About Us

Opella, the Consumer Healthcare business unit of Sanofi, is the purest and third-largest player globally in the Over-The-Counter (OTC) & Vitamins, Minerals & Supplements (VMS) market.

We have an unshakable belief in the power of self-care and the role it can play in creating a healthier society and a healthier planet. That's why we want to make self-care as simple as it should be by being consumer-led always, with science at our core.

Inside Opella, the Advanced Analytics team is focused on improving business processes using Data-driven Strategies, Mathematical Modelling, Machine Learning and Artificial Intelligence solutions.

Among all the current projects, the main focal points can be classified among Econometric and Optimization Model, Forecasting and GenAI initiatives.

The issues outlined in this document pertain to a project that falls within the realm of Econometric Modelling, specifically Marketing Mix Modeling.

Introduction

Marketing Mix Modelling is an econometric exercise that leverages historical data to quantify the impact of various marketing tactics on specific business KPIs, such as sales. The primary objective of these projects is to optimize the advertising mix and promotional strategies to achieve goals ranging from maximizing profit or gross margin to minimizing total expenditure.

Typically, marketing mix analyses are conducted using linear regression models, with nonlinear effects accounted for by transforming the independent variables through functions that capture diminishing returns on investment.

While Frequentist approaches have traditionally dominated this field, the landscape of marketing analytics has been significantly reshaped by the growing importance of Bayesian methods. By adopting a probabilistic approach to manage uncertainty in the inherently noisy environment of advertising, our team can incorporate domain expertise into the modelling process through Bayesian prior beliefs.

However, employing a Bayesian approach does not eliminate the challenges such as "Problem 1" and introduces its own set of drawbacks, such as "Problem 2."

Problem 1: Isolating variable effects under multicollinearity circumstances using Bayesian Regression

Description

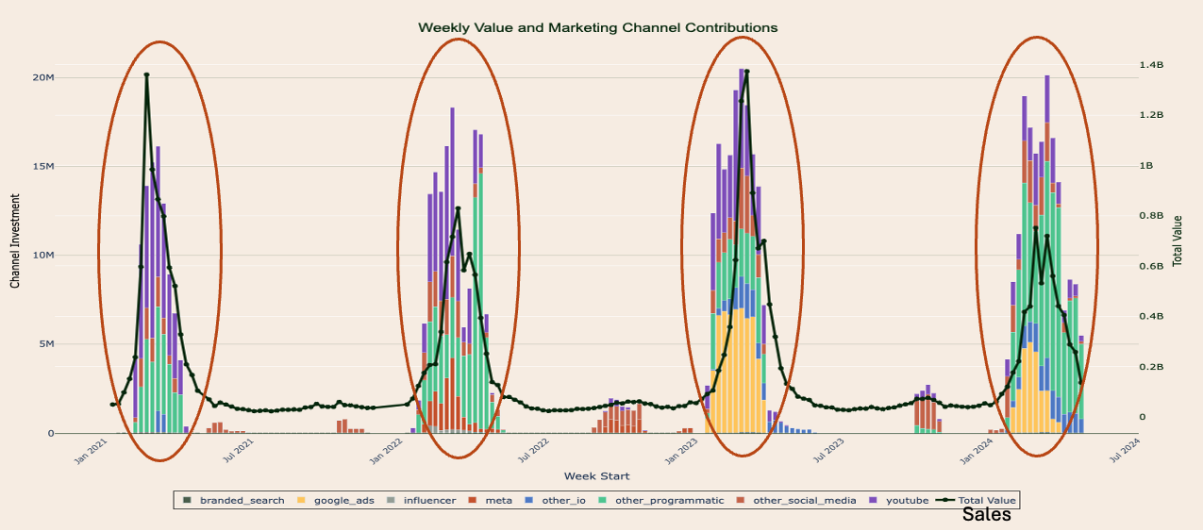
As with any Marketing Mix Modelling (MMM) project, several of our models encounter multicollinearity issues.

These issues primarily arise from variables dictated by the business rationale of our project, such as media channel investments. For instance, in seasonal brands (and sometimes even in non-seasonal ones), marketing departments tend to increase their investments across all channels during peak seasons and reduce them during off-seasons. This synchronized investment pattern results in highly similar trends, exacerbating multicollinearity problems.

Given that the impact of the estimated effect should remain minimally manipulated due to the nature of the project, we face several constraints:

- Cannot unify variables to estimate a single coefficient.
- Cannot apply dimensionality reduction techniques such as Principal Component Analysis (PCA) or Singular Value Decomposition (SVD).
- Cannot eliminate any variable from the model adjustment.
- Prefer to avoid regularization techniques that penalize the value of these coefficients.

These constraints necessitate alternative strategies to address multicollinearity without compromising the integrity of the estimated effects.



Assumptions

- Variable aggregation is not feasible because we aim to quantify the causal relationships between relevant business investments and total sales.
- Dimensionality reduction techniques, such as Principal Component Analysis (PCA) or Singular Value Decomposition (SVD), are unsuitable since we need to maintain the model's explainability, as this is an inferential rather than a predictive exercise.
- Variable elimination is not an option due to the business relevance of each variable.
- The Linear Regression model is fitted using a Markov Chain Monte Carlo (MCMC) method, specifically a Hamiltonian Monte Carlo (HMC) algorithm that employs the No-U-Turn Sampler (NUTS) technique.

Expected Results

The ideal outcome would be to develop a method for isolating the effects of each variable within a Bayesian framework (if possible although we are open to Frequentist proposals), while avoiding regularization techniques based on Normal or Laplacian priors.

Additionally, the solution should be seamlessly integrated into the model fitting process, rather than as a separate or preliminary step, to maintain the efficiency of the workflow.

Problem 2: Obtaining informative posteriors in experiments with few data points

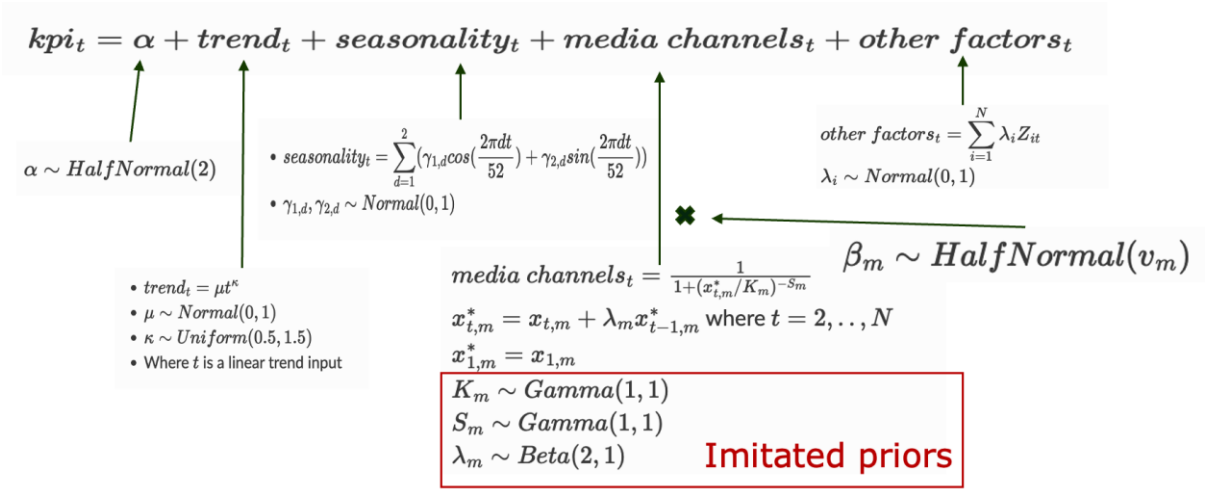
Description

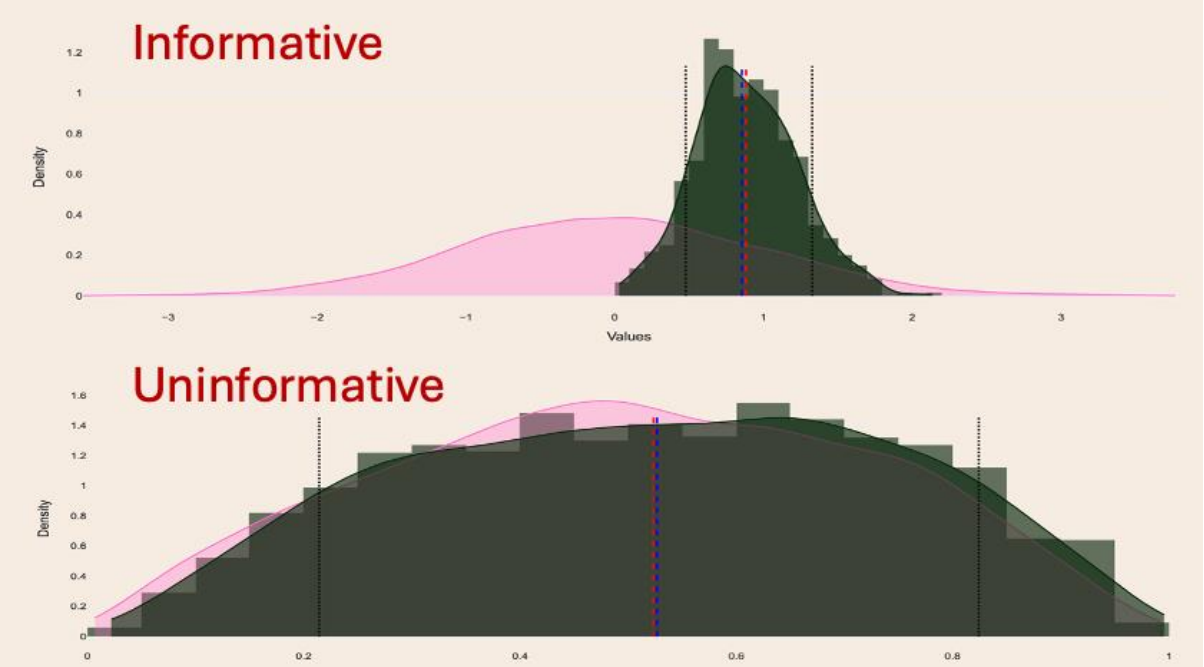
Given the constraints of limited data availability, our team aims to fit models with a small number of data points. This challenge is particularly relevant in the context of Marketing Mix Modeling (MMM), where understanding the impact of various marketing channels is crucial. A Bayesian approach is being used for dealing with these limitations by incorporating prior knowledge and probabilistic reasoning.

However, a probabilistic approach necessitates defining priors for a wide range of parameters, resulting in a parameter space that is vast relative to the amount of data available. This poses significant challenges for the Hamiltonian Monte Carlo (HMC) algorithm, which must navigate this extensive parameter space.

This imbalance results in a fitting process characterized by the following:

- The posteriors for the regression coefficients are updated based on the observations from the MCMC runs.
- However, the posteriors for the indirect parameters (red squared) remain non-informative. They tend to explore the entire allowed range, effectively behaving like uninformative priors.





Assumptions

The Linear Regression model is fitted using a Markov Chain Monte Carlo (MCMC) method, specifically a Hamiltonian Monte Carlo (HMC) algorithm that employs the No-U-Turn Sampler (NUTS) technique.

Expected Results

The ideal outcome of this proposal is to develop a method for estimating informative posteriors for parameters that are indirect in our regressions. Specifically, this refers to parameters of the nonlinear transformations rather than the regression coefficients themselves.