Exploring Marketing Mix Modeling (MMM) and Conversion Lift Experiment (CLE) blending

The Alshaya Group/H&M Case developed by Deloitte

Contents

Background to the Alshaya Group/H&M case

Alshaya Group ('Alshaya'), is a leading international franchise operator **managing franchises in a large spectrum of consumer sectors,** including apparel, health & beauty, hospitality and leisure.

Alshaya's portfolio extends across MENA, Russia, Turkey and Europe, with thousands of stores, cafes, restaurants and leisure destinations, as well as a growing online and digital business. Franchise partners include Starbucks, Mothercare, Debenhams and H&M, whose case will be discussed through this paper.

H&M's business idea is to offer fashion and quality at the best price in a sustainable way, and H&M group has 49 online markets and more than 4,900 stores in 72 markets, including in franchise markets. In the United Arab Emirates (UAE), Alshaya operates 30 H&M stores and the brand's online trading site.

H&M e-commerce in the UAE has been expanding, fueled by Alshaya's marketing strategy over time, which is the perfect starting point to develop complex analysis that allows Alshaya and H&M to have a better understanding about how marketing investments are driving sales. Nowadays, when analyzing a retail company, **a key point of any research is the measurement of both, the online and offline world,** to obtain a whole picture of the differences and synergies between them.

MMM gives the opportunity to have a comprehensive and strategic view of marketing activities effects and their impact on business

Specifically, **Alshaya was interested in measuring the effectiveness of its activity** from June 2018 to October 2019 **in Online Media** (Paid Search, Facebook, Instagram, GDN, Snapchat, Criteo, Social Affiliate and YouTube) and **Offline Media** (Magazines, Mall Branding, Outdoor, Radio and SMS), but also taking into account the contribution of **other drivers affecting simultaneously** the online and offline orders of H&M in the United Arab Emirates:

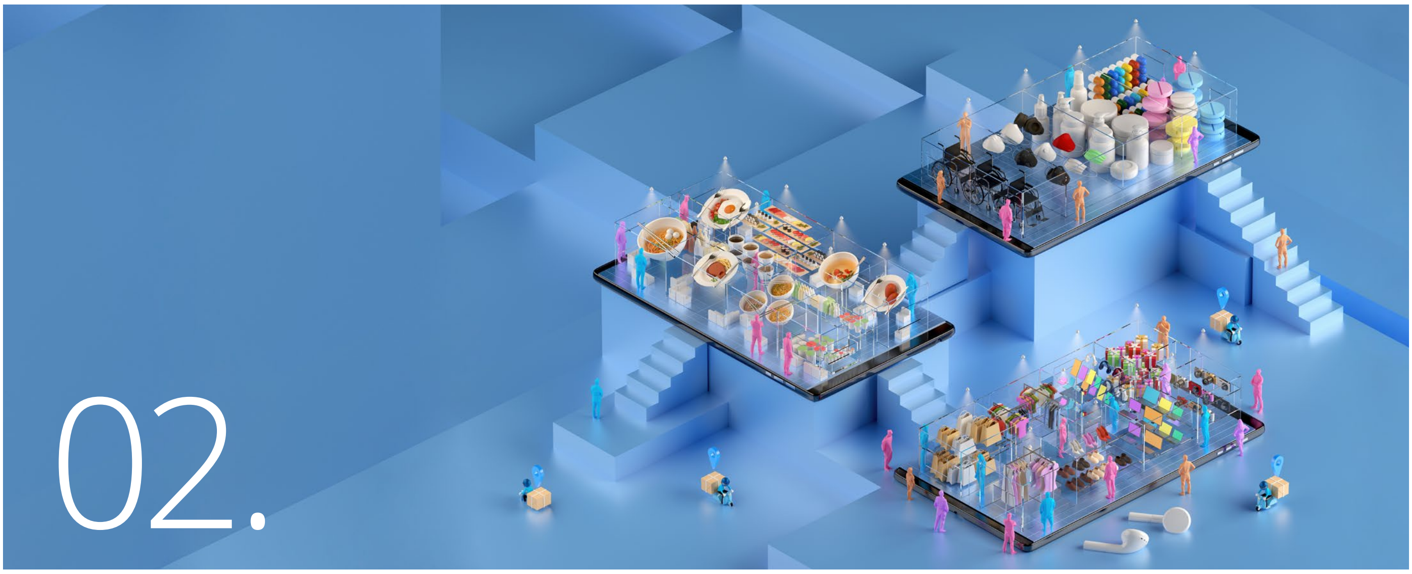
- Promotional activity clustered into different categories (low, mid, high and seasonal).
- Price and its impact on offline and online orders.
- Special releases of products.
- Other exogenous variables affecting time series such as seasonality, economic sentiment, unemployment rate, weather, etc.

To achieve this objective, **Deloitte used a holistic and global methodology,** which is able to consider in a single and unique measurement of all of these levers impacting the consecution of orders simultaneously. Marketing Mix Modeling (MMM), which consists of building statistical models analyzing the time series of one or several KPIs of interest (web sessions, online orders and offline orders in our case), was the methodology chosen. After modeling these KPIs, we converted order volume to revenue through the spending per order, so we were able to compare the efficiency of online and offline media.

MMM gives the opportunity to have a comprehensive and strategic view of marketing activities effects and their impact on business, with the possibility of analyzing its evolution over time. Additionally, it allows a second stage of analysis: MMM results can be used to optimize media budget, which ensures the best possible allocation of resources. It means we can look for the optimal distribution of media spending to find the split which ensures the maximization of revenues.

In this study we also tested an additional measurement approach, **Conversion Lift Experiment (CLE),** which is a kind of **lift test with which we can track conversions,** including standard and custom events, to understand the true value of Facebook advertising and how well it performs independent of other marketing efforts.

And, finally, we studied **the implementation of a methodology combining both methods.** Our objective was a first attempt to calibrate and harmonize results coming from different sources of measurement, avoiding silos and offering clients a **unified vision to better match ground truth and to make more informed marketing decisions.**



Top line methodologies (MMM & CLE)

What is MMM and what is CLE? What are the differences and how can they be blended? Both methodologies share the same objective, which is measuring the return of media investment (**ROAS:** Return On Advertising Spending) and its incrementality effect, but they use different approaches.

On the one hand, **MMM uses aggregate historical time series data to model sales (or other KPIs) outcomes over time,** as a function of advertising variables, other marketing variables, and control variables like weather, seasonalities, market competition and other external factors. 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 and Digital last year? 2) How should my media budget be allocated to maximize sales? 3) What will the sales be if we spend more money next year? Typically, 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.

Historically, marketers tested MMM with linear regression models. Regression models assume a linear relationship between the dependent variable (i.e. sales or other KPIs) and independent variables (i.e. Marketing activities and other external factors). The model splits sales into two main components: (1) Baseline, or the sales we would obtain without any marketing tactics (effect of structural variables as trends, seasonality, Brand strength. etc.) and (2) Marketing Contribution, or the incremental sales generated from each marketing tactic (i.e. offline and online media), where each coefficient measures the impact of every extra Euro invested in each marketing tactic onto sales.

Functional form of linear regression model in a marketing framework

$$y_t = \beta_0 + \sum_{j=1}^{m} \beta_j x_{jt} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma^2)$$

y_t: dependent variable (e.g. Sales) at time t
x_{jt}: advertising spend / activity for j^{th} marketing variable at time t
m: number of marketing variable
β_0: intercept of baseline in MMM
β_j: coefficient for j^{th} marketing variable
ϵ_t: error term, disturbance term, or noise. It captures all other factors which influence the dependent variable y_t other than the regressors x_{jt}.

As for any MMM case, one of the **key features that has to be modeled is the Baseline and its seasonality.** In this project, **Facebook's algorithm Prophet** was used for this purpose. Prophet relies on **Fourier series** to provide an accurate model of periodic effects, modeling in an accurate manner both the baseline and its seasonality. In order to infer trend and yearly seasonality, the Prophet forecasting model **decomposes time series into trend, seasonality and external regressors,** and does it in a powerful way. There are two trend models:

- Nonlinear saturating Growth
- Linear trend with changepoint detection

Although the previous model is intuitive, it is not able to account for modeling advertising effect:

- **Advertising saturation** (shape effect)
- **Non-immediate effect** (carryover effect)

The non-linear relationship between Advertising and Business can be measured through advanced modeling techniques

To account for these patterns, we needed to introduce non-linearities in the model and this fact immediately implies that obtaining estimates of the parameters will be non-trivial. We decided to combine this approach (Prophet algorithm) with a **model that allows us to measure shape effect and carryover.**

To model the **carryover effect of advertising,** we transformed the time series of the media variable in one channel through the adstock function:

Ad-stock function

$$x^*_{t,m} = adstock(x_{t-L+1}, \dots, x_{t,m}; w_m, L) = \frac{\sum_{l=0}^{L-1} w_m^d(l; \alpha_m; \theta_m) x_{t-l,m}}{\sum_{l=0}^{L-1} w_m^d(l; \alpha_m; \theta_m)}$$

Where:

$$w_m^d(l; \alpha_m; \theta_m) = \alpha_m^{(l-\theta_m)^2}$$

With $l = 0, 1, \dots L-1, 0 < \alpha_m < 1, 0 \le \theta_m \le L-1$

Here, we are estimating today's impact of media as a weighted average of its previous values while also introducing a delay effect.

To model the **shape effect** of advertising, media variables need to be transformed through a curvature function, such as the **Hill function.**

Curvature function: Hill function

$$Hill(x_{t,m}, K_m, S_m) = \frac{1}{1 + \frac{x_{t,m}}{K_m} - S_m}$$

With $x_{t,m} \geq 0$, $S_m > 0$ (slope), $K_m > 0$ (half saturation point)

In order to introduce non-linearities we use a **Bayesian approach** using Markov Chain Monte Carlo (MCMC) algorithms. The Bayesian framework allows us **to incorporate prior knowledge into model** estimation as prior distributions on the parameters. The prior knowledge may come from industry experience or previous media mix models of the same or similar advertisers. From a practical viewpoint, while in the classical approach we obtained an estimate of a parameter (a single number), maybe with a measure of statistical relevance or goodness of fitness (as a p-value or a confidence interval), now we will obtain:

- Subsequent distributions samples, i.e., a collection of values obtained sampling the distribution that characterizes the parameter.
- A credible interval to quantify the interval within which an unobserved parameter value falls with a specific probability.

In a few words, the input to Bayesian modeling is both the data and our a priori knowledge about the case we are modeling, also known as priors. Some examples of priors:

- We a priori know that **the effect of advertising is not immediate** and lasts over time.
- We a priori know that **the media investment will reach a saturation point.**
- We a priori know that **the impact of advertising over sales is non-negative** (the more investment, the more sales).

On the other hand, it is increasingly common to use **controlled experiments to measure media effectiveness and to maximize the incrementality of an advertising campaign.** In the most common variant —known as A/B, or split testing— the target population is divided into two twin groups (presenting equivalent socio-demographic traits, web navigation behavior, etc.): a test group, where members are shown adverts, and a control group, where members are not shown adverts. As the two groups differ only for having seen/not seen the advertising, the uplift in the metric of interest (e.g. total sales, or number of web registrations or app installs) between the test group and control group is the incrementality of the campaign.

Facebook offers advertisers the opportunity **to measure the incrementality of their campaigns via Conversion Lift Experiments (CLE) or lift studies.** Conversion Lift tests, which measure the performance of your advertising objectives, such as conversions collected with Facebook Pixel, between test and control groups of people who do and don't have the opportunity to see your advertising[1].

Summarizing: **MMM offers a global and strategic vision of measurement of all drivers affecting sales.** Working with aggregated data (at daily or weekly level) and with an inferential perspective, it gives a holistic and "large" view: anyway, it can suffer by multicollinearity or endogeneity problems, elements affecting the full accuracy of the measurement (i.e. models producing correlational, not causal results). On the other hand, **CLE is fed with more granular data: it tests a specific hypothesis and it can ensure this accuracy measurement,** gauging the true incremental causal uplift (it's the "gold standard"), but it works in a very specific and delimited perimeter (Facebook exposure in our case), with no visibility on the rest of drivers moving business. In other words: lift studies can be used to answer causal questions, but they are impractical or infeasible in the context of answering all the questions an advertiser may have about the effectiveness of their advertising channels[2].

It's clear each method presents strengths and weaknessess points: for this reason, the next step is to find a way to bring together these measurement methods to get a rounded view of their effectiveness. In other words: we search for a convergence between methodologies, trying to combine the best of both worlds and offer a **new unified vision of measurement[3]**.

Facebook offers advertisers the opportunity to measure the incrementality of their campaigns via Conversion Lift Experiments (CLE) or lift studies. Combining the results of CLE and MMM is a common challenge for the advertising industry

1. See Liu, C. H. Bryan, Elaine M. Bettaney, and Benjamin Paul Chamberlain (2018), *Designing Experiments to Measure Incrementality on Facebook,* arXiv E-Prints, June, arXiv:1806.02588.

2. There are two additional differences when we compare the two methodologies. The first one refers to the temporal vision: MMM looks at the past, using observational data viewed retrospectively, while CLE looks at the future, requiring forward planning and control over exactly how the advertising is delivered. The second one refers to the duration of media effectiveness: CLE is used to measure short term marketing impact, while MMM extends this measurement, incorporating also long-term effects.

3. The idea is to complement traditional MMM validating and calibrating results with experiments. Validation is the process of checking a model against ground truth, while calibration is the process of selecting or tuning a model to better match ground truth. Running experiments alongside MMM has two important benefits:

    - It informs the hypotheses the modeler has about the performance of ads
    - It enables calibration across measurement efforts.
    - See Facebook IQ, Measuring Facebook Accurately in Marketing Mix Models, available at: https://www.facebook.com/business/news/insights/considerations-for-creating-modern-marketing-mix-models.

03.

Calibration / validation method algorithm

How can we calibrate MMM with CLE results? We propose a methodological approach where we use lift studies to "constraint" MMM, helping it to learn and reflect the effect size measured in CLE. As CLEs are Randomized Experiments whose incrementality can also be translated to a ROAS estimate, we can expect that, whenever CLE results are available for some period included into the period when a MMM is fitted, they should allow us to calibrate properly the MMM. In this section, we outline an optimization methodology in order to perform this kind of calibration[4].

We present a simple case, with an MMM where the response ($\boldsymbol{Y}$, be it sales or conversions or whatever appropriated) to media, control and other predictive variables ($\boldsymbol{X}$) is assumed linear:

$$y_i = \beta_0 + \sum_{j=1}^{P} \beta_j \times x_{i,j} + \epsilon_i$$

The linear model estimation in matrix form is:

$$\hat{Y} = X \times \hat{\beta}$$

The most popular estimation method for linear models is Ordinary Least Squares (OLS), in which we pick the model coefficients to minimize the residual sum of squares:

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

What is achieved by $\hat{\beta} = (X^T X)^{-1} X^T Y$.

4. This methodology is a first approach to calibrate MMM and CLE: we don't expect it to be the "gold standard", but a first proposal to open the discussion around MMM and CLE integration, a field which started to be explored only in these last few years.

On the other hand, the estimated ROAS for each predictive variable x_j is:

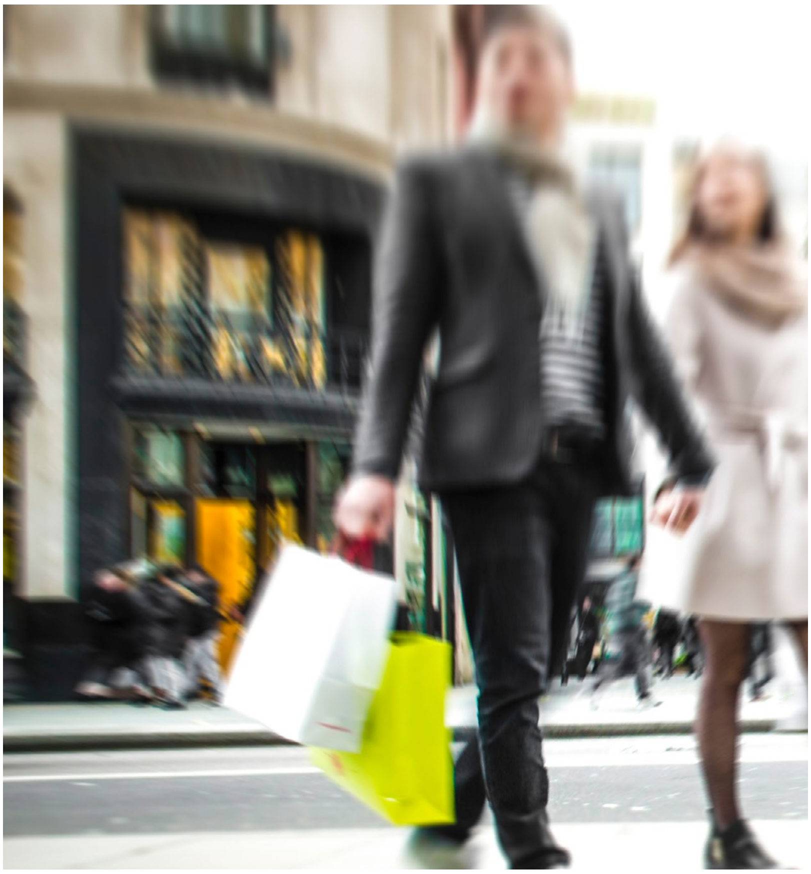
$$ROAS_j = \frac{\sum_i \beta_i \times x_{i,j}}{\sum_i x_{i,j}}$$

It is possible to modify OLS via a weighted scheme in order to grant more or less relevance to specific observations. To do so we must find the estimates of the β coefficients that minimize not the sum of the squared errors as before, but the weighted sum of the squared errors:

$$\sum_{i=1}^{N} w_i \times (y_i - \hat{y}_i)^2$$

We can take advantage of this to solve our problem using an optimization approach with weights to fit the MMM in order to increase the relevance of the observations in the CLE period. In other words, weights allow **minimizing the difference between the ROAS estimated by MMM and the ROAS estimated by CLE[5].** With this approach (which can be more sophisticated) we can obtain a convergence of ROAS between MMM and CLE, leading to find a first solution to blend and integrate both methodologies.

In any case, our suggestion is to **check carefully all results of MMM before applying the calibration with CLE.** Theoretically, we should expect a natural convergence of Facebook ROAS using separately MMM and CLE (they should provide similar results if both measurements are accurate). Anyway, in some cases, as we could not obtain this convergence (which could be explained by limitations in terms of granularity of some MMM data variables, multicollinearity problems, etc.), this calibration methodology can be an optimal balance to improve the quality and completeness of measurement.



Conversion Lift Experiment (CLE) is fed with more granular data than MMM. Based on CLE results, we can calibrate MMM results, minimizing the difference in the CLE period

5. To find the set of weights we can use standard optimization methods such as BFGS (quasi-Newton or variable metric algorithm), CG (conjugate gradients), L-BFGS-B (BFGS with a lower and/or upper bound), Finite-difference approximation, Simulated Annealing or Genetic Algorithms.

04.

Results for the Alshaya / H&M discussion

The results provided by the MMM study offered the client a full visibility on a series of issues that were difficult to evaluate previously, due to the complexity of the retail market environment and the simultaneous presence of several sets of variables interacting with each other, which makes it difficult coming up with a comprehensive evaluation of drivers moving the brand's business in the UAE.

After applying prophet algorithm and combining it with a Bayesian approach to identify Carryover and Shape Effects, it was possible to obtain **the following relevant insights:**

- **Baseline and price play a large role,** especially in offline orders.
- **Drivers as advertising, promotions and special releases have a greater impact on the online channel.**
- **Online advertising has a significant contribution on offline orders.** Although online advertising plays a large role on online sales, it also influences the number of offline sales.
- **Positive ROAs for all the invesments.**
- **Facebook is the medium that has generated more sales and has reached a higher ROAS than average.**
- When comparing both 2018 and 2019, **Facebook's ROAs has increased more than 20% in 2019.**
- **It is possible to increase the level of investment** before reaching saturation level.

Additionally, Facebook ran an experimental design with **randomized control trials** (the **CLE** study we presented in the previous paragraphs). The experiment, held in the period between 21st of May and 5th of June 2019, had the objective of providing an additional measurement of FB effectiveness, to be compared with MMM study. The experiment showed a lift in the test group quite aligned with the one provided by MMM study[6], making it not necessary the implementation of the calibration method explained above, and giving full robustness and reliability to the results obtained with both methodologies.

6. MMM ROAS for Facebook was around 40% higher with respect to CLE study, but we decided not to calibrate MMM with CLE results for two main reasons explaining this deviation: (1) there was not a complete overlap in terms of investment: performance Facebook ROAS measured with CLE and total Facebook ROAS measured with MMM; (2) MMM can obtain amplified results due to its additional focus on long term impact of media effectiveness.

05.

Applicability of calibration for other cases

The new digital platforms are evolving very quickly, generally faster than the methodological improvements of MMM. Even if MMM is nowadays much more sophisticated and accurate with respect to the past (and it's facing a phase of resurgence, as it is the most comprehensive technique embracing all businesses of marketers, with no relevant issues related with privacy and data protection), the focus it offers remains strategic and "macro". On the other hand, modern platforms such as Facebook's family of apps and services have dynamic, nuanced advertising possibilities and are constantly evolving to keep up with shifts in consumer behavior. This makes **measuring their true impact much more difficult** compared with established offline channels like linear TV or radio, and this constant evolution increases the importance of using innovative methodologies (Prophet algorithm, Bayesian models, ...) for measuring Facebook and other media in MMM.

For this reason, Deloitte recommends to work on methodological measurement improvements, that would be able to guarantee two relevant dimensions for the marketer: (1) to be sure of isolating all drivers of sales, giving the possibility of a full media measurement (not only one digital channel) and optimization scenarios; (2) to be sure of incorporating in our models the dynamic evolution of contributions over time and the complexity of the modern digital landscape.

Working with MMM and CLE is an opportunity to double check assumptions, to choose between models and to test again known outcomes. And the integration of both methodologies offers a solution to calibrate models to make them more accurate and better reflecting reality. In this way, the combination can clearly **boost the confidence in both outputs and add credibility to the effectiveness measurement results** provided to the client.

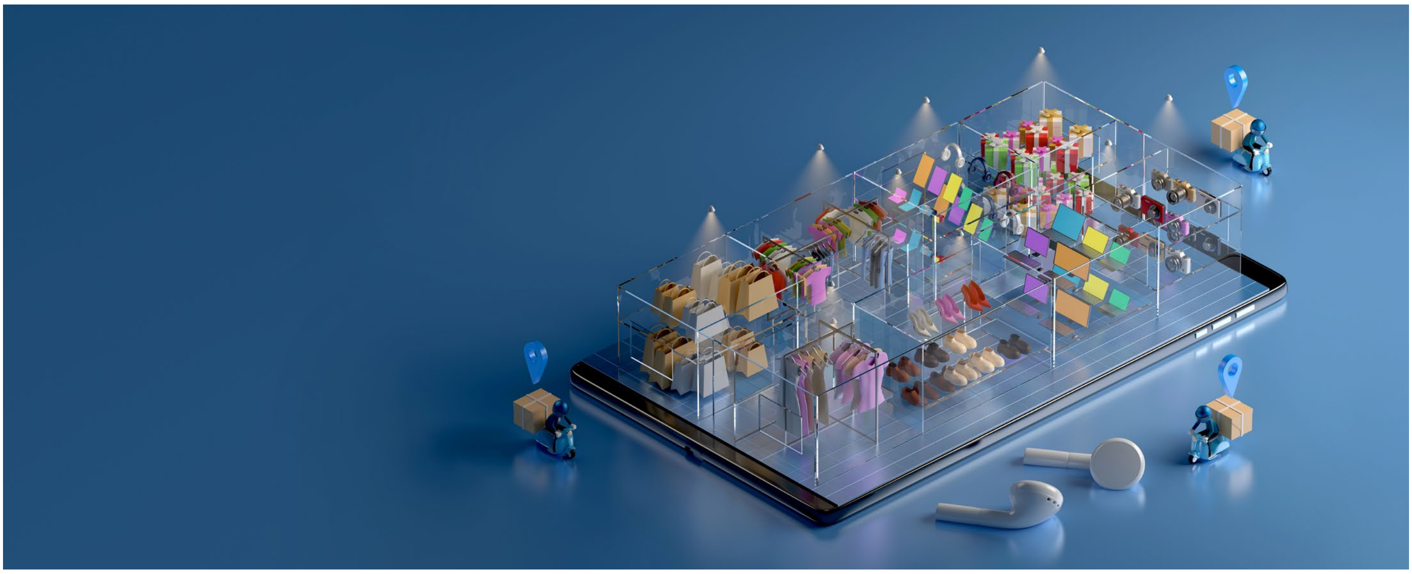
In any case, a few caveats have to be considered **before deciding to calibrate MMM with CLE:**

- **Check the complete temporal and investment overlap** between MMM and CLE measurement, to make a 100% complete comparison;
- **Consider long term effects of FB** that could not be measured entirely with CLE (effects usually measured by MMM[7], even if with a certain margin of error);
- **Apply the calibration with a conservative perspective,** testing it when there is a significant discrepancy between the results provided by MMM and CLE;
- **Control the impact of the calibration of MMM with CLE on the results of other media** effectiveness, to fully understand the implications of this refinement method for the case.
- **Think proactively,** before the beginning of the project, **on a methodological plan to integrate MMM and CLE,** to perfect align data and time requirement, the planning and the collaboration between the different stakeholders involved, etc. In many cases, this blending is conducted *a posteriori,* and it can complicate the achievement of an accurate picture of marketing effectiveness.

The pace of technological transformation is accelerating every day, and it's always more challenging to guarantee a precise measurement in a media landscape so fragmented (in terms of media offer, devices and platforms, advertising formats, audiences, etc.) and with a consumer behavior so volatile (in terms of media consumption, purchase channels, customer journeys, etc.). We are in the first steps of this new exciting journey, and there is a long way to go and to explore ahead. The blending of different methodologies, integrating strategic with tactical vision, seems to be the optimal path to continue ensuring the best methodological approach. Being aware perfection does not exist, even if we continue to strive for chasing it.

The blending of different methodologies, integrating strategic with tactical vision, seems to be the optimal path to continue ensuring the best methodological approach to measure ROAs

7. The methodologies used in MMM to measure long term effects of advertising are usually three: (1) media variables tested with higher levels of ad-stock, (2) Brand metrics, as Brand Awareness, Image or Consideration, modeled and integrated in intermediate outcomes before sales, creating Multi-stage models, (3) "baseline cleaning", applying techniques as UCM or BSTS to analyze residuals and identifying an additional long-term component of media effectiveness.



References

William Boulding, Eunkyu Lee, and Richard Staelin. "Mastering the Mix: Do Advertising, Promotion, and Sales Force Activities Lead to Differentiation?" In: Journal of Marketing Research 31.2 (1994), pp. 159–172. issn: 00222437.

Gerard J. Tellis. "Modeling Marketing Mix". In: The Handbook of Marketing Research (2006). doi: 10.4135/9781412973380.n24

Marnik G. Dekimpe and Dominique M. Hanssens. "The Persistence of Marketing Effects on Sales". In: Marketing Science 14.1 (Feb. 1995), pp. 1–21. issn: 1526-548X. doi: 10.1287/mksc.14.1.1.

Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, and Jim Koehler. Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects. Tech. rep. Google Inc.,2017.

Yunting Sun, Yueqing Wang, Yuxue Jin, David Chan, and Jim Koehler. Geo-level Bayesian Hierarchical Media Mix Modeling. Tech. rep. 2017.

David Chan and Mike Perry. Challenges and Opportunities in Media Mix Modeling. Tech. rep. Google, 2017.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Elements of statistical learning: data mining, inference and prediction. 2nd ed. Springer, 2009.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. isbn: 9781439840955.

P.E. Rossi, G.M. Allenby, and R. McCulloch. Bayesian Statistics and Marketing. Wiley Series in Probability and Statistics. Wiley, 2012. isbn: 9780470863688.

J.O. Rawlings, S.G. Pantula, and D.A. Dickey. Applied Regression Analysis: A Research Tool. Springer Texts in Statistics. Springer New York, 2001. isbn: 9780387984544.

Authors

Carlos Real Ugena
Senior Manager – Deloitte

Miguel Conde de Simón
Senior Specialist – Deloitte

Mujtaba Anwer
Head of Data Science – Alshaya

Davide Fabrizio
Director – Deloitte