The Value of Calibrating MMM with Lift Experiments

September 2022

Contents

The Why

Marketing Mix Modeling (MMM) helps to measure the contribution of offline and online marketing elements and external factors to Key Performance Indicator (Sales).

Conversion Lift Experiments (CLE) measure the actual number of conversions that are caused by specific digital ads by comparing two groups of people: a test group (who have seen the ads) and a control group (who have not seen the ads).

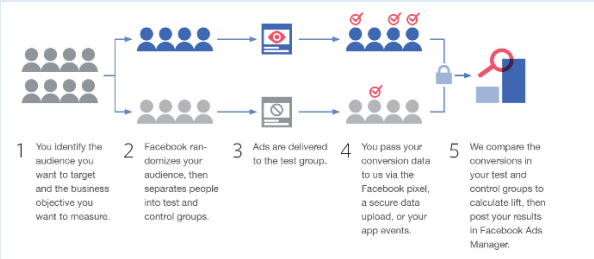
MMM works with aggregated data and has more of a macro-level view while CLE typically uses more granular data and a shorter duration.

This analysis aims to compare the results of these two approaches and explore how they can be used together for a more aligned and consistent understanding of advertising impact.

1.1. Context

The objective of this research analysis was to better understand the value of using lift studies as input for MMM calibration

Meta has a repository of incremental lifts based on experiments run on multiple KPIs and advertisers. These lift numbers are calculated using a ***test versus control*** framework.



Source:https://www.facebook.com/business/m/one-sheeters/conversion-lift

Analytic Edge has a repository of ***Marketing Mix Modeling*** studies across advertisers that have lifts from Meta campaigns.

1.2. Key Findings: Takeaway for Advertisers

Calibrating lift studies into the MMM measurement framework connected the two analyses. In 67% of the studies a change in the MMM ROI is observed, and on average, the ROI changed by 25%.

- Narasimha Rao, Vice President, Analytic Edge

What could it mean for advertisers?

Calibration impacts the Return on Advertising Spend (ROAS) which has implications for advertisers. ROAS is an important metric that is used when making decisions on budget allocation, and this means that if the **ROAS can reflect both the MMM and CLE perspective, it provides a balanced view towards budget decisions.**

After considering the differences in scope, in 67% of the studies, layering the experiments within MMMs impacted Meta ROIs by approximately 25%.

Accuracy of MMM is typically measured through out of sample validation. The results can further be fine-tuned by calibration. **Lift studies closely aligned to the MMM scope** are more comparable, and by doing this, experimental results can complement MMMs and generate results thar are closer to the ground truth.

2 in 3 of analyzed MMM studies significantly changed "Meta" ROI results after calibration

- **A clear (significant) improvement in prediction power when more than 2 lift studies are used for calibration** and hence, we recommend at least 2 studies for the calibration

On average, we observed a change of 25% in ROI post calibration.

- Depending on availability of multiple lift studies, MMM as a methodology (through calibration) can be much faster and much closer to ground truth.
- We observed ROI changes in the range of -21% to +29% post calibration

Criteria	Total Studies	Meta ROI Index Pre-Calibration	Meta ROI Index Post-Calibration
ROI Improved	6	1.97	2.55 (+29%)
ROI Declined	4	1.19	0.94 (-21%)

*ROI Index = Meta ROI / Total Media ROI

All the necessary factors like objectives, investment, purchase cycle etc. to be considered before calibration.

Figure 1: Impact of calibration

The What

2.1. Background

The general observation was that Return on Investment (ROI) from lift studies are lower compared to MMM results. The potential factors contributing to these differences could be on account of the following:

Product scope: MMM as a framework is holistic and analyzed typically at an aggregated brand level. Lift studies, on the other hand, could be very specific to a campaign focused on a particular product.

Sales channels covered: Lift studies are primarily based on online sales (single channel) while MMM primarily covers omnichannel sales. A coverage factor needs to be used to bridge the gap between these solutions. Online to offline halo impacts can be very important to ROAS/ROI.

Target audience: ROAS from lift studies in most cases reflect the impact from the campaign target audiences whereas MMM is based on impact from the total population.

Timing of the lift study: It is ideal to run more than one experiment at different points in time to help capture time-varying factors. Average impact might be different than what the lift study would say otherwise.

Long purchase cycle: If the test only runs for a short duration, then some of the payback might be under accounted.

Multiple campaigns (brand activation): Performance marketing versus brand activation campaigns need to be accounted for while comparing ROIs from lift and MMM studies.

Marginal impact: The full extent of campaign impact could go beyond the measured window in lift studies.

2.2. Scope

The study covered 9 markets across the US, EMEA, APAC and LATAM regions, with 15 MMM studies and nearly 200 lift studies. The timeframe of MMM studies stretched from January 2017 to September 2021. Projects had lift studies ranging from 1 to 21, with most projects having less than 4 lift studies.

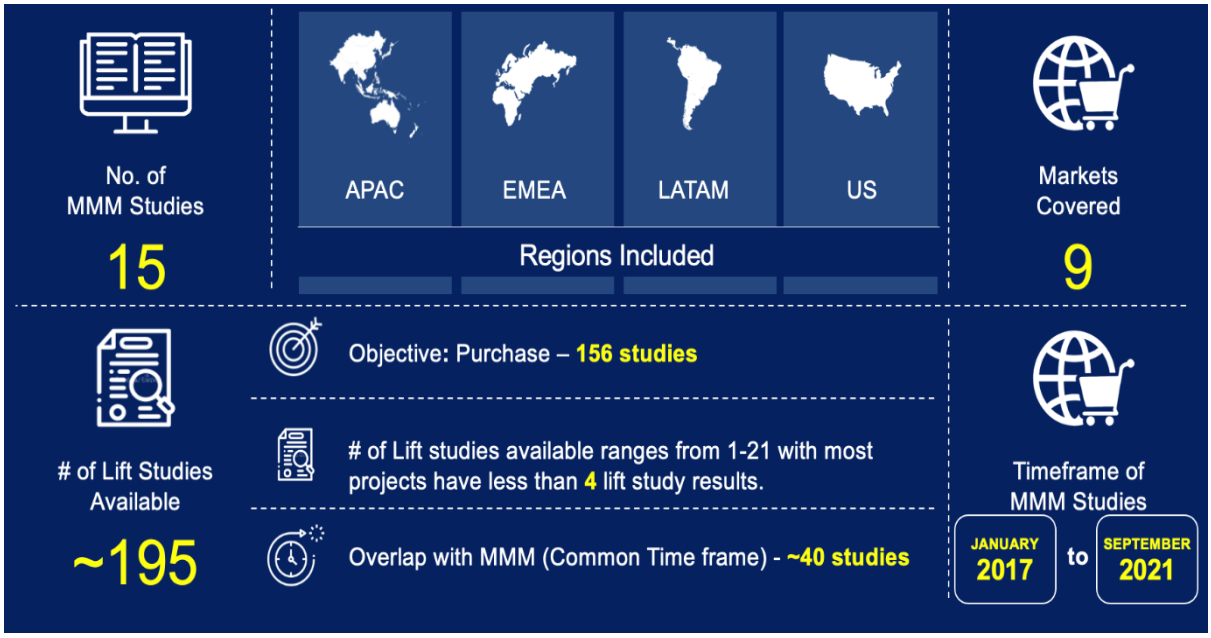


Figure 2: Scope of the project

The How

3.1. Approach

Overall, the objective of this research analysis was to better understand the value of using lift studies as input for MMM calibration. To calibrate different MMM results across multiple industries and regions, a unified approach was adopted.

- As inputs, multiple lift studies based on the factors typically available as part of any experiment study were used.
- The way the impact was evaluated was through measuring the improvement of model fit (R-square, MAPE) as well as the impact on Meta ROI.

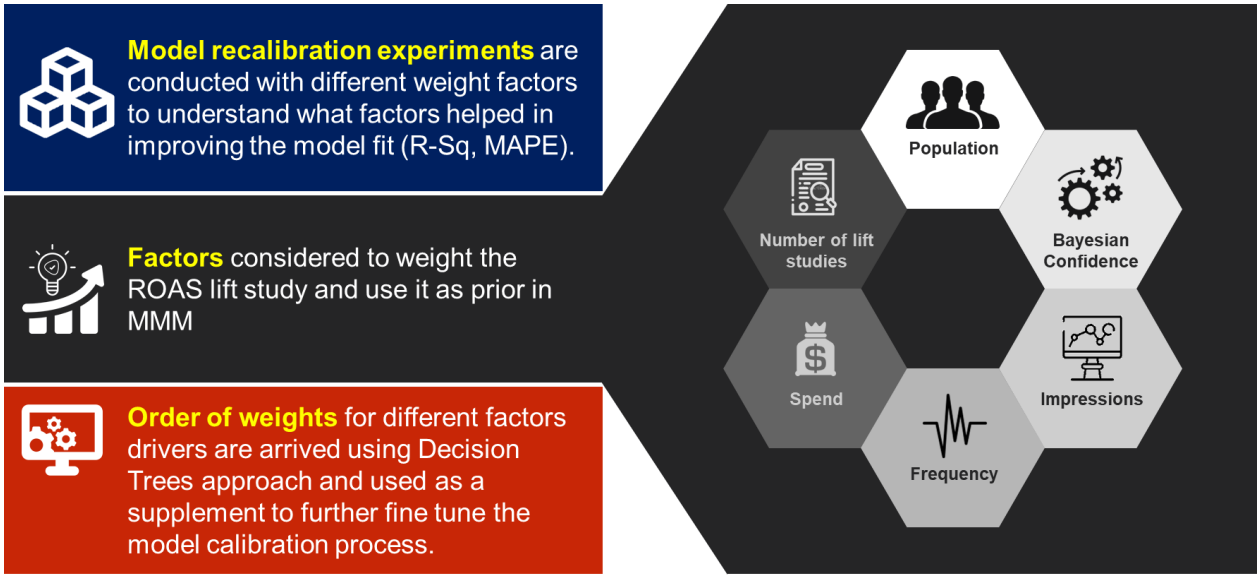


Figure 3: Factors considered for weighting

3.2. Approach: Analytical Framework

A 3-step analytical framework approach was adopted for the project.

It included alignment of objectives, product focus and time frames between lift studies and MMM results. Next, decision trees were built to understand the key factors and their weights to aggregate multiple lift studies to come up with unified priors. These priors were fed into the MMM analyses to measure the changes in Meta ROI.

i. For projects with multiple lift studies, the ***objectives, product focus and time frames*** were aligned with MMM.

ii. A decision tree approach was used to arrive at ***factors and their weights***. These were then used to estimate the weighted average ***ROI for calibration***.

iii. Recalibration of the model was done with the objective of improving ***model statistics*** for the same time period.

Figure 4: Approach

1

3.2.1. Data Harmonization

ROIs from multiple conversion lift studies were aggregated using Population, Bayesian Confidence, Impressions, Frequency, and Investment as weights. Lift studies conducted beyond MMM scope were also considered during the calibration process.

The key assumptions made during this step were as follows:

1. **Incremental sales** from lift studies are comparable to sales contributions from media in MMM studies.
2. **Metrics** – **Impressions, Spend & CPM** from lift studies are comparable to MMM.
3. **Filters** – Filter on studies where MMM results have been **signed off** by end clients.
4. Lift studies with very negative ROAS are not considered for this calibration exercise

Figure 5: Data harmonization

2

3.2.2. Decision Tree

One of the key challenges in this research was **aggregating multiple lift studies into one** without losing the information, and at the same time acknowledging the differences between lift studies and MMM.

Given that most of the advertisers had run more than one lift study, a machine learning analysis was used to determine the rules of aggregation and identify key factors and their weights through a trial-and-error approach. This unified approach enabled aggregation of lift studies to be done at scale, and in turn helped with the calibration process and to strike a balance between lift studies and MMM while measuring the short-term impact of media.

The step involves two parts:

- **Part 1 - Aggregating lift studies**

  The rationale of this part is 'prioritizing lift studies based on its level of importance' (e.g., based on how big the population of that lift study is or other factors like frequency, spend, etc.).

  Decision tree approach was used, and priority was assigned to lift studies in sequential order (tree structure) as shown in the figure below. The order of selecting the key factors to assign weights can vary depending on whether spend is considered as one of the factors or not during this step.

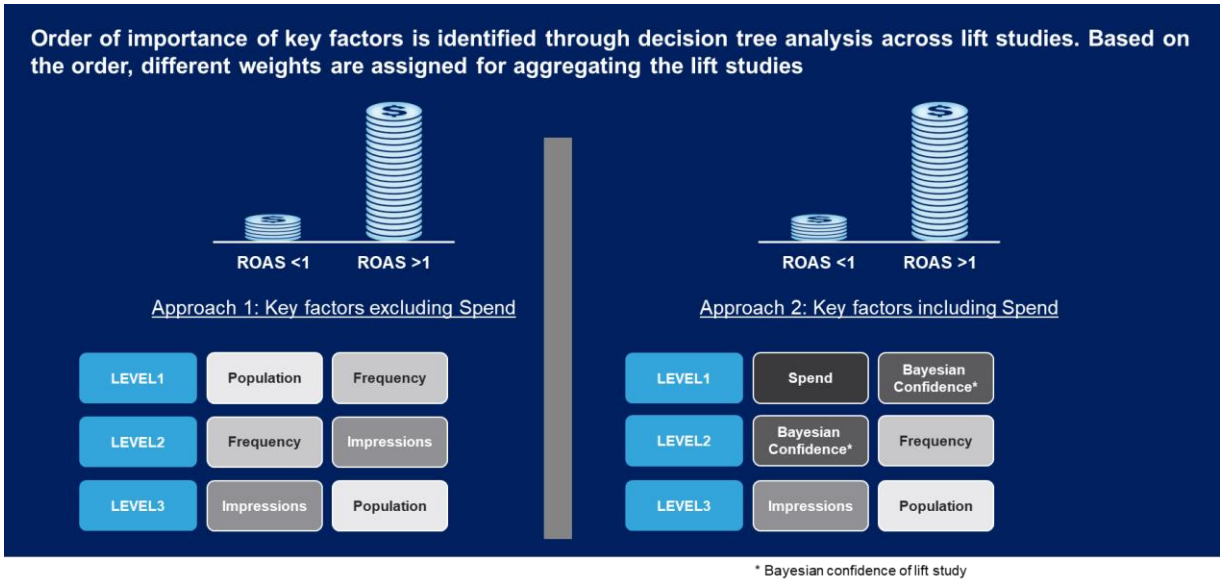


Figure 6: Key factors for lift study

***Part 1:* Aggregating lift studies**

Example

Let say the client has 8 conversion lift studies, 5 of which have ROAS <1, another 3 have ROAS >=1. Below is the approach to aggregate 8 lift studies:

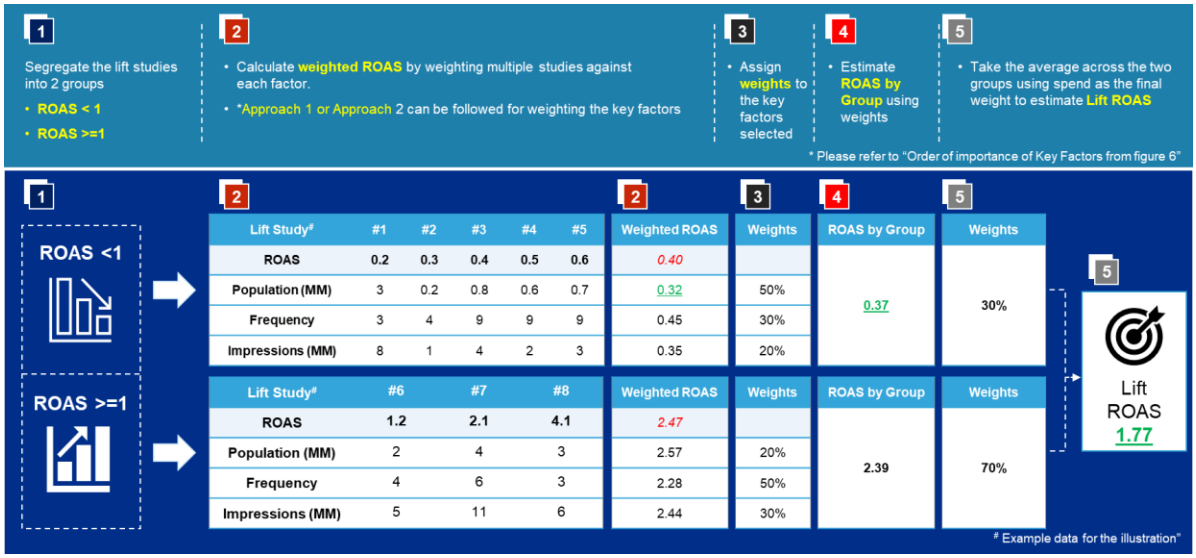


Figure 7: Decision Tree (Part 1) - Aggregating lift studies

Step 1: Split the lift studies into two groups using ROAS <1 and ROAS >=1 as criteria.

Step 2: Calculate factor level Weighted ROAS across different lift studies using values of each factor as weights.

- Population weighted ROAs of 0.32 for group ROAS <1 from figure 7 is derived as a weighted average of all ROAS from 5 lift studies using Population reach from each individual lift study as weights.

Step 3: Factor level Weights can be assigned based on order of importance from *"Key factors for lift study" from figure 6.* Key factors can include spend as well.

- 50% Weight is assigned to Population for ROAS < 1 group in the above example.

Step 4: ROAS by Group can be estimated by combining factor level Weighted ROAS and factor weights from steps 2 and 3.

- **0.37** for ROAS < 1 group is a weighted average of factor level Weighted ROAS 0.32, 0.45, 0.35 and factor level weights of 50%, 30% and 20%.

Step 5: Overall ROAS across lift study groups can be estimated by combining group level ROAS and group level weights. Spend can be used as a factor to determine the group level weights.

- **1.77** Lift ROAS is a weighted average of Group level ROAS 0.37 and 2.39 and weights of being 30% and 70%.

- **Part 2 – Finetuning aggregated Lift ROAS with MMM**

  This part aims at 'accounting for factors that could drive deviation between lift studies vs. MMM' before using lift study as prior for the calibration.

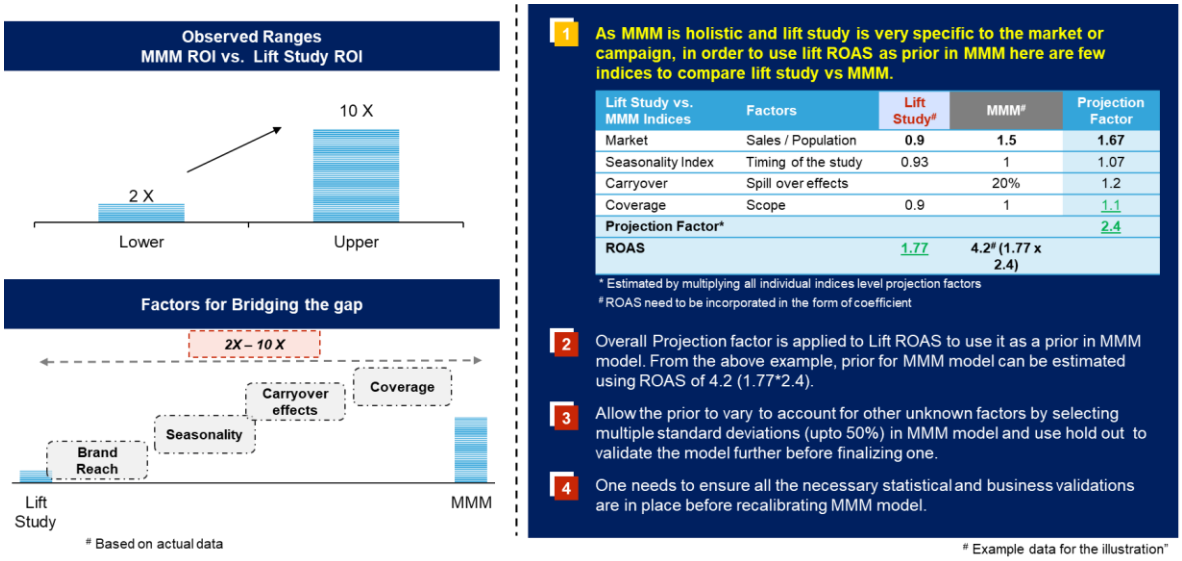


Figure 8: Decision Tree (Step 2) – Fine tuning aggregated ROAS with MMM

Part 2

<u>Example (cont'd)</u>

Overall lift ROAS of <u>1.77</u> from Part 1 was **projected** by considering differences in lift study scope vs. MMM scope through indices like 'Market, Seasonality Index, Carryover, and Product Coverage' etc.

<u>Step 1</u>: Compile data for all the key indices that are necessary while comparing lift study vs. MMM in terms of scope and coverage.

- Scope, Market Size, Seasonality Index, Carryover effects etc.

<u>Step 2</u>: Projection factors are derived for each factor by comparing the scope between lift study and MMM.

- Coverage for lift study is 0.9 in terms of overall MMM Scope and projection factor is estimated to be <u>1.1</u>

<u>Step 3</u>: Projection factors across different indices are aggregated (multiplied) and applied to lift study ROAS to estimate the prior for MMM.

- Weighted projection factor <u>2.4</u> was applied on lift study ROAS of 1.77and final projected ROAS of 4.2 was then used as prior for the calibration.

3

3.2.3. Recalibration Approach

The calibration is done by using the final lift study ROAS (from Part 2) as the prior value for the MMM model run. Along with this, there are additional constraints that have been incorporated during the calibration exercise.

- Meta ROAS from lift studies after applying necessary projections used as prior and were not allowing MMM ROAS after calibration to deviate more than 50% from this prior.
- Out of sample validation is used to validate the models.

 **Factors**	• Key elements considered from lift studies included Population, Bayesian Confidence, Impressions, Frequency, Spend, and number of lift studies. • Multiple approaches with several factors were tried for the purpose of imposing weights. • Appropriate weights were applied based on ROAS levels for being considered as priors in MMM. • Meta ROAS from lift studies were used as priors. MMM ROAS after calibration were not allowed to deviate more than 50% from these priors. • Out of sample validation was used to validate the model further before finalizing.
 **Watchouts**	• Lift studies with significantly low ROAS had to be treated with caution, especially when there were a limited number or only one study available for the calibration. • Across the modeled period, **two or more experimental results** (lift studies) are needed to calibrate. • It was ensured that experiments used for calibration were statistically significant (*80% cut off - Bayesian confidence used for this analysis).
 **Indices**	Multiple projection factors were created using the below listed factors versus lift study reach. These ratios were used to calibrate MMM with sales lift from experiments. • Population weights – Total population, number of households • Digital penetration – Meta users • Lift study coverage – Impressions, spend • Additional factors – Brand revenue vs. spend, brand market share, target reach, seasonality, purchase cycle etc.

4. Insights

The analysis tries to come up with a balanced view considering two different analyses exist which are solving for the same question.

Lift studies are typically shorter range, more targeted versus an MMM which takes a more long-term view and is more “brand” focussed. So, if both exist, then is there a way to layer in both and come up with a solution which may be representative of the strengths of both?

- The project clearly outlined possible advantages of calibration. Integration of experimental lift results with MMM brought models closer to reality. It is important to note that before considering the calibration, necessary factors like objectives, investment, purchase cycle etc. must be considered.
- One of the most important learnings from this calibration was that practitioners could be easily misled by the value of the calibration by just looking at the impact of model fit. The impact of calibration might look minimal from a fit POV (e.g., R-Square and MAPE) as this research analysis was supplemented with lifts from only one media channel (Meta).
- If calibration can be expanded to multiple channels, predictions and attribution can be further fine-tuned with this information.

Incrementality from experimentation may represent the short-term marketing strategy impact during the test period and might be difficult to use the test results alone for future budget allocation optimization. A unified approach through calibration rather than using only one approach would lead to better informed decisions but it will require more effort in terms of planning and budgets.

	Improved Accuracy	Stronger predictive ability of MMM measurement	Rsq MAPE
	**Better Business Validation**	Contributions established by MMM are sourced and supported by experiments – attributions are keener	Reduce misalignment of contributions to other tactics

Figure 9: Calibration advantages

In conclusion, for better and more confident decision making, it is recommended to calibrate MMMs with insights from lift studies following the approach illustrated in this paper.

- 2 in 3 of analyzed MMM studies significantly changed “Meta” ROI results after calibration and average variation of 25% in ROI is observed post calibration
- For more accurate incrementality measurement it is recommended to include 2 or more lift studies during calibration.

AUTHORS

Narasimha Rao

Vice President, Analytic Edge

ANALYTIC EDGE
enabling sharper insights

ANALYTIC EDGE provides technology-enabled analytics solutions in marketing & sales effectiveness.
We integrate technology, industry knowledge, and cutting-edge statistical techniques to deliver “fast, cost-efficient and actionable business insights”