NBER WORKING PAPER SERIES

LOCAL POLYNOMIAL ORDER IN REGRESSION DISCONTINUITY DESIGNS

Zhuan Pei
David S. Lee
David Card
Andrea Weber

Working Paper 27424
http://www.nber.org/papers/w27424

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

We thank Matias Cattaneo, Gordon Dahl, Guido Imbens, Jianqing Fan, Pat Kline, Pauline Leung, Doug Miller, Jack Porter, Yi Shen, Yan Song, Stefan Wager, and Vivian Wang and seminar participants at Brandeis, Cornell, and George Washington, as well as conference attendees at APPAM, CIRANO-CIREQ, Econometric Society Meetings, Jinan IESR, and SOLE for helpful comments. Camilla Adams, Amanda Eng, Samsun Knight, Suejin Lee, Carl Lieberman, Bailey Palmer, and Amy Tarczynski provided outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Local Polynomial Order in Regression Discontinuity Designs
Zhuan Pei, David S. Lee, David Card, and Andrea Weber
NBER Working Paper No. 27424
June 2020
JEL No. C21

ABSTRACT

Treatment effect estimates in regression discontinuity (RD) designs are often sensitive to the choice of bandwidth and polynomial order, the two important ingredients of widely used local regression methods. While Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014b) provide guidance on bandwidth, the sensitivity to polynomial order still poses a conundrum to RD practitioners. It is understood in the econometric literature that applying the argument of bias reduction does not help resolve this conundrum, since it would always lead to preferring higher orders. We therefore extend the frameworks of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014b) and use the asymptotic mean squared error of the local regression RD estimator as the criterion to guide polynomial order selection. We show in Monte Carlo simulations that the proposed order selection procedure performs well, particularly in large sample sizes typically found in empirical RD applications. This procedure extends easily to fuzzy regression discontinuity and regression kink designs.

Zhuan Pei
Dept. of Policy Analysis and Management
Cornell University
Ithaca, NY 14853
peizhuan@gmail.com

David S. Lee
Industrial Relations Section
Louis A. Simpson International Bldg.
Princeton University
Princeton, NJ 08544
and NBER
davidlee@princeton.edu

David Card
Department of Economics
549 Evans Hall, #3880
University of California, Berkeley
Berkeley, CA 94720-3880
and NBER
card@econ.berkeley.edu

Andrea Weber
Central European University
Department of Economics and Business
Nador u. 9
1051 Budapest
Hungary
webera@ceu.edu

1 Introduction

Regression discontinuity designs (RD designs or RDD) have been widely used in empirical social science research in recent years. Two important reasons for its appeal are that the research design permits clear and transparent identification of causal parameters of interest, and the design itself has testable implications similar in spirit to those in a randomized experiment (Lee, 2008 and Lee and Lemieux, 2010).

Although the identification strategy is both transparent and credible in principle, many methods can be used to estimate the same causal parameter of interest. The key challenge is to estimate the values of the conditional expectation functions at the discontinuity cutoff without making strong assumptions about the shape of that function.

Typical practice in applied research is to employ a nonparametric local regression estimator. We surveyed leading economics journals between 1999 and 2017 (which included *American Economic Review*, *American Economic Journals*, *Econometrica, Journal of Political Economy*, *Journal of Business and Economic Statistics*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Review of Economics and Statistics*), and found that of the 110 studies employing RDD, 76 use a local polynomial regression as their main specification (Appendix Table A.1). Among these 76 studies, local linear is the modal choice and is applied as the main specification in 45 studies, but the remaining 31 (about 40%) choose a different order.

As a practical matter, researchers often report results from using different polynomial orders, and feel re-assured when their estimates are robust. But what are they to do when their conclusions *are* sensitive to polynomial order? This question mirrors the motivation behind optimal bandwidth proposals by Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*), and it is the focus of the present paper.

Reasoning grounded in bias reduction of the RD estimator provides no guidance on this question. As both Hahn, Todd and Van der Klaauw (2001) and Porter (2003) point out, higher order polynomials have a smaller asymptotic bias than lower orders. On the other hand, Gelman and Imbens (2019) argue that high order polynomials can perform poorly in certain contexts.

In this paper, we propose to extend the now widely-used theoretical framework and data-driven approach of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*)—which use estimated asymptotic mean squared error (AMSE or asymptotic MSE) of the RD estimator as an optimality criterion for bandwidth choice—to guide polynomial order selection.[1] Thus, the proposed procedure is based on a

[1] Although the recent empirical literature continues to adopt the approach of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*), we do note that the recent work of Armstrong and Kolesár (2018*a*,*b*) has proposed an alternative

local (as opposed to global) optimality criterion, as advocated by Gelman and Imbens (2019). Intuitively, the procedure can choose a local linear specification when the true underlying conditional expectation function has small curvature near the threshold, but a higher order polynomial when the curvature is large.

Our proposal is complementary to the recent work by Hall and Racine (2015), who call into question the practice of choosing the polynomial order ad hoc for nonparametric estimation at an interior point, and suggest a cross-validation method to select the polynomial order jointly with the bandwidth.[2] Instead of cross-validation, we provide a formal justification for the application of a suggestion by Fan and Gijbels (1996) to RD designs, paralleling Imbens and Kalyanaraman (2012).

In order to assess the potential usefulness of the proposed procedure, we conduct Monte Carlo simulations based on two well-known examples (Lee, 2008 and Ludwig and Miller, 2007), where we use the exact same parameters as the simulations conducted by Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*). First, we illustrate the nature of the conundrum that researchers face in practice. Unsurprisingly, we find that in some cases the local linear specification performs the best, but in many other configurations, alternative polynomials fare better in terms of their MSE, coverage rate of the 95% confidence interval (CI), and size-adjusted CI length. Second, we find that the estimator chosen by comparing estimated AMSEs performs well in terms of MSE, CI coverage rate, and CI length, especially in larger sample sizes we often see employed in RD applications.

Finally, we compute the AMSE of the fuzzy RD estimator, the sharp and fuzzy estimators in the regression kink design (RK design or RKD), and the bias-corrected estimator of Calonico, Cattaneo and Titiunik (2014*b*) in all these contexts. We have implemented these computations in a Stata package `rdmse`. The installation instruction is available at https://sites.google.com/site/peizhuan/programs/.

The remainder of the paper is organized as follows. Section 2 briefly summarizes the typical theoretical arguments used for local polynomial RD estimators and establishes consistency of our proposed polynomial order selection procedure. Section 3 presents simulation results. In section 4, we discuss the extensions of our proposal to fuzzy RD designs and regression kink designs. Section 5 concludes.

---

paradigm for inference in regression discontinuity designs. In their alternative framework, the user supplies a bound on second derivatives of the underlying conditional expectation function and inference is conducted uniformly over the class of data-generating processes that satisfy the bound. In those analyses, the question of which polynomial order to use also arises (the current focus is on the most commonly utilized order in practice—local linear), but addressing the issue of polynomial order choice within this alternative framework is beyond the scope of this paper.

[2]Polynomial order choice is also discussed in the literature of sieve methods, but as reviewed by Chen (2007), only the rate at which polynomial order increases with sample size is specified, which does not readily translate into practical advice.

2 Local Polynomial Order in RD Designs: Theoretical Considerations

In this section, we review and re-examine the theoretical justification for the choices in nonparametric RD estimation. In a sharp RD design, the binary treatment D is a discontinuous function of the running variable X: $D = 1_{[X \geqslant 0]}$ where we normalize the policy cutoff to 0. Hahn, Todd and Van der Klaauw (2001) and Lee (2008) show that under smoothness assumptions, the estimand:

$$\lim_{x \to 0^+} E[Y|X=x] - \lim_{x \to 0^-} E[Y|X=x] \tag{1}$$

identifies the treatment effect $\tau \equiv E[Y_1 - Y_0|X=0]$, where Y_1 and Y_0 are the potential outcomes. To estimate (1), researchers typically use a polynomial regression framework to separately estimate $\lim_{x \to 0^+} E[Y|X=x]$ and $\lim_{x \to 0^-} E[Y|X=x]$. Specifically, they solve the minimization problem using only observations above the cutoff as denoted by the + superscript:

$$\min_{\{\tilde{\beta}_j^+\}} \sum_{i=1}^{n^+} \{Y_i^+ - \tilde{\beta}_0^+ - \tilde{\beta}_1^+ X_i^+ - \ldots - \tilde{\beta}_p^+ (X_i^+)^p\}^2 K(\frac{X_i^+}{h}). \tag{2}$$

The resulting $\hat{\beta}_0^+$ is the estimator for $\lim_{x \to 0^+} E[Y|X=x]$, and the estimator $\hat{\beta}_0^-$ for $\lim_{x \to 0^-} E[Y|X=x]$ is defined analogously. The RD treatment effect estimator is $\hat{\tau}_p \equiv \hat{\beta}_0^+ - \hat{\beta}_0^-$, where we emphasize its dependence on p by the subscript.

Any nonparametric RD estimator is generally biased in finite samples. Expressions for the exact bias require knowledge of the true underlying conditional expectation functions; thus, the econometric literature has focused on first-order asymptotic approximations for the bias and variance. Applying these ideas, Lemma 1 of Calonico, Cattaneo and Titiunik (2014*b*) derives the AMSE of the pth order local polynomial estimator $\hat{\tau}_p$ as a function of bandwidth as:

$$\text{AMSE}_{\hat{\tau}_p}(h) = h^{2p+2} \text{B}_p^2 + \frac{1}{nh} \text{V}_p \tag{3}$$

where B_p and V_p are unknown constants. The first term is the approximate squared bias, and the second term the approximate variance. B_p depends on the $(p+1)$th derivatives of the conditional expectation function $E[Y|X=x]$ on two sides of the cutoff, and V_p on the conditional variance $Var(Y|X=x)$ on two sides of the cutoff as well as the density of X at the cutoff.

First-order approximations like the one above have been used in the literature in two ways. First, Hahn, Todd and Van der Klaauw (2001) argue in favor of the local linear RD estimator ($p = 1$) over the kernel regression estimator ($p = 0$) for its smaller order of asymptotic bias—the biases of the two different estimators are $h^2\mathrm{B}_1$ and $h\mathrm{B}_0$ and are of orders $O(h^2)$ and $O(h)$ respectively. However, by the same logic, the asymptotic bias of the local quadratic estimator ($p = 2$) is of order $O(h^3)$, and the bias of the local cubic is of order $O(h^4)$. More generally, the bias of the pth order estimator is of order $O(h^{p+1})$. Therefore, if researchers were exclusively focused on the maximal shrinkage rate of the asymptotic bias, they would choose p to be as large as possible. Hahn, Todd and Van der Klaauw (2001) recommend $p = 1$, implicitly recognizing that factors beyond bias shrinkage rate should also be taken into consideration.

Second, expression (3) is used as a criterion to determine the optimal bandwidth for a chosen order p. Since the AMSE is a convex function of h, one can solve for the optimal bandwidth that leads to the smallest value of AMSE: $h_{opt}(p) \equiv \arg\min_h \mathrm{AMSE}_{\hat{\tau}_p}(h)$. Imbens and Kalyanaraman (2012) do precisely this to propose a bandwidth selector for local linear estimation (henceforth IK bandwidth) and Calonico, Cattaneo and Titiunik (2014*b*) further extend the selector to polynomial estimators of alternative orders (henceforth CCT bandwidth).

We now highlight that there is no theoretical ground to always prefer a specific polynomial order across all empirical contexts. By evaluating expression (3) at $h_{opt}(p)$, which is of order $O(n^{-\frac{1}{2p+3}})$, $\mathrm{AMSE}_{\hat{\tau}_p}(h_{opt}(p))$ is equal to $\mathrm{C}_p \cdot n^{-\frac{2p+2}{2p+3}}$ with C_p being a function of the constants B_p and V_p. Therefore, as the sample size n increases, $\mathrm{AMSE}_{\hat{\tau}_p}(h_{opt}(p))$ shrinks faster for a larger p and will eventually, for the same n, fall *below* that of a lower-order polynomial. Intuitively, if $E[Y|X = x]$ is close to being linear on both sides of the cutoff, then the local linear specification will provide an adequate approximation, and consequently $\hat{\tau}_1$ will have a smaller AMSE than that of $\hat{\tau}_2$ for a large range of sample sizes. On the other hand, if the curvature of $E[Y|X = x]$ is large near the cutoff, a higher p will have a lower AMSE, possibly even for small sample sizes. Although we expect higher-order polynomials to have lower AMSE in sufficiently large samples, the precise sample size threshold at which that happens depends on the data generating process (DGP) through the constant C_p.[3]

[3] Sun (2005) also studies the choice of polynomial order in an RDD, but focuses on the case where the degree of smoothness in the underlying conditional expectation function is unknown. Sun (2005) shows that it is rate-suboptimal to use an order that is too high given the amount of smoothness assumed, which Armstrong and Kolesár (2018*b*) echo when examining the issue in a general nonparametric regression framework. Our argument is distinct from theirs: we are willing to assume sufficient smoothness, but we do not just consider the convergence rates when choosing the polynomial order. We also account for the constant C_p, which often plays an important role empirically.

This point is concretely illustrated in Figure 1, using the two DGPs we rely on for subsequent simulations, which are based on Lee (2008) and Ludwig and Miller (2007) and described in greater detail in Appendix B.1. Since we know the parameters of the underlying DGPs, we can analytically compute the quantities on the right hand side of equation (3). Using Lemma 1 of Calonico, Cattaneo and Titiunik (2014*b*), we plot $\text{AMSE}_{\hat{\tau}_p}$ as a function of sample size n for $p = 1,2$, which are shown in Panels (A) and (B) of Figure 1 for the two DGPs respectively (see Appendix C.1 for details).

In Panel (A), we see that at small sample sizes, $\text{AMSE}_{\hat{\tau}_1}$ is marginally below $\text{AMSE}_{\hat{\tau}_2}$, but is larger at sample sizes over $n = 1,167$. Therefore, for the actual number of observations in the analysis sample of Lee (2008), $n_{actual} = 6,558$, local quadratic should be preferred to local linear based on the AMSE comparison—the associated reduction in AMSE is 9%. In Panel (B), the difference between $p = 1$ and $p = 2$ is much larger, and $\text{AMSE}_{\hat{\tau}_2}$ dominates $\text{AMSE}_{\hat{\tau}_1}$ for all n under $7,000$. At the actual number of observations in Ludwig and Miller (2007), $n_{actual} = 3,105$, the local quadratic estimator reduces the AMSE by a considerable 38%.[4] It is worth noting that at n_{actual}, the AMSE closely matches the MSE from our simulations in section 3 below, which are marked by a cross for the local linear estimator and a circle for local quadratic.

In practice, equation (3) cannot be directly applied because it depends on unknown derivatives of the conditional expectation function, unknown conditional variances, and the density of X. Thus, Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*) use the empirical analog of (3):

$$\widehat{\text{AMSE}}_{\hat{\tau}_p}(h) = h^{2p+2}\hat{\text{B}}_p^2 + \frac{1}{nh}\hat{\text{V}}_p \tag{4}$$

where the quantities B_p and V_p in (3) are replaced by consistent estimators $\hat{\text{B}}_p$ and $\hat{\text{V}}_p$, and the optimal feasible bandwidth is defined as $\hat{h}(p) \equiv \arg\min_h \widehat{\text{AMSE}}_{\hat{\tau}_p}(h)$. The two studies differ in how they arrive at the estimates of B_1 and V_1. Additionally, Calonico, Cattaneo and Titiunik (2014*b*) generalize Imbens and Kalyanaraman (2012) by proposing bandwidth selectors for $\hat{\tau}_p$ for any given p. Both bandwidth selectors include a regularization term, which reflects the variance in bias estimation and prevents the selection of large bandwidths. Even though the regularization term is asymptotically negligible, it often plays an important role empirically: for the fuzzy regression kink design application in Card et al. (2015*a*) and Card et al. (2017), the inclusion of the regularization term leads to 30 - 70% reductions in bandwidths and has a considerable impact on the point estimate and confidence interval of the treatment effect parameter. In our Monte

[4] We conduct the same exercise for other values of p in a previous working paper—see Tables 3-4, A.3-A.4 in Card et al. (2014).

Carlo simulations below, we experiment with the CCT bandwidth both with and without regularization.

In this paper, we simply extend the logic that justifies the optimal bandwidth by noting that we should choose the polynomial order corresponding to the lowest estimated AMSE. That is, we can define

$$\hat{p} \equiv \arg\min_{p \in \Omega} \widehat{\text{AMSE}}_{\hat{\tau}_p}\left(\hat{h}(p)\right),$$

where Ω consists of a finite number of candidate polynomial orders (Ω can contain as few as two elements if a researcher is just choosing between two orders). For the AMSE of $\hat{\tau}_p$, no new quantities need to be computed beyond the estimators $\hat{\text{B}}_p$ and $\hat{\text{V}}_p$ and the optimal $\hat{h}(p)$, which must already be calculated when implementing, for example, the CCT bandwidth.

In summary, once one has already chosen an estimator (and the corresponding AMSE-minimizing bandwidth selector such as CCT), then it is straightforward to also report the resulting $\widehat{\text{AMSE}}_{\hat{\tau}_p}$ for any given p and compare $\widehat{\text{AMSE}}_{\hat{\tau}_p}$ across different candidate polynomial orders. Appendix C.2 provides the exact expressions needed from Calonico, Cattaneo and Titiunik (2014*b*) for the calculation of the AMSE of $\hat{\tau}_p$, which is implemented in the Stata package `rdmse`.[5]

Although this simple order selection approach was suggested by Fan and Gijbels (1996) for general local polynomial regression, to the best of our knowledge, a formal theoretical justification for the suggestion has yet to be discussed, and the approach has yet to be applied to RD designs. We investigate the asymptotic property of the procedure in section 2.1 and report on its finite sample performance in section 3.

2.1 Theoretical Justification: Consistency of $\hat{p}$

This subsection presents a theoretical justification of choosing p on the basis of estimated AMSE. The justification parallels previous results on bandwidth selection, e.g. Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*), which prove the consistency of the bandwidth selector $\hat{h}(p)$ for $h_{opt}(p)$. There are two alternative asymptotic frameworks employed in the literature, and we show the consistency of $\hat{p}$ in both. The first asymptotic framework adopts bandwidths that shrink at the optimal rates. This is the framework that has been used to argue for the use of $p = 1$ over $p = 0$, as mentioned at the beginning of section 2; it is also the framework for the discussion about Figure 1. In the second framework,

[5] In addition, Appendix C.2 details the calculation of the AMSE of the *bias-corrected estimator* of Calonico, Cattaneo and Titiunik (2014*b*) (denoted by $\hat{\tau}_p^{bc}$), which is also implemented in the Stata package `rdmse`.

which is used by Calonico, Cattaneo and Titiunik (2014*b*) to derive their key inference results, we assume that bandwidths for polynomial estimators of different orders shrink at the same rate as a function of sample size.

We first define

$$p_{opt} \equiv \arg\min_{p \in \Omega} \text{MSE}_{\hat{\tau}_p}(h(p))$$

as the MSE-optimal polynomial order in the candidate set Ω, where $h(p)$ denotes the bandwidth choice for the pth order local regression estimator. In general, p_{opt} is a function of n, and consistency means $\hat{p}/p_{opt} \xrightarrow{\mathbb{P}} 1$. Using p_{max} to denote the largest candidate polynomial order ($p_{max} \equiv \max\{p|p \in \Omega\}$)—which can be as low as 1 if a researcher is choosing between local constant and local linear specifications—we state our assumptions.

Assumption 1. *p_{max} is constant.*

Assumption 2. *a) Assumptions 1 and 2 in Calonico, Cattaneo and Titiunik (2014*b*) hold with $S = p_{max} + 1$; b) $\hat{B}_p$ and $\hat{V}_p$ in equation (4) are consistent estimators for B_p and V_p in equation (3).*[6]

Assumption 3. *$h(p) = H_p \cdot n^{-\frac{1}{2p+3}}$ with $H_p > 0$ and $\hat{h}(p)/h(p) \xrightarrow{\mathbb{P}} 1$.*

Assumption 1 states that p_{max} does not increase with n. This assumption is not restrictive in practice since the researcher may always pick a large enough p_{max} a priori regardless of n.[7] Part a) of Assumption 2 consists of standard regularity conditions that allow for the asymptotic approximation of MSE, and part b) encompasses the estimators $\hat{\text{B}}_p$ and $\hat{\text{V}}_p$ in Imbens and Kalyanaraman (2012) for $p = 1$ and Calonico, Cattaneo and Titiunik (2014*b*) as special cases. Note that a larger p_{max} translates to a higher degree of smoothness

[6]Assumption 1 in Calonico, Cattaneo and Titiunik (2014*b*) consists of regularity conditions for the fourth moment of Y given X, the density of X, and the conditional expectation and variance functions of the potential outcomes given X. In particular, the conditional expectation functions of the potential outcomes are assumed to be S-times differentiable in a neighborhood around zero. Assumption 2 in Calonico, Cattaneo and Titiunik (2014*b*) requires the kernel function $K(\cdot)$ in the minimization problem (2) to have compact support, be nonnegative, and be continuous.

[7]Calonico, Cattaneo and Titiunik (2015) and Gelman and Imbens (2019) connect their concerns regarding a high-order *global* polynomial estimator to the Runge phenomenon. The Runge phenomenon arises in the polynomial *interpolation* of a function $f(x)$ over an interval $[a,b]$: using a polynomial of order n to interpolate a function through $n+1$ equispaced knots when n is large does not imply uniform convergence to f. In fact, large departures from the function may result outside the interpolation knots, especially toward the edge of $[a,b]$. One textbook remedy (Ch. 4 of Dahlquist and Björck, 2008 and Ch. 8 of Björck, 1996) to guard against the Runge phenomenon is to employ least squares regression as opposed to interpolation. As a rule of thumb, the recommendation is to use a polynomial order no larger than $2\sqrt{n}$ where n is the number of (equispaced) observations, which still leaves many polynomial order candidates. In practice, researchers typically choose polynomial orders from a much smaller set, and the concern of considering too-high a polynomial order is further alleviated by our consideration of *local*, as opposed to *global*, polynomials. In previous working papers Card et al. (2014) and Pei et al. (2018), we also follow Gelman and Imbens (2019) and plot the regression weights for various local estimators using the actual Lee (2008) and Ludwig and Miller (2007) data, and the weights appear reasonable for all of the polynomial orders we consider, up to a quintic specification.

in Assumption 2, which may seem undesirable ostensibly. But it is also arbitrary to assume, for example, that the conditional expectation functions $E[Y_1|X=x]$ and $E[Y_0|X=x]$ have continuous second derivatives ($S=2$) but not continuous third derivatives ($S=3$). The technicality of Assumption 2 notwithstanding, for all practical purposes, we treat these conditional expectation functions as infinitely smooth.

Assumption 3 is the key assumption of the first asymptotic framework we consider. It states that the theoretical bandwidth for each p shrinks at the optimal rate and that the bandwidth selector is consistent. The CCT bandwidth selector, for example, satisfies this property.

Proposition 1. *Under Assumptions 1, 2 and 3,* $p_{opt} \to p_{max}$ *and* $\hat{p}/p_{opt} \xrightarrow{\mathbb{P}} 1$.

The proofs of the Propositions are provided in Appendix A. Proposition 1 says that under standard asymptotics as provided by Assumption 3, a) the optimal polynomial order is the "corner solution" p_{max} when the sample size is large; b) the order we select will also converge to p_{max} in probability. Point a) echos the insight from Porter (2003) and our discussion above that a higher order estimator will dominate in a sufficiently large sample when using optimal bandwidths. However, to reiterate our point made at the beginning of section 2, which we will illustrate again in section 4 using an empirical example from RKD, the "corner solution" here reflects the theoretical property that $\text{AMSE}_{\hat{\tau}_p}$ decreases at a higher rate as a function of the sample size when p is larger. As we highlight earlier, although p_{opt} converges to p_{max} asymptotically, p_{opt} may not coincide with p_{max} in any finite sample. This is true even in sample sizes conventionally considered to be large, as is the case with our RKD example below in section 4. It is worth emphasizing that this is not a statement about the finite sample performance of $\hat{p}$—p_{opt} is not subject to sampling variation; instead, as argued in the discussion of Figure 1, it is about the important role of the constants (B_p and V_p for $p \in \Omega$) in determining p_{opt}, beyond the asymptotic rates that push p_{opt} toward p_{max}.

To highlight the role of these constants, we consider a second, alternative, asymptotic framework used in the literature, in which p_{opt} can be an "interior solution". That is, even in the limit as the sample size tends to infinity, we can still have $p_{opt} < p_{max}$. The key assumption of this alternative asymptotic framework is:

Assumption 4. $h(p) = H_p \cdot n^{-\alpha}$ *with* $H_p > 0$ *and* $\alpha \in (0,1)$ *for all* $p \in \Omega$.

Unlike in Assumption 3, all bandwidths shrink at the same rate in Assumption 4 regardless of the polynomial order p. It is analogous to the defining assumption of the asymptotic framework in Calonico, Cattaneo and Titiunik (2014*b*): for their inference result, Calonico, Cattaneo and Titiunik (2014*b*) assume that the bandwidth for estimating the bias and the bandwidth for estimating the treatment effect shrink at the same rate. Calonico, Cattaneo and Titiunik (2014*b*) maintain this assumption even though the bias term contains higher order derivatives of the conditional expectation functions than the treatment effect and that their corresponding bandwidth selectors in Calonico, Cattaneo and Titiunik (2014*b*) shrink at different rates as the sample size increases.

We now establish the consistency of $\hat{p}$ in this alternative asymptotic framework.

Proposition 2. *Under Assumptions 1, 2 and 4, $\hat{p}/p_{opt} \xrightarrow{\mathbb{P}} 1$ provided that p_{opt} is unique asymptotically.*

Under Assumption 4, $\text{MSE}_{\hat{\tau}_p}$ shrinks at the same rate for all p. Therefore, the limit of p_{opt} is generally not p_{max}, and the MSE of $\hat{\tau}_{p_{max}}$ does not always dominate that of alternative polynomial orders as is the case under Assumption 3. Instead, the optimal polynomial order depends on the magnitudes of the constants B_p and V_p from equation (3). To see this, consider the case where $\alpha = 1/5$: in the limit as $n \to \infty$, $\text{AMSE}_{\hat{\tau}_2}(h(2))/\text{AMSE}_{\hat{\tau}_1}(h(1))$ depends on the constants H_p, B_p, and V_p for $p = 1,2$ and can be larger than one, whereas under Assumption 3, $\text{AMSE}_{\hat{\tau}_2}(h(2))/\text{AMSE}_{\hat{\tau}_1}(h(1))$ goes to zero. There is another implication of this observation: under Assumption 3, p_{opt} is unique asymptotically, but there exist DGPs for which the AMSEs are the same for different p. Because of this implication, we assume the uniqueness of p_{opt} in Proposition 2. If the uniqueness assumption is relaxed, we still have the asymptotic no-regret property of $\hat{p}$ as per Li (1987) and Imbens and Kalyanaraman (2012): there is no loss asymptotically by using $\hat{p}$, as compared to any of the optimal orders that deliver the lowest MSE.

In summary, Propositions 1 and 2 establish the consistency of our polynomial order selection procedure in two asymptotic frameworks that have been invoked in the literature. In the first and more conventional framework, p_{opt} converges asymptotically to the corner solution p_{max}, the largest polynomial order in the candidate set. But even in a sample typically considered large, p_{opt} may not coincide with p_{max} depending on the bias and variance constants (B_p and V_p for $p \in \Omega$). Our second asymptotic framework, which is analogous to that of Calonico, Cattaneo and Titiunik (2014*b*), further emphasizes the role of the constants, which justifies $\hat{p}$ as consistent for p_{opt} when p_{opt} is distinct from p_{max}. We now move on to assess the practical performance of $\hat{p}$ in Monte Carlo exercises.

3 Monte Carlo Results

Although AMSE provides the theoretical basis for bandwidth selection and our complementary proposal for polynomial order selection, it is nevertheless a first-order asymptotic approximation of the true MSE. In this section, we conduct Monte Carlo simulations to examine the finite sample performance of local polynomial estimators of various orders—which themselves utilize the CCT bandwidth selectors—and our proposed order selection procedure.

We employ DGPs from two well-known empirical examples, Lee (2008) and Ludwig and Miller (2007), and the specifications of these DGPs follow *exactly* those in Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*). The conditional expectation functions are specified as piecewise quintic polynomials (see Appendix B.1 for details). Because of the 5th-order specification of the DGPs, the highest polynomial order we allow is $p_{max} = 4$ so that we do not mechanically favor estimators from correctly specified regressions.

Our simulations draw 10,000 repeated samples from the two DGPs. Below, we present results using a uniform kernel; results from the triangular kernel are available in a previous working paper Card et al. (2014), and the qualitative conclusions are the same.

The simulation results are organized as follows. Tables 1-4 report on the performances of conventional RD estimators ($\hat{\tau}_p$) applied to the two DGPs respectively, while Tables 5-8 report on the bias-corrected RD estimators ($\hat{\tau}_p^{bc}$) and the associated robust confidence intervals as per Calonico, Cattaneo and Titiunik (2014*b*). Results corresponding to two sample sizes are displayed in the eight tables: the actual sample size in Tables 1, 3, 5, 7, and large sample size in Tables 2, 4, 6, 8. The actual sample size is that of the analysis sample in the two empirical studies: $n_{actual} = 6,558$ for Lee (2008) and $n_{actual} = 3,105$ for Ludwig and Miller (2007). We set the large sample size to $n_{large} = 60,000$ for the Lee DGP and $n_{large} = 30,000$ for Ludwig-Miller. n_{large} is about $10 \times n_{actual}$ in both studies, and it is comparable or lower than the number of observations in many recent empirical papers.

In part (a) of each table, we show the summary statistics for the local linear estimator with three bandwidth choices: i) the (infeasible) theoretical optimal bandwidth (h_{opt}), which minimizes AMSE using knowledge of the underlying DGP,[8] ii) the default CCT bandwidth selector from Calonico, Cattaneo and Titiunik

[8]Even though the theoretically optimal bandwidth is never known in an empirical application, we show simulation results for h_{opt} as a check on our theoretical intuition. As documented below, MSE decreases monotonically with p under h_{opt} with moderately large sample sizes. This is consistent with our discussion of the asymptotic behavior of $\text{AMSE}_{\hat{\tau}_p}(h_{opt}(p))$ in section 2. Moreover, comparing the simulation results under h_{opt} with those under data-driven bandwidth selectors gives a sense of how much the

(2014*b*) ($\hat{h}_{CCT}$), and iii) the CCT bandwidth selector without the regularization term ($\hat{h}_{CCT,noreg}$). We report averages and percentages across the simulations: the average bandwidth in column (2), average number of observations within the bandwidth in column (3), bias in column (4), variance in column (5), MSE in column (6), coverage rate of the 95% CI in column (7), the average CI length in columns (8), and the average *size-adjusted* CI length in columns (9). We make three remarks on these statistics that we report. First, although the emphasis of Calonico, Cattaneo and Titiunik (2014*b*) is on inference, examining the MSE of the bias-corrected estimator $\hat{\tau}_p^{bc}$ is informative—holding constant the coverage rate, an estimator that delivers a lower MSE is preferable. Second, the conventional 95% CI associated with the estimator $\hat{\tau}_p$ ignores the bias term, but we can justify it by treating the bandwidth selector as shrinking slightly faster than at the optimal rate as in Card et al. (2012). And third, while the other statistics are standard in Monte Carlo exercises, the size-adjusted CI length warrants further explanation. Size-adjustment is necessary because not all 95% CIs achieve the nominal coverage rate, in which case no standard metric tells us how to trade off a lower coverage rate for a shorter confidence interval. Therefore, we adapt the size-adjusted power proposal from Zhang and Boos (1994) to calculate size-adjusted 95% CIs. Specifically, instead of using 1.96 as the critical value for constructing the 95% CI, we find the smallest critical value so that the resulting size-adjusted 95% CI has the nominal coverage rate in the simulation. We simply report the average length of these size-adjusted CIs in column (9).

In part (b) of each table, we present the same statistics for different polynomial orders; in columns (4)-(9), we express the quantities as a ratio to the quantity in the local linear specification.

3.1 Performances of Alternative Polynomials

The set of polynomial orders we assess is limited by the piecewise quintic specification of the two DGPs. As mentioned above, since the kth order derivative of the conditional expectation function is zero at the cutoff for $k > 5$, the highest-order estimator we allow is local quartic to ensure the finiteness of the theoretical optimal bandwidth. For the Lee DGP, the alternative polynomial orders are $p = 0, 2, 3, 4$, as well as the order $\hat{p}$ selected from the set $\{0, 1, 2, 3, 4\}$ that minimizes estimated AMSE. For Ludwig-Miller, we exclude $p = 0$ from the simulations under the actual sample size, because h_{opt} for $p = 0$ is so small (0.004) that the average effective sample size is only 17.[9]

behavior of $\hat{\tau}_p$ is driven by the bandwidth selector.

[9] The bias-corrected estimator of Calonico, Cattaneo and Titiunik (2014*b*) of order p is known to be equivalent to the conventional estimator of order $(p+1)$ when the main and pilot bandwidths are the same. We do not use the same main and pilot

We highlight several findings from Tables 1 to 8. First, although the *de facto* local linear estimator performs competitively in some cases (e.g., Lee DGP with data-driven bandwidth selectors as seen from Tables 1 and 5), it does not deliver the lowest MSE. Looking down column (6) in part (b) of every table, there is at least one alternative estimator for which the MSE ratio is less than one. In these cases, the reduction in MSE ranges from 5.5% (local quadratic with $\hat{h}_{CCT}$ in Table 5) to 73% (local quartic with h_{opt} in Table 4), and lower bias contributes more to the decrease in MSE than variance.

Second, from column (7) in all tables, alternative estimators may improve upon the local linear in terms of its 95% CI coverage rate. It is worth noting that the coverage rate of the local linear CI is close to the nominal level in many instances, in which case the improvement by alternative estimators is small. But the improvement can be substantial in other cases. Given the analysis of Calonico, Cattaneo and Titiunik (2014*b*), it is not surprising that the conventional local linear CI sometimes undercover. The undercoverage is more serious under the Lee DGP: The local linear CI coverage rate is as low as 66% in simulations with n_{actual} and when the larger $\hat{h}_{CCT,noreg}$ is used (Table 1(a)). But this undercoverage is alleviated with the use of any alternative order, and the local quartic estimator has the highest coverage rate of $1.389 \times 66.0\% = 91.7\%$. The robust local linear CI has coverage rates closer to the nominal level, although it once again significantly undercovers in simulations with the Lee DGP, n_{actual}, and $\hat{h}_{CCT,noreg}$ – the coverage rate is 84.6% as shown in Table 5(a). By comparison, the local cubic and quartic robust CIs cover the true treatment effect parameter between 94% and 95% of the time.

Third, since all else equal, researchers prefer tighter confidence intervals, we compare the length of confidence intervals across different choices of p. Tables 7 and 8 show that the coverage rates are close to the nominal 95% for all robust confidence intervals for the Ludwig-Miller DGP, and almost all of the polynomial orders greater than one yield confidence intervals that are smaller, and substantially so (above 35 percent) in some cases. In Tables 1 to 6, the coverage rates of both local linear and higher-order polynomials are noticeably below the nominal 95% rate. Thus, we rely on size-adjusted confidence intervals in column (7) to compare the precision of the estimates on equal footing. Of the 54 specifications that use higher-order polynomials in those tables, 52 of them have shorter size-adjusted confidence intervals than local linear.

Our final observation about the performance of higher-order polynomials is that the optimal bandwidths

bandwidth in our simulation, and therefore the 4th-order bias-corrected estimator differs from the 5th-order conventional estimator, and we choose $p_{max} = 4$ for both the bias-corrected and conventional estimators for simplicity. If one is still concerned with the 4th-order bias-corrected estimator used in the simulation being "close" to being a 5th order conventional estimator and that it is being mechanically favored, we used $p_{max} = 3$ for the bias-corrected estimator in our previous working paper Card et al. (2014), and our conclusions were not affected.

for each of those orders suggest that the intuition that RD designs should only use observations "close" to the discontinuity threshold can be misleading. The intuition is reasonable for the hypothetical infinite sample, but in practice, with a finite sample, the optimal (in the AMSE sense) bandwidth may be relatively large. For example, in Tables 1 and 2, the order with the lowest MSE is $p = 4$. With $p = 4$, the (theoretically) optimal bandwidth implies using an average of 5,227 of the 6,558 observations under n_{actual}, or 39,983 of the 60,000 observations under n_{large}. The same pattern consistently holds true for the remaining tables: the better performing estimators use higher-order polynomials, which in turn imply larger optimal bandwidths, and therefore use a substantial fraction of the sample.

3.2 Performance of the Polynomial Order Selection Procedure

We have thus far provided both theoretical arguments and Monte Carlo evidence that point toward a more flexible view regarding the choice of p. We have presented simulation results on the performance of estimators that take p as given and use existing methods for choosing the $\widehat{\text{AMSE}}$-minimizing h, conditional on the given p. The evidence of the local linear specification performing well in some cases but not in others underscores the polynomial-order-choice conundrum researchers sometimes face.

We now turn to the performance of our proposed order selection procedure. Specifically, we designate our candidate set Ω to contain all of the polynomial orders considered in section 3.1, and for a particular Monte Carlo draw, we compute the RD estimator for each p in Ω and their corresponding $\widehat{\text{AMSE}}_{\hat{\tau}_p}$. For that same draw, we choose the p with the lowest $\widehat{\text{AMSE}}$. By repeating this process over the Monte Carlo draws, we can examine how well this procedure performs in terms of MSE, coverage, and the length of the confidence interval.

We report the results in the rows labeled "$\hat{p}$" below the quartic in Tables 1 to 8. Overall, our procedure tends to select a polynomial specification that performs well. Although the selected polynomial order varies across repeated sample draws, the modal value of $\hat{p}$ coincides with the lowest MSE order in the majority of cases. In fact, this happens for all 12 permutations (2 DGPs times 3 bandwidth selectors times 2 estimators) under the large sample size, n_{large}. Sometimes, our procedure leads to the local linear specification being the modal choice, but when it does not, it results in an estimator with an MSE improvement over local linear. In these cases, the reduction in MSE can be more than 40% for the Lee DGP and more than 70% for the Ludwig-Miller DGP, and the proportional decrease in bias is generally larger than that in variance. We see qualitatively similar results for the $\hat{p}$-selected estimator in terms of its CI coverage rate and length: when the procedure does not select linear as the modal choice, it improves both the CI coverage rate and length relative to local linear, with the exception of one case (Lee DGP, bias-corrected estimator, and theoretical bandwidth—Table 5) where the coverage rate decreases by a small amount (0.1 percentage point).

We show additional results in Appendix Tables A.2 to A.4 for the sample size $n_{small} = 500$. This is the sample size used in the simulations of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*). We see from Table A.2 that because $p = 1$ minimizes the MSE of the conventional estimator $\hat{\tau}_p$ under the Lee DGP, our polynomial selection procedure fares worse than always using local linear. As shown in Table A.4, $\hat{p}$ does better for the bias-corrected estimator $\hat{\tau}_p^{bc}$, for which local constant is MSE-minimizing(!), leading to comparable or lower MSEs, but the corresponding CI may undercover. This somewhat underwhelming performance of $\hat{p}$ in small sample size is an important caveat, but we note that it is rare to find RD studies that rely on 500 or fewer observations. In our survey of 110 studies, only three papers use fewer than 500 observations, a third of the papers use fewer than 6,000 observations, and the median sample size is 21,561. A sample size of 60,000, the largest sample size used in our simulations, sits at the 63rd percentile. Therefore, it is fairly common to see studies with sample size at or larger than 60,000, much more so than seeing studies with about 500 observations. But even with 500 observations, our selection procedure performs well under the Ludwig-Miller DGP as shown in Tables A.3 and A.5: the modal $\hat{p}$ always coincides with the MSE minimizing polynomial order, and relative to local linear, our procedure leads to an estimator with improved MSE, CI coverage rate, and CI length.

To summarize, we have implemented simulations under two DGPs (Lee and Ludwig-Miller), three bandwidth choices (h_{opt}, h_{CCT}, and $h_{CCT,noreg}$), two types of estimators (conventional and bias-corrected), and three sample sizes (n_{small}, n_{actual}, and n_{large}). We see that the best performing polynomial order varies across context: the MSE minimizing specification ranges from local constant to local quartic (the highest order we consider), and local linear does perform best in some cases. We also find that our polynomial selection procedure generally performs well, especially in larger sample sizes typically used in RD studies.

4 Extensions: Fuzzy RD and RKD

In this section, we briefly discuss how AMSE-based local polynomial order choice applies to two popular extensions of the sharp RD design. The first extension is the fuzzy RD design, where the treatment assignment rule is not strictly followed. In the same way that we can estimate the AMSE of a sharp RD estimator, we can rely on Lemma 2 and Theorem A.2 of Calonico, Cattaneo and Titiunik (2014*b*) to estimate the AMSE of a fuzzy RD estimator (consistent with the prevailing practice of using the same bandwidth for the first stage and outcome equations in a fuzzy RD design, we use the same polynomial order for both in our Stata implementation `rdmse`).

The same principle can be applied to the regression kink design proposed and explored by Nielsen, Sørensen and Taber (2010) and Card et al. (2015*a*). For RKD, Calonico, Cattaneo and Titiunik (2014*b*) and Gelman and Imbens (2019) recommend using local quadratic ($p = 2$) by extending the Hahn, Todd and Van der Klaauw (2001) argument. But similar to our RD discussion, the AMSE of local quadratic may or may not be lower than alternative orders, depending on the sample size and DGP characteristics.

To illustrate this once again, but in the case of fuzzy RKD, we use the bottom-kink and top-kink samples of the application in Card et al. (2015*b*) to approximate the actual first-stage and reduced-form conditional expectation functions with global quintic specifications on each side of the cutoff (see Appendix B.2 for details). The specification of these approximating DGPs again allows us to compute $\text{AMSE}_{\hat{\tau}_p}$ as a function of sample size for different p (which is slightly different from equation (3) for the RD case):

$$\text{AMSE}_{\hat{\tau}_p}(h) = h^{2p+2}\text{B}_p^2 + \frac{1}{nh^3}\text{V}_p. \tag{5}$$

As shown in Panel (C) of Figure 1, the AMSE of the local quadratic fuzzy estimator is asymptotically smaller. However, it takes about 88 million observations for the local quadratic to dominate local linear. In Panel (D) of Figure 1, the local linear fuzzy estimator dominates its local quadratic counterpart for sample sizes up to 200 million observations; in fact, the threshold sample size that tips in favor of the local quadratic estimator is 61 trillion. Even though we had the universe of the Austrian unemployed workers over a span of 12 years, the number of observations is about 270,000 for both the top- and bottom-kink samples. In this case, these calculations give reason to prefer the local linear fuzzy RK estimator.

Once again, the results here highlight the importance of accounting for the bias and variance constants in the AMSE expression (analogous to RD, one can derive B_p and V_p for the pth order RK estimator) and not just the asymptotic shrinkage rate when selecting an estimator. Through simple algebraic manipulations of (5), we can show that the sample size equalizing the AMSEs of the local linear and quadratic RK estimators (i.e. location of the dashed line in Panels (C) and (D) of Figure 1) has the form

$$\text{Equalizing Sample Size} = 1.06 \cdot \left(\frac{V_2}{B_2^2}\right)\left(\frac{V_2}{V_1}\right)^{\frac{7}{3}}\left(\frac{B_2}{B_1}\right)^{7}. \tag{6}$$

First note that the equalizing sample size is larger when V_2 is larger relative to B_2^2, because V_2 carries more weight relative to B_2^2 in $\text{AMSE}_{\hat{\tau}_2}$ than V_1 relative to B_1^2 in $\text{AMSE}_{\hat{\tau}_1}$. More importantly, equation (6) reveals that the equalizing sample size increases with the ratios V_2/V_1 and B_2/B_1: higher variance and bias constants for the quadratic estimator require more observations to overcome them. As we emphasize throughout this paper, the optimal polynomial order depends on these DGP specific constants and therefore varies across empirical contexts.

5 Conclusion

This paper is motivated by the question of what researchers should do when their RD estimates are sensitive to the choice of polynomial order used in local regressions. Since the existing literature does not provide a practical answer, we propose to extend the logic of the widely-used approach of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*) and use the estimated AMSE to guide polynomial order selection. In Monte Carlo simulations based on two well-known RD examples, we see that the best polynomial ranges from local constant to quartic (the maximum order we allow) and varies across sample size and DGP characteristics. Our proposed order selection procedure performs reasonably well, especially in larger sample sizes we typically see in RD applications.

We make two concluding remarks. First, we choose to incorporate polynomial order selection into the framework of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*) because it is the modal paradigm empirical studies now follow. As mentioned in the introduction, however, recent papers by Armstrong and Kolesár (2018*a*,*b*) propose alternative estimation and inference strategies under uniformity.[10] Studying the choice of polynomial in Armstrong and Kolesár (2018*a*,*b*) could be a future avenue of research.

Second, we view the proposed polynomial selection procedure as a complement—not a substitute—to

[10] Similarly, Imbens and Wager (2019) propose a linear RD estimator through numerical convex optimization that minimizes the worst-case-scenario mean-squared-error over the class of DGPs with a known global bound on the second derivative. In a previous working paper Pei et al. (2018), we compare the performances of the procedures in Armstrong and Kolesár (2018*a*) and Imbens and Wager (2019) to that of Calonico, Cattaneo and Titiunik (2014*b*)—see Appendix D of Pei et al. (2018) for details.

analyses that explore result robustness to order choice. In many cases, different polynomial orders may yield substantively similar results, and the procedure will not be needed. But when researchers are confronted with estimate sensitivity with respect to polynomial order, the procedure can be used to rule out suboptimal estimators which yield drastically different results, as in the RKD context of Card et al. (2017).

References

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

Armstrong, Timothy B., and Michal Kolesár. 2018*a*. "Optimal inference in a class of regression models." *Econometrica*, 86(2): 655–683.

Armstrong, Timothy B., and Michal Kolesár. 2018*b*. "Simple and Honest Confidence Intervals in Nonparametric Regression." mimeograph.

Björck, Åke. 1996. *Numerical Methods for Least Squares Problems.* Philadelphia: SIAM.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014*a*. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *Stata Journal*, 14(4): 909–946.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014*b*. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2015. "Optimal Data-Driven Regression Discontinuity Plots." *Journal of the American Statistical Association*, 110(512): 1753–1769.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2012. "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design." NBER Working Paper 18564.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2014. "Local Polynomial Order in Regression Discontinuity Designs." Brandeis University Department of Economics Working Paper 81.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015*a*. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica*, 83(6): 2453–2483.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015*b*. "Inference on Causal Effects in a Generalized Regression Kink Design." Upjohn Institute Working Paper 15-218.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2017. "Regression Kink Design: Theory and Practice." In *Regression Discontinuity Designs: Theory and Applications*. Vol. 38 of *Advances in Econometrics*, , ed. Matias D. Cattaneo and Juan Carlos Escanciano, Chapter 5, 341–382. Emrald.

Chen, Xiaohong. 2007. "Large Sample Sieve Estimation of Semi-Nonparametric Models." In *Handbook of Econometrics*. Vol. 6, , ed. James J. Heckman and Edward E. Leamer, Chapter 76. Elsevier.

Dahlquist, Germund, and Åke Björck. 2008. *Numerical Methods in Scientific Computing, Volume I.* Society for Industrial and Applied Mathematics.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Applications.* Chapman and Hall.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–209.

Hall, Peter G., and Jeffrey S. Racine. 2015. "Infinite Order Cross-Validated Local Polynomial Regression." *Journal of Econometrics*, 185: 510–525.

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

Imbens, Guido W., and Stefan Wager. 2019. "Optimized Regression Discontinuity Designs." *Review of Economics and Statistics*, 101(2): 264–278.

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

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

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

Ludwig, Jens, and Douglas 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.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber. 2010. "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform." *American Economic Journal: Economic Policy*, 2(2): 185–215.

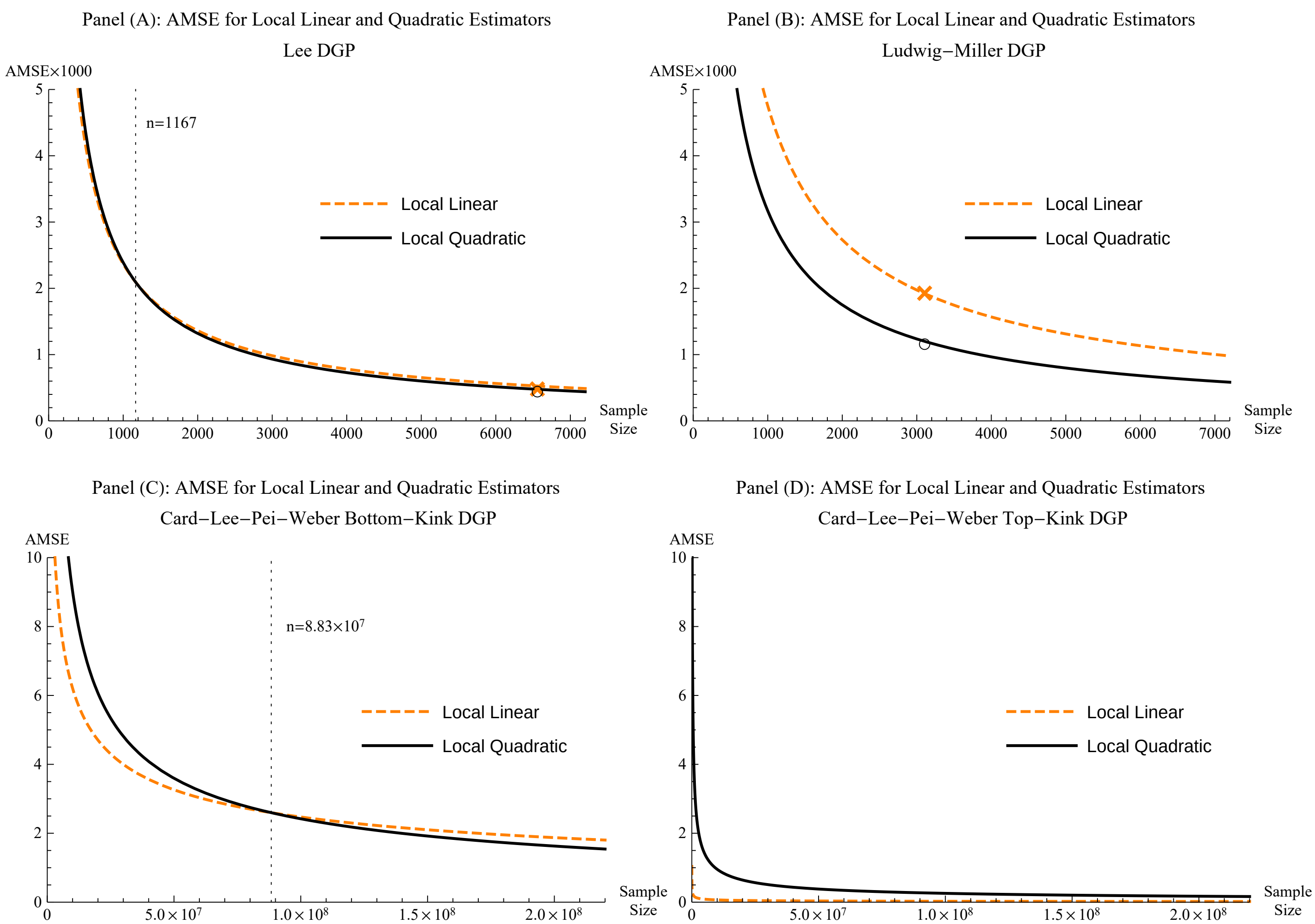
Pei, Zhuan, David S. Lee, David Card, and Andrea Weber. 2018. "Local Polynomial Order in Regression Discontinuity Designs." Princeton University Industrial Relations Section Working Paper No. 662.

Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." mimeographed.

Sun, Yixiao. 2005. "Adaptive Estimation of the Regression Discontinuity Model." mimeograph.

Zhang, Ji, and Dennis D. Boos. 1994. "Adjusted Power Estimates in Monte Carlo Experiments." *Communications in Statistics - Simulation and Computation*, 23(1): 165–173.

Figure 1: Asymptotic Mean-Squared-Error as a Function of Sample Size



Note: In Panels (A) and (B), we superimpose the simulated MSEs of the local linear (cross) and quadratic (circle) estimators with the theoretical optimal bandwidth. These MSEs are taken from Tables 1 and 3. At the actual sample size of the two studies, the theoretical AMSEs appear to be quite close to the corresponding MSEs.

Table 1: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Lee DGP, Actual Sample Size (n=6,558)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance× 1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.078	637	0.008	0.419	0.481	0.934	0.081	0.086
CCT	1	0.111	908	0.013	0.402	0.571	0.822	0.068	0.100
CCT w/o reg.	1	0.167	1335	0.018	0.437	0.768	0.660	0.060	0.138

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.015	127	1.969	1.275	1.610	0.960	1.123	1.271
	2	0.180	1477	0.723	0.975	0.917	1.008	0.987	0.964
	3	0.353	2888	0.631	0.902	0.837	1.008	0.945	0.918
	4	0.663	5227	0.672	0.780	0.738	1.009	0.879	0.856
	$\hat{p}$			0.627	0.816	0.762	1.008	0.891	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .194, .806)								
CCT	0	0.023	190	1.722	1.058	1.623	0.895	1.085	1.198
	2	0.207	1699	0.561	1.077	0.851	1.096	1.093	0.892
	3	0.298	2443	0.187	1.209	0.861	1.146	1.214	0.862
	4	0.348	2843	0.007	1.574	1.108	1.151	1.407	0.983
	$\hat{p}$			0.926	1.097	1.026	0.996	1.002	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.001, .779, .182, .038, 0)								
CCT w/o reg.	0	0.025	208	1.334	0.941	1.302	1.037	1.187	0.893
	2	0.274	2221	0.705	1.126	0.855	1.187	1.116	0.852
	3	0.425	3427	0.516	1.303	0.856	1.287	1.188	0.778
	4	0.539	4243	0.149	1.213	0.701	1.389	1.335	0.660
	$\hat{p}$			1.013	1.122	1.081	0.967	0.996	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.004, .646, .173, .123, .054)								

Table 2: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Lee DGP, Large Sample Size (n=60,000)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

Bandwidth	(1) p	(2) Avg. h	(3) Avg. n	(4) Bias	(5) Variance× 1000	(6) MSE ×1000	(7) Coverage Rate	(8) Avg. CI Length	(9) Avg. Size-adj. CI length
Theo. Optimal	1	0.050	3746	0.004	0.073	0.086	0.927	0.033	0.036
CCT	1	0.064	4778	0.005	0.069	0.098	0.847	0.030	0.040
CCT w/o reg.	1	0.069	5157	0.006	0.070	0.107	0.811	0.029	0.043

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.007	554	2.146	1.677	2.115	0.960	1.301	1.438
	2	0.131	9852	0.690	0.844	0.789	1.013	0.926	0.889
	3	0.276	20702	0.555	0.721	0.659	1.013	0.855	0.819
	4	0.542	39983	0.545	0.590	0.546	1.013	0.772	0.740
	$\hat{p}$			0.545	0.590	0.546	1.013	0.772	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, 0, 1)								
CCT	0	0.009	692	1.764	1.557	2.016	0.961	1.311	1.367
	2	0.151	11344	0.637	0.906	0.759	1.060	0.974	0.869
	3	0.280	20988	0.370	0.830	0.626	1.097	0.956	0.766
	4	0.357	26656	0.039	0.977	0.690	1.118	1.063	0.792
	$\hat{p}$			0.484	0.953	0.741	1.061	0.948	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .038, .172, .713, .078)								
CCT w/o reg.	0	0.009	704	1.585	1.506	1.853	0.997	1.345	1.285
	2	0.167	12501	0.730	0.959	0.813	1.050	0.964	0.889
	3	0.321	23959	0.562	0.986	0.756	1.085	0.931	0.800
	4	0.519	37787	0.347	1.042	0.725	1.093	0.935	0.833
	$\hat{p}$			0.541	0.969	0.737	1.055	0.903	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .02, .075, .349, .556)								

Table 3: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Actual Sample Size (n=3,105)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.045	175	0.018	1.609	1.926	0.922	0.155	0.174
CCT	1	0.050	195	0.022	1.550	2.055	0.886	0.147	0.182
CCT w/o reg.	1	0.051	198	0.023	1.520	2.061	0.882	0.145	0.181

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.151	587	0.576	0.654	0.601	1.015	0.817	0.778
	3	0.352	1362	0.364	0.503	0.442	1.021	0.718	0.666
	4	0.723	2653	0.220	0.397	0.340	1.028	0.642	0.583
	$\hat{p}$			0.209	0.412	0.351	1.026	0.648	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, 0, .082, .918)								
CCT	2	0.167	646	0.599	0.689	0.608	1.027	0.823	0.773
	3	0.293	1135	0.171	0.679	0.519	1.060	0.832	0.701
	4	0.341	1321	0.013	0.896	0.676	1.067	0.965	0.793
	$\hat{p}$			0.307	0.699	0.550	1.039	0.811	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, .291, .701, .009)								
CCT w/o reg.	2	0.173	670	0.642	0.685	0.614	1.023	0.816	0.781
	3	0.390	1495	0.397	0.672	0.537	1.015	0.739	0.732
	4	0.536	1996	0.059	0.701	0.518	1.057	0.806	0.697
	$\hat{p}$			0.347	0.625	0.493	1.027	0.733	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, .062, .689, .249)								

Table 4: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Large Sample Size (n=30,000)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.029	1072	0.007	0.253	0.304	0.924	0.062	0.069
CCT	1	0.031	1158	0.008	0.250	0.318	0.901	0.060	0.071
CCT w/o reg.	1	0.031	1168	0.008	0.248	0.319	0.900	0.060	0.070

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.002	76	2.899	3.596	4.395	0.964	1.889	2.077
	2	0.109	4102	0.568	0.584	0.541	1.015	0.767	0.724
	3	0.273	10245	0.364	0.416	0.369	1.021	0.649	0.602
	4	0.588	21534	0.231	0.314	0.271	1.023	0.562	0.518
	$\hat{p}$			0.231	0.314	0.271	1.023	0.562	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, 0, 1)								
CCT	0	0.002	80	2.623	3.570	4.279	0.971	1.912	2.054
	2	0.119	4450	0.612	0.591	0.545	1.015	0.766	0.731
	3	0.274	10257	0.295	0.457	0.378	1.037	0.675	0.607
	4	0.356	13288	0.000	0.547	0.430	1.052	0.744	0.642
	$\hat{p}$			0.258	0.470	0.383	1.035	0.675	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .881, .119)								
CCT w/o reg.	0	0.002	80	2.578	3.586	4.266	0.972	1.918	2.066
	2	0.120	4515	0.624	0.587	0.543	1.014	0.764	0.734
	3	0.296	11070	0.390	0.461	0.392	1.020	0.654	0.621
	4	0.534	19349	0.102	0.445	0.349	1.029	0.628	0.584
	$\hat{p}$			0.213	0.434	0.348	1.022	0.619	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .345, .655)								

Table 5: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Lee DGP, Actual Sample Size (n=6,558)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.078	637	0.004	0.495	0.507	0.950	0.088	0.088
CCT	1	0.111	906	0.007	0.461	0.514	0.900	0.077	0.092
CCT w/o reg.	1	0.165	1332	0.011	0.518	0.629	0.846	0.076	0.109

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.015	127	1.121	1.306	1.305	0.997	1.136	1.148
	2	0.180	1476	0.708	0.908	0.898	0.999	0.952	0.954
	3	0.353	2887	0.544	0.799	0.786	1.003	0.892	0.882
	4	0.663	5227	-0.010	0.751	0.732	1.002	0.865	0.856
	$\hat{p}$			0.299	0.804	0.786	0.999	0.880	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .567, .433)								
CCT	0	0.023	190	1.281	0.971	1.041	1.006	1.031	1.009
	2	0.207	1698	0.378	1.037	0.945	1.041	1.063	0.940
	3	0.299	2443	0.015	1.190	1.067	1.053	1.166	0.990
	4	0.347	2835	-0.031	1.568	1.406	1.054	1.335	1.132
	$\hat{p}$			1.111	0.982	1.008	0.994	0.987	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.297, .57, .12, .013, 0)								
CCT w/o reg.	0	0.025	208	1.055	0.831	0.881	1.033	0.995	0.890
	2	0.274	2224	0.481	1.117	0.960	1.076	1.070	0.888
	3	0.423	3415	0.230	1.130	0.940	1.112	1.201	0.879
	4	0.54	4247	0.057	1.626	1.339	1.122	1.532	1.070
	$\hat{p}$			1.005	0.935	0.948	0.985	0.935	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.323, .427, .134, .109, .007)								

Table 6: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Lee DGP, Large Sample Size (n=60,000)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.050	3746	0.001	0.083	0.084	0.945	0.035	0.036
CCT	1	0.064	4778	0.002	0.074	0.078	0.928	0.032	0.035
CCT w/o reg.	1	0.069	5157	0.002	0.073	0.079	0.920	0.032	0.035

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.007	554	1.210	1.669	1.665	1.000	1.297	1.293
	2	0.131	9852	0.690	0.807	0.800	1.003	0.905	0.890
	3	0.276	20702	0.476	0.667	0.658	1.004	0.821	0.806
	4	0.542	39983	-0.060	0.550	0.539	1.004	0.745	0.728
	$\hat{p}$			-0.033	0.568	0.557	1.003	0.749	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .058, .943)								
CCT	0	0.009	692	1.240	1.512	1.513	1.003	1.257	1.240
	2	0.151	11344	0.499	0.898	0.861	1.012	0.965	0.928
	3	0.280	20988	0.178	0.840	0.794	1.020	0.940	0.875
	4	0.357	26656	-0.021	1.022	0.963	1.025	1.043	0.956
	$\hat{p}$			0.285	0.860	0.815	1.012	0.932	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .014, .203, .751, .032)								
CCT w/o reg.	0	0.009	704	1.121	1.498	1.480	1.011	1.273	1.223
	2	0.167	12501	0.614	0.931	0.891	1.010	0.958	0.935
	3	0.321	23959	0.355	0.867	0.813	1.020	0.928	0.872
	4	0.519	37787	0.219	1.199	1.116	1.028	1.142	1.040
	$\hat{p}$			0.275	0.808	0.755	1.014	0.894	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .004, .09, .666, .241)								

Table 7: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Actual Sample Size (3,105)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.045	175	0.004	1.698	1.715	0.941	0.160	0.168
CCT	1	0.050	195	0.005	1.610	1.632	0.935	0.154	0.162
CCT w/o reg.	1	0.051	199	0.005	1.589	1.613	0.936	0.152	0.162

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.151	587	0.777	0.654	0.654	1.001	0.813	0.809
	3	0.352	1361	0.434	0.495	0.492	1.006	0.709	0.689
	4	0.723	2653	0.051	0.459	0.454	1.011	0.689	0.655
	$\hat{p}$			0.292	0.494	0.490	1.004	0.701	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, 0, .677, .323)								
CCT	2	0.167	647	0.683	0.685	0.682	1.004	0.830	0.824
	3	0.292	1133	0.199	0.746	0.736	1.010	0.863	0.839
	4	0.342	1322	0.086	0.979	0.965	1.008	0.989	0.969
	$\hat{p}$			0.516	0.681	0.675	1.005	0.821	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, .693, .306, .001)								
CCT w/o reg.	2	0.173	670	0.719	0.672	0.669	1.001	0.822	0.819
	3	0.390	1496	0.767	0.941	0.935	1.004	0.883	0.860
	4	0.537	1998	0.213	1.261	1.242	1.008	1.125	1.096
	$\hat{p}$			0.381	0.626	0.618	1.002	0.776	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, .306, .651, .043)								

Table 8: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Large Sample Size (n=30,000)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.029	1072	0.001	0.264	0.265	0.950	0.063	0.063
CCT	1	0.031	1158	0.001	0.253	0.254	0.946	0.062	0.063
CCT w/o reg.	1	0.031	1168	0.001	0.252	0.252	0.947	0.061	0.062

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.002	76	1.459	3.540	3.537	0.998	1.875	1.889
	2	0.109	4102	0.871	0.582	0.582	1.000	0.766	0.764
	3	0.273	10245	0.509	0.410	0.409	1.002	0.645	0.638
	4	0.588	21534	-0.189	0.324	0.324	1.001	0.572	0.568
	$\hat{p}$			-0.149	0.347	0.346	0.997	0.579	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .109, .891)								
CCT	0	0.002	80	1.378	3.550	3.545	1.000	1.883	1.894
	2	0.119	4450	0.876	0.593	0.593	1.001	0.771	0.768
	3	0.274	10257	0.223	0.481	0.480	1.005	0.695	0.682
	4	0.356	13288	-0.195	0.602	0.601	1.003	0.772	0.762
	$\hat{p}$			0.220	0.483	0.482	1.004	0.694	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, .008, .949, .043)								
CCT w/o reg.	0	0.002	80	1.306	3.568	3.562	0.999	1.889	1.894
	2	0.120	4515	0.861	0.588	0.588	1.000	0.768	0.768
	3	0.296	11070	0.449	0.484	0.483	1.002	0.671	0.668
	4	0.534	19349	0.361	0.698	0.696	1.001	0.845	0.843
	$\hat{p}$			0.107	0.450	0.448	0.999	0.657	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, .007, .75, .243)								

Appendix (For Online Publication Only)

A Proofs of Propositions 1 and 2

Proof of Proposition 1: First we show that $p_{opt} \to p_{max}$. For each p, Assumptions 2 and 3 ensure that

$$\frac{\text{MSE}_{\hat{\tau}_p}(h(p))}{\text{AMSE}_{\hat{\tau}_p}(h(p))} = 1 + o(1) \tag{A1}$$

by Lemma A1 of Calonico, Cattaneo and Titiunik (2014*b*), where $\text{MSE}_{\hat{\tau}_p}(h(p))$ is defined as the MSE of the estimator $\hat{\tau}_p$ with bandwidth $h(p)$. As mentioned in section 2, Assumption 3 also implies that

$$\text{AMSE}_{\hat{\tau}_p}(h(p)) = \text{C}_p \cdot n^{-\frac{2p+2}{2p+3}},$$

where C_p is a constant for each p and does not depend on n. It follows that for any $p \neq p_{max}$

$$\frac{\text{MSE}_{\hat{\tau}_{p_{max}}}(h(p_{max}))}{\text{MSE}_{\hat{\tau}_p}(h(p))} = \frac{\text{AMSE}_{\hat{\tau}_{p_{max}}}(h(p_{max}))}{\text{AMSE}_{\hat{\tau}_p}(h(p))} + o(1) \to 0$$

as $n \to \infty$. In other words, the MSE of using $\hat{\tau}_{p_{max}}$ is asymptotically smaller than using a lower-order polynomial estimator, when the bandwidths of the estimators shrink at the optimal rate. Therefore, $p_{opt} \to p_{max}$.

Next we show that $\hat{p} \xrightarrow{\mathbb{P}} p_{max}$. Under part b) of Assumption 2,

$$\frac{\widehat{\text{AMSE}}_{\hat{\tau}_p}(h(p))}{\text{AMSE}_{\hat{\tau}_p}(h(p))} \xrightarrow{\mathbb{P}} 1$$

for each $p \in \Omega$. Therefore,

$$\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{max}}}(h(p_{max}))}{\text{AMSE}_{\hat{\tau}_p}(h(p))} \xrightarrow{\mathbb{P}} 0$$

for any $p \neq p_{max}$, which implies that $\hat{p} \xrightarrow{\mathbb{P}} p_{max}$. Since $p_{opt} \to p_{max}$, $\hat{p} \xrightarrow{\mathbb{P}} p_{opt}$ as $n \to \infty$.

Proof of Proposition 2: We show that the probability $\Pr(\hat{p}/p_{opt} \neq 1)$ is arbitrarily small as $n \to \infty$.

$$\begin{aligned}
&\Pr\left(\frac{\hat{p}}{p_{opt}} \neq 1\right) \\
=&\Pr\left(\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{\hat{p}}}(h(\hat{p}))}{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} < 1\right) \\
\leqslant& \sum_{p \neq p_{opt}} \Pr\left(\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{p}}(h(p))}{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} < 1\right) \\
=& \sum_{p \neq p_{opt}} \Pr\left(\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{p}}(h(p))}{\text{MSE}_{\hat{\tau}_{p}}(h(p))} \frac{\text{MSE}_{\hat{\tau}_{p}}(h(p))}{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} \frac{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))}{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} < 1\right)
\end{aligned} \tag{A2}$$

Now we will examine the three fractions inside the probability statement of (A2) one by one. For the first fraction, Lemma A1 of Calonico, Cattaneo and Titiunik (2014*b*) and the consistency of $\hat{\text{B}}_p$ and $\hat{\text{V}}_p$ imply that

$$\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{p}}(h(p))}{\text{MSE}_{\hat{\tau}_{p}}(h(p))} \xrightarrow{\mathbb{P}} 1 \tag{A3}$$

for all p. The second fraction

$$\frac{\text{MSE}_{\hat{\tau}_{p}}(h(p))}{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} > 1$$

for all p by the definition and uniqueness of p_{opt}. For the third fraction, notice that

$$\Pr\left(\left|\frac{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))}{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} - 1\right| > \varepsilon\right) \leqslant \sum_{p \in \Omega} \Pr\left(\left|\frac{\text{MSE}_{\hat{\tau}_{p}}(h(p))}{\widehat{\text{AMSE}}_{\hat{\tau}_{p}}(h(p))} - 1\right| > \varepsilon\right). \tag{A4}$$

By Assumption 1 and condition (A3), the right hand side of can be made arbitrarily small by choosing a large enough sample size. It follows that

$$\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))}{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} \xrightarrow{\mathbb{P}} 1.$$

Putting all three fractions together, we know that, for each p,

$$\Pr\left(\frac{\widehat{\text{AMSE}}_{\hat{\tau}_{p}}(h(p))}{\text{MSE}_{\hat{\tau}_{p}}(h(p))} \frac{\text{MSE}_{\hat{\tau}_{p}}(h(p))}{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} \frac{\text{MSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))}{\widehat{\text{AMSE}}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} < 1\right)$$

can be made arbitrarily small by choosing a large enough sample size. It follows that $\Pr(\hat{p}/p_{opt} \neq 1) \to 0$ as $n \to \infty$ and that $\hat{p}/p_{opt} \xrightarrow{\mathbb{P}} 1$.

B Specifications of Data Generating Processes

B.1 Lee and Ludwig-Miller DGPs

To obtain the conditional expectation functions in the Lee and Ludwig-Miller DGPs, Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*) first discard the outliers in the empirical data (i.e. observations for which the absolute value of the running variable is very large) and then fit a separate quintic function on each side of the cutoff to the remaining observations. The conditional expectation functions are

$$\text{Lee: } E[Y|X=x] = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0 \\ 0.52 + 0.84x - 3.00x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{if } x \geqslant 0 \end{cases} \tag{A5}$$

$$\text{Ludwig-Miller: } E[Y|X=x] = \begin{cases} 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < 0 \\ 0.26 + 18.49x - 54.81x^2 + 74.30x^3 - 45.02x^4 + 9.83x^5 & \text{if } x \geqslant 0. \end{cases} \tag{A6}$$

Equations (A5) and (A6) are graphed in Appendix Figure A.1. As seen in the formulations above and as presented graphically, the Ludwig-Miller DGP has very large slope and curvature above the cutoff as compared to the Lee DGP.

The assignment variable X is specified as following the distribution $2\mathscr{B}(2,4) - 1$, where $\mathscr{B}(\alpha,\beta)$ denotes a beta distribution with shape parameters α and β. The outcome variable is given by $Y = E[Y|X = x] + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.1295$.

B.2 Card-Lee-Pei-Weber DGPs

The process of specifying the Card-Lee-Pei-Weber DGPs are described in section 4.4.3 of Card et al. (2017). Below we state the parameters in the bottom- and top-kink DGPs respectively.

B.2.1 Bottom-kink DGP

The first-stage and reduced-form conditional expectation functions for the bottom-kink DGP are specified as

$$\text{First-stage: } E[B|X=x] = \begin{cases} \beta_0 + \beta_1^+ x + \beta_2^+ x^2 + \beta_3^+ x^3 + \beta_4^+ x^4 + \beta_5^+ x^5 & \text{if } x < 0 \\ \beta_0 + \beta_1^- x + \beta_2^- x^2 + \beta_3^- x^3 + \beta_4^- x^4 + \beta_5^- x^5 & \text{if } x \geqslant 0 \end{cases} \tag{A7}$$

$$\text{Reduced-form: } E[Y|X=x] = \begin{cases} \gamma_0 + \gamma_1^+ x + \gamma_2^+ x^2 + \gamma_3^+ x^3 + \gamma_4^+ x^4 + \gamma_5^+ x^5 & \text{if } x < 0 \\ \gamma_0 + \gamma_1^- x + \gamma_2^- x^2 + \gamma_3^- x^3 + \gamma_4^- x^4 + \gamma_5^- x^5 & \text{if } x \geqslant 0 \end{cases} \tag{A8}$$

where

- $\beta_0 = 3.17$
- $\beta_1^+ = 3.14 \times 10^{-5}$; $\beta_1^- = 8.40 \times 10^{-6}$
- $\beta_2^+ = 5.30 \times 10^{-9}$; $\beta_2^- = -1.21 \times 10^{-8}$
- $\beta_3^+ = -3.82 \times 10^{-12}$; $\beta_3^- = -1.01 \times 10^{-11}$
- $\beta_4^+ = 9.54 \times 10^{-16}$; $\beta_4^- = -7.56 \times 10^{-16}$
- $\beta_5^+ = -8.00 \times 10^{-20}$; $\beta_5^- = 7.89 \times 10^{-19}$
- $\gamma_0 = 4.51$
- $\gamma_1^+ = -1.76 \times 10^{-5}$; $\gamma_1^- = -4.75 \times 10^{-5}$
- $\gamma_2^+ = 7.00 \times 10^{-9}$; $\gamma_2^- = 1.64 \times 10^{-7}$
- $\gamma_3^+ = -5.00 \times 10^{-12}$; $\gamma_3^- = 3.04 \times 10^{-10}$
- $\gamma_4^+ = 1.00 \times 10^{-15}$; $\gamma_4^- = 1.82 \times 10^{-13}$
- $\gamma_5^+ = -2.00 \times 10^{-19}$; $\gamma_5^- = 3.53 \times 10^{-17}$
- The conditional variances of B given X just above and below the cutoff are 2.05×10^{-4} and 2.07×10^{-4}, respectively.
- The conditional variances of Y given X just above and below the cutoff are 1.51 and 1.49, respectively.

- The density f_X evaluated at 0 is: 1.53×10^{-4}.

B.2.2 Top-kink DGP

The first-stage and reduced-form conditional expectation functions for the top-kink DGP are specified as quintic functions on both sides of the cutoff as in equations (A7) and (A8). The coefficients are:

- $\beta_0 = 3.65$
- $\beta_1^+ = -3.70 \times 10^{-6}$; $\beta_1^- = 1.03 \times 10^{-5}$
- $\beta_2^+ = 1.25 \times 10^{-8}$; $\beta_2^- = -3.18 \times 10^{-9}$
- $\beta_3^+ = -6.17 \times 10^{-12}$; $\beta_3^- = -5.72 \times 10^{-13}$
- $\beta_4^+ = 1.16 \times 10^{-15}$; $\beta_4^- = -4.83 \times 10^{-17}$
- $\beta_5^+ = -7.43 \times 10^{-20}$; $\beta_5^- = -1.42 \times 10^{-21}$
- $\gamma_0 = 4.65$
- $\gamma_1^+ = -1.29 \times 10^{-5}$; $\gamma_1^- = 1.51 \times 10^{-5}$
- $\gamma_2^+ = 2.35 \times 10^{-8}$; $\gamma_2^- = -5.69 \times 10^{-9}$
- $\gamma_3^+ = -1.42 \times 10^{-11}$; $\gamma_3^- = -1.07 \times 10^{-12}$
- $\gamma_4^+ = 3.04 \times 10^{-15}$; $\gamma_4^- = -8.49 \times 10^{-17}$
- $\gamma_5^+ = -2.06 \times 10^{-19}$; $\gamma_5^- = -2.65 \times 10^{-21}$
- The conditional variances of B given X just above and below the cutoff are 1.20×10^{-3} and 9.60×10^{-4}, respectively.
- The conditional variances of Y given X just above and below the cutoff are 1.62 and 1.63, respectively.
- The density f_X evaluated at 0 is: 2.35×10^{-5}.

C AMSE Calculation and Estimation

C.1 Theoretical AMSE Calculation

After the full specification of a data generating process, we can calculate $\text{AMSE}_{\hat{\tau}_p}(h)$ by applying Lemma 1 of Calonico, Cattaneo and Titiunik (2014*b*) in a sharp design and Lemma 2 in a fuzzy design. The lemmas provide the expressions for the constants in the squared-bias and variance terms, B_p^2 and V_p, that make up $\text{AMSE}_{\hat{\tau}_p}(h)$ according to equation (3). Specifically, B_p^2 depends on the $(p+1)$th derivatives on both sides of the cutoff, and V_p depends on the conditional variances on both sides of the cutoff as well as the density of the running variable at the cutoff. With B_p^2 and V_p computed, we can calculate the infeasible optimal bandwidth h_{opt} for a given sample size, which is simply a function of B_p^2 and V_p. Finally, plugging h_{opt} back into $\text{AMSE}_{\hat{\tau}_p}(h)$ yields the AMSE for that given sample size, and Figure 1 is the graphical representation of this mapping across different sample sizes (the Mathematica program used to generate the figure is available at https://sites.google.com/site/peizhuan/programs/).

C.2 AMSE Estimation

To estimate $\text{AMSE}_{\hat{\tau}_p}$, we rely on the proposed procedure in Calonico, Cattaneo and Titiunik (2014*a*,*b*). Our program `rdmse_cct2014` takes user-specified bandwidths as inputs and estimates $\hat{\text{B}}_p^2$ and $\hat{\text{V}}_p$ for the conventional estimator in the same way as Calonico, Cattaneo and Titiunik (2014*b*). We also provide another program `rdmse`, which speeds up the computation in `rdmse_cct2014` by modifying variance estimations. As with Calonico, Cattaneo and Titiunik (2014*a*), `rdmse` implements a nearest-neighbor estimator as per Abadie and Imbens (2006) and sets the number of neighbors to three. However, in the event of a tie, while Calonico, Cattaneo and Titiunik (2014*a*) selects all of the closest neighbors, we randomly select three neighbors. We adopt the same modification in Card et al. (2015*a*).

Additionally, `rdmse` estimates the AMSE of the bias-corrected RD or RK estimator $\hat{\tau}_p^{bc}$:

$$\widehat{\text{AMSE}}_{\hat{\tau}_p^{bc}}(h,b) = \left(\tilde{\text{B}}_p^{bc}(h,b)\right)^2 + \tilde{\text{V}}_p^{bc}(h,b),$$

where b is the pilot bandwidth used in Calonico, Cattaneo and Titiunik (2014*b*) to estimate the bias of $\hat{\tau}_p$. According to Theorems A.1 and A.2 of Calonico, Cattaneo and Titiunik (2014*b*), the bias of $\hat{\tau}_p^{bc}$ has two terms: the first term is the higher-order approximation error post bias-correction, and the second term captures the bias in estimating the bias of $\hat{\tau}_p$. These two terms involve the $(p+2)$th derivatives of the conditional expectation function on both sides of the cutoff, which are estimated via local polynomial regressions in the CCT bandwidth selection procedure for the sharp design, and in the "fuzzy CCT" bandwidth selection procedure of Card et al. (2015*a*). We follow the same algorithm to arrive at $\tilde{\text{B}}_p^{bc}$. $\tilde{\text{V}}_p^{bc}$ is simply the estimated variance of $\hat{\tau}_p^{bc}$, and its computation is covered in detail in Calonico, Cattaneo and Titiunik (2014*b*). In Table A.6, we provide details on the AMSE calculations in our software implementation by presenting the correspondence between the expressions in this paper and those in Calonico, Cattaneo and Titiunik (2014*a*,*b*).

Finally, as mentioned in Appendix A, our AMSE estimator is consistent for the true MSE in a sharp design. Consistency in the fuzzy design and for $\widehat{\text{AMSE}}_{\hat{\tau}_p^{bc}}(h,b)$ can be similarly established.

Figure A.1: Conditional Expectation Functions in RDD DGPs

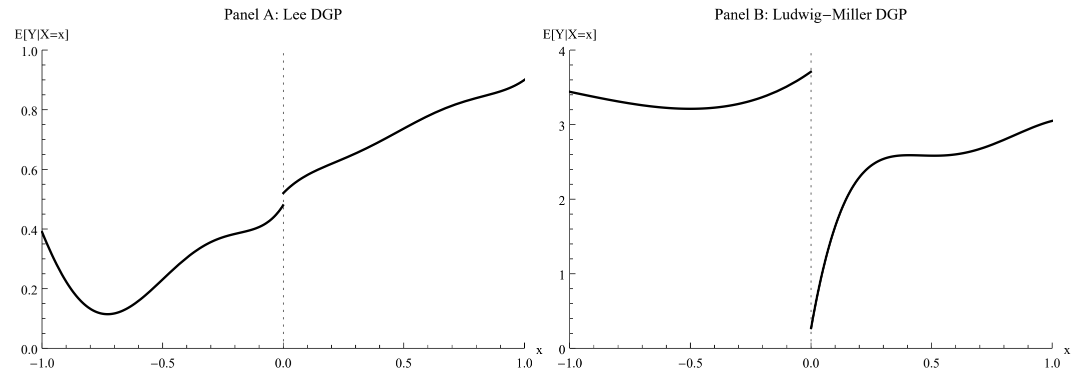


Figure A.2: Conditional Expectation Functions in RKD DGPs

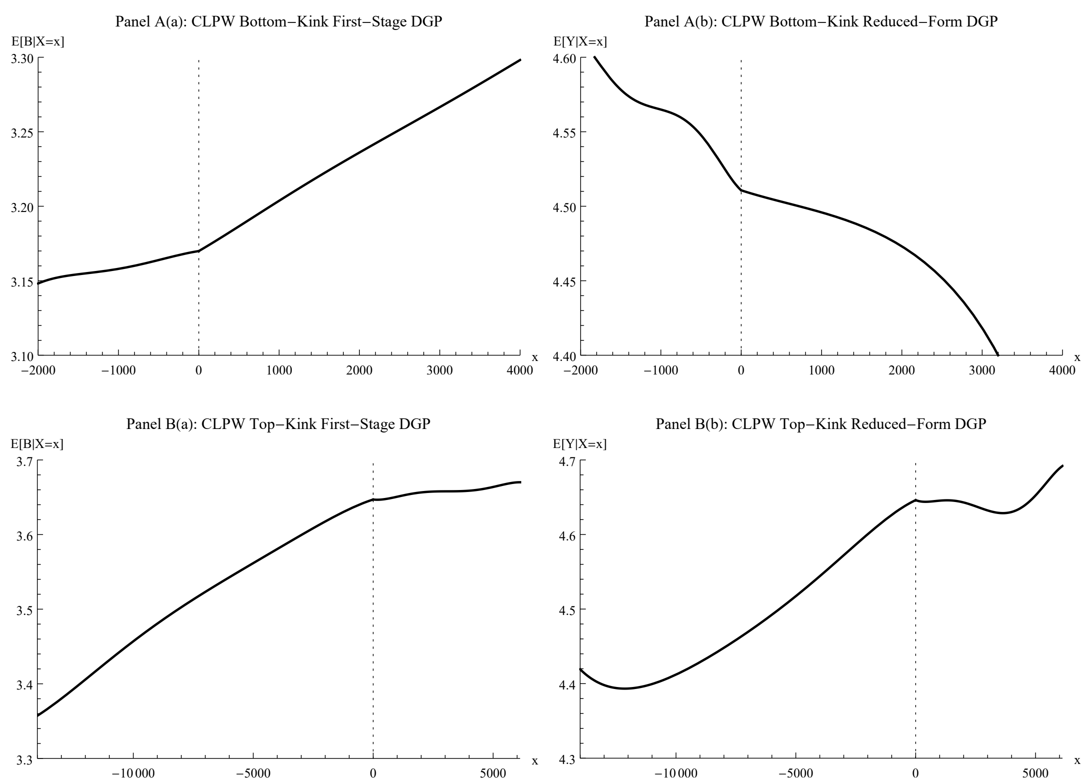


Table A.1: Main Specification of RD Papers Published in Leading Journals

Main Specification	Number of Papers	1999-2010	2011-2017
Local constant	11	8	3
Local linear	45	9	36
Local quadratic	6	1	5
Local cubic	5	4	1
Local quartic	2	2	0
Local 7th-order	1	1	0
Local 8th-order	1	0	1
Local but did not mention preferred polynomial	5	0	5
Total local	76	25	51
Global linear	4	1	3
Global quadratic	4	0	4
Global cubic	11	5	6
Global quartic	4	2	2
Global 5th-order	1	0	1
Global 8th-order	1	0	1
Global but did not mention preferred polynomial	1	0	1
Total global	26	8	18
Did not mention preferred specification	8	2	6
Total	110	35	75

Note: Our survey includes empirical RD papers published between 1999 and 2017 in the following leading journals: *American Economic Review*, *American Economic Journals*, *Econometrica, Journal of Political Economy*, *Journal of Business and Economic Statistics*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Review of Economics and Statistics* in our survey.

Table A.2: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Lee DGP, Small Sample Size (n=500)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.130	81	0.019	3.570	3.922	0.927	0.230	0.256
CCT	1	0.159	99	0.021	3.496	3.952	0.902	0.211	0.248
CCT w/o reg.	1	0.387	218	0.029	2.460	3.300	0.803	0.155	0.224

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.059	35	1.814	0.917	1.130	0.949	0.940	1.092
	2	0.260	162	0.811	1.128	1.086	1.007	1.064	1.035
	3	0.470	291	0.762	1.138	1.088	1.006	1.066	1.041
	4	0.838	472	0.844	1.104	1.069	1.007	1.053	1.031
	$\hat{p}$			1.163	1.129	1.149	0.949	0.945	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.468, .397, .04, .048, .047)								
CCT	0	0.059	37	2.382	0.686	1.263	0.803	0.816	1.140
	2	0.226	141	0.469	1.513	1.364	1.028	1.256	1.181
	3	0.275	172	0.107	2.213	1.959	1.027	1.528	1.436
	4	0.313	195	0.004	3.134	2.772	1.022	1.810	1.709
	$\hat{p}$			2.054	0.797	1.193	0.829	0.827	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.717, .28, .003, 0, 0)								
CCT w/o reg.	0	0.085	51	2.146	1.003	1.920	0.743	0.974	1.542
	2	0.424	252	0.733	1.667	1.380	1.077	1.336	1.201
	3	0.464	280	0.469	2.418	1.858	1.117	1.665	1.414
	4	0.491	297	0.064	3.303	2.464	1.139	2.015	1.614
	$\hat{p}$			1.598	1.053	1.435	0.868	0.902	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.387, .566, .046, .002, 0)								

Table A.3: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Small Sample Size (n=500)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

Bandwidth	(1) p	(2) Avg. h	(3) Avg. n	(4) Bias	(5) Variance x1000	(6) MSE ×1000	(7) Coverage Rate	(8) Avg. CI Length	(9) Avg. Size-adj. CI length
Theo. Optimal	1	0.065	41	0.037	7.863	9.215	0.922	0.366	0.412
CCT	1	0.076	47	0.049	7.449	9.821	0.881	0.329	0.422
CCT w/o reg.	1	0.079	49	0.053	7.333	10.099	0.869	0.320	0.424

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.196	123	0.594	0.695	0.645	1.016	0.827	0.777
	3	0.431	268	0.384	0.565	0.504	1.023	0.742	0.682
	4	0.854	477	0.150	0.497	0.427	1.028	0.702	0.628
	$\hat{p}$			0.241	0.555	0.482	1.019	0.716	
	Fraction of time $\hat{p}$=(1,2,3,4): (.001, .003, .486, .511)								
CCT	2	0.206	129	0.488	0.792	0.658	1.047	0.890	0.772
	3	0.280	175	0.074	1.014	0.770	1.066	1.040	0.855
	4	0.319	199	0.016	1.431	1.086	1.066	1.254	1.028
	$\hat{p}$			0.462	0.825	0.677	1.041	0.887	
	Fraction of time $\hat{p}$=(1,2,3,4): (.022, .876, .101, .001)								
CCT w/o reg.	2	0.240	150	0.660	0.814	0.711	1.020	0.841	0.821
	3	0.448	272	0.245	0.804	0.600	1.050	0.854	0.755
	4	0.504	303	0.019	1.062	0.771	1.070	1.030	0.845
	$\hat{p}$			0.479	0.796	0.641	1.024	0.811	
	Fraction of time $\hat{p}$=(1,2,3,4): (.007, .434, .501, .058)								

Table A.4: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Lee DGP, Small Sample Size (n=500)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.130	81	0.012	4.607	4.752	0.935	0.262	0.280
CCT	1	0.160	100	0.015	4.671	4.892	0.920	0.245	0.273
CCT w/o reg.	1	0.387	219	0.026	4.376	5.067	0.869	0.240	0.305

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.059	23	1.123	0.987	0.995	0.962	0.980	1.036
	2	0.260	162	0.795	1.010	0.998	1.000	1.006	0.994
	3	0.470	291	0.646	0.982	0.964	1.000	0.990	0.981
	4	0.838	472	0.062	1.207	1.170	1.000	1.089	1.083
	$\hat{p}$			0.922	1.017	1.012	0.966	0.949	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.496, .218, .125, .157, .005)								
CCT	0	0.059	37	1.865	0.625	0.754	0.976	0.824	0.886
	2	0.226	141	0.263	1.398	1.338	1.013	1.215	1.181
	3	0.276	172	-0.026	1.967	1.878	1.014	1.451	1.409
	4	0.314	195	-0.076	2.694	2.573	1.011	1.704	1.653
	$\hat{p}$			1.859	0.637	0.765	0.974	0.821	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.932, .067, .001, 0, 0)								
CCT w/o reg.	0	0.085	52	1.326	0.651	0.802	0.962	0.734	0.794
	2	0.424	252	0.320	1.544	1.347	1.053	1.276	1.151
	3	0.466	281	0.074	2.434	2.102	1.071	1.644	1.407
	4	0.495	299	0.020	5.586	4.824	1.073	2.033	1.721
	$\hat{p}$			1.313	0.588	0.743	0.953	0.718	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.877, .118, .005, 0, 0)								

Table A.5: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Small Sample Size (n=500)

(a): Simulation Statistics for the Local Linear Estimator (p=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bandwidth	p	Avg. h	Avg. n	Bias	Variance x1000	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.065	41	0.010	8.475	8.565	0.941	0.381	0.398
CCT	1	0.076	47	0.012	7.854	7.990	0.933	0.353	0.379
CCT w/o reg.	1	0.079	49	0.013	7.629	7.799	0.932	0.345	0.370

(b): Simulation Statistics for Other Polynomial Orders as Compared to p=1

Bandwidth	p	Avg. h	Avg. n	Ratio of Biases	Ratio of Variances	Ratio of MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.196	123	0.705	0.688	0.686	1.006	0.821	0.804
	3	0.431	268	0.363	0.563	0.558	1.007	0.738	0.717
	4	0.854	477	0.081	0.656	0.649	1.008	0.808	0.783
	$\hat{p}$			0.375	0.597	0.592	1.002	0.745	
	Fraction of time $\hat{p}$=(1,2,3,4): (.006, .113, .834, .047)								
CCT	2	0.206	128	0.353	0.837	0.825	1.013	0.918	0.876
	3	0.281	175	0.019	1.135	1.116	1.011	1.081	1.037
	4	0.319	199	0.088	1.629	1.601	1.008	1.293	1.252
	$\hat{p}$			0.436	0.855	0.843	1.004	0.908	
	Fraction of time $\hat{p}$=(1,2,3,4): (.168, .796, .035, .001)								
CCT w/o reg.	2	0.240	149	0.610	0.895	0.884	1.006	0.882	0.859
	3	0.448	273	0.450	1.387	1.361	1.013	1.181	1.130
	4	0.508	305	0.111	2.544	2.488	1.012	1.569	1.503
	$\hat{p}$			0.399	0.789	0.775	1.000	0.855	
	Fraction of time $\hat{p}$=(1,2,3,4): (.068, .737, .189, .007)								

Table A.6: Correspondence to the Expressions in Calonico, Cattaneo and Titiunik (2014*a*,*b*)

Expression in this paper	Expression in Calonico, Cattaneo and Titiunik (2014*a*,*b*) for the case of Sharp RD ($\nu = 0$)/RK ($\nu = 1$)	Fuzzy RD ($\nu = 0$)/RK ($\nu = 1$)
B_p	$\mathrm{B}_{\nu,p,p+1,0}$[SAp.38]	$\mathrm{B}_{\mathrm{F},\nu,p,p+1}$[SAp.39]
V_p	$\mathrm{V}_{\nu,p}$[SAp.38]	$\mathrm{V}_{\mathrm{F},\nu,p}$[SAp.39]
$\hat{\mathrm{B}}_p$	$\hat{B}_{n,p,q}$[SJp.920]	$\hat{B}_{n,p,q}$[SJp.920]
$\hat{\mathrm{V}}_p$	$\hat{V}_p$[SJp.920]	$\hat{V}_p$[SJp.920]
$\tilde{\mathrm{B}}_p^{bc}(h,b)$	Estimator of $h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+1}(h_n) - h_n^{p+1-\nu} b_n^{q-p}\mathrm{B}^{\mathrm{bc}}_{\nu,p,q}(h_n,b_n)$[p.2321]	Estimator of $h_n^{p+2-\nu}\mathrm{B}_{\mathrm{F},\nu,p,p+1}(h_n) - h_n^{p+1-\nu} b_n^{q-p}\mathrm{B}^{\mathrm{bc}}_{\mathrm{F},\nu,p,q}(h_n,b_n)$[p.2323]
$\tilde{\mathrm{V}}_p^{bc}(h,b)$	$\hat{V}^{\mathrm{bc}}_{n,p,q}$[SJp.922]	$\hat{V}^{\mathrm{bc}}_{n,p,q}$[SJp.922]

Note: The number after "p." and "SAp." refers to the page on which the particular expression appears in the main article or the Supplemental Appendix of Calonico, Cattaneo and Titiunik (2014*b*), respectively. The number after "SJp." refers to the page on which the particular expression appears in Calonico, Cattaneo and Titiunik (2014*a*). We set $q = p + 1$ for all of our estimators, which is the default used by Calonico, Cattaneo and Titiunik (2014*b*).