WORKING PAPER #622
PRINCETON UNIVERSITY
INDUSTRIAL RELATIONS SECTION
http://arks.princeton.edu/ark:/88435/dsp01v118rh27h

Local Polynomial Order in Regression Discontinuity Designs[1]

Zhuan Pei[2]
Cornell University and IZA

David Card[4]
UC Berkeley, NBER and IZA

David S. Lee[3]
Princeton University and NBER

Andrea Weber[5]
Central European University and IZA

August 6, 2018

Abstract

It has become standard practice to use local linear regressions in regression discontinuity designs. This paper highlights that the same theoretical arguments used to justify local linear regression suggest that alternative local polynomials could be preferred. We show in simulations that the local linear estimator is often dominated by alternative polynomial specifications. Additionally, we provide guidance on the selection of the polynomial order. The Monte Carlo evidence shows that the order-selection procedure (which is also readily adapted to fuzzy regression discontinuity and regression kink designs) performs well, particularly with large sample sizes typically found in empirical applications.

Keywords: Regression Discontinuity Design; Regression Kink Design; Local Polynomial Estimation; Polynomial Order

---

[1]We thank Gordon Dahl, Guido Imbens, Pat Kline, Pauline Leung, Doug Miller, Jack Porter, Yan Song, Stefan Wager, and seminar participants at Brandeis, Cornell, and George Washington, as well as the CIRANO-CIREQ, Econometric Society World Congress, Jinan IESR, and SOLE conferences for helpful comments. Michael Daly, Amanda Eng, Samsun Knight, Suejin Lee, Carl Lieberman, and Bailey Palmer provided outstanding research assistance.

[2]Department of Policy Analysis and Management, Cornell University, Ithaca, NY 14853, U.S.A.; zhuan.pei@cornell.edu

[3]Department of Economics, Princeton University, Louis A. Simpson International Building, Princeton, NJ 08544, U.S.A.; davidlee@princeton.edu

[4]Department of Economics, UC Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720, U.S.A.; card@berkeley.edu

[5]Department of Economics, Central European University, Nador utca 9, 1051 Budapest, Hungary; WeberA@ceu.edu

1 Introduction

Regression discontinuity designs (RD designs or RDD) are widely used in empirical social science research in recent years. Among the important reasons for its appeal are that the research design permits clear and transparent identification of causal parameters of interest, and that 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 that 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, the violation of which could lead to misleading inferences.

Typical practice in applied research is to employ a local regression estimator, with local linear being the most common. We surveyed leading economics journals between 1999 and 2017, and found that of the 110 studies that employed RDD, 76 use a local polynomial regression as their main specification, compared to 26 studies that report a global regression estimator as their main specification.[1] Out of the 76 studies that use local regression estimators, local linear is the modal choice and applied as the main specification in 45 studies (59%). Empirically, the reliance on local linear appears to be accelerating: out of the 51 studies published since 2011 with a local main specification, 36 apply local linear (71%). Meanwhile, recent econometric studies on RD estimation, whether focusing on bandwidth selection (e.g. Imbens and Kalyanaraman, 2012) or improving inference (e.g. Calonico, Cattaneo and Titiunik, 2014*b* and Armstrong and Kolesár, forthcoming), have taken the polynomial order as given, often treating local linear as the default.[2]

Despite the theoretical and practical focus on local linear, it has long been recognized that the choice of polynomial order can in principle be as consequential as the choice of bandwidth (e.g., see discussion in Fan and Gijbels, 1996). More recently, Hall and Racine (2015) call into question the practice of choosing the polynomial order in an ad hoc fashion, and suggest instead a cross-validation method to choose the poly-

[1]We include *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. We cannot classify whether the main specification is local or global for eight of the 110 studies; see Appendix Table A.1 for detailed tabulation.

[2]Another recent paper by Imbens and Wager (forthcoming) proposes an alternative to the local polynomial regression approach. Specifically, it finds the linear RD estimator through numerical convex optimization that minimizes the worst-case-scenario mean-squared-error over the class of functions with a known global bound on the second derivative. A more detailed discussion is in Appendix section C.

nomial order jointly with the bandwidth.[3] With this as background, the relevant questions for researchers interested in applying RD are: 1) To what extent does the theory of nonparametric estimation give us reasons to expect local linear to outperform other polynomial orders? 2) To what extent does local linear, in practice, perform better (or worse) than other orders? and 3) What data-based procedure or diagnostic could help guide the order choice?

This paper addresses these questions. First, we revisit the theoretical arguments used to justify a particular polynomial order. The seminal work of Hahn, Todd and Van der Klaauw (2001) correctly states that the local linear estimator has an asymptotically smaller bias than the kernel regression estimator. But as Porter (2003) points out, by the same argument, any higher-order polynomial will have an even smaller asymptotic bias than the local linear. Using the asymptotic approximation for mean-squared-error used in bandwidth calculations, we show that the linear specification may perform better or worse than alternative polynomials, depending on the sample size. Therefore, the theoretical frameworks used in the past do not uniformly point to the preference of any specific polynomial order, and if anything, remind us that the polynomial order is as much of a choice as the bandwidth.

Second, because existing methods are primarily based on asymptotic approximations, we explore the finite sample performance of local linear estimators against alternative orders of polynomials – in each case using various notions of optimal bandwidths – through Monte Carlo simulations based on two well-known examples (Lee, 2008 and Ludwig and Miller, 2007). In fact, we use the exact same data generating processes (DGP's) employed in the simulations of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*). We find that in most of our simulations, higher-order alternatives outperform $p = 1$ for the two specific DGP's we consider. In particular, they tend to outperform local linear in terms of their mean-squared-error (henceforth MSE), coverage rate of the 95% confidence interval (CI), and the size-adjusted CI length, especially when the sample size is large.

Finally, while we consider it advisable to explore the sensitivity of results to different polynomial orders, we additionally propose a polynomial order selection procedure for RD designs that is in the spirit of a suggestion by Fan and Gijbels (1996), and is a natural extension of the approach based on asymptotic MSE (henceforth AMSE) used in recent developments on bandwidth choices (e.g., Imbens and Kalyanaraman, 2012 and Calonico, Cattaneo and Titiunik, 2014*b*). In our Monte Carlo simulations, we find that the esti-

[3] Hall and Racine (2015) study nonparametric estimation at an interior point and propose a leave-one-out cross-validation procedure. 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 advice for applied researchers.

mator chosen by the proposed order selector improves upon local linear in terms of MSE, CI coverage rate, and CI length. The improvements are substantial when the sample size is large.

The remainder of the paper is organized as follows. Section 2 revisits the theoretical considerations behind the use of local polynomial estimators for RD. Section 3 presents simulation results. In section 4, we show that the AMSE-based methods for choosing the polynomial order can be easily extended to the fuzzy RD design and the regression kink design (RKD). Section 5 concludes.

2 Local Polynomial Order in RD Designs: Theoretical Considerations

Here, 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:

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

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

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. Lemma 1 of Calonico, Cattaneo and Titiunik (2014*b*) proves the AMSE of the p-th order local polynomial estimator $\hat{\tau}_p$ as a function of bandwidth as

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

under standard regularity conditions, 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 estimators are $h^2\text{B}_1$ and $h\text{B}_0$ and are of orders $O(h^2)$ and $O(h)$ respectively. However, by the very same logic, the asymptotic bias is of order $O(h^3)$ for the local quadratic estimator ($p=2$), $O(h^4)$ for local cubic ($p=3$), and $O(h^{p+1})$ for the p-th order estimator, $\hat{\tau}_p$, under standard regularity conditions. Therefore, if researchers were exclusively focused on the maximal shrinkage rate of the asymptotic bias, this argues for choosing p to be as large as possible. Hahn, Todd and Van der Klaauw (2001) recommends $p=1$, implicitly recognizing that factors beyond the 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 h that leads to the smallest value of AMSE, considering as optimal the bandwidth defined by $h_{opt}(p) \equiv \arg\min_h \text{AMSE}_{\hat{\tau}_p}(h)$. 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 higher-order polynomial estimators (henceforth CCT bandwidth). It follows from Lemma 1 of Calonico, Cattaneo and Titiunik (2014*b*) that $h_{opt}(p)$ for $\hat{\tau}_p$ is of order $O(n^{-\frac{1}{2p+3}})$.

Here we point out the natural implication of this. By evaluating expression (3) at h_{opt}, $\text{AMSE}_{\hat{\tau}_p}(h_{opt}(p))$ is equal to $\text{C}_p \cdot n^{-\frac{2p+2}{2p+3}}$, where C_p is a function of the constants B_p and V_p. Therefore, as the sample size n increases, $\text{AMSE}_{\hat{\tau}_p}(h_{opt}(p))$ 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, 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's through the constant C_p.

This point is concretely illustrated in Figure 1, using the two DGP's we rely on for subsequent simulations, which are taken from Lee (2008) and Ludwig and Miller (2007) and described in greater detail in Appendix section C. Since we know the parameters of the underlying DGP's, 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 $\mathrm{AMSE}_{\hat{\tau}_p}$ 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 DGP's respectively (see Appendix section B.1 for details).

In Panel (A), we see that at small sample sizes, $\mathrm{AMSE}_{\hat{\tau}_1}$ is marginally below $\mathrm{AMSE}_{\hat{\tau}_2}$, but is larger at sample sizes over $n = 1167$. Therefore, for the actual number of observations in the analysis sample of Lee (2008), $n_{actual} = 6558$, 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 $\mathrm{AMSE}_{\hat{\tau}_2}$ dominates $\mathrm{AMSE}_{\hat{\tau}_1}$ for all n under 7000. At the actual number of observations in Ludwig and Miller (2007), $n_{actual} = 3105$, 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 the cross for the local linear estimator and 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):

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

where the quantities B_p and V_p in (3) are replaced by estimators $\hat{\mathrm{B}}_p$ and $\hat{\mathrm{V}}_p$, and the optimal feasible bandwidth is defined as $\hat{h}(p) \equiv \arg\min_h \widehat{\mathrm{AMSE}}_{\hat{\tau}_p}(h)$. The two studies differ in how they arrive at the estimates of B_1 and V_1; Calonico, Cattaneo and Titiunik (2014*b*) also generalizes 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 (see discussions in Card et al., 2015*a* and Card et al., 2017). In our Monte Carlo simulations below,

[4] The same exercise can be performed for other values of p. See Tables 3-4, A.3-A.4 in Card et al. (2014) for more threshold sample sizes.

we experiment with the CCT bandwidth both with and without regularization. We additionally examine the performance of the *bias-corrected estimator* of Calonico, Cattaneo and Titiunik (2014*b*), denoted by $\hat{\tau}_p^{bc}$.[5]

Here, we extend the logic that justifies the optimal bandwidth by noting that we can further minimize the estimated AMSE by searching over different values of p. That is, we can define

$$\hat{p} \equiv \underset{p}{\arg\min} \widehat{\text{AMSE}}_{\hat{\tau}_p} \left(\hat{h}(p) \right).$$

No new quantities need be computed beyond the estimators $\hat{\text{B}}_p$ and $\hat{\text{V}}_p$ and the optimal $\hat{h}(p)$, which must be calculated when implementing, for example, the CCT bandwidth.

In summary, our observation is that if 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. The recommendation here is to try different values of p and choose the p with the lowest $\widehat{\text{AMSE}}_{\hat{\tau}_p}$. The exact expressions needed from Calonico, Cattaneo and Titiunik (2014*b*) for implementation and the description of a Stata command `rdmse` to perform the calculations are provided in Appendix section B.2. Although this simple order selection approach was suggested by Fan and Gijbels (1996) for general local polynomial regression, to the best of our knowledge, it has not yet been applied to RD designs, and its performance as an order selector for RD has not been assessed.[6] We report on the finite sample performance of this polynomial order selector below.

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. Therefore, 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. We focus on how the *de facto* standard in the literature of local linear compares to alternative orders, as well as the performance of the order selection procedure proposed above.

[5] This estimator is constructed by first estimating the asymptotic bias of the local RD estimator $\hat{\tau}_p$ with a local regression of order $p+1$, then subtracting it from $\hat{\tau}_p$. Additionally, Calonico, Cattaneo and Titiunik (2014*b*) propose to calculate the *robust confidence interval* by centering it at $\hat{\tau}_p^{bc}$ and accounting for the variance in estimating the bias.

[6] The procedure in Fan and Gijbels (1996) chooses the polynomial order that minimizes the estimated AMSE among alternative local polynomial estimators of the conditional expectation function evaluated at an interior point. Our adaptation of this procedure mirrors how Imbens and Kalyanaraman (2012) modify standard nonparametric bandwidth selection for RD designs.

We utilize DGP's from two well-known empirical examples, Lee (2008) and Ludwig and Miller (2007), and the specifications of these DGP's follow *exactly* those in Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014*b*); see Appendix section A.1 for details. Our simulations draw 10000 repeated samples from the two DGP's. Below, we present results using a uniform kernel; results from the triangular kernel are available in our previous working paper Card et al. (2014), and the qualitative conclusions are the same.[7]

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 DGP's respectively, while Tables 3 and 4 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*). 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: $n_{actual} = 6558$ for Lee (2008) and $n_{actual} = 3105$ for Ludwig and Miller (2007). We set the large sample size to $n_{large} = 60000$ for the Lee DGP and $n_{large} = 30000$ 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.[8]

In part (a) of each panel, 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, ii) the default CCT bandwidth selector from Calonico, Cattaneo and Titiunik (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), MSE in column (4), coverage rate of the 95% CI in column (5), the average CI length in columns (6), and the average *size-adjusted* CI length in columns (7).[9] In part (b) of each panel, we present the same statistics for different polynomial orders; in columns (4)-(7),

[7]Cheng, Fan and Marron (1997) have shown that the triangular kernel $K(u) = (1-u) \cdot 1_{u \in [-1,1]}$ is boundary optimal (and hence optimal for RD designs). In practice, the uniform kernel $K(u) = \frac{1}{2} \cdot 1_{u \in [-1,1]}$ is popular among practitioners for its convenience and for the ease of reconciling estimates with graphical evidence.

[8]We keep fixed the distributions of X and ε as well as $E[Y|X=x]$, while we vary the sample size. The simulation exercise does not speak to the situation where the researcher collects additional years of data in which the support of X changes.

[9]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% CI's 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% CI's. 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 CI's in column (7).

we express the quantities as a ratio to the quantity in the local linear specification.[10]

3.1 Local Linear versus Alternative Polynomials

We now highlight the following findings from Tables 1 to 4. First, the local linear estimator never delivers the lowest MSE in any of the tables. Looking down column (4) in part (b) of every panel, there is at least one alternative estimator for which the MSE ratio is less than one. The reduction in MSE from using a different polynomial order ranges from 5.5% (local quadratic with $\hat{h}_{CCT}$ in Panel A of Table 3) to 73% (local quartic with h_{opt} in Panel B of Table 2).

Second, from column (5) in all tables, the cubic and quartic estimators always improve upon the local linear in terms of its 95% CI coverage rate. In many instances, the coverage rate of the local linear CI is close to the nominal level, in which case the improvement by cubic and quartic 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 (Panel A(a) of Table 1). 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 Panel A(a) of Table 3. By comparison, the local cubic and quartic robust CI's 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. Table 4 shows that the coverage rates are close to the nominal 95% for all robust confidence intervals, 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 3, 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, only

[10] Since the k-th 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.

two of them have longer size-adjusted confidence intervals than local linear.

Our final observation about the performance of higher-order polynomials is that the optimal bandwidths 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 both Panels A and B of Table 1, the order with the lowest MSE is $p = 4$. With $p = 4$, the (theoretically) optimal bandwidth implies using an average of 5227 of the 6558 observations under n_{actual}, or 39983 of the 60000 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.

So an important lesson is that while nonparametric estimation and inference imagines only using data in a “close neighborhood” of the threshold *asymptotically*, optimal bandwidth procedures may yield wide bandwidths that amount to using a substantial fraction of the data in practice, particularly for a higher-order local polynomial estimator. There is a numerical equivalence here: the high-order local polynomial estimator (with uniform kernel) is numerically identical to trimming the tails of X and running “global polynomials” on the trimmed data. One can thus think of the bandwidth selector as providing a theoretical justification and guidance for the degree of trimming. Equivalently, a global polynomial that uses the entire range of X yields estimates that are numerically equivalent to a local high-order polynomial that uses a bandwidth covering the entire range of X.

Since our Monte Carlo results suggest that higher-order polynomials with somewhat wide bandwidths perform the best – and are numerically equivalent to some trimming of the distribution of X – we consider our findings in light of the recent study by Gelman and Imbens (forthcoming), who raise concerns about using high-order global polynomials to estimate the RD treatment effect. One concern in particular is that global polynomial estimators may assign too much weight (henceforth GI weights) to observations far away from the RD cutoff.

This does not appear to be the case for the applications we study. Using the *actual* Lee and Ludwig-Miller data, Appendix Figures A.3-A.4 plot the GI weights for the left and right intercept estimators that make up $\hat{\tau}_p$ for p between zero and five. For brevity, we only display the weights under the uniform kernel with bandwidth $\hat{h}_{CCT}$, and we see similar patterns in both figures. As desired, observations far away from the cutoff receive little weight under high-order polynomials compared to those close to the cutoff. Our previous version, Card et al. (2014), displays results with alternative kernel and bandwidth choices, and the patterns are similar.[11]

3.2 Performance of the Polynomial Order Selector

We have thus far provided both theoretical arguments and Monte Carlo evidence that point toward a broader 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 that p.

We now turn to the performance of the proposed order selector – choosing the $\widehat{\text{AMSE}}$-minimizing p. Specifically, for each Monte Carlo draw, we compute the RD estimator for multiple values of p and their corresponding $\widehat{\text{AMSE}}$. 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.

The results are found in the rows labeled "$\hat{p}$" below the quartic in Tables 1 to 4. Overall, they show that using the order $\hat{p}$ leads to comparable, or in many cases, considerably lower MSE than always choosing $p = 1$. For the Lee DGP with n_{actual}, we see from Panel A of Tables 1 and 3 that the ratio of $\text{MSE}_{\hat{p},h_{opt}}$ over $\text{MSE}_{1,h_{opt}}$ is 0.762 for the conventional estimator and 0.786 for the bias-corrected estimator. Out of the 12 permutations (3 bandwidth selectors times 2 estimators times 2 sample sizes), in three cases the MSE is slightly larger for the order selector $\hat{p}$ than that for local linear. In all other cases, the MSE from using $\hat{p}$ is lower, and in most cases significantly lower. The order selector $\hat{p}$ performs better with the larger sample sizes (Panel B of each table). For the Ludwig-Miller DGP, $\hat{p}$-selected estimator outperforms local linear in all simulations by selecting higher polynomial orders as shown in Tables 2 and 4. The reduction in MSE ranges from 33% to 73%.

We see qualitatively similar results for the $\hat{p}$-selected estimator in terms of its CI coverage rate vis-à-vis local linear. The $\hat{p}$-selected estimator has comparable or better coverage rates under the Lugwig-Miller DGP. Its performance under the Lee DGP once again depends on the sample size: It tends to have somewhat lower coverage rates than local linear when the sample size is n_{actual}, but the coverage rates are closer to the nominal level when the sample size is n_{large}. Overall, the improvement or reduction on the CI coverage rate tends to be moderate – less than 6% for the conventional CI and less than 2% for the CCT robust CI.

[11] We present additional simulation results using procedures from Armstrong and Kolesár (forthcoming) and Imbens and Wager (forthcoming). See Appendix section C for details.

The tables also show, in the three out of four scenarios where the $\hat{p}$-selected estimator enjoys a (weak) size advantage – 1) Lee DGP with n_{large}, 2) Ludwig-Miller DGP with n_{actual}, and 3) Ludwig-Miller DGP with n_{large} – the CI length is shorter and considerably so in many cases. The reduction in the CI length ranges from 5% to 25% in scenario 1), to between 18% and 35% in scenario 2), to above 30% in scenario 3). In short, the $\hat{p}$-selected estimator appears to have *both* a size advantage and a power advantage in these three scenarios.

We show additional results in Appendix Tables A.3 and A.4 for the sample size $n_{small} = 500$. This is the sample size used in the simulations of Imbens and Kalyanaraman (2012), Calonico, Cattaneo and Titiunik (2014*b*), and Armstrong and Kolesár (forthcoming). We see from Panel A of Table A.3 that $\hat{\tau}_1$ has the lowest MSE under the Lee DGP, and that using $\hat{p}$ leads to at least a 15% increase in MSE and lower CI coverage rates for all three bandwidth choices. As shown in Panel A of Table A.4, $\hat{p}$ does better for the bias-corrected estimator under the Lee DGP, leading to comparable or lower MSE's, although the CI coverage rates are still lower than $p = 1$. This 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 6000 observations, and the median sample size is 21561. A sample size of 60000, which is 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 60000, much more so than those with just 500 observations. But even with 500 observations, $\hat{p}$ unambiguously outperforms local linear under the Ludwig-Miller DGP. As shown in Panel B of Tables A.3 and A.4, using $\hat{p}$ improves upon its local linear counterpart across all three performance measures: MSE, CI coverage rate, and CI length.

In summary, we implemented simulations under two DGP's (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 find that local linear is dominated by other polynomial order choices in most cases. In these simulation exercises, using $\hat{p}$ as the polynomial order tends to improve upon always choosing local linear, especially when the sample size is large.

4 Extensions: Fuzzy RD and RKD

In this section, we briefly discuss how (A)MSE-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 existing RD literature, $p = 1$ is still the default choice in the fuzzy RD. But by a similar argument as above, local linear is not necessarily the best estimator in all applications. 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.

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*).[12] For RKD, Calonico, Cattaneo and Titiunik (2014*b*) and Gelman and Imbens (forthcoming) recommend using $p = 2$ by extending the Hahn, Todd and Van der Klaauw (2001) argument, but as we similarly discussed for the case for RD, the AMSE for $p = 2$ (while using the corresponding AMSE-minimizing h) may or may not be lower, depending on the sample size in any particular empirical application.

To illustrate this once again, but in the case of 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 section A.2 for details). The specification of these approximating DGP's again allows us to compute $\text{AMSE}_{\hat{\tau}_p}$ as a function of the sample size for different polynomial orders. 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 270000 for both the top- and bottom-kink samples. In this case, these calculations give reason to prefer the local linear fuzzy estimator.[13]

[12] Our software `rdmse` implements the AMSE estimation for all conventional/bias-corrected sharp/fuzzy RD/RK estimators.

[13] We make this same point in Card et al. (2017) through estimated AMSE's and Monte Carlo simulation results that compare alternative estimators.

5 Conclusion

The local linear estimator has become the standard specification in the regression discontinuity literature. In this paper, we re-examine the reasoning in favor of $p = 1$ and show that 1) the theoretical arguments that suggest the dominance of $p = 1$ over $p = 0$ also suggest the dominance of any $p > 1$ over $p = 1$, and 2) the dominance is dependent on the shape of the underlying data generating processes and sample sizes.

We concretely illustrate these points with simulations based on two well-known RD examples. In these exercises, $p = 1$ tends to be dominated by alternative polynomial specifications across bandwidth selectors, estimators (conventional and bias-corrected), and across sample-sizes. Our proposed order selector, which is simply a logical and complementary extension of the theoretical justification behind widely-used bandwidth selectors, performs reasonably well. It does particularly well in large sample sizes, comparable to the sample sizes we see employed in empirical applications today.

References

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

Armstrong, Timoth B., and Michal Kolesár. forthcoming. "Optimal inference in a class of regression models." *Econometrica*.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2016. "Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs." mimeographed.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2018. "Coverage Error Optimal Confidence Intervals." mimeographed.

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.

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.

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

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.

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

Gelman, Andrew, and Guido Imbens. forthcoming. "Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*.

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. forthcoming. "Optimized Regression Discontinuity Designs." *Review of Economics and Statistics*.

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.

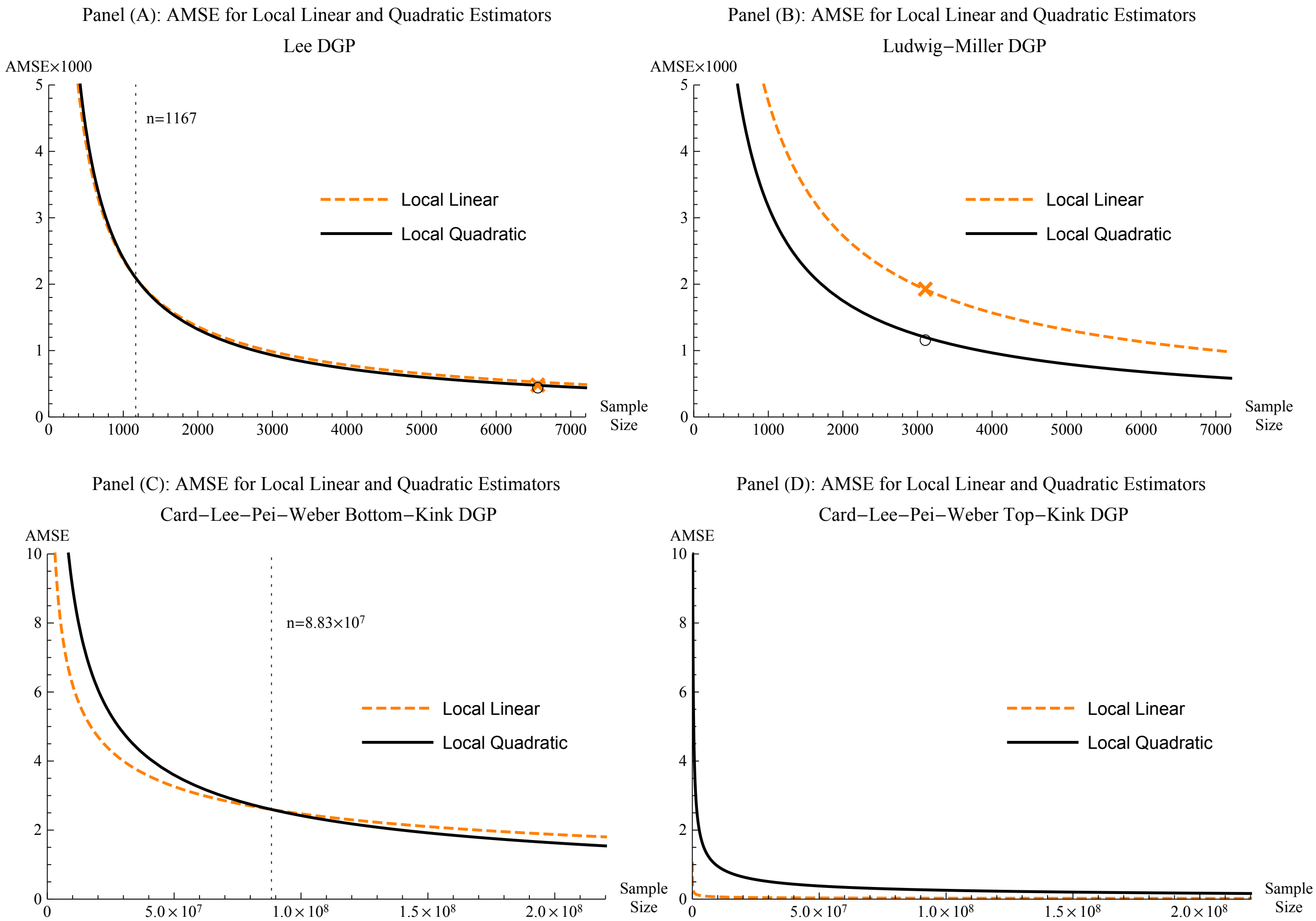
Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics*, 122(1): 159–208.

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.

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

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 MSE's of the local linear (cross) and quadratic (circle) estimators with the theoretical optimal bandwidth. These MSE's are taken from Tables 1 and 2. At the actual sample size of the two studies, the theoretical AMSE's appear to be quite close to the corresponding MSE's.

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=6,558)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.078	637	0.481	0.934	0.081	0.086
CCT	1	0.111	908	0.571	0.822	0.068	0.100
CCT w/o reg.	1	0.167	1335	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.015	127	1.610	0.960	1.123	1.271
	2	0.180	1477	0.917	1.008	0.987	0.964
	3	0.353	2888	0.837	1.008	0.945	0.918
	4	0.663	5227	0.738	1.009	0.879	0.856
	$\hat{p}$	0.603	4771	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	1.623	0.895	1.085	1.198
	2	0.207	1699	0.851	1.096	1.093	0.892
	3	0.298	2443	0.861	1.146	1.214	0.862
	4	0.348	2843	1.108	1.151	1.407	0.983
	$\hat{p}$	0.141	1159	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	1.302	1.037	1.187	0.893
	2	0.274	2221	0.855	1.187	1.116	0.852
	3	0.425	3427	0.856	1.287	1.188	0.778
	4	0.539	4243	0.701	1.389	1.335	0.660
	$\hat{p}$	0.263	2120	1.081	0.967	0.996	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.004, .646, .173, .123, .054)						

Panel B: Large Sample Size (n=60,000)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.050	3746	0.086	0.927	0.033	0.036
CCT	1	0.064	4778	0.098	0.847	0.030	0.040
CCT w/o reg.	1	0.069	5157	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.007	554	2.115	0.960	1.301	1.438
	2	0.131	9852	0.789	1.013	0.926	0.889
	3	0.276	20702	0.659	1.013	0.855	0.819
	4	0.542	39983	0.546	1.013	0.772	0.740
	$\hat{p}$	0.542	39983	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	2.016	0.961	1.311	1.367
	2	0.151	11344	0.759	1.060	0.974	0.869
	3	0.280	20988	0.626	1.097	0.956	0.766
	4	0.357	26656	0.690	1.118	1.063	0.792
	$\hat{p}$	0.266	19904	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	1.853	0.997	1.345	1.285
	2	0.167	12501	0.813	1.050	0.964	0.889
	3	0.321	23959	0.756	1.085	0.931	0.800
	4	0.519	37787	0.725	1.093	0.935	0.833
	$\hat{p}$	0.436	32182	0.737	1.055	0.903	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .02, .075, .349, .556)						

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=3,105)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.045	175	1.926	0.922	0.155	0.174
CCT	1	0.050	195	2.055	0.886	0.147	0.182
CCT w/o reg.	1	0.051	198	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.151	587	0.601	1.015	0.817	0.778
	3	0.352	1362	0.442	1.021	0.718	0.666
	4	0.723	2653	0.340	1.028	0.642	0.583
	p̂	0.692	2546	0.351	1.026	0.648	
	Fraction of time p̂=(1,2,3,4): (0, 0, .082, .918)						
CCT	2	0.167	646	0.608	1.027	0.823	0.773
	3	0.293	1135	0.519	1.060	0.832	0.701
	4	0.341	1321	0.676	1.067	0.965	0.793
	p̂	0.269	1041	0.550	1.039	0.811	
	Fraction of time p̂=(1,2,3,4): (0, .291, .701, .009)						
CCT w/o reg.	2	0.173	670	0.614	1.023	0.816	0.781
	3	0.390	1495	0.537	1.015	0.739	0.732
	4	0.536	1996	0.518	1.057	0.806	0.697
	p̂	0.446	1687	0.493	1.027	0.733	
	Fraction of time p̂=(1,2,3,4): (0, .062, .689, .249)						

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

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.029	1072	0.304	0.924	0.062	0.069
CCT	1	0.031	1158	0.318	0.901	0.060	0.071
CCT w/o reg.	1	0.031	1168	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.002	76	4.395	0.964	1.889	2.077
	2	0.109	4102	0.541	1.015	0.767	0.724
	3	0.273	10245	0.369	1.021	0.649	0.602
	4	0.588	21534	0.271	1.023	0.562	0.518
	p̂	0.588	21534	0.271	1.023	0.562	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, 0, 0, 1)						
CCT	0	0.002	80	4.279	0.971	1.912	2.054
	2	0.119	4450	0.545	1.015	0.766	0.731
	3	0.274	10257	0.378	1.037	0.675	0.607
	4	0.356	13288	0.430	1.052	0.744	0.642
	p̂	0.292	10914	0.383	1.035	0.675	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, 0, .881, .119)						
CCT w/o reg.	0	0.002	80	4.266	0.972	1.918	2.066
	2	0.120	4515	0.543	1.014	0.764	0.734
	3	0.296	11070	0.392	1.020	0.654	0.621
	4	0.534	19349	0.349	1.029	0.628	0.584
	p̂	0.474	17367	0.348	1.022	0.619	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, 0, .345, .655)						

Table 3: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Lee DGP, Actual and Large Sample Sizes

Panel A: Actual Sample Size (n=6,558)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.078	637	0.507	0.950	0.088	0.088
CCT	1	0.111	906	0.514	0.900	0.077	0.092
CCT w/o reg.	1	0.165	1332	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.015	127	1.305	0.997	1.136	1.148
	2	0.180	1476	0.898	0.999	0.952	0.954
	3	0.353	2887	0.786	1.003	0.892	0.882
	4	0.663	5227	0.732	1.002	0.865	0.856
	$\hat{p}$	0.488	3900	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	1.041	1.006	1.031	1.009
	2	0.207	1698	0.945	1.041	1.063	0.940
	3	0.299	2443	1.067	1.053	1.166	0.990
	4	0.347	2835	1.406	1.054	1.335	1.132
	$\hat{p}$	0.106	865	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	0.881	1.033	0.995	0.890
	2	0.274	2224	0.960	1.076	1.070	0.888
	3	0.423	3415	0.940	1.112	1.201	0.879
	4	0.540	4247	1.339	1.122	1.532	1.070
	$\hat{p}$	0.170	1385	0.948	0.985	0.935	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.323, .427, .134, .109, .007)						

Panel B: Large Sample Size (n=60,000)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.050	3746	0.084	0.945	0.035	0.036
CCT	1	0.064	4778	0.078	0.928	0.032	0.035
CCT w/o reg.	1	0.069	5157	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.007	554	1.665	1.000	1.297	1.293
	2	0.131	9852	0.800	1.003	0.905	0.890
	3	0.276	20702	0.658	1.004	0.821	0.806
	4	0.542	39983	0.539	1.004	0.745	0.728
	$\hat{p}$	0.527	38874	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	1.513	1.003	1.257	1.240
	2	0.151	11344	0.861	1.012	0.965	0.928
	3	0.280	20988	0.794	1.020	0.940	0.875
	4	0.357	26656	0.963	1.025	1.043	0.956
	$\hat{p}$	0.261	19538	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	1.480	1.011	1.273	1.223
	2	0.167	12501	0.891	1.010	0.958	0.935
	3	0.321	23959	0.813	1.020	0.928	0.872
	4	0.519	37787	1.116	1.028	1.142	1.040
	$\hat{p}$	0.356	26489	0.755	1.014	0.894	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .004, .09, .666, .241)						

Table 4: Simulation Statistics for the Bias-corrected Estimator of Various Polynomial Orders: Ludwig-Miller DGP, Actual and Large Sample Sizes

Panel A: Actual Sample Size (n=3,105)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.045	175	1.715	0.941	0.160	0.168
CCT	1	0.050	195	1.632	0.935	0.154	0.162
CCT w/o reg.	1	0.051	199	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.151	587	0.654	1.001	0.813	0.809
	3	0.352	1361	0.492	1.006	0.709	0.689
	4	0.723	2653	0.454	1.011	0.689	0.655
	p̂	0.472	1778	0.490	1.004	0.701	
	Fraction of time p̂=(1,2,3,4): (0, 0, .677, .323)						
CCT	2	0.167	647	0.682	1.004	0.830	0.824
	3	0.292	1133	0.736	1.010	0.863	0.839
	4	0.342	1322	0.965	1.008	0.989	0.969
	p̂	0.216	839	0.675	1.005	0.821	
	Fraction of time p̂=(1,2,3,4): (0, .693, .306, .001)						
CCT w/o reg.	2	0.173	670	0.669	1.001	0.822	0.819
	3	0.390	1496	0.935	1.004	0.883	0.860
	4	0.537	1998	1.242	1.008	1.125	1.096
	p̂	0.329	1270	0.618	1.002	0.776	
	Fraction of time p̂=(1,2,3,4): (0, .306, .651, .043)						

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

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.029	1072	0.265	0.950	0.063	0.063
CCT	1	0.031	1158	0.254	0.946	0.062	0.063
CCT w/o reg.	1	0.031	1168	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.002	76	3.537	0.998	1.875	1.889
	2	0.109	4102	0.582	1.000	0.766	0.764
	3	0.273	10245	0.409	1.002	0.645	0.638
	4	0.588	21534	0.324	1.001	0.572	0.568
	p̂	0.554	20303	0.346	0.997	0.579	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, 0, .109, .891)						
CCT	0	0.002	80	3.545	1.000	1.883	1.894
	2	0.119	4450	0.593	1.001	0.771	0.768
	3	0.274	10257	0.480	1.005	0.695	0.682
	4	0.356	13288	0.601	1.003	0.772	0.762
	p̂	0.280	10469	0.482	1.004	0.694	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, .008, .949, .043)						
CCT w/o reg.	0	0.002	80	3.562	0.999	1.889	1.894
	2	0.120	4515	0.588	1.000	0.768	0.768
	3	0.296	11070	0.483	1.002	0.671	0.668
	4	0.534	19349	0.696	1.001	0.845	0.843
	p̂	0.357	13272	0.448	0.999	0.657	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, .007, .75, .243)						

Appendix

A Specifications of Data Generating Processes

A.1 Lee and Ludwig-Miller DGP's

To obtain the conditional expectation functions in the Lee and Ludwig-Miller DGP's, 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

$$\text{Lee: } E[Y|X=x] = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0 \\ 0.52 + 0.84x - 3.00x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{if } x \geqslant 0 \end{cases} \tag{A1}$$

$$\text{Ludwig-Miller: } E[Y|X=x] = \begin{cases} 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < 0 \\ 0.26 + 18.49x - 54.81x^2 + 74.30x^3 - 45.02x^4 + 9.83x^5 & \text{if } x \geqslant 0. \end{cases} \tag{A2}$$

Equations (A1) and (A2) 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 $\sigma_\varepsilon = 0.1295$.

A.2 Card-Lee-Pei-Weber DGP's

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

A.2.1 Bottom-kink DGP

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

$$\text{First-stage: } E[B|X=x] = \begin{cases} \beta_0 + \beta_1^+ x + \beta_2^+ x^2 + \beta_3^+ x^3 + \beta_4^+ x^4 + \beta_5^+ x^5 & \text{if } x < 0 \\ \beta_0 + \beta_1^- x + \beta_2^- x^2 + \beta_3^- x^3 + \beta_4^- x^4 + \beta_5^- x^5 & \text{if } x \geqslant 0 \end{cases} \tag{A3}$$

$$\text{Reduced-form: } E[Y|X=x] = \begin{cases} \gamma_0 + \gamma_1^+ x + \gamma_2^+ x^2 + \gamma_3^+ x^3 + \gamma_4^+ x^4 + \gamma_5^+ x^5 & \text{if } x < 0 \\ \gamma_0 + \gamma_1^- x + \gamma_2^- x^2 + \gamma_3^- x^3 + \gamma_4^- x^4 + \gamma_5^- x^5 & \text{if } x \geqslant 0 \end{cases} \tag{A4}$$

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}.

A.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 (A3) and (A4). 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}.

B AMSE Calculation and Estimation

B.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).[1] 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.[2]

B.2 AMSE Estimation

To estimate $\text{AMSE}_{\hat{\tau}_p}$, we rely on the proposed procedure in Calonico, Cattaneo and Titiunik (2014*a*) and Calonico, Cattaneo and Titiunik (2014*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*).[3] Again, the correspondences between $\hat{\text{B}}_p$ and $\hat{\text{V}}_p$ in this paper and their notations in Calonico, Cattaneo and Titiunik (2014*b*) are laid out in Table A.5. 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 $\hat{\tau}_p^{bc}$:

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

[1] In Table A.5, we summarize the correspondence between the expressions in this paper to those in Calonico, Cattaneo and Titiunik (2014*b*).

[2] The program used to generate Figure 1 is available at https://sites.google.com/site/peizhuan/programs/.

[3] The installation instruction is available at https://sites.google.com/site/peizhuan/programs/.

where b is the pilot bandwidth used in Calonico, Cattaneo and Titiunik (2014b) to estimate the bias of $\hat{\tau}_p$. According to Theorems A.1 and A.2 of Calonico, Cattaneo and Titiunik (2014b), 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 actually 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. (2015a). We follow the same algorithm to arrive at $\tilde{\mathrm{B}}_p^{bc}$. $\tilde{V}_p^{bc}$ is simply the estimated variance of $\hat{\tau}_p^{bc}$, and its computation is covered in detail in Calonico, Cattaneo and Titiunik (2014b).

Finally, we point out that our AMSE estimator is consistent for the true MSE. In the case of the conventional estimator, this means that

$$\frac{\widehat{\mathrm{AMSE}}_{\hat{\tau}_p}(h)}{\mathrm{MSE}_{\hat{\tau}_p}(h)} \xrightarrow{p} 1 \tag{A5}$$

under the regularity conditions in Calonico, Cattaneo and Titiunik (2014b). The consistency result of (A5) is established by noting: 1) $\hat{\mathrm{B}}_p$ and $\hat{\mathrm{V}}_p$ are consistent estimators of B_p and V_p so that $\widehat{\mathrm{AMSE}}_{\hat{\tau}_p}(h)/\mathrm{AMSE}_{\hat{\tau}_p}(h) \xrightarrow{p} 1$, and 2) $\mathrm{AMSE}_{\hat{\tau}_p}(h)/\mathrm{MSE}_{\hat{\tau}_p}(h) \to 1$ as $h \to 0$, following Lemmas 1 and 2 of Calonico, Cattaneo and Titiunik (2014b) for the sharp and fuzzy cases respectively. The consistency of $\widehat{AMSE}_{\hat{\tau}_p^{bc}}(h,b)$ can be similarly established.

C Simulation Results from Armstrong and Kolesár (forthcoming) and Imbens and Wager (forthcoming)

For comparison purposes, we additionally present simulation results using alternative procedures for estimation and inference as proposed by Armstrong and Kolesár (forthcoming) (henceforth AK) and Imbens and Wager (forthcoming) (henceforth IW). We apply the procedures to the Lee and Ludwig-Miller DGP's under three sample sizes: n_{actual} and n_{large} as in Tables 1-4, as well as $n_{small} = 500$, the sample size used in the simulations of Imbens and Kalyanaraman (2012), Calonico, Cattaneo and Titiunik (2014b), and Armstrong and Kolesár (forthcoming). Each method is implemented using the authors' R packages.[4] For these

[4] To improve the numerical stability in applying the IW package `optrdd`, we trim the data generated from the Ludwig-Miller DGP by restricting observations to the range of $|x| < 0.1$ (leaving about 13% of the observations), outside which the data points would have been assigned small weights in constructing the estimator anyway. In addition, we use the `mosek` optimizer option when calling `optrdd`. We thank Stefan Wager for these suggestions.

procedures, it is necessary to specify a bound on the magnitude of the second derivative of the conditional expectation function. For the purposes of these simulations, we simply use the true maximum of the second derivative magnitude from the DGP's (A1) and (A2), while keeping in mind that in practice this quantity is unknown, and IW and AK suggest examining the results for a range of different values for this bound.[5] We tabulate the coverage rates and the average lengths of the resulting 95% confidence intervals over 1000 repeated Monte Carlo samples in Appendix Table A.2. Since Calonico, Cattaneo and Titiunik (2014*b*) is another paper that focuses on inference improvement, we repeat the coverage rates and average lengths of the 95% local linear CCT robust confidence intervals using the CCT bandwidth selector in Table A.2, for ease of comparison.

Appendix Table A.2 reveals two patterns. First, the AK and IW procedures do very well in terms of coverage rates: their CI's have coverage rates generally higher than CCT.[6] At the same time, the length of the CCT robust CI's are somewhat shorter than their AK and IW counterparts. The reduction in length ranges from 8.2% to about 30% for the Lee DGP, and from 20% to about 60% for the Ludwig-Miller DGP.

The fact that the AK and IW CI's – in comparison to the CCT robust CI's – have comparable or higher coverage rates but at the expense of efficiency is to be expected. The AK and IW procedures intend to maintain coverage over a specified class of functions. The larger that class of functions is, the more "agnostic" one can be about the shape of the underlying DGP, leading to longer confidence intervals. Conversely, one would expect that as that class of functions is assumed to be more restrictive, the CI's would become shorter. Cheng, Fan and Marron (1997) describe the class of functions for which local polynomial estimators with a triangular kernel enjoys minimax properties. This suggests that in practice, there could be cases in which CCT performs reasonably well, as it does in this case. For n_{large}, CCT has 94.6% coverage for the Ludwig-Miller DGP and has the shortest CI length, and as we see in Tables 3 and 4, the CCT robust CI based on a higher-order local polynomial regression may lead to further improvements.

[5]When applying the IW approach to the Lugwig-Miller DGP, we choose the maximum second derivative of the conditional expectation function over the range of $|x| < 0.1$. This maximum is higher than that imposed in the AK approach, which is only the maximum of the second derivatives of the conditional expectation function evaluated on both sides of $x = 0$.

[6]Calonico, Cattaneo and Farrell (2016) and Calonico, Cattaneo and Farrell (2018) propose confidence intervals that minimize the coverage error, which may improve upon those presented in Appendix Table A.2.

Figure A.1: Conditional Expectation Functions in RDD DGP's

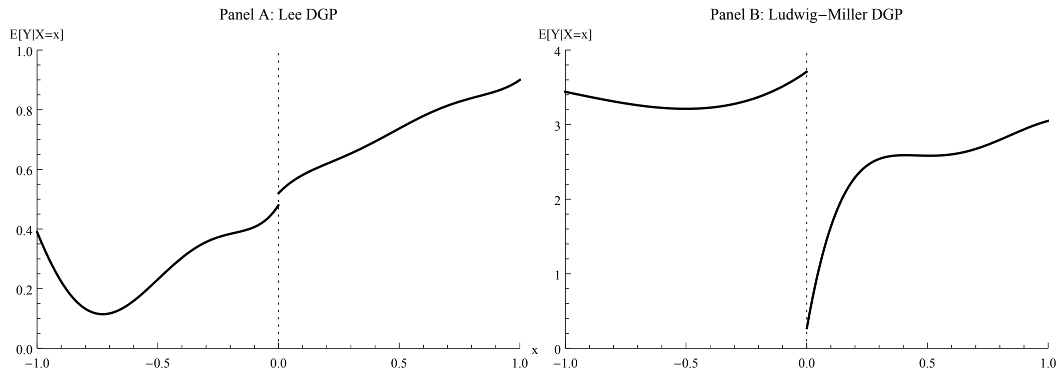


Figure A.2: Conditional Expectation Functions in RKD DGP's

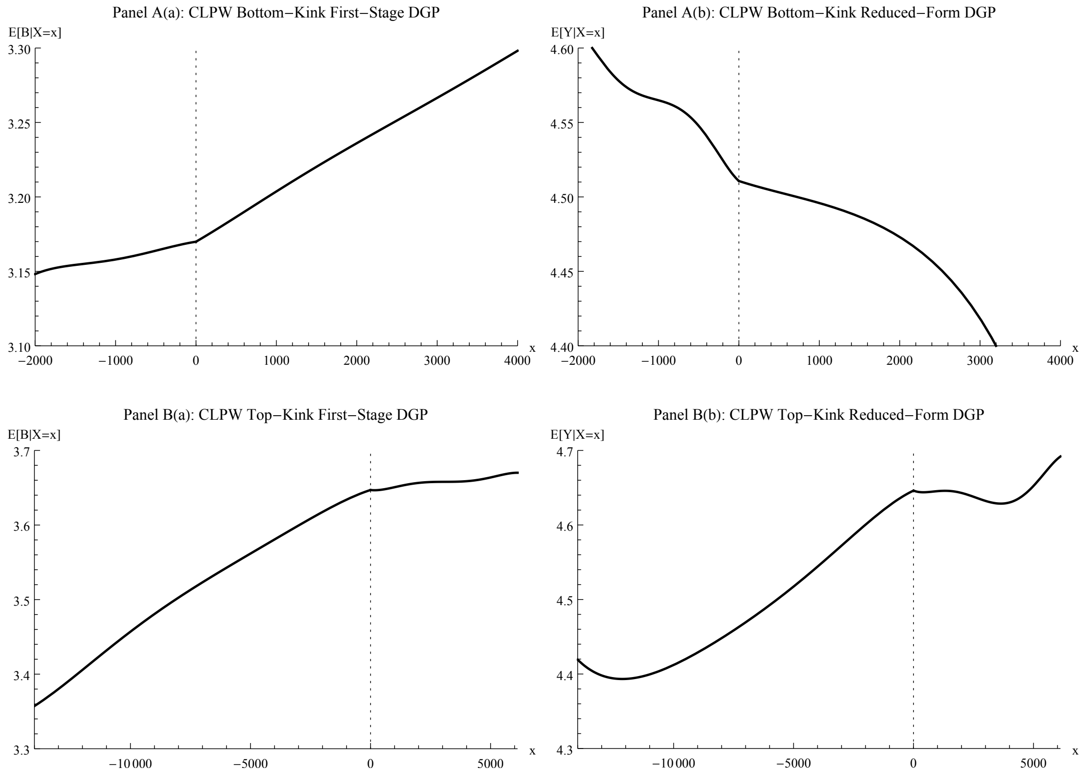
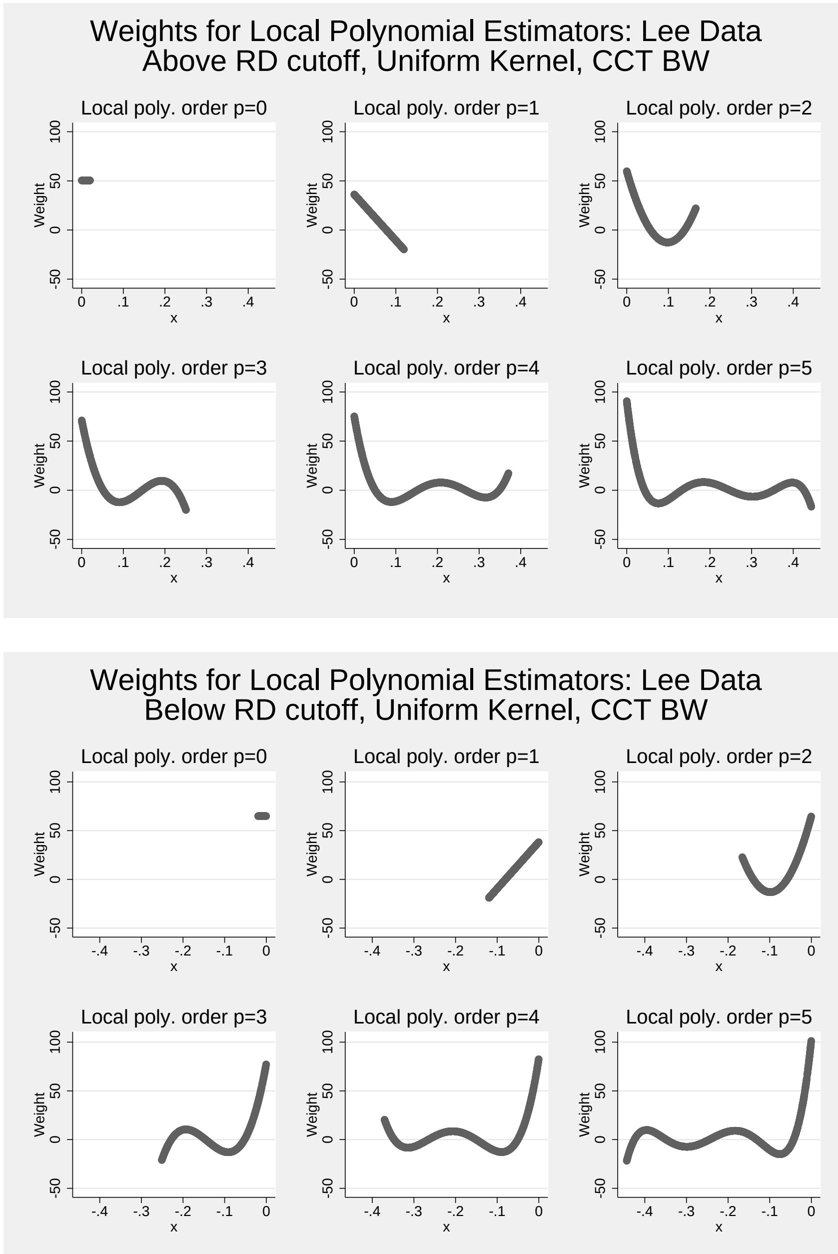
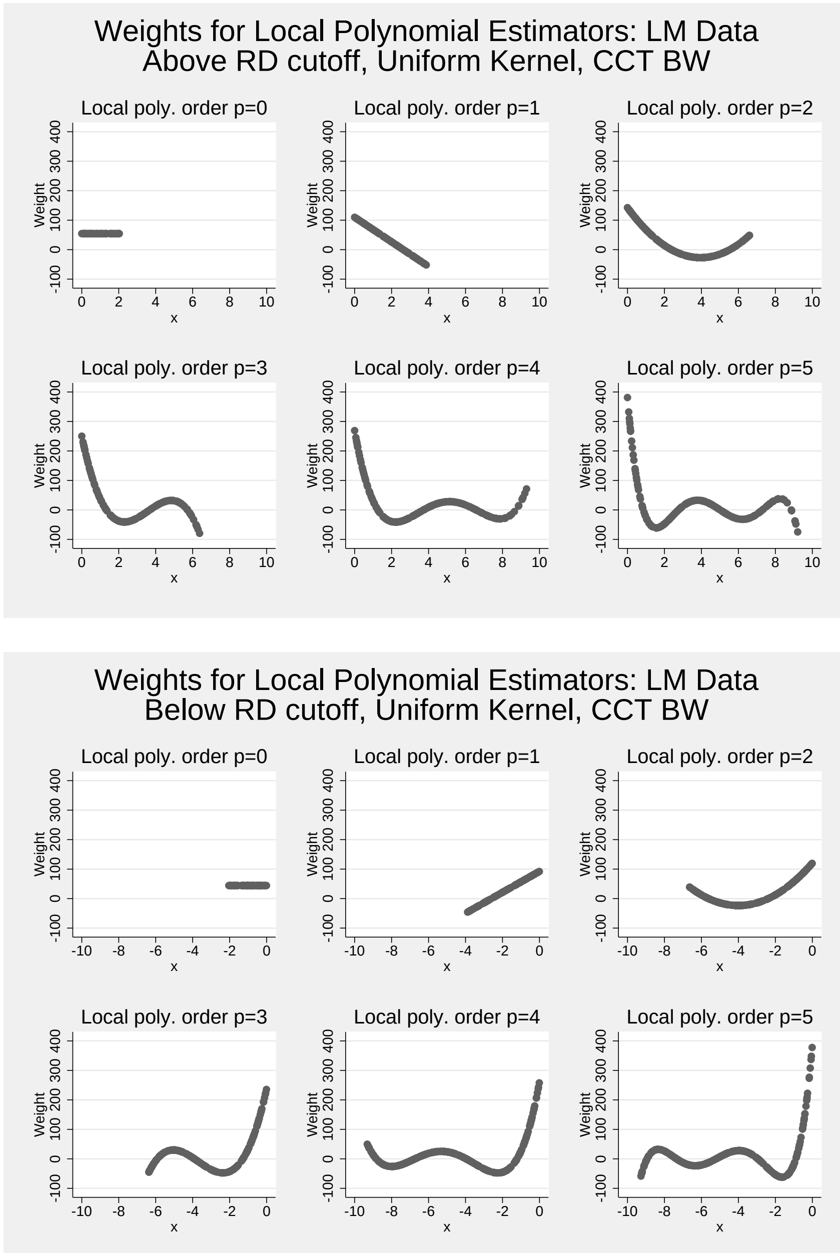


Figure A.3: Weights for Local Polynomial Estimators: Lee Data, CCT Bandwidth



Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the default CCT Bandwidth, h_{CCT}, for $p = 0, 1, ..., 5$.

Figure A.4: Weights for Local Polynomial Estimators: Ludwig-Miller Data, CCT Bandwidth



Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the default CCT Bandwidth, h_{CCT}, for $p = 0, 1, ..., 5$.

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 Armstrong and Kolesár (forthcoming) and Imbens and Wager (forthcoming) Confidence Intervals

		Coverage Rate of 95% CI			Average 95% CI Length		
		AK	IW	CCT Robust	AK	IW	CCT Robust
Lee DGP	n_{small}	0.956	0.976	0.920	0.319	0.267	0.245
	n_{actual}	0.962	0.964	0.900	0.110	0.093	0.077
	n_{large}	0.960	0.957	0.928	0.046	0.038	0.032
Ludwig-Miller DGP	n_{small}	0.976	0.932	0.933	0.565	0.425	0.353
	n_{actual}	0.973	0.964	0.935	0.228	0.205	0.154
	n_{large}	0.969	0.941	0.946	0.090	0.082	0.063

Table A.3: Simulation Statistics for the Conventional Estimator of Various Polynomial Orders: Lee and Ludwig-Miller DGP, Small Sample Size

Panel A: Lee DGP, Small Sample Size (n=500)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.130	81	3.922	0.927	0.230	0.256
CCT	1	0.159	99	3.952	0.902	0.211	0.248
CCT w/o reg.	1	0.387	218	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.059	35	1.130	0.949	0.940	1.092
	2	0.260	162	1.086	1.007	1.064	1.035
	3	0.470	291	1.088	1.006	1.066	1.041
	4	0.838	472	1.069	1.007	1.053	1.031
	$\hat{p}$	0.153	92	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	1.263	0.803	0.816	1.140
	2	0.226	141	1.364	1.028	1.256	1.181
	3	0.275	172	1.959	1.027	1.528	1.436
	4	0.313	195	2.772	1.022	1.810	1.709
	$\hat{p}$	0.096	60	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	1.920	0.743	0.974	1.542
	2	0.424	252	1.380	1.077	1.336	1.201
	3	0.464	280	1.858	1.117	1.665	1.414
	4	0.491	297	2.464	1.139	2.015	1.614
	$\hat{p}$	0.319	186	1.435	0.868	0.902	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.387, .566, .046, .002, 0)						

Panel B: 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)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.065	41	9.215	0.922	0.366	0.412
CCT	1	0.076	47	9.821	0.881	0.329	0.422
CCT w/o reg.	1	0.079	49	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.196	123	0.645	1.016	0.827	0.777
	3	0.431	268	0.504	1.023	0.742	0.682
	4	0.854	477	0.427	1.028	0.702	0.628
	$\hat{p}$	0.646	374	0.482	1.019	0.716	
	Fraction of time $\hat{p}$=(1,2,3,4): (.001, .003, .486, .511)						
CCT	2	0.206	129	0.658	1.047	0.890	0.772
	3	0.280	175	0.770	1.066	1.040	0.855
	4	0.319	199	1.086	1.066	1.254	1.028
	$\hat{p}$	0.218	136	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	0.711	1.020	0.841	0.821
	3	0.448	272	0.600	1.050	0.854	0.755
	4	0.504	303	0.771	1.070	1.030	0.845
	$\hat{p}$	0.396	242	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 and Ludwig-Miller DGP, Small Sample Size

Panel A: Lee DGP, Small Sample Size (n=500)

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.130	81	4.752	0.935	0.262	0.280
CCT	1	0.160	100	4.892	0.920	0.245	0.273
CCT w/o reg.	1	0.387	219	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.059	23	0.995	0.962	0.980	1.036
	2	0.260	162	0.998	1.000	1.006	0.994
	3	0.470	291	0.964	1.000	0.990	0.981
	4	0.838	472	1.170	1.000	1.089	1.083
	$\hat{p}$	0.169	99	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	0.754	0.976	0.824	0.886
	2	0.226	141	1.338	1.013	1.215	1.181
	3	0.276	172	1.878	1.014	1.451	1.409
	4	0.314	195	2.573	1.011	1.704	1.653
	$\hat{p}$	0.069	43	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	0.802	0.962	0.734	0.794
	2	0.424	252	1.347	1.053	1.276	1.151
	3	0.466	281	2.102	1.071	1.644	1.407
	4	0.495	299	4.824	1.073	2.033	1.721
	$\hat{p}$	0.119	72	0.743	0.953	0.718	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.877, .118, .005, 0, 0)						

Panel B: 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)
Bandwidth	p	Avg. h	Avg. n	MSE ×1000	Coverage Rate	Avg. CI Length	Avg. Size-adj. CI length
Theo. Optimal	1	0.065	41	8.565	0.941	0.381	0.398
CCT	1	0.076	47	7.990	0.933	0.353	0.379
CCT w/o reg.	1	0.079	49	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 MSE's	Ratio of Coverage Rates	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.196	123	0.686	1.006	0.821	0.804
	3	0.431	268	0.558	1.007	0.738	0.717
	4	0.854	477	0.649	1.008	0.808	0.783
	$\hat{p}$	0.422	260	0.592	1.002	0.745	
	Fraction of time $\hat{p}$=(1,2,3,4): (.006, .113, .834, .047)						
CCT	2	0.206	128	0.825	1.013	0.918	0.876
	3	0.281	175	1.116	1.011	1.081	1.037
	4	0.319	199	1.601	1.008	1.293	1.252
	$\hat{p}$	0.193	121	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	0.884	1.006	0.882	0.859
	3	0.448	273	1.361	1.013	1.181	1.130
	4	0.508	305	2.488	1.012	1.569	1.503
	$\hat{p}$	0.272	169	0.775	1.000	0.855	
	Fraction of time $\hat{p}$=(1,2,3,4): (.068, .737, .189, .007)						

Table A.5: Correspondence to the Expressions in Calonico, Cattaneo and Titiunik (2014*b*)

Expression in this paper	Expression in Calonico, Cattaneo and Titiunik (2014*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*).