Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs*

Sebastian Calonico[†] Matias D. Cattaneo [‡] Rocio Titiunik[§]

July 15, 2013

Abstract

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 severely biased, with empirical coverage well below their nominal target. We propose new, more robust, theory-based confidence interval estimators for average treatment effects in sharp RD, kink RD, fuzzy RD and fuzzy kink RD designs. Our proposed confidence intervals rely on a recentered RD estimator together with a novel standard-error estimator. For practical implementation, we propose a consistent standard-error estimator that does not require an additional bandwidth choice, as well as valid bandwidth choices compatible with our underlying large-sample theory. In a simulation study, we find that our novel data-driven confidence intervals exhibit close-to-correct empirical coverage and good empirical interval length on average, remarkably improving upon the alternatives available in the literature. We illustrate the performance of our proposed methods with household data from Progresa/Oportunidades, a conditional cash transfer program in Mexico. All the results in this paper are readily available in STATA using our companion package (`rdrobust`) described in Calonico, Cattaneo, and Titiunik (2013).

Keywords: regression discontinuity, local polynomials, bias correction, robust inference, alternative asymptotics.

JEL: C12, C14, C21.

---

*This paper is a revised and extended version of our paper previously entitled "Robust Nonparametric Inference in the Regression Discontinuity Design" (first draft: November 2011). We are specially thankful to Elie Tamer for encouraging us to write this new version. For helpful comments and suggestions we also thank Manuela Angelucci, Alberto Abadie, Josh Angrist, Ivan Canay, Victor Chernozhukov, Patrik Guggenberger, Hide Ichimura, Guido Imbens, Justin McCrary, Michael Jansson, Anna Mikusheva, Whitney Newey, Andres Santos, Jeff Smith, Jeff Wooldridge, as well as seminar and conference participants at many institutions. Finally, we also thank Luis Alejos Marroquin, Habiba Djebbari, Paul Gertler, Dan Gilligan, Jens Ludwig and Doug Miller for sharing, and helping us with, different datasets used in this paper.

[†]Department of Economics, University of Michigan.

[‡]Department of Economics, University of Michigan.

[§]Department of Political Science and Center for Political Studies, University of Michigan.

1 Introduction

The regression discontinuity (RD) design has become one of the leading non-experimental empirical strategies in Economics, Political Science and many other social and behavioral sciences.[1] In this design, units are assigned to treatment based on their value of an observed covariate, 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 certain 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. Flexible estimation of RD treatment effects approximates the outcome's (and treatment status') regression function given the score near the cutoff for control and treated groups separately, and computes the estimated effect as the difference (or ratio of differences) of the appropriate 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 average treatment effects. This estimation strategy involves approximating the regression functions above and below the cutoff by means of weighted local polynomial regressions, typically of order one, 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. As a consequence, the resulting data-driven confidence intervals for RD treatment effects may be biased, having empirical coverage well below their nominal target. This implies that, for example, these conventional confidence intervals may substantially over-reject the null hypothesis of no treatment effect in empirical applications.

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-square-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 in the literature, however, conventional bias-correction alone delivers very poor finite-sample performance because it relies on a low-quality distributional approximation. Thus, to improve the quality of the distributional approximation of the bias-corrected t-statistic, we also introduce a novel standard-error formula to account for the additional variability introduced by the estimated bias; that is, we rescale the bias-corrected t-statistic. The new

[1] See, among others, van der Klaauw (2008), Imbens and Lemieux (2008), Lee and Lemieux (2010) and Dinardo and Lee (2011) for recent reviews.

standard-error formula is theoretically justified by a non-standard 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 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; e.g., see Remark 2 below), but also continue to offer correct coverage rates in large samples even when the conventional confidence intervals do not (e.g., see Remark 1 below).

The main discussion focuses on the construction of robust confidence intervals for the RD average treatment effect at the cutoff in four settings: sharp RD, kink RD, fuzzy RD and fuzzy kink RD designs. 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 ones. The new confidence intervals are demonstrably more robust because they are valid under strictly weaker bandwidth conditions than those required by their conventional counterparts. In addition, we find in an empirically motivated simulation study that our proposed data-driven confidence intervals exhibit close-to-correct empirical coverage and good empirical interval length on average, remarkably improving upon the alternatives available in the literature. We also illustrate the performance of our proposed confidence intervals, as well as several of the conventional alternatives, in an empirical application that studies the effects of Progresa/Oportunidades, a large-scale anti-poverty conditional cash transfer program in Mexico, on households' consumption outcomes. Our illustration shows that in some, but not all, cases the conclusions drawn from conventional methods are not supported when our robust inference procedures are employed.

Our paper contributes to the emerging literature on inference for treatment effects in the RD design. Hahn, Todd, and van der Klaauw (2001) and Lee (2008) develop identification results, Porter (2003) gives optimality results of local polynomial estimators, McCrary (2008) studies specification testing, Imbens and Kalyanaraman (2012) develop bandwidth selection procedures for local-linear estimators, Otsu and Xu (2011) study empirical likelihood methods applied to local-linear estimators, Frandsen, Frölich, and Melly (2012) consider quantile treatment effects, Card, Lee, Pei, and Weber (2012), Dong (2012) and Dong and Lewel (2012) study the so-called kink RD designs, Marmer, Feir, and Lemieux (2012) discuss robust to weak-IV inference in fuzzy RD designs, and Cattaneo, Frandsen, and Titiunik (2013) propose randomization-inference methods.[2]

The rest of the paper is organized as follows. Section 2 describes the basic sharp RD design, reviews conventional results, provides simulation evidence to motivate our approach, and outlines the details of our proposed robust confidence intervals. Section 3 discusses extensions of the approach to kink RD, fuzzy RD and fuzzy kink RD designs. Mean-square-error optimal bandwidths and their theoretical validity when using our approach is discussed in Section 4 (e.g., see Remark

[2]From a more general perspective, our results also contribute to the literature on asymptotic approximations for nonparametric local polynomial estimators (Fan and Gijbels (1996)), which are useful in econometrics (see, e.g., Ichimura and Todd (2007) and references therein).

6 below), while valid standard-errors are proposed in Section 5. Results from the simulation study and empirical illustration are given in Sections 6 and 7, respectively, while Section 8 concludes. Our general theoretical results are summarized in the Appendix, but most technical details such as proofs and other derivations are relegated to the online supplemental appendix. Our new confidence intervals together with all other results presented in this paper are implemented in the companion STATA package (`rdrobust`) described in Calonico, Cattaneo, and Titiunik (2013).[3]

2 Sharp RD Design: Local-Linear Estimator, Its Potential Pitfalls and The New Robust Alternative

We first focus attention on constructing confidence intervals for the average treatment effect in the sharp RD framework where the probability of treatment assignment changes from zero to one at the threshold, but Section 3 discusses the extension of our main approach to other settings of empirical interest.

In the canonical sharp RD design, we assume that $(Y_i(0), Y_i(1), X_i)'$, $i = 1, 2, \ldots, n$, is a random sample from the triplet of random variables $(Y(0), Y(1), X)'$ with $f(x)$ the Lebesgue density of X_i. Given a known threshold $\bar{x}$, which we 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 outcome with and without treatment, respectively. As a consequence, the observed random sample is $\{(Y_i, X_i)' : i = 1, 2, \ldots, n\}$ with

$$Y_i = Y_i(0) \cdot \mathbf{1}(X_i < 0) + Y_i(1) \cdot \mathbf{1}(X_i \geq 0) = \begin{cases} Y_i(0) & \text{if } X_i < 0 \\ Y_i(1) & \text{if } X_i \geq 0 \end{cases},$$

where $\mathbf{1}(\cdot)$ denotes the indicator function.

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

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

where here, and elsewhere, we drop the evaluation point of functions whenever possible to simplify notation. Estimation in RD designs naturally focuses on the flexible approximation of the regression functions $\mu_-(x) = \mathbb{E}[Y_i(0)|X_i = x]$ and $\mu_+(x) = \mathbb{E}[Y_i(1)|X_i = x]$ near the cutoff $\bar{x} = 0$. Section A.1 in the Appendix describes the conventional assumptions on the basic RD model employed in this paper. In particular, near the cutoff, for all $x \in [-\kappa_0, \kappa_0]$ with $\kappa_0 > 0$, we assume continuity of $f(x)$ (which rules out discrete-valued running variables; see, e.g., Lee and Card (2008)) and smoothness

[3] Computer code in `R` is also available upon request.

of the underlying regression functions $\mu_-(x)$ and $\mu_+(x)$. Since the higher-order derivatives of these unknown regression functions are closely related to the bias of the RD estimators considered, we also introduce notation to describe these derivatives at either side of the threshold:

$$\mu_+^{(\nu)} = \lim_{x\to 0^+} \frac{\mathrm{d}^\nu}{\mathrm{d}x^\nu}\mu_+(x) \qquad \text{and} \qquad \mu_-^{(\nu)} = \lim_{x\to 0^-} \frac{\mathrm{d}^\nu}{\mathrm{d}x^\nu}\mu_-(x).$$

(By definition, $\mu_+ = \mu_+^{(0)}$ and $\mu_- = \mu_-^{(0)}$.)

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

$$\hat{\tau}_{\mathtt{SRD}}(h_n) = \hat{\mu}_{+,1}(h_n) - \hat{\mu}_{-,1}(h_n),$$

$$(\hat{\mu}_{+,1}(h_n), \hat{\mu}_{+,1}^{(1)}(h_n))' = \arg\min_{b_0,b_1\in\mathbb{R}} \sum_{i=1}^{n} \mathbf{1}(X_i \geq 0)(Y_i - b_0 - X_i b_1)^2 \; k\left(\frac{X_i}{h_n}\right),$$

$$(\hat{\mu}_{-,1}(h_n), \hat{\mu}_{-,1}^{(1)}(h_n))' = \arg\min_{b_0,b_1\in\mathbb{R}} \sum_{i=1}^{n} \mathbf{1}(X_i < 0)(Y_i - b_0 - X_i b_1)^2 \; k\left(\frac{-X_i}{h_n}\right),$$

where h_n denotes the bandwidth chosen and $k(x)$ denotes the kernel function supported on $[0, \kappa]$ for some $\kappa > 0$. Indeed, we will employ local polynomial regression estimators of various orders to approximate unknown regression functions throughout the paper, as these estimators are particularly well-suited for inference in the RD design (Fan and Gijbels (1996) and Cheng, Fan, and Marron (1997)). Section A.2 in the Appendix describes these estimators in detail. Our results cover all commonly used kernels, including the triangular kernel $k(u) = (1-u)\mathbf{1}(0 \leq u \leq 1)$ and uniform kernel $k(u) = \mathbf{1}(0 \leq u \leq 1)$.

In the sharp RD design, the local-linear estimator $\hat{\tau}_{\mathtt{SRD}}(h_n)$ is arguably the preferred and most common choice in practice. Conventional approaches to constructing confidence intervals for $\tau_{\mathtt{SRD}}$ using this estimator rely on the following large-sample approximation for the standardized t-statistic (see Lemma A1 in the Appendix for the general result):

Lemma 1. Suppose Assumptions A1–A2 hold with $S \geq 3$. If $nh_n^5 \to 0$ and $nh_n \to \infty$, then

$$T_{\mathtt{SRD}}(h_n) = \frac{\hat{\tau}_{\mathtt{SRD}}(h_n) - \tau_{\mathtt{SRD}}}{\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}} \to_d \mathcal{N}(0,1), \qquad \mathsf{V}_{\mathtt{SRD}}(h_n) = \mathbb{V}[\hat{\tau}_{\mathtt{SRD}}(h_n) | X_1, X_2, \cdots, X_n].$$

Conventional (infeasible) $100(1-\alpha)$-percent confidence intervals for $\tau_{\mathtt{SRD}}$ are theoretically justified from this result, and take the familiar form

$$I_{\mathtt{SRD}}(h_n) = \left[\hat{\tau}_{\mathtt{SRD}}(h_n) \pm \Phi_{1-\alpha/2}^{-1}\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}\right],$$

where Φ_α^{-1} is the upper α-quantile of the standard normal distribution (e.g., $\Phi_{0.95}^{-1} \approx 1.96$). In practice, of course, a standard-error estimator is needed to construct feasible confidence intervals

Figure 1: Regression Functions for Models 1–3 in simulations.

because $\mathsf{V}_{\mathtt{SRD}}(h_n)$ involves unknown quantities, but here we assume $\mathsf{V}_{\mathtt{SRD}}(h_n)$ is known and postpone the issue of standard-errors until Section 5.[4] Even in this simplified 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 A1 in the Appendix. Thus, in general, the confidence intervals $I_{\mathtt{SRD}}(h_n)$ will have correct asymptotic coverage only if the bandwidth h_n is chosen to be "small" enough so that the bias-condition $nh_n^5 \to 0$ is satisfied.

Several approaches are available in the literature to select h_n, including plug-in rules and cross-validation procedures. Imbens and Kalyanaraman (2012) give a recent account of the state-of-the-art in bandwidth selection for RD designs. Unfortunately, most (if not all) of these approaches lead to bandwidths that are too "large" because they do not satisfy the bias-condition just described. For example, minimizing the asymptotic mean squared error (MSE) of $\hat{\tau}_{\mathtt{SRD}}(h_n)$ gives the optimal plug-in bandwidth choice $h_{\mathtt{MSE}} = C_{\mathtt{MSE}}\ n^{-1/5}$ with $C_{\mathtt{MSE}}$ a constant, which by construction implies that $n(h_{\mathtt{MSE}})^5 \to c \in (0, \infty)$ and hence leads to a first-order bias in the distributional approximation.[5] 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, e.g., 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.

To illustrate the potential pitfalls of the conventional RD confidence intervals based on the t-statistic $T_{\mathtt{SRD}}(h_n)$ presented above and its data-driven version $T_{\mathtt{SRD}}(\hat{h}_n)$ with $\hat{h}_n$ a bandwidth estimate, we briefly summarize some results from a Monte Carlo study further discussed in Section

[4] Note that because $\hat{\tau}_{\mathtt{SRD}}(h_n)$ is a linear weighted least-squares estimator, an estimator of $\mathsf{V}_{\mathtt{SRD}}(h_n)$ takes the familiar form of Eicker-Huber-White heteroskedasticy-robust standard-errors after "plugging in" estimated residuals.

[5] This is a well-known problem in the nonparametric curve estimation literature (see, e.g., Fan and Gijbels (1996)).

Table 1: Empirical Coverage of different 95% Confidence Intervals

	Conventional		Robust Approach		Bandwidths	
		EC (%)		EC (%)	h_n	b_n
Model 1						
	$T_{\texttt{SRD}}(h_{\texttt{MSE}})$	93.9	$T^{\text{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	94.7	0.166	0.319
	$T_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	84.4	$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	93.3	0.335	0.337
	$T_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	83.1	$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	93.1	0.381	0.381
			$T^{\text{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	94.9	0.166	0.166
			$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	94.7	0.335	0.335
Model 2						
	$T_{\texttt{SRD}}(h_{\texttt{MSE}})$	92.5	$T^{\text{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	94.9	0.082	0.191
	$T_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	24.1	$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	91.2	0.185	0.296
	$T_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	79.1	$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	94.8	0.119	0.119
			$T^{\text{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	94.8	0.082	0.082
			$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	94.8	0.185	0.185
Model 3						
	$T_{\texttt{SRD}}(h_{\texttt{MSE}})$	85.8	$T^{\text{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	95.0	0.260	0.292
	$T_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	87.1	$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	95.1	0.231	0.340
	$T_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	93.9	$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	95.2	0.166	0.166
			$T^{\text{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	94.9	0.260	0.260
			$T^{\text{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	95.0	0.231	0.231

Notes: (i) EC = Empirical Coverage in percentage points, and (ii) columns under "Bandwidths" report the population and average estimated bandwidths choices, as appropriate, for main bandwidth h_n and pilot bandwidth b_n.

6.[6] Table 1 presents the results. We consider three alternative models for the regression function $\mu(x)$ illustrated in Figure 1. The first two models are motivated by empirical RD problems: Model 1 corresponds to a regression function implied by Lee (2008)'s dataset, and Model 2 corresponds to a regression function implied by Ludwig and Miller (2007)'s data. Model 3 is chosen to exhibit a different regression function with more curvature. All other features of the simulation study are held fixed, matching exactly the data generating process in Imbens and Kalyanaraman (2012).

Table 1 reports the empirical coverage of different 95% confidence intervals for each model under two distinct approaches. The first group of columns, labeled "Conventional", corresponds to the conventional approach based on Lemma 1. We consider three different bandwidth choices: (*i*) the infeasible MSE-optimal choice $h_{\texttt{MSE}}$, (*ii*) a data-driven, regularized choice $\hat{h}_{\texttt{IK}}$ proposed by Imbens and Kalyanaraman (2012), and (*iii*) a data-driven, cross-validation (CV) choice $\hat{h}_{\texttt{CV}}$ proposed by Ludwig and Miller (2007). The robust approach column in Table 1 is discussed further below.

The simulation results indeed show that the conventional confidence intervals constructed using Lemma 1 may have poor empirical coverage. In Models 1 and 2, the infeasible confidence intervals that use the MSE-optimal bandwidth ($h_{\texttt{MSE}} = 0.166$ and $h_{\texttt{MSE}} = 0.082$, respectively) have reasonably good empirical coverage (93.9% and 92.5%, respectively), but their data-driven counterparts that employ an estimated bandwidth exhibit substantial undercovering (e.g., for Model 1 the 95% confidence intervals based on $T_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$ and $T_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$ have empirical coverage of 84.4% and 83.1%,

[6] We use these simulation results for motivational purposes only. Our results, presented in the upcoming sections, are theory-based and enjoy certain demostrably superior theoretical properties when compared to the conventional ones.

respectively). In Model 3, which has a regression function with more curvature, even the infeasible confidence interval constructed using the MSE-optimal bandwidth ($h_{\texttt{MSE}} = 0.260$) is biased, showing an empirical coverage of 85.8%.

These simulations illustrate that Lemma 1 may not give a good approximation whenever the bandwidth employed is "large". Since applied researchers often estimate RD treatment effects employing MSE-optimal bandwidths in local-linear regressions and implicitly ignore the large-sample bias of the estimator, the poor coverage of conventional confidence intervals we highlight potentially affects many RD empirical applications. There are two main approaches to deal with this problem, neither of which is commonly used in practice. One is to undersmooth the estimator, that is, choose an ad-hoc "smaller" bandwidth. This approach, however, is not systematic and its effectiveness unavoidably relies on the unknown features of the underlying data generating process (i.e., it is difficult to know how much undersmoothing is needed in a given application). Moreover, from a theoretical perspective, it has the unsatisfactory effect of leading to a suboptimal rate of convergence for the resulting estimator, thereby affecting the local power properties of the associated hypothesis test. Practically, this means that less observations are effectively used for inference, leading to longer confidence intervals on average.

The second approach is to bias-correct the estimator. This approach is systematic and easy to justify theoretically, but is believed to have poor performance in finite samples. The basic idea of bias-correction is to remove the leading bias term by subtracting off a plug-in consistent estimator of it. As described in detail below, our contribution takes bias-correction as a starting point, and corrects this poor performance by providing a different asymptotic approximation that accounts for the added variability introduced by the bias estimate. Before turning to our results, however, we describe the details of bias-correction in the context of the RD estimate. The leading asymptotic bias of the local-linear estimator is

$$\mathbb{E}[\hat{\tau}_{\texttt{SRD}}(h_n)|X_1, X_2, \cdots, X_n] - \tau_{\texttt{SRD}} = h_n^2 \mathsf{B}_{\texttt{SRD}}(h_n)\{1 + o_p(1)\}$$

with

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

where $\mathcal{B}_{+,\texttt{SRD}}(h_n)$ and $\mathcal{B}_{-,\texttt{SRD}}(h_n)$ are observed quantities (function of $X_1, X_2, \cdots, X_n$, $k(\cdot)$ and h_n), which are asymptotically bounded. The exact forms of $\mathcal{B}_{+,\texttt{SRD}}(h_n)$ and $\mathcal{B}_{-,\texttt{SRD}}(h_n)$ are given in Lemma A1 in the Appendix. Therefore, an easy to implement plug-in bias-corrected estimator is given by

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

where

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

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.2 in the Appendix. Here, b_n is the so-called pilot bandwidth sequence, usually larger than h_n. (We employ the same kernel function $k(\cdot)$ to form all estimators for simplicity.) As mentioned above, bias-correction is theoretically appealing because, for example, it allows for a MSE-optimal choice of bandwidth h_n and hence leads to a faster convergence rate of $\hat{\tau}_{\mathtt{SRD}}(h_n)$.

Under weaker conditions, allowing in particular for "larger" bandwidths h_n, the bias-corrected (infeasible) t-statistic satisfies

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

which justifies theoretically confidence intervals for $\tau_{\mathtt{SRD}}$ of the form:

$$\begin{aligned} I^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n) &= \left[\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}\right] \\ &= \left[\left(\hat{\tau}_{\mathtt{SRD}}(h_n) - h_n^2 \hat{\mathsf{B}}_{\mathtt{SRD}}(h_n, b_n)\right) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}\right]. \end{aligned}$$

That is, the confidence intervals are re-centered to account for the presence of the (smoothing) bias. In practice, b_n may also be selected using an MSE-optimal choice, denoted $b_{\mathtt{MSE}}$, which can be implemented by a plug-in estimate, denoted $\hat{b}_{\mathtt{MSE}}$ (see Section 4 for details). As discussed in Section 6, these confidence intervals do not exhibit better empirical coverage than the conventional ones based on $T_{\mathtt{SRD}}(h_n)$, and they do underperform in many cases. Bias-correction is not particularly popular in empirical work, even though the bias-corrected statistic $T^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ may be preferred to the classical statistic $T_{\mathtt{SRD}}(h_n)$ due to its demonstrably better theoretical (asymptotic) properties. Our simulations are consistent with this empirical view on conventional bias-correction.

2.1 Robust Local-Linear RD Inference

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 $\hat{\mu}^{(2)}_{-,2}(b_n)$. In other words, this large-sample approximation relies on carefully tailored assumptions on the bandwidth sequences h_n and b_n that make the variability of the bias-correction estimate disappear asymptotically. We propose an alternative asymptotic approximation for bias-corrected local polynomial estimators that leads to confidence intervals for RD treatment effects capturing this additional sampling variability. This is the main contribution of our paper.

To highlight the differences in our approach, note that the conventional approach to bias-correction assumes that the bias-estimate is a consistent estimator of the asymptotic bias, and thus forces $\hat{\mathsf{B}}_{\mathtt{SRD}}(h_n, b_n)$ to be a "consistent" estimator of $\mathsf{B}_{\mathtt{SRD}}(h_n)$ (i.e., $\hat{\mathsf{B}}_{\mathtt{SRD}}(h_n, b_n)/\mathsf{B}_{\mathtt{SRD}}(h_n) \to_p 1$). Specifically, under the usual regularity conditions, if

$$n h_n^5 \left(\hat{\mathsf{B}}_{\mathtt{SRD}}(h_n, b_n) - \mathsf{B}_{\mathtt{SRD}}(h_n)\right)^2 \to_p 0 \tag{2}$$

then (1) holds. This approach allows for potentially "larger" bandwidths h_n because the leading asymptotic bias is manually removed from the distributional approximation, but the resulting distributional approximation for this bias-corrected estimator tends to provide a poor characterization of the finite sample variability of the statistic. The approximation does not account for the bias-correction component: condition (2) makes researchers proceed "as if" the leading bias is known. In finite samples, however, the bias-correction component will affect the sampling distribution of the estimator $\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$, which implies that the conventional distributional approximation may not accurately represent the finite-sample distribution of $T^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$.

We propose an alternative asymptotic theory that accounts for the potential contribution of the bias-correction estimate to the large sample distributional approximation of the sampling distribution of $T^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$. The idea is to allow for bandwidth sequences, entering in the bias-estimate $\hat{\mathsf{B}}_{\mathtt{SRD}}(h_n, b_n)$, that potentially make the bias-correction term in $\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ as important as the main estimator $\hat{\tau}_{\mathtt{SRD}}(h_n)$, even asymptotically. These bandwidth sequences weaken the condition (2) in an intuitive way, and lead to an alternative distributional approximation for $\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ with a different asymptotic variance in general. The resulting distributional approximation therefore potentially includes both the contribution of the main estimator $\hat{\tau}_{\mathtt{SRD}}(h_n)$ as well as the contribution of the bias-correction estimate.

The intuition behind our result is quite simple. We have

$$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n) = \frac{\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n) - \tau_{\mathtt{SRD}}}{\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}} = \Upsilon_{\mathtt{SRD}}(h_n) - \Upsilon^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n),$$

where

$$\Upsilon_{\mathtt{SRD}}(h_n) = \frac{\hat{\tau}_{\mathtt{SRD}}(h_n) - \tau_{\mathtt{SRD}} - h_n^2 \mathsf{B}_{\mathtt{SRD}}(h_n)}{\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}} \qquad \text{and} \qquad \Upsilon^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n) = \frac{h_n^2 \left(\hat{\mathsf{B}}_{\mathtt{SRD}}(h_n, b_n) - \mathsf{B}_{\mathtt{SRD}}(h_n)\right)}{\sqrt{\mathsf{V}_{\mathtt{SRD}}(h_n)}}.$$

It is easy to see that $\Upsilon_{\mathtt{SRD}}(h_n) \to_d \mathcal{N}(0,1)$. In addition, under appropriate conditions,

$$\Upsilon^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n) = \sqrt{n h_n^5} O_p\left(\frac{1}{\sqrt{n b_n^5}} + b_n\right) = O_p\left(\left(\frac{h_n}{b_n}\right)^{5/2} + \sqrt{n h_n^5 b_n^2}\right),$$

implying that $\Upsilon^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ is asymptotically negligible if (and only if)

$$\frac{h_n}{b_n} \to 0 \qquad \text{and} \qquad n h_n^5 b_n^2 \to 0. \tag{3}$$

The conditions in (3) specialize the high-level condition (2) underlying the classical approach to bias-correction. Specifically, the restriction $h_n/b_n \to 0$ controls the additional variability that the bias-correction term introduces to $\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$, while the condition $n h_n^5 b_n^2 \to 0$ ensures that the bias-correction term is asymptotically unbiased after proper scaling. In finite samples, however, h_n/b_n is never zero. Thus, to capture the (possibly first-order) effect of the bias-correction to the distributional approximation, we study the alternative large-sample approximation for the (properly centered and scaled) estimator $\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ based on the condition

$$\rho_n = \frac{h_n}{b_n} \to \rho \in [0, \infty],$$

which in particular allows for a pilot bandwidth b_n of the same order of (and potentially equal to) the main bandwidth h_n. This approach implies that the bias-correction term will not be consistent for its population counterpart in general, and whenever inconsistent will converge in distribution to a centered at zero normal random variable, provided the asymptotic bias is small enough.

This idea is formalized in the following theorem.

Theorem 1. Suppose Assumptions 1–2 hold with $S \geq 3$. 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

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

where

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

provided $\kappa \max\{h_n, b_n\} < \kappa_0$. The exact variance formula $\mathsf{V}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ is notationally cumbersome and thus given in the Appendix in equation (A-1).

Theorem 1 shows that by standardizing the bias-corrected estimator by its (conditional) variance, the asymptotic distribution of the resulting bias-corrected statistic $T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n)$ is Gaussian even when the condition $h_n/b_n \to 0$ is violated. This leads to a different asymptotic variance for the bias-corrected estimator $\hat{\tau}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ in general, which depends on the behavior of $\rho_n = h_n/b_n$. In the new variance $\mathsf{V}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ provided in this theorem, $\mathsf{C}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ may be interpreted as a standard-error correction to account for the variability of the estimated bias-correction term.

The key practical implication of Theorem 1 is that it justifies the more robust, theory-based $100(1-\alpha)$-percent confidence intervals:

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

The group of columns in Table 1 labeled "Robust Approach" exhibits the performance of the new confidence intervals employing $T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n)$ for different bandwidth choices, which perform remarkably well when compared to the other alternatives. We also present results for the choice $h_n = b_n$ (i.e., $\rho = 1$) because, as mentioned in Remark 3 below, in this special case $T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, h_n)$ coincides with the statistic constructed using a simple local-quadratic estimator without bias correction. Thus, this choice of bandwidths gives a simple possible implementation for our approach: choose h_n to be the cross-validated or MSE-optimal bandwidth choice for the local-linear RD estimator, but form confidence intervals for $\tau_{\mathtt{SRD}}$ using the local-quadratic RD estimator. We summarize important features of our main result in the remarks below.

Remark 1. The distributional approximation in Theorem 1 permits one bandwidth (but not both) to be fixed, provided it is not too large; i.e., both must satisfy $\kappa \max\{h_n, b_n\} < \kappa_0$, but only one needs to vanish.

Remark 2. Three main limiting cases are obtained depending on the limit $\rho_n \to \rho \in [0, \infty]$.

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

Case 2: $\rho \in (0, \infty)$. In this case $h_n = \rho b_n$ and $\mathsf{C}^{\mathsf{bc}}_{\mathsf{SRD}}(h_n, b_n) \asymp_p \mathsf{V}_{\mathsf{SRD}}(h_n)$. This is the knife-edge case where both $\hat{\tau}_{\mathsf{SRD}}(h_n)$ and $h_n^2 \hat{\mathsf{B}}_{\mathsf{SRD}}(h_n, b_n)$ contribute asymptotically and thus $\mathsf{V}^{\mathsf{bc}}_{\mathsf{SRD}}(h_n, b_n)$ captures the additional contribution of the bias-correction.

Case 3: $\rho = \infty$. In this case $b_n = o(h_n)$ and $\mathsf{V}_{\mathsf{SRD}}(h_n) = o_p(\mathsf{C}^{\mathsf{bc}}_{\mathsf{SRD}}(h_n, b_n))$, which implies that the bias-estimate is first-order while the actual estimator $\hat{\tau}_{\mathsf{SRD}}(h_n)$ is of smaller order. Here $\mathsf{V}^{\mathsf{bc}}_{\mathsf{SRD}}(h_n, b_n) \asymp_p \mathbb{V}[h_n^2 \hat{\mathsf{B}}_{\mathsf{SRD}}(h_n, b_n) | X_1, \cdots, X_n]$.

Remark 3. If $h_n = b_n$ (and the same kernel function $k(\cdot)$ is used), then $\hat{\tau}^{\mathsf{bc}}_{\mathsf{SRD}}(h_n, h_n)$ is numerically equivalent to the local-quadratic estimator ($p = 2$) of τ_{SRD}. This gives a simple relationship between local polynomial estimators of order p and $p + 1$, and their relation to manual bias-correction. See the appendix and the online supplemental appendix for further details and generalizations.

Remark 4. Theorem 1 and Remark 3 give a simple, formal justification for an approach based on the order of the local polynomial: a theoretically valid choice of h_n is to select the MSE-optimal bandwidth for the local-linear estimator, but construct confidence intervals using the local-quadratic estimator instead. This approach corresponds exactly to the case $h_n = b_n$ in Theorem 1. See the appendix and the online supplemental appendix for further details and generalizations.

Remark 5. All the results in this paper apply immediately when different bandwidths ($h_{+,n}$, $h_{-,n}$, $b_{+,n}$, $b_{-,n}$, say) are employed to construct the estimators $\hat{\mu}_{+,p}(h_{+,n})$, $\hat{\mu}_{-,p}(h_{-,n})$ and their associated bias-correction terms.

3 Other RD Designs

In this section, we discuss three applications of our main idea to other practically relevant settings: sharp kink RD, fuzzy RD and fuzzy kink RD designs. All the results briefly summarized here are special cases of Theorems A1 and A2 presented 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 the robust confidence intervals.

3.1 Sharp Kink RD

In this setting, the interest lies on the difference of the first derivative of the regression functions at the cutoff, as opposed to the differences in the levels of those functions. For details on identification and inference procedures using conventional approaches see, e.g., Card, Lee, Pei, and Weber (2012), Dong (2012), Dong and Lewel (2012), and references therein. The estimand of interest is

$$\tau_{\texttt{SKRD}} = \mu_+^{(1)} - \mu_-^{(1)}.$$

Although a local-linear estimator could still be used in this context, it is perhaps more appropriate to employ a local-quadratic estimator due to boundary-bias considerations. Thus, letting $\hat{\mu}_{+,2}^{(1)}(h_n)$ and $\hat{\mu}_{-,2}^{(1)}(h_n)$ denote local-quadratic estimators of $\mu_+^{(1)}$ and $\mu_-^{(1)}$ (see Section A.2 in the Appendix), we focus on the local-quadratic RD estimator

$$\hat{\tau}_{\texttt{SKRD}}(h_n) = \hat{\mu}_{+,2}^{(1)}(h_n) - \hat{\mu}_{-,2}^{(1)}(h_n).$$

Lemma A1 in the Appendix gives conditions so that

$$T_{\texttt{SKRD}}(h_n) = \frac{\hat{\tau}_{\texttt{SKRD}}(h_n) - \tau_{\texttt{SKRD}}}{\sqrt{\mathsf{V}_{\texttt{SKRD}}(h_n)}} \to_d \mathcal{N}(0,1), \qquad \mathsf{V}_{\texttt{SKRD}}(h_n) = \mathbb{V}[\hat{\tau}_{\texttt{SKRD}}(h_n)|X_1, X_2, \cdots, X_n],$$

which corresponds to the conventional distributional approximation. Following Imbens and Kalyanaraman (2012), a MSE-optimal bandwidth choice for $\tau_{\texttt{SKRD}}$ can be derived (see Lemma 2 in Section 4). This choice, among others, will again lead to a non-negligible first-order bias. Proceeding as before, we also have $\mathbb{E}[\hat{\tau}_{\texttt{SKRD}}(h_n)|X_1, X_2, \cdots, X_n] - \tau_{\texttt{SKRD}} \approx h_n^2 \mathsf{B}_{\texttt{SKRD}}(h_n)$ with

$$\mathsf{B}_{\texttt{SKRD}}(h_n) = \frac{\mu_+^{(3)}}{3!}\mathcal{B}_{+,\texttt{SKRD}}(h_n) - \frac{\mu_-^{(3)}}{3!}\mathcal{B}_{-,\texttt{SKRD}}(h_n),$$

where $\mathcal{B}_{+,\texttt{SKRD}}(h_n)$ and $\mathcal{B}_{-,\texttt{SKRD}}(h_n)$ are asymptotically bounded observed quantities (function of $X_1, X_2, \cdots, X_n$, $k(\cdot)$ and h_n), also given in Lemma A1. Therefore, a bias-corrected local-quadratic estimator of $\tau_{\texttt{SKRD}}$, now using a local-cubic bias-correction, is given by

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

with

$$\hat{\mathsf{B}}_{\texttt{SKRD}}(h_n, b_n) = \frac{\hat{\mu}_{+,3}^{(3)}(b_n)}{3!}\mathcal{B}_{+,\texttt{SKRD}}(h_n) - \frac{\hat{\mu}_{-,3}^{(3)}(b_n)}{3!}\mathcal{B}_{-,\texttt{SKRD}}(h_n),$$

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, as discussed in Section A.2 of the Appendix.

With these preliminaries, we can present our main result for the kink RD bias-corrected estimator $\hat{\tau}_{\texttt{SKRD}}^{\texttt{bc}}(h_n, b_n)$.

Theorem 2. Suppose Assumptions A1–A2 hold with $S \geq 4$. If $n \min\{h_n^7, b_n^7\} \max\{h_n^2, b_n^2\} \to 0$

and $n \min\{h_n, b_n\} \to \infty$, then

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

provided $\kappa \max\{h_n, b_n\} < \kappa_0$. The exact form of $\mathsf{V}^{\texttt{bc}}_{\texttt{SKRD}}(h_n, b_n)$ is given in Theorem A1 in the Appendix in equation (A-1).

This theorem is an analogue of Theorem 1 for the sharp kink RD design. In particular, it derives the new asymptotic variance formula $\mathsf{V}^{\texttt{bc}}_{\texttt{SKRD}}(h_n, b_n)$ capturing the additional contribution of the bias-correction to the sampling variability. In this case, the new variance also takes the form $\mathsf{V}^{\texttt{bc}}_{\texttt{SKRD}}(h_n, b_n) = \mathsf{V}_{\texttt{SKRD}}(h_n) + \mathsf{C}^{\texttt{bc}}_{\texttt{SKRD}}(h_n, b_n)$, where $\mathsf{C}^{\texttt{bc}}_{\texttt{SKRD}}(h_n, b_n)$ is the correction term. This result theoretically justifies the following more robust $100(1-\alpha)$-percent confidence interval for $\tau_{\texttt{SKRD}}$:

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

3.2 Fuzzy RD

In the fuzzy design, actual treatment status may differ from treatment assignment and is thus only partially determined by the running variable. Identification and conventional inference approaches are discussed in Hahn, Todd, and van der Klaauw (2001) and Porter (2003). To handle this case, we introduce the following additional notation: $(Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)'$, $i = 1, 2, \ldots, n$, is a random sample from the random vector $(Y(0), Y(1), T(0), T(1), X)'$, where in this case treatment status for each unit is determined by

$$T_i = T_i(0) \cdot \mathbf{1}(X_i < 0) + T_i(1) \cdot \mathbf{1}(X_i \geq 0) = \begin{cases} T_i(0) & \text{if } X_i < 0 \\ T_i(1) & \text{if } X_i \geq 0 \end{cases},$$

with $T_i(0), T_i(1) \in \{0, 1\}$, and the corresponding observed outcome variable is

$$Y_i = Y_i(0) \cdot (1 - T_i) + Y_i(1) \cdot T_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}.$$

The observed random sample now is $\{(Y_i, T_i, X_i)' : i = 1, 2, \ldots, n\}$. The estimand of interest is

$$\tau_{\texttt{FRD}} = \frac{\mathbb{E}[Y_i(1)|X = 0] - \mathbb{E}[Y_i(0)|X = 0]}{\mathbb{E}[T_i(1)|X = 0] - \mathbb{E}[T_i(0)|X = 0]},$$

provided that $\mathbb{E}[T_i(1)|X = 0] - \mathbb{E}[T_i(0)|X = 0] \neq 0$. For further discussion on the interpretation of $\tau_{\texttt{FRD}}$ also see Imbens and Lemieux (2008). Under appropriate conditions, this estimand is also nonparametrically identifiable as

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

where here, and elsewhere as needed, we make explicit the outcome variable underlying the population parameter. That is, $\tau_{Y,\mathtt{SRD}} = \mu_{Y+} - \mu_{Y-}$ with

$$\mu_{Y+} = \lim_{x\to 0^+} \mu_Y(x), \qquad \mu_{Y-} = \lim_{x\to 0^-} \mu_Y(x), \qquad \mu_Y(x) = \mathbb{E}[Y_i|X_i = x],$$

and $\tau_{T,\mathtt{SRD}} = \mu_{T+} - \mu_{T-}$ with

$$\mu_{T+} = \lim_{x\to 0^+} \mu_T(x), \qquad \mu_{T-} = \lim_{x\to 0^-} \mu_T(x), \qquad \mu_T(x) = \mathbb{E}[T_i|X_i = x].$$

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

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

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.2 in the Appendix for details.

To describe the large-sample results for $\hat{\tau}_{\mathtt{FRD}}(h_n)$ we employ the additional standard Assumption A3 in Section A.1 of the Appendix. Under Assumptions A1–A3, and appropriate bandwidth conditions, the conventional large-sample properties of $\hat{\tau}_{\mathtt{FRD}}$ are characterized by noting that

$$\hat{\tau}_{\mathtt{FRD}}(h_n) - \tau_{\mathtt{FRD}} = \tilde{\tau}_{\mathtt{FRD}}(h_n) + R_n$$

with

$$\tilde{\tau}_{\mathtt{FRD}}(h_n) = \frac{1}{\tau_{T,\mathtt{SRD}}}(\hat{\tau}_{Y,\mathtt{SRD}}(h_n) - \tau_{Y,\mathtt{SRD}}) - \frac{\tau_{Y,\nu}}{\tau^2_{T,\mathtt{SRD}}}(\hat{\tau}_{T,\mathtt{SRD}}(h_n) - \tau_{T,\mathtt{SRD}})$$

and $R_n = o_p((\hat{\tau}_{T,\mathtt{SRD}}(h_n) - \tau_{T,\mathtt{SRD}})^2 + (\hat{\tau}_{Y,\mathtt{SRD}}(h_n) - \tau_{Y,\mathtt{SRD}})(\hat{\tau}_{T,\mathtt{SRD}}(h_n) - \tau_{T,\mathtt{SRD}}))$. This shows that, to first-order, the fuzzy RD estimator behaves like a linear combination of two sharp RD estimators. Thus, as Lemma A2 in the appendix shows,

$$T_{\mathtt{FRD}}(h_n) = \frac{\hat{\tau}_{\mathtt{FRD}}(h_n) - \tau_{\mathtt{FRD}}}{\sqrt{\mathsf{V}_{\mathtt{FRD}}(h_n)}} \to_d \mathcal{N}(0,1), \qquad \mathsf{V}_{\mathtt{FRD}}(h_n) = \mathbb{V}[\tilde{\tau}_{\mathtt{FRD}}(h_n)|X_1, X_2, \cdots, X_n].$$

In this case, the leading (smoothing) bias of the local-linear fuzzy RD estimator $\hat{\tau}_{\mathtt{FRD}}(h_n)$ is given by $\mathbb{E}[\tilde{\tau}_{\mathtt{FRD}}(h_n)|X_1, X_2, \cdots, X_n] \approx h_n^2 \mathsf{B}_{\mathtt{FRD}}(h_n)$ with

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

where $\mathcal{B}_{+,\mathtt{FRD}}(h_n)$ and $\mathcal{B}_{-,\mathtt{FRD}}(h_n)$ are also asymptotically bounded observed quantities (function of $X_1, X_2, \cdots, X_n$, $k(\cdot)$ and h_n) and given in Lemma A2. Therefore, we construct a bias-corrected estimator of $\tau_{\texttt{SRD}}$ employing a local-quadratic estimate of the leading biases, which is given by

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

with

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

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

With these preliminaries, we obtain the following theorem resembling the previous discussion for sharp RD designs.

Theorem 3. Suppose Assumptions A1–A3 hold with $S \geq 3$, and $\tau_{T,\texttt{SRD}} \neq 0$. 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

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

provided that $h_n \to 0$ and $\kappa b_n < \kappa_0$. The exact form of $\mathsf{V}^{\texttt{bc}}_{\texttt{FRD}}(h_n, b_n)$ is given in Theorem A2 in equation (A-2).

3.3 Fuzzy Kink RD

Our final extension considers confidence intervals for average treatment effects in the fuzzy kink RD design (see, e.g., Card, Lee, Pei, and Weber (2012) and references therein for further discussion). We retain the notation and assumptions introduced for the fuzzy RD design above. In this setting the parameter of interest is

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

and therefore a natural estimator based on two local-quadratic (reduced form) estimates is

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

where, as before, the formal definitions of these local polynomial estimators are given in Section A.2 in the Appendix.

The linearization argument given for the fuzzy RD estimator applies here as well, leading to

$$\hat{\tau}_{\texttt{FKRD}}(h_n) - \tau_{\texttt{FKRD}} = \tilde{\tau}_{\texttt{FKRD}}(h_n) + R_n$$

with

$$\tilde{\tau}_{\texttt{FKRD}}(h_n) = \frac{1}{\tau_{T,\texttt{SKRD}}}(\hat{\tau}_{Y,\texttt{SKRD}}(h_n) - \tau_{Y,\texttt{SKRD}}) - \frac{\tau_{Y,\nu}}{\tau^2_{T,\texttt{SKRD}}}(\hat{\tau}_{T,\texttt{SKRD}}(h_n) - \tau_{T,\texttt{SKRD}})$$

and $R_n = o_p((\hat{\tau}_{T,\texttt{SKRD}}(h_n) - \tau_{T,\texttt{SKRD}})^2 + (\hat{\tau}_{Y,\texttt{SKRD}}(h_n) - \tau_{Y,\texttt{SKRD}})(\hat{\tau}_{T,\texttt{SKRD}}(h_n) - \tau_{T,\texttt{SKRD}}))$. Employing Lemma A2 in the appendix once more, we verify that

$$T_{\texttt{FKRD}}(h_n) = \frac{\hat{\tau}_{\texttt{FKRD}}(h_n) - \tau_{\texttt{FKRD}}}{\sqrt{\mathsf{V}_{\texttt{FKRD}}(h_n)}} \to_d \mathcal{N}(0,1), \qquad \mathsf{V}_{\texttt{FKRD}}(h_n) = \mathbb{V}[\tilde{\tau}_{\texttt{FKRD}}(h_n)|X_1, X_2, \cdots, X_n],$$

and $\mathbb{E}[\tilde{\tau}_{\texttt{FKRD}}(h_n)|X_1, X_2, \cdots, X_n] \approx h_n^2 \mathsf{B}_{\texttt{FKRD}}(h_n)$ with

$$\begin{aligned} \mathsf{B}_{\texttt{FKRD}}(h_n) &= \left(\frac{1}{\tau_{T,\texttt{SKRD}}}\frac{\mu^{(3)}_{Y+}}{3!} - \frac{\tau_{Y,\texttt{SKRD}}}{\tau^2_{T,\texttt{SKRD}}}\frac{\mu^{(3)}_{T+}}{3!}\right)\mathcal{B}_{+,\texttt{FKRD}}(h_n) \\ &\quad - \left(\frac{1}{\tau_{T,\texttt{SKRD}}}\frac{\mu^{(3)}_{Y-}}{3!} - \frac{\tau_{Y,\texttt{SKRD}}}{\tau^2_{T,\texttt{SKRD}}}\frac{\mu^{(3)}_{T-}}{3!}\right)\mathcal{B}_{-,\texttt{FKRD}}(h_n), \end{aligned}$$

where $\mathcal{B}_{+,\texttt{FKRD}}(h_n)$ and $\mathcal{B}_{-,\texttt{FKRD}}(h_n)$ are asymptotically bounded observed quantities (function of $X_1, X_2, \cdots, X_n$, $k(\cdot)$ and h_n), also given in Lemma A2.

Thus, we propose a plug-in bias-corrected estimator of $\tau_{\texttt{FKRD}}$ employing a local-cubic estimate of the leading biases, which gives the bias-corrected estimator

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

with

$$\begin{aligned} \hat{\mathsf{B}}_{\texttt{FKRD}}(h_n, b_n) &= \left(\frac{1}{\hat{\tau}_{T,\texttt{SKRD}}(h_n)}\frac{\hat{\mu}^{(3)}_{Y+,3}(b_n)}{3!} - \frac{\hat{\tau}_{Y,\texttt{SKRD}}(h_n)}{\hat{\tau}^2_{T,\texttt{SKRD}}(h_n)}\frac{\hat{\mu}^{(3)}_{T+,3}(b_n)}{3!}\right)\mathcal{B}_{+,\texttt{FKRD}}(h_n) \\ &\quad - \left(\frac{1}{\hat{\tau}_{T,\texttt{SKRD}}(h_n)}\frac{\hat{\mu}^{(3)}_{Y-,3}(b_n)}{3!} - \frac{\hat{\tau}_{Y,\texttt{SKRD}}(h_n)}{\hat{\tau}^2_{T,\texttt{SKRD}}(h_n)}\frac{\hat{\mu}^{(3)}_{T-,3}(b_n)}{3!}\right)\mathcal{B}_{-,\texttt{FKRD}}(h_n). \end{aligned}$$

The following theorem describes our result for the case of the fuzzy kink RD design.

Theorem 4. Suppose Assumptions 1–3 hold with $S \geq 4$, and $\tau_{T,\texttt{SKRD}} \neq 0$. 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

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

provided that $h_n \to 0$ and $\kappa b_n < \kappa_0$. The exact form of $\mathsf{V}^{\mathtt{bc}}_{\mathtt{FKRD}}(h_n, b_n)$ is given in Theorem A2 in equation (A-2).

4 Validity of MSE-Optimal Bandwidths Selectors

The purpose of this paper is to present more robust confidence intervals for RD estimands, based on bias-correction techniques and an alternative asymptotic approximation. These results were obtained using bandwidth sequences that in practice need to be chosen in some way. Imbens and Kalyanaraman (2012) give a recent overview on bandwidth selection in the RD design, where MSE-minimizing and cross-validation procedures are described.

In this section we derive MSE-optimal bandwidth choices for h_n and b_n, which apply to the main four RD settings of interest discussed previously, and show that these choices are fully compatible with our asymptotic distribution results (but not with the conventional ones). In the appendix, we also propose direct plug-in, data-driven bandwidth selectors for the sharp designs. In Section 6, we explore in a simulation study the performance of these bandwidth selectors as well as several alternatives available in the literature, while in Section 7 we employ them in an empirical illustration.

4.1 Sharp Designs

Assuming $\nu \le p$, the estimands in the sharp RD designs can be written as

$$\tau_\nu = \mu_+^{(\nu)} - \mu_-^{(\nu)},$$

where, in particular, $\tau_{\mathtt{SRD}} = \tau_0$ and $\tau_{\mathtt{SKRD}} = \tau_1$. As described in Section A.2 in the Appendix, the corresponding p-th order local-polynomial estimators are

$$\hat{\tau}_{\nu,p}(h_n) = \hat{\mu}_{+,p}^{(\nu)}(h_n) - \hat{\mu}_{-,p}^{(\nu)}(h_n),$$

where, in particular, $\hat{\tau}_{\mathtt{SRD}}(h_n) = \hat{\tau}_{0,1}(h_n)$ and $\hat{\tau}_{\mathtt{SKRD}}(h_n) = \hat{\tau}_{1,2}(h_n)$.

Therefore, we consider the generic MSE objective function

$$\mathsf{MSE}_{\nu,p}(h_n) = \mathbb{E}\left[(\hat{\tau}_{\nu,p}(h_n) - \tau_\nu)^2 \middle| X_1, X_2, \cdots, X_n \right].$$

Lemma 2. Suppose Assumptions 1–2 hold with $S \ge p+1$. Let $\nu \in \mathbb{N}$ with $\nu \le p$.

(**MSE**) If $h_n \to 0$ and $nh_n \to \infty$, then

$$\mathsf{MSE}_{\nu,p}(h_n) = h_n^{2(p+1-\nu)} \; \left[\mathsf{B}^2_{\nu,p,p+1} + o_p(1)\right] + \frac{1}{nh_n^{1+2\nu}} \; \left[\mathsf{V}_{\nu,p} + o_p(1)\right],$$

where

$$\mathsf{B}_{\nu,p,r} = \frac{\mu_+^{(r)} - \mu_-^{(r)}}{r!} e_\nu' \Gamma_p^{-1} \vartheta_{p,r}, \qquad \mathsf{V}_{\nu,p} = \frac{\sigma_-^2 + \sigma_+^2}{f} e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu.$$

The exact form of the other matrices is given in Section A.3 in the Appendix.

(**OB**) If $\mu_+^{(p+1)} \neq \mu_-^{(p+1)}$, then the (asymptotic) MSE-optimal bandwidth is

$$h_{\mathtt{MSE},\nu,p} = C_{\mathtt{MSE},\nu,p}\; n^{-\frac{1}{2p+3}}, \qquad C_{\mathtt{MSE},\nu,p} = \left(\frac{(1+2\nu)\mathsf{V}_{\nu,p}}{2(p+1-\nu)\mathsf{B}^2_{\nu,p,p+1}}\right)^{\frac{1}{2p+3}}.$$

This lemma justifies a set of MSE-optimal (infeasible) choices for h_n and b_n: $h_n = h_{\mathtt{MSE},0,1}$ and $b_n = h_{\mathtt{MSE},2,2}$ for Theorem 1, and $h_n = h_{\mathtt{MSE},1,2}$ and $b_n = h_{\mathtt{MSE},3,3}$ for Theorem 2. This result generalizes the work in Imbens and Kalyanaraman (2012), who proposed the choice $h_{\mathtt{MSE},0,1}$ for local-linear estimators. In Section A.6 in the Appendix we also construct direct plug-in (DPI) selectors for h_n and b_n based on these choices. Our construction employs the fact that $\mathsf{V}_{\nu,p} = \lim_{n\to\infty} nh_n^{1+2\nu}\mathbb{V}[\hat{\tau}_{\nu,p}(h_n)|X_1, X_2, \cdots, X_n]$ if $h_n \to 0$ and $nh_n \to \infty$, together with the standard-errors estimators proposed in the next section, to construct consistent plug-in estimates of the variance terms (Theorem 5), which thus avoids using consistent estimators of σ_+^2, σ_-^2 and f directly. Following Imbens and Kalyanaraman (2012), we also incorporate "regularization" to avoid small denominators. The online supplemental appendix contains a detailed discussion of our approach, and a comparison to other methods available in the RD literature. Theorem A-3 in the Appendix also shows that our bandwidth selectors are consistent and optimal in the sense of Li (1987).

Remark 6. The MSE-optimal bandwidth choices for the sharp designs are fully compatible with our asymptotic approximations given above, as they satisfy the rate-restrictions in Theorems 1–2. For example, in the case of Theorem 1, $n\min\{h_{\mathtt{MSE},0,1}, b_{\mathtt{MSE},2,2}\} \to \infty$, $n\min\{h^5_{\mathtt{MSE},0,1}, b^5_{\mathtt{MSE},2,2}\}\max\{h^2_{\mathtt{MSE},0,1}, b^2_{\mathtt{MSE},2,2}\} \to 0$.

Remark 7. The MSE-optimal bandwidth choices satisfy $\rho_n = h_{\mathtt{MSE},\nu,p}/b_{\mathtt{MSE},p+1,q} \to 0$. It remains an open question whether the choice $\rho_n \to 0$ is "optimal" from a distributional approximation point of view. Although beyond the scope of this paper, research on this question is underway.

4.2 Fuzzy Designs

Assuming $\nu \leq p$, the estimands in the fuzzy RD designs can be written as

$$\varsigma_\nu = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}}, \qquad \tau_{Y,\nu} = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}, \qquad \tau_{T,\nu} = \mu_{T+}^{(\nu)} - \mu_{T-}^{(\nu)},$$

where, in particular, $\tau_{\mathtt{FRD}} = \varsigma_0$ and $\tau_{\mathtt{FKRD}} = \varsigma_1$. As described in Section A.2 in the Appendix, the corresponding p-th order local-polynomial estimators are

$$\hat{\varsigma}_{\nu,p}(h_n) = \frac{\hat{\tau}_{Y,\nu}(h_n)}{\hat{\tau}_{T,\nu}(h_n)}, \qquad \hat{\tau}_{Y,\nu}(h_n) = \hat{\mu}^{(\nu)}_{Y+,p}(h_n) - \hat{\mu}^{(\nu)}_{Y-,p}(h_n), \qquad \hat{\tau}_{T,\nu}(h_n) = \hat{\mu}^{(\nu)}_{T+,p}(h_n) - \hat{\mu}^{(\nu)}_{T-,p}(h_n),$$

and its first-order linear approximation is

$$\tilde{\varsigma}_{\nu,p}(h_n) = \frac{1}{\tau_{T,\nu}}(\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu}) - \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu}).$$

Notice that, in particular, $\hat{\tau}_{\mathtt{FRD}}(h_n) = \hat{\varsigma}_{0,1}(h_n)$ and $\hat{\tau}_{\mathtt{FKRD}}(h_n) = \hat{\varsigma}_{1,2}(h_n)$.

Following Imbens and Kalyanaraman (2012), we consider the generic MSE objective function

$$\mathsf{MSE}_{\mathtt{F},\nu,p}(h_n) = \mathbb{E}\left[(\tilde{\varsigma}_{\nu,p}(h_n))^2 \middle| X_1, X_2, \cdots, X_n \right].$$

Lemma 3. Suppose Assumptions 1–3 hold with $S \geq p+1$. Let $\nu \in \mathbb{N}$ with $\nu \leq p$.

(**MSE**) If $h_n \to 0$ and $nh_n \to \infty$, then

$$\mathsf{MSE}_{\mathtt{F},\nu,p}(h_n) = h_n^{2(p+1-\nu)} \; \left[\mathsf{B}^2_{\mathtt{F},\nu,p,p+1} + o_p(1)\right] + \frac{1}{nh_n^{1+2\nu}} \; \left[\mathsf{V}_{\mathtt{F},\nu,p} + o_p(1)\right],$$

where

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

$$\mathsf{V}_{\mathtt{F},\nu,p} = \left(\frac{1}{\tau_{T,\nu}} \frac{\sigma^2_{YY-} + \sigma^2_{YY+}}{f} - \frac{2\tau_{Y,\nu}}{\tau^3_{T,\nu}} \frac{\sigma^2_{YT-} + \sigma^2_{YT+}}{f} + \frac{\tau^2_{Y,\nu}}{\tau^4_{T,\nu}} \frac{\sigma^2_{TT-} + \sigma^2_{TT+}}{f} \right) e'_\nu \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu.$$

The exact form of the other matrices is also given in Section A.3 in the Appendix.

(**OB**) If $\mathsf{B}_{\mathtt{F},\nu,p,p+1} \neq 0$, then the (asymptotic) MSE-optimal bandwidth is

$$h_{\mathtt{MSE},\mathtt{F},\nu,p} = C_{\mathtt{MSE},\mathtt{F},\nu,p} \; n^{-\frac{1}{2p+3}}, \qquad C_{\mathtt{MSE},\mathtt{F},\nu,p} = \left(\frac{(2\nu+1)\mathsf{V}_{\mathtt{F},\nu,p}}{2(p+1-\nu)\mathsf{B}^2_{\mathtt{F},\nu,p,p+1}} \right)^{\frac{1}{2p+3}}.$$

Proceeding as in the sharp RD cases, and using Lemma 2, infeasible bandwidth choices for h_n and b_n in Theorems 3–4 are readily available: $h_n = h_{\mathtt{MSE},\mathtt{F},0,1}$ and $b_n = h_{\mathtt{MSE},\mathtt{F},2,2}$ for Theorem 3, and $h_n = h_{\mathtt{MSE},\mathtt{F},1,2}$ and $b_n = h_{\mathtt{MSE},\mathtt{F},3,3}$ for Theorem 4. Feasible versions could also be developed along the lines discussed in Section A.6 in the Appendix. Importantly, just as in the sharp RD cases, the resulting optimal bandwidth choices are fully compatible with our asymptotic theory.

5 Standard Errors

In this section we propose valid standard-error estimators to implement the infeasible statistics presented in the previous sections. The exact formulas for the new proposed variances $\mathsf{V}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ [sharp RD], $\mathsf{V}^{\mathtt{bc}}_{\mathtt{SKRD}}(h_n, b_n)$ [sharp kink RD], $\mathsf{V}^{\mathtt{bc}}_{\mathtt{FRD}}(h_n, b_n)$ [fuzzy RD] and $\mathsf{V}^{\mathtt{bc}}_{\mathtt{FKRD}}(h_n, b_n)$ [fuzzy kink RD] in Theorems 1–4, respectively, are straightforward to derive but notationally cumbersome. They all have the same structure as they are derived by computing the conditional variance of (linear combinations of) weighted linear least-squares estimators. Thus, the only unknowns are diagonal matrices whose diagonal terms contain different conditional covariances depending on the setting under consideration.

In general, for all the cases considered in this paper, the key matrices containing unknown quantities are

$$\Psi_{YY+,p,q}(h_n,b_n),\qquad \Psi_{YT+,p,q}(h_n,b_n),\qquad \Psi_{TT+,p,q}(h_n,b_n),$$

$$\Psi_{YY-,p,q}(h_n,b_n),\qquad \Psi_{YT-,p,q}(h_n,b_n),\qquad \Psi_{TT-,p,q}(h_n,b_n),$$

with $p,q\in\mathbb{N}_+$, and with the generic notation

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

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

$$\sigma^2_{UV}(x)=\mathbb{C}ov[U,V|X=x].$$

Here U and V are placeholders for either Y_i or T_i. This generality is required to handle the fuzzy designs, where the covariances between Y_i and T_i naturally arise. Theorems A-1 and A-2 in the Appendix give the exact formulas for the standard-errors, 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 computed exactly in the same way as, and are actually a generalization of the middle matrix in, the traditional Huber-Eicker-White heteroskedasticity-robust standard-error formula from linear models. In fact, an analogue of these standard-errors could be constructed by plugging in the corresponding estimated residuals. 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 are more robust in finite-samples because they do not depend on kernel-based regression estimators and thus avoid the (implicit) bandwidth choices. Specifically, we define

$$\hat{\Psi}_{UV+,p,q}(h_n,b_n)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i\geq 0)K_{h_n}(X_i)K_{b_n}(X_i)r_p\left(\frac{X_i}{h_n}\right)r_q\left(\frac{X_i}{b_n}\right)'\hat{\sigma}^2_{UV+}(X_i),$$

$$\hat{\Psi}_{UV-,p,q}(h_n,b_n)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i< 0)K_{h_n}(X_i)K_{b_n}(X_i)r_p\left(\frac{X_i}{h_n}\right)r_q\left(\frac{X_i}{b_n}\right)'\hat{\sigma}^2_{UV-}(X_i),$$

with

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

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

where $\ell_j^+(i)$ is the j-th closest unit to unit i among $\{X_i : X_i \geq 0\}$ and $\ell_j^-(i)$ is the j-th closest unit to unit i among $\{X_i : X_i < 0\}$.[7] These estimators are asymptotically valid for any choice of $J \in \mathbb{N}_+$, because they are approximately conditionally unbiased (even though inconsistent when the number of nearest-neighbors J is kept fixed).

The following result gives consistency of the proposed estimators to their infeasible counterparts.

Theorem 5. Suppose $\sigma^2(x)$ is Lipschitz continuous on $(-\kappa_0, 0]$ and on $[0, \kappa_0)$.

(**1**) If the conditions in Theorems 1-2 hold, then

$$\hat{\Psi}_{YY+,p,q}(h_n, b_n) = \Psi_{YY+,p,q}(h_n, b_n) + o_p\left(\min\{h_n^{-1}, b_n^{-1}\}\right),$$

$$\hat{\Psi}_{YY-,p,q}(h_n, b_n) = \Psi_{YY-,p,q}(h_n, b_n) + o_p\left(\min\{h_n^{-1}, b_n^{-1}\}\right).$$

(**2**) If the conditions in Theorem 3-4 hold, then

$$\hat{\Psi}_{UV+,p,q}(h_n, b_n) = \Psi_{UV+,p,q}(h_n, b_n) + o_p\left(\min\{h_n^{-1}, b_n^{-1}\}\right),$$

$$\hat{\Psi}_{UV-,p,q}(h_n, b_n) = \Psi_{UV-,p,q}(h_n, b_n) + o_p\left(\min\{h_n^{-1}, b_n^{-1}\}\right),$$

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

This theorem implies that 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 $\Psi_{UV-,p,q}(h_n, b_n)$, as appropriate in each case, to construct the estimators $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$, $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SKRD}}(h_n, b_n)$, $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{FRD}}(h_n, b_n)$ and $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{FKRD}}(h_n, b_n)$ lead to consistent standard-error estimators. For example, in Theorem 1, our standard-error formula $\hat{\mathsf{V}}^{\mathtt{bc}}_{\mathtt{SRD}}(h_n, b_n)$ involves the following estimators:

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

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

leading to the feasible (for bandwidth choices h_n and b_n) confidence intervals:

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

The other confidence intervals can be constructed analogously.

[7] Alternatively, we could use "local sample covariances" as proposed in Abadie and Imbens (2010).

6 Simulation Evidence

We explored the main implications of our theoretical results in a Monte Carlo experiment. To facilitate comparisons, we employed the data generating process proposed in Imbens and Kalyanaraman (2012, henceforth IK), focusing only on the sharp RD design. We conducted $S = 10,000$ replications, and for each replication we generated a random sample $\{(X_i, \varepsilon_i)' : i = 1, ..., n\}$ 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 $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.1295$. We considered the three regression functions plotted in Figure 1, which are denoted $\mu_1(x)$, $\mu_2(x)$ and $\mu_3(x)$, respectively, and thus generated $Y_i = \mu_j(X_i) + \varepsilon_i$, $i = 1, 2, \cdots, n$, for each regression model $j = 1, 2, 3$. The exact functional form of these regression functions and all other details are given in the online supplemental appendix.

We focused on local-linear RD estimators with local-quadratic bias-correction, $\hat{\tau}^{\texttt{rbc}}_{\texttt{SRD}}(h_n, b_n)$, as discussed in Section 2. We investigated the empirical coverage and interval length of the following three competing 95% confidence intervals for a variety of possible bandwidth choices:

$$\begin{aligned}
\text{“Conventional” } (\hat{T}_{\texttt{SRD}}(h_n)) &\quad : \quad \hat{\tau}_{\texttt{SRD}}(h_n) \pm 1.96 \cdot \sqrt{\hat{\mathsf{V}}_{\texttt{SRD}}(h_n)}, \\
\text{“Bias-Corrected” } (\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n)) &\quad : \quad \hat{\tau}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n) \pm 1.96 \cdot \sqrt{\hat{\mathsf{V}}_{\texttt{SRD}}(h_n)}, \\
\text{“Robust Approach” } (\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_n, b_n)) &\quad : \quad \hat{\tau}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n) \pm 1.96 \cdot \sqrt{\hat{\mathsf{V}}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n)},
\end{aligned}$$

where the estimators $\hat{\mathsf{V}}_{\texttt{SRD}}(h_n)$ and $\hat{\mathsf{V}}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n)$ are constructed using the nearest-neighbor procedure discussed in Section 5 with $J = 3$. For comparison, we also report the infeasible versions of these confidence intervals employing $\mathsf{V}_{\texttt{SRD}}(h_n)$ and $\mathsf{V}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n)$.

To choose the main bandwidth h_n we consider the following alternatives: (i) the infeasible MSE-optimal choice $h_{\texttt{MSE},0,1}$, denoted $h_{\texttt{MSE}}$; (ii) a plug-in, regularized MSE-optimal selector proposed by IK, denoted $\hat{h}_{\texttt{IK}}$; (iii) the infeasible, each-side-squared MSE-optimal choice proposed by DesJardins and McCall (2009), denoted $h_{\texttt{DM}}$; (iv) a plug-in, each-side-squared MSE-optimal selector, denoted $\hat{h}_{\texttt{DM}}$; (v) a cross-validation estimator proposed by Ludwig and Miller (2007), denoted $\hat{h}_{\texttt{CV}}$; and (vi) our plug-in choice proposed in Section 4, denoted $\hat{h}_{\texttt{CCT}}$. Similarly, to choose the pilot bandwidth b_n, we constructed the appropriately modified versions of the choices enumerated above, with the exception of $\hat{h}_{\texttt{CV}}$ because it is not available for derivative estimation; these choices are denoted $b_{\texttt{MSE}}$, $\hat{b}_{\texttt{IK}}$, $b_{\texttt{DM}}$, $\hat{b}_{\texttt{DM}}$, and $\hat{b}_{\texttt{CCT}}$, respectively. The online supplemental appendix provides a detailed description of each of these procedures.

Tables 2–3 present the main results. Table 2 employs the infeasible standard-errors based on $\mathsf{V}_{\texttt{SRD}}(h_n)$ and $\mathsf{V}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n)$, while Table 3 employs the fully data-driven standard-errors $\hat{\mathsf{V}}_{\texttt{SRD}}(h_n)$ and $\hat{\mathsf{V}}^{\texttt{bc}}_{\texttt{SRD}}(h_n, b_n)$. The simulation results across both tables are qualitative very similar but, as expected, the feasible versions of the 95% confidence intervals exhibit slightly more empirical coverage distortion and longer intervals on average. In the online supplemental appendix, we also report results employing the traditional standard-error estimators constructed using plug-in estimated residuals (also mentioned in Section 5), which lead to even more undercoverage in our simulations. In all cases, the robust standard-error estimators lead to important improvements in empirical coverage with only moderate increments in the average empirical length of the resulting confidence intervals. The choice $\rho_n = 1$ is not only simple and intuitive, but also performed well in our simulation setup. In terms of actual results, these tables suggest that the empirical coverage of intervals based on $T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_n, b_n)$ exhibits an improvement of about 10-15 percentage points on average, depending on the particular data-driven bandwidths employed. Although not the main goal of this paper, we also found that our two-stage direct plug-in rule selector of h_n performs very well relative to the other plug-in selectors, and on par with the cross-validation bandwidth selector.

In sum, based on our theoretical results and the simulation evidence presented, we recommend employing the new robust standard-error estimates introduced in this paper when constructing confidence intervals for treatment effects in the RD design.

7 Empirical Illustration

We illustrate the performance of our methods and compare them to other conventional alternatives employing household data from Oportunidades (formerly known as Progresa), a well-known large-scale anti-poverty conditional cash transfer program in Mexico. Our goal is to show how the different methods perform in a substantive, realistic empirical application. Most details regarding data construction and implementation, as well as other results not reported in this section, are given in the online supplemental appendix to conserve space. All estimates and figures were constructed using the STATA package `rdrobust`, described in Calonico, Cattaneo, and Titiunik (2013).

Progresa/Oportunidades was first instituted in rural communities in 1998, and later was expanded to urban areas in 2003. This social program is best known for its experimental component: treatment was initially randomly assigned at the locality level in rural areas.[8] Indeed, its experimental features have spiked a huge body of work focusing on a variety of economic, health and related outcomes.[9] In order to target the program to poor households in both rural and urban areas, Mexican officials constructed a pre-intervention (at baseline) household poverty-index that determined each household's eligibility. In rural communities, seven distinct poverty cutoffs were used depending on the geographic area, while one common cutoff was used in all urban localities. Thus, Progresa/Oportunidades' eligibility assignment rule naturally leads to eight sharp (intention-to-treat) regression-discontinuity designs. Buddelmeyer and Skoufias (2004) were the first to note the RD features of this social program.

We illustrate our methods employing data from the urban RD design and one of the seven rural RD designs (the one corresponding to the median household population size, Region 3, Sierra-Negra-Zongolica-Mazateca). We do not pool the RD designs, nor we compare them with each other or to the experimental estimates from the rural areas, since without further (strong) assumptions the

[8] In urban areas, however, treatment was not randomly assigned.

[9] Recent examples include Attanasio, Meghir, and Santiago (2011), Behrman, Gallardo-García, Parker, Todd, and Vélez-Grajales (2012), Djebbari and Smith (2008), Dubois, de Janvry, and Sadoulet (2012), Fernald, Gertler, and Neufeld (2009), among many others. These papers also include references to early reviews and research work.

associated estimands need not to coincide with each other. Instead, we treat the RD designs as different examples, which vary in observable, and possibly unobservable, characteristics.

Our empirical exercise investigates the program treatment effect on two mutually exclusive measures of household consumption expenditures: food and non-food consumption.[10] Related literature on this topic include Hoddinott and Skoufias (2004), Angelucci and Attanasio (2009), Angelucci and De Giorgi (2009), Gertler, Martinez, and Rubio-Codina (2012) and Angelucci and Attanasio (2013), who have also investigated the effect of Oportunidades/Progresa on consumption using experimental methods (in rural areas) and non-experimental matching methods (in urban areas). Our illustrative results therefore contribute to this literature by presenting new empirical evidence based on non-experimental RD estimates. In this application, X_i denotes the household's poverty-index, $\bar{x} = 0$ denotes the centered cutoff for each RD design, and Y_i denotes the two different measures of household consumption.

Our final database contains 691 control households ($X_i < 0$) and $2,118$ intention-to-treat households ($X_i \geq 0$) in the urban RD design ($n = 2,809$, $X_i \in [-2.25\ ,\ 4.11]$), and 315 control households ($X_i < 0$) and 618 intention-to-treat households ($X_i \geq 0$) in the rural RD design ($n = 933$, $X_i \in [-456.6\ ,\ 338.4]$). In the online supplemental appendix, we address the empirical validity of these RD designs by conducting standard balance and falsification tests on pre-intervention covariates. These results give empirical support for the RD assumptions. Figures 2 and 3 present, respectively, the usual RD plots for the urban and rural areas (c.f. Figure 1). In these figures, the solid lines correspond to distinct fourth-order global polynomial fits for control and treatments units, and the solid dots correspond to sample averages of the outcome variable for each bin (or partition) of the running variable. The number of bins was chosen using an integrated mean-square error formula derived in Cattaneo and Farrell (2013), as explained in Calonico, Cattaneo, and Titiunik (2013, Section 2.7).

Our main empirical results are reported in Table 4. Panel A and B correspond, respectively, to the urban and rural RD designs. We consider three time periods: pre-intervention (as a falsification test), one year after the program started (1-year Treatment), and two years after the program started (2-year Treatment). Thus, each panel reports six groups of RD estimates (i.e., 2 outcomes $\times$ 3 periods). For each combination of outcome and time period, we conduct RD estimation and inference employing the same setup as in our simulation study: local-linear estimator of $\tau_{\texttt{SRD}}$, conventional confidence interval and robust confidence interval (with local-quadratic bias-correction), each implemented with the three different data-driven bandwidth choices $\hat{h}_{\texttt{CCT}}$, $\hat{h}_{\texttt{IK}}$ and $\hat{h}_{\texttt{CV}}$. To be specific, for each panel, outcome, period and bandwidth selection method we report $\hat{\tau}_{\texttt{SRD}}(\hat{h}_n)$, $\hat{I}_{\texttt{SRD}}(\hat{h}_n)$, $\hat{I}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_n, \hat{b}_n)$, $\hat{h}_n$ and $\hat{b}_n$.

This empirical exercise offers an array of interesting examples to discuss the performance of our proposed methods. First of all, using the pre-intervention data (columns 1–3, Panels A and B), we find no effects of the program in any case (i.e., food or non-food consumption in urban or rural localities).[11] This result gives additional evidence in favor of the validity of the RD designs, since

[10] In the online supplemental appendix we also examine total consumption expenditures for all the RD designs.

[11] In rural areas, pre and post-intervention food consumption data differs in two main aspects. First, the pre-

households in control and treatment areas exhibit on average the same levels of pre-intervention consumption. In the 1-year after treatment data, we find statistically significant effects of the program on food consumption in rural areas (columns 4–6, Panel B). This result is present in all cases when using both the conventional as well as the robust confidence intervals. On the other hand, in the same period, we find no statistically significant effects on non-food consumption in rural areas (columns 4–6, Panel B) nor on any of the outcomes in urban areas (columns 4–6, Panel A). These results are consistent across inference procedures.

The results from the 2-year after treatment data are the most interesting. In this case, for food consumption in urban areas (columns 7–9, Panel A), we find statistically significant results when using the conventional confidence intervals but these results are not statistically significant when using the robust confidence intervals proposed in this paper. This empirical example offers an instance where the conventional inference approach suggests the presence of a strong positive treatment effect, but our methods cast doubt on such a conclusion. On the other hand, when examining non-food consumption in urban areas (still columns 7–9, Panel A) the results appear to be more robust, as they are statistically significant at standard levels when using both the conventional and the robust confidence intervals. Finally, in the case of the rural RD design (columns 7–9, Panel B), we find no statistically significant effects on food consumption using either method, but we find a statistically significant (10-percent level) treatment effect on non-food consumption when using conventional confidence intervals. The latter result, however, is not particularly robust based on our proposed confidence intervals.

To summarize, the findings from the small empirical illustration suggest that the program Progresa/Oportunidades had (*i*) a positive, significant effect on non-food consumption in urban areas two years after its introduction, and (*ii*) a positive, significant effect on food consumption in rural areas one year after its introduction. Both results appear to be robust according to our proposed methods. In addition, the empirical findings using conventional methods suggest that the program had positive, significant effects on food consumption in urban areas and on non-food consumption in rural areas two years after its introduction, but these findings are not robust according to our proposed inference procedures.

8 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 simu-

intervention survey only provides information on expenditures (that is, it omits home production). Second, it reports expenditures only by food groups rather than asking detailed item-by-item questions, as in later waves. See, e.g., Angelucci and De Giorgi (2009) for further details.

lations, 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. In addition, in an empirical illustration addressing the program effects of Progresa/Oportunidades on consumption, we found that in some cases the conventional confidence intervals offer results that are validated by our proposed confidence intervals, while in other cases our robust confidence intervals cast doubt on the results obtained by conventional methods.

A 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. All the results stated in this Appendix, as well as other results discussed in the main text, are proven in the online supplemental appendix.

A.1 Assumptions

We employ the following conventional assumption on the basic sharp RD model.

Assumption A1. For some $\kappa_0 > 0$, the following holds in the neighborhood $(-\kappa_0, \kappa_0)$ around the cutoff $\bar{x} = 0$:
(a) $\mathbb{E}[Y_i^4|X_i] < \infty$.
(b) $f(x)$, the density of X_i, is continuous and bounded away from zero.
(c) For some $S \geq 1$, $\mu_-(x) = \mathbb{E}[Y_i(0)|X_i = x]$ and $\mu_+(x) = \mathbb{E}[Y_i(1)|X_i = x]$ are S-times continuously differentiable.
(d) $\sigma^2(x) = \mathbb{V}[Y_i|X_i = x]$ is bounded away from zero, bounded, and right and left continuous at $x = 0$.

Part (a) in Assumption 1 imposes existence of moments. Part (b) requires that the running variable X_i be continuously distributed near the cutoff, and also ensures the presence of observations arbitrarily close to the cutoff in large samples. Part (c) 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. Finally, part (d) imposes standard restrictions on the conditional variance of the observed outcome, but allows it to be potentially different at either side of the threshold. Thus, we set $\sigma_+^2 = \lim_{x \to 0^+} \sigma^2(x)$ and $\sigma_-^2 = \lim_{x \to 0^-} \sigma^2(x)$.

Throughout the paper we employed local polynomial regression estimators of various orders to approximate unknown regression functions. The following standard assumption on the kernel function is used to construct the estimators.

Assumption A2. For some $\kappa > 0$, the kernel function $k(\cdot) : [0, \kappa] \mapsto \mathbb{R}$ is bounded and nonnegative on $[0, \kappa]$, positive and continuous on $(0, \kappa)$, and zero outside its support.

This assumption permits all kernels commonly used in empirical work. Although our results extend to the case where possibly different kernels are used at either side of the threshold, to simplify the exposition we set $K(u) = k(-u) \cdot \mathbf{1}(u < 0) + k(u) \cdot \mathbf{1}(u \geq 0)$, implying that, for $\kappa > 0$ given in Assumption A2, $K(\cdot)$ is symmetric, bounded and nonnegative on $[-\kappa, \kappa]$, positive and continuous on $(-\kappa, \kappa)$, and zero outside its support.

Finally, to handle the fuzzy RD designs we impose the following additional assumption.

Assumption A3. For some $\kappa_0 > 0$, the following holds in the neighborhood $(-\kappa_0, \kappa_0)$ around the cutoff $\bar{x} = 0$:
(a) For some $S \geq 1$, $\mu_{T-}(x) = \mathbb{E}[T_i(0)|X_i = x]$ and $\mu_{T+}(x) = \mathbb{E}[T_i(1)|X_i = x]$ are S-times continuously differentiable.
(b) $\sigma_T^2(x) = \mathbb{V}[T_i|X_i = x]$ is bounded away from zero and right and left continuous at $x = 0$.

This assumption is analogous to Assumption A1, but involving as outcome variable the treatment assignment and treatment status for each unit.

A.2 Local Polynomial Estimators

For any $\nu, p \in \mathbb{N}$ with $\nu \leq p$, the p-th order local polynomial estimators of the ν-th order derivatives $\mu_+^{(\nu)}$ and $\mu_-^{(\nu)}$ are given by

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

with

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

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

where $r_p(x) = (1, x, \cdots, x^p)'$, 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.) For example, $\hat{\beta}_{+,1}(h_n)$ is a bivariate vector containing the two coefficients (intercept and slope) obtained from using weighted least-squares to estimate a linear model employing only observations with X_i to the right and near the discontinuity $\bar{x} = 0$ (i.e., for observations with $X_i \in [0, \kappa h_n]$). Similarly, $\hat{\beta}_{+,2}(h_n)$ and $\hat{\beta}_{+,3}(h_n)$ correspond, respectively, to local-quadratic and local-cubic regression coefficient vectors using the same observations.

It is well-known that, under appropriate regularity conditions and bandwidth restrictions, $\hat{\beta}_{+,p}(h_n) \to_p \beta_{+,p} = (\mu_+, \mu_+^{(1)}/1!, \mu_+^{(2)}/2!, \cdots, \mu_+^{(p)}/p!)'$ and $\hat{\beta}_{-,p}(h_n) \to_p \beta_{-,p} = (\mu_-, \mu_-^{(1)}/1!, \mu_-^{(2)}/2!, \cdots, \mu_-^{(p)}/p!)'$, 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). In the sequel, we set $\hat{\mu}_{+,p}(h_n) = \hat{\mu}_{+,p}^{(0)}(h_n)$ and $\hat{\mu}_{+,p}(h_n) = \hat{\mu}_{+,p}^{(0)}(h_n)$ to improve notation.

Note that, whenever possible, we drop the outcome variable subindex notation from $\hat{\mu}_{Y+,p}^{(\nu)}(h_n)$, $\hat{\mu}_{Y-,p}^{(\nu)}(h_n)$, $\hat{\beta}_{Y+,p}(h_n)$, $\hat{\beta}_{Y-,p}(h_n)$, etc.

A.3 Further Notation

We employ the following notation: $Y = [Y_1, \cdots, Y_n]'$, $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_n]'$ with $\varepsilon_i = Y_i - \mu(X_i)$, $\mathcal{X}_n = [X_1, \cdots, X_n]'$, and $\Sigma = \mathbb{E}[\varepsilon\varepsilon'|\mathcal{X}_n] = \text{diag}(\sigma^2(X_1), \cdots, \sigma^2(X_n))$ where $\text{diag}(a_1, ..., a_n)$ denotes the $(n \times n)$ diagonal matrix with diagonal elements $a_1, ..., a_n$.

We also set:

$$X_p(h) = [r_p(X_1/h), \cdots, r_p(X_n/h)]', \quad S_p(h) = [(X_1/h)^p, \cdots, (X_n/h)^p]',$$

$$W_+(h) = \text{diag}(\mathbf{1}(X_1 \geq 0)K_h(X_1), \cdots, \mathbf{1}(X_n \geq 0)K_h(X_n)),$$

$$W_-(h) = \text{diag}(\mathbf{1}(X_1 < 0)K_h(X_1), \cdots, \mathbf{1}(X_n < 0)K_h(X_n)).$$

In addition, we define the following (scaled) matrices

$$\Gamma_{+,p}(h) = X_p(h)'W_+(h)X_p(h)/n, \quad \Gamma_{-,p}(h) = X_p(h)'W_-(h)X_p(h)/n,$$

$$\vartheta_{+,p,q}(h) = X_p(h)'W_+(h)S_q(h)/n, \quad \vartheta_{-,p,q}(h) = X_p(h)'W_-(h)S_q(h)/n,$$

$$\Psi_{+,p,q}(h,b) = X_p(h)'W_+(h)\Sigma W_+(b)X_q(b)/n, \quad \Psi_{-,p,q}(h,b) = X_p(h)'W_-(h)\Sigma W_-(b)X_q(b)/n,$$

where we set for brevity $\Psi_{+,p}(h) = \Psi_{+,p,p}(h,h)$ and $\Psi_{-,p}(h) = \Psi_{-,p,p}(h,h)$.

We will also use repeatedly the large sample matrices

$$\Gamma_p = \int_0^\infty K(u)r_p(u)r_p(u)'\mathrm{d}u, \quad \vartheta_{p,q} = \int_0^\infty K(u)u^q r_p(u)\mathrm{d}u, \quad \Psi_p = \int_0^\infty K(u)^2 r_p(u)r_p(u)'\mathrm{d}u.$$

Letting $H_p(h) = \text{diag}(1, h^{-1}, \cdots, h^{-p})$, it follows that

$$\hat{\beta}_{+,p}(h_n) = H_p(h_n)\Gamma_{+,p}^{-1}(h_n)X_p(h_n)'W_+(h_n)Y/n,$$

$$\hat{\beta}_{-,p}(h_n) = H_p(h_n)\Gamma_{+,p}^{-1}(h_n)X_p(h_n)'W_+(h_n)Y/n.$$

Finally, recall that in the fuzzy designs we add a further subindex to denote the underlying outcome(s) used whenever appropriate. That is, for random variables U and V, we set $\Sigma_{UV} = \text{diag}(\sigma^2_{UV}(X_1), \cdots, \sigma^2_{UV}(X_n))$ with $\sigma^2_{UV}(x) = \mathbb{C}ov[U, V|X = x] = \mathbb{E}[(U - \mathbb{E}[U|X])(V - \mathbb{E}[V|X])|X = x]$, and similarly for other parameters (and also the estimators).

A.4 Sharp RD Designs

As in the main text, in this section we drop the notational dependence on the outcome variable Y. The general estimand of interest in the sharp RD design is

$$\tau_\nu = \mu_+^{(\nu)} - \mu_-^{(\nu)}, \qquad \mu_+^{(\nu)} = \nu! e_\nu' \beta_{+,p}, \qquad \mu_-^{(\nu)} = \nu! e_\nu' \beta_{-,p} \qquad\qquad (\nu \leq p),$$

and recall that $\tau_{\mathtt{SRD}} = \tau_0$ and $\tau_{\mathtt{SKRD}} = \tau_1$.

In this context, for any $\nu \leq p$, the conventional p-th order local polynomial RD estimator is

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

and recall that $\hat{\tau}_{\mathtt{SRD}}(h_n) = \hat{\tau}_{0,1}(h_n)$ and $\hat{\tau}_{\mathtt{SKRD}}(h_n) = \hat{\tau}_{1,2}(h_n)$.

A.4.1 Lemma A1

This lemma describes the asymptotic bias, variance and distribution of $\hat{\tau}_{\nu,p}(h_n)$. This result follows from known results in the local polynomial literature applied to the RD context (e.g., Fan and Gijbels (1996)).

Lemma A1. *Suppose Assumptions A1–A2 hold with $S \geq p + 2$. Let $\nu, r \in \mathbb{N}$ with $\nu \leq p$.*

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

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

where

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

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

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

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

$$\mathsf{V}_{\nu,p}(h_n) = \mathcal{V}_{+,\nu,p}(h_n) + \mathcal{V}_{-,\nu,p}(h_n),$$

$$\mathcal{V}_{+,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}} \nu!^2 e_\nu' \Gamma_{+,p}^{-1}(h_n) \Psi_{+,p}(h_n) \Gamma_{+,p}^{-1}(h_n) e_\nu = \frac{1}{nh_n^{1+2\nu}} \frac{\sigma_+^2}{f} \nu!^2 e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu [1 + o_p(1)],$$

$$\mathcal{V}_{-,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}} \nu!^2 e_\nu' \Gamma_{-,p}^{-1}(h_n) \Psi_{-,p}(h_n) \Gamma_{-,p}^{-1}(h_n) e_\nu = \frac{1}{nh_n^{1+2\nu}} \frac{\sigma_-^2}{f} \nu!^2 e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu [1 + o_p(1)].$$

(D) *If $nh_n^{2p+5} \to 0$ and $nh_n \to \infty$, then*

$$\frac{\hat{\tau}_{\nu,p}(h_n) - \tau_\nu - h_n^{p+1-\nu} \mathsf{B}_{\nu,p,p+1}(h_n)}{\sqrt{\mathsf{V}_{\nu,p}(h_n)}} \to_d \mathcal{N}(0, 1).$$

Therefore, for any $p < q$, the q-th order local polynomial bias-corrected estimator is

$$\hat{\tau}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n) = \hat{\tau}_p(h_n) - h_n^{p+1}\hat{\mathsf{B}}_{\nu,p,q}(h_n, b_n),$$

with

$$\hat{\mathsf{B}}_{\nu,p,q}(h_n, b_n) = (e'_{p+1}\hat{\beta}_{+,q}(b_n))\mathcal{B}_{+,\nu,p,p+1}(h_n) - (e'_{p+1}\hat{\beta}_{-,q}(b_n))\mathcal{B}_{-,\nu,p,p+1}(h_n).$$

A.4.2 Theorem A1

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

Theorem A1. Suppose Assumptions A1–A2 hold with $S \geq q + 1$ and $q \geq p + 1$. Let $\nu \in \mathbb{N}$ with $\nu \leq p$.

(B) If $\max\{h_n, b_n\} \to 0$ and $n \min\{h_n, b_n\} \to \infty$, then

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

where

$$\mathsf{B}^{\mathsf{bc}}_{\nu,p,q}(h, b) = \frac{\mu_+^{(q+1)}}{(q+1)!}\mathcal{B}_{+,p+1,q,q+1}(b)\frac{\mathcal{B}_{+,\nu,p,p+1}(h)}{(p+1)!} - \frac{\mu_-^{(q+1)}}{(q+1)!}\mathcal{B}_{-,p+1,q,q+1}(b)\frac{\mathcal{B}_{-,\nu,p,p+1}(h)}{(p+1)!}.$$

(V) If $n \min\{h_n, b_n\} \to \infty$, then $\mathbb{V}[\hat{\tau}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n)|\mathcal{X}_n] = \mathsf{V}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n)$, where

$$\mathsf{V}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n) = \mathcal{V}^{\mathsf{bc}}_{+,\nu,p,q}(h_n, b_n) + \mathcal{V}^{\mathsf{bc}}_{-,\nu,p,q}(h_n, b_n),$$

$$\mathcal{V}^{\mathsf{bc}}_{+,\nu,p,q}(h, b) = \mathcal{V}_{+,\nu,p}(h) - 2h^{p+1-\nu}\mathcal{C}_{+,\nu,p,q}(h, b)\frac{\mathcal{B}_{+,\nu,p,p+1}(h)}{(p+1)!} + h^{2(p+1-\nu)}\mathcal{V}_{+,p+1,q}(b)\frac{\mathcal{B}^2_{+,\nu,p,p+1}(h)}{(p+1)!^2},$$

$$\mathcal{V}^{\mathsf{bc}}_{-,\nu,p,q}(h, b) = \mathcal{V}_{-,\nu,p}(h) - 2h^{p+1-\nu}\mathcal{C}_{-,\nu,p,q}(h, b)\frac{\mathcal{B}_{-,\nu,p,p+1}(h)}{(p+1)!} + h^{2(p+1-\nu)}\mathcal{V}_{-,p+1,q}(b)\frac{\mathcal{B}^2_{-,\nu,p,p+1}(h)}{(p+1)!^2},$$

$$\mathcal{C}_{+,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}}\nu!(p+1)!e'_\nu\Gamma^{-1}_{+,p}(h)\Psi_{+,p,q}(h, b)\Gamma^{-1}_{+,q}(b)e_{p+1},$$

$$\mathcal{C}_{-,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}}\nu!(p+1)!e'_\nu\Gamma^{-1}_{-,p}(h)\Psi_{-,p,q}(h, b)\Gamma^{-1}_{-,q}(b)e_{p+1},$$

for $\dim(e_0) = p$ and $\dim(e_{p+1}) = q$.

(D) If $n \min\{h_n^{2p+3}, b_n^{2p+3}\} \max\{h_n^2, b_n^{2(q-p)}\} \to 0$ and $n \min\{h_n, b_n\} \to \infty$, then

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

provided $\kappa \max\{h_n, b_n\} < \kappa_0$.

From this theorem we obtain for Theorems 1 and 2:

$$\mathsf{V}^{\mathsf{bc}}_{\mathsf{SRD}}(h_n, b_n) = \mathsf{V}^{\mathsf{bc}}_{0,1,2}(h_n, b_n) \qquad \text{and} \qquad \mathsf{V}^{\mathsf{bc}}_{\mathsf{SKRD}}(h_n, b_n) = \mathsf{V}^{\mathsf{bc}}_{1,2,3}(h_n, b_n). \tag{A-1}$$

Remark A1. Remark 1 in the main text remains true: the distributional approximation in Theorem A1 permits one bandwidth (but not both) to be fixed, provided it is not too "large"; i.e., both must satisfy $\kappa \max\{h_n, b_n\} < \kappa_0$, but only one needs to vanish.

Remark A2. Remark 2 in the main text generalizes as follows. Three main limiting cases are obtained depending on the limit $\rho_n \to \rho \in [0, \infty]$.

Case 1: $\rho = 0$. In this case $h_n = o(b_n)$ and

$$\mathbb{V}[\hat{\tau}^{\mathtt{bc}}_{\nu,p,q}(h_n, b_n)|\mathcal{X}_n] = \mathbb{V}[\hat{\tau}_{\nu,p}(h_n)|\mathcal{X}_n]\{1 + o_p(1)\} = \frac{1}{nh_n^{1+2\nu}} \frac{\sigma_+^2 + \sigma_-^2}{f} (e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_0)\{1 + o_p(1)\},$$

which is the classical approach to bias-correction.

Case 2: $\rho \in (0, \infty)$. In this case $h_n = \rho b_n$ and

$$\begin{aligned}
&\mathbb{V}[\hat{\tau}^{\mathtt{bc}}_{p,q}(h_n, b_n)|\mathcal{X}_n] \\
&= \frac{1}{nh_n^{1+2\nu}} \left[\frac{\sigma_+^2 + \sigma_-^2}{f} (e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_0) + \rho^{2p+3} \frac{\sigma_+^2 + \sigma_-^2}{f} (e_p' \Gamma_q^{-1} \Psi_q \Gamma_q^{-1} e_{p+1})(e_\nu' \Gamma_p^{-1} \vartheta_{p,p+1})^2 \right. \\
&\qquad \left. - \rho^{p+2} \left(e_\nu' \Gamma_p^{-1} \left(\frac{\sigma_+^2}{f} \Psi_{p,q}(\rho) + \frac{\sigma_-^2}{f} \Psi_{p,q}(-\rho) \right) \Gamma_q^{-1} e_{p+1} \right) (e_\nu' \Gamma_p^{-1} \vartheta_p) \right] \{1 + o_p(1)\},
\end{aligned}$$

with $\Psi_{p,q}(\rho) = \int_0^\infty K(u) K(\rho u) r_p(u) r_q(\rho u)' \mathrm{d}u$. For conventional choices of kernel $K(\cdot)$, the limiting variance is increasing in ρ.

Case 3: $\rho = \infty$. In this case $b_n = o(h_n)$ and

$$\begin{aligned}
\mathbb{V}[\hat{\tau}^{\mathtt{bc}}_{p,q}(h_n, b_n)|\mathcal{X}_n] &= h_n^{2(p+1-\nu)} \mathbb{V}[\hat{\mathsf{B}}_{\nu,p,q}(h_n, b_n)|\mathcal{X}_n]\{1 + o_p(1)\} \\
&= \frac{\rho_n^{2(p+1-\nu)}}{nb_n^{1+2\nu}} \frac{\sigma_+^2 + \sigma_-^2}{f} (e_{p+1}' \Gamma_q^{-1} \Psi_q \Gamma_q^{-1} e_{p+1})(e_\nu' \Gamma_p^{-1} \vartheta_{p,p+1})^2 \{1 + o_p(1)\},
\end{aligned}$$

which implies that the bias-estimate is first-order while the actual estimator $\hat{\tau}_p(h_n)$ is of smaller order.

Remark A3. If $h_n = b_n$ (and the same kernel function $K(\cdot)$ is used), then $\hat{\tau}^{\mathtt{bc}}_{\nu,p,p+1}(h_n, h_n) = \hat{\tau}_{\nu,p+1}(h_n)$. This gives a simple relationship between local polynomial estimators of order p and $p + 1$, and their relation to manual bias-correction. This implies that $T^{\mathtt{rbc}}_{\nu,p,p+1}(h_n, h_n) = T_{\nu,p+1}(h_n)$. The result extends to $\hat{\tau}^{\mathtt{bc}}_{\nu,p,p+r}(h_n, h_n) = \hat{\tau}_{\nu,p+r}(h_n)$ and $T^{\mathtt{rbc}}_{\nu,p,p+r}(h_n, h_n) = T_{\nu,p+r}(h_n)$ when the natural generalization of the bias-correction estimate is used. See the supplemental appendix for details.

Remark A4. It is well known that bias-correction can be seen as another way of undersmoothing the original estimator. An interesting implication of Remark 3 is that our approach provides a formalization of this idea. In particular, it justifies a simple approach based on the order of the local polynomial: a systematic choice of h_n that leads to undersmoothing is to select the MSE-optimal bandwidth for the estimator $\hat{\tau}_p(h_n)$, but construct confidence intervals using the estimator $\hat{\tau}_{p+1}(h_n)$. This is the special case $\rho_n = h_n/b_n = 1$ in Theorem 1.

Remark A5. The previous results can be described using the *Equivalent Kernel Representation* of local polynomials (e.g., Fan and Gijbels (1996, Section 3.2.2)). For simplicity, consider the one-sided bias-corrected estimate of μ_+: $\hat{\tau}^{\mathtt{bc}}_{+,0,p,q}(h_n, b_n) = \hat{\mu}_{+,p}(h_n) - h_n^{p+1}(e_{p+1}' \hat{\beta}_{+,q}(b_n))\mathcal{B}_{+,0,p,p+1}(h_n)$. Letting $h_n = \rho b_n$ with $\rho \in (0, \infty)$,

$$\hat{\tau}^{\mathtt{bc}}_{+,0,p,q}(h_n, b_n) = \frac{1}{nh_n f} \sum_{i=1}^n \mathbf{1}(X_i \geq 0) \mathcal{K}_{p,q}\left(\frac{X_i}{h_n}; \rho \right) Y_i \ \{1 + o_p(1)\},$$

$$\mathcal{K}_{p,q}(x; \rho) = \mathcal{K}_p(x) - \rho^{p+2} \mathcal{K}^{bc}_{p,q}(\rho x),$$

where $\mathcal{K}_p(x) = e_0' \Gamma_p^{-1} r_p(x) K(x)$ is the equivalent kernel of the local polynomial estimator $\hat{\mu}_{+,p}(h_n)$, and $\mathcal{K}^{\mathtt{bc}}_{p,q}(x) = (e_{p+1}' \Gamma_q^{-1} r_q(x))(e_0' \Gamma_p^{-1} \vartheta_{p,p+1}) K(x)$ is the equivalent kernel induced by the bias-correction estimate $(e_{p+1}' \hat{\beta}_{+,q}(b_n))\mathcal{B}_{+,0,p,p+1}(h_n)$.

(i) Because $h_n = \rho b_n$, the asymptotics in Theorem A1 "convexify" the kernel function employed to construct the estimator $\hat{\tau}^{\mathtt{bc}}_{+,0,p,q}(h_n, b_n)$, because

$$\lim_{\rho\to 0^+} \mathcal{K}_{p,q}(x;\rho) = \mathcal{K}_p(x) \qquad \text{and} \qquad \mathcal{K}_{p,p+1}(x;1) = \mathcal{K}_{p+1}(x).$$

(ii) The asymptotic bias and variance reduce to

$$\mathbb{E}[\hat{\tau}^{\mathtt{bc}}_{+,0,p,q}(h_n, h_n/\rho)|\mathcal{X}_n] = \mu_+ + h_n^{p+2}\frac{\mu_+^{(p+2)}}{(p+2)!}\mathfrak{B}_p(\rho)\ \{1+o_p(1)\},$$

$$\mathbb{V}[\hat{\tau}^{\mathtt{bc}}_{+,0,p,q}(h_n, h_n/\rho)|\mathcal{X}_n] = \frac{1}{nh_n}\frac{\sigma_+^2}{f}\mathfrak{V}_p(\rho)\ \{1+o_p(1)\},$$

$$\mathfrak{B}_p(\rho) = \int_0^\infty x^{p+2}\mathcal{K}_{p,p+1}(x;\rho)\,\mathrm{d}x, \qquad \mathfrak{V}_p(\rho) = \int_0^\infty (\mathcal{K}_{p,p+1}(x;\rho))^2\mathrm{d}x,$$

For conventional choices of kernel $K(\cdot)$, $\mathfrak{B}_p(\rho)$ is decreasing and $\mathfrak{V}_p(\rho)$ is increasing in ρ. See the supplemental appendix for further details.

Remark A6. Cheng, Fan, and Marron (1997) study the optimal choice of boundary kernel of order p in a conditional MSE minimax sense for one-sided nonparametric regression estimation at a boundary point. Although not the focus of this paper, from this point estimation perspective, the induced equivalent kernel $\mathcal{K}_{p-1,p}(x;\rho)$ dominates $\mathcal{K}_p(x)$ for an appropriate choice of ρ, when a conventional kernel $K(\cdot)$ is used. (For $\rho > 0$, $\mathcal{K}_{p-1,p}(x;\rho)$ is also a boundary kernel of order p or larger.) See the supplemental appendix for further details.

A.5 Fuzzy RD Designs

Here we consider the ν-th fuzzy RD estimand

$$\varsigma_\nu = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}}, \qquad \tau_{Y,\nu} = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}, \qquad \tau_{T,\nu} = \mu_{T+}^{(\nu)} - \mu_{T-}^{(\nu)},$$

provided that $\nu \leq S$. Note that $\tau_{\mathtt{FRD}} = \varsigma_0$ and $\tau_{\mathtt{FKRD}} = \varsigma_1$.

The fuzzy RD estimator based on the p-th order local polynomial estimators $\hat{\tau}_{Y,\nu,p}(h_n)$ and $\hat{\tau}_{T,\nu,p}(h_n)$ therefore is

$$\hat{\varsigma}_{\nu,p}(h_n) = \frac{\hat{\tau}_{Y,\nu,p}(h_n)}{\hat{\tau}_{T,\nu,p}(h_n)}, \qquad \hat{\tau}_{Y,\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{Y+,p}(h_n) - \hat{\mu}^{(\nu)}_{Y-,p}(h_n), \qquad \hat{\tau}_{T,\nu,p}(h_n) = \hat{\mu}^{(\nu)}_{T+,p}(h_n) - \hat{\mu}^{(\nu)}_{T-,p}(h_n),$$

with

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

$$\hat{\mu}^{(\nu)}_{T+,p}(h_n) = \nu! e_\nu' \hat{\beta}_{T+,p}(h_n), \qquad \hat{\mu}^{(\nu)}_{T-,p}(h_n) = \nu! e_\nu' \hat{\beta}_{T-,p}(h_n),$$

and with the notation

$$\hat{\beta}_{Y+,p}(h_n) = H_p(h_n)\Gamma^{-1}_{+,p}(h_n)X_p(h_n)'W_+(h_n)Y/n, \qquad \hat{\beta}_{Y-,p}(h_n) = H_p(h_n)\Gamma^{-1}_{-,p}(h_n)X_p(h_n)'W_-(h_n)Y/n,$$

$$\hat{\beta}_{T+,p}(h_n) = H_p(h_n)\Gamma^{-1}_{+,p}(h_n)X_p(h_n)'W_+(h_n)T/n, \qquad \hat{\beta}_{T-,p}(h_n) = H_p(h_n)\Gamma^{-1}_{-,p}(h_n)X_p(h_n)'W_-(h_n)T/n.$$

Note that $\hat{\tau}_{\mathtt{FRD}}(h_n) = \hat{\varsigma}_{0,1}(h_n)$ and $\hat{\tau}_{\mathtt{FKRD}}(h_n) = \hat{\varsigma}_{1,2}(h_n)$.

A.5.1 Lemma A2

This lemma gives an analogue of Lemma A1 for fuzzy designs. Using the expansion

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a}-a) - \frac{a}{b^2}(\hat{b}-b) + \frac{a}{b^2\hat{b}}(\hat{b}-b)^2 - \frac{1}{b\hat{b}}(\hat{a}-a)(\hat{b}-b)$$

we obtain

$$\hat{\varsigma}_{\nu,p}(h_n) - \varsigma_\nu = \tilde{\varsigma}_{\nu,p}(h_n) + R_n$$

with

$$\tilde{\varsigma}_{\nu,p}(h_n) = \frac{1}{\tau_{T,\nu}}(\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu}) - \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})$$

$$R_n = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2 \hat{\tau}_{T,\nu,p}(h_n)}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})^2 - \frac{1}{\tau_{T,\nu}\hat{\tau}_{T,\nu,p}(h_n)}(\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu})(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu}).$$

Lemma A2. Suppose Assumptions A1–A3 hold with $S \geq p+2$. Let $\nu, r \in \mathbb{N}$ with $\nu \leq p$.

(**R**) If $h_n \to 0$ and $nh_n^{1+2\nu} \to \infty$, then

$$R_n = O_p\left(\frac{1}{nh_n^{1+2\nu}} + h_n^{2(p+1-\nu)}\right).$$

(B) If $h_n \to 0$ and $nh_n \to \infty$, then

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

where

$$\mathsf{B}_{\mathtt{F},\nu,p,r}(h_n) = \frac{1}{\tau_{T,\nu}}\mathsf{B}_{Y,\nu,p,r}(h_n) - \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\mathsf{B}_{T,\nu,p,r}(h_n),$$

with

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

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

(V) If $h_n \to 0$ and $nh_n \to \infty$, then $\mathbb{V}[\tilde{\varsigma}_{\nu,p}(h_n)|\mathcal{X}_n] = \mathsf{V}_{\mathtt{F},\nu,p}(h_n)$, where

$$\mathsf{V}_{\mathtt{F},\nu,p}(h_n) = \mathsf{V}_{\mathtt{F},+,\nu,p}(h_n) + \mathsf{V}_{\mathtt{F},-,\nu,p}(h_n),$$

where

$$\mathsf{V}_{\mathtt{F},+,\nu,p}(h_n) = \frac{1}{\tau_{T,\nu}^2}\mathcal{V}_{YY+,\nu,p}(h_n) - \frac{2\tau_{Y,\nu}}{\tau_{T,\nu}^3}\mathcal{V}_{YT+,\nu,p}(h_n) + \frac{\tau_{Y,\nu}^2}{\tau_{T,\nu}^4}\mathcal{V}_{TT+,\nu,p}(h_n),$$

with

$$\mathcal{V}_{YY+,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu'\Gamma_{+,p}^{-1}(h_n)\Psi_{YY+,p}(h_n)\Gamma_{+,p}^{-1}(h_n)e_\nu = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{YY+}^2}{f}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu[1+o_p(1)],$$

$$\mathcal{V}_{YT+,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu'\Gamma_{+,p}^{-1}(h_n)\Psi_{YT+,p}(h_n)\Gamma_{+,p}^{-1}(h_n)e_\nu = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{YT+}^2}{f}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu[1+o_p(1)],$$

$$\mathcal{V}_{TT+,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu'\Gamma_{+,p}^{-1}(h_n)\Psi_{TT+,p}(h_n)\Gamma_{+,p}^{-1}(h_n)e_\nu = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{TT+}^2}{f}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu[1+o_p(1)],$$

and

$$\mathsf{V}_{\mathtt{F},-,\nu,p}(h_n) = \frac{1}{\tau_{T,\nu}^2}\mathcal{V}_{YY-,\nu,p}(h_n) - \frac{2\tau_{Y,\nu}}{\tau_{T,\nu}^3}\mathcal{V}_{YT-,\nu,p}(h_n) + \frac{\tau_{Y,\nu}^2}{\tau_{T,\nu}^4}\mathcal{V}_{TT-,\nu,p}(h_n)$$

with

$$\mathcal{V}_{YY-,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu'\Gamma_{-,p}^{-1}(h_n)\Psi_{YY-,p}(h_n)\Gamma_{-,p}^{-1}(h_n)e_\nu = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{YY-}^2}{f}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu[1+o_p(1)],$$

$$\mathcal{V}_{YT-,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu' \Gamma_{-,p}^{-1}(h_n)\Psi_{YT-,p}(h_n)\Gamma_{-,p}^{-1}(h_n)e_\nu = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{YT-}^2}{f}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu[1+o_p(1)],$$

$$\mathcal{V}_{TT-,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu' \Gamma_{-,p}^{-1}(h_n)\Psi_{TT-,p}(h_n)\Gamma_{-,p}^{-1}(h_n)e_\nu = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{TT-}^2}{f}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu[1+o_p(1)].$$

(D) If $nh_n^{2p+5} \to 0$ and $nh_n \to \infty$, then

$$\frac{\tilde{\varsigma}_{\nu,p}(h_n) - h_n^{p+1-\nu}\mathsf{B}_{\mathtt{F},\nu,p,p+1}(h_n)}{\sqrt{\mathsf{V}_{\mathtt{F},\nu,p}(h_n)}} \to_d \mathcal{N}(0,1).$$

A.5.2 Theorem A2

This theorem gives an analogue of Theorem A1 for the bias-corrected fuzzy RD estimator

$$\hat{\varsigma}_{\nu,p,q}^{\mathsf{bc}}(h_n,b_n) = \hat{\varsigma}_{\nu,p}(h_n) - h_n^{p+1-\nu}\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n),$$

with

$$\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) = \frac{1}{\hat{\tau}_{T,\nu,p}(h_n)}\left((e_{p+1}'\hat{\beta}_{Y+,q}(b_n))\mathcal{B}_{+,\nu,p,p+1}(h_n) - (e_{p+1}'\hat{\beta}_{Y-,q}(b_n))\mathcal{B}_{-,\nu,p,p+1}(h_n)\right)$$
$$- \frac{\hat{\tau}_{Y,\nu,p}(h_n)}{\hat{\tau}_{T,\nu,p}(h_n)^2}\left((e_{p+1}'\hat{\beta}_{T+,q}(b_n))\mathcal{B}_{+,\nu,p,p+1}(h_n) - (e_{p+1}'\hat{\beta}_{T-,q}(b_n))\mathcal{B}_{-,\nu,p,p+1}(h_n)\right).$$

Linearizing the estimator we obtain

$$\begin{aligned}
\hat{\varsigma}_{\nu,p,q}^{\mathsf{bc}}(h_n,b_n) - \varsigma_\nu &= \hat{\varsigma}_{\nu,p}(h_n) - h_n^{p+1-\nu}\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) - \varsigma_\nu \\
&= \tilde{\varsigma}_{\nu,p}(h_n) + R_n - h_n^{p+1-\nu}\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) \\
&= \tilde{\varsigma}_{\nu,p,q}^{\mathsf{bc}}(h_n,b_n) + R_n - h_n^{p+1-\nu}\left(\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) - \check{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n)\right) \\
&= \tilde{\varsigma}_{\nu,p,q}^{\mathsf{bc}}(h_n,b_n) + R_n - R_n^{\mathsf{bc}}
\end{aligned}$$

with

$$\tilde{\varsigma}_{\nu,p,q}^{\mathsf{bc}}(h_n,b_n) = \frac{1}{\tau_{T,\nu}}(\hat{\tau}_{Y,\nu,p,q}^{\mathsf{bc}}(h_n,b_n) - \tau_{Y,\nu}) - \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}(\hat{\tau}_{T,\nu,p,q}^{\mathsf{bc}}(h_n,b_n) - \tau_{T,\nu}),$$

$$R_n = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2\hat{\tau}_{T,\nu,p}(h_n)}(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu})^2 - \frac{1}{\tau_{T,\nu}\hat{\tau}_{T,\nu,p}(h_n)}(\hat{\tau}_{Y,\nu,p}(h_n) - \tau_{Y,\nu})(\hat{\tau}_{T,\nu,p}(h_n) - \tau_{T,\nu}),$$

$$\begin{aligned}
\check{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) \;\;=\;\; & \frac{1}{\tau_{T,\nu}}\left((e_{p+1}'\hat{\beta}_{Y+,q}(b_n))\mathcal{B}_{+,\nu,p,p+1}(h_n) - (e_{p+1}'\hat{\beta}_{Y-,q}(b_n))\mathcal{B}_{-,\nu,p,p+1}(h_n)\right) \\
& -\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\left((e_{p+1}'\hat{\beta}_{T+,q}(b_n))\mathcal{B}_{+,\nu,p,p+1}(h_n) - (e_{p+1}'\hat{\beta}_{T-,q}(b_n))\mathcal{B}_{-,\nu,p,p+1}(h_n)\right),
\end{aligned}$$

$$R_n^{\mathsf{bc}} = h_n^{p+1-\nu}\left(\hat{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n) - \check{\mathsf{B}}_{\mathtt{F},\nu,p,q}(h_n,b_n)\right).$$

The following theorem summarizes the asymptotic bias, variance and distribution of $\hat{\varsigma}_{\nu,p,q}^{\mathsf{bc}}(h_n,b_n)$. Theorems 3 and 4 are special cases with $(\nu,p,q) = (0,1,2)$ and $(\nu,p,q) = (1,2,3)$, respectively.

Theorem A2. Suppose Assumptions A1–A3 hold with $S \geq p+2$. Let $\nu, r \in \mathbb{N}$ with $\nu \leq p$.

($\mathbf{R}^{\mathsf{bc}}$) If $\max\{h_n,b_n\} \to 0$, $nh_n^{1+2\nu} \to \infty$ and $nb_n \to \infty$, then

$$R_n^{\mathsf{bc}} = O_p\left(\frac{h_n^{p+1-\nu}}{\sqrt{nh_n^{1+2\nu}}} + h_n^{2(p+1-\nu)}\right)O_p\left(1 + \frac{1}{\sqrt{nb_n^{3+2p}}}\right).$$

(B) If $\max\{h_n, b_n\} \to 0$ and $n \min\{h_n, b_n\} \to \infty$, then

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

where

$$\mathsf{B}^{\mathtt{bc}}_{\mathtt{F},\nu,p,q}(h, b) = \frac{1}{\tau_{T,\nu}} \mathsf{B}^{\mathtt{bc}}_{Y,\nu,p,q}(h_n, b_n) - \frac{\tau_{Y,\nu}}{\tau^2_{T,\nu}} \mathsf{B}^{\mathtt{bc}}_{T,\nu,p,q}(h_n, b_n),$$

with

$$\mathsf{B}^{\mathtt{bc}}_{Y,\nu,p,q}(h, b) = \frac{\mu^{(q+1)}_{Y+}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}(b) \frac{\mathcal{B}_{+,\nu,p,p+1}(h)}{(p+1)!} - \frac{\mu^{(q+1)}_{Y-}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1}(b) \frac{\mathcal{B}_{-,\nu,p,p+1}(h)}{(p+1)!},$$

$$\mathsf{B}^{\mathtt{bc}}_{T,\nu,p,q}(h, b) = \frac{\mu^{(q+1)}_{T+}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}(b) \frac{\mathcal{B}_{+,\nu,p,p+1}(h)}{(p+1)!} - \frac{\mu^{(q+1)}_{T-}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1}(b) \frac{\mathcal{B}_{-,\nu,p,p+1}(h)}{(p+1)!}.$$

(V) If $\max\{h_n, b_n\} \to 0$ and $n \min\{h_n, b_n\} \to \infty$, then $\mathbb{V}[\tilde{\varsigma}^{\mathtt{bc}}_{\nu,p,q}(h_n, b_n)|\mathcal{X}_n] = \mathsf{V}^{\mathtt{bc}}_{\mathtt{F},\nu,p,q}(h_n, b_n)$, where

$$\mathsf{V}^{\mathtt{bc}}_{\mathtt{F},\nu,p,q}(h_n, b_n) = \mathsf{V}^{\mathtt{bc}}_{\mathtt{F},+,\nu,p,q}(h_n, b_n) + \mathsf{V}^{\mathtt{bc}}_{\mathtt{F},-,\nu,p,q}(h_n, b_n),$$

where

$$\mathsf{V}^{\mathtt{bc}}_{\mathtt{F},+,\nu,p,q}(h, b) = \mathsf{V}_{\mathtt{F},+,\nu,p}(h) - 2h^{p+1-\nu} \mathcal{C}_{\mathtt{F},+,\nu,p,q}(h, b) \frac{\mathcal{B}_{+,\nu,p,p+1}(h)}{(p+1)!} + h^{2p+2-2\nu} \mathsf{V}_{\mathtt{F},+,p+1,q}(b) \frac{\mathcal{B}^2_{+,\nu,p,p+1}(h)}{(p+1)!^2},$$

$$\mathsf{V}^{\mathtt{bc}}_{\mathtt{F},-,\nu,p,q}(h, b) = \mathsf{V}_{\mathtt{F},-,\nu,p}(h) - 2h^{p+1-\nu} \mathcal{C}_{\mathtt{F},-,\nu,p,q}(h, b) \frac{\mathcal{B}_{-,\nu,p,p+1}(h)}{(p+1)!} + h^{2p+2-2\nu} \mathsf{V}_{\mathtt{F},-,p+1,q}(b) \frac{\mathcal{B}^2_{-,\nu,p,p+1}(h)}{(p+1)!^2},$$

$$\mathcal{C}_{\mathtt{F},+,\nu,p,q}(h, b) = \frac{1}{\tau^2_{T,\nu}} \mathcal{C}_{YY+,\nu,p,q}(h, b) - \frac{2\tau_{Y,\nu}}{\tau^3_{T,\nu}} \mathcal{C}_{YT+,\nu,p,q}(h, b) + \frac{\tau^2_{Y,\nu}}{\tau^4_{T,\nu}} \mathcal{C}_{TT+,\nu,p,q}(h, b),$$

$$\mathcal{C}_{\mathtt{F},-,\nu,p,q}(h, b) = \frac{1}{\tau^2_{T,\nu}} \mathcal{C}_{YY-,\nu,p,q}(h, b) - \frac{2\tau_{Y,\nu}}{\tau^3_{T,\nu}} \mathcal{C}_{YT-,\nu,p,q}(h, b) + \frac{\tau^2_{Y,\nu}}{\tau^4_{T,\nu}} \mathcal{C}_{TT-,\nu,p,q}(h, b),$$

where

$$\mathcal{C}_{YY+,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e'_\nu \Gamma^{-1}_{+,p}(h) \Psi_{YT+,p,q}(h, b) \Gamma^{-1}_{+,q}(b) e_{p+1},$$

$$\mathcal{C}_{YT+,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e'_\nu \Gamma^{-1}_{+,p}(h) \Psi_{YT+,p,q}(h, b) \Gamma^{-1}_{+,q}(b) e_{p+1},$$

$$\mathcal{C}_{TT+,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e'_\nu \Gamma^{-1}_{+,p}(h) \Psi_{TT+,p,q}(h, b) \Gamma^{-1}_{+,q}(b) e_{p+1},$$

$$\mathcal{C}_{YY-,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e'_\nu \Gamma^{-1}_{-,p}(h) \Psi_{YY-,p,q}(h, b) \Gamma^{-1}_{-,q}(b) e_{p+1},$$

$$\mathcal{C}_{YT-,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e'_\nu \Gamma^{-1}_{-,p}(h) \Psi_{YT-,p,q}(h, b) \Gamma^{-1}_{-,q}(b) e_{p+1},$$

$$\mathcal{C}_{TT-,\nu,p,q}(h, b) = \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e'_\nu \Gamma^{-1}_{-,p}(h) \Psi_{TT-,p,q}(h, b) \Gamma^{-1}_{-,q}(b) e_{p+1}.$$

(D) If $n \min\{h_n^{2p+3}, b_n^{2p+3}\} \max\{h_n^2, b_n^{2(q-p)}\} \to 0$ and $n \min\{h_n^{1+2\nu}, b_n\} \to \infty$, then

$$\frac{\hat{\varsigma}^{\mathtt{bc}}_{\nu,p,q}(h_n, b_n) - \varsigma_\nu}{\sqrt{\mathsf{V}^{\mathtt{bc}}_{\mathtt{F},\nu,p,q}(h_n, b_n)}} \to_d \mathcal{N}(0, 1),$$

provided that $h_n \to 0$ and $\kappa b_n < \kappa_0$.

From this theorem we obtain for Theorems 3 and 4:

$$\mathsf{V}^{\mathsf{bc}}_{\mathtt{FRD}}(h_n, b_n) = \mathsf{V}^{\mathsf{bc}}_{\mathtt{F},0,1,2}(h_n, b_n) \qquad \text{and} \qquad \mathsf{V}^{\mathsf{bc}}_{\mathtt{FKRD}}(h_n, b_n) = \mathsf{V}^{\mathsf{bc}}_{\mathtt{F},1,2,3}(h_n, b_n). \tag{A-2}$$

A.6 Sharp RD Bandwidth Selectors

Using the results in Section 5 we propose data-driven bandwidth selector for sharp RD designs. For any $\nu \leq p$, we denote

$$\hat{\mathsf{V}}_{\nu,p}(h_n) = \hat{\mathcal{V}}_{+,\nu,p}(h_n) + \hat{\mathcal{V}}_{-,\nu,p}(h_n),$$

$$\hat{\mathcal{V}}_{+,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu' \Gamma_{+,p}^{-1}(h_n)\hat{\Psi}_{YY+,p}(h_n)\Gamma_{+,p}^{-1}(h_n)e_\nu,$$

$$\hat{\mathcal{V}}_{-,\nu,p}(h_n) = \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu' \Gamma_{-,p}^{-1}(h_n)\hat{\Psi}_{YY-,p}(h_n)\Gamma_{-,p}^{-1}(h_n)e_\nu,$$

where $\hat{\Psi}_{YY+,p}(h_n)$ and $\hat{\Psi}_{YY-,p}(h_n)$ are constructed as described in Section 5.

Plug-in Bandwidths Selectors. Fix $p, q \in \mathbb{N}$ with $q \geq p+1$. Let $\mathcal{B}_{\nu,p} = e_\nu' \Gamma_p^{-1} \vartheta_{p,p+1}$.

Step 0: Initial Bandwidths (v_n, c_n).
(i) Suppose $v_n \to_p 0$ and $nv_n \to_p \infty$. In particular, let $v_n = 2.58 \cdot \omega \cdot n^{-1/5}$ with

$$\omega = \min\left\{S_X,\ \frac{IQR_X}{1.349}\right\},$$

where S_X^2 denotes the sample variance of X_i, and IQR_X is the interquartile range of X_i.

(ii) Suppose $c_n \to_p 0$ and $nc_n^{2q+3} \to_p \infty$. In particular, let $c_n = \widehat{C}_{q+1,q+1}\ n^{-1/(2q+5)}$ with

$$\widehat{C}_{q+1,q+1} = \left(\frac{(2q+3)nv_n^{2q+3}\hat{\mathsf{V}}_{q+1,q+1}(v_n)}{2\mathcal{B}_{q+1,q+1}^2\left(e_{q+2}'\check{\beta}_{+,q+2} - e_{q+2}'\check{\beta}_{-,q+2}\right)^2}\right)^{1/(2q+5)},$$

where $\check{\beta}_{+,p}$ and $\check{\beta}_{-,p}$ denote the estimated coefficients of a $(p+1)$-th order global polynomial fit at either side of the threshold; i.e.,

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

$$\check{\beta}_{-,p} = \arg\min_{\beta\in\mathbb{R}^p}\sum_{i=1}^{n}\mathbf{1}(X_i < 0)(Y_i - r_p(X_i)'\beta)^2.$$

Step 1: Pilot Bandwidth b_n. Compute $\hat{b}_{p+1,q} = \widehat{C}_{p+1,q}\ n^{-1/(2q+3)}$ with

$$\widehat{C}_{p+1,q} = \left(\frac{(2p+3)nv_n^{2p+3}\hat{\mathsf{V}}_{p+1,q}(v_n)}{2(q-p)\mathcal{B}_{p+1,q}^2\left\{\left(e_{q+1}'\hat{\beta}_{+,q+1}(c_n) - e_{q+1}'\hat{\beta}_{-,q+1}(c_n)\right)^2 + 3\hat{\mathsf{V}}_{q+1,q+1}(c_n)\right\}}\right)^{1/(2q+3)}.$$

Step 2: Main Bandwidth h_n. Let $b_n = \hat{b}_{p+1,q}$, and compute $\hat{h}_{\nu,p} = \widehat{C}_{\nu,p}\ n^{-1/(2p+3)}$ with

$$\widehat{C}_{\nu,p} = \left(\frac{(2\nu+1)nv_n\hat{\mathsf{V}}_{p,0}(v_n)}{2(p+1-\nu)\mathcal{B}_{\nu,p}^2\left\{\left(e_{p+1}'\hat{\beta}_{+,q}(b_n) - e_{p+1}'\hat{\beta}_{-,q}(b_n)\right)^2 + 3\hat{\mathsf{V}}_{p+1,q}(b_n)\right\}}\right)^{1/(2p+3)}.$$

The selectors $\hat{h}_{\nu,p}$ and $\hat{b}_{p+1,q}$ are constructed following the idea of an ℓ-stage DPI bandwidth selector for density estimation (resp. with $\ell = 2$ and $\ell = 1$). See, e.g., Wand and Jones (1995, Section 3.6) for further discussion. The following theorem shows that these bandwidths selectors are consistent, and also optimal in the sense of Li (1987).

Theorem A3. Suppose Assumptions 1–2 hold with $S \geq q+1$ and $q \geq p+1$. In addition, suppose $e'_{q+2}\check{\beta}_{+,q+2} - e'_{q+2}\check{\beta}_{-,q+2} \to_p c \neq 0$ and $\nu \leq p$.

(**Step 1**) If $\mu_+^{(q+1)} \neq \mu_-^{(q+1)}$, then

$$\frac{\hat{b}_{p+1,q}}{b_{\mathtt{MSE},p+1,q}} \to_p 1 \qquad \text{and} \qquad \frac{\mathsf{MSE}_{p+1,q}(\hat{b}_{p+1,q})}{\mathsf{MSE}_{p+1,q}(b_{\mathtt{MSE},p+1,q})} \to_p 1.$$

(**Step 2**) If $\mu_+^{(p+1)} \neq \mu_-^{(p+1)}$, then

$$\frac{\hat{h}_{\nu,p}}{h_{\mathtt{MSE},\nu,p}} \to_p 1 \qquad \text{and} \qquad \frac{\mathsf{MSE}_{\nu,p}(\hat{h}_{\nu,p})}{\mathsf{MSE}_{\nu,p}(h_{\mathtt{MSE},\nu,p})} \to_p 1.$$

References

Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267.

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

Angelucci, M., and O. Attanasio (2009): "Program Effect on Consumption, Low Participation, and Methodological Issues," *Economic Development and Cultural Change*, 57(3), 479–506.

——— (2013): "The Demand for Food of Poor Urban Mexican Households: Understanding Policy Impacts Using Structural Models," *American Economic Journal: Economic Policy*, 5(1), 146–178.

Angelucci, M., and G. De Giorgi (2009): "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?," *American Economic Review*, 99(1), 486–508.

Attanasio, O., C. Meghir, and A. Santiago (2011): "Education choices in Mexico: Using a structural model and a randomized experiment to evaluate Progresa," *Review of Economic Studies*, 79(1), 37–66.

Behrman, J. R., J. Gallardo-García, S. W. Parker, P. E. Todd, and V. Vélez-Grajales (2012): "Are conditional cash transfers effective in urban areas? Evidence from Mexico," *Education Economics*, 20(3), 233–259.

Buddelmeyer, H., and E. Skoufias (2004): "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA," World Bank Policy Research Working Paper 3386.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2013): "Robust Data-Driven Inference in the Regression-Discontinuity Design," revision requested by *Stata Journal*.

Card, D., D. S. Lee, Z. Pei, and A. Weber (2012): "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design," Working paper, University of California at Berkeley.

Cattaneo, M. D., and M. H. Farrell (2013): "Optimal Convergence Rates, Bahadur Representation, and Asymptotic Normality of Partitioning Estimators," *Journal of Econometrics*, 174(2), 127–143.

Cattaneo, M. D., B. Frandsen, and R. Titiunik (2013): "Randomization Inference in the Regression Discontinuity Design: An Application to the Study of Party Advantages in the U.S. Senate," Working paper, University of Michigan.

Cheng, M.-Y., J. Fan, and J. S. Marron (1997): "On Automatic Boundary Corrections," *Annals of Statistics*, 25(4), 1691–1708.

DesJardins, S. L., and B. P. McCall (2009): "The Impact of the Gates Millennium Scholars Program on the Retention, College Finance- and Work-Related Choices, and Future Educational Aspirations of Low-Income Students," working paper, University of Michigan.

Dinardo, J., and D. S. Lee (2011): "Program Evaluation and Research Designs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4A, pp. 463–536. Elsevier Science B.V.

Djebbari, H., and J. Smith (2008): "Heterogeneous impacts in PROGRESA," *Journal of Econometrics*, 145(1-2), 64–80.

Dong, Y. (2012): "Jumpy or Kinky? Regression Discontinuity Without The Discontinuity," working paper, University of California at Irvine.

Dong, Y., and A. Lewel (2012): "Regression Discontinuity Marginal Threshold Treatment Effects," working paper, Boston College.

Dubois, P., A. de Janvry, and E. Sadoulet (2012): "Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico," *Journal of Labor Economics*, 30(3), 555–589.

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

Fernald, L. C. H., P. J. Gertler, and L. M. Neufeld (2009): "10-year eff ect of Oportunidades, MexicoŠs conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study," *Lancet*, 374, 1997–2005.

Frandsen, B., M. Frölich, and B. Melly (2012): "Quantile treatments effects in the regression discontinuity design," *Journal of Econometrics*, 168(2), 382–395.

Gertler, P. J., S. W. Martinez, and M. Rubio-Codina (2012): "Investing Cash Transfers to Raise Long-Term Living Standards," *American Economic Journal: Applied Economics*, 4(1), 164–192.

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.

Hoddinott, J., and E. Skoufias (2004): "The Impact of PROGRESA on Food Consumption," *Economic Development and Cultural Change*, 53(1), 37–61.

Ichimura, H., and P. E. Todd (2007): "Implementing Nonparametric and Semiparametric Estimators," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B of *Handbook of Econometrics*, chap. 74. Elsevier.

Imbens, G., and T. Lemieux (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142(2), 615–635.

Imbens, G. W., and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933–959.

Lee, D. S. (2008): "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics*, 142(2), 675–697.

Lee, D. S., and D. Card (2008): "Regression discontinuity inference with specification error," *Journal of Econometrics*, 142(2), 655–674.

Lee, D. S., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

Li, K.-C. (1987): "Asymptotic Optimality for C_p, C_L, Cross-validation and Generalized Cross-validation: Discrete Index Set," *Annals of Statistics*, 15(3), 958–975.

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.

Marmer, V., D. Feir, and T. Lemieux (2012): "Weak Identification in Fuzzy Regression Discontinuity Designs," working paper, Univeristy of British Columbia.

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

Otsu, T., and K.-L. Xu (2011): "Empirical Likelihood for Regression Discontinuity Design," working paper, Yale University.

Porter, J. (2003): "Estimation in the Regression Discontinuity Model," working paper, University of Wisconsin.

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.

van der Klaauw, W. (2008): "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics," *Labour*, 22(2), 219–245.

Wand, M., and M. Jones (1995): *Kernel Smoothing*. Chapman & Hall/CRC, Florida.

Table 2: Empirical Coverage and Average Interval Lenght of different 95% Confidence Intervals (Infeasible Asymptotic Variance)

	Conventional			**Bias-Corrected**			**Robust Approach**			**Bandwidths**	
		EC (%)	IL		EC (%)	IL		EC (%)	IL	h_n	b_n
Model 1											
	$T_{\mathtt{SRD}}(h_{\mathtt{MSE}})$	93.9	0.225	$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	91.4	0.225	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	94.7	0.254	0.166	0.319
	$T_{\mathtt{SRD}}(h_{\mathtt{DM}})$	93.4	0.213	$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, b_{\mathtt{DM}})$	88.7	0.213	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, b_{\mathtt{DM}})$	94.8	0.261	0.184	0.271
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}})$	84.4	0.159	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	77.7	0.159	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	93.3	0.244	0.335	0.337
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}})$	79.6	0.137	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{b}_{\mathtt{DM}})$	63.8	0.137	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{b}_{\mathtt{DM}})$	89.8	0.267	0.496	0.423
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}})$	83.1	0.152	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	78.1	0.152	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	93.1	0.224	0.381	0.381
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}})$	91.2	0.205	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	88.0	0.205	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	93.0	0.242	0.205	0.336
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	81.0	0.225	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	94.9	0.338	0.166	0.166
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, h_{\mathtt{DM}})$	81.3	0.213	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, h_{\mathtt{DM}})$	95.1	0.319	0.184	0.184
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	81.3	0.159	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	94.7	0.235	0.335	0.335
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{h}_{\mathtt{DM}})$	71.2	0.137	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{h}_{\mathtt{DM}})$	89.9	0.200	0.496	0.496
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	81.6	0.205	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	95.3	0.306	0.205	0.205
Model 2											
	$T_{\mathtt{SRD}}(h_{\mathtt{MSE}})$	92.5	0.327	$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	92.5	0.327	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	94.9	0.354	0.082	0.191
	$T_{\mathtt{SRD}}(h_{\mathtt{DM}})$	92.2	0.323	$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, b_{\mathtt{DM}})$	92.5	0.323	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, b_{\mathtt{DM}})$	94.9	0.352	0.084	0.190
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}})$	24.1	0.213	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	85.0	0.213	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	91.2	0.254	0.185	0.296
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}})$	14.8	0.206	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{b}_{\mathtt{DM}})$	82.8	0.206	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{b}_{\mathtt{DM}})$	89.1	0.245	0.196	0.319
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}})$	79.1	0.269	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	80.5	0.269	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	94.8	0.411	0.119	0.119
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}})$	87.6	0.300	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	92.0	0.300	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	94.2	0.325	0.097	0.226
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	79.0	0.327	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	94.8	0.513	0.082	0.082
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, h_{\mathtt{DM}})$	79.2	0.323	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, h_{\mathtt{DM}})$	94.8	0.505	0.084	0.084
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	81.0	0.213	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	94.8	0.319	0.185	0.185
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{h}_{\mathtt{DM}})$	80.8	0.206	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{h}_{\mathtt{DM}})$	94.6	0.309	0.196	0.196
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	79.6	0.300	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	94.8	0.464	0.097	0.097
Model 3											
	$T_{\mathtt{SRD}}(h_{\mathtt{MSE}})$	85.8	0.179	$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	84.7	0.179	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, b_{\mathtt{MSE}})$	95.0	0.246	0.260	0.292
	$T_{\mathtt{SRD}}(h_{\mathtt{DM}})$	87.4	0.182	$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, b_{\mathtt{DM}})$	86.1	0.182	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, b_{\mathtt{DM}})$	95.0	0.241	0.251	0.305
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}})$	87.1	0.190	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	88.8	0.190	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{b}_{\mathtt{IK}})$	95.1	0.235	0.231	0.340
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}})$	91.4	0.200	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{b}_{\mathtt{DM}})$	91.4	0.200	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{b}_{\mathtt{DM}})$	95.2	0.229	0.209	0.390
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}})$	93.9	0.226	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	81.9	0.226	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CV}}, \hat{h}_{\mathtt{CV}})$	95.2	0.340	0.166	0.166
	$T_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}})$	92.1	0.217	$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	91.4	0.217	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{b}_{\mathtt{CCT}})$	95.2	0.250	0.181	0.322
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	81.7	0.179	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{MSE}}, h_{\mathtt{MSE}})$	94.9	0.266	0.260	0.260
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, h_{\mathtt{DM}})$	81.8	0.182	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(h_{\mathtt{DM}}, h_{\mathtt{DM}})$	95.0	0.271	0.251	0.251
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	81.8	0.190	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{IK}}, \hat{h}_{\mathtt{IK}})$	95.0	0.283	0.231	0.231
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{h}_{\mathtt{DM}})$	81.8	0.200	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{DM}}, \hat{h}_{\mathtt{DM}})$	94.9	0.298	0.209	0.209
				$T^{\mathtt{bc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	82.0	0.217	$T^{\mathtt{rbc}}_{\mathtt{SRD}}(\hat{h}_{\mathtt{CCT}}, \hat{h}_{\mathtt{CCT}})$	95.2	0.325	0.181	0.181

Notes: (i) EC = Empirical Coverage in percentage points, (ii) IL = Average Interval Length, (iii) columns under "Bandwidths" report the population and average estimated bandwidths choices, as appropriate, for main bandwidth h_n and pilot bandwidth b_n.

Table 3: Empirical Coverage and Average Interval Lenght of different 95% Confidence Intervals (Estimated Asymptotic Variance with $J = 3$ nearest-neighbors)

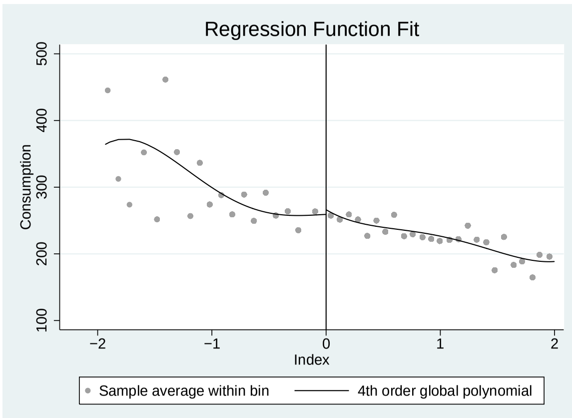
	Conventional			Bias-Corrected			Robust Approach			Bandwidths	
		EC (%)	IL		EC (%)	IL		EC (%)	IL	h_n	b_n
Model 1											
	$\hat{T}_{\texttt{SRD}}(h_{\texttt{MSE}})$	92.3	0.223	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	89.7	0.223	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	93.4	0.251	0.166	0.319
	$\hat{T}_{\texttt{SRD}}(h_{\texttt{DM}})$	92.0	0.211	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{DM}}, b_{\texttt{DM}})$	87.2	0.211	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{DM}}, b_{\texttt{DM}})$	93.4	0.259	0.184	0.271
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	83.2	0.158	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	76.7	0.158	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	91.8	0.241	0.335	0.337
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}})$	78.7	0.136	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{b}_{\texttt{DM}})$	63.3	0.136	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{b}_{\texttt{DM}})$	88.6	0.265	0.496	0.423
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	81.8	0.151	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	77.4	0.151	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	91.5	0.222	0.381	0.381
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}})$	89.9	0.201	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	86.7	0.201	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	91.7	0.238	0.205	0.336
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	79.4	0.223	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	92.3	0.332	0.166	0.166
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{DM}}, h_{\texttt{DM}})$	79.7	0.211	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{DM}}, h_{\texttt{DM}})$	92.6	0.313	0.184	0.184
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	80.4	0.158	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	93.4	0.232	0.335	0.335
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{h}_{\texttt{DM}})$	70.7	0.136	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{h}_{\texttt{DM}})$	88.6	0.198	0.496	0.496
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	80.1	0.201	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	92.6	0.298	0.205	0.205
Model 2											
	$\hat{T}_{\texttt{SRD}}(h_{\texttt{MSE}})$	91.2	0.354	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	91.6	0.354	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	93.5	0.384	0.082	0.191
	$\hat{T}_{\texttt{SRD}}(h_{\texttt{DM}})$	91.1	0.349	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{DM}}, b_{\texttt{DM}})$	91.5	0.349	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{DM}}, b_{\texttt{DM}})$	93.6	0.381	0.084	0.190
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	28.0	0.224	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	85.4	0.224	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	91.0	0.268	0.185	0.296
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}})$	17.9	0.216	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{b}_{\texttt{DM}})$	83.2	0.216	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{b}_{\texttt{DM}})$	89.1	0.258	0.196	0.319
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	79.7	0.287	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	81.3	0.287	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	93.4	0.450	0.119	0.119
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}})$	87.5	0.318	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	91.1	0.318	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	93.2	0.345	0.097	0.226
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	79.5	0.354	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	92.9	0.567	0.082	0.082
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{DM}}, h_{\texttt{DM}})$	79.7	0.349	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{DM}}, h_{\texttt{DM}})$	93.0	0.559	0.084	0.084
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	81.5	0.224	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	94.0	0.344	0.185	0.185
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{h}_{\texttt{DM}})$	81.4	0.216	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{h}_{\texttt{DM}})$	94.0	0.332	0.196	0.196
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	80.3	0.318	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	93.1	0.504	0.097	0.097
Model 3											
	$\hat{T}_{\texttt{SRD}}(h_{\texttt{MSE}})$	84.7	0.178	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	83.7	0.178	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, b_{\texttt{MSE}})$	93.5	0.244	0.260	0.292
	$\hat{T}_{\texttt{SRD}}(h_{\texttt{DM}})$	86.4	0.181	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{DM}}, b_{\texttt{DM}})$	84.9	0.181	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{DM}}, b_{\texttt{DM}})$	93.7	0.239	0.251	0.305
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}})$	86.1	0.189	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	87.6	0.189	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{b}_{\texttt{IK}})$	93.8	0.233	0.231	0.340
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}})$	90.0	0.198	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{b}_{\texttt{DM}})$	89.8	0.198	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{b}_{\texttt{DM}})$	93.7	0.227	0.209	0.390
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}})$	92.2	0.223	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	79.9	0.223	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CV}}, \hat{h}_{\texttt{CV}})$	92.7	0.333	0.166	0.166
	$\hat{T}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}})$	90.5	0.213	$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	89.8	0.213	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{b}_{\texttt{CCT}})$	93.4	0.246	0.181	0.322
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	80.7	0.178	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{MSE}}, h_{\texttt{MSE}})$	93.3	0.262	0.260	0.260
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(h_{\texttt{DM}}, h_{\texttt{DM}})$	80.6	0.181	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(h_{\texttt{DM}}, h_{\texttt{DM}})$	93.3	0.267	0.251	0.251
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	80.6	0.189	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{IK}}, \hat{h}_{\texttt{IK}})$	92.9	0.279	0.231	0.231
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{h}_{\texttt{DM}})$	80.5	0.198	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{DM}}, \hat{h}_{\texttt{DM}})$	92.9	0.294	0.209	0.209
				$\hat{T}^{\texttt{bc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	80.3	0.213	$\hat{T}^{\texttt{rbc}}_{\texttt{SRD}}(\hat{h}_{\texttt{CCT}}, \hat{h}_{\texttt{CCT}})$	92.7	0.317	0.181	0.181

Notes: (i) EC = Empirical Coverage in percentage points, (ii) IL = Average Interval Length, (iii) columns under "Bandwidths" report the population and average estimated bandwidths choices, as appropriate, for main bandwidth h_n and pilot bandwidth b_n.

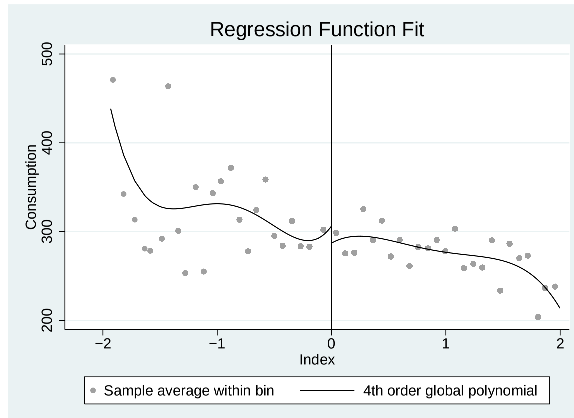
Figure 2: RD Plots of Progresa/Oportunidades on Food and Non-Food Consumption, Urban Localities.

Panel A: Food Consumption

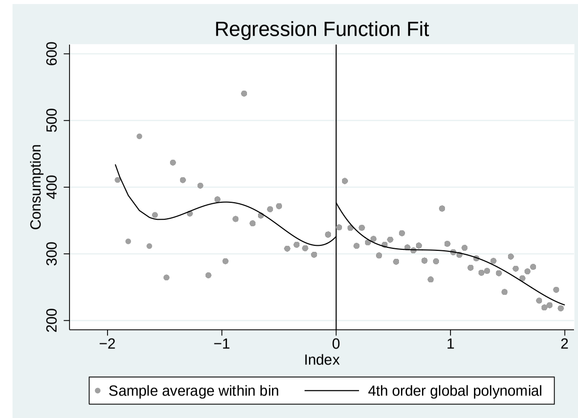
(a) Pre-Intervention



(b) 1-year Treatment

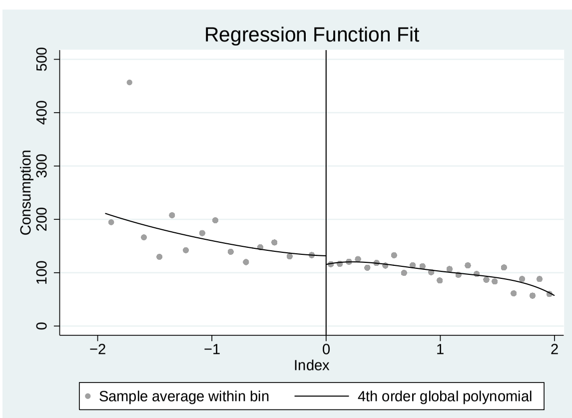


(c) 2-year Treatment

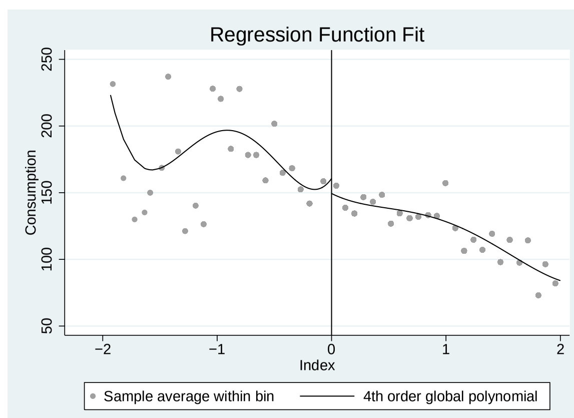


Panel B: Non-Food Consumption

(d) Pre-Intervention



(e) 1-year Treatment



(f) 2-year Treatment

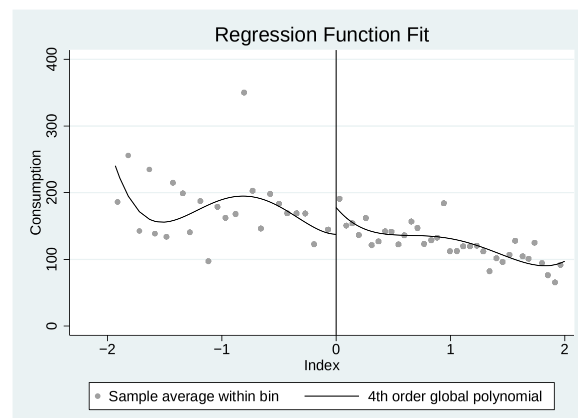
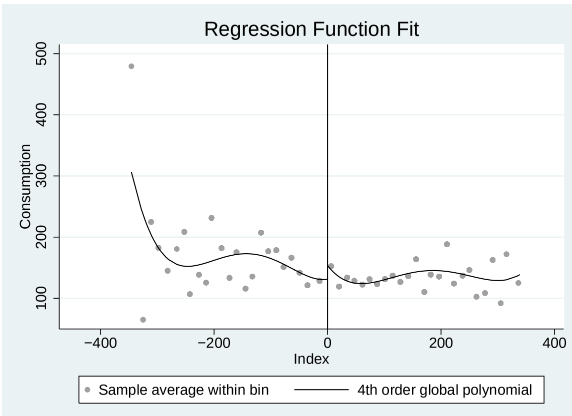


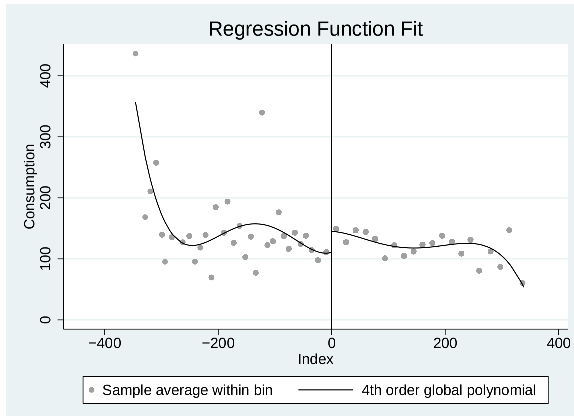
Figure 3: RD Plots of Progresa/Oportunidades on Food and Non-Food Consumption, Rural Localities.

Panel A: Food Consumption

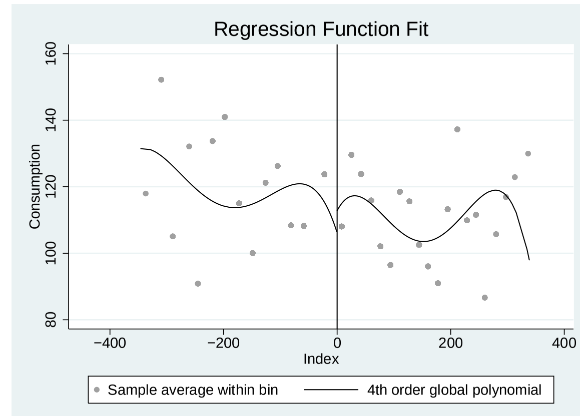
(a) Pre-Intervention



(b) 1-year Treatment

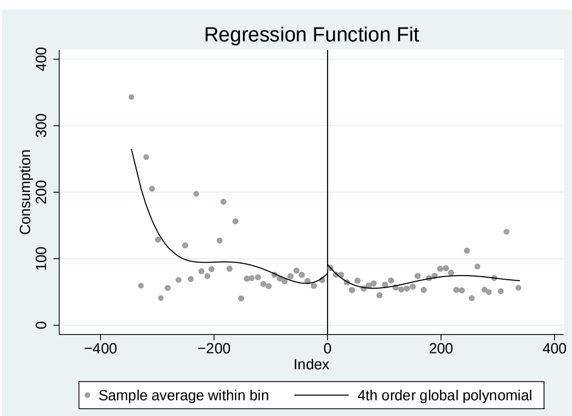


(c) 2-year Treatment

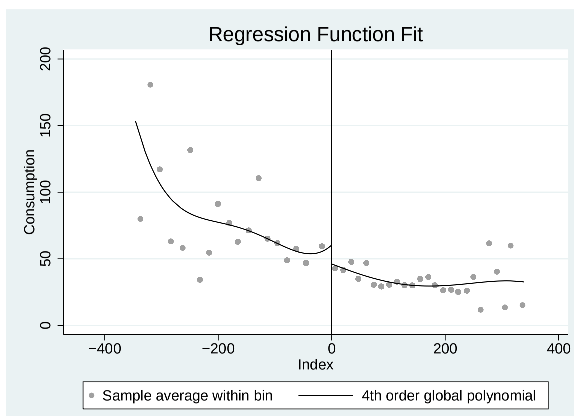


Panel B: Non-Food Consumption

(d) Pre-Intervention



(e) 1-year Treatment



(f) 2-year Treatment

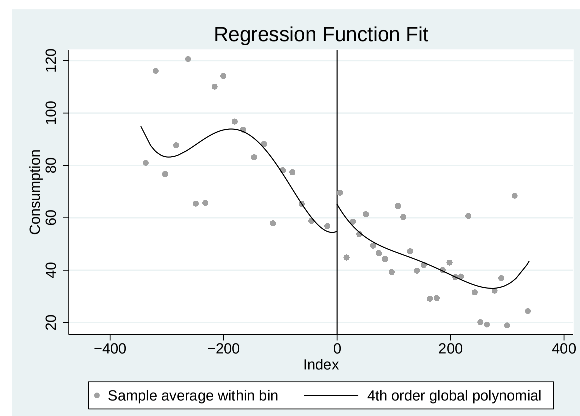


Table 4: Sharp RD Treatment Effect Estimates of Progresa/Oportunidades on Consumption.

PANEL A: URBAN LOCALITIES

	Pre-Intervention			**1-year Treatment**			**2-year Treatment**		
	BW-CCT	BW-IK	BW-CV	BW-CCT	BW-IK	BW-CV	BW-CCT	BW-IK	BW-CV
Food									
	3.4	3.8	6.9	−11.6	6.5	6.1	50.3	48.3	49.0
	(−29.0 , 35.9)	(−24.9 , 32.4)	(−16.3 , 30.0)	(−44.8 , 21.5)	(−16.1 , 29.1)	(−17.0 , 29.2)	(−8.8 , 109.4)*	(5.2 , 91.3)**	(1.4 , 96.5)**
	[−36.9 , 39.5]	[−40.1 , 41.7]	[−30.9 , 35.8]	[−54.7 , 20.8]	[−35.3 , 25.9]	[−39.6 , 24.0]	[−20.1 , 121.3]	[−14.7 , 107.9]	[−19.5 , 117.5]
	$\hat{h}_{\texttt{CCT}} = 0.57$	$\hat{h}_{\texttt{IK}} = 0.76$	$\hat{h}_{\texttt{CV}} = 1.25$	$\hat{h}_{\texttt{CCT}} = 0.44$	$\hat{h}_{\texttt{IK}} = 1.22$	$\hat{h}_{\texttt{CV}} = 1.13$	$\hat{h}_{\texttt{CCT}} = 0.43$	$\hat{h}_{\texttt{IK}} = 0.76$	$\hat{h}_{\texttt{CV}} = 0.64$
	$\hat{b}_{\texttt{CCT}} = 0.91$	$\hat{b}_{\texttt{IK}} = 0.74$		$\hat{b}_{\texttt{CCT}} = 0.80$	$\hat{b}_{\texttt{IK}} = 1.26$		$\hat{b}_{\texttt{CCT}} = 0.59$	$\hat{b}_{\texttt{IK}} = 0.82$	
Non-Food									
	−10.1	−8.7	−10.5	−8.8	−0.7	−0.8	40.6	32.8	36.2
	(−34.1 , 14.0)	(−27.9 , 10.6)	(−31.5 , 10.4)	(−38.0 , 20.3)	(−21.5 , 20.0)	(−21.8 , 20.2)	(0.3 , 80.9)**	(5.1 , 60.4)**	(5.8 , 66.7)**
	[−37.8 , 18.0]	[−45.9 , 21.9]	[−39.8 , 16.2]	[−46.2 , 21.0]	[−37.0 , 20.6]	[−36.1 , 20.8]	[−7.6 , 88.0]*	[4.7 , 77.3]**	[−1.4 , 87.1]*
	$\hat{h}_{\texttt{CCT}} = 0.56$	$\hat{h}_{\texttt{IK}} = 1.11$	$\hat{h}_{\texttt{CV}} = 0.84$	$\hat{h}_{\texttt{CCT}} = 0.37$	$\hat{h}_{\texttt{IK}} = 0.95$	$\hat{h}_{\texttt{CV}} = 0.91$	$\hat{h}_{\texttt{CCT}} = 0.40$	$\hat{h}_{\texttt{IK}} = 0.82$	$\hat{h}_{\texttt{CV}} = 0.68$
	$\hat{b}_{\texttt{CCT}} = 0.90$	$\hat{b}_{\texttt{IK}} = 0.80$		$\hat{b}_{\texttt{CCT}} = 0.63$	$\hat{b}_{\texttt{IK}} = 0.90$		$\hat{b}_{\texttt{CCT}} = 0.55$	$\hat{b}_{\texttt{IK}} = 1.06$	

PANEL B: RURAL LOCALITIES

	Pre-Intervention			**1-year Treatment**			**2-year Treatment**		
	BW-CCT	BW-IK	BW-CV	BW-CCT	BW-IK	BW-CV	BW-CCT	BW-IK	BW-CV
Food									
	16.2	1.6	6.6	33.4	33.0	38.2	13.1	0.9	3.5
	(−22.3 , 54.7)	(−21.3 , 24.5)	(−18.5 , 31.6)	(6.6 , 60.3)**	(13.1 , 52.9)***	(17.1 , 59.3)***	(−7.8 , 34.0)	(−21.6 , 23.4)	(−21.5 , 28.6)
	[−32.8 , 57.1]	[−13.2 , 55.5]	[−11.6 , 60.6]	[−0.3 , 61.4]*	[23.1 , 74.6]***	[22.3 , 73.5]***	[−8.4 , 39.8]	[−19.8 , 31.8]	[−11.7 , 34.3]
	$\hat{h}_{\texttt{CCT}} = 76$	$\hat{h}_{\texttt{IK}} = 249$	$\hat{h}_{\texttt{CV}} = 200$	$\hat{h}_{\texttt{CCT}} = 63$	$\hat{h}_{\texttt{IK}} = 242$	$\hat{h}_{\texttt{CV}} = 200$	$\hat{h}_{\texttt{CCT}} = 66$	$\hat{h}_{\texttt{IK}} = 275$	$\hat{h}_{\texttt{CV}} = 165$
	$\hat{b}_{\texttt{CCT}} = 122$	$\hat{b}_{\texttt{IK}} = 241$		$\hat{b}_{\texttt{CCT}} = 111$	$\hat{b}_{\texttt{IK}} = 228$		$\hat{b}_{\texttt{CCT}} = 125$	$\hat{b}_{\texttt{IK}} = 237$	
Non-Food									
	18.8	8.3	17.5	−11.7	−8.4	−8.4	14.4	10.0	10.3
	(−5.6 , 43.3)	(−7.9 , 24.5)	(−6.2 , 41.2)	(−34.2 , 10.7)	(−22.3 , 5.5)	(−22.9 , 6.1)	(−1.8 , 30.5)*	(−1.9 , 21.9)*	(−0.7 , 21.4)*
	[−5.9 , 50.4]	[−10.5 , 44.5]	[−6.4 , 63.2]	[−40.1 , 12.2]	[−27.0 , 10.1]	[−30.9 , 13.1]	[−1.8 , 35.1]*	[−4.7 , 26.1]	[−3.8 , 26.7]
	$\hat{h}_{\texttt{CCT}} = 107$	$\hat{h}_{\texttt{IK}} = 275$	$\hat{h}_{\texttt{CV}} = 115$	$\hat{h}_{\texttt{CCT}} = 97$	$\hat{h}_{\texttt{IK}} = 293$	$\hat{h}_{\texttt{CV}} = 245$	$\hat{h}_{\texttt{CCT}} = 84$	$\hat{h}_{\texttt{IK}} = 191$	$\hat{h}_{\texttt{CV}} = 235$
	$\hat{b}_{\texttt{CCT}} = 183$	$\hat{b}_{\texttt{IK}} = 223$		$\hat{b}_{\texttt{CCT}} = 165$	$\hat{b}_{\texttt{IK}} = 374$		$\hat{b}_{\texttt{CCT}} = 157$	$\hat{b}_{\texttt{IK}} = 223$	

Notes:
(i) BW-CCT, BW-IK and BW-CV correspond to estimation methods using bandwidth selectors proposed in Section A.6, IK and by cross-validation, respectively.
(ii) For each bandwidth selection method, table reports RD local-linear point estimator, conventional 95% confidence intervals in parenthesis, robust 95% confidence intervals in brackets, and bandwidths values. All confidence intervals are also robust to heteroskedasticity.
(iii) For each confidence interval, accompanying stars denote associated null hypothesis of no-treatment effect rejected at:
* statistically significant at 10% level,
** statistically significant at 5% level, and
*** statistically significant at 1% level.