Griffin Documentation

info@griffin-analytics.com

Abstract

Griffin MMM (Media Mix Modelling) is a Bayesian analytical tool designed to help marketers optimise their return on investment by analysing the effectiveness of various media channels. This document serves as a guide to Griffin MMM, covering installation, configuration, data preprocessing, core modelling, and advanced customisations.

The framework employs advanced statistical techniques, such as adstock transformations to model delayed marketing effects and saturation functions to account for diminishing returns. By integrating Facebook's Prophet, Griffin MMM incorporates seasonality and holiday effects, enabling nuanced analyses of temporal fluctuations in consumer behaviour.

This guide provides a step-by-step process for configuring Griffin MMM to meet specific business requirements and interpreting results through its robust visualisation tools. Features like lift testing and cross-validation empower users to validate marketing strategies with statistical rigour.

Griffin MMM offers intuitive outputs, including saturation curves, adstock effects, and channel contribution analyses, making insights accessible to stakeholders. The Bayesian inference framework provides posterior distributions for key metrics, allowing marketers to assess uncertainty and make informed decisions.

With its comprehensive reporting capabilities and robust statistical foundation, Griffin MMM is an invaluable tool for data scientists, analysts, and technically-minded marketers seeking to optimise marketing performance and derive actionable insights.

Contents

1 Introduction **7**

2 Getting Started **7**
2.1 Installation . . . 7
2.2 Directory Structure . . . 7
2.3 Basic Setup . . . 8
2.3.1 Configuration File Setup . . . 8
2.3.2 Data Requirements . . . 8

3 Configuration Guide **9**
3.1 Configuration File Example . . . 9
3.2 Configuration Details . . . 10
3.2.1 Data Handling . . . 10
3.2.2 Column Definitions . . . 10
3.2.3 Media Channels . . . 10
3.2.4 Model Parameters . . . 10
3.2.5 Custom Priors . . . 11
3.3 Updating Configuration . . . 11

4 Data Handling and Preprocessing **11**
4.1 Loading Data . . . 11
4.2 Data Transformations . . . 11
4.2.1 Normalisation and Scaling . . . 11
4.2.2 Handling Missing Data . . . 12
4.2.3 Feature Engineering . . . 12
4.2.4 Seasonality Adjustments . . . 12
4.2.5 Adstock Transformation . . . 12
4.2.6 Saturation Effects . . . 12
4.3 Data Validation . . . 12

5 Core Modeling **13**
5.1 Adstock Transformations . . . 13
5.2 Saturation Transformations . . . 13
5.2.1 Logistic Saturation Function . . . 13
5.2.2 Implementation in Griffin MMM . . . 14
5.3 Model Structure and Bayesian Inference . . . 14
5.4 Markov Chain Monte Carlo (MCMC) Sampling . . . 14
5.5 Base Model . . . 14

6 Validation **14**
6.1 Validation Functions . . . 14
6.1.1 Target Variable Validation . . . 14
6.1.2 Date Column Validation . . . 15
6.1.3 Media Spend and Impressions Validation . . . 15
6.1.4 Extra Features Validation . . . 15
6.1.5 Cross-Validation Setup . . . 15
6.1.6 Priors and Hyperparameter Validation . . . 15
6.1.7 Visual Validation Reports . . . 15
6.2 Ensuring Data Suitability . . . 16

7 Visualization 17
7.1 Plotting Saturation Curves . . . 17
7.1.1 Technical Details of Saturation Curve Calculation . . . 17
7.2 Plotting Adstock Effects . . . 18
7.3 Lift Test Analysis Plots . . . 18
7.4 Channel Contribution Visualisation . . . 18
7.5 Stacked Bar Chart of Channel Contributions . . . 18
7.6 Uncertainty and Posterior Distributions . . . 19
7.7 Posterior Distribution Plots . . . 19
7.8 Practical Uses of Visualizations . . . 19

8 Seasonality 21
8.1 Prophet Integration . . . 21
8.2 Adding Seasonality Components . . . 21
8.2.1 Yearly and Weekly Seasonality . . . 22
8.2.2 Holiday Effects . . . 22
8.3 Advanced Trend Modelling with Prophet . . . 22
8.3.1 Additive and Multiplicative Seasonality . . . 22
8.3.2 Changepoint Detection . . . 23
8.4 Practical Applications of Prophet Integration . . . 23

9 Example Workflow 24

10 API 25
10.1 API Overview . . . 25
10.2 Common Issues and Troubleshooting . . . 25
10.2.1 Data-Related Issues . . . 25
10.2.2 Configuration Errors . . . 26
10.2.3 Model Fitting Challenges . . . 26
10.3 Useful Tips for Efficient Modelling . . . 26

11 Marketing Mix Modelling: Concepts and Methodology 27
11.1 Econometric Foundations . . . 27
11.2 Adstock and Lag Effects . . . 28
11.3 Saturation and Diminishing Returns . . . 28
11.4 Hierarchical Bayesian Framework . . . 28
11.5 Endogeneity and Control Variables . . . 29
11.6 Model Calibration and Validation . . . 29

12 Key Concepts 30
12.1 Response Functions and Shape Effects . . . 30
12.1.1 Hill-Saturation Response . . . 30
12.1.2 Michaelis-Menten Kinetics . . . 30
12.2 Saturation Effects and Adstock Transformations . . . 30
12.2.1 Geometric Adstock . . . 30
12.2.2 Weibull Adstock . . . 31
12.3 Competitive Effects and Market Dynamics . . . 31
12.3.1 Cross-Channel Elasticity . . . 31
12.3.2 Market Share Dynamics . . . 31
12.4 Bayesian Implementation . . . 31
12.4.1 Prior Specifications . . . 31
12.4.2 Hierarchical Structure . . . 32

12.5 Model Diagnostics 32
12.5.1 Convergence Assessment 32
12.5.2 Predictive Checks 32

List of Figures

1 Example Saturation Curve Showing Diminishing Returns 17
2 Example Adstock Effect Showing Delayed Impact of Media Spend 18
3 Example Lift Test Analysis Showing Incremental Gains 19
4 Example Stacked Bar Chart of Media Channel Contributions 20
5 Example Posterior Distribution Showing ROI Estimates for Media Channels 21

List of Tables

Documentation Control

Version	Date	Reason for issue	Issued By
1.0.0	28-Oct-24	Initial version	Griffin Developers
1.0.1	20-Dec-24	Minor revisions	Griffin Developers

1. Introduction

Griffin MMM (Media Mix Modelling) is a Bayesian analytical tool designed to help marketers evaluate and optimise the performance of various marketing channels. In today's complex, multi-channel marketing environment, understanding the contribution of each channel to overall objectives is critical. Griffin MMM provides a sophisticated solution for quantifying these contributions, enabling businesses to allocate budgets effectively and maximise return on investment (ROI).

What sets Griffin MMM apart is its advanced Bayesian modelling approach, which incorporates uncertainty into its estimates, offering more reliable insights. Unlike traditional media mix models, Griffin MMM delivers full posterior distributions for each parameter, providing a comprehensive view of the data and potential outcomes. By leveraging Bayesian methods, the tool allows users to incorporate prior knowledge, enhancing model performance in scenarios with sparse or incomplete data.

This documentation serves as a comprehensive guide to setting up, using, and customising Griffin MMM to meet specific business requirements. Whether you are new to media mix modelling or an experienced analyst, this guide equips you with the knowledge needed to fully utilise Griffin MMM. Key features include advanced adstock and saturation functions, intuitive configuration options, automated data preprocessing, and integrated lift testing capabilitiesâĂŤall designed to deepen your understanding of marketing impact.

Griffin MMM is particularly suited for businesses aiming to improve their decision-making around marketing spend. It offers insights into the incremental impact of each channel, uncovers synergies between channels, and simulates budget scenarios to determine optimal resource allocation. This data-driven approach moves beyond simple heuristics, empowering businesses to adopt evidence-based marketing strategies.

The tool also provides robust visualisation and reporting features, enabling users to interpret and communicate model outcomes effectively. Visual outputs, such as saturation curves and lift analyses, offer a clear understanding of how marketing activities translate into results, making it easier to justify marketing investments to stakeholders. With its flexibility, scalability, and practical features, Griffin MMM is an ideal solution for small agencies and large enterprises alike, supporting data-driven decision-making at every level.

2. Getting Started

2.1. Installation

To begin using Griffin MMM, clone the repository and install the necessary dependencies. The tool is provided as an automated package, ensuring a straightforward installation process. Refer to the demo folder in the GitHub repository: `https://github.com/griffin-analytics/griffin-mmm-demo`.

The demo Jupyter workbook is optimised for deployment on Google Colab. The demo folder includes essential files, such as the configuration file (`config.yaml`) and the holidays file, both of which are editable to fit specific requirements.

2.2. Directory Structure

Organising your project with a clear directory structure is essential for efficient use of Griffin MMM. The recommended structure is as follows:

```
project_directory/
          config.yaml
          input_data.csv
          results/
          notebooks/
              griffin_demo.ipynb
```

Each component in this structure serves a specific purpose:

- `config.yaml`: Contains configuration settings, including hyperparameters, priors, and data mappings. This file allows users to customise the modelling process.

- `input_data.csv`: The primary input data file, which should include dates, media spends, and target metrics. Column names must match those defined in `config.yaml`.
- `results/`: Stores outputs such as diagnostic plots, posterior estimates, and model summaries. Keeping results organised ensures easy access for review.
- `notebooks/griffin_demo.ipynb`: An interactive Jupyter notebook demonstrating Griffin MMM's workflow. Use this notebook to experiment with the tool and test custom configurations.

2.3. Basic Setup

After installation, ensure the following key files and folders are present:

- `config.yaml`: For model configuration.
- `input_data.csv`: As the source of input data.
- `results/`: To store outputs from the model.

2.3.1. Configuration File Setup

The `config.yaml` file is critical for defining model behaviour. Key parameters include:

- **Model Name**: Identifies the model run, aiding in the organisation of multiple experiments.
- **Data Granularity**: Specifies whether the data is daily, weekly, or monthly, informing the model's time resolution.
- **Train-Test Split**: Determines the proportion of data for training versus testing (e.g., 90-10 split).
- **Column Mappings**: Maps key columns such as `date_col`, `target_col`, and `media`, defining where to locate dates, target metrics, and media spends in the input CSV.
- **Priors and Hyperparameters**: Allows the specification of priors for Bayesian estimation, including distributions and parameters. This feature is especially valuable for users with prior knowledge, such as expected ROI for specific channels.

2.3.2. Data Requirements

Griffin MMM requires well-structured input data to ensure optimal performance. The `input_data.csv` file should adhere to the following guidelines:

- **Date Column**: Ensure the date column is formatted as `YYYY-MM-DD` and correctly specified in `config.yaml`.
- **Media Spend Columns**: Media spend columns must match the names defined in the configuration file, contain numerical values, and have no missing entries.
- **Target Column**: Clearly define the target metric (e.g., sales or conversions) and maintain consistency across the dataset.

3. Configuration Guide

Griffin MMM uses a configuration file (`config.yaml`) to define key settings such as model structure, data paths, and hyperparameters. This file provides a customisable setup, allowing users to control how the model processes input data, applies Bayesian inference, and utilises marketing data to optimise results.

3.1. *Configuration File Example*

Below is an example of a YAML configuration file. Users should modify this template to suit their specific data and requirements:

```yaml
################################
# General Settings
################################

data_rows:
  total: 171
  start_date: 2019-07-28
  end_date: 2022-10-30

raw_data_granularity: weekly  # Temporal resolution: daily or weekly
train_test_ratio: 1.0         # Train-test split (default: 1.0 for full training)

################################
# Column Definitions
################################

# Columns to ignore during analysis
ignore_cols:
  - "price"
  - "media_imp_5"
  - "media_cost_5"

# Key columns for modelling
date_col: "date"              # Date column in ISO 8601 format (YYYY-MM-DD)
target_col: "subscribers"     # Target metric to optimise
target_type: "conversion"     # Type of target: revenue or conversion

# External features influencing the target metric
extra_features_cols:
  - "covid_index"
  - "competitor_spend"
extra_features_impact:
  "competitor_spend": "negative"

# Media channel definitions
media:
  - display_name: "Media Channel 1"
    impressions_col: media_imp_1
    spend_col: media_cost_1
  - display_name: "Media Channel 2"
    impressions_col: media_imp_2
    spend_col: media_cost_2

################################
# Model Parameters
################################

# MCMC sampling parameters
tune: 2000
draws: 2000
chains: 4
ad_stock_max_lag: 8
target_accept: 0.95
```

```yaml
# Prophet integration for seasonality and holidays
prophet:
  include_holidays: true
  holiday_country: 'US'
  yearly_seasonality: true
  weekly_seasonality: true

# Seed for reproducibility
seed: 42
```

3.2. Configuration Details

The `config.yaml` file is a flexible and comprehensive tool that allows users to customise every aspect of the modelling process. Below are detailed explanations of its key components.

3.2.1. Data Handling

- **raw_data_granularity**: Defines the temporal resolution of the input data (`daily` or `weekly`). This parameter ensures that the model accurately interprets time-series relationships.
- **train_test_ratio**: Specifies the proportion of data used for training versus testing. A value of `1.0` uses the entire dataset for training, which is useful for forecasting without immediate validation.

3.2.2. Column Definitions

- **ignore_cols**: Lists columns to exclude from analysis. These may include irrelevant features or confounding variables.
- **date_col**: Specifies the date column, formatted as `YYYY-MM-DD`. This is a required field for time-series data.
- **target_col** and **target_type**: `target_col` defines the dependent variable (e.g., `subscribers`), while `target_type` specifies its nature (`revenue` or `conversion`).
- **extra_features_cols** and **extra_features_impact**: These represent external factors (e.g., economic indicators) that influence the target metric. The `extra_features_impact` field indicates whether these features have a `positive` or `negative` relationship with the target.

3.2.3. Media Channels

- Each media channel is defined under the `media` section, with the following fields:
  - `display_name`: A descriptive name for the channel.
  - `impressions_col`: The column for impressions data (optional).
  - `spend_col`: The column for spend data (required).
  This structure enables the model to process multiple advertising channels, distinguishing between them based on their unique attributes.

3.2.4. Model Parameters

- **tune** and **draws**: Control the number of samples in the Markov Chain Monte Carlo (MCMC) process. Higher values improve model convergence but increase computational requirements.
- **ad_stock_max_lag**: Sets the maximum lag for adstock effects, capturing delayed impacts of marketing activities.
- **prophet**: Configures seasonality, trend, and holiday effects using Facebook's Prophet. Options include:
  - `holiday_country`: Specifies the region for public holidays.
  - `yearly_seasonality`, `weekly_seasonality`, and `trend`: Enable these components to reflect regular fluctuations and long-term trends.

3.2.5. Custom Priors

- The custom_priors section allows users to define prior distributions for key parameters. This feature is ideal for incorporating domain knowledge or biases about expected media performance.
- Examples:
  - intercept: A LogNormal prior can define baseline behaviour.
  - adstock_alpha: A Beta prior can set expectations for adstock decay.
  - saturation_beta: Adjusts diminishing returns behaviour for media spend using a Gamma distribution.

3.3. Updating Configuration

Griffin MMM's YAML-based configuration makes updates straightforward, enabling users to adapt the model for new campaigns or data sources. Common updates include:

- Adding new media channels under the media section.
- Modifying seasonality settings (e.g., updating holiday_country for region-specific holidays).
- Adjusting priors based on new insights or changes in expected channel performance.

The flexibility of config.yaml ensures that Griffin MMM remains adaptable, making it suitable for dynamic and evolving marketing environments.

4. Data Handling and Preprocessing

4.1. Loading Data

Griffin MMM requires input data in CSV format. This dataset forms the foundation for media mix modelling and must include the following columns:

- date: The date for each observation, formatted in ISO 8601 (YYYY-MM-DD).
- impressions: Media impressions per channel, quantifying the reach of advertisements across different channels.
- spend: Media spend per channel, indicating the budget allocated to each channel over time.
- target: The dependent variable (e.g., sales or conversions) that the model aims to predict or optimise.
- extra features: Optional columns capturing external factors such as economic indicators, competitor activity, or promotional events, which may influence the target metric.

Ensure these columns are properly named and formatted to match the mappings in config.yaml. Perform initial checks to confirm the dataset is complete, with no missing or misformatted date entries.

4.2. Data Transformations

Griffin MMM applies several transformations to prepare the data for modelling, enhancing performance and accuracy.

4.2.1. Normalisation and Scaling

Media spend and impressions often differ in scale, which can hinder model convergence. Griffin MMM standardises data using:

- **Standard Scaling**: Transforms data to have a mean of zero and a standard deviation of one, minimising the influence of outliers.
- **Min-Max Scaling**: Rescales data into a range (e.g., 0 to 1), particularly effective for non-Gaussian distributions.

4.2.2. Handling Missing Data

Incomplete datasets can reduce model effectiveness. Griffin MMM offers strategies to address missing values:

- **Mean/Median Imputation**: Fills missing values with the mean or median of the column, depending on the data distribution.
- **Dropping Missing Values**: Removes rows with missing entries if their occurrence is minimal, ensuring the dataset remains robust.

4.2.3. Feature Engineering

Griffin MMM supports creating additional features to enhance model performance:

- **Lag Features**: Captures delayed effects by including lagged versions of impressions or spend.
- **Interaction Terms**: Highlights synergies between channels (e.g., TV and digital spend interactions).
- **External Variables**: Incorporates factors like weather, competitor activity, or macroeconomic indicators to differentiate marketing-driven effects from external influences.

4.2.4. Seasonality Adjustments

Griffin MMM accounts for seasonal patterns that influence consumer behaviour:

- **Yearly Seasonality**: Adjusts for annual patterns such as holidays or sales events.
- **Weekly Seasonality**: Captures day-of-the-week trends (e.g., increased weekend activity).
- **Prophet Integration**: Adds advanced seasonality and trend components, decomposing time series into seasonal, trend, and residual parts.

4.2.5. Adstock Transformation

Griffin MMM uses adstock transformations to model the delayed effects of marketing activities. A decay function captures the persistence of media impact over time:

- The `ad_stock_max_lag` parameter defines the maximum time lag.
- The decay rate is controlled through the `adstock_alpha` prior, which can be customised based on empirical data or domain knowledge.

4.2.6. Saturation Effects

The model incorporates saturation effects, recognising diminishing returns at higher spend levels:

- Logistic and exponential functions model non-linear relationships between spend and impact.
- The `saturation_beta` and `saturation_lam` parameters control the shape and rate of saturation for each channel.

4.3. Data Validation

Griffin MMM performs rigorous validation checks to ensure data readiness:

- **Date Consistency**: Verifies dates are continuous and correctly formatted, with no gaps in the time series.
- **Value Checks**: Confirms that spend, impressions, and target values are non-negative, avoiding potential errors in model output.
- **Column Verification**: Ensures all columns specified in `config.yaml` are present in the dataset, preventing runtime errors.

Data handling and preprocessing are critical to achieving high-quality model outputs. By following these guidelines and leveraging Griffin MMM's built-in transformation and validation features, users can ensure reliable and actionable insights.

5. Core Modeling

Griffin MMM leverages advanced modelling techniques to represent real-world marketing dynamics, including adstock and saturation transformations. These features provide a nuanced understanding of how marketing spend impacts business outcomes, enabling marketers to optimise budgets effectively.

5.1. Adstock Transformations

Adstock transformations model the delayed and cumulative effects of marketing activities, recognising that advertising does not yield immediate and complete returns. Instead, its impact decays gradually over time, influencing consumer behaviour even after a campaign ends.

The transformation relies on two key parameters:

- **Decay Rate** (`adstock_alpha`): Represents how quickly the advertising effect diminishes over time. Values close to 1 indicate a prolonged impact, while values near 0 suggest a rapid decline.
- **Maximum Lag Period** (`ad_stock_max_lag`): Defines the maximum number of time periods over which the advertising effect persists.

The adstock effect at time t is calculated as:

$$A_t = X_t + \alpha A_{t-1} \tag{1}$$

where:

- A_t: Adstock effect at time t.
- X_t: Media spend at time t.
- α: Decay factor controlling carryover.

Adstock transformations are critical for accurately capturing the long-term impact of campaigns, particularly for channels like TV, where effects may persist for weeks. Users can customise `adstock_alpha` and `ad_stock_max_lag` to reflect empirical insights or domain expertise.

5.2. Saturation Transformations

Saturation transformations address diminishing returns, where additional marketing spend yields progressively smaller impacts. Griffin MMM uses non-linear functions to represent this phenomenon, ensuring realistic modelling of spending dynamics.

5.2.1. Logistic Saturation Function

The logistic function is commonly used to model saturation effects:

$$S(x) = \frac{\lambda}{1 + e^{-\beta(x-\theta)}} \tag{2}$$

where:

- $S(x)$: Saturation-adjusted effectiveness.
- λ: Maximum potential impact (upper limit).
- β: Growth rate, controlling the speed of saturation.
- θ: Midpoint, indicating when saturation becomes significant.

5.2.2. *Implementation in Griffin MMM*

Griffin MMM allows users to define unique saturation parameters for each media channel, reflecting their individual behaviours. For instance, digital channels may saturate faster than traditional media. Parameters such as `saturation_beta` and `saturation_lam` can be adjusted in the configuration file to account for these differences.

5.3. *Model Structure and Bayesian Inference*

Griffin MMM employs Bayesian inference to estimate model parameters, offering a robust framework for quantifying uncertainty. The Markov Chain Monte Carlo (MCMC) method is used to sample from posterior distributions, providing a probabilistic understanding of channel effects.

5.4. *Markov Chain Monte Carlo (MCMC) Sampling*

MCMC sampling generates posterior distributions of parameters, enabling comprehensive insights into their likely range. Two key phases govern the process:

- **Tuning Phase** (`tune`): The sampler adjusts its parameters to explore the space effectively. Higher tuning values (e.g., `2000`) improve convergence for complex models.
- **Drawing Samples** (`draws`): Once tuned, the sampler generates samples from the posterior distribution. These samples underpin model predictions and channel effect estimates.

5.5. *Base Model*

The base model integrates adstock, saturation, and Bayesian inference to estimate the effectiveness of each marketing channel. By providing a properly configured `config.yaml` file and formatted data, users can generate outputs including:

- **Channel Contributions**: Quantifies each channel's impact on the target metric.
- **Effectiveness Estimates**: Assesses ROI and the efficiency of marketing spend.
- **Saturation Insights**: Highlights diminishing returns for channels nearing their performance limits.

These core modelling capabilities empower marketers to make data-driven decisions, balancing immediate and long-term effects while accounting for diminishing returns. Griffin MMM's flexibility and rigorous statistical foundation ensure reliable insights for budget optimisation.

6. Validation

Griffin MMM includes robust validation utilities to ensure data consistency, accuracy, and suitability for model training. Effective validation helps catch errors early, ensuring that the input data is clean, reliable, and structured to maximise model performance.

6.1. *Validation Functions*

Griffin MMM performs automated checks across key data components to guarantee proper formatting and minimise issues that could compromise model outputs.

6.1.1. *Target Variable Validation*

The target variable is critical for model training. Griffin MMM ensures:

- **Non-Empty Target Column**: Verifies that the target metric (e.g., sales or conversions) has no missing entries. Missing values are either imputed or flagged for removal.
- **Statistical Summary Checks**: Generates a summary of the target variable to identify anomalies such as outliers or zero-variance scenarios that could skew predictions.

6.1.2. Date Column Validation

Time-series data requires a continuous and correctly formatted date column. Griffin MMM validates:

- **Uniqueness of Dates**: Ensures no duplicate date entries, which can disrupt time-series calculations.
- **Date Format and Continuity**: Confirms that dates follow the `YYYY-MM-DD` ISO format and checks for gaps in the sequence.
- **Handling Missing Dates**: Detects gaps and offers options for interpolation or manual correction.

6.1.3. Media Spend and Impressions Validation

Media spend and impressions are core drivers of the model. Griffin MMM performs:

- **Non-Negative Values**: Ensures all spend and impressions values are non-negative, flagging potential data entry errors.
- **Range Checks**: Identifies unusually high or low values for further inspection, avoiding unrealistic inputs that could distort results.
- **Missing Data Handling**: Addresses missing values through forward-filling or interpolation to maintain temporal consistency.

6.1.4. Extra Features Validation

External factors such as economic indicators and promotions are validated to ensure relevance:

- **Impact Definition**: Confirms that each feature's expected impact (`positive` or `negative`) aligns with the model configuration.
- **Correlation Analysis**: Evaluates correlations between extra features and the target variable, excluding features with minimal relevance to improve model parsimony.

6.1.5. Cross-Validation Setup

Cross-validation ensures the robustness of model results. Griffin MMM checks:

- **Train-Test Splitting**: Validates that the training set is sufficiently large for parameter estimation based on the `train_test_ratio`.
- **Temporal Consistency**: Maintains chronological order by ensuring training data precedes testing data, preventing information leakage.

6.1.6. Priors and Hyperparameter Validation

Custom priors and hyperparameters are validated to ensure appropriate settings:

- **Priors Consistency Check**: Verifies that priors match the expected conditions for their distributions (e.g., positive parameters for `Beta` distributions).
- **Hyperparameter Bounds**: Confirms that hyperparameters, such as `adstock_alpha` and `saturation_beta`, fall within realistic ranges to prevent instability during training.

6.1.7. Visual Validation Reports

Griffin MMM generates visual summaries to highlight potential issues:

- **Missing Data Heatmaps**: Visualise patterns of missing data across features to guide corrective actions.
- **Distribution Plots**: Display distributions of target, spend, and impressions to identify outliers or irregularities.

6.2. *Ensuring Data Suitability*

These validation checks ensure that data is properly structured, reducing the likelihood of errors during training and enhancing model reliability. By incorporating rigorous validation, Griffin MMM provides users with confidence in the accuracy and quality of model outputs, supporting informed and data-driven marketing decisions.

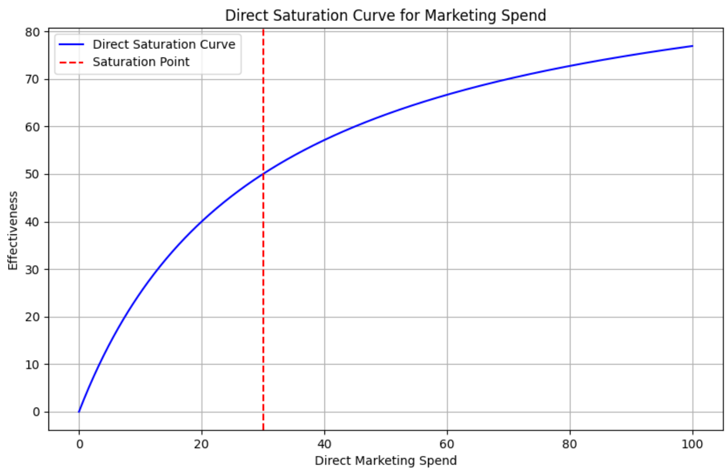
7. Visualization

Visualizing the model results is essential for interpreting and communicating the insights derived from Griffin MMM. Proper visual representation allows marketers to better understand the relationships between media spend and business outcomes, thereby informing strategic decision-making regarding budget allocation and channel effectiveness.

7.1. Plotting Saturation Curves

Saturation curves are a core visual output of Griffin MMM, providing insight into the concept of diminishing returns on marketing investment. Griffin MMM generates these curves for each media channel, illustrating how the effectiveness of marketing efforts changes as spend increases. This visualisation allows users to identify the optimal point of spend – the point where further increases yield little to no additional impact, signalling saturation.

Figure 1: Example Saturation Curve Showing Diminishing Returns



The plot in Figure 1 shows a typical saturation effect where effectiveness grows rapidly at lower levels of spend but begins to plateau as the channel becomes saturated. The curve typically follows an S-shape (sigmoid) or a logarithmic form, depending on the channel dynamics and market conditions. The point at which the curve flattens indicates diminishing returns, meaning that additional spend no longer provides proportional gains in output.

7.1.1. Technical Details of Saturation Curve Calculation

Griffin MMM uses a combination of non-linear transformations, including logistic or exponential functions, to represent saturation. The calculation takes into account:

- **Saturation Parameters**: The `saturation_beta` and `saturation_lam` parameters in the configuration file control the rate and shape of the saturation effect for each media channel. These parameters are estimated as part of the Bayesian inference process, allowing the model to capture how quickly each channel reaches its maximum effectiveness.
- **Interaction with Spend**: The saturation curve is derived by applying these transformations to the adstocked spend data. This means that the plotted curves incorporate both the delayed effect (through adstock) and the non-linear returns (through saturation), giving a holistic view of the channel's performance.

The flexibility of saturation modelling in Griffin MMM allows users to adapt the shape of the curve based on empirical insights or to allow the model to learn the saturation behaviour directly from the data.

7.2. Plotting Adstock Effects

Visualising adstock effects helps illustrate the temporal influence of marketing spend. Griffin MMM generates adstock plots that show how the impact of a media spend decays over time, capturing the lingering effect of advertising even after the initial campaign ends.

Figure 2: Example Adstock Effect Showing Delayed Impact of Media Spend

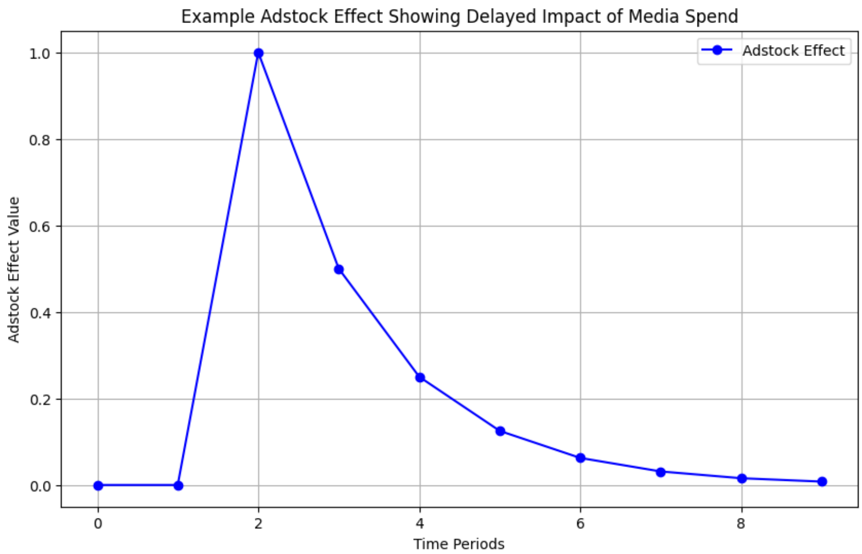


Figure 2 represents the adstock effect for a particular media channel. The curve illustrates how the initial impact of a campaign diminishes gradually rather than dropping off instantly, reflecting real-world consumer response behaviours. By understanding these lagged effects, users can better time their media investments, especially for channels where delayed responses are prominent (e.g., TV and radio).

7.3. Lift Test Analysis Plots

Lift test analysis plots are used to evaluate the incremental impact of specific marketing campaigns or interventions. These plots provide a visual representation of the treatment versus control performance, highlighting the effectiveness of a given marketing effort.

Figure 3 shows an example of a lift test analysis. The difference between the control and treatment lines represents the incremental gain achieved due to the intervention. Shaded areas around the lines indicate confidence intervals, providing a sense of the uncertainty around the estimates.

7.4. Channel Contribution Visualisation

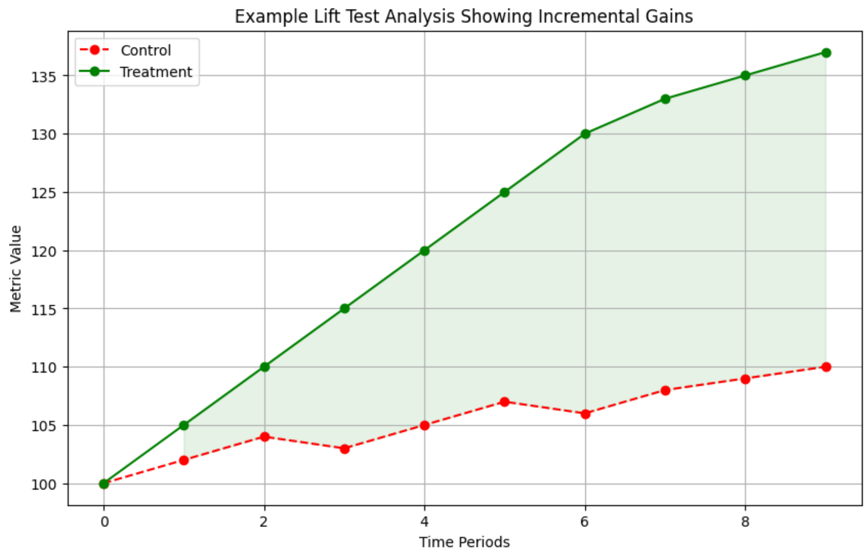
Griffin MMM also provides visualisations that display the contribution of each media channel to the overall target metric (e.g., total sales). These stacked bar charts or line graphs help break down the aggregate impact of all channels into individual contributions, making it easier for marketers to understand which channels are driving performance.

7.5. Stacked Bar Chart of Channel Contributions

The stacked bar chart visualises how much each media channel contributes to the target metric over time. This chart is particularly useful for seeing how different channels perform during specific periods, such as product launches or promotional campaigns.

Figure 4 demonstrates how contributions from various channels evolve over time. The visualisation helps pinpoint which channels are providing the most value and when their impact peaks, enabling marketers to optimise spend allocation.

Figure 3: Example Lift Test Analysis Showing Incremental Gains



7.6. *Uncertainty and Posterior Distributions*

One of the advantages of using a Bayesian framework is the ability to visualise uncertainty in the model's predictions. Griffin MMM generates plots that show the posterior distributions of key parameters, such as the return on investment (ROI) for each channel.

7.7. *Posterior Distribution Plots*

Posterior distribution plots help users understand the range and uncertainty associated with each estimated parameter:

- **ROI Distributions**: These plots show the distribution of ROI estimates for each media channel, providing insights into the variability and reliability of the returns. Channels with narrow distributions indicate more certainty around their effectiveness, while broader distributions suggest more variability.
- **Parameter Uncertainty**: Posterior plots for parameters like `adstock_alpha` or `saturation_beta` allow users to see the range of likely values for these parameters, which is crucial for understanding the potential impact of different modelling assumptions.

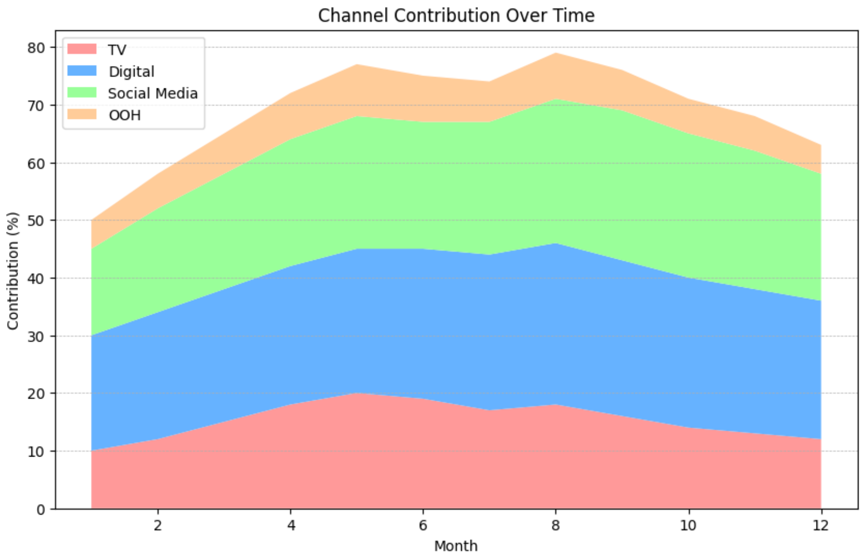
Figure 5 provides an example of a posterior distribution for media channel ROI. The distribution captures the full range of possible values, allowing marketers to make more informed decisions that take uncertainty into account.

7.8. *Practical Uses of Visualizations*

Visualisations generated by Griffin MMM serve several practical purposes:

- **Budget Allocation**: By understanding saturation points and channel contributions, marketers can make informed decisions about where to allocate their marketing budgets to maximise ROI.
- **Stakeholder Communication**: Complex modelling outputs are translated into intuitive visual formats, making it easier to communicate results and recommendations to stakeholders who may not have a technical background.

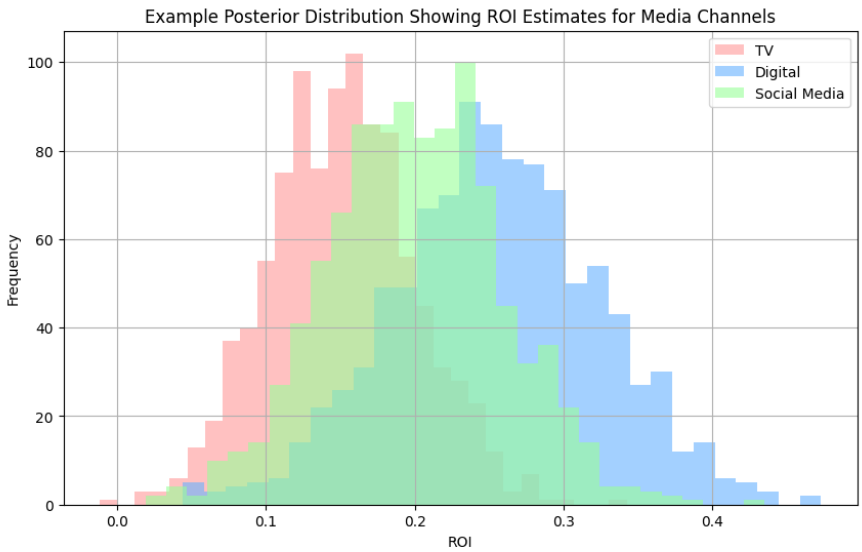
Figure 4: Example Stacked Bar Chart of Media Channel Contributions



- **Campaign Planning**: Adstock and lift visualisations help determine the timing and duration of future campaigns based on observed lag effects and measured incremental gains.

Effective visualisation is crucial to extracting and communicating actionable insights from the Griffin MMM model. By providing a wide range of plots, Griffin MMM ensures that both technical and non-technical stakeholders can derive value from the results, ultimately leading to better marketing decisions and improved business outcomes.

Figure 5: Example Posterior Distribution Showing ROI Estimates for Media Channels



8. Seasonality

8.1. Prophet Integration

Griffin MMM integrates with Facebook's Prophet, a powerful forecasting tool designed to handle time series data that includes seasonal effects and holiday influences. Prophet is particularly well-suited for business data that exhibits strong seasonal trends and periodic changes, such as weekly and yearly variations in consumer behaviour. By leveraging Prophet, Griffin MMM is able to provide more accurate and realistic estimates by accounting for temporal patterns that impact marketing effectiveness.

The integration with Prophet allows users to:

- **Incorporate Seasonality**: Adding yearly, weekly, and even daily seasonal components helps the model to better capture predictable patterns that occur at regular intervals, thus improving the overall predictive accuracy.
- **Adjust for Holidays and Special Events**: Many industries see significant variations in customer activity during holidays or special events. Griffin MMM, with Prophet's holiday integration, allows users to model these effects, accounting for spikes or drops in consumer behaviour during these periods.
- **Detect Trends**: Prophet helps capture underlying trends beyond seasonal fluctuations, such as gradual increases or decreases in consumer activity, which may be due to market growth, changing consumer preferences, or other macroeconomic factors.

Prophet's flexibility allows Griffin MMM to dynamically adjust seasonal components, resulting in a model that adapts to changing business conditions and provides a more nuanced understanding of marketing performance across different time horizons.

8.2. Adding Seasonality Components

Griffin MMM users can add seasonality components like yearly and weekly trends through the `config.yaml` file. This enables a more fine-tuned adjustment for recurring patterns that might significantly affect marketing performance. Below, we explore the specifics of configuring and using these seasonality components.

8.2.1. Yearly and Weekly Seasonality

Adding yearly and weekly seasonality helps Griffin MMM account for repeating patterns observed over these periods. The `prophet` section in the configuration file allows users to include seasonality components:

```
prophet:
  include_holidays: true
  holiday_country: 'US'
  yearly_seasonality: true
  trend: true
  weekly_seasonality: true
```

- **Yearly Seasonality**: Yearly seasonality accounts for annual patterns, such as increased sales during holidays or reduced marketing efficiency during off-peak seasons. These seasonal components help explain variance in the target metric that aligns with recurring annual events, such as Christmas, summer holidays, or fiscal year-end boosts in corporate spending.
- **Weekly Seasonality**: Weekly seasonality helps capture consumer behaviours that vary by day of the week, such as increased activity on weekends or higher conversion rates during weekdays. Including this component is especially helpful for businesses where customer engagement is known to be cyclical within a week.

These seasonal adjustments are applied during the preprocessing phase, and their effects are incorporated into the model's baseline trends. Including such components helps isolate the effects of marketing spend from predictable temporal variations, thereby providing a more accurate estimate of marketing-driven uplift.

8.2.2. Holiday Effects

Holiday effects can have a major impact on marketing performance, often causing spikes or drops in consumer engagement that must be accounted for to avoid skewed results:

- **Holiday Integration**: The `include_holidays` option allows Griffin MMM to automatically adjust for major holidays in the specified `holiday_country`. This is especially useful for global brands running campaigns across different markets, each with unique holiday schedules.
- **Custom Holiday Definitions**: In addition to built-in holidays, users can also define custom events that are significant for their business (e.g., product launches, promotional events). These custom holidays can be added to the Prophet configuration, allowing for bespoke seasonal adjustments that reflect unique business dynamics.

Including holiday effects ensures that the spikes or troughs in the time series caused by external factors do not wrongly influence the estimated effectiveness of marketing activities. This improves the reliability of the model by distinguishing between seasonality-driven and spend-driven variations in the target metric.

8.3. Advanced Trend Modelling with Prophet

In addition to seasonality, Prophet provides advanced trend modelling capabilities that can be leveraged by Griffin MMM to capture non-linear trends and structural changes in the target metric:

8.3.1. Additive and Multiplicative Seasonality

Griffin MMM allows users to choose between additive and multiplicative models of seasonality, depending on how they expect the seasonal effects to interact with the overall level of the target metric:

- **Additive Seasonality**: This is suitable when the magnitude of the seasonal effect does not depend on the level of the target metric. For example, a consistent seasonal boost of 10 units during a particular month regardless of the base level of the metric.

- **Multiplicative Seasonality**: This is used when the seasonal impact scales with the level of the target metric. For instance, a 10% increase in sales during a holiday season that scales with the overall baseline.

Choosing the right type of seasonality is crucial for accurately capturing the nature of recurring trends in the data.

8.3.2. Changepoint Detection

Prophet has built-in changepoint detection that is integrated into Griffin MMM to help identify moments in the time series when the trend shifts significantly. These changepoints can be caused by a variety of factors, such as market shifts, regulatory changes, or significant marketing interventions:

- **Automatic Changepoint Detection**: Prophet automatically identifies potential changepoints where the time series trajectory alters significantly. These changepoints are included in the model, allowing it to better fit the data around periods of structural change.
- **Manual Changepoint Specification**: Users can also manually specify changepoints if they have prior knowledge of significant events. This feature is particularly useful for incorporating known disruptions, such as competitor launches or major internal changes, directly into the model.

Changepoint detection helps Griffin MMM adapt to significant shifts in the underlying trend, providing a more flexible and accurate representation of how marketing activities are influencing the target metric over time.

8.4. Practical Applications of Prophet Integration

The integration of Prophet with Griffin MMM offers several practical benefits for marketers:

- **Forecasting with Seasonality and Trend Components**: The combined power of MMM and Prophet allows marketers to not only measure historical effectiveness but also forecast future performance by projecting seasonal and trend components forward in time.
- **Campaign Planning**: By understanding the expected seasonal and trend fluctuations, marketers can better plan their campaigns to take advantage of periods with high baseline activity or avoid times when customer interest is predicted to be low.
- **Scenario Analysis**: Users can simulate the impact of upcoming holidays or expected seasonal peaks and troughs to determine how best to allocate marketing spend in the future, ultimately helping to optimise campaign timing and effectiveness.

Prophet integration provides Griffin MMM with a robust framework to adjust for temporal complexities, ensuring that seasonality and other recurring patterns are correctly accounted for. This results in a more accurate attribution of marketing impact, supporting data-driven decision-making for budget allocation and campaign optimisation.

9. Example Workflow

To run a complete Griffin MMM model, follow these steps:

1. **Prepare Your Data**: Load your data into the correct CSV format, ensuring it matches the expected structure. The dataset should include columns for `date`, media spend per channel, impressions (if available), extra features (such as economic indicators or promotional events), and the target metric (e.g., sales or conversions). Each column must be clearly mapped in the `config.yaml` file to ensure proper preprocessing and model input.
2. **Configure the Model**: Set parameters in the `config.yaml` file, including channels, target variables, and time granularity. Make sure to:
   - Define `media` channels with their respective `spend_col` and `impressions_col`.
   - Set `adstock_alpha` and `saturation_beta` for each channel to control the decay and saturation behaviours.
   - Include seasonality and holiday components under the `prophet` section to improve accuracy.
   - Specify custom priors to reflect domain knowledge or adjust hyperparameters for specific campaigns.
3. **Preprocess the Data**: Run the preprocessing utilities provided by Griffin MMM to handle missing values, scale data appropriately, and add any lagged or interaction terms. Proper preprocessing ensures that the model receives data in a format that maximises the efficiency of Bayesian inference.
4. **Fit the Model**: Execute the model run using the provided tools. Griffin MMM will utilise the configured parameters and input data to perform Bayesian analysis of marketing effectiveness. The MCMC sampling process will estimate the impact of each channel while also quantifying uncertainty. During this phase, the model uses the specified `tune` and `draws` parameters to ensure robust parameter convergence.
5. **Evaluate and Visualize Results**: Utilize the visual outputs generated by Griffin MMM to gain insights into channel performance and effectiveness:
   - **Saturation Curves**: Plot saturation curves for each channel to understand diminishing returns and identify optimal spend levels.
   - **Adstock Effects**: Visualise adstock curves to assess the delayed impact of marketing activities.
   - **Lift Analysis**: Perform lift testing to validate the incremental impact of specific marketing interventions.
   - **Channel Contribution**: Review channel contribution graphs to see how different channels contribute to overall target metric performance over time.
6. **Adjust and Iterate**: Based on the evaluation, adjust the model configuration and repeat the workflow as necessary. Iteration is key to refining the accuracy of the model - tweak hyperparameters, adjust priors, or modify seasonality settings to achieve better alignment with observed results.
7. **Report Insights**: Once the model is satisfactorily tuned and validated, generate a report summarising the key insights. Include visualisations, ROI estimates, and performance comparisons across channels. This report can then be shared with stakeholders to inform strategic decisions regarding future marketing investments.

The example workflow provided above is iterative, meaning that marketers and analysts should expect to refine and re-run the model multiple times to optimise performance. Each iteration allows for improved parameter tuning, better data handling, and ultimately a more precise understanding of how marketing investments impact business outcomes.

10. API

10.1. API Overview

Griffin MMM provides a set of methods and tools internally to facilitate the transformation, modelling, and visualisation of marketing data. Advanced users who wish to extend the functionality of Griffin MMM or integrate it into other workflows can access the API for greater control over each modelling step. Below is an overview of key components:

Griffin MMM exposes several core classes and functions that advanced users can interact with directly:

- **Data Handling API**: Methods for loading and validating data are part of the `InputData` class. Functions include `validate_data()`, `transform_input_generic()`, and `load_csv()`. These functions help ensure data readiness for modelling.
- **Model Configuration and Setup**: The `MMM` class is used for initiating model fitting. It accepts user-provided configuration parameters from `config.yaml` and is responsible for setting up model priors, hyperparameters, and constraints.
- **Custom Transformations and Features**: Users can apply custom transformations on input data using the `transformers` module. This module includes predefined adstock, decay, and interaction transformations that can be applied to media data to enrich the feature set.
- **Bayesian Inference Methods**: The `fit()` method leverages PyMC's MCMC capabilities to estimate model parameters. This method includes options for specifying tuning iterations, draw counts, and chains to achieve convergence.
- **Visualisation Tools**: The `plot()` functions from the `plot.py` module generate key visual insights, including saturation curves, adstock effects, and channel contributions, making it easier for advanced users to incorporate these visuals into custom reporting.

For advanced users, detailed API documentation can be requested, which includes information on how to extend existing classes, customise priors, and modify the internal behaviour of the MMM model.

10.2. Common Issues and Troubleshooting

Griffin MMM is designed to handle a wide variety of input data and configurations, but users may still encounter issues during setup and modelling. This section provides a guide to addressing common problems:

10.2.1. Data-Related Issues

- **Missing Date Column**: Ensure that the `date_col` specified in `config.yaml` matches the actual date column in your dataset. Dates should be formatted in ISO 8601 (`YYYY-MM-DD`) to ensure compatibility with Prophet and time series processing.
- **Negative Channel Values**: Verify that all media spend and impressions values are non-negative. Negative values can indicate data entry errors and lead to inaccurate or non-convergent models. Use the preprocessing utilities in `transform.py` to clean and standardise such data before model fitting.
- **Duplicate Dates**: The date column must contain unique values for each time period. Duplicate entries will cause inconsistencies in time series processing and result in errors during the model run. Griffin MMM provides a validation utility that automatically checks for duplicate date entries.
- **Handling Missing Values**: Missing data in key columns can negatively affect model performance. Griffin MMM offers several imputation methods, such as forward-filling or interpolation, to handle these missing values. Users can specify imputation strategies in the `config.yaml` file or preprocess the data manually.

10.2.2. Configuration Errors

- **Incorrect Media Channel Configuration**: Each media channel must be properly defined with `spend_col` and `impressions_col` in `config.yaml`. Omitting either of these definitions can lead to errors during the model's adstock and saturation transformations. Double-check the configuration file for missing or incorrectly named attributes.
- **Priors and Hyperparameters**: Incorrect specification of priors or unrealistic hyperparameter values can result in poor model convergence or extremely high/low posterior estimates. It is advisable to start with default priors and gradually adjust them based on the data. If the model fails to converge, consider simplifying the priors by using distributions with broader support, such as switching from `Beta` to `Uniform` distributions.
- **Prophet Integration Issues**: If Prophet integration fails, check that the `prophet` options in `config.yaml` are correctly configured. Ensure that `include_holidays` is set to a valid country code, and that no conflicting seasonality options are specified (e.g., using both yearly and custom seasonalities incorrectly).

10.2.3. Model Fitting Challenges

- **Non-Convergence of MCMC Chains**: Bayesian modelling relies on MCMC sampling, which may sometimes fail to converge. This can be due to improper prior specifications, a small number of `tune` or `draws`, or highly collinear features. Increase the number of `tune` steps and use diagnostic tools like trace plots to verify that all chains converge to the same posterior distribution.
- **Runtime Performance**: Model fitting may be slow, especially for datasets with high granularity (e.g., daily data with hundreds of features). Consider reducing the number of features by excluding less relevant ones or aggregating data to a weekly level. Optimising hyperparameters such as `chains` and `draws` can also help balance accuracy with runtime performance.

10.3. Useful Tips for Efficient Modelling

- **Start with a Simple Model**: Begin with a limited number of channels and features to ensure that the model runs correctly. Once validated, incrementally add more complexity.
- **Use Visual Diagnostics**: Make use of the posterior distribution plots and diagnostic plots (e.g., trace plots) provided by Griffin MMM. These visualisations can help identify problems such as non-convergence, poor mixing of chains, or extreme parameter values.
- **Leverage Cross-Validation**: Use the built-in cross-validation tools to validate the model's predictive performance on out-of-sample data. This is particularly important for ensuring that the model generalises well beyond the training dataset.
- **Custom Priors**: Tailor the priors to fit your specific use case. For example, if you know that a certain channel has historically had a minimal impact, use a tighter prior distribution to reflect this domain knowledge.

By following these guidelines, users can ensure smoother implementation and achieve more reliable marketing insights.

11. Marketing Mix Modelling: Concepts and Methodology

Marketing serves as a vital catalyst for company growth, often representing significant financial commitments. Consequently, assessing the efficiency and fine-tuning the distribution of marketing funds is crucial for those in the field. Marketing Mix Modelling (MMM) has long been a key strategic instrument used by marketers to meet these objectives. Recent shifts towards consumer privacy, such as Apple's IDFA changes in iOS 14, highlight the critical need for marketers to future-proof their measurement strategies using tools like MMM.

While experimental designs and causal analytical methods are frequently applied to infer causality, they may be impractical or prohibitively expensive in certain contexts. MMM provides an alternative by utilizing aggregate time-series data alongside regression analysis to delineate the impact of marketing on sales. Additionally, MMM can be customized to account for variables like seasonal patterns, trends, and other external factors, and to incorporate geographic hierarchies. Importantly, the primary goal of MMM is often to assess the incremental impact of various marketing activities, not just sales forecasting.

Several complexities must be navigated when developing MMM. The rapid evolution of advertising media demands continuous adaptation to new marketing channels, presenting modelers with the "large p, small n" challenge. Moreover, aiming for detailed data granularity for actionable insights can lead to sparse data and anomalies. Modelers are tasked with balancing the availability of reliable historical data against the need for detailed data analysis. The sequential nature of data can also lead to correlated errors, undermining the foundational assumptions of ordinary least squares modelling. Additionally, the frequent practice of basing marketing budgets on projected revenues introduces endogeneity and multicollinearity issues, exacerbating difficulties with channel spend correlation. Notably, the tendency of demand-focused channels to exhibit self-selection bias necessitates careful treatment to avoid overstated results. Furthermore, the complexity of MMM involves considerable investment and diverse stakeholder engagement, raising the stakes for model clarity and interpretability. Lastly, traditional machine learning techniques like cross-validation are not always applicable for MMM parameter tuning and model selection due to data limitations and the potential non-representativeness of holdout samples for forecasting.

MMM models are instrumental in providing insight for future allocations of budgets. The influence of advertising is subject to fluctuations based on seasonality and various other factors, such as competitive, shape, carryover, and lag effects. Among these, the shape effect is particularly crucial for marketers. It illustrates how sales react to varying levels of advertising intensity, a phenomenon also known as the saturation effect. This effect hypothesises that consumer responses tend to plateau when advertising investments are ramped up. A widely accepted notion in this context is that brands typically see minimal responsiveness to low advertising investments, peak responsiveness at moderate investment levels, and diminishing returns at higher levels.

Most current models suggest an S-curve pattern in sales response to increased advertising. Additionally, the competitive effect is key, gauging how a brand's advertising efficiency stacks up against its market competitors. The carryover effect represents the residual impact of advertising beyond its immediate deployment, whereas the lag effect denotes the time delay in consumer response to specific advertising efforts. These effects have been extensively researched for three decades. However, the emergence of online media channels has strengthened our comprehension of consumer interactions with advertisements.

MMM is grounded in well-established economic and statistical theories that provide a framework for understanding how marketing activities influence consumer behaviour and business outcomes. At its core, MMM applies econometric modelling to estimate the relationship between marketing inputs and a desired business metric, such as sales, revenue, or brand awareness. This section delves into the theoretical aspects that underpin MMM, including the foundational principles, types of transformations used, and methodological considerations.

11.1. Econometric Foundations

MMM is based on econometric principles that model the relationship between independent variables (e.g., media spend across various channels) and a dependent variable (e.g., sales). The core econometric approach used is often a form of multiple linear regression, adjusted for complexities such as temporal dependencies, seasonality, and interactions among channels. The Bayesian framework, frequently applied in MMM, allows for the incorporation of prior beliefs and uncertainties, making the model robust to variations in marketing activities and external influences.

The equation for a typical MMM model can be represented as follows:

$$y_t = \alpha + \sum_{m=1}^{M} \beta_m x_{t,m} + \sum_{c=1}^{C} \gamma_c z_{t,c} + \epsilon_t,$$

where:

- y_t: Dependent variable at time t, representing the KPI (e.g., sales).
- β_m: Coefficient representing the effect of marketing channel m.
- $x_{t,m}$: Marketing spend for channel m at time t.
- γ_c: Coefficient for control variable c.
- $z_{t,c}$: Control variables (e.g., price, promotions, external economic indicators).
- ϵ_t: Error term capturing residual variance not explained by the model.

11.2. Adstock and Lag Effects

A critical concept in MMM is the delayed impact of advertising, which is modelled using adstock transformations. Adstock helps represent the cumulative effect of past advertising efforts on the current performance, recognising that marketing impact is not always immediate but can decay over time. This decay is captured using an adstock rate θ, which determines how quickly the effect of advertising diminishes. Mathematically, the adstocked value $x'_{t,m}$ for media channel m is defined recursively as:

$$x'_{t,m} = x_{t,m} + \theta x'_{t-1,m},$$

where $0 \leqslant \theta \leqslant 1$ controls the decay rate.

11.3. Saturation and Diminishing Returns

Another key theoretical aspect of MMM is the concept of diminishing returns, often represented using saturation functions. The relationship between marketing spend and its effect on sales is typically nonlinear; higher levels of spend often yield smaller incremental gains, which is referred to as saturation. The Hill function is commonly used to model this saturation effect:

$$\text{Hill}(x; K, S) = \frac{1}{1 + (x/K)^{-S}},$$

where K is the half-saturation point and S controls the steepness of the curve. This function provides a realistic depiction of how the response to advertising saturates at high levels of spend.

11.4. Hierarchical Bayesian Framework

MMM often employs a hierarchical Bayesian framework to account for the variability across different products, regions, or time periods. This framework allows for the modelling of parameters that vary by group, while also sharing information across groups to improve parameter estimation. The hierarchical structure is represented as follows:

$$\beta_{m,g} \sim \mathcal{N}(\mu_{\beta_m}, \tau^2),$$

where μ_{β_m} represents the average effect of media channel m across all groups, and τ controls the variability of this effect across groups g.

11.5. Endogeneity and Control Variables

Endogeneity, often caused by the correlation between media spend and unobserved factors influencing sales, is a significant challenge in MMM. For example, media spend might increase during periods when sales are expected to be high (e.g., holidays), creating a bias in the estimated effect of advertising. To mitigate this, MMM includes control variables such as pricing, promotions, and macroeconomic indicators. Including these factors helps isolate the true impact of marketing activities on the target metric.

11.6. Model Calibration and Validation

The robustness of an MMM model is ensured through calibration and validation processes. Cross-validation is often used to assess the model's ability to generalise to unseen data, while out-of-sample testing helps validate the model's predictive performance. Due to the temporal nature of the data, MMM models often use time-series cross-validation, which respects the sequential structure of the data.

Calibration of parameters like adstock rate θ and saturation parameters K and S is performed by iteratively adjusting them to minimise the difference between predicted and actual values. Bayesian calibration, which uses prior distributions for these parameters, is particularly effective in scenarios with limited historical data.

12. Key Concepts

12.1. Response Functions and Shape Effects

The shape effect in Media Mix Modelling characterises the non-linear relationship between marketing input and consumer response. We formalise this through multiple response functions:

12.1.1. Hill-Saturation Response

The Hill function, adapted from biochemistry, provides a flexible S-shaped response curve:

$$f(x) = \frac{\sigma}{1 + e^{-\beta(x-\lambda)}} \tag{3}$$

where:

- σ represents the maximum attainable response
- β controls the steepness of the response curve
- λ determines the inflexion point

12.1.2. Michaelis-Menten Kinetics

For diminishing returns scenarios, we employ the Michaelis-Menten function:

$$f(x) = \alpha \cdot \frac{x}{\lambda + x} \tag{4}$$

where:

- α represents the asymptotic maximum effect
- λ is the half-saturation constant

12.2. Saturation Effects and Adstock Transformations

12.2.1. Geometric Adstock

The temporal carryover effect is modelled through geometric adstock:

$$A_t = \sum_{l=0}^{L} \theta^l x_{t-l} \tag{5}$$

where:

- A_t is the adstock value at time t
- θ is the decay parameter $(0 \leqslant \theta \leqslant 1)$
- L is the maximum lag considered
- x_{t-l} represents the marketing input at lag l

12.2.2. *Weibull Adstock*

For more flexible decay patterns, we implement the Weibull adstock:

$$w(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} \tag{6}$$

where:

- k is the shape parameter
- λ is the scale parameter

12.3. *Competitive Effects and Market Dynamics*

12.3.1. *Cross-Channel Elasticity*

We quantify competitive effects through cross-channel elasticity matrices:

$$\eta_{ij} = \frac{\partial \log(y_i)}{\partial \log(x_j)} \tag{7}$$

where:

- η_{ij} represents the elasticity of brand i's response to brand j's activity
- y_i is brand i's performance metric
- x_j is brand j's marketing input

12.3.2. *Market Share Dynamics*

The competitive landscape is modelled through a market share attraction model:

$$s_i = \frac{A_i}{\sum_{j=1}^{n} A_j} \tag{8}$$

where:

- s_i is brand i's market share
- A_i is brand i's market attraction
- n is the number of competitors

12.4. *Bayesian Implementation*

12.4.1. *Prior Specifications*

We employ informative priors for key parameters:

$$\beta \sim \text{LogNormal}(\mu_\beta, \sigma_\beta) \tag{9}$$
$$\lambda \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda) \tag{10}$$
$$\theta \sim \text{Beta}(\alpha_\theta, \beta_\theta) \tag{11}$$

12.4.2. Hierarchical Structure

The model implements a hierarchical structure for channel-specific parameters:

$$\beta_j \sim \text{LogNormal}(\mu_j, \sigma_j) \tag{12}$$
$$\mu_j \sim \text{Normal}(\mu_0, \tau_0) \tag{13}$$
$$\sigma_j \sim \text{HalfNormal}(\nu_0) \tag{14}$$

where j indexes the marketing channels.

12.5. Model Diagnostics

12.5.1. Convergence Assessment

We employ multiple diagnostics:

- Gelman-Rubin statistic ($\hat{R}$)
- Effective sample size (ESS)
- Monte Carlo standard error (MCSE)

12.5.2. Predictive Checks

Model validation includes:

$$\text{ELPD} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta^s) \right) \tag{15}$$

where:

- ELPD is the expected log pointwise predictive density
- S is the number of posterior samples
- θ^s represents the parameter values in sample s