Robust Nonparametric Bias-Corrected Inference in the Regression Discontinuity Design*

Sebastian Calonico† Matias D. Cattaneo‡ Rocio Titiunik§

June 22, 2012

Abstract

In the regression discontinuity design units are assigned treatment based on whether their value of an observed covariate exceeds a known cutoff. Local polynomial estimators are now routinely employed to construct confidence intervals for treatment effects in this design. 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. In this paper, we show that these interval estimators may be improved by employing bias-corrected local polynomial estimators together with a novel standard-error estimator. Our innovation is to consider an alternative distributional approximation for the bias-corrected estimator that explicitly accounts for the (possibly first-order) contribution of the bias-estimate to the variability of the statistic. This alternative asymptotics lead to a new standard-error formula which is used for studentization purposes. We also propose a standard-error estimator that does not require an additional bandwidth choice, and develop valid bandwidth choices compatible with our asymptotic 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.

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

JEL: C12, C14, C21.

---

*For helpful comments we thank Guido Imbens, Michael Jansson, Anna Mikusheva, Whitney Newey, Andres Santos, seminar participants at Boston College, BU, Berkeley, Carleton, Cornell, Harvard, MIT, Rice, Texas A&M, UC-Santa Barbara, and conference participants at the "Junior Festival on New Developments in Microeconometrics" (Northwestern) and at the "Workshop on Inference in Nonstandard Problems" (Princeton/NSF).

†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 quasi-experimental empirical strategies in Economics, Political Science and many other social and behavioral sciences.[1] In this design, units are assigned treatment based on their value of an observed covariate, with the probability of treatment assignment jumping discontinuously at a known cutoff. Nonparametric local polynomial estimators of treatment effects have received great attention in the recent RD literature, becoming a standard choice in both empirical and theoretical work. These kernel-based estimators require a choice of bandwidth for implementation, and several bandwidth selectors are now available in the literature. Most bandwidth selectors are obtained by balancing squared-bias and variance, and typically lead to bandwidth choices that are too "large" for the usual distributional approximations invoked in the literature to be valid.[2] As a consequence, the resulting data-driven confidence intervals for RD treatment effects may be biased, having an 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 applications.

In this paper, we propose new data-driven confidence intervals for RD treatment effects, constructed using bias-corrected local polynomial estimators together with a new standard-error formula. Our innovation is to consider a non-standard asymptotic distributional approximation that accounts for the potential contribution that bias estimation may have to the finite-sample distribution of the statistic of interest. This idea leads naturally to a new standard-error estimate. Our alternative asymptotic approximation allows for "larger" bandwidth choices by virtue of the (possibly inconsistent) bias-correction estimate, permitting in particular a mean-square optimal bandwidth choice. In a simulation study, we find that our new confidence intervals exhibit close-to-correct empirical coverage and good empirical interval length on average, remarkably improving over currently available procedures.

Our proposed inferential framework follows the traditional approach in nonparametric curve estimation of forming confidence intervals using "optimal" bandwidths together with bias-correction (e.g., Fan and Gijbels (1996, Section 4.4)), but departs from it in that it relies on an alternative asymptotic approximation to characterize the sampling distribution of the resulting statistic. The traditional local polynomial bias-correction approach requires choosing two bandwidths: the main bandwidth to construct the original estimator (denoted h_n) and the pilot bandwidth to construct the bias-estimate (denoted b_n). In this classical approach, the bias estimate is required to be consistent for its population counterpart, requiring in particular the condition $h_n/b_n \to 0$. This approach preserves the same limiting distribution that would have been obtained if the leading bias was known, even when h_n is chosen to be optimal in a mean-square sense. The condition $h_n/b_n \to 0$ guarantees that the variability of the bias estimate is "small" relative to the variability of the estimator without bias-correction, ensuring that the bias-correction term has no effect

[1] See, among others, van der Klaauw (2008), Imbens and Lemieux (2008), Lee and Lemieux (2010) and Dinardo and Lee (2011) for recent reviews with emphasis in economics, and Cook (2008) for an interdisciplinary review.

[2] These optimal bandwidth selectors lead to a non-negligible bias in the distributional approximation.

asymptotically. This assumption, however, may not accurately capture the finite-sample behavior of the bias-corrected estimator because the distributional approximation ignores, by construction, the additional variability of the bias-estimate.

Motivated by this potential drawback of bias-correction, we consider a more general asymptotic approximation for the bias-corrected local polynomial estimators where the bias-estimate is not required to be consistent for its population counterpart. We implement this idea by considering the condition $h_n/b_n \to \alpha \in [0, \infty]$, which leads to a different asymptotic variance in general. This asymptotic variance reduces to the classical one when $\alpha = 0$, but is otherwise larger when conventional kernels are employed. The asymptotic experiment leads to new confidence intervals capturing the distributional effects of bias-correction, as we further discuss below.[3] The ideas behind our proposed alternative asymptotic approximation for bias-corrected estimators could also be applied in other nonparametric and semiparametric contexts involving bias-correction.

Our paper contributes to the emerging literature on inference for treatment effects in the RD design. Thistlethwaite-Campbell (1960) proposed the original design, while 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, Dong and Lewel (2012) study marginal treatment effects, and Cattaneo, Frandsen, and Titiunik (2012) propose randomization-inference methods. 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).

The rest of the paper is organized as follows. Section 2 introduces the RD model, reviews known results and motivates our approach. Section 3 presents our main result, while Section 4 discusses standard-error estimation and bandwidth(s) selection. Section 5 reports the results from a simulation study. Section 6 concludes. Technical and other details are collected in the supplemental appendix.

2 Setup, Motivation and Overview

2.1 Model and Estimator

We focus on inference for the average treatment effect in the so-called sharp RD framework (natural extensions are briefly outlined in Section 6). 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 $(X_i \geq 0)$

[3] From a practical point of view, our asymptotics also give theoretical justification to empirical recommendations sometimes found in the literature such as $h_n = \alpha b_n$ with $\alpha \approx 0.8$.

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, in this sharp RD design the observed random sample is $\{(Y_i, X_i)' : i = 1, 2, \dots, n\}$ with $Y_i = Y_i(0) \cdot \mathbf{1}(X_i < 0) + Y_i(1) \cdot \mathbf{1}(X_i \geq 0)$, where $\mathbf{1}(\cdot)$ denotes the indicator function. We set $\mu(x) = \mathbb{E}[Y_i|X_i = x]$ and $\sigma^2(x) = \mathbb{V}[Y_i|X_i = x]$.

The average treatment effect at the threshold is $\tau = \mathbb{E}[Y(1) - Y(0)|X_i = \bar{x}]$. Under mild regularity conditions, this parameter is nonparametrically identifiable as the difference of two conditional expectations evaluated at the (induced) boundary point $\bar{x} = 0$, that is, $\tau = \mu_+ - \mu_-$ with $\mu_+ = \lim_{x \to 0^+} \mu(x)$ and $\mu_- = \lim_{x \to 0^-} \mu(x)$. (We drop the evaluation point of functions whenever possible to simplify notation.) Following Hahn, Todd, and van der Klaauw (2001) and Porter (2003), a popular estimator of τ is constructed using kernel-based local polynomials at either side of the threshold. These estimators are particularly well-suited for inference in the RD design; see, e.g., Fan and Gijbels (1996) and Cheng, Fan, and Marron (1997) for details. The local polynomial RD estimator of order p is

$$\hat{\tau}_p(h_n) = \hat{\mu}_{+,p}(h_n) - \hat{\mu}_{-,p}(h_n), \qquad \hat{\mu}_{+,p}(h_n) = e_0'\hat{\beta}_{+,p}(h_n), \qquad \hat{\mu}_{-,p}(h_n) = e_0'\hat{\beta}_{-,p}(h_n),$$

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

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

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 $p = 2$), $K_h(u) = K(u/h)/h$ with $K(\cdot)$ a kernel function, and h_n is a positive bandwidth sequence. We set $\mu_+^{(s)} = \lim_{x \to 0^+} \partial^s \mu_+(x)/\partial x^s$ with $\mu_+(x) = \mathbb{E}[Y(1)|X_i = x]$, $\mu_-^{(s)} = \lim_{x \to 0^-} \partial^s \mu_-(x)/\partial x^s$ with $\mu_-(x) = \mathbb{E}[Y(0)|X_i = x]$, $s \in \mathbb{N}$, $\sigma_+^2 = \lim_{x \to 0^+} \sigma^2(x)$ and $\sigma_-^2 = \lim_{x \to 0^-} \sigma^2(x)$.

The local-linear RD estimator $\hat{\tau}_1(h_n)$ is the preferred and most common choice in practice. To motivate our approach, in the remainder of this section we first present an overview of the potential limitations of inference based on this estimator, and then briefly outline an example of our proposed alternative. Our discussion in this section is heuristic; Section 3 provides all the details and regularity conditions.

2.2 Local-Linear RD Inference and Its Potential Pitfalls

In the RD design, conventional approaches to inference using the local-linear RD estimator employ the following well-known asymptotic result.

Result 1. If $nh_n^7 \to 0$ and $nh_n \to \infty$, then

$$\sqrt{nh_n}\left(\hat{\tau}_1(h_n) - \tau - h_n^2\,\frac{\mu_+^{(2)} - \mu_-^{(2)}}{2}\,\mathfrak{B}_1\right) \to_d \mathcal{N}(0, \mathsf{V}_1(0)), \qquad \mathsf{V}_1(0) = \frac{\sigma_+^2 + \sigma_-^2}{f}\,\mathfrak{V}_1(0),$$

where $\mathfrak{B}_1$ and $\mathfrak{V}_1(0)$ are fixed constants that depend only on the kernel function.

Conventional confidence intervals for the RD treatment effect τ are directly justified from this result. Specifically, if $nh_n^5 \to 0$ and $nh_n \to \infty$, then

$$T_1(h_n) = \frac{\sqrt{nh_n}\,(\hat{\tau}_1(h_n) - \tau)}{\sqrt{\mathsf{V}_1(0)}} \to_d \mathcal{N}(0,1), \tag{1}$$

leading to the asymptotically valid (but infeasible) confidence interval $\hat{\tau}_1(h_n) \pm q_{1-\alpha/2}\sqrt{\mathsf{V}_1(0)/nh_n}$ with q_α the α-percentile of the $\mathcal{N}(0,1)$. In practice, of course, a standard-error estimator is needed to construct feasible confidence intervals, but in this section we assume $\mathsf{V}_1(0)$ is known for simplicity. Even in this simplified case, the choice of bandwidth h_n is crucial. The bias-condition $nh_n^5 \to 0$ is explicitly imposed to eliminate the contribution to the distributional approximation of the leading bias, which depends on the unknown quantities $\mu_+^{(2)}$ and $\mu_-^{(2)}$. Thus, to construct valid confidence intervals using (1), the bandwidth h_n should be "small" enough to satisfy the bias-condition.

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. Unfortunately, most (if not all) of these approaches lead to bandwidths that are too "large" because they do not satisfy the bias-condition required by the asymptotic theory. For example, minimizing the asymptotic mean-squared error (MSE) of $\hat{\tau}_1(h_n)$ gives the optimal plug-in bandwidth choice $h_{MSE} = C_{MSE}\, n^{-1/5}$ with C_{MSE} a constant, which by construction implies that $n(h_{MSE})^5 \to \alpha \in (0,\infty)$ and hence leads to a first-order bias in the distributional approximation. This is a well-known problem in the nonparametric curve estimation literature. Moreover, implementing this MSE-optimal bandwidth choice in practice is likely to introduce additional variability in the chosen bandwidth that may lead to "large" bandwidths as well. Cross-validation procedures also typically lead to "large" bandwidth choices; see, e.g., Ichimura and Todd (2007) and references therein.

To illustrate the potential pitfalls of the conventional RD confidence intervals based on $T_1(h_n)$, and its data-driven version $T_1(\hat{h}_n)$ with $\hat{h}_n$ a bandwidth estimate, we briefly summarize some results from a Monte Carlo study further discussed in Section 5. 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 employed by Imbens and Kalyanaraman (2012).

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

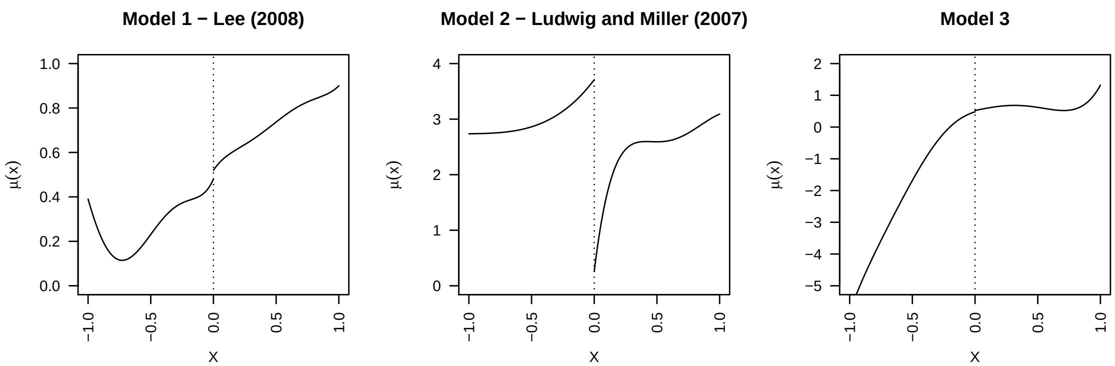

Figure 1: Regression Functions for Models 1–3 in simulations.

Table 1: Empirical Coverage of different 95% Confidence Intervals
(Infeasible Asymptotic Variance)

	Conventional		**Bias-Corrected**		**Robust Approach**		**Bandwidths**	
		EC (%)		EC (%)		EC (%)	h_n	b_n
Model 1								
	$T_1(h_{MSE})$	93.9	$T^{bc}_{1,2}(h_{MSE}, b_{MSE})$	91.4	$T^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	94.7	0.166	0.319
	$T_1(\hat{h}_{IK})$	84.4	$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	77.7	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	93.3	0.335	0.337
	$T_1(\hat{h}_{CV})$	83.1	$T^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	78.1	$T^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	93.1	0.381	0.381
			$T^{bc}_{1,2}(h_{MSE}, h_{MSE})$	81.0	$T^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	94.9	0.166	0.166
			$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.3	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	94.7	0.335	0.335
Model 2								
	$T_1(h_{MSE})$	92.5	$T^{bc}_{1,2}(h_{MSE}, b_{MSE})$	92.5	$T^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	94.9	0.082	0.191
	$T_1(\hat{h}_{IK})$	24.1	$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	85.0	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	91.2	0.185	0.296
	$T_1(\hat{h}_{CV})$	79.1	$T^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	80.5	$T^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	94.8	0.119	0.119
			$T^{bc}_{1,2}(h_{MSE}, h_{MSE})$	79.0	$T^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	94.8	0.082	0.082
			$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.0	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	94.8	0.185	0.185
Model 3								
	$T_1(h_{MSE})$	85.8	$T^{bc}_{1,2}(h_{MSE}, b_{MSE})$	84.7	$T^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	95.0	0.260	0.292
	$T_1(\hat{h}_{IK})$	87.1	$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	88.8	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	95.1	0.231	0.340
	$T_1(\hat{h}_{CV})$	93.9	$T^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	81.9	$T^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	95.2	0.166	0.166
			$T^{bc}_{1,2}(h_{MSE}, h_{MSE})$	81.7	$T^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	94.9	0.260	0.260
			$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.8	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	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.

The simulation results indeed show that the conventional confidence intervals constructed using (1) may have poor empirical coverage. In Models 1 and 2, the infeasible confidence intervals that use the MSE-optimal bandwidth (resp., $h_{MSE} = 0.166$ and $h_{MSE} = 0.082$) have reasonably good empirical coverage (resp., 93.9% and 92.5%), but their data-driven counterparts that employ an estimated bandwidth exhibit substantial undercovering (e.g., for Model 1 the 95% confidence intervals based on $T_1(\hat{h}_{MSE})$ and $T_1(\hat{h}_{CV})$ have empirical coverage of 84.4% and 83.1%, respectively). In Model 3, which has a regression function with more curvature, even the infeasible confidence intervals constructed using the MSE-optimal bandwidth ($h_{MSE} = 0.260$) is biased, showing an empirical coverage of 85.8%.

These simulations illustrate that (1) may not be a good approximation whenever the bandwidth employed is "large". There are two main approaches to deal with this problem. 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.

The other 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 idea is to remove the leading bias term by constructing a plug-in consistent estimator of this term. Specifically, given a "pilot" bandwidth b_n, the local-quadratic estimators of the unknown constants $\mu_+^{(2)}$ and $\mu_-^{(2)}$ in the leading bias are $\hat{\mu}_{+,2}^{(2)}(b_n) = 2e_2'\hat{\beta}_{+,2}(b_n)$ and $\hat{\mu}_{-,q}^{(2)}(b_n) = 2e_2'\hat{\beta}_{-,2}(b_n)$, respectively. Thus, under the conditions of Result 1, and if $h_n/b_n \to 0$ and $nh_n^5 b_n^2 \to 0$, then

$$T_{1,2}^{bc}(h_n, b_n) = \frac{\sqrt{nh_n}\left(\hat{\tau}_1(h_n) - \tau - h_n^2\left(\hat{\mu}_{+,2}^{(2)}(b_n)/2 - \hat{\mu}_{-,2}^{(2)}(b_n)/2\right)\mathfrak{B}_1\right)}{\sqrt{\mathsf{V}_1(0)}} \to_d \mathcal{N}(0,1). \tag{2}$$

This result theoretically allows for "larger" bandwidths h_n, but requires selecting a second bandwidth b_n that should be "larger" relative to h_n (i.e., $h_n/b_n \to 0$). In practice, b_n may also be selected using an MSE-optimal choice, denoted b_{MSE}, which, as discussed in Section 4, can be implemented by a plug-in estimate, denoted $\hat{b}_{MSE}$. The second group of columns in Table 1, labeled "Bias-Corrected", summarizes some simulation results for the 95% confidence intervals based on $T_{1,2}^{bc}(h_n, b_n)$, both infeasible and feasible. In our simulations, these confidence intervals do not exhibit better empirical coverage than the conventional ones based on $T_1(h_n)$, and they do underperform in many cases.

Bias-correction is not particularly popular in empirical work, even though the bias-corrected statistic $T_{1,2}^{bc}(h_n, b_n)$ may be preferred to the classical statistic $T_1(h_n)$ because it has some demonstrable better theoretical (asymptotic) properties. Our simulation results are indeed consistent with the empirical view on conventional bias-correction.

2.3 Our Approach to Local-Linear RD Inference

A potential problem with the conventional large-sample approximation (2) 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 $\hat{\mu}^{(2)}_{-,2}(b_n)$. This 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. Thus, we propose an alternative asymptotic approximation for bias-corrected local polynomial estimators that leads to confidence intervals for RD treatment effects that capture this additional variability. In the special case of the local-linear RD estimator, our alternative asymptotics lead to the following result, which we present here to provide a simplified overview of our approach and to offer a comparison to the conventional approaches discussed above.

Result 2. If $nh_n^5 b_n^2 \to 0$, $h_n/b_n \to \alpha \in [0,\infty)$ and $nh_n \to \infty$, then

$$\sqrt{nh_n}\left(\hat{\tau}_1(h_n) - \tau - h_n^2\left(\hat{\mu}^{(2)}_{+,2}(b_n)/2 - \hat{\mu}^{(2)}_{-,2}(b_n)/2\right)\mathfrak{B}_1\right) \to_d \mathcal{N}(0, \mathsf{V}_1(\alpha)),$$

where $\mathsf{V}_1(\alpha) = \frac{\sigma_+^2+\sigma_-^2}{f}\mathfrak{V}_1(\alpha)$ with $\mathfrak{V}_1(\alpha)$ given in Remark 3 in Section 3.2.

This result is a special case of our Theorem 1 below. It shows that the distributional approximation (2) is only valid when $\alpha = 0$ because $\mathsf{V}_1(\alpha) \neq \mathsf{V}_1(0)$ otherwise. Moreover, because $\mathsf{V}_1(\alpha) > \mathsf{V}_1(0)$ for all $\alpha > 0$ when a conventional kernel $K(\cdot)$ is used, Result 2 also implies that the variability of the bias-corrected estimator will be larger than usual. This suggests that confidence intervals based on $T^{bc}_{1,2}(h_n, b_n)$ will be too short whenever h_n/b_n is not "small", even if the leading bias term is small. This result naturally leads to the "robust" bias-corrected statistic

$$T^{rbc}_{1,2}(h_n, b_n) = \frac{\sqrt{nh_n}\left(\hat{\tau}_1(h_n) - \tau - h_n^2\left(\hat{\mu}^{(2)}_{+,2}(b_n)/2 - \hat{\mu}^{(2)}_{-,2}(b_n)/2\right)\mathfrak{B}_1\right)}{\sqrt{\mathsf{V}_1(h_n/b_n)}} \to_d \mathcal{N}(0,1).$$

The group of columns in Table 1 labeled "Robust Approach" exhibits the performance of the new confidence intervals employing $T^{rbc}_{1,2}(h_n, b_n)$ 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., $\alpha = 1$) because in this special case $T^{rbc}_{1,2}(h_n, h_n)$ 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 MSE-optimal bandwidth for the local-linear RD estimator, but form confidence intervals for τ using the local-quadratic RD estimator. We discuss this and other implications of our approach further below.

3 Robust Bias-Corrected RD Inference

3.1 Assumptions and Notation

We impose the following two assumptions, which are slightly stronger than needed in order to simplify the exposition. Let $\mathcal{N}(x) = (-x, x)$ and $\bar{\mathcal{N}}(x) = [-x, x]$, with $x > 0$.

Assumption 1. On the neighborhood $\mathcal{N}(\kappa_0)$ for some $\kappa_0 > 0$:

(a) $\mathbb{E}[Y_i^4|X_i] < \infty$, and X_i is continuously distributed with density $f(x)$.

(b) $f(x)$ is continuous and bounded away from zero.

(c) $\mu_-(x)$ and $\mu_+(x)$ are S-times continuously differentiable.

(d) $\sigma^2(x)$ is bounded away from zero, bounded, and right and left continuos at $x = 0$.

Assumption 2. For some $\kappa > 0$, $K(\cdot)$ is symmetric, bounded and nonnegative on $\bar{\mathcal{N}}(\kappa)$, and positive and continuous on $\mathcal{N}(\kappa)$.

We introduce 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:

$$\beta_{+,p} = \left[\mu_+, \mu_+^{(1)}, \mu_+^{(2)}/2, \cdots, \mu_+^{(p)}/p!\right]', \quad \beta_{-,p} = \left[\mu_-, \mu_-^{(1)}, \mu_-^{(2)}/2, \cdots, \mu_-^{(p)}/p!\right]',$$

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

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

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

In addition, we define the following (scaled) matrices

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

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

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

and their large sample analogues

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

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

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

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

3.2 Distributional Approximation

The next lemma describes the asymptotic bias, variance and distribution of $\hat{\tau}_p(h_n)$. This result follows from known results in the local polynomial literature applied to the RD context. We present it here to facilitate the upcoming discussion, and to introduce useful notation. (Qualifiers such as "for n large enough" or "with probability approaching 1" are omitted to save space.)

Lemma 1. Suppose Assumptions 1–2 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 $\mathbb{E}[\hat{\tau}_p(h_n)|\mathcal{X}_n] = \tau + h_n^{p+1}\mathsf{B}_{p,0,p+1}(h_n) + O_p(h_n^{p+2})$, where

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

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

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

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

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

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

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

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

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

This lemma formalizes Result 1 in the previous section for the local-linear RD estimator $\hat{\tau}_1(h_n)$. Specifically, if $nh_n^{2p+3} \to 0$ and $nh_n \to \infty$, then

$$T_p(h_n) = \frac{\hat{\tau}_p(h_n) - \tau}{\sqrt{\mathsf{V}_{p,0}(h_n)}} \to_d \mathcal{N}(0,1). \tag{3}$$

This is the distributional approximation typically employed in empirical work to construct confidence intervals for the treatment effect τ in the RD design, usually with $p = 1$ and with a conventional Huber-Eicker-White-type standard-error estimate for studentization purposes. This approach assumes away the effect of the leading bias term by restricting the bandwidth behavior: it requires a "small" enough bandwidth ($nh_n^{2p+3} \to 0$) so that the leading bias is negligible.

We investigate instead the distributional properties of the bias-corrected RD estimator

$$\hat{\tau}^{bc}_{p,q}(h_n, b_n) = \hat{\tau}_p(h_n) - h_n^{p+1}\hat{\mathsf{B}}_{p,0,q,p+1}(h_n, b_n),$$

where, for some $q \geq p+1$,

$$\hat{\mathsf{B}}_{p,0,q,p+1}(h_n, b_n) = \left(e'_{p+1}\hat{\beta}_{+,q}(b_n)\right)\mathcal{B}_{+,p,0,p+1}(h_n) - \left(e'_{p+1}\hat{\beta}_{-,q}(b_n)\right)\mathcal{B}_{-,p,0,p+1}(h_n)$$

and 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 converge rate of the resulting estimator.

The conventional approach to bias-correction assumes that the bias-estimate is a consistent estimator of the asymptotic bias, and thus relies on Lemma 1 to characterize the asymptotic distribution of $\hat{\tau}^{bc}_{p,q}(h_n, b_n)$. Specifically, under the conditions of Lemma 1 ($nh_n^{2p+5} \to 0$ and $nh_n \to \infty$), if

$$nh_n^{2p+3}\left(\hat{\mathsf{B}}_{p,0,q,p+1}(h_n, b_n) - \mathsf{B}_{p,0,p+1}(h_n)\right)^2 \to_p 0 \tag{4}$$

then

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

This approach allows for potentially "larger" bandwidths h_n because the leading asymptotic bias is manually removed from the distributional approximation. However, as discussed above, the resulting distributional approximation for this bias-corrected estimator tends to provide a poor characterization of the finite sample variability of the statistic. A potential drawback of this approach is that the approximation does not account for the bias-correction component: condition (4) makes researchers proceed "as if" the leading bias is known, thereby obtaining the distributional approximation directly from Lemma 1(D). In finite samples, however, the bias-correction component will (at least partially) affect the sampling distribution of the estimator $\hat{\tau}^{bc}_{p,q}(h_n, b_n)$, which implies that the conventional distributional approximation may not accurately represent the finite-sample distribution of $T^{bc}_{p,q}(h_n, b_n)$.

This is the key motivation for our paper. 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 $T^{bc}_{p,q}(h_n, b_n)$. The idea is to allow for bandwidth sequences, entering in the bias-estimate $\hat{\mathsf{B}}_{p,0,q,p+1}(h_n, b_n)$, that potentially make the bias-correction term in $\hat{\tau}^{bc}_{p,q}(h_n, b_n)$ as important as the main term $\hat{\tau}_p(h_n)$, even asymptotically. These bandwidth sequences weaken the condition (4) in an intuitive way, and lead to an alternative distributional approximation for $\hat{\tau}^{bc}_{p,q}(h_n, b_n)$ with a different asymptotic variance in general. The resulting distributional approximation therefore potentially includes both the contribution of the main estimator $\hat{\tau}_p(h_n)$ as well as the contribution of the bias-correction estimate.

The intuition behind our result is quite simple. We have

$$\sqrt{nh_n}\left(\hat{\tau}^{bc}_{p,q}(h_n,b_n)-\tau\right)=\Upsilon_p(h_n)-\Upsilon_{p,q}(h_n,b_n),$$

where

$$\Upsilon_p(h_n)=\sqrt{nh_n}\left(\hat{\tau}_p(h_n)-\tau-h_n^{p+1}\mathsf{B}_{p,0,p+1}(h_n)\right),$$

$$\Upsilon_{p,q}(h_n,b_n)=\sqrt{nh_n^{2p+3}}\left(\hat{\mathsf{B}}_{p,0,q,p+1}(h_n,b_n)-\mathsf{B}_{p,0,p+1}(h_n)\right).$$

Lemma 1 implies that $\Upsilon_p(h_n)$ converges in distribution. Under appropriate conditions,

$$\Upsilon_{p,q}(h_n,b_n)=\sqrt{nh_n^{2p+3}}O_p\left(\frac{1}{\sqrt{nb_n^{2p+3}}}+b_n^{q-p}\right)=O_p\left(\left(\frac{h_n}{b_n}\right)^{p+3/2}+\sqrt{nh_n^{2p+3}b_n^{2(q-p)}}\right),$$

implying that $\Upsilon_{p,q}(h_n,b_n)$ is asymptotically negligible if (and only if)

$$\frac{h_n}{b_n}\to 0 \qquad \text{and} \qquad nh_n^{2p+3}b_n^{2(q-p)}\to 0. \tag{5}$$

The conditions in (5) specialize the high-level condition (4) 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 $\hat{\tau}^{bc}_{p,q}(h_n,b_n)$, while the condition $nh_n^{2p+3}b_n^{2(q-p)}\to 0$ ensures that the bias-correction term is asymptotically unbiased after proper scaling. 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 $\hat{\tau}^{bc}_{p,q}(h_n,b_n)$ based on the condition

$$\alpha_n=\frac{h_n}{b_n}\to\alpha\in[0,\infty],$$

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, provided the asymptotic bias is small enough.

This idea is formalized in the following theorem. We employ the following additional notation

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

with, in particular, $\Psi_{+,p,p}(h,h)=\Psi_{+,p}(h)$ and $\Psi_{-,p,p}(h,h)=\Psi_{-,p}(h)$.

Theorem 1. Suppose Assumptions 1–2 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

$$\mathbb{E}[\hat{\tau}^{bc}_{p,q}(h_n,b_n)|\mathcal{X}_n]=\tau+h_n^{p+2}\mathsf{B}_{p,0,p+2}(h_n)\;[1+o_p(1)]-h_n^{p+1}b_n^{q-p}\mathsf{B}^{bc}_{p,0,q,p+1}(h_n,b_n)\;[1+o_p(1)],$$

where

$$\mathsf{B}^{bc}_{p,0,q,p+1}(h,b) = \frac{\mu_+^{(q+1)}}{(q+1)!}\mathcal{B}_{+,q,p+1,q+1}(b)\mathcal{B}_{+,p,0,p+1}(h) - \frac{\mu_-^{(q+1)}}{(q+1)!}\mathcal{B}_{-,q,p+1,q+1}(b)\mathcal{B}_{-,p,0,p+1}(h).$$

(V) If $n\min\{h_n, b_n\} \to \infty$, then $\mathbb{V}[\hat{\tau}^{bc}_{p,q}(h_n,b_n)|\mathcal{X}_n] = \mathsf{V}^{bc}_{p,0,q,p+1}(h_n,b_n)$, where

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

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

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

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

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

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

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

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

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

Theorem 1 shows that by standardizing the bias-corrected estimator by its (conditional) variance, the asymptotic distribution of the resulting bias-corrected statistic $T^{rbc}_{p,q}(h_n,b_n)$ 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 $\hat{\tau}^{bc}_{p,q}(h_n,b_n)$ in general, which depends on the behavior of $\alpha_n = h_n/b_n$.

Remark 1. The distributional approximation in Theorem 1(D) 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 $\alpha_n \to \alpha \in [0,\infty]$. (A fourth case also covered by the asymptotics is $\underline{\lim}_{n\to\infty}\alpha_n < \overline{\lim}_{n\to\infty}\alpha_n$.)

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

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

which is the classical approach to bias-correction.

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

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

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

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

$$\begin{aligned}
\mathbb{V}[\hat{\tau}^{bc}_{p,q}(h_n, b_n)|\mathcal{X}_n] &= h_n^{2p+2}\mathbb{V}[\hat{\mathsf{B}}_{p,0,q,p+1}(h_n, b_n)|\mathcal{X}_n]\{1 + o_p(1)\} \\
&= \frac{\alpha_n^{2p+2}}{nb_n}\frac{\sigma^2_+ + \sigma^2_-}{f}(e_p'\Gamma_q^{-1}\Psi_q\Gamma_q^{-1}e_p)(e_0'\Gamma_p^{-1}\vartheta_p)^2\{1 + o_p(1)\},
\end{aligned}$$

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

Remark 3. If $h_n = b_n$ (and the same kernel function $K(\cdot)$ is used), then $\hat{\tau}^{bc}_{p,p+1}(h_n, h_n) = \hat{\tau}_{p+1}(h_n)$. 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 $T^{rbc}_{p,p+1}(h_n, h_n) = T_{p+1}(h_n)$. The result extends to $\hat{\tau}^{bc}_{p,p+r}(h_n, h_n) = \hat{\tau}_{p+r}(h_n)$ and $T^{rbc}_{p,p+r}(h_n, h_n) = T_{p+r}(h_n)$ when the natural generalization of the bias-correction estimate $\hat{\mathsf{B}}_{p,0,q,r}(h_n, b_n)$ is used. See the supplemental appendix for details.

Remark 4. 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 $\alpha_n = h_n/b_n = 1$ in Theorem 1.

Remark 5. 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 $\hat{\tau}^{bc}_{+,p,q}(h_n, b_n) = \hat{\mu}_{+,p}(h_n) - h_n^{p+1}(e_{p+1}'\hat{\beta}_{+,q}(b_n))\mathcal{B}_{+,p,0,p+1}(h_n)$. Letting $h_n = \alpha b_n$ with $\alpha \in (0, \infty)$,

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

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

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

(i) Because $h_n = \alpha b_n$, the asymptotics in Theorem 1 "convexify" the kernel function employed to construct the estimator $\hat{\tau}^{bc}_{+,p,q}(h_n, b_n)$, because

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

(ii) The asymptotic bias and variance reduce to

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

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

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

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

Remark 6. 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 $\mathcal{K}_{p-1,p}(x;\alpha)$ dominates $\mathcal{K}_p(x)$ for an appropriate choice of α, when a conventional kernel $K(\cdot)$ is used. (For $\alpha > 0$, $\mathcal{K}_{p-1,p}(x;\alpha)$ is also a boundary kernel of order p or larger.) See the supplemental appendix for further details.

Remark 7. All the results in this paper extend immediately when different bandwidths ($h_{+,n}$, $h_{-,n}$, $b_{+,n}$, $b_{-,n}$, say) are employed to construct the estimators $\hat{\beta}_{+,p}(h_{+,n})$, $\hat{\beta}_{-,p}(h_{-,n})$ and their associated bias-correction terms.

4 Implementation Issues

We discuss valid standard-error estimates and bandwidth choices to implement in applications the infeasible statistics discussed in the previous sections: $T_p(h_n)$, $T^{bc}_{p,q}(h_n, b_n)$, $T^{rbc}_{p,q}(h_n, b_n)$.

4.1 Standard Errors

The only unknown matrix in $\mathsf{V}_{p,0}(h_n)$ and $\mathsf{V}^{bc}_{p,0,q,p+1}(h_n, b_n)$ is $\Sigma = \mathbb{E}[\varepsilon\varepsilon'|\mathcal{X}_n]$. This motivates the Huber-Eicker-White heteroskedasticity-robust analogue estimators

$$\hat{\Psi}_{+,p,q}(h_n, b_n) = X_p(h_n)'W_+(h_n)\hat{\Sigma}_+W_+(b_n)X_q(b_n)/n,$$

$$\hat{\Psi}_{-,p,q}(h_n, b_n) = X_p(h_n)'W_-(h_n)\hat{\Sigma}_- W_-(b_n)X_q(b_n)/n,$$

for an appropriate choice of $\hat{\Sigma}_+$ and $\hat{\Sigma}_+$. A popular choice is to use the estimators

$$\hat{\Sigma}_+(c_n) = \text{diag}(\hat{\varepsilon}^2_{+,1}(c_n), \cdots, \hat{\varepsilon}^2_{+,n}(c_n)), \qquad \hat{\varepsilon}_{+,i}(c_n) = \mathbf{1}(X_i \geq 0)(Y_i - \hat{\mu}_{+,p}(c_n)),$$

$$\hat{\Sigma}_-(c_n) = \text{diag}(\hat{\varepsilon}^2_{-,1}(c_n), \cdots, \hat{\varepsilon}^2_{-,n}(c_n)), \qquad \hat{\varepsilon}_{-,i}(c_n) = \mathbf{1}(X_i < 0)(Y_i - \hat{\mu}_{-,p}(c_n)),$$

where the estimators $\hat{\mu}_{+,p}(c_n)$ and $\hat{\mu}_{-,p}(c_n)$ employ a possibly different bandwidth c_n. In practice, typically $c_n = h_n$ because this standard-error estimator is automatically obtained from computing the RD estimator $\hat{\tau}_{p+1}(h_n)$ as a weighted least-squares problem. This choice, although simple and convenient, may be suboptimal for the object of interest and may not perform well in finite-samples. For further discussion and other proposals see, e.g., Fan and Gijbels (1996, Section 4.4).

Following Abadie and Imbens (2006), we propose an alternative standard-error estimator based on nearest-neighbor estimators with a fixed tuning parameter, which may be more robust in finite-samples because they do not require an additional bandwidth choice. We define

$$\hat{\Sigma}_+ = \text{diag}(\hat{\sigma}^2_+(X_1), \cdots, \hat{\sigma}^2_+(X_n)), \qquad \hat{\sigma}^2_+(X_i) = \mathbf{1}(X_i \geq 0)\frac{J}{J+1}\left(Y_i - \frac{1}{J}\sum_{j=1}^{J} Y_{\ell_{+,j}(i)}\right)^2,$$

$$\hat{\Sigma}_- = \text{diag}(\hat{\sigma}^2_-(X_1), \cdots, \hat{\sigma}^2_-(X_n)), \qquad \hat{\sigma}^2_-(X_i) = \mathbf{1}(X_i < 0)\frac{J}{J+1}\left(Y_i - \frac{1}{J}\sum_{j=1}^{J} Y_{\ell_{-,j}(i)}\right)^2,$$

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 $\{X_i : X_i < 0\}$. 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).

Theorem 2. Suppose the conditions in Theorem 1(D) hold. If $\sigma^2(x)$ is Lipschitz continuos on $(-\kappa_0, 0)$ and on $[0, \kappa_0)$, then

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

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

Therefore, under the conditions stated in Theorems 1 and 2, we obtain

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

where $\hat{\mathsf{V}}^{bc}_{p,\nu,q,p+1}(h_n, b_n)$ is given by $\mathsf{V}^{bc}_{p,\nu,q,p+1}(h_n, b_n)$ but with $\Psi_{+,p,q}(h_n, b_n)$ and $\Psi_{+,p,q}(h_n, b_n)$ replaced by the nearest-neighbor-based plug-in estimators $\hat{\Psi}_{+,p,q}(h_n, b_n)$ and $\hat{\Psi}_{-,p,q}(h_n, b_n)$, respectively. Similarly, under the conditions stated in Lemma 1 and Theorem 2,

$$\hat{T}_p(h_n) = \frac{\hat{\tau}_p(h_n) - \tau}{\sqrt{\hat{\mathsf{V}}_{p,0}(h_n)}} \to_d \mathcal{N}(0,1),$$

where $\hat{\mathsf{V}}_{p,\nu}(h)$ is given by $\mathsf{V}_{p,\nu}(h_n)$ but with $\Psi_{+,p}(h_n)$ and $\Psi_{+,p}(h_n)$ replaced by the nearest-neighbor-based plug-in estimators $\hat{\Psi}_{+,p}(h_n)$ and $\hat{\Psi}_{-,p}(h_n)$, respectively.

4.2 Choice of Bandwidths

We derive MSE-optimal bandwidth choices for h_n and b_n, and propose direct plug-in, data-driven bandwidth selectors. For a recent review on bandwidth selection in the RD design see Imbens and Kalyanaraman (2012). In the next section, we explore in a simulation study the performance of these bandwidth selectors as well as several alternatives available in the literature. Define

$$\mathsf{MSE}_{p,\nu}(h_n) = \mathbb{E}\left[\left(e_\nu'(\hat{\beta}_{+,p}(h_n) - \hat{\beta}_{+,p}(h_n)) - e_\nu'(\beta_{+,p} - \beta_{-,p}) \right)^2 \middle| \mathcal{X}_n \right].$$

Lemma 2. Suppose Assumptions 1–2 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

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

where

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

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

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

Using this lemma, we construct direct plug-in (DPI) selectors for h_n and b_n. Our construction employs the fact that $\mathsf{V}_{p,\nu} = \lim_{n\to\infty} nh_n^{1+2\nu}\mathsf{V}_{p,\nu}(h_n)$ if $h_n \to 0$ and $nh_n \to \infty$, together with Theorem 2, to construct consistent plug-in estimates of the variance terms. This 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 supplemental appendix contains a detailed discussion of our approach, and a comparison to other methods available in the RD literature.

Plug-in Bandwidths Selectors. Fix $p, q \in \mathbb{N}$ with $q \geq p+1$, and recall that $\hat{\mathsf{V}}_{p,\nu}(h)$ is the estimator described in Section 4.1. Let $\mathcal{B}_{p,\nu} = e_\nu' \Gamma_p^{-1} \vartheta_{p,p+1}$.

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

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

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 $\hat{b}_{q,p+1} = \widehat{C}_{q,p+1}\ n^{-1/(2q+3)}$ with

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

Step 2: Main Bandwidth h_n. Let $b_n = \hat{b}_{q,p+1}$, and compute $\hat{h}_{p,0} = \widehat{C}_{p,0}\ n^{-1/(2p+3)}$ with

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

The selectors $\hat{h}_{p,0}$ and $\hat{b}_{q,p+1}$ 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 3. Suppose Assumptions 1–2 hold with $S \geq q+1$ and $q \geq p+1$. In addition, suppose $e'_{q+2}\check{\beta}_{+,q+2} - e'_{q+2}\check{\beta}_{-,q+2} \to_p c \neq 0$.

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

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

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

$$\frac{\hat{h}_{p,0}}{h_{MSE,p,0}} \to_p 1 \qquad \text{and} \qquad \frac{\mathsf{MSE}_{p,0}(\hat{h}_{p,0})}{\mathsf{MSE}_{p,0}(h_{MSE,p,0})} \to_p 1.$$

Remark 8. The MSE-optimal bandwidth choices $h_{MSE,p,0}$ and $b_{MSE,q,p+1}$ (and their feasible versions $\hat{h}_{p,0}$ and $\hat{b}_{q,p+1}$) are fully compatible with our asymptotic approximations given above, as they satisfy the rate-restrictions in Theorems 1 and 2: $n \min\{h_{MSE,p,0}, b_{MSE,q,p+1}\} \to \infty$, $n \min\{h_{MSE,p,0}^{2p+3}, b_{MSE,q,p+1}^{2p+3}\} \max\{h_{MSE,p,0}^2, b_{MSE,q,p+1}^{2(q-p)}\} \to 0$.

Remark 9. The MSE-optimal bandwidth choices satisfy $\alpha_n = h_{MSE,p,0}/b_{MSE,q,p+1} \to 0$. It remains an open question whether the choice $\alpha_n \to 0$ is optimal from a distributional approximation point of view. Although beyond the scope of this paper, research on this question is underway.

5 Simulation Evidence

We explored the main implications of our theoretical results in a Monte Carlo experiment. To facilitate the comparison, we employed the data generating process proposed in Imbens and Kalyanaraman (2012, henceforth IK). 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 supplemental appendix.

For concreteness, we focused on local-linear RD estimators, $p = 1$, with local-quadratic bias-correction, $q = 2$. We investigated the empirical coverage and interval length of the following three competing 95% confidence intervals for a variety of possible bandwidth choices:

$$\begin{aligned}
\text{“Conventional” } (\hat{T}_1(h_n)) &: \quad \hat{\tau}_p(h_n) \pm 1.96 \cdot \sqrt{\hat{\mathsf{V}}_{p,0}(h_n)}, \\
\text{“Bias-Corrected” } (\hat{T}_{1,2}^{bc}(h_n, b_n)) &: \quad \hat{\tau}_{p,q}^{bc}(h_n, b_n) \pm 1.96 \cdot \sqrt{\hat{\mathsf{V}}_{p,0}(h_n)}, \\
\text{“Robust Approach” } (\hat{T}_{1,2}^{rbc}(h_n, b_n)) &: \quad \hat{\tau}_{p,q}^{bc}(h_n, b_n) \pm 1.96 \cdot \sqrt{\hat{\mathsf{V}}_{p,0,q,p+1}^{bc}(h_n, b_n)},
\end{aligned}$$

where the estimators $\hat{\mathsf{V}}_{p,0}(h_n)$ and $\hat{\mathsf{V}}_{p,0,q,p+1}^{bc}(h_n, b_n)$ are constructed using the nearest-neighbor procedure discussed in Section 4.1 with $J = 3$. For comparison, we also report the infeasible versions of these confidence intervals employing $\mathsf{V}_{p,0}(h_n)$ and $\mathsf{V}^{bc}_{p,0,q,p+1}(h_n, b_n)$.

To choose the main bandwidth h_n we consider the following alternatives: (i) the infeasible MSE-optimal choice $h_{MSE,p,0}$, denoted h_{MSE}; (ii) a plug-in, regularized MSE-optimal selector proposed by IK, denoted $\hat{h}_{IK}$; (iii) the infeasible, each-side-squared MSE-optimal choice proposed by DesJardins and McCall (2009), denoted h_{DM}; (iv) a plug-in, each-side-squared MSE-optimal selector, denoted $\hat{h}_{DM}$; (v) a cross-validation estimator proposed by Ludwig and Miller (2007), denoted $\hat{h}_{CV}$; and (vi) our plug-in choice proposed in Section 4.2, denoted $\hat{h}_{CCT}$. Similarly, to choose the pilot bandwidth b_n, we constructed the appropriately modified versions of the choices enumerated above, with the exception of $\hat{h}_{CV}$ because it is not available for derivative estimation, which are denoted b_{MSE}, $\hat{b}_{IK}$, b_{DM}, $\hat{b}_{DM}$, and $\hat{b}_{CCT}$, respectively. The 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 $\mathsf{V}_{p,0}(h_n)$ and $\mathsf{V}^{bc}_{p,0,q,p+1}(h_n, b_n)$, while Table 3 employs the fully data-driven standard-errors $\hat{\mathsf{V}}_{p,0}(h_n)$ and $\hat{\mathsf{V}}^{bc}_{p,0,q,p+1}(h_n, b_n)$. 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 supplemental appendix we also report results employing the traditional standard-error estimators constructed using $\hat{\Sigma}_+(h_n)$ and $\hat{\Sigma}_-(h_n)$ as in Section 4.2, which led 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 $\alpha_n = 1$ 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 $T^{rbc}_{1,2}(h_n, b_n)$ exhibit 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.

6 Final Remarks

We introduced an alternative asymptotic theory for bias-corrected local polynomial estimators of average treatment effects in the context of the sharp RD design. This distributional approximation leads to a different asymptotic variance in general, which we use to propose a new standard-error estimator. We found that the resulting data-driven confidence intervals performed very well in simulations, suggesting in particular that they provide a "robust" (to the choice of bandwidths) alternative when compared to the conventional confidence intervals routinely employ in empirical work.

In the specific context of the RD design, our results could be extended to the case of (i) quantile treatment effects (replace Y_i by $\mathbf{1}(Y_i \leq y)$ as in Frandsen, Frölich, and Melly (2012)), (ii) marginal treatment effects (replace $\hat{\tau}_{p,0}(h_n) = e_0'\hat{\beta}_{+,p}(h_n) - e_0'\hat{\beta}_{-,p}(h_n)$ by $\hat{\tau}_{p,1}(h_n) = e_1'\hat{\beta}_{+,p}(h_n) - e_1'\hat{\beta}_{-,p}(h_n)$ as in Dong and Lewel (2012)), and (iii) mean, quantiles and marginal treatment effects in the so-called fuzzy design (as discussed, e.g., in Porter (2003)). To conserve space, we do not provide details for these extensions here but we plan to address them in upcoming work (Calonico, Cattaneo, and Titiunik (2012)).

From a more general perspective, the main idea in this paper is to consider tuning parameter sequences that allow for a bias-correction estimate to be of (at least) the same order than the original (uncorrected) estimator. Thus, we expect our approach to also be applicable to other problems where non-parametric and semi-parametric statistics involving bias-correction are employed. Further research along this line in underway.

References

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

Calonico, S. F., M. D. Cattaneo, and R. Titiunik (2012): "Robust Inference in the Regression-Discontinuity using STATA," work in progress, University of Michigan.

Cattaneo, M. D., B. Frandsen, and R. Titiunik (2012): "Randomization Inference in the Regression Discontinuity Design," 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.

Cook, T. D. (2008): ""Waiting for Life to Arrive": A history of the regression-discontinuity design in Psychology, Statistics and Economics," *Journal of Econometrics*, 142(2), 636–654.

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.

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

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

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

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.

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," forthcoming in *Review of Economic Studies.*

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 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.

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-Campbell, J. (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											
	$T_1(h_{MSE})$	93.9	0.225	$T^{bc}_{1,2}(h_{MSE}, b_{MSE})$	91.4	0.225	$T^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	94.7	0.254	0.166	0.319
	$T_1(h_{DM})$	93.4	0.213	$T^{bc}_{1,2}(h_{DM}, b_{DM})$	88.7	0.213	$T^{rbc}_{1,2}(h_{DM}, b_{DM})$	94.8	0.261	0.184	0.271
	$T_1(\hat{h}_{IK})$	84.4	0.159	$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	77.7	0.159	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	93.3	0.244	0.335	0.337
	$T_1(\hat{h}_{DM})$	79.6	0.137	$T^{bc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	63.8	0.137	$T^{rbc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	89.8	0.267	0.496	0.423
	$T_1(\hat{h}_{CV})$	83.1	0.152	$T^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	78.1	0.152	$T^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	93.1	0.224	0.381	0.381
	$T_1(\hat{h}_{CCT})$	91.2	0.205	$T^{bc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	88.0	0.205	$T^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	93.0	0.242	0.205	0.336
				$T^{bc}_{1,2}(h_{MSE}, h_{MSE})$	81.0	0.225	$T^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	94.9	0.338	0.166	0.166
				$T^{bc}_{1,2}(h_{DM}, h_{DM})$	81.3	0.213	$T^{rbc}_{1,2}(h_{DM}, h_{DM})$	95.1	0.319	0.184	0.184
				$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.3	0.159	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	94.7	0.235	0.335	0.335
				$T^{bc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	71.2	0.137	$T^{rbc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	89.9	0.200	0.496	0.496
				$T^{bc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	81.6	0.205	$T^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	95.3	0.306	0.205	0.205
Model 2											
	$T_1(h_{MSE})$	92.5	0.327	$T^{bc}_{1,2}(h_{MSE}, b_{MSE})$	92.5	0.327	$T^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	94.9	0.354	0.082	0.191
	$T_1(h_{DM})$	92.2	0.323	$T^{bc}_{1,2}(h_{DM}, b_{DM})$	92.5	0.323	$T^{rbc}_{1,2}(h_{DM}, b_{DM})$	94.9	0.352	0.084	0.190
	$T_1(\hat{h}_{IK})$	24.1	0.213	$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	85.0	0.213	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	91.2	0.254	0.185	0.296
	$T_1(\hat{h}_{DM})$	14.8	0.206	$T^{bc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	82.8	0.206	$T^{rbc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	89.1	0.245	0.196	0.319
	$T_1(\hat{h}_{CV})$	79.1	0.269	$T^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	80.5	0.269	$T^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	94.8	0.411	0.119	0.119
	$T_1(\hat{h}_{CCT})$	87.6	0.300	$T^{bc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	92.0	0.300	$T^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	94.2	0.325	0.097	0.226
				$T^{bc}_{1,2}(h_{MSE}, h_{MSE})$	79.0	0.327	$T^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	94.8	0.513	0.082	0.082
				$T^{bc}_{1,2}(h_{DM}, h_{DM})$	79.2	0.323	$T^{rbc}_{1,2}(h_{DM}, h_{DM})$	94.8	0.505	0.084	0.084
				$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.0	0.213	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	94.8	0.319	0.185	0.185
				$T^{bc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	80.8	0.206	$T^{rbc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	94.6	0.309	0.196	0.196
				$T^{bc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	79.6	0.300	$T^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	94.8	0.464	0.097	0.097
Model 3											
	$T_1(h_{MSE})$	85.8	0.179	$T^{bc}_{1,2}(h_{MSE}, b_{MSE})$	84.7	0.179	$T^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	95.0	0.246	0.260	0.292
	$T_1(h_{DM})$	87.4	0.182	$T^{bc}_{1,2}(h_{DM}, b_{DM})$	86.1	0.182	$T^{rbc}_{1,2}(h_{DM}, b_{DM})$	95.0	0.241	0.251	0.305
	$T_1(\hat{h}_{IK})$	87.1	0.190	$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	88.8	0.190	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	95.1	0.235	0.231	0.340
	$T_1(\hat{h}_{DM})$	91.4	0.200	$T^{bc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	91.4	0.200	$T^{rbc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	95.2	0.229	0.209	0.390
	$T_1(\hat{h}_{CV})$	93.9	0.226	$T^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	81.9	0.226	$T^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	95.2	0.340	0.166	0.166
	$T_1(\hat{h}_{CCT})$	92.1	0.217	$T^{bc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	91.4	0.217	$T^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	95.2	0.250	0.181	0.322
				$T^{bc}_{1,2}(h_{MSE}, h_{MSE})$	81.7	0.179	$T^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	94.9	0.266	0.260	0.260
				$T^{bc}_{1,2}(h_{DM}, h_{DM})$	81.8	0.182	$T^{rbc}_{1,2}(h_{DM}, h_{DM})$	95.0	0.271	0.251	0.251
				$T^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.8	0.190	$T^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	95.0	0.283	0.231	0.231
				$T^{bc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	81.8	0.200	$T^{rbc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	94.9	0.298	0.209	0.209
				$T^{bc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	82.0	0.217	$T^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	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											
	$\hat{T}_1(h_{MSE})$	92.3	0.223	$\hat{T}^{bc}_{1,2}(h_{MSE}, b_{MSE})$	89.7	0.223	$\hat{T}^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	93.4	0.251	0.166	0.319
	$\hat{T}_1(h_{DM})$	92.0	0.211	$\hat{T}^{bc}_{1,2}(h_{DM}, b_{DM})$	87.2	0.211	$\hat{T}^{rbc}_{1,2}(h_{DM}, b_{DM})$	93.4	0.259	0.184	0.271
	$\hat{T}_1(\hat{h}_{IK})$	83.2	0.158	$\hat{T}^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	76.7	0.158	$\hat{T}^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	91.8	0.241	0.335	0.337
	$\hat{T}_1(\hat{h}_{DM})$	78.7	0.136	$\hat{T}^{bc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	63.3	0.136	$\hat{T}^{rbc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	88.6	0.265	0.496	0.423
	$\hat{T}_1(\hat{h}_{CV})$	81.8	0.151	$\hat{T}^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	77.4	0.151	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	91.5	0.222	0.381	0.381
	$\hat{T}_1(\hat{h}_{CCT})$	89.9	0.201	$\hat{T}^{bc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	86.7	0.201	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	91.7	0.238	0.205	0.336
				$\hat{T}^{bc}_{1,2}(h_{MSE}, h_{MSE})$	79.4	0.223	$\hat{T}^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	92.3	0.332	0.166	0.166
				$\hat{T}^{bc}_{1,2}(h_{DM}, h_{DM})$	79.7	0.211	$\hat{T}^{rbc}_{1,2}(h_{DM}, h_{DM})$	92.6	0.313	0.184	0.184
				$\hat{T}^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	80.4	0.158	$\hat{T}^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	93.4	0.232	0.335	0.335
				$\hat{T}^{bc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	70.7	0.136	$\hat{T}^{rbc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	88.6	0.198	0.496	0.496
				$\hat{T}^{bc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	80.1	0.201	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	92.6	0.298	0.205	0.205
Model 2											
	$\hat{T}_1(h_{MSE})$	91.2	0.354	$\hat{T}^{bc}_{1,2}(h_{MSE}, b_{MSE})$	91.6	0.354	$\hat{T}^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	93.5	0.384	0.082	0.191
	$\hat{T}_1(h_{DM})$	91.1	0.349	$\hat{T}^{bc}_{1,2}(h_{DM}, b_{DM})$	91.5	0.349	$\hat{T}^{rbc}_{1,2}(h_{DM}, b_{DM})$	93.6	0.381	0.084	0.190
	$\hat{T}_1(\hat{h}_{IK})$	28.0	0.224	$\hat{T}^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	85.4	0.224	$\hat{T}^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	91.0	0.268	0.185	0.296
	$\hat{T}_1(\hat{h}_{DM})$	17.9	0.216	$\hat{T}^{bc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	83.2	0.216	$\hat{T}^{rbc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	89.1	0.258	0.196	0.319
	$\hat{T}_1(\hat{h}_{CV})$	79.7	0.287	$\hat{T}^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	81.3	0.287	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	93.4	0.450	0.119	0.119
	$\hat{T}_1(\hat{h}_{CCT})$	87.5	0.318	$\hat{T}^{bc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	91.1	0.318	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	93.2	0.345	0.097	0.226
				$\hat{T}^{bc}_{1,2}(h_{MSE}, h_{MSE})$	79.5	0.354	$\hat{T}^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	92.9	0.567	0.082	0.082
				$\hat{T}^{bc}_{1,2}(h_{DM}, h_{DM})$	79.7	0.349	$\hat{T}^{rbc}_{1,2}(h_{DM}, h_{DM})$	93.0	0.559	0.084	0.084
				$\hat{T}^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	81.5	0.224	$\hat{T}^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	94.0	0.344	0.185	0.185
				$\hat{T}^{bc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	81.4	0.216	$\hat{T}^{rbc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	94.0	0.332	0.196	0.196
				$\hat{T}^{bc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	80.3	0.318	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	93.1	0.504	0.097	0.097
Model 3											
	$\hat{T}_1(h_{MSE})$	84.7	0.178	$\hat{T}^{bc}_{1,2}(h_{MSE}, b_{MSE})$	83.7	0.178	$\hat{T}^{rbc}_{1,2}(h_{MSE}, b_{MSE})$	93.5	0.244	0.260	0.292
	$\hat{T}_1(h_{DM})$	86.4	0.181	$\hat{T}^{bc}_{1,2}(h_{DM}, b_{DM})$	84.9	0.181	$\hat{T}^{rbc}_{1,2}(h_{DM}, b_{DM})$	93.7	0.239	0.251	0.305
	$\hat{T}_1(\hat{h}_{IK})$	86.1	0.189	$\hat{T}^{bc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	87.6	0.189	$\hat{T}^{rbc}_{1,2}(\hat{h}_{IK}, \hat{b}_{IK})$	93.8	0.233	0.231	0.340
	$\hat{T}_1(\hat{h}_{DM})$	90.0	0.198	$\hat{T}^{bc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	89.8	0.198	$\hat{T}^{rbc}_{1,2}(\hat{h}_{DM}, \hat{b}_{DM})$	93.7	0.227	0.209	0.390
	$\hat{T}_1(\hat{h}_{CV})$	92.2	0.223	$\hat{T}^{bc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	79.9	0.223	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CV}, \hat{h}_{CV})$	92.7	0.333	0.166	0.166
	$\hat{T}_1(\hat{h}_{CCT})$	90.5	0.213	$\hat{T}^{bc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	89.8	0.213	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{b}_{CCT})$	93.4	0.246	0.181	0.322
				$\hat{T}^{bc}_{1,2}(h_{MSE}, h_{MSE})$	80.7	0.178	$\hat{T}^{rbc}_{1,2}(h_{MSE}, h_{MSE})$	93.3	0.262	0.260	0.260
				$\hat{T}^{bc}_{1,2}(h_{DM}, h_{DM})$	80.6	0.181	$\hat{T}^{rbc}_{1,2}(h_{DM}, h_{DM})$	93.3	0.267	0.251	0.251
				$\hat{T}^{bc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	80.6	0.189	$\hat{T}^{rbc}_{1,2}(\hat{h}_{IK}, \hat{h}_{IK})$	92.9	0.279	0.231	0.231
				$\hat{T}^{bc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	80.5	0.198	$\hat{T}^{rbc}_{1,2}(\hat{h}_{DM}, \hat{h}_{DM})$	92.9	0.294	0.209	0.209
				$\hat{T}^{bc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	80.3	0.213	$\hat{T}^{rbc}_{1,2}(\hat{h}_{CCT}, \hat{h}_{CCT})$	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.