Challenges And Opportunities In Media Mix Modeling

David Chan and Michael Perry

Google Inc.

Abstract

Advertisers have a need to understand the effectiveness of their media spend in driving sales in order to optimize budget allocations. Media mix models are a common and widely used approach for doing so. We outline the various challenges such models encounter in consistently providing valid answers to the advertiser's questions on media effectiveness. We also discusses opportunities for improvements in media mix models that can produce better inference.

1 Introduction

Media mix models (MMMs) are statistical models used by advertisers to measure the effectiveness of their advertising spend and have been around in various forms since the 1960s (Borden, 1964; McCarthy, 1978). MMMs use aggregate historical time series data to model sales outcomes over time, as a function of advertising variables, other marketing variables, and control variables like weather, seasonality, and market competition. Metrics such as return on advertising spend (ROAS) and optimized advertising budget allocations are derived from these models, based on the assumption that these models provide valid causal results.

MMMs attempt to answer causal questions for the advertiser. For example: 1) What was my ROAS on TV last year? 2) What would my sales be if more or less money were spent next year? 3) How should my media budgets be allocated to maximize sales? Typically though, MMMs are regression models based on a limited amount of aggregated observational data and such models produce correlational, not causal results. It is only under certain narrow conditions that these estimates can be considered causal.

We discuss the challenges associated with MMMs in attempting to provide trustworthy answers to these types of questions and covers potential opportunities to improve the ability of MMMs to provide valid inference. An overview of the paper is as follows.

Section 2 motivates MMM analysis by first describing two common approaches that could be used to answer causal questions and highlights why they are impractical or infeasible in the context of answering all the questions an advertiser may have about the effectiveness of their advertising channels. Hence the reason advertisers turn to MMMs.

Section 3 gives context to the reader when discussing challenges and opportunities in MMMs, by introducing a regression model form that is widely used in MMMs today. The section describes

a particular model specification which attempts to address carryover and diminishing returns and briefly covers the type of data that is typically available to the modeler to fit such models.

Section 4 goes over the challenges that a modeler may face in attempting to produce reliable estimates from a regression MMM. These challenges can be grouped broadly into three main areas; data limitations, selection bias, and modeling.

Section 5 turns to a more optimistic note and highlights some opportunities to address the challenges pointed out in Section 4. We conclude with some final remarks around the need to acknowledge uncertainty in the modeling process, the need for transparency between the modeler and the end user of the model results, and the need to educate end-users of MMMs on their capabilities and limitations.

2   Causal Inference

This section discuss two approaches that can be used by the advertiser to answer causal questions around their advertising effectiveness. It covers why these approach are infeasible or impractical for advertisers, which in turn motivates why advertisers turn to using MMMs.

2.1   Randomized experiments

The generally accepted gold standard for answering causal questions is to perform a randomized experiment. A randomized experiment answers the question of what will happen if an advertiser did action X, by randomly splitting a population into a test group, where the action X is performed and a control group, where no action is taken. The randomization controls for all other sources of variation, so that statistically, the only difference between the test and control groups is the action X. We can then say that action X *caused* any observe change in outcome between the two group. See Hinkelmann and Kempthorne (2008) for a general introduction to experiments.

Randomized experiments can be performed at various levels, such as at a user-level, store-level or geographic-level (see Vaver and Koehler (2011)). The level of the experiment is constrained by the granularity at which it is possible to target action X and to track outcomes. While randomized experiments are the gold standard, certain factors prevent them being used more widely to provide answers an MMM might give. In order to answer a similar set of questions that MMMs answer, the advertiser may need to run many experiments with many different conditions over time.

For example, suppose an advertiser expects a MMM to inform on the effectiveness of an ad channel at various levels of ad spend, not just at one level of ad spend. The advertiser also uses a MMM to optimize media budgets which will require many experiments to answer. For most, if not all advertisers, this is infeasible. Other barriers to adoption of randomized experiments include technical hurdles in implementation, lost opportunity costs from having a control group, costs of having a test group, and weak advertising effects which may require very large sample sizes (see Lewis and Rao (2015)).

2.2   Potential outcomes

If advertisers can't run randomized experiments at scale and are therefore reliant on historical data, then they could try to derive causal conclusions via the the potential outcomes framework,

also known as the Rubin causal model for causal inference (see Imbens and Rubin (2015)). For discussions of this potential outcome framework in the context of digital advertising, see Chan, Ge, Gershony, Hesterberg and Lambert (2010) and Lambert and Pregibon (2007).

Suppose we want to estimate the ROAS of a single ad channel for an advertiser and the ad channel is either on or off during each time period t. Let y_t be the sales for time period t and x_t be an indicator of whether the ad channel was on or off. For any time period t, there are two potential sales outcomes:

$$y_t^1 \quad : \quad \text{potential outcome if ad channel turned on}$$
$$y_t^0 \quad : \quad \text{potential outcome if ad channel turned off}$$

The causal ROAS over the time periods is given by

$$\mathbb{E}(y_t^1 - y_t^0).$$

The challenge is that only one potential outcome is observed for each time period. A naive estimate would be to compare sales in the periods when the ad channel was on, to sales in the periods when the ad channel was off. This would give:

$$\mathbb{E}(y_t \mid x_t = 1) - \mathbb{E}(y_t \mid x_t = 0) \quad = \quad \left[\mathbb{E}(y_t^1 \mid x_t = 1) - \mathbb{E}(y_t^0 \mid x_t = 1) \right] +$$
$$\left[\mathbb{E}(y_t^0 \mid x_t = 1) - \mathbb{E}(y_t^0 \mid x_t = 0) \right].$$

The $\mathbb{E}(\cdot)$ operator here indicates the sample average in the data. The term in the first set of square brackets represents the causal effect of interest during the treated time periods. It is the difference between the average sales that occurred when $x_t = 1$ and the average sales that would have occurred during the same time periods if $x_t = 0$.

The term in the second set of square brackets represents selection bias. In a randomized experiment, this term would have an expected value of zero by design. The phrase 'selection bias' is used here to refer to any biases in the treatment selection mechanism that are also correlated with the outcome (sales), which in this context means potentially everything that determines ad spend. Selection bias may occur due to actions the advertiser takes, actions that potential purchasers take or actions that competitive advertisers take. For example, if the advertiser turns ads on according to the seasonality in demand (selection bias), then

$$\mathbb{E}(y_t^0 \mid x_t = 1) \neq \mathbb{E}(y_t^0 \mid x_t = 0).$$

It is this selection bias that makes trying to answer causal questions with observational data one of the most demanding problems in applied statistics. This selection bias needs to be accounted for in order to produce a valid causal result. One way to control for selection bias is through the use of the matching estimator described below.

Let $\mathbf{z}_t$ be the vector of control variables which could potentially affect sales at time t. These control variables would include other ad channels. The matching estimator can be defined as

$$\mathbb{E}(y_t^1 - y_t^0 \mid x_t = 1) = \sum_{\mathbf{z} \mid x_t = 1} \left[\mathbb{E}(y_t \mid x_t = 1, \mathbf{z}) - \mathbb{E}(y_t \mid x_t = 0, \mathbf{z}) \right] \times P(\mathbf{z}_t = \mathbf{z} \mid x_t = 1). \quad (1)$$

$P(\cdot)$ indicates the empirical probability. The matching estimator assumes that there are observations where $x_t = 0$ for all combinations of control variable values that occur in the case when $x_t = 1$. However this may not always be a valid assumption, for example in the case of correlated media variables. As the number of control variables increase, the number of possible combinations of different control values increases exponentially. In a typical MMM situation, where the number of data points is low, this assumption is nearly impossible to satisfy in order to use the matching estimator to provide the full range of ROAS estimates.

With both randomized experiments and use of matching estimators being infeasible or impractical, advertisers often turn to regression models to answer their questions around advertising effectiveness.

3 Regression

Before going into details on the regression model, we take a moment to discuss the type of data that is generally available in a typical MMM dataset and note some challenges even in the data definition and collection process.

3.1 MMM Data

Data used to fit MMMs are historical weekly or monthly aggregated national data, although geo-level or even store level data can be used. The data includes:

- **response data**, which are typically sales but can be other KPIs such as store visits,

- **media metrics** in the different media channels, such as impressions, clicks, GRPs, with media spend being the most common,

- **marketing metrics** such as price, promotion, product distribution, and

- **control factors** such as seasonality, weather and market competition.

MMMs are typically produced on contract by a third-party for an advertiser although some advertisers do them in-house. In either case, MMM presents formidable challenges in data collection and data quality. Apart from the logistics of collecting and collating all the necessary data for an MMM, the data itself could be of varying quality and granularity. First, in order to fit an MMM, the response data needs to be at the same level as the ads spend data. So while an advertiser may have very precise SKU level or store level sales data, advertising is typically done at a brand or product level and usually over an entire country. There is often an unclear mapping between the product for which the MMM is desired and the advertising spend. Halo effects from related brand advertising are also difficult to account for in the MMM. Therefore, assignment of particular ad campaigns as relevant to particular SKUs is partially subjective and could lead to error through under- or over-attributing advertising to particular SKUs.

Data used to fit MMMs usually end up at the lowest common denominator of geographical granularity. Meaning that whichever ad channel is available at the least granular level determines the granularity of the whole model. Recent work by Sun, Wang, Jin, Chan and Koehler (2017) on

Bayesian hierarchical models show that in some situations, it is possible to have a mix of data in the same model.

For the response data, typically sales, advertisers generally have a robust data collection mechanism in place. In most industries where MMMs are widely used, there are also third-party providers of sales, price and promotion data. So in addition to the advertiser's internal source of sales, price and promotion data for its own brands, advertisers may rely on these third parties to also provide competitor data. However, it is often difficult to get competitor variables for pricing, promotion and distribution and these variables are often omitted in MMMs.

The ads exposure data is more challenging to collect, as ad campaigns are often planned and executed through several intermediaries, such as agencies. Advertisers may have to go through many separate entities in order to get data on the ad campaigns they have funded. The complexity of this process can easily lead to missing or misinterpreting some important data. And while the ad dollars spent can be obtained, the ad exposure data can be more difficult to collect and are often estimated in different ways depending on the media and the vendor providing the data. This is particularly true of offline media. For example in print media, while circulation numbers for publication can be provided, it is not always a good proxy for the actual number of people that see the ad. Both the complex collection process and the use of proxy variables is another potential source of inaccuracy in the data.

3.2   A general regression model

A regression MMM specifies a parameterized sales function chosen by the modeler, e.g.,

$$y_t = \mathrm{F}(\mathbf{x}_{t-L+1}, \ldots \mathbf{x}_t, \mathbf{z}_{t-L+1}, \ldots, \mathbf{z}_t; \Phi) \quad t = 1, \ldots T, \tag{2}$$

where y_t is the sales at time t, $\mathrm{F}(\cdot)$ is the regression function, $\mathbf{x}_t = \{x_{t,m}, m = 1, \ldots, M\}$ is a vector of ad channel variables at time t, $\mathbf{z}_t = \{z_{t,c}, c = 1, \ldots, C\}$ is a vector of control variables at time t and Φ is the vector of parameters in the model. L indicates the longest lag effect that media or control variables has on sales.

In order to enable optimization of media budgets and to capture diminishing returns, the response of sales to a change in one ad channel can be specified by a one dimensional curve which is called the *response curve* for that channel. A common approach is to have the media variables enter in the model additively. Additionally, the control variables are often parameterized linearly with no lag effects, so that an MMM may look like:

$$y_t = \sum_{m=1}^{M} \beta_m f_m(x_{t-L+1,m}, \ldots, x_{t,m}) + \boldsymbol{\gamma}^T \mathbf{z}_t + \epsilon_t, \tag{3}$$

β_m is a channel-specific coefficient, $\boldsymbol{\gamma}$ is a column vector of coefficients on the column vector of control variables $\mathbf{z}_t$. $f_m(\cdot)$ captures nonlinearity in the effect of a single channel due to reach/frequency effects. The response y_t could also be transformed prior to use in the model above. ϵ_t is an error term that captures variation in y_t unexplained by the input variables. Such a function can be converted to a likelihood and fit to data using maximum likelihood estimation, Bayesian inference or other methods. See Jin, Wang, Sun, Chan and Koehler (2017) for further discussion of model fitting approaches.

While in general, it can't be claimed that results from a model such as above are causal, estimates from these models are generally more trustworthy when

- There are enough data needed to estimate all the parameters in the model

- There is useful variability in the advertising levels and control variables

- Model inputs vary independently

- The model accounts for all the important drivers that might impact sales

- The model captures the causal relationship between variables

4   Challenges

This section steps through a number of different issues that challenge the reliability of results from MMMs fitted to observational data. These are frequently encountered by the modeler, but are often not openly acknowledged or discussed with the end-user of the MMM. There are three broad areas of challenges involving data limitations, selection bias, and modeling.

4.1   Data limitations

4.1.1   Limited amount of data

The data quantity available to the modeler is often limited. A typical MMM dataset, consisting of three years of national weekly data is only 156 data points. From this, the modeler is expected to produce a MMM often with 20 or more ad channels. If flexible functional response forms are required as in Equation (3) for each ad channel, which is typically the case, the number of parameters in the model could exceed the number of data points available. To adequately model a lagged effect and a diminishing return might require 3 - 4 parameters for each channel. A rule-of-thumb for a minimum number of data points for a stable linear regression (putting aside whether causal effects are well-estimated) are 7-10 data points per parameter, of which typical MMMs fall short.

4.1.2   Correlated input variables

Advertisers often allocate their spend across ad channels in a correlated way, which may make sense from the perspective of maximizing ad effectiveness. Additionally, these correlated ad choices can interact with other marketing choices to lead to broad sets of correlated input variables. For example, a particular ad channel can only be observed at a high level when another marketing variable is at a high level or only during a certain season.

When fitting a linear regression model, highly correlated input variables can lead to coefficient estimates with high variance. This in turn, can lead to bad attribution of sales to the ad channel. For an illustration, consider a simplified version of a typical situation faced by a modeler, shown in Figure 1.

The figure has two fitted response surfaces, that each fit the data well and that have different slopes with respect to each ad channel. In this dataset, the modeler may find that many surfaces provide
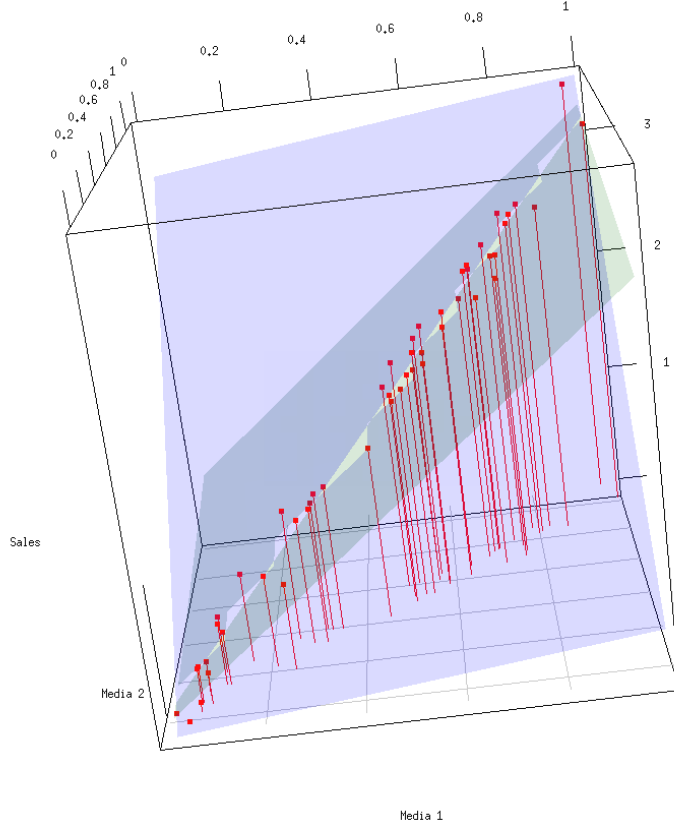
Figure 1: Two estimated response surface in the presence of correlated variables. Sales is on the z-axis, and there are two ad channels, with the spend of each on the other two axes. The ad channel spend levels are strongly correlated with each other and each plane fits the observed data well despite having different slopes.

good predictive accuracy of out-of-sample sales, but perform poorly when the advertiser deviates from previous spend patterns. This is because the data contains little information about sales outcomes when one of the ad channels moves independently of the other. Another consequence is that the estimated relationship can change radically due to small changes in the data or the addition or subtraction of seemingly unrelated variables in the model.

4.1.3   Limited range of data

Advertisers have often settled on a range of ad spend that accommodates their business needs. This is a problem when models are fit to this limited historical data and the model is expected to provide insights outside of the range of this data. For example, the advertiser may be interested in knowing the answer to 'what if I double my ad spend next year?'.

Fitting a model on a limited range of data exposes the advertiser to high levels of extrapolation uncertainty as shown in Figure 2. The right panel shows four fitted response curves, which all fit equally well to the observed range of data, but show very different sales outcomes when ad spend is increased.

Another common form of extrapolation is 'what happens if I cease my ad spend?'. This extrapola-
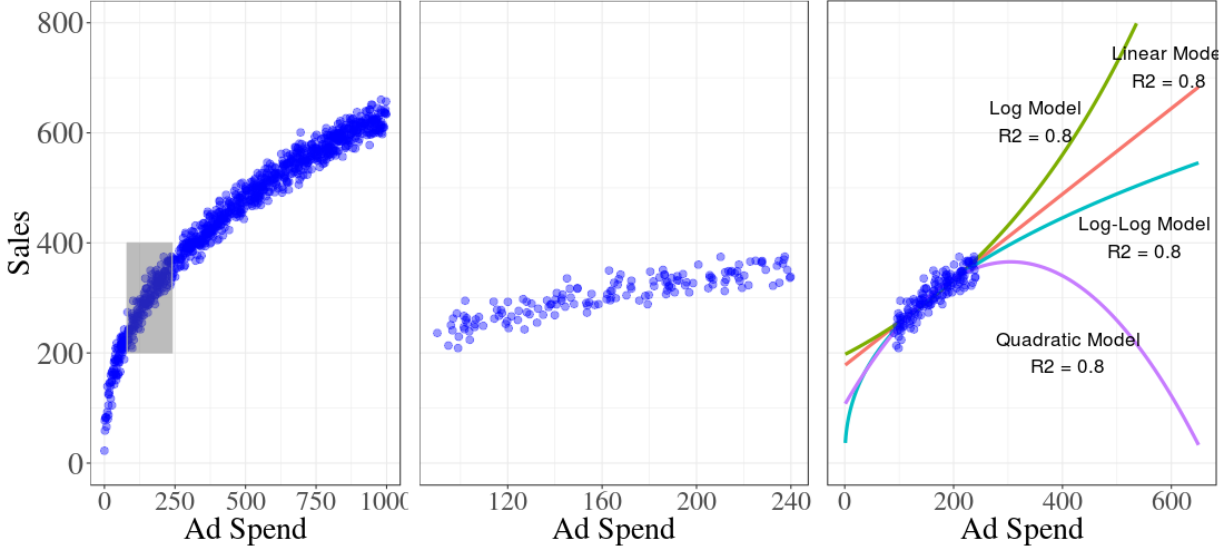
Figure 2: Illustration of extrapolation uncertainty when models are fitted to a limited range of data. The left panel in the figure represents what the response curve might be if the advertiser had spent in those ranges. The middle panel would be what is available to the modeler due to the advertiser spending only in a limited range. The right panel shows four fitted response curves.

tion occurs as the advertiser desires to know what the average ROAS is for an ad channel, which requires extrapolation back to zero spend. In such situations, the fitted model might provide reasonable estimates of the marginal ROAS, due to the data available around current spend levels, but poor estimates of the average ROAS.

4.2 Selection bias

Selection bias from MMMs perhaps represents the largest hurdle to MMMs providing valid estimates of advertising effectiveness. Selection bias occurs when an input media variable is correlated with an unobservable demand variable d_t, which in turn drives sales. When this variable d_t is omitted from the regression, the model has no way to attribute sales between those due to the media channel or to the underlying demand. We discuss selection bias that occurs in a number of different situations.

4.2.1 Ad targeting

Selection bias arises in the case of ad targeting as the ads are targeting a segment of the population which has already shown an interest in the product. This underlying interest or demand of the targeted population is not observable and not included in the model. Ad targeting is particularly acute in digital channels.

Two examples of ad targeting which can have severe selection bias are re-marketing ads and paid search ads. Re-marketing ads are targeted to a population of users that has already shown interest in the product through a visit to the advertisers website. Paid search ads are targeted to a population of users that has shown interest via a related query on a search engine.

4.2.2 Seasonality

Selection bias arises when media is targeted towards the seasonality in demand. An example of this is the demand for cold medicine, which we expect to fluctuate with the cold and flu season. Advertisers would increase or decrease their ad spend according to their understanding of the underlying demand.

MMMs do generally try to control for seasonality by including proxies for the demand seasonality in the model. Common proxies include category sales, but there is no guarantee these proxies are an accurate reflection of the underlying demand for the product. Inaccurate proxies in turn lead to biased estimates of the regression coefficients.

Other examples of targeted media lead to omitted variable bias, including advertisers choosing particular times and places to advertise based on fashions or events that can't be modeled accurately; or advertising in reaction to competitor activities that the modeler may not have data to model but that directly affect sales.

4.2.3 Funnel effects

Selection bias arises when there are funnel effects in the media and the model is mis-specified. When an ad channel also impacts the level of another ad channel, using a model like in Equation (3), which simultaneously estimate the impact of all ad channels in one equation, will lead to biased estimates. An example is a TV campaign driving more related queries, which in turn increase volume of paid search ads. For a derivation of the size of the bias, see Angrist and Krueger (1999). The underlying reason for the bias in the mis-specified model is that the downstream ads were affected by the TV ads. In assessing the ROAS of TV, the linear regression model does not account for the changes in paid search ads caused by TV.

Downstream ads should not be included with exogenously-determined ads in a single regression equation. Alternatives include graphical models (Pearl (2009)) and structural equation models. However, the problem isn't only a matter of model form. Both these alternatives require estimating the causal effect of the upstream ads on the downstream ads, which has just as stringent data requirements as estimating the effect of an ad channel on sales.

A broader problem is that many important variables may be partially determined by the advertiser's own ad choices, which means that including them as control variables leads to the same problem as in the TV and search example, although through a different mechanism. For example, price clearly has a strong effect on sales and therefore it is important to control for it in some way. However, if the advertiser makes some pricing decisions in response to its own advertising decisions, such as supporting an ad campaign with a discount, then including price in a regression equation will not properly control for this relationship.

4.3 Model selection and uncertainty

The modeler faces the challenge that the functional form of $F(\mathbf{x}_t, \mathbf{z}_t; \Phi)$ and the members of $\mathbf{x}_t$ and $\mathbf{z}_t$ to include in the model are ambiguous. Uncertainty of the functional form for $F(\mathbf{x}_t, \mathbf{z}_t; \Phi)$ has been long-acknowledged (see Quandt (1964)) and is due to the complexity of the sales response process.[1]

---

[1]MMMs fall into the broader category of demand modeling which has a long history in the field of economics (e.g. Deaton and Muellbauer (1980); Berry, Levinsohn and Pakes (1995)). The demand modeling literature provides

9

We further discuss the inherent challenges associated with making these modeling assumptions and the impact such modeling assumptions have on the final output.

4.3.1 Model selection

In order to decide on a model form, the modeler may use a model selection process based on accuracy metrics such as R^2 or predictive error, and to treat the model with the highest accuracy as a valid causal model. However, the number of data points is usually low relative to the dimension of the input space $(\mathbf{x}_t, \mathbf{z}_t)$.

The problem of model selection based on predictive accuracy is further exacerbated by the fact that MMM datasets typically have low signal-to-noise, where the ad variables are the signals of interest. The explanatory power of a few input variables–such as seasonal proxies, price and distribution–are often good enough to achieve high predictive power in absence of ad spend variables. This makes it easy to find models with high predictive accuracy by trying different specifications. Concrete examples of specification choices include whether to choose sales volume or log sales volume as the outcome variable, how to control for price and how to capture lag effects and diminishing returns. There are dozens of such choices underlying any single MMM.

In real datasets that we have encountered, the variance in the sales is usually much greater than the variance in the media spend. Lewis and Rao (2015) discusses the difficulty of measuring ad effectiveness in this situation in the context of digital advertising. Given that ad spend is a weak signal, the question then is what role the ad variables play in model selection. Generally a focus on predictive accuracy will not allow a modeler to eliminate models that produce inaccurate ROAS estimates.

4.3.2 Model uncertainty

Consider an example of model uncertainty in a real MMM dataset. The dataset is fairly typical in that it includes weekly data for the US, collected over several years. Although there are roughly a dozen ad channels, four are of most interest and this discussion focuses on those four ad channels. Five equally plausible models have been fitted. There could have been other plausible models and from discussions with practitioners, there typically are.

Model 1 is a baseline model and each of the others is a variation of Model 1. Model 2 uses a different proxy variable to capture seasonality. Model 3 only includes the most recent two years in the dataset because the response surface was believed to have been different before that. Model 4 excludes branded search ads as an ad channel, under the hypothesis that those estimates are too biased to be useful and that their effects are best captured by our control variables, which include search query volume. Note that branded search ads are not one of the channels focused on in the figures, so this can be viewed as the decision to exclude a nuisance parameter. Model 5 includes a linear time variable to capture a time-trend in the data.

Each model achieves an R^2 value of either 0.98 or 0.99. Models 1-4 achieve an out-of-sample MAPE of 6-8% in cross-validation using the last 12 weeks of data, with Model 5 doing worse (15%) because the time trend didn't remain consistent out-of-sample.[2]

---

a starting point for considering what variables to include. However, it does not provide strong guidance on the functional form of the model or the proper control variables.

[2]The scope for meaningful cross-validation of an MMM is typically limited due to small datasets and the time-series

The two figures show relevant results for a decision-maker. Figure 3 shows predicted sales under the current media mix and varying total ad budgets. The x-axis shows varying total media budget levels and the y-axis shows predicted sales; each colored line represents the prediction of one of the models. The models differ in their sales predictions by up to 50% and Model 2 and 3 have much higher slopes than the others, indicating a higher total ROAS across ad channels than implied by the other models. So despite fitting the data equally well, the models disagree on what sales are achievable under different ad spends.
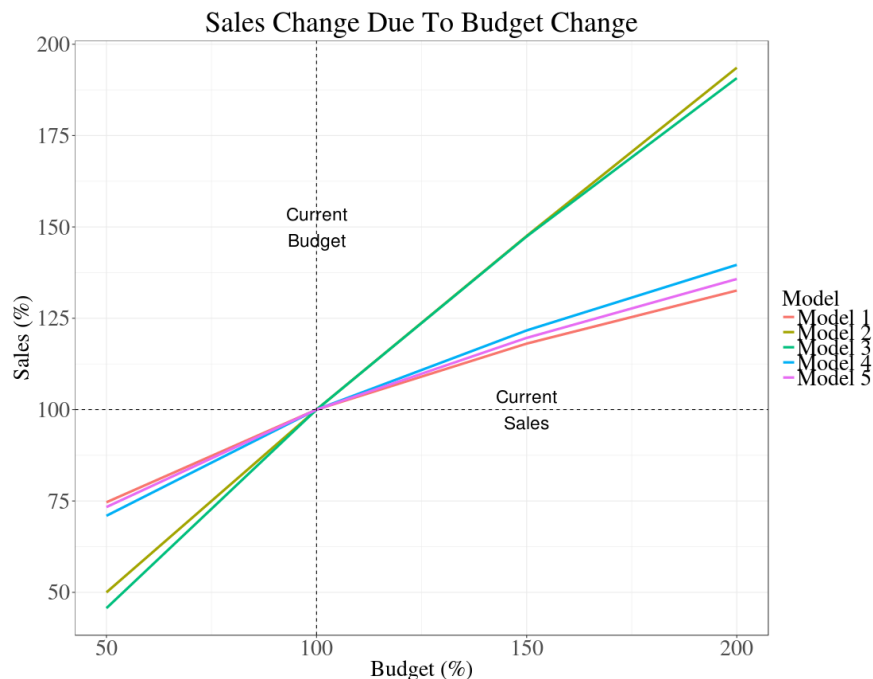
Figure 3: Predicted sales under current media mix and varying budgets in five plausible models.

Figure 4 shows the recommended budget allocations across media under varying total ad budgets. The x-axis again shows total ad budgets and the y-axis shows the ad spend for a given ad channel; each colored line is recommended spend for a particular channel under the varying total budget and the five panels show those recommendations under the five models. The models don't agree on how to allocate budget across media channels. Most importantly, they disagree on ad channel A, which is by far the largest channel for this advertiser. Model 2 implies that the advertiser should allocate the largest budget to channel A across the whole range of total ad budgets, while the other models do not agree. They do not agree about channels B, C or D either.

In this example, there are five plausible models that fit the data well, but that lead to different conclusions about the overall value of the ad channels and about how budget should be allocated.

5   Opportunities

In the previous section, we have covered the many challenges MMMs face in consistently providing valid inference to the advertiser. There are opportunities to improve the reliability of MMMs.

_____

nature of the data which doesn't allow for simple random sampling to produce a validation set.
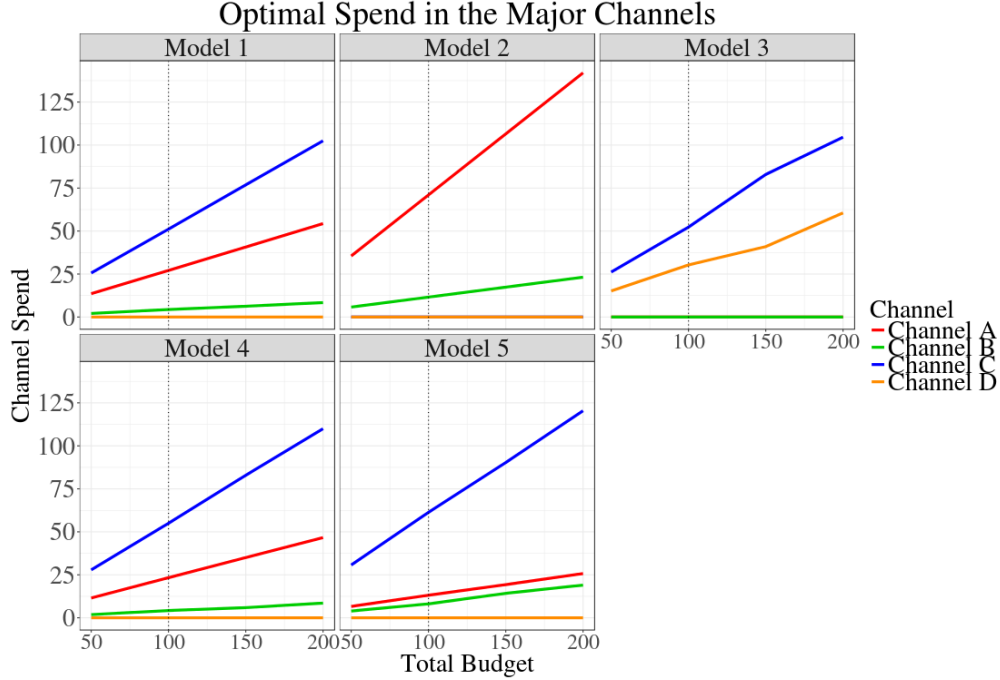
Figure 4: Optimal spend in major channels under varying total media budgets. Note that the budgets for the four ad channels in each panel don't add up to the total budget on the x-axis because only the top four ad channels are shown. Also, in the second and the third panels, two of the ad channels overlap so it's hard to distinguish their lines. In both cases, they overlap close to the line $y = 0$.

Opportunities exists in three broad areas: 1) Better data 2) Better models 3) Model evaluation through simulations.

Better data comes in two flavors. More accurate data and more granular data. As noted earlier, the collection process for data to use in a MMM is itself a challenge. A common approach has been for the advertiser to source their ads exposure data through the agencies which manage their ad campaigns. This could lead to data inconsistencies and inaccuracies. We propose that the industry adopt a standardized data and reporting format for MMMs. A tighter coordination between the intermediaries who execute the ad campaigns and the publishers on whose platforms the ads are served can enable more accurate data collection for the purpose of MMMs.

Apart from better data, better models also have a potentially large impact on the reliability of MMM. We discuss a number of model approaches which can improve the reliability of MMMs.

5.1 Bayesian modeling

Although the model in Equation (3) can be fitted via a number of different methods, we have found that a Bayesian framework for model fitting can offer improvements. Benefits of the Bayesian approach include:

- Ability to use informative priors for the parameters, where the informative priors can come from a variety of sources

- Ability to handle complicated models

- Ability to report on both parameter and model uncertainty

- Ability to propagate uncertainty to optimization statements

See Jin et al. (2017) for a discussion of the Bayesian framework that does the above. A strong candidate method for developing informative priors is by fitting category level MMMs across brands within a category using a Bayesian hierarchical model. We discuss this type of model further below.

5.2 Category models

A typical dataset for a single brand, consisting of national-weekly aggregated data, generally has too limited quantity and information to reliably estimate the parameters of the MMM. A way of introducing more data and independent variability into the model is to pool data from different brands within the same product category. We expect more independent variability in the category dataset as different brands tend to act independently of each other in terms of their advertising preferences and execution.

See Wang, Jin, Sun, Chan and Koehler (2017) for further discussion on category models and how they can be fitted using a hierarchical Bayesian approach. The brand level results can be used directly or the posteriors results from the category model can be used as informative priors for a Bayesian MMM for a brand that wasn't in the original category model. Wang et al. (2017) demonstrate with both simulation and real case studies that such an approach can improve parameter estimation and reduce the uncertainty of model prediction and extrapolation.

5.3 Geo models

Another approach to enable a higher sample size and more independent variation in the data is to use sub-national data when available. This can be in the form of city, county, province or state depending on the country. Along with geo-level data, it is possible fit geo-level models to better take advantage of the data available. Sun et al. (2017) demonstrate a hierarchical Bayesian model can be fitted to geo-level data and show that such models tend to reduce the variance of the parameter estimates. Increasing the information available may also reduce the amount of model uncertainty by reducing the number of models that fit the data well. Parameter estimates can be reported on an individual geo-level or at the national-level and budget optimization can be done at either level as well.

A further advantage of such a model is its ability to handle a mix of both geo and national level data. The national level data is disaggregated to the geo-level based on population metrics and the hierarchical Bayesian model can be fitted as usual. A further extension of such an approach is to fit category-geo-level models when such data is available.

5.4 Control variables

Selection bias has been discussed at length earlier in the paper as the largest hurdle to obtaining causal estimates from a MMM. Paid search advertising represents a severe form of selection bias in estimating ROAS. This selection bias occurs because the paid search advertising is targeting related query volume which is being driven by an unobservable demand metric which also affects sales.

In order to address this issue, the search volumes for relevant queries can be used as a control variable for the component of underlying demand from people likely to perform relevant product-related searches. Chen et al. (2017) shows how search query volume can be used in a MMM to make the estimate of paid search impact be more causal. This approach is not restricted to just paid search, but can be considered for other media where a control variable for the underlying demand can be utilized. For example, in the case of remarketing ads, the volume of website visits which triggered the re-marketing ads may act as a suitable control variable.

5.5 Graphical models

Figure 5 is a simple example of the media funnels one could expect when fitting MMMs. To best capture the complicated conditional dependencies due to issues such as funnel effects of media, a graphical model can be useful. A graphical model is a way of expressing the dependencies between variables, both observed and unobserved (Pearl (2009)). It may allow for clearer communication about the underlying causal structure of the media effect. This allows a more realistic model than what is typically achieved through a one- or two-stage regression framework. Graphical models can be fitted for example using the STAN package (Stan Development Team (2015).
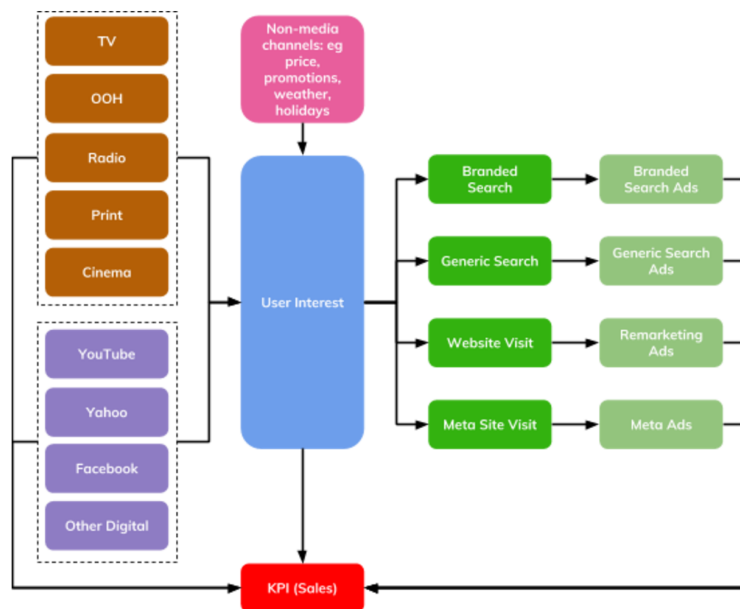
Figure 5: Example of media funnel effects on sales.

5.6 Model validation

Finally, it is useful to evaluate how accurately different types of models perform under different hypothesized scenarios such as different types of selection bias. This can be done using a simulator, which is a software package that produces simulated MMM datasets defined by the user to capture important aspects of the real world. See Zhang and Vaver (2017) for discussion of one such simulator.

While experiments are a source of ground truth, they usually provide a point estimate. Simulation

allows for more control of the conditions of the marketing environment, and provides a flexible and inexpensive environment from which to conduct virtual experiments. This provides insight into the marketing environment and how well different models capture different aspects of it.

Simulation helps the modeler build models that are robust to different assumptions. For example, a series of simulated datasets with increasing lagged media effects, or increasingly correlated media variables could be produced and the modeler could fit MMMs to these datasets to understand the models changing performance as the underlying environment changes. Because the datasets are simulated, the true underlying ROAS is known and the model output can be compared to a ground truth.

Advertisers could also incorporate simulated data as part of their request-for-proposals process. Simulated data, where the ground truth is known but withheld, can be provided to third parties in order to evaluate the performance their modeling process.

6   Final remarks

It is important for the modeler to be forthright about the types of questions a MMM can and can't answer. It is especially important to also acknowledge the uncertainty in the modeling process and the need for transparency between the modeler and the end user of the model results. As part of that transparency, the modeler should be forthright about the capabilities and limitations of MMMs.

If at all possible, metrics around the parameter uncertainty and the model uncertainty should be discussed with the end user. They should be aware of the process by which a model was chosen, the amount of model uncertainty and the degree to which final results are data-driven versus driven by the modeler's chosen priors and filters.

Acknowledgment

References

Angrist, J. & Krueger, A. B. (1999). Empirical strategies in labor economics. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 1277–1366).

Berry, S., Levinsohn, J. & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890. doi:10.2307/2171802

Borden, N. H. (1964). The concept of the marketing mix. *Journal of advertising research*, *4*(2), 2–7.

Chan, D., Ge, R., Gershony, O., Hesterberg, T. & Lambert, D. (2010). Evaluating online ad campaigns in a pipeline: causal models at scale. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 7–16). KDD '10. Washington, DC, USA: ACM. doi:10.1145/1835804.1835809

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*.

Deaton, A. & Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, *70*(3), 312–326.

Hinkelmann, K. & Kempthorne, O. (2008). *Design and analysis of experiments, volume 1, introduction to experimental design* (2nd ed.). John Wiley and Sons.

Imbens, G. W. & Rubin, D. M. (2015). *Causal inference for statistics, social, and biomedical sciences: an introduction* (1st ed.). Cambridge University Press.

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*.

Lambert, D. & Pregibon, D. (2007). More bang for their bucks: assessing new features for online advertisers. In *Proceedings of the 1st International Workshop on Data Mining and Audience Intelligence for Advertising* (pp. 7–15). ADKDD '07. San Jose, CA: ACM. doi:10.1145/1348599.1348601

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

McCarthy, J. E. (1978). *Basic marketing: a managerial approach* (6th ed.). Homewood, Il: R.D. Irwin.

Pearl, J. (2009). *Causality: models, reasoning and inference* (2nd ed.). Cambridge University Press.

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

Stan Development Team. (2015). *Stan modeling language user's guide and reference manual, version 2.10.0*. Retrieved from http://mc-stan.org/

Sun, Y., Wang, Y., Jin, Y., Chan, D. & Koehler, J. (2017). Geo-level bayesian hierarchical media mix modeling. *research.google.com*.

Vaver, J. & Koehler, J. (2011). Measuring ad effectiveness using geo experiments. *research.google.com*. Retrieved from https://research.google.com/pubs/pub38355.html

Wang, Y., Jin, Y., Sun, Y., Chan, D. & Koehler, J. (2017). A hierarchical bayesian approach to improve media mix models using category data. *research.google.com*.

Zhang, S. S. & Vaver, J. (2017). Introduction to the Aggregate Marketing System Simulator. *research.google.com*.