Econometrica, Vol. 82, No. 6 (November, 2014), 2295–2326

NOTES AND COMMENTS

ROBUST NONPARAMETRIC CONFIDENCE INTERVALS FOR REGRESSION-DISCONTINUITY DESIGNS

BY SEBASTIAN CALONICO, MATIAS D. CATTANEO, AND ROCIO TITIUNIK[1]

In the regression-discontinuity (RD) design, units are assigned to treatment based on whether their value of an observed covariate exceeds a known cutoff. In this design, local polynomial estimators are now routinely employed to construct confidence intervals for treatment effects. The performance of these confidence intervals in applications, however, may be seriously hampered by their sensitivity to the specific bandwidth employed. Available bandwidth selectors typically yield a "large" bandwidth, leading to data-driven confidence intervals that may be biased, with empirical coverage well below their nominal target. We propose new theory-based, more robust confidence interval estimators for average treatment effects at the cutoff in sharp RD, sharp kink RD, fuzzy RD, and fuzzy kink RD designs. Our proposed confidence intervals are constructed using a bias-corrected RD estimator together with a novel standard error estimator. For practical implementation, we discuss mean squared error optimal bandwidths, which are by construction not valid for conventional confidence intervals but are valid with our robust approach, and consistent standard error estimators based on our new variance formulas. In a special case of practical interest, our procedure amounts to running a quadratic instead of a linear local regression. More generally, our results give a formal justification to simple inference procedures based on increasing the order of the local polynomial estimator employed. We find in a simulation study that our confidence intervals exhibit close-to-correct empirical coverage and good empirical interval length on average, remarkably improving upon the alternatives available in the literature. All results are readily available in R and STATA using our companion software packages described in Calonico, Cattaneo, and Titiunik (2014d, 2014b).

KEYWORDS: Regression discontinuity, local polynomials, bias correction, robust inference, alternative asymptotics.

1. INTRODUCTION

THE REGRESSION-DISCONTINUITY (RD) DESIGN has become one of the leading quasi-experimental empirical strategies in economics, political sci-

[1]This paper is a revised and extended version of our paper previously entitled "Robust Nonparametric Bias-Corrected Inference in the Regression Discontinuity Design" (first draft: November 2011). We are specially thankful to the co-editor for encouraging us to write this new version as well as for giving us very useful comments and suggestions on the previous drafts. In addition, for thoughtful comments and suggestions we thank the three anonymous referees, Alberto Abadie, Josh Angrist, Ivan Canay, Victor Chernozhukov, Max Farrell, Patrik Guggenberger, Hide Ichimura, Guido Imbens, Michael Jansson, Justin McCrary, Anna Mikusheva, Whitney Newey, Zhuan Pei, Andres Santos, Jeff Smith, Jeff Wooldridge, as well as seminar and conference participants at many institutions. Finally, we also thank Manuela Angelucci, Luis Alejos Marroquin, Habiba Djebbari, Paul Gertler, Dan Gilligan, Jens Ludwig, and Doug Miller for sharing and helping us with different data sets used in this research project. The authors gratefully acknowledge financial support from the National Science Foundation (SES 1357561).

 DOI: 10.3982/ECTA11757

ence, education, and many other social and behavioral sciences (see van der Klaauw (2008), Imbens and Lemieux (2008), Lee and Lemieux (2010), and DiNardo and Lee (2011) for reviews). In this design, units are assigned to treatment based on their value of an observed covariate (also known as score or running variable), with the probability of treatment assignment jumping discontinuously at a known cutoff. For example, in its original application, Thistlethwaite and Campbell (1960) used this design to study the effects of receiving an award on future academic achievement, where the award was given to students whose test scores were above a cutoff. The idea of the RD design is to study the effects of the treatment using only observations near the cutoff to control for smoothly varying unobserved confounders. In the simplest case, flexible estimation of RD treatment effects approximates the regression function of the outcome given the score near the cutoff for control and treated groups separately, and computes the estimated effect as the difference of the values of the regression functions at the cutoff for each group.

Nonparametric local polynomial estimators have received great attention in the recent RD literature, and have become the standard choice for estimation of RD treatment effects. This estimation strategy involves approximating the regression functions above and below the cutoff by means of weighted polynomial regressions, typically of order 1 or 2, with weights computed by applying a kernel function on the distance of each observation's score to the cutoff. These kernel-based estimators require a choice of bandwidth for implementation, and several bandwidth selectors are now available in the literature. These bandwidth selectors are obtained by balancing squared-bias and variance of the RD estimator, a procedure that typically leads to bandwidth choices that are too "large" to ensure the validity of the distributional approximations usually invoked; that is, these bandwidth selectors lead to a non-negligible bias in the distributional approximation of the estimator.[2] As a consequence, the resulting confidence intervals for RD treatment effects may be biased, having empirical coverage well below their nominal target. This implies that conventional confidence intervals may substantially over-reject the null hypothesis of no treatment effect.

To address this drawback in conventional RD inference, we propose new confidence intervals for RD treatment effects that offer robustness to "large" bandwidths such as those usually obtained from cross-validation or asymptotic mean squared error minimization. Our proposed confidence intervals are constructed as follows. We first bias-correct the RD estimator to account for the effect of a "large" bandwidth choice; that is, we recenter the usual t-statistic with an estimate of the leading bias. As it is well known, however, conventional bias correction alone delivers very poor finite-sample performance because it relies on a low-quality distributional approximation. Thus, in order to improve

[2]For example, for the local-linear RD estimator, "small" and "large" bandwidths refer, respectively, to $nh_n^5 \to 0$ and $nh_n^5 \nrightarrow 0$ (e.g., $nh_n^5 \to c \in \mathbb{R}_{++}$), where h_n is the bandwidth and n is the sample size. Section 2 discusses this case in detail, while the general case is given in the Appendix.

the quality of the distributional approximation of the bias-corrected t-statistic, we rescale it with a novel standard error formula that accounts for the additional variability introduced by the estimated bias. The new standardization is theoretically justified by a nonstandard large-sample distributional approximation of the bias-corrected estimator, which explicitly accounts for the potential contribution that bias correction may add to the finite-sample variability of the usual t-statistic. Altogether, our proposed confidence intervals are demonstrably more robust to the bandwidth choice ("small" or "large"), as they are not only valid when the usual bandwidth conditions are satisfied (being asymptotically equivalent to the conventional confidence intervals in this case), but also continue to offer correct coverage rates in large samples even when the conventional confidence intervals do not (see Remarks 2 and 3 below). These properties are illustrated with an empirically motivated simulation study, which shows that our proposed data-driven confidence intervals exhibit close-to-correct empirical coverage and good empirical interval length on average.

Our discussion focuses on the construction of robust confidence intervals for the RD average treatment effect at the cutoff in four settings: sharp RD, sharp kink RD, fuzzy RD, and fuzzy kink RD designs. These are special cases of our main theorems given in the Appendix. In all cases, the bias-correction technique follows the standard approach in the nonparametrics literature (e.g., Fan and Gijbels (1996, Section 4.4, p. 116)), but our standard error formulas are different because they incorporate additional terms not present in the conventional formulas currently used in practice. The resulting confidence intervals allow for mean squared optimal bandwidth selectors and, more generally, enjoy demonstrable improvements in terms of allowed bandwidth sequences, coverage error rates, and, in some cases, interval length (see Remarks 2, 4, and 5 below). As a particular case, our results also justify confidence intervals estimators based on a local polynomial estimator of an order higher than the order of the polynomial used for point estimation, a procedure that is easy to implement in applications (see Remark 7 below). The new confidence intervals may be used both for inference on treatment effects (when the outcome of interest is used as an outcome in the estimation) as well as for falsification tests that look for null effects (when pretreatment or "placebo" covariates are used as outcomes in the estimation).

This paper contributes to the emerging methodological literature on RD designs. See Hahn, Todd, and van der Klaauw (2001) and Lee (2008) for identification results, Porter (2003) for optimality results of local polynomial estimators, McCrary (2008) for specification testing, Lee and Card (2008) for inference with discrete running variables, Imbens and Kalyanaraman (2012) for bandwidth selection procedures for local-linear estimators, Frandsen, Frölich, and Melly (2012) for quantile treatment effects, Otsu, Xu, and Matsushita (2014) for empirical likelihood methods, Card, Lee, Pei, and Weber (2014) and Dong (2014) for kink RD designs, Marmer, Feir, and Lemieux (2014) for weak-IV robust inference in fuzzy RD designs, Cattaneo, Frandsen, and Titiunik (2014) for randomization inference methods, Calonico, Cattaneo, and Titiunik

(2014a) for optimal RD plots, and Keele and Titiunik (2014) for geographic RD. More broadly, our results also contribute to the literature on asymptotic approximations for nonparametric local polynomial estimators (Fan and Gijbels (1996)), which are useful in econometrics (Ichimura and Todd (2007))—see Remark 8 and Calonico, Cattaneo, and Farrell (2014) for further discussion.

The rest of the paper is organized as follows. Section 2 describes the sharp RD design, reviews conventional results, and outlines our proposed robust confidence intervals. Section 3 discusses extensions to kink RD, fuzzy RD, and fuzzy kink RD designs. Mean squared error optimal bandwidths and their validity are examined in Section 4, while valid standard error estimators are discussed in Section 5. Section 6 presents our simulation study, and Section 7 concludes. In the Appendix, we summarize our general theoretical results, including extensions to arbitrary polynomial orders and higher-order derivatives, while in the Supplemental Material (Calonico, Cattaneo, and Titiunik (2014c)) we collect the main mathematical proofs, other methodological and technical results such as consistent bandwidth selection, additional simulation evidence, and an empirical illustration employing household data from Progresa/Oportunidades. Companion `R` and `STATA` software packages are described in Calonico, Cattaneo, and Titiunik (2014d, 2014b).

2. SHARP RD DESIGN

In the canonical sharp RD design, $(Y_i(0), Y_i(1), X_i)'$, $i = 1, 2, \dots, n$, is a random sample and X_i has density $f(x)$ with respect to the Lebesgue measure. Given a known threshold $\bar{x}$, set to $\bar{x} = 0$ without loss of generality, the observed score or forcing variable X_i determines whether unit i is assigned treatment ($X_i \geq 0$) or not ($X_i < 0$), while the random variables $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes with and without treatment, respectively. The observed random sample is $(Y_i, X_i)'$, $i = 1, 2, \dots, n$, where $Y_i = Y_i(0) \cdot (1 - T_i) + Y_i(1) \cdot T_i$ with $T_i = \mathbf{1}(X_i \geq 0)$ and $\mathbf{1}(\cdot)$ is the indicator function.

The parameter of interest is $\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = \bar{x}]$, the average treatment effect at the threshold. Under a mild continuity condition, Hahn, Todd, and van der Klaauw (2001) showed that this parameter is nonparametrically identifiable as the difference of two conditional expectations evaluated at the (induced) boundary point $\bar{x} = 0$:

$$\tau_{\text{SRD}} = \mu_+ - \mu_-, \quad \mu_+ = \lim_{x \to 0^+} \mu(x), \quad \mu_- = \lim_{x \to 0^-} \mu(x),$$
$$\mu(x) = \mathbb{E}[Y_i | X_i = x].$$

Throughout the paper, we drop the evaluation point of functions whenever possible to simplify notation. Estimation in RD designs naturally focuses on flexible approximation, near the cutoff $\bar{x} = 0$, of the regression functions $\mu_-(x) = \mathbb{E}[Y_i(0)|X_i = x]$ (from the left) and $\mu_+(x) = \mathbb{E}[Y_i(1)|X_i = x]$ (from the right). We employ the following assumption on the basic sharp RD model.

ASSUMPTION 1: *For some* $\kappa_0 > 0$, *the following hold in the neighborhood* $(-\kappa_0, \kappa_0)$ *around the cutoff* $\bar{x} = 0$:

(a) $\mathbb{E}[Y_i^4|X_i = x]$ *is bounded*, *and* $f(x)$ *is continuous and bounded away from zero*.

(b) $\mu_-(x) = \mathbb{E}[Y_i(0)|X_i = x]$ *and* $\mu_+(x) = \mathbb{E}[Y_i(1)|X_i = x]$ *are* S *times continuously differentiable*.

(c) $\sigma_-^2(x) = \mathbb{V}[Y_i(0)|X_i = x]$ *and* $\sigma_+^2(x) = \mathbb{V}[Y_i(1)|X_i = x]$ *are continuous and bounded away from zero*.

Part (a) in Assumption 1 imposes existence of moments, requires that the running variable X_i be continuously distributed near the cutoff, and ensures the presence of observations arbitrarily close to the cutoff in large samples. Part (b) imposes standard smoothness conditions on the underlying regression functions, which is the key ingredient used to control the leading biases of the RD estimators considered in this paper. Part (c) puts standard restrictions on the conditional variance of the observed outcome, which may be different at either side of the threshold. We set $\sigma_+^2 = \lim_{x\to 0^+} \sigma^2(x)$ and $\sigma_-^2 = \lim_{x\to 0^-} \sigma^2(x)$, where $\sigma^2(x) = \mathbb{V}[Y_i|X_i = x]$. Higher-order derivatives of the unknown regression functions are denoted by $\mu_+^{(\nu)}(x) = \mathrm{d}^\nu \mu_+(x)/\mathrm{d}x^\nu$ and $\mu_-^{(\nu)}(x) = \mathrm{d}^\nu \mu_-(x)/\mathrm{d}x^\nu$, for $\nu < S$ (with S in Assumption 1(b)). We also set $\mu_+^{(\nu)} = \lim_{x\to 0^+} \mu_+^{(\nu)}(x)$ and $\mu_-^{(\nu)} = \lim_{x\to 0^-} \mu_-^{(\nu)}(x)$; by definition, $\mu_+ = \mu_+^{(0)}$ and $\mu_- = \mu_-^{(0)}$.

REMARK 1—Discrete Running Variable: Assumption 1(a) rules out discrete-valued running variables. In applications where X_i exhibits many mass points near the cutoff, this assumption may still give a good approximation and our results might be used in practice. However, when X_i exhibits few mass points, our results do not apply directly without further assumptions and modifications, and other assumptions and inference approaches may be more appropriate; see, for example, Cattaneo, Frandsen, and Titiunik (2014).

Throughout the paper, we employ local polynomial regression estimators of various orders to approximate unknown regression functions (Fan and Gijbels (1996)). These estimators are particularly well-suited for inference in the RD design because of their excellent boundary properties (Cheng, Fan, and Marron (1997)). Section A.1 in the Appendix describes these estimators in full generality and introduces detailed notation not employed in the main text to ease the exposition. We impose the following assumption on the kernel function employed to construct these estimators.

ASSUMPTION 2: *For some* $\kappa > 0$, *the kernel function* $k(\cdot): [0, \kappa] \mapsto \mathbb{R}$ *is bounded and nonnegative*, *zero outside its support*, *and positive and continuous on* $(0, \kappa)$.

Assumption 2 permits all kernels commonly used in empirical work, including the triangular kernel $k(u) = (1 - u)\mathbf{1}(0 \le u \le 1)$ and the uniform kernel $k(u) = \mathbf{1}(0 \le u \le 1)$. Our results apply when different kernels are used on either side of the threshold, but we set $K(u) = k(-u) \cdot \mathbf{1}(u < 0) + k(u) \cdot \mathbf{1}(u \ge 0)$ for concreteness. This implies that, for $\kappa > 0$ in Assumption 2, $K(\cdot)$ is symmetric, bounded and nonnegative on $[-\kappa, \kappa]$, zero otherwise, and positive and continuous on $(-\kappa, \kappa)$. For simplicity, we employ the same kernel function $k(\cdot)$ to form all estimators in the paper.

2.1. *Robust Local-Linear Confidence Intervals*

Following Hahn, Todd, and van der Klaauw (2001) and Porter (2003), we consider confidence intervals based on the popular local-linear estimator of τ_{SRD}, which is the difference in intercepts of two first-order local polynomial estimators, one from each side of the threshold. Formally, for a positive bandwidth h_n,

$$\begin{aligned}
&\hat{\tau}_{\text{SRD}}(h_n) = \hat{\mu}_{+,1}(h_n) - \hat{\mu}_{-,1}(h_n),\\
&\big(\hat{\mu}_{+,1}(h_n), \hat{\mu}^{(1)}_{+,1}(h_n)\big)'\\
&\quad = \arg\min_{b_0, b_1 \in \mathbb{R}} \sum_{i=1}^{n} \mathbf{1}(X_i \ge 0)(Y_i - b_0 - X_i b_1)^2 K(X_i/h_n),\\
&\big(\hat{\mu}_{-,1}(h_n), \hat{\mu}^{(1)}_{-,1}(h_n)\big)'\\
&\quad = \arg\min_{b_0, b_1 \in \mathbb{R}} \sum_{i=1}^{n} \mathbf{1}(X_i < 0)(Y_i - b_0 - X_i b_1)^2 K(X_i/h_n).
\end{aligned}$$

Conventional approaches to constructing confidence intervals for τ_{SRD} using the local-linear estimator rely on the following large-sample approximation for the standardized t-statistic (see Lemma A.1(D) in the Appendix for the general result): if $nh_n^5 \to 0$ and $nh_n \to \infty$, then

$$\begin{aligned}
&T_{\text{SRD}}(h_n) = \frac{\hat{\tau}_{\text{SRD}}(h_n) - \tau_{\text{SRD}}}{\sqrt{\mathsf{V}_{\text{SRD}}(h_n)}} \to_d \mathcal{N}(0, 1),\\
&\mathsf{V}_{\text{SRD}}(h_n) = \mathbb{V}\big[\hat{\tau}_{\text{SRD}}(h_n) | \mathcal{X}_n\big], \quad \mathcal{X}_n = [X_1, \dots, X_n]'.
\end{aligned}$$

This justifies the conventional (infeasible) $100(1 - \alpha)$-percent confidence interval for τ_{SRD} given by

$$I_{\text{SRD}}(h_n) = \Big[\hat{\tau}_{\text{SRD}}(h_n) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathsf{V}_{\text{SRD}}(h_n)}\Big],$$

with Φ^{-1}_{α} the appropriate α-quantile of the standard normal distribution. In practice, a standard error estimator is needed to construct feasible confidence intervals because the variance $\mathsf{V}_{\text{SRD}}(h_n)$ involves unknown quantities, but for now we assume $\mathsf{V}_{\text{SRD}}(h_n)$ is known and postpone the issue of standard error estimation until Section 5. Even in this simplified known-variance case, the choice of the bandwidth h_n is crucial. The condition $nh_n^5 \to 0$ is explicitly imposed to eliminate the contribution of the leading bias to the distributional approximation, which depends on the unknown second derivatives $\mu_+^{(2)}$ and $\mu_-^{(2)}$, as described in Lemma A.1(B) in the Appendix. This means that, in general, the confidence intervals $I_{\text{SRD}}(h_n)$ will have correct asymptotic coverage only if the bandwidth h_n is "small" enough to satisfy the bias condition $nh_n^5 \to 0$.

Several approaches are available in the literature to select h_n, including plug-in rules and cross-validation procedures; see Imbens and Kalyanaraman (2012) for a recent account of the state of the art in bandwidth selection for RD designs. Unfortunately, these approaches lead to bandwidths that are too "large" because they do not satisfy the bias condition $nh_n^5 \to 0$: minimizing the asymptotic mean squared error (MSE) of $\hat{\tau}_{\text{SRD}}(h_n)$ gives the optimal plug-in bandwidth choice $h_{\text{MSE}} = C_{\text{MSE}} n^{-1/5}$ with C_{MSE} a constant, which by construction implies that $n(h_{\text{MSE}})^5 \to c \in (0, \infty)$ and hence leads to a first-order bias in the distributional approximation. This is a well-known problem in the nonparametric curve estimation literature. Moreover, implementing this MSE-optimal bandwidth choice in practice is likely to introduce additional variability in the chosen bandwidth that may lead to "large" bandwidths as well. Similarly, cross-validation bandwidth selectors tend to have low convergence rates, and thus also typically lead to "large" bandwidth choices; see, for example, Ichimura and Todd (2007) and references therein. These observations suggest that commonly used local-linear RD confidence intervals may not exhibit correct coverage in applications due to the presence of a potentially first-order bias in their construction, a phenomenon we illustrate with simulation evidence in Section 6. Since applied researchers often estimate RD treatment effects using local-linear regressions with MSE-optimal bandwidths and implicitly ignore the asymptotic bias of the estimator, the poor coverage of conventional confidence intervals we highlight potentially affects many RD empirical applications.

We propose a novel approach to inference based on bias correction to address this problem. Conventional bias correction seeks to remove the leading bias term of the statistic by subtracting off a consistent bias estimate, thus removing the impact of the potentially first-order bias. While systematic and easy to justify theoretically, this approach usually delivers poor performance in finite samples. We propose an alternative large-sample distributional approximation that takes bias correction as a starting point, but improves its performance in finite samples by accounting for the added variability introduced by the bias estimate.

To describe our approach formally, consider first the conventional bias-correction approach. The leading asymptotic bias of the local-linear estimator is

$$\mathbb{E}\big[\hat{\tau}_{\text{SRD}}(h_n)|\mathcal{X}_n\big] - \tau_{\text{SRD}} = h_n^2 \mathsf{B}_{\text{SRD}}(h_n)\big\{1 + o_p(1)\big\},$$

$$\mathsf{B}_{\text{SRD}}(h_n) = \frac{\mu_+^{(2)}}{2!}\mathcal{B}_{+,\text{SRD}}(h_n) - \frac{\mu_-^{(2)}}{2!}\mathcal{B}_{-,\text{SRD}}(h_n),$$

where $\mathcal{B}_{+,\text{SRD}}(h_n)$ and $\mathcal{B}_{-,\text{SRD}}(h_n)$ are asymptotically bounded, observed quantities (function of $\mathcal{X}_n$, $k(\cdot)$, and h_n) explicitly given in Lemma A.1(B) in the Appendix. Therefore, a plug-in bias-corrected estimator is

$$\hat{\tau}^{\text{bc}}_{\text{SRD}}(h_n, b_n) = \hat{\tau}_{\text{SRD}}(h_n) - h_n^2 \hat{\mathsf{B}}_{\text{SRD}}(h_n, b_n),$$

$$\hat{\mathsf{B}}_{\text{SRD}}(h_n, b_n) = \frac{\hat{\mu}^{(2)}_{+,2}(b_n)}{2!}\mathcal{B}_{+,\text{SRD}}(h_n) - \frac{\hat{\mu}^{(2)}_{-,2}(b_n)}{2!}\mathcal{B}_{-,\text{SRD}}(h_n),$$

with $\hat{\mu}^{(2)}_{+,2}(b_n)$ and $\hat{\mu}^{(2)}_{-,2}(b_n)$ denoting conventional local-quadratic estimators of $\mu_+^{(2)}$ and $\mu_-^{(2)}$, as described in Section A.1 in the Appendix. Here, b_n is the so-called pilot bandwidth sequence, usually larger than h_n. As shown in the Appendix for the general case, if $nh_n^7 \to 0$ and $h_n/b_n \to 0$, and other regularity conditions hold, then the bias-corrected (infeasible) t-statistic satisfies

$$T^{\text{bc}}_{\text{SRD}}(h_n, b_n) = \frac{\hat{\tau}^{\text{bc}}_{\text{SRD}}(h_n, b_n) - \tau_{\text{SRD}}}{\sqrt{\mathsf{V}_{\text{SRD}}(h_n)}} \to_d \mathcal{N}(0, 1),$$

which justifies confidence intervals for τ_{SRD} of the form

$$I^{\text{bc}}_{\text{SRD}}(h_n, b_n) = \Big[\big(\hat{\tau}_{\text{SRD}}(h_n) - h_n^2 \hat{\mathsf{B}}_{\text{SRD}}(h_n, b_n)\big) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathsf{V}_{\text{SRD}}(h_n)}\Big].$$

That is, in the conventional bias-correction approach, the confidence intervals are recentered to account for the presence of the bias. This approach allows for potentially "larger" bandwidths h_n, such as the MSE-optimal choice, because the leading asymptotic bias is manually removed from the distributional approximation. In practice, b_n may also be selected using an MSE-optimal choice, denoted b_{MSE}, which can be implemented by a plug-in estimate, denoted $\hat{b}_{\text{MSE}}$; see Section 4 for details. While bias correction is an appealing theoretical idea, a natural concern with the conventional large-sample approximation for the bias-corrected local-linear RD estimator is that it does not account for the additional variability introduced by the bias estimates $\hat{\mu}^{(2)}_{+,2}(b_n)$ and $\hat{\mu}^{(2)}_{-,2}(b_n)$, and thus the distributional approximation given above tends to provide a poor characterization of the finite-sample variability of the statistic. This large-sample approximation relies on the carefully tailored condition $h_n/b_n \to 0$, which makes the variability of the bias-correction estimate disappear asymptotically. However, h_n/b_n is never zero in finite samples.

Our alternative asymptotic approximation for bias-corrected local polynomial estimators removes the restriction $h_n/b_n \to 0$, leading to alternative confidence intervals for RD treatment effects capturing the (possibly first-order) effect of the bias correction on the distributional approximation. The alternative large-sample approximation we propose for the (properly centered and scaled) estimator $\hat{\tau}^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ allows for the more general condition $\rho_n = h_n/b_n \to \rho \in [0, \infty]$, which in particular permits a pilot bandwidth b_n of the same order of (and possibly equal to) the main bandwidth h_n. This approach implies that the bias-correction term may not be asymptotically negligible (after appropriate centering and scaling) in general, in which case it will converge in distribution to a centered at zero normal random variable, provided the asymptotic bias is small. Thus, the resulting distributional approximation includes the contribution of both the point estimate $\hat{\tau}_{\text{SRD}}(h_n)$ and the bias estimate, leading to a different asymptotic variance in general. This idea is formalized in the following theorem.

THEOREM 1: *Let Assumptions 1–2 hold with* $S \geq 3$. *If* $n \min\{h_n^5, b_n^5\} \times \max\{h_n^2, b_n^2\} \to 0$ *and* $n \min\{h_n, b_n\} \to \infty$, *then*

$$T^{\text{rbc}}_{\text{SRD}}(h_n, b_n) = \frac{\hat{\tau}^{\text{bc}}_{\text{SRD}}(h_n, b_n) - \tau_{\text{SRD}}}{\sqrt{\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, b_n)}} \to_d \mathcal{N}(0, 1),$$

$$\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, b_n) = \mathsf{V}_{\text{SRD}}(h_n) + \mathsf{C}^{\text{bc}}_{\text{SRD}}(h_n, b_n),$$

provided $\kappa \max\{h_n, b_n\} < \kappa_0$. *The exact form of* $\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ *is given in Theorem* A.1(V) *in the Appendix.*

Theorem 1 shows that by standardizing the bias-corrected estimator by its (conditional) variance, the asymptotic distribution of the resulting bias-corrected statistic $T^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ is Gaussian even when the condition $h_n/b_n \to 0$ is violated. The standardization formula $\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ depends explicitly on the behavior of $\rho_n = h_n/b_n$, and $\mathsf{C}^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ may be interpreted as a correction to account for the variability of the estimated bias-correction term. The additional term $\mathsf{C}^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ depends on the (asymptotic) variability of the bias-correction estimate as well as on its correlation with the original RD estimator $\hat{\tau}_{\text{SRD}}(h_n)$. The key practical implication of Theorem 1 is that it justifies the more robust, theory-based $100(1 - \alpha)$-percent confidence intervals:

$$I^{\text{rbc}}_{\text{SRD}}(h_n, b_n) = \Big[\big(\hat{\tau}_{\text{SRD}}(h_n) - h_n^2 \hat{\mathsf{B}}_{\text{SRD}}(h_n, b_n)\big) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathsf{V}_{\text{SRD}}(h_n) + \mathsf{C}^{\text{bc}}_{\text{SRD}}(h_n, b_n)}\Big].$$

We summarize important features of our main result in the remarks below.

REMARK 2—Robustness: The distributional approximation in Theorem 1 permits one bandwidth (but not both) to be fixed, provided this bandwidth is not too "large"; that is, both must satisfy $\kappa \max\{h_n, b_n\} < \kappa_0$ for all n large enough, but only one needs to vanish. This theorem allows for all conventional bandwidth sequences and, in addition, permits other bandwidth sequences that would make $I_{\text{SRD}}(h_n)$ and $I^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ invalid (i.e., $\mathbb{P}[\tau_{\text{SRD}} \in I_{\text{SRD}}(h_n)] \nrightarrow 1-\alpha$ and $\mathbb{P}[\tau_{\text{SRD}} \in I^{\text{bc}}_{\text{SRD}}(h_n)] \nrightarrow 1-\alpha$).

REMARK 3—Asymptotic Variance: Three limiting cases are obtained depending on $\rho_n \to \rho \in [0, \infty]$.

Case 1: $\rho = 0$. In this case, $h_n = o(b_n)$ and $\mathsf{C}^{\text{bc}}_{\text{SRD}}(h_n, b_n) = o_p(\mathsf{V}_{\text{SRD}}(h_n))$, thus making our approach asymptotically equivalent to the standard approach to bias correction: $\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, b_n)/\mathsf{V}_{\text{SRD}}(h_n) \to_p 1$.

Case 2: $\rho \in (0, \infty)$. In this case, $h_n = \rho b_n$, a knife-edge case, where both $\hat{\tau}_{\text{SRD}}(h_n)$ and $\hat{\mathsf{B}}_{\text{SRD}}(h_n, b_n)$ contribute to the asymptotic variance.

Case 3: $\rho = \infty$. In this case, $b_n = o(h_n)$ and $\mathsf{V}_{\text{SRD}}(h_n) = o_p(\mathsf{C}^{\text{bc}}_{\text{SRD}}(h_n, b_n))$, implying that the bias estimate is first-order while the actual estimator $\hat{\tau}_{\text{SRD}}(h_n)$ is of smaller order: $\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, b_n)/\mathbb{V}[h_n^2 \hat{\mathsf{B}}_{\text{SRD}}(h_n, b_n)|\mathcal{X}_n] \to_p 1$.

REMARK 4—Higher-Order Implications: Under the smoothness assumptions imposed, if h_n and b_n are such that the confidence intervals have correct asymptotic coverage, then the coverage error of $I^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ decays at a faster rate than the coverage error of $I_{\text{SRD}}(h_n)$. See Calonico, Cattaneo, and Farrell (2014) for further details.

REMARK 5—Interval Length: If $\rho_n = h_n/b_n \to \rho \in [0, \infty)$, then $I^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ and $I_{\text{SRD}}(h_n)$ have interval length proportional to $1/\sqrt{nh_n}$. If, in addition, h_n and b_n are chosen so that the confidence intervals have correct asymptotic coverage, then $I^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ will have shorter interval length than $I_{\text{SRD}}(h_n)$ for n large enough. However, because the proportionality constant is larger for $I^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ than for $I_{\text{SRD}}(h_n)$, the interval $I_{\text{SRD}}(h_n)$ may be shorter than $I^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ in small samples. See Section 6 for simulation evidence, and Calonico, Cattaneo, and Farrell (2014) for further details.

REMARK 6—Bootstrap: Bootstrapping $\hat{\tau}_{\text{SRD}}(h_n)$ or $T_{\text{SRD}}(h_n)$ will not improve the performance of the conventional confidence intervals because the bootstrap distribution is centered at $\mathbb{E}[\hat{\tau}_{\text{SRD}}(h_n)|\mathcal{X}_n]$. Bootstrapping $\hat{\tau}^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ or $T^{\text{bc}}_{\text{SRD}}(h_n, b_n)$ is possible, but these statistics are not asymptotically pivotal in general. Bootstrapping the asymptotically pivotal statistic $T^{\text{rbc}}_{\text{SRD}}(h_n, b_n)$ is possible, as an alternative to the Gaussian approximation. See Horowitz (2001) for further details.

REMARK 7—Special Case $h_n = b_n$: If $h_n = b_n$ (and the same kernel function $k(\cdot)$ is used), then $\hat{\tau}^{\text{bc}}_{\text{SRD}}(h_n, h_n)$ is numerically equivalent to the (not bias-corrected) local-quadratic estimator of τ_{SRD}, and $\mathsf{V}^{\text{bc}}_{\text{SRD}}(h_n, h_n)$ coincides with the variance of the latter estimator. This is true for any polynomial order used (see the Appendix and Supplemental Material), which gives a simple connection between local polynomial estimators of order p and $p+1$ and manual bias correction. Thus, this result provides a formal justification for an inference approach based on increasing the order of the RD estimator: choose h_n to be the MSE-optimal bandwidth for the local-linear estimator, but construct confidence intervals using a t-statistic based on the local-quadratic estimator instead. This approach corresponds to the case $h_n = b_n$ in Theorem 1.

REMARK 8—Nonparametrics and Undersmoothing: Our results apply more broadly to nonparametric kernel-based curve estimation problems, and also offer a new theoretical perspective on the tradeoffs and connections between undersmoothing (i.e., choosing an ad hoc “smaller” bandwidth) and explicit bias correction. See Calonico, Cattaneo, and Farrell (2014) for further details.

REMARK 9—Different Bandwidths: All our results may be extended to allow for different bandwidths entering the estimators for control and treatment units. In this case, the different bandwidth sequences should satisfy the conditions imposed in the theorems.

3. OTHER RD DESIGNS

We discuss three extensions of our approach to other empirically relevant settings: sharp kink RD, fuzzy RD, and fuzzy kink RD designs. The results presented here are special cases of Theorems A.1 and A.2 in the Appendix. In all cases, the construction follows the same logic: (i) the conventional large-sample distribution is characterized, (ii) the leading bias is presented and a plug-in bias correction is proposed, and (iii) the alternative large-sample distribution is derived to obtain robust confidence intervals.

3.1. *Sharp Kink RD*

In the sharp kink RD design, interest lies on the difference of the first derivative of the regression functions at the cutoff, as opposed to the difference in the levels of those functions (see, e.g., Card et al. (2014), Dong (2014), and references therein). The estimand is, up to a known scale factor, $\tau_{\text{SKRD}} = \mu_+^{(1)} - \mu_-^{(1)}$.

Although a local-linear estimator could still be used in this context, it is more appropriate to employ a local-quadratic estimator due to boundary-bias considerations. Thus, we focus on the local-quadratic RD estimator $\hat{\tau}_{\text{SKRD}}(h_n) = \hat{\mu}_{+,2}^{(1)}(h_n) - \hat{\mu}_{-,2}^{(1)}(h_n)$, where $\hat{\mu}_{+,2}^{(1)}(h_n)$ and $\hat{\mu}_{-,2}^{(1)}(h_n)$ denote local-quadratic estimators of $\mu_+^{(1)}$ and $\mu_-^{(1)}$, respectively; see Section A.1 in the Appendix. Lemma A.1(D) in the Appendix gives $T_{\text{SKRD}}(h_n) = (\hat{\tau}_{\text{SKRD}}(h_n) - \tau_{\text{SKRD}})/\sqrt{\mathsf{V}_{\text{SKRD}}(h_n)} \to_d \mathcal{N}(0,1)$ with $\mathsf{V}_{\text{SKRD}}(h_n) = \mathbb{V}[\hat{\tau}_{\text{SKRD}}(h_n)|\mathcal{X}_n]$, which corresponds to the conventional distributional approximation. The MSE-optimal bandwidth choice for $\hat{\tau}_{\text{SKRD}}(h_n)$ is derived in Lemma 1 in Section 4. This choice, among others, will again lead to a non-negligible first-order bias. Proceeding as before, we have $\mathbb{E}[\hat{\tau}_{\text{SKRD}}(h_n)|\mathcal{X}_n] - \tau_{\text{SKRD}} = h_n^2 \mathsf{B}_{\text{SKRD}}(h_n)\{1 + o_p(1)\}$ with $\mathsf{B}_{\text{SKRD}}(h_n) = \mu_+^{(3)} \mathcal{B}_{+,\text{SKRD}}(h_n)/3! - \mu_-^{(3)} \mathcal{B}_{-,\text{SKRD}}(h_n)/3!$, where $\mathcal{B}_{+,\text{SKRD}}(h_n)$ and $\mathcal{B}_{-,\text{SKRD}}(h_n)$ are asymptotically bounded observed quantities (function of $\mathcal{X}_n$, $k(\cdot)$, and h_n), also given in Lemma A.1(B).

A bias-corrected local-quadratic estimator of τ_{SKRD} is $\hat{\tau}^{\text{bc}}_{\text{SKRD}}(h_n, b_n) = \hat{\tau}_{\text{SKRD}}(h_n) - h_n^2 \hat{\mathsf{B}}_{\text{SKRD}}(h_n, b_n)$ with $\hat{\mathsf{B}}_{\text{SKRD}}(h_n, b_n) = \hat{\mu}^{(3)}_{+,3}(b_n)\mathcal{B}_{+,\text{SKRD}}(h_n)/3! - \hat{\mu}^{(3)}_{-,3}(b_n)\mathcal{B}_{-,\text{SKRD}}(h_n)/3!$, where $\hat{\mu}^{(3)}_{+,3}(b_n)$ and $\hat{\mu}^{(3)}_{-,3}(b_n)$ are the local-cubic estimators of $\mu_+^{(3)}$ and $\mu_-^{(3)}$, respectively; see Section A.1 in the Appendix for details.

THEOREM 2: *Let Assumptions 1–2 hold with $S \geq 4$. If $n \min\{h_n^7, b_n^7\} \times \max\{h_n^2, b_n^2\} \to 0$ and $n \min\{h_n, b_n\} \to \infty$, then*

$$T^{\text{rbc}}_{\text{SKRD}}(h_n, b_n) = \frac{\hat{\tau}^{\text{bc}}_{\text{SKRD}}(h_n, b_n) - \tau_{\text{SKRD}}}{\sqrt{\mathsf{V}^{\text{bc}}_{\text{SKRD}}(h_n, b_n)}} \to_d \mathcal{N}(0, 1),$$

provided $\kappa \max\{h_n, b_n\} < \kappa_0$. The exact form of $\mathsf{V}^{\text{bc}}_{\text{SKRD}}(h_n, b_n)$ is given in Theorem A.1(V) *in the Appendix.*

This theorem is analogous to Theorem 1, and derives the new variance formula $\mathsf{V}^{\text{bc}}_{\text{SKRD}}(h_n, b_n)$ for the sharp kink RD design capturing the additional contribution of the bias correction to the sampling variability. The new variance also takes the form $\mathsf{V}^{\text{bc}}_{\text{SKRD}}(h_n, b_n) = \mathsf{V}_{\text{SKRD}}(h_n) + \mathsf{C}^{\text{bc}}_{\text{SKRD}}(h_n, b_n)$, where $\mathsf{C}^{\text{bc}}_{\text{SKRD}}(h_n, b_n)$ is the correction term. This result theoretically justifies the following more robust $100(1-\alpha)$-percent confidence interval for τ_{SKRD}: $I^{\text{rbc}}_{\text{SKRD}}(h_n, b_n) = [\hat{\tau}^{\text{bc}}_{\text{SKRD}}(h_n, b_n) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathsf{V}^{\text{bc}}_{\text{SKRD}}(h_n, b_n)}]$.

3.2. *Fuzzy RD*

In the fuzzy RD design, actual treatment status may differ from treatment assignment and is thus only partially determined by the running variable. We introduce the following notation: $(Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)'$, $i = 1, 2, \ldots, n$, is a random sample where now treatment status for each unit is $T_i = T_i(0) \cdot \mathbf{1}(X_i < 0) + T_i(1) \cdot \mathbf{1}(X_i \geq 0)$, with $T_i(0), T_i(1) \in \{0, 1\}$. The observed random sample is $\{(Y_i, T_i, X_i)' : i = 1, 2, \ldots, n\}$. The estimand of interest is $\tau_{\text{FRD}} = (\mathbb{E}[Y_i(1)|X = 0] - \mathbb{E}[Y_i(0)|X = 0])/(\mathbb{E}[T_i(1)|X = 0] - \mathbb{E}[T_i(0)|X = 0])$, provided that $\mathbb{E}[T_i(1)|X = 0] - \mathbb{E}[T_i(0)|X = 0] \neq 0$. Under appropriate conditions, this estimand is nonparametrically identifiable as

$$\tau_{\text{FRD}} = \frac{\tau_{Y,\text{SRD}}}{\tau_{T,\text{SRD}}} = \frac{\mu_{Y+} - \mu_{Y-}}{\mu_{T+} - \mu_{T-}},$$

where here, and elsewhere as needed, we make explicit the outcome variable underlying the population parameter. That is, $\tau_{Y,\text{SRD}} = \mu_{Y+} - \mu_{Y-}$ with $\mu_{Y+} = \lim_{x\to 0^+} \mu_Y(x)$ and $\mu_{Y-} = \lim_{x\to 0^-} \mu_Y(x)$, $\mu_Y(x) = \mathbb{E}[Y_i|X_i = x]$, and $\tau_{T,\text{SRD}} = \mu_{T+} - \mu_{T-}$ with $\mu_{T+} = \lim_{x\to 0^+} \mu_T(x)$ and $\mu_{T-} = \lim_{x\to 0^-} \mu_T(x)$, $\mu_T(x) = \mathbb{E}[T_i|X_i = x]$. We employ the following additional assumption.

ASSUMPTION 3: *For $\kappa_0 > 0$, the following hold in the neighborhood $(-\kappa_0, \kappa_0)$ around the cutoff $\bar{x} = 0$:*
(a) $\mu_{T-}(x) = \mathbb{E}[T_i(0)|X_i = x]$ *and* $\mu_{T+}(x) = \mathbb{E}[T_i(1)|X_i = x]$ *are S times continuously differentiable*.
(b) $\sigma^2_{T-}(x) = \mathbb{V}[T_i(0)|X_i = x]$ *and* $\sigma^2_{T+}(x) = \mathbb{V}[T_i(1)|X_i = x]$ *are continuous and bounded away from zero*.

A popular estimator in this setting is the ratio of two reduced-form, sharp local-linear RD estimators:

$$\hat{\tau}_{\text{FRD}}(h_n) = \frac{\hat{\tau}_{Y,\text{SRD}}(h_n)}{\hat{\tau}_{T,\text{SRD}}(h_n)} = \frac{\hat{\mu}_{Y+,1}(h_n) - \hat{\mu}_{Y-,1}(h_n)}{\hat{\mu}_{T+,1}(h_n) - \hat{\mu}_{T-,1}(h_n)},$$

again now making explicit the outcome variable being used in each expression. That is, for a random variable U (equal to either Y or T), we set $\hat{\mu}_{U+,1}(h_n)$ and $\hat{\mu}_{U-,1}(h_n)$ to be the local-linear estimators employing U_i as outcome variable; see Section A.1 in the Appendix for details.

Under Assumptions 1–3, and appropriate bandwidth conditions, the conventional large-sample properties of $\hat{\tau}_{\text{FRD}}$ are characterized by noting that $\hat{\tau}_{\text{FRD}}(h_n) - \tau_{\text{FRD}} = \tilde{\tau}_{\text{FRD}}(h_n) + R_n$ with $\tilde{\tau}_{\text{FRD}}(h_n) = (\hat{\tau}_{Y,\text{SRD}}(h_n) - \tau_{Y,\text{SRD}})/\tau_{T,\text{SRD}} - \tau_{Y,\text{SRD}}(\hat{\tau}_{T,\text{SRD}}(h_n) - \tau_{T,\text{SRD}})/\tau^2_{T,\text{SRD}}$ and R_n a higher-order reminder term. This shows that, to first order, the fuzzy RD estimator behaves like a linear combination of two sharp RD estimators. Thus, as Lemma A.2(D) in the Appendix shows,

$$T_{\text{FRD}}(h_n) = \frac{\hat{\tau}_{\text{FRD}}(h_n) - \tau_{\text{FRD}}}{\sqrt{\mathsf{V}_{\text{FRD}}(h_n)}} \to_d \mathcal{N}(0, 1),$$

$$\mathsf{V}_{\text{FRD}}(h_n) = \mathbb{V}\big[\tilde{\tau}_{\text{FRD}}(h_n)|\mathcal{X}_n\big].$$

The bias (after linearization) of the local-linear fuzzy RD estimator $\hat{\tau}_{\text{FRD}}(h_n)$ is $\mathbb{E}[\tilde{\tau}_{\text{FRD}}(h_n)|\mathcal{X}_n] = h_n^2 \mathsf{B}_{\text{FRD}}(h_n)\{1 + o_p(1)\}$ with

$$\mathsf{B}_{\text{FRD}}(h_n) = \left(\frac{1}{\tau_{T,\text{SRD}}}\frac{\mu^{(2)}_{Y+}}{2!} - \frac{\tau_{Y,\text{SRD}}}{\tau^2_{T,\text{SRD}}}\frac{\mu^{(2)}_{T+}}{2!}\right)\mathcal{B}_{+,\text{FRD}}(h_n) - \left(\frac{1}{\tau_{T,\text{SRD}}}\frac{\mu^{(2)}_{Y-}}{2!} - \frac{\tau_{Y,\text{SRD}}}{\tau^2_{T,\text{SRD}}}\frac{\mu^{(2)}_{T-}}{2!}\right)\mathcal{B}_{-,\text{FRD}}(h_n),$$

where $\mathcal{B}_{+,\text{FRD}}(h_n)$ and $\mathcal{B}_{-,\text{FRD}}(h_n)$ are also asymptotically bounded observed quantities (function of $\mathcal{X}_n$, $k(\cdot)$, and h_n) and are given in Lemma A.2(B). A bias-corrected estimator of τ_{FRD} employing a local-quadratic estimate of the leading biases is $\hat{\tau}^{\text{bc}}_{\text{FRD}}(h_n, b_n) = \hat{\tau}_{\text{FRD}}(h_n) - h_n^2 \hat{\mathsf{B}}_{\text{FRD}}(h_n, b_n)$ with

$$\begin{aligned}
&\hat{\mathsf{B}}_{\text{FRD}}(h_n, b_n) \\
&\quad = \left(\frac{1}{\hat{\tau}_{T,\text{SRD}}(h_n)} \frac{\hat{\mu}^{(2)}_{Y+,2}(b_n)}{2!} - \frac{\hat{\tau}_{Y,\text{SRD}}(h_n)}{\hat{\tau}^2_{T,\text{SRD}}(h_n)} \frac{\hat{\mu}^{(2)}_{T+,2}(b_n)}{2!}\right)\mathcal{B}_{+,\text{FRD}}(h_n) \\
&\quad\quad - \left(\frac{1}{\hat{\tau}_{T,\text{SRD}}(h_n)} \frac{\hat{\mu}^{(2)}_{Y-,2}(b_n)}{2!} - \frac{\hat{\tau}_{Y,\text{SRD}}(h_n)}{\hat{\tau}^2_{T,\text{SRD}}(h_n)} \frac{\hat{\mu}^{(2)}_{T-,2}(b_n)}{2!}\right)\mathcal{B}_{-,\text{FRD}}(h_n).
\end{aligned}$$

We propose to bias-correct the fuzzy RD estimator using its first-order linear approximation, as opposed to directly bias-correct $\hat{\tau}_{Y,\text{SRD}}(h_n)$ and $\hat{\tau}_{T,\text{SRD}}(h_n)$ separately in the numerator and denominator of $\hat{\tau}_{\text{FRD}}(h_n)$. The former approach seems more intuitive, as it captures the leading bias of the actual estimator of interest.

THEOREM 3: *Let Assumptions 1–3 hold with* $S \geq 3$, *and* $\tau_{T,\text{SRD}} \neq 0$. *If* $n \min\{h_n^5, b_n^5\} \max\{h_n^2, b_n^2\} \to 0$ *and* $n \min\{h_n, b_n\} \to \infty$, *then*

$$T^{\text{rbc}}_{\text{FRD}}(h_n, b_n) = \frac{\hat{\tau}^{\text{bc}}_{\text{FRD}}(h_n, b_n) - \tau_{\text{FRD}}}{\sqrt{\mathsf{V}^{\text{bc}}_{\text{FRD}}(h_n, b_n)}} \to_d \mathcal{N}(0, 1),$$

provided that $h_n \to 0$ *and* $\kappa b_n < \kappa_0$. *The exact form of* $\mathsf{V}^{\text{bc}}_{\text{FRD}}(h_n, b_n)$ *is given in Theorem* A.2(V).

3.3. *Fuzzy Kink RD*

We retain the notation and assumptions introduced above for the fuzzy RD design. In the fuzzy kink RD, the parameter of interest and plug-in estimators are, respectively,

$$\tau_{\text{FKRD}} = \frac{\tau_{Y,\text{SKRD}}}{\tau_{T,\text{SKRD}}} = \frac{\mu^{(1)}_{Y+} - \mu^{(1)}_{Y-}}{\mu^{(1)}_{T+} - \mu^{(1)}_{T-}}$$

and

$$\hat{\tau}_{\text{FKRD}}(h_n) = \frac{\hat{\tau}_{Y,\text{SKRD}}(h_n)}{\hat{\tau}_{T,\text{SKRD}}(h_n)} = \frac{\hat{\mu}^{(1)}_{Y+,2}(h_n) - \hat{\mu}^{(1)}_{Y-,2}(h_n)}{\hat{\mu}^{(1)}_{T+,2}(h_n) - \hat{\mu}^{(1)}_{T-,2}(h_n)},$$

where $\hat{\tau}_{\text{FKRD}}(h_n)$ is based on two local-quadratic (reduced-form) estimates; see Section A.1 in the Appendix.

The linearization argument given for the fuzzy RD estimator applies here as well. Employing Lemma A.2(D) in the Appendix once more, we verify that $T_{\text{FKRD}}(h_n) = (\hat{\tau}_{\text{FKRD}}(h_n) - \tau_{\text{FKRD}})/\sqrt{\mathsf{V}_{\text{FKRD}}(h_n)} \to_d \mathcal{N}(0,1)$ with $\mathsf{V}_{\text{FKRD}}(h_n) = \mathbb{V}[\tilde{\tau}_{\text{FKRD}}(h_n)|\mathcal{X}_n]$, and $\mathbb{E}[\tilde{\tau}_{\text{FKRD}}(h_n)|\mathcal{X}_n] = h_n^2\mathsf{B}_{\text{FKRD}}(h_n)\{1 + o_p(1)\}$ with $\mathsf{B}_{\text{FKRD}}(h_n) = (\mu_{Y+}^{(3)}/\tau_{T,\text{SKRD}} - \tau_{Y,\text{SKRD}}\mu_{T+}^{(3)}/\tau_{T,\text{SKRD}}^2)\mathcal{B}_{+,\text{FKRD}}(h_n)/3! - (\mu_{Y-}^{(3)}/\tau_{T,\text{SKRD}} - \tau_{Y,\text{SKRD}}\mu_{T-}^{(3)}/\tau_{T,\text{SKRD}}^2)\mathcal{B}_{-,\text{FKRD}}(h_n)/3!$, where $\mathcal{B}_{+,\text{FKRD}}(h_n)$ and $\mathcal{B}_{-,\text{FKRD}}(h_n)$ are also given in Lemma A.2. A plug-in bias-corrected estimator of τ_{FKRD} employing local-cubic estimates of the leading biases is $\hat{\tau}_{\text{FKRD}}^{\text{bc}}(h_n, b_n) = \hat{\tau}_{\text{FKRD}}(h_n) - h_n^2\hat{\mathsf{B}}_{\text{FKRD}}(h_n, b_n)$, where $\hat{\mathsf{B}}_{\text{FKRD}}(h_n, b_n) = (\hat{\mu}_{Y+,3}^{(3)}(b_n)/\hat{\tau}_{T,\text{SKRD}}(h_n) - \hat{\tau}_{Y,\text{SKRD}}(h_n) \times \hat{\mu}_{T+,3}^{(3)}(b_n)/\hat{\tau}_{T,\text{SKRD}}^2(h_n))\mathcal{B}_{+,\text{FKRD}}(h_n)/3! - (\hat{\mu}_{Y-,3}^{(3)}(b_n)/\hat{\tau}_{T,\text{SKRD}}(h_n) - \hat{\tau}_{Y,\text{SKRD}}(h_n) \times \hat{\mu}_{T-,3}^{(3)}(b_n)/\hat{\tau}_{T,\text{SKRD}}^2(h_n))\mathcal{B}_{-,\text{FKRD}}(h_n)/3!$.

THEOREM 4: *Let Assumptions 1–3 hold with* $S \geq 4$, *and* $\tau_{T,\text{SKRD}} \neq 0$. *If* $n\min\{h_n^7, b_n^7\}\max\{h_n^2, b_n^2\} \to 0$ *and* $n\min\{h_n^3, b_n\} \to \infty$, *then*

$$T_{\text{FKRD}}^{\text{rbc}}(h_n, b_n) = \frac{\hat{\tau}_{\text{FKRD}}^{\text{bc}}(h_n, b_n) - \tau_{\text{FKRD}}}{\sqrt{\mathsf{V}_{\text{FKRD}}^{\text{bc}}(h_n, b_n)}} \to_d \mathcal{N}(0,1),$$

provided that $h_n \to 0$ *and* $\kappa b_n < \kappa_0$. *The exact form of* $\mathsf{V}_{\text{FKRD}}^{\text{bc}}(h_n, b_n)$ *is given in Theorem* A.2(V).

4. VALIDITY OF MSE-OPTIMAL BANDWIDTH SELECTORS

Following Imbens and Kalyanaraman (2012), we derive MSE-optimal bandwidth choices for h_n (bandwidth for RD point estimator) and b_n (bandwidth for bias estimator), which can be used to implement in practice all of the results discussed previously. As explained above, these bandwidth choices are not valid when conventional distributional approximations are used, but they are fully compatible with our distributional approach. Data-driven consistent bandwidth selectors are discussed in detail in Calonico, Cattaneo, and Titiunik (2014c, Section S.2.6).

Let $\nu, p, q \in \mathbb{Z}_+$, with $\nu \leq p < q$. Unless otherwise noted, throughout the paper ν will denote the derivative of interest, p will denote the order of the local polynomial point estimator, and q will denote the order of the local polynomial bias estimator. Define $\Gamma_p = \int_0^\infty K(u)r_p(u)r_p(u)'\,\mathrm{d}u$, $\vartheta_{p,q} = \int_0^\infty K(u)u^q r_p(u)\,\mathrm{d}u$, and $\Psi_p = \int_0^\infty K(u)^2 r_p(u)r_p(u)'\,\mathrm{d}u$, where $r_p(x) = (1, x, \dots, x^p)'$, and let e_ν be a conformable $(\nu+1)$th unit vector. See Section A.1 in the Appendix for more details.

4.1. *Sharp Designs*

To handle the sharp RD and sharp kink RD designs together, as well as the choice of pilot bandwidths, we introduce more general notation. The estimands in the sharp RD designs are $\tau_\nu = \mu_+^{(\nu)} - \mu_-^{(\nu)}$ with, in particular, $\tau_{\text{SRD}} = \tau_0$

and $\tau_{\text{SKRD}} = \tau_1$. The pth-order local polynomial RD estimators are $\hat{\tau}_{\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{+,p}(h_n) - \hat{\mu}^{(\nu)}_{-,p}(h_n)$ with $\nu \leq p$ with, in particular, $\hat{\tau}_{\text{SRD}}(h_n) = \hat{\tau}_{0,1}(h_n)$ and $\hat{\tau}_{\text{SKRD}}(h_n) = \hat{\tau}_{1,2}(h_n)$. From Lemma A.1(B) in the Appendix,

$$\mathbb{E}\big[\hat{\tau}_{\nu,p}(h_n)|\mathcal{X}_n\big] - \tau_\nu = h_n^{p+1-\nu}\mathsf{B}_{\nu,p,p+1,0}\big\{1 + o_p(1)\big\},$$
$$\mathsf{B}_{\nu,p,r,s} = \frac{\mu_+^{(r)} - (-1)^{\nu+r+s}\mu_-^{(r)}}{r!}\nu! e_\nu' \Gamma_p^{-1}\vartheta_{p,r},$$

which shows that, in general, the bias of the RD estimator could depend on a difference or a sum of higher-order derivatives. Thus, to develop MSE-optimal bandwidth choices for both RD estimators and their bias estimates, we consider the following generic MSE objective function:

$$\mathsf{MSE}_{\nu,p,s}(\mathsf{h}_n) = \mathbb{E}\Big[\Big(\big(\hat{\mu}^{(\nu)}_{+,p}(\mathsf{h}_n) - (-1)^s\hat{\mu}^{(\nu)}_{-,p}(\mathsf{h}_n)\big) - \big(\mu_+^{(\nu)} - (-1)^s\mu_-^{(\nu)}\big)\Big)^2\Big|\mathcal{X}_n\Big], \quad \nu, p, s \in \mathbb{Z}_+,$$

where h_n denotes a generic vanishing bandwidth sequence. The following lemma gives the general result.

LEMMA 1: *Suppose Assumptions 1–2 hold with $S \geq p+1$, and $\nu \leq p$. If $\mathsf{h}_n \to 0$ and $n\mathsf{h}_n \to \infty$, then*

$$\mathsf{MSE}_{\nu,p,s}(\mathsf{h}_n) = \mathsf{h}_n^{2(p+1-\nu)}\big[\mathsf{B}^2_{\nu,p,p+1,s} + o_p(1)\big] + \frac{1}{n\mathsf{h}_n^{1+2\nu}}\big[\mathsf{V}_{\nu,p} + o_p(1)\big],$$

where $\mathsf{V}_{\nu,p} = (\sigma_-^2 + \sigma_+^2)\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu/f$. If $\mathsf{B}_{\nu,p,p+1,s} \neq 0$, then the (asymptotic) MSE-optimal bandwidth is

$$\mathsf{h}_{\text{MSE},\nu,p,s} = \mathsf{C}^{1/(2p+3)}_{\text{MSE},\nu,p,s} n^{-1/(2p+3)}, \qquad \mathsf{C}_{\text{MSE},\nu,p,s} = \frac{(1+2\nu)\mathsf{V}_{\nu,p}}{2(p+1-\nu)\mathsf{B}^2_{\nu,p,p+1,s}}.$$

This lemma justifies a set of MSE-optimal (infeasible) choices for h_n and b_n, which are determined by the estimand (ν and s) and estimator (p). For example, in the case of Theorem 1, $h_n = \mathsf{h}_{\text{MSE},0,1,0}$ is MSE-optimal for $\hat{\tau}_{\text{SRD}}(h_n)$ and $b_n = \mathsf{h}_{\text{MSE},2,2,2}$ is MSE-optimal for the bias estimate (of $\hat{\tau}_{\text{SRD}}(h_n)$). Similarly, for Theorem 2, $h_n = \mathsf{h}_{\text{MSE},1,2,0}$ and $b_n = \mathsf{h}_{\text{MSE},3,3,3}$ are the MSE-optimal choices for the local-quadratic estimator $\hat{\tau}_{\text{SKRD}}(h_n)$ and its bias estimate, respectively. For the general case, depending on the choice of (ν, p, q), see Calonico, Cattaneo, and Titiunik (2014c, Section S.2.6).

REMARK 10—Bandwidths Validity: The MSE-optimal bandwidth choices are fully compatible with our confidence intervals because they satisfy the rate restrictions in Theorems 1–2. For example, in Theorem 1, $n\min\{\mathsf{h}_{\text{MSE},0,1,0}, \mathsf{h}_{\text{MSE},2,2,2}\} \to \infty$ and $n\min\{\mathsf{h}^5_{\text{MSE},0,1,0}, \mathsf{h}^5_{\text{MSE},2,2,2}\}\max\{\mathsf{h}^2_{\text{MSE},0,1,0}, \mathsf{h}^2_{\text{MSE},2,2,2}\} \to 0$.

REMARK 11—Estimated Bandwidths: Section S.2.6 of the Supplemental Material describes general data-driven direct plug-in (DPI) bandwidth selectors for sharp RD designs based on Lemma 1. Following Imbens and Kalyanaraman (2012), our proposed bandwidths incorporate "regularization" to avoid small denominators. But our bandwidth selectors are different from the selectors proposed by Imbens and Kalyanaraman (2012) in two ways: (i) our estimator of $\mathsf{V}_{\nu,p}$ avoids estimating σ^2_+, σ^2_-, and f directly, and (ii) pilot bandwidths are chosen to be MSE-optimal and thus the final bandwidth selectors are of the ℓ-stage DPI variety (Wand and Jones (1995, Section 3.6)). Our final bandwidth selectors are consistent and optimal in the sense of Li (1987); see Calonico, Cattaneo, and Titiunik (2014c, Section S.2.6, Theorem A.4).

REMARK 12—Optimal ρ_n: The MSE-optimal bandwidth choices imply $\rho_n \to 0$. In research underway, we are investigating whether this is an optimal choice from a distributional approximation perspective. See Remarks 4 and 5, and Calonico, Cattaneo, and Farrell (2014) for related discussion.

4.2. *Fuzzy Designs*

Let $\varsigma_\nu = \tau_{Y,\nu}/\tau_{T,\nu}$ with $\tau_{Y,\nu} = \mu^{(\nu)}_{Y+} - \mu^{(\nu)}_{Y-}$ and $\tau_{T,\nu} = \mu^{(\nu)}_{T+} - \mu^{(\nu)}_{T-}$. In particular, $\tau_{\text{FRD}} = \varsigma_0$ and $\tau_{\text{FKRD}} = \varsigma_1$. The pth-order local polynomial estimators are $\hat{\varsigma}_{\nu,p}(h_n) = \hat{\tau}_{Y,\nu,p}(h_n)/\hat{\tau}_{T,\nu,p}(h_n)$ with $\nu \le p$, $\hat{\tau}_{Y,\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{Y+,p}(h_n) - \hat{\mu}^{(\nu)}_{Y-,p}(h_n)$, and $\hat{\tau}_{T,\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{T+,p}(h_n) - \hat{\mu}^{(\nu)}_{T-,p}(h_n)$; see Section A.1 in the Appendix. In particular, $\hat{\tau}_{\text{FRD}}(h_n) = \hat{\varsigma}_{0,1}(h_n)$ and $\hat{\tau}_{\text{FKRD}}(h_n) = \hat{\varsigma}_{1,2}(h_n)$. The first-order linear approximation of $\hat{\varsigma}_{\nu,p}(h_n)$ is $\tilde{\varsigma}_{\nu,p}(h_n) = (\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu})/\tau_{T,\nu} - \tau_{Y,\nu}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})/\tau^2_{T,\nu}$, which we employ to construct the (approximate) MSE objective function for the main RD point estimators.

LEMMA 2: *Suppose Assumptions 1–3 hold with $S \ge p+1$, and $\nu \le p$. If $h_n \to 0$ and $nh_n \to \infty$, then*

$$\mathbb{E}\big[\big(\tilde{\varsigma}_{\nu,p}(h_n)\big)^2|\mathcal{X}_n\big] = h_n^{2(p+1-\nu)}\big[\mathsf{B}^2_{\text{F},\nu,p,p+1} + o_p(1)\big] + n^{-1}h_n^{-1-2\nu}\big[\mathsf{V}_{\text{F},\nu,p} + o_p(1)\big],$$

where

$$\mathsf{B}_{\text{F},\nu,p,r} = \left(\frac{1}{\tau_{T,\nu}}\frac{\mu^{(r)}_{Y+} - (-1)^{\nu+r}\mu^{(r)}_{Y-}}{r!} - \frac{\tau_{Y,\nu}}{\tau^2_{T,\nu}}\frac{\mu^{(r)}_{T+} - (-1)^{\nu+r}\mu^{(r)}_{T-}}{r!}\right)\nu! e'_\nu \Gamma_p^{-1}\vartheta_{p,r}$$

and $\mathsf{V}_{\text{F},\nu,p} = ((\sigma^2_{YY-} + \sigma^2_{YY+})/\tau^2_{T,\nu} - 2\tau_{Y,\nu}(\sigma^2_{YT-} + \sigma^2_{YT+})/\tau^3_{T,\nu} + \tau^2_{Y,\nu}(\sigma^2_{TT-} + \sigma^2_{TT+})/\tau^4_{T,\nu})\nu!^2 e'_\nu \Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu/f$. *If* $\mathsf{B}_{\text{F},\nu,p,p+1} \neq 0$, *then the (asymptotic) MSE-*

optimal bandwidth is

$$h_{\mathrm{MSE,F},\nu,p} = C_{\mathrm{MSE,F},\nu,p}^{1/(2p+3)} n^{-1/(2p+3)}, \quad C_{\mathrm{MSE,F},\nu,p} = \frac{(2\nu+1)\mathsf{V}_{\mathrm{F},\nu,p}}{2(p+1-\nu)\mathsf{B}^2_{\mathrm{F},\nu,p,p+1}}.$$

Valid bandwidth choices of h_n for fuzzy RD designs are also readily available using Lemma 2: $h_n = h_{\mathrm{MSE,F},0,1}$ for Theorem 3 and $h_n = h_{\mathrm{MSE,F},1,2}$ for Theorem 4. Following the logic outlined for sharp RD designs, it is possible to develop an MSE-optimal choice of the bandwidth b_n entering the bias estimator of the fuzzy RD estimators. As a simpler alternative, Lemma 1 can be used directly on the estimators of $\mu^{(r)}_{Y+} - (-1)^{\nu+r}\mu^{(r)}_{Y-}$ and $\mu^{(r)}_{T+} - (-1)^{\nu+r}\mu^{(r)}_{T-}$, separately. Feasible versions of these bandwidth selectors for fuzzy RD designs can also be developed along the lines of Section S.2.6 in the Supplemental Material. Importantly, just as in the sharp RD cases (Remark 10), these MSE-optimal bandwidth choices will be fully compatible with our asymptotic approximations.

5. STANDARD ERRORS

The exact formulas for the new variances $\mathsf{V}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n, b_n)$ [sharp RD], $\mathsf{V}^{\mathrm{bc}}_{\mathrm{SKRD}}(h_n, b_n)$ [sharp kink RD], $\mathsf{V}^{\mathrm{bc}}_{\mathrm{FRD}}(h_n, b_n)$ [fuzzy RD], and $\mathsf{V}^{\mathrm{bc}}_{\mathrm{FKRD}}(h_n, b_n)$ [fuzzy kink RD] in Theorems 1–4, respectively, are straightforward to derive but notationally cumbersome. They all have the same structure because they are derived by computing the conditional variance of (linear combinations of weighted) linear least-squares estimators. The only unknowns in these variance matrices are, depending on the setting under consideration (sharp or fuzzy RD designs), the matrices $\Psi_{YY+,p,q}(h_n, b_n)$, $\Psi_{YT+,p,q}(h_n, b_n)$, $\Psi_{TT+,p,q}(h_n, b_n)$, $\Psi_{YY-,p,q}(h_n, b_n)$, $\Psi_{YT-,p,q}(h_n, b_n)$, and $\Psi_{TT-,p,q}(h_n, b_n)$, with $p, q \in \mathbb{N}_+$ and the generic notation

$$\Psi_{UV+,p,q}(h_n, b_n) = \sum_{i=1}^{n} \mathbf{1}(X_i \geq 0) K_{h_n}(X_i) K_{b_n}(X_i) \times r_p(X_i/h_n) r_q(X_i/b_n)' \sigma^2_{UV+}(X_i)/n,$$

$$\Psi_{UV-,p,q}(h_n, b_n) = \sum_{i=1}^{n} \mathbf{1}(X_i < 0) K_{h_n}(X_i) K_{b_n}(X_i) \times r_p(X_i/h_n) r_q(X_i/b_n)' \sigma^2_{UV-}(X_i)/n,$$

where $\sigma^2_{UV+}(x) = \mathbb{C}\mathrm{ov}[U(1), V(1)|X = x]$ and $\sigma^2_{UV-}(x) = \mathbb{C}\mathrm{ov}[U(0), V(0)| X = x]$, and U and V are placeholders for either Y or T. This generality is required to handle the fuzzy designs, where the covariances between Y_i and T_i arise naturally. Theorems A.1 and A.2 in the Appendix give the exact standard error formulas, showing how the matrices $\Psi_{UV+,p,q}(h_n, b_n)$ and $\Psi_{UV-,p,q}(h_n, b_n)$ are employed.

The $(p+1)\times(q+1)$ matrices $\Psi_{UV+,p,q}(h_n,b_n)$ and $\Psi_{UV-,p,q}(h_n,b_n)$ are a generalization of the middle matrix in the traditional Huber–Eicker–White heteroskedasticity-robust standard error formula for linear models, and thus an analogue of this standard error estimator can be constructed by plugging in the corresponding estimated residuals. See Calonico, Cattaneo, and Titiunik (2014c, Section S.2.4) for further details. This choice, although simple and convenient, may not perform well in finite-samples because it implicitly employs the bandwidth choices used to construct the estimates of the underlying regression functions. As an alternative, following Abadie and Imbens (2006), we propose standard error estimators based on nearest-neighbor estimators with a fixed tuning parameter, which may be more robust in finite samples. Specifically, we define

$$\begin{aligned}\hat{\Psi}_{UV+,p,q}(h_n,b_n) &= \sum_{i=1}^{n}\mathbf{1}(X_i\geq 0)K_{h_n}(X_i)K_{b_n}(X_i)\\ &\quad\times r_p(X_i/h_n)r_q(X_i/h_n)'\hat{\sigma}^2_{UV+}(X_i)/n,\\ \hat{\Psi}_{UV-,p,q}(h_n,b_n) &= \sum_{i=1}^{n}\mathbf{1}(X_i< 0)K_{h_n}(X_i)K_{b_n}(X_i)\\ &\quad\times r_p(X_i/h_n)r_q(X_i/h_n)'\hat{\sigma}^2_{UV-}(X_i)/n,\end{aligned}$$

with

$$\begin{aligned}\hat{\sigma}^2_{UV+}(X_i) &= \mathbf{1}(X_i\geq 0)\frac{J}{J+1}\\ &\quad\times\left(U_i-\sum_{j=1}^{J}U_{\ell_{+,j}(i)}/J\right)\left(V_i-\sum_{j=1}^{J}V_{\ell_{+,j}(i)}/J\right),\\ \hat{\sigma}^2_{UV-}(X_i) &= \mathbf{1}(X_i< 0)\frac{J}{J+1}\\ &\quad\times\left(U_i-\sum_{j=1}^{J}U_{\ell_{-,j}(i)}/J\right)\left(V_i-\sum_{j=1}^{J}V_{\ell_{-,j}(i)}/J\right),\end{aligned}$$

where $\ell_j^+(i)$ is the jth closest unit to unit i among $\{X_i : X_i\geq 0\}$ and $\ell_j^-(i)$ is the jth closest unit to unit i among $\{X_i : X_i<0\}$, and J denotes the number of neighbors. ("Local sample covariances" could be used instead; see Abadie and Imbens (2010).)

In the Supplemental Material (Calonico, Cattaneo, and Titiunik (2014c, Section S.2.4)), we show that these estimators are asymptotically valid for any choice of $J\in\mathbb{N}_+$, because they are approximately conditionally unbiased (even though inconsistent for fixed nearest-neighbors $J\geq 1$). This justifies employing $\hat{\Psi}_{UV+,p,q}(h_n,b_n)$ and $\hat{\Psi}_{UV-,p,q}(h_n,b_n)$ in place of $\Psi_{UV+,p,q}(h_n,b_n)$ and $\Psi_{UV-,p,q}(h_n,b_n)$ to construct the estimators $\hat{\mathsf{V}}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n,b_n)$, $\hat{\mathsf{V}}^{\mathrm{bc}}_{\mathrm{SKRD}}(h_n,b_n)$,

$\hat{\mathsf{V}}^{\mathrm{bc}}_{\mathrm{FRD}}(h_n, b_n)$, and $\hat{\mathsf{V}}^{\mathrm{bc}}_{\mathrm{FKRD}}(h_n, b_n)$. For example, in Theorem 1, feasible confidence intervals are

$$\hat{I}^{\mathrm{rbc}}_{\mathrm{SRD}}(h_n, b_n) = \left[\hat{\tau}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n, b_n) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\hat{\mathsf{V}}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n, b_n)}\right],$$

where $\hat{\mathsf{V}}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n, b_n)$ is constructed using

$$\hat{\Psi}_{YY+,1,1}(h_n, b_n), \quad \hat{\Psi}_{YY+,1,2}(h_n, b_n), \quad \hat{\Psi}_{YY+,2,1}(h_n, b_n),$$
$$\hat{\Psi}_{YY+,2,2}(h_n, b_n), \quad \hat{\Psi}_{YY-,1,1}(h_n, b_n), \quad \hat{\Psi}_{YY-,1,2}(h_n, b_n),$$
$$\hat{\Psi}_{YY-,2,1}(h_n, b_n), \quad \text{and} \quad \hat{\Psi}_{YY-,2,2}(h_n, b_n).$$

The other feasible confidence intervals are constructed analogously.

6. SIMULATION EVIDENCE

We report the main results of a Monte Carlo experiment. We conducted 5000 replications, and for each replication we generated a random sample $\{(X_i, \varepsilon_i)' : i = 1, \dots, n\}$ with size $n = 500$, $X_i \sim 2\mathcal{B}(2, 4) - 1$ with $\mathcal{B}(p_1, p_2)$ denoting a beta distribution with parameters p_1 and p_2, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ with $\sigma_\varepsilon = 0.1295$. We considered three regression functions (Figure 1), denoted $\mu_1(x)$, $\mu_2(x)$, and $\mu_3(x)$, and labeled Model 1, 2, and 3, respectively. The outcome was generated as $Y_i = \mu_j(X_i) + \varepsilon_i$, $i = 1, 2, \dots, n$, for each regression model $j = 1, 2, 3$. The exact functional form of $\mu_1(x)$ and $\mu_2(x)$ was obtained from the data in Lee (2008) and Ludwig and Miller (2007), respectively, while $\mu_3(x)$ was chosen to exhibit more curvature. All other features of the simulation study were held fixed, matching exactly the data generating process in Imbens and Kalyanaraman (2012). For further details, see Calonico, Cattaneo, and Titiunik (2014c, Section S.3).

We consider confidence intervals for τ_{SRD} (sharp RD), employing a local-linear RD estimator ($p = 1$) with local-quadratic bias correction ($q = 2$), de-

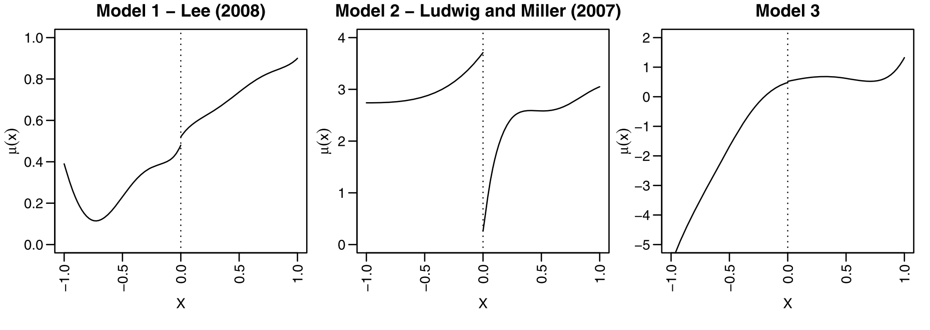

FIGURE 1.—Regression functions for models 1–3 in simulations.

noted $\hat{\tau}^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n)$ as in Section 2. We report empirical coverage and interval length of conventional (based on $T_{\mathtt{SRD}}(h_n)$) and robust (based on $T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n)$) 95% confidence intervals for different bandwidth choices:

$$\hat{I}_{\mathtt{SRD}}(h_n) = \left[\hat{\tau}_{\mathtt{SRD}}(h_n) \pm 1.96\sqrt{\hat{\mathsf{V}}_{\mathtt{SRD}}(h_n)}\right]$$

and

$$\hat{I}^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n) = \left[\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n) \pm 1.96\sqrt{\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)}\right],$$

where the estimators $\hat{\mathsf{V}}_{\mathtt{SRD}}(h_n)$ and $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ are constructed using the nearest-neighbor procedure discussed in Section 5 with $J = 3$. For comparison, we also report infeasible confidence intervals employing infeasible standard errors ($\mathsf{V}_{\mathtt{SRD}}(h_n)$ and $\mathsf{V}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$), and those constructed using the standard "plug-in estimated residuals" approach, which we denote $\check{\mathsf{V}}_{\mathtt{SRD}}(h_n)$ and $\check{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$.

Table I presents the main simulation results. The main bandwidth h_n is chosen in four different ways: (i) infeasible MSE-optimal choice $h_{\mathtt{MSE},0,1}$, denoted $h_{\mathtt{MSE}}$; (ii) plug-in, regularized MSE-optimal selector as described in Imbens and Kalyanaraman (2012, Section 4.1), denoted $\hat{h}_{\mathtt{IK}}$; (iii) cross-validation as described in Imbens and Kalyanaraman (2012, Section 4.5), denoted $\hat{h}_{\mathtt{CV}}$; and (iv) plug-in choice proposed in Section 4 (Remark 11) above, denoted $\hat{h}_{\mathtt{CCT}}$. Similarly, to choose the pilot bandwidth b_n, we construct modified versions of the choices just enumerated, with the exception of $\hat{h}_{\mathtt{CV}}$ because cross-validation is not readily available for derivative estimation; these choices are denoted $b_{\mathtt{MSE}}$, $\hat{b}_{\mathtt{IK}}$, and $\hat{b}_{\mathtt{CCT}}$, respectively. For further results, including other bandwidth selectors and test statistics, see Calonico, Cattaneo, and Titiunik (2014c, Section S.3).

The simulation results show that the robust confidence intervals lead to important improvements in empirical coverage (EC) with moderate increments in average empirical interval length (IL). The empirical coverage of the interval estimator $\hat{I}^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n)$ exhibits an improvement of about 10–15 percentage points on average with respect to the conventional interval $I_{\mathtt{SRD}}(h_n)$, depending on the particular model, standard error estimator, and bandwidth choices considered. As expected, the feasible versions of the confidence intervals exhibit slightly more empirical coverage distortion and longer intervals than their infeasible counterparts. The conventional plug-in residual standard error estimators ($\check{\mathsf{V}}_{\mathtt{SRD}}(h_n)$ and $\check{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$) tend to exhibit more undercoverage in our simulations than the proposed fixed-neighbor standard error estimators ($\hat{\mathsf{V}}_{\mathtt{SRD}}(h_n)$ and $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$). The choice $\rho_n = 1$ (equivalent to employing a local-quadratic estimator) is not only simple and intuitive (Remark 7), but also performed well in our simulations. Although not the main goal of this paper,

TABLE I

EMPIRICAL COVERAGE AND AVERAGE INTERVAL LENGTH OF DIFFERENT 95% CONFIDENCE INTERVALS[a]

	Conventional Approach						Robust Approach							Bandwidths	
	EC (%)			IL				EC (%)			IL				
	v	$\hat{\mathsf{v}}$	$\check{\mathsf{v}}$	v	$\hat{\mathsf{v}}$	$\check{\mathsf{v}}$		v^{bc}	$\hat{\mathsf{v}}^{\mathsf{bc}}$	$\check{\mathsf{v}}^{\mathsf{bc}}$	v^{bc}	$\hat{\mathsf{v}}^{\mathsf{bc}}$	$\check{\mathsf{v}}^{\mathsf{bc}}$	h_n	b_n
Model 1															
$I_{\texttt{SRD}}(h_{\texttt{MSE}})$	93.5	92.0	91.0	0.225	0.223	0.213	$I^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	94.5	93.0	92.2	0.273	0.270	0.258	0.166	0.251
$I_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	83.4	82.3	81.5	0.153	0.152	0.149	$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	92.4	91.4	91.1	0.270	0.267	0.262	0.375	0.350
$I_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	80.8	79.7	79.0	0.145	0.144	0.141	$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	91.8	90.5	90.0	0.213	0.211	0.206	0.428	0.428
$I_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}})$	90.7	89.4	88.4	0.206	0.203	0.195	$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	92.7	91.6	90.7	0.243	0.239	0.231	0.204	0.332
							$I^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	94.7	92.4	92.0	0.339	0.332	0.315	0.166	0.166
							$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	92.8	91.5	91.2	0.226	0.223	0.219	0.375	0.375
							$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	94.8	92.8	92.4	0.308	0.300	0.288	0.204	0.204
Model 2															
$I_{\texttt{SRD}}(h_{\texttt{MSE}})$	92.7	91.3	86.4	0.327	0.355	0.290	$I^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	94.8	93.6	89.9	0.355	0.386	0.315	0.082	0.189
$I_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	27.2	30.3	30.1	0.214	0.225	0.223	$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	89.3	89.5	90.1	0.247	0.261	0.262	0.184	0.325
$I_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	76.8	77.3	72.8	0.264	0.281	0.249	$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	94.6	93.5	91.6	0.401	0.439	0.376	0.124	0.124
$I_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}})$	87.4	87.3	80.8	0.300	0.319	0.265	$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	94.1	93.2	90.5	0.326	0.347	0.289	0.097	0.223
							$I^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	95.2	93.3	89.3	0.513	0.569	0.441	0.082	0.082
							$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	94.1	93.6	94.4	0.320	0.345	0.344	0.184	0.184
							$I^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	94.7	93.2	90.5	0.465	0.508	0.399	0.097	0.097

(*Continues*)

TABLE I—*Continued*

	Conventional Approach						Robust Approach							Bandwidths	
	EC (%)			IL				EC (%)			IL				
	V	$\hat{\mathsf{V}}$	$\check{\mathsf{V}}$	V	$\hat{\mathsf{V}}$	$\check{\mathsf{V}}$		V^{bc}	$\hat{\mathsf{V}}^{\mathrm{bc}}$	$\check{\mathsf{V}}^{\mathrm{bc}}$	V^{bc}	$\hat{\mathsf{V}}^{\mathrm{bc}}$	$\check{\mathsf{V}}^{\mathrm{bc}}$	h_n	b_n
Model 3															
$I_{\mathtt{SRD}}(h_{\mathtt{MSE}})$	85.8	84.6	84.0	0.179	0.178	0.175	$I^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	94.7	93.5	93.6	0.235	0.233	0.229	0.260	0.322
$I_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}})$	85.7	84.2	83.6	0.187	0.185	0.181	$I^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	94.8	93.6	93.5	0.234	0.231	0.227	0.241	0.352
$I_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}})$	93.1	91.6	90.8	0.219	0.217	0.207	$I^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	94.9	92.6	92.2	0.329	0.322	0.307	0.177	0.177
$I_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}})$	91.4	89.8	89.1	0.216	0.213	0.205	$I^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	95.0	93.3	92.6	0.249	0.245	0.236	0.183	0.329
							$I^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	94.9	93.2	93.4	0.266	0.262	0.258	0.260	0.260
							$I^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	94.9	93.2	93.2	0.278	0.274	0.268	0.241	0.241
							$I^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	95.4	93.1	92.5	0.324	0.316	0.302	0.183	0.183

[a](i) EC denotes empirical coverage in percentage points; (ii) IL denotes empirical average interval length; (iii) columns under "Bandwidths" report the population and average estimated bandwidth choices, as appropriate, for main bandwidth h_n and pilot bandwidth b_n; (iv) $\mathsf{V} = \mathsf{V}(h_n)$ and $\mathsf{V}^{\mathrm{bc}} = \mathsf{V}^{\mathrm{bc}}(h_n, b_n)$ denote infeasible variance estimators using the population variance of the residuals, $\hat{\mathsf{V}} = \hat{\mathsf{V}}(h_n)$ and $\hat{\mathsf{V}}^{\mathrm{bc}} = \hat{\mathsf{V}}^{\mathrm{bc}}(h_n, b_n)$ denote variance estimators constructed using nearest-neighbor standard errors with $J = 3$, and $\check{\mathsf{V}} = \check{\mathsf{V}}(h_n)$ and $\check{\mathsf{V}}^{\mathrm{bc}} = \check{\mathsf{V}}^{\mathrm{bc}}(h_n, b_n)$ denote variance estimators constructed using the conventional plug-in estimated residuals.

we also found that our two-stage direct plug-in rule selector of h_n performs well relative to the other plug-in selectors, and on par with the cross-validation bandwidth selector.

7. CONCLUSION

We introduced new confidence interval estimators for several regression-discontinuity estimands that enjoy demonstrably superior robustness properties. The results cover the sharp (level or kink) and fuzzy (level or kink) RD designs. Our confidence intervals were constructed using an alternative asymptotic theory for bias-corrected local polynomial estimators in the context of RD designs, which leads to a different asymptotic variance in general and thus justifies a new standard error estimator. We found that the resulting data-driven confidence intervals performed very well in simulations, suggesting in particular that they provide a robust (to the choice of bandwidths) alternative when compared to the conventional confidence intervals routinely employed in empirical work.

APPENDIX

In this appendix, we summarize our main results for arbitrary order of local polynomials. Here p denotes the order of main RD estimator, while q denotes the order in the bias correction. Proofs and other results are given in the Supplemental Material (Calonico, Cattaneo, and Titiunik (2014c)).

A.1. *Local Polynomial Estimators and Other Notation*

For $\nu, p \in \mathbb{N}$ with $\nu \leq p$, the pth-order local polynomial estimators of the νth-order derivatives $\mu_{Y+}^{(\nu)}$ and $\mu_{Y-}^{(\nu)}$ are

$$\hat{\mu}_{Y+,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\beta}_{Y+,p}(h_n),$$

$$\hat{\beta}_{Y+,p}(h_n) = \arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \mathbf{1}(X_i \geq 0)\big(Y_i - r_p(X_i)'\beta\big)^2 K_{h_n}(X_i),$$

$$\hat{\mu}_{Y-,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\beta}_{Y-,p}(h_n),$$

$$\hat{\beta}_{Y-,p}(h_n) = \arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \mathbf{1}(X_i < 0)\big(Y_i - r_p(X_i)'\beta\big)^2 K_{h_n}(X_i),$$

where $r_p(x) = (1, x, \ldots, x^p)'$, e_ν is the conformable $(\nu+1)$th unit vector (e.g., $e_1 = (0, 1, 0)'$ if $p = 2$), $K_h(u) = K(u/h)/h$, and h_n is a positive bandwidth sequence. (We drop the evaluation point of functions at $\bar{x} = 0$ to simplify notation.) Let $Y = (Y_1, \ldots, Y_n)'$, $X_p(h) = [r_p(X_1/h), \ldots, r_p(X_n/h)]'$, $S_p(h) = [(X_1/h)^p, \ldots, (X_n/h)^p]'$, $W_+(h) = \operatorname{diag}(\mathbf{1}(X_1 \geq 0)K_h(X_1), \ldots, \mathbf{1}(X_n \geq 0)K_h(X_n))$, $W_-(h) = \operatorname{diag}(\mathbf{1}(X_1 < 0)K_h(X_1), \ldots, \mathbf{1}(X_n < 0)K_h(X_n))$, $\Gamma_{+,p}(h) = X_p(h)'W_+(h)X_p(h)/n$, and $\Gamma_{-,p}(h) = X_p(h)'W_-(h)X_p(h)/n$, with

$\text{diag}(a_1, \dots, a_n)$ denoting the $(n \times n)$ diagonal matrix with diagonal elements $a_1, \dots, a_n$. It follows that $\hat{\beta}_{Y+,p}(h_n) = H_p(h_n)\Gamma_{+,p}^{-1}(h_n)X_p(h_n)'W_+(h_n)Y/n$ and $\hat{\beta}_{Y-,p}(h_n) = H_p(h_n)\Gamma_{-,p}^{-1}(h_n)X_p(h_n)'W_-(h_n)Y/n$, with $H_p(h) = \text{diag}(1, h^{-1}, \dots, h^{-p})$. We set $\hat{\mu}_{Y+,p}(h_n) = \hat{\mu}_{Y+,p}^{(0)}(h_n)$ and $\hat{\mu}_{Y-,p}(h_n) = \hat{\mu}_{Y-,p}^{(0)}(h_n)$ and, whenever possible, we also drop the outcome variable subindex notation. Under conditions given below, $\hat{\beta}_{+,p}(h_n) \to_p \beta_{+,p} = (\mu_+, \mu_+^{(1)}/1!, \mu_+^{(2)}/2!, \dots, \mu_+^{(p)}/p!)'$ and $\hat{\beta}_{-,p}(h_n) \to_p \beta_{-,p} = (\mu_-, \mu_-^{(1)}/1!, \mu_-^{(2)}/2!, \dots, \mu_-^{(p)}/p!)'$, implying that local polynomial regression estimates consistently the level of the unknown regression function (μ_+ and μ_-) as well as its first p derivatives (up to a known scale).

We also employ the following notation: $\vartheta_{+,p,q}(h) = X_p(h)'W_+(h)S_q(h)/n$ and $\vartheta_{-,p,q}(h) = X_p(h)'W_-(h)S_q(h)/n$, and $\Psi_{UV+,p,q}(h, b) = X_p(h)'W_+(h) \times \Sigma_{UV}W_+(b)X_q(b)/n$ and $\Psi_{UV-,p,q}(h, b) = X_p(h)'W_-(h)\Sigma_{UV}W_-(b)X_q(b)/n$ with $\Sigma_{UV} = \text{diag}(\sigma_{UV}^2(X_1), \dots, \sigma_{UV}^2(X_n))$ with $\sigma_{UV}^2(X_i) = \mathbb{C}\text{ov}[U_i, V_i|X_i]$, where U and V are placeholders for Y or T. We set $\Psi_{UV+,p}(h) = \Psi_{UV+,p,p}(h, h)$ and $\Psi_{UV-,p}(h) = \Psi_{UV-,p,p}(h, h)$ for brevity, and drop the outcome variable subindex notation whenever possible. Recall that $\Gamma_p = \int_0^\infty K(u)r_p(u)r_p(u)'\,\mathrm{d}u$, $\vartheta_{p,q} = \int_0^\infty K(u)u^q r_p(u)\,\mathrm{d}u$, and $\Psi_p = \int_0^\infty K(u)^2 r_p(u)r_p(u)'\,\mathrm{d}u$.

A.2. *Sharp RD Designs*

As in the main text, in this section we drop the notational dependence on the outcome variable Y. The general estimand is $\tau_\nu = \mu_+^{(\nu)} - \mu_-^{(\nu)}$ with $\mu_+^{(\nu)} = \nu! e_\nu' \beta_{+,p}$ and $\mu_-^{(\nu)} = \nu! e_\nu' \beta_{-,p}$, $\nu \le p$. Recall that $\tau_{\text{SRD}} = \tau_0$ and $\tau_{\text{SKRD}} = \tau_1$. For any $\nu \le p$, the conventional pth-order local polynomial RD estimator is $\hat{\tau}_{\nu,p}(h_n) = \hat{\mu}_{+,p}^{(\nu)}(h_n) - \hat{\mu}_{-,p}^{(\nu)}(h_n)$ with $\hat{\mu}_{+,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\beta}_{+,p}(h_n)$ and $\hat{\mu}_{-,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\beta}_{-,p}(h_n)$. Recall that $\hat{\tau}_{\text{SRD}}(h_n) = \hat{\tau}_{0,1}(h_n)$ and $\hat{\tau}_{\text{SKRD}}(h_n) = \hat{\tau}_{1,2}(h_n)$.

The following lemma describes the asymptotic bias, variance, and distribution of $\hat{\tau}_{\nu,p}(h_n)$.

LEMMA A.1: *Suppose Assumptions 1–2 hold with $S \ge p + 2$, and $nh_n \to \infty$. Let $r \in \mathbb{N}$ and $\nu \le p$.*

(B) *If $h_n \to 0$, then*

$$\mathbb{E}\big[\hat{\tau}_{\nu,p}(h_n)|\mathcal{X}_n\big] = \tau_\nu + h_n^{p+1-\nu}\mathsf{B}_{\nu,p,p+1}(h_n) + h_n^{p+2-\nu}\mathsf{B}_{\nu,p,p+2}(h_n) + o_p\big(h_n^{p+2-\nu}\big),$$

where

$$\mathsf{B}_{\nu,p,r}(h_n) = \mu_+^{(r)}\mathcal{B}_{+,\nu,p,r}(h_n)/r! - \mu_-^{(r)}\mathcal{B}_{-,\nu,p,r}(h_n)/r!,$$

$$\mathcal{B}_{+,\nu,p,r}(h_n) = \nu! e_\nu' \Gamma_{+,p}^{-1}(h_n)\vartheta_{+,p,r}(h_n) = \nu! e_\nu' \Gamma_p^{-1}\vartheta_{p,r} + o_p(1),$$

$$\mathcal{B}_{-,\nu,p,r}(h_n) = \nu! e_\nu' \Gamma_{-,p}^{-1}(h_n)\vartheta_{-,p,r}(h_n) = (-1)^{\nu+r}\nu! e_\nu' \Gamma_p^{-1}\vartheta_{p,r} + o_p(1).$$

(V) *If* $h_n \to 0$, *then* $\mathsf{V}_{\nu,p}(h_n) = \mathbb{V}[\hat{\tau}_{\nu,p}(h_n)|\mathcal{X}_n] = \mathcal{V}_{+,\nu,p}(h_n) + \mathcal{V}_{-,\nu,p}(h_n)$ *with*

$$\begin{aligned}\mathcal{V}_{+,\nu,p}(h_n) &= n^{-1}h_n^{-2\nu}\nu!^2 e_\nu' \Gamma_{+,p}^{-1}(h_n)\Psi_{+,p}(h_n)\Gamma_{+,p}^{-1}(h_n)e_\nu \\ &= n^{-1}h_n^{-1-2\nu}\sigma_+^2\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu/f\{1+o_p(1)\}, \\ \mathcal{V}_{-,\nu,p}(h_n) &= n^{-1}h_n^{-2\nu}\nu!^2 e_\nu' \Gamma_{-,p}^{-1}(h_n)\Psi_{-,p}(h_n)\Gamma_{-,p}^{-1}(h_n)e_\nu \\ &= n^{-1}h_n^{-1-2\nu}\sigma_-^2\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu/f\{1+o_p(1)\}.\end{aligned}$$

(D) *If* $nh_n^{2p+5} \to 0$, *then* $(\hat{\tau}_{\nu,p}(h_n) - \tau_\nu - h_n^{p+1-\nu}\mathsf{B}_{\nu,p,p+1}(h_n))/\sqrt{\mathsf{V}_{\nu,p}(h_n)} \to_d \mathcal{N}(0,1)$.

A qth-order $(p < q)$ local polynomial bias-corrected estimator is $\hat{\tau}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n) = \hat{\tau}_p(h_n) - h_n^{p+1}\hat{\mathsf{B}}_{\nu,p,q}(h_n,b_n)$ with

$$\begin{aligned}\hat{\mathsf{B}}_{\nu,p,q}(h_n,b_n) &= \big(e_{p+1}'\hat{\beta}_{+,q}(b_n)\big)\mathcal{B}_{+,\nu,p,p+1}(h_n) \\ &\quad - \big(e_{p+1}'\hat{\beta}_{-,q}(b_n)\big)\mathcal{B}_{-,\nu,p,p+1}(h_n).\end{aligned}$$

The following theorem gives the asymptotic bias, variance, and distribution of $\hat{\tau}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)$. Theorems 1 and 2 are special cases with $(\nu,p,q) = (0,1,2)$ and $(\nu,p,q) = (1,2,3)$, respectively.

THEOREM A.1: *Suppose Assumptions 1–2 hold with* $S \geq q+1$, *and* $n\min\{h_n,b_n\} \to \infty$. *Let* $\nu \leq p < q$.
(B) *If* $\max\{h_n,b_n\} \to 0$, *then*

$$\begin{aligned}\mathbb{E}\big[\hat{\tau}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)|\mathcal{X}_n\big] &= \tau_\nu + h_n^{p+2-\nu}\mathsf{B}_{\nu,p,p+2}(h_n)\{1+o_p(1)\} \\ &\quad - h_n^{p+1-\nu}b_n^{q-p}\mathsf{B}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)\{1+o_p(1)\},\end{aligned}$$

where

$$\begin{aligned}\mathsf{B}^{\mathrm{bc}}_{\nu,p,q}(h,b) &= \big[\mu_+^{(q+1)}\mathcal{B}_{+,p+1,q,q+1}(b)\mathcal{B}_{+,\nu,p,p+1}(h) \\ &\quad - \mu_-^{(q+1)}\mathcal{B}_{-,p+1,q,q+1}(b)\mathcal{B}_{-,\nu,p,p+1}(h)\big] \\ &\quad /\big[(q+1)!(p+1)!\big].\end{aligned}$$

(V) $\mathsf{V}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n) = \mathbb{V}[\hat{\tau}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)|\mathcal{X}_n] = \mathcal{V}^{\mathrm{bc}}_{+,\nu,p,q}(h_n,b_n) + \mathcal{V}^{\mathrm{bc}}_{-,\nu,p,q}(h_n,b_n)$ *with*

$$\begin{aligned}\mathcal{V}^{\mathrm{bc}}_{+,\nu,p,q}(h,b) &= \mathcal{V}_{+,\nu,p}(h) \\ &\quad - 2h^{p+1-\nu}\mathcal{C}_{+,\nu,p,q}(h,b)\mathcal{B}_{+,\nu,p,p+1}(h)/(p+1)! \\ &\quad + h^{2(p+1-\nu)}\mathcal{V}_{+,p+1,q}(b)\mathcal{B}^2_{+,\nu,p,p+1}(h)/(p+1)!^2,\end{aligned}$$

$$\begin{aligned}\mathcal{V}^{\mathrm{bc}}_{-,\nu,p,q}(h,b) = \mathcal{V}_{-,\nu,p}(h) &- 2h^{p+1-\nu}\mathcal{C}_{-,\nu,p,q}(h,b)\mathcal{B}_{-,\nu,p,p+1}(h)/(p+1)! \\ &+ h^{2(p+1-\nu)}\mathcal{V}_{-,p+1,q}(b)\mathcal{B}^2_{-,\nu,p,p+1}(h)/(p+1)!^2,\\ \mathcal{C}_{+,\nu,p,q}(h,b) = n^{-1}h^{-\nu}b^{-p-1}\nu!(p+1)! &\times e'_\nu \Gamma^{-1}_{+,p}(h)\Psi_{+,p,q}(h,b)\Gamma^{-1}_{+,q}(b)e_{p+1},\\ \mathcal{C}_{-,\nu,p,q}(h,b) = n^{-1}h^{-\nu}b^{-p-1}\nu!(p+1)! &\times e'_\nu \Gamma^{-1}_{-,p}(h)\Psi_{-,p,q}(h,b)\Gamma^{-1}_{-,q}(b)e_{p+1}.\end{aligned}$$

(D) *If* $n\min\{h_n^{2p+3}, b_n^{2p+3}\}\max\{h_n^2, b_n^{2(q-p)}\} \to 0$, *and* $\kappa \max\{h_n, b_n\} < \kappa_0$, *then* $T^{\mathrm{rbc}}_{\nu,p,q}(h_n,b_n) = (\hat{\tau}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n) - \tau_\nu)/\sqrt{\mathsf{V}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)} \to_d \mathcal{N}(0,1)$.

In Theorem 1, $\mathsf{V}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n,b_n) = \mathsf{V}^{\mathrm{bc}}_{0,1,2}(h_n,b_n)$, $\mathsf{V}_{\mathrm{SRD}}(h_n) = \mathsf{V}_{0,1}(h_n) = \mathcal{V}_{+,0,1}(h_n) + \mathcal{V}_{-,0,1}(h_n)$, and $\mathsf{C}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n,b_n) = \mathsf{V}^{\mathrm{bc}}_{\mathrm{SRD}}(h_n,b_n) - \mathsf{V}_{\mathrm{SRD}}(h_n) = \mathsf{V}^{\mathrm{bc}}_{0,1,2}(h_n,b_n) - \mathsf{V}_{0,1}(h_n)$. In Theorem 2, $\mathsf{V}^{\mathrm{bc}}_{\mathrm{SKRD}}(h_n,b_n) = \mathsf{V}^{\mathrm{bc}}_{1,2,3}(h_n,b_n)$, $\mathsf{V}_{\mathrm{SKRD}}(h_n) = \mathsf{V}_{1,2}(h_n) = \mathcal{V}_{+,1,2}(h_n) + \mathcal{V}_{-,1,2}(h_n)$, and $\mathsf{C}^{\mathrm{bc}}_{\mathrm{SKRD}}(h_n,b_n) = \mathsf{V}^{\mathrm{bc}}_{\mathrm{SKRD}}(h_n,b_n) - \mathsf{V}_{\mathrm{SKRD}}(h_n) = \mathsf{V}^{\mathrm{bc}}_{1,2,3}(h_n,b_n) - \mathsf{V}_{1,2}(h_n)$.

A.3. *Fuzzy RD Designs*

The νth fuzzy RD estimand is $\varsigma_\nu = \tau_{Y,\nu}/\tau_{T,\nu}$ with $\tau_{Y,\nu} = \mu^{(\nu)}_{Y+} - \mu^{(\nu)}_{Y-}$ and $\tau_{T,\nu} = \mu^{(\nu)}_{T+} - \mu^{(\nu)}_{T-}$, provided that $\nu \le S$. Note that $\tau_{\mathrm{FRD}} = \varsigma_0$ and $\tau_{\mathrm{FKRD}} = \varsigma_1$. The fuzzy RD estimator based on the pth-order local polynomial estimators $\hat{\tau}_{Y,\nu,p}(h_n)$ and $\hat{\tau}_{T,\nu,p}(h_n)$ is $\hat{\varsigma}_{\nu,p}(h_n) = \hat{\tau}_{Y,\nu,p}(h_n)/\hat{\tau}_{T,\nu,p}(h_n)$ with $\hat{\tau}_{Y,\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{Y+,p}(h_n) - \hat{\mu}^{(\nu)}_{Y-,p}(h_n)$ and $\hat{\tau}_{T,\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{T+,p}(h_n) - \hat{\mu}^{(\nu)}_{T-,p}(h_n)$, where $\hat{\mu}^{(\nu)}_{Y+,p}(h_n) = \nu! \times e'_\nu\hat{\beta}_{Y+,p}(h_n)$, $\hat{\mu}^{(\nu)}_{Y-,p}(h_n) = \nu! e'_\nu\hat{\beta}_{Y-,p}(h_n)$, $\hat{\mu}^{(\nu)}_{T+,p}(h_n) = \nu! e'_\nu\hat{\beta}_{T+,p}(h_n)$, and $\hat{\mu}^{(\nu)}_{T-,p}(h_n) = \nu! e'_\nu\hat{\beta}_{T-,p}(h_n)$. Note that $\hat{\tau}_{\mathrm{FRD}}(h_n) = \hat{\varsigma}_{0,1}(h_n)$ and $\hat{\tau}_{\mathrm{FKRD}}(h_n) = \hat{\varsigma}_{1,2}(h_n)$.

The following lemma gives an analogue of Lemma A.1 for fuzzy RD designs. Note that $\hat{\varsigma}_{\nu,p}(h_n) - \varsigma_\nu = \tilde{\varsigma}_{\nu,p}(h_n) + R_n$ with $\tilde{\varsigma}_{\nu,p}(h_n) = (\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu})/\tau_{T,\nu} - \tau_{Y,\nu}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})/\tau^2_{T,\nu}$ and $R_n = \tau_{Y,\nu}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})^2/(\tau^2_{T,\nu}\hat{\tau}_{T,\nu,p}(h_n)) - (\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu})(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})/(\tau_{T,\nu}\hat{\tau}_{T,\nu,p}(h_n))$.

LEMMA A.2: *Suppose Assumptions 1–3 hold with* $S \ge p+2$, *and* $nh_n \to \infty$. *Let* $r \in \mathbb{N}$ *and* $\nu \le p$.

(R) *If* $h_n \to 0$ *and* $nh_n^{1+2\nu} \to \infty$, *then* $R_n = O_p(n^{-1}h_n^{-1-2\nu} + h_n^{2p+2-2\nu})$.

(B) *If* $h_n \to 0$, *then*

$$\mathbb{E}\big[\tilde{\varsigma}_{\nu,p}(h_n)|\mathcal{X}_n\big] = h_n^{p+1-\nu}\mathsf{B}_{\mathrm{F},\nu,p,p+1}(h_n) + h_n^{p+2-\nu}\mathsf{B}_{\mathrm{F},\nu,p,p+2}(h_n) + o_p\big(h_n^{p+2-\nu}\big),$$

where

$$\mathsf{B}_{\mathrm{F},\nu,p,r}(h_n) = \mathsf{B}_{Y,\nu,p,r}(h_n)/\tau_{T,\nu} - \tau_{Y,\nu}\mathsf{B}_{T,\nu,p,r}(h_n)/\tau^2_{T,\nu},$$
$$\mathsf{B}_{Y,\nu,p,r}(h_n) = \mu^{(r)}_{Y+}\mathcal{B}_{+,\nu,p,r}(h_n)/r! - \mu^{(r)}_{Y-}\mathcal{B}_{-,\nu,p,r}(h_n)/r!,$$
$$\mathsf{B}_{T,\nu,p,r}(h_n) = \mu^{(r)}_{T+}\mathcal{B}_{+,\nu,p,r}(h_n)/r! - \mu^{(r)}_{T-}\mathcal{B}_{-,\nu,p,r}(h_n)/r!.$$

(V) *If* $h_n \to 0$, *then* $\mathsf{V}_{\mathrm{F},\nu,p}(h_n) = \mathbb{V}[\tilde{\varsigma}_{\nu,p}(h_n)|\mathcal{X}_n] = \mathsf{V}_{\mathrm{F},+,\nu,p}(h_n) + \mathsf{V}_{\mathrm{F},-,\nu,p}(h_n)$ *with*

$$\begin{aligned}\mathsf{V}_{\mathrm{F},+,\nu,p}(h_n) &= \big(1/\tau^2_{T,\nu}\big)\mathcal{V}_{YY+,\nu,p}(h_n) - \big(2\tau_{Y,\nu}/\tau^3_{T,\nu}\big)\mathcal{V}_{YT+,\nu,p}(h_n)\\ &\quad + \big(\tau^2_{Y,\nu}/\tau^4_{T,\nu}\big)\mathcal{V}_{TT+,\nu,p}(h_n),\\ \mathsf{V}_{\mathrm{F},-,\nu,p}(h_n) &= \big(1/\tau^2_{T,\nu}\big)\mathcal{V}_{YY-,\nu,p}(h_n) - \big(2\tau_{Y,\nu}/\tau^3_{T,\nu}\big)\mathcal{V}_{YT-,\nu,p}(h_n)\\ &\quad + \big(\tau^2_{Y,\nu}/\tau^4_{T,\nu}\big)\mathcal{V}_{TT-,\nu,p}(h_n),\end{aligned}$$

where, for $U = Y, T$ *and* $V = Y, T$,

$$\begin{aligned}\mathcal{V}_{UV+,\nu,p}(h_n) &= n^{-1}h_n^{-2\nu}\nu!^2 e'_\nu \Gamma^{-1}_{+,p}(h_n)\Psi_{UV+,p}(h_n)\Gamma^{-1}_{+,p}(h_n)e_\nu\\ &= n^{-1}h_n^{-1-2\nu}\sigma^2_{UV+}\nu!^2 e'_\nu\Gamma^{-1}_p\Psi_p\Gamma^{-1}_p e_\nu/f\{1+o_p(1)\},\\ \mathcal{V}_{UV-,\nu,p}(h_n) &= n^{-1}h_n^{-2\nu}\nu!^2 e'_\nu \Gamma^{-1}_{-,p}(h_n)\Psi_{UV-,p}(h_n)\Gamma^{-1}_{-,p}(h_n)e_\nu\\ &= n^{-1}h_n^{-1-2\nu}\sigma^2_{UV-}\nu!^2 e'_\nu\Gamma^{-1}_p\Psi_p\Gamma^{-1}_p e_\nu/f\{1+o_p(1)\}.\end{aligned}$$

(D) *If* $nh_n^{2p+5}\to 0$ *and* $nh_n^{1+2\nu}\to\infty$, *then* $(\hat{\varsigma}_{\nu,p}(h_n) - \varsigma_\nu - h_n^{p+1-\nu}\mathsf{B}_{\mathrm{F},\nu,p,p+1}(h_n))/\sqrt{\mathsf{V}_{\mathrm{F},\nu,p}(h_n)} \to_d \mathcal{N}(0,1)$.

The following theorem gives an analogue of Theorem A.1 for fuzzy RD designs; Theorems 3 and 4 are special cases with $(\nu, p, q) = (0,1,2)$ and $(\nu, p, q) = (1,2,3)$, respectively. This theorem summarizes the asymptotic bias, variance, and distribution of the bias-corrected fuzzy RD estimator:

$$\begin{aligned}\hat{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n) &= \hat{\varsigma}_{\nu,p}(h_n) - h_n^{p+1-\nu}\hat{\mathsf{B}}_{\mathrm{F},\nu,p,q}(h_n,b_n),\\ \hat{\mathsf{B}}_{\mathrm{F},\nu,p,q}(h_n,b_n) &= \big[\big(e'_{p+1}\hat{\beta}_{Y+,q}(b_n)\big)\mathcal{B}_{+,\nu,p,p+1}(h_n)\\ &\qquad - \big(e'_{p+1}\hat{\beta}_{Y-,q}(b_n)\big)\mathcal{B}_{-,\nu,p,p+1}(h_n)\big]/\hat{\tau}_{T,\nu,p}(h_n)\\ &\qquad - \hat{\tau}_{Y,\nu,p}(h_n)\big[\big(e'_{p+1}\hat{\beta}_{T+,q}(b_n)\big)\mathcal{B}_{+,\nu,p,p+1}(h_n)\\ &\qquad - \big(e'_{p+1}\hat{\beta}_{T-,q}(b_n)\big)\mathcal{B}_{-,\nu,p,p+1}(h_n)\big]/\hat{\tau}_{T,\nu,p}(h_n)^2.\end{aligned}$$

Linearizing the estimator, we obtain

$$\begin{aligned}
\hat{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)-\varsigma_\nu &= \tilde{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)+R_n-R^{\mathrm{bc}}_n,\\
\tilde{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n) &= \big(\hat{\tau}^{\mathrm{bc}}_{Y,\nu,p,q}(h_n,b_n)-\tau_{Y,\nu}\big)/\tau_{T,\nu}\\
&\quad -\tau_{Y,\nu}\big(\hat{\tau}^{\mathrm{bc}}_{T,\nu,p,q}(h_n,b_n)-\tau_{T,\nu}\big)/\tau^2_{T,\nu},\\
R_n &= \tau_{Y,\nu}\big(\hat{\tau}_{T,\nu,p}(h_n)-\tau_{T,\nu}\big)^2/\big(\tau^2_{T,\nu}\hat{\tau}_{T,\nu,p}(h_n)\big)\\
&\quad -\big(\hat{\tau}_{Y,\nu,p}(h_n)-\tau_{Y,\nu}\big)\big(\hat{\tau}_{T,\nu,p}(h_n)-\tau_{T,\nu}\big)/\big(\tau_{T,\nu}\hat{\tau}_{T,\nu,p}(h_n)\big),\\
R^{\mathrm{bc}}_n &= h_n^{p+1-\nu}\big(\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n)-\check{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n)\big),\\
\check{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) &= \big[\big(e'_{p+1}\hat{\beta}_{Y+,q}(b_n)\big)\mathcal{B}_{+,\nu,p,p+1}(h_n)\\
&\quad -\big(e'_{p+1}\hat{\beta}_{Y-,q}(b_n)\big)\mathcal{B}_{-,\nu,p,p+1}(h_n)\big]/\tau_{T,\nu}\\
&\quad -\tau_{Y,\nu}\big[\big(e'_{p+1}\hat{\beta}_{T+,q}(b_n)\big)\mathcal{B}_{+,\nu,p,p+1}(h_n)\\
&\quad -\big(e'_{p+1}\hat{\beta}_{T-,q}(b_n)\big)\mathcal{B}_{-,\nu,p,p+1}(h_n)\big]/\tau^2_{T,\nu}.
\end{aligned}$$

THEOREM A.2: *Suppose Assumptions 1–3 hold with $S \ge p+2$, and $n\min\{h_n,b_n\}\to\infty$. Let $\nu \le p < q$.*

(R^{bc}) *If $h_n \to 0$ and $nh_n^{1+2\nu}\to\infty$, and provided that $\kappa b_n < \kappa_0$, then $R^{\mathrm{bc}}_n = O_p(n^{-1/2}h_n^{p+1/2}+h_n^{2p+2-2\nu})O_p(1+n^{-1/2}b_n^{-3/2-p})$.*

(B) *If $\max\{h_n,b_n\}\to 0$, then*

$$\begin{aligned}
\mathbb{E}\big[\tilde{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)|\mathcal{X}_n\big] &= h_n^{p+2-\nu}\mathsf{B}_{\mathtt{F},\nu,p,p+2}(h_n)\big\{1+o_p(1)\big\}\\
&\quad + h_n^{p+1-\nu}b_n^{q-p}\mathsf{B}^{\mathrm{bc}}_{\mathtt{F},\nu,p,q}(h_n,b_n)\big\{1+o_p(1)\big\},
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{B}^{\mathrm{bc}}_{\mathtt{F},\nu,p,q}(h,b) &= \mathsf{B}^{\mathrm{bc}}_{Y,\nu,p,q}(h_n,b_n)/\tau_{T,\nu}-\tau_{Y,\nu}\mathsf{B}^{\mathrm{bc}}_{T,\nu,p,q}(h_n,b_n)/\tau^2_{T,\nu},\\
\mathsf{B}^{\mathrm{bc}}_{Y,\nu,p,q}(h,b) &= \big[\mu^{(q+1)}_{Y+}\mathcal{B}_{+,p+1,q,q+1}(b)\mathcal{B}_{+,\nu,p,p+1}(h)\\
&\quad -\mu^{(q+1)}_{Y-}\mathcal{B}_{-,p+1,q,q+1}(b)\mathcal{B}_{-,\nu,p,p+1}(h)\big]\\
&\quad /\big[(q+1)!(p+1)!\big],\\
\mathsf{B}^{\mathrm{bc}}_{T,\nu,p,q}(h,b) &= \big[\mu^{(q+1)}_{T+}\mathcal{B}_{+,p+1,q,q+1}(b)\mathcal{B}_{+,\nu,p,p+1}(h)\\
&\quad -\mu^{(q+1)}_{T-}\mathcal{B}_{-,p+1,q,q+1}(b)\mathcal{B}_{-,\nu,p,p+1}(h)\big]\\
&\quad /\big[(q+1)!(p+1)!\big].
\end{aligned}$$

(V) $\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},\nu,p,q}(h_n,b_n)=\mathbb{V}[\tilde{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)|\mathcal{X}_n]=\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},+,\nu,p,q}(h_n,b_n)+\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},-,\nu,p,q}(h_n,b_n)$ *with*

$$\begin{aligned}
\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},+,\nu,p,q}(h,b) &= \mathsf{V}_{\mathrm{F},+,\nu,p}(h) \\
&\quad - 2h^{p+1-\nu}\mathcal{C}_{\mathrm{F},+,\nu,p,q}(h,b)\mathcal{B}_{+,\nu,p,p+1}(h)/(p+1)! \\
&\quad + h^{2p+2-2\nu}\mathsf{V}_{\mathrm{F},+,p+1,q}(b)\mathcal{B}^2_{+,\nu,p,p+1}(h)/(p+1)!^2, \\
\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},-,\nu,p,q}(h,b) &= \mathsf{V}_{\mathrm{F},-,\nu,p}(h) \\
&\quad - 2h^{p+1-\nu}\mathcal{C}_{\mathrm{F},-,\nu,p,q}(h,b)\mathcal{B}_{-,\nu,p,p+1}(h)/(p+1)! \\
&\quad + h^{2p+2-2\nu}\mathsf{V}_{\mathrm{F},-,p+1,q}(b)\mathcal{B}^2_{-,\nu,p,p+1}(h)/(p+1)!^2, \\
\mathcal{C}_{\mathrm{F},+,\nu,p,q}(h,b) &= \left(1/\tau^2_{T,\nu}\right)\mathcal{C}_{YY+,\nu,p,q}(h,b) - \left(2\tau_{Y,\nu}/\tau^3_{T,\nu}\right)\mathcal{C}_{YT+,\nu,p,q}(h,b) \\
&\quad + \left(\tau^2_{Y,\nu}/\tau^4_{T,\nu}\right)\mathcal{C}_{TT+,\nu,p,q}(h,b), \\
\mathcal{C}_{\mathrm{F},-,\nu,p,q}(h,b) &= \left(1/\tau^2_{T,\nu}\right)\mathcal{C}_{YY-,\nu,p,q}(h,b) - \left(2\tau_{Y,\nu}/\tau^3_{T,\nu}\right)\mathcal{C}_{YT-,\nu,p,q}(h,b) \\
&\quad + \left(\tau^2_{Y,\nu}/\tau^4_{T,\nu}\right)\mathcal{C}_{TT-,\nu,p,q}(h,b),
\end{aligned}$$

where, for $U=Y,T$ and $V=Y,T$,

$$\begin{aligned}
\mathcal{C}_{UV+,\nu,p,q}(h,b) &= n^{-1}h^{-\nu}b^{-p-1}\nu!(p+1)! \\
&\quad \times e'_\nu \Gamma^{-1}_{+,p}(h)\Psi_{UV+,p,q}(h,b)\Gamma^{-1}_{+,q}(b)e_{p+1}, \\
\mathcal{C}_{UV-,\nu,p,q}(h,b) &= n^{-1}h^{-\nu}b^{-p-1}\nu!(p+1)! \\
&\quad \times e'_\nu \Gamma^{-1}_{-,p}(h)\Psi_{UV-,p,q}(h,b)\Gamma^{-1}_{-,q}(b)e_{p+1}.
\end{aligned}$$

(D) *If* $n\min\{h_n^{2p+3},b_n^{2p+3}\}\max\{h_n^2,b_n^{2(q-p)}\}\to 0$ *and* $n\min\{h_n^{1+2\nu},b_n\}\to\infty$, *and* $h_n\to 0$ *and* $\kappa b_n<\kappa_0$, *then* $T^{\mathrm{rbc}}_{\mathrm{F},\nu,p,q}(h_n,b_n)=(\hat{\varsigma}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)-\varsigma_\nu)/\sqrt{\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},\nu,p,q}(h_n,b_n)}\to_d \mathcal{N}(0,1)$.

In Theorem 3, $\mathsf{V}^{\mathrm{bc}}_{\mathrm{FRD}}(h_n,b_n)=\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},0,1,2}(h_n,b_n)$, $\mathsf{V}_{\mathrm{FRD}}(h_n)=\mathsf{V}_{\mathrm{F},0,1}(h_n)=\mathcal{V}_{\mathrm{F},+,0,1}(h_n)+\mathcal{V}_{\mathrm{F},-,0,1}(h_n)$, and $\mathsf{C}^{\mathrm{bc}}_{\mathrm{FRD}}(h_n,b_n)=\mathsf{V}^{\mathrm{bc}}_{\mathrm{FRD}}(h_n,b_n)-\mathsf{V}_{\mathrm{FRD}}(h_n)=\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},0,1,2}(h_n,b_n)-\mathsf{V}_{\mathrm{F},0,1}(h_n)$. In Theorem 4, $\mathsf{V}^{\mathrm{bc}}_{\mathrm{FKRD}}(h_n,b_n)=\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},1,2,3}(h_n,b_n)$, $\mathsf{V}_{\mathrm{FKRD}}(h_n)=\mathsf{V}_{\mathrm{F},1,2}(h_n)=\mathcal{V}_{\mathrm{F},+,1,2}(h_n)+\mathcal{V}_{\mathrm{F},-,1,2}(h_n)$, and $\mathsf{C}^{\mathrm{bc}}_{\mathrm{FKRD}}(h_n,b_n)=\mathsf{V}^{\mathrm{bc}}_{\mathrm{FKRD}}(h_n,b_n)-\mathsf{V}_{\mathrm{FKRD}}(h_n)=\mathsf{V}^{\mathrm{bc}}_{\mathrm{F},1,2,3}(h_n,b_n)-\mathsf{V}_{\mathrm{F},1,2}(h_n)$.

REFERENCES

ABADIE, A., AND G. W. IMBENS (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74 (1), 235–267. [2313]

——— (2010): "Estimation of the Conditional Variance in Paired Experiments," *Annales d'Économie et de Statistique*, 91 (1), 175–187. [2313]

CALONICO, S., M. D. CATTANEO, AND M. H. FARRELL (2014): "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Estimation," Working Paper, University of Michigan. [2298,2304,2305,2311]

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014a): "Optimal Data-Driven Regression Discontinuity Plots," Working Paper, University of Michigan. [2297]

——— (2014b): "Robust Data-Driven Inference in the Regression-Discontinuity Design," *Stata Journal* (forthcoming). [2295,2298]

——— (2014c): "Supplement to 'Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs'," *Econometrica Supplemental Material*, 82, http://dx.doi.org/10.3982/ECTA11757. [2298,2309-2311,2313-2315,2318]

——— (2014d): "`rdrobust`: An R Package for Robust Inference in Regression-Discontinuity Designs," Working Paper, University of Michigan. [2295,2298]

CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2014): "Inference on Causal Effects in a Generalized Regression Kink Design," Working Paper, UC-Berkeley. [2297,2305]

CATTANEO, M. D., B. FRANDSEN, AND R. TITIUNIK (2014): "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference* (forthcoming). [2297,2299]

CHENG, M.-Y., J. FAN, AND J. S. MARRON (1997): "On Automatic Boundary Corrections," *The Annals of Statistics*, 25 (4), 1691–1708. [2299]

DINARDO, J., AND D. S. LEE (2011): "Program Evaluation and Research Designs," in *Handbook of Labor Economics*, Vol. 4A, ed. by O. Ashenfelter and D. Card. New York: Elsevier Science B.V., 463–536. [2296]

DONG, Y. (2014): "Jumpy or Kinky? Regression Discontinuity Without the Discontinuity," Working Paper, UC-Irvine. [2297,2305]

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC. [2297-2299]

FRANDSEN, B., M. FRÖLICH, AND B. MELLY (2012): "Quantile Treatments Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168 (2), 382–395. [2297]

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69 (1), 201–209. [2297, 2298,2300]

HOROWITZ, J. L. (2001): "The Bootstrap," in *Handbook of Econometrics*, Vol. V, ed. by J. J. Heckman and E. Leamer. New York: Elsevier Science B.V., 3159–3228. [2304]

ICHIMURA, H., AND P. E. TODD (2007): "Implementing Nonparametric and Semiparametric Estimators," in *Handbook of Econometrics*, Vol. 6B, ed. by J. J. Heckman and E. E. Leamer. New York: Elsevier Science B.V., 5369–5468. [2298,2301]

IMBENS, G., AND T. LEMIEUX (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142 (2), 615–635. [2296]

IMBENS, G. W., AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79 (3), 933–959. [2297,2301,2309,2311, 2314,2315]

KEELE, L. J., AND R. TITIUNIK (2014): "Geographic Boundaries as Regression Discontinuities," *Political Analysis* (forthcoming). [2298]

LEE, D. S. (2008): "Randomized Experiments From Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142 (2), 675–697. [2297,2314]

LEE, D. S., AND D. CARD (2008): "Regression Discontinuity Inference With Specification Error," *Journal of Econometrics*, 142 (2), 655–674. [2297]

LEE, D. S., AND T. LEMIEUX (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48 (2), 281–355. [2296]

LI, K.-C. (1987): "Asymptotic Optimality for C_p, C_L, Cross-Validation and Generalized Cross-Validation: Discrete Index Set," *The Annals of Statistics*, 15 (3), 958–975. [2311]

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 (1), 159–208. [2314]

MARMER, V., D. FEIR, AND T. LEMIEUX (2014): "Weak Identification in Fuzzy Regression Discontinuity Designs," Working Paper, University of British Columbia. [2297]

MCCRARY, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142 (2), 698–714. [2297]

OTSU, T., K.-L. XU, AND Y. MATSUSHITA (2014): "Empirical Likelihood for Regression Discontinuity Design," *Journal of Econometrics* (forthcoming). [2297]

PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," Working Paper, University of Wisconsin. [2297,2300]

THISTLETHWAITE, D. L., AND D. T. CAMPBELL (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment," *Journal of Educational Psychology*, 51 (6), 309–317. [2296]

VAN DER KLAAUW, W. (2008): "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics," *Labour*, 22 (2), 219–245. [2296]

WAND, M. P., AND M. C. JONES (1995): *Kernel Smoothing*. Boca Raton, FL: Chapman & Hall/CRC. [2311]

Dept. of Economics, University of Miami, 5250 University Dr., Coral Gables, FL 33124, U.S.A.; scalonico@bus.miami.edu,

Dept. of Economics, University of Michigan, 611 Tappan Ave., Ann Arbor, MI 48109, U.S.A.; cattaneo@umich.edu,

and

Dept. of Political Science, University of Michigan, 505 S. State St., Ann Arbor, MI 48109, U.S.A.; titiunik@umich.edu.

Manuscript received July, 2013; final revision received June, 2014.