What is MMM?

MMM is an econometric model that aims to quantify the incremental impact of marketing and non-marketing activities on a pre-defined KPI (like sales or website visits). This is a holistic model used to understand how to allocate a marketing budget across marketing channels, products and regions and can help forecast the impact of future events or campaigns.

When running an MMM study, there are multiple steps involved. See below for the steps and approximate timings - note that these are typical timings for when you run the first study and that this can be shortened for refresh models and/or the method of implementation.

Step	Description	Approximate Timings
Define business questions and scope	As with any research or analysis, it's important to define the objective and key business questions that you want to be answered by the study. This will help shape how the MMM will be designed and executed, which in turn will help you get the best value out of an MMM.	1-2 weeks
Data collection	Data collection is one of the most important but time consuming steps in an MMM study. Time-series data for sales (or another KPI), media, non-media marketing (e.g. promotions) and macroeconomic factors will need to be collected, cleaned and processed for MMM use. Collecting accurate and quality data is critical - if inaccurate or poor quality data inputs are used in an MMM, it will produce inaccurate and poor quality data outputs.	4-6 weeks
Data review	In addition to data collection, data review is one of the most important steps in an MMM study. This involves thoroughly checking all the cleaned and processed data sets for its accuracy, just before modeling begins. Data review may also uncover data gaps or errors, which may require more data collection. As mentioned above, the repercussions of inaccurate or poor quality data inputs means inaccurate and poor quality data outputs.	1-2 weeks
Modeling	Once happy with the accuracy of all data inputs, the next step is to run the models. Expect modeling to be an iterative process - that is, consistently re-running and fine-tuning the models. The time and effort of this step will significantly vary based on the scope and complexity.	4-8 weeks (dependent on MMM scope)
Analysis and recommendations	Once happy with the model, the next step is to analyse the data outputs and results and make actionable recommendations based on the original business questions in scope. There are a wide range of different outputs and metrics that come out of an MMM.	2-4 weeks

Data Collection

Dependent vs. Independent Variables:

In terms of what data sets to collect, it's important to distinguish between and determine the dependent and independent variables to be measured in the MMM:

- **Dependent variable:** This will be the primary KPI / metric that the MMM will measure against. The most commonly used data for the dependent variable is sales, although different data can be used depending on the vertical (e.g. account sign-ups for a telco business, home loan applications for a bank).

- **Independent variables:** These will be the variables or factors to be included in the MMM that should have an impact on the dependent variable. The most commonly used data sets for independent variables include media, non-media marketing (e.g. promotions, discounts), seasonality (e.g. weather, holidays) and macroeconomic factors (e.g. economic growth).

Media activity:

Data collected for media ideally should reflect how many “eyeballs” have seen or been exposed to the media (e.g. impressions, GRPs). Spends should also be collected in order to calculate Return On Investment, however it is best practice to use exposure metrics as direct inputs into the model, as this is a better representation than spends of how media activity has been consumed by consumers.

- For digital activity, the most commonly used metrics are impressions. Avoid using clicks, as clicks do not account for view through conversions, and it is just as likely that someone can view an ad and convert.
- For TV and radio, the most commonly used metrics are Gross Rating Points (GRPs) or Target Audience Rating Points (TARPs).
- For print (e.g. newspapers or magazines), the most commonly used metrics are readership.
- As mentioned above, aim to collect data that reflects “eyeballs” or impressions for all other channels.

Paid and organic variables:

Historically it was best practice to collect and model data for paid activity only, as this would produce more actionable results (given that it is difficult to control organic activity). However with more options to interact with consumers with organic content, there could be good reasons to include it in the model

- Non-media marketing activity (e.g. promotions, discounts): Commonly used metrics include time-series pricing data or the use of dummy variables to indicate promotions. If it is required to measure each type of promotion separately (e.g. buy one get one free vs. gift with purchase), ensure separate dummy variables are created for each promotion.
- Seasonality and holidays: External factors that have a big impact on the dependent variable, such as seasonality and holidays, should be included in the model.
- Throughout the seasons (e.g. summer vs. winter), temperature is commonly used. For key periods or specific events (e.g. Christmas, government policy change that impacts sales), dummy variables are commonly used. Similar to above, ensure separate dummy variables are created if it is required to measure each specific event separately.
- Macroeconomic factors: While this will depend on how these factors impact your business, commonly used metrics and data sets include GDP growth, unemployment, inflation.

Variable Type	Bucket	Variable
Dependent variable	Sales	Sales
Independent variables	Promotions	Price Discounts
		Bundle Offers
	Media - Online	Facebook
		Google Search
		YouTube
		Display
	Media - Offline	TV
		Radio
		Print
		OOH
	Competitor activity	Competitor media activity
	Macroeconomic Factors	Seasonality / Holidays
		Economic Growth
		Government Policy Changes
		COVID-19 Impact

Data Review

Once all data has been collected, a data review should be conducted to review it for its accuracy and whether it aligns with the media plan, expectations and if it can answer the desired business questions.

Along with data collection, data review is one of the most critical stages in an MMM study. If the collected data is not reviewed, ends up being inaccurate and is used in the MMM, then the data outputs will also be inaccurate and not reliable.

Reviewing many different time-series data sets for its accuracy can be a difficult and time-consuming process. There is no right or wrong way to run a data review - checking the data very thoroughly at the most granular level can give the most confidence, but it can be the most time consuming. That said, checking the data at a topline level can be very quick, but it might not establish the most confidence.

Ultimately at the end of the data review process, we want to feel confident that all the collected data is clean and accurate and so the data outputs can be trusted as accurate.

Modeling Phase

Feature Engineering

Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data (more details on feature engineering here). Today's MMMs are more sophisticated and introduce different feature engineering techniques. Feature engineering can be as critical as data collection, as the process transforms the way raw data is informed in modeling. Therefore it is very important to understand the underlying assumptions in each technique, otherwise the result can be misleading.

Prophet Seasonality Decomposition:

Depending on the vertical, the dependent variable (e.g. sales, conversions) will likely be impacted by underlying seasonality trends. Given that MMM uses time-series data, a time-series analysis can be used to determine the seasonality trends which in turn can be incorporated in the final model.

Prophet is a Meta open source code for forecasting time series data. Prophet has automatically been included to decompose the data into trend, seasonality, holiday and weekday impacts, in order to improve the model fit and ability to forecast. Traditionally it would be required to collect and model seasonality and holiday data as separate dummy variables in the model. However, Prophet makes this process much easier, especially for those new to Robyn and MMM.

- Trend: Long-term and slowly evolving movement (either increasing or decreasing direction) over time. In marketing, the trend can be generated by factors which can create long-term momentum (e.g. general market growth in specific categories, gradual economic downturn, etc.)
- Seasonality: The repeating behavior that can be captured in a short-term cycle, usually yearly. For example - depending on the vertical or category, specific brands can have more sales in summer than winter.
- Weekday: The repeating behavior that can be captured within a week. Note that this is only usable when daily data is available.
- Holiday/Event: Holidays or other events that highly impact your dependent variable. (e.g. national holiday, mega sales day, etc.)

If you do not already have trend/seasonality data of your own, we would recommend you consider using Prophet for at least the trend and seasonality components. However as the complexity of the model and industry being measured increases, it may be worthwhile exploring additional ways to account for time-based trends e.g. collecting other data sources.

Model Design

Designing a model is a balance between answering all the specified business questions and correctly specifying the model. Model overspecification occurs when too many independent variables have been included in a model and the model produces one or more redundant predictor variables. That is, the model has difficulty in accurately calculating the coefficient or impact of one or more independent variables on the dependent variable.

Categorize variables into Paid Media, Organic and Context variables: There are three types of input variables in Robyn i.e. paid media, organic and context variables. Check below for general considerations on how to categorise each variable into these three buckets:

- **Paid Media:** Any media variables with a clear marketing spend falls into this category. Transformation techniques will be applied to paid media variables to reflect carryover effects (i.e. adstock) and saturation (see the Data Transformation Techniques section for more details). For these variables, it is recommended to use metrics that better reflect media exposures such as impressions, clicks or GRPs instead of spend. If not available, spends can be used as a last resort.
- **Organic Variable:** Any marketing activities without a clear marketing spend fall into this category. Typically this may include newsletters, push notifications, social media posts, etc. As organic variables are expected to have similar carryover (adstock) and saturating behavior as paid media variables, similar transformation techniques will be also applied to organic variables.

- **Contextual Variable:** These include other variables that are not paid or organic media that can help explain the dependent variable. The most common examples of context variables include competitor activity, price & promotional activity, macroeconomic factors like unemployment rate, etc. These variables will not undergo any transformation techniques and are expected to have a direct impact on dependent variables.

Data Transformation Techniques

Part of MMM's appeal is that it is grounded in key marketing principles, such as adstock and saturation, where these principles are further reflected in Robyn:

- **Adstock:** This technique is very useful for a better and more accurate representation of the real carryover effect of marketing campaigns. Moreover, it helps us understand the decay effects and how this can be used in campaign planning. Adstock reflects the theory that the effects of advertising can lag and decay following an initial exposure. In other words, not all effects of advertising are felt immediately - memory builds and people sometimes delay action until following weeks, where this awareness diminishes over time.
- **Saturation:** The theory of saturation entails that each additional unit of advertising exposure increases the response, but at a declining rate. This is a key marketing principle that is reflected in MMM and Robyn as a variable transformation.

Adstock

There are two adstock techniques you may choose from in Robyn, each with its pros and cons. In order to find the approach that best fits your model objectives and business purposes, we recommend testing various transformations.

Geometric

- The biggest advantage of the Geometric transformation is its simplicity. It only requires one parameter called 'theta' that can be quite intuitive. For example, an ad-stock of theta = 0.75 means that 75% of the impressions in period 1 were carried over to period 2. This can make it much easier to communicate results to non-technical stakeholders. In addition, Geometric is much faster to run than Weibull, which has two parameters to optimize.
- However, Geometric can be considered as too simple and often not suitable for digital media transformations, as shown in this study. When it comes to setting hyperparameters for a Geometric transformation technique, theta is the only parameter that can be adjusted, which reflects the fixed decay rate. For example, assuming TV spend on day 1 is 100€ and theta = 0.7, then day 2 has 100x0.7=70€ worth of effect carried-over from day 1, day 3 has 70x0.7=49€ from day 2 etc. A general rule-of-thumb for common media channels are:

    - TV = c(0.3, 0.8)
    - OOH/Print/Radio = c(0.1, 0.4)
    - Digital = c(0, 0.3)

Weibull

- While the traditional exponential adstock is very popular, it was recently reported by Ekimetrics & Annalect that the Weibull survival function / Weibull distribution can better fit modern media activity such as Facebook. The Weibull survival function / Weibull distribution provides significantly more flexibility in the shape and scale of the distribution. However, Weibull can take more time to run than Geometric, as it optimizes two parameters (i.e. shape and scale) and it can often be difficult to explain to non-technical stakeholders without charting.

- When it comes to setting hyperparameters for the Weibull transformation technique, this will depend on which type of Weibull transformation
- Weibull CDF adstock: The Cumulative Distribution Function of Weibull has two parameters - shape & scale. It also has a flexible decay rate, whereas Geometric adstock assumes a fixed decay rate.
- The shape parameter controls the shape of the decay curve, where the recommended bound is c(0.0001, 2). Note that the larger the shape, the more S-shape and the smaller the shape, the more L-shape.
- The scale parameter controls the inflexion point of the decay curve. We recommend a very conservative bound of c(0, 0.1), because scale can significantly increase the adstock's half-life.
- Weibull PDF adstock: The Probability Density Function of the Weibull technique also has two parameters in shape & scale, and also has a flexible decay rate as Weibull CDF. The difference to Weibull CDF is that Weibull PDF offers lagged effects.
- For the shape parameter:
    - When shape > 2, the curve peaks after x = 0 and has NULL slope at x = 0, enabling lagged effect and sharper increase and decrease of adstock, while the scale parameter indicates the limit of the relative position of the peak at x axis;
    - When 1 < shape < 2, the curve peaks after x = 0 and has infinite positive slope at x = 0, enabling lagged effect and slower increase and decrease of adstock, while scale has the same effect as above;
    - When shape = 1, the curve peaks at x = 0 and reduces to exponential decay, while scale controls the inflexion point;
    - When 0 < shape < 1, the curve peaks at x = 0 and has increasing decay, while scale controls the inflexion point.
    - While all possible shapes are relevant, we recommend c(0.0001, 10) as bounds for shape. When only strong lagged effects are of interest, we recommend c(2.0001, 10) as bound for shape.
- When it comes to scale, we recommend a conservative bound of c(0, 0.1) for scale.
- Due to the great flexibility of Weibull PDF and more freedom in hyperparameter spaces for Nevergrad to explore, it also requires a large number of iterations for modeling.

Saturation

Robyn utilizes the Hill function to reflect the saturation of each media channel. A Hill function is a two-parametric function in Robyn with alpha and gamma:

- Alpha controls the shape of the curve between exponential and s-shape. We recommend a bound of c(0.5, 3) - note that the larger the alpha, the more S-shape and the smaller the alpha, the more C-shape.
- Gamma controls the inflexion point. We recommend a bound of c(0.3, 1) - note that the larger the gamma, the later the inflection point in the response curve.

Modeling Techniques

Once you finish setting the parameters in feature engineering, it is now time to run your first model. Fortunately once parameters have been set, Robyn's modeling process is automated and the results will be automatically generated according to the model specification you have made. Therefore, less intervention is required in this part, however it is important to understand what is happening behind the scenes and how to interpret the results.

Ridge Regression

MMM uses regression modeling, which aims to derive an equation that describes the dependent variable. The model aims to assign a coefficient to each independent variable, where only the variables that are statistically significant stay in the model.

In very simple terms, the following model shows how the KPI is affected by changes in all the factors you have data for (you can find a more detailed equation for the Robyn model here):

$KPIt = \beta 0 + \beta 1x1 + \beta 2x2 + \beta 3x3 + \beta 4x4 + \ldots + \beta nxn$

KPIt : The KPI at the time (t) you want to model.

β0: The base performance, or what performance would be if all other factors were at their minimum.

β: The coefficients, or what a change in the variable (x) means for the KPI.

Overfitting and multicollinearity are commonly addressed issues in regression analysis. Overfitting the model will take away from it's predictive powers as you are taking away the flexibility with the model. On the other hand, not having enough variables and underfitting can lead to improper reads into certain channels. In order to address multicollinearity among many regressors and prevent overfitting we apply a regularization technique to reduce variance at the cost of introducing some bias. This approach tends to improve the predictive performance of MMMs.

The most common regularization and the one used in Robyn is Ridge regression - for more details on the mathematics behind it, see this page. In Robyn, Ridge regression is automatically built in and there are fortunately not many settings that you need to configure - however, if keen to configure further, refer to the tip below.

Regularization technique in Ridge regression is achieved by penalizing the regression model. This means that Ridge regression will shrink coefficients towards zero if those variables have a minor contribution to the dependent variable.

Model Selection

Model selection is a core part of the modeling process in Robyn. Building MMMs manually can be a very time-consuming process because MMMs are likely to contain a high cardinality of parameters to adjust (please refer to the Data Transformation Technique section part in the Feature Engineering section). Adjusting different hyperparameters can involve many subjective decisions, modeling experience and trial and error over hundreds of iterations, where it can take months to build an MMM from scratch. Fortunately, Robyn is able to automate a large portion of the modeling process and thus reduce the time taken to run a model as well as the "analyst-bias".

Robyn provides a semi-automated model selection process by automatically returning a set of optimal results. To accomplish this, Robyn leverages the multi-objective optimization capacity of Meta's evolutionary optimization platform Nevergrad.

Applications of the Model

One of the benefits of MMM and Robyn is that it can measure the effectiveness of all your advertising and also provide actionable insights to improve the effectiveness of various marketing activities. To achieve this, you must be able to correctly interpret the outputs of the model.

Interpret Marketing Mix Model Outputs

Model fit:

- In order to have an accurate model, the fit of the modelled data has to be accurately relative to the actual data provided. As mentioned in the Model Design section, overspecifying the model will erode its predictive powers, however not having enough variables and underspecifying can lead to improper reads into certain channels.
- Models with a low R squared values can likely be improved upon. Common ways to improve it include having a more comprehensive set of independent variables - that is, split up larger paid media channels or include additional baseline (non-media) variables that may explain the dependent variable.

Volume Contribution: In addition to being able to compare actual vs. modeled data, we can also see the incremental volume contributions from each modelled independent variable.

Response curve: Sometimes called saturation curves, response curves indicate if a specific media channel's spend is at an optimal level or if it is approaching saturation and therefore suggest potential budget reallocation. The faster the curves reach an inflection point and to a horizontal/flat slope, the quicker the media channel will saturate with each extra ($) spent.

Effect Share: It gives information about the percentage share of media spend variables responsible for generating conversions or revenue (here only the target generated through media activity will be considered).