LOCAL LINEAR TIE-BREAKER DESIGNS

By

Dan M. Kluger
Art B. Owen

Technical Report No. 2021-01
January 2021

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

LOCAL LINEAR TIE-BREAKER DESIGNS

By

Dan M. Kluger
Art B. Owen
Stanford University

Technical Report No. 2021-01
January 2021

**This research was supported in part by
National Science Foundation grant IIS 1837931.**

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://statistics.stanford.edu

Local linear tie-breaker designs

Dan M. Kluger
Stanford University

Art B. Owen
Stanford University

January 2021

Abstract

Tie-breaker experimental designs are hybrids of Randomized Control Trials (RCTs) and Regression Discontinuity Designs (RDDs) in which subjects with moderate scores are placed in an RCT while subjects with extreme scores are deterministically assigned to the treatment or control group. The design maintains the benefits of randomization for causal estimation while avoiding the possibility of excluding the most deserving recipients from the treatment group. The causal effect estimator for a tie-breaker design can be estimated by fitting local linear regressions for both the treatment and control group, as is typically done for RDDs. We study the statistical efficiency of such local linear regression-based causal estimators as a function of Δ, the radius of the interval in which treatment randomization occurs. In particular, we determine the efficiency of the estimator as a function of Δ for a fixed, arbitrary bandwidth under the assumption of a uniform assignment variable. To generalize beyond uniform assignment variables and asymptotic regimes, we also demonstrate on the Angrist and Lavy (1999) classroom size dataset that prior to conducting an experiment, an experimental designer can estimate the efficiency for various experimental radii choices by using Monte Carlo as long as they have access to the distribution of the assignment variable. For both uniform and triangular kernels, we show that increasing the radius of randomized experiment interval will increase the efficiency until the radius is the size of the local-linear regression bandwidth, after which no additional efficiency benefits are conferred.

1 Introduction

The regression discontinuity design (RDD) introduced by Thistlethwaite and Campbell (1960) has become a mainstay of causal inference in recent years, especially in econometrics (Angrist and Pischke, 2014) and the social sciences (Imbens and Rubin, 2015). The logic in an RDD is as follows. Some different treatment is offered to subject i depending on whether the value x_i of an assignment variable (also called a running variable) exceeds a threshold t, or not. Then if the future value of a quantity Y_i has a different expected value for x_i just barely larger than t than for x_i just barely smaller than t it becomes quite credible to interpret the difference causally, assuming that there is no a priori reason for $\mathbb{E}(Y \mid x)$ to have a step discontinuity at $x = t$.

In most uses of RDD, the assignment to treatment or control cannot be influenced by the investigator. In some settings, however, the investigator can inject randomness into the treatment assignments around the threshold value of $x = t$. The goal is to better measure the effect of the treatment variable Z on Y. This design is called a tie-breaker design. For instance, letting x be a measure of high-school performance and the treatment be awarding of a university scholarship, Angrist et al. (2014) use a tie-breaker design to measure the impact of the scholarship on students' academic outcomes. The top performing students all get the scholarship, the bottom performing students do not, and there is a group in the middle where a random selection is used. Aiken et al. (1998) use a tie-breaker design to offer remedial English classes based on a test given to students prior to matriculation. Companies offering loyalty programs to their best customers can also include some randomness in the reward, to better learn the impact of those rewards. Tie-breaker designs are also known as cutoff designs in clinical trial settings (Trochim and Cappelleri, 1992).

In our setting it is convenient to represent the treatment by $Z \in \{-1, 1\}$, with the level -1 for the control group. In a tie-breaker design, the investigator proceeds in these steps:

1) collect assignment variable values x_i for subjects $i = 1, \dots, N$,
2) determine a distribution for treatment variables $Z_i \in \{-1, 1\}$,
3) assign treatments Z_i to subjects,
4) observe corresponding Y_i, and
5) infer the treatment effect from (x_i, Z_i, Y_i) values.

The Y_i values are not available to the investigator at the time the distribution for Z_i is chosen. As a result the treatment decisions have to be chosen in part based on a guess as to what model might be used to fit the Y_i values. We are interested in settings where Y_i does not become available fast enough to employ bandit methods. For instance there might be a year long delay in the customer loyalty setting or six years in the educational attainment setting.

Owen and Varian (2020) study the statistical efficiency of tie-breaker designs. In the simplest formulation, the assignment variables are sorted and scaled so that $x_i = (2i - 1)/(2N) \in (-1, 1)$. Then for $0 \leqslant \Delta \leqslant 1$,

$$\Pr(Z_i = 1 \mid x_i) = \begin{cases} 0, & x_i \leqslant -\Delta \\ 1/2, & |x_i| < \Delta \\ 1, & x_i \geqslant \Delta. \end{cases} \tag{1}$$

This tie-breaker design interpolates between an RDD with $t = 0$ when $\Delta = 0$ and an RCT for $\Delta = 1$. They consider a regression model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 Z_i + \beta_3 x_i Z_i + \varepsilon_i \tag{2}$$

where ε_i are IID random variables with mean 0 and variance σ^2. They find statistical efficiency improves with Δ. We will describe below how efficiency is defined when we introduce our version. They also find no efficiency benefit from using a sliding scale for $\Pr(Z = 1 \mid x)$ instead of using just the three levels in (1). Model (2) is simple enough to analyze; they also describe algorithmic ways to study more general alternative models at step 2 when there may be vectors $\boldsymbol{x}_i$ available for each subject.

A weakness of model (2) is that it is global over all subjects. In RDDs, it is now more common to use nonparametric regression methods (Hahn et al., 2001; Calonico et al., 2014). Let

$$\mu_+(x) = \mathbb{E}(Y \mid x, Z = 1) \quad \text{and} \quad \mu_-(x) = \mathbb{E}(Y \mid x, Z = -1),$$

both assumed to be smooth in a neighborhood of $x = t$. The RDD is used to estimate $\mu_+(t) - \mu_-(t)$. A kernel regression $\hat{\mu}_-$ is fit to (x_i, Y_i) pairs with $Z_i = -1$ and hence $x_i \leqslant t$ and extrapolated to $\hat{\mu}_-(t)$. Similarly, a kernel regression estimate $\hat{\mu}_+(t)$ based on (x_i, Y_i) data with $x_i \geqslant t$ is produced and then the causal impact of treatment at $x = t$ is estimated by $\hat{\mu}_+(t) - \hat{\mu}_-(t)$.

Our goal in this paper is to show that the tie-breaker design is still superior to the RDD even when a local regression is used. The local regression will use a bandwidth parameter h chosen at step 5 above. That bandwidth is not known to the investigator at step 2 where Δ is chosen. To handle this, we show that the tie-breaker has a statistical advantage at all $h > 0$.

An outline of this paper is as follows. Section 2 introduces the kernel regressions we consider and gives an asymptotic formula for the variance of the estimated coefficients. Section 3 gives integral approximations to that variance and defines our notion of efficiency of the tie-breaker design with respect to the regression discontinuity design. Section 4 gives expressions for the efficiency of the tie-breaker design for two kernels of interest: a uniform boxcar kernel and an asymptotically optimal triangular kernel. The efficiency of the tie-breaker design depends on the ratio $\delta = \Delta/h$ of the size of the experimental region to the bandwidth used. Efficiency ranges from 1 at $\delta = 0$ to 4 at $\delta = 1$ (for the boxcar kernel) or to 3.6 at $\delta = 1$ for the triangular kernel. The efficiency is monotone in δ. To get an interpretable theory we have worked with a uniformly spaced assignment variable. Section 5 shows how one can compute efficiency empirically using one's actual assignment variables, focusing on the Israeli classroom size data from Angrist and Lavy (1999) as an example. The efficiency curves are quite similar to the ones we have for uniformly spaced assignment variables but they show slightly greater efficiency gains than the ones in our theorems for uniform spacings. Section 6 presents a discussion and an Appendix contains one of our proofs.

2 Kernels and problem formulation

In keeping with current RDD practice we will use a kernel smoothed version of (2). The parameter vector β is estimated by

$$\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^4} \sum_{i=1}^{N} K\Bigl(\frac{x_i - t}{h}\Bigr)\bigl(Y_i - (\beta_1 + \beta_2 x_i + \beta_3 Z_i + \beta_4 x_i Z_i)\bigr)^2 \qquad (3)$$

for a symmetric kernel function $K(\cdot) \geqslant 0$ and a bandwidth parameter $h > 0$. We suppose that $K(\cdot)$ is non-negative and symmetric. We have special interest in a uniform (boxcar) kernel $K_{\mathrm{BC}}(x) = 1_{|x|\leqslant 1}$ because it is a local version of the regression model (2). We are also interested in a triangular spike kernel $K_{\mathrm{TS}}(x) = (1 - |x|)_+$ where $z_+ = \max(0, z)$. This triangular kernel was shown by Cheng et al. (1997) to optimize a bias-variance tradeoff for extrapolation from $x_i > t$ to $\mathbb{E}(Y \,|\, x = t)$ and has been advocated for RDD analysis by Imbens and Kalyanaraman (2012) and Calonico et al. (2014) among others.

In the model (2), the treatment effect at x is

$$(\beta_0 + \beta_2 x + \beta_3 + \beta_4 x) - (\beta_0 + \beta_2 x - \beta_3 - \beta_4 x) = 2(\beta_3 + \beta_4 x).$$

An RDD is used for $x = t$ and there the treatment effect is $2(\beta_3 + \beta_4 t)$. We will shift the assignment variable so as to make $t = 0$ and then focus on β_3.

The kernel regression estimator from (3) has a bias and variance that both depend on the bandwidth h. Larger h typically bring greater bias because the true regression is not precisely linear over a region centered on t. Smaller h bring greater variance because then fewer data points are in the regression. Calonico et al. (2014) advocate for choosing smaller h than is mean square optimal. Making the bias negligible compared to the standard deviation has the effect of making it much easier to get a confidence interval for the treatment effect. Confidence intervals are important when the estimate is to be used for policy purposes where quantifying uncertainty is critical. For this reason, we will study the variance of $\hat{\beta}_3$ given h and not consider the bias. That is, we are assuming that the user will purposefully undersmooth the regression as recommended by Calonico et al. (2014).

The design matrix for the regression is $\mathcal{X} \in \mathbb{R}^{N\times 4}$ with i'th row $(1, x_i, Z_i, x_i Z_i)$. The response is $\mathcal{Y} = (Y_1, \dots, Y_N)^\mathsf{T}$. With $t = 0$, the kernel weights are $K(x_i/h)$ and we let $\mathcal{W} = \mathcal{W}(h) \in \mathbb{R}^{N\times N} = \mathrm{diag}(K(x_i/h))$. Then

$$\hat{\beta} = \hat{\beta}(\Delta) = (\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{Y} \qquad (4)$$

and under the assumption that $\mathrm{var}(\mathcal{Y} \,|\, \mathcal{X}) = \sigma^2 I_N$ we have

$$\mathrm{var}(\hat{\beta} \,|\, \mathcal{X}; \Delta) = (\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^\mathsf{T}\mathcal{W}^2\mathcal{X}(\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{X})^{-1}\sigma^2. \qquad (5)$$

Formula (4) for $\hat{\beta}$ matches the familiar generalized least squares formula for the case where $\mathrm{var}(\mathcal{Y} \mid \mathcal{X}) = \mathcal{W}\sigma^2$. Here $\mathcal{W}$ arises from weights that are not of inverse variance type and hence the formula for $\mathrm{var}(\hat{\beta} \mid \mathcal{X}; \Delta)$ involves a $\mathcal{W}^2$

factor and less cancellation than we might have expected. The boxcar kernel is special because then $K(x_i/h) \in \{0,1\}$ equals its own square. In that case $\mathrm{var}(\hat\beta\,|\,\mathcal{X};\Delta) = (\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{X})^{-1}\sigma^2$.

The estimand in regression discontinuity is $2\beta_3$. Therefore we study $\mathrm{var}(\hat\beta_3\,|\,\mathcal{X};\Delta)$ under a tie-breaker design as $(\mathrm{var}(\hat\beta\,|\,\mathcal{X};\Delta))_{3,3}$ using the expression in (5).

3 Asymptotic approximation via integrals

Our present analysis takes place before step 1 from the outline in the introduction: we want to study $\mathrm{var}(\hat\beta_3\,|\,\mathcal{X};\Delta)$, but we do not yet have the x_i. Given any list of x_i some numerical methods described in Owen and Varian (2020) can be adapted to the kernel regression setting, but that does not give theoretical insight. Without access to x_i we will study the uniformly spaced setting with $x_i = (2i-1)/(2N)$. This case is simple enough to illustrate the gains from a tie-breaker design. In practice we could apply it by replacing x_i by their centered and scaled ranks. In Section 6 we explain using asymptotic theory from Fan and Gijbels (1996) why this case models the most important features of the problem and in Section 5 we show numerical results on a real data set that is not equispaced.

For $x_i = (2i-1)/(2N)$, the matrices $\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{X}/N$ and $\mathcal{X}^\mathsf{T}\mathcal{W}^2\mathcal{X}/N$ contain elements that can be approximated by integrals of the form

$$\mathcal{I}^{rst} = \mathcal{I}^{rst}(\Delta,h,K) \equiv \frac{1}{2}\int_{-1}^{1} x^r\,\mathbb{E}(Z^s\,|\,x;\Delta)K\Bigl(\frac{x}{h}\Bigr)^t\,\mathrm{d}x \tag{6}$$

for integer exponents r, s and t. Our expressions will simplify somewhat because $Z^2=1$ making every $\mathcal{I}^{r,2,t} = \mathcal{I}^{r,0,t}$ and also because both x and $\mathbb{E}(Z \mid x;\Delta)$ are antisymmetric functions of x making them orthogonal to $K(x/h)$ which we have assumed to be symmetric. The error in those moment approximations is $O_p(N^{-1/2})$ if the Z_i are independent random variables. The error can be much less with other sampling schemes. For instance, we could use stratified sampling, forming pairs of subjects $(i,i+1)$ in the experimental region and randomly setting $Z_i=\pm1$ and $Z_{i+1}=-Z_i$. We will use $\approx$ to describe approximations that are $O_p(N^{-1/2})$ or better.

Applying first $Z^2=1$ and then using symmetry and anti-symmetry

$$\frac{1}{N}\mathcal{X}^\mathsf{T}\mathcal{W}\mathcal{X} \approx \begin{array}{c} \\ 1 \\ x \\ z \\ xz \end{array}\!\!\begin{array}{c} \begin{array}{cccc} \;\;1\;\; & \;\;x\;\; & \;\;z\;\; & \;\;xz\;\; \end{array} \\ \begin{bmatrix} \mathcal{I}^{001} & \mathcal{I}^{101} & \mathcal{I}^{011} & \mathcal{I}^{111} \\ \mathcal{I}^{101} & \mathcal{I}^{201} & \mathcal{I}^{111} & \mathcal{I}^{211} \\ \mathcal{I}^{011} & \mathcal{I}^{111} & \mathcal{I}^{021} & \mathcal{I}^{121} \\ \mathcal{I}^{111} & \mathcal{I}^{211} & \mathcal{I}^{121} & \mathcal{I}^{221} \end{bmatrix}\end{array} = \begin{bmatrix} \mathcal{I}^{001} & 0 & 0 & \mathcal{I}^{111} \\ 0 & \mathcal{I}^{201} & \mathcal{I}^{111} & 0 \\ 0 & \mathcal{I}^{111} & \mathcal{I}^{001} & 0 \\ \mathcal{I}^{111} & 0 & 0 & \mathcal{I}^{201} \end{bmatrix}.$$

Because $K^2(\cdot)$ is also a symmetric function we also get

$$\frac{1}{N}\mathcal{X}^{\mathsf{T}}\mathcal{W}^2\mathcal{X} \approx \begin{bmatrix} \mathcal{I}^{002} & 0 & 0 & \mathcal{I}^{112} \\ 0 & \mathcal{I}^{202} & \mathcal{I}^{112} & 0 \\ 0 & \mathcal{I}^{112} & \mathcal{I}^{002} & 0 \\ \mathcal{I}^{112} & 0 & 0 & \mathcal{I}^{202} \end{bmatrix}.$$

From all of the symmetries involved in the 32 components of these two matrices, we need to consider at most six distinct integrals. We rewrite them, beginning with

$$\frac{1}{N}\mathcal{X}^T\mathcal{W}\mathcal{X} \approx \begin{bmatrix} \nu_0 & 0 & 0 & \phi(\Delta) \\ 0 & \nu_2 & \phi(\Delta) & 0 \\ 0 & \phi(\Delta) & \nu_0 & 0 \\ \phi(\Delta) & 0 & 0 & \nu_2 \end{bmatrix} \tag{7}$$

where

$$\begin{aligned} \nu_0 &= \frac{1}{2}\int_{-1}^{1} K\Big(\frac{x}{h}\Big)\,\mathrm{d}x, \quad \nu_2 = \frac{1}{2}\int_{-1}^{1} x^2 K\Big(\frac{x}{h}\Big)\,\mathrm{d}x, \quad \text{and} \\ \phi(\Delta) &= \frac{1}{2}\int_{-1}^{-\Delta}(-x)K\Big(\frac{x}{h}\Big)\,\mathrm{d}x + \frac{1}{2}\int_{\Delta}^{1} xK\Big(\frac{x}{h}\Big)\,\mathrm{d}x = \int_{\Delta}^{1} xK\Big(\frac{x}{h}\Big)\,\mathrm{d}x. \end{aligned} \tag{8}$$

Note that ν_0 and ν_2 may depend on h but they do not depend on Δ. A similar argument shows that

$$\frac{1}{N}\mathcal{X}^T\mathcal{W}^2\mathcal{X} \approx \begin{bmatrix} \pi_0 & 0 & 0 & \psi(\Delta) \\ 0 & \pi_2 & \psi(\Delta) & 0 \\ 0 & \psi(\Delta) & \pi_0 & 0 \\ \psi(\Delta) & 0 & 0 & \pi_2 \end{bmatrix} \tag{9}$$

for

$$\begin{aligned} \pi_0 &= \frac{1}{2}\int_{-1}^{1} K^2\Big(\frac{x}{h}\Big)\,\mathrm{d}x, \quad \pi_2 = \frac{1}{2}\int_{-1}^{1} x^2 K^2\Big(\frac{x}{h}\Big)\,\mathrm{d}x, \quad \text{and} \\ \psi(\Delta) &= \int_{\Delta}^{1} xK^2\Big(\frac{x}{h}\Big)\,\mathrm{d}x. \end{aligned} \tag{10}$$

Now we are ready to describe the asymptotic variance of $\hat\beta_3$.

Theorem 1. *Let $x_i = (2i-1)/(2N)$, select $Z_i \in \{-1,1\}$ by the tie-breaker equation* (1)*. Let Y_i be uncorrelated random variables with common variance σ^2, conditionally on $\mathcal{X} = ((1, x_1, Z_1, x_1Z_1), \cdots, (1, x_N, Z_N, x_NZ_N))$. Next, for a symmetric kernel $K(\cdot) \geqslant 0$ with $0 < \int_{-\infty}^{\infty} x^2K(x)\,\mathrm{d}x < \infty$ and a bandwidth $h > 0$, let $\hat\beta$ be estimated by the kernel weighted regression* (3)*. Then*

$$N\mathrm{var}(\hat\beta_3 \,|\, \mathcal{X}; \Delta) = \frac{\sigma^2\big(\nu_2^2\pi_0 - 2\nu_2\phi(\Delta)\psi(\Delta) + \pi_2\phi^2(\Delta)\big)}{\big(\nu_0\nu_2 - \phi^2(\Delta)\big)^2} + O_p\Big(\frac{1}{\sqrt{N}}\Big), \tag{11}$$

where ν_0, ν_2 *and* $\phi(\Delta)$ *are defined in* (8) *and* π_0, π_2 *and* $\psi(\Delta)$ *are defined in* (10).

Proof. Reordering the components of β we find after substituting equations (7) and (9) into (5) that $\sqrt{N}(\hat\beta_1, \hat\beta_4, \hat\beta_2, \hat\beta_3)$ has variance

$$\begin{pmatrix} \nu_0 & \phi & 0 & 0 \\ \phi & \nu_2 & 0 & 0 \\ 0 & 0 & \nu_0 & \phi \\ 0 & 0 & \phi & \nu_2 \end{pmatrix}^{-1} \begin{pmatrix} \pi_0 & \psi & 0 & 0 \\ \psi & \pi_2 & 0 & 0 \\ 0 & 0 & \pi_0 & \psi \\ 0 & 0 & \psi & \pi_2 \end{pmatrix} \begin{pmatrix} \nu_0 & \phi & 0 & 0 \\ \phi & \nu_2 & 0 & 0 \\ 0 & 0 & \nu_0 & \phi \\ 0 & 0 & \phi & \nu_2 \end{pmatrix}^{-1} \sigma^2 + O_p\Bigl(\frac{1}{\sqrt{N}}\Bigr).$$

Now (11) follows directly by matrix inversion and multiplication. □

The variance formula in Theorem 1 does not require the linear model (2) to hold. When it does not hold there will generally be some bias where $\mathbb{E}(2\hat\beta_3 \mid \mathcal{X};\Delta) \neq \mu_+(0) - \mu_-(0)$. We suppose that the user will choose an undersmoothed h making bias smaller than the standard error, as recommended by Calonico et al. (2014). That undersmoothing takes place in our step 5 above and is not available when Δ is chosen.

We are primarily interested in comparing the asymptotic variance of $\hat\tau_0 = 2\hat\beta_3$ for various choices of Δ. We especially want to compare the efficiency of tie-breaker designs with $\Delta > 0$ to the RDD with $\Delta = 0$. To do this we consider the efficiency ratio

$$\mathrm{Eff}^{(N)}(\Delta) \equiv \frac{\mathrm{var}(\hat\beta_3 \,|\, \mathcal{X}; 0)}{\mathrm{var}(\hat\beta_3 \,|\, \mathcal{X}; \Delta)}. \tag{12}$$

Using Theorem 1, $\mathrm{Eff}^{(N)}(\Delta)$ converges in probability to the asymptotic efficiency ratio

$$\mathrm{Eff}(\Delta) = \frac{\bigl(\nu_2^2\pi_0 - 2\nu_2\phi(0)\psi(0) + \pi_2\phi^2(0)\bigr)\bigl(\nu_0\nu_2 - \phi^2(\Delta)\bigr)^2}{\bigl(\nu_2^2\pi_0 - 2\nu_2\phi(\Delta)\psi(\Delta) + \pi_2\phi^2(\Delta)\bigr)\bigl(\nu_0\nu_2 - \phi^2(0)\bigr)^2} \tag{13}$$

using quantities that we defined at (8) and (10).

4 Efficiency with boxcar and triangular kernels

In this section we present the efficiency ratios under the conditions of Theorem 1 for the two kernels of greatest interest: the boxcar kernel and the triangular kernel. We work with $x_i = (2i - N)/(2N)$ throughout this section.

For the boxcar kernel $K_{\mathrm{BC}}(x) = 1_{|x|\leqslant 1}$, we can assume without loss of generality that $h \leqslant 1$ because there are no data with $|x_i - t| = |x_i| > 1$, and then any $h > 1$ will give the same estimate as $h = 1$. We find for this kernel that

$$\nu_0 = \pi_0 = h, \quad \nu_2 = \pi_2 = \frac{h^3}{3}, \quad \text{and} \quad \phi(\Delta) = \psi(\Delta) = \frac{(h^2 - \Delta^2)_+}{2}. \tag{14}$$

Using some foresight, we define the local tie-breaker constant $\delta = \Delta/h$. This is the fraction of the local regression region in which the treatment was assigned at random.

Proposition 1. *Under the conditions of Theorem 1 and using the boxcar kernel* K_{BC}*, the asymptotic efficiency of the tie-breaker design is*

$$\mathrm{Eff}_{\mathrm{BC}} = 1 + 6\delta^2 - 3\delta^4 \tag{15}$$

for $\delta = \Delta/h \leqslant 1$*. If* $\delta > 1$*, then* $\mathrm{Eff}_{\mathrm{BC}} = 4$.

Proof. Because many quantities from (14) are identical, substituting them into (13) produces numerous simplifications that yield

$$\mathrm{Eff}_{\mathrm{BC}} = \frac{\nu_0\nu_2 - \phi^2(\Delta)}{\nu_0\nu_2 - \phi^2(0)} = \frac{\frac{h^4}{3} - \frac{(h^2-\Delta^2)_+^2}{4}}{\frac{h^4}{3} - \frac{h^4}{4}} = 4 - 3(1-\delta^2)_+^2.$$

For $0 \leqslant \delta < 1$ formula (15) follows from expanding the quadratic while for $\delta > 1$ the positive part term vanishes. □

Choosing $h = 1$ makes the local regression a global one. We then get the same efficiency ratio as in equation (6) from Owen and Varian (2020). By taking derivatives it is easy to show that the efficiency ratio in (15) is strictly increasing as the local amount of experimentation δ varies over the interval $0 < \delta < 1$. Figure 1 plots $\mathrm{Eff}_{\mathrm{BC}}$ versus δ.

The triangular spike kernel $K_{\mathrm{TS}}(x) = (1-|x|)_+$ (triangular kernel for short) is more complicated than the boxcar kernel because for it, K^2 is not proportional to K. Once again, we assume that $h \in [0,1]$. For this kernel we compute

$$\nu_0 = \frac{h}{2}, \quad \nu_2 = \frac{h^3}{12}, \quad \pi_0 = \frac{h}{3}, \quad \text{and} \quad \pi_2 = \frac{h^3}{30}$$

and then using $\delta = \Delta/h$, we get

$$\phi(\Delta) = \frac{h^2}{6}(1 - 3\delta^2 + 2\delta^3) \quad \text{and} \quad \psi(\Delta) = \frac{h^2}{12}(1 - 6\delta^2 + 8\delta^3 - 3\delta^4).$$

Proposition 2. *Under the conditions of Theorem 1 and using the triangular kernel* K_{TS}*, the asymptotic efficiency of the tie-breaker design is*

$$\mathrm{Eff}_{\mathrm{TS}} = \frac{2\bigl(3 - 2(1-3\delta^2+2\delta^3)^2\bigr)^2}{5 - 5(1-3\delta^2+2\delta^3)(1-6\delta^2+8\delta^3-3\delta^4) + 2(1-3\delta^2+2\delta^3)^2} \tag{16}$$

for $\delta = \Delta/h \leqslant 1$.

Proof. When we substitute values into the efficiency formula (13) we get some simplifications from $\pi_0 = (2/3)\nu_0$ and $\pi_2 = (2/5)\nu_2$. The constant term in $N\mathrm{var}(\hat\beta_3)/\sigma^2$ becomes

$$\frac{(2/3)\nu_2^2\nu_0 - 2\nu_2\phi(\Delta)\psi(\Delta) + (2/5)\nu_2}{(\nu_0\nu_2 - \phi^2(\Delta))^2} = 15\nu_2\frac{10\nu_2\nu_0 - 30\phi(\Delta)\psi(\Delta) + 6}{(\nu_0\nu_2 - \phi^2(\Delta))^2}$$

so that the efficiency ratio is

$$\mathrm{Eff}_{\mathrm{TS}} = \frac{5\nu_0\nu_2 - 15\phi(0)\psi(0) + 3\phi^2(0)}{5\nu_0\nu_2 - 15\phi(\Delta)\psi(\Delta) + 3\phi^2(\Delta)} \times \frac{(\nu_0\nu_2 - \phi^2(\Delta))^2}{(\nu_0\nu_2 - \phi^2(0))^2} \tag{17}$$

after cancelling a common factor of $30\nu_2$. Next $5\nu_0\nu_2 - 15\phi(\Delta)\psi(\Delta) + 3\phi^2(\Delta)$ equals

$$\frac{5h^4}{24} - 15\frac{h^4}{72}(1-3\delta^2+2\delta^3)(1-6\delta^2+8\delta^3-3\delta^4) + \frac{h^4}{12}(1-3\delta^2+2\delta^3)^2$$

and so the first factor in (17) is

$$\frac{2}{5-5(1-3\delta^2+2\delta^3)(1-6\delta^2+8\delta^3-3\delta^4)+2(1-3\delta^2+2\delta^3)^2}.$$

Turning to the second factor

$$\nu_0\nu_2 - \phi^2(\Delta) = \frac{h^4}{24} - \frac{h^4}{36}(1-3\delta^2+2\delta^3)^2 = \frac{h^4}{72}\bigl(3-2(1-3\delta^2+2\delta^3)^2\bigr)$$

and so the second factor equals $(3-2(1-3\delta^2+2\delta^3)^2)^2$, establishing (16). ☐

The second panel in Figure 1 shows $\mathrm{Eff}_{\mathrm{TS}}$ versus the local experiment size δ. The efficiency curve has a similar monotone increasing shape as we saw for the boxcar kernel. The maximum efficiency ratio, at $\delta = 1$, is $18/5 = 3.6$ instead of 4. The efficiency ratio is a rational function of δ with a numerator of degree 12 and a denominator of degree 7. It is strictly increasing on the interval $0 < \delta < 1$, though the proof is lengthy enough to move to the Appendix.

Proposition 3. *The derivative of* $\mathrm{Eff}_{\mathrm{TS}}$ *with respect to* δ *is positive for* $0 < \delta < 1$.

Proof. See the Appendix. ☐

5 Classroom size data

We explored the efficiency ratio for the tie-breaker design for x_i with a uniform distribution. While that can be arranged by using ranks, in other situations we might prefer to use the original value of a running variable and those might not be uniformly distributed. We demonstrate how this would be done using a dataset from Angrist and Lavy (1999) on classroom sizes.

Angrist and Lavy (1999) studied the causal effect of classroom size on test performance of elementary school students in Israel. In Israel, the Maimonides rule mandates that elementary school classes cannot exceed 40 students. If a school has 41 students enrolled in a particular grade that grade must be split into two classes. Note that grades that have 40 or fewer enrolled students are allowed to split into multiple classes and that grades with slightly more than 40

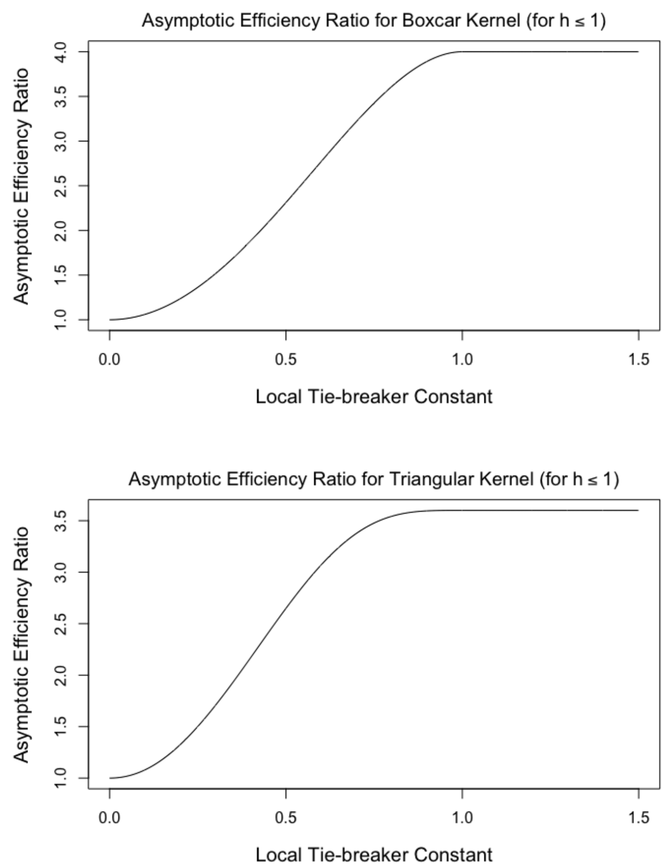


Figure 1: The top panel shows the efficiency of local tie-breaker design for uniform x_i and a boxcar kernel as a function of $\delta = \Delta/h$. The lower panel shows this efficiency for a triangular kernel.

students occasionally violate the Maimonides rule and do not split into multiple classes. Despite this, we can consider this a setting for RDD where the treatment variable is whether or not the school is legally mandated to split a particular grade into smaller classes.

The dataset, published on the Harvard Dataverse (Angrist and Lavy, 2009), has verbal and math scores for 3rd, 4th and 5th graders across Israel. We chose to focus exclusively on 4th grade verbal scores as our response variable and 4th grade enrollments as our assignment variable because Angrist and Lavy (1999) suggest that a slightly significant effect of the treatment on 4th grade verbal scores exists. A case could be made for a tie-breaker design in this setting because had one been used, the treatment effect might have been more

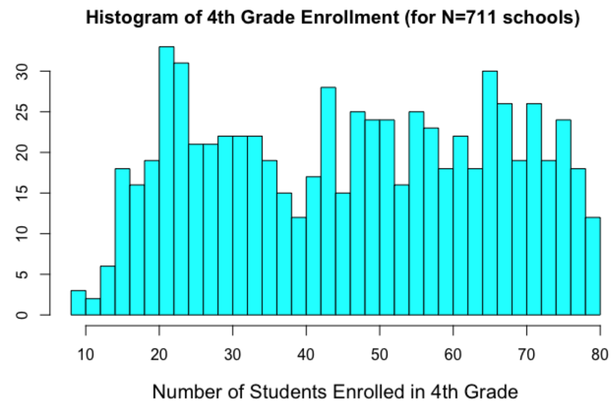
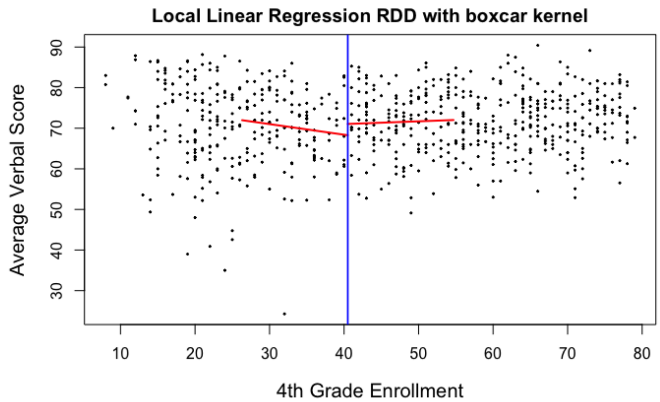


Figure 2: A histogram of 4th grade enrollments for our filtered dataset (with schools exceeding 80 4th grade students or three 4th grade classes removed)

accurately estimated.

To simplify the analysis, we removed all schools that either had more than 80 students or more than two 4th-grade classes from the dataset. We further removed all schools that had NA entries for either class size or verbal scores, leaving $N = 711$ schools in our filtered dataset. See Figure 2 for a visualization of the distribution of the 4th grade enrollments and Figure 3 for visualizations of the local linear regression based-RDD on this dataset using boxcar and triangular kernels. We use the bandwidths h_{IK} given by the Imbens and Kalyanaraman (2012) procedure. The apparent benefit from smaller classrooms is positive but small and it turns out, not statistically significant in this analysis. The 95% confidence interval (assuming homoscedastic errors) for the effect size at the boundary of the local linear regression-based RDD was $(-1.4, 7.0)$ when a boxcar kernel with bandwidth $h_{\mathrm{IK,BC}} = 14.18$ was used. The 95% confidence interval for the effect size at boundary of this RDD was $(-2.4, 9.4)$ when a triangular kernel with bandwidth $h_{\mathrm{IK,TS}} = 9.02$ was used.

Next we illustrate how an investigator can estimate the efficiency of tie-breaker designs as a function of Δ on sample values of the assignment variable. First we center the data, replacing x_i by $x_i - 40.5$ to move the target from $t = 40.5$ to $t = 0$. Next, for each Δ of interest we use 1000 Monte Carlo samples to estimate $\mathrm{var}(\hat{\beta}_3 \mid \mathcal{X}; \Delta)$ and also $\mathrm{var}(\hat{\beta}_3 \mid \mathcal{X}; 0)$, both up to a constant σ^2. That gives us 1000 efficiency ratios $\mathrm{Eff}^{(N)}(\Delta) = \mathrm{var}(\hat{\beta}_3 \,|\, \mathcal{X}; 0)/\mathrm{var}(\hat{\beta}_3 \,|\, \mathcal{X}; \Delta)$ for each Δ. In each of our 1000 samples, we simulate random assignments for a tie-breaker design at the given experimental radius Δ. The random assignments are stratified: in each consecutive pair of classroom sizes in the experimental region, one was randomly chosen to have $Z = 1$ and the other got $Z = -1$. The random Z_i let us compute the matrices $\mathcal{X}$ and $\mathcal{W}$ defined in the beginning of Section 2, from which we compute a non-asymptotic $\mathrm{var}(\hat{\beta}_3 \mid \mathcal{X}; \Delta)$. Note that we do not need to simulate any Y values to do this, because, in this initial



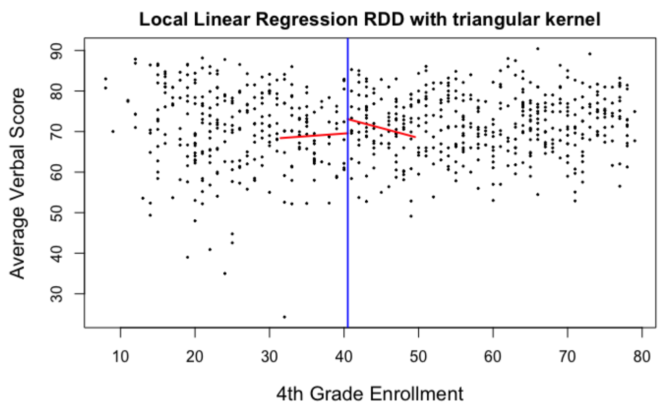


Figure 3: RDD fit to the 4th grader verbal scores from the Angrist and Lavy (2009) dataset when using a boxcar kernel (top) and a triangular kernel (bottom). For these two fits, the bandwidths $h_{\mathrm{IK,BC}}$ and $h_{\mathrm{IK,TS}}$ were chosen by the procedure in Imbens and Kalyanaraman (2012).

analysis, we are retaining the bandwidths from the Imbens and Kalyanaraman (2012) procedure on the original data.

Figure 4 shows boxplots of 1000 simulated $\mathrm{Eff}^{(N)}(\Delta)$ values for various choices of $\Delta \in \mathbb{N}$ to plot the full efficiency curve. It is clear from Figure 4 that with stratified allocations the efficiency is very reproducible. Figure 5 shows results for different bandwidths, ranging from $h_{\mathrm{IK}}/2$ to $3h_{\mathrm{IK}}$. Because the efficiencies are so reproducible given the bandwidth, we just plot curves of the mean and standard deviations of estimated Eff values. For the boxcar kernel we see that the tie-breaker design is reproducibly more efficient than the

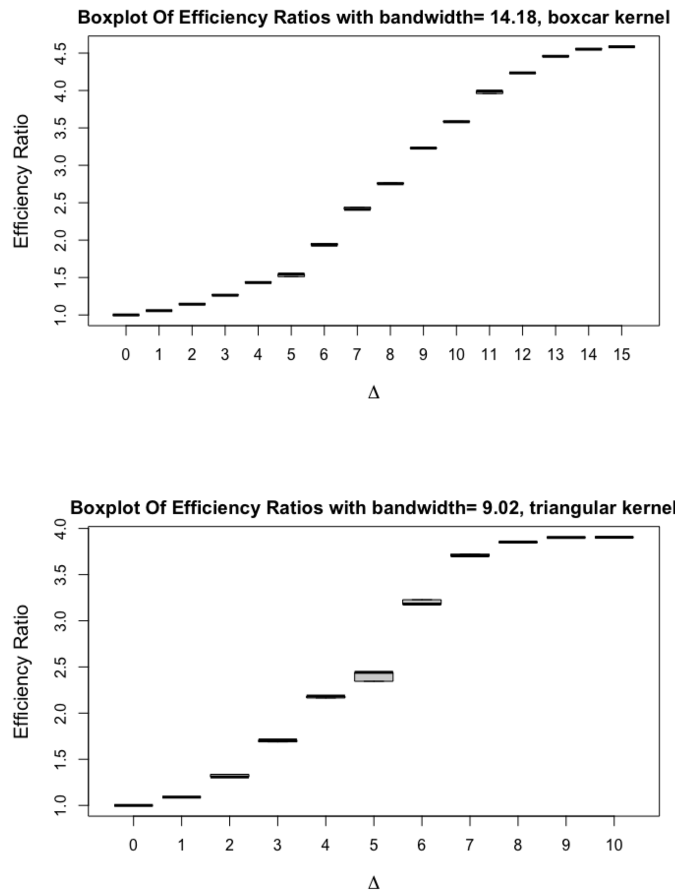


Figure 4: Boxplots of the Monte-Carlo efficiency ratio estimates for various values of $\Delta \in \mathbb{N}$ when using a boxcar kernel (top) and a triangular kernel (bottom). For both the boxcar kernel and the triangular kernel, we used the same bandwidth as in Figure 3, namely $h_{\mathrm{IK,BC}} = 14.18$ and $h_{\mathrm{IK,TS}} = 9.02$.

RDD and the effect increases as $\delta = \Delta/h$ increases for all h we studied. For the triangular kernel we see much the same thing apart from one value of δ and the smallest bandwidth where the tie-breaker comes out less efficient than RDD. For that point the experimental region consisted of just 17 data points, 8 with a class of 40 students and 9 with a class of 41 students.

For a further discussion of the Maimonides rule, see Angrist et al. (2019). They consider different data sets and also investigate the possibility that the class sizes are sometimes manipulated to be above the threshold triggering a classroom split.

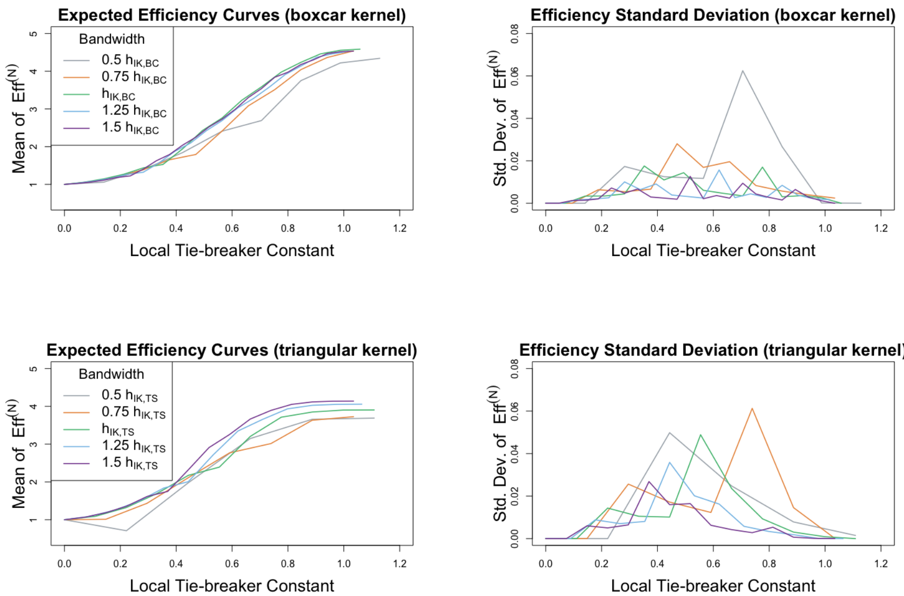


Figure 5: Monte-Carlo based estimates of the expected value (left) and standard deviation (right) of $\mathrm{Eff}^{(N)}(\Delta)$ versus Δ/h for the Angrist and Lavy (2009) dataset of 4th grader verbal scores. For these plots a boxcar kernel (top) and a triangular kernel (bottom) were used. The bandwidths plotted are different scalar multiples of $h_{\mathrm{IK,BC}}$ and $h_{\mathrm{IK,TS}}$ from the procedure of Imbens and Kalyanaraman (2012). The legend for the plots on the right are the same as those for the expected efficiency curves. Note that the curves are not smooth because, to avoid redundancy, only points that corresponded to integer values of Δ were used.

6 Discussion

Owen and Varian (2020) found an efficiency advantage for the tie-breaker in a global regression, wherein the estimation variance decreased monotonically with the amount of experimentation. This paper provides a comparable finding for the now more standard local linear regression approach. For any fixed bandwidth h, we see a theoretical efficiency that increases with the amount Δ of experimentation. We have not investigated the effect of Δ on the subsequent choice of h.

Our theoretical analysis is for a uniformly spaced assignment variable. Imbens and Wager (2019) consider how to optimally tune kernel weights in a regression discontinuity problem to a given set of data. Owen and Varian (2020) consider numerical optimization of tie-breaker designs on given data.

Here we offer one explanation for why the empirical efficiencies on nonuniformly distributed data look so similar to the theoretical ones for uniformly distributed data. We use some results about non-parametric regression from Fan and Gijbels (1996, Table 2.1). Nonparametric regression estimates $\hat{\mu}(t)$ typically have an asymptotic variance where the leading term is proportional to $1/f(t)$ where f is the probability density of the x_i. This arises because the local sample size is asymptotically proportional to $f(t)$. Hence, when considering nonuniform distributions, the $1/f(t)$ factors in the leading order variance terms will cancel out when computing the efficiency ratios. Some of the nonparametric regression estimators, such as the Nadaraya-Watson estimator, have a lead term in their bias that depends on the derivative $f'(t)$ and while $f'(t) = 0$ for uniformly distributed data it is not zero in general. Kernel weighted least squares methods (with symmetric $K(\cdot)$) does not have a dependency on $f'(t)$ in its bias. There is a curvature bias from $\mu''(t)$ but that is not related to the sampling distribution of the x_i. The lead terms in bias and variance for kernel regressions do not distinguish between distributions with the same value of $f(t)$ but different $f'(t)$.

We close by comparing the tie-breaker setting to some other similar sounding ones. In fuzzy RDDs (Campbell, 1969) the threshold varies perhaps randomly because it depends on some additional variables that are not available to the data analyst. There are also settings where the assigment variable is subject to manipulation. For instance if a passing grade is 50% there may be no candidates with recorded scores in the interval from 47% to 50%. See McCrary (2008) for more on manipulation and Rosenman and Rajkumar (2019) for a mitigation strategy. The tie-breaker setting is special because the investigator is able to control the treatment allocation.

Acknowledgments

This work was supported by the U.S. National Science Foundation under grant IIS-1837931. We thank Hal Varian for commenting on the paper.

References

Aiken, L. S., West, S. G., Schwalm, D. E., Carroll, J. L., and Hsiung, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation: Efficacy of a university-level remedial writing program. *Evaluation Review*, 22(2):207–244.

Angrist, J., Hudson, S., and Pallais, A. (2014). Leveling up: Early results from a randomized evaluation of post-secondary aid. Technical report, National Bureau of Economic Research.

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

Angrist, J. D. and Lavy, V. (2009). Replication data for: Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement.

Angrist, J. D., Lavy, V., Leder-Luis, J., and Shany, A. (2019). Maimonides' rule redux. *American Economic Review: Insights*, 1(3):309–24.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering Metrics*. Princeton Univerity Press, Princeton.

Borchers, H. W. (2019). *pracma: Practical Numerical Math Functions*. R package version 2.2.9.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Campbell, D. T. (1969). Reforms as experiments. *American psychologist*, 24(4):409.

Cheng, M.-Y., Fan, J., and Marron, J. S. (1997). On automatic boundary corrections. *The Annals of Statistics*, 25(4):1691–1708.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, Boca Raton, FL.

Hahn, J., Todd, P., and der Klaauw, W. V. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Higham, N. J. (2002). *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, second edition.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79:933–959.

Imbens, G. and Wager, S. (2019). Optimized regression discontinuity designs. *Review of Economics and Statistics*, 101(2):264–278.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

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

Owen, A. B. and Varian, H. (2020). Optimizing the tie-breaker regression discontinuity design. *Electronic Journal of Statistics*, 14(2):4004–4027.

Rosenman, E. and Rajkumar, K. (2019). Optimized partial identification bounds for regression discontinuity designs with manipulation. Technical Report arXiv:1910.02170, Stanford University.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, 51(6):309.

Trochim, W. M. K. and Cappelleri, J. C. (1992). Cutoff assignment strategies for enhancing randomized clinical trials. *Controlled Clinical Trials*, pages 190–212.

Appendix: Proof of Proposition 3

We want to show that this function

$$\mathrm{Eff}_{\mathrm{TS}}(\delta) = \frac{2\bigl(3-2(1-3\delta^2+2\delta^3)^2\bigr)^2}{5-5(1-3\delta^2+2\delta^3)(1-6\delta^2+8\delta^3-3\delta^4)+2(1-3\delta^2+2\delta^3)^2}$$

has a positive derivative for $0<\delta<1$. The numerator has degree 12 and the denominator has degree 7. The customary formula for the derivative of a rational function produces a rational function with a non-negative denominator and a numerator of degree 18. We will work through a sequence of steps reducing the degree of this polynomial to show that the numerator must be positive on $(0,1)$. That then rigorously establishes the monotonicity of $\mathrm{Eff}_{\mathrm{TS}}(\delta)$ which is visually apparent.

It is convenient to work instead with $x=1-\delta$. Then $1-3\delta^2+2\delta^3=3x^2-2x^3$ and $1-6\delta^2+8\delta^3-3\delta^4=4x^3-3x^4$. Therefore $\mathrm{Eff}_{\mathrm{TS}}(\delta)=f(1-\delta)$ where f is a function given by

$$\begin{aligned} f(x) &= \frac{2(3-2(3x^2-2x^3)^2)^2}{5-5(3x^2-2x^3)(4x^3-3x^4)+2(3x^2-2x^3)^2}\\ &= \frac{2(3-2x^4(3-2x)^2)^2}{5-5x^5(3-2x)(4-3x)+2x^4(3-2x)^2}\\ &= \frac{2(3-2x^4(3-2x)^2)^2}{5+[x^4(3-2x)][6-24x+15x^2]}\\ &= \frac{2(3-2g(x)(3-2x))^2}{5+g(x)(6-24x+15x^2)} \end{aligned}$$

for $g(x)=x^4(3-2x)$ and having replaced δ by $x=1-\delta$ we will show that $f'(x)<0$.

In the usual formula for the derivative of a ratio, $f'(x)$ has this numerator

$$\begin{aligned} n_1(x) &= 4(3-2g(x)(3-2x))(4g(x)-2g'(x)(3-2x))[5+g(x)(6-24x+15x^2)]\\ &\quad -2(3-2g(x)(3-2x))^2[g'(x)(6-24x+15x^2)+g(x)(-24+30x)]. \end{aligned}$$

Notice that $0 \leqslant g(x)(3-2x) \leqslant 1$ for $x \in [0,1]$ and so $3-2g(x)(3-2x) > 0$. As a result, the sign of $n_1(x)$ is preserved by dividing it by $2(3-2g(x)(3-2x))$, yielding

$$\begin{aligned} n_2(x) = {} & 2(4g(x) - 2g'(x)(3-2x))[5 + g(x)(6 - 24x + 15x^2)] \\ & - (3 - 2g(x)(3-2x))[g'(x)(6 - 24x + 15x^2) + g(x)(-24 + 30x)]. \end{aligned}$$

Now since $g(x) = x^3[x(3-2x)]$ and $g'(x) = x^3[12 - 10x]$ and $x \in (0,1)$ we can divide $n_2(x)$ by $x^3/6$ finding that it has the same sign as

$$\begin{aligned} n_3(x) = {} & \frac{1}{3}\big(4x - 2(12-10x)\big)(3-2x)[5 + g(x)(6 - 24x + 15x^2)] \\ & - \frac{1}{6}\big(3 - 2g(x)(3-2x)\big)\big[(12-10x)(6-24x+15x^2) + x(3-2x)(-24+30x)\big] \\ = {} & -8(1-x)(3-2x)[5 + g(x)(6 - 24x + 15x^2)] \\ & - (3 - 2g(x)(3-2x))[-35x^3 + 93x^2 - 70x + 12] \\ = {} & -8(1-x)(3-2x)[5 + g(x)(6 - 24x + 15x^2)] \\ & - (3 - 2g(x)(3-2x))(1-x)(35x^2 - 58x + 12). \end{aligned}$$

We can divide $n_3(x)$ by $-(1-x)$ getting a polynomial $n_4(x)$ with the opposite sign from n_3. This yields

$$\begin{aligned} n_4(x) = {} & 8(3-2x)[5 + g(x)(6 - 24x + 15x^2)] + (3 - 2g(x)(3-2x))(35x^2 - 58x + 12) \\ = {} & 8(3-2x)[-30x^7 + 93x^6 - 84x^5 + 18x^4 + 5] \\ & + (-8x^6 + 24x^5 - 18x^4 + 3)(35x^2 - 58x + 12) \\ = {} & 8[60x^8 - 276x^7 + 447x^6 - 288x^5 + 54x^4 - 10x + 15] \\ & + (-280x^8 + 1304x^7 - 2118x^6 + 1332x^5 - 216x^4 + 105x^2 - 174x + 36) \\ = {} & 200x^8 - 904x^7 + 1458x^6 - 972x^5 + 216x^4 + 105x^2 - 254x + 156. \end{aligned}$$

Note that the coefficient of x^3 in n_4 is a zero that must not be left out when entering the coefficients into symbolic differentiation codes.

We want to show that $n_4(x) > 0$ for $x \in (0,1)$ which then makes $n_3(x) < 0$ and ultimately $f'(x) < 0$ so that $\mathrm{Eff}'_{\mathrm{TS}}(\delta) > 0$ for $\delta = 1 - x \in (0,1)$.

Our next step is to show that $n_4''(x) > 0$ for all $x \in [0,1]$. Note that

$$n_4''(x) = 11200x^6 - 37968x^5 + 43740x^4 - 19440x^3 + 2592x^2 + 210.$$

Graphing $n_4''(x)$ versus x and evaluating it numerically, makes it is clear that $148 < n_4''(x) < 369$. Our next few steps are just to eliminate even an unreasonable doubt about the sign of n_4''. Some readers might prefer to skip that and go instead to the subsection marked "Conclusion of the proof".

Positivity of n_4''

We begin by writing

$$n_4'''(x) = 48x(1400x^4 - 3955x^3 + 3645x^2 - 1215x + 108).$$

Then by the triangle inequality,

$$|n_4'''(x)| \leqslant 48(1400 + 3955 + 3645 + 1215 + 108) = 495504 \leqslant 2^{19} \tag{18}$$

for all $x \in [0, 1]$.

Now for $k \in \{0, 1, \ldots, 2^{12}\}$ define $x_k = k2^{-12}$. For each k we will let $\widehat{n_4''(x_k)}$ be the numerical evaluation for the polynomial $n_4''(x_k)$ computed using Horner's method with double-precision in R, implemented with the 'horner' function in the **pracma** R package of Borchers (2019).

By formula (5.3) in Higham (2002), the absolute error in Horner's method for the polynomial $\sum_{r=0}^{n} a_r x^r$ is at most $\gamma_{2n}\tilde{p}(|x|)$ where $\gamma_{2n} \equiv 2nu/(1 - 2nu)$, u is the unit roundoff, and $\tilde{p}(|x|) = \sum\limits_{r=0}^{n} |a_r||x|^r$.

Applying that bound to our 6th degree polynomial $n_4''(x)$, and noting that for each $x \in [0, 1]$, $\tilde{p}(|x|)$ will be at most the sum of the absolute value of the coefficients we find that

$$\max_{0 \leqslant k \leqslant 2^{12}} \left| n_4''(x_k) - \widehat{n_4''(x_k)} \right| \leqslant \gamma_{2\times 6}\tilde{p}(|x_k|) \leqslant \frac{12u}{1 - 12u} \times 115{,}150. \tag{19}$$

We need not worry about floating point error induced by evaluating x_k, because each x_k is a floating point number. For double precision in R, the unit roundoff is $u = 2^{-53} \leqslant 2 \times 10^{-15}$ from which $12u/(1-12u) \leqslant 10^{-13}$, and so the maximum error in (19) is at most 10^{-7}. The smallest value of $\widehat{n_4''(x_k)}$ among all $2^{12} + 1$ evaluation points x_k was 148.5743 and so $\min_{0 \leqslant k \leqslant 2^{12}} n_4''(x_k) \geqslant 148$.

Now we are ready to prove that $n_4''(x) > 0$ for all $x \in [0, 1]$. For any $x \in [0, 1]$ there exists $k_* \in \{0, 1, \ldots, 2^{12}\}$ with $|x - x_{k_*}| \leqslant 2^{-13}$. By (18) we know that n_4'' is Lipschitz continuous on $[0, 1]$ with Lipschitz constant 2^{19}. Therefore

$$n_4''(x) \geqslant n_4''(x_{k_*}) - |n_4''(x_{k_*}) - n_4''(x)| \geqslant 148 - 2^{19} \times 2^{-13} > 0,$$

holds for all $x \in [0, 1]$.

Conclusion of the proof

We have shown above that $n_4''(x) > 0$ for any $x \in [0, 1]$. Now since $n_4'(1) = -20$, and $n_4''(x) > 0$ for all $x \in [0, 1]$, it follows that $n_4'(x) < 0$ for all $x \in [0, 1]$. Finally since $n_4(1) = 5$ and $n_4'(x) < 0$ for all $x \in [0, 1]$, it follows that $n_4(x) > 0$ for all $x \in [0, 1]$. This completes the proof. □