Improving Targeting Policies by Learning Across Marketing Campaigns

Marat Ibragimov[*] Duncan Simester[†] Artem Timoshenko[‡]

March 26, 2026

The performance of targeting policies depends on the amount of available training data. Firms often have broad archives of past campaigns to augment model training, but differences in timing, design, and audiences complicate transfer learning. We propose an approach that pools information from past campaigns using customer response patterns to capture differences between campaigns. Our approach extends a matrix factorization framework to account for the varying precision of campaign information. We evaluate the approach using 391 direct-mail field experiments conducted by a major apparel retailer. Compared with relying only on focal-campaign data, the proposed approach increases incremental profit by 19−29% and improves customer ranking. We demonstrate that the model effectively learns how to combine information across campaigns from the response patterns, replacing the need for explicit covariates. We investigate boundary conditions on the data requirements. For source campaigns, the model can leverage information from older and seemingly dissimilar past campaigns, and it obtains substantial performance improvements with as few as five source datasets. Data requirements for the focal campaign depend on the targeting objective: ranking customers requires relatively little focal data, whereas optimizing profit requires larger samples. Together, the findings provide a conceptual framework and practical guidance for improving targeting decisions by transferring information across marketing campaigns.

Keywords: personalized targeting, retail, transfer learning, matrix factorization, Bayesian models

[*]Goizueta Business School, Emory University, marat.ibragimov@emory.edu
[†]MIT Sloan School of Management, MIT, simester@mit.edu
[‡]Kellogg School of Management, Northwestern University, artem.timoshenko@northwestern.edu

1. INTRODUCTION

Firms often design targeting policies using experimental data from the prior year's version of the same campaign. This is a sensible starting point because the prior-year campaign is often a close analogue to the current context. However, this approach is inherently limited by the scale of the prior-year experiment. This highlights a tension in targeted marketing: performance improves when training data closely resembles the focal campaign, but using only the most relevant data source limits the quality of targeting decisions. To illustrate, consider a retailer deciding which customers should receive its 2026 Holiday catalog. If the firm conducted experiments for the 2025 Holiday mailing, it can use that data to train the 2026 targeting policy. However, the resulting policy will be constrained by the scale of the 2025 experiment.

A natural response is to augment the focal training data with information from other campaigns. For the 2026 Holiday catalog, this might include prior Valentine's Day promotions, Mother's Day mailings, back-to-school catalogs, and earlier Holiday campaigns. These campaigns could all contain useful information about customer responsiveness, but they are not identical to the focal campaign. They differ in timing, product assortment, promotional offers, creative execution, and eligible audiences. Critically, these differences are often poorly documented, preventing firms from simply pooling experiments or matching campaigns based on recorded characteristics alone. This paper addresses the question: How can firms leverage a broad library of past marketing campaigns to improve targeting for a focal campaign when candidate data sources are related but not identical?

We propose an approach that improves targeting policies by pooling information across past campaigns, even when those campaigns differ in design, timing, or product focus. The central idea is to use customer responsiveness as the primary signal for determining how information should transfer across campaigns. Campaigns need not be exact matches to the focal campaign to be informative; instead, they are useful when they reveal similar patterns

in how customers respond to marketing actions. The approach learns this common structure and uses it to improve predictions for the focal campaign.

To implement this idea, we extend the probabilistic matrix factorization framework. The model takes as inputs a set of treatment effect estimates for different campaigns and customer segments. Because these estimates vary in precision (for example due to sample size), the model explicitly accounts for this variation. It then combines treatment effect measures with available customer and campaign covariates to infer latent embeddings. This allows the model to pool information across campaigns, placing greater weight on more precise information.

We evaluate the approach using 391 direct-mail field experiments conducted by a major apparel retailer over a ten-year period. Our model substantially outperforms benchmark methods that rely only on focal-campaign data, including individual-level machine-learning methods. The gains are economically meaningful. Compared with the strongest benchmarks, the proposed approach increases incremental profit by $19-29\%$, while also improving the accuracy of customer rankings by responsiveness.

The targeting improvements arise primarily because the model captures variation in customer responsiveness and uses this information to pool information across campaigns. Remarkably, a specification that entirely omits observed customer and campaign covariates performs nearly as well as the full model. This suggests that much of the relevant structure can be recovered directly from observed response patterns, reducing the need for meticulously coded campaign archives. It makes the approach robust to incomplete documentation, which is common in industry applications.

Firms can improve targeting by borrowing information from a much broader set of past campaigns than managers might initially expect. While more recent campaigns are more informative, older campaigns still provide substantial incremental value, even when conducted more than two years before the focal campaign. Likewise, aligning source and focal campaigns by season or campaign type provides only modest performance benefits compared to

randomly selecting source campaigns. In practical terms, a retailer choosing targets for a Holiday catalog may still benefit from information contained in Valentine's Day, back-to-school, or other campaigns, even when they appear dissimilar.

We further investigate the data requirements for transfer learning. For the focal campaign, these requirements depend on the targeting objective: ranking customers by responsiveness is feasible with relatively little focal data, whereas optimizing profit requires larger focal samples to calibrate the magnitude of the predictions. For the source campaigns, campaign size is more important than the total number of campaigns. In our empirical setting, the proposed approach recovers most of the available gains with only five large source campaigns. In contrast, targeting performance is lower when incorporating information from many small source campaigns. In other words, transferring information from fewer, well-powered source experiments outperforms aggregating evidence from many underpowered ones.

The paper makes conceptual, methodological and empirical contributions. Conceptually, we formulate the problem of learning across marketing experiments in a matrix factorization framework. We use customer responsiveness as the primary signal for pooling information across campaigns, allowing firms to improve targeting policies even when metadata are incomplete or unreliable. Methodologically, we extend Bayesian probabilistic matrix factorization to account for variation in the precision of model inputs, recognizing that marketing experiments often substantially vary in scale and statistical precision. Empirically, we validate the approach using an extensive archive of field experiments, establish when information transfer is most valuable, and quantify the data requirements. Together, these contributions provide both a conceptual framework and actionable guidance for leveraging historical experiments to improve targeting decisions.

The paper continues in Section 2 with a review of the related literature. In Section 3 we formally state the business problem and introduce our model. Section 4 describes the empirical application and data, while Section 5 presents the main results by comparing our

proposed method against relevant benchmarks. Section 6 investigates the method's boundary conditions, identifying when information transfer is most valuable and quantifying the data requirements. Section 7 concludes with a summary of the findings and opportunities for future research.

2. RELATED LITERATURE

Our findings contribute to a growing literature in marketing focusing on the training of targeting (uplift) policies. Recent research has studied: the design and evaluation of targeting methods (Simester, Timoshenko, and Zoumpoulis 2020; Hitsch, Misra, and Zhang 2024), welfare implications of personalized pricing (Dubé and Misra 2023), explainable targeting policies and unintended bias (Ascarza and Israeli 2022; Zhang 2024). These studies primarily focus on identifying customers where treatment effects offset the cost of a promotion.

Estimating treatment effects contrasts with traditional customer lifetime value (CLV) approaches, which quantify expected customer value (Fader, Hardie, and Lee 2005; Gupta et al. 2006). Comprehensive reviews of the CLV research are provided by Kumar and Reinartz (2018), Ascarza et al. (2018), and Venkatesan, Farris, and Wilcox (2021). Recently, Lemmens and Gupta (2020) bridged these perspectives by introducing a profit-based framework that models the impact of marketing retention campaigns on CLV. While we also consider customer profit as our outcome, our study primarily contributes to uplift modeling by addressing the data-scarcity challenges inherent in training these policies.

The business setting in our paper is similar to the problem in Simester, Timoshenko, and Zoumpoulis (2025), who study the size of the focal experiment required to train and certify a targeting policy for direct mail campaigns. Like our study, their approach assumes an existing customer segmentation and models policy learning within the Bayesian framework. However, unlike our paper, they do not use information from multiple source campaigns. We propose and empirically validate a method to supplement the focal data with information from past experiments to improve the performance of targeting policies.

In concurrent research, Ellickson, Kar, and Reeder III (2023) and Huang, Ascarza, and Israeli (2024) also propose using data from source campaigns to target a focal campaign. Ellickson, Kar, and Reeder III (2023) study email promotions and decompose the impact of different keywords and different types of promotional offers. Huang, Ascarza, and Israeli (2024) model the incremental effects of coupons using observable characteristics of consumer packaged goods (CPG) brands. There are two important differences between their studies and our study. First, they use observable design features of marketing campaigns, and identify similarities between campaigns from these features. In contrast, our approach uses the variation in treatment effects across customer segments and campaigns as a primary signal to identify campaign similarities. Our method can incorporate observable covariates, but does not require these covariates, and so does not depend upon accurate documentation of customer and campaign design characteristics. On the other hand, we require that a sample of training data is available for the focal campaign, which their approaches do not need.

The second major distinction is that we account for uncertainty in the treatment effect estimates. When transferring information across campaigns, our approach recognizes that treatment effects in different campaigns capture customer responsiveness to potentially different marketing actions (design characteristics), and this responsiveness is measured with different precision. For example, the precision can depend on the size of the focal and source campaigns. We show that accounting for the varying precision across campaigns improves the accuracy of treatment effect predictions for the focal campaign. Our method extends the probabilistic matrix factorization framework (PMF; Salakhutdinov and Mnih 2008) by explicitly accounting for the precision of the treatment effect measures used as inputs. This extension is important to our application, and requires a change in both the definition and estimation of the PMF model.

Within the PMF literature, our proposed extension to the PMF model can be compared with two papers. Lakshminarayanan, Bouchard, and Archambeau (2011) introduce Robust

Bayesian Matrix Factorization to handle outliers and atypical customer behavior in the Netflix problem (predicting movie ratings for Netflix users). Their paper models heteroskedastic noise across user:movie observations by multiplying a global precision parameter by user-specific and movie-specific scaling factors. The scaling factors are estimated together with the user and movie embeddings. In contrast, Yang (2017) incorporate heteroskedasticity information as an input to the PMF framework (instead of estimating it). The paper aims to predict the conversion rates of digital ads, and to mitigate the sparsity in responses. It defines the precision as a function of the average conversion rates and the number of impressions. Similar to these two papers, our proposed extension focuses on heteroskedastic noise. We decompose the variation in observed treatment effects into two components: random deviation from the embedding structure and imprecision in the measurement. We assume that the first component is homoskedastic, and estimate it along with customer and campaign embeddings. The second component varies across observations. We assume that uncertainties associated with the treatment effect measurements are known and provide them as inputs to the model.

Obtaining sufficient data to train targeting policies can be a substantial obstacle to both the feasibility and performance of policies designed to target marketing actions. Industry and marketing academics have started to recognize that a firm's library of past campaigns could provide a rich source of training data. However, this introduces a new challenge: understanding how to summarize this past information and combine it with information about the current campaign. We propose a solution to this problem and, using a large set of field experiments, show that the solution can greatly improve targeting performance.

3. PROPOSED METHOD: BAYESIAN MATRIX FACTORIZATION

Our approach extends a matrix factorization framework to estimate treatment effects across marketing campaigns and customer segments. We parameterize the treatment effects using campaign and segment embeddings, and infer these embeddings from the available covariates

and (noisy) treatment effect measures. Our extension accounts for the precision of these measures.

3.1. Model Setup

Consider a firm that operates in a market with S customer segments and distributes C marketing campaigns to these customers. We index segments by $s \in \{1, \ldots, S\}$ and marketing campaigns by $c \in \{1, \ldots, C\}$. The effectiveness of marketing campaigns varies across segments, and we assume that the customer segmentation is known and is not updated during model training. We denote a conditional average treatment effect (CATE) of campaign c in segment s as τ_{cs}.

The firm's goal is to predict CATEs for each customer segment that is eligible for the focal campaign. To design a targeting policy, the firm can use the predicted CATEs to identify segments for which the expected response exceeds the cost of the marketing action. In our empirical application, we also consider ranking customer segments based upon their predicted responsiveness to the focal campaign.

For the focal and source campaigns, the firm has information about the CATEs in different segments. These measures could be obtained from pilot experiments, experiments conducted the previous year, natural experiments, or observational methods. The measures are only available in the customer segments that were eligible for a given campaign; $q_{cs} \in \{0, 1\}$ indicates whether segment s was eligible for campaign c. Our method aims to improve (update) these measures by pooling information across campaigns and customer segments. We denote the observed treatment effect measure for campaign c in segment s as y_{cs}, assuming a known precision λ_{cs} and a zero-mean Gaussian error distribution:

$$\mathbb{P}(y_{cs}|\tau_{cs}, \lambda_{cs}^{-1}) = \mathcal{N}(y_{cs}|\tau_{cs}, \lambda_{cs}^{-1}) \tag{1}$$

The notation $\mathcal{N}(x|\mu, \lambda^{-1})$ indicates a Gaussian random variable x with mean μ and variance λ^{-1}, where λ is a precision parameter.

For every customer segment s, the firm observes a vector of features $\mathbf{x}_s \in \mathbb{R}^{D_x \times 1}$ that reflect observable differences between segments, such as demographics, typical spending levels, and store visit frequency. Similarly, for each campaign c, the firm observes a vector of features $\mathbf{z}_c \in \mathbb{R}^{D_z \times 1}$ that characterize observable differences between campaigns, such as featured product categories, seasonality, and recency.

We denote three $C \times S$ matrices for the treatment effect estimates $\mathbf{Y} = [y_{cs}]_{C \times S}$, the precision of these estimates $\mathbf{\Lambda}_{obs} = [\lambda_{cs}]_{C \times S}$, and for the measurement availability $\mathbf{Q} = [q_{cs}]_{C \times S}$. Likewise, two matrices summarize all available segment and campaign characteristics: $\mathbf{X} = [\mathbf{x}_s']_{S \times D_x}$ and $\mathbf{Z} = [\mathbf{z}_c']_{C \times D_z}$ respectively. The available data includes $\mathcal{D} = (\mathbf{Y}, \mathbf{\Lambda}_{obs}, \mathbf{Q}, \mathbf{X}, \mathbf{Z})$. Our proposed approach attempts to infer unobserved information about true treatment effects τ_{cs} using the observable data $\mathcal{D}$. A complete table of notations is provided in Appendix A.

3.2. Model Structure

We represent the mean of true conditional treatment effect τ_{cs} as a dot product of segment embeddings $\mathbf{u}_s \in \mathbb{R}^{K \times 1}$ and campaign embeddings $\mathbf{v}_c \in \mathbb{R}^{K \times 1}$, and we capture the deviations from the mean by α:

$$\mathbb{P}\left(\tau_{cs} | \mathbf{u}_s, \mathbf{v}_c, \alpha\right) = \mathcal{N}\left(\tau_{cs} | \mathbf{u}_s' \mathbf{v}_c, \alpha^{-1}\right) \tag{2}$$

where $\mathbf{u}_s' \mathbf{v}_c$ and α^{-1} indicate the mean and variance of the normal distribution.

The segment and campaign embeddings are informed by observable characteristics ($\mathbf{X}$ and $\mathbf{Z}$). Specifically, we assume the following distributions for the embeddings:

$$\begin{aligned} \mathbb{P}(\mathbf{U} | \mathbf{B}_u, \mathbf{X}, \mathbf{\Sigma}_u) &= \prod_s \mathcal{N}(\mathbf{u}_s | \mathbf{B}_u' \mathbf{x}_s, \mathbf{\Sigma}_u) \\ \mathbb{P}(\mathbf{V} | \mathbf{B}_v, \mathbf{Z}, \mathbf{\Sigma}_v) &= \prod_c \mathcal{N}(\mathbf{v}_c | \mathbf{B}_v' \mathbf{z}_c, \mathbf{\Sigma}_v) \end{aligned} \tag{3}$$

where $\mathbf{B}_u \in \mathbb{R}^{D_x \times K}$ and $\mathbf{B}_v \in \mathbb{R}^{D_z \times K}$ are matrices of random coefficients. This linear specification allows the observable characteristics to systematically shape the latent (embedding)

space. For example, if treatment effects vary substantially across different seasons, the model can learn to place the embeddings for campaigns in different seasons further apart by adjusting the corresponding coefficients.

Finally, we place matrix normal priors on the weights of customer and segment embeddings:

$$\begin{aligned}
\mathbb{P}(\mathbf{B}_{\mathrm{u}}, \mathbf{\Sigma}_{\mathrm{u}}) &= \mathbb{P}(\mathbf{B}_{\mathrm{u}}|\mathbf{\Sigma}_{\mathrm{u}})\mathbb{P}(\mathbf{\Sigma}_{\mathrm{u}}) = \mathcal{MN}\left(\mathbf{B}_{\mathrm{u}}|\mathbf{M}_{\mathrm{u},0}, \mathbf{W}_{\mathrm{u},0}^{-1}, \mathbf{\Sigma}_{\mathrm{u}}\right) \mathcal{W}^{-1}(\mathbf{\Sigma}_{\mathrm{u}}|\mathbf{\Lambda_0}, \nu_0) \\
\mathbb{P}(\mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}}) &= \mathbb{P}(\mathbf{B}_{\mathrm{v}}|\mathbf{\Sigma}_{\mathrm{v}})\mathbb{P}(\mathbf{\Sigma}_{\mathrm{v}}) = \mathcal{MN}\left(\mathbf{B}_{\mathrm{v}}|\mathbf{W}_{\mathrm{v},0}, \mathbf{W}_{\mathrm{v},0}^{-1}, \mathbf{\Sigma}_{\mathrm{v}}\right) \mathcal{W}^{-1}(\mathbf{\Sigma}_{\mathrm{v}}|\mathbf{\Lambda_0}, \nu_0) \\
\mathbb{P}(\alpha) &= \mathcal{G}(\alpha|\mathrm{a}_0, \mathrm{b}_0)
\end{aligned} \tag{4}$$

where $\mathcal{MN}$ is a matrix normal distribution with mean matrix $\mathbf{M}_{\mathrm{u},0} \in \mathbb{R}^{\mathrm{D_x} \times \mathrm{K}}$ and positive-definite scale matrices $\mathbf{W}_{\mathrm{u},0} \in \mathbb{R}^{\mathrm{D_x} \times \mathrm{D_x}}$ and $\mathbf{\Sigma}_{\mathrm{u}} \in \mathbb{R}^{\mathrm{K} \times \mathrm{K}}$; $\mathcal{W}^{-1}$ is the inverse-Wishart distribution with ν_0 degrees of freedom and positive definite scale matrix $\mathbf{\Lambda_0} \in \mathbb{R}^{\mathrm{K} \times \mathrm{K}}$; $\mathcal{G}$ is a Gamma distribution with shape a_0 and rate b_0.

Figure 1 illustrates our model using plate notations.

3.3. Inference

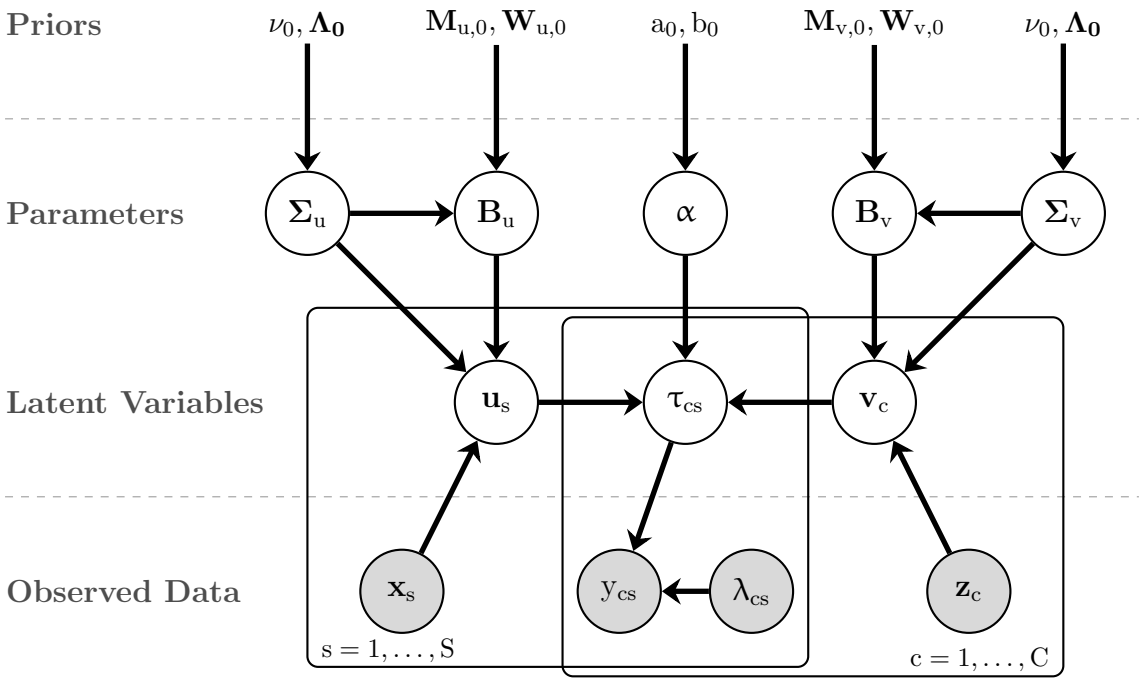
To estimate the model, we derive a closed form expression for $\mathbb{P}(\tau_{\mathrm{cs}}|\mathbf{U}, \mathbf{V}, \alpha, \mathcal{D})$:

$$\mathbb{P}(\tau_{\mathrm{cs}}|\mathbf{U}, \mathbf{V}, \alpha, \mathcal{D}) = \begin{cases} \mathcal{N}\left(\tau_{\mathrm{cs}}|\mathbf{u}_{\mathrm{s}}'\mathbf{v}_{\mathrm{c}}, \alpha^{-1}\right) & \text{if } \mathrm{q_{cs}} = 0 \\ \mathcal{N}\left(\tau_{\mathrm{cs}}\middle|\frac{\mathbf{u}_{\mathrm{s}}'\mathbf{v}_{\mathrm{c}}\alpha + \mathrm{y_{cs}}\lambda_{\mathrm{cs}}}{\lambda_{\mathrm{cs}}+\alpha}, \frac{1}{\lambda_{\mathrm{cs}}+\alpha}\right) & \text{if } \mathrm{q_{cs}} = 1 \end{cases} \tag{5}$$

We then rely on a Markov Chain Monte Carlo (MCMC) approach to draw samples from the joint distribution $\mathbb{P}(\mathbf{U}, \mathbf{V}, \alpha, \mathbf{B}_{\mathrm{u}}, \mathbf{\Sigma}_{\mathrm{u}}, \mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}}|\mathcal{D})$, and approximate the posterior predictive distribution of treatment effects $\mathbb{P}(\tau_{\mathrm{cs}}|\mathcal{D})$ using an average over these samples:

$$\mathbb{P}(\tau_{\mathrm{cs}}|\mathcal{D}) \simeq \frac{1}{\mathrm{T}} \sum_{\mathrm{t}} \mathbb{P}(\tau_{\mathrm{cs}}|\mathbf{U}^{\mathrm{t}}, \mathbf{V}^{\mathrm{t}}, \alpha^{\mathrm{t}}, \mathcal{D}) \tag{6}$$

Figure 1: Graphical Representation of the Model



Notes: The figure summarizes the design of the model using plate notation.

Our choice of conjugate priors allows us to derive analytical, closed-form expressions for the conditional distributions of the embeddings and their weights (for covariates). This ensures that the model remains computationally efficient and highly scalable across hundreds of campaigns. The only exception is the global precision parameter α, which lacks a closed-form posterior. To sample α, we incorporate a Hamiltonian Monte Carlo (HMC) step that utilizes the derivative of the log-likelihood function to achieve faster, more robust convergence. We provide a complete MCMC estimation algorithm in Appendix B, and mathematical derivations of the conditional distributions in the Web Appendix.

Our model approximates $\mathbb{P}(\tau_{cs}|\mathcal{D})$ with a Gaussian mixture distribution with T components with equal weights, where each component follows the normal distribution specified in

Equation (5). We use the mean of this Gaussian mixture as a point prediction:

$$\widehat{\tau}_{cs} = \begin{cases} \frac{1}{T}\sum_t \mathbf{u}_s^{t\prime}\mathbf{v}_c^t & \text{if } q_{cs} = 0 \\ \frac{1}{T}\sum_t \left(\lambda_{cs} + \alpha^t\right)^{-1} \cdot \left(\alpha^t \mathbf{u}_s^{t\prime}\mathbf{v}_c^t + \lambda_{cs} y_{cs}\right) & \text{if } q_{cs} = 1 \end{cases} \tag{7}$$

The estimator in Equation (7) has an intuitive interpretation. If measurement y_{cs} is not available for a segment-campaign, the model uses $\frac{1}{T}\sum_t \mathbf{u}_s^{t\prime}\mathbf{v}_c^t$ for prediction. If measurement y_{cs} is available ($q_{cs} = 1$), the model combines y_{cs} and $\mathbf{u}_s^{t\prime}\mathbf{v}_c^t$ weighted by the precision of each component, and then averages these predictions across T samples. If the measurement in the campaign-segment is very precise ($\lambda_{cs} \to +\infty$), the prediction converges to the measurement: $\widehat{\tau}_{cs} \to y_{cs}$; and vice versa, $\lim_{\lambda_{cs}\to 0} \widehat{\tau}_{cs} = \frac{1}{T}\sum_t \mathbf{u}_s^{t\prime}\mathbf{v}_c^t$.

3.4. Discussion

It may appear that the dot product specification assumes a specific relationship between customer and campaign characteristics. However, the parametrization of the treatment effects with a dot product of two K-dimensional dense vectors (embeddings) provides enough flexibility to approximate non-linear patterns in treatment effects across segments and campaigns.[1] Furthermore, the matrix factorization framework allows for efficient inference; when using the full sample in our empirical application we obtain stable posterior estimates in approximately seven seconds.

When implementing the proposed method, cross-validation can be used to help choose hyperparameters, including the dimensionality of the embeddings (K). In our empirical application, we select K = 3. The dimensionality of the embeddings balances a tradeoff: fewer dimensions restrict model flexibility, whereas higher dimensions increase the computational burden without yielding out-of-sample performance gains. We can also use cross-validation

[1] We used simulations to confirm that the dot product specification can closely approximate treatment effects even when the data generating process includes different types of non-linear relationships with the customer and campaign embeddings.

to assess whether additional information from past marketing campaigns helps to improve out-of-sample targeting performance, and whether the model consistently outperforms alternative methods.

In Related Literature, we acknowledged that our approach builds on the PMF formulation (Salakhutdinov and Mnih 2008). When the PMF model was initially proposed, the 'Netflix Problem' was used as motivation. The goal is to predict Netflix movie ratings for a new user : movie combination, based on the ratings by other users, and ratings by that user for other movies. There are two important differences between this problem and our problem. First, the training data in our problem represents noisy measures of the treatment effects, instead of the "true" movie ratings. To incorporate the associated uncertainty, we extended the PMF framework by introducing the λ_{cs} terms, and consequently adjusting the inference process (for α). Second, in the Netflix problem the goal is to predict movie ratings where no rating exists. In our setting, the focal data contains noisy measures of the treatment effects, which changes the objective from value imputation to value updating.

4. DATA OVERVIEW

Our empirical application uses data provided by a major US apparel retailer. The retailer sells through both direct and physical retail channels. The assortment spans men's and women's apparel, footwear, and home goods.

The retailer regularly distributes direct mail campaigns to its existing customers to promote new products and encourage customer spending. The focus of these campaigns varies and can include specific audiences, product categories, seasons, or special events. Some campaigns occur only once, while others recur annually at approximately the same time, such as the winter Holiday catalog.

To measure the overall impact of direct mail campaigns on profits (ATEs), the retailer routinely conducts field experiments in which qualified households are randomly assigned to different experimental conditions. The campaigns are categorized into two distinct ex-

perimental designs: incrementality campaigns that evaluate the effect of "mail" versus "no mail," and promotional campaigns that compare "mail + coupon" against "mail only."[2] For ease of exposition, in the incrementality campaigns, we will use the label "Treatment" to refer to the "mail" condition and "Control" to refer to the "no mail" condition. In the promotional campaigns, we will use the term "Treatment" to refer to the 'mail + coupon" condition and "Control" to refer to the "mail only" condition.

Our dataset consists of 391 marketing campaigns conducted between 2013 and 2023. Out of these 391 campaigns, 198 have a "pair": a similar marketing campaign conducted in the subsequent year.[3] For each campaign, our data include the internal name (e.g. "Spring Womens" campaign), the represented product categories (the "product group"), the in-home date, the circulation files identifying households randomly assigned to each condition, and whether the campaign included a discount code.

Figure 2 documents the variation in campaign characteristics across our sample. Approximately 68% of the campaigns are incrementality campaigns. Campaigns are distributed throughout the year, with higher frequency in the fourth quarter, coinciding with the Holiday season. Campaign size also varies significantly, ranging from 50,000 to 1,000,000 households. This sample size variation directly impacts the precision of segment-level treatment effect estimates for each campaign.

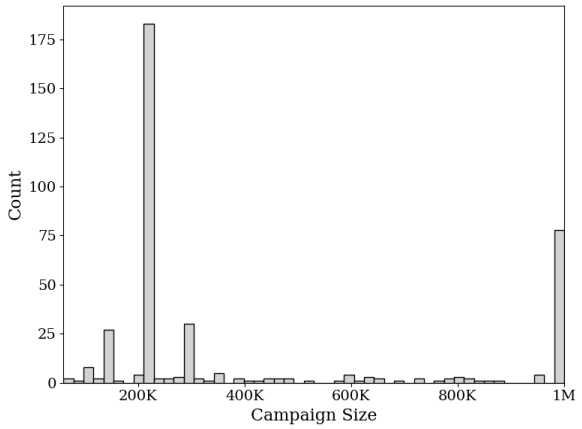
We performed a series of randomization checks to verify the integrity of the retailer's field experiments. These diagnostics ensure that treatment assignments are orthogonal to pre-existing customer characteristics, supporting the internal validity of the estimated treatment effects. The experiments pass the randomization checks. A detailed description is provided in the Web Appendix.

We define customer segments based on individual customer spending during the three years preceding each campaign's in-home date. The population is partitioned into 50 ap-

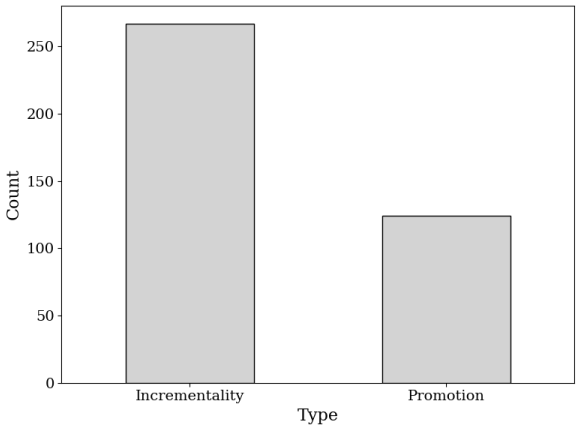
[2]The coupons include a single use promotional code (SUPC) redeemable for a discount on the customer's order.

[3]We manually defined a "pair" as two campaigns for which in-home dates fall within a two-week window in consecutive years, and the internal name and target product group are the same in both years.

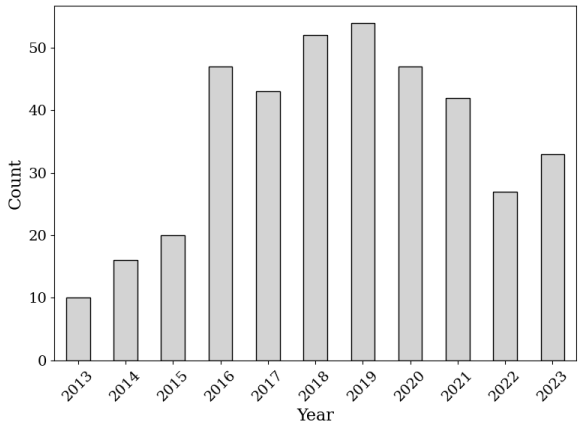
Figure 2: Summary Statistics of Direct Mail Campaigns



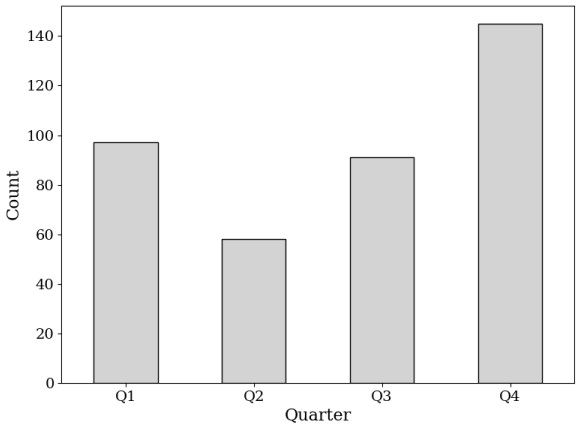
(a) Distribution of Campaign Sizes



(b) Campaigns by Type



(c) Campaigns by Year



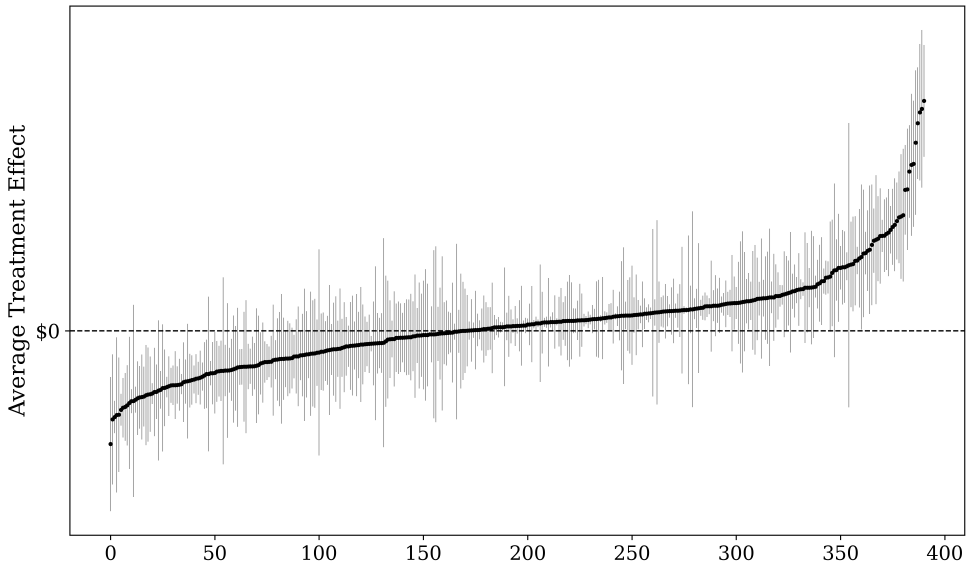
(d) Campaigns by Calendar Quarter

Notes: The figures report histograms summarizing the distribution of campaign characteristics. In each panel the unit of observation is a single campaign and the sample size is 391 (campaigns).

proximately equal segments, applying a rounding constraint to the nearest $5 increment. To mitigate the influence of outliers, we excluded customers with three-year spending exceeding $10,000, who represent less than 0.05% of the total observations. In our empirical setting, this segmentation captures sufficient variation to perform nearly as well as individual targeting.

The primary outcome measure is total profit realized within four weeks following the campaign's in-home date. We construct this variable using the retailer's transaction data, which identifies individual customers across both online and offline channels. A mailing cost is included in every experimental condition except the "no mail" condition.

Figure 3: Average Treatment Effect Across All Marketing Campaigns



Notes: The figure reports the ATE of the 391 marketing campaigns included in the study. The campaigns are sorted by the magnitude of the ATE. Each dot represents the estimated ATE of a single campaign (measured in dollars per customer). The y-axis labels are omitted to protect confidentiality. The error bars indicate 95% confidence intervals.

Figure 3 shows the distribution of Average Treatment Effects (ATE) across all marketing campaigns, sorted by magnitude. While on average campaigns generate positive incremental profit (the average ATE is 6.14¢ per customer), there is substantial variation in the effectiveness of the retailer's marketing actions across campaigns.

5. EMPIRICAL PERFORMANCE

We investigate whether the proposed model effectively leverages information from past marketing campaigns to improve targeting outcomes. We first describe how we construct training and validation datasets for model evaluation, and then introduce performance metrics and benchmark models. We present the findings in two steps: a performance comparison of the proposed and benchmark methods, and an ablation study decomposing the performance improvements by different components of the proposed approach.

5.1. Evaluation Approach

We use two approaches to construct training and testing data. In the **Next Year** approach, we restrict attention to campaigns that have a "pair" and train all models using the full focal campaign data from year $t-1$. We then evaluate the performance of the trained policy using data from the associated campaign in year t. Using the prior year's campaign to train the current year's policy is a common practical application, which enhances the managerial relevance of this evaluation approach. However, potential non-stationarity between years due to changes in firm actions and consumer behavior can make this evaluation approach challenging.

In the **Same Period** approach, we randomly split the focal campaign data into two subsamples. We use 70% of the data to train all models and the remaining 30% as a holdout sample to evaluate performance. This provides an empirical setting where validation is unaffected by non-stationarity, focusing the evaluation on model performance. However, the practical equivalent of this evaluation approach assumes the retailer can conduct a pilot study before a full-scale rollout, which may not always be feasible.

Both evaluation approaches maintain a strict separation between training and testing data. Under each method, we reserve the first 30 campaigns as source campaigns and then iterate the evaluation process over the (remaining) focal campaigns in our sample and report

the average performance across these iterations.

We report two performance metrics commonly used in the targeting literature. The Qini metric evaluates the models' ability to rank customers by their predicted treatment effects (Yadlowsky et al. 2025). Higher Qini indicates a policy that prioritizes customers more accurately. We also report the expected profit per customer ("*Profit*"). We estimate the expected profit using the Horvitz-Thompson estimator. Because the estimation relies on the experimental data, the resulting profit metric is qualitatively equivalent to evaluating the targeting policies in a field experiment Hitsch, Misra, and Zhang (2024). We index this profit measure by subtracting the profit earned in the "Control" condition.

5.2. Main Result

We initially compare the proposed method against four benchmark models. *Blanket Treatment* is a baseline policy in which the treatment is uniformly assigned to everyone in each segment. *Focal Only* is a model-free benchmark that uses just the focal training data to choose treatments separately in each segment. Specifically, within each segment we use the focal training data to separately calculate the average profit earned from each treatment, and then choose the treatment with the largest average performance. *Best ML Baseline* is the top-performing model among Lasso, XGBoost, and Causal Forest, selected by maximizing the respective metric (Qini or Profit) using 5-fold cross-validation on the focal training data. For these machine learning benchmarks, we use individual-level targeting based on the individual past purchase histories. The complete list of covariates is provided in the Web Appendix.

Table 1 reports the performance of each method. In this analysis, we only consider focal campaigns that have at least 30 (source) campaigns with earlier in-home dates. The proposed *Transfer* model outperforms all benchmarks across both evaluation frameworks and both metrics. The improvement over the *Focal Only* and *Best ML Baseline* confirm that information extracted from past marketing campaigns can be valuable, and can improve

targeting policies compared to just using data for the focal campaign. It also confirms that the proposed model is effective at both summarizing information from the past campaigns and transferring this information to the focal campaign.

Table 1: Performance of Proposed Model and Baseline Methods

	Same Period		**Next Year**	
Method	**Qini**	**Profit**	**Qini**	**Profit**
Blanket Treatment	—	5.07¢	—	6.29¢
		(2.24)		(2.83)
Focal Only	4.81	16.18¢	4.34	15.25¢
	(0.57)	(1.80)	(0.55)	(1.94)
Best ML Baseline (Profit)	6.05	18.76¢	4.06	15.68¢
	(0.66)	(1.81)	(0.51)	(2.06)
Best ML Baseline (QINI)	6.55	17.75¢	4.20	14.69¢
	(0.69)	(1.77)	(0.56)	(1.94)
Transfer (Proposed)	10.12	22.30¢	9.15	20.29¢
	(0.63)	(1.96)	(0.81)	(2.44)

Notes: The table compares the performance of the proposed model and benchmark methods using the *Qini* and *Profit* metrics. The **Same Period** comparison compares findings across 361 campaigns, and the **Next Year** comparison includes 198 campaigns. Standard errors are reported in parentheses. All performance differences between the methods are statistically significant with $p < 0.01$ in paired sample t-tests.

There are several additional findings of interest. First, the positive *Profit* metric for the *Blanket Treatment* indicates that uniformly treating each household outperforms the uniform control policy. This confirms that the retailer's marketing interventions are profitable on average in the considered sample, which is consistent with the positive overall average ATE reported for the campaigns in Figure 3.

Second, the *Focal Only* model outperforms the *Blanket Treatment*, highlighting the presence of significant treatment effect heterogeneity across segments (Shchetkina and Berman 2024). However, the *Best ML Baselines* do not substantially improve over the *Focal Only* policy. Recall that the ML models all exploit individual variation in treatment effects. The

comparable performance of the *Focal Only* model with these benchmarks suggests that the proposed segmentation effectively summarizes much of the individual variation explainable by the purchase histories.

The performance differences between the *Focal Only*, *Best ML Baseline* and *Transfer* policies are consistent across the **Same Period** and the **Next Year** approaches. This suggests that in our empirical setting, customer responsiveness to marketing campaigns does not substantially vary year-over-year, justifying the current business practice of using past-year experiments as training data and highlighting that our approach is robust to modest non-stationarity.

5.3. Decomposing the Source of the Performance Improvements

To understand the drivers of the performance improvement, in Table 2 we perform an ablation study by systematically removing different components of the proposed model and observing the resulting impact on performance. In particular, we compare the performance of the proposed model with the following additional benchmarks:

- *No Covariates* is the proposed model where we omit covariates. We separately remove campaign covariates, segment covariates, or both.
- *Embeddings Only* estimates treatment effects using segment and campaign embeddings, and does not combine the embedding-based predictions with the observed outcomes in the focal training data; see Equation (7). To compare with *Standard PMF*, we estimate the *Embeddings Only* with no covariates.
- *Standard PMF* differs from the proposed model in three ways: (a) it does not incorporate campaign-specific or segment-specific covariates, (b) it does not update its model-based predictions using the observed focal training data, and (c) it does not account for the uncertainty in the treatment effect measures.

The findings in Table 2 reveal that the *Standard PMF* model outperforms a random baseline, but lags significantly behind the *Embeddings Only* specification. The disparity stems from how the two models handle uncertainty: *Standard PMF* assumes equal weights

for each segment-campaign treatment effect measure, causing it to overfit to noisy model inputs. In contrast, the *Embeddings Only* model accounts for the precision of treatment effect measures, allowing it to recover a more accurate latent structure.

Table 2: Decomposing the Source of the Performance Improvements

	Same Period		**Next Year**	
Specification	**Qini**	**Profit**	**Qini**	**Profit**
Full Covariates	10.12	22.30¢	9.15	20.29¢
	(0.63)	(1.96)	(0.81)	(2.44)
Only Campaign Covariates	10.09	21.97¢	9.32	20.29¢
	(0.61)	(1.97)	(0.83)	(2.50)
Only Segment Covariates	10.09	22.26¢	9.11	20.30¢
	(0.63)	(1.96)	(0.81)	(2.44)
No Covariates	9.74	21.84¢	9.00	19.70¢
	(0.62)	(1.97)	(0.84)	(2.46)
Embeddings Only	8.11	16.41¢	6.90	14.94¢
	(0.68)	(1.85)	(0.91)	(2.44)
Standard PMF	2.91	6.04¢	2.88	6.22¢
	(0.51)	(1.58)	(0.55)	(1.94)

Notes: The **Same Period** comparison compares findings across 361 campaigns, and the **Next Year** comparison includes 198 campaigns. Standard errors are reported in parentheses.

We also see that the proposed model with *No Covariates* outperforms the *Embeddings Only* model. This highlights the importance of the updating rule described in Equation (7). The updating rule allows the model to adaptively balance the global latent structure with segment-specific evidence, which is enabled by the Bayesian framework.

Finally, adding covariates to create the *Full Covariates* model provides a directional improvement in performance, although the performance improvement is not statistically significant in this setting. This suggests that the campaign and customer covariates do not contribute a lot of incremental information compared to estimating embeddings using only customer responsiveness. This has an important practical implication: if the differences

between campaigns are sufficiently well captured by the responsiveness-based embeddings then firms do not need to maintain detailed records of the design characteristics of each campaign.[4]

To further investigate how the model captures this information, we examine the inferred campaign and segment embeddings. In the Web Appendix, we show that the model recovers meaningful structural similarities directly from the customer responsiveness, even when covariates are omitted from model training. For example, segments with similar historical spending are located closer to one another in the embedding space. Likewise, a campaign's nearest neighbors in the embedding space tend to share campaign types, seasonality, and temporal proximity. These findings confirm the internal validity of our approach and illustrate how the embeddings effectively capture the underlying campaign and segment differences from the customer responsiveness information.

In summary, this section demonstrates the practical value of combining information across marketing campaigns to improve targeting policies. Because the model learns from the covariation in responses across segments, rather than relying solely on campaign-level average treatment effects, it extracts substantial information from each source campaign. The method's ability to achieve these gains critically depends on the incorporation of measurement precision and the adaptive updating rule. These components allow the model to distinguish between signal and noise more effectively than traditional methods. Having established these internal drivers of model performance, we next investigate the boundary conditions by systematically varying the set of source campaigns used for transfer learning.

6. BOUNDARY CONDITIONS

We investigate boundary conditions for the proposed approach by varying three dimensions: (1) source campaign characteristics (recency, seasonality, and campaign type), (2) focal dataset size, and (3) the number and size of source campaigns. To isolate the effect of

[4]We caution that further research is required to investigate whether including covariates in different retail settings or including covariates using more-flexible functional forms can have a larger impact on targeting performance.

data size while holding campaigns constant, we focus on large campaigns and downsample the customer data used for training. These analyses document when information transfer is most valuable and reveal the practical constraints on the method's performance.

In the previous section we measured targeting performance using both the **Next Year** and **Same Period** frameworks. Hereafter, we exclusively use the **Same Period** approach, which eliminates confounding effects from non-stationarity and nearly doubles the available sample of focal campaigns. This larger sample provides sufficient statistical power to detect performance differences when evaluating subsets of source campaigns.

6.1. Source Campaign Characteristics

We start by exploring how the recency of the source campaign relative to the focal campaign affects performance. To reduce variation and improve comparability, we restrict attention to focal campaigns with at least 30 source campaigns available in each time window and then randomly select exactly 30 source campaigns from the corresponding window for each evaluation.

Table 3: Recency of the Source Campaigns

Recency of Source Campaigns	Qini	Profit
0–1 Years	9.14 (0.84)	23.12¢ (2.72)
1–2 Years	8.88 (0.85)	22.76¢ (2.73)
2–3 Years	8.68 (0.84)	22.54¢ (2.72)
Focal Only	4.26 (0.76)	16.72¢ (2.51)
Blanket Treatment	–	7.09¢ (3.14)

Notes: The table reports performance statistics averaged across the 221 campaigns, which have at least 30 source campaigns in each recency window. Standard errors are reported in parentheses.

Table 3 reports the performance of the proposed method according to the time interval between the in-home dates of the focal and source campaigns. The findings reveal that recent source campaigns are significantly more effective at increasing performance than older ones (using paired sample t-tests). This aligns with our intuition regarding non-stationarity;

consumer behavior and campaign designs evolve, making recent data a more accurate proxy for the focal task. Nevertheless, older campaigns still provide substantial incremental value over the baseline, even if they occurred over two years prior. This suggests that retailers with few recent campaigns can still successfully extract signal from older campaigns.

We also investigated the seasonality and types of the source and focal campaigns. For seasonality, we grouped campaigns according to the calendar quarter in which they were mailed. We then evaluated whether performance improved if there was alignment in the quarters in which the source and focal campaigns were mailed. With one exception, seasonal alignment between source and focal campaigns does appear to improve targeting performance. However, the performance differences are relatively small.

For the campaign type, we compared the outcomes of the 235 focal incrementality campaigns ("mail" versus "no mail") and the 82 promotional campaigns ("mail + coupon" versus "mail only"). We trained targeting policies for each set separately using three distinct pools of source campaigns: (a) strictly incrementality, (b) strictly promotional, and (c) randomly selected. As expected, targeting policies are more profitable when the focal and source campaigns align in type. However, this alignment yields only modest performance improvements compared to using misaligned or randomly selected source campaigns. Although paired-sample t-tests indicate that some of these differences are statistically significant, they are generally not large enough to be managerially meaningful. The findings for seasonality and campaign type are reported in the Web Appendix.

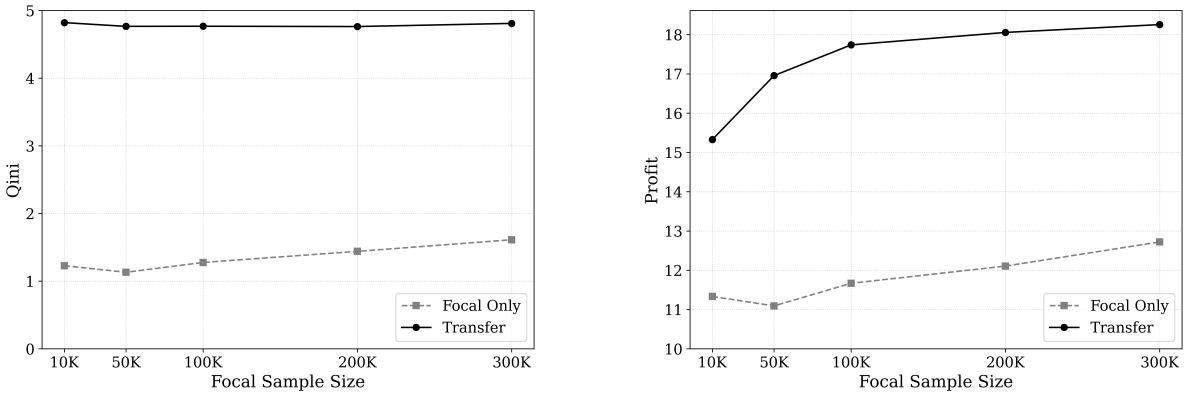
We conclude that the proposed transfer approach can effectively improve targeting policies even when source campaigns appear substantially different from the focal campaign.

6.2. Size of the Focal Training Dataset

In this analysis, we focus on large campaigns with at least 300,000 customers in the focal training dataset. This yields 95 focal campaigns. We then randomly draw without replacement subsets of the focal training data to implement both the proposed model and the *Focal*

Only benchmark. Specifically, we select 10,000, 50,000, 100,000, 200,000 or 300,000 customers to comprise the focal training datasets. This approach allows us to vary the size of the focal training dataset while holding the set of focal campaigns fixed.

Figure 4: Varying the Size of the Focal Training Dataset



(a) Qini **(b)** Profit

Notes: The figures report performance statistics averaged across 95 campaigns.

Figure 4 reports the average performance of the two models. Across both performance metrics, the *Focal Only* model's performance improves as a log-linear function of sample size, a result consistent with the theoretical properties of the difference-in-means estimator. However, the performance of the proposed *Transfer* model varies depending on the metric. While Profit initially improves rapidly when the focal training data is small and plateaus at approximately 100k customers, the Qini metric is notably stable when varying the size of the focal data. This suggests that the relative ranking of customer segments (captured by Qini) is highly transferable across campaigns, whereas the absolute values of treatment effects used for profit optimization are more campaign-specific and so require more focal data. For practitioners, this implies that ranking-based targeting requires substantially less focal data than targeting based on profit thresholds.

It is helpful to recognize that even with little focal training data the proposed model does not collapse to a random prediction. The model can use the information contained in the

source campaigns to learn the embedding structure. This in turn provides information to inform predictions of the typical responsiveness of customers in the focal campaign.

6.3. Number and Size of Source Campaigns

To investigate the model performance when varying the number of source campaigns, we restrict attention to focal campaigns that have at least 50 source campaigns available (ruling out the very earliest campaigns in our dataset). For each focal campaign we randomly draw without replacement subsets of 5, 10, 20, 30, 40 or 50 campaigns to use in the training data.

Table 4: Varying the Number of Source Campaigns

Nbr. Source Campaigns	Qini	Profit
50	9.78 (0.65)	22.37¢ (2.07)
40	9.79 (0.65)	22.41¢ (2.07)
30	9.68 (0.65)	22.33¢ (2.08)
20	9.58 (0.66)	22.21¢ (2.08)
10	9.43 (0.66)	21.97¢ (2.08)
5	9.12 (0.66)	21.61¢ (2.07)
Focal Only	4.66 (0.59)	16.35¢ (1.90)
Blanket Treatment	–	5.36¢ (2.37)

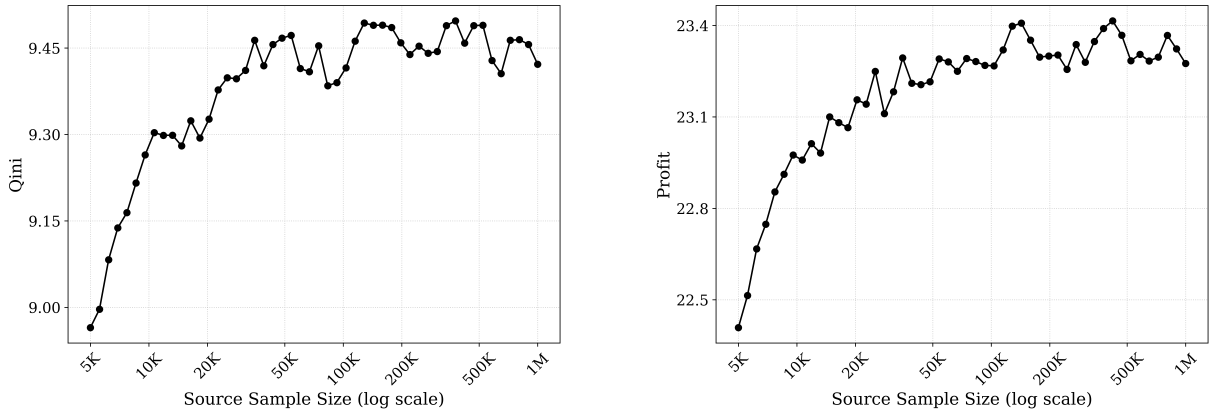
Notes: The table reports performance statistics averaged across the 341 campaigns, for which at least 50 source campaigns are available. Standard errors are reported in parentheses.

Table 4 reports the results. As expected, targeting performance improves with the number of source campaigns included in the estimation. However, even with just 5 source campaigns, we obtain a large proportion of the performance improvement available when using 50 campaigns. This finding suggests that relatively few source campaigns are sufficient to infer the embedding structure in our approach. Consequently, applications are not restricted to large retailers with hundreds of past experiments. The proposed model could also be used by smaller retailers who may only run campaigns quarterly or during holiday seasons.

Next we investigate the size of source campaigns. We focus on source campaigns with at least 1 million customers, and the 309 focal campaigns with at least 30 such sources. We

then randomly downsample source campaigns to different sample sizes and use these data to train the proposed model. This approach allows us to compare performance while holding the set of source and focal campaigns fixed. The findings are reported in Figure 5. Larger source campaigns improve model performance at a relatively consistent rate until source campaigns contain 100,000 customers, at which point the improvement plateaus.

Figure 5: Varying the Size of the Source Data



(a) Qini **(b)** Profit

Notes: The figure reports performance statistics averaged across 309 campaigns.

We can compare the results in Figure 5 with those in Table 4. Both Qini and Profit metrics require *large source datasets*, but neither requires a *large number* of source campaigns. This comparison sheds light on how the embeddings use source information. A key input to the model is the noisy estimate of treatment effects in each source campaign. Our adjustment for measurement error allows the method to place greater weight on more precise source estimates. As a result, a small number of source campaigns can be sufficient to recover accurate embeddings, provided that those campaigns are large enough to estimate treatment effects precisely. By contrast, when we downsample the source campaigns, we introduce noise into all of their estimated treatment effects, which makes the embeddings less precise and reduces targeting performance.

To further understand the source and focal dataset requirements, it is helpful to recall

how the proposed method uses these data. We can simplify the model training into three tasks: (1) constructing embeddings for source campaigns and customer segments, (2) locating the focal campaign within the embedding space, and (3) updating predictions from the embeddings using the focal data. The global accuracy of the embeddings (Task 1) depends heavily on the information in the *source* datasets. In contrast, accurately locating the focal campaign (Task 2) and updating the segment-specific predictions (Task 3) rely almost entirely on the information in the *focal* training data.

This framework directly explains our empirical results. The Qini metric, which captures customer ranking, depends primarily on the quality of the global segment embeddings (Task 1). Consequently, Qini remains remarkably stable even with minimal focal data but deteriorates when source data is restricted (Figures 4a and 5a). In contrast, optimizing Profit requires identifying whether treatment effects exceed the treatment costs. This demands more precise focal campaign location and prediction updates (Tasks 2 and 3), explaining why Profit depends on both the focal and source data (Figures 4b and 5b).

6.4. Summary

The investigation of the boundary conditions yields important practical insights regarding transfer learning and data requirements. Our findings show that the proposed matrix factorization approach is remarkably robust to variations in source campaign characteristics. While incorporating more recent campaigns or aligning by season and campaign type provides modest performance benefits, the method successfully extracts valuable signal even from older or seemingly dissimilar campaigns.

Furthermore, our analysis clarifies the specific data requirements for both the source and focal datasets. We find that the size of the source campaigns is substantially more important than the total number of them. The model can capture a large proportion of the available performance gains with as few as five source campaigns, provided those campaigns are large enough to yield precise segment-level treatment effect estimates. For the focal

campaign, data requirements depend heavily on the targeting objective: accurately ranking customers (Qini metric) is feasible with minimal focal training data, whereas optimizing absolute profit thresholds requires larger focal samples to properly calibrate the magnitude of the predictions.

7. CONCLUSION

Targeting marketing promotions is an important application of machine learning in marketing. The performance of a targeting policy depends upon how much training data is available to reliably estimate treatment effects for different customers. Traditionally, firms have used information from either the same campaign conducted in a prior period, or a pilot experiment. We propose a method that augments these data by transferring information from past marketing campaigns to improve targeting decisions for a focal campaign.

A primary challenge in cross-campaign transfer is that past marketing actions, timing, and eligibility criteria vary significantly and are often poorly documented. We address this issue by reformulating the information transfer problem as a matrix factorization task. We extend the probabilistic matrix factorization (PMF) model to explicitly account for the varying precision of treatment effect estimates across segments and incorporate observable covariates into the hierarchical prior. Our results demonstrate that accounting for the accuracy of treatment effect estimates is important as it allows the model to distinguish responsiveness signals from noise inherent in retail experiments.

We use a large dataset comprising 391 randomized field experiments conducted by a large apparel retailer to evaluate the proposed approach. We consider settings in which the retailer has access to a (current) pilot study and settings in which it uses the prior year's campaign to provide training data for the focal campaign. In both settings, our approach consistently outperforms benchmarks. We investigate boundary conditions, and show how the performance of the proposed model depends upon the number and characteristics of source campaigns, and the size of the past experiments.

There are several avenues for future research. Currently, the model assumes that customer segments are pre-defined based on historical spending. Future extensions could integrate the segmentation process directly into the Bayesian framework and investigate how to extend the model to customer-level targeting (Kim, Bradlow, and Iyengar 2023; Zhang and Misra 2024). Additionally, while our application focused on direct mail at an apparel retailer, future studies could explore the model's boundaries across alternative industries, marketing channels and value propositions.

Our proposed method incorporates observable covariates for customer segments and marketing campaigns using linear functional forms. This structure preserves the analytical forms of the conditional distributions, and estimation of the model remains efficient. However, future research could investigate alternative approaches that use nonlinear specifications. For example, Adams, Dahl, and Murray (2010) incorporates covariates into the standard PMF model using Gaussian processes.

By treating a firm's experimental history as a collective repository of insights rather than a series of isolated events, our approach amplifies the value of marketing data and offers a robust, scalable solution for designing targeting policies.

REFERENCES

Adams, Ryan Prescott, George E Dahl, and Iain Murray (2010), "Incorporating side information in probabilistic matrix factorization with gaussian processes," *arXiv preprint arXiv:1003.4944*.

Ascarza, Eva and Ayelet Israeli (2022), "Eliminating unintended bias in personalized policies using bias-eliminating adapted trees (BEAT)," *Proceedings of the National Academy of Sciences*, 119 (11), e2115293119.

Ascarza, Eva, Scott A Neslin, Oded Netzer, Zachery Anderson, Peter S Fader, Sunil Gupta, Bruce G S Hardie, Aurélie Lemmens, Barak Libai, David Neal et al. (2018), "In pursuit of enhanced customer retention management: Review, key issues, and future directions," *Customer Needs and Solutions*, 5 (1), 65–81.

Dubé, Jean-Pierre and Sanjog Misra (2023), "Personalized pricing and consumer welfare," *Journal of Political Economy*, 131 (1), 131–189.

Ellickson, Paul B, Wreetabrata Kar, and James C Reeder III (2023), "Estimating marketing component effects: Double machine learning from targeted digital promotions," *Marketing Science*, 42 (4), 704–728.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), ""Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275–284.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V Kumar, Nathaniel Lin, Nalini Ravishanker, and S Sriram (2006), "Modeling customer lifetime value," *Journal of Service Research*, 9 (2), 139–155.

Hitsch, Günter J, Sanjog Misra, and Walter W Zhang (2024), "Heterogeneous treatment effects and optimal targeting policy evaluation," *Quantitative Marketing and Economics*, 22 (2), 115–168.

Huang, Ta-Wei, Eva Ascarza, and Ayelet Israeli (2024), "Incrementality Representation Learning: Synergizing Past Experiments for Intervention Personalization," *Available at SSRN 4859809*.

Kim, Mingyung, Eric Bradlow, and Raghuram Iyengar (2023), "A Bayesian Dual-Network Clustering Approach for Selecting Data and Parameter Granularities," *Available at SSRN 4497834*.

Kumar, V. and Werner Reinartz (2018), *Customer Relationship Management: Concept, Strategy, and Tools* Springer Berlin, Heidelberg, 3 edition.

Lakshminarayanan, Balaji, Guillaume Bouchard, and Cedric Archambeau "Robust Bayesian matrix factorisation," "Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics," pages 425–433, JMLR Workshop and Conference Proceedings (2011).

Lemmens, Aurélie and Sunil Gupta (2020), "Managing churn to maximize profits," *Marketing Science*, 39 (5), 956–973.

Salakhutdinov, Ruslan and Andriy Mnih "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," "Proceedings of the 25th international conference on Machine learning," pages 880–887 (2008).

Shchetkina, Anya and Ron Berman "When Is Heterogeneity Actionable for Targeting?," "Proceedings of the 25th ACM Conference on Economics and Computation," pages 778–779 (2024).

Simester, Duncan, Artem Timoshenko, and Spyros I Zoumpoulis (2020), "Efficiently evaluating targeting policies: Improving on champion vs. challenger experiments," *Management Science*, 66 (8), 3412–3424.

Simester, Duncan, Artem Timoshenko, and Spyros I Zoumpoulis (2025), "A sample size calculation for training and certifying targeting policies," *Management Science*.

Venkatesan, Rajkumar, Paul W Farris, and Ronald T Wilcox (2021), *Marketing Analytics: Essential Tools for Data-Driven Decisions* University of Virginia Press.

Yadlowsky, Steve, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager (2025), "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects," *Journal of the American Statistical Association*, 120 (549), 38–51 Epub 2024 Oct 11.

Yang, Hongxia "Bayesian heteroscedastic matrix factorization for conversion rate prediction," "Proceedings of the 2017 ACM on Conference on Information and Knowledge Management," pages 2407–2410 (2017).

Zhang, Walter and Sanjog Misra "Coarse personalization," "Proceedings of the 25th ACM Conference on Economics and Computation," pages 1206–1208 (2024).

Zhang, Walter Wang *Optimal comprehensible targeting* Ph.D. thesis, The University of Chicago (2024).

A. TABLE OF NOTATIONS

Symbol	Description
Model Structure & Data	
$s \in \{1, \ldots, S\}$	Index for customer segments.
$c \in \{1, \ldots, C\}$	Index for marketing campaigns.
K	Dimension of the latent embedding space.
D_x and D_z	Dimensions of observable feature vectors for segments and campaigns.
τ_{cs}	True Conditional Average Treatment Effect for campaign c in segment s.
y_{cs}	Observed noisy measure of the CATE for campaign c in segment s. Matrix form: $\mathbf{Y}$.
λ_{cs}	Known measurement precision (inverse variance) of y_{cs}. Matrix form: $\mathbf{\Lambda}_{obs}$.
q_{cs}	Binary indicator: 1 if data exists for pair (c, s), 0 otherwise.
$\mathbf{u}_s$	$K \times 1$ latent embedding vector for segment s. Matrix form: $\mathbf{U}$.
$\mathbf{v}_c$	$K \times 1$ latent embedding vector for campaign c. Matrix form: $\mathbf{V}$.
α	Global precision parameter for the structural error term.
$\mathbf{x}_s$	$D_x \times 1$ vector of observable covariates for segment s.
$\mathbf{z}_c$	$D_z \times 1$ vector of observable covariates for campaign c.
Prior Parameters (Hierarchical)	
$\mathbf{B}_u$	$D_x \times K$ matrix of regression coefficients linking $\mathbf{x}_s$ to $\mathbf{u}_s$.
$\mathbf{B}_v$	$D_z \times K$ matrix of regression coefficients linking $\mathbf{z}_c$ to $\mathbf{v}_c$.
$\mathbf{\Sigma}_u$	$K \times K$ covariance matrix for segment embeddings $\mathbf{u}_s$.
$\mathbf{\Sigma}_v$	$K \times K$ covariance matrix for campaign embeddings $\mathbf{v}_c$.
$\mathbf{M}_{u,0}$ and $\mathbf{M}_{v,0}$	Prior mean matrix for coefficients $\mathbf{B}_u$ and $\mathbf{B}_v$ (Matrix Normal).
$\mathbf{W}_{u,0}$ and $\mathbf{W}_{v,0}$	Prior scale matrix for coefficients $\mathbf{B}_u$ and $\mathbf{B}_v$ (Matrix Normal).
$\mathbf{\Lambda_0}$	Prior scale matrix for covariance $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_v$ (Inverse-Wishart).
ν_0	Degrees of freedom for the Inverse-Wishart prior.
a_0 and b_0	Shape and rate parameters for the Gamma prior on α.
Model Inference (Posterior Parameters)	
$\boldsymbol{\mu}_s$	Posterior mean vector of segment embedding $\mathbf{u}_s$.
$\mathbf{S}_s$	Posterior covariance matrix of segment embedding $\mathbf{u}_s$.
$\boldsymbol{\mu}_c$	Posterior mean vector of campaign embedding $\mathbf{v}_c$.
$\mathbf{S}_c$	Posterior covariance matrix of campaign embedding $\mathbf{v}_c$.
$\nu_{u,n}$ and $\nu_{v,n}$	Posterior degrees of freedom for $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_v$ (updating ν_0).
$\mathbf{\Lambda}_{u,n}$ and $\mathbf{\Lambda}_{v,n}$	Posterior scale matrix for covariances $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_v$.
$\mathbf{W}_{u,n}$ and $\mathbf{W}_{v,n}$	Posterior scale matrix for coefficients $\mathbf{B}_u$ and $\mathbf{B}_v$.
$\mathbf{M}_{u,n}$ and $\mathbf{M}_{v,n}$	Posterior mean matrix for coefficients $\mathbf{B}_u$ and $\mathbf{B}_v$.

B. ALGORITHM FOR MODEL INFERENCE

Algorithm 1 Proposed Method: Bayesian Matrix Factorization

Inputs: $\mathbf{Y}, \mathbf{\Lambda}_{\text{obs}}, \mathbf{Q}, \mathbf{X}, \mathbf{Z}$

Initialize: $\mathbf{U}^0 \sim \mathcal{U}(-1,1)^{\text{K}}, \mathbf{V}^0 \sim \mathcal{U}(-1,1)^{\text{K}}, \alpha^0 = 1, \mathbf{\Lambda_0} = \mathbf{I}_{\text{K}}, \nu_0 = \text{K}, \text{a} = 1, \text{b} = 1$
$\mathbf{M}_{\text{v},0} = \mathbf{0}_{\text{D}_\text{z}\times\text{K}}, \mathbf{W}_{\text{v},0} = \mathbf{I}_{\text{D}_\text{z}}, \mathbf{M}_{\text{u},0} = \mathbf{0}_{\text{D}_\text{x}\times\text{K}}, \mathbf{W}_{\text{u},0} = \mathbf{I}_{\text{D}_\text{x}}$

for $\text{t} = 1$ to $\text{t}_0 + \text{T}$ **do**

Sample α^{t} using HMC algorithm from Algorithm 2
$\alpha^{\text{t}} \sim \mathbb{P}(\alpha|\mathbf{U}^{\text{t}-1}, \mathbf{V}^{\text{t}-1}, \mathcal{D})$

Compute posterior parameters for $(\mathbf{B}_{\text{u}}, \mathbf{\Sigma}_{\text{u}})$
$\mathbf{W}_{\text{u},n} = \mathbf{X}'\mathbf{X} + \mathbf{W}_{\text{u},0}$
$\mathbf{M}_{\text{u},n} = \mathbf{W}_{\text{u},n}^{-1}(\mathbf{X}'\mathbf{U}^{\text{t}-1} + \mathbf{W}_{\text{u},0}\mathbf{M}_{\text{u},0})$
$\mathbf{\Lambda}_{\text{u,n}} = \mathbf{\Lambda_0} + (\mathbf{U}^{\text{t}-1} - \mathbf{X}\mathbf{M}_{\text{u},n})'(\mathbf{U}^{\text{t}-1} - \mathbf{X}\mathbf{M}_{\text{u},n}) + (\mathbf{M}_{\text{u},n} - \mathbf{M}_{\text{u},0})'\mathbf{W}_{\text{u},0}(\mathbf{M}_{\text{u},n} - \mathbf{M}_{\text{u},0})$

Sample $(\mathbf{B}_{\text{u}}, \mathbf{\Sigma}_{\text{u}})$ sequentially
$\mathbf{\Sigma}_{\text{u}}^{\text{t}} \sim \mathcal{W}^{-1}(\mathbf{\Lambda}_{\text{u,n}}, \nu_0 + \text{S})$
$\mathbf{B}_{\text{u}}^{\text{t}} \sim \mathcal{MN}(\mathbf{M}_{\text{u},n}, \mathbf{W}_{\text{u},n}^{-1}, \mathbf{\Sigma}_{\text{u}}^{\text{t}})$.

Compute posterior parameters for $(\mathbf{B}_{\text{v}}, \mathbf{\Sigma}_{\text{v}})$
$\mathbf{W}_{\text{v},n} = \mathbf{Z}'\mathbf{Z} + \mathbf{W}_{\text{v},0}$
$\mathbf{M}_{\text{v},n} = \mathbf{W}_{\text{v},n}^{-1}(\mathbf{Z}'\mathbf{V}^{\text{t}-1} + \mathbf{W}_{\text{v},0}\mathbf{M}_{\text{v},0})$
$\mathbf{\Lambda}_{\text{v,n}} = \mathbf{\Lambda_0} + (\mathbf{V}^{\text{t}-1} - \mathbf{Z}\mathbf{M}_{\text{v},n})'(\mathbf{V}^{\text{t}-1} - \mathbf{Z}\mathbf{M}_{\text{v},n}) + (\mathbf{M}_{\text{v},n} - \mathbf{M}_{\text{v},0})'\mathbf{W}_{\text{v},0}(\mathbf{M}_{\text{v},n} - \mathbf{M}_{\text{v},0})$

Sample $(\mathbf{B}_{\text{v}}, \mathbf{\Sigma}_{\text{v}})$ sequentially
$\mathbf{\Sigma}_{\text{v}}^{\text{t}} \sim \mathcal{W}^{-1}(\mathbf{\Lambda}_{\text{v,n}}, \nu_0 + \text{C})$
$\mathbf{B}_{\text{v}}^{\text{t}} \sim \mathcal{MN}(\mathbf{M}_{\text{v},n}, \mathbf{W}_{\text{v},n}^{-1}, \mathbf{\Sigma}_{\text{v}}^{\text{t}})$

Compute posterior parameters for and sample $\mathbf{u}_{\text{s}}^{\text{t}}$
for $\text{s} = 1$ to S **do**
$\mathbf{S}_{\text{s}} = [(\mathbf{\Sigma}_{\text{u}}^{\text{t}})^{-1} + \sum_{\text{c}} \text{q}_{\text{cs}}(\lambda_{\text{cs}}^{-1} + (\alpha^{\text{t}})^{-1})^{-1}\mathbf{v}_{\text{c}}^{\text{t}-1}(\mathbf{v}_{\text{c}}^{\text{t}-1})']^{-1}$
$\boldsymbol{\mu}_{\text{s}} = \mathbf{S}_{\text{s}}\,[\sum_{\text{c}} \text{q}_{\text{cs}}(\lambda_{\text{cs}}^{-1} + (\alpha^{\text{t}})^{-1})^{-1}\mathbf{v}_{\text{c}}^{\text{t}-1}\text{y}_{\text{cs}} + (\mathbf{\Sigma}_{\text{u}}^{\text{t}})^{-1}(\mathbf{B}_{\text{u}}^{\text{t}}\mathbf{x}_{\text{s}})]$
$\mathbf{u}_{\text{s}}^{\text{t}} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{s}}, \mathbf{S}_{\text{s}})$.

Compute posterior parameters for and sample $\mathbf{v}_{\text{c}}^{\text{t}}$
for $\text{c} = 1$ to C **do**
$\mathbf{S}_{\text{c}} = [(\mathbf{\Sigma}_{\text{v}}^{\text{t}})^{-1} + \sum_{\text{s}} \text{q}_{\text{cs}}(\lambda_{\text{cs}}^{-1} + (\alpha^{\text{t}})^{-1})^{-1}\mathbf{u}_{\text{s}}^{\text{t}}(\mathbf{u}_{\text{s}}^{\text{t}})']^{-1}$
$\boldsymbol{\mu}_{\text{c}} = \mathbf{S}_{\text{c}}\,[\sum_{\text{s}} \text{q}_{\text{cs}}(\lambda_{\text{cs}}^{-1} + (\alpha^{\text{t}})^{-1})^{-1}\mathbf{u}_{\text{s}}^{\text{t}}\text{y}_{\text{cs}} + (\mathbf{\Sigma}_{\text{v}}^{\text{t}})^{-1}(\mathbf{B}_{\text{v}}^{\text{t}}\mathbf{z}_{\text{c}})]$
$\mathbf{v}_{\text{c}}^{\text{t}} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{c}}, \mathbf{S}_{\text{c}})$.

Compute predicted CATEs using T samples
for $\text{c} = 1$ to C and $\text{s} = 1$ to S **do**

$$\widehat{\tau}_{\text{cs}} = \begin{cases} \frac{1}{\text{T}} \sum_{\text{t}=\text{t}_0}^{\text{t}_0+\text{T}} \mathbf{u}_{\text{s}}^{\text{t}\prime}\mathbf{v}_{\text{c}}^{\text{t}} & \text{if } \text{q}_{\text{cs}} = 0 \\ \frac{1}{\text{T}} \sum_{\text{t}=\text{t}_0}^{\text{t}_0+\text{T}} (\lambda_{\text{cs}} + \alpha^{\text{t}})^{-1} \cdot \left(\alpha^{\text{t}}\mathbf{u}_{\text{s}}^{\text{t}\prime}\mathbf{v}_{\text{c}}^{\text{t}} + \lambda_{\text{cs}}\text{y}_{\text{cs}}\right) & \text{if } \text{q}_{\text{cs}} = 1 \end{cases}$$

Notes: $\mathbf{0}_{M\times N}$ is $M \times N$ matrix with zeros, $\mathbf{I}_N$ is $N \times N$ identity matrix.

Algorithm 2 Hamiltonian Monte Carlo Sampling

function SAMPLE_$\alpha(\alpha^{t-1}, \mathbf{U}^{t-1}, \mathbf{V}^{t-1}, \mathcal{D}, t, t_0, a_0, b_0)$

Persistent: $\overline{\delta}, \overline{\epsilon}, \epsilon$

Initialize: $\epsilon_0 = 0.1, \tau = 10, \kappa_0 = 0.75, \gamma_0 = 0.05, \delta_0 = 0.8$

if $t = 1$ **then**

$\overline{\delta} \leftarrow 0, \quad \overline{\epsilon} \leftarrow 1, \quad \epsilon \leftarrow \epsilon_0$

Sample ω and use leapfrog integrator

$\omega \sim \mathcal{N}(\omega \mid 0, 1)$

$L^t \leftarrow \lceil 1/\epsilon \rceil$

$\alpha \leftarrow \alpha^{t-1}$

$\widetilde{\omega} \leftarrow \omega$

for $l = 1, \ldots, L^t$ **do**

Update $\widetilde{\omega} \leftarrow \widetilde{\omega} + \frac{\epsilon}{2} \cdot \frac{\partial \log \pi(\alpha)}{\partial \alpha}$.

Update $\alpha \leftarrow \alpha + \epsilon \cdot \widetilde{\omega}$.

Update $\widetilde{\omega} \leftarrow \widetilde{\omega} + \frac{\epsilon}{2} \cdot \frac{\partial \log \pi(\alpha)}{\partial \alpha}$.

Sample u and make metropolis accept step

$u \sim \mathcal{U}(0, 1)$.

$$\alpha^t \leftarrow \begin{cases} \alpha & \text{if } u \leq \min\left\{1, \frac{\pi(\alpha)}{\pi(\alpha^{t-1})} \exp\left[\frac{\omega^2 - \widetilde{\omega}^2}{2}\right]\right\}, \\ \alpha^{t-1} & \text{otherwise.} \end{cases}$$

Update time interval ϵ during warm-up

if $t \leq t_0$ **then**

$\overline{\delta} \leftarrow \left(1 - \frac{1}{t+\tau}\right)\overline{\delta} + \frac{1}{t+\tau}(\delta_0 - \alpha^t)$

$\epsilon \leftarrow 10\epsilon_0 \exp\left(-\frac{\sqrt{t}}{\gamma_0}\overline{\delta}\right)$

$\log \overline{\epsilon} \leftarrow t^{-\kappa_0} \log \epsilon + (1 - t^{-\kappa_0}) \log \overline{\epsilon}$

else

$\epsilon \leftarrow \overline{\epsilon}$

return α^t

Notes: $\log \pi(\alpha) = -\sum_{c,s} \frac{q_{cs}}{2}\left[\frac{(y_{cs} - \mathbf{u}_s^{t-1}\mathbf{v}_c^{t-1})^2}{\lambda_{cs}^{-1} + \alpha^{-1}} + \log\left(\lambda_{cs}^{-1} + \alpha^{-1}\right)\right] + \log \mathcal{G}(\alpha | a_0, b_0)$

Web Appendix

Improving Targeting Policies by Learning Across Marketing Campaigns

by Marat Ibragimov, Duncan Simester, and Artem Timoshenko

Table of Contents

Web Appendix WA1: Closed Form Solution for Treatment Effects

To find the closed-form solution for $\mathbb{P}(\tau_{cs}|\mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y})$ in Equation (5) we use Bayes' theorem to decompose the original probability:

$$\mathbb{P}(\tau_{cs}|\mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y}) = \frac{\mathbb{P}(\tau_{cs}, \mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y})}{\mathbb{P}(\mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y})} = \frac{\mathbb{P}(\mathbf{Y}|\tau_{cs}, \mathbf{U}, \mathbf{V}, \alpha)\mathbb{P}(\tau_{cs}|\mathbf{u}_s, \mathbf{v}_c, \alpha)}{\mathbb{P}(\mathbf{Y}|\mathbf{U}, \mathbf{V}, \alpha)} \tag{WA1}$$

Equation (WA1) depends on whether y_{cs} is available ($q_{cs} = 1$) or unobserved ($q_{cs} = 0$). *Case 1:* $q_{cs} = 0$. There is no available estimate of the treatment effect for campaign c and segment s. In this case $\mathbf{Y}$ does not contain y_{cs}, thus $\mathbb{P}(\mathbf{Y}|\tau_{cs}, \mathbf{U}, \mathbf{V}, \alpha) = \mathbb{P}(\mathbf{Y}|\mathbf{U}, \mathbf{V}, \alpha)$ and

$$\mathbb{P}(\tau_{cs}|\mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y}) = \mathbb{P}(\tau_{cs}|\mathbf{u}_s, \mathbf{v}_c, \alpha) = \mathcal{N}\left(\tau_{cs}\middle|\mathbf{u}_s'\mathbf{v}_c, \alpha^{-1}\right) \tag{WA2}$$

Case 2: $q_{cs} = 1$. Only y_{cs} would be impacted by conditioning on τ_{cs}. Thus,

$$\begin{aligned} \mathbb{P}(\tau_{cs}|\mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y}) &= \frac{\mathbb{P}(\tau_{cs}, \mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y})}{\mathbb{P}(\mathbf{U}, \mathbf{V}, \alpha, \mathbf{Y})} = \frac{\mathbb{P}(y_{cs}|\tau_{cs})\mathbb{P}(\tau_{cs}|\mathbf{u}_s, \mathbf{v}_c, \alpha)}{\mathbb{P}(y_{cs}|\mathbf{u}_s, \mathbf{v}_c, \alpha)} \\ = \frac{\mathcal{N}\left(y_{cs}|\tau_{cs}, \lambda_{cs}^{-1}\right)\mathcal{N}\left(\tau_{cs}|\mathbf{u}_s'\mathbf{v}_c, \alpha^{-1}\right)}{\mathcal{N}\left(y_{cs}|\mathbf{u}_s'\mathbf{v}_c, (\lambda_{cs}+\alpha)^{-1}\right)} &= \mathcal{N}\left(\tau_{cs}\middle|\frac{\mathbf{u}_s'\mathbf{v}_c\alpha + y_{cs}\lambda_{cs}}{\lambda_{cs}+\alpha}, \frac{1}{\lambda_{cs}+\alpha}\right) \end{aligned} \tag{WA3}$$

After combining Equations (WA2) and (WA3) we obtain Equation (5).

Web Appendix WA2: Derivation of Sampling Distribution

Using the model structure discussed in Section 3 and the law of total probability we can express the posterior predictive distribution $\mathbb{P}(\tau_{cs}|\mathcal{D})$:

$$\mathbb{P}(\tau_{cs}|\mathcal{D}) = \iint \mathbb{P}(\tau_{cs}|\mathbf{U},\mathbf{V},\alpha,\mathcal{D})\mathbb{P}(\mathbf{U},\mathbf{V},\alpha,\mathbf{B}_u,\boldsymbol{\Sigma}_u,\mathbf{B}_v,\boldsymbol{\Sigma}_v|\mathcal{D})\mathrm{d}\mathbf{U}\mathrm{d}\mathbf{V}\mathrm{d}\alpha\mathrm{d}\mathbf{B}_u\mathrm{d}\boldsymbol{\Sigma}_u\mathrm{d}\mathbf{B}_v\mathrm{d}\boldsymbol{\Sigma}_v \tag{WA4}$$

The integrand consists of two components. In Appendix WA1, we derive a closed-form expression for the first component:

$$\mathbb{P}(\tau_{cs}|\mathbf{U},\mathbf{V},\alpha,\mathcal{D}) = \begin{cases} \mathcal{N}\left(\tau_{cs}|\mathbf{u}_s'\mathbf{v}_c,\alpha^{-1}\right) & \text{if } q_{cs}=0 \\ \mathcal{N}\left(\tau_{cs}\middle|\frac{\mathbf{u}_s'\mathbf{v}_c\alpha+y_{cs}\lambda_{cs}}{\lambda_{cs}+\alpha},\frac{1}{\lambda_{cs}+\alpha}\right) & \text{if } q_{cs}=1 \end{cases} \tag{WA5}$$

Since the distribution $\mathbb{P}(\mathbf{U},\mathbf{V},\alpha,\mathbf{B}_u,\boldsymbol{\Sigma}_u,\mathbf{B}_v,\boldsymbol{\Sigma}_v|\mathcal{D})$ lacks a known closed-form expression, we employ a Gibbs sampling algorithm to draw samples from the distribution. The algorithm draws each variable iteratively, conditioning on the current values of all other variables. We next derive all conditional probabilities.

Conditional Distributions for U and V

We start our derivations with the conditional distribution of $\mathbf{U}$:

$$\begin{aligned} \mathbb{P}(\mathbf{U}|\mathbf{V},\alpha,\mathbf{B}_u,\boldsymbol{\Sigma}_u,\mathbf{B}_v,\boldsymbol{\Sigma}_v,\mathcal{D}) &= \frac{\mathbb{P}(\mathbf{U},\mathbf{V},\alpha,\mathbf{B}_u,\boldsymbol{\Sigma}_u,\mathbf{B}_v,\boldsymbol{\Sigma}_v,\mathcal{D})}{\int \mathbb{P}(\mathbf{U},\mathbf{V},\alpha,\mathbf{B}_u,\boldsymbol{\Sigma}_u,\mathbf{B}_v,\boldsymbol{\Sigma}_v,\mathcal{D})\mathrm{d}\mathbf{U}} \\ &= \frac{\mathbb{P}(\mathcal{D}|\mathbf{U},\mathbf{V},\alpha)\mathbb{P}(\mathbf{U}|\mathbf{B}_u,\boldsymbol{\Sigma}_u)}{\int \mathbb{P}(\mathcal{D}|\mathbf{U},\mathbf{V},\alpha)\mathbb{P}(\mathbf{U}|\mathbf{B}_u,\boldsymbol{\Sigma}_u)\mathrm{d}\mathbf{U}} \\ &\propto \left[\prod_{c,s}\mathcal{N}\left(y_{cs}|\mathbf{u}_s'\mathbf{v}_c,\alpha^{-1}+\lambda_{cs}^{-1}\right)\right]\left[\prod_s \mathcal{N}(\mathbf{u}_s|\mathbf{B}_u\mathbf{x}_s,\boldsymbol{\Sigma}_u)\right] \\ &\propto \prod_s\left[\mathcal{N}(\mathbf{u}_s|\mathbf{B}_u\mathbf{x}_s,\boldsymbol{\Sigma}_u)\prod_c \mathcal{N}\left(y_{cs}|\mathbf{u}_s'\mathbf{v}_c,\alpha^{-1}+\lambda_{cs}^{-1}\right)\right] \end{aligned} \tag{WA6}$$

Equation (WA6) demonstrates that all components $\mathbf{u}_s$ are conditionally independent, so the sampling can be done in parallel for each $\mathbf{u}_s$ separately. Furthermore, for each component

$\mathbf{u}_s$, we can use the properties of the conjugate priors to derive the sampling distribution:

$$\begin{aligned} &\mathbb{P}(\mathbf{U}|\mathbf{V}, \alpha, \mathbf{B}_u, \mathbf{\Sigma}_u, \mathbf{B}_v, \mathbf{\Sigma}_v, \mathcal{D}) \\ &= \prod_s \mathbb{P}(\mathbf{u}_s \mid \mathbf{V}, \alpha, \mathbf{B}_u, \mathbf{\Sigma}_u, \mathcal{D}) = \prod_s \mathcal{N}(\mathbf{u}_s \mid \boldsymbol{\mu}_s, \mathbf{S}_s) \end{aligned} \tag{WA7}$$

where the parameters of the distribution can be expressed analytically:

$$\begin{aligned} \boldsymbol{\mu}_s &= \mathbf{S}_s \left[\sum_c \frac{q_{cs}}{\lambda_{cs}^{-1} + \alpha^{-1}} \mathbf{v}_c y_{cs} + \mathbf{\Sigma}_u^{-1} (\mathbf{B}_u \mathbf{x}_s) \right] \\ \mathbf{S}_s &= \left[\mathbf{\Sigma}_u^{-1} + \sum_c \frac{q_{cs}}{\lambda_{cs}^{-1} + \alpha^{-1}} \mathbf{v}_c \mathbf{v}_c' \right]^{-1} \end{aligned} \tag{WA8}$$

By symmetry, we derive the closed-form expression for the campaign embeddings $\mathbf{v}_c$:

$$\begin{aligned} &\mathbb{P}(\mathbf{V}|\mathbf{U}, \alpha, \mathbf{B}_v, \mathbf{\Sigma}_v, \mathbf{B}_u, \mathbf{\Sigma}_u, \mathcal{D}) \\ &= \prod_c \mathbb{P}(\mathbf{v}_c \mid \mathbf{U}, \alpha, \mathbf{B}_v, \mathbf{\Sigma}_v, \mathcal{D}) = \prod_c \mathcal{N}(\mathbf{v}_c \mid \boldsymbol{\mu}_c, \mathbf{S}_c) \end{aligned} \tag{WA9}$$

with the following parameters of the posterior distribution:

$$\begin{aligned} \boldsymbol{\mu}_c &= \mathbf{S}_c \left[\sum_s \frac{q_{cs}}{\lambda_{cs}^{-1} + \alpha^{-1}} \mathbf{u}_s y_{cs} + \mathbf{\Sigma}_v^{-1} (\mathbf{B}_v \mathbf{z}_c) \right] \\ \mathbf{S}_c &= \left[\mathbf{\Sigma}_v^{-1} + \sum_s \frac{q_{cs}}{\lambda_{cs}^{-1} + \alpha^{-1}} \mathbf{u}_s \mathbf{u}_s' \right]^{-1} \end{aligned} \tag{WA10}$$

Conditional Distributions for $(\mathbf{B}_u, \mathbf{\Sigma}_u)$ and $(\mathbf{B}_v, \mathbf{\Sigma}_v)$

Similarly, we can compute the distribution $(\mathbf{B}_u, \mathbf{\Sigma}_u)$:

$$\begin{aligned} &\mathbb{P}(\mathbf{B}_u, \mathbf{\Sigma}_u \mid \mathbf{U}, \mathbf{V}, \alpha, \mathbf{B}_v, \mathbf{\Sigma}_v, \mathcal{D}) \propto \mathbb{P}(\mathbf{U} \mid \mathbf{B}_u, \mathbf{\Sigma}_u, \mathcal{D}) \mathbb{P}(\mathbf{B}_u, \mathbf{\Sigma}_u) \\ &\propto \left[\prod_s \mathcal{N}(\mathbf{u}_s \mid \mathbf{B}_u \mathbf{x}_s, \mathbf{\Sigma}_u) \right] \mathcal{MN}\left(\mathbf{B}_u | \mathbf{M}_{u,0}, \mathbf{W}_{u,0}^{-1}, \mathbf{\Sigma}_u\right) \mathcal{W}^{-1}(\mathbf{\Sigma}_u | \mathbf{\Lambda_0}, \nu_0) \\ &= \mathcal{MN}\left(\mathbf{B}_u | \mathbf{M}_{u,n}, \mathbf{W}_{u,n}^{-1}, \mathbf{\Sigma}_u\right) \mathcal{W}^{-1}(\mathbf{\Sigma}_u | \mathbf{\Lambda}_{u,n}, \nu_{u,n}) \end{aligned} \tag{WA11}$$

where the parameters of the posterior distribution can be expressed analytically:

$$\begin{aligned}
&\nu_{\mathrm{u,n}} = \nu_0 + \mathrm{S} \\
&\mathbf{W}_{\mathrm{u},n} = \mathbf{X}'\mathbf{X} + \mathbf{M}_{\mathrm{u},0} \\
&\mathbf{M}_{\mathrm{u},n} = \mathbf{W}_{\mathrm{u},n}^{-1}\left(\mathbf{X}'\mathbf{U} + \mathbf{W}_{\mathrm{u},n}\mathbf{M}_0\right) \\
&\mathbf{\Lambda}_{\mathrm{u,n}} = \mathbf{\Lambda_0} + (\mathbf{U} - \mathbf{X}\mathbf{M}_{\mathrm{u},n})'(\mathbf{U} - \mathbf{X}\mathbf{M}_{\mathrm{u},n}) + (\mathbf{M}_{\mathrm{u},n} - \mathbf{M}_0)'\mathbf{W}_{\mathrm{u},n}(\mathbf{M}_{\mathrm{u},n} - \mathbf{M}_0)
\end{aligned} \tag{WA12}$$

The structure of the distribution suggests a two-stage sampling of $(\mathbf{B}_{\mathrm{u}}, \mathbf{\Sigma}_{\mathrm{u}})$: (1) sample $\mathbf{\Sigma}_{\mathrm{u}}$ according to the inverse-Wishart distribution; (2) given the realization of $\mathbf{\Sigma}_{\mathrm{u}}$ sample $\mathbf{B}_{\mathrm{u}}$ according to the matrix-normal distribution. The sampling distribution only depends upon the segment embedding $\mathbf{U}$, which contains all of the information about $(\mathbf{B}_{\mathrm{u}}, \mathbf{\Sigma}_{\mathrm{u}})$ (see Figure 1).

We can derive the distribution for the parameters $(\mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}})$ by symmetry:

$$\begin{aligned}
\mathbb{P}(\mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}} \mid \mathbf{V}, \mathbf{V}, \alpha, \mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}}, \mathcal{D}) &\propto \mathbb{P}(\mathbf{V} \mid \mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}}, \mathcal{D})\mathbb{P}(\mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}}) \\
\propto \left[\prod_{\mathrm{s}} \mathcal{N}(\mathbf{v}_{\mathrm{c}} \mid \mathbf{B}_{\mathrm{v}}\mathbf{z}_{\mathrm{c}}, \mathbf{\Sigma}_{\mathrm{v}})\right] &\mathcal{MN}\left(\mathbf{B}_{\mathrm{v}} | \mathbf{M}_{\mathrm{v},0}, \mathbf{W}_{\mathrm{v},0}^{-1}, \mathbf{\Sigma}_{\mathrm{v}}\right) \mathcal{W}^{-1}(\mathbf{\Sigma}_{\mathrm{v}} | \mathbf{\Lambda_0}, \nu_0) \\
&= \mathcal{MN}\left(\mathbf{B}_{\mathrm{v}} | \mathbf{M}_{\mathrm{v},n}, \mathbf{W}_{\mathrm{v},n}^{-1}, \mathbf{\Sigma}_{\mathrm{v}}\right) \mathcal{W}^{-1}(\mathbf{\Sigma}_{\mathrm{v}} | \mathbf{\Lambda}_{\mathrm{v,n}}, \nu_{\mathrm{v,n}})
\end{aligned} \tag{WA13}$$

where the parameters of the posterior distribution can be expressed analytically:

$$\begin{aligned}
&\nu_{\mathrm{v,n}} = \nu_0 + \mathrm{C} \\
&\mathbf{W}_{\mathrm{v},n} = \mathbf{Z}'\mathbf{Z} + \mathbf{M}_{\mathrm{v},0} \\
&\mathbf{M}_{\mathrm{v},n} = \mathbf{W}_{\mathrm{v},n}^{-1}\left(\mathbf{Z}'\mathbf{V} + \mathbf{W}_{\mathrm{v},n}\mathbf{M}_0\right) \\
&\mathbf{\Lambda}_{\mathrm{v,n}} = \mathbf{\Lambda_0} + (\mathbf{V} - \mathbf{Z}\mathbf{M}_{\mathrm{v},n})'(\mathbf{V} - \mathbf{Z}\mathbf{M}_{\mathrm{v},n}) + (\mathbf{M}_{\mathrm{v},n} - \mathbf{M}_0)'\mathbf{W}_{\mathrm{v},n}(\mathbf{M}_{\mathrm{v},n} - \mathbf{M}_0)
\end{aligned} \tag{WA14}$$

Conditional Distribution for α

Lastly, we derive the conditional distribution for α:

$$\mathbb{P}(\alpha | \mathbf{U}, \mathbf{V}, \mathbf{B}_{\mathrm{u}}, \mathbf{\Sigma}_{\mathrm{u}}, \mathbf{B}_{\mathrm{v}}, \mathbf{\Sigma}_{\mathrm{v}}, \mathcal{D}) \propto \left[\prod_{\mathrm{c,s}} \mathcal{N}\left(\mathrm{y}_{\mathrm{cs}} | \mathbf{u}_{\mathrm{s}}'\mathbf{v}_{\mathrm{c}}, \alpha^{-1} + \lambda_{\mathrm{cs}}^{-1}\right)^{\mathrm{q_{cs}}}\right] \mathcal{G}(\alpha | \mathrm{a}_0, \mathrm{b}_0) \tag{WA15}$$

Equation (WA15) does not support a closed-form expression for a posterior distribution. Moreover, we could not identify a conjugate prior for α that would absorb variation in λ_{cs} across the campaign-segment combinations to ensure an analytical form for the posterior distribution. We thus rely on the Markov Chain Monte Carlo (MCMC) approach to draw

samples from the conditional distribution $\mathbb{P}(\boldsymbol{\alpha}|\mathbf{U}, \mathbf{V}, \mathbf{B}_{\mathrm{u}}, \boldsymbol{\Sigma}_{\mathrm{u}}, \mathbf{B}_{\mathrm{v}}, \boldsymbol{\Sigma}_{\mathrm{v}}, \mathbf{Y})$.

We can improve the MCMC sampling procedure for Equation (WA15) by incorporating the derivative of the log-likelihood function. In particular, the logarithm of Equation (WA15) and its derivative can be expressed as follows:

$$\log \mathbb{P}(\boldsymbol{\alpha}|\mathbf{U}, \mathbf{V}, \mathbf{B}_{\mathrm{u}}, \boldsymbol{\Sigma}_{\mathrm{u}}, \mathbf{B}_{\mathrm{v}}, \boldsymbol{\Sigma}_{\mathrm{v}}, \mathcal{D}) = -\sum_{\mathrm{c,s}} \frac{\mathrm{q_{cs}}}{2} \left[\frac{(\mathrm{y_{cs}} - \mathbf{u}_{\mathrm{s}}' \mathbf{v}_{\mathrm{c}})^2}{\lambda_{\mathrm{cs}}^{-1} + \alpha^{-1}} + \log\left(\lambda_{\mathrm{cs}}^{-1} + \alpha^{-1}\right) \right] \tag{WA16}$$
$$\log \mathrm{Z} + \log \mathcal{G}(\boldsymbol{\alpha}|\mathrm{a_0}, \mathrm{b_0})$$

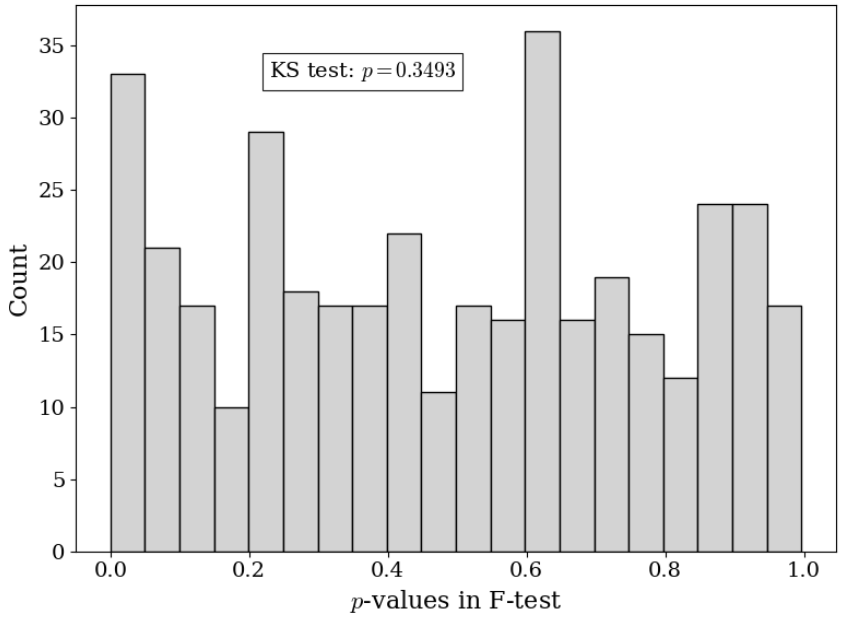
$$\frac{\partial \log \mathbb{P}(\boldsymbol{\alpha}|\mathbf{U}, \mathbf{V}, \mathbf{B}_{\mathrm{u}}, \boldsymbol{\Sigma}_{\mathrm{u}}, \mathbf{B}_{\mathrm{v}}, \boldsymbol{\Sigma}_{\mathrm{v}}, \mathcal{D})}{\partial \boldsymbol{\alpha}} = -\sum_{\mathrm{c,s}} \frac{\mathrm{q_{cs}}}{2} \left[\frac{(\mathrm{y_{cs}} - \mathbf{u}_{\mathrm{s}}' \mathbf{v}_{\mathrm{c}})^2}{(\alpha/\lambda_{\mathrm{cs}} + 1)^2} - \frac{1}{\alpha\,(\alpha/\lambda_{\mathrm{cs}} + 1)} \right] \tag{WA17}$$
$$+ \frac{\mathrm{a_0} - 1}{\alpha} - \mathrm{b_0}$$

The derivative in Equation (WA17) does not include the normalization constant Z, which allows us to use the Hamiltonian Monte Carlo (HMC) sampling approach. Similarly to the Metropolis-Hastings algorithm, the HMC generates samples from a target distribution, $\mathbb{P}(\mathrm{x}) \sim \pi(\mathrm{x})/\mathrm{Z}$, by first sampling random variables from a simpler distribution, $\mathrm{q}(\mathrm{x}^{\mathrm{t}}|\mathrm{x}^{\mathrm{t}-1})$, and then accepting (or rejecting) new samples using a stochastic rule. The stochastic rule is defined so that over time, a sequence of accepted samples closely approximates the target distribution $\mathbb{P}(\mathrm{x})$. HMC improves on the Metropolis-Hastings algorithm by utilizing the derivative of the log-likelihood function. This allows for more efficient exploration of the parameter space, and leads to faster convergence (burn-in) and better approximation of the target distribution.

Web Appendix WA3: Randomization Test

We conduct randomization checks using an F-test approach. Specifically, for each marketing campaign, we estimate a linear regression where the binary treatment indicator is regressed on pretreatment covariates (See Web Appendix WA4). We estimate a separate regression for each campaign, record the p-value of the F-test in each regression, and then summarize the distribution in Figure WA1. A Kolmogorov-Smirnov (KS) test does not reject the null hypothesis that the p-values are uniformly distributed; the p-value of the KS test is 0.35. This suggests that the treatment assignment is independent of pretreatment variables, supporting the validity of the randomization.

Figure WA1: Distribution of p-values in the Randomization Test



Notes: The figure reports histograms of the distribution of p-values in the randomization test for each of 391 campaigns. The y-axis represents the count of campaigns, and the x-axis represents p-values from the F-test with five pretreatment variables.

Web Appendix WA4: Targeting Variables

We use individual covariates described in Table WA1 to estimate machine-learning baselines. The covariates are (customer×campaign)-specific. For the Transfer (Proposed) model, we aggregate these covariates by averaging across customers within each segment. Furthermore, we reduce the covariates in Section 6 to four variables: *Monetary Value* (159 weeks), *Order Frequency* (159 weeks), *Recency*, and *Never Purchased*. This improves model convergence with fewer source campaigns. Table WA2 confirms that the Transfer model with reduced covariates performs on par with the full-covariates specification.

Table WA1: Summary of Individual Customer Covariates

Variable	**Description**
Monetary Value	Cumulative spending (USD) in the week preceding the campaign in-home date. Same for 2, 4, 8, 13, 26, 53, 106, and 159 weeks.
Order Frequency	Number of transactions in the week preceding the campaign in-home date. Same for 2, 4, 8, 13, 26, 53, 106, and 159 weeks.
Recency	Number of days between the customer's most recent purchase relative to the campaign in-home date.
Tenure	Number of days between the initial transaction and the campaign in-home date.
Return Rate	The proportion of total units returned relative to total units purchased over the customer's lifetime prior to the campaign.
Never Purchased	A binary indicator for customers with no transaction history prior to the current campaign. This happens when the customer purchased the product before the start of our observational period.

Table WA2: Proposed Model with Full and Reduced Covariates

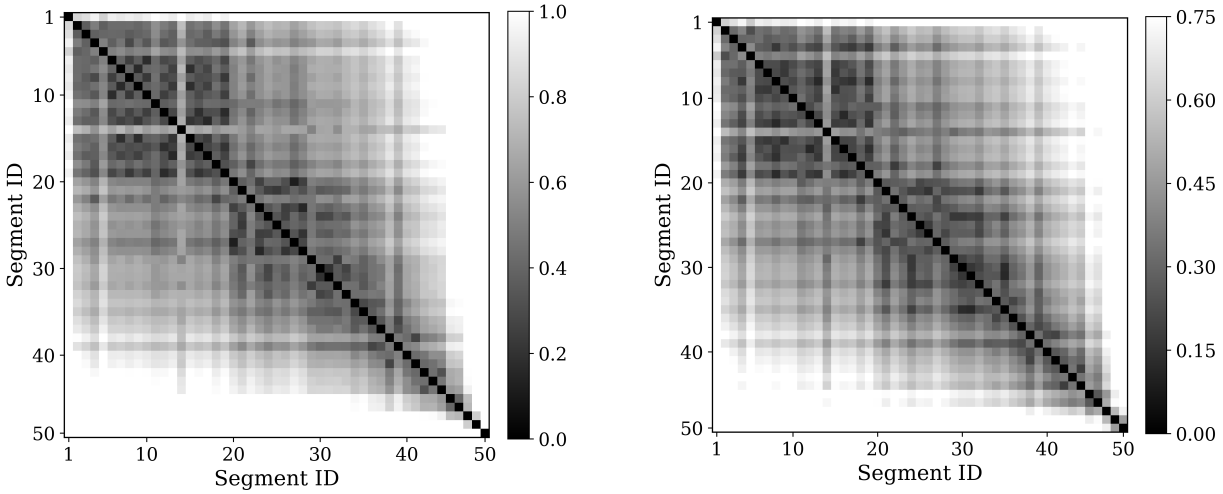
	Same Period		**Next Year**	
Method	**Qini**	**Profit**	**Qini**	**Profit**
Transfer: Full Covariates	9.83	21.96¢	8.75	20.03¢
	(0.64)	(1.96)	(0.83)	(2.45)
Transfer: Reduced Covariates	9.93	21.85¢	9.15	20.29¢
	(0.62)	(1.96)	(0.84)	(2.48)

Notes: The table compares the performance of the proposed model with full and reduced covariates using the Qini and *Profit* metrics. The **Same Period** comparison compares findings across 361 campaigns, and the **Next Year** comparison includes 198 campaigns. Standard errors are reported in parentheses.

Web Appendix WA5: Segment and Campaign Embeddings

In this section, we evaluate the internal validity of our proposed approach by examining the segment and campaign embeddings. To obtain the embeddings, we estimate the model using all available campaigns. Figure WA2 illustrates the Euclidean distance between the segment embeddings. The segments are ordered by their historical 3-year spending, ranging from no purchases to over $3,000. We observe a block-diagonal pattern, indicating that segments with similar past spending are located closer to each other in the embedding space.

Figure WA2: Euclidean Distance between Segment Embeddings



(a) Full Model **(b)** No Covariates

Notes: The figures report euclidean distances between the embeddings of customer segments. The segments are ordered according to the 3-year past spending, where the first segment includes customers with no purchases and the last segment (50) includes customers who spent over $3,000.

Table WA3 evaluates the campaign embeddings by comparing the characteristics of a focal campaign's nearest neighbors in the embedding space against a random selection of campaigns. The results show that a campaign's nearest neighbors are substantially more similar in terms of campaign type, seasonality, and recency than the random baseline. Notably, these patterns hold both when the model is estimated with the full set of covariates and when it is estimated without any covariates. This confirms that the model infers the underlying segment and campaign characteristics directly from observed customer responsiveness, even when explicit covariates are entirely omitted from model training.

Table WA3: Comparison of Random and Nearest Campaigns

	Random Campaigns	**Nearest Campaigns**	
		Full Model	**No Covariates**
Match on Campaign Type	57.19%	88.13%	69.82%
Match on Seasonality	26.14%	42.71%	31.66%
Match on Campaign Name	2.76%	10.13%	6.14%
Conducted Within 2 Years	39.44%	49.67%	43.63%

Notes: This table compares the characteristics of the five random campaigns against the five nearest campaigns in the embedding space.

Web Appendix WA6: Seasonality

Table WA4: Evaluating the Role of Seasonality: Qini

Focal	**#**	**Source Campaign**					**Focal Only**
Campaign	**Focals**	**Q1**	**Q2**	**Q3**	**Q4**	**Random**	**Baseline**
Q1	44	**9.85** (1.58)	9.65 (1.53)	9.22 (1.67)	8.86 (1.67)	9.58 (1.59)	4.34 (1.23)
Q2	26	10.57 (3.20)	**10.73** (3.14)	9.90 (3.19)	9.46 (3.28)	10.29 (3.19)	6.43 (3.07)
Q3	44	9.06 (1.88)	9.18 (1.84)	**9.56** (1.84)	8.43 (1.94)	9.28 (1.84)	3.99 (1.76)
Q4	70	8.73 (1.63)	8.14 (1.69)	8.69 (1.63)	**8.34** (1.63)	8.93 (1.61)	3.82 (1.49)

Notes: The table reports the Qini metric averaged across **# Focals** campaigns within each row. Standard errors are reported in parentheses.

Table WA5: Evaluating the Role of Seasonality: Profit

Focal	**#**	**Source Campaign**					**Focal Only**
Campaign	**Focals**	**Q1**	**Q2**	**Q3**	**Q4**	**Random**	**Baseline**
Q1	44	**20.31¢** (4.32)	19.85¢ (4.34)	19.65¢ (4.38)	18.92¢ (4.43)	20.14¢ (4.35)	11.59¢ (4.13)
Q2	26	30.79¢ (10.25)	**30.79¢** (10.33)	29.86¢ (10.36)	29.23¢ (10.36)	30.22¢ (10.34)	23.40¢ (10.29)
Q3	44	23.80¢ (6.52)	23.77¢ (6.46)	**24.03¢** (6.44)	23.12¢ (6.58)	23.70¢ (6.52)	17.68¢ (6.08)
Q4	70	24.95¢ (5.51)	24.03¢ (5.58)	25.22¢ (5.53)	**24.53¢** (5.53)	25.15¢ (5.50)	19.22¢ (4.86)

Notes: The table reports the *Profit* metric averaged across **# Focals** campaigns within each row. Standard errors are reported in parentheses.

Web Appendix WA7: Campaign Type: Incrementality vs. Promotions

Table WA6: Source Campaign Type: Qini

Focal Campaign	**# Focals**	**Source Campaign**			**Focal Only**
		Incr	**Promo**	**Random**	**Baseline**
Incrementality	235	**12.02**	11.27	11.94	5.98
		(0.85)	(0.88)	(0.85)	(0.81)
Promotion	82	3.86	**4.02**	4.06	1.26
		(0.85)	(0.83)	(0.82)	(0.64)

Notes: The table reports the Qini metric averaged across **# Focals** campaigns within each row. Standard errors are reported in parentheses.

Table WA7: Source Campaign Type: Profit

Focal Campaign	**# Focals**	**Source Campaign**			**Focal Only**
		Incr	**Promo**	**Random**	**Baseline**
Incrementality	235	**24.22¢**	23.70¢	24.17¢	17.68¢
		(2.87)	(2.86)	(2.87)	(2.63)
Promotion	82	20.68¢	**21.11¢**	21.06¢	15.75¢
		(2.23)	(2.23)	(2.23)	(2.02)

Notes: The table reports *Profit* measures averaged across **# Focals** campaigns within each row. Standard errors are reported in parentheses.