Building and Validating Media Mix Models

by Erica Mason | emason@thirdlove.com

THIRDLOVE

1. Introduction

In e-commerce, measuring and understanding marketing effectiveness holistically across channels is essential to the success of a company. Doing this properly is complex as e-commerce companies may advertise on multiple platforms, both online or out of home, or work with multiple agencies over time. Further, with increasing concerns for customer privacy and data regulation, attaining this holistic understanding has become more difficult as advertising platforms are, rightly, less inclined or able to share customer level data. While all businesses have varying objectives, the underlying goal for most marketers is the same: understand the true return on investment comparability for all marketing channels. Typically this is achieved through some combination of click attribution, experimentation, and media mix models (MMMs). Each have their own strengths and weaknesses for understanding ground truth and cross-channel effectiveness, which we will expand on. This paper will offer a case for marrying two approaches, time series MMMs and experimentation, to provide a validation methodology that can yield a higher degree of accuracy than relying on any one approach in isolation. We will show that in particular, building an MMM alone, without validating the model against ground truth results from experiments, can lead to incorrect conclusions. In the example provided, three models built from the same dataset and slightly different assumptions conclude that a different marketing channel has the lowest cost per acquisition. Ultimately the goal of this paper is to offer a solution that can help advance the precision of cross-channel measurement throughout the marketing science community. Industry innovation and standardization requires cooperation and willingness to test new methodologies. We hope this white paper provides a meaningful step in that direction.

The paper is laid out as follows. Section 2 defines measurement terminology that will be used throughout the rest of the paper. Section 3 gives an overview of three methods for measuring marketing efficiency: click attribution, experimentation, and MMMs. Section 4 describes how to build an MMM and the insights that can be gleaned from the model. Section 5 describes how to validate an MMM using backtest and experiment results by way of an example with a fictitious company and dataset. Section 6 concludes the paper. We provide links to sample code for the reader in the Appendix.

2. Terminology

There are many ways to think about marketing spend and many metrics that are used to quantify the efficiency of the spend, namely, Customer Acquisition Cost, Cost Per Thousand Impressions, Cost Per Click, Cost per Order, Cost Per Acquisition, or another term entirely. This can get confusing and definitions can get muddled. For our purposes, we will be using the following three metrics with respect to each advertising channel:

- **Cost Per Acquisition (CPA):** The total dollars spent on a channel divided by the total acquisitions attributed to this same marketing channel (more details on how to find this denominator to follow in the next section). This gives credit to the marketing channel regardless of whether those acquisitions may have occurred anyway in the absence of marketing, e.g. due to organic growth from word of mouth.

- **Cost Per Incremental Acquisition (CPIA):** The total dollars spent on a channel divided by the total incremental acquisitions, which is defined as the number of acquisitions that occurred due to the marketing spend. This does not include acquisitions that occurred due to organic growth. (Again the next section will provide more details on how to estimate these incremental acquisitions).

- **Marginal Cost Per Acquisition (MCPA):** The cost to acquire an additional customer at the current spend level. If we assumed a linear relationship between marketing spend and acquisitions then this is equal to the CPIA. However, if we assume diminishing returns with increasing ad spend, which is a more realistic assumption, then the more we spend, the more the MCPA increases and diverges from the CPIA.

These three metrics are shown visually in Figure 1 where we have a nonlinear relationship between spend and acquisition. The relationship shown in the figure where CPA ≤ CPIA ≤ MCPA should always hold. The first equality will only hold if all acquisitions are incremental and the second equality will only hold if the relation between spend and acquisition is linear, but neither of these conditions usually occur in real applications.

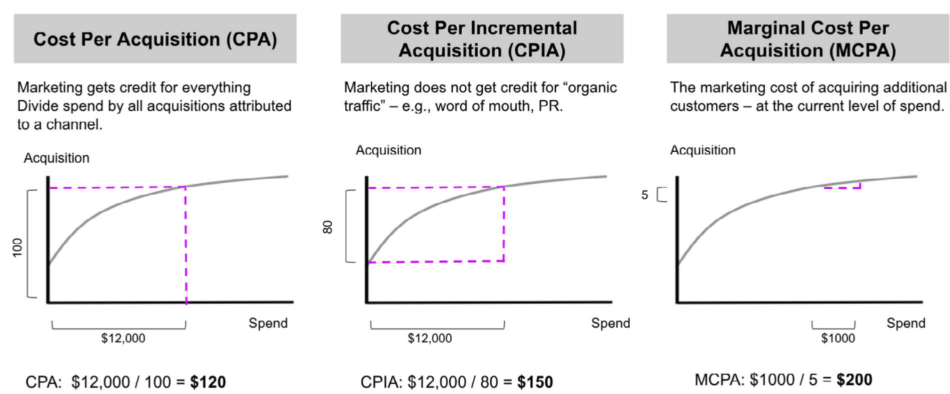

Figure 1: Illustrative figure to show difference between average cost, average incremental cost, and marginal cost in marketing. Left: Average blended costs is the total marketing spend / total acquisitions. Middle: Average incremental cost is the total marketing spend / total acquisitions due to marketing, i.e. total incremental acquisitions that would not have occurred without marketing as opposed to acquisitions from 'organic growth'. Right: Marginal cost is the amount it takes to get a single further acquisition at some level of spend.

The big question now is why do we care about these three separate metrics? The CPA is the most common and what can easily be observed, whereas the second and third give us a much better sense of the health of our acquisitions and the sustainability of our marketing spend. To have a sustainable business, we should not spend more to acquire anyone than they will spend on our product in their lifetime. As such, we should try to cap MCPA at our customer lifetime value (LTV), when spending on marketing. If instead we cap CPA at LTV, it is very likely that a portion of our customers are acquired at a higher cost than their value. Now that we have an understanding of these metrics, the next section will discuss three different ways to estimate them.

3. Methods for Measuring Marketing Efficiency

When it comes to marketing measurement, there is no one source of truth that can tell us the efficiency of all of our marketing channels. In general, marketers and marketing data science/analytics teams have three tools at their disposal for measuring marketing efficiency, none of which should be used in isolation to make spend allocation decisions. These three tools are described below, in order of complexity.

3.1 Click Attribution

Click attribution is the most commonly used methodology for gauging marketing efficiency by channel. Advertising platforms will simply embed a pixel in the vendor's checkout page and every time a customer who has clicked on (or viewed) an ad from the advertising platform makes a purchase, the advertising platform 'counts' this purchase as their own. This count is then used to report total purchases and the CPA to the advertiser. This metric is commonly used because it is easy to understand, does not require modeling or advanced analytics to compute, and therefore can be easily communicated with business stakeholders. Further, this methodology allows the advertiser to easily see trends over time.

While this method is simple, there are obvious drawbacks. First, there can be double counting of purchases across different advertising platforms if the customers click or view ads from multiple platforms. For example, Facebook and Pinterest could be counting the same purchase towards their totals if a customer clicked on both ads before purchasing. This limitation can be addressed by deduplicating purchase attribution with some logic, such as only counting the last click or view before a purchase (last click attribution), but such an approach can inherently favor high intent channels such as search, and harm 'view through' channels such as video. Multi Touch attribution [1], in which different weights are applied to different marketing touchpoints in the customer journey, was recently purported to be the holy grail to fix these issues, but data sharing hurdles have prevented this from realizing its potential [2].

The second drawback of click attribution is that the costs reported are CPAs, and as such do not provide the cost per *incremental* acquisition nor the *marginal* cost, which are key metrics that need to be known to understand whether marketing spend is being used effectively, and to decide where to spend the next dollar. Another way to say this is that CPA does not give us an idea of of the causal effect of marketing. The third drawback to click attribution is in the name— it depends on clicks, which means it is not applicable for offline marketing channels such as TV, radio, podcast, direct mail, and print advertising. While click attribution is easy to report and great for understanding general trends, given these drawbacks, it does not give a full picture of marketing efficiency and should not be used as the sole tool in a marketing analytics toolbook.

3.2 Experimentation

Marketers can run many different types of tests to estimate the incremental lift that advertising provides. The gold standard for online channels is randomized controlled tests (RCTs) in which a randomly selected subset of the population, the control group, is held out from advertising while the remaining population, the test group, is exposed. The difference in purchase rate between the test and control groups provides the incremental lift provided by that channel and this can be used to calculate the CPIA [3]. The CPIA gives the marketers a much better understanding of whether their marketing efforts are revenue positive, as this number can be compared to the LTV of their customers.

While RCTs are the gold standard in terms of determining causality between advertising and purchases, there are some issues with experimentation that mean, as with click attribution, that marketers should not solely rely on experimentation results to make spend decisions. First, experiment results are true for points time relative to whatever else is going on in the world and may not always be applicable. For example, if ThirdLove runs a Facebook RCT at the same time that there is a big TV campaign going on, the lift may not be as large had the TV campaign not been running. That creates the need for constant testing, which presents an opportunity cost for holding out potential purchasers from the advertising campaigns. Second, as with click attribution, RCTs can not be performed for all channels such as TV, print, or podcast, as it is not possible to control who is or is not exposed to the test. In cases like these other experimentation methods, such as geo-based testing [4], must be conducted, which generally have very high variance results [3]. Because of these drawbacks, while RCTs do provide insights in terms of the true costs of some primary marketing channels, they are not totally prescriptive in terms of how to spend a marketing budget holistically.

3.3 Media Mix Modeling

The final methodology for measuring marketing efficiency is building an MMM, a regression model that intends to find the relationship between marketing channel spend and some business level outcome such as acquisitions or total orders. The strength of this approach over click attribution and experimentation is that, as a top-down model, the only inputs are spend and this overall outcome— no clicks or views required. Consequently, all channels, online and offline, can be addressed in a uniform fashion. Further, these models can be used to quantify the impact between each marketing channel and the overall outcome, and for forecasting as they can predict the outcome as a function of different spend inputs. Also, unlike the other two methodologies, MMMs are prescriptive as given a model, spend can be optimized to maximize the business outcome. This feature gives marketers an explicit spend plan for all channels to follow, instead of having to come up with a plan given different trends or experiment results for each channel.

While MMMs may seem like the perfect solution, there are again drawbacks to this approach and it is not advisable to use an MMM in isolation without the other two methods. First, this approach is the most ‘black box’ of the three, and therefore there is often an education barrier that must be overcome for business stakeholders to feel comfortable using the recommendations of the model. Second, as a regression model, the outputs are only as good as the inputs. Often marketing spend is correlated as businesses make ‘big pushes’ or pull back in many channels at the same time. This makes it difficult, or impossible, to tease out any relationships. Further, we are trying to draw causal relationships from this model, when in reality there may be no causal mechanism and we are merely picking up on correlations between certain spending and the outcome variable. Finally, we are often working with limited data. In young businesses, the state of the company today looks nothing like it looked two years ago, so the timeframe of useable data is often limited. Unlike other state-of-the-art predictive models, MMMs are often built from hundreds of data points, rather than hundreds of thousands [5].

While there is no model or methodology that provides perfect recommendations for marketers in isolation, using all three of these approaches in tandem results in a holistically informed view of marketing efficiency. That’s exactly what we do at ThirdLove: 1) with click attribution we can spot trends, identify issues quickly, and micro optimize across different campaigns within the same channel; 2) with an MMM we can estimate our marginal costs, forecast our acquisitions, and optimize our spending plans, and 3) with experimentation, we measure ground truth values for acquisition costs that allow us to validate our MMM and help us understand how much of our click attributed acquisitions are actually from organic traffic.

	Click Attribution	**Experimentation**	**Media Mix Models**
Pros	Very easy to calculate and understand Easy to spot trends over time	Gold standard for inference so gives a ground truth for the incremental value of marketing Easy to implement on many advertising platforms	Incorporates online and offline channels and can control for non marketing factors Model is prescriptive in that it provides an optimal media mix
Cons	Only applicable for online channels Penalizes ‘view-through’ channels and favors ‘demand capture’ channels	Not always possible for all channels Opportunity cost of not marketing to the holdout group Results dependent on a point in time media mix	Quality of the results is highly dependent on quality of inputs. Need years of data Difficult to distinguish between correlation and causation Model can appear as a black box to business stakeholders

Table 1: Table to show pros and cons of three marketing measuring techniques: Click Attribution, Experimentation, and Media Mix Models.

4. Building a Media Mix Model

The rest of this paper will specifically focus on how to build an MMM and how to validate the model using the other two approaches as well as traditional backtesting. This section focuses on the data required to build a model, the model itself, and the insights that can be drawn from the model once it is built.

4.1 The Data

To build an MMM we need time series data for the following: the outcome variable, which we will assume is acquisitions from this point forward, the marketing spend by channel, and any other variables outside of marketing spend that we may need to control for. In terms of how far back in time to go for data, there is a trade off: on the one side, the further you go back, the more data you get to train the model, while on the other side, the further you go back the less relevant that data is to the current state of your business. At ThirdLove, we go back approximately two years with daily spend data to give us ~730 data points. We have found that daily data, albeit noisy, helps us build a more robust model as opposed to weekly data where we would only have ~104 data points (or be forced to go further back in time). Further, we have found that there is less correlation between our channel spends at the daily level. In terms of the other variables we want to control for, these should include anything that has an impact on acquisitions outside of marketing spend. For e-commerce applications, these can include day of week or other types of seasonality, changes in product offering or quality over time, significant changes in acquisition strategy, sales events, and big press releases, among others.

Another thing to consider is that the input spend data may need to be transformed before we try to fit a model to it. For example, if we only have monthly totals of how much we spent for some channel, there needs to be a strategy for spreading this spend daily over the month. Further we may want to tailor how we scale the spend in different channels according to the size of their target markets or cost per impressions. We can also decay the spend over time according to some function if we believe that certain channels have a delayed impact on acquisition. These choices are very application specific and can feel arbitrary. In Section 5 we will address how to validate the model and assess whether these transformations are appropriate.

4.2 The Model

Once the data is collected and transformed, we want to create a model that we can use for inference about the impacts of the marketing channels on the business. MMMs lend themselves to time series modeling with a log-log regression component because of the ability of these models to capture dynamics such as diminishing returns, the multiplicative effect of marketing, seasonality, and underlying trend. The functional form of the model is then essentially the Cobb-Douglas production function,

$$y_t = A_t(f(x_{1,t}) + 1)^{\beta_1}(f(x_{2,t}) + 1)^{\beta_2}...(f(x_{N,t}) + 1)^{\beta_N}$$

Where,

- y_t is the number of acquisitions at time t, where t has a daily cadence.

- A_t is a scaling constant that accounts for acquisitions not explained by marketing variables. When we solve we explicitly break out this A_t into contributions from seasonality, controls for changes in the business, and organic acquisitions. A_t changes over time as a function of the previous timestep.

- $x_{i,t}$ is the amount of dollars spent in marketing channel i at time t. We have an x variable for each of the N marketing channels. These spend amounts can be transformed before they are input into the model to account for adstock lag and channel saturations [5]. These transformations are represented in the above equation by $f(\cdot)$.

- The +1 is added to each term for dealing with time periods with instances of zero spend in one or more channels.

- β_i is the sensitivity for channel i . This dictates the relationship between channel spend and acquisitions. This exponent should be between 0 and 1, and the sum of all of the β values for all i should also be less than 1. This is ultimately what we are solving for to understand the relative efficiencies of our different channels.

At ThirdLove we use the BSTS R package to find the model parameters [7] (a link to example code is provided in the Appendix).

One important note on the model is that it is essential to apply some sort of regularization because the spend inputs are often correlated. It is unlikely that you will have a data set without correlations between channel spends as marketers will often pull back or increase on many channels together depending on the needs of the business. As mentioned, we have found that using daily data helps with this decorrelation. Also encouraging the marketing team to fluctuate spend in independent channels periodically provides richer material for the model to learn from. In the example code we regularize by putting Gaussian priors on our coefficients.

4.3 The Insights

Once we have our solved for our model parameters, we now have a function that outputs number of acquisitions given a media mix. We can use this function to perform a number of tasks to help us better understand the efficiency of our marketing spend and make better decisions. For example, we can now infer the average incremental cost and marginal cost of marketing spend, we can forecast out acquisitions given different media mixes, and we can optimize our media mix such that we maximize acquisitions given a budget, or minimize budget given an acquisition target.

4.3.1 Marketing Costs

To estimate the average incremental value a channel, let's say channel 2, we can calculate the number of acquisitions we will get at a certain media mix

$x_1, x_2, x_3, ..., x_N$ and then recalculate the number given that channel 2 has no spend. The difference between these two values is the number of incremental acquisitions provided by spending in channel 2 and the average incremental cost for this channel can be calculated as:

Average Incremental Cost for Channel 2 = $\dfrac{x_2}{f(x_1, x_2, ..., x_N) - f(x_1, 0, ..., x_N)}$

Similarly, we can estimate the marginal cost for channel 2 by finding the difference in acquisitions at a certain media mix and at that same media mix plus one more dollar in channel 2. This gives us the number of acquisitions per dollar in channel 2 at this medix mix, of which we can take the reciprocal to find the number of channel 2 dollars per additional acquisition, shown below:

Marginal Cost for Channel 2 = $\dfrac{1}{f(x_1, x_2 + 1, ..., x_N) - f(x_1, x_2, ..., x_N)}$

4.3.2 Forecasting Media Mix Scenarios

Using the model it is very straightforward to plug in various media mixes and find the number of acquisitions predicted in each case. At ThirdLove we provide this ability as a tool to our marketers to enable them to play out different scenarios and help them make budgeting decisions.

4.3.3 Media Mix Optimization

In addition to plugging in spend amounts and finding the acquisitions, we can also set up optimization problems that will optimize some outcome such as acquisitions or total spend. Since our model is a convex function (as long as our beta coefficients are between 0 and 1 and sum to less than or equal to 1), we can easily solve for optimal spend amounts using a solver to find a global optimum. At ThirdLove, we use the CVXPY solver [7]. Below is the formulation for a maximization problem to find the spend amounts that will maximize total acquisitions given a certain budget.

$$
\begin{aligned}
&maximize \quad (x_1 + 1)^{\beta_1}(x_2 + 1)^{\beta_2}...(c_N + 1)^{\beta_N} \\
&subject\ to \\
&\sum_{i=1}^{N} x_i \leq \ Total\ Budget \\
&0 \leq x_i \leq \ Channel\ Budget_i \quad \forall i
\end{aligned}
$$

This is the simplest version of this optimization. We could add more constraints such as a separate budget for online and offline channels or any other specific budget constraints. We could also maximize over multiple time periods to get the optimal media mix over different days of the week. At ThirdLove we provide this optimization functionality to our marketing teams for maximizing orders and for minimizing spend subject to revenue goals which allows them to look at the problem in two different ways.

5. Model Validation

All of these insights are obviously very dependent on the values of the parameters found by solving the time series model. So how can we make sure the parameters are right? First, we can make sure that we have ‘good data,’ that is, daily spend data by channel for the last two years that has some independent fluctuations of channel spends. But deciding how much fluctuation is necessary to ensure that we are picking up the channel signals appropriately can seem arbitrary.

At ThirdLove we use two techniques to validate our model. The first is an obvious choice for a time series model: we perform backtests where we train the model up to some point in time and examine the difference between the actuals and the predictions in the holdout period. Note that we are not holding out a subset of our customer base, but a portion of our input data over a period of time.

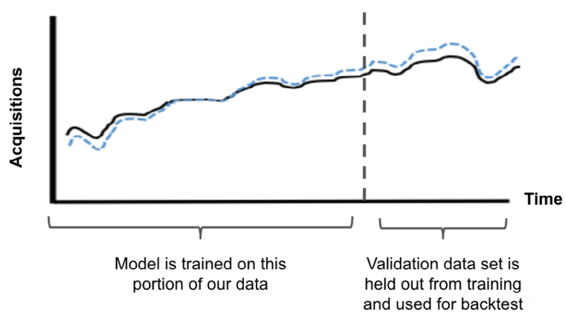

Figure 2: Illustrative figure to show backtesting of a time series model. The model is trained with data to the left of the dotted vertical line. The remaining input data is fed into the model and the acquisition predictions are compared to the true historical data to find the model error.

The second validation method we perform is to compare our experimental results with the average incremental cost per acquisition that we infer from the model. We have found that it is very important to perform both checks when validating the model, because two models with similar backtest results may have very different average incremental cost estimates, and without having ground truth experimental results, it is unclear which model is better reflecting reality. We will demonstrate an example of this discrepancy in the next section.

5.1 XYZ Enterprises Example

Imagine we have a company called *XYZ Enterprises* and for the last two years we have been collecting data on how much we spent in each of our four marketing channels and our subsequent acquisitions. We have been diligently collecting this data every day, shown in Table 2 and Figure 3, and now we are ready to build our first MMM so that we can optimize our marketing budget. (Note: the code to generate this synthetic data set is provided in the companion code). In this example we have four channels and channels 1 and 3 are offline channels, while channels 2 and 4 are online channels.

Date	Channel 1	Channel 2	Channel 3	Channel 4	Acquisitions
2016-05-01	$843	$30,073	$17,526	$2,924	6339
2016-05-02	$116	$35,476	$11,795	$8,269	5044
2016-05-03	$2,016	$17,214	$11,322	$7,245	5193
2016-05-04	$1,456	$2,232	$14,295	$5,724	4547
2016-05-05	$2,056	$12,455	$26,495	$5,313	5542
...	...	...	...	...	...

Table 2: Table to show the first five rows of simulated marketing spend and resulting acquisitions for XYZ Enterprises

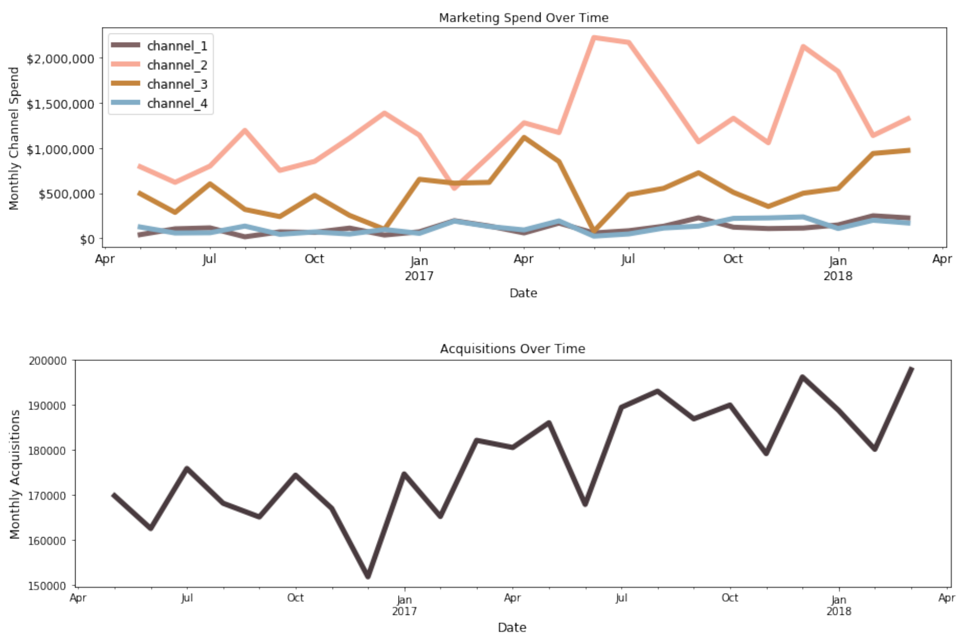

Figure 3: Figure to show the simulated spend by channel and acquisitions for XYZ Enterprises. Simulated daily data has been aggregated to the month level.

When building our model, we used the same functional form described in Section 4.2, including investigating different schemes for preprocessing the spend data before putting it into the MMM. The ways that we can preprocess the data are: 1) scaling the spend data before using it in the model to account for the different rate of saturations of our channels due to their different audience sizes and 2) by adding decay functions to our spend to account for the lag between the time the ads are served (and paid for) and the time that we acquire the users. Incorporating saturation and adstock decay components is extremely important in properly modeling the media mix.

Here are three different scenarios in which different modeling of the scaling and decay functions produce significantly different results. In the first scheme we do not scale the spend at all before inputting it into the model, but we do add a seven day decay to each of the channels so that the daily spend is split between the day we spent it and the six following days using an exponential decay. In the second scheme we do not add any decay but we scale all of the spend by 5000 before inputting it into the model to create a more gradual saturation of spend. In the third scheme we add four day spend decays for channels 1 and 3 only, since they are the offline channels and potential customers may be exposed to the ads when they are out and about. We also used different scaling constants for channels 1 through 3, because we believe these may have bigger audiences that will saturate slower with respect to increasing spend than channel 4. All of these choices may or may not be reasonable given the evidence that we have for the time it takes for people to see an ad and make a purchase as well as the relative audiences sizes of these different channels. However, these choices can start to seem arbitrary if we do not validate them properly.

	Channel 1	Channel 2	Channel 3	Channel 4
1	No scaling Add 7 day decay	No scaling Add 7 day decay	No scaling Add 7 day decay	No scaling Add 7 day decay
2	Scale by 5000 No decay	Scale by 5000 No decay	Scale by 5000 No decay	No scaling Add 7 day decay
3	Scale by 15000 Add 4 day decay	Scale by 15000 No decay	Scale by 15000 Add 4 day decay	No scaling Add 7 day decay

Table 3: Table to show the preprocessing treatment to deal with saturation and acquisition delay for each channel for three models

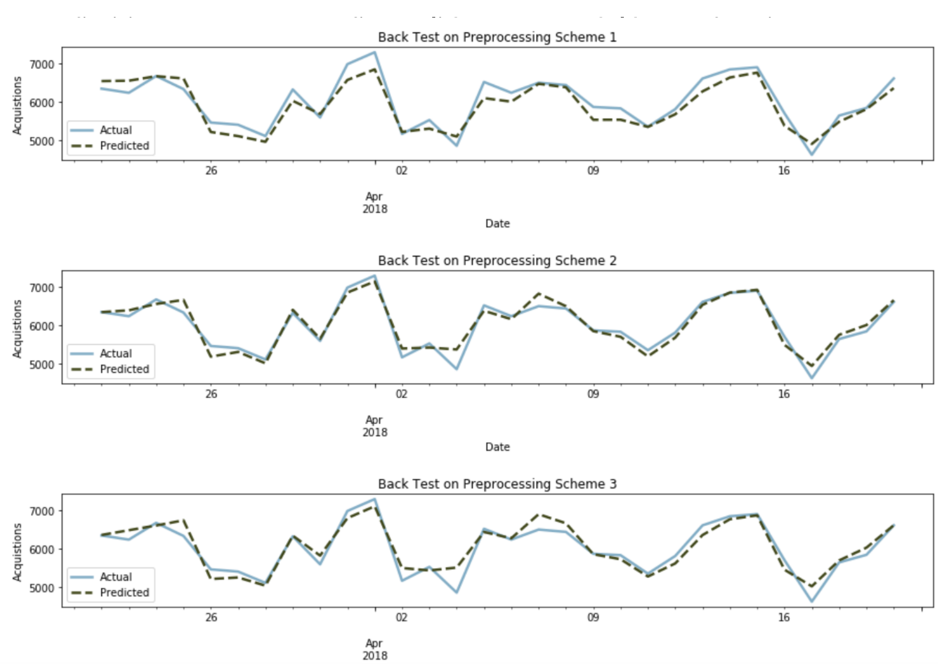

Figure 4: Figure to show backtest for three models with three different preprocessing schemes shown in Table 3

The three schemes above provide backtest mean absolute percentage errors (MAPE) of 3.6%, 2.5%, and 3.1% respectively. At this point it is tempting to think that we are done and choose the second preprocessing scheme because it fits the data the best. We could also go further and try to minimize the backtest MAPE over a whole grid of different decay lengths and scalings, as we might do in hyperparameter tuning for a traditional supervised learning problem.

However, if we want to do more than just predict the the number of new customers our company will acquire for spend amounts our model has already seen, that is, if we want to infer the cost per incremental acquisition, marginal cost per acquisitions, and use the model for optimizing our media mix, we need to further validate. Because of the high noise in our data and the frequently correlated channel spends we can get very similar backtest MAPEs with very different cost estimates. For example, using the three models generated with our three preprocessing schemes we can go back to a period in February 2018 and calculate, using the equations in section 4.3.1, the average and marginal costs per channel. These are shown in Table 4 for each model. We can see that for the three models we generated, while the MAPEs only range from 2.5-3.6%, the estimates for Channel 1 CPIA ranges from $20 - $68 for the same period of time. The magnitude of this range is high for all channels and for the MCPA values as well. We could draw wildly different conclusions about whether our ad spend is efficient and about which channels we should invest more in depending on which scheme we go with. For example, with the second model, channels 2 and 3 have the lowest marginal costs, so our optimization would conclude that we could take some of the money we are spending in channel 4 and put it into these two channels. However, if we used the third model we would choose to move money into channels 1 and 3 instead.

	Backtest Result - Daily MAPE	Channel 1 CPIA/MCPA	Channel 2 CPIA/MCPA	Channel 3 CPIA/MCPA	Channel 4 CPIA/MCPA
Model 1	3.6%	$20 / $229	$12 / $117	$16 / $153	$24 / $269
Model 2	2.5%	$69 / $148	$52 / $127	$50 / $125	$148 / $257
Model 3	3.1%	$48 / $65	$65 / $131	$41 / $70	$130 / $226

Table 4: Cost Per Incremental Acquisition (CPIA) and Marginal Cost Per Acquisition (MCPA) estimates for each channel from 2018-02-01 to 2018-03-01 for three models built with different preprocessing of spend data according to Table 3.

In light of this observation, relying on the model backtest alone is not a viable option. Luckily, XYZ Enterprises has been trying to measure the marketing efficiency in other ways over the past two years, with click attribution and experimentation at the channel level. Let's say that during February 2018, XYZ Enterprises conducted three holdout tests for channels 2,3, and 4, and we also have the CPAs reported from channels 2 and 4. We can use these results to validate our model, because (1) the CPAs reported by the channel platforms should always be lower than the model estimates of CPIA as the reported CPAs include non incremental orders and potentially cannibalized orders from other channels, and (2) the holdout test results CPIA values should be close to our model estimates for CPIAs.

	Click Attribution from Channel Platform - CPA	**Holdout Test Results - CPIA**
Channel 1	n/a offline channel	no test data
Channel 2	$36	$66
Channel 3	n/a offline channel	$39
Channel 4	$96	$128

Table 5: Cost Per Incremental Acquisition (CPIA) and Marginal Cost Per Acquisition (MCPA) estimates for each channel from 2018-02-01 to 2018-03-01 for three models built with different preprocessing of spend data according to Table 3.

Comparing the click attribution and test results with the backtest results, XYZ Enterprises comes to the following conclusions:

- With regards to the CPA numbers reported from ad platforms, model 1 gives completely unreasonable cost estimates as the model CPIAs are lower than the click attributed CPAs. On the other hand, both models 2 and 3 seem reasonable as the model CPIAs are greater than the click attributed CPAs. With these click attributed CPAs, we can only do this directional comparison, but it helps to rule out model options that are not viable.

- For the holdout tests, we can can do more than a directional assessment. In a perfect world, the model CPIAs would perfectly match the holdout test results. If we had more channel test results and more models, we could look at the RMSE between the model CPIAs and the holdout results. In this case, we can quickly identify that model 3 has a smaller margin of error when comparing the the holdout test CPIAs to the model CPIA estimates.

With these two conclusions XYZ Enterprises identifies model 3 as the best choice of the three, even though it did not provide the best back test results. With this, we can either move forward with model 3 or we can keep iterating on the way the data is preprocessed until we are satisfied that we have minimized with error between the test results and the model CPIAs. Either way, we still won't know if our model is accurately capturing the dynamics of channel 1, for which we have no test results. This is the dilemma of the marketing data scientist, but the only solution is to keep tracking, keep testing, and to keep updating the models as more time goes by. This will likely be an iterative process where the model must be recalibrated as more test results come in. Further, holdout test results in one channel can change based on how much is being spent in other channels so iterative testing and retraining is valuable even if results are available for every channel.

6. Conclusion

In conclusion, the MMM is an insightful tool which can help companies optimize their marketing spend. For maximum accuracy, a media mix model needs to be validated and iterated upon. Validation can come from using click attribution measured from online advertising platforms as a lower bound for modeled CPIAs, and performing holdout tests to find ground truth CPIAs. These ground truth CPIAs also validate the MMM preprocessing parameters such as those that control adstock lag and saturation. Without tools for outside validation, the MMM can seem accurate through backtesting but fail to correctly infer attribution from marketing spend, which in turn can lead to faulty predictions for new media mixes.

7. Acknowledgements

Thank you to Kinga Dobolyi, Tessa Johnson, Kim Larsen, Megan Cartwright, and Andrew Hakanson for reviewing drafts of this paper.

8. Appendix

The companion code used for generating the XYZ Enterprises MMM validation example is provided at https://github.com/mecommerce/ThirdLove-Tech-Blog. It is composed of two notebooks, written in Python 3.6:

0_TL_Whitepaper_Companion_Code_Data_Generatation contains the code for generating synthetic marketing spend and sales data

1_TL_Whitepaper_Companion_Code_MMM contains the code for building the media mix model, backtesting, and cost inference

9. References

1. Bernman, R. (2018) Beyond the Last Touch: Attribution in Online Advertising; Marketing Science; Vol. 37, No. 5; https://doi.org/10.1287/mksc.2018.1104

2. AdExchanger (2018) Did Google Just Kill Independent Attribution? https://adexchanger.com/analytics/did-google-just-kill-independent-attribution/

3. Gordon, B. Zettelmeyer, F. Bhargava, N. Chapsky, D. (2016) A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook; https://www.kellogg.northwestern.edu/faculty/gordon_b/files/kellogg_fb_whitepaper.pdf

4. Kerman, J. Wang, P. Vaver, J. (2017) Estimating Ad Effectiveness using Geo Experiments in a Time-Based Regression Framework; https://research.google.com/pubs/archive/45950.pdf

5. Accenture Digital (2018) Exploring Granular Data in MMM: Updated Models, Better Insights

6. Larsen, K. (2018) Data Science Can't Replace Human Marketers Just Yet — Here's Why; https://www.thirdlove.com/blogs/bimodal/data-science-can-t-replace-human-marketers-just-yet-here-s-why

7. Scott, S. (2018) Package 'bsts' Bayesian Structural Time Series https://cran.r-project.org/web/packages/bsts/bsts.pdf

8. Diamond, S., Boyd, D (2016) A Python-Embedded Modeling Language for Convex Optimization; Journal of Machine Learning Research; Vol. 17, No. 83 https://www.cvxpy.org/