MARKETING MIX MODELING

WHY MODELING?

Many companies spend a lot of money in various advertising channels but also in other sales related activities e.g. special promotions and offers or trainings of their sales force. So it is only natural, they want to find out which activities have the greatest effect on sales. To assess and differentiate between the effects of various activities, statistical methods have to be applied.

In general, there are two approaches to find out which channel or activity has the greatest impact:

1. If there is data on individual level for both contacts with advertisements and whether this person has bought the product or service or not, then the attribution modeling approach can be used. This is often the case with e-commerce or other online-focused sales activities.

2. If the data is only available on aggregated level, i.e. media spendings per week or units sold per week in a region, then the more general marketing mix modeling approach should be first choice.

This white-paper focuses on the latter approach as this is the more common approach in retail, automotive or IT.
Regression is still widely used

In general, the marketing mix modeling approach is based on finding the statistical model (in many cases, but not exclusively linear regression or general linear models) that best captures the channels' and activities' influence as independent variables on the sales figures as dependent variable while also taking into account activities or advertising spend of competitors and even external influences (weather, holiday seasons, general economic or political climate) either as numerical or dummy variables.

Main objectives of a marketing mix model

The main objectives of a marketing mix modeling project are to find out:

- How much do the various activities contribute to the overall sales (i.e. a decomposition of total sales into the channels)?
- What is the ROI of these activities (i.e. the return on investment per channel)?
- What is the optimal budget allocation to the channels?
- How can effects of decreasing efficiency be avoided?

In addition, the models also allow to simulate various scenarios consisting of increasing, decreasing or removing activities.

DATA HARMONIZATION AND CONSOLIDATION

Usually, a marketing mix modeling does not start with one clean dataset, but with a number of datasets in very diverse
formats. So, all the different data sources have to be harmonized first. Only after this very time-consuming step, they can be integrated into one dataset that can be used to calculate the model.

This step often involves creating import routines to correctly open and read the files (CSV, ASCII, SPSS etc.) into a statistical software.

**Data transformation can take some time**

Then, the datasets have to be transformed so that they all refer to a time variable on the same level (day or calendar week).

After this, the variables (e.g. amount of money spent on online and TV advertising, price promotions or special activities) have to be recoded so that they can be used as features (independent variables) in the model selection process.

**DATA VALIDATING AND CLEANING**

In many cases, there will be (a lot of) missing values, errors and implausible values in the raw data. This step will check the datasets and assess possible flaws in the data.

The client often receives a short summary of the data quality and cleaning procedures that will be applied to the data before calculating the models.

**MODEL SELECTION AND FEATURE SELECTION**

After the data has been transformed, integrated and cleaned, the next step is the selection of the best model for this
question. As this process is quite unique to each project and depends on the data and more importantly on the subject matter, only a high-level description can be given in this white-paper.

In general, we put an emphasis on model simplicity because a “simple” multivariate linear (OLS) regression model is in many cases much more robust than a highly multidimensional, nonlinear model.

**Precision versus explainability**

But even more important, these kinds of models can be described and understood by subject matter experts, i.e. you, and therefore its validity can also be assessed by your own experience with the topic.

The model will be constructed to take into account:

- effects of marketing and sales activities of the client and media agencies but also variables like price levels, special promotions or stocking measures (highly dependent on the market segment)
- decay curves for these activities (e.g. decay or ad-stock effects)
- external influences not under control of the company (e.g. weather, competitors’ activities, industry trends)
- repeating seasonal trends (e.g. national holidays)
- long term drifts and trends

In many cases, the result is a multivariate linear / log-linear model that can also include interaction terms for variables influencing each other.
At least two or three years of data should be available

To assess seasonal trends and variations, the data should at least contain two or three full years. The longer the timeline of the datasets, the better the ability to test the model with historical data and avoid overfitting.

Often, the client will be provided not only with one, but a few alternative models, because the model with the best fit and lowest cross-validation error cannot always be interpreted very easily.

So, sometimes, we also discuss alternative models with the client that may sacrifice a small degree of precision in order to be much simpler and more intuitively understood.

MODEL EVALUATION AND DOCUMENTATION

After the model and features are defined and optimized, the model results will be calculated and often documented in an Excel spreadsheet and a PowerPoint presentation.

In our reports, we usually include detailed information about the methods used and the model quality – with statistical
quality measures and by comparing the predicted sales results with the observed result. We believe that modeling does not have to be a black box for the clients.

**SIMULATION**

Additionally, in many cases, we create a web-based interactive simulation based on the model. This allows the client to change various input parameters (e.g. reducing, increasing or stopping certain marketing and sales activities) and to explore the resulting predicted sales numbers in different scenarios.
ABOUT DATALION

DataLion is a web-based software to analyze and visualize enterprise, media and consumer data. More than 400 companies in Europe are using DataLion to analyze survey data, create interactive dashboards and reports or automate their analytics processes.

Register now your free DataLion account: https://datalion.com/try

DataLion GmbH
Herzog-Wilhelm-Straße 1
80331 Munich
Germany

Tel.: +49 89 716 772 100