of category analysis is the ability to incorporate competitive factors - impact across brands - into MMMs. Developing MMMs for a single brand can suffer from omitted variables, of which competitive factors, such as competitor price and promotion, are common ones (Borden, 1964; Ehrenberg & Barnard, 2000; Quandt, 1964). In category analysis, impact from competitor activities on a brand of interest can be explicitly included in the model to help reduce bias in parameter estimates.

One barrier to directly using the MMM results based on the category datasets is data sharing restrictions among different and often competing advertisers. Often, an advertiser would allow its data to participate in establishing a category benchmark, but not to derive any brand-specific results other than for its own brand. In fact, such restriction is the most common based on our experience interacting with advertisers and MMM vendors. Where restrictions apply, we propose condensing the information learned from a category dataset into Bayesian informative priors, which can then be used in a brand-specific MMM without directly accessing other brands’ data.

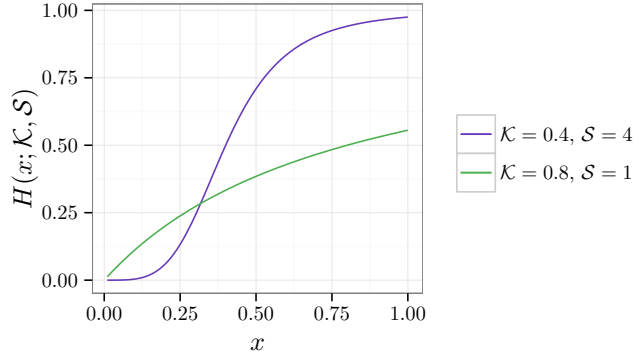
The remainder of the paper is structured as follows. Section 2 introduces a framework of MMMs, which we use as an example to demonstrate the advantages of pooling multiple brand datasets in an MMM study. Sections 3 and 4 discuss a hierarchical Bayesian model using a pooled dataset, extracting informative priors from the category results and utilizing the priors in brand-specific models. Section 5 discusses analysis and model comparisons in four simulation scenarios. Section 6 and 7 apply our method on two real case categories and compare brand MMMs using informative priors derived from the category versus weak priors. Summary comments and discussion are included in Section 8.

## 2 A basic media mix model

We begin by outlining a media mix model for a single brand, building upon the model form introduced by Jin et al. (2017).

For time  $t = 1, \dots, T$ , we use the notation  $y_t$  for the time series of the response variable. In MMMs,  $y_t$  is usually a type of KPI, such as dollar sales or volume sales. We use the notation  $x_{t,m}$  to denote the media variable for media channel  $m$  at time  $t$ , for  $m = 1, \dots, M$ , e.g.,  $x_{t,m}$ , could be the advertising spend or Gross Rating Points (GRPs) of TV ads, advertising spend or number of impressions of online display ads over a week period, etc. Lastly, we use the notation  $z_{t,c}$ ,  $c = 1, \dots, C$ , for control variables. Common control variables include the product price, All Commodity Volume (ACV) weighted distribution of product and retailer promotion, weather, average competitor price, etc. It is up to each modeler to choose relevant control variables, depending on the business structure of the category and the influence of the control variable on the response.

Instead of a linear relationship between response  $y_t$  and  $x_{t,m}$ , MMMs often incorporate flexibility to account for the nonlinear aspects of media effects on KPI, of which three main ones include the carryover effect of media exposures, diminishing returns of media investment and the necessity to build a certain level of awareness before significant returns from media spend realized. The first one



**Figure 1:** Illustration of Hill transformation under two sets of Hill parameters.

is also referred to as the lag structure (or adstock) of media impact, while the latter two combined to form the shape or curvature (an “S” curve) of sales response to media exposure (Little, 1979; Tellis, 2006).

Several functional forms to account for media carryover effects have been proposed, such as geometric adstock, delayed adstock, or polynomial distributed lags (Jastram, 1955; Jin et al., 2017; Palda, 1965). As for the shape effect, the log transformation, Hill transformation, or logistic growth function is sometimes used to capture media’s diminishing returns (Cain, 2005; Jin et al., 2017; Little, 1979).

In this paper, we follow the prior literature by Jin et al. (2017) to use geometric adstock and Hill transformation for media carryover and shape effect respectively, as well as values of some fixed parameters and choices of weak priors.

The geometric adstock function is defined as,

$$GA(x_t; \alpha, L) = \frac{\sum_{l=0}^L x_{t-l} \alpha^l}{\sum_{l=0}^L \alpha^l} \quad (1)$$

where the carryover rate  $\alpha \in (0, 1)$ . For simplicity, we follow the example in Jin et al. (2017) and set the length of carryover effects at 13 weeks for all media channels. Note that the denominator in the above definition makes the output of the function lie within the range of  $x$ .

The Hill transformation function is defined as,

$$H(x; \mathcal{K}, \mathcal{S}) = \frac{1}{1 + \left(\frac{x}{\mathcal{K}}\right)^{-\mathcal{S}}}$$

where  $\mathcal{K} > 0$  and  $\mathcal{S} > 0$ . The Hill equation originates from quantitative pharmacology (Gesztelyi et al., 2012). It maps the real line to  $(0, 1)$  and reaches  $1/2$ , the half saturation point, when  $x = \mathcal{K}$ . Thus, the parameter  $\mathcal{K}$  is often referred to as the half maximal Effective Concentration ( $EC$  or  $EC_{50}$ ). The parameter  $\mathcal{S}$  is also known as the Hill coefficient, interpreted as the largest absolute value of the slope of the curve. Figure 1 illustrates two example shapes representable by the Hill transformation: the ‘S’ curve and diminishing-returns-only. Jin et al. (2017) discusses the usage of Hill transformation in more detail.

To simplify the specification of weak priors for  $\mathcal{K}$  across different media, we first scale media

variables to be between 0 and 1,

$$x_{t,m} = \frac{x_{t,m}^{(0)} - \min_t(x_{t,m}^{(0)})}{\max_t(x_{t,m}^{(0)}) - \min_t(x_{t,m}^{(0)})}$$

where  $x_{t,m}^{(0)}$  denotes the original media variable of channel  $m$  at time  $t$ . In this paper, we restrict  $\mathcal{K}$  to between 0 and 1 in order to avoid non-identifiability of the model (Jin et al., 2017) and to achieve faster convergence.

Then, a basic MMM that allows for geometric carryover effects and a flexible shape structure, can be written as Jin et al. (2017):

$$\Theta(y_t) = \tau_0 + \sum_{m=1}^M \beta_m h_m(x_{t,m}) + \sum_{c=1}^C \gamma_c z_{t,c} + \epsilon_t,$$

where

$$h_m(x) = \text{H}(\text{GA}(x; \alpha_m, L_m); \mathcal{K}_m, \mathcal{S}_m) \quad (2)$$

and  $\epsilon_t \sim \text{Normal}(0, \sigma^2)$  for time  $t = 1, \dots, T^1$ . Common choices of transformation  $\Theta$  on the response variable include the identity and the logarithm function. The specific choice depends on the distribution of the response variable.

The above model specification adds 5 parameters (or 4, if  $L_m$  is preset) for every media variable included in the model. Estimating these parameters is non-trivial given that MMMs are often based on weekly observations of a single brand over less than 5 years. Due to the lack of quantity and information content in MMM datasets relative to model complexity, media variable coefficients are often estimated as insignificant (wide confidence intervals), significantly negative, or too large to be true.

### 3 Deriving informative priors using category data

In this section, we first introduce our Bayesian hierarchical model which utilizes data from all brands within the category. In the next section, we discuss how to derive informative priors from the category results and use them in an MMM for a specific brand of interest.

We first introduce notation for multiple brands within the category. We use  $\tau_b$  to denote brand-specific intercepts. Let  $x_{t,m,b}$  indicate the spend or exposure variable of media channel  $m$  for brand  $b$  at time  $t$ , and  $z_{t,c,b}$  represent control variable  $c$  for brand  $b$  at time  $t$ . We use  $\beta_{m,b}$ ,  $m = 1, \dots, M$  to denote brand-specific coefficients for the  $M$  media variables, and  $\gamma_{c,b}$ ,  $c = 1, \dots, C$  to denote brand-specific coefficients for the  $C$  control variables. For  $b = 1, \dots, B$ , brand-specific coefficients for the same variable share a category-wide prior distribution:

$$\begin{aligned} \beta_{m,b} &\sim \text{Normal}^+(\beta_m, \eta_m^2), & m = 1, \dots, M \\ \gamma_{c,b} &\sim \text{Normal}(\gamma_c, \xi_c^2), & c = 1, \dots, C. \end{aligned} \quad (3)$$

---

<sup>1</sup>Sometimes interaction terms can or should be added to the model to represent combined impact of media variables, or between media and control variables. For simplicity and a focus on the benefits of category analysis, we leave out interaction terms in our models in this paper.

Weak or noninformative hyperpriors can then be specified for the category hyperparameters  $\{\beta_m, \eta_m\}$  and  $\{\gamma_c, \xi_c\}$ .

It can be counterintuitive to observe a media channel<sup>2</sup> on which more spend would lead to less sales. Sometimes models with unconstrained priors output negative media effect estimates due to omitted variables, rather than a negatively influencing media channel. In this paper, we work with non-negative weak priors on  $\beta_m$  and  $\beta_{m,b}$ , as a representation of advertisers expectation of a non-negative incremental impact of media effects. Non-negative priors are not the only reasonable choice of prior; in fact, Jin et al. (2017) explores several reasonable priors with slightly different assumptions for media parameters. For the other media parameters such as  $\mathcal{K}_m$  and  $\mathcal{S}_m$ , we follow the weak priors used in Jin et al. (2017).

Similar to the basic MMM, we need to scale the media variables to between 0 and 1 to be consistent with the support of  $\mathcal{K}$  parameter in the Hill transformation. In category analysis, we scale across all brand datasets for each of the  $M$  media channels. In particular, the scaled media variables are obtained as follows,

$$x_{t,m,b} = \frac{x_{t,m,b}^{(0)} - \min_{t,b}(x_{t,m,b}^{(0)})}{\max_{t,b}(x_{t,m,b}^{(0)}) - \min_{t,b}(x_{t,m,b}^{(0)})} \quad (4)$$

where  $x_{t,m,b}^{(0)}$  indicates the original media variable. Scaling within a media channel but across all brands provides us a possibility to observe different sections of the media spend spectrum, which in turn helps us better estimate the shape of media response<sup>3</sup>.

Then for brand  $b = 1, \dots, B$ , at time  $t = 1, \dots, T$ , the hierarchical category MMM (HCM) can be written as,

$$\Theta(y_{t,b}) \sim Normal(\mu_{t,b}, \sigma^2) \quad (5)$$

where

$$\mu_{t,b} = \tau_b + \sum_{m=1}^M \beta_{m,b} h_m(x_{t,m,b}) + \sum_{c=1}^C \gamma_{c,b} z_{t,c,b}. \quad (6)$$

In (6), the Hill transformation parameters  $\mathcal{K}_m$  and  $\mathcal{S}_m$  are shared across brands in the category for each  $m$ , while  $\beta$  and  $\gamma$ , as described in (3), are brand-specific but follow a same category-wide distribution. The above model specification is one of many possibilities; it represents the modeler’s prior knowledge that a certain level of similarity exists in media effects across brands within the category, namely, similar shape and carryover but potentially varying magnitude. There can be situations where one would specify the model to allow for varying shape effects, for example. We discuss more on this topic in Section 5.3.

The model described in (3) and (5) is a standard and widely used Bayesian hierarchical model. Its fundamental idea is to approach a complex problem by breaking it into smaller parts through decomposing the joint distribution of a set of random variables into a series of conditional models. Bayesian hierarchical models have been applied to many research areas, such as marketing (Rossi, Allenby, & McCulloch, 2005), data processing (Li & Perona, 2005), genetics (Foll & Gaggiotti, 2008, 2), speech recognition (Yildiz, von Kriegstein, & Kiebel, 2013, 9), etc.

<sup>2</sup>A negative impact can be observed for a certain advertising campaign for a certain brand. But an overall negative impact for a media channel over multiple years and numerous campaigns is quite rare.

<sup>3</sup>Note that the above scaling induces a data-dependency between the prior on  $\mathcal{K}$  and the actual spend levels observed in the data which may not be desirable unless it reflects an actual set of background knowledge about the observed spend levels.

## 4 Utilizing informative priors for a single brand

When there are no restrictions on sharing data across brands, one can skip the steps described in this section and directly use the brand-specific parameters estimated by the HCM. For example, when a category model is built using all the brands of a similar product category owned by one company.

If the category dataset is not available to generate brand-specific MMM results directly, we propose an alternative approach of building a brand-specific MMM using only the data of the brand of interest and informative Bayesian priors in the form of posterior samples of the category-wide (hyper)parameters extracted from the category model. This way, the joint distribution of the category-wide media parameters learned from the HCM is preserved. At the same time, these parameters are not specific to any particular brand and thus anonymous to a certain extent.

In particular, for brand  $b^*$  within the category of interest, a brand-specific model can be described as follows. For time  $t$ ,  $t = 1, \dots, T$ ,

$$\Theta(y_{t,b^*}) \sim \text{Normal}(\mu_t^*, \sigma^2) \quad (7)$$

where,

$$\mu_t^* = \tau^* + \sum_{m=1}^M \beta_m^* h_m(x_{t,m,b^*}) + \sum_{c=1}^C \gamma_c^* z_{t,c,b^*},$$

and

$$\beta_m^* \sim \text{Normal}^+(\beta_m, \eta_m^2).$$

We use  $\{\Phi_m\}^{(c)}$  to represent the posterior samples from the category (c) model, where  $\Phi_m = \{\beta_m, \eta_m, \alpha_m, \mathcal{K}_m, \mathcal{S}_m\}$ . In each MCMC iteration of the brand-specific model, instead of estimating  $\Phi_m$  using only the brand dataset, one randomly draws a sample from the joint empirical distribution approximated by  $\{\Phi_m\}^{(c)}$ . By incorporating informative priors in the format of  $\{\Phi_m\}^{(c)}$  in the brand-specific MMMs, we can preserve maximal information inherited from the category model to be passed onto brand models, while maintaining a certain level of anonymity for individual brands' datasets. It is important to extract the joint posterior of  $\Phi_m$  from the HCM, instead of the marginals. As Jin et al. (2017) pointed out, the media parameters, especially  $\beta_m$ ,  $\mathcal{K}_m$ , and  $\mathcal{S}_m$ , are often highly correlated as they can trade off each other to represent similar media responses. In fact, in the above proposed sampling approach for a brand-specific model, the information exchange between the brand-specific parameters and  $\{\Phi_m\}^{(c)}$  is equivalent to that between the brand-specific parameters and  $\Phi_m$  in the HCM. Therefore, the parameter estimates from a brand-specific model using informative priors in the format of  $\{\Phi_m\}^{(c)}$  are equivalent to those of the brand-specific parameters directly from the HCM.

Besides the posterior samples of the category-wide parameters  $\Phi_m$ , the range of media variables of the category needs to be passed from category to brand analysis, in order to maintain the same scaling transformation in the brand MMM as the category MMM in (4), which is essential for the informative priors derived from category model to be meaningful to the brand-specific models. The minimum values of media variables of a category are usually 0, while the maximum values are often not sensitive data and can be shared.

If an advertiser prohibits its data from participating in developing the HCM, a category model can be built using other brands in the category whose datasets are accessible. The resulted informative

priors can be used in a similar manner as discussed above, as long as it is reasonable to assume the media responses of the brand is similar to the brands used in the category model.

We hope by demonstrating the benefits of a category analysis in this paper, advertisers are encouraged to relax data sharing restrictions to at least allow for usage in developing a category model.

## 5 Simulation studies

In this section, we illustrate some of the key benefits of our category analysis through simulation studies. Because MMMs are being developed and used in real practice for individual brands, we focus our attention on comparing the brand-specific models, using weak priors versus using informative priors represented by  $\{\Phi_m\}^{(c)}$ .

In order to implement the simulation scenarios discussed in this section, we introduce a sequential simulation process. It allows dependencies between covariates in addition to the dependency of the response variable on the covariates. For example, one might want to simulate different brands' media spends based on each brand's underlying media planning behaviors, which is further correlated with the product's underlying seasonality. A sequential simulation process enables us to fulfill a chain of dependencies as is described above. Appendix A contains more detail of the data simulation process.

The rest of this section is structured as follows. We first look at two common scenarios where category analysis can provide an advantage over a single brand MMM: Section 5.1 discusses the simulation scenario where media variables exhibit a larger variation across brands, compared to within a single brand; Section 5.2 investigates the scenario where the competitive factor among brands has a non-trivial impact on KPIs. We then demonstrate the importance of a category-brand hierarchy when non-trivial variation exists among brands in Section 5.3.

### 5.1 When across-brand media variation is larger than within-brand

We first discuss a scenario where pooling multiple datasets would potentially be the most beneficial: larger variation of media variables across different brands than within a brand. In this setting, shrinkage across brands helps estimate the shape parameters and model coefficients more accurately and enables us to extrapolate (to a certain extent) with more confidence.

We first introduce the data simulation specifications and model setup in Section 5.1.1. Then, Section 5.1.2 and 5.1.3 compares the brand-specific results in two perspectives (via estimations of ROAS and response curves) and discusses the differences in these two types of model performance metrics.

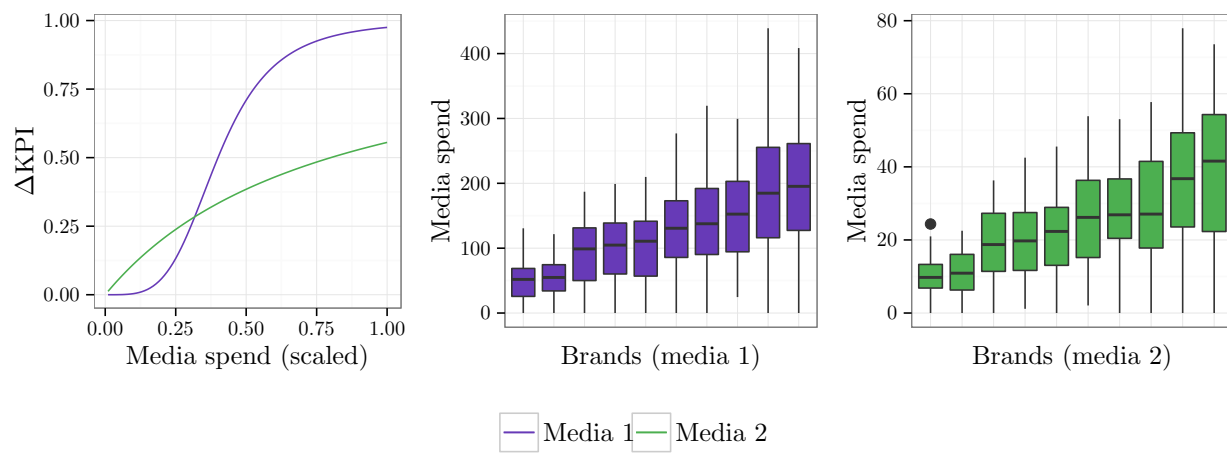
#### 5.1.1 Data simulation and model setup

For this simulation scenario, we simulated  $M = 2$  media channels with the specifications listed in Table 1. The Hill transformations under these specifications are illustrated in the left panel of Figure 2. We simulated 100 datasets, each of which contains weekly observations of 10 brands over



**Table 1:** Specification of media impact in simulation Section 5.1.

	Media 1	Media 2
$\mathcal{K}$	0.4	0.8
$\mathcal{S}$	4	1
Coefficient $\beta_{m,\cdot}$	$\beta_{1,\cdot} \sim N(0.5, 0.01^2)$	$\beta_{2,\cdot} \sim N(0.2, 0.01^2)$



**Figure 2:** Left panel: illustration of two simulated media shapes (left panel). Middle and right panels: simulated media spend on two channels across the 10 brands of an example dataset, ordered by the median spend of media 1.

104 weeks. For each simulated dataset, we ran 2000 iterations with 4 parallel MCMC chains using the *RStan* language (Stan Development Team, 2015).

The middle and right panels of Figure 2 illustrate the variation in media spend across brands in one example dataset. In both plots, brands are ordered by the median of media spend in media channel 1. The media spend in channel 2 roughly, though not exactly, follows the same order. It is because in our simulation, media spend is correlated with each brand’s base size: the bigger the company, the more media spend.

Compared to the category dataset, the small brands’ datasets individually would only correspond to the lower section of media’s response curve, while the big brands’ datasets could concentrate on the upper section of a media’s response curve. By pooling the brand datasets together, small brands can benefit from the observations from bigger brands to extrapolate with improved accuracy; and vice versa for bigger brands. There is still a limit on extrapolating from an estimated response curve, as accuracy may deteriorate quickly beyond what we have observed in the category.

For each simulated dataset, we first develop a category-level model specified in (3) and (5), using all simulated data of  $B = 10$  brands and  $T = 104$  weeks. For  $m = 1, 2$ , we use the following weak priors in the category model,

$$\mathcal{K}_m \sim \text{Beta}(2, 2), \mathcal{S}_m \sim \text{Gamma}(3, 1), \text{ and } \beta_m \sim \text{Uniform}(0, 5). \quad (8)$$

For demonstration purpose, we focus on media shape and coefficients; no carryover effects were simulated, and no lag structure was included in the models. In this and the following two simulation scenarios, we use a logarithm transformation on the response variable in both the data simulation and the models, i.e.  $\Theta(y) = \log(y)$ .

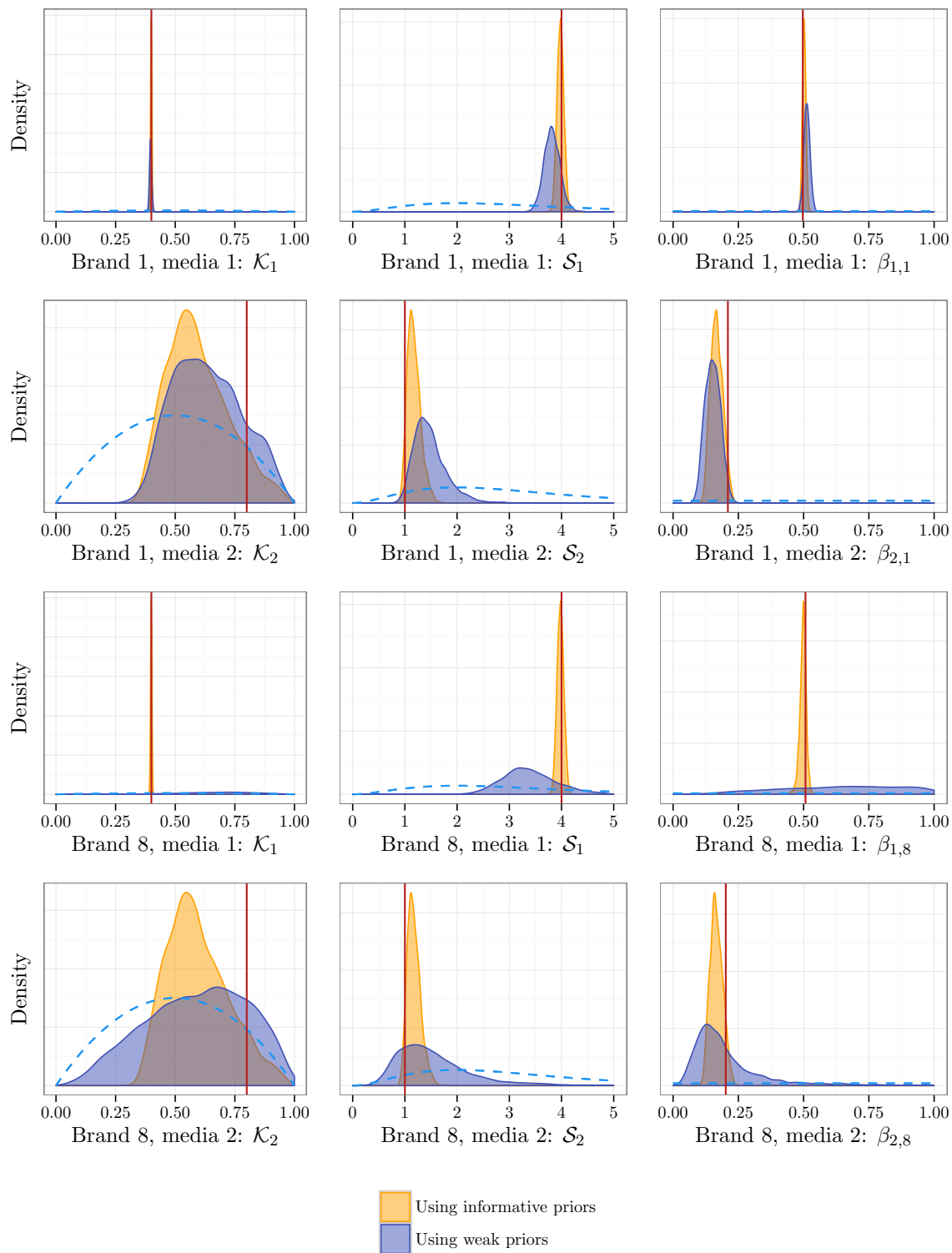
For each of the 100 simulations, we extract the joint posterior samples  $\{\Phi_m\}^{(c)}$  from the category model, to be passed onto the brand-specific MMMs.

### 5.1.2 Comparisons of ROAS and mROAS estimated by brand-specific models

In this section, we compare the results from brand models described in (7) using informative priors represented by  $\{\Phi_m\}^{(c)}$  with that using weak priors listed in (8). Taking one simulation dataset as an example, Figure 3 compares the marginal density of posterior samples of the media parameters estimated by the brand models with the true values (vertical red lines). We also plot the weak priors used (dashed light blue lines) for reference. To keep the figure readable, only two simulated brands are shown.

Several patterns are evident in Figure 3. In some cases, like  $\mathcal{K}_1$  of brand 1, the informative priors do not seem to make a difference, and both brand models estimate the parameter well. In some cases, e.g., the coefficients of both media for brand 8, the informative priors substantially reduce posterior uncertainty. In some cases, e.g.  $\mathcal{S}_1$  for both brand 1 and brand 8, the informative priors also improve the accuracy of the point estimate: the brand model with informative priors provides a more accurate estimate of  $\mathcal{S}_1$  with less uncertainty, compared to the brand model using weak priors. In other cases, both models show low estimation accuracy, e.g.  $\mathcal{S}_2$ .

The preceding example shows that informative priors derived from the category model can help improve the estimation accuracy and reduce uncertainty by passing on learnings obtained from a richer dataset. The compromise between data and priors is a standard example of the posterior



**Figure 3:** Marginal density plots of media parameters posterior samples estimated by brand models using informative priors (in gold), compared to those using weak priors (in blue), using an example dataset, as well as the true values (vertical red lines) and the weak priors used (dashed light blue lines).

distribution as a result of both the data and the priors: when the information in the data is weak, an informative prior has more influence on the posterior. Using the category dataset to derive informative priors can supplement the lack of information content of a single brand dataset.

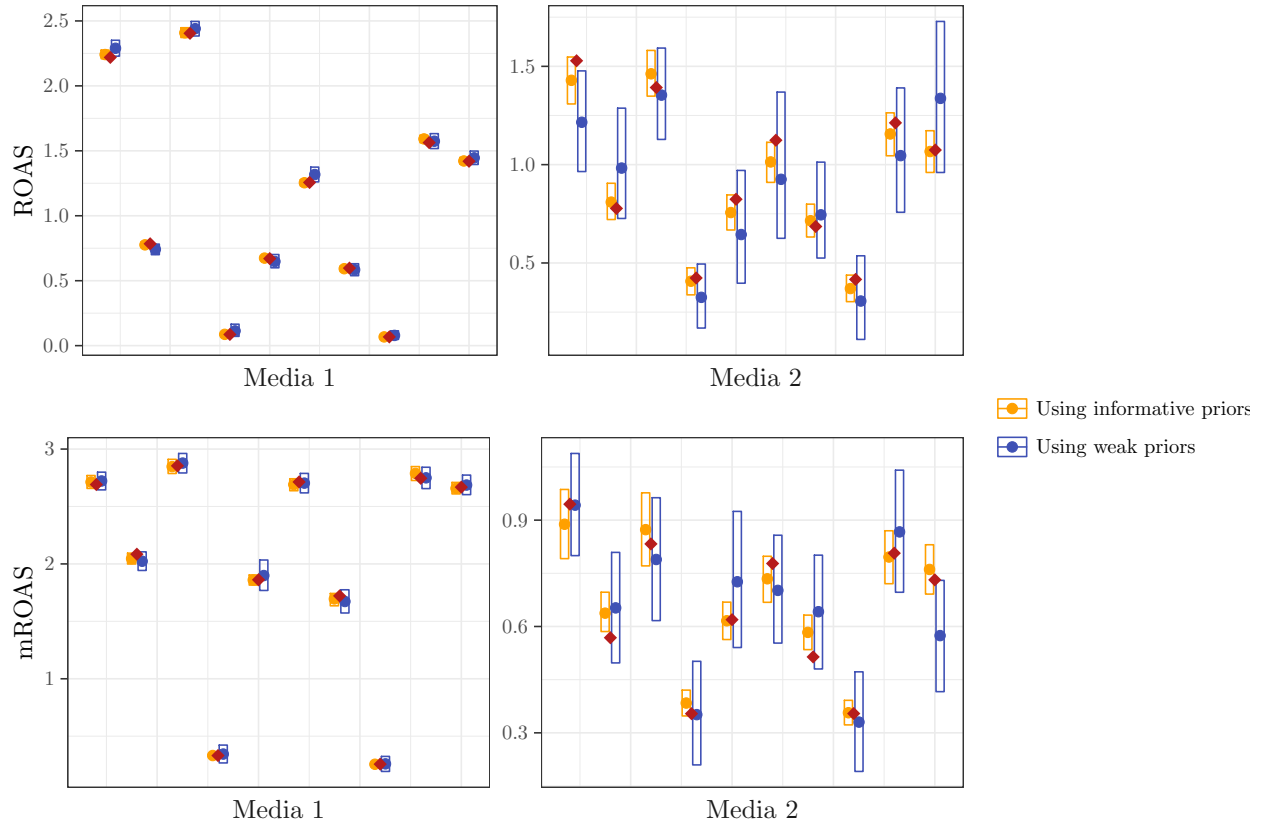
Further, when informative priors are used, the three parameters of the first media (first and third row of Figure 3) are estimated with narrow uncertainty and good accuracy, while those for the second media (in second and fourth row) have wider uncertainty with lower accuracy. The different estimation precision reflects different transformation parameter values used in the simulation. Setting  $\mathcal{K}_1 = 0.4$  indicates the saturation effect of media channel 1 is observed; setting  $\mathcal{K}_2 = 0.8$ , however, is equivalent to assuming we only observe a little more than the first half of the ‘‘S’’ curve. In such cases, the Hill transformation can be over-parameterized, resulting in more flexibility than the data can identify (Jin et al., 2017).

The brand-specific model estimates can also be compared in terms of (average) Return On Ad Spend (ROAS) and marginal Return On Ad Spend (mROAS), which are summary metrics familiar to advertisers. The precise definition of these metrics is given in Appendix B. One can calculate  $\text{ROAS}_{m,b,i}$  and  $\text{mROAS}_{m,b,i}$  for each media channel  $m$  and each brand  $b$ , estimated using the  $i$ th MCMC sample of the media parameters. The variation in these metrics can then be used to measure uncertainty. Figure 4 summarizes the ROAS metrics. For the 10 brands in the same example dataset as in Figure 2, they are calculated using samples drawn from the 2000 MCMC iterations (after warm-up). In general, the estimates are more accurate for media channel 1. The informative priors reduce the uncertainty in estimated ROAS and mROAS.

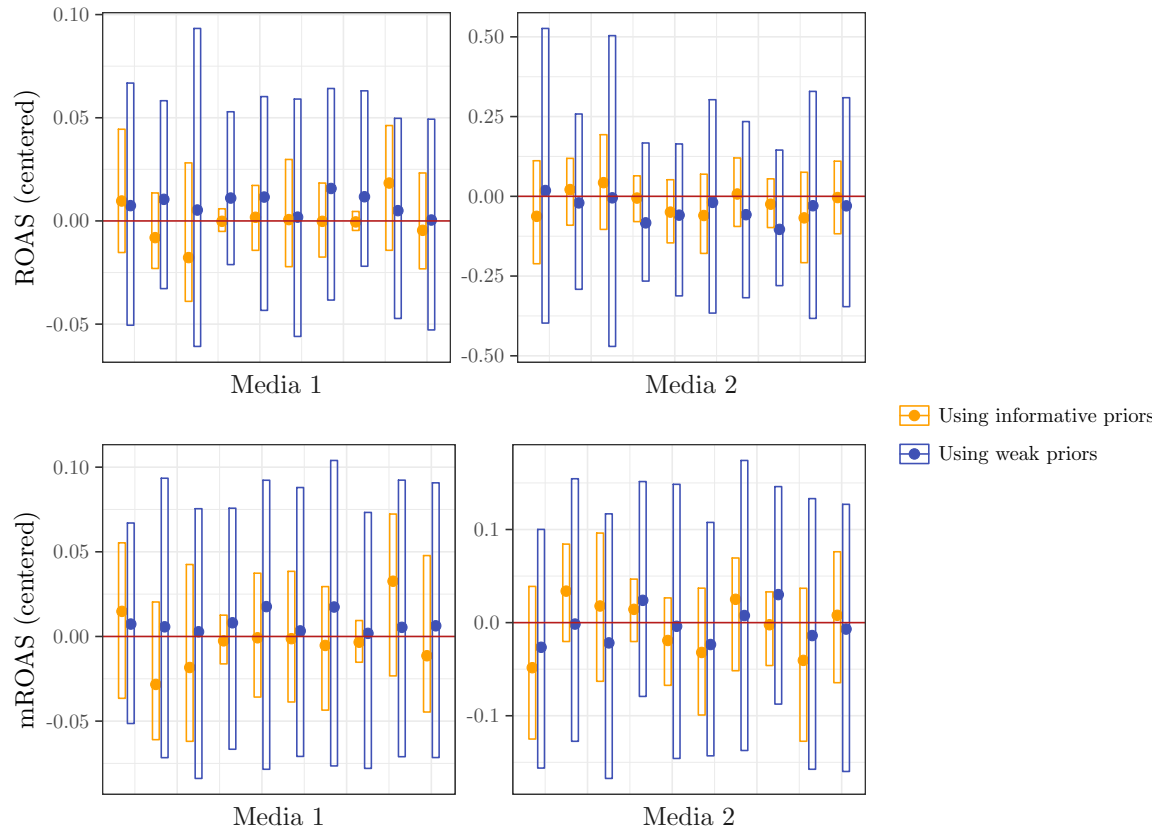
For media 1, the variation of ROAS and mROAS across brands is large compared to the variation within a single brand. By definition, ROAS measures the average performance of each media channel over its historical spend level and thus reflects the different media spend levels of the brands.

For an example of the impact of actual media spend on estimated ROAS, consider a small brand whose media spend mostly resides on the lower end of the media response curve. ROAS only measures the media channel performance restricted to the section of the response curve observed for this brand; it does not tell us any information on the accuracy of potential extrapolation. Meanwhile, if a brand has media spend level that varies substantially over time, we would expect ROAS calculated over different time periods to have large variation due to advertising that corresponds to different sections of the media response curve. Therefore, the variation in estimated ROAS caused by different levels of media spend can sometimes be confused with the variation introduced by model estimation. The other metric, mROAS, shares the same behavior, for it only measures the model performance induced by a small (1%) change in media variables and is partially influenced by the value at which the small change is applied.

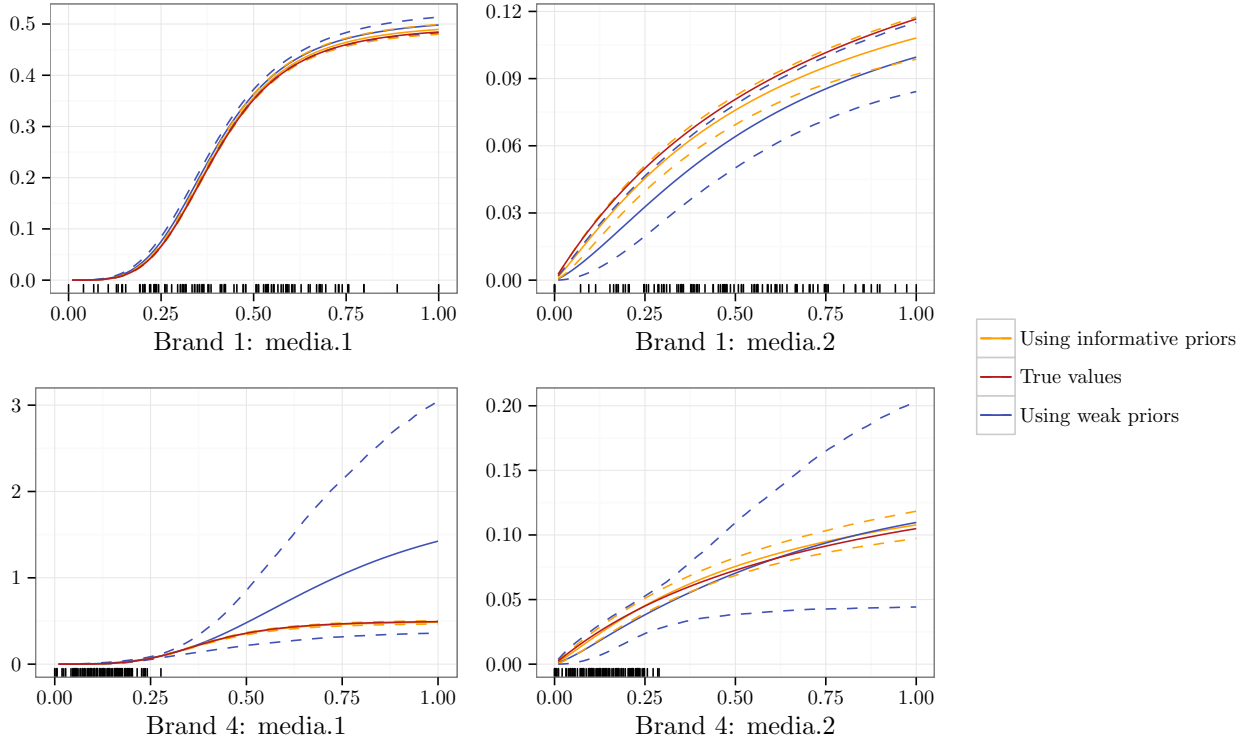
We repeat the above simulation 100 times using the same category-level parameters specified in Table 1, as well as fixed overall brand sizes. Each simulation dataset contains 10 brands. We summarize in Figure 5 the estimated average ROAS and mROAS after subtracting the true values. Each data point summarized in Figure 5 is posterior mean estimates over all MCMC iteration for a simulated dataset. The benefits of using informative priors are consistent across datasets with similar characteristics.



**Figure 4:** Comparison of estimated average ROAS (top) and mROAS (bottom) for each of the 10 brands simulated in an example dataset, of media 1 (left) and media 2 (right), against the true values (red dots). The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean. Another version of this figure for media 1 is included in Appendix D (Figure 26), where the true values are subtracted from the estimations.



**Figure 5:** Comparison of estimated average ROAS (top) and mROAS (bottom) summarized over the 100 simulated datasets, of media 1 (left) and media 2 (right), after subtracting the true values. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.



**Figure 6:** Response curves estimated by the brand-specific models using informative priors with 5th and 95th percentiles (in dotted lines), compared with that using weak priors with 5th and 95th percentiles (in dotted lines), and the true response curve (in red). Tick marks on the bottom indicate values of the observed media variables of each brand.

### 5.1.3 Comparisons of response curves estimated by brand models

To isolate the media parameters from the variation in the dataset, we define a media response curve as  $R(x) = \beta_m H(x, \mathcal{K}_m, \mathcal{S}_m)$  for  $x \in [0, 1]$ .

Though the average ROAS and mROAS are critical in reporting the overall effectiveness of media, a reasonable estimation of the response curve is required to extrapolate beyond the range of individual brand's media spend with good confidence, and therefore, is critical to obtaining a reasonable estimate of optimal media mix. We calculate the response values using each of the MCMC samples at each sampled values of  $x \in [0, 1]$  and then use the 5th and 95th percentiles of the response values at each evaluation point of  $x$ , i.e., the pointwise 90% credible interval, to indicate the uncertainty in estimating the response curve.

Figure 6 displays the response curves for two of the 10 brands. The brand model using informative priors often provides a narrower credible interval, as well as a smaller error. We observe three types of patterns in Figure 6. Both brand models offer a reasonable estimate of the response curve for media 1 of brand 1. For media 2 of brand 1, the weak priors produce estimates with larger deviation from the true response curve than the category informative priors, but both still have fairly similar shape. Lastly, for media 1 of brand 4, the response curve estimated using weak priors is highly deviated from the true response curve and has a wrong trajectory.

The improvement realized by using informative priors can be explained by the enriched variation from the category dataset, compared to a single brand’s dataset. The tick marks along the base of each plot in Figure 6 indicate the locations of each media spend observation of the brand plotted. Substantial improvements are seen in regions where the advertiser has little data. For media 1 of brand 1 and 4, we see that within the range of the brand’s own observations, the response curve estimated by the brand model with weak priors agrees with the true response curve. It is the section beyond the range of each brand’s media spend where the informative priors provide the most increase in estimation accuracy and confidence. A growing brand could potentially be interested in that section, for extrapolation purposes, when they consider expanding and increasing their media spend in certain media, e.g., online channels.

Figure 6 shows that the informative priors derived from the category model can contain information learned from other brands, and therefore help brands to estimate the later section of the response curves better than they could with their own data. The improvement lies not only in the accuracy of point estimates (solid lines in Figure 6), but also in the width of the credible intervals (dashed lines).

In particular, the improvement occurs because of the trade-off among the media parameters (Jin et al., 2017). The category model uses a larger range of values of media variables to better narrow down the media parameters. By providing the brand model with an informative prior derived from the category model, we improve the estimation accuracy of the media parameters, even beyond the range of media spend one brand observes. On the other hand, Brand 1 has a wide range of spend levels. Its response curve estimated using weak priors already agrees with the true curve well.

In general, brands that have a good variation in their own media spend would see less improvement in extrapolation accuracy and estimation confidence from incorporating category informative priors. However, they may still benefit from category analysis in media channels where other brands have spent differently, and also from accurately capturing competitor effects (see Section 5.2).

To inspect the generality of our above observations, we plot the distribution of mean response curve estimated based on each of the 100 simulated datasets for all 10 brands in Figure 7. Indeed, the improvement of using informative priors is not an isolated case.

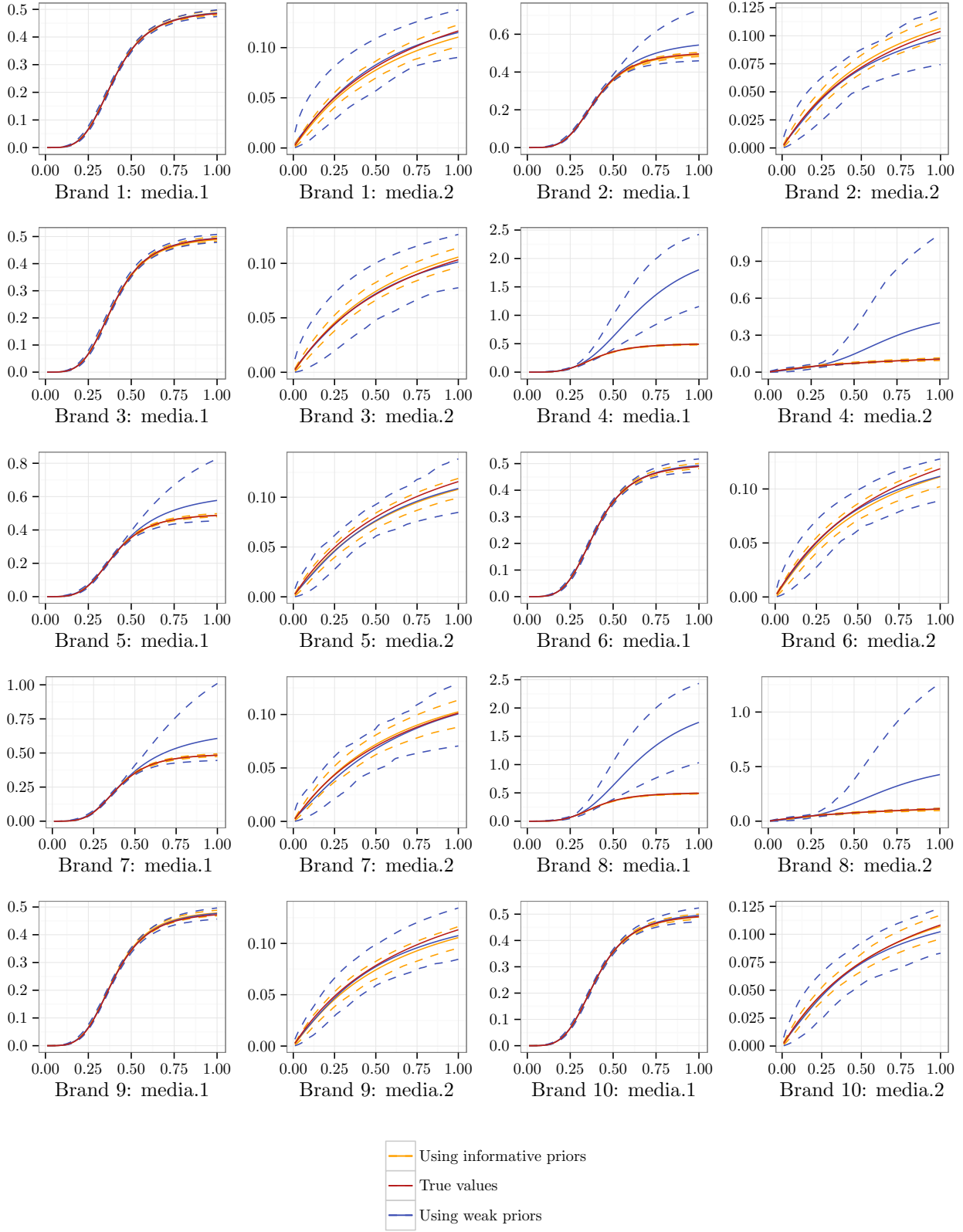
Through the above simulation study, we see that pooling different brands’ datasets can improve the estimation accuracy of response curves of media impact, as the cross-brands variation in media variables can be the key to better inference. Our conclusion applies to both media variables and control variables.

This simulation study is a much simplified version of what the real data might be like. In reality, pooling datasets across brands may improve the accuracy of parameter estimation, but it does not guarantee the estimation be unbiased, nor the estimation uncertainty reasonably small. Bias or large uncertainties could be introduced by other factors such as omitted variables.

## 5.2 When competitive factors have a non-trivial impact on KPIs

In this section, we inspect the potential benefits of including competitive factors in the HCM. In particular, we use price of the brand and of its competitors as our example control variables. The following section is structured similarly to Section 5.1.





**Figure 7:** Comparison of estimated response curves summarized over the 100 simulated datasets with 5th and 95th percentiles (in dashed lines) with the true response curves (in red), of media 1 (left) and media 2 (right).

**Table 2:** Specification of media and price impact in Section 5.2.

	Media 1	Price	Competitor price
Transformation	$H(x, \mathcal{K} = 0.4, \mathcal{S} = 4)$	$\log(x)$	$\log(x)$
Coefficient	$\beta_{1,\cdot} \sim N(0.5, 0.01^2)$	$\gamma_{p,\cdot} \sim N(-1, 0.05^2)$	$\gamma_{cp,\cdot} \sim N(0.5, 0.05^2)$

### 5.2.1 Data simulation and model setup

In this scenario, we simulate only one media variable, in order to focus on the effects of competitive factors. The price variable is simulated as an autoregressive process of order 4, to imitate what we observe in the soda case study (see Figure 22 in later sections). We calculate the competitor price variables as described in Appendix C. The simulated response depends on a control variable that we assume is known to the category dataset but withheld from any individual brand alone. We simulated 100 datasets using the same specifications listed in Table 2, each of which contains weekly observations of 10 brands over 104 weeks. For each simulation instance, we ran 2000 MCMC iterations using *RStan* with 4 parallel chains.

For each simulated dataset, we first develop a HCM specified in (3) and (5), using the category data of  $B = 10$  brands and  $T = 104$  weeks. We use the same weak priors described in (8) for the media parameters. We also include the two control variables, price and competitor price, in the model. We use the following weak priors in the category model for the price coefficient  $\gamma_p$  and the competitor price coefficient  $\gamma_{cp}$ :

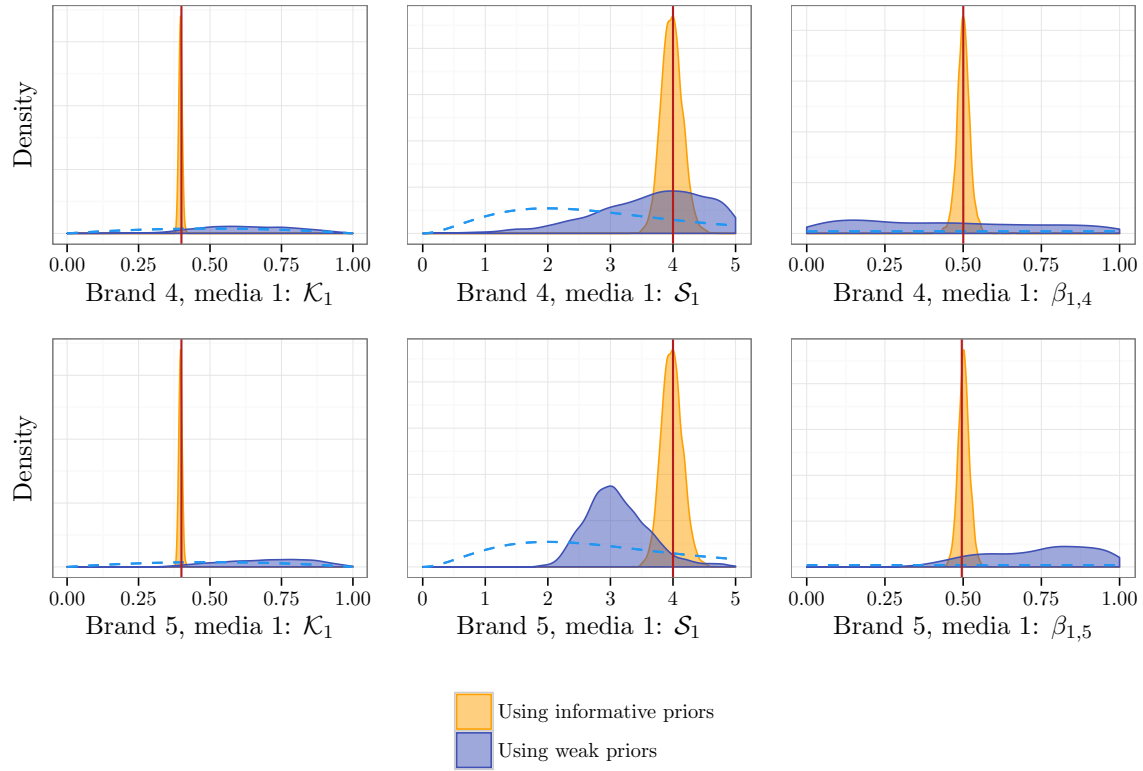
$$\gamma_p \sim \text{Uniform}(-5, 5), \gamma_{cp} \sim \text{Uniform}(-5, 5).$$

### 5.2.2 Comparisons of brand-specific models using informative v.s. weak priors

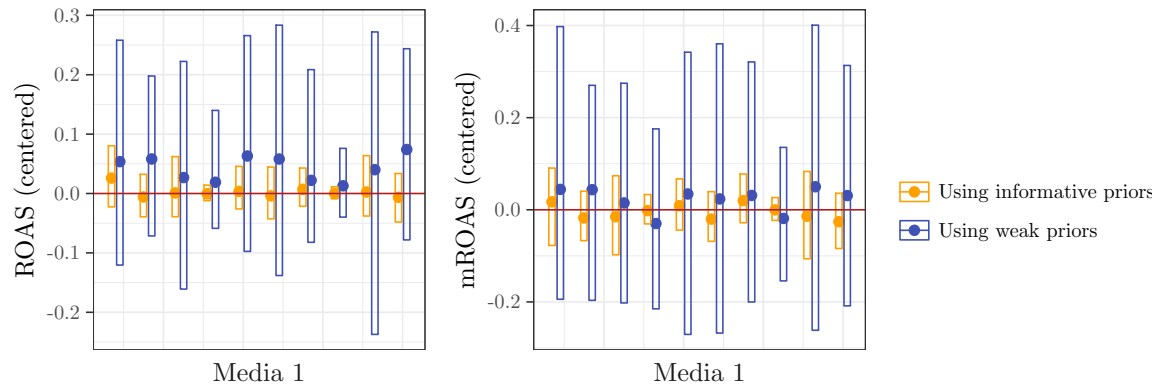
We develop two sets of brand-specific models: one set uses the same weak priors as the category model, the other uses the informative priors derived from the category model results. Each brand-specific model uses only the single brand’s data of the one media variable and the brand’s own price variable. The competitor price variable is omitted in the brand-specific models. If competitor variables are available to individual brands, we suggest using a full category model.

Figure 8 compares the individual media parameter estimates by the brand-specific models against the true values for two of the 10 brands in one of the 100 simulated datasets. The brand models using the weak priors show relatively low estimation accuracy and large uncertainties, partially due to the omitted competitor variable. Figure 8 shows that informative priors can improve the estimation accuracy (e.g., brand 4) and confidence (for both brands shown). For readers’ information, we include more results for this example in Appendix D, i.e., the centered ROAS and mROAS comparisons (Figure 27) and response curve comparison (Figure 28). We now move on to examine the brand-specific model comparisons summarized over the 100 simulated datasets for a more general study.

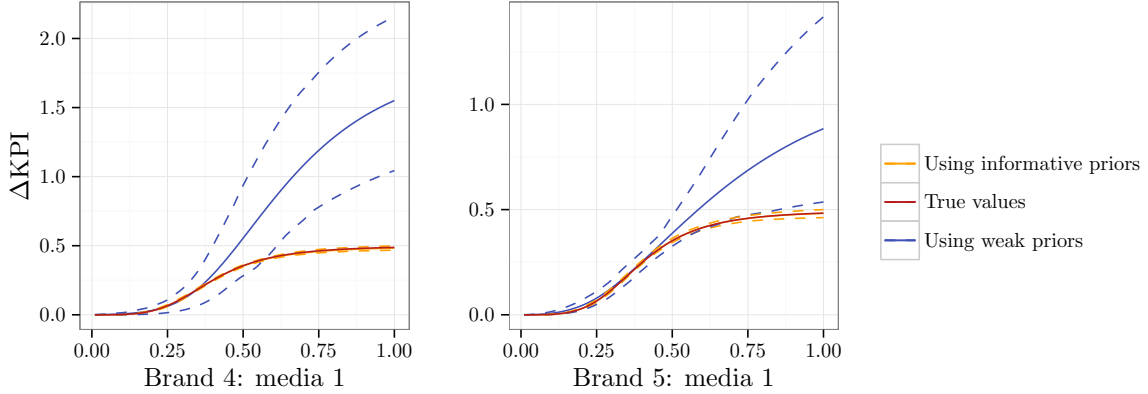
Figure 9 compares the average ROAS and mROAS centered by true values summarized over the 100 datasets. Figure 10 compares the estimated media response curves with the true response curve, summarized over the 100 datasets. We can see that informative priors consistently improves estimation accuracy as well as estimation confidence, especially for the estimated response curves.



**Figure 8:** Marginal density plots of media parameters posterior samples estimated by brand models using informative priors (in gold), compared to those using weak priors (in blue), using an example dataset, as well as the true values (vertical red lines) and the weak priors used (dashed light blue lines).



**Figure 9:** Comparison of estimated average ROAS (left) and mROAS (right) summarized over the 100 simulated datasets of media 1, after subtracting the true values. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.



**Figure 10:** Comparison of estimated response curves summarized over the 100 simulated datasets with 5th and 95th percentiles (in dashed lines) with the true response curves (in red), of media 1 for two example brands.

**Table 3:** Specification of media impact in Section 5.3.

	Media 1	Media 2
$\mathcal{K}$	0.4	0.8
$\mathcal{S}$	4	1
Coefficient $\beta_{m,\cdot}$	$\beta_{1,\cdot} \sim N(0.5, 0.1^2)$	$\beta_{2,\cdot} \sim N(0.2, 0.05^2)$

### 5.3 When non-trivial variation exists among brands

In the following scenario, we demonstrate the importance of a category-brand hierarchy in developing a category model, when there is non-trivial variation among brands. Our findings from this study can be extended to variation in control variable coefficients without loss of generality.

#### 5.3.1 Data simulation and model setup

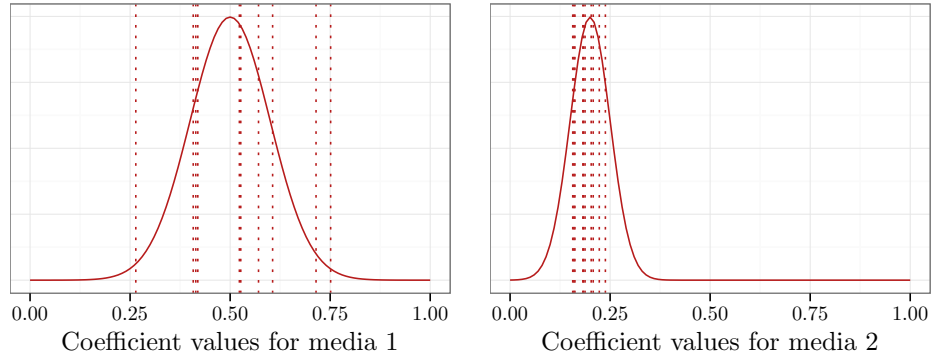
We simulate two media channels with the specifications listed in Table 3. The standard deviation of the media coefficients are increased from previous simulations to 20-25% of the mean value, while the other media parameters remain the same as in Section 5.1. We simulate 100 datasets using the same specifications, each of which contains weekly observations of 10 brands over 104 weeks.

Figure 11 illustrates the distributions used to sample the brand-specific media coefficients in solid lines, and the 10 values sampled for the 10 brands in dotted vertical lines, for one simulated dataset.

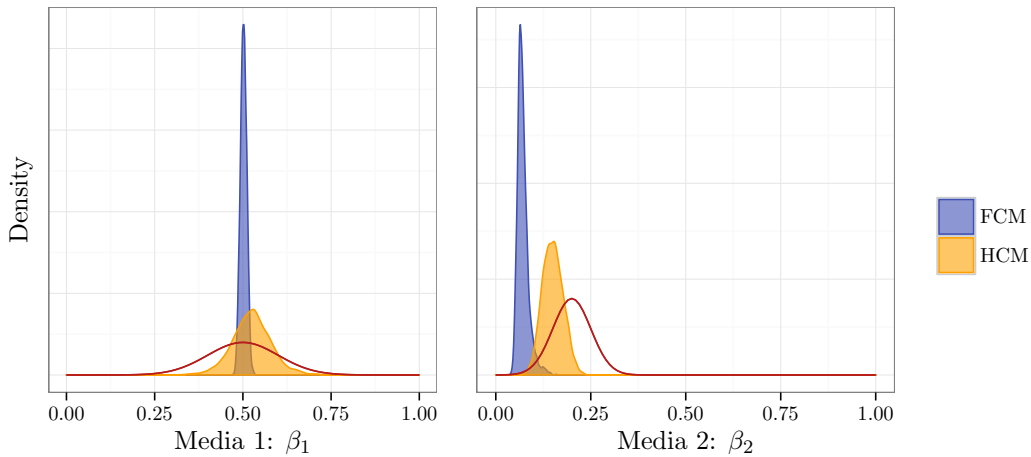
In this simulation, we compare brand-specific models using informative priors derived from the HCM with informative priors derived from a category model without a hierarchy. The latter can be specified similarly to the HCM, but with the following additional assumption:

$$\beta_{m,1} = \dots = \beta_{m,B} = \beta_m, \quad (9)$$

for  $m = 1, \dots, M$ . We refer to a category model under (9) as a flat category model (FCM) from here on.



**Figure 11:** Distribution (solid lines) of brand-specific media coefficients for media 1 (left) and media 2 (right), as well as the 10 brand-specific coefficients (dashed lines).



**Figure 12:** Marginal density plots of media parameters posterior samples estimated by the HCM (in gold) and FCM (in blue), compared to the true distribution (in red) used to generate the brand-level media coefficients  $\beta_{m,b}$ 's.

### 5.3.2 Summaries of derived informative priors

We first examine the posterior samples of the category-level media coefficients  $\beta_m$  estimated by the two category models. Figure 12 displays such a comparison using one simulated dataset. It is not entirely an equitable comparison, considering that the posterior samples of  $\beta_m$  in HCM represents the distribution of the mean parameter of the distribution from which the brand-specific coefficients  $\beta_{m,\cdot}$  are drawn, while the posterior samples of  $\beta_m$  in FCM represents the distribution of the brand-specific  $\beta_{m,\cdot}$ , due to (9). Here we attempt to conduct a comparison in order to understand the difference in the two category model structures.

Figure 12 shows that the posterior distribution of  $\beta_1$  estimated by the FCM is noticeably tighter than that from both the HCM and the distribution from which  $\beta_{1,b}$ 's are drawn. The plotted posterior samples are estimates of the mean parameter of the distribution from which the brand-specific media coefficients are sampled. The FCM under the assumption in (9) is equivalent to a HCM with a prior distribution on  $\beta_{m,b}$  with a standard deviation of 0, i.e.  $\beta_{m,b} \sim N(\beta_m, 0^2)$ . The FCM posterior estimates, being a combination of the prior and data, are thus a lot tighter towards

the sample mean than the HCM with a prior distribution whose standard deviation is greater than 0.

We also notice that the  $\beta_2$  estimated by both the HCM and the FCM are shifted towards 0, compared to the true distribution we used to sample the  $\beta_{2,b}$ 's. In fact, Figure 11 shows that such underestimation is mostly due to chance. The 10 randomly-drawn values for the brand-specific coefficients,  $\{\beta_{2,b}\}_{b=1}^{B=10}$ , are not evenly distributed: more than half of the 10 values are less than the mean  $\beta_2$ . Our simulation case shows us the limitation of category analysis when the category is small. It serves as motivational evidence to accumulate datasets from and conduct a analysis over a large category composed of many similar brands.

In this study, we denote the posterior samples of the media parameters estimated by the HCM as  $\{\Phi_m\}^{(hc)}$  and those by the FCM as  $\{\Phi_m\}^{(fc)}$ . For every brand of each simulated dataset, we then build two brand-specific models using informative priors in the format of  $\{\Phi_m\}^{(hc)}$  and  $\{\Phi_m\}^{(fc)}$  respectively.

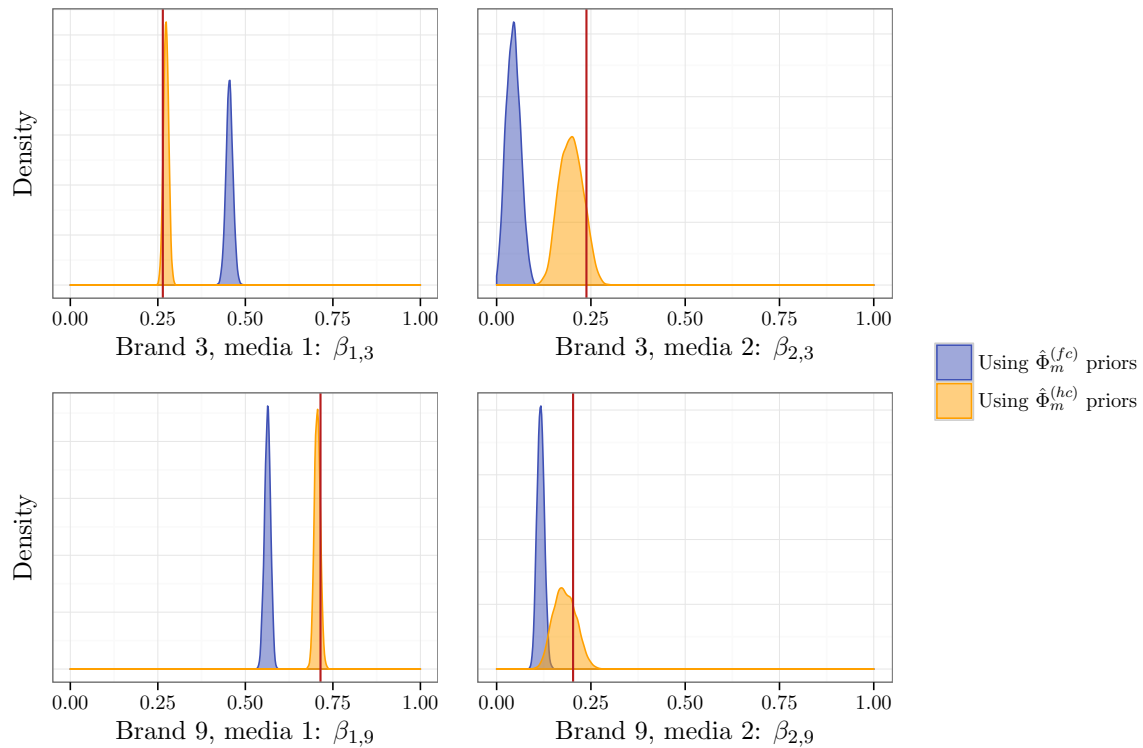
### 5.3.3 Comparisons of brand-specific models using priors from the HCM v.s. FCM

Figure 13 compares the posterior estimates of brand-specific media coefficients using  $\{\Phi_m\}^{(hc)}$  or  $\{\Phi_m\}^{(fc)}$  as informative priors, with the true values indicated by the red lines, for one simulated dataset as an example. The informative priors represented by  $\{\Phi_m\}^{(hc)}$  allow the brand-specific model to adapt to the underlying variation of  $\beta_{1,b}$ 's, while  $\{\Phi_m\}^{(fc)}$  leads to high estimation error. The average ROAS and mROAS comparison for this simulated dataset (Figure 29) is included in Appendix D, along with the comparison of estimated response curves (Figure 30).

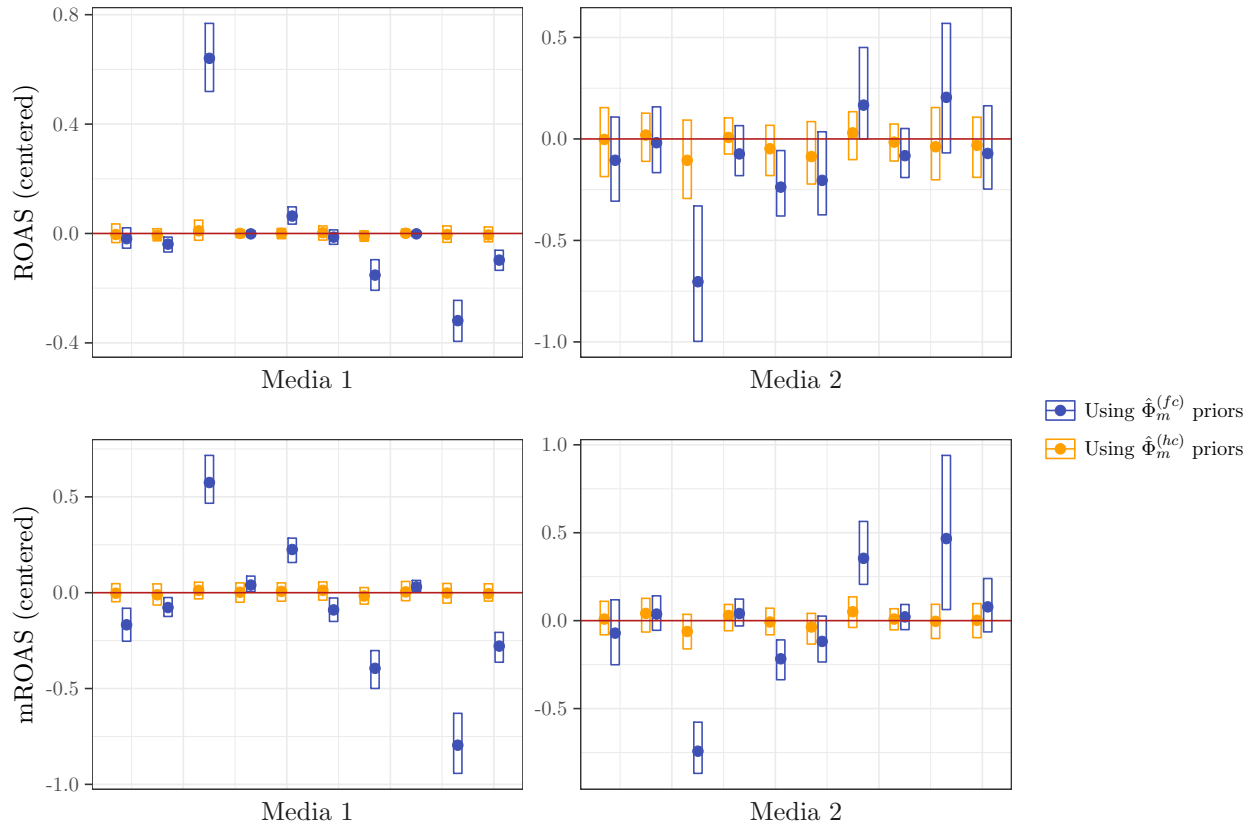
We repeat the simulation for 100 times using the same specifications summarized in Table 3 and the same 10 values of the brand-specific coefficients displayed in Figure 11. Figure 14 shows the distribution of the mean of estimated average ROAS and mROAS from each simulated dataset after subtracting the true values. We see that for some brands, the incorrectly flat structure of the FCM leads to significantly lower estimation accuracy of average ROAS and mROAS, compared to the uncertainties introduced by the data when calculating these two metrics. For some brands that behave similarly to the category mean, i.e. when  $\beta_{m,b} \approx \beta_m$ , the estimation accuracy is understandably better.

Figure 15 compares the pointwise mean of estimated response curves from each of the 100 simulated datasets for two example brands. The curves estimated by brand-specific models using  $\{\Phi_m\}^{(fc)}$  show both high estimation error and large uncertainties, which confirms the importance to incorporate a category-brand hierarchy when there is variation among the brands.

Through this simulation scenario, we see the importance of allowing for a category-brand hierarchy in the category model when there is variation among the brands. We also see the benefits of incorporating as many brands as possible in a category study. Yet, this simulation only explores nontrivial variation in media variable coefficients and assumes the same shape parameters across brands within the same category. In order to gain benefits from pooling different brands together, the brands have to share similarity on some level. If there is significant variation among all media parameters across brands - such that the response curves don't share the same basic shape - we go back to the same parameter to observation ratio as fitting a single media mix model using a single dataset. Large distinctions among brands may be the case in some categories, and in those cases category analysis is not likely to provide much improvement from analyses of individual

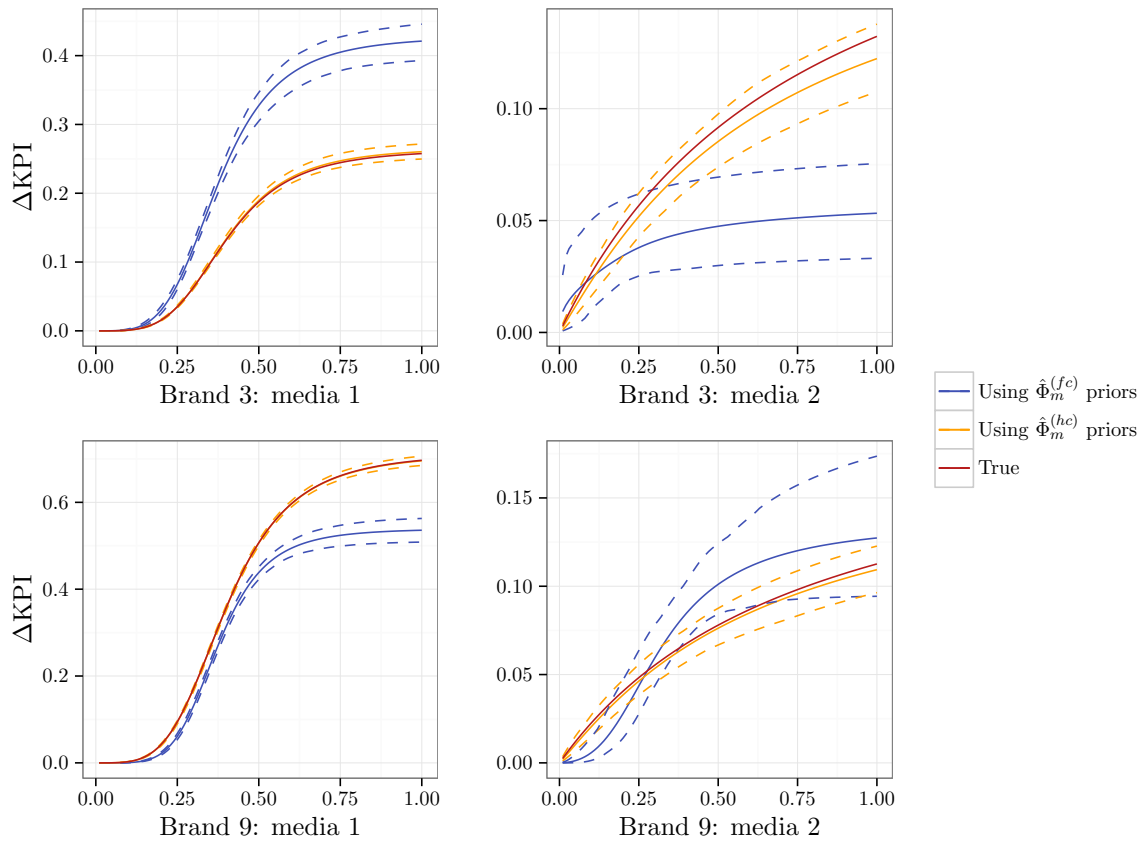


**Figure 13:** Marginal density plots of media parameters posterior samples estimated by brand models using informative priors derived from the HCM (in gold) and those from the FCM (in blue), as well as the brand-specific true values (vertical red lines) using an example dataset.



**Figure 14:** Comparison of estimated average ROAS (top) and mROAS (bottom) summarized over the 100 simulated datasets, of media 1 (left) and media 2 (right), after subtracting the true values. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.





**Figure 15:** Comparison of estimated response curves summarized over the 100 simulated datasets with 5th and 95th percentiles (in dashed lines) with the true response curves (in red), of media 1 (left) and media 2 (right) for two example brands.

brands. The complexity of our model is always restricted to the amount of and the information content within our data and more and/or better data can support more complex models. If there is sufficient information content in the datasets, e.g., through a category of significantly more brands than we’ve simulated here, it is worthwhile to explore variations across brands in terms of media response in a more complex manner.

## 6 A case study of the shampoo category

In this section, we use data from the shampoo category to provide an example of category-brand analysis, as well as to discuss some challenges we face in real world MMM analysis. Google used data consolidated by Neustar MarketShare. The data included sources such as Kantar Media, IRI, ITG, JD Power, and Rentrak.

We first introduce the dataset and the model setup in Section 6.1. Section 6.2 then compares the brand-specific results using informative priors v.s. weak priors. In the end, we briefly compares the across-brand variation in coefficients of the control variables to those of the media variables (Section 6.3).

### 6.1 Data and model setup

The dataset includes  $B = 14$  shampoo brands and covers 2.5 years ( $T = 130$  weeks) of weekly data from April of 2012, through September of 2014. All weekly observations are aggregated at the national level. These 14 brands make up about 60% of the total volume sales of the US shampoo industry over these 2.5 years. The average weekly volume sales ranges from 0.6 million oz. to 44.3 million oz. Our dataset includes some of the major supermarket brands (low price, e.g., brand 10), as well as some salon brands (high price, e.g., brand 13) (Figure 16<sup>4</sup>). From here on, we refer to these 14 brands as brand 1 to brand 14, sorted by media spend in a descending order.

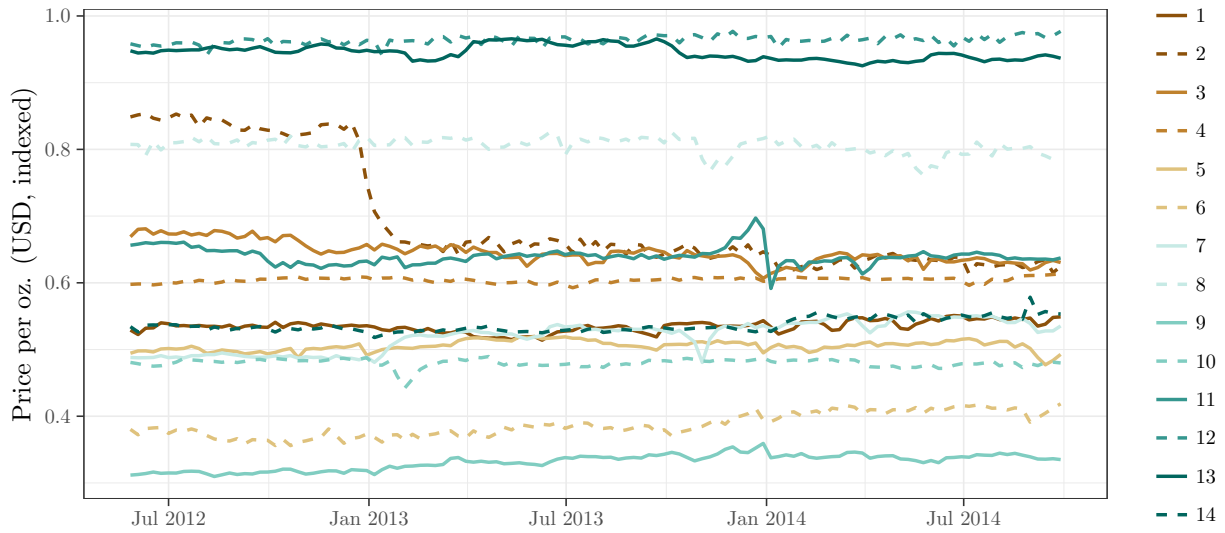
During the timespan of our dataset, 57.8% of the total media spend (in US dollars) of the 14 shampoo brands<sup>5</sup> was on TV, with 36.7% on magazines, 3.7% on Internet display (including Google Display Network and non-Google platforms), 1.3% on YouTube, 0.4% on Internet search (including Google and non-Google search), and the rest 0.2% among other channels, such as out-of-home (OOH), newspaper, etc. Figure 17 displays the distribution of overall media spend among the major channels split by brands.

We first develop the category model described in (3) and (5) for the shampoo category. The media variables are spend (in US dollars) in the major six media channels: TV, magazine, Internet display, YouTube (split into MastHead v.s. non-MastHead ads), and Internet search. Alternatively, one can also use media exposure variables instead of spend, such as TV GRPs, digital display impressions, etc. In this case study, we don’t have reliable exposure data for all major media channels and thus resort to using media spend variables.

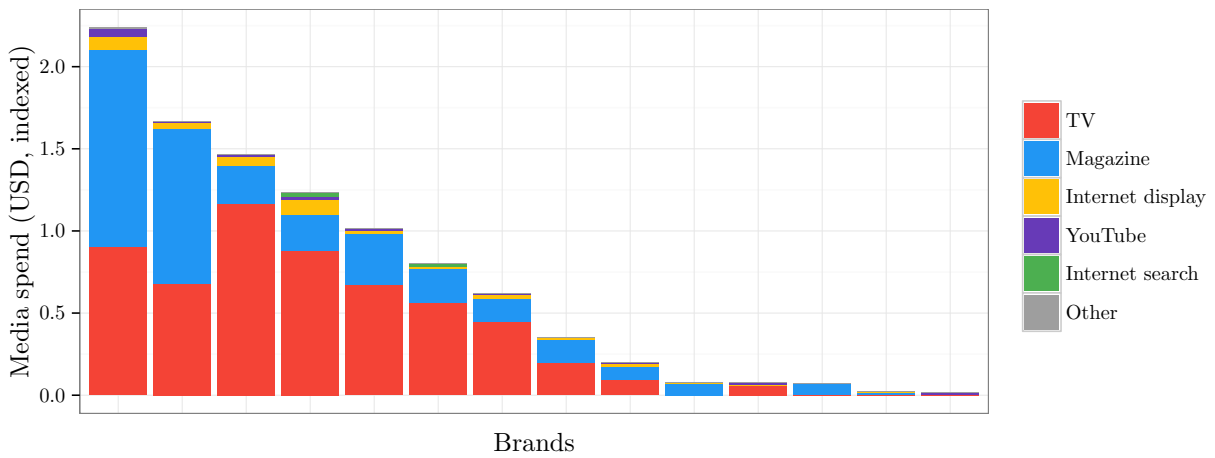
---

<sup>4</sup>The average price of brand 2 dropped significantly at the end of 2012 due to the addition of several lines of lower-tier shampoo products under the same brand.

<sup>5</sup>Particularly, the media spend data collected cover both the shampoo products and conditioner products, and sometimes other relevant hair products (e.g. hair spray, hair cream) of the 14 brands in our study, as such hair products are often advertised together and their advertising expenses inseparable. We are aware of this data issue and understand that our estimated media impact is only a partial impact, i.e. the impact of media spend for total hair product on the sales of shampoo products.



**Figure 16:** Average weekly price per volume (indexed) for the 14 shampoo brands.



**Figure 17:** Media spend (indexed) split by channels for the 14 shampoo brands (in descending order of total media spend over 2.5 years).

We also incorporate the following control variables: price per 16 oz. (in US dollars), All Commodity Volume (ACV) weighted distribution of product, ACV weighted distribution of retailer feature and/or display promotions, the competitor equivalent of these three merchandising variables, as well as the number of social mentions split by sentiment (positive, neutral, and negative). We use volume sales as the response variable.

When calculating the competitor variables, such as price and promotional distribution, we first group the 14 shampoo brands into three clusters by their weekly price using a  $k$ -means algorithm and then calculate the competitor variables within a cluster (direct competition) and across different clusters (indirect competition). Our grouping of brands is motivated by the difference in shampoo brand targeting between supermarket brands and salon brands; we use price as a proxy to the brands’ market targeting. There can be other methods to obtain a meaningful clustering of direct competition within a category, and we see in Section 7, price is not always a good proxy for this purpose. We include the details of how we constructed the competitor variables in Appendix C.

For  $m = 1, \dots, M$ , we use a  $Beta(2, 2)$  prior on the media shape parameters  $\mathcal{K}_m$  and a  $Gamma(3, 1)$  on  $\mathcal{S}_m$ , as well as a  $Uniform(0, 5)$  prior on the coefficients of the media variables. We use a  $N(0, 3^2)$  prior on the coefficients of the control variables. Model training was implemented in the *RStan* language. We use a multiplicative model form by applying the logarithm transformation on the response variable, volume sales of each shampoo brand. We ran 4 parallel chains, each with 2000 MCMC iterations and a warm up phase of 1000 iterations.

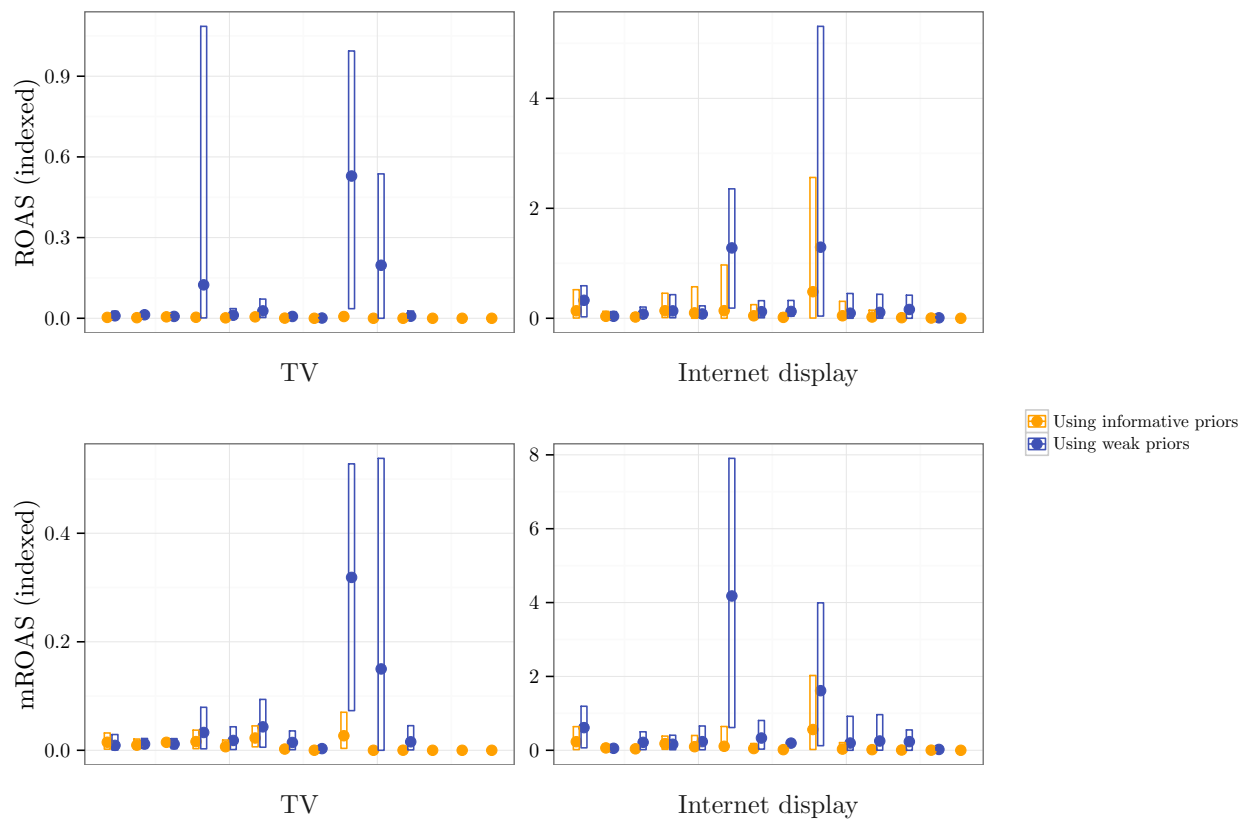
## 6.2 Comparison of brand-specific models using informative v.s. weak priors

To understand the benefits of deploying informative priors derived from the category model, we first develop baseline brand-specific models, using weak priors, similar to the ones we use in the category model. We also exclude all the competitor variables in the brand-specific models. We then develop the brand-specific models using the informative priors in the format of the joint posterior samples estimated by the HCM,  $\{\Phi_m\}^{(c)}$ .

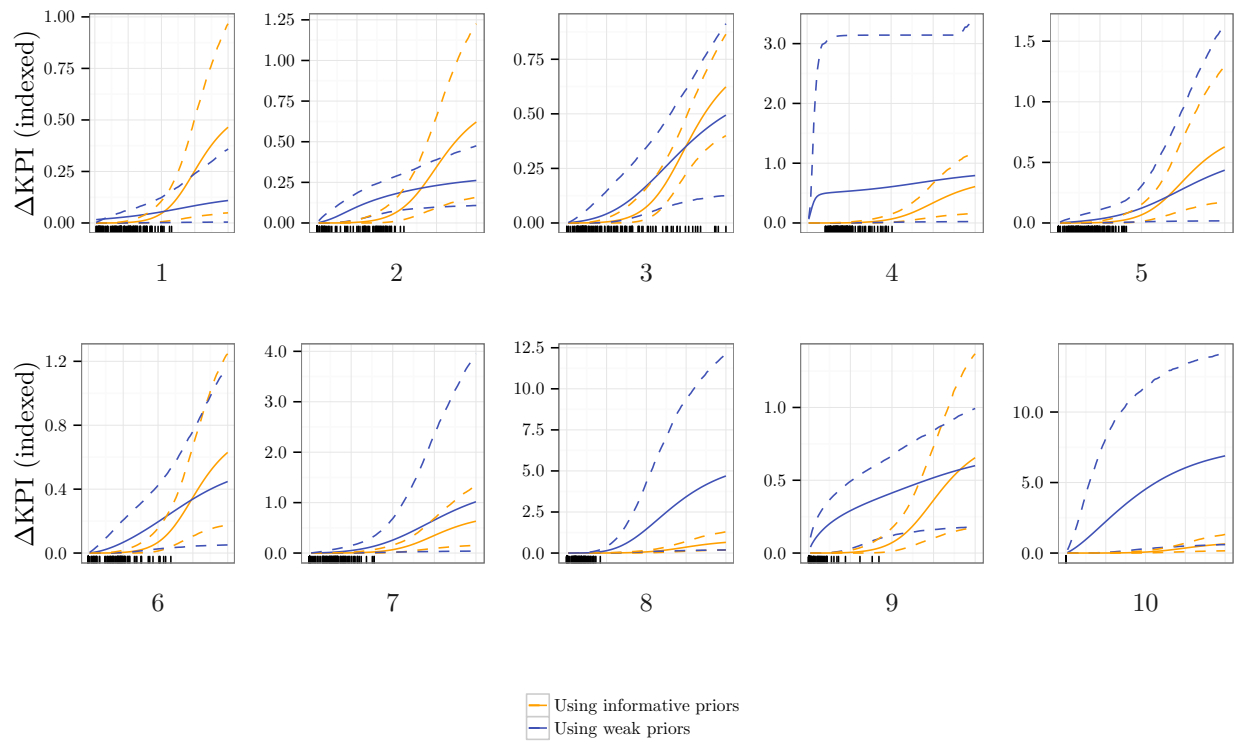
We first compare the average ROAS and mROAS estimates for all 14 brands from the brand-specific models. Figure 18 displays the comparison for two media channels: TV and Internet display. For both media channels, using informative priors derived from the category analysis help reduce the estimation uncertainty of ROAS and mROAS metrics. The results of brand-specific models using informative priors display a larger similarity among brands, compared to that using weak priors, which is consistent with our assumption and model design.

Figure 19 compares the estimated response curves for TV across the 10 of the 14 shampoo brands with non-zero TV spend. The informative priors help reduce the estimation uncertainty, as indicated by narrower credible intervals of the response curves (dotted lines in Figure 19). The level of uncertainty reduction varies across brand. Similarly to what we observed in the simulation studies, brands with a smaller range of media spend benefit more from the category-derived informative priors. E.g., brand 10, a small brand, has limited media spend. When using only the brand’s own data and weak priors, the estimated response curve has quite wide uncertainty (blue dotted lines in Figure 19). In comparison, the informative priors do not seem to influence the results for TV of brand 3 as much.

We also notice that brand 4 yields a strange estimated response curve when using weak priors: a



**Figure 18:** Comparison of estimated average ROAS (top, indexed) and mROAS (bottom, indexed) for TV (left) and Internet display (right) for the 14 shampoo brands. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.



**Figure 19:** Response curves estimated by the brand-specific models using informative priors with 5th and 95th percentiles (in dotted lines), compared with that using weak priors with 5th and 95th percentiles (in dotted lines), for TV of ten example brands of the shampoo category.

sharp increase in impact at the early section of the curve and then an almost-flat section. This strange pattern is likely from the brand’s lack of observations where the media spend is small, as indicated by the tick marks on the  $x$ -axis. Therefore, without borrowing strength from other brands with small media spend, the model for brand 4 with weak priors cannot well tease out the absolute impact of this media channel.

At the same time, we do observe differences in estimated response curves among brands. Its interpretation, however, is not necessarily clear. One explanation could be that the same media used by different brands have different effects, which is plausible, as well-designed and executed ad campaigns may have more impact on audience purchasing behavior. Another explanation could be that our category data are insufficient to develop a strong informative prior, so that the brand-level results are largely influenced by noise or bias in the brand-level data.

### 6.3 Other learnings

Across the brands, we see a great deal of similarity in media effects of TV and Internet display. The top rows of Figure 20 displays the posterior density of brand-specific media coefficients  $\beta_{m,b}$ ,  $b = 1, \dots, 14$ , for TV and Internet display, estimated by the HCM. A similar comparison across brands, for control variables, such as price per volume and retailer promotion distribution, displays a much larger diversity across brands (bottom row of Figure 20). The similarity of estimated media coefficients can be interpreted as the media effects share more similarity across brands, or that the pooled category dataset is insufficient in distinguishing the brand-effect of media variables. The control variables, however, have a much stronger signal, and thus are easier to distinguish.

## 7 A case study of the soda category

In this section, we present another real case study using the soda category data and focus on what is different compared to the shampoo category. Though both are part of the CPG industry, the uniqueness of these two categories can result in different modeling decisions. The source for the soda category data is the same as those for the shampoo category and this section is structured similarly to the previous section.

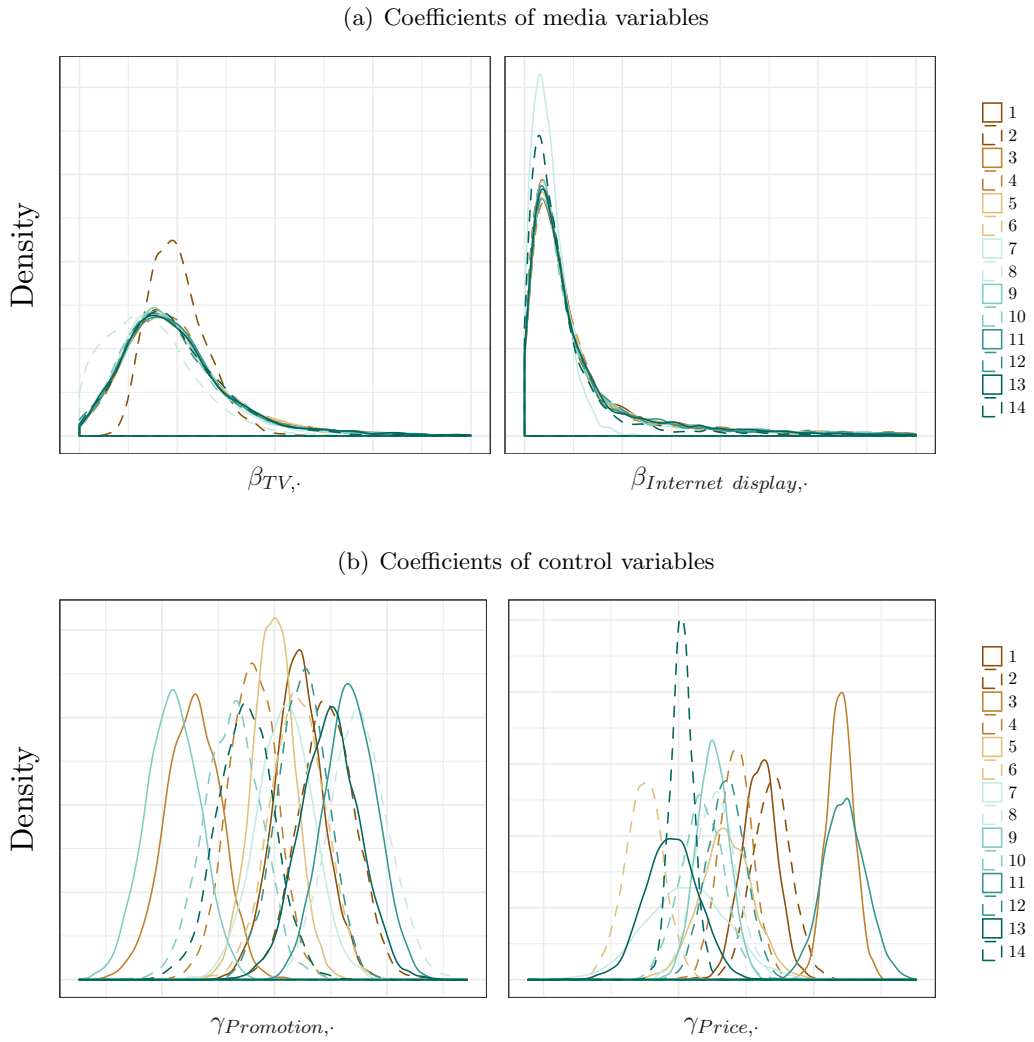
### 7.1 Data and model setup

The dataset includes  $B = 10$  soda brands and covers  $T = 130$  weeks of weekly observations from January of 2012 to September of 2014. The average weekly volume sales<sup>6</sup> ranges from 16 million oz. to 181 million oz. From here on, we refer to these 10 soda brands as brand 1 to brand 10, sorted by media spend in a descending order. During the 2.5 years of our observation, 86.1% of the total media spend (in US dollars) of the 10 soda brands were spent on TV, with 4.5% on radio, 3.5% on magazines, 2.1% on Internet display, 1.7% on YouTube, and 2.1% on other media channels (0.09% on Internet search, 0.07% on newspaper, and 1.9% on business-to-business), as summarized in Figure 21.

We summarize the following traits of the soda category that are different from the shampoo category:

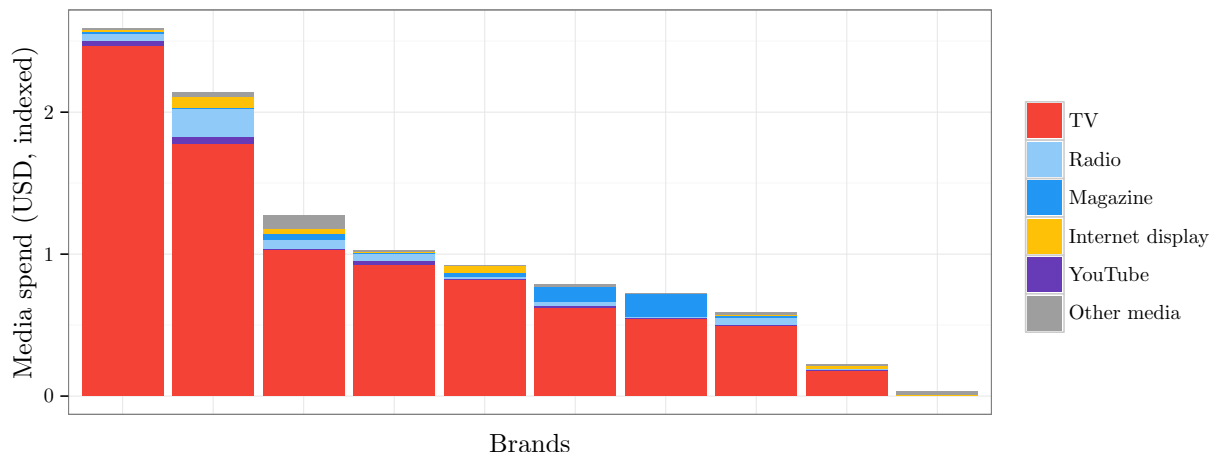
---

<sup>6</sup>The data of soda sales collected by IRI cover only retail sales, not sales through restaurant and bars.



**Figure 20:** Posterior distribution of brand-specific media and control coefficients compared across brands, estimated by the HCM for the shampoo category.

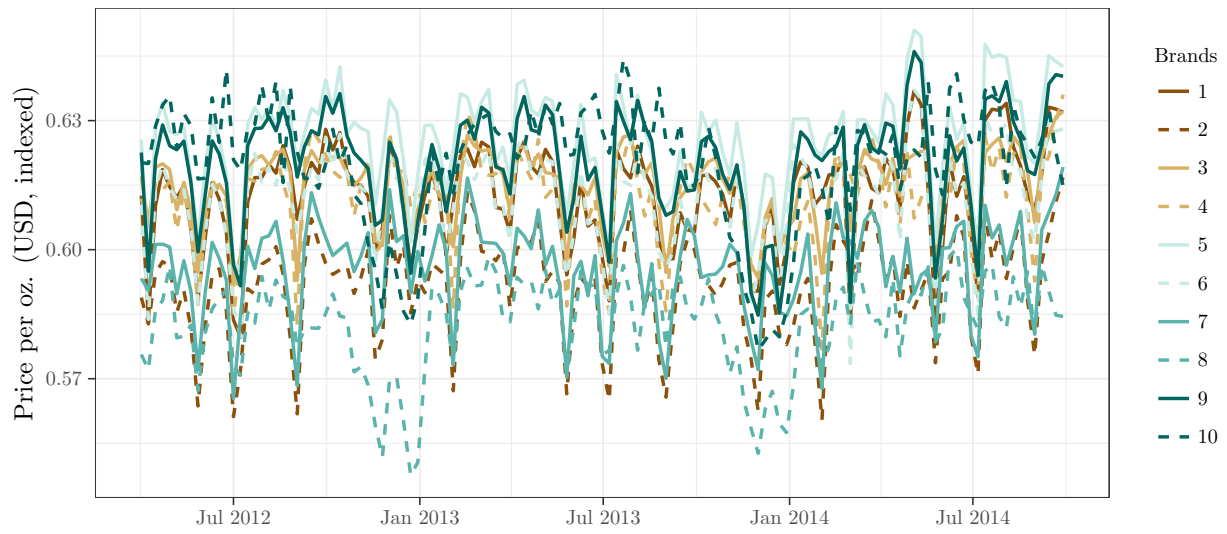




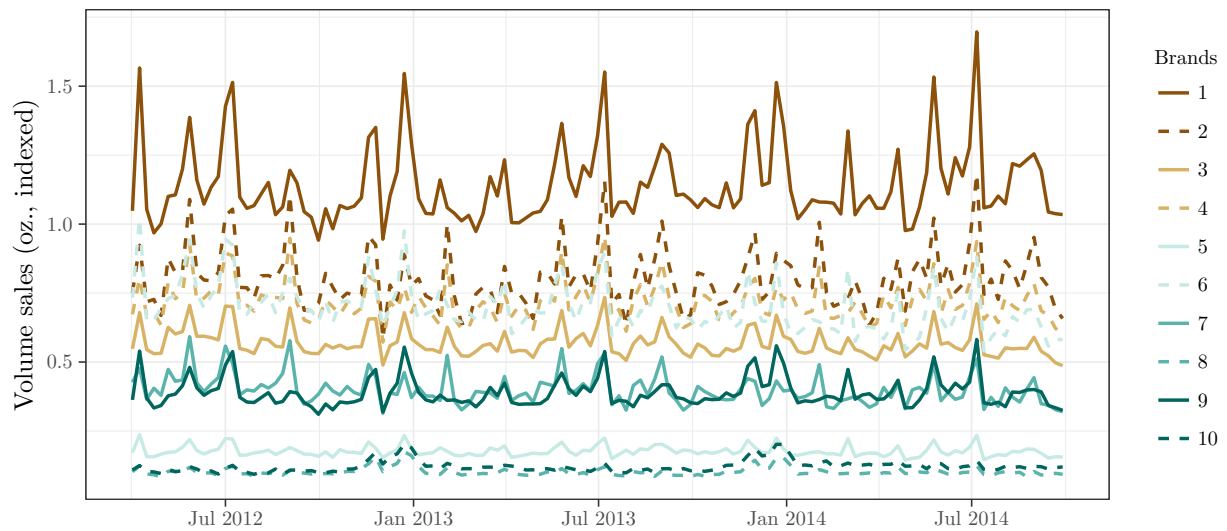
**Figure 21:** Media spend (indexed) split by channels for the 10 soda brands (in descending order of total media spend over 2.5 years).

1. Gathering complete sales data of the soda category is difficult because sales through restaurants and bars are hard to track and not included in our dataset.
2. The soda category has one dominating major media channel, TV. Further, the distribution of weekly spend on TV is extremely long-tailed: the maximum weekly spend on TV is about 11 times the 90th-percentile and more than three times the 99th-percentile. The long tail is from large brands spending much more than smaller brands, as well as large amounts of budget spent concentrated within a small number of weeks due to flighted campaigns.
3. There is no major separation among brands in terms of retail prices (Figure 22), unlike retail prices of shampoo brands that we have seen in previous Section (Figure 16).
4. Strong seasonality exists in volume sales of the soda category (Figure 23), possibly due to soda consumption patterns, such as over major sports events and holiday seasons.
5. The 10 soda brands include sub-brands of the same main brand. For example, data for *Diet Coke* and *Coke Zero* are gathered separately when possible. Still, this introduces unique traits of the soda category that are not observed among the shampoo category, e.g., potential halo effects of advertising among the sub-brands of the main brand (Nisbett & Wilson, 1977).
6. The *Diet*-typed sodas have distinctively different target demographics (female) than the other brands (mostly gender neutral), and hence an often adopted media channel, magazines, compared to radio for the other soda brands. These *Diet*-typed sodas are not direct competitors with the other brands, and vice versa.

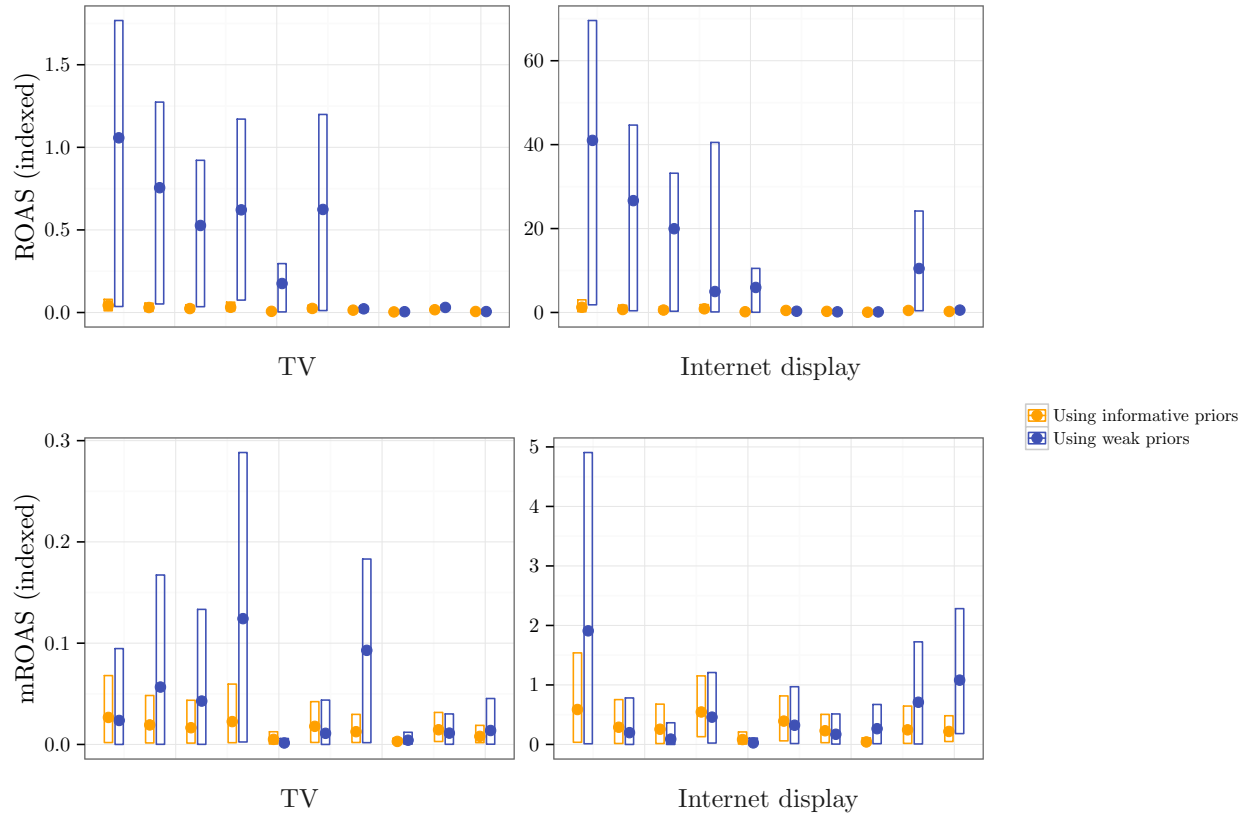
Based on the above observations we made of the soda category, we apply logarithm transformation to the media variables to redistribute the long tails. The category total weekly volume sales is used as a proxy of category seasonality in the model. We identify soda brands which share a parent brand, and incorporate “sibling” brand media variables into the model. Instead of identifying direct competitors by price in shampoo category, we use the gender targeting of the soda brands to establish direct competition.



**Figure 22:** Average weekly price per volume (indexed) for the 10 soda brands.



**Figure 23:** Weekly volume sales (indexed) of the 10 soda brands.



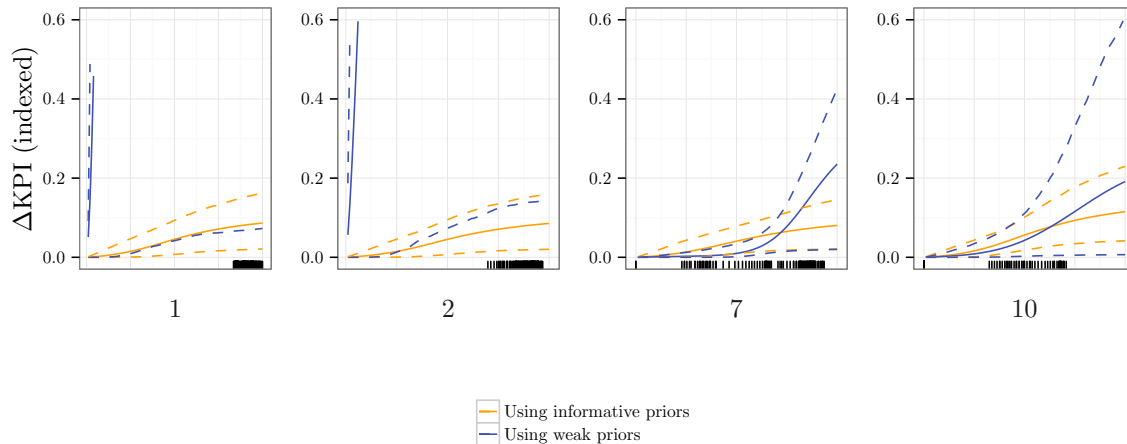
**Figure 24:** Comparison of estimated average ROAS (left, indexed) and mROAS (right, indexed) for TV for the 10 soda brands. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.

There is more than one reasonable approach to address the unique features of the soda category; each modeler would have their own preference based on their experiences and prior knowledge. In this paper, we merely offer one reasonable choice as an example.

## 7.2 Comparison of brand-specific models using informative v.s. weak priors

To understand how information extracted from the hierarchical category MMM impact brand-specific estimates, we again develop two sets of MMMs using individual brand's data: the first set uses the informative priors derived from the category, the second set uses the same weak priors we used in the category model.

Figure 24 compares the ROAS and mROAS for TV for all  $B = 10$  brands, estimated using weak priors (WP) or informative priors (IP) derived from the category model, while Figure 25 compares the corresponding response curves for four example brands. The estimates made with the weak priors have very large uncertainties compared to those made with the informative priors. The reduced uncertainties by using informative priors could be due to improved estimates from pooling different brands' information and better seasonality estimates using the category total sales.



**Figure 25:** Response curves estimated by the brand-specific models using informative priors with 5th and 95th percentiles (in dotted lines), compared with that using weak priors with 5th and 95th percentiles (in dotted lines), for TV of four example brands of the soda category.

Comparing the response curves of TV estimated from the soda category (Figure 25) to that from shampoo (Figure 19), we see that soda brands spent almost twice on TV over the same period of time as shampoo brands did, yet the effects of TV in the soda category estimated using the informative priors are much smaller than that in the shampoo category. Such low estimated impact of TV could result from a higher brand stability and awareness of the soda brands - all of the 10 soda brands have existed for many years with high levels of brand awareness. As a result, TV campaigns mostly aim to retain that brand awareness, to “remind people of their brands”. Such a long-term effect of media is not captured by the MMM structure we use in this paper. Several studies focus on the long-term effects of marketing efforts (Ataman, Van Heerde, & Mela, 2010; Dekimpe & Hanssens, 1999; Leeflang et al., 2009; Mela, Gupta, & Lehmann, 1997).

In comparison, several of the shampoo brands are relatively new; even for the well-known brands, some of them have introduced significantly different lines of products in the time period we studied. The functionality of TV ads to introduce new brands or products potentially leads to more short-term impact of TV campaigns on consumers, which can be captured by our models.

Unlike the response curves of TV for shampoo brands displaying an “S” curve (Figure 19), those for soda brands yield a curve similar to an effective reach curve (Figure 25), which can be approximated using the Hill transformation when fixing  $\mathcal{S}$  at 1 Jin et al. (2017). When there is not enough information in the dataset, one choice a modeler can make is to reduce the complexity of the model, i.e. reduce the number of parameters to estimate in the model.

From the above discussion and results, we see that different product categories can be quite different in terms of their business models, how media affects sales, interactions between brands, and relevant control variables. We cannot stress enough the importance of understanding the category and customizing the analysis in any applied setting. The domain knowledge, coupled with our proposed method of pooling different brands together, helps develop a meaningful MMM.

## 8 Discussion

Media mix modelers are often faced with challenges from insufficient data quantity and information content relative to the model complexity. In this paper, we propose pooling datasets of different brands within the same product category to achieve more useful variation in the data and improved range of media observations, compared to using an individual brand’s dataset. It is difficult for a single brand to greatly vary their media spend pattern over time; even if it does, it takes a couple of years to obtain enough observations for an MMM analysis. Such variation in media spend is relatively easier to obtain with multiple brands. We demonstrate that a hierarchical Bayesian model can be used to learn certain aspects of media effects across brands. Such learnings can then be passed onto brand-specific MMMs via informative Bayesian priors, which have the advantage of anonymity of the brand-specific data. Our approach of category analysis is not limited to the exact model specification, as long as the category model specification is consistent with that of brand-specific models.

Through three scenarios of simulated data (Section 5) and two case studies (Section 6 and 7), we see that the informative priors derived from the hierarchical category model can both improve the accuracy and reduce the uncertainty of estimating the media response curve, and thus render more accurate ROAS and media optimization results. Such benefits are large for small brands within the category for estimating and extrapolating media effects, as well as for large brands that always maintain a certain level of media spend in particular channels. Furthermore, the category dataset can also be used to better understand the product’s intrinsic trend and seasonality that is independent of each brand’s media activities. In this paper, we gave an example in the soda category, of using the category total sales as a proxy of the seasonality of underlying demand for soda. Modelers can also use the category dataset to better understand the impact of competitor activities on brand KPIs. In the third simulation scenario, we gave an example of how including competitor activities in the category model can improve our understanding of media effects. The improvement in estimation accuracy and uncertainty then propagates to brand-specific models via informative priors. Even when the brand models lack access to competitor data, the priors developed using the dataset that did include competitor data help reduce the bias caused by the omitted variables. We hope our study serve as a motivation for advertisers to allow their datasets to participate in category analysis to improve our learnings on media effects.

Through the case studies of shampoo and soda categories, we see the importance and necessity of understanding the unique features of each category and accounting for them in a reasonable way in the category model. Such customization of category analyses includes, but is not limited to, transformation on the explanatory variables, specification of inter-brand relationships (competition and halo effects). For categories with a more complex structure, one might consider extending the category - brand hierarchy. For example, for the automobile category, it is worthwhile to investigate using a hierarchy of category - segments (CUV v.s. conventional) - brands instead.

Our study focuses mostly on improving the lack of variation in MMM datasets by pooling different brands together. For challenges not addressable by introducing more observational variation, such as bias arising from ad targeting of consumer demand and lack of representation of media funnel effects in MMMs, the category analysis likely offers little or no improvement over a single brand MMM. Chan and Perry (2017) and Chen et al. (2017) offer more in-depth discussions on such topics.

## Acknowledgment

The authors would like to thank Tony Fagan and Penny Chu for their encouragement and support, Zhe Chen, Xiaojing Huang, Conor Sontag, Stephanie Zhang, and Shi Zhong for numerous helpful discussions and collaboration, especially Steve Scott for his guidance and great feedback to this work, Luis Gonzalez Perez for several insightful discussions, and Michael Perry for his cross-functional support in this project and for his editorial contribution to the paper.

## Appendices

### A Simulation data generation process

We generate the simulated data by the following process, for a category of  $B$  brands,  $M$  media channels over  $T$  weeks.

- Step 1. For given  $(\beta_m, \eta_m)$ ,  $m = 1, \dots, M$ , randomly sample brand-specific coefficients  $\beta_{m,b} \sim N^+(\beta_m, \eta_m^2)$ .
- Step 2. Simulate underlying demand of products of a category with seasonality of  $T$  weeks using a sinusoidal function.
- Step 3. Randomly assign brand size  $\mathcal{B}_b$  to brand  $b = 1, \dots, B$ , for example,  $\mathcal{B}_b \sim N(100, 50^2)$ .
- Step 4. For each brand  $b$ ,

- Simulate its media planning seasonality pattern that is correlated with the product demand seasonality, for example, with a correlation of 0.8.
- Simulate  $M$  media variables that are correlated with the media planning seasonality patterns and scaled proportional to the brand's size.
- Simulate  $C$  control variables. E.g. price variables can be of a monthly pattern as many CPG products are.

Step 5. For each brand  $b$ , calculate competitor variables based on the other brands within the category.

Step 6.

- For each brand  $b$  and media channel  $m$ , calculate rate of incremental sales:

$$r_{t,m,b} = \beta_{m,b} h_m(x_{t,m,b})$$

- For each brand  $b$  and control variable  $c$ , including competitor variables, calculate rate of incremental sales:

$$r_{t,c,b} = r_{c,b} z_{t,c,b}$$

Step 7. Calculate sales as a product of brand size and media incremental rate of sales:

$$y_{t,b} = \mathcal{B}_b \exp\left(\sum_{m=1}^M r_{t,m,b} + \sum_{c=1}^C r_{t,c,b} + N(0, \sigma^2)\right)$$

for given  $\sigma$ .

The above simulation is based on the following assumptions: media variables impact sales in a multiplicative model form; each brand's media expenditure is correlated with the size of the brand.

## B Calculation of ROAS and mROAS

Using estimated model parameters, the average ROAS for media  $m$ , brand  $b$ , over the  $T$  weeks of simulated data can be calculated as follows,

$$ROAS_{m,b} = \frac{\sum_{t=1}^T (\hat{y}_t(X_{t,m,b} = x_{t,m,b}) - \hat{y}_t(X_{t,m,b} = 0))}{\sum_{t=1}^T x_{t,m,b}},$$

where  $\hat{y}_t(X_{t,m,b} = x)$  denotes model predicted response when media variable  $X_{t,m,b}$  takes value  $x$ . Similarly, we can calculate the average mROAS at 1% multiplicative incremental on the media variable  $m$  for brand  $b$  as,

$$mROAS_{m,b} = \frac{\sum_{t=1}^T (\hat{y}_t(X_{t,m,b} = 1.01 \times x_{t,m,b}) - \hat{y}_t(X_{t,m,b} = x_{t,m,b}))}{0.01 \times \sum_{t=1}^T x_{t,m,b}}.$$

Because our simulation assumes a logarithm transformation on the response variable and no media lag, using a short hand  $\hat{y}_t$  for  $\hat{y}_t(X_{t,m,b} = x_{t,m,b})$ , we have,

$$ROAS_{m,b} = \frac{\sum_{t=1}^T \hat{y}_t \cdot (1 - \frac{\hat{y}_t(X_{t,m,b}=0)}{\hat{y}_t})}{\sum_{t=1}^T x_{t,m,b}} = \frac{\sum_{t=1}^T \hat{y}_t \cdot (1 - \exp(-\hat{\beta}_{m,b} h_m(x_{t,m,b})))}{\sum_{t=1}^T x_{t,m,b}}$$

and ,

$$mROAS_{m,b} = \frac{\sum_{t=1}^T \hat{y}_t \cdot (\exp\{\hat{\beta}_{m,b} (h_m(1.01 \times x_{t,m,b}) - h_m(x_{t,m,b}))\} - 1)}{0.01 \times \sum_{t=1}^T x_{t,m,b}}.$$

## C Calculation of competitor variables

Sometimes a change in a brand's KPI is not due to anything the brand initiated, but rather its competitor's activities, such as price changes, new product launches, massive media spends, etc. The goal is to include competitive factors into the category-level model, while at the same time to reduce the dimension of competitor variables, which is on the scale of number of brands within the category. This appendix discusses one approach of summarizing competitor variables.

We first group the brands  $b = 1, \dots, B$  within a category into several direct-competing clusters  $\{C^1, \dots, C^g\}$ . This can be done differently based on the business model of a category. E.g., we use price to determine direct competitors in the shampoo category, and targeted demographics in the soda category. There can be other reasonable clustering methods.

Denote  $b \in C_b$ . So if brands  $1, \dots, 5$  are clustered into two groups  $\{C^1 = \{1, 2\}, C^2 = \{3, 4, 5\}\}$ , we can write  $1 \in C_1 = C^1$  and similarly  $5 \in C_5 = C^2$ . After clustering the brands, for each brand  $b$ , we put all other brands  $c, c \notin b$  into two groups: direct competitors (which are in the same price cluster as brand  $b$ ), and the rest as indirect competitors. In this manner, for each competitive variable, we can reduce its dimension from number of brands to two: a variable for direct competitors and a variable for indirect competitors. In our case studies of the shampoo and soda categories, the model seems to select direct competitor variables as important, and tends to render the indirect competitor variables insignificant.

## C.1 Competitor price

For brand  $b = 1, \dots, B$ , we calculate a weighted average of the direct competitor prices as

$$CP_{t,b}^{DR} = \frac{\sum_{c \in C_b, c \neq b} P_{t,c} S_{t,c}}{\sum_{c \in C_b, c \neq b} S_{t,c}}$$

and indirect competitor prices as

$$CP_{t,b}^{IDR} = \frac{\sum_{c \notin C_b} P_{t,c} S_{t,c}}{\sum_{c \notin C_b} S_{t,c}}. \quad (10)$$

The calculation of indirect competitors' weighted average price is potentially misleading, when brands of a category are grouped into more than two price groups. For example, when three clusters are formed by price (high-priced, mid-priced, and low-priced), the meaningful way to measure the level of competitiveness of the mid-priced group against the other two groups is by the level of price separation between them. In other words, a more expensive high-priced group and a cheaper low-priced group means less competition for the mid-priced group. The weighted average price calculated in (10), however, can fail to distinguish the levels of such separation. In such cases, we can use a weighted average price in relative terms, by calculating the absolute distance between prices of different brands.

## C.2 Competitor media

For brand  $b = 1, \dots, B$ , we calculate a normalized sum of competitor media variables as follows:

$$CM_{t,b}^{DR} = \frac{\sum_{c \in C_b, c \neq b} X_{t,c}}{\sum_t \sum_{c \in C_b} X_{t,c}},$$

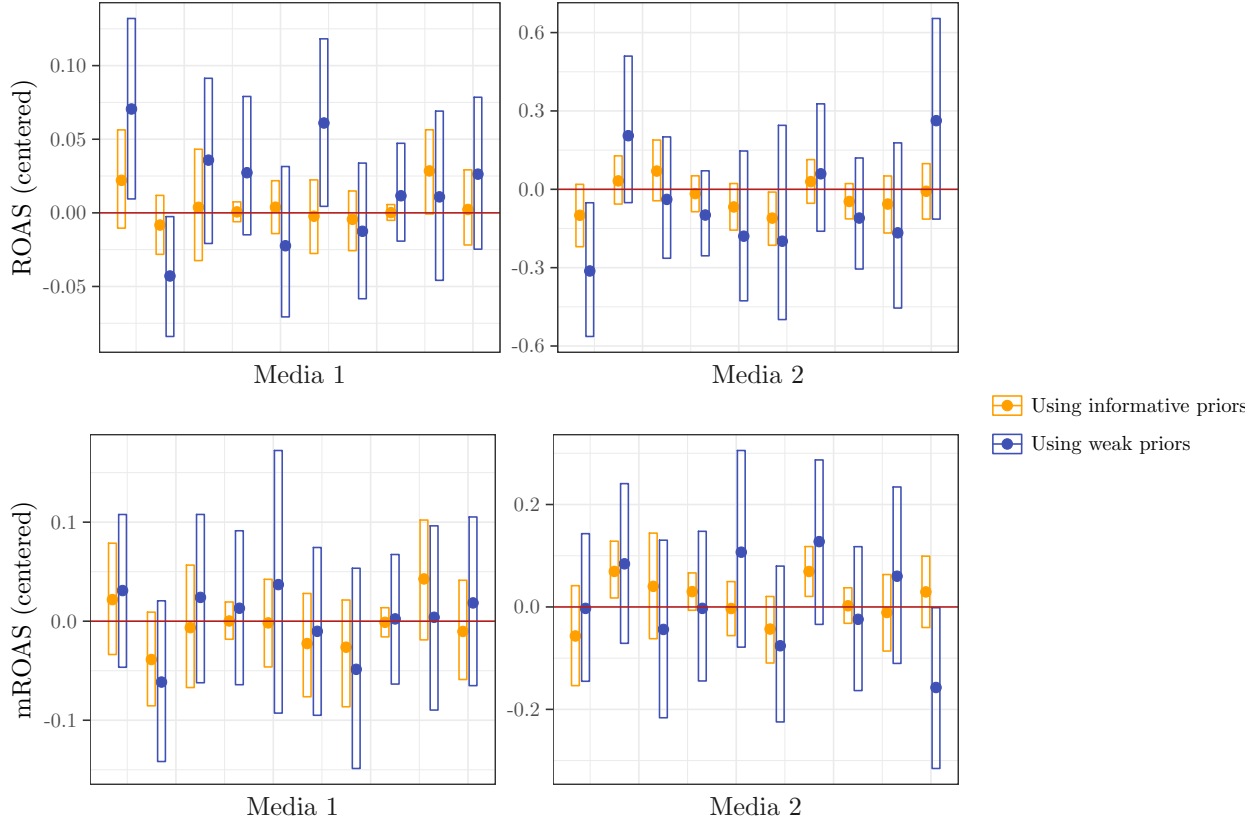
where  $X_{t,c}$  is total media spend of direct competitor  $c$  of brand  $b$ . We normalize the sum by the total media spend of the brand cluster  $C_b$ , so that this competitor media variable is comparable across brand clusters of different sizes, i.e. containing various number of brands. Similarly, indirect competitor media variable can be defined as

$$CM_{t,b}^{IDR} = \frac{\sum_{c \notin C_b} X_{t,c}}{\sum_t \sum_{c \notin C_b} X_{t,c}}.$$

## C.3 Competitor distribution

The product or promotional distribution variables of competitor products behave similarly to competitor media variables, in that it is additive among competitors, and that we need to normalize the variable by each brand cluster, so that the competitor distribution variables are comparable across brand clusters of different sizes.





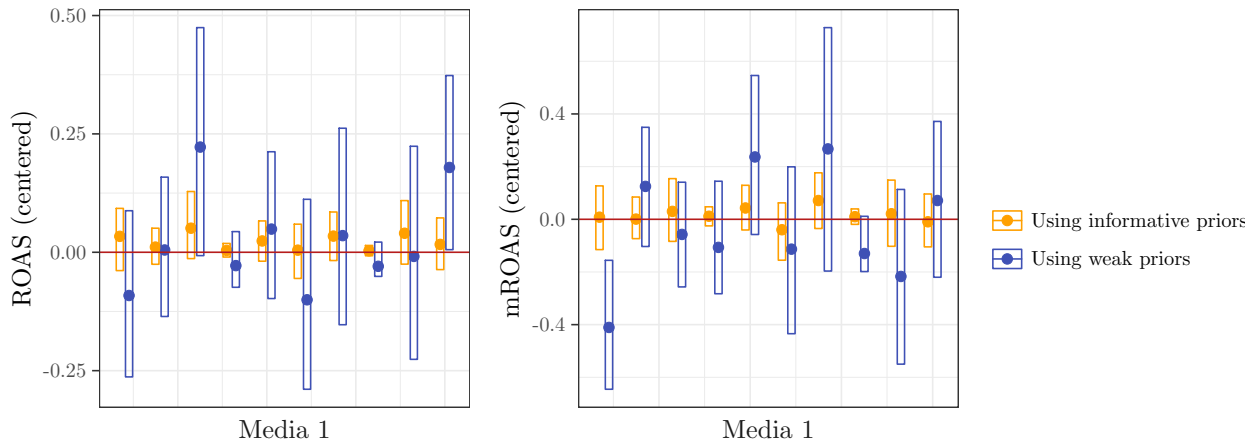
**Figure 26:** Another view of Figure 4: comparison of estimated ROAS (left) and mROAS (right) of media 1 and 2 for the 10 simulated brands in Section 5.1.2, after subtracting the true values. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.

Therefore, for brand  $b = 1, \dots, B$ , we calculate a normalized sum of competitor distribution variables as follows:

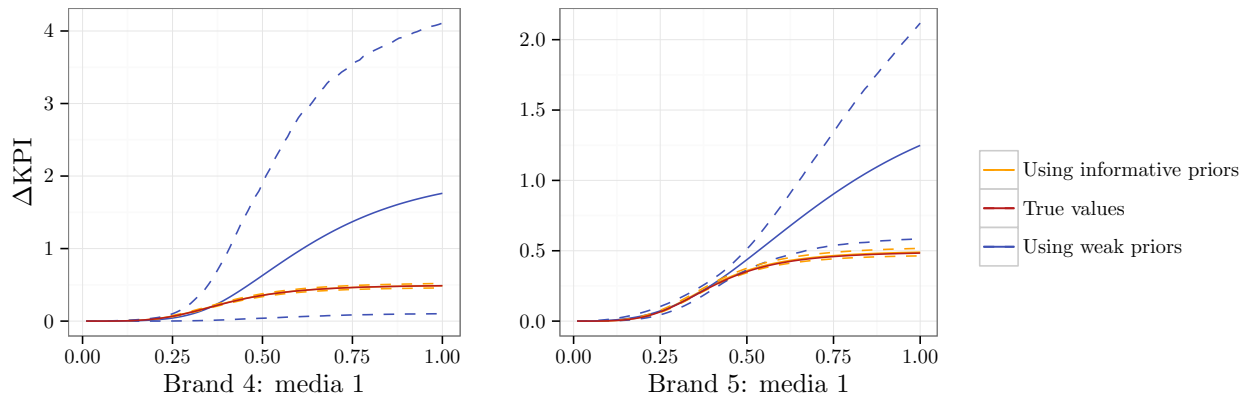
$$CD_{t,b}^{DR} = \frac{\sum_{c \in C_b, c \neq b} D_{t,c}}{\sum_t \sum_{c \in C_b} D_{t,c}},$$

$$CD_{t,b}^{IDR} = \frac{\sum_{c \notin C_b} D_{t,c}}{\sum_t \sum_{c \notin C_b} D_{t,c}}.$$

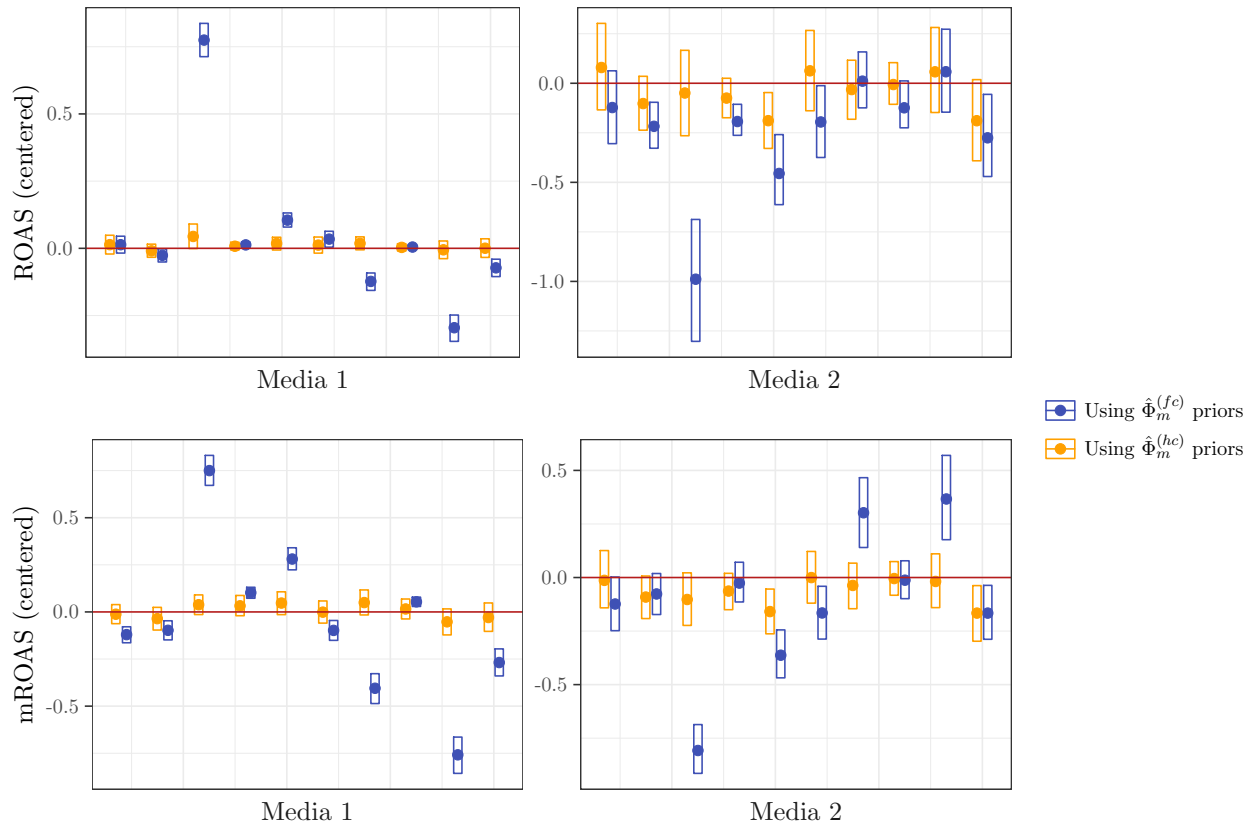
## D Additional figures



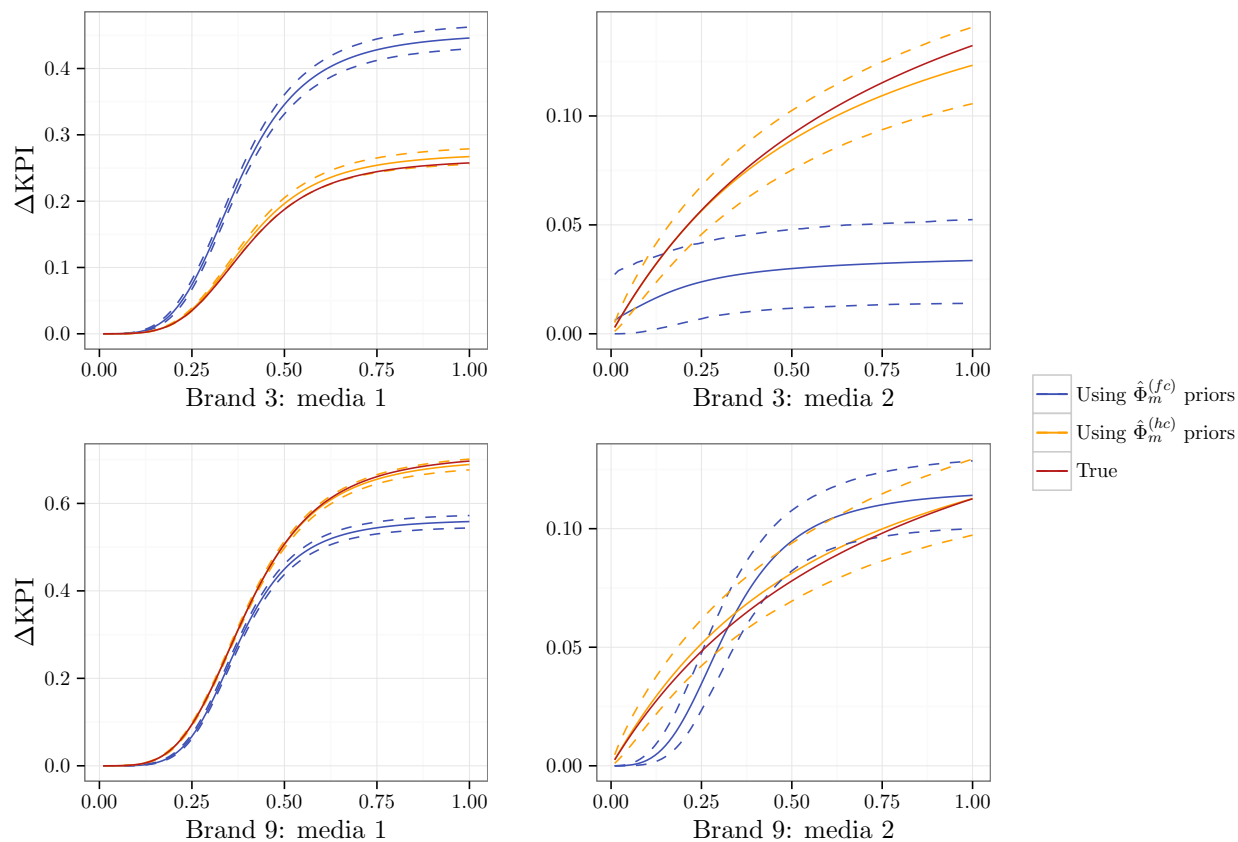
**Figure 27:** Comparison of estimated average ROAS (left) and mROAS (right) of media 1 for the 10 brands simulated in an example dataset in Section 5.2.2, after subtracting the true values. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.



**Figure 28:** Response curves of media 1 estimated by the brand-specific models using informative priors with 5th and 95th percentiles (in dotted lines), compared with that using weak priors with 5th and 95th percentiles (in dotted lines), and the true response curve (in red), for an example dataset in Section 5.2.2.



**Figure 29:** Comparison of estimated average ROAS (left) and mROAS (right) of media 1 and 2 for the 10 brands simulated in an example dataset in Section 5.3.3, after subtracting the true values. The bottom level of each bar represents the 5th percentile of the estimated values, while the top level the 95th percentile and the dot the posterior mean.



**Figure 30:** Response curves of media 1 and 2 estimated by the brand-specific models using informative priors from the HCM and the FCM with 5th and 95th percentiles (in dotted lines), compared with the true response curve (in red), for an example dataset in Section 5.3.3.

## References

- Ataman, M. B., Van Heerde, H. J., & Mela, C. F. (2010). The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research*, *47*(5), 866–882. doi:10.1509/jmkr.47.5.866
- Blake, T., Nosko, C., & Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: a large-scale field experiment. *Econometrica*, *83*(1), 155–174. doi:10.3982/ECTA12423
- Borden, N. H. (1964). The concept of the marketing mix. *Journal of advertising research*, *4*(2), 2–7.
- 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*.
- Dekimpe, M. G. & Hanssens, D. M. (1999). Sustained spending and persistent response: a new look at long-term marketing profitability. *Journal of Marketing Research*, *36*(4), 397–412. doi:10.2307/3151996
- Dekimpe, M. G. & Hanssens, D. M. (2000). Time-series models in marketing: past, present and future. *International Journal of Research in Marketing*, *17*(2-3), 183–193. doi:10.1016/S0167-8116(00)00014-8
- Ehrenberg, A. & Barnard, N. (2000). Problems with marketing’s ‘decision’ models. *ANZMAC*.
- Foll, M. & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, *180*, 977–993. doi:10.1534/genetics.108.092221
- Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge University Press.
- Gesztelyi, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B., & Tosaki, A. (2012). The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences*, *66*(4), 427–438. doi:10.1007/s00407-012-0098-5
- 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*.
- Leeflang, P. S. H., Bijmolt, T. H. A., Van Doorn, J., Hanssens, D. M., Van Heerde, H. J., Verhoef, P. C., & Wieringa, J. E. (2009). Creating lift versus building the base: current trends in marketing dynamics. *International Journal of Research in Marketing*, *26*(1), 13–20. doi:10.1016/j.ijresmar.2008.06.006
- 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

- Li, F.-F. & Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In *Proceedings of the 2005 ieee computer society conference on computer vision and pattern recognition (cvpr'05) - volume 2 - volume 02* (pp. 524–531). CVPR '05. Washington, DC, USA: IEEE Computer Society. doi:10.1109/CVPR.2005.16
- Little, J. D. C. (1979). Aggregate advertising models: the state of the art. *Operations Research*, 27(4), 629–667.
- Mela, C. F., Gupta, S., & Lehmann, D. R. (1997, May). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, 34(2), 248–261. doi:10.2307/3151862
- Nisbett, R. E. & Wilson, T. D. (1977). The halo effect: evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, 35(4), 250–256. doi:10.1037/0022-3514.35.4.250
- Palda, K. S. (1965). The measurement of cumulative advertising effects. *The Journal of Business*, 38(2), 162–179. doi:10.1086/294759
- Quandt, R. E. (1964). Estimating the effectiveness of advertising: some pitfalls in econometric methods. *Journal of Marketing Research*, 1(2), 51–60. doi:10.2307/3149922
- Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian statistics and marketing*. Chichester, England: Wiley-Interscience.
- Stan Development Team. (2015). *Stan modeling language user's guide and reference manual, version 2.10.0*. Retrieved from <http://mc-stan.org/>
- Stephens-Davidowitz, S., Varian, H., & Smith, M. D. (2017). Super returns to Super Bowl ads? *Quantitative Marketing and Economics*, 15(1), 1ffdfdfdfdf28. doi:10.1007/s11129-016-9179-0
- Sun, Y., Wang, Y., Jin, Y., Chan, D., & Koehler, J. (2017). Geo-level bayesian hierarchical media mix modeling. *research.google.com*.
- 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>
- Yildiz, I. B., von Kriegstein, K., & Kiebel, S. J. (2013). From birdsong to human speech recognition: bayesian inference on a hierarchy of nonlinear dynamical systems. *PLOS Computational Biology*, 9. doi:10.1371/journal.pcbi.1003219