

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

Wesley Hartmann* Harikesh Nair† Sridhar Narayanan‡

September 2009[§]

WORK-IN-PROGRESS - INCOMPLETE

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 exploit 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 show theoretically and using simulations the conditions under which the RD estimator may be valid, considering in particular, the role of strategic consumer self-selection. 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), and at the Marketing Science Conferences (2007 & 2009), for their helpful feedback. All errors are our own.

1 Introduction

Targeted marketing has become a 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. 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. 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 targeted marketing.

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 assignment is 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, Chay et al. 2005); to housing (Chay and Greenstone 2005, Cellini et al. 2008); to the social effects of imprisonment (Chen and Shapiro 2007); and auction design (Yuan 2008) amongst others. In contrast, applications to Marketing and Industrial Organization contexts have been very sparse. These include Busse et al. (2008 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 great 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. A second aspect is that targeting policies of firms often involve trigger rules. Marketing allocation policies often involve “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-specific threshold of a continuous score variable. For instance, catalogs might be mailed based on levels of underlying ‘RFM’ score variables, credit card promotions may be given based on cutoffs of FICO[®] scores, detailing calls may be made to a physician based on whether she is in specific prescription-based deciles, price discounts may be given to people above or below certain age cutoffs, etc. The ubiquity of such trigger-rules generate a wealth of discontinuity-based contexts that facilitate nonparametric identification of treatment effects (in this instance, the effect of the marketing activity) using an RD design.

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, 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 (e.g. Rossi et al. 2003). While this approach is feasible if full information is available to the analyst about how the firm allocates its marketing efforts, it is prone to misspecification bias in the absence of such information. In some situations, such misspecification bias can be large. More generally, a common practice to solve endogeneity problems is to augment the model with a parametric assumption about the nature of firm’s targeting policies. By imposing specific functional forms that relate all of the observed units within the empirical analysis, parametric approaches can obtain inference from much less data. However,

functional forms and parametric statistical distributions impose strong assumptions that have the potential to misinform researchers about actual causal effects.

By contrast, the regression discontinuity approach does not depend on the use or validity of instrumental variables.¹ Additionally, the estimates obtained are nonparametric, thus mitigating concerns about misspecification biases. Unlike parametric approaches, which combine information from substantively different observations, the nonparametric approach takes advantage of extensive data only in the region where variation in the marketing mix is observed, so as to obtain local, but more credible, estimates of causal effects. A further advantage is that the estimator is easy to implement. Estimation 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 treatment levels between the left and the right. The estimator is also favorable to firms as they can take advantage of pre-existing discontinuous rules-of-thumb to identify causal 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, more parametric assumptions may be required. Second, the estimator provides a local treatment effect which is relevant for the subpopulation of consumers at the threshold, and not globally.² A third caveat is that, like any other alternative, the conditions for the validity of the estimator have to be carefully assessed depending on the context. We consider the last aspect 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 an underlying formal model of agent behavior which considers how consumers sort at the threshold.

We present two illustrations of the identification conditions that cover the range of targeting marketing contexts observed in practice. In each, we start with an underlying model of strategic decision-making by customers, which we link to continuity conditions for the application of an RD estimator. We consider in detail the ability of consumers to *precisely* self-select into either side of the trigger, and its implica-

¹Indeed, 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 cutoff.

²In 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.

tion for RD applications of targeting. Precise strategic selection has been one of the main challenges in the application of RD designs in practice (see Lee and Card 2007). 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). Second, we present a detailed illustration to a loyalty program in which selection is problematic, and discuss approaches to mitigate the selection bias. If customers are uncertain about the exact score, the threshold, or both, they will be unable to select around the threshold to achieve preferential treatment.³ First, we illustrate that uncertainty about the thresholds in the reward program retains the validity of the RD estimator.

We then apply the RD estimator to two empirical applications 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 that 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 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 non-parametrically using local linear regression (Fan and Gijbels 1996). We find that in both cases, controlling for 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 identification. Second, we present a detailed illustration of the identifiability of causal

³For 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.

marketing effects using the design, and show theoretically the conditions under which the RD estimator may be valid, 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, $\mathcal{E}[y_i(1) - y_i(0)]$. The RD design implies that treatment is assigned depending on whether a continuous score, x_i , crosses a threshold, \bar{x} , i.e., $d_i = I(x_i > \bar{x})$. Then, the observed size of the discontinuity in the outcome in a neighborhood of \bar{x} are,

$$y^- = \lim_{x \rightarrow \bar{x}^-} \mathcal{E}[y_i(0) | x_i = \bar{x}] \text{ and } y^+ = \lim_{x \rightarrow \bar{x}^+} \mathcal{E}[y_i(1) | x_i = \bar{x}] \quad (1)$$

and of the treatment are,

$$d^- = \lim_{x \rightarrow \bar{x}^-} \mathcal{E}[d_i | x_i = \bar{x}] \text{ and } d^+ = \lim_{x \rightarrow \bar{x}^+} \mathcal{E}[d_i | x_i = \bar{x}] \quad (2)$$

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

That is, to obtain a causal effect of the treatment, we simply estimate the discontinuity in the outcomes, and weigh it down by the discontinuity in treatments.⁴ The interpretation of size of the discontinuity in outcomes as a treatment effect hinges

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.

crucially on the continuity conditions. The continuity of $\mathcal{E}[y_i(1)]$, and $\mathcal{E}[y_i(0)]$ in x_i enables us to interpret the outcomes just below \bar{x} as a valid counterfactual for the outcomes just above \bar{x} , thereby facilitating a relevant control group for measurement of a treatment.

An assesment of the validity of an RD applciation thus depends on whether continuity in counterfactual outcomes 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.

We now illustrate the importance of structural modeling features in establishing the identifiability of the RD estimator in various context. 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. While this point is fairly straightforward, it allows us to provide an example of selection and potential identification problems. We then alter the model to characterize purchase history targeting. In this case, a customer can select their score by making a binary purchase decision that increases their score by a fixed amount. An important feature of this model is randomness in the purchase decision which can prevent some individuals from obtaining the cutoff. We show that if the customer knows the cutoff, the RD estimator is invalidated because the distribution of customers just below and just above the cutoff is different. However, if customers have a continuous distribution of uncertainty around the true cutoff, then the RD estimator is valid. In this case, the structure of the model required specification of information available to the agent (i.e. do they know the cutoff?), the beliefs of the agent (i.e. properties of their beliefs about the distribution of the cutoff), and whether or not the agent knows the unobservables entering their choice equation. These two model specifications illustrate the margins at which our empirical specifications are identified and the importance of structure.

2.1 Illustrative Selection Model for RD

The model involves two stages. In the first stage, the customer makes a selection decision that can affect the score and treatment. In the second stage, the customer makes a decision about the outcome of interest. We consider the outcome stage first, then work back to the selection stage

2.2 Outcome Stage

The outcome Y is a binary variable indicating whether or not an individual makes a purchase. Treatment is indicated by the binary variable d . An individual receives treatment if the value of their score, z' , is greater than a threshold, \bar{z} . We model the individual's outcome as a random utility model, where:

$$\begin{aligned} u_1(z', X, R, \varepsilon_1 | \beta, \bar{z}) &= v(X, R = 1 | \beta) d(z' \geq \bar{z}) + v(X, R = 0 | \beta) d(z' < \bar{z}) + \varepsilon_1 \\ u_0(\varepsilon_0) &= 0 + \varepsilon_0 \end{aligned}$$

v -s indicate the non-stochastic portion of the individuals utility of choosing to purchase, while the non-stochastic utility of not purchasing is normalized to zero. We assume that receiving treatment is preferred, such that $v(X, R = 1 | \beta) > v(X, R = 0 | \beta)$.

2.3 Selection Stage

In the first stage, each customer is endowed with some initial value of the score, z . The score relevant for treatment, z' , is defined:

$$z' = z + m$$

In words, the customer can choose a value m that shifts the score to a value that might yield preferential treatment. Moving the score is not however costless. There are fixed and marginal costs totalling $C = F + \tau m$.

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

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

2.4 Identification

We now consider the identification conditions from HTV regarding the existence of the expected outcomes at the limits and the continuity of the control outcome at the cutoff.

The limit of the outcomes from above the score exist as long as z has positive mass just above the score. The limit from below will not however exist if the marginal customer z^* is below the cutoff, \bar{z} . The limit therefore does not exist if:

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

Another possible way for the limit to exist is if there are heterogeneous individuals such that the fixed costs of selection are sufficiently large to prevent selection for at least some of the individuals. However, we show next that, unless all customers have sufficiently large fixed costs, the continuity in the expected control outcome fails in the presence of this type of heterogeneity.

To evaluate the continuity of the expected control outcomes, we define the expected control taking the limit from below and above the cutoff.

At the limit below the cutoff, \bar{z}^- , the only customers there will be those for which $z = \bar{z}^-$ and chose not to move. Above the cutoff, there is a combination of those customers with $z = \bar{z}^+$ that chose not to move and those customers with some z less than the cutoff but close enough to select up to \bar{z}^+ . This implies that, even when there is an identical distribution of z above and below the cutoff, there will be more customers above the cutoff because of the selection. Integrating over types then implies that the fraction of customers above may be greater for some types, such that the expected outcome either above or below differs because the type distribution is different.

$$\begin{aligned}
\lim_{z' \rightarrow \bar{z}^-} E_\varepsilon [Y^{R=0} | z' = z + m] &= \frac{\int E [Y^{R=0} | \theta] f(z = \bar{z}^- | \theta) h(\theta) d\theta}{\int f(z = \bar{z}^- | \theta) h(\theta) d\theta} \\
&= \int E [Y^{R=0} | \theta] \frac{f(z = \bar{z}^- | \theta) h(\theta)}{\kappa^-} d\theta \\
\lim_{z' \rightarrow \bar{z}^+} E_\varepsilon [Y^{R=0} | z' = z + m] &= \frac{\int E [Y^{R=0} | \theta] [f(z = \bar{z}^+ | \theta) + F(z^- | \theta) - F(z^*(\theta) | \theta)] h(\theta) d\theta}{\int [f(z = \bar{z}^+ | \theta) + F(z^- | \theta) - F(z^*(\theta) | \theta)] h(\theta) d\theta} \\
&= \int E [Y^{R=0} | \theta] \frac{[f(z = \bar{z}^+ | \theta) + F(z^- | \theta) - F(z^*(\theta) | \theta)] h(\theta)}{\kappa^+} d\theta
\end{aligned}$$

This can be seen more clearly if we are willing to make a simplifying assumptions. Suppose $f(z|\theta) \sim U(0, \hat{z})$. as

$$\begin{aligned}
\lim_{z' \rightarrow \bar{z}^-} E_\varepsilon [Y^{R=0} | z' = z + m] &= \frac{\frac{1}{\hat{z}} \int E [Y^{R=0} | \theta] h(\theta) d\theta}{\frac{1}{\hat{z}} \int h(\theta) d\theta} = \int E [Y^{R=0} | \theta] h(\theta) d\theta \\
\lim_{z' \rightarrow \bar{z}^+} E_\varepsilon [Y^{R=0} | z' = z + m] &= \frac{\int E [Y^{R=0} | \theta] \left(\frac{1}{\hat{z}} + \int_{z^*(\theta)}^{\bar{z}^-} \frac{1}{\hat{z}} dz \right) h(\theta) d\theta}{\int \left(\frac{1}{\hat{z}} + \int_{z^*(\theta)}^{\bar{z}^-} \frac{1}{\hat{z}} dz \right) d\theta} \\
&= \int E [Y^{R=0} | \theta] \frac{\left(\frac{1}{\hat{z}} + \frac{\bar{z}^- - z^*(\theta)}{\hat{z}} \right)}{\kappa} h(\theta) d\theta
\end{aligned}$$

We see that the limit from below is the expected outcome of each type, θ , weighted by their mass, $h(\theta)$. Above the cutoff, the weight is adjusted by how many of these types were likely to select up from below the cutoff.

Note that in all the equations above, the expectations are conditioned on X and β , though we have dropped these conditioning arguments to simplify the notation.

The above analysis suggests that geographic targeting will typically be a valid RD application because preferential marketing treatment 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 turn to such a model where the score variable is a customer's purchase history, i.e. something much less costly to change.

2.5 History-Based Targeting

We now consider the economic forces outlined in the stylized model above using a detailed example of a loyalty program. The discussion below generalizes the previous model to situations where targeting may also be based on the past behavioral history of the consumers. An important consideration is that forward-looking consumers may adjust current *consumption* so as to obtain treatments in the future, thereby generating discontinuous sorting at the threshold, and thus invalidating the RD design. We illustrate circumstances that may mitigate these concerns. Throughout, we underscore the importance of having a formal model of behavior, and careful consideration of consumers' information sets and incentives to establish the validity or lack thereof, of an RD-based analysis.

This section is divided into three parts. We start by presenting a canonical example of a frequent-flier program which we use to illustrate the consequences of precise selection around the score. We start with an underlying model of consumer behavior and relate it to continuity conditions of HTV. We then discuss the mitigating role of consumer uncertainty about the threshold. The analysis suggests that RD is not likely to be viable in traditional frequency reward programs, where thresholds are communicated to customers as purchase incentives. On the other hand, when firms analyze purchase histories to define targeted groups for their "non-contractual" promotion activity, the fact that customers are unaware of thresholds for better treatment prevents the type of selection that invalidates RD.

2.5.1 Selection in a frequent-flier program

We extend the model in the previous section, such that a customer selects their score by making a binary purchase decision, indicated by y . In this case, the score is now

defined:

$$z' = z + m \times y(m, z, \varepsilon) \quad (3)$$

where we assume m is the exogenously determined points in this purchase decision. In an airline program, m would be the miles a business traveler is required to fly for a meeting, while z is the travelers existing miles and y is whether or not the traveler purchases from the airline with which it has the potential to earn a reward. Keeping with our previous setup, we however assume that the reward only affects purchase in the second period, i.e. the outcome stage defined above.

Note that we have defined the purchase decision to be a function of the variables that will determine whether a reward is received, m and z , as well as a random component ε that might give the customer a large incentive or disincentive to purchase from the firm (e.g. the customer's preferred airline has multiple connections on this route).

To define a customer's selection/purchase decision, we begin by specifying the expected future payoffs of the decision:

$$\begin{aligned} EV(z', R|z, m, y) &= E_{\bar{z}, \varepsilon} [\max \{v(X, R, z', \bar{z}|\beta) + \varepsilon_1, 0 + \varepsilon_0\} | z, m, y] \\ &= \left[\begin{array}{l} \Pr(z + my \geq \bar{z}) E_{\varepsilon} [\max \{v(X, R = 1|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \\ + \Pr(z + my < \bar{z}) E_{\varepsilon} [\max \{v(X, R = 0|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \end{array} \right] \\ &= \left[\begin{array}{l} G(z + my) E_{\varepsilon} [\max \{v(X, R = 1|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \\ + (1 - G(z + my)) E_{\varepsilon} [\max \{v(X, R = 0|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \end{array} \right] \end{aligned}$$

where G is the c.d.f. of the customers beliefs about \bar{z} , and all other variables are defined as above. In the typical frequency reward program, G has mass 1 at the true value of \bar{z} , because the firm communicates thresholds to customers to incentivize purchase. When customers receive targeted offers based on their purchase history, they typically do not know the criteria the firm uses to decide who gets which offer. In these cases customers are likely to have some continuous belief distribution about the thresholds. As we show below, the continuity of G can assure that selection does not invalidate RD.

We now specify the choice specific values that determine purchase in the selection stage:

$$\begin{aligned} V_1(z, X_0|\beta) &= v_{1,0}(X_0|\beta) + \delta EV(z', R|z, m, y = 1) + \varepsilon_{1,0} \\ V_0(z, X_0|\beta) &= 0 + \delta EV(z', R|z, m, y = 0) + \varepsilon_{1,0} \end{aligned}$$

where the additional 0 subscript on some variables denotes the selection stage as opposed to the outcome stage. The probability of purchase is therefore defined as follows:

$$\Pr(\varepsilon_{1,0} - \varepsilon_{0,0} > -[v_{1,0}(X_0|\beta) + \delta[Q(z + m) - Q(z)]])$$

where

$$Q(z) = \begin{aligned} & G(z) E_\varepsilon [\max \{v(X, R = 1|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \\ & + (1 - G(z)) E_\varepsilon [\max \{v(X, R = 0|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \end{aligned}$$

It can be seen that:

$$Q(z+m) - Q(z) = [G(z+m) - G(z)] \cdot \begin{bmatrix} E_\varepsilon [\max \{v(X, R = 1|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \\ -E_\varepsilon [\max \{v(X, R = 0|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \end{bmatrix}$$

Thus, we see 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. $G(z+m) - G(z)$.

We now evaluate how the selection/purchase probability affects the continuity of the control outcome at the cutoff. In the equations below, we drop the conditioning arguments on X and β to simplify the notation.

$$\begin{aligned} \lim_{z' \rightarrow \bar{z}^+} E_\varepsilon [Y^{R=0}|z] &= \frac{\int E_\varepsilon [Y^{R=0}] \left[\frac{\Pr(y=0|z=\bar{z}^+, \theta) f(z=\bar{z}^+|\theta)}{+\Pr(y=1|z=\bar{z}^+ - m, \theta) f(z=\bar{z}^+ - m|\theta)} \right] h(\theta) d\theta}{\int \left[\frac{\Pr(y=0|z=\bar{z}^+, \theta) f(z=\bar{z}^+|\theta)}{+\Pr(y=1|z=\bar{z}^+ - m, \theta) f(z=\bar{z}^+ - m|\theta)} \right] h(\theta) d\theta} \\ &= \int E_\varepsilon [Y^{R=0}] \left[\frac{\Pr(y=0|z=\bar{z}^+, \theta) f(z=\bar{z}^+|\theta)}{+\Pr(y=1|z=\bar{z}^+ - m, \theta) f(z=\bar{z}^+ - m|\theta)} \right]_{\kappa^+} h(\theta) d\theta \\ \lim_{z' \rightarrow \bar{z}^-} E_\varepsilon [Y^{R=0}|z] &= \frac{\int E_\varepsilon [Y^{R=0}] \left[\frac{\Pr(y=0|z=\bar{z}^-, \theta) f(z=\bar{z}^-|\theta)}{+\Pr(y=1|z=\bar{z}^- - m, \theta) f(z=\bar{z}^- - m|\theta)} \right] h(\theta) d\theta}{\int \left[\frac{\Pr(y=0|z=\bar{z}^-, \theta) f(z=\bar{z}^-|\theta)}{+\Pr(y=1|z=\bar{z}^- - m, \theta) f(z=\bar{z}^- - m|\theta)} \right] h(\theta) d\theta} \\ &= \int E_\varepsilon [Y^{R=0}] \left[\frac{\Pr(y=0|z=\bar{z}^-, \theta) f(z=\bar{z}^-|\theta)}{+\Pr(y=1|z=\bar{z}^- - m, \theta) f(z=\bar{z}^- - m|\theta)} \right]_{\kappa^-} h(\theta) d\theta \end{aligned}$$

We analyze the above expressions under two scenarios: unknown and known cutoffs. First, when the cutoff is unknown, the selection/purchase probability just below and just above a point such as \bar{z} or $\bar{z} - m$ converge to the same amount. Figure X graphically depicts this scenario. Continuity of G around both \bar{z} and $\bar{z} - m$ assure that the purchase regions in the top panel of the figure will be the same on both sides of these points, such that the implied distribution of z' is the same on both sides of \bar{z} .

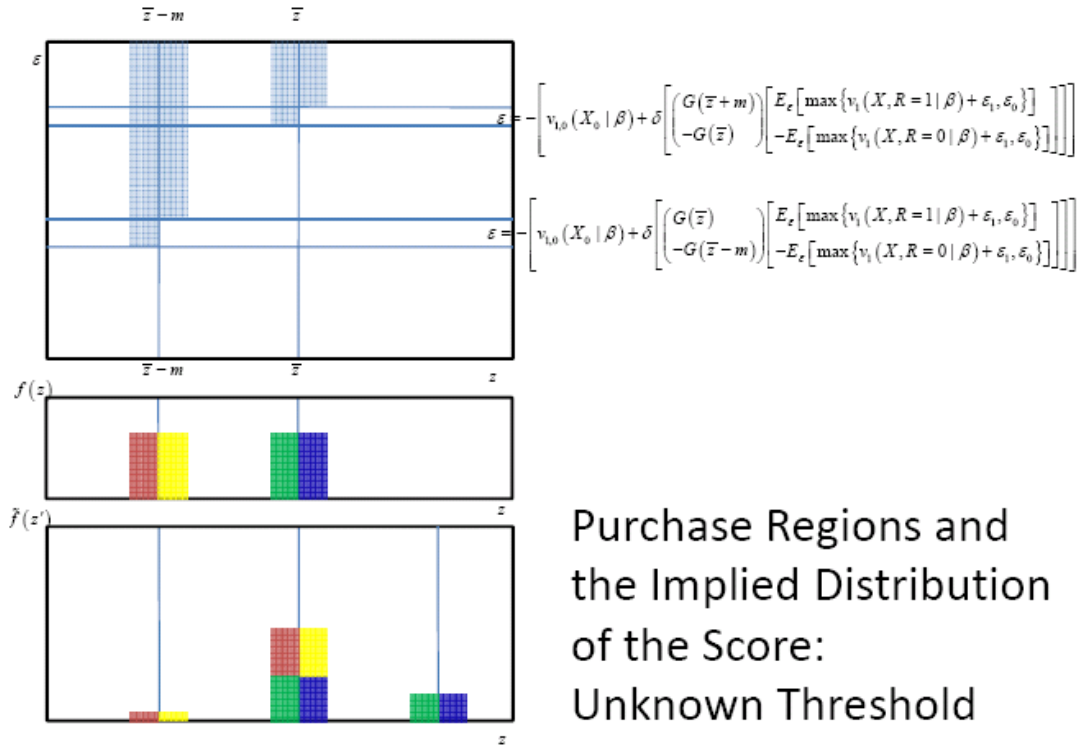


Figure 1:

Second, when the cutoff is known, the purchase regions are distinctly different. In this case, $G(z + m) - G(z) = 1$ if $z = \bar{z}^-$ because a selection/purchase moves a customer from no treatment to treatment with perfect certainty. The same holds for all values of z down to $\bar{z}^+ - m$. On the other hand, $G(z + m) - G(z) = 0$ if $z = \bar{z}^+$ because the customer obtains the preferential treatment whether she purchases or not. At $z = \bar{z}^- - m$, the expression also equals zero because the customer has no possibility of receiving the treatment, whether purchasing or not. Figure Y depicts this pattern and we see that it implies a discontinuous distribution of z' around \bar{z} .

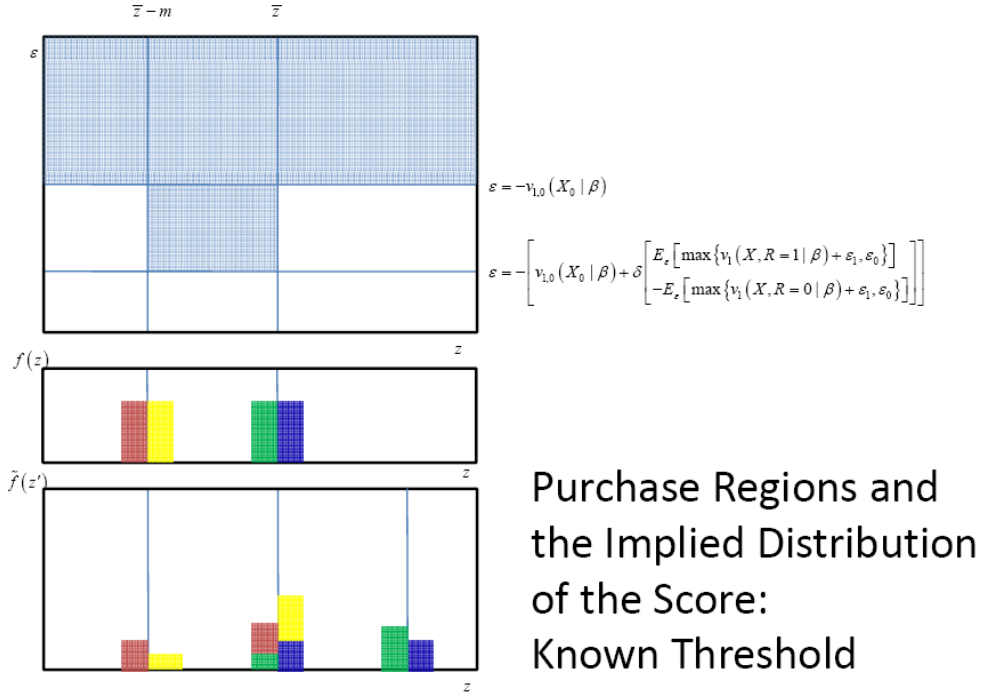


Figure 2:

The discontinuity of the distribution of z' around \bar{z} breaks the continuity of the expected control outcome around \bar{z} because customers with different types, θ , have different relative masses just below and just above the cutoff (i.e. the size of the discontinuity depends on θ). The expected control outcome just below the cutoff, for instance, therefore becomes a weighted average of all types, where those types that value the reward most are weighted less because they are less likely to stay at \bar{z}^- .

The above analysis therefore documents that traditional reward programs where the cutoffs are communicated to customers are not valid RD applications because the discrete change in purchase incentives at the cutoffs creates a discontinuity in expected outcomes in the absence of treatment. However, firms targeted marketing based on purchase histories, where cutoffs are not communicated to customers are a viable RD application as long as the customers have a continuous distribution of beliefs about the location of the cutoff around the true cutoff, \bar{z} , and the true cutoff less the additional points available to get there, i.e. $\bar{z} - m$.

2.5.2 Uncertainty about the score vs. the cutoff

While the above analysis documents how uncertainty about the cutoff provides the continuity restrictions necessary to use RD, we now show that uncertainty about the score works in a similar manner. Returning to Equation XXXX above, we merely need to redefine the distribution of the customer's beliefs, G , to be with regard to their score as opposed to the threshold. Modifying the equation, we obtain:

$$\begin{aligned}
 EV(z', R|z, m, y) &= E_{\bar{z}, \varepsilon} [\max \{v(X, R, z', \bar{z}|\beta) + \varepsilon_1, 0 + \varepsilon_0\} | z, m, y] \\
 &= \left[\begin{array}{l} \Pr(z + my \geq \bar{z}) E_{\varepsilon} [\max \{v(X, R = 1|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \\ + \Pr(z + my < \bar{z}) E_{\varepsilon} [\max \{v(X, R = 0|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \end{array} \right] \\
 &= \left[\begin{array}{l} (1 - G(\bar{z} - my)) E_{\varepsilon} [\max \{v(X, R = 1|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \\ + G(\bar{z} - my) E_{\varepsilon} [\max \{v(X, R = 0|\beta) + \varepsilon_1, 0 + \varepsilon_0\}] \end{array} \right]
 \end{aligned}$$

Continuity in G at $z \in \{\bar{z}, \bar{z} - m\}$ assures that choice specific value functions at \bar{z}^- and \bar{z}^+ (as well as $\bar{z}^- - m$ and $\bar{z}^+ - m$), will be the same. One way to interpret this is to redefine the score to be a function of a variable v , that is unobservable to the agent.

$$z' = z + m \times y(m, z, \varepsilon) + v \quad (4)$$

Note that in defining the indicator for treatment, whether the uncertainty is with respect to the threshold or the score, merely involves moving v to either side of the inequality sign:

$$\text{Unknown Score} : d(z' \geq \bar{z}) = d(z + m \times y(m, z, \varepsilon) + v \geq \bar{z}) \quad (5)$$

$$\text{Unknown Cutoff} : d(z' \geq \bar{z}) = d(z + m \times y(m, z, \varepsilon) \geq \bar{z} + v) \quad (6)$$

In our customer based targeting casino application below, identification of the RD estimator is provided by both of these features. 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 cutoffs 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.

2.5.3 Discussion

The discussion above highlights the role of selection, and illustrates the difference between precise and imprecise sorting for the validity of RD application. We conclude this section by laying out the main implication of the above results for applied researchers and firms looking to implement the RD design in practice.

Precision about timing and uncertainty We wish to emphasize the importance of being very precise about agents’s information sets, and the timing of resolution of uncertainty. In the example above, mere randomness is not enough, i.e., the presence of an error term ε does not solve the selection-induced discontinuity, as ε is assumed to be observed by consumers prior to making the selection decision. Observability of ε implies agents can sort precisely given their preferences and information available to them; but only imprecise sorting will generate the needed smoothing. Observability of ε is the standard assumption in most canonical discrete-choice models; hence, these models of behavior will by definition, generate selection-induced discontinuities, if the analyst believes those to be the models generating the data.⁵ The analyst justifying the RD design in the presence of selection on the basis of randomness thus has to make a very specific case why there exists uncertainty for the agent when making the selection decision. This is likely to be context specific.

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

⁵Interestingly, in the context of this model, making ε unknown to consumers (and no v) may not be enough due to the inherent discreteness of y , which prevents smoothing via ε . One could envisage others contexts in which y is continuous, e.g., $y = (x, z)' \beta + \varepsilon$, where unobservability of ε is sufficient.

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 casinos’ best estimate of the average revenue the casino can earn from the customer per day of his visit, after controlling for luck.⁶ 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 cutoffs 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 daily win (ADW) they expect from the customer.⁷

We observe four such mailings between January and September 2006 (we report only one of the mailing 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 (1) holds across mailings

⁶The 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.

⁷Note 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.

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

(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 an 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 Table (2) 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 outcomes 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. 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 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

	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

	Tiers				
	0 to 1	1 to 2	2 to 3	3 to 4	4 to 5
ADW Cutoff	\$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

a visit and 0 otherwise.⁸ To obtain unconditional (of visit) effects, the the 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 an effect on inducing consumers to visit the casino 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.

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 (2) 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 (3) 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

⁸We 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.

at which we intend to measure treatment effects of the promotions.

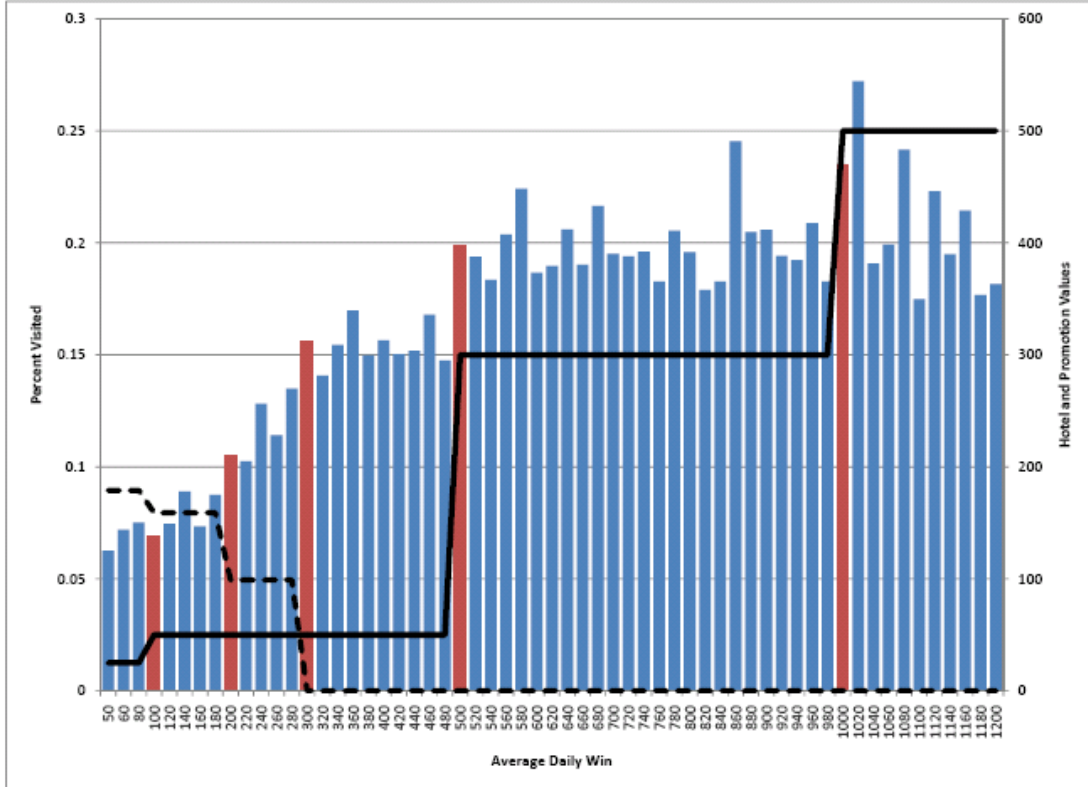


Figure 3: Percent Visiting Casino as a Function of ADW

Looking at Figure (3), 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 cutoffs. 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 are working for that set of customers. Figure (4) presents an analogous plot for the theoretical win. Eyeballing Figure (4), 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.

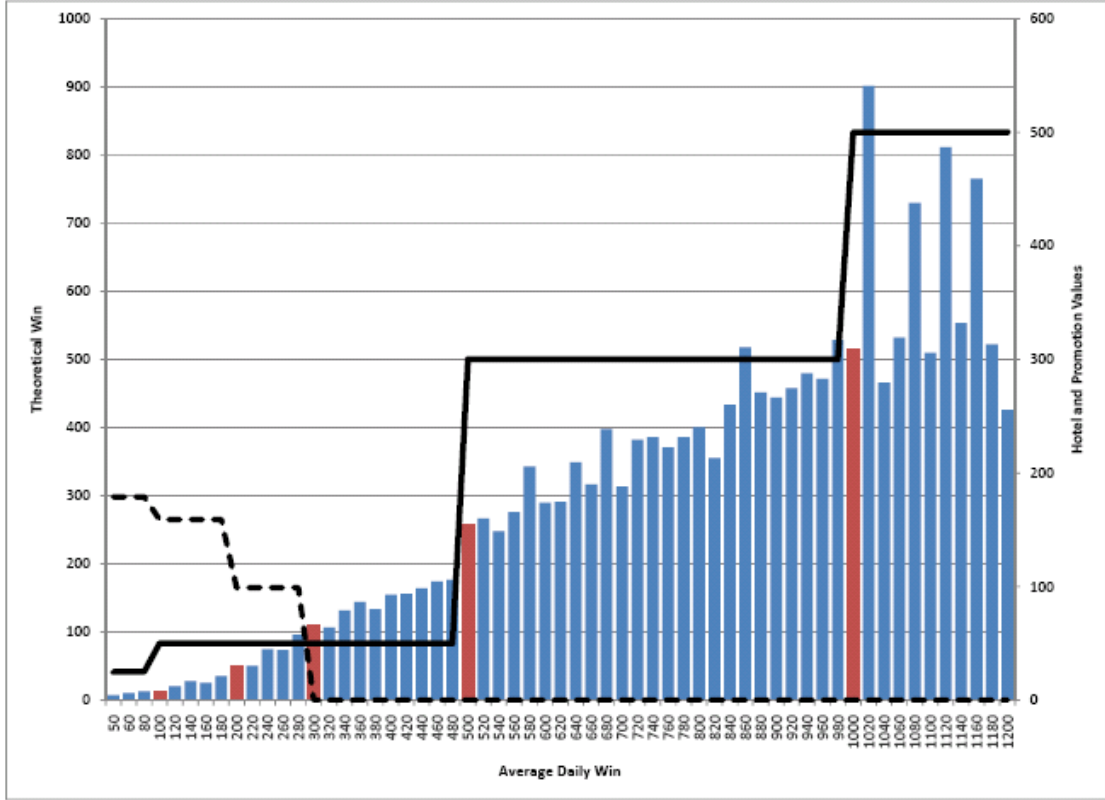


Figure 4: Theoretical Win as a Function of ADW

Figure (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.⁹ Following the suggestion of Imbens and Lemiux (2008), we estimate these non-parametrically using a rectangular kernel, using observations within a bandwidth of size h on either side of each cutoff. Because there are six different tiers, and five different ADW 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 cutoffs. An advantage of this approach is that the standard errors of the rectangular kernel are the same as robust standard errors available for least squares. We also estimated the RD using local linear regression (not reported), and

⁹As 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.

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
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

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 band-width of ADW that minimizes the mean square error across the five cutoffs. For completeness, we also present estimates using the same band-width used to generate the plots in Figures (3 & 4).

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, 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 $ADW = \$1000$, where there is a change of tiers from 0 to 1. From Figure (4), which uses a bin-width of \$20, it may appear that there is a large jump in theoretical win at \$1000. In Figure (5), we plot the theo values at $ADW = \$1000$, using a bin-width of \$2. We see that evidence of a jump disappears locally. Robustness to band-width allows the analyst to guard against these errors.

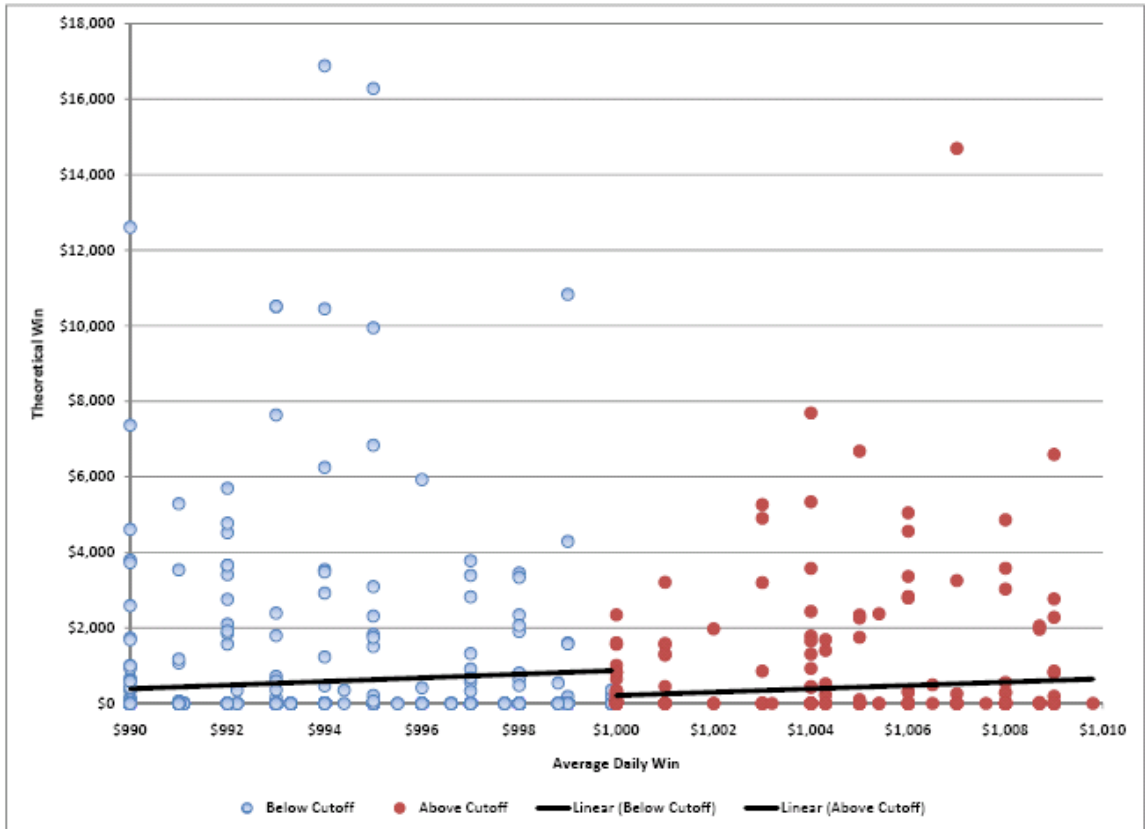


Figure 5: Theoretical Win at $ADW=1000$

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,

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

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.

3.2 Direct-Mail Activity by a B2C Firm

In this section we consider an 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 is most likely to respond. Importantly, the firm decides the choice of customers to target based on thresholds 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.

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

$$y = \tau \times I(s > \bar{s}) + \alpha + \beta s + \gamma s \times I(s > \bar{s}) + \varepsilon \quad (7)$$

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. We calibrate the bandwidth 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. Though not reported, 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 competitors. Hence, they view additional direct-mail from the focal company as a nuisance.

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 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. The approach utilizes the crude thresholds defining targeted groups 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 sometimes have unintended effects that may not be evident otherwise.

We consider our approach to be complementary to other more parametric approaches. 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 these approaches, even if firms are unable to implement the individual-level policies that can be suggested by individual-level models. Parametric methods that enable pooling are also essential in sparse-data situations. In data-rich environments, we hope that this paper encourages further exploration of the use of non-parametric methods that 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. Fan, J., Gijbels, I., (1996). "Local Polynomial Modelling and its Applications." Chapman & Hall, London.
4. Hahn, J., Todd, P., Van Der Klaauw, W., (2001). "Identification and estimation of treatment effects with a regression discontinuity design." *Econometrica* 69, 201–209.
5. Hardle, W., (1990). "Applied Nonparametric Regression." Cambridge University Press, Cambridge.
6. Heckman, J.J., Hotz, J., (1989). "Alternative methods for evaluating the impact of training programs (with discussion)." *Journal of the American Statistical Association* 84, 862–874.
7. Hendel, I. and Aviv Nevo. (2006). "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, November 74(6), 1637-73.
8. 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.
9. Imbens, G., Angrist, J., (1994). "Identification and estimation of local average treatment effects." *Econometrica* 61, 467–476.
10. Lal, Rajiv, and Patricia Carrolo (2001). "Harrah's Entertainment Inc.," Harvard Business School Case 502-011.
11. Lee, D.S., (2008). "Randomized experiments from non-random selection in U.S. house elections." *Journal of Econometrics*, 142(2), February.
12. Lee, D.S., Card, D., (2008). "Regression discontinuity inference with specification error." *Journal of Econometrics*, 142(2), February.
13. Lee, D. S. and Thomas Lemieux, (2009). "Regression Discontinuity Designs in Economics," NBER Working Papers 14723, National Bureau of Economic Research.
14. Ludwig, J., Miller, D., (2007). "Does head start improve children's life chances? Evidence from a regression discontinuity design." *Quarterly Journal of Economics* 122 (1), 159–208.
15. Matsudaira, J., (2008). "Mandatory summer school and student achievement." *Journal of Econometrics*, 142(2), February.

16. McCrary, J., (2008). "Testing for manipulation of the running variable in the regression discontinuity design." *Journal of Econometrics*, 142(2), February.
17. Porter, J., (2003). "Estimation in the Regression Discontinuity Model." Mimeo. Department of Economics, University of Wisconsin.
18. Rosenbaum, P., Rubin, D., (1983). "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70, 41–55.
19. Rubin, D., (1974). "Estimating causal effects of treatments in randomized and non-randomized studies." *Journal of Educational Psychology* 66, 688–701.
20. Thistlewaite, D., Campbell, D., (1960). "Regression-discontinuity analysis: an alternative to the ex-post facto experiment." *Journal of Educational Psychology* 51, 309–317.
21. Trochim, W., (1984). "Research Design for Program Evaluation: The Regression-Discontinuity Design." Sage Publications, Beverly Hills, CA.