POLITECNICO
MILANO 1863

Bayesian Marketing Mix Modeling

Bayesian Statistics

Montagna Rosita
Peracchione Zeno
Pertusi Federica
Riboni Alessandro
Tomasi Francesco
Vozza Daniele

Tutor: Emanuele Fedrigolli

https://github.com/Zenopera/MMM_Bayesian_Project

February 14, 2025

Contents

1 Introduction to Marketing Mix Modeling

1.1 Overview

Marketing Mix Modeling (MMM) is a statistical approach used to optimize marketing budget allocation across different media channels. Traditional heuristic-based methods often rely on arbitrary rules, such as setting marketing budgets as a fixed percentage of expected revenues. However, these methods introduce significant uncertainty and lack adaptability. Instead, **Bayesian MMM** provides a data-driven framework that estimates the impact of each marketing channel on key performance indicators (KPIs) such as sales and customer acquisition. This enables businesses to make informed decisions and adjust their strategies based on measurable outcomes.

1.2 Key Benefits of Marketing Mix Modeling

Marketing Mix Modeling (MMM), with their structured approach, provides businesses a way to optimizing marketing investments. One key advantage is its ability to measure the effectiveness of different media channels, allowing for data-driven budget allocation to maximize returns. Unlike traditional methods, MMM refines ROI forecasting, offering precise projections that guide strategic decisions.

Additionally, MMM uncovers synergies between channels, enabling more integrated marketing strategies while adapting dynamically to market changes. It also justifies marketing investments with concrete data, enhancing accountability and alignment with business objectives. By optimizing the timing and frequency of campaigns, running scenario analyses, and tailoring efforts to specific customer segments, MMM empowers companies to improve efficiency, boost ROI, and sustain long-term growth.

1.3 Historical Background

The origins of Marketing Mix Models trace back to the 1950s and have evolved significantly over time:

- **1950s-1960s:** Initial development of MMMs to quantify advertising impacts on sales.
- **1970s-1980s:** Increased adoption with the introduction of advanced statistical methods.
- **1990s:** Introduction of scanner data and loyalty programs for granular insights.
- **2000s:** The rise of digital marketing added complexity to MMMs.
- **2010s:** The integration of big data and machine learning improved modeling capabilities.
- **2020s:** Bayesian MMMs enable robust uncertainty modeling, lift-test calibration, and interpretable results.

2 Application to Fiber-Optic Telecommunications

2.1 Industry Context and Challenges

In recent years, **the fiber-optic telecommunications** sector has experienced exponential growth, driven by the increasing demand for high-speed connectivity and continuous technological advancements. In this highly competitive landscape, companies in the industry face increasingly complex challenges in maximizing the return on marketing investments, optimizing pricing strategies, and responding effectively to competitor actions.

Instead of relying solely on large amounts of data, we incorporate prior knowledge and marketing assumptions to develop a Bayesian framework. This approach helps mitigate overfitting and underfitting by integrating expert-driven priors, allowing the model to generalize effectively even with limited observations.

2.2 Problem Addressed

This report presents a **Marketing Mix Modeling (MMM)** project for a leading company in the fiber-optic telecommunications industry, aiming to predict sales revenue based on marketing expenditures across various channels and key indicators related to pricing and competitor activities. The challenge lies in the intrinsic complexity of these relationships, which are characterized by non-linear effects, temporal dynamics, and variable interactions. Traditional MMM models, based on linear regression techniques or more sophisticated variants, often suffer from significant limitations, including:

- **Assumption of linearity:** Many traditional models assume that the effect of independent variables on sales is linear, ignoring possible saturation thresholds or diminishing marginal returns.
- **Difficulty in handling uncertainty:** Classical frameworks often provide point estimates without adequate uncertainty measures, making it difficult to assess the robustness of strategic decisions.
- **Challenges in identifying effects:** Traditional models struggle to distinguish non-linear effects, where marketing response saturates or diminishes over time, from dynamic effects that result from seasonality or delayed impacts.
- **Lack of adaptability to market changes:** Marketing strategies evolve rapidly, and static models struggle to capture these changes in a timely manner.

2.3 Overcoming Limitations with the Developed Framework

To address these challenges, our approach leverages a **hierarchical Bayesian model**, which enables:

- **Quantifying and incorporating uncertainty:** The Bayesian approach allows results to be expressed as probability distributions, improving the understanding of risks associated with marketing decisions.
- **Integrating non-linear and dynamic effects:** Instead of treating them separately, we adopt a model that jointly considers these aspects, reducing the risk of identification issues and providing a more realistic picture of marketing strategy impacts.
- **Effectively handling missing data:** The Bayesian structure facilitates modeling uncertainties related to missing data, reducing the risk of bias in estimates.
- **Enhancing business interpretability:** The chosen approach ensures a balance between statistical robustness and practical applicability, providing valuable insights for defining marketing and pricing strategies.

Our starting point was the model proposed in [3], which adopts a more conservative approach in distinguishing non-linear and dynamic effects. In contrast, our framework integrates them, recognizing their key role in MMM for the telecommunications sector. This approach enables a more realistic and informative modeling process, directly impacting business decision-making.

2.4 Modeling Techniques and Data Requirements

The Bayesian MMM framework developed in this study incorporates advanced modeling techniques to better capture marketing effectiveness:

- **Nonlinear Effects:** The model captures diminishing returns on marketing investments through logistic saturation functions and accounts for carryover effects over time using decay functions such as geometric and Pascal decay.
- **Time-Varying Effects:** Seasonality, unexpected events, evolving marketing strategies and time-varying coefficients are modeled through Gaussian processes.

These techniques are described in detail in Section 5.

2.5 Implementation in PyMC Library

To facilitate implementation, we used **PyMC**, a Bayesian modeling library that provides:

- Priors customization to incorporate expert knowledge.
- MCMC-based inference for parameter estimation
- Support for hierarchical models and time series analysis.
- Inclusion of Gaussian Processes to model time-varying effects.
- Integration with JAX for accelerated computations.
- Parallel chain sampling for computational efficiency.
- Model validation and diagnostics through visualization techniques.

By combining these techniques, we ensure a robust and adaptive modeling approach suitable for the complexities of fiber-optic telecommunications marketing.

3 Dataset

3.1 Dataset Description

Our dataset consists of 418 weekly observations covering the period from January 1, 2017, to December 29, 2024. However, due to limitations in data collection, records are available only until March 17, 2024, leaving the last 38 weeks unfilled. The dataset presents a challenge of data scarcity, with a relatively limited number of observations for model fitting and a non-negligible amount of missing data.

The response variable represents the number of fiber-optic offers sold each week. The dataset includes 89 variables, covering multiple aspects of marketing activity, pricing strategies, and competitor influence. The key categories of variables are:

- **Temporal Features:** Data are structured as **Time Series** with weekly observations.
- **Marketing Exposure Metrics:** Contains data on digital advertising, including display ads, online video promotions, and search engine marketing. Audio advertisements from streaming platforms, introduced in recent years, are also included.
- **Television Advertising:** Covers TV campaigns related to fiber offers, general brand awareness, mobile services, and bundled packages. Some advertising categories were merged where their effects were interchangeable.
- **Competitor Advertising:** Includes data on advertising efforts from multiple competitors, with varying degrees of impact on fiber and mobile sales. Certain competitors have had stronger erosive effects in recent periods.
- **Pricing Information:** Tracks fiber pricing in physical stores and online platforms, along with competitor pricing across different channels to measure relative market positioning.
- **Additional Indicators:** Includes consumer engagement metrics, promotional activity, and aggregated advertising spending across multiple platforms.

3.2 Metodology

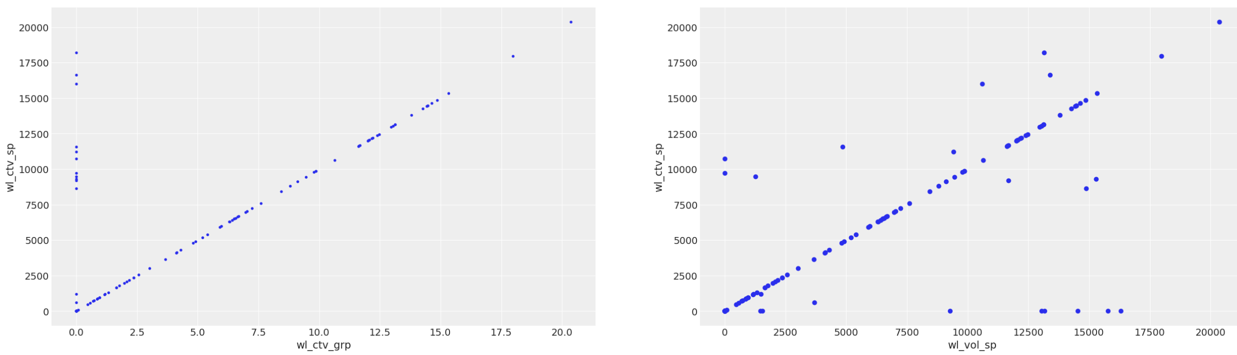
Given the challenges of data sparsity and differences in scale, several preprocessing steps were performed to refine the dataset and ensure the stability and interpretability of the Bayesian model. This phase of the analysis is explored in 4, and the key steps are:

- **Handling Missing and Zero Output Observations:** Observations where no sales were recorded were removed from the dataset.
- **Removing Redundant and Highly Correlated Variables:** Certain variables were found to be either perfectly correlated or strongly redundant. Variables that were completely dependent on others were removed, while highly correlated ones capturing similar effects were aggregated to reduce redundancy while preserving their overall impact.
- **Adjusting TV Advertising Variables:** To ensure consistency across different spending categories, some advertising categories were merged due to their interchangeable effects, and spending values in television advertising were rescaled to align with other marketing investments.
- **Price Variables and Competitors Influence:** To better capture competitive dynamics, relative price differences were computed instead of using absolute price levels. Aggregated indicators were also introduced to represent overall competitor advertising intensity across different categories.
- **Final Variable Selection:** This refined dataset allows the Bayesian model to effectively capture both structural and time-varying effects, reducing noise while retaining key marketing insights.

4 Preprocessing

4.1 Variables Correlation

Given the large number of variables relative to the number of observations, we observed high correlation (>0.85) among certain variables. To mitigate issues related to multicollinearity and to simplify the model, we reduced these highly correlated variables to a single representative variable, which was retained in the dataset.



(a) The variables are linearly dependent. `wl_ctv_sp` is a linear transformation of `wl_ctv_grp`.

(b) Strong collinearity (0.8794) between `wl_vol_sp` and `wl_ctv_sp`. We sum them into `wl_vol_ctv_sp`.

Figure 1: Examples of strong collinearity between advertising variables.

4.2 Variables Aggregation

With the guidance of our tutor, we aggregated similar advertising channels by summing their respective variables and we kept only the most relevant metrics. This process reduced the number of features while preserving relevant information for output forecasting. Variables selection was performed as described in Section 4.6; jointly with that, groups of similar variables were created, in order to perform the same transformations and the same inclusion in the model based on the variable type. Specifically we created new variables that represent for each week the maximal price difference between our client and the competitors both in store and online. Morevoer we considered the total tv spent of all our competitors adding their individual values on each advertising channel.

4.3 Dummy Variables

To develop the model's ability to capture seasonal effects and external events that influence sales, we introduced a set of dummy variables (binary variables with value 1 only in selected weeks or periods). These variables helped the model differentiate specific time periods that exhibit distinct behavioral patterns.

The dummy variables created include:

- **Christmas Week**: Identifies weeks corresponding to the Christmas holiday period, when consumer behavior and advertising impact are different from regular weeks; the result is a lower sales amount.
- **Ferragosto Week**: Captures the effect of the mid-August holiday, a period characterized by lower economic activity in Italy. In cases where Ferragosto fell at the end of the week

(Friday, Saturday, or Sunday), its impact was distributed across two weeks: the week including Ferragosto received a dummy value of 0.8, while the following week was assigned a value of 0.2. This adjustment aligns with the marketing concept of the *extended weekend effect*, which considers shifts in consumer behavior and economic activity due to prolonged holiday periods.

- **COVID Period**: Flags the weeks affected by COVID-19 restrictions, during which marketing effectiveness and consumer spending patterns were significantly increased due to the heightened demand for fiber networks driven by home confinement during quarantine.
- **Promotion Weeks**: Highlights weeks in which the company launched major promotional campaigns, helping to isolate the impact of price discounts and special offers on sales performance, which are associated with a significant increase in sales that cannot be explained by other factors.

This results in an additional set of control variables related to time, which helped in understanding seasonal and temporal variations:

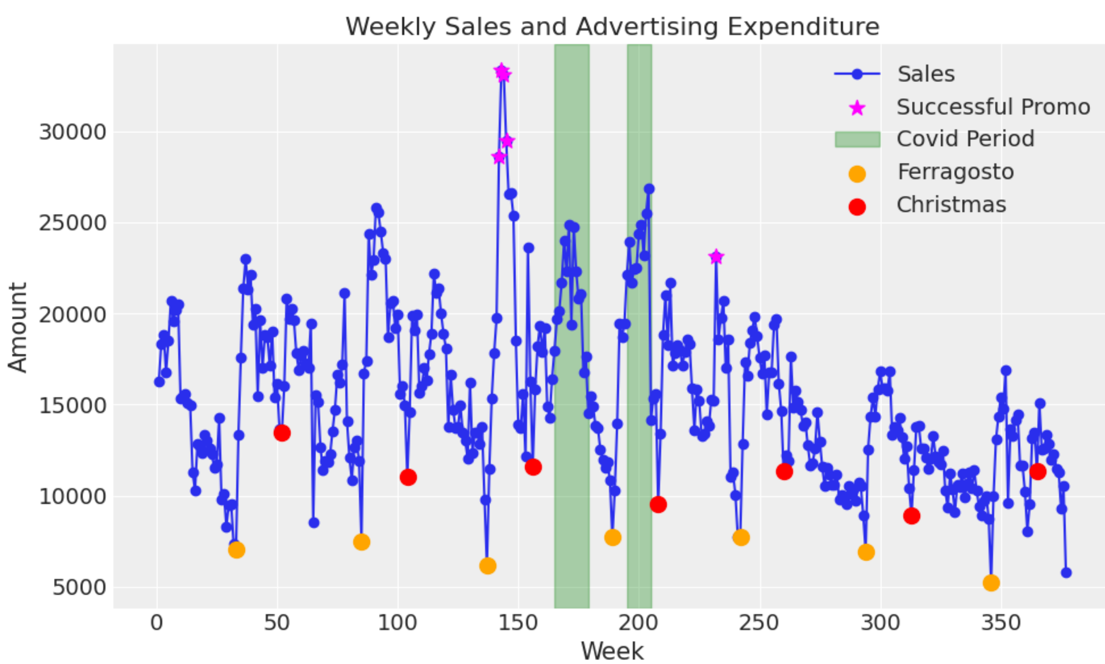


Figure 2: Visualization of Dummy Variables Effects on Sales

4.4 Train-Test Splitting

The dataset was split into a training set and a test set, following a chronological order due to the time series nature of the dataset. From a marketing perspective, the division was on January 1, 2023: this aligns with the strategy where past data are used to predict future trends starting from the beginning of the year.

A validation set was not used, as the goal was to leverage all available training data for model estimation, given the limited number of observations. Moreover, the perspective adopted for the task was of an unsupervised forecasting of future data.

4.5 Range-Aware Scaling

A transformation of the variables was necessary, due to differences in scale and measurement units, to lighten the computational weight and facilitate the convergence of the models employed. The principal transformations considered were **Min-Max Scaling** and Normalization. Scaling was chosen because the non-linear functions applied require input values greater than zero. However, a key challenge was defining a scaling method that accounted for the original range of the variables. This was particularly important for marketing channel (Media and TV) spending variables, where the range served as a preliminary indicator of variable importance. To address this, Media and TV variables were scaled not based on their individual minimum and maximum values, but using the overall minimum and maximum separately for the two variables groups. This ensured that the relative importance of these variables was preserved in the transformation process. The transformation was made firstly to the training set, and then applied to the test set using the same maximum and minimum values, in order to treat the test set as unobserved and keep consistency.

4.6 Variables Selection

Variable selection was one of the main challenges of this project, especially because the reasonable choices were different whether we considered the statistical or marketing perspective. From a marketing point of view, the typical strategy is to avoid pre-model variable selection or to use models with specific regularization terms, and instead assess the significance of variables post-model through posterior distributions or p-values. From a statistical perspective, some approaches were explored, including priors with regularization.

Variable selection was **performed in multiple ways**, at different stages of the pipeline. Some decisions were based on the intrinsic characteristics of the variables, such as their range. Others relied on the relationship between the variables and the response, through correlation analysis or an **Ordinary Least Squares** model. Furthermore, post-model selection was performed through posterior distributions analysis. These techniques, jointly with the marketing-driven Tutor's insights, showed that many variables were redundant or not particularly useful. In the end, to maintain consistency from a marketing perspective, we decided not to use prior distributions that performed variable selection, and we kept only pre-model and post-model selection.

These are the variables included in our final models, associated in groups based on their scope.

```
media_columns = ['display_sp', 'wl_vol_ctv_sp', 'search_sp']

tv_columns = ['tv_fibra_cross', 'tv_istituzionale', 'tv_mobile']

channel_columns = [media_columns, tv_columns]

competitors_columns =['tot_tv_comp_fibra', 'tot_tv_comp_mobile',
                     'tot_tv_comp_istituzionale', 'tv_comp1_mobile']

price_columns = ['web_prezzo_ottico', 'max_delta_store_comp',
                'max_delta_web_comp', 'delta_web_comp1', 'delta_store_comp1' ]

control_columns = ['covid19','special_promo','ferragosto','christmas']
```

5 Models

5.1 Bayesian Approach

A key challenge in Marketing Mix Modeling (MMM) is the **identification problem**, which arises due to the presence of multiple valid solutions. This issue stems from three main factors: a limited number of observations, correlated spending patterns, and a high number of variables. Typically, MMM relies on weekly sales and advertising spend data spanning 2 to 5 years, resulting in only 100 to 300 observations for model fitting. Such a constrained dataset makes it difficult to accurately estimate the impact of each marketing channel.
Correlated spending patterns add further complexity. Marketing budgets often exhibit low variability over time, meaning that ad spend remains relatively stable across weeks. Additionally, collinearity between different marketing channels complicates the identification of each channel's individual contribution. Seasonal trends can further amplify these issues by creating a strong correlation between marketing spend and demand, making attribution even more challenging.
As a result, multiple models may fit the data well while suggesting different budget allocation strategies, making it difficult to determine the truly optimal marketing mix.
A Bayesian approach helps overcome these challenges by incorporating domain expertise directly into the model through prior distributions and multiple levels of hierarchy. These priors are applied not only to standard parameters but also to **non-linear (NL)** and **time-varying (TV)** effects, guiding the model towards more plausible and stable solutions. By integrating prior knowledge, Bayesian MMM reduces the risk of overfitting, mitigates collinearity issues, and provides robust uncertainty estimates. This ensures that marketing decisions are based on a combination of data-driven insights and expert judgment, leading to more reliable and actionable strategies for budget optimization.

5.2 Non Linear Components

In the scientific literature the response of revenue to advertising (and more in general to marketing actions) has long been hypothesized to be non-linear. To enhance the predictive capability of our models we decide to apply non-linear transformations to the advertising spend variables to represents two key phenomena: saturation and carryover effect. These transformations make the model more realistic by incorporating both the effect of diminishing return and the persistence of advertising.

Saturation

Beyond a certain threshold, increasing spend on a promotional channel yields progressively lower marginal impact, reflecting the phenomenon of market saturation. To model the diminishing returns of advertising investments we decided to use the **Logistic Saturation** defined as:

$$f(x) = \frac{1 - e^{-\lambda x}}{1 + e^{-\lambda x}}$$

where the parameter $\lambda > 0$ controls the rate at which saturation is reached.

Carryover effect

The carryover effect consists in the delayed effect of advertising on sales. An advertising investment can have a prolonged impact over time, as in the case of a TV commercial whose jingle remains in consumers' memory even weeks or months later. To model this phenomenon we use **AdStock** with

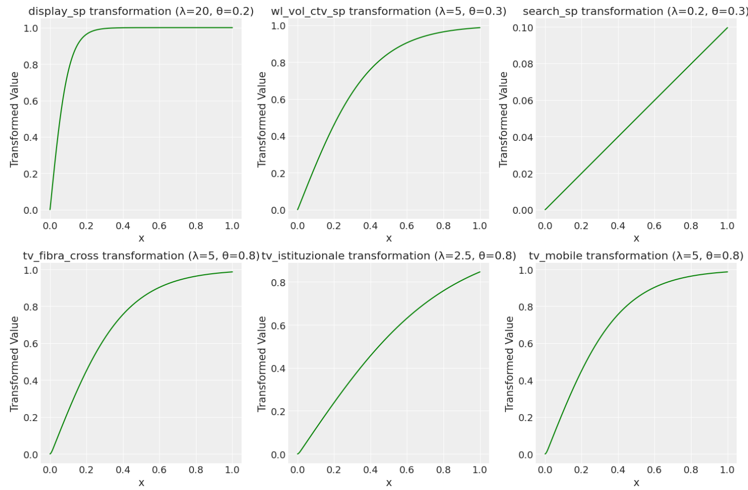


Figure 3: Non Linear transformations, with prior parameters, applied to `channel_columns`.

Geometric Decay where the effect of advertising over time is computed as a convolution of the time series of the advertising spend with an exponentially decaying weight sequence:

$$\text{AdStock}(x_t) = \frac{\sum_{l=0}^{L} \theta^l x_{t-l}}{\sum_{l=0}^{L} \theta^l}$$

where the parameter θ (with $0 \leq \theta \leq 1$) represents the decay rate. A higher θ value implies a more persistent advertising effect over time. L is the maximum memory window considered, which is fixed L=10, as we considered looking back 10 weeks sufficient for our model without introducing an additional parameter.

Implementation details

The parameters for the AdStock and Saturation functions, and the order in which they are applied, were deterministically set based on visual inspection of the plots of each variable against the output and theoretical assumption found in scientific literature. The selected values reflect observed patterns in diminishing returns (saturation effects) and carryover effects. The goal was to visually assess the nonlinear dependencies as precisely as possible before modeling, in order to maximize the linear relationship between the variables and the response.

For each variable in the dataset, the transformations were applied as follows:

- For digital media variables, the *logistic saturation* transformation was applied first, followed by *adstock*. This accounts for immediate saturation effects before modeling the carryover effect.
- For TV-related variables, the order was reversed: *adstock* was applied first to smooth the time series, followed by *logistic saturation* to capture diminishing returns after the delayed effect.

To validate the transformations, we visualized the variables before and after applying adstock and saturation. Additionally, we computed the correlation of the transformed variables with the output to ensure that the transformations improved (or at least preserved) their explanatory power within the model. We then used these informations to set up the priors of the model.

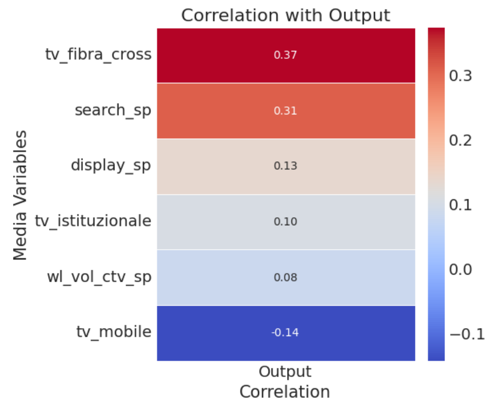


Figure 4: Before NL transformations.

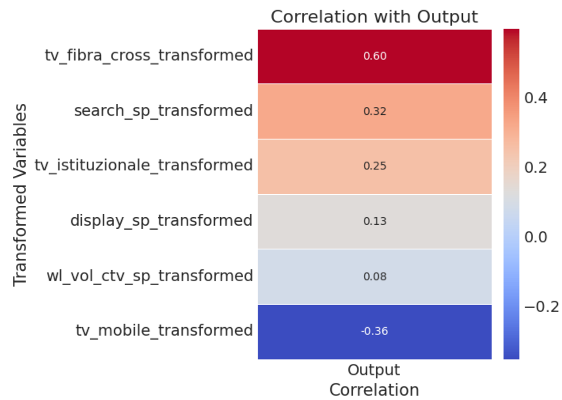


Figure 5: After NL transformations.

5.3 Time Varying Components

Another key concept is that the market changes over time. Specifically our field of investigation is subject to a strong seasonality. To model this effect we implement two different time varying components that worked at two different levels.

Sales seasonality

To represent the general seasonality of the market we decided to use a **Fourier Basis** that, unlike the introduction of other dummy variables on weekly/monthly basis, gives us a continuous and more parsimonious representation of the phenomena. The Fourier basis functions model the periodicity and seasonality in the data through a linear combination of sine and cosine terms, capturing repeating seasonal patterns (such as weekly or yearly cycles). In this case, we chose to model the annual seasonality with a period of 52 weeks, using six sinusoidal terms at 3 different frequencies; this term is added in the model with an apposite coefficient.

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

Advertising effectiveness seasonality

The effect of advertising on the sales can be different depending on the time of the year. To model this, we decided to use a **Gaussian Process (GP)** which allows us to capture dependencies over time. The GP acts as a latent process that increases or decreases simultaneously the effect of all the channel of advertising, since it multiplies the linear combination of `channel_columns`. This interaction enables the model to capture how the impact of advertising and promotional activities changes dynamically over time, adapting to different market conditions. We use a `Periodic` kernel (also known as ExpSineSquared Kernel) to correctly model the periodic nature of the time series and we tuned the smoothness of the GP with the parameter `length_scale`.

$$k(x, x') = \sigma^2 \exp\left(- \frac{2\sin^2\left(\frac{\pi|x-x'|}{52}\right)}{\lambda^2} \right)$$

Implementation details

The two aforementioned techniques influence response prediction in different ways: the Fourier basis is added independently, while the GP is linked to marketing spending variables. As a result, they can be used jointly in the model. However, since they aim to explain similar and highly correlated effects, we decided to test models using each technique individually as well as both together. Furthermore, the sampling of models with a GP with length scale defined by a prior distribution is usually associated with long computation times and divergencies in the chains. For this reason in our first model we decided to explore the possibility of not implementing the Gaussian Process creating a model that accounts for seasonality only with Fourier basis functions (Model 1).

We then removed the Fourier basis from our model and we introduced the GP with length scale as an hyperparameter, with a prior Gamma distribution. After understanding a plausibile value for this hyperparameter we decided to implement the second model with a fixed length scale, in order to save computational weight (Model 2).

We then combined the results of the two models to address both effects (Model 3). Since the complete model was very complex and prone to problematic interpretation from a business point of view, we decided to use informative priors for both components, using the Fourier and GP posteriors means of the first two models as priors for the third one. In this case we have chosen a larger length scale: the resulting posteriors for the two temporal effects are reported in Figure 6.

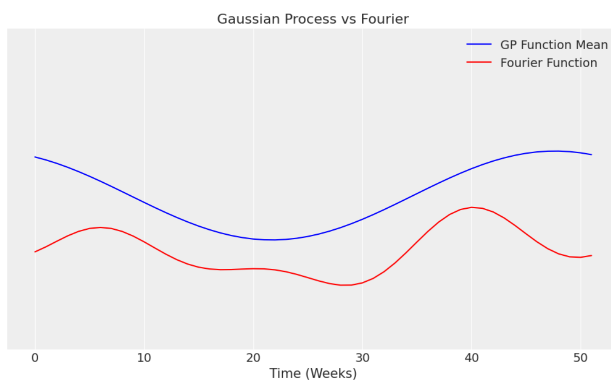


Figure 6: GP and Fourier function have a similar trend as seen in this comparison from Model 3

5.4 Model Formulation

All the models tried are structured as a hierarchical Bayesian linear regression model with transformed advertising variables; the key difference in the different versions is the way in which the time effects are considered: with Fourier basis and/or Gaussian Process. Several formulations were tried, with different prior distributions configurations. The following is the formulation that lead to the best performances, keeping consistency between statistical and marketing interpretations. As a consequence, all these priors can be considered **informative**.

The likelihood of the model is obtained by:

$$y_t \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_t, \sigma), \quad t = 1, \dots, T \tag{1}$$

where

$$
\begin{aligned}
\mu_t = \beta_0 + f_t \cdot \left(\sum_{i=1}^{n_{\text{channels}}} X_{\text{transformed},i,t} \cdot \beta_{\text{channel},i} \right) + \\
+ \sum_{i=1}^{n_{\text{price}}} (X_{\text{price},i,t} \cdot \beta_{\text{price},i}) - \sum_{i=1}^{n_{\text{competitors}}} (X_{\text{competitors,i,t}} \cdot \beta_{\text{competitors, i}}) + \\
+ \sum_{i=1}^{n_{\text{control}}} (X_{\text{control},i,t} \cdot \beta_{\text{control},i}) + \sum_{i=1}^{n_{\text{fourier}}} (X_{\text{fourier},i,t} \cdot \beta_{\text{fourier},i})
\end{aligned}
\tag{2}
$$

and

$$\sigma \sim \text{HalfNormal}(0.1) \tag{3}$$

and the following prior distributions:

- Saturation and Adstock parameters were defined with apposite distributions that respected their range, defining a small standard deviation and a different mean for each variable, previously chosen through plots and correlation with the target variable.
  - $\lambda \overset{\text{iid}}{\sim} \text{Gamma}(\mu = \text{lambda\_params}, \sigma = 0.05)$. Saturation parameter, $\lambda > 0$.
  - $\theta \overset{\text{iid}}{\sim} \text{Beta}(\mu = \text{theta\_params}, \sigma = 0.05)$. Adstock parameter, $\theta \in (0, 1)$.
- $\beta_{\text{channel}} \overset{\text{iid}}{\sim} HalfNormal(0, 0.08)$.

  $X_{\text{transformed}}$ is the matrix containing the non-linear transformations of `media_columns` and `tv_columns`, as described in Section 5.2. They are assumed to be directly proportional to the response.
- $\beta_{\text{competitors}} \overset{\text{iid}}{\sim} HalfNormal(0, 0.1)$.

  Since this group of variables is subtracted in μ_t, it is assumed to be inversely proportional to the response, meaning that an higher difference between the company's price and the competitors' price leads to a sales reduction.
- $\beta_0 \sim HalfNormal(\mu_{\text{intercept}}, 0.1)$

  Since it represents the sales baseline, we assumed it positive.
- The remaining coefficients were not restricted by sign, but the prior means were the coefficients of an OLS model, with frequentist approach, prevoiusly computed.
  - $\beta_{\text{price}} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\text{price}}, 0.1)$
  - $\beta_{\text{control}} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\text{control}}, 0.1)$
  - $\beta_{\text{fourier}} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\text{fourier}}, 0.01)$
- The prior distribution of the Gaussian Process was:

  $f_t \sim \mathcal{GP}(0, k(t, t'))$,

  where $k(t, t')$ is the Periodic kernel defined as shown in Section 5.3 with length scale originally set with prior $\lambda_{\text{scale}} \sim \text{Gamma}(2, 4)$ and then fixed to $\lambda_{\text{scale}} = 0.7$. We fixed $\lambda_{\text{scale}} = 10$ in the final model to obtain a lighter and more stable GP.

5.5 Bayesian Framework of the Model

The joint likelihood derived from Equation 1, over all time points is the following:

$$p(y|\mu_t, \sigma) = \prod_{t=1}^{T} \mathcal{N}(y_t|\mu_t, \sigma) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \mu_t)^2}{2\sigma^2}\right) \tag{4}$$

Although this function is analytically known, evaluating directly the posterior and marginal distributions is not possible, due to the nonlinear transformations (saturation and adstock functions), the Gaussian Process, and the hierarchical structure of the parameters.

The posterior distribution of the parameters is given by:

$$p(\gamma|y) \propto p(y|\gamma)p(\gamma) \tag{5}$$

While he marginal distribution of y_t is obtained by integrating over the posterior distribution of the parameters:

$$p(y_t) = \int p(y_t|\gamma)p(\gamma)d\gamma \tag{6}$$

Where γ includes all model parameters, such as regression coefficients β, the Gaussian Process f_t, the saturation parameter λ, and the adstock parameter θ. Since these computations involve high-dimensional integrals and nonlinear dependencies, they cannot be computed in closed form. Instead, Bayesian inference is performed using MCMC sampling to approximate the posterior distribution.

5.6 Model Pipeline

All trials with the three model structures are configured following these key steps. The specific differences between the best version of the three model structures are detailed directly in Section 6. In addition to the plots reported in this section as examples, further plots are present for all three versions in the Appendix A.

- **Priors Identification - Nonlinear Functions:** Using scatterplots and correlation values, plausible prior means are identified for the two hyperparameters included in Adstock and Saturation functions, θ and λ.
- **Fourier Basis Creation:** Fourier basis columns are created accordingly to 5.3.
- **Priors Identification - Beta Coefficients:** To reduce the possibility of an identification problem, for some coefficients we found reasonable prior means through a non-Bayesian linear regression, also for the Fourier basis if included in the model.
- **Model Definition:** Priors, nonlinear transformations of advertising variables, Fourier basis and/or Gaussian Process are defined in the model, through PyMC library (2.5); the transformed variables are combined in a linear equation.
- **Prior Predictive Check:** The prior predictive distribution is sampled to validate the prior choices before inference; generally, all observed response values have to stay in 95% Confidence Intervals (CI).

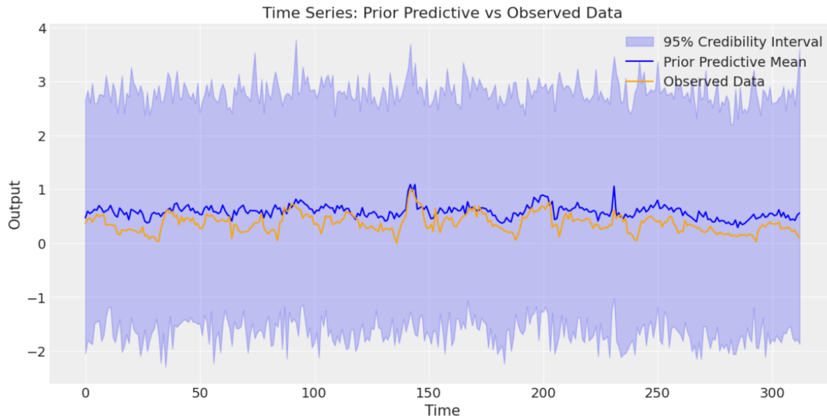


Figure 7: Prior Predictive Check of Model 3

- **Sampling:** The sampling process, in most of the cases, involved 2-3000 posterior samples with 2-3000 tuning steps; 4 chains, each run in parallel; target acceptance rate of 0.99-0.995 to ensure stable sampling.
- **Convergence Analysis:** The stability of the model is verified by ensuring the absence of divergences in the sampling chains and by checking the Rhat values for each parameter, to ensure the belonging to the established range, typically $[1, 1.01]$.
- **Trace Plot Analysis:** The trace plot consists of two key visualizations for each parameter: Posterior Density Distributions and Markov Chain Trace Plots. In the first, convergence is indicated when the distributions across chains align well. In the second, sampled values are shown, for each parameter across iterations: well-mixed traces (without trends or drifts) suggest good convergence.
- **Posterior Predictive Distribution:** To assess the predictive performance of the model, the prior predictive, posterior predictive, and observed data are compared in a plot, with respective 95% CI. The posterior predictive has to adapt as well as possible to the observed data.

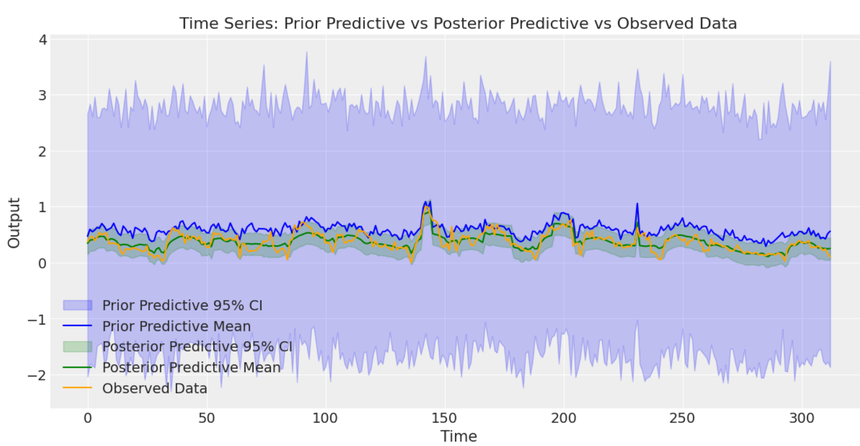


Figure 8: Posterior Predictive Distribution of Model 3

- **Contribution Analysis:** To evaluate the impact of different factors on the predicted response, we decomposed the model's output into individual components: each contribution is computed by applying the estimated posterior means of the regression coefficients to the corresponding transformed input variables, including variables and seasonality. The results are visualized over time to illustrate their relative influence on the response variable.

Figure 9: The Total Contributions of the 3 different Macro-Components in Model 3

- **Test Set Prediction:** Prior and posterior predictive distributions are extended in the unobserved test set, to assess the prediction capabilities and the possible presence of overfitting. 95% Confidence Intervals are included to assess prediction uncertainty.

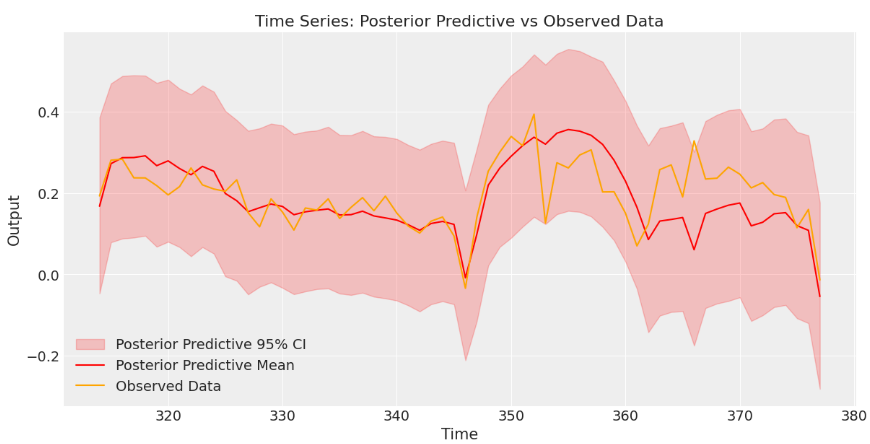


Figure 10: Test Set Predictions of Model 3

- **Evaluation Metrics:** Predictive performance metrics are computed, such as:
  - **MAPE**: Mean absolute percentage error; it is the most useful metric as it has a well-defined range and it can be minimized with a common threshold of 10%.

– **MAE**: Mean absolute error, showing the average deviation from actual values. Since it is an absolute value, it is again difficult to interpret because it is not restricted in a fixed range.

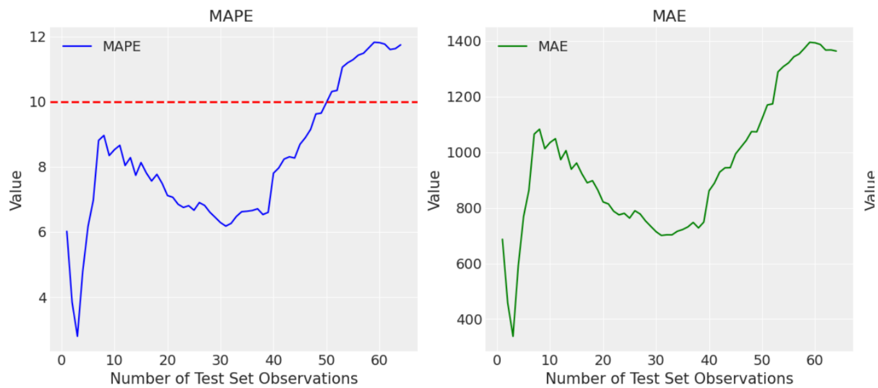


Figure 11: Evaluation Metrics of Model 3

- **Test Set Rescaling:** To interpret the predictions in their original scale, we apply inverse min-max scaling, using the training set's minimum and maximum values. This transformation can be applied to the true and predicted values and to the upper and lower bounds given by CIs:

$$y = y \times (\max(y_{\text{train}}) - \min(y_{\text{train}})) + \min(y_{\text{train}})$$

  The rescaled predictions provide a more interpretable comparison from the marketing point of view.

6 Results

6.1 Model Evaluation

The model's performance was assessed through multiple statistical metrics and plots. The most important results are collected in the next table:

Model	**Fourier Basis**	**Gaussian Process**	**MAPE**	**MAE**
Model 1	Yes	No	13.0005%	1463.2249
Model 2	No	Yes	13.3468%	1538.4798
Model 3	Yes	Yes	11.7472%	1363.9828

Table 1: Comparison of the final versions of the three model structures.

Details and considerations for the three models are the following:

- **Model 1:** Time effects are considered only with a **Fourier basis**, independently added to the other components. This model is able to understands the seasonal trend with great accuracy and without the use of hyperparameters making the chain sampling quick. However it doesn't model the coefficients changes over time assuming the β parameters of the channel variables constant.

- **Model 2:** Time effects are considered only with a **Periodic Gaussian Process**, but multiplied to the spending variables. Including a GP, the elasticity (measured by the β coefficient) of each channel can change over time, giving a relevant insight from the marketing point of view. Since this technique is more complex and computationally heavy, the length scale parameter has an important role and the best fixed value or prior distribution has to be found through several trials and considerations.

- **Model 3:** The effects of the Fourier basis and the GP are **combined**, using as prior means the posterior means extracted from Model 1 and 2. Including both the time effects adds detail and interpretability to the model, but caution is needed: overfitting risk becomes higher. Using reasonable and more informative hyperparameters and priors the result is the best achieved, since the seasonality is taken in all its different forms.

To summarize the results obtained, among the three models tested, the combined model (Fourier Basis + Gaussian Process) performs the best in terms of MAPE. All three models have a solid foundation due to the presence of non-linear components through the variable transformations and time-varying effect modelization. However, what makes Model 3 stand out is the synergy between the Fourier basis and the Gaussian Process. The Fourier basis captures the overall seasonal trends in the market, while the Gaussian Process adapts to capture the varying impact of advertising on sales, depending on the time of year. This combination enables the model to detect future trends more precisely, leading to more accurate predictions.
The MAPE of 11.75% is comparable with the recommended benchmark of 10%, and represents a strong result, as confirmed by our tutor, and demonstrates the model's robustness and effectiveness in sales forecasting.
By analyzing the plots tracking the evaluation metrics (MAE and MAPE) in Figure 11, it is clear that performance gets worse in the most recent time periods. For example considering only the first 50 weeks of the test set the MAPE of Model 3 is 10% (and this value decreases up to 7% with only 40 weeks). This suggests that the model predicts the response well without excessive overfitting, but potential changes in the market, marketing strategy, or variable collection method may introduce instability in the predictions. During the project the expertise of our tutor helped us to individuate and isolate the most important ones like shift in the marketing and advertising strategy of our client (or special offers resulting in dummy variables) resulting in better results both in train and test set.

6.2 Variable Contributions

The models presented in this paper highlights the relations and dependencies between advertising and sales. Both the macro groups of channels (i.e. **Media and TV**) demonstrated their relevance. Observing the plots (Figures 22, 33, 41), given the presented adv spent of our client, it's clear that the general impact of Media and Tv on their sales is comparable. This shows the importance of differentiate between different channels to reach the biggest possible number of people. Of all the advertising channel the most correlated with the Output turned out to be `display_sp`.
As regards the relationship with **competitors**, consistently with what was expected, it was found that having less expensive products than the competition favors sales. The contribution to sales given by the price difference was positive for the entire first part of the period considered. The change in the price relationship with competitors that occurred around week 250 was one of the causes of the general decrease in sales observed in the following weeks. At the same time, it was seen that the influence of the competetitor's advertising expenses repeated the same pattern throughout almost all the period and appears to have a less significant impact on our client's revenues.
Finally, the contribution plots also indicate the importance of direct representation of **seasonality**

(for example in Figure 21). The intercept, used in combination with the Fourier seasonality (if present), represents the baseline of sales in a given period of the year. However as shown in the plot the magnitude of this formulation is not enough to completely explain all the sales. The temporal aspect is still crucial and the dummy variables were important to explain the anomalous data of the festive weeks. The dummies for particularly effective promotions and for the periods of restrictions following the Covid-19 pandemic were even more important, with the limitation of being **special events** for which the modeling does not derive directly from the data but from our previous knowledge. This constitutes a limit for the forecasts on future data.

7 Conclusions

7.1 Models Comparison

The performance of the models presented in this report are comparable and the choice between one of them strictly depends on the needs of the user. The results presented in this paper show that it's possible to represent seasonality, and more in general time-varying effects, at two different levels without significant loss in terms of performance. Both Fourier function and Gaussian Process correctly identified the correct seasonal trend of the Market with low values corresponding at low-seasonality period like the Summer vacations and high-values corresponding at high-seasonality like the "Back-to-School" period.
Model 1 is surely the simpler model proposed and it's also the most efficient in terms of both time of execution and implementation. If the goal of the client is simply to predict the general trend of sales it's definitely the best choice.
Conversely in a business context, when the primary reason is the understanding of the relations between advertising and sales, having a bigger insight on the role of the effect of the channels is crucial. At the beginning of the project, when we were using another library, `pymc-marketing`, which is less customizable, the model assigned excessive weight to seasonality and intercept. This led to a challenging interpretation from a marketing perspective, as it did not provide predictive value for the variables related to marketing spends, suggesting the uselessness of advertising! The seasonality expressed by Fourier function is in fact an external and not changeable factor: model this seasonal variability in the channel parameters through a GP is a bigger improvement. However, as mentioned multiple times, using a Gaussian Process has some negative aspects. **Model 2** is the best choice if one prefers more "informative" predictions even if slightly less precise. Our results show that the best results are obtained when the smoothness is comparable with the 3-basis Fourier implemented in Model 1, underlining once again the close relationship between the two formulations.
Model 3 combined both this approaches in a single model. Given its complexity it might lead to not clear or misleading conclusions (the well known identification problem). To solve this it's necessary the implementation of strict priors obtained by the previous computations of the others model. Using the weighted posterior of model 1 and 2 let the model to combine the seasonality of the sales with a smoother version of the GP in coherent way. The result is a model able to obtain the best performance describing the trend of the sales both as an external uncontrollable phenomenon and as a business-dependent one.

7.2 Key Findings and Limitations of MMM

The results of this project highlight the relationship between marketing investments and sales, showing the important impact on the sales of the advertising channels. The Bayesian approach used for the creation of this Marketing Mixing Model offers a robust and data-driven framework for decision-

making that helped us to distinguish and understand the different phenomena that characterizes the optical-fiber market. We observed the importance of all the considered adv channels and we discovered that the impact of **Media and Tv advertising** is comparable.
We also faced (and learned to overcome) the problems linked with the choice of the parameters of **Saturation and AdStock**, key to describe the persistence of a marketing campaign in the consumers' mind. These transformations are highly sensitive to their hyperparameter and, without an accurate analysis, significant transformation issues would have arisen, making the variables unusable and leading to identification problems, with all the weight assigned to the intercept and seasonal trend, making the model useless by a business point of view.
The representation of the **Seasonality** was the core of our work: we have shown that can be represented directly with Fourier basis, indirectly investigating the temporal variation of the effectiveness of the advertising with a Gaussian Process and with a combination of the two. A wrong choice of the smoothness of these functions can lead to poor performance and/or bad results.
The analysis of **Competitors'** data highlighted the influence of their advertising campaigns and the importance of offering competitive prices. Adjusting the promotional strategies considering the competitor movements can enhance the market positioning.
Unfortunately special promotions, although very impactful on the sales, are very difficult to be modeled: the presence of special situations, if known, must be highlighted and isolated. If significant external factors (e.g., economic shifts, new competitors) change market dynamics, the model may lose his predictive ability and retraining may be necessary. The biggest limitation of a **Marketing Mix Modeling** is in fact the assumption that the period considered for the training of the model is representative of the market situation also in the following months/years and therefore it assumes a certain degree of **Market Stationarity**.
Another limitation of the model is that, due to its complexity and hierarchical structure, it does not allow for closed-form solutions for updating the posterior parameters and deriving the marginal distributions, which makes the implementation of the Bayesian framework more computationally demanding and less interpretable.

8 Bibliography

References

[1] N. H. Borden. The concept of the marketing mix. *Journal of Advertising Research*, 4(2):2–7, 1964.

[2] D. Chan and M. Perry. Challenges and opportunities in media mix modeling. `https://research.google.com`, 2017.

[3] Ryan Dew, Nicolas Padilla, and Anya Shchetkina. Your mmm is broken: Identification of nonlinear and time-varying effects in marketing mix models, 2024.

[4] Facebook Experimental. Analyst's Guide to Marketing Mix Modeling (MMM).

[5] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984.

[6] R. Gesztelyi, J. Zsuga, A. Kemeny-Beke, B. Varga, B. Juhasz, and A. Tosaki. The hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences*, 66(4):427–438, 2012.

[7] Chang Gong, Di Yao, Lei Zhang, Sheng Chen, Wenbin Li, Yueyang Su, and Jingping Bi. Causalmmm: Learning causal structure for marketing mix modeling. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, page 238–246, New York, NY, USA, 2024. Association for Computing Machinery.

[8] Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, and Jim Koehler. Bayesian methods for media mix modeling with carryover and shape effects. Technical report, Google Inc., 2017.

[9] T. Lorenzen. In response to 'your mmm is broken', a viable estimation strategy. *ScanMarqed Blog*, 2024.

[10] Edwin Ng, Zhishi Wang, and Athena Dai. Bayesian time varying coefficient model with applications to marketing mix modeling, 2024.

[11] PyMC Marketing. Marketing Mix Modeling: ROAS.

[12] PyMC Marketing. Marketing Mix Modeling: Time-Varying Media Example.

[13] PyMC Marketing. Marketing Mix Modeling: Time-Varying Parameters Example.

[8][10][4][11][13][12][1][2][5][6][9][7]

A Appendix - Plots

A.1 Model 1

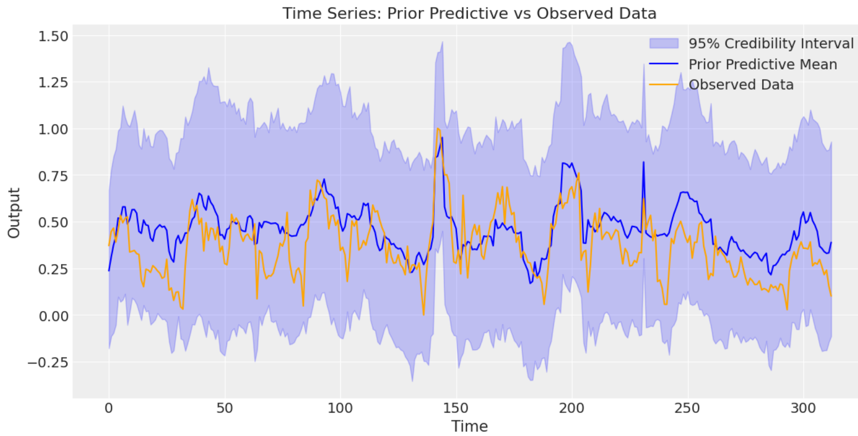


Figure 12: Prior Predictive Check

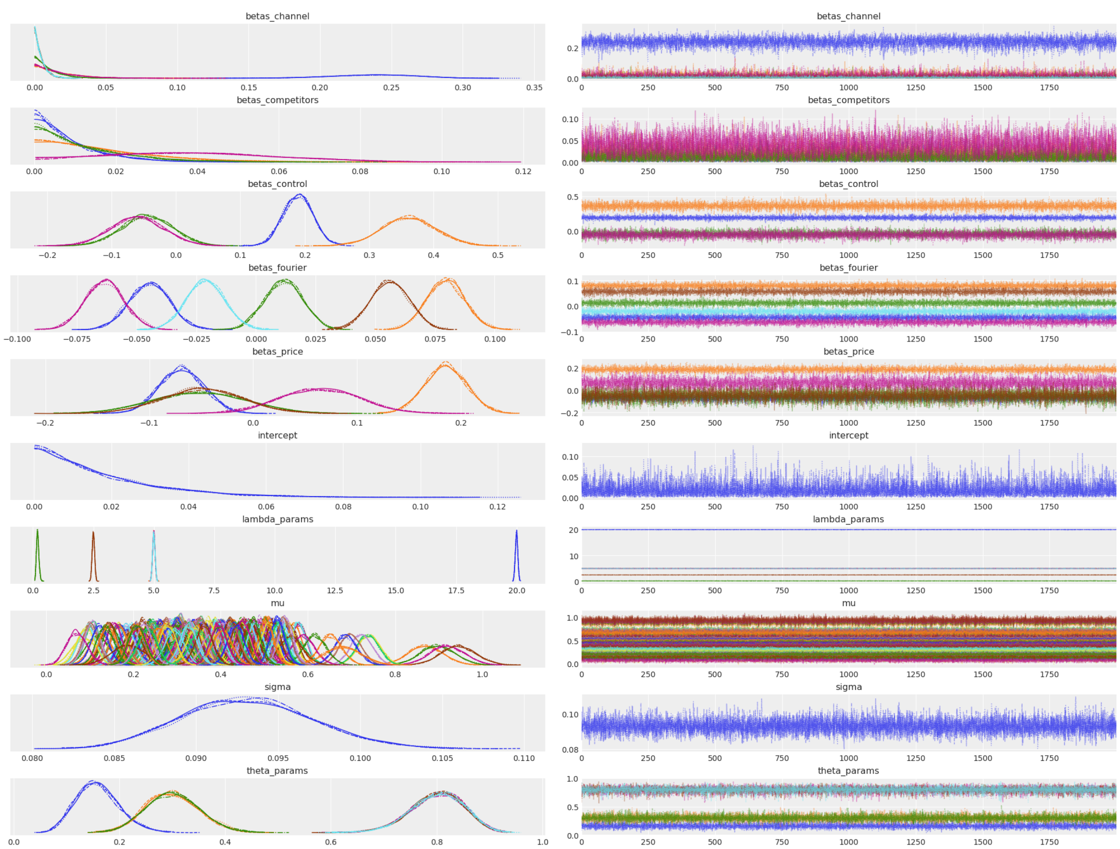


Figure 13: Trace Plots

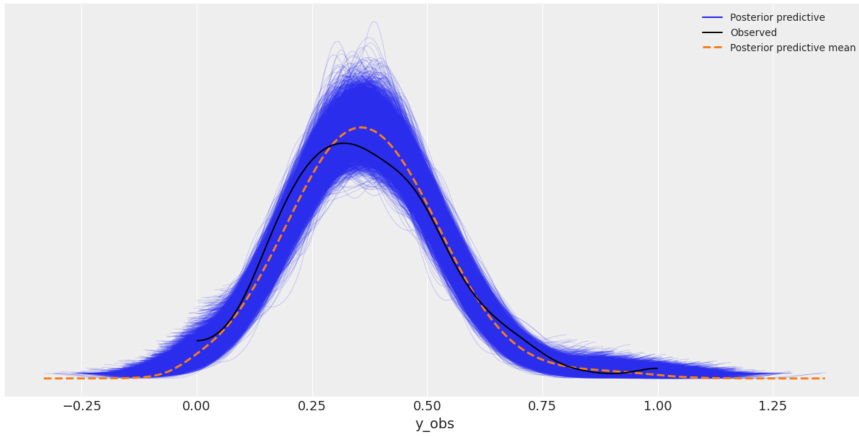


Figure 14: Posterior Predictive Check

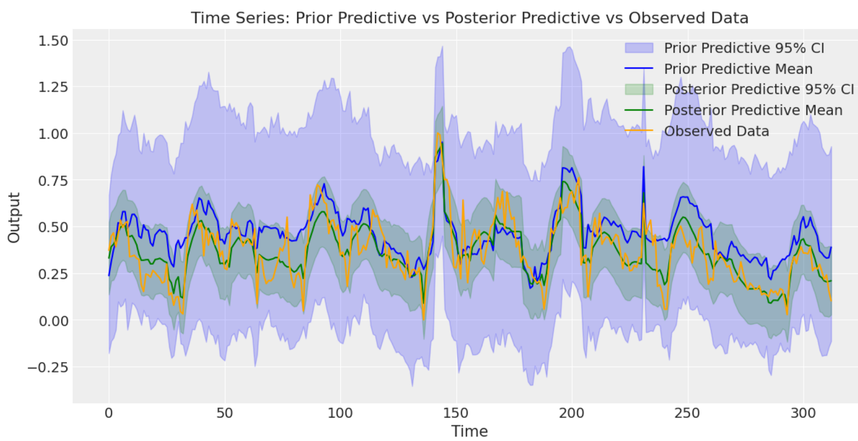


Figure 15: Prior vs Posterior

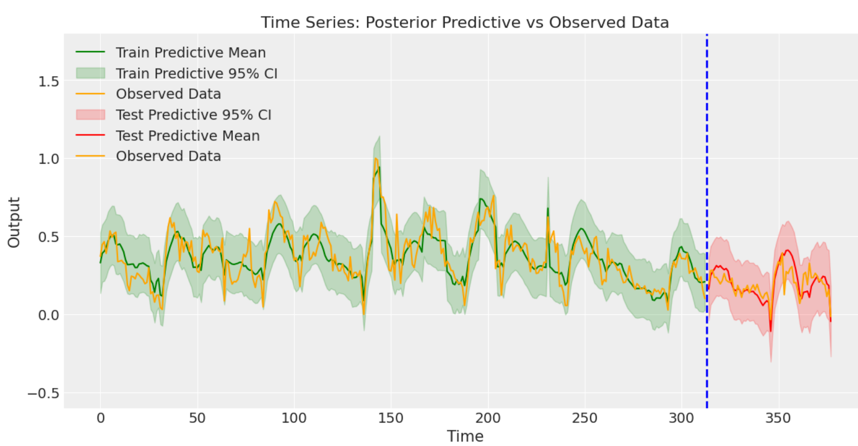


Figure 16: Posterior Train+Test

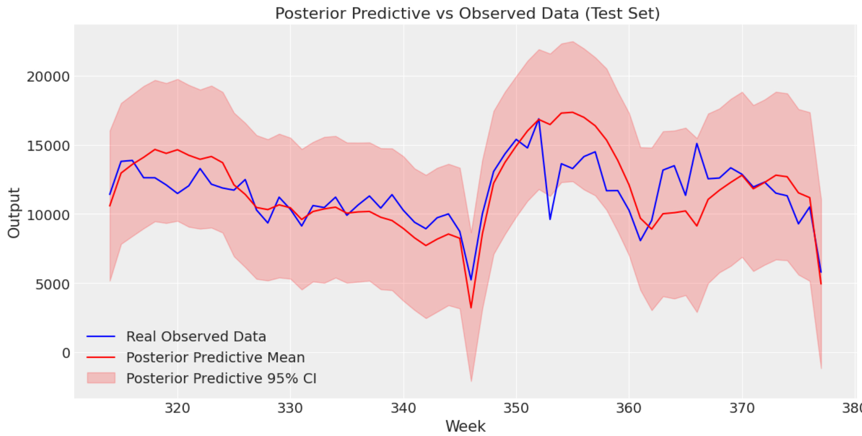


Figure 17: Test Set Predictions on Rescaled Data

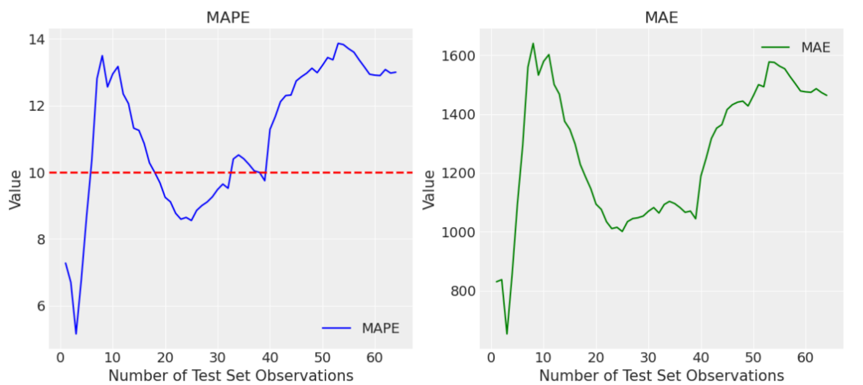


Figure 18: Evaluation Metrics

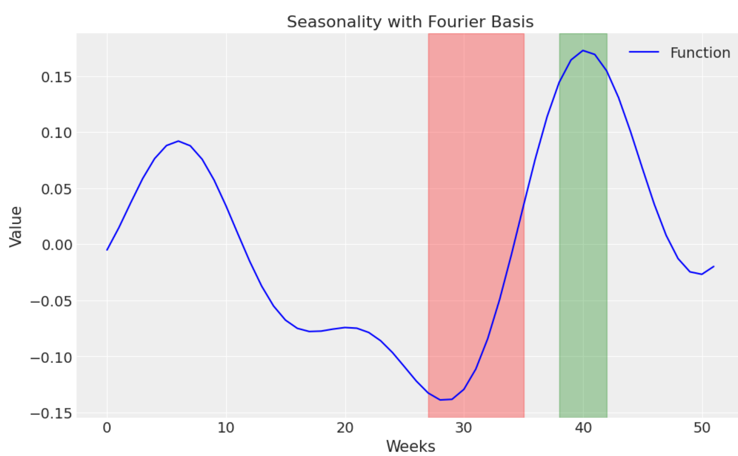


Figure 19: Fourier Seasonality

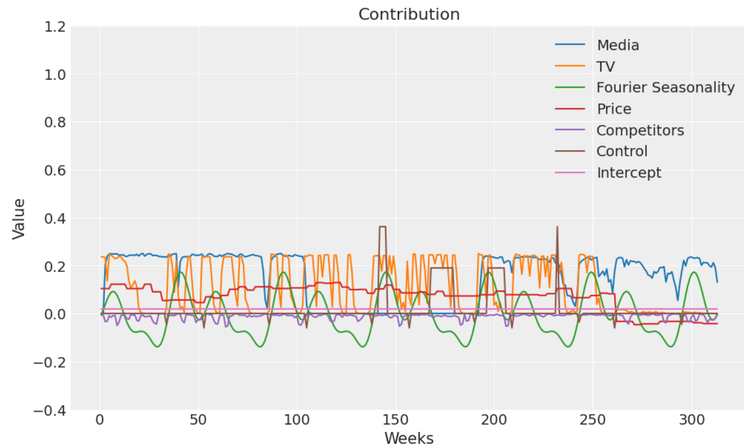


Figure 20: Contribution Details

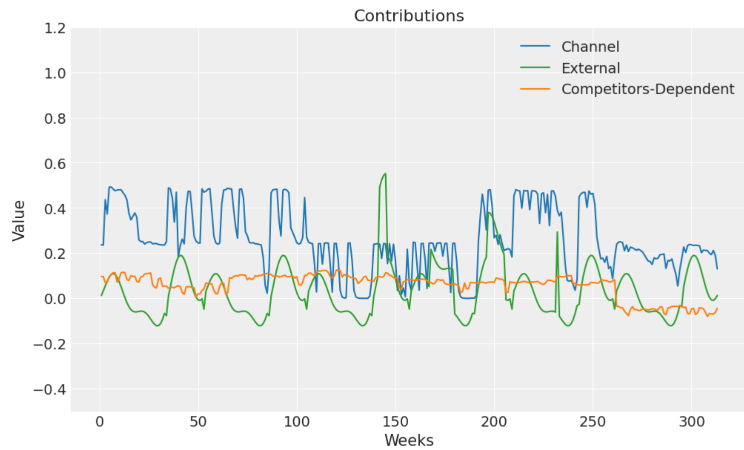


Figure 21: Macro Component Contributions

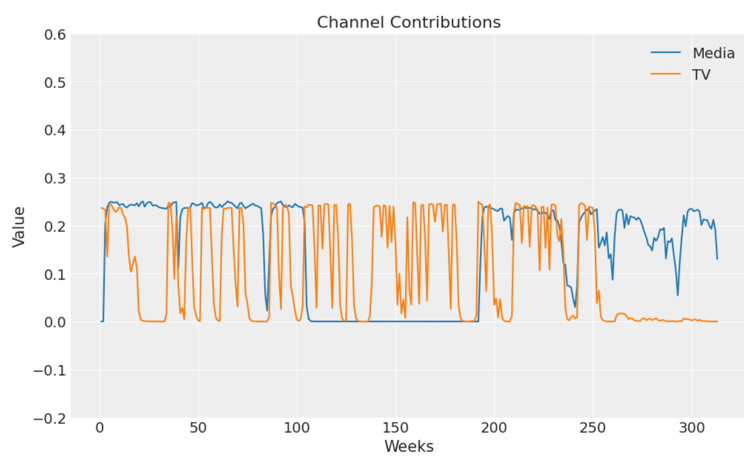


Figure 22: Channel Contributions

A.2 Model 2

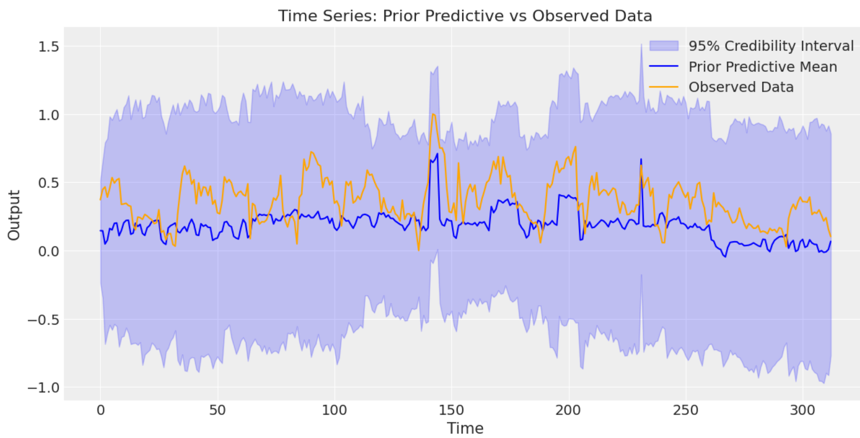


Figure 23: Prior Predictive Check

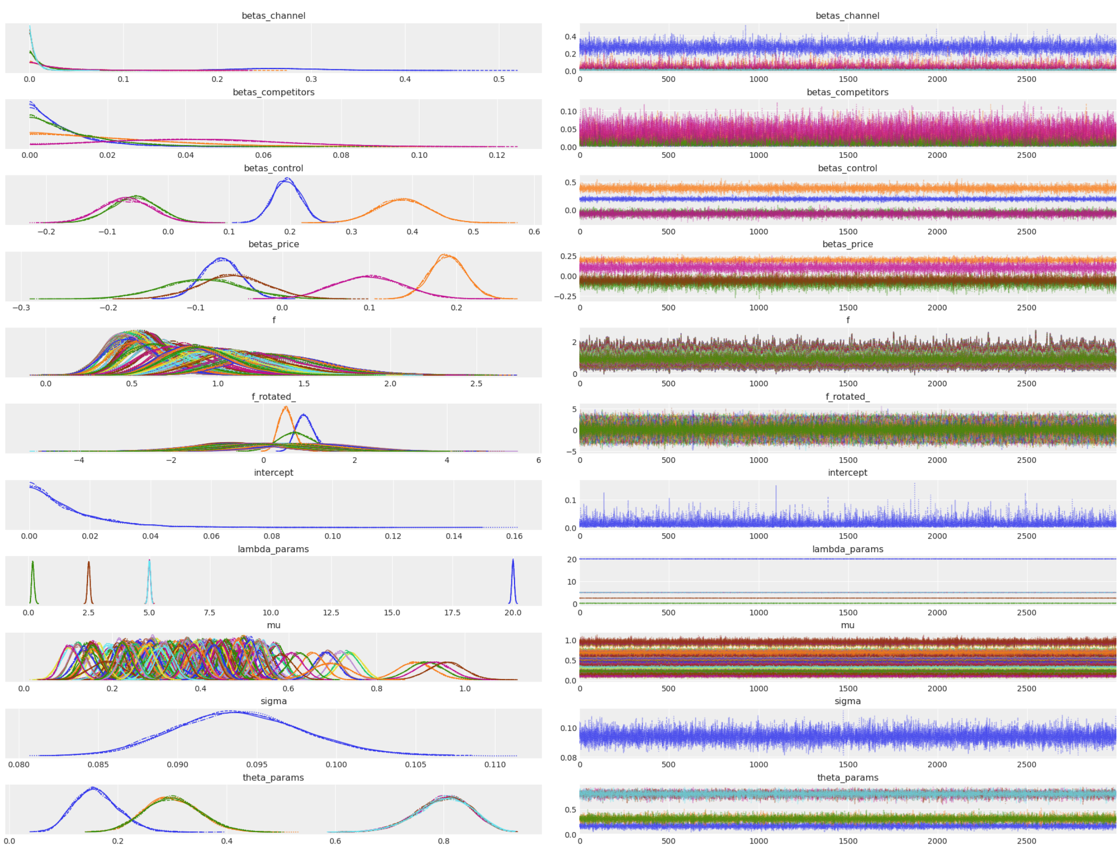


Figure 24: Trace Plots

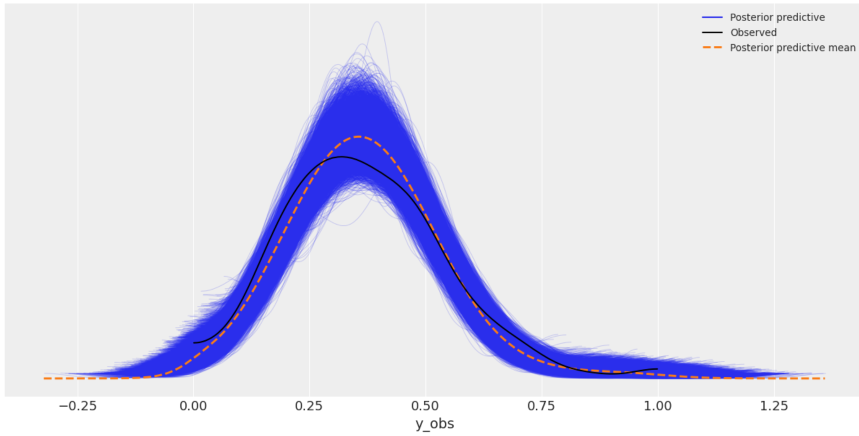


Figure 25: Posterior Predictive Check

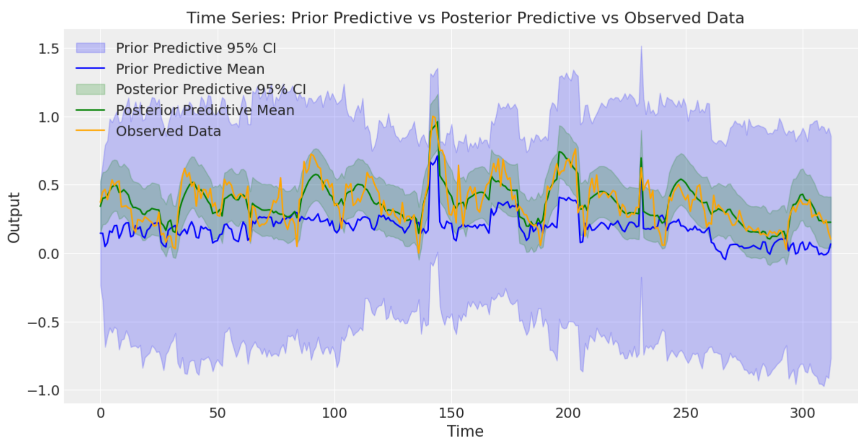


Figure 26: Prior vs Posterior

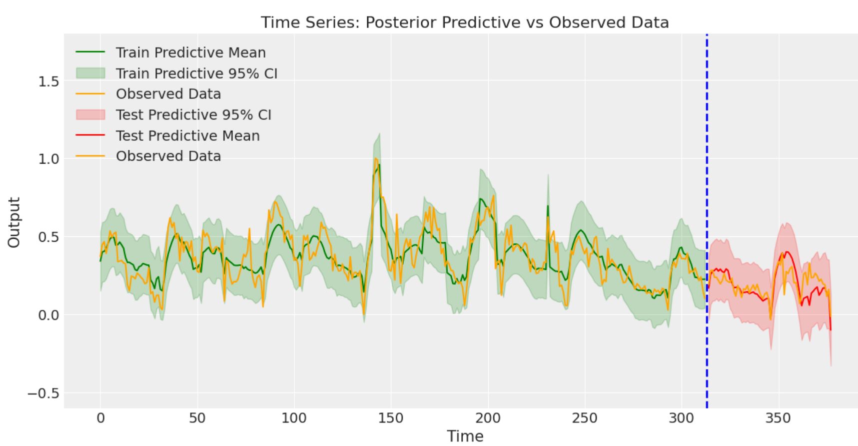


Figure 27: Posterior Train+Test

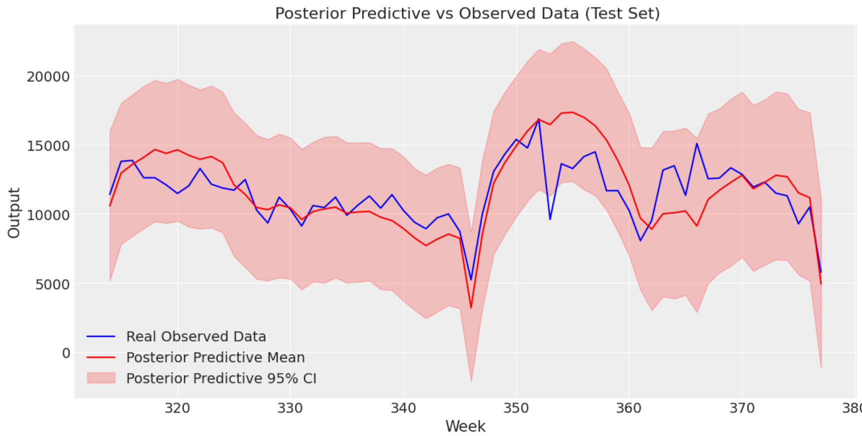


Figure 28: Test Set Predictions on Rescaled Data

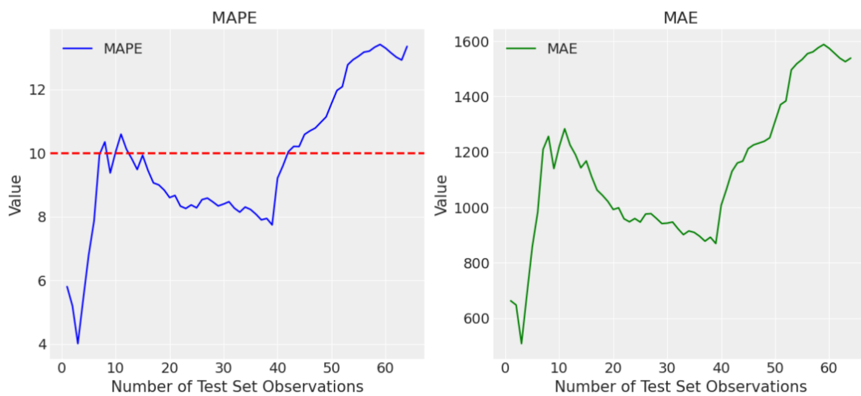


Figure 29: Evaluation Metrics

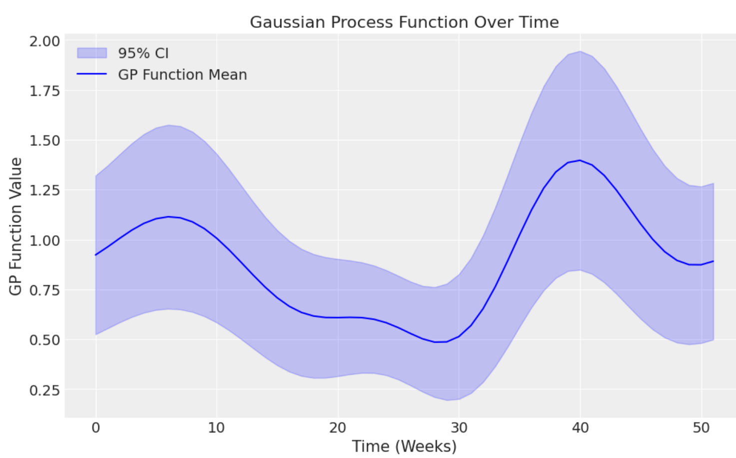


Figure 30: Gaussian Process

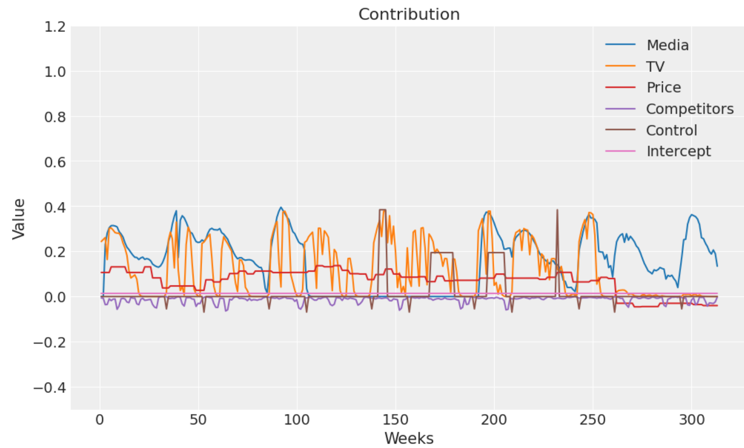


Figure 31: Contribution Details

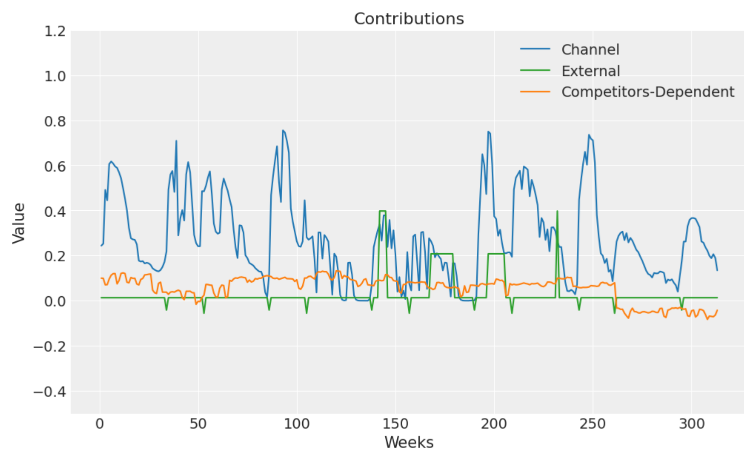


Figure 32: Macro Component Contributions

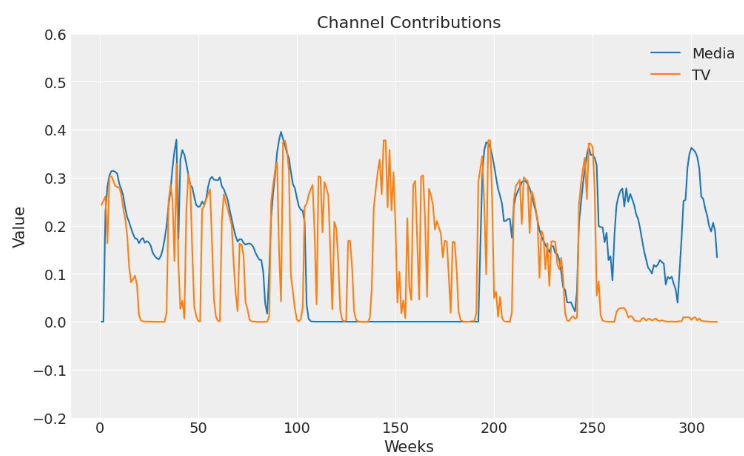


Figure 33: Channels Contributions

A.3 Model 3

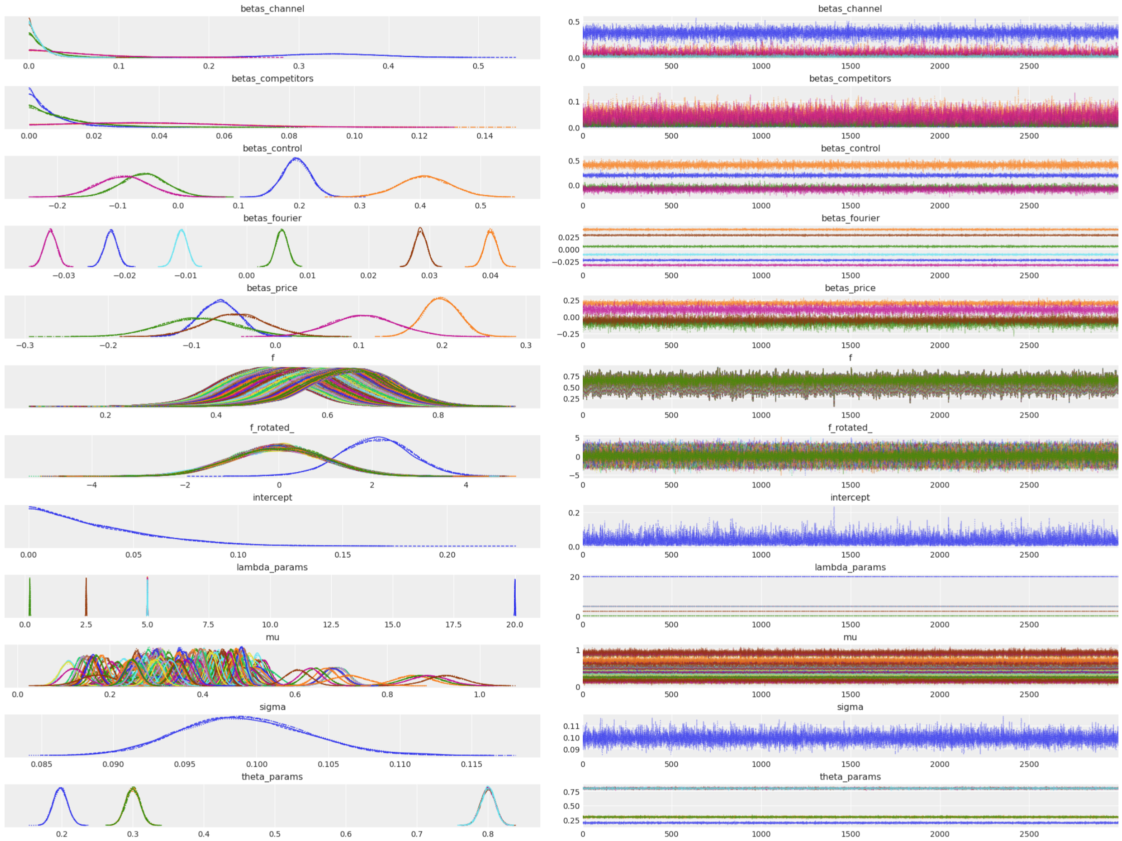


Figure 34: Trace Plots

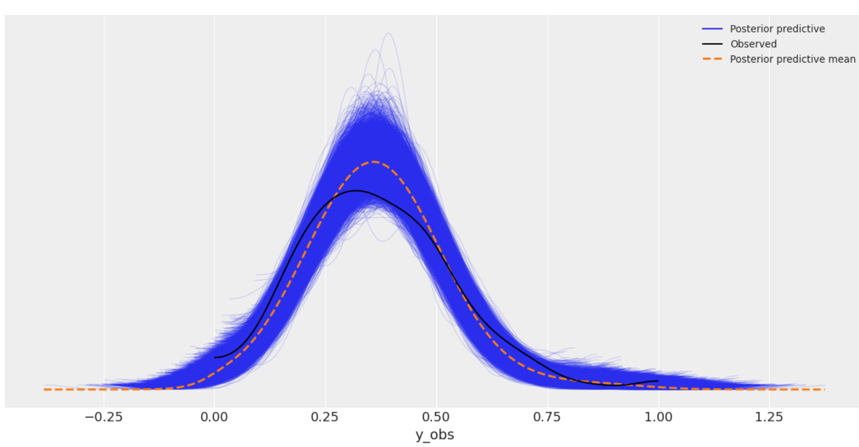


Figure 35: Posterior Predictive Check

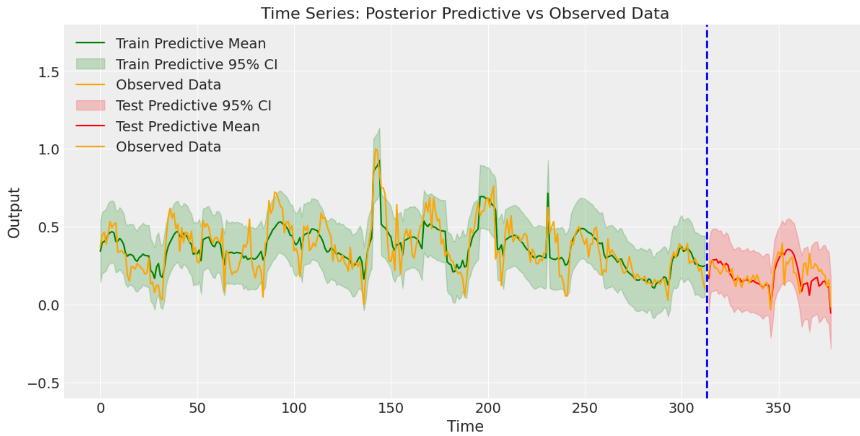


Figure 36: Posterior Train+Test

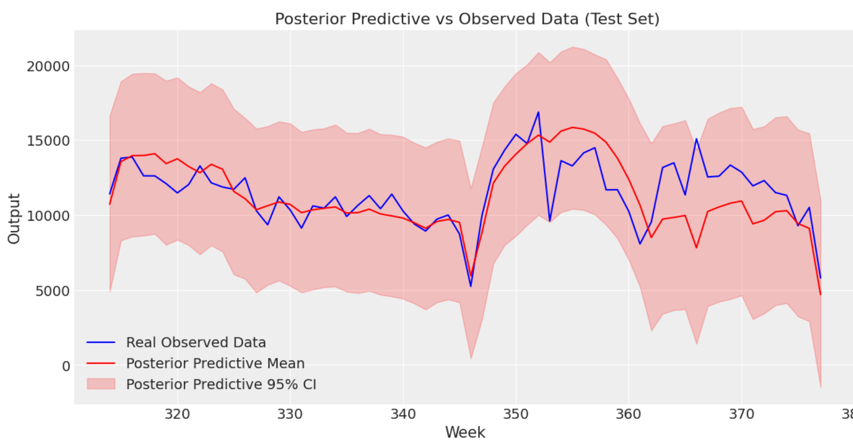


Figure 37: Test Set Predictions on Rescaled Data

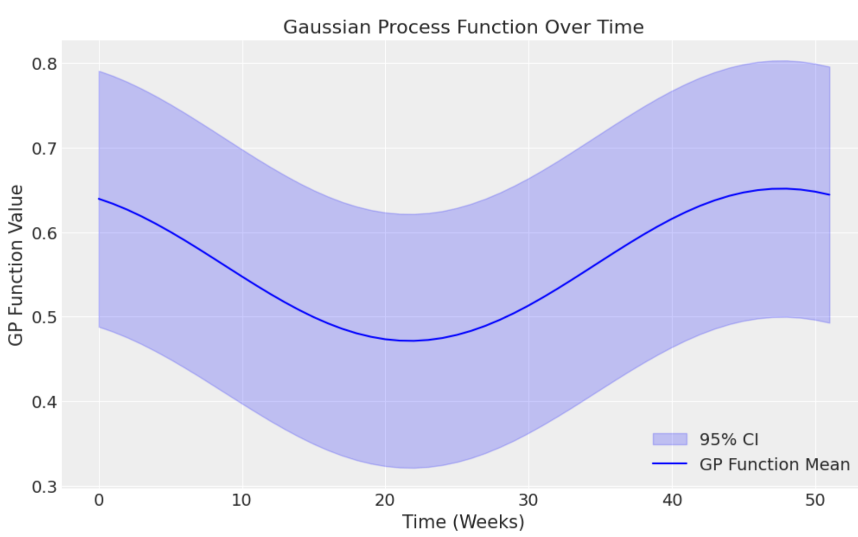


Figure 38: Gaussian Process

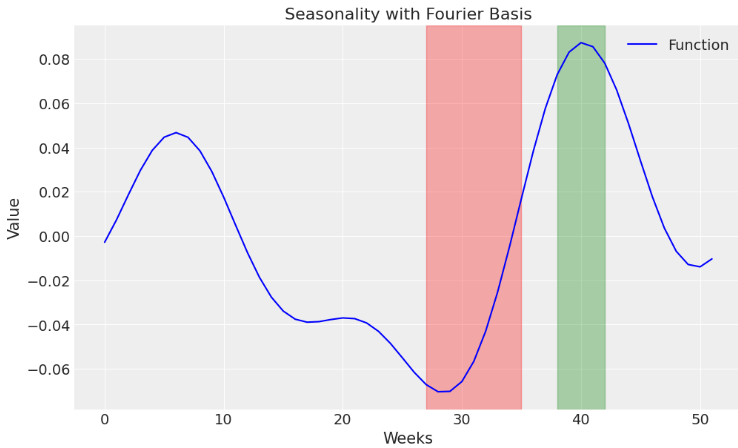


Figure 39: Fourier Seasonality

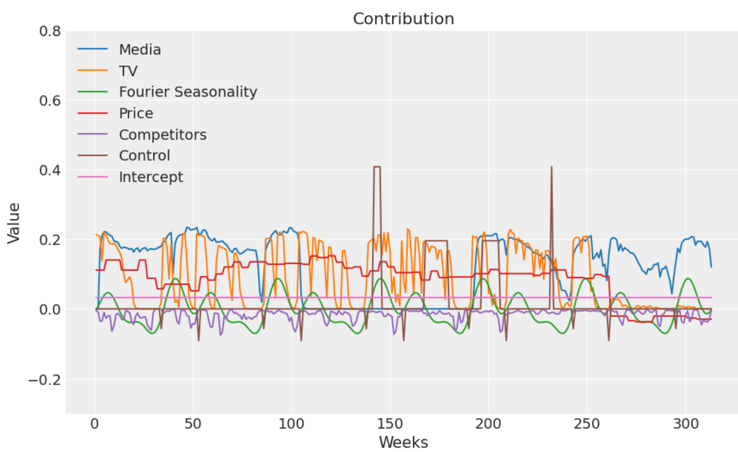


Figure 40: Contribution Details

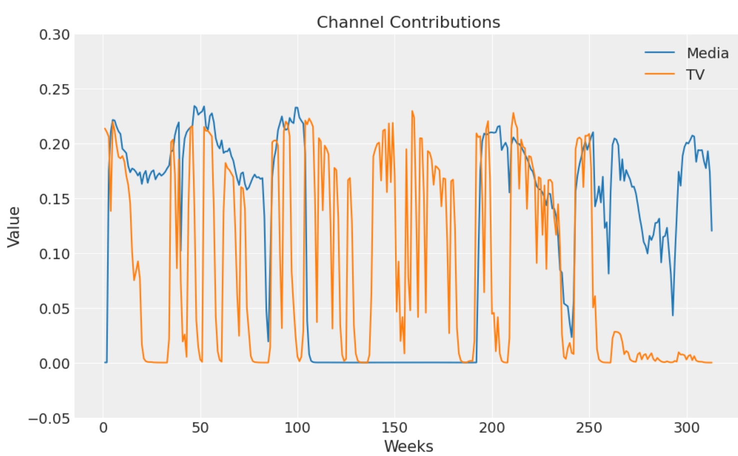


Figure 41: Channel