Master's thesis

Master's Programme in Data Science

Bayesian Experimental Design in Marketing Mix Modeling

Tri Luu

January 30, 2026

Supervisor(s): Assoc. Professor Luigi Acerbi

Examiner(s): Assoc. Professor Luigi Acerbi
Dr. Mikko Ervasti

University of Helsinki
Faculty of Science

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Data Science

Tekijä — Författare — Author

Tri Luu

Työn nimi — Arbetets titel — Title

Bayesian Experimental Design in Marketing Mix Modeling

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Master's thesis	January 30, 2026	65

Tiivistelmä — Referat — Abstract

Marketing mix modeling (MMM) is a statistical method for measuring marketing effectiveness, with Bayesian MMMs among the most widely adopted. MMMs provide insights into historical marketing performance and guidance for future budget allocation planning. Still, MMMs do not necessarily recommend what actions to take to improve the current knowledge and understanding of marketing performance. On the other hand, marketing experiments provide insights into specific questions that help gain more information and understanding, but they are limited in scope due to operational complexity. Bayesian experimental design (BED) emerges as a theoretically grounded framework that utilizes the current state of knowledge reflected in the Bayesian MMMs and guides the design of marketing experiments under resource constraints.

The thesis demonstrates how the BED framework can be applied to design information-optimal investment allocations for marketing experiments using two simulated scenarios where (i) the performance of one out of four channels in the mix is uncertain, and (ii) the performance of all four channels is uncertain. The optimal allocations in the first scenario are shown to reduce the uncertainty in the posterior distributions and shift the posterior modes closer to the ground truth, while the optimal allocations in the second scenario provide limited improvements in the posterior uncertainty compared to the models fitted without any experimental investments. In both scenarios, the optimal experiments are considered extreme relative to the typical budget allocation, so budget constraints are required to ensure the practicality of the experiments in real-world scenarios. Meanwhile, the results demonstrate that, under reasonable prior specification and budget constraints, controlled investment variation experiments optimized by the BED framework have promising potential to provide additional insights into marketing performance.

ACM Computing Classification System (CCS):
Applied computing → Operations research → Marketing
Mathematics of computing → Probability and statistics → Probabilistic inference problems → Bayesian computation

Avainsanat — Nyckelord — Keywords

Bayesian experimental design, marketing mix modeling, Bayesian inference

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

Preface

Bayesian inference has always been an interesting topic to me, as Bayes' theorem is so intuitive that it quantifies the uncertainty and confidence in human knowledge and represents our learning process using statistics. While Bayesian inference courses at the university have captivated me with the intuition behind the Bayesian framework, Sellforte has provided me with the opportunity to see how Bayesian inference is applied to real-world marketing analytics at a large scale. The amount of data and the number of features used in marketing analytics make Bayesian inference more difficult and yet more intriguing than ever.

Therefore, I would like to thank Juha Nuutinen and Mikko Ervasti, two co-founders of Sellforte, for allowing me to work at the company and for believing in me to solve challenging problems. Moreover, I would like to thank my thesis advisors at Sellforte, Mikko Ervasti (again) and Kacper Solarski, and my thesis supervisor, Luigi Acerbi, for their helpful feedback on my work and amazing insights about Bayesian inference and marketing analytics.

ChatGPT and Grammarly have been utilized for checking grammatical errors, finding synonyms, and improving the coherence and cohesion of the thesis.

1. Introduction

Marketing is an indispensable function of modern firms; yet, measuring marketing effectiveness remains one of the most challenging problems in both academic research and industry practice. This chapter provides an overview of how marketing and its measurement methods have evolved over the years and how marketing mix modeling and Bayesian experimental design can be combined to form a measurement module for marketing effectiveness.

1.1 Background & Context

Marketing plays a crucial role in many aspects of a business, from launching a new feature to entering a new market. The practice of marketing and product promotion can be traced back to ancient Greece and Rome, where archaeologists have discovered evidence of advertising, branding, and promotional messages (Jones and Shaw, 2002). Marketing has also become a formal subject in universities for over a century, with some of the earliest courses taught at the University of Pennsylvania in 1905 and the University of Wisconsin in 1910 (Bartels et al., 1976). Throughout this long history, marketers have always wanted to know which marketing channels perform the best and how much return each channel generates, a question at the core of marketing accountability and productivity research (Rust et al., 2004). However, the complexity of customer behaviors and the non-stationary nature of market environments make it difficult to measure the performance of marketing channels with high certainty (Rust et al., 2004). Moreover, skepticism about the effectiveness of marketing has long existed among some business executives. For example, Phil Knight, the co-founder of Nike, recounted in his memoir that he once asked his marketing agency to prove in concrete numbers that their marketing efforts and investments had generated additional revenues for the company, but he did not receive persuasive responses (Knight, 2016). Knight even mentioned that he was not a firm believer in marketing and even expressed negative opinions about marketing graduates (Knight, 2016).

Phil Knight's memoir reflected the context of marketing around the 1970s and 1980s, during which marketing was mainly done via channels like television, print, magazines, or sponsorships. We categorize these marketing channels as "offline channels", in contrast to "online channels" that emerged later with the rise of the Internet. A key limitation of offline channels was the restricted amount of diagnostic data that could be collected to attribute exposures to outcomes with precision (Vakratsas and Ambler, 1999). Marketers often relied on reach estimates or post-exposure surveys, such as asking buyers whether they had seen specific ads. However, self-reports in those surveys introduced validity and bias concerns and did not identify the incremental sales effects (Vakratsas and Ambler, 1999), while attribution measurement in marketing in general and offline channels specifically could be further complicated by external factors such as brand awareness and macroeconomic trends.

During the 1990s and 2000s, the emergence of the Internet facilitated a new wave of marketing methods that are widely used to the present day, which we categorize as "online marketing". Online marketing is centered around websites with high traffic and rich user-level data, such as search engines like Google or Bing and social platforms like Facebook, Snapchat, or LinkedIn. These platforms provide marketers with targeting opportunities to reach specific audiences that offline channels could not offer (Goldfarb and Tucker, 2011a). While this ensured improved targeting efficiency, it did not necessarily resolve the uncertainty in marketing effectiveness measurement. A method known as the "last-click attribution model" was devised to address this challenge. Last-click attribution attributes a conversion event, such as a product purchase or a subscription sign-up, to the user's final activity right before the conversion (Anderl et al., 2016; Li and Kannan, 2014). A critical enabling technology for last-click attribution models was the Urchin Tracking Module (UTM), introduced by Urchin Software (later acquired by Google) in the early 2000s. UTM parameters are tags appended to the URLs, for example: `utm_source=google&utm_medium=cpc&utm_campaign=marketing_campaign`, to track the traffic sources and user web sessions across web properties (Clifton, 2012). UTM tracking has been widely adopted and has become the industry standard, helping advertisers and ad platforms better monitor their marketing performance and understand whether an activity is attributed to an ad or not.

Over time, the success of online marketing collided with the increasing concerns over data privacy (Goldfarb and Tucker, 2011b). Major ad platforms are criticized for intensively collecting and monetizing users' data (Degeling et al., 2019). New legislation, such as the European Union's General Data Protection Regulation (GDPR), restricts cross-site tracking and enforces mandatory user consent, helping limit data collection and bring more transparency to websites' data usage (Goldfarb and Tucker, 2011b; Degeling et al., 2019). These criticisms and restrictions have been shown to reduce the effectiveness of cookie-based tracking and introduce biases to the attribution models (Miller and Skiera, 2024). As a result, UTM tracking and last-click attribution became less reliable for measuring marketing performance, and marketers started to explore alternatives to measure their marketing effectiveness, with marketing mix modeling being one of the famous options.

Marketing mix modeling (MMM) is a statistical method that uses historical sales and marketing data to measure the impact of various marketing activities in the mix on some key performance indicators (KPIs) such as revenues, website traffic, or number of new subscriptions. The methods can range from Bayesian models (Brodersen et al., 2015; Sun et al., 2017) to complex deep learning models (Gong et al., 2024; Mulc et al., 2025). Among those MMM methods, Bayesian models have become more popular as they allow marketers to utilize their domain knowledge of the company's products, customers, and market dynamics, and incorporate such knowledge into the model's prior distributions and quantify the uncertainty level of marketing performance (Gelman et al., 2014). Unlike digital attribution methods like last-click attribution that are mainly built for online interactions and website traffic, MMM incorporates both offline and online marketing channels along with other external factors such as seasonality or macro-economic events, providing insights into the historical marketing performance and helping with future investment budget planning (Brodersen et al., 2015; Clifton, 2012; Sun et al., 2017). However, MMM provides little guidance to marketing practitioners on what investment strategies they could implement to further improve their current knowledge and understanding of marketing performance, thereby enhancing the results of the model itself. Moreover, MMM results are often sensitive to the assumptions made in the model or, in the case of Bayesian models, the prior distributions of the parameters (Gelman et al., 2014), and incorrect assumptions can make the model's insights irrelevant to the marketers.

On the other hand, marketing experiments, including A/B tests, geo-lift experiments, or controlled investment variation experiments, provide strong causal inferences by isolating the incremental effects of a marketing intervention (Athey and Imbens, 2017). In a geo-lift experiment, for instance, marketers can increase the budget of one marketing channel in a treatment location (the test group) while keeping the budgets of other comparable regions unchanged (the control group). The sales recorded in the control group can then be used to estimate the sales of the test group as if the test group's budget were not adjusted, or as if the intervention were not carried out. The difference between the estimated sales and the actual recorded sales in the test group of the experiment is considered the incremental sales of the studied channel. The advantage of these experiments is that they can directly answer marketing questions posed by practitioners, offer precise and convincing evidence of marketing effectiveness, and provide additional information and learning that MMM often struggles to deliver.

However, such marketing experiments are costly and operationally complex. They often require significant efforts to avoid the spillover effects between the treatment and control groups (Eckles et al., 2018). Coming back to the geo-lift test example above, if the treatment location is close enough to, or even overlaps with, other regions in the control group, the increased budget of the studied channel can also affect the sales in the control group, harming the effectiveness of the causal estimation. Moreover, marketing experiments are often narrow in their scope. They are typically designed to answer specific questions about individual campaigns or channels and do not provide insights about the cross-channel interactions and long-term effectiveness that MMM offers (Brodersen et al., 2015; Sun et al., 2017). Hence, there is a contrast between MMM and marketing experiments. MMM offers scalability and long-term insights, but with limited actionable recommendations to improve the current knowledge and understanding of marketing performance. Meanwhile, experiments provide answers to specific questions about marketing performance at the cost of complex implementation and limited scope.

In such contexts, Bayesian experimental design (BED) emerges as a promising connection between marketing experiments and MMM. BED applies Bayesian decision theory to approach experiment design problems by maximizing the expected specified utility, often defined as expected information gain (Ryan et al., 2016). The benefits of the BED framework become more apparent when being put into consideration with MMM and marketing experiments. BED utilizes prior knowledge, encapsulated as MMM's estimates, to design the new experiments, ensuring the limited resources are allocated optimally so that they generate the most information and learning. On the other hand, experiments designed by the BED framework can, in turn, be used to refine the MMM model, further enhancing the model's capabilities and robustness. In this way, MMM, marketing experiments, and BED are neither competing methods nor alternatives to each other, but they are complementary components that can work together to deliver the most value in marketing measurement.

1.2 Thesis Contribution

To summarize, MMM is a structured, aggregated, and data-driven marketing effectiveness measurement method that provides insights into historical marketing performance and helps with future investment budget planning, but lacks actionable recommendations on how to improve the marketing knowledge and understanding. Meanwhile, marketing experiments offer answers to specific marketing questions and validation of the assumptions made in the models, helping gain more information and learning about marketing performance. However, it is not straightforward to design a marketing experiment under resource constraints like investment budgets or the maximum length of the experimental period. This thesis studies how the BED framework can be utilized to help marketers design information-optimal investment allocations for marketing experiments under different states of knowledge and confidence encapsulated by the MMM. Such optimal experiments can, in turn, be used to fine-tune the model, forming a closed self-improvement iterative module that helps marketing practitioners gain insights into their marketing effectiveness and performance of the channels in the mix.

The thesis is structured as follows. Chapter 2 discusses the technical background of the MMM and the BED framework. Chapter 3 presents the methodology used in the thesis: how the MMM is built and fitted, what utility function is used in the BED framework, and how the synthetic data used throughout the thesis is generated. Chapter 4 illustrates how the BED framework can be utilized in the context of MMM to design optimal investment allocations using two simulated scenarios where (i) the performance of one channel in the mix is uncertain and (ii) the performance of all channels in the mix is uncertain. These two scenarios aim to represent two different levels of uncertainty about marketing performance and study what the optimal investment allocations in each case are to gain the most information and learning. Chapter 5 reflects the limitations of the thesis and discusses potential ideas and improvements upon its work, and Chapter 6 summarizes the contributions and the insights of the thesis.

2. Background

This chapter discusses the background of marketing mix modeling (MMM) and Bayesian experimental design (BED). Section 2.1 presents the mathematical definition of MMM and two typical assumptions made in the model about how marketing works. Section 2.2 discusses the BED framework and its main practical challenges.

2.1 Marketing Mix Modeling

Marketing mix modeling (MMM) is a statistical method used to measure the contribution of various marketing activities to some key performance indicators (KPIs), most commonly sales or revenue. MMM provides marketing practitioners with an empirical foundation for understanding how different marketing channels, including both offline and online channels, and external factors, such as seasonality, competition, or macroeconomic indicators, affect the interested KPIs (Hanssens et al., 2001; Jin et al., 2017). With statistical insights, MMM helps marketers evaluate the efficiency of their investments, supporting the budget allocation decision processes.

MMM is typically formulated as a regression model, where the target variable y_t at time t is expressed as a function of marketing inputs along with other control variables:

$$y_t = f(x_{t,1}, x_{t,2}, \cdots, x_{t,M}, z_t; \theta) + \epsilon_t. \tag{2.1}$$

Here, $x_{t,m}$ is the marketing inputs of Channel m (with $m = 1, 2, \cdots, M$), z_t represents the control factors, such as seasonality or trend, $f(\cdot)$ is a nonlinear response function parameterized by θ, and ϵ_t represents the stochastic noise, most commonly modeled using Gaussian distributions $\mathcal{N}(0, \sigma^2)$. Two defining features of MMM are the carryover effects and the diminishing returns or saturation effects (Broadbent, 1979; Hanssens et al., 2001; Jin et al., 2017). Respectively, these effects capture the temporal persistence of marketing effects and the nonlinear relationships between marketing inputs and outputs.

Carryover effects refer to the delayed and decaying impact of marketing over time. More specifically, instead of exerting all observable influences solely at the time of ex-

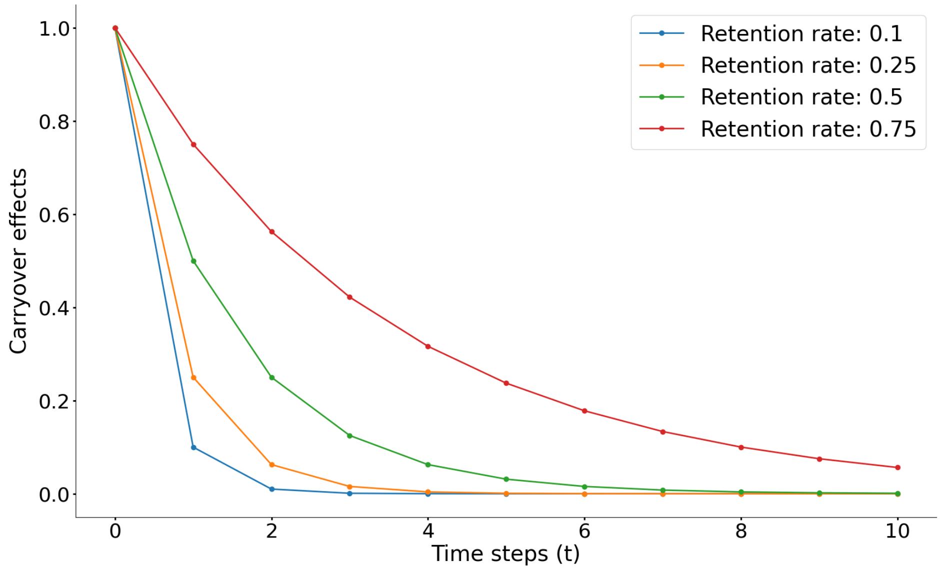

Figure 2.1: Simple adstock function: $x^*_{t,m} = x_{t,m} + \lambda x^*_{t-1,m}$ with different retention rates. A marketing exposure of Channel m is carried out at the time 0 with the recorded value $x_{0,m} = 1$. The carryover effects are recorded over a time period of 10, during which no other marketing exposures are carried out, such that $x_{t,m} = 0$ for any $t > 0$. Different retention rates affect how fast the carryover effects decay to zero. Channels with small retention rates decay faster than those with large retention rates.

posure, a marketing activity continues to affect customer behavior and KPIs for subsequent periods, even after the activity is stopped, but with diminishing effectiveness (Broadbent, 1979). Such diminishing persistence is attributed to customer memory, brand awareness, and purchasing cycle (Naik and Raman, 2003). Carryover effects in MMM can be modeled using the adstock function (Broadbent, 1979; Franses, 2025). A simple formulation of an adstock function is given by

$$x^*_{t,m} = x_{t,m} + \lambda x^*_{t-1,m}, \tag{2.2}$$

where, at the time t, $x^*_{t,m}$ is the adstock-transformed marketing variable of Channel m, $x_{t,m}$ is the recorded value of the marketing variable, and $\lambda \in (0, 1)$ is the decay parameter or the retention rate of the carryover effects. Figure 2.1 illustrates the carryover effects using Equation (2.2) with different values for the retention rate λ. Figure 2.1 shows the exposure of a marketing activity at time 0, after which no marketing activities are carried out for a time period of 10. Figure 2.1 reveals how different choices of the retention rate reflect the marketers' assumptions of how a channel works. High retention rates correspond to long-lived effects, which are suitable for brand campaigns, while low retention rates describe short-lived effects suitable for promotional campaigns (Broadbent, 1979).

Equation (2.2) indicates that, theoretically, the adstock effects of a marketing activity never decay to zero. However, this is not always true in practice, and a MMM model often imposes a finite max lag length to limit the model complexity and better reflect domain knowledge. For example, Jin et al. (2017) parameterized the adstock function of one marketing channel using its spend, a max length parameter, and a geometric decay weight function as follows:

$$x^*_{t,m} = \frac{\sum_{l=0}^{L-1} \alpha_m^l x_{t-l,m}}{\sum_{l=0}^{L-1} \alpha_m^l}, \tag{2.3}$$

where L is the maximum look-back period, and α_m is the retention rate of the channel m. This structure allows practitioners to capture the persistent marketing effects while ensuring the model's robustness. On the other hand, it is important to note that a long adstock window may blur the distinction between short-term media effects and long-term baseline growth, a problem known as the adstock illusion (Cain, 2025). Carefully specifying decay structures or using hierarchical models helps avoid this confounding problem.

The second key marketing assumption in MMM is diminishing returns, which are applicable in many fields, including economics, finance, and energy. In the marketing context, diminishing returns refer to the scenario that, when the investment in one channel is increased, the effectiveness of the channel, measured by, for example, the amount of incremental sales it generates, will increase at a decreasing rate (Simon and Arndt, 1980). After a specific investment threshold, the channel stops bringing any additional effects or benefits, regardless of how much more investment is put into the channel. Such a threshold is called "the saturation point", and the channel is said to be saturated when its investment reaches the saturation point. In MMM, diminishing return effects can be modeled using the Hill function (Jin et al., 2017):

$$\text{Hill}(x_{t,m}; \alpha_m, \beta_m) = \frac{1}{1 + (x_{t,m}/\alpha_m)^{-\beta_m}}, \tag{2.4}$$

where $x_{t,m}$ is the value of the marketing variable of Channel m at the time t, $\alpha_m > 0$ is the half saturation point, which is the value of $x_{t,m}$ that results in 50% of the saturation, and β_m is the shape, or slope, parameter. Figure 2.2 illustrates the diminishing returns modeled using the Hill function. Figure 2.2 shows that, as we increase the value of $x_{t,m}$, the value of the Hill function asymptotically approaches 1, capturing the saturation effects. An alternative way to include the diminishing returns is to use the inverse-negative exponential (INE) function:

$$\text{INE}(x; \beta_m) = 1 - \exp\left(\frac{-x}{\beta_m}\right), \tag{2.5}$$

where β_m is the shape of the curve. Intuitively speaking, the shape of the INE curve can be interpreted as the amount of investments needed to reach $1 - 1/e \approx 0.632$ of

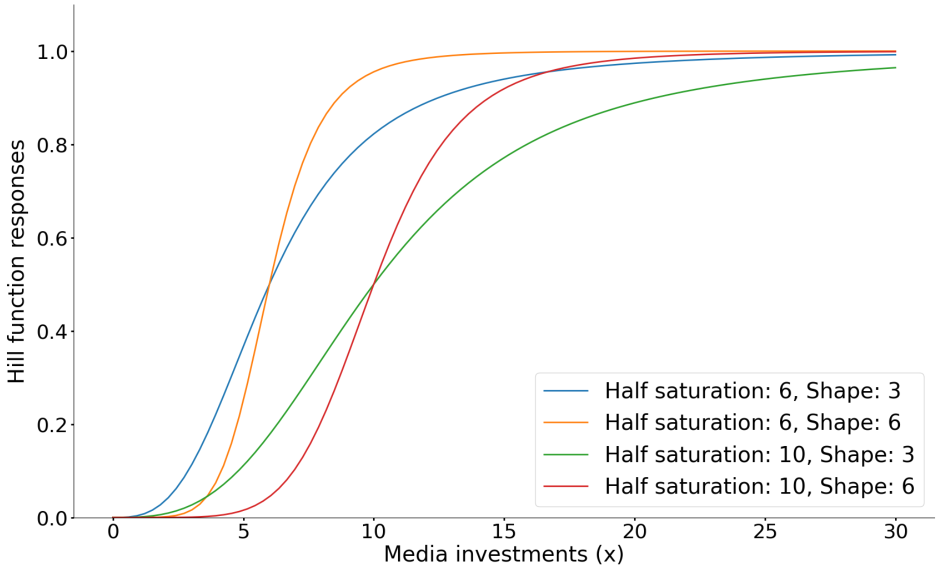

Figure 2.2: Hill function with different values for the half saturation point α and the shape parameter β. As the amount of investment x is increased, the effectiveness of a channel is increased at a diminishing rate, asymptotically reaching 1 as the saturation level. Channels with higher half-saturation points α have shallower progression and require higher investment values x to get close to the saturation level. On the other hand, channels with higher values for the shape parameter β have steeper progression and require smaller investment values x to get close to the saturation level.

the saturation level. Figure 2.3 illustrates the diminishing returns modeled using the INE function. Figure 2.3 shows that a channel with a smaller value for the shape parameter is faster to be saturated than a channel with a larger value for the shape parameter. The INE is a concave function. Hence, it is computationally simpler and often preferred when MMM outputs are used for budget optimization tasks because they ensure the convexity of the optimization problem.

2.2 Bayesian Experimental Design

Bayesian experimental design (BED) theory is based on the foundation of decision theory, an approach for solving a wide range of applied statistical problems. In this section, we will describe the BED framework, its marketing use cases, and practical challenges. This section is mainly based on Bernardo et al. (1994) and Gelman et al. (2014).

There are three concepts central to decision theory: states, utilities, and decisions. States are the possible states of the world or underlying conditions, such as unknown factors or model parameters, that represent one's current knowledge or uncertainty. Utilities are the decision maker's preferences, quantifying the value gained or loss incurred by an action. Decisions are the actions or choices available to the

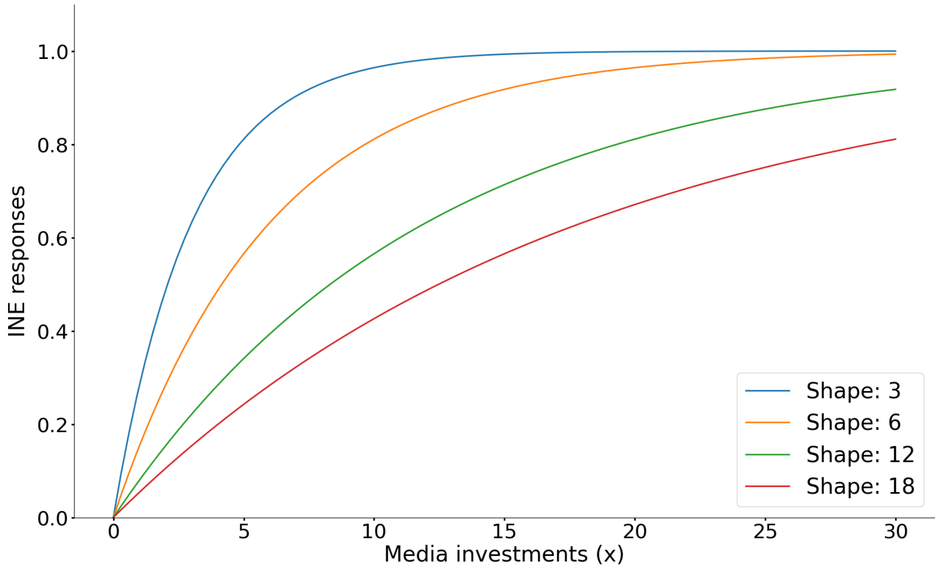

Figure 2.3: Inverse-negative exponential (INE) function with different values for the shape parameter β. A curve with a lower shape has a steeper slope, while the one with a larger shape has a flatter slope. The shape parameter β determines how fast a channel gets saturated, as it represents how much investment is required to reach approximately 63% of the saturation level. Channels with small values for the shape parameter saturate faster than those with large values for the shape parameter.

decision maker to be taken based on the given state and associated utilities. A decision problem involves taking actions (decisions) that maximize the decision maker's preferences (utilities) given their current state of knowledge. Uncertainty of a decision problem can stem from unknown states of the world or from variability in the utility outcomes of a decision. In the context of Bayesian experimental design, decisions are choices of experiments, such as experiment configurations and settings, and the optimal experiment is the one that maximizes the specified utility function given the current knowledge. Formally, the Bayesian experimental design problem can be written as an optimization problem:

$$\hat{d} = \arg\max_{d \in \mathcal{D}} \overline{U}(d), \tag{2.6}$$

where d denotes a candidate experiment from the space of all possible designs $\mathcal{D}$ with $\hat{d}$ being the optimal experiment, and $\overline{U}(d)$ is the expected utility function over all uncertain states. The expected utility function is defined by integrating, or averaging, the utility over all of the unknown states θ:

$$\overline{U}(d) = E_\theta \left[U(d, \theta) \right] = \int_{\theta \in \Omega} U(d, \theta) p(\theta | d) d\theta, \tag{2.7}$$

where Ω is the space of all possible world states, $U(d, \theta)$ is the utility of choosing the design d under the state θ, and $p(\theta|d)$ reflects any prior information about θ, which, before data collection of the experiment, is typically the prior $p(\theta)$ since the design d itself does not change the prior distributions. In short, the goal of BED is to choose the experiment d that maximizes the expected utility $\overline{U}(d)$ given the decision maker's current knowledge.

A major strength of the BED framework lies in its flexibility: the ability to incorporate any suitable utility functions $U(\cdot)$ and the design space $\mathcal{D}$ based on one's problems. For example, one might choose a utility function that rewards precision in parameter estimation or one that rewards improved predictions for future observations (Ryan et al., 2016). One of the most common utility functions used in BED is the expected information gain (EIG) (Ryan et al., 2016; Foster et al., 2019), which is an information-based measure of how much an experiment is expected to increase the current knowledge. EIG is also the main utility function used in the thesis.

In practice, with EIG, an experiment is considered to be optimal if it is expected to produce data that most reduces our uncertainty about the model parameters of interest. More formally, the optimal design $\hat{d}$ is the one that maximizes the Kullback-Leibler (KL) divergence between the prior and posterior distributions after observing the experiment's outcome (Kullback and Leibler, 1951; Ryan et al., 2016; Foster et al., 2019). Intuitively speaking, the most optimal design $\hat{d}$ is expected to yield observations that make the posterior distribution as different from the prior distribution as possible. In the Bayesian context, the KL divergence criterion is equivalent to maximizing the mutual information between the experimental data y and the model parameters θ. Mathematically, the expected information gain utility can be expressed as the expected KL divergence between the prior and posterior distributions over all possible observations y of the experiment and all possible model parameters θ:

$$U(d) = E_{\theta|y,d} E_{y|d} \left[\log \frac{p(\theta|y,d)}{p(\theta)} \right]. \tag{2.8}$$

Notably, Equation (2.8) involves a nested expectation: one over all possible experimental observations y and the other over all possible states or model parameters θ. Hence, in most cases, there is no analytical solution to calculate the double integrals in the expected information gain, especially in the case of MMM, where the model involves special functions representing carryover effects and diminishing returns as mentioned in Section 2.1. Different numerical approximation methods, such as Monte Carlo algorithms and information sampling, have been used in the literature to estimate the expected information gain for experimental design problems (Ryan et al., 2016).

Finally, it is worth noting the flexibility in how one can define the space of possible experiments $\mathcal{D}$ in the BED framework. In marketing applications, experiments can take many forms, from A/B tests to geo-lift experiments. Designs of such experiments can also vary in granular details. For example, an experiment could consist of showing a specific advertisement to only a particular user segment in an A/B test or allocating different budget levels to different marketing channels in different geographic markets in a geo-lift experiment. The BED framework can, in principle, accommodate all such variants by letting the design space $\mathcal{D}$ represent the set of all possible implementation choices.

3. Methodology

This chapter describes the methodology adopted in the thesis[†]. Firstly, Section 3.1 presents the formulation of the marketing mix model (MMM) along with its structural assumptions, observation model, and prior distributions under the Bayesian inference framework. Secondly, Section 3.2 introduces the utility function used in the Bayesian experimental design (BED), with a particular focus on expected information gain and its connection to the MMM. Finally, Section 3.3 describes the procedure used to generate the synthetic data employed throughout the analysis discussed in Chapter 4 of the thesis.

3.1 Marketing Mix Model

As discussed in Section 2.1, MMMs typically rely on two core assumptions about the effects of marketing activities: carryover effects and diminishing returns. Firstly, carryover effects are modeled with a geometric adstock function (Jin et al., 2017; Hanssens et al., 2001). At any given time period, the effective media exposure of a channel is defined as a weighted average of current and past marketing exposures, where the weights decay geometrically over time. The function's retention rate parameter controls how quickly the effects of past exposures diminish. The geometric adstock function can be formulated as:

$$\text{adstock}(x^*_{t,m}; r_m, L) = \sum_{l=0}^{L-1} r^l_m x_{t-l}, \tag{3.1}$$

where $x^*_{t,m} = \{x_{t',m} \,|\, t' \leq t\}$ is the time series of the marketing inputs $x_{t',m}$ of Channel m up until week t, r_m is the retention rate parameter of Channel m, and L is a hyperparameter representing the maximum look-back period of the carryover effects. In the thesis, L is set to be 13 weeks by default, which is assumed to be sufficiently long to capture all significant carryover effects while maintaining computational tractability. The observed marketing inputs $x_{t,m}$ are considered the investments of Channel m measured in monetary units.

[†]Python implementation of the methodology in the thesis can be found at https://version.helsinki.fi/luutri/bedmmm and https://github.com/triluu03/bedmmm.

Secondly, diminishing returns are modeled using the inverse-negative exponential (INE) function. This function maps adstocked-transformed marketing investments to a bounded response that asymptotically approaches a saturation level. The shape parameter determines how quickly the response approaches the saturation and can be interpreted as the level of investments required to reach approximately 63% of the saturation level. The INE function can be formulated as:

$$\text{INE}(x'_{t,m}; \beta_m) = 1 - \exp\left(\frac{-x'_{t,m}}{\beta_m}\right), \tag{3.2}$$

where $x'_{t,m}$ is the adstock-transformed marketing investments at week t and β_m is the shape parameter of Channel m.

Combining the carryover effects, diminishing returns, baseline demand, and control variables, the target output is modeled as the total of channel-specific contributions, a baseline component, linear control effects, and stochastic noise:

$$y_t = \sum_{m=1}^{M} \alpha_m \, \text{INE}(\text{adstock}(x^*_{t,m}; r_m, L); \beta_m) + b_t + \sum_{c=1}^{C} \gamma_c z_{t,c} + \epsilon_t, \tag{3.3}$$

in which, at week t, y_t is the target metric, α_m is the saturation level of Channel m ($m = 1, \cdots, M$), b_t is the baseline metric, $z_{t,c}$ is the control variable ($c = 1, \cdots, C$) with the control effect parameter γ_c, and ϵ_t is the added white noise.

One of the most common control variables is seasonality, representing the nature of a company's products or customer behaviors. For example, gift cards or presents are typically sold more during the holidays, such as Christmas or New Year's Eve, than on a normal day, while winter coats are typically sold more during the winter time than the summer time. In this thesis, seasonality is also included as the sole control variable of the model.

In Equation (3.3) that defines the MMM, the observed data includes the marketing time series $x^*_{t,m}$ and the control variables $z_{t,c}$. Besides the hyperparameter L, which is set to 13 by default in the thesis, all other parameters are modeled using Bayesian inference. The retention rate r_m of Channel m is assumed to lie between zero and one and, therefore, assigned to have a uniform prior distribution with the lower bound lb_m and the upper bound ub_m that satisfy $0 \leq lb_m \leq ub_m < 1$. The saturation parameter α_m is strictly positive and is modeled using a truncated normal distribution with a lower bound at zero and the mean and standard deviation denoted as $\mu_{\alpha,m}$ and $\sigma_{\beta,m}$, respectively. Similarly, the shape parameter β_m is also strictly positive and assigned a truncated normal distribution with lower bound at zero and the mean and standard deviation denoted as $\mu_{\beta,m}$ and $\sigma_{\beta,m}$, respectively. The control coefficient follows a zero-mean normal distribution with a standard deviation denoted as $\sigma_{\gamma,c}$. The baseline is also modeled with a normal prior distribution with a mean of μ_b and a

Table 3.1: Prior distributions of the parameters in the MMM fitted with Bayesian inference.

Parameter name	Symbol	Distributions
Retention rate	r_m	$\text{Uniform}(lb_m, ub_m)$
Shape	β_m	$\text{TruncatedNormal}(\mu_{\beta,m}, \sigma^2_{\beta,m})$
Saturation	α_m	$\text{TruncatedNormal}(\mu_{\alpha,m}, \sigma^2_{\alpha,m})$
Control's effect	γ_c	$\text{Normal}(0, \sigma^2_{\gamma,c})$
Baseline	b_t	$\text{Normal}(\mu_b, \sigma^2_b)$
Stochastic noise	ϵ_t	$\text{Normal}(0, \sigma^2_\epsilon)$

standard deviation of σ_b. Lastly, the stochastic noise is generated based on the normal distribution with zero mean and the standard deviation of σ_ϵ. Table 3.1 summarizes the prior formulation of all parameters in the model.

The model is fitted to a historical dataset consisting of observed target metrics, media investments, and control variables over time. Mathematically, the dataset used in MMM can be formulated as:

$$\mathcal{D} = \{(y_t, \mathbf{x}_t, \mathbf{z}_t) \,|\, t = 1, 2, \cdots, T\}, \tag{3.4}$$

where y_t is the target metric, $\mathbf{x}_t$ is a vector containing the marketing investments of all M media channels, $\mathbf{z}_t$ is a vector containing the values of all C control variables, and T is the length of the time-series dataset measured in weeks. In the thesis, the model is fitted using the No-U-Turn Sampler (NUTS) algorithm (Hoffman et al., 2014) with the implementation in the Python library PyMC (Patil et al., 2010).

3.2 Utility Function for Bayesian Experimental Design

Bayesian experimental design (BED) aims to select experimental conditions that maximize the expected information gain (EIG), defined in Equation (2.8), about model parameters of interest. In the thesis, the experimental variables correspond to media investments of marketing channels in the mix over a pre-defined period of time, and the objective is to maximize the learning about marketing performance given the current state of knowledge. Notably, the current state of knowledge is considered after the historical dataset $\mathcal{D}$ mentioned in Equation (3.4) has been observed, and the Bayesian MMM has been fitted. In other words, the priors over parameters in the BED framework are the posteriors obtained from fitting the MMM to the historical data $\mathcal{D}$. The likelihood of future observations is determined by the generative model defined by the MMM and the proposed media investments. Therefore, the EIG can be formulated with a dependency on the observed historical dataset $\mathcal{D}$ as follows:

$$U(\mathbf{x}|\mathcal{D}) = E_{\theta|\mathbf{y},\mathbf{x},\mathcal{D}} E_{\mathbf{y}|\mathbf{x},\mathcal{D}} \left[\log \frac{p(\theta|\mathbf{y},\mathbf{x},\mathcal{D})}{p(\theta|\mathcal{D})} \right], \tag{3.5}$$

where $\mathbf{x} \in \mathbb{R}^{K \times M}$ consists of the media investments of all M channels over the experimental period of K weeks, θ denotes for all parameters related to the performance of marketing channels in the study, and $\mathbf{y} \in \mathbb{R}^{K}$ are the response observations generated by the investments $\mathbf{x}$ over the experimental period of K weeks. The optimal media investments $\hat{\mathbf{x}}$ are the ones that maximize the above EIG given a pre-defined experimental space:

$$\begin{aligned} \max_{\mathbf{x}} \quad & U(\mathbf{x}|\mathcal{D}) \\ \text{s.t.} \quad & \sum_{t=1}^{K} \sum_{m=1}^{M} x_{t,m} \leq B \\ & 0 \leq x_{t,m} \leq ub_{t,m} \quad \forall t \in \{1, \cdots, K\} \text{ and } m \in \{1, \cdots, M\}, \end{aligned} \tag{3.6}$$

where B is the total budget allowed for all M channels during the K-week experimental period and $ub_{t,m}$ is the upper bound representing the maximum investment value allowed for Channel m at week t of the experiment. Based on the Bayes theorem and the assumption that all K observations encapsulated in the target variable $\mathbf{y}$ are independent of each other, Equation (3.5) of the EIG can be reformulated as:

$$\begin{aligned} U(\mathbf{x}|\mathcal{D}) &= E_{\theta|\mathbf{y},\mathbf{x},\mathcal{D}} E_{\mathbf{y}|\mathbf{x},\mathcal{D}} \left[\log \frac{p(\theta|\mathbf{y},\mathbf{x},\mathcal{D})}{p(\theta|\mathcal{D})} \right] \\ &= E_{\theta|\mathbf{y},\mathbf{x},\mathcal{D}} E_{\mathbf{y}|\mathbf{x},\mathcal{D}} \left[\log \frac{p(\theta,\mathbf{y}|\mathbf{x},\mathcal{D})}{p(\theta|\mathcal{D})p(\mathbf{y}|\mathbf{x},\mathcal{D})} \right] \\ &= E_{\theta|\mathbf{y},\mathbf{x},\mathcal{D}} E_{\mathbf{y}|\mathbf{x},\mathcal{D}} \left[\log \frac{p(\mathbf{y}|\theta,\mathbf{x},\mathcal{D})}{p(\mathbf{y}|\mathbf{x},\mathcal{D})} \right] \\ &= E_{\theta|\mathbf{y},\mathbf{x},\mathcal{D}} E_{\mathbf{y}|\mathbf{x},\mathcal{D}} \left[\log \frac{\prod_{t=1}^{K} p(y_t|\theta,\mathbf{x},\mathcal{D})}{\prod_{t=1}^{K} p(y_t|\mathbf{x},\mathcal{D})} \right], \end{aligned}$$

which is estimated using the Nested Monte Carlo (NMC) method with $N \times S$ samples of the model's parameter θ and N samples of the target response y_t for each week t during the K-week experimental period (Myung et al., 2013; Foster et al., 2019):

$$\mathrm{U}(\mathbf{x}|\mathcal{D}) \approx \frac{1}{N} \sum_{n=1}^{N} \log \frac{\prod_{t=1}^{K} p(y_{t,n}|\theta_{n,0},\mathbf{x})}{\frac{1}{S} \sum_{s=1}^{S} \prod_{t=1}^{K} p(y_{t,n}|\theta_{n,s},\mathbf{x})}, \tag{3.7}$$

where $\theta_{n,s} \sim p(\theta|\mathcal{D})$ is a sample of the model's parameter θ and $y_{t,n} \sim p(y_{t,n}|\theta = \theta_{n,0}, \mathbf{x})$ is a generated target response of week t in the experimental period given the investment $\mathbf{x}$ and a sample of the model's parameter $\theta_{n,0}$. Since $p(\theta|\mathcal{D})$ is the posterior distribution of the model's parameters θ after having observed the dataset $\mathcal{D}$, the samples $\theta_{n,s}$ are essentially the posterior samples drawn from the NUTS algorithm used to fit the MMM. Therefore, those posterior samples are reused to simulate future observations, evaluate the likelihoods, and estimate the utility function. This approach helps reduce the computational cost when estimating the utility function, as no additional sampling is required beyond the original Bayesian inference step. Mathematically, given Equation (3.7), the optimal media investments $\hat{\mathbf{x}}$ in Equation (3.6) can be elaborated as:

$$\hat{\mathbf{x}} = \arg\max U(\mathbf{x}|\mathcal{D}) \approx \arg\max \left(\frac{1}{N} \sum_{n=1}^{N} \log \frac{\prod_{t=1}^{K} p(y_{t,n}|\theta_{n,0}, \mathbf{x})}{\frac{1}{S} \sum_{s=1}^{S} \prod_{t=1}^{K} p(y_{t,n}|\theta_{n,s}, \mathbf{x})} \right). \tag{3.8}$$

Equation (3.8) defines an optimization problem over the space of media investments $\mathbf{x}$, and the utility function is optimized using the Constrained Optimization By Linear Approximation (COBYLA) algorithm (Powell, 1994) with the implementation in the Python library SciPy (Virtanen et al., 2020). The experimental period K is set to 4 weeks in all simulated scenarios presented in Chapter 4.

Importantly, the experimental design in the thesis is not optimized with respect to all parameters in the model. Instead, it is only optimized with respect to two subsets of parameters directly related to two core assumptions made in the model regarding the marketing performance: (i) the saturation and shape parameters representing the diminishing returns and (ii) the retention rate parameter representing the carryover effects. This restriction is motivated by the fact that these two subsets are both of primary interest and most directly influenced by experimental budget variation, whereas optimizing over all parameters would increase computational complexity while yielding limited additional learning for decision-making. All remaining non-marketing parameters are fixed at their posterior means when evaluating the utility function defined in Equation (3.8).

3.3 Synthetic Data

The synthetic dataset used in the thesis is generated with $T = 104$ weeks, $C = 1$ control variable representing seasonality, and $M = 4$ marketing channels. Seasonality is generated as a sinusoidal wave (Jin et al., 2017; Sun et al., 2017; Zhang et al., 2024) with added Gaussian noise:

$$z_t = \cos\left(2\pi(t-12)/52\right) + \epsilon_{z,t}, \tag{3.9}$$

where, at time t, z_t is the control variable seasonality and $\epsilon_{z,t}$ is the white noise such that $\epsilon_{z,t} \sim \mathcal{N}(0,1)$. The media investment of each channel is assumed to be correlated with the seasonality, reflecting real-life scenarios in which marketing budgets increase during periods of high expected demand and vice versa (Jin et al., 2017; Sun et al.,

Table 3.2: Marketing channel parameters used to generate the synthetic data.

	Retention rate (r_m)	Saturation (α_m)	Shape (β_m)
Channel 1	0.2	5.0	3.0
Channel 2	0.2	2.0	1.0
Channel 3	0.8	6.0	3.0
Channel 4	0.8	2.0	1.0

Table 3.3: Non-marketing parameters used to generate the synthetic data.

Parameter name	Parameter symbol	Values
Control effect	γ_c	Fixed at 1.0
Baseline	b_t	Fixed at 10.0
Stochastic noise	ϵ_t	Drawn from $\mathcal{N}(0, 3)$

2017; Zhang et al., 2024). The media spends on marketing channels in the synthetic data are then generated with:

$$x_{t,m} = u_m + \rho_m z_t + \sqrt{1 - \rho_m^2}\epsilon_{m,t}, \tag{3.10}$$

where $x_{t,m}$ is the generated investment of Channel m at time t, u_m is the lower bound of the investment values that is set to 2.0 for all channels in the dataset, ρ_m is the correlation between the media spends and the seasonality z_t such that $\rho_m \sim \text{Uniform}(0, 0.8)$, and $\epsilon_{m,t}$ is the added white noise such that $\epsilon_{m,t} \sim \mathcal{N}(0, 1)$. The synthetic data contains $M = 4$ marketing channels with distinct combinations of retention rates, saturation levels, and shape parameters:

1. Low retention rate, high saturation, and high shape
2. Low retention rate, low saturation, and low shape
3. High retention rate, high saturation, and high shape
4. High retention rate, high saturation, and high shape

These configurations are chosen to represent heterogeneous channel behaviors in the mix, including differences in persistence, scalability, and responsiveness to investment. Table 3.2 presents the details of each channel's parameters generated in the synthetic data. These parameters are then used in Equation (3.3) along with a control effect of 1.0, a fixed baseline of 10.0, and an added noise $\epsilon \sim \mathcal{N}(0, 3)$ to produce the target metric y_t. Table 3.3 summarizes the values of non-marketing parameters used to generate the synthetic data.

The resulting dataset of 104 weeks is treated as historical data used to fit the MMM. Additionally, 17 weeks of data are also generated using the same process above. This additional period is not used during the model fitting step but, instead, simulates the future marketing budget planning and its corresponding observations that help define the experiment space, guide the optimization of the utility function within the BED framework, and analyze the results in the simulated scenarios discussed in Chapter 4. This additional 17-week period of data will be referred to as simulated future budget planning in the thesis.

Figure 3.1 visualizes both the historical data over 104 weeks and the future budget planning over 17 weeks with the control variable, total media investments, and the target KPI variable. Figure 3.1 shows that the control variable affects the target KPI variable both directly as the seasonal effects, and indirectly via the generated media investments. Figure 3.2 decomposes the proportion of media investments of each channel in the synthetic data for each week. Figure 3.2 indicates that the shares of investments are not stable over time, as different channels have different levels of correlation with the control variable in terms of their investments.

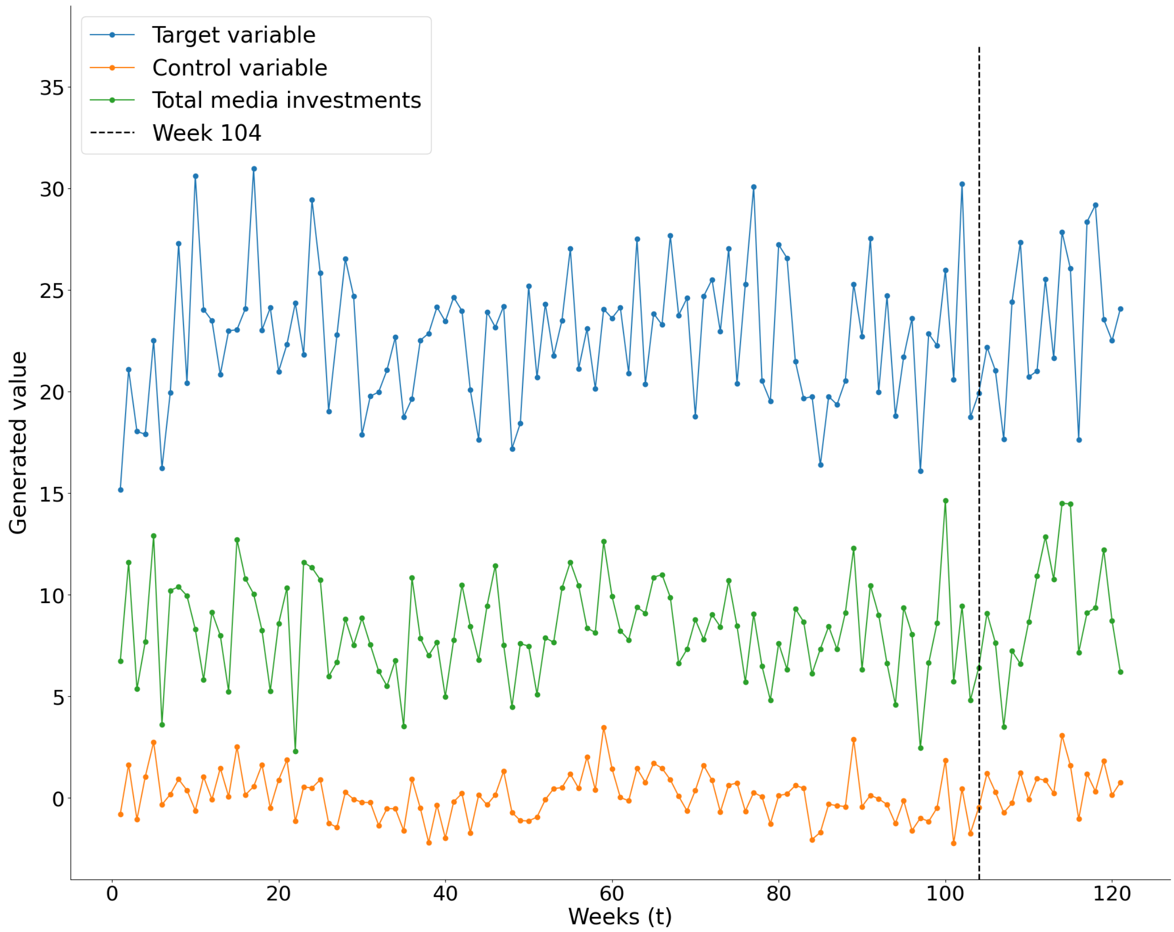

Figure 3.1: Generated time-series synthetic data containing 104 weeks of historical data and 17 weeks of future budget planning. The orange line represents the control variable seasonality z_t, the green line represents the total media investments of $M = 4$ media channels $\sum_{m=1}^{M} x_{t,m}$, and the blue line represents the target KPI variable y_t. The vertical black dashed line marks Week 104 when the historical data ends.

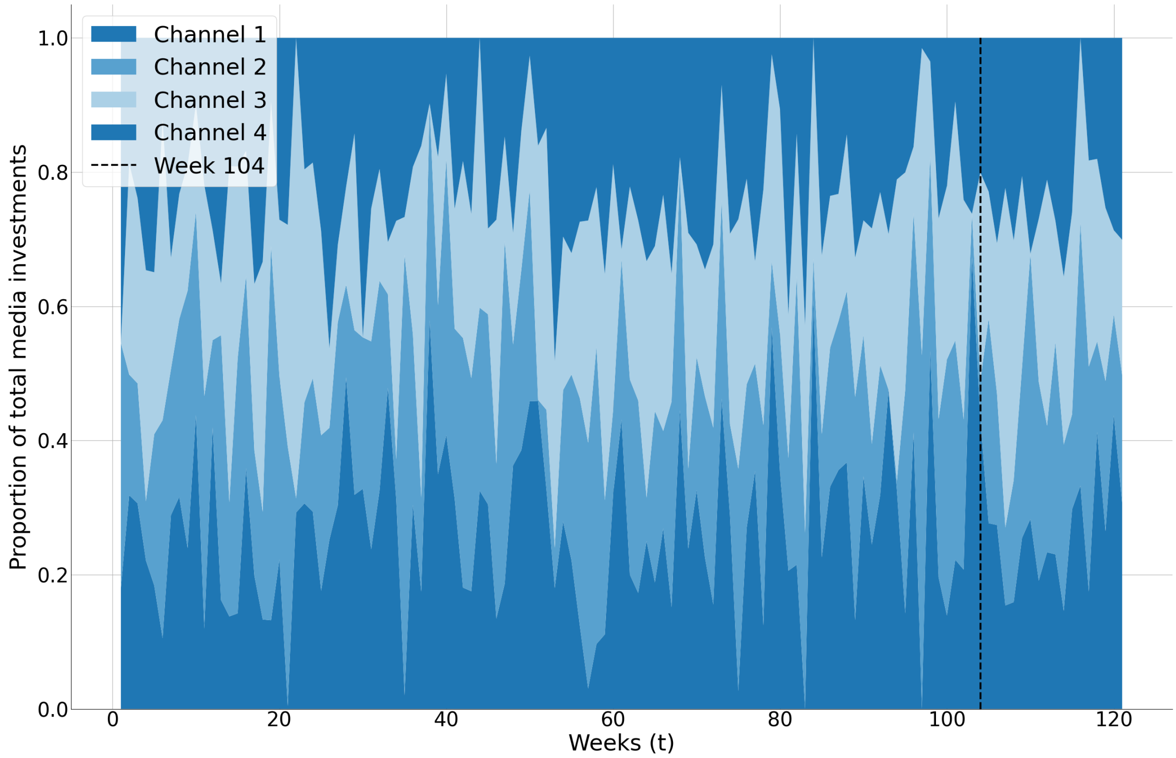

Figure 3.2: The decomposed media investment proportion in each week of four channels over 104 weeks of historical data and 17 weeks of future budget planning. From the bottom up to the top, the areas represent the investments of Channels 1, 2, 3, and 4, respectively. The black dashed line marks Week 104 when the historical data ends.

4. Experiments & Results

This chapter presents simulated experiments that apply the Bayesian experimental design (BED) framework to marketing mix modeling (MMM). Two settings of the experiments are considered. First, Section 4.1 studies the case where the performance of one channel in the mix (Channel 1) is uncertain while the performance of all other channels (Channels 2, 3, and 4) is certain. Second, Section 4.2 considers the scenario where the performance of all channels in the mix is uncertain. Prior distributions of the MMM are utilized to represent such performance uncertainty. Specifically, uncertain channels have weak prior distributions and large prior standard deviations for their marketing parameters and vice versa. The two scenarios considered in this chapter aim to study the design of information-optimal investment allocations under different states of knowledge and confidence levels about marketing performance.

In both settings, the workflow is the same. First, a MMM is fitted to the historical synthetic data consisting of 104 weeks to establish the pre-experiment posterior distributions that reflect the current state of knowledge. Then, the BED framework is applied using the pre-experiment posterior samples to select the optimal media investment allocation over the experimental period of 4 weeks to maximize the expected information gain (EIG) for a targeted subset of parameters. Two subsets of parameters are considered: (i) saturation and shape parameters, representing the diminishing returns, and (ii) the retention rate parameter, representing the carryover effects. Finally, the recommended experimental investments are utilized as the investments of 4 subsequent time points (Weeks 105 - 108), and, using the same prior specification as the pre-experiment model's, two post-experiment MMMs are refitted with both historical data and simulated future budget planning: one with the experimental investments and the other without the experimental investments. These with-experiment and without-experiment models are then compared against each other to evaluate how the proposed experiment changes the uncertainty and posterior distributions of the marketing parameters.

4.1 One Uncertain Channel

This section considers the scenario where the performance of Channel 1 is uncertain while the performance of Channels 2, 3, and 4 is assumed to be known with high confidence. The scenario is intended to represent a practical setting in which marketing practitioners have a well-established understanding of most channels in the mix, with uncertainty remaining for only a limited subset of channels. This section contains three subsections. Subsection 4.1.1 describes the pre-experiment MMM fitting and the resulting posterior distributions of Channel 1's parameters. Subsection 4.1.2 and Subsection 4.1.3 then apply the BED framework to find the optimal experimental investment allocation with respect to (i) the saturation and shape parameters and (ii) the retention rate parameter, respectively, to enhance the current state of knowledge about the performance of Channel 1. In this case, the BED framework is limited to optimizing the marketing investments of Channel 1 only, while the investments of the other three channels are kept fixed during the experimental period using the investments in simulated future budget planning.

4.1.1 Pre-experiment MMM

The synthetic dataset is generated from known hyperparameters, which are treated as the ground truth values for channel performance as summarized in Table 3.2. Based on these ground truth values, the prior distributions are constructed to reflect the intended knowledge state of different channels in the mix. To be more specific, the saturation and shape parameters of Channels 2, 3, and 4 are assigned informative normal distributions centered at the ground truth values themselves with standard deviation 0.1, and their retention rate parameters are assigned uniform distributions with lower bounds 0.01 lower than the ground truth values and upper bounds 0.01 higher than the ground truth values. On the other hand, Channel 1 is assigned weaker priors. Its saturation parameter is assigned a normal distribution centered at 7.0 while the ground truth is 5.0, and the shape parameter is assigned a normal distribution centered at 5.0 while the ground truth is 3.0. The priors of both the saturation and shape parameters of Channel 1 have standard deviations of 3.0, representing limited certainty about the channel's diminishing return behavior. Channel 1's retention rate parameter is assigned a wide uniform distribution with lower bound 0.1 and upper bound 0.9, reflecting that it is unknown about Channel 1's carryover effects.

For the remaining components in the model, priors are assigned based on the hyperparameters used in the synthetic data generation process, summarized in Table 3.3. Specifically, the control effect γ_c follows a Gaussian prior with mean 1.0 and

Table 4.1: The prior specifications of the MMM in the scenario "one uncertain channel". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation. Stochastic noise does not have a ground truth value, and its values are drawn from the same distribution used in the synthetic data generation.

Parameter name	Parameter symbol	Distributions
Retention rate	r_1 (0.2)	Uniform$(0.1, 0.9)$
	r_2 (0.2)	Uniform$(0.19, 0.21)$
	r_3 (0.8)	Uniform$(0.79, 0.81)$
	r_4 (0.8)	Uniform$(0.79, 0.81)$
Shape	β_1 (3.0)	TruncatedNormal$(5, 3^2)$
	β_2 (1.0)	TruncatedNormal$(1, 0.1^2)$
	β_3 (3.0)	TruncatedNormal$(3, 0.1^2)$
	β_4 (1.0)	TruncatedNormal$(1, 0.1^2)$
Saturation	α_1 (5.0)	TruncatedNormal$(7, 3^2)$
	α_2 (2.0)	TruncatedNormal$(2, 0.1^2)$
	α_3 (6.0)	TruncatedNormal$(6, 0.1^2)$
	α_4 (2.0)	TruncatedNormal$(2, 0.1^2)$
Control effect	γ_c (1.0)	Normal$(1, 1^2)$
Baseline	b_t (10.0)	Normal$(10, 0.1^2)$
Stochastic noise	ϵ_t	Normal$(0, 3^2)$

standard deviation 1.0; the baseline b_t is more informative with mean 10.0 and standard deviation 0.1; the stochastic noise ϵ_t has mean 0.0 and standard deviation 3.0. Table 4.1 summarizes the prior specification used in this experiment, along with each parameter's ground truth value.

The MMM is fitted using the NUTS algorithm (Hoffman et al., 2014) with 5000 tuning iterations and 15000 iterations as the posterior draws. Table 4.2 reports posterior summary statistics for the key performance parameters of Channel 1: saturation, shape, and retention rate. Figure 4.1 shows the posterior distributions of those parameters together with their ground truth values used in synthetic data generation ($\alpha_1 = 5.0$, $\beta_1 = 3.0$, and $r_1 = 0.2$). All three posterior distributions place non-negligible probability densities around the ground truth. However, the posteriors remain relatively diffuse, and the ground truth values are not yet assigned the highest probability densities in the distributions. This observation is consistent with the intentionally simulated weak state of knowledge for Channel 1.

Out of three parameters representing the performance of a marketing channel, saturation and shape parameters jointly define the saturation curve of a channel, representing diminishing returns. Figure 4.2 visualizes the posterior predictive uncertainty

Table 4.2: The summary statistics of the pre-experiment posterior distributions of Channel 1's parameters in the scenario "one uncertain channel". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation.

Parameter	Mean	Standard deviation	94% HDI
α_1 (5.0)	7.959	1.916	4.446 – 11.526
β_1 (3.0)	6.319	2.067	2.499 – 10.189
r_1 (0.2)	0.329	0.145	0.1 – 0.578

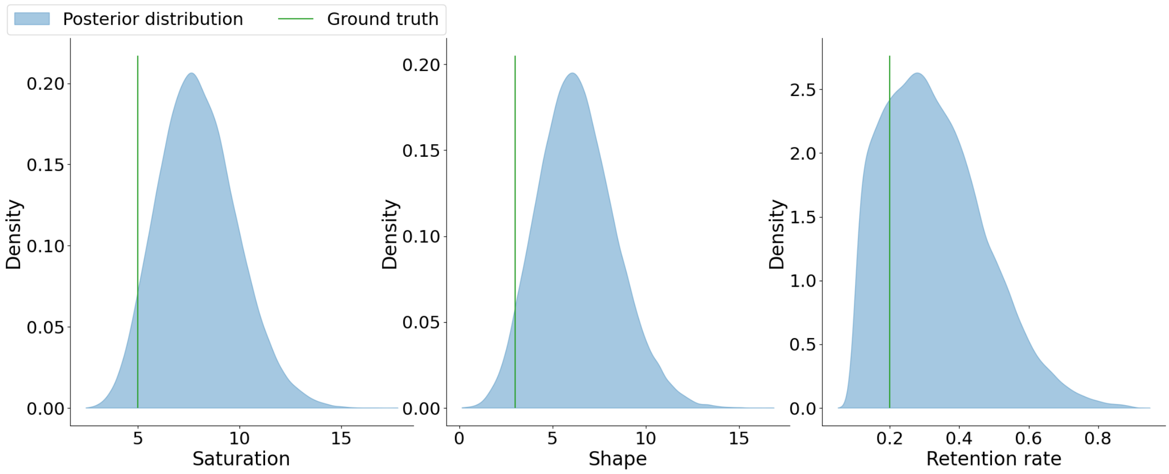

Figure 4.1: Posterior distributions of the pre-experiment MMM in the scenario "one uncertain channel" and the ground truth values of Channel 1's three parameters. The blue areas represent the posterior distributions of, from left to right, saturation, shape, and retention rate, respectively. The green vertical lines represent the ground truth values of the corresponding parameters, which are the hyperparameters used in the synthetic data generation. Channel 1 data is generated with saturation $\alpha_1 = 5$, shape $\beta_1 = 3$, and retention rate $r_1 = 0.2$.

of Channel 1's saturation curve along with the ground truth curve computed from the hyperparameters used in the synthetic data generation. The ground truth curve lies within the 94% HDI, indicating that the fitted model is broadly consistent with the true data-generating process. However, the ground truth curve is not the mode of the posterior distributions, especially for low investments closer to zero or high investments close to the saturated level.

With the pre-experiment MMM fitted, the BED framework is then applied to propose experimental investments for 4 subsequent weeks: Week 105, 106, 107, and 108. In the design optimization, based on the simulated future budget planning, the investments of Channels 2, 3, and 4 are kept constant at their planned budgets for the respective weeks, while the investments of Channel 1 are optimized to maximize the utility function. Channel 1's investment level is constrained to be non-negative and at the maximum two times its planned budget for the experimental period from Week

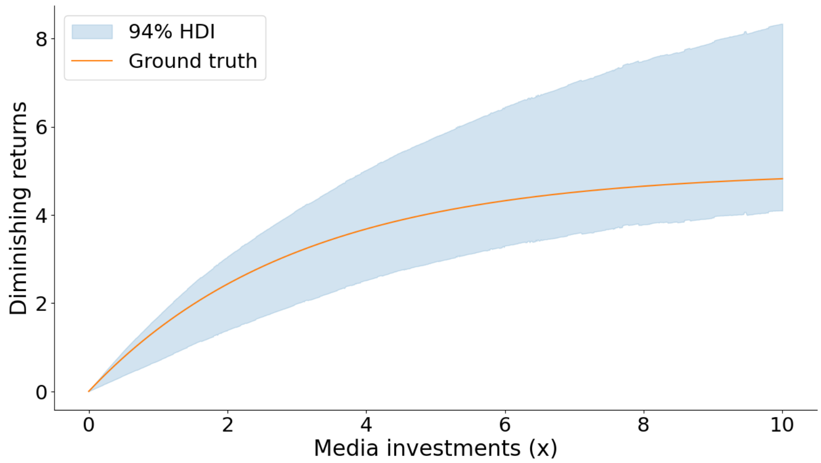

Figure 4.2: Posterior saturation curve of Channel 1 along with the ground truth curve. The blue area represents the posterior 94% HDI of the saturation curve. The orange line represents the ground truth curve, which is calculated using the hyperparameters used in the synthetic data generation. Channel 1 data is generated with saturation $\alpha_1 = 5.0$ and shape $\beta_1 = 3.0$.

105 to Week 108.

Two cases are considered when optimizing the experimental budget allocation. The first case involves optimizing the experiments with respect to the saturation curve's parameters: saturation and shape, and the second case involves optimizing with respect to the carryover effects parameter: retention rate. The resulting recommended investments are used as the investments of the period from Week 105 to Week 108. Two post-experiment MMMs are then refitted: one with the experimental investments and the other without the experimental investments. Both of them are assigned the same prior specification as the pre-experiment model as presented in Table 4.1.

4.1.2 Post-experiment optimized for saturation and shape

Figure 4.3 presents the planned budget of Channel 1 and its experimental investments when optimizing the allocation with respect to the saturation and shape parameters. With this allocation, the optimal experiment places all of Channel 1's budget for the 4-week experimental period in the first week and forces the investments of the last three weeks to zero. Intuitively, the anomalously high investment in Week 105 helps gain insights about the saturation parameter, as such a high investment is likely to reach the saturated level of Channel 1. On the other hand, the following zero investments for the last three weeks help separate the carryover effects from the diminishing returns. Such zero investments could also help gain insights about the retention rate parameter, even though the experiment is not optimized with respect to the retention rate. However, the information gained might be limited as a substantial amount of marketing effects can

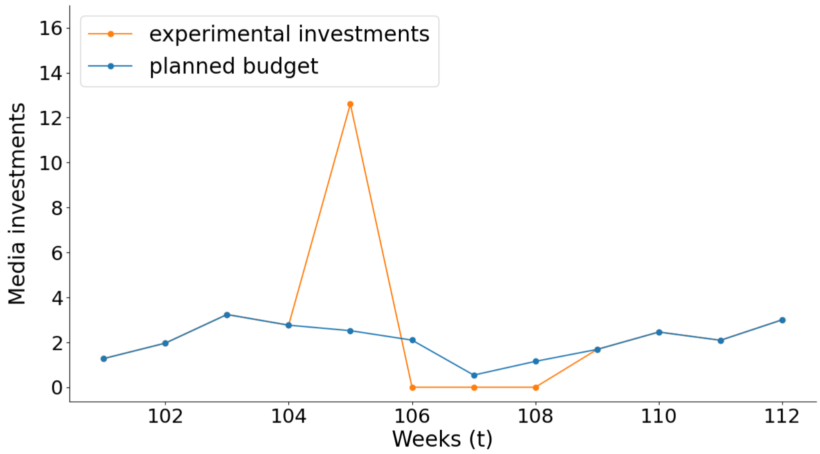

Figure 4.3: Channel 1's planned budget and experimental investments optimized with respect to the saturation and shape parameters in the scenario "one uncertain channel". The blue line represents the planned budget of Channel 1, while the orange line represents the experimental investments during the experimental period: Weeks 105 - 108.

Table 4.3: The summary statistics of the post-experiment posterior distributions of Channel 1's parameters from two MMMs fitted with and without experimental investments optimized for the saturation and shape parameters in the scenario "one uncertain channel". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation.

Parameter	Model	Mean	Standard deviation	94% HDI
α_1 (5.0)	with exp.	7.092	1.696	4.037 – 10.252
	without exp.	7.791	1.897	4.333 – 11.3
β_1 (3.0)	with exp.	5.356	2.041	1.715 – 9.117
	without exp.	6.236	2.103	2.273 – 10.074
r_1 (0.2)	with exp.	0.296	0.134	0.1 – 0.530
	without exp.	0.323	0.150	0.1 – 0.585

be carried over from the atypically high investment in the first week of the experimental period.

These values from the optimal design replace the originally generated investments of Channel 1 during the same four weeks in the planned budget, while other channels' investments remain unchanged. Two new MMMs are then refitted with both historical data and simulated future budget planning in the synthetic dataset. One model is fitted with the experimental investments for Channel 1 from Week 105 to Week 108, while the other model is fitted with the originally generated dataset. Both of these models use the same prior specifications presented in Table 4.1.

Table 4.3 summarizes the resulting post-experiment posterior distributions of Channel 1's parameters for the above two models. Compared to the model fitted without the experimental investments, the uncertainty in the saturation and shape

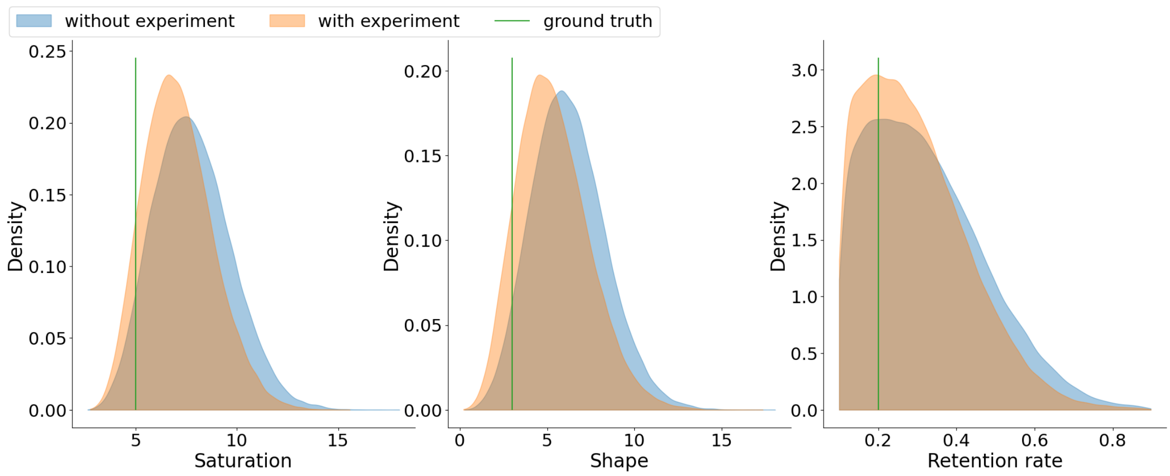

Figure 4.4: Posterior distributions of the models fitted with and without the experiment optimized for the saturation and shape parameters, along with the ground truth values of, from left to right, Channel 1's saturation, shape, and retention rate parameters, respectively. The blue areas represent the posterior distributions without the experiment, while the orange areas represent the ones with the experiment. The green vertical lines represent the ground truth values of the corresponding parameters, which are the hyperparameters used in the synthetic data generation.

parameters of the with-experiment model slightly decreases; standard deviations reduce from 1.897 to 1.696 for the saturation α_1 and from 2.103 to 2.041 for the shape β_1. The retention rate also shows a slight decrease in the uncertainty for the model fitted with the experiment (standard deviation reduces from 0.15 to 0.134). Moreover, the with-experiment model has posterior means for Channel 1's parameters closer to the ground truth values than the without-experiment model does, given that the ground truth values of Channel 1's parameters are $\alpha_1 = 5.0$ for the saturation, $\beta_1 = 3.0$ for the shape, and $r_1 = 0.2$ for the retention rate.

Figure 4.4 visualizes the differences in posterior distributions between these two models using the posterior distributions for three marketing parameters of Channel 1 obtained from the model fitted with the experiment and the one fitted without the experiment, along with the ground truth values of the parameters used in the synthetic data generation. The posterior distributions of the with-experiment model are shown to shift towards the ground truth values for both the saturation and shape parameters and assign higher probability densities around the ground truth compared to the model fitted without the experiment. Figure 4.4 also indicates that the posterior distribution of the retention rate parameter from the with-experiment model also shifts towards the ground truth, and the distribution is more concentrated around the ground truth value 0.2 compared to the without-experiment model.

Figure 4.5 visualizes the posterior saturation curves of Channel 1 obtained from the models fitted with and without the experiment, along with the ground truth curve

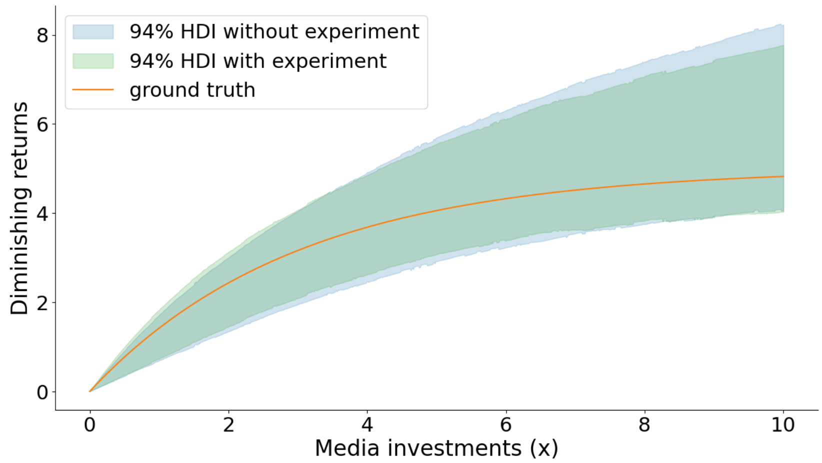

Figure 4.5: Posterior saturation curves of Channel 1 with and without the experiment optimized for the saturation and shape parameters, along with the ground truth curve in the scenario "one uncertain channel". The blue area represents the posterior 94% HDI of the saturation curve without the experiment, while the green area represents the posterior 94% HDI of the curve with the experiment. The orange line represents the ground truth curve, which is calculated with the hyperparameters used in the synthetic data generation.

calculated from the hyperparameters used in the synthetic data generation. The saturation curve from the with-experiment model is shown to have a narrower posterior HDI than the without-experiment model's posterior curve, especially for high investment levels. This indicates that the uncertainty of the saturation curve is reduced in the model fitted with experimental data compared to the model fitted without it. Such a reduction in uncertainty can be attributed to the atypically high investment of 12.6 in the first week of the experimental period. For low investment levels, the posterior HDI of the model fitted with the experiment shifts slightly upwards to higher values for the diminishing returns along the y-axis. Even though this does not necessarily suggest a reduction in uncertainty for the saturation curve around low investment levels, this shows that the posterior saturation curve obtained from the model fitted with the experiment shifts closer to the ground truth for the low investment levels than the one fitted without the experiment.

To summarize, in the scenario where only the performance of Channel 1 is uncertain, the optimal investment allocation optimized by the BED framework with respect to the saturation and shape parameters recommends placing all of the experimental budget in the first week for Channel 1 and forcing the investments of the last three weeks to zero. Such an allocation is shown to reduce the uncertainty in all three Channel 1's marketing parameters based on the comparison between the MMM fitted with the experiment and the MMM fitted without the experiment. The with-experiment model also shows a shift of posterior distributions towards the ground truth values of

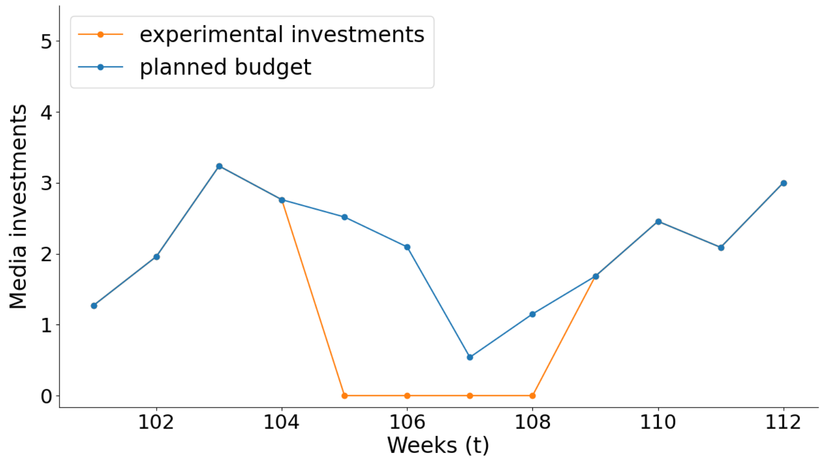

Figure 4.6: Channel 1's planned budget and experimental investments optimized with respect to the retention rate parameter in the scenario "one uncertain channel". The blue line represents the planned budget of Channel 1, while the orange line represents the experimental investments during the experimental period: Weeks 105 - 108.

the mentioned marketing parameters compared to the without-experiment MMM.

4.1.3 Post-experiment optimized for retention rate

Figure 4.6 presents the planned budget of Channel 1 and its experimental investments when optimizing the experiments with respect to the retention rate parameter. In this case, the optimal design recommends forcing the investments of Channel 1 to zero during the whole experimental period from Week 105 to Week 108. Given the assumptions made in the MMM, the marketing signals during the weeks with zero investments are all carried over from past investments. In other words, the target metrics driven by marketing during those weeks are mainly controlled by carryover effects encapsulated by the retention rate parameter. Therefore, intuitively, forcing the investments of Channel 1 to zero for the whole experimental period helps gain insights about the retention rate parameter. On the other hand, zero-investment observations might not help enhance the current knowledge about the saturation and shape parameters, as, based on the INE function, the diminishing returns of zero investments are always zero regardless of what values the saturation and shape parameters have.

These zero-investment values from the optimal design replace the originally generated investments of Channel 1 for those four weeks in the future budget, while other channels' investments remain unchanged. Two new MMMs are then refitted with both historical data and simulated future budget planning in the synthetic dataset. One model is fitted with the experimental investments for Channel 1 from Week 105 to Week 108, while the other model is fitted with the originally generated dataset. Both of these models use the same prior specifications presented in Table 4.1.

Table 4.4: The summary statistics of the post-experiment posterior distributions of Channel 1's parameters from two MMMs fitted with and without experimental investments optimized for the retention rate parameter in the scenario "one uncertain channel". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation.

Parameter	Model	Mean	Standard deviation	94% HDI
α_1 (5.0)	with exp.	8.038	1.882	4.616 − 11.560
	without exp.	7.791	1.897	4.333 − 11.3
β_1 (3.0)	with exp.	6.035	2.006	2.370 − 9.818
	without exp.	6.236	2.103	2.273 − 10.074
r_1 (0.2)	with exp.	0.283	0.127	0.1 − 0.505
	without exp.	0.323	0.150	0.1 − 0.585

Table 4.4 summarizes the resulting post-experiment posterior distributions of Channel 1's parameters: saturation, shape, and retention rate for the above two models. Compared to the model fitted without the experimental investments, the uncertainty in the saturation and shape parameters of the model with the experimental investments remains roughly the same; standard deviations change from 1.897 to 1.882 for the saturation α_1 and from 2.103 to 2.006 for the shape β_1. The retention rate also shows a slight decrease in the uncertainty for the model fitted with the experimental investments (standard deviation reduces from 0.15 to 0.127). On the other hand, the with-experiment model has the posterior mean farther from the ground truth of 5.0 for the saturation, closer to the ground truth of 3.0 for the shape, and closer to the ground truth of 0.2 for the retention rate.

Figure 4.4 visualizes the differences in posterior distributions between these two models using the posterior distributions for all three parameters of Channel 1 obtained from the model with and the one without the experimental investments, along with the ground truth values of the parameters used in the synthetic data generation. The posterior distributions of both saturation and shape parameters of the with-experiment model are shown not to have noticeable changes compared to the model fitted without the experiment. Meanwhile, the model fitted with the experiment has higher posterior probability concentration for the retention rate parameter around the ground truth than the model fitted without the experiment. Figure 4.4 also indicates that the posterior distribution with the experiment shifts towards the ground truth value of the retention rate, placing a higher probability density on the ground truth.

Figure 4.8 visualizes the posterior saturation curves of Channel 1 obtained from the models fitted with and without the experiment, along with the ground truth curve calculated from the hyperparameters used in the synthetic data generation. The posterior HDI of the saturation curve from the model fitted with the experiment is shown to

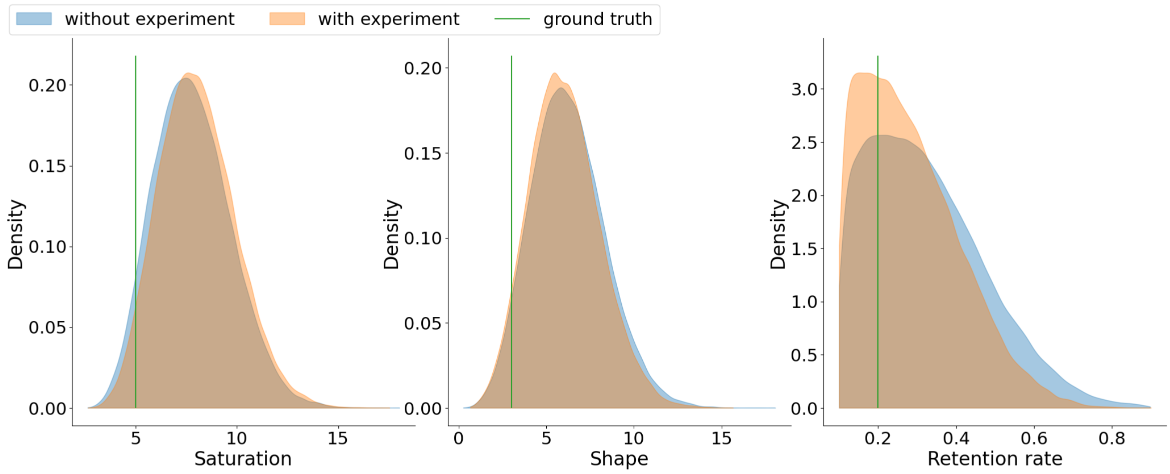

Figure 4.7: Posterior distributions of the models fitted with and without the experiment optimized for the retention rate parameter, along with the ground truth values of, from left to right, Channel 1's saturation, shape, and retention rate parameters, respectively. The blue areas represent the posterior distributions without the experiment, while the orange areas represent the ones with the experiment. The green vertical lines represent the ground truth values of the corresponding parameters, which are the hyperparameters used in the synthetic data generation.

shift to the higher diminishing returns along the y-axis, indicating a narrower posterior curve. However, as previously mentioned, the investments suggested by the experiment are all zeroes for the whole experimental period. Mathematically, the saturation curve returns 0.0 as the diminishing return at 0.0 investment regardless of what values the saturation and shape parameters have, so an observation at (or near) zero investment provides limited information about the diminishing returns behaviors.

To summarize, in the scenario where only the performance of Channel 1 is uncertain, the optimal investment allocation optimized for the retention rate parameter recommends forcing the investments of Channel 1 to zero during the whole 4-week experimental period. Such an allocation helps reduce the uncertainty in the retention rate parameter and shift the posterior distributions towards their ground truth values. However, the effects of the experiment on the saturation and shape parameters are inconsistent regarding both uncertainty and the ground truth value probability. This is expected as the experiment is optimized with respect to the retention rate parameter, but not the saturation and shape parameters.

4.2 All Uncertain Channels

This section considers the scenario where the performance of all four channels is uncertain. The scenario is meant to represent a practical setting where marketing practitioners do not have a well-established understanding of most channels in the mix. Similar

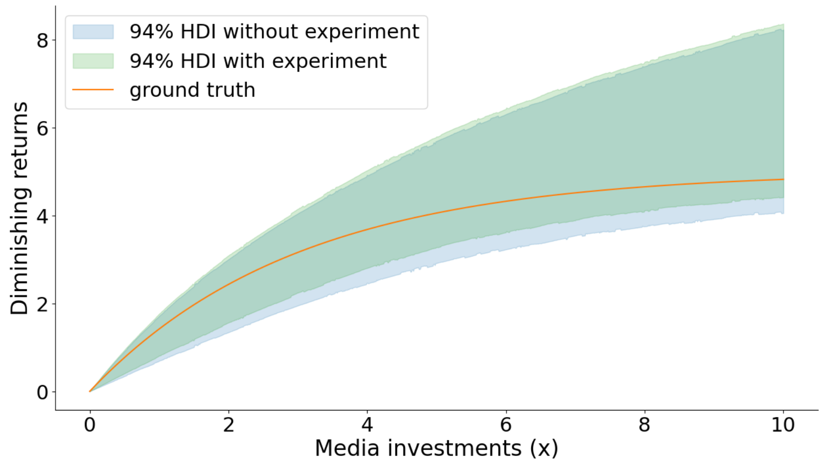

Figure 4.8: Posterior saturation curves of Channel 1 with and without the experiment optimized for the retention rate parameter, along with the ground truth curve in the scenario "one uncertain channel". The blue area represents the posterior 94% HDI of the saturation curve without the experiment, while the green area represents the posterior 94% HDI of the curve with the experiment. The orange line represents the ground truth curve, which is calculated with the hyperparameters used in the synthetic data generation.

to Section 4.1, the uncertainty is reflected using the prior distributions of the model, and every channel is assigned the same broad uncertainty for saturation, shape, and retention rate. This section contains three sub-sections. Subsection 4.2.1 describes the pre-experiment MMM fitting and the resulting posterior distributions of all channels. Subsection 4.2.2 and Subsection 4.2.3 then apply the BED framework to design the optimal experiments targeting two subsets of parameters: (i) the saturation and shape parameters, and (ii) the retention rate parameter, respectively. In this case, the BED framework is set to optimize the experiments based on the marketing investments of all four channels in the mix.

4.2.1 Pre-experiment MMM

To represent the low prior certainty across the entire media mix, all channels share the same prior distributions for the saturation and shape parameters. More specifically, the saturation parameter α_m (for any $m \in \{1, 2, 3, 4\}$) is assigned a truncated normal distribution with mean 4.0 and standard deviation 1.5. The prior distribution for the same parameter also follows a truncated normal distribution with mean 2.0 and standard deviation 1.5. On the other hand, some prior knowledge of the carryover effects is assumed in this scenario: Channels 1 and 2's retention rate parameters are assigned a prior uniform distribution for short carryover channels with lower bound 0.1 and upper bound 0.5, while Channels 3 and 4's retention rates are assigned a prior uniform distribution for long carryover channels with lower bound 0.5 and upper bound

Table 4.5: The prior specifications of the MMM in the scenario "all uncertain channels". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation. Stochastic noise does not have a ground truth value, and its values are drawn from the same distribution used in the synthetic data generation.

Parameter name	Parameter symbol	Distributions
Retention rate	r_1 (0.2)	$\text{Uniform}(0.1, 0.5)$
	r_2 (0.2)	$\text{Uniform}(0.1, 0.5)$
	r_3 (0.8)	$\text{Uniform}(0.5, 0.9)$
	r_4 (0.8)	$\text{Uniform}(0.5, 0.9)$
Shape	β_1 (3.0)	$\text{TruncatedNormal}(2.0, 1.5^2)$
	β_2 (1.0)	$\text{TruncatedNormal}(2.0, 1.5^2)$
	β_3 (3.0)	$\text{TruncatedNormal}(2.0, 1.5^2)$
	β_4 (1.0)	$\text{TruncatedNormal}(2.0, 1.5^2)$
Saturation	α_1 (5.0)	$\text{TruncatedNormal}(4.0, 1.5^2)$
	α_2 (2.0)	$\text{TruncatedNormal}(4.0, 1.5^2)$
	α_3 (6.0)	$\text{TruncatedNormal}(4.0, 1.5^2)$
	α_4 (2.0)	$\text{TruncatedNormal}(4.0, 1.5^2)$
Control effect	γ_c (1.0)	$\text{Normal}(1, 1^2)$
Baseline	b_t (10.0)	$\text{Normal}(10, 0.1^2)$
Stochastic noise	ϵ_t	$\text{Normal}(0, 3^2)$

0.9.

For other non-marketing components in the model, similar to Section 4.1, the priors are assigned based on the hyperparameters used in the synthetic data generation process summarized in Table 3.3. Specifically, the control effect γ_c follows a normal prior distribution with mean 1.0 and standard deviation 1.0; the baseline b_t is more informative with mean 10.0 and standard deviation 0.1; the stochastic noise ϵ_t has mean 0.0 and standard deviation 3.0. Table 4.5 summarizes the full prior specification used in this scenario along with each parameter's ground truth value.

The MMM is fitted using the NUTS algorithm (Hoffman et al., 2014) with 5000 tuning iterations and 15000 iterations as the posterior draws. Table 4.6 summarizes the posterior distributions for each channel's saturation, shape, and retention rate parameters. Figure 4.9 shows the posterior distributions of those parameters together with their ground truth values used in the synthetic data generation. All fitted posterior distributions cover the ground truth values without placing negligible probability densities around the ground truth. However, the posteriors remain relatively diffuse, and the ground truth values are not yet assigned the highest probability densities in all of the distributions. Especially for the retention rate parameter, the posterior distribu-

Table 4.6: The summary statistics of the pre-experiment posterior distributions of all marketing parameters in the model in the scenario "all uncertain channels". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation.

	Parameter	Mean	Standard deviation	94% HDI
Channel 1	α_1 (5.0)	4.919	1.186	2.707 – 7.149
	β_1 (3.0)	2.893	1.025	1.026 – 4.787
	r_1 (0.2)	0.257	0.107	0.1 – 0.446
Channel 2	α_2 (2.0)	3.636	1.186	1.402 – 5.848
	β_2 (1.0)	2.558	1.213	0.383 – 4.802
	r_2 (0.2)	0.324	0.112	0.136 – 0.5
Channel 3	α_3 (6.0)	4.272	1.211	2.005 – 6.569
	β_3 (3.0)	2.894	1.274	0.402 – 5.213
	r_3 (0.8)	0.689	0.111	0.5 – 0.869
Channel 4	α_4 (2.0)	3.572	1.2	1.366 – 5.881
	β_4 (1.0)	2.444	1.349	0.000 – 4.646
	r_4 (0.8)	0.746	0.108	0.551 – 0.9

tions are not clearly identifiable, representing weak separation between the carryover effects and diminishing returns in attributing the target variable to marketing effects.

Figure 4.10 shows the posterior predictive uncertainty of the saturation curves for all marketing channels in the mix, along with their ground truth curves calculated using the hyperparameters used in the synthetic data generation. The ground truth curves all lie within the posterior 94% HDI, indicating that the fitted model is consistent with the data-generating process.

With the pre-experiment MMM fitted, the BED framework is then applied to propose experimental investments for the subsequent 4-week period: Week 105, 106, 107, and 108. During the design optimization, the investments of all four channels during those four weeks are optimized jointly, each of which is constrained to be non-negative and at a maximum of two times the total budget planned for the mentioned 4-week period. Moreover, based on the simulated future budget planning, the total budget allowed for all channels during the experimental period is also two times the planned budget from Week 105 to Week 108.

Similar to Section 4.1, two cases are considered for the optimal experiments. The first case involves optimizing the experiments for the saturation and shape parameters that jointly define the saturation curve. The second case involves optimizing the retention rate parameter representing the carryover effects. The resulting recommended investments are used as the investments of the period from Week 105 to Week 108.

Two post-experiment MMMs are then refitted: one with the experimental investments and the other without the experimental investments. Both of them are assigned the same prior specification as that used for the pre-experiment model presented in Table 4.5.

4.2.2 Post-experiment optimized for saturation and shape

Figure 4.11 presents the planned budget of all channels in the mix and their corresponding experimental investments when optimizing the experiments with respect to the saturation and shape parameters. With this allocation, the optimal experiment increases the investment of Channels 1 and 2 to a higher level than the planned typical allocation while stopping all of the investments in Channels 3 and 4. Intuitively, high investments help gain insights about the saturation parameter, especially if they are around the saturated levels of the channels, while zero investments help separate the carryover effects from the diminishing returns. With the constrained budget, the optimal experiment reallocates the budgets of Channels 3 and 4 to Channels 1 and 2.

These values from the optimal design replace the originally generated investments of the period from Week 105 to Week 108 in the simulated future budget of all channels in the mix. Two new MMMs are then refitted with both historical data and simulated future budget planning in the synthetic data set. One model is fitted with the experimental investments mentioned above, while the other model is fitted with the originally generated dataset. Both of these models use the same prior specification presented in Table 4.5.

Table 4.7 presents the summary statistics of the resulting post-experiment posterior distributions of all marketing parameters from the above two models. With absolute terms, except for the shape of Channel 4, the posterior standard deviations of saturation and shape parameters of all channels from the model fitted with the experimental investments are reduced compared to those from the model fitted without the experiment. However, the reduction is arguably negligible apart from Channel 2's saturation parameter. The same observation applies to the posterior means. The differences in the posterior means between the two models are negligible except for the saturation parameter of Channel 2. The posterior mean of Channel 2's saturation from the model fitted with the experiment is 3.375, farther from the ground truth value 2.0 than the posterior mean of 2.329 obtained from the model fitted without the experiment.

Figure 4.12 illustrates the posterior distributions of marketing parameters from those two models along with their ground truth values. Some reduction in the uncertainty is shown in the posterior distributions for the saturation and shape parameters as the model fitted with the experiment places has slightly more concentrated distributions than the one fitted without the experiment. On the other hand, there are no clear shifts in the posterior distributions towards the ground truth values in the model fitted with the experiment. Especially for the retention rate parameter, the posterior distributions are indicated to shift away from the ground truth values as the model fitted with the experiment places lower probability densities around the ground truth than the model fitted without the experiment does.

Figure 4.13 visualizes the posterior saturation curves of all channels from the models fitted with and without the experiment, along with their ground truth curves calculated with the hyperparameters used in the synthetic data generation. There is no clear reduction in the uncertainty of the posterior saturation curves between the models fitted with and without the experimental investments. More specifically, the posterior 94% HDIs from two models are visually the same with some small shifts around high investment levels. In other words, the posterior 94% HDIs of the model fitted with the experiment are not systematically narrower than the ones from the model fitted without the experiment.

Based on the observations when comparing the posterior distributions between two models fitted with and without the experiment, the experiment provides limited information and learning, while the marketing parameters, especially the retention rate, in the model remain weakly identified. One possible reason for this is due to the weak prior specification used in this simulated scenario, where all marketing parameters share the same prior distributions with high uncertainty. Therefore, the signals from the 4-week experimental observations are not yet strong enough to guide the posterior distributions towards the desired results. One suggestion to address this situation is to widen the experiment space by either increasing the allowed budget for the experiment or running the experiment for a longer period. However, enlarging the scope of the experiment might not be practical for real-world use cases, as experiments optimized using the BED framework are not business-optimal budget planning but information-optimal stimulus. Therefore, running experiments comes with the risks of high costs and revenue loss for marketing practitioners. Depending on the use cases, marketers can be reluctant to run highly extreme experiments, which intuitively are among the most information-optimal experiments.

In summary, in the scenario where the performance of all four channels in the mix is uncertain, the investment allocation optimized by the BED framework with respect to the saturation and shape parameters recommends moving the planned budget from Channels 3 and 4 to Channels 1 and 2, and also increasing their investments to a level higher than the typical level. Such an allocation does not clearly reduce the uncertainty in the marketing parameters of the MMM fitted with the experimental investments compared to the MMM fitted without the experiment. This indicates that 4-week experimental observations do not provide strong informational signals to guide the posterior distributions towards the ground truth values.

4.2.3 Post-experiment optimized for retention rate

Figure 4.14 presents the planned budget of all channels in the mix and their corresponding experimental investments when optimizing the experiments with respect to the retention rate parameter. When optimizing the experiments with respect to the retention rate parameter, for Channels 1 and 2, the optimal design recommends investments of 12.61 for Channel 1 and 11.98 for Channel 2 in the first week of the experiment, while setting the investments of the following three weeks to zero. For Channels 3 and 4, the optimal experiment recommends forcing the investments to zero during the whole experimental period. Intuitively, zero-investment observations help gain insights about the retention rate parameter, as the marketing effects during such zero-investment periods are solely carried over from past investments based on the carryover effects. On the other hand, zero-investment observations might not help gain learning about the saturation and shape parameters, as the diminishing returns of zero investments are always zero regardless of what values the saturation and shape parameters are. Meanwhile, the atypically high investments during the first experimental week for Channels 1 and 2 provide a foundation for the following three weeks, with zero investments for the carryover effects. Such high investments could also provide insights about the saturation curve as they are likely around the saturated level of Channels 1 and 2.

These values from the optimal experiment replace the originally generated investments of the period from Week 105 to Week 108 in the future budget of all channels in the mix. Two new MMMs are then refitted with both historical and simulated future budget planning in the synthetic dataset. One model is fitted with the experimental investments mentioned above, while the other model is fitted with the originally generated dataset. Both of these models use the same prior specification as the pre-experiment MMM presented in Table 4.5.

Table 4.8 presents the summary statistics of the resulting post-experiment posterior distributions of all marketing parameters from the above two models. The differences of the posterior means and standard deviations between the models fitted with and without the experimental investments are negligible. There is a small shift of the posterior mean towards the ground truth value for Channel 2's shape parameter in the model fitted with the experiment. Meanwhile, the posterior mean of Channel 2's saturation in the model fitted with the experiment is further away from the ground truth than the one from the model fitted without the experiment. For other parameters, there are no considerable reductions in the posterior standard deviations or clear shifts in the posterior means towards the ground truth values of any marketing variables.

Figure 4.15 illustrates the posterior distributions of marketing parameters from those two models, along with their ground truth values, which indicate the same observations as in Table 4.8. There are small shifts in the posterior distributions of the model fitted with the experiment, with the most visible change in Channel 2's shape parameter. On the other hand, the effects of the experiments on the posterior distributions of the retention rate are minimal and even counterintuitive, given that the experiment is optimized with respect to the retention rate parameter. For example, the model fitted with the experiment places lower probability densities around the ground truth values of Channel 3's and Channel 4's retention rates than the model without the experiment does.

Figure 4.16 visualizes the posterior saturation curves of all channels from the models fitted with and without the experiment, along with their ground truth curves computed with the hyperparameters used in the synthetic data generation, showing small differences between the curves obtained from the two models. However, there is no clear reduction in the uncertainty of posterior saturation curves in any channels from the model fitted with the experiment. This is intuitively expected as the experiment is not optimized towards the saturation curve's parameters, but with respect to the retention rate parameter.

Based on the observations discussed about when comparing the posterior distributions between two models fitted with and without the experiment, similar to the observations in Subsection 4.2.2, the experiment optimized with respect to the retention rate in this case also provides limited information and learning, and the marketing parameters in the model remain weakly identified given the weak prior specification representing high uncertainty of the model.

In summary, in the scenario where the performance of all channels in the mix is uncertain, the investment allocation optimized for the retention rate parameter recommends placing anomalously high investments in the first week of the experimental period for Channels 1 and 2 and halting the investments to zero during other periods of Channels 1 and 2, and also for the other two channels during the whole experimental period. Such an allocation does not seem to help enhance the current state of knowledge about marketing parameters in the model when comparing the posterior distributions between the model fitted with the experiment and the one fitted without the experiment. This is most probably because the 4-week experiment does not provide strong enough observations to guide the model towards the ground truth values of the parameters.

4.3 Summary

This chapter simulates two different scenarios about the uncertainty of marketing performances and discusses how the BED framework can be applied to help enhance the current state of knowledge, with optimal experiments running for 4 weeks under a restricted budget. The first scenario simulates the situation where the performance of Channel 1 is uncertain while the performances of the other three channels in the mix are certain. The optimal designs yield modest reductions in the posterior uncertainty for the targeted subsets of parameters. The second scenario simulates the situation where the performance of all channels in the mix is uncertain. In this case, the optimal design provides limited information and changes in the posterior uncertainty of the marketing parameters in the model, most likely due to the weak prior specification used for the model in this case.

Two scenarios discussed in this chapter highlight several important insights regarding optimized investment allocation for marketing experiments. First, investment patterns recommended by the BED framework tend to be relatively extreme compared to typical budget allocations. This reflects the fact that information-oriented experiments prioritize exploration over short-term returns, favoring previously untested investment allocations to maximize information gain. Second, the effectiveness of such optimized experiments is highly dependent on the existing state of knowledge. When prior beliefs about marketing performance are relatively precise and confident, controlled investment variation can be an effective and operationally simple approach to generating meaningful additional learning. In contrast, under conditions of high uncertainty, simple budget reallocation experiments provide limited additional information, as the signal generated within a short experimental window is insufficient to meaningfully update the current state of knowledge. In such cases, more complex experimental designs, such as geo-lift tests, are more suitable to substantially reduce uncertainty. However, once a sufficiently confident baseline understanding is established, controlled investment allocation experiments remain a valuable and efficient tool for refining estimates of marketing effectiveness.

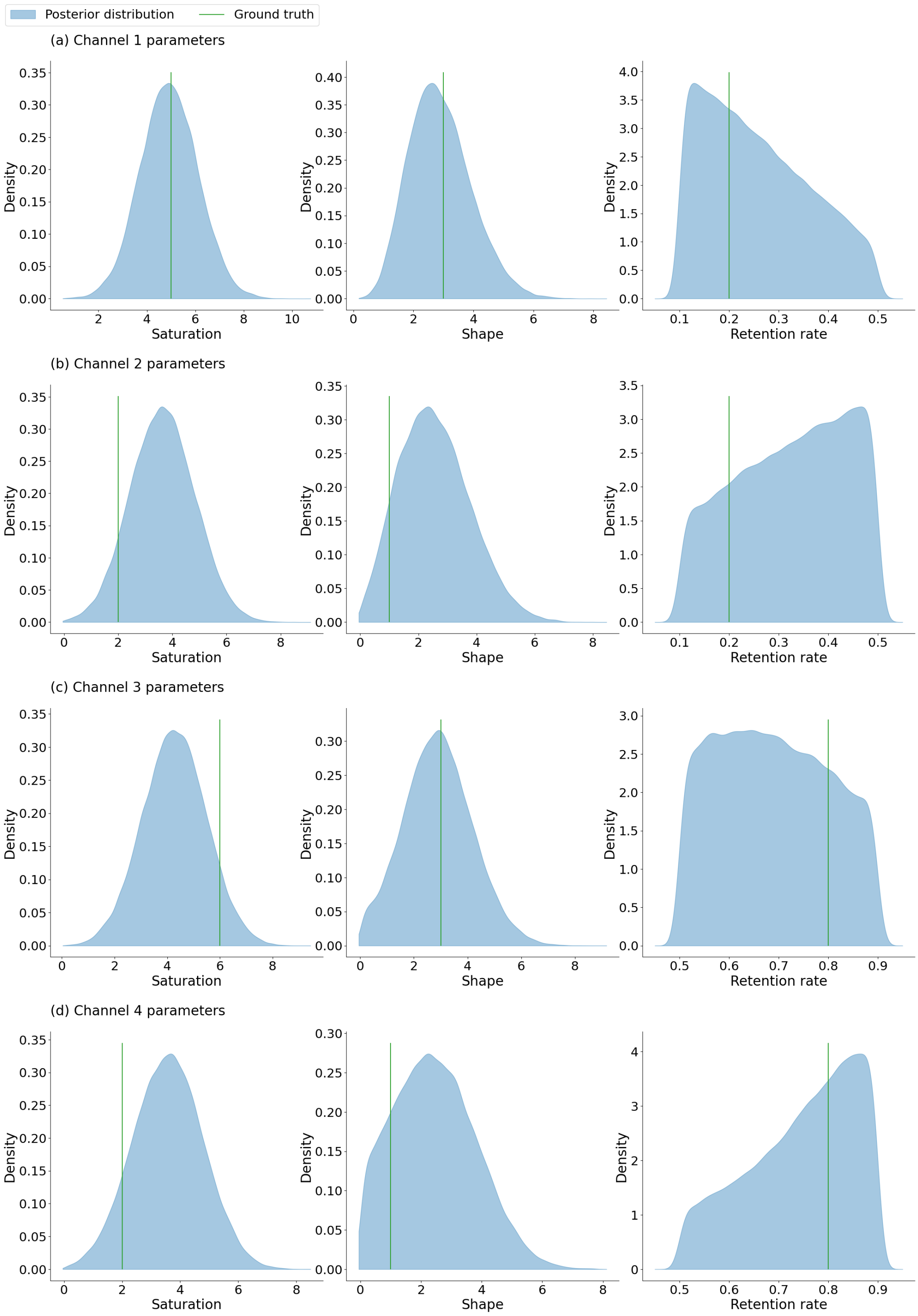

Figure 4.9: Posterior distributions of the pre-experiment MMM in the scenario “all uncertain channels” and the ground truth values of three marketing parameters for all four channels in the mix. The blue areas represent the posterior distributions of, from left to right, saturation, shape, and retention rate, respectively. The green vertical lines represent the ground truth values, which are the hyperparameters used in the synthetic data generation.

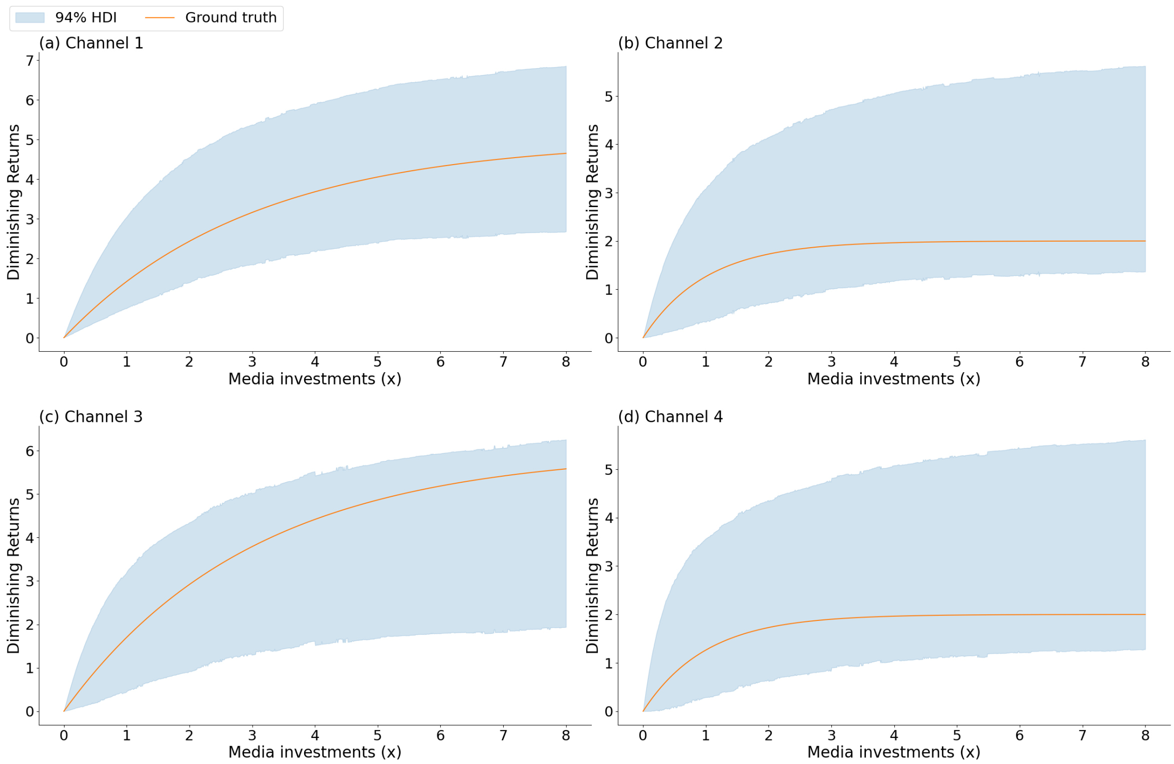

Figure 4.10: Posterior saturation curves of all four channels fitted by the pre-experiment model in the scenario "all uncertain channels" along with their ground truth curves. The blue areas represent the posterior 94% HDI of the curves. The orange line represents the ground truth curves, which are calculated using the hyperparameters used in the synthetic data generation.

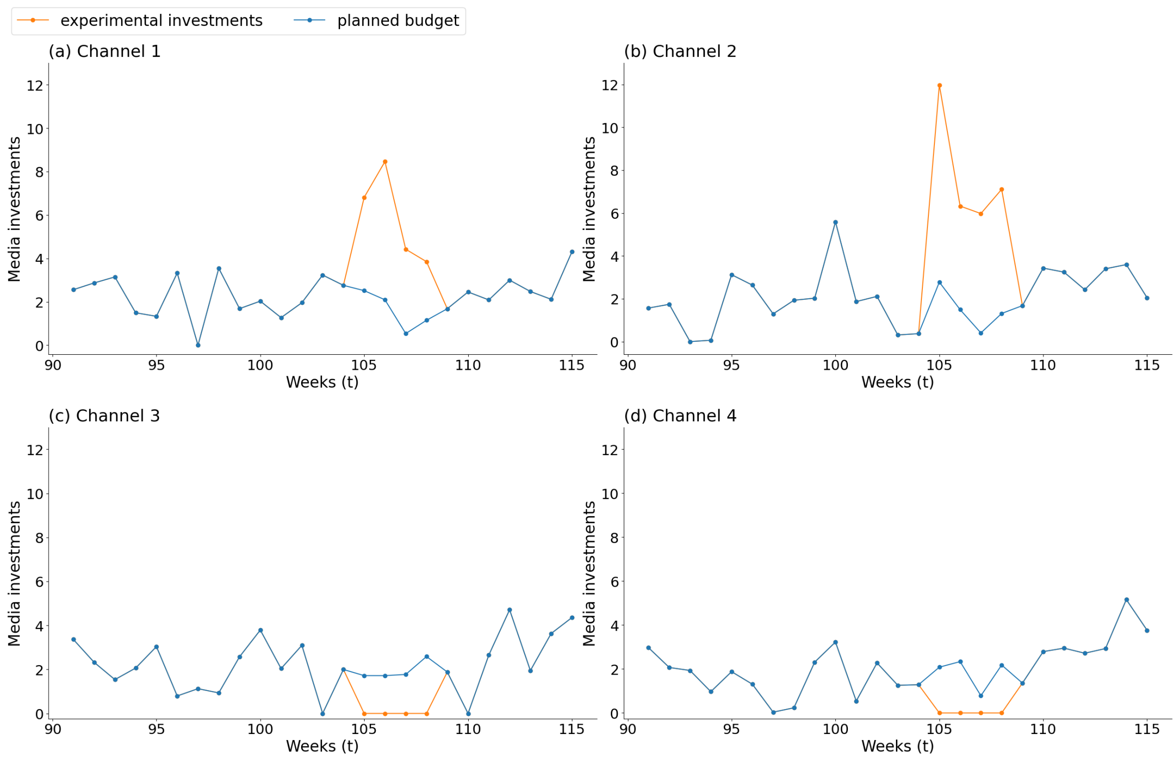

Figure 4.11: The planned budgets and experimental investments optimized with respect to the saturation and shape parameters of all four channels in the scenario “all uncertain channel”. The blue lines represent the planned budget, while the orange lines represent the experimental investments during the experimental period: Weeks 105 - 108 of the corresponding channels.

Table 4.7: The summary statistics of the posterior distributions of all marketing parameters from two post-experiment MMMs fitted with and without the experimental investments optimized for the saturation and shape parameters in the scenario "all uncertain channels". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation.

	Parameter	Model	Mean	Standard deviation	94% HDI
Channel 1	α_1 (5.0)	with exp.	4.754	1.134	2.573 – 6.842
		without exp.	4.755	1.160	2.573 – 6.929
	β_1 (3.0)	with exp.	2.825	1.031	1.002 – 4.775
		without exp.	2.849	1.050	0.933 – 4.768
	r_1 (0.2)	with exp.	0.267	0.109	0.100 – 0.455
		without exp.	0.249	0.107	0.100 – 0.443
Channel 2	α_2 (2.0)	with exp.	3.375	1.134	1.177 – 5.442
		without exp.	2.329	1.249	0.075 – 4.488
	β_2 (1.0)	with exp.	2.268	1.191	0.184 – 4.442
		without exp.	2.329	1.249	0.075 – 4.488
	r_2 (0.2)	with exp.	0.341	0.109	0.150 – 0.500
		without exp.	0.331	0.111	0.141 – 0.500
Channel 3	α_3 (6.0)	with exp.	4.342	1.179	2.124 – 6.565
		without exp.	4.373	1.218	2.062 – 6.650
	β_3 (3.0)	with exp.	2.939	1.227	0.680 – 5.279
		without exp.	2.976	1.243	0.589 – 5.303
	r_3 (0.8)	with exp.	0.675	0.106	0.500 – 0.852
		without exp.	0.684	0.110	0.500 – 0.864
Channel 4	α_4 (2.0)	with exp.	3.501	1.205	1.237 – 5.768
		without exp.	3.586	1.223	1.319 – 5.888
	β_4 (1.0)	with exp.	2.303	1.338	0.002 – 4.524
		without exp.	2.326	1.334	0.002 – 4.530
	r_4 (0.8)	with exp.	0.748	0.108	0.551 – 0.900
		without exp.	0.757	0.105	0.560 – 0.900

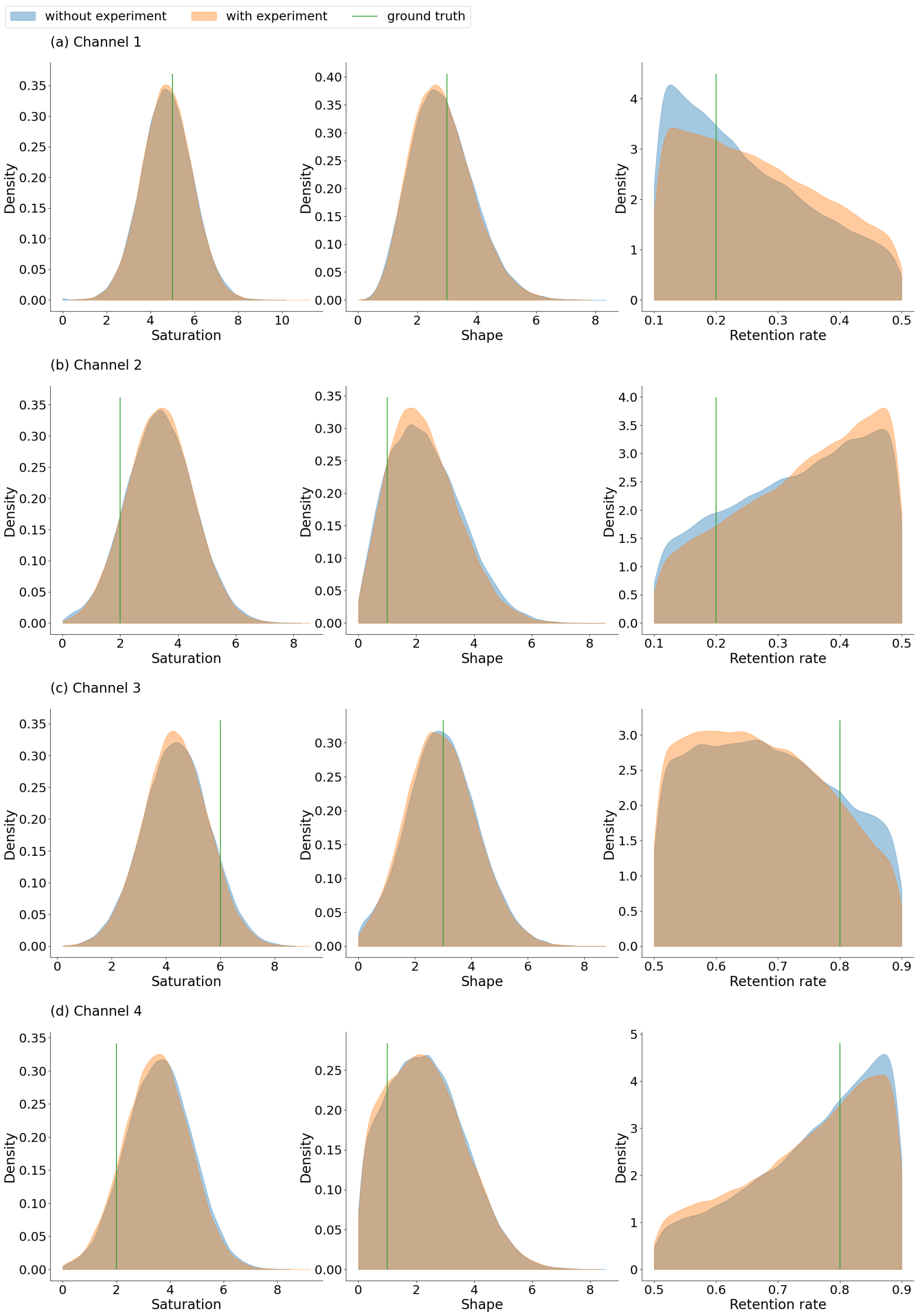

Figure 4.12: Posterior distributions of the models fitted with and without the experiment optimized for the saturation and shape parameters, along with the ground truth values of, from left to right, the saturation, shape, and retention rate parameters of all four channels in the mix in the scenario "all uncertain channels". The blue areas represent the posterior distributions without the experiment, while the orange areas represent the ones with the experiment. The green vertical lines represent the ground truth values of the corresponding parameters, which are the hyperparameters used in the synthetic data generation.

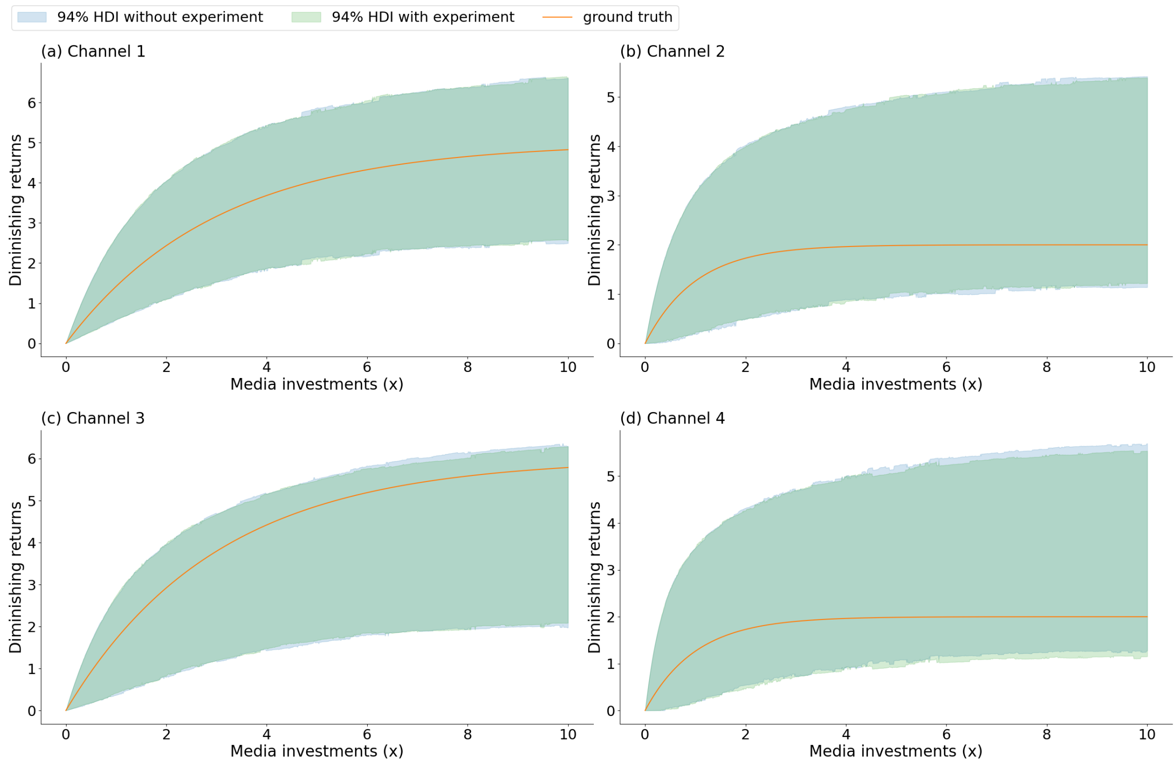

Figure 4.13: Posterior saturation curves of all four channels with and without the experiment optimized for the saturation and shape parameters, along with the ground truth curves in the scenario "all uncertain channels". The blue areas represent the posterior 94% HDIs of the saturation curves without the experiment, while the green areas represent the posterior 94% of the curves with the experiment. The orange lines represent the ground truth curves, which are calculated with the hyperparameters used in the synthetic data generation.

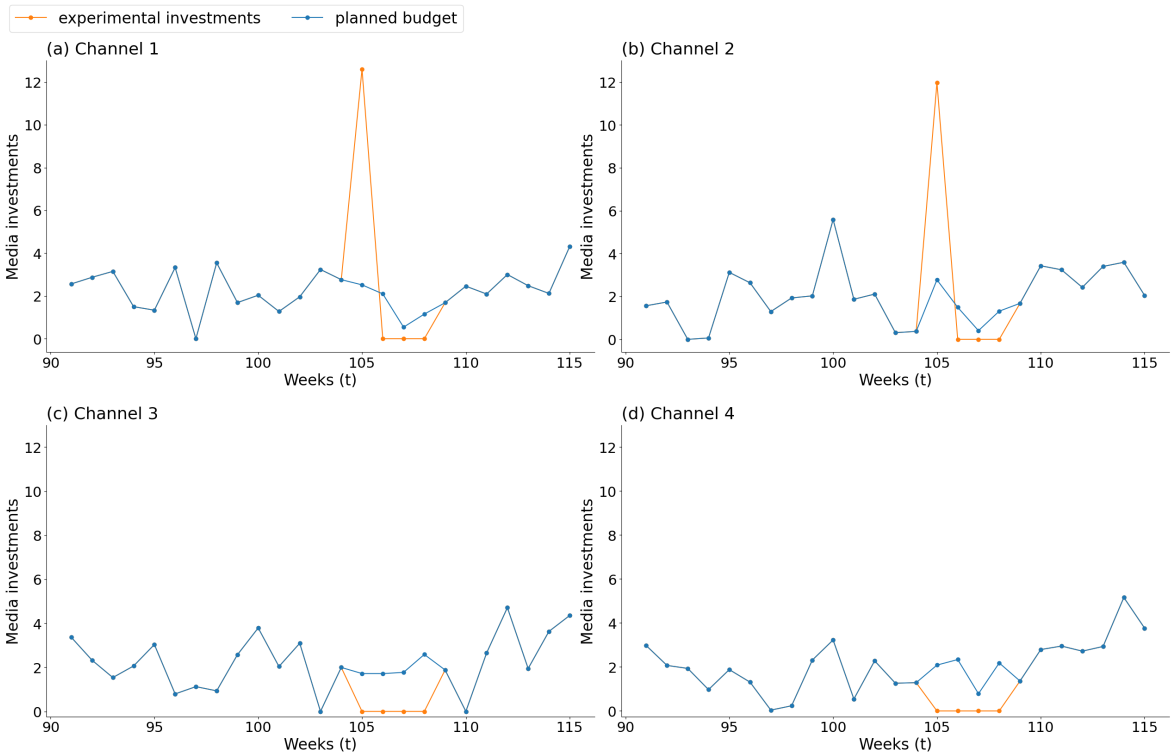

Figure 4.14: The planned budgets and experimental investments optimized with respect to the retention rate parameter of all four channels in the scenario "all uncertain channel". The blue lines represent the planned budget, while the orange lines represent the experimental investments during the experimental period: Weeks 105 - 108 of the corresponding channels.

Table 4.8: The summary statistics of the post-experiment posterior distributions of all marketing parameters from two MMMs fitted with and without the experimental investments optimized for the retention rate parameter in the scenario "all uncertain channels". The value shown in parentheses next to each parameter corresponds to its ground truth used in the synthetic data generation.

	Parameter	Model	Mean	Standard deviation	94% HDI
Channel 1	α_1 (5.0)	with exp.	4.755	1.123	2.637 – 6.853
		without exp.	4.755	1.160	2.573 – 6.929
	β_1 (3.0)	with exp.	2.664	1.022	0.835 – 4.556
		without exp.	2.849	1.050	0.933 – 4.768
	r_1 (0.2)	with exp.	0.248	0.104	0.100 – 0.435
		without exp.	0.249	0.107	0.100 – 0.443
Channel 2	α_2 (2.0)	with exp.	3.402	1.136	1.289 – 5.574
		without exp.	2.329	1.249	0.075 – 4.488
	β_2 (1.0)	with exp.	2.125	1.226	0.047 – 4.292
		without exp.	2.329	1.249	0.075 – 4.488
	r_2 (0.2)	with exp.	0.327	0.110	0.143 – 0.500
		without exp.	0.331	0.111	0.141 – 0.500
Channel 3	α_3 (6.0)	with exp.	4.243	1.185	2.000 – 6.450
		without exp.	4.373	1.218	2.062 – 6.650
	β_3 (3.0)	with exp.	2.884	1.229	0.558 – 5.143
		without exp.	2.976	1.243	0.589 – 5.303
	r_3 (0.8)	with exp.	0.677	0.107	0.500 – 0.857
		without exp.	0.684	0.110	0.500 – 0.864
Channel 4	α_4 (2.0)	with exp.	3.433	1.204	1.152 – 5.660
		without exp.	3.586	1.223	1.319 – 5.888
	β_4 (1.0)	with exp.	2.246	1.337	0.001 – 4.481
		without exp.	2.326	1.334	0.002 – 4.530
	r_4 (0.8)	with exp.	0.749	0.109	0.549 – 0.900
		without exp.	0.757	0.105	0.560 – 0.900

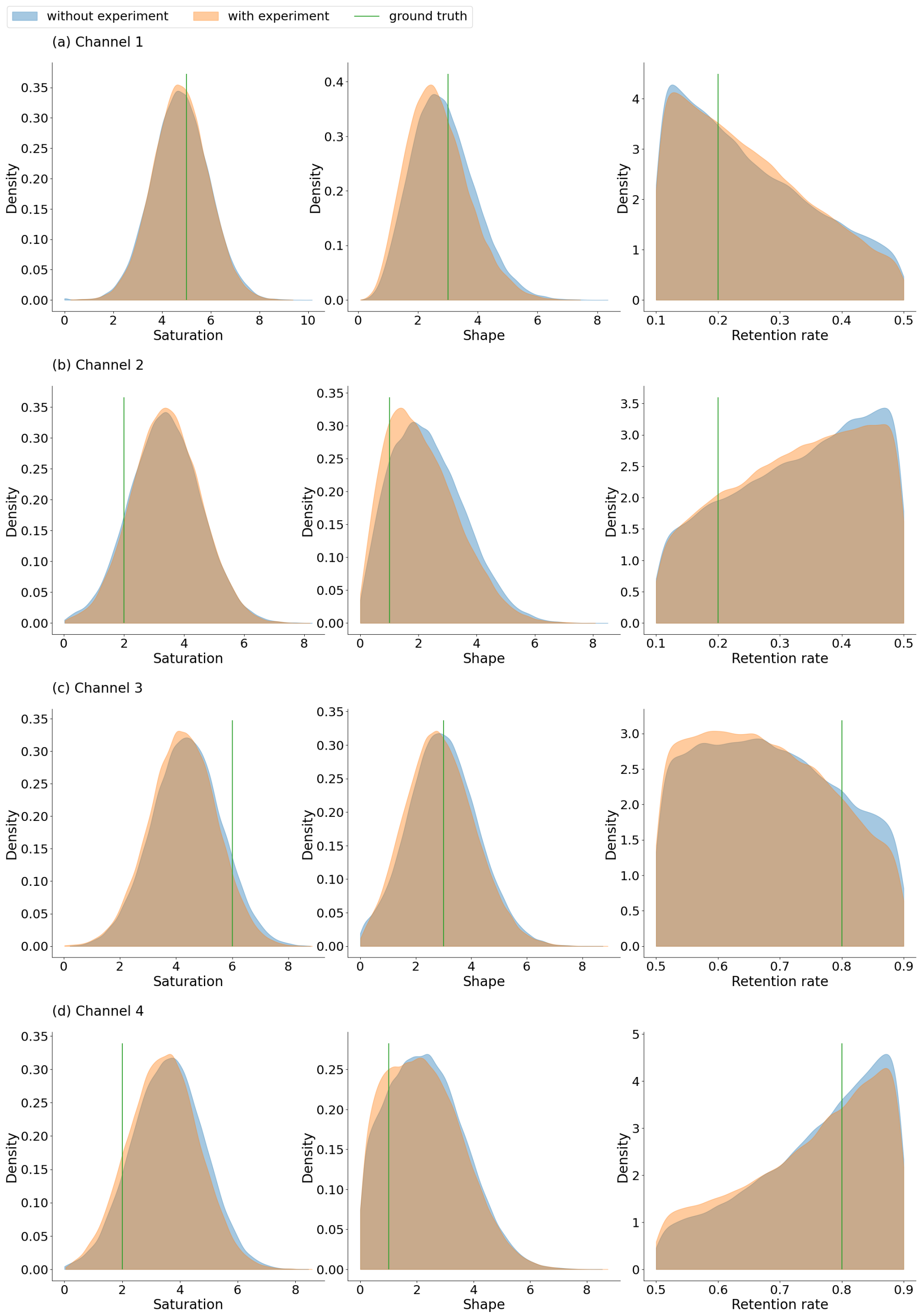

Figure 4.15: Posterior distributions of the models fitted with and without the experiment optimized for the retention rate parameter, along with the ground truth values of, from left to right, the saturation, shape, and retention rate parameters of all four channels in the mix in the scenario "all uncertain channels". The blue areas represent the posterior distributions without the experiment, while the orange areas represent the ones with the experiment. The green vertical lines represent the ground truth values of the corresponding parameters, which are the hyperparameters used in the synthetic data generation.

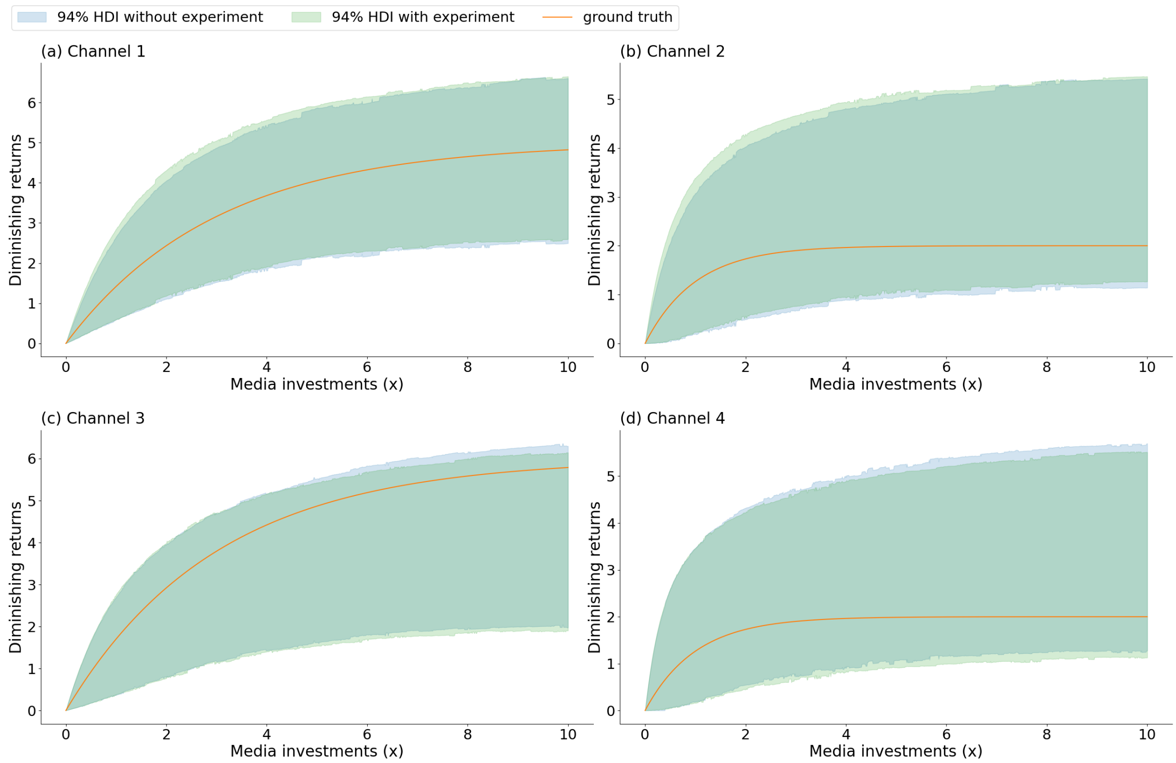

Figure 4.16: Posterior saturation curves of all four channels with and without the experiment optimized for the retention rate parameter, along with the ground truth curves in the scenario “all uncertain channels”. The blue areas represent the posterior 94% HDIs of the saturation curves without the experiment, while the green areas represent the posterior 94% of the curves with the experiment. The orange lines represent the ground truth curves, which are calculated with the hyperparameters used in the synthetic data generation.

5. Discussion

This chapter discusses the limitations of the thesis in Section 5.1 and potential extensions and improvements to the work presented in the two following sections to fully utilize the Bayesian experimental design (BED) framework in the context of marketing mix modeling (MMM). Specifically, Section 5.2 explores the use of regularized utility functions to mitigate the impractical extremity of experiments derived from the expected information gain (EIG). Section 5.3 discusses the opportunities and challenges associated with applying BED to more complex MMM formulations, including hierarchical and interaction-based models.

5.1 Thesis Limitations

The thesis has several limitations related to its methodology and applicability in real-world use cases. First, the posterior distributions of the retention rate parameters are weakly identified, particularly in the "all-uncertain-channels" scenario discussed in Section 4.2. Since the retention rate controls the carryover effects, different values of retention rate can lead to substantially different adstock-transformed signal patterns over the experimental period. As a consequence, weak identification of the retention rate undermines the optimization process in the BED framework: identical budget allocations may correspond to different effective exposure patterns, thereby limiting the amount of additional learning and information the experimental budgets provide.

Second, the model used in the thesis is simplified and does not reflect the complexity level of the most common models used in the industry. For example, models used in real-world use cases typically have more channels in the mix (instead of four as in the thesis) and more complicated assumptions about how marketing works, in addition to the carryover effects and the diminishing returns. Section 5.3 discusses some of these additional assumptions in more detail. On the other hand, due to the limited availability of public datasets for MMMs, the results in the thesis are solely based on the synthetic data that can introduce optimistic biases, as the data generation process matches perfectly the generative model defined in the MMM. In real-world scenarios, the assumptions encapsulated by the generative model are not guaranteed to be correct. Therefore, even though the thesis has illustrated how the BED framework can be applied to improve the MMM results, it does not provide fully concrete insights into the potential challenges the BED framework might have in real-world use cases.

Finally, to marketing practitioners, the variables optimized in marketing experiments are not always the investments of the channels in the mix but can be platform-specific parameters of individual marketing campaigns. For example, in Google Ads, a campaign can be configured to have a target Return on Ad Spend (ROAS) of 4.5 and the daily or weekly investments are optimized by the Google platform itself. Therefore, the target ROAS parameters are more suitable experimental variables in this case than the investment levels are. However, to optimize based on such platform-specific parameters, the BED framework needs to take into account how the optimization algorithms of marketing platforms work and how different parameters affect the performance of the interested campaigns and channels, which are not publicly available in most cases.

5.2 Regularized Utility Function

When EIG is used as the utility function for designing marketing experiments without any budget constraints, the resulting optimal designs are often extreme relative to typical marketing budget investments, such as investing two to three times as much as the typical investment level or halting the investments to zero. While such designs are optimal from a purely informational perspective, they might be considered impractical in real-world settings, as marketing practitioners will be unwilling to implement such extreme investments due to the associated operational risks and potential revenue loss.

To address this issue, a regularization term can be added to the utility function to penalize deviations from a baseline of investment level. One simple approach is to incorporate a penalty based on the absolute difference between the experimental investments and the planned budgets. Mathematically, this regularized utility function can be formulated along with the expected information gain (EIG) as:

$$\mathrm{EIG}(x) + \lambda \log \left(|\mathbf{x} - \mathbf{x}_0| \right), \tag{5.1}$$

where $\mathbf{x} \in \mathbb{R}^{K \times M}$ is the media spends in an experiment design, $\mathbf{x}_0 \in \mathbb{R}^{K \times M}$ is the baseline media spends, representing the planned marketing investments in the absence of experimentation, and λ is the regularization weight that controls the degree of conservativeness: larger values correspond to a stronger preference for designs that remain close to the baseline level and vice versa.

However, a key limitation of this approach is that it does not account for the heterogeneity in channel performances. Penalizing deviations solely based on absolute differences treats all channels similarly, regardless of their growth potential or expected marginal returns. In practice, it may be reasonable to substantially increase investments in a high-potential channel even if such an increase deviates significantly from the historical investment levels, but the absolute difference regularization would penalize such behavior, potentially leading to overly conservative designs.

An alternative regularization strategy is to incorporate the marketing performance directly into the utility function using the expected outcomes generated by the experimental investments, net of the additional investment cost. Mathematically, the utility function with such a regularization strategy can be formulated as:

$$\mathrm{EIG}(x) + \lambda \left(\mathbb{E} \left[\sum_{m=1}^{M} \alpha_m \, \mathrm{INE}(\mathrm{adstock}(x^*_{t,m}; r_m, L); \beta_m) \right] - \left(\sum_{m=1}^{M} x_m - \sum_{m=1}^{M} x_{m,0} \right) \right). \tag{5.2}$$

In this case, the regularization term can be interpreted as the expected profits from the experiment, where revenue is modeled via the MMM response function and costs are proportional to deviations from the baseline spend. As a result, channels with high growth potential but limited historical investments may receive disproportionately large experimental budgets, aligning more closely with practical marketing intuitions.

A drawback of this approach lies in the mismatch of scales between the two components of the utility function as presented in Equation (5.2). Expected information gain is measured on a logarithmic scale, while the revenue-based regularization term is expressed on a linear scale. Therefore, the choice of the regularization weight is neither straightforward nor universal and must be tuned based on the specific context and use cases.

5.3 Complex Marketing Mix Model

The marketing mix model used in this thesis is intentionally simple, serving primarily to illustrate how the BED framework can be applied in the context of marketing mix modeling. In both academic research and industrial applications, however, MMMs are often more complex and incorporate more assumptions about how marketing works in practice. This section briefly discusses two common extensions to the MMM, geo-hierarchical models and models with cross-channel interactions, and how the BED framework can be adapted to these models.

A common extension of standard MMMs is the use of hierarchical geo-location models (Sun et al., 2017). Many companies operate across multiple regions or countries, where marketing strategies share a common structure but differ in magnitude or effectiveness due to local factors. In a Bayesian geo-hierarchical MMM, channel-level performance parameters are drawn from higher-level prior distributions that represent overall marketing effectiveness, while region-specific parameters capture local deviations from the shared global parameters.

With geo-hierarchical models, the BED experiment space can be extended to include geographic dimensions. For example, experiments may be restricted to specific regions where the operational risk of extreme interventions is relatively lower compared to other regions. This allows marketing practitioners to preserve stable investment levels in core markets while conducting more aggressive experiments in other test regions. Importantly, the information is shared across regions through the hierarchical structure of the MMM. Experiments conducted in a single geo-location can provide insights not only for that region but also for all other locations in the model.

Despite these advantages, integrating BED with geo-hierarchical MMMs introduces several challenges. Firstly, the experimental design space becomes high-dimensional, integrating both channel and geographic dimensions. Approximating the utility function in such spaces can be computationally demanding. Secondly, in real-world scenarios, the marketing activities often exhibit spill-over effects, where investments in one region can influence the outcomes in other neighboring or connected regions. Such interdependencies may lead to biased inferences or misleading experimental recommendations.

Another extension to the standard MMMs is the cross-channel interactions of marketing channels. In practice, marketing channels are not independent of each other. For example, advertising on social media platforms like Facebook or Snapchat may increase brand awareness, thereby improving the effectiveness of search advertising in Google or Bing. Such cross-channel effects can be incorporated into MMMs through interaction terms and modeled using either linear or non-linear functions.

Applying the BED framework in the presence of channel interactions can pose some challenges. Interactions increase model complexity and introduce nonlinear dependencies between experimental design variables and model parameters. As a result, the posterior distributions and the expected information gain become more difficult to approximate. Moreover, examining the contributions of individual channels becomes less straightforward as changes in one channel could affect the inference of parameters associated with other channels. These challenges complicate both the optimization of the experimental design and the interpretation of the resulting recommendations.

6. Conclusions

This thesis investigates how Bayesian experimental design (BED) can be integrated with Marketing Mix Modeling (MMM) to design investment allocations for marketing experiments under budget constraints that optimally improve the current state of knowledge about marketing performance. Using controlled simulation studies, the thesis demonstrates how the posterior distributions from a Bayesian MMM can be utilized as the prior knowledge in the BED framework to design experiments that maximize the expected information gain (EIG) about key marketing parameters representing two core assumptions in MMMs, carryover effects and diminishing returns, and how the experimental investments can, in turn, contribute to the improvement of the MMM with its atypical observations.

Two simulated scenarios are considered in the thesis, reflecting two different states regarding the prior knowledge of marketing performance. The first scenario assumes that the performance of one marketing channel in the mix is uncertain, while the performance of others is well identified. In this setting, the experiment optimized with respect to the diminishing returns recommends high investments for the initial week, followed by zero-investment weeks, while the experiment optimized for the carryover effects recommends halting the investments to zero for the whole experimental period. These designs are shown to shift posterior distributions closer to the ground truth values and reduce uncertainty relative to models fitted without experimental investments. This scenario illustrates that, when prior knowledge is reasonably informative and uncertainty is localized, investment allocations optimized by the BED framework can meaningfully improve learning about specific channel characteristics.

The second scenario simulates that the performance of all channels in the mix is uncertain. In this case, the BED framework frequently recommends abrupt reallocation of the budget, concentrating most of the investment in a subset of channels while forcing the investments of others to zero. The optimized experiments produce limited improvements in posterior uncertainty and even lead the posterior distributions to shift away from the ground truth. Such observations highlight an important limitation that, when parameter identifiability is weak and the priors are diffuse across all channels in the mix, 4-week experiments do not provide sufficiently strong signals to meaningfully enhance the model performance.

Across both scenarios, a consistent pattern emerges: experiments that are information-optimal are extreme relative to typical marketing budget allocations. Specifically, high investments are informative for learning saturation behavior, while zero investments help isolate the carryover effects. While these recommendations are theoretically and intuitively justified with respect to the expected information gain criterion, they may be considered impractical in real-world scenarios. Therefore, depending on specific use cases, different budget constraints used to control the experiment space can be enforced to follow operational conditions, risk aversion, or potential revenue loss when running the experiments.

On the other hand, the thesis also has several limitations. The results in the study are solely based on the synthetic data generated with the same generative model used in the fitted MMM, ensuring the consistency between the data and assumptions made in the model. Meanwhile, the MMM used in the thesis is simplified and does not reflect the complexity level of the models used in the industry. Moreover, the BED framework in the thesis focuses on optimizing the experimental investments of the channels in the mix, while marketing practitioners might be more interested in campaign parameters configured in specific platforms.

In conclusion, the thesis demonstrates that Bayesian experimental design provides a theoretically robust framework to design investment allocations for marketing experiments informed by MMM uncertainty. While it faces practical challenges when being applied in real-world scenarios, the framework offers insights into the trade-offs between information learning and business risks in marketing measurement. When applying appropriate constraints and domain expertise, BED has the potential to contribute to the iterative refinement of marketing mix models specifically and marketing measurement in general.

Bibliography

Anderl, E., Becker, I., von Wangenheim, F., and Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, 33(3):457–474.

Athey, S. and Imbens, G. (2017). Chapter 3 - The Econometrics of Randomized Experiments. In *Handbook of Field Experiments*, volume 1, pages 73–140. Elsevier.

Bartels, R. et al. (1976). The History of Marketing Thought. Grid Columbus, OH.

Bernardo, J. M., Smith, A. F., and Berliner, M. (1994). *Bayesian theory*, volume 586. Wiley Online Library.

Broadbent, S. (1979). One way TV advertisements work. *Journal of the Market Research Society*, 21(3):139–166.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring Causal Impact Using Bayesian Structual Time-series Models. *The Annals of Applied Statistics*, 9(1):247–274.

Cain, P. (2025). Long-term advertising effects: The Adstock illusion. *Applied Marketing Analytics*, 11(1):23–42.

Clifton, B. (2012). *Advanced web metrics with Google Analytics*. John Wiley & Sons.

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., and Holz, T. (2019). We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *Proceedings of the Network and Distributed System Security Symposium 2019*. Internet Society.

Eckles, D., Gordon, B. R., and Johnson, G. A. (2018). Field studies of psychologically targeted ads face threats to internal validity. *Proceedings of the National Academy of Sciences*, 115(23):E5254–E5255.

Foster, A., Jankowiak, M., Bingham, E., Horsfall, P., Teh, Y. W., Rainforth, T., and Goodman, N. (2019). Variational Bayesian Optimal Experimental Design. In *Proceedings of Advances in Neural Information Processing Systems*, volume 32.

Franses, P. H. (2025). Adstock revisited. *Applied Economics*, 57(8):882–886.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press.

Goldfarb, A. and Tucker, C. (2011a). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science*, 30(3):389–404.

Goldfarb, A. and Tucker, C. E. (2011b). Privacy Regulation and Online Advertising. *Management Science*, 57(1):57–71.

Gong, C., Yao, D., Zhang, L., Chen, S., Li, W., Su, Y., and Bi, J. (2024). CausalMMM: Learning Causal Structure for Marketing Mix Modeling. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 238–246.

Hanssens, D. M., Parsons, L. J., and Schultz, R. L. (2001). *Market Response Models*. Springer New York.

Hoffman, M. D., Gelman, A., et al. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Jin, Y., Wang, Y., Sun, Y., Chan, D., and Koehler, J. (2017). Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects. *Google Research*, pages 1–34.

Jones, D. B. and Shaw, E. H. (2002). A History of Marketing Thought. *Handbook of Marketing*, pages 39–65.

Knight, P. (2016). *Shoe Dog - A Memoir by the Creator of Nike*. Simon & Schuster.

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Li, H. A. and Kannan, P. (2014). Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment. *Journal of Marketing Research*, 51(1):40–56.

Miller, K. M. and Skiera, B. (2024). Economic consequences of online tracking restrictions: Evidence from cookies. *International Journal of Research in Marketing*, 41(2):241–264.

Mulc, T., Anderson, M., Cubre, P., Zhang, H., Liu, I., and Kumar, S. (2025). NNN: Next-Generation Neural Networks for Marketing Measurement.

Myung, J. I., Cavagnaro, D. R., and Pitt, M. A. (2013). A tutorial on adaptive design optimization. *Journal of Mathematical Psychology*, 57(3):53–67.

Naik, P. A. and Raman, K. (2003). Understanding the Impact of Synergy in Multimedia Communications. *Journal of Marketing Research*, 40(4):375–388.

Patil, A., Huard, D., and Fonnesbeck, C. J. (2010). PyMC: Bayesian Stochastic Modelling in Python. *Journal of Statistical Software*, 35(4):1–81.

Powell, M. J. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis*, pages 51–67. Springer.

Rust, R. T., Ambler, T., Carpenter, G. S., Kumar, V., and Srivastava, R. K. (2004). Measuring Marketing Productivity: Current Knowledge and Future Directions. *Journal of Marketing*, 68(4):76–89.

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2016). A Review of Modern Computational Algorithms for Bayesian Optimal Design. *International Statistical Review*, 84(1):128–154.

Simon, J. L. and Arndt, J. (1980). The Shape of the Advertising Response Function. *Journal of Advertising Research*, 20(4):11–28.

Sun, Y., Wang, Y., Jin, Y., Chan, D., and Koehler, J. (2017). Geo-level Bayesian Hierarchical Media Mix Modeling. *Google Research*.

Vakratsas, D. and Ambler, T. (1999). How Advertising Works: What Do We Really Know? *Journal of Marketing*, 63(1):26–43.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Zhang, Y., Wurm, M., Li, E., Wakim, A., Kelly, J., Price, B., and Liu, Y. (2024). Media Mix Model Calibration With Bayesian Priors. *Google Research*.