

Nonparametric Estimation of Marketing-Mix Effects Using a Regression Discontinuity Design

Wesley Hartmann* Harikesh S. Nair† Sridhar Narayanan‡

This version: November, 2009[§]

Abstract

We discuss how regression discontinuity designs arise naturally in settings where firms target marketing activity at consumers, and discuss how this aspect may be exploited for econometric inference of causal effects of marketing effort. Our main insight is to use commonly observed discreteness 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 illustrate theoretically the conditions under which the RD estimator may be valid. Specifically, we argue 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 present two empirical applications: the first, to measuring the effect of casino e-mail promotions targeted to customers based on ranges of their expected profitability; and the second, to measuring the effect of direct mail targeted by a B2C company to zip-codes based on thresholds 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 fully control for the endogeneity induced by the targeting rule.

Keywords: regression discontinuity, nonparametric identification, treatment effects, targeted marketing, selection, endogeneity, casinos, direct-mail.

*Stanford GSB, Email: hartmann_wesley@gsb.stanford.edu

†Stanford GSB, Email: harikesh.nair@stanford.edu.

‡Stanford GSB, Email: sridhar.narayanan@stanford.edu.

§We thank seminar participants at IIM (Bangalore), Goethe University (Frankfurt), University of Rochester and at the Marketing Science Conferences (2007 & 2009), for their helpful feedback. All errors are our own.

1 Introduction

Targeted marketing is an ubiquitous element of firms' marketing strategies. The explosion of data on consumers has made it possible for firms to tailor prices, advertising and other elements of the marketing mix to specific consumers or groups of consumers (e.g., Rossi, McCulloch and Allenby 1996). The measurement of the causal effects of such targeted marketing is however tricky. A first-order complication arises as 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 the use of the Regression Discontinuity design to nonparametrically measure the causal effects of marketing effort, taking advantage of heuristic rules often used by firms for targeting.

Regression Discontinuity (henceforth RD) was first introduced by Thistlethwaite and Campbell (1960) in the evaluation literature (e.g. Cook and Campbell 1979, Trochim 1984), and has become increasingly popular in program evaluation in economics. Essentially, an RD design arises when treatment is assigned based on whether an underlying continuous *score* variable crosses a discrete threshold. The discontinuity induced by this treatment rule induces a discontinuity in the outcomes for individuals at the threshold. In an important paper, Hahn, Todd and Van der Klaauw (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 threshold. Under these conditions, observations immediately to one side of the threshold act as a valid control for observations on the other side, and RD achieves local quasi-randomization. This facilitates 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); and the social effects of imprisonment (Chen and Shapiro 2007) amongst others. In contrast, applications to Marketing and Industrial Organization contexts have been very sparse. These include Busse et al. (2006 and 2009) who measure 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 (2007), Imbens and Lemieux (2008) and Lee and Lemieux (2009) 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 policy-relevant history crosses a pre-specified threshold of a continuous score variable. For instance, catalogs might be mailed based on cutoffs of underlying ‘RFM’ score variables, credit card promotions may be given based on thresholds 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 thresholds, etc. The ubiquity of such trigger-rules generate a wealth of discontinuity-based contexts that facilitate nonparametric identification of marketing effects using an RD design, which have previously been unexploited in the literature. Estimation of the effect mainly involves obtaining nonparametric estimates of the mean outcomes from the left and from the right of the trigger values, and comparing these to the difference in marketing levels between the left and the right.

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 strategic consumer selection to delineate a set of viable and non-viable 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 therefore often be used to measure marketing effects under geographic targeting (i.e. high fixed costs of moving to receive the treatment), or situations where targeting is based on scores that cannot be changed, such as age-based marketing (i.e. infinite fixed cost of selection). Second, we present a detailed illustration of targeting in a canonical loyalty program in which selection is problematic, and discuss approaches to mitigate the selection bias. We illustrate that uncertainty about the exact score, the threshold, or both, or uncertainty about the details of the program more generally, can ensure the validity of the RD estimator even in the presence of selection.¹ The novel aspects of our analysis relates the identification restrictions to underlying primitives like agent preferences and beliefs as well as the econometrician’s assumptions about the extent to which agents have knowledge of stochastic errors entering the model. Finally,

¹For 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.

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 and the belief structures leading to these decisions. An important takeaway from these analyses is that the identification conditions for applying an RD estimator need to be evaluated with respect to an underlying structural model of behavior of agents.

We consider the RD design to be complementary to several alternative methods focused on uncovering causal effects. In targeted marketing applications in particular, there is a sense in which RD designs, and when available, may be more viable and credible than other methods. One popular alternative is to use instrumental variables. In practice, such variables are hard to obtain in targeting situations, and even if available, often do not have the variation required to estimate the treatment effect precisely, or rely on ad-hoc exclusion arguments for validity. Customer-side variables are hard to justify as valid instruments due to their likely effect on outcomes. Cost-side variables, which are often used as instruments in other contexts, are not feasible in the context of targeted marketing, since costs are typically segment and consumer specific and are functions of consumer characteristics, and hence cannot be considered excluded from the outcome regression. 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 (see Manchanda et al. 2004; Kao et al. 2005 who have outlined this approach). 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 (though see Ellickson and Misra 2007 for an approach that controls for selectivity nonparametrically). In the absence of such information, the analysis is sensitive to misspecification bias.

By contrast, the regression discontinuity approach does not depend on the use or validity of instrumental variables.² Further, only limited knowledge of the firm’s marketing allocation rule, i.e. details of the discontinuity, is required. The estimates obtained are also nonparametric, thus mitigating concerns about specification biases. A more subtle advantage is that the nonparametric approach takes advantage of extensive data only in the region where variation in the marketing mix is actually observed. The estimates so obtained, are local, but more credible. Parametric approaches impose functional forms that hold globally, and thus combine information from substantively different observations for inference. This enables inference from much less data, but at the cost of imposing global assumptions that may or may not hold locally, generating unknown biases. A further advantage of the RD estimator is that it is easy to implement. The estimator is also favorable to firms as they can take advantage of pre-existing discontinuous rules-of-thumb to identify causal

²Indeed, as shown by HTV, the “fuzzy” RD estimator has a local instrumental variable interpretation, in the sense of Imbens and Angrist (1994). In this sense, the RD design has a “built-in” instrument: whether the score is to the right or left of the threshold.

effects, which is less resource intensive than organizing randomized trials. Thus, on the whole, the approach is attractive to researchers due to its robustness and to practitioners due to its ease of use.

The main caveats for adopting the RD approach are three-fold. First, by its nonparametric nature, the estimator is data intensive and requires many observations on consumer behavior at the threshold; in sparse-data situations, parametric approaches are more suitable. Second, the estimator provides a local treatment effect which is relevant only for the subpopulation of consumers at the threshold, and not globally.³ 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 especially crucial. The HTV conditions on identification are stated in terms of continuity of counterfactual outcomes at the threshold. 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 agent behavior which considers how consumers sort at the threshold.

Based on the model analyses described above, we find two viable empirical applications of RD in the context of targeted marketing. The first application assesses marketing efforts directed by a casino towards members of its loyalty program and the second application measures the effect of direct mail for a B2C company. The casino targets its marketing efforts towards high-rollers. The firm's targeting rule is not observed in its entirety, but we know it is discontinuous based on thresholds on a score variable, which is the average level of gambling activity by the consumer in the past. These thresholds are not known to the consumers and hence they cannot self-select into preferential treatment. We observe both the score variable and the thresholds. In the second application, 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 threshold. In both applications, there is no feasible instrument, but RD is feasible. We estimate the effect of the marketing programs in both applications nonparametrically using local linear regression (Fan and Gijbels 1996). We find that in both cases, controlling for the endogeneity has large implications on the conclusions drawn from the analysis. In particular, we find in both cases that a naive estimate has an altogether different sign than the RD estimate.

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. Further, we point out that such rules-of-thumb, which have been typically treated as nuisance issues to be dealt with, are a source of

³In many situations, this may precisely be 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.

identification of the causal effects of marketing activities. Second, we present detailed illustrations of the identifiability of causal marketing effects using the design, and show theoretically the conditions under which the RD estimator may be valid in marketing contexts, considering in particular, the role of strategic consumer self-selection. Finally, we demonstrate the utility of the RD approach through two empirical applications, with counter-intuitive conclusions that would not be possible to obtain through a naive or even more parametric analysis.

The rest of the 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

We start with a short review of the key identification conditions from HTV, which we use to motivate our discussion below of the identifiability of marketing mix effects using RD-based designs. 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, we focus 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 threshold, \bar{z} , i.e., $d_i = \mathcal{I}(z_i \geq \bar{z})$. Then, the observed size of the discontinuity in the outcome in a neighborhood of \bar{z} are,

$$Y^- = \lim_{z \rightarrow \bar{z}^-} \mathbb{E}[Y_i(0) | z_i = \bar{z}] \text{ and } Y^+ = \lim_{z \rightarrow \bar{z}^+} \mathbb{E}[Y_i(1) | z_i = \bar{z}]$$

and of the treatment are,

$$d^- = \lim_{z \rightarrow \bar{z}^-} \mathbb{E}[d_i | z_i = \bar{z}] \text{ and } d^+ = \lim_{z \rightarrow \bar{z}^+} \mathbb{E}[d_i | z_i = \bar{z}] \quad (1)$$

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

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.⁴ Estimation of the size of the discontinuities

⁴In 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.

can be achieved by nonparametrically estimating the limit values (d^-, d^+, Y^-, Y^+) (see for e.g., Porter 2003 for approaches to estimation of these limits). The interpretation of 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. To set the stage for what follows, note that 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 $z_i = \bar{z}$. Essentially, one implies the other. To see this, let θ index consumer types. Note that, $\mathbb{E}_\theta[Y(0|\theta) | z = \bar{z}] = \int Y(0|\theta) h(\theta|z = \bar{z}) d(\theta)$, where $h(\cdot)$ is the density of θ given z (equivalently for $Y(1|\theta)$). If $h(\theta|z)$ is not continuous at $z = \bar{z}$, we obtain that,

$$\lim_{z \rightarrow \bar{z}^-} \int Y(0|\theta) h(\theta|z) d(\theta) \neq \lim_{z \rightarrow \bar{z}^+} \int Y(0|\theta) h(\theta|z) d(\theta) \quad (2)$$

Thus, continuity of counterfactual outcomes at the threshold is violated, and the RD is invalid.⁵

Corollary 2 *If the distribution of the score z given the type θ is not continuous at $z = \bar{z}$, the RD design is invalid*

Proof. By Bayes' rule, the conditional density of types given the score is,

$$h(\theta|z) = \frac{f(z|\theta) f(\theta)}{f(z)}$$

This implies that if the distribution of the score z given the types, $f(z|\theta)$, is not continuous at $z = \bar{z}$, then, $h(\theta|z)$ will not be continuous as well. Using the argument above, discontinuity in $h(\theta|z)$ invalidates the RD. ■

An assessment of the validity of an RD application thus depends on whether continuity conditions hold in that context. We now consider what kind of underlying behavior by agents is required for these identifying conditions to hold in targeted marketing settings. In the subsections that follow, we consider models of strategic customer behavior related to our two empirical applications.

2.1 Illustrative Selection Model for RD

As described in the previous section, the validity of the RD design in a specific context is dependent on whether the outcome is continuous in the score at the point of discontinuity of the score. A potential reason for discontinuity of the outcome in the score, and one that is particularly of relevance to many Marketing and Industrial Organization contexts is strategic self-selection. The intuition is

⁵To recall the definition of continuity, $h(\theta|z)$ is continuous at $z = \bar{z}$, if $\lim_{z \rightarrow \bar{z}^-} h(\theta|z) = \lim_{z \rightarrow \bar{z}^+} h(\theta|z) = h(\theta|\bar{z})$. Hence, discontinuity of $h(\theta|z)$ implies (2).

simple. The treatment is assigned based on whether the score variable for an agent is higher than a threshold. If agents can strategically self-select to be on the right of this threshold, then agents immediately to the left of the threshold cannot be considered a valid control group for those who lie immediately to the right. This invalidates the regression discontinuity design. In this section, we explore more formally the consequence of such strategic self-selection for identification of an RD estimator. We also explore if heterogeneity can mitigate the effects of selection.

We begin by specifying a simple illustrative model of selection. In the context of this model, we show that geographic targeting can be typically analyzed using RD because the costs of moving will generally be substantially greater than the benefits of receiving a preferential marketing treatment. The model involves two stages. Initially, consumers are endowed with a score z . We can think of z as the customer's location on a Hotelling line. In the first stage, the customer makes a selection decision to move his location to \tilde{z} . We refer to \tilde{z} as the manipulated score. 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 the manipulated score \tilde{z} and his treatment eligibility. We consider the outcome stage first, then present the selection stage.

2.1.1 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,

$$\begin{aligned} u_1 &= v(X, R = 1|\beta) \mathcal{I}(\tilde{z} \geq \bar{z}) + v(X, R = 0|\beta) \mathcal{I}(\tilde{z} < \bar{z}) + \eta_1 \\ u_0 &= \eta_0 \end{aligned} \tag{3}$$

Here, $v(\cdot)$ -s indicate the non-stochastic portion of the individuals utility of choosing to purchase, and $\eta = (\eta_1, \eta_0)$ are mean-zero unobservables (to the econometrician) that affect purchases. We assume that receiving treatment is preferred, such that $v(X, R = 1|\beta) > v(X, R = 0|\beta)$.

2.1.2 Stage 1: Selection

In the first stage, each customer can choose to manipulate his current score z to the score relevant for treatment, \tilde{z} ,

$$\tilde{z} = z + m \tag{4}$$

That is, the customer can choose to move his score by a value m in order to obtain preferential treatment. Changing the score is however not costless. We model the total cost of moving as having a fixed and marginal component, $C = F + \tau m$. A consumer at z would move to \bar{z} if the expected

value from obtaining the treatment is greater than the cost,

$$\mathbb{E}_\eta [u_1 - u_0] = v(X, R = 1|\beta) - v(X, R = 0|\beta) \geq F + \tau(\bar{z} - z) \quad (5)$$

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

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

2.1.3 Identification

We now consider whether an RD applied to this context is valid. As shown in Corollary 2, the RD design is invalid if the distribution of the score is not continuous at the threshold. Hence, we discuss identification in terms of whether continuity of the manipulated score \tilde{z} is violated at the threshold, \bar{z} . Continuity of \tilde{z} depends on whether the marginal consumer has a score z^* less than \bar{z} . If, $z^* < \bar{z}$, all consumers between $[z^*, \bar{z})$, including those just to the left of \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} , and would be discontinuous, implying the RD is invalid. Essentially, with selection, the limit of the counterfactual outcome just to the left of \bar{z} does not exist.

Formally, the condition for RD to be invalid, $z^* < \bar{z}$, implies,

$$\frac{1}{\tau} (F + \tau\bar{z} - [v(X, R = 1|\beta) - v(X, R = 0|\beta)]) < \bar{z}$$

i.e., that,

$$F < [v(X, R = 1|\beta) - v(X, R = 0|\beta)] \quad (7)$$

Intuitively, if the fixed costs of moving are not higher than the gain from moving, selection can invalidate an RD application.

Heterogeneity We now consider if 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, we show that unless *all* customers have sufficiently large fixed costs, the continuity in the counterfactual outcome fails in the presence of this type of heterogeneity.

We first demonstrate this graphically to show the intuition behind why this is the case and then also demonstrate this more formally. Consider a situation where there are two types of consumers, one type with sufficiently high fixed costs, such that they would not satisfy the condition in (7) above. The second type of consumers has fixed cost such that a fraction of them with $z > z^*$ move

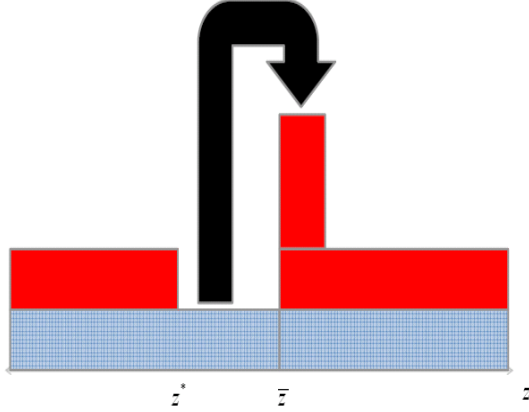


Figure 1: Selection Above threshold by One Consumer Type

to the right of the threshold \bar{z} , with z^* as defined early. Figure 1 graphically depicts this situation. Consumers of the first type are lightly shaded. None of these consumers select into the treatment group. Type 2 consumers are represented in dark shaded boxes and as can be seen in the picture, a proportion of these consumers select into treatment. This is a situation where the limit of the outcome to the left and right of the threshold (\bar{z}) exist. However because some of the consumers of the second type have incentives to select to be on the right of the threshold, there is a discontinuity of types and consequently a discontinuity in outcomes at the threshold. Thus, even if there are some consumers of this second type, the conditions for identification of RD are violated.

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,

$$\lim_{\tilde{z} \rightarrow \bar{z}^-} \mathbb{E}[Y(0|\theta) | \tilde{z}] = \lim_{\tilde{z} \rightarrow \bar{z}^-} \int \int Y(X, R = 0, \tilde{z}, \eta, \theta) d\mathcal{F}_{\theta|\tilde{z}}(\theta | \tilde{z} < \bar{z}) d\mathcal{F}_{\eta}(\eta) \quad (8)$$

while the limit from the right of \bar{z} is,

$$\lim_{\tilde{z} \rightarrow \bar{z}^+} \mathbb{E}[Y(0|\theta) | \tilde{z}] = \lim_{\tilde{z} \rightarrow \bar{z}^+} \int \int Y(X, R = 0, \tilde{z}, \eta, \theta) d\mathcal{F}_{\theta|\tilde{z}}(\theta | \tilde{z} \geq \bar{z}) d\mathcal{F}_{\eta}(\eta) \quad (9)$$

In the presence of selection, the set of consumers to the right of the threshold would have lower τ and F , and higher β than those to the left. Hence, $\mathcal{F}_{\theta|\tilde{z}}(\theta | \tilde{z} < \bar{z}) \neq \mathcal{F}_{\theta|\tilde{z}}(\theta | \tilde{z} \geq \bar{z})$ and the expected value of the counterfactual outcome to the left (equation (8)) is different from the expected value of the counterfactual outcome to the right (equation (9)). Hence, heterogeneity in types does not guarantee validity of the RD design. The only way in which this is guaranteed is 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 .

2.1.4 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. The above analysis however demonstrates that applying RD to geographic targeting relies on an underlying model of strategic 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. We now discuss such a model where the score variable is a customer’s purchase history, which is much less costly for a consumer to change.

2.2 History-Based Targeting

We now consider a canonical example of a frequent-flier loyalty program. This analysis generalizes the previous model to situations where targeting may also be based on the past behavioral history of the consumers. Additionally, we introduce a new element into the model representing randomness that drives the selection decision of the consumer. We ask whether such randomness can smooth out the discontinuity induced by selection into treatment. We show that the answer depends on the *nature of the randomness, and the precise details of how it affects behavior*. When the score and threshold are known with certainty, we show that randomness in the selection decision is not enough to smooth out selection-induced discontinuities. This analysis suggests that RD is not likely to be viable in traditional frequency reward programs, where thresholds or scores are communicated to customers as purchase incentives. Further, we show that uncertainty about the magnitude of the reward is not enough to smooth the discontinuity. On the other hand, we show that randomness that arises due to consumer uncertainty about the score or the threshold can make the RD design valid even in the presence of selection.

2.2.1 Selection in a frequent-flier program

Consider a simplified frequent-flier program that provides a mileage reward (e.g. premium seating) R to consumers if their accumulated miles flown, z , crosses a threshold, \bar{z} . Assume that consumers have no rewards at the beginning of the period. Here, R is the treatment, z is the score and \bar{z} is the threshold at which a discontinuity arises. The econometrician wishes to measure a treatment effect of the frequent-flier program on flight demand by comparing the flying behavior of consumers just to the right of \bar{z} to those just to the left of \bar{z} . Consumers are forward-looking, and are assumed to know the reward R as well as the threshold \bar{z} , and have an incentive to fly to earn the reward when their current miles $z < \bar{z}$. The timing of the actions each period is set up as follows. At the beginning of

the period, consumers are endowed with a mileage z . Based on z, R and their characteristics, they evaluate whether to self-select and fly in order to earn their reward. Flying adds m miles to z . In contrast to the previous model, we allow a random shock ε to their flying decision to be realized (this can represent factors like whether the customer’s preferred airline has multiple connections on a route). This shock is unobserved to the econometrician. Consumers condition on the realization of ε , and then make a decision to fly and change their score. Denote the selection decision by an indicator $y = y(m, z, R, \varepsilon)$. The flying decision determines their new mileage, $\tilde{z} = \tilde{z}(m, z, y(m, z, R, \varepsilon))$. Those with $\tilde{z} \geq \bar{z}$ obtain the reward R . Finally, at the end of the period, conditioning on their treatment status, consumers make a decision of whether to fly, denoted as $Y = Y(\tilde{z}, R, \eta)$, where η is an unobservable (to the econometrician) that affects the outcome of interest.⁶ The econometrician observes $\{Y, \tilde{z}\}$ across a sample of consumers in the neighborhood of \bar{z} . As before, z is the “true” score, \tilde{z} is the “manipulated” score, y is the action that generates selection, and Y is the outcome of interest. Note that the action of “flying” is both a selection action as well as an outcome of interest, albeit at two different points of time. Thus, this setup builds on that in section (2.1) but has two important differences. First, we now consider a situation where actions at the selection stage are discrete, and also allow for randomness (ε) to influence the selection decision.

Second, we recognize that the manipulated score, \tilde{z} , can be written as,

$$\tilde{z} = z + m \times y(m, z, R, \varepsilon) \quad (10)$$

i.e., selecting to fly can earn m more miles via rewards. We start with examining how y is determined. Following a canonical discrete-choice set-up, 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 ,

$$y = \mathcal{I}(f(m, z, R) + \varepsilon > 0) \quad (11)$$

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 flying and earning a reward and that for not flying. As is typical, ε is assumed to be additive, and observed by the consumer prior to choice at the selection stage. Consider consumer 1 with miles z_1 just to the left of \bar{z} such that $z_1 \in [\bar{z} - m, \bar{z})$. Given selection, the induced distribution of his manipulated score \tilde{z}_1 is,

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

To examine the implications of selection for the induced distribution of \tilde{z} , note that in general, we expect that $f(m, z, R | z < \bar{z}) > f(m, z, R | z \geq \bar{z})$, as we expect that consumers who do not have

⁶To be clear, ε is an unobservable that affects the selection decision, while η is an unobservable that affects the outcome decision.

the reward but are close to it derive a net value from self-selecting and flying to earn the reward which 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 threshold will jump to have higher mass to the right of the threshold. As the proportion of consumers who purchase at the selection stage is higher just below the threshold than just above, the density of the manipulated score has a discontinuity at the threshold. A discontinuity in the density of the manipulated score at the threshold implies a discontinuity in the types just to the left and the right of \bar{z} , and RD is invalidated by selection.

This analysis demonstrates how selection in a typical loyalty program invalidates the RD design. Further, randomness in the selection decision (ε) does not generate the smoothing required for an RD analysis to be valid. In the remainder of this sub-section, we demonstrate this intuition graphically in figure 2. For simplicity, we consider a situation where z has a uniform distribution, but the analysis generalizes to any continuous distribution. Also, the entire analysis is conducted for a consumer of a given type. We show that selection causes discontinuity at the threshold for any given consumer type, and therefore, by extension, there is a discontinuity at the threshold across the distribution of types as well.

As seen in Corollary 2, the validity of the RD design can be assessed by checking whether continuity of \tilde{z} is violated at \bar{z} . We first recognize that consumers with a manipulated score \tilde{z} near \bar{z} must have an initial score z that is in the neighborhood of either \bar{z} or $\bar{z} - m$. This is because consumers in the neighborhood of $\bar{z} - m$ who choose to purchase in the selection stage add m miles to their score and end up in the neighborhood of \bar{z} . At the same time, consumers who are initially in the neighborhood of \bar{z} remain there if they choose not to purchase in the selection stage.

We now consider the selection decision in three regions, (1) $z < \bar{z} - m$, (2) $\bar{z} - m < z < \bar{z}$, and (3) $z > \bar{z}$. Amongst these, selection is not a relevant force in regions (1) and (3), as adding m miles when in region (1) does not achieve the reward, while the reward has already been attained when at region (3). However, when in region (2), a decision to fly obtains the reward, which generates a strong incentive to select. This incentive creates a jump in the probability of purchasing for selection at both $\bar{z} - m$ and \bar{z} . Discontinuities in the purchase decision then induce a discontinuity in the manipulated score. To see this, refer to Figure 2. The top panel depicts the purchase incentives at the selection stage over the space of the error ε and score z . The decision to purchase and select is determined by the realization of ε . Consumers in regions (1) and (3) will select to fly if ε large

⁷This can be seen from the fact that the value of selecting to fly at the selection stage 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 fly at the selection stage. For consumers with $z < \bar{z}$, the probability of getting treatment increases when the consumer selects to fly. However, for consumers with $z > \bar{z}$, this is not true.

enough such that,

$$\varepsilon > -f(m, z, R|z < \bar{z} - m)$$

$$\varepsilon > -f(m, z, R|z > \bar{z})$$

To isolate the selection incentives in region (2), note that we set $f(m, z, R|z < \bar{z} - m) = f(m, z, R|z > \bar{z})$ because below $\bar{z} - m$ and above \bar{z} , there is no additional reward earned due to purchase at the selection stage. Hence, in the top panel of Figure 2, the cutoff heights on ε for selection to occur are the same for region (1) and (3).

Now, consider consumers in the box labeled 2. These consumers have an added incentive to purchase in the selection stage to earn the reward. Hence, their ε need only be as large as,

$$\varepsilon > -f(m, z, R|\bar{z} - m < z < \bar{z}) < -f(m, z, R|\bar{z} - m) = -f(m, z, R|z > \bar{z})$$

Hence, in the top panel of Figure 2, the cutoff heights on ε for selection to occur in region (2), are lower than that for region (1) and (3). As the space of ε in which selection can occur is larger in region (2), the probability of selection is also larger. The middle panel of figure 2 depicts four sets of consumers. The boxes labeled A and B depict consumers with an initial score z just below and just above respectively of $\bar{z} - m$ and the boxes C and D depict consumers just below and above the threshold \bar{z} . Assume that the widths of these boxes are infinitesimally small. By assumption, the distribution of z , denoted by $f(z)$ is continuous at both $\bar{z} - m$ and at \bar{z} and hence the heights of the boxes on either side of these values i.e. A and B, and C and D respectively are equal (under the assumption of infinitesimally small widths of these boxes).

Now consider the bottom panel, that depicts the distribution of $f(\tilde{z})$ after the selection stage. Some of the consumers in each of the four boxes A through D have an incentive to purchase in the selection stage since they have a high enough realization of the error ε that they purchase irrespective of the reward. However, the incentives to purchase are higher for consumers in boxes B and C than in A and D because for these consumers there is the added incentive of obtaining a reward. Thus, the proportion of consumers in B and C who purchase is higher than those for A and D respectively. The bottom panel of the figure depicts the manipulated score. As can be seen, the combined effect of a greater proportion of consumers in B who purchase than in A and a smaller proportion in C than in D who stay at their original score causes a discontinuity at the threshold \bar{z} .

2.2.2 Mitigating Selection: Uncertainty

Uncertainty about the manipulation of 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.

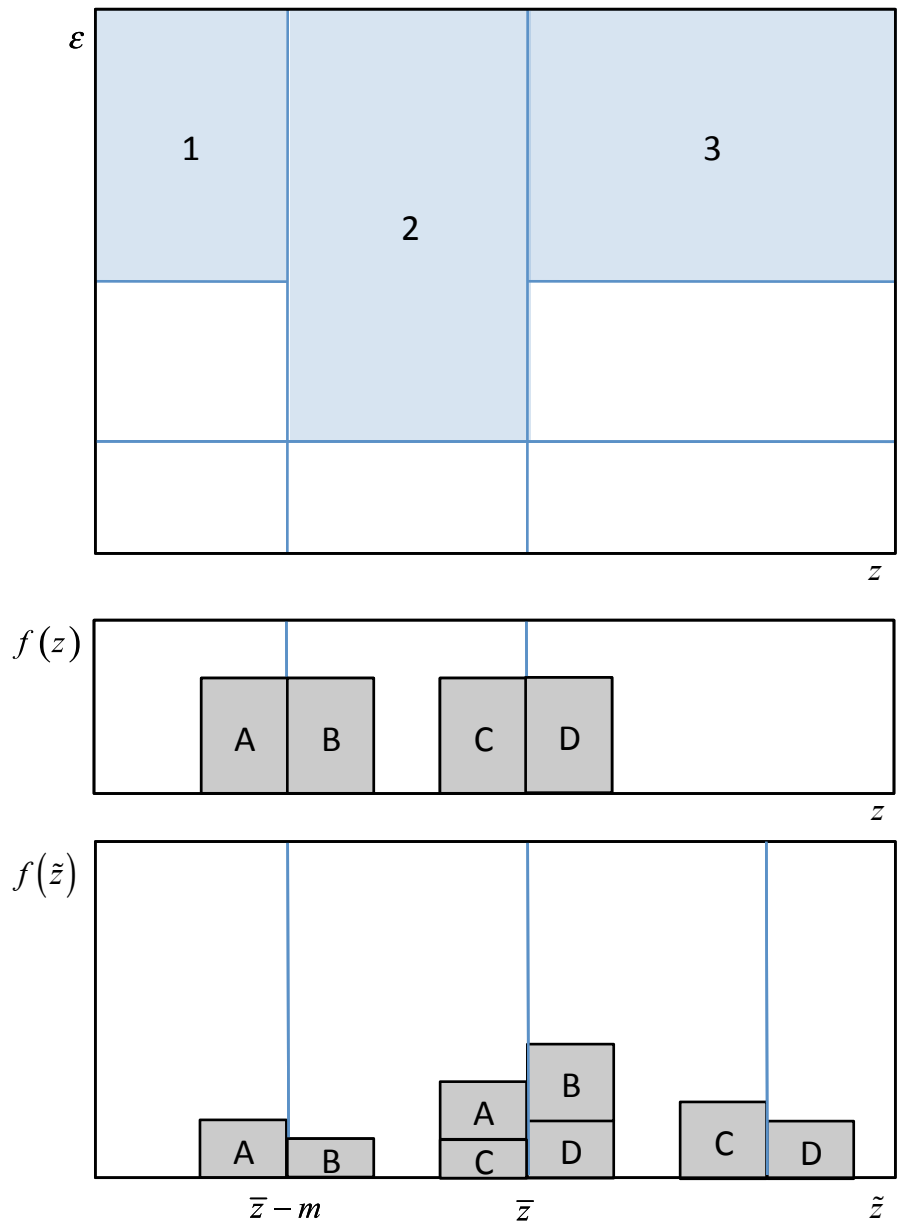


Figure 2: Discontinuity due to selection

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. An example of such a situation is that of a hypothetical loyalty program where consumers who choose to purchase are entered into a lottery to earn points. Thus, m is zero with some probability and positive 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 threshold still have no incentives to fly 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 threshold will have an incentive to select that consumers to the right do not face. This will cause a discontinuity at the threshold, invalidating the RD design.

To see this formally, 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,

$$\tilde{z} = \begin{cases} z + m \times y(m, z, R, \varepsilon) & \text{w.p. } p \\ z & \text{w.p. } (1 - p) \end{cases} \quad (12)$$

Consider a consumer to the right of the threshold. 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 threshold 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 12,

$$\tilde{z}_1 = \begin{cases} z_1 + m & \text{w.p. } p \times \Pr(y = 1 | m = m, z_1, R, \varepsilon_1) \\ z_1 & \text{w.p. } 1 - [p \times \Pr(y = 1 | m = m, z, R, \varepsilon)] \end{cases} \quad (13)$$

As long as the conditions described in section 2.2.1 hold, $\Pr(y = 1 | m = m, z_1, R, \varepsilon_1)$ is non-zero. 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 threshold \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

$$\tilde{z}_1 = \begin{cases} z_1 + m & \text{w.p. } \Pr(y = 1|m, z_1, R, \varepsilon_1)g(m) \\ z_1 & \text{w.p. } \Pr(y = 0|m, z, R, \varepsilon)g(m) \end{cases} \quad (14)$$

Once again, as long as $g(\cdot)$ is any continuous density and the conditions described in 2.2.1 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.

Uncertainty about the Score We add a different kind of uncertainty into the model. We generate some the uncertainty from the consumer's perspective into the score \tilde{z} 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. The main difference of this scenario from the one with continuous m is that here, the uncertainty operates directly on the score, while there, the uncertain quantity, m , affected the score only via the realization of the discrete selection action y . We show that this change is sufficient to mitigate the effect of selection. The equation below makes this clear. Consider a modification to equation (10) as follows,

$$\tilde{z} = z + m \times y(m, z, R, \varepsilon) + w \quad (15)$$

The key addition w , is an error term that adds randomness to the score. w is assumed to be not under the control of the consumer when he makes the selection decision. Note that making w additively separable effectively requires that it is a source of pre-selection randomness. This may occur, for instance, if consumers may forget their mileage and cannot look it up, or if consumers' do not know the exact score used by the firm.⁸ 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 threshold used for estimation. While 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 threshold \bar{z} . Following the intuition proposed in Lee (2008), we illustrate below that 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 threshold 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 (15). We can think of the distribution of \tilde{z}_1 induced by Equation (15) as the following

⁸Or 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.

mixing distribution,

$$\tilde{z}_1 = \begin{cases} z_1 + m + w & \text{w.p. } \Pr(y = 1|z_1 < \bar{z}, x_1) \\ z_1 + w & \text{w.p. } \Pr(y = 0|z_1 < \bar{z}, x_1) \end{cases} \quad (16)$$

Thus, the distribution of \tilde{z}_1 is obtained by taking a weighted average of the 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 $\Pr(y = 1|z_1 < \bar{z},) > 0$. However, due to 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 (16), except the weighting probabilities are evaluated at the right of the threshold, e.g., $\Pr(y = 1|z_2 \geq \bar{z})$. 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 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 threshold. Consequently, the RD design is now valid. To the extent that such randomness is plausible in many contexts, RD designs maybe considered very similar to quasi-randomized experiments (see for e.g., Lee and Lemieux 2009).⁹ 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 Threshold A common feature of targeting contexts is that consumers may not know the exact threshold used for treatment, even if they know other aspects of the firm’s targeting policy. In this case consumers may likely have a continuous belief distribution over the thresholds. We now consider whether uncertainty about the exact threshold at which reward miles may be earned may be enough to ensure the validity of the RD design. Analyzing this aspect is more complicated, as it requires us to be more explicit about the choice-specific value functions that generate the function $f(m, z, R)$ in Equation (11).

We start with the basic model without other sources of randomness (i.e. no w) with $\tilde{z} = z + my$. To define a customer’s selection/purchase decision, we begin by specifying $f(m, z, R) =$

⁹Note the implicit requirement that the randomness w has full support in the region $(\bar{z} - h, \bar{z} + h)$ of the threshold. For instance, suppose the airline never changes its mileage rules, and hence, those to the right of the threshold (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.

$\mathcal{V}_1(m, z, R|\beta) - \mathcal{V}_0(m, z, R|\beta)$, where, the choice-specific value functions, $\{\mathcal{V}_1, \mathcal{V}_0\}$, are defined as,

$$\begin{aligned}\mathcal{V}_1(m, z, R|\beta) &= v_1(R|\beta) + \delta \mathbb{E}\mathcal{V}(\tilde{z}, R|m, z, y = 1) + \varepsilon_1 \\ \mathcal{V}_0(m, z, R|\beta) &= 0 + \delta \mathbb{E}\mathcal{V}(\tilde{z}, R|m, z, y = 0) + \varepsilon_0\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\mathcal{V}(\tilde{z}, R|m, z, y) &= \int \int [\mathcal{I}(\tilde{z} \geq \bar{z}) [\max\{v_1(R|\beta) + \eta_1, \eta_0\}] \\ &\quad + \mathcal{I}(\tilde{z} < \bar{z}) [\max\{v_1(0|\beta) + \eta_1, \eta_0\}]] d\mathcal{F}_\eta(\eta_1, \eta_0) d\mathcal{G}_{\bar{z}}(\bar{z})\end{aligned}\tag{17}$$

That is, if the consumer is able to cross the required threshold by choosing y today, $\mathcal{I}(\tilde{z} \geq \bar{z}) = 1$, and he will have the reward R and miles $\tilde{z} = z + my$ tomorrow. If he chooses to fly tomorrow, he obtains payoff $v_1(R|\beta) + \eta_1$; 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 threshold by selecting to fly today, $\mathcal{I}(\tilde{z} < \bar{z}) = 1$, the value from the best 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 (17) the consumers' beliefs over the threshold \bar{z} is represented by a continuous density, $\mathcal{G}_{\bar{z}}(\bar{z})$, and his beliefs over the random shocks η by the density $\mathcal{F}_\eta(\eta_1, \eta_0)$. Moving the integration into the brackets, and noting that $\Pr(\tilde{z} \geq \bar{z}|m, z, y) \equiv \mathcal{G}_{\bar{z}}(\tilde{z})$, we can write Equation (17) as,

$$\mathbb{E}\mathcal{V}(\tilde{z}, R|m, z, y) = G_{\bar{z}}(\tilde{z}) \mathbb{E}_\eta [\max\{v_1(R|\beta) + \eta_1, \eta_0\}]\tag{18}$$

$$+ [1 - G_{\bar{z}}(\tilde{z})] \mathbb{E}_\eta [\max\{v_1(0|\beta) + \eta_1, \eta_0\}]\tag{19}$$

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, $\Omega(z + m, z|R, \beta) = \mathbb{E}\mathcal{V}(\tilde{z}, R|m, z, y = 1) - \mathbb{E}\mathcal{V}(\tilde{z}, R|m, z, y = 0)$, the relative expected future value from selection versus not. We can evaluate Equation (18) at $y = (1, 0)$ ¹⁰ to obtain,

$$\begin{aligned}\Omega(z + m, z|R, \beta) &= [G_{\bar{z}}(z + m) - G_{\bar{z}}(z)] \\ &\quad \times [\mathbb{E}_\eta \max\{v_1(R|\beta) + \eta_1, \eta_0\} - \mathbb{E}_\eta [\max\{v_1(0|\beta) + \eta_1, \eta_0\}]]\end{aligned}\tag{20}$$

¹⁰Remember that at $y = 0$, $\tilde{z} = z$ and at $y = 1$, $\tilde{z} = z + m$

Intuitively, Equation (20) implies that the future component of the incentive to select/purchase is the difference in expected utility under treatment and not, weighted by the increase in the probability of receiving treatment that is due to the m incremental points, i.e. $G_{\bar{z}}(z+m) - G_{\bar{z}}(z)$. The decision to select is given now as,

$$\Pr(y=1) = \Pr(v_1(R|\beta) + \delta\Omega(z+m, z|R, \beta) + \varepsilon_1 - \varepsilon_0 > 0) \quad (21)$$

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

$$\tilde{z} = \begin{cases} z+m & \text{w.p. } \Pr(y=1|m, z, R) \\ z & \text{w.p. } \Pr(y=0|m, z, R) \end{cases} \quad (22)$$

Proposition 3 *The distribution of \tilde{z} is continuous at the true threshold. Hence, the RD is valid.*

Proof. First, fix the value of the threshold actually used by the firm at \bar{z} . Now note that the density of \tilde{z} will be continuous at the true threshold \bar{z} if the mass of \tilde{z} that piles up to the left and right of \bar{z} due to selection is the same. This will be the case if, (1) $\Pr(y=1|m, z, R)$ is the same just to the left and to the right of $z = \bar{z} - m$, and, (2) $\Pr(y=0|m, z, R)$ is the same just to the left and to the right of $z = \bar{z}$. Note from Equation 20 above that z affects the probabilities only through the difference in cumulative densities, $[G_{\bar{z}}(z+m) - G_{\bar{z}}(z)]$. 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 $z = \bar{z} - m$, and $z = \bar{z}$. To see that this is the case, note that $G_{\bar{z}}(z)$ is the *marginal* c.d.f. of the random variable \bar{z} evaluated at the value z , prior to selection. As consumers do not know the true \bar{z} , this function is continuous at all z (by the primitive assumption), including at $z = \bar{z} - m$, and $z = \bar{z}$. The difference between two continuous functions is also continuous. Hence, $[G_{\bar{z}}(z+m) - G_{\bar{z}}(z)]$ is also continuous. Q.E.D. ■

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

$$\lim_{z \rightarrow \bar{z}^-} G_{\bar{z}}(z+m) = 1 \text{ and } \lim_{z \rightarrow \bar{z}^-} G_{\bar{z}}(z) = 0 \quad (23)$$

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

$$\lim_{z \rightarrow \bar{z}^+} G_{\bar{z}}(z+m) = 1 \text{ and } \lim_{z \rightarrow \bar{z}^+} G_{\bar{z}}(z) = 1 \quad (24)$$

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

To clarify the intuition behind how uncertainty about the threshold validates the RD design, we now depict the scenario graphically, expanding the discussion around Figure 2 earlier, which was in a situation where the threshold was known. Assume, for the sake of simplicity once again, that z is uniformly distributed and that the uncertainty about the threshold also has a uniform distribution. In other words, the true threshold $\bar{z} = \bar{z} + w$, where $w \sim Uniform[-\bar{w}, \bar{w}]$. Consumers do not know what w is but know its distribution.

Consider the top panel in figure 3. This depicts the purchase incentives for consumers in the space of $\varepsilon = \varepsilon_1 - \varepsilon_0$ and z . Consumers in the box labeled 1 have an initial score lying between $\bar{z} - m - \bar{w}$ and $\bar{z} - m + \bar{w}$. Since the true threshold \bar{z} is uniformly distributed between $\bar{z} - \bar{w}$ and $\bar{z} + \bar{w}$, these consumers have an incentive to purchase in the selection stage in order to receive the reward, described by the inequality,

$$\varepsilon_1 - \varepsilon_0 > -[\mathcal{V}_1(m, z, R|\beta, \bar{z} - m - \bar{w} < z < \bar{z} - m + \bar{w}) - \mathcal{V}_0(m, z, R|\beta, \bar{z} - m - \bar{w} < z < \bar{z} - m + \bar{w})]$$

Similarly, consumers in the box labeled 2 have an incentive to purchase in the selection stage given by the inequality

$$\varepsilon_1 - \varepsilon_0 > -[\mathcal{V}_1(m, z, R|\beta, \bar{z} - \bar{w} < z < \bar{z} + \bar{w}) - \mathcal{V}_0(m, z, R|\beta, \bar{z} - \bar{w} < z < \bar{z} + \bar{w})]$$

The middle panel shows the initial scores of consumers in the neighborhood of $\bar{z} - m$ (boxes A and B) and \bar{z} (boxes C and D). Consider, as before, that these boxes are of infinitesimal width. By continuity, A and B are of equal height, and similarly, C and D are of equal height.

The bottom panel of the figure shows the distribution of the manipulated score. Note that by continuity, the proportion of consumers in boxes A and B (all of whom correspond to box 1 in the top panel) who will purchase to potentially receive the reward will be equal. This is shown by the equal heights of the boxes labeled A and B stacked at \bar{z} in the bottom panel. Similarly, the proportion of consumers in boxes C and D (all of whom correspond to box 2 in the top panel) who move are equal. Thus, the customers who remain at \bar{z} are shown by boxes of equal height stacked at \bar{z} . The heights of the stacked boxes on either side of \bar{z} are equal, demonstrating the continuity that uncertainty about \bar{z} ensures.

Discussion The above analysis documents that traditional reward programs where the thresholds 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 thresholds are viable RD applications, as long as consumers have a continuous distribution of beliefs about the uncertainty. Nevertheless, uncertainty about the value of rewards does not restore the validity of the

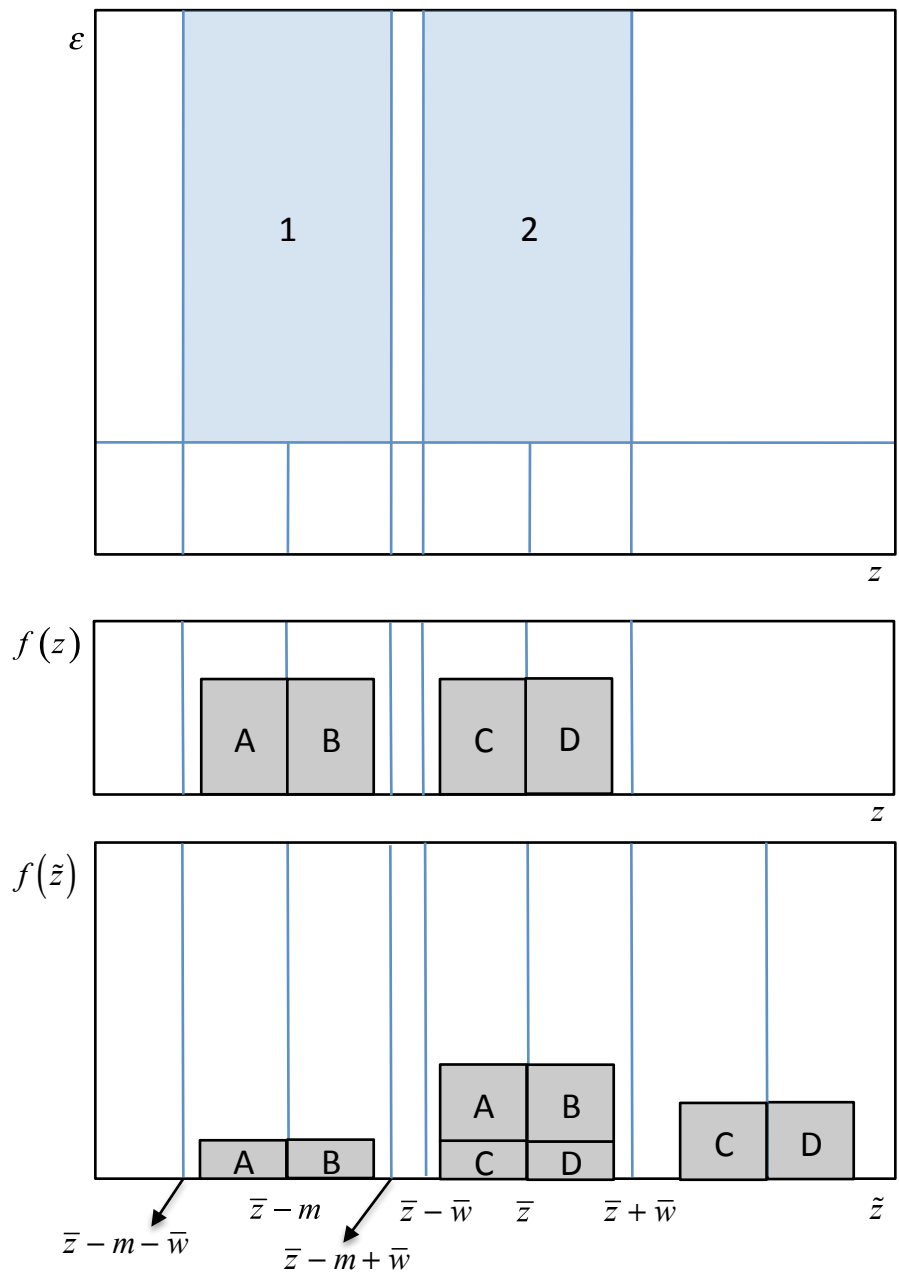


Figure 3: Selection with an unknown threshold

design. In our customer based targeting casino application below, identification of the RD estimator is provided by both uncertainty about the score as well as the threshold. 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. Similarly, the program involves offers sent to customers based on thresholds that were not provided to the public. Customers may not even know that they will get these offers, or that they differ, but if they suspect this, they would not know what threshold the firm used for determining preferential treatment.

2.2.3 Extensions: Purchase Timing as a Score Variable

To close this discussion, note the models in the previous section were developed in the context of a frequent flier program, but the intuition carries over to others contexts as well. A natural extension is to situations where the *timing of purchase* is the score variable. For instance, in Busse et al. 2006 and Busse et al. 2009, automobile prices fall discontinuously during a promotion period. We can think of analyzing this using an RD design with time of purchase as a score variable. “Before-After” contexts of this sort are typically plagued by the concern that outcomes in the “after” period arise due to time-varying unobservables that are unrelated to the treatment. The RD design is able to address this concern by focusing on outcomes within the small window of the time of promotion during which these unobservables may plausibly be the same. The extent to which selection may invalidate such RD applications depends on what the outcomes of interest are conditioned on. We first consider a simple case of myopic consumer demand to illustrate that the validity of RD depends on whether the analysis is conditioned on an initial decision such as purchase or a store visit. Next, we extend the analysis to dynamic demand contexts where state-dependence can exaggerate biases from the myopic case and show that validity of RD in this cases depends on what consumers know when.

Consider a simple discrete choice model where consumers maximize their current payoffs in each period in continuous time to decide whether to buy or not at the outcome stage. The score variable is time itself and is denoted as before by z . Let the consumer decide whether to purchase based on the maximization: $\max\{v(R_z|\beta) + \eta_{1z}, \eta_{0z}\}$, where η_{1z} and η_{0z} are unobservables (to the researcher) in the payoffs for purchase and no purchase respectively. Let R_z denote whether there is a price discount at the time. Note that this payoff is similar to the one earlier set up in equation (3) except that time is the score. The RD design arises because price discontinuously drops at $z = \bar{z}$. If we assume that a consumer can potentially buy with equal probability at all time periods, then the distribution of the score $f(z)$ is uniform. In this case there is no manipulation of the score as time itself is the score and every consumer in the analysis is a potential purchaser all the time. By the arguments outlined previously, the RD estimator is valid because the distribution of the score is

continuous at the threshold.

Now, consider how the validity changes if we analyze an outcome conditional on an initial decision such as visit to the store. The initial selection decision is indicated by $y_z = \mathcal{I}(f(R_z|\theta) + \varepsilon_z > 0)$. $f(R_z|\theta)$ is the customer's payoff for visiting the store, where the dependence of $f(\cdot)$ on R_z indicates that the customer knows the period z price when deciding whether to visit the store in period z . θ (parameters in the store visit decision) includes β (parameters in the payoff for the purchase decision) because the preference factors that lead one to visit the store are related to those that lead one to purchase. ε_z is an unobservable (to the researcher) that affects the store visit decision. For simplicity we assume that η and ε are uncorrelated and that η is not known before entering 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 $f(R_z|\theta) + \varepsilon_z > 0$ 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} . Let $f(\tilde{z})$ be the distribution of the manipulated score. For any discontinuous change in R , $f(R_z|\theta)$ will change discontinuously and consequently, $Pr[f(R_z|\theta) + \varepsilon_z > 0]$, and $f(\tilde{z})$. By Corollary 2, a discontinuity in the score invalidates RD.

We depict this graphically in figure 4. Regions 1 and 2 in the top panel represent the store visit probabilities before and after the price change at \bar{z} . Note that the visit probability is higher in region 2, reflecting the fact that the visit decision is positively affected by the reward. The second panel reflects the assumption that all customers are viable potential customers in all periods, such that the distribution of the initial score (actual time) is uniform. Boxes A and B represent this distribution before and after the price change respectively. These two are of identical heights due to the assumption of uniform distribution of the score. The third panel of the figure represents the distribution of the manipulated score induced by the selection decision. We see that after the price change, there is a greater propensity to enter the store, such that there is a greater probability that any given customer is in the store just after the price change relative to just before the price change. This discontinuity in the probability of any type being above and below the cutoff, as pointed out in section 2, leads to a discontinuity in the types above and below the cutoff and therefore invalidates RD.

The above analysis also holds when there are dynamics in demand that might arise from durability, storability or lasting effects of consumption. However, the potential selection problems when conditioning on a previous decision like store visit are exaggerated. For example, if a customer intertemporally substitutes demand and expects a future price decrease, he will be even less likely to visit the store before the actual price change. This vertically shrinks region 1 in Figure 4, thereby increasing the discontinuity. While this intuition gives a preview of the potential implications of

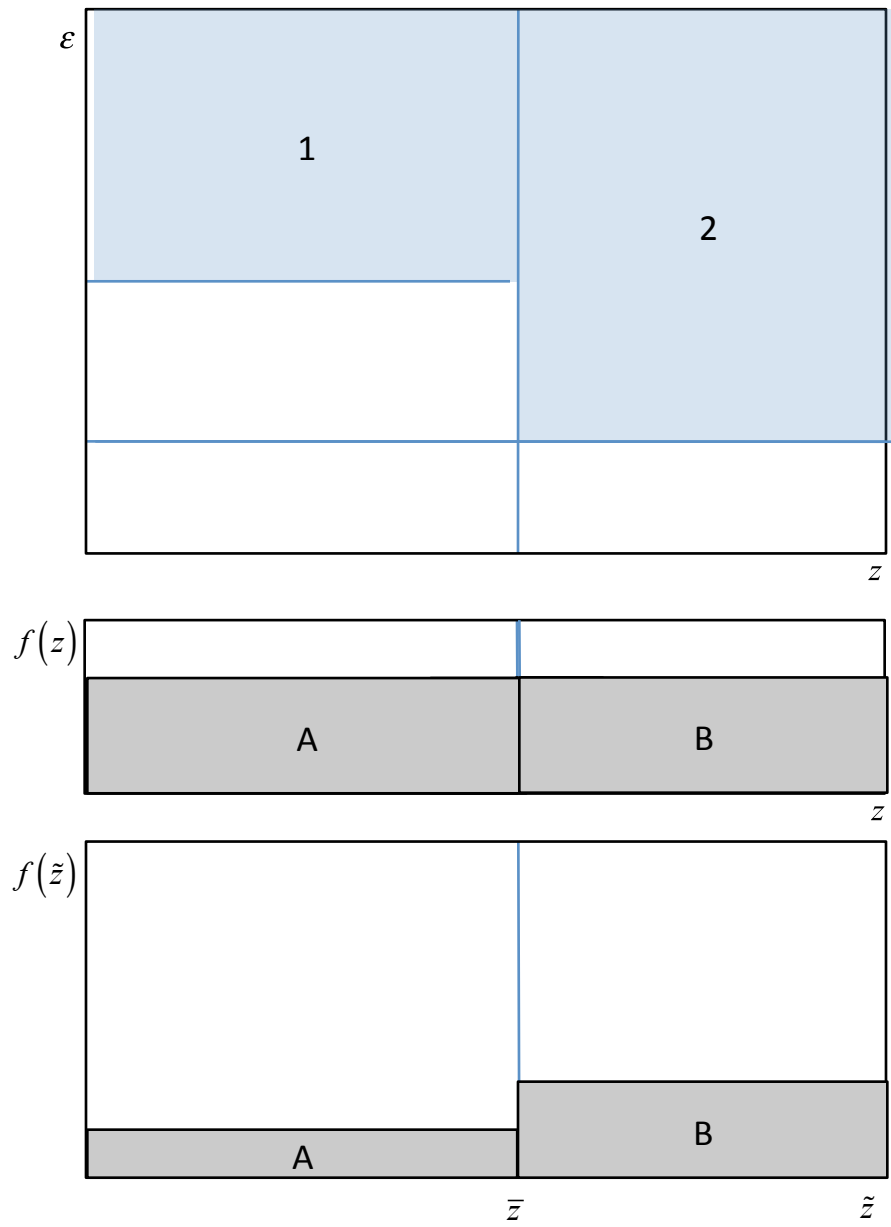


Figure 4: Selection with timing of purchase as the score variable

state-dependence, price expectations and purchase timing add some additional considerations to the analysis that we now consider more formally.

We can analyze this RD design using the logic developed in the previous section. To relate the analyses, we represent the time as a score as z as before. The econometrician is interested in analyzing outcomes around a score, \tilde{z} that is manipulated by some initial (selection) decision. In the earlier analysis in this section, we considered this to be store visit. Here we consider the initial decision to be purchase, where the outcome of interest is some other outcome, Y , that is only analyzed within the set of customers that make a purchase. Therefore the relevant score, \tilde{z} , is purchase timing. Let p_H be the high price prevailing prior to the sale. The “reward” R in this model is a low price, $R = p_L$, which will be available after a threshold time, \bar{z} . Let $y = y(p_H, R, \varepsilon)$ denote the decision to select and *to wait* for the next period; and \tilde{z} denote the manipulated score, which represents the actual time of purchase observed by the econometrician. Letting m denote the time increment, we can write \tilde{z} as,

$$\tilde{z} = z + m \times y(p_H, R, \varepsilon) \quad (25)$$

From the discussion in section 2.2.1, if the exact time at which prices will fall, \bar{z} , is known with certainty, selection invalidates the RD (the presence of the randomness ε in y is not enough to mitigate selection). Now consider the case where consumers do not know the timing of the price cut (i.e. \bar{z}) but have a belief about it that has a continuous distribution as before. If we analyze outcomes conditional on knowledge of the threshold after the promotion starts, the RD approach is invalid. Essentially, if there is no uncertainty about the threshold to the right of the threshold (i.e. after the start of the promotion), the uncertainty about the threshold to its left is insufficient to make the RD design valid. To see this, note that in this situation

$$\lim_{z \rightarrow \bar{z}^+} G_{\tilde{z}}(z + m) = 1 \text{ and } \lim_{z \rightarrow \bar{z}^+} G_{\tilde{z}}(z) = 1 \quad (26)$$

Hence as z approaches \bar{z} from the right, the difference between the c.d.f’s $G_{\tilde{z}}(z + m) - G_{\tilde{z}}(z)$. goes to 0.

Now consider the limits from the left. Since the actual timing of promotion \bar{z} is unknown to consumers before it is announced, these limits do not go to one. For any finite m , by continuity of $G(\cdot)$ and by the fact that it is a c.d.f. which for a continuous distribution is strictly increasing, we have

$$G_{\tilde{z}}(z + m) > G_{\tilde{z}}(z) \quad \forall z < \bar{z}, m > 0$$

Hence, as $z \rightarrow \bar{z}$ from the left, the difference between the c.d.f’s $G_{\tilde{z}}(z + m) - G_{\tilde{z}}(z)$ approaches a value strictly greater than 0. Since we have shown earlier that the difference in c.d.f’s approach zero

when we $z \rightarrow \bar{z}$ from the right, the difference in c.d.f.'s has a discontinuous jump $z = \bar{z}$. This, as discussed in Proposition 3 is sufficient to show that the RD design is invalid.

If, on the other hand, the outcomes are analyzed without conditioning on knowledge of the threshold after the promotion starts, we have a valid RD design. Such a situation might be present if consumers arrive at the store without knowledge of the promotion, for instance, or if the outcome is not conditioned on arrival at the store. In such a case, the consumers on both sides of the threshold have a continuous distribution of beliefs about the timing of the price cut (\bar{z}). This is enough to smooth the discontinuity induced by selection as per the discussion in section 2.2.2. Thus, the validity of the RD design depends crucially on a structural object, viz. the price expectations of consumers and on whether the outcome that is analyzed is conditioned on knowledge about the promotion in the post-promotional period or not.

Discussion 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 the design is valid. Even though the RD design fits into the “reduced-form causal effect” literature, we believe 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 consumers’ information sets and incentives to establish the validity or lack thereof, of an RD-based analysis. We illustrated these in the context of examples related to our empirical applications. We now discuss these.

3 RD Applied to Targeted Marketing

In this section, we apply the regression discontinuity approach to two applications that involve firms targeting elements of the marketing mix to specific groups of customers. The first application assesses the marketing program at a casino. The casino targets room rates, dining coupons and promotional credits to members of their loyalty program based on the expected daily amount the customer would gamble. The second application assesses the efficacy of direct-mail activity and involves a B2C company soliciting contacts from potential customers whose probability of response (represented by a one dimensional score) falls within a pre-specified range. We describe the details of each of the applications below and illustrate that the regression discontinuity approach reveals negative promotion effects that might otherwise have been interpreted as beneficial.

3.1 Targeted Promotions Emailed by a Casino

Two reasons make our application to marketing activity at casinos particularly 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. Historically, promotions at casinos, referred to in industry parlance as “comps”, were given out by pit-bosses, managers and wait staff at their discretion to reward gamblers spending time and money at tables or slots. The introduction of sophisticated database management systems and cards that perfectly track past behavior led to two changes in promotion activity at U.S. casinos. First, reward programs became more formal, with casinos defining contracts specifying comps corresponding to pre-specified cumulative levels of gaming. Second, discretionary or non-contractual comps became part of database marketing programs that leveraged customer information to retain and/or develop customers based on their observed historical behavior. Our data comprises details of one such database marketing program from an anonymous casino in Las Vegas.

The casino sends emails to its customers offering a package that may include discounted room rates, show tickets, dining credits, and promotional credits. The casino recognizes that some customers do not gamble enough for a promotion to ever generate the incremental revenue required to offset the cost of the promotion. Consequently, they target less costly and more restrictive offers to “low-rollers” and more costly and more lucrative offers to “higher-rollers”. Importantly, the classification of consumers into low or high-rollers, and the subsequent allocation of comps to these tiers is determined on the basis of a one dimensional, continuous score, referred to as the “Average Daily Win” (henceforth ADW). As the name suggests, 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.¹¹ The ADW is determined on the basis of a proprietary algorithm that incorporates customer characteristics as well as past play history. We have access to the ADW for all customers in the data, but not the algorithm that generated the ADW. In addition, we also have access to the threshold of ADW on the basis of which customers are sorted into tiers. While 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, nor the ADW-specific thresholds that generate sorting into tiers. This aspect mitigates selection concerns in this application. In Table 1 we provide an example of a casino mailing to 79,419 customers in which the offer depends on the casino’s calculation of the average

¹¹The 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 \$1000 on their first play and the other lost all \$100 on their first play.

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 1: Example of Tiered Email Promotion Sent by the Casino

daily win (ADW) they expect from the customer.¹²

We observe four such mailings between January and September 2006 (we report only one of the mailings to conserve space). All mailings are identical except in terms of the show ticket offerings, which vary across mailings. Further, the first mailing is only sent to the top two tiers. Importantly 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 through 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, viz. whether the customer visited the casino (*Trip*) and the casino’s expected win from the customer (*Theoretical Win*). The theoretical win is similar to ADW in that it adjusts for the customers’ 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 2. We see from the table that customers in the top two tiers arrive with about 23% probability, while 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 while customers in the top tier have theoretical wins that average \$617.

3.1.1 Analysis: Correlational Effects

One obvious pattern from Table 2 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 this to be an effect of the promotion itself. As a benchmark, we start by regressing the outcome variables on tier fixed effects. The tier fixed effect implicitly captures the effect of changing a promotion from the base tier to those for that tier. We pool observations across

¹²Note that the highest rollers, customers with ADW above \$2,500, are not included, as the casino deals with such customers on a one-to-one basis. Customers with ADW less than \$50 are sent emails but do not receive special offers such as discounted rooms or credits.

	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

Table 2: Summary Statistics of Casino Outcomes by Tier

ADW Cutoff	Tiers				
	0 to 1	1 to 2	2 to 3	3 to 4	4 to 5
	\$1,000	\$500	\$300	\$200	\$100
Promotion Categories					
Regular Room	-	-	\$99-Free	\$159-\$99	\$179-\$159
Room on Value Day	-	-	-	\$79-Free	\$99-\$79
Show Tickets	-	B to A	-	-	-
Dining Credits	\$50	-	\$25	-	-
Promo Credits	\$200	\$250	-	-	\$25
OLS					
Visit	0.15***	0.15***	0.09***	0.05***	0.008**
Theoretical Win	487.34***	196.67***	124.16***	53.01***	10.99**

Table 3: OLS Regressions of Casino Outcomes on Tier

all the four mailings while including period fixed effects to account for differences across the timing of mailings. Table 3 presents the results.

The OLS estimates with visit and theoretical win as the dependent variables are listed in the last two rows of Table 3. The visit variable is coded as 1 in case of a visit and 0 otherwise.¹³ 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 3, we find that it appears that higher tiers arrive more frequently. It appears that better offers are generally increasing with visit probabilities. Tiers 0, 1, and 2 appear to arrive at greater rates, perhaps indicating a response to the reduced room rates or additional credits. Overall, the visit model suggests that promotions have a significant effect on inducing consumers to visit the casino

¹³We choose OLS to illustrate the pitfalls of not accounting for the endogeneity. More sophisticated statistical specifications (e.g. count models, or 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.

and that these effects are highest for the top customers. The results on theoretical win replicate these findings. In particular, the top tiers tend to spend an incremental \$487 more in response to the promotions they receive, relative to the promotions the next lower tier receives. Assuming a maximum costs of about \$250 per promotion, the marginal effects from this correlational analysis suggest that the targeting program is working fairly well. Moving to a more complicated non-linear model will not alter this conclusion, as any such model will ultimately reflect this correlation in the data.

We now consider an RD estimator applied to the data that uncovers true causal effects of the targeting program.

3.1.2 Analysis: Causal Effects

The OLS estimates reported in Table 3 are problematic due to 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.

To set up the estimator, we start by plotting the outcomes against the score variables to see if there is an evidence of a discontinuity. Figure 5 presents a plot of the percentage of consumers visiting the casino (solid bars, left y -axis) against the ADW (x -axis). To illustrate the equivalent variation of promotions, both the promotional credits (solid black line) and the price charged for rooms (dotted black line) are also plotted against ADW (values read along the right y -axis). We focus on these two as they can be expressed in equivalent dollar terms. Finally, the ADW values at which tier status changes are colored in red. These represent the threshold at which we intend to measure treatment effects of the promotions. Looking at Figure 5, we find clear discontinuities in the allocation of promotional credits and room prices. Surprisingly however, we find little discontinuities in the outcome variable - visit probability - at the thresholds. There is an overall pattern in which visit probability is increasing in ADW, but no robust evidence for jumps in visit probability along this pattern. We see that there is likely a discontinuous shift at $ADW = 500$, suggesting that the promotions might be working for that set of customers.

Figure 6 presents an analogous plot for the theoretical win. Eyeballing Figure 6, we see that there is a discontinuous jump in theo at $ADW = 500$; however, apart from this local effect, the pattern mirrors that of the visit probability plot. It is possible that the choice of bin-width is obscuring true effects in these plots; hence we now discuss estimates from a formal RD estimator applied to these

data. To address the issue with bin-size, we present specifications in which changes to the bandwidth are explored.

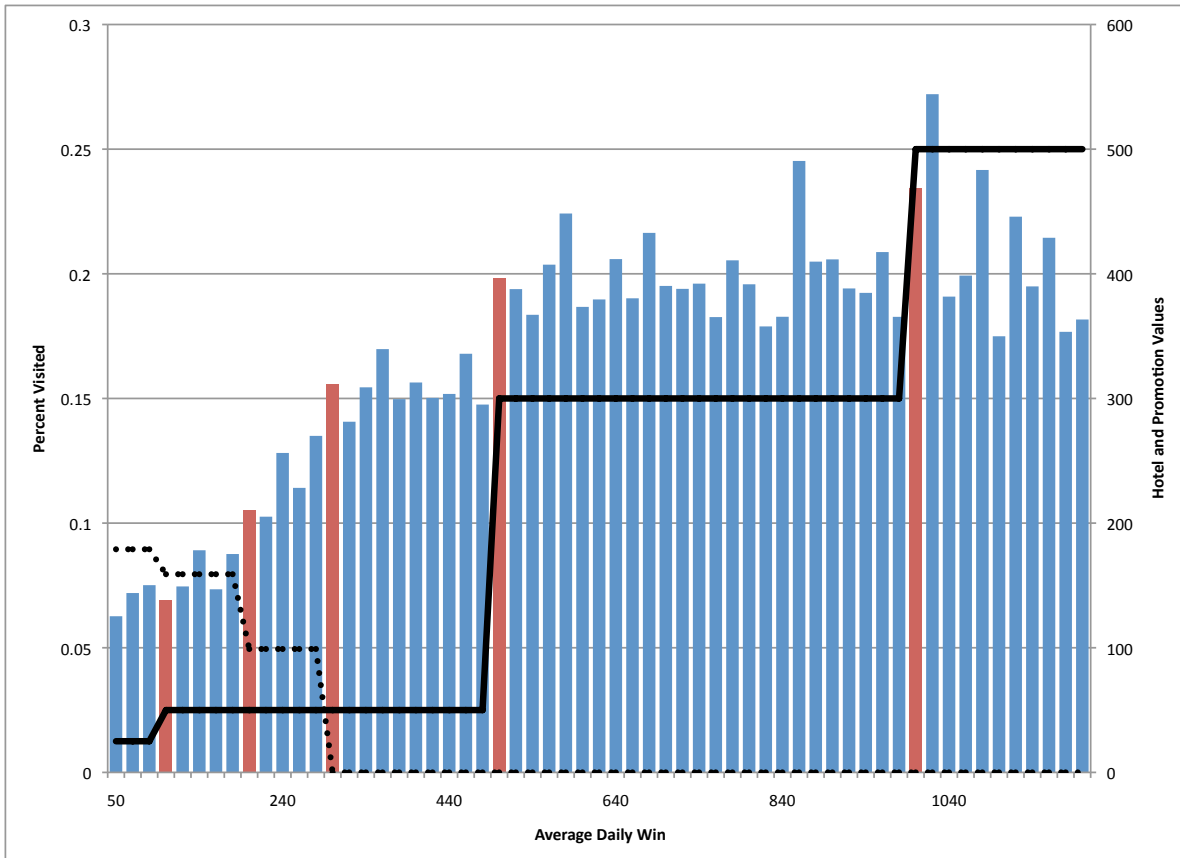


Figure 5: Percent Visiting Casino as a Function of ADW

Table 4 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.¹⁴ Following the suggestion of Imbens and Lemiux (2008), we estimate these nonparametrically using a rectangular kernel, using observations within a bandwidth of size h on either side of each threshold. Because there are six different tiers, and five different ADW thresholds, we estimate five different regression discontinuity specifications focusing on each tier threshold. 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 thresholds. An advantage of this approach is that the standard errors of the rectangular kernel are the same as

¹⁴As Lee and Lemiux (2009) notes, 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. Further, 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 ADW.

robust standard errors available for least squares. We also estimated the RD using 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 ADW that minimizes the mean square error. For completeness, we also present estimates using the same bandwidth used to generate the plots in Figures 5 and 6.

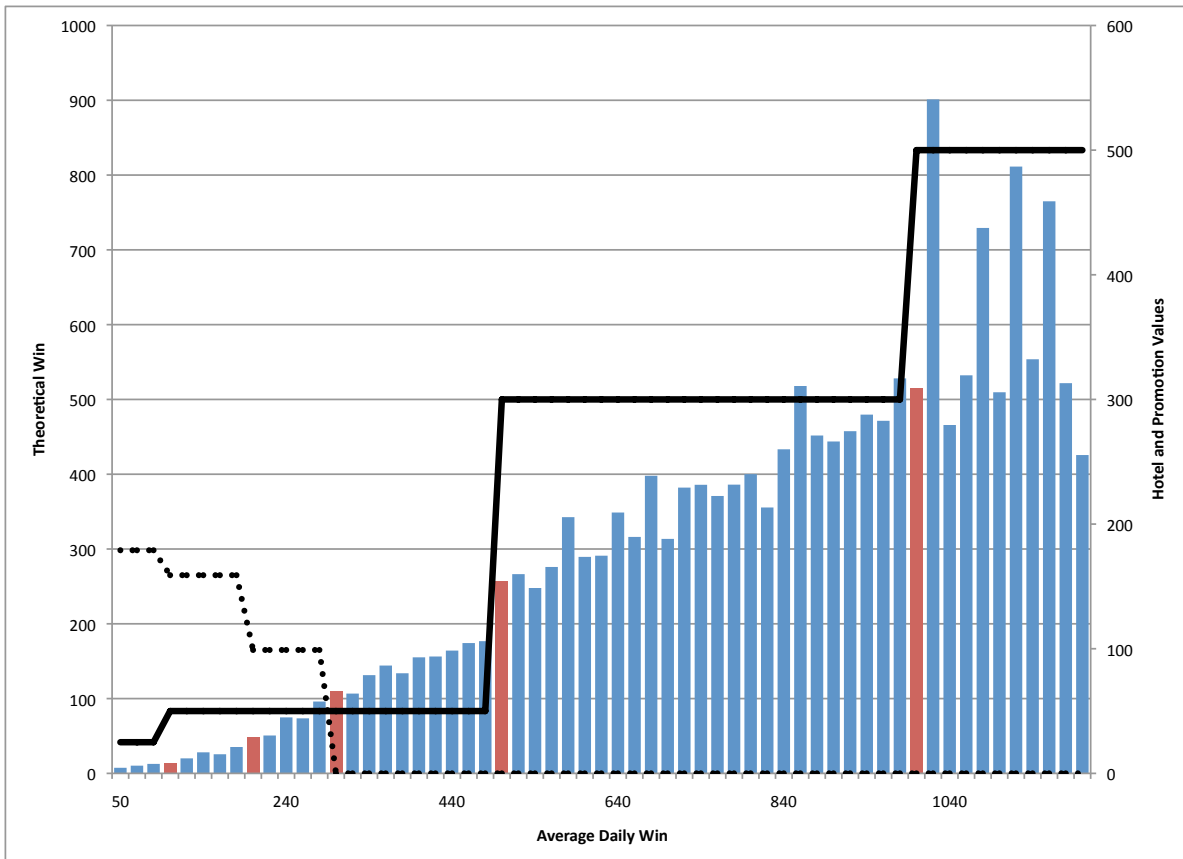


Figure 6: Theoretical Win as a Function of ADW

Referring to Table 4, 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 4 to 5 ($ADW = \100). However, both 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,

ADW Cutoff	Tiers				
	0 to 1 \$1,000	1 to 2 \$500	2 to 3 \$300	3 to 4 \$200	4 to 5 \$100
Promotion Categories					
Regular Room	-	-	\$99-Free	\$159-\$99	\$179-\$159
Room on Value Day	-	-	-	\$79-Free	\$99-\$79
Show Tickets	-	B to A	-	-	-
Dining Credits	\$50	-	\$25	-	-
Promo Credits	\$200	\$250	-	-	\$25
Effect on Outcome (Bandwith Minimizes MSE)					
Visit	0.043	-0.078**	-0.008	0.018	-0.005
Theoretical Win	-556	-207**	-41	5	-4**
Effect on Outcome (Bandwith = \$20)					
Visit	0.047	0.029	0.008	0.014*	-0.015**
Theoretical Win	-333	26.06	-11	8	-4**

Table 4: RD Estimates of Promotional Effects

most of the effects are negative or insignificant. A negative effect is plausible, as provision of dining credits and show tickets can draw customers away from the slots and the gambling tables. We thus conjecture that these credits are substituting for actual gaming. Using a band-width of \$20 for the ADW, we see that there is a positive, but insignificant, jump in theoretical win at $ADW = \$500$. But under the optimal band-width, which is smaller, we find that evidence reverses: there is a large negative effect of the increased promotions when moving from Tier 1 to Tier 2. This reversal also underscores importance of considering robustness to choices of band-width, 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 0 to 1. From Figure 6, which uses a bin-width of \$20, it may appear that there is a large jump in theoretical win at \$1000. In Figure 7, we plot the theo values at $ADW = \$1000$, within a bin-width of \$10. We see that evidence of a jump disappears locally. Robustness to band-width allows the analyst to guard against these errors.

Overall, while the general relationship between theoretical win and ADW is positive, in the limit as one explores the variation in a neighborhood of the discontinuities, the effects are either not significant or negative. The pattern for other tiers is similar. The bottom-line is that the analyses reveals that the marketing-program 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.

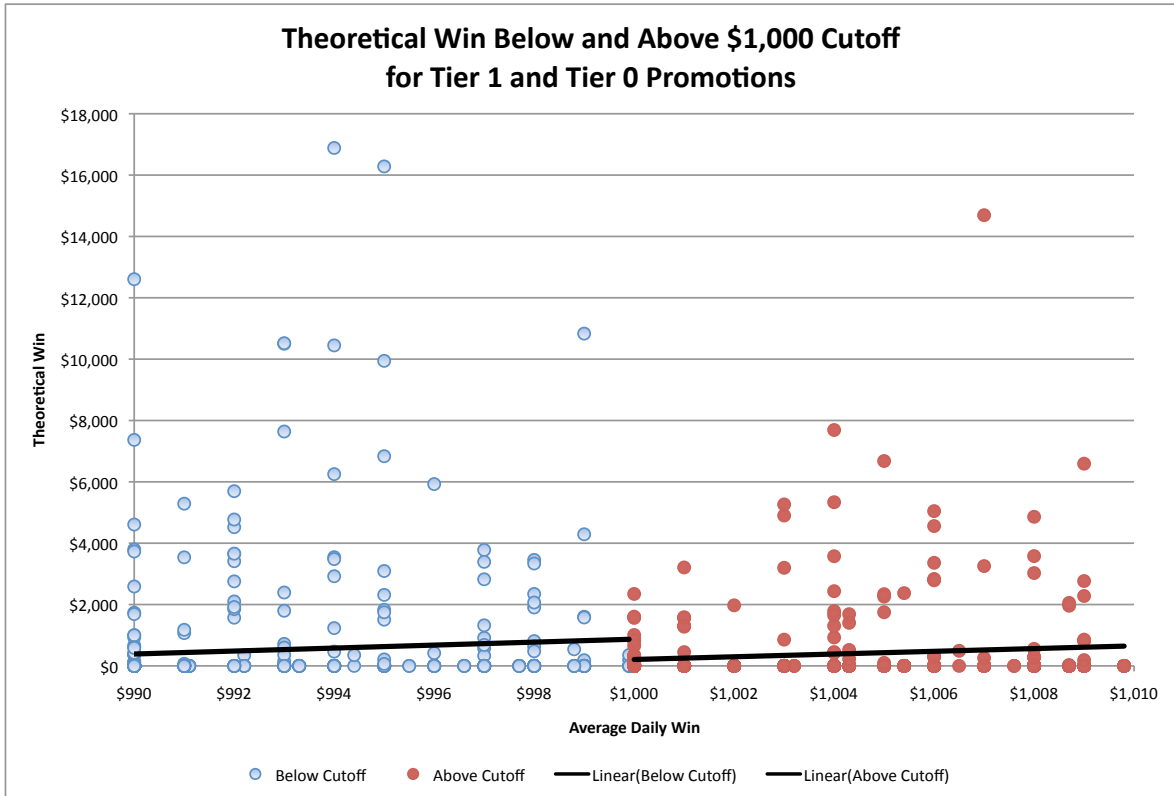


Figure 7: Theoretical Win at ADW=1000

3.2 Direct-Mail Activity by a B2C Firm

In this section we consider a second application where a B2C company sends direct-mail to customers to solicit a request for contact with the company. Once the customer contacts the company (either online or via the 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, as customers are not randomly allocated mail solicitations. Rather, as response rates are small for this mode of marketing (of the order of 1-2%), the firm tends to send direct-mail to customers that it anticipates are most likely to respond. Importantly, the firm decides the choice of customers to target based on thresholds of a one dimensional score variable, which is an (unknown to us) function of customer characteristics, zip-code features, and past response history. We observe the score variable, the threshold, as well as the responses of customers from 6 different states in the US.

We summarize the results of the analysis in Table 5. From Table 5, the difference in means appear to suggest greater response rates to the solicitation for customers with greater scores. We now explore how much of this pattern is driven by the targeting rule as opposed to the causal effect

State	Score 1			Score 2		
	AZ	WA	NJ	MN	TN	WI
Difference in Means						
Promotion Received	0.0051 (0.0012)***	0.0020 (0.0002)***	0.0019 (0.0003)***	0.0011 (0.0005)**	0.0023 (0.0005)***	0.0034 (0.0012)***
OLS Controlling for Score & Score Interacted w/ Above Cutoff						
Promotion Received	0.0040 (0.0025)	0.0005 (0.0003)*	-0.0010 (0.0006)*	-0.0005 (0.0008)	0.0011 (0.0008)	-0.0011 (0.0023)
RD Estimate with Bandwidth that Minimizes MSE						
Promotion Received	-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 Bandwidth						
Promotion Received	-0.0090 (0.0070)	-0.0001 (0.0009)	-0.0001 (0.0011)	-0.0010 (0.0014)	0.0049 (0.0017)***	-0.0078 (0.0037)**

*** is significant at 1% level, ** at 5% level and * at 10% level

Table 5: Results from Analysis of Direct-Mail Program

of direct-mail on response. As a first cut, we present results from an OLS specification that is very similar to an RD,

$$y = \tau \times I(s > \bar{s}) + \alpha + \beta s + \gamma s \times I(s > \bar{s}) + \varepsilon \quad (27)$$

where s is the score variable, \bar{s} is the threshold, and τ measures the treatment effect of interest. We estimate this specification separately for each state using only observations that lie in a small neighborhood of the threshold for each state. The table reports estimates of τ from this regression. We find that the lift from direct mail drops significantly when we focus only on individuals near the threshold, where the variation in whether a solicitation was received exists. This first-cut provides the evidence that the endogeneity bias in the naive OLS estimator is large.

The last panel presents results from a regression discontinuity specification which uses the bandwidth that minimizes the mean squared error. The identification assumption for applying the RD is that consumers do not choose housing locations so as to receive preferred direct-mail. We calibrate the band-width separately by state. Looking at the last two rows of Table 5, we see that direct-mail induces a significant lift in response only in Tennessee, and generates a significant decrease in Arizona and Wisconsin. We find the Arizona effect is not robust to increasing the bandwidth by 50%. The negative effect in Wisconsin is surprising; we conjecture, but cannot verify, that these may reflect satiation effects in that state. If Wisconsin has been heavily blanketed by direct-mail in the past, it is possible that consumers react adversely to being subject to more direct-mail activity. Another explanation is that the score and the threshold in Wisconsin picked out customers who have also been subject to heavy direct-mail by competitors¹⁵.

¹⁵This may appear to be problematic since it would violate the continuity assumption required for validity of RD.

3.3 Discussion

The RD approach in our two applications revealed several null, or negative effects. The reader should not infer this to imply that the approach will yield null effects always; the right inference is that much of the observed positive correlation in the data in both contexts happen to be 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 threshold defining a tier implies the firm can do just as well in terms of marketing by merging the adjacent tiers together. Similarly, large differences in effects across various thresholds imply that the firm may want to consider a finer classification of consumers into tiers on the basis of the current scoring variable. An interesting open question is what the *optimal* score should be in case the firm is interested in sorting consumers, or in implementing an RD. 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. But this remains an open question for future research.

4 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 exploit the thresholds defining targeted groups in an RD design to obtain precise nonparametric estimates of marketing mix effects. We discuss conditions under which the application of the estimator to targeted marketing contexts is valid. We discuss the important role of consumer self-selection, and how it can be mitigated. Two empirical applications illustrating the implementation of the method are presented. An important finding from the empirical analysis is that our quasi-experimental approach reveals that marketing mix variables

However, the treatment effect in this case can be interpreted as the net effect of the promotion of the focal firm and its competitors, and this would be a valid RD design.

sometimes have unintended effects that may not be evident otherwise.

We expect this approach to controlling for the endogeneity to be used in conjunction with other approaches developed for understanding demand under targeting. For instance, parametric models of heterogeneous demand can provide the continuous representation of heterogeneity that firms could use to better define thresholds 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.

5 References

1. Angrist, J.D., Lavy, V., (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics* 114, 533–575.
2. Black, S., (1999). “Do better schools matter? Parental valuation of elementary education.” *Quarterly Journal of Economics* 114, 577–599.
3. Busse, Meghan, Jorge Silva-Risso and Florian Zettelmeyer. (2006). “\$1000 Cash Back: The Pass-Through of Auto Manufacturer Promotions.” *American Economic Review* 96(4): 1253–1270.
4. Busse, Meghan, Duncan Simester and Florian Zettelmeyer (2009), “The Best Price You’ll Ever Get: The 2005 Employee Discount Pricing Promotions in the U.S. Automobile Industry,” *Marketing Science*, forthcoming.
5. Chen, M. Keith and Jesse M. Shapiro, (2007) “Does Prison Harden Inmates? A Discontinuity-based Approach,” Working Paper 1450, Cowles Foundation.
6. Chay, Kenneth Y. and Michael Greenstone (2005) “Does Air Quality Matter? Evidence from the Housing Market,” *Journal of Political Economy*, April, 113 (2), 376–424.
7. Conlisk, J. Gerstner, E. and Sobel, J. (1984). “Cyclic Pricing by a Durable Good Monopolist”, *Quarterly Journal of Economics* (August), 489-505.
8. Ellickson, Paul and Misra, Sanjog (2007), “Enriching Interactions: Incorporating Outcome Data into Static Discrete Games,” working paper, University of Rochester.

9. Fan, J., Gijbels, I., (1996). "Local Polynomial Modelling and its Applications." Chapman & Hall, London.
10. Hahn, J., Todd, P., Van Der Klaauw, W., (2001). "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design." *Econometrica* 69, 201–209.
11. Hardle, W., (1990). "Applied Nonparametric Regression." Cambridge University Press, Cambridge.
12. Huang, Ka-Wei (2009). "Optimal Criteria for Selecting Price Discrimination Metrics When Buyers Have Log-normally Distributed Willingness-to-Pay," *Quantitative Marketing and Economics*, 7(3), 321-341.
13. Imbens, G., Angrist, J., (1994). "Identification and Estimation of local average treatment effects." *Econometrica* 61, 467–476.
14. Kao, L. J, Chiu, C-C., Gilbride, T.J., Otter, T., and Allenby, G.M. (2005). "Evaluating the Effectiveness of Marketing Expenditures," working paper, Ohio State University.
15. Lal, Rajiv, and Patricia Carrolo (2001). "Harrah's Entertainment Inc.," Harvard Business School Case 502-011.
16. Lee, D., Enrico Moretti, and Matthew Butler (2004). "Do Voters Affect or Elect Policies? Evidence from the U.S. House," *Quarterly Journal of Economics*, August, 119 (3), 807–859.
17. Lee, D.S., (2008). "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics*, 142(2), February.
18. Lee, D. S. and Thomas Lemieux, (2009). "Regression Discontinuity Designs in Economics," NBER Working Papers 14723, National Bureau of Economic Research.
19. Manchanda, P., P. E. Rossi and P.K. Chintagunta (2004), "Response Modeling with Non-Random Marketing Mix Variables," *Journal of Marketing Research*, 41 (November), 467-478.
20. Porter, J., (2003). "Estimation in the Regression Discontinuity Model." Mimeo. Department of Economics, University of Wisconsin.
21. Rossi, Peter E., G. Allenby and R. McCulloch (1996). "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15, 321-340.
22. Thistlewaite, D., Campbell, D., (1960). "Regression-discontinuity analysis: an alternative to the ex-post facto experiment." *Journal of Educational Psychology* 51, 309–317.

23. Trochim, W., (1984). "Research Design for Program Evaluation: The Regression-Discontinuity Design." Sage Publications, Beverly Hills, CA.