Appendix (For Online Publication Only)

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

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

We first define

$$p_{opt} \equiv \underset{p \in \Omega}{\arg\min} \text{AMSE}_{\hat{\tau}_p}(h(p))$$

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

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

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

[1] Assumption 1 in Calonico, Cattaneo and Titiunik (2014*b*) consists of regularity conditions for the fourth moment

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

Assumption 1 states that p_{max} does not change with n. This is consistent with the standard approach in other contexts, such as choosing the order of a time series autoregression from a fixed candidate set by the Akaike or Bayesian Information Criterion that penalizes model complexity (Stock and Watson, 2011), or selecting from a fixed set of covariate polynomial terms in propensity score matching or LASSO (the candidate set in Imbens and Rubin, 2015 for propensity score matching and Chernozhukov et al., 2018 for LASSO consists of linear, linear interactions and quadratic terms). In addition, Assumption 1 is not restrictive by itself as the researcher may always pick a large enough p_{max} a priori regardless of n. However, care is needed as Calonico, Cattaneo and Titiunik (2015) and Gelman and Imbens (2019) express concerns regarding high-order global RD polynomial estimators related to the Runge phenomenon. The Runge phenomenon arises in the polynomial interpolation of a function $f(x)$ over an interval $[a,b]$: using a polynomial of order n to interpolate a function through $n+1$ equispaced knots when n is large does not imply uniform convergence to f. In fact, large departures from the function may result outside the interpolation knots, especially toward the edge of $[a,b]$. One textbook remedy (Ch. 4 of Dahlquist and Björck, 2008 and Ch. 8 of Björck, 1996) to guard against the Runge phenomenon is to employ least squares regression as opposed to interpolation. As a rule of thumb, the textbooks recommend using a polynomial order no larger than $2\sqrt{n}$ where n is the number of (equispaced) observations. However, this rule of thumb does not cater to local RD estimators, and researchers typically choose polynomial orders from a set with a much smaller p_{max}. The Stata package `rdrobust` (Calonico, Cattaneo and Titiunik, 2014*a* and Calonico et al., 2017) caps the polynomial order at 8, and 107 of the 110 RD papers we surveyed use local orders no larger than 5. Given the concerns voiced by Calonico, Cattaneo and Titiunik (2015) and Gelman and Imbens (2019) and the status quo of the econometric and applied literature, it is advisable to always limit p_{max} to be at or below 8 and in

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

most cases at or below 5.[2]

Part a) of Assumption 2 consists of standard regularity conditions that allow for the asymptotic approximation of MSE, and part b) encompasses the estimators $\hat{\mathrm{B}}_p$ and $\hat{\mathrm{V}}_p$ in Imbens and Kalyanaraman (2012) for $p = 1$ and Calonico, Cattaneo and Titiunik (2014*b*) as special cases. Note that a larger p_{max} translates to a higher degree of smoothness in Assumption 2, which may seem undesirable ostensibly. But it is also arbitrary to assume, for example, that the conditional expectation functions $E[Y_1|X=x]$ and $E[Y_0|X=x]$ have continuous second derivatives ($S=2$) but not continuous third derivatives ($S=3$). The technicality of Assumption 2 notwithstanding, for all practical purposes, we treat these conditional expectation functions as infinitely smooth.

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

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

Proof. First we show that $p_{opt} \to p_{max}$. As mentioned in section 2, Assumption 3 implies that

$$\mathrm{AMSE}_{\hat{\tau}_p}(h(p)) = \mathrm{C}_p \cdot n^{-\frac{2p+2}{2p+3}}, \tag{A1}$$

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

$$\frac{\mathrm{AMSE}_{\hat{\tau}_{p_{max}}}(h(p_{max}))}{\mathrm{AMSE}_{\hat{\tau}_p}(h(p))} \to 0$$

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

Next we show that $\hat{p} \xrightarrow{\mathbb{P}} p_{max}$. Under part b) of Assumption 2, Lemma A1 of Calonico, Cattaneo

[2] Another practical consideration is multicollinearity. As the polynomial order increases, multicollinearity is more likely when executing the regression. Using the Lee (2008) data, for example, `rdrobust` reports issues in bandwidth computation when $p = 7$.

and Titiunik (2014*b*) shows that

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

for each $p \in \Omega$. It follows that

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

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

Remark. We can calculate the rate at which p_{opt} converges to p_{max} based on equation (A1). For example, if a researcher is choosing between local linear and quadratic i.e. $\Omega = \{1,2\}$ as in Figure 1, quadratic is MSE-optimal when

$$n^{2/35} > \text{C}_2/\text{C}_1.$$

In this case, $|p_{opt} - p_{max}| = |p_{opt} - 2| < (\text{C}_2/\text{C}_1)n^{-2/35}$. Similarly, we can derive that for a general Ω where $\tilde{p}$ is the highest order in the set other than p_{max}, $|p_{opt} - p_{max}| = O(n^{-2(p_{max}-\tilde{p})/(2p_{max}+3)(2\tilde{p}+3)})$.

Proposition 1 says that under standard asymptotics as provided by Assumption 3, a) the optimal polynomial order is the "corner solution" p_{max} when the sample size is large; b) the order we select will also converge to p_{max} in probability. Point a) echos the insight from Porter (2003) and our discussion above that a higher order estimator will dominate in a sufficiently large sample when using optimal bandwidths. However, to reiterate our point made at the beginning of section 2, which we illustrate again in section 4 using the RKD example, the "corner solution" here reflects the theoretical property that $\text{AMSE}_{\hat{\tau}_p}$ decreases at a higher rate as a function of the sample size when p is larger.

It is clear from Figure 1 and the remark above that although p_{opt} converges to p_{max} asymptotically, p_{opt} may not coincide with p_{max} in any finite sample. This is true even in sample sizes conventionally considered to be large, as is the case with our RKD example in section 4. It is worth emphasizing that this is not a statement about the finite sample performance of $\hat{p}$—p_{opt} is not subject to sampling variation; instead, as discussed in section 2, it is about the important role of the constants (B_p and V_p for $p \in \Omega$) in determining p_{opt}, beyond the asymptotic rates in the remark above that push p_{opt} toward p_{max}.

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

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

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

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

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

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

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

Now we examine the three fractions inside the probability statement of (A2) one by one. For the first fraction, Lemma A1 of Calonico, Cattaneo and Titiunik (2014*b*) again implies that

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

for all p. The second fraction

$$\frac{\mathrm{AMSE}_{\hat{\tau}_p}(h(p))}{\mathrm{AMSE}_{\hat{\tau}_{p_{opt}}}(h(p_{opt}))} > 1$$

for all p by the definition and uniqueness of p_{opt}. For the third fraction, notice that for any $\varepsilon > 0$

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

By Assumption 1 and condition (A3), the right hand side of (A4) can be made arbitrarily small by choosing a large enough sample size. It follows that

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

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

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

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

Under Assumption 4, $\mathrm{AMSE}_{\hat{\tau}_p}$ shrinks at the same rate for all p. Therefore, the limit of p_{opt} is generally not p_{max}, and the AMSE of $\hat{\tau}_{p_{max}}$ does not always dominate that of alternative polynomial orders as is the case under Assumption 3. Instead, the optimal polynomial order depends on the magnitudes of the constants B_p and V_p from equation (3) even asymptotically. In another contrast with Assumption 3, under which p_{opt} is unique as $n \to \infty$, there exist DGPs for which the AMSEs are the same for different p. We therefore assume the uniqueness of p_{opt} in Proposition 2, but even if the uniqueness assumption is relaxed, $\hat{p}$ still has the desirable asymptotic no-regret per Li (1987) and Imbens and Kalyanaraman (2012). Namely, there is no loss asymptotically by using $\hat{p}$, as compared to any of the optimal orders that deliver the lowest MSE.

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

B Specifications of Data Generating Processes

B.1 Lee and Ludwig-Miller DGPs

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

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

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

Equations (A5) and (A6) are graphed in Appendix Figure A.1. The assignment variable X is specified as following the distribution $2\mathscr{B}(2,4)-1$, where $\mathscr{B}(a,b)$ denotes a beta distribution with shape parameters a and b. The outcome variable is given by $Y = E[Y|X=x]+\varepsilon$, where $\varepsilon \sim N(0,\sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.1295$.

B.2 Card-Lee-Pei-Weber DGPs

The process of specifying the Card-Lee-Pei-Weber DGPs are described in section 4.4.3 of Card et al. (2017). In both the bottom- and top-kink DGPs, the first-stage and reduced-form conditional expectation functions are specified as

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

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

We also specify $\sigma_B^2(0^+)$, $\sigma_B^2(0^-)$, $\sigma_Y^2(0^+)$, and $\sigma_Y^2(0^-)$, which are the conditional variances of B and Y given X just above and below the cutoff. Finally, we specify $f_X(0)$, the density of X at the cutoff. The values of these parameters are provided in Appendix Table B.5.

C AMSE Calculation and Estimation

C.1 Theoretical AMSE Calculation

After the full specification of a data generating process, we can calculate $\mathrm{AMSE}_{\hat{\tau}_p}(h)$ 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 $\mathrm{AMSE}_{\hat{\tau}_p}(h)$ according to equation (3). Specifically, B_p^2 depends on the $(p+1)$th derivatives on both sides of the cutoff, and V_p depends on the conditional variances on both sides of the cutoff as well as the density of the running variable at the cutoff. With B_p^2 and V_p computed, we can calculate the infeasible optimal bandwidth h_{opt} for a given sample size, which is simply a function of B_p^2 and V_p. Finally, plugging h_{opt} back into $\mathrm{AMSE}_{\hat{\tau}_p}(h)$ yields the AMSE for that given sample size, and Figure 1 is the graphical representation of this mapping across different sample sizes.

C.2 AMSE Estimation

To estimate $\mathrm{AMSE}_{\hat{\tau}_p}$, we rely on the proposed procedure in Calonico, Cattaneo and Titiunik (2014*a*,*b*). Our program `rdmse_cct2014` takes user-specified bandwidths as inputs and estimates $\hat{\mathrm{B}}_p^2$ and $\hat{\mathrm{V}}_p$ for the conventional estimator in the same way as Calonico, Cattaneo and Titiunik (2014*b*). The correspondences between $\hat{\mathrm{B}}_p$ and $\hat{\mathrm{V}}_p$ in this paper and their notations in Calonico, Cattaneo and Titiunik (2014*b*) are laid out in Table C.1. 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*b*), `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*b*) selects all of the closest neighbors, we randomly select three neighbors. We adopt the same modification in Card et al. (2015).

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

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

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

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

References

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

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

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

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

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

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "`rdrobust`: Software for Regression Discontinuity Designs." *Stata Journal*, 17: 372–404.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015. "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. 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.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *Econometrics Journal*, 21(1): C1–C68.

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

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

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press.

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

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

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

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

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics.* . 3rd ed., Pearson.

Figure A.1: Conditional Expectation Functions in RDD DGPs

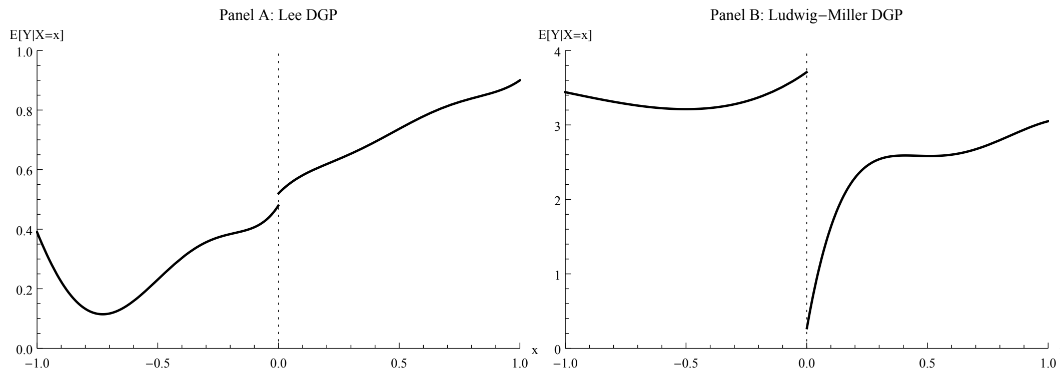

Figure A.2: Conditional Expectation Functions in RKD DGPs

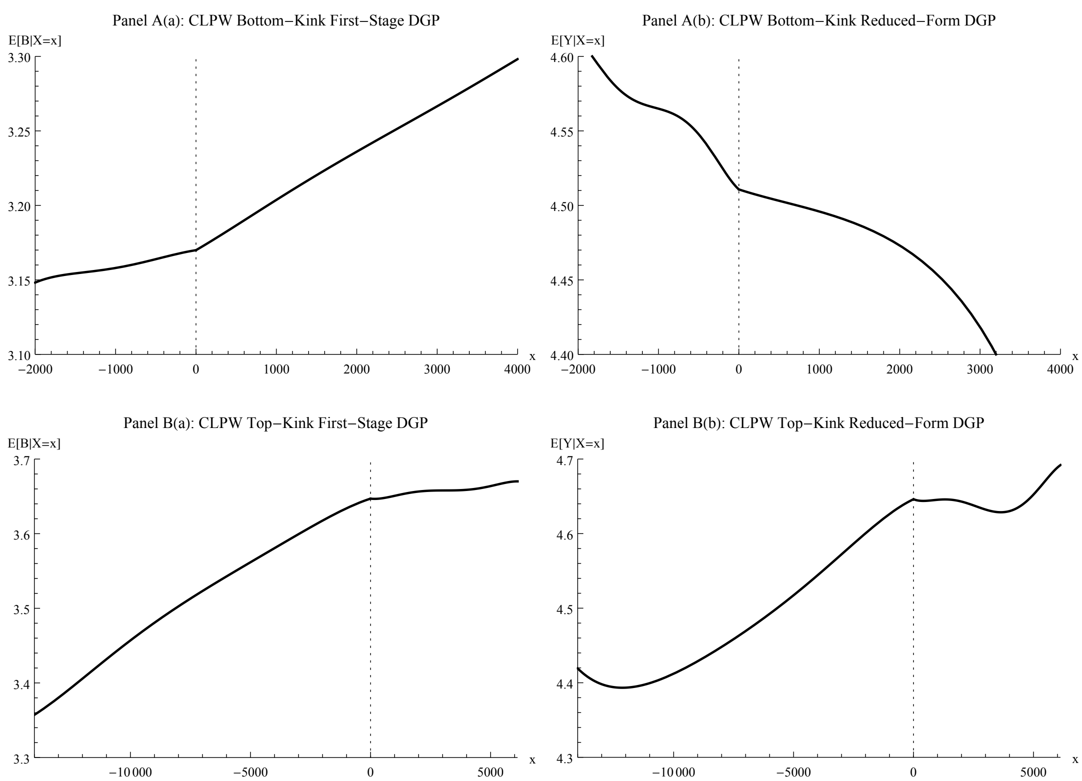

Table A.1: Main Specification of RD Papers Published in Leading Journals

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

Note: Our survey includes empirical RD papers published between 1999 and 2017 in the following 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*.

Table B.1: 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	0.099	811	0.483	0.947	0.085	0.086
CCT	1	0.139	1140	0.481	0.906	0.077	0.089

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.022	183	1.298	0.945	1.132	1.145
	2	0.216	1766	0.905	0.949	0.959	0.951
	3	0.407	3321	0.800	0.952	0.904	0.887
	4	0.747	5739	0.770	0.952	0.890	0.873
	$\hat{p}$			0.814	0.948	0.898	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .607, .392)						
CCT	0	0.032	266	1.056	0.909	1.034	1.025
	2	0.248	2030	0.980	0.938	1.061	0.953
	3	0.344	2808	1.141	0.946	1.166	1.018
	4	0.390	3180	1.503	0.945	1.337	1.182
	$\hat{p}$			1.009	0.903	0.993	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.219, .668, .101, .012, 0)						

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	0.064	4766	0.080	0.945	0.034	0.035
CCT	1	0.080	6020	0.075	0.931	0.032	0.034

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.011	798	1.624	0.948	1.290	1.283
	2	0.157	11782	0.815	0.950	0.909	0.894
	3	0.319	23847	0.677	0.948	0.829	0.822
	4	0.610	44516	0.568	0.947	0.763	0.757
	$\hat{p}$			0.596	0.944	0.766	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, 0, 0, .054, .946)						
CCT	0	0.013	983	1.493	0.935	1.248	1.237
	2	0.181	13539	0.883	0.942	0.964	0.936
	3	0.323	24147	0.827	0.946	0.941	0.893
	4	0.400	29856	1.021	0.950	1.047	0.976
	$\hat{p}$			0.832	0.942	0.934	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (0, .002, .196, .780, .023)						

Table B.2: 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	0.057	222	1.617	0.939	0.155	0.164
CCT	1	0.064	247	1.562	0.935	0.151	0.162

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.181	702	0.665	0.940	0.819	0.815
	3	0.406	1566	0.507	0.945	0.720	0.699
	4	0.814	2881	0.484	0.946	0.709	0.684
	p̂			0.515	0.941	0.715	
	Fraction of time p̂=(1,2,3,4): (0, 0, .700, .300)						
CCT	2	0.198	770	0.692	0.938	0.836	0.819
	3	0.337	1304	0.769	0.941	0.870	0.848
	4	0.384	1484	1.011	0.939	0.998	0.978
	p̂			0.685	0.939	0.828	
	Fraction of time p̂=(1,2,3,4): (0, .706, .294, .001)						

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	0.036	1364	0.251	0.947	0.062	0.063
CCT	1	0.039	1469	0.244	0.945	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	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.003	109	3.456	0.943	1.849	1.871
	2	0.131	4904	0.590	0.951	0.771	0.759
	3	0.315	11802	0.422	0.950	0.652	0.644
	4	0.662	23874	0.340	0.951	0.588	0.576
	p̂			0.373	0.944	0.594	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, 0, .105, .896)						
CCT	0	0.003	115	3.425	0.943	1.849	1.859
	2	0.141	5301	0.602	0.948	0.776	0.766
	3	0.315	11785	0.499	0.949	0.702	0.691
	4	0.399	14892	0.627	0.947	0.781	0.773
	p̂			0.499	0.948	0.701	
	Fraction of time p̂=(0,1,2,3,4): (0, 0, .010, .961, .029)						

Table B.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	0.166	103	3.692	0.922	0.222	0.250
CCT	1	0.205	128	3.901	0.893	0.202	0.246

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.053	33	1.101	0.878	0.929	1.077
	2	0.311	194	1.121	0.927	1.076	1.050
	3	0.542	333	1.137	0.928	1.087	1.056
	4	0.943	496	1.150	0.928	1.092	1.066
	$\hat{p}$			1.187	0.871	0.954	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.525, .430, .013, .018, .015)						
CCT	0	0.084	52	1.162	0.729	0.811	1.123
	2	0.271	169	1.386	0.919	1.269	1.194
	3	0.318	198	1.977	0.920	1.551	1.457
	4	0.351	219	2.778	0.918	1.848	1.729
	$\hat{p}$			1.138	0.742	0.831	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.731, .269, .001, 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	0.082	51	8.618	0.910	0.354	0.419
CCT	1	0.097	60	9.377	0.869	0.319	0.430

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.235	147	0.658	0.933	0.843	0.762
	3	0.497	307	0.521	0.944	0.762	0.664
	4	0.961	498	0.463	0.951	0.733	0.618
	$\hat{p}$			0.519	0.940	0.741	
	Fraction of time $\hat{p}$=(1,2,3,4): (.004, .005, .519, .473)						
CCT	2	0.246	154	0.660	0.915	0.904	0.764
	3	0.323	201	0.798	0.933	1.062	0.856
	4	0.357	222	1.141	0.930	1.287	1.046
	$\hat{p}$			0.677	0.912	0.903	
	Fraction of time $\hat{p}$=(1,2,3,4): (.014, .913, .072, .001)						

Table B.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	0.166	103	4.514	0.928	0.252	0.277
CCT	1	0.205	128	4.831	0.913	0.239	0.274

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	0	0.053	33	0.987	0.921	0.985	1.018
	2	0.311	194	1.026	0.932	1.017	1.003
	3	0.542	333	1.013	0.932	1.013	0.995
	4	0.943	496	1.282	0.933	1.139	1.128
	$\hat{p}$			0.999	0.920	0.971	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.626, .227, .032, .115, 0)						
CCT	0	0.084	52	0.725	0.900	0.825	0.859
	2	0.271	169	1.354	0.928	1.219	1.182
	3	0.318	198	1.895	0.925	1.461	1.416
	4	0.351	219	2.628	0.924	1.720	1.689
	$\hat{p}$			0.729	0.899	0.825	
	Fraction of time $\hat{p}$=(0,1,2,3,4): (.977, .023, 0, 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	0.082	51	8.026	0.933	0.372	0.401
CCT	1	0.097	60	7.562	0.928	0.346	0.385

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

Bandwidth	p	Avg. h	Avg. n	Ratio of MSE's	Coverage Rate	Ratio of Avg. CI Lengths	Ratio of Avg. Size-adj. CI lengths
Theo. Optimal	2	0.235	147	0.708	0.942	0.835	0.802
	3	0.497	307	0.586	0.947	0.757	0.712
	4	0.961	498	0.716	0.946	0.846	0.801
	$\hat{p}$			0.630	0.940	0.762	
	Fraction of time $\hat{p}$=(1,2,3,4): (.021, .124, .843, .012)						
CCT	2	0.246	154	0.862	0.939	0.928	0.873
	3	0.323	201	1.188	0.937	1.096	1.044
	4	0.357	222	1.681	0.934	1.315	1.272
	$\hat{p}$			0.866	0.932	0.919	
	Fraction of time $\hat{p}$=(1,2,3,4): (.186, .793, .021, 0)						

Table B.5: Parameter Values in the Card-Lee-Pei-Weber DGPs

	Bottom-Kink DGP		Top-Kink DGP	
Parameter	Above Cutoff	Below Cutoff	Above Cutoff	Below Cutoff
β_0	3.17	3.17	3.65	3.65
β_1	3.14×10^{-5}	8.40×10^{-6}	-3.70×10^{-6}	1.03×10^{-5}
β_2	5.30×10^{-9}	-1.21×10^{-8}	1.25×10^{-8}	-3.18×10^{-9}
β_3	-3.82×10^{-12}	-1.01×10^{-11}	-6.17×10^{-12}	-5.72×10^{-13}
β_4	9.54×10^{-16}	-7.56×10^{-16}	1.16×10^{-15}	-4.83×10^{-17}
β_5	-8.00×10^{-20}	7.89×10^{-19}	-7.43×10^{-20}	-1.42×10^{-21}
γ_0	4.51	4.51	4.65	4.65
γ_1	-1.76×10^{-5}	-4.75×10^{-5}	-1.29×10^{-5}	1.51×10^{-5}
γ_2	7.00×10^{-9}	1.64×10^{-7}	2.35×10^{-8}	-5.69×10^{-9}
γ_3	-5.00×10^{-12}	3.04×10^{-10}	-1.42×10^{-11}	-1.07×10^{-12}
γ_4	1.00×10^{-15}	1.82×10^{-13}	3.04×10^{-15}	-8.49×10^{-17}
γ_5	-2.00×10^{-19}	3.53×10^{-17}	-2.06×10^{-19}	-2.65×10^{-21}
σ_B^2	2.05×10^{-4}	2.07×10^{-4}	1.20×10^{-3}	9.60×10^{-4}
σ_Y^2	1.51	1.49	1.62	1.63
f_X	1.53×10^{-4}	1.53×10^{-4}	2.35×10^{-5}	2.35×10^{-5}

Note: For $j = 1, \ldots, 5$, the values of β_j, γ_j, σ_B^2, and σ_Y^2 above the cutoff correspond, respectively, to those of β_j^+, γ_j^+, $\sigma_B^2(0^+)$, and $\sigma_Y^2(0^+)$, which are defined in Appendix B.2. The values of β_j, γ_j, σ_B^2, and σ_Y^2 below the cutoff correspond, respectively, to those of β_j^-, γ_j^-, $\sigma_B^2(0^-)$, and $\sigma_Y^2(0^-)$. By construction, the values of β_0, γ_0, and f_X are the same on both sides of the cutoff.

Table C.1: Correspondence to the Expressions in Calonico, Cattaneo and Titiunik (2014*a*,*b*)

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

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