Regression Discontinuity Designs with a Mismeasured Assignment Variable

Rodrigo Carril* Andres Gonzalez-Lira†

February 3, 2022

PRELIMINARY

This paper studies regression discontinuity (RD) designs in settings where the assignment variable is mismeasured. In addition to the standard sources of classical measurement error, we allow the assignment variable to be affected by the treatment. We first establish how to recover the RD parameters of interest in this case, given the distributions of measurement error and treatment effects. We then show that these objects can be nonparametrically recovered from the densities of the mismeasured running variable conditional on treatment status. In the absence of mismeasurement, the conditional densities of the running variable for treatment and control units should each be sharply discontinuous at the threshold. The difference between these sharp benchmarks and the observed densities reveals the extent of mismeasurement, and can be used to adjust the RD estimates. Imposing further assumptions, identification is possible even in the presence of "manipulation" of the running variable. We develop a method to estimate treatment effects in this context based on our results. We discuss examples of settings that fit within our framework, and illustrate our method with a particular empirical application on the effect of a policy to increase competition for public procurement contracts.

*Universitat Pompeu Fabra and Barcelona School of Economics. Email: rodrigo.carril@upf.edu.
†Yale University. Email: andres.gonzalez-lira@yale.edu.

1 Introduction

Regression Discontinuity Designs (RDDs) are pervasive in applied research evaluating the effects of a wide range of policies and programs. Invoking mild continuity assumptions on conditional mean functions, this method leverages the discontinuous nature of treatment assignment probabilities at an arbitrary cutoff of a so-called assignment variable. This allows identification of a local treatment effect by comparing observations that are just above and just below the discontinuity. Identification relies heavily on the econometrician's ability to observe accurately the assignment variable. In many cases, however, the assignment variable may be observed with error, which invalidates the canonical design.

A popular application involving RDDs with a mismeasured assignment variable involves the estimation of the causal effect of public health insurance where eligibility is a discontinuous function of individuals' income (see, e.g., Hullegie and Klein, 2010; Koch, 2013; De La Mata, 2012; and Pei and Shen, 2017). These papers explicitly address the issue of measurement error, since they do not observe the income variable used by the authority to determine eligibility, but rather a self-reported value coming from survey data. Another recent application involves the effect of publicity rules governing procurement contracts, which change sharply at arbitrary cutoffs of expected award amounts (Carril, Gonzalez-Lira, and Walker, 2022). In this case, the authors do not observe the agencies' ex-ante estimates, but rather only the ex-post awards.

The literature has focused on a case where the observed assignment variable corresponds to the true latent assignment variable plus some mean-zero classical error. While reasonable in many applications, we argue that there often might be an *additional* source of measurement error which has two important characteristics: (i) it may not have a mean equal to zero, and (ii) it may affect units differentially according to their treatment status. In particular, we consider a case where receiving the treatment may have a causal effect on the value of the observed assignment variable. These treatment effects may not be zero on average, and by definition they will only affect observations in the treatment group. In the case of the first applications, this occurs if receiving health insurance has a causal effect on the income that individuals self-report in the survey. In the second application, it manifests if publicizing contract opportunities has an effect on the value of the awarded price.

In this paper, we propose a framework that extends the canonical setting of the RDD to incorporate these two types of measurement error in the assignment variable. We start by describing how these errors affect the observed distributions of the running variable for the treatment and control group. We then propose a method to recover the densities of the latent running variable from the observed data. As a result, we nonparametrically recover the distributions of the measurement errors, which we show can be used to recover the standard RDD estimands of interest, such as the reduced form effect or the local average treatment effect. We also use recent developments in partial identification in the case of manipulation of the running variable (Gerard et al., 2020) to show that our method is robust to the presence of strategic bunching.

Intuitively, our method uses the idea that measurement error will reduce the sharpness of the discontinuity in the densities of treated and control units, and that these effects will manifest only within a window around the cutoff. Using methods developed in the bunching literature to interpolate and extrapolate densities under standard smoothness assumptions, we are then able to recover the underlying latent densities that should be sharply discontinuous. Finally, the distribution of the measurement errors are inferred from the difference between the sharp counterfactuals and the observed distributions. Crucially, our ability to separately observe the densities of the treatment and control group allow us to identify potentially distinct types of errors affecting each group.

We apply our method to the procurement setting studied by Carril, Gonzalez-Lira, and Walker (2022). Contracts expected to exceed \$25,000 are subject to a mandate of openly publicizing the solicitation. Despite our inability to observe these ex-ante estimates, we recover their distribution from the observe density of ex-post realized awards. We estimate that publicizing contracts reduces observed ex-post award prices by 6%. Accounting for this error in the assignment variable generates a modest increase in the RDD estimates, relative to a naive procedure that ignores the presence of measurement error. The results are not affected much by considering the possibility of strategic bunching responses.

We make two main contributions to the literature on RDDs with measurement error (Hullegie and Klein, 2010; Koch, 2013; De La Mata, 2012; and Pei and Shen, 2017). First, we consider identification with a more general structure of error that may differ between the treatment and control groups. As we argued above, we think this may be relevant in a wide range of applications. Second, we address the empirical challenge of measurement error by nonparametrically identifying the distribution of errors based only on the information in the densities of treated and control observations. In contrast, most existing papers either assume a functional form for the distribution of the error, or rely on external information to recover its shape. Our empirical framework does this by building upon existing papers that use density analysis to achieve nonparametric identification of behavioral parameters (Saez, 2010; Kleven and Waseem, 2013; Kleven, 2016; Chernozhukov et al., 2013). Finally, our method extends recent work by (Jales and Yu, 2017) emphasizing that density distributions around policy thresholds can be used to identify the effects of policy.

The rest of the paper proceeds as follows. Section 2 presents our model, the challenges posed to the standard RDD framework, and how we solve this based on our density analysis. Section 3 describes our estimation method and Section 4 applies it to a particular empirical setting. Section 5 concludes.

2 Model

2.1 Baseline Statistical Model

We start from a standard regression discontinuity design model. There is a series of i.i.d. observations or units $i \in \{1, ..., N\}$, for which we observe an outcome of interest Y_i, and whether the unit received a treatment ($D_i \in \{0,1\}$). We depart from the standard model by introducing both a *latent* assignment variable, and an *observed* assignment variable. The latent assignment variable $\tilde{X}_i$ influences the probability that observation i receives a treatment ($D_i = 1$). In particular, $\Pr\left(D_i = 1|\tilde{X}_i\right)$ is discontinuous (only) at a cutoff value c. Without loss of generality, we make the normalization that $c = 0$,[1] and that the probability of treatment experiences a positive jump at the threshold. This implies that compliers are not treated below $c = 0$, and do get the treatment above $c = 0$.

We assume a piece-wise linear specification for the conditional mean of the outcome given the latent assignment variable, so that our model is described by the following two key equations:

$$E\left[Y_i|\,\tilde{X}_i\right] = \begin{cases} \alpha_0 + \beta_0 \cdot \tilde{X}_i & \text{if } \tilde{X}_i \leq 0 \\ \alpha_1 + \beta_1 \cdot \tilde{X}_i & \text{if } \tilde{X}_i > 0 \end{cases} \tag{1}$$

$$E\left[D_i|\,\tilde{X}_i\right] = \begin{cases} \tilde{\pi}(\tilde{X}_i) & \text{if } \tilde{X}_i \leq 0 \\ \tilde{\pi}(\tilde{X}_i) + \delta & \text{if } \tilde{X}_i > 0 \end{cases} \tag{2}$$

for some function $\tilde{\pi}(\cdot)$. As usual, the main parameter of interest is the local average treatment effect (LATE), given by:

$$\tau_{LATE} \equiv \frac{\lim_{c\to 0^+} E\left[Y_i|\tilde{X}_i = c\right] - \lim_{c\to 0^-} E\left[Y_i|\tilde{X}_i = c\right]}{\lim_{c\to 0^+} E\left[D_i|\tilde{X}_i = c\right] - \lim_{c\to 0^-} E\left[D_i|\tilde{X}_i = c\right]} = \frac{\alpha_1 - \alpha_0}{\delta} \tag{3}$$

For ease of exposition, however, in much of what follows we focus instead on the reduced form parameter:

$$\tau_{RF} = \alpha_1 - \alpha_0 \tag{4}$$

without scaling by the first stage, and leave for the Appendix the results involving τ_{LATE}.

The problem we face is that the econometrician does not observe $\tilde{X}_i$, but only X_i, which we refer to as the observed assignment variable. The observed assignment variable depends on both the latent assignment variable and on actual treatment status. Let $X_i^d(\tilde{X}_i)$ be the *potential* observed assignment variable that would be observed for i, given $\tilde{X}_i$ and a treatment status $D_i = d$, with $d \in \{0,1\}$. We can then write the observed assignment variable as a function of potential observed assignment variables and treatment status:

$$\begin{aligned} X_i &= D_i \cdot X_i^1(\tilde{X}_i) + (1 - D_i) \cdot X_i^0(\tilde{X}_i) \\ &= X_i^0(\tilde{X}_i) + D_i \cdot \left[X_i^1(\tilde{X}_i) - X_i^0(\tilde{X}_i)\right] \end{aligned} \tag{5}$$

Of course, in this setting we cannot construct sample analogs of $E[Y_i|\tilde{X}_i]$ or $E[D_i|\tilde{X}_i]$, but rather only of $E[Y_i|X_i]$ and $E[D_i|X_i]$. A "naive RD" estimator that assumes that $X_i = \tilde{X}_i$ will not, in general,

[1] Alternatively, we can think of $\tilde{X}_i$ as the latent assignment variable re-centered around the cutoff c. That is, $\tilde{X}_i = \tilde{X}_i^* - c$, where $\tilde{X}_i^*$ is the "true" latent assignment variable.

be consistent for τ_{RF} or τ_{LATE}. Our goal is to show that under certain conditions we can still generate a consistent estimator for our parameters of interest using only information from the observed data (Y_i, D_i, X_i).

2.2 Parameters of Interest in the Baseline Model

We make the following assumptions:

A1 Values of the latent assignment variable $\tilde{X}_i$ are i.i.d. draws from a distribution with a smooth density $f_{\tilde{X}}(\cdot)$.

A2 The probability of treatment given a value of the latent assign variable is continuous everywhere, except at the cutoff. Given (2), this implies assuming that the latent function $\tilde{\pi}(\cdot)$ is continuous.

A3 For the control group, (i.e. $D_i = 0$), the observed assignment variable is equal to the latent assignment variable plus a mean-zero random error. That is, $X_i^0(\tilde{X}_i) = \tilde{X}_i + \xi_i$, with $\xi_i \sim F_{\xi}(\cdot)$ for some symmetric F_{ξ}, $E[\xi_i] = 0$, and $\xi_i \perp \tilde{X}_i$.

A4 For the treatment group, (i.e. $D_i = 1$), the observed assignment variable is equal to the latent assignment variable plus a random error, which may not have mean zero. The measurement error in the treatment group is orthogonal to that in the control group. That is, $X_i^1(\tilde{X}_i) = \tilde{X}_i + \gamma_i$, with $\gamma_i \sim F_{\gamma}(\cdot)$ for some symmetric F_{γ}, $\gamma_i \perp \tilde{X}_i$, and $\gamma_i \perp \xi_i$.

The first assumption is standard in the bunching literature, requiring the density of the latent assignment variable to be smooth. We need this since we will rely on interpolations and extrapolations of observed density functions to estimate latent densities near the cutoff. The second assumption is also technical and standard for RD models. We require that the latent function $\tilde{\pi}(\tilde{X}_i)$ is continuous everywhere, so that the jump in the treatment probability occurs only at the cutoff, and that the size of this jump can be attributed to the treatment. Importantly, we impose no additional parametric restrictions on $f_{\tilde{X}}(\cdot)$ and $\tilde{\pi}(\cdot)$.

Assumptions **A3** and **A4** impose linear measurement errors of treatment and control groups, and that both of these errors are orthogonal to each other and to the latent assignment variable. The linearity assumption may not be too restrictive, as one could possibly redefine the latent assignment variable. For example, if proportional measurement errors were more plausible in a particular application, the latent assignment variable $\tilde{X}_i$ could be redefined to be measured in natural logarithms. On the other hand, the orthogonality between the measurement errors in the treatment and control groups may be more or less plausible depending on the application.

The following proposition makes precise what are the objects that need to be observed or estimated in order to recover the structural parameters of interest.

Proposition 1. *The conditional expectation of outcomes given the observed assignment variable, $E[Y_i|X_i]$, can be expressed as an explicit linear function of the structural parameters $(\alpha_0, \alpha_1, \beta_0, \beta_1)$. In particular:*

$$E[Y_i|X_i] = \alpha_0 \cdot \phi_1(X_i) + \beta_0 \cdot \phi_2(X_i) + \alpha_1 \cdot \phi_3(X_i) + \beta_1 \cdot \phi_4(X_i) \tag{6}$$

where $\phi_k(\cdot)$, $k \in \{1, 2, 3, 4\}$ are explicit functions of the observed assignment variable (X_i), expected treatment probabilities at given values of the observed assignment variable $(E[D_i|X_i])$, and moments of the distributions of measurement errors evaluated at given values of the observed assignment variable $(F_\gamma(X_i), F_\xi(X_i))$. In particular:

$$\phi_1(X_i) \equiv [1 - F_\xi(X_i)] \cdot [1 - E[D_i|X_i]] + [1 - F_\gamma(X_i)] \cdot E[D_i|X_i]$$
$$\phi_2(X_i) \equiv F_\xi(X_i) \cdot [1 - E[D_i|X_i]] + F_\gamma(X_i) \cdot E[D_i|X_i]$$
$$\phi_3(X_i) \equiv \phi_1(X_i) \cdot X_i - [1 - F_\gamma(X_i)] \cdot E[D_i|X_i] \cdot E[\gamma|\gamma \geq X_i] - [1 - F_\xi(X_i)] \cdot [1 - E[D_i|X_i]] \cdot E[\xi|\xi \geq X_i]$$
$$\phi_4(X_i) \equiv \phi_2(X_i) \cdot X_i - F_\gamma(X_i) \cdot E[D_i|X_i] \cdot E[\gamma|\gamma < X_i] - F_\xi(X_i) \cdot [1 - E[D_i|X_i]] \cdot E[\xi|\xi < X_i]$$

Proof. See Appendix A. □

The main implication of Proposition 1 is that, since $E[D_i|X_i]$ is observable, knowledge of the two distributions of measurement errors is sufficient to construct $\phi_k(X_i)$, for $k \in \{1, 2, 3, 4\}$. It follows that the structural parameters of interest—in particular, α_0 and α_1—can be easily recovered by estimating (6) via OLS.[2] In Section 2.4, we propose a strategy to estimate F_ξ and F_γ from the densities of the observed assignment variable conditional on treatment status.

2.3 Applications

There are many settings in which a policy is introduced discontinuously based on an assignment variable that is observed with error. Consider, for example, the case of measuring the causal effect of health insurance coverage, where insurance eligibility is based on an income threshold, and said income is measured with error by the researcher. This application has received substantial attention in the literature, including by Hullegie and Klein (2010), Koch (2013), De La Mata (2012), and Pei and Shen (2017). All these papers use survey data to measure income. In this case, $\tilde{X}_i$ is the income observed by the authority that determines insurance eligibility—which the researcher cannot observe—, and X_i is the income reported in the survey. The treatment corresponds to receiving health insurance $(D_i = 1)$, and being treated is discontinuously more likely for individuals whose actual income $\tilde{X}_i$ falls below a certain cutoff c.

The measurement error in the assignment variable comes from two sources in this case. First, income observed in survey data is a proxy of real income. This introduces a type of measurement error that affects all units, and which existing applications treated as a mean-zero classical error. Second, the fact that health insurance may itself have a causal effect on income, induces another form of measurement error *only on treated units*, and which is not mean-zero. In our model, the

[2] In Appendix B, we extend the argument to obtain the analog of a first stage specification that allows us to estimate δ.

measurement error induced by the survey is captured by ξ, while γ captures the combination of the survey error and the error induced by the treatment. Importantly, all existing applications consider only the first type of error, but not the second one. Because health insurance mechanically reduces out-of-pocket medical expenditures, ignoring the effect of the treatment on the observed assignment variable may lead to bias.

Our second example is the setting studied by Carril, Gonzalez-Lira, and Walker (2022). In US federal procurement, contract opportunities expected to exceed $c = 25,000$ dollars in award value should be openly publicized on a government's website ($D_i = 1$). The authors do not observe the contracting officer's ex-ante estimate of the award $(\tilde{X}_i)$, but rather the ex-post realized award value (X_i). In practice, compliance with the mandate is imperfect, which makes this a fuzzy RD setting.

Again, the observed assignment variable X_i is subject to two sources of measurement error. First, officers may have inaccurate expectations about award values, which introduces a mean-zero measurement error in *all* observations. Additionally, the fact that a contract is publicized may have a causal effect on the observed ex-post award. Indeed, there is a strong theoretical case for the treatment to have a negative effect on average award prices: publicizing the solicitation increases the number of potential bidders who learn about the opportunity, thereby increasing the expected number of bids and—due to more intense competition—decreasing award values. This "error" is an object of interest in and of itself, as it reflects the causal effect of publicizing procurement contracts on award prices. On the other hand, failing to account for the fact that treated units are subject to this non-mean-zero measurement error will invalidate an RD analysis of the effect of the treatment on other (non-price) outcomes.

2.4 Recovering Distribution of Errors from Density Analysis

We now argue that F_γ and F_ξ can be identified from the densities of the treated ($D_i = 1$) and control ($D_i = 0$). We start with an intuitive discussion, which we later formalize.

2.4.1 Intuition

First, consider a baseline scenario in which there is no measurement error in neither the treatment nor the control group. We can think of this as a particular case of our model, where $V(\xi_i) \equiv \sigma^2_\xi = 0$, $E(\gamma_i) \equiv \mu_\gamma = 0$, and $V(\gamma_i) \equiv \sigma^2_\gamma = 0$, implying that $\gamma_i = 0$ and that $\xi_i = 0$ for all i. It follows from **A3** and **A4** that both X^0_i and X^1_i will be exactly equal to the latent assignment variable $\tilde{X}_i$. This will lead to frequency distributions of treatment and control groups, such as the ones depicted in Figure 1 (a). The graph also shows the total frequency distribution, which is the vertical sum of the two. In this case, the observed distributions of X_i match exactly the latent distributions of $\tilde{X}_i$. The jump in the probability of treatment at the threshold generates a positive jump in the frequency of treated units, which is compensated one-to-one with a drop in the frequency of control units. This exact compensation implies that the frequency of all units is perfectly smooth, consistent with **A1**.

Now consider the effect of measurement error in the control group, i.e., $\sigma_\xi > 0$, as depicted in Figure 1 (b). Because the treatment group is not affected by ξ_i, the observed frequency distribution for treated units remains unchanged and is equal to the latent distribution. For the control group, however, the sharp discontinuity at the threshold gets smoothed out by the measurement error. Because now the sharp increase in the frequency of treated units is not compensated exactly by a drop in control units at the cutoff, the observed total frequency distribution is now discontinuous at $X_i = c$.

Figure 1 (c) and Figure 1 (d) consider the additional effect of measurement error in the treatment group. Figure 1 (c) shows a case where γ_i is also mean-zero, i.e., $\mu_\gamma = 0$ and $\sigma_\gamma > 0$. This mean-zero γ_i has an analogous effect on the observed frequency of treated units as the one ξ_i had on the control distribution, namely it smooths out the sharp discontinuity at the cutoff. Because now the distributions of both treated and control units are continuous, the observed total frequency distribution is also continuous. Depending on the relative size of σ_ξ and σ_γ, the total frequency distribution may be more or less smooth, as shown more clearly in Appendix Figure A1.

Finally, Figure 1 (d) shows the general case where $\sigma_\xi^2 > 0$, $\mu_\gamma \neq 0$, and $\sigma_\gamma^2 > 0$. That is, a case where the control group is subject to mean-zero measurement error and the treatment group is subject to non-mean-zero measurement error. Relative to panel (c), the non-zero mean of γ_i introduces a horizontal shift in the observed frequency distribution of treated units. In the Figure, we have depicted a case where $\mu_\gamma < 0$, so that the frequency of treated units is shifted to the left. This generates an excess mass in the observed total frequency distribution to the left of the cutoff, and a corresponding missing mass to the right.

The purpose of this discussion is to highlight that the unobserved measurement errors ξ_i and γ_i affect the observed densities of treated and control units in predictable ways. Our main identification argument rests on the fact that we can observe these distributions separately for treated and control units, and can therefore distinguish between different types of errors. Indeed, our estimation method proposes to "reverse-engineer" these effects, recovering the latent distributions from the observed ones.

2.4.2 Formalization: Convolution of Densities

We now formalize the previous argument by characterizing the density of the observed assignment variable X_i conditional on treatment status $D_i = d \in \{0, 1\}$, showing how these relate to the latent distributions of $\tilde{X}_i$, ξ_i, and γ_i.

Consider first the density of the observed assignment variable for control units, $f_{X|D=0}(x)$. Because the observed assignment variable is given by the sum of two independent random variables—$\tilde{X}_i$ and ξ_i, see **A3**—, its density is given by the convolution of the densities $f_{\tilde{X}|D=0}$ and f_ξ. That is, for $X_i = x$:

$$f_{X|D=0}(x) = \int_{-\infty}^{\infty} f_{\tilde{X}|D=0}(x - \xi) f_\xi(\xi) d\xi \qquad (7)$$

Using Bayes' rule and some simple manipulations we have:

$$\begin{aligned}
f_{X|D=0}(x) &= \int_{-\infty}^{\infty} \frac{\Pr(D=0|\tilde{X}_i = x-\xi)\cdot f_{\tilde{X}}(x-\xi)\cdot f_{\xi}(\xi)}{\Pr(D_i=0)} d\xi \\
&= \int_{-\infty}^{\infty} \frac{[1-\tilde{\pi}(x-\xi) - \delta \cdot \mathbf{1}[x-\xi>0]]\cdot f_{\tilde{X}}(x-\xi)\cdot f_{\xi}(\xi)}{\Pr(D_i=0)} d\xi \\
&= \int_{-\infty}^{\infty} \frac{[1-\tilde{\pi}(x-\xi)]\cdot f_{\tilde{X}}(x-\xi)}{\Pr(D_i=0)} f_{\xi}(\xi) d\xi - \int_{-\infty}^{x} \frac{\delta \cdot f_{\tilde{X}}(x-\xi)}{\Pr(D_i=0)} f_{\xi}(\xi) d\xi
\end{aligned}$$

Given this, we can rewrite observed density of the assignment variable for control units as:

$$f_{X|D=0}(x) = \widehat{f}_{\tilde{X}|D=0}(x) - \int_{-\infty}^{x} \frac{\delta \cdot f_{\tilde{X}}(x-\xi)\cdot f_{\xi}(\xi)}{\Pr(D_i=0)} \cdot d\xi \tag{8}$$

where:

$$\widehat{f}_{\tilde{X}|D=0}(x) \equiv E_{\xi}\left[\frac{[1-\tilde{\pi}(x-\xi)]\cdot f_{\tilde{X}}(x-\xi)}{\Pr(D_i=0)}\right] \tag{9}$$

and we call $\widehat{f}_{\tilde{X}|D=0}(x)$ the latent *quasi-density* of control units, since this function is closely related to the true density of the latent assignment variable for control units, $f_{\tilde{X}|D=0}(x)$.

To understand these expressions, first consider a value of the observed assignment variable sufficiently below the threshold, so that the probability that the latent assignment variable was above the threshold is negligible, i.e., $X_i = x \ll 0$, so that $f_{\xi}(x) = 0$. In this case, the second term in Equation (8) is zero. On the other hand, by Bayes' rule, $\frac{[1-\tilde{\pi}(z)]\cdot f_{\tilde{X}}(z)}{\Pr(D_i=0)} = f_{\tilde{X}|D=0}(z)$ when $z < 0$, so that the first term in (8) is $\widehat{f}_{\tilde{X}|D=0}(x) = E_{\xi}[f_{\tilde{X}|D=0}(x-\xi)]$. This means that the latent quasi-density is the expectation of the true density of the latent assignment variable, shifted by the (mean-zero) measurement error ξ. As depicted in Appendix Figure A2, if the true density is linear, then it will be exactly equal to the quasi-density. On the contrary, if the true density is very non-smooth, then the quasi-density will correspond to a smoothed-out version of the former. However, because of our smoothness assumption in **A1**, the quasi-density will be sufficiently close to the true density, therefore implying that $f_{X|D=0}(x) = \widehat{f}_{\tilde{X}|D=0}(x) \approx f_{\tilde{X}|D=0}(x)$ for $x \ll 0$. In words, sufficiently below the threshold, the density of the observed assignment variable for the control group is approximately equal to the unobserved density of the latent assignment variable. It follows that the expected number of units with observed assignment variable $X_i = x$ equals the expected number of units with latent assignment variable $\tilde{X}_i = x$, for $x \ll 0$.

As we move closer to the threshold from below, the second term in Equation (8) becomes positive. This corresponds to the excess density of units, relative to the counterfactual density in the first term. Intuitively, this term is given by the mass of units with latent assignment variable to the right of the cutoff $c = 0$ that receive a sufficiently negative ξ so as to end up with $X_i < 0$. Now consider $X_i = x$ closely below the threshold, so that $f_{\tilde{X}}(x-\xi)$ is approximately constant, with $\delta \cdot f_{\tilde{X}}(x-\xi)/\Pr(D_i=0) = \Delta$. It follows that:

$$\widehat{f}_{\tilde{X}|D=0}(x) - f_{X|D=0}(x) = \Delta \cdot F_{\xi}(x) \tag{10}$$

for x approaching the cutoff from below.

Figure 2 shows graphically the logic of Equation (10). Consider units in the control group and closely below the cutoff. The left-hand side of (10) is the missing mass of units in the observed density, relative to the latent (quasi-)density. Such gap is explained by observations that received a positive value of ξ such that their observed assignment variable crosses the cutoff. This is equivalent to the size of the discontinuity in the latent density of control units at the threshold (Δ), times the probability of having a latent assignment variable that crossed the cutoff ($\Pr(x - \xi > 0) = F_{\xi}(x)$).

A symmetric argument can be given for X_i approaching the cutoff from above. In this case, the second term becomes the excess mass of units in the observed density, relative to the latent (quasi-)density. Once we get to a high enough value of $X_i \gg 0$, once again $f_{\xi}(X_i)$ goes to zero and this excess mass disappears, with observed and latent densities converging.

The argument for treated units is analogous. Using the convolution of $\tilde{X}_i$ and γ, and applying Bayes' rule and simple manipulations we obtain:

$$f_{X|D=1}(x) = \widehat{f}_{\tilde{X}|D=1}(x) + \int_{-\infty}^{x} \frac{\delta \cdot f_{\tilde{X}}(x-\gamma) \cdot f_{\gamma}(\gamma)}{\Pr(D_i = 1)} \cdot d\gamma \tag{11}$$

where:

$$\widehat{f}_{\tilde{X}|D=1}(x) \equiv E_{\gamma}\left[\frac{\tilde{\pi}(x-\gamma) \cdot f_{\tilde{X}}(x-\gamma)}{\Pr(D_i = 1)}\right] \tag{12}$$

As we will see, the main difference from the case of control units is that, since γ_i may have a non-zero mean, the latent quasi-density $\widehat{f}_{\tilde{X}|D=1}(x)$ is shifted horizontally from the true density of the latent assignment variable $f_{\tilde{X}|D=1}(x)$. The size of this shift corresponds to the mean of the measurement error in the treatment group, $\mu_{\gamma} \equiv E[\gamma]$. Apart from this, the argument is mostly equivalent.

Consider $X_i = x \ll 0$ so that the second term in (11) is zero. For these values we have that $\widehat{f}_{\tilde{X}|D=1}(x) = E_{\gamma}\left[f_{\tilde{X}|D=1}(x-\gamma)\right] \approx f_{\tilde{X}|D=1}(x) - \mu_{\gamma}$, where the equality follows from Bayes rule and the fact that $\frac{\tilde{\pi}(z) \cdot f_{\tilde{X}}(z)}{\Pr(D_i=1)} = f_{\tilde{X}|D=1}(z)$ when $z < 0$. The approximation is exact for a linear density of the latent assignment variable, and in general is guaranteed to be close given **A1**. This makes clear that, for treated units, the latent quasi-density is shifted horizontally by μ_{γ}, relative to the true density of the latent assignment variable. Moreover, (11) implies that sufficiently below $c = 0$, the observed density of the assignment variable will be equal to the latent density of $\tilde{X}_i$ shifted horizontally by the mean of the measurement error γ_i.

As we approach $c = 0$, the second term becomes positive and corresponds to the excess mass in the observed density of the treatment group, relative to the (shifted) latent density of $\tilde{X}_i$ for these units. Now, note that we know that the discontinuities in the latent densities of the treatment and control group should be equal to each other at the cutoff $c = 0$. Therefore, we know that if $\delta \cdot f_{\tilde{X}}(x-\xi)/\Pr(D_i = 0) = \Delta$ for $x \approx 0$, then $\delta \cdot f_{\tilde{X}}(x-\gamma)/\Pr(D_i = 1) = \Delta$ for $x \approx \mu_{\gamma}$. So considering a value close to μ_{γ}, such that we can assume that $f_{\tilde{X}}(x-\gamma)$ is approximately constant, we obtain:

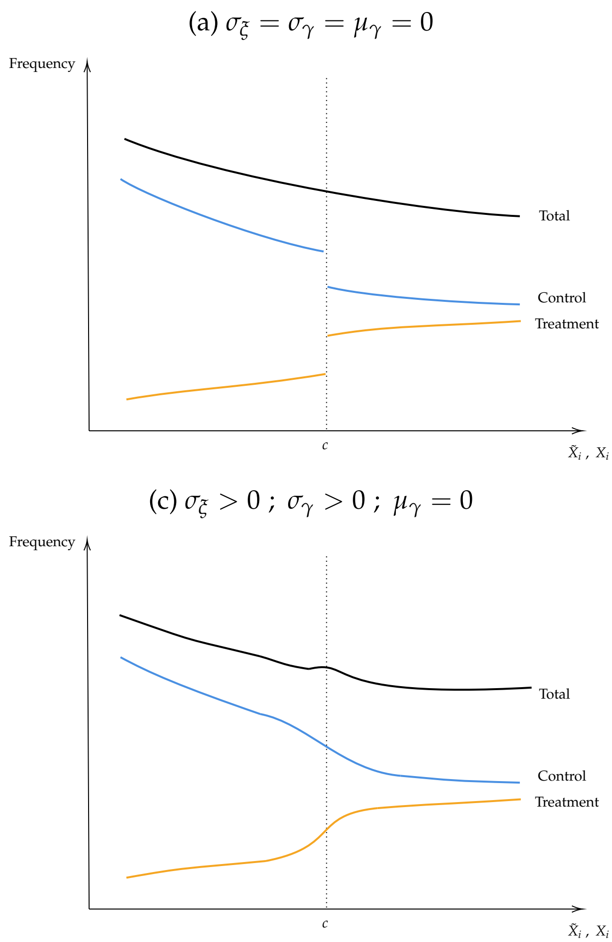
$$f_{X|D=1}(x) - \widehat{f}_{\tilde{X}|D=1}(x) = \Delta \cdot F_\gamma(x) \tag{13}$$

for x approaching μ_γ from below.

Above μ_γ, the logic extends in a straightforward way. The left-hand side of (13) becomes a measure of *missing* mass, and sufficiently above the cutoff we reach a point where observed and (shifted) latent densities converge. Figure 3 summarizes these relationships. Panel (a) provides a graphical representation of equation (11), while Panel (b) zooms in around μ_γ to show the relationship between excess/missing masses and the size of the discontinuity in $f_{\tilde{X}|D=1}$ described by (13). One take-away of this Figure is that γ_i generates analogous effects as on the densities of treated units as ξ_i generates for control units, except for the horizontal shift explained by the non-zero mean of γ_i.

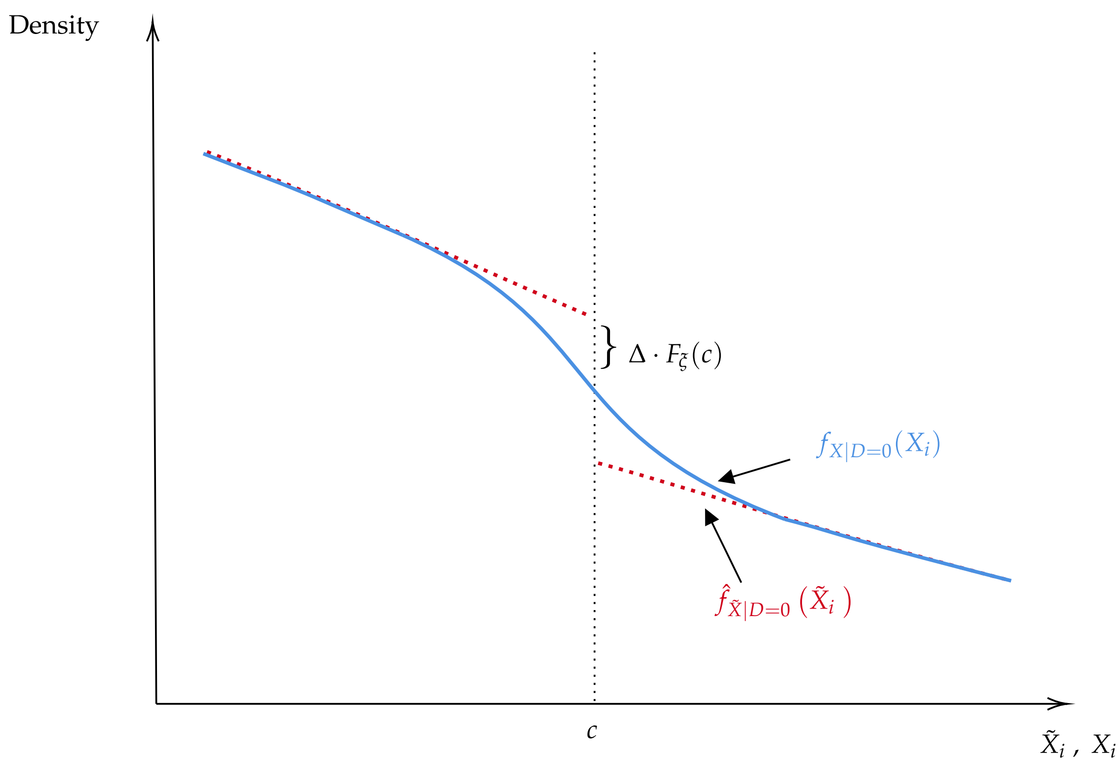
The overall take-away of this and the preceding section is that if we were either able to observe or estimate the different relevant densities of X_i and $\tilde{X}_i$, then we could recover the distributions of the measurement errors that affect each group of units, which—according to Proposition 1— is what we need for estimating the structural parameters of interest in our RD model.

Figure 1: Impact of measurement errors on observed densities



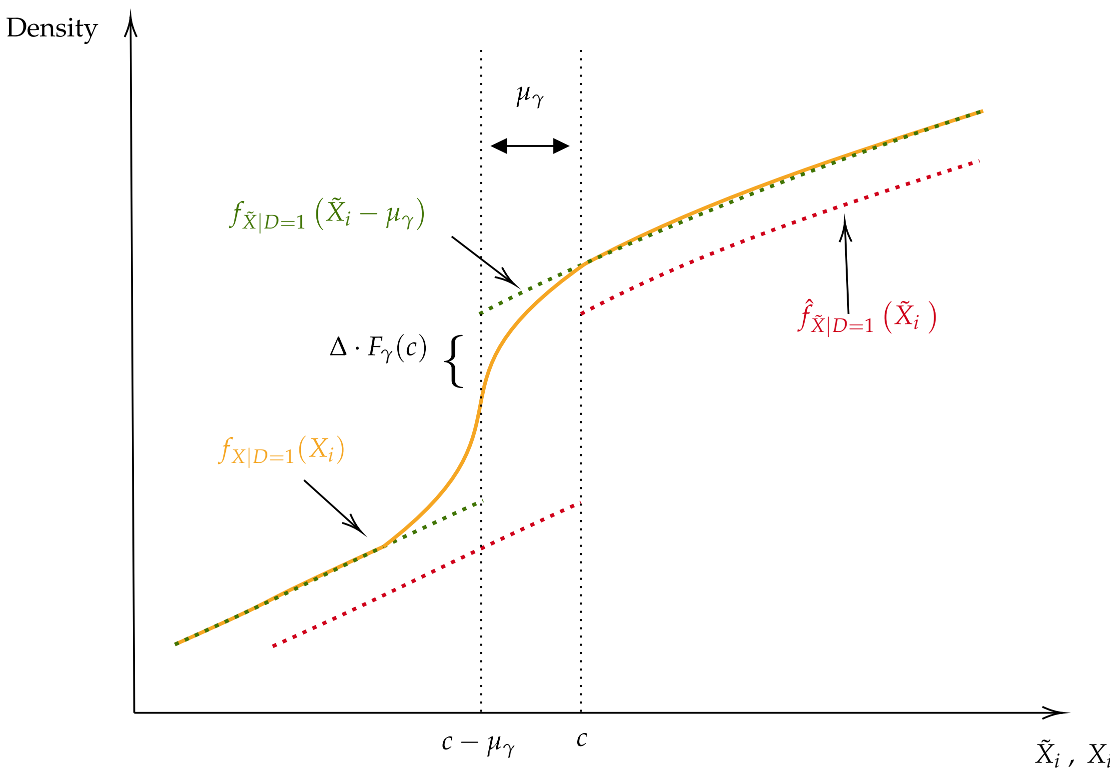
Notes: This figure shows an example of how measurement errors affect the observed distributions of treated and control observations, as well as the total distribution corresponding to the sum of the first two. Panel (a) shows a case with no measurement errors ($\sigma_\xi = \sigma_\gamma = \mu_\gamma = 0$). The total distribution is continuous by Assumption A1. The distributions of treated and controls are each discontinuous at the cutoff. The jump in the number of treated observations is equal to the drop in the number of controls. Panel (b) shows a case with mean-zero measurement error affecting the control group ($\sigma_\xi > 0$). Measurement error smooths-out the control distribution, generating a discontinuity in the total distribution. Panel (c) also introduces mean-zero measurement error affecting the treatment group ($\sigma_\gamma > 0$). Measurement error smooths-out the treatment distribution, making the total distribution continuous again (although not necessarily smooth). Panel (d) shows the general case of mean-zero measurement error in the control group and non-mean-zero measurement error in the treatment group. Relative to Panel (c), the distribution of treated observations is shifted vertically by μ_γ. The total distribution exhibits excess mass below the cutoff as a result.

Figure 2: The effect of measurement error for control units near the cutoff



Notes: This figure shows the graphical representation of Equation (10). Measurement error smooths-out the observed density near the cutoff. The difference between the observed and latent density that contains no measurement error is equal to the size of the discontinuity in the latent density times the CDF of the measurement error evaluated at the cutoff.

Figure 3: The effect of measurement error for treated units near the cutoff



Notes: This figure shows the graphical representation of Equation (13). Measurement error smooths-out the observed density near the cutoff, and shifts it horizontally by the error mean. The difference between the observed and latent density that contains no measurement error is equal to the size of the discontinuity in the latent density times the CDF of the measurement error evaluated at the cutoff.

3 Estimation Approach

Our estimation method relies on "reverse-engineering" the effects highlighted in the previous section, to obtain key latent magnitudes from observed data. Based on these ideas, and building upon bunching methods originally developed to measure behavioral responses to taxes (see Kleven (2016)), our approach nonparametrically recovers the distributions of the latent assignment variable and of measurement errors.

3.1 Empirical Distributions

Discretization and Binning. In order to implement the estimation method, we start by discretizing the space of the assignment variable to work with binned frequency distributions. Consider the division of the range of possible assignment variable values into a set of equally-sized and right-inclusive bins around the cutoff $b \in \{-R, (-R+1), ..., -1, 0, 1, ..., (R-1), R\}$. Note that bin $b = 0$ includes units right at, or slightly below, the cutoff c.

Let $\{n_b^d\}_{b=-R}^R$ be the frequency distribution of observed units conditional on treatment status $D_i = d$, for $d \in \{0, 1\}$, so that n_b^d denotes the number of observations with treatment status d and observed assignment variable $X_i \in b$. Likewise, let $\{\tilde{n}_b^d\}_{b=-R}^R$ represent the (unobserved) frequency distribution of the assignment variable $\tilde{X}_i$. We also denote the distribution of *all* units—both treated and control—by simply omitting the superscript. That is, $n_b = n_b^0 + n_b^1$, and $\tilde{n}_b = \tilde{n}_b^0 + \tilde{n}_b^1$.

We also define a *shifted* distribution of treatment units, $\{n_b^{1,s}(\mu_\gamma)\}_{b=-R}^R$, which is obtained by subtracting a value of μ_γ to every treated ($D_i = 1$) observation. That is, $n_b^{1,s}(\mu_\gamma)$ denotes the number of treated units with observed assignment variable X_i such that $(X_i + \mu_\gamma) \in b$.

Theoretical Implications. The analysis from Section 2 yields a series of implications about the relationship between observable densities, latent densities, and measurement errors. We now state these implications in terms of binned frequency distribution.

First, we have that for treated units far enough from the cutoff, the distribution of the observed assignment variable, appropriately shifted to cancel out the mean measurement error, coincides with the distribution of the latent assignment variable. That is, there exist some $(\underline{b}^1, \bar{b}^1)$ such that:

$$E[\tilde{n}_b^1] = E[n_b^{s,1}(\mu_\gamma)] \tag{14}$$

for $\mu_\gamma = E[\gamma_i]$, $b < \underline{b}^1 < 0$ and $b > \bar{b}^1 > 0$.

An analogous result follows for control units, without the horizontal shift. Far enough from the cutoff, the distributions of the observed and latent assignment variable conditional on $D_i = 0$ coincide. That is, there exist some $(\underline{b}^0, \bar{b}^0)$ such that:

$$E[\tilde{n}_b^0] = E[n_b^0] \tag{15}$$

for $b < \underline{b}^0 < 0$, and $b > \bar{b}^0 > 0$.

A corollary from these two results is that far from the cutoff we can easily compute the *unconditional* frequency of the latent assignment variable by adding the frequency of control units and the shifted frequency of treated units. That is:

$$E[\tilde{n}_b] = E[n_b^0 + n_b^{s,1}(\mu_\gamma)] \tag{16}$$

for $\mu_\gamma = E[\gamma_t]$, $b < \underline{b} = \min\{\underline{b}^0, \underline{b}^1\} < 0$ and $b > \bar{b} = \max\{\bar{b}^0, \bar{b}^1\} > 0$.

Another important result is that the frequency distributions should satisfy an integration constraint: the total number of observations in the observed and latent unconditional frequency distributions should be equal. This can be restated in terms of the difference below the two frequencies, and implies that any excess (or missing) mass below the cutoff must equal the missing (or excess) mass above the cutoff. That is:

$$\sum_{b \leq 0} (\tilde{n}_b - n_b) = \sum_{b > 0} (n_b - \tilde{n}_b) \tag{17}$$

Finally, the discrete equivalent to equations (10) and (13) corresponds to the following expressions:

$$\Delta_n \cdot F_{\gamma'}(x) = E[n_{b_x}^{1,s}(\mu_\gamma)] - E[\tilde{n}_{b_x}^1] \tag{18}$$

$$\Delta_n \cdot F_\xi(x) = E[n_{b_x}^0] - E[\tilde{n}_{b_x}^0] \tag{19}$$

for $x \in b_x$, $b_x \leq 0$, $\gamma' = \gamma - \mu_\gamma$, and where Δ_n is the change in the *number* of observations that change treatment status at the cutoff.

3.2 Estimation

The estimation procedure consists of two steps that stem directly from the theoretical results. The first step relies on the integration constraint. The second step hinges on matching the discontinuity of the treatment and control groups.

3.2.1 Step 1

We start by assuming knowledge of the mean measurement error in the treatment group, allowing us to infer the total frequency distribution of the latent assignment variable far from the cutoff. Given the assumption of smoothness (**A1**), we then use standard bunching methods to infer the shape of the distribution near the cutoff, fitting a flexible function from both sides.

Suppose that we know that $E[\gamma_t] = \mu_\gamma$, then we can construct $\{n_b^{1,s}(\mu_\gamma)\}$ by shifting the observed frequency distribution of treated observations $\{n_b^1\}$ to the right.[3] Because of result (16),

[3] This is done by simply adding μ_γ to value of the observed assignment variable X_i of every treated observation.

we can then construct the unobserved distribution $\{\tilde{n}_b\}$ sufficiently far from the cutoff by adding $\{n_b^{1,s}(\mu_\gamma)\}$ and $\{n_b^0\}$. Finally, we fit a polynomial function through our constructed distribution $\{n_b^0 + n_b^{1,s}(\mu_\gamma)\}$, ignoring the contribution of the bins close to the cutoff.

More concretely, given a polynomial degree Q and a set of excluded bins around the cutoff $\{\underline{b}, ..., 0, ..., \overline{b}\}$, we estimate the following specification:

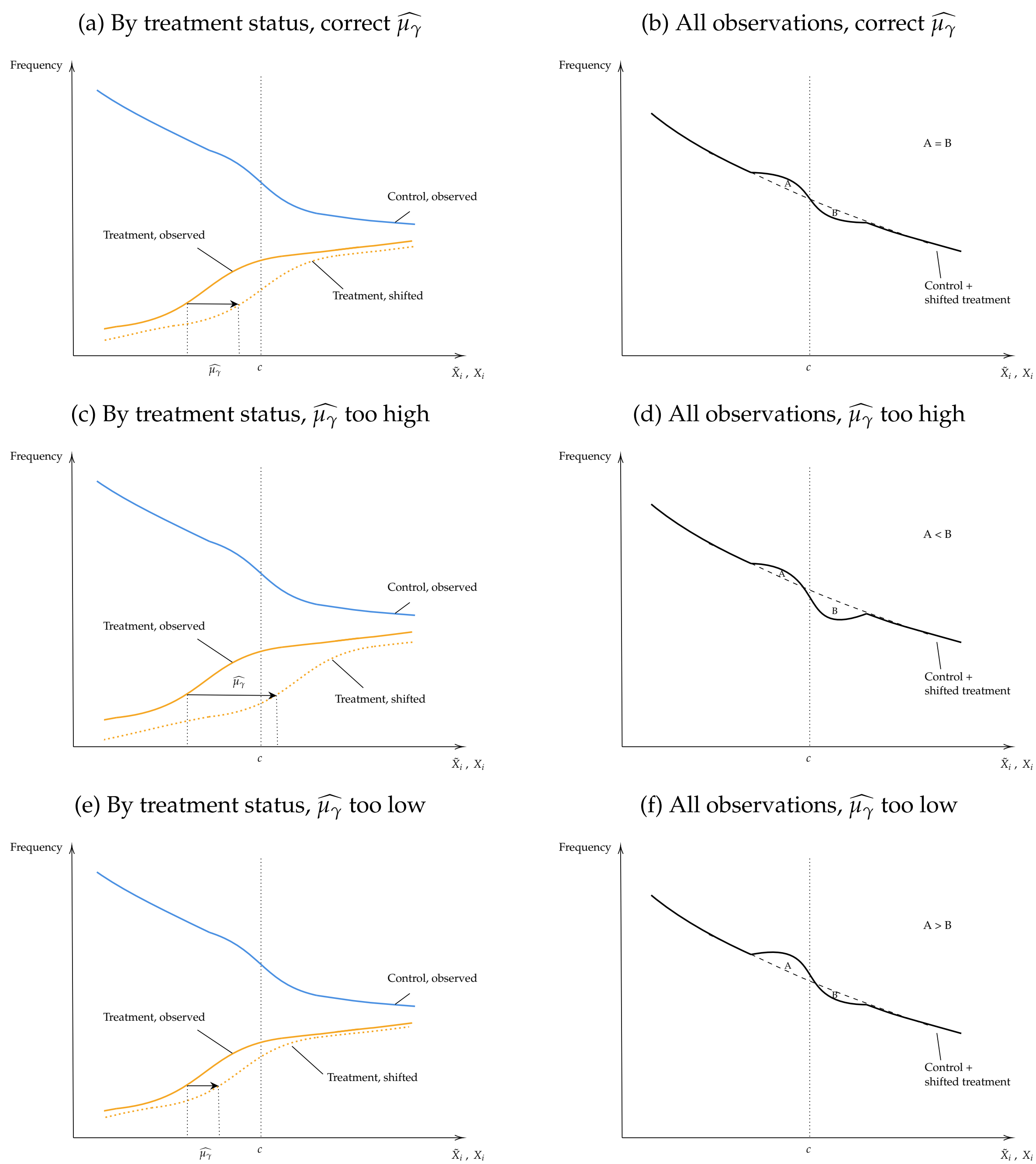
$$\left[n_b^0 + n_b^{1,s}(\widehat{\mu_\gamma})\right] = \sum_{x=0}^{Q} \alpha_x \cdot b^x + \sum_{j=\underline{b}}^{\overline{b}} \gamma_j \cdot \mathbf{1}[b=j] + \nu_b, \quad \text{for } b = \{-R, ..., R\}. \tag{20}$$

and obtain fitted values:

$$\widehat{n_b} = \sum_{x=0}^{Q} \widehat{\alpha_x} \cdot b^x, \quad \text{for } b = \{-R, ..., R\}.$$

Now, in practice, μ_γ is an unknown parameter. In order to estimate it, we rely on the integration constraint (17). As the intuition from Figure 4 shows, the integration constraints will bind only when we shift the distribution of treated units according to the right value of $\mu_\gamma = E[\gamma_i]$. Therefore, we start from an initial guess of $\widehat{\mu_\gamma}$, and iterate until the constraint (17) is satisfied.

Figure 4: Intuition of step 1



Notes: This figure provides graphical intuition of the procedure to estimate the mean measurement error in the treatment group (μ_γ). Panels (a), (c), and (e) display frequency distributions of the treatment and control groups. Panels (b), (d), and (f) show distributions that add the control group observations with the shifted distribution of treated observations. This step is based on an integration constraint condition on this latter distribution: the excess mass below the cutoff must equal the missing mass above the cutoff. The key intuition is that the integration constraint condition is only met if the distribution of publicized contracts is re-centered by the correct mean of μ_γ.

3.2.2 Step 2

The second step seeks to estimate separate latent distributions by treatment status, i.e. $\{\widehat{\tilde{n}_b^0}\}$ and $\{\widehat{\tilde{n}_b^1}\}$.

Recall that in the absence of any measurement error, the distributions of treated and control units should be continuous everywhere except at the cutoff, where we should see a discontinuous jump in treated observations mirrored by a discontinuous dip in control observations. Suppose that we knew the size of this change, which we denote as Δ_n. Knowledge of Δ_n would allow us to undo these discontinuities by shifting the right part of each distribution vertically. Indeed, the distributions $\{n_b^0 + \Delta \cdot \mathbf{1}[b > 0]\}$ and $\{n_b^1 - \Delta \cdot \mathbf{1}[b > 0]\}$ should be continuous if ξ_i and γ_i are equal to zero for all i.

In the presence of measurement error, these vertical shifts will not make the observed distributions continuous. However, just as in the discussion above, the effects of ξ_i and γ_i should only manifest itself within some window around the cutoff. Using the same previous logic, we can use a polynomial interpolation to estimate the latent distributions around the cutoff.

First, we construct distributions that are vertically shifted by Δ_n to the right of the cutoff: $\{n_b^0 + \Delta_n \cdot \mathbf{1}[b > 0]\}_{b=-R}^{R}$ and $\{n_b^{1,s}(\widehat{\mu_\gamma}) - \Delta_n \cdot \mathbf{1}[b > 0]\}_{b=-R}^{R}$. We then separately estimate the following two specifications:

$$\left(n_b^0 + \Delta_n \cdot \mathbf{1}[b > 0]\right) = \sum_{x=0}^{Q} \alpha_x^0 \cdot b^x + \sum_{j=\underline{b^0}}^{\overline{b^0}} \gamma_j^0 \cdot \mathbf{1}[b = j] + \nu_b^0, \quad \text{for } b = \{-R, ..., R\} \tag{21}$$

$$\left(n_b^{1,s}(\widehat{\mu_\gamma}) - \Delta_n \cdot \mathbf{1}[b > 0]\right) = \sum_{x=0}^{Q} \alpha_x^1 \cdot b^x + \sum_{j=\underline{b^1}}^{\overline{b^1}} \gamma_j^1 \cdot \mathbf{1}[b = j] + \nu_b^1, \quad \text{for } b = \{-R, ..., R\} \tag{22}$$

and compute fitted values ignoring the contribution of the bins within the excluded window:

$$\widehat{n_b^{*0}} = \sum_{x=0}^{Q} \widehat{\alpha_x^0} \cdot b^x, \quad \text{for } b = \{-R, ..., R\}$$

$$\widehat{n_b^{*1}} = \sum_{x=0}^{Q} \widehat{\alpha_x^1} \cdot b^x, \quad \text{for } b = \{-R, ..., R\}$$

Finally, our estimates of the latent distributions do incorporate the discontinuous effect of the treatment. We estimate these by re-adding the shift that we originally removed:

$$\widehat{n_b^0} = \widehat{n_b^{*0}} - \Delta_n \cdot \mathbf{1}[b > 0] \quad \text{for } b = \{-R, ..., R\}$$

$$\widehat{n_b^1} = \widehat{n_b^{*1}} + \Delta_n \cdot \mathbf{1}[b > 0] \quad \text{for } b = \{-R, ..., R\}$$

Again, the difficulty in practice is that we do not know the value of Δ_n. Since, in practice, this is not directly observed, our method iterates over guesses of $\widehat{\Delta_n}$. The convergence criterion in this case is based on the fit of the interpolations outside the excluded window. Indeed, if the vertical shift we guess is too low or too high, the polynomial interpolation will fit poorly just outside of the excluded area. Figure 5 shows this intuition graphically.

Given a guess of $\widehat{\Delta_n}$, we compute the residuals for each of the two regressions (21) and (22). We then search over $\widehat{\Delta_n}$ to minimize:

$$W\left(\widehat{\Delta_n}\right) = 0.5 \cdot \sum_{b \neq Z^0} \widehat{\nu_b^0}\left(\widehat{\Delta_n}\right)^2 + 0.5 \cdot \sum_{b \neq Z^1} \widehat{\nu_b^1}\left(\widehat{\Delta_n}\right)^2$$

where $Z^0 = \{\underline{b^0}, ..., \overline{b^0}\}$ and $Z^1 = \{\underline{b^1}, ..., \overline{b^1}\}$ correspond to the excluded regions around the cutoff.

3.2.3 Step 3

The final step relies on (18) and (19), and our previous estimates of the latent distributions to compute the CDF of the measurement errors. In particular,

$$\widehat{F_{\gamma'}}(x_1) = \frac{n^{1,s}_{b_{x1}}(\widehat{\mu_\gamma}) - \widehat{\tilde{n}^1_{b_{x1}}}}{\widehat{\Delta_n}}$$

$$\widehat{F_\xi}(x_0) = \frac{n^0_{b_{x0}} - \widehat{\tilde{n}^0_{b_{x0}}}}{\widehat{\Delta_n}}$$

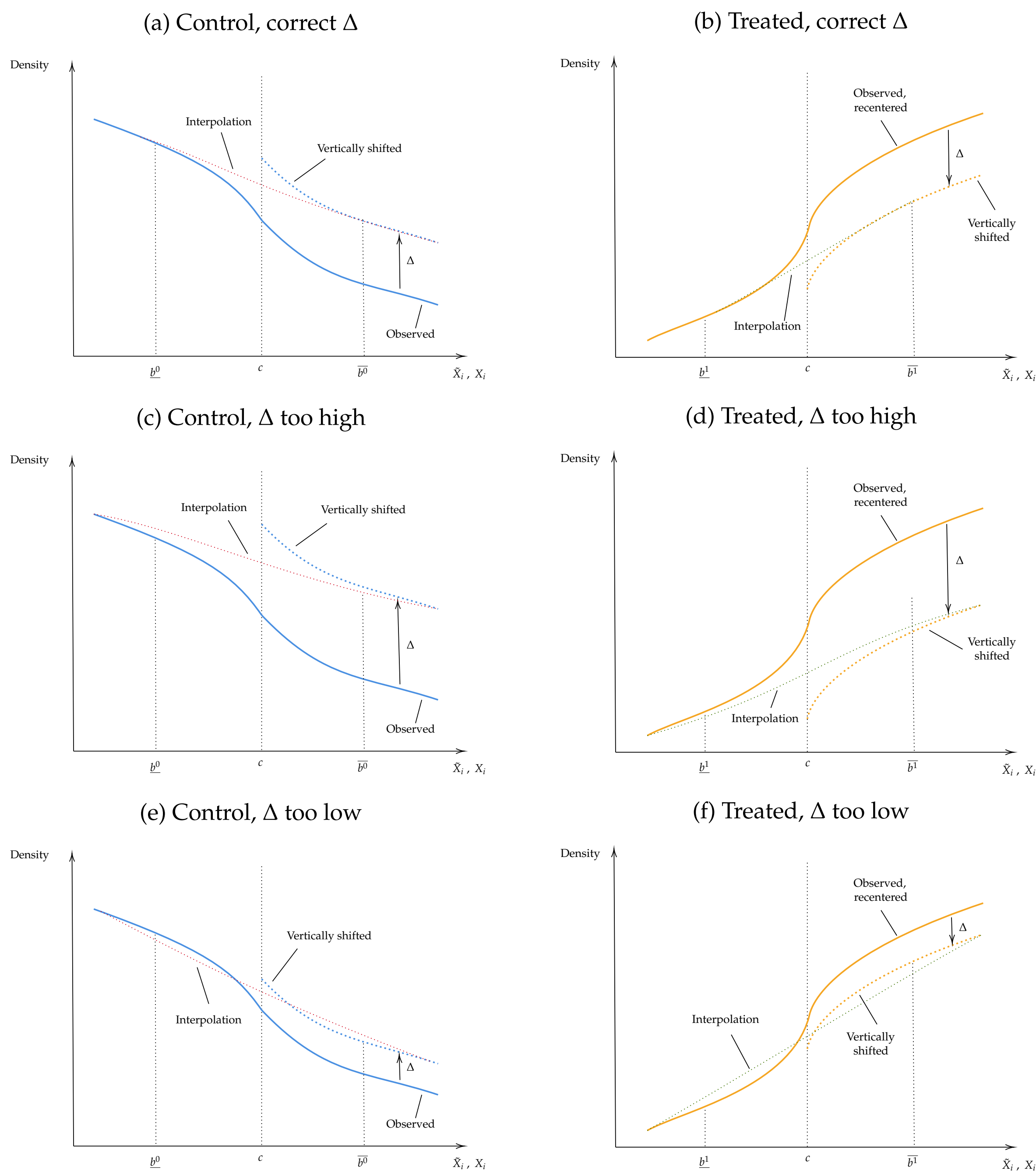
for $x_1 \in b_{x1}$, $b_{x1} \in \{\underline{b}^1, ..., 0\}$, $x_0 \in b_{x0}$, $b_{x0} \in \{\underline{b}^0, ..., 0\}$ and $\gamma' = \gamma - \widehat{\mu_\gamma}$. Note that all estimated magnitudes from the right-hand side are obtained in steps 1 and 2.

For values below the excluded regions, we impose $F_{\gamma'} = 0$ and $F_\xi = 0$, since below the excluded region there is no longer any systematic influence of measurement error. Finally, we obtain estimates for the CDF evaluated at points above the cutoff by imposing symmetry, so that $F_{\gamma'}(x) = 1 - F_{\gamma'}(-x)$ and $F_\xi(x) = 1 - F_\xi(-x)$.

3.2.4 Standard errors

For all of our estimates, we compute standard errors via bootstrap. We sample with replacement from the original distribution of contracts, and implement steps 1 through 3, obtaining a set of estimates $\hat{\theta}$. We repeat this process H times. The standard errors correspond to the empirical standard deviation of $\hat{\theta}^{(h)}$, for $h = \{1, 2, ..., H\}$.

Figure 5: Intuition of step 2



Notes: This figure provides graphical intuition of the procedure to estimate the size of the discontinuity in the latent distributions of the true assignment variable. Panels (a), (c), and (e) display frequency distributions of the control group. Panels (b), (d), and (f) show distributions of the treated group. In all cases, we also shift vertically the observed distribution above the cutoff. This step is based on the observation that, in the absence of measurement error, the distributions of treated and controls should be smooth and continuous, except for the discrete jump of size Δ at the cutoff. The key intuition is that a linear interpolation will fit poorly when the vertical shift above the cutoff does not reflect the correct Δ.

3.3 Accounting for Manipulation of the Assignment Variable

3.3.1 Introduction

We now consider an extension of the baseline model in which units may potentially "manipulate" the observed assignment variable. If units have some control over this variable, they may strategically affect it to influence their treatment status. In the health insurance example, individuals may have incentives to declare (both to authorities and to survey enumerators) an income just below the eligibility cutoff. In the procurement example, contracting officers may generate an estimate of the contract that puts them right below the publicity requirement threshold.

It is well known that this behavior invalidates the standard RD design and for this reason researchers routinely test for the absence of manipulation by checking the continuity of the assignment variable.[4] Recently, Gerard et al. (2020) showed that partial identification is generally possible under the presence of manipulation, and derive sharp bounds on the treatment effects of interest. We build on their results and adapt them to our framework with measurement error.

We restrict the manipulation behavior in three important ways. As Gerard et al. (2020), we allow only for one-sided selection. That is, we assume that units that manipulate their assignment variable always do it in one direction. For example, individuals may strategically under-report income to qualify for health insurance eligibility, but will never strategically over-report. Similarly, contracting officers may decrease but not increase their estimates. The "right" direction is of course context-dependent, but given our examples we assume, without loss of generality, that units may engage in downward manipulation only.

Second, we assume that units that engage in manipulation are successful in obtaining the treatment status that is consistent with their observed assignment variable. In other words, if units manipulate their assignment variable downwards in order to obtain the treatment (as in the health insurance example), then all manipulated units are treated. If instead the manipulation is driven by an incentive to avoid the treatment (as in the procurement contracts example), then all units that manipulate downwards are not treated. The plausibility of these behavioral assumptions is also context-dependent, but likely to be relevant in many applications, including our examples. As Gerard et al. (2020) show, imposing these simplifies the bounds of interest significantly.

Our third restriction on manipulation behavior is that it results in "bunching" right below the cutoff, and that this excess mass comes from units that are relatively close to the threshold from above. In other words, manipulation occurs only within a window around the cutoff. This assumption is standard in the bunching literature, and enables the use of interpolation methods to estimate the counterfactual density in the absence of bunching. Furthermore, it can be easily micro-founded by the existence of a cost of manipulation that is increasing in the size of the distortion.

[4] E.g., implementing the test developed by McCrary (2008).

3.3.2 Model Extension, Bounds, and Identification

We now make precise how these extensions affect the model. Let B_i be an indicator function that takes the value of 1 if unit i manipulates the observed assignment variable to bunch below the threshold. Allowing for this behavior, the relationship between observed and potential assignment variables becomes:

$$X_i = X_i^0(\tilde{X}_i) + D_i \cdot \left[X_i^1(\tilde{X}_i) - X_i^0(\tilde{X}_i)\right] + B_i \cdot (1 - D_i) \cdot \left[X_i^B(\tilde{X}_i) - X_i^0(\tilde{X}_i)\right] \tag{23}$$

where $X_i^B(\tilde{X}_i)$ is the observed assignment variable for units that engage in manipulation. Note that we have already impose that manipulated units are not treated (i.e., $B_i = 1 \implies D_i = 0$). We state the two other restrictions as an additional assumption to our model:

A4 Manipulated units bunch in some small region below the cutoff, and manipulation is not possible for units sufficiently above from the cutoff. That is, $X_i^B(\tilde{X}_i) \approx -\varepsilon$ for some small $\varepsilon \geq 0$, and there exists some $\bar{X}^B > 0$ such that $B_i = 0$ for all i with $\tilde{X}_i > \bar{X}^B$.

Following Gerard et al. (2020), it is straightforward to obtain sharp bounds on the local average treatment effects among complier units that do not manipulate their assignment variable.[5] All this requires is an estimate of the share of units right below the cutoff that engage in manipulation, which can be calculated as an excess mass relative to a counterfactual without manipulation. Bounds are then obtained by assuming that manipulated units are those with the most extreme values of the outcome variable, in the sense that they minimize or maximize the size of the treatment effect when these units are excluded. This implies chopping the tails of the distribution of outcomes Y_t for units below the threshold and in the control group. We show this implementation in Section 4.

The main hurdle introduced by this extension is obtaining an estimate of excess bunching right below the cutoff in the control group. The issue now is that the excess mass below the cutoff in the control group confounds strategic bunching and measurement error. If the treatment probability increases (decreases) at the cutoff, measurement error attenuates (strengthens) excess bunching coming from manipulation.

However, similar identification arguments as the ones presented above can be used to recover the latent distribution. Intuitively, the key insight is that the excess mass generated by bunching is manifested only in the distribution of control contracts, *and* that this behavior occurs only in one-direction, whereas measurement error generates symmetric noise moving observations both upwards and downwards. Once again, the identification argument critically relies on our ability to observe separately the densities of treated ($D_i = 1$) and control units ($D_i = 0$).

[5] All details can be found in Gerard et al. (2020). The assumptions we impose make our model fit with their special case in Appendix C.3., which has advantage that the computation of bounds does not involve optimization.

4 Empirical Application: Competition in Procurement Contracts

We illustrate how to implement our method in the context of an empirical application. We do so in the context of the discontinuous publicity mandates in US federal procurement, as studied by Carril, Gonzalez-Lira, and Walker (2022).

4.1 Setting

We observe a series of i.i.d. observations of procurement contracts. Outcomes Y_i observed includes the number of bids received in the solicitation, characteristics of the winning bidder, and post-award performance (e.g. delays, cost overruns). We also observe whether the solicitation was publicized in a government website or not ($D_i \in \{0, 1\}$), and are interested in the effect of this "treatment" on the observed outcomes.

Regulation favors the use of publicity if the expected value of the award ($\tilde{X}_i$) exceeds a cutoff of $c = 25,000$ dollars. We do not observe this latent assignment variable, but only ex-post realized award prices (X_i). Relative to ($\tilde{X}_i$), prices of treated contracts are measured with error, since publicizing them may have a causal effect on the value of the award ex-post (γ_i). We call these the price effects of publicity. For simplicity, we assume that there is no measurement error in the control group (i.e., $\xi_i = 0$ for all i), but we do allow for manipulation of the running variable. In particular, contracting officers may award a value right at or below the cutoff in order to avoid the publicizing requirement.

4.2 Empirical implementation details

We measure the assignment variable as the natural log of the award price, and re-center it around the cutoff. That is, if contract i is awarded for p_i, then $X_i = \log p_i - \log 25,000$. We then use bins of constant width of 0.01 log-points to generate our observations. Bin $b = 0$ includes all contracts with price greater than approximately $24,751[6] and smaller than or equal to $25,000. Our estimation is performed on a total set of 150 bins centered around zero, from -0.75 to 0.75. In dollar terms, this corresponds to contracts between $11,809 and $52,925.

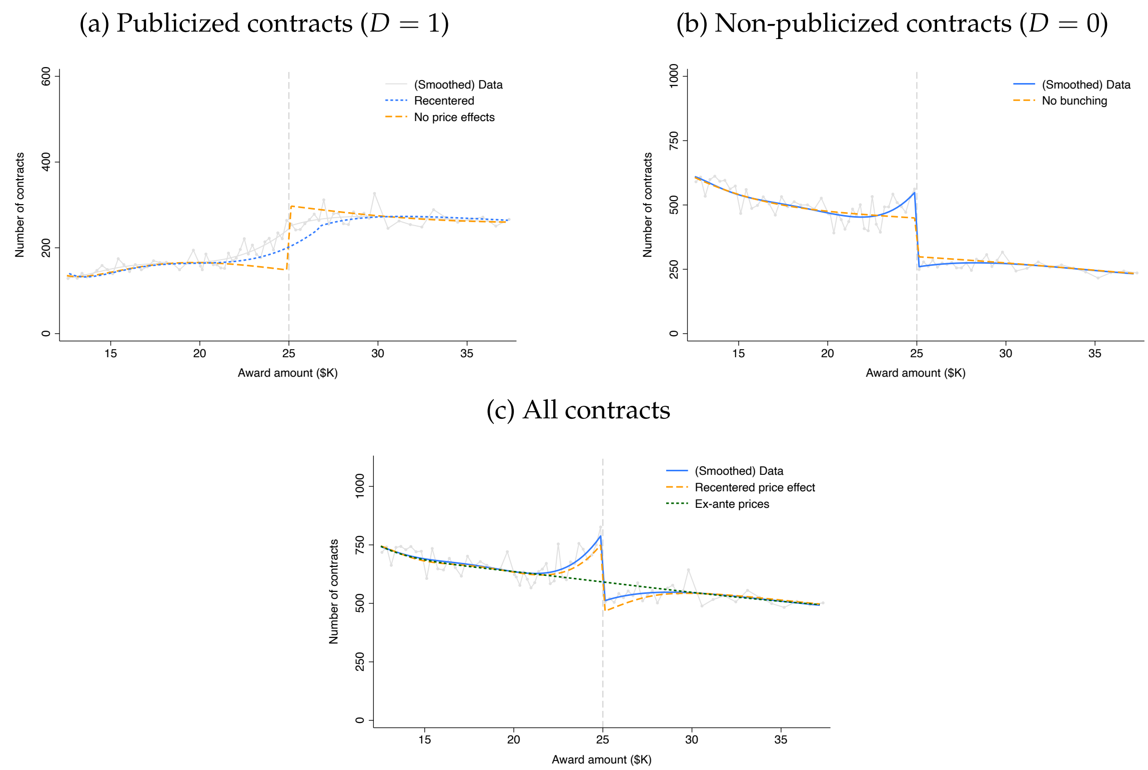
For the estimation, we use a fifth-degree polynomial, i.e. $Q = 5$. The excluded window for step 1 is symmetric, excluding 12 bins below zero and 12 bins above. In dollar terms, the excluded window consists of contracts between $22,173 and $28,187. Since bunching is asymmetric, for step 2 we choose only 5 bins below the cutoff and 12 bins above for Z^0, keeping the symmetric window of 12 bins above and below for Z^1.

[6] $\log(x) - \log(25,000) = 0.01 \iff x = 25,000 \cdot \exp(-0.01)$

4.3 Results of density analysis

Figure 6 shows the implementation of our method, yielding estimated distributions of the latent assignment variable, i.e., ex-ante price estimates. Panel (a) shows the distribution of the treatment group, i.e., publicized contracts, along with the latent distribution in which price effects (γ_i) are eliminated. Under our modelling assumptions, this coincides with the ex-ante price estimates observed by the buyers. Price effects smooth-out the distribution at the cutoff, and the observed excess mass to the right of the cutoff represents the set of contracts that had expected awards above the threshold, but were brought below it because of the competitive effect of publicizing the solicitations.

Figure 6: Results from density analysis

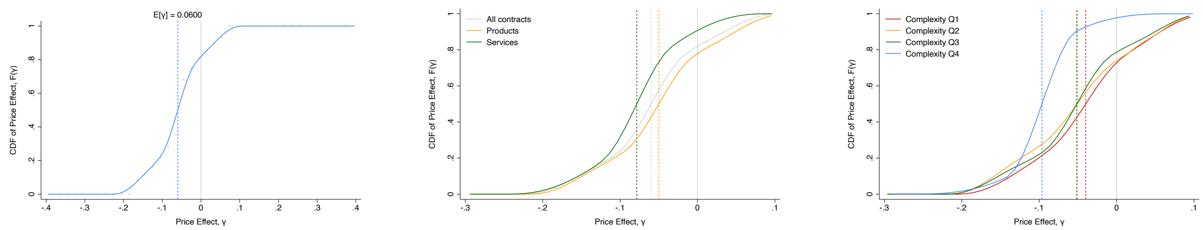


Notes: This figure shows the empirical distribution of the number of contracts at different price bins. Panel (a) shows the distribution of non-publicized contracts ($D = 0$). Panel (b) displays the distribution of publicized contracts ($D = 1$). Panel (c) displays the overall distribution, i.e., the sum of publicized and non-publicized contracts at every price. The blue line corresponds to a polynomial fit of degree five. The blue line corresponds to a polynomial fit of degree five. The orange dashed lines in panels (b) and (c) represent the counterfactual distributions in the absence of price effects and bunching. The counterfactual distributions stem from the proposed framework. In panel (a), The comparison between the solid blue and the dashed orange lines provide a visual interpretation of the mass of bunched contracts. The comparison between the dashed blue and the dashed orange lines in panel (b) inform visually about the distribution of price effects.

Panel (b) repeats the exercise for the control group, i.e., non-publicized contracts, this time shutting down the possibility of bunching responses. Again, this counterfactual distribution reflects the ex-ante awards. As expected, some of the mass just below the cutoff relocates to the right, while

Figure 7: Distribution of Price Effects

(a) Full Distribution (b) Good vs. Service (c) Quartile of Complexity



Notes: This figure presents the estimated CDF of the price effect parameters γ. The panel (a) shows the cumulative distribution function (CDF) of all contracts in the sample. Every gray dots show actual point estimates given a discretization of the support of γ. The blue line corresponds to a kernel fit. This estimation procedure builds upon comparing the empirical densities to a counterfactual distribution of publicized contracts assuming no price effect. The counterfactual distribution is generated from the interpolation of a polynomial of degree 5. The dashed vertical line corresponds to the estimated mean effect. The panel (b) shows the CDF of price effects separating contracts of goods and services. The panel (c) describes the CDF of price effects by quartile of complexity.

the rest goes to the publicized density to satisfy the integration constraint.

Finally, we present the effect of each of the two forces (bunching and price effects) on the total density of contracts in Panel (c). Based on this analysis, we can conclude that one-third of the total excess mass below the cutoff is due to price effects of competition, whereas the rest is explained by strategic bunching. Removing both effects gives us an estimated distribution of ex-ante award estimates that is completely smooth.

Once we know the density of the latent running variable, we can easily back out the distributions of the price effects γ_i. Figure 7(a) depicts the (nonparametric) estimate of the CDF of γ_t, along with a local polynomial smoothing. We find that publicity leads to an average reduction in award price of 0.06 log-points (SE: 0.02), equivalent to $ 1,456 at the discontinuity. From the full distribution, we see that publicizing contract opportunities reduces award prices for 83% of the contracts. Table 1 provides more details on the mean and variance of price effects and displays sub-group analyses. We find that price effects are higher for services, and the effects are larger for contracts that are more complex. [7]

4.4 Results of RD and Corrections

We now leverage the discontinuous nature of the publicity requirements to gauge the effects of publicity on a set of other relevant outcomes, including the level of competition, characteristics of the winning bidder, and post-award contractor performance. We use the estimates of price effects and bunching to adjust the RDD estimates accounting for these factors.

[7] By complexity, we refer to the average cost-overrun for all contracts in the product category valued under $20,000.

Table 1: Estimated Price Effects

Estimate / Sample	All	Goods	Services	Complexity Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean (μ_γ)	0.0595	0.0498	0.0782	0.0397	0.0505	0.0510	0.0962
	(0.0201)	(0.0622)	(0.0596)	(0.0475)	(0.1692)	(0.1908)	(0.0920)
Standard Deviation (σ_γ)	0.0643	0.0670	0.0534	0.0669	0.0739	0.0680	0.0369
	(0.0075)	(0.0084)	(0.0202)	(0.0140)	(0.0760)	(0.0295)	(0.0280)

Notes: This table shows the estimates corresponding to the effect of publicity on contract prices. The estimates result from analyzing the observed contract price density distribution relative to a counterfactual distribution. The observed densities are generated using bins of width \$250. The counterfactual distribution stems from a polynomial interpolation of degree 5. The standard deviation is calculated over the non-parametric distribution of γ. The standard errors are calculated through bootstrap. The subgroup analysis is performed independently for each group.

Naive RDD Results. We start with the naive RDD results which ignore the presence of measurement error or bunching responses, and then apply the relevant corrections. In our baseline specifications, we use a simple linear fit at both sides of the cutoff and no controls, but also present results from the robust local polynomial approach proposed by Calonico, Cattaneo, and Titiunik (2014). We present these naive RDD results visually, by plotting binned scatters of the first stage and reduced form relationships.

The results for the first stage are presented graphically in Figure 8(a). We see that the use of FedBizOpps jumps sharply past the \$25,000 threshold of award amounts. The share of contracts that are publicly solicited in the government platform increases from roughly 30% at or slightly below \$25,000, to 50% right above this cutoff.

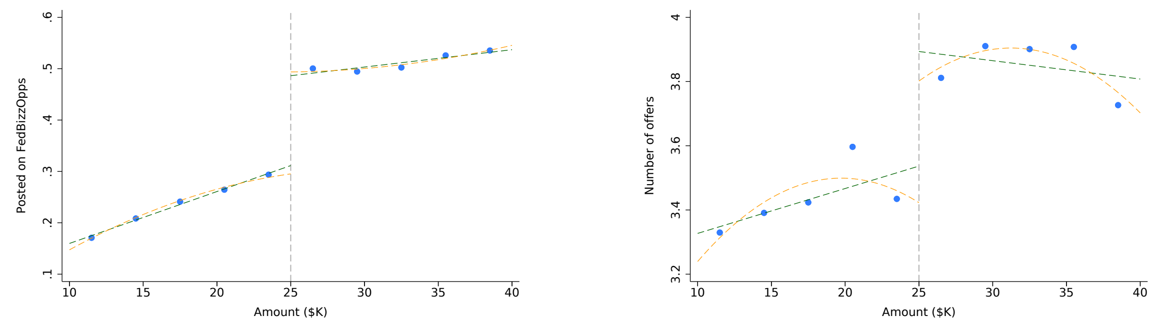
We show reduced form specifications for three sets of outcomes: the intensity of competition, winning vendor characteristics (including its relationship with the awarding office), and post-award performance. Figure 8(b) shows how posting solicitations on FedBizOpps impacts the number of offers that a contract receives around the cutoff. Contracts right above \$25,000 (which are more likely to be publicly solicited), receive roughly 0.4 more bids. The magnitude of the increase in the number of offers is considerable given that the policy only changes the likelihood of a publicized solicitation by around 20 p.p.

Figure 9 shows reduced form specifications of winner characteristics. In Panel (a), we see that publicized contracts are awarded to vendors that are relatively larger, as measured by a reduction of the probability of awarding the contract to a small firm. Panel (b) and Panel (c) show that publicized contracts are more likely to be awarded to foreign firms or firms that are located geographically more distant from the contracting office location.

To measure the impact , we use two measures that are commonly used in the literature: cost

Figure 8: Publicity requirement and intensity of competition

(a) Use of FedBizOpps (b) Number of offers received



Notes: Panels (a) shows the fraction of contracts posted on FedBizOpps. Panel (b) shows the number of offers received. Each graph includes the averages by bins of award amounts (blue dots), as well as linear and quadratic fits at each side of $25,000. The data sources are FBO.gov and the Federal Procurement Data System-Next Generation. The sample consists of competitive, non-R&D, definitive contracts and purchase orders, with award values between $ 10,000 and $ 40,000, awarded by the Department of Defense in fiscal years 2015 through 2019. Award amounts are discretized into right-inclusive bins of $3,000 dollars length.

overruns and delays (e.g. Decarolis, 2014; Kang and Miller, 2017; Decarolis et al., 2020; Carril, 2019). Because the data contain the total sum of payments and completion date expected at the time of the award for each contract, we can construct measures of cost overruns and delays by comparing these expectations to the realized payments and duration.

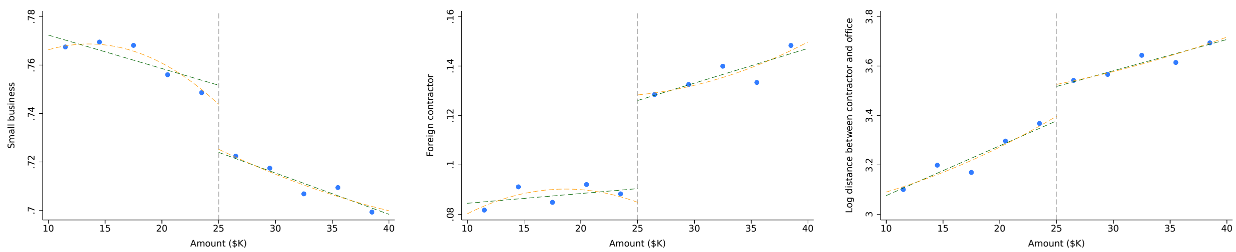
Figure 10 presents results on two measures of post-award contract performance: cost overruns and delays. We find that the share of contracts with overruns and the share of contracts with delays increase by 2 p.p. and 1.5 p.p., respectively. These differences are statistically and economically considering the magnitude of the first stage.

Adjusted RDD Results. Now we consider refinements to our naive RDD results. First, we explore robustness to our baseline linear specification with the estimator proposed by Calonico, Cattaneo, and Titiunik (2014), which uses robust local polynomial fits. Second, building upon the results of our density analysis, we adjust the baseline RDD estimates to account for the observed running variable (award price) being subject to both measurement error due to treatment effects (price effects of publicity) and potential manipulation (bunching).

Table 2 presents reduced-form estimates for each relevant outcome variable. The first column shows the coefficient of our naive linear RDD using ordinary least squares (OLS). These results replicate the RDD plots discussed earlier. Column (2) presents Calonico et al. (2014)'s local polynomial estimates with robust bias-corrected standard errors. Overall, non-linear estimates are similar in magnitude and significance to simple OLS estimates. The third column present estimates that account for price effects in the treatment group (i.e. publicized contracts), following our correction method. Accounting for price effects has a relatively modest impact, and in most cases tends to amplify the naive results. This is consistent with the fact that price effects smooth-out the discontinuity for the treatment group. Thus, under naive estimation, some publicized contracts are

Figure 9: Publicity and the characteristics of the winning firm

(a) Contractor is a small business (b) Foreign firm (c) Distance to the office



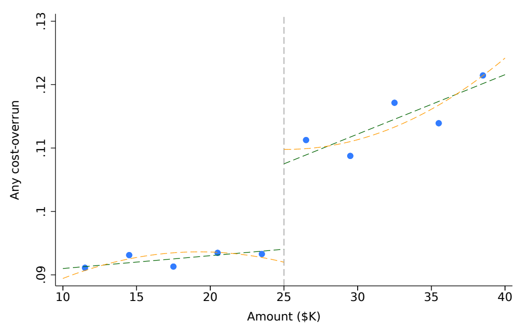
Notes: This figure presents four binned scatter plots, which depict an average outcome by bins of award amounts, as well as linear and quadratic fits at each side of $25,000. The outcome in each Panel is as follows: (a) indicator equal to one if the awarded contractor is a small business (based on SBA); (b) an indicator equal to one if the contract is awarded to a foreign vendor; (c) the natural logarithm of the distance (in miles) from the contracting office's location and the vendor location. The data source is the Federal Procurement Data System-Next Generation. The sample consists of non-R&D, definitive, and competitively awarded contracts and purchase orders, with award values between $ 10,000 and $ 40,000, awarded by the Department of Defense in fiscal years 2015 through 2019. Award amounts are discretized into right-inclusive bins of $3,000 dollars length.

observed below the threshold when their original (ex-ante) price was above it.

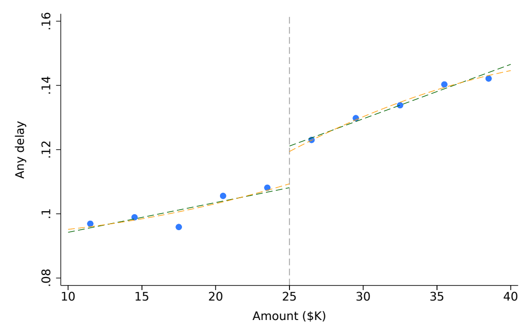
The next two columns present partial identification estimates that account for bunching responses. Column (4) shows lower and upper bounds without accounting for price effects, while the fifth column shows bounds that do adjust for price effects. Notably, since the magnitude of bunching is modest in our context, the bounds presented are relatively narrow, which tells us that bunching does not pose a serious threat to the interpretation of our results. Interestingly, the lower bounds in Column (5) tend to be very close to our baseline estimates. This implies that the downward bias introduced by price effects on the naive estimates of Column (1) is of similar magnitude than the worst-case upward bias introduced by bunching responses.

Figure 10: Publicity and post-award contract performance

(a) Share of contracts experiencing cost-overruns



(b) Share of delayed contracts



Notes: This figure presents four binned scatter plots, which depict an average outcome by bins of award amounts, as well as linear and quadratic fits at each side of $25,000. The outcome in each Panel is as follows: (a) the share of contracts that the actual obligated contract dollars exceed expected total obligations at the time of the award (i.e., cost-overruns); (b) the share of contracts whose actual days of contract duration exceed the expected days of duration at the time of the award (i.e., delays). The data source is the Federal Procurement Data System-Next Generation. The sample consists of non-R&D, definitive, and competitive contracts and purchase orders, with award values between $ 10,000 and $ 40,000, awarded by the Department of Defense in fiscal years 2015 through 2019. Award amounts are discretized into right-inclusive bins of $3,000 dollars length.

Table 2: Reduced-form RDD Estimates and Corrections

Dependent Variable	OLS (1)	CCT (2)	Price Effect Adjustment (3)	Manipulation Bounds (4)	Price Effect + Manip. Bounds (5)
Number of offers	0.3569 (0.0677)	0.5447 (0.1053)	0.3526	[0.2762 , 0.5344]	[0.3073 , 0.4506]
One offer	-0.0191 (0.0064)	-0.0235 (0.0108)	-0.0204	[-0.0272 , 0.0052]	[-0.0248 , -0.0070]
Log distance firm-office	0.1392 (0.0481)	0.1199 (0.0817)	0.1909	[0.0290 , 0.2688]	[0.1304 , 0.2619]
Foreign firm	0.0357 (0.0045)	0.0508 (0.0078)	0.0375	[0.0328 , 0.0520]	[0.0358 , 0.0465]
New firm	0.0175 (0.0075)	0.0185 (0.0126)	0.0247	[0.0023 , 0.0350]	[0.0164 , 0.0344]
Small business	-0.0277 (0.0065)	-0.0295 (0.0110)	-0.0265	[-0.0523 , -0.0195]	[-0.0399 , -0.0220]
Any cost-overrun	0.0135 (0.0045)	0.0246 (0.0077)	0.0144	[0.0103 , 0.0263]	[0.0127 , 0.0216]
Cost-overruns (relative dollars)	0.0095 (0.0058)	0.0161 (0.0100)	0.0127	[0.0053 , 0.0179]	[0.0104 , 0.0174]
Any delay	0.0130 (0.0047)	0.0151 (0.0080)	0.0143	[0.0094 , 0.0270]	[0.0123 , 0.0222]
Delays (days)	2.3262 (2.0388)	4.0361 (3.4935)	2.7491	[1.2278 , 5.3554]	[2.1492 , 4.4660]
Number of modifications	0.0375 (0.0173)	0.0619 (0.0300)	0.0395	[0.0204 , 0.0926]	[0.0300 , 0.0701]

Notes: This table shows Regression Discontinuity Design (RDD) estimates of the reduced-form relationship between a series of outcome variables and an indicator of whether a contract award price exceeds $25,000. Coefficients in column (1) use a linear fit above and below the discontinuity. Coefficients in column (2) correspond to the robust local polynomial method proposed by Calonico, Cattaneo, and Titiunik (2014). Column (3) applies a correction to the estimates in column (1), accounting for the existence of price-effects, following our correction method. Column (4) shows bounds on the reduced-form coefficient in column (1), accounting for the possibility of "running variable manipulation" (i.e. bunching). Column (5) shows bounds on the adjusted reduced-form coefficient in column (4), accounting for both the existence of price-effects and the possibility of "running variable manipulation" (i.e. bunching). Standard errors for the coefficients in columns (1) and (2) are shown in parentheses.

5 Conclusion

In this paper, we study a Regression Discontinuity Design (RDD) framework in which the observed running variable is subject to two sources of measurement error. On top of standard mean-zero classical measurement error affecting all units, we allow treated observations to be subject to an additional non-mean-zero measurement error, reflecting the potential influence of treatment effects in the observed assignment variable. We discuss how this affects the standard framework and we prove that the RDD parameters of interest can be recovered given the measurement error distributions. We then argue that the distributions of these errors can be identified from the observed densities of treated and control units. Based on these ideas, we develop a method to estimate the error distributions, which allows us to recover the estimands of interest of the RDD. We show the implementation of our method in the context of the empirical setting studied by Carril, Gonzalez-Lira, and Walker (2022).

References

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

Carril, R. (2019). Rules Versus Discretion in Public Procurement.

Carril, R., A. Gonzalez-Lira, and M. S. Walker (2022). Competition under Incomplete Contracts and the Design of Procurement Policies. *Mimeo*.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013). Inference on Counterfactual Distributions. *Econometrica 81*(6), 2205–2268.

De La Mata, D. (2012). The effect of medicaid eligibility on coverage, utilization, and children's health. *Health economics 21*(9), 1061–1079.

Decarolis, F. (2014). Awarding price, contract performance, and bids screening: Evidence from procurement auctions. *American Economic Journal: Applied Economics 6*(1 A), 108–132.

Decarolis, F., L. M. Giuffrida, E. Iossa, V. Mollisi, and G. Spagnolo (2020). Bureaucratic Competence and Procurement Outcomes. *The Journal of Law, Economics, and Organization No. 24201*.

Gerard, F., M. Rokkanen, and C. Rothe (2020). Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. *Quantitative Economics 11*(3), 839–870.

Hullegie, P. and T. J. Klein (2010). The effect of private health insurance on medical care utilization and self-assessed health in germany. *Health economics 19*(9), 1048–1062.

Jales, H. and Z. Yu (2017). Identification and estimation using a density discontinuity approach. *Advances in Econometrics 38*, 29–72.

Kang, K. and R. A. Miller (2017). Winning by Default: Why is there so Little Competition in Government Procurement?

Kleven, H. (2016). Bunching. *Annual Review of Economics*.

Kleven, H. J. and M. Waseem (2013). Using Notches To Uncover Optimization Frictions And Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics 128*(December), 669–723.

Koch, T. G. (2013). Using rd design to understand heterogeneity in health insurance crowd-out. *Journal of Health Economics 32*(3), 599–611.

McCrary, J. (2008, February). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Pei, Z. and Y. Shen (2017). The devil is in the tails: Regression discontinuity design with measurement error in the assignment variable. *Advances in Econometrics 38*(October), 455–502.

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

Appendix

A Proof of Proposition 1

The goal is to derive an expression for $E[Y_i|X_i]$ that depends on structural parameters $(\mathbf{ff}, \mathbf{fi})$ and objects that we can either observe or estimate.

A.1 Key equations of the model

We start by stating the key equations that characterize the model. Rewrite (1) and (2) as:

$$E[Y_i|\tilde{X}_i] = \alpha_0 + \beta_0 \cdot \tilde{X}_i + \left[(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot \tilde{X}_i\right] \cdot \mathbf{1}[\tilde{X}_i > 0] \tag{24}$$

$$E[D_i|\tilde{X}_i] = \Pr(D_i = 1|\tilde{X}_i) = \tilde{\pi}(\tilde{X}_i) + \delta \cdot \mathbf{1}[\tilde{X}_i > 0] \tag{25}$$

And note that (5), **A2** and **A3** imply:

$$X_i = \tilde{X}_i + (1 - D_i) \cdot \xi_i + D_i \cdot \gamma_i \tag{26}$$

A.2 Useful definitions

We now define a series of terms that will later prove useful in finding expressions for our objects of interest. We also express them in terms of observable or estimable objects using the law of iterated expectations, the law of total probability and/or Bayes' rule.

First, let $\Lambda \equiv \Pr(\tilde{X}_i \leq 0|X_i)$. Note that:

$$\begin{aligned}
\Lambda &= \Pr(\tilde{X}_i \leq 0|X_i) \\
&= \Pr(\tilde{X}_i \leq 0|X_i, D_i = 0) \cdot \Pr(D_i = 0|X_i) + \Pr(\tilde{X}_i \leq 0|X_i, D_i = 1) \cdot \Pr(D_i = 1|X_i) \\
&= \Pr(X_i - \xi_i \leq 0|X_i) \cdot [1 - E[D_i|X_i]] + \Pr(X_i - \gamma_i \leq 0|X_i, D_i = 1) \cdot E[D_i|X_i] \\
\Lambda &= \left[1 - F_{\xi}(X_i)\right] \cdot [1 - E[D_i|X_i]] + [1 - F_{\gamma}(X_i)] \cdot E[D_i|X_i]
\end{aligned} \tag{27}$$

Second, let $\Pi \equiv \Pr(D_i = 1|X_i, \tilde{X}_i \leq 0)$. We have:

$$\begin{aligned}
\Pi &= \Pr(D_i = 1|X_i, \tilde{X}_i \leq 0) \\
&= \frac{\Pr(\tilde{X}_i \leq 0|X_i, D_i = 1) \cdot \Pr(D_i = 1|X_i)}{\Pr(\tilde{X}_i \leq 0|X_i)} \\
&= \frac{\Pr(X_i - \gamma_i \leq 0|X_i) \cdot E[D_i|X_i]}{\Lambda} \\
\Pi &= \frac{[1 - F_{\gamma}(X_i)] \cdot E[D_i|X_i]}{\Lambda}
\end{aligned} \tag{28}$$

Third, let $\Sigma \equiv \Pr(D_i = 1|X_i, \tilde{X}_i > 0)$. We have:

$$\begin{aligned}
\Sigma &= \Pr(D_i = 1|X_i, \tilde{X}_i > 0) \\
&= \frac{\Pr(\tilde{X}_i > 0|X_i, D_i = 1) \cdot \Pr(D_i = 1|X_i)}{\Pr(\tilde{X}_i > 0|X_i)} \\
&= \frac{\Pr(X_i - \gamma_i > 0|X_i) \cdot E[D_i|X_i]}{1 - \Lambda} \\
\Sigma &= \frac{F_\gamma(X_i) \cdot E[D_i|X_i]}{1 - \Lambda} \qquad (29)
\end{aligned}$$

Finally, let $\Omega_0 \equiv E\left[\tilde{X}_i \middle| X_i, \tilde{X}_i \leq 0\right]$ and $\Omega_1 \equiv E\left[\tilde{X}_i \middle| X_i, \tilde{X}_i > 0\right]$. We have:

$$\begin{aligned}
\Omega_0 &= E\left[\tilde{X}_i \middle| X_i, \tilde{X}_i \leq 0\right] \\
&= E\left[\tilde{X}_i|X_i, \tilde{X}_i \leq 0, D_i = 1\right] \cdot \Pi + E\left[\tilde{X}_i|X_i, \tilde{X}_i \leq 0, D_i = 0\right] \cdot (1 - \Pi) \\
&= E\left[X_i - \gamma|X_i, X_i \leq \gamma\right] \cdot \Pi + E\left[X_i - \xi|X_i, X_i \leq \xi\right] \cdot (1 - \Pi) \\
&= X_i - \Pi \cdot E\left[\gamma|\gamma \geq X_i\right] - (1 - \Pi) \cdot E\left[\xi|\xi \geq X_i\right] \qquad (30)
\end{aligned}$$

$$\begin{aligned}
\Omega_1 &= E\left[\tilde{X}_i \middle| X_i, \tilde{X}_i > 0\right] \\
&= E\left[\tilde{X}_i|X_i, \tilde{X}_i > 0, D_i = 1\right] \cdot \Sigma + E\left[\tilde{X}_i|X_i, \tilde{X}_i > 0, D_i = 0\right] \cdot (1 - \Sigma) \\
&= E\left[X_i - \gamma|X_i, X_i > \gamma\right] \cdot \Sigma + E\left[X_i - \xi|X_i, X_i > \xi\right] \cdot (1 - \Sigma) \\
&= X_i - \Sigma \cdot E\left[\gamma|\gamma < X_i\right] - (1 - \Sigma) \cdot E\left[\xi|\xi < X_i\right] \qquad (31)
\end{aligned}$$

A.3 Conditional expectation

We now find the expression for $E[Y_i|X_i]$ that is stated in Proposition 1. Using the Law of Total Probability:

$$E[Y_i|X_i] = \underbrace{\Pr(\tilde{X}_i \leq 0|X_i)}_{\Lambda} \cdot E[Y_i|X_i, \tilde{X}_i \leq 0]) + \underbrace{\Pr(\tilde{X}_i > 0|X_i)}_{1-\Lambda} \cdot E[Y_i|X_i, \tilde{X}_i > 0] \qquad (32)$$

Using the law of iterated expectations and equation (24):

$$\begin{aligned}
E[Y_i|X_i, \tilde{X}_i \leq 0] &= E\left[E[Y_i|\tilde{X}_i, X_i, \tilde{X}_i \leq 0]|X_i, \tilde{X}_i \leq 0\right] \\
&= E\left[\alpha_0 + \beta_0 \tilde{X}_i|X_i, \tilde{X}_i \leq 0\right] \\
&= \alpha_0 + \beta_0 \cdot \Omega_0 \qquad (33)
\end{aligned}$$

Analogously:

$$\begin{aligned}
E[Y_i|X_i, \tilde{X}_i > 0] &= E\left[E[Y_i|\tilde{X}_i, X_i, \tilde{X}_i > 0]|X_i, \tilde{X}_i > 0\right] \\
&= E\left[\alpha_1 + \beta_1 \tilde{X}_i|X_i, \tilde{X}_i > 0\right] \\
&= \alpha_1 + \beta_1 \cdot \Omega_1 \qquad (34)
\end{aligned}$$

Combining (32), (33), and (34):

$$E[Y_i|X_i] = \alpha_0 \cdot \Lambda + \alpha_1 \cdot (1-\Lambda) + \beta_0 \cdot \Lambda\Omega_0 + \beta_1 \cdot (1-\Lambda)\Omega_1 \tag{35}$$

And finally, using the definitions of Λ, Π, Σ, Ω_0, and Ω_1, we obtain:

$$E[Y_i|X_i] = \alpha_0 \cdot \phi_1(X_i) + \alpha_1 \cdot \phi_2(X_i) + \beta_0 \cdot \phi_3(X_i) + \beta_1 \cdot \phi_4(X_i) \tag{36}$$

where:

$$\begin{aligned}
\phi_1(X_i) &\equiv \left[1 - F_\xi(X_i)\right] \cdot [1 - E[D_i|X_i]] + [1 - F_\gamma(X_i)] \cdot E[D_i|X_i] \\
\phi_2(X_i) &\equiv F_\xi(X_i) \cdot [1 - E[D_i|X_i]] + F_\gamma(X_i) \cdot E[D_i|X_i] \\
\phi_3(X_i) &\equiv \phi_1(X_i) \cdot X_i - [1 - F_\gamma(X_i)] \cdot E[D_i|X_i] \cdot E\left[\gamma|\gamma \geq X_i\right] - \left[1 - F_\xi(X_i)\right] \cdot [1 - E[D_i|X_i]] \cdot E\left[\xi|\xi \geq X_i\right] \\
\phi_4(X_i) &\equiv \phi_2(X_i) \cdot X_i - F_\gamma(X_i) \cdot E[D_i|X_i] \cdot E\left[\gamma|\gamma < X_i\right] - F_\xi(X_i) \cdot [1 - E[D_i|X_i]] \cdot E\left[\xi|\xi < X_i\right]
\end{aligned} \tag{37}$$

□

B From Reduced Form to LATE: Accounting for First Stage

Proposition 1 implies that we may, in principle, be able to estimate α_0 and α_1 via OLS, which is sufficient to compute an estimate of τ_{RF}. However, we may often be interested in the causal parameter τ_{LATE}, which requires an estimate of δ to scale the reduced form. Here we show that this is possible by deriving an expression for $E\left[D_i|\, X_i\right]$ that depends on the structural parameter δ, as well as observable and estimable objects. The resulting expression is not linear, however, so a nonlinear estimation method must be used (e.g., nonlinear least squares).

Consider $E[D_i|X_i]$. Using the law of iterated expectations, and then the law of total probability twice, we have:

$$\begin{aligned}
E[D_i|X_i] &= E\left[\, E[D_i|\tilde{X}_i, X_i] \middle|\, X_i\right] \\
&= E\left[\, \tilde{\pi}(\tilde{X}_i) + \delta \cdot \mathbf{1}[\tilde{X}_i > 0] \middle|\, X_i\right] \\
&= \Lambda \cdot E\left[\, \tilde{\pi}(\tilde{X}_i) \middle|\, X_i, \tilde{X}_i \leq 0\right] + (1-\Lambda) \cdot E\left[\, \tilde{\pi}(\tilde{X}_i) + \delta \middle|\, X_i, \tilde{X}_i > 0\right] \\
&= \Lambda \cdot \left\{\Pi \cdot E\left[\, \tilde{\pi}(X_i - \gamma)|\, X_i, Xi \leq \gamma\right] + (1-\Pi) \cdot E\left[\, \tilde{\pi}(X_i - \xi)|\, X_i, Xi \leq \xi\right] \right\} \\
&\quad + (1-\Lambda) \cdot \left\{\Sigma \cdot E\left[\, \tilde{\pi}(X_i - \gamma) + \delta|\, X_i, Xi > \gamma\right] \right. \\
&\qquad\qquad \left. + (1-\Sigma) \cdot E\left[\, \tilde{\pi}(X_i - \xi) + \delta|\, X_i, Xi > \xi\right] \right\}
\end{aligned} \tag{38}$$

Using the definitions of Λ and Σ, and rearranging:

$$\begin{aligned}
E[D_i|X_i] = {} & [1 - F_\gamma(X_i)] \cdot E[D_i|X_i] \cdot E\left[\,\tilde{\pi}(X_i - \gamma)|\, X_i, Xi \leq \gamma\right] \\
& + \left[1 - F_\xi(X_i)\right] \cdot [1 - E[D_i|X_i]] \cdot E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i, Xi \leq \xi\right] \\
& + F_\gamma(X_i) \cdot E[D_i|X_i] \cdot E\left[\,\tilde{\pi}(X_i - \gamma) + \delta|\, X_i, Xi > \gamma\right] \\
& + F_\xi(X_i) \cdot [1 - E[D_i|X_i]] \cdot E\left[\,\tilde{\pi}(X_i - \xi) + \delta|\, X_i, Xi > \xi\right] \} \\
= {} & \left[1 - F_\xi(X_i)\right] \cdot E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i, Xi \leq \xi\right] + F_\gamma(X_i) \cdot E\left[\,\tilde{\pi}(X_i - \gamma) + \delta|\, X_i, Xi > \gamma\right] \\
& - E[D_i|X_i] \cdot \{ \left[1 - F_\xi(X_i)\right] \cdot E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i, Xi \leq \xi\right] \\
& \qquad + F_\xi(X_i) \cdot E\left[\,\tilde{\pi}(X_i - \xi) + \delta|\, X_i, Xi > \xi\right] \\
& \qquad - [1 - F_\gamma(X_i)] \cdot E\left[\,\tilde{\pi}(X_i - \gamma)|\, X_i, Xi \leq \gamma\right] \\
& \qquad - F_\gamma(X_i) \cdot E\left[\,\tilde{\pi}(X_i - \gamma) + \delta|\, X_i, Xi > \gamma\right] \} \\
= {} & \left[1 - F_\xi(X_i)\right] \cdot E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i, Xi \leq \xi\right] + F_\gamma(X_i) \cdot E\left[\,\tilde{\pi}(X_i - \gamma) + \delta|\, X_i, Xi > \gamma\right] \\
& - E[D_i|X_i] \cdot \left\{E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i\right] - E\left[\,\tilde{\pi}(X_i - \gamma)|\, X_i\right] + \delta\left[F_\xi(X_i) - F_\gamma(X_i)\right]\right\}
\end{aligned}$$

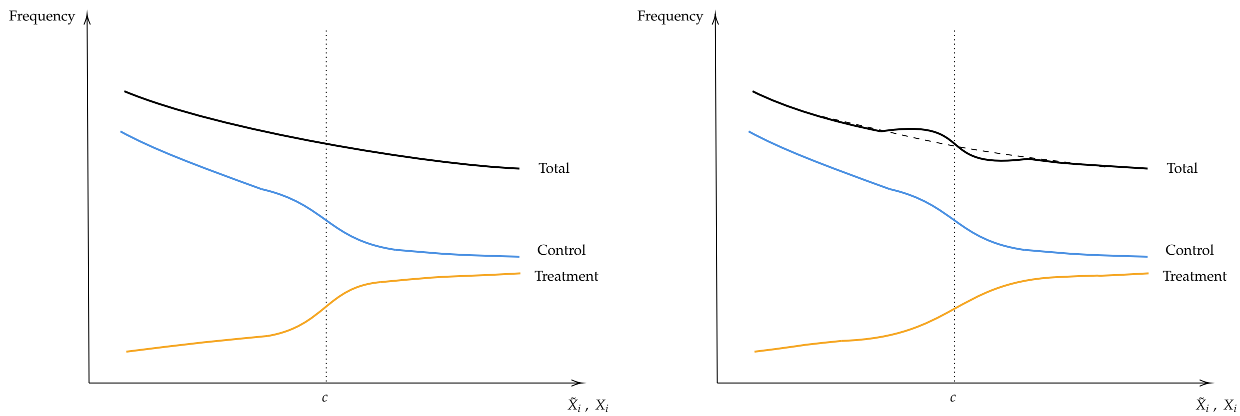
$$E[D_i|X_i] = \frac{\left[1 - F_\xi(X_i)\right] \cdot E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i, Xi \leq \xi\right] + F_\gamma(X_i) \cdot E\left[\,\tilde{\pi}(X_i - \gamma)|\, X_i, Xi > \gamma\right]}{1 + E\left[\,\tilde{\pi}(X_i - \xi)|\, X_i\right] - E\left[\,\tilde{\pi}(X_i - \gamma)|\, X_i\right] + \delta\left[F_\xi(X_i) - F_\gamma(X_i)\right]} \tag{39}$$

The expression implies that knowledge of F_γ, F_ξ and $\tilde{\pi}$ implies that δ can be estimated with a nonlinear estimation method (e.g., nonlinear least squares).

C Additional Figures

Figure A1: Impact of measurement errors depending on relative size of their variance

(a) $\sigma_\gamma = \sigma_\xi > 0 \; ; \; \mu_\gamma = 0$ (b) $\sigma_\gamma > \sigma_\xi > 0 \; ; \; \mu_\gamma = 0$

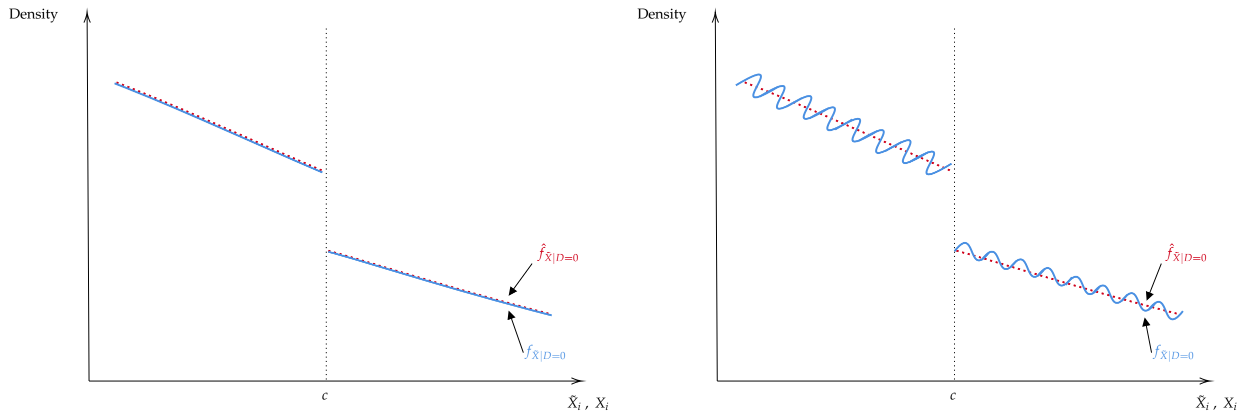


Notes: This figure shows an example of how measurement errors affect the observed distributions of treated and control observations, as well as the total distribution corresponding to the sum of the first two. The depicted cases involve mean-zero measurement errors affecting both the treatment and the control group. In Panel(a), we show a case in which both errors have equal variance. As a result, the total distribution is perfectly smooth. In Panel(b), we show a case in in which the error in the treatment group has a larger variance than the error in the control group. As a result, the total distribution, while continuous, is not smooth around the cutoff.

Figure A2: Quasi-density and curvature

(a) Smooth case

(b) Non-smooth case



Notes: This figure shows the relationship between the curvature of the latent density and the estimated quasi-density. As shown in Panel (a), if the latent density is close to linear, the estimated quasi-density will closely approximate it. As shown in Panel (b), if the latent density is non-smooth, then the estimated quasi-density will reflect a smooth approximation of the latent density.