



POLITECNICO
MILANO 1863

Bayesian Marketing Mix Modeling

Time-Varying Non-Linear Models for Predicting Optical Fiber Sales

R. Montagna, Z. Peracchione, F. Pertusi, A. Riboni, F. Tomasi, D. Vozza

14/02/2025

What is MMM?

- A statistical technique used to measure the impact of marketing activities on sales.
- Helps identify which advertising channels generate the best ROI.

What is it used for?

- Optimizes marketing budget allocation.
- Improves sales forecasting.
- Supports strategic decisions based on data.

In this industry context the key challenges in budget allocation are:

- **Attribution difficulty:** Multiple advertising channels are used simultaneously.
- **Non-linear effects:** More advertising doesn't always mean more sales (saturation).
- **External factors:** Seasonality, competitor pricing, market trends.

Example: If a TV promotion boosts sales today, was it due to the ad itself or a competitor lowering prices?

Dataset Description

- 418 weekly observations (2017-2024).
- Includes sales data, marketing spend, pricing strategies, and competitor activities.
- Response variable: weekly fiber-optic offers sold.
- 89 variables across different categories:
 - ▶ **Marketing Exposure:** Digital Ads, TV campaigns, Audio Ads.
 - ▶ **Competitor Advertising:** Rival ad spending across channels.
 - ▶ **Pricing Information:** Own & competitor pricing.
 - ▶ **Consumer Engagement & Promotions.**

Steps Taken to Prepare the Data:

- **Variable Correlation Analysis**
- **Feature Aggregation**
- **Dummy Variables**
 - ▶ Christmas & Ferragosto (seasonality effects).
 - ▶ COVID impact (shift in consumer demand).
 - ▶ Promotional weeks (special offers).
- **Train-Test Splitting**
- **Range-Aware Scaling**
- **Variable Selection**

- Incorporates prior distributions to regularize estimates.
- Quantifies uncertainty via probability distributions.

Carryover Effect → AdStock with Geometric decay:

$$\text{AdStock}(x_t) = \frac{\sum_{l=0}^L \theta^l x_{t-l}}{\sum_{l=0}^L \theta^l} \quad \theta \stackrel{\text{iid}}{\sim} \text{Beta}(\mu = \text{theta_params}, \sigma = 0.05).$$

Saturation Effect → Logistic Function:

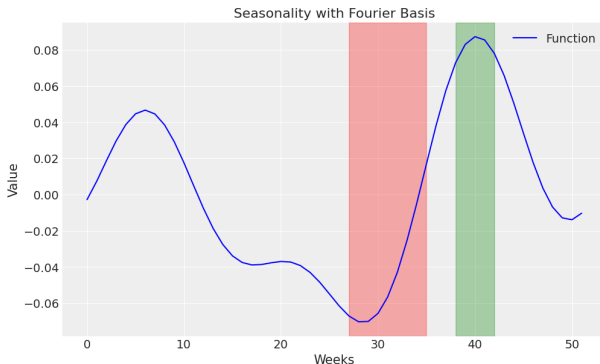
$$f(x) = \frac{1 - e^{-\lambda x}}{1 + e^{-\lambda x}} \quad \lambda \stackrel{\text{iid}}{\sim} \text{Gamma}(\mu = \text{lambda_params}, \sigma = 0.05).$$

The likelihood of the model is obtained by: $y_t | \mu_t, \sigma \stackrel{\text{ind}}{\sim} \mathcal{N}(\mu_t, \sigma)$

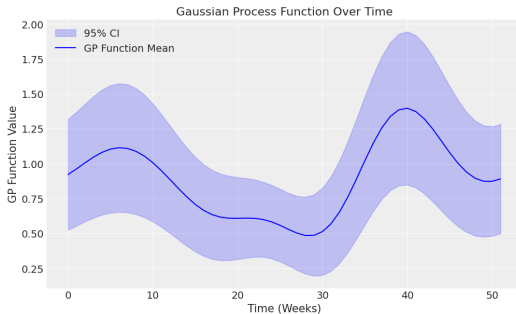
$$\begin{aligned} \mu_t = & \beta_0 + f_t \cdot \left(\sum_{i=1}^{n_{\text{channels}}} X_{\text{transformed},i,t} \cdot \beta_{\text{channel},i} \right) + \\ & + \sum_{i=1}^{n_{\text{price}}} (X_{\text{price},i,t} \cdot \beta_{\text{price},i}) - \sum_{i=1}^{n_{\text{competitors}}} (X_{\text{competitors},i,t} \cdot \beta_{\text{competitors},i}) + \\ & + \sum_{i=1}^{n_{\text{control}}} (X_{\text{control},i,t} \cdot \beta_{\text{control},i}) + \sum_{i=1}^{n_{\text{fourier}}} (X_{\text{fourier},i,t} \cdot \beta_{\text{fourier},i}) \end{aligned} \quad (1)$$

$$\sigma \sim \text{HalfNormal}(0.1).$$

$$f(t) = \sum_{k=1}^3 \left[a_k \cos\left(\frac{2\pi kt}{52}\right) + b_k \sin\left(\frac{2\pi kt}{52}\right) \right]$$



$$f_t \sim \mathcal{GP}(0, k(t, t')) \quad k(t, t') = \exp \left(- \frac{2 \sin^2 \left(\frac{\pi |t - t'|}{52} \right)}{\lambda^2} \right)$$



The prior distributions of the model parameters are the following:

- $\beta_0 \sim \text{HalfNormal}(\mu_{\text{intercept}}, 0.1)$
- $\beta_{\text{channel}} \stackrel{\text{iid}}{\sim} \text{HalfNormal}(0, 0.08).$
- $\beta_{\text{price}} \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_{\text{price}}, 0.1)$
- $\beta_{\text{competitors}} \stackrel{\text{iid}}{\sim} \text{HalfNormal}(0, 0.1).$
- $\beta_{\text{control}} \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_{\text{control}}, 0.1)$
- $\beta_{\text{fourier}} \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_{\text{fourier}}, 0.01)$

Three principal models: with Fourier Basis and/or Gaussian Process.

Common pipeline for each model:

1. Priors Identification:
 - 1.1 Correlation and Plot Analysis
 - 1.2 Non Bayesian OLS
2. Model Definition and Prior Predictive Check
3. Sampling and Convergence Analysis
4. Posterior Predictive Distribution
5. Rescaling
6. Contribution Analysis
7. Evaluation Metrics

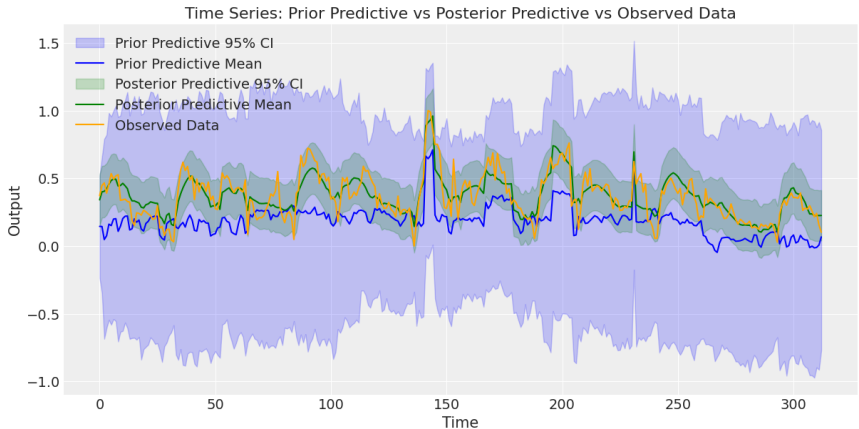


Figure: Prior vs Posterior Predictive Distribution - Model 2

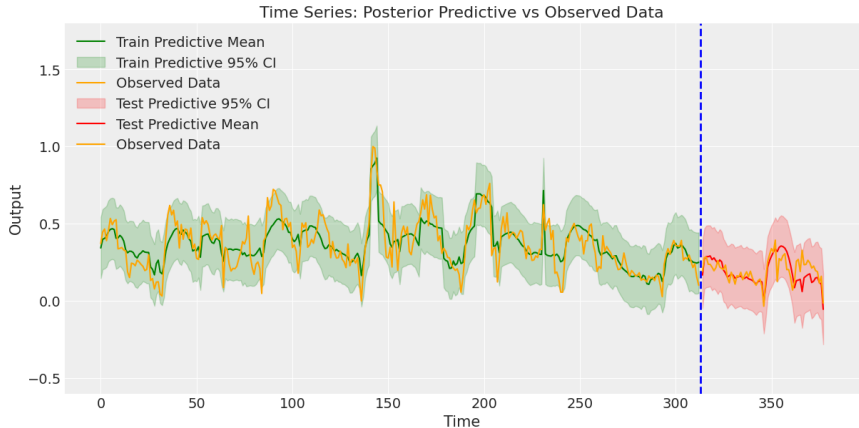


Figure: Posterior Predictive Distribution - Training vs Test - Model 3

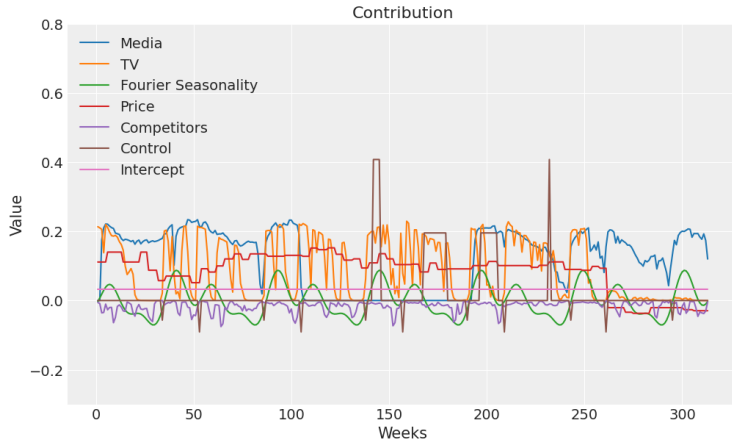


Figure: Contribution Analysis - Model 3

Model	Fourier Basis	Gaussian Process	MAPE
Model 1	Yes	No	13.001%
Model 2	No	Yes	13.347%
Model 3	Yes	Yes	11.747%

Table: Results of the final versions of the three model structures.

- The combination of both techniques gives the best performance.
- Good trade-off between desired accuracy and overfitting reduction.
- Needed more informative and restrictive priors.
- Performance reduction in the last timestamps.

- Variables impact on sales.
 - Bayesian framework advantages:
 - ▶ Uncertainty quantification.
 - ▶ Inclusion of experts knowledge.
 - ▶ Identification problem resolution.
 - Marketing explainability.
- Challenges in distinguishing the different effects.
 - Strategy and market changes.
 - Data availability.
 - Not available closed-form solutions for the posterior and marginal distributions.

- Explore alternative priors and model structures.
- Obtain the ideal threshold: 10%.
- ROI optimization, budget allocation.