When Can We Ignore Measurement Error in the Running Variable?*

Yingying Dong†
University of California Irvine

Michal Kolesár‡
Princeton University

February 1, 2023

Abstract

In many applications of regression discontinuity designs, the running variable used by the administrator to assign treatment is only observed with error. We show that, provided the observed running variable (i) correctly classifies the treatment assignment, and (ii) affects the conditional means of the potential outcomes smoothly, ignoring the measurement error nonetheless yields an estimate with a causal interpretation: the average treatment effect for units whose observed running variable equals to the cutoff. We show that, possibly after doughnut trimming, these assumptions accommodate a variety of settings where support of the measurement error is not too wide. We propose to conduct inference using bias-aware methods, which remain valid even when discreteness or irregular support in the observed running variable may lead to partial identification. We illustrate the results for both sharp and fuzzy designs in an empirical application.

*We thank Tim Armstrong, Xiaohong Chen, and Christoph Rothe, conference participants at the ASSA 2023 Annual Meeting, and seminar participants at the University of Toronto for helpful comments. Kolesár acknowledges support by the Sloan Research Fellowship and by the National Science Foundation Grant SES-22049356. All errors are our own.

†email: yyd@uci.edu

‡email: mkolesar@princeton.edu

1 Introduction

The key characteristic of regression discontinuity (RD) designs is that assignment of units to treatment is determined by whether the value of a particular covariate X^*, called the running variable, exceeds a fixed threshold c. Under weak continuity conditions, comparing units on either side of the threshold identifies the average treatment effect (ATE) for units with $X^* = c$. However, in many cases, researchers only observe a noisy version X of the true running variable X^*. In our survey of papers published in leading economics journals that featured RD designs, 23% of them used a running variable measured with error.[1] Most commonly, the noise arose due to rounding or grouping, such as when researchers observe age in years, income reported in income brackets, or job tenure in months, while the administrator assigning treatment uses the exact birthdate, income, or job tenure.

The prevalence of this problem has given rise to a growing literature on measurement error in RD designs (see, among others, Barreca et al., 2016; Bartalotti et al., 2021; Davezies and Le Barbanchon, 2017; Dong, 2015; Hullegie and Klein, 2010; Pei and Shen, 2017; or Dieterle et al., 2020). As this literature points out, ignoring the measurement error can lead to inconsistent estimates of the usual RD estimand, the ATE for units with $X^* = c$.[2] This motivated the development of a variety of alternative estimation and inference procedures to recover this ATE; the solutions depend on the particular auxiliary assumptions about the form of the measurement error, or the availability of auxiliary datasets.

This paper makes the simple point that under easily interpretable conditions, existing RD techniques provide inference for a slightly different estimand, the ATE for units with $X = c$. That is, one could interpret the RD analysis as if X were the true running variable. For instance, suppose X^* is birthdate, but we only observe the year of birth X. While standard RD analysis does not yield valid inference for the RD estimand associated with X^*, the ATE for individuals born on the cutoff date, it does provide valid inference for the ATE for individuals born in the cutoff year.[3]

This result relies on two key conditions: (i) using X as a running variable correctly classifies the treatment assignment, and (ii) the conditional means of the potential outcomes are smooth in X. The first condition holds automatically for certain types of rounding

[1] See Appendix A.3 for details.

[2] As we discuss in more detail in Section 2.3, an exception is Battistin et al. (2009), who, in the context of a fuzzy RD design, consider measurement error with a point mass at zero, and give conditions under which ignoring the measurement error nonetheless yields a consistent estimate of a fuzzy RD analog of this estimand.

[3] Equivalently, this estimand is a weighted average of ATEs conditional on X^*, with weights given by the density of the measurement error conditional on $X = c$. In the example, it corresponds to a weighted average of ATEs for individuals born on each day of the cutoff year, weighted by the relative birthdate frequencies.

error; more generally it requires removing observations at or in the immediate vicinity of the threshold, resulting in a "doughnut design". Our approach is thus most relevant when the support of the measurement error is relatively narrow (which includes most settings with rounding or grouping error); otherwise the doughnut trimming may end up removing too many observations and preclude informative inference.

For the second condition, we give a formal result showing it holds under *weaker* conditions than those needed for inference on the ATE conditional on $X^* = c$, if we were to observe X^*: intuitively, the measurement error smooths out kinks or other irregularities in the conditional mean of the outcome given X^*. Inference can be conducted using bias-aware inference methods (e.g. Armstrong & Kolesár, 2018, 2020; Noack & Rothe, 2021), which automatically adapt to the potentially irregular support of X. In particular, we can ignore the measurement error in the sense that estimation and inference on the trimmed data can proceed as if X were the running variable used by the administrator to assign treatment.

An appealing feature of focusing on the ATE for units with $X = c$ is that valid inference relies only on assumptions about the smoothness of the conditional mean of the potential outcomes given X. Such assumptions are easy to interpret, and are (partially) testable. Furthermore, inference is standard in that we can directly apply existing methods. In contrast, for inference on the average treatment effect for units with $X^* = c$, one needs to either make specific assumptions about latent objects, such as the distribution of the measurement error given X or X^*, or have access to auxiliary data; furthermore, the form of the estimator depends on the exact form of these specific assumptions.

The rest of the paper proceeds as follows. Section 2 gives the setup and the main results, and discusses their applicability under different types of measurement error. Section 3 illustrates the results in an empirical application. Section 4 concludes. Auxiliary results appear in the appendix.

2 Setup and main results

We are interested in the effect of a treatment T on an outcome Y. Let $Y(t)$ denote the potential outcomes, $t \in \{0, 1\}$. The observed outcome is given by $Y = Y(0)+T(Y(1)-Y(0))$. An administrator assigns individuals to treatment if their running variable X^* crosses a threshold c, which we normalize to $c = 0$. Let $Z = \mathrm{I}\{X^* \geq 0\}$ denote the indicator for treatment assignment. In a sharp RD design, all individuals comply with the treatment assignment, so that $T = Z$. Let $g_t^*(x) := E[Y(t) \mid X^* = x]$ denote the conditional mean of the potential outcomes given X^*, and let $g^*(x) := E[Y \mid X^* = x] = \mathrm{I}\{x \geq 0\}g_1^*(x) + \mathrm{I}\{x < 0\}g_0^*(x)$ denote the conditional mean of the observed outcome.

If the conditional means g_t^* are continuous at 0, the jump in g^* at 0 identifies an average treatment effect for units at the threshold (Hahn et al., 2001),

$$\tau^* := E[Y(1) - Y(0) \mid X^* = 0] = \lim_{x \downarrow 0} g^*(x) - \lim_{x \uparrow 0} g^*(x). \tag{1}$$

If X^* were observed, and we strengthen the continuity assumption by placing nonparametric smoothness assumptions on g^*, several methods for estimation and inference on τ^* are available (see, e.g., Armstrong & Kolesár, 2018, 2020; Calonico et al., 2014; Imbens & Kalyanaraman, 2012; Imbens & Wager, 2019). Under parametric restrictions on g^*, we can leverage standard parametric regression methods for estimation and inference.

We do not observe X^* directly, however; instead we observe $X = X^* - e$, where e is measurement error.[4] As our leading example, we focus on the case where e represents rounding or grouping error. By *rounding*, we mean that the observed running variable can be written as $X = r(X^*)$, where the rounding function r is a monotone step function that is idempotent (rounding a number twice is the same as rounding it once, $r(r(X^*)) = r(X^*)$), and the steps are equal-sized. For instance, if we observe a rounded-down version of X^*, then the rounding function corresponds to the floor function, $r(X^*) = \lfloor X^* \rfloor$, and $e = X^* - \lfloor X^* \rfloor$. Rounding is a special case of *grouping*. Grouping allows $r(\cdot)$ to be any idempotent step function, not necessarily with equal-sized steps; it arises when the sample space for X^* is partitioned into subsets, and $X = r(X^*)$ is a numeric value representing the subset (Heitjan & Rubin, 1991). For instance, under interval reporting, X may correspond to the midpoint or one of the endpoints of the interval that X^* belongs to. In spatial RD, it is common to observe the centroid of the unit's ZIP code or county instead of its exact location.

Our framework, outlined in Section 2.1 for sharp RD, and extended to fuzzy RD settings in Section 2.3, only imposes high-level conditions on e that allow the measurement error to take many other forms besides grouping; X may be discrete, continuous, or mixed. Section 2.2 discusses sufficient low-level conditions in particular settings. In addition to grouping, we consider classical measurement error (e is independent of $Y(1), Y(0)$ and X^*), Berkson measurement error (e is independent of $Y(1), Y(0)$ and X), as well as *heaping*. While under grouping, the grouping function $r(X)$ is the same for each unit (so that the conditional distribution of X given X^* is degenerate), under heaping one may round the data according one of several rounding functions $r_u(\cdot)$, $u = 1, 2, \ldots, K$, with different ranges; we observe $X = r_U(X^*)$, where U is a random variable, possibly correlated with potential outcomes or X^*, determining which rounding function is used (Heitjan & Rubin, 1991). For instance,

[4] Our setting is distinct from that in Eckles et al. (2022), where the running variable X observed by the researcher is also the variable used by the administrator, and it can be thought of as a noisy measure of some latent variable X^* that affects potential outcomes.

some parents may report ages of their children rounded to the nearest year or nearest half-year, while others may not round at all and report the exact date of birth (e.g. Heitjan & Rubin, 1990). The degree of rounding may depend on characteristics of the parents or the age of the child (ages of very young children and children of parents with more education are usually more likely to be precisely reported). The density of the running variable then displays "heaps" at ages that are multiples of 0.5.

2.1 Estimation and inference with measurement error

Our approach is based on the observation that any variable X can serve as a running variable, provided that it correctly classifies the treatment assignment, and provided that the conditional mean functions of the potential outcomes given X are continuous at the cutoff:

(*C1*) $\mathrm{I}\{X \geq 0\} = Z$ almost surely.

(*C2*) $g_t(x) := E[Y(t) \mid X = x]$ is continuous at 0 for $t = 0, 1$.

We give a detailed discussion of these conditions in our setting, where X is a mismeasured version of X^*, in Section 2.2 below. Condition (*C2*) corresponds to the standard RD continuity condition from Hahn et al. (2001), except it is applied to X rather than X^*.

To interpret the estimand we consider, let $g^*_t(X^*, e) := E[Y(t) \mid X^*, e]$ denote the conditional mean of the potential outcome $Y(t)$, $t = 0, 1$, given *both* the true value of the running variable and the measurement error, and let $\tau^*(X^*, e) = g^*_1(X^*, e) - g^*_0(X^*, e) = E[Y(1) - Y(0) \mid X^*, e]$ denote the conditional ATE, conditional on both X^* and e. We denote the conditional ATE, conditional on X^* only, by $\tau^*(x) = g^*_1(x) - g^*_0(x) = E[Y(1) - Y(0) \mid X^* = x]$, so that $\tau^* = \tau^*(0)$. While, as we discuss further below, our approach does allow the measurement error to affect potential outcomes, to clearly link the estimand we consider to the usual RD estimand, it is useful to rule this possibility out and assume that the measurement error doesn't affect the conditional ATE, at least when $X = 0$ (note $E[Y(1) - Y(0) \mid X^* = x, X = 0] = \tau^*(x, x)$):

(*C3*) $\tau^*(x, x) = \tau^*(x)$.

This condition is slightly weaker than the requirement that the measurement error be non-differential, i.e., independent of $(Y(1), Y(0))$ given X^* (Carroll et al., 2006, Chapter 2.6); for instance, rounding, grouping, and classical measurement errors are all non-differential.

Lemma 2.1. *Suppose that $T = Z$ and conditions* (C1) *and* (C2) *hold. Then the jump in the conditional mean function $g(x) := E[Y \mid X = x]$ at 0 identifies the ATE for units with*

$X = 0$,

$$\tau := E[Y(1) - Y(0) \mid X = 0] = \lim_{x \downarrow 0} g(x) - \lim_{x \uparrow 0} g(x). \tag{2}$$

If, in addition, condition (C3) holds, then $\tau = \int \tau^*(e) dF_{e|X}(e \mid 0)$, *where* $F_{e|X}(e \mid x)$ *is the conditional distribution of* $e = X^* - X$ *given* $X = x$.

Proof. Under condition *(C1)* $g(x) = \mathrm{I}\{X \geq 0\} g_1(x) + \mathrm{I}\{X < 0\} g_0(x)$, so that $\lim_{x \downarrow 0} g(x) - \lim_{x \uparrow 0} g(x) = \lim_{x \downarrow 0} g_1(x) - \lim_{x \uparrow 0} g_0(x)$, which equals $g_1(0) - g_0(0)$ by condition *(C2)*. By iterated expectations, $\tau = E\left[E\left[Y(1) - Y(0) \mid X^*, X = 0\right] \mid X = 0\right]$, which gives the second claim. □

By Lemma 2.1, we can effectively "ignore" the measurement error in the observed running variable X, in that we can conduct the analysis as if X were the running variable used by the administrator to assign treatment, provided that we align the target of inference accordingly, setting it to τ. Under condition *(C3)*, we can interpret τ as a weighted average of ATEs conditional on $X^* = x$, $\tau^*(x)$; in contrast τ^* simply corresponds to the conditional ATE at 0, $\tau^*(0)$. If condition *(C3)* doesn't hold, then both estimands may be expressed as weighted averages of conditional ATEs, conditional on both X^* and e, but the weighting functions are different, which makes them a little harder to compare. Specifically, by iterated expectations, $\tau = \int \tau^*(e, e) dF_{e|X}(e \mid 0)$, while $\tau^* = \int \tau^*(0, e) dF_{e|X^*}(e \mid 0)$, where $F_{e|X^*}$ is the conditional distribution of e given X^*.

Remark 1 (Comparison of τ and τ^*)**.** The parameter τ^* corresponds to an ATE for units with the latent running variable X^* equal to 0. If X^* is the birthdate of an individual, for instance, then τ^* is the ATE for those born on the cutoff date. If X is month of birth, then τ corresponds to the ATE for those born in the same month as the individuals born on the cutoff date. Since the measurement error is non-differential, τ can be expressed as a weighted average of $\tau^*(x)$, the ATE for individuals born on day x, and the same month the individuals born on the cutoff date. The exact weights depend on the distribution of births in that month. If, for instance, birthdate is uniformly distributed within the month, then the weights are uniform. The estimands τ and τ^* are generally different unless $\tau^*(x)$ is constant on the support of e. For example, if $\tau^*(x) = a + bx$, and $E[e \mid X = 0] = 1/2$ (say when e is uniform), which appears to be consistent with the results of our empirical application, then $\tau^* = a$, while $\tau = a + b/2$. Which parameter is more policy relevant depends on the particular policy counterfactual one has in mind.[5]

[5] One may object that the parameter τ is reverse-engineered in that Lemma 2.1 shows τ is the parameter that our analysis happens to identify. The same criticism may be leveled at the result of Hahn et al. (2001) that RD analysis in the absence of measurement error identifies τ^*. We view both results as useful in separating the internal and external validity of the analysis.

For estimation and inference, we need to strengthen condition (C2) by assuming that g satisfies appropriate parametric or nonparametric smoothness conditions. As a simple parametric approach, one may assume that $g(x)$ takes the form of a polynomial of degree q on either side of the threshold for values of x within distance h of the threshold. Then one could estimate τ by a local polynomial regression of Y onto $m_q(X) = (\mathrm{I}\{X \geq 0\}, \mathrm{I}\{X \geq 0\}X, \dots, \mathrm{I}\{X \geq 0\}X^q, 1, \dots, X^q)$ (a polynomial in X interacted with treatment assignment), using ordinary least squares (OLS).[6] Specifically, given a sample $\{Y_i, X_i\}_{i=1}^n$, the estimator is defined as

$$\hat{\tau}^Y_{h,q} = (1, 0, \dots, 0)' \left(\sum_{i=1}^{n} \mathrm{I}\{|X_i| \leq h\} m_q(X_i) m_q(X_i)' \right)^{-1} \sum_{i=1}^{n} \mathrm{I}\{|X_i| \leq h\} m_q(X_i) Y_i. \quad (3)$$

Under i.i.d. sampling, inference can be conducted using Eicker-Huber-White (EHW) standard errors, provided that X has at least $q+1$ support points on either side of the threshold.

A limitation of the parametric approach is that if $g(x)$ is not exactly polynomial inside the estimation window, the estimator will be biased; consequently, confidence intervals (CIs) based on EHW standard errors will undercover τ. To address this issue, as the preferred inference method, we propose to use the bias-aware (or "honest") inference approach developed in Armstrong and Kolesár (2018, 2020) and Kolesár and Rothe (2018). This approach enlarges the CIs by taking into account the potential finite sample bias of the estimator. In particular, letting $\hat{\sigma}(\hat{\tau}^Y_{h,q})$ denote the standard error, a CI with level $1-\alpha$ takes the form

$$\hat{\tau}^Y_{h,q} \pm \mathrm{cv}_\alpha \left(B(\hat{\tau}^Y_{h,q}) / \hat{\sigma}(\hat{\tau}^Y_{h,q}) \right) \cdot \hat{\sigma}(\hat{\tau}^Y_{h,q}). \quad (4)$$

Here $\mathrm{cv}_\alpha(t)$ is the $1-\alpha$ quantile of a folded normal distribution $|N(t,1)|$, and $B(\hat{\tau}^Y_{h,q})$ is a bound on the finite-sample conditional (on X) bias of the estimator. As a baseline assumption to bound the potential bias, we replace the parametric assumption that g is polynomial with the weaker non-parametric assumption that $g \in \mathcal{F}_{RD}(M)$, where

$$\mathcal{F}_{RD}(M) = \{ f_1(x)\, \mathrm{I}\{x \geq 0\} + f_0(x)\, \mathrm{I}\{x < 0\} \colon \|f_0'\|_{C^1} \leq M, \|f_1'\|_{C^1} \leq M \}. \quad (5)$$

Here $\|f\|_{C^1} = \sup_{x \neq x'} |f(x) - f(x')| / |x - x'|$ is the Lipschitz constant of f (if f is differentiable, then the constant is the maximum of its derivative, $\|f\|_{C^1} = \sup_x |f'(x)|$). The parameter space for g thus corresponds to (the closure of) a family of functions that are twice differentiable on either side of the cutoff, with the second derivative bounded in abso-

[6] While this covers a global approach by setting $h = \infty$, as discussed in Gelman and Imbens (2019), such an approach may perform poorly relative to local approaches.

lute value by M, but are potentially discontinuous at 0. Under this assumption, it is optimal to run a local linear regression, i.e. use the estimator $\hat{\tau}^Y_{h,1}$. The (conditional on X) bias of the estimator is maximized at the function $h(x) = Mx^2(\mathrm{I}\{x < 0\} - \mathrm{I}\{x \geq 0\})/2$, so that $B(\hat{\tau}^Y_{h,1})$ is given by eq. (3), with Y_i replaced by $h(X_i)$. See Armstrong and Kolesár (2020) and Kolesár and Rothe (2018) for details. An appealing feature of the bias-aware CI is that because it accounts for the exact finite-sample bias of the estimator, it is valid under any bandwidth sequence, including using a fixed bandwidth; for example, the bandwidth h may be selected to minimize the (worst-case over $\mathcal{F}_{RD}(M)$) mean squared error, or the length of the resulting confidence interval.[7]

2.2 Conditions for validity of proposed approach

Unlike existing approaches that seek to do inference on τ^* even in presence of measurement error (e.g. Davezies and Le Barbanchon, 2017; Dong, 2015; Hullegie and Klein, 2010; or Pei and Shen, 2017), we do not impose specific assumptions on the measurement error distribution or require auxiliary data. Instead, our approach is based on the observation that we can use existing parametric or nonparametric methods for inference on τ provided that condition *(C1)* holds, and we strengthen condition *(C2)* by assuming that the conditional mean functions g is smooth (in the sense that it is exactly polynomial inside the estimation window, or else $g \in \mathcal{F}_{RD}(M)$). These assumptions are high-level in that they are exactly the conditions needed to interpret the RD design with the observed running variable as a valid RD design. In the following remarks, we discuss in detail sufficient conditions for these assumptions in specific measurement errors settings. We also discuss related practical issues.

Remark 2 (Correct classification of treatment assignment)**.** In a few special cases, such as when X corresponds to X^* rounded down to the nearest integer, and the threshold c is an integer, condition *(C1)* holds automatically.[8] In general, however, measurement error in X may induce misclassification of the treatment assignment for values of X equal to the cutoff or in its immediate vicinity. In such cases, condition *(C1)* requires dropping observations with such values of X, resulting in a "doughnut" design (e.g. Almond et al., 2011; Barreca et al., 2011).

[7] We implement this method in the empirical application in Section 3, where we also discuss the choice of M, the key tuning parameter. See Kolesár and Rothe (2018), Imbens and Wager (2019), and Armstrong and Kolesár (2020) for a more detailed discussion, including a discussion of implementation issues.

[8] Another example when condition *(C1)* holds in the full sample is when only values of X outside the immediate vicinity of the cutoff are error-ridden. Specifically, let $X^* - U$ be an initial noisy measurement, where U is pure measurement error with bounded support. We observe a follow-up exact measurement, $X = X^*$ with probability $p(X^* - U)$, and observe the noisy measurement otherwise. The probability $p(\cdot)$ varies smoothly and equals 1 near the cutoff, when it is necessary to wait for the follow-up measurement to determine treatment assignment.

The exact form of such doughnut trimming depends on the support of the measurement error. Under grouping error, we need to remove observations corresponding to the subset containing 0. For instance, under ordinary rounding, or rounding up to the nearest integer, we need to remove observations with $X = 0$. Under interval measurement, we need to remove X that falls into the interval containing 0. Under Berkson, classical, or other types of measurement error with bounded support $[s_0, s_1]$, with $s_0 \leq 0 \leq s_1$, we need to remove observations with $X \in [-s_1, -s_0)$.

A limitation of our approach is that it does not handle settings in which the support e is unknown or unbounded. Furthermore, if the support of e is wide, the doughnut trimming may result in removing many observations, and preclude informative inference. One way to proceed in such cases is focus on inference about τ^* under parametric assumptions about the measurement error distribution and the form of g^*, as in Hullegie and Klein (2010).[9] If the researcher observes the treatment assignment Z, in addition to Y and X, we can easily infer which units are misclassified. Pei and Shen (2017) discuss how to use this information to recover τ^* without parametric restrictions on g^* or on the measurement error distribution.[10]

Remark 3 (Smoothness of g)**.** Our proposed approach requires smoothness of g in the sense that $g \in \mathcal{F}_{RD}(M)$. To discuss this condition, suppose that condition *(C1)* holds, so that we may write $g(x) = \mathrm{I}\{x \geq 0\}g_1(x) + \mathrm{I}\{x < 0\}g_0(x)$, where, by iterated expectations,

$$g_t(x) = E[Y(t) \mid X = x] = E[g_t^*(X^*, e) \mid X = x]. \tag{6}$$

In Lemma A.1 in Appendix A.1, we give a formal result showing that if (a) the conditional distribution of the measurement error $F_{e|X}(e \mid x)$ is smooth in x; and (b) the effect of the measurement error on potential outcomes is smooth, so that $g_t^*(X^*, e)$ is smooth in the second argument, then the condition $g \in \mathcal{F}_{RD}(M)$ is *weaker* than the analogous smoothness requirement needed for inference on τ^* if X^* were observed: the smoothness of g_t in X is *greater* than the smoothness of $g_t(X^*, e)$ in X^*. Consequently, eq. (5) will hold for g even in settings where it may not hold for $g^*(x) = \mathrm{I}\{x \geq 0\}g_1^*(x) + \mathrm{I}\{x < 0\}g_0^*(x)$. Condition (b) holds trivially for Berkson measurement error, since then $F_{e|X}(e \mid x)$ doesn't depend on x at all; it also holds for classical measurement error if the density of X^* is smooth.[11]

[9] Inference on τ^* without parametric restrictions on g^* is challenging, because τ^* is generally unidentified unless the measurement error distribution is completely known. Furthermore, even if the distribution is known, the rates of convergence can be very slow. For example, if e is classical Gaussian measurement error, the lower bounds in Fan and Truong (1993) suggest that the rate is logarithmic in the sample size.

[10] In the context of fuzzy RD, Davezies and Le Barbanchon (2017) develop a nonparametric approach to estimation of τ^* that is likewise flexible about the form and support of the measurement error distribution, but they require the econometrician to observe both X and X^* in the subsample of treated individuals.

[11] Specifically, by Bayes' rule, the conditional density of e given $X = x$ may be written as $f(e; x) = f_{X^*}(x+$

To gain intuition for this result, suppose first that the measurement error is non-differential, so that $g_t^*(X^*, e) = g_t^*(X^*)$, and condition (b) above holds trivially. Then we may write eq. (6) as $g_t(x) = E[g_t^*(X^*) \mid X = x]$. Lemma A.1 then formalizes the notion that the conditional expectation $E[\cdot \mid X = x]$ "smooths out" non-linearities in g_t^*.[12] For example, if g_t^* contains kinks (so that it has smoothness index 1 and eq. (5) fails for g^*), these kinks will be smoothed out by the measurement error provided that the conditional density of e given X is continuous with a bounded slope (in which case g_t will have smoothness index 2, and eq. (5) will hold for g). This smoothing effect is greatest under grouping error, or more generally whenever X is discrete: under the interpretation in Remark 4 below, eq. (5) always holds for appropriately chosen M (so Lemma A.1 is not needed), since we can always smoothly interpolate $g_t(x)$ through the support points of X, using, say, spline interpolation (e.g. Späth, 1995).

The result in Lemma A.1 also goes through under differential measurement error, provided the error affects the potential outcomes smoothly—this is analogous to the result that $g_t^*(X^*)$ remains to be smooth in X^* even if agents can manipulate their running variable, so long as the manipulation is not perfect (Lee, 2008).

For the parametric approach, if condition *(C1)* holds, a sufficient condition for g to be polynomial of degree q on either side of the threshold is that $g_0^*(X^*, e)$ and $g_1^*(X^*, e)$ are multivariate polynomials of degree q, and $E[e^j \mid X]$ for $j = 1, \ldots, q$ are polynomials of degree j. This follows directly from the binomial theorem.

While conditions (a) and (b) in Remark 3 are relatively mild, they necessitate subpopulation analysis under heaping. For example, consider using birthweight as a running variable to identify the effects of hospital care on infant health as in Almond et al. (2010). Suppose that some (but not all) hospitals report rounded rather than precise birthweight—then condition (a) fails since the distribution $F_{e|X}(e \mid x)$ changes discontinuously as at round values of x. Furthermore, suppose that hospitals with fewer resources are more likely report rounded rather than precise birthweight, as argued in Barreca et al. (2016). Then condition (b) also fails, since $g_t^*(X^*, e)$ is then potentially discontinuous at $e = 0$ due to different hospital composition under rounded vs exact reporting. Suppose, however, that the conditional mean functions g_t are smooth in the subpopulation of hospitals that don't round birthweight. Con-

$e)f_e(e)/\int f_{X^*}(x+e)f_e(e)de$, where f_{X^*} and f_e denote the density of X^* and e, respectively. This expression is smooth in x so long as the density of X^* is smooth and the marginal density of X, $\int f_{X^*}(x+e)f_e(e)de$, is bounded away from zero.

[12] See, for example, Newey (2013, Section 4) for a discussion in the context of nonparametric instrumental variables regression, where X^* plays the role of an endogenous variable, and X plays the role of an instrument. One consequence of this smoothing in the current context is that the measurement error may smooth out the discontinuity of g^* at the cutoff, making g continuous, and causing condition *(C1)* to fail unless we restrict the sample as discussed in Remark 2.

ditions (a) and (b) then hold if we drop the heaping points, only keeping observations with non-round values of X, as suggested by Barreca et al. (2016). The estimand τ corresponds to a conditional ATE for units with $X = X^* = 0$ born in a hospital that reports exact birthweight.

In some instances of heaping, all values of X are rounded, but the rounding precision differs across units. Say some individuals round age to the nearest month, while others round to the nearest year or half-year. To ensure conditions (a) and (b) hold in such cases, we drop individuals with age in months that is a multiple of 6. The remaining sample then only contains those who report age in months. We can interpret the estimand as the ATE for those born in the cutoff month in the subpopulation of individuals who report the running variable with the greatest precision.

Since the conditional mean function $g(x)$ is identified over the support of X, smoothness assumptions such as $g \in \mathcal{F}_{RD}(M)$ are testable. Problems such as heaping are often apparent from simple plots of undersmoothed binned averages of the outcome against X (see, e.g., Figure 1 in Barreca et al., 2011), and one can also conduct more formal specification tests (see, e.g., Kolesár & Rothe, 2018, Appendix S.3).

Remark 4 (Irregular support of X)**.** The measurement error may result in a coarsening of the support of the observed running variable X relative to X^*. Under grouping, for example, X becomes discrete even if X^* is continuously distributed. Furthermore, there may be a gap in the support around 0 due to doughnut trimming (see Remark 2). In such "irregular" cases, since conditional mean functions are only well-defined over the support of the conditioning variable, following Kolesár and Rothe (2018) and Imbens and Wager (2019), we interpret smoothness assumptions such as eq. (5) to mean that there exists a function $g(x) \in \mathcal{F}_{RD}(M)$ with domain $\mathbb{R}$ such that $E[Y \mid X] = g(X)$ with probability one. With discrete X or under a doughnut design, there will be multiple functions g satisfying this condition, and the parameter τ will only be partially identified.

An advantage of bias-aware inference is that the estimator and CI construction remains the same whether the support of X is continuous, discrete, or otherwise irregular, and whether τ is point or partially identified. Under irregular support of X, the finite-sample bias of the estimator may be large, but the CI will automatically reflect it via a larger critical value (in such cases, the interval will converge to the identified set as the sample size $n \to \infty$). We illustrate these points in the empirical application in Section 3, where we show that under rounding error, confidence intervals for τ tend to be longer than confidence intervals for τ^* that one would obtain using the same construction if X^* were observed.[13]

[13]In large samples, CIs for τ will be wider than the corresponding CIs for τ^* if X^* were observed, because the former don't converge to a point, while the latter do. In finite samples, the variability of the estimators,

2.3 Fuzzy designs

In fuzzy RD designs, only a subset of the individuals complies with the treatment assignment, so that $T \neq Z$. In this case, Hahn et al. (2001) show that the fuzzy RD parameter can be interpreted as a local average treatment effect for individuals who comply with the treatment assignment. Let us reconsider their argument when we use a variable X as the running variable, not necessarily equal to the running variable X^* used by the administrator.[14]

Let $T(1)$ denote the potential treatment status of the individual if they are assigned to treatment, and let $T(0)$ denote their status if they are not assigned to treatment. The observed treatment is given by $T = T(Z)$, and the observed outcome is given by $Y = Y(T(Z)) = Y(0) + T(Z)(Y(1) - Y(0))$. Let $\mathfrak{C}$ denote the event that an individual is a complier, that is $T(1) > T(0)$. Finally, in analogy to the conditional means g and g_t, let $p(x) = E[T \mid X = x]$ and $p_z(x) = E[T(z) \mid X = x]$ for $z \in \{0, 1\}$.

We replace the sharp RD condition that all individuals comply with the treatment assignment ($T = Z$) with the weaker condition that a non-zero fraction of individuals complies with it, and that nobody defies the treatment assignment (in analogy with the monotonicity condition in Imbens and Angrist, 1994):

(*F1*) $P(T(1) \geq T(0) \mid X = 0) = 1$, and $P(T(1) > T(0) \mid X = 0) > 0$.

Next, we replace the continuity assumption (*C2*) with a continuity assumption on the first stage and reduced form regression functions:[15]

(*F2*) $p_z(x)$ and $E\left[Y\left(T\left(z\right)\right) \mid X = x\right]$, $z = 0, 1$, are continuous at 0.

Intuitively, if treatment eligibility Z did not change at the cutoff but was instead fixed, this condition implies that the observed outcome $Y = Y(T(Z))$ would be continuous at 0. As a result, any discontinuity must be due to change in treatment eligibility, which allows for identification of causal effects. Conditions (*F1*) and (*F2*) are analogous to the standard fuzzy RD assumptions, but applied to X rather than X^*.

Finally, to link the estimand we consider to the usual RD estimand, analogous to condition (*C3*), it is useful to assume that the measurement error has no effect on the compliance probability or the ATE for compliers once we control for X^*.

which in general cannot be ranked, also matters, and the CI for τ^* may end up being wider.

[14] The original argument in Hahn et al. (2001) involved defining potential treatments under counterfactual values of the running variable. However, the running variable may not be manipulable (e.g. when X^* corresponds to a birthdate). We therefore use a slightly different argument, based on manipulation of the treatment assignment. The treatment assignment is typically manipulable, say by moving the cutoff.

[15] Analogous to an instrumental variables regression that uses Z as an instrument, these are (non-parametric) regressions of T and Y, respectively, onto Z and X.

(*F3*) $P(\mathfrak{C} \mid X^* = x, X = 0) = P(\mathfrak{C} \mid X^* = x)$ and $E[Y(1) - Y(0) \mid \mathfrak{C}, X^* = x, X = 0] = E[Y(1) - Y(0) \mid \mathfrak{C}, X^* = x]$.

This is a slightly weaker requirement that the measurement error e be non-differential, i.e. independent of $(Y(1), Y(0), T(1), T(0))$ given X^*. In analogy to condition (*C3*) in the sharp case, condition (*F3*) is helpful for interpreting the estimand, but it is not necessary for validity of our approach.

With this setup, we obtain a fuzzy RD analog of Lemma 2.1.

Lemma 2.2. *Suppose that conditions* (C1)*,* (F1)*, and* (F2) *hold. Then*

$$\tau_F := E[Y(1) - Y(0) \mid \mathfrak{C}, X = 0] = \frac{\lim_{x\downarrow 0} g(x) - \lim_{x\uparrow 0} g(x)}{\lim_{x\downarrow 0} p(x) - \lim_{x\uparrow 0} p(x)}.$$

If, in addition, condition (F3) *holds, then* $\tau_F = \int \tau_F^*(e)\omega(e)dF_{e|X}(e \mid 0)$*, where* $\tau_F^*(x) := E[Y(1) - Y(0) \mid \mathfrak{C}, X^* = x]$*, and* $\omega(e) = \frac{P(\mathfrak{C}|X^*=e)}{\int P(\mathfrak{C}|X^*=e)dF_{e|X}(e|0)}$.

Proof. Observe that

$$\begin{aligned}
\lim_{x\downarrow 0} g(x) - \lim_{x\uparrow 0} g(x) &= \lim_{x\downarrow 0} E[Y(T(1)) \mid X = x] - \lim_{x\uparrow 0} E[Y(T(0)) \mid X = x] \\
&= E[Y(T(1)) - Y(T(0)) \mid X = 0] \\
&= E[(Y(1) - Y(0))(T(1) - T(0)) \mid X = 0] \\
&= E[Y(1) - Y(0) \mid X = 0, \mathfrak{C}]P(\mathfrak{C} \mid X = 0),
\end{aligned}$$

where the first equality uses the fact that $Y = Y(T(Z))$, and that by condition (*C1*), $T = T(1)$ for individuals with $X \geq 0$, and $T = T(0)$ for those with $X < 0$, the second equality uses condition (*F2*), the third uses $Y(T(z)) = Y(0) + T(z)(Y(1) - Y(0))$, and the last equality uses iterated expectations and condition (*F1*). By analogous arguments, $\lim_{x\downarrow 0} p(x) - \lim_{x\uparrow 0} p(x) = P(\mathfrak{C} \mid X = 0)$. The second claim follows by applying iterated expectations to the numerator and denominator of $\tau_F = \frac{E[(Y(1)-Y(0))\mathfrak{C}|X=0]}{E[\mathfrak{C}|X=0]}$, and using condition (*F3*). $\square$

Under perfect compliance, $T = Z$, Lemma 2.2 reduces to Lemma 2.1. In analogy to the sharp case, any variable satisfying conditions (*C1*), (*F1*), and (*F2*) can be used as a running variable. With $X = X^*$, we obtain the standard result that

$$\tau_F^* := \tau_F^*(0) = \frac{\lim_{x\downarrow 0} g^*(x) - \lim_{x\uparrow 0} g^*(x)}{\lim_{x\downarrow 0} p^*(x) - \lim_{x\uparrow 0} p^*(x)},$$

where $p^*(x) = E[T \mid X^* = x]$.

Unless the local average treatment effects $\tau_F^*(x)$ are constant on the support of e, $\tau_F \neq \tau_F^*$. Since $\tau_F^*(x)$ is given by the ratio of the reduced form effect $E[Y(T(1)) - Y(T(0)) \mid X^* = x]$ to the first stage effect $E[T(1) - T(0) \mid X^* = x]$, whether $\tau_F^*(x)$ is locally constant depends on heterogeneity in both the reduced form and the first stage conditional mean functions. Our empirical results in Section 3, for example, are consistent with the reduced form effect being approximately constant, while the first stage effect is approximately linear, i.e., $E[Y(T(1)) - Y(T(0)) \mid X^* = x] \approx a$ and $E[T(1) - T(0) \mid X^* = x] \approx b + cx$; further, the measurement error is approximately uniform on $[0, 1]$. So $\tau_F \approx a/(b + c/2)$, while $\tau_F^* \approx a/b$.

If the measurement error is differential, and condition *(F3)* doesn't hold, then both τ_F^* and τ_F may be expressed as weighted averages of conditional ATEs for compliers, conditional on both X^* and e, $\tau_F^*(X^*, e) := E[Y(1) - Y(0) \mid \mathfrak{C}, X^*, e]$. Specifically, by iterated expectations, $\tau_F = \frac{\int \tau_F^*(e,e) p_{\mathfrak{C}}(e,e) dF_{e|X}(e|0)}{\int p_{\mathfrak{C}}(e,e) dF_{e|X}(e|0)}$, while $\tau_F^* = \frac{\int \tau_F^*(0,e) p_{\mathfrak{C}}(0,e) dF_{e|X^*}(e|0)}{\int p_{\mathfrak{C}}(0,e) dF_{e|X^*}(e|0)}$, where $p_{\mathfrak{C}}(X^*, e) := P(\mathfrak{C} \mid X^*, e)$. Lemma 2.2 is related to the result in Battistin et al. (2009) who show that if we replace condition *(C1)* with the assumption that the measurement error is non-differential, and has a point mass at zero but is otherwise smooth, $\tau_F = \tau_F^*$. If the measurement error is differential, but affects the potential outcomes and potential treatments smoothly, the arguments in Battistin et al. (2009) imply that $\tau_F = \tau_F^*(0, 0)$; this was shown in Card et al. (2015) in the context of fuzzy regression kink designs.

Similarly to the sharp case, if we assume that the conditional mean functions $g(x)$ and $p(x)$ are polynomial inside a window h of the threshold, then we can estimate τ_F as a ratio of local polynomial estimators

$$\hat{\tau}_{h,q} = \hat{\tau}^Y_{h,q} / \hat{\tau}^T_{h,q}, \tag{7}$$

with $\hat{\tau}^Y_{h,q}$ defined in eq. (3), and $\hat{\tau}^T_{h,q}$ defined analogously.[16] If there are at least $q + 1$ support points for X on either side of the threshold and inside the estimation window, then under i.i.d. sampling, standard errors for $\hat{\tau}_{h,q}$ can be constructed based on the EHW covariance matrix for $(\hat{\tau}^Y_{h,q}, \hat{\tau}^T_{h,q})$ using the delta method.

Our preferred approach weakens the polynomial assumptions on $g(x)$ and $p(x)$ by instead assuming that $g \in \mathcal{F}_{RD}(M_y)$, and $p \in \mathcal{F}_{RD}(M_t)$. While this assumption only delivers set identification if the support of X is irregular (see Remark 4), we can use the bias-aware inference approach for constructing CIs that are asymptotically valid whether τ_F is point identified, set identified, or unidentified.[17] In particular, following Noack and Rothe (2021),

[16] Equivalently, as noted in Hahn et al. (2001), the estimator can be computed as a two-stage least squares estimator in a regression of Y onto T using $\mathrm{I}\{X \geq 0\}$ and instrument, and the remaining elements of $m_q(X)$ as exogenous covariates, using observations inside the estimation window.

[17] The parameter τ_F is unidentified if the instrument $\mathrm{I}\{X \geq 0\}$ is irrelevant in the sense that $P(T(1) > T(0) \mid X = 0) = 0$. Since the expression for τ_F in Lemma 2.2 not well-defined in this case, one can define τ_F in an arbitrary way.

we can test the hypothesis $H_0\colon \tau_F = \tau_{F,0}$ by checking whether 0 is in the bias-aware confidence interval based on $\hat{\tau}_{h,1}^{Y-\tau_{F,0}T}$, and noting that the smoothness assumptions on g and p imply $E[Y - \tau_{F,0}T \mid X = x] \in \mathcal{F}_{RD}(M_y + |\tau_{F,0}|M_t)$. The confidence set for τ_F is constructed by collecting all values of $\tau_{F,0}$ that are not rejected, similar to the construction of Anderson and Rubin (1949) confidence set in standard linear instrumental variables model.[18]

3 Empirical Application

In this section, we use data from Holbein and Hillygus (2016) to estimate the impact of preregistration on youth turnout in an election. Holbein and Hillygus (2016) leverage the fact that in Florida, individuals who were ineligible to vote in the 2008 election (those born after November 4, 1990) were nonetheless eligible to preregister to be added to the voter rolls for the next election. Those born before November 4, 1990 were already eligible to register regularly and vote in 2008. This motivates a fuzzy RD design, where the treatment T is an indicator for preregistering, the outcome Y is an indicator for voting in the 2012 election, and the running variable X^* is the proximity to the eligibility cutoff in days.

To illustrate the effects of measurement error in the running variable, we compare this design to a fuzzy RD design in which we (pretend to) only observe individuals' month of birth, and hence use proximity to November 1990 in months, X, as a running variable. We discard individuals born in November 1990, since their eligibility cannot be determined by month of birth alone (see Remark 2). We show that, consistent with the discussion in Remarks 1 and 4, (i) using proximity in days vs months yields different estimates, reflecting the impact of the rounding error on the estimand, and (ii) using month of birth generally leads to wider CIs.

We first visualize both versions of the RD design. In each case, the sample size is 186,575, consisting of individuals born within 6 months of the eligibility cutoff. Figure 1 presents the first stage, plotting preregistration rate against proximity in days (panel (a)) or in months (panel (b)). For ineligible individuals, the preregistration rate is essentially 0, while for eligible individuals, the preregistration rate is downward slopping: those born further away from the cutoff preregister with lower probability. There is a clear jump in the registration rate at the eligibility threshold in either panel. Figure 2 shows the reduced form, plotting the proportion who voted in the 2012 election against proximity to eligibility. In both panels, there is a small jump in the voting probability at the cutoff.

[18] We implement this method in our empirical application in Section 3, where we also discuss the choice of the smoothness constants M_t and M_y. See Noack and Rothe (2021) for a detailed discussion of implementation issues.

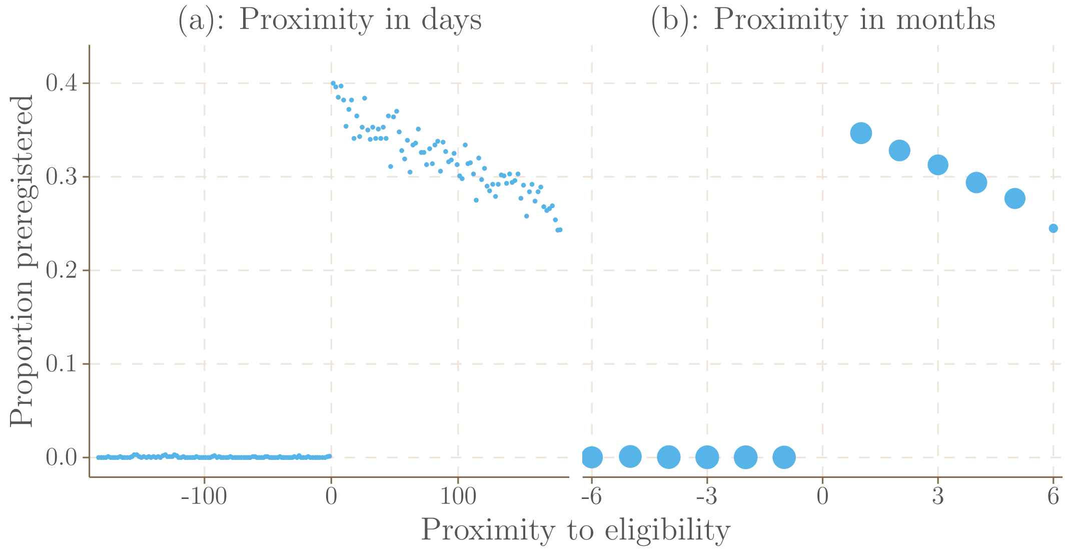

Figure 1: Effect of proximity on preregistering.

Notes: In panel (a), proximity is measured in days, and each point corresponds to an average of 1,000 individuals. In panel (b), proximity is measured in months, and each point corresponds to an average across all individuals born in a given month.

We use five specifications to compute the fuzzy RD estimator in eq. (7), the sharp RD estimators of the first stage and reduced form effects, and the associated confidence intervals. For ease of comparison across specifications, all specifications use a uniform kernel and local linear regression ($q = 1$). The first specification follows Holbein and Hillygus (2016), and uses bandwidth set to $h = 60$ days (or $h = 2$ months), and the confidence intervals to not account for the potential bias of the estimator. The second specification differs only in that it uses a slightly larger bandwidth, $h = 90$ days (or $h = 3$ months).[19] The third specification uses the robust bias correction (RBC) method of Calonico et al. (2014). For proximity in days, we use the default "MSE optimal" bandwidth provided by their software package; for proximity in months we use $h = 3$.[20]

The last two methods implement the bias-aware approach. We use confidence intervals given in eq. (4) for the first stage and reduced form effects; for inference on the fuzzy RD estimand, we use the Noack and Rothe (2021) construction. Implementing these methods requires a choice of smoothness bounds for the first stage (M_t) and the reduced form (M_y).

[19] These specifications can be interpreted as imposing a parametric linear functional form inside the estimation window. Alternatively, one can justify them by an "undersmoothing" argument: the specifications implicitly assume that the constants M_y and M_t are small enough so that the bias is negligible at these bandwidth choices.

[20] The formal arguments justifying the RBC method and the default bandwidth selector require the running variable to be continuous, which is not the case in either design. When proximity is measured in months, the discreteness causes implementation issues with the default "MSE optimal" bandwidth calculations.

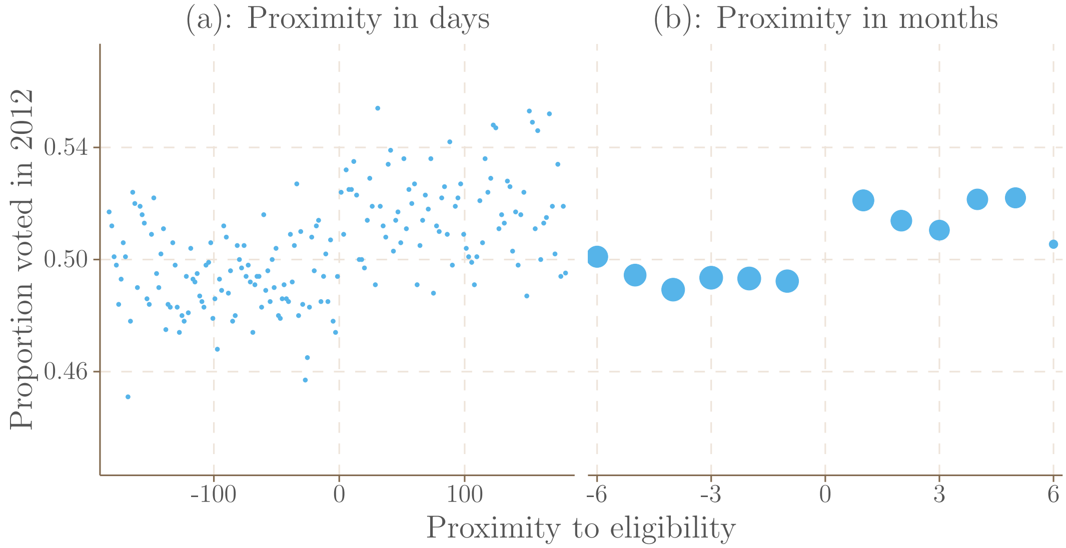

Figure 2: Effect of proximity on voting.

Notes: In panel (a), proximity is measured in days, and each point corresponds to an average of 1,000 individuals. In panel (b), proximity is measured in months, and each point corresponds to an average across all individuals born in a given month.

The results of Low (1997) and Armstrong and Kolesár (2018) imply that picking M_t and M_y in a data-driven way without violating coverage requires further non-convex restrictions on the parameter spaces $\mathcal{F}_{RD}(M_y)$ and $\mathcal{F}_{RD}(M_t)$ for g and p.[21] A natural way of doing this is to relate the global smoothness of g and p to the local smoothness constants M_y and M_t. We consider two ways of formalizing how the global and local smoothness relate. In particular, the fourth specification assumes, following the proposal in Armstrong and Kolesár (2020), that M_y is bounded by the smoothness of a global quartic approximation to g on either side of the cutoff, as measured by the largest (in absolute value) second derivative of the fitted line; we impose an analogous assumption on M_t and p. The fifth specification follows the suggestion in Imbens and Wager (2019) to use a global quadratic regression instead, and, additionally, multiply the largest second derivative of the fitted line by some moderate factor, taken here to be 2.

Since there are many other reasonable ways of formalizing the idea that the local and global smoothness are related, we view these methods as merely rules of thumb (ROTs) for selecting the smoothness constants. To assess these rules, we use the visualization approach

[21] The problem of choosing the smoothness constants is essentially a non-parametric model selection problem. Echoing the difficulties with conducting valid post-model selection inference in parametric contexts (e.g. Leeb & Pötscher, 2005), Armstrong and Kolesár (2018) show that bias-aware confidence intervals that assume the worst-case smoothness are in fact highly efficient at smooth functions. Thus, there is little scope for improvement by using data-driven choices of the smoothness constants.

proposed in Noack and Rothe (2021), described and implemented in Appendix A.2. These visualizations suggest that the Armstrong and Kolesár (2020) ROT is quite conservative, and allows for g and p to be quite non-smooth. The second ROT delivers more optimistic smoothness bounds that generate reasonably smooth conditional mean functions. To make the smoothness constants comparable across the specifications, we report the implied smoothness constants after rescaling the running variable to have support $[-1, 1]$ (which amounts to multiplying the original smoothness constants by $\max_i X_i^2$ and $\max_i (X_i^*)^2$, respectively). Given a choice of the smoothness constants, the bandwidth is selected so that the point estimate defined in eq. (3) minimizes the worst-case (over the chosen smoothness class) finite-sample MSE of the estimator.

3.1 Results

Table 1 presents the first stage estimates, the effect of the preregistration eligibility on preregistration. The estimates are stable across the specifications, in the range of 38–40% when using proximity in days; the estimates using proximity in months are slightly lower, in the range 36–37%, but still indicating a clear jump in the preregistration rate at the eligibility threshold. This is consistent with our theory, discussed in Remark 1, and reflects the difference between the parameters $E[T(1)-T(0) \mid X^* = 0]$ (the effect for those born on the cutoff date, November 4) and $E[T(1) - T(0) \mid X = 0]$ (the effect for those born in November, the cutoff month). In particular, the latter estimand averages over individuals born further away from the cutoff date, and Figure 1 suggests that the preregistration probability is decreasing with the distance to the cutoff. Since X is discrete, the parameter $E[T(1)-T(0) \mid X = 0]$ is not point identified. The confidence intervals for the bias-aware specifications, which account for this, are correspondingly wider than those in panel A, albeit they still remain quite tight.

Table 2 presents the reduced form estimates, the effect of preregistration eligibility on voting. The point estimates are about 3% for both designs and stable across specifications. In line with the discussion in Remark 4, the CIs are produced by the bias-aware specifications are wider when using proximity in months, reflecting the loss of point identification.

Table 3 presents the fuzzy RD estimates of the effect of preregistration on voting. When eligibility is measured in months, the smaller first stage estimates in panel B of Table 1 translate to larger estimates of the effect of preregistration on voting, around 10%, compared to 7–8% when eligibility is measured in days. When eligibility is measured in months, the fuzzy RD estimand, τ_F, is the ATE for compliers born in November 1990, and thus averages over individuals born further away from the cutoff than the estimand τ_F^* when eligibility is measured in days, which corresponds to the ATE for compliers born on November 4, 1990.

Table 1: First stage estimates: effect of eligibility on preregistration.

	OLS		RBC	Bias-aware inference	
	(1)	(2)	(3)	(4)	(5)
Panel A: Proximity in days					
Estimate	0.384	0.379	0.393	0.396	0.381
SE	0.006	0.005	0.008	0.009	0.005
95% CI	$(0.373, 0.396)$	$(0.370, 0.388)$	$(0.378, 0.409)$	$(0.377, 0.414)$	$(0.371, 0.391)$
Bandwidth	60	90	43	28	86
Eff. obs.	$63,220$	$94,118$	$43,538$	$28,274$	$89,776$
Rescaled M_t				1.015	0.061
Panel B: Proximity in months					
Estimate	0.365	0.363	0.368	0.365	0.365
SE	0.009	0.006	0.017	0.009	0.009
95% CI	$(0.348, 0.382)$	$(0.351, 0.375)$	$(0.335, 0.402)$	$(0.303, 0.428)$	$(0.345, 0.385)$
Bandwidth	2	3	3	2	2
Eff. obs.	$64,011$	$94,662$	$94,662$	$64,011$	$64,011$
Rescaled M_t				0.865	0.092

Notes: Column (1) uses local linear regression with bandwidth equal to 60 days (panel A) or 2 months (panel B), without any bias corrections. Column (2) is analogous, but uses bandwidth equal to 90 days or 3 months. Column (3) uses the RBC procedure, with the default "MSE optimal" bandwidth in panel A, and bandwidth equal to 3 months in panel B. Columns (4) and (5) report bias-aware confidence intervals, with bandwidth chosen to minimize the worst-case MSE. Column (4) uses the ROT of Armstrong and Kolesár (2020) to choose the smoothness constant M_t, while column (5) uses the ROT of Imbens and Wager (2019). Eff. obs refers to the number of observations inside the estimation window. The smoothness constants are reported after rescaling the running variable to have support $[-1, 1]$.

If the treatment effect for compliers born x days from the eligibility threshold, $\tau_F^*(x) = E[Y(1) - Y(0) \mid \mathfrak{C}, X^* = x]$, is increasing in x, then τ_F will be larger than τ_F^*, which is consistent with the results in Table 3. However, the bias-aware confidence intervals are fairly wide, and also consistent with $\tau_F^*(x)$ being constant.

4 Summary and conclusions

Measurement error is a common feature of RD applications. We show that its presence does not have deleterious effects on the validity of existing inference methods, provided that one employs doughnut trimming to ensure that the observed running variable X correctly

Table 2: Reduced form estimates: effect of eligibility on voting.

	OLS		RBC	Bias-aware inference	
	(1)	(2)	(3)	(4)	(5)
Panel A: Proximity in days					
Estimate	0.028	0.027	0.032	0.031	0.028
SE	0.008	0.007	0.012	0.012	0.007
95% CI	$(0.012, 0.044)$	$(0.014, 0.040)$	$(0.008, 0.056)$	$(0.005, 0.058)$	$(0.013, 0.044)$
Bandwidth	60	90	38	29	83
Eff. obs.	$63,220$	$94,118$	$39,195$	$29,285$	$86,881$
Rescaled M_y				1.401	0.099
Panel B: Proximity in months					
Estimate	0.037	0.034	0.042	0.037	0.034
SE	0.013	0.009	0.025	0.013	0.009
95% CI	$(0.012, 0.062)$	$(0.017, 0.051)$	$(-0.007, 0.090)$	$(-0.085, 0.159)$	$(0.009, 0.059)$
Bandwidth	2	3	3	2	3
Eff. obs.	$64,011$	$94,662$	$94,662$	$64,011$	$94,662$
Rescaled M_y				1.818	0.121

Notes: See Table 1.

classifies the treatment assignment. Care needs to be taken when interpreting the estimand: it corresponds to the ATE for units with the observed running variable X equal to the cutoff, rather than the usual parameter, the ATE for units with the latent running variable X^* equal to the cutoff. We illustrate this point in an empirical application.

References

Almond, D., Doyle, J. J., Kowalski, A. E., & Williams, H. (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *Quarterly Journal of Economics*, *125*(2), 591–634. https://doi.org/10.1162/qjec.2010.125.2.591

Almond, D., Doyle, J. J., Kowalski, A. E., & Williams, H. (2011). The role of hospital heterogeneity in measuring marginal returns to medical care: A reply to Barreca, Guldi, Lindo, and Waddell. *The Quarterly Journal of Economics*, *126*(4), 2125–2131. https://doi.org/10.1093/qje/qjr037

Table 3: Fuzzy RD estimates of the effect of preregistration on voting.

	OLS		RBC	Bias-aware inference	
	(1)	(2)	(3)	(4)	(5)
Panel A: Proximity in days					
Estimate	0.073	0.072	0.076	0.080	0.074
SE	0.021	0.018	0.032	0.031	0.018
95% CI	$(0.031, 0.114)$	$(0.037, 0.106)$	$(0.013, 0.139)$	$(0.012, 0.143)$	$(0.034, 0.122)$
Bandwidth	60	90	36	29	83
Eff. obs.	$63,220$	$94,118$	$35,785$	$29,285$	$86,881$
Rescaled M_y				1.401	0.099
Rescaled M_t				1.015	0.061
Panel B: Proximity in months					
Estimate	0.101	0.094	0.113	0.101	0.094
SE	0.035	0.024	0.068	0.035	0.024
95% CI	$(0.034, 0.169)$	$(0.047, 0.141)$	$(-0.020, 0.246)$	$(-0.268, 0.505)$	$(0.023, 0.180)$
Bandwidth	2	3	3	2	3
Eff. obs.	$64,011$	$94,662$	$94,662$	$64,011$	$94,662$
Rescaled M_y				1.818	0.121
Rescaled M_t				0.865	0.092

Notes: See Table 1.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63. https://doi.org/10.1214/aoms/1177730090

Armstrong, T. B., & Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica*, *86*(2), 655–683. https://doi.org/10.3982/ECTA14434

Armstrong, T. B., & Kolesár, M. (2020). Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics*, *11*(1), 1–39. https://doi.org/10.3982/QE1199

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic Inquiry*, *54*(1), 268–293. https://doi.org/10.1111/ecin.12225

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving babies? revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, *126*(4), 2117–2123. https://doi.org/10.1093/qje/qjr042

Bartalotti, O., Brummet, Q., & Dieterle, S. (2021). A correction for regression discontinuity designs with group-specific mismeasurement of the running variable. *Journal of Business & Economic Statistics*, *39*(3), 833–848. https://doi.org/10.1080/07350015.2020.1737081

Battistin, E., Brugiavini, A., Rettore, E., & Weber, G. (2009). The retirement consumption puzzle: Evidence from a regression discontinuity approach. *American Economic Review*, *99*(5), 2209–2226. https://doi.org/10.1257/aer.99.5.2209

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Card, D., Lee, D. S., Pei, Z., & Weber, A. (2015). Inference on causal effects in a generalized regression kink design. *Econometrica*, *83*(6), 2453–2483. https://doi.org/10.3982/ECTA11224

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models* (2nd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9781420010138

Davezies, L., & Le Barbanchon, T. (2017). Regression discontinuity design with continuous measurement error in the running variable. *Journal of Econometrics*, *200*(2), 260–281. https://doi.org/10.1016/j.jeconom.2017.06.010

Dieterle, S., Bartalotti, O., & Brummet, Q. (2020). Revisiting the effects of unemployment insurance extensions on unemployment: A measurement-error-corrected regression discontinuity approach. *American Economic Journal: Economic Policy*, *12*(2), 84–114. https://doi.org/10.1257/pol.20160439

Dong, Y. (2015). Regression discontinuity applications with rounding errors in the running variable. *Journal of Applied Econometrics*, *30*(3), 422–446. https://doi.org/10.1002/jae.2369

Eckles, D., Ignatiadis, N., Wager, S., & Wu, H. (2022). *Noise-induced randomization in regression discontinuity designs* [arXiv: 2004.09458].

Fan, J., & Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, *21*(4). https://doi.org/10.1214/aos/1176349402

Gelman, A., & Imbens, G. W. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

Hahn, J., Todd, P. E., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209. https://doi.org/doi.org/10.1111/1468-0262.00183

Heitjan, D. F., & Rubin, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, *85*(410), 304–314. https://doi.org/10.1080/01621459.1990.10476202

Heitjan, D. F., & Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics*, *19*(4), 2244–2253. https://doi.org/10.1214/aos/1176348396

Holbein, J. B., & Hillygus, D. S. (2016). Making young voters: The impact of preregistration on youth turnout. *American Journal of Political Science*, *60*(2), 364–382. https://doi.org/10.1111/ajps.12177

Hullegie, P., & Klein, T. J. (2010). The effect of private health insurance on medical care utilization and self-assessed health in Germany. *Health Economics*, *19*(9), 1048–1062. https://doi.org/10.1002/hec.1642

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, *79*(3), 933–959. https://doi.org/10.1093/restud/rdr043

Imbens, G. W., & Wager, S. (2019). Optimized regression discontinuity designs. *The Review of Economics and Statistics*, *101*(2), 264–278. https://doi.org/10.1162/rest_a_00793

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475. https://doi.org/10.2307/2951620

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304. https://doi.org/10.1257/aer.20160945

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, *142*(2), 675–697. https://doi.org/10.1016/j.jeconom.2007.05.004

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655–674. https://doi.org/10.1016/j.jeconom.2007.05.003

Leeb, H., & Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, *21*(1), 21–59. https://doi.org/10.1017/S0266466605050036

Low, M. G. (1997). On nonparametric confidence intervals. *The Annals of Statistics*, *25*(6), 2547–2554. https://doi.org/10.1214/aos/1030741084

Newey, W. K. (2013). Nonparametric instrumental variables estimation. *American Economic Review*, *103*(3), 550–556. https://doi.org/10.1257/aer.103.3.550

Noack, C., & Rothe, C. (2021). *Bias-aware inference in fuzzy regression discontinuity designs* [arXiv: 1906.04631].

Pei, Z., & Shen, Y. (2017). The devil is in the tails: Regression discontinuity design with measurement error in the assignment variable. In M. D. Cattaneo & J. C. Escanciano (Eds.), *Regression discontinuity designs* (pp. 455–502). Emerald Group Publishing. https://doi.org/10.1108/S0731-905320170000038019

Späth, H. (1995). *One dimensional spline interpolation algorithms*. A K Peters. https://doi.org/10.1201/9781439864715

van der Vaart, A., & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer. https://doi.org/10.1007/978-1-4757-2545-2

Appendix A Auxiliary results

A.1 Effect of measurement error on smoothness of conditional mean

We now formalize the notion that measurement error smooths out non-linearities in g_t^*. To this end, first we introduce some definitions. For a real-valued function on a bounded set in $\mathbb{R}^d$, and a multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$, let $D^\alpha f = \partial^{\sum_{j=1}^d \alpha_j} f / \partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}$. For an integer k, let $\|f\|_{C^{k+1}} = \sum_{|\alpha| \le k} \sup_x |D^\alpha f(x)| + \sum_{|\alpha|=k} \sup_{x \neq y} |f^{(k)}(x) - f^{(k)}(y)| / \|x - y\|$ denote the Hölder norm, with the convention that $\|f\|_{C^0} = \sup_x |f(x)|$, and that $\|f\|_{C^k} = \infty$ if f is not $(k-1)$-times differentiable. We say that f has Hölder smoothness index k if $\|f\|_{C^k} < \infty$ (e.g. van der Vaart & Wellner, 1996, Section 2.7.1). This quantifies the "smoothness" of f (k is also called the Hölder exponent; for simplicity we focus attention on exponents that are integers). In other words, f has smoothness k if it is k times differentiable almost everywhere, with the derivatives bounded.

The next result shows that if (a) the conditional density of $f(e; x)$ of e given $X = x$ is sufficiently smooth in the second argument, and (b) $g_t^*(x, e)$ is also sufficiently smooth in the second argument, then the smoothness of g_t is given by the *sum* of the smoothness indices of $x \mapsto g_t^*(x, e)$ and that of $e \mapsto f(e; x)$. This makes precise the notion that measurement error "smooths out" the non-linearities in g_t^*.

Lemma A.1. *Suppose that (X, e) has bounded support, and that the distribution of e given $X = x$ is continuous with bounded density $f(e; x)$ such that $\|\partial^s f / \partial x^s\|_{C^r} < \infty$ for some non-negative integers r, s. Let $h(x, e)$ be a function such that $\sup_e \|\partial^r h(\cdot, e) / \partial e^r\|_{C^s} < \infty$. Then $g(x) := E[h(X + e, e) \mid X = x]$ has smoothness $s + r$.*

Proof. Since the lower-order derivatives $D^{(0,k-1)}f(e;x)$ and $D^{(k-1,0)}h(x+e,e)$ exist and are Lipschitz continuous for $k \leq s$, by dominated convergence theorem, we can take a derivative under the integral sign using the Leibniz product rule, so that, for all x,

$$g^{(s)}(x) = \int \sum_{k=0}^{s} \binom{s}{k} D^{(k,0)}h(x+e,e)D^{(0,s-k)}f(e;x)de.$$

Thus,

$$\begin{aligned}
g^{(r+s)}(x) &= \sum_{k=0}^{s} \binom{s}{k} \frac{d^r}{dx^r} \int D^{(k,0)}h(y,y-x)D^{(0,s-k)}f(y-x;x)dy \\
&= \sum_{k=0}^{s}\sum_{u=0}^{r}\sum_{v=0}^{u} \binom{s}{k}\binom{r}{u}\binom{u}{v}(-1)^{r-u+v} \int D^{(k,r-u)}h(y,y-x)D^{(v,s-k+u-v)}f(y-x;x)dy,
\end{aligned}$$

where the first line follows by change of variables, and the second line by the dominated convergence theorem and Leibnitz product rule. Since $D^{(k,r-u)}h(y,y-x)$ and $D^{(v,s-k+u-v)}f(y-x;x)$ are bounded, it follows that $g^{(r+s)}$ is bounded. □

A.2 Visualization of smoothness constants

Here we assess the smoothness constants suggested by the ROTs using the visualization approach proposed in Noack and Rothe (2021). To explain the approach, suppose that we are interested in a sharp RD regression of an outcome $\tilde{Y}_i$ on a running variable $\tilde{X}_i$, and make the assumption that the conditional mean satisfies $E[\tilde{Y}_i \mid \tilde{X}_i] \in \mathcal{F}_{RD}(M)$. To assess the plausibility of the smoothness bound M, we regress $\tilde{Y}_i$ on a basis function transformation $g(\tilde{X}_i)$ of $\tilde{X}_i$, and the interaction of $g(\tilde{X}_i)$ with $\mathrm{I}\{\tilde{X}_i \geq 0\}$. To ensure that the estimated regression function lies in $\mathcal{F}_{RD}(M)$, we minimize the sum of squared residuals subject to the constraint that the second derivative of the estimated regression function be no larger than M, and equal M at the cutoff. If the basis is sufficiently flexible, the estimated regression function will tend to overfit the data, and therefore represent an extremal element of $\mathcal{F}_{RD}(M)$. If the estimated regression function appears relatively smooth, this thus is an indicator that the choice of M is quite optimistic; if we are clearly overfitting the data, it signals that the choice of M is conservative—it is unlikely that $E[\tilde{Y}_i \mid \tilde{X}_i]$ lies outside $\mathcal{F}_{RD}(M)$.

We use this method to assess the plausibility of the ROTs that we used to calibrate the bounds M_y and M_t in the first stage and reduced form sharp RD regressions. In the former, the outcome $\tilde{Y}_i$ corresponds to the treatment variable T_i, while in the latter, $\tilde{Y}_i = Y_i$. To implement the method, as a basis function, we use a quadratic spline with 21 knots on each

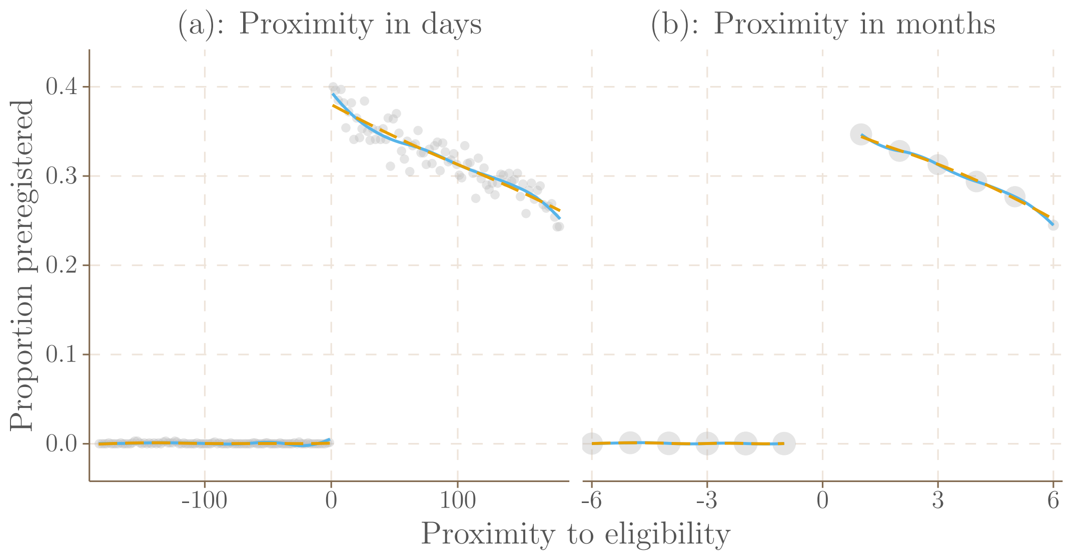

Figure A.1: Visualization of extreme conditional mean functions in the class $\mathcal{F}_{RD}(M_t)$, for different choices of the first stage smoothness constant M_t.

Notes: Orange dotted line visualizes the choice of M_t based on the ROT proposed by Imbens and Wager (2019). Blue solid line visualizes M_t based on the ROT proposed by Armstrong and Kolesár (2020). Values of these constants are given in columns (4) and (5) of Table 1. In panel (a), proximity is measured in days, and each point corresponds to an average of 1,000 individuals. In panel (b), proximity is measured in months, and each point corresponds to an average across all individuals born in a given month.

side of the cutoff when proximity is measured in days, and with 6 knots when it is measured in months.

Figure A.1 visualizes the choices for the first stage smoothness constant M_t, as estimated by the ROTs proposed by Armstrong and Kolesár (2020) and Imbens and Wager (2019). Both choices ROTs appear reasonable based on the figure. Figure A.2 gives an analogous visualization for the choices for the reduced form smoothness constant M_y. Here the ROT proposed by Armstrong and Kolesár (2020) is quite conservative, while the Imbens and Wager (2019) ROT is more optimistic.

A.3 Survey of empirical literature

To understand the prevalence of measurement error issues in the applied papers that use RD designs, we surveyed articles in 7 leading journals (*American Economic Journal: Applied Economics, American Economic Journal: Economic Policy, American Economic Review, Quarterly Journal of Economics, Journal of Political Economy*, *Review of Economics and Statistics*, and *Review of Economic Studies*) published between 2005 and 2020. We identified 139 papers that used RD design, of which 32 papers (23%) featured a running variable

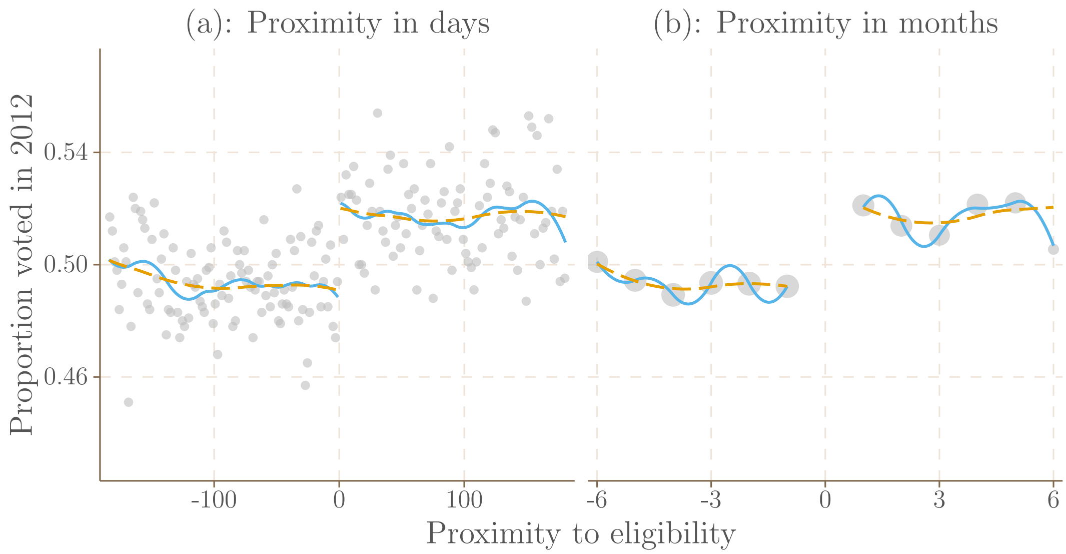

Figure A.2: Visualization of extreme conditional mean functions in the class $\mathcal{F}_{RD}(M_y)$, for different choices of the reduced form smoothness constant M_y.

Notes: Orange dotted line visualizes the choice of M_y based on the ROT proposed by Imbens and Wager (2019). Blue solid line visualizes M_y based on the ROT proposed by Armstrong and Kolesár (2020). Values of these constants are given in columns (4) and (5) of Table 2. In panel (a), proximity is measured in days, and each point corresponds to an average of 1,000 individuals. In panel (b), proximity is measured in months, and each point corresponds to an average across all individuals born in a given month.

measured with error. For each of these articles, we classified the type of measurement error, whether Assumption *(C1)* holds (i) outright, (ii) after doughnut trimming; or, under heaping error, (iii) after dropping the heaping points. Finally, we noted the type of correction employed in the paper.

Table A.1 reports the survey results. 27 out of the 32 papers (84%) feature rounding or grouping error. Out of these, 17 require no explicit measurement error correction, provided we interpret the estimand correctly. 3 papers correctly employ doughnut trimming (or else include a dummy for the cutoff month, which has the same effect). Several papers don't quite deal with measurement error issues correctly, either by failing to create a doughnut hole, or else by accounting for the discreteness of the rounded running variable by clustering the standard errors by the running variable.[22] No paper discusses the implications of the measurement error for the interpretation of the estimand.

Our survey indicates that even though grouping error is fairly common in practice, there is a lack of clarity among applied researchers in how to account for it. Since our survey focuses on the most selective journals, it is likely that due to selection bias, the prevalence of measurement error is even higher than our 23% estimate. Likewise, while grouping error is very common, it likely accounts for a lower share of measurement error types than 84%, as we find in our survey, since other types of measurement error are more difficult to deal with.

[22] As argued in Kolesár and Rothe (2018), clustering exacerbates, rather than solves, any inference issues that the discreteness causes.

Table A.1: Survey of empirical RD papers with measurement error in the running variable.

Paper	Running variable	(*C1*) holds	Error type	Correction
Snyder and Evans (2006)	quarter of birth	Y	R	Not needed
Black et al. (2007)	rounded risk score	Y	R	Not needed
Card et al. (2007)	job tenure in months	Y	R	Not needed
Anderson et al. (2012)	age in months	Y	R	Not needed
Magruder (2012)	distance from magisterial district to bargaining council regime border	Y	R	Not needed
Clark and Royer (2013)	month-year of birth	Y	R	Not needed
Borghans et al. (2014)	month-year of birth	Y	R	Not needed
Anderson et al. (2014)	age in months	Y	R	Not needed
Kumar (2018)	distance from county centroid to Texas border	Y	R	Not needed
Avdic and Karimi (2018)	month-year of birth	Y	R	Not needed
Wherry et al. (2018)	month-year of birth	Y	R	Not needed
Malamud and Pop-Eleches (2010)	month-year of birth	Y	R	CRV
Chetty et al. (2014)	income bins	Y	R	CRV
Haggag and Paci (2014)	interval-reported taxi fare	Y	R	CRV
Erten and Keskin (2018)	month-year of birth	Y	R	CRV
Erten and Keskin (2020)	month-year of birth	Y	R	CRV
Dieterle et al. (2020)	distance from county centroid to state border	Y	R	Bartalotti et al. (2021)
Lalive (2007)	month-year of birth	YD	R	
Stancanelli and van Soest (2012)	month-year of birth	YD	R	
Li et al. (2015)	age in years	D	R	Doughnut
Carpenter and Dobkin (2017)	age in months	D	R	Dummy for cutoff month
Kreiner et al. (2020)	age in months	D	R	Dummy for cutoff month, CRV
Oreopoulos (2006)	year of birth	D	R	CRV
Johnston and Mas (2018)	week of unemployment insurance claim	D	R	
Davis (2008)	calendar month or year	D	R	
Lleras-Muney (2005)	year of birth, sometimes reported in multiples of 10	YD/H	R/H	
Almond et al. (2011)	birth weight	H	H	Series of doughnuts
Barreca et al. (2011)	birth weight	H	H	Series of doughnuts
Almond et al. (2010)	birth weight	H	H	
González (2013)	month of birth / month of abortion / estimated month of conception	Y/D/D	R/R/O	Discussed possible attenuation bias when using estimated month of conception
Battistin et al. (2009)	age in years	N	O	Not needed (under maintained assumptions)
Becker et al. (2013)	revised regional GDP per capita	N	O	

Notes: Whether Assumption (*C1*) holds is coded as: Y—Yes; D—Yes, after doughnut trimming; YD—Yes, possibly after doughnut trimming (there is insufficient information in the paper to determine whether doughnut trimming is needed); H—Yes after dropping heaping points; N—No. Error type is coded as: R—Rounding or grouping error; H—Heaping; O—Other. CRV: clustered standard errors by running variable following Lee and Card (2008).