Latent Theme-based Decomposition of the Causal Impact of Marketing Interventions

Khaled Boughanmi, Raghuram Iyengar, and Young-Hoon Park*

Abstract

Quantifying and interpreting the causal impact of marketing interventions across large product assortments is a critical concern in both marketing theory and practice. The level of analysis (e.g., product or category) plays a central role in the usefulness of the results for downstream decisions. While product-level analyses are often impractical due to data sparsity, category-level approaches rely on rigid, firm-defined taxonomies that may not align with how customers perceive and use products. We propose a modeling framework that integrates causal inference with unsupervised probabilistic machine learning to determine the level of analysis based on behaviorally grounded themes. Specifically, we combine regression analysis with a mixed-membership model based on Latent Dirichlet Allocation to jointly infer latent themes and estimate theme-level treatment effects. Applying our framework to data from a randomized field experiment involving a price coupon, we find that while the intervention is effective in lifting overall spending, only a few latent themes drive the observed effect. Treatment effects also vary across customers and are stronger for themes they had previously engaged with, revealing a reinforcement dynamic often overlooked by category-based analyses. Finally, we show that theme-based targeting significantly outperforms conventional approaches in driving returns, highlighting the practical value of our framework for more effective and customer-centric marketing.

Keywords: Causal Machine Learning, Latent Purchase Themes, Bayesian Inference, Heterogeneous Treatment Effects, Promotions, Marketing Interventions, Customer Targeting

*Khaled Boughanmi is an Assistant Professor of Marketing at the Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY, kb746@cornell.edu. He is the corresponding author. Raghuram Iyengar is the Miers-Busch W'1885 Professor and Professor of Marketing, Department of Marketing, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, riyengar@wharton.upenn.edu.Young-Hoon Park is the Sung-Whan Suh Professor of Management and Professor of Marketing, Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, yp34@cornell.edu.

1 Introduction

Rigorous analysis of marketing interventions such as advertising campaigns and price promotions is fundamental for advancing both marketing theory and practice. Field experiments serve as a powerful method for identifying causal effects and refining strategies (e.g., Lodish et al. 1995; Anderson and Simester 2004; Feit and Berman 2019; Gordon et al. 2019; Shapiro et al. 2021). Beyond assessing whether an intervention is effective, analyzing how its impact differs across customers reveals deeper insight into the mechanisms driving its effectiveness (e.g., Hutchinson et al. 2000). However, such experiments can be costly and time-consuming, often requiring close collaboration and coordination among stakeholders (e.g., Anderson and Simester 2011). These challenges underscore the importance of maximizing the value of each experiment, motivating the use of advanced econometric methods for more efficient and informative analyses (e.g., Li et al. 2015; Athey and Imbens 2017).

One area of particular interest is quantifying and interpreting the causal impact of marketing interventions on demand across a firm's product assortment. In this context, managers often seek to answer two key questions: (1) Are treatment effects primarily limited to specific products, or do they spill over across the assortment? and (2) Which types of customers respond, and to what kinds of products do they respond to? (e.g., Goli and Chintagunta 2021; Seiler et al. 2021; Ye and Shankar 2024). These questions are difficult to address in settings with large and diverse product assortments, where sparse purchase data at the individual product level often hinders reliable estimation of causal effects (e.g., Lenk and Orme 2009; Jaikumar et al. 2024).

To address data sparsity, firms often group products into predefined categories and evaluate treatment effects at the category level. However, these groupings typically reflect the firm's internal taxonomy (e.g., skincare, makeup, or body care for a company offering personal care products) rather than how customers perceive and use products. As a result, category-level analyses can obscure meaningful behavioral variation and mischaracterize the true impact of an intervention on customer behavior. For example, a single skincare product may serve multiple customer needs such as daily hygiene and intensive hydration that are not captured by rigid, firm-defined classifications.

A parallel stream of research in probabilistic machine learning has explored data-driven approaches to grouping products by inferring latent structure in consumer behavior (e.g., Dew et al. 2024). Studies have shown that understanding customers' underlying purchase motivations can improve the predictions of future purchases (e.g., Jacobs et al. 2016). The notion is similar in spirit to the "jobs to be done" framework, which suggests that consumers purchase collections of products to fulfill specific goals (e.g., Christensen et al. 2016). This perspective underscores the importance of moving beyond surface-level product categories to uncover deeper behavioral patterns that shape demand.

Despite their benefits for predictive modeling, such data-driven methods for grouping products have rarely been applied in causal inference settings. Yet integrating them holds significant potential. First, grouping products based on latent themes—behaviorally meaningful patterns inferred from customer purchase data—can offer a more flexible and customer-centric foundation for analyzing causal effects. Second, theme-based groupings can help address data sparsity by aggregating products based on inferred themes rather than firm-defined categories. Third, decomposing treatment effects across themes can yield more granular and managerially relevant insights, particularly for personalization and targeting strategies.

In this paper, we propose a modeling framework that integrates unsupervised probabilistic machine learning with causal inference to estimate both average and heterogeneous treatment effects across latent themes. Specifically, we combine a regression analysis for causal estimation with a mixed-membership model based on Latent Dirichlet Allocation (LDA) to uncover latent themes and assess their responsiveness to marketing interventions. This unified framework enables the joint identification of latent themes and the decomposition of treatment effects, such as those of a price coupon, across those themes.

Our approach contributes to both marketing theory and practice in several ways. First, it enables a more granular understanding of treatment effects than traditional category-based analyses. Second, it supports more precise customer targeting by estimating individual-level heterogeneity not only in aggregate but also within each latent theme. Third, it bridges two distinct domains—causal inference and probabilistic machine learning—within a single, coherent framework. While causal inference ensures unbiased estimation of treatment effects, probabilistic modeling captures the complexity of customer-product relationships. To the best of our knowledge, our framework is among the first to integrate these approaches for analyzing treatment effects at the latent theme level.

The proposed framework addresses several technical and conceptual challenges while combining the two areas of research. Because products can belong to multiple themes, the model must decompose treatment effects without double-counting. We do so by assigning product-level outcomes to themes probabilistically, using posterior weights that preserve additivity. To satisfy causal identification assumptions, we estimate latent themes using only pre-treatment data, thereby avoiding contamination from post-treatment behavior (e.g., Imbens and Rubin 2015; Athey and Imbens 2017). Our Bayesian estimation procedure accounts for uncertainty in product-theme assignments, customer-theme preferences, and model parameters, enhancing the robustness of downstream inference. Finally, the framework accommodates heterogeneous treatment effects at both the aggregate and theme levels, enabling firms to tailor interventions based on individual pre-treatment behaviors.

This paper contributes to a growing literature that applies machine learning to causal inference in marketing (e.g., Ascarza 2018; Hagen et al. 2020; Goli et al. 2021; Iyengar et al. 2022; Ellickson et al. 2023; Li and Sonnier 2023; Unal and Park 2023; Turjeman and Feinberg 2024). While these studies offer valuable insights, they primarily estimate treatment effects at the aggregate level or conditional on observable customer characteristics, and often rely on firm-defined product categories. In contrast, we infer latent themes from observed purchase behavior and estimate theme-specific treatment effects, uncovering heterogeneity that is unobserved yet managerially actionable.

We also extend research on probabilistic machine learning in marketing—see Dew et al. (2024) for an in-depth overview. LDA has been applied to a variety of marketing contexts, including online reviews (e.g., Tirunillai and Tellis 2014), search queries (e.g., Liu and Toubia 2018), social media content (e.g., Zhong and Schweidel 2020), and browsing histories (e.g., Trusov et al. 2016). Boughanmi and Ansari (2021) extend this line of work by using a nonparametric mixed-membership model to extract experiential music themes from user-generated tags. While LDA has also been used to improve predictive performance in purchase modeling (e.g., Jacobs et al. 2016; Kim and Zhang 2023), its application to causal inference—especially for estimating treatment effects at the theme level—remains limited. Our approach addresses this gap by using LDA not only to uncover latent behavioral patterns but also as a foundation for causal decomposition and customer-level targeting.

To demonstrate the utility of our framework, we apply it to data from a field experiment conducted by a major Asian personal care retailer, originally reported in Gopalakrishnan and Park (2021).[1] The retailer offers over 10,000 products across broad categories such as skincare and makeup. In the experiment, customers were randomly assigned to one of three conditions: a premium coupon ($10 off a minimum purchase of $20), a standard coupon ($7 off a minimum purchase of $20), or no coupon (control). For our analysis, we focus on the comparison between the premium-coupon and no-coupon groups, as this comparison yielded the most pronounced behavioral differences and provides a clear setting to illustrate how our framework uncovers heterogeneous treatment effects across latent themes.

This application presents several challenges that make it well-suited for our approach. First, the retailer's large assortment leads to sparse data at the individual product level, making conventional analyses difficult to implement reliably. Second, firm-defined product

[1] We thank the authors for generously sharing their data.

categories may not align how customers actually perceive and use products, particularly when items serve multiple purposes. Third, identifying which customers respond to specific themes is critical for improving targeting and optimizing promotional effectiveness.

Our analysis yields four key findings. First, we identify ten latent purchase themes that capture a broad range of customer needs, including functional daily care, deep hydration, and wellness. Many items appear across multiple themes, underscoring the limitations of rigid, firm-defined categories. Second, a small subset of themes—three out of ten, all related to distinct skincare needs—exhibit statistically significant treatment effects. While a traditional category-level analysis highlights the importance of skincare, it fails to reveal the specific needs driving these effects. Third, we observe substantial heterogeneity in treatment effects that customers are more responsive to themes they had previously engaged with, suggesting a reinforcement dynamic. Fourth, theme-based targeting of customers significantly outperforms conventional methods in generating returns. Together, these findings demonstrate the benefits of integrating causal inference with probabilistic machine learning: causal inference provides robust estimates of treatment effects, while the latent theme framework offers the behavioral granularity needed for more effective targeting.

The remainder of the paper is organized as follows. Section 2 presents our modeling framework for estimating treatment effects at the aggregate and theme levels. Section 3 outlines our inference procedure that jointly infers latent themes and estimates treatment effects. Section 4 describes a field experiment and data. Section 5 presents the results, including theme identification, treatment effects, heterogeneity, and targeting performance. Section 6 concludes with contributions, limitations, and directions for future research.

2 Treatment Effects at Aggregate and Theme Levels

This section presents our modeling framework for estimating the effects of marketing interventions on customer behavior across products. Section 2.1 outlines a standard approach for estimating aggregate-level treatment effects and discusses its limitations when analyzing effects at more disaggregated levels. Section 2.2 introduces the concept of latent purchase themes. Section 2.3 then explains how aggregate-level treatment effects can be decomposed into theme-level treatment effects. Section 2.4 details how these themes are inferred using the Latent Dirichlet Allocation framework. Finally, Section 2.5 extends the model to incorporate both observed and unobserved sources of customer heterogeneity.

2.1 Treatment Effects Across Levels of Analysis

Consider a firm (e.g., Sephora) offering a broad assortment of products (J) that seeks to determine the causal impact of a marketing intervention—such as a price coupon—on customer behavior (similar to the setting analyzed in our application). To this end, the firm conducts a randomized controlled trial (RCT), assigning customers to either a treatment group (N_1), which receives the intervention, or a control group (N_0), which does not.[2] Customer purchases are then tracked across the entire product assortment. Such experimental paradigm is common in practice, as firms routinely test various marketing interventions to influence customer behavior and improve performance—at the firm, product category, and individual product level.

A common approach to estimating the average treatment effect (ATE) in this context is through the following model specification:

$$y_i = \tau_0 + \tau_1 \cdot w_i + \varepsilon_i, \tag{1}$$

where y_i denotes the outcome of interest for customer i (e.g., total spend across the product assortment during the campaign period), and w_i is a binary indicator equal to 1 if customer i is assigned to the treatment group and 0 otherwise. The parameter τ_0 represents the expected outcome for the control group, and τ_1 captures the ATE of the intervention. The error term ε_i is assumed to be normally distributed, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

A key managerial decision when applying the above model is selecting the appropriate level of analysis—whether to use aggregated outcomes (e.g., total spend across all products per customer) or more disaggregated measures (e.g., customer spending by product category or individual product). The value of the analysis depends on how well the chosen level aligns with both managerial objectives and customer behavior. Each level of disaggregation involves trade-offs that affect the interpretability and usefulness of the results.

In practice, most firms organize their products into predefined categories (e.g., skincare, makeup, or body care, in the case of a personal care retailer) for purposes of inventory management, promotions, and reporting. Analyzing treatment effects at the category level is operationally convenient and often aligned with how firms make decisions. However, this firm-centric approach may mask meaningful variation in how customers respond to specific products within a category, potentially leading to overly broad or misleading conclusions. Moreover, rigid, predefined categories may not reflect the diverse ways in which customers

[2] Our framework can easily be extended to accommodate multiple treatment conditions.

use products to meet different needs. For example, a skincare product might be part of a daily skincare routine while also serving as an intensive hydrator, which are not captured by conventional classifications. While product-level analyses could, in theory, reveal such nuance, they often lack statistical power due to sparse purchase data at the item level.

These limitations underscore the need for a more flexible, data-driven framework that can uncover latent patterns in customer preferences while supporting downstream causal inference. We leverage a behaviorally grounded construct—purchase theme—that captures customer needs across the product assortment.

2.2 Purchase Themes

Extant research suggests that underlying purchase themes play a significant role in shaping customer behavior (e.g., Urban and Hauser 2004; Jacobs et al. 2016). Identifying these latent themes can uncover nuanced differences in customer preferences that are often overlooked by traditional category- or product-level analyses, and help managers develop more effective marketing strategies (e.g., Jacobs et al. 2021). In our framework, purchase themes refer to behaviorally meaningful groupings of products, capturing structured patterns in customer purchase behavior across the entire product assortment. Each customer may engage with multiple themes to varying degrees, with certain themes more prominent depending on individual preferences.

In what follows, for the ease of exposition, we first assume that purchase themes are observed and show how aggregate-level ATEs can be decomposed into theme-level ATEs. Let $y_{i,j}$ denote the outcome (e.g., spending) on product j by customer i following the treatment. If no purchase is made, then $y_{i,j} = 0$ for $j \in J$. Suppose there are K themes, and let $y_{i,k}$ denote the outcome associated with theme k for customer i. Then:

$$y_{i,k} = \sum_{j=1}^{J} \mathbb{I}_k(z_{ij})\, y_{i,j}, \tag{2}$$

where $\mathbb{I}_k(\cdot)$ is an indicator function equal to 1 if the theme associated with product j for customer i, denoted by z_{ij}, is equal to k and 0 otherwise.[3] Under this formulation, the outcome for theme k is the sum of the outcomes for all products that customer i associates with that theme.

[3] While z_{ij} could be simplified to z_j, indicating a hard association between a product and a theme, our model allows z_{ij} to vary by customer, capturing individual differences in how products are associated with themes.

Note that y_i, the total spending for customer i across all themes and products, can be expressed as:

$$y_i = \sum_{k=1}^{K} y_{i,k}. \tag{3}$$

Substituting Equation 2 into Equation 3 confirms that total spending is fully accounted for by the sum of spending across all themes and products:

$$y_i = \sum_{k=1}^{K} y_{i,k} = \sum_{k=1}^{K}\sum_{j=1}^{J} \mathbb{I}_k(z_{ij}) y_{i,j} = \sum_{j=1}^{J} y_{i,j}. \tag{4}$$

2.3 Theme-level Treatment Effects

When themes are observable, theme-level ATEs, $\tau_{1,k}$, can be estimated using the following model:

$$y_{i,k} = \tau_{0,k} + \tau_{1,k} \cdot w_i + \varepsilon_{i,k}, \tag{5}$$

where $\tau_{0,k}$ denotes the expected outcome for the control group, and $\tau_{1,k}$ capture treatment effect for theme k. The error term $\varepsilon_{i,k}$ is assumed to be normally distributed, $\varepsilon_{i,k} \sim \mathcal{N}(0, \sigma_k^2)$.

Because total spending is the sum of spending across all themes, the aggregate-level ATE, τ_1, can be expressed as the sum of the theme-level ATEs, $\tau_{1,k}$:

$$\tau_1 = \sum_{k=1}^{K} \tau_{1,k}. \tag{6}$$

This relationship can be confirmed by aggregating Equation 5 across all themes:

$$y_i = \sum_{k=1}^{K} y_{i,k} = \sum_{k=1}^{K} \tau_{0,k} + \left(\sum_{k=1}^{K} \tau_{1,k}\right) w_i + \sum_{k=1}^{K} \varepsilon_{i,k}. \tag{7}$$

By defining the sums, we can rewrite the equation as:

$$y_i = \tilde{\tau}_0 + \tilde{\tau}_1 \cdot w_i + \tilde{\varepsilon}_i, \tag{8}$$

where $\tilde{\tau}_0 = \sum_{k=1}^{K} \tau_{0,k}$, $\tilde{\tau}_1 = \sum_{k=1}^{K} \tau_{1,k}$ and $\tilde{\varepsilon}_i \sim \mathcal{N}(0, \tilde{\sigma}^2)$, with $\tilde{\sigma}^2 = \sum_{k=1}^{K} \sigma_k^2 + 2\sum_{k<k'} \text{cov}(\varepsilon_{i,k}, \varepsilon_{i,k'})$. The overall variance $\tilde{\sigma}^2$ incorporates both the individual theme-level variances and their pair-wise covariances. This confirms that $\tau_1 = \tilde{\tau}_1 = \sum_{k=1}^{K} \tau_{1,k}$. Given the linearity of the model

and the consistency of the estimators in Equations 1 and 5, the sum of the estimated theme-level effects, $\hat{\tau}_{1,k}$, converges asymptotically to the estimated aggregate-level treatment effect, $\hat{\tau}_1$, as the sample size increases.[4]

The above description assumes that purchase themes are directly observed. In most practical applications, however, themes are latent and must be inferred from customer data. We next describe how latent themes can be estimated and how this estimation modifies the identification of theme-level causal treatment effects.

2.4 Latent Purchase Themes

The model for inferring themes should satisfy several important criteria. First, it should reveal interpretable themes to enable managers to derive actionable insights. Second, it should infer both the number and composition of themes directly from the data, without relying on strong prior assumptions. Third, it must allow direct inference of customer-specific product-theme allocations (z_{ij}). Finally, it should flexibly assign each product to one or more themes with varying probabilities, reflecting the complexity of customer behavior and overlapping product preferences.

We adopt the Latent Dirichlet Allocation (LDA) model (Blei et al. 2003), which satisfies all of the criteria outlined above. LDA is well-suited to our context due to its transparency, interpretability, and ability to support probabilistic inference and uncertainty quantification through posterior sampling methods such as Gibbs sampling (Griffiths 2004). Compared to more complex models like neural networks or autoencoders, it is particularly advantageous for managerial interpretation and causal analysis. That said, LDA relies on several assumptions that may be critical depending on the application. Specifically, it assumes a fixed number of themes that remain constant post-treatment and adopts a bag-of-products representation, treating the order in which a customer purchases products as irrelevant. This order independence applies both to the sequence of products and to the customers themselves. In contexts where temporal dynamics or evolving preferences are critical, alternative models may be more appropriate for uncovering latent themes.

In the LDA framework, latent theme k is represented by a J-dimensional theme-product probability vector, $\boldsymbol{\phi}_k$, where each element ϕ_{kj} captures the strength of association between theme k and product j. A higher value of ϕ_{kj} indicates a stronger association, while a

[4]When $K = J$, such that each product defines its own theme, the model reduces to a full product-level decomposition of the treatment effect. In this case, the aggregate-level treatment effect τ_1 equals the sum of the product-level treatment effects $\tau_{1,j}$, such that $\tau_1 = \sum_{j=1}^{J} \tau_{1,j}$.

lower value reflects a weaker link. Although themes are common across customers, their relevance varies at the individual level. Customer i is characterized by a K-dimensional customer–theme probability vector, $\boldsymbol{\theta}_i$, where each element θ_{ik} denotes the importance of theme k to customer i. Larger values of θ_{ik} indicate more relevant themes, while smaller values reflect less relevant themes.

Finally, we assume that each customer's purchase aligns with a latent theme specific to that transaction. For each purchase n by customer i, a latent theme η_{in} is sampled from the customer's theme preference vector:

$$P(\eta_{in} = k \mid \boldsymbol{\theta}_i) = \theta_{ik}. \tag{9}$$

Given the selected theme η_{in}, the purchased product v_{in} is then drawn from the product distribution associated with that theme:

$$P(v_{in} = j \mid \boldsymbol{\phi}_{\eta_{in}}) = \phi_{\eta_{in} j}. \tag{10}$$

To analyze outcomes at the theme level, we leverage the probability that product j is associated with theme k for customer i, denoted $P(z_{ij} = k \mid \boldsymbol{\theta}_i, \boldsymbol{\phi}_k)$. This association is estimated using the posterior distributions of $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_k$ (Griffiths 2004). Under the conditional independence assumptions of the LDA model, this probability does not depend on a specific purchase instance n and can be expressed as:

$$P(z_{ij} = k \mid \boldsymbol{\theta}_i, \boldsymbol{\phi}_k) = P(\eta_{in} = k \mid v_{in} = j, \boldsymbol{\theta}_i, \boldsymbol{\phi}_k) \propto \theta_{ik} \cdot \phi_{kj}. \tag{11}$$

This probabilistic assignment acknowledges the inherent uncertainty in how products relate to themes, especially when products serve multiple customer needs. Our approach retains uncertainty in customer–product–theme relationships, and avoids the distortions that can bias theme-level outcome measures and causal estimates.[5]

[5] Note that η_{in} and z_{ij} are distinct but related latent variables. η_{in} denotes the theme associated with a specific purchase instance n by customer i, whereas z_{ij} captures the overall association between customer i and product j across all purchases. Under the conditional independence assumptions of the LDA model, both variables have the same posterior structure given $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}$, which allows us to equate their probabilities. Importantly, we retain these associations as probabilistic to maintain uncertainty in theme assignment and ensure valid inference.

2.5 Heterogeneity in Treatment Effects

Our framework accounts for both observed and unobserved sources of customer heterogeneity when estimating treatment effects. Let $\boldsymbol{x}_i$ represent a vector of M observed pre-treatment characteristics for customer i (e.g., demographics, prior purchase patterns), and let $\boldsymbol{d}_i$ denote a vector of S latent pre-treatment characteristics (e.g., prior thematic spending patterns). We define the full vector of customer characteristics as $\tilde{\boldsymbol{x}}_i = (\boldsymbol{x}_i, \boldsymbol{d}_i)$.

The heterogeneous treatment effect for customer i (HTE) at the aggregate level, denoted $\tau_{1,i}$, is modeled as:

$$\tau_{1,i} = \boldsymbol{\tau}_1^{h\top}(1, \tilde{\boldsymbol{x}}_i) = \tau_{1,0}^h + \underbrace{\sum_{m=1}^{M} \tau_{1,m}^h \tilde{x}_{i,m}}_{\text{Observed heterogeneity}} + \underbrace{\sum_{s=1}^{S} \tau_{1,M+s}^h \tilde{x}_{i,M+s}}_{\text{Unobserved heterogeneity}}, \tag{12}$$

where $\tau_{1,0}^h$ captures the baseline treatment effect, and the remaining terms account for heterogeneity across both observed and unobserved dimensions. We estimate these heterogeneous effects using the following outcome model:

$$y_i = \tau_{0,i} + \tau_{1,i} \cdot w_i + \varepsilon_i^h \tag{13}$$

where $\tau_{0,i}$ is the baseline outcome under control and $\varepsilon_i^h \sim \mathcal{N}(0, \sigma^{2h})$.

We extend Equation 12 to estimate HTEs at the theme level. For customer i and theme k, the treatment effect, $\tau_{1,i,k}$, is modeled as:

$$\tau_{1,i,k} = \boldsymbol{\tau}_{1,k}^{h\top}(1, \tilde{\boldsymbol{x}}_i) = \tau_{1,0,k}^h + \underbrace{\sum_{m=1}^{M} \tau_{1,m,k}^h \tilde{x}_{i,m}}_{\text{Observed heterogeneity}} + \underbrace{\sum_{s=1}^{S} \tau_{1,M+s,k}^h \tilde{x}_{i,M+s}}_{\text{Unobserved heterogeneity}}. \tag{14}$$

The corresponding theme-level outcomes are modeled as:

$$y_{i,k} = \tau_{0,i,k} + \tau_{1,i,k} \cdot w_i + \varepsilon_{i,k}^h, \tag{15}$$

where $\varepsilon_{i,k}^h \sim \mathcal{N}(0, \sigma_k^{2h})$. Consistent with the ATEs decomposition, this framework allows for a parallel decomposition of HTEs across themes, such that the aggregate-level heterogeneity vector can be expressed as the sum of theme-specific components: $\boldsymbol{\tau}_1^h = \sum_{k=1}^{K} \boldsymbol{\tau}_{1,k}^h$.

Table 1 summarizes how our framework distinguishes between aggregate- and theme-level analyses. It also clarifies the distinction between ATEs and HTEs at each level, providing deeper insight into how interventions influence different customers and behavioral themes.

Table 1: Comparison of Treatment Effects at Aggregate and Theme Levels

	Outcomes	Average Treatment Effect	Heterogeneous Treatment Effect
Aggregate-level	y_i	τ_1 (Equation 1)	$\boldsymbol{\tau}_1^h$ (Equation 13)
Theme-level	$y_{i,k}$	$\tau_{1,k}$ (Equation 5)	$\boldsymbol{\tau}_{1,k}^h$ (Equation 15)

3 Inference

This section outlines our inference procedure, which jointly identifies latent themes and decomposes treatment effects across them. The approach addresses key methodological challenges, including accounting for uncertainty in theme estimation and ensuring a valid causal interpretation of the resulting treatment effects.

To ensure the causal validity of our estimates, it is essential that the inference of latent themes is not influenced by the treatment. If theme estimation were based on post-treatment behavior, the resulting themes could capture changes caused by the treatment, thereby biasing the theme structure and undermining the causal interpretation of theme-level effects (e.g., Angrist and Pischke 2009; Imbens and Rubin 2015). To avoid this, we estimate latent themes and their associated parameters using only pre-treatment data. This ensures that theme definitions are exogenous to the treatment, allowing for a clean and unbiased decomposition of treatment effects across latent purchase themes.

Furthermore, our inference procedure explicitly incorporates three key sources of uncertainty, each of which plays a critical role in accurately estimating theme-level outcomes and treatment effects. First, we account for uncertainty in individual customer-theme probabilities ($\boldsymbol{\theta}_i$), which capture how likely each customer is to engage with various themes. Second, we incorporate uncertainty in the theme-product distributions ($\boldsymbol{\phi}_k$), which reflect how representative each product is of a given theme. Third, we model product-to-theme allocations (z_{ij}) as latent variables, using their posterior probabilities $P(z_{ij} = k \mid \boldsymbol{\theta}_i, \boldsymbol{\phi}_k)$. These probabilities capture the uncertainty in how each customer's purchases relate to different themes. Together, these three components form the foundation of our estimation framework. By accounting for uncertainty at every level of the model, we improve the robustness and credibility of our causal estimates. In contrast, ignoring uncertainty in these latent parameters can lead to biased or misleading conclusions about treatment effects.

3.1 Algorithm

A key contribution of our inference procedure is its joint estimation of latent themes and theme-level treatment effects. We first infer LDA parameters from pre-treatment data, then estimate treatment effects based on posterior samples of the inferred structure. This joint estimation is implemented via Gibbs sampling, which allows uncertainty in the latent variables to be fully propagated through to the treatment effect estimates.

Following the standard LDA formulation (Blei et al. 2003), we place symmetrical Dirichlet priors on customer-theme and theme-product distributions. Specifically, customer-theme probabilities are modeled as $\boldsymbol{\theta}_i \sim \text{Dirichlet}(\alpha, \ldots, \alpha)$, where α is the concentration parameter. Similarly, theme-product probabilities are modeled as $\boldsymbol{\phi}_k \sim \text{Dirichlet}(\beta, \ldots, \beta)$, with β as the corresponding concentration parameter.

For each sampling iteration ℓ, we begin by sampling theme assignments for observed purchases based on pre-treatment data. This step ensures that themes inference is not influenced by the treatment. Specifically, for customer i and each of their D_i pre-treatment purchases, we sample the theme allocation η_{in} for purchase n using:

$$P(\eta_{in} = k | \eta_{\neg in}, \boldsymbol{v}, \alpha, \beta) \propto (n^c_{ik,\neg in} + \alpha) \cdot \frac{(n^f_{kv_{in},\neg in} + \beta)}{\sum_{j=1}^{J} n^f_{kj,\neg in} + J\beta}, \tag{16}$$

where $n^c_{ik,\neg in}$ is the number of purchases by customer i assigned to theme k, and $n^f_{kj,\neg in}$ is the number of times product j is assigned to theme k across all customers, both excluding purchase instance n of customer i. For further details on collapsed Gibbs sampling for the LDA model, see Griffiths (2004).

Next, we utilize $\{\{\eta_{in}\}_{n=1}^{D_i}\}_{i=1}^{N}$ to update the customer-theme counts n^c_{ik} and product-theme counts n^f_{kj}. As in Heinrich (2005), these counts are then leveraged to sample $\boldsymbol{\theta}_i$ for customer i using:

$$\theta_i \sim \text{Dirichlet}\left(\alpha + \boldsymbol{n}^c_i\right), \tag{17}$$

and $\boldsymbol{\phi}_k$ for each theme k using:

$$\phi_k \sim \text{Dirichlet}\left(\beta + \boldsymbol{n}^f_k\right). \tag{18}$$

Here, $\boldsymbol{n}^c_i = (n^c_{i1}, \cdots, n^c_{iK})$ is the vector of counts of purchases by customer i assigned to theme k, and $\boldsymbol{n}^f_k = (n^f_{k1}, \cdots, n^f_{kJ})$ is the vector of counts of product j assigned to theme k across all customers at iteration ℓ.

This step accounts for uncertainty in the posterior estimates of customer preferences and theme distributions. Based on these posterior estimates, we sample product-theme assignments z_{ij} for each customer and product, accounting for the uncertainty in product-to-theme allocations, as specified in Equation 11.

Theme-level outcomes for iteration ℓ, $y_{i,k}$, are computed based on the sampled product-theme assignments, incorporating the uncertainty in how product-level outcomes map to themes:

$$y_{i,k} = \sum_{j=1}^{J} \mathbb{I}(z_{ij})\ y_{i,j}.$$

Additionally, recall that a customer's latent characteristics, $\boldsymbol{d}_i$, are inferred from pre-treatment data. The specific composition of $\boldsymbol{d}_i$ can vary by context to provide the most actionable insights for managers. In our application, we define $\boldsymbol{d}_i$ as the customer's pre-treatment spending across the K latent themes. Thus, $\boldsymbol{d}_i$ is a K-dimensional vector (i.e., $S = K$) that captures how pre-treatment thematic expenditures influence treatment effects.

Note that uncertainty in these latent characteristics should be accounted for. To this end, we consider samples of $\boldsymbol{d}_i$ within each MCMC iteration, given by:

$$d_{i,k} = \sum_{j=1}^{J} \mathbb{I}_k(z_{ij})\, y_{i,j}^{pre}, \tag{19}$$

where $y_{i,j}^{pre}$ indicates the total amount that customer i spent on product j prior to the treatment. Here $\mathbb{I}_k$ is an indicator function equal to 1 if product j is assigned to theme k, and 0 otherwise.

Finally, we sample the regression parameters, including treatment effects for each theme. This inference procedure enables the joint estimation of latent themes and their effects, while maintaining independence from treatment during theme estimation. Additionally, it accounts for the uncertainty from the LDA model parameters into the causal inference process, ensuring robust and interpretable results. Algorithm 1 outlines our inference procedure.[6]

[6] Appendix A provides details on the full conditionals for the regression parameters. Appendix B presents simulation results that demonstrate the effectiveness of our approach in identifying latent theme compositions and estimating treatment effects.

Algorithm 1 Joint Estimation of Latent Themes and Treatment Effects

Initialize: Randomly assign purchases to themes η_{in}
Compute initial counts:
  n^c_{ik}: number of customer i's products assigned to theme k
  n^f_{kj}: number of times product j is assigned to theme k

for iteration $\ell = 1, \ldots, I$ **do**
  for customer $i = 1, \ldots, N$ **do** ▷ Purchase-theme assignments
    for purchase $n = 1, \ldots, D_i$ **do**
      Sample η_{in} using Equation 16
      Update counts $n^c_{i\eta_{in}}$ and $n^f_{\eta_{in}j}$
    end for
  end for

  Sample $\boldsymbol{\theta}_i$ for $i = 1, \ldots, N$ using Equation 17 ▷ Customer-theme probabilities
  Compute $\boldsymbol{\phi}_k$ for $k = 1, \ldots, K$ using Equation 18 ▷ Product-theme probabilities

  for customer $i = 1, \ldots, N$ **do**
    for product $j = 1, \ldots, J$ **do**
      Sample z_{ij} using Equation 11 ▷ Product-theme assignments
    end for
    for theme $k = 1 = 1, \ldots, K$ **do**
      Compute $y_{i,k}$ using Equation 2 ▷ Theme-level outcomes
      Compute $d_{i,k}$ using Equation 19 ▷ Theme-level pre-treatment spending
    end for
  end for
  ▷ Regression parameters
  Sample τ_0, τ_1, and σ based on y_i ▷ Use Equations A.2 and A.3 in Web Appendix A
  Sample $\boldsymbol{\tau}^h_0$, $\boldsymbol{\tau}^h_1$, and σ^h based on y_i
  for theme $k = 1, \ldots, K$ **do**
    Sample $\tau_{0,k}$, $\tau_{1,k}$, and σ_k based on $y_{i,k}$
    Sample $\boldsymbol{\tau}^h_{0,k}$, $\boldsymbol{\tau}^h_{1,k}$, and σ^h_k based on $y_{i,k}$
  end for
end for

4 Application: A Price Coupon Experiment

We demonstrate the utility of our proposed framework using data from a randomized field experiment on coupon promotions, originally reported in Gopalakrishnan and Park (2021). Briefly, the experiment was conducted by a leading Asian personal care retailer to evaluate the impact of coupon-based promotions on various customer outcomes along the online conversion funnel, such as website visits, product searches, and purchases, across a heterogeneous customer base. To this end, they conducted a series of experiments where customers were randomly assigned to one of three conditions: a standard coupon ($7 off a minimum purchase of $20), a premium coupon ($10 off a minimum purchase of $20), or no coupon (control group). For the purposes of our analysis, we focus on the comparison that generated the largest monetary impact on customer behavior: the premium-coupon group (treatment) versus the no-coupon group (control).

Although the experiment was originally designed to examine the impact of coupons on customer behavior along the online conversion funnel, it also provides an ideal setting to demonstrate the benefits of our proposed framework in analyzing more granular aspects of customer behavior. The retailer offers a wide assortment of over 10,000 SKUs, spanning a broad range of price points from budget-friendly to premium, with products organized into mutually exclusive, predefined categories (e.g., skincare, makeup).

The dataset for our analysis includes 4,247 customers, of whom 3,446 received the premium coupon and 801 received no promotional message. Randomization checks reported in Table 1 (Panel B) of Gopalakrishnan and Park (2021) support the validity of the random assignment. In Section 5, we introduce the latent themes identified from customer purchase data and present additional randomization checks by comparing the theme distributions across treatment and control groups.

The data consist of two components: (1) detailed transaction records at both the individual and product levels, including timestamps, purchase amounts, product descriptions, brands, and categories, and (2) demographic information such as age and gender, which allows us to account for customer heterogeneity. We observe customer purchase behavior over a 12-month pre-treatment period and during the one-week campaign window. Because our goal is to decompose the treatment effect of the coupon into underlying thematic responses, we use total spending during the campaign as the primary outcome measure.[7]

[7] All purchase amounts were originally recorded in the local currency and converted to U.S. dollars using the average exchange rate over the data period.

5 Results

This section presents our results in four parts. First, we identify ten latent themes from pre-treatment transactions and show how they capture meaningful variation in customer needs that extend beyond the firm's existing product taxonomy. Second, we examine ATEs, both in aggregate and at the theme level, and compare them to estimates based on conventional product categories. Third, we analyze HTEs by observed and unobserved customer characteristics, uncovering patterns of promotional responsiveness. Finally, we assess the managerial relevance of our framework by evaluating its effectiveness in customer targeting.

5.1 Latent Purchase Themes

We begin by identifying latent themes that characterize how customers relate to different products based on their actual purchase behavior. To this end, we apply an LDA model (Blei et al. 2003) to pre-treatment transaction data. Based on a balance between predictive performance and interpretability, we select ten themes that offer a parsimonious yet expressive representation of customer-product associations.[8] This theme-based perspective offers a behaviorally grounded alternative to the firm's existing product taxonomy, which includes five broad categories—skincare, makeup, hair care, bath and body care, and a consolidate all "other" category comprising fragrances, tools, and accessories.[9] While useful for operational purposes, these predefined groupings may overlook how customers combine products to meet their specific needs. In contrast, the latent themes revealed by our model uncover demand structures that are influenced by how customers actually shop.

Table 2 summarizes the ten identified latent themes along with representative product types.[10] Unlike the firm's predefined taxonomy, which assigns each product to a single category (e.g., skincare, makeup, hair care), these themes emerge from actual co-purchase behavior and often span multiple categories, capturing holistic customer routines. This approach reveals both category-focused needs (e.g., intensive hydration) and broader lifestyle patterns (e.g., natural indulgence and gifting). For example, Theme 1 (Natural Soothing Care) aligns primarily with skincare, emphasizing calming and hydrating routines using natural ingredients such as sheet masks and toners. Theme 2 (Functional Daily Care) spans

[8] Appendix C provides additional details on model estimation, including MCMC convergence diagnostics and the procedure used for selecting the number of themes.

[9] As is common in the industry, these categories are further subdivided into subcategories—e.g., cleansers and moisturizers within skincare—to cover a wide range of items.

[10] Due to a non-disclosure agreement with the collaborating firm, products are reported by type rather than specific brand or item name. The themes were validated in consultation with the retailer.

skincare, body care, and hair care, reflecting practical, results-oriented routines for everyday use (e.g., cleansers, body lotions, hair treatments). Theme 4 (Intensive Hydration) is rooted in skincare, featuring deep moisturizers, sleeping masks, and emulsions for skin barrier repair. Theme 6 (Makeup & Daily Essentials) integrates makeup and and basic skincare, supporting streamlined daily routines with cushions, lip tints, and cleansers. Finally, Theme 7 (Natural Indulgence & Gifing) includes aesthetically appealing, family-friendly products across skincare and body care, oriented toward treat-yourself and gifting occasions (e.g., body mists, creams, hydrating ampoules).

Table 2: Summary of Latent Themes and Top Product Types

	Theme Name	Top Product Types
1	Natural Soothing Care	moisturizing mask, nourishing mask, soothing mask, mask pack, brightening mask, massage mask, hydrating mask, firming mask, air cushion, toner
2	Functional Daily Care	essence mask, fit mask, pouch pack, ampoule mask, hand mask, blackhead nose patch, cleanser, hair treatment, body lotion, soothing gel
3	Gentle & Sensitive Skincare	hydrogel mask, herb soap, toothbrush, hair treatment, toner, emulsion, vitamin supplement, balancing water, emulsion, shampoo
4	Intensive Hydration	air cushion, moisturizer, sun protector, skin refiner, foundation, essence, toner, emulsion, emulsion, sleeping mask
5	Purifying & Wellness Boost	blackhead patch, health supplement, foundation, energy ampoule, cleansing oil, cleanser, toner, scrub, slimming supplements, hair serum
6	Makeup & Daily Essentials	cushion puff, toner, lip tint, cleanser, BB cushion, eyeshadow, emulsion, sunscreen, foundation, cleansing wipes
7	Natural Indulgence & Gifting	body mist, cream, facial mask, promotional gift, cleanser, hydrating ampoule, cleansing oil, nail polish, bag & pouch, sun cushion
8	Oral Hygiene & Scalp Health	toothpaste, shampoo, mouthwash, hair treatment, shampoo, shampoo, toothpaste, shampoo, hair pack, hair treatment
9	Botanical Skincare Rituals	facial mask, toner, cream, cushion, emulsion, essence, sun protector, cleanser, serum, cleansing oil
10	Beauty Accessories & Tools	cushion puff, cotton pads, nail polish, lip & eye remover, hair mask, eyeshadow, eyelashes, nail polish, nail color, foundation

To visualize how these themes differ from the firm's conventional product categories, Figure 1 presents a two-dimensional t-Distributed Stochastic Neighbor Embedding (t-SNE) projection (Maaten and Hinton 2008) of the posterior theme-product distribution for skincare products.[11] Each point represents a skincare product and is color-coded by its most likely theme assignment.[12] The presence of distinct clusters suggests coherent thematic groupings, while overlaps among clusters indicate that many products span multiple themes, reflecting their varied roles skincare items play in customer behavior. As shown in Figure 1, Theme 4 (12% of skincare purchases), Theme 5 (8%), and Theme 9 (11%) all overlap with the skincare category but differ substantially in terms of customer needs and usage patterns.

[11] t-SNE is a dimensionality reduction technique that projects high-dimensional data into two dimensions, allowing for the visualization of product-theme relationships and the underlying structure among themes.

[12] Prominent themes are color-coded to enhance visual interpretability. Products associated with less prominent themes are displayed in gray with distinct markers, indicating lower assignment certainty.

These overlapping yet distinct clusters highlight the added granularity of our theme-based approach, which captures nuanced patterns in customer perception and product use that are not evident from the firm's standard category taxonomy.

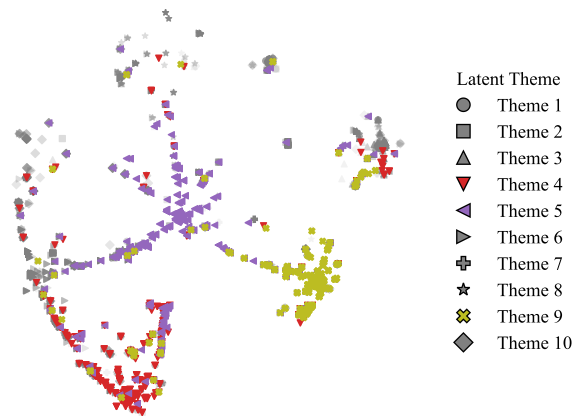

Figure 1: t-SNE Visualization of Skincare Products

To further examine this flexibility, we analyze the distribution of theme associations at the product level. For each product, we identify the theme with the highest posterior probability and tally how often each product appears across different themes. As shown in Figure 2, approximately 52% of the products are associated with more than one theme. For example, a skincare item like *Repair Serum* appears in seven themes (Themes 1, 2, 4, 5, 6, 8, and 10), ranging from everyday care to specialized hydration. These overlaps underscore the multidimensional roles that products play in meeting diverse customer needs—insights that are often obscured by rigid, one-dimensional categorizations.

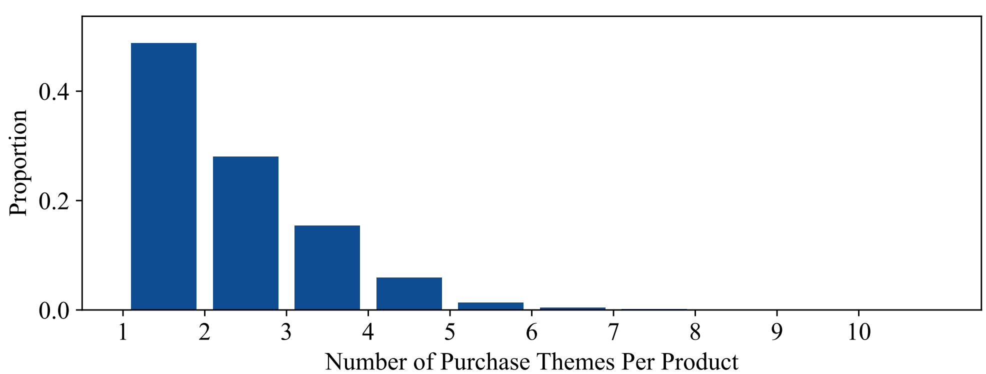

Figure 2: Distribution of Theme Frequency per Product

Having established the latent themes, we next assess whether the treatment and control groups are balanced with respect to the inferred customer-theme probabilities. As an additional diagnostic, we compare the distribution of the ten latent themes between the two groups. As shown in Figure 3, there are no statistically significant differences for any of the themes, confirming that randomization achieves balance at baseline with respect to latent themes. This result also indicates that our inference procedure is agnostic to treatment assignment, thereby supporting causal interpretability of the estimated treatment effects.

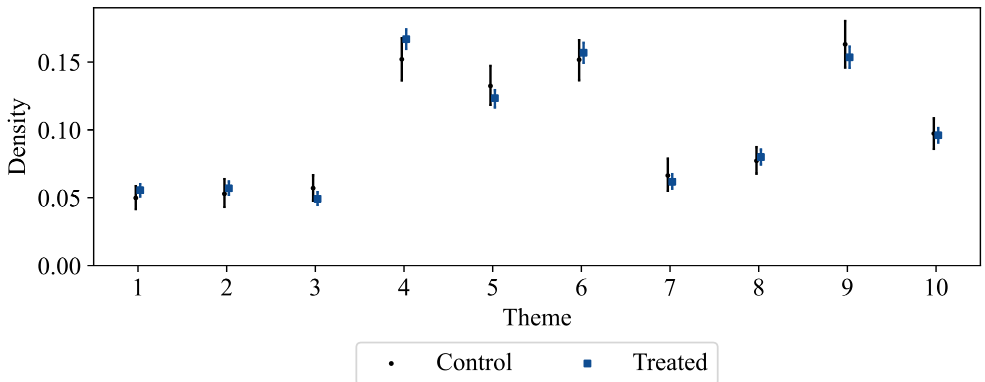

Figure 3: Distribution of Latent Themes Across Groups

In summary, these results highlight that the inferred themes capture meaningful behavioral structure while maintaining balance across experimental conditions. By leveraging this nuanced structure, our theme-based framework provides a flexible foundation for downstream causal analysis. As we show in the following sections, these themes not only summarize purchase behavior more effectively but also highlight which customer needs are most responsive to marketing interventions.

5.2 Average Treatment Effects

We examine the ATEs to assess how the coupon promotion influences overall customer spending. Our aggregate-level analysis shows that customers in the treatment group spend \$1.93 more than those in the control group, with a 95% credible interval of [1.07, 2.80]. This estimate is consistent with the findings reported by Gopalakrishnan and Park (2021).

To better understand the sources of the observed effect, we decompose the ATEs using our theme-based framework. Table 3 reports the estimated ATEs for the ten latent themes inferred from pre-treatment transaction data.[13] Interestingly, only a subset of themes exhibit statistically significant treatment effects at the 95% credible level. Theme 9 shows the largest

[13]The posterior mean of the theme-level ATEs sums to \$1.94 (95% CI: [1.14, 2.75]), closely aligning with the aggregate-level results.

effect, with an estimated lift \$0.64 (95% CI: [0.16, 1.15]), followed by Theme 4 (\$0.41; 95% CI: [0.02, 0.87]) and Theme 5 (\$0.35; 95% CI: [0.01, 0.67]). All three themes relate to skincare but reflect distinct underlying customer motivations. Theme 9 focuses on botanical skincare routines and accounts for over one-third of the total treatment effect. Theme 4 is associated with advanced moisturizing products, while Theme 5 emphasizes holistic, wellness-oriented care. These results illustrate that customer responsiveness to the intervention is shaped not just by product types, but by specific needs customers seek to fulfill through those products.

Table 3: Theme-level Average Treatment Effects

Theme	Estimate	95% Credible Interval
1	0.08	[−0.04, 0.18]
2	0.04	[−0.07, 0.14]
3	0.08	[−0.14, 0.30]
4	0.41*	[0.02, 0.87]
5	0.35*	[0.01, 0.67]
6	0.19	[−0.05, 0.44]
7	0.06	[−0.07, 0.17]
8	0.06	[−0.09, 0.22]
9	0.64*	[0.16, 1.15]
10	0.03	[−0.12, 0.18]

Note: * indicates that the 95% Bayesian interval excludes 0.

Moreover, we compare the ATEs from our approach to those based on the firm's five product categories. Table 4 shows that the skincare category exhibits a significant increase of \$1.52 (95% CI: [0.78, 2.20]), accounting for nearly 80% of the overall treatment effect. A smaller but still significant increase of \$0.33 (95% CI: [0.07, 0.64]) is observed in the "other" category, while the remaining categories show no statistically significant effects.[14]

Together, Tables 3 and 4 highlight the advantages of our theme-based decomposition. While category-level analysis captures broad patterns in customer response, it lacks the granularity to reveal which specific customer needs are driving these effects. In contrast, our theme-based approach shows that only a few skincare-related themes respond strongly to the promotion—offering a more behaviorally grounded and actionable view of customer response. This insight is especially valuable for managers seeking to evaluate the causal

[14] We also conducted additional analyses based on the firm's 64 subcategories and 185 sub-subcategories, but found limited insight due to data sparsity. As product categorization becomes more granular, the number of statistically significant treatment effects declines substantially, reflecting the increasingly sparse data within each group. Results based on these finer categorizations are available from the authors upon request.

Table 4: Category-level Average Treatment Effects

Category	Estimate	95% Credible Interval
Skin Care	1.51*	[0.07, 0.64]
Makeup	0.01	[−0.22, 0.26]
Hair Care	0.00	[−0.00, 0.00]
Bath & Body Care	0.08	[−0.14, 0.30]
Other	0.33*	[0.07, 0.64]

Note: * indicates that the 95% Bayesian interval excludes 0.

impact of marketing interventions.

5.3 Heterogeneous Treatment Effects

To explore whether treatment effects vary across customers, we assess HTEs using both theme- and category-level frameworks.[15] Our analysis includes five observed characteristics: gender, age, and RFM (recency, frequency, and moneytary) metrics—as well as unobserved heterogeneity captured through pre-treatment spending across the ten inferred themes.[16]

Figure 4 presents theme-level HTEs by observed customer characteristics. While none of the differences is statistically significant at the 95% credible level, several directional patterns suggest potential behavioral variation. For example, more recent customers appear more responsive to Theme 9, and female customers show higher estimated effects in Themes 3 ($0.48), 4 ($0.99), and 5 ($0.86), but lower responsiveness in Theme 9.

Next, we examine unobserved heterogeneity by evaluating how pre-treatment spending on each theme influences treatment effects. Figure 5 presents theme-level HTEs on pre-treatment theme spending. The positive diagonal elements indicate that customers tend to respond more strongly to themes they previously engaged with. In Themes 1, 5, and 9, these within-theme effects are statistically significant: each additional dollar of pre-treatment spending corresponds to approximately a one-cent increase in the treatment effect. This pattern reveals a self-reinforcing dynamic in promotional response: coupon interventions are more likely to reinforce existing preferences than to encourage exploration of new themes. While promotions can effectively deepen engagement within a given theme, they appear less

[15] Results for the aggregate-level analysis are available from the authors upon request.

[16] Gender is coded as a binary indicator (1 if female). The remaining variables are dichotomized based on median splits. For recency, customers with more recent interactions are coded as 1. Note that the median split on monetary value breaks its collinearity with pre-treatment theme-level spending, allowing the latter to be estimated.

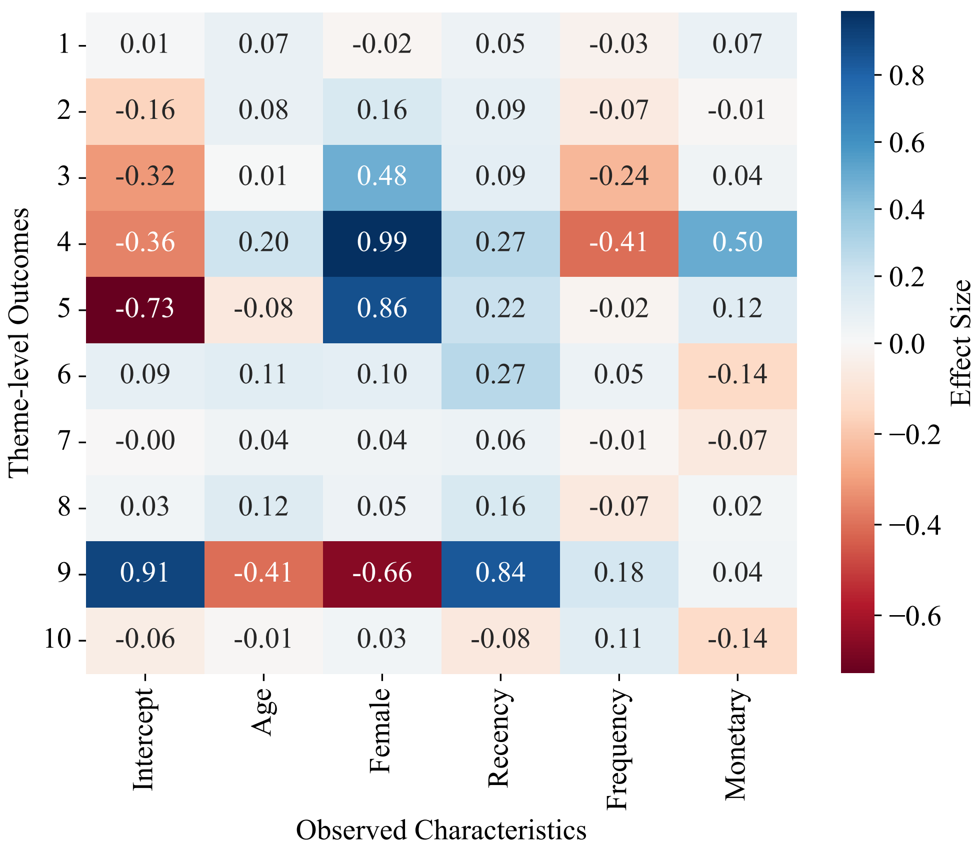

Figure 4: Observed Heterogeneity in Treatment Effects

Note: * indicates that the 95% Bayesian credible interval excludes 0.

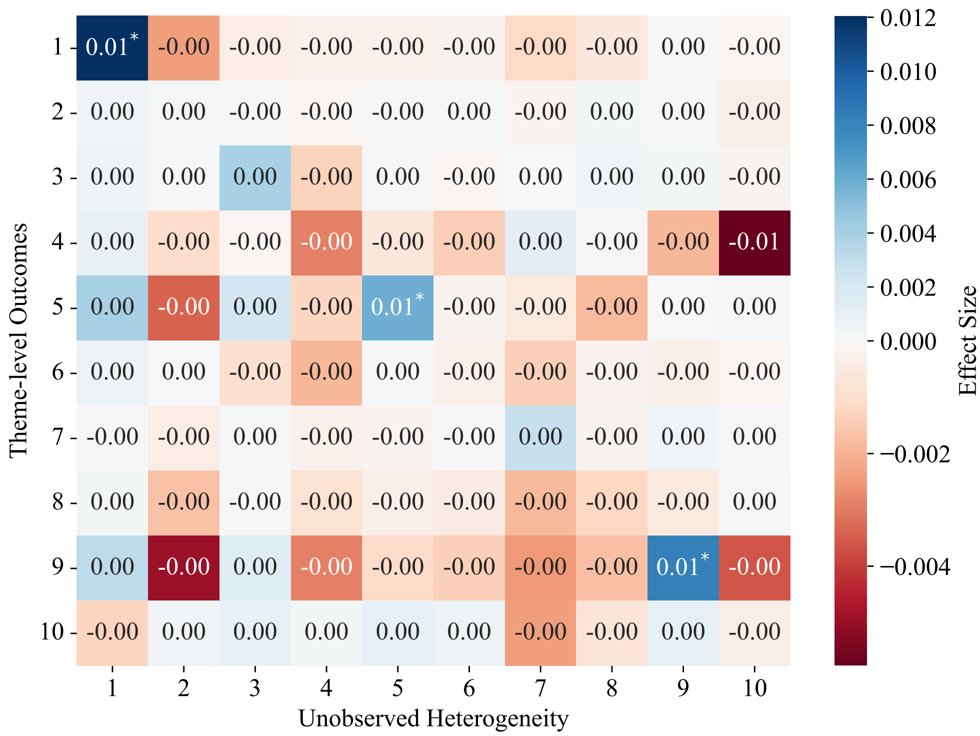

Figure 5: Unobserved Heterogeneity in Treatment Effects

Note: * indicates that the 95% Bayesian credible interval excludes 0.

effective at driving cross-theme discovery.

We compare these results with those obtained from using the firm's predefined categories. Figure 6 presents observed HTEs at the category level. The makeup category shows a statistically significant increase ($0.62) among recent customers, while other subgroup differences, including $1.49 increase in skincare among female customers, are not statistically significant.

A similar pattern emerges in Figure 7, which presents category-level HTEs based on pre-treatment spending within each product category.[17] In this case, the patterns are weak and difficult to interpret. For instance, higher body care spending is modestly associated with greater treatment response in the makeup category, while pre-treatment hair care spending corresponds to a slight decline in effects within the "other" category. These inconsistent patterns stand in sharp contrast to the significant diagonal elements observed in the theme-based analysis.

In summary, these results underscore the diagnostic value of our theme-based framework. While the category-level analysis offers limited insight into which subgroups or behavioral

[17] The median split on monetary value breaks its collinearity with pre-treatment category-level spending, allowing the latter to be estimated.

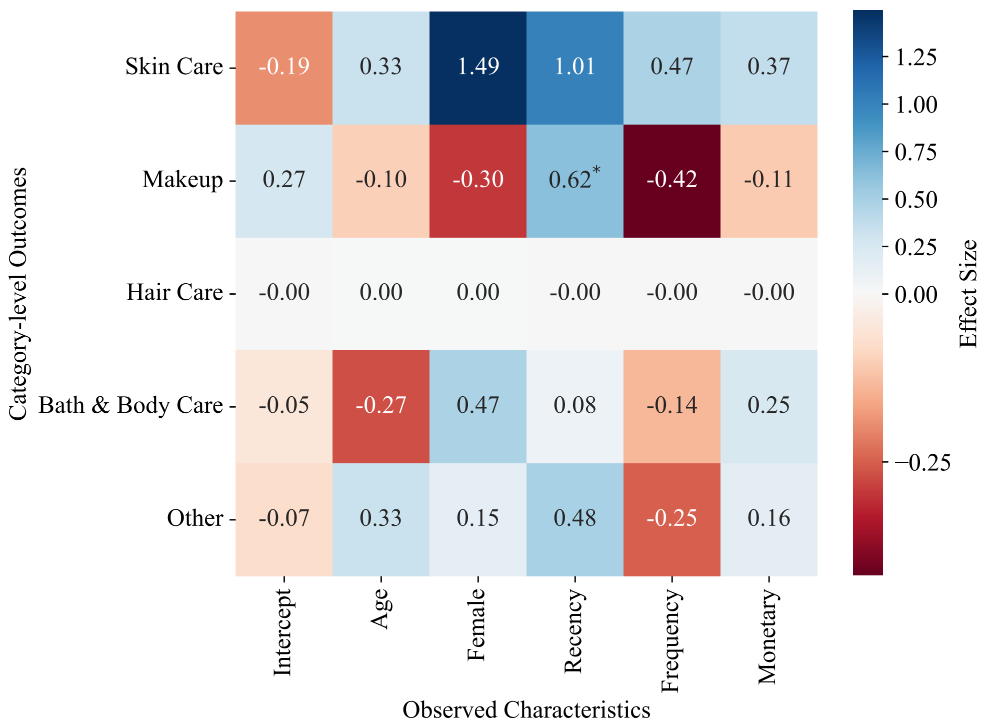

Figure 6: Observed Heterogeneity in Category-Level Treatment Effects

Note: * indicates that the 95% Bayesian credible interval excludes 0.

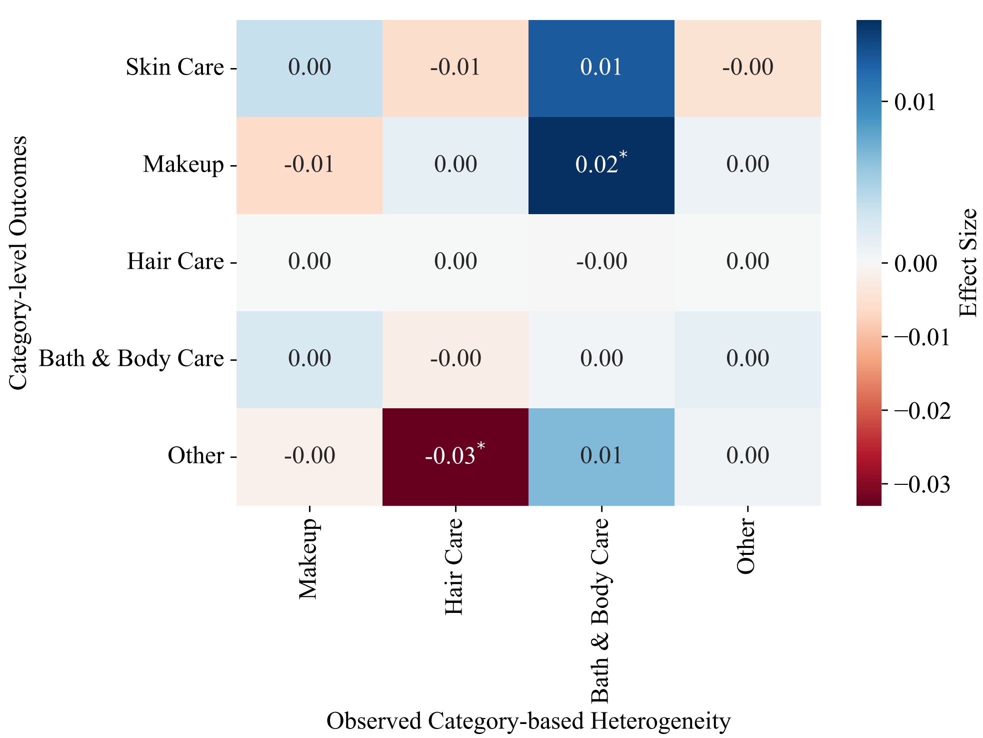

Figure 7: Category Heterogeneity in Category-Level Treatment Effects

Note: * indicates that the 95% Bayesian credible interval excludes 0.

themes respond to coupon promotions, the theme-level HTEs reveal two key findings. First, not all themes are equally responsive, highlighting the value of moving beyond broad product categories to uncover actionable substructure. Second, customer response tend to reinforce existing preferences rather than shift them, suggesting that promotions often strengthen established theme-level purchase patterns. These insights provide a more nuanced understanding of whom to target and how to tailor marketing campaigns.

5.4 Customer Targeting

A key advantage of RCTs is that they enable firms to conduct model-based evaluations of alternative targeting strategies. In this section, we evaluate the managerial value of our theme-based framework by comparing its targeting performance to that of conventional approaches based on customer demographics and RFM measures, and category-level spending. Our approach infers each customer's pre-treatment engagement with latent themes, yielding a detailed behavioral profile of thematic spending patterns prior to the intervention. We then evaluate whether this richer characterization leads to improved targeting performance.

Figure 8 compares the cumulative revenue generated by four targeting strategies: (1) random targeting, which selects customers without using any information and serves as a baseline; (2) demographics-based targeting, which uses customer demographic variables and RFM metrics; (3) category-based targeting, which augments demographics and RFM variables with pre-treatment spending across predefined product categories; and (4) theme-based targeting (our approach), which incorporates pre-treatment theme-level spending alongside demographics and RFM variables. For each of the three model-based approaches (excluding random targeting), we estimate individual-level treatment effects ($\tau_{1,i}$) and rank customers in descending order of predicted responsiveness. To ensure comparability, cumulative revenue is computed using treatment effect estimates from our theme-based model for all four strategies.

The results show that theme-based targeting consistently outperforms the alternatives. For example, when targeting the top 50% of customers, our approach yields an expected revenue of \$7,148, compared to \$5,879 for the category-based model, \$5,360 for the demographics-only model, and \$3,990 under random targeting. These differences represent gains of 79% over random targeting, 33% over demographics-only, and 22% over category-based model.

The superior performance of our approach stems from its ability to uncover behaviorally meaningful themes that reflect the underlying motivations behind customer purchases. These themes not only capture past behavior with greater granularity but also improve predictions

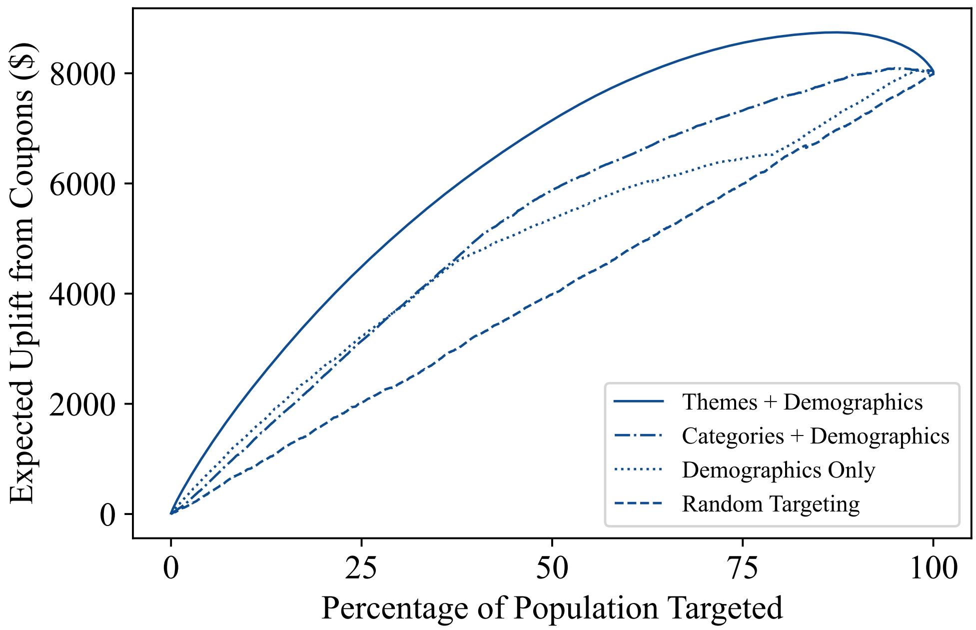

Figure 8: Cumulative Uplift of Customer Targeting

of promotional responsiveness—revealing patterns that conventional approaches often overlook. In contrast, demographics-only and category-based models rely exclusively on observed customer characteristics.

6 Conclusion

Understanding the causal impact of marketing interventions is essential for informed decision-making. Equally important is identifying which types of products—or more precisely, which underlying customer needs—are most responsive to these interventions. Such insights enable firms to focus their marketing efforts where they are most likely to generate meaningful returns. However, traditional approaches often fall short in this regard. Product-level analyses can provide detailed insights but are typically impractical at scale due to data sparsity and limited statistical power. Category-level analyses mitigate some of these challenges by grouping products into broader classifications, but they rely on rigid, predefined taxonomies that may mask meaningful within-category variation and fail to reflect how customers actually perceive and use products.

This paper presents a general modeling framework that addresses these limitations by leveraging a probabilistic machine learning approach to analyze treatment effects across latent purchase themes in settings with large product assortments. These latent themes represent customer-centric, behaviorally grounded groupings of products inferred from observed purchase patterns. The framework jointly estimates both the latent themes and treatment effects—average and heterogeneous—while explicitly accounting for uncertainty in product-to-theme assignments.

We apply this framework to data from a randomized field experiment originally reported in Gopalakrishnan and Park (2021). Although the coupon intervention leads to a statistically significant increase in overall spending, only a subset of latent themes drive the observed effects. These themes—primarily within skincare—capture distinct needs such as functional daily care, deep hydration, and wellness. Treatment effects also vary across customers and are stronger for themes they had previously engaged with—patterns often overlooked by category-based analyses. Building on these insights, we demonstrate that theme-based targeting approach yields significantly higher returns than conventional methods, underscoring the framework's practical relevance. More broadly, our results underscore the value of integrating causal inference with probabilistic machine learning: causal inference enables robust estimation of treatment effects, while the latent theme framework provides the behavioral granularity needed for more effective targeting.

There are several limitations of the proposed model that warrant consideration and offer avenues for future research. First, the number of themes is assumed to be fixed and known a priori. While non-parametric alternatives such as the Hierarchical Dirichlet Process can relax this assumption, they introduce significant computational complexity. Second, the current framework treats themes as static, which is reasonable given the one-week duration of the campaign. Incorporating temporal dynamics could provide a more nuanced view of evolving customer behavior—particularly over an extended post-treatment period—but it also raises challenges for causal interpretation, especially if post-treatment data are used in theme estimation. Balancing the benefits of dynamic modeling with the requirements of causal identification remains an important direction for future research. Finally, while our empirical application is based on an RCT, the framework is broadly applicable to other contexts where firms seek to evaluate the impact of marketing interventions at a granular level. Potential applications include promotions, advertising, and branding, and the framework can be extended to quasi-experimental settings with appropriate identification strategies.

We hope our framework provides a useful tool for both researchers and practitioners seeking to understand the causal impact of marketing interventions in high-dimensional product spaces. By moving beyond predefined product categories to uncover latent, behaviorally meaningful themes, it provides a richer and more actionable perspective on how—and why—customers respond to marketing interventions.

Funding and Competing Interests: All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

References

Anderson ET, Simester D (2011) A step-by-step guide to smart business experiments. *Harvard business review* 89(3):98–105.

Anderson ET, Simester DI (2004) Long-run effects of promotion depth on new versus established customers: three field studies. *Marketing Science* 23(1):4–20.

Angrist JD, Pischke JS (2009) *Mostly harmless econometrics: An empiricist's companion* (Princeton university press).

Ascarza E (2018) Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research* 55(1):80–98.

Athey S, Imbens GW (2017) The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives* 31(2):3–32.

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *Journal of Machine Learning Research* 3(Jan):993–1022.

Boughanmi K, Ansari A (2021) Dynamics of musical success: A machine learning approach for multimedia data fusion. *Journal of Marketing Research* 58(6):1034–1057.

Christensen CM, Hall T, Dillon K, Duncan DS (2016) Know your customers' jobs to be done. *Harvard business review* 94(9):54–62.

Dew R, Padilla N, Luo LE, Oblander S, Ansari A, Boughanmi K, Braun M, Feinberg F, Liu J, Otter T, et al. (2024) Probabilistic machine learning: New frontiers for modeling consumers and their choices. *International Journal of Research in Marketing* .

Ellickson PB, Kar W, Reeder III JC (2023) Estimating marketing component effects: Double machine learning from targeted digital promotions. *Marketing Science* 42(4):704–728.

Feit EM, Berman R (2019) Test & roll: Profit-maximizing a/b tests. *Marketing Science* 38(6):1038–1058.

Goli A, Chintagunta P, Sriram S (2021) Effect of payment on user engagement in online courses. *Journal of Marketing Research* 59(1):11–34.

Goli A, Chintagunta PK (2021) What happens when a retailer drops a product category? investigating the consequences of ending tobacco sales. *Marketing Science* 40(6):1169–1198.

Gopalakrishnan A, Park YH (2021) The impact of coupons on the visit-to-purchase funnel. *Marketing Science* 40(1):48–61.

Gordon BR, Zettelmeyer F, Bhargava N, Chapsky D (2019) A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science* 38(2):193–225.

Griffiths TL (2004) Finding scientific topics. *PNAS* .

Hagen L, Uetake K, Yang N, Bollinger B, Chaney AJ, Dzyabura D, Etkin J, Goldfarb A, Liu L, Sudhir K, et al. (2020) How can machine learning aid behavioral marketing research? *Marketing Letters* 31:361–370.

Heinrich G (2005) Parameter estimation for text analysis. Technical report, Technical report Darmstadt, Germany.

Hutchinson JW, Kamakura WA, Lynch Jr JG (2000) Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research. *Journal of consumer research* 27(3):324–344.

Imbens GW, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences,* (Cambridge University Press).

Iyengar R, Park YH, Yu Q (2022) The impact of subscription programs on customer purchases. *Journal of Marketing Research* 59(6):1101–1119.

Jacobs B, Fok D, Donkers B (2021) Understanding large-scale dynamic purchase behavior. *Marketing Science* 40(5):844–870.

Jacobs BJD, Donkers B, Fok D (2016) Model-based purchase predictions for large assortments. *Marketing Science* 35(3):389–404.

Jaikumar S, Chintagunta PK, Sahay A (2024) Do no harm? unintended consequences of pharmaceutical price regulation in india. *Journal of Marketing* 88(6):1–23.

Kim M, Zhang J (2023) Discovering online shopping preference structures in large and frequently changing store assortments. *Journal of Marketing Research* 60(4):665–686.

Lenk P, Orme B (2009) The value of informative priors in bayesian inference with sparse data. *Journal of Marketing Research* 46(6):832–845.

Li JQ, Rusmevichientong P, Simester D, Tsitsiklis JN, Zoumpoulis SI (2015) The value of field experiments. *Management Science* 61(7):1722–1740.

Li KT, Sonnier GP (2023) Statistical inference for the factor model approach to estimate causal effects in quasi-experimental settings. *Journal of Marketing Research* 60(3):449–472.

Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science* 37(6):930–952.

Lodish LM, Abraham M, Kalmenson S, Livelsberger J, Lubetkin B, Richardson B, Stevens ME (1995) How tv advertising works: A meta-analysis of 389 real world split cable tv advertising experiments. *Journal of marketing research* 32(2):125–139.

Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of machine learning research* 9(Nov):2579–2605.

Seiler S, Tuchman A, Yao S (2021) The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. *Journal of Marketing Research* 58(1):22–49.

Shapiro BT, Hitsch GJ, Tuchman AE (2021) Tv advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica* 89(4):1855–1879.

Tirunillai S, Tellis GJ (2014) Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of marketing research* 51(4):463–479.

Trusov M, Ma L, Jamal Z (2016) Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science* 35(3):405–426.

Turjeman D, Feinberg FM (2024) When the data are out: Measuring behavioral changes following a data breach. *Marketing Science* 43(2):440–461.

Unal M, Park YH (2023) Fewer clicks, more purchases. *Management Science* 69(12):7317–7334.

Urban GL, Hauser JR (2004) "listening in" to find and explore new combinations of customer needs. *Journal of Marketing* 68(2):72–87.

Ye T, Shankar V (2024) Close a store to open a pandora's box? the effects of store closure on sales, omnichannel shopping, and mobile app usage. *Marketing Science* .

Zhong N, Schweidel DA (2020) Capturing changes in social media content: A multiple latent changepoint topic model. *Marketing Science* 39(4):827–846.

Appendix

These materials are supplied by the authors to aid in understanding the paper.

A Full Conditionals for Regression Models

We provide the mathematical details for the full conditionals for the regression models. For the ease of exposition, we consider a generic setting with a target outcome y_i^{target} and customer characteristic vector $\boldsymbol{\chi}_i^{\text{target}}$, such that

$$y_i^{\text{target}} = \boldsymbol{\tau}^{\text{target}\top} \boldsymbol{\chi}_i^{\text{target}} + \varepsilon_i^{\text{target}}, \tag{A.1}$$

with $\varepsilon_i^{\text{target}} \sim \mathcal{N}(0, \sigma^{2\text{target}})$. Note that all the regression models in Equations 1, 5, 13, and 15 are special cases of the model presented in Equation A.1. As shown in Table A.1, the choice of the target outcome y_i^{target} and the target vector $\boldsymbol{\chi}_i^{\text{target}}$ determines whether the regression model operates at an aggregate-level or theme-level, capturing either average or heterogeneous treatment effects, respectively. For example, setting y_i^{target} to y_i and $\boldsymbol{\chi}_i^{\text{target}}$ to $(1, w_i)$, results in the model for an aggregate-level ATE analysis, and setting $\boldsymbol{\chi}_i^{\text{target}}$ to $(1, \tilde{\boldsymbol{x}}_i, w_i, \tilde{\boldsymbol{x}}_i \cdot w_i)$ allows us to conduct an aggregate-level HTE analysis. In the estimation procedure, we update the theme-level outcomes at each iteration ℓ to incorporate uncertainty in the product-theme allocation, setting y_i^{target} to $y_{i,k}$. We also update the unobserved heterogeneity vector at each iteration to $\boldsymbol{d}_i$ to account for its uncertainty.

Table A.1: Aggregate and Theme-level Regression Models

Model	y_i^{target}	$\boldsymbol{\chi}_i^{\text{target}}$	$\boldsymbol{\tau}^{\text{target}}$	$\sigma^{2\text{target}}$
ATE				
Aggregate-level	y_i	$(1, w_i)$	$\boldsymbol{\tau} = (\tau_0, \tau_1)$	σ^2
Theme-level	$y_{i,k}$	$(1, w_i)$	$\boldsymbol{\tau}_k = (\tau_{0,k}, \tau_{0,k})$	σ_k^2
HTE				
Aggregate-level	y_i	$(1, \tilde{\boldsymbol{x}}_i, w_i, \tilde{\boldsymbol{x}}_i \cdot w_i)$	$\boldsymbol{\tau}^h = (\boldsymbol{\tau}_0^h, \boldsymbol{\tau}_1^h)$	σ^{2h}
Theme-level	$y_{i,k}$	$(1, \tilde{\boldsymbol{x}}_i, w_i, \tilde{\boldsymbol{x}}_i \cdot w_i)$	$\boldsymbol{\tau}_k^h = (\boldsymbol{\tau}_{0,k}^h, \boldsymbol{\tau}_{1,k}^h)$	σ_k^{2h}

Now, we provide details of the full conditionals of the regression parameters. For each of the target parameters in Table A.1, we consider the following standard priors:

$$\begin{aligned}\boldsymbol{\tau}^{\text{target}} &\sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \\ \sigma^{2\text{target}} &\sim \text{Inv-Gamma}(a_0, b_0).\end{aligned}$$

Given these priors, and the likelihood of the data, the full conditional distributions of the target parameters are given below.

- Full conditional for $\boldsymbol{\tau}^{\text{target}}$:

$$\boldsymbol{\tau}^{\text{target}} \mid \boldsymbol{y}^{\text{target}}, \boldsymbol{\chi}^{\text{target}}, \sigma^{2\text{target}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{A.2}$$

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^{2\text{target}}} \boldsymbol{\chi}^{\text{target}\top} \boldsymbol{\chi}^{\text{target}} \right)^{-1}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \frac{1}{\sigma^{2\text{target}}} \boldsymbol{\chi}^{\text{target}\top} \boldsymbol{y}^{\text{target}} \right)$$

- Full conditional for $\sigma^{2\text{target}}$:

$$\sigma^{2\text{target}} \mid \boldsymbol{y}^{\text{target}}, \boldsymbol{\chi}^{\text{target}}, \boldsymbol{\tau}^{\text{target}} \sim \text{Inv-Gamma}(a, b) \tag{A.3}$$

$$a = a_0 + \frac{n}{2}, \quad b = b_0 + \frac{1}{2} \left(\boldsymbol{y}^{\text{target}} - \boldsymbol{\chi}^{\text{target}} \boldsymbol{\tau}^{\text{target}} \right)^\top \left(\boldsymbol{y} - \boldsymbol{\chi}^{\text{target}} \boldsymbol{\tau}^{\text{target}} \right)$$

In our empirical application, we consider weakly informative priors by setting $\boldsymbol{\mu}_0$ to a null vectors and $\boldsymbol{\Sigma}_0$ to a diagonal matrix with flat variances of 10^2. Finally, the shape and scale parameters a_0 and b_0 of the inverse-gamma priors are both set to 1.

B Simulations

We use simulations to demonstrate the effectiveness of our proposed approach in identifying latent theme compositions and estimating treatment effects. Importantly, our procedure avoids generating non-existent treatment effects. We present two key simulations. The first is a small-scale simulation that illustrates the recovery of themes and treatment effects, while the second shows that our model is robust to large product assortments, similar to those in our empirical application.

B.1 Small Number of Products

We simulate data for $N = 100$ individuals by sampling distributions over $K = 3$ latent themes from a symmetrical Dirichlet distribution with $\alpha = 0.1$. We also sample theme distributions over $J = 10$ products from a symmetrical Dirichlet distribution with $\beta = 0.01$. For each individual, and for pre-treatment and post-treatment, we sample products based on the individual's latent themes. Outcomes with *null* treatment effects are also sampled. We apply our inference algorithm to recover the theme distributions and treatment effects. We run an MCMC chain for 2,000 iterations, discarding the first 1,000 iterations as burn-in and retaining the latter 1,000 iterations for analysis.

We begin by assessing whether the themes are well recovered. To this end, we report the posterior mean distribution $\hat{\Phi} = (\hat{\boldsymbol{\phi}}_1, \ldots, \hat{\boldsymbol{\phi}}_K)$ for themes. The true theme distributions over products, $\Phi = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)^\top$, are given by the following matrix:

$$\Phi = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.51 & 0.00 & 0.49 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}.$$

The estimated distributions of themes, $\hat{\boldsymbol{\phi}}_k$, are given by:

$$\hat{\Phi} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.55 & 0.00 & 0.44 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}.$$

We observe that the theme distributions over items exhibit perfect recovery, as the estimated values closely match the true values. Figure B.1 presents the MCMC draws of the estimated treatment effects at the theme level. After discarding the first 1,000 iterations (burn-in), the posterior 95% confidence intervals for all treatment effects contain zero. Therefore, we conclude that the treatment effects are recovered with no bias.

B.2 Large Product Assortment

The purpose of this simulation is to mimic the scale of our empirical application, involving $N = 1,000$ customers and $J = 2,000$ products. Each customer makes five purchases pre and post treatment. We also consider outcomes with *null* treatment effects. Figure B.2

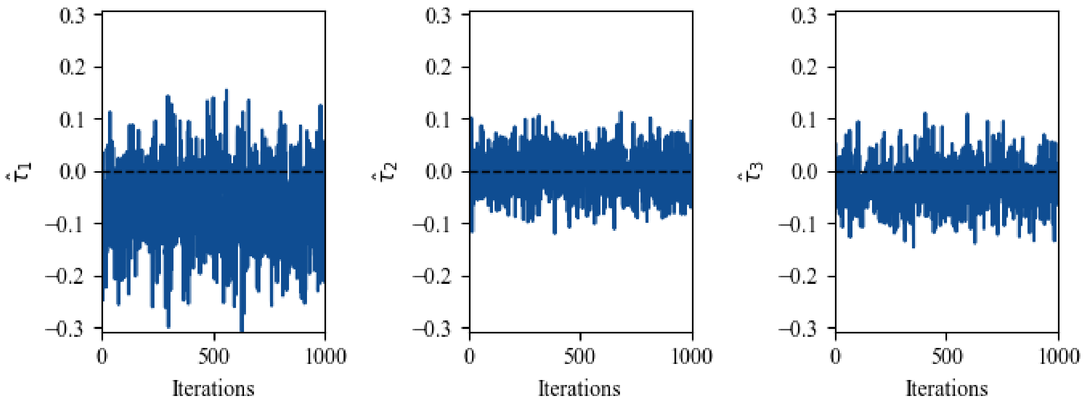

Figure B.1: MCMC Draws of Theme-level Treatment Effects

shows the posterior estimates for the theme-level treatment effects. These estimates are all non-significant, demonstrating that the simulation accurately recovers the original treatment effects with no bias in the causal estimation.

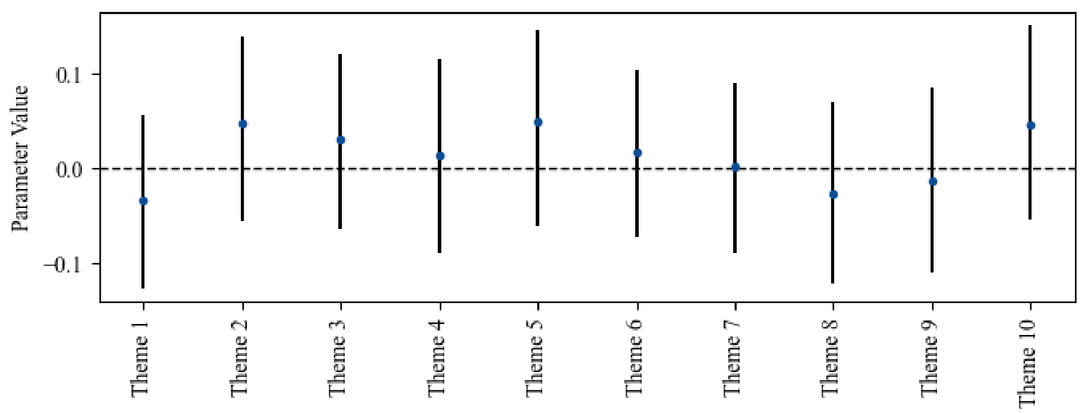

Figure B.2: Posterior Estimates of Theme-level Treatment Effects

C Model Selection and Convergence Checks

We provide details on how we determine the number of latent themes, as well as the convergence checks of the MCMC chains for theme estimation and treatment effect inference.

C.1 Number of Themes

To infer latent purchase patterns in customer behavior before the intervention, we estimate a set of latent themes using pre-treatment transaction data. Because the true number of themes is unknown, we treat the pre-treatment data as a training set and estimate the LDA models across a range of possible theme counts, from 5 to 30.

To evaluate the performance of each theme configuration, we compute perplexity, a standard metric in topic modeling that measures how well the model predicts unseen data. Perplexity is inversely related to the likelihood of observing a product in a customer's purchase history, given the inferred themes and model parameters (Blei et al. 2003). Lower perplexity indicates better predictive fit. For this purpose, we randomly split pre-treatment data into a 90% training set for model estimation and a 10% test set for holdout validation.

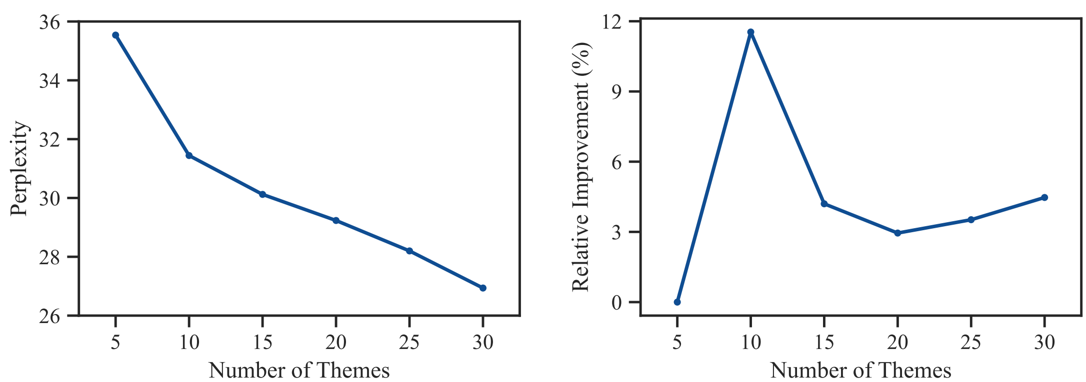

Figure C.1: Model Fit across Theme Counts

Figure C.1 shows the perplexity values for the holdout data across different numbers of themes. As the number of themes increases, model fit improves (left panel), but the relative gains diminish significantly after ten themes (right panel). To balance predictive accuracy and interpretability, especially for managerial applications, we select ten themes as the optimal configuration for subsequent analysis.

C.2 Convergence Checks

We ran Algorithm 1 for 2,000 iterations using ten themes, discarding the first 1,000 draws as burn-in and retaining the remaining draws for analysis. Visual assessment of the MCMC chains indicates convergence and good mixing as we show next.

We used the time series trace plot to check the convergence of the MCMC chains. Figure C.2 shows the trace plots of the chains for 2,000 iterations for the average log-likelihood of the LDA model. The chain is stationary and indicates good mixing and stability of themes discovery.

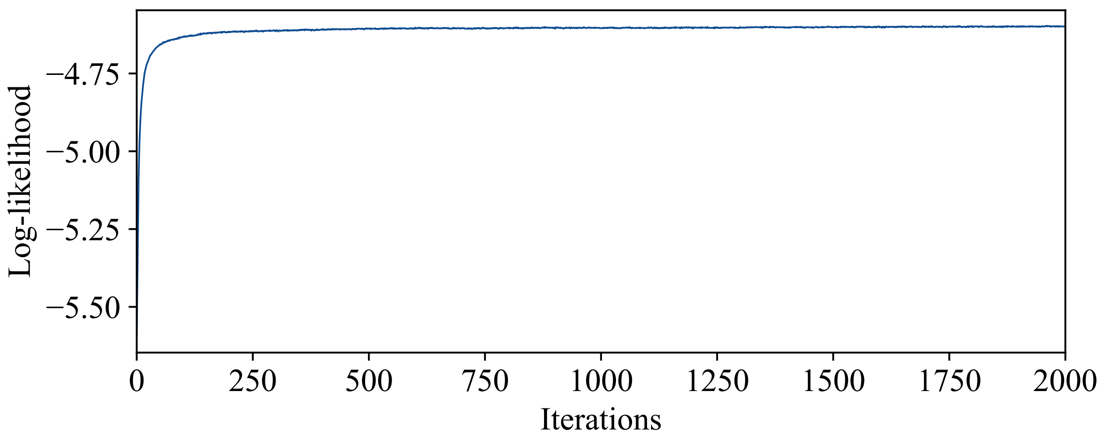

Figure C.2: MCMC Draws for Average Log-Likelihood of the LDA Model

Figure C.3 shows the trace plots of the chains for 2,000 iterations for the theme-level ATEs. The chains appear stationary, indicating good mixing. Initial 1,000 draws were dropped as burn-in sample in the assessment of the treatment effect estimates.

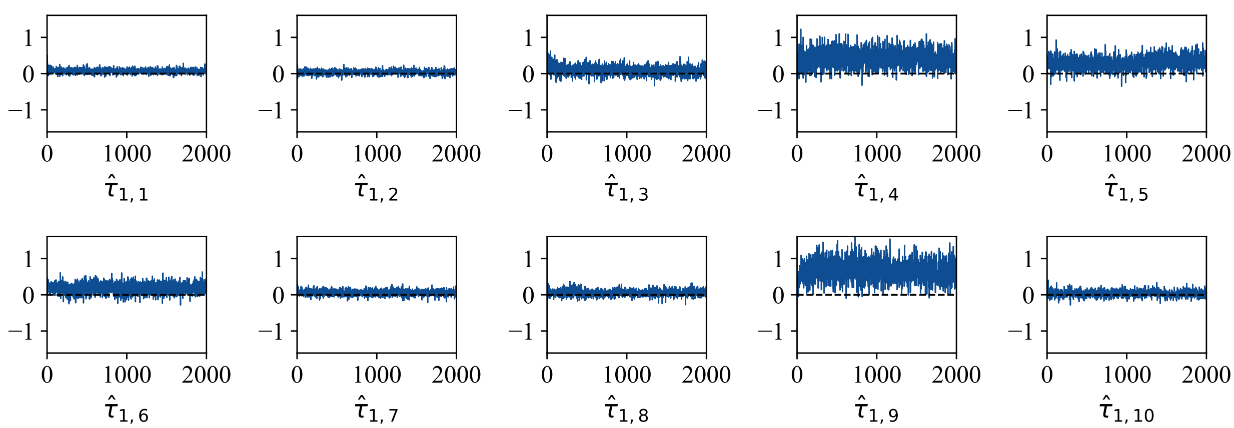

Figure C.3: MCMC Draws for Theme-level ATEs