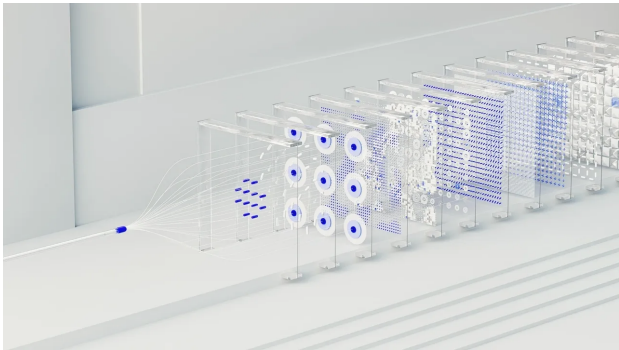


◀ All Posts

A Comprehensive Guide to Bayesian Marketing Mix Modeling

9 May 2023 | Data, Methodologies



As cookies become less reliable, marketers are turning to sophisticated analytical methods to evaluate their campaigns. Marketing Mix Modeling (MMM) has resurfaced as a vital tool for this purpose. This resurgence is driven by influential research and discussions, such as "Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects" (2017) and "Bayesian Media Mix Modeling using PyMC3, for Fun and Profit" (2020). The accessibility of open-source software (OSS) like

Sign up for blog updates!

Join my email list to receive updates and information.

Email address

SIGN UP

Recent Posts

- A De...
12 Mar
2024

Embr...
20 Jan
2024

Opti mi...
8 Jan
2024

PyMC with its specialized package PyMC-Marketing, and Stan, known for its language-agnostic interface, has been pivotal. These OSS tools facilitate Bayesian MMM, making it a preferred methodology for analysts and agencies. Marketing executives are increasingly leveraging Bayesian MMM to scrutinize and optimize their marketing strategies.

Bayesian Marketing Mix Modeling (MMM) is an advanced statistical approach that allows marketing managers and decision-makers to leverage the power of Bayesian statistics in evaluating the performance and effectiveness of their marketing strategies, as well as optimizing their media investments. This comprehensive guide will introduce the key concepts, components, techniques, and benefits of Bayesian MMM, while discussing how to select appropriate prior distributions, likelihood functions, and advanced modeling techniques. These elements help capture the complexities of marketing data and inform better decision-making processes.

The Weather Analogy

We can liken Bayesian Marketing Mix Modeling (MMM) to predicting the probability of rain outside when you have your bedroom curtains closed in the morning. In this analogy, having the curtains closed represents the initial state of uncertainty about the effectiveness of different marketing channels. Since we haven't opened the curtains yet, we have little understanding of the probability of rain. However, given prior experience of the time of year or your location, we may have a prior belief of the probability of it raining today. Within Bayesian Modeling, this is intuitively called the prior. The prior belief represents the probability you attribute to something occurring, or in the context of marketing mix modeling, the probability of the effectiveness of a media campaign.

For instance, imagine that you live in London, and it's April. Based on historical weather data and your experience, you know there's a 40% chance of rain on any given day in London during this month. This can be compared to having prior knowledge about the effectiveness of various marketing channels from previous campaigns or industry benchmarks.

Now, you decide to open the curtains and look outside. You see dark clouds gathering, which provides new evidence about the likelihood of rain. In the context of Bayesian Modeling, this is naturally referred to as the likelihood.

Considering the dark clouds, you now believe there's an 80% chance of rain today. You've updated your initial prior about the probability of rain by combining your prior knowledge (40% chance) with this new information (80% chance). This process is akin to how Bayesian MMM updates its probability distributions based on new sales, marketing data, and external factors.

After performing Bayesian inference, you have a better understanding of the probability of rain, allowing you to make more informed decisions, such as whether to take an umbrella or not. This is referred to as the posterior. The posterior represents the outcome probability of rainfall, considering both the domain expertise and the data. The posterior is a powerful component of Bayesian models, as it includes the full uncertainty of the event occurring. In essence, it represents how sure you are of that outcome

happening. Similarly, Bayesian MMM provides marketing decision-makers with probability distributions that capture the uncertainty of marketing channel effectiveness, enabling them to make better-informed decisions about their marketing strategies and the potential magnitude of impact from their media investments.

The Prior

In Bayesian statistics, we can adjust our priors to explicitly express the confidence of a result. For instance, you can be very confident about a result or have no insight at all. Take the weather analogy; we may have been very certain about the probability of rainfall being 40%, or alternatively may have had a broader range of prior beliefs about the probability of rainfall being between 30%-50%. We categorize these priors as informative or uninformative, and they can take various shapes. The choice of prior can be crucial in Bayesian analysis, as it can influence the outcome of the posterior probability distribution. However, one major benefit of Bayesian Marketing Mix Modeling is that it enforces

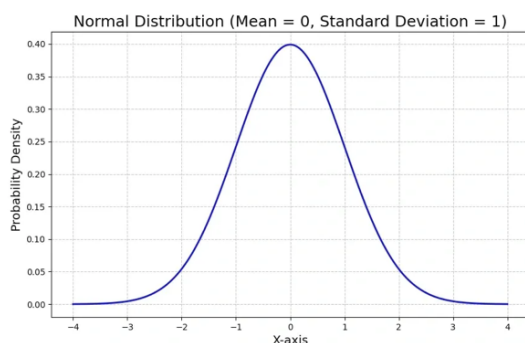
transparency in your subjectivity and biases – which, when shared across a company, can be a powerful tool for aligning across stakeholders and collaboration. The types of prior can be broken down into two broad categories:

- **Informative priors:** These priors incorporate existing knowledge about the parameters, which could be derived from previous studies, domain expertise, or industry benchmarks. Informative priors are useful when there is a reasonable amount of certainty regarding the parameters.
- **Uninformative priors:** These priors express minimal or no prior knowledge about the parameters. They are used when there is a lack of information or when the analyst wants to avoid introducing any biases from prior beliefs. Uninformative priors typically have a flat or uniform distribution, assigning equal probability to all possible values of the parameters.

Additionally, priors may also take different shapes. The list of potential probability distributions is

very extensive; however, in marketing mix modeling, there are a few core set of priors which are particularly useful:

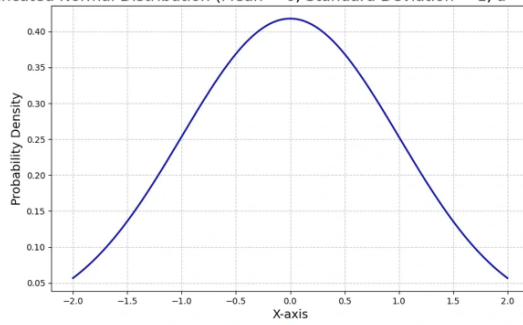
- **Normal distribution:** This is a continuous probability distribution that is symmetric around its mean. Normal priors are often used for parameters where the prior belief is centered around a particular value, with the possibility of variation on either side. These are most commonly used for regression coefficients.



Graph of a normal distribution prior

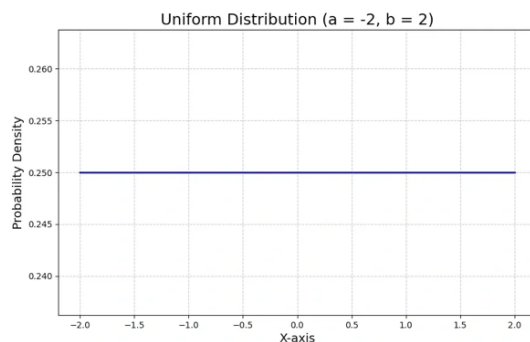
- **Truncated normal distribution:** This is a normal distribution but limited to a specific range of values. Truncated normal priors are useful when parameter values are expected to fall within a certain range, such as price or distribution elasticities.

Truncated Normal Distribution (Mean = 0, Standard Deviation = 1, $a = -2$, $b = 2$)



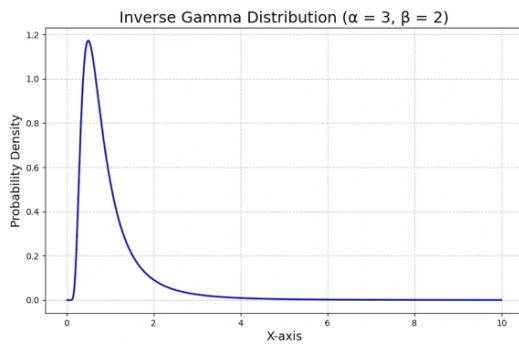
Graph of a truncated normal distribution prior

- **Uniform distribution:** This is a continuous distribution where all values within a specified range have an equal probability of occurrence. Uniform priors are often used as uninformative priors when there is no prior knowledge about the parameters.



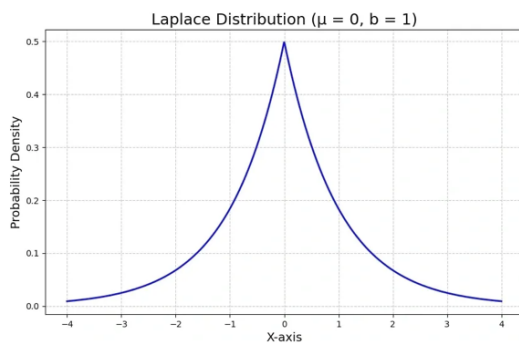
Graph of a uniform distribution prior

- **Inverse gamma distribution:**
The inverse gamma distribution is a strictly positive skewed distribution. Due to its positivity constraint, the inverse gamma is commonly used for measuring media impacts.



Graph of an inverse gamma distribution prior

- **Laplace distribution:** This is a continuous probability distribution characterized by its sharp peak at the mean and heavy tails. Laplace priors are useful in marketing mix modeling when there is a need to induce sparsity, for instance, in the case of a large number of predictors in the model.



Graph of a laplace distribution prior

Selecting the appropriate prior distribution for marketing mix modeling is crucial for obtaining meaningful results. The choice of prior should be guided by the specific problem, the available prior knowledge, and the desired level of

influence on the posterior distribution. However, it is important to note that as the sample size of your modeled data increases, the level of influence the prior has on the model diminishes. Intuitively, this makes sense because, with more data and information available, we would expect the model to be guided more by the data itself.

The Likelihood

The likelihood function plays a crucial role in Bayesian analysis, representing the probability of observing data given specific parameters. To illustrate this concept with a simple marketing-related analogy, imagine running an advertising campaign with an unknown effectiveness at increasing sales (represented by a parameter α). The likelihood function would express the probability of observing a particular set of sales data given the campaign's effectiveness (α). You may be familiar with the concept of maximum likelihood, which is the point where the likelihood function reaches its peak, indicating the most likely set of parameter values (in this case, α) given the observed

data. Under certain conditions, such as the normality of residuals, OLS estimates in Frequentist MMM are equivalent to the maximum likelihood estimate. In marketing mix modeling, selecting the appropriate likelihood function depends on the data's nature and assumptions about the response variable's distribution.

Below is a concise guide to choosing the right likelihood function for marketing mix modeling, including some common likelihood distributions:

- **Normal (Gaussian)**
distribution: This distribution is widely employed for modeling continuous response variables symmetrically distributed around the mean. In marketing mix modeling, the normal distribution is frequently used for modeling value sales or other continuous metrics like transformed dependent variables. It's suitable when the data shows a linear relationship with predictors and a constant variance (homoscedasticity).
- **Student's t-distribution:** This distribution is similar to the normal distribution, but with

heavier tails, essentially indicating a higher probability of outlier, resulting in a more robust model. In marketing mix modeling, the t-distribution can be used for modeling continuous response variables with potential outliers, such as store-level revenue with high levels of dispersion. It's a good choice when data exhibits mild deviations from the assumptions of normality and constant variance.

- **Poisson distribution:** This distribution models count data, i.e., non-negative integer values representing an event's number of occurrences. In marketing mix modeling, the Poisson distribution can be used for modeling the number of new customers, orders, or other count-based metrics. It's appropriate when the mean equals the variance, although this assumption often doesn't hold.
- **Negative binomial distribution:** An extension of the Poisson distribution, the negative binomial distribution accounts for overdispersion (when the variance is greater than the mean). In marketing mix modeling, the negative

binomial distribution can be used for modeling count data with overdispersion, such as new customers, orders, or other count-based metrics with high variability. Suitable when the Poisson distribution's assumptions don't hold.

In selecting a likelihood function for marketing mix modeling, it's

[HOME](#) [TECH ▼](#)

[ABOUT](#) [CONTACT](#)

function's assumptions and verifying whether they hold for the data is crucial. In some cases, data transformations or robust regression techniques might be necessary to meet the assumptions and obtain reliable results. If you're unsure, several model validation tools, such as Posterior Predictive Checks (which will be discussed later), can be used to determine if the correct likelihood has been chosen in Bayesian Modeling.

The Posterior: Combining the prior and likelihood

Combining the prior and likelihood using Bayes' rule is the fundamental principle of Bayesian analysis. The output, known as the posterior probability, contains all the

essential information to drive decision-making in Bayesian Marketing Mix Modeling Bayes' rule states that the posterior probability is proportional to the product of the likelihood and the prior probability. Mathematically, it can be written as:

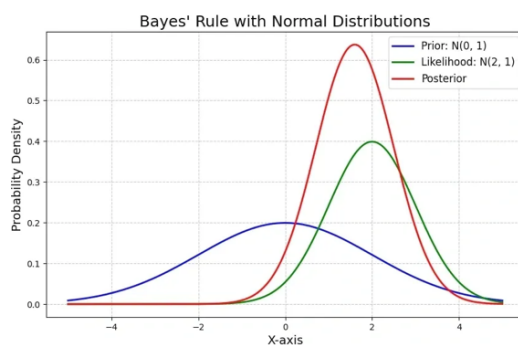
$$P(\theta|D) \propto P(D|\theta) * P(\theta)$$

Here, $P(\theta|D)$ is the posterior probability (the output), $P(D|\theta)$ is the likelihood, and $P(\theta)$ is the prior probability. However, in order to correctly combine these components, we require a normalizing factor, this is essentially to ensure that the posterior probabilities sum to one, otherwise our output posterior is not a valid probability distribution. The denominator in Bayes' rule, $P(D)$, is the marginal likelihood, which is the probability of observing the data, irrespective of the parameters.

Referring back to the weather example, suppose you have a prior belief about the probability of rain in London in April, and you observe dark clouds outside your window. The likelihood function represents the probability of observing dark clouds given that it is raining. By applying Bayes' rule, you can update your prior belief about the probability of rain by incorporating

the observed evidence (dark clouds) to obtain a posterior probability of rain.

In marketing mix modeling the process is similar. You start with a prior distribution for the parameters based on your prior knowledge or beliefs, and you have a likelihood function that represents the probability of observing the data given the parameters. You can then use Bayes' rule to update your beliefs about the parameters based on the observed data.



Graph of the impact of bayes' rule

In Bayesian analysis, we typically evaluate the posterior probability distribution at the mean rather than the maximum to obtain regression coefficients and other parameter estimates. This contrasts with the Ordinary Least Squares (OLS) method, which focuses on minimizing the sum of squared differences between observed and predicted values. By centering on the mean of the posterior

distribution, Bayesian approaches provide a more nuanced understanding of parameter uncertainty, incorporating prior beliefs alongside the observed data.

Bayesian inference may seem conceptually simple to calculate due to the use of Bayes' rule, but it often involves complex integration. With each additional variable comes another integral, and closed-form solutions are incredibly difficult beyond double integrals. To perform Bayesian analysis and obtain the posterior distribution, you can use probabilistic programming languages such as PyMC or Stan. These libraries implement advanced algorithms like Hamiltonian Monte Carlo Method. (HMC) with NUTS (No-U-Turn-Sampler) being the fastest HMC method currently available for efficiently sampling from the posterior distribution. By using these algorithms, you can approximate the posterior distribution without calculating the marginal likelihood directly, which is often intractable.

The Bayesian Workflow

The Bayesian workflow is a

structured, principled process to ensure a full understanding of your model and the model is robust enough to deliver the relevant insights you require for your business questions. The Bayesian workflow should be applied to marketing mix models in order to give credibility to your analysis. The workflow involves several iterative processes that allow for adjustments in the model based on different concepts and diagnostics. In the context of marketing mix models, these adjustments can include changing the likelihood, priors, model structure, reparameterization and other components to better capture the underlying data patterns. Andrew Gelman (2020) sets out a comprehensive Bayesian Workflow, the workflow also includes iterative model building, model checking, validation and troubleshooting of computational problems, model understanding, and model comparison. Here are some key concepts and diagnostics to consider in a well-structured Bayesian Workflow:

- **Sensitivity analysis:** This involves assessing how sensitive the model's results are to changes in the priors or

model structure. If your model's results change significantly with small changes in the priors, it may indicate a lack of robustness or an issue with the model's structure. In this case, you might need to reconsider the choice of priors, modify the model structure, or even consider an alternative modeling approach.

- **Posterior predictive checks:**
These checks involve comparing the model's predicted outcomes with the observed data. If there are substantial discrepancies between the predictions and the data, it might suggest that the model is not capturing the underlying patterns in the data effectively. You may need to adjust the likelihood, priors, or model structure to improve the model's fit.
- **Pareto k and Leave-One-Out Cross-Validation (LOO-CV):**
Pareto k is a diagnostic that helps identify potential issues with the model's fit, while LOO-CV is a method for model comparison and selection. If you observe high Pareto k values, it indicates there are points of high leverage and

you should consider adjusting the likelihood or model structure. LOO-CV can be used to compare different models and select the one with the best predictive performance.

- **Rhat and divergences:** Rhat is a diagnostic that measures the convergence of the Markov Chain Monte Carlo (MCMC) algorithm, while divergences indicate potential issues with the Hamiltonian Monte Carlo (HMC) sampling process. If Rhat values are above 1.2 or if there are many divergences, it might suggest problems with the MCMC algorithm or the model's structure. In such cases, you might need to reparametrize the model, adjust priors, or change the likelihood to improve convergence and sampling efficiency.
- **Credible Intervals:** Credible intervals are an essential aspect of Bayesian analysis, providing a range of values within which a parameter is likely to fall within the posterior distribution. In the context of marketing mix models, credible intervals offer insights into the uncertainty surrounding the estimated

effects of marketing channels, helping decision-makers assess the level of confidence they can place in the model's predictions. Unlike confidence intervals in frequentist statistics, credible intervals have a more intuitive interpretation, as they directly describe the probability that a parameter lies within the specified range. When interpreting the results of a Bayesian marketing mix model, it's crucial to consider the credible intervals alongside the point estimates, as they provide a more comprehensive understanding of the model's performance and the associated uncertainty. By incorporating credible intervals into the decision-making process, marketers can make more informed choices about resource allocation and strategy, while accounting for the inherent uncertainty in their data and models.

Bayesian MMM in Practice

Basic Regression

Bayesian Marketing Mix Modeling

(MMM) can start with a relatively simple linear regression model to understand the impacts of various factors on sales or other key KPIs relevant to your model. For example, in a CPG context, you can begin by estimating the coefficients for the impact of price changes on sales volume. In this simple scenario, there is prior knowledge that the elasticity of the relationship between price and sales volume is negative. Therefore, we can set a negative prior distribution, which can be combined with the observed sales data, such as from IRI or Nielsen, to generate more comprehensive estimate of the coefficients.

However, advertising and marketing are complex fields, and simple linear regression models often fall short in capturing the nuances of these relationships. This is where advanced concepts in Bayesian MMM, as introduced in various papers, become particularly valuable.

Time-Varying Parameters (TVPs) in Marketing Mix Modeling

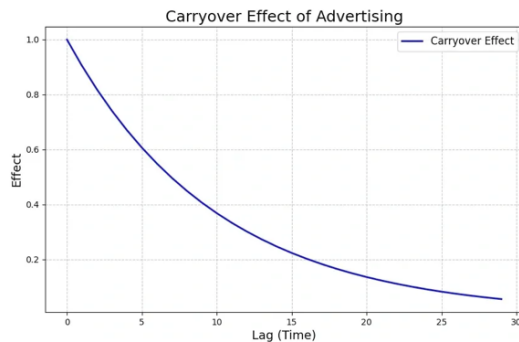
In the evolving landscape of marketing, Time-Varying Parameters (TVPs) have emerged as a pivotal tool in Bayesian Marketing Mix Modeling (MMM).

These models account for how the effect of marketing efforts changes over time, reflecting the dynamic nature of consumer behavior and market conditions. Unlike traditional models with static coefficients, TVPs allow for coefficients that adapt with the shifting effectiveness of marketing strategies. They enable the MMM to capture the influence of an advertising campaign that might be highly effective during a specific event, such as a major sports event, but less so afterwards. The incorporation of TVPs in Bayesian MMM offers a flexible, nuanced approach to estimating marketing effectiveness, crucial for making informed, timely decisions in today's fast-paced market.

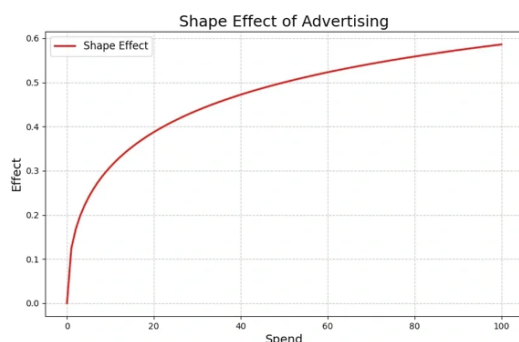
Media Estimation

In the paper by Jin et al. (2017), titled "Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects," the authors propose a media mix model that incorporates both carryover and shape effects of advertising. The carryover effect refers to the lag or delayed response to advertising, while the shape effect captures ad saturation and diminishing returns at high levels of spend. By modeling these effects, the authors provide a

more accurate and comprehensive understanding of the relationship between advertising spend and sales outcomes.



The carryover in marketing mix models



The shape effect in marketing mix models

Hierarchical Effects

Another important concept in Bayesian MMM is hierarchical modeling across channels and/or regions. In the paper "Geo-level Bayesian Hierarchical Media Mix Modeling" by Jin et al. (2017), the authors introduce a geo-level Bayesian hierarchical media mix

model (GBHMMM). This approach leverages sub-national data when available, providing more accurate estimates with tighter credible intervals compared to models based solely on national-level data. Hierarchical modeling allows for better understanding and optimization of marketing efforts by capturing the complex relationships between various marketing channels or regions.

Multivariate Priors

Multivariate priors are indeed used to explicitly model the correlations between variables in the context of Marketing Mix Modeling (MMM). They allow us to incorporate prior knowledge about the relationships between different parameters or variables, which can lead to more accurate estimates and improved decision-making.

In the context of MMM, multivariate priors can be particularly useful in the following ways:

- **Capturing correlations between marketing channels:** Advertising channels may exhibit correlated effects on sales, either due to similarities in their target audience, content, or other factors. By using a multivariate prior to

model, the joint distribution of the effects of different channels, we can account for these correlations and obtain more accurate estimates of their individual impacts.

- **Understanding cross-channel interactions:** In some cases, the effectiveness of one marketing channel might depend on the presence or absence of another channel. For example, the impact of an online ad campaign might be amplified when combined with a complementary TV campaign. Multivariate priors can help capture these interactions by modeling the joint distribution of the effects of different channels, enabling better insights into the true impact of each channel and more informed decisions about marketing budget allocation.
- **Incorporating domain knowledge:** Domain knowledge, such as industry-specific insights or information from previous campaigns, can be integrated into a multivariate prior to improve the model's performance. This can be particularly valuable when data is limited or noisy,

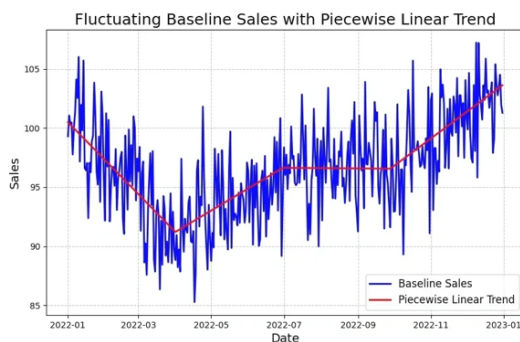
as the prior knowledge can help guide the estimation of parameters and lead to more accurate results.

Special Variables

Bayesian marketing mix models can be enhanced by incorporating various special variables and mathematical constructs to capture complex patterns and trends in the data more effectively. Some of the key aspects to consider are the organic base (or baseline sales), seasonality, and other factors that influence sales performance.

- **Long-Term Base** (baseline sales): Baseline represents the sales level in the absence of any marketing activities. Accurately modeling the organic base is crucial for understanding the incremental impact of marketing efforts. In a Bayesian marketing mix model, you can employ various mathematical constructs and functional forms to model the baseline sales, including the use of a Gaussian process for its flexibility and precision in capturing the nuances of sales patterns. While Facebook's Prophet forecasting model

uses a piecewise linear trend to represent baseline sales, a Bayesian approach allows for the seamless integration of trends through Gaussian processes, as discussed in the blog "[Measuring Marketing Effectiveness Over the Long-Term.](#)" This method provides a comprehensive framework for understanding and predicting baseline sales with greater accuracy.



The fluctuating base in marketing mix models

- **Seasonality:** Similarly seasonal patterns can play a significant role in driving sales. Fourier terms, derived from Fourier series, can be used to capture these cyclical patterns in a Bayesian marketing mix model. Facebook's Prophet model, for instance, uses Fourier terms to model seasonality in a flexible manner. By including Fourier terms, you can more

accurately capture the seasonal effects, leading to better model performance and more informed marketing decisions.

- **Unobserved Components Model (UCM):** The Unobserved Components Model (UCM) is a broader framework that can be used to model various aspects of the data, including trends, seasonality, and cyclical components. Through incorporating elements from the UCM into a Bayesian marketing mix model, you can leverage its flexibility and sophistication to capture complex patterns in the data. This can lead to improved model performance and help you gain deeper insights into the underlying drivers of sales, enabling better decision-making in marketing strategy.

Benefits of Marketing-Mix Models

Bayesian Marketing Mix Models offer a substantial range of benefits compared to their Frequentist counterparts, making them an attractive choice for marketing professionals and decision-makers.

By leveraging the power of Bayesian statistics, these models provide a more flexible, robust, and nuanced understanding of marketing effectiveness. While we have touched upon some of these advantages already, it is worth exploring the extensive list of merits:

- **Establishing a Cohesive Measurement Ecosystem:** Bayesian MMM creates a unified and comprehensive measurement ecosystem that seamlessly integrates results from uplift tests, attribution, and previous studies, while learning from new data sources. This single source of truth allows businesses to make well-informed decisions, as opposed to frequentist MMM, which only provides inferences based on the data being analyzed.
- **Incorporating Domain Expertise:** One key advantage of Bayesian MMM is its ability to include domain expertise, such as established relationships like negative price elasticities. Moreover, it accurately models aspects like distribution and stocking variables that can be

overstated in frequentist MMMs.

- **Delivering Robust Performance with Unreliable Data:** Bayesian MMM thrives across various markets and industries, even when data is unreliable or sporadic. Its adaptability ensures steadfast performance across different scenarios, making it a valuable tool for businesses dealing with inconsistent data.
- **Emphasizing Uncertainty for Informed Decision-making:** Bayesian MMM focuses on uncertainty, allowing the model to generate future scenarios that incorporate the full probabilistic impact of potential outcomes. This feature enables businesses to make fully informed decisions against all possible scenarios.
- **Streamlined Data Requirements:** Unlike frequentist MMM, Bayesian MMM can work with considerably shorter time periods, allowing businesses to initiate the measurement of marketing endeavors without delay. As the dataset expands, the certainty of the estimates progressively increases.
- **Leveraging Hierarchical**

Learning for Products and Regions:

Bayesian MMM enables regions and/or products to learn from one another by implementing a hierarchical relationship. This is particularly useful for pooling products or when introducing a new product with a limited number of observations.

- **Achieving Hyper-granular Measurement of Small Spend Channels:** By imposing a hierarchical relationship across ad creatives, Bayesian MMM attains hyper-granular measurement of small spend channels, helping businesses comprehend the impact at the most intricate level.
- **Precise Measurement in Synergistic Campaigns:** Media campaigns often operate across multiple channels concurrently to exploit synergy. Bayesian MMM accurately apportions impact in these instances by incorporating special priors (multivariate) to account for correlations between variables within the model structure.
- **Embracing Bayesian Decision-making and Optimization under Uncertainty:** Bayesian

decision-making, including optimization under uncertainty, represents a significant advantage of Bayesian MMM. This approach enables businesses to make strategic decisions by contemplating a range of potential outcomes and their associated probabilities, ensuring they are well-equipped for various scenarios.

Final Remarks

In this comprehensive guide, we explore the intricate components of Bayesian marketing mix modeling, a field poised for continued growth. Bayesian Marketing Mix Modeling (MMM) equips marketing managers and decision-makers with a powerful statistical framework to assess and optimize their marketing strategies. By incorporating prior knowledge, domain expertise, and flexible modeling techniques, Bayesian MMM delivers a more robust and nuanced understanding of marketing effectiveness, particularly when faced with unreliable or limited data.

This guide highlights the fact that Bayesian MMM involves various key

components such as the choice of priors, likelihood functions, and advanced modeling techniques, which help capture the complexities of marketing data and inform better decision-making processes. By adopting an iterative, structured Bayesian Workflow, practitioners can ensure the credibility and reliability of their analysis, leading to well-informed strategic decisions in marketing resource allocation and campaign planning.

As the cookie crumbles and data-driven decision-making becomes increasingly critical, Bayesian Marketing Mix Modeling emerges as the preferred approach for agencies and analysts seeking a comprehensive understanding of marketing performance. By leveraging the inherent flexibility and sophistication of Bayesian statistics, marketers can gain deeper insights into the impact of their media investments and make more informed choices about their marketing strategies.

Author: Niall Oulton, Company: 1749

Gain deep insights and strategic foresight. Enhance your business with advanced analytics. We offer everything from standard to custom