Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Local Polynomial Order in Regression Discontinuity Designs

Zhuan Pei, David S. Lee, David Card & Andrea Weber

To cite this article: Zhuan Pei, David S. Lee, David Card & Andrea Weber (2021): Local Polynomial Order in Regression Discontinuity Designs, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2021.1920961

To link to this article: https://doi.org/10.1080/07350015.2021.1920961

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2021, VOL. 00, NO. 0, 1–9
https://doi.org/10.1080/07350015.2021.1920961

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Local Polynomial Order in Regression Discontinuity Designs

Zhuan Pei[a,b], David S. Lee[c,d], David Card[b,d,e], and Andrea Weber[b,f]

[a]Department of Policy Analysis and Management, Cornell University, Ithaca, NY; [b]IZA Institute of Labor Economics, Bonn, Germany; [c]Department of Economics, Princeton University, Princeton, NJ; [d]National Bureau of Economic Research, Cambridge, MA; [e]Department of Economics, UC Berkeley, Berkeley, CA; [f]Department of Economics and Business, Central European University, Budapest, Hungary

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 and Calonico, Cattaneo, and Titiunik provided 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 and Calonico, Cattaneo, and Titiunik 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.

ARTICLE HISTORY
Received June 2020
Accepted April 2021

KEYWORDS
Local polynomial estimation; Polynomial order; Regression discontinuity design; Regression kink design;

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; 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 and found that of the 110 studies employing RDD, 76 use a local polynomial regression as their main specification (Table A1, supplementary material). 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 reassured 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), and it is the focus of the present article.

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) pointed out, higher order polynomials have a smaller asymptotic bias than lower orders. On the other hand, Gelman and Imbens (2019) argued that high-order polynomials can perform poorly in certain contexts.

In this article, 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)—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. Thus, the proposed procedure is based on a local (as opposed to global) optimality criterion, as advocated by Gelman and Imbens (2019).

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. 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

CONTACT Zhuan Pei zhuan.pei@cornell.edu Department of Policy Analysis and Management, Cornell University, Ithaca, NY 14853.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

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). 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, 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) in all these contexts. We have implemented these computations in a Stata package `rdmse`. The installation instruction and program documentation are available online at *https://peizhuan.github.io/programs/*.

The remainder of the article is organized as follows. Section 2 summarizes the theory of local polynomial RD estimators, and the corresponding Appendix A (appendix, supplementary material) shows the 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 RDDs and RKDs. Section 5 concludes.

2. RD Local Polynomial Order: 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) showed 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 local polynomial regressions to separately estimate its two terms. Specifically, they solve the minimization problem using only observations above the cutoff as denoted by the + superscript:

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

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 the first-order asymptotic approximations for the bias and variance. Applying these ideas, Lemma 1 of Calonico, Cattaneo, and Titiunik (2014) derived the AMSE of the pth order local polynomial estimator $\hat{\tau}_p$ as a function of bandwidth

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

where B_p and V_p are unknown constants that depend on the properties of the data-generating process (DGP). The AMSE approximates the conditional MSE of $\hat{\tau}_p$ with bandwidth h: $\text{MSE}_{\hat{\tau}_p}(h) \equiv E[(\hat{\tau}_p(h) - \tau)^2|\mathbf{X}]$, where $\mathbf{X} = [X_1, \ldots, X_n]$ consists of X of all n sample observations. The first term of the AMSE is the approximate squared bias, and the second term the approximate variance.

First-order approximations like the one above have been used in the literature in two ways. First, Hahn, Todd, and Van der Klaauw (2001) argued 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 $h^2 B_1$ and hB_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) recommended $p = 1$, 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: $h_{\text{opt}}(p) \equiv \arg\min_h \text{AMSE}_{\hat{\tau}_p}(h)$. Imbens and Kalyanaraman (2012) did precisely this to propose a bandwidth selector for local linear estimation (henceforth IK bandwidth) and Calonico, Cattaneo, and Titiunik (2014) further extended 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 $h_{\text{opt}}(p)$, which is of order $O(n^{-\frac{1}{2p+3}})$, $\text{AMSE}_{\hat{\tau}_p}(h_{\text{opt}}(p))$ is equal to $\text{C}_p \cdot n^{-\frac{2p+2}{2p+3}}$ with C_p being a function of the constants B_p and V_p. Therefore, as the sample size n increases, $\text{AMSE}_{\hat{\tau}_p}(h_{\text{opt}}(p))$ 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 DGP through the constant C_p.

This point is concretely illustrated in Figure 1, using the two DGPs we employ for subsequent simulations. The DGPs are based on Lee (2008) and Ludwig and Miller (2007) and

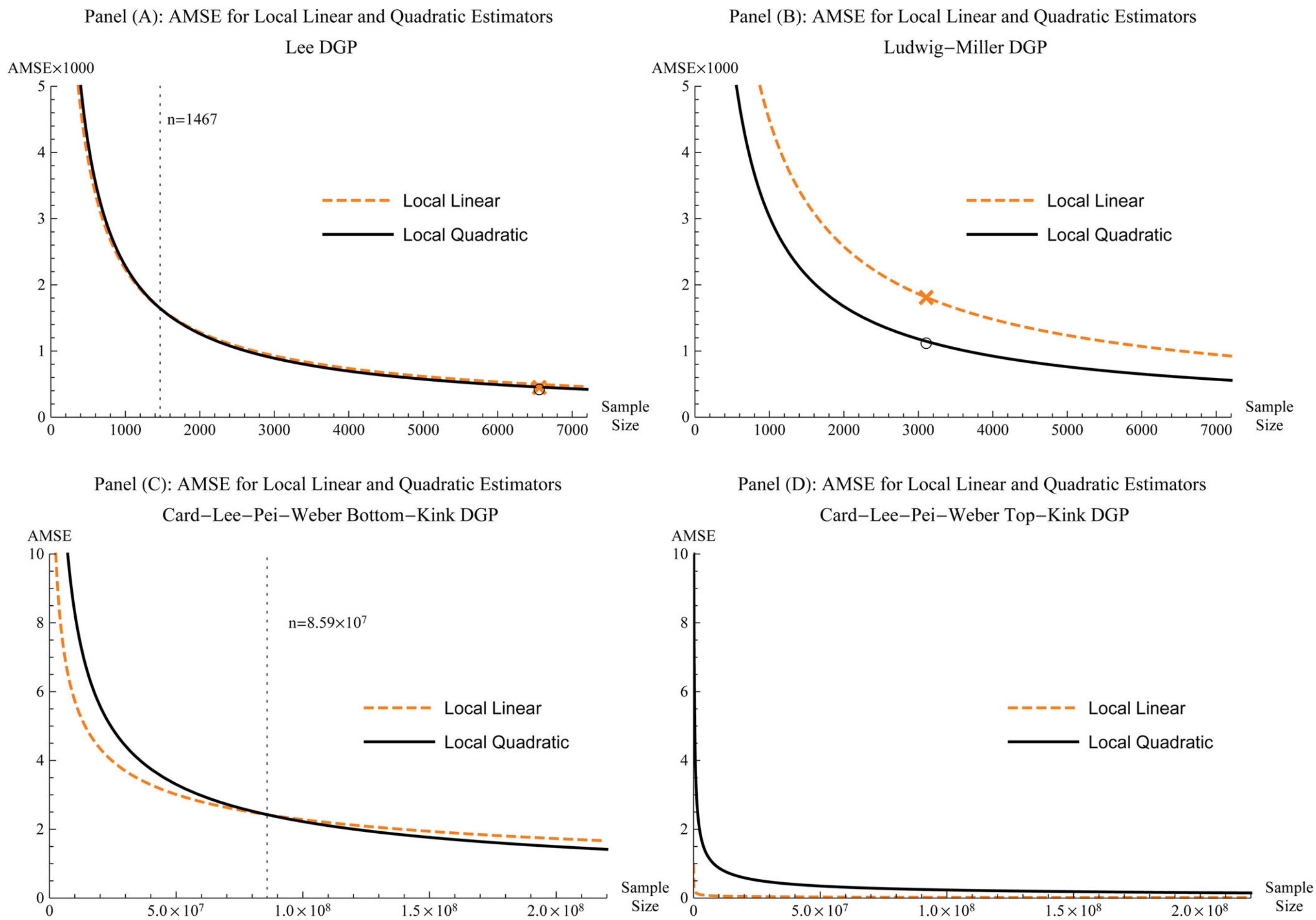

Figure 1. Asymptotic mean-squared-error as a function of sample size.
Note: We plot theoretical AMSEs as functions of sample size in two RD and two RK DGPs. We calculate the AMSEs for local linear and quadratic estimators with triangular kernel and the theoretical MSE-optimal bandwidth. In Panels (A) and (B), we superimpose the simulated MSEs of the local linear (cross) and quadratic (circle) estimators. These MSEs are taken from Tables 1 and 2. We discuss the rate at which the MSE-optimal polynomial order increases with sample size in the Remark below Proposition 1 in Appendix A (supplementary material).

described in greater detail in Appendix B.1 (supplementary material). 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), we plot in Figure 1 the $\text{AMSE}_{\hat{\tau}_p}$ under the triangular kernel in the two DGPs as a function of sample size n for $p = 1, 2$ (see Appendix C.1 for details, supplementary material).

For the Lee (2008) DGP in Panel (A), $\text{AMSE}_{\hat{\tau}_1}$ is marginally below $\text{AMSE}_{\hat{\tau}_2}$ at small sample sizes but is larger at sample sizes over $n = 1467$. Therefore, for the actual sample size in Lee (2008), $n_{\text{actual}} = 6558$, local quadratic should be preferred to local linear based on the AMSE comparison—the associated reduction in AMSE is 8%. For the Ludwig and Miller (2007) DGP 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 7000. At the actual sample size in Ludwig and Miller (2007), $n_{\text{actual}} = 3105$, the local quadratic estimator reduces the AMSE by a considerable 37%. It is worth noting that at n_{actual}, the AMSEs closely match the MSEs 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 B_p and V_p depend on unknown quantities such as the derivatives of the conditional expectation function, conditional variances, and the density of X. Thus, Imbens and Kalyanaraman (2012) and Calonico, Cattaneo, and Titiunik (2014) used the empirical analog of Equation (3) for the local linear estimator

$$\widehat{\text{AMSE}}_{\hat{\tau}_1}(h) = h^4 \hat{\text{B}}_1^2 + \frac{1}{nh}\hat{\text{V}}_1, \tag{4}$$

where B_1 and V_1 are replaced by consistent estimators $\hat{\text{B}}_1$ and $\hat{\text{V}}_1$, and the MSE-optimal feasible bandwidth is defined as $\hat{h}(1) \equiv \arg\min_h \widehat{\text{AMSE}}_{\hat{\tau}_1}(h)$. The two studies differ in how they arrive at the estimates of B_1 and V_1. Additionally, Calonico, Cattaneo, and Titiunik (2014) generalized Imbens and Kalyanaraman (2012) by proposing bandwidth selectors for $\hat{\tau}_p$ for any p.

In this article, we simply extend the logic that justifies the optimal bandwidth by noting that we can choose the polynomial order corresponding to the lowest estimated AMSE. That is, we 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; see Appendix A (supplementary material) for more discussion of Ω). For the AMSE of $\hat{\tau}_p$, no new quantities need to be computed beyond the estimators $\hat{\mathrm{B}}_p$ and $\hat{\mathrm{V}}_p$ and the optimal $\hat{h}\left(p\right)$, 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 $\widehat{\mathrm{AMSE}}_{\hat{\tau}_p}$ for any given p and compare $\widehat{\mathrm{AMSE}}_{\hat{\tau}_p}$ across different candidate polynomial orders. Appendix C.2 (supplementary material) provides the exact expressions needed from Calonico, Cattaneo, and Titiunik (2014) for the calculation of the AMSE of $\hat{\tau}_p$, which is implemented in the Stata package `rdmse`.

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 has yet to be discussed, and the approach has yet to be applied to RD designs. In Appendix A (supplementary material), we investigate the asymptotic property of the procedure and prove the consistency of $\hat{p}$ in two asymptotic frameworks that have been invoked in the literature.

Before proceeding to examine the finite sample performance of $\hat{p}$, we make several remarks.

Remark 1. We can also estimate the AMSE of the *bias-corrected estimator* of Calonico, Cattaneo, and Titiunik (2014) (denoted by $\hat{\tau}_p^{bc}$). Appendix C.2 (supplementary material) provides details, and the calculation is also implemented in the Stata package `rdmse`.

Remark 2. We can allow for different polynomial orders on two sides of the threshold, similar to recent developments that permit different bandwidths. Calonico et al. (2017, 2019) implemented bandwidths that are optimal for the left and right intercept estimators, respectively. Following this line of reasoning, our Stata package can calculate the AMSE of each intercept estimator of a given polynomial order with the Calonico et al. (2017, 2019) bandwidths. Another recent study by Arai and Ichimura (2018) proposed simultaneous left and right bandwidth selectors that are MSE-optimal for the sharp RD estimator. It is also possible to extend Arai and Ichimura (2018) and jointly select left and right polynomial orders that are optimal for the RD estimator itself. In fact, the finiteness of the polynomial choice set makes the exercise easier than the bandwidth selection by Arai and Ichimura (2018), who have to innovate to avoid a degenerate optimization problem.

Remark 3. Calonico et al. (2019) considered the identification, estimation, and inference in local RD regressions with covariates. Among other contributions, they propose covariate-adjusted MSE-optimal bandwidth selectors, which require the estimation of covariate-adjusted biases and variances. This article can be extended to select polynomial orders after covariate incorporation by building on Calonico et al. (2019).

Remark 4. Our MSE-optimal polynomial order selection procedure stems from the perspective of point estimation and not inference. Calonico, Cattaneo, and Farrell (2020) recently showed that the inference-optimal bandwidth that minimizes confidence interval coverage error rate is different from the MSE-optimal bandwidth (the former shrinks faster as a function of n). The same may also be true for polynomial order choices. Future work can study inference-optimal polynomial orders by building on the Edgeworth expansion approach in Calonico, Cattaneo, and Farrell (2020).

Remark 5. There exist alternative econometric estimation and inference approaches to the local polynomial paradigm, but many still require a polynomial order as input. Our proposal is applicable to frequentist approaches based on local approximation, for example, Otsu, Xu, and Matsushita (2015), whose empirical likelihood procedure relies on moment conditions formulated from the local linear RD estimator. One could adapt the Otsu, Xu, and Matsushita (2015) procedure by starting with our MSE-optimal polynomial order. In contrast, our way of calculating the AMSE does not apply to the local randomization approach by Cattaneo, Titiunik, and Vazquez-Bare (2017), where the polynomial choice amounts to a parametric assumption, or the order of global polynomial fit lines in RD graphs (Calonico, Cattaneo, and Titiunik 2015), for which Lee and Lemieux (2010) suggested selection procedures based on goodness-of-fit criteria. Finally, our frequentist proposal does not apply to the Bayesian RD approach of Geneletti et al. (2015), and we leave the polynomial order choice therein as an open question. (In another Bayesian study, Branson et al. 2019 largely circumvented this polynomial choice by modeling the potential outcome means conditional on X as Gaussian processes; although the mean functions of the Gaussian processes are still specified as polynomials, their choice is shown to be inconsequential in examples.)

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. In this section, we conduct Monte Carlo simulations to examine the finite sample performance of local polynomial estimators of various orders—which themselves use 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). The conditional expectation functions are specified as piecewise quintic polynomials (see Appendix B.1 for details, supplementary material). Because of the fifth-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 triangular kernel; additional results using the uniform kernel are available in the previous working article (Pei et al. 2020), and the qualitative conclusions are the same.

The simulation results are organized as follows. Tables 1 and 2 report on the performances of conventional RD estimators ($\hat{\tau}_p$) applied to the two DGPs, respectively, while Tables B.1 and B.2 (appendix, supplementary material) report on the bias-corrected RD estimators ($\hat{\tau}_p^{\text{bc}}$) and the associated robust confidence intervals as per Calonico, Cattaneo, and Titiunik (2014). Each of the four tables displays results corresponding to two sample sizes: the actual sample size in Panel A and large sample size in Panel B. The actual sample size is that of the analysis sample in the two empirical studies: $n_{\text{actual}} = 6558$ for Lee (2008) and $n_{\text{actual}} = 3105$ for Ludwig and Miller (2007). We set the large sample size to $n_{\text{large}} = 60{,}000$ for the Lee DGP and $n_{\text{large}} = 30{,}000$ for Ludwig–Miller. n_{large} is about $10 \times n_{\text{actual}}$ in both studies, and it is comparable to or lower than the n in many empirical papers.

In part (a) of each panel, we show the summary statistics for the local linear estimator with two bandwidths. The first bandwidth is the (infeasible) theoretical optimal bandwidth (h_{opt}), which minimizes AMSE using knowledge of the underlying DGP. Even though the theoretically optimal bandwidth is never known in an empirical application, we present 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, which is consistent with our discussion of the asymptotic behavior of $\text{AMSE}_{\hat{\tau}_p}(h_{\text{opt}}(p))$ in Section 2. The second bandwidth is the default CCT bandwidth selector from Calonico, Cattaneo, and Titiunik (2014) ($\hat{h}_{\text{CCT}}$).

We report averages and percentages across the simulations: average bandwidth in column (2), average number of observations within the bandwidth in column (3), MSE in column (4), coverage rate of the 95% CI in column (5), the average CI length in column (6), and the average size-adjusted CI length in column (7). 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 (7).

In part (b) of each table, we present the same statistics for different polynomial orders. In columns (4), (6), and (7), we express the quantities as a ratio to the quantity in the local linear specification for ease of comparison.

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.

We highlight several findings from the four tables. First, although the *de facto* local linear estimator performs competitively in some cases (e.g., Lee DGP with CCT bandwidth selectors in Panel A of Table 1 and Table B.1, supplementary material), it does not deliver the lowest MSE. Looking down column (4) 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 2% (local quadratic with $\hat{h}_{\text{CCT}}$ in Panel A of Table 2) to 72% (local quartic with h_{opt} in Panel B of Table 2).

Second, from column (5) 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 sometimes be substantial. Given the analysis of Calonico, Cattaneo, and Titiunik (2014), it is not surprising that the conventional local linear CI sometimes undercover. The undercoverage is more serious under the Lee DGP: for example, the local linear CI coverage rate is 83% in simulations with n_{actual} and $\hat{h}_{\text{CCT}}$ (Part (a) of Panel A in Table 1). But this undercoverage is alleviated with the use of higher order alternatives, and the local quadratic, cubic and quartic estimators all lead to a coverage rate of at least 90%. The robust local linear CI has better coverage rates as shown in the appendix (Tables B.1 and B.2, supplementary material), and the use of alternative orders may bring further improvement.

Finally, we compare the length of confidence intervals across different choices of p. Table B.2 (supplementary material) shows that the coverage rates are close to the nominal 95% for all robust confidence intervals for the Ludwig-Miller DGP, and all of the polynomial orders greater than one yield confidence intervals that are smaller, and substantially so in many cases. In Tables 1, 2, and B.1, the CI coverage rates of local linear can fall 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 36 specifications that use higher order polynomials in those tables, 33 of them have shorter size-adjusted confidence intervals than local linear.

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.

Table 1. Simulation statistics for the conventional estimator of various polynomial orders: Lee DGP, actual and large sample sizes.

Panel A: Actual sample size ($n = 6558$)

(a): Simulation statistics for the local linear estimator ($p = 1$)

Bandwidth	(1) p	(2) Avg. h	(3) Avg. n	(4) MSE $\times$ 1000	(5) Coverage Rate	(6) Avg. CI Length	(7) Avg. Size-adj. CI length
Theo. Optimal	1	0.099	811	0.450	0.934	0.078	0.084
CCT	1	0.139	1140	0.518	0.831	0.067	0.097

(b): Simulation statistics for other polynomial orders as compared to $p = 1$

Bandwidth	p	Avg. h	Avg. n	Ratio of MSEs	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.022	183	1.583	0.896	1.109	1.245
	2	0.216	1766	0.932	0.942	0.995	0.965
	3	0.407	3321	0.853	0.945	0.958	0.913
	4	0.747	5739	0.764	0.942	0.898	0.867
	$\hat{p}$			0.790	0.941	0.911	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0.001, 0, 0.228, 0.771)						
CCT	0	0.032	266	1.609	0.751	1.081	1.199
	2	0.248	2030	0.893	0.900	1.094	0.893
	3	0.344	2808	0.932	0.941	1.222	0.870
	4	0.390	3180	1.230	0.940	1.421	1.021
	$\hat{p}$			1.025	0.827	1.002	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0.868, 0.105, 0.027, 0)						

Panel B: 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) MSE $\times$ 1000	(5) Coverage Rate	(6) Avg. CI Length	(7) Avg. size-Adj. CI length
Theo. Optimal	1	0.064	4766	0.081	0.927	0.032	0.035
CCT	1	0.080	6020	0.091	0.851	0.029	0.039

(b): Simulation statistics for other polynomial orders as compared to $p = 1$

Bandwidth	p	Avg. h	Avg. n	Ratio of MSEs	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.011	798	2.051	0.893	1.287	1.420
	2	0.157	11782	0.805	0.939	0.933	0.899
	3	0.319	23847	0.681	0.939	0.865	0.835
	4	0.610	44516	0.568	0.939	0.786	0.755
	$\hat{p}$			0.568	0.939	0.786	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, 0, 1)						
CCT	0	0.013	983	1.987	0.820	1.301	1.355
	2	0.181	13539	0.786	0.896	0.977	0.873
	3	0.323	24147	0.660	0.927	0.964	0.784
	4	0.400	29856	0.745	0.946	1.075	0.807
	$\hat{p}$			0.733	0.907	0.958	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0.011, 0.102, 0.825, 0.062)						

Table 2. Simulation statistics for the conventional estimator of various polynomial orders: Ludwig-Miller DGP, actual and large sample sizes.

Panel A: Actual Sample Size (n=3105)

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

Bandwidth	(1) p	(2) Avg. h	(3) Avg. n	(4) MSE × 1000	(5) Coverage Rate	(6) Avg. CI Length	(7) Avg. Size-adj. CI length
Theo. Optimal	1	0.057	222	1.809	0.918	0.150	0.170
CCT	1	0.064	247	1.942	0.884	0.142	0.178

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSEs	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.181	702	0.618	0.930	0.824	0.789
	3	0.406	1566	0.457	0.938	0.729	0.681
	4	0.814	2881	0.354	0.948	0.660	0.589
	$\hat{p}$			0.373	0.944	0.665	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, 0, 0.082, 0.918)						
CCT	2	0.198	770	0.610	0.904	0.831	0.780
	3	0.337	1304	0.542	0.934	0.844	0.719
	4	0.384	1484	0.717	0.937	0.980	0.834
	$\hat{p}$			0.563	0.917	0.823	
	Fraction of time $\hat{p}$=(1,2,3,4): (0, 0.271, 0.726, 0.003)						

Panel B: Large Sample Size (n=30,000)

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

Bandwidth	(1) p	(2) Avg. h	(3) Avg. n	(4) MSE × 1000	(5) Coverage Rate	(6) Avg. CI Length	(7) Avg. Size-adj. CI length
Theo. Optimal	1	0.036	1364	0.286	0.923	0.060	0.067
CCT	1	0.039	1158	0.318	0.901	0.060	0.071

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSEs	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.003	109	4.299	0.886	1.861	2.065
	2	0.131	4904	0.546	0.938	0.773	0.728
	3	0.315	11802	0.379	0.942	0.658	0.612
	4	0.662	23874	0.279	0.947	0.573	0.523
	$\hat{p}$			0.279	0.947	0.573	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, 0, 1)						
CCT	0	0.003	115	4.171	0.869	1.882	2.014
	2	0.141	5301	0.548	0.915	0.772	0.729
	3	0.315	11785	0.389	0.934	0.684	0.607
	4	0.399	14892	0.451	0.945	0.755	0.648
	$\hat{p}$			0.395	0.933	0.684	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0.001, 0.905, 0.094)						

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 each table. 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 8 permutations (2 DGPs $\times$ 2 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 always results in an estimator with improved MSE over local linear. In these cases, the reduction in MSE ranges between 17% and 43% for the Lee DGP and between 46% than 72% for the Ludwig–Miller DGP. 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 maintains the coverage rate if the local linear CI coverage rate is close to 95%, and it improves coverage if the local linear CI undercovers. The procedure also helps to reduce the CI length relative to local linear, especially for the Ludwig–Miller DGP. As emphasized in Remark 4, however, our procedure is not theoretically grounded in inference, and the good performance of the CI here may not generalize to other contexts.

We show additional results in the appendix (Tables B.3 to B.4, supplementary material) for the sample size $n_{\text{small}} = 500$. This is the sample size used in the simulations of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo, and Titiunik (2014). We see from Panel A of Table B.3 (supplementary material) 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 Panel A of Table B.4, $\hat{p}$ does better for the bias-corrected estimator $\hat{\tau}_p^{\text{bc}}$, 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 sizes 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 6000 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 $n \geqslant 60,000$, 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 Panel B of Tables B.3 and B.4 (supplementary material): the modal $\hat{p}$ always coincides with the MSE minimizing polynomial order, and relative to local linear, our procedure leads to improved MSE.

To summarize, we have implemented simulations under two DGPs (Lee and Ludwig–Miller), two bandwidth choices (h_{opt} and h_{CCT}), 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). 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. We rely on Lemma 2 and Theorem A.2 of Calonico, Cattaneo, and Titiunik (2014) to estimate the AMSE of a fuzzy RD estimator by first linearizing it, and we implement the calculation in the Stata package `rdmse`.

The second extension is the regression kink design (Nielsen, Sørensen, and Taber 2010; Card et al. 2015a), which our Stata implementation also accommodates. For RKD, Calonico, Cattaneo, and Titiunik (2014) 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 specify DGPs based on the bottom- and top-kink samples of the application in Card et al. (2015b) (see Appendix B.2 for details, supplementary material). These DGPs again allow us to compute $\text{AMSE}_{\hat{\tau}_p}$ as a function of sample size for different p. As shown in Panel (C) of Figure 1, the AMSE of the local quadratic fuzzy estimator is asymptotically smaller. However, it takes about 86 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. Since these threshold sample sizes are far larger than the 270,000 observations in both the bottom- and top-kink samples, they give reason to prefer the local linear RK estimator.

5. Conclusion

This article 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) 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 typically seen in RD applications.

As a concluding remark, we view the proposed polynomial selection procedure as a complement—not a substitute—to 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).

Supplemental Material

The supplemental materials contain the Appendix and replication programs.

Acknowledgments

We thank the editor, Jianqing Fan, an anonymous associate editor, three anonymous referees, Matias Cattaneo, Gordon Dahl, Guido Imbens, Pat Kline, Pauline Leung, Doug Miller, Jack Porter, Yi Shen, Yan Song, Stefan Wager, Vivian Wong, and participants at various seminars and conferences for helpful comments. Camilla Adams, Amanda Eng, Sarah Frick, Katie Guyot, Samsun Knight, Suejin Lee, Carl Lieberman, Bailey Palmer, and Amy Tarczynski provided outstanding research assistance.

References

Arai, Y., and Ichimura, H. (2018), "Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator," *Quantitative Economics*, 9, 441–482. [4]

Branson, Z., Rischard, M., Bornn, L., and Miratrix, L. W. (2019), "A Nonparametric Bayesian Methodology for Regression Discontinuity Designs," *Journal of Statistical Planning and Inference*, 202, 14–30. [4]

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020), "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs," *Econometrics Journal*, 23, 192–210. [4]

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82, 2295–2326. [1,2,3,4,5,8]

——— (2015), "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association*, 110, 1753–1769. [4]

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017), "`rdrobust`: Software for Regression Discontinuity Designs," *Stata Journal*, 17, 372–404. [4]

——— (2019), "Regression Discontinuity Designs Using Covariates," *Review of Economics and Statistics*, 101, 442–451. [4]

Card, D., Lee, D. S., Pei, Z., and Weber, A. (2015a), "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 83, 2453–2483. [8]

——— (2015b), "Inference on Causal Effects in a Generalized Regression Kink Design," Upjohn Institute Working Paper 15-218. [8]

——— (2017), "Regression Kink Design: Theory and Practice," in *Regression Discontinuity Designs: Theory and Applications*. Vol. 38 of *Advances in Econometrics*, ed. M. D. Cattaneo and J. C. Escanciano, Bingley: Emerald Publishing Ltd. [9]

Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017), "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality," *Journal of Policy Analysis and Management*, 36, 643–681. [4]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Boca Raton: Chapman and Hall. [1,4]

Gelman, A., and Imbens, G. W. (2019), "Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 37, 447–456. [1,8]

Geneletti, S., O'Keeffe, A. G., Sharples, L. D., Richardson, S., Baio, G. (2015), "Bayesian Regression Discontinuity Designs: Incorporating Clinical Knowledge in the Causal Analysis of Primary Care Data," *Statistics in Medicine*, 34, 2334–2352. [4]

Hahn, J., Todd, P., and Van der Klaauw, W. (2001), "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69, 201–209. [1,2,8]

Hall, P. G., and Racine, J. S. (2015), "Infinite Order Cross-Validated Local Polynomial Regression," *Journal of Econometrics*, 185, 510–525. [1]

Imbens, G. W., and Kalyanaraman, K. (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933 – 959. [1,2,3,4,8]

Lee, D. S. (2008), "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697. [1,2,3,4,5]

Lee, D. S., and Thomas, L. (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355. [1,4]

Ludwig, J., and Miller, Douglas, L. (2007), "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics*, 122, 159–208. [2,3,4,5]

Nielsen, H. S., Sørensen, T., and Taber, C. R. (2010), "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform," *American Economic Journal: Economic Policy*, 2, 185–215. [8]

Otsu, T., Xu, K.-L., and Matsushita, Y. (2015), "Empirical Likelihood for Regression Discontinuity Design," *Journal of Econometrics*, 186, 94–112. [4]

Pei, Z., Lee, D. S., Card, D., and Weber, A. (2020), "Local Polynomial Order in Regression Discontinuity Designs," NBER Working Paper 27424. [4]

Porter, J. (2003), "Estimation in the Regression Discontinuity Model," mimeographed. [1]

Zhang, J., and Boos, D. D. (1994), "Adjusted Power Estimates in Monte Carlo Experiments," *Communications in Statistics – Simulation and Computation*, 23, 165–173. [5]