Wild bootstrap for fuzzy regression discontinuity designs: obtaining robust bias-corrected confidence intervals

YANG HE* AND OTÁVIO BARTALOTTI†

*Microsoft, 1 Memorial Dr., Cambridge, MA 02142, Cambridge, MA 02142.
Email: y.he0802@gmail.com

†Department of Economics, Iowa State University and IZA, 468F Heady Hall, Ames, IA, 50011.
Email: bartalot@iastate.edu

First version received: 28 March 2019; final version accepted: 20 November 2019.

Summary: This paper develops a novel wild bootstrap procedure to construct robust bias-corrected valid confidence intervals for fuzzy regression discontinuity designs, providing an intuitive complement to existing robust bias-corrected methods. The confidence intervals 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.

Keywords: Fuzzy regression discontinuity, robust confidence intervals, wild bootstrap, average treatment effect.

JEL codes: C01, C14, C21.

1. INTRODUCTION

Regression discontinuity (RD) designs are one of the leading empirical approaches in economics, political science, and public policy evaluation, being used extensively 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 the 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 to 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

¹ Imbens and Lemieux (2008) and Lee and Lemieux (2010) provided reviews of this literature with many examples.
are self-selected to treatment on the basis of anticipated gains, Hahn et al. (2001) showed 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. 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) minimises the asymptotic mean squared error (AMSE) of the treatment effect estimator.

In an influential paper, Calonico et al. (2014) (henceforth ‘CCT’), showed 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. CCT proposed a solution to this problem by obtaining a valid estimate of the leading bias term and re-centring the conventional point estimator. Furthermore, the additional variability introduced by the bias estimation needs to be considered when CIs are constructed. This approach is referred to as the robust bias-corrected (RBC) inference method, and it 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 the present 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 that of CCT, and simulations demonstrate that 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) developed valid coverage error expansions for RBC CIs for general nonparametric inference and the RD design–specific case, respectively. They showed 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. The expansions were then used to develop new coverage-error–optimal bandwidths for forming CIs. In recent work, Calonico et al. (2019b) characterised 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 on 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

---

2 See Fan (1992) and Hahn et al. (2001) for a detailed discussion of the properties of the local polynomial estimator 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.
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 estimators is then used to construct CIs. This procedure is in line with RBC literature, where the variance of the estimated bias term and the covariance between the estimated bias and the 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 with 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, as described in Calonico et al. (2017), and could naturally serve as an alternative.

Moreover, the results in the present 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 is 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) evaluated nonlinear functions involving four discontinuity parameters estimated jointly, requiring the researcher to evaluate the bias and covariances of all parameters, while relying on linearisations and the delta method to obtain analytical formulas. In this particular case, the bootstrap method we propose is very practical compared with the need to derive formulas for RBC CIs.

Additionally, because the bootstrap mimics the true DGP, 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 are the main strengths of the proposed bootstrap.

Compared with Bartalotti et al. (2017), who developed a similar iterated bootstrap procedure for robust inference in the special case of sharp RD designs, the present paper provides important generalisations in several dimensions. First, it adapts bootstrapping instrumental variable models to fuzzy RD designs. Second, its validity is extended to and theoretically proved in general local polynomials and higher-order 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 are confirmed by simulation studies.

Although we focus on the RD case because 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 whether 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 present scope.

Concomitantly and independently, Chiang et al. (2019) proposed a multiplier bootstrap procedure for fuzzy RD and many related settings that is based on a Bahadur representation of a general class of Wald estimators. Their paper, however, had a more specific focus on uniform inference for quantile treatment effects in RD designs. Both the procedure and proofs in that paper differ

© 2020 Royal Economic Society.
from the ones proposed here and could potentially serve as alternatives in the cases covered by both approaches, if one keeps in mind that their approach offers robust uniform inference.5

In a related work, Cattaneo et al. (forthcoming) independently considered a single-step wild bootstrap for the RBC $t$ statistics of interest in the context of nonparametric regression using series estimators, and they proved the validity of the wild bootstrap in that setting. Those 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 was shown to provide robust uniformly valid confidence bands.

The results of Cattaneo et al. (forthcoming) and Chiang et al. (2019) suggest that an iterated bootstrap like the one proposed here might achieve uniform coverage, even though a formal analysis of that possibility is beyond the scope of the present paper. Nevertheless, our pointwise 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 of the performance of 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 by 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 with consideration of the impact of the nonrandom worst-case bias in estimation in that class of functions. In closely related work, Imbens and Wager (2019) examined the RD estimator as a linear combination of the observed outcomes and obtained weights for each observation by numerical optimisation. The minimax linear estimator minimises the variance and worst-case bias for a similarly defined class of functions.

Honest CIs offer 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 the present paper because we do not base the bootstrap procedure on obtaining worst-bias characterisations. Nevertheless, a bootstrap procedure that would achieve the properties of the CIs described in Armstrong and Kolesár (2018) or Imbens and Wager (2019) would be an interesting topic for future research that can build upon the developments presented here.

The paper is organised 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 CIs. Asymptotic properties are discussed and summarised 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.

5 Section 4 presents a comparison between our proposed wild bootstrap and the multiplier bootstrap in Chiang et al. (2019).
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 that determines treatment assignment, \(T_i\) is a binary variable that 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 undergoes 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, they are \(T_i(0)\) 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 themselves are 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

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

where the symbol \(\mathbb{E}\) represents the expectation, and \(\tau_Y\) and \(\tau_T\) represent the sharp RD estimators, i.e., the 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, \(\tau_Y\), reveals the intent-to-treat effect. The treatment effect is recovered by dividing the intent-to-treat effect by the difference in treatment probabilities. When the treatment effect is not constant across units, \(\zeta\) should be interpreted with caution. If treatment status is independent of treatment effects at the cutoff, \(\zeta\) is the average treatment effect at the cutoff. This assumption rules out self-selection on the basis of anticipated gain. Hahn et al. (2001) showed that under a less restrictive assumption that the treatment effect and status are jointly independent of the running variable around the cutoff, and under a local weak monotonicity assumption, the LATE is identified.\(^6\)

Equation (2.1) presents \(\zeta\) as a ratio of two sharp RD estimators. Because of 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 \(\sigma_{Z+}^2(x)\) and \(\sigma_{Z-}^2(x)\), the \(\eta^{th}\)-order derivative of conditional expectations \(\mu_{Z+}^{(\eta)}(x)\) and \(\mu_{Z-}^{(\eta)}(x)\), and their

\(^6\) 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.

© 2020 Royal Economic Society.
limits. Formally, they are defined as
\[ \mu_{Z+}(x) = \mathbb{E}[Z_i(1) \mid X_i = x] \quad \mu_{Z-}(x) = \mathbb{E}[Z_i(0) \mid X_i = x] \]
\[ \sigma_{Z+}^2(x) = \mathbb{V}[Z_i(1) \mid X_i = x] \quad \sigma_{Z-}^2(x) = \mathbb{V}[Z_i(0) \mid X_i = x] \]
\[ \mu_{Z+}^{(q)}(0) = \frac{d^q \mu_{Z+}(x)}{dx^n} \quad \mu_{Z-}^{(q)}(0) = \frac{d^q \mu_{Z-}(x)}{dx^n} \]
where the symbol \( \mathbb{V}(\cdot) \) represents variance. The treatment effect \( \tau \) is nonparametrically estimable because \( \mu_{Z-} \) and \( \mu_{Z+} \) 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 \( \{X_i, T_i, Y_i\}_{i=1}^n \) form a random sample of size \( n \). There exists a positive number \( \kappa_0 \) 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) \( E[Z_i^+ \mid X_i = x] \) is bounded; (c) \( \mu_{Z-}(x) \) and \( \mu_{Z+}(x) \) are three times continuously differentiable; (d) \( \sigma_{Z-}^2(x) \) and \( \sigma_{Z+}^2(x) \) are continuous and bounded away from zero; and (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 conditions 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 fuzzy RD 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 because of 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
\[ \hat{\tau}(h) = \frac{\hat{\tau}_Y(h)}{\hat{\tau}_T(h)} = \frac{\hat{\mu}_{Y+}(h) - \hat{\mu}_{Y-}(h)}{\hat{\mu}_{T+}(h) - \hat{\mu}_{T-}(h)}, \]
with
\[ \hat{\mu}_{Z+}(h) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n \mathbb{I}(X_i \geq 0)(Z_i - \beta_0 - X_i\beta_1)^2 \frac{1}{h} K \left( \frac{X_i}{h} \right) \]
\[ \hat{\mu}_{Z-}(h) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n \mathbb{I}(X_i < 0)(Z_i - \beta_0 - X_i\beta_1)^2 \frac{1}{h} K \left( \frac{X_i}{h} \right). \]

\(^7\) Throughout the main text we focus on the case in which the researcher implements a local linear model to estimate \( \tau_Z \) and a quadratic model to approximate the bias term. The proofs presented in the online appendix for the validity of the proposed bootstraps include the general case in which higher-order polynomials can be used to obtain \( \tau_Z \) 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.
The conditional expectations \( \mu_{Z^+} \) and \( \mu_{Z^-} \) are consistently estimated by \( \hat{\mu}_{Z^+}(h) \) and \( \hat{\mu}_{Z^-}(h) \) when \( h \to 0 \). The asymptotic distribution of the quotient estimator \( \frac{\hat{\mu}_{Z^+}(h)}{\hat{\mu}_{Z^-}(h)} \) 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.,

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

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

Let \( V(h) = \mathbb{V}(\hat{\tau}_T(h) | X_1, ..., X_n) \). Then, \( \hat{\tau}_T(h) - \tau_T \xrightarrow{d} N\left(0, 1\right) \), and the CIs can be constructed as

\[
\hat{\tau}_T(h) \pm q_{1-\alpha/2} V(h)^{1/2},
\]

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 \( \hat{\tau}_T(h) \) 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) showed that the asymptotic distribution is normal but not centred at zero. Using (2.2) 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) developed a plug-in bandwidth selector for RD estimators, which is optimal in the sense that the AMSE of the point estimator is minimised.

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 normalisation so that the additional variability is accounted for.

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

\[
\mathbb{E}[\hat{\tau}_T(h) | X_1, ..., X_n] - \tau_T = h^2\left(\frac{1}{\tau_T}B_+(h) - \frac{\tau_Y}{\tau_T^2}B_T(h)\right)(1 + o_p(1)),
\]

with

\[
B_+(h) = \frac{\mu_{Z^+}}{2}B_+(h) - \frac{\mu_{Z^-}}{2}B_-(h).
\]

The terms \( B_+(h) \) and \( B_-(h) \), explicitly defined in the 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 \( \tau_T, \mu_{Z^+}^{(2)}, \) and \( \mu_{Z^-}^{(2)} \). Among these, \( \tau_T \) is consistently estimated by the local linear regression with bandwidth \( h \). CCT proposed 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

---

9 Unless otherwise stated, all limits in the present paper are assumed to hold as \( n \to \infty \).
produces the bias-corrected estimator
\[ \hat{\zeta}^{bc}(h, b) = \hat{\zeta}(h) - \Delta(h, b), \]
with
\[ \Delta(h, b) = h^2 \left( \frac{1}{\hat{\tau}_T(h)} \hat{B}_Y(h, b) - \frac{\hat{\tau}_Y(h)}{\hat{\tau}_T(h)} \hat{B}_T(h, b) \right). \]
\[ \hat{B}_Z(h, b) = \frac{\mu(2)}{2} \frac{B_+ + \mu(2)}{2} \frac{B_+ - \mu(2)}{2} \]
\[ V^{bc}(h, b) = V(h) + C(h, b), \]
\[ C(h, b) \text{ captures the adjustment to variance introduced by the bias-correction term.} \]

Notice that the bias \( \Delta(h, b) \) is estimated with uncertainty. As a result, the variance of the bias-corrected estimator \( \hat{\zeta}^{bc}(h, b) \) is different from \( V(h) \). CCT proposed a new formula for the variance of the bias-corrected estimator and used it for normalisation:
\[ \frac{\hat{\zeta}^{bc}(h, b) - \zeta}{V^{bc}(h, b)^{1/2}} \overset{d}{\to} N(0, 1), \] (2.3)

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 the present 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 requires only 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 of kernel functions commonly used in practice.

To simplify notation, let \( m = \min(h, b) \), and define the scaled and truncated kernel functions:
\[ K_{+,h}(x) = \frac{1}{h} K(x/h) \mathbb{1}[x \geq 0], \quad K_{-,h}(x) = \frac{1}{h} K(x/h) \mathbb{1}[x < 0], \]
\[ K_{+,b}(x) = \frac{1}{b} K(x/b) \mathbb{1}[x \geq 0], \quad K_{-,b}(x) = \frac{1}{b} K(x/b) \mathbb{1}[x < 0]. \]

Recent studies have further established RBC inference as the standard method in the RD literature. Calonico et al. (2018a) developed valid coverage error expansions for RBC CIs for general nonparametric inference, while Calonico et al. (2019a) obtained similar results in the context of RD designs. They showed 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) provided 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 on the basis of the insight provided by 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 used 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 in which the bandwidths used for estimating \(\tau\) and the bias are the same, \(h = b\), which translates to the bandwidth condition \(nb^7 \rightarrow 0\) under Assumption 2.2. By the same argument that \(h = o_p(n^{-1/5})\) is required for valid inference in an RD design estimated by local linear regression, \(b = o_p(n^{-1/7})\) is required in an 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 are specifically tailored for RBC inference in RD designs. Alternatively, mean squared error (MSE)–optimal bandwidths, which are also valid (but suboptimal) in terms of coverage error, could be used, as discussed in that paper. Both bandwidth selectors are available in the statistical package \texttt{rdrobust} for STATA and R, as described in Calonico et al. (2015, 2017).\(^{10}\)

Algorithm 3.1 consistently estimates the bias term.

\textbf{Algorithm 1 (Bias Estimation)}. Assume \(h\) and \(b\) are bandwidths as defined by Assumption 2.2.

\textbf{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:

\[
\hat{g}_{Z-}(x) = \hat{\beta}_{Z-,0} + \hat{\beta}_{Z-,1}x + \hat{\beta}_{Z-,2}x^2, \quad \hat{g}_{Z+}(x) = \hat{\beta}_{Z+,0} + \hat{\beta}_{Z+,1}x + \hat{\beta}_{Z+,2}x^2
\]

with

\[
(\hat{\beta}_{Z-,0}, \hat{\beta}_{Z-,1}, \hat{\beta}_{Z-,2})' = \min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^{n} (Z_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2)^2 K_{-,b}(X_i)
\]

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

\(^{10}\) A package for R called \texttt{frdboot} implements the procedures described below and is available at: https://github.com/yhe0802/FRD-bootstrap/blob/master/frdboot_0.1.0.zip

© 2020 Royal Economic Society.
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 \( \hat{\varepsilon}_Z \) for all \( i \).

**Step 2.** Repeat the following steps \( B_1 \) times to produce the bootstrap estimates \( \hat{\tau}^{*}_{Z,b_1}(h), \ldots, \hat{\tau}^{*}_{Z,b_B}(h) \). For the \( k \)th replication:

1. Draw independent and identically distributed (i.i.d.) random variables \( e_i^* \) with mean zero, variance one, and bounded fourth moments independent of the original data, and construct

\[ \varepsilon_i^* = \hat{\varepsilon}_Z e_i^*, \quad Z_i^* = \hat{g}_Z(X_i) + \varepsilon_i^* \]

for all \( i \). Note that both \( \hat{\varepsilon}_Y \) and \( \hat{\varepsilon}_T \) are multiplied by the same \( e_i^* \).

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. Save \( \hat{\xi}_k^*(h) = \frac{\hat{\mu}_{Z^-}^*(h) - \hat{\mu}_{Z^+}^*(h)}{\hat{\mu}_{T^-}^*(h) - \hat{\mu}_{T^+}^*(h)}. \)

**Step 3.** Estimate the bias as

\[ \Delta^*(h, b) = \frac{1}{B_1} \sum_{k=1}^{B_1} \hat{\xi}_k^*(h) - \frac{\hat{g}_Y^+(0) - \hat{g}_Y^-(0)}{\hat{g}_T^+(0) - \hat{g}_T^-(0)}. \] (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 realisation of random number \( e_i^* \) to preserve the correlation between \( Y_i \) and \( T_i \). The last step calculates the bias from the local linear estimator in the bootstrap by definition. Under Assumptions 2.1 and 2.2 and with \( B_1 \) 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 (2.3). 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,

\[ \frac{\hat{\xi}(h) - \Delta^*(h, b) - \xi}{V^{bc}(h, b)^{1/2}} \rightarrow^d N(0, 1), \] (3.2)

where \( \Delta^*(h, b) \) is defined by Equation (3.1).
Algorithm 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 $\hat{Z}_i$ 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 $k$th replication:

1. Draw i.i.d. random variables $\varepsilon_i^*$ with mean zero, variance one, and bounded fourth moments independent of the original data, and construct $\hat{\varepsilon}_{Z_i}^* = \hat{\varepsilon}_{Z_i} + \varepsilon_i^*$, $Z_i^* = \hat{g}_Z(X_i) + \varepsilon_i^*$, for all $i = 1, \ldots, n$. Note that both $\hat{\varepsilon}_{Z_i}$ and $\hat{\varepsilon}_{T_i}$ are multiplied by the same $\varepsilon_i^*$.

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:

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

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

3. Apply Algorithm 3.1 to the bootstrapped data set $(X_1, T_i^*, Y_i^*), \ldots, (X_n, T_n^*, Y_n^*)$ 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.

4. Save the estimator $\hat{\zeta}_i^*(h) = \frac{\hat{\mu}_{Z+}^*(h) - \hat{\mu}_{Z-}^*(h)}{\hat{\mu}_{T+}^*(h) - \hat{\mu}_{T-}^*(h)}$, and its bias $\Delta^*_i(h, b)$.

**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 Cumulative Distribution Function (CDF) of $\hat{\zeta}_{i+}^*(h) - \Delta_{i+}^*(h, b) - \zeta^*$, $\hat{\zeta}_{ij}^*(h) - \Delta_{ij}^*(h, b) - \zeta^*$ 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 Algorithm 3.1 is applied. The last step uses the empirical distribution of the bias-corrected estimator to construct CIs. As before, $B_2$ is assumed to be large enough 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,
\[
\forall^* (\hat{\zeta}^*(h) - \Delta^{**}(h, b))/V^{bc}(h, b) \to^p 1,
\]
and
\[
\sup_x \left| \Pr\left[ \frac{\hat{\zeta}^*(h) - \Delta^{**}(h, b) - \zeta^*}{\sqrt{V^*(\hat{\zeta}^*(h) - \Delta^{**}(h, b))}} \leq x \right] - \Pr\left[ \frac{\hat{\zeta}(h) - \Delta^{*}(h, b) - \zeta}{\sqrt{V^{bc}(h, b)}} \leq x \right] \right| \to^p 0.
\]

Theorem 3.2 enables one to construct CIs in the following form:
\[
(\hat{\zeta}(h) - \Delta^{*}(h, b) + \zeta^* - (\hat{\zeta}^*(h) - \Delta^{**}(h, b))_{1-a/2}, \hat{\zeta}(h) - \Delta^{*}(h, b) + \zeta^* + (\hat{\zeta}^*(h) - \Delta^{**}(h, b))_{a/2}),
\]
where all the terms with superscript * are defined in Algorithm 3.2. This CI is not centred at the bias-corrected point estimator.

Remark 3.1. The proposed bias correction differs from RBC’s formula in finite samples. Although the analytical bias is obtained by linearising \( E\left[ \hat{\tau}_Y(h) - \tau_Y \right] \) and then evaluating only its first-order terms, Algorithm 3.1 directly estimates \( E\left[ \hat{\tau}_Y(h) \right] \) through bootstrap. Both methods consistently estimate the bias. By avoiding the linearisation 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).

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 for two reasons. First, the wild bootstrap creates new residuals but leaves the regressors unchanged, which means the design matrices need to be computed only once even when they are repeatedly used in fitting local polynomials. Second, the number of data points actually used in estimation is a lot smaller than the full sample because of the local nature of the estimation.

4. MONTE CARLO SIMULATIONS

This section summarises 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 DGPs used and the 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

---

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

12 To fit local polynomials is equivalent to estimate weighted least squares—i.e., the estimated parameter is \((X'KX)^{-1}XKY\), where \(X\) is a matrix of regressors and \(K\) is a weighting matrix determined by kernel function. Both \(X\) and \(K\) are not affected by the bootstrap, so one need compute \((X'KX)^{-1}XK\) just once and then reuse it in the bootstrap calculations. Then, each bootstrap replication requires just a single matrix-vector multiplication.
Wild bootstrap for fuzzy regression discontinuity

(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 motivated this as an attempt to generate a plausible model with sizable distortion when a conventional t test is performed. The true treatment effects for these DGPs are $\zeta_1 = 0.04$, $\zeta_2 = -3.45$, and $\zeta_3 = 0.04$, 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 that affect the outcome. This is modelled by the correlation, $\rho$, 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 fuzzy RD, the plug-in RBC as implemented in the $rdrobust$ package described in Calonico et al. (2015), and the conventional estimators. Simulation results are presented in Table 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 in the empirical coverage and interval length columns. The properties of the point estimators in the first three columns, though 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 four to ten times larger bias. This is not surprising because 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 cost 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 Root Mean Square Error (RMSE).

Panels B and C present results for when the treatment is endogenous, which is likely the primary reason to choose RD designs as the identification strategy. The case with positive (negative) $\rho$ 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

---

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.

© 2020 Royal Economic Society.
Table 1. Empirical coverage and average interval length.

<table>
<thead>
<tr>
<th>DGP</th>
<th>Method</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>EC(%)</th>
<th>IL</th>
<th>$h_{CEOpt}$</th>
<th>$b_{CEOpt}$</th>
<th>$h_{MSE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wild bootstrap</td>
<td>0.010</td>
<td>0.057</td>
<td>0.058</td>
<td>94.9</td>
<td>0.217</td>
<td>0.140</td>
<td>0.323</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiplier bootstrap</td>
<td>0.003</td>
<td>0.070</td>
<td>0.074</td>
<td>93.3</td>
<td>0.274</td>
<td>0.140</td>
<td>0.323</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plug-in RBC</td>
<td>0.010</td>
<td>0.058</td>
<td>0.059</td>
<td>93.1</td>
<td>0.210</td>
<td>0.140</td>
<td>0.323</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>0.042</td>
<td>0.032</td>
<td>0.053</td>
<td>68.1</td>
<td>0.116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Wild bootstrap</td>
<td>0.026</td>
<td>0.062</td>
<td>0.067</td>
<td>91.7</td>
<td>0.234</td>
<td>0.117</td>
<td>0.299</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>Multiplier bootstrap</td>
<td>0.023</td>
<td>0.324</td>
<td>0.336</td>
<td>96.3</td>
<td>1.269</td>
<td>0.117</td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plug-in RBC</td>
<td>0.026</td>
<td>0.062</td>
<td>0.068</td>
<td>90.3</td>
<td>0.231</td>
<td>0.117</td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>0.215</td>
<td>0.079</td>
<td>0.229</td>
<td>2.6</td>
<td>0.186</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Wild bootstrap</td>
<td>0.003</td>
<td>0.059</td>
<td>0.059</td>
<td>95.4</td>
<td>0.231</td>
<td>0.115</td>
<td>0.317</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>Multiplier bootstrap</td>
<td>0.001</td>
<td>0.076</td>
<td>0.081</td>
<td>92.4</td>
<td>0.297</td>
<td>0.115</td>
<td>0.317</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plug-in RBC</td>
<td>0.003</td>
<td>0.059</td>
<td>0.059</td>
<td>94.1</td>
<td>0.224</td>
<td>0.115</td>
<td>0.317</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>-0.025</td>
<td>0.044</td>
<td>0.050</td>
<td>87.3</td>
<td>0.157</td>
<td></td>
<td></td>
<td>0.398</td>
</tr>
</tbody>
</table>

Panel A: $\rho = 0$

Panel B: $\rho = 0.9$

Panel C: $\rho = -0.9$

Note: EC denotes empirical coverage, and IL denotes average interval length based on 5,000 simulations; nominal coverage probabilities are 95% for each estimator. Sample size is 1,000. 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 bandwidths. The bootstrap approach uses $B_1 = 500$ replications to compute bias and $B_2 = 999$ replications to obtain the empirical distribution of the bias-corrected estimator.
correlation has little effect on the bias because 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 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 summarise, the wild bootstrap approach proposed in the present 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.  

### 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. Clustered data are very common in empirical studies, and there is a vast literature on handling it. 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) and 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 their availability in most statistical software packages. Their 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) pointed out that the existence of joint density of the running variable and the clustering variable ensures that cluster dependence vanishes asymptotically, not being captured by the usual approximations.

Bartalotti and Brummet (2017) developed 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) provided 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 \texttt{rdrobust} and \texttt{RDD} packages provide options to take this dependence into consideration. Both the \texttt{rdrobust} and \texttt{RDD} packages...

---

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

15 Another interesting extension of the bootstrap proposed in the present 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 by using the particular 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 the present paper.

16 See Wooldridge (2003), Cameron et al. (2011), and Cameron and Miller (2015) for an overview of this topic.

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

© 2020 Royal Economic Society.
used in the present paper offer the option to specify a clustering variable, as explained by Calonico et al. (2017). Chiang et al. (2019) offered 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) provided a comprehensive survey of bootstrap methods and showed 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 the present 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,

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

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.\(^{18}\)

Simulation results for \( G = 5, 10, 25 \) are reported in Table 2.\(^{19}\) 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 proposed wild bootstrap procedure 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 were estimated by using the ‘Maimonides’ rule’ as instrument.

As described by Angrist and Lavy (1999), the Maimonides’ rule holds that the maximum class size is 40. The rule 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. Because 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 fit into a fuzzy RD design.\(^{21}\)

We consider the first discontinuity in class size for the fourth grade. The sample used in this application includes 1,164 classes from schools with enrollments no larger than 80. The outcome variables are average verbal and math test scores at the class level. The discontinuities in class

---

\(^{18}\) The design ensures that the individual errors have the same standard errors as the baseline case presented in Section 4 for easy comparison.

\(^{19}\) \( G \) denotes the number of clusters on each side of the cutoff.

\(^{20}\) The data are available at [http://economics.mit.edu/faculty/angrist/data1/data/anglavy99](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.
Table 2. Empirical coverage and average interval length (clustered data).

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

Note: EC denotes empirical coverage, and IL denotes average interval length based on 5,000 simulations; nominal coverage probabilities are 95% for each estimator. Sample size is 1,000. 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 bandwidths. The bootstrap approach uses $B_1 = 500$ replications to compute bias and $B_2 = 999$ replications to obtain empirical distribution of the bias-corrected estimator.

Table 2 shows the results of empirical coverage and average interval length for different methods and DGP settings. The table includes columns for bias, standard deviation (SD), root mean squared error (RMSE), empirical coverage (EC), and average interval length (IL). 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 bandwidths. The bootstrap approach uses $B_1 = 500$ replications to compute bias and $B_2 = 999$ replications to obtain empirical distribution of the bias-corrected estimator.

The table reveals that the Wild bootstrap method provides the best balance between coverage and interval length, especially for larger DGP settings. The Plug-in RBC method also performs well, while the Conventional method appears to underestimate the bandwidth, leading to wider intervals and lower coverage.

Three methods are applied to estimate the effect of class size on average verbal/math scores, and results are shown in Table 3. The first column lists the original point estimates from 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

size and outcomes against enrollment are visualised 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 3. The first column lists the original point estimates from 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

© 2020 Royal Economic Society.
Y. He and O. Bartalotti

Figure 1. Class size (left), average verbal score (middle), and average math score (right).

Table 3. The effect of class size on average verbal score and average math score.

<table>
<thead>
<tr>
<th></th>
<th>Average treatment effect</th>
<th>95% CI</th>
<th>$h_{\text{CEOpt}}$</th>
<th>$b_{\text{CEOpt}}$</th>
<th>$h_{\text{MSE}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Corrected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Average verbal score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild bootstrap</td>
<td>$-0.495$</td>
<td>$-0.547$</td>
<td>$(-1.138, 0.213)$</td>
<td>$8.706$</td>
<td>$18.278$</td>
</tr>
<tr>
<td>Plug-in RBC</td>
<td>$-0.495$</td>
<td>$-0.568$</td>
<td>$(-1.212, 0.075)$</td>
<td>$8.706$</td>
<td>$18.278$</td>
</tr>
<tr>
<td>Conventional</td>
<td>$-0.488$</td>
<td></td>
<td>$(-1.104, 0.129)$</td>
<td></td>
<td>$7.952$</td>
</tr>
<tr>
<td>Panel B: Average math score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild bootstrap</td>
<td>$-0.198$</td>
<td>$-0.224$</td>
<td>$(-0.953, 0.586)$</td>
<td>$8.159$</td>
<td>$17.683$</td>
</tr>
<tr>
<td>Plug-in RBC</td>
<td>$-0.198$</td>
<td>$-0.251$</td>
<td>$(-1.002, 0.500)$</td>
<td>$8.159$</td>
<td>$17.683$</td>
</tr>
<tr>
<td>Conventional</td>
<td>$-0.202$</td>
<td></td>
<td>$(-0.802, 0.398)$</td>
<td></td>
<td>$9.200$</td>
</tr>
</tbody>
</table>

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 from 0.198 to 0.224-0.251 for average 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 those 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 that was established by Calonico et al. (2014) and has been 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

This paper is derived from Yang He’s work on his doctoral dissertation while at Iowa State University. The authors are grateful to Gray Calhoun for his invaluable insight, help, and support during the preparation of the original version of this manuscript. We thank Harold Chiang for kindly providing the code used in Chiang et al. (2019). We are also thankful to Marinho Bertanha, Beatriz Pereira, Ying-Ying Lee, Brent Kreider, Cindy Yu, the Editor, and four anonymous referees for valuable comments. Finally, Niklaus Julius provided invaluable research assistance during the paper revision process.

REFERENCES


© 2020 Royal Economic Society.


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher’s website.

Online Appendix
Replication package

*Co-editor Petra Todd handled this manuscript.*