# Causal Inference under Threshold Manipulation: A Bayesian Mixture Approach

Kohsuke Kubota
NTT DOCOMO, INC.
Tokyo, Japan
kousuke.kubota.xt@nttdocomo.com

Shonosuke Sugasawa
Keio University
Tokyo, Japan
sugasawa@econ.keio.ac.jp

## Abstract

Many marketing applications, including credit card incentive programs, offer rewards to customers exceeding specific spending thresholds to encourage increased consumption. Quantifying the causal effect of these thresholds on customers is crucial for effective marketing strategy design. While regression discontinuity design is a common method for such causal inference tasks, its assumptions can be violated when customers, aware of the thresholds, strategically manipulate their spending to qualify for the rewards. To address this issue, we propose a novel framework for estimating the causal effect of thresholds on customers under their manipulation. The core idea is to model the observed spending distribution as a mixture of two distributions: one representing customers strategically affected by the threshold and the other representing those unaffected. To fit the mixture model, we adopt a Bayesian approach, which enables valid causal effect estimation with proper uncertainty quantification. Furthermore, we extend this framework to a hierarchical Bayesian setting to estimate heterogeneous causal effects across customer subgroups, allowing for stable inference even with small subgroup sample sizes. We demonstrate the effectiveness of our proposed methods through simulation studies and show that our proposed framework yields more accurate estimates of the causal effect of thresholds on customers compared to naive regression discontinuity design methods.

## CCS Concepts

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

## Keywords

Causal Inference, Bunching Estimation, Bayesian Estimation, Markov Chain Monte Carlo

**ACM Reference Format:**
Kohsuke Kubota and Shonosuke Sugasawa. 2025. Causal Inference under Threshold Manipulation: A Bayesian Mixture Approach . In *Proceedings of KDD '25*. ACM, Toronto, Canada, 7 pages. https://doi.org/XXXXXXX.XXXXXXX


*KDD '25, Toronto, Canada*

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

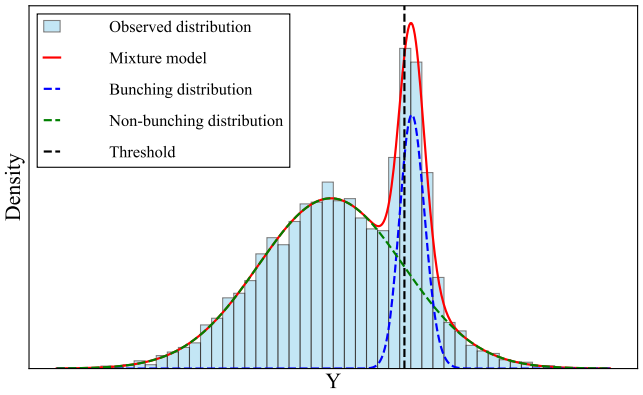


**Figure 1: Conceptual illustration of our proposed mixture model (red). Modeling the observed distribution (skyblue) as a mixture of a non-bunching distribution (green, unaffected by the threshold) and a strategic bunching distribution (blue, distorted near the threshold).**

## 1 Introduction

Estimating causal effects of thresholds is important for effective marketing strategy design. Many marketing applications, including loyalty programs [12, 16] and credit card incentive programs [1], offer rewards to customers who exceed specific spending thresholds to encourage increased consumption. These thresholds are often set to incentivize customers to spend more, and understanding their causal effects on customer behavior is crucial to optimizing marketing strategies. To estimate the causal effect of thresholds on customer behavior, regression discontinuity design (RDD) is a representative method [8, 23]. RDD can estimate the causal effect of a threshold by using local randomization around the threshold, which means that the treatment assignment for customers around the threshold is assumed to be random. Many studies have applied RDD because this method is "one of the most credible research designs in the social, behavioral, biomedical and statistical sciences for program evaluation and causal inference in the absence of an experimentally assigned treatment" [5] and is "the perception that the identification assumptions are quite weak" [15].

However, the assumptions of RDD can be violated when customers, aware of the thresholds, strategically manipulate their

[1] https://creditcards.chase.com/rewards-credit-cards/sapphire/preferred

spending to qualify for the rewards [10, 14, 15]. For example, incentive programs where rewards are offered to customers who exceed a spending threshold are likely to lead to strategic behavior. The assumption of local randomization is supported by the *continuity* condition [8], which states that the potential outcomes of the customers just below and above the threshold are the same. This condition cannot be satisfied when customers can manipulate their behavior to reach the thresholds because they may have different potential outcomes depending on whether they are above or below the threshold [10, 14, 15]. To address the manipulation problem, an alternative causal inference framework called bunching estimation [3] is advocated in the economics literature, by assuming a discontinuous change in the slope of the function at the threshold. However, such an approach relies on the assumption that a segment of the distribution just before the threshold, where the probability density function becomes zero, is what gets transformed into the observed "bunch" at the threshold. Consequently, the practical applicability of these methods can be limited in scenarios where such a zero-density region cannot be assumed, which would often arise in marketing applications.

To overcome the limitations of causal inference under manipulation, we propose a novel framework for estimating the causal effect of thresholds on customers under their manipulation. A key feature of the proposed method is to employ a mixture of two distributions: one representing customers strategically affected by the threshold and the other representing those unaffected (see Figure 1). We adopt a Bayesian approach to fit the mixture model, which enables valid causal effect estimation with proper uncertainty quantification. Hence, we call the proposal *Bayesian Modeling of Threshold Manipulation via Mixtures* (BMTM) in what folllows. This approach allows us to successfully capture the treated and untreated populations without relying on implausible assumptions. Furthermore, we can natually extend BMTM to a hierarchical Bayesian setting to estimate heterogeneous causal effects across customer subgroups, which we call hierarchical BMTM (HBMTM). Although the demand for pursuing heterogeneous causal effects is increasing in marketing, simply applying the proposed method to subgroups separately may lead to unstable estimates when the sample sizes within subgroups are small. By adopting a fully Bayesian framework, we can naturally introduce a hierarchical structure that borrows information across subgroups, enabling stable inference for heterogeneous causal effects.

We demonstrate the effectiveness of our proposed framework through simulation studies. These studies show that our proposed methods achieve more accurate and reliable estimates of causal effects than the RDD methods. In particular, HBMTM provides stable estimates of heterogeneous causal effects across customer subgroups in scenarios with small subgroup sample sizes.

## 2 Related Work

In this section, we review related work on bunching estimation and methods to estimate heterogeneous causal effects using RDD. The bunching estimation examines the discontinuity in the distribution of the running variable. In the literature, there are two types of discontinuities: kinks, where the slope of an incentive schedule changes [e.g. 19] and notches, where the level of the incentive changes discontinuously [e.g. 20]. Marketing incentive programs, which offer a discrete reward upon reaching a threshold, typically fall into the notch category. In addition to the literature in Section 1, the existing bunching estimation methods [e.g. 13] assume sharp bunching, where individuals can precisely target and locate themselves at the threshold. However, in real-world marketing scenarios, this assumption is often too strong because customers may not be fully aware of the exact threshold value or may face constraints and imperfect control over their spending behavior. In contrast to existing approaches, the proposed method provides a simple method for estimating the causal effect under plausible assumptions in marketing applications.

In RDD (without threshold manipulation), there are some methods to estimate the heterogeneous treatment effect between subgroups. For example, Sugasawa et al. [21] developed a hierarchical model to estimate the subgroup-specific causal effect, and Alcantara et al. [1] and Tao et al. [22] employed the Bayesian additive regression tree to capture heterogeneous effects. However, these methods cannot be applied to the situation under threshold manipulation. To our knowledge, no attempts have been made to estimate the heterogeneous causal effects under threshold manipulation. Hence, this paper would be the first to provide an effective method to estimate heterogeneous causal effects under threshold manipulation.

## 3 Causal Inference under Threshold Manipulation

Consider the problem of estimating the causal effect of a threshold $K$ on the behavior of customer spending with the potential outcome framework [9, 18]. Let $Y(1)$ be the potential outcome with threshold $K$ and $Y(0)$ be the potential outcome without threshold. Assume that there are two types of customers: those who strategically exceed a specific threshold $K$ (*i.e.,* bunching customers) and those who do not exceed the threshold (*i.e.,* non-bunching customers). The bunching customers are those who, in the presence of a threshold, can exceed that threshold and, by doing so, obtain greater benefits than in a situation without any threshold. These customers are likely to increase their spending to exceed the threshold or adjust their spending just to exceed the threshold to maximize their benefits. Conversely, non-bunching customers are those who, in the presence of a threshold, either cannot exceed it or cannot gain any additional benefit by exceeding it compared to a situation without any threshold. Among the non-bunching customers are customers for whom the "pain" (or cost) of increasing spending to surpass the threshold exceeds the utility gained from receiving the incentive, and customers who would already surpass the threshold without any threshold. Consequently, it is assumed that the bunching customers are those whose potential outcome without the threshold, $Y(t)$ (for $t \in \{0, 1\}$), falls within the neighborhood of the threshold $K$. Non-bunching customers, in contrast, are those whose $Y(t)$ lies outside this neighborhood.

In the framework, the average treatment effect of the threshold on customers is defined as the difference in their potential outcomes with and without the threshold as follows.

$$\tau \coloneqq \mathbb{E}\left[Y(1) - Y(0) \mid \underline{Y}^I \leq Y(t) \leq \overline{Y}^I,\ t \in \{0, 1\}\right] \tag{1}$$

where the interval $[\underline{Y}^I, \overline{Y}^I]$ defines the neighborhood of the threshold $K$. This definition can be interpreted as the average treatment effect of the threshold on bunching customers, which is conceptually similar to the average treatment effect on the treated (ATT) in the potential outcome framework.

Furthermore, to account for heterogeneity between customer subgroups, this definition can easily be extended to estimate subgroup-specific causal effects. Let customers be divided into $G$ subgroups and let $Y_g(1)$ and $Y_g(0)$ be the potential outcomes with and without the threshold for the $g$th subgroup, respectively. For the $i$th customer in the $g$th subgroup, where $i = 1, \ldots, n_g$ and $g = 1, \ldots, G$, the heterogeneous causal effect of the threshold on customers is defined as follows:

$$\tau_g \coloneqq \mathbb{E}\left[Y_g(1) - Y_g(0) \mid \underline{Y_g}^I \le Y(t) \le \overline{Y_g}^I,\ t \in \{0,1\}\right] \tag{2}$$

where the interval $[\underline{Y_g}^I, \overline{Y_g}^I]$ defines the neighborhood of the threshold $K$ for the $g$th subgroup. This definition allows us to estimate the heterogeneous causal effect of a threshold on customers in each subgroup, which is important for understanding how different subgroups of customers respond to the threshold.

# 4 Bayesian Modeling of Threshold Manipulation via Mixtures

## 4.1 Mixture Model for Bunching and Non-bunching Distributions

To estimate the causal effect of a threshold on bunching customers $\tau$, the observed data $y_i$ for each customer $i = 1, \ldots, n$ are assumed to be independent and identically distributed samples from a mixture distribution. The mixture distribution is assumed to be a two-component mixture distribution representing non-bunching customers (called non-bunching distribution) and bunching customers (called bunching distribution). The non-bunching distribution is assumed to be the counterfactual distribution of the observed data $y_i$ that would have been observed had the threshold not existed. Since the bunching distribution consists of bunching customers, it is assumed that the observed data $y_i$ of the bunching distribution clusters are just above the threshold $K$.

To model the observed data, we adopt a mixture model that captures the non-bunching and bunching distributions. The non-bunching distribution is modeled using a normal distribution, a common choice for modeling continuous data. The normal distribution is characterized by its mean and standard deviation, which can be estimated from the observed data. The bunching distribution is modeled using a skew normal distribution [2], which is a flexible distribution that can capture the skewness and high concentration of data observed near the threshold $K$ (fixed) resulting from customer manipulation.

For the observed data $y_i$, we assume the following mixture model:

$$y_i \sim \pi \cdot f_{SN}(y_i \mid K, \sigma_1, \alpha_1) + (1-\pi) \cdot f_N(y_i \mid \mu_2, \sigma_2), \tag{3}$$

where $f_{SN}$ is the probability density function of the skew normal distribution, and $f_N$ is the probability density function of the normal distribution. The parameters of the skew normal distribution are $K$, $\sigma_1$, and $\alpha_1$, where $K$ is the fixed threshold, $\sigma_1$ is the scale parameter, and $\alpha_1$ is the shape parameter. The parameters of the normal distribution are $\mu_2$ and $\sigma_2$, where $\mu_2$ is the mean and $\sigma_2$ is the standard deviation. In the model (3), $\pi \in (0,1)$ is the mixing proportion, which represents the proportion of bunching customers in the observed data. Note that the unknown parameters in the model (3) are $\Theta = (\sigma_1, \alpha_1, \mu_2, \sigma_2, \pi)$.

By introducing a binary latent variable $z_i \in \{0,1\}$, the mixture model (3) can also be expressed as

$$y_i|(z_i = 1) \sim f_{SN}(y_i \mid K, \sigma_1, \alpha_1), \quad y_i|(z_i = 0) \sim f_N(y_i \mid \mu_2, \sigma_2)$$

with $P(z_i = 1) = 1 - P(z_i = 0) = \pi$. Here $z_i$ can be interpreted as a latent (unobserved) binary indicator representing that the $i$th customer is included in the buncing ($z_i = 1$) and non-bunching ($z_i = 0$) groups. Hence, the observed data $y_i$ can be regarded as $Y_i(z_i)$, where $Y_i(1)$ and $Y_i(0)$ are potential outcomes of the $i$th customer. This formulation implies that either of the potential outcomes can be observed, but we do not know which outcome is observed, since the latent group membership $z_i$ is not observed. This contrasts with the standard potential outcome framework in causal inference, where the treatment assignment is observed.

## 4.2 Average Treatment Effect and Bayesian Inference

To obtain the average treatment effect $\tau$ defined in (1), we need to consider the interval $[\underline{Y}^I, \overline{Y}^I]$ as the neighborhood of the threshold $K$. Since we can identify the bunching distribution $f_{SN}(y_i \mid K, \sigma_1, \alpha_1)$, a reasonable choice would be the interval $[\underline{Y}^I_c(\sigma_1, \alpha_1), \overline{Y}^I_c(\sigma_1, \alpha_1)]$ determined by the $(1-c)$% highest density interval (HDI) of the bunching distribution for arbitrary small $c$ (e.g. $c = 0.01$). Note that the interval is a function of the unknown parameters $(\sigma_1, \alpha_1)$, in the bunching distribution. Then, the expectation of treated potential outcome is

$$\begin{aligned}&\mathbb{E}\left[Y(1) \mid \underline{Y}^I_c(\sigma_1, \alpha_1) \le Y(1) \le \overline{Y}^I_c(\sigma_1, \alpha_1)\right] \\ &= (1-c)^{-1}\left(K + \sigma_1 \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \cdot \sqrt{\frac{2}{\pi}}\right),\end{aligned} \tag{4}$$

using the expectation of the skew normal distribution [2] and that fact that $\mathbb{P}(\underline{Y}^I_c(\sigma_1, \alpha_1) \le Y(1) \le \overline{Y}^I_c(\sigma_1, \alpha_1)) = 1-c$ from the definition of the interval. On the other hand, the expectation of the non-bunching distribution in the interval $[\underline{Y}^I, \overline{Y}^I]$ (i.e., the expectation of the untreated potential outcome) can also be obtained in the close form as a function of $(\mu_2, \sigma_2)$. Therefore, under the mixture model (3), the average treatment effect (1) can be expressed as

$$\begin{aligned}\tau(\Theta) \equiv (1-c)^{-1}&\left(K + \sigma_1 \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \cdot \sqrt{\frac{2}{\pi}}\right) - \mu_2 \\ &- \sigma_2 \frac{\phi(\underline{Y}^I_c(\sigma_1, \alpha_1); \mu_2, \sigma_2) - \phi(\overline{Y}^I_c(\sigma_1, \alpha_1); \mu_2, \sigma_2)}{\Phi(\overline{Y}^I_c(\sigma_1, \alpha_1); \mu_2, \sigma_2) - \Phi(\underline{Y}^I_c(\sigma_1, \alpha_1); \mu_2, \sigma_2)},\end{aligned} \tag{5}$$

where $\phi(\cdot; a, b)$ and $\Phi(\cdot; a, b)$ are the probability density and cumulative distribution function of the normal distributions with mean $a$ and variance $b$, respectively.

For fitting the mixture model (3), we adopt a Bayesian approach to estimate the parameters of the model. Specifically, we

employ the prior distributions $\sigma_1 \sim N(m_{\sigma_1}, s^2_{\sigma_1})$, $\alpha_1 \sim N(m_{\alpha_1}, s^2_{\alpha_1})$, $\mu_2 \sim N(m_{\mu_2}, s^2_{\mu_2})$, $\sigma_2 \sim N(m_{\sigma_2}, s^2_{\sigma_2})$, $\pi \sim \text{Beta}(\alpha_\pi, \beta_\pi)$, where $m_{\sigma_1}$, $s^2_{\sigma_1}$, $m_{\alpha_1}$, $s^2_{\alpha_1}$, $m_{\mu_2}$, $s^2_{\mu_2}$, $m_{\sigma_2}$, $s^2_{\sigma_2}$, $\alpha_\pi$, and $\beta_\pi$ are hyperparameters specified by a user. To effectively sample from the posterior distribution of the parameters $\Theta$, we use the Markov Chain Monte Carlo (MCMC) algorithm. For MCMC sampling, we use the MCMC algorithm implemented in the probabilistic programming language `stan` [4]. Stan is a probabilistic programming language widely used for Bayesian modeling, which takes advantage of the Hamiltonian Monte Carlo algorithm [6] to efficiently sample complex probability distributions. Based on the posterior samples of $\Theta$, one can approximate the posterior distribution of the causal effect $\tau(\Theta)$, which gives not only a point estimate but also a measure of uncertainty such as the credible interval 95%.

In this context, there are three advantages to using the Bayesian approach for mixture models. First, the Bayesian approach can incorporate prior information about the parameters. In the mixture model (3), the bunching distribution is expected to be concentrated around the threshold $K$, and the normal distribution is expected to be more spread. We can incorporate this prior information into the model by specifying appropriate prior distributions for the parameters. Second, the Bayesian approach can provide a natural way to quantify uncertainty in the parameter estimates and their functions. Although the causal effect $\tau(\Theta)$ given in (5) is a rather complicated function of $\Theta$, its uncertainty can be easily quantified through posterior distributions of $\Theta$, which will be useful for making more informed decisions based on the estimates. Finally, the Bayesian approach can easily be extended to hierarchical models, which allows us to estimate heterogeneous causal effects across customer subgroups $\tau_g$, as discussed in the subsequent section.

## 4.3 Pursuing heterogeneous treatment effect via hierarchical Bayesian modeling

The proposed method can be extended to a hierarchical Bayesian setting to estimate heterogeneous causal effects between subgroups of customers. In what follows, we assume that the threshold is common to all subgroups and is $K$. For $g = 1, \ldots, G$ with $G$ being the number of subgroups, the mixture model for the $g$th subgroup can be defined as follows:

$$y_i^{(g)} \sim \pi_g \cdot f_{SN}(y_i \mid K, \sigma_1^{(g)}, \alpha_1^{(g)}) + (1 - \pi_g) \cdot f_N(y_i \mid \mu_2^{(g)}, \sigma_2^{(g)}), \quad (6)$$

where $y_i^{(g)}$ is the observed data for the $i$th customer in the $g$th subgroup, and $\pi_g$ is the mixing proportion for the $g$th subgroup. Here $\sigma_1^{(g)}$, $\alpha_1^{(g)}$, $\mu_2^{(g)}$ and $\sigma_2^{(g)}$ are unknown subgroup-specific parameters. The mixing proportion $\pi_g$ represents the proportion of bunching customers in the observed data for the $g$th subgroup. The causal effect of the threshold on customers $\tau_g$ can be defined in the same way as (5) as a function of subgroup-specific parameters, $\Theta_g = (\sigma_1^{(g)}, \alpha_1^{(g)}, \mu_2^{(g)}, \sigma_2^{(g)}, \pi_g)$.

The subgroup-specific parameters are hierarchically modeled to borrow information from subgroups for more stable estimates. A critical aspect in specifying such hierarchical models, particularly for Bayesian estimation using MCMC algorithms, is the choice between centered and non-centered parameterizations [e.g. 17, 24]. To enhance sampling efficiency and mitigate potential strong correlations between hierarchical levels of parameters, especially when group variances are small or data per group are limited, we employ a non-centered parameterization for several of these group-specific parameters. The detailed specifications for these parameters are as follows.

$$\sigma_1^{(g)} = \exp(\mu_{\sigma_1} + \sigma_{\sigma_1} \cdot z_{\sigma_1}^{(g)}), \quad \alpha_1^{(g)} = \mu_{\alpha_1} + \sigma_{\alpha_1} \cdot z_{\alpha_1}^{(g)},$$
$$\sigma_2^{(g)} = \exp(\mu_{\sigma_2} + \sigma_{\sigma_2} \cdot z_{\sigma_2}^{(g)}), \quad \mu_2^{(g)} = \mu_{\mu_2} + \sigma_{\mu_2} \cdot z_{\mu_2}^{(g)},$$
$$\pi_g = \text{logit}^{-1}(\mu_\pi + \sigma_\pi \cdot z_\pi^{(g)}),$$

where $z_{\sigma_1}^{(g)}, z_{\alpha_1}^{(g)}, z_{\mu_2}^{(g)}, z_{\sigma_2}^{(g)}, z_\pi^{(g)} \sim N(0, 1)$. The above formulation indicates that the group-specific parameters are different but generated from a common distribution, known as random effects. For unknown parameters of the mixture model, we assign prior distributions as follows:

$$\mu_{\sigma_1} \sim N(m_{\mu\sigma_1}, s^2_{\mu\sigma_1}), \quad \sigma_{\sigma_1} \sim N(m_{\sigma\sigma_1}, s^2_{\sigma\sigma_1}), \quad \mu_{\alpha_1} \sim N(m_{\mu\alpha_1}, s^2_{\mu\alpha_1}),$$
$$\sigma_{\alpha_1} \sim N(m_{\sigma\alpha_1}, s^2_{\sigma\alpha_1}), \quad \mu_{\sigma_2} \sim N(m_{\mu\sigma_2}, s^2_{\mu\sigma_2}), \quad \sigma_{\sigma_2} \sim N(m_{\sigma\sigma_2}, s^2_{\sigma\sigma_2}),$$
$$\mu_{\mu_2} \sim N(m_{\mu\mu_2}, s^2_{\mu\mu_2}), \quad \sigma_{\mu_2} \sim N(m_{\sigma\mu_2}, \quad s^2_{\sigma\mu_2}), \quad \mu_\pi \sim N(m_{\mu\pi}, s^2_{\mu\pi}),$$
$$\sigma_\pi \sim N(m_{\sigma\pi}, s^2_{\sigma\pi}).$$

Given the prior distributions, the posterior distribution of the group-specific parameters $\Theta_g$ as well as the parameters in the random-effects distributions can be approximated by an MCMC algorithm. Then, the posterior samples of $\Theta_g$ give the posterior distribution of the subgroup-specific causal effect $\tau(\Theta_g)$.

## 4.4 Isolating the Non-Bunching Distribution by Excluding the Bunching Region

As a variation to our primary mixture model, we explore an alternative approach designed to more explicitly isolate the parameters of the non-bunching distribution by systematically excluding observations likely belonging to the bunching distribution. This approach potentially provides a more robust estimation of the underlying non-bunching distribution, which is the counterfactual distribution used to calculate the causal effect.

The procedures for this alternative approach are as follows. First, we estimate a two-component mixture model with the bunching distribution and the non-bunching distribution using the entire dataset described in Equation (3). Second, we identify and exclude the region predominantly influenced by the bunching behavior. This is achieved by defining the interval $[\underline{Y}^I, \overline{Y}^I]$ based on the estimated parameters of the bunching distribution. Observations falling within this interval are considered to be influenced by the bunching behavior and are excluded from the subsequent analysis. Third, the remaining observations, those falling outside the interval, are then used to re-estimate the parameters of the non-bunching distribution, which now serves as our primary estimate of the counterfactual distribution. By systematically excluding the bunching region, we aim to obtain a more accurate representation of the non-bunching distribution, which is crucial for estimating the threshold's causal effect on customers. Finally, the causal effect of the threshold on customers $\tau$ is calculated as the difference between the mean of the

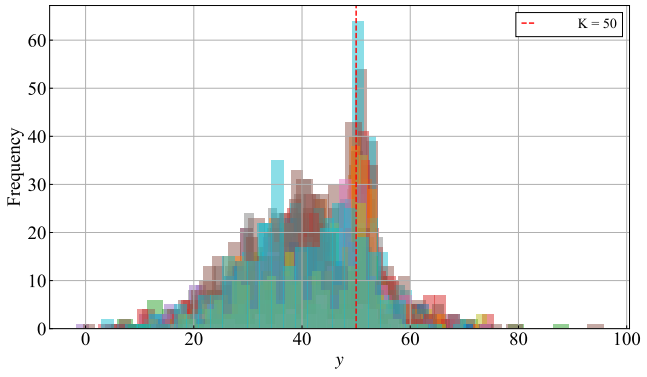


**Figure 2: Simulation dataset from a specific random seed, featuring a mixture model of bunching and non-bunching distributions (threshold $K = 50$), with data from $G = 100$ subgroups overlaid.**

bunching distribution and the mean of the non-bunching distribution in the interval $[\underline{Y}^I, \overline{Y}^I]$ as in Section 4.1. The procedures for the alternative approach can be applied to the hierarchical Bayesian model in Section 4.3 in a similar way.

## 5 Experiments

We demonstrate the effectiveness of our proposed methods by comparing their performance with the standard RDD method. We assume that the customer threshold $K$ is known and set to 50, and the observed data originate from a population structured into $G = 100$ distinct subgroups. The observed data for each subgroup $g$ ($g = 1, \ldots, 100$) are generated by a mixture model consisting of a bunching distribution and a non-bunching distribution, weighted by subgroup-specific proportions $\pi_g$ and $1 - \pi_g$, respectively. The parameters defining the distributions of components, including the proportions, can vary from subgroup to subgroup. We equally divided $G$ groups into four clusters and set the same number of group-specific sample sizes $n_g$ ($g = 1, \ldots, G$) to the same values within the same clusters. The sample sizes of the clusters are set to (50, 100, 200, 300).

In this simulation, the parameters of Equation (6) are set as follows:

$$\pi_g = 1/(1 + \exp(-\eta_g)), \quad \eta_g \sim N(\mu_\eta, \sigma_\eta^2),$$
$$\mu_\eta \sim N(-2, 0.1^2), \quad \sigma_\eta \sim N^+(0.5, 0.1^2),$$
$$\sigma_1^{(g)} \sim N(\mu_{\sigma_1}, \sigma_{\sigma_1}^2), \quad \mu_{\sigma_1} \sim N(2, 0.1^2), \quad \sigma_{\sigma_1} \sim N^+(0.25, 0.05^2),$$
$$\alpha_1^{(g)} \sim N(\mu_{\alpha_1}, \sigma_{\alpha_1}^2), \quad \mu_{\alpha_1} \sim N(1, 0.1^2), \quad \sigma_{\alpha_1} \sim N^+(0.25, 0.05^2),$$
$$\mu_2^{(g)} \sim N(\mu_{\mu_2}, \sigma_{\mu_2}^2), \quad \mu_{\mu_2} \sim N(40, 0.1^2), \quad \sigma_{\mu_2} \sim N^+(2, 0.05^2),$$
$$\sigma_2^{(g)} \sim N(\mu_{\sigma_2}, \sigma_{\sigma_2}^2), \quad \mu_{\sigma_2} \sim N(10, 0.1^2), \quad \sigma_{\sigma_2} \sim N^+(1, 0.05^2),$$

where $N^+(\cdot)$ denotes a truncated normal distribution with lower bound 0. Figure 2 shows the generated data set from a specific random seed, featuring a mixture model of bunching and non-bunching distributions (threshold $K = 50$), with data from subgroups $G = 100$ overlaid.

To evaluate the performance of our proposed method, we perform simulation replications of $M = 100$ with different random seeds. The evaluation focuses on the methods' point and interval estimation performance. To evaluate the models in terms of point estimation performance, we use the mean squared error (MSE) given by

$$\text{MSE} = \frac{1}{M} \sum_{g=1}^{G} \left(\tau_g - \widehat{\tau_g}\right)^2, \tag{7}$$

To evaluate the Bayesian models in terms of their interval estimation performance, we use the interval score (IS) [7] given by

$$\text{IS} = \frac{1}{G} \sum_{g=1}^{G} \left\{ (u_g - l_g) + \frac{2}{\alpha}(l_g - \tau_g)\mathbb{1}(\tau_g < l_g) + \frac{2}{\alpha}(\tau_g - u_g)\mathbb{1}(\tau_g > u_g) \right\}, \tag{8}$$

where $l_g$ and $u_g$ are the lower and upper bounds of the $100(1-\alpha)\%$ HDI for each subgroup $g$, respectively, and $\mathbb{1}(\cdot)$ is the indicator function. The IS increases when the prediction interval generated by the model is too broad and increases when the observed value falls outside of this interval. Therefore, a smaller IS value indicates that a more appropriate interval estimation is achieved. In this paper, $\alpha$ is 0.05.

For the simulated dataset, we fit the proposed BMTM and HBMTM models. Each model is implemented using two distinct estimation strategies: primary and alternative approaches. The primary approach uses the mixture model described in Equation (3) and (6), while the alternative approach uses the procedures described in Section 4.4. These methods are denoted as BMTM-P and HBMTM-P for the primary approach and BMTM-A and HBMTM-A for the alternative approach. Bayesian estimation is performed using the MCMC algorithm with four chains, each with 3000 iterations, a warm-up period of 3000 iterations, and adapt delta to 0.80.

For BMTM, we set the priors for each relevant parameter as follows:

$$m_{\sigma_1} = 0, \quad s_{\sigma_1} = 2, \quad m_{\alpha_1} = 0, \quad s_{\alpha_1} = 5, \quad m_{\mu_2} = 0, \quad s_{\mu_2} = 30,$$
$$m_{\sigma_2} = 0, \quad s_{\sigma_2} = 30, \quad \alpha_\pi = 3, \quad \beta_\pi = 7,$$

For HBMTM, we set the priors for each relevant parameter as follows:

$$m_{\mu\sigma_1} = 0, \quad s_{\mu\sigma_1} = 0.3, \quad m_{\sigma\sigma_1} = 0, \quad s_{\sigma\sigma_1} = 0.3,$$
$$m_{\mu\alpha_1} = 0, \quad s_{\mu\alpha_1} = 1, \quad m_{\sigma\alpha_1} = 0, \quad s_{\sigma\alpha_1} = 1, \quad m_{\mu\sigma_2} = 0, \quad s_{\mu\sigma_2} = 1,$$
$$m_{\sigma\sigma_2} = 0, \quad s_{\sigma\sigma_2} = 0.3, \quad m_{\mu\mu_2} = 40, \quad s_{\mu\mu_2} = 10, \quad m_{\sigma\mu_2} = 0, \quad s_{\sigma\mu_2} = 10,$$
$$m_{\mu\pi} = -1.4, \quad s_{\mu\pi} = 2, \quad m_{\sigma\pi} = 0, \quad s_{\sigma\pi} = 2,$$

For comparison, we also fit an RDD method. In our research problem, the variable $y$ (see Figure 2) functions simultaneously as the outcome variable and, in effect, as the running variable, the assignment variable for the treatment. This makes it difficult to distinguish the outcome and the running variables distinctly, as is typically done in standard RDD. Therefore, we employ an RDD-based method that attempts to capture the magnitude of the threshold effect on the distribution of $y$ by evaluating the discontinuity in the distribution of the variable $y$ around the threshold $K$. Specifically, kernel density estimation is performed for the variable $y$ immediately before and after the threshold $K$, and the difference between the resulting density estimates is used as the causal effect

**Table 1: Simulation results of the proposed and baseline methods in the subgroup setting. The boldface values indicate the best performance among the methods.**

| Method | MSE | IS |
|---|---|---|
| RDD | 6.773 | |
| BMTM-P | 0.678 | 3.424 |
| BMTM-A | 0.662 | 3.138 |
| **HBMTM-P** | **0.185** | **1.177** |
| HBMTM-A | 0.186 | 1.180 |

of the threshold on customers. Furthermore, when kernel density estimation is applied, since the threshold $K$ represents a distinct boundary, a boundary correction is applied to mitigate boundary bias [11]. To perform the RDD analysis, the bandwidth is established as 15 from the threshold (range: 35 to 65), considering the sample size needed for analytical stability around this point.

Table 1 summarizes the average simulation results (over $M = 100$ random seeds) for the proposed methods and the baseline method in the subgroup setting. In general, our proposed methods demonstrated superior performance to the RDD baseline method. HBMTM-P performed best in point estimation (MSE) and interval estimation (IS). HBMTM-A also delivered a strong performance, with its results eminently comparable to those of HBMTM-P. These results suggest that in the subgroup setting, HBMTM-P and HBMTM-A effectively estimate the causal effect of the threshold on customers, even when the sample size of each subgroup is small.

One notable finding from our simulation studies, particularly in the subgroup setting, was the enhanced performance of the BMTM-A approach over the BMTM-P approach in terms of point and interval estimation. This difference arises primarily from the challenges associated with accurately estimating the non-bunching distribution when applying these methods to subgroups with limited sample sizes. The BMTM-A approach mitigates challenges in small subgroups through its sequential estimation strategy. Identifying and excluding the bunching distribution isolates cleaner data for a more stable and accurate characterization of the non-bunching component, ultimately improving parameter and interval estimates.

While Table 1 summarizes the overall performance metrics, Figure 3 provides a visual comparison of the subgroup-specific causal effects estimated $\tau_g$. This figure explicitly contrasts the point estimates and the credible intervals 95% obtained from BMTM-P and HBTM-P for each subgroup, providing deeper insight into their performance at the subgroup level. As illustrated in Figure 3, HBMTM-P consistently yields more accurate point estimates of $\tau_g$ and narrower 95% credible intervals across the various subgroups compared to BMTM-P. These visual comparisons further emphasize the effectiveness of our proposed HBTM-P method in estimating the threshold's causal effect on customers, even in the presence of small sample sizes within subgroups.

## 6 Conclusion

This study addresses the critical challenge of accurately estimating the causal effects of thresholds in marketing, particularly when customers strategically manipulate their behavior to qualify for incentives. It is a common scenario that violates the assumptions of standard methods such as RDD. We propose a novel framework called

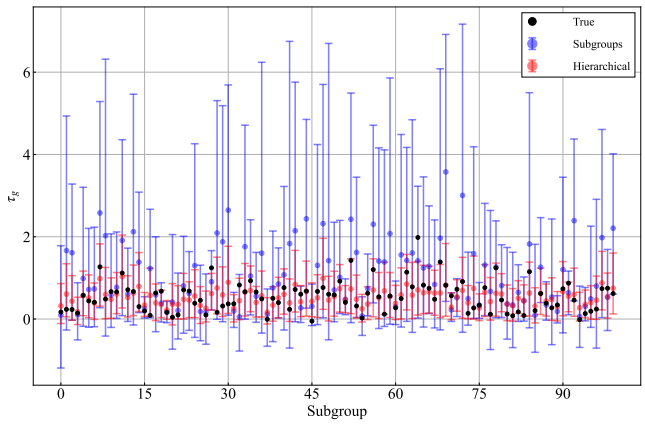


**Figure 3: Point estimates and 95% credible intervals of $\tau_g$ by BMTM-P and HBMTM-P.**

*Bayesian Modeling of Threshold Manipulation via Mixtures* (BMTM), which models the observed spending distribution as a mixture of two latent distributions: one representing customers unaffected by the threshold and another representing those strategically affected, exhibiting a distortion near the threshold value. Furthermore, we extend BMTM to Hierarchical BMTM (HBMTM) to robustly estimate heterogeneous causal effects between customer subgroups, even with limited sample sizes.

Our simulation studies demonstrate the effectiveness of our proposed methods, showing that they outperform standard RDD methods in estimating causal effects. HBMTM effectively estimates the causal effect because the hierarchical Bayesian approach allows borrowing strength across subgroups, leading to more stable estimates.

The findings of this study offer significant implications both for theory and practice in applying causal inference methods in marketing. Theoretically, our framework contributes to the growing literature on causal inference in the presence of strategic customer behavior, providing a flexible and robust approach to disentangle manipulation effects from true causal impacts. Marketers can use BMTM and HBMTM to gain more precise insight into how threshold-based incentives influence various customer segments. This understanding can inform the design of more effective and profitable marketing strategies, optimizing threshold placements, and personalizing offers to maximize customer response and return on investment.

Although our study provides a solid foundation for estimating causal effects under threshold manipulation, there are several avenues for future research. The current framework assumes specific parametric forms for the mixture components, and future work could explore more flexible, semi-parametric, or non-parametric approaches to capturing complex spending behaviors. Furthermore, investigating the performance of our methods with diverse real-world datasets, potentially incorporating dynamic aspects of customer behavior over time, would further validate and extend their applicability.

## References


[1] Rafael Alcantara, Meijia Wang, P Richard Hahn, and Hedibert Lopes. 2024. Modified bart for learning heterogeneous effects in regression discontinuity designs. *arXiv preprint arXiv:2407.14365* (2024).

[2] Adelchi Azzalini. 1985. A class of distributions which includes the normal ones. *Scandinavian journal of statistics* (1985), 171–178.

[3] Marinho Bertanha, Carolina Caetano, Hugo Jales, and Nathan Seegert. 2024. Bunching estimation methods. *Handbook of Labor, Human Resources and Population Economics* (2024), 1–44.

[4] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76 (2017).

[5] Matias D Cattaneo, Nicolás Idrobo, and Rocío Titiunik. 2024. *A practical introduction to regression discontinuity designs: Extensions.* Cambridge University Press.

[6] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. 1987. Hybrid Monte Carlo. *Physics Letters B* 195, 2 (1987), 216–222.

[7] Tilmann Gneiting and Adrian E Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association* 102, 477 (2007), 359–378.

[8] Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. 2001. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica* 69, 1 (2001), 201–209.

[9] Guido W Imbens and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge university press.

[10] Takuya Ishihara and Masayuki Sawada. 2024. Manipulation-Robust Regression Discontinuity Designs. arXiv:2009.07551 [econ.EM] https://arxiv.org/abs/2009.07551

[11] M Chris Jones. 1993. Simple boundary correction for kernel density estimation. *Statistics and computing* 3 (1993), 135–146.

[12] Ran Kivetz, Oleg Urminsky, and Yuhuang Zheng. 2006. The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention. *Journal of Marketing Research* 43, 1 (2006), 39–58.

[13] Henrik J. Kleven and Mazhar Waseem. 2013. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan *. *The Quarterly Journal of Economics* 128, 2 (04 2013), 669–723.

[14] David S. Lee. 2008. Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142, 2 (2008), 675–697. The regression discontinuity design: Theory and applications.

[15] Justin McCrary. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142, 2 (2008), 698–714. The regression discontinuity design: Theory and applications.

[16] Alina Nastasoiu, Neil T Bendle, Charan K Bagga, Mark Vandenbosch, and Salvador Navarro. 2021. Separating customer heterogeneity, points pressure and rewarded behavior to assess a retail loyalty program. *Journal of the Academy of Marketing Science* 49 (2021), 1132–1150.

[17] Omiros Papaspiliopoulos, Gareth O Roberts, and Martin Sköld. 2007. A general framework for the parametrization of hierarchical models. *Statist. Sci.* (2007), 59–73.

[18] Donald B Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* 66, 5 (1974), 688.

[19] Emmanuel Saez. 2010. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy* 2, 3 (August 2010), 180–212.

[20] Joel Slemrod. 2013. Buenas notches: lines and notches in tax system design. *Ejtr* 11 (2013), 259.

[21] Shonosuke Sugasawa, Takuya Ishihara, and Daisuke Kurisu. 2023. Hierarchical Regression Discontinuity Design: Pursuing Subgroup Treatment Effects. *arXiv preprint arXiv:2309.01404* (2023).

[22] Kevin Tao, Y Samuel Wang, and David Ruppert. 2025. Bayesian analysis of regression discontinuity designs with heterogeneous treatment effects. *arXiv preprint arXiv:2504.10652* (2025).

[23] Donald L Thistlethwaite and Donald T Campbell. 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology* 51, 6 (1960), 309.

[24] Yaming Yu and Xiao-Li Meng. 2011. To center or not to center: That is not the question—an Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics* 20, 3 (2011), 531–570.