The Econometrics Journal

Econometrics Journal (2020), volume **23**, pp. 192–210.
doi: 10.1093/ectj/utz022

Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs

SEBASTIAN CALONICO[†], MATIAS D. CATTANEO[‡] AND MAX H. FARRELL[††]

[†]*Columbia University, Department of Health Policy and Management, 722 West 168th Street, Office R482 New York, NY 10032, USA.*
sebastian.calonico@columbia.edu

[‡]*Princeton University, Department of Operations Research and Financial Engineering, 98 Charlton St, Sherred Hall 227, Princeton NJ 08540, USA.*
cattaneo@princeton.edu

[††]*University of Chicago, Booth School of Business, 5807 South Woodlawn Ave, USA.*
max.farrell@chicagobooth.edu

First version received: 8 October 2019; final version accepted: 16 October 2019.

Summary: Modern empirical work in regression discontinuity (RD) designs often employs local polynomial estimation and inference with a mean square error (MSE) optimal bandwidth choice. This bandwidth yields an MSE-optimal RD treatment effect estimator, but is by construction invalid for inference. Robust bias-corrected (RBC) inference methods are valid when using the MSE-optimal bandwidth, but we show that they yield suboptimal confidence intervals in terms of coverage error. We establish valid coverage error expansions for RBC confidence interval estimators and use these results to propose new inference-optimal bandwidth choices for forming these intervals. We find that the standard MSE-optimal bandwidth for the RD point estimator is too large when the goal is to construct RBC confidence intervals with the smaller coverage error rate. We further optimize the constant terms behind the coverage error to derive new optimal choices for the auxiliary bandwidth required for RBC inference. Our expansions also establish that RBC inference yields higher-order refinements (relative to traditional undersmoothing) in the context of RD designs. Our main results cover sharp and sharp kink RD designs under conditional heteroskedasticity, and we discuss extensions to fuzzy and other RD designs, clustered sampling, and pre-intervention covariates adjustments. The theoretical findings are illustrated with a Monte Carlo experiment and an empirical application, and the main methodological results are available in `R` and `Stata` packages.

Keywords: *Edgeworth expansions*, *coverage error*, *local polynomial regression*, *tuning parameter selection*, *treatment effects*.

JEL codes: *C14*, *C21*.

1. INTRODUCTION

The regression discontinuity (RD) design is widely used in program evaluation, causal inference, and treatment effect settings. (For general background on these settings, see Imbens and Rubin (2015) and Abadie and Cattaneo (2018), and references therein.) In recent years, RD has become one of the prime research designs for the analysis and interpretation of observational studies in

social, behavioural, biomedical, and statistical sciences. For introductions to RD designs, literature reviews, and background references, see Imbens and Lemieux (2008), Lee and Lemieux (2010), Cattaneo and Escanciano (2017), and Cattaneo et al. (2019, 2020).

Modern empirical work in RD designs often employs a mean square error (MSE) optimal bandwidth for local polynomial estimation of and inference on treatment effects.[1] This MSE-optimal bandwidth choice yields an MSE-optimal RD point estimator, but is by construction invalid for inference. Robust bias-corrected (RBC) inference methods provide a natural solution to this problem: RBC confidence intervals and related inference procedures remain valid even when the MSE-optimal bandwidth is used (Calonico et al., 2014, 2019). In this paper, we show that this choice of bandwidth is suboptimal when the goal is to construct RBC confidence intervals with minimal coverage error (CE), and we establish a new bandwidth choice delivering CE-optimal RBC confidence interval estimators or, analogously, minimizing the error in rejection probability of the associated hypothesis testing procedures for RD treatment effects.

Our main results are valid coverage error expansions for local polynomial RBC confidence interval estimators. The precise characterization offered by these expansions allows us to study bandwidth selection in detail, and to propose several novel bandwidth choices that are optimal for inference. First and foremost, we derive a CE-optimal bandwidth choice designed to minimize coverage error of the *interval* estimator, which is a fundamentally different goal than minimizing the MSE of the *point* estimator. The MSE- and CE-optimal bandwidths are therefore complementary, as both can be used in empirical work to construct, respectively, optimal point estimators and optimal inference procedures for RD treatment effects. For example, we find that in the case of the popular local linear RD estimator, if the sample size is $n = 500$, then shrinking the MSE-optimal bandwidth by approximately 27% leads to RBC confidence intervals with the fastest coverage error decay rate. Further, we use our expansions to derive bandwidth choices that trade off coverage error against interval length, which is conceptually analogous to trading size and power of the associated statistical tests, while retaining asymptotically correct coverage (or size control). Finally, by examining the leading constant terms of our coverage error expansions, we can deliver novel optimal choices for the auxiliary bandwidth required for RBC inference. We also provide plug-in, data-driven bandwidth selectors for use in practice, and illustrate their performance with real and simulated data.

Our theoretical results prove that RBC confidence interval estimators have coverage error strictly smaller (i.e., vanishing faster) than those of interval estimators based on undersmoothing, as long as enough smoothness of the underlying conditional expectation functions is available to at least characterize the MSE of the RD point estimator, the most natural case in empirical applications. RBC intervals are as good as their undersmoothed counterparts when no additional smoothness is available beyond what is needed to quantify the asymptotic bias of the t-test statistic. These results, coupled with our bandwidth selectors, provide precise theory-based guidance for empirical practice employing RD designs: RBC confidence interval estimators constructed with the CE-optimal, and even with the MSE-optimal, bandwidth choice dominate the alternative procedures in terms of coverage error performance.

Our main theoretical results focus on sharp RD designs with heteroskedastic data, covering both levels (standard sharp RD design) as well as derivatives (kink and higher-order RD designs). The latter case is of interest in, for example, Card et al. (2015, 2017), Dong and Lewbel (2015), Cerulli et al. (2017), and Ganong and Jäger (2018). We also discuss extensions to fuzzy, geographic, multi-

[1] See Imbens and Kalyanaraman (2012), Calonico et al. (2014), Arai and Ichimura (2016, 2018), Calonico et al. (2019), and references therein. Cattaneo and Vazquez-Bare (2016) give a general discussion of bandwidth/neighbourhood selection methods in RD designs.

score, and multi-cutoff RD designs (Hahn et al., 2001; Papay et al., 2011; Keele and Titiunik, 2015; Cattaneo et al., 2016), as well as to clustered data and/or inclusion of pre-intervention covariates (Lee and Card, 2008; Bartalotti and Brummet, 2017; Calonico et al., 2019). Our results can also be applied to other RD methods and settings such as those considered in Xu (2017), Dong (2019), Dong et al. (2019), Chiang et al. (2019), and He and Bartalotti (2019).

Finally, we remark that our discussion of inference-optimal bandwidth selection, as well as all treatments of MSE-optimal choices are within the context of local polynomial methods (Fan and Gijbels, 1996) under continuity assumptions of the underlying conditional expectation functions. CE- and MSE-optimal bandwidth choices should not be used when the goal is to employ local randomization assumptions in the context of RD designs (Cattaneo et al., 2015), because in this setting the underlying assumptions are different and the targeted neighbourhood around the cutoff is conceptually distinct. As such, the appropriate neighbourhood under local randomization cannot be generated by MSE- or CE-optimal bandwidth choices, and other methods are more appropriate: see Section 3 in Cattaneo et al. (2015) for one example. For further discussion of these different assumptions and methodologies, as well as comparisons between neighbourhood selectors, see Cattaneo and Vazquez-Bare (2016), Cattaneo et al. (2017), and Sekhon and Titiunik (2017).

The rest of the paper proceeds as follows. Section 2 presents the RD setup and gives a brief, but self-contained, introduction to standard estimation and inference methods. Section 3 gives the main results of the paper: valid higher-order coverage error expansions for commonly used confidence intervals as well as CE-optimal and related bandwidth choices. Section 4 discusses implementation and other practical issues. Section 5 briefly outlines several extensions, while numerical results using real and simulated data are reported in Section 6. Finally, Section 7 concludes. The supplemental appendix (SA, hereafter) contains all technical details and proofs, as well as more discussion of methodological, implementation, and numerical issues. Calonico et al. (2017) details general-purpose `Stata` and `R` software packages implementing our main methodological results.

2. SETUP

We assume that the researcher observes a random sample $(Y_i, T_i, X_i)'$, $i = 1, 2, \ldots, n$, where Y_i denotes the outcome variable of interest, T_i denotes treatment status, and X_i denotes an observed continuous score or running random variable, which determines treatment assignment for each unit in the sample. In the canonical sharp RD design, all units with X_i not smaller than a known threshold c are assigned to the treatment group and take up treatment, while all units with X_i smaller than c are assigned to the control group and do not take up treatment, so that $T_i = \mathbb{1}(X_i \geq c)$. Using the potential outcomes framework, $Y_i = Y_i(0) \cdot (1 - T_i) + Y_i(1) \cdot T_i$, with $Y_i(1)$ and $Y_i(0)$ denoting the potential outcomes with and without treatment, respectively, for each unit. The parameters of interest in sharp RD designs are either the average treatment effect at the cutoff or its derivatives:

$$\tau_\nu = \tau_\nu(c) = \frac{\partial^\nu}{\partial x^\nu} \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x] \Big|_{x=c},$$

where here and elsewhere we drop evaluation points of functions when it causes no confusion. With this notation, τ_0 corresponds to the standard sharp RD estimand, while τ_1 denotes the sharp kink RD estimand (up to scale). In Section 5, we discuss imperfect treatment compliance (i.e.,

fuzzy RD designs) and other extensions of this basic RD setup. Identification of τ_ν, as well as estimation and inference using local polynomial regression methods, proceed under the following standard regularity conditions.

ASSUMPTION 2.1 (RD) *For all $x \in [x_l, x_u]$, where $x_l < c < x_u$, and $t \in \{0, 1\}$: $\mathbb{E}[Y_i(t)|X_i = x]$ is $S \geq \min\{1, \nu\}$ times continuously differentiable with an S^{th} derivative that is Hölder-continuous with exponent $a \in (0, 1]$; the Lebesgue density of X_i, $f(x)$, and $\mathbb{V}[Y_i(t)|X_i = x]$ are positive and continuous; $\mathbb{E}[|Y_i(t)|^\delta|X_i = x]$, $\delta > 8$, is continuous; and the Lebesgue density of $(Y(t), X)$, $f_{y_t,x}(x)$, is positive and continuous.*

'Flexible' (i.e., nonparametric) local polynomial least squares estimators are the standard approach for estimation and inference in RD designs. The idea is to first choose a neighbourhood around the cutoff c via a positive bandwidth choice h, and then employ (local) weighted polynomial regression using only observations with score X_i lying within the selected neighbourhood. That is,

$$\hat{\tau}_\nu(h) = \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_{+,p}(h) - \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_{-,p}(h), \qquad \nu = 0, 1, 2, \ldots, p,$$

where $\boldsymbol{e}_\nu$ denotes the conformable $(\nu + 1)$th unit vector, and $\hat{\boldsymbol{\beta}}_{-,p}(h)$ and $\hat{\boldsymbol{\beta}}_{+,p}(h)$ correspond to the weighted least squares coefficients given by

$$\hat{\boldsymbol{\beta}}_{-,p}(h) = \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(c > X_i) \left(Y_i - \boldsymbol{r}_p(X_i - c)'\boldsymbol{\beta}\right)^2 K_h(X_i - c),$$

$$\hat{\boldsymbol{\beta}}_{+,p}(h) = \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(c \leq X_i) \left(Y_i - \boldsymbol{r}_p(X_i - c)'\boldsymbol{\beta}\right)^2 K_h(X_i - c),$$

with $\boldsymbol{r}_p(x) = (1, x, \cdots, x^p)'$ and $K_h(\cdot) = K(\cdot/h)/h$ for a kernel (weighting) function $K(\cdot)$. The kernel is assumed to obey the following regularity conditions.

ASSUMPTION 2.2 (KERNEL) $K(u) = \mathbb{1}(u < 0)k(-u) + \mathbb{1}(u \geq 0)k(u)$, *where $k(\cdot) : [0, 1] \mapsto \mathbb{R}$ is bounded and continuous on its support, positive (0,1), zero outside its support, and either is constant or $(1, K(u)\boldsymbol{r}_{3(p+1)}(u)')$ is linearly independent on $(-1, 1)$.*

The kernel and bandwidth serve to localize the regression fit near the cutoff. The choice of bandwidth, h, is the key parameter when implementing the RD estimator, and we discuss this choice in detail below. The most popular choices of kernel are the uniform kernel and the triangular kernel, which give equal weighting and linear down-weighting to the observations with $X_i \in [c - h, c + h]$, respectively. Finally, although our results cover any choice of $p \geq 0$, the preferred choice of polynomial order for point estimation is $p = 1$ (i.e., local-linear RD treatment effect estimator) because of the poor behaviour of higher-order polynomial approximations at or near boundary points. See Section 2.1.1 of Calonico et al. (2015) and Gelman and Imbens (2019) for more discussion.

2.1. MSE-optimal bandwidth choice and point estimation

Selecting the bandwidth h or, equivalently, the neighbourhood around the cutoff c, is challenging in applications. The default approach in modern empirical work is to minimize an approximation to the MSE of the point estimator $\hat{\tau}_\nu(h)$, or some other closely related quantity. Under standard regularity conditions, the conditional MSE of $\hat{\tau}_\nu(h)$ can be approximated as $h \to 0$ and $nh \to \infty$

as follows:

$$\mathbb{E}[(\hat{\tau}_\nu(h) - \tau_\nu)^2 | X_1, \ldots, X_n] \approx_{\mathbb{P}} h^{2p+2-2\nu} \mathscr{B}^2 + \frac{1}{nh^{1+2\nu}} \mathscr{V}, \tag{2.1}$$

where $\approx_{\mathbb{P}}$ denotes an approximation in probability (see the SA for a precise statement), and where $\mathscr{V}$ and $\mathscr{B}$ denote, respectively, approximations to the variance and bias of the $\hat{\tau}_\nu(h)$.

Using (), the MSE-optimal bandwidth choice for the RD treatment effect estimator $\hat{\tau}_\nu(h)$ is

$$h_{\mathtt{MSE}} = \left[\frac{(1+2\nu)\mathscr{V}}{2(1+p-\nu)\mathscr{B}^2} \right]^{1/(2p+3)} n^{-1/(2p+3)}, \tag{2.2}$$

where, of course, it is assumed that $\mathscr{B} \neq 0$. Further details and exact formulas are given in the SA to conserve space.

The infeasible MSE-optimal bandwidth choice $h_{\mathtt{MSE}}$ can be used to construct an MSE-optimal point estimator of the RD treatment effect τ_ν, given by $\hat{\tau}_\nu(h_{\mathtt{MSE}})$. In practice, because $\mathscr{V}$ and $\mathscr{B}$ involve unknown quantities, researchers rely on a plug-in estimator of the MSE-optimal bandwidth $h_{\mathtt{MSE}}$, say $\hat{h}_{\mathtt{MSE}}$, which is constructed by forming plug-in estimators $(\hat{\mathscr{V}}(b), \hat{\mathscr{B}}(b))$ of $(\mathscr{V}, \mathscr{B})$, for some preliminary bandwidth $b \to 0$; the formulas for $\hat{\mathscr{V}}(b)$ and $\hat{\mathscr{B}}(b)$ are also given in the SA. This approach gives a feasible, asymptotically MSE-optimal, RD point estimator $\hat{\tau}_\nu(\hat{h}_{\mathtt{MSE}})$, and is commonly used in empirical work. All other MSE-optimal bandwidth choices available in the literature are also proportional to $n^{-1/(2p+3)}$, where the factor of proportionality depends on the specific MSE objective function being optimized and/or other specific methodological choices. See Imbens and Kalyanaraman (2012), Calonico et al. (2014), Arai and Ichimura (2016, 2018), and Calonico et al. (2019) for concrete examples, and Cattaneo and Vazquez-Bare (2016) for more general discussion.

2.2. *Robust bias-corrected inference*

The infeasible estimator $\hat{\tau}_\nu(h_{\mathtt{MSE}})$ and its data-driven counterpart $\hat{\tau}_\nu(\hat{h}_{\mathtt{MSE}})$ are MSE-optimal point estimators of τ_ν in large samples. In empirical work, these point estimators are used not only to construct the 'best guess' of the unknown RD treatment effect τ_ν, but also to conduct statistical inference, in particular for forming confidence intervals for τ_ν. The standard approach employs a Wald test statistic under the null hypothesis, and inverts it to form the confidence intervals. Specifically, for some choice of bandwidth h, the naïve t-test statistic takes the form

$$T(h) = \frac{\hat{\tau}_\nu(h) - \tau_\nu}{\sqrt{\hat{\mathscr{V}}(h)/(nh^{1+2\nu})}},$$

where it is assumed that $T(h) \sim \mathcal{N}(0, 1)$, at least in large samples, and hence the corresponding confidence interval estimator for τ_ν is

$$I_{\mathtt{US}}(h) = \left[\hat{\tau}_\nu(h) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\mathscr{V}}(h)}{nh^{1+2\nu}}} \ , \ \hat{\tau}_\nu(h) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\mathscr{V}}(h)}{nh^{1+2\nu}}} \right],$$

where z_α denotes the (100α)-percentile of the standard normal distribution. Crucially, the confidence interval $I_{\mathtt{US}}(h)$ will only have correct asymptotic coverage, in the sense of $\mathbb{P}[\tau_\nu \in I_{\mathtt{US}}(h)] = 1 - \alpha + o(1)$, if h obeys $nh^{2p+3} \to 0$, that is, the bandwidth is 'small enough'. In particular, the MSE-optimal bandwidth is 'too large': it is easy to show that $\mathbb{P}[\tau_\nu \in I_{\mathtt{US}}(h_{\mathtt{MSE}})] \not\to 1 - \alpha$, rendering inference and confidence intervals based on the naïve t-test statistic $T(h_{\mathtt{MSE}})$ invalid.

An approach to resolve the invalidity of the confidence interval $I_{\texttt{US}}(h_{\texttt{MSE}})$ is to undersmooth (hence the 'US' notation) by selecting a bandwidth 'smaller' than $h_{\texttt{MSE}}$, or than $\hat{h}_{\texttt{MSE}}$ in practice, when constructing the interval estimator. This approach, however, has at least two empirical and theoretical drawbacks: (i) interval length is enlarged (that is, power is decreased) because fewer observations are used; and (ii) undersmoothing is suboptimal in terms of the coverage error of $I_{\texttt{US}}(h)$. The first drawback is methodologically obvious, and we will discuss it further after the new CE-optimal bandwidth choice is presented. The second drawback is formally established for RD designs in the following section as part of our main results, using novel valid coverage error expansions.

Bias correction is an alternative to undersmoothing. In the context of RD designs, Calonico et al. (2014) introduced a robust bias-correction method to conduct statistical inference in general, and to form confidence intervals in particular, which in its simplest form is given as follows:

$$T_{\texttt{RBC}}(h) = \frac{\hat{\tau}_{\nu,\texttt{BC}}(h) - \tau_\nu}{\sqrt{\hat{\mathscr{V}}_{\texttt{BC}}(h)/(nh^{1+2\nu})}}, \qquad \hat{\tau}_{\nu,\texttt{BC}}(h) = \hat{\tau}_\nu(h) - h^{1+p-\nu}\hat{\mathscr{B}}(b),$$

and

$$I_{\texttt{RBC}}(h) = \left[\, \hat{\tau}_{\nu,\texttt{BC}}(h) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\mathscr{V}}_{\texttt{BC}}(h)}{nh^{1+2\nu}}}\,,\ \hat{\tau}_{\nu,\texttt{BC}}(h) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\mathscr{V}}_{\texttt{BC}}(h)}{nh^{1+2\nu}}} \,\right],$$

where again exact formulas for $\hat{\mathscr{B}}(b)$ and $\hat{\mathscr{V}}_{\texttt{BC}}(h)$ are discussed in the SA. For inference, a key feature is that $\hat{\mathscr{V}}_{\texttt{BC}}(h)$ is an estimator of the variance of $\hat{\tau}_{\nu,\texttt{BC}}(h)$, not of the variance of $\hat{\tau}_\nu(h)$. For implementation, $\hat{\mathscr{B}}(b)$ depends on a local polynomial regression of order $p+1$.

An important empirical and theoretical property of $I_{\texttt{RBC}}(h)$ is that $\mathbb{P}[\tau_\nu \in I_{\texttt{RBC}}(h_{\texttt{MSE}})] \to 1-\alpha$, where the same bandwidth is used for both (optimal) point estimation and (suboptimal yet valid) statistical inference. Furthermore, Calonico et al. (2014) showed that the interval estimator remains valid under a wider set of bandwidth sequences, even when minimal additional smoothness of the unknown regression functions is assumed, and it was found to perform much better than other methods in both simulations and replication studies (Ganong and Jäger, 2018; Hyytinen et al., 2018). In this paper we offer principled, theoretical results that explain the good numerical properties of $I_{\texttt{RBC}}(h)$, and we also provide new concrete ways to improve its implementation further. In the upcoming sections we present the following main results.

(a) We establish that $I_{\texttt{RBC}}(h)$ has asymptotic coverage error that is no larger than $I_{\texttt{US}}(h)$, and is strictly smallern most practically relevant cases, even when the corresponding best possible bandwidth is used to construct each confidence interval.
(b) We show that employing the MSE-optimal bandwidth $h_{\texttt{MSE}}$ to construct $I_{\texttt{RBC}}(h)$ is valid but suboptimal in terms of coverage error.
(c) We derive new optimal bandwidth choices that minimize the coverage error of the RBC confidence intervals. We discuss the consequences for interval length and how length can be further optimized, including automatic, optimal auxiliary bandwidths.

We also discuss the implications of these results for empirical work and explore them numerically with real and simulated data.

3. MAIN RESULTS

Our main theoretical results are valid coverage error expansions for both $I_{\text{US}}(h)$ and $I_{\text{RBC}}(h)$. These are based on generic, valid Edgeworth expansions in the context of RD designs, which could be used for other purposes, such as studying the error in rejection probability of hypothesis tests. The generic results, and other technical details, are given in the SA.

To state our first main result, recall that $I_{\text{US}}(h)$ is constructed using $\hat{\mathscr{V}}(b)$, while $I_{\text{RBC}}(h)$ is constructed using both $\hat{\mathscr{B}}(b)$ and $\hat{\mathscr{V}}_{\text{BC}}(b)$, all of which are precisely described in the SA (heuristically, they are consistent estimators of higher-order biases and variances of the RD point estimator). In particular, b denotes the bandwidth used to construct the bias-correction estimate $\hat{\mathscr{B}}(b)$ and the associated variance estimate $\hat{\mathscr{V}}_{\text{BC}}(b)$. An important quantity is $\rho = h/b$, which we discuss in detail further below.

THEOREM 3.1 (COVERAGE ERROR EXPANSIONS) *Suppose Assumptions 2.1 and 2.2 hold, that* $nh^{1+2v}/\log(nh)^{2+\eta} \to \infty$ *for* $\eta > 0$, *and* $\rho = h/b$ *is bounded and bounded away from zero.*

(a) If $S \geq p+1$ *and* $nh^{2p+3}\log(nh)^{1+\eta} \to 0$, *then*

$$\mathbb{P}[\tau_v \in I_{US}(h)] - (1-\alpha) = \frac{1}{nh}\mathscr{Q}_{US,1} + nh^{3+2p}\mathscr{Q}_{US,2} + h^{1+p}\mathscr{Q}_{US,3} + \epsilon_{US}$$

$$\mathbb{P}[\tau_v \in I_{RBC}(h)] - (1-\alpha) = \frac{1}{nh}\mathscr{Q}_{RBC,1} + \epsilon_{US},$$

where $\epsilon_{US} = o(n^{-1}h^{-1}) + O(nh^{3+2p+2a} + h^{1+p+a})$.

(b) If $S \geq p+2$ *and* $nh^{2p+5}\log(nh)^{1+\eta} \to 0$, *then*

$$\mathbb{P}[\tau_v \in I_{RBC}(h)] - (1-\alpha) = \frac{1}{nh}\mathscr{Q}_{RBC,1} + nh^{5+2p}\mathscr{Q}_{RBC,2} + h^{2+p}\mathscr{Q}_{RBC,3} + \epsilon_{RBC},$$

where $\epsilon_{RBC} = o(n^{-1}h^{-1}) + O(nh^{5+2p+2a} + h^{2+p+a})$.

The n-varying, bounded quantities $(\mathscr{Q}_{US,\ell}, \mathscr{Q}_{RBC,\ell})$, $\ell = 1, 2, 3$, *are cumbersome and hence further discussed in the SA.*

This theorem establishes higher-order coverage error characterizations for the RD confidence intervals $I_{\text{US}}(h)$ and $I_{\text{RBC}}(h)$, under two distinct smoothness regimes, controlled by S. (Coverage error expansions under $\rho \to 0$ are given in the SA because they require additional regularity conditions.) In the first case, described in part (a), the two confidence intervals are compared when the same level of smoothness is allowed. Specifically, we consider the setting where smoothness is exhausted after the leading higher-order terms of the RD point estimator $\hat{\tau}_v$ are characterized, which is the minimal smoothness needed to compute the MSE-optimal bandwidth h_{MSE}, as commonly done in practice (see (2.1) and ()). Thus, in this regime, $I_{\text{RBC}}(h_{\text{MSE}})$ can be formed, but no additional smoothness is available, which gives the least favourable setting for robust bias-correction techniques. Part (a) shows, nonetheless, that even in this case, $I_{\text{RBC}}(h)$ is never worse in terms of asymptotic coverage error than $I_{\text{US}}(h)$, an important practical and theoretical result.

From a practical point of view, researchers first select a polynomial order (usually $p = 1$), and then form confidence intervals using some bandwidth choice (often an empirical implementation of h_{MSE}). It is rarely the case that the underlying regression functions are not smoother than what is exploited by the procedure. Part (b) discusses this case, and shows that $I_{\text{RBC}}(h)$ is strictly superior to $I_{\text{US}}(h)$ in terms of coverage error rates when additional smoothness is available. To

be specific, comparing parts (a) and (b), it is shown that the coverage error of $I_{\text{RBC}}(h)$ vanishes faster than that of $I_{\text{US}}(h)$. This result gives strong theoretical justification for employing $I_{\text{RBC}}(h)$ in empirical work.

The derivations in the SA also show that both $I_{\text{US}}(h)$ and $I_{\text{RBC}}(h)$ exhibit higher-order boundary carpentry thanks to the specific fixed-n variance estimators used; see Calonico et al. (2018) for more discussion. This result is empirically important because it shows that the good boundary properties possessed by local polynomial estimators in point estimation carry over to inference under proper Studentization. Thus, our results formalize the crucial importance of using fixed-n standard error formulas, as sometimes implemented in software for RD designs (Calonico et al., 2017).

Finally, the expansions given in Theorem 3.1, as well as the underlying technical work presented in the SA, are new to the literature. They cannot be deduced from results already available (Calonico et al., 2018, 2019) because they apply to the difference of two local polynomial estimates, $\hat{\tau}_\nu(h) = \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_{+,p}(h) - \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_{-,p}(h)$, and higher-order terms of these differences are not trivially expressible as differences or sums of terms for each component, unlike the case when analysing first-order asymptotic approximations or MSE expansions. It is possible to upgrade the results in the SA to also show that $I_{\text{RBC}}(h)$ is a coverage error optimal confidence interval estimator, uniformly over empirically relevant classes of data-generating processes, employing the optimality framework presented in Calonico et al. (2019). We do not provide details on this result for brevity.

3.1. CE-optimal bandwidths and methodological implications

We now employ Theorem 3.1 to develop a CE-optimal bandwidth choice for RD designs. This bandwidth choice will be made feasible in Section 4, where we address implementation issues in detail. The following theorem is our second main result.

THEOREM 3.2 (COVERAGE ERROR OPTIMALITY) *Suppose the conditions of Theorem 3.1(b) hold. If $\mathcal{Q}_{RBC,2} \neq 0$ or $\mathcal{Q}_{RBC,3} \neq 0$, then the robust bias-corrected CE-optimal confidence interval is $I_{RBC}(h_{RBC})$, where*

$$h_{RBC} = \mathscr{H}\, n^{-1/(3+p)}, \qquad \mathscr{H} = \operatorname{argmin}_{H>0} \left| \frac{1}{H}\mathcal{Q}_{RBC,1} + H^{5+2p}\mathcal{Q}_{RBC,2} + H^{2+p}\mathcal{Q}_{RBC,3} \right|.$$

The coverage obeys $\mathbb{P}[\tau_\nu \in I_{RBC}(h_{RBC})] = 1 - \alpha + O(n^{-(2+p)/(3+p)})$.

This theorem gives the CE-optimal bandwidth choice, h_{RBC}, and the corresponding CE-optimal RBC confidence interval estimator, $I_{\text{RBC}}(h_{\text{RBC}})$. The optimal rate for the bandwidth sequence is $h_{\text{RBC}} \propto n^{-1/(3+p)}$, along with the associated optimal constant $\mathscr{H}$, which cannot be given in closed form (c.f. (2.2)). An analogous result is given in the SA for $I_{\text{US}}(h)$, where it is shown that the corresponding CE-optimal bandwidth choice is $h_{\text{US}} \propto n^{-1/(2+p)}$, and with a different constant of proportionality. Furthermore, this shows that $\mathbb{P}[\tau_\nu \in I_{\text{US}}(h_{\text{US}})] = 1 - \alpha + O(n^{-(1+p)/(2+p)})$, and therefore the RBC confidence interval estimator $I_{\text{RBC}}(h_{\text{RBC}})$ has a faster coverage error rate than the best possible undersmoothed confidence interval $I_{\text{US}}(h_{\text{US}})$.

Our results establish that $h_{\text{MSE}} \neq h_{\text{RBC}} \neq h_{\text{US}}$ in rates (and constants, of course) for all $p \geq 1$, and $h_{\text{MSE}} \asymp h_{\text{RBC}} \neq h_{\text{US}}$ for $p = 0$. That is, a bandwidth different from the MSE-optimal one should be used when the goal is to construct confidence intervals with small asymptotic coverage error whenever $p \geq 1$. More generally, focusing on the bandwidth choice and its consequences for

coverage (interval length is addressed in the next section), we can offer three key methodological conclusions for inference in RD designs.

(1) **MSE-optimal bandwidth**. In this case, the researcher chooses $h = h_{\texttt{MSE}} \propto n^{-1/(3+2p)}$. This choice of bandwidth is simple and very popular, but leads to first-order bias, rendering $I_{\texttt{US}}(h)$ invalid. On the other hand, $T_{\texttt{RBC}}(h_{\texttt{MSE}}) \sim \mathcal{N}(0, 1)$ in large samples, and hence $I_{\texttt{RBC}}(h)$ is still asymptotically valid. Theorem 3.1 quantifies the rate of coverage error decay precisely, and we find:

$$\mathbb{P}[\tau_\nu \in I_{\texttt{US}}(h_{\texttt{MSE}})] - (1-\alpha) \asymp 1,$$
$$\mathbb{P}[\tau_\nu \in I_{\texttt{RBC}}(h_{\texttt{MSE}})] - (1-\alpha) \asymp n^{-\min\{2,2+p\}/(2+p)}.$$

(2) **CE-optimal bandwidth for $I_{\texttt{US}}(h)$**. While ad-hoc undersmoothing of $h_{\texttt{MSE}}$ is a possible method for fixing the first-order coverage distortion of $I_{\texttt{US}}(h)$, a more theoretically founded choice is to use $h = h_{\texttt{US}} \propto n^{-1/(2+p)}$, which is also a valid choice for $I_{\texttt{RBC}}(h)$. In fact, this choice yields the same coverage error rate for both intervals:

$$\mathbb{P}[\tau_\nu \in I_{\texttt{US}}(h_{\texttt{US}})] - (1-\alpha) \asymp n^{-(1+p)/(2+p)},$$
$$\mathbb{P}[\tau_\nu \in I_{\texttt{RBC}}(h_{\texttt{US}})] - (1-\alpha) \asymp n^{-(1+p)/(2+p)}.$$

(3) **CE-optimal bandwidth for $I_{\texttt{RBC}}(h)$**. Finally, the researcher can also choose $h = h_{\texttt{RBC}} \propto n^{-1/(3+p)}$. This bandwidth choice is again too 'large' for $I_{\texttt{US}}(h)$, and hence leads to a first-order coverage distortion, but is optimal for $I_{\texttt{RBC}}(h)$:

$$\mathbb{P}[\tau_\nu \in I_{\texttt{US}}(h_{\texttt{RBC}})] - (1-\alpha) \asymp 1,$$
$$\mathbb{P}[\tau_\nu \in I_{\texttt{RBC}}(h_{\texttt{RBC}})] - (1-\alpha) \asymp n^{-(2+p)/(3+p)}.$$

The first point formalizes that an MSE-optimal bandwidth is always a valid choice for robust bias-correction inference, with the coverage error rates depending on the polynomial order p. Crucially, for any $p \geq 1$, the robust bias-corrected interval $I_{\texttt{RBC}}(h_{\texttt{MSE}})$ will never achieve the fastest decay in coverage error, and therefore $h_{\texttt{MSE}}$ must always be undersmoothed if the goal is to construct confidence intervals for the RD treatment effect with the fastest vanishing coverage error rate. In Sections 4 and 5, we employ this insight to propose simple rule-of-thumb CE-optimal bandwidth choices.

The last two points above re-emphasize the advantages of RBC inference: $I_{\texttt{US}}(h_{\texttt{US}})$ and $I_{\texttt{RBC}}(h_{\texttt{US}})$ exhibit the same coverage error rates, which are suboptimal relative to $I_{\texttt{RBC}}(h_{\texttt{RBC}})$. In other words, $I_{\texttt{RBC}}(h_{\texttt{RBC}})$ should be preferred to all the other alternatives discussed above, when the goal is to construct CE-optimal confidence intervals in RD designs where smoothness of the underlying regression functions is not binding. This is one of the main theoretical and practical findings of this paper.

3.2. Interval length

An obvious concern is that the improvements in coverage offered by robust bias correction may come at the expense of larger (average) interval length. However, we now show that this is not the case. By symmetry, the (squared) length of the intervals $I_{\texttt{US}}(h)$ and $I_{\texttt{RBC}}(h)$ take the same form:

$$|I_{\texttt{US}}(h)|^2 = 4 \cdot z_{\frac{\alpha}{2}}^2 \cdot \frac{\hat{\mathscr{V}}(h)}{nh^{1+2\nu}} \qquad \text{and} \qquad |I_{\texttt{RBC}}(h)|^2 = 4 \cdot z_{\frac{\alpha}{2}}^2 \cdot \frac{\hat{\mathscr{V}}_{\texttt{BC}}(h)}{nh^{1+2\nu}}.$$

Table 1. L_2-optimal ρ.

p	Kernel		
	Triangular	Epanechnikov	Uniform
0	0.778	0.846	1
1	0.850	0.898	1
2	0.887	0.924	1
3	0.909	0.940	1
4	0.924	0.950	1

Note: Computed by minimizing the L_2 distance between the RBC induced equivalent kernel and the optimal variance-minimizing equivalent kernel obtained by Cheng et al. (1997) for $\nu = 0$.

Thus, comparing asymptotic length amounts to examining the rate of contraction, $n^{-1}h^{-1-2\nu}$, and the limiting variance constants, $\hat{\mathscr{V}}(h) \to_{\mathbb{P}} \mathscr{V}$ and $\hat{\mathscr{V}}_{\mathtt{BC}}(h) \to_{\mathbb{P}} \mathscr{V}_{\mathtt{BC}}$, which we show in the SA depend on the 'equivalent kernel' function induced by the choice of $K(\cdot)$ and ρ (and p). See Fan and Gijbels (1996, Section 3.2.2) for more discussion on equivalent kernels in local polynomial estimation.

First, regarding the contraction rate of the confidence intervals, the formal comparison follows directly from the discussion above: robust bias correction can accommodate, and will optimally employ, a slower vanishing bandwidth (i.e., h is 'larger') than undersmoothing, and hence $I_{\mathtt{RBC}}(h)$ will contract more quickly (i.e., $nh^{1+2\nu} \to \infty$ faster than with undersmoothing). This result formalizes the heuristic idea that using a larger bandwidth leads to more observations being used and hence improved power. To be precise, we have $|I_{\mathtt{RBC}}(h_{\mathtt{RBC}})|^2 \asymp n^{-(2+p)/(3+p)}$ compared with $|I_{\mathtt{US}}(h_{\mathtt{US}})|^2 \asymp n^{-(1+p)/(2+p)}$. It is also instructive to note that $|I_{\mathtt{RBC}}(h_{\mathtt{MSE}})|^2 \asymp n^{-(2p+2)/(2+p)}$ and $|I_{\mathtt{RBC}}(h_{\mathtt{US}})|^2 \asymp n^{-(1+p)/(2+p)}$, which agrees with the above discussion regarding the impact of using $h_{\mathtt{MSE}}$, $h_{\mathtt{US}}$, and $h_{\mathtt{RBC}}$ to construct the interval estimators. The intervals $I_{\mathtt{US}}(h_{\mathtt{MSE}})$ and $I_{\mathtt{US}}(h_{\mathtt{RBC}})$ do not have correct asymptotic coverage.

Second, it is possible to optimize the asymptotic variance constant entering the length of the RBC confidence interval, as a function of $K(\cdot)$ and the quantity $\rho = h/b$. We can then select these two optimally to minimize the asymptotic constant portion of interval length. Specifically, Cheng et al. (1997) show that the asymptotic variance of a local polynomial point estimator at a boundary point is minimized by employing the uniform kernel $K(u) = \mathbb{1}(|u| \leq 1)$. If $I_{\mathtt{RBC}}(h)$ is formed choosing $K(u)$ to be uniform, it follows immediately that $\rho = 1$ is optimal, as with this choice the induced equivalent kernel becomes pointwise equal to the optimal equivalent kernel. For other choices of kernel $K(\cdot)$ we can derived the optimal choice of ρ, depending on p, by minimizing the L_2 distance between the induced equivalent kernel and the optimal variance-minimizing equivalent kernel. See the SA for all technical details.

In particular, for $\nu = 0$, we computed the L_2-optimal ρ for two popular kernels in RD applications: the triangular kernel $K(u) = (1 - |u|)\mathbb{1}(|u| \leq 1)$, which Cheng et al. (1997) show is MSE-optimal (i.e., optimal from a point estimation perspective), and the Epanechnikov kernel. Table 1 gives the results. These ρ^* optimal choices do not depend on the data, and thus are immediately implementable. For example, in the leading empirical case of $p = 1$ and triangular weighting, $\rho^* = 0.850$ is the recommended choice minimizing the asymptotic variance and hence the interval length of $I_{\mathtt{RBC}}(h)$. We explore the numeric properties of these choices in Section 6.

4. DATA-DRIVEN IMPLEMENTATIONS

We now discuss several implementable CE-optimal and related bandwidth selectors, building on our theoretical and methodological results. We focus exclusively on data-driven implementations of $I_{\mathtt{RBC}}(h)$, that is, in constructing a data-driven version of $h_{\mathtt{RBC}}$ and other related bandwidth selectors for RBC inference. We first present two main approaches to selecting the CE-optimal bandwidth choice: (a) a rule-of-thumb (ROT) based on an implementation of the MSE-optimal choice $h_{\mathtt{MSE}}$, generically denoted by $\hat{h}_{\mathtt{MSE}}$; and (b) a direct plug-in (DPI) rule based on estimating the unknown quantities $\mathscr{Q}_{\mathtt{RBC},\ell}$, $\ell = 1, 2, 3$, and solving the optimization problem in Theorem 3.2. We then discuss other choices that trade off coverage error and interval length, leveraging our coverage error expansions (Theorem 3.1).

The discussion below focuses on the main bandwidth h, which is the crucial choice in applications. For $\rho = h/b$, i.e., the auxiliary bandwidth b, we consider three choices: (a) $\rho = 1$, for any kernel, which corresponds to the practically relevant case $h = b$; (b) $\rho = \rho^*$ discussed above (Table 1); and (c) $\rho = h/b$ estimated from the data by replacing h and b with plug-in estimators, $\hat{h}_{\mathtt{MSE}}$ and $\hat{b}_{\mathtt{MSE}}$, of the MSE-optimal choices for the point estimator and the bias correction, respectively. The first two choices of ρ are fully automatic once h is chosen; the third requires a data-driven implementation of b as well. The form of $b_{\mathtt{MSE}}$ can be found by selecting (ν, p) appropriately and referring to (2.2). For example, for τ_0 and $p = 1$, $(\nu, p) = (2, 2)$ when a quadratic approximation is used for bias correction. See the SA for details.

4.1. ROT bandwidth choice

A simple strategy to construct a feasible bandwidth selector that yields the optimal coverage error decay rate is to rescale an existing choice so that the rate agrees with $h_{\mathtt{RBC}}$. We call this the ROT approach. For $\hat{h}_{\mathtt{MSE}}$, a data-driven implementation of $h_{\mathtt{MSE}}$, we simply set

$$\hat{h}^{\mathtt{rot}}_{\mathtt{RBC}} = n^{-p/((2p+3)(p+3))}\, \hat{h}_{\mathtt{MSE}}.$$

It is immediate that $\hat{h}^{\mathtt{rot}}_{\mathtt{RBC}} \propto h_{\mathtt{RBC}}$, and therefore this empirical choice has the optimal rate of decay and yields an interval $I_{\mathtt{RBC}}(\hat{h}^{\mathtt{rot}}_{\mathtt{RBC}})$ with the fastest possible coverage error decay. As an example, for the popular local-linear RD estimator ($p = 1$) and a sample of size $n = 500$, the MSE-optimal bandwidth selector $\hat{h}_{\mathtt{MSE}}$ is shrunk by $100(1 - n^{-1/20})\% \approx 27\%$ to obtain RBC confidence intervals with the fastest coverage error decay rate.

Feasible MSE-optimal bandwidths are widely available in software: see Calonico et al. (2017), and references therein, for second-generation plug-in choices satisfying $\hat{h}_{\mathtt{MSE}}/h_{\mathtt{MSE}} \to_{\mathbb{P}} 1$. Following this, ρ is selected according to the options above ($\rho = 1$, $\rho = \rho^*$, or $\rho = \hat{\rho}$). It is worth noting that, despite the constants being suboptimal in this approach, the 'direction' of the trade-off is still correct in the sense that if the bias is small relative to higher moments, the CE- and MSE-optimal bandwidths will increase, and $\hat{h}^{\mathtt{rot}}_{\mathtt{RBC}}$ reflects this.

4.2. DPI bandwidth choice

Our second approach to constructing fully data-driven CE-optimal bandwidth choices employs plug-in estimators of the unknown constants underlying $h_{\mathtt{RBC}}$ in Theorem 3.2. While this bandwidth choice does not have a closed-form solution in general, it is easy to form plug-in (consistent) estimators of the quantities $\mathscr{Q}_{\mathtt{RBC},\ell}$, $\ell = 1, 2, 3$, for any ν, p, kernel, and ρ. Given these estimators,

the DPI bandwidth selector yielding CE-optimal RBC inference is

$$\hat{h}_{\mathrm{RBC}} = \hat{\mathscr{H}}\, n^{-1/(3+p)}, \qquad \hat{\mathscr{H}} = \operatorname{argmin}_{H>0} \left| \frac{1}{H} \hat{\mathscr{Q}}_{\mathrm{RBC},1} + H^{5+2p} \hat{\mathscr{Q}}_{\mathrm{RBC},2} + H^{2+p} \hat{\mathscr{Q}}_{\mathrm{RBC},3} \right|,$$

where $\hat{\mathscr{Q}}_{\mathrm{RBC},\ell} \to_{\mathbb{P}} \mathscr{Q}_{\mathrm{RBC},\ell}$, $\ell = 1, 2, 3$, are discussed in the SA. Again, ρ is chosen afterwards according to the three options above ($\rho = 1$, $\rho = \rho^*$, or $\rho = \hat{\rho}$).

Estimating the quantities $\mathscr{Q}_{\mathrm{RBC},\ell}$, $\ell = 1, 2, 3$, is straightforward. These are expressed in pre-asymptotic form, so constructing the estimators boils down to replacing marginal expectations by sample averages and employing pilot bandwidth choices. Natural choices of pilot bandwidths are the corresponding MSE-optimal bandwidth selectors, already implemented in the literature. It is easy to show (see the SA for discussion) that, under regularity conditions, the DPI bandwidth selector will be consistent in the sense that $\hat{h}_{\mathrm{RBC}}/h_{\mathrm{RBC}} \to_{\mathbb{P}} 1$. The resulting data-driven RBC confidence intervals will be CE-optimal, given the choice of point estimator and enough smoothness of the unknown regression functions.

4.3. Coverage error and interval length trade-off

It is natural to have a preference for shorter intervals that still have good coverage properties. Our main results allow us to discuss formally such a trade-off, and to propose alternative bandwidth choices reflecting it. Larger bandwidths (i.e., smaller values of γ when $h = Hn^{-\gamma}$) yield on average shorter intervals: as already highlighted, one of the strengths of RBC inference is that it allows for, and will optimally employ, a larger bandwidth relative to the best undersmoothing procedure. Thus, we may seek to use a bandwidth larger than h_{RBC} that reduces interval length, while still retaining good coverage properties.

We consider the generic bandwidth choice $h_{\mathrm{TO}} = H_{\mathrm{TO}} n^{-\gamma_{\mathrm{TO}}}$, for constants $H_{\mathrm{TO}} > 0$ and $\gamma_{\mathrm{TO}} > 0$, where 'TO' stands for 'trade-off'. First we set the exponent γ_{TO}. For valid inference, Theorem 3.1 requires that γ_{TO} lie in $(1/(5+2p), 1)$, and Theorem 3.2 gives $h_{\mathrm{RBC}} \asymp n^{-1/(p+3)}$. For any bandwidth smaller than this ($h \ll h_{\mathrm{RBC}}$), both coverage error and length can be reduced with a larger bandwidth, and hence we restrict attention to

$$\frac{1}{5+2p} < \gamma_{\mathrm{TO}} \leq \frac{1}{3+p}. \tag{4.1}$$

Any choice in this range is valid in the sense that coverage error vanishes asymptotically and length is reduced compared with what $\gamma_{\mathrm{RBC}} = 1/(p+3)$ would give.

To choose the constant H_{TO} we characterize more precisely the trade-off we are making. It is perhaps not surprising that this will be about balancing, in a certain way, bias- and variance-type terms. This is also true for CE and MSE minimization, because all three methods deal with, at heart, similar fundamental quantities, but in every case the specific manifestation is different. The particulars in this case are described as follows.

Recall from Section 3.2 that the length of $I_{\mathrm{RBC}}(h)$ does not depend on the bias, only on the variance, and, more precisely, scales as the standard deviation. Thus, squared length is proportional to variance and is therefore analogous to the first term in coverage error, which captures variance (and other centred moment) errors, but not bias. Furthermore, for the range in (4.3), the third term of coverage error is of higher order relative to the other two. Therefore, we can view a trade-off of interval length against coverage error as comparing the second term of coverage error (the squared scaled bias) against a variance-type term: the square of interval length, which not only changes the constants involved but also properly adjusts for any $\nu \geq 0$

because $|I_{\text{RBC}}(h)|^2 \asymp n^{-1}h^{-1-2\nu}$. The leading constant portions of coverage error and length are $H^{5+2p}\hat{\mathscr{Q}}_{\text{RBC},2}$ and $4z^2_{\frac{\alpha}{2}}\hat{\mathscr{V}}_{\text{BC}}H^{-1-2\nu}$, respectively, where $\hat{\mathscr{V}}_{\text{BC}}$ and $\hat{\mathscr{Q}}_{\text{RBC},2}$ are preliminary feasible estimators of $\mathscr{V}_{\text{BC}}$ and $\mathscr{Q}_{\text{RBC},2}$. Therefore, we select the constant H_{TO} in $h_{\text{TO}} = H_{\text{TO}}n^{-\gamma_{\text{TO}}}$ as

$$\hat{H}_{\text{TO}} = \arg\min_{H>0} W \times H^{5+2p}\hat{\mathscr{Q}}_{\text{RBC},2} + (1-W) \times H^{-1-2\nu}4z^2_{\frac{\alpha}{2}}\hat{\mathscr{V}}_{\text{BC}}$$

$$= \left(\frac{1-W}{W}\frac{1+2\nu}{5+2p}\frac{4z^2_{\frac{\alpha}{2}}\hat{\mathscr{V}}_{\text{BC}}}{\hat{\mathscr{Q}}_{\text{RBC},2}} \right)^{\frac{1}{6+2p+2\nu}},$$

for a researcher-chosen weight $W \in (0,1)$. In Section 6 we find that $\hat{h}_{\text{TO}} = \hat{H}_{\text{TO}}n^{-\gamma_{\text{TO}}}$ (and its infeasible counterpart h_{TO}) behaves as expected, with the natural choice of $W = 1/2$ and $\gamma_{\text{TO}} = 0.1964$, the midpoint of (4.3) for $p = 1$.

5. EXTENSIONS

We briefly discuss several extensions of our main results. Unlike results based on first-order asymptotic approximations, establishing valid higher-order Edgeworth expansions in the settings of this section would require non-trivial additional work beyond the scope of this paper. Nevertheless, following the logic and results above, we can provide simple ROT bandwidth choices targeting inference, based on already-available MSE-optimal bandwidth selectors.

5.1. Other RD designs

In the context of fuzzy (and fuzzy kink) RD designs, the estimand and estimator are ratios of sharp RD design estimands and estimators, respectively. First-order asymptotic approximations follow directly from standard linearization methods, and although the validity of the coverage error expansion can be similarly proven to hold, this is no help in computing the terms of the expansion. That is, even though the linearization error has no effect on the first-order asymptotic approximation, it can have a direct effect on the Edgeworth and coverage error expansions. Without capturing the effect of the linearization, full derivation of inference optimal bandwidths is not possible. However, in this context we propose the following ROT bandwidth:

$$\check{h}^{\texttt{rot}}_{\text{RBC}} = n^{-p/((2p+3)(p+3))}\,\check{h}_{\text{MSE}},$$

where $\check{h}_{\text{MSE}}$ denotes an implementation of the MSE-optimal bandwidth for the fuzzy (or fuzzy kink) RD estimator. Sharp, fuzzy, and kink RD designs also arise in geographic, multi-score, and multi-cutoff RD settings (Papay et al., 2011; Keele and Titiunik, 2015; Cattaneo et al., 2016), and the results in this paper can also be used in those cases directly.

5.2. Clustered data

When the data exhibit clustering, first-order asymptotic results can be easily extended to account for clustered sampling where (a) each unit i belongs to exactly one of G clusters and (b) $G \to \infty$ and $Gh \to \infty$ (see Bartalotti and Brummet, 2017; Calonico et al., 2019). Since MSE-optimal

bandwidth choices in this context are available and fully implemented, the corresponding ROT implementation is

$$\check{h}_{\mathtt{RBC}}^{\mathtt{rot}} = G^{-p/((2p+3)(p+3))}\, \check{h}_{\mathtt{MSE}},$$

where now G denotes the number of clusters, and $\check{h}_{\mathtt{MSE}}$ denotes an implementation of the MSE-optimal bandwidth accounting for clustering. RBC confidence intervals are formed using this bandwidth choice, together with appropriate (cluster-robust) standard error estimators.

5.3. Pre-intervention covariates

Calonico et al. (2019) employ first-order asymptotics to characterize formally the implications of including pre-intervention covariates in the estimation of and inference for RD treatment effects. Again, this is not sufficient for higher-order expansions, and the inclusion of covariates will impact the coverage error expansion, making it impossible to derive a fully optimal bandwidth from existing results. However, a ROT bandwidth selector in this context is

$$\tilde{h}_{\mathtt{RBC}}^{\mathtt{rot}} = n^{-p/((2p+3)(p+3))}\, \tilde{h}_{\mathtt{MSE}},$$

where n denotes the sample size and $\tilde{h}_{\mathtt{MSE}}$ denotes an implementation of the MSE-optimal bandwidth accounting for the inclusion of additional pre-intervention covariates. RBC confidence intervals are formed using this bandwidth choice, together with appropriate covariate-adjusted standard error estimators.

6. NUMERICAL RESULTS

We present empirical evidence highlighting the performance of the new RD bandwidth selection and inference methods developed. We consider a Monte Carlo experiment and an empirical application, both employing the dataset of Ludwig and Miller (2007) used to study the effect of Head Start assistance on child mortality. This canonical dataset was employed before by Calonico et al. (2014), Cattaneo et al. (2017) and Calonico et al. (2019), where further institutional and descriptive information is provided.

6.1. Monte Carlo experiment

The simulations use $n = 500$ i.i.d. draws, $i = 1, 2, ..., n$, from the model

$$Y_i = m(X_i) + \varepsilon_i, \qquad X_i \sim 2\mathcal{B}(2, 4) - 1, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2),$$

where $\mathcal{B}(\alpha, \beta)$ denotes a beta distribution with parameters α and β, and the regression function $m(x)$ is obtained from the Head Start data. Specifically, we estimate the regression function using a fifth-order polynomial with separate coefficients for $X_i < c$ and $X_i > c$, where X_i is a poverty index and $c = 59.1984$ is the RD cutoff point. This estimation leads to

$$m(x) = \begin{cases} 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < c \\ 0.26 + 18.49x - 54.81x^2 + 74.30x^3 - 45.02x^4 + 9.83x^5 & \text{if } x \geq c \end{cases},$$

with $\sigma_\varepsilon = 0.6136$.

We consider 5,000 replications, and report empirical coverage and average interval length for a variety of inference procedures. Specifically, Table 2 considers undersmoothing ($I_{\mathtt{US}}(h)$) and

Table 2. Empirical coverage and average interval length of 95% confidence intervals.

	Bandwidth		Empirical coverage					Interval length				
	h	$\widehat{\rho}$	US	RBC:	$\widehat{\rho}$	ρ^*	$\rho=1$	US	RBC:	$\widehat{\rho}$	ρ^*	$\rho=1$
h_{MSE}	0.154	0.520	92.7		94.5	93.6	92.8	1.14		1.28	1.50	1.64
$\widehat{h}_{MSE}$	0.174	0.571	88.7		93.7	93.4	93.0	1.08		1.24	1.42	1.55
$\widetilde{h}_{MSE}$	0.173	0.571	88.9		93.7	93.6	93.0	1.08		1.24	1.43	1.56
h_{RBC}	0.145	0.492	93.1		94.6	93.5	92.6	1.17		1.31	1.55	1.69
$\widehat{h}_{RBC}$	0.163	0.535	88.7		93.6	91.0	90.5	1.15		1.32	1.50	1.64
h^{rot}_{RBC}	0.113	0.381	94.0		94.7	93.0	91.9	1.35		1.46	1.79	1.94
$\widehat{h}^{rot}_{RBC}$	0.127	0.418	92.4		94.3	93.4	92.2	1.28		1.39	1.69	1.83
$\widetilde{h}^{rot}_{RBC}$	0.127	0.416	92.5		94.3	93.1	92.1	1.28		1.39	1.69	1.84
h_{TO}	0.203	0.686	88.8		94.2	93.9	93.4	0.98		1.19	1.30	1.42
$\widehat{h}_{TO}$	0.172	0.566	87.3		93.6	90.9	90.7	1.11		1.31	1.46	1.59

Note: US denotes undersmoothed confidence interval, $I_{\mathtt{US}}(h)$, and RBC denotes robust bias-corrected confidence interval, $I_{\mathtt{RBC}}(h)$. Procedures are computed using the triangular kernel, $p=1$, and HC_3 variance estimation. Recall that $\hat{\rho}=\hat{h}/\hat{b}_{\mathtt{MSE}}$ for corresponding bandwidth selectors $\hat{h}$ (given in the table), and $\rho^*=0.850$ (Table 1).

RBC ($I_{\mathtt{RBC}}(h)$) confidence intervals for different choices of bandwidths h and parameter ρ. In all cases, we consider a local-linear RD estimator ($p=1$) with the triangular kernel and 'HC3' heteroskedasticity consistent standard errors, motivated by the fact that the least-squares residuals are on average too small (see the SA for more).

The results in Table 2 are organized as follows. The table presents three groups by row: (a) procedures employing MSE-optimal bandwidth choices ($h_{\mathtt{MSE}}$, $\hat{h}_{\mathtt{MSE}}$, $\tilde{h}_{\mathtt{MSE}}$); (b) procedures employing CE-optimal bandwidth choices ($h_{\mathtt{RBC}}$, $\hat{h}_{\mathtt{RBC}}$, $h^{\mathtt{rot}}_{\mathtt{RBC}}$, $\hat{h}^{\mathtt{rot}}_{\mathtt{RBC}}$, $\tilde{h}^{\mathtt{rot}}_{\mathtt{RBC}}$); and (c) procedures employing trade-off bandwidth choices ($h_{\mathtt{TO}}$, $\hat{h}_{\mathtt{TO}}$). Quantities without hats or tildes correspond to infeasible bandwidth choices, quantities with hats denote feasible implementations (DPI without label, or ROT with corresponding label), and quantities with tildes denote feasible implementations with covariate adjustment. For the latter the model includes, as a predetermined covariate, percentage of urban population in 1960.

The table also presents three groups by columns: (a) 'Bandwidth' reports infeasible or average feasible bandwidth choices (recall $\hat{\rho}=\hat{h}/\hat{b}_{\mathtt{MSE}}$ with $\hat{h}$ as appropriate); (b) 'Empirical coverage' reports coverage of $I_{\mathtt{US}}(h)$ and of $I_{\mathtt{RBC}}(h)$ for three choices of $\rho=h/b$; and (c) 'Interval length' reports the average length of the same four distinct confidence intervals (undersmoothing and three implementations of RBC indexed by the choice of ρ). Further implementation details are given in the SA.

All the findings emerging from the simulation study are in qualitative agreement with the main theoretical results from our paper. Confidence intervals based on undersmoothing, $I_{\mathtt{US}}(h)$, did not exhibit good coverage properties, while those based on RBC, $I_{\mathtt{RBC}}(h)$, performed well. While the MSE-optimal bandwidth selectors also worked well, the CE-optimal bandwidth selectors offered some empirical refinements in terms of coverage error. Furthermore, the bandwidth selector based on coverage error and interval length trade-off discussed in Section 4.3 also performed well. Other empirical findings are in line with our theoretical and methodological discussions.

Table 3. Head Start empirical application.

	Point estimate	Bandwidth		RBC confidence intervals		
		$\widehat{h}$	$\widehat{\rho}$	$\widehat{\rho}$	ρ^*	$\rho = 1$
$\widehat{h}_{MSE}$	−2.409	6.81	0.635	[−5.46 , −0.1]	[−5.88 , −0.44]	[−6.41 , −1.09]
$\widetilde{h}_{MSE}$	−2.473	6.98	0.651	[−5.21 , −0.37]	[−5.78 , −0.69]	[−6.54 , −1.39]
$\widehat{h}_{RBC}$	−3.311	4.467	0.416	[−6.14 , −0.82]	[−6.53 , −1.09]	[−6.23 , −0.27]
$\widehat{h}^{rot}_{RBC}$	−3.273	4.581	0.427	[−6.12 , −0.78]	[−6.57 , −1.16]	[−6.26 , −0.39]
$\widetilde{h}^{rot}_{RBC}$	−3.526	4.696	0.438	[−6.13 , −1.25]	[−7.06 , −1.75]	[−6.93 , −1.23]

Note: Procedures are computed using the triangular kernel, $p = 1$, and HC_3 variance estimation. Recall that $\hat{\rho} = \hat{h}/\hat{b}_{\text{MSE}}$ for bandwidth selectors $\hat{h}$ (given in the table), and $\rho^* = 0.850$ (Table 1).

6.2. Empirical application

To complement the Monte Carlo experiment, we also employed the Head Start data to illustrate the performance of our new bandwidth selection and inference methods using a realistic empirical application. Specifically, we study the RD treatment effect of Head Start assistance on child mortality following the original work of Ludwig and Miller (2007). See also Cattaneo et al. (2017) for a recent re-examination of the empirical findings using modern RD methodology.

In this application, the unit of analysis is a U.S. county, and eligibility into Head Start assistance was based on each county's poverty index in 1960. The RD design naturally emerges by the assignment rule to the program: $T_i = \mathbb{1}(X_i \geq c)$, where X_i denotes the 1960 poverty index of county i and $c = 59.1984$ was the federally mandated cutoff point. The outcome variable considered is mortality rates per 100,000 for children between 5 and 9 years old, with Head Start-related causes, during the period 1973–1983.

The main empirical results are presented in Table 3. We first report the sharp RD treatment effect estimator using a local-linear estimator ($p = 1$) with triangular kernel and MSE-optimal bandwidth. In line with previous findings, we obtain $\hat{\tau}_0(\hat{h}_{\text{MSE}}) = -2.409$ with $\hat{h}_{\text{MSE}} = 6.81$. Next, we compute several RBC confidence intervals with different choices of bandwidths h and ρ, including the new inference procedures proposed in this paper. In all cases, the empirical results are in qualitative agreement and confirm an RD treatment effect that is statistically different from zero.

7. CONCLUSION

This paper presented two main results for RD designs, which have concrete, practical implications for empirical work. First, we established valid coverage error expansions of naïve and RBC confidence intervals for RD treatment effects, and showed that the latter confidence intervals never have asymptotically larger coverage errors and can indeed offer higher-order refinements whenever the underlying regression functions are smooth enough (arguably the most relevant case in applications). Thus, this result offers concrete guidance for empirical work in RD designs by ranking competing confidence interval estimators encountered in practice.

Second, using our coverage error expansions, we also developed CE-optimal bandwidth choices and discussed how to implement them in practice. In the same way that MSE-optimal bandwidths

deliver MSE-optimal point estimators for RD treatment effects, our new CE-optimal bandwidth choices deliver inference-optimal confidence intervals in the sense that their coverage error is the smallest possible given the choice of point estimator used. This second result also offers concrete empirical guidance for applied work using RD designs, providing a companion bandwidth choice to be used when forming confidence intervals for RD treatment effects.

ACKNOWLEDGEMENTS

A preliminary version of this paper circulated under the title 'Coverage error optimal confidence intervals for regression discontinuity designs' (first draft: June 26, 2016). We thank Josh Angrist, Zhuan Pei, Rocio Titiunik, and Gonzalo Vazquez-Bare for comments. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES 1357561 and SES 1459931). Farrell gratefully acknowledges financial support from the Richard N. Rosett and John E. Jeuck Fellowships.

REFERENCES

Abadie, A. and M. D. Cattaneo (2018). Econometric methods for program evaluation. *Annual Review of Economics 10*, 465–503.

Arai, Y. and H. Ichimura (2016). Optimal bandwidth selection for the fuzzy regression discontinuity estimator. *Economic Letters 141*, 103–6.

Arai, Y. and H. Ichimura (2018). Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quantitative Economics 9*, 441–82.

Bartalotti, O. and Q. Brummet (2017). Regression discontinuity designs with clustered data. In Cattaneo, M. D. and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Vol. 38)*, 383–420. Emerald Group Publishing, Howard House, Wagon Lane, Bingley BD16 1WA, UK.

Calonico, S., M. D. Cattaneo and M. H. Farrell (2018). 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 (2019). Coverage error optimal confidence intervals for local polynomial regression. (arXiv:1808.01398).

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 101*, 442–51.

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

Calonico, S., M. D. Cattaneo and R. Titiunik (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association 110*, 1753–69.

Card, D., D. S. Lee, Z. Pei and A. Weber (2015). Inference on causal effects in a generalized regression kink design. *Econometrica 83*, 2453–83.

Card, D., D. S. Lee, Z. Pei and A. Weber (2017). Regression kink design: Theory and practice. In Cattaneo, M. D. and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Vol. 38)*, 341–82. Emerald Group Publishing, Howard House, Wagon Lane, Bingley BD16 1WA, UK.

Cattaneo, M. D. and J. C. Escanciano (2017). *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Vol. 38)*. Emerald Group Publishing.

Cattaneo, M. D., B. Frandsen and R. Titiunik (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate. *Journal of Causal Inference 3*, 1–24.

Cattaneo, M. D., N. Idrobo and R. Titiunik (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Cambridge Elements: Quantitative and Computational Methods for Social Science*. Cambridge University Press, Cambridge.

Cattaneo, M. D., N. Idrobo and R. Titiunik (2020). A Practical Introduction to Regression Discontinuity Designs: Extensions. *Cambridge Elements: Quantitative and Computational Methods for Social Science*. Cambridge University Press.

Cattaneo, M. D., L. Keele, R. Titiunik and G. Vazquez-Bare (2016). Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics 78*, 1229–48.

Cattaneo, M. D., R. Titiunik and G. Vazquez-Bare (2017). Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality. *Journal of Policy Analysis and Management 36*, 643–81.

Cattaneo, M. D. and G. Vazquez-Bare (2016). The choice of neighborhood in regression discontinuity designs. *Observational Studies 2*, 134–46.

Cerulli, G., Y. Dong, A. Lewbel and A. Poulsen (2017). Testing stability of regression discontinuity models. In Cattaneo, M. D. and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Vol. 38)*, 317–39. Emerald Group Publishing, Howard House, Wagon Lane, Bingley BD16 1WA, UK.

Cheng, M.-Y., J. Fan and J. S. Marron (1997). On automatic boundary corrections. *Annals of Statistics 25*, 1691–708.

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.

Dong, Y. (2019). Regression discontinuity designs with sample selection. *Journal of Business and Economic Statistics 37*, 171–86.

Dong, Y., Y.-Y. Lee and M. Gou (2019). Regression discontinuity designs with a continuous treatment, SSRN working paper No. 3167541, University of California, Irvine.

Dong, Y. and A. Lewbel (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics 97*, 1081–92.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC.

Ganong, P. and S. Jäger (2018). A permutation test for the regression kink design. *Journal of the American Statistical Association 113*, 494–504.

Gelman, A. and G. W. Imbens (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics 37*, 447–56.

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.

He, Y. and O. Bartalotti (2019). Wild bootstrap for fuzzy regression discontinuity designs: Obtaining robust bias-corrected confidence intervals. *Working paper, Iowa State University*.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen and J. Tukiainen (2018). When does regression discontinuity design work? Evidence from random election outcomes. *Quantitative Economics 9*, 1019–51.

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

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

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, Cambridge, 32 Avenue of the Americas, New York, NY 10013, USA.

Keele, L. J. and R. Titiunik (2015). Geographic boundaries as regression discontinuities. *Political Analysis 23*, 127–55.

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.

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.

Papay, J. P., J. B. Willett and R. J. Murnane (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics 161*, 203–7.

Sekhon, J. and R. Titiunik (2017). On interpreting the regression discontinuity design as a local experiment. In Cattaneo, M. D. and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Vol. 38)*, 1–28. Emerald Group Publishing, Howard House, Wagon Lane, Bingley BD16 1WA, UK.

Xu, K.-L. (2017). Regression discontinuity with categorical outcomes. *Journal of Econometrics 201*, 1–18.

SUPPORTING INFORMATION

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

Online Supplement

Co-editor Victor Chernozhukov handled this manuscript.