DISCUSSION PAPER SERIES

Wild Bootstrap for Fuzzy Regression Discontinuity Designs: Obtaining Robust Bias-Corrected Confidence Intervals

Yang He
Otávio Bartalotti

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

Wild Bootstrap for Fuzzy Regression Discontinuity Designs: Obtaining Robust Bias-Corrected Confidence Intervals

Yang He
Microsoft Research Cambridge

Otávio Bartalotti
Iowa State University and IZA

NOVEMBER 2019

ABSTRACT

Wild Bootstrap for Fuzzy Regression Discontinuity Designs: Obtaining Robust Bias-Corrected Confidence Intervals

This paper develops a novel wild bootstrap procedure to construct robust bias-corrected (RBC) valid confidence intervals (CIs) for fuzzy regression discontinuity designs, providing an intuitive complement to existing RBC methods. The CIs generated by this procedure are valid under conditions similar to the procedures proposed by Calonico et al. (2014) and related literature. Simulations provide evidence that this new method is at least as accurate as the plug-in analytical corrections when applied to a variety of data generating processes featuring endogeneity and clustering. Finally, we demonstrate its empirical relevance by revisiting Angrist and Lavy (1999) analysis of class size on student outcomes.

Corresponding author:
Otávio Bartalotti
260 Heady Hall
Department of Economics
Iowa State University
Ames, IA, 50011
USA

E-mail: bartalot@iastate.edu

1   Introduction

Regression discontinuity (RD) designs are one of the leading empirical approaches in economics, political science, and public policy evaluation, being extensively used to estimate the causal effects of treatments or policies under transparent assumptions.[1] The identification in RD designs exploits the fact that many policies and programs use a threshold based on a score, also called a "running variable," to assign treatment to individuals or firms. If the researcher credibly believes that subject's position relative to the threshold is not related to unobserved characteristics driving the outcome of interest, we can attribute the differences between units slightly above and below the cutoff as caused by treatment alone. When the running variable does not entirely determine the treatment, there are both treated and untreated units on each side of the cutoff, a situation referred to as the fuzzy RD design. Directly comparing the outcomes on both sides of the cutoff results in an intent-to-treat effect, and the average treatment effect at the cutoff can be recovered by taking the ratio of difference in outcomes and difference in treatment probabilities at the threshold, as in a Wald formulation of the treatment effect in the instrumental variable setting. Even when units are self-selected to treatment based on anticipated gains, Hahn et al. (2001) show that this ratio can be interpreted as the local average treatment effect (LATE) under proper assumptions.

The identification of RD designs occurs exactly at the cutoff, and in practice the treatment effect is typically estimated by fitting local linear models above and below the threshold, which are extrapolated to the exact point of discontinuity.[2][3] The choice of the bandwidth (h) in these nonparametric estimators is an important econometric issue, determining the trade-off between bias and variance. One popular bandwidth selector proposed by Imbens and Kalyanaraman (2012) minimizes the asymptotic mean squared error (AMSE) of the treatment effect estimator.[4]

In an influential paper, Calonico et al. (2014), henceforth CCT, show that the AMSE-optimal bandwidth shrinks slowly enough that the leading bias term in the local polynomial estimators will be non-negligible, affecting the asymptotic distribution of the estimator. Consequently, the usual confidence intervals (CIs) for the RD treatment effects are invalid, and have empirical coverage well below their nominal levels. Calonico et al. (2014) propose a solution to this problem by obtaining a valid estimate of the leading bias-term and re-centering the conventional point estimator. Furthermore, the additional variability introduced by the bias estimation needs to be considered when constructing CIs. This approach is referred to as the robust bias corrected (RBC) inference method, and results in an asymptotically normal bias-corrected point estimator under weaker assumptions on the bandwidth shrinkage rates. CIs based on this method are valid even when AMSE optimal bandwidths are used.

In this paper, we contribute to this growing RBC literature by proposing a wild bootstrap procedure as an alternative to the plug-in analytical RBC inference methods for fuzzy RD designs. The new bootstrap procedure is asymptotically equivalent to CCT's and simulations demonstrate it performs well in finite samples.

Recent studies have further developed the original idea of CCT, establishing RBC inference as the standard method in the RD literature. Calonico et al. (2018a) and Calonico et al. (2019a) develop valid coverage error expansions for RBC CIs for general nonparametric inference and the RD design specific case, respectively. They show that RBC CIs achieve higher-order refinements in terms of coverage error, outperforming intervals produced by undersmoothing, while

---

[1]Imbens and Lemieux (2008) and Lee and Lemieux (2010) provide reviews of this literature with many examples.

[2]See Fan (1992); Hahn et al. (2001) for a detailed discussion on local polynomial estimator properties and its use in RD designs.

[3]The following discussion is similar to the description in Bartalotti et al. (2017).

[4]This bandwidth selector is $h = O_p(n^{-1/5})$, where n is the number of observations.

preserving robustness to the choice of bandwidth parameter. The expansions are then used to develop new coverage-error optimal bandwidths for forming CIs. In current work, Calonico et al. (2019b) characterizes CIs that obtain minimum "worst-case" coverage error uniformly over plausible distributions of the data, providing bandwidth and kernel selectors that achieve minimax coverage error and interval length optimal CIs. Finally, Cattaneo et al. (Forthcoming) have extended the RBC framework for partitioning-based series estimators.

We propose a wild bootstrap procedure that builds upon the RBC literature's insight by resampling from higher order local polynomials. In particular, the local linear models are estimated as usual for both outcome and treatment, resulting in a conventional biased estimator. To estimate the bias, additional local quadratic models are used, and the potentially correlated residuals on both the outcome and treatment equations serve as the "true" data generating process (DGP) for the bootstrap. The bias of the conventional estimator is therefore known under this bootstrap DGP and can be calculated by averaging the error of the linear model's estimates across bootstrap replications. Any remaining bias converges to zero at a faster rate, allowing the bias of the local linear model to be estimated. This approach is described in Algorithm 3.1 and the resulting bias-corrected estimator is shown to be asymptotically normal with mean zero in Theorem 3.1.

To account for the additional variability introduced by the bias correction, we propose an iterated bootstrap procedure: generate many bootstrap datasets from local quadratic models and calculate the bias-corrected estimate for each of them. The resulting empirical distribution of bias-corrected estimator is then used to construct CIs. This procedure is in line with RBC literature, where the variance of estimated bias term and the covariance between estimated bias and original point estimator are derived analytically. This complex adjustment to the original variance is automatically embedded in the iterated bootstrap. The bootstrap implementation is described in Algorithm 3.2, and the resulting CIs are shown to be asymptotically valid in Theorem 3.2.

Compared to existing implementations of the RBC approach in the literature, the proposed procedure is straightforward and does not require intensive derivations. This bootstrap complements the methods readily available in statistical software described in Calonico et al. (2017), and could naturally serve as an alternative.

Moreover, the results in this paper provide an additional tool for practitioners, which can be particularly useful in situations for which the analytical results or the statistical software code has not been developed or are particularly cumbersome. Some cases include more complex RD designs with objects of interest involving multiple cutoffs and nonlinear functions of discontinuity parameters. For a practical example, Bertanha and Imbens (2019) evaluate non-linear functions involving four discontinuity parameters estimated jointly, requiring the researcher to evaluate the bias and covariances of all parameters, while relying on linearizations and the delta method to obtain analytical formulas. In this particular case, the bootstrap method we propose is very practical compared to the need to derive formulas for RBC CIs.

Additionally, since the bootstrap mimics the true data generating process, it has the flexibility to accommodate dependent (clustered) data by adjusting the resampling algorithm accordingly. We demonstrate how the bootstrap procedure can be applied to such clustered data. This flexibility and adaptability to the particular needs of different empirical objectives is the main strength of the bootstrap proposed.

Compared to Bartalotti et al. (2017), which developed a similar iterated bootstrap procedure for robust inference in special case of sharp RD designs, the current paper provides important generalizations in several dimensions. First, it adapts bootstrapping IV models to fuzzy RD designs. Second, its validity is extended and theoretically proved to general local polynomials and higher order of derivatives of interest, which could be used in the context of "Kink" RD designs, for example. Last, its flexibility and capability to accommodate clustered data is

confirmed by simulation studies.

While we focus on the RD case, as it is widely used and most relevant for practitioners in the social sciences, a similar iterated wild bootstrap strategy could be used to successfully implement RBC methods in other relevant settings. One such case is nonparametric local estimation as described in Calonico et al. (2018a), and our paper can serve as a step in that direction. Naturally, exploring if the wild bootstrap achieves the higher order refinements described by Calonico et al. (2018a) when evaluating plug-in RBC methods also offers an interesting path for future research, which is beyond our scope.

Concomitantly and independently, Chiang et al. (2019) proposed a multiplier bootstrap procedure for fuzzy RD and many related settings based on a Bahadur representation of a general class of Wald estimators. Their paper, however, has a more specific focus on uniform inference for quantile treatment effects in RD designs. Both the procedure and proofs in that paper differ from the ones proposed here and could potentially serve as alternatives in the cases covered by both approaches, keeping in mind that their approach offers robust uniform inference.[5]

In a related work, Cattaneo et al. (Forthcoming) independently consider a single-step wild bootstrap for the RBC t-statistics of interest in the context of nonparametric regression using series estimators, and prove the validity of the wild bootstrap in that setting. These results are closely related to ours, and underscore the relevance and flexibility of wild bootstrap to perform RBC inference in nonparametric regression for different estimation approaches. Furthermore, the bootstrap in that case is shown to provide robust uniformly valid confidence bands.

The results on Cattaneo et al. (Forthcoming) and Chiang et al. (2019) suggest that a iterated bootstrap like the one proposed here might achieve uniform coverage, even though a formal analysis of that possibility is beyond the scope of this paper. Nevertheless, our point-wise inference procedure benefits from its very intuitive nature, easy implementation and flexibility as exemplified in Section 5 when dealing with dependent (clustered) data, and multiple discontinuities or functions of parameters. In Section 4, we present useful comparisons in performance by the proposed wild bootstrap, the multiplier bootstrap in Chiang et al. (2019), and the plug-in RBC implementations.

Recently, honest CIs were proposed as an alternative approach to RBC inference, notably Armstrong and Kolesár (2018) and Armstrong and Kolesár (Forthcoming). Honest CIs guarantee good coverage properties uniformly over all possible functions (within a family) for the conditional expectation of the outcome. The class of functions considered is determined by their smoothness, which is assumed to be bounded, and the CIs are built considering the impact of the non-random worst-case bias in estimation in that class of functions. In closely related work, Imbens and Wager (2019) examine the RD estimator as a linear combination of the observed outcomes and obtains weights for each observation by numerical optimization. The minimax linear estimator minimizes the variance and worst-case bias for a similarly defined class of functions.

This approach offers an attractive alternative to RBC that can be useful in many circumstances, especially when the researcher cannot rely on local arguments (bandwidth shrinking) to control misspecification of the outcome's conditional expectation function close to the cutoff, for example when the running variable is discrete, see (Kolesár and Rothe, 2018), or measured with error as in (Bartalotti et al., 2019).

Advancing this literature is beyond the scope of this paper, as we do not base the bootstrap procedure on obtaining worst bias characterizations. Nevertheless, a bootstrap procedure that would achieve the properties of the CIs described in Armstrong and Kolesár (2018) or Imbens

---

[5]Section 4 presents a comparison between our proposed wild bootstrap and the multiplier bootstrap in Chiang et al. (2019).

and Wager (2019) would be an interesting topic of future research that can build upon the developments presented here.

The paper is organized as follows. Section 2 describes the basic fuzzy RD approach, its usual implementation, and the basic RBC inference approach. Section 3 presents the proposed bootstrap procedures to estimate bias and construct confidence intervals. Asymptotic properties are discussed and summarized in two theorems. Section 4 provides simulation evidence that the bootstrap procedure effectively reduces bias and generates valid CIs. Implementation of the bootstrap to clustered data is discussed in Section 5. Section 6 demonstrates the applied relevance of this bootstrap procedure by applying it to the scholastic achievement data used by Angrist and Lavy (1999). Section 7 provides a brief conclusion.

2 Background

This section provides additional details of identification assumptions and estimation methods in fuzzy RD designs. It also briefly introduces the RBC approach proposed by CCT and subsequent papers. Notations defined in this and following sections are consistent with CCT.

In a typical fuzzy RD setting, researchers are interested in the local causal effect of treatment at a given cutoff. For any unit i, (X_i, T_i, Y_i) is observed, where X_i is a continuous running variable which determines treatment assignment, T_i is a binary variable which indicates actual treatment status and Y_i is the outcome. In sharp RD designs, the treatment actually received is the same as the assigned treatment, i.e., $T_i = \mathbb{1}(X_i \geq c)$, with c being the cutoff. In fuzzy RD designs, however, the received treatment is not a deterministic function of running variable X_i. Instead, the probability $\Pr(T_i = 1 \mid X_i)$ is between zero and one in both sides but experiences a sudden change at the cutoff. Without loss of generality, the cutoff c can be reset to zero. If assigned to treatment ($X_i \geq 0$), unit i's actual treatment status and outcome are represented by functions $T_i(1)$ and $Y_i(1)$, otherwise $T_i(0)$ and $Y_i(0)$. Thus the observed treatment status and outcome are

$$T_i = T_i(0)\,\mathbb{1}(X_i < 0) + T_i(1)\,\mathbb{1}(X_i \geq 0)$$
$$Y_i = Y_i(0)\,\mathbb{1}(X_i < 0) + Y_i(1)\,\mathbb{1}(X_i \geq 0).$$

For each unit i's outcome, either $Y_i(0)$ or $Y_i(1)$ is observed. The data itself is uninformative in terms of treatment effect because the counterfactual outcome could be arbitrary. However, under continuity and smoothness conditions on $T_i(0), Y_i(0), T_i(1)$ and $Y_i(1)$ around the cutoff $X_i = 0$, it is possible to identify the treatment effect for units just at the cutoff and the estimand of interest is

$$\zeta = \frac{\tau_Y}{\tau_T} = \frac{\mathrm{E}[Y_i(1) \mid X_i = 0] - \mathrm{E}[Y_i(0) \mid X_i = 0]}{\mathrm{E}[T_i(1) \mid X_i = 0] - \mathrm{E}[T_i(0) \mid X_i = 0]}, \tag{2.1}$$

where the symbol E represents the expectation and τ_Y and τ_T represent the sharp RD estimators, i.e., difference in expectations at the cutoff. Intuitively, this is a Wald estimator in the limit where the assigned treatment serves as an instrument. The reduced-form difference in expected outcome, τ_Y, reveals the "intent-to-treat" (ITT) effect. The treatment effect is recovered by dividing ITT effect by the difference in treatment probabilities. When the treatment effect is not constant across units, ζ should be interpreted with caution. If treatment status is independent of treatment effects at the cutoff, ζ is the average treatment effect (ATE) at the cutoff. This assumption rules out self-selection based on anticipated gain. Hahn et al. (2001) show that under a less restrictive assumption that the treatment effect and status are jointly independent of the running variable around the cutoff, and a local weak monotonicity

assumption, the LATE is identified.[6]

Equation 2.1 presents ζ as a ratio of two sharp RD estimators. Due to this symmetry, we use "Z" as a placeholder for either outcome variable Y or treatment variable T to ease the notation. In addition, denote the conditional expectations $\mu_{Z+}(x)$ and $\mu_{Z-}(x)$, conditional variances $\sigma^2_{Z+}(x)$ and $\sigma^2_{Z-}(x)$, the η-th order derivative of conditional expectations $\mu^{(\eta)}_{Z+}(x)$ and $\mu^{(\eta)}_{Z-}(x)$ and their limits. Formally, they are defined as

$$\mu_{Z+}(x) = \mathrm{E}[Z_i(1) \mid X_i = x] \qquad\qquad \mu_{Z-}(x) = \mathrm{E}[Z_i(0) \mid X_i = x]$$
$$\sigma^2_{Z+}(x) = \mathbb{V}[Z_i(1) \mid X_i = x] \qquad\qquad \sigma^2_{Z-}(x) = \mathbb{V}[Z_i(0) \mid X_i = x]$$
$$\mu^{(\eta)}_{Z+}(x) = \frac{d^\eta \mu_{Z+}(x)}{dx^\eta} \qquad\qquad \mu^{(\eta)}_{Z-}(x) = \frac{d^\eta \mu_{Z-}(x)}{dx^\eta}$$
$$\mu^{(\eta)}_{Z+} = \lim_{x \to 0^+} \mu^{(\eta)}_{Z+}(x) \qquad\qquad \mu^{(\eta)}_{Z-} = \lim_{x \to 0^-} \mu^{(\eta)}_{Z-}(x)$$

where the symbol $\mathbb{V}(\cdot)$ represents variance. The treatment effect ζ is nonparametrically estimable because μ_{Z-} and μ_{Z+} can be estimated consistently under Assumption 2.1, which lists standard conditions in the fuzzy RD literature.[7] (See, in particular, Hahn et al., 2001, Porter, 2003 and CCT.)

Assumption 2.1 (Behaviour of the DGP near the cutoff) *The random variables $\{X_i, T_i, Y_i\}^n_{i=1}$ form a random sample of size n. There exists a positive number κ_0 such that the following conditions hold for all x in the neighbourhood $(-\kappa_0, \ \kappa_0)$ around zero: (a) The density of X_i is continuous and bounded away from zero at x; (b) $\mathrm{E}[Z_i^4 \mid X_i = x]$ is bounded; (c) $\mu_{Z-}(x)$ and $\mu_{Z+}(x)$ are three times continuously differentiable; (d) $\sigma^2_{Z-}(x)$ and $\sigma^2_{Z+}(x)$ are continuous and bounded away from zero; (e) $\mu_{T-}(0) \neq \mu_{T+}(0)$.*

Assumption 2.1(a) ensures that the number of data points arbitrarily close to the cutoff increases as the sample size grows. Part (c) imposes necessary smoothness condition to allow an approximation by second order polynomials. Parts (b) and (d) put standard restrictions on moments to ensure that the estimated local polynomials are well behaved. Part (e) requires that the treatment assignment as an instrument is valid, in the sense that it induces a first stage difference in treatment probability. Furthermore, the usual LATE identification and interpretation for the treatment effect in FRD requires weak monotonicity of treatment with respect to the running variable, as described in Hahn et al. (2001) Assumption (A3).

In practice, local polynomial regression is widely used to estimate RD designs due to its boundary properties.[8] As an illustration, consider the local linear regression using kernel function $K(\cdot)$ with a common bandwidth, h, used for both the outcome and the treatment at both sides of the cutoff. The estimated treatment effect is

$$\hat{\zeta}(h) = \frac{\hat{\tau}_Y(h)}{\hat{\tau}_T(h)} = \frac{\hat{\mu}_{Y+}(h) - \hat{\mu}_{Y-}(h)}{\hat{\mu}_{T+}(h) - \hat{\mu}_{T-}(h)}, \tag{2.2}$$

---

[6]In Hahn et al. (2001) this is described in Assumption (A3) and is required for the identification and interpretation of the estimand as a LATE.

[7]Throughout the main text we focus on the case where the researcher implements a local linear model to estimate τ_Z and a quadratic model to approximate the bias term. The proofs presented in the online appendix for the validity of the bootstraps proposed include the general case in which higher-order polynomials can be used to obtain τ_Z or a higher-order bias correction is implemented, e.g., Bartalotti (2018).

[8]See Fan and Gijbels (1996) for discussions on the boundary properties of local polynomial regression. See Gelman and Imbens (2018) for discussions on the choices of global and local polynomial regression and its order.

with

$$\hat{\mu}_{Z+}(h) = \arg\min_{\beta_0} \min_{\beta_1} \sum_{i=1}^{n} \mathbb{1}\{X_i \geq 0\}(Z_i - \beta_0 - X_i\beta_1)^2 \frac{1}{h} K\left(\frac{X_i}{h}\right)$$

$$\hat{\mu}_{Z-}(h) = \arg\min_{\beta_0} \min_{\beta_1} \sum_{i=1}^{n} \mathbb{1}\{X_i < 0\}(Z_i - \beta_0 - X_i\beta_1)^2 \frac{1}{h} K\left(\frac{X_i}{h}\right).$$

The conditional expectations μ_{Z+} and μ_{Z-} are consistently estimated by $\hat{\mu}_{Z+}(h)$ and $\hat{\mu}_{Z+}(h)$ when $h \to 0$.[9] The asymptotic distribution of the quotient estimator $\frac{\hat{\tau}_Y(h)}{\hat{\tau}_T(h)}$ can be derived by applying the delta method. Let V_Z be the asymptotic variance of $\hat{\tau}_Z(h)$ and C_{YT} be the asymptotic covariance between $\hat{\tau}_Y(h)$ and $\hat{\tau}_T(h)$, i.e.,

$$\left(\begin{array}{c} \sqrt{nh}(\hat{\tau}_Y(h) - \tau_Y) \\ \sqrt{nh}(\hat{\tau}_T(h) - \tau_T) \end{array} \right) \xrightarrow{d} N\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} V_Y & C_{YT} \\ C_{YT} & V_T \end{array} \right)\right),$$

$$\sqrt{nh}(\hat{\zeta}(h) - \zeta) \xrightarrow{d} N\left(0, \frac{1}{\tau_T^2} V_Y - \frac{2\tau_Y}{\tau_T^3} C_{YT} + \frac{\tau_Y^2}{\tau_T^4} V_T\right).$$

Let $V(h) = \mathbb{V}(\hat{\zeta}(h) \mid X_1, ..., X_n)$, then $\frac{\hat{\zeta}(h) - \zeta}{\sqrt{V(h)}} \xrightarrow{d} N(0, 1)$ and the CIs can be constructed as

$$\hat{\zeta}(h) \pm q_{1-\alpha/2} V(h)^{1/2} \tag{2.3}$$

where $q_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

The above asymptotic distribution is valid only when bandwidth h shrinks fast enough such that the bias of $\hat{\zeta}_Z(h)$ is negligible relative to $\sqrt{V(h)}$. Formally, $h = o_p(n^{-1/5})$ is required. With a bandwidth of order $O_p(n^{-1/5})$, Hahn et al. (2001) show that the asymptotic distribution is normal but not centered at zero. Using (2.3) to construct CIs without considering this first-order bias in distributional approximation leads to coverage rates that differ from the nominal level. Imbens and Kalyanaraman (2012) develop plug-in bandwidth selector for RD estimators, which is optimal in the sense that AMSE of the point estimator is minimized.

Two different approaches are commonly adopted in empirical studies. One is undersmoothing. In this case, instead of using the AMSE-optimal bandwidth, researchers may want to choose a smaller bandwidth in order to meet the requirement of $h = o_p(n^{-1/5})$. However, this often leads to a series of ad-hoc bandwidths without theoretical basis. Another approach is bias correction, in which the leading bias is consistently estimated to remove the distortion of the asymptotic approximation. However, this approach does not perform well initially because the estimated bias introduces additional variability. The RBC approach is based on bias correction, but derives an alternative asymptotic variance component for normalization so that the additional variability is accounted for.

For any bandwidth $h \to 0$, the first-order bias of fuzzy RD estimator from local linear regression is

$$\mathrm{E}[\hat{\zeta}(h) \mid X_1, ..., X_n] - \zeta = h^2 \left(\frac{1}{\tau_T} \boldsymbol{B}_Y(h) - \frac{\tau_Y}{\tau_T^2} \boldsymbol{B}_T(h)\right)(1 + o_p(1)), \tag{2.4}$$

with

$$\boldsymbol{B}_Z(h) = \frac{\mu_{Z+}^{(2)}}{2} \mathfrak{B}_+(h) - \frac{\mu_{Z-}^{(2)}}{2} \mathfrak{B}_-(h).$$

---

[9]Unless otherwise stated, all limits in this paper are assumed to hold as $n \to \infty$.

The terms $\mathfrak{B}_+(h)$ and $\mathfrak{B}_-(h)$, explicitly defined in online appendix, are observed quantities that depend on the kernel, bandwidth and running variable. To explicitly calculate the first-order bias, one needs to estimate τ_Z, $\mu_{Z+}^{(2)}$ and $\mu_{Z-}^{(2)}$. Among them τ_Z is consistently estimated by the local linear regression with bandwidth h. CCT propose a local second-order regression with a (potentially) different bandwidth, b, to estimate the second order derivatives $\mu_{Z+}^{(2)}$ and $\mu_{Z-}^{(2)}$. This produces the bias-corrected estimator

$$\hat{\zeta}^{bc}(h,b) = \hat{\zeta}(h) - \Delta(h,b),$$

with

$$\Delta(h,b) = h^2 \Big(\frac{1}{\hat{\tau}_T(h)} \hat{\boldsymbol{B}}_Y(h,b) - \frac{\hat{\tau}_Y(h)}{\hat{\tau}_T^2(h)} \hat{\boldsymbol{B}}_T(h,b) \Big),$$

$$\hat{\boldsymbol{B}}_Z(h,b) = \frac{\hat{\mu}_{Z+}^{(2)}(b)}{2} \mathfrak{B}_+(h) - \frac{\hat{\mu}_{Z-}^{(2)}(b)}{2} \mathfrak{B}_-(h).$$

Notice that the bias $\Delta(h,b)$ is estimated with uncertainty. As a result, the variance of bias-corrected estimator $\hat{\zeta}^{bc}(h,b)$ is different from $V(h)$. CCT propose a new formula for the variance of bias-corrected estimator and use it for normalization:

$$\frac{\hat{\zeta}^{bc}(h,b) - \zeta}{V^{bc}(h,b)^{1/2}} \xrightarrow{d} N(0,1), \tag{2.5}$$

where $V^{bc}(h,b) = V(h) + C(h,b)$ and $C(h,b)$ captures the adjustment to variance introduced by the bias-correction term. This distributional approximation is valid even when $h = O_p(n^{-1/5})$, as long as certain conditions on h and b are satisfied. Assumption 2.2 specifies the bandwidth and kernel conditions assumed by CCT, which are also used in this paper.

Assumption 2.2 (Bandwidth and kernel) *Let h be the bandwidth used to estimate the local linear model and let b be the bandwidth used to estimate the local quadratic model used to estimate the bias correction. Then $nh \to \infty$, $nb \to \infty$, and $n \times \min(h,b)^5 \times \max(h,b)^2 \to 0$ as $n \to \infty$. The kernel function $K(\cdot)$ is positive, bounded, and continuous on the interval $[-\kappa, \kappa]$ and zero outside that interval for some $\kappa > 0$.*

Assumption 2.2 does not require $nh^{1/5} \to 0$. Instead, it only requires that $nh^{1/5}b^{1/2} \to 0$ when $h < b$ or $nb^{1/5}h^{1/2} \to 0$ when $h > b$. This assumption also allows for the vast majority kernel functions commonly used in practice.

To simplify notation, let $m = \min(h,b)$ and define the scaled and truncated kernel functions

$$K_{+,h}(x) = \frac{1}{h} K(x/h) \,\mathbb{1}\{x \geq 0\}, K_{-,h}(x) = \frac{1}{h} K(x/h) \,\mathbb{1}\{x < 0\},$$

$$K_{+,b}(x) = \frac{1}{b} K(x/b) \,\mathbb{1}\{x \geq 0\}, K_{-,b}(x) = \frac{1}{b} K(x/b) \,\mathbb{1}\{x < 0\}.$$

Recent studies have further established RBC inference as the standard method in the RD literature. Calonico et al. (2018a) develop valid coverage error expansions for RBC CIs for general nonparametric inference, while Calonico et al. (2019a) obtains similar results in the context of RD designs. They show that RBC CIs achieve higher-order refinements in terms of coverage error, outperforming intervals produced by undersmoothing, while preserving robustness to the choice of bandwidth parameter.

Furthermore, Calonico et al. (2019a) provides coverage-error optimal bandwidths for forming CIs that we will exploit in the bootstrap implementation described in Section 4, where a simple bootstrap procedure is proposed to construct CIs based on the insight provided these recent advances in the RBC literature. This bootstrap procedure is straightforward in the sense that no derivation of analytical formulas for the bias, variance and covariance terms is required. The bias-corrected estimator and the associated CIs are numerically different from the analytical alternative but asymptotically equivalent.

3   Bootstrap Algorithm and Validity

In this section, two bootstrap algorithms are proposed to obtain the RBC point estimator and CIs for the fuzzy RD design, extending the results in Bartalotti et al. (2017). Their validity is justified in two theorems and proved in the online appendix. The idea behind both algorithms is to use higher-order local polynomials to approximate the joint behaviour of (X_i, T_i, Y_i) around the cutoff. These polynomials, together with the empirical variance structure, serve as the "true" DGP in the bootstrap under which we evaluate the bias of the local linear estimator employed by the researcher when implementing RD. Assumption 2.2 guarantees that the estimated "true" DGP is close to the unknown DGP in the sense that the distributional approximation derived from the "true" DGP is asymptotically valid. This can be best illustrated from the special case where the bandwidths used for estimating τ and the bias are the same, $h = b$, which translates to the bandwidth condition $nb^7 \to 0$ under Assumption 2.2. By the same argument that $h = o_p(n^{-1/5})$ is required for valid inference in a RD design estimated by local linear regression, $b = o_p(n^{-1/7})$ is required in a RD design estimated by local quadratic regression, see CCT, Calonico et al. (2018a) and Calonico et al. (2019a) for in-depth discussions.

In practice, the preliminary bandwidth, b, used to estimate the bias and the main bandwidth for the point estimate of the treatment effect, h, need to be chosen. A natural choice in the RBC setting would be to implement the coverage error optimal bandwidths for b and h proposed by Calonico et al. (2019a), which is specifically tailored for RBC inference in RD designs. Alternatively, MSE-optimal bandwidths which are also valid (but sub-optimal) in terms of coverage error could be used, as discussed in that paper. Both bandwidth selectors are available in the statistical package *rdrobust* for STATA and R described in Calonico et al. (2015, 2017).[10]

Algorithm 3.1 consistently estimates the bias term.

Algorithm 3.1 (Bias estimation) Assume h and b are bandwidths as defined by Assumption 2.2.

STEP *1.* Using the preliminary bandwidth, b, estimate local second order polynomials $\hat{g}_{Z-}$ and $\hat{g}_{Z+}$ with least squares using $K_{-,b}$ and $K_{+,b}$ for weights:

$$\hat{g}_{Z-}(x) = \hat{\beta}_{Z-,0} + \hat{\beta}_{Z-,1}x + \hat{\beta}_{Z-,2}x^2, \qquad \hat{g}_{Z+}(x) = \hat{\beta}_{Z+,0} + \hat{\beta}_{Z+,1}x + \hat{\beta}_{Z+,2}x^2$$

with

$$(\hat{\beta}_{Z-,0}, \hat{\beta}_{Z-,1}, \hat{\beta}_{Z-,2})' = \underset{\beta_0,\beta_1,\beta_2}{\arg\min} \sum_{i=1}^{n} (Z_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2)^2 K_{-,b}(X_i)$$

$$(\hat{\beta}_{Z+,0}, \hat{\beta}_{Z+,1}, \hat{\beta}_{Z+,2})' = \underset{\beta_0,\beta_1,\beta_2}{\arg\min} \sum_{i=1}^{n} (Z_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2)^2 K_{+,b}(X_i).$$

Let

$$\hat{g}_Z(x) = \begin{cases} \hat{g}_{Z-}(x) & \text{if } x < 0 \\ \hat{g}_{Z+}(x) & \text{otherwise} \end{cases}$$

and calculate the residuals $\hat{\varepsilon}_{Zi} = Z_i - \hat{g}_Z(X_i)$ for all i.

STEP *2.* Repeat the following steps B_1 times to produce the bootstrap estimates $\hat{\tau}_{Z,1}^*(h), \ldots, \hat{\tau}_{Z,B_1}^*(h)$. For the kth replication:

---

[10]A package for R called *frdboot* implements the procedures described below and is available at: `https://github.com/yhe0802/FRD-bootstrap/blob/master/frdboot_0.1.0.zip`

2.1. Draw i.i.d. random variables e_i^* with mean zero, variance one, and bounded fourth moments independent of the original data and construct

$$\varepsilon_{Zi}^* = \hat{\varepsilon}_{Zi} e_i^*,$$
$$Z_i^* = \hat{g}_Z(X_i) + \varepsilon_{Zi}^*$$

for all i. Note that both $\hat{\varepsilon}_{Yi}$ and $\hat{\varepsilon}_{Ti}$ are multiplied by the same e_i^*.

2.2. Using the main bandwidth, h, calculate $\hat{\mu}_{Z+}^*(h)$ and $\hat{\mu}_{Z-}^*(h)$ by estimating the local linear model on the bootstrap data set using $K_{+,h}$ and $K_{-,h}$ for weights:

$$\hat{\mu}_{Z-}^*(h) = \arg\min_{\mu} \min_{\beta} \sum_{i=1}^{n} (Z_i^* - \mu - \beta X_i)^2 K_{-,h}(X_i)$$

$$\hat{\mu}_{Z+}^*(h) = \arg\min_{\mu} \min_{\beta} \sum_{i=1}^{n} (Z_i^* - \mu - \beta X_i)^2 K_{+,h}(X_i).$$

2.3. Save $\hat{\zeta}_k^*(h) = \frac{\hat{\mu}_{Y+}^*(h) - \hat{\mu}_{Y-}^*(h)}{\hat{\mu}_{T+}^*(h) - \hat{\mu}_{T-}^*(h)}$.

STEP *3.* Estimate the bias as

$$\Delta^*(h,b) = \frac{1}{B_1} \sum_{k=1}^{B_1} \hat{\zeta}_k^*(h) - \frac{\hat{g}_{Y+}(0) - \hat{g}_{Y-}(0)}{\hat{g}_{T+}(0) - \hat{g}_{T-}(0)}. \tag{3.1}$$

Algorithm 3.1 consists of three steps. The first step estimates the bootstrap DGP, which is captured by second order local polynomials. The second step creates a series of new samples through wild bootstrap and finds the traditional fuzzy RD estimate for each sample. Crucial for the procedure is that pairs of residuals are multiplied by the same realization of random number e^* to preserve the correlation between Y_i and T_i. The last step calculates the bias from local linear estimator in the bootstrap by definition. Under Assumptions 2.1, 2.2 and B_1 large enough, the procedure described by Algorithm 3.1 consistently estimates the bias component, resulting in a bias-corrected estimator that has the same asymptotic distribution as in Equation (2.5). This conclusion is formally given in Theorem 3.1.

Theorem 3.1 *Under Assumptions 2.1 and 2.2, as n and B_1 go to infinity,*

$$\frac{\hat{\zeta}(h) - \Delta^*(h,b) - \zeta}{V^{bc}(h,b)^{1/2}} \to^d N(0,1), \tag{3.2}$$

where $\Delta^*(h,b)$ is defined by Equation (3.1).

Theorem 3.1 enables one to construct valid confidence interval in the form of $\hat{\zeta}(h) - \Delta^*(h,b) \pm V^{bc}(h,b)^{1/2}$. However, the term $V^{bc}(h,b)$ still needs to be calculated. The second algorithm circumvents the analytical derivation of $V^{bc}(h,b)$ through an iterated bootstrap. In particular, the first layer bootstrap is designed to mimic the randomness due to sampling error and the second layer bootstrap, as described in Algorithm 3.1, is designed to estimate bias due to model misspecification. The additional variability introduced by the bias correction term will be automatically accounted for by this iterated bootstrap. The detailed procedure is given in Algorithm 3.2.

Algorithm 3.2 (Distribution) Assume h and b are bandwidths as defined by Assumption 2.2 and Algorithm 3.1.

STEP *1.* Using the preliminary bandwidth, b, estimate $\hat{g}_{Z+}$ and $\hat{g}_{Z-}$ and generate $\hat{g}_Z(\cdot)$ and the residuals $\hat{\varepsilon}_{Zi}$ just as in Algorithm 3.1.

STEP 2. Repeat the following steps B_2 times to produce bootstrap estimates of the bias-corrected estimate. For the kth replication:

    2.1. Draw i.i.d. random variables e_i^* with mean zero, variance one, and bounded fourth moments independent of the original data and construct

$$\begin{aligned}\varepsilon_{Zi}^* &= \hat{\varepsilon}_{Zi}e_i^*,\\ Z_i^* &= \hat{g}_Z(X_i) + \varepsilon_{Zi}^*.\end{aligned}$$

    for all $i = 1, \ldots, n$. Note that both $\hat{\varepsilon}_{Yi}$ and $\hat{\varepsilon}_{Ti}$ are multiplied by the same e_i^*.

    2.2. Using the main bandwidth, h, calculate $\hat{\mu}_{Z+}^*(h)$ and $\hat{\mu}_{Z-}^*(h)$ by estimating the local linear model on the bootstrap data set using $K_{+,h}$ and $K_{-,h}$ for weights:

$$\hat{\mu}_{Z-}^*(h) = \arg\min_{\mu}\min_{\beta}\sum_{i=1}^{n}(Z_i^* - \mu - \beta X_i)^2 K_{-,h}(X_i),$$

$$\hat{\mu}_{Z+}^*(h) = \arg\min_{\mu}\min_{\beta}\sum_{i=1}^{n}(Z_i^* - \mu - \beta X_i)^2 K_{+,h}(X_i).$$

    2.3. Apply Algorithm 3.1 to the bootstrapped data set $(X_1, T_1^*, Y_1^*), \ldots, (X_n, T_n^*, Y_n^*)$ using the same bandwidths h and b that are used in the rest of this algorithm but re-estimating all of the local polynomials on the bootstrap data. Generate B_1 new bootstrap samples and let $\Delta^{**}(h, b)$ represent the bias estimator returned by Algorithm 3.1.

    2.4. Save the estimator $\hat{\zeta}_k^*(h) = \frac{\hat{\mu}_{Y+}^*(h) - \hat{\mu}_{Y-}^*(h)}{\hat{\mu}_{T+}^*(h) - \hat{\mu}_{T-}^*(h)}$, and its bias $\Delta_k^{**}(h, b)$.

STEP 3. Define $\zeta^* = \frac{\hat{g}_{Y+}(0) - \hat{g}_{Y-}(0)}{\hat{g}_{T+}(0) - \hat{g}_{T-}(0)}$ and use the empirical CDF of $\hat{\zeta}_1^*(h) - \Delta_1^{**}(h, b) - \zeta^*, \ldots, \hat{\zeta}_{B_2}^*(h) - \Delta_{B_2}^{**}(h, b) - \zeta^*$ as the sampling distribution of $\hat{\zeta}(h) - \Delta^*(h, b) - \zeta$.

Algorithm 3.2 also consists of three steps. The first step estimates the bootstrap DGP, which is captured by second order local polynomials. The second step creates a series of new samples, to which the Algorithm 3.1 is applied. The last step uses the empirical distribution of bias-corrected estimator to construct CIs. As before, B_2 is assumed large enough so that simulation error can be ignored. The validity of Algorithm 3.2 is established in the following theorem.

Theorem 3.2 *Under Assumptions 2.1 and 2.2, as n, B_1 and B_2 go to infinity,*

$$\mathbb{V}^*(\hat{\zeta}^*(h) - \Delta^{**}(h, b))/V^{bc}(h, b) \to^p 1$$

and

$$\sup_x \left| \Pr^*\left[\frac{\hat{\zeta}^*(h) - \Delta^{**}(h, b) - \zeta^*}{\mathbb{V}^*(\hat{\zeta}^*(h) - \Delta^{**}(h, b))^{1/2}} \le x \right] - \Pr\left[\frac{\hat{\zeta}(h) - \Delta^*(h, b) - \zeta}{V^{bc}(h, b)^{1/2}} \le x \right] \right| \to^p 0.$$

Theorem 3.2 enables one to construct CIs in the following form:

$$\left(\hat{\zeta}(h) - \Delta^*(h, b) + \zeta^* - (\hat{\zeta}^*(h) - \Delta^{**}(h, b))_{1-\alpha/2}, \hat{\zeta}(h) - \Delta^*(h, b) + \zeta^* + (\hat{\zeta}^*(h) - \Delta^{**}(h, b))_{\alpha/2} \right),$$

where all the terms with superscript $*$ are defined in Algorithm 3.2. This CI is not centred at the bias-corrected point estimator.

11

Remark 3.1 The proposed bias correction differs from RBC's formula in finite samples. While the analytical bias is obtained by linearizing $\mathrm{E}\left[\frac{\hat{\tau}_Y(h)}{\hat{\tau}_T(h)} - \frac{\tau_Y}{\tau_T}\right]$ and then only evaluating its first order terms, Algorithm 3.1 directly estimates $\mathrm{E}\left[\frac{\hat{\tau}_Y^*(h)}{\hat{\tau}_T^*(h)} - \frac{\tau_Y^*}{\tau_T^*}\right]$ through bootstrap. Both methods consistently estimate the bias. By avoiding the linearization of the estimand, we remove one source of approximation error, potentially leading to performance improvements relative to the RBC analytical formulas or the multiplier bootstrap proposed by Chiang et al. (2019).[11]

Remark 3.2 When the original treatment is binary, the bootstrap sample will no longer have binary treatment. Though it creates some difficulty for interpretation, it does not invalidate the estimation and inference because its conditional expectation and covariance with outcome variable remain unchanged.

Remark 3.3 The iterated bootstrap is less computationally intensive than it might initially appear due to two reasons. First, the wild bootstrap creates new residuals but leaves the regressors unchanged, which means the design matrices only need to be computed once even when they are repeatedly used in fitting local polynomials.[12] Second, the number of data points actually used in estimation is a lot smaller than the full sample due to the local nature of the estimation.

4   Monte Carlo Simulations

This section summarizes the result of Monte Carlo experiments designed to evaluate the finite sample performance of the bootstrap procedures proposed in Section 3 relative to the existing plug-in alternatives. The details about the data generating processes (DGP) used and implementation are provided in the online appendix.

The conditional mean functions used in the simulations are similar to the ones used by CCT, adapted to the fuzzy RD context. For concreteness, the first mean function (DGP 1) is designed to match features of U.S. congressional election data from Lee (2008). The second mean function (DGP 2) matches the relationship between child mortality rate and county poverty rate from the analysis of Head Start data in Ludwig and Miller (2007). The last mean function (DGP 3) is similar to the first one except for some coefficients. CCT motivates this as an attempt to generate a plausible model with sizable distortion when conventional t-test is performed. The true treatment effects for these DGPs are $\zeta_1 = 0.04, \zeta_2 = -3.45, \zeta_3 = 0.04$, respectively.

To accommodate different endogeneity structures found in empirical data, we consider three cases. In the baseline case the treatment status is exogenous, i.e., there is no correlation between treatment assignment and the outcome. In the two endogenous cases the treatment status is correlated with unobserved characteristics which affect the outcome. This is modeled by the correlation, ρ, between the error terms on the outcome and treatment status equations as described in the online appendix.

Besides the proposed bootstrap, three additional approaches are estimated for comparison: the multiplier bootstrap proposed by Chiang et al. (2019) for FRD,[13] the plug-in RBC as implemented in the *rdrobust* package described in Calonico et al. (2015), and the conventional

---

[11]We thank a referee for bringing this point to our attention.

[12]To fit local polynomials is equivalent to estimate weighted least square, i.e., the estimated parameter is $(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{Y}$, where $\mathbf{X}$ is matrix of regressors and $\mathbf{K}$ is weighting matrix determined by kernel function. Both $\mathbf{X}$ and $\mathbf{K}$ are not affected by the bootstrap so one just need to compute $(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}$ once and then reuse it in the bootstrap calculations. Then each bootstrap replication just requires a single matrix-vector multiplication.

[13]We are thankful to Harold Chiang for providing their code and helping us implement it. Any potential mistakes in the implementation are our own.

estimators. Simulation results are presented in Table 4.1. The first three columns report the estimated treatment effect's bias, standard error and root MSE. The fourth and fifth columns present the CI's empirical coverage and average length. The last three columns list the average coverage error optimal bandwidths proposed by Calonico et al. (2019a), which are used for the three RBC methods (h_{CEOpt}, b_{CEOpt}) and the conventional method (h_{MSE}). Naturally, both the RBC methods described previously and the bootstrap procedures proposed are designed for inference, so the interest lies mainly on the empirical coverage and interval length columns. The properties of the point estimators in the first three columns, while interesting are not the main focus of the analysis.

The baseline case is listed in Panel A. The three robust methods, wild bootstrap, multiplier bootstrap and plug-in RBC approaches, generate point estimates with very similar bias and standard errors, even though the multiplier bootstrap seems to achieve lower bias at the cost of higher standard errors, especially on DGP 2. In contrast, the conventional approach reports 4-10 times larger bias. This is not surprising since the robust methods explicitly correct the bias. The conventional method also fails to deliver a valid CI (coverage rates are 68.1%, 2.6% and 87.3% for the three DGPs respectively). Robust methods achieve improvements by reducing bias and increasing interval length. Except for DGP 2, they generate intervals with empirical coverage close to the nominal level and the wild bootstrap performs well for DGPs 1 and 3. However, for DGP 2, the wild bootstrap and the plug-in RBC methods report some size distortion. This is because DGP 2 has great curvature around the cutoff and makes precise fitting challenging. Still, both the wild bootstrap and the plug-in RBC methods improve significantly on the coverage obtained by the conventional method (from 2.6% to around 91%) at the sacrifice of slightly longer intervals (from 0.186 to 0.23). The multiplier bootstrap obtains better coverage, at the cost of significantly higher interval length and RMSE.

Panels B and C present results when the treatment is endogenous, which is likely the primary reason to choose RD designs as the identification strategy. The case with positive (negative) ρ is listed in Panel B (C). Again, the conventional estimator has significantly larger bias than the RBC methods. As for CIs, the robust approaches work reasonably well in all cases. The conventional method performs significantly worse, with empirical coverage rate as low as 1.7%. The sign of correlation has little effect on the bias since it is caused by model misspecification rather than an imperfect instrumental variable.

As discussed in Pei et al. (2017), bias correction adds variability that can sometimes dominate the bias reduction obtained, which is relevant to researchers interested in the point estimates of the treatment effect. Table 4.1 allows a similar analysis to that of Pei et al. (2017) by comparing the RMSE under the conventional and RBC based methods in all the DGPs considered. The results do not indicate any systematic loss in terms of RMSE when using the wild bootstrap or plug-in RBC, while coverage improves markedly in all cases.

To summarize, the wild bootstrap approach proposed in this paper performs significantly better than the conventional method and is at least on par with the plug-in RBC and multiplier bootstrap methods. This wild bootstrap procedure automatically accommodates various types of covariance structure and thus is a simple alternative to obtain valid CIs in RD designs. [14]

5   Extension: Clustered Data

This section explores the application of the bootstrap procedure to clustered data in RD designs and provides evidence for its usefulness.[15] Clustered data are very common in empirical studies

---

[14]While it would be plausible that the wild bootstrap proposed achieves the higher order refinements described by Calonico et al. (2018a), evaluating that formally is beyond the scope of this work.

[15]Another interesting extension of the bootstrap proposed in this paper would be including covariate-adjustments as proposed by Calonico et al. (2019). One could generate the DGP for the bias in Algorithm 3.1 using the particular

and there is a vast literature on handling it.[16] Units within the same cluster are usually dependent and ignoring this dependence is likely to invalidate statistical inference. In short, one can either explicitly estimate the dependence structure with some additional specifications, such as random coefficient models, or account for the dependence when performing inference by, for example using a cluster-robust variance estimator, see Liang and Zeger (1986); Arellano (1987).

Cluster-robust variance estimators are very popular in part because they do not require assumptions on the dependence structure and partly because of its availability in most statistical software. Its validity is based on asymptotics when the number of clusters, G, grows to infinity, which is, unfortunately, not trivial to establish in nonparametric models. The main obstacle is that shrinking bandwidths is likely to destroy the dependence structure. For local polynomial regressions, Wang (2003) and Chen et al. (2008) point out that the existence of joint density of running variable and clustering variable ensures that cluster dependence vanishes asymptotically, not being captured by the usual approximations.[17]

Bartalotti and Brummet (2017) develop analytical approximations for the distribution and MSE-optimal bandwidth selector for the conventional RD estimator in a fixed-G setting, sidestepping the issue. Calonico et al. (2019a) provides similar results for the RBC approach, including coverage error optimal bandwidth selectors that we implement below. Available software provides options to take this dependence into consideration. Both the *rdrobust* and *RDD* packages used in this paper offer the option to specify a clustering variable as explained by Calonico et al. (2017). Chiang et al. (2019) offer a cluster-robust version of their multiplier bootstrap under high level assumptions, relying on the Bahadur representation of the Wald ratio without describing implementation details.

Naturally, our wild bootstrap approach could offer an intuitive and easy to implement alternative to the analytical approximations described. In fact, Cameron et al. (2008) provide a comprehensive survey of bootstrap methods and show that proper bootstrap procedures outperform the conventional cluster-robust variance estimator when the number of clusters is small (five to thirty).

To highlight the flexibility and robustness of the wild bootstrap procedure proposed in this paper, we revise the resampling algorithm to accommodate clustering and test its performance with clustered data. Following Brownstone and Valletta (2001) and Cameron et al. (2008), the wild bootstrap procedure for clustered data is quite straightforward: for units in the same cluster, their residuals are multiplied by the same random number drawn from the auxiliary distribution. For example,

$$Z_{gi}^* = \hat{g}_Z(X_{gi}) + \hat{\varepsilon}_{Z_{gi}} e_g^*,$$

where e_g^*, a random number from distribution with zero mean and unit variance, is shared by all units in the same group. For the purpose of simulation, it is assumed that errors in the outcome equation are clustered according to a random effect model, in particular, $u_{ygi} = u_{yg}^* + u_{yi}^*$ with $g = 1, 2, \ldots, G$ being a cluster indicator.[18]

---

covariate choices and constraints imposed by the researcher as discussed in Calonico et al. (2019) in their Section 2. The wild bootstrap based on transforming the residuals described here should, intuitively, hold. A formal analysis of that case is beyond the scope of this paper.

[16]See Wooldridge (2003); Cameron et al. (2011); Cameron and Miller (2015) for an overview on this topic.

[17]A special case where this does not happen is that clustering occurs at the running variable level as discussed by Chen and Jin (2005). For example, in panel data where each individual are observed for multiple times and the running variable is at individual level, each individual is a cluster and will not vanish with shrinking bandwidth. Lee and Card (2008) consider another example in RD designs where clustering occurs at the running variable level and cluster-robust variance estimator is recommended for inference.

[18]The design ensures that the individual errors have the same standard errors as the baseline case presented in Section 4 for easy comparison.
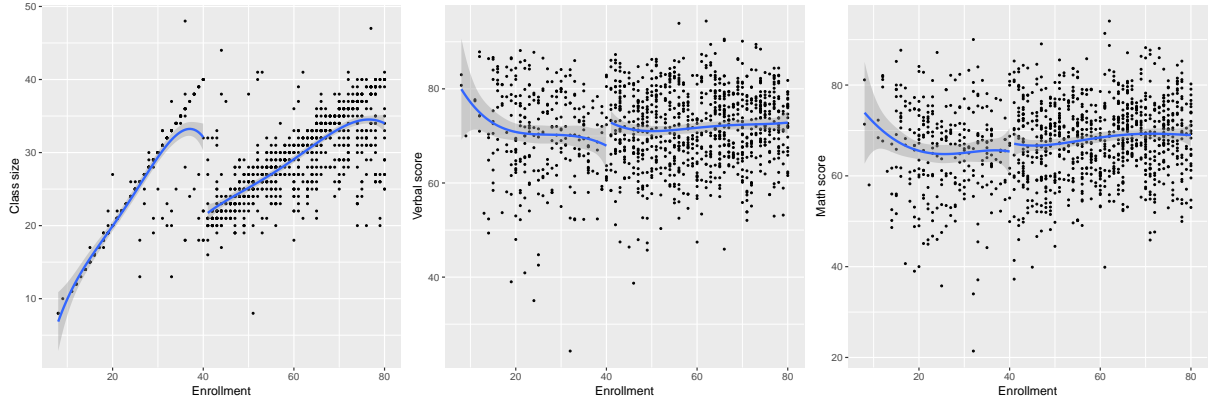
Figure 1: Class size, average verbal and math scores

Simulation results for $G = 5, 10, 25$ are reported in Table 5.1.[19] The wild bootstrap approach consistently outperforms the conventional method, closely matching the coverage from the plug-in RBC approach. This simple experiment shows that the wild bootstrap procedure proposed can also be easily applied to clustered data with slight adjustment to its resampling algorithm.

6 Empirical Illustration

In this section, we apply the bootstrap procedure to the data used in Angrist and Lavy (1999).[20] In that paper, the effects of class size on scholastic achievement are estimated using the "Maimonides' rule" as instrument.

As described by Angrist and Lavy (1999) the "Maimonides' rule" holds that the maximum class size is 40, and has been adopted by Israeli public schools to determine the division of enrollment cohorts into classes since 1969. Following this rule, when enrollment increases and passes multiples of 40, an additional class is required. Since the total enrollment is roughly evenly divided into all classes, an additional class causes a sudden drop in class sizes. Ideally, when the enrollment grows from 40 to 41, class size will drop by almost half. Because of student turnover and imperfect enforcement of this rule, the empirical data fits into a fuzzy RD design.[21]

We consider the first discontinuity in class size for the 4th grade. The sample used in this application includes 1164 classes from schools with enrollments no larger than 80. The outcome variables are average verbal and math test scores at class level. The discontinuities in class size and outcomes against enrollment are visualized in Figure 1. Each dot in these plots represents a class and the regression lines are fitted by fourth order polynomials. The shaded areas indicate CIs. The first plot clearly shows the discontinuity in class size, which is exploited for identification of the class size effect. The second plot suggests a discontinuity in average verbal score, but not as important as that in class size. The last plot does not provide much evidence for a discontinuity in average math score.

Three methods are applied to estimate the effect of class size on average verbal/math scores and results are shown in Table 6.1. The first column lists the original point estimates from

---

[19]G denotes the number of clusters on each side of the cutoff.

[20]The data is available at http://economics.mit.edu/faculty/angrist/data1/data/anglavy99.

[21]The treatment variable, class size, is multivalued as opposed to the discrete binary case discussed in Section 2.

local linear regression, which depends only on the bandwidth choice. The second column lists the bias-corrected point estimates based on bootstrap and analytical plug-in bias corrections. The estimates are very close to each other but differ meaningfully from the original estimates: the magnitude increases from 0.495 to 0.547~0.568 for average verbal score and 0.198 to 0.224~0.251 for math score.

Consistent with Figure 1, none of the CIs for the treatment effect on both average verbal and math scores excludes zero. The CIs from the wild bootstrap are wider than that from the plug-in RBC approach, in line with the simulations presented.

7   Conclusion

A new wild bootstrap procedure is proposed to obtain RBC point estimates and valid CIs in fuzzy RD designs. This new method provides an easy to implement alternative to the analytical RBC approach established by Calonico et al. (2014) and further advanced in recent years. The procedure is implemented through a novel iterated bootstrap that extends the developments in Bartalotti et al. (2017), also serving as an intuitive alternative to the multiplier bootstrap in Chiang et al. (2019). This new procedure is proved to be theoretically valid and empirically supported by simulation studies, performing as well as analytical plug-in alternatives, including in the presence of clustered data. An empirical illustration is provided, confirming the procedure's applied relevance.

Acknowledgements

References

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics 114*, 533–75.

Arellano, M. (1987). Practitioners' corner: computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics 49*, 431–34.

Armstrong, Timothy and Michal Kolesár. Forthcoming. Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics, Forthcoming*.

Armstrong, Timothy B. and Michal Kolesár. 2018. Optimal inference in a class of regression models. *Econometrica* 86 (2):655–83.

Bartalotti, O. (2018). Regression discontinuity and heteroskedasticity robust standard errors: Evidence from a fixed-bandwidth approximation. *Journal of Econometric Methods 8*(1).

Bartalotti, O. and Q. Brummet (2017). Regression discontinuity designs with clustered data. In M. D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory*

and Applications (Advances in Econometrics), Volume 38, pp. 383–420. Emerald Publishing Limited.

Bartalotti, O., Q. Brummet, and S. Dieterle (2019). A Correction for regression discontinuity designs with group-specific mismeasurement of the running variable. *IZA Discussion Paper*, n. 12366.

Bartalotti, O., G. Calhoun, and Y. He (2017). Bootstrap confidence intervals for sharp regression discontinuity designs. In M. D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics)*, Volume 38, pp. 421–53. Emerald Publishing Limited.

Bertanha, M., and G. Imbens (2019). External validity in fuzzy regression discontinuity designs. *Journal of Business and Economic Statistics 0*, 1–39.

Brownstone, D. and R. Valletta (2001). The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. *Journal of Economic Perspectives 15*(4), 129–41.

Calonico, S., M. D. Cattaneo, and M. H. Farrell, (2018a). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association 113*, 767–79.

Calonico, S., M. D. Cattaneo, and M. H. Farrell, (2019a). Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. *Econometrics Journal, Forthcoming*.

Calonico, S., M. D. Cattaneo, and M. H. Farrell, (2019b). Coverage error optimal confidence intervals for local polynomial regression. *arXiv:1808.01398*

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2015). rdrobust: An R package for robust nonparametric inference in regression-discontinuity designs. *R Journal 7*, 38–51.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *Stata Journal 17*, 372–404.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics 3*, 442–51.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*, 2295–326.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics 90*, 414–27.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business and Economic Statistics 29*, 238–49.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*, 317–72.

Cattaneo, M. D., , M. H. Farrell, and Y. Feng (Forthcoming). Large sample properties of partitioning-based series estimators. *Annals of Statistics, Forthcoming*.

Chen, K., J. Fan, and Z. Jin (2008). Design-adaptive minimax local linear regression for longitudinal/clustered data. *Statistica Sinica*, 515–34.

Chen, K. and Z. Jin (2005). Local polynomial regression analysis of clustered data. *Biometrika 92*, 59–74.

Chiang, H. D., Y.-C. Hsu, and Y. Sasaki (2019). Robust uniform inference for quantile treatment effects in regression discontinuity designs. *Journal of Econometrics 211*, 589–618.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association 87*, 998–1004.

Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Volume 66. CRC Press, New York, NY.

Gelman, A. and G. Imbens (2018). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics 0*, 1–10.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*, 201–09.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies 79*, 933–59.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*, 615–35.

Imbens, G. W. and W. Wager (2019). Optimized regression discontinuity designs. *Review of Economics and Statistics 101*, 264–78.

Kolesár, M. and C. Rothe (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review 108*, 2277–304.

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics 142*, 675–97.

Lee, D. S. and D. Card (2008). Regression discontinuity inference with specification error. *Journal of Econometrics 142*, 655–74.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Ludwig, J. and D. L. Miller (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *Quarterly Journal of Economics 122*, 159–208.

Pei, Z., D. Card , S. Lee and A. Weber (2017). Regression kink designs: theory and practice. In M. D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics)*, Volume 38, pp. 341–82. Emerald Publishing Limited.

Porter, J. (2003). Estimation in the regression discontinuity model. *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison*, 5–19.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika 90*, 43–52.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review 93*(2), 133–38.

Table 4.1: Empirical coverage and average interval length

DGP	Method	Bias	SD	RMSE	EC(%)	IL	h_{CEOpt}	b_{CEOpt}	h_{MSE}
					Panel A: $\rho = 0$				
1	Wild bootstrap	0.010	0.057	0.058	94.9	0.217	0.140	0.323	
	Multiplier bootstrap	0.003	0.070	0.074	93.3	0.274	0.140	0.323	
	Plug-in RBC	0.010	0.058	0.059	93.1	0.210	0.140	0.323	
	Conventional	0.042	0.032	0.053	68.1	0.116			0.400
2	Wild bootstrap	0.026	0.062	0.067	91.7	0.234	0.117	0.299	
	Multiplier bootstrap	0.023	0.324	0.336	96.3	1.269	0.117	0.299	
	Plug-in RBC	0.026	0.062	0.068	90.3	0.231	0.117	0.299	
	Conventional	0.215	0.079	0.229	2.6	0.186			0.216
3	Wild bootstrap	0.003	0.059	0.059	95.4	0.231	0.115	0.317	
	Multiplier bootstrap	0.001	0.076	0.081	92.4	0.297	0.115	0.317	
	Plug-in RBC	0.003	0.059	0.059	94.1	0.224	0.115	0.317	
	Conventional	-0.025	0.044	0.050	87.3	0.157			0.205
					Panel B: $\rho = 0.9$				
1	Wild bootstrap	0.010	0.057	0.058	96.3	0.223	0.139	0.323	
	Multiplier bootstrap	0.005	0.070	0.073	93.8	0.273	0.140	0.323	
	Plug-in RBC	0.013	0.059	0.060	94.6	0.214	0.139	0.323	
	Conventional	0.043	0.033	0.054	70.7	0.121			0.398
2	Wild bootstrap	0.025	0.066	0.070	94.5	0.246	0.119	0.302	
	Multiplier bootstrap	0.030	0.343	0.354	96.4	1.343	0.118	0.302	
	Plug-in RBC	0.041	0.067	0.078	89.7	0.233	0.119	0.302	
	Conventional	0.226	0.092	0.244	3.0	0.207			0.222
3	Wild bootstrap	0.003	0.067	0.067	95.5	0.255	0.114	0.316	
	Multiplier bootstrap	0.003	0.076	0.081	92.8	0.297	0.114	0.317	
	Plug-in RBC	0.006	0.063	0.063	94.7	0.230	0.114	0.316	
	Conventional	-0.024	0.043	0.049	86.5	0.156			0.204
					Panel C: $\rho = -0.9$				
1	Wild bootstrap	0.011	0.063	0.063	93.6	0.223	0.141	0.324	
	Multiplier bootstrap	0.000	0.072	0.076	93.8	0.282	0.140	0.324	
	Plug-in RBC	0.008	0.059	0.060	93.1	0.211	0.141	0.324	
	Conventional	0.042	0.031	0.052	65.7	0.113			0.402
2	Wild bootstrap	0.025	0.057	0.062	91.7	0.226	0.114	0.296	
	Multiplier bootstrap	0.037	0.304	0.314	95.6	1.193	0.114	0.296	
	Plug-in RBC	0.024	0.057	0.062	90.8	0.218	0.114	0.296	
	Conventional	0.201	0.064	0.211	1.7	0.165			0.208
3	Wild bootstrap	0.003	0.059	0.059	95.8	0.234	0.116	0.317	
	Multiplier bootstrap	-0.001	0.078	0.082	93.0	0.305	0.116	0.319	
	Plug-in RBC	0.001	0.059	0.059	94.8	0.227	0.116	0.317	
	Conventional	-0.027	0.045	0.052	89.1	0.160			0.207

Note: EC denotes empirical coverage and IL denotes average interval length based on 5000 simulations; nominal coverage probabilities are 95% for each estimator. Sample size is 1000. Following Chiang et al. (2019), the multiplier bootstrap uses the Epanechnikov kernel. Otherwise, the triangular kernel is used. The columns h_{CEOpt} and b_{CEOpt} list average coverage error optimal bandwidths following Calonico et al. (2019a). The column h_{MSE} lists average MSE-optimal bandwidth. The bootstrap approach uses $B_1 = 500$ replications to compute bias and $B_2 = 999$ replications to obtain the empirical distribution of bias-corrected estimator.

Table 5.1: Empirical coverage and average interval length (clustered data).

DGP	Method	Bias	SD	RMSE	EC(%)	IL	h_{CEOpt}	b_{CEOpt}	h_{MSE}
					Panel A: $G = 5$				
1	Wild bootstrap	0.017	0.081	0.083	87.2	0.269	0.224	0.318	
	Plug-in RBC	0.017	0.082	0.083	86.9	0.275	0.224	0.318	
	Conventional	0.043	0.071	0.083	83.7	0.249			0.392
2	Wild bootstrap	0.033	0.085	0.092	84.1	0.278	0.147	0.297	
	Plug-in RBC	0.034	0.086	0.093	84.8	0.292	0.147	0.297	
	Conventional	0.214	0.101	0.237	22.5	0.275			0.216
3	Wild bootstrap	0.007	0.081	0.081	89.0	0.273	0.179	0.312	
	Plug-in RBC	0.007	0.081	0.081	89.2	0.278	0.179	0.312	
	Conventional	-0.023	0.076	0.080	87.5	0.261			0.202
					Panel B: $G = 10$				
1	Wild bootstrap	0.016	0.069	0.071	90.5	0.243	0.198	0.321	
	Plug-in RBC	0.016	0.069	0.071	89.5	0.240	0.198	0.321	
	Conventional	0.043	0.055	0.070	83.8	0.200			0.396
2	Wild bootstrap	0.031	0.071	0.078	88.7	0.256	0.143	0.299	
	Plug-in RBC	0.032	0.072	0.079	87.6	0.258	0.143	0.299	
	Conventional	0.213	0.089	0.231	12.9	0.239			0.216
3	Wild bootstrap	0.005	0.068	0.069	92.7	0.248	0.160	0.316	
	Plug-in RBC	0.005	0.069	0.069	91.5	0.245	0.160	0.316	
	Conventional	-0.025	0.062	0.067	88.0	0.220			0.204
					Panel C: $G = 25$				
1	Wild bootstrap	0.014	0.062	0.064	92.4	0.223	0.175	0.323	
	Plug-in RBC	0.014	0.063	0.064	90.5	0.217	0.175	0.323	
	Conventional	0.043	0.043	0.060	78.7	0.157			0.399
2	Wild bootstrap	0.031	0.066	0.072	90.5	0.238	0.136	0.300	
	Plug-in RBC	0.032	0.066	0.073	89.2	0.238	0.136	0.300	
	Conventional	0.214	0.084	0.230	6.4	0.210			0.216
3	Wild bootstrap	0.004	0.062	0.062	94.1	0.230	0.143	0.317	
	Plug-in RBC	0.003	0.063	0.063	92.8	0.226	0.143	0.317	
	Conventional	-0.025	0.053	0.059	86.6	0.186			0.205

Note: EC denotes empirical coverage and IL denote average interval length based on 5000 simulations; nominal coverage probabilities are 95% for each estimator. Sample size is 1000. The triangular kernel is used. The columns h_{CEOpt} and b_{CEOpt} list average coverage error optimal bandwidths for clustered data following Calonico et al. (2019a).The column h_{MSE} lists average MSE-optimal bandwidth. The bootstrap approach uses $B_1 = 500$ replications to compute bias and $B_2 = 999$ replications to obtain empirical distribution of bias-corrected estimator.

Table 6.1: The effect of class size on average verbal score and average math score.

	ATE		95% CI		h_{CEOpt}	b_{CEOpt}	h_{MSE}
	Original	Corrected					
Panel A: Average verbal score							
Wild bootstrap	-0.495	-0.547	(-1.138	0.213)	8.706	18.278	
Plug-in RBC	-0.495	-0.568	(-1.212	0.075)	8.706	18.278	
Conventional	-0.488		(-1.104	0.129)			7.952
Panel B: Average math score							
Wild bootstrap	-0.198	-0.224	(-0.953	0.586)	8.159	17.683	
Plug-in RBC	-0.198	-0.251	(-1.002	0.500)	8.159	17.683	
Conventional	-0.202		(-0.802	0.398)			9.200