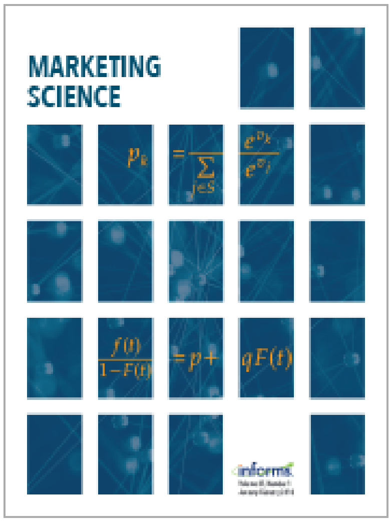


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Identifying Causal Marketing Mix Effects Using a Regression Discontinuity Design

Wesley Hartmann, Harikesh S. Nair, Sridhar Narayanan,

MARKETING SCIENCE
CELEBRATING 30 YEARS
Vol. 30, No. 6, November–December 2011, pp. 1079–1097
ISSN 0732-2399 | EISSN 1526-548X | 11 | 3006 | 1079

informs®
http://dx.doi.org/10.1287/mksc.1110.0670

Identifying Causal Marketing Mix Effects Using a Regression Discontinuity Design

Wesley Hartmann, Harikesh S. Nair, Sridhar Narayanan
Stanford Graduate School of Business, Stanford University, Stanford, California 94305
{hartmann_wesley@gsb.stanford.edu, harikesh.nair@stanford.edu, sridhar.narayanan@stanford.edu}

We discuss how regression discontinuity designs arise naturally in settings where firms target marketing activity at consumers, and we illustrate how this aspect may be exploited for econometric inference of causal effects of marketing effort. Our main insight is to use commonly observed discontinuities and kinks in the heuristics by which firms target such marketing activity to consumers for nonparametric identification. Such kinks, along with continuity restrictions that are typically satisfied in marketing and industrial organization applications, are sufficient for identification of local treatment effects. We review the theory of regression discontinuity estimation in the context of targeting and explore its applicability to several marketing settings. We discuss identifiability of causal marketing effects using the design and show that consideration of an underlying model of strategic consumer behavior reveals how identification hinges on model features such as the specification and value of structural parameters as well as belief structures. We emphasize the role of selection for identification. We present two empirical applications: the first measures the effect of casino e-mail promotions targeted to customers based on ranges of their expected profitability, and the second measures the effect of direct mail targeted by a business-to-consumer company to zip codes based on cutoffs of expected response. In both cases, we illustrate that exploiting the regression discontinuity design reveals negative effects of the marketing campaigns that would not have been uncovered using other approaches. Our results are nonparametric, easy to compute, and control for the endogeneity induced by the targeting rule.

Key words: regression discontinuity; nonparametric identification; treatment effects; targeted marketing; selection; endogeneity; casinos; direct mail
History: Received: November 25, 2009; accepted: June 2, 2011; Eric Bradlow and then Preyas Desai served as the editor-in-chief and Sanjog Misra served as associate editor for this article.

1. Introduction

Targeting is a ubiquitous element of firms' marketing strategies. The advent of database marketing has made it possible for firms to tailor prices, advertising, and other elements of the marketing mix to consumers based on their type (e.g., Rossi et al. 1996). The measurement of the causal effects of such targeted marketing is, however, tricky. A first-order complication arises because observed correlation in the data between outcome variables and marketing activities is driven both by any causal effects of marketing and by the targeting rule, leading to an endogeneity problem in estimation. The commonly used solution of instrumental variables may be infeasible in such contexts because a good instrument, a variable that is correlated with the marketing effort but otherwise uncorrelated with the outcome variable, may be hard, if not impossible, to obtain. In this paper, we propose utilizing heuristic rules often used by firms for targeting as a regression discontinuity design to nonparametrically measure the causal effects of marketing effort.

Regression discontinuity (henceforth, RD) was first introduced by Thistlethwaite and Campbell (1960) in the evaluation literature (e.g., Cook and Campbell 1979) and has become increasingly popular in program evaluation in economics. An RD design arises when treatment is assigned based on whether an underlying continuous *score* variable crosses a cutoff. The discontinuity induced by this treatment rule induces a discontinuity in the outcomes for individuals at the cutoff. Hahn et al. (2001; henceforth, HTV) formally showed that this discontinuity nonparametrically identifies a local average treatment effect if the counterfactual outcomes for agents with and without treatment are also *continuous* functions of the score in the neighborhood of the cutoff. Under these conditions, observations immediately to one side of the cutoff act as a control for observations on the other side, facilitating measurement of a causal effect of the treatment. RD designs have now been used to study treatment effects in a variety of contexts, from education (Black 1999, Angrist and Lavy 1999), to housing (Chay and Greenstone 2005), to voting (Lee et al. 2004) amongst others. In contrast, applications to marketing and industrial organization contexts have been sparse. Notable applications include Busse et al. (2006, 2010), who measured the effect of manufacturer promotion on automobile prices and sales using a design in which calendar time is the score variable, albeit not in a targeted marketing context. Van der Klaauw (2008), Imbens and Lemieux (2008), and Lee and Lemiuex (2010) are excellent summary papers on RD that discuss the method, its variants, and applications in detail.

We believe targeted marketing contexts are particularly well suited for the use of RD methods for two reasons. First, firms often target groups of customers with similar treatments. Even though firms face a continuous distribution of consumer types, it is common in actual business practice to allocate similar marketing interventions to groups of customers. The reasons for this bunching include menu or implementation costs, or the inherent difficulty of tracking historical information required for targeting at the individual-customer level. Second, targeting policies of firms often involve trigger rules. Marketing allocation often involves "rules of thumb" whereby groups of consumers obtain similar marketing levels based on whether a relevant function of their characteristics or historical behavior crosses a prespecified cutoff. For instance, catalogs might be mailed based on cutoffs of underlying "RFM" (recency, frequency, monetary) score variables, credit card promotions may be given based on cutoffs of FICO© scores, detailing calls may be made to a physician based on whether he is in specific prescription-based deciles, price discounts may be given to people above or below certain age cutoffs, etc. The ubiquity of such trigger rules generates a wealth of discontinuity-based contexts that facilitate nonparametric identification of marketing effects using an RD design, which have previously been unexploited in the marketing literature.

Applying RD in marketing and industrial organization contexts, where theoretical and empirical models of strategic choice are abundant, naturally leads us to consider the extent to which these models relate to the identifiability of RD. We consider permutations on simple models of consumer selection to delineate a set of viable and nonviable applications for RD. First, we present a Hotelling-style model to show that if customers face sufficiently high costs of selecting, RD is valid. The model illustrates that RD can often be used to measure marketing effects under geographic targeting (i.e., high fixed costs of moving to receive the treatment) or in situations where targeting is based on scores that cannot be changed, such as age-based marketing (i.e., infinite fixed cost of selection). We then apply RD to a geographic targeting example where the score variable is a function incorporating the probability of response at the zip-code level. Direct mail is sent to a zip code if the probability of response is above a cutoff.

Second, we present a detailed illustration of targeting based on past purchases. When past purchase behavior crosses a threshold, customers qualify for preferential treatment. We show that applicability of RD hinges on whether or not customers are uncertain about the exact score, the cutoff, or both. The implication is that canonical reward programs where thresholds and scores are communicated to customers have selection effects that invalidate RD. On the other hand, database marketing programs, where typically both the score and cutoff are unknown to consumers, are valid RD applications even in the presence of selection.[1] To illustrate the value of RD in database marketing, we analyze data from a casino's marketing efforts to members of its loyalty program. The casino uses a targeting rule that is discontinuous based on cutoffs in the average level of past gambling activity. These cutoffs are not known to the consumers, and hence, they cannot self-select into preferential treatment. We estimate the effect of both the database marketing and geographic targeting applications nonparametrically using local linear regression (Fan and Gijbels 1996). We find in both cases that controlling for the endogeneity has large implications on the conclusions drawn from the analysis. In particular, we find that a naive estimate has an altogether different sign than the RD estimate.

Finally, we also formally consider time as a score. We illustrate that the validity of the RD in the timing case hinges on whether or not the estimation is conditional on selection decisions such as purchase or store visitation as well as the belief structures leading to these decisions. We discuss the role of dynamics induced by durability or storability in the interpretation and identification of treatment effects in a time-based design. An important takeaway from these analyses is that the identification conditions cannot be evaluated without consideration of an explicit structural model of behavior of agents that is representative of the underlying data-generating process. Our analysis illuminates an "RD paradox": the design is often thought of as "atheoretic" or "assumption-less," but identification often relies on crucial, sometimes nontransparent, primitive assumptions regarding behavior.

We consider the RD design to be complementary to several alternative methods focused on uncovering

[1] For instance, pharmaceutical firms use volumetric deciles of physicians to decide the number of detailing calls made to doctors. These deciles are category specific, and doctors are unlikely to know their own prescription volumes relative to all other physicians for each category. Similarly, consumers are unlikely to know their RFM score or the trigger values used for targeted mailing of catalogs.

causal effects. Randomized variation through experiments are ideal, but firms are often unable or unwilling to conduct randomized trials because of considerations of cost, time, and potential backlash from consumers not receiving preferred treatment. When experimentation is unavailable, RD may be more viable for targeted marketing applications. One popular alternative is to use instrumental variables, but such variables are hard to obtain because customer-side variables typically fail the required exclusion restrictions and cost-side variables typically do not vary by the segment the firm uses for targeting. Another alternative is to augment the analysis with a model of how firms allocate marketing efforts and to incorporate the restrictions implied by this model in estimation (e.g., see Manchanda et al. 2004, Otter et al. 2011). These authors are careful to point out that this approach is feasible only if full information is available to the analyst about how the firm allocates its marketing efforts. In the absence of such information, the analysis is sensitive to misspecification bias.

The main caveats for adopting the RD approach are threefold. First, by its nonparametric nature, the estimator is data intensive and requires many observations on consumer behavior at the cutoff; in sparse-data situations, parametric approaches are more suitable. Second, the estimator provides a local treatment effect that is relevant only for the subpopulation of consumers at the cutoff, not globally.[2] A third caveat is that, like any other alternative, the conditions for the validity of the estimator have to be carefully assessed depending on the context. We consider the last aspect to be especially crucial. The HTV conditions on identification are stated in terms of continuity of counterfactual outcomes at the cutoff. We discuss in detail how these conditions can be translated in practice to several commonly observed targeted marketing situations. A key point we wish to emphasize is that the validity of the RD design has to be based on a formal model of a data-generating process by explicitly considering how consumers sort at the cutoff.

To summarize, this paper makes three contributions. First, we identify the ready application of the RD design to typical targeted marketing contexts. Our goal is not to present new estimators per se, but to point out how discontinuous rules of thumb, which are pervasive in real-world marketing situations, may be used to achieve nonparametric identification. Furthermore, we point out that such rules of thumb, which have been typically treated as nuisance issues to be dealt with, are a source of identification of the causal effects of marketing activities. Second, we present detailed illustrations of the identifiability of causal marketing effects using the design, and we show theoretically the conditions under which the RD estimator may be valid in marketing contexts, considering in particular the role of consumer self-selection. The link to a structural model, and the treatment of identification in the context of such a framework, is new to the RD literature. Finally, we demonstrate the utility of the RD approach through two empirical applications, with counterintuitive conclusions that are hard to uncover through a naive analysis.

The rest of this paper proceeds as follows. We first provide a brief review of the identification conditions for the RD estimator. We then discuss our theoretical and simulation results on identification of marketing mix effects under specific targeting situations. We then present our two empirical applications. The last section concludes.

2. Identification of Marketing Mix Effects Using an RD Design

In this section, we review identification conditions from HTV, Lee (2008), and Lee and Lemieux (2010), and we discuss the role of local inference in marketing contexts.

2.1. Identification Conditions

The following describes the key identification conditions from HTV. To set up the notation, let d_i indicate exposure to marketing, and let $Y_i(1)$ and $Y_i(0)$ be the potential outcomes for individual i with and without marketing. The treatment effect, $Y_i(1) - Y_i(0)$, cannot be directly estimated, as only $Y_i = d_i Y_i(1) + (1-d_i) Y_i(0)$ is observed by the analyst for each i. Instead, the focus is on measuring an average treatment effect, $\mathbb{E}[Y_i(1) - Y_i(0)]$, where the expectation $\mathbb{E}(\cdot)$ is taken over individuals (or the density of individual types). The RD design implies that treatment is assigned depending on whether a continuous score, z_i, crosses a cutoff, $\bar{z}$; i.e., $d_i = \mathcal{I}(z_i \geq \bar{z})$. Then, the observed size of the discontinuity in the outcome (Equation 1) and the treatment (Equation 2) in a neighborhood of $\bar{z}$ are

$$Y^- = \lim_{z \to \bar{z}^-} \mathbb{E}[Y_i(0) \mid z_i = \bar{z}] \quad \text{and} \qquad Y^+ = \lim_{z \to \bar{z}^+} \mathbb{E}[Y_i(1) \mid z_i = \bar{z}], \tag{1}$$

$$d^- = \lim_{z \to \bar{z}^-} \mathbb{E}[d_i \mid z_i = \bar{z}] \quad \text{and} \qquad d^+ = \lim_{z \to \bar{z}^+} \mathbb{E}[d_i \mid z_i = \bar{z}]. \tag{2}$$

[2] In many situations, this may be precisely the object of interest for inference. Measurement of treatment effects for the entire population would require more assumptions, or the restrictions from a formal model of behavior.

Theorem 1 (HTV). *Suppose that*(1) d^- *and* d^+ *exist, and* $d^+ \neq d^-$; *and* (2) Y^- *and* Y^+ *are continuous in* z_i *at* $z_i = \bar{z}$. *Then, the quantity* $\beta = (Y^+ - Y^-)/(d^+ - d^-)$ *measures the average treatment effect at* $\bar{z}$; *i.e.,* $\beta = \mathbb{E}[Y_i(1) - Y_i(0) \mid z_i = \bar{z}]$.

That is, to obtain a causal effect of the treatment, we estimate the discontinuity in the outcomes and weigh it down by the discontinuity in treatments.[3] Estimation of the size of the discontinuities can be achieved by nonparametrically estimating the limit values (d^-, d^-, Y^-, Y^+) (see, e.g., Porter 2003). The interpretation of the size of the discontinuity in outcomes as a treatment effect hinges crucially on the continuity conditions. The continuity of $\mathbb{E}[Y_i(1)]$ and $\mathbb{E}[Y_i(0)]$ in z_i enables us to interpret the outcomes just below $\bar{z}$ as a valid counterfactual for the outcomes just above $\bar{z}$, thereby facilitating a relevant control group for measurement of a treatment. The condition of continuity of counterfactual outcomes is equivalent to requiring either continuity in the density of z_i, or continuity in the density of consumer types given z_i at $z_i = \bar{z}$. In essence, one implies the other. Letting θ index consumer types, and taking the expectation over θ, note that $\mathbb{E}_\theta[Y(0 \mid \theta) \mid z = \bar{z}] = \int Y(0 \mid \theta) h(\theta \mid z = \bar{z})\, d(\theta)$, where $h(\cdot)$ is the density of θ given z (equivalently for $Y(1 \mid \theta)$). If $h(\theta \mid z)$ is not continuous at $z = \bar{z}$, we obtain that

$$\lim_{z \to \bar{z}^-} \int Y(0 \mid \theta) h(\theta \mid z)\, d(\theta) \neq \lim_{z \to \bar{z}^+} \int Y(0 \mid \theta) h(\theta \mid z)\, d(\theta). \tag{3}$$

Thus, continuity of counterfactual outcomes at the cutoff is violated, and the RD is invalid.[4] By implication, the RD design is equivalently invalid if the distribution of the score z given type θ is not continuous at $z = \bar{z}$. By Bayes rule, the conditional density of types given the score is $h(\theta \mid z) = (f(z \mid \theta) f(\theta))/f(z)$, implying that a discontinuity in $f(z \mid \theta)$ will make $h(\theta \mid z)$ discontinuous as well.

The continuity condition is invalidated in the context of selection. Assume that agents have control over their score. If they can precisely control their score and have an incentive to select into treatment, the distribution of the score z will not be continuous at the cutoff $\bar{z}$, because the agents immediately to the right of the cutoff are those who chose to select into treatment, whereas those to the left of the cutoff chose not to be treated. Nevertheless, Lee (2008) and Lee and Lemieux (2010) introduce the important idea that when the score includes some random noise, the continuity conditions may still be satisfied and a valid RD design obtained. Stated formally, let the score now have a component to the utility that is not predictable and not controllable by the agent. Thus, $z = x + w$, where x is the systematic part of the score that the agent can predict and can take actions to control, and w is an exogenous random chance component to the score that cannot be predicted or controlled by the agent. If w has a continuous density, the distribution of z at the cutoff $\bar{z}$ is locally continuous, thus validating the RD design. This results from the fact that agents are unable to precisely select into treatment. Thus, in the neighborhood of the cutoff $\bar{z}$, the distribution of agents on either side of the cutoff is the same. In the above notation, $h(\theta \mid z)$ is continuous at $z = \bar{z}$, validating the RD design. If, on the other hand, w is either zero (i.e., there is no random, unpredictable component to the score) or is discontinuous at $\bar{z}$, then the randomness cannot validate the design.

The continuity condition is also violated if the cutoffs on the score variable that define treatment is chosen at a point of discontinuity in the score. In some marketing contexts, for instance, this may happen because competitive promotions use the same cutoffs for treatment or because firms use natural points of discontinuity in the score as cutoffs for assigning consumers to the treatment. The latter is typically associated with sparse data in the neighborhood of the cutoff or the use of structural breaks in the score to decide the cutoffs.

2.2. Local Inference and Marketing

As previously mentioned, RD estimates an effect that is local to a particular value of the score. This could be problematic when designing a marketing policy for the entire distribution of customers in a database. If a firm were to only set one cutoff, we might worry that the firm would not set that cutoff exactly where the effect of the marketing policy is largest, or even where the magnitude is expected to be the average effect. Therefore, if firms find large local effects, their optimal response is to move the cutoff to include more customers until the marginal benefit of the treatment is roughly equal to the incremental cost of the treatment.

How does such a strategic cutoff-setting process affect inference? It remains true, per the preceding subsection, that the effects are valid treatment effects at the cutoffs. However, it is clear that the treatment effects are not randomly sampled points from the distribution of the score. Practitioners should therefore be cautious in interpreting the results. We recommend the use of multiple cutoffs and movement of cutoffs in response to either negative marginal effects or

[3] In a sharp RD, $d^+ = 1$ and $d^- = 0$, so the discontinuity in outcomes itself is the treatment effect. In a "fuzzy" RD, $(d^+, d^-) \in (0, 1)$; i.e., treatment is not certain if the score is crossed.

[4] Recall, by the definition of continuity, $h(\theta \mid z)$ is continuous at $z = \bar{z}$ if $\lim_{z \to \bar{z}^-} h(\theta \mid z) = \lim_{z \to \bar{z}^+} h(\theta \mid z) = h(\theta \mid \bar{z})$. Hence, discontinuity of $h(\theta \mid z)$ implies (3).

large marginal effects. Through such a process, firms will retain the simplicity of classifying customers into a limited number of targeted groups for marketing while also reducing the loss associated with under- or overincentivizing some set of customers.

3. Geographic Targeting

We begin by considering an example of geographic targeting. This simple application is instructive for two reasons. First, the requirement for identification is simply that the cost of moving residences is substantially larger than benefits of preferential marketing. Second, this model allows us to clearly illustrate the identifying conditions from HTV. We wrap up this section with an empirical analysis of geographic targeting using RD.

3.1. Model

We define a model that reflects our empirical application where consumers are targeted preferential direct mail based on their location. The model involves two stages. Initially, each consumer is endowed with a score z, which can be thought of as his location on a Hotelling line. In the first stage, the customer makes a selection decision to move his location to $\tilde{z}$, which we call the "manipulated score." If the consumer decides not to move to a new location, his manipulated score $\tilde{z}$ remains the same as his initially endowed score z. If $\tilde{z} \geq \bar{z}$, the consumer is eligible for the treatment. In the second stage, the customer makes a decision about the outcome of interest, conditional on his treatment eligibility. The rational consumer takes the effect of treatment on outcomes in the second stage into account when making a selection decision in the first stage. We consider the two stages of the model in reverse, considering the second stage first and then using the optimality condition from the second stage into account in solving for the optimal decision in the first stage.

3.1.1. Stage 2: Outcome. The outcome Y is a binary variable indicating whether or not an individual makes a purchase. Treatment is indicated by the binary variable $R = \mathcal{I}(\tilde{z} \geq \bar{z})$. We model the individual's outcome as a random utility model $Y = \mathcal{I}(u_1 > u_0)$, where

$$\begin{aligned} u_1 &= v(X, R=1 \mid \beta)\mathcal{I}(\tilde{z} \geq \bar{z}) \\ &\quad + v(X, R=0 \mid \beta)\mathcal{I}(\tilde{z} < \bar{z}) + \eta_1, \qquad (4) \\ u_0 &= \eta_0. \end{aligned}$$

Here, $v(\cdot)$s indicate the nonstochastic portion of the individual's utility of choosing to purchase, and $\eta = (\eta_1, \eta_0)$ are mean-zero unobservables (to the econometrician) that affect purchases. We introduce these unobservables so as to clarify the separate role played by unobservables affecting selection versus those affecting purchase, in ensuring identification.

3.1.2. Stage 1: Selection. In the first stage, each customer can choose to manipulate his current score to $\tilde{z} = z + m$. Changing the score is not costless. The total cost of moving has a fixed and marginal component, $C = F + \tau m$. A consumer at z would move to the cutoff $\bar{z}$ if the expected value from obtaining the treatment is greater than the cost:

$$\begin{aligned} \mathbb{E}_\eta[u_1 - u_0] &= v(X, R=1 \mid \beta) - v(X, R=0 \mid \beta) \\ &\geq F + \tau(\bar{z} - z). \qquad (5) \end{aligned}$$

The marginal customer that selects into treatment is defined as z^* such that

$$z^* = \frac{1}{\tau}\big(F + \tau\bar{z} - [v(X, R=1 \mid \beta) - v(X, R=0 \mid \beta)]\big). \qquad (6)$$

3.1.3. Identification. We now consider whether an RD applied to this context is valid. Validity depends on whether continuity of the manipulated score $\tilde{z}$ is violated at the cutoff $\bar{z}$. Continuity of $\tilde{z}$ depends on whether the marginal consumer has a score z^* less than z. If $z^* < \bar{z}$, all consumers between $[z^*, \bar{z})$ would move. Hence, the score would have positive mass to the right of $\bar{z}$ but no mass just to the left of $\bar{z}$. Thus, the distribution of $\tilde{z}$ would jump at $\bar{z}$, invalidating the RD. Another intuition is to note that with such selection, the limit of the counterfactual outcome just to the left of $\bar{z}$ does not exist. Formally stated, the condition for RD to be invalid, $z^* < \bar{z}$, implies from Equation (6) that $F < [v(X, R=1 \mid \beta) - v(X, R=0 \mid \beta)]$. Intuitively, if the fixed costs of moving are not higher than the gain from moving, selection can invalidate an RD application by violating continuity in the score relevant for treatment.

3.1.4. Heterogeneity. We now consider whether heterogeneity of consumer types can resolve this identification problem. For instance, heterogeneity in $\theta = (F, \tau, \beta)$ could imply that there exist at least some mass of consumers to the left of $\bar{z}$, who may not move (for example, individuals with very high fixed costs). This ensures that the limit of the counterfactual outcome from the left exists. However, unless *all* customers have sufficiently large fixed costs, the RD is not valid. Mathematically, it is easier to see this in terms of checking the continuity of the counterfactual outcome Y in the absence of treatment: $R = 0$. The limit of the counterfactual outcome from the left of $\bar{z}$ is

$$\begin{aligned} &\lim_{\tilde{z} \to \bar{z}^-} \mathbb{E}[Y(0 \mid \theta) \mid \tilde{z}] \\ &= \lim_{\tilde{z} \to \bar{z}^-} \iint Y(X, R=0, \tilde{z}, \eta, \theta)\, d\mathcal{F}_{\theta \mid \tilde{z}}(\theta \mid \tilde{z} < \bar{z})\, d\mathcal{F}_\eta(\eta), \qquad (7) \end{aligned}$$

whereas the limit from the right of $\bar{z}$ is

$$\lim_{\tilde{z}\to\bar{z}^+}\mathbb{E}[Y(0\mid\theta)\mid\tilde{z}]$$
$$=\lim_{\tilde{z}\to\bar{z}^+}\iint Y(X,R=0,\tilde{z},\eta,\theta)d\mathcal{F}_{\theta\mid\tilde{z}}(\theta\mid\tilde{z}\geq\bar{z})\,d\mathcal{F}_{\eta}(\eta). \tag{8}$$

In the presence of selection, the set of consumers to the right of the cutoff would have lower τ and F and higher β than those to the right. Hence, $\mathcal{F}_{\theta|\tilde{z}}(\theta\mid\tilde{z}<\bar{z})\neq\mathcal{F}_{\theta|\tilde{z}}(\theta\mid\tilde{z}\geq\bar{z})$, and the left-hand sides of Equations (7) and (8) are not the same. Hence, heterogeneity does not guarantee validity of the RD design. This can be mitigated only if θ is such that no one moves in order to obtain treatment, which is likely if the fixed costs of moving are large enough compared to the benefits of obtaining the reward R.

3.1.5. Discussion. The above analysis suggests that geographic targeting will plausibly be a valid RD application because preferential marketing treatment (e.g., receipt of catalogs) is unlikely to ever be large enough to outweigh the costs of moving. However, applying RD to geographic targeting relies on an underlying model of customer selection as well as the definition and magnitude of structural parameters such as moving costs. Moving outside of the geographic space as the underlying score variable may involve much smaller "moving" costs that could invalidate RD applications.

3.2. Direct Mail Activity by a Direct Marketing Firm

We consider a canonical marketing problem: measuring the causal effects of direct mail. Our application involves a direct marketing firm sending direct mail to customers to solicit a request for contact with the company. Once the customer contacts the company (either online or by phone), further promotions and prices are offered in order to acquire the customer. We focus on whether or not the customer contacts the company as the response variable of interest. Measuring the effect of direct mail is not straightforward, because the firm does not randomly choose consumers to whom to send the direct mail solicitations. Rather, because response rates are small for this mode of marketing (of the order of 1%–2%), the firm tends to send direct mail to customers it anticipates are most likely to respond. The firm's targeting is at the level of a zip code; i.e., it chooses to send direct mail to all consumers in the zip code or to none. It is important to note that the firm decides the choice of zip codes in which to send direct mail to customers based on cutoffs on a one-dimensional score variable, which is a (unknown to us) function of customer characteristics, past response histories, and other features of the zip code. The question of interest is whether direct mail solicitations causally affect the number of customer contacts. We observe the score variable, the cutoff, a treatment variable indicating whether or not direct mail was sent, as well as the customer contacts for each of the zip codes in six different states in the United States. The dependent variable is whether or not a customer in a zip code contacts the company.

We conduct four kinds of analysis on these data to illustrate the application of RD to this geographical targeting context and report these results in Table 1. First, we test for differences in mean customer contact rates for those zip codes in which consumers received the mail solicitation versus zip codes in which consumers did not receive mail solicitations, and we do this separately for each of the six states in our data. These mean differences are identical to the slope coefficients for an ordinary least squares (OLS) regression of direct mail solicitations on the customer contact rates. These results, reported in the top panel of the table, would seem to suggest a significantly positive effect of direct mail solicitation on customer contact rates. However, this is a naive analysis that does not account for the fact that zip codes with higher scores are selected for the direct mail campaign, presumably because they have higher expected customer response rates.

We next discuss the RD estimates for the effect of direct mail solicitations on customer contact rates. Specifically, we compare the limiting values of the customer contact rates for zip codes with scores in the neighborhood of the cutoff and on the two sides of it to measure the causal effect of the direct mail solicitation. The identifying assumption for the causal effect using RD in this application is that the benefits from receiving the direct mail solicitation are small compared to the fixed costs of moving to a different zip code, and hence it is implausible that consumers select into treatment. We compute the limiting values of the customer contact rates using local linear regressions on the two sides of the cutoff. In the middle panel of Table 1, we report the RD estimates with optimal bandwidth computed separately for each of the six states in our data, because the cutoffs differ by state. In any RD application, we need to choose the bandwidth around the cutoff in which the analysis is done. We focus on the bandwidth that minimizes mean squared error (MSE).

The results of the RD estimates computed for the optimal bandwidth show that the causal effect of direct mail is positive only in Tennessee and is actually negative in Arizona and Wisconsin. The remaining three states have no significant effect of the direct mail solicitation on response rates. Whereas the null effects in Washington, New Jersey, and Minnesota are not necessarily surprising, the negative effects in

Arizona and Wisconsin are. We conjecture, but cannot verify, that these may reflect an adverse reaction to direct mail activity because of heavy direct mail activity by the firm in the past. Another explanation is that in Wisconsin, for instance, consumers who were picked out for this campaign were subject to heavy direct mail by competitors.

We assess sensitivity of the results to the bandwidth selected. In the bottom panel of the table, we report the RD estimates when the bandwidth is 50% higher than the optimal bandwidth. We find that the estimates for Arizona are not robust to the change of bandwidth, whereas those in Wisconsin are. This illustrates the fact that in RD designs, bandwidth selection is an important part of the process of inference.

4. Targeting Based on Past Purchases

We discuss a model of history-based targeting. Such an approach is valid for RD if individuals do not know either their score or cutoff. The former is an application of the approach in Lee (2008) where a random shifter of the score is realized postselection. Uncertainty about the cutoff, however, is more involved and requires a formal derivation of agents belief structures because identification hinges on continuity of these beliefs in particular neighborhoods. A precursor to deriving each is a model with known score and cutoff. In this case, it becomes clear why a random variable a known to the consumer, such as the realization of a logit shock to outcomes, is insufficient to make RD valid. This rules out RD for reward programs where firms communicate point accumulations and payoff schedules. Structural consideration of this application also allows us to illustrate how selection satisfying Lee's criteria affects the interpretation of treatment effects. To better deal with treatment effects on selected samples, we suggest and analyze an application with multiple cutoffs.

4.1. Model

Once again, the model reflects one of our empirical applications where we consider a reward program in which a casino offers short-lived promotions to gamblers based on a score computed as a function of their past gambling activity. Unlike in a geographical targeting application, fixed costs to consumers of changing gambling amounts are likely quite low, so the validity conditions in the previous section do not apply. However, in the context of this application, we show that an RD design continues to be valid if consumers have uncertainty about the exact score or cutoff used by the firm to target promotions. This requires augmenting the model to allow for uncertainty and randomness *at the selection stage*. Our analysis reveals that the role of this randomness is subtle and context specific. We show that the extent to which such randomness can smooth out the discontinuity induced by selection depends on the *nature of the randomness and the precise details of how it affects behavior*. We build on the contribution of Lee (2008), which pointed out that randomness in the score can validate RD designs, by discussing how this aspect requires a careful consideration of an underlying behavioral model for assessment.

4.1.1. Setup. Consider a simplified promotion program that provides a reward R to consumers if an index of their past outcomes, z, crosses a cutoff, $\bar{z}$. In the casino application, R can represent play credits given to consumers, and z can represent an index of the dollar amount a consumer has played with the casino in the past (we present specific details in §4.2). Here, R is the treatment, z is the score, and $\bar{z}$ is the cutoff at which a discontinuity arises. The econometrician wishes to measure a treatment effect of the reward on current outcomes (e.g., how much the consumer plays today) by comparing the behavior of consumers just to the right of $\bar{z}$ to those just to the left of $\bar{z}$. Consumers are forward looking, know the reward R, and initially also know the cutoff $\bar{z}$ and their score, z. Customers have an incentive to play to earn the reward when their current score is below the cutoff but a reduced incentive to play with the objective of earning a reward when their current score already exceeds the cutoff.[5]

As before, we consider a two-stage model. In the first stage, the consumer makes a selection decision. He enters the first stage with no rewards and with score z. Based on z, R, and his characteristics, he evaluates whether to self-select and play in the first stage in order to earn a reward in the second. Selection adds m to z, and the manipulated score is then $\tilde{z} = z + m$. If the consumer chooses to not select into treatment, the manipulated score remains the same as the original score; i.e., $\tilde{z} = z$.

In stage 2, those consumers with manipulated score $\tilde{z} \geq \bar{z}$ obtain the reward R. Then, conditioning on their treatment status, all consumers make an outcome decision denoted as $Y = Y(\tilde{z}, R, \eta)$, where η is an unobservable (to the econometrician).

The econometrician observes $\{Y, \tilde{z}\}$ across a sample of consumers in the neighborhood of $\bar{z}$ and wants to estimate the effect of the treatment R on outcomes Y. Note that the action of "playing" is both a selection action and an outcome of interest, albeit at two different stages.

[5] Such incentives have been shown to be nontrivial empirically—for instance, Drèze and Nunes (2011) document that airline consumers just below the 25,000-mile cutoff are significantly more likely to fly to earn a frequent-flier reward than those just above. As we show, this type of discontinuity invalidates RD applications in the case of frequency reward programs.

Table 1 Results from Analysis of Direct Mail Program

State	AZ	WA	NJ	MN	TN	WI
Difference in means	0.0051 (0.0012)***	0.0020 (0.0002)***	0.0019 (0.0003)***	0.0011 (0.0005)**	0.0023 (0.0005)***	0.0034 (0.0012)***
RD estimate with bandwith that minimizes MSE	−0.0168 (0.0073)**	−0.0002 (0.0012)	0.0000 (0.0016)	0.0008 (0.0017)	0.0063 (0.0021)***	−0.0103 (0.0048)**
RD estimate with 50% increase in bandwith	−0.0090 (0.0070)	−0.0001 (0.0009)	−0.0001 (0.0011)	−0.0010 (0.0014)	0.0049 (0.0017)***	−0.0078 (0.0037)**

***Significant at the 1% level; **significant at the 5% level.

In contrast to the previous model, we now allow for some randomness. We allow selection to be stochastic by introducing a shock ε that represents random events in the casino that affect the consumer's selection decisions. This shock is unobserved to the econometrician, but observed by consumers. To reflect the dependence on ε, we denote the selection decision by an indicator $y = y(m, z, R, \varepsilon)$.[6] Thus, this setup builds on that in §3 by (a) allowing ε to influence the selection decision and (b) making actions at the selection stage discrete.

The manipulated score, $\tilde{z}$, can be written as

$$\tilde{z} = z + m \times y(m, z, R, \varepsilon). \tag{9}$$

Following a canonical discrete-choice setup, we assume that y is determined based on an inequality condition involving consumer's type, his state z, and the realization of the error ε:

$$y = \mathcal{I}(f(m, z, R) + \varepsilon > 0). \tag{10}$$

We do not explicitly write out the deterministic component, $f(m, z, R)$, but implicitly, this would be the difference between the choice-specific value function associated with playing and earning a reward and that for not playing. Consider a consumer 1 with z just to the left of $\bar{z}$ such that $z \in [\bar{z} - m, \bar{z})$. Given selection, the induced distribution of his manipulated score $\tilde{z}$ is

$$\tilde{z} = \begin{cases} z + m & \text{w.p. } \Pr(y = 1 \mid z < \bar{z}), \\ z & \text{w.p. } \Pr(y = 0 \mid z < \bar{z}). \end{cases}$$

To examine the implications of selection for the induced distribution of $\tilde{z}$, note that in general, we expect that $f(m, z, R \mid z < \bar{z}) > f(m, z, R \mid z \geq \bar{z})$, because we expect that consumers who do not have the reward but are close to it derive a net value from self-selecting that is *higher* than that derived by those who already have the reward.[7] This implies that post-selection, the manipulated score for consumers who start just below the cutoff will jump to have higher mass to the right of the cutoff. Because the proportion of consumers who purchase at the selection stage is higher just below the cutoff than just above, the density of the manipulated score has a discontinuity at the cutoff. A discontinuity in the density of the manipulated score at the cutoff implies a discontinuity in the types just to the left and the right of $\bar{z}$, and RD is invalidated by selection. This intuition is demonstrated graphically in Figure 1.

Discussion. We just demonstrated that selection in a typical reward program invalidates the RD design. Recent literature has suggested that random factors that drive outcomes (such as ϵ) can make RD designs valid even in the presence of selection. Casual intuition may suggest that if random events shift consumer's gambling amounts around a known cutoff, it could be that the distribution of types at the cutoff is "as if it were randomized," and the RD may be valid. Here, we show that this intuition is not general and does not hold in the example shown above. Specifically, it fails because, as Lee (2008) suggests, the realization of the randomness to the agent must occur *after* selection for the randomness to smooth out the discontinuity in scores. The above analysis makes this intuition specific.

4.1.2. Mitigating Selection: Uncertainty.

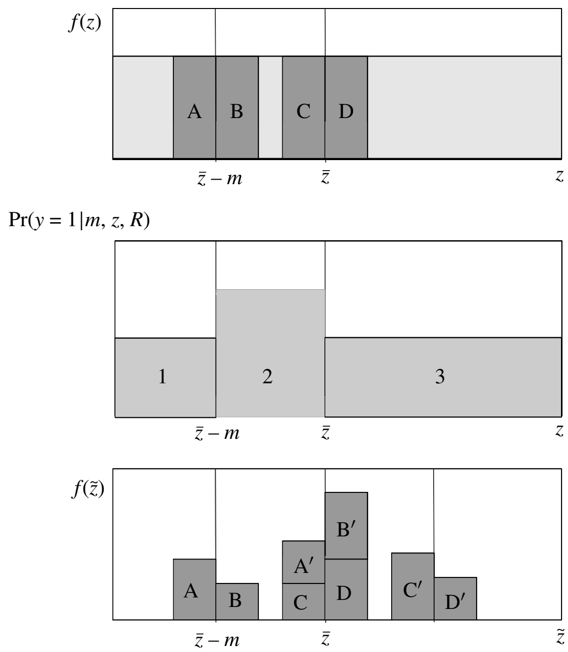
Uncertainty About the Score. Now we consider if uncertainty (from the consumer's perspective) in the manipulation of the score can mitigate the issues raised by selection. We generate the uncertainty by adding an additive shock to the manipulated score. This shock represents aspects of the score that cannot be controlled by the consumer at the selection stage. We assume this uncertainty has a continuous density.[8]

[6] To be clear, ε is an unobservable that affects the selection decision y, whereas η is an unobservable that affects the outcome decision Y.

[7] This can be seen from the fact that the value of selecting is a function of the additional utility derived from getting the treatment, weighted by the increase in probability of treatment that comes from choosing to play at the selection stage. For consumers with $z < \bar{z}$, the probability of getting treatment increases when the consumer selects. However, for consumers with $z > \bar{z}$, this is not true.

[8] In principle, the RD estimate finds the limits of the outcomes for consumers on two sides of the cutoff, and therefore the distribution

Figure 1 Discontinuity Because of Selection



Notes. The top panel depicts a continuous distribution of the score, $f(z)$, chosen to be uniform for convenience. Four consumers, A–D, are shown. The middle panel depicts the probability of selection as a function of z. The probability is higher for those in region 2: $[\bar{z}-m, \bar{z})$. The bottom panel depicts how the discontinuous incentives in the middle panel led the otherwise continuously distributed individuals from the top panel to be discontinuously distributed around the cutoff $\bar{z}$. There are four relevant types of customers within a small bandwidth of the cutoff: A′ are those from region 1 that chose y; B′ are those from region 2 that chose y with the added incentive m that put them across the cutoff; C are those just below the cutoff who, despite an increased incentive to move, did not chose y; and D are those just above the cutoff that also did not chose y.

We therefore modify Equation (9) to include a random term, w:

$$\tilde{z} = z + m \times y(m, z, R, \varepsilon) + w. \tag{11}$$

The key addition in this specification, w, is an error term that represents the consumer's uncertainty about the eventual realization of the score. This may occur, for instance, if consumers forget their past play since it has been too long or if consumers do not know the exact score used by the firm.[9] Assume that w has a continuous density with full support over $(\bar{z}-h, \bar{z}+h)$, where h is the bandwidth defining the neighborhood of the cutoff used for estimation. Whereas the econometrician does not observe either ε or w, the consumer observes ε prior to the selection decision, but *not* w. Introduction of w thus removes the ability of agents to sort *precisely* around the cutoff $\bar{z}$.[10] Following the intuition proposed in Lee (2008), we illustrate how this aspect restores the validity of the RD in spite of selection.

Consider two individuals in a neighborhood of $\bar{z}$, such that individual 1 lies close to the left and individual 2 to the right of the cutoff; i.e., $z_1 < \bar{z}$ and $z_2 \geq \bar{z}$. Consider the distribution of the manipulated score $\tilde{z}_1$ for individual 1 implied by the modified score determination rule in Equation (11). We can think of the distribution of $\tilde{z}_1$ induced by Equation (11) as the following mixing distribution:

$$\tilde{z}_1 = \begin{cases} z_1 + m + w & \text{w.p. } \Pr(y = 1 \mid z_1 < \bar{z}, x_1), \\ z_1 + w & \text{w.p. } \Pr(y = 0 \mid z_1 < \bar{z}, x_1). \end{cases} \tag{12}$$

Thus, the distribution of $\tilde{z}_1$ is obtained by taking a weighted average of the probability density function (pdf) of w, evaluated at the translated location parameters $z_1 + m$ or z_1, and weighted by the probabilities that $y = 1$ or 0, given that $z_1 < \bar{z}$. If one observes individual 1 in repeated trials, his manipulated score will accumulate mass to the right of $\bar{z}$ as long as $\Pr(y = 1 \mid z_1 < \bar{z}) > 0$. However, because of the additional density of w, this distribution will be smooth. The distribution of the score for individual 2 to the right of $\bar{z}$ will analogously be determined as in Equation (12), except the weighting probabilities are evaluated at the right of the cutoff, e.g., $\Pr(y = 1 \mid z_2 \geq \bar{z})$. The implication of the additional source of uncertainty can now be clarified. We see that randomness in w makes the distribution of $\tilde{z}$ smooth at $\bar{z}$ for *every* individual. The distribution across individuals is a weighted average of the distribution for each individual. Because the distribution for every individual is smooth and continuous at $\bar{z}$, the distribution of $\tilde{z}$ across all individuals will also be smooth and continuous at $\bar{z}$. Continuity implies that the distribution of types just to the left and right of $\tilde{z} = \bar{z}$ is the same, and hence it is as if there is local randomization at the cutoff. Consequently, the RD design is now valid. To the extent that such randomness is plausible in many

of w can have infinitesimally small variance. In practice, because RD estimates involve comparing consumers in a certain bandwidth on two sides of the cutoff, we assume that the distribution of w needs to have sufficiently large bandwidth relative to this bandwidth for the purpose of this discussion.

[9] Or, alternatively, the firm may induce some uncertainty. For example, the firm informs all consumers who are close to, or have just earned, a reward that they are enrolled in a lottery for potential miles.

[10] Note the implicit requirement that the randomness w has full support in the region $(\bar{z}-h, \bar{z}+h)$ of the cutoff. For instance, suppose the casino never changes its reward rules, and hence, those to the right of the cutoff (individual 2 in the example above) can never forfeit their earned reward. In this case w has only positive support for individuals 3 and 4, and the distribution of $\tilde{z}_2$ will be truncated below at $\bar{z}$. This generates a discontinuity in the distribution of $\tilde{z}$ across all individuals at $\bar{z}$, and RD is invalid. Continuity of the density of w is also important. Obviously, if the density of w is discontinuous, the required smoothing is not achieved.

contexts, RD designs may be considered very similar to quasi-randomized experiments (see, e.g., Lee and Lemiuex 2010). Note that the validity of the RD depends crucially on a *structural element of the model,* i.e., the beliefs of consumers about the score.

Uncertainty About the Cutoff. Although an astute customer may occasionally be able to perfectly know their score, it is much less likely that a customer knows the cutoffs in a given database marketing program. In this case, consumers may have a continuous belief distribution over the cutoffs. We now consider whether uncertainty about the exact cutoff at which rewards may be earned may be enough to ensure the validity of the RD design even in situations where consumers observe their score perfectly. Analyzing this aspect is more complicated, because it requires us to be more explicit about the choice-specific value functions that generate the function $f(m, z, R)$ in Equation (10).

We start with the basic model without other sources of randomness (i.e., no w); i.e., $\tilde{z} = z + my$. To define a customer's selection/play decision, we begin by specifying $f(m, z, R) = \mathscr{V}_1(m, z, R \mid \beta) - \mathscr{V}_0(m, z, R \mid \beta)$, where the choice-specific value functions, $\{\mathscr{V}_1, \mathscr{V}_0\}$, are defined as

$$\mathscr{V}_1(m,z,R|\beta) = v_1(R|\beta) + \delta\mathbb{E}\mathscr{V}(\tilde{z},R|m,z,y=1) + \varepsilon_1,$$
$$\mathscr{V}_0(m,z,R|\beta) = 0 + \delta\mathbb{E}\mathscr{V}(\tilde{z},R|m,z,y=0) + \varepsilon_0.$$

Here, $v_1(\cdot)$ is the deterministic component of the per-period value from playing that depends on whether the consumer has a reward or not, and the εs are stochastic unobservables (to the econometrician) as before. The consumer's discount rate is δ. The expected future value from choosing the action y is

$$\begin{aligned}&\mathbb{E}\mathscr{V}(\tilde{z}, R \mid m, z, y)\\ &= \iint \big[\mathscr{I}(\tilde{z} \geq \bar{z})[\max\{v_1(R \mid \beta) + \eta_1, \eta_0\}] + \mathscr{I}(\tilde{z} < \bar{z})\\ &\quad \cdot [\max\{v_1(0 \mid \beta) + \eta_1, \eta_0\}]\big] d\mathscr{F}_\eta(\eta_1, \eta_0)\, d\mathscr{G}_{\bar{z}}(\bar{z}). \quad (13)\end{aligned}$$

That is, if the consumer is able to cross the required cutoff by choosing y today, $\mathscr{I}(\tilde{z} \geq \bar{z}) = 1$, and he will have the reward R and $\tilde{z} = z + my$ tomorrow. If he chooses to play tomorrow, he obtains payoff $v_1(R \mid \beta) + \eta_1$; otherwise, he obtains only η_0. The value from the best-possible action tomorrow conditional on earning the reward is the maximum of these two payoffs. If, on the other hand, he is unable to cross the cutoff by selecting to play today, $\mathscr{I}(\tilde{z} < \bar{z}) = 1$, the value from the best-possible action tomorrow is analogously the maximum of the two payoffs but evaluated at $R = 0$. However, $\bar{z}$ and $\eta = (\eta_1, \eta_0)$ are unknown at the time of selection. Hence, the future value involves integrating out $\bar{z}$ and η over the consumer's beliefs over these variables. In Equation (13) the consumer's beliefs over the cutoff $\bar{z}$ is represented by a continuous density, $\mathscr{G}_{\bar{z}}(\bar{z})$, and his beliefs over the random shocks η by the density $\mathscr{F}_\eta(\eta_1, \eta_0)$. Moving the integration into the brackets, and noting that $\Pr(\tilde{z} \geq \bar{z} \mid m, z, y) \equiv \mathscr{G}_{\bar{z}}(\tilde{z})$, we can write Equation (13) as

$$\begin{aligned}&\mathbb{E}\mathscr{V}(\tilde{z}, R \mid m, z, y)\\ &\quad = G_{\bar{z}}(\tilde{z})\mathbb{E}_\eta[\max\{v_1(R \mid \beta) + \eta_1, \eta_0\}]\\ &\qquad + [1 - G_{\bar{z}}(\tilde{z})]\mathbb{E}_\eta[\max\{v_1(0 \mid \beta) + \eta_1, \eta_0\}], \quad (14)\end{aligned}$$

where $G_{\bar{z}}(\bar{z})$ represents the cumulative density function of the consumer's beliefs about $\bar{z}$.

With some abuse of notation, let $\Omega(z+m, z \mid R, \beta) = \mathbb{E}\mathscr{V}(\tilde{z}, R \mid m, z, y = 1) - \mathbb{E}\mathscr{V}(\tilde{z}, R \mid m, z, y = 0)$, the relative expected future value from selection versus not. We can evaluate Equation (14) at $y = (1, 0)$[11] to obtain

$$\begin{aligned}\Omega(z+m, z|R,\beta) &= [G_{\bar{z}}(z+m) - G_{\bar{z}}(z)]\\ &\quad \times \big[\mathbb{E}_\eta \max\{v_1(R|\beta)+\eta_1, \eta_0\}\\ &\quad - \mathbb{E}_\eta[\max\{v_1(0|\beta)+\eta_1, \eta_0\}]\big]. \quad (15)\end{aligned}$$

Intuitively, Equation (15) implies that the future component of the incentive to select/purchase is the difference in expected utility under treatment and not under treatment, weighted by the increase in the probability of receiving treatment that is due to adding m to the score, i.e., $G_{\bar{z}}(z+m) - G_{\bar{z}}(z)$. The decision to select is given now as

$$\begin{aligned}\Pr(y=1) &= \Pr(v_1(R \mid \beta) + \delta\Omega(z+m, z \mid R, \beta)\\ &\quad + \varepsilon_1 - \varepsilon_0 > 0). \quad (16)\end{aligned}$$

We can represent the distribution induced by this type of selection on the manipulated score $\tilde{z}$ as

$$\tilde{z} = \begin{cases} z + m & \text{w.p. } \Pr(y = 1 \mid m, z, R), \\ z & \text{w.p. } \Pr(y = 0 \mid m, z, R). \end{cases} \quad (17)$$

Proposition 1. *The distribution of $\tilde{z}$ is continuous at the true cutoff. Hence, the RD is valid.*

Proof. First, fix the value of the cutoff actually used by the firm at $\bar{\mathfrak{z}}$. Now note that the density of $\tilde{z}$ will be continuous at the true cutoff $\bar{\mathfrak{z}}$ if the mass of $\tilde{z}$ that piles up to the left and right of $\bar{\mathfrak{z}}$ as a result of selection is the same. This will be the case if (1) $\Pr(y = 1 \mid m, z, R)$ is the same just to the left and to the right of $z = \bar{\mathfrak{z}} - m$, and (2) $\Pr(y = 0 \mid m, z, R)$ is the same just to the left and to the right of $z = \bar{\mathfrak{z}}$. Note from Equation (15) above that z affects $[G_{\bar{z}}(z + m) - G_{\bar{z}}(z)]$. Hence, (1) and (2) will be satisfied if the limit of $[G_{\bar{z}}(z+m) - G_{\bar{z}}(z)]$ from the left and the right is the same at $z = \bar{\mathfrak{z}} - m$ and $z = \bar{\mathfrak{z}}$. To see that this is the

[11] Remember that at $y = 0$, $\tilde{z} = z$, and at $y = 1$, $\tilde{z} = z + m$.

case, note that $G_{\bar{z}}(z)$ is the *marginal* cumulative distribution function of the random variable $\bar{z}$ evaluated at the value z prior to selection. Because consumers do not know the true $\bar{\mathfrak{z}}$, this function is continuous at all z (by the primitive assumption), including at $z=\bar{\mathfrak{z}}-m$ and $z=\bar{\mathfrak{z}}$. The difference between two continuous functions is also continuous. Hence, $[G_{\bar{z}}(z+m)-G_{\bar{z}}(z)]$ is also continuous. □

It is interesting to contrast this with the situation where the true cutoff is known. In the typical frequency reward program, $\mathcal{G}_{\bar{z}}$ has mass 1 at the true value of $\bar{\mathfrak{z}}$, because the firm communicates cutoffs to customers to incentivize purchase. In this case, note that

$$\lim_{z\to\bar{\mathfrak{z}}^-} G_{\bar{z}}(z+m)=1 \quad \text{and} \quad \lim_{z\to\bar{\mathfrak{z}}^-} G_{\bar{z}}(z)=0 \tag{18}$$

because selection moves a customer from no treatment to treatment *with perfect certainty* if he moves the score by m from below the known cutoff (contrast this with the case when the cutoff is unknown, where this cannot be known for sure). From the right of $\bar{\mathfrak{z}}$, we have that

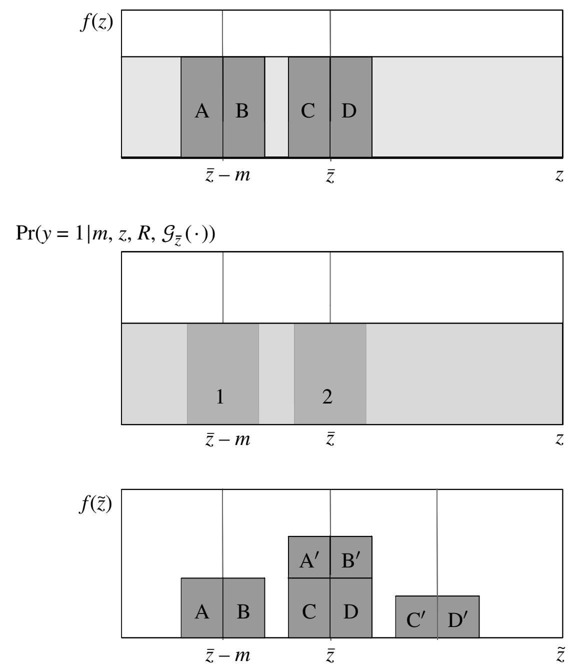
$$\lim_{z\to\bar{\mathfrak{z}}^+} G_{\bar{z}}(z+m)=1 \quad \text{and} \quad \lim_{z\to\bar{\mathfrak{z}}^+} G_{\bar{z}}(z)=1 \tag{19}$$

as the consumer on the right gets the reward for sure. Hence, with a known cutoff, $G_{\bar{z}}(z+m)-G_{\bar{z}}(z)$ jumps from 1 to 0, as one moves from the left to the right of $\bar{\mathfrak{z}}$. In essence, the smoothing generated by the continuity of the density $\mathcal{G}_{\bar{z}}(\cdot)$ is lost.

This intuition is depicted in Figure 2, which is analogous to Figure 1, which depicted the situation with a known cutoff. In Figure 2, we assume for the sake of simplicity that z is uniformly distributed and that the uncertainty about the cutoff also has a uniform distribution. In other words, the true cutoff $\bar{\mathfrak{z}}=\bar{z}+w$, where $w\sim \text{Uniform}[-\bar{w},\bar{w}]$. Consumers do not know what w is but know its distribution. Figure 2 shows that the induced distribution of $\tilde{z}$ is smooth, thereby ensuring the validity of the RD.

Uncertainty About m**.** Another form of uncertainty could be uncertainty about m, or in other words, the effect of the selection decision on the score. We discuss this situation in Appendix B for the purpose of keeping this brief, but we demonstrate there that this kind of uncertainty is insufficient to make the RD design valid in the presence of self-selection. This underscores the importance of considering the specific nature of uncertainty and its implications on consumer behavior to assess whether uncertainty or randomness can more generally resolve the identification issues in an RD context in the presence of selection.

Figure 2 Selection with an Unknown Cutoff



Notes. The top panel depicts a continuous distribution of customers along the score, $f(z)$. The middle panel depicts the probability of y, which adds m to the score, z. R is a reward if $z+m>\bar{z}$. $\mathcal{G}_{\bar{z}}(\cdot)$ represents the agent's beliefs about the unknown cutoff. Assuming continuity in $\mathcal{G}_{\bar{z}}(\cdot)$, y is equally likely just above and below the cutoff, as well as just above and below $\bar{z}-m$. The bottom panel depicts how the continuous incentives in the middle panel led the continuously distributed individuals from the top panel to be continuously distributed around the cutoff $\bar{z}$. There are four relevant types of customers within a small bandwidth of the cutoff: A′ and B′ customers were equally likely to choose y because of their uncertainty about the cutoff, as were C and D. The heights of the stacked boxes on either side of $\bar{z}$ are equal, demonstrating the continuity that uncertainty about the cutoff ensures.

4.1.3. What Is the Composition of Agents for Whom Treatment Is Measured? Although the identification conditions laid out in §2 define when a treatment effect can be measured for those agents clustered around the cutoff, the composition of those agents can be very difficult for researchers to know. Consider the previous example of uncertainty about the cutoff as depicted in Figure 2. Graphically, we depicted the types of agents who move to a neighborhood of the cutoff as a mix of A/B and C/D. However, we cannot characterize who A/B or C/D types are or know the relative proportion of A/B versus C/D consumers without actually writing down the true structural model and belief structure, $\mathcal{G}_{\bar{z}}(\cdot)$, of each agent type. Beliefs are well known to be elusive to researchers. Therefore, we have a treatment effect that is averaged across a sample of individuals; we cannot hone in on the specific subpopulation for whom this treatment effect is relevant without further assumptions. Although this challenge is not solvable without a structural model and the elusive belief data, we suggest using applications where there are many cutoffs to learn about the distribution of treatment effects across types. Our empirical application of casino promotions includes five separate cutoffs, providing us with the ability to find the treatment effect of casino promotions for consumers of different profiles, albeit locally at these cutoffs.

4.1.4. Discussion. The above analysis documents that traditional reward programs where the cutoffs are communicated to customers are not valid RD applications. However, targeted marketing based on purchase histories in which there exists uncertainty about the program, the scores or cutoffs are viable RD applications, as long as consumers have a continuous distribution of beliefs about the uncertainty. Uncertainty about the value of rewards does not restore the validity of the design. In our casino application below, identification of the RD estimator is provided by uncertainty about the score as well as that about the cutoff. Customers are unlikely to know their score because their past gaming behavior is affected by their luck and the casino's algorithm for adjusting their expected worth for luck. This score was not communicated to consumers either. Furthermore, the program involves offers sent to customers based on cutoffs that were not provided to the public. Customers may not even know that they would get these offers or that they differ from other customers in the nature of the offers they receive, but even if they suspected this, they would not know what cutoff the firm used for determining preferential treatment.

4.2. Targeted Promotions E-mailed by a Casino

In this section, we apply the regression discontinuity approach to assess a casino's database marketing program. Two reasons make our application to marketing activity at casinos compelling. First, targeted marketing is an important component of customer management at casinos (Lal and Carrolo 2001). Second, casino-based applications are particularly data rich and thus well suited for application of nonparametric methods. Casinos typically send offers to casinos either by direct mail or e-mail, offering packages that include discounted room rates, show tickets, dining credits, and promotional credits. Consumers are classified as "low rollers" and "high rollers," with more lucrative offers to the high rollers. It is important to note that the classification of consumers into low or high rollers, and the subsequent allocation of promotional offers to these tiers, is determined on the basis of a one-dimensional continuous score, referred to as the "average daily win" (henceforth, ADW). The ADW is the casino's best estimate of the average revenue the casino can earn from the customer per day of his visit, after controlling for luck.[12] We have access to the ADW for all customers in the data but not the proprietary algorithm that generated the ADW. In addition, we also have access to the cutoff of the ADW on the basis of which customers are sorted into tiers. Although consumers are aware that more play will move them into higher tiers and earn them more comps, they are unaware of the exact definition of the ADW or the ADW-specific cutoffs that generate sorting into tiers. This aspect mitigates selection concerns in this application. In Table 2 we provide an example of a casino mailing to 79,419 customers in which the offer depends on the casino's calculation of the ADW it expects from the customer.[13]

We observe four such mailings between January and September 2006. All mailings are identical except in terms of the show ticket offerings, which vary across mailings. Furthermore, the first mailing is only sent to the top two tiers. Important for the subsequent analysis, the pattern of higher tiers obtaining superior offers reflected in Table 2 holds across mailings (i.e., tiers 0 and 1 receive one offer, and tiers 2–5 receive a different offer that is inferior to the one received by the top two tiers). This systematic targeting is one reason why comps and subsequent play will be positively correlated.

We wish to measure the causal effect of comps and promotions. We consider two outcome variables that are relevant to the casino—namely, whether the customer visited the casino (*Trip*) and the casino's expected win from the customer (*Theoretical win*, or Theo for short). The theoretical win is similar to the ADW in that it adjusts for the customer's luck. It differs in that it recalculates the spending on a given occasion as opposed to providing a measure of the expected spending on any given day. These are summarized, by tier, in Table 3. We see from Table 3 that customers in the top two tiers arrive with about 23% probability, whereas customers in the bottom tier only arrive with 7.5% probability. There are also substantial differences in spending by tier, with the bottom tier having theoretical wins averaging only $10, whereas customers in the top tier have theoretical wins that average $617.

[12] The control for luck incorporates the role of randomness in past play behavior. For example, two identical customers willing to spend $100 may actually spend very different amounts during the day if one customer won $1,000 on their first play and the other lost all $100 on their first play.

[13] Note that the highest rollers, customers with an ADW above $2,500, are not included, because the casino deals with such customers on a one-to-one basis. Customers with an ADW less than $50 are sent e-mails but do not receive special offers such as discounted rooms or credits.

Table 2 Example of Tiered E-mail Promotions Sent by the Casino

Tier	Minimum ADW ($)	Maximum ADW ($)	Price of standard room ($)	Price of room on value day ($)	Show tickets	Dining credits	Promotional credits	Individuals mailed
0	1,000	2,500	0	0	2 free	100	500	3,600
1	500	999	0	0	2 free	50	300	8,975
2	300	499	0	0	2 for $22	50	50	12,848
3	200	299	99	0	2 for $ 22	25	50	12,249
4	100	199	159	79	2 for $22	25	50	23,116
5	50	99	179	99	2 for $ 22	25	25	18,631

Table 3 Summary Statistics of Casino Outcomes by Tier

	ADW range ($)	*N*	Mean	Std. dev.
Trip				
Tier 0	1,000 to 2,500	18,236	0.229	0.420
Tier 1	500 to 999	45,219	0.229	0.420
Tier 2	300 to 499	48,165	0.171	0.377
Tier 3	200 to 299	43,932	0.126	0.332
Tier 4	100 to 199	90,747	0.083	0.276
Tier 5	50 to 99	83,355	0.075	0.263
Theoretical win				
Tier 0	1,000 to 2,500	18,236	$616.70	$2,576.50
Tier 1	500 to 999	45,219	$328.17	$997.59
Tier 2	300 to 499	48,165	$135.52	$514.04
Tier 3	200 to 299	43,932	$64.11	$295.45
Tier 4	100 to 199	90,747	$21.68	$150.88
Tier 5	50 to 99	83,355	$10.08	$69.86

4.2.1. Analysis: Correlational Effects. One obvious pattern from Table 3 is that both outcome variables are increasing in the tiers. A pure correlational analysis that does not control for this targeting rule would pick up this positive correlation and falsely infer it as an effect of the promotion. As a benchmark, we start by regressing the outcome variables on tier fixed effects, which implicitly capture the effect of changing a promotion from the base tier to those for that tier. We pool observations across all four mailings while including period fixed effects to account for differences across the timing of mailings. Table 4 presents the results.

The OLS estimates with visit and theoretical win as the dependent variables are listed in the last two rows of Table 4. The visit variable is coded as 1 in case of a visit and 0 otherwise.[14] To obtain unconditional (of visit) effects, the Theo variable for a customer who does not visit is included as a 0 in the regression. For ease of interpretation, the incremental difference in the actual promotions when going from tier τ to $(\tau+1)$ are also listed in the top panel. Looking at Table 4, it would appear that the promotions have strongly positive effects, with higher tiers having higher visit probabilities and higher theoretical wins. Moving to a nonlinear model would not fundamentally alter these results because they are driven by the underlying correlations between the tiers and the two dependent variables.

4.2.2. Analysis: Causal Effects. The OLS estimates reported in Table 4 are problematic for two reasons. First, the estimation does not control for the targeting employed by the casino, thereby overstating the effect of the promotion. This induces a classic endogeneity bias into measurement. Second, by imposing a (linear) functional form to hold globally across all types of customers and tiers, the estimator leverages the information from all of the observations to learn about the promotional effects, despite the fact that the variation in the offers really only occurs at ADW levels of $100, $200, $300, $500, and $1,000. This results in misspecification bias of an unknown form. We address both issues with the RD estimator. Table 5 presents estimates of the RD estimator applied to these data. As in the previous analysis, we pool across all four mailings to estimate the treatment effects using the RD design while controlling for period fixed effects.[15] Following the suggestion of Imbens and Lemieux (2008), we estimate these nonparameterically using a rectangular kernel, using observations within a bandwidth of size h on either side of each cutoff. Because there are six different tiers and five different ADW cutoffs, we estimate five different regression discontinuity specifications focusing on each tier cutoff. In practice, this turns out to be least squares estimators using only observations lying in the neighborhood defined by the bandwidth on either side of each of the cutoffs. An advantage of this approach is that the standard errors of the rectangular kernel are the same as robust standard errors available for least squares. We also estimated the RD using

[14] We choose OLS to illustrate the pitfalls of not accounting for the endogeneity. More sophisticated statistical specifications (e.g., count models, discrete link functions) will not change the message from this analysis as long as they interpret the full correlation of outcomes and tier status as effects of the promotion.

[15] As Lee and Lemieux (2010) note, the panel aspect of the mailings does not add anything to the identification of the RD estimator, except to reduce the variance of the estimate. Furthermore, we do not expect the repeated treatment of individuals to affect the estimation because we expect that the firm has accounted for past promotion/treatment when calculating an ADW.

Table 4 OLS Regressions of Casino Outcomes on Tier

	Changes in promotions when moving between tiers				
ADW cutoff	5 to 4: $100	4 to 3: $200	3 to 2: $300	2 to 1: $500	1 to 0: $1,000
Promotion categories					
Regular room	$179 to $159	$159 to $99	$99 to free	—	—
Room on value day	$99 to $79	$79 to free	—	—	—
Show tickets	—	—	—	B to A	—
Dining credits	—	—	$25 to $50	—	$ 50 to $100
Promo credits	$25 to $50	—	—	$50 to $300	$300 to $500
OLS					
Visit	0.008**	0.05***	0.09***	0.15***	0.15***
Theoretical win	10.99**	53.01***	124.16***	196.67***	487.34***

***Significant at the 1% level; **significant at the 5% level.

local linear regression (not reported) and found the results to be similar. As in all nonparametric applications, an important aspect of the estimation is correct choice of the bandwidth. We choose the bandwidth by cross validation. We conduct a search for the bandwidth of an ADW that minimizes the mean squared error across the five cutoffs. We also present the estimates for an arbitrarily chosen bandwidth of $20.

Referring to Table 5, we see that effects of the marketing program on visit probabilities are all insignificant except for the transition from tier 2 to tier 1 (ADW = $500) and from tier 4 to tier 5 (ADW = $100). However, both of these negative effects are not robust to changes in the size of the bandwidth. We therefore conclude that the casino offers are neither increasing nor decreasing the probability that a customer visits the casino. This is surprising because there is substantial variation in the price of a room. However, elasticities may be low because of other substantial costs of visiting a casino in Las Vegas or because the room is only a small part of the expenses of these gamblers.

We see similar patterns when analyzing the theoretical win of the customers. Most notably, most of the effects are negative or insignificant. A negative effect is plausible, because provision of dining credits and show tickets can draw customers away from the slots and the gambling tables. We conjecture that these credits may substitute for actual gaming. Using a bandwidth of $20 for the ADW, we see that there is a positive, but insignificant, jump in theoretical win at ADW = $500. However, under the optimal bandwidth, which is smaller, we find that the evidence reverses: there is a large negative effect of the increased promotions when moving from tier 2 to tier 1. This reversal also underscores the importance of considering robustness to choices of bandwidth, and the dangers of pooling data across very dissimilar observations. To reiterate the point, we present a plot of theoretical win at ADW = $1,000, where there is a change of tiers from 1 to 0.

Overall, although the general relationship between theoretical win and ADW is positive, in the limit exploring the variation in a neighborhood of the discontinuities, the effects are either not significant or negative. One interpretation of the results is that comps and hotel credits are not significant drivers of

Table 5 RD Estimates of Promotional Effects

	Changes in promotions when moving between tiers				
ADW cutoff	5 to 4: $100	4 to 3: $200	3 to 2: $300	2 to 1: $500	1 to 0: $1,000
Promotion categories					
Regular room	$179 to $159	$159 to $99	$99 to free	—	—
Room on value day	$99 to $79	$79 to free	—	—	—
Show tickets	—	—	—	B to A	—
Dining credits	—	—	$25 to $50	—	$50 to $100
Promo credits	$25 to $50	—	—	$50 to $300	$300 to $500
Effect on outcome (bandwith minimizes MSE)					
Visit	−0.005	0.018	−0.008	−40.078**	0.043
Theoretical win	−4**	5	−41	−207**	−556
Effect on outcome (bandwith = $20)					
Visit	−0.015**	0.014*	0.008	0.029	0.047
Theoretical win	4**	8	−11	26.06	−333

**Significant at the 5% level; *significant at the 10% level.

consumer's decisions to visit or to play at the casino, but competition forces the casino to continue marketing despite the low returns. Alternatively, comps affect long-run considerations such as building loyalty to the casino, even though in the short run they do little to shift visits or play. The other interpretation of the results is that they provide some evidence of short-term inefficient marketing decision making at the firm. The bottom line is that the analysis reveals that the marketing program we analyzed at the casino is not working: the comps are either ineffective or losing money. This is a significantly different conclusion from the previous correlational analysis. Nevertheless, it is possible that such an ineffective program may be a necessary competitive response to other casinos that offer similar programs. Our analysis can only speak to the profitability of the parameters of the program, as opposed to the need for the program as whole.

4.3. Discussion

The RD approach in our two applications revealed several null, or negative, effects. The reader should not infer that the approach will always yield null effects; the right inference is that much of the observed positive correlation in the data in both contexts is due to the targeting rule and not due to the marketing activity. It is important to point out that null effects for the direct mail company (and the casino) are not necessarily bad. A null effect at a cutoff defining a tier implies that the firm can do just as well in terms of marketing by merging the adjacent tiers together. Similarly, large differences in effects across various cutoffs imply that the firm may want to consider a finer classification of consumers into tiers on the basis of the scoring variable. An interesting open question is what the *optimal* score should be. It is unlikely that researchers will be able to answer this question credibly unless they have detailed data on the constraints or information asymmetries that forced firms to sort consumers on the basis of the observed score in the first place, rather than on the basis of their full range of characteristics and history. Some progress has been made in limited contexts. For instance, Huang (2009) shows that when willingness to pay is log-normally distributed, the optimal sorting score may be determined on the basis of their explanatory power in regressions of willingness to pay. However, this remains an open question for future research.

5. Purchase Timing as a Score Variable

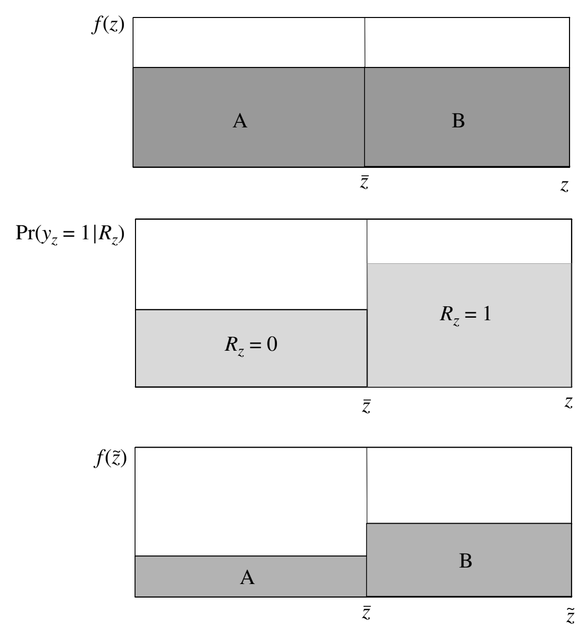
RD naturally extends to situations where the score variable is, or relates to, time. An RD with time as a score is a "before–after" design. Before–after contexts are typically plagued by the concern that outcomes in the "after" period arise because of time-varying unobservables that are unrelated to the treatment. The RD design is able to address this concern because these time-varying unobservables are unlikely to drop discontinuously at the exact moment of the treatment. Hence, the discontinuity in outcomes after the treatment is not due to these unobservables. Researchers in marketing and industrial organization have used RD with time as a score to analyze causal effects of price promotions (Busse et al. 2006, 2010; Fong et al. 2010) and service guarantees (Chen et al. 2009).

We discuss two aspects of the design that are relevant to time-based RDs. The first relates the nature of the outcome being measured and the second, to demand-side effects flowing from the type of product considered.

5.1. Nature of Outcome Being Measured

To understand this point, consider analyzing data from an initial decision and then a conditional decision by consumers. Assume a continuous distribution of customers over time, $f(z)$ (the top panel in Figure 3), where z denotes time. Let the first decision be indicated by $y_z = \mathcal{I}(g(R_z \mid \theta) + \varepsilon_z > 0)$. To fix ideas, call it the store visit decision. R_z denotes whether there is

Figure 3 **Time as a Score**



Notes. The top panel depicts a continuous distribution of customers over time, $f(z)$. Comparing A and B at $\bar{z}$, we see that continuity is maintained at the cutoff. The middle panel depicts how a discontinuous change in an incentive R_z at $\bar{z}$ results in a discontinuous change in the probability of an initial decision (e.g., a store visit). RD can validly estimate any such effect because of the continuous distribution of the score, $f(z)$. However, RD is invalid when estimating any outcome conditional on the initial decision, because the score then becomes the timing of the first decision, $\tilde{z}$. We see in the bottom panel that the manipulated score (a combination of the top two panels) changes discontinuously from A to B at $\bar{z}$.

an increased incentive (e.g., a price discount for some product at the store) that makes customers more likely to visit at time z. $g(R_z \mid \theta)$ is the nonstochastic portion of a customer's payoff for this choice. This payoff is similar to Equation (10) except that time is the score. θ includes parameters for the store choice decision as well as parameters for any conditional decisions, such as purchase. ε_z is an unobservable (to the researcher) that affects the store visit decision. The potential for RD arises when the incentive R_z discontinuously changes at $z = \bar{z}$. Such a discontinuity in the incentive induces a discontinuity in the probability of the store visit as depicted in the middle panel of the figure. The treatment effect of the change in R_z on the store visit decision is identified by RD because the distribution of the score $f(z)$ is continuous at the cutoff.

Next, consider any outcome conditional on having visited the store, such as whether or not to purchase. In terms of empirical implementation, this would entail analyzing data only on those consumers who visited the store. The RD estimator in this case will analyze the purchase decisions of only those customers that entered the store just before and after the price change, i.e., consumers for whom $g(R_z \mid \theta) + \varepsilon_z > 0$ for $z = \bar{z} \pm h$, where h is a small bandwidth of time. The manipulated score in this case is the timing of the store visit, $\tilde{z}$. Together, the distribution of the initial score (top panel) and the store visit probability (middle panel) generate a manipulated score, $f(\tilde{z})$ (bottom panel). The distribution of this manipulated score is discontinuous because $g(R_z \mid \theta)$ changes discontinuously, and consequently, $\Pr[g(R_z \mid \theta) + \varepsilon_z > 0]$ and $f(\tilde{z})$ also become discontinuous. As shown before, a discontinuity in the score invalidates RD. When measuring treatment effects on an outcome conditional on visit, the only case in which RD with time as a score is valid is when *all* consumers visit the store with no precise knowledge of whether the promotion is in effect or not—both before and after the promotion begins. In such a case, the first (visit) decision is unaffected by the promotion and RD is valid.

Finally, to close the discussion, consider estimating the effect of the promotion on the joint outcome of visiting the store and purchasing. In terms of empirical implementation, this would entail analyzing the demand data of all consumers who could potentially visit the store (i.e., the demand of those who did not visit should be included as a zero in the estimation data set). The relevant score is now time itself, and the RD validly measures the effect of the promotion on demand. The key insight here is that what is being measured matters: RD designs work when the outcomes are unconditional on timing decisions by consumers but may not when analyzing outcomes conditional on a timing chosen by customers, depending on the nature of consumer beliefs.

5.2. Nature of Product and Dynamics

The second aspect to consider is the nature of the product for which we measure promotion effects. Busse et al. (2006) point out that if the good is durable or storable, forward-looking and price-sensitive consumers will tend to queue up after previous promotions as in a Sobel-style model (Sobel 1984). The promotion occurs at time $\bar{z}$. At time $\bar{z} - h$, only those customers who cannot wait for the next promotion visit the store (and make purchase decisions). However, at $\bar{z} + h$, all customers who would have potentially bought since the past promotion visit. Revisiting our analysis, the store visit effect is identified. However, the purchase decision conditional on visiting at $\bar{z} - h$ versus $\bar{z} + h$ compares two different types of customers: price-insensitive or impatient consumers who buy prior to promotion and price-sensitive or patient consumers who wait for the promotion. Generally speaking, forward-looking behavior and expectations accentuate issues related to dynamic selection. Assuaging dynamic effects is important to validate time-based RD designs.

We stated above that RD estimates a valid treatment effect when time itself is the score but not when the score is a timing chosen by the customer. However, a vexing issue is that both cases measure effects that are relevant to a subpopulation selected by dynamic considerations: the RD's treatment effect is a function of the distribution of the state variables for the consumers who visit. The consumer's states are a function of the firm's past actions, which makes comparing the estimated treatment effects across studies difficult unless the firm's history is held constant across studies. For instance, in a storable good model, the key state affecting purchases is inventory, which is a function of past promotions. Using historical data on promotions, Fong et al. (2010) find that the RD estimand over time does not perform well compared to a randomized experiment in which base prices across products and stores are changed for storable grocery goods. The RD application is valid because time itself is the score, not a timing chosen by customers. However, the measured treatment effects are a function of inventories realized at the time of the temporary price changes (for the RD) or the base-price change (for the experiment). For the random experiment and the RD to yield the same results, the researcher would have to ensure the firm's promotion history prior to both the RD and the experiment are the same. This is difficult to ensure in practice. Of course, this is the case for comparing any two estimators with state-dependent treatment effects.

5.3. Discussion

Overall, our discussion has highlighted the different aspects an analyst contemplating an RD application in a targeted marketing context needs to consider in order to ensure that the design is valid. Even though the RD design fits into the "reduced-form causal effect" literature, we believe that its validity in marketing situations cannot be established without specifying a clearly articulated structural model of behavior. Our detailed analysis underscores the importance of formally specifying consumer's information sets and incentives to establish the validity, or lack thereof, of an RD-based analysis. Furthermore, we have illustrated in both purchase history targeting and time as a score that even when the RD estimates a valid treatment effect, it can easily be for an unknown population whose selection is determined by how state variables and beliefs pass their way through a structural model.

6. Conclusions

This paper illustrates the use of regression discontinuity techniques in targeted marketing and industrial organization applications. To the best of our knowledge, we are unaware of other targeted marketing applications that have exploited the identification enabled by the rules of thumb and heuristics pervasive in marketing practice. These heuristics had previously been thought of as a "nuisance" issue that had to be dealt with in estimation by researchers or as evidence of inefficient marketing decision making by firms. Here, we show that the heuristics actually aid estimation by facilitating identification and are also useful to firms as they enable credible measurement of the return on investment on their marketing spends. We illustrate the approach using two empirical applications. Our quasi-experimental approach reveals negative effects of marketing mix variables that are not easily uncovered otherwise.

We expect our approach to control for the endogeneity to be used in conjunction with other approaches for understanding demand under targeting. Treatment effects obtained via the design may be combined with estimates from structural models to improve or audit results (e.g., see Khwaja et al. 2011 for the case of a matching estimator). Furthermore, parametric models of heterogeneous demand can provide the continuous representation of heterogeneity that firms could use to better define cutoffs when using group-level targeting. Better measures of heterogeneity will improve heuristic rules of thumb used for targeting, even if firms are unable to implement the individual-level policies that can be suggested by individual-level models. Parametric methods for solving the endogeneity that enable pooling are also essential in sparse-data situations. In data-rich environments, we hope this paper encourages further exploration of the use of nonparametric methods to facilitate optimal marketing mix allocation.

Acknowledgments

The authors thank the review team, Anand Bodapati, Florian Zettlemeyer, seminar participants at Goethe University (Frankfurt), Indian Institute of Management (Bangalore), NYU Stern School of Business, Olin Business School at Washington University in St. Louis, Yale School of Management, and at the Marketing Science (2007, 2009), SICS (2010), and QME (2010) conferences for their helpful feedback. All errors are their own.

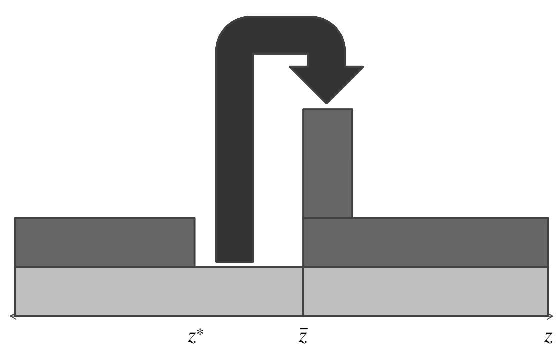
Appendix A. Heterogeneity in Types Does Not Mitigate Selection

Here, we demonstrate graphically why heterogeneity in types does not guarantee validity of the RD design in the presence of selection. Consider a situation where there are two types of consumers: one type has sufficiently high fixed costs such that they would not satisfy the fixed costs condition in §3.1.3. The second type of consumers has fixed costs such that a fraction of them with $z > z^*$ move to the right of the cutoff $\bar{z}$, with z^* as defined previously. Figure A.1 graphically depicts this situation. Consumers of the first type are represented by lightly shaded boxes. None of these consumers selects into the treatment group. Type 2 consumers are represented by darkly shaded boxes, and as can be seen in the figure, a proportion of these consumers selects into treatment. This is a situation where the limit of the outcome to the left and right of the cutoff ($\bar{z}$) exists. However, because some of the consumers of the second type have incentives to select to be on the right of the cutoff, there is a discontinuity of types and consequently a discontinuity in outcomes at the cutoff. Thus, even if there are some consumers of this second type, the conditions for identification of RD are violated.

Appendix B. Uncertainty About Value of Reward Does Not Mitigate Selection

Following the model sketched in §4.1.2, we consider a situation where there is some uncertainty in m itself; i.e., the consumer can choose y but does not know how much the score will move by his actions. One example may be a casino reward program where the consumer does not know the exact weights via which the casino translates his play at slots, tables, and other games into the score used for targeting. Another is of a hypothetical program where consumers who choose to purchase are entered into a lottery to earn points. Thus, m is zero with some probability and positive

Figure A.1 Selection Causes One of the Consumer Types to Move



with some other probability. A moment's reflection reveals that this sort of uncertainty does not make the RD design valid, because consumers to the right of the cutoff still have no incentives to gamble in order to earn the reward (they already have the reward). However, as long as m is positive with some probability, consumers just to the left of the cutoff will have an incentive to select that consumers to the right do not face. This will cause a discontinuity at the cutoff, invalidating the RD design.

To see this formally stated, consider a situation where m is discrete. Let the random variable $m = m$ (>0) with some probability p and $m = 0$ with probability $(1-p)$. Thus, the manipulated score can be rewritten as

$$\tilde{z} = \begin{cases} z + m \times y(m, z, R, \varepsilon) & \text{w.p. } p, \\ z & \text{w.p. } (1-p). \end{cases} \tag{B1}$$

Consider a consumer to the right of the cutoff. This consumer already has the reward and therefore has no incentive to select to fly. Now consider an individual with a score in the neighborhood of the cutoff and to the left of it. Specifically, this individual has a score

$$z_1 \in [\bar{z} - m, \bar{z}).$$

The manipulated score for this individual has the following distribution induced by Equation (B1):

$$\tilde{z}_1 = \begin{cases} z_1 + m & \text{w.p. } p \times \Pr(y = 1 \mid m = m, z_1, R, \varepsilon_1), \\ z_1 & \text{w.p. } 1 - [p \times \Pr(y = 1 \mid m = m, z, R, \varepsilon)]. \end{cases} \tag{B2}$$

As long as the conditions described in §4.1.1 hold, $\Pr(y = 1 \mid m = m, z_1, R, \varepsilon_1)$ is nonzero. Thus, if there is any probability p that m takes a positive value, the distribution of $\tilde{z}_1$ will accumulate more mass just to the right of the cutoff $\bar{z}$ relative to the left, both because of consumers initially on the right of $\bar{z} - m$ ending up to the right of $\bar{z}$ and consumers initially just to the left of $\bar{z}$ leaving and ending up to the right of $\bar{z}$. This will cause a discontinuity in the distribution of the score, invalidating the RD design.

Continuous m. One might argue that the discrete distribution of m drives the discontinuity in the distribution of the manipulated score. Hence, we now consider a change in the model above in which m has a continuous distribution. Let the pdf of m be $g(\cdot)$. Modifying the discussion for a discrete m, the manipulated score for the individual is

$$\tilde{z}_1 = \begin{cases} z_1 + m & \text{w.p. } \Pr(y = 1 \mid m, z_1, R, \varepsilon_1) g(m), \\ z_1 & \text{w.p. } \Pr(y = 0 \mid m, z, R, \varepsilon) g(m). \end{cases} \tag{B3}$$

Once again, as long as $g(\cdot)$ is any continuous density and the conditions described in 4.1.1 hold, the incentives of consumers in the region $[\bar{z} - m, \bar{z})$ to select in order to receive the reward (i.e., $\tilde{z}_1 = z_1 + m$) will be higher than those in regions (1) and (3). This, as discussed earlier, causes a discontinuity in the distribution of the manipulated score $\tilde{z}$, invalidating the RD approach.

Appendix C. Some Examples of Potential RD Applications in Marketing

Regression discontinuity designs could potentially be applied to a variety of marketing contexts. Table C.1 gives examples of some potential RD applications in marketing.

Table C.1 Examples of Potential RD Applications in Marketing

Industry/Firm	Score	Treatment	Outcome	Discontinuity
Cataloging	RFM (recency, frequency, monetary) value	Targeted catalogs to zip codes	Purchase, response	Catalogs are be mailed to households based on cutoffs of RFM scores.
Credit cards	FICO (Fair Issac Corporation) credit score	APR, credit lines	Sign-up for card, default	APRs and credit limits are assigned based on cutoffs of FICO scores.
Insurance	Company-specific risk score	Premiums	Sign-up for policy, renewal, satisfaction	Premiums are assigned based on cutoffs of company-specific risk scores that capture driving, health risk.
Services	Inactive time	New offers, initiate contact	Whether customer churns	In many services settings, prolonged periods of inactivity by customers may trigger promotions to prevent churn. Examples include value-added account offers by banks. In such settings, the effect of the promotion can be measured by comparing outcomes for those on either side of the discrete cutoffs of inactive time.
Rentals (e.g., Netflix)	Time since last rental	Promotions	Retention	Free-rental offers may be assigned to inactive customers based on cutoffs of inactive time.
Banking (e.g., Mint.com)	Amount in savings account	Suggestions for better investment options	Take-up of offer	Personal financial aggregator firms offer suggestions for higher-return investment options based on amounts deposited in lower-return options such as savings accounts.

Table C.1 (Cont'd.)

Industry/Firm	Score	Treatment	Outcome	Discontinuity
Airlines	Number of "Premier" passengers who arrived	Offer of Economy-Plus seat	Future flying of premier customers, satisfaction with mileage program	Airlines like United may upgrade their Premier status fliers to Economy-Plus seats, based on availability determined by the number of Economy-Plus seats sold and the number of Premier passengers on the flight. Premier passengers do not know the cutoffs and cannot self-select into flights.
Hotels	Current reward points	Offer upgrades	Future hotel stays	Upgrades (next-in-class rooms) often offered based on cutoffs of past history/accumulated points.

References

Angrist, J. D., V. Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quart. J. Econom.* **114**(2) 533–575.

Black, S. E. 1999. Do better schools matter? Parental valuation of elementary education. *Quart. J. Econom.* **114**(2) 577–599.

Busse, M., J. Silva-Risso, F. Zettelmeyer. 2006. $1000 cash back: The pass-through of auto manufacturer promotions. *Amer. Econom. Rev.* **96**(4) 1253–1270.

Busse, M. R., D. I. Simester, F. Zettelmeyer. 2010. "The best price you'll ever get": The 2005 employee discount pricing promotions in the U.S. automobile industry. *Marketing Sci.* **29**(2) 268–290.

Chay, K. Y., M. Greenstone. 2005. Does air quality matter? Evidence from the housing market. *J. Political Econom.* **113**(2) 376–424.

Chen, X., G. John, J. M. Hays, A. V. Hill, S. E. Geurs. 2009. Learning from a service guarantee quasi-experiment. *J. Marketing Res.* **46**(5) 584–596.

Cook, T. D., D. T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis for Field Settings*. Rand McNally, Chicago.

Drèze, X., J. C Nunes. 2011. Recurring goals and learning: The impact of successful reward attainment on purchase behavior. *J. Marketing Res.* **48**(2) 268–281.

Fan, J., I. Gijbels. 1996. *Local Polynomial Modeling and Its Applications*. Chapman & Hall, London.

Fong, N. M., D. I. Simester, E. T. Anderson. 2010. Private label vs. national brand price sensitivity: Evaluating non-experimental identification strategies. Working paper, Massachusetts Institute of Technology, Cambridge.

Hahn, J., P. Todd, W. Van Der Klaauw. 2001. Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica* **69**(1) 201–209.

Huang, K.-W. 2009. Optimal criteria for selecting price discrimination metrics when buyers have log-normally distributed willingness-to-pay. *Quant. Marketing Econom.* **7**(3) 321–341.

Imbens, G. W., T. Lemieux. 2008. Regression discontinuity designs: A guide to practice. *J. Econometrics* **142**(2) 807–828.

Khwaja, A., G. Picone, M. Salm, J. G. Trogdon. 2011. A comparison of treatment effects estimators using a structural model of AMI treatment choices and severity of illness information from hospital charts. *J. Appl. Econometrics* **26**(5) 825–853.

Lal, R., P. Carrolo. 2001. Harrah's Entertainment Inc. HBS Case 502-011, Harvard Business School, Boston.

Lee, D. S. 2008. Randomized experiments from non-random selection in U.S. House elections. *J. Econometrics* **142**(2) 675–697.

Lee, D. S., T. Lemieux. 2010. Regression discontinuity designs in economics. NBER Working Paper 14723, National Bureau of Economic Research, Cambridge, MA.

Lee, D., E. Moretti, M. J. Butler. 2004. Do voters affect or elect policies? Evidence from the U.S. House. *Quart. J. Econom.* **119**(3) 807–859.

Manchanda, P., P. E. Rossi, P. K. Chintagunta. 2004. Response modeling with non-random marketing mix variables. *J. Marketing Res.* **41**(4) 467–478.

Otter, T., T. J. Gilbride, M. G. Allenby. 2011. Testing models of strategic behavior characterized by conditional likelihoods. *Marketing Sci.* **30**(4) 686–701.

Porter, J. 2003. Estimation in the regression discontinuity model. Mimeo, Department of Economics, University of Wisconsin–Madison, Madison.

Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Sobel, J. 1984. The timing of sales. *Rev. Econom. Stud.* **51**(3) 353–358.

Thistlethwaite, D. L., D. T. Campbell. 1960. Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *J. Ed. Psych.* **51**(6) 309–317.

Van der Klaauw, W. 2008. Regression discontinuity analysis: A survey of recent developments in economics. *Labour: Rev. Labour Econom. Indust. Relations* **22**(2) 219–245.