Regression Discontinuity Designs Using Covariates: Supplemental Appendix*

Sebastian Calonico† Matias D. Cattaneo‡ Max H. Farrell§ Rocio Titiunik¶

April 24, 2019

Abstract

This supplemental appendix contains the proofs of the main results, several extensions, additional methodological and technical results, and further simulation details, not included in the main paper to conserve space.

*Cattaneo gratefully acknowledges financial support from the National Science Foundation through grants SES-1357561 and SES-1459931, and Titiunik gratefully acknowledges financial support from the National Science Foundation through grant SES-1357561.

†Mailman School of Public Health, Columbia University.

‡Department of Operations Research and Financial Engineering, Princeton University.

§Booth School of Business, University of Chicago.

¶Department of Politics, Princeton University.

Contents

Part I

Omitted Details from Main Paper

This section briefly summarizes omitted details from the main paper concerning the Sharp RD design, and also reports a an overview of the main results, analogous to those reported in the main paper, for other RD designs. All remaining details are given in Part II and Part III below.

1 Sharp RD Design Main Formulas

We give a very succinct account of the main expressions for sharp RD designs, which were omitted in the main paper to avoid overwhelming notation. These formulas are derived in Part II below, where also the notation is introduced. The main goal of this section is to give a quick self-contained account of the main expressions, but most details on notation are postponed to the following parts of this supplemental appendix.

Let $\mathbf{R}_p(h) = [(\mathbf{r}_p((X_1-\bar{x})/h),\cdots,\mathbf{r}_p((X_n-\bar{x})/h))']$ be the $n\times(1+p)$ design matrix, and $\mathbf{K}_-(h) = \mathrm{diag}(\mathbb{1}(X_i<\bar{x})K_h(X_i-\bar{x}) : i=1,2,\cdots,n)$ and $\mathbf{K}_+(h) = \mathrm{diag}(\mathbb{1}(X_i\geq\bar{x})K_h(X_i-\bar{x}) : i=1,2,\cdots,n)$ be the $n\times n$ weighting matrices for control and treatment units, respectively. We also define $\boldsymbol{\mu}_{S-}^{(a)} := (\mu_{Y-}^{(a)}, \boldsymbol{\mu}_{Z-}^{(a)}{}')'$, $\boldsymbol{\mu}_{S+}^{(a)} := (\mu_{Y+}^{(a)}, \boldsymbol{\mu}_{Z+}^{(a)}{}')'$, $a\in\mathbb{Z}_+$, where $g^{(s)}(x) = \partial^s g(x)/\partial x^s$ for any sufficiently smooth function $g(\cdot)$. Let $\boldsymbol{\sigma}_{S-}^2 := \mathbb{V}[\mathbf{S}_i(0)|X_i=\bar{x}]$ and $\boldsymbol{\sigma}_{S+}^2 := \mathbb{V}[\mathbf{S}_i(1)|X_i=\bar{x}]$, and recall that $\mathbf{S}_i(t) = (Y_i(t),\mathbf{Z}_i(t))'$, $t\in\{0,1\}$. Let $\mathbf{e}_\nu$ denote a conformable $(1+\nu)$-th unit vector. Finally, recall that $\mathbf{s}(h) = (1,-\tilde{\boldsymbol{\gamma}}_Y(h))'$ and $\mathbf{s} = (1,-\boldsymbol{\gamma}_Y)'$.

The pre-asymptotic bias $\mathcal{B}_{\tilde{\tau}}(h) = \mathcal{B}_{\tilde{\tau}+}(h) - \mathcal{B}_{\tilde{\tau}-}(h)$ and its asymptotic counterpart $\mathcal{B}_{\tilde{\tau}} := \mathcal{B}_{\tilde{\tau}+} - \mathcal{B}_{\tilde{\tau}-}$ are characterized by

$$\mathcal{B}_{\tilde{\tau}-}(h) := \mathbf{e}_0'\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p}(h)\frac{\mathbf{s}'\boldsymbol{\mu}_{S-}^{(p+1)}}{(p+1)!} \to_{\mathbb{P}} \mathcal{B}_{\tilde{\tau}-} := \mathbf{e}_0'\boldsymbol{\Delta}_{p,-}\frac{\mathbf{s}'\boldsymbol{\mu}_{S-}^{(p+1)}}{(p+1)!}$$

$$\mathcal{B}_{\tilde{\tau}+}(h) := \mathbf{e}_0'\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\vartheta}_{+,p}(h)\frac{\mathbf{s}'\boldsymbol{\mu}_{S+}^{(p+1)}}{(p+1)!} \to_{\mathbb{P}} \mathcal{B}_{\tilde{\tau}+} := \mathbf{e}_0'\boldsymbol{\Delta}_{p,+}\frac{\mathbf{s}'\boldsymbol{\mu}_{S+}^{(p+1)}}{(p+1)!}$$

where, with the (slightly abusive) notation $\mathbf{v}^k = (v_1^k, v_2^k,\cdots,v_n^k)'$, $\boldsymbol{\iota}_n = (1,\cdots,1)'\in\mathbb{R}^n$, $\boldsymbol{\Gamma}_{-,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{R}_p(h)/n$ and $\boldsymbol{\vartheta}_{-,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)(\mathbf{X}-\bar{x}\boldsymbol{\iota}_n/h)^{p+1}/n$, $\boldsymbol{\Gamma}_{+,p}(h)$ and $\boldsymbol{\vartheta}_{+,p}(h)$ defined analogously after replacing $\mathbf{K}_-(h)$ with $\mathbf{K}_+(h)$, and

$$\boldsymbol{\Delta}_{p,-} := \left(\int_{-\infty}^0 \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du\right)^{-1}\left(\int_{-\infty}^0 \mathbf{r}_p(u)u^{1+p}K(u)du\right),$$

$$\boldsymbol{\Delta}_{p,+} := \left(\int_0^{\infty} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du\right)^{-1}\left(\int_0^{\infty} \mathbf{r}_p(u)u^{1+p}K(u)du\right).$$

The pre-asymptotic variance $\mathcal{V}_{\tilde{\tau}}(h) = \mathcal{V}_{\tilde{\tau}-}(h) + \mathcal{V}_{\tilde{\tau}+}(h)$ and its asymptotic counterpart $\mathcal{V}_{\tilde{\tau}} :=$

$\mathcal{V}_{\tilde{\tau}-} + \mathcal{V}_{\tilde{\tau}+}$ are characterized by

$$\mathcal{V}_{\tilde{\tau}-}(h) := [\mathbf{s}' \otimes \mathbf{e}_0'\mathbf{P}_{-,p}(h)]\mathbf{\Sigma}_{S-}[\mathbf{s} \otimes \mathbf{P}_{-,p}(h)\mathbf{e}_0] \ \to_{\mathbb{P}} \ \mathcal{V}_{\tilde{\tau}-} := \frac{\mathbf{s}'\boldsymbol{\sigma}^2_{S-}\mathbf{s}}{f}\mathbf{e}_0'\mathbf{\Lambda}_{p,-}\mathbf{e}_0$$

$$\mathcal{V}_{\tilde{\tau}+}(h) := [\mathbf{s}' \otimes \mathbf{e}_0'\mathbf{P}_{+,p}(h)]\mathbf{\Sigma}_{S+}[\mathbf{s} \otimes \mathbf{P}_{+,p}(h)\mathbf{e}_0] \ \to_{\mathbb{P}} \ \mathcal{V}_{\tilde{\tau}+} := \frac{\mathbf{s}'\boldsymbol{\sigma}^2_{S+}\mathbf{s}}{f}\mathbf{e}_0'\mathbf{\Lambda}_{p,+}\mathbf{e}_0$$

where $\mathbf{P}_{-,p}(h) = \sqrt{h}\mathbf{\Gamma}^{-1}_{-,p}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)/\sqrt{n}$ and $\mathbf{P}_{+,p}(h) = \sqrt{h}\mathbf{\Gamma}^{-1}_{+,p}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)/\sqrt{n}$, and

$$\mathbf{\Lambda}_{p,-} := \left(\int_{-\infty}^{0} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du\right)^{-1} \left(\int_{-\infty}^{0} \mathbf{r}_p(u)\mathbf{r}_p(u)'^2du\right) \left(\int_{-\infty}^{0} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du\right)^{-1},$$

$$\mathbf{\Lambda}_{p,+} := \left(\int_{0}^{\infty} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du\right)^{-1} \left(\int_{0}^{\infty} \mathbf{r}_p(u)\mathbf{r}_p(u)'^2du\right) \left(\int_{0}^{\infty} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du\right)^{-1}.$$

To construct pre-asymptotic estimates of the bias terms, we replace the only unknowns, $\boldsymbol{\mu}^{(p+1)}_{S-}$ and $\boldsymbol{\mu}^{(p+1)}_{S+}$, by q-th order ($p < q$) local polynomial estimates thereof, using the preliminary bandwidth b. This leads to the pre-asymptotic feasible bias estimate $\tilde{\mathcal{B}}_{\tilde{\tau}}(b) := \tilde{\mathcal{B}}_{\tilde{\tau}+}(b) - \tilde{\mathcal{B}}_{\tilde{\tau}-}(b)$ with

$$\tilde{\mathcal{B}}_{\tilde{\tau}-}(b) := \mathbf{e}_0'\mathbf{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p}(h)\frac{\mathbf{s}(h)'\tilde{\boldsymbol{\mu}}^{(p+1)}_{S-,q}(b)}{(p+1)!} \quad \text{and} \quad \tilde{\mathcal{B}}_{\tilde{\tau}+}(b) := \mathbf{e}_0'\mathbf{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p}(h)\frac{\mathbf{s}(h)'\tilde{\boldsymbol{\mu}}^{(p+1)}_{S+,q}(b)}{(p+1)!}$$

where $\tilde{\boldsymbol{\mu}}^{(p+1)}_{S-,q}(b)$ and $\tilde{\boldsymbol{\mu}}^{(p+1)}_{S+,q}(b)$ collect the q-th order local polynomial estimates of the $(p+1)$-th derivatives using as outcomes each of the variables in $\mathbf{S}_i = (Y_i, \mathbf{Z}_i')'$ for control and treatment units. Therefore, the bias-corrected covariate-adjusted sharp RD estimator is

$$\tilde{\tau}^{\mathsf{bc}}(h) = \frac{1}{\sqrt{nh}}[\mathbf{s}(h)' \otimes \mathbf{e}_0'(\mathbf{P}^{\mathsf{bc}}_{+,p}(h,b) - \mathbf{P}^{\mathsf{bc}}_{-,p}(h,b))]\mathbf{S},$$

with $\mathbf{S} = (\mathbf{Y}, \mathrm{vec}(\mathbf{Z})')'$, $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)'$, and

$$\mathbf{P}^{\mathsf{bc}}_{-,p}(h,b) = \sqrt{h}\mathbf{\Gamma}^{-1}_{-,p}(h)\left[\mathbf{R}_p(h)'\mathbf{K}_-(h) - \rho^{1+p}\boldsymbol{\vartheta}_{-,p}(h)\mathbf{e}_{p+1}'\mathbf{\Gamma}^{-1}_{-,q}(b)\mathbf{R}_q(b)'\mathbf{K}_-(b)\right]/\sqrt{n},$$

$$\mathbf{P}^{\mathsf{bc}}_{+,p}(h,b) = \sqrt{h}\mathbf{\Gamma}^{-1}_{+,p}(h)\left[\mathbf{R}_p(h)'\mathbf{K}_+(h) - \rho^{1+p}\boldsymbol{\vartheta}_{+,p}(h)\mathbf{e}_{p+1}'\mathbf{\Gamma}^{-1}_{+,q}(b)\mathbf{R}_q(b)'\mathbf{K}_+(b)\right]/\sqrt{n},$$

where $\tilde{\mathbf{P}}^{\mathsf{bc}}_{-,p}(h,b)$ and $\tilde{\mathbf{P}}^{\mathsf{bc}}_{-,p}(h,b)$ are directly computable from observed data, given the choices of bandwidth h and b, with $\rho = h/b$, and the choices of polynomial order p and q, with $p < q$.

The exact form of the (pre-asymptotic) heteroskedasticity-robust or cluster-robust variance estimator follows directly from the formulas above. All other details such preliminary bandwidth selection, plug-in data-driven MSE-optimal bandwidth estimation, and other extensions and results, are given in the upcoming parts of this supplemental appendix.

2 Other RD designs

As we show below, our main results extend naturally to cover other popular RD designs, including fuzzy, kink, and fuzzy kink RD. Here we give a short overview of the main ideas, deferring all details to the upcoming Parts II and III below. There are two wrinkles to the standard sharp RD design discussed so far that must be accounted for: ratios of estimands/estimators for fuzzy designs and derivatives in estimands/estimators for kink designs.

2.1 Fuzzy RD Designs

The distinctive feature of fuzzy RD designs is that treatment compliance is imperfect. This implies that $T_i = T_i(0) \cdot \mathbb{1}(X_i < \bar{x}) + T_i(1) \cdot \mathbb{1}(X_i \geq \bar{x})$, that is, the treatment status T_i of each unit $i = 1, 2, \cdots, n$ is no longer a deterministic function of the running variable X_i, but $\mathbb{P}[T_i = 1 | X_i = x]$ still changes discontinuously at the RD threshold level $\bar{x}$. Here, $T_i(0)$ and $T_i(1)$ denote the two potential treatment status for each unit i when, respectively, $X_i < \bar{x}$ (not offered treatment) and $X_i \geq \bar{x}$ (offered treatment).

To analyze the case of fuzzy RD designs, we first recycle notation for potential outcomes and covariates as follows:

$$\begin{aligned} Y_i(t) &:= Y_i(0) \cdot (1 - T_i(t)) + Y_i(1) \cdot T_i(t) \\ \mathbf{Z}_i(t) &:= \mathbf{Z}_i(0) \cdot (1 - T_i(t)) + \mathbf{Z}_i(1) \cdot T_i(t) \end{aligned}$$

for $t = 0, 1$. That is, in this setting, potential outcomes and covariates are interpreted as their "reduced form" (or intention-to-treat) counterparts. Giving causal interpretation to covariate-adjusted instrumental variable type estimators is delicate; see e.g. Abadie (2003) for more discussion. Nonetheless, the above re-definitions enable us to use the same notation, assumptions, and results, already given for the sharp RD design, taking the population target estimands as simply the probability limits of the RD estimators.

We employ Assumption SA-5 (in Part III below), which complements Assumption SA-3 (in Part II below). The standard fuzzy RD estimand is

$$\varsigma = \frac{\tau_Y}{\tau_T}, \qquad \tau_Y = \mu_{Y+} - \mu_{Y-}, \qquad \tau_T = \mu_{T+} - \mu_{T-},$$

where recall that we continue to omit the evaluation point $x = \bar{x}$, and we have redefined the potential outcomes and additional covariates to incorporate imperfect treatment compliance. Furthermore, now τ has a subindex highlighting the outcome variable being considered (Y or T), and hence $\tau = \tau_Y$ by definition.

The standard estimator of ς, without covariate adjustment, is

$$\hat{\varsigma}(h) = \frac{\hat{\tau}_Y(h)}{\hat{\tau}_T(h)}, \qquad \hat{\tau}_V(h) = \mathbf{e}_0'\hat{\boldsymbol{\beta}}_{V+,p}(h) - \mathbf{e}_0'\hat{\boldsymbol{\beta}}_{V-,p}(h),$$

with $V \in \{Y, T\}$, where the exact definitions are given below. Similarly, the covariate-adjusted fuzzy RD estimator is

$$\tilde{\varsigma}(h) = \frac{\tilde{\tau}_Y(h)}{\tilde{\tau}_T(h)}, \qquad \tilde{\tau}_V(h) = \mathbf{e}_0'\tilde{\boldsymbol{\beta}}_{V+,p}(h) - \mathbf{e}_0'\tilde{\boldsymbol{\beta}}_{V-,p}(h),$$

with $V \in \{Y, T\}$, where the exact definitions are given below. Our notation makes clear that the fuzzy RD estimators, with or without additional covariates, are simply the ratio of two sharp RD estimators, with or without covariates.

The properties of the standard fuzzy RD estimator $\hat{\varsigma}(h)$ were studied in great detail before, while the covariate-adjusted fuzzy RD estimator $\tilde{\varsigma}(h)$ has not been studied in the literature before.

Let Assumptions SA-1, SA-3, and SA-5 hold. If $nh \to \infty$ and $h \to 0$, then

$$\tilde{\varsigma}(h) \to_{\mathbb{P}} \frac{\tau_Y - [\boldsymbol{\mu}_{Z+} - \boldsymbol{\mu}_{Z-}]'\boldsymbol{\gamma}_Y}{\tau_T - [\boldsymbol{\mu}_{Z+} - \boldsymbol{\mu}_{Z-}]'\boldsymbol{\gamma}_T},$$

where $\boldsymbol{\gamma}_V = (\boldsymbol{\sigma}^2_{Z-} + \boldsymbol{\sigma}^2_{Z+})^{-1}\mathbb{E}[(\mathbf{Z}_i(0) - \boldsymbol{\mu}_{Z-}(X_i))V_i(0) + (\mathbf{Z}_i(1) - \boldsymbol{\mu}_{Z+}(X_i))V_i(1)|X_i = \bar{x}]$ with $V \in \{Y, T\}$.

Under the same conditions, when no additional covariates are included, it is well known that $\hat{\varsigma}(h) \to_{\mathbb{P}} \varsigma$. Thus, this result clearly shows that both probability limits will coincide under the same sufficient condition as in the sharp RD design: $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z+}$. Therefore, at least asymptotically, a (causal) interpretation for the probability limit of the covariate-adjusted fuzzy RD estimator can be deduced from the corresponding (causal) interpretation for the probability limit of the standard fuzzy RD estimator, whenever the condition $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z+}$ holds.

Since the fuzzy RD estimators are constructed as a ratio of two sharp RD estimators, their asymptotic properties can be characterized by studying the asymptotic properties of the corresponding sharp RD estimators, which have already been analyzed in previous sections. Specifically, the asymptotic properties of covariate-adjusted fuzzy RD estimator $\tilde{\zeta}(h)$ can be characterized by employing the following linear approximation:

$$\tilde{\varsigma}(h) - \varsigma = \mathbf{f}_{\varsigma}'(\tilde{\boldsymbol{\tau}}(h) - \boldsymbol{\tau}) + \epsilon_{\tilde{\varsigma}},$$

with

$$\mathbf{f}_{\varsigma} = \begin{bmatrix} \frac{1}{\tau_T} \\ -\frac{\tau_Y}{\tau_T^2} \end{bmatrix}, \qquad \tilde{\boldsymbol{\tau}}(h) = \begin{bmatrix} \tilde{\tau}_Y(h) \\ \tilde{\tau}_T(h) \end{bmatrix}, \qquad \boldsymbol{\tau} = \begin{bmatrix} \tau_Y \\ \tau_T \end{bmatrix},$$

and where the term $\epsilon_{\tilde{\varsigma}}$ is a quadratic (high-order) error. Therefore, it is sufficient to study the asymptotic properties of the bivariate vector $\tilde{\boldsymbol{\tau}}(h)$ of covariate-adjusted sharp RD estimators, provided that $\epsilon_{\tilde{\varsigma}}$ is asymptotically negligible relative to the linear approximation, which is proven below in this supplemental appendix. As before, while not necessary for most of our results, we continue to assume that $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z+}$ so the standard RD estimand is recovered by the covariate-adjusted fuzzy RD estimator.

Employing the linear approximation and parallel results as those discussed above for the sharp RD design (now also using T_i as outcome variable as appropriate), it is conceptually straightforward to conduct inference in fuzzy RD designs with covariates. All the same results outlined in the previous section are established for this case: in this supplemental appendix we present MSE expansions, MSE-optimal bandwidth, MSE-optimal point estimators, consistent bandwidth selectors, robust bias-corrected distribution theory and consistent standard errors under either heteroskedasticity or clustering, for the covariate-robust fuzzy RD estimator $\tilde{\varsigma}(h)$. All details are given in Part III below, and these results are implemented in the general purpose software packages for `R` and `Stata` described in Calonico, Cattaneo, Farrell, and Titiunik (2017).

2.2 Kink RD Designs

Our final extension concerns the so-called kink RD designs. See Card, Lee, Pei, and Weber (2015) for a discussion on identification and Calonico, Cattaneo, and Titiunik (2014b) for a discussion on estimation and inference, both covering sharp and fuzzy settings without additional covariates. We briefly outline identification and consistency results when additional covariates are included in kink RD estimation (i.e., derivative estimation at the cutoff), but relegate all other inference results to the upcoming parts of this supplemental appendix.

The standard sharp kink RD parameter is (proportional to)

$$\tau_{Y,1} = \mu_{Y+}^{(1)} - \mu_{Y-}^{(1)},$$

while the fuzzy kink RD parameter is

$$\varsigma_1 = \frac{\tau_{Y,1}}{\tau_{T,1}}$$

where $\tau_{T,1} = \mu_{T+}^{(1)} - \mu_{T-}^{(1)}$. In the absence of additional covariates in the RD estimation, these RD treatment effects are estimated by using the local polynomial plug-in estimators:

$$\hat{\tau}_{Y,1}(h) = \mathbf{e}_1'\hat{\boldsymbol{\beta}}_{Y+,p}(h) - \mathbf{e}_1'\hat{\boldsymbol{\beta}}_{Y-,p}(h) \qquad \text{and} \qquad \hat{\varsigma}_1(h) = \frac{\hat{\tau}_{Y,1}(h)}{\hat{\tau}_{T,1}(h)},$$

where $\mathbf{e}_1$ denote the conformable 2nd unit vector (i.e., $\mathbf{e}_1 = (0, 1, 0, 0, \cdots, 0)'$). Therefore, the covariate-adjusted kink RD estimators in sharp and fuzzy settings are

$$\tilde{\tau}_{Y,1}(h) = \mathbf{e}_1'\tilde{\boldsymbol{\beta}}_{Y+,p}(h) - \mathbf{e}_1'\tilde{\boldsymbol{\beta}}_{Y-,p}(h)$$

and

$$\tilde{\varsigma}_1(h) = \frac{\tilde{\tau}_{Y,1}(h)}{\tilde{\tau}_{T,1}(h)}, \qquad \tilde{\tau}_{V,1}(h) = \mathbf{e}_1'\tilde{\boldsymbol{\beta}}_{V+,p}(h) - \mathbf{e}_1'\tilde{\boldsymbol{\beta}}_{V-,p}(h), \qquad V \in \{Y, T\},$$

respectively. The following lemma gives our main identification and consistency results.

Let Assumptions SA-1, SA-3, and SA-5 hold. If $nh \to \infty$ and $h \to 0$, then

$$\tilde{\tau}_{Y,1}(h) \to_{\mathbb{P}} \tau_{Y,1} - [\boldsymbol{\mu}_{Z+}^{(1)} - \boldsymbol{\mu}_{Z-}^{(1)}]'\boldsymbol{\gamma}_Y$$

and

$$\tilde{\varsigma}_1(h) \to_{\mathbb{P}} \frac{\tau_{Y,1} - [\boldsymbol{\mu}_{Z+}^{(1)} - \boldsymbol{\mu}_{Z-}^{(1)}]'\boldsymbol{\gamma}_Y}{\tau_{T,1} - [\boldsymbol{\mu}_{Z+}^{(1)} - \boldsymbol{\mu}_{Z-}^{(1)}]'\boldsymbol{\gamma}_T},$$

where $\boldsymbol{\gamma}_Y$ and $\boldsymbol{\gamma}_T$ are defined in the upcoming sections, and recall that $\boldsymbol{\mu}_{Z-}^{(1)} = \boldsymbol{\mu}_{Z-}^{(1)}(\bar{x})$ and $\boldsymbol{\mu}_{Z+}^{(1)} = \boldsymbol{\mu}_{Z+}^{(1)}(\bar{x})$ with $\boldsymbol{\mu}_{Z-}^{(1)}(x) = \partial\boldsymbol{\mu}_{Z-}(x)/\partial x$ and $\boldsymbol{\mu}_{Z+}^{(1)}(x) = \partial\boldsymbol{\mu}_{Z+}(x)/\partial x$.

As before, in this setting it is well known that $\hat{\tau}_{Y,1}(h) \to_{\mathbb{P}} \tau_{Y,1}$ (sharp kink RD) and $\hat{\varsigma}_1(h) \to_{\mathbb{P}} \varsigma_1$ (fuzzy kink RD), formalizing once again that the estimand when covariates are included is in general different from the standard kink RD estimand without covariates. In this case, a sufficient condition for the estimands with and without covariates to agree is $\boldsymbol{\mu}_{Z+}^{(1)} = \boldsymbol{\mu}_{Z-}^{(1)}$ for both sharp and fuzzy kink RD designs.

While the above results are in qualitative agreement with the sharp and fuzzy RD cases, and therefore most conclusions transfer directly to kink RD designs, there is one interesting difference concerning the sufficient conditions guaranteeing that both estimands coincide: a sufficient condition now requires $\boldsymbol{\mu}_{Z+}^{(1)} = \boldsymbol{\mu}_{Z-}^{(1)}$. This requirement is not related to the typical falsification test conducted in empirical work, that is, $\boldsymbol{\mu}_{Z+} = \boldsymbol{\mu}_{Z-}$, but rather a different feature of the conditional distributions of the additional covariates given the score—the first derivative of the regression function at the cutoff. Therefore, this finding suggests a new falsification test for empirical work in kink RD designs: testing for a zero sharp *kink* RD treatment effect on "pre-intervention" covariates. For example, this can be done using standard sharp kink RD treatment effect results, using each covariate as outcome variable.

As before, inference results follow the same logic already discussed (see Parts II and III for details). All the results are fully implemented in the `R` and `Stata` software described by Calonico, Cattaneo, Farrell, and Titiunik (2017).

Part II

Sharp RD Designs

Let $|\cdot|$ denote the Euclidean matrix norm, that is, $|\mathbf{A}|^2 = \text{trace}(\mathbf{A}'\mathbf{A})$ for scalar, vector or matrix $\mathbf{A}$. Let $a_n \precsim b_n$ denote $a_n \leq Cb_n$ for positive constant C not depending on n, and $a_n \asymp b_n$ denote $C_1 b_n \leq a_n \leq C_2 b_n$ for positive constants C_1 and C_2 not depending on n. When a subindex $\mathbb{P}$ is present in the notation, the corresponding statements refer to "in probability". In addition, statements such as "almost surely", "for h small enough" or "for n large enough" (depending on the specific context) are omitted to simplify the exposition. Throughout the paper and supplemental appendix $\nu, p, q \in \mathbb{Z}_+$ with $\nu \leq p < q$ unless explicitly noted otherwise.

3 Setup

3.1 Notation

Recall the basic notation introduced in the paper for Sharp RD designs. The outcome variable and other covariates are

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$$

$$\mathbf{Z}_i = T_i \cdot \mathbf{Z}_i(1) + (1 - T_i) \cdot \mathbf{Z}_i(0)$$

with $(Y_i(0), Y_i(1))$ denoting the potential outcomes, T_i denoting treatment status, X_i denoting the running variable, and $(\mathbf{Z}_i(0)', \mathbf{Z}_i(1)')$ denoting the other (potential) covariates, $\mathbf{Z}_i(0) \in \mathbb{R}^d$ and $\mathbf{Z}_i(1) \in \mathbb{R}^d$. In sharp RD designs, $T_i = \mathbb{1}(X_i \geq \bar{x})$.

We also employ the following vectors and matrices:

$$\mathbf{Y} = [Y_1, \cdots, Y_n]', \qquad \mathbf{X} = [X_1, \cdots, X_n]',$$

$$\mathbf{Z} = [\mathbf{Z}_1, \cdots, \mathbf{Z}_n]', \qquad \mathbf{Z}_i = [Z_{i1}, Z_{i2}, \cdots, Z_{id}]', \qquad i = 1, 2, \cdots, n,$$

$$\mathbf{Y}(0) = [Y_1(0), \cdots, Y_n(0)]', \qquad \mathbf{Y}(1) = [Y_1(1), \cdots, Y_n(1)]',$$

$$\mathbf{Z}(0) = [\mathbf{Z}_1(0), \cdots, \mathbf{Z}_n(0)]', \qquad \mathbf{Z}(1) = [\mathbf{Z}_1(1), \cdots, \mathbf{Z}_n(1)]',$$

$$\boldsymbol{\mu}_{Y-}(\mathbf{X}) = \mathbb{E}[\mathbf{Y}(0)|\mathbf{X}], \qquad \boldsymbol{\mu}_{Y+}(\mathbf{X}) = \mathbb{E}[\mathbf{Y}(1)|\mathbf{X}],$$

$$\boldsymbol{\Sigma}_{Y-} = \mathbb{V}[\mathbf{Y}(0)|\mathbf{X}], \qquad \boldsymbol{\Sigma}_{Y+} = \mathbb{V}[\mathbf{Y}(1)|\mathbf{X}],$$

$$\boldsymbol{\mu}_{Z-}(\mathbf{X}) = \mathbb{E}[\text{vec}(\mathbf{Z}(0))|\mathbf{X}], \qquad \boldsymbol{\mu}_{Z+}(\mathbf{X}) = \mathbb{E}[\text{vec}(\mathbf{Z}(1))|\mathbf{X}].$$

$$\boldsymbol{\Sigma}_{Z-} = \mathbb{V}[\text{vec}(\mathbf{Z}(0))|\mathbf{X}], \qquad \boldsymbol{\Sigma}_{Z+} = \mathbb{V}[\text{vec}(\mathbf{Z}(1))|\mathbf{X}].$$

Recall that $\mathbf{e}_\nu$ denotes the conformable $(\nu+1)$-th unit vector, which may take different dimensions in different places.

We also define:

$$\mu_{Y-}(x) = \mathbb{E}[Y_i(0)|X_i = x], \qquad \mu_{Y+}(x) = \mathbb{E}[Y_i(1)|X_i = x],$$

$$\sigma^2_{Y-}(x) = \mathbb{V}[Y_i(0)|X_i = x], \qquad \sigma^2_{Y+}(x) = \mathbb{V}[Y_i(1)|X_i = x],$$

and

$$\boldsymbol{\mu}_{Z-}(x) = \mathbb{E}[\mathbf{Z}_i(0)|X_i = x], \qquad \boldsymbol{\mu}_{Z+}(x) = \mathbb{E}[\mathbf{Z}_i(1)|X_i = x],$$

$$\boldsymbol{\sigma}^2_{Z-}(x) = \mathbb{V}[\mathbf{Z}_i(0)|X_i = x], \qquad \boldsymbol{\sigma}^2_{Z+}(x) = \mathbb{V}[\mathbf{Z}_i(1)|X_i = x],$$

where

$$\mu_{Z_\ell -}(x) = \mathbb{E}[Z_{i\ell}(0)|X_i = x], \qquad \mu_{Z_\ell +}(x) = \mathbb{E}[Z_{i\ell}(1)|X_i = x],$$

for $\ell = 1, 2, \cdots, d$.

In addition, to study sharp RD designs with covariates, we need to handle the joint distribution of the outcome variable and the additional covariates. Thus, we introduce the following additional notation:

$$\mathbf{S}_i = \left[Y_i, \mathbf{Z}_i'\right]', \qquad \mathbf{S}_i(0) = \left[Y_i(0), \mathbf{Z}_i(0)'\right]', \qquad \mathbf{S}_i(1) = \left[Y_i(1), \mathbf{Z}_i(1)'\right]',$$

$$\mathbf{S} = [\mathbf{Y}, \mathbf{Z}], \qquad \mathbf{S}(0) = [\mathbf{Y}(0), \mathbf{Z}(0)], \qquad \mathbf{S}(1) = [\mathbf{Y}(1), \mathbf{Z}(1)],$$

$$\boldsymbol{\mu}_{S-}(\mathbf{X}) = \mathbb{E}[\operatorname{vec}(\mathbf{S}(0))|\mathbf{X}], \qquad \boldsymbol{\mu}_{S+}(\mathbf{X}) = \mathbb{E}[\operatorname{vec}(\mathbf{S}(1))|\mathbf{X}],$$

$$\boldsymbol{\Sigma}_{S-} = \mathbb{V}[\operatorname{vec}(\mathbf{S}(0))|\mathbf{X}], \qquad \boldsymbol{\Sigma}_{S+} = \mathbb{V}[\operatorname{vec}(\mathbf{S}(1))|\mathbf{X}],$$

$$\boldsymbol{\mu}_{S-}(x) = \mathbb{E}[\mathbf{S}_i(0)|X_i = x], \qquad \boldsymbol{\mu}_{S+}(x) = \mathbb{E}[\mathbf{S}_i(1)|X_i = x],$$

$$\boldsymbol{\sigma}^2_{S-}(x) = \mathbb{V}[\mathbf{S}_i(0)|X_i = x], \qquad \boldsymbol{\sigma}^2_{S+}(x) = \mathbb{V}[\mathbf{S}_i(1)|X_i = x].$$

3.2 Assumptions

We employ the following assumptions, which are exactly the ones discussed in the main paper.

Assumption SA-1 (Kernel) *The kernel function $k(\cdot) : [0, 1] \mapsto \mathbb{R}$ is bounded and nonnegative, zero outside its support, and positive and continuous on $(0, 1)$. Let*

$$K(u) = \mathbb{1}(u < 0)k(-u) + \mathbb{1}(u \geq 0)k(u),$$

$$K_{\mathbf{h}}(u) = \mathbb{1}(u < 0)k_{h_-}(-u) + \mathbb{1}(u \geq 0)k_{h_+}(u), \qquad k_h(u) = \frac{1}{h}k\left(\frac{u}{h}\right), \qquad \mathbf{h} = (h_-, h_+)'.$$

In what follows, h denotes a generic bandwidth (e.g., $h = h_-$ or $h = h_+$ depending on the context). Whenever necessarily, we assume throughout that $h_- \propto h_+$ for simplicity.

Assumption SA-2 (SRD, Standard) *For $\varrho \geq 1$, $x_l, x_u \in \mathbb{R}$ with $x_l < \bar{x} < x_u$, and all $x \in [x_l, x_u]$:*

(a) The Lebesgue density of X_i, denoted $f(x)$, is continuous and bounded away from zero.
(b) $\mu_{Y-}(x)$ and $\mu_{Y+}(x)$ are ϱ times continuously differentiable.
(c) $\sigma^2_{Y-}(x)$ and $\sigma^2_{Y+}(x)$ are continuous and invertible.
(d) $\mathbb{E}[|Y_i(t)|^4|X_i = x]$, $t \in \{0,1\}$, are continuous.

Assumption SA-3 (SRD, Covariates) *For $\varrho \geq 1$, $x_l, x_u \in \mathbb{R}$ with $x_l < \bar{x} < x_u$, and all $x \in [x_l, x_u]$:*
(a) $\mathbb{E}[\mathbf{Z}_i(0)Y_i(0)|X_i = x]$ and $\mathbb{E}[\mathbf{Z}_i(1)Y_i(1)|X_i = x]$ are continuously differentiable.
(b) $\boldsymbol{\mu}_{S-}(x)$ and $\boldsymbol{\mu}_{S+}(x)$ are ϱ times continuously differentiable.
(c) $\boldsymbol{\sigma}^2_{S-}(x)$ and $\boldsymbol{\sigma}^2_{S+}(x)$ are continuous and invertible.
(d) $\mathbb{E}[|\mathbf{S}_i(t)|^4|X_i = x]$, $t \in \{0,1\}$, are continuous.
(e) $\boldsymbol{\mu}^{(\nu)}_{Z-}(\bar{x}) = \boldsymbol{\mu}^{(\nu)}_{Z+}(\bar{x})$.

4 Standard Sharp RD

The main properties of the standard sharp RD estimator have been already analyzed in Calonico, Cattaneo, and Titiunik (2014b) and Calonico, Cattaneo, and Farrell (2018, 2019) in great detail. In this supplement we only reproduce the results needed to study the covariate-adjusted RD estimator.

Under Assumption SA-2, the standard (without covariate adjustment) sharp RD estimand for $\nu \leq S$ is:

$$\tau_{Y,\nu} = \mu^{(\nu)}_{Y+} - \mu^{(\nu)}_{Y-},$$

$$\mu^{(\nu)}_{Y+} = \mu^{(\nu)}_{Y+}(\bar{x}) = \frac{\partial^\nu}{\partial x^\nu}\mu_{Y+}(x)\bigg|_{x=\bar{x}}, \qquad \mu^{(\nu)}_{Y-} = \mu^{(\nu)}_{Y-}(\bar{x}) = \frac{\partial^\nu}{\partial x^\nu}\mu_{Y-}(x)\bigg|_{x=\bar{x}},$$

where we set $\mu_{Y-} = \mu^{(0)}_{Y-}$ and $\mu_{Y+} = \mu^{(0)}_{Y+}$. Define

$$\boldsymbol{\beta}_{Y-,p} = \boldsymbol{\beta}_{Y-,p}(\bar{x}) = \left[\mu_{Y-}\ ,\ \frac{1}{1!}\mu^{(1)}_{Y-}\ ,\ \frac{1}{2!}\mu^{(2)}_{Y-}\ ,\ \cdots\ ,\ \frac{1}{p!}\mu^{(p)}_{Y-}\right]',$$

$$\boldsymbol{\beta}_{Y+,p} = \boldsymbol{\beta}_{Y+,p}(\bar{x}) = \left[\mu_{Y+}\ ,\ \frac{1}{1!}\mu^{(1)}_{Y+}\ ,\ \frac{1}{2!}\mu^{(2)}_{Y+}\ ,\ \cdots\ ,\ \frac{1}{p!}\mu^{(p)}_{Y+}\right]'.$$

The standard, without covariate adjustment, sharp RD estimator for $\nu \leq p$ is:

$$\hat{\tau}_{Y,\nu}(\mathbf{h}) = \hat{\mu}^{(\nu)}_{Y+,p}(h_+) - \hat{\mu}^{(\nu)}_{Y-,p}(h_-),$$

$$\hat{\mu}^{(\nu)}_{Y+,p}(h) = \nu!\mathbf{e}'_\nu\hat{\boldsymbol{\beta}}_{Y+,p}(h), \qquad \hat{\mu}^{(\nu)}_{Y-,p}(h) = \nu!\mathbf{e}'_\nu\hat{\boldsymbol{\beta}}_{Y-,p}(h),$$

$$\hat{\boldsymbol{\beta}}_{Y-,p}(h) = \underset{\boldsymbol{\beta}\in\mathbb{R}^{1+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}(X_i < \bar{x})(Y_i - \mathbf{r}_p(X_i - \bar{x})'\boldsymbol{\beta})^2 k_h(-(X_i - \bar{x})),$$

$$\hat{\boldsymbol{\beta}}_{Y+,p}(h) = \underset{\boldsymbol{\beta}\in\mathbb{R}^{1+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}(X_i \geq \bar{x})(Y_i - \mathbf{r}_p(X_i - \bar{x})'\boldsymbol{\beta})^2 k_h(X_i - \bar{x}),$$

where $\mathbf{r}_p(x) = (1, x, \cdots, x^p)'$, $\mathbf{e}_\nu$ is the conformable $(\nu+1)$-th unit vector, $k_h(u) = k(u/h)/h$ with $k(\cdot)$ the kernel function, and h is a positive bandwidth sequence. This gives

$$\hat{\boldsymbol{\beta}}_{Y-,p}(h) = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\Upsilon}_{Y-,p}(h), \qquad \hat{\boldsymbol{\beta}}_{Y,+,p}(h) = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\Upsilon}_{Y+,p}(h),$$

$$\boldsymbol{\Gamma}_{-,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{R}_p(h)/n, \qquad \boldsymbol{\Upsilon}_{Y-,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{Y}/n,$$

$$\boldsymbol{\Gamma}_{+,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbf{R}_p(h)/n, \qquad \boldsymbol{\Upsilon}_{Y+,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbf{Y}/n,$$

where

$$\mathbf{R}_p(h) = \left[\mathbf{r}_p\left(\frac{X_1-\bar{x}}{h}\right), \mathbf{r}_p\left(\frac{X_2-\bar{x}}{h}\right), \cdots, \mathbf{r}_p\left(\frac{X_n-\bar{x}}{h}\right)\right]'_{n\times(1+p)},$$

$$\mathbf{H}_p(h) = \operatorname{diag}(h^j : j = 0, 1, \cdots, p) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & h & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h^p \end{bmatrix}_{(1+p)\times(1+p)},$$

$$\mathbf{K}_-(h) = \operatorname{diag}(\mathbb{1}(X_i < \bar{x})k_h(-(X_i - \bar{x})) : i = 1, 2, \cdots, n),$$

$$\mathbf{K}_+(h) = \operatorname{diag}(\mathbb{1}(X_i \geq \bar{x})k_h(X_i - \bar{x}) : i = 1, 2, \cdots, n).$$

We introduce the following additional notation:

$$\boldsymbol{\mu}_{Y-}(\mathbf{X}) = [\mu_{Y-}(X_1), \mu_{Y-}(X_2), \cdots, \mu_{Y-}(X_n)]',$$

$$\boldsymbol{\mu}_{Y+}(\mathbf{X}) = [\mu_{Y+}(X_1), \mu_{Y+}(X_2), \cdots, \mu_{Y+}(X_n)]',$$

and, with the (slightly abusive) notation $\mathbf{v}^k = (v_1^k, v_2^k, \cdots, v_n^k)$ for $\mathbf{v} \in \mathbb{R}^n$,

$$\boldsymbol{\vartheta}_{-,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)((\mathbf{X} - \bar{x}\boldsymbol{\iota}_n)/h)^{1+p}/n,$$

$$\boldsymbol{\vartheta}_{+,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_+(h)((\mathbf{X} - \bar{x}\boldsymbol{\iota}_n)/h)^{1+p}/n,$$

where $\boldsymbol{\iota}_n = (1, 1, \cdots, 1)' \in \mathbb{R}^n$.

Finally, to save notation, set

$$\begin{aligned} \mathbf{P}_{-,p}(h) &= \sqrt{h}\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)/\sqrt{n}, \\ \mathbf{P}_{+,p}(h) &= \sqrt{h}\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)/\sqrt{n}, \end{aligned}$$

which gives

$$\hat{\boldsymbol{\beta}}_{Y-,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)\mathbf{Y},$$

$$\hat{\boldsymbol{\beta}}_{Y+,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)\mathbf{Y}.$$

4.1 Hessian Matrices and Invertibility

The following lemma handles the Hessian matrices $\boldsymbol{\Gamma}_{-,p}$ and $\boldsymbol{\Gamma}_{+,p}$. This result is used below to give conditions for asymptotic invertibility, thereby making local polynomial estimators well defined in large samples.

Lemma SA-1 *Let Assumptions SA-1 and SA-2 hold. If $nh \to \infty$ and $h \to 0$, then*

$$\boldsymbol{\Gamma}_{-,p}(h) = \mathbb{E}[\boldsymbol{\Gamma}_{-,p}(h)] + o_{\mathbb{P}}(1) \qquad \text{and} \qquad \boldsymbol{\Gamma}_{+,p}(h) = \mathbb{E}[\boldsymbol{\Gamma}_{+,p}(h)] + o_{\mathbb{P}}(1),$$

with

$$\mathbb{E}[\boldsymbol{\Gamma}_{-,p}(h)] = \boldsymbol{\Gamma}_{-,p}\{1+o(1)\}, \qquad \boldsymbol{\Gamma}_{-,p} = f\int_{-\infty}^{0}\mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du,$$

$$\mathbb{E}[\boldsymbol{\Gamma}_{+,p}(h)] = \boldsymbol{\Gamma}_{+,p}\{1+o(1)\}, \qquad \boldsymbol{\Gamma}_{+,p} = f\int_{0}^{\infty}\mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)du,$$

where recall that $f = f(\bar{x})$.

Proof of Lemma SA-1. Recall that for any random variable, vector or matrix $\mathbf{A}_n$, Markov inequality implies $\mathbf{A}_n = \mathbb{E}[\mathbf{A}_n] + O_{\mathbb{P}}(|\mathbb{V}[\mathbf{A}_n]|)$. Thus, the result follows by noting that $|\mathbb{V}[\boldsymbol{\Gamma}_{-,p}(h)]|^2 = O(n^{-1}h^{-1})$ and similarly $|\mathbb{V}[\boldsymbol{\Gamma}_{+,p}(h)]|^2 = O(n^{-1}h^{-1})$. The second part follows by changing variables and taking limits. ∎

4.2 Conditional Bias

We characterize the smoothing bias of the standard RD estimator $\hat{\tau}_{Y,\nu}(\mathbf{h})$. We have

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{Y-,p}(h)|\mathbf{X}] = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbb{E}[\mathbf{Y}(0)|\mathbf{X}]/n,$$

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{Y+,p}(h)|\mathbf{X}] = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbb{E}[\mathbf{Y}(1)|\mathbf{X}]/n.$$

Lemma SA-2 *Let Assumptions SA-1 and SA-2 hold with $\varrho \geq p+2$. If $nh \to \infty$ and $h \to 0$, then*

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{Y-,p}(h)|\mathbf{X}] = \boldsymbol{\beta}_{Y-,p} + \mathbf{H}_p^{-1}(h)\left[h^{1+p}\mathbf{B}_{Y-,p,p}(h) + h^{2+p}\mathbf{B}_{Y-,p,1+p}(h) + o_{\mathbb{P}}(h^{2+p})\right],$$

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{Y+,p}(h)|\mathbf{X}] = \boldsymbol{\beta}_{Y+,p} + \mathbf{H}_p^{-1}(h)\left[h^{1+p}\mathbf{B}_{Y+,p,p}(h) + h^{2+p}\mathbf{B}_{Y+,p,1+p}(h) + o_{\mathbb{P}}(h^{2+p})\right],$$

with

$$\mathbf{B}_{Y-,p,a}(h) = \boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p,a}(h)\frac{\mu_{Y-}^{(1+a)}}{(1+a)!} \to_{\mathbb{P}} \mathbf{B}_{Y-,p,a} = \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\vartheta}_{-,p,a}\frac{\mu_{Y-}^{(1+a)}}{(1+a)!},$$

$$\mathbf{B}_{Y+,p,a}(h) = \boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\vartheta}_{+,p,a}(h)\frac{\mu_{Y+}^{(1+a)}}{(1+a)!} \to_{\mathbb{P}} \mathbf{B}_{Y+,p,a} = \boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\vartheta}_{+,p,a}\frac{\mu_{Y+}^{(1+a)}}{(1+a)!},$$

and where

$$\boldsymbol{\vartheta}_{-,p,a} = f\int_{-\infty}^{0} \mathbf{r}_p(u)u^{1+a}K(u)du, \qquad \boldsymbol{\vartheta}_{+,p,a} = f\int_{0}^{\infty} \mathbf{r}_p(u)u^{1+a}K(u)du,$$

$\mu_{Y-}^{(1+p)} = \mu_{Y-}^{(1+p)}(\bar{x})$, $\mu_{Y+}^{(1+p)} = \mu_{Y+}^{(1+p)}(\bar{x})$ *and* $f = f(\bar{x})$.

Proof of Lemma SA-2. A Taylor series expansion of $\mu_{Y-}(x)$ at $x = \bar{x}$ gives

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}_{Y-,p}(h)|\mathbf{X}] &= \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\boldsymbol{\mu}_{Y-}(\mathbf{X})/n \\
&= \boldsymbol{\beta}_{Y-,p} + \mathbf{H}_p^{-1}(h)\left[h^{1+p}\mathbf{B}_{Y-,p,p}(h) + h^{2+p}\mathbf{B}_{Y-,p,1+p}(h) + o_{\mathbb{P}}(h^{2+p})\right],
\end{aligned}$$

and similarly for $\mathbb{E}[\hat{\boldsymbol{\beta}}_{Y+,p}(h)|\mathbf{X}]$, verifying the first two results. For the last two results, Lemma SA-1 gives $\boldsymbol{\Gamma}_{-,p}^{-1}(h) = \boldsymbol{\Gamma}_{-,p}^{-1} + o_{\mathbb{P}}(1)$ and $\boldsymbol{\Gamma}_{+,p}^{-1}(h) = \boldsymbol{\Gamma}_{+,p}^{-1} + o_{\mathbb{P}}(1)$, while by proceeding as in the proof of that lemma we have $\boldsymbol{\vartheta}_{-,p,a}(h) = \mathbb{E}[\boldsymbol{\vartheta}_{-,p,a}(h)] + o_{\mathbb{P}}(1)$ and $\boldsymbol{\vartheta}_{-,p,a}(h) = \mathbb{E}[\boldsymbol{\vartheta}_{-,p,a}(h)] + o_{\mathbb{P}}(1)$, and by changing variables and taking limits we obtain $\mathbb{E}[\boldsymbol{\vartheta}_{-,p,a}(h)] \to \boldsymbol{\vartheta}_{-,p,a}$ and $\mathbb{E}[\boldsymbol{\vartheta}_{+,p,a}(h)] \to \boldsymbol{\vartheta}_{+,p,a}$. ∎

4.3 Conditional Variance

We characterize the exact, fixed-n (conditional) variance formulas of the standard RD estimator $\hat{\tau}_{Y,\nu}(\mathbf{h})$. These terms are $\mathbb{V}[\hat{\boldsymbol{\beta}}_{Y-,p}(h)|\mathbf{X}]$ and $\mathbb{V}[\hat{\boldsymbol{\beta}}_{Y+,p}(h)|\mathbf{X}]$.

Lemma SA-3 *Let Assumptions SA-1 and SA-2 hold. If* $nh \to \infty$ *and* $h \to 0$*, then*

$$\begin{aligned}
\mathbb{V}[\hat{\boldsymbol{\beta}}_{Y-,p}(h)|\mathbf{X}] &= \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\boldsymbol{\Sigma}_{Y-}\mathbf{K}_-(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{H}_p^{-1}(h)/n^2 \\
&= \frac{1}{nh}\mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)\boldsymbol{\Sigma}_{Y-}\mathbf{P}_{-,p}(h)'\mathbf{H}_p^{-1}(h),
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}[\hat{\boldsymbol{\beta}}_{Y+,p}(h)|\mathbf{X}] &= \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)\boldsymbol{\Sigma}_{Y+}\mathbf{K}_+(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{H}_p^{-1}(h)/n^2 \\
&= \frac{1}{nh}\mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)\boldsymbol{\Sigma}_{Y+}\mathbf{P}_{+,p}(h)'\mathbf{H}_p^{-1}(h),
\end{aligned}$$

with

$$nh\mathbf{H}_p(h)\mathbb{V}[\hat{\boldsymbol{\beta}}_{Y-,p}(h)|\mathbf{X}]\mathbf{H}_p(h) \to_{\mathbb{P}} \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Psi}_{Y-,p}\boldsymbol{\Gamma}_{-,p}^{-1},$$

$$nh\mathbf{H}_p(h)\mathbb{V}[\hat{\boldsymbol{\beta}}_{Y+,p}(h)|\mathbf{X}]\mathbf{H}_p(h) \to_{\mathbb{P}} \boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Psi}_{Y+,p}\boldsymbol{\Gamma}_{+,p}^{-1},$$

and where

$$\boldsymbol{\Psi}_{Y-,p} = f\sigma_{Y-}^2\int_{-\infty}^{0} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2du, \qquad \boldsymbol{\Psi}_{Y+,p} = f\sigma_{Y+}^2\int_{0}^{\infty} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2du,$$

$\sigma_{Y-}^2 = \sigma_{Y-}^2(\bar{x})$, $\sigma_{Y+}^2 = \sigma_{Y+}^2(\bar{x})$ *and* $f = f(\bar{x})$.

Proof of Lemma SA-3. The first two equalities follow directly. Lemma SA-1 gives $\boldsymbol{\Gamma}_{-,p}^{-1}(h) = \boldsymbol{\Gamma}_{-,p}^{-1} + o_{\mathbb{P}}(1)$ and $\boldsymbol{\Gamma}_{+,p}^{-1}(h) = \boldsymbol{\Gamma}_{+,p}^{-1} + o_{\mathbb{P}}(1)$. Set

$$\boldsymbol{\Psi}_{Y-,p}(h) = h\mathbf{R}_p(h)'\mathbf{K}_-(h)\boldsymbol{\Upsilon}_{Y-}\mathbf{K}_-(h)\mathbf{R}_p(h)/n$$

and

$$\boldsymbol{\Psi}_{Y+,p}(h) = h\mathbf{R}_p(h)'\mathbf{K}_+(h)\boldsymbol{\Upsilon}_{Y+}\mathbf{K}_+(h)\mathbf{R}_p(h)/n,$$

and, by proceeding as before, we have $\boldsymbol{\Psi}_{Y-,p}(h) = \mathbb{E}[\boldsymbol{\Psi}_{Y-,p}(h)] + o_{\mathbb{P}}(1)$ and $\boldsymbol{\Psi}_{Y+,p}(h) = \mathbb{E}[\boldsymbol{\Psi}_{Y+,p}(h)] + o_{\mathbb{P}}(1)$, and also $\mathbb{E}[\boldsymbol{\Psi}_{Y-,p}(h)] \to \boldsymbol{\Psi}_{Y-,p}$ and $\mathbb{E}[\boldsymbol{\Psi}_{Y+,p}(h)] \to \boldsymbol{\Psi}_{Y+,p}$, by changing variables and taking limits. ∎

5 Covariates Sharp RD

We also employ repeatedly properties and results for sharp RD regressions for the covariates $\mathbf{Z}_i$. Define

$$\boldsymbol{\mu}_{Z-}^{(\nu)} = \boldsymbol{\mu}_{Z-}^{(\nu)}(\bar{x}) = \frac{\partial^\nu}{\partial x^\nu}\boldsymbol{\mu}_{Z-}(x)\bigg|_{x=\bar{x}}, \qquad \boldsymbol{\mu}_{Z+}^{(\nu)} = \boldsymbol{\mu}_{Z+}^{(\nu)}(\bar{x}) = \frac{\partial^\nu}{\partial x^\nu}\boldsymbol{\mu}_{Z+}(x)\bigg|_{x=\bar{x}},$$

and where, with exactly the same notation logic as above,

$$\boldsymbol{\mu}_{Z-}^{(\nu)\prime} = \nu!\mathbf{e}_\nu'\boldsymbol{\beta}_{Z-}, \qquad \boldsymbol{\beta}_{Z-} = [\boldsymbol{\beta}_{Z_1-,p}\ ,\ \boldsymbol{\beta}_{Z_2-,p}\ ,\ \cdots\ ,\ \boldsymbol{\beta}_{Z_d-,p}]_{(1+p)\times d},$$

$$\boldsymbol{\mu}_{Z+}^{(\nu)\prime} = \nu!\mathbf{e}_\nu'\boldsymbol{\beta}_{Z+}, \qquad \boldsymbol{\beta}_{Z+} = [\boldsymbol{\beta}_{Z_1+,p}\ ,\ \boldsymbol{\beta}_{Z_2+,p}\ ,\ \cdots\ ,\ \boldsymbol{\beta}_{Z_d+,p}]_{(1+p)\times d},$$

$$\boldsymbol{\beta}_{Z_\ell-,p} = \boldsymbol{\beta}_{Z_\ell-,p}(\bar{x}) = \left[\mu_{Z_\ell-}\ ,\ \frac{\mu_{Z_\ell-}^{(1)}}{1!}\ ,\ \frac{\mu_{Z_\ell-}^{(2)}}{2!}\ ,\ \cdots\ ,\ \frac{\mu_{Z_\ell-}^{(p)}}{p!}\right]',$$

$$\boldsymbol{\beta}_{Z_\ell+,p} = \boldsymbol{\beta}_{Z_\ell+,p}(\bar{x}) = \left[\mu_{Z_\ell+}\ ,\ \frac{\mu_{Z_\ell+}^{(1)}}{1!}\ ,\ \frac{\mu_{Z_\ell+}^{(2)}}{2!}\ ,\ \cdots\ ,\ \frac{\mu_{Z_\ell+}^{(p)}}{p!}\right]',$$

$\mu_{Z_\ell-} = \mu_{Z_\ell+}(\bar{x}) = \mu_{Z_\ell-}^{(0)} = \mu_{Z_\ell+}^{(0)}(\bar{x})$ and $\mu_{Z_\ell+} = \mu_{Z_\ell+}(\bar{x}) = \mu_{Z_\ell+}^{(0)} = \mu_{Z_\ell+}^{(0)}(\bar{x})$, for $\ell = 1, 2, \cdots, d$.

Therefore, following the same notation as in the standard sharp RD, we introduce the sharp RD estimators:

$$\hat{\boldsymbol{\beta}}_{Z-,p}(h) = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\Upsilon}_{Z-,p}(h), \qquad \boldsymbol{\Upsilon}_{Z-,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{Z}/n,$$

$$\hat{\boldsymbol{\beta}}_{Z+,p}(h) = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\Upsilon}_{Z+,p}(h), \qquad \boldsymbol{\Upsilon}_{Z+,p}(h) = \mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbf{Z}/n.$$

Observe that

$$\hat{\boldsymbol{\beta}}_{Z-,p}(h) = [\hat{\boldsymbol{\beta}}_{Z_1-,p}(h)\ ,\ \hat{\boldsymbol{\beta}}_{Z_2-,p}(h)\ ,\ \cdots\ ,\ \hat{\boldsymbol{\beta}}_{Z_d-,p}(h)]_{(1+p)\times d},$$

$$\hat{\boldsymbol{\beta}}_{Z+,p}(h) = [\hat{\boldsymbol{\beta}}_{Z_1+,p}(h)\ ,\ \hat{\boldsymbol{\beta}}_{Z_2+,p}(h)\ ,\ \cdots\ ,\ \hat{\boldsymbol{\beta}}_{Z_d+,p}(h)]_{(1+p)\times d},$$

which are simply the least-square coefficients from a multivariate regression, that is, $\hat{\boldsymbol{\beta}}_{Z_\ell-,p}(h)$ and

$\hat{\boldsymbol{\beta}}_{Z_\ell+,p}(h)$ are $((1+p)\times 1)$ vectors given by

$$\hat{\boldsymbol{\beta}}_{Z_\ell-,p}(h) = \underset{\mathbf{b}\in\mathbb{R}^{1+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}(X_i < \bar{x})(Z_{i\ell} - \mathbf{r}_p(X_i-\bar{x})'\mathbf{b})^2 k_h(-(X_i-\bar{x})),$$

$$\hat{\boldsymbol{\beta}}_{Z_\ell+,p}(h) = \underset{\mathbf{b}\in\mathbb{R}^{1+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}(X_i \geq \bar{x})(Z_{i\ell} - \mathbf{r}_p(X_i-\bar{x})'\mathbf{b})^2 k_h(X_i-\bar{x}),$$

for $\ell = 1, 2, \cdots, d$.

Note that

$$\hat{\boldsymbol{\beta}}_{Z-,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)\mathbf{Z}, \qquad \hat{\boldsymbol{\beta}}_{Z+,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)\mathbf{Z},$$

or, in vectorized form,

$$\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z-,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)]\operatorname{vec}(\mathbf{Z}),$$

$$\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z+,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)]\operatorname{vec}(\mathbf{Z}),$$

using $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$ (for conformable matrices A, B and C).

Finally, the (placebo) RD treatment effect estimator for the additional covariates is

$$\hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h}) = \hat{\boldsymbol{\mu}}^{(\nu)}_{Z+,p}(h_+) - \hat{\boldsymbol{\mu}}^{(\nu)}_{Z-,p}(h_-)$$

with

$$\hat{\boldsymbol{\mu}}^{(\nu)}_{Z-,p}(h)' = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{Z-,p}(h), \qquad \hat{\boldsymbol{\mu}}^{(\nu)}_{Z+,p}(h)' = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{Z+,p}(h).$$

5.1 Conditional Bias

We characterize the smoothing bias of the standard RD estimators using the additional covariates as outcomes. We have

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{Z-,p}(h)|\mathbf{X}] = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbb{E}[\mathbf{Z}(0)|\mathbf{X}]/n,$$

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{Z+,p}(h)|\mathbf{X}] = \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbb{E}[\mathbf{Z}(1)|\mathbf{X}]/n.$$

Lemma SA-4 *Let Assumptions SA-1, SA-2 and SA-3 hold with $\varrho \geq p+2$. If $nh \to \infty$ and $h \to 0$, then*

$$\mathbb{E}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z-,p}(h))|\mathbf{X}] = \operatorname{vec}(\boldsymbol{\beta}_{Z-,p}) + [\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)]\left[h^{1+p}\mathbf{B}_{Z-,p,p}(h) + h^{2+p}\mathbf{B}_{Z-,p,1+p}(h) + o_{\mathbb{P}}(h^{2+p})\right],$$

$$\mathbb{E}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z+,p}(h))|\mathbf{X}] = \operatorname{vec}(\boldsymbol{\beta}_{Z+,p}) + [\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)]\left[h^{1+p}\mathbf{B}_{Z+,p,p}(h) + h^{2+p}\mathbf{B}_{Z+,p,1+p}(h) + o_{\mathbb{P}}(h^{2+p})\right],$$

where

$$\mathbf{B}_{Z-,p,a}(h) = [\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,a}(h)]\frac{\boldsymbol{\mu}^{(1+a)}_{Z-}}{(1+a)!} \to_{\mathbb{P}} \mathbf{B}_{Z-,p,a} = [\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{-,p}\boldsymbol{\vartheta}_{-,p,a}]\frac{\boldsymbol{\mu}^{(1+a)}_{Z-}}{(1+a)!},$$

$$\mathbf{B}_{Z+,p,a}(h) = [\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p,a}(h)]\frac{\boldsymbol{\mu}^{(1+a)}_{Z+}}{(1+a)!} \to_{\mathbb{P}} \mathbf{B}_{Z+,p,a} = [\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{+,p}\boldsymbol{\vartheta}_{+,p,a}]\frac{\boldsymbol{\mu}^{(1+a)}_{Z+}}{(1+a)!},$$

$\boldsymbol{\mu}^{(1+p)}_{Z-} = \boldsymbol{\mu}^{(1+p)}_{Z-}(\bar{x})$ *and* $\boldsymbol{\mu}^{(1+p)}_{Z+} = \boldsymbol{\mu}^{(1+p)}_{Z+}(\bar{x})$.

Proof of Lemma SA-4. The proof is analogous to the one of Lemma SA-2. We only prove the left-side case to save space. First, a Taylor series expansion of $\boldsymbol{\mu}_{Z-}(x)$ at $x = \bar{x}$ gives

$$\begin{aligned}
&\mathbb{E}[\hat{\boldsymbol{\beta}}_{Z-,p}(h)|\mathbf{X}]\\
&= \mathbf{H}^{-1}_p(h)\boldsymbol{\Gamma}^{-1}_{-,p}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\boldsymbol{\mu}_{Z-}(\mathbf{X})\\
&= \boldsymbol{\beta}_{Z-,p} + \mathbf{H}^{-1}_p(h)\left[h^{1+p}\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,p}(h)\frac{\boldsymbol{\mu}^{(1+p)\prime}_{Z-}}{(1+p)!} + h^{2+p}\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,p+1}(h)\frac{\boldsymbol{\mu}^{(2+p)\prime}_{Z-}}{(2+p)!} + o_{\mathbb{P}}(h^{2+p})\right],
\end{aligned}$$

and similarly for $\mathbb{E}[\hat{\boldsymbol{\beta}}_{Z+,p}(h)|\mathbf{X}]$. Second, note that

$$\operatorname{vec}\left(\mathbf{H}^{-1}_p(h)\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,a}(h)\frac{\boldsymbol{\mu}^{(1+a)\prime}_{Z-}}{(1+a)!}\right) = [\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,a}(h)]\frac{\boldsymbol{\mu}^{(1+a)}_{Z-}}{(1+a)!},$$

where $\operatorname{vec}(\boldsymbol{\mu}^{(1+a)\prime}_{Z-}) = \boldsymbol{\mu}^{(1+a)}_{Z-}$ and $[\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,a}(h)] = [\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)][\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p,a}(h)]$. The rest follows directly, as in Lemma SA-2. ∎

5.2 Conditional Variance

We characterize the exact, fixed-n (conditional) variance formulas of the standard RD estimators using the additional covariates as outcomes. These terms are $\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z-,p}(h))|\mathbf{X}]$ and $\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z+,p}(h))|\mathbf{X}]$.

Lemma SA-5 *Let Assumptions SA-1, SA-2 and SA-3 hold. If $nh \to \infty$ and $h \to 0$, then*

$$\begin{aligned}
\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z-,p}(h))|\mathbf{X}] &= [\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)\boldsymbol{\Gamma}^{-1}_{-,p}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)]\boldsymbol{\Sigma}_{Z-}[\mathbf{I}_d \otimes \mathbf{K}_-(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}^{-1}_{-,p}(h)\mathbf{H}^{-1}_p(h)]/n^2\\
&= \frac{1}{nh}[\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)][\mathbf{I}_d \otimes \mathbf{P}_{-,p}(h)]\boldsymbol{\Sigma}_{Z-}[\mathbf{I}_d \otimes \mathbf{P}_{-,p}(h)'][\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)],
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z+,p}(h))|\mathbf{X}] &= [\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)\boldsymbol{\Gamma}^{-1}_{+,p}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)]\boldsymbol{\Sigma}_{Z+}[\mathbf{I}_d \otimes \mathbf{K}_+(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}^{-1}_{+,p}(h)\mathbf{H}^{-1}_p(h)]/n^2\\
&= \frac{1}{nh}[\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)][\mathbf{I}_d \otimes \mathbf{P}_{+,p}(h)]\boldsymbol{\Sigma}_{Z+}[\mathbf{I}_d \otimes \mathbf{P}_{+,p}(h)'][\mathbf{I}_d \otimes \mathbf{H}^{-1}_p(h)],
\end{aligned}$$

with

$$nh[\mathbf{I}_d \otimes \mathbf{H}_p(h)]\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z-,p}(h))|\mathbf{X}][\mathbf{I}_d \otimes \mathbf{H}_p(h)] \to_{\mathbb{P}} [\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{-,p}]\boldsymbol{\Psi}_{Z-,p}[\mathbf{I}_d \otimes \boldsymbol{\Gamma}^{-1}_{-,p}],$$

$$nh[\mathbf{I}_d \otimes \mathbf{H}_p(h)]\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z+,p}(h))|\mathbf{X}][\mathbf{I}_d \otimes \mathbf{H}_p(h)] \to_{\mathbb{P}} [\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{+,p}^{-1}]\boldsymbol{\Psi}_{Z+,p}[\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{+,p}^{-1}],$$

where

$$\boldsymbol{\Psi}_{Z-,p} = f(\bar{x})\left[\boldsymbol{\sigma}_{Z-}^2 \otimes \int_{-\infty}^{0} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2 du\right], \qquad \boldsymbol{\sigma}_{Z-}^2 = \boldsymbol{\sigma}_{Z-}^2(\bar{x}) = \mathbb{V}[\mathbf{Z}_i(0)|X_i = \bar{x}],$$

and

$$\boldsymbol{\Psi}_{Z+,p} = f(\bar{x})\left[\boldsymbol{\sigma}_{Z+}^2 \otimes \int_{0}^{\infty} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2 du\right], \qquad \boldsymbol{\sigma}_{Z+}^2 = \boldsymbol{\sigma}_{Z+}^2(\bar{x}) = \mathbb{V}[\mathbf{Z}_i(1)|X_i = \bar{x}].$$

Proof of Lemma SA-5. We have

$$\begin{aligned}
\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z-,p}(h)) &= [\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)]\operatorname{vec}(\mathbf{Z}(0)) \\
&= [\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)][\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{-,p}^{-1}(h)][\mathbf{I}_d \otimes \mathbf{R}_p(h)'\mathbf{K}_-(h)]\operatorname{vec}(\mathbf{Z}(0))
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{vec}(\hat{\boldsymbol{\beta}}_{Z+,p}(h)) &= [\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)]\operatorname{vec}(\mathbf{Z}(1)) \\
&= [\mathbf{I}_d \otimes \mathbf{H}_p^{-1}(h)][\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{+,p}^{-1}(h)][\mathbf{I}_d \otimes \mathbf{R}_p(h)'\mathbf{K}_+(h)]\operatorname{vec}(\mathbf{Z}(1))
\end{aligned}$$

and thus the first two equalities follow directly. Lemma SA-1 gives $\boldsymbol{\Gamma}_{-,p}^{-1}(h) = \boldsymbol{\Gamma}_{-,p}^{-1} + o_{\mathbb{P}}(1)$ and $\boldsymbol{\Gamma}_{+,p}^{-1}(h) = \boldsymbol{\Gamma}_{+,p}^{-1} + o_{\mathbb{P}}(1)$. Set

$$\boldsymbol{\Psi}_{Z-,p}(h) = h[\mathbf{I}_d \otimes \mathbf{R}_p(h)'\mathbf{K}_-(h)]\boldsymbol{\Sigma}_{Z-}[\mathbf{I}_d \otimes \mathbf{K}_-(h)\mathbf{R}_p(h)]/n$$

and

$$\boldsymbol{\Psi}_{Z+,p}(h) = h[\mathbf{I}_d \otimes \mathbf{R}_p(h)'\mathbf{K}_+(h)]\boldsymbol{\Sigma}_{Z+}[\mathbf{I}_d \otimes \mathbf{K}_+(h)\mathbf{R}_p(h)]/n$$

and by proceeding as before we have $\boldsymbol{\Psi}_{Z-,p}(h) = \mathbb{E}[\boldsymbol{\Psi}_{Z-,p}(h)] + o_{\mathbb{P}}(1)$ and $\boldsymbol{\Psi}_{Z+,p}(h) = \mathbb{E}[\boldsymbol{\Psi}_{Z+,p}(h)] + o_{\mathbb{P}}(1)$, and by changing variables and taking limits we obtain $\mathbb{E}[\boldsymbol{\Psi}_{Z-,p}(h)] \to \boldsymbol{\Psi}_{Z-,p}$ and $\mathbb{E}[\boldsymbol{\Psi}_{Z+,p}(h)] \to \boldsymbol{\Psi}_{Z+,p}$. ∎

Finally, observe that

$$\begin{aligned}
[\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{-,p}^{-1}]\boldsymbol{\Psi}_{Z-,p}[\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{-,p}^{-1}] &= \left[\boldsymbol{\sigma}_{Z-}^2(\bar{x}) \otimes f(\bar{x})\boldsymbol{\Gamma}_{-,p}^{-1}\left(\int_{-\infty}^{0} \mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2 du\right)\boldsymbol{\Gamma}_{-,p}^{-1}\right] \\
&= \frac{\boldsymbol{\sigma}_{Z-}^2(\bar{x})}{\sigma_{Y-}^2(\bar{x})} \otimes \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Psi}_{Y-,p}\boldsymbol{\Gamma}_{-,p}^{-1},
\end{aligned}$$

and similarly

$$[\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{+,p}^{-1}]\boldsymbol{\Psi}_{Z+,p}[\mathbf{I}_d \otimes \boldsymbol{\Gamma}_{+,p}^{-1}] = \frac{\boldsymbol{\sigma}_{Z+}^2(\bar{x})}{\sigma_{Y+}^2(\bar{x})} \otimes \boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Psi}_{Y+,p}\boldsymbol{\Gamma}_{+,p}^{-1}.$$

6 Covariate-Adjusted Sharp RD

For $\nu \leq p$, the Covariate-Adjusted sharp RD estimator implemented with bandwidths $\mathbf{h} = (h_-, h_+)$ is:

$$\tilde{\tau}_{Y,\nu}(\mathbf{h}) = \nu! \mathbf{e}'_{2+p+\nu} \tilde{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}) - \nu! \mathbf{e}'_{\nu} \tilde{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}),$$

$$\tilde{\boldsymbol{\theta}}_{Y,p}(\mathbf{h}) = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}) \\ \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \end{bmatrix}, \qquad \tilde{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}) \in \mathbb{R}^{2+2p}, \qquad \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \in \mathbb{R}^{d},$$

$$\tilde{\boldsymbol{\theta}}_{Y,p}(\mathbf{h}) = \underset{\mathbf{b}_-, \mathbf{b}_+, \boldsymbol{\gamma}}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - \mathbf{r}_{-,p}(X_i - \bar{x})' \mathbf{b}_- - \mathbf{r}_{+,p}(X_i - \bar{x})' \mathbf{b}_+ - \mathbf{Z}'_i \boldsymbol{\gamma})^2 K_{\mathbf{h}}(X_i - \bar{x}),$$

where $\mathbf{b}_- \in \mathbb{R}^{1+p}$, $\mathbf{b}_+ \in \mathbb{R}^{1+p}$, $\boldsymbol{\gamma} \in \mathbb{R}^d$, and

$$\mathbf{r}_{-,p}(u) := \mathbb{1}(u < 0) \mathbf{r}_p(u), \qquad \mathbf{r}_{+,p}(u) := \mathbb{1}(u \geq 0) \mathbf{r}_p(u).$$

Using partitioned regression algebra, we have

$$\tilde{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}) = \hat{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}) - \hat{\boldsymbol{\beta}}_{Z,p}(\mathbf{h}) \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}), \qquad \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) = \tilde{\boldsymbol{\Gamma}}_p^{-1}(\mathbf{h}) \tilde{\boldsymbol{\Upsilon}}_{Y,p}(\mathbf{h}),$$

where

$$\hat{\boldsymbol{\beta}}_{Y,p}(\mathbf{h}) = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{Y-,p}(h_-) \\ \hat{\boldsymbol{\beta}}_{Y+,p}(h_+) \end{bmatrix}_{(2+2p) \times 1}, \qquad \hat{\boldsymbol{\beta}}_{Z,p}(\mathbf{h}) = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{Z-,p}(h_-) \\ \hat{\boldsymbol{\beta}}_{Z+,p}(h_+) \end{bmatrix}_{(2+2p) \times d},$$

and

$$\begin{aligned} \tilde{\boldsymbol{\Gamma}}_p(\mathbf{h}) &= \mathbf{Z}' \mathbf{K}_-(h_-) \mathbf{Z}/n - \boldsymbol{\Upsilon}_{Z-,p}(h_-)' \boldsymbol{\Gamma}_{-,p}^{-1}(h) \boldsymbol{\Upsilon}_{Z-,p}(h_-) \\ &\quad + \mathbf{Z}' \mathbf{K}_+(h_+) \mathbf{Z}/n - \boldsymbol{\Upsilon}_{Z+,p}(h_+)' \boldsymbol{\Gamma}_{+,p}^{-1}(h_+) \boldsymbol{\Upsilon}_{Z+,p}(h_+), \end{aligned}$$

$$\begin{aligned} \tilde{\boldsymbol{\Upsilon}}_{Y,p}(\mathbf{h}) &= \mathbf{Z}' \mathbf{K}_-(h_-) \mathbf{Y}/n - \boldsymbol{\Upsilon}_{Z-,p}(h_-)' \boldsymbol{\Gamma}_{-,p}^{-1}(h) \boldsymbol{\Upsilon}_{Y-,p}(h_-) \\ &\quad + \mathbf{Z}' \mathbf{K}_+(h_+) \mathbf{Y}/n - \boldsymbol{\Upsilon}_{Z+,p}(h_+)' \boldsymbol{\Gamma}_{+,p}^{-1}(h_+) \boldsymbol{\Upsilon}_{Y+,p}(h_+). \end{aligned}$$

Therefore, the above representation gives

$$\begin{aligned} \tilde{\tau}_{Y,\nu}(\mathbf{h}) &= \hat{\tau}_{Y,\nu}(\mathbf{h}) - \hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h})' \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \\ &= \tilde{\mu}^{(\nu)}_{Y+,p}(h_+; \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) - \tilde{\mu}^{(\nu)}_{Y-,p}(h_-; \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) \end{aligned}$$

with

$$\tilde{\mu}^{(\nu)}_{Y-,p}(h_-; \boldsymbol{\gamma}) = \hat{\mu}^{(\nu)}_{Y-,p}(h_-) - \hat{\boldsymbol{\mu}}^{(\nu)}_{Z-,p}(h_-)' \boldsymbol{\gamma},$$

$$\tilde{\mu}^{(\nu)}_{Y+,p}(h_-; \boldsymbol{\gamma}) = \hat{\mu}^{(\nu)}_{Y+,p}(h_-) - \hat{\boldsymbol{\mu}}^{(\nu)}_{Z+,p}(h_+)' \boldsymbol{\gamma}.$$

6.1 Hessian Matrix, Invertibility and Consistency

The estimators $\hat{\mu}^{(\nu)}_{Y-,p}(h)$, $\hat{\mu}^{(\nu)}_{Y+,p}(h)$, $\hat{\boldsymbol{\mu}}^{(\nu)}_{Z-,p}(h)$ and $\hat{\boldsymbol{\mu}}^{(\nu)}_{Z+,p}(h)$, are all standard (two-sample) local polynomial estimators without additional covariates, and therefore are well defined, with probability approaching one, if the matrices $\boldsymbol{\Gamma}_{-,p}(h)$ and $\boldsymbol{\Gamma}_{+,p}(h)$ are (asymptotically) invertible. This follows from Lemma SA-1, and conventional results from the local polynomial literature (e.g., Fan and Gijbels (1996)).

Therefore, the covariate-adjusted estimator $\tilde{\tau}_{Y,\nu}(\mathbf{h})$ will be well defined in large samples, provided $\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})$ is well defined. The following lemma gives the probability limit of the two components of $\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})$.

Lemma SA-6 *Let Assumptions SA-1, SA-2 and SA-3 hold. If* $n\min\{h_-,h_+\}\to\infty$ *and* $\max\{h_-,h_+\}\to 0$, *then*

$$\tilde{\boldsymbol{\Gamma}}_p(\mathbf{h}) = \kappa\left(\boldsymbol{\sigma}^2_{Z-}+\boldsymbol{\sigma}^2_{Z+}\right)+o_{\mathbb{P}}(1),$$

and

$$\tilde{\boldsymbol{\Upsilon}}_{Y,p}(\mathbf{h}) = \kappa(\mathbb{E}[(\mathbf{Z}_i(0)-\boldsymbol{\mu}_{Z-}(X_i))Y_i(0)|X_i=\bar{x}]+\mathbb{E}[(\mathbf{Z}_i(1)-\boldsymbol{\mu}_{Z+}(X_i))Y_i(1)|X_i=\bar{x}])+o_{\mathbb{P}}(1),$$

where

$$\kappa = f\int_{-\infty}^{0}K(u)du = f\int_{0}^{\infty}K(u)du,$$

$\boldsymbol{\mu}_{Z-}=\boldsymbol{\mu}_{Z-}(\bar{x})$, $\boldsymbol{\mu}_{Z+}=\boldsymbol{\mu}_{Z+}(\bar{x})$, $\boldsymbol{\sigma}^2_{Z-}=\boldsymbol{\sigma}^2_{Z-}(\bar{x})$, $\boldsymbol{\sigma}^2_{Z+}=\boldsymbol{\sigma}^2_{Z+}(\bar{x})$, *and* $f=f(\bar{x})$.

Proof of Lemma SA-6. Analogous to the proof of Lemma SA-1, which gives $\boldsymbol{\Gamma}^{-1}_{-,p}(h) = (\mathbb{E}[\boldsymbol{\Gamma}_{-,p}(h)])^{-1}+o_{\mathbb{P}}(1) = \boldsymbol{\Gamma}^{-1}_{-,p}+o_{\mathbb{P}}(1)$ and $\boldsymbol{\Gamma}^{-1}_{+,p}(h) = (\mathbb{E}[\boldsymbol{\Gamma}_{+,p}(h)])^{-1}+o_{\mathbb{P}}(1) = \boldsymbol{\Gamma}^{-1}_{+,p}+o_{\mathbb{P}}(1)$. In particular, Markov inequality implies $\mathbf{Z}'\mathbf{K}_-(h_-)\mathbf{Z}/n = \mathbb{E}[\mathbf{Z}'\mathbf{K}_-(h_-)\mathbf{Z}/n]+o_{\mathbb{P}}(1)$, $\boldsymbol{\Upsilon}_{Z-,p}(h_-) = \mathbb{E}[\boldsymbol{\Upsilon}_{Z-,p}(h_-)]+o_{\mathbb{P}}(1)$, $\boldsymbol{\Upsilon}_{Z-,p}(h_-) = \mathbb{E}[\boldsymbol{\Upsilon}_{Z-,p}(h_-)]+o_{\mathbb{P}}(1)$, $\mathbf{Z}'\mathbf{K}_+(h_+)\mathbf{Z}/n = \mathbb{E}[\mathbf{Z}'\mathbf{K}_+(h_+)\mathbf{Z}/n]+o_{\mathbb{P}}(1)$, $\boldsymbol{\Upsilon}_{Z+,p}(h_+) = \mathbb{E}[\boldsymbol{\Upsilon}_{Z+,p}(h_+)]+o_{\mathbb{P}}(1)$, $\boldsymbol{\Upsilon}_{Z+,p}(h_+) = \mathbb{E}[\boldsymbol{\Upsilon}_{Z+,p}(h_+)]+o_{\mathbb{P}}(1)$. Next, changing variables and taking limits as $h\to 0$,

$$\begin{aligned}\tilde{\boldsymbol{\Gamma}}_p(\mathbf{h}) &= \mathbb{E}[\mathbf{Z}_i(0)\mathbf{Z}_i(0)'|X_i=\bar{x}]+\mathbb{E}[\mathbf{Z}_i(1)\mathbf{Z}_i(1)'|X_i=\bar{x}]\\ &\quad -\boldsymbol{\mu}_{Z-}\boldsymbol{\kappa}'_{-,p}\boldsymbol{\Gamma}_{-,p}\boldsymbol{\kappa}_{-,p}\boldsymbol{\mu}'_{Z-}-\boldsymbol{\mu}_{Z+}\boldsymbol{\kappa}'_{+,p}\boldsymbol{\Gamma}_{+,p}\boldsymbol{\kappa}_{+,p}\boldsymbol{\mu}'_{Z+},\\ &= \kappa\mathbb{V}[\mathbf{Z}_i(0)|X_i=\bar{x}]+\kappa\mathbb{V}[\mathbf{Z}_i(1)|X_i=\bar{x}] = \kappa\left(\boldsymbol{\sigma}^2_{Z-}+\boldsymbol{\sigma}^2_{Z+}\right),\end{aligned}$$

where $\boldsymbol{\kappa}_{-,p} = f\int_{-\infty}^{0}\mathbf{r}_p(u)K(u)du$, $\boldsymbol{\kappa}_{+,p} = f\int_{0}^{\infty}\mathbf{r}_p(u)K(u)du$ and $\kappa = \mathbf{e}'_0\boldsymbol{\kappa}_{-,p} = \mathbf{e}'_0\boldsymbol{\kappa}_{+,p}$, and because $\boldsymbol{\Gamma}_{-,p}\mathbf{e}_0 = \boldsymbol{\kappa}_{-,p}$ and $\boldsymbol{\Gamma}_{+,p}\mathbf{e}_0 = \boldsymbol{\kappa}_{+,p}$ and hence $\boldsymbol{\Gamma}^{-1}_{-,p}\boldsymbol{\kappa}_{-,p} = \mathbf{e}_0$, $\boldsymbol{\Gamma}^{-1}_{+,p}\boldsymbol{\kappa}_{+,p} = \mathbf{e}_0$ and $\boldsymbol{\kappa}'_{-,p}\boldsymbol{\Gamma}^{-1}_{-,p}\boldsymbol{\kappa}_{-,p} = \kappa = \boldsymbol{\kappa}'_{+,p}\boldsymbol{\Gamma}^{-1}_{+,p}\boldsymbol{\kappa}_{+,p}$.

The second result is proved using the same arguments. ∎

The previous lemma shows that $\tilde{\boldsymbol{\Gamma}}_p(\mathbf{h})$ is asymptotically invertible, given our assumptions, and hence the covariate-adjusted sharp RD estimator $\tilde{\tau}_{Y,\nu}(\mathbf{h}) = \hat{\tau}_{Y,\nu}(\mathbf{h})-\hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h})\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})$ is well defined in large samples. Moreover, because $\hat{\tau}_{Y,\nu}(\mathbf{h}) \to_{\mathbb{P}} \tau_{Y,\nu}$ by Lemmas SA-2 and SA-3, $\hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h}) \to_{\mathbb{P}} \boldsymbol{\tau}_{Z,\nu}$ by Lemmas SA-4 and SA-5, under the conditions of Lemma SA-6, we also obtain the following lemma.

Lemma SA-7 *Let Assumptions SA-1, SA-2 and SA-3 hold with $\varrho \geq p$. If $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$ and $\max\{h_-, h_+\} \to 0$, then*

$$\tilde{\tau}_{Y,\nu}(h) \to_{\mathbb{P}} \tau_{Y,\nu} - \left[\boldsymbol{\mu}_{Z+}^{(\nu)} - \boldsymbol{\mu}_{Z-}^{(\nu)}\right]' \boldsymbol{\gamma}_Y,$$

with

$$\boldsymbol{\gamma}_Y = \left[\boldsymbol{\sigma}_{Z-}^2 + \boldsymbol{\sigma}_{Z+}^2\right]^{-1} \left[\mathbb{E}[(\mathbf{Z}_i(0) - \boldsymbol{\mu}_{Z-}(X_i))Y_i(0)|X_i = \bar{x}] + \mathbb{E}[(\mathbf{Z}_i(1) - \boldsymbol{\mu}_{Z+}(X_i))Y_i(1)|X_i = \bar{x}]\right],$$

where recall that $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z-}(\bar{x})$, $\boldsymbol{\mu}_{Z+} = \boldsymbol{\mu}_{Z+}(\bar{x})$, $\boldsymbol{\sigma}_{Z-}^2 = \boldsymbol{\sigma}_{Z-}^2(\bar{x})$, and $\boldsymbol{\sigma}_{Z+}^2 = \boldsymbol{\sigma}_{Z+}^2(\bar{x})$.

Proof of Lemma SA-7. Follows directly from Lemmas SA-2–SA-6. ∎

6.2 Linear Representation

Using the fixed-n representation

$$\begin{aligned} \tilde{\tau}_{Y,\nu}(\mathbf{h}) &= \hat{\tau}_{Y,\nu}(\mathbf{h}) - \hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h})'\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \\ &= \tilde{\mu}_{Y+,p}^{(\nu)}(h_+; \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) - \tilde{\mu}_{Y-,p}^{(\nu)}(h_-; \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) \end{aligned}$$

with

$$\tilde{\mu}_{Y-,p}^{(\nu)}(h; \boldsymbol{\gamma}) = \hat{\mu}_{Y-,p}^{(\nu)}(h) - \hat{\boldsymbol{\mu}}_{Z-,p}^{(\nu)}(h)'\boldsymbol{\gamma},$$

$$\tilde{\mu}_{Y+,p}^{(\nu)}(h; \boldsymbol{\gamma}) = \hat{\mu}_{Y+,p}^{(\nu)}(h) - \hat{\boldsymbol{\mu}}_{Z+,p}^{(\nu)}(h)'\boldsymbol{\gamma},$$

we have

$$\tilde{\tau}_{Y,\nu}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})' \operatorname{vec}(\hat{\boldsymbol{\beta}}_{S,p}(\mathbf{h})),$$

where

$$\mathbf{s}_{S,\nu}(\mathbf{h}) = \begin{bmatrix} \nu!\mathbf{e}_\nu \\ -\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \otimes \nu!\mathbf{e}_\nu \end{bmatrix} = \begin{bmatrix} 1 \\ -\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \end{bmatrix} \otimes \nu!\mathbf{e}_\nu,$$

$$\hat{\boldsymbol{\beta}}_{S,p}(\mathbf{h}) = \hat{\boldsymbol{\beta}}_{S+,p}(h_+) - \hat{\boldsymbol{\beta}}_{S-,p}(h_-),$$

with

$$\hat{\boldsymbol{\beta}}_{S-,p}(h) = [\hat{\boldsymbol{\beta}}_{Y-,p}(h), \hat{\boldsymbol{\beta}}_{Z-,p}(h)], \qquad \hat{\boldsymbol{\beta}}_{S+,p}(h) = [\hat{\boldsymbol{\beta}}_{Y+,p}(h), \hat{\boldsymbol{\beta}}_{Z+,p}(h)].$$

Recall that $\hat{\boldsymbol{\beta}}_{Y-,p}(h)$, $\hat{\boldsymbol{\beta}}_{Y+,p}(h)$, $\hat{\boldsymbol{\beta}}_{Z-,p}(h)$ and $\hat{\boldsymbol{\beta}}_{Z+,p}(h)$ denote the one-sided RD regressions discussed previously.

Furthermore, note that

$$\hat{\boldsymbol{\beta}}_{S-,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)\mathbf{S}, \qquad \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S-,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)]\mathbf{S},$$

$$\hat{\boldsymbol{\beta}}_{S+,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)\mathbf{S} \qquad \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S+,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)]\mathbf{S}.$$

Finally, by Lemma SA-7, it follows that

$$\mathbf{s}_{S,\nu}(\mathbf{h}) \to_{\mathbb{P}} \mathbf{s}_{S,\nu} = \begin{bmatrix} \nu!\mathbf{e}_\nu \\ -\boldsymbol{\gamma}_Y \otimes \nu!\mathbf{e}_\nu \end{bmatrix},$$

and therefore it is sufficient to study the asymptotic properties of $\hat{\boldsymbol{\beta}}_{S+,p}(h)$ and $\hat{\boldsymbol{\beta}}_{S-,p}(h)$, under the assumption of $\boldsymbol{\tau}_{Z,\nu} = 0$ (pre-intervention covariates). Finally, we also define

$$\boldsymbol{\beta}_{S-,p}(x) = [\boldsymbol{\beta}_{Y-,p}(x), \boldsymbol{\beta}_{Z-,p}(x)], \qquad \boldsymbol{\beta}_{S+,p}(x) = [\boldsymbol{\beta}_{Y+,p}(x), \boldsymbol{\beta}_{Z+,p}(x)],$$

with the notation $\boldsymbol{\beta}_{S-,p} = \boldsymbol{\beta}_{S-,p}(\bar{x})$ and $\boldsymbol{\beta}_{S+,p} = \boldsymbol{\beta}_{S+,p}(\bar{x})$, as above.

7 Inference Results

In this section we study the asymptotic properties of $\tilde{\tau}_{Y,\nu}(\mathbf{h})$. First we derive the bias and variance of the estimator, and then discuss bandwidth selection and distribution theory under the assumption that $\boldsymbol{\tau}_{Z,\nu} = 0$ (pre-intervention covariates). Note that our results do not impose any structure on $\mathbb{E}[Y_i(t)|X_i, \mathbf{Z}_i(t)]$, $t \in \{0, 1\}$, and hence $\tilde{\boldsymbol{\gamma}}_{Y,p}(h)$ has a generic best linear prediction interpretation.

7.1 Conditional Bias

We characterize the smoothing bias of $\hat{\boldsymbol{\beta}}_{S-,p}(h)$ and $\hat{\boldsymbol{\beta}}_{S+,p}(h)$, the main ingredients entering the covariate-adjusted sharp RD estimator $\tilde{\tau}_{Y,\nu}(\mathbf{h})$. Observe that

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{S-,p}(h)|\mathbf{X}] = [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)]\mathbb{E}[\mathbf{S}(0)|\mathbf{X}]/n,$$

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{S+,p}(h)|\mathbf{X}] = [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)]\mathbb{E}[\mathbf{S}(1)|\mathbf{X}]/n.$$

Lemma SA-8 *Let assumptions SA-1, SA-2 and SA-3 hold with $\varrho \geq p+2$, and $nh \to \infty$ and $h \to 0$. Then,*

$$\begin{aligned} &\mathbb{E}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S-,p}(h))|\mathbf{X}] \\ &= \mathrm{vec}(\boldsymbol{\beta}_{S-,p}) + [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)]\left[h^{1+p}\mathbf{B}_{S-,p,p}(h) + h^{2+p}\mathbf{B}_{S-,p,p+1}(h) + o_{\mathbb{P}}(h^{2+p})\right], \end{aligned}$$

$$\begin{aligned}&\mathbb{E}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{S+,p}(h))|\mathbf{X}]\\&=\operatorname{vec}(\boldsymbol{\beta}_{S+,p})+[\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)]\left[h^{1+p}\mathbf{B}_{S+,p,p}(h)+h^{2+p}\mathbf{B}_{S+,p,p+1}(h)+o_{\mathbb{P}}(h^{2+p})\right],\end{aligned}$$

where

$$\mathbf{B}_{S-,p,a}(h)=[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p,a}(h)]\frac{\boldsymbol{\mu}_{S-}^{(1+a)}}{(1+a)!}\to_{\mathbb{P}}\mathbf{B}_{S-,p,a}=[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\vartheta}_{-,p,a}]\frac{\boldsymbol{\mu}_{S-}^{(1+a)}}{(1+a)!},$$

$$\mathbf{B}_{S+,p,a}(h)=[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\vartheta}_{+,p,a}(h)]\frac{\boldsymbol{\mu}_{S+}^{(1+a)}}{(1+a)!}\to_{\mathbb{P}}\mathbf{B}_{S+,p,a}=[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\vartheta}_{+,p,a}]\frac{\boldsymbol{\mu}_{S+}^{(1+a)}}{(1+a)!}.$$

Proof of Lemma SA-8. Follows exactly as in Lemma SA-4 but now using $\mathbf{S}$ instead of $\mathbf{Z}$ as outcome variable. ∎

7.2 Conditional Variance

We characterize the exact, fixed-n (conditional) variance formulas of the main ingredients entering the covariate-adjusted sharp RD estimator $\tilde{\tau}_{Y,\nu}(\mathbf{h})$. These terms are $\mathbb{V}[\hat{\boldsymbol{\beta}}_{S-,p}(h)|\mathbf{X}]$ and $\mathbb{V}[\hat{\boldsymbol{\beta}}_{S+,p}(h)|\mathbf{X}]$.

Lemma SA-9 *Let assumptions SA-1, SA-2 and SA-3 hold, and $nh\to\infty$ and $h\to 0$. Then,*

$$\begin{aligned}&\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{S-,p}(h))|\mathbf{X}]\\&=[\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)]\boldsymbol{\Sigma}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{K}_-(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{H}_p^{-1}(h)]/n^2\\&=\frac{1}{nh}[\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)][\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h)]\boldsymbol{\Sigma}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h)'][\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)],\end{aligned}$$

$$\begin{aligned}&\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{S+,p}(h))|\mathbf{X}]\\&=[\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)]\boldsymbol{\Sigma}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{K}_+(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{H}_p^{-1}(h)]/n^2\\&=\frac{1}{nh}[\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)][\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h)]\boldsymbol{\Sigma}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h)'][\mathbf{I}_{1+d}\otimes\mathbf{H}_p^{-1}(h)],\end{aligned}$$

with

$$nh[\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{S-,p}(h))|\mathbf{X}][\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\to_{\mathbb{P}}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}]\boldsymbol{\Psi}_{S-,p}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}],$$

$$nh[\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\beta}}_{S+,p}(h))|\mathbf{X}][\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\to_{\mathbb{P}}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}]\boldsymbol{\Psi}_{S+,p}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}],$$

where

$$\boldsymbol{\Psi}_{S-,p}=f(\bar{x})\left[\boldsymbol{\sigma}_{S-}^2\otimes\int_{-\infty}^0\mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2du\right],\qquad\boldsymbol{\sigma}_{S-}^2=\boldsymbol{\sigma}_{S-}^2(\bar{x})=\mathbb{V}[\mathbf{S}_i(0)|X_i=\bar{x}],$$

and

$$\mathbf{\Psi}_{S+,p} = f(\bar{x}) \left[\boldsymbol{\sigma}^2_{S+} \otimes \int_0^\infty \mathbf{r}_p(u)\mathbf{r}_p(u)' K(u)^2 du \right], \qquad \boldsymbol{\sigma}^2_{S+} = \boldsymbol{\sigma}^2_{S+}(\bar{x}) = \mathbb{V}[\mathbf{S}_i(1)|X_i = \bar{x}].$$

Proof of Lemma SA-9. Follows exactly as in Lemma SA-5 but now using $\mathbf{S}$ instead of $\mathbf{Z}$ as outcome variable. ∎

7.3 Convergence Rates

In the rest of Part I (Sharp RD designs) of this supplemental appendix, we assume the conditions of Lemmas SA-2–SA-9 hold, unless explicitly noted otherwise.

The results above imply that

$$[\mathbf{I}_{1+d} \otimes \mathbf{H}_p(h)](\hat{\boldsymbol{\beta}}_{S-,p}(h) - \boldsymbol{\beta}_{S-,p}) = O_{\mathbb{P}}\left(h^{1+p} + \frac{1}{\sqrt{nh}}\right),$$

$$[\mathbf{I}_{1+d} \otimes \mathbf{H}_p(h)](\hat{\boldsymbol{\beta}}_{S+,p}(h) - \boldsymbol{\beta}_{S+,p}) = O_{\mathbb{P}}\left(h^{1+p} + \frac{1}{\sqrt{nh}}\right),$$

and therefore, because $\boldsymbol{\mu}^{(\nu)}_{Z-}(\bar{x}) = \boldsymbol{\mu}^{(\nu)}_{Z+}(\bar{x})$ by assumption,

$$\begin{aligned} \tilde{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu} &= \hat{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu} - \hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h})' \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) \\ &= O_{\mathbb{P}}\left(h^{1+p-\nu} + \frac{1}{\sqrt{nh^{1+2\nu}}}\right) = o_{\mathbb{P}}(1). \end{aligned}$$

Furthermore, we have

$$\tilde{\mu}^{(\nu)}_{Y-,p}(h; \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) - \mu^{(\nu)}_{Y-,p}(\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) = O_{\mathbb{P}}\left(h^{1+p-\nu} + \frac{1}{\sqrt{nh^{1+2\nu}}}\right) = o_{\mathbb{P}}(1),$$

$$\tilde{\mu}^{(\nu)}_{Y-,p}(h; \tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) - \mu^{(\nu)}_{Y-,p}(\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) = O_{\mathbb{P}}\left(h^{1+p-\nu} + \frac{1}{\sqrt{nh^{1+2\nu}}}\right) = o_{\mathbb{P}}(1),$$

where

$$\tilde{\mu}^{(\nu)}_{Y-,p}(h; \boldsymbol{\gamma}) = \hat{\mu}^{(\nu)}_{Y-,p}(h) - \hat{\boldsymbol{\mu}}^{(\nu)}_{Z-,p}(h)' \boldsymbol{\gamma}, \qquad \mu^{(\nu)}_{Y-,p}(\boldsymbol{\gamma}) = \mu^{(\nu)}_{Y-,p} - \boldsymbol{\mu}^{(\nu)\prime}_{Z-,p} \boldsymbol{\gamma},$$

$$\tilde{\mu}^{(\nu)}_{Y+,p}(h; \boldsymbol{\gamma}) = \hat{\mu}^{(\nu)}_{Y+,p}(h) - \hat{\boldsymbol{\mu}}^{(\nu)}_{Z+,p}(h) \boldsymbol{\gamma}, \qquad \mu^{(\nu)}_{Y+,p}(\boldsymbol{\gamma}) = \mu^{(\nu)}_{Y+,p} - \boldsymbol{\mu}^{(\nu)\prime}_{Z+,p} \boldsymbol{\gamma}.$$

7.4 Bias Approximation

We give the bias approximations for each of the estimators, under the conditions imposed above (Lemmas SA-1–SA-9).

7.4.1 Standard Sharp RD Estimator

We have

$$\mathbb{E}[\hat{\mu}^{(\nu)}_{Y-,p}(h)|\mathbf{X}] - \mu^{(\nu)}_{Y-} = h^{1+p-\nu}\mathcal{B}_{Y-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

$$\mathbb{E}[\hat{\mu}^{(\nu)}_{Y+,p}(h)|\mathbf{X}] - \mu^{(\nu)}_{Y+} = h^{1+p-\nu}\mathcal{B}_{Y+,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

where

$$\mathcal{B}_{Y-,\nu,p}(h) = \nu!\mathbf{e}'_\nu\mathbf{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p}(h)\frac{\mu^{(1+p)}_{Y-}}{(1+p)!} \to_{\mathbb{P}} \mathcal{B}_{Y-,\nu,p} = \nu!\mathbf{e}'_\nu\mathbf{\Gamma}^{-1}_{-,p}\boldsymbol{\vartheta}_{-,p}\frac{\mu^{(1+p)}_{Y-}}{(1+p)!},$$

$$\mathcal{B}_{Y+,\nu,p}(h) = \nu!\mathbf{e}'_\nu\mathbf{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p}(h)\frac{\mu^{(1+p)}_{Y+}}{(1+p)!} \to_{\mathbb{P}} \mathcal{B}_{Y+,\nu,p} = \nu!\mathbf{e}'_\nu\mathbf{\Gamma}^{-1}_{+,p}\boldsymbol{\vartheta}_{+,p}\frac{\mu^{(1+p)}_{Y+}}{(1+p)!},$$

where we set $\boldsymbol{\vartheta}_{-,p}(h) := \boldsymbol{\vartheta}_{-,p,p}(h)$, $\boldsymbol{\vartheta}_{+,p}(h) := \boldsymbol{\vartheta}_{+,p,p}(h)$, $\boldsymbol{\vartheta}_{-,p} := \boldsymbol{\vartheta}_{-,p,p}$ and $\boldsymbol{\vartheta}_{+,p} := \boldsymbol{\vartheta}_{+,p,p}$ to save notation.

Therefore,

$$\mathbb{E}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}] - \tau_\nu = h_+^{1+p-\nu}\mathcal{B}_{Y+,\nu,p}(h_+) - h_-^{1+p-\nu}\mathcal{B}_{Y-,\nu,p}(h_-) + o_{\mathbb{P}}(\max\{h_-^{2+p-\nu}, h_+^{2+p-\nu}\}).$$

7.4.2 Covariate-Adjusted Sharp RD Estimator

Using the linear approximation, we define

$$\mathsf{Bias}[\tilde{\mu}^{(\nu)}_{Y-,p}(h)] = \mathbb{E}[\mathbf{s}'_{S,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S-,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S-,p})]|\mathbf{X}],$$

$$\mathsf{Bias}[\tilde{\mu}^{(\nu)}_{Y+,p}(h)] = \mathbb{E}[\mathbf{s}'_{S,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S+,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S+,p})]|\mathbf{X}],$$

and therefore

$$\mathsf{Bias}[\tilde{\mu}^{(\nu)}_{S-,p}(h)] = h^{1+p-\nu}\mathcal{B}_{S-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

$$\mathsf{Bias}[\tilde{\mu}^{(\nu)}_{S+,p}(h)] = h^{1+p-\nu}\mathcal{B}_{S-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

where

$$\mathcal{B}_{S-,\nu,p}(h) = \mathbf{s}'_{S,\nu}\mathbf{B}_{S-,p}(h) \to_{\mathbb{P}} \mathcal{B}_{S-,\nu,p} = \mathbf{s}'_{S,\nu}\mathbf{B}_{S-,p},$$

$$\mathcal{B}_{S+,\nu,p}(h) = \mathbf{s}'_{S,\nu}\mathbf{B}_{S+,p}(h) \to_{\mathbb{P}} \mathcal{B}_{S-,\nu,p} = \mathbf{s}'_{S,\nu}\mathbf{B}_{S+,p},$$

where we set $\mathbf{B}_{S-,p}(h) := \mathbf{B}_{S-,p,p}(h)$, $\mathbf{B}_{S+,p}(h) := \mathbf{B}_{S+,p,p}(h)$, $\mathbf{B}_{S-,p} := \mathbf{B}_{S-,p,p}$, and $\mathbf{B}_{S+,p} := \mathbf{B}_{S+,p,p}$.

Therefore, we define

$$\mathsf{Bias}[\tilde{\tau}_{Y,\nu}(\mathbf{h})] = \mathbb{E}[\mathbf{s}'_{S,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S,p}(\mathbf{h})) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S,p})]|\mathbf{X}]$$

and, using the results above,

$$\mathsf{Bias}[\tilde{\tau}_{Y,\nu}(\mathbf{h})] = h_+^{1+p-\nu}\mathcal{B}_{S+,\nu,p}(h_+) - h_-^{1+p-\nu}\mathcal{B}_{S-,\nu,p}(h_-) + o_{\mathbb{P}}(\max\{h_-^{2+p-\nu}, h_+^{2+p-\nu}\}).$$

7.5 Variance Approximation

We give the variance approximations for each of the estimators, under the conditions imposed above (Lemmas SA-1–SA-9).

7.5.1 Standard Sharp RD Estimator

We have

$$\mathbb{V}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{Y-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{Y+,\nu,p}(h_+)$$

with

$$\begin{aligned}\mathcal{V}_{Y-,\nu,p}(h) &= \nu!^2 h\mathbf{e}_\nu'\mathbf{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{\Sigma}_{Y-}\mathbf{K}_-(h)\mathbf{R}_p(h)\mathbf{\Gamma}_{-,p}^{-1}(h)\mathbf{e}_\nu/n \\ &= \nu!^2\mathbf{e}_\nu'\mathbf{P}_{-,p}(h)\mathbf{\Sigma}_{Y-}\mathbf{P}_{-,p}(h)'\mathbf{e}_\nu,\end{aligned}$$

$$\begin{aligned}\mathcal{V}_{Y+,\nu,p}(h) &= \nu!^2 h\mathbf{e}_\nu'\mathbf{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbf{\Sigma}_{Y+}\mathbf{K}_+(h)\mathbf{R}_p(h)\mathbf{\Gamma}_{+,p}^{-1}(h)\mathbf{e}_\nu/n \\ &= \nu!^2\mathbf{e}_\nu'\mathbf{P}_{+,p}(h)\mathbf{\Sigma}_{Y+}\mathbf{P}_{+,p}(h)'\mathbf{e}_\nu.\end{aligned}$$

Furthermore, we have

$$\mathcal{V}_{Y-,\nu,p}(h) \to_{\mathbb{P}} \nu!^2\mathbf{e}_\nu'\mathbf{\Gamma}_{-,p}^{-1}\mathbf{\Psi}_{Y-,p}\mathbf{\Gamma}_{-,p}^{-1}\mathbf{e}_\nu =: \mathcal{V}_{Y-,\nu,p},$$

$$\mathcal{V}_{Y+,\nu,p}(h) \to_{\mathbb{P}} \nu!^2\mathbf{e}_\nu'\mathbf{\Gamma}_{+,p}^{-1}\mathbf{\Psi}_{Y+,p}\mathbf{\Gamma}_{+,p}^{-1}\mathbf{e}_\nu =: \mathcal{V}_{Y+,\nu,p}.$$

7.5.2 Covariate-Adjusted Sharp RD Estimator

Using the linear approximation, we define

$$\begin{aligned}\mathsf{Var}[\tilde{\tau}_{Y,\nu}(\mathbf{h})] &= \mathbb{V}[\mathbf{s}_{S,\nu}'[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S,p}(\mathbf{h})) - \mathrm{vec}(\boldsymbol{\beta}_{S,p})]|\mathbf{X}] \\ &= \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{S-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{S+,\nu,p}(h_+)\end{aligned}$$

with

$$\mathcal{V}_{S-,\nu,p}(h) = \mathbf{s}_{S,\nu}'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h)]\mathbf{\Sigma}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h)']\mathbf{s}_{S,\nu},$$

$$\mathcal{V}_{S+,\nu,p}(h) = \mathbf{s}_{S,\nu}'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h)]\mathbf{\Sigma}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h)']\mathbf{s}_{S,\nu}.$$

Furthermore,

$$\mathcal{V}_{S-,\nu,p}(h) \to_{\mathbb{P}} \mathbf{s}_{S,\nu}'[\mathbf{I}_{1+d}\otimes\mathbf{\Gamma}_{-,p}^{-1}]\mathbf{\Psi}_{S-,p}[\mathbf{I}_{1+d}\otimes\mathbf{\Gamma}_{-,p}^{-1}]\mathbf{s}_{S,\nu} =: \mathcal{V}_{S-,\nu,p},$$

$$\mathcal{V}_{S+,\nu,p}(h) \to_{\mathbb{P}} \mathbf{s}_{S,\nu}'[\mathbf{I}_{1+d} \otimes \mathbf{\Gamma}_{+,p}^{-1}]\mathbf{\Psi}_{S+,p}[\mathbf{I}_{1+d} \otimes \mathbf{\Gamma}_{+,p}^{-1}]\mathbf{s}_{S,\nu} =: \mathcal{V}_{S+,\nu,p},$$

7.6 MSE Expansions

Using the derivations above, we give asymptotic MSE expansions and optimal bandwidth choices for the estimators considered. All the expressions in this section are justified as asymptotic approximations under the conditions $nh^{1+2\nu} \to \infty$ and $h \to 0$, with $\nu \leq p$, and the assumptions imposed throughout. We discuss the estimation of the unknown constants in the following sections, where these constants are also used for bias correction and standard error estimation.

For related results see Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014b), Arai and Ichimura (2018), and references therein.

7.6.1 Standard Sharp RD Estimator

- **MSE expansion: One-sided**. We have:

$$\begin{aligned}\mathbb{E}[(\hat{\mu}_{Y-,p}^{(\nu)}(h) - \mu_{Y-}^{(\nu)})^2|\mathbf{X}] &= h^{2(1+p-\nu)}\mathcal{B}_{Y-,\nu,p}^2(h)\{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{Y-,\nu,p}(h)\\ &= h^{2(1+p-\nu)}\mathcal{B}_{Y-,\nu,p}^2\{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{Y-,\nu,p}\{1 + o_{\mathbb{P}}(1)\}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(\hat{\mu}_{Y+,p}^{(\nu)}(h) - \mu_{Y+}^{(\nu)})^2|\mathbf{X}] &= h^{2(1+p-\nu)}\mathcal{B}_{Y+,\nu,p}^2(h)\{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{Y+,\nu,p}(h)\\ &= h^{2(1+p-\nu)}\mathcal{B}_{Y+,\nu,p}^2\{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{Y+,\nu,p}\{1 + o_{\mathbb{P}}(1)\}.\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{Y-,\nu,p} \neq 0$ and $\mathcal{B}_{Y+,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{Y-,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)}\frac{\mathcal{V}_{Y-,\nu,p}/n}{\mathcal{B}_{Y-,\nu,p}^2}\right]^{\frac{1}{3+2p}} \qquad \textit{and} \qquad \mathfrak{h}_{Y+,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)}\frac{\mathcal{V}_{Y+,\nu,p}/n}{\mathcal{B}_{Y+,\nu,p}^2}\right]^{\frac{1}{3+2p}}.$$

- **MSE expansion: Sum/Difference**. Let $h = h_+ = h_-$. Then, we have:

$$\begin{aligned}&\mathbb{E}[(\hat{\mu}_{Y+,p}^{(\nu)}(h) \pm \hat{\mu}_{Y-,p}^{(\nu)}(h) - (\mu_{Y+}^{(\nu)} \pm \mu_{Y-}^{(\nu)}))^2|\mathbf{X}]\\ &= h^{2(1+p-\nu)}\left[\mathcal{B}_{Y+,\nu,p}(h) \pm \mathcal{B}_{Y-,\nu,p}(h)\right]^2\{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\left[\mathcal{V}_{Y-,\nu,p}(h) + \mathcal{V}_{Y+,\nu,p}(h)\right]\\ &= h^{2(1+p-\nu)}\left[\mathcal{B}_{Y+,\nu,p} \pm \mathcal{B}_{Y-,\nu,p}\right]^2\{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\left[\mathcal{V}_{Y-,\nu,p} + \mathcal{V}_{Y+,\nu,p}\right]\{1 + o_{\mathbb{P}}(1)\}.\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{Y+,\nu,p} \pm \mathcal{B}_{Y-,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{\Delta Y,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)}\frac{(\mathcal{V}_{Y-,\nu,p} + \mathcal{V}_{Y+,\nu,p})/n}{(\mathcal{B}_{Y+,\nu,p} - \mathcal{B}_{Y-,\nu,p})^2}\right]^{\frac{1}{3+2p}},$$

$$\mathfrak{h}_{\Sigma Y,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\mathcal{V}_{Y-,\nu,p}+\mathcal{V}_{Y+,\nu,p})/n}{(\mathcal{B}_{Y+,\nu,p}+\mathcal{B}_{Y-,\nu,p})^2}\right]^{\frac{1}{3+2p}}.$$

7.6.2 Covariate-Adjusted Sharp RD Estimator

- **MSE expansion: One-sided**. We define

$$\mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y-,p}(h)] = \mathbb{E}[(\mathbf{s}'_{Y,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S-,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S-,p})])^2|\mathbf{X}],$$

$$\mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y+,p}(h)] = \mathbb{E}[(\mathbf{s}'_{Y,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S+,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S+,p})])^2|\mathbf{X}].$$

Then, we have:

$$\begin{aligned}\mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y-,p}(h)] &= h^{2(1+p-\nu)}\mathcal{B}^2_{S-,\nu,p}(h)\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{S-,\nu,p}(h)\\ &= h^{2(1+p-\nu)}\mathcal{B}^2_{S-,\nu,p}\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{S-,\nu,p}\{1+o_{\mathbb{P}}(1)\}\end{aligned}$$

and

$$\begin{aligned}\mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y+,p}(h)] &= h^{2(1+p-\nu)}\mathcal{B}^2_{S+,\nu,p}(h)\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{S+,\nu,p}(h)\\ &= h^{2(1+p-\nu)}\mathcal{B}^2_{S+,\nu,p}\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{S+,\nu,p}\{1+o_{\mathbb{P}}(1)\}\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{S-,\nu,p} \neq 0$ and $\mathcal{B}_{S+,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{S-,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{\mathcal{V}_{S-,\nu,p}/n}{\mathcal{B}^2_{S-,\nu,p}}\right]^{\frac{1}{3+2p}} \qquad \textit{and} \qquad \mathfrak{h}_{S+,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{\mathcal{V}_{S+,\nu,p}/n}{\mathcal{B}^2_{S+,\nu,p}}\right]^{\frac{1}{3+2p}}.$$

- **MSE expansion: Sum/Difference**. Let $h = h_+ = h_-$. We define

$$\mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y+,p}(h) \pm \tilde{\mu}^{(\nu)}_{Y-,p}(h)] = \mathbb{E}[(\mathbf{s}'_{Y,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{S,p}(h)) \pm \mathrm{vec}(\hat{\boldsymbol{\beta}}_{S,p})])^2|\mathbf{X}]$$

Then, we have:

$$\begin{aligned}&\mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y+,p}(h) \pm \tilde{\mu}^{(\nu)}_{Y-,p}(h)]\\ &= h^{2(1+p-\nu)}\left[\mathcal{B}_{S+,\nu,p}(h) \pm \mathcal{B}_{S-,\nu,p}(h)\right]^2\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\left[\mathcal{V}_{S-,\nu,p}(h)+\mathcal{V}_{S+,\nu,p}(h)\right]\\ &= h^{2(1+p-\nu)}\left[\mathcal{B}_{S+,\nu,p} \pm \mathcal{B}_{S-,\nu,p}\right]^2\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\left[\mathcal{V}_{S-,\nu,p}+\mathcal{V}_{S+,\nu,p}\right]\{1+o_{\mathbb{P}}(1)\}.\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{S+,\nu,p} \pm \mathcal{B}_{S-,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{\Delta S,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\mathcal{V}_{S-,\nu,p}+\mathcal{V}_{S+,\nu,p})/n}{(\mathcal{B}_{S+,\nu,p}-\mathcal{B}_{S-,\nu,p})^2}\right]^{\frac{1}{3+2p}},$$

$$\mathfrak{h}_{\Sigma S,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)}\frac{(\mathcal{V}_{S-,\nu,p}+\mathcal{V}_{S+,\nu,p})/n}{(\mathcal{B}_{S+,\nu,p}+\mathcal{B}_{S-,\nu,p})^2}\right]^{\frac{1}{3+2p}}.$$

Note that

$$\mathsf{MSE}[\tilde{\tau}_{Y,\nu}(h)] = \mathsf{MSE}[\tilde{\mu}^{(\nu)}_{Y+,p}(h) - \tilde{\mu}^{(\nu)}_{Y-,p}(h)].$$

7.7 Bias Correction

Using the derivations above, we give bias-correction formulas for the estimators considered. Recall that $\nu \leq p < q$.

7.7.1 Standard Sharp RD Estimator

For completeness, we present first the bias-correction for the sharp RD estimator without covariates. This case was already analyzed in detail by Calonico, Cattaneo, and Titiunik (2014b) and Calonico, Cattaneo, and Farrell (2018, 2019). The bias-corrected estimator in sharp RD designs without covariates is

$$\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \hat{\tau}_{Y,\nu}(\mathbf{h}) - \left[h_+^{1+p-\nu}\hat{\mathcal{B}}_{Y+,\nu,p,q}(h_+,b_+) - h_-^{1+p-\nu}\hat{\mathcal{B}}_{Y-,\nu,p,q}(h_-,b_-)\right],$$

where

$$\hat{\mathcal{B}}_{Y-,\nu,p,q}(h,b) = \nu!\mathbf{e}_\nu'\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p}(h)\frac{\hat{\mu}^{(1+p)}_{Y-,q}(b)}{(1+p)!},$$

$$\hat{\mathcal{B}}_{Y+,\nu,p,q}(h,b) = \nu!\mathbf{e}_\nu'\boldsymbol{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p}(h)\frac{\hat{\mu}^{(1+p)}_{Y+,q}(b)}{(1+p)!}.$$

Recall that $\hat{\tau}_{Y,\nu}(h)$ can be written as

$$\begin{aligned}
\hat{\tau}_{Y,\nu}(\mathbf{h}) &= \hat{\mu}^{(\nu)}_{Y+,p}(\mathbf{h}) - \hat{\mu}^{(\nu)}_{Y-,p}(\mathbf{h}) \\
&= \nu!\mathbf{e}_\nu'\left[\frac{1}{n^{1/2}h_+^{1/2}}\mathbf{H}_p^{-1}(h_+)\mathbf{P}_{+,p}(h_+) - \frac{1}{n^{1/2}h_-^{1/2}}\mathbf{H}_p^{-1}(h_-)\mathbf{P}_{-,p}(h_-)\right]\mathbf{Y} \\
&= \left[\frac{1}{n^{1/2}h_+^{1/2+\nu}}\nu!\mathbf{e}_\nu'\mathbf{P}_{+,p}(h_+) - \frac{1}{n^{1/2}h_-^{1/2+\nu}}\nu!\mathbf{e}_\nu'\mathbf{P}_{-,p}(h_-)\right]\mathbf{Y}
\end{aligned}$$

with

$$\mathbf{P}_{-,p}(h) = \sqrt{h}\boldsymbol{\Gamma}^{-1}_{-,p}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)/\sqrt{n},$$

$$\mathbf{P}_{+,p}(h) = \sqrt{h}\boldsymbol{\Gamma}^{-1}_{+,p}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)/\sqrt{n}.$$

The bias-corrected standard sharp RD estimator can also be represented in an analogous way. Setting $\rho = h/b$, we have

$$\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \hat{\mu}^{(\nu)\mathsf{bc}}_{Y+,p,q}(h_+,b_+) - \hat{\mu}^{(\nu)\mathsf{bc}}_{Y-,p,q}(h_-,b_-)$$

with

$$
\begin{aligned}
\hat{\mu}^{(\nu)\mathtt{bc}}_{Y-,p,q}(h,b) &= \hat{\mu}^{(\nu)}_{Y-,p}(h) - h^{1+p-\nu}\hat{\mathcal{B}}_{Y-,\nu,p,q}(h,b) \\
&= \nu!\mathbf{e}_\nu'\mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{Y}/n - h^{1+p-\nu}\nu!\mathbf{e}_\nu'\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p}(h)\frac{\hat{\mu}^{(1+p)}_{Y-,q}(b)}{(1+p)!} \\
&= \nu!\mathbf{e}_\nu'\mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\left[\mathbf{R}_p(h)'\mathbf{K}_-(h) - \rho^{1+p}\boldsymbol{\vartheta}_{-,p}(h)\mathbf{e}_{1+p}'\boldsymbol{\Gamma}_{-,q}^{-1}(b)\mathbf{R}_q(b)'\mathbf{K}_-(b)\right]\mathbf{Y}/n \\
&= \frac{1}{n^{1/2}h^{1/2+\nu}}\nu!\mathbf{e}_\nu'\mathbf{P}^{\mathtt{bc}}_{-,p,q}(h,b)\mathbf{Y}
\end{aligned}
$$

with

$$
\mathbf{P}^{\mathtt{bc}}_{-,p,q}(h,b) = \sqrt{h}\boldsymbol{\Gamma}_{-,p}^{-1}(h)\left[\mathbf{R}_p(h)'\mathbf{K}_-(h) - \rho^{1+p}\boldsymbol{\vartheta}_{-,p}(h)\mathbf{e}_{1+p}'\boldsymbol{\Gamma}_{-,q}^{-1}(b)\mathbf{R}_q(b)'\mathbf{K}_-(b)\right]/\sqrt{n},
$$

and, similarly,

$$
\hat{\mu}^{(\nu)\mathtt{bc}}_{Y+,p,q}(h,b) = \frac{1}{n^{1/2}h^{1/2+\nu}}\nu!\mathbf{e}_\nu'\mathbf{P}^{\mathtt{bc}}_{+,p,q}(h,b)\mathbf{Y}
$$

with

$$
\mathbf{P}^{\mathtt{bc}}_{+,p,q}(h,b) = \sqrt{h}\boldsymbol{\Gamma}_{+,p}^{-1}(h)\left[\mathbf{R}_p(h)'\mathbf{K}_+(h) - \rho^{1+p}\boldsymbol{\vartheta}_{+,p}(h)\mathbf{e}_{1+p}'\boldsymbol{\Gamma}_{+,q}^{-1}(b)\mathbf{R}_q(b)'\mathbf{K}_+(b)\right]/\sqrt{n}.
$$

Therefore,

$$
\hat{\tau}^{\mathtt{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \nu!\mathbf{e}_\nu'\left[\frac{1}{n^{1/2}h_+^{1/2+\nu}}\mathbf{P}^{\mathtt{bc}}_{+,p,q}(h_+,b_+) - \frac{1}{n^{1/2}h_-^{1/2+\nu}}\mathbf{P}^{\mathtt{bc}}_{-,p,q}(h_-,b_-)\right]\mathbf{Y}.
$$

7.7.2 Covariate-Adjusted Sharp RD Estimator

The bias-corrected covariate-adjusted sharp RD estimator is

$$
\tilde{\tau}^{\mathtt{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \tilde{\tau}_{Y,\nu}(\mathbf{h}) - \left[h_+^{1+p-\nu}\hat{\mathcal{B}}_{S+,p,q}(h_+,b_+) - h_-^{1+p-\nu}\hat{\mathcal{B}}_{S+,p,q}(h_-,b_-)\right],
$$

$$
\hat{\mathcal{B}}_{S-,\nu,p,q}(h,b) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{S-,q}(b)}{(1+p)!},
$$

$$
\hat{\mathcal{B}}_{S+,\nu,p,q}(h,b) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\vartheta}_{+,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{S+,q}(b)}{(1+p)!}.
$$

Recall that

$$
\tilde{\tau}_{Y,\nu}(\mathbf{h}) = \hat{\tau}_{Y,\nu}(\mathbf{h}) - \hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h})'\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) = \tilde{\mu}^{(\nu)}_{Y+,p}(h_+;\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) - \tilde{\mu}^{(\nu)}_{Y-,p}(h_-;\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}))
$$

and hence

$$
\tilde{\tau}^{\mathtt{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \tilde{\mu}^{(\nu)\mathtt{bc}}_{Y+,p}(h_+;\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) - \tilde{\mu}^{(\nu)\mathtt{bc}}_{Y-,p}(h_-;\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}))
$$

with

$$\tilde{\mu}^{(\nu)\mathsf{bc}}_{Y-,p}(h;\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) = \frac{1}{n^{1/2}h^{1/2+\nu}}\mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)]\mathbf{S},$$

$$\tilde{\mu}^{(\nu)\mathsf{bc}}_{Y+,p}(h;\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})) = \frac{1}{n^{1/2}h^{1/2+\nu}}\mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)]\mathbf{S}.$$

Therefore,

$$\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \mathbf{s}_{S,\nu}(\mathbf{h})'\left[\frac{1}{n^{1/2}h_+^{1/2+\nu}}[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)] - \frac{1}{n^{1/2}h_-^{1/2+\nu}}[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)]\right]\mathbf{S}.$$

7.8 Distributional Approximations

We study the classical and the robust bias-corrected standardized statistics based on the three estimators considered in the paper. We establish the asymptotic normality of the statistics allowing for (but nor requiring that) $\rho = h/b \to 0$, and hence our results depart from the traditional bias-correction approach in the nonparametrics literature; see Calonico, Cattaneo, and Titiunik (2014b) and Calonico, Cattaneo, and Farrell (2018, 2019) for more discussion.

7.8.1 Standard Sharp RD Estimator

The two standardized statistics are:

$$T_{Y,\nu}(\mathbf{h}) = \frac{\hat{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu}}{\sqrt{\mathbb{V}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}]}} \qquad \text{and} \qquad T^{\mathsf{bc}}_{Y,\nu}(h,b) = \frac{\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(h,b) - \tau_{Y,\nu}}{\sqrt{\mathbb{V}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(h,b)|\mathbf{X}]}}$$

where

$$\mathbb{V}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{Y-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{Y+,\nu,p}(h_+),$$

$$\mathcal{V}_{Y-,\nu,p}(h) = \nu!^2\mathbf{e}_\nu'\mathbf{P}_{-,p}(h)\boldsymbol{\Sigma}_{Y-}\mathbf{P}_{-,p}(h)'\mathbf{e}_\nu,$$

$$\mathcal{V}_{Y+,\nu,p}(h) = \nu!^2\mathbf{e}_\nu'\mathbf{P}_{+,p}(h)\boldsymbol{\Sigma}_{Y+}\mathbf{P}_{+,p}(h)'\mathbf{e}_\nu,$$

and

$$\mathbb{V}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{Y-,\nu,p,q}(h_-,b_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{Y+,\nu,p,q}(h_+,b_+),$$

$$\mathcal{V}^{\mathsf{bc}}_{Y-,\nu,p,q}(h,b) = \nu!^2\mathbf{e}_\nu'\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)\boldsymbol{\Sigma}_{Y-}\mathbf{P}^{\mathsf{bc}}_{-,\nu,q}(h,b)'\mathbf{e}_\nu,$$

$$\mathcal{V}^{\mathsf{bc}}_{Y+,\nu,p,q}(h,b) = \nu!^2\mathbf{e}_\nu'\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)\boldsymbol{\Sigma}_{Y+}\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)'\mathbf{e}_\nu.$$

As shown above, $\mathcal{V}_{Y-,\nu,p}(h_-) \asymp_{\mathbb{P}} 1$, $\mathcal{V}_{Y+,\nu,p}(h_+) \asymp_{\mathbb{P}} 1$, $\mathcal{V}^{\mathsf{bc}}_{Y-,\nu,p,q}(h_-,b_-) \asymp_{\mathbb{P}} 1$ and $\mathcal{V}^{\mathsf{bc}}_{Y+,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$, provided $\overline{\lim}_{n\to\infty}\max\{\rho_-,\rho_+\} < \infty$ and the other assumptions and bandwidth conditions hold. The following lemma gives asymptotic normality of the standardized statistics, and make precise the assumptions and bandwidth conditions required.

Lemma SA-10 *Let assumptions SA-1, SA-2 and SA-3 hold with $\varrho \geq 1+q$, and $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$.*
(1) If $nh_-^{2p+3} \to 0$ and $nh_+^{2p+3} \to 0$, then

$$T_{Y,\nu}(\mathbf{h}) \to_d \mathcal{N}(0,1).$$

(2) If $nh_-^{2p+3}\max\{h_-^2, b_-^{2(q-p)}\} \to 0$, $nh_+^{2p+3}\max\{h_+^2, b_+^{2(q-p)}\} \to 0$ and $\overline{\lim}_{n\to\infty}\max\{\rho_-,\rho_+\} < \infty$, then

$$T^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) \to_d \mathcal{N}(0,1).$$

Proof of Lemma SA-10. This theorem is an special case of lemma SA-11 below (i.e., when covariates are not included). ∎

7.8.2 Covariate-Adjusted Sharp RD Estimator

The two standardized statistics are:

$$T_{S,\nu}(\mathbf{h}) = \frac{\tilde{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu}}{\sqrt{\mathsf{Var}[\tilde{\tau}_{Y,\nu}(\mathbf{h})]}} \qquad \text{and} \qquad T^{\mathsf{bc}}_{S,\nu}(\mathbf{h},\mathbf{b}) = \frac{\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) - \tau_{Y,\nu}}{\sqrt{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}}$$

where

$$\mathsf{Var}[\tilde{\tau}_{Y,\nu}(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{S-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{S+,\nu,p}(h_+),$$

$$\mathcal{V}_{S-,\nu,p}(h) = \mathbf{s}'_{S,\nu}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h)]\mathbf{\Sigma}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h)']\mathbf{s}_{S,\nu},$$

$$\mathcal{V}_{S+,\nu,p}(h) = \mathbf{s}'_{S,\nu}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h)]\mathbf{\Sigma}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h)']\mathbf{s}_{S,\nu}.$$

and

$$\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{S-,\nu,p,q}(h_-,b_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{S+,\nu,p,q}(h_+,b_+),$$

$$\mathcal{V}^{\mathsf{bc}}_{S-,\nu,p,q}(h,b) = \mathbf{s}'_{S,\nu}[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)]\mathbf{\Sigma}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)']\mathbf{s}_{S,\nu},$$

$$\mathcal{V}^{\mathsf{bc}}_{S+,\nu,p,q}(h,b) = \mathbf{s}'_{S,\nu}[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)]\mathbf{\Sigma}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)']\mathbf{s}_{S,\nu}.$$

As shown above, $\mathcal{V}_{S-,\nu,p}(h) \asymp_{\mathbb{P}} 1$, $\mathcal{V}_{S+,\nu,p}(h) \asymp_{\mathbb{P}} 1$, $\mathcal{V}^{\mathsf{bc}}_{S-,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$ and $\mathcal{V}^{\mathsf{bc}}_{S+,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$, provided $\overline{\lim}_{n\to\infty}\max\{\rho_-,\rho_+\} < \infty$ and the other assumptions and bandwidth conditions hold. The following lemma gives asymptotic normality of the standardized statistics, and make precise the assumptions and bandwidth conditions required.

Lemma SA-11 *Let assumptions SA-1, SA-2 and SA-3 hold with $\varrho \geq 1+q$, and $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$.*

(1) If $nh_-^{2p+3} \to 0$ *and* $nh_+^{2p+3} \to 0$, *then*

$$T_{S,\nu}(\mathbf{h}) \to_d \mathcal{N}(0,1).$$

(2) If $nh_-^{2p+3}\max\{h_-^2, b_-^{2(q-p)}\} \to 0$, $nh_+^{2p+3}\max\{h_+^2, b_+^{2(q-p)}\} \to 0$ *and* $\overline{\lim}_{n\to\infty}\max\{\rho_-, \rho_+\} < \infty$, *then*

$$T^{\mathsf{bc}}_{S,\nu}(\mathbf{h},\mathbf{b}) \to_d \mathcal{N}(0,1).$$

Proof of Lemma SA-11. Both parts follow from the Linderberg-Feller's triangular array central limit theorem. Here we prove only part (2), as part (1) is analogous. Only for simplicity we assume that $h = h_- = h_+$ and $b = b_- = b_+$.

First, recall that $\tilde{\tau}_{Y,\nu}(\mathbf{h}) = \hat{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu} - \hat{\boldsymbol{\tau}}_{Z,\nu}(\mathbf{h})'\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})$, and hence define

$$\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) - \tau_{Y,\nu} - \hat{\boldsymbol{\tau}}^{\mathsf{bc}}_{Z,\nu}(\mathbf{h},\mathbf{b})'\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}),$$

where $\hat{\boldsymbol{\tau}}^{\mathsf{bc}}_{Z,\nu}(\mathbf{h},\mathbf{b})$ denotes the bias-corrected standard RD estimator using the additional covariates as outcome variables (c.f., $\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})$). Then, since $\boldsymbol{\tau}_{Z,\nu} = \mathbf{0}$ by assumption,

$$T^{\mathsf{bc}}_{S,\nu}(\mathbf{h},\mathbf{b}) = \frac{\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) - \tau_{Y,\nu} - \hat{\boldsymbol{\tau}}^{\mathsf{bc}}_{Z,\nu}(\mathbf{h},\mathbf{b})'\boldsymbol{\gamma}_{Y,p}}{\sqrt{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}} + o_{\mathbb{P}}(1)$$

because $nh^{1+2\nu}\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})] \asymp_{\mathbb{P}} 1$ and $\sqrt{nh^{1+2\nu}}\hat{\boldsymbol{\tau}}^{\mathsf{bc}}_{Z,\nu}(\mathbf{h},\mathbf{b})'[\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) - \boldsymbol{\gamma}_{Y,p}] = O_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$.

Second, let

$$\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S,p,q}(\mathbf{h},\mathbf{b}) = \hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S+,p,q}(h_+,b_+) - \hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S-,p,q}(h_-,b_-),$$
$$\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S-,p,q}(h,b) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}^{\mathsf{bc}}_{-,p}(h)\mathbf{S},$$
$$\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S+,p,q}(h,b) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}^{\mathsf{bc}}_{+,p}(h)\mathbf{S},$$

and therefore

$$T^{\mathsf{bc}}_{S,\nu}(\mathbf{h},\mathbf{b}) = \frac{\mathbf{s}'_{S,\nu}\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S,p,q}(\mathbf{h},\mathbf{b}) - \mathbb{E}[\mathbf{s}'_{S,\nu}\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S,p,q}(\mathbf{h},\mathbf{b})|\mathbf{X}]}{\sqrt{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}} + o_{\mathbb{P}}(1),$$

because, using the previous results and the structure of the bias-corrected estimator, we have

$$\frac{\mathbb{E}[\mathbf{s}'_{S,\nu}\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S,p,q}(\mathbf{h},\mathbf{b})|\mathbf{X}] - \tau_{Y,\nu}}{\sqrt{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}} = O_{\mathbb{P}}\left(\sqrt{n}h^{1/2+p+2}\right) + O_{\mathbb{P}}\left(\sqrt{n}h^{1/2+1+p}b^{(q-p)}\right) = o_{\mathbb{P}}(1).$$

Finally, we have

$$T^{\mathsf{bc}}_{S,\nu}(\mathbf{h},\mathbf{b}) = \frac{\mathbf{s}'_{S,\nu}\left[\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S,p,q}(\mathbf{h},\mathbf{b}) - \mathbb{E}[\hat{\boldsymbol{\beta}}^{\mathsf{bc}}_{S,p,q}(\mathbf{h},\mathbf{b})|\mathbf{X}]\right]}{\sqrt{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}} + o_{\mathbb{P}}(1) \to_d \mathcal{N}(0,1)$$

using a triangular array CLT for mean-zero variance-one independent random variables, provided that $nh \to \infty$. ∎

7.9 Variance Estimation

The only unknown matrices in the asymptotic variance formulas derived above are:

- Standard Estimator: $\boldsymbol{\Sigma}_{Y-} = \mathbb{V}[\mathbf{Y}(0)|\mathbf{X}]$ and $\boldsymbol{\Sigma}_{Y+} = \mathbb{V}[\mathbf{Y}(1)|\mathbf{X}]$.
- Covariate-Adjusted Estimator: $\boldsymbol{\Sigma}_{S-} = \mathbb{V}[\mathbf{S}(0)|\mathbf{X}]$ and $\boldsymbol{\Sigma}_{S+} = \mathbb{V}[\mathbf{S}(1)|\mathbf{X}]$.

All these matrices are assumed to be diagonal matrices, since we impose conditional heteroskedasticity of unknown form. In the following section we discuss the case where these matrices are block diagonal, that is, under clustered data, which requires only a straightforward extension of the methodological work outlined in this appendix.

In the heteroskedastic case, each diagonal element would contain the unit's specific conditional variance terms for units to the left of the cutoff (controls) and for units to the right of the cutoff (treatments). Thus, simple plug-in variance estimators can be constructed using estimated residuals, as it is common in heteroskedastic linear model settings. In this section we describe this approach in some detail.

We consider two alternative type of standard error estimators, based on either a Nearest Neighbor (NN) and plug-in residuals (PR) approach. For $i = 1, 2, \cdots, n$, define the "estimated" residuals as follows.

- **Nearest Neighbor (NN) approach**:

$$\hat{\varepsilon}_{V-,i}(J) = \mathbb{1}(X_i < \bar{x})\sqrt{\frac{J}{J+1}}\left(V_i - \frac{1}{J}\sum_{j=1}^{J} V_{\ell_{-,j}(i)}\right),$$

$$\hat{\varepsilon}_{V+,i}(J) = \mathbb{1}(X_i \geq \bar{x})\sqrt{\frac{J}{J+1}}\left(V_i - \frac{1}{J}\sum_{j=1}^{J} V_{\ell_{+,j}(i)}\right),$$

where $V \in \{Y, Z_1, Z_2, \cdots, Z_d\}$, and $\ell_{+,j}(i)$ is the index of the j-th closest unit to unit i among $\{X_i : X_i \geq \bar{x}\}$ and $\ell_{-,j}(i)$ is the index of the j-th closest unit to unit i among $\{X_i : X_i < \bar{x}\}$, and J denotes a (fixed) the number of neighbors chosen.

- **Plug-in Residuals (PR) approach**:

$$\hat{\varepsilon}_{V-,p,i}(h) = \mathbb{1}(X_i < \bar{x})\sqrt{\omega_{-,p,i}}(V_i - \mathbf{r}_p(X_i - \bar{x})'\hat{\boldsymbol{\beta}}_{V-,p}(h)),$$

$$\hat{\varepsilon}_{V+,p,i}(h) = \mathbb{1}(X_i \geq \bar{x})\sqrt{\omega_{+,p,i}}(V_i - \mathbf{r}_p(X_i - \bar{x})'\hat{\boldsymbol{\beta}}_{V+,p}(h)),$$

where again $V \in \{Y, Z_1, Z_2, \cdots, Z_d\}$ is a placeholder for the outcome variable used, and the additional weights $\{(\omega_{-,p,i}, \omega_{+,p,i}) : i = 1, 2, \cdots, n\}$ are introduced to handle the different variants of heteroskedasticity-robust asymptotic variance constructions (e.g., Long and Ervin (2000), MacKinnon (2012), and references therein). Typical examples of these weights are

	HC_0	HC_1	HC_2	HC_3
$\omega_{-,p,i}$	1	$\frac{N_-}{N_- - 2\,\text{tr}(\mathbf{Q}_{-,p}) + \text{tr}(\mathbf{Q}_{-,p}\mathbf{Q}_{-,p})}$	$\frac{1}{\mathbf{e}_i'\mathbf{Q}_{-,p}\mathbf{e}_i}$	$\frac{1}{(\mathbf{e}_i'\mathbf{Q}_{-,p}\mathbf{e}_i)^2}$
$\omega_{+,p,i}$	1	$\frac{N_+}{N_+ - 2\,\text{tr}(\mathbf{Q}_{+,p}) + \text{tr}(\mathbf{Q}_{+,p}\mathbf{Q}_{+,p})}$	$\frac{1}{\mathbf{e}_i'\mathbf{Q}_{+,p}\mathbf{e}_i}$	$\frac{1}{(\mathbf{e}_i'\mathbf{Q}_{+,p}\mathbf{e}_i)^2}$

where

$$N_- = \sum_{i=1}^{n} \mathbb{1}(X_i < \bar{x}) \qquad \text{and} \qquad N_+ = \sum_{i=1}^{n} \mathbb{1}(X_i \geq \bar{x}),$$

and $(\mathbf{Q}_{-,p}, \mathbf{Q}_{+,p})$ denote the corresponding "projection" matrices used to obtain the estimated residuals,

$$\mathbf{Q}_{-,p} = \mathbf{R}_p(h)\boldsymbol{\Gamma}_{-,p}^{-1}\mathbf{R}_p(h)'\mathbf{K}_-(h)/n, \qquad \mathbf{Q}_{+,p} = \mathbf{R}_p(h)\boldsymbol{\Gamma}_{+,p}^{-1}\mathbf{R}_p(h)'\mathbf{K}_+(h)/n,$$

and $\mathbf{e}_i'\mathbf{Q}_-\mathbf{e}$ and $\mathbf{e}_i'\mathbf{Q}_+\mathbf{e}_i$ are the corresponding i-th diagonal element.

7.9.1 Standard Sharp RD Estimator

Define the estimators

$$\hat{\boldsymbol{\Sigma}}_{Y-}(J) = \text{diag}(\hat{\varepsilon}^2_{Y-,1}(J), \hat{\varepsilon}^2_{Y-,2}(J), \cdots, \hat{\varepsilon}^2_{Y-,n}(J)),$$

$$\hat{\boldsymbol{\Sigma}}_{Y+}(J) = \text{diag}(\hat{\varepsilon}^2_{Y+,1}(J), \hat{\varepsilon}^2_{Y+,2}(J), \cdots, \hat{\varepsilon}^2_{Y+,n}(J)),$$

and

$$\hat{\boldsymbol{\Sigma}}_{Y-,p}(h) = \text{diag}(\hat{\varepsilon}^2_{Y-,p,1}(h), \hat{\varepsilon}^2_{Y-,p,2}(h), \cdots, \hat{\varepsilon}^2_{Y-,p,n}(h)),$$

$$\hat{\boldsymbol{\Sigma}}_{Y+,p}(h) = \text{diag}(\hat{\varepsilon}^2_{Y+,p,1}(h), \hat{\varepsilon}^2_{Y+,p,2}(h), \cdots, \hat{\varepsilon}^2_{Y+,p,n}(h)).$$

- **Undersmoothing NN Variance Estimator**:

$$\check{\mathbb{V}}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}_{Y-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}_{Y+,\nu,p}(h_+),$$

$$\check{\mathcal{V}}_{Y-,\nu,p}(h) = \nu!^2\mathbf{e}_\nu'\mathbf{P}_{-,p}(h)\hat{\boldsymbol{\Sigma}}_{Y-}(J)\mathbf{P}_{-,p}(h)'\mathbf{e}_\nu,$$

$$\check{\mathcal{V}}_{Y+,\nu,p}(h) = \nu!^2\mathbf{e}_\nu'\mathbf{P}_{+,p}(h)\hat{\boldsymbol{\Sigma}}_{Y+}(J)\mathbf{P}_{+,p}(h)'\mathbf{e}_\nu.$$

- **Undersmoothing PR Variance Estimator**:

$$\hat{\mathbb{V}}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}_{Y-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}_{Y+,\nu,p}(h_+),$$

$$\hat{\mathcal{V}}_{Y-,\nu,p}(h) = \nu!^2\mathbf{e}_\nu'\mathbf{P}_{-,p}(h)\hat{\mathbf{\Sigma}}_{Y-,p}(h)\mathbf{P}_{-,p}(h)'\mathbf{e}_\nu,$$

$$\hat{\mathcal{V}}_{Y+,\nu,p}(h) = \nu!^2\mathbf{e}_\nu'\mathbf{P}_{+,p}(h)\hat{\mathbf{\Sigma}}_{Y+,p}(h)\mathbf{P}_{+,p}(h)'\mathbf{e}_\nu.$$

- **Robust Bias-Correction NN Variance Estimator**:

$$\check{\mathbb{V}}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}^{\mathsf{bc}}_{Y-,\nu,p,q}(h_-,b_-) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}^{\mathsf{bc}}_{Y+,\nu,p,q}(h_+,b_+),$$

$$\check{\mathcal{V}}^{\mathsf{bc}}_{Y-,\nu,p,q}(h,b) = \nu!^2\mathbf{e}_\nu'\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)\hat{\mathbf{\Sigma}}_{Y-}(J)\mathbf{P}^{\mathsf{bc}}_{-,\nu,q}(h,b)'\mathbf{e}_\nu$$

$$\check{\mathcal{V}}^{\mathsf{bc}}_{Y+,\nu,p,q}(h,b) = \nu!^2\mathbf{e}_\nu'\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)\hat{\mathbf{\Sigma}}_{Y+}(J)\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)'\mathbf{e}_\nu.$$

- **Robust Bias-Correction PR Variance Estimator**:

$$\hat{\mathbb{V}}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}^{\mathsf{bc}}_{Y-,\nu,p,q}(h_-,b_-) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}^{\mathsf{bc}}_{Y+,\nu,p,q}(h_+,b_+),$$

$$\mathcal{V}^{\mathsf{bc}}_{Y-,\nu,p,q}(h,b) = \nu!^2\mathbf{e}_\nu'\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)\hat{\mathbf{\Sigma}}_{Y-,q}(h)\mathbf{P}^{\mathsf{bc}}_{-,\nu,q}(h,b)'\mathbf{e}_\nu$$

$$\mathcal{V}^{\mathsf{bc}}_{Y+,\nu,p,q}(h,b) = \nu!^2\mathbf{e}_\nu'\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)\hat{\mathbf{\Sigma}}_{Y+,q}(h)\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)'\mathbf{e}_\nu.$$

The following lemma gives the consistency of these asymptotic variance estimators.

Lemma SA-12 *Suppose the conditions of Lemma SA-10 hold. If, in addition,* $\max_{1\leq i\leq n}|\omega_{-,p,i}| = O_{\mathbb{P}}(1)$ *and* $\max_{1\leq i\leq n}|\omega_{+,p,i}| = O_{\mathbb{P}}(1)$*, and* $\sigma^2_{S+}(x)$ *and* $\sigma^2_{S-}(x)$ *are Lipschitz continuous, then*

$$\frac{\check{\mathbb{V}}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}]}{\mathbb{V}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}]} \to_{\mathbb{P}} 1, \quad \frac{\check{\mathbb{V}}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}]}{\mathbb{V}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathbb{V}}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}]}{\mathbb{V}[\hat{\tau}_{Y,\nu}(\mathbf{h})|\mathbf{X}]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathbb{V}}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}]}{\mathbb{V}[\hat{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})|\mathbf{X}]} \to_{\mathbb{P}} 1.$$

The first part of the lemma was proven in Calonico, Cattaneo, and Titiunik (2014b), while the second part follows directly from well known results in the local polynomial literature (e.g., Fan and Gijbels (1996)). We do not include the proof to conserve same space.

7.9.2 Covariate-Adjusted Sharp RD Estimator

Define the estimators

$$\check{\boldsymbol{\Sigma}}_{S-}(J) = \begin{bmatrix} \check{\boldsymbol{\Sigma}}_{YY-}(J) & \check{\boldsymbol{\Sigma}}_{YZ_1-}(J) & \check{\boldsymbol{\Sigma}}_{YZ_2-}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{YZ_d-}(J) \\ \check{\boldsymbol{\Sigma}}_{Z_1Y-}(J) & \check{\boldsymbol{\Sigma}}_{Z_1Z_1-}(J) & \check{\boldsymbol{\Sigma}}_{Z_1Z_2-}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{Z_1Z_d-}(J) \\ \check{\boldsymbol{\Sigma}}_{Z_2Y-}(J) & \check{\boldsymbol{\Sigma}}_{Z_2Z_1-}(J) & \check{\boldsymbol{\Sigma}}_{Z_2Z_2-}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{Z_2Z_d-}(J) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \check{\boldsymbol{\Sigma}}_{Z_dY-}(J) & \check{\boldsymbol{\Sigma}}_{Z_dZ_1-}(J) & \check{\boldsymbol{\Sigma}}_{Z_dZ_2-}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{Z_dZ_d-}(J) \end{bmatrix}$$

and

$$\check{\boldsymbol{\Sigma}}_{S+}(J) = \begin{bmatrix} \check{\boldsymbol{\Sigma}}_{YY+}(J) & \check{\boldsymbol{\Sigma}}_{YZ_1+}(J) & \check{\boldsymbol{\Sigma}}_{YZ_2+}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{YZ_d+}(J) \\ \check{\boldsymbol{\Sigma}}_{Z_1Y+}(J) & \check{\boldsymbol{\Sigma}}_{Z_1Z_1+}(J) & \check{\boldsymbol{\Sigma}}_{Z_1Z_2+}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{Z_1Z_d+}(J) \\ \check{\boldsymbol{\Sigma}}_{Z_2Y+}(J) & \check{\boldsymbol{\Sigma}}_{Z_2Z_1+}(J) & \check{\boldsymbol{\Sigma}}_{Z_2Z_2+}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{Z_2Z_d+}(J) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \check{\boldsymbol{\Sigma}}_{Z_dY+}(J) & \check{\boldsymbol{\Sigma}}_{Z_dZ_1+}(J) & \check{\boldsymbol{\Sigma}}_{Z_dZ_2+}(J) & \cdots & \check{\boldsymbol{\Sigma}}_{Z_dZ_d+}(J) \end{bmatrix}$$

where the matrices $\check{\boldsymbol{\Sigma}}_{VW-}(J)$ and $\check{\boldsymbol{\Sigma}}_{VW+}(J)$, $V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$, are $n \times n$ matrices with generic (i,j)-th elements, respectively,

$$\left[\check{\boldsymbol{\Sigma}}_{VW-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V-,i}(J)\hat{\varepsilon}_{W-,i}(J),$$

$$\left[\check{\boldsymbol{\Sigma}}_{VW+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V+,i}(J)\hat{\varepsilon}_{W+,i}(J),$$

for all $1 \leq i, j \leq n$, and for all $V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$.

Similarly, define the estimators

$$\hat{\boldsymbol{\Sigma}}_{S-,p}(h) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{YY-,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{YZ_d-,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{Z_1Y-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_1Z_d-,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{Z_2Y-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_2Z_d-,p}(h) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{\Sigma}}_{Z_dY-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_dZ_d-,p}(h) \end{bmatrix}$$

and

$$\hat{\boldsymbol{\Sigma}}_{S+,p}(h) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{YY+,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{YZ_d+,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{Z_1Y+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_1Z_d+,p}(h) \\ \check{\boldsymbol{\Sigma}}_{Z_2Y+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_2Z_d+,p}(h) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{\Sigma}}_{Z_dY+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_dZ_d+,p}(h) \end{bmatrix}$$

where the matrices $\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)$ and $\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)$, $V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$, are $n \times n$ matrices with generic (i,j)-th elements, respectively,

$$\Big[\hat{\mathbf{\Sigma}}_{VW-,p}(h)\Big]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V-,p,i}(h)\hat{\varepsilon}_{W-,p,j}(h),$$

$$\Big[\hat{\mathbf{\Sigma}}_{VW+,p}(h)\Big]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V+,p,i}(h)\hat{\varepsilon}_{W+,p,j}(h),$$

for all $1 \leq i,j \leq n$, and for all $V,W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$.

- **Undersmoothing NN Variance Estimator**:

$$\check{\mathsf{V}}\mathsf{ar}[\tilde{\tau}_{Y,\nu}(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}_{S-,\nu,p}(\mathbf{h}) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}_{S+,\nu,p}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{S-,\nu,p}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h_-)]\check{\mathbf{\Sigma}}_{S-}(J)[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h_-)']\mathbf{s}_{S,\nu}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{S+,\nu,p}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h_+)]\check{\mathbf{\Sigma}}_{S+}(J)[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h_+)']\mathbf{s}_{S,\nu}(\mathbf{h}).$$

- **Undersmoothing PR Variance Estimator**:

$$\hat{\mathsf{V}}\mathsf{ar}[\tilde{\tau}_{Y,\nu}(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}_{S-,\nu,p}(\mathbf{h}) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}_{S+,\nu,p}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{S-,\nu,p}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h_-)]\hat{\mathbf{\Sigma}}_{S-,p}(h_-)[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(h_-)']\mathbf{s}_{S,\nu}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{S+,\nu,p}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h_-)]\hat{\mathbf{\Sigma}}_{S+,p}(h_+)[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(h_-)']\mathbf{s}_{S,\nu}(\mathbf{h}).$$

- **Robust Bias-Correction NN Variance Estimator**:

$$\check{\mathsf{V}}\mathsf{ar}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\mathbf{h},\mathbf{b}) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}_{S+,\nu,p,q}(\mathbf{h},\mathbf{b}),$$

$$\check{\mathcal{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\mathbf{h},\mathbf{b}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)]\check{\mathbf{\Sigma}}_{S-}(J)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)']\mathbf{s}_{S,\nu}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{S+,\nu,p,q}(\mathbf{h},\mathbf{b}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)]\check{\mathbf{\Sigma}}_{S+}(J)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)']\mathbf{s}_{S,\nu}(\mathbf{h}).$$

- **Robust Bias-Correction PR Variance Estimator**:

$$\hat{\mathsf{V}}\mathsf{ar}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\mathbf{h},\mathbf{b}) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}_{S+,\nu,p,q}(\mathbf{h},\mathbf{b}),$$

$$\hat{\mathcal{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\mathbf{h},\mathbf{b}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)]\hat{\mathbf{\Sigma}}_{S-,q}(h_-)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)']\mathbf{s}_{S,\nu}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{S+,\nu,p,q}(\mathbf{h},\mathbf{b}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)]\hat{\mathbf{\Sigma}}_{S+,q}(h_+)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)']\mathbf{s}_{S,\nu}(\mathbf{h}).$$

The following lemma gives the consistency of these asymptotic variance estimators.

Lemma SA-13 *Suppose the conditions of Lemma SA-11 hold. If, in addition,* $\max_{1\leq i\leq n}|\omega_{-,i}| = O_{\mathbb{P}}(1)$ *and* $\max_{1\leq i\leq n}|\omega_{+,i}| = O_{\mathbb{P}}(1)$, *and* $\sigma^2_{S+}(x)$ *and* $\sigma^2_{S-}(x)$ *are Lipschitz continuous, then*

$$\frac{\check{\mathsf{V}}\mathsf{ar}[\tilde{\tau}_{Y,\nu}(\mathbf{h})]}{\mathsf{Var}[\tilde{\tau}_{Y,\nu}(\mathbf{h})]]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathsf{V}}\mathsf{ar}[\tilde{\tau}_{Y,\nu}(\mathbf{h})]}{\mathsf{Var}[\tilde{\tau}_{Y,\nu}(\mathbf{h})]} \to_{\mathbb{P}} 1, \quad \frac{\check{\mathsf{V}}\mathsf{ar}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathsf{V}}\mathsf{ar}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]}{\mathsf{Var}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b})]} \to_{\mathbb{P}} 1.$$

The proof of this result is also standard. For example, the first result reduces to showing

$$\left|\mathbf{R}_p(h)'\mathbf{K}_-(h)\check{\mathbf{\Sigma}}_{VW-}(J)\mathbf{K}_-(h)\mathbf{R}_p(h) - \mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{\Sigma}_{VW-}\mathbf{K}_-(h)\mathbf{R}_p(h)\right| = o_{\mathbb{P}}(n),$$

$$\left|\mathbf{R}_p(h)'\mathbf{K}_+(h)\check{\mathbf{\Sigma}}_{VW+}(J)\mathbf{K}_+(h)\mathbf{R}_p(h) - \mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbf{\Sigma}_{VW+}\mathbf{K}_+(h)\mathbf{R}_p(h)\right| = o_{\mathbb{P}}(n),$$

$$\mathbf{\Sigma}_{VW-} = \mathbb{C}ov[\mathbf{V}(0),\mathbf{W}(0)|\mathbf{X}], \quad \mathbf{\Sigma}_{VW+} = \mathbb{C}ov[\mathbf{V}(1),\mathbf{W}(1)|\mathbf{X}],$$

$$V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\},$$

which can be established using bounding calculations under the assumptions imposed. The other results are proven the same way.

7.10 Extension to Clustered Data

As discussed in the main text, it is straightforward to extend the results above to the case where the data exhibits a clustered structured. All the derivations and results obtained above remain valid, with the only exception of the asymptotic variance formulas, which now would depend on the particular form of clustering. In this case, the asymptotics are conducted assuming that the number of clusters, G, grows ($G \to \infty$) satisfying the usual asymptotic restriction $Gh \to \infty$. For a review on cluster-robust inference see Cameron and Miller (2015).

For brevity, in this section we only describe the asymptotic variance estimators with clustering, which are now implemented in the upgraded versions of the `Stata` and `R` software described in Calonico, Cattaneo, and Titiunik (2014a, 2015). Specifically, we assume that each unit i belongs to one (and only one) cluster g, and let $\mathcal{G}(i) = g$ for all units $i = 1, 2, \cdots, n$ and all clusters $g = 1, 2, \cdots, G$. Define

$$\omega_{-,p} = \frac{G}{G-1}\frac{N_- - 1}{N_- - 1 - p}, \qquad \omega_{+,p} = \frac{G}{G-1}\frac{N_+ - 1}{N_+ - 1 - p}.$$

The clustered-consistent variance estimators are as follows. We recycle notation for convenience, and to emphasize the nesting of the heteroskedasticity-robust estimators into the cluster-robust ones.

7.10.1 Standard Sharp RD Estimator

Redefine the matrices $\check{\mathbf{\Sigma}}_{Y-}(J)$ and $\check{\mathbf{\Sigma}}_{Y+}(J)$, respectively, to now have generic (i,j)-th elements

$$\left[\check{\mathbf{\Sigma}}_{Y-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{Y-,i}(J)\hat{\varepsilon}_{Y-,i}(J),$$

$$\left[\check{\mathbf{\Sigma}}_{Y+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{Y+,i}(J)\hat{\varepsilon}_{Y+,i}(J),$$

$$1 \leq i, j \leq n.$$

Similarly, redefine the matrices $\hat{\mathbf{\Sigma}}_{Y-,p}(h)$ and $\hat{\mathbf{\Sigma}}_{Y+,p}(h)$, respectively, to now have generic (i,j)-th elements

$$\left[\hat{\mathbf{\Sigma}}_{Y-,p}(h)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{Y-,p,i}(h)\hat{\varepsilon}_{Y-,p,j}(h),$$

$$\left[\hat{\mathbf{\Sigma}}_{Y+,p}(h)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{Y+,p,i}(h)\hat{\varepsilon}_{Y+,p,j}(h),$$

$$1 \leq i, j \leq n.$$

With these redefinitions, the clustered-robust variance estimators are as above. In particular, if each cluster has one observation, then the estimators reduce to the heteroskedastic-robust estimators with $\omega_{-,p,i} = \omega_{+,p,i} = 1$ for all $i = 1, 2, \cdots, n$.

7.10.2 Covariate-Adjusted Sharp RD Estimator

Redefine the matrices $\check{\mathbf{\Sigma}}_{VW-}(J)$ and $\check{\mathbf{\Sigma}}_{VW+}(J)$, respectively, to now have generic (i,j)-th elements

$$\left[\check{\mathbf{\Sigma}}_{VW-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V-,i}(J)\hat{\varepsilon}_{W-,i}(J),$$

$$\left[\check{\mathbf{\Sigma}}_{VW+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V+,i}(J)\hat{\varepsilon}_{W+,i}(J),$$

$$1 \leq i, j \leq n, \qquad V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}.$$

Similarly, redefine the matrices $\hat{\mathbf{\Sigma}}_{VW-,p}(h)$ and $\hat{\mathbf{\Sigma}}_{VW+,p}(h)$, respectively, to now have generic (i,j)-th elements

$$\left[\hat{\mathbf{\Sigma}}_{VW-,p}(h)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V-,p,i}(h)\hat{\varepsilon}_{W-,p,j}(h),$$

$$\left[\hat{\mathbf{\Sigma}}_{VW+,p}(h)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V+,p,i}(h)\hat{\varepsilon}_{W+,p,j}(h),$$

$$1 \leq i, j \leq n, \qquad V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}.$$

With these redefinitions, the clustered-robust variance estimators are as above. In particular, if each cluster has one observation, then the estimators reduce to the heteroskedastic-robust estimators with $\omega_{-,p,i} = \omega_{+,p,i} = 1$ for all $i = 1, 2, \cdots, n$.

8 Estimation using Treatment Interaction

Consider now the following treatment-interacted covariate-adjusted sharp RD estimator:

$$\check{\eta}_{Y,\nu}(\mathbf{h}) = \nu! \mathbf{e}_\nu' \check{\boldsymbol{\beta}}_{Y+,p}(h_-) - \nu! \mathbf{e}_\nu' \check{\boldsymbol{\beta}}_{Y-,p}(h_+),$$

$$\check{\boldsymbol{\theta}}_{Y-,p}(h) = \begin{bmatrix} \check{\boldsymbol{\beta}}_{Y-,p}(h) \\ \check{\boldsymbol{\gamma}}_{Y-,p}(h) \end{bmatrix} = \underset{\mathbf{b}\in\mathbb{R}^{1+p},\boldsymbol{\gamma}\in\mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}(X_i < \bar{x})(Y_i - \mathbf{r}_p(X_i-\bar{x})'\mathbf{b} - \mathbf{Z}_i'\boldsymbol{\gamma})^2 K_h(X_i-\bar{x}),$$

$$\check{\boldsymbol{\theta}}_{Y+,p}(h) = \begin{bmatrix} \check{\boldsymbol{\beta}}_{Y+,p}(h) \\ \check{\boldsymbol{\gamma}}_{Y+,p}(h) \end{bmatrix} = \underset{\mathbf{b}\in\mathbb{R}^{1+p},\boldsymbol{\gamma}\in\mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}(X_i \geq \bar{x})(Y_i - \mathbf{r}_p(X_i-\bar{x})'\mathbf{b} - \mathbf{Z}_i'\boldsymbol{\gamma})^2 K_h(X_i-\bar{x}).$$

In words, we study now the estimator that includes first order interactions between the treatment variable T_i and the additional covariates Z_i. Using well known least-squares algebra, this is equivalent to fitting the two separate "long" regressions $\check{\boldsymbol{\theta}}_{Y-,p}(h)$ and $\check{\boldsymbol{\theta}}_{Y+,p}(h)$.

Using partitioned regression algebra, we have

$$\check{\boldsymbol{\beta}}_{Y-,p}(h) = \hat{\boldsymbol{\beta}}_{Y-,p}(h) - \hat{\boldsymbol{\beta}}_{Z-,p}(h)\check{\boldsymbol{\gamma}}_{Y-,p}(h),$$

$$\check{\boldsymbol{\beta}}_{Y+,p}(h) = \hat{\boldsymbol{\beta}}_{Y+,p}(h) - \hat{\boldsymbol{\beta}}_{Z+,p}(h)\check{\boldsymbol{\gamma}}_{Y+,p}(h),$$

and

$$\check{\boldsymbol{\gamma}}_{Y-,p}(h) = \check{\boldsymbol{\Gamma}}_{-,p}^{-1}(h)\check{\boldsymbol{\Upsilon}}_{Y-,p}(h),$$

$$\check{\boldsymbol{\gamma}}_{Y+,p}(h) = \check{\boldsymbol{\Gamma}}_{+,p}^{-1}(h)\tilde{\boldsymbol{\Upsilon}}_{Y+,p}^{\perp}(h),$$

where

$$\check{\boldsymbol{\Gamma}}_{-,p}(h) = \mathbf{Z}'\mathbf{K}_-(h)\mathbf{Z}/n - \boldsymbol{\Upsilon}_{Z-,p}(h)'\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\Upsilon}_{Z-,p}(h),$$

$$\check{\boldsymbol{\Gamma}}_{+,p}(h) = \mathbf{Z}'\mathbf{K}_+(h)\mathbf{Z}/n - \boldsymbol{\Upsilon}_{Z+,p}(h)'\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\Upsilon}_{Z+,p}(h),$$

$$\check{\boldsymbol{\Upsilon}}_{Y-,p}(h) = \mathbf{Z}'\mathbf{K}_-(h)\mathbf{Y}/n - \boldsymbol{\Upsilon}_{Z-,p}(h)'\boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\Upsilon}_{Y-,p}(h),$$

$$\check{\boldsymbol{\Upsilon}}_{Y+,p}(h) = \mathbf{Z}'\mathbf{K}_+(h)\mathbf{Y}/n - \boldsymbol{\Upsilon}_{Z+,p}(h)'\boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\Upsilon}_{Y+,p}(h).$$

This gives

$$\check{\eta}_{Y,\nu}(\mathbf{h}) = \hat{\tau}_{Y,\nu}(\mathbf{h}) - \left[\hat{\boldsymbol{\mu}}_{Z+,p}^{(\nu)}(h_+)'\tilde{\boldsymbol{\gamma}}_{Y+,p}(h_+) - \hat{\boldsymbol{\mu}}_{Z-,p}^{(\nu)}(h_-)'\tilde{\boldsymbol{\gamma}}_{Y-,p}(h_-)\right],$$

with

$$\hat{\boldsymbol{\mu}}_{Z-,p}^{(\nu)}(h)' = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{Z-,p}(h), \qquad \hat{\boldsymbol{\mu}}_{Z+,p}^{(\nu)}(h)' = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{Z+,p}(h).$$

8.1 Consistency and Identification

Recall that we showed that $\hat{\tau}_{Y,\nu}(\mathbf{h}) \to_{\mathbb{P}} \tau_{Y,\nu}$ and $\tilde{\tau}_{Y,\nu}(\mathbf{h}) \to_{\mathbb{P}} \tau_{Y,\nu}$, under the conditions of Lemma SA-7. In this section we show, under the same minimal continuity conditions, that $\check{\eta}_{Y,\nu}(\mathbf{h}) \to_{\mathbb{P}} \eta_{Y,\nu} \neq \tau_{Y,\nu}$ in general, and give a precise characterization of the probability limit.

Lemma SA-14 *Let the conditions of Lemma SA-7 hold. Then,*

$$\check{\eta}_{Y,\nu}(\mathbf{h}) \to_{\mathbb{P}} \eta_{Y,\nu} := \tau_{Y,\nu} - \left[\boldsymbol{\mu}_{Z+}^{(\nu)\prime}\boldsymbol{\gamma}_{Y+} - \boldsymbol{\mu}_{Z-}^{(\nu)\prime}\boldsymbol{\gamma}_{Y-}\right],$$

with

$$\gamma_{Y-} = \sigma_{Z-}^{-1}\mathbb{E}\left[(\mathbf{Z}_i(0) - \mu_{Z-}(X_i))Y_i(0) \middle| X_i = \bar{x} \right],$$

$$\gamma_{Y+} = \sigma_{Z+}^{-1}\mathbb{E}\left[(\mathbf{Z}_i(1) - \mu_{Z+}(X_i))Y_i(1) \middle| X_i = \bar{x} \right],$$

where recall that $\mu_{Z-} = \mu_{Z-}(\bar{x})$, $\mu_{Z+} = \mu_{Z+}(\bar{x})$, $\sigma^2_{Z-} = \sigma^2_{Z-}(\bar{x})$, *and* $\sigma^2_{Z+} = \sigma^2_{Z+}(\bar{x})$.

Proof of Lemma SA-14. We only prove the right-hand-side case (subindex "+"), since the other case is identical. Recall that the partitioned regression representation gives

$$\check{\beta}_{Y+,p}(h) = \hat{\beta}_{Y+,p}(h) - \hat{\beta}_{Z+,p}(h)\tilde{\gamma}_{Y+,p}(h),$$

where $\hat{\beta}_{Y+,p}(h) \to_{\mathbb{P}} \beta_{Y+,p}$ by Lemmas SA-2 and SA-3, and $\hat{\beta}_{Z+,p}(h) \to_{\mathbb{P}} \beta_{Z+,p}(h)$ by Lemmas SA-4 and SA-5. Therefore, it remains to show that $\tilde{\gamma}_{Y+,p}(h) = \check{\Gamma}^{-1}_{+,p}(h)\check{\Upsilon}_{Y+,p}(h) \to_{\mathbb{P}} \gamma_{Y+}$.

First, proceeding as in Lemmas SA-1, we have $\check{\Gamma}_{+,p}(h) \to_{\mathbb{P}} \kappa\sigma^2_{Z+}$. Second, proceeding analogously, we also have

$$\mathbf{Z}'\mathbf{K}_+(h)\mathbf{Y}/n \to_{\mathbb{P}} \kappa\mathbb{E}\left[\mathbf{Z}_i(1)Y_i(1) \middle| X_i = \bar{x} \right]$$

and

$$\Upsilon_{Z+,p}(h)'\Gamma^{-1}_{+,p}(h)\Upsilon_{Y+,p}(h) \to_{\mathbb{P}} \mu_Z\kappa'_{+,p}\Gamma^{-1}_{+,p}\kappa_{+,p}\mu_Y = \kappa\mu_Z\mu_Y.$$

The last two results imply

$$\begin{aligned}
\check{\Upsilon}_{Y+,p}(h) &= \mathbf{Z}'\mathbf{K}_+(h)\mathbf{Y}/n - \Upsilon_{Z+,p}(h)'\Gamma^{-1}_{+,p}(h)\Upsilon_{Y+,p}(h) \\
&= \kappa_+\mathbb{E}\left[(\mathbf{Z}_i(1) - \mu_Z(X_i))\mu_{Y+}(X_i, \mathbf{Z}_i(1)) \middle| X_i = \bar{x} \right] + o_{\mathbb{P}}(1).
\end{aligned}$$

This gives the final result. ∎

Example 1 *If, in addition, we assume*

$$\mathbb{E}[Y_i(0)|X_i = x, \mathbf{Z}_i(0)] = \xi_{Y-}(x) + \mathbf{Z}_i(0)'\delta_{Y-},$$

$$\mathbb{E}[Y_i(1)|X_i = x, \mathbf{Z}_i(1)] = \xi_{Y+}(x) + \mathbf{Z}_i(1)'\delta_{Y+},$$

which only needs to hold near the cutoff, we obtain the following result:

$$\eta_{Y,\nu} = \tau_{Y,\nu} - \left[\mu^{(\nu)\prime}_{Z+}\delta_{Y+} - \mu^{(\nu)\prime}_{Z-}\delta_{Y-} \right]$$

because

$$
\begin{aligned}
\boldsymbol{\gamma}_{Y+} &= \boldsymbol{\sigma}_{Z+}^{-1}\mathbb{E}\left[(\mathbf{Z}_i(1)-\boldsymbol{\mu}_{Z+}(X_i))Y_i(1)\middle|\, X_i=\bar{x}\right] \\
&= \boldsymbol{\sigma}_{Z+}^{-1}\mathbb{E}\left[(\mathbf{Z}_i(1)-\boldsymbol{\mu}_{Z+}(X_i))\mu_{Y+}(X_i,\mathbf{Z}_i(1))\middle|\, X_i=\bar{x}\right] \\
&= \boldsymbol{\sigma}_{Z+}^{-1}\mathbb{E}\left[(\mathbf{Z}_i(1)-\boldsymbol{\mu}_{Z+}(X_i))(\xi_{Y+}(X_i)+\mathbf{Z}_i(1)'\boldsymbol{\delta}_{Y+})\middle|\, X_i=\bar{x}\right] \\
&= \boldsymbol{\sigma}_{Z+}^{-1}\mathbb{E}\left[(\mathbf{Z}_i(1)-\boldsymbol{\mu}_{Z+}(X_i))\mathbf{Z}_i(1)'\middle|\, X_i=\bar{x}\right]\boldsymbol{\delta}_{Y+} \\
&= \boldsymbol{\delta}_{Y+},
\end{aligned}
$$

and, analogously, $\boldsymbol{\gamma}_{Y-}=\boldsymbol{\delta}_{Y-}$.

8.2 Demeaning Additional Regressors ($\nu=0$)

Let $\nu=0$. Consider now the following demeaned treatment-interacted covariate-adjusted sharp RD estimator:

$$\dot{\eta}_{Y,0}(\mathbf{h})=\mathbf{e}_0'\dot{\boldsymbol{\beta}}_{Y+,p}(h_-)-\mathbf{e}_0'\dot{\boldsymbol{\beta}}_{Y-,p}(h_+),$$

$$\begin{bmatrix}\dot{\boldsymbol{\beta}}_{Y-,p}(h)\\ \dot{\boldsymbol{\gamma}}_{Y-,p}(h)\end{bmatrix}=\underset{\mathbf{b}\in\mathbb{R}^{1+p},\boldsymbol{\gamma}\in\mathbb{R}^d}{\operatorname{argmin}}\sum_{i=1}^n \mathbb{1}(X_i<\bar{x})(Y_i-\mathbf{r}_p(X_i-\bar{x})'\mathbf{b}-(\mathbf{Z}_i-\bar{\mathbf{Z}})'\boldsymbol{\gamma})^2K_h(X_i-\bar{x}),$$

$$\begin{bmatrix}\dot{\boldsymbol{\beta}}_{Y+,p}(h)\\ \dot{\boldsymbol{\gamma}}_{Y+,p}(h)\end{bmatrix}=\underset{\mathbf{b}\in\mathbb{R}^{1+p},\boldsymbol{\gamma}\in\mathbb{R}^d}{\operatorname{argmin}}\sum_{i=1}^n \mathbb{1}(X_i\geq\bar{x})(Y_i-\mathbf{r}_p(X_i-\bar{x})'\mathbf{b}-(\mathbf{Z}_i-\bar{\mathbf{Z}})'\boldsymbol{\gamma})^2K_h(X_i-\bar{x}),$$

where

$$\bar{\mathbf{Z}}=\frac{1}{N}\sum_{i=1}^n \mathbb{1}(h_-\leq X_i-\bar{x}\leq h_+)\mathbf{Z}_i,\qquad N=\sum_{i=1}^n \mathbb{1}(h_-\leq X_i-\bar{x}\leq h_+).$$

This implies that

$$\bar{\mathbf{Z}}=\bar{\mathbf{Z}}_-+\bar{\mathbf{Z}}_+\to_{\mathbb{P}}\frac{1}{2}\boldsymbol{\mu}_{Z-}+\frac{1}{2}\boldsymbol{\mu}_{Z+},$$

because

$$\bar{\mathbf{Z}}_-=\frac{1}{N}\sum_{i=1}^n \mathbb{1}(h_-\leq X_i-\bar{x}<0)\mathbf{Z}_i\to_{\mathbb{P}}\frac{1}{2}\boldsymbol{\mu}_{Z-},$$

$$\bar{\mathbf{Z}}_+=\frac{1}{N}\sum_{i=1}^n \mathbb{1}(0\leq X_i-\bar{x}<h_+)\mathbf{Z}_i\to_{\mathbb{P}}\frac{1}{2}\boldsymbol{\mu}_{Z+}.$$

By standard least-squares algebra, we have

$$\check{\eta}_{Y,\nu}(\mathbf{h})=\dot{\eta}_{Y,0}(\mathbf{h})-\bar{\mathbf{Z}}'(\dot{\boldsymbol{\gamma}}_{Y+,p}(h_+)-\dot{\boldsymbol{\gamma}}_{Y-,p}(h_-)),\qquad \dot{\boldsymbol{\gamma}}_{Y-,p}(h)=\check{\boldsymbol{\gamma}}_{Y-,p}(h),\qquad \dot{\boldsymbol{\gamma}}_{Y+,p}(h)=\check{\boldsymbol{\gamma}}_{Y+,p}(h),$$

because only the first element of $\mathbf{b}$ (the intercept) is affected. Using the results in the previous sections, we obtain

$$
\begin{aligned}
\dot{\eta}_{Y,0}(\mathbf{h}) &= \check{\eta}_{Y,0}(\mathbf{h}) - \bar{\mathbf{Z}}'(\dot{\boldsymbol{\gamma}}_{Y+,p}(h_+) - \dot{\boldsymbol{\gamma}}_{Y-,p}(h_-)) \\
&= \hat{\tau}_{Y,0}(\mathbf{h}) - \left[\hat{\boldsymbol{\mu}}^{(0)}_{Z+,p}(h_+)'\dot{\boldsymbol{\gamma}}_{Y+,p}(h_+) - \hat{\boldsymbol{\mu}}^{(0)}_{Z-,p}(h_-)'\dot{\boldsymbol{\gamma}}_{Y-,p}(h_-)\right] + \bar{\mathbf{Z}}'(\dot{\boldsymbol{\gamma}}_{Y+,p}(h_+) - \dot{\boldsymbol{\gamma}}_{Y-,p}(h_-)) \\
&= \hat{\tau}_{Y,0}(\mathbf{h}) - \left[(\hat{\boldsymbol{\mu}}^{(0)}_{Z+,p}(h_+) - \bar{\mathbf{Z}})'\dot{\boldsymbol{\gamma}}_{Y+,p}(h_+) - (\hat{\boldsymbol{\mu}}^{(0)}_{Z-,p}(h_-) - \bar{\mathbf{Z}})'\dot{\boldsymbol{\gamma}}_{Y-,p}(h_-))\right].
\end{aligned}
$$

Therefore, using the results from the previous subsection, and assuming $\boldsymbol{\tau}_Z = \boldsymbol{\mu}_{Z+} - \boldsymbol{\mu}_{Z-} = \mathbf{0}$, we obtain

$$\dot{\eta}_{Y,0}(\mathbf{h}) = \hat{\tau}_{Y,0}(\mathbf{h}) + o_{\mathbb{P}}(1) \to_{\mathbb{P}} \tau_{Y,0},$$

provided that $\bar{\mathbf{Z}} \to_{\mathbb{P}} \boldsymbol{\mu}_{Z+} = \boldsymbol{\mu}_{Z-}$.

Establishing a valid distributional approximation, and other higher-order asymptotic results, for the estimator $\dot{\eta}_{Y,0}(\mathbf{h})$ is considerably much harder because of the presence of the (kernel regression) estimator $\bar{\mathbf{Z}}$. Furthermore, the above adjustment does not work for $\nu > 0$ (kink RD designs) because in this case the slopes should be appropriately demeaned.

Part III

Fuzzy RD Designs

We now allow for imperfect treatment compliance, and hence

$$T_i = T_i(0) \cdot \mathbb{1}(X_i < \bar{x}) + T_i(1) \cdot \mathbb{1}(X_i \geq \bar{x}),$$

that is, the treatment status T_i is no longer a deterministic function of the forcing variable X_i, but $\mathbb{P}[T_i = 1|X_i]$ still jumps at the RD threshold level $\bar{x}$. To be able to employ the same notation, assumptions and results given above for sharp RD designs, we recycle notation as follows:

$$Y_i(t) := Y_i(0) \cdot (1 - T_i(t)) + Y_i(1) \cdot T_i(t), \qquad t = 0, 1,$$

and

$$\mathbf{Z}_i(t) := \mathbf{Z}_i(0) \cdot (1 - T_i(t)) + \mathbf{Z}_i(1) \cdot T_i(t), \qquad t = 0, 1.$$

Through this section we employ the same notation introduced for the case of sharp RD designs. The main change is in the subindex indicating which outcome variable, Y or T, is being used to form estimands and estimators. In other words, we now have either 2 outcomes (standard RD setup) or $2 + d$ outcomes (covariate-adjusted RD setup).

To conserve some space, we do not provide proofs of the results presented in this section. They all follow the same logic used for sharp RD designs, after replacing the outcome variable and linear combination vector as appropriate.

9 Setup

9.1 Additional Notation

We employ the following additional vector notation

$$\mathbf{T} = [T_1, \cdots, T_n]', \qquad \mathbf{T}(0) = [T_1(0), \cdots, T_n(0)]', \qquad \mathbf{T}(1) = [T_1(1), \cdots, T_n(1)]'.$$

We then collect the outcome variables Y and T together:

$$\mathbf{U}_i = [Y_i, T_i]', \qquad \mathbf{U}_i(0) = [Y_i(0), T_i(0)]', \qquad \mathbf{U}_i(1) = [Y_i(1), T_i(1)]',$$

$$\mathbf{U} = [\mathbf{Y}, \mathbf{T}], \qquad \mathbf{U}(0) = [\mathbf{Y}(0), \mathbf{T}(0)], \qquad \mathbf{U}(1) = [\mathbf{Y}(1), \mathbf{T}(1)],$$

$$\boldsymbol{\mu}_{U-}(\mathbf{X}) = \mathbb{E}[\mathrm{vec}(\mathbf{U}(0))|\mathbf{X}], \qquad \boldsymbol{\mu}_{U+}(\mathbf{X}) = \mathbb{E}[\mathrm{vec}(\mathbf{U}(1))|\mathbf{X}],$$

$$\boldsymbol{\Sigma}_{U-} = \mathbb{V}[\mathrm{vec}(\mathbf{U}(0))|\mathbf{X}], \qquad \boldsymbol{\Sigma}_{U+} = \mathbb{V}[\mathrm{vec}(\mathbf{U}(1))|\mathbf{X}],$$

$$\boldsymbol{\mu}_{U-}(x) = \mathbb{E}[\mathbf{U}_i(0)|X_i = x], \qquad \boldsymbol{\mu}_{U+}(x) = \mathbb{E}[\mathbf{U}_i(1)|X_i = x],$$

$$\boldsymbol{\sigma}^2_{U-}(x) = \mathbb{V}[\mathbf{U}_i(0)|X_i = x], \qquad \boldsymbol{\sigma}^2_{U+}(x) = \mathbb{V}[\mathbf{U}_i(1)|X_i = x].$$

Recall that $\mathbf{e}_\nu$ denotes the conformable $(\nu+1)$-th unit vector, which may take different dimensions in different places. We also define:

$$\mu_{T-}(x) = \mathbb{E}[T_i(0)|X_i = x], \qquad \mu_{T+}(x) = \mathbb{E}[T_i(1)|X_i = x],$$

$$\sigma^2_{T-}(x) = \mathbb{V}[T_i(0)|X_i = x], \qquad \sigma^2_{T+}(x) = \mathbb{V}[T_i(1)|X_i = x].$$

In addition, to study fuzzy RD designs with covariates, we need to handle the joint distribution of the outcome variable and the additional covariates. Thus, we introduce the following additional notation:

$$\mathbf{F}_i = \left[Y_i, T_i, \mathbf{Z}_i'\right]', \qquad \mathbf{F}_i(0) = \left[Y_i(0), T_i(0), \mathbf{Z}_i(0)'\right]', \qquad \mathbf{F}_i(1) = \left[Y_i(1), T_i(0), \mathbf{Z}_i(1)'\right]',$$

$$\mathbf{F} = [\mathbf{Y}, \mathbf{T}, \mathbf{Z}], \qquad \mathbf{F}(0) = [\mathbf{Y}(0), \mathbf{T}(0), \mathbf{Z}(0)], \qquad \mathbf{F}(1) = [\mathbf{Y}(1), \mathbf{T}(1), \mathbf{Z}(1)],$$

$$\boldsymbol{\mu}_{F-}(\mathbf{X}) = \mathbb{E}[\mathrm{vec}(\mathbf{F}(0))|\mathbf{X}], \qquad \boldsymbol{\mu}_{F+}(\mathbf{X}) = \mathbb{E}[\mathrm{vec}(\mathbf{F}(1))|\mathbf{X}],$$

$$\boldsymbol{\Sigma}_{F-} = \mathbb{V}[\mathrm{vec}(\mathbf{F}(0))|\mathbf{X}], \qquad \boldsymbol{\Sigma}_{F+} = \mathbb{V}[\mathrm{vec}(\mathbf{F}(1))|\mathbf{X}],$$

$$\boldsymbol{\mu}_{F-}(x) = \mathbb{E}[\mathbf{F}_i(0)|X_i = x], \qquad \boldsymbol{\mu}_{F+}(x) = \mathbb{E}[\mathbf{F}_i(1)|X_i = x],$$

$$\boldsymbol{\sigma}^2_{F-}(x) = \mathbb{V}[\mathbf{F}_i(0)|X_i = x], \qquad \boldsymbol{\sigma}^2_{F+}(x) = \mathbb{V}[\mathbf{F}_i(1)|X_i = x].$$

9.2 Assumptions

We employ the following additional Assumption in this setting.

Assumption SA-4 (FRD, Standard) *For $\varrho \geq 1$, $x_l, x_u \in \mathbb{R}$ with $x_l < \bar{x} < x_u$, and all $x \in [x_l, x_u]$:*
(a) The Lebesgue density of X_i, denoted $f(x)$, is continuous and bounded away from zero.
(b) $\mu_{U-}(x)$ and $\mu_{U+}(x)$ are ϱ times continuously differentiable.
(c) $\sigma^2_{U-}(x)$ and $\sigma^2_{U+}(x)$ are continuous and invertible.
(d) $\mathbb{E}[|\mathbf{U}_i(t)|^4|X_i = x]$, $t \in \{0, 1\}$, are continuous.
(e) $\mu_{T-}(\bar{x}) \neq \mu_{T+}(\bar{x})$.

Assumption SA-5 (FRD, Covariates) *For $\varrho \geq 1$, $x_l, x_u \in \mathbb{R}$ with $x_l < \bar{x} < x_u$, and all $x \in [x_l, x_u]$:*
(a) $\mathbb{E}[\mathbf{Z}_i(0)\mathbf{U}_i(0)'|X_i = x]$ and $\mathbb{E}[\mathbf{Z}_i(1)\mathbf{U}_i(1)'|X_i = x]$ are continuously differentiable.
(b) $\boldsymbol{\mu}_{F-}(x)$ and $\boldsymbol{\mu}_{F+}(x)$ are ϱ times continuously differentiable.
(c) $\boldsymbol{\sigma}^2_{F-}(x)$ and $\boldsymbol{\sigma}^2_{F+}(x)$ are continuous and invertible.
(d) $\mathbb{E}[|\mathbf{F}_i(t)|^4|X_i = x]$, $t \in \{0, 1\}$, are continuous.
(e) $\boldsymbol{\mu}^{(\nu)}_{Z-}(\bar{x}) = \boldsymbol{\mu}^{(\nu)}_{Z+}(\bar{x})$.

9.3 Standard Fuzzy RD

Under Assumptions SA-2 and SA-4, the standard fuzzy RD estimand is

$$\varsigma_\nu = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}}, \qquad \tau_{Y,\nu} = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}, \qquad \tau_{T,\nu} = \mu_{T+}^{(\nu)} - \mu_{T-}^{(\nu)}, \qquad \nu \le S,$$

where, using the same notation introduced above,

$$\mu_{V+}^{(\nu)} = \mu_{V+}^{(\nu)}(\bar{x}) = \frac{\partial^\nu}{\partial x^\nu}\mu_{V+}(x)\bigg|_{x=\bar{x}}, \qquad \mu_{V-}^{(\nu)} = \mu_{V-}^{(\nu)}(\bar{x}) = \frac{\partial^\nu}{\partial x^\nu}\mu_{V-}(x)\bigg|_{x=\bar{x}}, \qquad V \in \{Y, T\}.$$

The standard fuzzy RD estimator for $\nu \le p$ is:

$$\hat{\varsigma}_\nu(\mathbf{h}) = \frac{\hat{\tau}_{Y,\nu}(\mathbf{h})}{\hat{\tau}_{T,\nu}(\mathbf{h})}, \qquad \hat{\tau}_{Y,\nu}(\mathbf{h}) = \hat{\mu}_{Y+,p}^{(\nu)}(h_+) - \hat{\mu}_{Y-,p}^{(\nu)}(h_-), \qquad \hat{\tau}_{T,\nu}(\mathbf{h}) = \hat{\mu}_{T+,p}^{(\nu)}(h_+) - \hat{\mu}_{T-,p}^{(\nu)}(h_-),$$

where, also using the same notation introduced above,

$$\hat{\mu}_{V+,p}^{(\nu)}(h) = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{V+,p}(h), \qquad \hat{\mu}_{V-,p}^{(\nu)}(h) = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{V-,p}(h), \qquad V \in \{Y, T\}.$$

9.4 Covariate-Adjusted Fuzzy RD

The covariate-adjusted fuzzy RD estimator for $\nu \le p$ is:

$$\tilde{\varsigma}_\nu(\mathbf{h}) = \frac{\tilde{\tau}_{Y,\nu}(\mathbf{h})}{\tilde{\tau}_{T,\nu}(\mathbf{h})}, \qquad \tilde{\tau}_{Y,\nu}(\mathbf{h}) = \tilde{\mu}_{Y+,p}^{(\nu)}(\mathbf{h}) - \tilde{\mu}_{Y-,p}^{(\nu)}(\mathbf{h}), \qquad \tilde{\tau}_{T,\nu}(\mathbf{h}) = \tilde{\mu}_{T+,p}^{(\nu)}(\mathbf{h}) - \tilde{\mu}_{T-,p}^{(\nu)}(\mathbf{h}),$$

where, also using the same notation introduced above,

$$\tilde{\mu}_{V-,p}^{(\nu)}(\mathbf{h}) = \nu!\mathbf{e}_\nu'\tilde{\boldsymbol{\theta}}_{V,p}(\mathbf{h}), \qquad \tilde{\mu}_{V+,p}^{(\nu)}(\mathbf{h}) = \nu!\mathbf{e}_{2+p+\nu}'\tilde{\boldsymbol{\theta}}_{V,p}(h),$$

with

$$\tilde{\boldsymbol{\theta}}_{V,p}(\mathbf{h}) = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_{V,p}(\mathbf{h}) \\ \tilde{\boldsymbol{\gamma}}_{V,p}(\mathbf{h}) \end{bmatrix}, \qquad V \in \{Y, T\}.$$

10 Inference Results

The fuzzy RD estimators, $\hat{\varsigma}_\nu(h)$ and $\tilde{\varsigma}_\nu(h)$, are a ratio of two sharp RD estimators, $(\hat{\tau}_{Y,\nu}(h), \hat{\tau}_{T,\nu}(h))$ and $(\tilde{\tau}_{Y,\nu}(h), \tilde{\tau}_{T,\nu}(h))$, respectively. Therefore, the fuzzy RD estimators are well defined whenever their underlying sharp RD estimators are, and the results from the previous section can be applied directly to established asymptotic invertibility of the corresponding matrices.

10.1 Consistency

Under Assumptions SA-1 and SA-4, and if $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$ and $\max\{h_-, h_+\} \to 0$, then

$$\hat{\varsigma}_\nu(\mathbf{h}) \to_{\mathbb{P}} \varsigma_\nu = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}},$$

which has been established in the literature before (e.g., Calonico, Cattaneo, and Titiunik (2014b)). Similarly, proceeding as in Lemma SA-7, after imposing assumptions SA-4 and SA-5, we obtain the following lemma.

Lemma SA-15 *Let Assumptions SA-1, SA-4 and SA-5 hold with $\varrho \geq p$. If $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$ and $\max\{h_-, h_+\} \to 0$, then*

$$\tilde{\varsigma}_\nu(\mathbf{h}) \to_{\mathbb{P}} \tilde{\varsigma}_\nu = \frac{\tau_{Y,\nu} - \left[\boldsymbol{\mu}_{Z+}^{(\nu)} - \boldsymbol{\mu}_{Z-}^{(\nu)}\right]' \boldsymbol{\gamma}_Y}{\tau_{T,\nu} - \left[\boldsymbol{\mu}_{Z+}^{(\nu)} - \boldsymbol{\mu}_{Z-}^{(\nu)}\right]' \boldsymbol{\gamma}_T},$$

with

$$\boldsymbol{\gamma}_V = \left[\boldsymbol{\sigma}_{Z-}^2 + \boldsymbol{\sigma}_{Z+}^2\right]^{-1} \left[\mathbb{E}[(\mathbf{Z}_i(0) - \boldsymbol{\mu}_{Z-}(X_i))V_i(0)|X_i = \bar{x}] + \mathbb{E}[(\mathbf{Z}_i(1) - \boldsymbol{\mu}_{Z+}(X_i))V_i(1)|X_i = \bar{x}]\right],$$

for $V \in \{Y, T\}$, and where recall that $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z-}(\bar{x})$, $\boldsymbol{\mu}_{Z+} = \boldsymbol{\mu}_{Z+}(\bar{x})$, $\boldsymbol{\sigma}_{Z-}^2 = \boldsymbol{\sigma}_{Z-}^2(\bar{x})$, and $\boldsymbol{\sigma}_{Z+}^2 = \boldsymbol{\sigma}_{Z+}^2(\bar{x})$.

10.2 Linear Approximation

To obtain MSE approximations, MSE-optimal bandwidths, and large sample distribution theory for the fuzzy RD estimators, we employ a linear approximation for these estimators. This approach gives a representation of the fuzzy RD estimators based on linear combinations of the underlying sharp RD estimators.

Specifically, using the identity

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) + \frac{a}{b^2\hat{b}}(\hat{b} - b)^2 - \frac{1}{b\hat{b}}(\hat{a} - a)(\hat{b} - b),$$

we have the following linearizations.

10.2.1 Standard Fuzzy RD Estimator

We have

$$\hat{\varsigma}_\nu(\mathbf{h}) - \varsigma_\nu = \frac{\hat{\tau}_{Y,\nu}(\mathbf{h})}{\hat{\tau}_{T,\nu}(\mathbf{h})} - \frac{\tau_{Y,\nu}}{\tau_{T,\nu}} = \mathbf{f}_{U,\nu}' \operatorname{vec}(\hat{\boldsymbol{\beta}}_{U,p}(\mathbf{h}) - \boldsymbol{\beta}_{U,p}) + \epsilon_{\varsigma,\nu},$$

with

$$\mathbf{f}_{U,\nu} = \begin{bmatrix} \frac{1}{\tau_{T,\nu}} \\ -\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2} \end{bmatrix} \otimes \nu! \mathbf{e}_\nu$$

and

$$\epsilon_{\varsigma,\nu} = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2 \hat{\tau}_{T,\nu}(\mathbf{h})}(\hat{\tau}_{T,\nu}(\mathbf{h}) - \tau_{T,\nu})^2 - \frac{1}{\tau_{T,\nu}\hat{\tau}_{T,\nu}(\mathbf{h})}(\hat{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu})(\hat{\tau}_{T,\nu}(\mathbf{h}) - \tau_{T,\nu}).$$

Therefore, under the assumptions above, it follows from previous lemmas that

$$\epsilon_{\varsigma,\nu} = O_{\mathbb{P}}\left(\frac{1}{nh^{1+2\nu}} + h^{2(1+p-\nu)}\right) = o_{\mathbb{P}}(1),$$

provided that $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$ and $\max\{h_-, h_+\} \to 0$, and the assumptions imposed hold.

Recall that

$$\hat{\boldsymbol{\beta}}_{U,p}(\mathbf{h}) = \hat{\boldsymbol{\beta}}_{U+,p}(h_+) - \hat{\boldsymbol{\beta}}_{U-,p}(h_-), \qquad \boldsymbol{\beta}_{U,p} = \boldsymbol{\beta}_{U+,p} - \boldsymbol{\beta}_{U-,p},$$

with

$$\hat{\boldsymbol{\beta}}_{U-,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)\mathbf{U}, \qquad \hat{\boldsymbol{\beta}}_{U+,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)\mathbf{U},$$

or, in vectorized form,

$$\operatorname{vec}(\hat{\boldsymbol{\beta}}_{U-,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_2 \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)]\operatorname{vec}(\mathbf{U}),$$

$$\operatorname{vec}(\hat{\boldsymbol{\beta}}_{U+,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_2 \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)]\operatorname{vec}(\mathbf{U}).$$

Thus, we have

$$\hat{\boldsymbol{\mu}}_{U-,p}^{(\nu)}(h)' = [\hat{\mu}_{Y-,p}^{(\nu)}(h), \hat{\mu}_{T-,p}^{(\nu)}(h)] = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{U-,p}(h),$$

$$\hat{\boldsymbol{\mu}}_{U+,p}^{(\nu)}(h)' = [\hat{\mu}_{Y+,p}^{(\nu)}(h), \hat{\mu}_{T+,p}^{(\nu)}(h)] = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{U+,p}(h),$$

$$\boldsymbol{\mu}_{U-}^{(\nu)\prime} = [\mu_{Y-}^{(\nu)}, \mu_{T-}^{(\nu)}] = \nu!\mathbf{e}_\nu'\boldsymbol{\beta}_{U-,p}, \qquad \boldsymbol{\mu}_{U+}^{(\nu)\prime} = [\mu_{Y+}^{(\nu)}, \mu_{T+}^{(\nu)}] = \nu!\mathbf{e}_\nu'\boldsymbol{\beta}_{U+,p}.$$

10.2.2 Covariate-Adjusted Fuzzy RD estimator

We have

$$\tilde{\varsigma}_\nu(\mathbf{h}) - \varsigma_\nu = \frac{\tilde{\tau}_{Y,\nu}(\mathbf{h})}{\tilde{\tau}_{T,\nu}(\mathbf{h})} - \frac{\tau_{Y,\nu}}{\tau_{T,\nu}} = \mathbf{f}_{F,\nu}(\mathbf{h})'\operatorname{vec}(\hat{\boldsymbol{\beta}}_{F,p}(\mathbf{h}) - \boldsymbol{\beta}_{F,p}) + \epsilon_{\tilde{\varsigma},\nu},$$

with (see Lemma SA-15)

$$\mathbf{f}_{F,\nu}(\mathbf{h}) = \begin{bmatrix} \frac{1}{\tau_{T,\nu}} \\ -\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2} \\ -\frac{1}{\tau_{T,\nu}}\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h}) + \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\tilde{\boldsymbol{\gamma}}_{T,p}(\mathbf{h}) \end{bmatrix} \otimes \nu!\mathbf{e}_\nu \to_{\mathbb{P}} \mathbf{f}_{F,\nu} = \begin{bmatrix} \frac{1}{\tau_{T,\nu}} \\ -\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2} \\ -\frac{1}{\tau_{T,\nu}}\boldsymbol{\gamma}_{Y,p} + \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\boldsymbol{\gamma}_{T,p} \end{bmatrix} \otimes \nu!\mathbf{e}_\nu$$

and

$$\epsilon_{\tilde{\varsigma},\nu} = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2\tilde{\tau}_{T,\nu}(\mathbf{h})}(\tilde{\tau}_{T,\nu}(\mathbf{h}) - \tau_{T,\nu})^2 - \frac{1}{\tau_{T,\nu}\tilde{\tau}_{T,\nu}(\mathbf{h})}(\tilde{\tau}_{Y,\nu}(\mathbf{h}) - \tau_{Y,\nu})(\tilde{\tau}_{T,\nu}(\mathbf{h}) - \tau_{T,\nu}).$$

Therefore, under the assumptions above, it follows from previous lemmas that

$$\epsilon_{\tilde{\varsigma},\nu} = O_{\mathbb{P}}\left(\frac{1}{nh^{1+2\nu}} + h^{2(1+p-\nu)}\right) = o_{\mathbb{P}}(1),$$

provided that $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$ and $\max\{h_-, h_+\} \to 0$, and the assumptions imposed hold.

Recall that

$$\hat{\boldsymbol{\beta}}_{F,p}(\mathbf{h}) = \hat{\boldsymbol{\beta}}_{F+,p}(h_+) - \hat{\boldsymbol{\beta}}_{F-,p}(h_-), \qquad \boldsymbol{\beta}_{F,p} = \boldsymbol{\beta}_{F+,p} - \boldsymbol{\beta}_{F-,p},$$

with

$$\hat{\boldsymbol{\beta}}_{F-,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)\mathbf{F}, \qquad \hat{\boldsymbol{\beta}}_{F+,p}(h) = \frac{1}{\sqrt{nh}}\mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)\mathbf{F},$$

or, in vectorized form,

$$\operatorname{vec}(\hat{\boldsymbol{\beta}}_{F-,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_2 \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{-,p}(h)] \operatorname{vec}(\mathbf{F}),$$

$$\operatorname{vec}(\hat{\boldsymbol{\beta}}_{F+,p}(h)) = \frac{1}{\sqrt{nh}}[\mathbf{I}_2 \otimes \mathbf{H}_p^{-1}(h)\mathbf{P}_{+,p}(h)] \operatorname{vec}(\mathbf{F}).$$

Thus, we have

$$\hat{\boldsymbol{\mu}}_{F-,p}^{(\nu)}(h)' = [\hat{\mu}_{Y-,p}^{(\nu)}(h), \hat{\mu}_{T-,p}^{(\nu)}(h), \hat{\boldsymbol{\mu}}_{Z-,p}^{(\nu)}(h)'] = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{F-,p}(h),$$

$$\hat{\boldsymbol{\mu}}_{F+,p}^{(\nu)}(h)' = [\hat{\mu}_{Y+,p}^{(\nu)}(h), \hat{\mu}_{T+,p}^{(\nu)}(h), \hat{\boldsymbol{\mu}}_{Z+,p}^{(\nu)}(h)'] = \nu!\mathbf{e}_\nu'\hat{\boldsymbol{\beta}}_{F+,p}(h),$$

$$\boldsymbol{\mu}_{F-}^{(\nu)\prime} = [\mu_{Y-}^{(\nu)}, \mu_{T-}^{(\nu)}, \boldsymbol{\mu}_{Z-}^{(\nu)\prime}] = \nu!\mathbf{e}_\nu'\boldsymbol{\beta}_{F-,p}, \qquad \boldsymbol{\mu}_{F+}^{(\nu)\prime} = [\mu_{Y+}^{(\nu)}, \mu_{T+}^{(\nu)}, \boldsymbol{\mu}_{Z+}^{(\nu)\prime}] = \nu!\mathbf{e}_\nu'\boldsymbol{\beta}_{F+,p}.$$

Therefore, all the results discussed for covariate-adjusted sharp RD designs can be applied to fuzzy RD designs, provided that the vector of outcome variables $\mathbf{S}_i$ is replaced by $\mathbf{F}_i$, and the appropriate linear combination is used (e.g., $\mathbf{s}_{S,\nu}(\mathbf{h})$ is replaced by $\mathbf{f}_{F,\nu}(\mathbf{h})$).

10.3 Conditional Bias

We characterize the smoothing bias of $\{\hat{\boldsymbol{\beta}}_{U-,p}(h), \hat{\boldsymbol{\beta}}_{U+,p}(h)\}$ and $\{\hat{\boldsymbol{\beta}}_{F-,p}(h), \hat{\boldsymbol{\beta}}_{F+,p}(h)\}$, the main ingredients entering the standard fuzzy RD estimator $\varsigma_\nu(\mathbf{h})$ and the covariate-adjusted sharp RD estimator $\tilde{\varsigma}_\nu(\mathbf{h})$, respectively. Observe that

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{V-,p}(h)|\mathbf{X}] = [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)]\mathbb{E}[\mathbf{V}(0)|\mathbf{X}]/n,$$

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_{V+,p}(h)|\mathbf{X}] = [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)]\mathbb{E}[\mathbf{V}(1)|\mathbf{X}]/n,$$

for $V \in \{U, F\}$.

Lemma SA-16 *Let assumptions SA-1, SA-4 and SA-5 hold with $\varrho \geq p+2$, and $nh \to \infty$ and $h \to 0$. Then, $V \in \{U, F\}$,*

$$\begin{aligned}
&\mathbb{E}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{V-,p}(h))|\mathbf{X}] \\
&= \mathrm{vec}(\boldsymbol{\beta}_{V-,p}) + [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)] \left[h^{1+p}\mathbf{B}_{V-,p,p}(h) + h^{2+p}\mathbf{B}_{V-,p,p+1}(h) + o_{\mathbb{P}}(h^{2+p})\right],
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{V+,p}(h))|\mathbf{X}] \\
&= \mathrm{vec}(\boldsymbol{\beta}_{V+,p}) + [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)] \left[h^{1+p}\mathbf{B}_{V+,p,p}(h) + h^{2+p}\mathbf{B}_{V+,p,p+1}(h) + o_{\mathbb{P}}(h^{2+p})\right],
\end{aligned}$$

where

$$\mathbf{B}_{V-,p,a}(h) = [\mathbf{I}_{1+d} \otimes \boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p,a}(h)]\frac{\boldsymbol{\mu}_{V-}^{(1+a)}}{(1+a)!} \to_{\mathbb{P}} \mathbf{B}_{V-,p,a} = [\mathbf{I}_{1+d} \otimes \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\vartheta}_{-,p,a}]\frac{\boldsymbol{\mu}_{V-}^{(1+a)}}{(1+a)!},$$

$$\mathbf{B}_{V+,p,a}(h) = [\mathbf{I}_{1+d} \otimes \boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\vartheta}_{+,p,a}(h)]\frac{\boldsymbol{\mu}_{V+}^{(1+a)}}{(1+a)!} \to_{\mathbb{P}} \mathbf{B}_{V+,p,a} = [\mathbf{I}_{1+d} \otimes \boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\vartheta}_{+,p,a}]\frac{\boldsymbol{\mu}_{V+}^{(1+a)}}{(1+a)!}.$$

10.4 Conditional Variance

We characterize the exact, fixed-n (conditional) variance formulas of the main ingredients entering the standard fuzzy RD estimator $\varsigma_\nu(\mathbf{h})$ and the covariate-adjusted sharp RD estimator $\tilde{\varsigma}_\nu(\mathbf{h})$. These terms are $\mathbb{V}[\hat{\boldsymbol{\beta}}_{V-,p}(h)|\mathbf{X}]$ and $\mathbb{V}[\hat{\boldsymbol{\beta}}_{V+,p}(h)|\mathbf{X}]$, for $V \in \{U, F\}$.

Lemma SA-17 *Let assumptions SA-1, SA-2 and SA-3 hold, and $nh \to \infty$ and $h \to 0$. Then, for $V \in \{U, F\}$,*

$$\begin{aligned}
&\mathbb{V}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{V-,p}(h))|\mathbf{X}] \\
&= [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)]\boldsymbol{\Sigma}_{V-}[\mathbf{I}_{1+d} \otimes \mathbf{K}_-(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{H}_p^{-1}(h)]/n^2 \\
&= \frac{1}{nh}[\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)][\mathbf{I}_{1+d} \otimes \mathbf{P}_{-,p}(h)]\boldsymbol{\Sigma}_{V-}[\mathbf{I}_{1+d} \otimes \mathbf{P}_{-,p}(h)'][\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)],
\end{aligned}$$

$$\begin{aligned}
&\mathbb{V}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{V+,p}(h))|\mathbf{X}] \\
&= [\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)]\boldsymbol{\Sigma}_{V+}[\mathbf{I}_{1+d} \otimes \mathbf{K}_+(h)\mathbf{R}_p(h)\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{H}_p^{-1}(h)]/n^2 \\
&= \frac{1}{nh}[\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)][\mathbf{I}_{1+d} \otimes \mathbf{P}_{+,p}(h)]\boldsymbol{\Sigma}_{V+}[\mathbf{I}_{1+d} \otimes \mathbf{P}_{+,p}(h)'][\mathbf{I}_{1+d} \otimes \mathbf{H}_p^{-1}(h)],
\end{aligned}$$

with

$$nh[\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\mathbb{V}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{V-,p}(h))|\mathbf{X}][\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\to_{\mathbb{P}}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}]\boldsymbol{\Psi}_{V-,p}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{-,p}^{-1}],$$

$$nh[\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\mathbb{V}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{V+,p}(h))|\mathbf{X}][\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)]\to_{\mathbb{P}}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}]\boldsymbol{\Psi}_{V+,p}[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}_{+,p}^{-1}],$$

where

$$\boldsymbol{\Psi}_{V-,p}=f(\bar{x})\left[\boldsymbol{\sigma}_{V-}^2\otimes\int_{-\infty}^0\mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2du\right],\qquad\boldsymbol{\sigma}_{V-}^2=\boldsymbol{\sigma}_{V-}^2(\bar{x})=\mathbb{V}[\mathbf{V}_i(0)|X_i=\bar{x}],$$

and

$$\boldsymbol{\Psi}_{V+,p}=f(\bar{x})\left[\boldsymbol{\sigma}_{V+}^2\otimes\int_0^{\infty}\mathbf{r}_p(u)\mathbf{r}_p(u)'K(u)^2du\right],\qquad\boldsymbol{\sigma}_{V+}^2=\boldsymbol{\sigma}_{V+}^2(\bar{x})=\mathbb{V}[\mathbf{V}_i(1)|X_i=\bar{x}].$$

10.5 Convergence Rates

Furthermore, because the results in the previous section apply immediately to the numerator and denominator of the fuzzy RD estimators. Furthermore, the results above imply that

$$[\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)](\hat{\boldsymbol{\beta}}_{V-,p}(h)-\boldsymbol{\beta}_{V-,p})=O_{\mathbb{P}}\left(h^{1+p}+\frac{1}{\sqrt{nh}}\right),$$

$$[\mathbf{I}_{1+d}\otimes\mathbf{H}_p(h)](\hat{\boldsymbol{\beta}}_{V+,p}(h)-\boldsymbol{\beta}_{V+,p})=O_{\mathbb{P}}\left(h^{1+p}+\frac{1}{\sqrt{nh}}\right),$$

for $V\in\{U,F\}$.

Furthermore, the vector of linear combinations satisfy

$$\mathbf{f}_{F,\nu}(\mathbf{h})=\begin{bmatrix}\frac{1}{\tau_{T,\nu}}\\-\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\\-\frac{1}{\tau_{T,\nu}}\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})+\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\tilde{\boldsymbol{\gamma}}_{T,p}(\mathbf{h})\end{bmatrix}\otimes\nu!\mathbf{e}_\nu\to_{\mathbb{P}}\mathbf{f}_{F,\nu}=\begin{bmatrix}\frac{1}{\tau_{T,\nu}}\\-\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\\-\frac{1}{\tau_{T,\nu}}\boldsymbol{\gamma}_{Y,p}+\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\boldsymbol{\gamma}_{T,p}\end{bmatrix}\otimes\nu!\mathbf{e}_\nu$$

and

$$\hat{\mathbf{f}}_{F,\nu}(\mathbf{h})=\begin{bmatrix}\frac{1}{\tilde{\tau}_{T,\nu}(\mathbf{h})}\\-\frac{\tilde{\tau}_{Y,\nu}(\mathbf{h})}{\tilde{\tau}_{T,\nu}^2(\mathbf{h})}\\-\frac{1}{\tau_{T,\nu}}\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})+\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\tilde{\boldsymbol{\gamma}}_{T,p}(\mathbf{h})\end{bmatrix}\otimes\nu!\mathbf{e}_\nu\to_{\mathbb{P}}\mathbf{f}_{F,\nu}=\begin{bmatrix}\frac{1}{\tau_{T,\nu}}\\-\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\\-\frac{1}{\tau_{T,\nu}}\boldsymbol{\gamma}_{Y,p}+\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\boldsymbol{\gamma}_{T,p}\end{bmatrix}\otimes\nu!\mathbf{e}_\nu$$

provided that Assumptions SA-1, SA-4 and SA-5 hold, and $nh^{1+2\nu}\to\infty$ and $h\to0$. Similarly, under the same conditions,

$$\hat{\mathbf{f}}_{U,\nu}(\mathbf{h})=\begin{bmatrix}\frac{1}{\hat{\tau}_{T,\nu}(\mathbf{h})}\\-\frac{\hat{\tau}_{Y,\nu}(\mathbf{h})}{\hat{\tau}_{T,\nu}^2(\mathbf{h})}\end{bmatrix}\otimes\nu!\mathbf{e}_\nu\to_{\mathbb{P}}\mathbf{f}_{U,\nu}=\begin{bmatrix}\frac{1}{\tau_{T,\nu}}\\-\frac{\tau_{Y,\nu}}{\tau_{T,\nu}^2}\end{bmatrix}\otimes\nu!\mathbf{e}_\nu.$$

10.6 Bias Approximation

10.6.1 Standard Sharp RD Estimator

We have

$$\mathsf{Bias}[\hat{\varsigma}_{-,\nu}(h)] = \mathbb{E}[\mathbf{f}'_{U,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U-,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{U-,p})]|\mathbf{X}],$$

$$\mathsf{Bias}[\hat{\varsigma}_{+,\nu}(h)] = \mathbb{E}[\mathbf{f}'_{U,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U+,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{U+,p})]|\mathbf{X}],$$

and therefore

$$\mathsf{Bias}[\hat{\varsigma}^{(\nu)}_{-,p}(h)] = h^{1+p-\nu}\mathcal{B}_{U-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

$$\mathsf{Bias}[\hat{\varsigma}^{(\nu)}_{+,p}(h)] = h^{1+p-\nu}\mathcal{B}_{U-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

where

$$\mathcal{B}_{U-,\nu,p}(h) = \mathbf{f}'_{U,\nu}\mathbf{B}_{U-,p}(h) \to_{\mathbb{P}} \mathcal{B}_{U-,\nu,p} = \mathbf{f}'_{U,\nu}\mathbf{B}_{U-,p},$$

$$\mathcal{B}_{U+,\nu,p}(h) = \mathbf{f}'_{U,\nu}\mathbf{B}_{U+,p}(h) \to_{\mathbb{P}} \mathcal{B}_{U-,\nu,p} = \mathbf{f}'_{U,\nu}\mathbf{B}_{U+,p}.$$

Therefore, we define

$$\mathsf{Bias}[\tilde{\varsigma}_{\nu}(\mathbf{h})] = \mathbb{E}[\mathbf{f}'_{U,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U,p}(\mathbf{h})) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{U,p})]|\mathbf{X}]$$

and, using the results above,

$$\mathbb{E}[\hat{\varsigma}_{\nu}(\mathbf{h})|\mathbf{X}] = h_+^{1+p-\nu}\mathcal{B}_{U+,\nu,p}(h_+) - h_-^{1+p-\nu}\mathcal{B}_{U-,\nu,p}(h_-) + o_{\mathbb{P}}(\max\{h_-^{2+p-\nu}, h_+^{2+p-\nu}\}).$$

10.6.2 Covariate-Adjusted Fuzzy RD Estimator

Using the linear approximation, we define

$$\mathsf{Bias}[\tilde{\varsigma}^{(\nu)}_{-,p}(h)] = \mathbb{E}[\mathbf{f}'_{F,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F-,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F-,p})]|\mathbf{X}],$$

$$\mathsf{Bias}[\tilde{\varsigma}^{(\nu)}_{+,p}(h)] = \mathbb{E}[\mathbf{f}'_{F,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F+,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F+,p})]|\mathbf{X}],$$

and therefore

$$\mathsf{Bias}[\tilde{\varsigma}^{(\nu)}_{-,p}(h)] = h^{1+p-\nu}\mathcal{B}_{F-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

$$\mathsf{Bias}[\tilde{\varsigma}^{(\nu)}_{+,p}(h)] = h^{1+p-\nu}\mathcal{B}_{F-,\nu,p}(h) + o_{\mathbb{P}}(h^{1+p-\nu}),$$

where

$$\mathcal{B}_{F-,\nu,p}(h) = \mathbf{f}'_{F,\nu}\mathbf{B}_{F-,p}(h) \to_{\mathbb{P}} \mathcal{B}_{S-,\nu,p} = \mathbf{f}'_{F,\nu}\mathbf{B}_{F-,p},$$

$$\mathcal{B}_{F+,\nu,p}(h) = \mathbf{f}'_{F,\nu}\mathbf{B}_{F+,p}(h) \to_{\mathbb{P}} \mathcal{B}_{S-,\nu,p} = \mathbf{f}'_{F,\nu}\mathbf{B}_{F+,p}.$$

Therefore, we define

$$\mathsf{Bias}[\tilde{\varsigma}_{\nu}(\mathbf{h})] = \mathbb{E}[\mathbf{f}'_{F,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p}(\mathbf{h})) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p})]|\mathbf{X}]$$

and, using the results above,

$$\mathsf{Bias}[\tilde{\varsigma}_\nu(\mathbf{h})] = h_+^{1+p-\nu}\mathcal{B}_{F+,\nu,p}(h_+) - h_-^{1+p-\nu}\mathcal{B}_{F-,\nu,p}(h_-) + o_{\mathbb{P}}(\max\{h_-^{2+p-\nu}, h_+^{2+p-\nu}\}).$$

10.7 Variance Approximation

10.7.1 Standard Fuzzy RD Estimator

We define

$$\begin{aligned}\mathsf{Var}[\hat{\varsigma}_\nu(\mathbf{h})] &= \mathbb{V}[\mathbf{f}'_{U,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U,p}(\mathbf{h})) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{U,p})]|\mathbf{X}] \\ &= \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{U-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{U+,\nu,p}(h_+)\end{aligned}$$

with

$$\mathcal{V}_{U-,\nu,p}(h) = \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \mathbf{P}_{-,p}(h)]\boldsymbol{\Sigma}_{U-}[\mathbf{I}_2 \otimes \mathbf{P}_{-,p}(h)']\mathbf{f}_{U,\nu},$$

$$\mathcal{V}_{U+,\nu,p}(h) = \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \mathbf{P}_{+,p}(h)]\boldsymbol{\Sigma}_{U+}[\mathbf{I}_2 \otimes \mathbf{P}_{+,p}(h)']\mathbf{f}_{U,\nu}.$$

Furthermore,

$$\mathcal{V}_{U-,\nu,p}(h) \to_{\mathbb{P}} \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{-,p}^{-1}]\boldsymbol{\Psi}_{U-,p}[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{-,p}^{-1}]\mathbf{f}_{U,\nu} =: \mathcal{V}_{U-,\nu,p},$$

$$\mathcal{V}_{U+,\nu,p}(h) \to_{\mathbb{P}} \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{+,p}^{-1}]\boldsymbol{\Psi}_{U+,p}[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{+,p}^{-1}]\mathbf{f}_{U,\nu} =: \mathcal{V}_{U+,\nu,p},$$

10.7.2 Covariate-Adjusted Fuzzy RD Estimator

We define

$$\begin{aligned}\mathsf{Var}[\tilde{\varsigma}_\nu(\mathbf{h})] &= \mathbb{V}[\mathbf{f}'_{F,\nu}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p}(\mathbf{h})) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p})]|\mathbf{X}] \\ &= \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{F-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{F+,\nu,p}(h_+)\end{aligned}$$

with

$$\mathcal{V}_{F-,\nu,p}(h) = \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h)]\boldsymbol{\Sigma}_{F-}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h)']\mathbf{f}_{F,\nu},$$

$$\mathcal{V}_{F+,\nu,p}(h) = \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h)]\boldsymbol{\Sigma}_{F+}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h)']\mathbf{f}_{F,\nu}.$$

Furthermore,

$$\mathcal{V}_{F-,\nu,p}(h) \to_{\mathbb{P}} \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \boldsymbol{\Gamma}_{-,p}^{-1}]\boldsymbol{\Psi}_{F-,p}[\mathbf{I}_{2+d} \otimes \boldsymbol{\Gamma}_{-,p}^{-1}]\mathbf{f}_{F,\nu} =: \mathcal{V}_{F-,\nu,p},$$

$$\mathcal{V}_{F+,\nu,p}(h) \to_{\mathbb{P}} \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \boldsymbol{\Gamma}_{+,p}^{-1}]\boldsymbol{\Psi}_{F+,p}[\mathbf{I}_{2+d} \otimes \boldsymbol{\Gamma}_{+,p}^{-1}]\mathbf{f}_{F,\nu} =: \mathcal{V}_{F+,\nu,p},$$

10.8 MSE Expansions

For related results see Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014b), Arai and Ichimura (2016), and references therein.

10.8.1 Standard Fuzzy RD Estimator

- **MSE expansion: One-sided**. We define

$$\mathsf{MSE}[\hat{\varsigma}_{-,\nu}(h)] = \mathbb{E}[(\mathbf{f}'_{U,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U-,p}(h)) - \mathrm{vec}(\boldsymbol{\beta}_{U-,p})])^2|\mathbf{X}],$$

$$\mathsf{MSE}[\hat{\varsigma}_{+,\nu}(h)] = \mathbb{E}[(\mathbf{f}'_{U,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U+,p}(h)) - \mathrm{vec}(\boldsymbol{\beta}_{U+,p})])^2|\mathbf{X}].$$

Then, we have:

$$\begin{aligned}\mathsf{MSE}[\hat{\varsigma}_{-,\nu}(h)] &= h^{2(1+p-\nu)}\mathcal{B}^2_{U-,\nu,p}(h)\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{U-,\nu,p}(h)\\ &= h^{2(1+p-\nu)}\mathcal{B}^2_{U-,\nu,p}\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{U-,\nu,p}\{1+o_{\mathbb{P}}(1)\}\end{aligned}$$

and

$$\begin{aligned}\mathsf{MSE}[\hat{\varsigma}_{+,\nu}(h)] &= h^{2(1+p-\nu)}\mathcal{B}^2_{U+,\nu,p}(h)\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{U+,\nu,p}(h)\\ &= h^{2(1+p-\nu)}\mathcal{B}^2_{U+,\nu,p}\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{U+,\nu,p}\{1+o_{\mathbb{P}}(1)\}\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{U-,\nu,p} \neq 0$ and $\mathcal{B}_{U+,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{U-,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)}\frac{\mathcal{V}_{U-,\nu,p}/n}{\mathcal{B}^2_{U-,\nu,p}}\right]^{\frac{1}{3+2p}} \qquad \textit{and} \qquad \mathfrak{h}_{U+,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)}\frac{\mathcal{V}_{U+,\nu,p}/n}{\mathcal{B}^2_{U+,\nu,p}}\right]^{\frac{1}{3+2p}}.$$

- **MSE expansion: Sum/Difference**. Let $h = h_+ = h_-$. We define

$$\mathsf{MSE}[\hat{\varsigma}_{+,\nu}(h) \pm \hat{\varsigma}_{-,\nu}(h)] = \mathbb{E}[(\mathbf{f}'_{U,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{U,p}(h)) \pm \mathrm{vec}(\boldsymbol{\beta}_{U,p})])^2|\mathbf{X}]$$

Then, we have:

$$\begin{aligned}&\mathsf{MSE}[\hat{\varsigma}_{+,\nu}(h) \pm \hat{\varsigma}_{-,\nu}(h)]\\ &= h^{2(1+p-\nu)}\left[\mathcal{B}_{U+,\nu,p}(h) \pm \mathcal{B}_{U-,\nu,p}(h)\right]^2\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\left[\mathcal{V}_{U-,\nu,p}(h) + \mathcal{V}_{U+,\nu,p}(h)\right]\\ &= h^{2(1+p-\nu)}\left[\mathcal{B}_{U+,\nu,p} \pm \mathcal{B}_{U-,\nu,p}\right]^2\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\left[\mathcal{V}_{U-,\nu,p} + \mathcal{V}_{U+,\nu,p}\right]\{1+o_{\mathbb{P}}(1)\}.\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{U+,\nu,p} \pm \mathcal{B}_{U-,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{\Delta U,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\mathcal{V}_{U-,\nu,p} + \mathcal{V}_{U+,\nu,p})/n}{(\mathcal{B}_{U+,\nu,p} - \mathcal{B}_{U-,\nu,p})^2}\right]^{\frac{1}{3+2p}},$$

$$\mathfrak{h}_{\Sigma U,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\mathcal{V}_{U-,\nu,p} + \mathcal{V}_{U+,\nu,p})/n}{(\mathcal{B}_{U+,\nu,p} + \mathcal{B}_{U-,\nu,p})^2}\right]^{\frac{1}{3+2p}}.$$

Note that, when $h = h_+ = h_-$,

$$\mathsf{MSE}[\hat{\varsigma}_\nu(\mathbf{h})] = \mathsf{MSE}[\hat{\varsigma}_{+,\nu}(h) - \hat{\varsigma}_{-,\nu}(h)].$$

10.8.2 Covariate-Adjusted Fuzzy RD Estimator

- **MSE expansion: One-sided**. We define

$$\mathsf{MSE}[\tilde{\varsigma}_{-,\nu}(h)] = \mathbb{E}[(\mathbf{f}'_{F,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F-,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F-,p})])^2|\mathbf{X}],$$

$$\mathsf{MSE}[\tilde{\varsigma}_{+,\nu}(h)] = \mathbb{E}[(\mathbf{f}'_{F,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F+,p}(h)) - \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F+,p})])^2|\mathbf{X}].$$

Then, we have:

$$\begin{aligned}\mathsf{MSE}[\tilde{\varsigma}_{-,\nu}(h)] &= h^{2(1+p-\nu)}\mathcal{B}^2_{F-,\nu,p}(h)\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{F-,\nu,p}(h) \\ &= h^{2(1+p-\nu)}\mathcal{B}^2_{F-,\nu,p}\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{F-,\nu,p}\{1+o_{\mathbb{P}}(1)\}\end{aligned}$$

and

$$\begin{aligned}\mathsf{MSE}[\tilde{\varsigma}_{+,\nu}(h)] &= h^{2(1+p-\nu)}\mathcal{B}^2_{F+,\nu,p}(h)\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{F+,\nu,p}(h) \\ &= h^{2(1+p-\nu)}\mathcal{B}^2_{F+,\nu,p}\{1+o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}}\mathcal{V}_{F+,\nu,p}\{1+o_{\mathbb{P}}(1)\}\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{F-,\nu,p} \neq 0$ and $\mathcal{B}_{F+,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{F-,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{\mathcal{V}_{F-,\nu,p}/n}{\mathcal{B}^2_{F-,\nu,p}}\right]^{\frac{1}{3+2p}} \quad \textit{and} \quad \mathfrak{h}_{F+,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{\mathcal{V}_{F+,\nu,p}/n}{\mathcal{B}^2_{F+,\nu,p}}\right]^{\frac{1}{3+2p}}.$$

- **MSE expansion: Sum/Difference**. Let $h = h_+ = h_-$. We define

$$\mathsf{MSE}[\tilde{\varsigma}_{+,\nu}(h) \pm \tilde{\varsigma}_{-,\nu}(h)] = \mathbb{E}[(\mathbf{f}'_{F,p}[\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p}(h)) \pm \mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p})])^2|\mathbf{X}]$$

Then, we have:

$$\begin{aligned}
&\mathsf{MSE}[\tilde{\varsigma}_{+,\nu}(h) \pm \tilde{\varsigma}_{-,\nu}(h)] \\
&= h^{2(1+p-\nu)} \left[\mathcal{B}_{F+,\nu,p}(h) \pm \mathcal{B}_{F-,\nu,p}(h)\right]^2 \{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}} \left[\mathcal{V}_{F-,\nu,p}(h) + \mathcal{V}_{F+,\nu,p}(h)\right] \\
&= h^{2(1+p-\nu)} \left[\mathcal{B}_{F+,\nu,p} \pm \mathcal{B}_{F-,\nu,p}\right]^2 \{1 + o_{\mathbb{P}}(1)\} + \frac{1}{nh^{1+2\nu}} \left[\mathcal{V}_{F-,\nu,p} + \mathcal{V}_{F+,\nu,p}\right] \{1 + o_{\mathbb{P}}(1)\}.
\end{aligned}$$

Under the additional assumption that $\mathcal{B}_{F+,\nu,p} \pm \mathcal{B}_{F-,\nu,p} \neq 0$, we obtain

$$\mathfrak{h}_{\Delta F,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\mathcal{V}_{F-,\nu,p} + \mathcal{V}_{F+,\nu,p})/n}{(\mathcal{B}_{F+,\nu,p} - \mathcal{B}_{F-,\nu,p})^2}\right]^{\frac{1}{3+2p}},$$

$$\mathfrak{h}_{\Sigma F,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\mathcal{V}_{F-,\nu,p} + \mathcal{V}_{F+,\nu,p})/n}{(\mathcal{B}_{F+,\nu,p} + \mathcal{B}_{F-,\nu,p})^2}\right]^{\frac{1}{3+2p}}.$$

Note that, when $h = h_+ = h_-$,

$$\mathsf{MSE}[\tilde{\varsigma}_\nu(\mathbf{h})] = \mathsf{MSE}[\tilde{\varsigma}_{+,\nu}(h) - \tilde{\varsigma}_{-,\nu}(h)].$$

10.9 Bias Correction

10.9.1 Standard Fuzzy RD Estimator

The bias-corrected covariate-adjusted fuzzy RD estimator is

$$\hat{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b}) = \hat{\varsigma}_\nu(\mathbf{h}) - \left[h_+^{1+p-\nu}\hat{\mathcal{B}}_{U+,p,q}(h_+,b_+) - h_-^{1+p-\nu}\hat{\mathcal{B}}_{U+,p,q}(h_-,b_-)\right],$$

$$\hat{\mathcal{B}}_{U-,\nu,p,q}(h,b) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{U-,q}(b)}{(1+p)!},$$

$$\hat{\mathcal{B}}_{U+,\nu,p,q}(h,b) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{U+,q}(b)}{(1+p)!}.$$

Therefore, we have

$$\begin{aligned}
\hat{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b}) \;=\; & \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'\left[\frac{1}{n^{1/2}h_+^{1/2+\nu}}[\mathbf{I}_{2+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)] - \frac{1}{n^{1/2}h_-^{1/2+\nu}}[\mathbf{I}_{2+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)]\right]\mathbf{F} \\
& + \epsilon_{\varsigma,\nu} + (\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}) - \mathbf{f}_{U,\nu})'\operatorname{vec}(\hat{\boldsymbol{\beta}}_{U,p}(\mathbf{h}) - \boldsymbol{\beta}_{U,p}).
\end{aligned}$$

10.9.2 Covariate-Adjusted Fuzzy RD Estimator

The bias-corrected covariate-adjusted fuzzy RD estimator is

$$\tilde{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b}) = \tilde{\varsigma}_\nu(\mathbf{h}) - \left[h_+^{1+p-\nu}\hat{\mathcal{B}}_{F+,p,q}(h_+,b_+) - h_-^{1+p-\nu}\hat{\mathcal{B}}_{F+,p,q}(h_-,b_-)\right],$$

$$\hat{\mathcal{B}}_{F-,\nu,p,q}(h,b) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{F-,q}(b)}{(1+p)!},$$

$$\hat{\mathcal{B}}_{F+,\nu,p,q}(h,b) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{F+,q}(b)}{(1+p)!}.$$

Therefore, we have

$$\begin{aligned}\tilde{\varsigma}^{\mathsf{bc}}_{\nu}(\mathbf{h},\mathbf{b}) \quad &= \quad \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'\left[\frac{1}{n^{1/2}h_+^{1/2+\nu}}[\mathbf{I}_{2+d} \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)] - \frac{1}{n^{1/2}h_-^{1/2+\nu}}[\mathbf{I}_{2+d} \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)]\right]\mathbf{F} \\ &\quad +\epsilon_{\tilde{\varsigma},\nu} + (\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}) - \mathbf{f}_{F,\nu})'\,\mathrm{vec}(\hat{\boldsymbol{\beta}}_{F,p}(\mathbf{h}) - \boldsymbol{\beta}_{F,p}).\end{aligned}$$

10.10 Distributional Approximations

10.10.1 Standard Fuzzy RD Estimator

The two standardized statistics are:

$$T_{U,\nu}(\mathbf{h}) = \frac{\varsigma_\nu(\mathbf{h}) - \varsigma_\nu}{\sqrt{\mathsf{Var}[\varsigma_\nu(\mathbf{h})]}} \qquad \text{and} \qquad T^{\mathsf{bc}}_{U,\nu}(\mathbf{h},\mathbf{b}) = \frac{\varsigma^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b}) - \varsigma_\nu}{\sqrt{\mathsf{Var}[\varsigma^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b})]}}$$

where

$$\mathsf{Var}[\varsigma_\nu(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{U-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{U+,\nu,p}(h_+),$$

$$\mathcal{V}_{U-,\nu,p}(h) = \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \mathbf{P}_{-,p}(h)]\mathbf{\Sigma}_{U-}[\mathbf{I}_2 \otimes \mathbf{P}_{-,p}(h)']\mathbf{f}_{F,\nu},$$

$$\mathcal{V}_{U+,\nu,p}(h) = \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \mathbf{P}_{+,p}(h)]\mathbf{\Sigma}_{U+}[\mathbf{I}_2 \otimes \mathbf{P}_{+,p}(h)']\mathbf{f}_{F,\nu}.$$

and

$$\mathsf{Var}[\varsigma^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{U-,\nu,p,q}(h_-,b_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{U+,\nu,p,q}(h_+,b_+),$$

$$\mathcal{V}^{\mathsf{bc}}_{U-,\nu,p,q}(h,b) = \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)]\mathbf{\Sigma}_{S-}[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)']\mathbf{f}_{U,\nu},$$

$$\mathcal{V}^{\mathsf{bc}}_{U+,\nu,p,q}(h,b) = \mathbf{f}'_{U,\nu}[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)]\mathbf{\Sigma}_{S+}[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)']\mathbf{f}_{U,\nu}.$$

As shown above, $\mathcal{V}_{U-,\nu,p}(h) \asymp_{\mathbb{P}} 1$, $\mathcal{V}_{U+,\nu,p}(h) \asymp_{\mathbb{P}} 1$, $\mathcal{V}^{\mathsf{bc}}_{U-,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$ and $\mathcal{V}^{\mathsf{bc}}_{U+,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$, provided $\overline{\lim}_{n\to\infty}\max\{\rho_-,\rho_+\} < \infty$ and the other assumptions and bandwidth conditions hold.

Lemma SA-18 *Let assumptions SA-1, SA-4 and SA-5 hold with $\varrho \geq 1+q$, and $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$.*
(1) If $nh_-^{2p+3} \to 0$ and $nh_+^{2p+3} \to 0$, then

$$T_{U,\nu}(\mathbf{h}) \to_d \mathcal{N}(0,1).$$

(2) If $nh_-^{2p+3}\max\{h_-^2, b_-^{2(q-p)}\} \to 0$, $nh_+^{2p+3}\max\{h_+^2, b_+^{2(q-p)}\} \to 0$ and $\overline{\lim}_{n\to\infty}\max\{\rho_-,\rho_+\} < \infty$,

then

$$T^{\mathsf{bc}}_{U,\nu}(\mathbf{h},\mathbf{b}) \to_d \mathcal{N}(0,1).$$

10.10.2 Covariate-Adjusted Fuzzy RD Estimator

The two standardized statistics are:

$$T_{F,\nu}(\mathbf{h}) = \frac{\tilde{\varsigma}_\nu(\mathbf{h}) - \varsigma_\nu}{\sqrt{\mathsf{Var}[\tilde{\varsigma}_\nu(\mathbf{h})]}} \qquad \text{and} \qquad T^{\mathsf{bc}}_{F,\nu}(\mathbf{h},\mathbf{b}) = \frac{\tilde{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b}) - \varsigma_\nu}{\sqrt{\mathsf{Var}[\tilde{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b})]}}$$

where

$$\mathsf{Var}[\tilde{\varsigma}_\nu(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}_{F-,\nu,p}(h_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}_{F+,\nu,p}(h_+),$$

$$\mathcal{V}_{F-,\nu,p}(h) = \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h)]\mathbf{\Sigma}_{F-}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h)']\mathbf{f}_{F,\nu},$$

$$\mathcal{V}_{F+,\nu,p}(h) = \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h)]\mathbf{\Sigma}_{F+}[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h)']\mathbf{f}_{F,\nu}.$$

and

$$\mathsf{Var}[\tilde{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h},\mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{F-,\nu,p,q}(h_-,b_-) + \frac{1}{nh_+^{1+2\nu}}\mathcal{V}^{\mathsf{bc}}_{F+,\nu,p,q}(h_+,b_+),$$

$$\mathcal{V}^{\mathsf{bc}}_{F-,\nu,p,q}(h,b) = \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)]\mathbf{\Sigma}_{F-}[\mathbf{I}_{2+d} \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h,b)']\mathbf{f}_{F,\nu},$$

$$\mathcal{V}^{\mathsf{bc}}_{F+,\nu,p,q}(h,b) = \mathbf{f}'_{F,\nu}[\mathbf{I}_{2+d} \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)]\mathbf{\Sigma}_{F+}[\mathbf{I}_{2+d} \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h,b)']\mathbf{f}_{F,\nu}.$$

As shown above, $\mathcal{V}_{F-,\nu,p}(h) \asymp_{\mathbb{P}} 1$, $\mathcal{V}_{F+,\nu,p}(h) \asymp_{\mathbb{P}} 1$, $\mathcal{V}^{\mathsf{bc}}_{F-,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$ and $\mathcal{V}^{\mathsf{bc}}_{F+,\nu,p,q}(h,b) \asymp_{\mathbb{P}} 1$, provided $\overline{\lim}_{n\to\infty} \max\{\rho_-,\rho_+\} < \infty$ and the other assumptions and bandwidth conditions hold.

Lemma SA-19 *Let assumptions SA-1, SA-4 and SA-5 hold with $\varrho \geq 1+q$, and $n\min\{h_-^{1+2\nu}, h_+^{1+2\nu}\} \to \infty$.*
(1) If $nh_-^{2p+3} \to 0$ and $nh_+^{2p+3} \to 0$, then

$$T_{F,\nu}(\mathbf{h}) \to_d \mathcal{N}(0,1).$$

(2) If $nh_-^{2p+3}\max\{h_-^2, b_-^{2(q-p)}\} \to 0$, $nh_+^{2p+3}\max\{h_+^2, b_+^{2(q-p)}\} \to 0$ and $\overline{\lim}_{n\to\infty}\max\{\rho_-,\rho_+\} < \infty$, then

$$T^{\mathsf{bc}}_{F,\nu}(\mathbf{h},\mathbf{b}) \to_d \mathcal{N}(0,1).$$

10.11 Variance Estimation

The only unknown matrices in the asymptotic variance formulas derived above are:

- Standard Estimator: $\mathbf{\Sigma}_{U-} = \mathbb{V}[\mathbf{U}(0)|\mathbf{X}]$ and $\mathbf{\Sigma}_{U+} = \mathbb{V}[\mathbf{U}(1)|\mathbf{X}]$.
- Covariate-Adjusted Estimator: $\mathbf{\Sigma}_{F-} = \mathbb{V}[\mathbf{F}(0)|\mathbf{X}]$ and $\mathbf{\Sigma}_{F+} = \mathbb{V}[\mathbf{F}(1)|\mathbf{X}]$.

We consider two alternative type of standard error estimators, based on either a Nearest Neighbor (NN) and plug-in residuals (PR) approach. For $i = 1, 2, \cdots, n$, define the "estimated" residuals as follows.

- **Nearest Neighbor (NN) approach**:

$$\hat{\varepsilon}_{V-,i}(J) = \mathbb{1}(X_i < \bar{x})\sqrt{\frac{J}{J+1}}\left(V_i - \frac{1}{J}\sum_{j=1}^{J} V_{\ell_{-,j}(i)}\right),$$

$$\hat{\varepsilon}_{V+,i}(J) = \mathbb{1}(X_i \geq \bar{x})\sqrt{\frac{J}{J+1}}\left(V_i - \frac{1}{J}\sum_{j=1}^{J} V_{\ell_{+,j}(i)}\right),$$

  where $V \in \{Y, T, Z_1, Z_2, \cdots, Z_d\}$, and $\ell_j^+(i)$ is the index of the j-th closest unit to unit i among $\{X_i : X_i \geq \bar{x}\}$ and $\ell_j^-(i)$ is the index of the j-th closest unit to unit i among $\{X_i : X_i < \bar{x}\}$, and J denotes a (fixed) the number of neighbors chosen.

- **Plug-in Residuals (PR) approach**:

$$\hat{\varepsilon}_{V-,p,i}(h) = \mathbb{1}(X_i < \bar{x})\sqrt{\omega_{-,p,i}}(V_i - \mathbf{r}_p(X_i - \bar{x})'\hat{\boldsymbol{\beta}}_{V-,p}(h)),$$

$$\hat{\varepsilon}_{V+,p,i}(h) = \mathbb{1}(X_i \geq \bar{x})\sqrt{\omega_{+,p,i}}(V_i - \mathbf{r}_p(X_i - \bar{x})'\hat{\boldsymbol{\beta}}_{V+,p}(h)),$$

  where again $V \in \{Y, T, Z_1, Z_2, \cdots, Z_d\}$ is a placeholder for the outcome variable used, and the additional weights $\{(\omega_{-,p,i}, \omega_{+,p,i}) : i = 1, 2, \cdots, n\}$ are described in the sharp RD setting above.

10.11.1 Standard Fuzzy RD Estimator

Define the estimators

$$\check{\mathbf{\Sigma}}_{U-}(J) = \begin{bmatrix} \check{\mathbf{\Sigma}}_{YY-}(J) & \check{\mathbf{\Sigma}}_{YT-}(J) \\ \check{\mathbf{\Sigma}}_{TY-}(J) & \check{\mathbf{\Sigma}}_{TT-}(J) \end{bmatrix}$$

and

$$\check{\mathbf{\Sigma}}_{U+}(J) = \begin{bmatrix} \check{\mathbf{\Sigma}}_{YY+}(J) & \check{\mathbf{\Sigma}}_{YT+}(J) \\ \check{\mathbf{\Sigma}}_{TY+}(J) & \check{\mathbf{\Sigma}}_{TT+}(J) \end{bmatrix}$$

where the matrices $\check{\mathbf{\Sigma}}_{VW-}(J)$ and $\check{\mathbf{\Sigma}}_{VW+}(J)$, $V, W \in \{Y, T\}$, are $(p+1) \times (p+1)$ matrices with generic (i, j)-th elements, respectively,

$$\left[\check{\mathbf{\Sigma}}_{VW-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(i = j)\hat{\varepsilon}_{V-,i}(J)\hat{\varepsilon}_{W-,i}(J),$$

$$\left[\check{\mathbf{\Sigma}}_{VW+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(i = j)\hat{\varepsilon}_{V+,i}(J)\hat{\varepsilon}_{W+,i}(J),$$

for all $1 \leq i, j \leq n$, and for all $V, W \in \{Y, T\}$.

Similarly, define the estimators

$$\hat{\mathbf{\Sigma}}_{U-,p}(h) = \begin{bmatrix} \hat{\mathbf{\Sigma}}_{YY-,p}(h) & \hat{\mathbf{\Sigma}}_{YT-,p}(h) \\ \hat{\mathbf{\Sigma}}_{TY-,p}(h) & \hat{\mathbf{\Sigma}}_{TT-,p}(h) \end{bmatrix}$$

and

$$\hat{\mathbf{\Sigma}}_{U+,p}(h) = \begin{bmatrix} \hat{\mathbf{\Sigma}}_{YY+,p}(h) & \hat{\mathbf{\Sigma}}_{YT+,p}(h) \\ \hat{\mathbf{\Sigma}}_{TY+,p}(h) & \hat{\mathbf{\Sigma}}_{TT+,p}(h) \end{bmatrix}$$

where the matrices $\hat{\mathbf{\Sigma}}_{VW-,p}(h)$ and $\hat{\mathbf{\Sigma}}_{VW+,p}(h)$, $V, W \in \{Y, T\}$, are $(p+1) \times (p+1)$ matrices with generic (i, j)-th elements, respectively,

$$\left[\hat{\mathbf{\Sigma}}_{VW-,p}(h)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(i = j)\hat{\varepsilon}_{V-,p,i}(h)\hat{\varepsilon}_{W-,p,j}(h),$$

$$\left[\hat{\mathbf{\Sigma}}_{VW+,p}(h)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(i = j)\hat{\varepsilon}_{V+,p,i}(h)\hat{\varepsilon}_{W+,p,j}(h),$$

for all $1 \leq i, j \leq n$, and for all $V, W \in \{Y, T\}$.

- **Undersmoothing NN Variance Estimator**:

$$\check{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}_\nu(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}_{U-,\nu,p}(\mathbf{h}) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}_{U+,\nu,p}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{U-,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{P}_{-,p}(h_-)]\check{\mathbf{\Sigma}}_{U-}(J)[\mathbf{I}_2 \otimes \mathbf{P}_{-,p}(h_-)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{U+,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{P}_{+,p}(h_+)]\check{\mathbf{\Sigma}}_{U+}(J)[\mathbf{I}_2 \otimes \mathbf{P}_{+,p}(h_+)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}).$$

- **Undersmoothing PR Variance Estimator**:

$$\hat{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}_\nu(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}_{U-,\nu,p}(\mathbf{h}) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}_{U+,\nu,p}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{U-,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h_-)]\hat{\mathbf{\Sigma}}_{U-,p}(h_-)[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h_-)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{U+,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h_-)]\hat{\mathbf{\Sigma}}_{U+,p}(h_+)[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h_-)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}).$$

- **Robust Bias-Correction NN Variance Estimator**:

$$\check{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}^{\mathsf{bc}}_\nu(\mathbf{h}, \mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}^{\mathsf{bc}}_{U-,\nu,p,q}(\mathbf{h}, \mathbf{b}) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}_{U+,\nu,p,q}(\mathbf{h}, \mathbf{b}),$$

$$\check{\mathcal{V}}^{\mathsf{bc}}_{U-,\nu,p,q}(\mathbf{h}, \mathbf{b}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-, b_-)]\check{\mathbf{\Sigma}}_{U-}(J)[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-, b_-)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{U+,\nu,p,q}(\mathbf{h}, \mathbf{b}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+, b_+)]\check{\mathbf{\Sigma}}_{U+}(J)[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+, b_+)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}).$$

- **Robust Bias-Correction PR Variance Estimator**:

$$\hat{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}^{\mathsf{bc}}_{\nu}(\mathbf{h},\mathbf{b})] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}^{\mathsf{bc}}_{U-,\nu,p,q}(\mathbf{h},\mathbf{b}) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}_{U+,\nu,p,q}(\mathbf{h},\mathbf{b}),$$

$$\hat{\mathcal{V}}^{\mathsf{bc}}_{U-,\nu,p,q}(\mathbf{h},\mathbf{b}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)]\hat{\mathbf{\Sigma}}_{U-,q}(h_-)[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{-,p,q}(h_-,b_-)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{U+,\nu,p,q}(\mathbf{h},\mathbf{b}) = \hat{\mathbf{f}}_{U,\nu}(\mathbf{h})'[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)]\hat{\mathbf{\Sigma}}_{U+,q}(h_+)[\mathbf{I}_2 \otimes \mathbf{P}^{\mathsf{bc}}_{+,p,q}(h_+,b_+)']\hat{\mathbf{f}}_{U,\nu}(\mathbf{h}).$$

Lemma SA-20 *Suppose the conditions of Lemma SA-10 hold. If, in addition,* $\max_{1\leq i\leq n}|\omega_{-,i}| = O_{\mathbb{P}}(1)$ *and* $\max_{1\leq i\leq n}|\omega_{+,i}| = O_{\mathbb{P}}(1)$*, and* $\sigma^2_{U+}(x)$ *and* $\sigma^2_{U-}(x)$ *are Lipschitz continuous, then*

$$\frac{\check{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}_{\nu}(\mathbf{h})]}{\mathsf{Var}[\hat{\varsigma}_{\nu}(\mathbf{h})]]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}_{\nu}(\mathbf{h})]}{\mathsf{Var}[\hat{\varsigma}_{\nu}(\mathbf{h})]} \to_{\mathbb{P}} 1, \quad \frac{\check{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}^{\mathsf{bc}}_{\nu}(\mathbf{h},\mathbf{b})]}{\mathsf{Var}[\hat{\varsigma}^{\mathsf{bc}}_{\nu}(\mathbf{h},\mathbf{b})]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathsf{V}}\mathsf{ar}[\hat{\varsigma}^{\mathsf{bc}}_{\nu}(\mathbf{h},\mathbf{b})]}{\mathsf{Var}[\hat{\varsigma}^{\mathsf{bc}}_{\nu}(\mathbf{h},\mathbf{b})]} \to_{\mathbb{P}} 1.$$

10.11.2 Covariate-Adjusted Fuzzy RD Estimator

Define the estimators

$$\check{\mathbf{\Sigma}}_{F-}(J) = \begin{bmatrix} \check{\mathbf{\Sigma}}_{YY-}(J) & \check{\mathbf{\Sigma}}_{YT-}(J) & \check{\mathbf{\Sigma}}_{YZ_1-}(J) & \check{\mathbf{\Sigma}}_{YZ_2-}(J) & \cdots & \check{\mathbf{\Sigma}}_{YZ_d-}(J) \\ \check{\mathbf{\Sigma}}_{TY-}(J) & \check{\mathbf{\Sigma}}_{TT-}(J) & \check{\mathbf{\Sigma}}_{TZ_1-}(J) & \check{\mathbf{\Sigma}}_{TZ_2-}(J) & \cdots & \check{\mathbf{\Sigma}}_{TZ_d-}(J) \\ \check{\mathbf{\Sigma}}_{Z_1Y-}(J) & \check{\mathbf{\Sigma}}_{Z_1T-}(J) & \check{\mathbf{\Sigma}}_{Z_1Z_1-}(J) & \check{\mathbf{\Sigma}}_{Z_1Z_2-}(J) & \cdots & \check{\mathbf{\Sigma}}_{Z_1Z_d-}(J) \\ \check{\mathbf{\Sigma}}_{Z_2Y-}(J) & \check{\mathbf{\Sigma}}_{Z_2T-}(J) & \check{\mathbf{\Sigma}}_{Z_2Z_1-}(J) & \check{\mathbf{\Sigma}}_{Z_2Z_2-}(J) & \cdots & \check{\mathbf{\Sigma}}_{Z_2Z_d-}(J) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \check{\mathbf{\Sigma}}_{Z_dY-}(J) & \check{\mathbf{\Sigma}}_{Z_dT-}(J) & \check{\mathbf{\Sigma}}_{Z_dZ_1-}(J) & \check{\mathbf{\Sigma}}_{Z_dZ_2-}(J) & \cdots & \check{\mathbf{\Sigma}}_{Z_dZ_d-}(J) \end{bmatrix}$$

and

$$\check{\mathbf{\Sigma}}_{F+}(J) = \begin{bmatrix} \check{\mathbf{\Sigma}}_{YY+}(J) & \check{\mathbf{\Sigma}}_{YT+}(J) & \check{\mathbf{\Sigma}}_{YZ_1+}(J) & \check{\mathbf{\Sigma}}_{YZ_2+}(J) & \cdots & \check{\mathbf{\Sigma}}_{YZ_d+}(J) \\ \check{\mathbf{\Sigma}}_{TY+}(J) & \check{\mathbf{\Sigma}}_{TT+}(J) & \check{\mathbf{\Sigma}}_{TZ_1+}(J) & \check{\mathbf{\Sigma}}_{TZ_2+}(J) & \cdots & \check{\mathbf{\Sigma}}_{TZ_d+}(J) \\ \check{\mathbf{\Sigma}}_{Z_1Y+}(J) & \check{\mathbf{\Sigma}}_{Z_1T+}(J) & \check{\mathbf{\Sigma}}_{Z_1Z_1+}(J) & \check{\mathbf{\Sigma}}_{Z_1Z_2+}(J) & \cdots & \check{\mathbf{\Sigma}}_{Z_1Z_d+}(J) \\ \check{\mathbf{\Sigma}}_{Z_2Y+}(J) & \check{\mathbf{\Sigma}}_{Z_2T+}(J) & \check{\mathbf{\Sigma}}_{Z_2Z_1+}(J) & \check{\mathbf{\Sigma}}_{Z_2Z_2+}(J) & \cdots & \check{\mathbf{\Sigma}}_{Z_2Z_d+}(J) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \check{\mathbf{\Sigma}}_{Z_dY+}(J) & \check{\mathbf{\Sigma}}_{Z_dT+}(J) & \check{\mathbf{\Sigma}}_{Z_dZ_1+}(J) & \check{\mathbf{\Sigma}}_{Z_dZ_2+}(J) & \cdots & \check{\mathbf{\Sigma}}_{Z_dZ_d+}(J) \end{bmatrix}$$

where the matrices $\check{\mathbf{\Sigma}}_{VW-}(J)$ and $\check{\mathbf{\Sigma}}_{VW+}(J)$, $V,W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$, are $n \times n$ matrices with generic (i,j)-th elements, respectively,

$$\left[\check{\mathbf{\Sigma}}_{VW-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V-,i}(J)\hat{\varepsilon}_{W-,i}(J),$$

$$\left[\check{\mathbf{\Sigma}}_{VW+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V+,i}(J)\hat{\varepsilon}_{W+,i}(J),$$

for all $1 \leq i,j \leq n$, and for all $V,W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$.

Similarly, define the estimators

$$\hat{\boldsymbol{\Sigma}}_{F-,p}(h) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{YY-,p}(h) & \hat{\boldsymbol{\Sigma}}_{YT-,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{YZ_d-,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{TY-,p}(h) & \hat{\boldsymbol{\Sigma}}_{TT-,p}(h) & \hat{\boldsymbol{\Sigma}}_{TZ_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{TZ_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{TZ_d-,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{Z_1Y-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1T-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_1Z_d-,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{Z_2Y-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2T-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_2Z_d-,p}(h) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{\Sigma}}_{Z_dY-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dT-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_1-,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_2-,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_dZ_d-,p}(h) \end{bmatrix}$$

and

$$\hat{\boldsymbol{\Sigma}}_{F+,p}(h) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{YY+,p}(h) & \hat{\boldsymbol{\Sigma}}_{YT+,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{YZ_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{YZ_d+,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{TY+,p}(h) & \hat{\boldsymbol{\Sigma}}_{TT+,p}(h) & \hat{\boldsymbol{\Sigma}}_{TZ_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{TZ_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{TZ_d+,p}(h) \\ \hat{\boldsymbol{\Sigma}}_{Z_1Y+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1T+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_1Z_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_1Z_d+,p}(h) \\ \check{\boldsymbol{\Sigma}}_{Z_2Y+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2T+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_2Z_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_2Z_d+,p}(h) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{\Sigma}}_{Z_dY+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dT+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_1+,p}(h) & \hat{\boldsymbol{\Sigma}}_{Z_dZ_2+,p}(h) & \cdots & \hat{\boldsymbol{\Sigma}}_{Z_dZ_d+,p}(h) \end{bmatrix}$$

where the matrices $\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)$ and $\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)$, $V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$, are $n \times n$ matrices with generic (i,j)-th elements, respectively,

$$\left[\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V-,p,i}(h)\hat{\varepsilon}_{W-,p,j}(h),$$

$$\left[\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(i=j)\hat{\varepsilon}_{V+,p,i}(h)\hat{\varepsilon}_{W+,p,j}(h),$$

for all $1 \leq i, j \leq n$, and for all $V, W \in \{Y, Z_1, Z_2, \cdots, Z_d\}$.

- **Undersmoothing NN Variance Estimator**:

$$\check{\mathsf{V}}\mathsf{ar}[\tilde{\varsigma}_\nu(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\check{\mathcal{V}}_{F-,\nu,p}(\mathbf{h}) + \frac{1}{nh_+^{1+2\nu}}\check{\mathcal{V}}_{F+,\nu,p}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{F-,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h_-)]\check{\boldsymbol{\Sigma}}_{F-}(J)[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h_-)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{F+,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h_+)]\check{\boldsymbol{\Sigma}}_{F+}(J)[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h_+)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}).$$

- **Undersmoothing PR Variance Estimator**:

$$\hat{\mathsf{V}}\mathsf{ar}[\tilde{\varsigma}_\nu(\mathbf{h})] = \frac{1}{nh_-^{1+2\nu}}\hat{\mathcal{V}}_{F-,\nu,p}(\mathbf{h}) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathcal{V}}_{F+,\nu,p}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{F-,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h_-)]\hat{\boldsymbol{\Sigma}}_{F-,p}(h_-)[\mathbf{I}_{2+d} \otimes \mathbf{P}_{-,p}(h_-)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{F+,\nu,p}(\mathbf{h}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h_-)]\hat{\boldsymbol{\Sigma}}_{F+,p}(h_+)[\mathbf{I}_{2+d} \otimes \mathbf{P}_{+,p}(h_-)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}).$$

- **Robust Bias-Correction NN Variance Estimator**:

$$\check{\mathsf{Var}}[\tilde{\varsigma}_{\nu}^{\mathsf{bc}}(\mathbf{h},\mathbf{b})] = \frac{1}{nh_{-}^{1+2\nu}}\check{\mathcal{V}}_{F-,\nu,p,q}^{\mathsf{bc}}(\mathbf{h},\mathbf{b}) + \frac{1}{nh_{+}^{1+2\nu}}\check{\mathcal{V}}_{F+,\nu,p,q}(\mathbf{h},\mathbf{b}),$$

$$\check{\mathcal{V}}_{F-,\nu,p,q}^{\mathsf{bc}}(\mathbf{h},\mathbf{b}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h}))'[\mathbf{I}_{2+d}\otimes\mathbf{P}_{-,p,q}^{\mathsf{bc}}(h_-,b_-)]\check{\mathbf{\Sigma}}_{F-}(J)[\mathbf{I}_{2+d}\otimes\mathbf{P}_{-,p,q}^{\mathsf{bc}}(h_-,b_-)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}),$$

$$\check{\mathcal{V}}_{F+,\nu,p,q}(\mathbf{h},\mathbf{b}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d}\otimes\mathbf{P}_{+,p,q}^{\mathsf{bc}}(h_+,b_+)]\check{\mathbf{\Sigma}}_{F+}(J)[\mathbf{I}_{2+d}\otimes\mathbf{P}_{+,p,q}^{\mathsf{bc}}(h_+,b_+)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}).$$

- **Robust Bias-Correction PR Variance Estimator**:

$$\hat{\mathsf{Var}}[\tilde{\varsigma}_{\nu}^{\mathsf{bc}}(\mathbf{h},\mathbf{b})] = \frac{1}{nh_{-}^{1+2\nu}}\hat{\mathcal{V}}_{F-,\nu,p,q}^{\mathsf{bc}}(\mathbf{h},\mathbf{b}) + \frac{1}{nh_{+}^{1+2\nu}}\hat{\mathcal{V}}_{F+,\nu,p,q}(\mathbf{h},\mathbf{b}),$$

$$\hat{\mathcal{V}}_{F-,\nu,p,q}^{\mathsf{bc}}(\mathbf{h},\mathbf{b}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d}\otimes\mathbf{P}_{-,p,q}^{\mathsf{bc}}(h_-,b_-)]\hat{\mathbf{\Sigma}}_{F-,q}(h_-)[\mathbf{I}_{2+d}\otimes\mathbf{P}_{-,p,q}^{\mathsf{bc}}(h_-,b_-)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}),$$

$$\hat{\mathcal{V}}_{F+,\nu,p,q}(\mathbf{h},\mathbf{b}) = \hat{\mathbf{f}}_{F,\nu}(\mathbf{h})'[\mathbf{I}_{2+d}\otimes\mathbf{P}_{+,p,q}^{\mathsf{bc}}(h_+,b_+)]\hat{\mathbf{\Sigma}}_{F+,q}(h_+)[\mathbf{I}_{2+d}\otimes\mathbf{P}_{+,p,q}^{\mathsf{bc}}(h_+,b_+)']\hat{\mathbf{f}}_{F,\nu}(\mathbf{h}).$$

Lemma SA-21 *Suppose the conditions of Lemma SA-11 hold. If, in addition,* $\max_{1\leq i\leq n}|\omega_{-,i}| = O_{\mathbb{P}}(1)$ *and* $\max_{1\leq i\leq n}|\omega_{+,i}| = O_{\mathbb{P}}(1)$, *and* $\sigma^2_{F+}(x)$ *and* $\sigma^2_{F-}(x)$ *are Lipschitz continuous, then*

$$\frac{\check{\mathsf{Var}}[\tilde{\varsigma}_{\nu}(\mathbf{h})]}{\mathsf{Var}[\tilde{\varsigma}_{\nu}(\mathbf{h})]]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathsf{Var}}[\tilde{\varsigma}_{\nu}(\mathbf{h})]}{\mathsf{Var}[\tilde{\varsigma}_{\nu}(\mathbf{h})]} \to_{\mathbb{P}} 1, \quad \frac{\check{\mathsf{Var}}[\tilde{\varsigma}_{\nu}^{\mathsf{bc}}(\mathbf{h},\mathbf{b})]}{\mathsf{Var}[\tilde{\varsigma}_{\nu}^{\mathsf{bc}}(\mathbf{h},\mathbf{b})]} \to_{\mathbb{P}} 1, \quad \frac{\hat{\mathsf{Var}}[\tilde{\varsigma}_{\nu}^{\mathsf{bc}}(\mathbf{h},\mathbf{b})]}{\mathsf{Var}[\tilde{\varsigma}_{\nu}^{\mathsf{bc}}(\mathbf{h},\mathbf{b})]} \to_{\mathbb{P}} 1.$$

10.12 Extension to Clustered Data

As discussed for sharp RD designs, it is straightforward to extend the results above to the case of clustered data. Recall that in this case asymptotics are conducted assuming that the number of clusters, G, grows ($G\to\infty$) satisfying the usual asymptotic restriction $Gh\to\infty$.

For brevity, we only describe the asymptotic variance estimators with clustering, which are now implemented in the upgraded versions of the `Stata` and `R` software described in Calonico, Cattaneo, and Titiunik (2014a, 2015). Specifically, we assume that each unit i belongs to one (and only one) cluster g, and let $\mathcal{G}(i)=g$ for all units $i=1,2,\cdots,n$ and all clusters $g=1,2,\cdots,G$. Define

$$\omega_{-,p} = \frac{G}{G-1}\frac{N_- - 1}{N_- - p - 1}, \qquad \omega_{+,p} = \frac{G}{G-1}\frac{N_+ - 1}{N_+ - p - 1}.$$

The clustered-consistent variance estimators are as follows. We recycle notation for convenience, and to emphasize the nesting of the heteroskedasticity-robust estimators into the cluster-robust ones.

10.12.1 Standard Fuzzy RD Estimator

Redefine the matrices $\check{\boldsymbol{\Sigma}}_{VW-}(J)$ and $\check{\boldsymbol{\Sigma}}_{VW+}(J)$, respectively, to now have generic (i,j)-th elements

$$\left[\check{\boldsymbol{\Sigma}}_{VW-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V-,i}(J)\hat{\varepsilon}_{W-,i}(J),$$

$$\left[\check{\boldsymbol{\Sigma}}_{VW+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V+,i}(J)\hat{\varepsilon}_{W+,i}(J),$$

$$1 \leq i,j \leq n, \qquad V,W \in \{Y,T\}.$$

Similarly, redefine the matrices $\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)$ and $\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)$, respectively, to now have generic (i,j)-th elements

$$\left[\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V-,p,i}(h)\hat{\varepsilon}_{W-,p,j}(h),$$

$$\left[\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V+,p,i}(h)\hat{\varepsilon}_{W+,p,j}(h),$$

$$1 \leq i,j \leq n, \qquad V,W \in \{Y,T\}.$$

With these redefinitions, the clustered-robust variance estimators are as above. In particular, if each cluster has one observation, then the estimators reduce to the heteroskedastic-robust estimators with $\omega_{-,p,i} = \omega_{+,p,i} = 1$ for all $i = 1, 2, \cdots, n$.

10.12.2 Covariate-Adjusted Fuzzy RD Estimator

Redefine the matrices $\check{\boldsymbol{\Sigma}}_{VW-}(J)$ and $\check{\boldsymbol{\Sigma}}_{VW+}(J)$, respectively, to now have generic (i,j)-th elements

$$\left[\check{\boldsymbol{\Sigma}}_{VW-}(J)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V-,i}(J)\hat{\varepsilon}_{W-,i}(J),$$

$$\left[\check{\boldsymbol{\Sigma}}_{VW+}(J)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V+,i}(J)\hat{\varepsilon}_{W+,i}(J),$$

$$1 \leq i,j \leq n, \qquad V,W \in \{Y,T,Z_1,Z_2,\cdots,Z_d\}.$$

Similarly, redefine the matrices $\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)$ and $\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)$, respectively, to now have generic (i,j)-th elements

$$\left[\hat{\boldsymbol{\Sigma}}_{VW-,p}(h)\right]_{ij} = \mathbb{1}(X_i < \bar{x})\mathbb{1}(X_j < \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V-,p,i}(h)\hat{\varepsilon}_{W-,p,j}(h),$$

$$\left[\hat{\boldsymbol{\Sigma}}_{VW+,p}(h)\right]_{ij} = \mathbb{1}(X_i \geq \bar{x})\mathbb{1}(X_j \geq \bar{x})\mathbb{1}(\mathcal{G}(i) = \mathcal{G}(j))\hat{\varepsilon}_{V+,p,i}(h)\hat{\varepsilon}_{W+,p,j}(h),$$

$$1 \leq i,j \leq n, \qquad V,W \in \{Y,T,Z_1,Z_2,\cdots,Z_d\}.$$

With these redefinitions, the clustered-robust variance estimators are as above. In particular, if each cluster has one observation, then the estimators reduce to the heteroskedastic-robust estimators with $\omega_{-,p,i} = \omega_{+,p,i} = 1$ for all $i = 1, 2, \cdots, n$.

11 Estimation using Treatment Interaction

As in the case of sharp RD designs, it is easy to show that the interacted covariate-adjusted fuzzy RD treatment effect estimator is not consistent in general for the standard fuzzy RD estimand. Recall that we showed that $\hat{\varsigma}_\nu(\mathbf{h}) \to_{\mathbb{P}} \varsigma_\nu$ and $\tilde{\varsigma}_\nu(\mathbf{h}) \to_{\mathbb{P}} \varsigma_\nu$, under the conditions of Lemma SA-15 and if $\boldsymbol{\mu}^{(\nu)}_{Z+} = \boldsymbol{\mu}^{(\nu)}_{Z-}$.

In this section we show, under the same minimal continuity conditions, that

$$\check{\zeta}_\nu(\mathbf{h}) := \frac{\check{\eta}_{Y,\nu}(\mathbf{h})}{\check{\eta}_{T,\nu}(\mathbf{h})} \to_{\mathbb{P}} \zeta_\nu \neq \varsigma_\nu$$

in general, and give a precise characterization of the probability limit.

Lemma SA-22 *Let the conditions of Lemma SA-15 hold. Then,*

$$\check{\zeta}_\nu(\mathbf{h}) \to_{\mathbb{P}} \zeta_\nu := \frac{\tau_{Y,\nu} - \left[\boldsymbol{\mu}^{(\nu)\prime}_{Z+}\boldsymbol{\gamma}_{Y+} - \boldsymbol{\mu}^{(\nu)\prime}_{Z-}\boldsymbol{\gamma}_{Y-}\right]}{\tau_{T,\nu} - \left[\boldsymbol{\mu}^{(\nu)\prime}_{Z+}\boldsymbol{\gamma}_{T+} - \boldsymbol{\mu}^{(\nu)\prime}_{Z-}\boldsymbol{\gamma}_{T-}\right]}$$

with, for $V \in \{Y, T\}$,

$$\boldsymbol{\gamma}_{V-} = \boldsymbol{\sigma}^{-1}_{Z-}\mathbb{E}\left[(\mathbf{Z}_i(0) - \boldsymbol{\mu}_{Z-}(X_i))V_i(0)\big|\, X_i = \bar{x}\right],$$

$$\boldsymbol{\gamma}_{V+} = \boldsymbol{\sigma}^{-1}_{Z+}\mathbb{E}\left[(\mathbf{Z}_i(1) - \boldsymbol{\mu}_{Z+}(X_i))V_i(1)\big|\, X_i = \bar{x}\right],$$

where recall that $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z-}(\bar{x})$, $\boldsymbol{\mu}_{Z+} = \boldsymbol{\mu}_{Z+}(\bar{x})$, $\boldsymbol{\sigma}^2_{Z-} = \boldsymbol{\sigma}^2_{Z-}(\bar{x})$, and $\boldsymbol{\sigma}^2_{Z+} = \boldsymbol{\sigma}^2_{Z+}(\bar{x})$.

Part IV

Implementation Details

We give details on our proposed bandwidth selection methods. We also discuss some of their basic asymptotic properties. Recall that $\nu \leq p < q$, and let $\mathbf{h} = (h_-, h_+)$, $\mathbf{b} = (b_-, b_+)$, $\mathbf{v} = (v_-, v_+)$, $\mathbf{d} = (d_-, d_+)$, denote possibly different vanishing bandwidth sequences. The implementation details described in this section are exactly the implementations in the companion general purpose `Stata` and `R` packages described in Calonico, Cattaneo, Farrell, and Titiunik (2017).

12 Sharp RD Designs

All the bandwidth choices in sharp RD settings rely on estimating the following pre-asymptotic constants.

- **Bias Constants**:

$$\mathcal{B}_{-,\nu,p}(\mathbf{h}) = \mathbf{O}_{-,\nu,p}(h_-)\frac{[1, -\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})']\boldsymbol{\mu}_{S-}^{(1+p)}}{(1+p)!}, \qquad \mathbf{O}_{-,\nu,p}(h) = [\mathbf{I}_{1+d} \otimes \nu!\mathbf{e}_\nu' \boldsymbol{\Gamma}_{-,p}^{-1}(h)\boldsymbol{\vartheta}_{-,p}(h)],$$

$$\mathcal{B}_{+,\nu,p}(\mathbf{h}) = \mathbf{O}_{+,\nu,p}(h_+)\frac{[1, -\tilde{\boldsymbol{\gamma}}_{Y,p}(\mathbf{h})']\boldsymbol{\mu}_{S+}^{(1+p)}}{(1+p)!}, \qquad \mathbf{O}_{+,\nu,p}(h) = [\mathbf{I}_{1+d} \otimes \nu!\mathbf{e}_\nu' \boldsymbol{\Gamma}_{+,p}^{-1}(h)\boldsymbol{\vartheta}_{+,p}(h)].$$

- **Variance Constants**:

$$\mathcal{V}_{S-,\nu,p}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d} \otimes \mathbf{P}_{-,p}(h_-)]\boldsymbol{\Sigma}_{S-}[\mathbf{I}_{1+d} \otimes \mathbf{P}_{-,p}(h_-)']\mathbf{s}_{S,\nu}(\mathbf{h}),$$

$$\mathcal{V}_{S+,\nu,p}(\mathbf{h}) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d} \otimes \mathbf{P}_{+,p}(h_+)]\boldsymbol{\Sigma}_{S+}[\mathbf{I}_{1+d} \otimes \mathbf{P}_{+,p}(h_+)']\mathbf{s}_{S,\nu}(\mathbf{h}).$$

  where $\boldsymbol{\Sigma}_{S-}$ and $\boldsymbol{\Sigma}_{S+}$ depend on whether heteroskedasticity or clustering is assumed, and recall that

$$\mathbf{P}_{-,\nu,p}(h) = \sqrt{h}\boldsymbol{\Gamma}_{-,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_-(h)/\sqrt{n},$$
$$\mathbf{P}_{+,\nu,p}(h) = \sqrt{h}\boldsymbol{\Gamma}_{+,p}^{-1}(h)\mathbf{R}_p(h)'\mathbf{K}_+(h)/\sqrt{n}.$$

We approximate all these constants by employing consistent (and sometimes optimal) preliminary bandwidth choices. Specifically, we consider two preliminary bandwidth choices to select the main bandwidth(s) $\mathbf{h}$: (i) $\mathbf{b} \to \mathbf{0}$ is used to estimate the unknown "misspecification DGP biases" ($\boldsymbol{\mu}_{S-}^{(1+p)}$ and $\boldsymbol{\mu}_{S+}^{(1+p)}$), and (ii) $\mathbf{v} \to \mathbf{0}$ is used to estimate the unknown "design matrices objects" ($\mathbf{O}_{-,\nu,p}(\cdot), \mathbf{O}_{+,\nu,p}(\cdot), \mathbf{P}_{-,\nu,p}(\cdot), \mathbf{P}_{+,\nu,p}(\cdot)$) and the variance terms. In addition, we construct MSE-optimal choices for bandwidth $\mathbf{b}$ using the preliminary bandwidth $\mathbf{v} \to \mathbf{0}$, and an approximation to the underlying bias of the "misspecification DGP biases" $\boldsymbol{\mu}_{S-}^{(1+p)}$ and $\boldsymbol{\mu}_{S+}^{(1+p)}$. Once the main bandwidths $\mathbf{h}$ and $\mathbf{b}$ are chosen, we employ them to conduct MSE-optimal point estimation and valid bias-corrected inference.

12.1 Step 1: Choosing Bandwidth v

We require $v \to 0$ and $nv \to \infty$ (or $Gv \to \infty$ in the clustered data case). For practice, we propose a rule-of-thumb based on density estimation:

$$\hat{v} = C_K \cdot C_{\texttt{sd}} \cdot n^{-1/5}, \qquad C_K = \left(\frac{8\sqrt{\pi} \int K(u)^2 du}{3 \left(\int u^2 K(u) du \right)^2} \right)^{1/5}, \qquad C_{\texttt{sd}} = \min \left\{ s,\ \frac{IQR}{1.349} \right\},$$

where s^2 denotes the sample variance and IQR denotes the interquartile range of $\{X_i : 1 \leq i \leq n\}$. This bandwidth choice is simple modification of Silverman's rule of thumb. In particular, $C_K = 1.059$ when $K(\cdot)$ is the Gaussian kernel, $C_K = 1.843$ when $K(\cdot)$ is the uniform kernel, and $C_K = 2.576$ when $K(\cdot)$ is the triangular kernel.

12.2 Step 2: Choosing Bandwidth b

Since the target of interest when choosing bandwidth b are linear combinations of either (i) $\boldsymbol{\mu}_{S+}^{(1+p)} - \boldsymbol{\mu}_{S-}^{(1+p)}$, (ii) $\boldsymbol{\mu}_{S-}^{(1+p)}$ and $\boldsymbol{\mu}_{S+}^{(1+p)}$, or (less likely) $\boldsymbol{\mu}_{S+}^{(1+p)} + \boldsymbol{\mu}_{S-}^{(1+p)}$, we can employ the optimal choices already developed in the paper for these quantities. This approach leads to the MSE-optimal infeasible selectors ($p < q$):

Under the regularity conditions imposed above, and if $\mathcal{B}_{S-,1+p,q} \neq 0$ and $\mathcal{B}_{S+,1+p,q} \neq 0$, we obtain

$$\mathfrak{b}_{S+,1+p,q} = \left[\frac{3+2p}{2(q-p)} \frac{\mathcal{V}_{S-,1+p,q}/n}{\mathcal{B}_{S-,1+p,q}^2} \right]^{\frac{1}{3+2q}},$$

$$\mathfrak{b}_{S+,1+p,q} = \left[\frac{3+2p}{2(q-p)} \frac{\mathcal{V}_{S+,1+p,q}/n}{\mathcal{B}_{S+,1+p,q}^2} \right]^{\frac{1}{3+2q}},$$

and if $\mathcal{B}_{S+,1+p,q} \pm \mathcal{B}_{S-,1+p,q} \neq 0$, we obtain

$$\mathfrak{b}_{\Delta S,1+p,q} = \left[\frac{3+2p}{2(q-p)} \frac{(\mathcal{V}_{S-,1+p,q} + \mathcal{V}_{S+,1+p,q})/n}{(\mathcal{B}_{S+,\nu,p} - \mathcal{B}_{S-,1+p,q})^2} \right]^{\frac{1}{3+2q}},$$

$$\mathfrak{b}_{\Sigma S,1+p,q} = \left[\frac{3+2p}{2(q-p)} \frac{(\mathcal{V}_{S-,1+p,q} + \mathcal{V}_{S+,1+p,q})/n}{(\mathcal{B}_{S+,1+p,q} + \mathcal{B}_{S-,1+p,q})^2} \right]^{\frac{1}{3+2q}}.$$

Therefore, the associated data-driven counterparts are:

$$\hat{\mathfrak{b}}_{S+,1+p,q} = \left[\frac{3+2p}{2(q-p)} \frac{\hat{\mathcal{V}}_{S-,1+p,q}/n}{\hat{\mathcal{B}}_{S-,1+p,q}^2} \right]^{\frac{1}{3+2q}},$$

$$\hat{\mathfrak{b}}_{+,1+p,q} = \left[\frac{3+2p}{2(q-p)}\frac{\hat{\mathcal{V}}_{S+,1+p,q}/n}{\hat{\mathcal{B}}^2_{S+,1+p,q}}\right]^{\frac{1}{3+2q}},$$

$$\hat{\mathfrak{b}}_{\Delta S,1+p,q} = \left[\frac{3+2p}{2(q-p)}\frac{(\hat{\mathcal{V}}_{S-,1+p,q}+\hat{\mathcal{V}}_{S+,1+p,q})/n}{(\hat{\mathcal{B}}_{S+,\nu,p}-\hat{\mathcal{B}}_{S-,1+p,q})^2}\right]^{\frac{1}{3+2q}},$$

$$\hat{\mathfrak{b}}_{\Sigma S,1+p,q} = \left[\frac{3+2p}{2(q-p)}\frac{(\hat{\mathcal{V}}_{S-,1+p,q}+\hat{\mathcal{V}}_{S+,1+p,q})/n}{(\hat{\mathcal{B}}_{S+,1+p,q}+\hat{\mathcal{B}}_{S-,1+p,q})^2}\right]^{\frac{1}{3+2q}}.$$

where the preliminary constant estimates are chosen as follows.

- **Variance Constants**:

$$\hat{\mathcal{V}}_{S-,1+p,q} = \mathbf{s}_{S,1+p}(\hat{\mathbf{c}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,q}(\hat{c})]\hat{\mathbf{\Sigma}}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,q}(\hat{c})']\mathbf{s}_{S,1+p}(\hat{\mathbf{c}}),$$

$$\hat{\mathcal{V}}_{S+,1+p,q} = \mathbf{s}_{S,1+p}(\hat{\mathbf{c}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,q}(\hat{c})]\hat{\mathbf{\Sigma}}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,q}(\hat{c})']\mathbf{s}_{S,1+p}(\hat{\mathbf{c}}),$$

$$\hat{\mathbf{c}} = (\hat{c},\hat{c}).$$

  with $\hat{\mathbf{\Sigma}}_{S-}$ and $\hat{\mathbf{\Sigma}}_{S+}$ denoting the estimators described above under heteroskedasticity or under clustering, using either a nearest neighbor approach $(\check{\mathbf{\Sigma}}_{S-}(J),\check{\mathbf{\Sigma}}_{S+}(J))$ or a plug-in estimated residuals approach $(\hat{\mathbf{\Sigma}}_{S-,q}(\hat{c}),\hat{\mathbf{\Sigma}}_{S+,q}(\hat{c}))$.

- **Bias Constants**:

$$\hat{\mathcal{B}}_{S-,1+p,q} = \mathbf{O}_{-,1+p,q}(\hat{c})\frac{[1,-\tilde{\boldsymbol{\gamma}}_{Y,q}(\hat{\mathbf{c}})']\boldsymbol{\mu}^{(1+p)}_{S-,q}(\hat{d}_-)}{(1+p)!},$$

$$\hat{\mathcal{B}}_{S+,1+p,q} = \mathbf{O}_{-,1+p,q}(\hat{c})\frac{[1,-\tilde{\boldsymbol{\gamma}}_{Y,q}(\hat{\mathbf{c}})']\boldsymbol{\mu}^{(1+p)}_{S+,q}(\hat{d}_+)}{(1+p)!},$$

  where $\hat{\mathbf{d}} = (\hat{d}_-,\hat{d}_+)\to\mathbf{0}$ denotes a preliminary (possibly different) bandwidth sequence chosen to approximate the underlying bias of the bias estimator.

To construct the preliminary bandwidth $\hat{\mathbf{d}} = (\hat{d}_-,\hat{d}_+)$, can use (recursively) MSE-optimal choices targeted to the corresponding "misspecification DGP biases": (i) $\boldsymbol{\mu}^{(1+q)}_{S+}-\boldsymbol{\mu}^{(1+q)}_{S-}$, (ii) $\boldsymbol{\mu}^{(1+q)}_{S-}$ and $\boldsymbol{\mu}^{(1+q)}_{S+}$, or (less likely) $\boldsymbol{\mu}^{(1+q)}_{S+}+\boldsymbol{\mu}^{(1+q)}_{S-}$. This idea gives the MSE-optimal choices are:

$$\mathfrak{d}_{S-,1+q,1+q} = \left[\frac{3+2q}{2}\frac{\mathcal{V}_{S-,1+q,1+q}/n}{\mathcal{B}^2_{S-,1+q,1+q}}\right]^{\frac{1}{5+2q}},$$

$$\mathfrak{d}_{S+,1+q,1+q} = \left[\frac{3+2q}{2}\frac{\mathcal{V}_{S+,1+q,1+q}/n}{\mathcal{B}^2_{S+,1+q,1+q}}\right]^{\frac{1}{5+2q}},$$

$$\mathfrak{d}_{\Delta S,1+q,1+q} = \left[\frac{3+2q}{2}\frac{(\mathcal{V}_{S-,1+q,1+q}+\mathcal{V}_{S+,1+q,1+q})/n}{(\mathcal{B}_{S+,1+q,1+q}-\mathcal{B}_{S-,1+q,1+q})^2}\right]^{\frac{1}{5+2q}},$$
$$\mathfrak{d}_{\Sigma S,1+q,1+q} = \left[\frac{3+2q}{2}\frac{(\mathcal{V}_{S-,1+q,1+q}+\mathcal{V}_{S+,1+q,1+q})/n}{(\mathcal{B}_{S+,1+q,1+q}+\mathcal{B}_{S-,1+q,1+q})^2}\right]^{\frac{1}{5+2q}}.$$

In turn, these choices are implemented in an ad-hoc manner as follows:

$$\hat{\mathfrak{d}}_{S-,1+q,1+q} = \left[\frac{3+2q}{2}\frac{\hat{\mathcal{V}}_{S-,1+q,1+q}/n}{\mathring{\mathcal{B}}^2_{S-,1+q,1+q}}\right]^{\frac{1}{5+2q}},$$

$$\hat{\mathfrak{d}}_{S+,1+q,1+q} = \left[\frac{3+2q}{2}\frac{\hat{\mathcal{V}}_{S+,1+q,1+q}/n}{\mathring{\mathcal{B}}^2_{S+,1+q,1+q}}\right]^{\frac{1}{5+2q}},$$

$$\hat{\mathfrak{d}}_{\Delta S,1+q,1+q} = \left[\frac{3+2q}{2}\frac{(\hat{\mathcal{V}}_{S-,1+q,1+q}+\hat{\mathcal{V}}_{S+,1+q,1+q})/n}{(\mathring{\mathcal{B}}_{S+,1+q,1+q}-\mathring{\mathcal{B}}_{S-,1+q,1+q})^2}\right]^{\frac{1}{5+2q}},$$
$$\hat{\mathfrak{d}}_{\Sigma S,1+q,1+q} = \left[\frac{3+2q}{2}\frac{(\hat{\mathcal{V}}_{S-,1+q,1+q}+\hat{\mathcal{V}}_{S+,1+q,1+q})/n}{(\mathring{\mathcal{B}}_{S+,1+q,1+q}+\mathring{\mathcal{B}}_{S-,1+q,1+q})^2}\right]^{\frac{1}{5+2q}},$$

where

$$\hat{\mathcal{V}}_{S-,1+q,1+q} = \mathbf{s}_{S,1+q}(\hat{\mathbf{c}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,1+q}(\hat{c})]\hat{\boldsymbol{\Sigma}}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,1+q}(\hat{c})']\mathbf{s}_{S,1+q}(\hat{\mathbf{c}}),$$
$$\hat{\mathcal{V}}_{S+,1+q,1+q} = \mathbf{s}_{S,1+q}(\hat{\mathbf{c}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,1+q}(\hat{c})]\hat{\boldsymbol{\Sigma}}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,1+q}(\hat{c})']\mathbf{s}_{S,1+q}(\hat{\mathbf{c}}),$$
$$\hat{\mathbf{c}} = (\hat{c},\hat{c}).$$

with $\hat{\boldsymbol{\Sigma}}_{S-}$ and $\hat{\boldsymbol{\Sigma}}_{S+}$ denoting the estimators described above under heteroskedasticity or under clustering, using either a nearest neighbor approach $(\check{\boldsymbol{\Sigma}}_{S-}(J),\check{\boldsymbol{\Sigma}}_{S+}(J))$ or a plug-in estimated residuals approach $(\hat{\boldsymbol{\Sigma}}_{S-,1+q}(\hat{c}),\hat{\boldsymbol{\Sigma}}_{S+,1+q}(\hat{c}))$, and

$$\mathring{\mathcal{B}}_{S-,1+q,1+q} = \left([1,-\tilde{\boldsymbol{\gamma}}_{Y,q}(\hat{\mathbf{c}})']\hat{\boldsymbol{\mu}}^{(1+q)}_{S-,1+q}(x_-)\right)\mathbf{O}_{-,1+q,1+q}(\hat{c}),$$

$$\mathring{\mathcal{B}}_{S+,1+q,1+q} = \left([1,-\tilde{\boldsymbol{\gamma}}_{Y,q}(\hat{\mathbf{c}})']\hat{\boldsymbol{\mu}}^{(1+q)}_{S+,1+q}(x_+)\right)\mathbf{O}_{+,1+q,1+q}(\hat{c}),$$

where x_- and x_+ denote, respectively, the third quartile of $\{X_i : X_i < \bar{x}\}$ and first quartile of $\{X_i : X_i \geq \bar{x}\}$. Notice that $\mathring{\mathcal{B}}_{S-,1+q,1+q}$ and $\mathring{\mathcal{B}}_{S+,1+q,1+q}$ are not consistent estimates of their population counterparts, but will be more stable in applications. Furthermore, the resulting bandwidth choices $(\hat{\mathfrak{d}}_{S-,1+q,1+q},\hat{\mathfrak{d}}_{S+,1+q,1+q},\hat{\mathfrak{d}}_{\Delta S,1+q,1+q},\hat{\mathfrak{d}}_{\Sigma S,1+q,1+q})$ will have the correct rates (though "incorrect" constants), and hence $\hat{\mathcal{B}}_{S-,1+p,q}$ and $\hat{\mathcal{B}}_{S+,1+p,q}$ will be consistent estimators of their population counterparts, under appropriate regularity conditions, if $\hat{\mathbf{d}} = (\hat{d}_-,\hat{d}_+) = (\hat{\mathfrak{d}}_{S-,1+q,1+q},\hat{\mathfrak{d}}_{S+,1+q,1+q})$ or $\hat{\mathbf{d}} = (\hat{d}_-,\hat{d}_+) = (\hat{\mathfrak{d}}_{\Delta S,1+q,1+q},\hat{\mathfrak{d}}_{\Delta S,1+q,1+q})$ or $\hat{\mathbf{d}} = (\hat{d}_-,\hat{d}_+) = (\hat{\mathfrak{d}}_{\Sigma S,1+q,1+q},\hat{\mathfrak{d}}_{\Sigma S,1+q,1+q})$.

The following lemma establishes the consistency of these choices. The result applies to the heteroskedasticity-consistent case, but it can be extended to the clustered-consistent case using the same ideas, after replacing n by G, as appropriate, to account for the effective sample size in the latter case.

Lemma SA-23 *Let the conditions of Lemma SA-13 hold with $\varrho \geq q+3$. In addition, suppose that* $\min\{\mathring{\mathcal{B}}_{S-,1+q,1+q}, \mathring{\mathcal{B}}_{S+,1+q,1+q}, \mathring{\mathcal{B}}_{S+,1+q,1+q} - \mathring{\mathcal{B}}_{S-,1+q,1+q}, \mathring{\mathcal{B}}_{S-,1+q,1+q} + \mathring{\mathcal{B}}_{S-,1+q,1+q}\} \rightarrow_{\mathbb{P}} C \in (0,\infty)$, *and $k(\cdot)$ is Lipschitz continuous on its support. Then, if* $\min\{\mathcal{B}_{S-,1+p,q}, \mathcal{B}_{S+,1+p,q}, \mathcal{B}_{S+,1+p,q} - \mathcal{B}_{S-,1+p,q}, \mathcal{B}_{S-,1+p,q} + \mathcal{B}_{S+,1+p,q}\} \neq 0$,

$$\frac{\hat{\mathfrak{b}}_{S-,1+p,q}}{\mathfrak{b}_{S-,1+p,q}} \rightarrow_{\mathbb{P}} 1, \qquad \frac{\hat{\mathfrak{b}}_{S+,1+p,q}}{\mathfrak{b}_{S+,1+p,q}} \rightarrow_{\mathbb{P}} 1, \qquad \frac{\hat{\mathfrak{b}}_{\Delta S,1+p,q}}{\mathfrak{b}_{\Delta S,1+p,q}} \rightarrow_{\mathbb{P}} 1, \qquad \frac{\hat{\mathfrak{b}}_{\Sigma S,1+p,q}}{\mathfrak{b}_{\Sigma S,1+p,q}} \rightarrow_{\mathbb{P}} 1.$$

The proof of Lemma SA-23 is long and tedious. Its main intuition is as follows. First, it is shown that both $\hat{v} \rightarrow_{\mathbb{P}} 0$ and $\hat{d} \rightarrow_{\mathbb{P}} 0$, with $\hat{d} \in \{\hat{\mathfrak{d}}_{S-,1+q,1+q}, \hat{\mathfrak{d}}_{S+,1+q,1+q}, \hat{\mathfrak{d}}_{\Delta S,1+q,1+q}, \hat{\mathfrak{d}}_{\Sigma S,1+q,1+q}\}$, satisfy the following properties: $\frac{\hat{v}-v}{v} \rightarrow_{\mathbb{P}} 0$ and $\mathbb{P}[C_1 v \leq \hat{v} \leq C_2 v] \rightarrow 1$, and $\frac{\hat{d}-d}{d} \rightarrow_{\mathbb{P}} 0$ and $\mathbb{P}[C_1 d \leq \hat{d} \leq C_2 d] \rightarrow 1$, for some positive constants $C_1 < C_2$. This may require "truncation" of the preliminary bandwidths, which is commonly done in practice. Second, the previous facts combined with the Lipschitz continuity $k(\cdot)$ allows to "replace" the random bandwidths by their non-random counterparts. Finally, consistency of the underlying constants of the bandwidths selectors in Lemma SA-23 follows by the results obtained in the sections above.

12.3 Step 3: Choosing Bandwidth h

With the assumptions, choices and results above, we have the following implementations:

$$\hat{\mathfrak{h}}_{S-,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{\hat{\mathcal{V}}_{S-,\nu,p}/n}{\hat{\mathcal{B}}^2_{S-,\nu,p}}\right]^{\frac{1}{3+2p}},$$

$$\hat{\mathfrak{h}}_{S+,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{\hat{\mathcal{V}}_{S+,\nu,p}/n}{\hat{\mathcal{B}}^2_{S+,\nu,p}}\right]^{\frac{1}{3+2p}},$$

$$\hat{\mathfrak{h}}_{\Delta S,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\hat{\mathcal{V}}_{S-,\nu,p}+\hat{\mathcal{V}}_{S+,\nu,p})/n}{(\hat{\mathcal{B}}_{S+,\nu,p}-\hat{\mathcal{B}}_{S-,\nu,p})^2}\right]^{\frac{1}{3+2p}},$$

$$\hat{\mathfrak{h}}_{\Sigma S,\nu,p} = \left[\frac{1+2\nu}{2(1+p-\nu)} \frac{(\hat{\mathcal{V}}_{S-,\nu,p}+\hat{\mathcal{V}}_{S+,\nu,p})/n}{(\hat{\mathcal{B}}_{S+,\nu,p}+\hat{\mathcal{B}}_{S-,\nu,p})^2}\right]^{\frac{1}{3+2p}},$$

where now the preliminary constant estimates are chosen as follows.

- **Bias Constants**:

$$\hat{\mathcal{B}}_{S-,\nu,p} = \mathbf{O}_{-,\nu,p}(\hat{c}) \frac{[1, -\tilde{\boldsymbol{\gamma}}_{Y,p}(\hat{\mathbf{b}})']\boldsymbol{\mu}^{(1+p)}_{S-,p}(\hat{b}_-)}{(1+p)!},$$

$$\hat{\mathcal{B}}_{S+,\nu,p} = \mathbf{O}_{+,\nu,p}(\hat{c})\frac{[1,-\tilde{\boldsymbol{\gamma}}_{Y,p}(\hat{\mathbf{b}})']\boldsymbol{\mu}^{(1+p)}_{S+,p}(\hat{b}_+)}{(1+p)!},$$

with $\hat{\mathbf{b}} = (\hat{b}_-, \hat{b}_+)$ chosen out of $\{\hat{\mathfrak{b}}_{-,1+p,q}, \hat{\mathfrak{b}}_{+,1+p,q}, \hat{\mathfrak{b}}_{\Delta,1+p,q}, \hat{\mathfrak{b}}_{\Sigma,1+p,q}\}$ as appropriate according to the target bandwidth selector.

- **Variance Constants**:

$$\hat{\mathcal{V}}_{S-,\nu,p} = \mathbf{s}_{S,\nu}(\hat{\mathbf{c}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(\hat{c})]\hat{\boldsymbol{\Sigma}}_{S-}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{-,p}(\hat{c})']\mathbf{s}_{S,\nu}(\hat{\mathbf{c}}),$$

$$\hat{\mathcal{V}}_{S+,\nu,p} = \mathbf{s}_{S,\nu}(\hat{\mathbf{c}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(\hat{c})]\hat{\boldsymbol{\Sigma}}_{S+}[\mathbf{I}_{1+d}\otimes\mathbf{P}_{+,p}(\hat{c})']\mathbf{s}_{S,\nu}(\hat{\mathbf{c}}),$$

$$\hat{\mathbf{c}} = (\hat{c},\hat{c}).$$

with $\hat{\boldsymbol{\Sigma}}_{S-}$ and $\hat{\boldsymbol{\Sigma}}_{S+}$ denoting the estimators described above under heteroskedasticity or under clustering, using either a nearest neighbor approach $(\check{\boldsymbol{\Sigma}}_{S-}(J), \check{\boldsymbol{\Sigma}}_{S+}(J))$ or a plug-in estimated residuals approach $(\hat{\boldsymbol{\Sigma}}_{S-,p}(\hat{c}), \hat{\boldsymbol{\Sigma}}_{S+,p}(\hat{c}))$.

The following lemma establishes the consistency of these choices under some regularity conditions.

Theorem SA-1 *Suppose the assumptions in Lemma SA-23 hold. Then, if* $\min\{\mathcal{B}_{S-,\nu,p}, \mathcal{B}_{S+,\nu,p}, \mathcal{B}_{S+,\nu,p}-\mathcal{B}_{S-,\nu,p}, \mathcal{B}_{S-,\nu,p}+\mathcal{B}_{S+,\nu,p}\} \neq 0$,

$$\frac{\hat{\mathfrak{h}}_{-,\nu,p}}{\mathfrak{h}_{-,\nu,p}} \to_p 1, \qquad \frac{\hat{\mathfrak{h}}_{+,\nu,p}}{\mathfrak{h}_{+,\nu,p}} \to_p 1, \qquad \frac{\hat{\mathfrak{h}}_{\Delta,\nu,p}}{\mathfrak{h}_{\Delta,\nu,p}} \to_p 1, \qquad \frac{\hat{\mathfrak{h}}_{\Sigma,\nu,p}}{\mathfrak{h}_{\Sigma,\nu,p}} \to_p 1.$$

The proof of this lemma is analogous to the proof of Lemma SA-23.

12.4 Bias-Correction Estimation

Once the bandwidths have been chosen, it is easy to implement the bias-correction methods. Specifically, the bias-corrected covariate-adjusted sharp RD estimator is

$$\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\mathbf{h},\mathbf{b}) = \tilde{\tau}_{Y,\nu}(\mathbf{h}) - \left[h_+^{1+p-\nu}\hat{\mathcal{B}}_{S+,p,q}(h_+,b_+) - h_-^{1+p-\nu}\hat{\mathcal{B}}_{S+,p,q}(h_-,b_-)\right],$$

$$\hat{\mathcal{B}}_{S-,\nu,p,q}(h,b) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}^{-1}_{-,p}(h)\boldsymbol{\vartheta}_{-,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{S-,q}(b)}{(1+p)!},$$

$$\hat{\mathcal{B}}_{S+,\nu,p,q}(h,b) = \mathbf{s}_{S,\nu}(\mathbf{h})'[\mathbf{I}_{1+d}\otimes\boldsymbol{\Gamma}^{-1}_{+,p}(h)\boldsymbol{\vartheta}_{+,p}(h)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{S+,q}(b)}{(1+p)!},$$

and thus its feasible version is

$$\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\hat{\mathbf{h}},\hat{\mathbf{b}}) = \tilde{\tau}_{Y,\nu}(\hat{\mathbf{h}}) - \left[\hat{h}_+^{1+p-\nu}\hat{\mathcal{B}}_{S+,p,q}(\hat{h}_+,\hat{b}_+) - \hat{h}_-^{1+p-\nu}\hat{\mathcal{B}}_{S+,p,q}(\hat{h}_-,\hat{b}_-)\right],$$

$$\hat{\mathscr{B}}_{S-,\nu,p,q}(\hat{h}_-,\hat{b}_-) = \mathbf{s}_{S,\nu}(\hat{\mathbf{h}})'[\mathbf{I}_{1+d}\otimes\mathbf{\Gamma}^{-1}_{-,p}(\hat{h}_-)\boldsymbol{\vartheta}_{-,p}(\hat{h}_-)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{S-,q}(\hat{b}_-)}{(1+p)!},$$

$$\hat{\mathscr{B}}_{S+,\nu,p,q}(\hat{h}_+,\hat{b}_+) = \mathbf{s}_{S,\nu}(\hat{\mathbf{h}})'[\mathbf{I}_{1+d}\otimes\mathbf{\Gamma}^{-1}_{+,p}(\hat{h}_+)\boldsymbol{\vartheta}_{+,p}(\hat{h}_+)]\frac{\hat{\boldsymbol{\mu}}^{(1+p)}_{S+,q}(\hat{b}_+)}{(1+p)!},$$

where $\hat{\mathbf{b}} = (\hat{b}_-,\hat{b}_+)$ is chosen as discussed in Step 2 above, and $\hat{\mathbf{h}} = (\hat{h}_-,\hat{h}_+)$ is chosen as discussed in Step 3 above. Notice that $\hat{\mathbf{c}}$ and $\hat{\mathbf{d}}$ are not used directly in this construction, only indirectly through $\hat{\mathbf{b}}$ and $\hat{\mathbf{h}}$.

12.5 Variance Estimation

Once the bandwidths have been chosen, the robust variance estimation (after bias-correction) is done by plug-in methods. Specifically, the robust variance estimator is as follows.

- **Robust Bias-Correction NN Variance Estimator**:

$$\check{\mathsf{V}}\mathsf{ar}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\hat{\mathbf{h}},\hat{\mathbf{b}})] = \frac{1}{n\hat{h}_-^{1+2\nu}}\check{\mathscr{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}) + \frac{1}{nh_+^{1+2\nu}}\check{\mathscr{V}}_{S+,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}),$$

$$\check{\mathscr{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}) = \mathbf{s}_{S,\nu}(\hat{\mathbf{h}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(\hat{h}_-,\hat{b}_-)]\check{\mathbf{\Sigma}}_{S-}(J)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(\hat{h}_-,\hat{b}_-)']\mathbf{s}_{S,\nu}(\hat{\mathbf{h}}),$$

$$\check{\mathscr{V}}_{S+,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}) = \mathbf{s}_{S,\nu}(\hat{\mathbf{h}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(\hat{h}_+,\hat{b}_+)]\check{\mathbf{\Sigma}}_{S+}(J)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(\hat{h}_+,\hat{b}_+)']\mathbf{s}_{S,\nu}(\hat{\mathbf{h}}).$$

- **Robust Bias-Correction PR Variance Estimator**:

$$\hat{\mathsf{V}}\mathsf{ar}[\tilde{\tau}^{\mathsf{bc}}_{Y,\nu}(\hat{\mathbf{h}},\hat{\mathbf{b}})] = \frac{1}{n\hat{h}_-^{1+2\nu}}\hat{\mathscr{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}) + \frac{1}{nh_+^{1+2\nu}}\hat{\mathscr{V}}_{S+,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}),$$

$$\hat{\mathscr{V}}^{\mathsf{bc}}_{S-,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}) = \mathbf{s}_{S,\nu}(\hat{\mathbf{h}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(\hat{h}_-,\hat{b}_-)]\hat{\mathbf{\Sigma}}_{S-,q}(\hat{h}_-)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{-,p,q}(\hat{h}_-,\hat{b}_-)']\mathbf{s}_{S,\nu}(\hat{\mathbf{h}}),$$

$$\hat{\mathscr{V}}_{S+,\nu,p,q}(\hat{\mathbf{h}},\hat{\mathbf{b}}) = \mathbf{s}_{S,\nu}(\hat{\mathbf{h}})'[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(\hat{h}_+,\hat{b}_+)]\hat{\mathbf{\Sigma}}_{S+,q}(\hat{h}_+)[\mathbf{I}_{1+d}\otimes\mathbf{P}^{\mathsf{bc}}_{+,p,q}(\hat{h}_+,\hat{b}_+)']\mathbf{s}_{S,\nu}(\hat{\mathbf{h}}).$$

where $\hat{\mathbf{b}} = (\hat{b}_-,\hat{b}_+)$ is chosen as discussed in Step 2 above, and $\hat{\mathbf{h}} = (\hat{h}_-,\hat{h}_+)$ is chosen as discussed in Step 3 above. Notice that $\hat{\mathbf{c}}$ and $\hat{\mathbf{d}}$ are not used directly in this construction, only indirectly through $\hat{\mathbf{b}}$ and $\hat{\mathbf{h}}$.

13 Fuzzy RD Designs

Follows exactly the same logic outlined for the sharp RD setting, after replacing $\mathbf{S}_i = (Y_i, \mathbf{Z}_i')'$ by $\mathbf{F}_i = (Y_i, T_i, \mathbf{Z}_i')'$, and the linear combination $\mathbf{s}_{S,\nu}(\cdot)$ by $\mathbf{f}_{F,\nu}(\cdot)$, as discussed previously for estimation and inference. We do not reproduce the implementation details here to conserve space. Nonetheless, all these results are also implemented in the companion general purpose `Stata` and `R` packages described in Calonico, Cattaneo, Farrell, and Titiunik (2017).

Part V

Simulation Results

We provide further details on the data generating processes (DGPs) employed in our simulation study and further numerical results not presented in the paper.

We consider four data generating processes constructed using the data of Lee (2008), who studies the incumbency advantage in U.S. House elections exploiting the discontinuity generated by the rule that the party with a majority vote share wins. The forcing variable is the difference in vote share between the Democratic candidate and her strongest opponent in a given election, with the threshold level set at $\bar{x} = 0$. The outcome variable is the Democratic vote share in the following election.

All DGPs employ the same basic simulation setup, with the only exception of the functional form of the regression function and a correlation parameter. Specifically, for each replication, the data is generated as i.i.d. draws, $i = 1, 2, ..., n$ with $n = 1,000$, as follows:

$$Y_i = \mu_{y,j}(X_i, Z_i) + \varepsilon_{y,i} \qquad Z_i = \mu_z(X_i) + \varepsilon_{z,i} \qquad X_i \sim (2\mathcal{B}(2,4) - 1)$$

where

$$\begin{pmatrix} \varepsilon_{y,i} \\ \varepsilon_{z,i} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_j\right), \qquad \mathbf{\Sigma}_j = \begin{pmatrix} \sigma_y^2 & \rho_j \sigma_y \sigma_z \\ \rho_j \sigma_y \sigma_z & \sigma_z^2 \end{pmatrix},$$

with $\mathcal{B}(a, b)$ denoting a beta distribution with parameters a and b. The regression functions $\mu_{y,j}(x, z)$ and $\mu_z(z)$, and the form of the variance-covariance matrix Σ_j, $j = 1, 2, 3, 4$, are discussed below.

- **Model 1** does not include additional covariates. The regression function is obtained by fitting a 5-th order global polynomial with different coefficients for $X_i < 0$ and $X_i > 0$. The resulting coefficients estimated on the Lee (2008) data, after discarding observations with past vote share differences greater than 0.99 and less than -0.99, leads to the following functional form:

$$\mu_{y,1}(x, z) = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0 \\ 0.52 + 0.84x - 3.00x^2 + 07.99x^3 - 09.01x^4 + 3.56x^5 & \text{if } x \geq 0 \end{cases}$$

  We also compute $\sigma_y = 0.1295$ and $\sigma_z = 0.1353$ from the same sample.

- **Model 2** includes one additional covariate (previous democratic vote share) and all parameters are also obtained from the real data. The regression function for the outcome is obtained by fitting a 5-th order global polynomial on X_i with different coefficients for $X_i < 0$ and $X_i > 0$, now with the addition of the covariate Z_i, leading to the following regression function:

$$\mu_{y,2}(x,z) = \begin{cases} 0.36 + 0.96x + 5.47x^2 + 15.28x^3 + 15.87x^4 + 5.14x^5 + 0.22z & \text{if } x < 0 \\ 0.38 + 0.62x - 2.84x^2 + 08.42x^3 - 10.24x^4 + 4.31x^5 + 0.28z & \text{if } x \geq 0 \end{cases}.$$

Similarly, we obtain the regression function for the covariate Z_i by fitting a 5-th order global polynomial on X_i on either side of the threshold:

$$\mu_z(x) = \begin{cases} 0.49 + 1.06x + 5.74x^2 + 17.14x^3 + 19.75x^4 + 7.47x^5 & \text{if } x < 0 \\ 0.49 + 0.61x + 0.23x^2 - 03.46x^3 + 06.43x^4 - 3.48x^5 & \text{if } x \geq 0 \end{cases}.$$

The only difference between models 2 to 4 is the assumed value of ρ, the correlation between the residuals $\varepsilon_{y,i}$ and $\varepsilon_{z,i}$. In Model 2, we use $\rho = 0.2692$ as obtained from the data.

- **Model 3** takes Model 2 but sets the residual correlation ρ between the outcome and covariate to zero.
- **Model 4** takes Model 2 but doubles the residual correlation ρ between the outcome and covariate equations.

We consider $5,000$ replications. We compare the standard RD estimator ($\hat{\tau}$) and the covariate-adjusted RD estimator ($\tilde{\tau}$), with both infeasible and data-driven MSE-optimal and CER-optimal bandwidth choices. To analyze the performance of our inference procedures, we report average bias of the point estimators, as well as average coverage rate and interval length of nominal 95% confidence intervals, all across the $5,000$ replications. In addition, we also explore the performance of our data-driven bandwidth selectors by reporting some of their main statistical features, such as mean, median and standard deviation. We report tables with estimates using triangular kernel with different standard errors estimators: nearest neighbor (NN) heteroskedasticity-robust, HC_1, HC_2 and HC_3 variance estimators.

The numerical results are given in the following tables, which follow the same structure as discussed in the paper. All findings are highly consistent with our large-sample theoretical results and the simulation results discussed in the paper.

Table SA-1: Simulation Results (MSE, Bias, Empirical Coverage and Interval Length), NN

	$\hat{\tau}$				$\tilde{\tau}$				**Change** (%)			
	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL
Model 1												
MSE-POP	0.045	0.014	0.934	0.198	0.046	0.014	0.936	0.197	0.4	−0.5	0.2	−0.6
MSE-EST	0.046	0.020	0.909	0.170	0.047	0.020	0.907	0.170	0.4	−0.8	−0.3	−0.3
CER-POP	0.053	0.008	0.932	0.240	0.053	0.008	0.928	0.238	0.5	−1.8	−0.4	−0.8
CER-EST	0.050	0.013	0.931	0.205	0.051	0.012	0.927	0.204	0.8	−2.2	−0.4	−0.5
Model 2												
MSE-POP	0.048	0.015	0.930	0.212	0.041	0.010	0.935	0.183	−16.4	−33.1	0.5	−13.6
MSE-EST	0.050	0.020	0.912	0.187	0.041	0.012	0.920	0.162	−18.2	−36.6	0.9	−13.4
CER-POP	0.056	0.009	0.928	0.257	0.048	0.006	0.930	0.221	−15.0	−34.6	0.3	−13.9
CER-EST	0.055	0.012	0.926	0.225	0.046	0.008	0.936	0.194	−16.0	−36.6	1.0	−13.6
Model 3												
MSE-POP	0.046	0.015	0.929	0.199	0.044	0.012	0.934	0.192	−5.0	−18.4	0.5	−3.7
MSE-EST	0.048	0.020	0.909	0.176	0.044	0.016	0.915	0.169	−7.1	−17.9	0.7	−4.0
CER-POP	0.053	0.009	0.929	0.241	0.051	0.007	0.927	0.232	−3.6	−20.1	−0.2	−3.9
CER-EST	0.052	0.012	0.928	0.212	0.049	0.010	0.931	0.203	−4.7	−19.0	0.4	−4.1
Model 4												
MSE-POP	0.051	0.015	0.932	0.224	0.035	0.008	0.938	0.159	−32.1	−48.4	0.6	−29.0
MSE-EST	0.052	0.020	0.914	0.197	0.035	0.009	0.929	0.142	−33.2	−54.7	1.6	−28.2
CER-POP	0.059	0.009	0.928	0.271	0.041	0.005	0.937	0.192	−31.2	−49.5	1.0	−29.2
CER-EST	0.058	0.013	0.930	0.237	0.039	0.006	0.942	0.170	−31.6	−54.5	1.3	−28.4

Notes:

(i) All estimators are computed using the triangular kernel, NN variance estimation, and two bandwidths (h and b).
(ii) Columns $\hat{\tau}$ and $\tilde{\tau}$ correspond to, respectively, standard RD estimation and covariate-adjusted RD estimation; columns "$\sqrt{MSE}$" report the square root of the mean square error of point estimator; columns "Bias" report average bias relative to target population parameter; and columns "EC" and "IL" report, respectively, empirical coverage and interval length of robust bias-corrected 95% confidence intervals.
(iii) Rows correspond to bandwidth method used to construct the estimator and inference procedures. Rows "MSE-POP" and "MSE-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven MSE-optimal bandwidths (without or with covariate adjustment depending on the column). Rows "CER-POP" and "CER-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven CER-optimal bandwidths (without or with covariate adjustment depending on the column).

Table SA-2: Simulation Results (MSE, Bias, Empirical Coverage and Interval Length), HC_1

	$\hat{\tau}$				$\tilde{\tau}$				**Change** (%)			
	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL
Model 1												
MSE-POP	0.045	0.014	0.935	0.196	0.046	0.014	0.933	0.195	0.4	−0.5	−0.2	−0.6
MSE-EST	0.046	0.020	0.910	0.169	0.046	0.020	0.909	0.169	0.4	−0.8	−0.1	−0.3
CER-POP	0.053	0.008	0.929	0.235	0.053	0.008	0.923	0.233	0.5	−1.8	−0.6	−0.8
CER-EST	0.050	0.013	0.930	0.202	0.051	0.012	0.925	0.201	0.9	−2.2	−0.5	−0.5
Model 2												
MSE-POP	0.048	0.015	0.929	0.210	0.041	0.010	0.935	0.181	−16.4	−33.1	0.7	−13.5
MSE-EST	0.050	0.020	0.911	0.186	0.041	0.012	0.921	0.161	−18.2	−36.6	1.2	−13.4
CER-POP	0.056	0.009	0.924	0.252	0.048	0.006	0.929	0.217	−15.0	−34.6	0.5	−13.7
CER-EST	0.055	0.012	0.928	0.222	0.046	0.008	0.933	0.192	−15.9	−36.0	0.6	−13.5
Model 3												
MSE-POP	0.046	0.015	0.929	0.197	0.044	0.012	0.932	0.190	−5.0	−18.4	0.3	−3.6
MSE-EST	0.048	0.019	0.910	0.175	0.044	0.016	0.916	0.168	−7.1	−17.8	0.7	−4.0
CER-POP	0.053	0.009	0.923	0.236	0.051	0.007	0.924	0.227	−3.6	−20.1	0.1	−3.8
CER-EST	0.052	0.012	0.927	0.209	0.049	0.010	0.928	0.200	−4.7	−18.7	0.1	−4.1
Model 4												
MSE-POP	0.051	0.015	0.929	0.222	0.035	0.008	0.938	0.157	−32.1	−48.4	0.9	−28.9
MSE-EST	0.052	0.020	0.913	0.196	0.035	0.009	0.930	0.141	−33.2	−54.6	1.8	−28.3
CER-POP	0.059	0.009	0.926	0.266	0.041	0.005	0.931	0.189	−31.2	−49.5	0.5	−29.1
CER-EST	0.058	0.013	0.929	0.235	0.039	0.006	0.936	0.168	−31.5	−54.0	0.8	−28.4

Notes:

(i) All estimators are computed using the triangular kernel, HC_1 variance estimation, and two bandwidths (h and b).
(ii) Columns $\hat{\tau}$ and $\tilde{\tau}$ correspond to, respectively, standard RD estimation and covariate-adjusted RD estimation; columns "$\sqrt{MSE}$" report the square root of the mean square error of point estimator; columns "Bias" report average bias relative to target population parameter; and columns "EC" and "IL" report, respectively, empirical coverage and interval length of robust bias-corrected 95% confidence intervals.
(iii) Rows correspond to bandwidth method used to construct the estimator and inference procedures. Rows "MSE-POP" and "MSE-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven MSE-optimal bandwidths (without or with covariate adjustment depending on the column). Rows "CER-POP" and "CER-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven CER-optimal bandwidths (without or with covariate adjustment depending on the column).

Table SA-3: Simulation Results (MSE, Bias, Empirical Coverage and Interval Length), HC_2

	$\hat{\tau}$				$\tilde{\tau}$				**Change** (%)			
	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL
Model 1												
MSE-POP	0.045	0.014	0.936	0.198	0.046	0.014	0.935	0.197	0.4	−0.5	−0.2	−0.6
MSE-EST	0.046	0.020	0.912	0.170	0.046	0.020	0.910	0.170	0.4	−0.8	−0.3	−0.3
CER-POP	0.053	0.008	0.934	0.239	0.053	0.008	0.928	0.237	0.5	−1.8	−0.6	−0.8
CER-EST	0.050	0.013	0.932	0.205	0.050	0.012	0.929	0.204	0.8	−2.1	−0.4	−0.5
Model 2												
MSE-POP	0.048	0.015	0.932	0.212	0.041	0.010	0.939	0.183	−16.4	−33.1	0.8	−13.5
MSE-EST	0.050	0.020	0.912	0.187	0.041	0.012	0.924	0.162	−18.2	−36.6	1.3	−13.4
CER-POP	0.056	0.009	0.927	0.255	0.048	0.006	0.932	0.220	−15.0	−34.6	0.6	−13.7
CER-EST	0.055	0.013	0.930	0.224	0.046	0.008	0.936	0.194	−15.9	−36.0	0.6	−13.5
Model 3												
MSE-POP	0.046	0.015	0.931	0.199	0.044	0.012	0.934	0.192	−5.0	−18.4	0.3	−3.6
MSE-EST	0.047	0.020	0.913	0.176	0.044	0.016	0.920	0.169	−7.2	−17.8	0.8	−4.0
CER-POP	0.053	0.009	0.928	0.240	0.051	0.007	0.931	0.231	−3.6	−20.1	0.3	−3.8
CER-EST	0.052	0.012	0.931	0.211	0.049	0.010	0.931	0.202	−4.7	−18.6	0.1	−4.1
Model 4												
MSE-POP	0.051	0.015	0.930	0.224	0.035	0.008	0.940	0.159	−32.1	−48.4	1.0	−29.0
MSE-EST	0.052	0.020	0.916	0.197	0.035	0.009	0.931	0.142	−33.2	−54.6	1.7	−28.3
CER-POP	0.059	0.009	0.929	0.270	0.041	0.005	0.933	0.191	−31.2	−49.5	0.5	−29.1
CER-EST	0.057	0.013	0.932	0.237	0.039	0.006	0.941	0.170	−31.5	−54.0	0.9	−28.4

Notes:

(i) All estimators are computed using the triangular kernel, HC_2 variance estimation, and two bandwidths (h and b).
(ii) Columns $\hat{\tau}$ and $\tilde{\tau}$ correspond to, respectively, standard RD estimation and covariate-adjusted RD estimation; columns "$\sqrt{MSE}$" report the square root of the mean square error of point estimator; columns "Bias" report average bias relative to target population parameter; and columns "EC" and "IL" report, respectively, empirical coverage and interval length of robust bias-corrected 95% confidence intervals.
(iii) Rows correspond to bandwidth method used to construct the estimator and inference procedures. Rows "MSE-POP" and "MSE-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven MSE-optimal bandwidths (without or with covariate adjustment depending on the column). Rows "CER-POP" and "CER-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven CER-optimal bandwidths (without or with covariate adjustment depending on the column).

Table SA-4: Simulation Results (MSE, Bias, Empirical Coverage and Interval Length), HC_3

	$\hat{\tau}$				$\tilde{\tau}$				**Change** (%)			
	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL	$\sqrt{MSE}$	Bias	EC	IL
Model 1												
MSE-POP	0.045	0.014	0.943	0.203	0.046	0.014	0.941	0.201	0.4	−0.5	−0.2	−0.6
MSE-EST	0.046	0.021	0.918	0.173	0.046	0.020	0.914	0.172	0.4	−0.8	−0.4	−0.3
CER-POP	0.053	0.008	0.939	0.247	0.053	0.008	0.936	0.245	0.5	−1.8	−0.3	−0.8
CER-EST	0.050	0.013	0.940	0.209	0.050	0.013	0.938	0.208	0.8	−2.1	−0.2	−0.5
Model 2												
MSE-POP	0.048	0.015	0.938	0.216	0.041	0.010	0.941	0.187	−16.4	−33.1	0.3	−13.6
MSE-EST	0.050	0.020	0.919	0.189	0.041	0.013	0.929	0.164	−18.2	−36.6	1.0	−13.4
CER-POP	0.056	0.009	0.937	0.263	0.048	0.006	0.937	0.227	−15.0	−34.6	0.0	−13.8
CER-EST	0.054	0.013	0.933	0.229	0.046	0.008	0.944	0.198	−15.9	−36.0	1.1	−13.5
Model 3												
MSE-POP	0.046	0.015	0.937	0.203	0.044	0.012	0.940	0.196	−5.0	−18.4	0.3	−3.6
MSE-EST	0.047	0.020	0.917	0.178	0.044	0.016	0.923	0.171	−7.2	−17.8	0.7	−4.0
CER-POP	0.053	0.009	0.937	0.247	0.051	0.007	0.937	0.238	−3.6	−20.1	0.0	−3.8
CER-EST	0.051	0.013	0.935	0.216	0.049	0.010	0.936	0.207	−4.8	−18.6	0.1	−4.1
Model 4												
MSE-POP	0.051	0.015	0.938	0.229	0.035	0.008	0.944	0.162	−32.1	−48.4	0.6	−29.0
MSE-EST	0.052	0.020	0.923	0.200	0.035	0.009	0.938	0.143	−33.2	−54.6	1.6	−28.3
CER-POP	0.059	0.009	0.936	0.278	0.041	0.005	0.940	0.197	−31.2	−49.5	0.4	−29.2
CER-EST	0.057	0.013	0.936	0.242	0.039	0.006	0.947	0.173	−31.5	−54.0	1.1	−28.4

Notes:

(i) All estimators are computed using the triangular kernel, HC_3 variance estimation, and two bandwidths (h and b).
(ii) Columns $\hat{\tau}$ and $\tilde{\tau}$ correspond to, respectively, standard RD estimation and covariate-adjusted RD estimation; columns "$\sqrt{MSE}$" report the square root of the mean square error of point estimator; columns "Bias" report average bias relative to target population parameter; and columns "EC" and "IL" report, respectively, empirical coverage and interval length of robust bias-corrected 95% confidence intervals.
(iii) Rows correspond to bandwidth method used to construct the estimator and inference procedures. Rows "MSE-POP" and "MSE-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven MSE-optimal bandwidths (without or with covariate adjustment depending on the column). Rows "CER-POP" and "CER-EST" correspond to, respectively, procedures using infeasible population and feasible data-driven CER-optimal bandwidths (without or with covariate adjustment depending on the column).

Table SA-5: Simulation Results (Data-Driven Bandwidth Selectors), NN

	Pop.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
Model 1								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.144	0.079	0.167	0.192	0.196	0.222	0.337	0.041
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.144	0.078	0.166	0.190	0.196	0.221	0.319	0.041
Model 2								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.085	0.170	0.194	0.200	0.226	0.328	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.158	0.079	0.171	0.198	0.202	0.231	0.333	0.042
Model 3								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.086	0.169	0.193	0.199	0.225	0.333	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.154	0.080	0.169	0.195	0.200	0.226	0.329	0.042
Model 4								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.084	0.170	0.195	0.200	0.227	0.321	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.161	0.087	0.172	0.200	0.203	0.232	0.340	0.043

Notes:

(i) All estimators are computed using the triangular kernel, NN variance estimation, and two bandwidths (h and b).
(ii) Column "Pop." reports target population bandwidth, while the other columns report summary statistics of the distribution of feasible data-driven estimated bandwidths.
(iii) Rows $\hat{\mathfrak{h}}_{\hat{\tau}}$ and $\tilde{\mathfrak{h}}_{\tilde{\tau}}$ corresponds to feasible data-driven MSE-optimal bandwidth selectors without and with covariate adjustment, respectively.

Table SA-6: Simulation Results (Data-Driven Bandwidth Selectors), HC_1

	Pop.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
Model 1								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.144	0.085	0.167	0.191	0.196	0.222	0.326	0.041
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.144	0.083	0.166	0.190	0.195	0.220	0.320	0.040
Model 2								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.088	0.169	0.195	0.199	0.226	0.322	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.158	0.087	0.171	0.198	0.202	0.230	0.337	0.042
Model 3								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.088	0.168	0.193	0.198	0.225	0.322	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.154	0.084	0.169	0.194	0.199	0.225	0.322	0.041
Model 4								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.087	0.169	0.195	0.200	0.227	0.324	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.161	0.092	0.172	0.199	0.202	0.230	0.333	0.043

Notes:

(i) All estimators are computed using the triangular kernel, HC_1 variance estimation, and two bandwidths (h and b).
(ii) Column "Pop." reports target population bandwidth, while the other columns report summary statistics of the distribution of feasible data-driven estimated bandwidths.
(iii) Rows $\hat{\mathfrak{h}}_{\hat{\tau}}$ and $\tilde{\mathfrak{h}}_{\tilde{\tau}}$ corresponds to feasible data-driven MSE-optimal bandwidth selectors without and with covariate adjustment, respectively.

Table SA-7: Simulation Results (Data-Driven Bandwidth Selectors), HC_2

	Pop.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
Model 1								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.144	0.085	0.168	0.192	0.197	0.223	0.326	0.041
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.144	0.084	0.167	0.191	0.196	0.221	0.320	0.040
Model 2								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.088	0.170	0.195	0.200	0.227	0.323	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.158	0.087	0.172	0.199	0.202	0.231	0.337	0.042
Model 3								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.088	0.169	0.194	0.199	0.226	0.324	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.154	0.085	0.170	0.195	0.200	0.226	0.321	0.041
Model 4								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.088	0.170	0.196	0.200	0.228	0.324	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.161	0.093	0.172	0.200	0.203	0.231	0.333	0.043

Notes:

(i) All estimators are computed using the triangular kernel, HC_2 variance estimation, and two bandwidths (h and b).
(ii) Column "Pop." reports target population bandwidth, while the other columns report summary statistics of the distribution of feasible data-driven estimated bandwidths.
(iii) Rows $\hat{\mathfrak{h}}_{\hat{\tau}}$ and $\tilde{\mathfrak{h}}_{\tilde{\tau}}$ corresponds to feasible data-driven MSE-optimal bandwidth selectors without and with covariate adjustment, respectively.

Table SA-8: Simulation Results (Data-Driven Bandwidth Selectors), HC_3

	Pop.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
Model 1								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.144	0.086	0.169	0.194	0.198	0.225	0.326	0.040
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.144	0.085	0.168	0.192	0.197	0.222	0.320	0.040
Model 2								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.089	0.171	0.197	0.201	0.229	0.325	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.158	0.088	0.173	0.200	0.204	0.232	0.338	0.042
Model 3								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.090	0.171	0.196	0.201	0.228	0.326	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.154	0.086	0.171	0.197	0.201	0.228	0.319	0.041
Model 4								
$\hat{\mathfrak{h}}_{\hat{\tau}}$	0.156	0.089	0.172	0.198	0.202	0.230	0.325	0.042
$\tilde{\mathfrak{h}}_{\tilde{\tau}}$	0.161	0.094	0.174	0.201	0.204	0.233	0.333	0.043

Notes:

(i) All estimators are computed using the triangular kernel, HC_3 variance estimation, and two bandwidths (h and b).
(ii) Column "Pop." reports target population bandwidth, while the other columns report summary statistics of the distribution of feasible data-driven estimated bandwidths.
(iii) Rows $\hat{\mathfrak{h}}_{\hat{\tau}}$ and $\tilde{\mathfrak{h}}_{\tilde{\tau}}$ corresponds to feasible data-driven MSE-optimal bandwidth selectors without and with covariate adjustment, respectively.

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231–263.

Arai, Y., and H. Ichimura (2016): "Optimal bandwidth selection for the fuzzy regression discontinuity estimator," *Economic Letters*, 141(1), 103–106.

——— (2018): "Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator," *Quantitative Economics*, 9(1), 441–482.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018): "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113(522), 767–779.

——— (2019): "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression," arXiv:1808.01398.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017): "`rdrobust`: Software for Regression Discontinuity Designs," *Stata Journal*, 17(2), 372–404.

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

——— (2014b): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

——— (2015): "`rdrobust`: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs," *R Journal*, 7(1), 38–51.

Cameron, A. C., and D. L. Miller (2015): "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50(2), 317–372.

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

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

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

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

Long, J. S., and L. H. Ervin (2000): "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model," *The American Statistician*, 54(3), 217–224.

MacKinnon, J. G. (2012): “Thirty years of heteroskedasticity-robust inference,” in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, ed. by X. Chen, and N. R. Swanson. Springer.