

Treatment Effect Analyses through Orthogonality Conditions Implied by a Fuzzy Regression Discontinuity Design, with Two Empirical Studies¹

Muzhe Yang²

November 2013

¹The first empirical study uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis. I thank Kenneth Chay for generously providing the data for the second empirical study of this paper.

²Assistant Professor, Department of Economics, Rauch Business Center, Lehigh University, 621 Taylor Street, Bethlehem, PA 18015. Phone: (610) 758-4962. Fax: (610) 758-4677. E-mail: muzheyang@lehigh.edu.

Abstract

This study proposes a new estimator for estimating a treatment effect in one particular fuzzy regression discontinuity (RD) setting, in which the treatment effect is homogeneous on the support of an assignment variable and the treatment assignment is exogenous conditional on that assignment variable. The estimator is constructed using orthogonality conditions and can be easily implemented by an instrumental variable (IV) estimation procedure. We use Monte Carlo experiments to show that the proposed estimator can substantially reduce the bias in estimating the treatment effect caused by misspecifying the regression model of the observed outcome. We also use two empirical studies to demonstrate the advantages of our proposed estimator over alternative estimators. Furthermore, we use the first empirical study to highlight a connection between our proposed estimator and propensity-score matching estimators. The second empirical study emphasizes that the proposed estimator can work in a fuzzy RD setting where the cutoff point is either unknown or not exactly known.

1 Introduction

A key problem of analyzing a treatment effect is the impossibility of observing an outcome in the treated and in the untreated states simultaneously. Thus, a treatment effect defined as the difference between potential outcomes in the presence and in the absence of a treatment has to be inferred under assumptions that allow for comparing the outcomes of different individuals with different treatment status (Holland, 1986; Rubin, 1974). Among various empirical strategies¹ employed to identify a treatment effect, random assignment of the treatment serves as an effective way to ensure individuals with different treatment status are comparable in their observable and unobservable characteristics. However, for many of the most general questions in the social sciences, random assignment is either too costly to implement or viewed as unethical. Nevertheless, in some empirical settings there can be compelling reasons to believe that individuals receive a treatment by chance in the absence of a formal random assignment. One of those empirical settings permits the use of a regression discontinuity (RD) design (Hahn, Todd and Van der Klaauw, 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010).²

One important feature of an RD design is that the treatment assignment is known to be dependent on an observed covariate, often referred to as the “assignment” variable (a.k.a., “running” or “forcing” variable) in the literature. Hahn, Todd and Van der Klaauw (2001) utilize this feature by focusing on the expected outcome and the treatment probability conditional on the assignment variable. With this focus they show that a treatment effect can be nonparametrically identified at the cutoff value of the assignment variable, at which the treatment probability changes discontinuously whereas the expected outcome in the absence of the treatment changes continuously. For a sharp RD, the treatment probability changes from 0 to 1 at the cutoff, which is exact and known to researchers. Although this knowledge

¹Imbens and Wooldridge (2009) provide an overview and discussions on the recent development of empirical strategies for treatment effect analyses.

²For a history and overview of RD design, see Cook (2008), Thistlethwaite and Campbell (1960), and Trochim (1984).

allows researchers to skip estimating the change in the treatment probability at the cutoff, using this knowledge inevitably requires knowing the exact change in the expected outcome also at the cutoff. For the latter, Hahn, Todd and Van der Klaauw (2001) propose nonparametric methods such as local linear regressions. Imbens and Kalyanaraman (2012) further discuss how to choose an optimal bandwidth required by nonparametric regressions applied to RD designs.

Our study focuses on a fuzzy RD design, where the treatment probability is known to change at the cutoff value of the assignment variable, but the exact magnitude is unknown. In this case researchers lack the exact knowledge about the treatment assignment. Previous studies on the fuzzy RD have emphasized the nonparametric identification of a treatment effect at the cutoff, which includes the sharp RD as a special case (Hahn, Todd and Van der Klaauw, 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Existing estimation methods are aimed at pinpointing the functional form of the expected outcome and the treatment probability conditional on the assignment variable. In contrast, our study focuses on a treatment effect away from the cutoff and uses orthogonality conditions to construct an estimator that can be easily implemented using an instrumental variable estimation procedure.

Specifically, our study utilizes a possibility that the fuzzy RD permits while the sharp RD forbids: in the fuzzy RD case two individuals with different treatment status may have the same treatment probability conditional on the assignment variable, whereas in the sharp RD case two individuals with the same value of the assignment variable cannot have different treatment status. Furthermore, we examine one particular fuzzy RD in which the following two assumptions hold: first, the treatment effect is homogeneous on the support of the assignment variable; and second, the treatment is randomly assigned near the cutoff conditional on the assignment variable.

The first assumption essentially allows us to identify a treatment effect not just *at*, but *local to* the cutoff, since the treatment effect is constant across the values of the assignment

variable, including the cutoff value. A recent study by Angrist and Rokkanen (2012) also use this assumption to explore the possibilities of identifying a treatment effect away from the cutoff, which, as they demonstrate by their study on “inframarginal applicants,” can be of greater policy interest. Different from their identification strategies, we use the second assumption aforementioned, which is also referred to as the “unconfoundedness” assumption, to derive orthogonality conditions that have not been used in the RD literature.

The central idea of our identification strategy is the use of the deviation between the actual treatment status and the treatment probability predicted by the assignment variable, which is infeasible for a sharp RD. Under the unconfoundedness assumption the deviation occurs exogenously for individuals who have the same value of the assignment variable. This means that the treatment assignment deviates from the assignment variable for reasons not related to the potential outcomes in the presence and in the absence of the treatment. When this is true, we show that the deviation is uncorrelated with both the unobservable (characterized by an error term) and the observable (characterized by a function of the assignment variable) in the regression model of the observed outcome. We derive two orthogonality conditions and use them to construct a new estimator, which estimates an average treatment effect for the population near the cutoff where these two assumptions are imposed: the unconfoundedness and the homogeneity of the treatment effect with respect to the assignment variable.

The implementation of our proposed estimator requires two stages: first, estimate the probability of being treated conditional on the assignment variable, which is essentially the propensity score of the treatment; and next, use the difference between the treatment status and the propensity score as an instrumental variable for the treatment status in the regression of the observed outcome on the treatment status only, without the intercept term. This proposed estimator is easy to implement and has several advantages over alternative estimators, which are demonstrated through Monte Carlo experiments and two empirical studies. Three key findings are summarized as follows.

First, our proposed estimator does not require specifying the regression model of the observed outcome, and thus it can avoid the specification error. We use Monte Carlo experiments to illustrate the possibility of incurring significant biases due to specification errors when using ordinary least squares (OLS) as opposed to our proposed estimator.

Second, the proposed estimator requires estimating a propensity score in the first stage. We use Monte Carlo experiments to show that a misspecified model for the propensity score can have small impacts on the treatment effect estimates—a very different result compared with the impact of misspecifying the regression model of the observed outcome. We also apply our proposed estimator to an empirical study on the effect of crisis-induced anxiety on marijuana use. In this empirical study we use the propensity score estimated in the first stage to conduct a balancing check on the observed covariates and also a sensitivity analysis. As Rosenbaum (2010) suggests, the balancing check helps to detect specification errors of the propensity score model, and the sensitivity analysis can be useful for assessing the magnitude of selection bias due to unobservables.

Third, our proposed estimator does not require knowing the exact cutoff value. Therefore, it allows researchers to deal with situations in which the cutoff is unknown, or not exactly known. To demonstrate this point, we use an empirical study previously done by Chay, McEwan and Urquiola (2005) on the evaluation of a school-based intervention program. One important feature of this intervention program is the use of multiple cutoff points for selection eligibility. Assuming those multiple cutoffs are unknown to researchers, we demonstrate the advantage of the proposed estimator over the OLS used by Chay, McEwan and Urquiola (2005) and also over the local linear regression estimator proposed by Hahn, Todd and Van der Klaauw (2001).

The rest of the paper proceeds as follows. Section 2 presents the model and derives the estimator, along with explanations on the inference, implementation and applicability of the proposed estimator. Section 3 examines the finite sample performance of the proposed estimator in comparison with alternative estimators through Monte Carlo experiments. The

advantages of the proposed estimator are further demonstrated by two empirical studies in Section 4, followed by conclusion remarks in Section 5.

2 The Model and Proposed Estimator

In this section we use a simple model to illustrate a case where orthogonality conditions are implied by a particular fuzzy RD. We use those orthogonality conditions to construct a new estimator, which can be easily implemented using a standard instrumental variable (IV) estimation procedure.

The model represents a case in which a treatment is assigned exogenously, and the effect of the treatment is homogeneous on the support of the assignment variable that has a cutoff value at which the treatment probability changes discontinuously. To fix this idea, we start with the following specification:³

$$y_j = \alpha_j + g(x) + u, \text{ where } \mathbb{E}(u|x) = 0 \text{ and } j = 0, 1. \quad (1)$$

In this equation y_j denotes an outcome in the treated ($j = 1$) or the untreated (i.e., the control) state ($j = 0$); α_j is an intercept term; u is an additive error term; and $g(x)$ is a continuous and smooth function of an observable variable x , henceforth referred to as the assignment variable. Based on equation (1), the treatment effect averaged across x is given by

$$\mathbb{E}(y_1 - y_0) = \mathbb{E}_x[\mathbb{E}(y_1 - y_0|x)] = \alpha_1 - \alpha_0 \equiv \theta. \quad (2)$$

Our study focuses on this particular kind of average treatment effect, which is homogeneous on the support of the assignment variable.

Note that y_1 and y_0 are not simultaneously observable. Thus, the observed outcome,

³The following discussions of identification and estimation of treatment effects do not consider additional covariates, to avoid unnecessary complications. Imbens and Lemieux (2008, p. 625) discuss the methods of including additional covariates.

denoted by y , can be represented by a linear combination of y_1 and y_0 through a dummy variable for treatment status:

$$y = y_0 + d(y_1 - y_0) = \alpha_0 + d\theta + g(x) + u, \quad (3)$$

where d equals 1 (or 0) if the treatment is received (or not). We use the following equation to further explain this treatment status variable:

$$d = p(x) + v, \text{ where } \mathbb{E}(v|x) = 0. \quad (4)$$

Here, v is the error term, which represents the unobservables in the treatment assignment process. One implication of equation (4) is that $p(x)$ equals $\mathbb{E}(d|x)$, which is $\Pr(d = 1|x)$. Using the term given by Rosenbaum and Rubin (1983), we will call $p(x)$, or $\Pr(d = 1|x)$, the propensity score henceforth. We assume that the propensity score has a discontinuous change at one value of x , denoted by x_0 . Our study focuses on a fuzzy RD, which means that the discontinuous change of $p(x)$ at x_0 is strictly between 0 and 1.

Next, we introduce a key assumption used throughout this paper: y_j (in equation 1) is independent of d (in equation 4) conditional on x . In other words, we assume that the treatment assignment (d) is exogenous conditional on the assignment variable (x). Note that this assumption is essentially the unconfoundedness assumption used by matching and inverse probability weighting (IPW) estimators (Imbens and Wooldridge, 2009). Different from these existing estimators, our proposed estimator uses two orthogonality conditions easily derived under the unconfoundedness assumption.

First, the unconfoundedness assumption implies $\mathbb{E}(u|v, x) = \mathbb{E}(u|x) = 0$, which also means that $\mathbb{E}(u|d, x) = \mathbb{E}(u|x) = 0$. This further implies⁴

$$\mathbb{E}[(d - p(x)) \cdot u] = 0. \quad (5)$$

⁴Note: $\mathbb{E}[(d - p(x))u] = \mathbb{E}_x\{\mathbb{E}[(d - p(x))u|x]\} = \mathbb{E}_x[\mathbb{E}(du|x) - p(x)\mathbb{E}(u|x)] = \mathbb{E}_x[p(x)\mathbb{E}(u|x) + \mathbb{E}(vu|x) - p(x)\mathbb{E}(u|x)] = \mathbb{E}_x\{v\mathbb{E}_v[\mathbb{E}(u|v, x)]\} = 0$.

Second, equation (4) gives $\mathbb{E}(d - p(x)|x) = 0$, which implies⁵

$$\mathbb{E}[(d - p(x)) \cdot g(x)] = 0. \quad (6)$$

Next, we apply the two orthogonality conditions (equations 5 and 6), together with equation (4), to equation (3) and obtain the following equation:

$$\mathbb{E}[(d - p(x)) \cdot y] = \underbrace{\mathbb{E}(d - p(x))}_{=0} \cdot \alpha_0 + \mathbb{E}[(d - p(x)) \cdot d] \cdot \theta + \underbrace{\mathbb{E}[(d - p(x)) \cdot g(x)]}_{=0} + \underbrace{\mathbb{E}[(d - p(x)) \cdot u]}_{=0}. \quad (7)$$

Thus, we have the following result:

$$\theta = \frac{\mathbb{E}[(d - p(x)) \cdot y]}{\mathbb{E}[(d - p(x)) \cdot d]}. \quad (8)$$

Using the analogy principle (Cameron and Trivedi, 2005, p. 135), we propose the following estimator based on equation (8) for a random sample ($i = 1, 2, \dots, N$):

$$\hat{\theta} = \frac{\sum_{i=1}^N (d_i - \hat{p}(x_i)) \cdot y_i}{\sum_{i=1}^N (d_i - \hat{p}(x_i)) \cdot d_i}. \quad (9)$$

A key feature of our proposed estimator is the use of the *difference* between the treatment status (d) and the propensity score ($p(x)$), which distinguishes it from matching and IPW estimators under the unconfoundedness assumption. In the following we give further explanations on the inference, implementation and applicability of this estimator.

First, our proposed estimator belongs to a class of estimators based on the method of moments, and as a result it is root- N consistent and asymptotically normal (Cameron and Trivedi, 2005, Chapter 6).

Second, the proposed estimator has the same formula as an IV estimator in the just-identification case: the difference between d and $p(x)$ is in the position of an IV for d in the

⁵Note: $\mathbb{E}[(d - p(x))g(x)] = \mathbb{E}_x\{\mathbb{E}[(d - p(x))g(x)|x]\} = \mathbb{E}_x[\mathbb{E}(d|x)g(x) - p(x)g(x)] = \mathbb{E}_x[p(x)g(x) - p(x)g(x)] = 0$.

regression of y on d without the intercept term. But note that our proposed estimator is derived under the assumptions different from those required by an IV estimator.

Third, there are two stages (or steps) to implement this estimator. In stage one, estimate the propensity score and obtain $\hat{p}(x)$. In stage two, run an IV regression of y on d without the intercept term, using the difference between d and $\hat{p}(x)$ as the IV for d . The standard error of $\hat{\theta}$ is obtained from the stage-two IV estimation. Note that in stage two the instrument ($d - \hat{p}(x)$) is actually estimated, but the generated instrument does not affect the asymptotic variance of $\hat{\theta}$, which is not true if a regressor is generated (Wooldridge, 2010, pp. 125–126).

Fourth, this estimator is applicable to a case where the treatment assignment (d) is exogenous conditional on the assignment variable (x) and the treatment effect (θ) is constant over the support of that assignment variable. These conditions can be arguably, and of course not exclusively, met in some RD settings: subjects with the assignment variable having values close to a cutoff point (i.e., $|x - x_0|$ being small) receive the treatment randomly, and the treatment effect does not vary with the assignment variable (x). Note that our proposed estimator cannot be applied to a sharp RD, where $p(x)$ takes only two values (0 and 1) and changes from 0 to 1 at the cutoff point (x_0). To emphasize these points just discussed, we refer to our proposed estimator as fuzzy RD estimator. In Section 4.1 we illustrate an application of our proposed estimator in comparison with alternative estimators.

Fifth, one implication of our proposed estimator is that the estimator can work in an RD setting where the cutoff is unknown or not exactly known. In those cases it is likely to misspecify the model for estimating the propensity score required in the first stage. In Section 3 we use Monte Carlo experiments to demonstrate that the impact of misspecifying the propensity score on the treatment effect estimates can be small. We also use an empirical application described in Section 4.2 to illustrate an advantage of using our proposed estimator over conventional estimators in a setting where the cutoff is not exactly known.

Sixth, our study focuses on one kind of average treatment effect, which is homogeneous on the support of the assignment variable. As a result, the treatment effect can be identified

at the cutoff point (x_0) and estimated using a local linear nonparametric regression method proposed by Hahn, Todd and Van der Klaauw (2001), which we refer to as the conventional RD estimator in this paper. In the following sections we use Monte Carlo experiments (Section 3) and two empirical applications (Section 4) to demonstrate that our proposed estimator can perform significantly better than the conventional RD estimator under the unconfoundedness assumption in finite samples.

Lastly, our proposed estimator is asymptotically equivalent to the two-stage estimator proposed by Robinson (1988) for partial linear regression models such as equation (3). The Robinson’s estimator uses the following result:

$$\theta = \frac{\mathbb{E}[(d - \mathbb{E}(d|x)) \cdot (y - \mathbb{E}(y|x))]}{\mathbb{E}[(d - \mathbb{E}(d|x)) \cdot (d - \mathbb{E}(d|x))]}, \quad (10)$$

which requires estimating both $\mathbb{E}(y|x)$ and $\mathbb{E}(d|x)$. In contrast, our proposed estimator does not require estimating $\mathbb{E}(y|x)$, which can reduce the computation burden significantly when $\mathbb{E}(y|x)$ is estimated nonparametrically.⁶ In Section 3 we use Monte Carlo experiments to confirm that our proposed estimator, without estimating $\mathbb{E}(y|x)$, can produce estimates very similar to the ones given by Robinson’s estimator in finite samples.

3 Monte Carlo Experiments

In this section we conduct a series of Monte Carlo experiments to investigate the finite sample performance of the proposed fuzzy RD estimator in comparison with alternative estimators.

The sample size considered varies from 100 to 2,000. All simulations use 1,000 repetitions.

⁶Our proposed estimator can be derived from Robinson’s estimator using these two equalities: (1) $\mathbb{E}[(d - \mathbb{E}(d|x))(y - \mathbb{E}(y|x))] = \mathbb{E}[(d - \mathbb{E}(d|x))y]$; and (2) $\mathbb{E}[(d - \mathbb{E}(d|x))(d - \mathbb{E}(d|x))] = \mathbb{E}[(d - \mathbb{E}(d|x))d]$. Furthermore, our proposed estimator and Robinson’s estimator are related to the slope parameter of a linear regression model as follows:

$$\theta = \frac{\mathbb{E}[(d - p(x)) \cdot y]}{\mathbb{E}[(d - p(x)) \cdot d]} = \frac{\mathbb{E}[(d - \mathbb{E}(d|x)) \cdot (y - \mathbb{E}(y|x))]}{\mathbb{E}[(d - \mathbb{E}(d|x)) \cdot (d - \mathbb{E}(d|x))]} = \frac{\mathbb{E}_x[Cov(d, y|x)]}{\mathbb{E}_x[Var(d|x)]}.$$

3.1 Design and Estimators

We specify a data-generating process (DGP) as follows. The assignment variable (x) is drawn from a uniform distribution with support between -1 and 1. Thus, the mean and the variance of x are 0 and 1/3, respectively. Without loss of generality, we assume that a discontinuous change in the probability of receiving a treatment occurs at the zero value of x . The treatment (d) is binary and defined by the following indicator function:

$$\begin{aligned} d &= 1\{-0.5 + z + x + v \geq 0\}, \text{ where} & (11) \\ z &= 1\{x \geq 0\} \text{ and } v \sim N(0, 1) \text{ independent of } x. \end{aligned}$$

Here, d equals 1 (or 0) for receiving the treatment (or not). The treatment probability jumps discontinuously at the zero value of x with the size equal to 0.383.⁷ This discontinuity is shown in the top panel of Figure 1.

The potential outcomes in the treated (y_1) and untreated (y_0) states are defined as follows:

$$y_0 = 1 + \frac{1}{e^{|x|}} + x^2 + x^3 + u; \quad (12)$$

$$y_1 = 2 + \frac{1}{e^{|x|}} + x^2 + x^3 + u; \text{ and} \quad (13)$$

$$u \sim \chi^2(1) - 1, \text{ independent of } v \text{ and } x.$$

Here, y_0 and y_1 are generated by the same nonlinear function of x and the additive error term u , which is positively skewed with a mean of 0 and variance of 2. The difference in the intercept term between y_1 and y_0 represents the treatment effect, which is equal to 1 (i.e., $y_1 - y_0$) and homogeneous on the support of x . The observed outcome, denoted by y , can

⁷ $\lim_{x \rightarrow 0^+} \Pr(d = 1|x) - \lim_{x \rightarrow 0^-} \Pr(d = 1|x) = \Phi(0.5) - \Phi(-0.5) = 0.383$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

be represented by a linear combination using equations (11)–(13):⁸

$$y = dy_1 + d(y_1 - y_0) = 1 + d + \frac{1}{e^{|x|}} + x^2 + x^3 + u. \quad (14)$$

There is a discontinuous change in the expected value of y conditional on x occurring at the zero value of x , with the size by design also equal to 0.383.⁹ This discontinuity is shown in the bottom panel of Figure 1. Given the DGP, the treatment effect can be identified by the two discontinuities as follows:

$$\frac{\lim_{x \rightarrow 0^+} \mathbb{E}(y|x) - \lim_{x \rightarrow 0^-} \mathbb{E}(y|x)}{\lim_{x \rightarrow 0^+} \mathbb{E}(d|x) - \lim_{x \rightarrow 0^-} \mathbb{E}(d|x)} = \frac{0.383}{0.383} = 1. \quad (15)$$

Next, we use four estimators to estimate the treatment effect. The first one is the fuzzy RD estimator proposed in this paper. In the first stage we estimate $\Pr(d = 1|x)$ using two different models: (1) a probit model of d on z and x ; and (2) a linear probability model (LPM) of d on z and x . In the second stage we use the difference between d and the estimated $\Pr(d = 1|x)$ from the first stage—the first-stage residual ($d - \widehat{\Pr}(d = 1|x)$)—as an IV for d , and we implement an IV estimation of y on d only, without the intercept term.

The second estimator is the OLS of regressing y on d and x . We choose not to use a higher order polynomial of x to approximate the nonlinearity of $\mathbb{E}(y|x)$ (equation 14) in order to gauge the impact of misspecifying $\mathbb{E}(y|x)$ on the treatment effect estimate.

The third estimator is the two-stage estimator proposed by Robinson (1988) for partial linear regression models such as the one described in equation (14). In the first stage we estimate $\mathbb{E}(y|x)$ and $\mathbb{E}(d|x)$ nonparametrically, using a Gaussian kernel-weighted local polynomial fit. In the second stage we run a linear regression of $y - \widehat{\mathbb{E}}(y|x)$ on $d - \widehat{\mathbb{E}}(d|x)$ without the intercept term.¹⁰

⁸We use such a nonlinear function of x to avoid a case where a linear regression model of y on d and x is correctly specified.

⁹ $\lim_{x \rightarrow 0^+} \mathbb{E}(y|x) - \lim_{x \rightarrow 0^-} \mathbb{E}(y|x) = 0.383.$

¹⁰We use *Stata* to implement Robinson’s two-stage estimator, with the command “semipar” written by Verardi and Debarsy (2012).

The final estimator is the one described by equation (15), which we refer to as the conventional RD estimator in this paper. To implement this estimator, we use local linear regression models for $\mathbb{E}(y|x)$ and $\mathbb{E}(d|x)$ on both sides of the cutoff (i.e., the zero value of x) with a triangle kernel, and the optimal bandwidth proposed by Imbens and Kalyanaraman (2012).¹¹

3.2 Results

We evaluate the finite-sample performance of the proposed fuzzy RD estimator in comparison with the other three estimators, using five criteria—mean bias, median bias, root mean square error (root MSE), median absolute error, and standard deviation. The results are reported in Tables 1–5 and Figures 2–3.

For the proposed fuzzy RD estimator there is a difference in the mean bias (Table 1) and the median bias (Table 2) between using a probit (column 1) and using a linear probability model (column 2) for the first-stage estimation of the treatment probability. But, the difference (in magnitude) in the biases is small, which is within one percentage point across all samples. Furthermore, the difference is ignorable for the median absolute error (columns 1 and 2 of Table 4). Given the DGP, we know that the probit model is the correct one for estimating the treatment probability while the LPM is not. Nonetheless, results of Tables 1, 2 and 4 suggest that the impact of misspecifying the model of treatment probability on the treatment effect estimate could be small. Moreover, there is little difference in the efficiency of the fuzzy RD estimator between using the probit and using the LPM in the first stage, which is suggested by the root MSE (columns 1 and 2 of Table 3) and by the standard deviation (columns 1 and 2 of Table 5). Figure 2 further demonstrates that the limit distribution (as sample size gets large) of the fuzzy RD estimator is quite similar between using probit and using LPM for the first-stage treatment probability estimation.

In comparison with the OLS that ignores the nonlinearity of $\mathbb{E}(y|x)$ in x (equation 14), our

¹¹We use *Stata* to implement this conventional RD estimator, with the command “rd” written by Nichols (2011).

proposed fuzzy RD estimator has significantly smaller bias in magnitude across all samples in the cases of mean bias (columns 1–3 of Table 1) and median bias (columns 1–3 of Table 2). Our proposed estimator also has smaller median absolute error than the OLS across all samples (columns 1–3 of Table 4), although the differences are not as substantial as the ones in the cases of mean bias and median bias. Overall, the results suggest that the impact of misspecifying $\mathbb{E}(y|x)$ on the treatment effect estimate could be large, and the benefit of using our proposed estimator, which does not require specifying $\mathbb{E}(y|x)$, could be substantive. Furthermore, Table 3 shows that in almost all samples the OLS incurs a larger root MSE than our proposed estimator, while Table 5 shows that the OLS is only slightly more efficient than our proposed estimator across all samples. This pattern suggests that there could be modest efficiency gain in choosing OLS over our proposed estimator, but the gain is at the cost of incurring a much larger bias in the treatment effect estimate.

We also compare our proposed estimator with the Robinson’s estimator. The finite-sample performances are quite similar between the two estimators (columns 1–2 and 4 of Tables 1–5). The similarity indeed confirms that for the given DGP estimating $\mathbb{E}(y|x)$ is not necessary and can be avoided, as our proposed estimator does. Furthermore, in many empirical applications nonparametric estimation of $\mathbb{E}(y|x)$ can be very challenging, especially when x is a vector including many control variables. In those situations our proposed estimator can be more suitable and feasible than the Robinson’s estimator.

Lastly, we examine the finite-sample performance of the conventional RD estimator. Figure 3 shows that in contrast to our proposed estimator the magnitudes of mean bias, median bias and median absolute error of the conventional RD estimator (also reported in column 5 of Tables 1, 2 and 4) are much larger than those of our proposed estimator even when the latter misspecifies the first-stage treatment probability model (column 2 of Tables 1, 2 and 4) across all samples. Moreover, Table 5 shows that the conventional RD estimator (column 5) produces very noisy estimates especially when the sample size is below 500. Similarly, the root MSE of the conventional RD estimator (column 5 of Table 3) is

huge relative to that of our proposed estimator in small samples, which is shown in Table 3 and Figure 3. These results show that our proposed estimator significantly outperforms the conventional RD estimator when the sample size is small in the given DGP.

To sum up, we find evidence suggesting that our proposed estimator outperforms the OLS in a case where the treatment variable is exogenous, but the regression model of the observed outcome is misspecified. In contrast, the impact of misspecifying the first-stage treatment probability model of our proposed estimator on the treatment effect estimate can be small. Our proposed estimator can also outperform the conventional RD estimator in terms of both bias and efficiency, and the gain can be substantial for small samples, under the unconfoundedness assumption and when the treatment effect is homogeneous on the support of the assignment variable.

4 Empirical Applications

In this section we use two empirical studies to illustrate the application of the proposed estimator. The first empirical study examines a case where a treatment is likely to be exogenously assigned conditional on the associated assignment variable, and the treatment effect is unlikely to vary with the value of that assignment variable. In the second empirical study we apply the proposed estimator to a case where the cutoff point required by an RD design is not exactly known. In both studies we find advantages of using the proposed estimator over several alternative estimators.

4.1 Crisis-Induced Anxiety and Marijuana Use: Evidence from the 9/11 Terrorist Attacks

The first empirical study examines the effect of anxiety on marijuana use. Specifically, we focus on a case where anxiety is likely induced by an external crisis—the September 11 terrorist attack (hereafter 9/11). Disorders from anxiety have been identified as influences

central to marijuana use, along with other contributing factors such as age, gender, exposure to family conflicts, and having peers who are marijuana users (Thompson, 2001). A study by Khoury et al. (2010) finds that exposure to traumatic experience during childhood is associated with substance abuse and substance dependence. We expand that study by considering exposure to a crisis such as 9/11 that occurs during young adulthood.

4.1.1 Empirical setting and research design

The empirical setting of our study is the same as the one used by Wang and Yang (2013). We use data from the third wave (Wave III) of the National Longitudinal Study of Adolescent Health (Add Health), which has altogether four waves of panel data designed to be nationally representative.¹² The Wave III sample includes 15,197 young adults who were interviewed between July 2001 and April 2002;¹³ the average age was 22, and about 98.17 percent of them were aged between 19 and 25. One unique feature of the Wave III Add Health is the continuation of the interview process during and after 9/11, except only for the transmission of biomarker data because of the closing of the Federal Express service.¹⁴ In Figure 4 we plot the distribution of the interview dates of Wave III, where the vertical line indicates September 11, 2001. The numbers of interviews are indeed similar immediately before and after 9/11, which is consistent with the fact that the Add Health surveys continued despite the occurrence of 9/11.

Our study uses this empirical setting to examine the effect of anxiety, likely induced by 9/11, on marijuana use frequencies of those who were interviewed just before and just after 9/11. However, we are unable to further examine the effect by the distance between the

¹²For one robustness check (in Figure 5) we use data from Wave II, conducted from April to August 1996. More detailed descriptions of the design and data collection of Add Health are provided in Wang and Yang (2013).

¹³A few Wave III in-home interviews (including 218 respondents) were conducted as pretests in April 2001.

¹⁴“A unique challenge faced by the project was the closing of the Federal Express service and airports in the week following the September 11, 2001 attack on the World Trade Center and the Pentagon. During this time, field interviewers were instructed to keep all the specimens that were collected in their refrigerator and to ship as soon as airports re-opened and the Federal Express service re-initiated. Despite these shipping delays, the project team decided to maintain the field work operation active, to the extent possible.” (Source: <http://www.cpc.unc.edu/projects/addhealth/data/guides/biomark.pdf>)

place of an Add Health interview and the locations of 9/11 attacks such as New York City, because there is no actual geographic identifiers in the Wave III data. Nevertheless, studies summarized in Marshall and Galea (2004) show that people not only in the attacked region but also across the country reported experiences similar to post-traumatic stress disorder (PTSD) in the first week after 9/11. “In all studies, having anxiety symptoms or meeting criteria for PTSD was strongly associated with number of hours of television watched on September 11 and in the days afterward.” (Marshall and Galea, 2004, p. 37)

In the following analysis we use the response to this survey question as a crude measure for having anxiety symptom: “Now, think about the past seven days. How often was the following true during the past seven days? You were bothered by things that usually don’t bother you.” Responses were recorded on a 0–3 scale: “0: never or rarely; 1: sometimes; 2: a lot of the time; 3: most of the time or all of the time.” We generate an indicator of whether an individual exhibited any anxiety symptom: it equals one if the recorded response is positive, and zero if the recorded response is zero.

In Wave III Add Health, marijuana use frequencies were measured by the number of times using marijuana in the past 30 days. In the following analysis we focus on the marijuana use relative frequency, which is the number of times using marijuana in the past 30 days divided by 30. Among 15,123 respondents for whom we have information about whether and how often they used marijuana, 98.77 percent of them have relative frequencies between zero and one. Thus, we also use those relative frequencies as proxies for the marijuana use probability measured on a monthly basis.

For each individual we have demographic information such as age, gender, race/ethnicity, education and income. Table 6 presents the summary statistics based on the full sample and three subsamples: within three days, within seven days, and within 30 days before and after 9/11. The mean of the “9/11” dummy variable represents the proportion of respondents who were interviewed on or after September 11, 2001. Note that in all three subsamples almost exactly 50 percent of the respondents were interviewed either before or after September 11,

2001, which is consistent with the fact that the Add Health interviews were not suspended because of 9/11. Similarly, in all three subsamples the proportions of individuals who, in the past seven days, were bothered by things that usually are not bothersome are also close to 50 percent. This pattern implies that within a short time frame, such as a week, being bothered or not by things that usually are not bothersome could be equally likely, or viewed as random.

Next, we provide evidence suggesting that anxiety symptoms measured by the question of being bothered by things that usually are not bothersome could be driven by an external crisis such as 9/11. Panel A of Figure 5 shows that those interviewed right after 9/11 on average have higher daily average responses to the question of being bothered by things that usually aren't bothersome than those who were interviewed just before 9/11. To check whether this post-9/11 increase is driven by any unobservable characteristics of the respondents, we plot in Panel B those same respondents' daily average responses to the "being bothered" question in Wave II when there was no 9/11. In contrast to Panel A, in Panel B we find little evidence of any significant difference, on average, in the responses to the "being bothered" question asked in Wave II of those who were interviewed just before and just after 9/11 in Wave III.

Table 7 presents the OLS estimates of the effect of 9/11 on anxiety across various specifications for the three subsamples and the full sample. Overall, the results suggest that there could be an increase of 4–12 percentage points in having anxiety symptoms for those who were interviewed just after 9/11. Taken together, Figure 5 and Table 7 suggest that anxiety symptoms measured by the "being bothered" question could be driven by an external crisis such as 9/11, as opposed to differences in individuals' intrinsic characteristics. For the latter, we provide further evidence in Table 8, and Figures 6 and 7.

Table 8 (columns 1 and 3) shows that for those who were interviewed within three or seven days before or after 9/11, the observed characteristics, except the proportion of male respondents, are similar between two groups: those who had anxiety (in the past seven days)

and those who did not have anxiety (in the past seven days);¹⁵ this result shows that the gender difference in having anxiety symptoms should be controlled for in our estimation model. Next, we use the propensity score of having anxiety symptoms conditional on the observed characteristics to assess the comparability of the two groups. For brevity, detailed explanations of the specification and estimation of the propensity score are provided in the footnotes of Table 8.¹⁶ Here, we follow the suggestion given by Rosenbaum (2010, p. 74) of checking on covariate balance as a diagnostic check of the specification of the propensity score estimation model. Results in columns (2) and (4) of Table 8 indicate the balance of the covariates included in the propensity score estimation, suggesting the correct specification of the estimation model.

In Figures 6 and 7 we check the overlap of the estimated propensity scores between the treated group (i.e., those who had anxiety symptoms) and the control group (i.e., those who did not have anxiety symptoms) for the subsamples including those interviewed within three days (Figure 6) or within seven days (Figure 7) before or after 9/11. Panels A and C of both figures confirm the overlap between the treatment and the control groups. To assess whether the overlap is sufficiently large, which would be consistent with the exogeneity of having anxiety symptoms in the past seven days conditional on the observables, we plot the distributions of another propensity score in Panels B and D of both figures: the conditional probability of being interviewed on or after September 11, 2001 (i.e., the treatment). Because interview dates of Add Health, from interviewees' perspective, are randomly assigned and also without being interrupted by 9/11, the distributions of this second propensity score indeed demonstrate very large overlap between the treated and the control groups. Overall, the comparisons between Panels A and B, and between Panels C and D of both figures indicate that there is a significant overlap in the propensity score of being bothered by things that usually are not bothersome, which provides evidence on the exogeneity of having

¹⁵We provided estimation details in the table footnotes.

¹⁶We use *Stata* to implement the propensity-score matching estimator, with the command “psmatch2” written by Leuven and Sianesi (2003).

anxiety symptoms (in the past seven days) conditional on the observables included in the propensity score estimations.

4.1.2 Results and discussions

Tables 9 and 10 present the estimates of the effect of anxiety likely triggered by 9/11 (i.e., the treatment) on the relative frequency of using marijuana of those interviewed within three days (Table 9) and within seven days (Table 10) before or after 9/11. The estimates obtained using our proposed fuzzy RD estimator suggest that crisis-induced anxiety could raise the likelihood of using marijuana by 4 to 5 percentage points. In Panels A and B of both tables we find whether or not including the dummy variable indicating the post-9/11 period in the first-stage probit estimation of the treatment probability has little impact on the estimated treatment effect. Furthermore, in Panel C of both tables we confirm the results previously obtained from the Monte Carlo experiments: there is little difference in the treatment effect estimates between using a probit and a linear probability model for the first-stage treatment probability estimation.

For comparison purpose we also estimate the treatment effect using OLS, 2SLS and the conventional RD estimator for the 3-day and 7-day samples; the results are reported in Tables 11 and 12, respectively. The OLS estimates (shown in Panel A of both tables) are very similar to the ones given by our proposed estimator (shown in Tables 9 and 10).¹⁷ Note that this result is very different from the pattern suggested by the Monte Carlo experiments, where the treatment effect estimates produced by OLS and the proposed estimator can be quite different when the OLS estimator misspecifies the expected outcome conditional on the observables. One possible explanation for the similarity between the OLS estimates and the ones given by the proposed estimator is that in this particular study most control variables used by the OLS are dummy variables—in this case the linear regression model is close

¹⁷Consistent with the pattern found in the Monte Carlo experiments, the standard errors of the OLS estimator used by this study are slightly smaller than those of the proposed estimator, suggesting a very modest efficiency gain from choosing the OLS over the proposed estimator.

to be exact (i.e., fully saturated), which reduced specification errors significantly, whereas specification errors were purposely preserved in the Monte Carlo experiments.

In Panels B and C of Tables 11 and 12 we report the estimates given by 2SLS and the conventional RD estimator.¹⁸ Following the identification strategy used by Wang and Yang (2013) and comparing with their 2SLS first-stage results, we find that the statistical power of the “9/11” instrument in this study has greatly reduced, likely because of the much narrower time frame used here than the one used by Wang and Yang (2013). As a result of the weak instrument indicated by the small value of the first-stage partial F statistic (shown in Panel B of both tables), the 2SLS has produced very noisy estimates that are likely to be severely biased. One implication from the comparison with 2SLS is that our proposed estimator could have significant advantage over 2SLS in a setting where it is necessary to restrict the sample closely around a cutoff point (such as the date of 9/11 of this study), but in doing so there is no sufficient statistical power of the instruments to be used by 2SLS.

In comparison with our proposed estimator we also find that the conventional RD estimator produces even more noisy estimates than the 2SLS; the results are reported in Panel C of Tables 11 and 12, with estimation details provided in the table footnotes for brevity. One likely reason for the noisy estimates given by the conventional RD estimator is the small sample size. Previous Monte Carlo experiments suggest that to have similar finite-sample performance, the conventional RD estimator would need much larger sample size than the proposed estimator that is valid under the unconfoundedness assumption and when the treatment effect does not vary with the assignment variable.

For our study the requirement for the treatment effect being invariant with respect to the assignment variable means that the effect of anxiety on marijuana use should not vary by the number of days between the interview date and September 11, 2001. In general, this requirement will not be satisfied because there are seasonal variations in the occurrences of anxiety and marijuana use (Kovalenko et al., 2000). As a result, the effect of anxiety on

¹⁸For brevity, we give detailed explanations of the estimation procedures in the two tables and the associated footnotes.

marijuana use can vary by season. However, our study focuses on two subsamples with a narrow time frame—within three or seven days before or after 9/11. With the narrow time frame, it can be reasonable to assume that the effect of anxiety on marijuana use is constant with respect to the number of days between an interview date and September 11, 2001, since the effect probably does not vary from one day to another day with at most two weeks in between.

Next, we demonstrate another advantage of using the proposed estimator over the conventional RD estimator when the unconfoundedness assumption is valid and the treatment effect does not vary with the assignment variable. For our proposed estimator the use of propensity scores highlights the importance of knowing well the process of assigning or selecting a treatment, which is arguably more transparent in an RD setting than other settings where rules of treatment assignment or selection can be completely unobserved. In addition, the use of propensity scores can be followed by the sensitivity analysis described by Rosenbaum (2010, Chapter 3, pp. 76–79) to further assess the validity of the unconfoundedness assumption.

We report the results of the sensitivity analysis in Tables 13 and 14.¹⁹ Implementation details are provided in the table footnotes for brevity.²⁰ In both tables we focus on the upper bound of the p -value for the null hypothesis of no treatment effect: if that upper bound is smaller than conventional significance levels (i.e., 10%, 5% or 1%), then the null hypothesis can be rejected. The sensitivity analysis uses Γ to parameterize the presence of unobservables affecting the treatment assignment: the odds ratio of treatment versus control is assumed to be bounded by Γ as follows:

$$\frac{1}{\Gamma} \leq \frac{\pi_i/(1 - \pi_i)}{\pi_j/(1 - \pi_j)} \leq \Gamma. \quad (16)$$

¹⁹We use *Stata* to implement the sensitivity analysis, with the command “rbounds” written by Gangl (2004).

²⁰Rosenbaum (2010, Chapter 3) also provides an example of the sensitivity analysis for testing the effect of welding fumes on DNA damage (pp. 79–85).

Here, π represents the probability of receiving the treatment conditional on all relevant covariates, either observable or unobservable; subject i has received the treatment, while subject j has not; subjects i and j have the same observed covariates and form a pair for the sensitivity analysis. Empirically, we use a one-to-one nearest neighbor propensity-score matching to form those pairs. One point worth stressing is that π is not the propensity score—the latter is defined in terms of observed covariates, not including unobserved covariates (Rosenbaum, 2010, p. 72). As a result, π cannot be estimated from the data and the bounds described by Γ cannot be tested but assumed. In the absence of unobservables affecting the treatment assignment, Γ equals 1 and the p -value is a single value. If there are unobservables affecting the treatment assignment, regardless of their impacts on the potential outcomes (i.e., the outcome in the presence or absence of the treatment), Γ becomes greater than 1 and the p -value is bounded by the two values calculated based on $1/\Gamma$ and Γ , respectively. In Table 13, results from the sensitivity analysis based on the 3-day sample are not informative, likely due to the small sample size.²¹ In contrast, for the 7-day sample the sensitivity analysis is informative. Specifically, in Table 14 we find that the conclusion of rejecting the null hypothesis of no treatment effect will be overturned, at the 10% significance level, when Γ (described in 16) exceeds 1.26. However, the decision to overturn the conclusion is based on a worst-case scenario. To illustrate this point we consider Γ being equal to 1.27. In this case there is indeed still no evidence proving the existence of unobservables. Furthermore, the upper bound of the p -value (0.101) indicates that the confidence interval for the treatment effect would include zero, with a probability of at least 0.899, if unobservables cause the odds ratio of treatment assignment (described in 16) to differ between two subjects in a treatment-control comparison pair by 1.27.

In summary, the sensitivity analysis does not test for the unconfoundedness assumption, but it provides one way to describe whether conclusions on the presence of a treatment effect can be qualitatively altered for specific deviations from the unconfoundedness assumption.

²¹Rosenbaum (2010) gives detailed discussions on the issue of the power of the sensitivity analysis in Chapter 14 (pp. 265–269).

Simply speaking, a large Γ at which the upper bound of the p -value exceeds conventional significance levels, say Γ^* , could indicate the presence of large effects of the unobservables. If the effects of the unobservables, as suggested by Γ^* , are incredibly large, we would view the study as unbiased from the unobservables. But, it is a judgment call how large Γ has to be to suggest that conclusions on the existence of a treatment effect should not be affected by the presence of unobservables.

To sum up, in this study we assembled a body of empirical evidence suggesting an effect of anxiety likely triggered by an external crisis on marijuana use among young adults. Specifically, a crisis-induced anxiety could increase the probability of using marijuana by 4 to 5 percentage points. Note that these results were obtained using a narrow time frame around 9/11—for a population interviewed within three or seven days before or after 9/11. According to Kovalenko et al.’s study (2000), marijuana use exhibits seasonal variation, with estimated peak occurring in between August and September. Thus, our study focused on the month of September could have examined marijuana use when it was already at the highest level. If this is true, then we could understate the actual effect of a crisis-induced anxiety on marijuana use, which requires off-peak months should be considered as well. In this study we also examined the applicability, performances and advantages of the proposed estimator in comparison with several alternative estimators.

4.2 Evaluation of a School-Based Intervention Program Assigned on the Basis of Test Score Cutoffs

One important implication of our proposed estimator is that the estimator can work in a fuzzy RD setting where the cutoff point is unknown or not exactly known. Without knowing the cutoff point precisely we would incorrectly specify the propensity score model needed for the first stage of the proposed estimator. Our Monte Carlo experiments suggest that specification errors of the propensity score model could have only small impacts on the treatment effect estimates. In this section we reexamine an empirical study previously

done by Chay, McEwan and Urquiola (2005) to further demonstrate the advantage of the proposed estimator over several alternatives in the setting where the cutoff points are not exactly known.

4.2.1 Empirical setting and research design

Chay, McEwan and Urquiola (2005) examined the “900 School Program” (henceforth “P900”) initiated by the Chilean government in 1990. P900 was a nationwide intervention aimed at about 900 low-performing and publicly funded schools. There are four interventions associated with this program: (1) infrastructure improvement, such as building repairs; (2) new instructional materials, including textbooks for students from grades 1 to 4, small classroom libraries, cassette recorders and copy machines; (3) training workshops (focusing on teaching language and mathematics) for school teachers conducted by local supervisors of the Ministry of Education; and (4) after-school tutoring workshops for third and fourth graders who did not perform well enough relative to their grade level. Each workshop was guided by two trained aides recruited from graduates of local secondary schools. Interventions (1) and (2) were the focus of the first two years (1990 and 1991), and P900 was expanded to include (3) and (4) in 1992.

The assignment of P900 was done in two stages to select about 900 schools. First, in 1988 the Ministry of Education administered nationwide achievement tests to the entire population of fourth graders. Officials then calculated each school’s average test scores in language and mathematics, and the average of both. The schools’ overall average scores were then ranked from highest to lowest in each of Chile’s 13 administrative regions. The Ministry determined separate cutoff scores for each region. Schools whose overall average scores fell below their region’s cutoff were eligible for participation. In the second stage, regional teams of officials added two criteria to remove some of the eligible schools: (a) to lower program costs, some very small or inaccessible schools were excluded, in part because there was a parallel program designed to accommodate them; (b) schools were also removed from the pre-selected list if

they had managerial problems, such as misreported enrollment. The regional teams using their own discretion also included certain schools that were ineligible according to the first-stage criteria. More detailed descriptions are provided by Chay, McEwan and Urquiola (2005, table 2, p. 1248).

From the school’s perspective, there was little incentive to forgo participation, because the government covered the full cost. In this empirical setting a school’s ability to self-select into P900 depended on whether it could manipulate its 1988 test score. But, as Chay, McEwan and Urquiola (2005) emphasized in their study, the schools’ 1988 test scores were collected under a different political regime, at a time when P900 was not contemplated. As a result, it is unlikely that schools manipulated their 1988 performance to qualify for P900 which was implemented by a new government in 1990. In Figure 8 we plot the distribution of the 1988 test scores. The distribution appears to be smooth at the region-specific cutoff score, which is consistent with the belief that schools did not manipulate their 1988 test scores to qualify for P900.

Chay, McEwan and Urquiola (2005) provide a body of evidence supporting the validity of using a fuzzy RD design to examine the impact of P900 on the gain score from 1988 to 1992, separately for mathematics and language. For the purpose of our study, we follow one approach used by their study—restricting the samples to include schools with 1988 test scores very close to the cutoff scores—to find schools that can be comparable for the treatment (i.e., receiving P900) effect evaluation. Specifically, we use the same narrow bands that Chay, McEwan and Urquiola (2005) use—within ± 3 points and within ± 5 points of the cutoff.

In Figure 9 we check the overlap of the estimated propensity scores of being selected into P900 between the treated group (i.e., participants) and the control group (i.e., nonparticipants) for those schools with their 1988 test scores within ± 3 points of the cutoff (Panels A and C), and within ± 5 points of the cutoff (Panels B and D). Overall, Figure 9 indicates that there is an overlap in the propensity score of being selected into P900 between the treated

and the control groups, which is consistent with the belief that the assignment of P900 for schools just above or just below the cutoff could be exogenous, conditional on observed school-level characteristics such as test performance, class size and average socioeconomic status (SES) of a school’s students.²²

For the purpose of our study we emphasize two facts about the cutoff score used by the P900 assignment, which have been thoroughly discussed in Chay, McEwan and Urquiola (2005). First, the exact cutoff score is known to the program administrators, but not observed by the researchers. To solve this problem, Chay, McEwan and Urquiola (2005) propose two ways of estimating the cutoff score (pp. 1247–1249), which are referred to as “cutoff definition 1” and “cutoff definition 2” in their study. For comparison purpose and to be consistent with their study, in the following analysis we use the “cutoff definition 2”²³ and also focus on urban schools with at least 15 students in the fourth grade as they did. Second, the cutoff score is specific to the region where the school is located. There are 13 cutoff scores, one for each of Chile’s 13 administrative regions. In the following analysis we demonstrate the consequence of ignoring this fact, when running an OLS regression like the one used by Chay, McEwan and Urquiola (2005).

4.2.2 Results and discussions

We follow and expand the OLS regression model used by Chay, McEwan and Urquiola’s (2005, table 4, p. 1253) and report the estimates in Tables 15 and 16, for the within 3-point and within 5-point samples, respectively. Like their study, we do not further investigate whether the effect of P900 on the 1988–1992 gain score varies with the assignment variable—the distance between the 1988 test score and the cutoff score. In other words, we consider the treatment effect to be constant over the support of the assignment variable (i.e., from -3 to 3, or from -5 to 5).

²²The SES index is scaled between 0 and 100, with higher values indicating higher SES.

²³The “cutoff definition 2” defines the cutoff score as the one that maximizes the proportion of schools being correctly allocated to P900 according to the selection criteria known to the researchers (Chay, McEwan and Urquiola, 2005, p. 1248).

In both tables we control for the 1988 test score (or a polynomial of it) in odd-numbered columns, whereas we control for the normalized 1988 test score—the difference between the 1988 test score and the region-specific cutoff score—in even-numbered columns. We list several striking findings as follows.

First, whether controlling for the 1988 test score or controlling for the normalized 1988 test score makes a surprising difference, and the difference becomes more salient when using a higher order polynomial of the control variable. Specifically, columns (1) and (2) of both tables show that the estimated P900 effects will change when replacing the 1988 test score with the normalized score in the regression model. Furthermore, columns (3) and (4) of both tables indicate that the discrepancy can become larger when a cubic polynomial of the control variable—either the 1988 test score or the normalized value—is included in the regression model. In this empirical setting there can be equally compelling reason for using either control variable. First, as a central point of Chay, McEwan and Urquiola’s study (2005), controlling for the 1988 test score is important for removing the bias from mean reversions of test scores for the P900 program evaluation. Second, also as Chay, McEwan and Urquiola (2005) point out, the empirical setting of P900 permits the use of a fuzzy RD design, which would use the normalized 1988 test score as the assignment variable. Without further information on the program assignment, a researcher will be put into a quandary.

Second, a solution to the quandary just described is the use of region fixed effects. In columns (5) and (6) of both tables we add 12 region dummies (with Region 13 used as the base category) to the regression model when controlling for a linear term of either the 1988 test score or the normalized score. Here, the point estimates of the P900 effect obtained from using either control variable become identical, once the region dummies are included in the OLS regressions. This finding highlights the importance of knowing exactly how cutoff scores are determined, which would allow a researcher to effectively control for the heterogeneity that arises when comparing schools near a region-specific cutoff and those schools may come from different regions.

Third, the solution just described will not be a perfect one in the case of using a polynomial of either control variable. Columns (7) and (8) show that the point estimates of the P900 effect diverge again when a cubic polynomial of either the 1988 test score or the normalized score is used in the OLS regression. This finding highlights the uncertainty of modeling the expectation of the gain score conditional on the assignment variable. The correct specification may involve a function of both the 1988 test score and the normalized score, but the correct specification in practice is hardly ever known to researchers.

Tables 17 and 18 present the estimates (in Panels A and B) given by our proposed estimator. In both tables comparing columns (1) versus (4), (2) versus (5), and (3) versus (6), we find that the estimates of the P900 effect still differ by whether including or excluding the eligibility dummy (equal to 1 for schools scoring below the cutoff and thus eligible for P900) in estimating the probability of receiving P900 in the first stage of the proposed estimator. However, the pattern is different from the one demonstrated by the OLS case in Tables 15 and 16. In the OLS case we find that the discrepancy gets larger when adding higher order polynomials of the control variable (either the 1988 test score or the normalized score) to the gain score regression model. In contrast, for our proposed estimator we do not need to specify the gain score regression model, but the model for the P900 assignment. We find that revising the model for the P900 assignment by adding higher order polynomials of the 1988 test score—the variable that is actually and completely observed by a researcher—could reduce the discrepancy arising from the inclusion or exclusion of the dummy variable for P900 eligibility—the variable that is actually estimated from the data by the researcher, since the cutoff is not exactly known.

Tables 17 and 18 also present the estimates (in Panel C) given by the conventional RD estimator, which is not used by Chay, McEwan and Urquiola’s study (2005). This conventional RD estimator is applied to the full sample, using the method given by Imbens and Kalyanaraman (2012) to select the optimal bandwidth. For brevity the estimation details are provided in the table footnotes. Here, we find the conventional RD estimates to

be quite noisy and even negatively signed—a finding that is completely opposite to what Chay, McEwan and Urquiola (2005) have found, and also a finding that would cast doubt on the performance of the conventional RD estimator in samples that are not sufficiently large.

To sum up, we reexamined Chay, McEwan and Urquiola’s (2005) study on the P900 program evaluation, using our proposed estimator. We find that the 1988–1992 gain score in language, but not in mathematics, of P900 schools is about 0.3 standard deviations²⁴ higher than that of the non-P900 schools, when comparing schools close to the P900 selection cutoff.²⁵ In contrast, Chay, McEwan and Urquiola (2005) find that the P900 program can increase the 1988–1992 gain scores of both mathematics and language by about 0.2 standard deviations. Usually, a gain in test scores between 0.2 and 0.5 standard deviations is viewed as modest, but as insignificant if the gain falls below 0.2 standard deviations. Thus, based on this rule of thumb there are meaningful differences between our findings and the ones given by Chay, McEwan and Urquiola (2005) regarding both the magnitude and the scope of the P900 effect. As demonstrated in our study, it is crucial to have the exact knowledge about the determination of the cutoff scores used for the P900 eligibility. Chay, McEwan and Urquiola (2005) use the estimated cutoff scores and thus estimated P900 eligibility. In contrast, our proposed estimator does not require using the P900 eligibility in the first-stage treatment probability estimation. Furthermore, our proposed estimator does not require specifying the regression model for the gain score, which is required for the OLS regressions used by Chay, McEwan and Urquiola (2005). In this study we also demonstrate that if researchers lack the information that the P900 selection cutoffs are region specific, then they would obtain different estimates of the P900 effect just by controlling for the normalized test score instead of the actual 1988 test score. Such a change in the estimates indeed reflects the presence of uncontrolled heterogeneities (e.g., the region fixed effects) among schools being

²⁴2.1(or 2.3)/7.391 \approx 0.3.

²⁵The calculation follows Chay, McEwan and Urquiola’s (2005) study. Specifically, we use the following two standard deviations for the interpretation of the P900 effect in terms of z -scores: (1) 7.791, the standard deviation of the 1988–1992 gain score in mathematics; and (2) 7.391, the standard deviation of the 1988–1992 gain score in language.

compared near the normalized cutoff. We use this study to illustrate that our proposed estimator could avoid such an omission of the heterogeneities when comparing subjects close to a normalized cutoff, since the proposed estimator requires specifying only the model of the treatment probability, not the model of the outcome.

5 Conclusion

In this study we propose a new estimator for one particular fuzzy RD design, in which we assume that the treatment effect is homogeneous with respect to the assignment variable and the treatment assignment is exogenous conditional on the assignment variable. Under these two assumptions we derive the orthogonality-condition-based estimator, which can be easily implemented using an IV estimation procedure. We use both Monte Carlo experiments and two empirical studies to examine the estimator’s performance in finite samples in comparison with alternative estimators that are also valid in those settings.

The Monte Carlo experiments enable us to gauge the bias in estimating the treatment effect when the data-generating process is fully controlled by the researcher. We find that the bias from misspecifying the treatment probability model required by the first stage of the proposed estimator is minimal, but the bias from misspecifying the regression model of the outcome is large, which can be avoided by the proposed estimator.

We use the first empirical study to highlight a connection between our proposed estimator and the propensity-score matching estimator, both of which require estimating a propensity score in the first stage. In particular, we use the balancing check on the observed covariates to detect specification errors of the propensity score and the sensitivity analysis to assess the magnitude of selection bias due to unobservables, both of which are suggested by Rosenbaum (2010). This empirical study focuses on the effect of anxiety likely triggered by 9/11 (i.e., the treatment) on marijuana use, which is also a case for a homogeneous treatment effect with respect to the assignment variable—the number of days between an interview date

and September 11, 2001. Here, the effect of anxiety on marijuana use can have seasonable variation, as suggested by Kovalenko et al.'s study (2000), but the effect is unlikely to vary from one day to another day for those who were interviewed within three or seven days before or after 9/11.

The second empirical study is a reexamination of a previous study done by Chay, McEwan and Urquiola (2005) on the evaluation of P900, which is a school-based intervention program assigned using region-specific cutoff scores. We use this feature to demonstrate that the proposed estimator can overcome the problem of not knowing, or not knowing exactly, how the multiple cutoff scores are defined in estimating the effect of P900 on test score gains. We find that the OLS used by Chay, McEwan and Urquiola (2005) produces estimates that are sensitive to how researchers control for those multiple cutoffs in the regression model of the gain score. In this study we also find that the local linear regression estimator proposed by Hahn, Todd and Van der Klaauw (2001) does not perform well in the estimation sample used by Chay, McEwan and Urquiola (2005).

Our proposed estimator hinges upon the validity of the two assumptions aforementioned. As shown by Angrist and Rokkanen (2012), when empirically justifiable, the assumption of a homogeneous treatment effect with respect to the assignment variable allows for identifying a treatment effect away from the cutoff. Also as Angrist and Rokkanen (2012) point out, the treatment effect away from the cutoff can have a broader policy implication than the effect only at the cutoff. The assumption of the exogeneity of a treatment conditional on the assignment variable (i.e., the unconfoundedness) emphasizes a randomization that is unique to RD designs and that is a result of imperfect control over the assignment variable (Lee, 2008). In both of our empirical studies, the cutoffs (i.e., the 9/11 date or region-specific test score cutoffs) are unexpected and determined exogenously, which can preclude the possibility that an individual manipulates the assignment variable to qualify for the treatment. However, for fuzzy RD designs the possibility of nonrandom treatment assignment may not be avoided when we focus on the population near the cutoff: the less-than-100% treatment

take-up rate of those who are near the cutoff and eligible for the treatment actually can be the result of selection based on potential gain from the treatment. In this case our proposed estimator will fail, and identification of a treatment effect only at the cutoff remains a distinct possibility, although the generalizability of that treatment effect can be questioned by certain policy-oriented empirical studies.

References

- Angrist, J. and M. Rokkanen (2012). “Wanna Get Away? RD Identification Away from the Cutoff.” *NBER Working Paper* No. 18662.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York, NY: Cambridge University Press.
- Chay, K. Y., P. J. McEwan and M. Urquiola (2005). “The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools.” *American Economic Review* 95(4): 1237–1258.
- Cook, T. D. (2008). “‘Waiting for Life to Arrive’: A History of the Regression-Discontinuity Design in Psychology, Statistics and Economics.” *Journal of Econometrics* 142(2): 636–654.
- Gangl, M. (2004). “RBOUNDS: Stata Module to Perform Rosenbaum Sensitivity Analysis for Average Treatment Effects on the Treated.” <http://econpapers.repec.org/software/bocbocode/s438301.htm> (accessed July 1, 2013).
- Hahn, J., P. Todd and W. Van der Klaauw (2001). “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” *Econometrica* 69(1): 201–209.
- Holland, P. W. (1986). “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396): 945–960.

- Imbens, G. W. and K. Kalyanaraman (2012). “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *The Review of Economic Studies* 79(3): 933–959.
- Imbens, G. W. and T. Lemieux (2008). “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics* 142(2): 615–635.
- Imbens, G. W. and J. M. Wooldridge (2009). “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature* 47(1): 5–86.
- Khoury, L., Y. L. Tang, B. Bradley, J. F. Cubells and K. Ressler (2010). “Substance Use, Childhood Traumatic Experience, and Posttraumatic Stress Disorder in an Urban Civilian Population.” *Depression and Anxiety* 27(12): 1077–1086.
- Kovalenko, P. A., C. W. Hoven, J. Wicks, R. E. Moore, D. J. Mandell and H. Liu (2000). “Seasonal Variations in Internalizing, Externalizing, and Substance Use Disorders in Youth.” *Psychiatry Research* 94(2): 103–119.
- Lee, D. S. (2008). “Randomized Experiments from Non-Random Selection in U.S. House Elections.” *Journal of Econometrics* 142(2): 675–697.
- Lee, D. S. and T. Lemieux (2010). “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48(2): 281–355.
- Leuven, E. and B. Sianesi (2003). “PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing” (version 4.0.6 May 17, 2012). <http://ideas.repec.org/c/boc/bocode/s432001.html> (accessed July 1, 2013).
- Marshall, R. D. and S. Galea (2004). “Science for the Community: Assessing Mental Health after 9/11.” *Journal of Clinical Psychiatry* 65(Supplement 1): 37–43.
- Nichols, A. (2011). “rd 2.0: Revised Stata Module for Regression Discontinuity Estimation.” <http://ideas.repec.org/c/boc/bocode/s456888.html> (accessed July 1, 2013).

- Robinson, P. M. (1988). “Root-N-Consistent Semiparametric Regression.” *Econometrica* 56(4): 931–954.
- Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York, NY: Springer.
- Rosenbaum, P. R. and D. B. Rubin (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70(1): 41–55.
- Rubin, D. B. (1974). “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology* 66(5): 688–701.
- Thistlethwaite, D. L. and D. T. Campbell (1960). “Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment.” *Journal of Educational Psychology* 51(6): 309–317.
- Thompson, K. M. (2001). “Marijuana Use among Adolescents: Trends, Patterns, and Influences.” *Minerva Pediatrica* August 53(4): 313–324.
- Trochim, W. M. K. (1984). *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Beverly Hills, CA: Sage Publications.
- Verardi, V. and N. Debarsy (2012). “Robinson’s Square Root of N Consistent Semiparametric Regression Estimator in Stata.” *The Stata Journal* 12(4): 726–735.
- Wang, Y. and M. Yang (2013). “Crisis-Induced Depression, Physical Activity and Dietary Intake among Young Adults: Evidence from the 9/11 Terrorist Attacks.” *Economics & Human Biology* 11(2): 206–220.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data (Second Edition)*. Cambridge, MA: MIT Press.

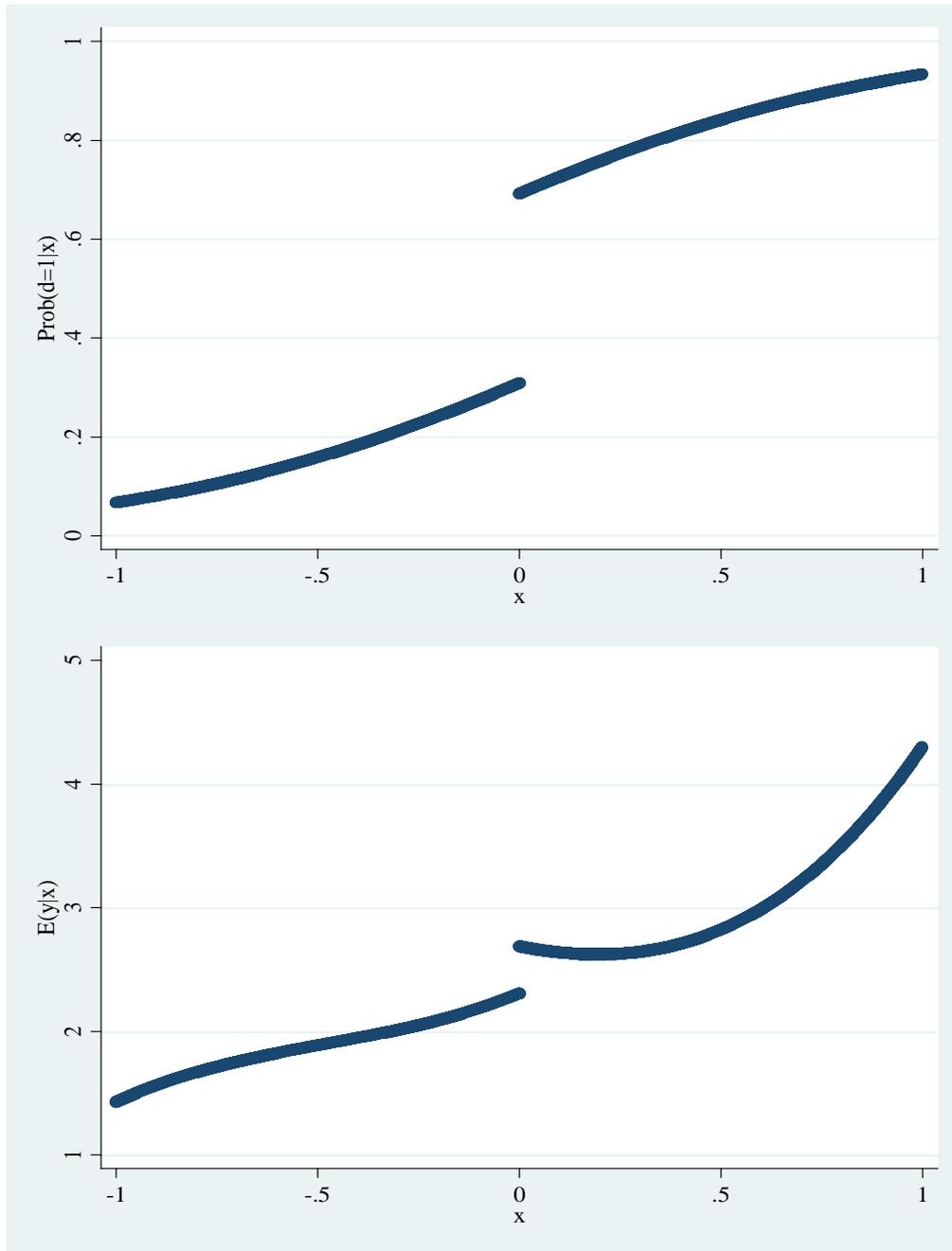


Figure 1: Simulated Outcome and Treatment Probability with Discontinuities at Zero

Notes: Both of the discontinuities in $E(y|x)$ and in $\text{Prob}(d=1|x)$ are designed to be equal to 0.383.

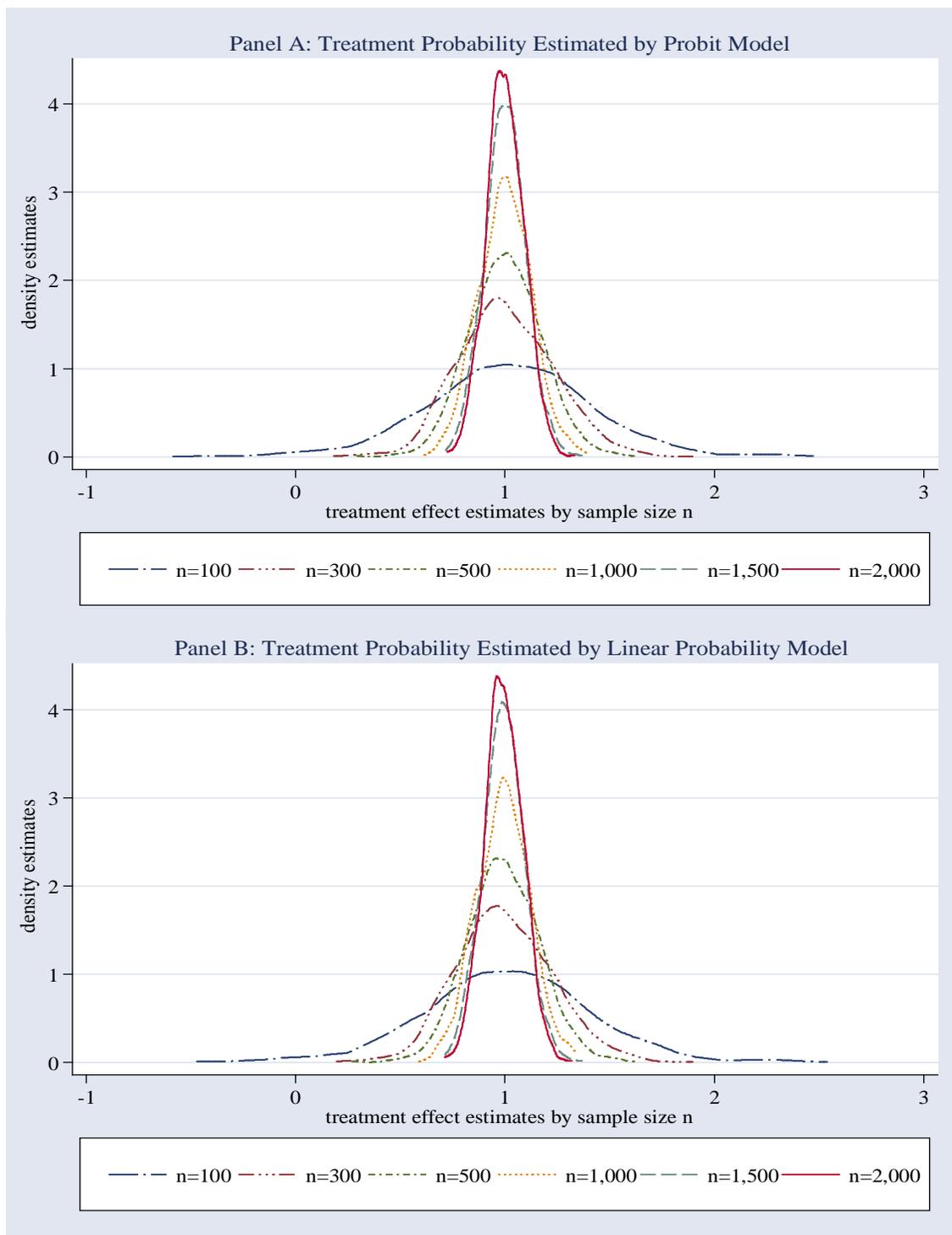


Figure 2: Estimated density based on Epanechnikov kernel

Notes: Simulations use 1,000 repetitions.

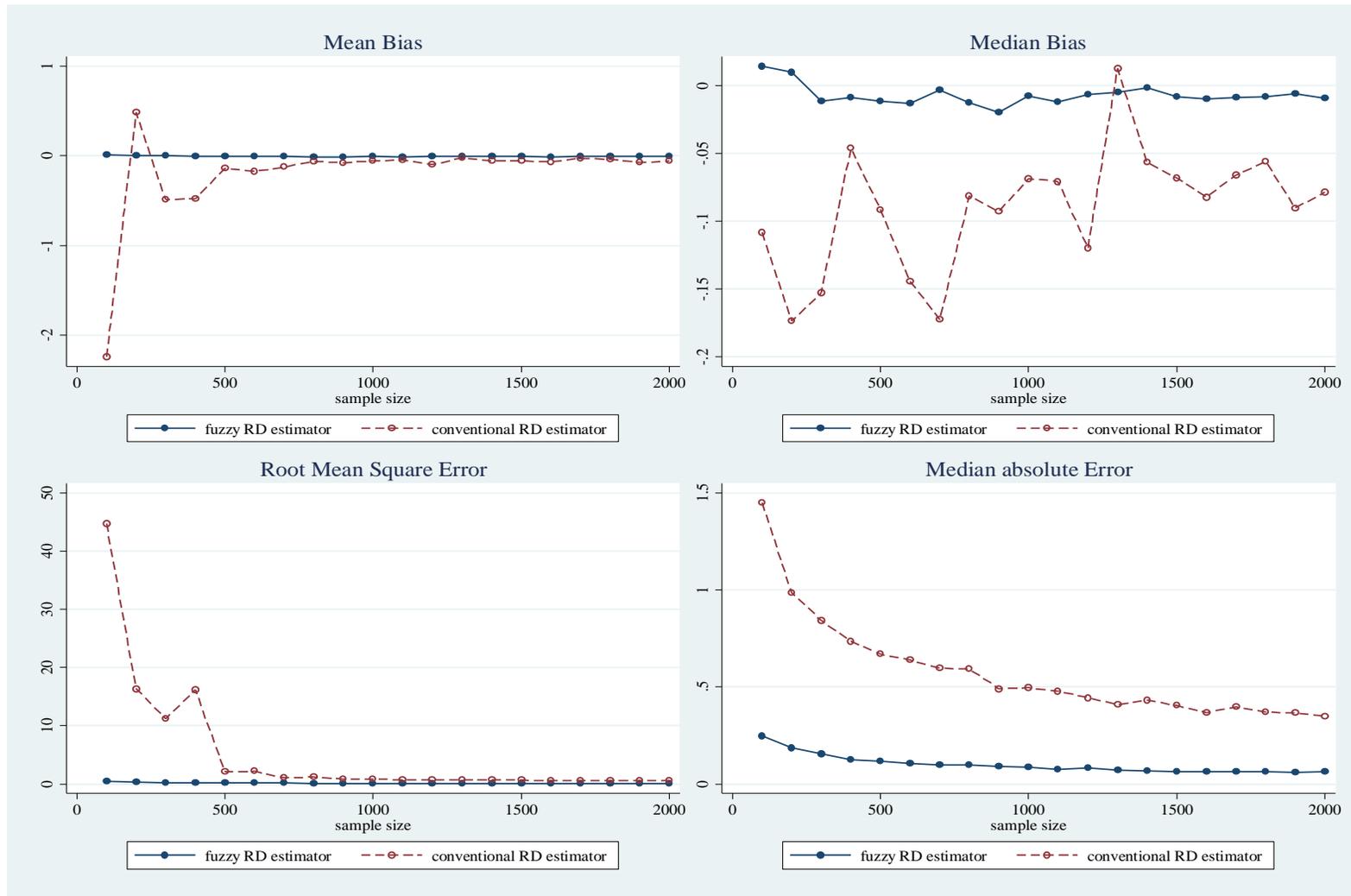


Figure 3: Comparisons of Finite Sample Properties

Notes: The treatment probability of the fuzzy RD estimator proposed by this paper is estimated by a linear probability model. Stata code for implementing a conventional RD estimator is “rd” written by Nichols (2011), which uses local linear regression models on both sides of the cutoff, with a triangle kernel, and the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). Simulations use 1,000 repetitions.

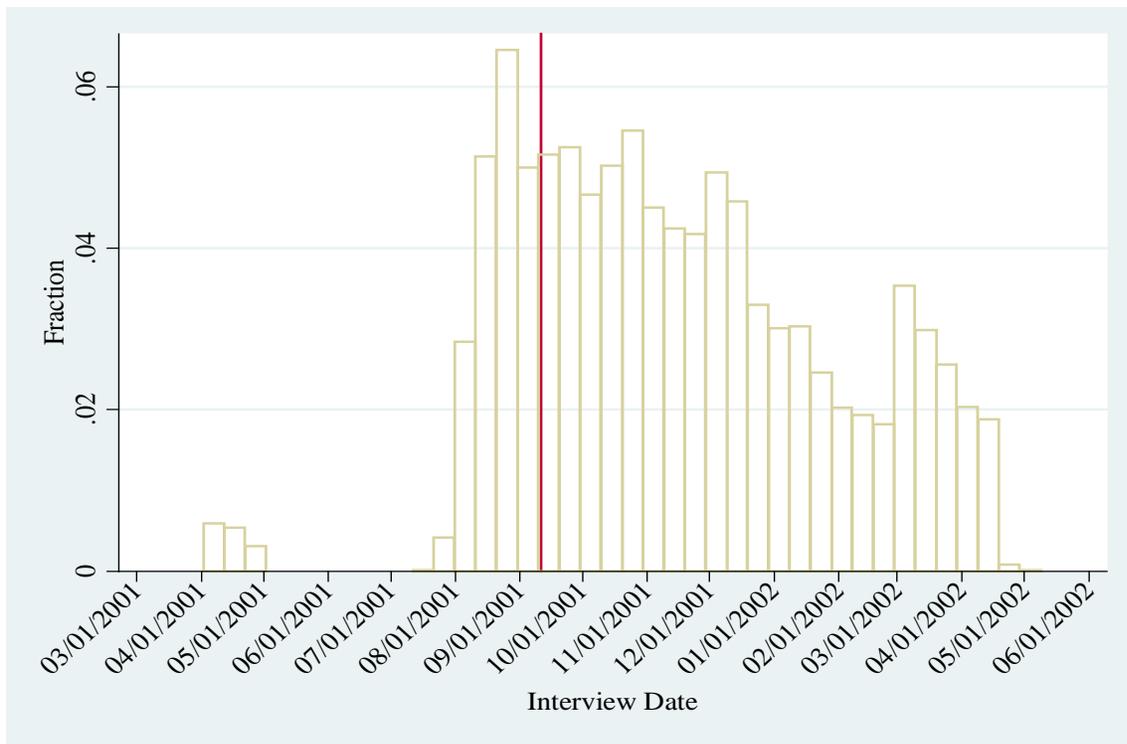


Figure 4: Interview Date Distribution

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. The vertical line indicates September 11, 2001. This figure is a replica of the figure 1 of Wang and Yang (2013).



Figure 5: Responses of Individuals Interviewed in Waves II and III of Add Health

Notes: Data are from Waves II and III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Depicted in the panels are the series of responses to the question (“during the past seven days, you were bothered by things that usually don’t bother you”) with the scale 0–3 (0: never or rarely; 1: sometimes; 2: a lot of the time; 3: most of the time or all of the time) averaged across respondents who were interviewed on the same date. The vertical line indicates September 11, 2001.

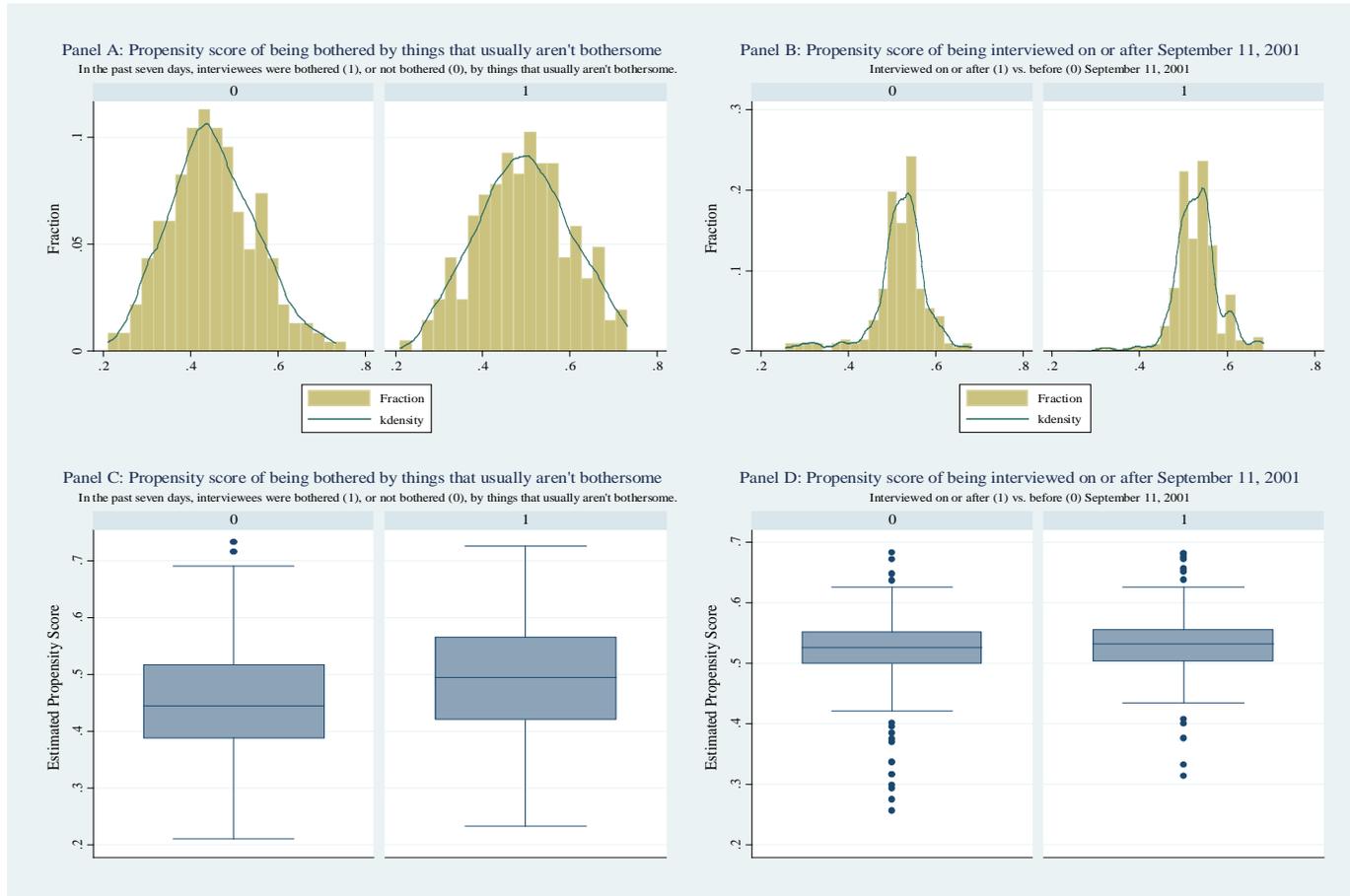


Figure 6: Checks on the Overlap between the Distributions of Propensity Scores of Those Interviewed Within 3 Days Before or After 9/11

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. The propensity score of being bothered by things that usually are not bothersome (1/0) is estimated by a probit model including the following covariates: a dummy variable of being interviewed before (0) or not before (1) 09/11/2001; a cubic polynomial of the number of days between 09/11/2001 and the interview date; age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health. The propensity score of being interviewed on or after 09/11/2001 (1/0) is estimated by a probit model including the following covariates: age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health.

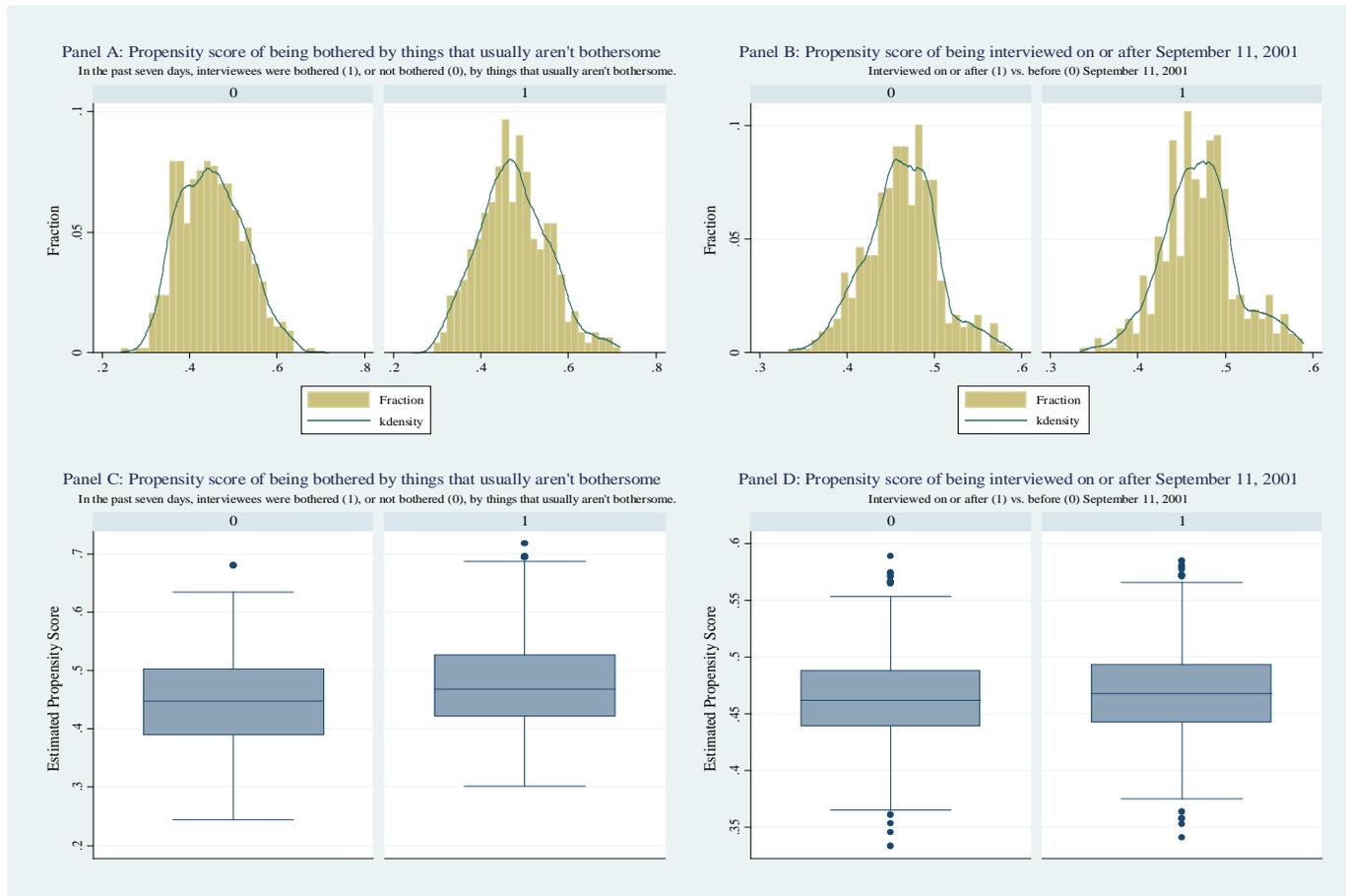


Figure 7: Checks on the Overlap between the Distributions of Propensity Scores of Those Interviewed Within 7 Days Before or After 9/11

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. The propensity score of being bothered by things that usually are not bothersome (1/0) is estimated by a probit model including the following covariates: a dummy variable of being interviewed before (0) or not before (1) 09/11/2001; a cubic polynomial of the number of days between 09/11/2001 and the interview date; age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health. The propensity score of being interviewed on or after 09/11/2001 (1/0) is estimated by a probit model including the following covariates: age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health.

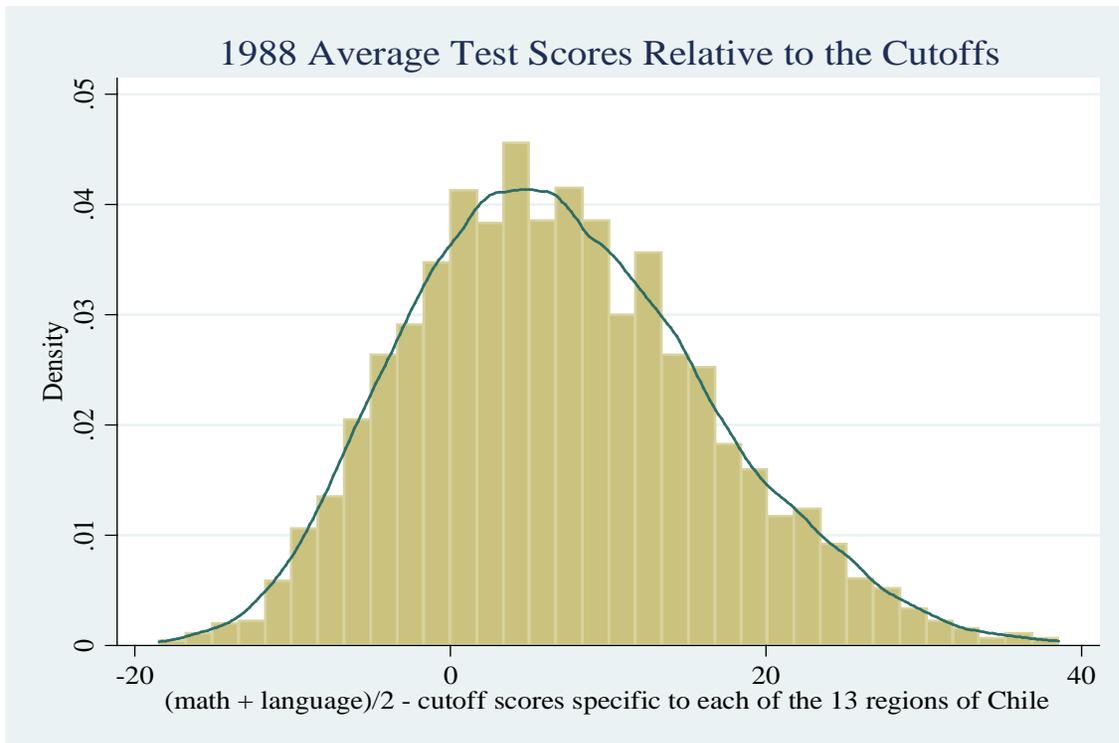


Figure 8: Distribution of the Normalized 1988 Average Test Scores

Notes: The sample includes urban schools with at least 15 students in the fourth grade in 1988. The density is estimated using the Epanechnikov kernel with the bandwidth set to be 1.677 (the one that minimizes the mean integrated squared error if the data were Gaussian and a Gaussian kernel were used).

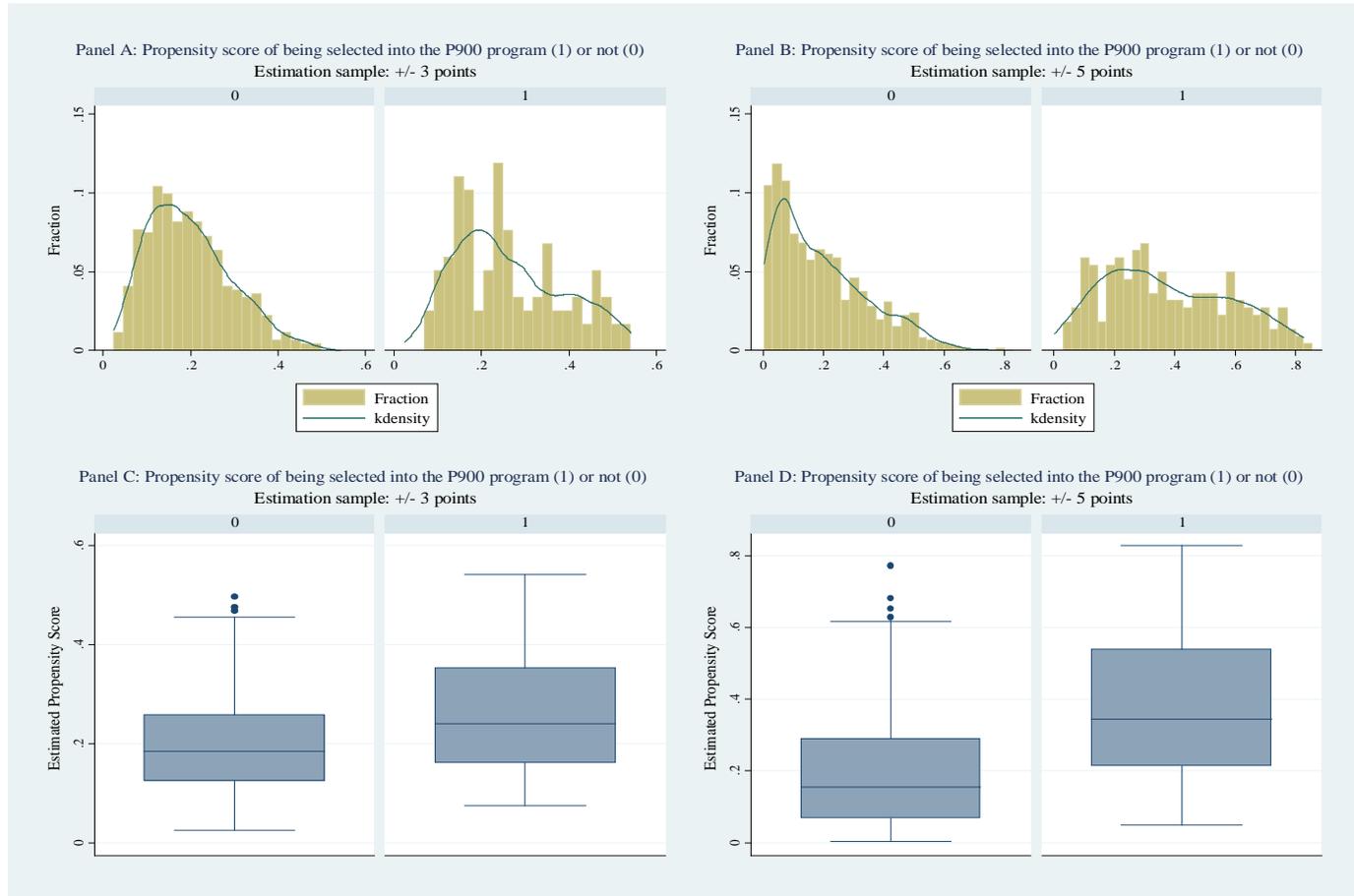


Figure 9: Checks on the Overlap between the Distributions of Propensity Scores of Schools with Test Scores near the Cutoff

Notes: The sample includes urban schools with at least 15 students in the fourth grade in 1988. The propensity score of being selected into the P900 program (1/0) is estimated by a probit model including the following covariates: average test score (over math and language) of a school in 1988, the number of test takers of the school in 1988, and the average socioeconomic status index of students of the school in 1990.

Table 1: Monte Carlo Results — Mean Bias

Sample size	(1)	(2)	(3)	(4)	(5)	(6)
	Fuzzy RD estimator treatment probability estimated by Probit	Fuzzy RD estimator treatment probability estimated by LPM	OLS estimator	Robinson's estimator	Conventional RD estimator	Optimal bandwidth used for the conventional RD estimator
100	0.021	0.012	-0.063	0.021	-2.241	0.659
200	0.011	0.004	-0.069	0.011	0.481	0.660
300	0.006	-0.002	-0.075	0.005	-0.492	0.645
400	0.005	-0.003	-0.073	0.007	-0.477	0.627
500	0.001	-0.007	-0.076	0.003	-0.143	0.637
600	-0.002	-0.009	-0.081	-0.001	-0.175	0.618
700	0.001	-0.006	-0.078	0.002	-0.125	0.616
800	-0.006	-0.014	-0.086	-0.004	-0.064	0.601
900	-0.007	-0.014	-0.086	-0.004	-0.077	0.598
1,000	-0.001	-0.009	-0.079	0.001	-0.056	0.596
1,100	-0.003	-0.011	-0.083	-0.001	-0.047	0.589
1,200	-0.001	-0.008	-0.080	0.000	-0.101	0.588
1,300	0.001	-0.007	-0.076	0.003	-0.020	0.581
1,400	0.001	-0.007	-0.079	0.002	-0.059	0.580
1,500	0.000	-0.007	-0.079	0.001	-0.058	0.573
1,600	-0.004	-0.012	-0.082	-0.002	-0.069	0.578
1,700	-0.003	-0.010	-0.079	-0.001	-0.028	0.565
1,800	-0.001	-0.009	-0.080	0.001	-0.042	0.560
1,900	0.004	-0.003	-0.074	0.006	-0.076	0.561
2,000	0.001	-0.006	-0.077	0.003	-0.059	0.562

Notes: Simulations use 1,000 repetitions. The true value of the treatment effect is 1. Columns (1) and (2) are based on the fuzzy RD estimator proposed by this paper with the treatment probability estimated by a probit and a linear probability model (LPM), respectively. Column (3) uses an OLS estimator of regressing y on an intercept, the treatment variable (d) and the assignment variable (x). Column (4) uses Robinson's (1988) estimator with Stata code written by Verardi and Debarsy (2012). Column (5) is based on a conventional RD estimator with Stata code written by Nichols (2011). Column (6) reports the optimal bandwidth, proposed by Imbens and Kalyanaraman (2012), used for the conventional RD estimator.

Table 2: Monte Carlo Results — Median Bias

Sample size	Fuzzy RD estimator treatment probability estimated by		(3)	(4)	(5)	(6)
	Probit	LPM	OLS estimator	Robinson's estimator	Conventional RD estimator	Optimal bandwidth used for the conventional RD estimator
100	0.019	0.014	-0.055	0.023	-0.108	0.659
200	0.015	0.010	-0.067	0.017	-0.174	0.660
300	-0.007	-0.012	-0.081	-0.002	-0.153	0.645
400	-0.002	-0.009	-0.076	0.007	-0.046	0.627
500	-0.001	-0.012	-0.077	-0.001	-0.092	0.637
600	-0.003	-0.013	-0.077	0.002	-0.144	0.618
700	0.002	-0.003	-0.075	0.003	-0.173	0.616
800	-0.008	-0.013	-0.088	-0.006	-0.082	0.601
900	-0.008	-0.020	-0.087	-0.003	-0.093	0.598
1,000	0.001	-0.008	-0.078	0.004	-0.069	0.596
1,100	-0.003	-0.012	-0.084	-0.004	-0.071	0.589
1,200	0.000	-0.007	-0.080	-0.001	-0.120	0.588
1,300	0.002	-0.005	-0.073	0.005	0.012	0.581
1,400	0.005	-0.002	-0.077	0.008	-0.057	0.580
1,500	0.000	-0.009	-0.077	0.002	-0.068	0.573
1,600	0.000	-0.010	-0.086	0.002	-0.082	0.578
1,700	-0.001	-0.009	-0.079	0.002	-0.066	0.565
1,800	-0.003	-0.008	-0.080	0.002	-0.056	0.560
1,900	0.001	-0.006	-0.077	0.003	-0.090	0.561
2,000	-0.002	-0.009	-0.077	0.001	-0.079	0.562

Notes: Simulations use 1,000 repetitions. The true value of the treatment effect is 1. Columns (1) and (2) are based on the fuzzy RD estimator proposed by this paper with the treatment probability estimated by a probit and a linear probability model (LPM), respectively. Column (3) uses an OLS estimator of regressing y on an intercept, the treatment variable (d) and the assignment variable (x). Column (4) uses Robinson's (1988) estimator with Stata code written by Verardi and Debarsy (2012). Column (5) is based on a conventional RD estimator with Stata code written by Nichols (2011). Column (6) reports the optimal bandwidth, proposed by Imbens and Kalyanaraman (2012), used for the conventional RD estimator.

Table 3: Monte Carlo Results — Root Mean Square Error

Sample size	(1)	(2)	(3)	(4)	(5)	(6)
	Fuzzy RD estimator treatment probability estimated by		OLS estimator	Robinson's estimator	Conventional RD estimator	Optimal bandwidth used for the conventional RD estimator
	Probit	LPM				
100	0.403	0.403	0.388	0.403	44.705	0.659
200	0.277	0.277	0.274	0.272	16.363	0.660
300	0.229	0.229	0.234	0.227	11.196	0.645
400	0.189	0.189	0.196	0.187	16.201	0.627
500	0.172	0.172	0.181	0.170	2.069	0.637
600	0.159	0.160	0.175	0.158	2.195	0.618
700	0.147	0.147	0.160	0.147	1.056	0.616
800	0.137	0.138	0.159	0.136	1.214	0.601
900	0.133	0.133	0.153	0.130	0.816	0.598
1,000	0.126	0.126	0.144	0.124	0.840	0.596
1,100	0.114	0.115	0.139	0.113	0.740	0.589
1,200	0.117	0.118	0.139	0.115	0.726	0.588
1,300	0.107	0.107	0.128	0.106	0.771	0.581
1,400	0.101	0.102	0.126	0.100	0.679	0.580
1,500	0.098	0.099	0.122	0.097	0.656	0.573
1,600	0.096	0.097	0.123	0.095	0.607	0.578
1,700	0.094	0.094	0.120	0.093	0.623	0.565
1,800	0.093	0.094	0.121	0.092	0.607	0.560
1,900	0.085	0.086	0.111	0.085	0.568	0.561
2,000	0.090	0.090	0.116	0.089	0.558	0.562

Notes: Simulations use 1,000 repetitions. The true value of the treatment effect is 1. Columns (1) and (2) are based on the fuzzy RD estimator proposed by this paper with the treatment probability estimated by a probit and a linear probability model (LPM), respectively. Column (3) uses an OLS estimator of regressing y on an intercept, the treatment variable (d) and the assignment variable (x). Column (4) uses Robinson's (1988) estimator with Stata code written by Verardi and Debarsy (2012). Column (5) is based on a conventional RD estimator with Stata code written by Nichols (2011). Column (6) reports the optimal bandwidth, proposed by Imbens and Kalyanaraman (2012), used for the conventional RD estimator.

Table 4: Monte Carlo Results — Median Absolute Error

Sample size	(1)	(2)	(3)	(4)	(5)	(6)
	Fuzzy RD estimator treatment probability estimated by		OLS estimator	Robinson's estimator	Conventional RD estimator	Optimal bandwidth used for the conventional RD estimator
	Probit	LPM				
100	0.249	0.249	0.244	0.249	1.451	0.659
200	0.186	0.185	0.181	0.180	0.986	0.660
300	0.153	0.155	0.166	0.150	0.842	0.645
400	0.131	0.126	0.134	0.128	0.735	0.627
500	0.116	0.116	0.125	0.113	0.668	0.637
600	0.105	0.105	0.120	0.104	0.638	0.618
700	0.098	0.098	0.105	0.097	0.597	0.616
800	0.095	0.097	0.116	0.097	0.592	0.601
900	0.089	0.091	0.108	0.089	0.492	0.598
1,000	0.087	0.087	0.094	0.086	0.496	0.596
1,100	0.075	0.075	0.098	0.071	0.479	0.589
1,200	0.080	0.083	0.099	0.079	0.445	0.588
1,300	0.074	0.071	0.088	0.074	0.409	0.581
1,400	0.070	0.068	0.090	0.068	0.431	0.580
1,500	0.064	0.065	0.086	0.063	0.404	0.573
1,600	0.066	0.065	0.092	0.065	0.367	0.578
1,700	0.064	0.064	0.088	0.063	0.397	0.565
1,800	0.063	0.063	0.088	0.062	0.372	0.560
1,900	0.058	0.059	0.083	0.059	0.365	0.561
2,000	0.061	0.062	0.085	0.058	0.349	0.562

Notes: Simulations use 1,000 repetitions. The true value of the treatment effect is 1. Columns (1) and (2) are based on the fuzzy RD estimator proposed by this paper with the treatment probability estimated by a probit and a linear probability model (LPM), respectively. Column (3) uses an OLS estimator of regressing y on an intercept, the treatment variable (d) and the assignment variable (x). Column (4) uses Robinson's (1988) estimator with Stata code written by Verardi and Debarsy (2012). Column (5) is based on a conventional RD estimator with Stata code written by Nichols (2011). Column (6) reports the optimal bandwidth, proposed by Imbens and Kalyanaraman (2012), used for the conventional RD estimator.

Table 5: Monte Carlo Results — Standard Deviation

Sample size	(1)	(2)	(3)	(4)	(5)	(6)
	Fuzzy RD estimator treatment probability estimated by Probit	Fuzzy RD estimator treatment probability estimated by LPM	OLS estimator	Robinson's estimator	Conventional RD estimator	Optimal bandwidth used for the conventional RD estimator
100	0.403	0.403	0.383	0.403	44.671	0.659
200	0.277	0.277	0.266	0.272	16.364	0.660
300	0.229	0.229	0.222	0.227	11.191	0.645
400	0.189	0.189	0.182	0.187	16.202	0.627
500	0.172	0.172	0.164	0.170	2.065	0.637
600	0.159	0.160	0.155	0.158	2.189	0.618
700	0.147	0.147	0.141	0.147	1.049	0.616
800	0.137	0.138	0.134	0.136	1.212	0.601
900	0.133	0.133	0.127	0.130	0.812	0.598
1,000	0.126	0.126	0.121	0.124	0.839	0.596
1,100	0.114	0.115	0.112	0.113	0.739	0.589
1,200	0.117	0.118	0.114	0.115	0.720	0.588
1,300	0.107	0.107	0.103	0.106	0.771	0.581
1,400	0.101	0.102	0.099	0.100	0.677	0.580
1,500	0.098	0.098	0.093	0.097	0.654	0.573
1,600	0.096	0.096	0.092	0.095	0.603	0.578
1,700	0.094	0.094	0.091	0.093	0.623	0.565
1,800	0.093	0.094	0.091	0.092	0.606	0.560
1,900	0.085	0.086	0.083	0.085	0.563	0.561
2,000	0.090	0.090	0.087	0.089	0.555	0.562

Notes: Simulations use 1,000 repetitions. The true value of the treatment effect is 1. Columns (1) and (2) are based on the fuzzy RD estimator proposed by this paper with the treatment probability estimated by a probit and a linear probability model (LPM), respectively. Column (3) uses an OLS estimator of regressing y on an intercept, the treatment variable (d) and the assignment variable (x). Column (4) uses Robinson's (1988) estimator with Stata code written by Verardi and Debarsy (2012). Column (5) is based on a conventional RD estimator with Stata code written by Nichols (2011). Column (6) reports the optimal bandwidth, proposed by Imbens and Kalyanaraman (2012), used for the conventional RD estimator.

Table 6: Summary Statistics

Variables	3 days before/after 9/11 N = 435		7 days before/after 9/11 N = 1,008		30 days before/after 9/11 N = 4,039		Full sample N = 12,023	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
9/11 (1/0)	0.524	0.500	0.465	0.499	0.468	0.499	0.769	0.422
Days	-0.189	2.006	-0.582	4.482	-1.304	17.526	63.465	74.795
Bothered (1/0)	0.471	0.500	0.462	0.499	0.442	0.497	0.434	0.496
Number of times using marijuana in the past 30 days	2.605	8.355	2.935	9.308	3.169	10.236	3.151	10.588
Marijuana use relative frequency	0.087	0.279	0.098	0.310	0.106	0.341	0.105	0.353
Age	21.669	1.693	21.715	1.648	21.707	1.703	22.005	1.762
Male (1/0)	0.448	0.498	0.448	0.498	0.451	0.498	0.482	0.500
White (1/0)	0.738	0.440	0.729	0.445	0.725	0.447	0.704	0.456
Black (1/0)	0.182	0.386	0.188	0.391	0.199	0.400	0.216	0.411
Native (1/0)	0.085	0.279	0.073	0.261	0.077	0.267	0.078	0.268
Asian (1/0)	0.101	0.302	0.103	0.304	0.088	0.284	0.087	0.281
Hispanic (1/0)	0.161	0.368	0.172	0.377	0.173	0.378	0.173	0.378
Personal income	13.248	15.638	13.424	13.806	12.954	14.400	14.075	16.264
Years of education	13.347	1.936	13.279	1.958	13.236	1.909	13.231	1.964

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. The variable “9/11” is a dummy variable equal to 1 (or 0) if an individual was interviewed on or after (or before) September 11, 2001. The variable “days” represents the number of days between 09/11/2001 and the interview date of Wave III. The variable “bothered” is a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The variable “marijuana use relative frequency” is calculated by the number of times using marijuana in the past 30 days divided by 30. The variable “age” is measured in years. The variable “personal income” is measured by the total personal income before tax during 2000–2001 (in \$1,000) at Wave III. The variable “years of education” measures the years of education as of Wave III of Add Health.

Table 7: 9/11 and Anxiety

Outcome (1/0): a respondent reported that during the past seven days he or she was bothered (1), or not bothered (0), by things that usually are not bothersome.

	(1)	(2)	(3)	(4)
<i>Panel A: 3 days before and after 9/11 (N = 435)</i>				
9/11 (1/0)	0.079* (0.048)	-0.141 (0.137)	-0.136 (0.142)	-0.186 (0.561)
<i>Panel B: 7 days before and after 9/11 (N = 1,008)</i>				
9/11 (1/0)	0.092*** (0.031)	-0.015 (0.087)	-0.025 (0.087)	-0.168 (0.174)
<i>Panel C: 30 days before and after 9/11 (N = 4,039)</i>				
9/11 (1/0)	0.037** (0.016)	0.122*** (0.043)	0.121*** (0.042)	0.116*** (0.043)
<i>Panel D: full sample (N = 12,023)</i>				
9/11 (1/0)	0.009 (0.011)	0.049*** (0.018)	0.053*** (0.018)	0.054*** (0.018)
Cubic polynomial of days	No	Yes	Yes	Yes
Individual characteristics	No	No	Yes	Yes
Day of week fixed effect	No	No	No	Yes

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Descriptions and definitions of variables are provided in the notes of Table 6. The variable “9/11” is a dummy variable equal to 1 (or 0) if an individual was interviewed on or after (or before) September 11, 2001. Individual characteristics controlled for are age, age squared, gender, race and ethnicity (White, Black, Native, Asian, or Hispanic). The day of week fixed effect is applied to Monday through Saturday. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 8: Differences in the Means between Those Who Had Anxiety and Those Who Did Not Have Anxiety in the Past Seven Days

Variables	3 days before and after 9/11 (<i>N</i> = 435)		7 days before and after 9/11 (<i>N</i> = 1,008)	
	(1)	(2)	(3)	(4)
Age	-0.075 (0.162)	[-0.45] {0.654}	-0.073 (0.104)	[-0.35] {0.724}
Male (1/0)	-0.101** (0.048)	[0.44] {0.664}	-0.076** (0.031)	[0.21] {0.838}
White (1/0)	0.007 (0.042)	[0.37] {0.713}	-0.027 (0.028)	[0.38] {0.703}
Black (1/0)	-0.021 (0.037)	[0.14] {0.888}	0.002 (0.025)	[-0.17] {0.863}
Native (1/0)	-0.004 (0.027)	[0.38] {0.703}	0.011 (0.017)	[-0.26] {0.795}
Asian (1/0)	0.012 (0.029)	[-0.19] {0.851}	0.032 (0.019)	[-0.93] {0.354}
Hispanic (1/0)	-0.046 (0.035)	[-0.15] {0.882}	-0.004 (0.024)	[-0.27] {0.788}
Personal income	-0.265 (1.511)	[-0.23] {0.817}	-0.151 (0.882)	[-0.02] {0.984}
Years of education	-0.094 (0.187)	[-0.41] {0.683}	-0.147 (0.123)	[0.05] {0.959}

Notes 1: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Descriptions and definitions of variables are provided in the notes of Table 6. We use the following procedure to obtain the differences in the means between those who reported to have been bothered by things (in the past seven days) that usually do not bother them and those who reported not to have been bothered by things (in the past seven days) that usually do not bother them. First, we generate a “bothered” dummy variable equal to 1 (or 0) if a respondent was (or was not) bothered by things (in the past seven days) that usually do not bother him or her. Next, we regress each variable listed in the table on an intercept and the “bothered” dummy variable. We reported those coefficient estimates for the “bothered” dummy variable in columns (1) and (3). Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Notes 2: Columns (2) and (4) report the results of balancing checks on the observable characteristics (listed in the table) based on estimated propensity scores. First, the propensity score of being bothered by things that usually are not bothersome (1/0) is estimated by a probit model including the following covariates: a dummy variable of being interviewed before (0) or not before (1) 09/11/2001; a cubic polynomial of the number of days between 09/11/2001 and the interview date; age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health. Second, we use a one-to-one nearest neighbor matching with the caliper set to be less than 0.1 to form two matched samples (“bothered” vs. “not bothered”). Third, we conduct t-tests for the equality of means in the two matched samples. T-tests are based on weighted regressions of each of the observable characteristics (listed in the table) on the “bothered” (1/0) dummy variable; the weights are the frequencies with which the observation is used as a match for the nearest neighbor matching. The t statistics are reported in [square brackets] of columns (2) and (4); the associated p-values for the two-tailed test are reported in {curly brackets} of columns (2) and (4).

Table 9: Fuzzy RD Estimates of the Effect of Anxiety on Marijuana Use for Those Interviewed Within 3 Days Before or After 9/11

Outcome: marijuana use relative frequency (i.e., the number of times using marijuana in the past 30 days divided by 30) with sample size $N = 435$

	(1)	(2)	(3)
<i>Panel A: Estimation of the probability of being bothered (1/0) uses a probit model including the "9/11" dummy variable and the following covariates:</i>			
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
Bothered (1/0)	0.049*	0.052*	0.052*
	(0.027)	(0.028)	(0.028)
<i>Panel B: Estimation of the probability of being bothered (1/0) uses a probit model including the following covariates but excluding the "9/11" dummy variable:</i>			
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
Bothered (1/0)	0.049*	0.053*	0.052*
	(0.027)	(0.029)	(0.028)
<i>Panel C: Estimation of the probability of being bothered (1/0) uses a linear probability model including the "9/11" dummy variable and the following covariates:</i>			
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
Bothered (1/0)	0.049*	0.052*	0.052*
	(0.027)	(0.028)	(0.028)

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Descriptions and definitions of variables are provided in the notes of Table 6. The variable “bothered” is a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The variable “9/11” is a dummy variable equal to 1 (or 0) if an individual was interviewed on or after (or before) September 11, 2001. Individual characteristics controlled for are age, age squared, gender, race and ethnicity (White, Black, Native, Asian, or Hispanic). The day of week fixed effect is applied to Monday through Saturday. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 10: Fuzzy RD Estimates of the Effect of Anxiety on Marijuana Use for Those Interviewed Within 7 Days Before or After 9/11

Outcome: marijuana use relative frequency (i.e., the number of times using marijuana in the past 30 days divided by 30) with sample size $N = 1,008$

	(1)	(2)	(3)
<i>Panel A: Estimation of the probability of being bothered (1/0) uses a probit model including the "9/11" dummy variable and the following covariates:</i>			
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
Bothered (1/0)	0.032 (0.020)	0.036* (0.020)	0.036* (0.020)
<i>Panel B: Estimation of the probability of being bothered (1/0) uses a probit model including the following covariates but excluding the "9/11" dummy variable:</i>			
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
Bothered (1/0)	0.032 (0.020)	0.036* (0.020)	0.035* (0.020)
<i>Panel C: Estimation of the probability of being bothered (1/0) uses a linear probability model including the "9/11" dummy variable and the following covariates:</i>			
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
Bothered (1/0)	0.032 (0.020)	0.036* (0.020)	0.036* (0.020)

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Descriptions and definitions of variables are provided in the notes of Table 6. The variable “bothered” is a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The variable “9/11” is a dummy variable equal to 1 (or 0) if an individual was interviewed on or after (or before) September 11, 2001. Individual characteristics controlled for are age, age squared, gender, race and ethnicity (White, Black, Native, Asian, or Hispanic). The day of week fixed effect is applied to Monday through Saturday. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 11: OLS, 2SLS, and Conventional RD Estimates of the Effect of Anxiety on Marijuana Use for Those Interviewed Within 3 Days Before or After 9/11

Outcome: marijuana use relative frequency (i.e., the number of times using marijuana in the past 30 days divided by 30) with sample size $N = 435$

	(1)	(2)	(3)
<i>Panel A: OLS</i>			
Bothered (1/0)	0.049*	0.053*	0.052*
	(0.027)	(0.027)	(0.027)
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
<i>Panel B: 2SLS using the "9/11" dummy variable as the instrument for "bothered (1/0)"</i>			
Bothered (1/0)	0.570	0.210	0.216
	(0.457)	(0.204)	(0.184)
First-stage partial F statistic and the p -value (shown below)	2.712	2.405	2.647
	0.100	0.049	0.033
Hansen's J statistic for the overidentification test and the p -value (shown below)	n/a	4.966	3.683
	n/a	0.174	0.298
Days, days squared, days cubed used as extra instruments	No	Yes	Yes
Cubic polynomial of days	No	No	No
Individual characteristics	No	No	Yes
<i>Panel C: Conventional RD approach</i>			
Bothered (1/0)		-1.236	
		[2.130]	
Optimal bandwidth of the assignment variable used		1.691	

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Descriptions and definitions of variables are provided in the notes of Table 6. The variable “bothered” is a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The variable “9/11” is a dummy variable equal to 1 (or 0) if an individual was interviewed on or after (or before) September 11, 2001. Individual characteristics controlled for are age, age squared, gender, race and ethnicity (White, Black, Native, Asian, or Hispanic). The day of week fixed effect is applied to Monday through Saturday. Heteroskedasticity-robust standard errors are reported in parentheses. In Panel C we use a conventional RD estimator with Stata code (“rd”) written by Nichols (2011). This estimator uses local linear regression models on both sides of the cutoff, with a triangle kernel, and the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). The “treatment” variable is “bothered (1/0)” and the “assignment” variable is the “days” variable — the number of days between 09/11/2001 and the interview date of Wave III. The analytical standard error is reported in [bracket]. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 12: OLS, 2SLS, and Conventional RD Estimates of the Effect of Anxiety on Marijuana Use for Those Interviewed Within 7 Days Before or After 9/11

Outcome: marijuana use relative frequency (i.e., the number of times using marijuana in the past 30 days divided by 30) with sample size $N = 1,008$

	(1)	(2)	(3)
<i>Panel A: OLS</i>			
Bothered (1/0)	0.032 (0.020)	0.036* (0.019)	0.035* (0.019)
Cubic polynomial of days	Yes	Yes	Yes
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
<i>Panel B: 2SLS using the "9/11" dummy variable as the instrument for "bothered (1/0)"</i>			
Bothered (1/0)	0.244 (0.211)	0.283 (0.208)	0.248 (0.199)
First-stage partial F statistic and the p -value (shown below)	2.879 0.022	2.823 0.024	2.823 0.024
Hansen's J statistic for the overidentification test and the p -value (shown below)	4.449 0.217	4.256 0.235	2.493 0.477
Days, days squared, days cubed used as extra instruments	Yes	Yes	Yes
Cubic polynomial of days	No	No	No
Individual characteristics	No	Yes	Yes
Day of week fixed effect	No	No	Yes
<i>Panel C: Conventional RD approach</i>			
Bothered (1/0)		-1.109 [2.492]	
Optimal bandwidth of the assignment variable used		2.089	

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. Descriptions and definitions of variables are provided in the notes of Table 6. The variable “bothered” is a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The variable “9/11” is a dummy variable equal to 1 (or 0) if an individual was interviewed on or after (or before) September 11, 2001. Individual characteristics controlled for are age, age squared, gender, race and ethnicity (White, Black, Native, Asian, or Hispanic). The day of week fixed effect is applied to Monday through Saturday. Heteroskedasticity-robust standard errors are reported in parentheses. In Panel C we use a conventional RD estimator with Stata code (“rd”) written by Nichols (2011). This estimator uses local linear regression models on both sides of the cutoff, with a triangle kernel, and the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). The “treatment” variable is “bothered (1/0)” and the “assignment” variable is the “days” variable — the number of days between 09/11/2001 and the interview date of Wave III. The analytical standard error is reported in [bracket]. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 13: Sensitivity Analysis of the Effect of Anxiety on Marijuana Use for Those Interviewed Within 3 Days Before or After 9/11

Γ (gamma)	<i>p</i> -value for testing the null hypothesis of no treatment effect		Γ (gamma)	<i>p</i> -value for testing the null hypothesis of no treatment effect	
	upper bound	lower bound		upper bound	lower bound
1	0.127	0.127			
1.01	0.136	0.119	1.26	0.404	0.020
1.02	0.144	0.112	1.27	0.416	0.019
1.03	0.153	0.105	1.28	0.428	0.017
1.04	0.162	0.098	1.29	0.440	0.016
1.05	0.171	0.092	1.30	0.452	0.015
1.06	0.181	0.086	1.31	0.464	0.014
1.07	0.191	0.080	1.32	0.475	0.013
1.08	0.201	0.075	1.33	0.487	0.012
1.09	0.211	0.070	1.34	0.498	0.011
1.10	0.221	0.065	1.35	0.510	0.010
1.11	0.232	0.061	1.36	0.521	0.009
1.12	0.243	0.057	1.37	0.533	0.008
1.13	0.254	0.053	1.38	0.544	0.008
1.14	0.265	0.049	1.39	0.555	0.007
1.15	0.276	0.046	1.40	0.566	0.007
1.16	0.287	0.043	1.41	0.576	0.006
1.17	0.299	0.040	1.42	0.587	0.006
1.18	0.310	0.037	1.43	0.598	0.005
1.19	0.322	0.034	1.44	0.608	0.005
1.20	0.333	0.032	1.45	0.618	0.004
1.21	0.345	0.029	1.46	0.628	0.004
1.22	0.357	0.027	1.47	0.638	0.004
1.23	0.369	0.025	1.48	0.648	0.003
1.24	0.381	0.023	1.49	0.658	0.003
1.25	0.393	0.022	1.50	0.667	0.003

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. The treatment refers to the variable “bothered” (1/0), a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The outcome variable is marijuana use relative frequency. The sensitivity analysis follows Rosenbaum (2010, Chapter 3). Gamma (Γ) represents the odds ratio that indicates differential treatment assignment due to unobservables, with $\Gamma > 1$ indicating the presence and $\Gamma = 1$ indicating the absence of differential treatment assignment due to unobservables. The sensitivity analysis uses two matched samples (“bothered” vs. “not bothered”) obtained by a one-to-one nearest neighbor matching with the caliper set to be less than 0.1. The propensity score of being bothered by things that usually are not bothersome (1/0) is estimated by a probit model including the following covariates: a dummy variable of being interviewed before (0) or not before (1) 09/11/2001; a cubic polynomial of the number of days between 09/11/2001 and the interview date; age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health.

Table 14: Sensitivity Analysis of the Effect of Anxiety on Marijuana Use for Those Interviewed Within 7 Days Before or After 9/11

Γ (gamma)	<i>p</i> -value for testing the null hypothesis of no treatment effect		Γ (gamma)	<i>p</i> -value for testing the null hypothesis of no treatment effect	
	upper bound	lower bound		upper bound	lower bound
1	0.003	0.003			
1.01	0.004	0.002	1.26	0.092	0.000
1.02	0.004	0.002	1.27	0.101	0.000
1.03	0.005	0.002	1.28	0.110	0.000
1.04	0.006	0.001	1.29	0.119	0.000
1.05	0.007	0.001	1.30	0.128	0.000
1.06	0.008	0.001	1.31	0.139	0.000
1.07	0.010	0.001	1.32	0.149	0.000
1.08	0.011	0.001	1.33	0.160	0.000
1.09	0.013	0.000	1.34	0.171	0.000
1.10	0.015	0.000	1.35	0.183	0.000
1.11	0.018	0.000	1.36	0.195	0.000
1.12	0.020	0.000	1.37	0.208	0.000
1.13	0.023	0.000	1.38	0.221	0.000
1.14	0.026	0.000	1.39	0.234	0.000
1.15	0.030	0.000	1.40	0.248	0.000
1.16	0.033	0.000	1.41	0.262	0.000
1.17	0.037	0.000	1.42	0.276	0.000
1.18	0.042	0.000	1.43	0.291	0.000
1.19	0.047	0.000	1.44	0.306	0.000
1.20	0.052	0.000	1.45	0.320	0.000
1.21	0.058	0.000	1.46	0.336	0.000
1.22	0.064	0.000	1.47	0.351	0.000
1.23	0.070	0.000	1.48	0.366	0.000
1.24	0.077	0.000	1.49	0.382	0.000
1.25	0.085	0.000	1.50	0.398	0.000

Notes: Data are from Wave III of the National Longitudinal Study of Adolescent Health (Add Health) in-home interviews. The treatment refers to the variable “bothered” (1/0), a dummy variable equal to 1 (or 0) if an individual reported that during the past seven days he or she was bothered (or not bothered) by things that usually are not bothersome. The outcome variable is marijuana use relative frequency. The sensitivity analysis follows Rosenbaum (2010, Chapter 3). Gamma (Γ) represents the odds ratio that indicates differential treatment assignment due to unobservables, with $\Gamma > 1$ indicating the presence and $\Gamma = 1$ indicating the absence of differential treatment assignment due to unobservables. The sensitivity analysis uses two matched samples (“bothered” vs. “not bothered”) obtained by a one-to-one nearest neighbor matching with the caliper set to be less than 0.1. The propensity score of being bothered by things that usually are not bothersome (1/0) is estimated by a probit model including the following covariates: a dummy variable of being interviewed before (0) or not before (1) 09/11/2001; a cubic polynomial of the number of days between 09/11/2001 and the interview date; age and age squared; male (1/0); White (1/0); Black (1/0); Native (1/0); Asian (1/0); Hispanic (1/0); total personal income before tax during 2000–2001 (in \$1,000) at Wave III; and years of education at Wave III of Add Health.

Table 15: OLS Estimates of the P900 Effects on 1988–1992 Gain Scores, within +/- 3 Points of the Selection Threshold

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Mathematics (N = 560)</i>								
P900 (1/0)	2.045*** (0.786)	2.135** (0.904)	1.672** (0.794)	1.906** (0.931)	1.201 (0.939)	1.201 (0.939)	1.013 (0.972)	0.846 (0.963)
1988 test score	Yes		Yes		Yes		Yes	
1988 test score squared			Yes				Yes	
1988 test score cubed			Yes				Yes	
1988 test score - cutoff		Yes		Yes		Yes		Yes
(1988 test score - cutoff) squared				Yes				Yes
(1988 test score - cutoff) cubed				Yes				Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies (regions 1–12)					Yes	Yes	Yes	Yes
<i>Panel B: Language (N = 560)</i>								
P900 (1/0)	2.431*** (0.704)	2.518*** (0.794)	2.112*** (0.692)	2.301*** (0.800)	1.984** (0.835)	1.984** (0.835)	1.763** (0.837)	1.688** (0.834)
1988 test score	Yes		Yes		Yes		Yes	
1988 test score squared			Yes				Yes	
1988 test score cubed			Yes				Yes	
1988 test score - cutoff		Yes		Yes		Yes		Yes
(1988 test score - cutoff) squared				Yes				Yes
(1988 test score - cutoff) cubed				Yes				Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies (regions 1–12)					Yes	Yes	Yes	Yes

Notes: The sample includes urban schools with at least 15 students in the fourth grade in 1988. The dependent variables are the 1988–1992 gain scores in mathematics and language. The variable “P900” (1/0) is a dummy variable indicating a school being selected into the P900 program (1) or not (0). Other control variables include the linear and quadratic terms of the number of test takers of the school in 1988, the product of the 1988 test score and the number of test takers of the school in 1988, the average SES of the school’s students in 1990, and the change in that average SES between 1990 and 1992. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 16: OLS Estimates of the P900 Effects on 1988–1992 Gain Scores, within +/- 5 Points of the Selection Threshold

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Mathematics (N = 938)</i>								
P900 (1/0)	2.114*** (0.672)	1.860** (0.772)	1.861*** (0.667)	1.986** (0.793)	1.230 (0.800)	1.230 (0.800)	1.060 (0.806)	1.285 (0.825)
1988 test score	Yes		Yes		Yes		Yes	
1988 test score squared			Yes				Yes	
1988 test score cubed			Yes				Yes	
1988 test score - cutoff		Yes		Yes		Yes		Yes
(1988 test score - cutoff) squared				Yes				Yes
(1988 test score - cutoff) cubed				Yes				Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies (regions 1–12)					Yes	Yes	Yes	Yes
<i>Panel B: Language (N = 938)</i>								
P900 (1/0)	2.365*** (0.578)	2.233*** (0.664)	2.150*** (0.576)	2.331*** (0.676)	1.818*** (0.699)	1.818*** (0.699)	1.630** (0.700)	1.905*** (0.712)
1988 test score	Yes		Yes		Yes		Yes	
1988 test score squared			Yes				Yes	
1988 test score cubed			Yes				Yes	
1988 test score - cutoff		Yes		Yes		Yes		Yes
(1988 test score - cutoff) squared				Yes				Yes
(1988 test score - cutoff) cubed				Yes				Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies (regions 1–12)					Yes	Yes	Yes	Yes

Notes: The sample includes urban schools with at least 15 students in the fourth grade in 1988. The dependent variables are the 1988–1992 gain scores in mathematics and language. The variable “P900” (1/0) is a dummy variable indicating a school being selected into the P900 program (1) or not (0). Other control variables include the linear and quadratic terms of the number of test takers of the school in 1988, the product of the 1988 test score and the number of test takers of the school in 1988, the average SES of the school’s students in 1990, and the change in that average SES between 1990 and 1992. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 17: Fuzzy RD Estimates of the P900 Effects on 1988–1992 Gain Scores, within +/- 3 Points of the Cutoff

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Mathematics (N = 560)</i>						
P900 (1/0)	1.904 (1.640)	1.663 (1.666)	1.731 (1.676)	2.140 (1.756)	1.641 (1.811)	1.633 (1.834)
<i>Covariates included in the probit model of estimating the probability of being selected into the P900 program (1/0)</i>						
1988 test score	Yes	Yes	Yes	Yes	Yes	Yes
1988 test score squared		Yes	Yes		Yes	Yes
1988 test score cubed			Yes			Yes
Eligible (1/0)				Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Language (N = 560)</i>						
P900 (1/0)	2.373 (1.447)	2.167 (1.460)	2.187 (1.461)	2.592 (1.582)	2.179 (1.603)	2.219 (1.611)
<i>Covariates included in the probit model of estimating the probability of being selected into the P900 program (1/0)</i>						
1988 test score	Yes	Yes	Yes	Yes	Yes	Yes
1988 test score squared		Yes	Yes		Yes	Yes
1988 test score cubed			Yes			Yes
Eligible (1/0)				Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel C: Conventional RD approach (full sample used, N = 2,591)</i>						
		<i>Mathematics</i>			<i>Language</i>	
P900 (1/0)		-2.247 [4.486]			-4.846 [4.938]	
Optimal bandwidth of the assignment variable used		5.402			4.451	

Notes: The sample includes urban schools with at least 15 students in the fourth grade in 1988. The dependent variables are the 1988–1992 gain scores in mathematics and language. The variable “P900” (1/0) is a dummy variable indicating a school being selected into the P900 program (1) or not (0). The “eligible” (1/0) is a dummy variable indicating whether the school has a 1988 average test score below or equal to (1), or above (0) the “cutoff definition 2” used by Chay, McEwan and Urquiola (2005). Other control variables include the linear and quadratic terms of the number of test takers of the school in 1988, the product of the 1988 test score and the number of test takers of the school in 1988, the average SES of the school’s students in 1990, and the change in that average SES between 1990 and 1992. In Panel C we use a conventional RD estimator with Stata code (“rd”) written by Nichols (2011). This estimator uses local linear regression models on both sides of the cutoff, with a triangle kernel, and the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). The “treatment” variable is “P900 (1/0)” and the “assignment” variable is the school’s 1988 average test score minus the cutoff score defined by “cutoff definition 2” of Chay, McEwan and Urquiola (2005). The analytical standard error is reported in [bracket]. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 18: Fuzzy RD Estimates of the P900 Effects on 1988–1992 Gain Scores, within +/- 5 Points of the Cutoff

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Mathematics (N = 938)</i>						
P900 (1/0)	1.858 (1.340)	1.821 (1.344)	1.846 (1.338)	2.056 (1.483)	1.872 (1.491)	1.798 (1.496)
<i>Covariates included in the probit model of estimating the probability of being selected into the P900 program (1/0)</i>						
1988 test score	Yes	Yes	Yes	Yes	Yes	Yes
1988 test score squared		Yes	Yes		Yes	Yes
1988 test score cubed			Yes			Yes
Eligible (1/0)				Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Language (N = 938)</i>						
P900 (1/0)	2.155* (1.161)	2.113* (1.164)	2.121* (1.155)	2.341* (1.310)	2.161* (1.313)	2.124 (1.313)
<i>Covariates included in the probit model of estimating the probability of being selected into the P900 program (1/0)</i>						
1988 test score	Yes	Yes	Yes	Yes	Yes	Yes
1988 test score squared		Yes	Yes		Yes	Yes
1988 test score cubed			Yes			Yes
Eligible (1/0)				Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel C: Conventional RD approach (full sample used, N = 2,591)</i>						
		<i>Mathematics</i>			<i>Language</i>	
P900 (1/0)		-2.247 [4.486]			-4.846 [4.938]	
Optimal bandwidth of the assignment variable used		5.402			4.451	

Notes: The sample includes urban schools with at least 15 students in the fourth grade in 1988. The dependent variables are the 1988–1992 gain scores in mathematics and language. The variable “P900” (1/0) is a dummy variable indicating a school being selected into the P900 program (1) or not (0). The “eligible” (1/0) is a dummy variable indicating whether the school has a 1988 average test score below or equal to (1), or above (0) the “cutoff definition 2” used by Chay, McEwan and Urquiola (2005). Other control variables include the linear and quadratic terms of the number of test takers of the school in 1988, the product of the 1988 test score and the number of test takers of the school in 1988, the average SES of the school’s students in 1990, and the change in that average SES between 1990 and 1992. In Panel C we use a conventional RD estimator with Stata code (“rd”) written by Nichols (2011). This estimator uses local linear regression models on both sides of the cutoff, with a triangle kernel, and the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). The “treatment” variable is “P900 (1/0)” and the “assignment” variable is the school’s 1988 average test score minus the cutoff score defined by “cutoff definition 2” of Chay, McEwan and Urquiola (2005). The analytical standard error is reported in [bracket]. Heteroskedasticity-robust standard errors are reported in parentheses. * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.