DISCUSSION PAPER SERIES

IZA

IZA DP No. 10320

The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable

Zhuan Pei
Yi Shen

October 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable

Zhuan Pei
Cornell University
and IZA

Yi Shen
University of Waterloo

Discussion Paper No. 10320
October 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 10320
October 2016

ABSTRACT

The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable*

Identification in a regression discontinuity (RD) design hinges on the discontinuity in the probability of treatment when a covariate (assignment variable) exceeds a known threshold. If the assignment variable is measured with error, however, the discontinuity in the first stage relationship between the probability of treatment and the observed mismeasured assignment variable may disappear. Therefore, the presence of measurement error in the assignment variable poses a challenge to treatment effect identification. This paper provides sufficient conditions for identification when only the mismeasured assignment variable, the treatment status and the outcome variable are observed. We prove identification separately for discrete and continuous assignment variables and study the properties of various estimation procedures. We illustrate the proposed methods in an empirical application, where we estimate Medicaid takeup and its crowdout effect on private health insurance coverage.

JEL Classification: C10, C18

Keywords: regression discontinuity design, measurement error

Corresponding author:

Zhuan Pei
Department of Policy Analysis and Management
Cornell University
134 Martha Van Rensselaer Hall
Ithaca, NY 14853
USA
E-mail: zhuan.pei@cornell.edu

* We thank two anonymous referees, Orley Ashenfelter, Eric Auerbach, Matias Cattaneo, Eleanor Choi, Damon Clark, Kirill Evdokimov, Hank Farber, Marjolaine Gauthier-Loiselle, Jian Kang, Marta Lachowska, Lars Lefgren, Pauline Leung, Jia Li, Carl Lieberman, Andrew Marder, Alex Mas, Jordan Matsudaira, Alexander Meister, Stephen Nei, Andrew Shephard, Lara Shore- Sheppard, and especially Bo Honoré and David Lee for helpful comments. We have also benefited from helpful suggestions given by the participants of the Princeton Labor Seminar, Princeton Political Science Methodology Seminar and the Advances in Econometrics conference at the University of Michigan. We thank David Card and Lara Shore-Sheppard for graciously sharing their data. Zhuan Pei gratefully acknowledges financial support from the Richard A. Lester Fellowship. Finally, we thank Yue Fang, Suejin Lee, Katherine Wen and Le Wu for outstanding research assistance.

1 Introduction

Over the past two decades, many studies in economics have relied on the Regression Discontinuity (RD) design to evaluate the effects of a wide range of policy programs. In an RD design, treatment is assigned based on whether an observed covariate (called the "assignment" or "running" variable) exceeds a known threshold. Provided that agents just above and just below the threshold share the same baseline characteristics, any difference in the outcomes between these two groups can be attributed to the causal effect of the treatment.[1]

A classical RD design depends crucially on the econometrician's ability to observe an accurate measure of the assignment variable. In many cases, however, only a noisy version of the assignment variable is available. This scenario is likely to occur when survey data are used, where the value of the assignment variable comes from self-reporting as opposed to an administrative source.

A typical example is an application that uses income as an assignment variable to study the effect of means-tested transfer programs where eligibility depends on whether income falls below a certain threshold. However, most administrative data cannot be used for an RD because they only include the treatment population, namely those who enroll in the program, and contain little information on the various outcomes for applicants who are denied benefits. Therefore, practitioners may be forced to rely on survey data in order to apply an RD design. For instance, Schanzenbach (2009) uses the Early Childhood Longitudinal Study to study the effect of school lunch on obesity and compares obesity rates for children below and above the reduced-price lunch cutoff for the National School Lunch Program.[2] de la Mata (2012) estimates the effect of Medicaid/CHIP coverage on children's health care utilization and health outcomes with the Panel Study of Income Dynamics (PSID) and its Child Development Study (CDS) supplement. Koch (2013) uses the Medical Expenditure Panel Survey (MEPS) to study health insurance crowdout by focusing on income cutoffs in the Children's Health Insurance Program (CHIP).

The above studies all use income data gathered from surveys as their assignment variable in the RD analyses, but measurement error in surveyed income has been widely documented (see Bound et al. (2001) for a review), and the presence of this measurement error threatens the identification in an RD design. Even if there is perfect compliance with the discontinuous treatment assignment rule (i.e. in a sharp RD design), there may *not* be a discontinuity in the probability of treatment conditional on the *observed* noisy assignment

[1] See Lee and Lemieux (2010) for a survey of the RD literature.

[2] In addition to the RD approach, Schanzenbach (2009) also conducts analyses using other research designs.

variable. Instead of a step function, the first stage relationship – the probability of treatment conditional on the noisy assignment variable – is smoothly S-shaped. This lack of discontinuity greatly complicates the identification and estimation of the program effect.

In this paper, we study the identification and estimation of the RD treatment effect in the presence of classical measurement error in the assignment variable. We consider separately the cases where the assignment variable and measurement error are discrete and continuous. In the discrete case (e.g., binned income), we show that when the assignment variable is bounded, not only can we identify the first stage and outcome conditional expectation functions without specifying the measurement error distribution, but we can also identify the true assignment variable distribution by exploiting the tail behavior of its observed counterpart. This is advantageous in the RD context, since a key appeal is the design's testability, which entails the smoothness of the true assignment variable distribution. In addition, recovery of the true assignment variable distribution may allow for the discovery of bunching and the estimation of meaningful economic quantities, even when the RD design is invalid.[3] The identification result leads to a simple minimum-distance estimation procedure for the true assignment variable distribution, and the estimation of the first stage and outcome relationships follows from an application of Bayes' Theorem. Standard reasoning implies that the resulting estimators are efficient, $\sqrt{N}$-consistent, and asymptotically normal.

We also explore the case where the assignment variable is continuous and propose three alternative approaches to semiparametrically or nonparametrically identify the RD treatment effect. The first approach assumes normal measurement error while remaining agnostic about the true assignment variable distribution; identification follows from the result of Schwarz and Bellegem (2010). The second approach is a novel identification-at-infinity strategy that exploits the tail behavior of the observed assignment variable distribution. The third approach adapts the nonparametric simulation-extrapolation (SIMEX) method of Carroll et al. (1999), which assumes that the variance of the measurement error is known. The last two approaches do not identify the true assignment variable distribution, but recover the RD treatment effect parameter under perfect compliance.

We illustrate our proposed methods in an empirical application where we study the takeup of Medicaid, a government health insurance program in the U.S., and the extent to which it crowds out private health insurance. Since Medicaid eligibility is only available to families with income below a strict cutoff, we apply

[3] As a caveat, however, the detection of nonsmoothness or bunching in the assignment variable distribution when the variable itself is discrete will not entail a nonparametric procedure. Rather, it will depend on the parametric functional form subjectively specified by the researcher.

an RD design exploiting this policy discontinuity. We show that the noisy income measures in the Survey of Income and Program Participation (SIPP) indeed lead to a smooth first stage relationship between Medicaid coverage and income, and then apply two modeling approaches to recover the true income distribution and the first stage and outcome relationships. The first approach discretizes income and uses the proposed minimum distance estimation strategy; the second approach treats income as continuous and adopts a parametric maximum likelihood estimation (MLE) framework. The two methods yield similar findings: Medicaid takeup rate for the barely eligible is between 10 and 25 percent, and there is little evidence of sorting around the threshold and private insurance crowdout. However, the less parametric discrete approach yields much noisier estimates, limiting its empirical appeal.

Several related studies have attempted to tackle the identification challenge posed by measurement error in RD designs. Hullegie and Klein (2010) adopt a parametric Berkson measurement error specification, in which the true assignment variable is the sum of the observed assignment variable and an independent normally distributed measurement error. This specification is attractive in that it can be easily implemented in practice, but it implies that the distribution of the true assignment variable is smooth and precludes the testing of density discontinuity. In this paper, we adopt the more conventional classical measurement error model (Bound et al. (2001)), which allows for nonsmoothness in the assignment variable distribution. Dong (2015) considers rounding (nonclassical) error in the assignment variable typically encountered in age-based RD designs. In three recent studies, Yu (2012) and Yanagi (2014) consider the identification problem assuming "small" measurement error variance; Davezies and Le Barbanchon (2014) assume the availability of an auxiliary dataset so that the true assignment variable distribution can be observed (for the treated population). In contrast to these studies, the measurement error distribution in our setup is not restricted to have small second moments, nor do we assume the observability of the true assignment variable distribution from an auxiliary dataset.[4]

The remainder of the paper is organized as follows. Section 2 introduces the statistical model. Section 3 and 4 study the identification and estimation in the discrete and continuous cases, respectively. Section 5 provides an empirical illustration by employing the proposed methods to estimate Medicaid takeup and crowdout. Section 6 concludes.

[4]A related but distinct problem is heaping in the assignment variable (Barreca et al. (2011), Almond et al. (2011), and Barreca et al. (2016)). In the heaping setup, treatment assignment is based on the observed assignment variable. For the problem at hand, we do not observe the variable determining treatment.

2 Baseline Statistical Model

In a conventional sharp RD design, the econometrician observes the assignment variable X^*, eligibility/treatment $D^* = 1_{[X^*<0]}$, and outcome Y:

$$Y = y(D^*, X^*, \varepsilon). \tag{1}$$

The function y is continuous in its second argument, ε is the unobserved error term, and the eligibility threshold is normalized to zero.[5] A standard result (e.g. Hahn et al. (2001)) is that the treatment effect $\delta_{sharp} = E[y(1,0,\varepsilon) - y(0,0,\varepsilon)|X^* = 0]$ is identified by

$$\delta_{sharp} = \lim_{x^*\uparrow 0} E[Y|X^* = x^*] - \lim_{x^*\downarrow 0} E[Y|X^* = x^*]$$

when the conditional expectation $E[y(D^*, X^*, \varepsilon)|X^* = x^*]$ is continuous at $x^* = 0$ for $D^* = 0, 1$.

An important restriction of the sharp design is the assumption of the observability of program *eligibility* D^* (or equivalently, perfect compliance with the treatment assignment rule). In most applied contexts (such as those in the studies cited above), this assumption is rarely satisfied. In all social programs, for example, the takeup rate of entitlement programs among eligible individuals and families is not 100 percent.[6] The imperfect compliance with the treatment assignment rule gives rise to the so-called fuzzy RD design, where the outcome

$$Y = y(D, X^*, \varepsilon) \tag{2}$$

depends on the actual program participation status D. It is a well-known result that under independence (Hahn et al. (2001)) or smoothness assumptions (DiNardo and Lee (2011)), the ratio

$$\delta_{fuzzy} = \frac{\lim_{x^*\uparrow 0} E[Y|X^* = x^*] - \lim_{x^*\downarrow 0} E[Y|X^* = x^*]}{\lim_{x^*\uparrow 0} E[D|X^* = x^*] - \lim_{x^*\downarrow 0} E[D|X^* = x^*]},$$

is the average treatment effect of D on Y for the "complier" population that takes up benefit when eligible.

[5]Note that in some applications $D^* = 1_{[X^*\geqslant 0]}$ is the treatment determining mechanism. However, the motivating examples in Section 1 follow $D^* = 1_{[X^*<0]}$, i.e. program eligibility depends on whether the assignment variable (income) falls *below* a known cutoff.

[6]See Currie (2006) for a survey on benefit takeup in social programs.

In this paper, we study a variant of these RD models, in which we only observe a noisy version of X^*. Let X denote this observed assignment variable, and $u \equiv X - X^*$ the measurement error. As mentioned in Section 1, a key assumption that distinguishes this study from Hullegie and Klein (2010) is that the measurement error is independent of the true assignment variable as opposed to the observed assignment variable. Formally,

Assumption 1 (Independence) $X^* \perp u$.

In the sharp design, the econometrician observes the joint distribution of (D^*, X, Y) with

$$\begin{aligned} D^* &= 1_{[X^*<0]} \\ X &= X^* + u, \end{aligned} \tag{3}$$

and Y is defined by equation (1). In the fuzzy design, D is observed instead of D^* and Y is defined by equation (2).

In both the sharp and fuzzy designs, the presence of u poses an identification challenge. For simplicity, we first focus on the sharp design (3). As seen in Figure 1, the so-called first stage relationship with the noisy X, $E[D^*|X]$, is smooth at $X = 0$, and we can no longer rely on the discontinuity in that relationship to identify a treatment effect.[7] In a way, this is an extreme form of the attenuation bias, which is typically associated with measurement error. Because of this lack of first stage discontinuity, we cannot treat the problem as a fuzzy design either. In fact, the same force that smoothes out the first stage relationship also smoothes out the outcome function $E[Y|X]$, and the fuzzy RD estimand simply becomes undefined. That said, it is possible to naively perform local regressions on both sides of the threshold and obtain nonzero discontinuity estimates for both $E[D^*|X]$ and $E[Y|X]$ through misspecification. Whether the resulting ratio attenuates or exaggerates the true RD treatment effect depends on the particular specification adopted.

The issue of attenuation, or the lack thereof, can also be seen by following an alternative definition of the sharp RD estimand: $\tilde{\delta}_{sharp} = \lim_{x\uparrow 0}[Y|X = x, D^* = 1] - \lim_{x\downarrow 0}[Y|X = x, D^* = 0]$. When X is measured without error, $\tilde{\delta}_{sharp}$ coincides with the conventional sharp RD estimand. However, unlike the conventional sharp RD estimand, which collapses to zero even when X is measured with very small error, $\tilde{\delta}_{sharp}$ gradually

[7] This is in contrast to the case where a first stage discontinuity exists despite the presence of measurement error in the assignment variable for a subset of observations. Battistin et al. (2009) show that a fuzzy design can still be applied to identify the causal effect in that case. See Card et al. (2015) for an analogous result in the regression kink design.

drifts apart from δ_{sharp} as the measurement error variance increases. In fact, this is an attractive property of $\tilde{\delta}_{sharp}$, which we exploit in Section 4. For now, we argue that $\tilde{\delta}_{sharp}$ is not necessarily attenuated, and the direction of the bias in $\tilde{\delta}_{sharp}$ depends on the particular functional form in the relationship between X^* and Y. Suppose, for example, Y is linear in X^* for both the treatment and control groups: $E[Y|X^*, D^* = 1] = \psi_0 + \psi_1 X^*$ and $E[Y|X^*, D^* = 0] = \psi_2 + \psi_3 X^*$. If we project Y on X within each group, standard OLS results indicate that the slope estimators $\hat{\psi}_1$ and $\hat{\psi}_3$ are both attenuated. This in turn implies that if $\psi_1, \psi_3 < 0$, then $\hat{\psi}_0$ is biased upward and $\hat{\psi}_2$ downward, and the resulting RD estimator $\hat{\psi}_0 - \hat{\psi}_2$ is biased upward. When $\delta_{sharp} = \psi_0 - \psi_2 > 0$, we actually arrive at an exaggerated, rather than attenuated, RD estimate! In short, the association of classical measurement error with attenuation comes from slope estimation in an error-in-variable model and does not generalize to intercept estimation, as is the case for RD.

In the rest of this paper, we study whether we can recover the RD treatment effect in the presence of u. The first step in our quest is the identification of the distribution of X^*. As mentioned in the Introduction, not only is the identification of the X^* distribution used to recover the RD treatment effect parameter, but it can also be used to test the validity of the RD design. However, it is not possible to identify the X^* distribution from the observed X distribution absent any other information. In the presence of measurement error, economists have traditionally relied on repeated measures of the true explanatory variable (e.g. Ashenfelter and Krueger (1994), Black et al. (2000), Hausman et al. (1991), Li and Vuong (1998), Schennach (2004)). However, such measures may not be available in the data. What is helpful in a sharp RD context is that observed program eligibility $D^* = 1_{[X^* < 0]}$, which is a deterministic function of X^*, is informative of the value of X^*. Therefore, it becomes an interesting question as to whether and under what additional assumptions we can use the joint distribution of (X, D^*) to identify the X^* distribution and δ_{sharp}. In a fuzzy RD model, program takeup D is no longer a deterministic function of true assignment variable X^*. As a consequence, we will impose additional assumptions on the u distribution for identification.

As pointed out by a referee, another deviation from model (3) that may arise in practice results from our inability to observe the true RD threshold. For example, the econometrician may not have all the information to precisely construct a family's eligibility threshold for a means-tested program. While it is difficult to disentangle the measurement error in threshold from that in income or to identify the true threshold a particular family faces, the problem of identifying the RD treatment effect is the same as above. To see this, suppose the true eligibility assignment mechanism is $D^* = 1_{[W^* < c^*]}$, where W^* is, say, the actual family income and c^* the monetary eligibility threshold (note that we use the income normalization

$X^* = W^* - c^*$ above, and the eligibility threshold is normalized to zero). Suppose the econometrician only sees proxies $W = W^* + u$ and $c = c^* + v$, but not W^* and c^* directly. In this case, we can rewrite the eligibility assignment mechanism as $D^* = 1_{[X-u+v<0]} = 1_{[X-\tilde{u}<0]}$ where $\tilde{u} \equiv u - v$. Suppose Assumption 1 holds and that $X^* \perp\!\!\!\perp v$, then $X^* \perp\!\!\!\perp \tilde{u}$, and we have a model isomorphic to (3).

3 Discrete Assignment Variable and Measurement Error

In this section, we study the identification of the assignment variable distribution and the RD treatment effect parameter when X^* and u are discrete. A discrete assignment variable setup may appear odd given the continuity assumption in identifying the treatment effect in a sharp RD design, but it is necessary in many policy contexts where an RD design appears compelling (Lee and Card (2008)). Even if the assignment variable is continuous (e.g. income), the discretization of X^* can be thought of as a binned-up approximation (this is common practice in graphical presentations of most RD applications). We should point out, however, that if we assume independence between the underlying *continuous* assignment variable and measurement error, then their respective discretized versions are not going to be independent.[8] However, we can also start with the assumption that the discretized X^* and u are independent, in which case their continuous versions will not be. Which one of these assumptions is correct? We are inclined to believe that *neither* assumption is likely to hold empirically, but it is important to gauge whether they reasonably approximate the data. We discuss an overidentification test to that effect in Subsection 3.4.

It is also worth noting that when the discrete X and X^* have the same support (e.g. when they denote age/birthdate, student enrollment), the measurement error cannot be independent to the true assignment variable – the extreme form of this is the misclassification error in a binary variable, which is known to be mean-reverting. Identification of the assignment variable distribution as well as the RD conditional expectation functions with this type of nonclassical measurement error is a corollary of the problem nicely solved by Chen et al. (2009).[9] Under appropriate rank conditions and shape restrictions on the conditional expectation function, Chen et al. (2009) employ the observed conditional characteristic function of Y and its derivative for identification.

[8] Dong (2015) makes a similar point in the case of age rounding.

[9] We thank a referee for suggesting this reference.

3.1 Identification of the True Assignment Variable Distribution: Perfect Compliance

In this subsection, we provide sufficient conditions for identifying the distributions of $X^*$and u from the joint distribution of X and D^* in model (3), where X^* and u are discrete and bounded. The identification result is shown in two steps: 1) identification of the support of X^* and u and 2) identification of the probability mass at each point in the supports of X^* and u. In addition to the independence between X^* and u, the identification result relies on the assumption of positive mass in the X^* distribution around the threshold, along with a technical rank condition discussed in detail below.

Denote the support of any random variable Z by support_Z, and denote $L_Z \equiv \min\{\text{support}_Z\}$ and $U_Z \equiv \max\{\text{support}_Z\}$ for a discrete and bounded Z. Without loss of generality, we consider the case where the supports of X^* and u belong to the set of integers. Formally, the discrete and bounded support assumption is written as

Assumption DB (Discrete and Bounded Support) $\text{support}_{X^*} \subseteq \{L_{X^*}, L_{X^*}+1, ..., U_{X^*}-1, U_{X^*}\}$ and $\text{support}_u \subseteq \{L_u, L_u+1, ..., U_u-1, U_u\}$, where $L_{X^*}, U_{X^*}, L_u, U_u \in \mathbb{Z}$.

The assumption of independence between X^* and u gives strong implications relating their respective supports to the observed support of X, conditional on D^*. Specifically,

$$\begin{aligned} \min\{\text{support}_{X|D^*=d}\} &= \min\{\text{support}_{X^*|D^*=d}\} + \min\{\text{support}_u\} \\ \max\{\text{support}_{X|D^*=d}\} &= \max\{\text{support}_{X^*|D^*=d}\} + \max\{\text{support}_u\} \text{ for } d = 0, 1 \end{aligned} \tag{4}$$

which imposes four restrictions on six unknowns. In order to identify the supports of X^* and u, we impose the additional assumption

Assumption 2 (Threshold Support) $-1, 0 \in \text{support}_{X^*}$.

Assumption 2 states that there exist agents with X^* right at and below the eligibility threshold 0. This is not a strong assumption and must be satisfied in all valid RD designs because the quasi-experimental variation of an RD design comes from agents around the threshold. The addition of this weak assumption is sufficient for identifying the supports.

Lemma 1 (**Support Identification in a Sharp Design**) Under Assumptions DB, 1 and 2, L_{X^*}, U_{X^*}, L_u and U_u are identified.

Proof of Lemma 1 is in Subsection A.1.1 of the online Supplemental Material (Pei and Shen (2016)). Intuitively, individuals who are in the program ($D^* = 1$) but appear ineligible ($X \geqslant 0$) have a positive measurement error $u > 0$. Analogously, those with $D^* = 0$ but $X < 0$ have a negative measurement error $u < 0$. This is essentially the insight behind equation (4) and the proof of Lemma 1.

With the support of X^* identified, we next derive the identification of the probability mass of X^* at every point in its support. Denote the probability mass of X^* by p_k at each integer k, and denote that of u by m_l at each integer l. Let the conditional probability masses of the observed assignment variable X be $q_j^1 \equiv \Pr(X = j|D^* = 1)$ and $q_j^0 \equiv \Pr(X = j|D^* = 0)$, for j in the support of the X distribution. Denote the marginal probabilities of D^* by $r^1 \equiv \Pr(D^* = 1)$ and $r^0 \equiv \Pr(D^* = 0)$.

Under the independence assumption of X^* and u, the distribution of $X|D^*$ is the convolution of the distribution of $X^*|D^*$ and that of u. In particular,

$$\begin{aligned} q_j^1 &= \frac{\sum_{k<0} p_k m_{j-k}}{\sum_{k<0} p_k} \\ q_j^0 &= \frac{\sum_{k\geqslant 0} p_k m_{j-k}}{\sum_{k\geqslant 0} p_k} \end{aligned} \tag{5}$$

Additionally, the marginal probabilities of D give rise to two more restrictions on the parameters of interest:

$$\begin{aligned} r^1 &= \sum_{k<0} p_k \\ r^0 &= \sum_{k\geqslant 0} p_k \end{aligned} \tag{6}$$

Note that $r^1, r^0 > 0$ under Assumption 2, and the q_j^1 and q_j^0's are thus well-defined. Note also that $\sum_k p_k = 1$ follows from $r^1 + r^0 = 1$ and (6), and $\sum_l m_l = 1$ follows from $\sum_j (q_j^1 r^1 + q_j^0 r^0) = 1$, and they are therefore redundant constraints.

Taken together, (5) and (6) represent $2K_u + K_{X^*}$ restrictions on $K_u + K_{X^*}$ parameters, where $K_{X^*} = |\{L_{X^*}, L_{X^*}+1, ..., U_{X^*}-1, U_{X^*}\}|$ and $K_u = |\{L_u, L_u+1, ..., U_u-1, U_u\}|$ denote the number of probability mass points to be identified in the X^* and u distributions. Even though there are more constraints than parameters, it is not clear that the X^* distribution is always identified because of the nonlinearity in (5). To formally investigate the identifiability of the parameters, we introduce the following notations: let $p_k^1 \equiv \frac{p_k}{r^1}$ for $k < 0$ and $p_k^0 \equiv \frac{p_k}{r^0}$ for $k \geqslant 0$. Define $Q^1(t) \equiv \sum_j q_j^1 e^{tj}$, $Q^0(t) \equiv \sum_j q_j^0 e^{tj}$, $P^1(t) \equiv \sum_k p_k^1 e^{tk}$, $P^0(t) \equiv \sum_k p_k^0 e^{tk}$ and $M(t) \equiv \sum_l m_l e^{tl}$ to be the moment generating functions (MGF's) of the random variables, $X|D = 1$,

$X|D=0$, $X^*|D=1$, $X^*|D=0$ and u.[10] It is a well-known result that the moment generating function of the sum of two independent random variables is the product of the moment generating functions of the two variables (see for example Chapter 10 of Grinstead and Snell (1997)). Consequently, equations (5) and (6) can be compactly represented as

$$\begin{aligned} Q^d(t) &= P^d(t)M(t) \text{ for all } t \neq 0 \text{ and } d = 0,1 \\ P^1(0) &= P^0(0) = 1 \end{aligned} \tag{7}$$

Because $P^1(t)$ and $P^0(t)$ are everywhere positive, equation (7) implies

$$M(t) = \frac{Q^1(t)}{P^1(t)} = \frac{Q^0(t)}{P^0(t)}.$$

It follows that

$$P^0(t)Q^1(t) = P^1(t)Q^0(t),$$

which eliminates the nuisance parameters associated with the measurement error distribution.[11] Matching the coefficients in $P^0(t)Q^1(t)$ to those in $P^1(t)Q^0(t)$ along with the constraint $P^1(0) = P^0(0) = 1$ results in the following *linear* system of equations in terms of the p_k^1's and p_k^0's:

$$\underbrace{\begin{bmatrix} q^1_{U_u-1} & 0 & \cdots & 0 & -q^0_{U_{X^*}+U_u} & 0 & \cdots & 0 \\ q^1_{U_u-2} & q^1_{U_u-1} & \cdots & 0 & -q^0_{U_{X^*}+U_u-1} & -q^0_{U_{X^*}+U_u} & \cdots & 0 \\ \vdots & q^1_{U_u-2} & \cdots & \vdots & \vdots & -q^0_{U_{X^*}+U_u-1} & \cdots & \vdots \\ q^1_{L_u+L_{X^*}} & \vdots & \cdots & 0 & -q^0_{L_u} & \vdots & \cdots & 0 \\ 0 & q^1_{L_u+L_{X^*}} & \cdots & q^1_{U_u-1} & 0 & -q^0_{L_u} & \cdots & -q^0_{U_{X^*}+U_u} \\ \vdots & 0 & \cdots & q^1_{U_u-2} & \vdots & 0 & \cdots & -q^0_{U_{X^*}+U_u-1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & q^1_{L_u+L_{X^*}} & 0 & 0 & \cdots & -q^0_{L_u} \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{bmatrix}}_{\mathbf{Q}:\ (K_{X^*}+K_u)\times K_{X^*}} \underbrace{\begin{bmatrix} p^0_{U_{X^*}} \\ p^0_{U_{X^*}-1} \\ \vdots \\ p^0_0 \\ p^1_{-1} \\ p^1_{-2} \\ \vdots \\ p^1_{L_{X^*}} \end{bmatrix}}_{\mathbf{p}:\ K_{X^*}\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{b}:\ (K_{X^*}+K_u)\times 1} \tag{8}$$

Standard results in linear algebra can be invoked to provide identification of the probability masses.

[10]Because of the bounded support assumption, the defined moment generating functions always exist and are positive for all t.

[11]The independence assumption implies that the measurement error distribution is invariant with respect to treatment status, and it plays a key role in the derivation above. This is certainly a strong restriction, since it is possible that the treatment status may affect measurement error, but we cannot achieve identification without this restriction.

Denote system (8) with the compact notation $\mathbf{Qp} = \mathbf{b}$, where $\mathbf{Q}$ is the $(K_{X^*} + K_u) \times K_{X^*}$ *data* matrix, $\mathbf{p}$ the $K_{X^*} \times 1$ *parameter* vector and $\mathbf{b}$ the $(K_{X^*} + K_u) \times 1$ *constant* vector of 0's and 1's. The parameter vector $\mathbf{p}$ is identified if and only if $\mathbf{Q}$ is of full rank. Note that there are more rows than columns in $\mathbf{Q}$, i.e. $K_{X^*} + K_u > K_{X^*}$, and we introduce the following assumption

Assumption 3 (Full Rank) $\mathbf{Q}$ in equation (8) has full column rank.

Note that Assumption 3 is not always satisfied, and an example is provided in Subsection A.2 of the Supplemental Material. At the same time, Assumption 3 is directly testable because $\text{rank}(\mathbf{Q}) = \text{rank}(\mathbf{Q}^\mathbf{T}\mathbf{Q})$, and $\mathbf{Q}$ is of full rank if and only if $\det(\mathbf{Q}^\mathbf{T}\mathbf{Q}) \neq 0$. The distribution of the determinant estimator can be obtained by the delta method because it is a polynomial in the q_j^0's and q_j^1's, the observed probability masses.

With Assumption 3, the $\mathbf{p}$ vector is identified. Because $p_k = r^1 p_k^1$ for $k < 0$ and $p_k = p_k^0 r^0$ for $k \geqslant 0$ and because r^1 and r^0 are observed, uniqueness of the p_k^1's and the p_k^0's implies the uniqueness of the p_k's. Although parameters of the the measurement error distribution are eliminated in (8), they are identified after the identification of the p_k's as shown in Subsection A.3 of the Supplemental Material. Formally, the identification of probability masses is summarized in the following Lemma.

Lemma 2 (**Probability Mass Identification in a Sharp Design**) Suppose $(\{p_k\}, \{m_l\})$ solves the system of equations consisting of (5) and (6). Then $(\{p_k\}, \{m_l\})$ is the unique solution if and only if Assumption 3 holds.

Combining Lemmas 1 and 2 implies the identification of the true underlying distributions of model (3):

Proposition 1 Under Assumption DB, 1, 2, and 3, the distributions of X^* and u are identified.

Intuitively, the identification result extracts information about u from the tails of the X^* distribution in the treatment and control groups. The discrete and bounded assumption reduces the dimensionality of the identification problem and fully specifies the statistical model with a finite number of parameters, even though we are completely agnostic about the shapes of the X^* and u distributions over their respective supports. After we have pinned down the set of parameters, we identify the X^* distribution by starting from the tails and working our way in, using the independent measurement error assumption.

When we relax boundedness, we no longer have the luxury of a finite dimensional parameter space. Consequently, the distributions of X^* and u are in general not identified from model (3), as shown through a constructive example in Subsection A.4 of the Supplemental Material.

3.2 Identification of the True Assignment Variable Distribution: Imperfect Compliance

As mentioned in Section 2, the assumption of perfect compliance or equivalently the observability of eligibility ($D^* = 1_{[X^*<0]}$) is not often satisfied, as is the case in almost all means-tested social programs. Instead, only a measure of program participation D may be available. In this subsection, we consider the more realistic case of imperfect compliance, or a fuzzy design, for discrete and bounded assignment variable X^* and measurement error u. We set out to identify the X^* distribution from the observed joint distribution (X,D).

The difficulty with imperfect compliance is that rather than having benefit receipt D as a deterministic step function of X^*, $\Pr(D=1|X^*)$ is potentially an unrestricted function in X^*, even though eligibility is still given by $D^* = 1_{[X^*<0]}$. In the extreme, program participation D could be independent from X^* and therefore would not provide any additional information. If this were the case, uniquely deconvolving X^* and u from the observed joint distribution (X,D) would not be possible. In many programs (typically means-tested transfer programs), however, it is the case that if an agent's true assignment variable is above the eligibility threshold, she is forbidden from participating in the program, that is

Assumption 4 (One-sided Fuzzy) $\Pr(D=1|X^*=x^*)=0$ for $x^* \geqslant 0$.

Under Assumption 4, any agent with $D=1$ has true assignment variable $X^*<0$. It follows that the upper end point in the $X|D=1$ distribution identifies U_u, provided that Assumptions 1 and 2 hold for the $D=1$ population. Unlike the perfect compliance scenario, where $X^* \perp\!\!\!\perp u$ conditional on D^* directly follows from Assumption 1, an additional assumption is needed to ensure the independence between X^* and u conditional on $D=1$:

Assumption 1F (Strong Independence) $u \perp\!\!\!\perp X^*, D$.

Identification also requires an extension of Assumption 2:

Assumption 2F (Threshold Support: Fuzzy) $-1 \in \text{support}_{X^*|D=1}$ and $0 \in \text{support}_{X^*}$.

Remark 1. A weaker version of Assumption 1F, that $u \perp\!\!\!\perp X^*$ conditional on D, suffices for the results below. However, it may be difficult to find situations where this weaker assumption is empirically justified but Assumption 1F is not. In fact, this weaker assumption does not even imply that $X^* \perp\!\!\!\perp u$ unconditionally as in Assumption 1. Therefore, we propose the stronger Assumption 1F.

Remark 2. Although Assumption 2F is stronger than Assumption 2, it needs to be satisfied in a valid fuzzy RD design without measurement error. That is, there must be nonzero takeup just below the eligibility cutoff, without which a first stage discontinuity does not exist.

Even with Assumptions IF and 2F, we still need to distinguish between nontakeup and ineligibility. An individual with $D = 0$ and $X = -1$ could have true income $X^* = 1$ (with an implied measurement error $u = -2$) and is not program eligible; or she could be eligible with income $X^* = -1$ (with an implied measurement error $u = 0$) but chooses not to participate in the program. On the one hand, if every observation with $D = 0$ is treated as ineligible, then the lower end point in the support of u, L_u, is that in the $X|D = 0$ distribution. On the other hand, if every observation with $D = 0$ is treated as an eligible nontakeup, then L_u is 0. Clearly, the two treatments imply different distributions. However, if the researcher believes that the identified length of the right tail in the u distribution sheds light on the length of its left tail, it may be reasonable to assume

Assumption 5 (Symmetry in Support) $L_u = -U_u$,

which is weaker than imposing symmetry in the measurement error distribution, as is conventional in the literature. With the additional Assumptions 4, 1F, 2F and 5, the supports of the X^* and u are identified:

Lemma 1F (**Support Identification in a Fuzzy Design**) Under Assumptions DB, 1F, 2F, 4 and 5, the support of u and the support of X^* conditional on $D = 0$ and $D = 1$ are identified.

The proof of Lemma 1F is provided in Subsection A.1.1 of the Supplemental Material. The intuition is similar to the sharp case: thanks to Assumption 4, the right tail of $X^* \geqslant 0$ in the $D = 1$ population provides information on the length of the right tail of the u distribution. The length of the left tail of the measurement error distribution is then obtained by symmetry (Assumption 5). The identification of probability masses can proceed analogously as in Subsection 3.1, but the matrix $\mathbf{Q}$ needs to be modified to a matrix with larger dimensions, which we denote by $\mathbf{Q_F}$. With the corresponding rank assumption

Assumption 3F (Full Rank: Fuzzy) $\mathbf{Q_F}$ has full column rank,

we have the following identification result in the fuzzy case:[12]

Proposition 1F Under Assumptions DB, 1F, 2F, 3F, 4 and 5, the distributions of X^* conditional on D and u are identified.

[12]The exact form of $\mathbf{Q_F}$, along with the proof of Proposition 1F, is provided in Subsection A.1.2 of the Supplemental Material.

Remark 3. Identification is possible in the absence of Assumptions 2F, 4 and 5, provided that we know the values of U_u and L_u. In this case, $L_{X^*|D=d}$ and $U_{X^*|D=d}$ can be recovered using this knowledge, and the probability masses are identified if the full rank condition is satisfied. In practice, this observation has little practical value since researchers rarely–if at all–know the true values of U_u and L_u.

Remark 4. Related to Remark 3, identification can be obtained with only the independence assumption (Assumption 1/1F) if we have explicit knowledge of the marginal distribution of X^*, say from a validation sample.[13] This is because, as it is easy to show, the marginal distribution of u is identified from the marginal distribution of X and X^* by an overidentified linear system of equations. It follows that the distribution of X^* conditional on D^* (D) in the sharp (fuzzy) case is identified from the observed $X|D^*$ ($X|D$) distribution and the identified u distribution.

In practice, however, a researcher is unlikely to obtain the marginal distribution of X^* in the case of a transfer program even if she has access to administrative earnings data. First, commonly used administrative earnings records are of *quarterly or annual* frequency, but program eligibility is usually based on *monthly* income. Second, the income used for determining program eligibility is typically earnings after certain deductions (child care or work related expenses, for example) plus unearned income. In that sense, the administrative earnings records are also a noisy version of the income for program eligibility determination, not to mention the fact that they may not perfectly measure true earnings either (e.g. Abowd and Stinson (2013)).

Remark 5. One might question our implicit assumption that benefit receipt D is accurately measured.[14] Errors in reporting program participation status in means-tested transfer programs have been documented in validation studies of survey data. Marquis and Moore (1990) estimate that the cash welfare underreporting rate (i.e. actual benefit recipients who report no benefits) in the 1984 SIPP panel could be as high as 50%. Underreporting of Medicaid coverage appears to be less severe – Card et al. (2004) estimate that the overall error rate in the 1990-93 SIPP Medicaid status is 15% for the state of California.

Underreporting, however, does not pose a threat to the identification of the X^* distribution, provided that those with $D = 1$ indeed received benefits and were therefore eligible. It follows that the supports of $X^*|D$ and u are identified correctly and that probability masses can be recovered, as long as Assumption

[13] This is the starting point of the approach taken by Davezies and Le Barbanchon (2014).

[14] For formal studies on the identification and estimation of regression models/treatment effects under misclassification of the treatment variable, see the seminal papers by Mahajan (2006) and Lewbel (2007).

1F holds. It will be problematic, however, if those who do not take part in the program mistakenly report participation. Fortunately, the rate of false-positive reporting of participation in transfer programs is very small empirically – around 0.2% in the Marquis and Moore (1990) study and 1.5% in Card et al. (2004). This suggests that the reporting error in D does not pose a big threat to the procedure above when applying an RD design using benefit discontinuities at the eligibility cutoff. Furthermore, trimming procedures can be undertaken to correct for the false-positive reporting problem, which we discuss in Subsection 3.5.

3.3 Identification of the Conditional Expectation Functions and the RD Treatment Effect

In this subsection, we show that the conditional expectation function $E[Y|X^*]$ in a sharp design and $E[Y|X^*]$ and $E[D|X^*]$ in a one-sided fuzzy design can be identified under conditional versions of Assumptions 1, 2 and 3. In essence, these assumptions allow the performance of the deconvolution exercise detailed in the previous subsections for each value of Y. In a sharp design, for example, once we identify $\Pr(X^*|Y)$ for each value of Y, we apply Bayes' Theorem to recover $\Pr(Y|X^*)$, which we then integrate to obtain $E[Y|X^*]$. Finally, as with any RD design with a discrete assignment variable, we can parametrically extrapolate the conditional expectation functions to recover the RD treatment effect.[15] For ease of exposition, we focus on the case with a binary Y and discuss identification with a general Y distribution in the Supplemental Material.

In the sharp RD model (1), the treatment effect is $\delta_{sharp} = \text{extp}_{c\uparrow 0} E[Y|X^* = c] - E[Y|X^* = 0]$ for discrete X^*. The first term is the left intercept of the $E[Y|X^*]$ function parametrically extrapolated using $E[Y|X^* = x^*]$ for $x^* < 0$, and we use "extp" to denote this extrapolation operator (analogous to the limit operator in the continuous X^* case). The second term $E[Y|X^* = 0]$ is directly observed from the data. In order to identify δ_{sharp}, we need to identify the conditional expectation function $E[Y|X^*]$, for which we propose the assumptions below. These assumptions imply that Assumptions 1, 2 and 3 hold conditional on Y:

Assumption 1Y (Nondifferential Measurement Error) $u \perp\!\!\!\perp X^*, Y$.[16]

Assumption 2Y (Threshold Support) $-1, 0 \in \text{support}_{X^*|Y=y}$ for each $y = 0, 1$.

[15] As pointed out by Lee and Card (2008), there may be a misspecification error in the parametric extrapolation of $E[Y|X^* = 0, D = 1]$. In this paper, we abstract away from this issue.

[16] Nondifferential measurement error is a common assumption in the literature (see Carroll et al. (2006)). As with Assumption 1F, a weaker version of Assumption 1Y, $X^* \perp\!\!\!\perp u$ conditional on Y, also delivers the identification results. However, we adopt Assumption 1Y for its simplicity in economic interpretation, and the same goes for Assumption 1FY below.

Note that Assumption 1Y and Assumption 2Y allow the formulation of the conditional counterparts of equation (8): $\mathbf{Q}_Y \mathbf{p}_Y = \mathbf{b}_Y$ for $Y = 0,1$, where $\mathbf{Q}_Y$ and $\mathbf{p}_Y$ consist of probability masses of q_j^1,q_j^0, p_k^1 and p_k^0 conditional on Y.

Assumption 3Y (Full Rank) The matrix $\mathbf{Q}_Y$ is of full rank for $Y = 0,1$.

Proposition 2 Under Assumptions DB, 1Y, 2Y and 3Y, the conditional expectation function $E[Y|X^*]$ is identified for model (1).

As stated at the beginning of this subsection, the proof of Proposition 2 (and of Proposition 2F below) follows from an application of the Bayes' Theorem. The details are provided in Subsection A.1.3 of the Supplemental Material. Once we pin down $E[Y|X^* = x^*]$, δ_{sharp} is subsequently identified by parametric extrapolation.

Analogous to a sharp design, we can obtain the identification of δ_{fuzzy}, which in the discrete case is defined as $\delta_{fuzzy} \equiv \frac{\text{extp}_{x^* \uparrow 0} E[Y|X^*=x^*] - E[Y|X^*=0]}{\text{extp}_{x^* \uparrow 0} E[D|X^*=x^*] - E[D|X^*=0]}$. The only difference is the need to recover the first stage relationship $E[D|X^*]$ in a fuzzy design. Again, assumptions underpinning Proposition 1F are extended to hold conditional on Y:

Assumption 1FY (Strong Independence and Nondifferential Measurement Error: Fuzzy) $u \perp\!\!\!\perp X^*, D, Y$.

Assumption 2FY (Threshold Support: Fuzzy) $-1 \in \text{support}_{X^*|D=1,Y=y}$ and $0 \in \text{support}_{X^*|Y=y}$ for each $y = 0,1$.

Assumption 3FY (Full Rank: Fuzzy) The matrix $\mathbf{Q}_{\mathbf{F}Y}$ is of full rank for $Y = 0,1$.[17]

Proposition 2F Under Assumptions DB, 1FY, 2FY, 3FY, 4 and 5, the conditional expectation functions $E[Y|X^*]$ and $E[D|X^*]$ are identified for model (2).

Remark 6. Propositions 2 and 2F can be easily generalized from the formulation with a binary Y – see Subsection A.5 of the Supplemental Material for details.

[17] Just like $\mathbf{Q}_Y$ is the conditional analog of $\mathbf{Q}$, $\mathbf{Q}_{\mathbf{F}Y}$ is the conditional analog of $\mathbf{Q}_\mathbf{F}$.

3.4 Estimators of the Assignment Variable Distribution and the RD Treatment Effect

In this subsection, we propose estimators for the assignment variable distribution and the RD treatment effect in the discrete and bounded case. As with identification, estimation of the X^* distribution follows two steps: estimation of its support and estimation of the probability masses at each point in its support. Support estimation follows the identification results in the proofs of Lemmas 1 and 1F, which derive from equation (4), with the population quantities replaced by sample analogs. We abstract away from the sampling error of support and simply assume that the sample is large enough to reveal the true support of the distribution. We present the case in the sharp design setting where we omit the F subscript for notational convenience. Results in the fuzzy case follow by replacing D^* by D.

Given the specification of probability model, the likelihood function can be explicitly written out by using the p_k^1's, p_k^0's, m_l's and the marginal probabilities r^1 and r^0. Formally, the likelihood for the joint distribution (X,D^*) is

$$\begin{aligned} L(X_i,D_i^*) &= L(X_i|D_i^*)L(D_i^*) \\ &= \{(\sum_k p^1_{X_i-k}m_k)r^1\}^{D_i^*}\{(\sum_k p^0_{X_i-k}m_k)r^0\}^{1-D_i^*} \end{aligned} \tag{9}$$

Researchers can directly maximize the logarithm of equation (9), and the resulting estimators are efficient provided that the parameters are in the interior of the parameter space, i.e. strictly between zero and one. However, the analytical solutions to maximizing the log likelihood do not appear to exist, and numerically optimizing (9) may become computationally intensive as the number of points in the supports of X^* and u increases.

An alternative strategy relies on the identification equation (8), which fits nicely into a standard minimum distance framework $f(\mathbf{q},\mathbf{p}) = \mathbf{Qp}-\mathbf{b} = 0$ (Kodde et al. (1990)) from which an estimator of $\mathbf{p}$ can be obtained easily.[18] Because of the linearity in (8), the parameter vector of interest $\mathbf{p}$ can be estimated analytically once an estimator of $\mathbf{q}$ is obtained. Estimation proceeds according to the following steps.

1. Obtain the estimators $\hat{q}_j^1 = \frac{\sum_i 1_{[X_i=j]}\cdot 1_{[D_i^*=1]}}{\sum_i 1_{[D_i^*=1]}}$, $\hat{q}_j^0 = \frac{\sum_i 1_{[X_i=j]}\cdot 1_{[D_i^*=0]}}{\sum_i 1_{[D_i^*=0]}}$ and $\hat{r}^1 = \frac{1}{N}\sum 1_{[D_i^*=1]}$ (N denotes the sample size), as well as $\hat{\Omega}$, which is a consistent estimator for the asymptotic variance covariance matrix Ω of $\hat{\mathbf{q}}$, (that is, $\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q}) \Rightarrow N(0,\Omega)$). Since $X|D^* = d$ follows a multinomial distribution for each d, $\hat{\Omega}$ is a

[18]Analogous to $\mathbf{p}$, which contains the parameters p_k^0's and p_k^1's, $\mathbf{q}$ is the vector that contains all the parameters q_j^0's and q_j^1's .

block-diagonal matrix $\hat{\Omega} = \begin{bmatrix} \hat{\Omega}^{11} & \mathbf{0} \\ \mathbf{0} & \hat{\Omega}^{00} \end{bmatrix}$, and

$$\hat{\Omega}^{dd}_{ij} = \begin{cases} (1-\hat{q}^d_i)\hat{q}^d_i/(d\hat{r}^1 + (1-d)(1-\hat{r}^1)) & \text{if } i = j \\ -\hat{q}^d_i\hat{q}^d_j/(d\hat{r}^1 + (1-d)(1-\hat{r}^1)) & \text{if } i \neq j \end{cases}$$

2. Form the estimator of the $\mathbf{Q}$ matrix under perfect compliance in (8), $\hat{\mathbf{Q}}$, by replacing the q^1_j and q^0_j in $\mathbf{Q}$ with their estimators;

3. Derive a consistent estimator of $\mathbf{p}$: $\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} f(\hat{\mathbf{q}},\mathbf{p})' f(\hat{\mathbf{q}},\mathbf{p}) = (\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-1}(\hat{\mathbf{Q}}'\mathbf{b})$;

4. Compute the optimal weighting matrix $\widehat{\mathbf{W}} = (\widehat{\nabla_{\mathbf{q}} f}\hat{\Omega}\widehat{\nabla_{\mathbf{q}} f}')^{-1}$ where $\hat{\Omega}$ is a consistent estimator for the variance covariance matrix of the $\mathbf{q}$ derived in step 1.[19] $\widehat{\nabla_{\mathbf{q}} f}$ is a consistent estimator for $\nabla_{\mathbf{q}} f$, the Jacobian of f with respect to $\mathbf{q}$. Because $\nabla_{\mathbf{q}} f$ depends on $\mathbf{p}$, step 3 is necessary for first obtaining a consistent estimate of $\mathbf{p}$. Since f is linear in $\mathbf{q}$, $\widehat{\nabla_{\mathbf{q}} f}$ can be computed analytically;

5. Arrive at the optimal estimator of $\mathbf{p}$: $\hat{\mathbf{p}}_{opt} = (\hat{\mathbf{Q}}'\hat{\mathbf{W}}\hat{\mathbf{Q}})^{-1}(\hat{\mathbf{Q}}'\hat{\mathbf{W}}\mathbf{b})$.

Provided that the true parameter lies in the interior of the parameter space:

Assumption 6 (Interior) $\mathbf{p} \in (0,1)^K$ where K is the length of $\mathbf{p}$

The derivation of the asymptotic distribution of $\hat{\mathbf{p}}_{opt}$ is standard. Specifically,

Proposition 3 Under Assumptions DB, 1, 2, 3 and 6 for the sharp case, $\sqrt{N}(\hat{\mathbf{p}}_{opt} - \mathbf{p}) \Rightarrow \mathcal{N}(\mathbf{0},\Sigma)$ where $\Sigma^{-1} \equiv \mathbf{Q}'(\nabla_{\mathbf{q}} f \Omega \nabla_{\mathbf{q}} f')^{-1}\mathbf{Q}$ with Ω being the asymptotic variance matrix of $\mathbf{q}$ and $\nabla_{\mathbf{q}} f \equiv \nabla_{\mathbf{q}}(\mathbf{Qp} - \mathbf{b})$.

Analogously, for the imperfect compliance case, we have

Proposition 3F Under Assumptions DB, 1F, 2F, 3F, 4, 5 and 6 for the fuzzy case, $\sqrt{N}(\hat{\mathbf{p}}_{\mathbf{F}opt} - \mathbf{p_F}) \Rightarrow \mathcal{N}(\mathbf{0},\Sigma_{\mathbf{F}})$ where $\Sigma_{\mathbf{F}}^{-1} \equiv \mathbf{Q_F}'(\nabla_{\mathbf{q_F}} f_{\mathbf{F}} \Omega_{\mathbf{F}} \nabla_{\mathbf{q_F}} f_{\mathbf{F}}')^{-1}\mathbf{Q_F}$ with $\Omega_{\mathbf{F}}$ being the asymptotic variance matrix of $\mathbf{q_F}$ and $\nabla_{\mathbf{q_F}} f_{\mathbf{F}} \equiv \nabla_{\mathbf{q_F}}(\mathbf{Q_F p_F} - \mathbf{b_F})$.

The main conclusion of Kodde et al. (1990) shows that $\hat{\mathbf{p}}_{opt}$ is efficient – it has the same asymptotic variance as the MLE estimator – if $\mathbf{p}$ is exactly or over-identified by $f(\mathbf{q},\mathbf{p}) = 0$, and $\hat{\mathbf{q}}$ is the MLE estimator. Since

[19] It turns out that $\nabla_{\mathbf{q}} f \Omega \nabla_{\mathbf{q}} f'$ is singular, which is analogous to the issue raised in Satchachai and Schmidt (2008). Satchachai and Schmidt (2008) advise against using the generalized inverse, which is confirmed by our own numerical investigation. Instead, they propose dropping one or more restrictions, but state that the problem of which restrictions to drop has yet to be solved.

both conditions are satisfied in our setup, we have obtained a computationally inexpensive estimator without sacrificing efficiency. Also note that the assumptions can be jointly tested by the overidentifying restrictions as is standard for minimum distance estimators. In particular, the test statistic $N\cdot(f(\hat{\mathbf{q}},\hat{\mathbf{p}}_{opt})'\hat{\mathbf{W}}f(\hat{\mathbf{q}},\hat{\mathbf{p}}_{opt}))$ in the sharp case or $N\cdot(f(\hat{\mathbf{q}}_{\mathbf{F}},\hat{\mathbf{p}}_{\mathbf{F}opt})'\hat{\mathbf{W}}_{\mathbf{F}}f(\hat{\mathbf{q}}_{\mathbf{F}},\hat{\mathbf{p}}_{\mathbf{F}opt}))$ in the fuzzy case follows a χ^2-distribution with degrees of freedom equal to K_u, the number of points in the support of the measurement error, when assumptions in Proposition 3 or Proposition 3F are satisfied.[20]

A concern arises because the optimal estimators of p_k^1 and p_k^0 may never sum to one by following the procedure above. Therefore, we modify the estimation strategy and incorporate the matrix constraint $\mathbf{Rp}=\mathbf{c}$ with $\mathbf{c}=\begin{bmatrix}1\\1\end{bmatrix}$ and $\mathbf{R}=\begin{bmatrix}1&\cdots&1&0&\cdots&0\\0&\cdots&0&1&\cdots&1\end{bmatrix}$. In step 3, the consistent estimator is instead $\hat{\mathbf{p}}=(\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-1}\mathbf{R}'\{\mathbf{R}(\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-1}\mathbf{R}'\}^{-1}\mathbf{c}$,[21] and in step 5, $\hat{\mathbf{p}}_{opt}=(\hat{\mathbf{Q}}'\hat{\mathbf{W}}\hat{\mathbf{Q}})^{-1}\mathbf{R}'\{\mathbf{R}(\hat{\mathbf{Q}}'\hat{\mathbf{W}}\hat{\mathbf{Q}})^{-1}\mathbf{R}'\}^{-1}\mathbf{c}$. The asymptotic variance matrix of $\hat{\mathbf{p}}_{opt}$ is given by $\mathbf{T}((\mathbf{QT})'\mathbf{W}(\mathbf{QT}))^{-1}\mathbf{T}'$, where $\mathbf{T}$ is a matrix whose columns are the first $K-2$ eigenvectors of the projection matrix $\mathbf{I}-\mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{R}$. The computation for $\hat{\mathbf{W}}$ is unaltered by the imposition of the linear constraints $\mathbf{Rp}=\mathbf{c}$.

In order to construct the asymptotic distribution of the RD treatment effect estimators, we need to estimate the variance matrix of $\widehat{E[Y|X^*]}$ in the sharp RD case and $\widehat{E[D|X^*]}$ and $\widehat{E[Y|X^*]}$ in the fuzzy RD case for each point in the support of X^*. As shown in equations (A5), (A6), (A7) and (A8) in the Supplemental Material, the estimands are differentiable functions of $\Pr(X^*=x^*|D=d,Y=y)$, $\Pr(D=d|Y=y)$ and $\Pr(Y=1)$ for $d,y=0,1$. The delta method can be directly applied, where the Jacobian of the transformations is derived analytically. The general expressions of the estimators, $\hat{\delta}_{sharp}$ and $\hat{\delta}_{fuzzy}$, cannot be obtained because they depend on the the functional forms of $E[Y|X^*]$ and $E[D|X^*]$, which vary from application to application.

Finally, we note that the restriction to a binary Y can be relaxed. First, estimation proceeds in exactly the same way when Y takes on a finite number of values. Second, we show in Subsection A.5 of the Supplemental Material that our estimators can be further generalized to the case where Y is continuously distributed, though the procedure is likely to be computationally burdensome and numerically sensitive. In the numerical and empirical illustrations below, therefore, we adhere to a binary outcome variable.

[20]Identification in Subsection 3.2 is established under the symmetry assumption (Assumption 5), and the model is not identified in its absence. Therefore, this overidentification test may not have power against the violation of Assumption 5.

[21]For a clear exposition of this result, see the "makecns" entry in Stata (2010).

3.5 Potential Issues in Practical Implementation

There are several issues in implementing the procedure described in Subsection 3.4. First of all, in order to have realistic support for the true assignment variable, the maximum value of X needs to be significantly larger for the $D = 0$ group than for the $D = 1$ group, since their difference is the upper end point in the X^* distribution. Also, the left tail of the X distribution may need to be significantly longer that of the right tail for the $D = 1$ group, since their difference is the lower bound of the X^* distribution (following Assumption 5). Because symmetry is a functional form assumption, which may not hold when the assignment variable is measured in levels (e.g. in the case of income), a transformation may be needed. In practice, we use a Box-Cox transformation and experiment with various transformation parameters. The overidentification test mentioned in the previous subsection can us help decide which transformation parameters to use.

It is also possible, as mentioned in Remark 5, that an ineligible individual with a very large observed X may actually report program participation ($D = 1$) by mistake. If this is the case, the supports will not be correctly identified, and using a Box-Cox transformation will not be sufficient to correct the problem. A trimming procedure could be adopted in practice, which drops outliers in both the left and right tails of the $X|D = 1$ and $X|D = 0$ distributions. As with the case of transformation parameters, we try out several trimming percentages and examine the sensitivity of the empirical results. Finally, a quadratic programming routine with inequality constraints can be used in practice to guarantee nonnegativity of the probability masses.

3.6 Illustration with a Simple Numerical Example

In this subsection, we illustrate the proposed estimation procedure in Subsection 3.4 with a simple numerical example. We focus on the more complicated fuzzy case and show that the true first stage and outcome functions as well as the X^* distribution can indeed be recovered when the assumptions in Propositions 2F and 3F are met. In the baseline example, we generate X^* following a uniform distribution on the set of integers from -10 to 10. u follows a uniform distribution between -3 and 3 and is therefore symmetric in its support (Assumption 5). The true first stage relationship is given by

$$E[D|X^*] = (\alpha_{D^*X^*}X^* + \alpha_{D^*})1_{[X^*<0]} = \alpha_{D^*}D^* + \alpha_{D^*X^*}D^*X^*, \tag{10}$$

which reflects the one-sided fuzzy assumption (Assumption 4), and the size of the first stage discontinuity is α_{D^*}. The outcome response function is given by the simple constant treatment effect specification

$$E[Y|X^*, D] = \delta_0 + \delta_1 X^* + \delta_{fuzzy} D \tag{11}$$

where the treatment effect to be identified is δ_{fuzzy}. Note that equations (10) and (11) together imply that the outcome relationship is

$$E[Y|X^*] = \beta_0 + \beta_{D^*} D^* + \beta_1 X^* + \beta_{D^*X^*} D^* X^* \tag{12}$$

where $\beta_0 = \delta_0$, $\beta_{D^*} = \alpha_{D^*} \cdot \delta_{fuzzy}$, $\beta_1 = \delta_1$ and $\beta_{D^*X^*} = \alpha_{D^*X^*} \cdot \delta_{fuzzy}$.

Figures 2 and 3 present graphical results based on a sample of 25,000 generated observations under the parameter values $\alpha_{D^*X^*} = -0.01$, $\alpha_{D^*} = 0.8$, $\delta_0 = 0.15$, $\delta_1 = -0.01$, and $\delta_{fuzzy} = 0.6$, with the implied coefficients in (12) being $\beta_0 = 0.15$, $\beta_{D^*} = 0.48$, $\beta_1 = -0.01$ and $\beta_{D^*X^*} = -0.006$. We choose $N = 25,000$ because it is about the average sample size in the relevant studies – 45,722 in Hullegie and Klein (2010), 34,949 in Koch (2013), 11,541 in Schanzenbach (2009) and 2,954 in de la Mata (2012). The top and bottom panels in Figure 2 plot the *observed* first stage and outcome relationships, i.e. $E[D|X]$ and $E[Y|X]$, respectively. Note that there is no visible discontinuity at the threshold, and the estimated first stage and outcome discontinuities based on these observed relationships cannot identify the true parameter values of α_{D^*} and β_{D^*}, which are 0.8 and 0.48 respectively.

Figure 3 plots the *estimated* first stage and outcome relationships based on procedures developed in Subsection 3.4 against the actual (10) and (11) specified with the parameter values above. As is evident from the graphs, the proposed procedures can correctly recover the $E[D|X^*]$ and $E[Y|X^*]$ of the underlying RD design. $\hat{\hat{\delta}}_{fuzzy}$, the RD treatment effect parameter, is obtained by fitting another linear minimum distance procedure on the estimated $E[D|X^*]$ and $E[Y|X^*]$ (as well as their estimated variance matrices) with the parametric restrictions (10) and (12). In 1,000 repeated samples, the average point estimate for α_{D^*} is 0.75 (true parameter value is 0.8), the average standard error is 0.063, and the coverage rate of the 95% confidence interval is 97%; the average point estimate for β_{D^*} is 0.48 (true parameter value is 0.48), the average standard error is 0.075 and the coverage rate of the 95% confidence interval is 98%.

To gauge the performances of the estimators in adverse settings, we test their sensitivity to 1) the violation of symmetry and 2) larger supports in X^* and u relative to the sample size. Unfortunately, the performance of the estimators quickly deteriorates when the symmetry assumption is violated, which is an important limitation of the proposed method. For example, when u is supported on the integers in $[-4,3]$ but the lower bound of its support is erroneously assumed to be -3, the average point estimates for α_{D^*} and β_{D^*} are 0.66 and 0.39, and the coverage rates of the 95% confidence interval are 0.56 and 0.74, respectively. In comparison, the behavior of the estimators is more robust with respect to alternative specifications of support_X: as the set of points in support_X increases from 27 in the numerical example above to 37, the coverage rates are still around 95%. The coverage rates fall to around 80% and 55% when $|\text{support}_X|$ is 47 and 67, respectively.

As mentioned above, the proposed procedure can also be used to assess the degree of nonsmoothness in the X^* distribution at the eligibility threshold. In another numerical exercise, we consider two alternative specifications: 1) the specification used in Figures 2 and 3, for which there is no discontinuity in the X^* distribution at the eligibility threshold; 2) X^* is still supported on the set of integers from -10 to 10 but with a discontinuity at the eligibility threshold: $\Pr(X^* = i) = 0.06$ for each $i < 0$ and $\Pr(X^* = i) = 0.036$ for $i \geqslant 0$. Figures 4 and 5 present the observed X and *estimated* X^* distribution for cases 1) and 2), respectively. Note that there is no obvious discontinuity in the observed X distribution at the eligibility threshold in case 2) (top panel of Figure 5), which again illustrates the problematic nature of using the observed assignment variable X for RD analyses. In both cases, we test for the threshold discontinuity by fitting a linear minimum distance procedure on the estimated X^* distribution with the restriction

$$\Pr(X^* = x^*) = \gamma_0 + \gamma_{D^*} D^* + \gamma_1 X^* + \gamma_{D^*X^*} D^* X^*.$$

Let $\gamma_{D^*}^{1)}$ and $\gamma_{D^*}^{2)}$ be the coefficients of γ_{D^*} in cases 1) and 2) respectively, and based on the specifications above, $\gamma_{D^*}^{1)} = 0$ and $\gamma_{D^*}^{2)} = 0.024$. In 10,000 repeated samples with 25,000 observations, the average value of $\gamma_{D^*}^{1)}$ is -0.003, and the coverage rate of the 95% confidence interval is 89%; the average value of the $\gamma_{D^*}^{2)}$ estimates is 0.031, and the coverage rate of the 95% confidence interval is 87%. The coverage rates for the density discontinuity confidence intervals improve to 90% or higher as the sample size exceeds 50,000. Overall, this simple numerical example verifies that the true assignment variable distribution and the RD treatment effect parameter can indeed be recovered using the proposed method, when the model assumptions are met.

4 Continuous Assignment Variable and Measurement Error

In this section, we study the identification in an RD design when X^* and u are continuously distributed and discuss the sharp and fuzzy cases separately in the two subsections below.[22] Before we proceed, we introduce the analog of Assumption 2 in the continuous case, a standard assumption in the RD literature.

Assumption 2C (Positive Density at Threshold) Let the p.d.f. of X^* be f_{X^*}. There exists $a > 0$ such that $f_{X^*}(x^*) > 0$ for $x^* \in (-a, a)$.

Assumption 2C ensures that the conditional expectation functions are well-defined in the sharp and fuzzy RD estimands as defined in Section 2.

4.1 Identification under Perfect Compliance

In this subsection, we focus on the case of perfect compliance and consider three distinct approaches.

Approach 1. In the first approach, we assume that u follows a normal distribution with mean zero and an unknown variance σ^2. Along with the classical measurement error assumption, the normality of u allows the distribution X^* to be identified. Provided that Assumption 2C holds and that the measurement error is nondifferential, the RD treatment effect is also identified.

Assumption 7 (Normality) $u \sim \phi(0, \sigma^2)$.

Proposition 4 (a) Under Assumptions 1 and 7, the distributions of X^* and u are identified. (b) Under Assumptions 1Y, 2C and 7, δ_{sharp} is identified.

Proposition 4(a) is a corollary of Theorem 2.1 of Schwarz and Bellegem (2010), and the proof of Proposition 4(b) is similar to that of Proposition 2. Details are provided in Subsection A.1.4 of the Supplemental Material.

Approach 2. In the second approach, we show that normality of the measurement error can be weakened without sacrificing the identification of the RD treatment effect under perfect compliance. Specifically, an identification-at-infinity strategy can be applied with a regularity condition governing the tail behavior of the measurement error.

[22] A case not considered in this paper is when X^* is continuous but u is discrete (or vice versa). A starting point in this setting is Dong and Lewbel (2011).

Proposition 5 If Assumption 2C holds, f_{X^*} is continuous at 0, and the c.d.f. of u, F_u, satisfies

$$\lim_{x\to\infty}\frac{1-F_u(x+v)}{1-F_u(x)}=0 \qquad \text{for all } v>0, \tag{13}$$

$$\lim_{x\to-\infty}\frac{F_u(x-v)}{F_u(x)}=0 \qquad \text{for all } v>0, \tag{14}$$

then the RD treatment effect is identified:

$$\lim_{x\to\infty}E[Y|X=x,D^*=1]-\lim_{x\to-\infty}E[Y|X=x,D^*=0]=\delta_{sharp}. \tag{15}$$

We prove Proposition 5 in Subsection A.1.5 of the Supplemental Material. Intuitively, when we see an observation with $X\gg 0$ and $D^*=1$, it is because X^* is close to zero or because u is very large. Condition (13) states that the right tail of the measurement error distribution needs to be "light" enough and makes u unlikely to be very large. Therefore, as X becomes large in the treatment group, we end up with observations for which X^* is just below zero. While the Laplace distribution violates (13) and (14), almost all of the common distributions whose tails are comparable to or lighter than the normal distribution satisfy these tail conditions.[23]

A visual illustration of the identification result in Proposition 5 is presented in Figure 6. We use the same conditional expectation function for $E[Y|X^*]$ as in Subsection 3.6. X^*, u and the error term in the outcome equation are all normally distributed. We split the simulated sample of 25,000 observations into two groups by the value of D^*. In the upper and lower panels, we show the raw scatter plot for the $D^*=1$ and $D^*=0$ groups, respectively. In each panel, we impose fit obtained from a simple local regression. As X becomes large in the $D^*=1$ group and as X becomes small in the $D^*=0$ group, we can see that the averages of the Y's (represented by the curved lines) approach the true intercepts of the $E[Y|X^*]$ function (represented by the horizontal lines).

We now discuss the relationship between the identification-at-infinity strategy proposed here and recent papers by Yu (2012) and Yanagi (2014) that study the same question. Yu (2012) assumes that the measurement error variance shrinks to zero as the sample size increases and proposes using a trimmed sample to recover the RD treatment effect, in which observations with $X\geqslant 0$ in the $D^*=1$ group and observations

[23] When u follows a Laplace distribution, e.g., $f_u(v)=\frac{1}{2}e^{-|v|}$ for $v\in\mathbb{R}$, $\lim_{x\to\infty}E[Y|X=x,D^*=1]=\int_0^\infty E[Y|X^*=-v]e^{-v}dv$, which is generally not equal to $\lim_{x^*\uparrow 0}E[Y|X^*=x^*]$.

with $X < 0$ in the $D^* = 0$ group are dropped. Specifically, the RD treatment effect estimand is the difference between the left intercept of $E[Y|X = x, D^* = 1]$ and the right intercept of $E[Y|X = x, D^* = 0]$. As mentioned in Section 2, dividing the population into treatment and control and separately estimating the intercept in each is attractive, because it results in a much smaller bias than naively fitting $E[Y|X]$ with local regressions. Our approach builds on the idea of Yu (2012) in this regard. The difference is that we put no restriction on the measurement error variance, and it may result in a smaller bias as a result. To see this, we illustrate the strategy of Yu (2012) in Figure 7, and the bias appears to be somewhat larger as compared to that in Figure 6.

Assuming bounded measurement error u, Yanagi (2014) creatively applies a small error variance approximation (Chesher (1991)) to $E[Y|X^* = x, D^* = d]$ in a sharp design:

$$\begin{aligned} E[Y|X^* = x, D^* = d] &= E[Y|X = x, D^* = d] \\ &\quad -\sigma^2 \left(\log^{(1)} f_{X|D}(x|d)\right) E^{(1)}[Y|X = x, D^* = d] \\ &\quad -\frac{\sigma^2}{2} E^{(2)}[Y|X = x, D^* = d] + o(\sigma^2) \end{aligned} \tag{16}$$

where σ^2 again denotes the variance of u. For the approximation to work, however, the derivatives of the conditional expectation and density functions need to be estimated. Moreover, σ needs to be small and known.

To illustrate how our identification results relate to Yanagi (2014), first note that the proof of Proposition 5 can be carried over to the case with bounded measurement error u. In fact, when support_u is bounded, the tail conditions (13) and (14) are no longer needed. Furthermore, we can identify $\text{support}_u \equiv [\underline{u}, \bar{u}]$ and identify the RD treatment effect with

$$\lim_{x \to \bar{u}} E[Y|X = x, D^* = 1] - \lim_{x \to \underline{u}} E[Y|X = x, D^* = 0] = \delta_{sharp}. \tag{17}$$

This "identification-at-boundary" strategy of (17) has the advantage of avoiding the derivatives in (16), which may be hard to estimate in practice. Furthermore, it does not place additional restrictions on the measurement error distribution, nor does it require σ to be known.

Approach 3. When σ *is* known (e.g. from an external source), we can also apply the SIMEX strategy of Cook and Stefanski (1994) and Carroll et al. (1999) to recover the RD treatment effect. It is the final approach we propose in this subsection.[24] The idea behind the method is most simply illustrated in the linear case where the conditional expectation function $E[Y|X^* = x^*, D^* = 1]$ is equal to $\psi_0 + \psi_1 X^*$. When we regress Y on X within the $D^* = 1$ population, the least squares slope estimator converges to the expression with the well-known attenuation factor, $\psi_1 \cdot \frac{var(X^*|D^*=1)}{var(X^*|D^*=1)+\sigma^2}$. The attenuation problem worsens as σ^2 increases, and the SIMEX approach of Cook and Stefanski (1994) makes use of this observation: it traces out the estimate as a function of the degree of measurement error contamination via simulation, and extrapolates this function to recover the true parameter ψ_1 as well as the target conditional expectation function. Formally, for any $\lambda > 0$, we can add additional noise $\tilde{u}_\lambda$ with variance $\sigma^2\lambda$ to X. The new measurement error is $u + \tilde{u}_\lambda$ with variance $\sigma^2(1+\lambda)$, and the corresponding population slope parameter is $g(\lambda) \equiv \psi_1 \cdot \frac{var(X^*|D^*=1)}{var(X^*|D^*=1)+\sigma^2(1+\lambda)}$. By choosing different λ's, we obtain the value of the g function at various points, which we can then use to extrapolate and recover $g(-1) = \psi_1$.

The linearity of $E[Y|X^* = x^*, D^* = 1]$ in the example above can be relaxed. As long as this target function is smooth, we can apply the nonparametric procedure proposed by Carroll et al. (1999). Specifically, for each λ we choose, we can estimate the value of the conditional expectation function of $\mu_d(\lambda) = E[Y|X + \tilde{u}_\lambda = 0, D^* = d]$ via a local linear regression, and use polynomials to extrapolate $\mu_d(\lambda)$ back to $\lambda = -1$ and recover the left ($d = 1$) and right ($d = 0$) intercepts in the RD estimand. The difference between μ_1 and μ_0 at $\lambda = -1$ identifies δ_{sharp}. In the next subsection, we consider whether and how the three approaches can be extended to a design with imperfect compliance.

4.2 Identification under Imperfect Compliance

Only one of the three approaches proposed in Subsection 4.1 carries over to an RD design with imperfect compliance. With the one-sided fuzzy assumption, Approach 1 still identifies the X^* distribution and the RD treatment effect δ_{fuzzy}. The idea is similar in spirit to that of Proposition 1F, which relies on Assumptions DB, 1F, 2F, 3F, 4 and 5. Analogous to Proposition 2F, we can identify the X^* distribution for each value of D and Y under strong independence and nondifferential measurement error, which allows the identification of $E[D|X^*]$ and $E[Y|X^*]$ by Bayes' Theorem and hence the RD treatment effect.

Proposition 4F (a) Under Assumptions 1F, 4 and 7, the distributions of X^* and u are identified. (b) Under Assumptions 1FY, 2C, 4 and 7, δ_{fuzzy} is identified.

[24]We thank a referee for suggesting this approach. See Chapter 5 of Carroll et al. (2006) for a detailed overview of the SIMEX method in the measurement error context.

Remark 7. Assumption 4 (one-sided fuzzy) is no longer needed if the researcher knows what σ is, which parallels the redundancy of Assumption 4 in Proposition 2F when the support of the measurement error is known.

Remark 8. The normal measurement error assumption plays a key role in Proposition 4F: it encapsulates symmetry (a stronger version of Assumption 5) and the tail restriction normality imposes renders the continuous analog of Assumption 2F unnecessary in the identification of the X^* distribution.

Unfortunately, Approaches 2 and 3 do not carry over to the fuzzy case, even with the one-sided fuzzy assumption. For the identification-at-infinity/boundary strategy (Approach 2), as X approaches $-\infty$, the control group ($D = 0$) consists of both compliers with X^* close to the threshold and never takers with a large negative X^*. Absent strong assumptions, it is not possible to disentangle the two groups and identify $\lim_{x^* \downarrow 0} E[Y|X^* = x^*]$. The SIMEX strategy (Approach 3) will not recover the discontinuity in the first stage relationship $E[D|X^*]$. Without observing D^*, it is not possible to extrapolate the left and right intercepts in the first stage relationship from separate regressions. Applying SIMEX to only the observed first stage relationship $E[D|X]$ also will not work because the target function $E[D|X^*]$ is potentially discontinuous at zero, a violation of the underlying smoothness assumption of SIMEX. Despite being more nonparametric, the inability of Approaches 2 and 3 to identify the RD treatment effect in a fuzzy design greatly limits their practical relevance.

4.3 Estimation

In this subsection, we discuss estimation for the three approaches introduced in Subsection 4.1. For Approach 1, Schwarz and Bellegem (2010) propose a conceptual minimum distance framework. Applied to our context, it will select the estimators for the distributions of $X^*|D^* = 1$ and u to fit the observed characteristic function of $X|D^* = 1$. The fact that the distribution of X^* given $D^* = 1$ has no support on the entire interval of $(0, \infty)$ satisfies the regularity condition of Schwarz and Bellegem (2010), which proves the consistency of the conceptual estimators. As for practical implementation, Schwarz and Bellegem (2010) suggest discretizing the $X^*|D^* = 1$ distribution with the number of support points increasing as $n \to \infty$, but defer the details to future research. Assuming that the Lapalace transform of the target density decays sufficiently quickly, Matias (2002) proposes an explicit estimator for σ and shows that the rate of convergence is slower than $\log n$. Matias (2002) also proves that the minimax mean squared error of the pointwise density estimator cannot decrease faster than $(\log n)^{-1}$ uniformly over a set of regular densities, which makes is difficult to apply empirically.

For Approach 2, one can adapt the estimator proposed by Andrews and Schafgans (1998) as a starting point for estimation and inference. Let $s(x)$ be a weight function that takes the value 0 for $x \leq 0$, the value 1 for $x \geq b$ for some positive constant b, and is smoothly increasing between 0 and b with bounded derivatives. γ_n is the bandwidth parameter, which goes to infinity with the sample size n. The Andrews-Schafgans estimator is defined as

$$\hat{\kappa}_1 = \frac{\sum_{i=1}^{N} Y_i D_i s(X_i - \gamma_N)}{\sum_{i=1}^{N} D_i s(X_i - \gamma_N)},$$

where $\kappa_1 \equiv \lim_{x\to\infty} E[Y|X = x, D^* = 1]$ is the quantity to be estimated. The consistency result $\hat{\kappa}_1 \xrightarrow{p} \kappa_1$ in Andrews and Schafgans (1998) is established under similar assumptions as in this paper, along with additional regularity conditions requiring 1) finite moments of the random variables and 2) an upper bound on the rate at which γ_N tends to infinity, which depends on the upper tail behavior of the X distribution. Furthermore, with one extra condition setting a lower bound for γ_N as $N \to \infty$, Andrews and Schafgans (1998) also show that the distribution of the estimator $\hat{\kappa}_1$ is asymptotically normal and centered around κ_1. The rate of convergence is $\sqrt{N}$ multiplied by a factor that depends on the tail of the X distribution and the exact choice of γ_N.

For Approach 3, Carroll et al. (1999) show that the error of their local linear nonparametric SIMEX estimator with bandwidth h is of order $O_p\{h^2 + (Nh)^{-\frac{1}{2}}\}$, provided that the polynomial extrapolant is exact. However, the performance of the estimator is sensitive in practice. Berry et al. (2002) note that alternative choices of smoothing parameters may give rise to great instability in estimates, and Staudenmayer and Ruppert (2004) find that the SIMEX estimates are not robust to alternative values within the 95% confidence interval for σ, which is estimated in an external validation study. Given the challenges in estimation described in this section, we will adopt a parametric approach in the subsequent empirical illustration.

5 Empirical Application: Medicaid Takeup and Crowdout

As one of the largest entitlement programs in the United States, Medicaid has received considerable attention in policy discussions. An important debate is on the extent to which Medicaid eligibility crowds out private insurance – that is, what fraction of eligible individuals drop private insurance and enroll in Medicaid. In a seminal paper, Cutler and Gruber (1996) use simulated instruments and estimate that Medicaid reduced private insurance coverage for children by 30 to 40 percent when Medicaid eligibility greatly expanded between 1988 and 1993. This high crowdout rate, however, is not the consensus of the literature. Several subsequent studies (e.g., Thorpe and Florence (1998), Yazici and Kaestner (2000) and Card and Shore-Sheppard (2004)) find smaller effects for the same time period using various research designs. As noted by Shore-Sheppard (2008), the question of crowdout has produced less consensus than disagreement.

In this section, we illustrate our proposed methods by using them to estimate Medicaid takeup and private insurance crowdout. Specifically, we apply an RD design by exploiting the Medicaid income eligibility cutoff rule. We use microdata constructed by Card and Shore-Sheppard (2004) from the full panel research files of the 1990-93 SIPP and make further restrictions to arrive at our analysis sample, as detailed in Subsection B.1 of the Supplemental Material.[25]

We apply two modeling approaches to estimate Medicaid takeup and private insurance crowdout around the income eligibility threshold. First, we use the proposed method in Section 3 based on a discretized income measure. Because of the difficulty in implementing the semi/nonparametric estimators in Section 4, in our second modeling strategy we treat income as continuous and adopt a parametric MLE framework. In particular, we specify 1) X^* as a function of a normal random variable that allows for bunching at the income eligibility threshold, 2) u to be mean zero normal, and 3) the one-sided fuzzy first stage and outcome relationships to be logistic in polynomials of X^* and D. Details of this parametric formulation are provided in Subsection B.2 of the Supplemental Material.

As mentioned in Subsection 3.5, we need to transform the income variable in order for Assumptions 1FY (strong independence) and 5 (symmetry in the support of u) to plausibly hold. Through experimentation, we find that a Box-Cox transformation with parameter ρ between 0.3 and 0.35 appears to be consistent with these assumptions as indicated by the overidentification tests. Therefore, we present the main results using $\rho = 0.33$, but also present estimates under other values of ρ as robustness checks. To help understand the meaning of the transformed assignment variable, Table A.1 in the Supplemental Material provides a mapping between the transformed income measures that are normalized against eligibility cutoffs and the actual family income amount. For example, a family with a child facing the 100% federal poverty line (FPL)

[25]Card and Shore-Sheppard (2004) study public health insurance coverage and private insurance crowdout by combining RD and difference-in-differences designs and leverage off both the age and income eligibility rules for Medicaid.

has an actual monthly income of \$1,117 (1991 dollars) if the transformed and normalized income measure is 0; the family's actual income is \$1,231 if the transformed and normalized income measure takes on the value 1. In the remainder of the section, we refer to this transformed and normalized income variable simply as income.

Figures 8a through 8c plot the density of the *observed* income, the first stage relationship between Medicaid coverage and income, and the outcome relationship between private insurance coverage and income. The distribution of income in Figure 8a is approximately symmetric and shows no visible sign of bunching around the cutoff. In Figure 8b, Medicaid coverage decreases as income rises, but there is no discontinuous drop-off at the cutoff, which is to be expected given the presence of measurement error. Correspondingly, Figure 8c shows that while private insurance coverage increases with income, the relationship is again smooth through the cutoff.

Figures 8d through 8f plot the *estimated true* income density, first stage, and outcome relationship. Estimates from the discrete and continuous models are juxtaposed together for ease of comparison. In Figure 8d, while the discrete approach yields a much noisier estimate of the density than the parametric continuous approach, the two models roughly agree for much of the distribution, especially near the cutoff. Turning to Figure 8e, we again find that the first stage estimated using the discrete model displays a similar trend as the continuous model near the cutoff, albeit with more noise. The graphical evidence suggests that Medicaid coverage increases about 10 to 20 percentage points at the income threshold.[26] Finally, the behavior of the discrete model estimates in Figure 8f matches their continuous counterparts close to the threshold, and no visible discontinuity is detected.

The numerical discontinuity estimates are presented in Tables A.2 and A.3 in the Supplemental Material for the discrete and continuous models, respectively. Consistent with visual evidence, in Table A.2 we generally find a statistically insignificant discontinuity in the income distribution, a significant discontinuity in Medicaid coverage between 15 and 25 percentage points, and insignificant discontinuity estimates for private insurance coverage. The resulting crowdout estimates are insignificant as well, but they are very imprecise – the 95% confidence intervals contain practically all estimates from the literature. In column (5), we report the p-values from the overidentification tests: since none of the models in the table are formally

[26]The one-sided fuzzy assumption is plausible in the Medicaid context. According to CMS (2014), only 3.1% of the cases were incorrectly assigned eligibility status when families applied to public health insurance, and the trimming we conduct further alleviates the concern. We should note, however, that the ongoing Medicaid participants might not have their income eligibility recertified every month, potentially casting doubt on the one-sided fuzzy assumption, but there is little evidence of income rebounding for these participants as documented by Pei (Forthcoming).

rejected at the 5% level, the model appears to fit the data reasonably well.

For brevity, we only present the continuous model results using the Box-Cox transformation parameter $\lambda = 0.33$ and 1% trimming in Table A.3. We find no evidence of bunching in the income distribution, a statistically significant 12.5% Medicaid takeup rate just below the eligibility cutoff, and a small and statistically insignificant crowdout effect. The parsimonious specification in the continuous model leads to much improved precision in the estimates, and even the upper bound of the 95% confidence interval excludes the crowdout estimates from Cutler and Gruber (1996). Finally, as documented in Subsection B.3 of the Supplemental Material, the continuous model provides reasonable fit to the data.

To summarize, the two models are broadly consistent in their estimates. We find that the Medicaid takeup rate just below the eligibility cutoff falls between 10 and 25 percent and that there is little evidence of income sorting around the threshold and the private insurance crowdout. However, due to the lack of functional form restrictions imposed by the discrete model, its estimates are quite imprecise.

6 Conclusion

This paper investigates identification and estimation in an RD design where the assignment variable is measured with error. This is a challenging problem because the presence of measurement error may smooth out the first stage discontinuity and eliminate the source of identification. We provide conditions for the identification of the true assignment variable distribution and the RD treatment effect, using only the mismeasured assignment variable and the binary treatment status. Our results contribute to the measurement error literature and shed light on the limitations of certain RD applications.

We first study the case where the assignment variable and the measurement error are discrete, and propose sufficient conditions for identification in both the sharp and fuzzy designs. A simple estimation procedure is proposed using a minimum distance framework. Following standard arguments, the resulting estimators are $\sqrt{N}$-consistent, asymptotically normal, and efficient. A numerical example verifies that when the model assumptions are met, the true assignment variable distribution and the RD treatment effect parameter can indeed be recovered using the proposed method.

We also explore the case where the assignment variable and measurement error are continuous and propose three different identification approaches. The first approach assumes normality of the measurement error, and identifies the assignment variable distribution and the RD treatment effect in sharp and one-sided fuzzy designs. The second approach adopts a novel identification-at-infinity strategy, and the third approach relies on the SIMEX method of Carroll et al. (1999), both of which identify the RD treatment effect in a sharp design. Because the first approach accommodates a fuzzy design, it potentially works for a larger range of applications, but it is handicapped by the slow convergence rate of the semiparametric estimator.

In our empirical application, we apply an RD design to estimate Medicaid takeup and private insurance crowdout using SIPP data. We exploit the discontinuity in the eligibility formula with respect to family income, but because income is measured with error, the first stage relationship between Medicaid takeup and reported income is not discontinuous at the eligibility cutoff. We use two approaches to recover the true income distribution and the RD treatment effect: our proposed method based on discretized income and a *parametric* MLE framework that treats income as continuous. The two approaches yield similar results: Medicaid takeup rate for the barely eligible is between 10 and 25 percent and there is little evidence of income sorting around the threshold and of private insurance crowdout. However, the estimates from the discrete approach are quite imprecise, which by itself is unlikely to deliver convincing policy conclusions. Therefore, we believe the parametric approaches adopted by this paper for classical measurement error and by Hullegie and Klein (2010) for Berkson measurement error to be more appealing alternatives for empirical research.

References

Abowd, John M. and Martha H. Stinson, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *The Review of Economics and Statistics*, 2013, *95* (5), 1451–1467.

Almond, Douglas, Joseph J. Doyle Jr., Amanda E. Kowalski, and Heidi Williams, "The Role of Hospital Heterogeneity in Measuring Marginal Returns to Medical Care: A Reply to Barreca, Guldi, Lindo, and Waddell," *Quarterly Journal of Economics*, November 2011, *126* (4), 2125–2131.

Andrews, Donald W. K.. and Marcia M. A. Schafgans, "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 1998, *65*, 497–517.

Ashenfelter, Orley and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins," *The American Economic Review*, 1994, *84* (5), 1157–1173.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell, "Heaping-Induced Bias in Regression-Discontinuity Designs," *Economic Inquiry*, 2016, *54* (1), 268–293.

_, Melanie Guldi, Jason M. Lindo, and Glen R. Waddell, "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification," *Quarterly Journal of Economics*, October 2011, *126* (4), 2117–2123.

Battistin, Erich, Agar Brugiavini, Enrico Rettore, and Guglielmo Weber, "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach," *American Economic Review*, 2009, *99* (5), 2209–2266.

Berry, Scott M., Raymond J. Carroll, and David Ruppert, "Bayesian Smoothing and Regression Splines for Measurement Error Problems," *Journal of the American Statistical Association*, 2002, *97* (457), 160–169.

Black, Dan A., Mark C. Berger, and Frank A. Scott, "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 2000, *95* (451), 739–748.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement Error in Survey Data," in J.J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, chapter 59, pp. 3705–3843.

Card, David and Lara D. Shore-Sheppard, "Using Discontinuous Eligibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low-Income Children," *The Review of Economics and Statistics*, 2004, *86* (3), 752–766.

_ , Andrew K. G. Hildreth, and Lara D. Shore-Sheppard, "The Measurement of Medicaid Coverage in the SIPP: Evidence from a Comparison of Matched Records," *Journal of Business & Economic Statistics*, 2004, *22* (4), 410–420.

_ , David S. Lee, Zhuan Pei, and Andrea Weber, "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, November 2015, *83* (6), 2453–2483.

Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceaunu, *Measurement Error in Nonlinear Models: A Modern Perspective*, 2 ed., Chapman & Hall/CRC, 2006.

_ , Jeffrey D. Maca, and David Ruppert, "Nonparametric Regression in the Presence of Measurement Error," *Biometrika*, 1999, *86* (3), 541–554.

Chen, Xiaohong, Yingyao Hu, and Arthur Lewbel, "Nonparametric Identification and Estimation of Nonclassical Errors-in-Variables Models without Additional Information," *Statistica Sinica*, 2009, *19*, 949–968.

Chesher, Andrew, "The Effect of Measurement Error," *Biometrika*, 1991, *78* (3), 451–462.

CMS, "Medicaid and CHIP 2014 Improper Payments Report," Technical Report, Center for Medicare and Medicaid Services 2014.

Cook, J.R. and Leonard A. Stefanski, "Simulation-Extrapolation Estimation in Parametric Measurement Error Models," *Journal of the American Statistical Association*, 1994, *89* (428), 1314–1328.

Currie, Janet, "The Take-up of Social Benefits," in Alan Auerbach, David Card, and John Quigley, eds., *Public Policy and the Income Distribution*, New York: Russell Sage, 2006, chapter 3, pp. 80–148.

Cutler, David M. and Jonathan Gruber, "Does Public Insurance Crowd Out Private Insurance," *Quarterly Journal of Economics*, 1996, *111* (2), 391–430.

Davezies, Laurent and Thomas Le Barbanchon, "Regression Discontinuity Design with Continuous Measurement Error in the Running Variable," November 2014. Working Paper.

de la Mata, Dolores, "The Effect of Medicaid Eligibility on Coverage, Utilization, and Children's Health," *Health Economics*, 2012, *21* (9), 1061–1079.

DiNardo, John and David S. Lee, "Program Evaluation and Research Designs," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Part A, Elsevier, 2011, chapter 5, pp. 463–536.

Dong, Yingying, "Regression Discontinuity Applications With Rounding Errors in the Running Variable," *Journal of Applied Econometrics*, 2015, *30*, 422–446.

_ **and Arthur Lewbel**, "Nonparametric Identification of a Binary Random Factor in Cross Section Data," *Journal of Econometrics*, 2011, *163* (2), 163–171.

Grinstead, Charles M. and J. Laurie Snell, *Introduction to Probability*, second revised ed., American Mathematical Society, July 1997.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Hausman, Jerry A., Whitney K. Newey, Hidehiko Ichimura, and James L. Powell, "Identification and Estimation of Polynomial Errors-in-variables Models," *Journal of Econometrics*, 1991, *50* (3), 273–295.

Hullegie, Patrick and Tobias J. Klein, "The Effect of Private Health Insurance on Medical Care Utilization and Self-assessed Health in Germany," *Health Economics*, 2010, *19* (9), 1048–1062.

Jales, Hugo and Zhengfei Yu, "Identification and Estimation in Density Discontinuity Designs," 2016. Working Paper.

Kleven, Henrik J. and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics*, 2013, *128* (2), 669–723.

Koch, Thomas G., "Using RD Design to Understand Heterogeneity in Health Insurance Crowd-out," *Journal of Health Economics*, 2013, *32* (3), 599–611.

Kodde, D. A., F. C. Plam, and G. A. Pfann, "Asymptotic Least-squares Estimation Efficiency Considerations and Applications," *Journal of Applied Econometrics*, 1990, *5* (3), 229–243.

Lee, David S. and David Card, "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 2008, *142* (2), 655–674. The regression discontinuity design: Theory and applications.

_ **and Thomas Lemieux**, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, June 2010, *48*, 281–355.

Lewbel, Arthur, "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 2007, *75* (2), 537–551.

Li, Qi and Jeffery Scott Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, 2007.

Li, Tong and Quang Vuong, "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 1998, *65* (2), 139–165.

Mahajan, Aprajit, "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 2006, *74* (3), 631–665.

Marquis, Kent H. and Jefferey C. Moore, "Measurement Errors in SIPP Program Reports," Technical Report, U.S. Census Bureau 1990.

Matias, Catherine, "Semiparametric Deconvolution with Unknown Noise Variance," *ESAIM: Probability and Statistics*, 2002, *6*, 271–292.

Pedace, Roberto and Nancy Bates, "Using Administrative Records to Assess Earnings Reporting Error in the Survey of Income and Program Participation," *Journal of Economic and Social Measurement*, 2000, *26* (3), 173–192.

Pei, Zhuan, "Eligibility Recertification and Dynamic Opt-in Incentives in Income-tested Social Programs: Evidence from Medicaid/CHIP," *American Economic Journal: Economic Policy*, Forthcoming.

_ **and Yi Shen**, "Supplemental Material to "The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable"," 2016. https://sites.google.com/site/peizhuan/files/rdme-supplemental-material.pdf.

Ruckdeschel, Peter and Matthias Khol, "General Purpose Convolution Algorithm in S4 Classes by Means of FFT," *Journal of Statistical Software*, 2014, *59* (4), 1–25.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2(3)*, 180–212.

Satchachai, Panutat and Peter Schmidt, "GMM with More Moment Conditions than Observations," *Economics Letters*, 2008, *99* (2), 252–255.

Schanzenbach, Diane Whitmore, "Do School Lunches Contribute to Childhood Obesity?," *J. Human Resources*, 2009, *44* (3), 684–709.

Schennach, Susanne M., "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 2004, *72* (1), 33–75.

Schwarz, Maik and Sebastien Van Bellegem, "Consistent Density Deconvolution under Partially Known Error Distribution," *Statistics and Probability Letters*, 2010, *80*, 236–241.

Shore-Sheppard, Lara D., "Stemming the Tide? The Effect of Expanding Medicaid Eligibility On Health Insurance Coverage," *The B.E. Journal of Economic Analysis & Policy: Advances*, 2008, *8*.

Stata, *Base Reference Manual: Stata 11 Documentation* Stata Press 2010.

Staudenmayer, John and David Ruppert, "Local Polynomial Regression and Simulation-extrapolation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2004, *66* (1), 17–30.

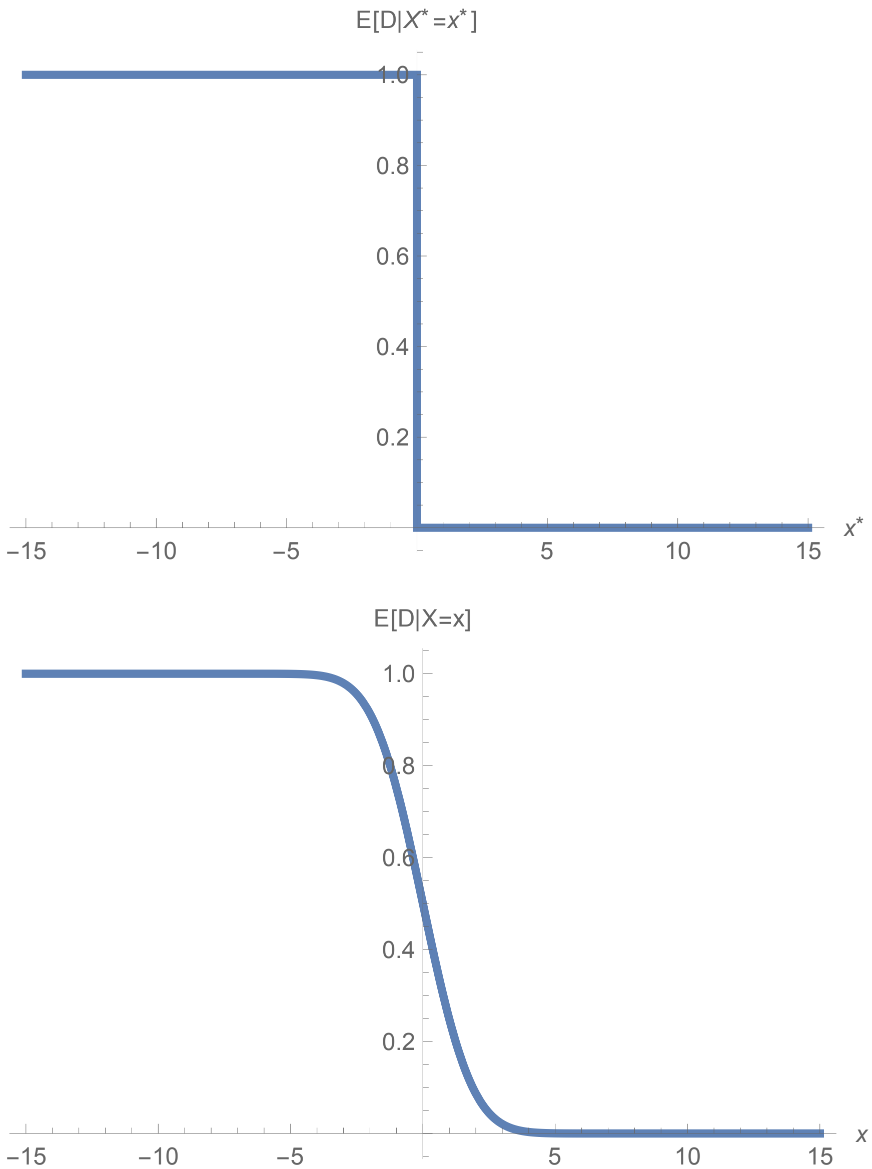
Thorpe, Kenneth E. and Curtis S. Florence, "Health Insurance among Children: The Role of Expanded Medicaid Coverage," *Inquiry*, 1998, *35* (4), 369–379.

Yanagi, Takahide, "The Effect of Measurement Error in the Sharp Regression Discontinuity Design," December 2014. Working Paper.

Yazici, Esel Y. and Robert Kaestner, "Medicaid Expansions and the Crowding Out of Private Health Insurance among Children," *Inquiry*, 2000, *37* (1), 23–32.

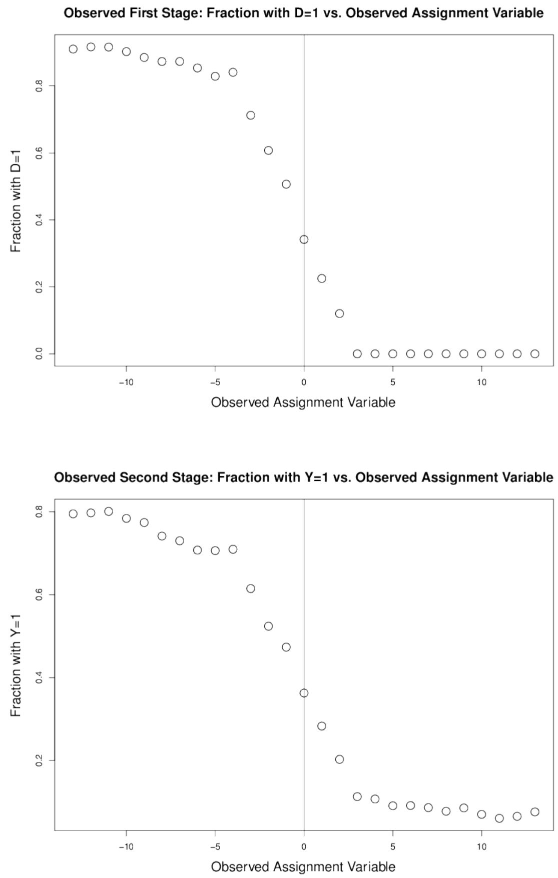
Yu, Ping, "Identification of Treatment Effects in Regression Discontinuity Designs with Measurement Error," April 2012. Working Paper.

Figure 1: Theoretical Effect of Smooth Measurement Error in the Assignment Variable



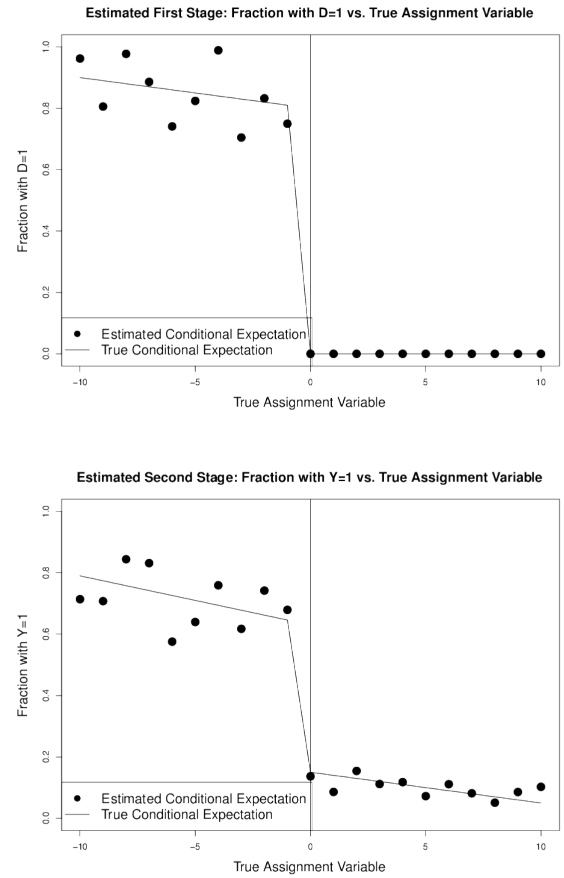
Notes: The upper panel plots the true first stage relationship $E[D|X^* = x^*]$ in a sharp RD. The lower panel plots the observed first stage relationship $E[D|X = x]$. The lower panel is generated by assuming that X^* and u are both normally distributed.

Figure 2: Observed First Stage and Outcome Relationships: Expectation of D and Y Conditional on the Noisy Assignment Variable X



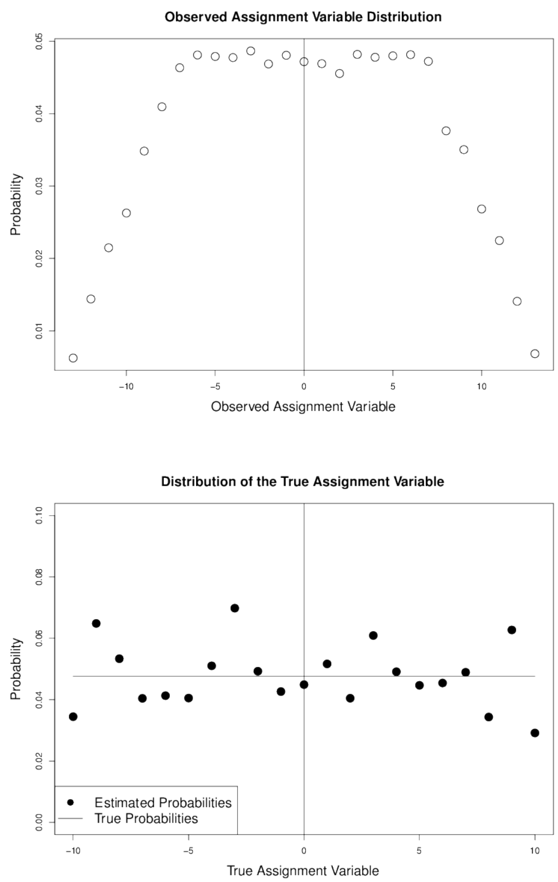
Notes: Illustrative example is based on a sample of size $N = 25,000$. X^* and u are uniformly distributed on the set of integers in $[-10, 10]$ and $[-3, 3]$, respectively. The true first stage and outcome response functions are $E[D|X^*] = (-0.01X^* + 0.8)D^*$ and $E[Y|X^*, D] = 0.15 + 0.6D - 0.01X^*$, respectively, which imply a true second stage equation of $E[Y|X^*] = 0.15 + 0.48D^* - 0.01X^* - 0.006D^*X^*$. Plotted are $E[D|X]$ and $E[Y|X]$ respectively where $X = X^* + u$.

Figure 3: Estimated First Stage and Outcome Relationships: Expectation of D and Y Conditional on the True Assignment Variable X^*



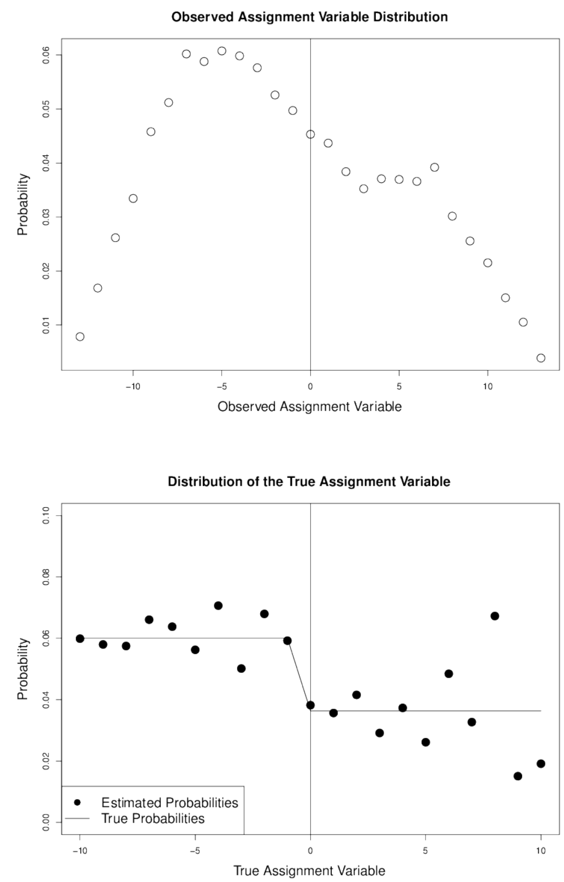
Notes: Same data generating process as Figure 2. Plotted are the estimated $E[D|X^*]$ and $E[Y|X^*]$ following procedures developed in Subsection 3.4 against the true conditional expectations.

Figure 4: Assignment Variable Distribution with Uniform X^* Distribution: Observed vs. Estimated



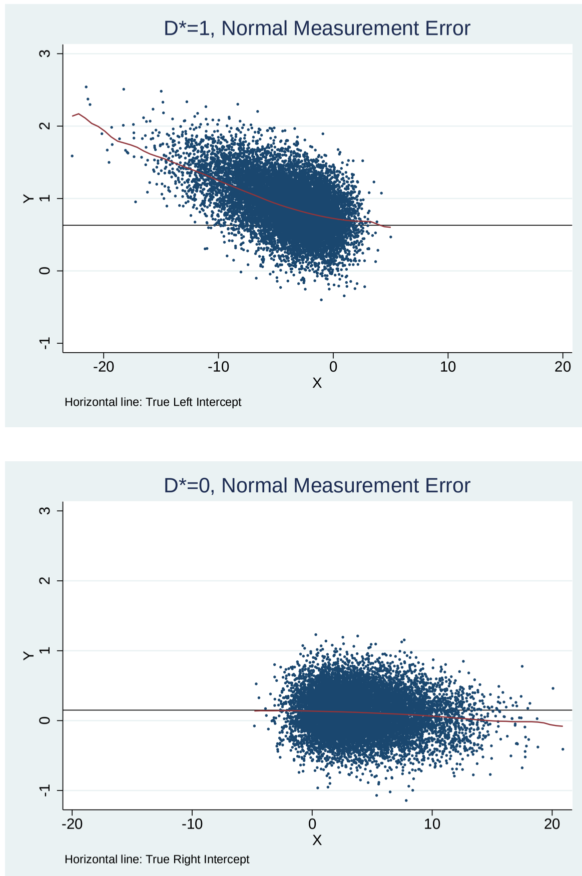
Notes: Illustrative example is based on a sample of size $N = 25,000$. X^* and u are uniformly distributed on the set of integers in $[-10,10]$ and $[-3,3]$, respectively. Plotted are the distributions of X and X^*, with the latter against the true uniform distribution specified.

Figure 5: Assignment Variable Distribution when True X^* Distribution is Not Smooth: Observed vs. Estimated



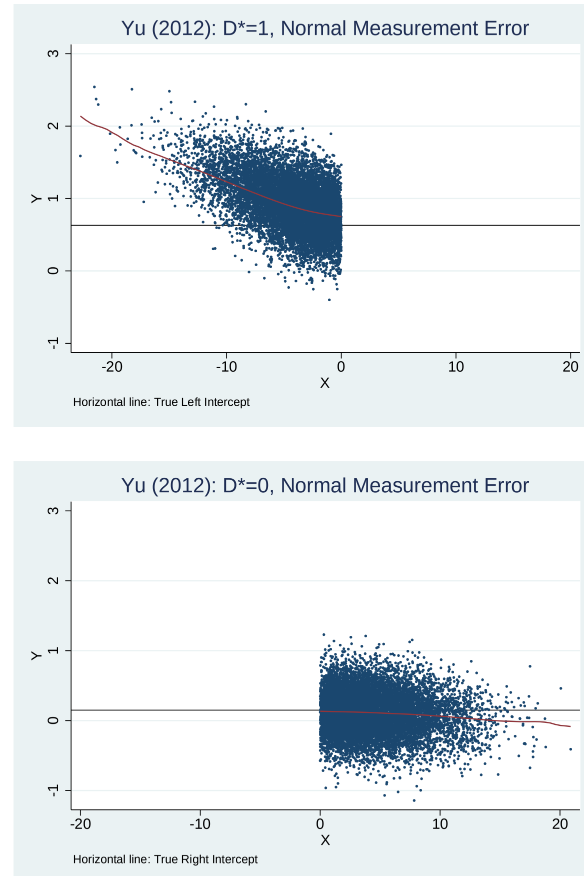
Notes: Illustrative example is based on a sample of size $N = 25,000$. X^* is supported on the set of integers in $[-10,10]$ with $\Pr(X^* = i) = 0.06$ for $i < 0$ and $\Pr(X^* = i) = 0.036$ for $i \geqslant 0$. Plotted are the distributions of X and X^*, with the latter against the true distribution specified above.

Figure 6: Identification-at-infinity Illustration



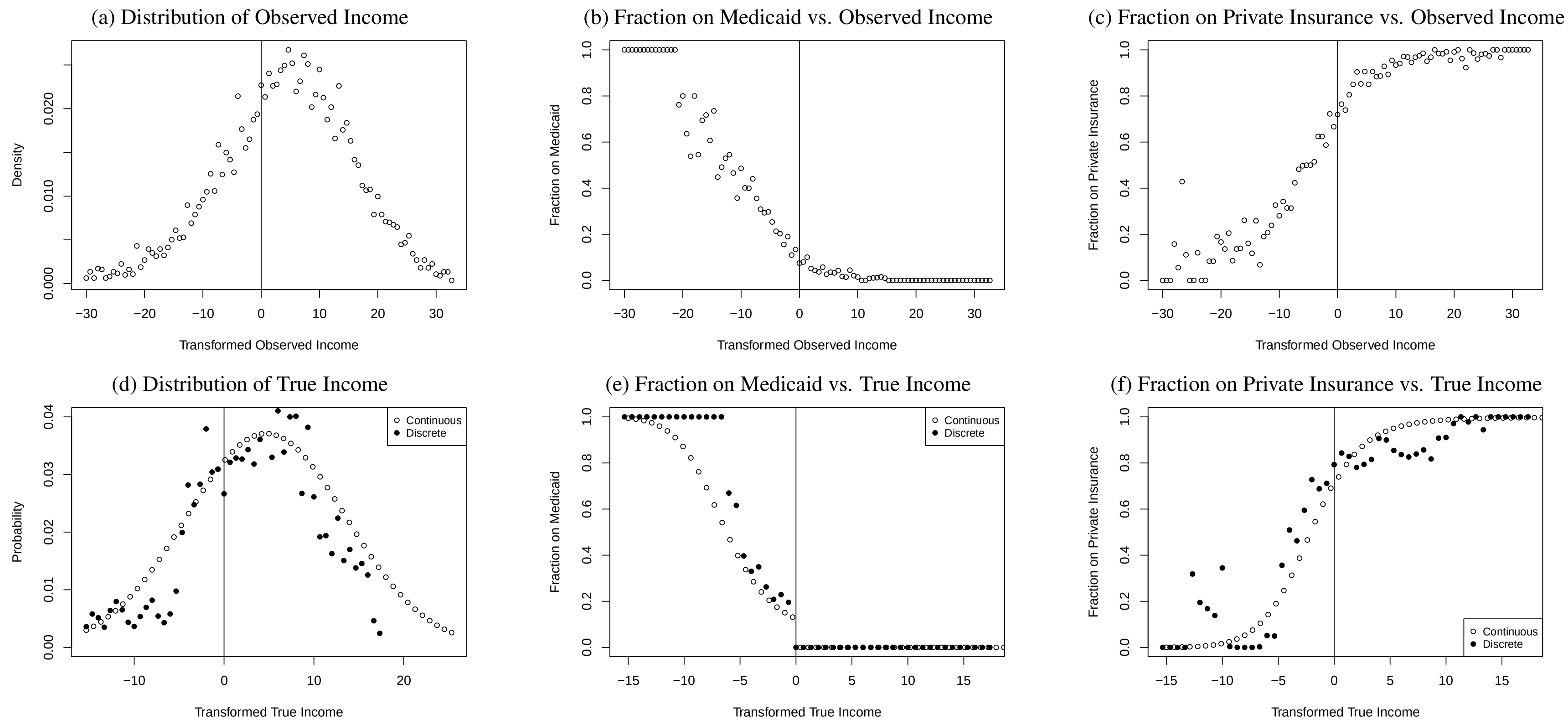
Notes: Scatter plot with 25,000 simulated observations (dots) and local constant smoothing (curve). Local smoother is implemented with the `lpoly` command in Stata with default options. The data generating process is $Y = 0.15 + 0.48D^* - 0.01X^* - 0.006D^*X^* + \varepsilon$, where $X^* \sim N(0,25)$, $\varepsilon \sim N(0,0.09)$ and $u \sim N(0,2)$.

Figure 7: Illustration of Identification in Yu (2012)



Notes: Same data and local smoothing method as Figure 6. Observations with $X \geqslant 0$ and $D = 1$ and with $X < 0$ and $D = 0$ are dropped per Yu (2012).

Figure 8: Observed and Estimated Income Distributions and First Stage and Outcome Relationships



Notes: Income is derived from a Box-Cox transformation of the actual family income with parameter 0.33, and is normalized against the transformed Medicaid Eligibility Threshold. Top and Bottom 1% of the normalized income is trimmed. Zero is the cutoff and a child is Medicaid eligible if the normalized income is less than zero. See Table A.1 for a mapping between the transformed and actual income values for various Medicaid cutoffs. Figures 8a through 8c plot the distribution of and first stage and outcome relationships with the observed income variable. Figures 8d through 8f plot the estimated distribution of and first stage and outcome relationships with the true income variable, where solid and hollow points represent estimates from the discrete and continuous modeling approaches, respectively.