Deloitte.

FACEBOOK

The future is modeled

A how-to guide for Advanced Marketing Mix Models

As Albert Einstein once said

“The crisis is the greatest blessing for people and nations, because the crisis brings progress. Creativity comes from anxiety as the day comes from the dark night”.

During the past years, businesses of all types and sizes have been experiencing a **deep digital transformation**

(the crisis Einstein was talking about).

Therefore, advertisers are more than ever willing to understand more about the marketing measurement solutions landscape (Econometric marketing response models, Attribution and Experiments, among others).

Certainly, one of the main concerns is how to implement effective measurement techniques for online and offline media investments, which are also **resilient to signal-loss scenarios and privacy-compliant.** Advertisers are eager to define a **unified measurement strategy** that can work on both, tactical and strategic, media decision making.

Marketing Mix Models (MMMs) are now still more crucial to respond to these needs of a converging and multi-technique approach, attracting Multi-touch attribution (MTA) and experiments to its orbit.

So... What are the main trends around MMM?

- The marketing industry is experiencing a boom in in-house **MMM development**, digital-native advertisers (and traditional ones) are turning in mass to MMM, creating a protected perimeter of development with internal expertise and pushing hard for MMM UI access and API. However, good MMMs are hard to build without proper guidance.
- **MMM crosses a phase of progressive "democratization"**. Gone are the times when the recipe of the modelling was a secret carefully protected by vendors: now we move in a co-creation and transparent approach (vendor, advertiser and media companies as partners), with initiatives of global **open source MMM code** releasing.
- **MMM works with more granular data than in the past,** taking advantage of more sophisticated modelling techniques, getting closer to insights previously provided only by MTA (splits by format, device, creative, audience, placement, campaign objective, etc.). This leads Direct response, Digital native, Gaming and App companies to turn in mass to MMM, which is perceived as the best compromise between the business perspective and a dynamic delivery of results.

Project Robyn[1], created by Facebook's Marketing Science Partners, as well as two of the open source solutions it includes: Facebook's open-source platform Nevergrad[2], and Facebook's Prophet[3] are nowadays three of the most relevant initiatives in this direction. Moreover, Facebook's MMM feed is an example of the data set ideally constructed for use in Marketing Mix Modelling providing high level granularity and level of details about the advertiser's activity across Facebook. Such datasets should be considered as an industry standard and demanded by advertisers across all channels enabling more actionable models built at a higher cadence.

In this renovated context, MMM can take several directions of evolution. Multiple inputs are positive to guarantee a continuous improvement of the quality of the insights provided. Anyway, the advertisers have to be aware of the risks that a fragmentation of methodologies, data, technologies and development process can create. Considering the ubiquitous access to information new technologies have brought with them, it is very difficult today to evaluate the quality and veracity of high volumes of information. In media effectiveness measurement, we face the same problem: are modellers and decision makers able to recognize or to build a good MMM? It is important to give an answer to this question, because the experience tells us it is hard to build valid and robust models without a clear guidance.

Guidance

A good MMM responds to a clear need of advertisers: to have a **fast and precise solution to help -and improve gradually- the marketing decision making process**. What are the main ingredients ensuring this outcome? In our vision there are **6 relevant recommendations to be followed:**

1. Consider modeling your funnel in full

"Model all the KPIs that matter to your business, not only sales" could be the quote summarizing this point. MMMs offer omni-channel capabilities like no other methodology in the measurement landscape: models can measure the impact of media on the upper funnel, such as Brand awareness or Website visits, or go down to the lower funnel reaching Sales in the online channel or in physical stores. This **global and holistic measurement**, capturing the whole business dynamics with its connection between different steps of the funnel, is fundamental if advertisers want to discover how media impact business generation in the short and long term and how different channels interact mutually to create synergies and multiplicative effects.

At the same time, we have to be aware that measurement becomes more complex each time models move spatially and temporally away from the advertising exposure, increasing the difficulty of the exercise, with potential impact on automation and the scalability of the models and the subsequent agility of decision making.

The **balance between completeness and quickness/agility** is one of the most important decisions to be made before starting any MMM project.

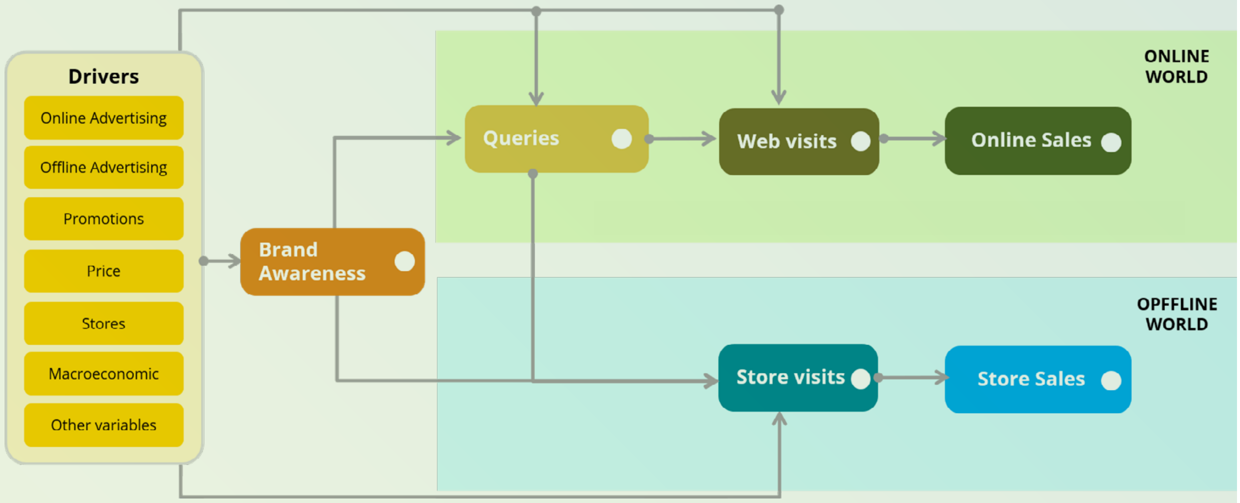
Full customer journey modeling

Why full Customer Journey modeling?

- The entire business, throughout the all levels of the funnel, and all media and content types can be analyzed.
- Marketers can achieve full-funnel marketing performance:

Measuring the effect of the media in the mid and long term, generating the brand.

Measuring the effects of campaigns with a sales objective and others whose objective is to generate a brand or traffic to the web.



To develop the complete customer journey, advertisers can work with **Stacking or Stacked Generalization, an ensemble machine learning algorithm that uses a meta-learning algorithm** to learn how to best combine the predictions from two or more base machine learning algorithms. The architecture of a stacking model involves two or more base models, often referred to as **level-0 models**, and a meta-model that combines the predictions of the base models, referred to as a **level-1 model**.

- **Level-0 Models (Base-Models):** models fit on the training data and whose predictions are compiled.
- **Level-1 Model (Meta-Model):** model that learns how to best combine the predictions of the base models (is trained on the predictions made by base models on out-of-sample data).

The **benefit of stacking** is that it **can harness the capabilities of a range of well-performing models on regression task and make predictions** that have better performance than any single model in the ensemble.

2. Customize the solution to suit the business.

MMM offers a remarkable added value: the combination of business and media perspective in one unique model, with the possibility of a global and combined understanding of structural and dynamic drivers of growth. MMM should follow a global architecture of modelling offering robustness in terms of methodology, but it also needs to be **customized and adapted to the specific context of the advertiser**.

The Bayesian framework approach helps, including flexibility and customization related with three dimensions: (1) ad-stock effect, which incorporates the fact sales uplift does not decrease immediately when we stop investing (due to the carry over effect of awareness in the following weeks); (2) diminishing return curve, as we know the relationship between media investment and sales is not linear and presents from a certain point a saturation effect, where an increase of investment produces only marginal incremental sales; (3) non media variables impact (pricing, promotions, competitors, new products releases, new business units, etc.), with the possibility of incorporating prior knowledge or restraints into the model to suit the advertiser business.

Bayesian framework approach

Why a Bayesian framework approach?

— It allows to incorporate prior knowledge into model estimation as prior distributions on the parameters.

— Bayesian inference treats the parameters as random variables and is based on their posterior distribution $p(\Phi|y,X,Z)$ given the data and the prior distribution $\pi(\Phi)$ (Bayes Theorem): $p(\Phi|y,X,Z) \propto L(y|X,Z,\Phi)\pi(\Phi)$

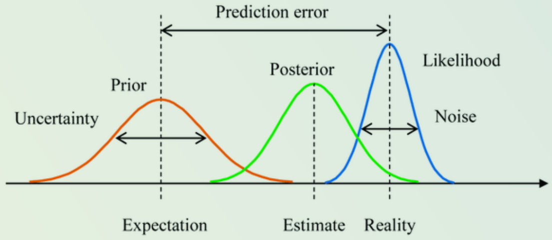
Where:

Φ denotes the vector of parameters in the model

X denotes all the media variables and Z all the media variables in the data.

$L(y|X,Z,\Phi)$ is the log-likelihood given the data and the parameters.

— **Bayesian models are extremely flexible:** you can model complex situations, like those modeled as non linear models, or those needing non linear variables with unknown parameters, otherwise ad-hoc, but which now can be estimated using non-normal distributions. In the case of regression, you can fit from simple linear regressions to multilevel generalized linear models.



We obtain:

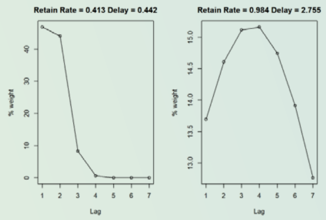
— subsequent distributions samples, i.e., a collection of values obtained sampling the distribution that characterizes the parameter;

— and a credible Interval to quantify the interval within which an unobserved parameter value falls with a specific probability.

— It allows to introduce non-linearities on media investment:

Ad-stock function

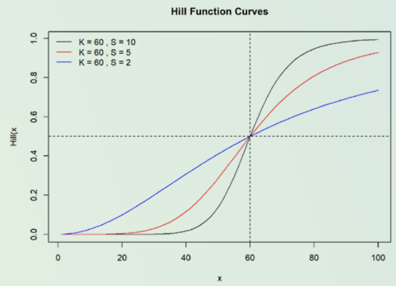
$$X^*_{t,m} = adstock(x_{t-L+1,\dots,}x_{t,m}; w_m, L) = \frac{\sum_{l=0}^{L-1} w_m^d\,(l;\alpha_m;\theta_m)x_{t-l,m}}{\sum_{l=0}^{L-1} w_m^d\,(l;\alpha_m;\theta_m)}$$



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)



3. Unlock the power of AI-automated code to increase scalability and reduce human bias

Reducing human bias is one of the biggest concerns of traditional MMMs: the use of **automated code**, which substitutes old manual code and modelling approach, is the present and the future of a more advanced approach.

The limitation of the biases produced by analyst subjectivity offers a more objective perspective, also making **MMM more transferable and scalable.** Scalability is one of the priorities of global groups or corporations owning different sets of brands, to guarantee homogeneity in terms of approach, comparability between different brands/markets and, in the end, a more efficient process and a better Cost/Benefit ratio. All these automation processes give another advantage: we can increase the number of models developed and the modelling cadence of the updates (even to weekly or daily level), providing a quasi-real-time solution to understand the performance of marketing activity and accelerate the relative decision making.

In econometric models, the work of analysts will anyways be required to understand which are the main business assumptions and the interpretation of model results, automation helps here to reduce bias and provide model alternatives that make business sense for analysts too.

Advanced MMMs should leverage automated code and techniques that are scalable and objective. Artificial Intelligence techniques such as those within **Facebook's Nevergrad open-source platform**, optimize the way to explore over possible model solutions reaching optimum results faster and in a more precise way than traditional algorithms. And **Facebook's Robyn, the open-source code for semiautomated MMM using Machine Learning techniques**, pursues a similar objective: increase scalability and objectivity of the analysis.

Robyn, Nevergrad and Prophet

Why Project Robyn?

- **Robyn is an open-sourced** code for semi-automated **Marketing Mix Model using Machine Learning techniques**.
- It is **advanced**, offering **automated** code, making the modeling process significantly faster to run. It can be calibrated and validated using real world experiments, fully customisable (adstock, automated seasonality, ridge regression).
- It is **controllable and scalable**: standardised and stable code to limit analyst bias and subjectivity, making models scalable and transferable.
- It is **actionable**: continuous modelling helps understand the performance of marketing activity in almost real time. No need to wait until the campaigns have finished.

Why Nevergrad?

- **Nevergrad is an easy-to-use optimization toolbox for AI researchers**, including those who aren't Python geeks. Optimizing any function takes only a couple of lines of code.
- **The platform provides a single, consistent interface to use a wide range of derivative-free algorithms**, including evolution strategies, differential evolution, particle swarm optimization, Cobyla, and Bayesian optimization. The platform also facilitates research on new derivative-free optimization methods, and novel algorithms can be easily incorporated into the platform.
- **Nevergrad also provides generic algorithms that can better adapt to the structure of a particular problem**, including by using specific mutations or recombination in evolutionary algorithms, through the new parametrization system.

Why Prophet?

- **Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit** with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.
- **Accurate and fast.** Prophet is used in many applications across Facebook for producing reliable forecasts for planning and goal setting. We've found it to perform better than any other approach in the majority of cases. Prophet fits models in Stan so that you get forecasts in just a few seconds.
- **Fully automatic.** Get a reasonable forecast on messy data with no manual effort. Prophet is robust to outliers, missing data, and dramatic changes in your time series.
- **Tunable forecasts.** The Prophet procedure includes many possibilities for users to tweak and adjust forecasts. You can use human-interpretable parameters to improve your forecast by adding your domain knowledge.
- **Available in R or Python.** The team has implemented the Prophet procedure in R and Python, but they share the same underlying Stan code for fitting. Use whatever language you're comfortable with to get forecasts.

4. Build models that are flexible and adapt to future scenarios

To improve model predictability we have to address a typical issue related with MMMs: the presence of **highly correlated explanatory variables**, usually known as multicollinearity, as in the case of simultaneous investment in different media channels. During the preparatory analysis it is imperative to understand the reason for the highly collinear data in order to advise appropriate remedy. There are typically a few common reasons why data can be correlated and hence result in multi-collinearity:

It can be due to “redundancy”, this is a situation when the analyst uses multiple variables to reflect the same or similar process (for example two or more variables to demonstrate economic conditions or seasonal shifts). Other situation is when there are “functional or direct dependencies” within the data set, this is when different variables influence each other (for example temperature and air pressure are dependent), or when variables are correlated due to formulas between them when they are measured (for example price is ratio of sales volume to revenue, and will always be negatively correlated to sales volume if this calculation was used to obtain the variable) The most common in marketing and business settings is the case of synchronised data, this is a situation when the variables each reflect a different process, but are moving in a synchronized way. This is the case of media channels during a marketing campaign; when all channels are phased on purpose to maximise combined effect of the communication. Of course, variables can be simply correlated just at random, which is something that can also happen but can be sometimes difficult to discern in a large data set, and accessed automatically For redundancy scenarios, the advice is to leave one of the variables out or or to replace them with a common factor, potentially a principal component. In case of functionally related data using the different techniques such as simultaneous equations can be advised.

For the highly correlated data in a synchronised marketing campaign, even if there is no magic and definitive solutions around this topic, regularization techniques, such as Lasso or Ridge regression, have demonstrated their reliability in order to face multicollinearity among many regressors, preventing overfitting and helping on variable mis-selection. This approach tends to **improve the predictive performance of MMMs**, providing more flexible models that allow deeper levels of detail and more robust results, with a better balance between analytical and business requirements of the projects.

Another main concern about flexible models that adapt to future scenarios is to be able to quickly feed with data on a weekly cadence in order to provide **updated MMM results at a faster pace** required by business decision makers. Rolling-window analyses for model's parameters stability is a good approach to ensure the stability of the marketing mix time-series model over time when refreshing data inputs more frequently. It is usual to suppose that times-series coefficients are constant in time, therefore, to examine the amount of instability in coefficients is essential. Below are the steps to take for a rolling window analysis:

1. Choose a rolling window size n. This will be the number of consecutive observations (let's assume weeks) per rolling window.
2. Define the number of increments between successive rolling windows, let's say it is one week in this case. Then you will have to divide the entire data set into T – n + 1 subsamples. The first rolling window contains observations for period 1 through n, the second rolling window contains observations for period 2 through n + 1 week, and so on.
3. Run the model using each rolling window subsamples.
4. Plot each estimate and point-wise confidence intervals over the rolling window index to see how the estimate changes with time. You should expect a little fluctuation for each, but large fluctuations or trends indicate that the parameter might be time varying.

Ridge regression

Why Ridge Regression?

— Allows analysing multiple regression data that suffer from multicollinearity.

— In this case, there are problems estimating the parameters. The usual least-squares regression equation is in trouble:

$\hat{\beta}=(X^TX)^{-1}X^Ty$ where XTX is singular or nearly so, and has some eigenvalues of zero or nearly zero, so inverting is not a good idea.

— Modified linear regression with least squares and by adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. Adding a ridge on the diagonal

$X^TX+\lambda I$ with $\lambda>0$, which increases all the eigenvalues by λ and takes the problem away: $\beta\hat{\lambda}=(X^TX+\lambda I)^{-1}X^Ty$

Ridge and linear regression

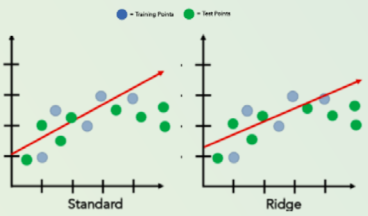
We can write out the optimization problem that ridge is solving, minimizing:

$$\sum_{i=1}^{n}\left(y_t-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

Where $RSS=\Sigma_{i=1}^{n}(y_t-\beta_0-\Sigma_{j=1}^{p}\beta_j x_{ij})^2$ (residual sum of squares)

$||\beta||_2=\sum_{j=1}^{p}\beta_j^2$ is an ℓ_2 penalty

And λ is the *tuning parameter*



Bayesian Linear Regression

The Bayesian form of a linear regression with regularization can be written as:

With **likelihood**: $\mu_i = \beta_0 + \sum_{n=1}^{N} \beta_n x_{in}$

$$Y \sim N(\mu, \sigma_Y^2)$$

And **priors**, e.g.:

$$\beta_n \sim N(0, (2\lambda)^{-1/2})$$
$$\beta_0 \sim N(0, \sigma_{02}^{\beta})$$
$$\sigma_Y \sim Inv\Gamma(\alpha_\Gamma, \beta_\Gamma)$$
$$\lambda^{-2} \sim \Gamma(\alpha_\lambda, \beta_\lambda)$$

Where $\beta n, \beta 0$,$\lambda{-}2$ and σY are **hyper-priors**.

5. Explore the relationship between experiments, multi-touch attribution and MMMs

Advanced MMMs offer a unique flexibility in the field of media effectiveness measurement: **they can be calibrated with experiments and attribution results**, to ensure a consolidated and blended solution containing incremental results.

There are three different ways to make these solutions converge into MMMs: (1) constraining MMM models with experiment and/or attribution results via priors, guiding modelling towards solutions that reflect the experiments' effect size; (2) selecting MMM models that minimize the distance between model predicted results and results from experiments or attribution and (3) validating MMM results by just comparing the predicted contribution for a channel compared to its experiment or attribution results within a certain period of time.

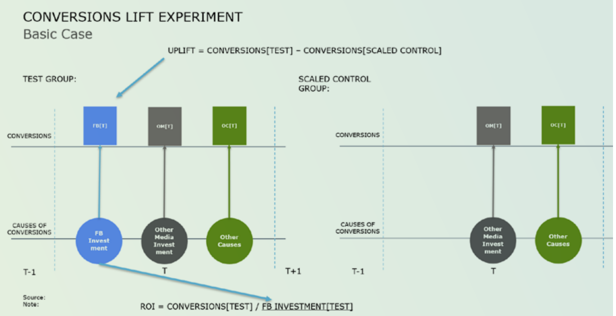
Working with different methodologies is an opportunity to double check assumptions, to choose between models and to test again known outcomes. And their integration offers a solution to **calibrate MMMs to make them more accurate and better reflect reality**. In this way, the combination can clearly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

MMM calibration with experiments

Why Calibration?

— Measure the performance of advertising objectives, such **as the difference in conversions collected** between people with media exposure as a treatment, versus non-exposed groups via **conversion lift, geo lift, etc.; or through MTA results**.

— Fine tune the model. It is straightforward to inject this new data via **model selection or prior knowledge** about the effects of the media variable during the period when the experiment took place.



Calibration process

1. Compute for that period the differences between the lift study/attribution results and the expected results from our (until now) uncalibrated model.

2. Add this new knowledge to the model.

Now we are going to sample simultaneously 2 posteriors from 2 interdependent log likelihoods:

$$\mu_t = \tau + \Sigma_{m=1}^{M} \beta_m Hill\,(x^*_{t,m}, K_m, S_m) + \Sigma_{c=1}^{C} \gamma_c z_{t,c};\; Y \sim N(\mu_t, \sigma_Y^2)$$

and

$$\mu_{fb} = \beta_{fb} Hill\,(x_{t,fb}, K_{fb}, {}^*S_{fb});\; Y_{fb} \sim N\,(\mu_{fb}, \sigma_{fb}^2)$$

Doing so, we will get the benefit of minimizing the error:

$$MAPEcalibration\ exp. = \frac{\Sigma_{k=1} |ROI_{MMM,k} - ROI_{exp,k}|}{n}$$

where $ROI_k = \dfrac{\Sigma_{m=1}^{M} \beta_m Hill\,(x^*_{k,m}, K_m, S_m)}{\Sigma_{m=1}^{M}\ x^*_{k,m}}$

6. Focus on actionable insights

More data and more sophisticated algorithms give the opportunity to “stress” the level of MMMs analysis, facilitating the **discovery of new patterns that some years ago remained latent and hidden under the surface**. Model different types of campaigns (e.g. prospecting vs. retargeting vs. modelling each campaign separately) is now an objective that can be achieved through MMM and not only MTA, obtaining actionable insights on performance and media budget allocation in almost real time, taking also advantage of the progress in terms of code automation. Using scorecards based on the use of context variables for media tactics within the modeling analysis period helps drive more actionability on results.

In any case models don’t have to lose the focus on the core of MMM: they still cover a **strategic perspective**, even if **models are moving progressively to an increasing coverage of a tactical one**. In this sense it’s also important to identify its limitations, the frontier that cannot be overtaken if modellers want to guarantee robust and reliable results. Finding the optimal balance between data granularity levels and insights robustness is a key element to fix constantly in our minds.

Finally, integrating a marketing budget optimizer with the ability to apply custom restraints that suit the business can provide great recommendations about how to switch budgets in your media mix for the upcoming cycles. Again: an additional tool in the direction of a **full actionability of the insights**.

Actionable Insights

Why actionable insights?

— The biggest challenge is finding a **marketing budget optimizer** facilitating decision-making, with the activation of media, platforms, objectives, etc. The workhorse for optimization is a **scenario simulator**:

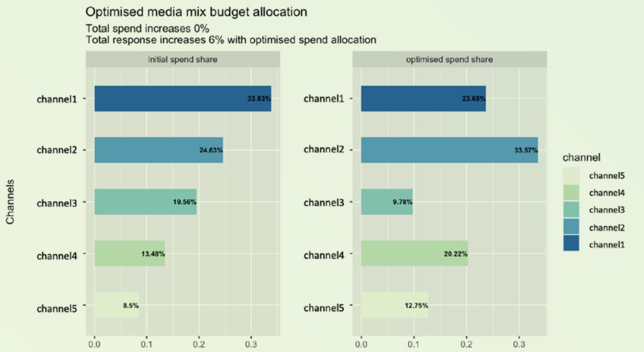
Inputs: a time period, the actual media investment for that period, and a proposed budget for each media.

Processing: the fitted models are run using the input data to obtain the forecasted contributions from each media with the new budget.

Output: forecasted contributions and ROI for each media.

— Later, the optimization uses the **scenario simulator as a cost function**. Using **Derivative-Free optimization algorithms** (do not require gradient information). Concretely, an implementation of the **Nelder-Mead** algorithm for derivative-free optimization.

— More importantly, they can be used to solve non-smooth optimization problems, and, on the other hand, they can also **handle box constraints on parameters**.



Finding balance on data granularity levels

The ideal would be to have as much breakdown as possible, to use the optimizer and have the entire budget placed on media, campaign objective, format, platform... but the **parameters to be measured dimensionality increases very quickly.** In addition, these new techniques as Bayesian approach, ad-stock, hill curves, ridge regression **imply an even greater number of parameters for each of these variables**, triggering remarkably the number of parameters to estimate. **There is a need to seek balance because we could find limitations:**

— The data **quantity available to the modeler is often limited**: the explanatory variables and the available observations. For example, a rule-of-thumb for a minimum number of data points for a stable linear regression are 7-10 data points per parameter. Otherwise, we can fall into **overfitting the model**, matrices close to the singularity.

— It is necessary to deal with a high dimensionality parameter space, where the greater the number of parameters and the greater the complexity of the model imply greater **computational requirements and an increase in training time**. Although, to achieve computational efficiency while maintaining the overall accuracy of the model, there are techniques and algorithms that can help.

— In the case of the Bayesian approach, we would need to experiment and test the different a priori conditions to be able to coherently fit the complex model.

In the era of big data and big computation these issues will be progressively addressed taking advantage of the always growing capacities to process as many data, dimensions, computing time and priors combinations as possible, without opposing the good business sense of the analysts sifting through the results thus obtained.

Conclusions

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.) **and unpredictable in this New Normal** during Covid pandemic.

In this context MMM presents the highest level of **flexibility and adaptability** to these changing dynamics, able to absorb new externalities and unexpected events and to recalibrate constantly its approach, in a sort of "Darwinian evolution" of media effectiveness measurement.

What are the advantages of using MMM? (1) It allows business to grow by transforming marketing practices grounded in data and science; (2) it helps all companies access advanced, privacy safe, efficient marketing effectiveness analytics in the context of evolving digital consumer behaviours and the changing digital marketing ecosystem; (3) it is transparent, offering customisable models that democratise econometrics and solve for analyst bias and subjectivity.

The future is modeled, and possibly "Marketing Mix" modeled.

Facebook's Authors

Leonel Sentana - Marketing Science Partner

Deloitte Analytics Authors

Davide Fabrizio – Director

Miguel Conde – Senior Specialist

María Eugenia Gómez – Senior Consultant

Carlos Real – Senior Manager

[1] Robyn is a Marketing Mix Modeling (MMM) code. It can be used to build end-to-end time series regression models and as an econometrics code library. Automated, built for very large data sets, and is suitable for digital and complex consumer behavior.

[2] Nevergrad is an open-source gradient-free optimization platform which efficiently optimizes the exploration over possible model solutions, reaching to optimum results faster and in a more precise way than traditional algorithms.

Deloitte.

FACEBOOK