Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://amstat.tandfonline.com/loi/ubes20

Regression Discontinuity Designs With Sample Selection

Yingying Dong

To cite this article: Yingying Dong (2017): Regression Discontinuity Designs With Sample Selection, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2017.1302880

To link to this article: https://doi.org/10.1080/07350015.2017.1302880

View supplementary material

Accepted author version posted online: 13 Mar 2017.
Published online: 08 Aug 2017.

Submit your article to this journal

Article views: 354

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://amstat.tandfonline.com/action/journalInformation?journalCode=ubes20

Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

Check for updates

Regression Discontinuity Designs With Sample Selection

Yingying Dong
Department of Economics, 3151 Social Science Plaza, University of California Irvine, Irvine, CA 92697-5100
(yyd@uci.edu, http://yingyingdong.com/)

This article extends the standard regression discontinuity (RD) design to allow for sample selection or missing outcomes. We deal with both treatment endogeneity and sample selection. Identification in this article does not require any exclusion restrictions in the selection equation, nor does it require specifying any selection mechanism. The results can therefore be applied broadly, regardless of how sample selection is incurred. Identification instead relies on smoothness conditions. Smoothness conditions are empirically plausible, have readily testable implications, and are typically assumed even in the standard RD design. We first provide identification of the "extensive margin" and "intensive margin" effects. Then based on these identification results and principle stratification, sharp bounds are constructed for the treatment effects among the group of individuals that may be of particular policy interest, that is, those always participating compliers. These results are applied to evaluate the impacts of academic probation on college completion and final GPAs. Our analysis reveals striking gender differences at the extensive versus the intensive margin in response to this negative signal on performance.

KEY WORDS: Extensive margin; Fuzzy design; Gender differences; Intensive margin; Missing outcomes; Performance standard; Regression discontinuity; Sample selection.

1. INTRODUCTION

One of the frequently encountered issues in empirical applications of regression discontinuity (RD) designs is the issue of sample selection or missing outcomes. Intuitively, identification in the standard RD design relies on comparability of observations right above and right below the RD threshold (Hahn, Todd, and van der Klaauw 2001; see also the discussion in Dong 2016). Differential sample selection or missing outcomes near the RD threshold may undermine such comparability and hence the standard RD design is not valid. Recent empirical studies highlighting this issue include Martorell and McFarlin (2011); McCrary and Royer (2011); Kim (2012) among others.

McCrary and Royer (2011) estimated the impacts of female education on fertility and infant health, using an RD design based on the age-at-school-entry policy. Infant health is observed only for those women who give birth, a selected sample where ample selection (the fertility decision) itself may depend on women's education. The selection bias is corrected by controlling for the inverse Mills ratio (Heckman 1979). No exclusion restriction is present in the selection equation. This approach therefore relies entirely on the distributional assumption requiring the error term in the sample selection equation and that in the outcome equation to follow a joint normal distribution. See also Martorell and McFarlin (2011) for a similar approach in their RD design, where sample selection arises because earnings are not observed for those who do not work.

In addition, Kim (2012) estimated the effects of taking remedial courses on students' performance in the subsequent main courses. Only those who take and complete the subsequent courses have available their performance measures. Following a similar approach to Lee (2009), Kim (2012) provided bounds on the treatment effects in his sharp RD design.

Various parametric, semi-parametric, or nonparametric estimators exist for sample selection models with or without endogeneity. See, for example, Heckman (1979, 1990), Ahn and Powell (1993), Andrews and Schafgans (1998), Das, Newey, and Vella (2003), and Lewbel (2007). See also Vella (1998) for a survey on estimation of sample selection models. Existing sample selection corrections typically require exclusion restrictions when not making functional form or distributional assumptions. They may not work well in the above empirical applications of RD designs due to the absence of plausible exclusion restrictions.

This article extends the standard RD design to allow for differential sample selection or missing outcomes above or below the RD cutoff. We focus on fuzzy designs, with sharp designs following as a special case. We deal with both treatment endogeneity and sample selection. To our best knowledge, so far there do not exist any studies that provide formal identification of treatment effects in RD designs when sample selection results in incomparability of observations near the RD threshold.

This article first provides point identification of the extensive and intensive margin effects on the observed outcome distribution. Then based on these point identification results, bounds are established on subgroup treatment effects. Identification

© 2017 American Statistical Association
Journal of Business & Economic Statistics
XXXX 2017, Vol. 0, No. 0
DOI: 10.1080/07350015.2017.1302880
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

here does not require any exclusion restrictions in the selection equation. The key assumptions are similar to those employed in the standard RD design. Identification here also does not require specifying any selection mechanism. Sample selection can result from nonparticipation (e.g., dropout or unemployment), survey nonresponse, or other reasons (e.g, censoring by death).

With nonnegative outcomes such as wage or health care utilization, the observed outcome for non-participants (those who do not work or do not use health care) is zero. In contrast, when the outcome is test score or other performance measure, the outcome for nonparticipants is truly not observed. Average treatment effects (ATEs) or local average treatment effects (LATEs) in general are not identified in the first place. We explicitly consider the latter case where outcomes are missing nonrandomly, but all our results apply to both cases.

Except for the standard RD literature, a few other studies are related.[1] Frandsen (2015) provided identification of treatment effects in a general model where the outcome is censored. Frandsen assumed random censoring, which we do not assume here. Staub (2014) proposed a framework to decompose the ATE for nonnegative outcomes, assuming that the LATE is already identified. In contrast, here ATEs or LATEs are not point identified, since we do not observe outcomes for nonparticipants, for example, test scores for dropouts. Staub also discussed bounds on subpopulation-specific ATEs by restricting the sign of the treatment effects, while we do not impose any sign restrictions.[2] In addition, Chen and Flores (2014) provided bounds on treatment effects in randomized experiments when both sample selection and noncompliance are present. Unlike their bounds, we provide sharp bounds.

We apply our identification results to evaluate the impacts of academic probation on college completion and final GPAs, using confidential data from a large Texas university. The proposed approach yields empirical evidence that is different from that by the standard RD design. We show striking gender differences in response to this negative signal on performance. Women are significantly more likely to drop out when placed on probation. In contrast, probation has little impacts on men's dropout probability. Men seem to cope with this negative signal by temporarily improving their GPAs to avoid being suspended.

The rest of the article proceeds as follows. Section 2 provides identification of the extensive and intensive margin effects. Section 3 provides sharp bounds on the treatment effect for the always participating compliers. Also discussed is identifying characteristics of subgroups of compliers. Section 4 presents the empirical application. Section 5 concludes. The main text focuses on bounds on average treatment effects. Proofs and additional bounds on the corresponding quantile treatment effects are provided in the Appendices.

2. IDENTIFICATION OF THE EXTENSIVE AND INTENSIVE MARGIN EFFECTS

Let T be a binary treatment, so $T = 1$ when one is treated and 0 otherwise. Let R be the so-called running or forcing variable that determines the assignment of the treatment. At a known threshold $R = r_0$, the treatment probability has a discrete change. Let Y^* be the outcome of interest, which is observed only for a non-randomly selected sample. Further let Y be the observed outcome and S be a binary sample selection indicator, so $Y = Y^*$ if $S = 1$, and Y is missing if $S = 0$. For example, T can be an indicator for placement on academic probation, and R can be the grade point average (GPA) used to determine placement on academic probation. Y^* can then be later performance, which is observed only for students who do not drop out, so S is an indicator for enrolling in school.

Given data on Y, S, T, and R, as a first step we are interested in identifying the treatment effect on the sample selection probability, the extensive margin effect. We are also interested in the intensive margin effect, that is, the treatment effect on the observed outcome conditional on being selected into the sample. Here, we take advantage of the RD design to address both treatment endogeneity and sample selection, so both the extensive and intensive margin effects are only identified locally at the RD cutoff $R = r_0$ among the so-called compliers.

Let Y_t^* for $t = 1, 0$, be an individual's potential outcome under treatment or no treatment, and $Y^* = Y_1^* T + Y_0^*(1 - T)$. Similarly define S_t for $t = 1, 0$ as the potential sample selection under treatment or no treatment.[3] The observed selection status is then $S = S_1 T + S_0(1 - T)$. Identification in this article does not require knowing the selection mechanism, so no selection model or DGP for S is specified.

Let r be a value R can take on. All the following discussion applies to $r \in (r_0 - \varepsilon, r_0 + \varepsilon)$ for some small $\varepsilon > 0$. Let $Z = 1(R \geq r_0)$, where $1(\cdot)$ is an indicator function equal to 1 if the expression in the bracket is true and 0 otherwise. Given $R = r$, define $T_z(r)$, $z = 1, 0$, as an individual's potential treatment status above or below the RD cutoff. For example, for an individual with the observed running variable $r > 0$, $T_1(r)$ is her observed treatment, while $T_0(r)$ is her counterfactual treatment if she were below the cutoff.[4] We can then define four types of individuals in a common probability space (Ω, F, P) (Angrist, Imbens, and Rubin 1996): always taker is the event $T_1(r) = T_0(r) = 1$; never taker is the event $T_1(r) = T_0(r) = 0$; complier is the event $T_1(r) - T_0(r) = 1$, and defier is the event $T_1(r) - T_0(r) = -1$. For notational convenience, we simply use

[1] Identification of the standard RD design has been discussed in Hahn, Todd, and van der Klaauw (2001), Lee (2008), and Dong (2016). Inference was discussed by Porter (2003), Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014), Cattaneo, Frandsen, and Titiunik (2015), Otsu, Xu, and Matsushita (2015), and Feir, Lemieux, and Marmer (2016). See Cattaneo, Titiuni, and Vazquez-Bare (2016) for a comparison of different inference approaches for the standard RD design.

[2] In particular, Staub (2014) discussed bounds under two alternative assumptions. The first assumption assumes that treatment effects are nonnegative for everyone. The second assumption assume that treatment effects are nonnegative for switchers and have the same sign for always participants, and further that one knows that ATE > 0 or ATE < 0.

[3] $Y_t^* \equiv Y^*(t, S_t)$ for $t = 0, 1$.

[4] Assume $T = h(R, V)$ for unobservables V, which can be a vector. Without loss of generality, one can write $T = h_1(R, V)Z + h_0(R, V)(1 - Z)$. The function $h_z(R, V)$ for $z = 0, 1$ describes the treatment assignment below or above the cutoff. Define then $T_z(r) \equiv h_z(r, V)$ for $z = 0, 1$.

T_1 and T_0 to denote $T_1(r)$ and $T_0(r)$, respectively. Note that, however, just as potential outcomes can depend on the running variable, individual types can implicitly be functions of the running variable.

Formally define the extensive margin effect as $E[S_1 - S_0|R = r_0, C]$ and the intensive margin effect as $E[Y_1^*|S_1 = 1, R = r_0, C] - E[Y_0^*|S_0 = 1, R = r_0, C]$. The extensive margin effect captures how the participation probability differs under treatment or no treatment, while the intensive margin effect captures how the observed outcomeis expected to differ in these two counterfactual states of treatment.

Unlike the extensive margin effect, the intensive margin effect in general does not represent a causal effect at the individual level. For example, the intensive margin effect typically is different from the treatment effect for the always participating individuals, since participation is likely to change with treatment. Instead, one may view the intensive margin effect as a causal parameter from the distributional point of view. This is similar to the distributional effects frequently estimated in the program evaluation literature. The distributional effects of a social program or treatment, represented by quantile treatment effects(QTEs), generally do not capture individual causal effects unless rank invariance or rank preservation holds (see, e.g., the discussion in Heckman, Smith, and Clements 1997 and the nonparametric tests for this assumption in Dong and Shen 2016). However, if what policy makers care about is how the outcome distribution changes with the treatment, then the QTE or similarly the intensive margin effect is the treatment effect of policy interest. In our empirical application, the outcome of primary interest is the final GPA in college. The extensive margin effect measures the impact of academic probation on the probability of completing college, while the intensive margin effect measures how academic probation affects the GPA (measuring quality or training) of college graduates, regardless of the composition change.

Let $F_{\cdot|\cdot}(\cdot|\cdot)$ or $F_{\cdot|\cdot}(\cdot)$ denote the conditional distribution function throughout the article.

Assumption 1. The following assumptions hold jointly with probability 1 for $r \in (r_0 - \varepsilon, r_0 + \varepsilon)$.

A1. (Discontinuity): $\lim_{r \downarrow r_0} E[T|R = r] \neq \lim_{r \uparrow r_0} E[T|R = r]$.
A2. (Monotonicity): $\Pr(D) = 0$.
A3. (Smoothness): $F_{Y_t^*, S_t|R,\Theta}(y, s|r)$ for $s, t \in \{0, 1\}$ and $\Theta \in \{A, N, C\}$ are continuous at r_0. $\Pr(\Theta|R = r)$ for $\Theta \in \{A, N, C\}$ is continuous at r_0. The density of R is continuous and strictly positive at r_0.

A1 and A2 are the standard RD identifying assumptions (Hahn, Todd, and van der Klaauw 2001). A1 requires a positive fraction of compliers at the RD threshold. A2 is a monotonicity assumption ruling out defiers. A2 can be weakened by the assumption that conditional on the values of potential outcomes, there are more compliers than defiers (de Chaisemartin 2014).

A3 requires that the conditional joint distribution of potential outcomes and potential sample selection conditional on the running variable is continuous.[5] The observed sample selection $S = S_0 + T(S_1 - S_0)$ is allowed to change at the RD cutoff. In contrast, in the standard RD design, only the conditional distribution of potential outcomes, $F_{Y_t^*|R,\Theta}(y|r)$ for $\Theta \in \{A, N, C\}$, is required to be continuous (see, e.g., Hahn, Todd, and van der Klaauw 2001; Dong 2016).

The smoothness conditions in A3 are imposed on the full sample of observations with or without missing outcomes. The standard RD argument applies that covariates are not needed for consistency in estimating unconditional treatment effects, though they can be useful for improving efficiency or for testing validity of the RD design. A3 is plausible given no precise manipulation of the running variable and hence no sorting—the typical argument for the standard RD design identification (Lee 2008).

A3 has readily testable implications. One can follow the standard RD validity tests to test smoothness of the density of the running variable (McCrary 2008; Cattaneo, Jansson, and Ma 2016) and smoothness of the conditional means of predetermined covariates at the RD cutoff.

Theorem 1. Let $g(\cdot)$ be any measurable real function such that $E|g(\cdot)| < \infty$. If Assumption 1 holds, then for $t = 0, 1$,

$$E\left[g\left(Y_t^*\right)|S_t = 1, R = r_0, C\right] = \frac{\lim_{r \downarrow r_0} E\left[1\left(T = t\right) g\left(Y^*\right) S|R = r\right] - \lim_{r \uparrow r_0} E\left[1\left(T = t\right) g\left(Y^*\right) S|R = r\right]}{\lim_{r \downarrow r_0} E\left[1\left(T = t\right) S|R = r\right] - \lim_{r \uparrow r_0} E\left[1\left(T = t\right) S|R = r\right]}, \tag{1}$$

and

$$E\left[S_t|R = r_0, C\right] = \frac{\lim_{r \downarrow r_0} E\left[1\left(T = t\right) S|R = r\right] - \lim_{r \uparrow r_0} E\left[1\left(T = t\right) S|R = r\right]}{\lim_{r \downarrow r_0} E\left[1\left(T = t\right)|R = r\right] - \lim_{r \uparrow r_0} E\left[1\left(T = t\right)|R = r\right]}. \tag{2}$$

Note that $g(Y^*)S$ in the above is observed and is equal to $g(Y)$ if $S = 1$, and 0 if $S = 0$. When $g(Y^*) = 1(Y^* \leq y)$ for y in $\mathbb{R}$ from the distribution of (Y, S, T, R), Equation (1) identifies $F_{Y_t^*|S_t=1,R=r_0,C}(y)$ for $t = 0, 1$, the counterfactual distribution of observed outcomes under treatment or no treatment. When $g(Y^*) = Y^*$, Equation (1) identifies $E[Y_t^*|S_t = 1, R = r_0, C]$ and hence the intensive margin $E[Y_1^*|S_1 = 1, R = r_0, C] - E[Y_0^*|S_0 = 1, R = r_0, C]$. In addition, given Equation (2), the extensive margin can be simplified to the standard fuzzy RD estimand,

$$E\left[S_1 - S_0|R = r_0, C\right] = \frac{\lim_{r \downarrow r_0} E\left[S|R = r\right] - \lim_{r \uparrow r_0} E\left[S|R = r\right]}{\lim_{r \downarrow r_0} E\left[T|R = r\right] - \lim_{r \uparrow r_0} E\left[T|R = r\right]}.$$

[5] Alternatively, one could assume that $F_{Y_t^*, S_t, \Theta|R}(y, s|r)$ for any $\Theta \in \{A, N, C\}$ is continuous at r_0.

If the probability of sample selection is smooth at the RD threshold, that is, $\lim_{r\downarrow r_0} E[S|R=r] = \lim_{r\uparrow r_0} E[S|R=r]$, then the extensive margin effect $E[S_1 - S_0|R=r_0, C] = 0$, and further for $t = 0, 1$,

$$E\left[g\left(Y_t^*\right)|S_t=1, R=r_0, C\right] = \frac{\lim_{r\downarrow r_0} E\left[g\left(Y^*\right)1\left(T=t\right)|R=r, S=1\right] - \lim_{r\uparrow r_0} E\left[g\left(Y^*\right)1\left(T=t\right)|R=r, S=1\right]}{\lim_{r\downarrow r_0} E\left[1\left(T=t\right)|R=r, S=1\right] - \lim_{r\uparrow r_0} E\left[1\left(T=t\right)|R=r, S=1\right]}.$$

Applying $1(T=1) = 1 - 1(T=0)$ yields

$$E\left[g\left(Y_1^*\right)|S_1=1, R=r_0, C\right] - E\left[g\left(Y_0^*\right)|S_0=1, R=r_0, C\right] = \frac{\lim_{r\downarrow r_0} E\left[g(Y)|R=r, S=1\right] - \lim_{r\uparrow r_0} E\left[g(Y)|R=r, S=1\right]}{\lim_{r\downarrow r_0} E\left[T|R=r, S=1\right] - \lim_{r\uparrow r_0} E\left[T|R=r, S=1\right]}. \quad (3)$$

That is, the intensive margin effect can be identified by the standard RD estimand using only the selected sample in this case.

Note that, however, even if $\lim_{r\downarrow r_0} E[S|R=r] = \lim_{r\uparrow r_0} E[S|R=r]$, Equation (3) in general does not identify $E[g(Y_1^*)|S=1, R=r_0, C] - E[g(Y_0^*)|S=1, R=r_0, C]$, a causal effect for the selected sample. If further $\lim_{r\downarrow r_0} E[g(Y_t^*)|S=1, R=r, C] = \lim_{r\downarrow r_0} E[g(Y_t^*)|S=1, R=r, C]$ for $t=0,1$, that is, $E[g(Y_t^*)|S=1, R=r, C]$ is continuous at $r=r_0$, then $E[g(Y_t^*)|S_t=1, R=r_0, C] = E[g(Y_t^*)|S=1, R=r_0, C]$, and hence Equation (3) would identify $E[g(Y_1^*)|S=1, R=r_0, C] - E[g(Y_0^*)|S=1, R=r_0, C]$. In particular,

$$\begin{aligned} E\left[g\left(Y_1^*\right)|S_1=1, R=r_0, C\right] &= \lim_{r\downarrow r_0} E\left[g\left(Y_1^*\right)|S_1=1, R=r, C\right] \\ &= \lim_{r\downarrow r_0} E\left[g\left(Y_1^*\right)|S=1, R=r, C\right] \\ &= E\left[g\left(Y_1^*\right)|S=1, R=r_0, C\right], \end{aligned}$$

where the first equality follows from Assumption A3, the second quality follows from the fact that $T=1$ for C when $r > r_0$, and $S = S_1$ when $T=1$, while the last quality follows from continuity of $E[g(Y_1^*)|S=1, R=r, C]$. One can similarly show $E[g(Y_0^*)|S_0=1, R=r_0, C] = E[g(Y_0^*)|S=1, R=r_0, C]$, given continuity of $E[g(Y_0^*)|S=1, R=r, C]$.[6]

To estimate the extensive and intensive margin effects, the standard RD estimation can be applied, since both parameters involve strictly conditional means at a boundary point. Let $g(Y^*) = Y^*$, local linear or polynomial regressions can be used to consistently estimate the four discontinuities in Equations (1) and (2). Bandwidth choices can follow the plug-in approaches of Imbens and Kalyanaraman (2012) or Calonico, Cattaneo and Titiunik (2014, CCT hereafter).

Alternatively, one can apply the standard fuzzy RD estimator to estimate the extensive margin effect $E[S_1 - S_0|R=r_0, C]$.[7] One can also apply the standard fuzzy RD estimator to estimate $E[Y_t^*|S_t=1, R=r_0, C]$ for $t=0,1$ and hence the difference or the intensive margin effect, using $1(T=t)Y^*S$ as the outcome and $1(T=t)S$ as the treatment. Standard errors can be obtained by bootstrap.

3. BOUNDS ON SUBGROUP TREATMENT EFFECTS

The previous section shows identification of the extensive and intensive margin effects. Sample composition may change with the treatment status, so those with $S_1 = 1$ are not necessarily the same individuals as those with $S_0 = 1$. For example, the subpopulation with $S_1 = 1$ would involve new participants if treatment increases participation, or would not include quitters if treatment reduces participation. This section further discusses identification of subgroup treatment effects.

The analysis extends the discussion in Angrist (2001). Angrist noted that in the case of nonnegative outcomes with a nontrivial fraction of zeros (e.g., wages or health care utilization), the conditional-on-positives (COP) effect does not measure the true causal impact of any treatment on participating individuals.

Following principle stratification (Frangakis and Rubin 2002), one can classify individuals into four subgroups based on their joint distribution of potential sample selection status: new participants ($S_0=0, S_1=1$), quitters ($S_0=1, S_1=0$), never participants ($S_0=S_1=0$) and always participants ($S_0=1, S_1=1$). Further note that the RD design only identifies treatment effects locally among compliers, so these four types are defined among compliers. That is, we essentially define principle strata based on the joint distribution of potential sample selection and potential treatment.

Nonparametrically, one cannot achieve point identification of the treatment effect for each subgroup of compliers. However, one may construct sharp bounds on the treatment effect of those always participating compliers, that is, $E[Y_1^* - Y_0^*|S_0=1, S_1=1, R=r_0, C]$. The treatment effect for this group measures the true causal effect of the treatment that is not due to changes in participation (Lee 2009). In the case of academic probation, this parameter measures the causal effect of academic probation among a stable group of students who would stay in college regardless of whether they are on probation or not. We focus on deriving bounds on average treatment effects. Bounds on the corresponding quantile treatment effects are provided in Online Supplemental Appendix I.[8]

Define $p_t \equiv E[S_t|R=r_0, C]$, $t=0,1$, which is identified by Equation (2) of Theorem 1. Further define $p_{jk} \equiv \Pr(S_0=j, S_1=k|, R=r_0, C)$, $j,k=0,1$. The identified distributions in Theorem 1 can be decomposed as

[6]That is, smoothness conditions need to hold for the selected sample in order for Equation (3) to identify a causal effect for the selected sample.

[7]In practice, the fuzzy RD estimator along with its robust bias-corrected inference can be conveniently implemented using the Stata command rdrobust.ado (*https://sites.google.com/site/rdpackages/rdrobust*).

[8]Zhang and Rubin (2003) and Imai (2008) discussed similar bounds in the context of randomized experiments with perfect compliance. See also Lee (2009), Blanco, Flores, and Flores-Lagunes (2013), and Chen and Flores (2014) for construction of bounds in evaluating the effects of Job Corps.

follows:

$$F_{Y_1^*|S_1=1,R=r_0,C}(y) = F_{Y_1^*|S_0=1,S_1=1,R=r_0,C}(y)\frac{p_{11}}{p_1} + F_{Y_1^*|S_0=0,S_1=1,R=r_0,C}(y)\frac{p_{01}}{p_1},$$

and

$$F_{Y_0^*|S_0=1,R=r_0,C}(y) = F_{Y_0^*|S_0=1,S_1=1,R=r_0,C}(y)\frac{p_{11}}{p_0} + F_{Y_0^*|S_0=1,S_1=0,R=r_0,C}(y)\frac{p_{10}}{p_0}.$$

The above expressions involve fractions of three types of individuals, p_{11}, p_{01}, and p_{10}. Without further assumptions, these fractions are not point identified. However, it is easy to show that assuming $p_{11} > 0$,

$$p_{11} \in \mathcal{P} \equiv (0, 1] \cap [p_0 + p_1 - 1, \min\{p_0, p_1\}].$$

General bounds for $E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C]$ can then be constructed by following the approach of Horowitz and Manski (1995). Define $Q_t(\tau) \equiv F^{-1}_{Y_t^*|S_t=1,R=r_0,C}(\tau)$ for $\tau \in (0, 1)$ and $t = 0, 1$. For simplicity, let $-\infty = \inf\{y{:}y \in \mathcal{Y}\}$ and $+\infty = \sup\{y{:}y \in \mathcal{Y}\}$, where $\mathcal{Y} \subseteq \mathbb{R}$ is the support of $Y_t^*|S_t = 1, r = r_0, C$, for $t = 0, 1$. In the worst-case (best-case) scenario, the smallest (largest) p_{11}/p_1 values of Y_1 in the conditional distribution $Y_1^*|S_1 = 1, R = r_0, C$ and the largest (smallest) p_{11}/p_0 values of Y_0 of in $Y_0^*|S_0 = 1, R = r_0, C$ belong to always participants. That is, $L \le E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C] \le U$, where

$$L \equiv \min_{p_{11}\in\mathcal{P}}\left(\frac{p_1}{p_{11}}\int_{-\infty}^{Q_1(p_{11}/p_1)} y dF_{Y_1^*|S_1=1,R=r_0,C}(y) - \frac{p_0}{p_{11}}\int_{Q_0(1-p_{11}/p_0)}^{+\infty} y dF_{Y_0^*|S_0=1,R=r_0,C}(y)\right),$$

and

$$U \equiv \max_{p_{11}\in\mathcal{P}}\left(\frac{p_1}{p_{11}}\int_{Q_1(1-p_{11}/p_1)}^{+\infty} y dF_{Y_1^*|S_1=1,R=r_0,C}(y) - \frac{p_0}{p_{11}}\int_{-\infty}^{Q_0(p_{11}/p_0)} y dF_{Y_0^*|S_0=1,R=r_0,C}(y)\right).$$

These bounds are typically too wide to be informative in practice. In the following, we consider two commonly employed assumptions to tighten the bounds for $E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C]$.

3.1 Bounds Under Monotonic Selection

Given $p_t \equiv E[S_t|R = r_0, C]$, $t = 0, 1$, ruling out one type of compliers allows one to identify the fractions of the remaining two types. We therefore impose the following monotonicity assumption.

Assumption 2 (Monotonic Selection): $\Pr(S_0 \ge S_1) = 1$.

Assumption 2 requires that treatment can only affect sample selection in "one direction," in particular, everyone is less likely to participate under treatment. Derivation for $S_1 \ge S_0$ is symmetric to that for $S_0 \ge S_1$, so for now we focus on $S_0 \ge S_1$. In our empirical scenario, this assumption assumes that academic probation induces individuals to quit rather than to participate, which is plausible. Existing studies have shown that probation increases the probability of dropout (Lindo, Sanders, and Oreopoulos 2010). Monotonic selection is frequently used in constructing bounds in similar settings (see, e.g. Zhang and Rubin 2003; Imai 2008; Lee 2009; Blanco, Flores, and Flores-Lagunes 2013; Chen and Flores 2014, in the context of randomized experiments). Such a monotonicity assumption is consistent with a latent index sample selection model with an additively separable latent error (Heckman 1979, 1990; Vytlacil 2002). Following similar arguments to those in de Chaisemartin (2014), one can alternatively assume that conditional on potential outcomes, there are more quitting than newly participating compliers.

Under Assumption 2, the subpopulation with $S_1 = 1$ consists of only always participants, that is, those having $S_0 = 1$ and $S_1 = 1$, while the subpopulation with $S_0 = 1$ consists of always participants ($S_0 = 1$, $S_1 = 1$) and quitters ($S_0 = 1$, $S_1 = 0$). Let $q = p_{10}/p_0$ denote the fraction of quitters among the subpopulation with $S_0 = 1$.

$$F_{Y_1^*|S_1=1,R=r_0,C}(y) = E\left[1\left(Y_1^* \le y\right)|S_1 = 1, S_0 = 1, R = r_0, C\right], \tag{4}$$

and

$$F_{Y_0^*|S_0=1,R=r_0,C}(y) = E\left[1\left(Y_0^* \le y\right)|S_1 = 1, S_0 = 1, R = r_0, C\right](1-q) + E\left[1\left(Y_0^* \le y\right)|S_1 = 0, S_0 = 1, R = r_0, C\right]q. \tag{5}$$

The worst-case (best-case) scenario is that the largest (smallest) $1 - q$ observations in the conditional distribution $Y_0^*|S_0 = 1, R = r_0, C$ belong to always participants and the smallest (largest) q observations belong to quitters. It follows that

$$\begin{aligned} &E\left[Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C\right] \\ &\quad \le E\left[Y_0^*|S_0 = 1, Y_0^* \geqslant Q_0(q), R = r_0, C\right], \text{ and} \\ &E\left[Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C\right] \\ &\quad \ge E\left[Y_0^*|S_0 = 1, Y_0^* \le Q_0(1-q), R = r_0, C\right]. \end{aligned} \tag{6}$$

The quantiles $Q_0(\tau)$ for $\tau = 1 - q$, q can be obtained from the identified conditional distribution $F_{Y_0^*|S_0=1,R=r_0,C}(y)$ by Theorem 1, once one knows q. In particular, $Q_0(\tau) = \inf\{y{:}F_{Y_0^*|S_0=1,R=r_0,C}(y) \ge \tau\}$. The following Lemma 1 provides identification of q.

Lemma 1. 1 If Assumptions 1 and 2 hold, then

$$q = \frac{\lim_{r\downarrow r_0} E[S|R = r] - \lim_{r\uparrow r_0} E[S|R = r]}{\lim_{r\downarrow r_0} E[S(1-T)|R = r] - \lim_{r\uparrow r_0} E[S(1-T)|R = r]}. \tag{7}$$

Further by the inequalities in (6), we obtain the following bounds.

Theorem 2. If Assumptions 1 and 2 hold, then $L^m \le E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C] \le U^m$, where

$$L^m \equiv E\left[Y_1^*|S_1 = 1, R = r_0, C\right] - \frac{1}{1-q}\int_{Q_0(q)}^{+\infty} y dF_{Y_0^*|S_0=1,R=r_0,C}(y)$$

$$= E\left[Y_1^*|S_1 = 1, R = r_0, C\right] - \frac{1}{1-q} E\left[1\left(Y_0^* \geq Q_0(q)\right) Y_0^* | S_0 = 1, R = r_0, C\right], \text{ and}$$

$$U^m \equiv E\left[Y_1^*|S_1 = 1, R = r_0, C\right] - \frac{1}{1-q}\int_{-\infty}^{Q_0(1-q)} y dF_{Y_0^*|S_0=1,R=r_0,C}(y)$$

$$= E\left[Y_1^*|S_1 = 1, R = r_0, C\right] - \frac{1}{1-q} E\left[1\left(Y_0^* \leq Q_0(1-q)\right) Y_0^* | S_0 = 1, R = r_0, C\right],$$

All the terms are identified by Theorem 1 and Lemma 1.

Conditional means in the first terms of the lower and upper bounds can be identified by Equation (1) of Theorem 1, setting $g(Y^*) = Y^*$ for $t = 1$, while those in the second terms can be identified by setting $g(Y^*) = 1(Y^* \geq Q_0(q))Y^*$ or $g(Y^*) = 1(Y^* \leq Q_0(1-q))Y^*$ for $t = 0$.

The above bounds fall in the class of "worst-case" bounds by Horowitz and Manski (1995) and hence are sharp by their Proposition 4. That is, L^m (U^m) is the largest (smallest) lower (upper) bound for $E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C]$ that is consistent with the observed data. Neither exclusion restriction nor bounded support of the outcome is required for these bounds. In contrast, the bounds proposed by Horowitz and Manski (2000) require that the support of the outcome is bounded so one can impute the missing data with either the largest or the smallest possible values.

Theorem 2 provides bounds on the treatment effect for the always participating compliers. The quitting compliers participate only under no treatment. Without making any assumptions on their counterfactual outcomes under treatment, bounds can be constructed only for their potential outcome under no treatment $E[Y_0^*|S_1 = 0, S_0 = 1, R = r_0, C]$. The upper bound is $\frac{1}{q}E[1(Y_0^* \geq Q_0(1-q))Y_0^*|S_0 = 1, R = r_0, C]$, and the lower bound is $\frac{1}{q}E[1(Y_0^* \leq Q_0(q))Y_0^*|S_0 = 1, R = r_0, C]$.

In addition, Theorem 2 assumes that $S_0 \geq S_1$ holds almost surely. If instead $S_1 \geq S_0$ holds almost surely, then $L^{m'} \leq E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C] \leq U^{m'}$, where

$$L^{m'} \equiv \frac{1}{1-q} E\left[1\left(Y_1^* \leq Q_1(1-q)\right) Y_1^* | S_1 = 1, R = r_0, C\right] - E\left[Y_0^*|S_0 = 1, R = r_0, C\right], \text{ and}$$

$$U^{m'} \equiv \frac{1}{1-q} E\left[1\left(Y_1^* \geq Q_1(q)\right) Y_1^* | S_1 = 1, R = r_0, C\right] - E\left[Y_0^*|S_0 = 1, R = r_0, C\right].$$

The bounds in Theorem 2 can be conveniently estimated by the following steps.

Step 1: Estimate $E[Y_1|S_1 = 1, R = r_0, C]$ by the standard fuzzy RD estimator, using Y^*ST as the outcome and ST as the treatment.

Step 2: Estimate q by the standard fuzzy RD estimator, using S as the outcome and $S(1-T)$ as the treatment. Denote the estimate as $\hat{q}$.

Step 3: Estimate $F_{Y_0^*|S_0=1,R=r_0,C}(y)$ by the standard fuzzy RD estimator, using $1(Y^* \leq y)S(1-T)$ as the outcome and $S(1-T)$ as the treatment. Then invert the estimated distribution to get the quantiles $\hat{Q}_0(\hat{q})$ and $\hat{Q}_0(1-\hat{q})$.[9]

Step 4: Estimate $E[1(Y_0^* \leq Q_0(1-q))Y_0^*|S_0 = 1, R = r_0, C]$ and $E[1(Y_0^* \geq Q_0(q))Y_0^*|S_0 = 1, R = r_0, C]$ by the standard fuzzy RD estimators, using $S(1-T)$ as the treatment and $1(Y^* \leq \hat{Q}_0(1-\hat{q}))Y^*S(1-T)$ and $1(Y^* \geq \hat{Q}_0(\hat{q}))Y^*S(1-T)$, respectively, as the outcomes.

Step 5: Construct bounds by replacing each term involved in Theorem 2 with their estimates from Steps 1 to 4.

By construction, the lower and upper bounds are ordered, that is, $L^m \leq U^m$, so confidence intervals for the true parameter can be constructed following Imbens and Manski (2004), using bootstrapped standard errors. Such confidence intervals are valid by Lemma 3 of Stoye (2009), and by noticing that estmators of the proposed bounds are smooth functions of asymptotically normal estimators in Steps 1 to 4 (Calonico, Cattaneo and Titiunik 2014; Frandsen, Frölich, and Melly 2012). If desired, one can also construct confidence intervals for the entire identification region, for example, by bootstrap, following Horowitz and Manski (2000).

3.2 Subgroup Characteristics and Testable Implications of Monotonic Selection

Monotonic selection stated in Assumption 2 plays an important role in obtaining the sharp bounds in Theorem 2. By blocking sample selection in one direction, this assumption also permits point identification of subgroup characteristics among compliers. Identifying subgroup characteristics provides important information regarding what types of individuals are more likely to quit (or participate) when they are under treatment. For example, in our empirical application, it is of policy interest to determine what types of students would quit when placed on academic probation.

Identifying subgroup characteristics also leads to the opportunity of verifying the monotonic selection assumption. Under Assumption 2, the identified probability distribution of characteristics for the quitting compliers should be bounded between 0 and 1. Otherwise, what is identified is a weighted difference in the probability distribution of characteristics between the quitting compliers and the newly participating compliers, and hence could lie outside of the interval [0, 1].

Let X with a support $\mathcal{X} \subseteq \mathbb{R}$ be some predetermined covariate other than the running variable. Following Theorem 1, immediately we have the following corollary.

[9]In practice, these quantiles can be conveniently estimated by using the RD quantile treatment effect estimator proposed by Frandsen, Frölich, and Melly (2012), after replacing T with ST and $(1-T)$ with $S(1-T)$ to deal with sample selection.

Corollary 1. Assume that A1 and A2 hold. Assume further that A3 holds after replacing Y_t^* with X. Then for $t = 0, 1$,

$$F_{X|S_t=1,R=r_0,C}(x) = \frac{\lim_{r\downarrow r_0} E\left[1\left(X \le x\right)1\left(T = t\right)S|R = r\right] - \lim_{r\uparrow r_0} E\left[1\left(X \le x\right)1\left(T = t\right)S|R = r\right]}{\lim_{r\downarrow r_0} E\left[1\left(T = t\right)S|R = r\right] - \lim_{r\uparrow r_0} E\left[1\left(T = t\right)S|R = r\right]}.$$

Analogous to Equations (4) and (5), the above identified distributions can be decomposed as follows.

Corollary 2. Assume that A1 and A2 hold. Assume further that A3 holds after replacing Y_t^* with X. Under Assumption 2,

$$F_{X|S_1=1,S_0=1,R=r_0,C}(x) = F_{X|S_1=1,R=r_0,C}(x), \text{ and} \tag{9}$$

$$F_{X|S_1=0,S_0=1,R=r_0,C}(x) = \frac{1}{q}F_{X|S_0=1,R=r_0,C}(x) - \frac{1-q}{q}F_{X|S_1=1,R=r_0,C}(x), \tag{10}$$

where q is identified by Lemma 1, and $F_{X|S_t=1,R=r_0,C}(x), t = 0, 1$ is identified by Corollary 1.

Equation (9) identifies the distribution of covariates for the always participating compliers, while Equation (10) identifies that for the quitting compliers. Assumption 2 implies

$$1 \ge \frac{1}{q}F_{X|S_0=1,R=r_0,C}(x) - \frac{1-q}{q}F_{X|S_1=1,R=r_0,C}(x) \ge 0 \quad \text{for all } x \in \mathcal{X}. \tag{11}$$

Equation (11) along with the inequality $E[S_1 - S_0|R = r_0, C] < 0$ under Assumption 2 can be easily tested by one-sided t tests. Equation (11) therefore provides a practical way of verifying the plausibility of monotonic selection in Assumption 2.

Kitigawa (2015) proposed a similar test for the LATE assumption of Imbens and Angrist (1994). Kitigawa's (2015) test uses the fact that given monotonicity along with the other LATE assumptions, the identified probability density distributions of potential outcomes for compliers should be nonnegative. Here we test covariates. Typically binary covariates, such as gender, race, or ethnicity indicators are available. The proposed tests can then be implemented by simply testing that the identified probabilities of these binary covariates for the quitting compliers are between 0 and 1.

If instead assuming $S_1 \ge S_0$, one can analygously identify characteristics of the always participating compliers and newly participating compliers. That is,

$$F_{X|S_1=0,S_0=1,R=r_0,C}(x) = E\left[1\left(X \le x\right)|S_0 = 1, R = r_0, C\right], \text{ and}$$

$$F_{X|S_1=1,S_0=1,R=r_0,C}(x) = \frac{1}{q}E\left[1\left(X \le x\right)|S_1 = 1, R = r_0, C\right] - \frac{1-q}{q}E\left[1\left(X \le x\right)|S_0 = 1, R = r_0, C\right].$$

3.3 Bounds Under Stochastic Dominance

When monotonic sample selection is not plausible, it is necessary to rely on alternative assumptions to construct bounds. This section provides sharp bounds for $E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C]$ under a different assumption than Assumption 2.

Assumption 3. (Stochastic Dominance):
$F_{Y_1^*|S_0=1,S_1=1,R=r_0,C}(y) \le F_{Y_1^*|S_0=0,S_1=1,R=r_0,C}(y)$ and
$F_{Y_0^*|S_0=1,S_1=1,R=r_0,C}(y) \le F_{Y_0^*|S_0=1,S_1=0,R=r_0,C}(y)$ for any $y \in \mathcal{Y}$.

(8)

Assumption 3 requires that the distribution of potential outcome Y_1^* (Y_0^*) for those always participating compliers weakly stochastically dominates that of newly participating (quitting) compliers. This assumption is plausible when those who participate regardless of treatment states have better outcomes than those who are induced to participate only in one treatment state (Blanco, Flores, and Flores-Lagunes 2013; Chen and Flores 2014). Only mean dominance, $E[Y_1^*|S_0 = 1, S_1 = 1, R = r_0, C] \ge E[Y_1^*|S_0 = 0, S_1 = 1, R = r_0, C]$ and $E[Y_0^*|S_0 = 1, S_1 = 1, R = r_0, C] \ge E[Y_0^*|S_0 = 1, S_1 = 0, R = r_0, C]$, is needed to derive sharp bounds on the average treatment effect of the always participating compliers. We impose a stronger assumption to also derive sharp bounds for the corresponding quantile treatment effects (provided in Online Supplemental Appendix I).

Theorem 3. Assume that $p_0 + p_1 > 1$. If Assumptions 1 and 3 hold, then
$L^s \le E[Y_1^* - Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C] \le U^s$, where

$$L^s \equiv E\left[Y_1^*|S_1 = 1, R = r_0, C\right] - \frac{p_0}{p_0 + p_1 - 1} \times E\left[1\left(Y_0^* \ge Q_0\left(\frac{1-p_1}{p_0}\right)\right)Y_0^*|S_0 = 1, R = r_0, C\right], \text{ and}$$

$$U^s \equiv \frac{p_1}{p_0 + p_1 - 1}E\left[1\left(Y_1^* \ge Q_1\left(\frac{1-p_0}{p_1}\right)\right) \times Y_1^*|S_1 = 1, R = r_0, C\right] - E\left[Y_0^*|S_0 = 1, R = r_0, C\right].$$

all the terms are identified by Theorem 1.

p_t, $t = 0, 1$ can be identified by Equation (2) of Theorem 1. Conditional means in the first terms of the lower or upper bounds can be identified by Equation (1), setting $g(Y^*) = Y^*$ or $g(Y^*) = 1(Y^* \ge Q_1(\frac{1-p_0}{p_1}))Y^*$ for $t = 1$, while those in the second terms can also be identified by Equation (1), setting $g(Y^*) = 1(Y^* \ge Q_0(\frac{1-p_1}{p_0}))Y^*$ or $g(Y^*) = Y^*$ for $t = 0$. Estimation and construction of confidence intervals follow analogously to those discussed in Section 3.1.

Finally, note that Assumptions 2 and 3 may be changed and combined, depending on their plausibility in a particular empirical application. For example, if both Assumptions 2 and 3 hold in addition to Assumption 1, then the sharp bounds in Theorem 2 can be tightened. In particular, stochastic dominance in Assumption 3 implies that $E[Y_0^*|S_1 = 1, S_0 = 1, R = r_0, C] \ge E[Y_0^*|S_0 = 1, R = r_0, C]$, while $E[Y_0^*|S_0 = 1, R = r_0, C] \ge \frac{1}{1-q}E[1(Y_0^* \le Q_0(1-q))Y_0^*|S_0 = 1, R = r_0, C]$. The lower bound is then $L^{ms} \equiv E[Y_1^*|S_1 = 1, R = r_0, C] - \frac{1}{1-q}E[1(Y_0^* \ge Q_0(q))Y_0^*|S_0 = 1, R = r_0, C]$, and the upper bound is $U^{ms} \equiv E[Y_1^*|S_1 = 1, R = r_0, C] - E[Y_0^*|S_0 = 1, R = r_0, C]$. That is, the extensive margin is the upper bound in this case.

4. EMPIRICAL APPLICATION: ACADEMIC PROBATION AND GENDER DIFFERENCES IN RESPONSES

This section applies the proposed approach to evaluate the impacts of academic probation. Nearly all colleges and universities in the U.S. adopt academic probation to motivate students to stay above a certain performance standard. Surprisingly little empirical evidence exists on how this popular policy affects students' outcomes.

Typically students are placed on academic probation if their GPAs fall below a certain threshold. Students on probation face the real threat of being suspended if their performance continues to fall below the required standard. In a seminal study, Lindo, Sanders, and Oreopoulos (2010) examined the effects of academic probation using data from a large Canadian university. Fletcher and Tokmouline (2010) performed similar analysis using the U.S. data. Both studies adopt the standard sharp RD design to evaluate the effects of the first-year (or first-term) probation. They show that placement on academic probation discourages some students from continuing in school while motivating others to perform better. That is, academic probation simultaneously increases the dropout probability yet improves the performance of those nondropouts.

Here, we investigate the effects of being ever placed on academic probation in college. Correctly evaluating the overall effects of academic probation requires dealing with attrition that differs right above and right below the probation threshold. We also investigate what type of students are induced to drop out when placed on probation. For example, although academic probation increases college attrition, it might be welfare improving if those who drop out are low performing students who would not gain much from staying in college anyways. Identifying the characteristics of dropouts is possible given our identification results on subgroup characteristics in Section 3.2.

Let Y^* be the cumulative GPA. Let S be a sample selection indicator which is 1 if a student does not drop out and 0 otherwise. Y^* is observed only if $S = 1$, that is, a student does not drop out by the time their performance is measured. Our main analysis focuses on the final GPAs of college graduates. We additionally look at GPAs at the end of the first, second, third, and fourth academic years. The treatment T is an indicator of whether a student has ever been on probation. The running variable R is the first semester GPA. Fuzzy RD designs are entailed, since students with the first semester GPA falling just above the probation threshold may still fail and be placed on probation later. One exception occurs when the outcome under consideration is performance at the end of the first year (second semester). In this case, probation is determined solely by the first semester GPA falling below the probation threshold and hence the RD design is sharp.

The analysis draws on confidential data from a large public university in Texas. These data are collected under the Texas Higher Education Opportunity Project (THEOP).[10] An undergraduate at this university is considered to be "scholastically deficient" if his or her GPA falls below 2.0. We do not observe the actual probation status. The treatment T is set to be 1 as long as a student's cumulative GPA is below the school-wide cutoff 2.0, that is, when a student is considered to be "scholastically deficient."[11] The data represent the entire population of the first-time freshmen cohorts between 1992 and 2002. Their college transcript information is available from 1992 to 2007. We include in our sample all students for whom we have complete records. The total sample size is 64,310.

Table 1 presents the sample summary statistics for the full sample and the sample with the first semester GPA falling between 1.5 and 2.5 (referred to as the close-to-cutoff sample).[12] The sample size for final GPA is much smaller, indicating serious sample selection or attrition. Compared with students who have never been placed on probation, those ever on probation are much less likely to complete college, 44.4% lower in the full sample or 30.4% lower in the close-to-cutoff sample. Among students who compete college, those ever on probation also have lower final GPAs, 0.627 lower for the full sample and 0.243 lower for the close-to-cutoff sample. However, these simple correlations do not represent the causal impacts of academic probation, since students ever on probation are expected to be poorer performers. For example, they have lower SAT scores on average. They are also less likely to be ranked among the top 25% of their high school classes and less likely to be a member of National Honors Society (NHS). In addition, students ever on probation are more likely to be male and less likely to be White. All these differences are statistically significant at the 1% level. The same general pattern holds true for the close-to-cutoff sample, even though not surprisingly all the differences are smaller. Still all but one of the differences, the NHS membership, are statistically significant at the 1% level for the close-to-cutoff sample.

Figure 1 plots the probabilities of probation conditional on the first semester GPA for the full sample, women, and men separately.[13] For those whose first semester GPAs fall below the probation threshold, the probability of being on probation is 1 by construction. This one-sided noncompliance implies no defiers, and hence Assumption A2 holds by design. The estimated discontinuity in the probation probability at the cutoff is 59.3% for the full sample, 66.3% for women, and 53.6% for men. These estimates are statistically significant at the 1% level. Therefore, Assumption A1 holds.

Now consider Assumption A3. There are no consistent tests for A3. However, one can test its implications, smoothness of the conditional means of predetermined covariates and smoothness of the density of the running variable. Figure 2(a) shows the conditional means of some pre-determined covariates, including SAT score, indicators for male, White, whether one is ranked among the top 25% of the high school class, whether one is an NHS member in high school, and whether one is from a feeder school. Figure 2(b) presents the density of the first semester

[10] Fletcher and Tokmouline (2010) also used the THEOP data, but all the data used in this article are obtained and processed independently.

[11] In practice, when a student is considered as scholastically deficient, he or she may only be given an academic warning. However, a quick survey administered to the relevant academic deans suggests that students are generally placed on probation in this case.

[12] The close-to-cutoff sample is used to produce sample summary statistics and figures only.

[13] All our figures are conveniently generated using the Stata command, rdplot.ado. Details can be found in Calonico, Cattaneo, and Titiunik, (2015).

Table 1. Sample descriptive statistics

	Ever on probation		Never on probation		
	N	Mean (SD)	N	Mean (SD)	Difference
			I: Full sample		
Final GPA	6,447	2.535	44,492	3.162	−0.627
		(0.323)		(0.439)	(0.006)***
College completion	14,398	0.448	49,912	0.891	−0.444
		(0.497)		(0.311)	(0.003)***
Male	14,398	0.579	49,912	0.461	0.117
		(0.494)		(0.499)	(0.005)***
White	14,398	0.726	49,912	0.836	−0.110
		(0.446)		(0.370)	(0.004)***
SAT score	14,369	1,112	49,825	1,182	−69.88
		(129.8)		(135.9)	(1.274)***
Top 25% of HS class	14,398	0.689	49,912	0.832	−0.111
		(0.359)		(0.440)	(0.003)***
HS NHS member	14,369	0.265	49,912	0.350	−0.085
		(0.441)		(0.477)	(0.004)***
Feeder school	14,369	0.121	49,912	0.180	−0.059
		(0.326)		(0.384)	(0.004)***
			II: 1st semester GPA = 2.0 ± 0.5		
Final GPA	4,607	2.565	7901	2.808	−0.243
		(0.324)		(0.323)	(0.006)***
College completion	8,512	0.541	9351	0.845	−0.304
		(0.498)		(0.362)	(0.006)***
Male	8,512	0.565	9357	0.465	0.100
		(0.496)		(0.499)	(0.007)***
White	8,512	0.746	9351	0.806	−0.059
		(0.435)		(0.396)	(0.006)***
SAT score	8,497	1,111	9336	1,124	−12.43
		(127.2)		(120.4)	(1.855)***
Top 25% of HS class	8,512	0.706	9351	0.778	−0.073
		(0.456)		(0.415)	(0.007)***
HS NHS member	8,512	0.265	9351	0.273	−0.008
		(0.442)		(0.446)	(0.007)
Feeder school	8,512	0.124	9351	0.147	−0.023
		(0.330)		(0.354)	(0.005)***

GPA.[14] No noticeable differences are observed in the average values of the covariates or in the density of the running variable at the probation threshold. More formally, we perform falsification tests, that is, test the impacts of academic probation on these covariates. We also test the discontinuity in the density of the running variable at the RD cutoff (McCrary 2008; Cattaneo, Jansson, and Ma 2016). Results from these tests are reported in Table 2. None of the estimates are statistically significant, supporting the validity of the research design here.

We then estimate the extensive and intensive margin effects based on Theorem 1. Figure 3 visualizes the probability of completing college (top row) and the final mean GPA (bottom row) given the first semester GPA. Women whose first semester GPAs fall just below 2.0 are much less likely to complete college than those whose GPAs fall just above. In sharp contrast, for men the probability of completing college does not differ much just above and just below the probation threshold. Note that in the bottom row of Figure 3, any discontinuities (or lack of discontinuities) in the observed GPA at the probation threshold can

Table 2. RD validity tests

I: RD effects of Academic Probation on Covariates			
Male	0.032 (0.045)	Top 25% of HS Class	− 0.040 (0.036)
White	0.005 (0.038)	HS NHS member	− 0.006 (0.033)
SAT score	0.158 (12.87)	Feeder school	0.025 (0.025)
II: Discontinuity in the Density of Running Variable			
	0.115 (0.600)		0.047 (0.041)

Note: In Panel I, the CCT bias-corrected estimates along with robust standard errors are reported. In Panel II, the first column reports the estimated discontinuity in logarithm of the empirical density of the running variable (with a bin width 0.01); the second column reports the estimated discontinuity by the nonparametric density estimator of Cattaneo, Jansson, and Ma (2016).

[14] Students whose first semester GPAs are exactly 2.0 are not included in our sample, considering possible rounding at this value. We assume that observations away from 2.0 are correctly measured.

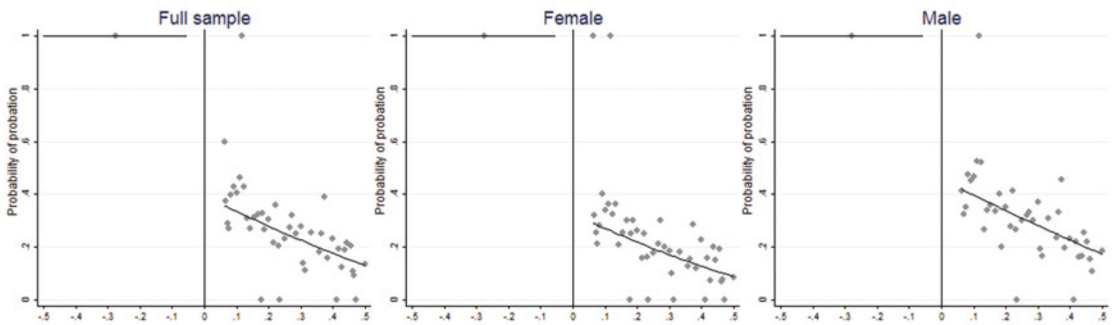

Figure 1. Probability of ever placement on probation against first semester GPA (centered at 2.0).

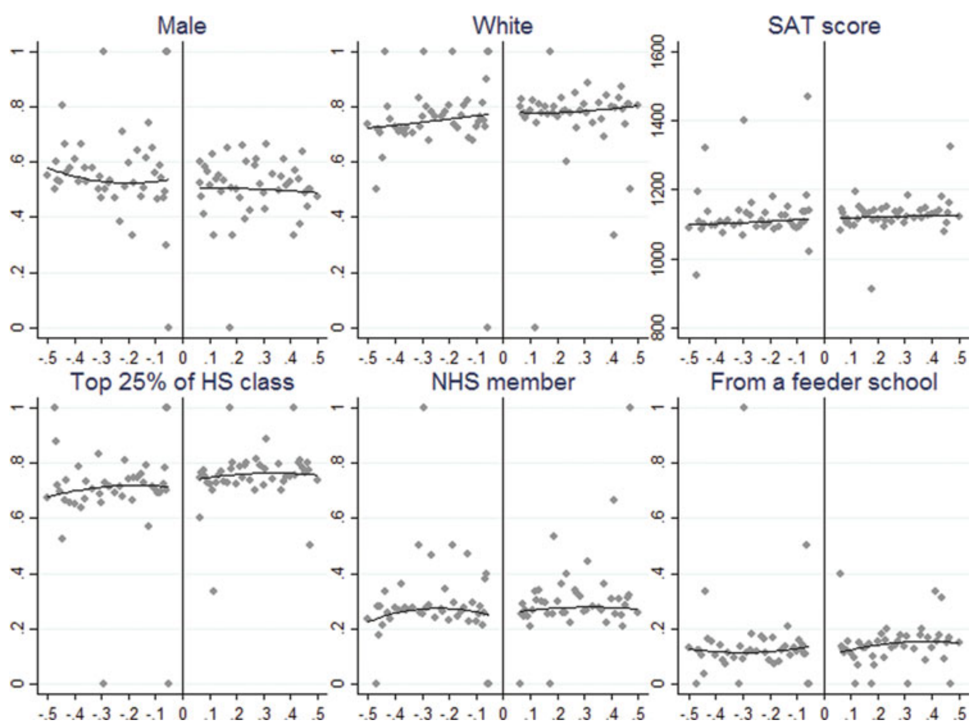

Figure 2(a). Conditional means of covariates conditional on first semester GPA.

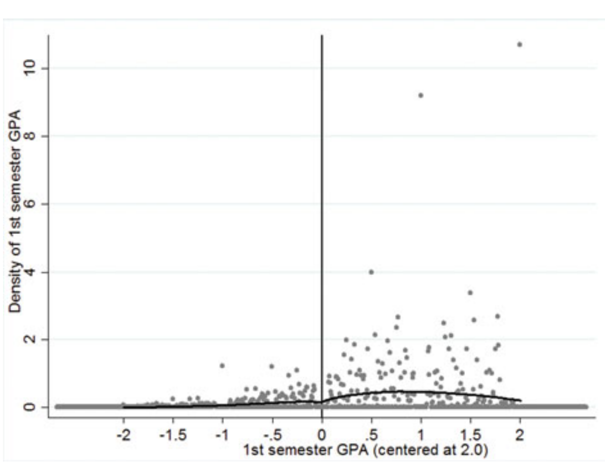

Figure 2b. Empirical density of the running variable (first semester GPA).

result from either changes in sample selection or real changes in the performance of those nondropouts.

Table 3 presents the main results.[15] The top panel of Table 3 reports the estimated extensive and intensive margin effects. For comparison purposes, the middle panel of Table 3 presents the estimated LATEs by the standard RD design. The bottom panel presents the estimated bounds on the probation effect of those always participating compliers. Discussion on these bounds is deferred until later. The probability for women to complete college is estimated to decrease by 18.2% if they have ever been placed on academic probation. This estimate is statistically significant at the 1% level. In sharp contrast, probation is estimated to have a small, positive, yet insignificant impact (5.6% with

[15]For notational convenience, in all the tables, I drop C and $R = r_0$ in the conditioning set. Nevertheless, all estimates are among the compliers at the probation threshold.

Table 3. Effects of academic probation on college completion and final GPAs

	Full sample		Female		Male	
	I: RDD with sample selection					
(1):Pr ($S_0 = 1$)	0.824	(0.018)***	0.834	(0.023)***	0.820	(0.025)***
(2):Pr ($S_1 = 1$)	0.773	(0.039)***	0.659	(0.064)***	0.875	(0.080)***
Extensive margin: (2)-(1)	– 0.051	(0.037)	– 0.182	(0.068)***	0.057	(0.090)
(3): E($Y_0 \vert S_0 = 1$)	2.727	(0.016)***	2.768	(0.020)***	2.686	(0.022)***
(4): E($Y_1 \vert S_1 = 1$)	2.771	(0.026)***	2.837	(0.039)***	2.716	(0.036)***
Intensive margin: (4)-(3)	0.045	(0.036)	0.069	(0.050)	0.030	(0.050)
	II: Standard RDD					
	0.029	(0.032)	0.049	(0.040)	0.047	(0.058)
	III: Bounds for always participating compliers					
Lower bound 1	– 0.011	(0.054)	– 0.010	(0.099)	0.030	(0.060)
Upper bound 1	0.209	(0.098)**	0.148	(0.121)	0.030	(0.102)
90% CI 1	[– 0.080	0.336]	[– 0.139	0.306]	[– 0.068	0.198]
Lower bound 2	0.045	(0.036)	0.069	(0.050)	0.030	(0.052)
Upper bound 2	0.209	(0.098)**	0.148	(0.121)	0.030	(0.102)
90% CI 2	[– 0.002	0.336]	[– 0.002	0.318]	[– 0.055	0.198]
N	64,310		32,952		31,358	

Notes: All estimates are conditional on compliers at the first semester GPA equal to 2.0; Estimation of the extensive and intensive margins, and the bounds follows the description in Sections 2 and 3.1, respectively. The CCT bias-corrected robust inference is used; in Panel III, 1 refers to the bounds under the monotonic sample selection assumption, while 2 refers to the bounds assuming additionally mean dominance, particularly $E(Y_0|S_0 = 1, S_1 = 0, C, R = r_0) \geq E(Y_0|S_0 = 1, S_1 = 1, C, R = r_0)$; bootstrapped standard errors are in the parentheses; Imbens and Manski's (2004) CIs are reported; ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

a standard error 0.09) on men's probability of completing college. The estimated effects at the intensive margin are small and insignificant for both men and women, so academic probation does not seem to promote the ultimate performance of college graduates. Note that by the standard RD design, the estimated effects of academic probation on final GPAs are all small and insignificant, hiding any significant changes at the extensive margin.

To further investigate gender differences in response to placement on probation, the top rows of Figures 4 and 5 show, respectively, the probabilities for women and men to stay in college till the end of the first, second, third, and fourth years. The bottom rows show correspondingly their cumulative GPAs. These figures reveal remarkable gender differences. In Figure 4, women who fall just below the (first-semester) probation threshold are increasingly more likely to drop out over academic years. In contrast, in Figure 5 the dropout probability for men in general does not differ much on either side of the probation threshold in all years. At the same time, the observed mean GPAs for men are always higher to the left of the threshold than those to the right. This visual evidence suggests that while women are more likely to drop out once being placed on probation, men seem to cope with this negative signal by improving their performance to avoid being suspended.

Tables 4 reports the estimated impacts on college persistence and the cumulative GPA for women. Consistent with the visual evidence in Figure 4, estimates in Table 4 show that placement on probation significantly reduces college persistence among women. Almost all women finish the first year of college, regardless of their probation status. The estimated impact on the probability of completing the first year is −1.1% and is not statistically significant. However, the probabilities of completing the second, third, and fourth years are estimated to decrease significantly by 11.7%, 16.2%, and 16.7%, respectively.

Table 5 reports the estimated impacts for men. Placement on probation has small and insignificant effects on their probability of staying in college in all years, yet it has positive effects on their observed college GPAs. The estimates range from 0.084 to 0.107 in the first three years and statistically significant. The estimated effect is 0.113 (with a standard error 0.072) at the end of the fourth year. These results suggest that men may temporarily improve their performance once they are on probation. No significant improvement is found in their final GPAs' by the time they complete college. Finally, it is worth noting that for men the standard RD design yields significant estimates that are largely similar to those estimated intensive margin effects. This is what one would expect when there are no significant changes at the extensive margin, or in the dropout probability for men in this case.

Do relatively low ability women drop out once they are placed on probation? This would be plausible if they form rational expectations and make optimal decisions based on their potential gains from staying in college. Table 6 reports the estimated average characteristics of those quitting and always participating compliers among women. Quitters are more likely to be White. They have slightly higher SAT scores on average. Note that SAT score is significantly positively correlated with college final GPA in our sample. For example, SAT score alone explains over 17% of the total sample variation in college final GPA among women. In addition, quitters are more likely to be ranked among the top 25% of their high school classes. Interestingly, quitters seem to be less likely from a feeder school, suggesting that they may have fewer high-school peers with whom they can share information.

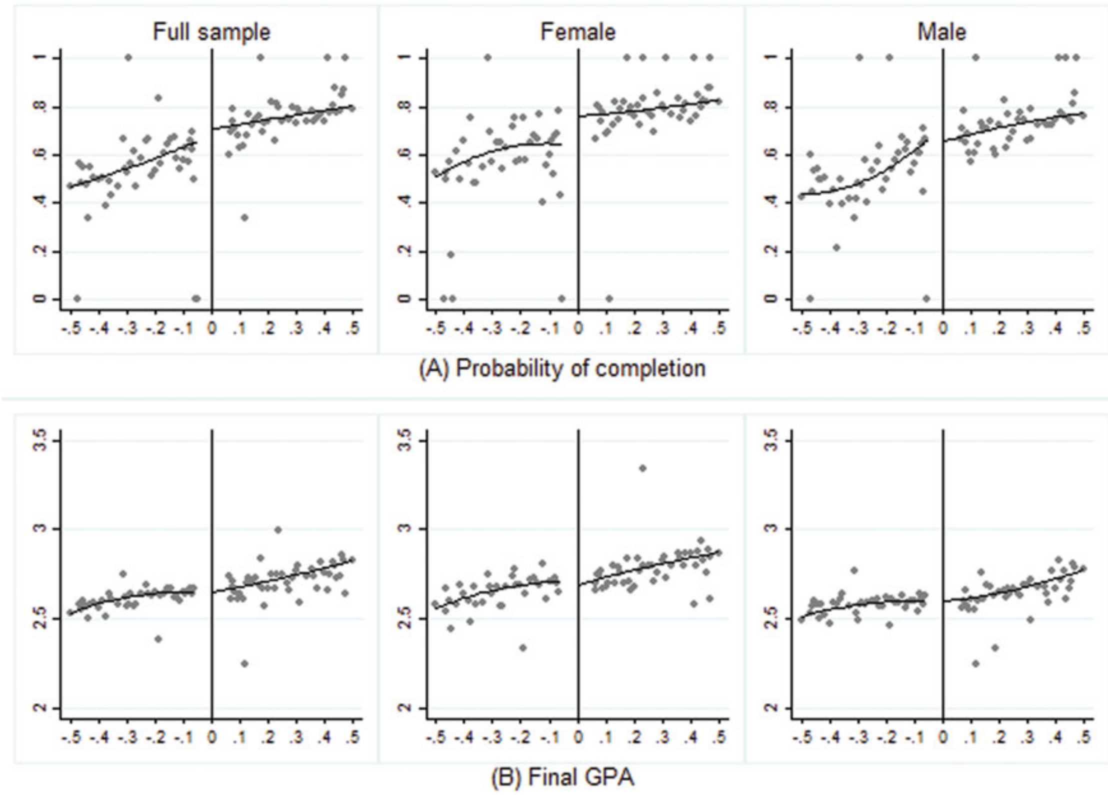

Figure 3. College completion and final GPAs against first semester GPA.

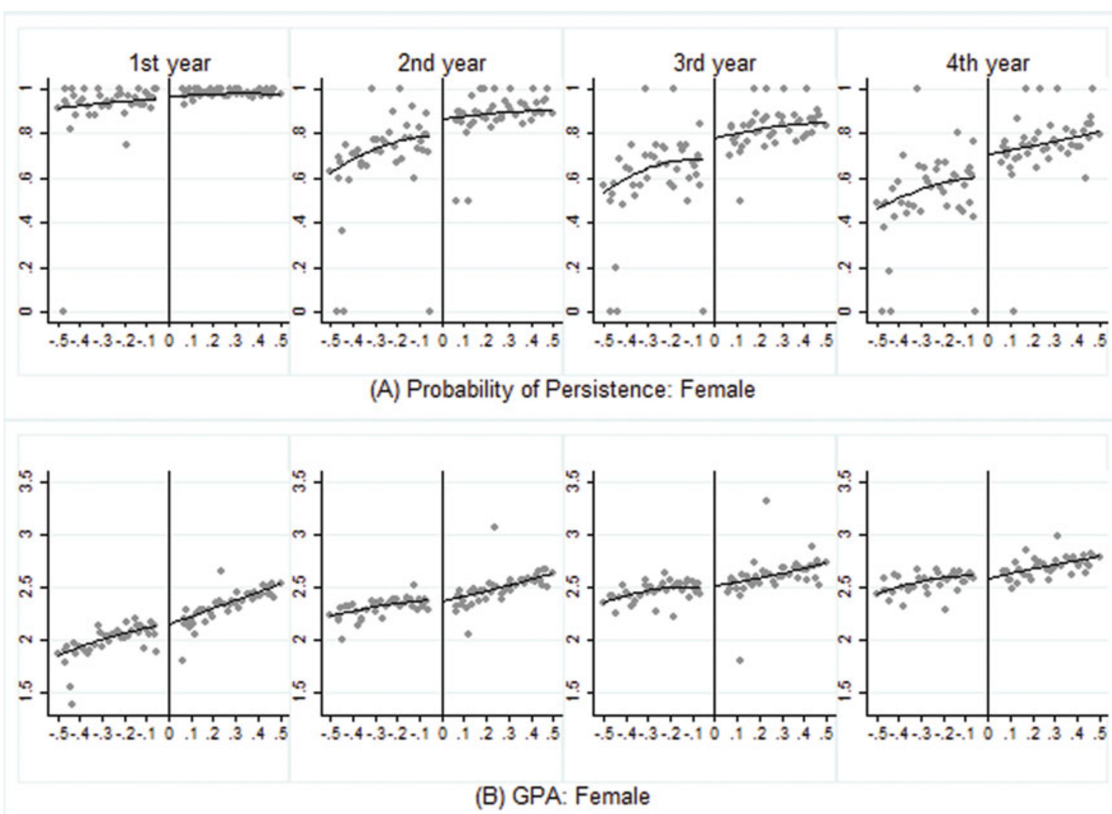

Figure 4. College persistence and GPAs for women.

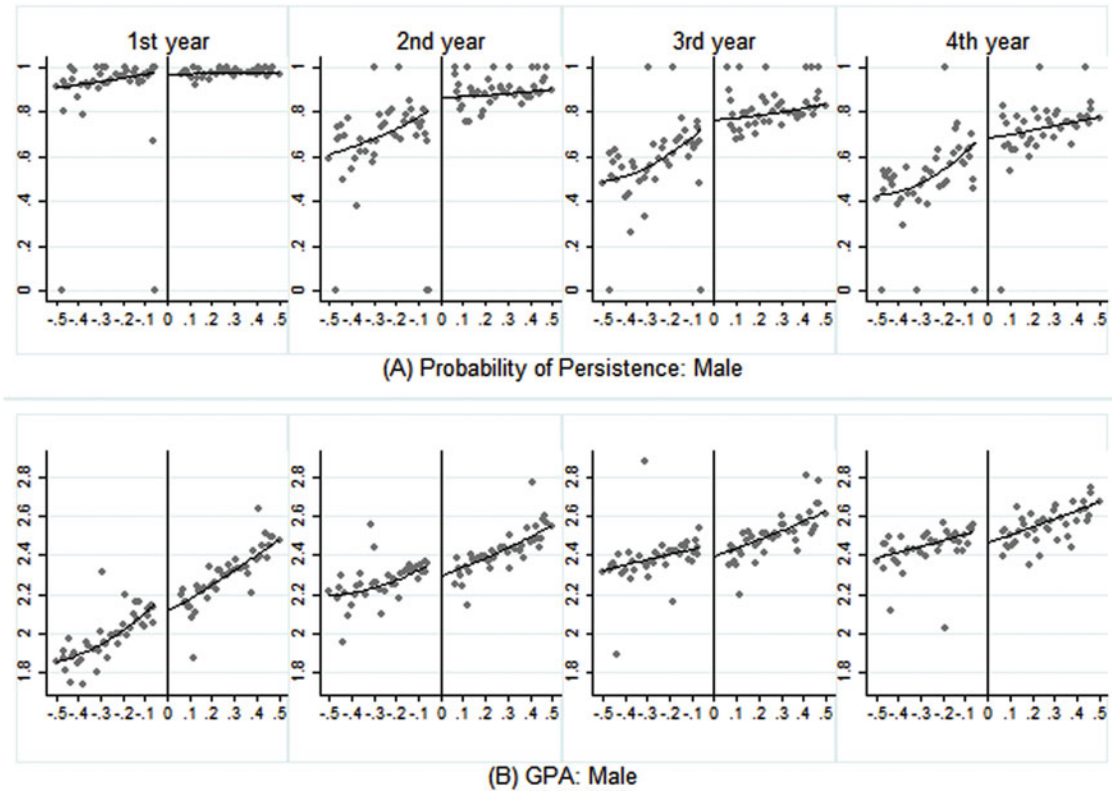

Figure 5. (A) Probability of persistence: Male; (B) GPA: Male.

Overall estimates in Table 6 do not suggest that those quitters have lower ability compared with the always participating compliers. Quitter characteristics are all estimated to carry the plausible positive sign. It is easy to test that the estimated probabilities are not negative or greater than 1, so monotonic selection is plausible. Assume that those quitters on average would perform at least the same as the always participants, had they not dropped out, that is, $E[Y_0|S_0 = 1, S_1 = 0, C, R = r_0] \geq E[Y_0|S_0 = 1, S_1 = 1, C, R = r_0]$. This is a mean dominance in the opposite direction than that implied by Assumption 3. Then analogous to the discussion at the end of Section 3.3, the upper bound on the intensive margin effect is $E[Y_1^*|S_1= 1, R = r_0, C] - \frac{1}{1-q}E[1(Y_0^* \leq Q_0(1 - q))Y_0^*|S_0= 1, R = r_0, C]$, while the lower bound is $E[Y_1^*|S_1= 1, R = r_0, C] - E[Y_0^*|S_0= 1, R = r_0, C]$. That is, the intensive margin serves as a lower bound for the true probation effect among those always participating compliers.

The bottom panels of Tables 3–5 report estimates of the bounds under monotonic selection and the above bounds under additionally the mean dominance $E[Y_0|S_0 = 1, S_1 = 0, C, R = r_0] \geq E[Y_0|S_0 = 1, S_1 = 1, C, R = r_0]$. Imbens and Manski's (2004) confidence intervals are reported, since what is of interest is the confidence interval for the true parameter, not that for the identification region. Adding the mean dominance assumption in general tightens the estimated bounds. The estimated intensive margin effects, or the lower bounds on the true probation effect for those always participating compliers, are small yet insignificant for women. At the same time, the lower ends of the 90% CIs for women are slightly below zero. That is, although we can rule out large negative impacts of placement on probation, there do not seem to be significant gains on average for women. For men, the probation effects are bounded above zero for the first three years. The lower end of the 90% confidence interval is slightly below zero in the fourth year and moves further below zero by the time they graduate. These results confirm again that men seem to temporarily improve their GPAs once they are on probation.

In Online Supplemental Appendix II, we report additional results for students who are ranked among the top 25% of their high school classes and those who are not. These additional results are consistent with a discouragement effect of placement on academic probation. In particular, those in the top quarter of their high school classes are more likely to be discouraged and hence to drop out once on probation. The impacts on the dropout rates are large (8.7%–11.6%) and statistically significant from the second academic year onward. In contrast, placement on probation has mostly positive yet insignificant impacts on college persistence among those who are not in the 25% of their high school classes. In addition, the estimated intensive margin effects are all positive. We can therefore rule out significant negative impacts of probation on GPAs, even though any positive effects might be small.

These empirical results reveal striking gender differences in response to placement on academic probation. College probation discourages women from completing college. The discouragement effect is particularly pronounced among those who perform relatively better in high school. Intuitively, placement on probation is likely to be a greater negative information shock for them. In contrast, men in general are not discouraged

Table 4. Effects on college persistence and GPAs (Women)

	1st year	2nd year	3rd year	4th year
	I: RDD with sample selection			
(1):Pr($S_0 = 1$)	0.974	0.898	0.848	0.801
	(0.005)***	(0.019)***	(0.023)***	(0.029)***
(2):Pr($S_1 = 1$)	0.956	0.779	0.686	0.641
	(0.015)***	(0.049)***	(0.065)***	(0.070)***
Extensive margin: (2)-(1)	−0.011	−0.117	−0.162	−0.167
	(0.017)	(0.052)**	(0.071)**	(0.073)**
(3): E($Y_0 \| S_0 = 1$)	2.149	2.480	2.608	2.683
	(0.011)***	(0.017)***	(0.018)***	(0.023)***
(4): E($Y_1 \| S_1 = 1$)	2.125	2.529	2.636	2.775
	(0.027)***	(0.035)***	(0.040)***	(0.056)***
Intensive margin: (4) and (3)	−0.024	0.049	0.029	0.092
	(0.042)	(0.050)	(0.070)	(0.073)
	II: Standard RDD			
	0.033	0.039	0.012	0.106
	(0.023)	(0.039)	(0.042)	(0.042)**
	III: Bounds for always participating compliers			
Lower bound 1	−0.024	−0.092	−0.228	−0.098
	(0.050)	(0.088)	(0.121)*	(0.101)
Upper bound 1	0.080	0.066	0.229	0.108
	(0.069)	(0.147)	(0.183)	(0.104)
90% CI 1	[−0.024 0.080]	[−0.209 0.262]	[−0.341 0.418]	[−0.229 0.242]
lower bound 2	−0.024	0.049	0.029	0.092
	(0.042)	(0.050)	(0.070)	(0.073)
Upper bound 2	0.080	0.066	0.229	0.108
	(0.069)	(0.147)	(0.183)	(0.104)
90% CI 2	[−0.079 0.169]	[−0.031 0.300]	[−0.063 0.421]	[−0.008 0.272]
N	51,374	51,115	48,128	40,921

Notes: All estimates are conditional on compliers at the first semester GPA equal to 2.0; estimation of the extensive and intensive margins, and the bounds follows the description in Sections 2 and 3.1, respectively. The CCT bias-corrected robust inference is used; In Panel III, 1 refers to the bounds under the monotonic sample selection assumption, while 2 refers to the bounds assuming additionally mean dominance, particularly $E(Y_0|S_0 = 1, S_1 = 0, C, R = r_0) \geq E(Y_0|S_0 = 1, S_1 = 1, C, R = r_0)$; Bootstrapped standard errors are in the parentheses; Imbens and Manski's (2004) CIs are reported; ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

by this negative signal on performance. Instead men temporarily improve their GPAs to avoid being suspended. These findings strongly suggest that, to make academic probation more beneficial, universities and colleges should take into account the discouragement effects, particularly for women.

It is worth mentioning that our findings of the gender differences differ from those documented by Lindo, Sanders, and Oreopoulos (2010). In particular, they showed that the dropout rate among men almost doubles when placed on probation while that among women has no significant changes. The differential findings could be due to different data (Canadian vs. U.S. school data) or different policy implementation. For example, students generally receive a notice about their probation status. Different universities may communicate the message differently. In addition, different universities impose different rules or restrictions for students who are on probation. They may also offer different services to assist these students. These variations may lead to different impacts on students. It is therefore of great policy interest to perform further analysis using more detailed data to investigate students' responses to this negative signal on performance.

5. CONCLUSION

This article discusses identification of treatment effects in RD designs when differential sample selection leads to incomparability of observations near the RD threshold. Sample selection or missing outcomes can frequently arise due to dropout, survey nonresponse, censoring, or many other reasons.

We deal with both treatment endogeneity and sample selection issues. Identification in this article does not require any exclusion restrictions in the selection equation, nor does it require specifying the selection mechanism. The proposed identification results can therefore be applied broadly. The key identifying assumption, smoothness of the conditional distribution of potential outcomes and potential sample selection status, is plausible under no sorting near the RD threshold. This type of smoothness conditions are typically assumed even in the standard RD design. They also have readily testable implications and can be easily verified.

This article first provides nonparametric identification of the extensive and intensive margin effects of the treatment. This article then constructs sharp bounds on the treatment effect among a well-defined subgroup of compliers, namely

Table 5. Effects on college persistence and GPAs (Men)

	1st year	2nd year	3rd year	4th year
	I: RDD with sample selection			
(1):Pr ($S_0 = 1$)	0.973	0.925	0.872	0.809
	(0.005)***	(0.017)***	(0.022)***	(0.027)***
(2):Pr ($S_1 = 1$)	0.975	0.856	0.872	0.847
	(0.015)***	(0.047)***	(0.056)***	(0.066)***
Extensive margin: (2)-(1)	0.000	−0.063	−0.008	0.036
	(0.017)	(0.050)	(0.065)	(0.078)
(3): E($Y_0 \| S_0 = 1$)	2.108	2.416	2.517	2.597
	(0.014)***	(0.018)***	(0.021)***	(0.026)***
(4): E($Y_1 \| S_1 = 1$)	2.192	2.523	2.611	2.710
	(0.024)***	(0.039)***	(0.038)***	(0.043)***
Intensive margin: (4)-(3)	0.084	0.107	0.094	0.113
	(0.027)***	(0.054)***	(0.054)*	(0.072)
	II: Standard RDD			
	0.078	0.103	0.098	0.142
	(0.025)***	(0.046)**	(0.049)**	(0.064)**
	III: Bounds for always participating compliers			
Lower bound 1	0.084	0.041	0.094	0.113
	(0.029)	(0.095)	(0.074)	(0.079)
Upper bound 1	0.084	0.188	0.094	0.113
	(0.088)	(0.138)	(0.142)	(0.194)
90% CI 1	[0.037 0.230]	[−0.086 0.371]	[−0.028 0.327]	[−0.017 0.429]
Lower bound 2	0.084	0.107	0.094	0.113
	(0.027)	(0.054)	(0.054)	(0.072)
Upper bound 2	0.084	0.188	0.094	0.113
	(0.088)	(0.138)	(0.142)	(0.194)
90% CI 2	[0.039 0.230]	[0.030 0.384]	[0.006 0.327]	[−0.006 0.429]
N	51,374	51,115	48,128	40,921

Notes: All estimates are conditional on compliers at the first semester GPA equal to 2.0; estimation of the extensive and intensive margins, and the bounds follow the description in Sections 2 and 3.1, respectively. The CCT bias-corrected robust inference is used; in Panel III, 1 refers to the bounds under the monotonic sample selection assumption, while 2 refers to the bounds assuming additionally mean dominance, particularly $E(Y_0|S_0 = 1, S_1 = 0, C, R = r_0) \geq E(Y_0|S_0 = 1, S_1 = 1, C, R = r_0)$; bootstrapped standard errors are in the parentheses; Imbens and Manski's (2004) CIs are reported; ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

Table 6. Mean characteristics of subgroups of compliers

	Always participants	Quitters
White	0.781 (0.067)***	0.893 (0.374)**
SAT score	1,093 (14.19)***	1,112 (106.4)***
Top 25% of HS class	0.774 (0.068)***	0.948 (0.460)**
HS NHS	0.268 (0.059)***	0.346 (0.450)
Feeder school	0.172 (0.055)***	0.005 (0.419)

Notes: Estimates are based on the sample of women; NHS means National Honors Society member; the CCT bias-corrected estimates are reported; bootstrapped standard errors are in the parentheses; ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

those always participating compliers. Further discussed is point identification of each subgroup characteristics among compliers.

Applying these identification results, we evaluate impacts of college probation and provide empirical evidence that is different from that by the standard RD design. We show that there are striking gender differences at the extensive versus the intensive margin in response to placement on probation in college. The probability for women to complete college decreases significantly if they have ever been placed on academic probation. Contrary to what one might expect, low ability women are not more likely to drop out. Instead those who are in the top percentiles of their high school classes are more likely to drop out once on probation. In contrast, placement on probation has little impacts on men's probability of dropping out of college. Men seem to cope with probation by temporarily improving their GPAs to avoid being suspended.

For simplicity, this article does not deal with covariates other than the running variable in developing theory and in the empirical analysis. The standard argument for RD designs applies, that is, covariates are not needed for consistency but may improve efficiency in estimating unconditional treatment effects. If desired, one can easily incorporate covariates as additional control variables in the local linear or polynomial regressions involved. In addition, this article deals with a single known cutoff. In some empirical applications, multiple cutoffs exist. For example, some colleges have floating probation thresholds that depend on the number of credit hours taken. Multiple cutoffs are also common in geographic RD designs (Keele and Titiunik 2015). This article's results can be applied by normalizing all the thresholds to be zero, providing that all the assumptions hold at each cutoff. Cattaneo et al. (2016) provided a detailed discussion on this approach in the standard RD design and the interpretation of the identified treatment effects. See also Bertanha (2016) for an alternative approach. We refer interested readers to these papers and references therein.

ACKNOWLEDGMENT

The author would like to thank anonymous referees for valuable comments.

[Received March 2016. Revised February 2017.]

REFERENCES

Ahn, H., and Powell, J. L. (1993), "Semiparametric Estimation of Censored Selection Models With a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29. [1]

Andrews, D. W. K., and Schafgans, M. (1998), "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65, 497–517. [1]

Angrist, J. D. (2001), "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business & Economic Statistics*, 19, 2–16. [4]

Angrist, J. D., Imbens, G., and Rubin, D. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455. [2]

Bertanha, M. (2016), "Regression Discontinuity Design With Many Thresholds," Working Paper. [16]

Blanco, G., Flores, C. A., and Flores-Lagunes, A. (2013), "The Effects of Job Corps Training on Wages of Adolescents and Young Adults," *American Economic Review: P&P*, 103, 418–422. [5,7]

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326. [4,6]

Cattaneo, M. D., Jansson, M., and Ma, X. (2016), "Simple Local Regression Distribution Estimators With an Application to Manipulation Testing," Working paper. [3,9,15]

Chen, X., and Flores, C. A. (2014), "Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps," Working Paper. [2,5,7]

Das, M., Newey, W. K., and Vella, F. (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58. [1]

De Chaisemartin, C. (2014), "Tolerating Defiance? Local Average Treatment Effects without Monotonicity," working paper. [3,5]

Dong, Y. (2016), "An Alternative Assumption to Identify LATE in Regression Discontinuity Models," working paper. [1,3]

Dong, Y., and Shen, S. (2016), "Testing for Rank Invariance or Similarity in Program Evaluation," working paper. [3]

Fletcher, J. M., and Tokmouline, M. (2010), "The Effects of Academic Probation on College Success: Lending Students a Hand or Kicking Them While They are Down," working paper. [8]

Frandsen, R. B. (2015), "Treatment Effects With Censoring and Endogeneity," *Journal of the American Statistical Association*, 110, 1745–1752. [2]

Frandsen, B. R., Frölich, M., and Melly, B. (2012), "Quantile Treatment Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168, 382–395. [6]

Frangakis, C. E., and Rubin, D. B. (2002), "Principal Stratification and Causal Inference," *Biometrics*, 58, 21–29. [4]

Hahn, J., Todd, P., and van der Klaauw, W. (2001), "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69, 201–209. [1,3]

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161. [1,5]

——— (1990), "Varieties of Selection Bias," *American Economic Review, P&P*, 80, 313–318. [1,5]

Heckman, J. J., Smith, J., and Clements, N. (1997), "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535. [3]

Horowitz, J. L., and Manski, C. F. (1995), "Identification and Robustness With Contaminated and Corrupted Data," *Econometrica*, 63, 281–302. [5,6]

——— (2000), "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77–84. [6]

Imai, K. (2008), "Sharp Bounds on the Causal Effects in Randomized Experiments With" Truncation-by-Death," *Statistics and Probability Letters*, 78, 144–149. [5]

Imbens, G. W., and Kalyanaraman, K. (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933–959. [4]

Imbens, G. W., and Manski, C. F. (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [6,13]

Keele, L. J., and Titiunik, R. (2015), "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155. [15]

Kim, B. M. (2012), "Do Developmental Mathematics Courses Develop the Mathematics?" working paper. [1]

Kitigawa, T. (2015), "A Test for Instrument Validity," *Econometrica*, 83, 2043–2063. [7]

Lee, D. S. (2008), "Randomized Experiments From Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697. [3]

——— (2009), "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102. [1,4,5]

Lewbel, A. (2007), "Endogenous Selection or Treatment Model Estimation," *Journal of Econometrics*, 141, 777–806. [1]

Lindo, M. J., Sanders, N. J., and Oreopoulos, P. (2010), "Ability, Gender, and Performance Standards: Evidence From Academic Probation," *American Economic Journal: Applied Economics*, 2, 95–117. [5,8,14]

McCrary, J. (2008), "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698–714. [3,9]

McCrary, J., and Royer, H. (2011), "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, 101, 158–195. [1]

Martorell, P., and McFarlin, I., Jr. (2011), "Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes," *Review of Economics and Statistics*, 93, 436–454. [1]

Staub, E. K. (2014), "A Causal Interpretation of Extensive and Intensive Margin Effects in Generalized Tobit Models," *Review of Economics and Statistics*, 96, 371–375. [2]

Stoye, J. (2009), "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315. [6]

Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. [5]

Zhang, J. L., and Rubin, D. B. (2003), "Estimation of Causal Effects via Principle Stratification When Some Outcomes are Truncated by "Death"," *Journal of Educational Behavioral Statistics*, 28, 353–368. [5]