Placebo Discontinuity Design

Rahul Singh
Harvard Department of Economics and Society of Fellows

Moses Stewart
Harvard Department of Economics

July 2025

Abstract

Standard regression discontinuity design (RDD) models rely on the continuity of expected potential outcomes at the cutoff. The standard continuity assumption can be violated by strategic manipulation of the running variable, which is realistic when the cutoff is widely known and when the treatment of interest is a social program or government benefit. In this work, we identify the treatment effect despite such a violation, by leveraging a placebo treatment and a placebo outcome. We introduce a local instrumental variable estimator. Our estimator decomposes into two terms: the standard RDD estimator of the target outcome's discontinuity, and a new adjustment term based on the placebo outcome's discontinuity. We show that our estimator is consistent, and we justify a robust bias-corrected inference procedure. Our method expands the applicability of RDD to settings with strategic behavior around the cutoff, which commonly arise in social science.

1 Introduction

Regression discontinuity design (RDD) strategies are widely used to estimate the effect of a treatment on an outcome of interest, when the probability of treatment assignment exhibits a discontinuous jump at a known threshold (Thistlethwaite and Campbell, 1960). Applications of RDD are broad and include studies in electoral politics (Hall, 2015; Spenkuch and Toniatti, 2018), education (Cook and Kang, 2016; Goodman et al., 2019), housing markets (Kumar, 2018), and market design (Abdulkadiroğlu et al., 2017; Kong, 2021), among others.

The main identifying assumption in standard RDD is the continuity of expected potential outcomes at the cutoff (Hahn et al., 2001). This assumption ensures that, in the absence of treatment, units just above and below the cutoff are comparable. As discussed in Remark 1, this condition is closely related to the requirement that the conditional distribution of unobserved confounding given the running variable is continuous at the cutoff (Lee and Lemieux, 2010). To empirically assess the plausibility of this assumption, researchers often conduct falsification tests using placebo variables.[1] For example, Imbens and Lemieux (2008) and Cattaneo et al. (2024) recommend checking for discontinuities in placebo outcomes at the cutoff as a diagnostic for invalid RDD. However, when these falsification tests indicate a failure of the RDD continuity assumption, current methods offer no path forward for recovering the treatment effect.

In this paper, we propose a new identification strategy that leverages a placebo treatment and a placebo outcome to recover the treatment effect even when the continuity of expected potential outcomes

[1] The literature also considers predetermined covariates, of which the placebo treatment may be viewed as a special case.

may fail. We introduce a local linear instrumental variable estimator that calculates the treatment effect as the difference between the right-limit of the observed outcome and an adjusted left-limit that accounts for unobserved confounding via placebo variables. The estimator decomposes into a standard RDD estimate for the target outcome, $\hat{\tau}^{y}_{\text{rdd}}$, plus an adjustment term $(\hat{\tau}^{w}_{\text{rdd}})^{\top}\hat{\gamma}_{-}$, where $\hat{\tau}^{w}_{\text{rdd}}$ is the discontinuity in the placebo outcome and $\hat{\gamma}_{-}$ is an appropriate weight.

Conceptually, we build on the insights and techniques of the placebo variable literature (also called the negative control literature) (Miao et al., 2018; Deaner, 2018; Tchetgen Tchetgen et al., 2024), to answer a salient question about RDD, which is an extremely popular method in social science. To the best of our knowledge, our identification result appears to be novel and motivates a new estimation problem. For this new estimation problem, we propose and analyze what appears to be a novel variation of the widely used local linear RDD estimator (Fan and Gijbels, 1992), and we derive bias-corrected inference (Calonico et al., 2014, 2019). Future work may develop bias-aware inference (Armstrong and Kolesár, 2018; Imbens and Wager, 2019; Noack and Rothe, 2024).

The rest of the paper proceeds as follows. Section 2 illustrates the main idea with a real world application. Section 3 presents our main identification assumptions and result. Section 4 derives our estimator. Section 5 proves consistency and normality with bias-corrected inference. Section 6 concludes. The main text focuses on the core insights, while formal proofs are deferred to the appendix.

2 Example: Maimonides rule

To motivate our approach, consider an example adapted from Angrist et al. (2019). Suppose we are interested in estimating the effect of transitioning from a large class (39-40 pupils) to a small class (20-21 pupils), denoted by the binary treatment A. The outcome of interest is student achievement Y. A Maimonides-style rule determines class size: schools must split classes when total enrollment in a grade level D exceeds a statutory threshold $d^* = 40$. For instance, a school with $D = 42$ students in a grade level receives funding for two classes of 21 students ($A = 1$), while a school with exactly $D = 40$ students in a grade level keeps one class of 40 students ($A = 0$).

To apply a conventional RDD estimator in this setting, one must assume that the expected potential outcomes $\mathbb{E}\{Y(1, d, U, \eta_y) \mid D = d\}$ and $\mathbb{E}\{Y(0, d, U, \eta_y) \mid D = d\}$ are continuous at the cutoff d^*. Here, $Y(A, D, U, \eta_y)$ denotes the test score of a student in a class size of 20-21 students ($A = 1$) or 39-40 students ($A = 0$), given the total enrollment in the grade level (D). We model unobserved heterogeneity through two components—school administrator sophistication U, which may influence total enrollment near the cutoff, and student-level heterogeneity η_y, which is random. The left panel of Figure 1 depicts a valid RDD scenario under this assumption.

Remark 1 (Continuity and unobserved confounding). As noted by Lee and Lemieux (2010), the standard RDD assumption, i.e. continuity of the expected potential outcomes at the cutoff, is implied by the following two conditions: for all (u, η_y), (i) the conditional distribution $d \mapsto \mathbb{P}(d \mid u, \eta_y)$ is continuous at d^* ; and (ii) the potential outcome function $d \mapsto Y(1, d, u, \eta_y)$ is bounded and continuous at d^*.

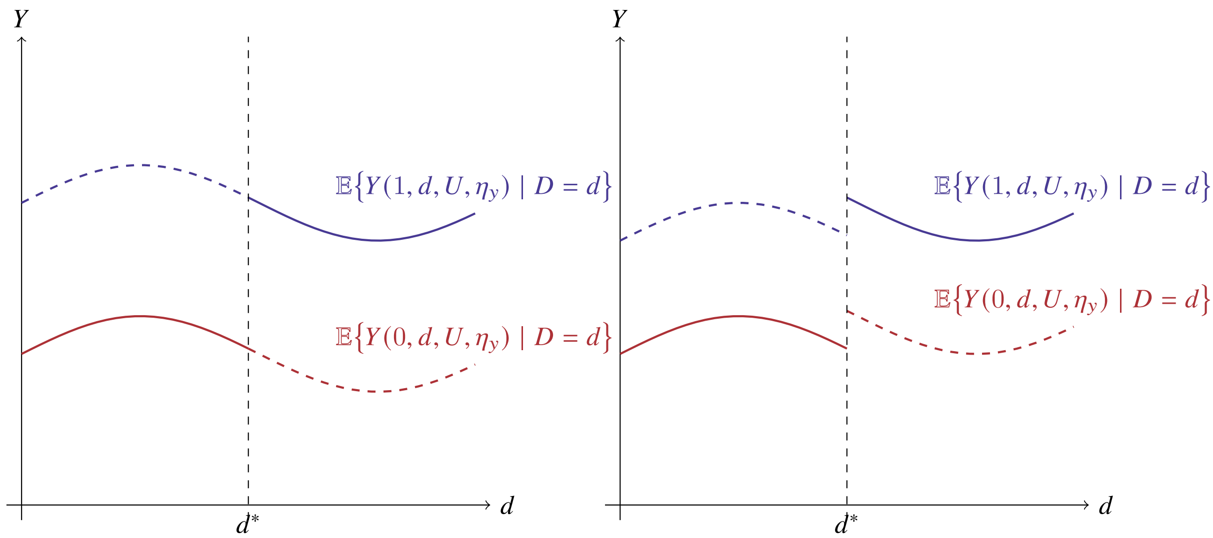
To see this, note that continuity of $\mathbb{P}(d \mid u, \eta_y)$ implies continuity of $\mathbb{P}(u, \eta_y \mid d)$ by Bayes' rule. Then, applying the dominated convergence theorem, we can exchange the limit and expectation, so that

$$
\begin{aligned}
&\lim_{\epsilon \downarrow 0} \mathbb{E}\{Y(1, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon\} = \lim_{\epsilon \downarrow 0} \int Y(1, d^* + \epsilon, u, \eta_y) \; \mathrm{d}\mathbb{P}(u, \eta_y \mid d^* + \epsilon) \\
&= \int Y(1, d^*, U, \eta_y) \; \mathrm{d}\mathbb{P}(u, \eta_y \mid d^*) = \mathbb{E}\{Y(1, d^*, U, \eta_y) \mid D = d^*\}.
\end{aligned}
$$

The same argument applies from the left.

However, this assumption may fail if school administrators behave strategically. For instance, more sophisticated school leaders may recognize the budgetary and pedagogical advantages of exceeding the class size threshold, perhaps by enrolling just above the cutoff. These administrators may also tend to run more effective schools with higher-achieving students, thereby inducing a spurious upward jump in test scores around d^* that is not due to class size per se (Angrist et al., 2019). In such a case, the unobserved aptitude of school leaders (U) confounds the relationship between enrollment (D) and student achievement (Y) near the cutoff. As a result, schools just above d^* may perform better because they are led by more capable administrators, invalidating the continuity assumption that would allow researchers to evaluate the effect of class size. This confounded scenario is depicted in the right panel of Figure 1.

Figure 1: Valid and invalid regression discontinuity design



Notes: Figure 1 illustrates the standard continuity assumption on the expected potential outcomes for a valid RDD (on the left) and an invalid RDD (on the right). The Y-axis gives the expected test score $\mathbb{E}\{Y(a, d, U, \eta_y) \mid D = d\}$ for a student in a class of 20-21 students ($A = 1$) or 39-40 students ($A = 0$) given $D = d$ enrolled students in the school. The x-axis gives the number of students enrolled in the school. The dotted regions of the line indicate the unobserved, counterfactual expected outcomes.

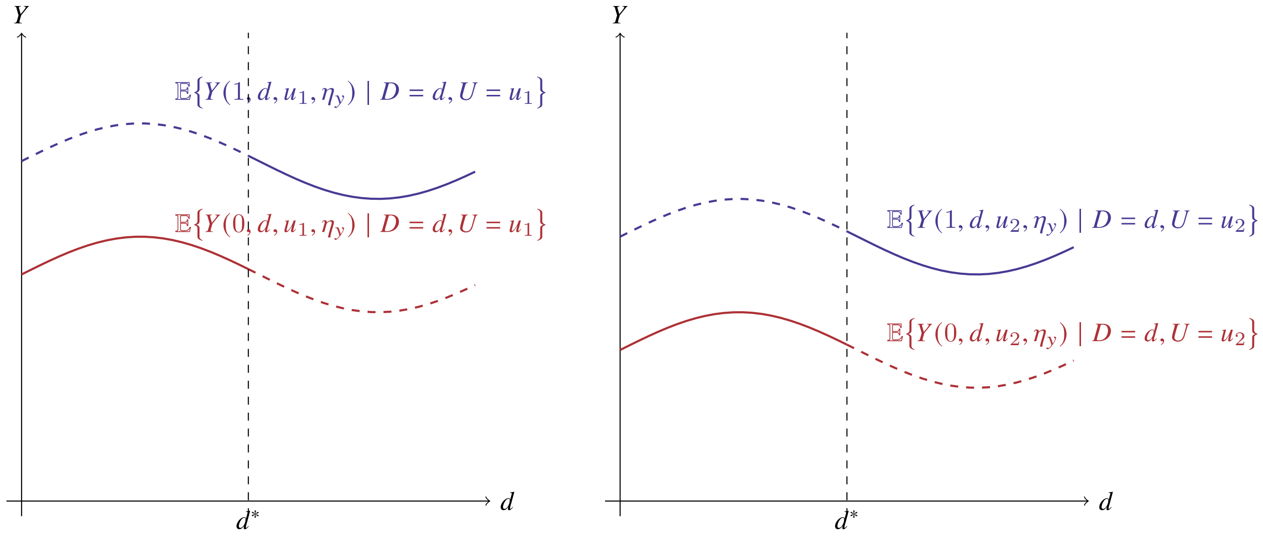
Our identification strategy recovers the treatment effect in such confounded settings by positing that continuity holds after further conditioning on the unobserved school-level aptitude U. As shown in Figure 2, we require that, conditional on both $D = d$ and $U = u$, the expected potential outcomes become continuous at the cutoff. We then exploit the statistical relationship between test scores in the prior year W, which serve as placebo outcomes, and the availability of transfer-eligible students based on birthdates Z, which act as placebo treatments, to approproately adjust for the unobserved confounder U.

Intuitively, test scores in a previous year are unaffected by enrollment (D) in the current year, making it a valid placebo outcome (W). In this sense, previous test scores (W) serve as a proxy for the school administrator aptitude (U).

Meanwhile, the availability of students able to transfer across grades should only affect student test scores (Y) through its influence on total enrollment in the grade (D), making it a valid placebo treatment

(Z). The availability of transfer students is like a local instrument, or a special type of pre-determined covariate, that can help to distill the variation due to school administrator aptitude.

Figure 2: Valid placebo discontinuity design



Notes: Figure 1 illustrates the relaxed expected potential outcomes assumption we place on the outcome. The Y-axis gives the test score $\mathbb{E}\{Y(a, d, u, \eta_y) \mid D = d, U = u\}$ for a student in a class of 20-21 students ($A = 1$) or 39-40 students ($A = 0$) enrolled in a school with $D = d$ pupils and administrator aptitude level $U = u$. The x-axis gives the number of students enrolled in the school. The dotted regions of the line indicate the unobserved, counterfactual outcome.

3 Identification: Placebo discontinuity

In this section, we formalize the model, interpret the key assumptions, and present our main identification result: how to recover the treatment effect in an RDD setting with strategic behavior around the cutoff. Standard practice uses placebo variables to detect this issue; as our first contribution, we use placebo variables to correct this issue.

3.1 Notation

Consider the nonseparable model

$$Y = Y(A, D, U, \eta_y),$$

where A is the treatment, D is the running variable, U is unobserved confounding, and η_y is unobserved heterogeneity. Notice that we decompose the unobservable into the component U for which we will have a model, and the component η_y which will be exogeneous conditional on U.

We will study the usual causal parameter in RDD,

$$\tau_0 = \mathbb{E}\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^*\},$$

which reflects the average causal effect of moving from untreated to treated at the cutoff d^*, conditional on the running variable.

In contrast with standard approaches, we allow for discontinuities in the conditional distribution of unobservables $f(u, \eta_y \mid d)$ at d^*, acknowledging that strategic behavior or sorting can lead to non-smooth densities near the cutoff. However, we require the density to remain strictly positive in a neighborhood of d^* to maintain identifiability.

We employ placebo variables (Z, W) to handle unobserved confounding. The full structural system is written as,

$$Y = Y(A, D, Z, U, \eta_y), \quad W = W(A, D, Z, U, \eta_w), \quad A = A(D, Z, U, \eta_a).$$

We summarize the invariances of the structural system in the next section. A graphical summary of the causal structure is previewed in Figure 3.

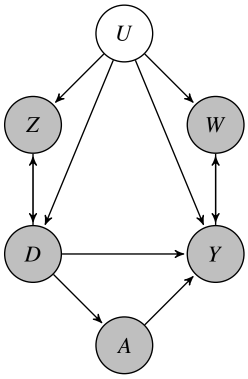


Figure 3: Placebo discontinuity design DAG

3.2 Assumptions

The following assumptions identify τ_0 even when the expected potential outcomes $d \mapsto \mathbb{E}\{Y(0, D, U, \eta_y) \mid D = d\}$ and $d \mapsto \mathbb{E}\{Y(1, D, U, \eta_y) \mid D = d\}$ are discontinuous at the cutoff $d = d^*$.

Throughout, we place assumptions in some neighborhood of the threshold d^* defined as,

$$\mathcal{D}_-(\epsilon) = (d^* - \epsilon, d^*), \quad \mathcal{D}_+(\epsilon) = (d^*, d^* + \epsilon), \quad \mathcal{D}(\epsilon) = \mathcal{D}_-(\epsilon) \cup \mathcal{D}_+(\epsilon).$$

Assumption 1 (RDD)**.** *Assume that the following limits exist and are unequal.*

$$\lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* - \epsilon\}], \quad \lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon\}]$$

The existence of this limit, for the expectation of the treatment assignment A given the running variable D, is standard in the RDD identification literature (Hahn et al., 2001; Lee and Lemieux, 2010).

Definition 1 (Sharp design). Under a "sharp" design, we assume that there is no heterogeneity in treatment assignment given the running variable (η_a does not exist) and for all $\epsilon \in \mathcal{D}(\epsilon)$,

$$\mathbb{E}\{A \mid D = d^* - \epsilon\} = 0, \quad \mathbb{E}\{A \mid D = d^* + \epsilon\} = 1.$$

A "fuzzy" design is one that is not sharp, i.e. one that does not satisfy the equalities above.

In a sharp design, A is fully determined by D, implying no heterogeneity in treatment assignment (i.e., no η_a). In this case, Assumption 1 is trivially satisfied.

We define $\mathcal{Z}(\epsilon)$ such that $\mathbb{P}(Z \in \mathcal{Z}(\epsilon) \mid D \in \mathcal{D}(\epsilon)) = 1$.

Assumption 2 (Placebo variables). *Assume there exists some $\epsilon > 0$ such that for all $d \in \mathcal{D}(\epsilon)$ and for all $z \in \mathcal{Z}(\epsilon)$,*

a *(Causal consistency): if $A = a$, $D = d$, and $Z = z$ then $Y = Y(a, d, z, U, \eta_y)$, and $W = W(a, d, z, U, \eta_w)$ almost surely. If $D = d$ and $Z = z$ then $A = A(d, z, U, \eta_a)$ almost surely.*

b *(Placebo selection on unobservables): $Z \perp\!\!\!\perp \eta_y, \eta_a \mid U, D$ and $\eta_w \perp\!\!\!\perp D, Z \mid U$.*

c *(Overlap): if $f(u) > 0$ then $f(a, d, z \mid u) > 0$, where we assume the densities exist.*

d *(Exclusion): $Y(a, d, z, U, \eta_y) = Y(a, d, U, \eta_y)$, $W(a, d, z, U, \eta_w) = W(U, \eta_w)$, and $A(d, z, U, \eta_a) = A(d, \eta_a)$ almost surely.*

In summary, we have the following simplification local to the cutoff:

$$Y = Y(A, D, U, \eta_y), \quad W = W(U, \eta_w), \quad A = A(D, \eta_a).$$

Causal consistency rules out network effects near the cut-off.

The main condition within placebo selection on observables is the implication that $Z \perp\!\!\!\perp \eta_y, \eta_w | D, U$: conditional upon the running variable and unobserved confounder, the placebo treatment is as good as random for the target outcome and placebo outcome.[2] In the sharp design, $Z \perp\!\!\!\perp \eta_a | D, U$ automatically holds; in the fuzzy design, it imposes that the placebo treatment is conditionally independent of the idiosyncratic aspect of treatment assignment. As shown in Appendix B, $Z \perp\!\!\!\perp \eta_y, \eta_a, \eta_w | D, U$ together with an additional continuity condition suffice for our argument to work. In the main text, to simplify the exposition, we strengthen $\eta_w \perp\!\!\!\perp Z \mid D, U$ to $\eta_w \perp\!\!\!\perp D, Z \mid U$ and thereby avoid the additional continuity.

Overlap ensures there is no stratum of unobserved confounding for which the treatment, running variable, or placebo treatment are deterministic. In the language of the RDD literature, we rule out complete control but allow imprecise control (Lee and Lemieux, 2010).

Exclusion encodes the idea that the placebo treatment does not directly cause the outcome of interest; instead, it affects the outcome via the running variable. Exclusion also encodes the idea that the placebo outcome is not caused by the treatment, running variable, or placebo treatment. Finally, exclusion imposes that the treatment is pinned down by only the running variable (and possibly idiosyncratic noise in the fuzzy setting).

In the running example, suppose class size A is determined by grade enrollment size D, and we study its effect on student achievement Y in year t. School administrator aptitude U is unobserved, and strategic behavior around the cutoff can cause a jump in $f(d \mid u, \eta_y)$ (and hence $f(u, \eta_y \mid d)$), since strategic administrators who are aware of the class-enrollment cutoff d^* may manipulate enrollment so that they exceed the threshold d^* to open an additional class. Let Z be the availability of students to transfer between grade levels (affecting D but not Y directly), and W be achievement in year $t - 1$ (unaffected by D or Z). Overlap implies school officials cannot perfectly control enrollment levels D. However, there is a jump in $d \mapsto f(d \mid u, \eta_y)$ that we wish to allow for and to deal with.

Assumption 3 (Conditional independence). *There exists some $\epsilon > 0$ such that for all $d \in \mathcal{D}(\epsilon)$*

a *$\eta_a \perp\!\!\!\perp \eta_y \mid D = d$ or $\eta_a \perp\!\!\!\perp \eta_y \mid D = d, U$;*

b *$\eta_a \perp\!\!\!\perp U \mid D = d$.*

[2] By weak union, $\eta_w \perp\!\!\!\perp D, Z \mid U$ implies $\eta_w \perp\!\!\!\perp Z \mid D, U$.

Assumption 3 is vacuously satisfied by the sharp design.

In the fuzzy design, it is analogous to Hahn et al. (2001, Condition C3): it implies that conditional on D, the randomness in the treatment assignment mechanism is unrelated to the idiosyncratic aspect of potential outcomes and the unobserved confounding. In other words, given an enrollment level D, whether a school has one or two classes does not depend on the heterogeneity in students' test scores or the administrator's underlying ability. This assumption in the fuzzy design may be relaxed by imposing homogeneity of effects; see Appendix C.

Assumptions 2 and 3 can be summarized by the results below:

Lemma 1 (Necessary independences). *Assumption 2 implies that for all $d \in \mathcal{D}(\epsilon)$ and $z \in \mathcal{Z}(\epsilon)$,*

a $Y \perp\!\!\!\perp Z \mid D, U$

b $W \perp\!\!\!\perp D, Z \mid U$

Additionally under Assumption 3b it follows that,

c $A \perp\!\!\!\perp U \mid D$

Assumption 4 (Continuity). *For all $d \in D(\epsilon)$, assume that the mappings $d \mapsto \mathbb{E}\big\{Y(1, D, U, \eta_y) \mid D = d, U = u\big\}$ and $d \mapsto \mathbb{E}\big\{Y(0, D, U, \eta_y) \mid D = d, U = u\big\}$ are continuous for all u.*

Assumption 4 weakens the standard continuity assumption of the RDD literature: continuity of potential outcomes conditional upon the running variable $D = d$ near the cutoff d^*. Instead, it requires continuity of potential outcomes conditional upon both $D = d$ and $U = u$.

While the density of student enrollment may jump at d^*, we anticipate that potential student achievement is continuous when fixing the administrator's aptitude $U = u$.

Lemma 2 (Limit). *Under a fuzzy design and Assumptions 1, 2, 3, and 4,*

$$\tau_0 = \frac{\mathbb{E}\big\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \mid D = d^*\big\}}{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon\} - \mathbb{E}\{A \mid D = d^* - \epsilon\}]}.$$

In a sharp design, and under Assumptions 1, 2, and 4:

$$\tau_0 = \mathbb{E}\left\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \mid D = d^*\right\}.$$

Lemma 2 solves for the treatment effect τ_0 in an interpretable way: for each value of the unobserved confounded $U = u$, conduct RDD conditional upon $U = u$, then average over U. In our running example, this is equivalent to averaging the treatment effect of a class size (A) on test scores in year-t (Y) for schools with enrollment levels at d^* over different administrator-ability levels U.

Clearly, this calculation is infeasible, since U is unobserved. We need additional assumptions to identify τ_0 in terms of observed variables.

3.3 Main identification result

Section 3.2 established that the causal estimand τ_0 can be written in terms of conditional expectations over unobserved confounding (U). The key challenge now is to express τ_0 in terms of observable variables. To do this, we adapt the "confounding bridge" approach (Tchetgen Tchetgen et al., 2024), which uses auxiliary variables to solve an integral equation that recovers the influence of U on Y.

Assumption 5 (Confounding bridge)**.** *For all $d \in \mathcal{D}_+(\epsilon)$, there exists a bounded solution to the following operator equation almost surely,*

$$\mathbb{E}\{Y \mid D = d, U\} = \int h_+(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D = d, U).$$

Similarly, for all $d \in \mathcal{D}_-(\epsilon)$ we have,

$$\mathbb{E}\{Y \mid D = d, U\} = \int h_-(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D = d, U).$$

The confounding bridge assumption states that, for fixed values of the running variable (D) and unobserved confounder (U), the conditional expectation of the outcome (Y) can be expressed as a function of a proxy variable (W) that is associated with (U). The functions h_+ and h_- act as weighting kernels that summarize how W captures the influence of U on Y on either side of the threshold.

In the running example, for schools with the same number of enrolled students ($D = d$), there exists a function $h_+(d - d^*, \cdot)$ such that we can recover expected student achievement using a reweighting of the distribution of test scores in year $t - 1$ (W). The function h_+ does not depend on U directly, but its integral with respect to the distribution of W conditional on U reconstructs the confounded conditional expectation of Y. This formulation allows us to bypass the unobserved U by using an observed variable W that is a proxy for U.

First we show that h_0 is a solution to an integral equation in terms of observed variables, though it may not be unique.

Lemma 3 (Factuals I)**.** *Lemma 1 and Assumption 5 imply that any h_0 that exists also solves, for all $d \in \mathcal{D}(\epsilon)$ and all $z \in \mathcal{Z}(\epsilon)$,*

$$\mathbb{E}\{Y \mid D = d, Z = z\} = \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D = d, Z = z)$$

Lemma 3 translates the bridge equation from a conditional expectation over U to one over observed variables Z and W. It shows that the same h_0 function also solves a second integral equation where both sides of the equation are estimable from the data.

Assumption 6 (Completeness)**.** *For all $d \in \mathcal{D}(\epsilon)$ and all $z \in \mathcal{Z}(\epsilon)$, $\mathbb{E}\{f(U) \mid D = d, Z = z\} = 0$ if and only if $f(U) = 0$ almost surely.*

Assumption 6 is necessary to guarantee a unique solution to the equation in Lemma 3 (factuals). Without this assumption, there could be multiple functions h that satisfy the same moment condition, making point identification of τ_0 more challenging.

Future work may build on Bennett et al. (2022) to give weaker conditions under which τ_0 can still be identified without assuming completeness.

Lemma 4 (Factuals II)**.** *Lemma 1 and Assumption 6 imply that if there exists a function h_0 that, for all $d \in \mathcal{D}(\epsilon)$ and all $z \in \mathcal{Z}(\epsilon)$ satisfies,*

$$\mathbb{E}\{Y \mid D = d, Z = z\} = \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D = d, Z = z),$$

then Assumption 5 holds.

Lemma 4 completes the identification argument by showing that the existence of a function h_0 that solves the observed-data version of the bridge equation implies the existence of a confounding bridge in the structural model. This allows us to write τ_0 in terms of functions of observed variables.

Theorem 1 (Placebo identification)**.** *Suppose Assumptions 1, 2, 3, 4, and 5 hold. Then*

$$\tau_0 = \frac{\lim_{\epsilon\downarrow 0}\int h_+(\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*) - \lim_{\epsilon\downarrow 0}\int h_-(-\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*)}{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon\} - \mathbb{E}\{A \mid D = d^* - \epsilon\}]}$$

If in addition Assumption 6 holds then the integrals of h_0 in the numerator are identified. Therefore τ_0 is identified.

In sharp design, under Assumptions 1, 2, 4, and 5, then this simplifies to:

$$\tau_0 = \lim_{\epsilon\downarrow 0}\int h_+(\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*) - \lim_{\epsilon\downarrow 0}\int h_-(-\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*)$$

Theorem 1 provides a complete identification strategy—it shows how the discontinuity in potential outcomes conditional on unobserved U can be corrected using an observable proxy (W) and an auxiliary variable (Z). The functions h_+ and h_- effectively "invert" the confounding effect of U using the relationship between U and W. This inversion is unique when completeness holds and yields a point-identified causal estimand.

In the running example, we estimate the effect of class size on test scores in year-t by using test scores in year $t-1$ (W) as a proxy for unobserved administrator ability (U), and availability of students to distort enrollment counts (Z) as a local instrument. If the bridge functions h_+ and h_- can be identified from observed data, then we can reconstruct the counterfactual outcomes at the cutoff and identify τ_0.

4 Estimation: Local instrumental variable regression

Theorem 1 identifies the RDD treatment effect by reweighting two confounding bridges $h_-(d, w)$ and $h_+(d, w)$. Recent literature to approximate these integral equations rely on kernel methods to non-parametrically estimate these confounding bridges. However, Fan and Gijbels (1992) show that the order of bias of these kernel estimators on boundary points (d^*) is high, and advocate the use of local linear regressions to estimate points on the boundary with less bias.

As our second contribution, we propose a local linear instrumental variable estimator, extending Fan and Gijbels (1992) to accommodate instruments. For tractability, we introduce a partially linear approximation to the confounding bridge in Assumption 5. The final estimator decomposes into two intepretable terms; it adjusts the discontinuity in the target outcome by the discontinuity in the placebo outcome.

For clarity, we focus on the sharp design, i.e. the numerator in Theorem 1. The fuzzy design is a straightforward extension, since the denominator in Theorem 1 is a standard RDD estimand.

For simplicity, we also focus on the exactly identified case where $\dim(z) = \dim(w)$. Our results naturally extend to the overidentified case $\dim(z) \geq \dim(w)$ by standard techniques.

4.1 Locally approximating the confounding bridge

In line with the identification strategy presented in Section 3, we estimate the treatment effect τ_0 using a two-step procedure. First, we solve for the confounding bridges $h_+(\cdot,\cdot)$ and $h_-(\cdot,\cdot)$ at the cutoff d^* using the moment condition from Lemma 3. Then, invoking Theorem 1, we compute the difference in their limits, $\tau_0 = \lim_{\epsilon\downarrow 0}\left[\int h_+(\epsilon, w)\, d\mathbb{P}(w \mid D = d^*) - \int h_-(-\epsilon, w)\, d\mathbb{P}(w \mid D = d^*)\right]$.

To motivate our estimator, consider the class of partially linear functions,

$$h_+(d-d^*,w) = g_+(d-d^*) + w^\top \gamma_+, \qquad h_-(d-d^*,w) = g_-(d-d^*) + w^\top \gamma_-, \tag{1}$$

where $g_+ \colon \mathcal{D}_+(\epsilon) \to \mathbb{R}$ and $g_- \colon \mathcal{D}_-(\epsilon) \to \mathbb{R}$ are any continuous functions, and γ_+ plays the role of the best linear projection of $h_+(\cdot,\cdot)$ onto W at $D = d^*$. Proposition 2 below gives a detailed characterization. Within this class, the moment condition in Lemma 3 becomes,

$$\mathbb{E}\left\{Y - g_+(D-d^*) - W^\top \gamma_+ \mid D = d, Z = z\right\} = 0. \tag{2}$$

This expression closely resembles a partially linear instrumental variable moment condition, where W plays the role of the endogenous regressor and Z serves as the instrument. Therefore, we propose a local instrumental variable regression approach to estimate the best approximation to $h_+(\cdot,\cdot)$ near d^*.

4.2 Algorithm

To define our estimator, we introduce some notation. Let $H_1 = \mathrm{Diag}(1, h_n)$ be a transformation matrix of the kernel bandwidth. Let $\alpha_+ = (\alpha_{+,0}, \alpha_{+,1})^\top = \left(g_+(0), g_+'(0)\right)^\top$ concatenate the nonlinear component of h_+ and its derivative at the cutoff. Finally, we concatenate all of the parameters to be estimated as $\nu_+^\top = (\alpha_{+,0}, \alpha_{+,1}, \gamma_+^\top)$.

We propose what appears to be a new local linear objective function, which we call local instrumental variable regression. Its first order condition, which yields the estimator $\hat{\nu}_+$, is

$$\frac{1}{n}\sum_{i=1}^n \omega_{i,+} \begin{bmatrix} R_{i,1} \\ Z_i \end{bmatrix} \left(Y_i - R_{i,1}^\top H_1 \hat{\alpha}_+ - W_i^\top \hat{\gamma}_+\right) = \mathbf{0},$$

where $\omega_{i,+} = \frac{1}{h_n}\mathbb{1}(D_i \geq d^*) K\left(\frac{|D_i - d^*|}{h_n}\right)$ are local kernel weights, $R_{i,1} = \left(1, \frac{D_i - d^*}{h_n}\right)^\top$ is a transformed vector of the running variable, and $Z_i \in \mathbb{R}^{\dim(w)}$ is a vector of placebo treatments. This empirical moment is clearly a sample analogue of the population moment implied by equation (2). The left limit objects are analogous.

We then use the coefficients estimated by local instrumental variable regression to construct our estimator of the treatment effect. The appropriate formula is immediate from Theorem 1 and equation (1). Formally, the approximate treatment effect and its estimator are

$$\begin{aligned} \tau_{\text{pdd}} &= g_+(0) + \lim_{\epsilon \downarrow 0} \mathbb{E}\left\{W^\top \mid D = d^* + \epsilon\right\}\gamma_+ - \left(g_-(0) + \lim_{\epsilon \downarrow 0} \mathbb{E}\left\{W^\top \mid D = d^* - \epsilon\right\}\gamma_-\right) \\ &= \alpha_{+,0} + (\beta_{+,0}^w)^\top \gamma_+ - \left\{\alpha_{-,0} + (\beta_{-,0}^w)^\top \gamma_-\right\} \\ \hat{\tau}_{\text{pdd}} &= \hat{\alpha}_{+,0} + (\hat{\beta}_{+,0}^w)^\top \hat{\gamma}_+ - \left\{\hat{\alpha}_{-,0} + (\hat{\beta}_{-,0}^w)^\top \hat{\gamma}_-\right\}, \end{aligned} \tag{3}$$

where $\hat{\beta}_{+,0}^w$ estimates the right-limit $\beta_{+,0}^w = \lim_{\epsilon \downarrow 0} \mathbb{E}\{W_i \mid D = d^* + \epsilon\}$ using the local linear estimator of Hahn et al. (2001):

$$H_1 \hat{\beta}_+^{w_j} = \arg\min_{\beta} \sum_{i=1}^n \omega_{i,+} (W_{i,j} - R_{i,1}^\top \beta)^2.$$

Then, we let $\hat{\beta}_{+,0}^w = \left(\hat{\beta}_{+,0}^{w_1}, \ldots, \hat{\beta}_{+,0}^{w_{\dim(w)}}\right)^\top \in \mathbb{R}^{\dim(w)}$ concatenate the first components of these estimated coefficients. The left limit objects are analogous.

Section 5 below shows that this estimator is consistent for τ_{pdd}. After bias correction, which we defer to Section 5, it is also asymptotically normal at the familiar rate of $n^{-2/5}$.

4.3 Equivalence: RDD with a new adjustment term

For any finite sample size, our proposed estimator is numerically equivalent to a highly interpretable procedure: take the standard RDD estimator based on discontinuity in the outcome (Y), and add an adjustment term based on the discontinuity in the placebo outcome (W).

Proposition 1 (Estimator equivalence). *Our proposed estimator in equation (3) is numerically equivalent to*

$$\hat{\tau}_{pdd} = \hat{\tau}^{y}_{rdd} + \left(\hat{\tau}^{w}_{rdd}\right)^{\top}\hat{\gamma}_{-},$$

where $\hat{\tau}^{y}_{rdd} = \hat{\beta}^{y}_{+,0} - \hat{\beta}^{y}_{-,0} \in \mathbb{R}$ and $\hat{\tau}^{w}_{rdd} = \left(\hat{\beta}^{w}_{+,0} - \hat{\beta}^{w}_{-,0}\right) \in \mathbb{R}^{\dim(w)}$ are constructed from the familiar RDD local linear objective functions, namely

$$H_1\hat{\beta}^{y}_{+} = \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \omega_{i,+}\left(Y_i - R_{i,1}^{\top}\beta\right)^2, \qquad H_1\hat{\beta}^{w_j}_{+} = \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \omega_{i,+}\left(W_{i,j} - R_{i,1}^{\top}\beta\right)^2,$$

and analogous objectives below the cutoff. Moreover, $\hat{\gamma}_{-}$ is the local instrumental variable regression of the residualized $Y^{\perp}$ on the residualized $W^{\perp}$, instrumenting with Z, below cutoff:

$$\hat{\gamma}_{-} = \left[\frac{1}{n}\sum_{i=1}^{n} \omega_{i,-}\{Z_i(W_i^{\perp})^{\top}\}\right]^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} \omega_{i,-}(Z_i Y_i^{\perp})\right\}.$$

Here, $(Y_i^{\perp})$ are the residuals from a local linear regression of (Y_i) onto (D_i), and $(W_i^{\perp})$ are the residuals from a local linear regression of (W_i) onto (D_i). In other words, we residualize the target outcome and the placebo outcome using the running variable.

The RDD literature, summarized recently by e.g. Cattaneo et al. (2024), advocates for testing whether $\hat{\tau}^{w}_{\mathrm{rdd}}$ is close to zero as a diagnostic for whether $f(d \mid u, \eta_y)$ (or, relatedly, $f(u, \eta_y \mid d)$) changes discontinuously at d^*. Our contribution is to show that this discontinuity is an essential ingredient for an adjustment term. If $\hat{\tau}^{w}_{\mathrm{rdd}}$ is indeed negligible, our adjustment term is zero and our estimator $\hat{\tau}_{\mathrm{pdd}}$ simplifies to the standard RDD estimate $\hat{\tau}^{y}_{\mathrm{rdd}}$.

Within our adjustment term, the placebo outcome discontinuity ($\hat{\tau}^{w}_{\mathrm{rdd}}$) is multiplied by weights ($\hat{\gamma}_{-}$).

These weights are increasing in the correlation between the residualized outcome ($Y^{\perp}$) and the placebo treatment (Z). Intuitively, if this correlation is stronger, then there is more unobserved confounding on the target outcome, and hence we increase the weights.

The weights are decreasing in the correlation between the residualized placebo outcome ($W^{\perp}$) and the placebo treatment (Z). Intuitively, if this correlation is weaker, then the placebo outcome is a weaker proxy for the unobserved confounding, and hence we compensate by increasing the weights on its estimated discontinuity.

Proposition 1 clarifies the danger of a very weak proxy, which would lead to numerical instability in this final aspect of the weights. Future work may study the behavior of our estimator when using a weak proxy.

Remark 2 (Only left-adjustment). The decomposition in Proposition 1 only estimates γ_{-}, and not γ_{+}, because our treatment effect of interest is the effect at the cutoff. As such, it is the effect on the treated, so only the untreated potential outcome needs adjustment.

In more detail, our estimator is anchored on the distribution of U just above the cutoff. It preserves the right-limit $\mathbb{E}[Y \mid D = d^* + \epsilon]$. However, it adjusts the left-limit using γ_-. The adjusted term allows us to estimate the counterfactual below the cutoff $\mathbb{E}[\mathbb{E}[Y \mid D = d^* - \epsilon, U] \mid D = d^*]$, in the thought experiment where U matches the distribution it would have followed above the cutoff.

An alternative causal parameter is $\tilde{\tau}_0 = \mathbb{E}[Y(1, d^*, U, \eta_y) - Y(0, d^*, U, \eta_y)]$, for which our estimator would have two adjustment terms: $\tilde{\tau}_{\text{pdd}} = \hat{\tau}^y_{\text{rdd}} + (\mathbb{E}\{W\} - \hat{\beta}^w_{+,0})^\top \hat{\gamma}_+ - (\mathbb{E}\{W\} - \hat{\beta}^w_{-,0})^\top \hat{\gamma}_-$.

5 Bias corrected inference

As our third contribution, we study the large sample properties of the estimator in Proposition 1. We prove it is consistent. Then, we prove that it is asymptotically normal at the familiar rate $n^{-2/5}$ after bias correction.

5.1 Consistency

While prior work used placebo outcomes and predetermined covariates to detect assumption violations, we propose a method to directly correct for them and thereby recover τ_0. Below we state sufficient conditions for consistency.

Let $\mu_{+,y}(d - d^*) = \mathbb{E}\{Y \mid D = d\}$ for $d \geq d^*$ and $\mu_{-,y}(d - d^*) = \mathbb{E}\{Y \mid D = d\}$ for $d < d^*$. Similarly, define $\mu_{+,w_j}(d - d^*)$, $\mu_{-,w_j}(d - d^*)$, $\mu_{+,z_j}(d - d^*)$, and $\mu_{-,z_j}(d - d^*)$..

Let $\sigma^2_{+,y}(d - d^*) = \text{Var}(Y \mid D = d)$ for $d \geq d^*$ and $\sigma^2_{-,y}(d - d^*) = \text{Var}(Y \mid D = d)$ for $d < d^*$. Similarly, define $\sigma^2_{+,w_j}(d - d^*)$ and $\sigma^2_{-,w_j}(d - d^*)$.

Assumption 7 (Regularity conditions)**.** *Let the following conditions hold.*

a *(Continuous mean) For all $d \in \mathcal{D}_+(\epsilon)$ and $j \in \{1, \dots, \dim(w)\}$, $\mu_{+,y}(\cdot)$ and $\mu_{+,w_j}(\cdot)$ are 3-times continuously differentiable and bounded. Additionally, $\mu_{+,z_j}(\cdot)$, $\mathbb{E}\{Z_i W_i^\top \mid D = d\}$, and $\mathbb{E}\{Z_i Y_i \mid D = d\}$ are continuous and bounded.*

   For all $d \in \mathcal{D}_-(\epsilon)$ and $j \in \{1, \dots, \dim(w)\}$, $\mu_{-,y}(\cdot)$ and $\mu_{-,w_j}(\cdot)$ are 3-times continuously differentiable and bounded. Additionally, $\mu_{-,z_j}(\cdot)$, $\mathbb{E}\{Z_i W_i^\top \mid D = d\}$, and $\mathbb{E}\{Z_i Y_i \mid D = d\}$ are continuously differentiable and bounded.

b *(Continuous variance) For all $d \in \mathcal{D}_+(\epsilon)$ and $j \in \{1, \dots, \dim(w)\}$, $\sigma^2_{+,y}(\cdot)$ and $\sigma^2_{+,w_j}(\cdot)$ are continuous and bounded away from 0. For all $d \in \mathcal{D}_-(\epsilon)$, $\sigma^2_{-,y}(\cdot)$ and $\sigma^2_{-,w_j}(\cdot)$ are continuous and bounded away from 0.*

c *(Bounded higher moments) For all $d \in \mathcal{D}(\epsilon)$ and $j \in \{1, \dots, \dim(w)\}$, there exists some $\zeta > 0$ and $M \in \mathbb{R}$ such that we have $\mathbb{E}\left\{\left|Y_i - \mu_{+,y}(d - d^*)\right|^{2+\zeta} \mid D_i = d\right\}$, $\mathbb{E}\left\{\left|Y_i - \mu_{-,y}(d - d^*)\right|^{2+\zeta} \mid D_i = d\right\}$, $\mathbb{E}\left\{\left|W_{i,j} - \mu_{+,w_j}(d - d^*)\right|^{2+\zeta} \mid D_i = d\right\}$, $\mathbb{E}\left\{\left|W_{i,j} - \mu_{-,w_j}(d - d^*)\right|^{2+\zeta} \mid D_i = d\right\} < M$.*

The regularity conditions in Assumption 7 are variations of standard assumptions in the RDD literature and are implied by those in Calonico et al. (2019).

Assumption 8 (Kernel conditions)**.** *Assume that the kernel function $K\colon \mathbb{R}^+ \mapsto \mathbb{R}^+$ is symmetric, continuous, and satisfies $\int_0^\infty u^5 K(u)^q \, \mathrm{d}u < \infty$ for $q \in [1, 3]$.*

Assumption 8 is satisfied by the window kernel $K(u) = \mathbb{1}(u \leq 1)$ and the triangle kernel $K(u) = (1 - u)\mathbb{1}(u \leq 1)$, which are the most frequently used kernels in the RDD literature. It also allows for the

Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{u^2}{2\lambda}\right\}$. Under these conditions we can characterize the probability limit of our local instrumental variable estimator for any vanishing bandwidth $1 \gg h_n \gg n^{-1}$.

Proposition 2 (Probability limit)**.** *Let Assumptions 7a and 8 hold, and assume that $nh_n \to \infty$ and $h_n \to 0$. Then, it follows that,*

$$\hat{\tau}_{pdd} \xrightarrow{p} \tau^y_{rdd} - (\tau^w_{rdd})^\top \gamma_-$$

where $\gamma_- = \lim_{\epsilon \downarrow 0} \operatorname{Cov}(Z_i, W_i^\top \mid D = d^* - \epsilon)^{-1} \operatorname{Cov}(Z_i, Y_i \mid D_i = d^* - \epsilon)$ is a best local linear approximation, i.e. it solves the local projection problem

$$Y = c + W^\top \gamma_- + \eta, \qquad \lim_{\epsilon \downarrow 0} \mathbb{E}\{\eta Z_i \mid D_i = d^* - \epsilon\} = 0, \qquad \lim_{\epsilon \downarrow 0} \mathbb{E}\{\eta \mid D_i = d^* - \epsilon\} = 0.$$

Building on Proposition 1 (estimator equivalence), which characterizes the finite sample estimate, Proposition 2 characterizes the asymptotic limit of the local estimator. Now, at the population level, we show that if the placebo outcome does not jump at the cutoff ($\tau^w_{\text{rdd}} = 0$), then the local instrumental variable estimand reduces to the traditional RDD estimand.

The population limit of the weights admits a similar interpretation as before. It also sheds light on the nature of our approximation. In particular, Proposition 2 shows that γ_- can be interpreted as a linear projection of Y onto the distribution of W, instrumenting with Z, when D approaches the cutoff from the left. In other words, γ_- is the parameter which best approximates the confounding bridge in Lemma 3 (factuals) as a partially linear function of the placebo outcome W.

If the confounding bridge $h_-(0, w)$ is very nonlinear in W then this approximation can be poor. However, we advocate for this approach because full nonparametric estimation over (D_i, W_i) would require us to uniformly estimate $h_-(0, w)$ as a function of w to characterize the asymptotic bias of the local instrumental variable estimator. There would be $\dim(w)$ additional kernels and bandwidths in the estimator. As discussed by Calonico et al. (2019), localizing weights for many variables would suffer from a curse of dimensionality, and hence empirical applications would become challenging. Our estimator does not suffer from this curse of dimensionality, and is therefore practical for empirical researchers. Future work may attempt to overcome these challenges, perhaps by placing additional structure on the effective dimension or sparsity of W (Singh, 2020; Noack et al., 2021).

Corollary 1 (Consistency)**.** *Let Assumptions 1, 2, 3, 4, 5, 7a, and 8 hold. Additionally, suppose that equation (1) holds almost surely. Then, if $nh_n \to \infty$ and $h_n \to 0$ it follows that $\hat{\tau}_{pdd} \xrightarrow{p} \tau_0$.*

Corollary 1 shows that when the confounding bridge lies in the class of partially linear functions in equation (1), then the local instrumental variable estimator is consistent for the treatment effect τ_0.

5.2 Bias correction and asymptotic normality

As shown in Proposition 1, our estimator can be decomposed into a sum of two components: (i) the RDD estimate of the outcome discontinuity $\hat{\tau}^y_{\text{rdd}}$, and (ii) the RDD estimate of the placebo outcome discontinuity, appropriately weighted as $(\hat{\tau}^w_{\text{rdd}})^\top \hat{\gamma}_-$. A key insight from Calonico et al. (2014) is that, when the bandwidth h_n is selected to optimize mean squared error (MSE), these discontinuity estimators suffer from non-negligible bias in their sampling distributions. This bias persists in large samples and leads to invalid inference, particularly when constructing confidence intervals.

To address this issue, we extend the robust bias correction technique developed by Calonico et al. (2014). In particular, we construct bias-corrected estimators that remain consistent and asymptotically normal, even when using MSE-optimal bandwidths. Specifically, we define the bias-corrected version of the local instrumental variable estimator as,

$$\hat{\tau}_{\text{pdd}}^{\text{bc}} = \hat{\tau}_{\text{rdd}}^{y,\text{bc}} - \left(\hat{\tau}_{\text{rdd}}^{w,\text{bc}}\right)^{\top} \hat{\gamma}_{-},$$

where $\hat{\tau}_{\text{rdd}}^{y,\text{bc}}$ and $\hat{\tau}_{\text{rdd}}^{w,\text{bc}}$ are bias-corrected estimators for the discontinuities of Y and W, respectively. In particular, we explicitly estimate and subtract the first-order bias terms associated with the local linear estimates $\hat{\tau}_{\text{rdd}}^{y}$ and $\hat{\tau}_{\text{rdd}}^{w}$. The explicit forms of these bias corrections are provided in Appendix F.

Theorem 2 (Limit distribution)**.** *Let Assumptions 7 and 8 hold. Additionally, assume that the bandwidths satisfy $nh_n \to \infty$, $nh_n^7 \to 0$, and $\lim_{n\to\infty}(h_n/b_n) < \infty$, where b_n is the bandwidth of the bias correction. Then, for $\tau_{pdd} = \tau_{rdd}^{y} - (\tau_{rdd}^{w})^{\top}\gamma_{-}$, it follows that,*

$$\sqrt{nh_n}\hat{V}_{bc}^{-1/2}\left(\hat{\tau}_{pdd}^{bc} - \tau_{pdd}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where $\hat{V}_{bc}^{-1/2}$ is given in Appendix F.

Theorem 2 establishes the asymptotic normality of the bias-corrected local instrumental variable estimator under the commonly used MSE-optimal bandwidth rate of $h_n \asymp n^{-1/5}$ (Imbens and Lemieux (2008)). The bias correction bandwidth may be of the same order, i.e. $b_n \asymp h_n$.

More generally, Theorem 2 tolerates any vanishing bandwidths in a wide range, i.e. any $n^{-1/7} \gg h_n \gg n^{-1}$ and any $b_n \gtrsim h_n$.

In practice, we recommend a data-driven bandwidth choice $\hat{h}_n$ along the lines of e.g. Imbens and Kalyanaraman (2012); Calonico et al. (2018, 2020). Our analysis extends accordingly. This choice is easily implemented with existing software for RDD.

Theorem 2 justifies the use of standard Wald-type inference for $\hat{\tau}_{\text{pdd}}^{\text{bc}}$, e.g. the construction of confidence intervals and hypothesis tests.

6 Discussion

This paper develops a new identification and estimation framework for regression discontinuity designs when the standard continuity of potential outcomes assumption fails due to unobserved confounding. We show that by leveraging a placebo treatment and a placebo outcome, it is possible to recover the causal parameter even when the running variable's distribution exhibits discontinuities at the threshold. Our identification argument relies on conditional continuity given an unobserved confounder, and it employs the concept of a confounding bridge to encode the relationship between the unobserved confounder and its observed proxy. Under a completeness condition, we show that the integrated bridge function is identified from observed data, and the treatment effect at the cutoff can be recovered using integrals of observable variables.

To operationalize this strategy, we propose a local instrumental variable estimator that approximates the confounding bridge using a partially linear specification. We demonstrate that the estimator decomposes into a standard RDD term and an adjustment term driven by the discontinuity in the placebo outcome. This decomposition yields an interpretable adjustment that allows researchers to correct for violations of the standard RDD assumptions, rather than discarding such designs altogether. We establish conditions for consistency and bias-corrected inference, allowing optimal bandwidths. Our results extend the applicability of RDD methods to settings with strategic behavior, providing researchers with a tractable and theoretically grounded approach for handling unobserved confounding near the cutoff.

References

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., Pathak, P. A., and Zarate, R. A. (2017). Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago's Exam Schools. *American Economic Review*, 107(5):240–245.

Angrist, J. D., Lavy, V., Leder-Luis, J., and Shany, A. (2019). Maimonides' Rule Redux. *American Economic Review: Insights*, 1(3):309–324.

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

Bennett, A., Kallus, N., Mao, X., Newey, W., Syrgkanis, V., and Uehara, M. (2022). Inference on strongly identified functionals of weakly identified functions. *arXiv:2208.08291*.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, 101(3):442–451.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2024). *A Practical Introduction to Regression Discontinuity Designs*. Cambridge Elements: Quantitative and Computational Methods for Social Science. Cambridge University Press., Cambridge University Press.

Cook, P. J. and Kang, S. (2016). Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation. *American Economic Journal: Applied Economics*, 8(1):33–57.

Deaner, B. (2018). Proxy controls and panel data. *arXiv:1810.00283*.

Fan, J. and Gijbels, I. (1992). Variable Bandwidth and Local Linear Regression Smoothers. *The Annals of Statistics*, 20(4).

Goodman, J., Melkers, J., and Pallais, A. (2019). Can Online Delivery Increase Access to Education? *Journal of Labor Economics*, 37(1):1–34.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Hall, A. B. (2015). What Happens When Extremists Win Primaries? *American Political Science Review*, 109(1):18–42.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.

Imbens, G. and Wager, S. (2019). Optimized regression discontinuity designs. *Review of Economics and Statistics*, 101(2):264–278.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Kong, Y. (2021). Sequential Auctions with Synergy and Affiliation across Auctions. *Journal of Political Economy*, 129(1):148–181.

Kumar, A. (2018). Do Restrictions on Home Equity Extraction Contribute to Lower Mortgage Defaults? Evidence from a Policy Discontinuity at the Texas Border. *American Economic Journal: Economic Policy*, 10(1):268–297.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Noack, C., Olma, T., and Rothe, C. (2021). Flexible covariate adjustments in regression discontinuity designs. *arXiv:2107.07942*.

Noack, C. and Rothe, C. (2024). Bias-aware inference in fuzzy regression discontinuity designs. *Econometrica*, 92(3):687–711.

Singh, R. (2020). Kernel methods for unobserved confounding: Negative controls, proxies, and instruments. *arXiv:2012.10315*.

Spenkuch, J. L. and Toniatti, D. (2018). Political Advertising and Election Results. *The Quarterly Journal of Economics*, 133(4):1981–2036.

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., and Miao, W. (2024). An introduction to proximal causal inference. *Statistical Science*, 39(3):375–390.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309.

A Identification proof

A.1 Necessary independences

Proof of Lemma 1 (necessary independences).

We proceed in steps.

1. First, we show that $A \perp\!\!\!\perp Z \mid D, U$. We write,

$$\begin{aligned}\mathbb{P}\{A = a \mid D = d, Z = z, U\} &= \mathbb{P}\{A(d, \eta_a) = a \mid D = d, Z = z, U\} \\ &= \mathbb{P}\{A(d, \eta_a) = a \mid D = d, U\},\end{aligned}$$

   where the first line follows from Assumption 2d (exclusion) and the second line follows from Assumption 2b (placebo selection on unobservables).

   To show that $Y \perp\!\!\!\perp Z \mid D, U$, first note that,

$$\begin{aligned}\mathbb{P}\{Y = y \mid D = d, Z = z, U\} &= \mathbb{P}\big\{Y(A, d, U, \eta_y) = y \mid D = d, Z = z, U\big\} \\ &= \mathbb{P}\big\{Y(A, d, U, \eta_y) = y \mid D = d, U\big\},\end{aligned}$$

   where the first line follows from Assumption 2d (exclusion) and the second line follows from Assumption 2b (placebo selection on unobservables) and the result we have already shown: $(A \perp\!\!\!\perp Z \mid D, U)$.

2. To show the second claim $(W \perp\!\!\!\perp D, Z \mid U)$, we have that,

$$\begin{aligned}\mathbb{P}\{W = w \mid D = d, Z = z, U\} &= \mathbb{P}\{W(U, \eta_w) = w \mid D = d, Z = z, U\} \\ &= \mathbb{P}\{W(U, \eta_w) = w \mid U\},\end{aligned}$$

   where the first line follows from Assumption 2d (exclusion) and the second line follows from Assumption 2b (placebo selection on unobservables).

3. Similarly, we prove the last claim $(A \perp\!\!\!\perp U \mid D)$:

$$\begin{aligned}\mathbb{P}\{A \mid D = d, U = u\} &= \mathbb{P}\{A(d, \eta_a) \mid D = d, U = u\} \\ &= \mathbb{P}\{A(d, \eta_a) \mid D = d\}.\end{aligned}$$

   where the first line follows from Assumption 2d (exclusion) and the second line follows from Assumption 3b $(\eta_a \perp\!\!\!\perp U \mid D)$.

□

A.2 Limit

Proof of Lemma 2 (limit).

Assume a fuzzy design. We proceed in steps.

1. Notice that under Assumption 2 (placebo variables), we are able to write,

$$Y = Y(0, D, U, \eta_y) + A(D, \eta_a)\big(Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y)\big).$$

2. Then, we have that,

$$\begin{aligned}
&\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\} \\
&= \mathbb{E}\big\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\big\} - \mathbb{E}\big\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \\
&\quad + \mathbb{E}\big\{A(D, \eta_a)\big(Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y)\big) \mid D = d^* + \epsilon, U\big\} \\
&\quad - \mathbb{E}\big\{A(D, \eta_a)\big(Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y)\big) \mid D = d^* - \epsilon, U\big\} \qquad \text{(I)} \\
&= \mathbb{E}\big\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\big\} - \mathbb{E}\big\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \\
&\quad + \mathbb{E}\{A(D, \eta_a) \mid D = d^* + \epsilon, U\}\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^* + \epsilon, U\big\} \\
&\quad - \mathbb{E}\{A(D, \eta_a) \mid D = d^* - \epsilon, U\}\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \qquad \text{(II)} \\
&= \mathbb{E}\big\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\big\} - \mathbb{E}\big\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \\
&\quad + \mathbb{E}\{A(D, \eta_a) \mid D = d^* + \epsilon\}\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^* + \epsilon, U\big\} \\
&\quad - \mathbb{E}\{A(D, \eta_a) \mid D = d^* - \epsilon\}\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \qquad \text{(III)},
\end{aligned}$$

   where (I) follows from the expansion in step 1. The factoring of $A(D, \eta_a)$ in (II) follows from Assumption 3 ($\eta_a \perp\!\!\!\perp \eta_y, U \mid D = d$) and weak union, whereby $\eta_a \perp\!\!\!\perp \eta_y \mid D = d, U$. Alternatively, one may simply assume the condition $\eta_a \perp\!\!\!\perp \eta_y \mid D = d, U$. Dropping the conditioning on U in (III) follows from Lemma 1c ($A \perp\!\!\!\perp U \mid D$).

3. By Assumptions 1 (RDD) and 4 (continuity), we can take the limit of the equality above to show,

$$\begin{aligned}
&\lim_{\epsilon \downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \\
&= \lim_{\epsilon \downarrow 0}\big[\mathbb{E}\big\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\big\} - \mathbb{E}\big\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\big\}\big] \\
&\quad + \lim_{\epsilon \downarrow 0}[\mathbb{E}\{A(D, \eta_a) \mid D = d^* + \epsilon\}] \lim_{\epsilon \downarrow 0}\big[\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^* + \epsilon, U\big\}\big] \\
&\quad - \lim_{\epsilon \downarrow 0}[\mathbb{E}\{A(D, \eta_a) \mid D = d^* - \epsilon\}] \lim_{\epsilon \downarrow 0}\big[\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^* - \epsilon, U\big\}\big] \\
&= \lim_{\epsilon \downarrow 0}[\mathbb{E}\{A(D, \eta_a) \mid D = d^* + \epsilon\} - \mathbb{E}\{A(D, \eta_a) \mid D = d^* - \epsilon\}] \\
&\quad \times \mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^*, U\big\}.
\end{aligned}$$

   Specifically, to go from the second line to the third line, the initial terms cancel, and the later terms factorize.

   Rearranging the above, we then find that,

$$\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^*, U\big\} = \frac{\lim_{\epsilon \downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}]}{\lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon\} - \mathbb{E}\{A \mid D = d^* - \epsilon\}]}.$$

4. This implies

$$\begin{aligned}
\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^*\big\} &= \mathbb{E}\big\{\mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^*, U\big\} \mid D = d^*\big\} \\
&= \frac{\mathbb{E}\big\{\lim_{\epsilon \downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \mid D = d^*\big\}}{\lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon\} - \mathbb{E}\{A \mid D = d^* - \epsilon\}]}
\end{aligned}$$

where first equality follows from the law of iterated expectations and the second equality substitutes in our result above.

The argument for the sharp design is the same, recognizing that Assumption 3 vacuously holds and simplifying the denominator. $\square$

A.3 Factuals

Lemma A.1 (Towards factuals). *Lemma 1 implies for all $d \in \mathcal{D}(\epsilon)$ and all $z \in \mathcal{Z}(\epsilon)$,*

$$\mathbb{P}(y \mid d, z) = \int \mathbb{P}(y \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z)$$
$$\mathbb{P}(w \mid d, z) = \int \mathbb{P}(w \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) = \int \mathbb{P}(w \mid u) \, \mathrm{d}\mathbb{P}(u \mid d, z)$$

Proof of Lemma A.1 (towards factuals).

To lighten notation, we abbreviate $(D = d, Z = z) = (d, z)$ and similarly for other variables. Then, we have that,

$$\mathbb{P}(y \mid d, z) = \int \mathbb{P}(y, u \mid d, z) \, \mathrm{d}u = \int \mathbb{P}(y \mid d, z, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) = \int \mathbb{P}(y \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z)$$
$$\mathbb{P}(w \mid d, z) = \int \mathbb{P}(w, u \mid d, z) \, \mathrm{d}u = \int \mathbb{P}(w \mid d, z, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) = \int \mathbb{P}(w \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) = \int \mathbb{P}(w \mid u) \, \mathrm{d}\mathbb{P}(u \mid d, z),$$

where we use that $Y \perp\!\!\!\perp Z \mid U, D$ and $W \perp\!\!\!\perp D, Z \mid U$ from Lemma 1. $\square$

Proof of Lemma 3 (factuals I).
We have that,

$$\begin{aligned}
\mathbb{E}\{Y \mid d, z\} &= \int y \, \mathrm{d}\mathbb{P}(y \mid d, z) \\
&= \int y \, \mathrm{d}\mathbb{P}(y \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) \qquad \text{(I)} \\
&= \int \mathbb{E}\{Y \mid d, u\} \, \mathrm{d}\mathbb{P}(u \mid d, z) \\
&= \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) \qquad \text{(II)} \\
&= \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid d, z) \qquad \text{(III)},
\end{aligned}$$

where (I) and (III) follow from Lemma A.1 (towards factuals), and (II) follows from Assumption 5 (confounding bridge). $\square$

Proof of Lemma 4 (factuals II).
Suppose a solution to $\mathbb{E}\{Y \mid d, z\} = \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid d, z)$ exists. Then by Lemma A.1 (towards factuals),

$$\begin{aligned}
\mathbb{E}\{Y \mid d, z\} &= \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid d, z) \\
&= \int h_0(d - d^*, w) \, \mathrm{d}\mathbb{P}(w \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z).
\end{aligned}$$

Furthermore, using Lemma A.1 (towards factuals), we can also write that,

$$\begin{aligned}\mathbb{E}\{Y \mid d, z\} &= \int y \, \mathrm{d}\mathbb{P}(y \mid d, z) \\ &= \int y \, \mathrm{d}\mathbb{P}(y \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z) \\ &= \int \mathbb{E}(Y \mid d, u) \, \mathrm{d}\mathbb{P}(u \mid d, z).\end{aligned}$$

Therefore by Assumption 6 (completeness), we can equate the objects within the integrals. □

A.4 Main result

Proof of Theorem 1 (placebo identification).

We proceed in steps.

1. First, note that

$$\begin{aligned}\mathbb{E}\left\{\int h_0(d-d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D=d, U) \,\middle|\, D=d'\right\} &= \mathbb{E}\left\{\int h_0(d-d^*, w) \, \mathrm{d}\mathbb{P}(w \mid U) \,\middle|\, D=d'\right\} \\ &= \mathbb{E}\left\{\int h_0(d-d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D=d', U) \,\middle|\, D=d'\right\} \\ &= \int h_0(d-d^*, w) \, \mathrm{d}\mathbb{P}(w \mid D=d'),\end{aligned}$$

where the first and second line follow from Lemma 1b ($W \perp\!\!\!\perp D \mid U$) and the last line uses the law of iterated expectations.

2. Also recall from the proof of Lemma 2 (limit) we showed that,

$$\begin{aligned}&\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D=d^*+\epsilon, U\}] - \lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D=d^*-\epsilon, U\}] \\ &= \lim_{\epsilon\downarrow 0}[\mathbb{E}\{A(D,\eta_a) \mid D=d^*+\epsilon\} - \mathbb{E}\{A(D,\eta_a) \mid D=d^*-\epsilon\}] \\ &\quad \times \mathbb{E}\big\{Y(1,D,U,\eta_y) - Y(0,D,U,\eta_y) \mid D=d^*, U\big\}.\end{aligned}$$

The limits on the right hand side exist due to Assumption 1 (RDD). Therefore the limits on the left hand side also exist.

3. Therefore, we can write the numerator of the expression from Lemma 2 (limit) as,

$$\begin{aligned}&\mathbb{E}\left\{\lim_{\epsilon\downarrow 0}\mathbb{E}\{Y \mid D=d^*+\epsilon, U\} - \lim_{\epsilon\downarrow 0}\mathbb{E}\{Y \mid D=d^*-\epsilon, U\} \,\middle|\, D=d^*\right\} \\ &= \mathbb{E}\left\{\lim_{\epsilon\downarrow 0}\int h_+(\epsilon, w) \, \mathrm{d}\mathbb{P}(w \mid D=d^*+\epsilon, U) \,\middle|\, D=d^*\right\} - \mathbb{E}\left\{\lim_{\epsilon\downarrow 0}\int h_-(-\epsilon, w) \, \mathrm{d}\mathbb{P}(w \mid D=d^*-\epsilon, U) \,\middle|\, D=d^*\right\} \\ &= \lim_{\epsilon\downarrow 0}\mathbb{E}\left\{\int h_+(\epsilon, w) \, \mathrm{d}\mathbb{P}(w \mid D=d^*+\epsilon, U) \,\middle|\, D=d^*\right\} - \lim_{\epsilon\downarrow 0}\mathbb{E}\left\{\int h_-(-\epsilon, w) \, \mathrm{d}\mathbb{P}(w \mid D=d^*-\epsilon, U) \,\middle|\, D=d^*\right\} \\ &= \lim_{\epsilon\downarrow 0}\int h_+(\epsilon, w) \, \mathrm{d}\mathbb{P}(w \mid D=d^*) - \lim_{\epsilon\downarrow 0}\int h_-(-\epsilon, w) \, \mathrm{d}\mathbb{P}(w \mid D=d^*)\end{aligned}$$

where the first equality follows from Assumption 5 (confounding bridge), the second equality uses the dominated convergence theorem and the boundedness of Assumption 5 (confounding bridge), and the last equality follows from step 1.

4. Finally, we argue that the overall expression is unique. To begin, note that $\mathbb{E}\{Y \mid d, U\}$ is unique. Furthermore, in the proof of Lemma 4 (factuals II), we showed that $\mathbb{E}\{Y \mid d, U\} = \int h_0(d - d^*, w)\, \mathrm{d}\mathbb{P}(w \mid d, U)$ almost surely by appealing to Assumption 6, so the latter integral is unique.

□

B Relaxing Assumption 2b

Assumption 2b implies the necessary condition $Z \perp\!\!\!\perp \eta_y, \eta_a, \eta_w | D, U$.

In this appendix, we show that this necessary condition, together with an additional continuity condition, lead to the same main result. Intuitively, we trade off an assumption on placebo exogeneity for an assumption on placebo continuity, i.e. independence for functional form.

Our goal is to define an alternative set of assumptions under which the conclusion of Theorem 1 continues to hold, replacing Assumption 2b with weaker conditions. We summarize the invariances of the full structural system in the next section. A graphical summary of the causal structure is provided in Figure 4.

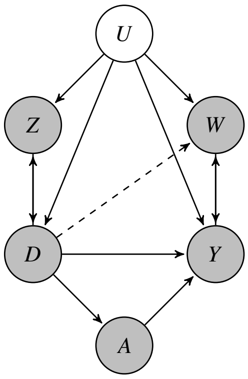


Figure 4: Additional continuity DAG

B.1 Extension

Assumption B.1 (Placebo variables)**.** *Assume there exists some $\epsilon > 0$ such that for all $d \in \mathcal{D}(\epsilon)$ and for all $z \in \mathcal{Z}(\epsilon)$,*

a (Causal consistency): if $A = a$, $D = d$, and $Z = z$ then $Y = Y(a, d, z, U, \eta_y)$, and $W = W(a, d, z, U, \eta_w)$ almost surely. If $D = d$ and $Z = z$ then $A = A(d, z, U, \eta_A)$ almost surely.

b (Placebo selection on unobservables): $Z \perp\!\!\!\perp \eta_y, \eta_a, \eta_w \mid U, D$.

c (Overlap): if $f(u) > 0$ then $f(a, d, z \mid u) > 0$, where we assume the densities exist.

d (Exclusion): $Y(a, d, z, U, \eta_y) = Y(a, d, U, \eta_y)$, $W(a, d, z, U, \eta_w) = W(U, \eta_w)$, and $A(d, z, U, \eta_a) = A(d, \eta_a)$ almost surely.

Assumptions B.1a,c,d are the same as those in Assumption 2. However, Assumption B.1c is weaker than Assumption 2c by weak union. This relaxation makes placebo selection on observables a condition that is exclusively about the placebo treatment.

Lemma B.1 (Necessary independences). *Assumption 2 implies that for all $d \in \mathcal{D}(\epsilon)$ and $z \in \mathcal{Z}(\epsilon)$,*

a $Y \perp\!\!\!\perp Z \mid D, U$

b $W \perp\!\!\!\perp Z \mid D, U$

Additionally under Assumption 3b it follows that,

c $A \perp\!\!\!\perp U \mid D$

Lemma B.1b is weaker than Lemma 1b by weak union. The rest of the lemma remains the same.

Assumption B.2 (Confounding bridge continuity). *Assume that the mapping $d \mapsto \mathbb{E}\{h_+(d - d^*, W) \mid D = d, U = u\}$ is continuous for all $d \in \mathcal{D}_+(\epsilon)$ and all u. Similarly, assume that $d \mapsto \mathbb{E}\{h_-(d - d^*, W) \mid D = d, U = u\}$ is continuous for all $d \in \mathcal{D}_-(\epsilon)$ and all u.*

This additional continuity condition generalizes the standard RDD continuity condition (Assumption 4). In particular, it replaces the potential outcome function with the confounding bridge function. Intuitively, if confounding is continuous, then we can allow more complex forms of it.

Theorem B.1 (Placebo identification). *Suppose Assumptions 1, 3, 4, 5, B.1, and B.2 hold. Then*

$$\tau_0 = \frac{\lim_{\epsilon \downarrow 0} \int h_+(\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*) - \lim_{\epsilon \downarrow 0} \int h_-(-\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*)}{\lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon\} - \mathbb{E}\{A \mid D = d^* - \epsilon\}]}$$

If in addition Assumption 6 holds then the integrals of h_0 in the numerator are identified. Therefore τ_0 is identified.

In sharp design, under Assumptions 1, 4, 5, B.1, and B.2, then this simplifies to:

$$\tau_0 = \lim_{\epsilon \downarrow 0} \int h_+(\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*) - \lim_{\epsilon \downarrow 0} \int h_-(-\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*)$$

Theorem B.1 recovers our main identification result (Theorem 1) under an alternative set of assumptions, as desired.

B.2 Proof of extension

Proof of Lemma B.1. Compared to the proof of Lemma 1, the only difference is in the second claim ($W \perp\!\!\!\perp Z \mid D, U$). We have that,

$$\begin{aligned}\mathbb{P}\{W = w \mid D = d, Z = z, U\} &= \mathbb{P}\{W(U, \eta_w) = w \mid D = d, Z = z, U\} \\ &= \mathbb{P}\{W(U, \eta_w) = w \mid D, U\},\end{aligned}$$

where the first line follows from Assumption B.1d (exclusion) and the second line follows from Assumption B.1b (placebo selection on unobservables). □

Lemma B.2 (Towards factuals). *Lemma 1 implies for all $d \in \mathcal{D}(\epsilon)$ and all $z \in \mathcal{Z}(\epsilon)$,*

$$\begin{aligned}\mathbb{P}(y \mid d, z) &= \int \mathbb{P}(y \mid d, u)\ \mathrm{d}\mathbb{P}(u \mid d, z) \\ \mathbb{P}(w \mid d, z) &= \int \mathbb{P}(w \mid d, u)\ \mathrm{d}\mathbb{P}(u \mid d, z).\end{aligned}$$

Proof. The argument is identical to Lemma A.1, excluding the final step. □

Proof of Theorem B.1. Notice that Lemma 2 remains unchanged under the conditions of this appendix, since it only appeals to Lemma 1c, which remains unchanged. Therefore, we proceed in steps similar to the proof of Theorem 1.

1. From the proof of Lemma 2 (limit) we showed that,

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\}] - \lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \\
&= \lim_{\epsilon\downarrow 0}[\mathbb{E}\{A(D,\eta_a) \mid D = d^* + \epsilon\} - \mathbb{E}\{A(D,\eta_a) \mid D = d^* - \epsilon\}] \\
&\quad \times \mathbb{E}\big\{Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y) \mid D = d^*, U\big\}.
\end{aligned}$$

   The limits on the right hand side exist due to Assumption 1 (RDD). Therefore the limits on the left hand side also exist.

2. Define $h_-(0, w) = \lim_{\epsilon\downarrow 0} h_-(\epsilon, w)$ for all w, where the existence of the limit follows from Assumption B.2. Then, we can write the numerator of the expression from Lemma 2 (limit) as,

$$\begin{aligned}
&\mathbb{E}\left\{\lim_{\epsilon\downarrow 0}\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \lim_{\epsilon\downarrow 0}\mathbb{E}\{Y \mid D = d^* - \epsilon, U\} \,\middle|\, D = d^*\right\} \\
&= \mathbb{E}\left\{\lim_{\epsilon\downarrow 0}\int h_+(\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^* + \epsilon, U) \,\middle|\, D = d^*\right\} - \mathbb{E}\left\{\lim_{\epsilon\downarrow 0}\int h_-(-\epsilon, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^* - \epsilon, U) \,\middle|\, D = d^*\right\} \\
&= \mathbb{E}\left\{\int h_+(0, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*, U) \,\middle|\, D = d^*\right\} - \mathbb{E}\left\{\int h_-(0, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*, U) \,\middle|\, D = d^*\right\} \\
&= \int h_+(0, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*) - \int h_-(0, w)\ \mathrm{d}\mathbb{P}(w \mid D = d^*)
\end{aligned}$$

   where the first equality follows from Assumption 5 (confounding bridge), the second equality uses Assumption B.2, and the third uses the law of iterated expectations.

3. Finally, we argue that the overall expression is unique. To begin, note that $\mathbb{E}\{Y \mid d, U\}$ is unique. Furthermore, in the proof of Lemma 4 (factuals II), we showed that $\mathbb{E}\{Y \mid d, U\} = \int h_0(d - d^*, w)\ \mathrm{d}\mathbb{P}(w \mid d, U)$ almost surely by appealing to Assumption 6, so the latter integral is unique.

□

C Relaxing Assumption 3

Assumption 3 holds trivially in the sharp design. In the fuzzy design, however, it imposes a nontrivial restriction on the treatment assignment mechanism.

In this appendix, we show that Assumption 3 can be relaxed in the fuzzy design if treatment effects are homogeneous. Intuitively, we trade off a restriction on the treatment mechanism for an assumption on the outcome model, i.e. independence for functional form.

Our goal is to define an alternative set of assumptions under which the conclusion of Lemma 2 continues to hold, replacing Assumption 3 with weaker conditions. We summarize the invariances of the full structural system in the next section. A graphical summary of the causal structure is provided in Figure 5.

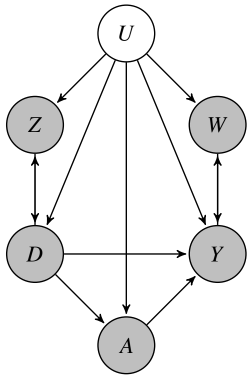


Figure 5: Homogeneous effect DAG

C.1 Extension

Assumption C.1 (RDD). *Assume that the following limits exist and are unequal.*

$$\lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* - \epsilon, U\}], \quad \lim_{\epsilon \downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon, U\}]$$

This condition modifies the standard RDD treatment assignment assumption (Assumption 1). It requires that a discontinuity in expected treatment probability at the cutoff after conditioning on U.

Assumption C.2 (Placebo variables). *Assume there exists some $\epsilon > 0$ such that for all $d \in \mathcal{D}(\epsilon)$ and for all $z \in \mathcal{Z}(\epsilon)$,*

- *a (Causal consistency): if $A = a$, $D = d$, and $Z = z$ then $Y = Y(a, d, z, U, \eta_y)$, $W = W(a, d, z, U, \eta_w)$. If $D = d$ and $Z = z$ then $A = A(d, z, U, \eta_A)$ almost surely.*
- *b (Selection on unobservables): $Z \perp\!\!\!\perp \eta_y, \eta_a \mid U$ and $\eta_w \perp\!\!\!\perp Z, D \mid U$.*
- *c (Overlap): if $f(u) > 0$ then $f(a, d, z \mid u) > 0$, where we assume the densities exist.*
- *d (Exclusion): $Y(a, d, z, U, \eta_y) = Y(a, d, U, \eta_y)$, $W(a, d, z, U, \eta_w) = W(U, \eta_w)$, and $A(d, z, U, \eta_a) = A(d, U, \eta_a)$ almost surely.*

Assumptions C.2a–c are equivalent to those in Assumption 2. However, Assumption C.2d relaxes Assumption 2d, allowing for unobserved confounding in the treatment mechanism. Under Assumption C.2, locally, our notation simplifies to

$$Y = Y(A, D, U, \eta_y), \quad W = W(U, \eta_w), \quad A = A(D, U, \eta_a),$$

and can be summarized by the result below.

Lemma C.1 (Necessary independences). *Assumption C.2 implies that for all $d \in \mathcal{D}(\epsilon)$ and $z \in \mathcal{Z}(\epsilon)$,*

- *a $A, Y \perp\!\!\!\perp Z \mid D, U$*
- *b $W \perp\!\!\!\perp D, Z \mid U$*

Assumption C.3 (Homogeneity). *Under homogenous treatment effects, we assume that for $d \in \mathcal{D}(\epsilon)$ the treatment effect is constant, i.e.*

$$\tau_0 = Y(1, d, U, \eta_y) - Y(0, d, U, \eta_y)$$

almost surely.

We use Assumption C.3 in place of Assumption 3 in Section 3.We impose that treatment effects are homogeneous across individuals near the cutoff, while allowing treatment assignment to depend on unobserved confounders that affect the potential outcomes.

Together with Assumptions C.1, C.2, and 4, this yields the following identification result:

Lemma C.2 (Limit)**.** *Under Assumptions C.1, C.2, C.3, and 4,*

$$\tau_0 = \frac{\mathbb{E}\big\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \mid D = d^*\big\}}{\mathbb{E}\big\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon, U\} - \mathbb{E}\{A \mid D = d^* - \epsilon, U\}] \mid D = d^*\big\}}$$

Lemma C.2 expresses the treatment effect τ_0 as the ratio of two discontinuities, each computed after conditioning on U and then averaging over U. This formulation permits identification even when the treatment (A) is confounded by unobservables.

For estimation, we can apply the same confounding bridge approach from Section 4 to recover the counterfactual term $\mathbb{E}\big\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{A \mid D = d^* - \epsilon, U\}] \mid D = d^*\big\}$ by using the observed proxies for U. The final estimator will look like a ratio of estimators in Section 4, using Y in the numerator and A in the denominator.

C.2 Proof of extension

Proof of Lemma C.1.
To show that $A \perp\!\!\!\perp Z \mid D, U$, we write

$$\begin{aligned}\mathbb{P}\{A = a \mid D = d, Z = z, U\} &= \mathbb{P}\{A(d, U, \eta_a) = a \mid D = d, Z = z, U\} \\ &= \mathbb{P}\{A(d, \eta_a) = a \mid D = d, U\},\end{aligned}$$

where the first line follows from Assumption C.2d (exclusion) and the second line follows from Assumption C.2b (placebo selection on unobservables). Otherwise, the result is the same as Lemma 1. □

Proof of Lemma C.2.
Assume a fuzzy design. We proceed in steps.

1. Notice that under Assumption C.2 (proxy controls), we are able to write,

$$Y = Y(0, D, U, \eta_y) + A(D, U, \eta_a)\big(Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y)\big).$$

2. Using the decomposition above, we can write

$$\begin{aligned}&\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\} \\ &= \mathbb{E}\big\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\big\} - \mathbb{E}\big\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \\ &\quad + \mathbb{E}\big\{A(d^* + \epsilon, U, \eta_a)\big(Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y)\big) \mid D = d^* + \epsilon, U\big\} \\ &\quad - \mathbb{E}\big\{A(d^* + \epsilon, U, \eta_a)\big(Y(1, D, U, \eta_y) - Y(0, D, U, \eta_y)\big) \mid D = d^* - \epsilon, U\big\} \\ &= \mathbb{E}\big\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\big\} - \mathbb{E}\big\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\big\} \\ &\quad + \tau_0\mathbb{E}\{A(d^* + \epsilon, U, \eta_a) \mid D = d^* + \epsilon, U\} - \tau_0\mathbb{E}\{A(d^* + \epsilon, U, \eta_a) \mid D = d^* - \epsilon, U\},\end{aligned}$$

where we use Assumption C.3 to derive the second equality.

3. By Assumptions C.1 (RDD) and 4 (continuity), we can take the limit of the equality above to show,

$$
\begin{aligned}
&\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \\
&= \lim_{\epsilon\downarrow 0}\left[\mathbb{E}\left\{Y(0, d^* + \epsilon, U, \eta_y) \mid D = d^* + \epsilon, U\right\} - \mathbb{E}\left\{Y(0, d^* - \epsilon, U, \eta_y) \mid D = d^* - \epsilon, U\right\}\right] \\
&\quad + \tau_0 \lim_{\epsilon\downarrow 0}[\mathbb{E}\{A(D, U, \eta_a) \mid D = d^* + \epsilon, U\} - \mathbb{E}\{A(D, U, \eta_a) \mid D = d^* - \epsilon, U\}] \\
&= \tau_0 \lim_{\epsilon\downarrow 0}[\mathbb{E}\{A(D, U, \eta_a) \mid D = d^* + \epsilon, U\} - \mathbb{E}\{A(D, U, \eta_a) \mid D = d^* - \epsilon, U\}].
\end{aligned}
$$

   Specifically, to go from the second line to the third line, the initial terms cancel, and the later terms factorize.

4. Taking conditional expectations of both sides, we have that

$$
\begin{aligned}
&\mathbb{E}\left\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \mid D = d^*\right\} \\
&= \tau_0 \mathbb{E}\left\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{A(D, U, \eta_a) \mid D = d^* + \epsilon, U\} - \mathbb{E}\{A(D, U, \eta_a) \mid D = d^* - \epsilon, U\}] \mid D = d^*\right\}.
\end{aligned}
$$

   Rearranging and gives

$$
\tau_0 = \frac{\mathbb{E}\left\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{Y \mid D = d^* + \epsilon, U\} - \mathbb{E}\{Y \mid D = d^* - \epsilon, U\}] \mid D = d^*\right\}}{\mathbb{E}\left\{\lim_{\epsilon\downarrow 0}[\mathbb{E}\{A \mid D = d^* + \epsilon, U\} - \mathbb{E}\{A \mid D = d^* - \epsilon, U\}] \mid D = d^*\right\}}.
$$

□

D Equivalence proof

D.1 Notation

Let $\mathbf{e}_0 = (1, 0)^\top$ and $\mathbf{e}_1 = (0, 1)^\top$. Furthermore, let $q = \dim(w)$ and define

$$
\begin{aligned}
\omega_{i,+} &= \frac{1}{h_n}\mathbb{1}(X_i \geq x^*)K\left(\frac{|D_i - d^*|}{h_n}\right) \in \mathbb{R} \\
R_{i,p} &= \left(1, \left(\frac{D_i - d^*}{h_n}\right), \ldots, \left(\frac{D_i - d^*}{h_n}\right)^p\right)^\top \in \mathbb{R}^{p+1} \\
H_p &= \mathrm{Diag}(1, \ldots, h_n^p) \in \mathbb{R}^{(p+1)\times(p+1)}
\end{aligned}
\qquad
\begin{aligned}
\mathbf{K}_+ &= \mathrm{Diag}(\omega_{1,+}, \ldots, \omega_{n,+}) \in \mathbb{R}^{n\times n} \\
\mathbf{R}_p &= \left(R_{1,p}, \ldots, R_{n,p}\right)^\top \in \mathbb{R}^{n\times(p+1)} \\
\mathbf{Z} &= (Z_1, \ldots, Z_n)^\top \in \mathbb{R}^{n\times q} \\
\mathbf{W} &= (W_1, \ldots, W_n)^\top \in \mathbb{R}^{n\times q}.
\end{aligned}
$$

D.2 Estimator recap

We estimate $\nu_+^\top = (\alpha_{+,0}, \alpha_{+,1}, \gamma_+^\top)$ by solving a local linear IV objective function with first order conditions,

$$
\frac{1}{n}\sum_{i=1}^{n}\omega_{i,+}\begin{bmatrix} R_{i,p} \\ Z_i \end{bmatrix}\left(Y_i - R_{i,p}^\top H_1\hat{\alpha}_+ - W_i^\top\hat{\gamma}_+\right) = \mathbf{0}.
$$

Then we have that,

$$
\hat{\nu}_+ = \left(\begin{bmatrix} \mathbf{R}_1^\top \\ \mathbf{Z}^\top \end{bmatrix}\mathbf{K}_+\begin{bmatrix} \mathbf{R}_1 & \mathbf{W} \end{bmatrix}\right)^{-1}\left(\begin{bmatrix} \mathbf{R}_1^\top \\ \mathbf{Z}^\top \end{bmatrix}\mathbf{K}_+\mathbf{Y}\right) = \begin{bmatrix} \mathbf{R}_1^\top\mathbf{K}_+\mathbf{R}_1 & \mathbf{R}_1^\top\mathbf{K}_+\mathbf{W} \\ \mathbf{Z}^\top\mathbf{K}_+\mathbf{R}_1 & \mathbf{Z}^\top\mathbf{K}_+\mathbf{W} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{R}_1^\top\mathbf{K}_+\mathbf{Y} \\ \mathbf{Z}^\top\mathbf{K}_+\mathbf{Y} \end{bmatrix}.
$$

We estimate $\beta_{+}^{w_j}$ using the local linear estimator of Hahn et al. (2001),

$$H_1 \hat{\beta}_{+}^{w_j} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \omega_{i,+} \left(W_{i,j} - R_{i,1}^{\top} \beta \right)^2 = (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{W}_{,j}.$$

Then, we let $\beta_{+,0}^{w} = \left(\beta_{+,0}^{w_1}, \ldots, \beta_{+,0}^{w_q} \right)^{\top} \in \mathbb{R}^q$ and $\beta_{+}^{w} = \left(\beta_{+}^{w_1}, \ldots, \beta_{+}^{w_q} \right) \in \mathbb{R}^{2 \times q}$.

D.3 Main result

Proof of Proposition 1 (estimator decomposition).
We proceed in steps, using regression algebra.

1. Using the formula for the inverse of a block matrix,

$$\begin{aligned}
\hat{\nu}_{+} &= \begin{bmatrix} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1 & \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{W} \\ \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{R}_1 & \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{W} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{Y} \\ \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{Y} \end{bmatrix} \\
&= \begin{bmatrix} (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} + (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{W} \, \mathbf{Q}_{+}^{-1} \, \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{R}_1 (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} & -(\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{W} \, \mathbf{Q}_{+}^{-1} \\ -\mathbf{Q}_{+}^{-1} \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{R}_1 (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} & \mathbf{Q}_{+}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{Y} \\ \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{Y} \end{bmatrix}
\end{aligned}$$

   where $\mathbf{Q}_{+} = \mathbf{Z}^{\top} \mathbf{K}_{+} \left\{ \mathbf{I} - \mathbf{R}_1 (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \right\} \mathbf{W}$ is the Schur complement. Then, it follows that for $H_1 \hat{\alpha}_{+} = (\alpha_{+,0}, h_n \alpha_{+,1})^{\top}$, we have that,

$$\begin{aligned}
H_1 \hat{\alpha}_{+} &= (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{Y} \\
&\quad + (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{W} \, \mathbf{Q}_{+}^{-1} \, \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{R}_1 (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{Y} \\
&\quad - (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{W} \, \mathbf{Q}_{+}^{-1} \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{Y} \\
&= H_1 \hat{\beta}_{+}^{y} + H_1 \hat{\beta}_{+}^{w} \left\{ \mathbf{Q}_{+}^{-1} \, \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{R}_1 (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{Y} - \mathbf{Q}_{+}^{-1} \mathbf{Z}^{\top} \mathbf{K}_{+} \mathbf{Y} \right\} \\
&= H_1 \hat{\beta}_{+}^{y} - H_1 \hat{\beta}_{+}^{w} \left[\mathbf{Q}_{+}^{-1} \mathbf{Z}^{\top} \mathbf{K}_{+} \{ \mathbf{I} - \mathbf{R}_1 (\mathbf{R}_1^{\top} \mathbf{K}_{+} \mathbf{R}_1)^{-1} \mathbf{R}_1^{\top} \mathbf{K}_{+} \} \mathbf{Y} \right] \\
&= H_1 \hat{\beta}_{+}^{y} - H_1 \hat{\beta}_{+}^{w} \hat{\gamma}_{+},
\end{aligned}$$

   where $\hat{\beta}_{+}^{y}$ and $\hat{\beta}_{+}^{w_j}$ solve the local linear objective functions,

$$H_1 \hat{\beta}_{+}^{y} = \arg\min_{\beta} \sum_{i=1}^{n} \omega_{i,+} \left(Y_i - R_{i,1}^{\top} \beta \right)^2, \qquad H_1 \hat{\beta}_{+}^{w_j} = \arg\min_{\beta} \sum_{i=1}^{n} \omega_{i,+} \left(W_{i,j} - R_{i,1}^{\top} \beta \right)^2.$$

2. We can go through the same steps to show that $H_1 \hat{\alpha}_{-} = H_1 \hat{\beta}_{-}^{y} - H_1 \hat{\beta}_{-}^{w} \hat{\gamma}_{-}$.

3. Plugging these expressions into our estimator for $\hat{\tau}_{\text{pdd}}$, we find that,

$$\begin{aligned}
\hat{\tau}_{\text{pdd}} &= \hat{\alpha}_{+,0} + \left(\hat{\beta}_{+,0}^{w} \right)^{\top} \hat{\gamma}_{+} - \left\{ \hat{\alpha}_{-,0} + \left(\hat{\beta}_{-,0}^{w} \right)^{\top} \hat{\gamma}_{-} \right\} \\
&= \hat{\beta}_{+,0}^{y} - \left(\hat{\beta}_{+,0}^{w} \right)^{\top} \hat{\gamma}_{+} + \left(\hat{\beta}_{+,0}^{w} \right)^{\top} \hat{\gamma}_{+} - \left(\hat{\beta}_{-,0}^{y} - \left(\hat{\beta}_{-,0}^{w} \right)^{\top} \hat{\gamma}_{-} + \left(\hat{\beta}_{+,0}^{w} \right)^{\top} \hat{\gamma}_{-} \right) \\
&= \hat{\beta}_{+,0}^{y} - \hat{\beta}_{-,0}^{y} + \left(\hat{\beta}_{+,0}^{w} - \hat{\beta}_{-,0}^{w} \right)^{\top} \hat{\gamma}_{-} \\
&= \hat{\tau}_{\text{rdd}}^{y} + \left(\hat{\tau}_{\text{rdd}}^{w} \right)^{\top} \hat{\gamma}_{-}.
\end{aligned}$$

4. Finally, we show the claim that $\hat{\gamma}_{-} = \left[\frac{1}{n} \sum_{i=1}^{n} \omega_{i,-} \{ Z_i (W_i^{\perp})^{\top} \} \right]^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \omega_{i,-} (Z_i Y_i^{\perp}) \right\}$. Expanding

out our expression for $\hat{\gamma}_-$, we have that,

$$
\begin{aligned}
\hat{\gamma}_- &= \mathbf{Q}_-^{-1}\mathbf{Z}^\top\mathbf{K}_-\{\mathbf{I}-\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\}\mathbf{Y} \\
&= \left[\mathbf{Z}^\top\mathbf{K}_-\{\mathbf{I}-\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\}\mathbf{W}\right]^{-1}\left[\mathbf{Z}^\top\mathbf{K}_-\{\mathbf{I}-\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\}\mathbf{Y}\right] \\
&= \left[\sum_{i=1}^n \omega_{i,-}Z_i\{W_i^\top - R_{i,1}^\top(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\mathbf{W}\}\right]^{-1}\left[\sum_{i=1}^n \omega_{i,-}Z_i\{Y_i - R_{i,1}^\top(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\mathbf{Y}\}\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^n \omega_{i,-}Z_i\left\{W_i^\top - R_{i,1}^\top\left(\frac{1}{n}\sum_{i=1}^n \omega_{i,-}R_{i,1}R_{i,1}^\top\right)^{-1}\frac{1}{n}\sum_{i=1}^n \omega_{i,-}R_{i,1}W_i^\top\right\}\right]^{-1} \\
&\quad\times\left[\frac{1}{n}\sum_{i=1}^n \omega_{i,-}Z_i\left\{Y_i - R_{i,1}^\top\left(\frac{1}{n}\sum_{i=1}^n \omega_{i,-}R_{i,1}R_{i,1}^\top\right)^{-1}\frac{1}{n}\sum_{i=1}^n \omega_{i,-}R_{i,1}Y_i\right\}\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^n \omega_{i,-}\{Z_i(W_i^\perp)^\top\}\right]^{-1}\left\{\frac{1}{n}\sum_{i=1}^n \omega_{i,-}(Z_iY_i^\perp)\right\}.
\end{aligned}
$$

□

E Consistency proof

E.1 Main result

Proof of Proposition 2 (estimand).

We proceed in steps.

1. By Proposition 1,

$$\hat{\tau}_{\text{pdd}} = \hat{\beta}_{+,0}^y - \hat{\beta}_{-,0}^y + \left(\hat{\beta}_{+,0}^w - \hat{\beta}_{-,0}^w\right)^\top \hat{\gamma}_-.$$

   Standard results from e.g. Hahn et al. (2001, Section 4.1) show that under our conditions,

$$
\begin{aligned}
\hat{\beta}_{+,0}^y - \hat{\beta}_{-,0}^y &\xrightarrow{p} \lim_{\epsilon\downarrow 0}\{\mathbb{E}\{Y \mid D = d^*+\epsilon\} - \mathbb{E}\{Y \mid D = d^*-\epsilon\}\} \\
\hat{\beta}_{+,0}^w - \hat{\beta}_{-,0}^w &\xrightarrow{p} \lim_{\epsilon\downarrow 0}\{\mathbb{E}\{W \mid D = d^*+\epsilon\} - \mathbb{E}\{W \mid D = d^*-\epsilon\}\}.
\end{aligned}
$$

   Therefore, appealing to the continuous mapping theorem, it suffices to show that $\hat{\gamma}_- \xrightarrow{p} \gamma_-$.

2. Within the proof of Proposition 1, we have shown that for $\mathbf{Q}_- = \mathbf{Z}^\top\mathbf{K}_-\left\{\mathbf{I}-\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\right\}\mathbf{W}$,

$$\hat{\gamma}_- = \mathbf{Q}_-^{-1}\mathbf{Z}^\top\mathbf{K}_-\{\mathbf{I}-\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\}\mathbf{Y}.$$

   Using Lemma E.1 ($\mathbf{Q}_-$-limit) and the continuous mapping theorem, we can write that,

$$n\mathbf{Q}_-^{-1} \xrightarrow{p} \kappa^{-1}\lim_{\epsilon\downarrow 0}\mathrm{Cov}\big(Z_i, W_i^\top \mid D = d^*-\epsilon\big)^{-1}, \qquad \kappa = f_d(d^*)\int_0^\infty K(u)\,\mathrm{d}u.$$

   By Lemma E.1 ($\mathbf{Q}_-$-limit), replacing $\mathbf{W}$ with $\mathbf{Y}$, we can analogously show that,

$$\frac{1}{n}\mathbf{Z}^\top\mathbf{K}_-\{\mathbf{I}-\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\}\mathbf{Y} \xrightarrow{p} \kappa\lim_{\epsilon\downarrow 0}\mathrm{Cov}(Z_i, Y_i \mid D_i = d^*-\epsilon).$$

Therefore, applying the continuous mapping and dominated convergence theorems, it follows that,

$$\hat{\gamma}_- \xrightarrow{p} \lim_{\epsilon\downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D = d^* - \epsilon\big)^{-1} \mathrm{Cov}(Z_i, Y_i \mid D_i = d^* - \epsilon).$$

3. We verify the claim that γ_- is the best linear predictor for

$$Y = c + W^\top \gamma_- + \eta, \qquad \lim_{\epsilon\downarrow 0} \mathbb{E}\{\eta Z_i \mid D_i = d^* - \epsilon\} = 0, \qquad \lim_{\epsilon\downarrow 0} \mathbb{E}\{\eta \mid D_i = d^* - \epsilon\} = 0.$$

Substituting for Y in the probability limit gives

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D_i = d^* - \epsilon\big)^{-1} \mathrm{Cov}(Z_i, Y_i \mid D_i = d^* - \epsilon) \\
&\quad = \lim_{\epsilon\downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D_i = d^* - \epsilon\big)^{-1} \big\{\mathrm{Cov}\big(Z_i, c + W_i^\top \gamma_- + \eta \mid D_i = d^* - \epsilon\big)\big\} \\
&\quad = \lim_{\epsilon\downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D_i = d^* - \epsilon\big)^{-1} \big\{\mathrm{Cov}\big(Z_i, W_i^\top \mid D_i = d^* - \epsilon\big)\gamma_- + \mathrm{Cov}(Z_i, \eta \mid D_i = d^* - \epsilon)\big\} \\
&\quad = \gamma_-.
\end{aligned}$$

□

Proof of Corollary 1 (consistency).

We proceed in steps. By symmetry, we focus only on showing convergence above the cutoff.[3] We want to show that,

$$\hat{\alpha}_{+,0} + (\hat{\beta}^w_{+,0})^\top \hat{\gamma}_+ \xrightarrow{p} \lim_{\epsilon\downarrow 0} \mathbb{E}\{h_+(\epsilon, W) \mid D = d^*\}.$$

1. We express the limit as a sum.

Substituting in $h_+(d - d^*, w)$ from Equation (1),

$$\begin{aligned}
\lim_{\epsilon\downarrow 0} \mathbb{E}\{h_+(\epsilon, W) \mid D = d^*\} &= g_+(0) + \mathbb{E}\big\{W^\top \mid D = d^*\big\}\gamma_+ \\
&= \alpha_{+,0} + \left(\beta^w_{+,0}\right)^\top \gamma_+.
\end{aligned}$$

2. It is helpful to characterize a partially linear representation for Y.

By Lemma 3 (factuals) and Equation (1), for all $d \in \mathcal{D}_+(\epsilon)$ recall that we can write,

$$\mathbb{E}\{Y - h_+(D - d^*, W) \mid D = d, Z = z\} = \mathbb{E}\big\{Y - g_+(D_i - d^*) - W^\top \gamma_+ \mid D = d, Z = z\big\} = 0.$$

Using this moment condition, it follows that for some $\delta \in [0, D - d^*]$ we can use a Taylor expansion to represent,

$$\begin{aligned}
Y &= g_+(D - d^*) + W^\top \gamma_+ + \tilde{\eta} \\
&= g_+(0) + g_+^{(1)}(0)(D - d^*) + W^\top \gamma_+ + \tilde{\eta} + \frac{1}{2} g_+^{(2)}(\delta)(D - d^*)^2 \\
&= \alpha_{+,0} + \alpha_{+,1}(D - d^*) + W^\top \gamma_+ + \tilde{\eta} + \frac{1}{2} g_+^{(2)}(\delta)(D - d^*)^2,
\end{aligned}$$

where $\mathbb{E}\{\tilde{\eta} \mid D = d, Z = z\} = 0$ for all $d \in \mathcal{D}_+(\epsilon)$ and $z \in \mathcal{Z}_+(\epsilon)$.

3. Using the representation in step 2, we now show that $\alpha_{+,0} = \beta^y_{+,0} - (\beta^w_{+,0})^\top \gamma_+$.

[3] We will not use cancellation, so that the same argument gives convergence below the cutoff.

Applying the conditional expectation $\mathbb{E}\{\cdot \mid Z, D = d^* + \epsilon\}$ on both sides,

$$\mathbb{E}\{Y \mid Z, D = d^* + \epsilon\} = \alpha_{+,0} + \alpha_{+,1}\epsilon + \mathbb{E}\{W^\top \mid Z, D = d^* + \epsilon\}\gamma_+ + \frac{1}{2}g_+^{(2)}(\delta)\epsilon^2.$$

Further taking the conditional expectation $\mathbb{E}\{\cdot \mid D = d^* + \epsilon\}$ on both sides and appealing to the law of iterated expectations,

$$\mathbb{E}\{Y \mid D = d^* + \epsilon\} = \alpha_{+,0} + \alpha_{+,1}\epsilon + \mathbb{E}\{W^\top \mid D = d^* + \epsilon\}\gamma_+ + \frac{1}{2}g_+^{(2)}(\delta)\epsilon^2.$$

Finally, taking the limit on both sides,

$$\lim_{\epsilon \downarrow 0} \mathbb{E}\{Y \mid D_i = d^* + \epsilon\} = \alpha_{+,0} + \lim_{\epsilon \downarrow 0} \mathbb{E}\{W_i^\top \mid D_i = d^* + \epsilon\}\gamma_+.$$

Rearranging gives $\alpha_{+,0} = \beta^y_{+,0} - (\beta^w_{+,0})^\top \gamma_+$.

4. Next, we show that $\hat{\gamma}_+ \xrightarrow{p} \gamma_+$.

   As argued in the proof of Proposition 2 (estimand) we have that,

$$\begin{aligned}
\hat{\gamma}_+ &\xrightarrow{p} \lim_{\epsilon \downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D = d^* + \epsilon\big)^{-1} \mathrm{Cov}(Z_i, Y_i \mid D_i = d^* + \epsilon) \\
&= \lim_{\epsilon \downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D = d^* + \epsilon\big)^{-1} \mathrm{Cov}\Big(Z_i, \alpha_{+,0} + \alpha_{+,1}(D_i - d^*) + W_i^\top \gamma_+ + \tilde{\eta} + \frac{1}{2}g_+^{(2)}(\delta)(D_i - d^*)^2 \mid D_i = d^* + \epsilon\Big) \\
&= \lim_{\epsilon \downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D = d^* + \epsilon\big)^{-1} \mathrm{Cov}\big(Z_i, W_i^\top \mid D_i = d^* + \epsilon\big)\gamma_+ \\
&\quad + \lim_{\epsilon \downarrow 0} \mathrm{Cov}\big(Z_i, W_i^\top \mid D = d^* + \epsilon\big)^{-1} \mathrm{Cov}(Z_i, \tilde{\eta} \mid D_i = d^* + \epsilon) \\
&= \gamma_+,
\end{aligned}$$

   where the first equality uses step 2 and the third equality uses that

$$\begin{aligned}
\mathrm{Cov}(Z_i, \tilde{\eta} \mid D_i = d^* + \epsilon) &= \mathbb{E}\{Z_i \tilde{\eta} \mid D_i = d^* + \epsilon\} - \mathbb{E}\{Z_i \mid D_i = d^* + \epsilon\}\mathbb{E}\{\tilde{\eta} \mid D_i = d^* + \epsilon\} \\
&= \mathbb{E}\{Z_i \mathbb{E}\{\tilde{\eta} \mid Z_i, D_i = d^* + \epsilon\}\} - \mathbb{E}\{Z_i \mid D_i = d^* + \epsilon\}\mathbb{E}\{\mathbb{E}\{\tilde{\eta} \mid Z_i, D_i = d^* + \epsilon\}\} \\
&= 0.
\end{aligned}$$

5. We also show that $\hat{\alpha}_{+,0} \xrightarrow{p} \alpha_{+,0}$.

   In the proof of Proposition 1 (estimator decomposition), we showed that $\hat{\alpha}_{+,0} = \hat{\beta}^y_{+,0} - (\hat{\beta}^w_{+,0})^\top \hat{\gamma}_+$. In step 3, we showed that $\alpha_{+,0} = \beta^y_{+,0} - (\beta^w_{+,0})^\top \gamma_+$.

   Recall that $\hat{\beta}^y_{+,0} \xrightarrow{p} \beta^y_{+,0}$ and $\hat{\beta}^w_{+,0} \xrightarrow{p} \beta^w_{+,0}$ by standard arguments (Hahn et al., 2001, Section 4.1). Using that $\hat{\gamma}_+ \xrightarrow{p} \gamma_+$ from step 4, the continuous mapping theorem implies that $\hat{\beta}^y_{+,0} - (\hat{\beta}^w_{+,0})^\top \hat{\gamma}_+ \xrightarrow{p} \beta^y_{+,0} - (\beta^w_{+,0})^\top \gamma_+$. Therefore, $\hat{\alpha}_{+,0} \xrightarrow{p} \alpha_{+,0}$.

6. Collecting results and applying the continuous mapping theorem we have shown that,

$$\begin{aligned}
\hat{\alpha}_{+,0} + (\hat{\beta}^w_{+,0})^\top \hat{\gamma}_+ &\xrightarrow{p} \alpha_{+,0} + (\beta^w_{+,0})^\top \gamma_+ \\
&= \lim_{\epsilon \downarrow 0} \mathbb{E}\{h_+(\epsilon, W) \mid D = d^*\}
\end{aligned}$$

   where the convergence uses steps 4 and 5, and the equality uses step 1.

$\square$

E.2 Technical lemma

Lemma E.1 ($\mathbf{Q}_-$-limit). *Let Assumptions 7a and 8 hold, and assume that $h \to 0$ and $nh \to \infty$. Then,*

$$\frac{1}{n}\mathbf{Q}_- \xrightarrow{p} \kappa \lim_{\epsilon \downarrow 0} \operatorname{Cov}(Z_i, W_i^\top \mid D = d^* - \epsilon),$$

for $\kappa = f_d(d^*) \int_0^\infty K(u)\, \mathrm{d}u$.

Proof of Lemma E.1.

We proceed in steps.

1. As shown in the proof of Proposition 1,

$$\mathbf{Q}_- = \mathbf{Z}^\top \mathbf{K}_- \{\mathbf{I} - \mathbf{R}_1(\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_-\} \mathbf{W}.$$

   Therefore, we have the decomposition

$$\begin{aligned}
\frac{1}{n}\mathbf{Q}_- &= \frac{1}{n}\sum_{i=1}^n \omega_{i,-} Z_i \{W_i^\top - R_{i,1}^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W}\} \\
&= \frac{1}{n}\sum_{i=1}^n \omega_{i,-} Z_i \Big(W_i^\top - \mathbf{e}_0^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W}\Big) \\
&\quad - \frac{1}{n}\sum_{i=1}^n \omega_{i,-} Z_i \bigg\{\bigg(\frac{D_i - d^*}{h_n}\bigg) \mathbf{e}_1^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W}\bigg\}.
\end{aligned}$$

   We analyze each term below.

2. Under Assumptions 7a and 8, results from Hahn et al. (2001, Section 4.1) imply that

$$\begin{aligned}
&\mathbf{e}_0^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W} \xrightarrow{p} (\beta_{-,0}^w)^\top = \lim_{\epsilon \downarrow 0} \mathbb{E}\{W_i^\top \mid D_i = d^* - \epsilon\} \\
&\frac{1}{h_n} \mathbf{e}_1^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W} \xrightarrow{p} (\beta_{-,1}^w)^\top.
\end{aligned}$$

3. Consider the first term in the decomposition. We will show that

$$\frac{1}{n}\sum_{i=1}^n \omega_{i,-} Z_i \Big(W_i^\top - \mathbf{e}_0^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W}\Big) \xrightarrow{p} \kappa \lim_{\epsilon \downarrow 0} \operatorname{Cov}(Z_i, W_i^\top \mid D_i = d^* - \epsilon).$$

   By step 2 and the weak law of large numbers,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \omega_{i,-} Z_i \Big(W_i^\top - \mathbf{e}_0^\top (\mathbf{R}_1^\top \mathbf{K}_- \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{K}_- \mathbf{W}\Big) \\
&= \frac{1}{n}\sum_{i=1}^n \omega_{i,-} Z_i (W_i - \beta_{-,0}^w)^\top + o_p(1) \\
&= \left(\frac{1}{nh_n}\sum_{i=1}^n \mathbb{1}(D_i < d^*) K\bigg(\frac{|D_i - d^*|}{h_n}\bigg) Z_i (W_i - \beta_{-,0}^w)^\top\right) + o_p(1) \\
&= \frac{1}{h_n}\mathbb{E}\bigg\{\mathbb{1}(D_i < d^*) K\bigg(\frac{|D_i - d^*|}{h_n}\bigg) Z_i (W_i - \beta_{-,0}^w)^\top\bigg\} + o_p(1).
\end{aligned}$$

We use the law of iterated expectations and u-substitution with $u = \frac{d-d^*}{h_n}$ to express the former term as

$$
\begin{aligned}
&\frac{1}{h_n}\mathbb{E}\left\{\mathbb{1}(D_i < d^*)K\left(\frac{|D_i - d^*|}{h_n}\right)\mathbb{E}\left\{Z_i(W_i - \beta^w_{-,0})^\top \,\middle|\, D_i\right\}\right\} \\
&= \frac{1}{h_n}\int_{-\infty}^{d^*} f_d(d)\mathbb{1}(d < d^*)K\left(\frac{|d - d^*|}{h_n}\right)\mathbb{E}\left\{Z(W - \beta^w_{-,0})^\top \,\middle|\, D = d\right\}\,\mathrm{d}d \\
&= \frac{1}{h_n}\int_{-\infty}^{0} f_d(d^* + uh_n)K(-u)\mathbb{E}\left\{Z(W - \beta^w_{-,0})^\top \,\middle|\, D = d^* + uh_n\right\}h_n\,\mathrm{d}u \\
&= \int_{-\infty}^{0} f_d(d^* + uh_n)K(-u)\mathbb{E}\left\{Z(W - \beta^w_{-,0})^\top \,\middle|\, D = d^* + uh_n\right\}\,\mathrm{d}u \\
&= \kappa \lim_{\epsilon\downarrow 0}\mathrm{Cov}\big(Z, W^\top \mid D = d^* - \epsilon\big) + o_p(1).
\end{aligned}
$$

The last line uses the dominated convergence theorem and the boundedness of $\mathbb{E}\{Z_i W_i^\top \mid D_i = \cdot\}$ and $\mu_{-,z}(\cdot)$ from Assumption 7, taking the limit as $h_n \to 0$. Specifically, we argue that

$$
\begin{aligned}
&\lim_{h_n\downarrow 0}\int_{-\infty}^{0} f_d(d^* + uh_n)K(-u)\mathbb{E}\left\{Z(W - \beta^w_{-,0})^\top \,\middle|\, D = d^* + uh_n\right\}\,\mathrm{d}u \\
&= \int_{-\infty}^{0} f_d(d^*)K(-u)\lim_{\epsilon\downarrow 0}\mathbb{E}\left\{Z(W - \beta^w_{-,0})^\top \,\middle|\, D = d^* - \epsilon\right\}\,\mathrm{d}u \\
&= f_d(d^*)\lim_{\epsilon\downarrow 0}\mathbb{E}\left\{Z(W - \beta^w_{-,0})^\top \,\middle|\, D = d^* - \epsilon\right\}\int_{-\infty}^{0} K(-u)\,\mathrm{d}u \\
&= \kappa \lim_{\epsilon\downarrow 0}\mathrm{Cov}\big(Z, W^\top \mid D = d^* - \epsilon\big).
\end{aligned}
$$

4. Consider the second term in the decomposition. We will show that

$$
\frac{1}{n}\sum_{i=1}^{n}\omega_{i,-}Z_i\left\{\left(\frac{D_i - d^*}{h_n}\right)\mathbf{e}_1^\top(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\mathbf{W}\right\} \xrightarrow{p} 0.
$$

By step 2 and the weak law of large numbers,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\omega_{i,-}Z_i\left\{\left(\frac{D_i - d^*}{h_n}\right)\mathbf{e}_1^\top(\mathbf{R}_1^\top\mathbf{K}_-\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{K}_-\mathbf{W}\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\omega_{i,-}Z_i(D_i - d^*)(\beta^w_{-,1})^\top + o_p(1) \\
&= \frac{1}{nh_n}\sum_{i=1}^{n}\mathbb{1}(D_i < d^*)K\left(\frac{|D_i - d^*|}{h_n}\right)Z_i(D_i - d^*)(\beta^w_{-,1})^\top + o_p(1) \\
&= \frac{1}{h_n}\mathbb{E}\left\{\mathbb{1}(D_i < d^*)K\left(\frac{|D_i - d^*|}{h_n}\right)Z_i(D_i - d^*)(\beta^w_{-,1})^\top\right\} + o_p(1).
\end{aligned}
$$

We use the law of iterated expectations and u-substitution with $u = \frac{d-d^*}{h_n}$ to express the former term as

$$
\begin{aligned}
&\frac{1}{h_n}\mathbb{E}\left\{\mathbb{1}(D_i < d^*)K\left(\frac{|D_i - d^*|}{h_n}\right)\mathbb{E}\{Z_i \mid D_i\}(D_i - d^*)\right\}(\beta^w_{-,1})^\top \\
&= \frac{1}{h_n}\int_{-\infty}^{d^*} f_d(d)\mathbb{1}(d < d^*)K\left(\frac{|d - d^*|}{h_n}\right)\mathbb{E}\{Z \mid D = d\}(d - d^*)\ \mathrm{d}d\,(\beta^w_{-,1})^\top \\
&= \frac{1}{h_n}\int_{-\infty}^{0} f_d(d^* + uh_n)K(-u)\,\mathbb{E}\{Z \mid D = d^* + uh_n\}\,uh_n\,h_n\ \mathrm{d}u\,(\beta^w_{-,1})^\top \\
&= h_n\int_{-\infty}^{0} f_d(d^* + uh_n)K(-u)\,\mathbb{E}\{Z \mid D = d^* + uh_n\}\,u\ \ \mathrm{d}u\,(\beta^w_{-,1})^\top \\
&= o_p(1).
\end{aligned}
$$

The last line uses the dominated convergence theorem and the boundedness of $\mu_{-,z}(\cdot)$ from Assumption 7, taking the limit as $h_n \to 0$. Specifically, we argue that

$$
\begin{aligned}
\lim_{h_n\downarrow 0}\int_{-\infty}^{0} f_d(d^* + uh_n)K(-u)\,\mathbb{E}\{Z \mid D = d^* + uh_n\}\,u\ \ \mathrm{d}u &= \int_{-\infty}^{0} f_d(d^*)K(-u)\,\mu_{-,z}(0)\,u\ \ \mathrm{d}u \\
&= f_d(d^*)\mu_{-,z}(0)\int_{-\infty}^{0} K(-u)\ u\ \ \mathrm{d}u \\
&= \kappa\mu_{-,z}(0).
\end{aligned}
$$

□

F Bias corrected inference proof

F.1 Notation

Let $\mathbf{e}_2 = (0, 0, 1)^\top$. Furthermore, define,

$$
\begin{aligned}
\mathbf{s} &= (1, \gamma_-^\top)^\top \otimes \mathbf{e}_0 = (1, 0, \gamma_{-,1}, 0, ..., \gamma_{-,q}, 0)^\top \in \mathbb{R}^{2(q+1)} \\
\hat{\mathbf{s}} &= (1, \hat{\gamma}_-^\top)^\top \otimes \mathbf{e}_0 = (1, 0, \hat{\gamma}_{-,1}, 0, ..., \hat{\gamma}_{-,q}, 0)^\top \in \mathbb{R}^{2(q+1)} \\
\mu_{+,s}(0) &= (\mu_{+,y}(0), \mu_{+,w_1}(0), \ldots, \mu_{+,w_q}(0))^\top \in \mathbb{R}^{q+1} \\
S_i &= (Y_i, W_i^\top)^\top \in \mathbb{R}^{q+1} \\
\mathbf{D} &= (D_1, \ldots, D_n)^\top \in \mathbb{R}^n
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{S} &= (S_1, \ldots, S_n)^\top \in \mathbb{R}^{n\times(q+1)} \\
\hat{\Gamma}_{+,1} &= \frac{1}{nh_n}\mathbf{R}_1^\top\mathbf{K}_+\mathbf{R}_1 \in \mathbb{R}^{2\times 2} \\
\hat{\Gamma}_{+,2} &= \frac{1}{nb_n}\mathbf{R}_2^\top\mathbf{K}_+\mathbf{R}_2 \in \mathbb{R}^{3\times 3} \\
\hat{\Lambda}_+ &= \frac{1}{nh_n}\mathbf{R}_1^\top\mathbf{K}_+(\mathbf{D} - d^*)^2 h_n^{-2} \in \mathbb{R}^2.
\end{aligned}
$$

For some vector $\mathbf{v}$, we use the convention $\mathbf{v}^2 = (v_1^2, \ldots, v_n^2)^\top$ to denote element-wise squaring. Here, $\otimes$ denotes the Kronecker product, and $\mathrm{vec}(\cdot)$ to define the operator which stacks columns of a matrix from left to right, i.e.,

$$
\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \otimes B = \begin{bmatrix} a_1 B & a_2 B \\ a_3 B & a_4 B \end{bmatrix}, \qquad \mathrm{vec}\Big(\vec{a}_1 \quad \vec{a}_2 \quad \ldots \quad \vec{a}_m\Big) = \begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_m \end{pmatrix}.
$$

Additionally, let $\hat{\beta}_+^s = [\hat{\beta}_+^y, \hat{\beta}_+^w] \in \mathbb{R}^{2\times(q+1)}$, where $\hat{\beta}_+^w = [\hat{\beta}_+^{w_1}, ..., \hat{\beta}_+^{w_q}] \in \mathbb{R}^{2\times q}$. Finally, we let $\mu_{+,s}^{(2)}(d) = \frac{\partial^2}{\partial d^2}\mu_{+,s}(d)$.

F.2 Estimator recap

By Proposition 1, we can express our bias-uncorrected estimator in a convenient way for analysis:

$$\hat{\tau}_{\mathrm{pdd}} = \mathbf{e}_0^\top \left\{ H_1\hat{\beta}_+^y - H_1\hat{\beta}_-^y + \left(H_1\hat{\beta}_+^w - H_1\hat{\beta}_-^w \right)\hat{\gamma}_- \right\} = \hat{\mathbf{s}}^\top \mathrm{vec}\left(H_1\hat{\beta}_+^s - H_1\hat{\beta}_-^s \right), \tag{4}$$

which multiplicatively separates the weights $\hat{\gamma}_-$ from the familiar discontinuity estimators. The second equality holds because

$$\begin{aligned}
\mathbf{e}_0^\top (H_1\hat{\beta}_+^y + H_1\hat{\beta}_+^w\hat{\gamma}_-) &= \mathbf{e}_0^\top H_1\hat{\beta}_+^y + \mathbf{e}_0^\top H_1\hat{\beta}_+^w\hat{\gamma}_- \\
&= 1 \cdot \mathbf{e}_0^\top H_1\hat{\beta}_+^y + \hat{\gamma}_{-,1} \cdot \mathbf{e}_0^\top H_1\hat{\beta}_+^{w_1} + \ldots + \hat{\gamma}_{-,q} \cdot \mathbf{e}_0^\top H_1\hat{\beta}_+^{w_q} \\
&= (1,0)^\top H_1\hat{\beta}_+^y + (\hat{\gamma}_{-,1},0)^\top H_1\hat{\beta}_+^{w_1} + \ldots + (\hat{\gamma}_{-,q},0)^\top H_1\hat{\beta}_+^{w_q} \\
&= \hat{\mathbf{s}}^\top \mathrm{vec}\left(H_1\hat{\beta}_+^s \right).
\end{aligned}$$

In line with Calonico et al. (2014), we choose to study the asymptotic normality of $\hat{\beta}_+^s - \hat{\beta}_-^s$. From the proof Proposition 2 (estimand), $\hat{\mathbf{s}} \xrightarrow{p} \mathbf{s}$. Slutsky's theorem then gives asymptotic normality for $\hat{\tau}_{\mathrm{pdd}}$.

The vectorization stacks several local linear regressions. For example, a component regression is

$$H_1\hat{\beta}_+^y = (\mathbf{R}_1^\top \mathbf{K}_+ \mathbf{R}_1)^{-1}\mathbf{R}_1^\top \mathbf{K}_+ \mathbf{Y} = \frac{1}{nh_n}\hat{\Gamma}_{+,1}^{-1}\mathbf{R}_1^\top \mathbf{K}_+ \mathbf{Y}.$$

Therefore, we can express the vectorization as the expanded regression

$$\mathrm{vec}(H_1\hat{\beta}_+^s) = \frac{1}{nh_n}\left(\mathbf{I}_{q+1} \otimes \hat{\Gamma}_{+,1}^{-1}\mathbf{R}_1^\top \mathbf{K}_+ \right)\mathrm{vec}(\mathbf{S}). \tag{5}$$

F.3 Conditional bias

Under Assumptions 7 (regularity conditions) and 8 (kernel), Lemma SA-8 of Calonico et al. (2019) implies that, when $nh \to \infty$ and $h \to 0$, we have

$$\begin{aligned}
\mathbb{E}\left\{ \mathrm{vec}\left(H_1\hat{\beta}_+^s \right) \mid \mathbf{D} \right\} &= \mathrm{vec}(H_1\beta_+^s) + \frac{1}{2}h_n^2\left(\mathbf{I}_{q+1} \otimes \hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+ \right)\mu_{+,s}^{(2)}(0) + O_p\left(h_n^3 \right) \\
&= \mathrm{vec}(H_1\beta_+^s) + h_n^2 B + O_p\left(h_n^3 \right),
\end{aligned}$$

where we define $B = \frac{1}{2}\left(\mathbf{I}_{q+1} \otimes \hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+ \right)\mu_{+,s}^{(2)}(0)$. Therefore, the conditional bias of $\mathrm{vec}\left(H_1\hat{\beta}_+^s \right)$ is given by $h_n^2 B + O_p(h_n^3)$.

We apply the 1-dimensional bias correction for $\hat{\beta}_+^y$ and $\hat{\beta}_+^{w_j}$, from Calonico et al. (2014). We estimate the $\hat{\mu}_{+,s}^{(2)}(0)$ using a local quadratic regression with bandwidth b_n, giving,

$$\begin{aligned}
\hat{B} &= \hat{B}_+ - \hat{B}_- \\
\hat{B}_+ &= \frac{1}{2}\left(\mathbf{I}_{q+1} \otimes \hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+ \right)\hat{\mu}_{+,s}^{(2)}(0) \\
\hat{\mu}_{+,s}^{(2)}(0) &= \frac{2}{nb_n^3}\left(\mathbf{I}_{q+1} \otimes \mathbf{e}_2^\top \hat{\Gamma}_{+,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_+ \right)\mathrm{vec}(\mathbf{S}),
\end{aligned}$$

with analogous objects below the cutoff.

In particular, using the mixed-product property of the Kronecker product, i.e. that $(A \otimes B)(C \otimes D) =$

$(AC) \otimes (BD)$, we write the leading conditional bias term as

$$
\begin{aligned}
h_n^2 \hat{B}_+ &= \frac{h_n^2}{nb_n^3}\Big(\mathbf{I}_{q+1} \otimes \hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+\Big)\Big(\mathbf{I}_{q+1} \otimes \mathbf{e}_2^\top \hat{\Gamma}_{+,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_+\Big)\mathrm{vec}(\mathbf{S}) \\
&= \frac{h_n^2}{nb_n^3}\Big(\mathbf{I}_{q+1} \otimes \hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+ \mathbf{e}_2^\top \hat{\Gamma}_{+,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_+\Big)\mathrm{vec}(\mathbf{S}).
\end{aligned}
\tag{6}
$$

The bias-corrected local instrumental variable estimator is then given by,

$$
\begin{aligned}
\hat{\tau}_{\mathrm{pdd}}^{\mathrm{bc}} &= \hat{\tau}_{\mathrm{rdd}}^{y,\mathrm{bc}} - (\hat{\tau}_{\mathrm{rdd}}^{w,\mathrm{bc}})^\top \hat{\gamma}_- \\
&= \hat{\mathbf{s}}^\top \{\mathrm{vec}(H_1 \hat{\beta}_+^s - H_1 \hat{\beta}_-^s) - h_n^2 \hat{B}\} \\
&= \hat{\mathbf{s}}^\top \{\mathrm{vec}(H_1 \hat{\beta}_+^s) - h_n^2 \hat{B}_+\} \\
&\quad - \hat{\mathbf{s}}^\top \{\mathrm{vec}(H_1 \hat{\beta}_-^s) - h_n^2 \hat{B}_-\} \\
&= \hat{\mathbf{s}}^\top \left[\mathbf{I}_{q+1} \otimes \left\{\frac{1}{nh_n}\hat{\Gamma}_{+,1}^{-1}\mathbf{R}_1^\top \mathbf{K}_+ - \frac{h_n^2}{nb_n^3}\hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+ \mathbf{e}_2^\top \hat{\Gamma}_{+,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_+\right\}\right]\mathrm{vec}(\mathbf{S}) \\
&\quad - \hat{\mathbf{s}}^\top \left[\mathbf{I}_{q+1} \otimes \left\{\frac{1}{nh_n}\hat{\Gamma}_{-,1}^{-1}\mathbf{R}_1^\top \mathbf{K}_- - \frac{h_n^2}{nb_n^3}\hat{\Gamma}_{-,1}^{-1}\hat{\Lambda}_- \mathbf{e}_2^\top \hat{\Gamma}_{-,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_-\right\}\right]\mathrm{vec}(\mathbf{S}) \\
&= \frac{1}{nh_n}\hat{\mathbf{s}}^\top \big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{\mathrm{bc}}\big)\mathrm{vec}(\mathbf{S}).
\end{aligned}
\tag{7}
$$

The first equality is how we define the estimator, the second equality uses equation (4), the third equality is how we define the bias correction, and the fourth equality uses equations (5) and (6). In the last line we define the empirical matrix

$$
\begin{aligned}
\mathbf{P}_{\mathrm{bc}} &= \mathbf{P}_{+,\mathrm{bc}} - \mathbf{P}_{-,\mathrm{bc}} \\
\mathbf{P}_{+,\mathrm{bc}} &= \hat{\Gamma}_{+,1}^{-1}\mathbf{R}_1^\top \mathbf{K}_+ - \left(\frac{h_n}{b_n}\right)^3 \hat{\Gamma}_{+,1}^{-1}\hat{\Lambda}_+ \mathbf{e}_2^\top \hat{\Gamma}_{+,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_+ \\
\mathbf{P}_{-,\mathrm{bc}} &= \hat{\Gamma}_{-,1}^{-1}\mathbf{R}_1^\top \mathbf{K}_- - \left(\frac{h_n}{b_n}\right)^3 \hat{\Gamma}_{-,1}^{-1}\hat{\Lambda}_- \mathbf{e}_2^\top \hat{\Gamma}_{-,2}^{-1}\mathbf{R}_2^\top \mathbf{K}_-.
\end{aligned}
$$

F.4 Conditional variance

We define the following diagonal matrices in $\mathbb{R}^{n(q+1)\times n(q+1)}$:

$$
\begin{aligned}
\Sigma_+ &= \mathrm{Diag}\Big(\sigma_{+,y}^2(D_1 - d^*), \ldots, \sigma_{+,y}^2(D_n - d^*); \sigma_{+,w_1}^2(D_1 - d^*), \ldots, \sigma_{+,w_1}^2(D_n - d^*); \ldots; \sigma_{+,w_q}^2(D_1 - d^*), \cdots, \sigma_{+,w_q}^2(D_n - d^*)\Big) \\
\hat{\Sigma}_+ &= \mathrm{Diag}\Big(\hat{\sigma}_{+,y}^2(D_1 - d^*), \ldots, \hat{\sigma}_{+,y}^2(D_n - d^*); \hat{\sigma}_{+,w_1}^2(D_1 - d^*), \ldots, \hat{\sigma}_{+,w_1}^2(D_n - d^*); \ldots; \hat{\sigma}_{+,w_q}^2(D_1 - d^*), \cdots, \hat{\sigma}_{+,w_q}^2(D_n - d^*)\Big)
\end{aligned}
$$

where $\sigma_{+,y}^2(D_i - d^*) = \mathrm{Var}(Y \mid D = D_i)$ and $\hat{\sigma}_{+,y}^2(D_i - d^*) = (Y_i - \hat{\beta}_{+,0}^{y,\mathrm{bc}})^2$.

In this notation, we write the conditional variance as

$$
\begin{aligned}
\mathrm{Var}\Big(\sqrt{nh_n}\hat{\tau}_{\mathrm{pdd}}^{\mathrm{bc}} \mid \mathbf{D}\Big) &= nh_n \mathrm{Var}\Big(\hat{\tau}_{\mathrm{pdd}}^{\mathrm{bc}} \mid \mathbf{D}\Big) \\
&= \frac{1}{nh_n}\hat{\mathbf{s}}^\top \big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{\mathrm{bc}}\big)\mathrm{Var}(\mathrm{vec}(\mathbf{S}) \mid \mathbf{D})\big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{\mathrm{bc}}\big)^\top \hat{\mathbf{s}} \\
&= \frac{1}{nh_n}\hat{\mathbf{s}}^\top \big\{\mathbf{I}_{q+1} \otimes (\mathbf{P}_{+,\mathrm{bc}} - \mathbf{P}_{-,\mathrm{bc}})\big\}\mathrm{Var}(\mathrm{vec}(\mathbf{S}) \mid \mathbf{D})\big\{\mathbf{I}_{q+1} \otimes (\mathbf{P}_{+,\mathrm{bc}} - \mathbf{P}_{-,\mathrm{bc}})\big\}^\top \hat{\mathbf{s}} \\
&= \frac{1}{nh_n}\hat{\mathbf{s}}^\top \big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{+,\mathrm{bc}}\big)\Sigma_+ \big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{+,\mathrm{bc}}\big)^\top \hat{\mathbf{s}} \\
&\quad + \frac{1}{nh_n}\hat{\mathbf{s}}^\top \big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{-,\mathrm{bc}}\big)\Sigma_- \big(\mathbf{I}_{q+1} \otimes \mathbf{P}_{-,\mathrm{bc}}\big)^\top \hat{\mathbf{s}}.
\end{aligned}
$$

The first equality uses the laws of variance, the second equality uses equation (7), the third equality uses a definition, and the fourth equality recognizes that cross terms in the quadratic form are zero by definition of the kernel weights above and below the threshold.

We estimate the variance using the empirical analogue, i.e.

$$\begin{aligned}\hat{V}_{\mathrm{bc}} = &\frac{1}{nh_n}\hat{\mathbf{s}}^\top\left(\mathbf{I}_{q+1}\otimes\mathbf{P}_{+,\mathrm{bc}}\right)\hat{\Sigma}_+\left(\mathbf{I}_{q+1}\otimes\mathbf{P}_{+,\mathrm{bc}}\right)^\top\hat{\mathbf{s}}\\ &+\frac{1}{nh_n}\hat{\mathbf{s}}^\top\left(\mathbf{I}_{q+1}\otimes\mathbf{P}_{-,\mathrm{bc}}\right)\hat{\Sigma}_-\left(\mathbf{I}_{q+1}\otimes\mathbf{P}_{-,\mathrm{bc}}\right)^\top\hat{\mathbf{s}}.\end{aligned}$$

F.5 Main result

Proof of Theorem 2 (limit distribution).
Recall that we showed in the proof of Proposition 2 that $\hat{\gamma}_- \xrightarrow{p} \gamma_-$, so therefore $\hat{\mathbf{s}} \xrightarrow{p} \mathbf{s}$.

Moreover, in this appendix we have provided new characterizations of the conditional bias and variance for the local instrumental variable regression estimator proposed in this work.

Apart from these differences, $\hat{\tau}^{\mathrm{bc}}_{\mathrm{pdd}}$ has the same overall form as the covariate-adjusted sharp RD estimator from Calonico et al. (2019). Therefore, the inference result follows by combining our characterizations together with Lemma SA-11 of Calonico et al. (2019). □