SUPPLEMENT TO "TESTING IDENTIFYING ASSUMPTIONS IN FUZZY REGRESSION DISCONTINUITY DESIGNS"

YOICHI ARAI[a] YU-CHIN HSU[b] TORU KITAGAWA[c] ISMAEL MOURIFIÉ [d] YUANYUAN WAN[e]

We describe how our test differs and complements existing tests in Appendix B. We discuss several extensions in Appendix C. We formally state the asymptotic validity of our test in Appendix D. All proofs are collected in Appendix E. Additional empirical results of Section 5 are provided in Appendix G.

Appendix B. Comparison between our approach and the existing approach

We provide a detailed analytical discussion of how our testing approach differs from and complements the existing approach in terms of assessing the local continuity (LC) assumption.

Let $X \in \mathcal{X} \subset \mathbb{R}^{d_x}$ be observable covariates. Assuming that all the probability densities in the following equations are well defined, we can write

$$f_{Y_d(r),T_{|r-r_0|}|R,X}(y,t|r,x) = \frac{f_{R|Y_d(r),T_{|r-r_0|},X}(r|y,t,x)}{f_{X|R}(x|r)f_R(r)} f_{Y_d(r),T_{|r-r_0|},X}(y,t,x), \tag{B.1}$$

where $f_{Y_d(r),T_{|r-r_0|}|R,X}(y,t|r,x)$ denotes the conditional density of $(Y_d(r), T_{|r-r_0|})$ given R, X. On the right-hand side of the equation, the continuity of $f_{R|Y_d(r),T_{|r-r_0|},X}(r|y,t,x)$ in r near r_0 is essentially Lee (2008)'s *stronger local continuity* (SLC) assumption (with different notation), which was originally introduced in the sharp RD framework and later discussed in Dong (2018) in the context of the FRD setting. Since the SLC assumption is not directly testable, the existing literature has derived

Date: Saturday 20th March, 2021.

a. School of Social Sciences, Waseda University, yarai@waseda.jp.
b. Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University, ychsu@econ.sinica.edu.tw.
c. Department of Economics, University College London, t.kitagawa@ucl.ac.uk.
d. Department of Economics, University of Toronto, ismael.mourifie@utoronto.ca.
e. Department of Economics, University of Toronto, yuanyuan.wan@utoronto.ca.

Acknowledgement: Financial support from the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001), the ERC through the ERC starting grant (grant number EPP-715940), the Japan Society for the Promotion of Science through the Grants-in-Aid for Scientific Research No. 15H03334, Ministry of Science and Technology of Taiwan (MOST107-2410-H-001-034-MY3), Academia Sinica Taiwan through the Career Development Award, and Waseda University Grant for Special Research Projects is gratefully acknowledged.

tests for two of its implications: (i) the continuity of $f_R(r)$, see for instance McCrary (2008), Otsu, Xu, and Matsushita (2013), Cattaneo, Jansson, and Ma (2020), and Bugni and Canay (2018), and (ii) the continuity of $F_{X|R}(x|r)$ in r, see Canay and Kamat (2018). We can see from equation (B.1) that LC can be considered as an implication of SLC. It is, however, important to note that LC neither implies nor is implied by continuity of $f_R(r)$ and continuity of $F_{X|R}(x|r)$ in r. As a result, there are important empirical scenarios in which the conclusions of the existing tests are not necessarily informative about validity or invalidity of LC, as we illustrate below.

Scenario 1: The existing approach accepts continuity of f_R and $F_{X|R}$ while our approach refutes LC:

This scenario corresponds to the case that the existing approach of testing continuity of f_R and $F_{X|R}$ at $r = r_0$ overlooks the failure of FRD-design, while our approach detects it. To be specific, consider an empirical context in which multiple programs share the same running variable R and the common threshold r_0, e.g., a household can participate in two social programs and both of them use the same poverty index and poverty line to determine eligibility. Let $D, Z \in \{0, 1\}$ denote the treatment statuses, respectively. The researcher is interested in the causal effect of treatment in the first programme, i.e., the effect of D. For simplicity, let us assume that the assignment of the second program is sharp, $Z = \mathbf{1}[R \geq r_0]$. In such a context, the potential outcome model can be written as

$$Y = \underbrace{\{Y_{11}Z + Y_{10}(1-Z)\}}_{Y_1} D + \underbrace{\{Y_{01}Z + Y_{00}(1-Z)\}}_{Y_0}(1-D), \tag{B.2}$$

where $Y_{dz}(r)$, $d \in \{1, 0\}$ and $z \in \{1, 0\}$, are the potential outcomes indexed by the two treatments. As can be seen, if the researcher is unaware of the second treatment, the potential outcome Y_d that she specifies would be $Y_d = Y_{d1}Z + Y_{d0}(1-Z)$. Suppose now $f_{R|Y_{dz}(r),T_{|r-r_0|},X}(r|y,t,x)$ is continuous in r for any y, t, x, then $f_R(r)$ and $F_{X|R}(x|r)$ are continuous. However, the density $f_{R|Y_d(r),T_{|r-r_0|},X}(r|y,t,x)$ can be discontinuous if the second treatment Z affects the outcome. Specifically, since we have (for $d = 1$)

$$\begin{aligned}
&\lim_{r \downarrow r_0} f_{R|Y_{11}Z+Y_{10}(1-Z),T_{|r-r_0|},X}(r|y,t,x) - \lim_{r \uparrow r_0} f_{R|Y_{11}Z+Y_{10}(1-Z),T_{|r-r_0|},X}(r|y,t,x) \\
&= \lim_{r \downarrow r_0} f_{R|Y_{11},T_{|r-r_0|},X}(r|y,t,x) - \lim_{r \uparrow r_0} f_{R|Y_{10},T_{|r-r_0|},X}(r|y,t,x) \\
&= \left[\lim_{r \to r_0} \frac{f_{Y_{11},T_{|r-r_0|}|R,X}(y,t|r,x)}{f_{Y_{11},T_{|r-r_0|}|X}(y,t|x)} - \lim_{r \to r_0} \frac{f_{Y_{10},T_{|r-r_0|}|R,X}(y,t|r,x)}{f_{Y_{10},T_{|r-r_0|}|X}(y,t|x)} \right] f_{R|X}(r_0|x),
\end{aligned}$$

where the first equality follows since the assignment of Z is sharp, and the second equality follows from Bayes rule and continuity of $f_{R|Y_{dz}(r),T_{|r-r_0|},X}(r|y,t,x)$ in r. The two terms in the brackets do not have to cancel out as Y_{10} and Y_{11} are two different potential outcomes, with and without the second treatment. Therefore, in this scenario, learning continuity of $f_R(r)$ and $F_{X|R}(x|r)$ does not inform about failure of LC. Our approach, in contrast, can detect violation of LC, if the distributional differences between Y_{d0} and Y_{d1} at the cut-off leads to violation of the testable implications of Theorem 1 (i) in the main text.

For ease of exposition, we consider binary Z here and interpret it as an unobservable treatment sharing the running variable and cut-off. It is straightforward to generalize the current argument to cases where Z is nonbinary and its discontinuity at r_0 is in terms of its conditional distribution given r. We can also interpret Z as any unobservable factor affecting the outcome, whose distribution changes discontinuously at the cut-off. For instance, the existence of such a Z is often suspected when geographical boundaries are used for regression discontinuity.

Scenario 2: The existing approach rejects continuity of f_R and $F_{X|R}$ while FRD-validity holds (so that our approach does not refute):

This scenario corresponds to the case that FRD-validity holds but the existing approach finds discontinuity of f_R and $F_{X|R}$ at the cut-off. This can happen for a data generating process in which the discontinuity of either $f_R(r)$ or $f_{X|R}(x|r)$ (or both) is compensated exactly by the discontinuity of $f_{R|Y_1(r),T_{|r-r_0|},X}(r|y,t,x)$ in such a way that $f_{Y_d(r),T_{|r-r_0|}|R,X}(y,t|r,x)$ remains continuous. This scenario is not pathological, and is likely to happen in empirical applications where the manipulation is made independently of the potential outcomes. For instance, in the context of the empirical application relating to Maimonides's rule in Israel, Angrist, Lavy, Leder-Luis, and Shany (2019) argues that the presence of discontinuity in the running variable (school enrollment) is mainly due to a school board administration objective to increase their budgets and was "unrelated to socioeconomic characteristics conditional on a few controls" (please refer to Section 5.1 for detailed discussion). This narrative evidence justifies $f_{R|Y_1(r),T_{|r-r_0|},X}(r|y,t,x) = f_{R|X}(r|x)$ in some local neighborhood of the cut-off. This reduces equation (B.1) to

$$f_{Y_1(r),T_{|r-r_0|}|R,X}(y,t|r,x) = \frac{f_{Y_1(r),T_{|r-r_0|},X}(y,t,x)}{f_X(x)}$$

in the local neighborhood of the cut-off, implying $f_{Y_1(r),T_{|r-r_0|}|R,X}(y,t|r,x)$ is continuous at r_0. However, either $f_R(r)$ or $F_{X|R}(x|r)$ (or both) is discontinuous at r_0. This example illustrates that even when the running variable's density is discontinuous, FRD-validity can hold.

Appendix C. Extensions

In this section, we briefly discuss several extensions. First, we discuss the relationship between FRD-validity considered in our paper and other FRD-identifying assumptions considered in the literature. Second, we show how to incorporate conditioning covariates in our test.

C.1. Relationship with other FRD identifying assumptions.

In the LATE framework, de Chaisemartin (2017) argues that the Wald (IV) estimand can have a well-defined causal interpretation under a weaker version of instrument monotonicity. A parallel of his weaker monotonicity condition in the FRD setting can be written as follows: there exists $\epsilon > 0$ such that

$$P(T_{|r-r_0|} = \mathbf{DF}|Y_d(r) = y, R = r) \leq P(T_{|r-r_0|} = \mathbf{C}|Y_d(r) = y, R = r), \quad d \in \{0,1\}, \quad y \in \mathcal{Y}$$

for all $r \in (r_0 - \epsilon, r_0 + \epsilon)$. It can be shown that our Theorem 1 holds by replacing Assumption 1 with this weaker monotonicity assumption and modifying Assumption 2 to include $T = \mathbf{DF}$. That is, inequalities 2 and 3 remain unimprovable testable implications under this weaker version of the local monotonicity assumption.

Bertanha and Imbens (2020) consider an alternative local monotonicity assumption that is more restrictive than Assumption 1.

Assumption C.1 (Strong local monotonicity)**.** *There exists $\epsilon > 0$ such that any individual in the population is classified into one of the following three types based on their treatment selection responses:*

$$T = \begin{cases} \mathbf{A}, & \text{if } D(r) = 1, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{C}, & \text{if } D(r) = 1\{r \geq r_0\}, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{N}, & \text{if } D(r) = 0, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon). \end{cases} \tag{C.1}$$

This monotonicity implies that in some neighborhood of the cut-off, compliance status is invariant for any given individual. It can be shown that strengthening Assumption 1 to Assumption C.1 does not yield further testable implications beyond those of Theorem 1 (i), i.e., Theorem 1 (ii) holds true even if Assumption 1 in the main text is replaced by Assumption C.1 above.[1]

The literature has considered the *local independence* assumption,[2] which is a stronger form of identifying assumption than LC.

Assumption C.2 (Local independence)**.** *There exists $\epsilon > 0$ such that for $d = 0, 1$, $(Y_d(r), D(r))$ is jointly independent of R in the neighborhood $(r_0 - \epsilon, r_0 + \epsilon)$ and $\lim_{r \downarrow r_0} Y_d(r) = \lim_{r \uparrow r_0} Y_d(r) \equiv Y_d(r_0)$ a.s.*

The statement of Theorem 1 (i) indeed holds even if LC is replaced by this local independence assumption.

C.2. **Incorporating Covariates.** The standard FRD design does not require covariates to identify treatment effect at the cut-off, but they are often included in practice to increase efficiency. See Imbens and Kalyanaraman (2012), Calonico, Cattaneo, Farrel, and Titiunik (2019), and Hsu and Shen (2019). Another motivation for incorporating the conditioning covariates arises if the potential outcomes can depend on the covariates whose distribution is discontinuous at the cut-off. In this case, RD analysis without conditioning on covariates leads to violation of the local continuity assumption (Frölich and Huber (2019)).

In what follows, we consider testing a version of FRD-validity where local monotonicity and local continuity are imposed conditional on a covariate vector $X \in \mathcal{X} \subset \mathbb{R}^{d_x}$. We allow X to be discrete or continuous. We assume that there are observations near the cut-off point conditioning on each realization x. The conditional version of FRD-validity is stated formally as follows:

Assumption C.3. *The limits $\pi^+(x) \equiv \lim_{r \downarrow r_0} P(D = 1 | R = r, X = x)$ and $\pi^-(x) \equiv \lim_{r \uparrow r_0} P(D = 1 | R = r, X = x)$ exist and $\pi^+(x) \neq \pi^-(x)$ for all $x \in \mathcal{X}$.*

[1] Our proof of Theorem 1 (ii) in Appendix E constructs a distribution of potential outcomes and selection types that, in fact, satisfies Assumption C.1.

[2] This assumption is slightly weaker than the HTV local independence assumption, which involves a local exclusion restriction that rules out causal dependence of Y_d on R in the neighborhood. See Dong and Lewbel (2015), and Dong (2018) for discussion comparing local continuity and HTV local independence, and the restrictions that HTV local independence impose on the distribution of observables.

Assumption C.4 (Local continuity conditional on X)**.** *For $d = 0, 1$, $t \in \{\mathbf{A}, \mathbf{C}, \mathbf{N}\}$, and $B \subseteq \mathcal{Y}$ be a measurable set, the conditional probability $P(Y_d(r) \in B, T_{|r-r_0|} = t | R = r, X = x)$ is continuous in r at r_0, for all $x \in \mathcal{X}$.*

Assumption C.5 (Local monotonicity conditional on X)**.** *Let $t \in \{\mathbf{DF}, \mathbf{I}\}$. There exists a small $\epsilon > 0$ such that $P(T_{|r-r_0|} = t | R = r, X = x) = 0$ for all $r \in (r_0 - \epsilon, r_0 + \epsilon)$ and for all $x \in \mathcal{X}$.*

Theorem 1 can be immediately extended to the conditional version of FRD-validity by conditioning additionally on X. To fit into our testing framework, it is convenient to rewrite the moment inequalities conditional on X in terms of moment inequalities unconditional on X.

To do so, let $Z = (Y, X)$ and $\mathcal{Z}$ be the support of Z. We obtain the following inequalities as the testable implications for Assumptions C.3-C.5: for C a hypercube in $\mathcal{Z}$:

$$\lim_{r \uparrow r_0} \mathbb{E}_P[1\{Z \in C\} D | R = r] - \lim_{r \downarrow r_0} \mathbb{E}_P[1\{Z \in C\} D | R = r] \leq 0, \tag{C.2}$$

$$\lim_{r \downarrow r_0} \mathbb{E}_P[1\{Z \in C\}(1-D) | R = r] - \lim_{r \uparrow r_0} \mathbb{E}_P[1\{Z \in C\}(1-D) | R = r] \leq 0. \tag{C.3}$$

In comparison to inequalities (2) and (3), the only difference is that the indicator functions in (C.2) and (C.3) index boxes in $\mathcal{Z}$ instead of the intervals in $\mathcal{Y}$. Accordingly, by defining a class of instrument functions as

$$\mathcal{G}_z = \{g_\ell(\cdot) = 1(\cdot \in C_\ell) : \ell \equiv (z, c) \in \mathcal{L}\}, \text{ where}$$
$$C_\ell = \times_{j=1}^{d_x+1} [z_j, z_j + c] \cap \mathcal{Z} \text{ and}$$
$$\mathcal{L} = \left\{(z, c) : \ c^{-1} = q, \text{ and } q \cdot z_j \in \{0, 1, 2, \cdots, (q-1)\}^{d_x+1} \text{ for } q = 1, 2, \cdots \right\}, \tag{C.4}$$

we can implement the testing procedure shown in the main text to assess the conditional version of FRD-validity.

C.3. **Joint test.** Our test complements the widely used continuity tests for the distribution of conditioning covariates. Since continuity of the conditional distribution of some covariates at the cut-off is often considered to be supporting evidence for no-selection around the cut-off, it is worthwhile to combine our test with a continuity test.

Suppose we want to test the continuity of the distribution of covariates X at the cut-off *jointly* with the testable implications of (2) and (3). Since continuity of the distribution of X is expressed as a set of local moment *equalities*, we can obtain a joint test by adding the additional set of equality constraints to the null hypothesis.

We hence consider testing the inequalities of Theorem 1 (i) in the main text and the set of equalities indexed by $j \in \mathcal{J}$,

$$v^x(j) \equiv \lim_{r \uparrow r_0} \mathbb{E}_P[1\{X \in C_j^x\}|R = r] - \lim_{r \downarrow r_0} \mathbb{E}_P[1\{X \in C_j^x\}|R = r] = 0.$$

where C_j^x is a hypercube or a quadrant in the space of covariates X, and $\mathcal{J}$ forms a countable collection thereof.

We estimate $v^x(j)$ by $\hat{v}^x(j)$, the difference of two local linear estimators. Following how Andrews and Shi (2013) incorporate moment equalities, we modify the KS test statistic as

$$\hat{S}_n^{joint} = \max\left\{ \sup_{d \in \{0,1\},\ \ell \in \mathcal{L}} \frac{\sqrt{nh} \cdot \hat{v}_{n,d}(\ell)}{\hat{\sigma}_{n,d,\xi}(\ell)}, \quad \sup_{j \in \mathcal{J}} \frac{\sqrt{nh} \cdot |\hat{v}^x(j)|}{\hat{\sigma}^x_{n,\xi}(j)} \right\},$$

where $\hat{\sigma}^x_{n,\xi}(j)$ is an estimator for the asymptotic standard deviation of $\sqrt{nh}(\hat{v}^x(j) - v^x(j))$. Critical values for this test statistic can be obtained by a procedure similar to Algorithm 1 in the main text. Some differences are that for the moment equality constraints, we do not have the moment selection step, and the absolute values are taken for the estimators $\hat{v}^x(j)$ and their bootstrap analogues.

APPENDIX D. ASYMPTOTIC PROPERTIES OF THE PROPOSED TEST

In this appendix, we spell out the regularity conditions and state the theorems that guarantee the asymptotic validity of our test. Their proofs are given in Appendix E.3.

We normalize the support of the observed outcome Y to $[0,1]$.[3] Let $\mathcal{P}$ be the collection of probability distributions of observables (Y, D, R). We denote the Lebesgue density of the running variable, R, by f_R.

Let $h_2(\cdot,\cdot)$ be a covariance kernel on $\mathcal{L} \times \mathcal{L}$. Let $\mathcal{H}_2$ be the collection of all possible covariance kernel functions on $\mathcal{L} \times \mathcal{L}$. For any pair of $h_2^{(1)} \in \mathcal{H}_2$ and $h_2^{(2)} \in \mathcal{H}_2$, we define the distance between them as

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{\{\ell_1, \ell_2 \in \mathcal{L}\}} |h_2^{(1)}(\ell_1, \ell_2) - h_2^{(2)}(\ell_1, \ell_2)|. \tag{D.1}$$

[3]This support normalization is without loss of generality. If not, we can define $\widetilde{Y} = \Phi(Y)$ where $\Phi(\cdot)$ is the CDF of standard normal, as in the first step of Algorithm 1.

Let $\sigma_{P,d}(\ell_1,\ell_2|r) = Cov_P(g_{\ell_1}(Y)D^d(1-D)^{1-d}, g_{\ell_2}(Y)D^d(1-D)^{1-d}|R=r)$ for $d=1,0$. We denote their right and left limits at r_0 by $\sigma_{P,d,+}(\ell_1,\ell_2) = \lim_{r\downarrow r_0}\sigma_{P,d}(\ell_1,\ell_2|r)$ and $\sigma_{P,d,-}(\ell_1,\ell_2) = \lim_{r\uparrow r_0}\sigma_{P,d}(\ell_1,\ell_2|r)$, respectively. Existence of these limits is implied by the set of assumptions in Assumption D.1, below.

For $j=0,1,2,\ldots$, let $\vartheta_j = \int_0^\infty u^j K(u)du$. Let $f_R^+(r_0) = \lim_{r\downarrow r_0} f_R(r_0)$ and $f_R^-(r_0) = \lim_{r\uparrow r_0} f_R(r_0)$. For $d=0,1$ and $\star = +,-$, define

$$h_{2,P,d,\star}(\ell_1,\ell_2) = \frac{\int_0^\infty(\vartheta_2 - u\vartheta_1)^2K^2(u)du\cdot\sigma_{P,d,\star}(\ell_1,\ell_2)}{c_\star f_R^\star(r_0)(\vartheta_2\vartheta_0-\vartheta_1^2)^2}, \tag{D.2}$$

which is the covariance kernel of the limiting process of $\sqrt{nh}(\hat{m}_{d,\star}(\ell) - m_{P,d,\star}(\ell))$, with undersmoothing bandwidths. It can be shown that the covariance kernel of the limiting processes of $\sqrt{nh}(\hat{\nu}_d(\ell) - \nu_{P,d}(\ell))$ is $h_{2,P,d} = h_{2,P,d,+} + h_{2,P,d,-}$.

We denote the v-th derivative of $m_{P,d}(\ell,r) = \mathbb{E}_P[g_\ell(Y)D^d(1-D)^{1-d}|R=r]$ with respect to the running variable by $m_{P,d}^{(v)}(\ell,r)$, $d=1,0$. The v-th derivative of f_R is denoted similarly. For $\delta>0$, define $\mathcal{N}_\delta(r_0) = \{r: |r-r_0|<\delta\}$ as a neighborhood of r around r_0. Let $\mathcal{N}_\delta^+(r_0) = \{r: 0<r-r_0<\delta\}$ and $\mathcal{N}_\delta^-(r_0) = \{r: 0<r_0-r<\delta\}$ be one-sided open neighborhoods excluding r_0.

Assumption D.1. *Let f_R be common for all $P\in\mathcal{P}$. There exist $\delta>0$, $\epsilon>0$, $0<\bar{f}_R<\infty$, and $0\le M<\infty$ such that for all $P\in\mathcal{P}$,*

(i) $f_R(r)>\epsilon$ *on* $\mathcal{N}_\delta(r_0)$.

(ii) $f_R(r)$ *is continuous and bounded from above by* $\bar{f}_R$ *on* $\mathcal{N}_\delta^+(r_0)\cup\mathcal{N}_\delta^-(r_0)$, *and* $f_R^+(r_0)$ *and* $f_R^-(r_0)$ *exist.*

(iii) $f_R(r)$ *is twice continuously differentiable in* r *on* $\mathcal{N}_\delta^+(r_0)\cup\mathcal{N}_\delta^-(r_0)$ *and* $|f_R^{(1)}(r)|\le M$ *and* $|f_R^{(2)}(r)|\le M$ *on* $\mathcal{N}_\delta^+(r_0)\cup\mathcal{N}_\delta^-(r_0)$;

(iv) for $d=0,1$ *and for all* $\ell\in\mathcal{L}$, $m_{P,d}(\ell,r)$ *is twice continuously differentiable in* r *on* $\mathcal{N}_\delta^+(r_0)\cup\mathcal{N}_\delta^-(r_0)$;

(v) for $d=0,1$ *and for all* $\ell\in\mathcal{L}$, $|m_{P,d}^{(1)}(\ell,r)|\le M$ *and* $|m_{P,d}^{(2)}(\ell,r)|\le M$ *on* $\mathcal{N}_\delta^+(r_0)\cup\mathcal{N}_\delta^-(r_0)$;

Assumption D.1 (*iii*)-(*v*) imply that with undersmoothing bandwidths, the bias terms of $\hat{\nu}_1(\ell)$ and $\hat{\nu}_0(\ell)$ are asymptotically negligible uniformly over $\ell\in\mathcal{L}$. Note that Assumption D.1 does not restrict the support of Y and allows Y to be discrete, continuous, or some mixture of the two. Note also that we allow $f_R(r)$ to be discontinuous at the cut-off, reflecting the fact that the testable implications of FRD-validity that we are focusing on do not require continuity of $f_R(r)$ at the cut-off.

Assumption D.2. *The kernel function $K(\cdot)$ and bandwidth h satisfy*

(i) $K(\cdot)$ *is nonnegative, symmetric, bounded by* $\bar{K} < \infty$*, and has a compact support (say* $[-1, 1]$*),*

(ii) $\int_{\mathbb{R}} K(u)du = 1$*, and* $\int_{\mathbb{R}} u^2 K(u)du > 0$*,*

(iii) $h \to 0$*,* $nh \to \infty$ *and* $nh^5 \to 0$ *as* $n \to \infty$*.*

Assumption D.2 is standard, and the triangular kernel used in our Monte Carlo studies and empirical applications satisfies this assumption. Note that $nh^5 \to 0$ as $n \to \infty$ corresponds to an undersmoothing choice of bandwidth so that the bias term of $\hat{v}_{n,d}$ converges to zero even after $\sqrt{nh}$ is multiplied.

Assumption D.3. *Let* $\{U_i : 1 \leq i \leq n\}$ *be a sequence of i.i.d. random variables* $E[U] = 0$*,* $E[U^2] = 1$*, and* $E[|U|^4] < M_1$ *for some* $M_1 < \infty$*, and* $\{U_i : 1 \leq i \leq n\}$ *is independent of the sample* $\{(Y_i, D_i, R_i) : 1 \leq i \leq n\}$*.*

Assumption D.3 is standard for the multiplier bootstrap (see, e.g., Hsu (2016)). Note the standard normal distribution for U satisfies Assumption D.3.

Assumption D.4. $\{a_n\}$ *is a sequence of nonnegative numbers satisfying* $\lim_{n\to\infty} a_n = \infty$ *and* $\lim_{n\to\infty} a_n/\sqrt{nh} = 0$*.* $\{B_n\}$ *is a sequence of nonnegative numbers that is nondecreasing,* $\lim_{n\to\infty} B_n = \infty$ *and* $\lim_{n\to\infty} B_n/a_n = 0$*.*

In our Monte Carlo study and empirical applications, we specify $a_n = (0.3\ln(n))^{1/2}$ and $B_n = (0.4\ln(n)/\ln\ln(n))^{1/2}$ following Andrews and Shi (2013, 2014).

Let $\mathcal{P}_0 \subset \mathcal{P}$ be the set of distributions of observables that satisfy the null hypothesis given in equation (5) in the main text. The next assumption states that $\mathcal{P}_0$ contains a distribution of data that satisfies the moment inequalities $\{v_{P,d}(\ell) : d = 0, 1, \ell \in \mathcal{L}\}$ with equality for some $\ell \in \mathcal{L}$.

Assumption D.5. *Let* $\mathcal{L}^o_{P,d} \equiv \{\ell \in \mathcal{L} : v_{P,d}(\ell) = 0\}$*. There exists* $P_c \in \mathcal{P}_0$ *such that* *(i) Either $\mathcal{L}^o_{P_c,1}$ or $\mathcal{L}^o_{P_c,0}$ under P_c is nonempty.*

(ii) For $d = 0, 1$, $h_{2,P_c,d,+} \in \mathcal{H}_{2,cpt}$ and $h_{2,P_c,d,-} \in \mathcal{H}_{2,cpt}$, where $\mathcal{H}_{2,cpt}$ is a compact subset of $\mathcal{H}_2$ with respect to the norm defined in equation (D.1).

(iii) Either $h_{2,P_c,1} = h_{2,P_c,1,+} + h_{2,P_c,1,-}$ restricted to $\mathcal{L}^o_{P_c,1} \times \mathcal{L}^o_{P_c,1}$ is not a zero function or $h_{2,P_c,0} = h_{2,P_c,0,+} + h_{2,P_c,0,-}$ restricted to $\mathcal{L}^o_{P_c,0} \times \mathcal{L}^o_{P_c,0}$ is not a zero function.

Theorem D.1. *Suppose Assumptions D.1-D.4 hold. Then, for every compact subset $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$, the following claims hold for the testing procedure presented in Algorithm 1:*

(a) $\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ d\in\{0,1\},h_{2,P,d,+},h_{2,P,d,-}\in\mathcal{H}_{2,cpt}\}} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) \leq \alpha.$

(b) If Assumption D.5 also holds, then

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ d\in\{0,1\},h_{2,P,d,+},h_{2,P,d,-}\in\mathcal{H}_{2,cpt}\}} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) = \alpha.$$

Theorem D.1 (a) shows that our test has asymptotically uniformly correct size over a compact set of covariance kernels. Theorem D.1 (b) shows that our test is at most infinitesimally conservative asymptotically when the null contains at least one P_c defined in Assumption D.5. Theorem D.1 extends Theorem 2 of Andrews and Shi (2013) and Theorem 5.1 of Hsu (2017) to local moment inequalities in the context of RD designs.

The next theorem shows consistency of our test against a fixed alternative.

Theorem D.2. *Suppose Assumptions D.1-D.4 hold and $\alpha < 1/2$. If there exists $\ell \in \mathcal{L}$ such that either $\nu_{P_1,1}(\ell) > 0$ or $\nu_{P_1,0}(\ell) > 0$, then $\lim_{n\to\infty} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) = 1$.*

We can also show that our test is unbiased against some $\sqrt{nh}$-local alternatives. We consider a sequence of $P_n \in \mathcal{P}\backslash\mathcal{P}_0$ such that

$$\nu_{P_n,d}(\ell) = \nu_{P_c,d}(\ell) + \frac{\delta_d(\ell)}{\sqrt{nh}}, \tag{D.3}$$

for $d = 1, 0$ and some $P_c \in \mathcal{P}_0$ defined in Assumption D.5. Here, $\delta_d(\ell) > 0$ specifies local violation of the null hypothesis inside the interval $\ell \in \mathcal{L}$. We consider local alternatives that satisfy the next set of assumptions:

Assumption D.6. *A sequence of local alternatives $\{P_n \in \mathcal{P}\backslash\mathcal{P}_0 :\ n \geq 1\}$ satisfies the following conditions:*

(i) (D.3) *holds under* P_n,

(ii) *for* $d = 0, 1$, $\delta_d(\ell) \geq 0$ *if* $\ell \in \mathcal{L}^o_{P_c,d}$,

(iii) *for* $d = 0, 1$, $\delta_d(\ell) > 0$ *for some* $\ell \in \mathcal{L}^o_{P_c,d}$.

(iv) *for* $d = 0, 1$, $\lim_{n\to\infty} d(h_{2,P_n,d,+}, h^*_{2,d,+}) = 0$ *and* $\lim_{n\to\infty} d(h_{2,P_n,d,-}, h^*_{2,d,-}) = 0$ *for some* $h^*_{2,d,+} \in \mathcal{H}_2$ *and* $h^*_{2,d,-} \in \mathcal{H}_2$.

Assumption D.6 (i) requires that the local alternatives converge to a boundary null P_c at rate $(nh)^{-1/2}$. Assumption D.6 (ii) ensures that our test is unbiased and Assumption D.6 (iii) makes sure that each P_n in the sequence is not in $\mathcal{P}_0$. Assumption D.6 (iv) restricts the asymptotic behavior of the covariance kernels as considered in LA1(c) of Andrews and Shi (2013).

The following theorem shows that the asymptotic local power of our test is greater than or equal to α when η tends to zero, i.e., our test is unbiased against those local alternatives that satisfy Assumption D.6.

Theorem D.3. *Suppose Assumptions D.1 to D.4 hold and* $\alpha < 1/2$. *If a sequence of local alternatives* $\{P_n : n \geq 1\}$ *satisfies Assumption D.6, then* $\lim_{\eta\to 0} \lim_{n\to\infty} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) \geq \alpha$.

Appendix E. Proofs

We first introduce a lemma that allows us to extend inequalities (2) and (3) to any Borel set in $\mathcal{Y}$.

Lemma E.1. *Under the conditions of Theorem 1 (i), inequalities (2) and (3) hold for any closed interval* $[y', y]$, $-\infty \leq y' \leq y \leq \infty$, *if and only if they hold for any Borel set in* $\mathcal{Y}$.

Proof. We focus on inequality (2). The claim concerning inequality (3) can be shown analogously. The "if" part is trivial. We apply Andrews and Shi (2013, Lemma C1) to show the "only if" part. Let $\mathcal{C} \equiv \{[y, y'] : -\infty \leq y \leq y' \leq \infty\}$ be the class of intervals and C be a generic element of $\mathcal{C}$. Let $\mu_1(\cdot) = \lim_{r\downarrow r_0} \mathbb{E}_P[1\{Y \in \cdot\}D|R = r] - \lim_{r\uparrow r_0} \mathbb{E}_P[1\{Y \in \cdot\}D|R = r]$, which is a well-defined set function if Assumptions 1 and 2 hold. See the proof of Theorem 1 (i) below for existence of the left and right limits of $E_P[1\{Y \in \cdot\}D|R = r]$. It then holds that $\mu_1 : \mathcal{C} \to \mathbb{R}$ is a bounded and countably additive set function satisfying $\mu_1(\emptyset) = 0$ and $\mu_1(C) \geq 0$ for any C. Applying Andrews and Shi (2013, Lemma C1), since the smallest σ-algebra generated by $\mathcal{C}$ coincides with the Borel σ-algebra $\mathcal{B}(\mathcal{Y})$, it follows that $\mu_1(C_\ell) \geq 0$ for any $C_\ell \in \mathcal{L}$ implies that $\mu_1(B) \geq 0$ for any $B \in \mathcal{B}(\mathcal{Y})$. □

E.1. **Proof of Theorem 1: Claim (i):** Let $B \subset \mathbb{R}$ be an arbitrary closed interval. We have

$$\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y \in B\}D|R = r_0+\epsilon] \geq \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} \in \{\mathbf{A},\mathbf{C}\}\}|R = r_0+\epsilon]$$
$$= \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{C}\}|R = r_0+\epsilon] + \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0+\epsilon],$$

where the first inequality follows since the set of selection types such that $D(r) = 1$ at $r_0 \leq r < r_0+\epsilon$ includes $\{\mathbf{A},\mathbf{C}\}$. On the other hand, we have

$$\begin{aligned}
&\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y \in B\}D|R = r_0+\epsilon] \\
&\quad \leq \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} \in \{\mathbf{A},\mathbf{C}\}\}|R = r_0+\epsilon] + \lim_{\epsilon\to 0}P(T_{|r-r_0|} = \mathbf{I}|R = r_0+\epsilon) \\
&= \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{C}\}|R = r_0+\epsilon] + \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0+\epsilon]
\end{aligned}$$

where the third line follows by Assumption 1. Hence,

$$\begin{aligned}
&\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y \in B\}D|R = r_0+\epsilon] \\
&= \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{C}\}|R = r_0+\epsilon] + \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0+\epsilon].
\end{aligned} \tag{E.1}$$

Similarly, we have $\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y \in B\}D|R = r_0-\epsilon] \geq \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0-\epsilon]$ and

$$\begin{aligned}
&\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y \in B\}D|R = r_0-\epsilon] \\
&\leq \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0-\epsilon] + \lim_{\epsilon\to 0}P(T_{|r-r_0|} \in \{\mathbf{I},\mathbf{DF}\}|R = r_0-\epsilon) \\
&= \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0-\epsilon],
\end{aligned}$$

implying

$$\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y \in B\}D|R = r_0-\epsilon] = \lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r) \in B, T_{|r-r_0|} = \mathbf{A}\}|R = r_0-\epsilon]. \tag{E.2}$$

Taking the difference of equation (E.1) and equation (E.2) and employing Assumption 2 leads to the desired inequality:

$$\begin{aligned}&\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y\in B\}D|R=r]-\lim_{r\uparrow r_0}\mathbb{E}_P[1\{Y\in B\}D|R=r]\\&=\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y_1(r)\in B, T_{|r-r_0|}\in\{\mathbf{C}\}\}|R=r]\geq 0.\end{aligned}\tag{E.3}$$

Similarly we can show that

$$\begin{aligned}&\lim_{r\uparrow r_0}\mathbb{E}_P[1\{Y\in B\}(1-D)|R=r]-\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y\in B\}(1-D)|R=r]\\&=\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y_0(r)\in B, T_{|r-r_0|}\in\{\mathbf{C}\}\}|R=r]\geq 0.\end{aligned}\tag{E.4}$$

Note that the proof is also valid when Assumption 1 is replaced by Assumption C.1.

Claim (ii): Suppose that the distribution of observables (Y, D, R) satisfies inequalities (2) and (3). By Lemma E.1, they hold for an arbitrary Borel set. By the absolute continuity assumption, we have the conditional density of (Y, D) given R denoted by $f_{Y,D|R}(y, d|r)$. We denote the left and right limits of $f_{Y,D|R}$ at r_0 by $f_{Y,D|R}(y, d|r_{0,-}) = \lim_{r\uparrow r_0} f_{Y,D|R}(y, d|r)$ and $f_{Y,D|R}(y, d|r_{0,+}) = \lim_{r\downarrow r_0} f_{Y,D|R}(y, d|r)$, respectively.

In what follows, we construct a joint distribution of potential variables $(\tilde{Y}_1(r), \tilde{Y}_0(r), \tilde{D}(r) : r \in \mathcal{R})$ that satisfies Assumptions 1 and 2 and matches with the given distribution of observables.

First, for $d \in \{0, 1\}$, consider outcome responses that are invariant to the running variable, $\tilde{Y}_d(r) = \tilde{Y}_d(r')$ for all $r, r' \in \mathcal{R}$, a.s., i.e., the running variable has no direct causal impact for anyone in the population. We can hence drop index r from the notation of the potential outcomes and reduce them to $(\tilde{Y}_1, \tilde{Y}_0) \in \mathcal{Y}^2$. For the treatment selection response to the running variable, consider that only the following selection responses are allowed in the population:

$$\tilde{D}(r)=\begin{cases}1\{r\geq r_0\}, & \text{labeled as } \tilde{T}=\mathbf{C}\\ 1, & \text{labeled as } \tilde{T}=\mathbf{A}\\ 0, & \text{labeled as } \tilde{T}=\mathbf{N}.\end{cases}$$

With these simplifications, we construct a joint distribution of $(\tilde{Y}_1(r), \tilde{Y}_0(r), \tilde{D}(r) : r \in \mathcal{R})$ given R by constructing a joint distribution of $(\tilde{Y}_1, \tilde{Y}_0, \tilde{T}) \in \mathcal{Y}^2 \times \{\mathbf{C}, \mathbf{A}, \mathbf{N}\}$ given R, where $\tilde{T}$ does not vary in $|r - r_0|$. To distinguish the probability law of observables corresponding to the given sampling process and the probability law of $(\tilde{Y}_1, \tilde{Y}_0, \tilde{T})$ to be constructed, we use P and f to denote the former probability law and its density, and $\mathbb{P}$ to denote the latter probability law.

Let $B \subset \mathbb{R}$ be an arbitrary Borel Set. For the always-takers' potential outcome distributions, consider

$$\mathbb{P}(\tilde{Y}_1 \in B, \tilde{T} = \mathbf{A}|r) = \begin{cases} P(Y \in B, D = 1|r), & \text{for } r < r_0, \\ \int_B \min \left\{ \begin{matrix} f_{Y,D|R}(y, D = 1|r_{0,-}), \\ f_{Y,D|R}(y, D = 1|r) \end{matrix} \right\} d\mu, & \text{for } r \geq r_0, \end{cases}$$

and

$$\mathbb{P}(\tilde{Y}_0 \in B, \tilde{T} = \mathbf{A}|r) = \begin{cases} Q(B)P(D = 1|r), & \text{for } r < r_0, \\ Q(B) \int_{\mathcal{Y}} \min \left\{ \begin{matrix} f_{Y,D|R}(y, D = 1|r_{0,-}), \\ f_{Y,D|R}(y, D = 1|r) \end{matrix} \right\} d\mu, & \text{for } r \geq r_0, \end{cases},$$

where $Q(\cdot)$ is an arbitrary probability measure on $\mathcal{Y}$. The joint distribution of $(\tilde{Y}_1, \tilde{Y}_0, \tilde{T} = \mathbf{A})$ can be constructed by coupling these distributions assuming, for instance, that $\tilde{Y}_1$ and $\tilde{Y}_0$ are independent conditional on $(\tilde{T}, R)$.

For the never-takers' potential outcome distributions, consider

$$\mathbb{P}(\tilde{Y}_0 \in B, \tilde{T} = \mathbf{N}|r) = \begin{cases} P(Y \in B, D = 0|r), & \text{for } r \geq r_0, \\ \int_B \min \left\{ \begin{matrix} f_{Y,D|R}(y, D = 0|r_{0,+}), \\ f_{Y,D|R}(y, D = 0|r) \end{matrix} \right\} d\mu, & \text{for } r < r_0, \end{cases}$$

and

$$\mathbb{P}(\tilde{Y}_1 \in B, \tilde{T} = \mathbf{N}|r) = \begin{cases} Q(B)P(D = 0|r), & \text{for } r \geq r_0, \\ Q(B) \int_{\mathcal{Y}} \min \left\{ \begin{matrix} f_{Y,D|R}(y, D = 0|r_{0,+}) \\ f_{Y,D|R}(y, D = 0|r) \end{matrix} \right\} d\mu, & \text{for } r < r_0. \end{cases}$$

For the compliers' potential outcome distributions, if $\pi^+ = \pi^-$, we specify that no compliers exist in the population. If $\pi^+ > \pi^-$, consider

$$\mathbb{P}(\tilde{Y}_1 \in B, \tilde{T} = \mathbf{C}|r)$$
$$= \begin{cases} P(Y \in B, D = 1|r) - \int_B \min \begin{Bmatrix} f_{Y,D|R}(y, D = 1|r_{0,-}), \\ f_{Y,D|R}(y, D = 1|r) \end{Bmatrix} d\mu, & \text{for } r \geq r_0, \\ (\pi^+ - \pi^-)^{-1} \left[P(D = 1|r) - \int_{\mathcal{Y}} \min \begin{Bmatrix} f_{Y,D|R}(y, D = 1|r_{0,-}), \\ f_{Y,D|R}(y, D = 1|r) \end{Bmatrix} d\mu \right] & \\ \times \left[\lim_{r \downarrow r_0} P(Y \in B, D = 1|r) - \lim_{r \uparrow r_0} P(Y \in B, D = 1|r) \right], & \text{for } r < r_0. \end{cases}$$

and

$$\mathbb{P}(\tilde{Y}_0 \in B, \tilde{T} = \mathbf{C}|r)$$
$$= \begin{cases} P(Y \in B, D = 0|r) - \int_B \min \begin{Bmatrix} f_{Y,D|R}(y, D = 0|r_{0,+}), \\ f_{Y,D|R}(y, D = 0|r) \end{Bmatrix} d\mu, & \text{for } r < r_0, \\ (\pi^+ - \pi^-)^{-1} \left[P(D = 1|r) + - \int_{\mathcal{Y}} \min \begin{Bmatrix} f_{Y,D|R}(y, D = 1|r_{0,-}), \\ f_{Y,D|R}(y, D = 1|r) \end{Bmatrix} d\mu \right] & \\ \times \left[\lim_{r \uparrow r_0} P(Y \in B, D = 0|r) - \lim_{r \downarrow r_0} P(Y \in B, D = 0|r) \right], & \text{for } r \geq r_0. \end{cases}$$

If the distribution of (Y, D, R) satisfies the testable implications shown in the first claim, then it can be shown that the conditional distribution of $(\tilde{Y}_1, \tilde{Y}_0, \tilde{T})$ given $R = r$ constructed in this way is a proper probability distribution (i.e., nonnegative, additive, and sum up to one) for all r. We can also confirm that the constructed distribution of $(\tilde{Y}_1, \tilde{Y}_0, \tilde{T})$ given R matches with the distribution of observables, i.e., it satisfies, for any $d = 1, 0$, $r \in \mathcal{R}$, and measurable set $B \subset \mathcal{Y}$,

$$P(Y \in B, D = d|r) = \sum_{\tilde{T}:\tilde{D}(r)=d} \mathbb{P}(\tilde{Y}_d \in B, \tilde{T}|r).$$

We now check the conditional distribution of $(\tilde{Y}_1, \tilde{Y}_0, \tilde{T})$ given R constructed above satisfies Assumptions 1 and 2. First, by the construction of treatment selection response, $\mathbb{P}(\tilde{T} = \{\mathbf{DF}, \mathbf{I}\}|r) = 0$ for any r. Hence, Assumption 1 holds.

To check Assumption 2, note that

$$\lim_{r\downarrow r_0}\mathbb{P}(\tilde{Y}_1\in B,\tilde{T}=\mathbf{A}|r)=\lim_{r\downarrow r_0}\int_B \min\left\{\begin{array}{c} f_{Y,D|R}(y,D=1|r_{0,-}),\\ f_{Y,D|R}(y,D=1|r)\end{array}\right\}d\mu$$

$$=\int_B \min\left\{\begin{array}{c} f_{Y,D|R}(y,D=1|r_{0,-}),\\ f_{Y,D|R}(y,D=1|r_{0,+})\end{array}\right\}d\mu=\int_B f_{Y,D|R}(y,D=1|r_{0,-})d\mu$$

$$=\lim_{r\uparrow r_0}P(Y\in B,D=1|r)=\lim_{r\uparrow r_0}\mathbb{P}(\tilde{Y}_1\in B,\tilde{T}=\mathbf{A}|r),$$

where the third equality follows by the assumption that the distribution of (Y,D,R) satisfies inequality (2). Hence, $\mathbb{P}(\tilde{Y}_1,\tilde{T}=\mathbf{A}|r)$ is continuous at $r=r_0$. Similar arguments apply to show that $\mathbb{P}(\tilde{Y}_0,\tilde{T}=\mathbf{A}|r)$, $\mathbb{P}(\tilde{Y}_1,\tilde{T}=\mathbf{N}|r)$, and $\mathbb{P}(\tilde{Y}_0,\tilde{T}=\mathbf{N}|r)$ are all continuous at r_0. For compliers, we have

$$\lim_{r\downarrow r_0}\mathbb{P}(\tilde{Y}_1\in B,\tilde{T}=\mathbf{C}|r)$$

$$=\lim_{r\downarrow r_0}\left[P(Y\in B,D=1|r)-\int_B \min\left\{\begin{array}{c} f_{Y,D|R}(y,D=1|r_{0,-}),\\ f_{Y,D|R}(y,D=1|r)\end{array}\right\}d\mu\right]$$

$$=\lim_{r\downarrow r_0}P(Y\in B,D=1|r)-\lim_{r\uparrow r_0}P(Y\in B,D=1|r).$$

Also, by noting $\lim_{r\downarrow r_0}\int_{\mathcal{Y}}\min\left\{\begin{array}{c} f_{Y,D|R}(y,D=1|r_{0,-}),\\ f_{Y,D|R}(y,D=1|r)\end{array}\right\}d\mu=\pi^{-}$, we obtain

$$\lim_{r\uparrow r_0}\mathbb{P}(\tilde{Y}_1\in B,\tilde{T}=\mathbf{C}|r)=\lim_{r\downarrow r_0}P(Y\in B,D=1|r)-\lim_{r\uparrow r_0}P(Y\in B,D=1|r).$$

Hence, we have shown that the constructed distribution of $(\tilde{Y}_1,\tilde{Y}_0,\tilde{T})$ given R satisfies Assumption 2. This completes the proof of the second claim.

E.2. Identification of the compliers' potential outcome distributions.

Proposition E.1. *If Assumptions 1 to 3 hold, then the compliers' potential outcome distributions at the cut-off,*

$$F_{Y_1(r_0)|\mathbf{C},R=r_0}(y)\equiv\lim_{r\to r_0}P\left(Y_1(r)\le y|T_{|r-r_0|}=\mathbf{C},R=r\right),$$

$$F_{Y_0(r_0)|\mathbf{C},R=r_0}(y)\equiv\lim_{r\to r_0}P\left(Y_0(r)\le y|T_{|r-r_0|}=\mathbf{C},R=r\right),$$

are identified by

$$F_{Y_1(r_0)|\mathbf{C},R=r_0}(y) = \frac{\lim\limits_{r\downarrow r_0}\mathbb{E}_P[1\{Y \le y\}D|R=r] - \lim\limits_{r\uparrow r_0}\mathbb{E}_P[1\{Y \le y\}D|R=r]}{\pi^+ - \pi^-},$$

$$F_{Y_0(r_0)|\mathbf{C},R=r_0}(y) = \frac{\lim\limits_{r\uparrow r_0}\mathbb{E}_P[1\{Y \le y\}(1-D)|R=r] - \lim\limits_{r\downarrow r_0}\mathbb{E}_P[1\{Y \le y\}(1-D)R=r]}{\pi^+ - \pi^-}.$$

Proof. We first note that under Assumptions 1 and 2, $\pi^+ - \pi^- = \lim_{r\to r_0} P(T_{|r-r_0|} = \mathbf{C}|\mathbf{R} = \mathbf{r})$. Based on (E.3) in the proof of Theorem 1, we have

$$\begin{aligned}
& \lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y \le y\}D|R=r] - \lim_{r\uparrow r_0}\mathbb{E}_P[1\{Y \le y\}D|R=r] \\
= \; & F_{Y_1(r_0)|\mathbf{C},R=r_0}(y) \cdot \lim_{r\to r_0} P(T_{|r-r_0|} = \mathbf{C}|\mathbf{R} = \mathbf{r}) \\
= \; & F_{Y_1(r_0)|\mathbf{C},R=r_0}(y) \cdot (\pi^+ - \pi^-)
\end{aligned}$$

Hence, the identification result of $F_{Y_1(r_0)|\mathbf{C},R=r_0}(y)$ is shown.

The identification result for $F_{Y_0(r_0)|\mathbf{C},R=r_0}(y)$ can be shown similarly by using equation (E.4). We omit the details for brevity. □

E.3. **Lemmas and Proofs for Theorems in Appendix D.** We show three lemmas that lead to the theorems in Appendix D.

We first present a lemma that shows a Bahadur representation for $\hat{m}_{d,\star}$, $d = 0, 1$ and $\star = +, -$, uniform in $\ell \in \mathcal{L}$ and $P \in \mathcal{P}$ subject to Assumption D.1. This lemma extends the undersmoothing case of Lemma 1 in Chiang, Hsu, and Sasaki (2017) by providing an approximation that is also uniform over the data generating processes $P \in \mathcal{P}$. It also modifies the undersmoothing case of Theorem 1 in Lee, Song, and Whang (2015) by focusing on the boundary point and uniformity over the class of intervals rather than quantiles.

Given a class of data generating processes $\mathcal{P}$, we say that a sequence of random variables $\{Z_n\}$ converges in probability to zero $\mathcal{P}$-uniformly if $\sup_{\{P\in\mathcal{P}\}} P(|Z_n| > \epsilon) \to 0$ as $n \to \infty$ for any $\epsilon > 0$, which we denote by $Z_n = o_{\mathcal{P}}(1)$.

Lemma E.2. *Let $\mathcal{P}$ be a class of data generating processes satisfying Assumption D.1, and $\hat{m}_{d,\star}$, $m_{P,d}$, and $m_{P,d,\star}$, $d = 1, 0$ and $\star = +, -$, be as defined in Appendix A. Under Assumption D.2,*

$$\sup_{\{\ell \in \mathcal{L}\}} \left| \sqrt{nh}(\hat{m}_{d,\star}(\ell) - m_{P,d,\star}(\ell)) - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i^{\star} \mathcal{E}_{d,i}(\ell) \right| = o_{\mathcal{P}}(1), \tag{E.5}$$

where

$$\begin{aligned} w_i^{+} &= \frac{\left[\vartheta_2 - \vartheta_1 \left(\frac{R_i - r_0}{h_+}\right)\right] K\left(\frac{R_i - r_0}{h_+}\right) 1\{R_i \geq r_0\}}{c_+ f_R^{+}(r_0)(\vartheta_0\vartheta_2 - \vartheta_1^2)}, \\ w_i^{-} &= \frac{\left[\vartheta_2 + \vartheta_1 \left(\frac{R_i - r_0}{h_-}\right)\right] K\left(\frac{R_i - r_0}{h_-}\right) 1\{R_i < r_0\}}{c_- f_R^{-}(r_0)(\vartheta_0\vartheta_2 - \vartheta_1^2)}, \\ \mathcal{E}_{d,i}(\ell) &= g_\ell(Y_i) D_i^d (1 - D_i)^{1-d} - m_{P,d}(\ell, R_i). \end{aligned}$$

Proof. We provide a proof for the case of $d = 1$ and $\star = +$ only, as the proofs for the other cases are similar. Substituting the mean value expansion, $g_\ell(Y_i) D_i = m_{P,1}(\ell, R_i) + \mathcal{E}_{1,i}(\ell) = m_{P,1,+}(\ell) + h_+ m_{P,1}^{(1)}(\ell, r_0)\left(\frac{R_i - r_0}{h_+}\right) + \frac{h_+^2}{2} m_{P,1}^{(2)}(\ell, \tilde{R}_i)\left(\frac{R_i - r_0}{h_+}\right)^2 + \mathcal{E}_{1,i}(\ell)$, $\tilde{R}_i \in [0, R_i]$, we obtain

$$\begin{aligned} &\sqrt{nh}\left[\hat{m}_{1,+}(\ell) - m_{P,1,+}(\ell)\right] \\ =&\sqrt{nh^3} \cdot c_+ \sum_{i=1}^{n} w_{n,i}^{+} m_{P,1}^{(1)}(\ell, r_0)\left(\frac{R_i - r_0}{h_+}\right) + \sqrt{nh^5} \cdot \frac{c_+^2}{2} \sum_{i=1}^{n} w_{n,i}^{+} m_{P,1}^{(2)}(\ell, \tilde{R}_i)\left(\frac{R_i - r_0}{h_+}\right)^2 \qquad \text{(E.6)} \\ &+ \sqrt{nh} \cdot \sum_{i=1}^{n} w_{n,i}^{+} \mathcal{E}_{1,i}(\ell) \qquad \text{(E.7)} \end{aligned}$$

The first order conditions for the local linear regression implies the first term in (E.6) is zero. By the boundedness of $m_{P,1}^{(2)}$ (Assumption D.1) (iv), the absolute value of the second term in (E.6) can be bounded uniformly in $\ell \in \mathcal{L}$ by $M\sqrt{nh^5}\frac{c_+^2}{2}\left|\sum_{i=1}^{n} w_{n,i}^{+}\left(\frac{R_i - r_0}{h_+}\right)^2\right|$. Since we have

$$\begin{aligned} \sum_{i=1}^{n} w_{n,i}^{+}\left(\frac{R_i - r_0}{h_+}\right)^2 &= \frac{(\hat{\vartheta}_2^{+})^2 - \hat{\vartheta}_1^{+}\hat{\vartheta}_3^{+}}{\hat{\vartheta}_2^{+}\hat{\vartheta}_0^{+} - (\hat{\vartheta}_1^{+})^2} \\ &= \frac{\vartheta_2^2 - \vartheta_1\vartheta_3}{\vartheta_2\vartheta_0 - \vartheta_1^2} + o_{\mathcal{P}}(1), \end{aligned}$$

where $w_{n,i}^{+}$ and $\hat{\vartheta}_j^{+}$ are as defined in Appendix A, and the second line follows by Lemma 2 in Fan and Gijbels (1992); for nonnegative finite j,

$$\hat{\vartheta}_j^{+} = f_R^{+}(r_0)\vartheta_j + o_{\mathcal{P}}(1) \tag{E.8}$$

holds where the $\mathcal{P}$-uniform convergence here follows by Assumption D.1 (i.e., $\mathcal{P}$ shares the common marginal distribution of R). Hence, combined with the undersmoothing bandwidth (Assumption D.2 (iii)), the second term in (E.6) is $o_{\mathcal{P}}(1)$.

The conclusion of the lemma is obtained by verifying $\sup_{\{\ell\in\mathcal{L}\}} \left| \sqrt{nh} \sum_{i=1}^{n} w_{n,i}^{+}\mathcal{E}_{1,i}(\ell) - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i^{+}\mathcal{E}_{1,i}(\ell) \right| = o_{\mathcal{P}}(1)$. Consider

$$\begin{aligned}
&\sup_{\{\ell\in\mathcal{L}\}} \left| \sqrt{nh} \sum_{i=1}^{n} w_{n,i}^{+}\mathcal{E}_{1,i}(\ell) - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i^{+}\mathcal{E}_{1,i}(\ell) \right| \\
\le & c_+^{-1} \underbrace{\left| \frac{\hat{\vartheta}_2^{+}}{\hat{\vartheta}_2^{+}\hat{\vartheta}_0^{+} - (\hat{\vartheta}_1^{+})^2} - \frac{\vartheta_2}{f_R^{+}(r_0)(\vartheta_2\vartheta_0 - \vartheta_1^2)} \right|}_{\text{(i)}} \cdot \underbrace{\sup_{\{\ell\in\mathcal{L}\}} \left| \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K\left(\frac{R_i - r_0}{h_+}\right) 1\{R_i \ge r_0\}\mathcal{E}_{1,i}(\ell) \right|}_{\text{(ii)}} \\
& + c_+^{-1} \underbrace{\left| \frac{\hat{\vartheta}_1^{+}}{\hat{\vartheta}_2^{+}\hat{\vartheta}_0^{+} - (\hat{\vartheta}_1^{+})^2} - \frac{\vartheta_1}{f_R^{+}(r_0)(\vartheta_2\vartheta_0 - \vartheta_1^2)} \right|}_{\text{(iii)}} \cdot \underbrace{\sup_{\{\ell\in\mathcal{L}\}} \left| \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K\left(\frac{R_i - r_0}{h_+}\right) \left(\frac{R_i - r_0}{h_+}\right) 1\{R_i \ge r_0\}\mathcal{E}_{1,i}(\ell) \right|}_{\text{(iv)}}.
\end{aligned} \tag{E.9}$$

Since (E.8) implies both terms (i) and (iii) in (E.9) are $o_{\mathcal{P}}(1)$, it suffices to show that the terms (ii) and (iv) in (E.9) are stochastically bounded uniformly in $P \in \mathcal{P}$. Let j be a nonnegative integer and

$$f_{n,i}^{(j)}(\ell) \equiv \frac{1}{\sqrt{h}} K\left(\frac{R_i - r_0}{h_+}\right) \left(\frac{R_i - r_0}{h_+}\right)^j 1\{R_i \ge r_0\}\mathcal{E}_{1,i}(\ell).$$

Consider obtaining a $\mathcal{P}$-uniform bound for $P(\sqrt{n}\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^{n} f_{n,i}^{(j)}(\ell)\right| > \epsilon)$ for $\epsilon > 0$ (i.e., term (ii) corresponds to $j=0$ and term (iv) corresponds to $j=1$). By Markov's inequality,

$$
\begin{aligned}
P\left(\sqrt{n}\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^{n} f_{n,i}^{(j)}(\ell)\right| > \epsilon\right) &\leq \epsilon^{-1}\sqrt{n}\mathbb{E}_P\left[\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^{n} f_{n,i}^{(j)}(\ell)\right|\right] \\
&= \epsilon^{-1}\sqrt{n}\left(\mathbb{E}_P\left[\max\left\{\sup_{\{\ell\in\mathcal{L}\}}\frac{1}{n}\sum_{i=1}^{n} f_{n,i}^{(j)}(\ell), \sup_{\{\ell\in\mathcal{L}\}}\frac{1}{n}\sum_{i=1}^{n}(-f_{n,i}^{(j)}(\ell))\right\}\right]\right) \\
&= \epsilon^{-1}\sqrt{n}\mathbb{E}_P\left[\sup_{\{f_{n,i}^{(j)}\in\mathcal{F}_n^{+}\cup\mathcal{F}_n^{-}\}}\frac{1}{n}\sum_{i=1}^{n} f_{n,i}^{(j)}\right], \qquad \text{(E.10)}
\end{aligned}
$$

where $\mathcal{F}_n^{+} \equiv \{f_{n,i}^{(j)}(\ell) : \ell \in \mathcal{L}\}$ and $\mathcal{F}_n^{-} \equiv \{-f_{n,i}^{(j)}(\ell) : \ell \in \mathcal{L}\}$. Note that $\mathcal{F}_n^{+}$ and $\mathcal{F}_n^{-}$ are VC-subgraph classes whose VC-dimensions are equal to 2 (see, e.g., Lemma A.1 in Kitagawa and Tetenov (2018)) with a uniform envelope $\bar{K}/\sqrt{h}$ and an $L_2(P)$ envelope,

$$
\sup_{\{\ell\in\mathcal{L}\}}\|f_{n,i}^{(j)}(\ell)\|_{L_2(P)} \leq \left[c_{+}\bar{f}_R\int_0^{\infty}K^2(u)u^{2j}du\right]^{1/2} < \infty.
$$

Since $\mathcal{F}_n^{+}\cup\mathcal{F}_n^{-}$ is a VC-subgraph class sharing the same uniform and $L_2(P)$ envelope, a maximal inequality for the VC-subgraph class of functions with bounded $L_2(P)$-envelope (Lemma A.5 in Kitagawa and Tetenov (2018)) applies and (E.10) can be bounded from above by

$$
C_1\left(c_{+}\bar{f}_R\int_0^{\infty}K^2(u)u^{2j}du\right)^{1/2} n^{-1/2}
$$

for all n satisfying $nh \geq \frac{C_2\bar{K}^2}{c_{+}\bar{f}_R\int_0^{\infty}K^2(u)u^{2j}du}$, where C_1 and C_2 are positive constants that do not depend on P or bandwidth. Since $nh \to \infty$, this maximal inequality with $j=0$ and $j=1$ imply term (ii) and term (iv) in (E.9) are stochastically bounded $\mathcal{P}$-uniformly. Hence,

$$
\sup_{\{\ell\in\mathcal{L}\}}\left|\sqrt{nh}\sum_{i=1}^{n} w_{n,i}^{+}\mathcal{E}_{1,i}(\ell) - \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i^{+}\mathcal{E}_{1,i}(\ell)\right| = o_{\mathcal{P}}(1) \qquad \text{(E.11)}
$$

holds. □

The next lemma shows $\mathcal{P}$-uniform convergence of the covariance kernel of $w_i^{\star}\mathcal{E}_{d,i}(\cdot)$, the summand in the Bahadur representation of Lemma E.2.

Lemma E.3. *Let $d = 1$ or 0, and $\star = +$ or $-$. For $\ell_1, \ell_2 \in \mathcal{L}$, define*

$$\hat{h}_{2,P,d,\star}(\ell_1, \ell_2) = \frac{1}{nh} \sum_{i=1}^{n} (w_i^\star)^2 \sigma_{P,d}(\ell_1, \ell_2 | R_i).$$

Let $\mathcal{P}$ be a class of data generating processes satisfying Assumption D.1 and assume that the kernel function and the bandwidth satisfy Assumption D.2. Then,

$$\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} \left| \hat{h}_{2,P,d,\star}(\ell_1, \ell_2) - h_{2,P,d,\star}(\ell_1, \ell_2) \right| = o_{\mathcal{P}}(1),$$

where $h_{2,P,d,\star}$ is as defined in equation (D.2) above.

Proof. We show the claim for the case of $d = 1$ and $\star = +$. The other cases can be proven similarly. Since

$$\begin{aligned} &\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} \left| \hat{h}_{2,P,1,+}(\ell_1, \ell_2) - h_{2,P,1,+}(\ell_1, \ell_2) \right| \\ \leq &\underbrace{\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} \left| \hat{h}_{2,P,1,+}(\ell_1, \ell_2) - \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1, \ell_2)] \right|}_{\text{(v)}} + \underbrace{\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} \left| \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1, \ell_2)] - h_{2,P,1,+}(\ell_1, \ell_2) \right|}_{\text{(vi)}}, \end{aligned}$$

we show $\mathcal{P}$-uniform convergences of term (v) and term (vi) separately.

First, by exploiting Assumption D.1, we can obtain a uniform upper bound of term (vi) as follows:

$$\left| \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1, \ell_2)] - h_{2,P,1,+}(\ell_1, \ell_2) \right| \leq \frac{5M\bar{f}_R \int_0^\infty (\vartheta_2 - \vartheta_1 u)^2 u K^2(u) du}{(f_R^+(r_0))^2 (\vartheta_0 \vartheta_2 - \vartheta_1^2)^2} h, \tag{E.12}$$

which converges to zero as $n \to \infty$ since $h \to 0$. Since the marginal distribution of R is common for $\mathcal{P}$, this convergence is uniform in $P \in \mathcal{P}$, so term (vi) is $o_{\mathcal{P}}(1)$.

Regarding term (v), Jensen's inequality bounds its mean by

$$\begin{aligned} &\mathbb{E}_P \left[\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} \left| \hat{h}_{2,P,1,+}(\ell_1, \ell_2) - \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1, \ell_2)] \right| \right] \\ \leq &\frac{1}{[c_+ f_R^+(r_0)(\vartheta_0 \vartheta_2 - \vartheta_1^2)]^2} \mathbb{E}_P \left[\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} \left| \frac{1}{n} \sum_{i=1}^{n} f_{n,i}(\ell_1, \ell_2) - \mathbb{E}_P(f_{n,i}(\ell_1, \ell_2)) \right| \right], \end{aligned} \tag{E.13}$$

where $f_{n,i}(\ell_1, \ell_2) \equiv \frac{1}{h} \left[\vartheta_2 - \vartheta_1 \left(\frac{R_i - r_0}{h_+} \right) \right]^2 K^2 \left(\frac{R_i - r_0}{h_+} \right) \cdot 1\{R_i \geq r_0\} \mathcal{E}_{1,i}(\ell_1) \mathcal{E}_{1,i}(\ell_2)$. Since $\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)$ can be viewed as the sum of three indicator functions for intervals (indexed by ℓ_1 and ℓ_2), $\{f_{n,i}(\ell_1, \ell_2) :$

$\ell_1, \ell_2 \in \mathcal{L}\}$ is a VC-subgraph class of functions with a uniform envelope $h^{-1}(\vartheta_2 + \vartheta_1)^2 \bar{K}^2$ and $L_2(P)$-envelope,

$$[\mathbb{E}_P(f_{n,i}^2(\ell_1, \ell_2))]^{1/2} \leq \frac{1}{\sqrt{h}} \left[c_+ \bar{f}_R \int_0^\infty (\vartheta_2 - \vartheta_1 u)^4 K^4(u)\, du \right]^{1/2}.$$

Applying Lemma A.5 in Kitagawa and Tetenov (2018), we obtain

$$\mathbb{E}_P \left[\sup_{\{\ell_1, \ell_2 \in \mathcal{L}\}} \left| \frac{1}{n} \sum_{i=1}^n f_{n,i}(\ell_1, \ell_2) - E_P(f_{n,i}(\ell_1, \ell_2)) \right| \right] \leq \frac{C_1}{\sqrt{nh}} \sqrt{c_+ \bar{f}_R \int_0^\infty (\vartheta_2 - \vartheta_1 u)^4 K^4(u)\, du} \tag{E.14}$$

for all $nh \geq \frac{C_2(\vartheta_2+\vartheta_1)^4 \bar{K}^4}{c_+ \bar{f}_R \int_0^\infty (\vartheta_2 - \vartheta_1 u)^4 K^4(u) du}$, where C_1 and C_2 are positive constraints that do not depend on P and h. Combining (E.13), (E.14), and Markov's inequality, we conclude that term (v) is $o_{\mathcal{P}}(1)$. $\square$

Exploiting the preceding two lemmas, the next lemma proves the functional central limit theorem for $\hat{m}_{n,d,\star}$ along sequences of the data generating processes in $\mathcal{P}$.

Lemma E.4. *Suppose that Assumptions D.1 and D.2 hold, and let $\{P_n\}$ be a sequence of data generating processes in $\mathcal{P}$. Then, for any subsequence $\{k_n\}$ of $\{n\}$ such that for $d = 0, 1$ and $\star = +, -$, $\lim_{n\to\infty} d(h_{2,P_{k_n},d,\star}, h^*_{2,d,\star}) = 0$ for some $h^*_{2,d,\star} \in \mathcal{H}_2$, we have*

$$\sqrt{k_n h}(\hat{m}_{d,\star}(\cdot) - m_{P_{k_n},d,\star}(\cdot)) \Rightarrow \Phi_{h^*_{2,d,\star}}(\cdot), \tag{E.15}$$

where Φ_{h_2} denotes a mean zero Gaussian process with covariance kernel h_2. In addition, we have for $d = 0, 1$,

$$\sqrt{k_n h}(\hat{v}_d(\cdot) - v_{P_{k_n},d}(\cdot)) \Rightarrow \Phi_{h^*_{2,d}}(\cdot),$$

where $h^*_{2,d} = h^*_{2,d,+} + h^*_{2,d,-}$.

Proof. To simplify notation, we show this theorem for a sequence $\{n\}$. All the arguments go through with $\{k_n\}$ in place of $\{n\}$.

By Lemma E.2, (E.15) follows if we show $\frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i^+ \mathcal{E}_{1,i}(\cdot) \Rightarrow \Phi_{h^*_{2,d,+}}(\cdot)$. For this purpose, we apply the functional central limit theorem (FCLT; Theorem 10.6 of Pollard (1990)) to the triangular array of independent processes, $\{f_{n,i}(\cdot) : 1 \leq i \leq n\}$, where $f_{n,i}(\ell) = \frac{1}{\sqrt{nh}} w_i^+ \mathcal{E}_{1,i}(\ell)$, $\ell \in \mathcal{L}$. Let their envelope functions be $\{F_{n,i} : 1 \leq i \leq n\}$ with $F_{n,i} = (nh)^{-1/2}|w_i^+|$. Define empirical processes indexed by $\ell \in \mathcal{L}$ as $\widehat{\Phi}_n^+(\ell) = \sum_{i=1}^n f_{n,i}(\ell)$. First, since $\{f_{n,i}(\ell) : \ell \in \mathcal{L}\}$ is a VC-subgraph class of functions (see, e.g., Lemma A.1 in Kitagawa and Tetenov (2018)), manageability of $\{f_{n,i}(\ell) : \ell \in \mathcal{L}, 1 \leq i \leq n\}$ (condition (i) of Theorem 10.6 in Pollard (1990)) is implied by a polynomial bound for the packing number of the VC-subgraph class of functions (see, e.g., Theorem 4.8.1 in Dudley (1999)). For condition (ii) of Theorem 10.6 in Pollard (1990), note that

$$\begin{aligned}
\mathbb{E}_{P_n}[\widehat{\Phi}_n^+(\ell_1)\widehat{\Phi}_n^+(\ell_2)] &= \frac{1}{h}\mathbb{E}_{P_n}\left[(w_i^+)^2\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)\right] \\
&= \mathbb{E}_{P_n}[\hat{h}_{2,P_n,1,+}(\ell_1,\ell_2)] = h_{2,P_n,1,+}(\ell_1,\ell_2) + o(1) \\
&\to h^*_{2,1,+}(\ell_1,\ell_2),
\end{aligned}$$

as $n \to \infty$, where the $o(1)$ term in the second line follows from the bound shown in (E.12), and the third line follows by the assumption on $\{P_n\}$ in the current lemma. Condition (iii) of Theorem 10.6 in Pollard (1990) can be shown by noting

$$\sum_{i=1}^n \mathbb{E}_{P_n}[F_{n,i}^2] = \frac{1}{h}\mathbb{E}_{P_n}[(w_i^+)^2] \leq \frac{\bar{f}_R\int_0^\infty(\vartheta_2-\vartheta_1 u)^2K^2(u)du}{c_+(f_R^+(r_0))^2(\vartheta_2\vartheta_0-\vartheta_1^2)^2}.$$

Condition (iv) of Theorem 10.6 in Pollard (1990) follows by that, for any $\epsilon > 0$,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}_{P_n}[F_{n,i}^2\cdot 1\{F_{n,i}>\epsilon\}] \leq \sum_{i=1}^n \mathbb{E}_{P_n}\left[\frac{F_{n,i}^4}{\epsilon^2}\right] &= \frac{1}{\epsilon^2 nh^2}E_{P_n}[(w_i^+)^4] \\
&\leq (nh)^{-1}\frac{\bar{f}_R\int_0^\infty(\vartheta_2-\vartheta_1 u)^4K^4(u)du}{\epsilon^2c_+^3[f_R^+(r_0)(\vartheta_0\vartheta_2-\vartheta_1^2)]^4} \to 0 \text{ as } n\to\infty,
\end{aligned}$$

where the first inequality holds because $1\{F_{n,i} > \epsilon\} \leq (F_{n,i}/\epsilon)^\varsigma$ for any $\varsigma > 0$ and we take $\varsigma = 2$ here.

To show condition (v) of Theorem 10.6 in Pollard (1990), note that

$$\begin{aligned}
\hat{\rho}^2_{1,+}(\ell_1,\ell_2) = &\sum_{i=1}^n E_{P_n}(f_{n,i}(\ell_1)-f_{n,i}(\ell_2))^2 \\
&= h_{2,P_n,1,+}(\ell_1,\ell_1) - 2h_{2,P_n,1,+}(\ell_1,\ell_2) + h_{2,P_n,1,+}(\ell_2,\ell_2) + o(1) \\
&\to h^*_{2,1,+}(\ell_1,\ell_1) - 2h^*_{2,1,+}(\ell_1,\ell_2) + h^*_{2,1,+}(\ell_2,\ell_2) \equiv \rho^2_{1,+}(\ell_1,\ell_2),
\end{aligned}$$

where the second line follows by (E.12). Note that the convergence in the last line holds uniformly over $\ell_1, \ell_2 \in \mathcal{L}$ by Lemma E.3, and this uniform convergence ensures condition (v) of Theorem 10.6 in Pollard (1990).

Hence, by the FCLT of Pollard (1990), we obtain $\sqrt{nh}(\hat{m}_{1,+}(\ell) - m_{P_n,1,+}(\ell)) \Rightarrow \Phi_{h_{2,1,+}}(\ell)$. Similarly, we can show $\sqrt{nh}(\hat{m}_{1,-}(\ell) - m_{P_n,1,-}(\ell)) \Rightarrow \Phi_{h_{2,1,-}}(\ell)$.

To show the second part, note that

$$\begin{aligned} &\sqrt{nh}(\hat{v}_1(\ell) - v_{P_n,1}(\ell)) = \sqrt{nh}(\hat{m}_{1,-}(\ell) - m_{P_n,1,-}(\ell)) - \sqrt{nh}(\hat{m}_{1,+}(\ell) - m_{P_n,1,+}(\ell)) \\ \Rightarrow &\Phi_{h^*_{2,1,-}+h^*_{2,1,+}}(\ell) = \Phi_{h^*_{2,1}}(\ell), \end{aligned}$$

where the weak convergence holds due to the fact that $\hat{m}_{n,1,+}(\ell)$ and $\hat{m}_{n,1,-}(\ell)$ are estimated from separate samples, so that the two processes are mutually independent. The same arguments apply to the $d = 0$ case. This completes the proof. □

Define, for $d = 1, 0$ and $\star = +, -$,

$$\widehat{\Phi}^u_{n,d,\star}(\ell) = \sum_{i=1}^{n} U_i \cdot \sqrt{nh} w^{\star}_{n,i} (g_\ell(Y_i) D_i^d (1 - D_i)^{1-d} - \hat{m}_{n,d,\star}(\ell)).$$

We denote weak convergence conditional on a sample generated from a sample size-dependent distribution of data P_n by $\overset{P_n}{\Rightarrow}$.[4] We denote convergence in probability along the sequence $\{P_n\}$ by $\overset{P_n}{\to}$.

Lemma E.5. *Suppose that Assumptions D.1-D.3 hold, and let $\{P_n\}$ be a sequence of data generating processes in $\mathcal{P}$. For a subsequence $\{k_n\}$ of $\{n\}$ such that for $d = 0, 1$ and $\star = +, -$, $\lim_{n\to\infty} d(h_{2,P_{k_n},d,\star}, h^*_{2,d,\star}) = 0$ holds for some $h^*_{2,d,\star} \in \mathcal{H}_2$, then $\widehat{\Phi}^u_{k_n,d,\star} \overset{P_{k_n}}{\Rightarrow} \Phi_{h^*_{2,d,\star}}$. In addition, for $d = 0, 1$, $\widehat{\Phi}^u_{v_1,k_n} \equiv \widehat{\Phi}^u_{n,1,-}(\ell) - \widehat{\Phi}^u_{n,1,+}(\ell) \overset{P_{k_n}}{\Rightarrow} \Phi_{h^*_{2,d}}$ and $\widehat{\Phi}^u_{v_0,k_n} \equiv \widehat{\Phi}^u_{n,0,+}(\ell) - \widehat{\Phi}^u_{n,0,-}(\ell) \overset{P_{k_n}}{\Rightarrow} \Phi_{h^*_{2,d}}$ hold with $h^*_{2,d} = h^*_{2,d,+} + h^*_{2,d,-}$ defined in (D.2).*

Proof. To simplify notation, we show this theorem for a sequence $\{n\}$, since all the arguments go through with $\{k_n\}$ in place of $\{n\}$. For the first part, it is sufficient to show the case of $\widehat{\Phi}^u_{n,1,+}$ since

[4] Extending the definition of conditional weak convergence given in Section 2.9 of Van Der Vaart and Wellner (1996) to a sequence of data distributions, $\widehat{\Phi}^u_n \overset{P_n}{\Rightarrow} \Phi$ means for any $\epsilon > 0$, $\lim_{n\to\infty} P_n(\sup_{\{f \in BL\}} |E_u(f(\widehat{\Phi}^u_n)) - E(f(\Phi))| > \epsilon) = 0$, where f maps random element $\Phi(\cdot)$ to $\mathbb{R}$, BL collects f with a bounded Lipschitz constant, and $E_u(\cdot)$ is the expectation of $(U_i : i = 1, \ldots, n)$ conditional on the data.

the arguments for the other cases are the same. We use the same arguments as the proof in Hsu (2016). We define $\hat{\phi}_{n,i,1,+}(\ell) = \sqrt{nh} w^+_{n,i}(g_\ell(Y_i)D_i - \hat{m}_{1,+}(\ell))$, so $\widehat{\Phi}^u_{n,1,+} = \sum_{i=1}^n U_i \cdot \hat{\phi}_{n,i,1,+}(\ell)$.

First, we note that the triangular array $\{\hat{f}_{n,i}(\ell) = U_i \cdot \hat{\phi}_{n,i,1,+}(\ell) : \ \ell \in \mathcal{L}, 1 \leq i \leq n\}$ is manageable with respect to envelope functions $\{\widehat{F}_{n,i} = 2\sqrt{nh}|U_i| \cdot |w^+_{n,i}| : 1 \leq i \leq n\}$. Define $\hat{h}_{2,1,+}(\ell_1, \ell_2) = \sum_{i=1}^n \hat{\phi}_{n,i,1,+}(\ell_1)\hat{\phi}_{n,i,1,+}(\ell_2)$. If we have

$$\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\hat{h}_{2,1,+}(\ell_1,\ell_2) - h^*_{2,1,+}(\ell_1,\ell_2)| \xrightarrow{P_n} 0, \tag{E.16}$$

and

$$nh \sum_{i=1}^n |w^+_{n,i}|^2 \xrightarrow{P_n} M_1, \tag{E.17}$$

$$n^3h^3 \sum_{i=1}^n |w^+_{n,i}|^4 \xrightarrow{P_n} M_2, \tag{E.18}$$

for $M_1, M_2 < \infty$, adopting the proof of Theorem 2.1 of Hsu (2016) yields $\widehat{\Phi}^u_{n,1,+}(\ell) \overset{P_u}{\Rightarrow} \Phi_{h^*_{2,1,+}}(\ell)$, and similarly for $\widehat{\Phi}^u_{n,1,-}(\ell) \overset{P_u}{\Rightarrow} \Phi_{h^*_{2,1,-}}(\ell)$. For the second part, note that $\widehat{\Phi}^u_{\nu_1,n}(\ell) = \widehat{\Phi}^u_{n,1,-}(\ell) - \widehat{\Phi}^u_{n,1,+}(\ell)$ and by the independence of the two simulated processes, we have $\widehat{\Phi}^u_{\nu_1,n}(\ell) \overset{P_u}{\Rightarrow} \Phi_{h^*_{2,1}}(\ell)$.

Hence, the rest of the proof focuses on verifying (E.16) - (E.18). For positive integer $j < \infty$ and nonnegative integer $k < \infty$, a straightforward extension of Lemma 2 in Fan and Gijbels (1992) gives

$$(nh)^{(j-1)} \sum_{i=1}^n |w^+_{n,i}|^j \left(\frac{R_i - r_0}{h_+}\right)^k = \frac{\int_0^\infty K^j(u)(\vartheta_2 - \vartheta_1 u)^j u^k du}{c_+^{j-1}[\vartheta_0\vartheta_2 - \vartheta_1^2]^j} + o_{\mathcal{P}}(1), \tag{E.19}$$

where the first term on the right-hand side is finite, and the assumption that $\mathcal{P}$ shares a fixed distribution for R leads to this convergence being uniform over $\mathcal{P}$. Hence, (E.17) and (E.18) hold, as $\{P_n\} \in \mathcal{P}$.

To show (E.16), it suffices to show

$$\begin{aligned}
&\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\hat{h}_{2,1,+}(\ell_1,\ell_2) - h_{2,P,1,+}(\ell_1,\ell_2)| \\
\leq &\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\hat{h}_{2,1,+}(\ell_1,\ell_2) - \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]| + \sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)] - h_{2,P,1,+}(\ell_1,\ell_2)| \\
=& o_{\mathcal{P}}(1).
\end{aligned} \tag{E.20}$$

The proof of Lemma E.3 shows that the second term in (E.20) converges to zero uniformly in $\mathcal{P}$. We hence focus on showing that the first term in (E.20) is $o_{\mathcal{P}}(1)$.

Rewrite $\hat{\phi}_{n,i,1,+}(\ell)$ as follows by applying the mean value expansion:

$$\begin{aligned}\hat{\phi}_{n,i,1,+}(\ell) &= \sqrt{nh} w_{n,i}^{+} \left[m_{P,1}(\ell, R_i) - \hat{m}_{1,+}(\ell) + \mathcal{E}_{1,i}(\ell) \right] \\ &= w_{n,i}^{+} \hat{a}_1(\ell) + \hat{a}_{2,i}(\ell) + \hat{a}_{3,i}(\ell),\end{aligned}$$

where

$$\begin{aligned}\hat{a}_1(\ell) &\equiv -\sqrt{nh}[\hat{m}_{1,+}(\ell) - m_{P,1,+}(\ell)] \\ \hat{a}_{2,i}(\ell) &\equiv \sqrt{nh} w_{n,i}^{+} \left[h_+ m_{P,+}^{(1)}(\ell, R_i) \left(\frac{R_i - r_0}{h_+} \right) + \frac{h_+^2}{2} m_{P,1}^{(2)}(\ell, \tilde{R}_i) \left(\frac{R_i - r_0}{h_+} \right)^2 \right], \\ \hat{a}_{3,i}(\ell) &\equiv \sqrt{nh} w_{n,i}^{+} \mathcal{E}_{1,i}(\ell).\end{aligned}$$

Then, we have

$$\begin{aligned}\hat{h}_{2,1,+}(\ell_1, \ell_2) = & \underbrace{\hat{a}_1(\ell_1)\hat{a}_1(\ell_2) \sum_{i=1}^{n} (w_{n,i}^{+})^2}_{\text{(i)}} + \underbrace{\sum_{i=1}^{n} \hat{a}_{2,i}(\ell_1)\hat{a}_{2,i}(\ell_2)}_{\text{(ii)}} + \underbrace{\sum_{i=1}^{n} \hat{a}_{3,i}(\ell_1)\hat{a}_{3,i}(\ell_2)}_{\text{(iii)}} \\ & + \underbrace{\sum_{i=1}^{n} w_{n,i}^{+} \left[\hat{a}_1(\ell_1)(\hat{a}_{2,i}(\ell_2) + \hat{a}_{3,i}(\ell_2)) + \hat{a}_1(\ell_2)(\hat{a}_{2,i}(\ell_1) + \hat{a}_{3,i}(\ell_1)) \right]}_{\text{(iv)}} \\ & + \underbrace{\sum_{i=1}^{n} \left[\hat{a}_{2,i}(\ell_1)\hat{a}_{3,i}(\ell_2) + \hat{a}_{2,i}(\ell_2)\hat{a}_{3,i}(\ell_2) \right]}_{\text{(v)}}.\end{aligned}$$

By Lemma E.4 and (E.19), term (i) is $o_{\mathcal{P}}(1)$ uniformly over $\ell_1, \ell_2 \in \mathcal{L}$. By Assumption D.1 (v), the absolute value of term (ii) can be bounded by $\left\{ 2M(nh) \sum_{i=1}^{n} (w_{n,i}^{+})^2 \left[\left(\frac{R_i - r_0}{h_+} \right) + \left(\frac{R_i - r_0}{h_+} \right)^2 \right] \right\} \cdot (h_+ \vee h_+^2)$ uniformly over $\ell_1, \ell_2 \in \mathcal{L}$, which is $o_{\mathcal{P}}(1)$ by (E.19) and $h_+ \to 0$. To examine term (iv), note that

$$
\begin{aligned}
&\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\sum_{i=1}^{n} w_{n,i}^{+}\hat{a}_1(\ell_1)\hat{a}_{2,i}(\ell_2)\right| \\
\leq &(nh)^{-1/2}\sup_{\{\ell\in\mathcal{L}\}}|\hat{a}_1(\ell)|\cdot 2M(nh)\sum_{i=1}(w_{n,i}^{+})^2\left|\left(\frac{R_i-r_0}{h_+}\right)+\left(\frac{R_i-r_0}{h_+}\right)^2\right|\cdot(h_+\vee h_+^2) \\
= &o_{\mathcal{P}}(1),
\end{aligned}
$$

where the final line follows by Lemma E.4, equation (E.19), $nh\to\infty$, and $h_+\to 0$. Note also that

$$
\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\sum_{i=1}^{n} w_{n,i}^{+}\hat{a}_1(\ell_1)\hat{a}_{3,i}(\ell_2)\right| \leq (nh)^{-1}\sup_{\{\ell\in\mathcal{L}\}}|\hat{a}_1(\ell)|\cdot\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}^{n}(w_{n,i}^{+})^2\mathcal{E}_{1,i}(\ell)\right|.
$$

The proof of (E.11) in Lemma E.2 can be extended to claim the following Bahadur representation:

$$
\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}^{n}(w_{n,i}^{+})^2\mathcal{E}_{1,i}(\ell)-\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(w_i^{+})^2\mathcal{E}_{1,i}(\ell)\right|=o_{\mathcal{P}}(1).
$$

As in the proof of Lemma E.4, FCLT applied to $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(w_i^{+})^2\mathcal{E}_{1,i}(\ell)$ shows $\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}^{n}(w_{n,i}^{+})^2\mathcal{E}_{1,i}(\ell)\right|$ is stochastically bounded uniformly in $\mathcal{P}$. Combining this with Lemma E.4 and $nh\to\infty$, we obtain $\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\sum_{i=1}^{n} w_{n,i}^{+}\hat{a}_1(\ell_1)\hat{a}_{3,i}(\ell_2)\right|=o_{\mathcal{P}}(1)$. This implies term (iv) is $o_{\mathcal{P}}(1)$. Regarding term (v), we have

$$
\begin{aligned}
&\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\sum_{i=1}^{n}\hat{a}_{2,i}(\ell_1)\hat{a}_{3,i}(\ell_2)\right| \\
\leq &M(nh)^{-1/2}(h_+\vee h_+^2)\times\left\{\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}(w_{n,i}^{+})^2\left(\frac{R_i-r_0}{h_+}\right)\mathcal{E}_{1,i}(\ell)\right|\right. \\
&\left.+\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}(w_{n,i}^{+})^2\left(\frac{R_i-r_0}{h_+}\right)^2\mathcal{E}_{1,i}(\ell)\right|\right\}. \qquad \text{(E.21)}
\end{aligned}
$$

Similar to the proof of (E.11) in Lemma E.2, the two terms in the curly brackets of (E.21) admit the following Bahadur representation: for positive integer $j<\infty$,

$$
\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}^{n}(w_{n,i}^{+})^2\left(\frac{R_i-r_0}{h_+}\right)^j\mathcal{E}_{1,i}(\ell)-\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(w_i^{+})^2\left(\frac{R_i-r_0}{h_+}\right)^j\mathcal{E}_{1,i}(\ell)\right|=o_{\mathcal{P}}(1).
$$

Similar to the proof of Lemma E.4, the FCLT applied to $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(w_i^+)^2\left(\frac{R_i-r_0}{h_+}\right)^j\mathcal{E}_{1,i}(\ell)$ shows that it is stochastically bounded uniformly in $\mathcal{P}$. Accordingly, since $nh\to\infty$ and $h_+\to 0$, the upper bound in (E.21) is $o_{\mathcal{P}}(1)$.

We now show term (iii) is the leading term such that $\sup_{\{\ell_1,\ell_2\}}\left|\sum_{i=1}^{n}\hat{a}_{3,i}(\ell_1)\hat{a}_{3,i}(\ell_2)-\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]\right| = o_{\mathcal{P}}(1)$ holds. Modifying the proof of (E.11) by replacing $f_{n,i}^j(\ell)$ with $f_{n,i}(\ell_1,\ell_2)$ defined in the proof of Lemma E.3, we obtain the Bahadur-type uniform approximation,

$$\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\sum_{i=1}^{n}\hat{a}_{3,i}(\ell_1)\hat{a}_{3,i}(\ell_2)-(nh)^{-1}\sum_{i=1}^{n}(w_i^+)^2\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)\right| = o_{\mathcal{P}}(1).$$

We hence aim to verify $\left|(nh)^{-1}\sum_{i=1}^{n}(w_i^+)^2\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)-\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]\right| = o_{\mathcal{P}}(1)$. Note that

$$\begin{aligned}&\mathbb{E}_P\left[\left|(nh)^{-1}\sum_{i=1}^{n}(w_i^+)^2\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)-\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]\right|\right]\\ \leq&\frac{1}{[c_+f_R^+(r_0)(\vartheta_0\vartheta_2-\vartheta_1^2)]^2}\mathbb{E}_P\left[\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^{n}f_{n,i}(\ell_1,\ell_2)-\mathbb{E}_P(f_{n,i}(\ell_1,\ell_2))\right|\right],\end{aligned}\tag{E.22}$$

where $f_{n,i}(\ell_1,\ell_2)$ is as defined in the proof of Lemma E.3. Note that this upper bound coincides with (E.13). Hence, the proof of Lemma E.3 yields $\left|(nh)^{-1}\sum_{i=1}^{n}(w_i^+)^2\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)-\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]\right| = o_{\mathcal{P}}(1)$. □

Lemma E.6. *Suppose that Assumptions D.1 and D.2 hold. Let $\{P_n\}$ be a sequence of data generating processes in $\mathcal{P}$. For any subsequence $\{k_n\}$ of $\{n\}$ such that for $d=0,1$ and $\star=+,-$, $\lim_{n\to\infty}d(h_{2,P_{k_n},d,\star},h^*_{2,d,\star})=0$ for some $h^*_{2,d,\star}\in\mathcal{H}_2$, then for $d=0,1$, $\sup_{\{\ell\in\mathcal{L}\}}|\hat{\sigma}^{-1}_{d,\xi}(\ell)-\sigma^{-1}_{d,P_{k_n},\xi}(\ell)|\overset{P_{k_n}}{\to}0$, where $\sigma_{d,P_{k_n},\xi}(\ell)\equiv\max\{h_{2,P_{k_n},d}(\ell,\ell),\xi\}$.*

Proof. Using the notation defined in the proof of Lemma E.5, we note, for $d=0,1$,

$$\hat{\sigma}_{d,\xi}(\ell)=\max\{\xi,\sqrt{\hat{h}^2_{2,d,+}(\ell,\ell)+\hat{h}^2_{2,d,-}(\ell,\ell)}\}.$$

The uniform convergence of (E.20) shown in the proof of Lemma E.5 implies

$$\sup_{\{\ell \in \mathcal{L}\}} \left| \sqrt{\hat{h}^2_{2,d,+}(\ell,\ell) + \hat{h}^2_{2,d,-}(\ell,\ell)} - h_{2,P_{k_n},d}(\ell,\ell) \right| \xrightarrow{P_{k_n}} 0.$$

Due to the fact that the maximum operator is a continuous functional and the fact that $\sigma_{d,k_n,\xi}$ is bounded away from zero, $\sup_{\{\ell \in \mathcal{L}\}} |\hat{\sigma}^{-1}_{d,\xi}(\ell) - \sigma^{-1}_{d,P_{k_n},\xi}(\ell)| \xrightarrow{P_{k_n}} 0$ follows by the continuous mapping theorem. □

Remark: Note that the results in Lemmas E.4 and E.5 hold jointly for $d = 0$ and $d = 1$. We omit the results and proofs for brevity.

Proof of Theorem D.1: Having shown Lemmas E.4 to E.6, we prove the current theorem adapting the proof of Theorem 2 in Andrews and Shi (2013). Let $\mathcal{H}_1$ denote the set of measurable functions mapping $\mathcal{L}$ to $[-\infty, 0]$. Let $h = (h_1, h_2)$, where $h_1 = (h_{1,0}, h_{1,1}) \in \mathcal{H}_1 \times \mathcal{H}_1$ and $h_2 = (h_{2,0}, h_{2,1}) \in \mathcal{H}_2 \times \mathcal{H}_2$. Define

$$T(h) = \sup_{\{d \in \{0,1\}, \ell \in \mathcal{L}\}} \frac{\Phi_{h_{2,d}}(\ell)}{\sigma_{d,P_{k_n},\xi}(\ell)} + h_{1,d}(\ell).$$

Define $c_0(h_1, h_2, \alpha)$ as the (1-α)-th quantile of $T(h)$. Similar to Lemma A2 of Andrews and Shi (2013), we can show that for any $\zeta > 0$,

$$\limsup_{n\to\infty} \sup_{\{P \in \mathcal{P}_0:\ d \in \{0,1\}, h_{2,P,d,+}, h_{2,P,d,-} \in \mathcal{H}_{2,cpt}\}} P\Big(\widehat{S}_n > c_0(h^P_{1,n}, h_{2,P}, \alpha) + \zeta\Big) \leq \alpha, \tag{E.23}$$

where $h^P_{1,n} = (h^P_{1,0,n}, h^P_{1,1,n})$ such that for $d = 0, 1$, $h^P_{1,d,n} = \sqrt{nh}\nu_{P,d}$ which belongs to $\mathcal{H}_1$ under $P \in \mathcal{P}_0$. Also, similar to Lemma A3 of Andrews and Shi (2013), we can show that for all $\alpha < 1/2$

$$\limsup_{n\to\infty} \sup_{\{P \in \mathcal{P}_0:\ d \in \{0,1\}, h_{2,P,d,+}, h_{2,P,d,-} \in \mathcal{H}_{2,cpt}\}} P\Big(c_0(\psi_n, h_{2,P}, \alpha) < c_0(h^P_{1,n}, h_{2,P}, \alpha)\Big) = 0, \tag{E.24}$$

where $\psi_n(\ell) = (\psi_{n,0}(\ell), \psi_{n,1}(\ell))$, $\ell \in \mathcal{L}$, as defined in Algorithm 1 in the main text. To complete the proof, it suffices to show that for all $0 < \zeta < \eta$,

$$\limsup_{n\to\infty} \sup_{\{P \in \mathcal{P}_0:\ d \in \{0,1\}, h_{2,P,d,+}, h_{2,P,d,-} \in \mathcal{H}_{2,cpt}\}} P\Big(\hat{c}_\eta(\alpha) < c_0(\psi_n, h_{2,P}, 1-\alpha) + \zeta\Big) = 0. \tag{E.25}$$

Let $\{P_n \in \mathcal{P}_0 : n \geq 1\}$ be a sequence for which the probability in equation (E.25) evaluated at P_n differs from its supremum over $P \in \mathcal{P}_0$ by $\delta_n > 0$ or less and $\lim_{n\to\infty} \delta_n = 0$. By the definition of lim sup, such a sequence always exists. Therefore, it is equivalent to show that for $0 < \zeta < \eta$,

$$\lim_{n\to\infty} P_n\left(\hat{c}_{n,\eta}(\alpha) < c_0(\psi_n, h_{2,P_n}, \alpha) + \zeta\right) = 0, \tag{E.26}$$

where $\hat{c}_{n,\eta}(\alpha)$ is $\hat{c}_\eta(\alpha)$ with its dependence on the sample size along the sequence of sampling distributions $\{P_n\}$ notated explicitly. The limit on the left-hand side of (E.26) exists by the construction of $\{P_n\}$, and we want to show it is equal to 0. Given that we restrict $h_{2,P,d,+}$ and $h_{2,P,d,-}$ to a compact set $\mathcal{H}_{2,cpt}$, there exists a subsequence $\{k_n\}$ of $\{n\}$ such that for $d = 0, 1$, $h_{2,P_{k_n},d,+}$ and $h_{2,P_{k_n},d,-}$ converge to $h^*_{2,d,+}$ and $h^*_{2,d,-}$, respectively, for some $h^*_{2,d,+}, h^*_{2,d,-} \in \mathcal{H}_{2,cpt}$, $d = 0, 1$.

By Lemmas E.4, E.5, and E.6,

$$\begin{aligned}
&\sqrt{k_n h}(\hat{\nu}_d(\cdot) - \nu_{P_{k_n},d}(\cdot)) \Rightarrow \Phi_{h^*_{2,d}}(\cdot),\\
&\widehat{\Phi}^u_{\nu_d,k_n}(\cdot) \overset{P_{k_n}}{\Rightarrow} \Phi'_{h^*_{2,d}}(\cdot),\\
&\sup_{\{\ell\in\mathcal{L}\}} |\hat{\sigma}^{-1}_{d,\xi}(\ell) - \sigma^{-1}_{d,P_{k_n},\xi}(\ell)| \overset{P_{k_n}}{\to} 0
\end{aligned}$$

for $d = 0, 1$, where $\Phi'_{h^*_{2,d}}(\ell)$ is an independent copy of $\Phi_{h^*_{2,d}}(\ell)$. By the almost sure representation theorem (e.g., Theorem 9.4 of Pollard (1990)), there exists a probability space and random objects $(\tilde{\nu}_d(\cdot), \widetilde{\Phi}^u_{\nu_d,k_n}(\cdot), \tilde{\sigma}_{d,\xi}(\cdot))$ and $(\widetilde{\Phi}_{h^*_{2,d}}(\cdot), \widetilde{\Phi}'_{h^*_{2,d}}(\cdot))$, $d = 0, 1$, defined on it, such that they have the same probability distribution as $(\hat{\nu}_d(\cdot), \widehat{\Phi}^u_{\nu_d,k_n}(\cdot), \hat{\sigma}_{d,\xi}(\cdot))$ and $(\Phi_{h^*_{2,d}}(\cdot), \Phi'_{h^*_{2,d}}(\cdot))$, $d = 0, 1$, and satisfy

$$\sup_{d\in\{0,1\},\ell\in\mathcal{L}} \left\| \begin{pmatrix} \sqrt{k_n h}(\tilde{\nu}_d(\cdot) - \nu_{P_{k_n},d}(\cdot)) \\ \widetilde{\Phi}^u_{\nu_d,k_n}(\ell) \\ \tilde{\sigma}_{d,\xi}(\ell) \end{pmatrix} - \begin{pmatrix} \widetilde{\Phi}_{h^*_{2,d}}(\ell) \\ \widetilde{\Phi}'_{h^*_{2,d}}(\ell) \\ \sigma_{d,P_{k_n},\xi}(\ell) \end{pmatrix} \right\| \to 0, \tag{E.27}$$

as $n \to \infty$, a.s. We also define an analogue of $\psi_{n,d}$ as

$$\tilde{\psi}_{k_n,d}(\cdot) = -B_{k_n} \cdot 1\left\{ \frac{\sqrt{k_n h} \cdot \tilde{\nu}_d(\cdot)}{\tilde{\sigma}_{d,\xi}(\ell)} < -a_{k_n} \right\},$$

and let $\tilde{c}_{k_n,\eta}(\alpha)$ be the $(1-\alpha+\eta)$-th quantile of $\sup_{d\in\{0,1\},\ell\in\mathcal{L}} \left\{ \frac{\widetilde{\Phi}^u_{\nu_d,k_n}(\ell)}{\tilde{\sigma}_{d,\xi}(\ell)} + \tilde{\psi}_{k_n,d}(\ell) \right\}$ plus η, which by construction shares the probability law with $\hat{c}_{k_n,\eta}(\alpha)$.

Let Ω_1 be the subset of the sample space such that the convergence of (E.27) holds. Following the proof of Theorem 1 of Andrews and Shi (2013), we can show an inequality analogous to (12.28) in Andrews and Shi (2013); for any sequence $\{\tilde{a}_{k_n}\} \in \mathbb{R}$ that may depend on h_1 and P, and for any $\zeta_1 > 0$,

$$\limsup_{n\to\infty} \sup_{\{h_{1,0},h_{1,1}\in\mathcal{H}_1\}} \mathbb{P}\left(\sup_{\{d\in\{0,1\},\ell\in\mathcal{L}\}} \frac{\widetilde{\Phi}^u_{\nu_d,k_n}(\ell)}{\tilde{\sigma}_{d,\xi}(\ell)} + h_{1,d}(\ell) \leq \tilde{a}_{k_n}\right) - \mathbb{P}\left(\sup_{\{d\in\{0,1\},\ell\in\mathcal{L}\}} \frac{\widetilde{\Phi}_{h^*_{2,d}}(\ell)}{\sigma_{d,P_{k_n},\xi}(\ell)} + h_{1,d}(\ell) \leq \tilde{a}_{k_n} + \zeta_1\right) \leq 0, \tag{E.28}$$

where $\mathbb{P}$ denotes the measure of the probability space that $(\tilde{\nu}_{d,k_n}, \widetilde{\Phi}^u_{\nu_d,k_n}(\cdot), \tilde{\sigma}_{d,\xi}(\cdot) : n = 1,2,\ldots,)$ are defined on. By (E.28) and a similar argument to Lemma A5 of Andrews and Shi (2013), we have that for all $0 < \zeta < \zeta_1 < \eta$ and $\omega \in \Omega_1$,

$$\liminf_{n\to\infty} \tilde{c}_{k_n,\eta}(\alpha)(\omega) \geq c_0(\tilde{\psi}_{k_n}, h_{2,P_{k_n}}, \alpha) + \zeta_1. \tag{E.29}$$

Given that $P(\Omega_1) = 1$, this implies

$$\lim_{n\to\infty} \mathbb{P}(\tilde{c}_{k_n,\eta}(\alpha) < c_0(\tilde{\psi}_{k_n}, h_{2,P_{k_n}}, \alpha) + \zeta) = 0. \tag{E.30}$$

Since $(\tilde{c}_{k_n,\eta}(\alpha), \tilde{\psi}_{k_n})$ share the probability law with $(\hat{c}_{k_n,\eta}(\alpha), \psi_{k_n})$, we also have

$$\lim_{n\to\infty} P_{k_n}(\hat{c}_{k_n,\eta}(\alpha) < c_0(\psi_{k_n}, h_{2,P_{k_n}}, \alpha) + \zeta) = 0. \tag{E.31}$$

For any convergent sequence $\{b_n\}$, if there exists a subsequence $\{b_{k_n}\}$ converging to b, then $\{b_n\}$ converges to b as well. Therefore, (E.31) is sufficient for (E.26). Theorem D.1(a) is shown by combining (E.23), (E.24) and (E.25).

We next show Theorem D.1(b). Under Assumption D.5, consider pointwise asymptotics under $P_c \in \mathcal{P}_0$. Similarly to the proof of Proposition 1 of Barrett and Donald (2003) and Lemma 1 of Donald and Hsu (2016), we can show $\widehat{S}_n \xrightarrow{d} \sup_{\{(d,\ell):\, \ell\in\mathcal{L}^o_{P_c,d}\}} \Phi_{h_{2,P_c,d}}(\ell)/\sigma_{d,P_c,\xi}(\ell)$ whose CDF is denoted by $H(a)$. By Tsirel'son (1975), if either $\Phi_{h_{2,P_c,0}}$ restricted to $\mathcal{L}^o_{P_c,0} \times \mathcal{L}^o_{P_c,0}$ or $\Phi_{h_{2,P_c,1}}$ restricted to $\mathcal{L}^o_{P_c,1} \times \mathcal{L}^o_{P_c,1}$ is not a zero function, then $H(a)$ is continuous and strictly increasing for $a \in (0,\infty)$ and $H(0) > 1/2$.

Following the proof of Theorem 2(b) of Andrews and Shi (2013), we can show that $\hat{c}_\eta(\alpha) \xrightarrow{P_c} q_c(1-\alpha+\eta)+\eta$ where $q_c(1-\alpha+\eta)$ denotes the $(1-\alpha+\eta)$-th quantile of $\sup_{(d,\ell):\ \ell\in\mathcal{L}^o_{P_c,d}} \Phi_{h_{2,P_c,d}}(\ell)/\sigma_{d,P_c,\xi}(\ell)$. Because $H(a)$ is continuous at $q_c(1-\alpha)$, we have $\lim_{\eta\to 0} q_c(1-\alpha+\eta)+\eta = q_c(1-\alpha)$. This suffices to show that $\lim_{n\to\infty} P_c(\widehat{S}_n > \hat{c}_\eta(\alpha)) = \alpha$. Combined with the claim of (a) in the current theorem, Theorem D.1(b) holds. □

Proof of Theorem D.2: Under any fixed alternative P_A, the exists (d,ℓ^*) such that $\nu_d(\ell^*) > 0$, so $\widehat{S}_n/\sqrt{nh} \geq \nu_d(\ell^*)/\sigma_{d,P_A,\xi}(\ell^*)$ in probability that implies that $\widehat{S}_n$ will diverge to positive infinity in probability. Also, the $\hat{c}_\eta(\alpha)$ is bounded in probability, so $\lim_{n\to\infty} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) = 1$. □

Proof of Theorem D.3: Define $\mathcal{L}^{++}_d = \{\ell \in \mathcal{L}^o_{P_c,d} : \delta_d(\ell) > 0\}$. For $d = 1, 0$, let $\sigma^*_{d,\xi}(\ell) \equiv \max\{\xi, \sqrt{(h^*_{2,d,+}(\ell,\ell))^2 + (h^*_{2,d,-}(\ell,\ell))^2}\}$ be the limiting trimmed variance along the sequence of local alternatives $\{P_n\}$. It can be shown that $\widehat{S}_n \overset{P_\eta}{\Rightarrow} \sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} (\Phi_{h^*_{2,d}}(\ell) + \delta_d(\ell))/\sigma^*_{d,\xi}(\ell)$ and $\hat{c}_\eta(\alpha) \xrightarrow{P_n} c_\eta + \eta$ where c_η is the $(1-\alpha+\eta)$-th quantile of $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} \Phi_{h^*_{2,d}}(\ell)/\sigma^*_{d,\xi}(\ell)$. Then, the limit of the local power is

$$P(\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} (\Phi_{h^*_{2,d}}(\ell) + \delta_d(\ell))/\sigma^*_{d,\xi}(\ell) \geq c_\eta + \eta).$$

We need to consider the following two cases: (a) both $h^*_{2,0}$ restricted to $\mathcal{L}^o_{P_c,0} \times \mathcal{L}^o_{P_c,0}$ and $h^*_{2,1}$ restricted to $\mathcal{L}^o_{P_c,1} \times \mathcal{L}^o_{P_c,1}$ are zero functions and (b) at least one of $h^*_{2,0}$ restricted to $\mathcal{L}^o_{P_c,0} \times \mathcal{L}^o_{P_c,0}$ or $h^*_{2,1}$ restricted to $\mathcal{L}^o_{P_c,1} \times \mathcal{L}^o_{P_c,1}$ is not a zero function.

For case (a), because $h^*_{2,0}$ restricted to $\mathcal{L}^o_{P_c,0} \times \mathcal{L}^o_{P_c,0}$ and $h^*_{2,1}$ restricted to $\mathcal{L}^o_{P_c,1} \times \mathcal{L}^o_{P_c,1}$ are zero functions, then $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} |\Phi_{h^*_{2,d}}(\ell)| \xrightarrow{P_\eta} 0$ and $\widehat{S}_n \xrightarrow{P_\eta} \sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} \delta_d(\ell)/\sigma^*_{d,\xi}(\ell) > 0$. Also, it is true that $c_\eta + \eta = \eta$ and when $\eta \to 0$, we have $P(\widehat{S}_n > \eta) = 1$ when η is small enough.

For case (b), when at least one of $h^*_{2,0}$ restricted to $\mathcal{L}^o_{P_c,0} \times \mathcal{L}^o_{P_c,0}$ or $h^*_{2,1}$ restricted to $\mathcal{L}^o_{P_c,1} \times \mathcal{L}^o_{P_c,1}$ is not a zero function, then by the continuity of the distribution of $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} (\Phi_{h^*_{2,d}}(\ell) + \delta_d(\ell))/\sigma^*_{d,\xi}(\ell)$ and $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} \Phi_{h^*_{2,d}}(\ell)/\sigma^*_{d,\xi}(\ell)$,

$$\lim_{\eta\to 0} P(\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} (\Phi_{h^*_{2,d}}(\ell) + \delta_d(\ell))/\sigma^*_{d,\xi}(\ell) \geq c_\eta + \eta) = P(\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} (\Phi_{h^*_{2,d}}(\ell) + \delta_d(\ell))/\sigma^*_{d,\xi}(\ell) \geq c),$$

where c is the $(1-\alpha)$-th quantile of $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}}\Phi_{h^*_{2,d}}(\ell)/\sigma^*_{d,\xi}(\ell)$. By assumption, $\delta_d(\ell)$ is nonnegative if $\ell\in\mathcal{L}^o_{P_c,d}$, so $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}}(\Phi_{h^*_{2,d}}(\ell)+\delta_d(\ell))/\sigma^*_{d,\xi}(\ell)$ first order stochastically dominates $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}}\Phi_{h^*_{2,d}}(\ell)/\sigma^*_{d,\xi}(\ell)$ and it follows that

$$P(\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}}(\Phi_{h^*_{2,d}}(\ell)+\delta_d(\ell))/\sigma^*_{d,\xi}(\ell)\geq c)\geq\alpha.$$

This completes the proof for Theorem D.3. □

Appendix F. Additional Simulation Results for Section 4

In this section, we report additional simulation results. Tables F.1 to F.3 report detailed results for the size properties of our test using data-driven choices of MSE-optimal bandwidths (AI, IK, and CCT). We consider undersmoothing, an MSE-optimal plus RBC implementation, and a CER-optimal plus RBC implementation, respectively. For the CER implementation, we also implement the data-driven plug-in bandwidth (reported in the DPI column of the relevant tables) as suggested by (see Calonico, Cattaneo, and Farrell, 2020, Section 4.2). Tables F.4 to F.6 show the simulated power properties for the four specifications violating the null.

Table F.1. Size Properties (with undersmoothing)

		AI			IK			CCT		
DGP	n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Size1	1000	0.013	0.060	0.112	0.005	0.02	0.061	0.004	0.019	0.054
	2000	0.025	0.071	0.126	0.003	0.025	0.064	0.005	0.034	0.065
	4000	0.018	0.078	0.140	0.003	0.038	0.084	0.006	0.035	0.074
	8000	0.021	0.065	0.120	0.012	0.045	0.095	0.01	0.033	0.085
Size2	1000	0.018	0.063	0.115	0.003	0.014	0.046	0.003	0.012	0.040
	2000	0.014	0.064	0.116	0.008	0.035	0.074	0.006	0.031	0.062
	4000	0.02	0.064	0.119	0.006	0.041	0.074	0.007	0.038	0.080
	8000	0.013	0.06	0.101	0.005	0.039	0.072	0.006	0.036	0.077

TABLE F.2. Size Properties (MSE-optimal + RBC)

DGP	n	AI			IK			CCT		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
Size1	1000	0.006	0.037	0.079	0.004	0.016	0.041	0.003	0.017	0.036
	2000	0.016	0.058	0.108	0.002	0.018	0.043	0.006	0.022	0.051
	4000	0.020	0.067	0.119	0.008	0.049	0.097	0.002	0.022	0.065
	8000	0.020	0.066	0.125	0.007	0.051	0.085	0.008	0.037	0.077
Size2	1000	0.013	0.039	0.083	0.002	0.008	0.034	0.002	0.015	0.029
	2000	0.012	0.051	0.122	0.004	0.033	0.061	0.002	0.021	0.044
	4000	0.013	0.077	0.130	0.010	0.040	0.078	0.007	0.037	0.077
	8000	0.016	0.065	0.125	0.006	0.042	0.090	0.003	0.040	0.085

TABLE F.3. Size Properties (CER-optimal + RBC)

DGP	n	AI+adjustment			IK+adjustment			CCT+adjustment			DPI		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Size1	1000	0.016	0.055	0.11	0.004	0.025	0.056	0.004	0.024	0.049	0.011	0.042	0.071
	2000	0.026	0.076	0.134	0.005	0.022	0.054	0.005	0.024	0.059	0.016	0.044	0.084
	4000	0.018	0.086	0.140	0.004	0.043	0.101	0.006	0.035	0.068	0.016	0.054	0.110
	8000	0.023	0.057	0.113	0.008	0.045	0.09	0.01	0.036	0.073	0.015	0.064	0.121
Size2	1000	0.018	0.067	0.109	0.002	0.015	0.043	0.003	0.014	0.037	0.005	0.034	0.073
	2000	0.013	0.06	0.117	0.009	0.032	0.078	0.005	0.024	0.055	0.016	0.059	0.121
	4000	0.017	0.06	0.118	0.011	0.042	0.083	0.008	0.039	0.078	0.017	0.062	0.115
	8000	0.013	0.057	0.101	0.008	0.044	0.079	0.007	0.035	0.085	0.017	0.057	0.105

TABLE F.4. Power Properties (undersmoothing)

DGP	n	AI 1%	AI 5%	AI 10%	IK 1%	IK 5%	IK 10%	CCT 1%	CCT 5%	CCT 10%
Power1	1000	0.096	0.215	0.303	0.074	0.174	0.244	0.047	0.111	0.183
	2000	0.235	0.439	0.557	0.218	0.403	0.504	0.113	0.256	0.359
	4000	0.561	0.753	0.816	0.541	0.744	0.824	0.373	0.604	0.693
	8000	0.908	0.962	0.977	0.926	0.975	0.987	0.785	0.907	0.94
Power2	1000	0.051	0.122	0.221	0.014	0.061	0.122	0.011	0.052	0.101
	2000	0.1	0.271	0.383	0.065	0.194	0.296	0.03	0.14	0.214
	4000	0.323	0.554	0.678	0.293	0.511	0.647	0.15	0.342	0.438
	8000	0.741	0.885	0.928	0.752	0.888	0.934	0.503	0.732	0.818
Power3	1000	0.052	0.164	0.28	0.037	0.123	0.197	0.023	0.078	0.134
	2000	0.128	0.299	0.432	0.114	0.257	0.393	0.062	0.17	0.261
	4000	0.359	0.573	0.693	0.283	0.51	0.638	0.187	0.383	0.494
	8000	0.758	0.883	0.938	0.724	0.87	0.922	0.526	0.734	0.831
Power4	1000	0.032	0.099	0.175	0.005	0.05	0.089	0.001	0.024	0.053
	2000	0.058	0.172	0.252	0.031	0.123	0.209	0.017	0.06	0.134
	4000	0.119	0.264	0.399	0.106	0.268	0.383	0.043	0.144	0.24
	8000	0.331	0.55	0.673	0.322	0.54	0.656	0.144	0.326	0.458

TABLE F.5. Power Properties (MSE optimal + RBC)

DGP	n	AI 1%	AI 5%	AI 10%	IK 1%	IK 5%	IK 10%	CCT 1%	CCT 5%	CCT 10%
Power1	1000	0.031	0.103	0.18	0.022	0.081	0.153	0.014	0.05	0.089
	2000	0.088	0.221	0.334	0.084	0.207	0.299	0.033	0.115	0.18
	4000	0.277	0.476	0.577	0.267	0.459	0.549	0.154	0.305	0.414
	8000	0.627	0.812	0.87	0.664	0.82	0.877	0.433	0.654	0.754
Power2	1000	0.024	0.086	0.149	0.005	0.023	0.056	0.002	0.022	0.05
	2000	0.04	0.135	0.233	0.029	0.099	0.173	0.009	0.052	0.095
	4000	0.13	0.293	0.406	0.1	0.246	0.363	0.043	0.142	0.226
	8000	0.396	0.624	0.725	0.4	0.622	0.724	0.186	0.391	0.528
Power3	1000	0.035	0.106	0.176	0.012	0.063	0.12	0.006	0.027	0.071
	2000	0.05	0.174	0.289	0.052	0.154	0.247	0.025	0.079	0.147
	4000	0.182	0.361	0.484	0.118	0.289	0.416	0.07	0.183	0.296
	8000	0.492	0.694	0.781	0.419	0.64	0.744	0.261	0.466	0.582
Power4	1000	0.011	0.057	0.114	0.001	0.024	0.053	0.001	0.017	0.038
	2000	0.03	0.118	0.204	0.017	0.06	0.134	0.01	0.042	0.104
	4000	0.066	0.181	0.288	0.043	0.144	0.24	0.016	0.079	0.168
	8000	0.149	0.341	0.478	0.144	0.326	0.458	0.067	0.201	0.311

TABLE F.6. Power Property (CER optimal + RBC)

DGP	n	AI+adjustment 1%	5%	10%	IK+adjustment 1%	5%	10%	CCT+adjustment 1%	5%	10%	DPI 1%	5%	10%
Power1	1000	0.088	0.225	0.312	0.048	0.143	0.222	0.031	0.09	0.15	0.038	0.102	0.162
	2000	0.228	0.439	0.551	0.171	0.316	0.423	0.081	0.205	0.292	0.071	0.174	0.263
	4000	0.562	0.749	0.816	0.425	0.629	0.719	0.281	0.486	0.609	0.174	0.354	0.487
	8000	0.918	0.964	0.977	0.83	0.935	0.957	0.666	0.831	0.877	0.485	0.658	0.743
Power2	1000	0.045	0.133	0.214	0.008	0.046	0.096	0.006	0.045	0.087	0.015	0.054	0.094
	2000	0.108	0.266	0.384	0.052	0.156	0.244	0.023	0.097	0.178	0.037	0.097	0.179
	4000	0.316	0.56	0.672	0.19	0.399	0.519	0.099	0.248	0.376	0.083	0.215	0.316
	8000	0.744	0.889	0.927	0.587	0.793	0.857	0.368	0.598	0.707	0.225	0.427	0.550
Power3	1000	0.065	0.159	0.273	0.026	0.107	0.171	0.014	0.061	0.118	0.022	0.091	0.142
	2000	0.129	0.306	0.439	0.091	0.209	0.319	0.048	0.128	0.221	0.051	0.151	0.242
	4000	0.368	0.581	0.693	0.205	0.421	0.548	0.143	0.321	0.426	0.122	0.277	0.383
	8000	0.751	0.888	0.933	0.597	0.781	0.859	0.432	0.64	0.734	0.264	0.490	0.606
Power4	1000	0.025	0.101	0.171	0.005	0.036	0.078	0.004	0.027	0.059	0.014	0.055	0.095
	2000	0.055	0.175	0.263	0.027	0.092	0.188	0.016	0.074	0.143	0.034	0.080	0.148
	4000	0.113	0.265	0.403	0.073	0.201	0.324	0.03	0.138	0.23	0.036	0.131	0.216
	8000	0.324	0.545	0.681	0.233	0.438	0.569	0.121	0.283	0.425	0.084	0.245	0.349

Appendix G. Additional Empirical Results for Section 5

Table G.1. Jump Size of the Propensity Score (different choice of h)

	3	5	AL	IK	CCT
Grade 4					
Cut-off 40	0.60	0.65	0.80	0.60	0.77
Cut-off 80	0.47	0.48	0.64	0.51	0.56
Cut-off 120	0.30	0.44	0.56	0.37	0.55
Grade 5					
Cut-off 40	0.53	0.47	0.66	0.47	0.47
Cut-off 80	0.42	0.42	0.55	0.42	0.49
Cut-off 120	0.31	0.29	0.44	0.31	0.42

Table G.2. Testing Results for Israeli School Data: p-values, $\xi = 0.0316$

	3	5	AL	IK	CCT	MSE-RBC	CER-RBC
g4math							
Cut-off 40	0.986	0.934	0.764	0.978	0.968	0.975	0.974
Cut-off 80	0.909	0.865	0.715	0.944	0.888	0.776	0.973
Cut-off 120	0.443	0.702	0.665	0.604	0.568	0.610	0.646
g4verb							
Cut-off 40	0.928	0.627	0.465	0.641	0.529	0.574	0.455
Cut-off 80	0.911	0.883	0.185	0.906	0.720	0.300	0.855
Cut-off 120	0.935	0.683	0.474	0.730	0.186	0.222	0.131
g5math							
Cut-off 40	0.876	0.282	0.482	0.631	0.609	0.903	0.241
Cut-off 80	0.516	0.446	0.930	0.482	0.765	0.814	0.708
Cut-off 120	0.939	0.827	0.626	0.883	0.838	0.832	0.731
g5verb							
Cut-off 40	0.594	0.893	0.953	0.900	0.938	0.960	0.957
Cut-off 80	0.510	0.692	0.504	0.519	0.929	0.956	0.979
Cut-off 120	0.696	0.811	0.601	0.699	0.774	0.729	0.762

TABLE G.3. Testing Results for Israeli School Data: p-values, $\xi = 0.1706$

	3	5	AL	IK	CCT	MSE-RBC	CER-RBC
g4math							
Cut-off 40	0.986	0.934	0.945	0.978	0.959	0.961	0.965
Cut-off 80	0.909	0.865	0.713	0.944	0.878	0.763	0.953
Cut-off 120	0.443	0.702	0.660	0.565	0.540	0.599	0.646
g4verb							
Cut-off 40	0.924	0.627	0.451	0.637	0.517	0.571	0.469
Cut-off 80	0.911	0.883	0.185	0.906	0.688	0.281	0.836
Cut-off 120	0.935	0.683	0.471	0.730	0.183	0.249	0.125
g5math							
Cut-off 40	0.861	0.275	0.481	0.623	0.600	0.887	0.236
Cut-off 80	0.516	0.429	0.916	0.479	0.762	0.797	0.714
Cut-off 120	0.939	0.827	0.624	0.883	0.836	0.808	0.746
g5verb							
Cut-off 40	0.594	0.893	0.953	0.934	0.938	0.944	0.946
Cut-off 80	0.510	0.671	0.496	0.513	0.946	0.946	0.974
Cut-off 120	0.696	0.811	0.594	0.699	0.757	0.740	0.852

TABLE G.4. Testing Results for Israeli School Data: p-values, $\xi = 0.5$

	3	5	AL	IK	CCT	MSE-RBC	CER-RBC
g4math							
Cut-off 40	0.984	0.934	0.940	0.978	0.950	0.957	0.956
Cut-off 80	0.907	0.853	0.832	0.936	0.893	0.774	0.956
Cut-off 120	0.443	0.683	0.633	0.557	0.519	0.592	0.580
g4verb							
Cut-off 40	0.907	0.599	0.450	0.637	0.499	0.503	0.422
Cut-off 80	0.907	0.880	0.165	0.906	0.760	0.266	0.939
Cut-off 120	0.935	0.668	0.449	0.719	0.164	0.194	0.130
g5math							
Cut-off 40	0.854	0.678	0.461	0.788	0.829	0.917	0.832
Cut-off 80	0.499	0.419	0.913	0.466	0.749	0.785	0.691
Cut-off 120	0.931	0.812	0.591	0.873	0.818	0.763	.732
g5verb							
Cut-off 40	0.955	0.875	0.946	0.926	0.936	0.945	0.945
Cut-off 80	0.499	0.664	0.930	0.504	0.938	0.946	0.974
Cut-off 120	0.665	0.795	0.708	0.688	0.750	0.673	0.825

TABLE G.5. Testing Results for Colombia's SR Data: p-values ($\xi = 0.00999$, full table)

Outcome variables	MPV Bandwidths 2	3	4	Other Bandwidth Choices AI	IK	CCT	MSE RBC	CER RBC
Risk protection, consumption smoothing and portfolio choice								
Individual inpatient medical spending	0.53	0.79	0.85	0.56	0.95	0.96	0.63	0.86
Individual outpatient medical spending	0.95	0.87	0.85	0.01	0.61	0.89	0.92	0.87
Variability of individual inpatient medical spending	0.50	0.79	0.87	0.66	0.94	0.96	0.67	0.86
Variability of individual outpatient medical spending	0.91	0.95	0.99	0.83	0.68	0.97	0.98	0.94
Individual education spending	0.15	0.19	0.17	0.02	0.90	0.09	0.21	0.14
Household education spending	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total spending on food	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total monthly expenditure	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Has car	0.97	0.76	0.87	0.99	0.72	0.99	0.87	0.94
Has radio	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medical care use								
Preventive physician visit	0.62	0.99	1.00	1.00	0.37	0.98	0.50	0.75
Number of growth development checks last year	0.73	0.93	0.96	0.92	0.65	0.99	0.85	0.96
Curative care use	0.98	0.96	0.96	0.98	0.97	0.95	0.99	0.94
Primary care	0.92	0.92	0.94	0.99	0.96	0.95	0.96	0.88
Medical visit-specialist	0.98	0.93	0.73	0.93	0.89	0.63	0.83	0.81
Hospitalization	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00
Medical visit for chronic disease	0.15	0.49	0.72	0.53	0.09	0.64	0.64	0.73
Curative care use among children	0.95	0.98	0.98	1.00	0.95	0.99	0.99	0.99
Health status								
Child days lost to illness	0.60	0.67	0.80	0.76	0.66	0.86	0.76	0.83
Cough, fever, diarrhea	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00
Any health problem	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.99
Birthweight (KG)	0.90	100	1.00	0.99	0.90	1.00	1.00	1.00
Behavioral distortions								
Drank alcohol during pregnancy	0.42	0.74	0.87	0.93	0.19	0.93	0.91	0.94
Number of drinks per week during pregnancy	0.78	0.88	0.91	0.85	0.75	0.84	0.85	0.87
Months child breastfed	0.94	0.95	0.92	0.86	0.94	0.87	0.95	0.90
Folic acid during pregnancy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Number months folic acid during pregnancy	0.92	0.95	0.92	0.96	0.94	0.75	0.85	0.80
Contributory regime enrollment (ECV)	0.55	0.54	0.37	0.00	0.74	0.32	0.49	0.61
Contributory regime enrollment (DHS)	0.97	0.99	0.99	1.00	0.63	1.00	1.00	1.00
Other insurance (ECV)	0.89	0.93	0.92	0.87	0.81	0.90	0.94	0.92
Other insurance (DHS)	0.91	0.96	0.97	0.95	0.88	0.96	0.96	0.97
Uninsured (ECV)	0.67	0.68	0.45	0.07	0.75	0.68	0.69	0.68
Uninsured (DHS)	0.99	1.00	1.00	1.00	0.79	0.97	0.83	0.91

TABLE G.6. Testing Results for Colombia's SR Data: p-values ($\xi = 0.0316$, full table)

Outcome variables	MPV Bandwidths 2	3	4	Other Bandwidth Choices AI	IK	CCT	MSE RBC	CER RBC
Risk protection, consumption smoothing and portfolio choice								
Individual inpatient medical spending	0.53	0.78	0.85	0.56	0.94	0.95	0.60	0.87
Individual outpatient medical spending	0.93	0.85	0.82	0.01	0.56	0.84	0.87	0.87
Variability of individual inpatient medical spending	0.49	0.77	0.86	0.64	0.93	0.95	0.64	0.85
Variability of individual outpatient medical spending	0.87	0.93	0.97	0.95	0.62	0.95	0.92	0.91
Individual education spending	0.15	0.18	0.17	0.02	0.89	0.09	0.18	0.14
Household education spending	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total spending on food	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total monthly expenditure	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Has car	0.97	0.76	0.86	0.99	0.72	0.99	0.88	0.95
Has radio	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medical care use								
Preventive physician visit	0.62	0.99	1.00	1.00	0.37	0.98	0.52	0.73
Number of growth development checks last year	0.72	0.93	0.96	0.92	0.64	0.99	0.88	0.96
Curative care use	0.98	0.96	0.96	0.98	0.97	0.95	0.97	0.95
Primary care	0.92	0.92	0.94	0.99	0.96	0.95	0.94	0.88
Medical visit-specialist	0.98	0.93	0.73	0.93	0.89	0.63	0.79	0.80
Hospitalization	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00
Medical visit for chronic disease	0.15	0.49	0.72	0.53	0.09	0.64	0.63	0.72
Curative care use among children	0.96	0.98	0.97	1.00	0.95	1.00	0.99	0.98
Health status								
Child days lost to illness	0.60	0.67	0.80	0.76	0.66	0.86	0.77	0.82
Cough, fever, diarrhea	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00
Any health problem	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.99
Birthweight (KG)	0.90	100	1.00	0.99	0.90	1.00	1.00	1.00
Behavioral distortions								
Drank alcohol during pregnancy	0.42	0.74	0.87	0.93	0.19	0.93	0.89	0.94
Number of drinks per week during pregnancy	0.78	0.88	0.90	0.85	0.72	0.84	0.81	0.86
Months child breastfed	0.94	0.95	0.92	0.86	0.94	0.87	0.95	0.90
Folic acid during pregnancy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Number months folic acid during pregnancy	0.92	0.95	0.92	0.96	0.94	0.75	0.87	0.79
Contributory regime enrollment (ECV)	0.55	0.54	0.37	0.00	0.74	0.32	0.50	0.65
Contributory regime enrollment (DHS)	0.97	0.99	0.99	1.00	0.63	1.00	1.00	1.00
Other insurance (ECV)	0.89	0.93	0.92	0.87	0.81	0.90	0.94	0.93
Other insurance (DHS)	0.91	0.96	0.97	0.95	0.88	0.96	0.95	0.98
Uninsured (ECV)	0.67	0.68	0.45	0.07	0.75	0.68	0.68	0.69
Uninsured (DHS)	0.99	1.00	1.00	1.00	0.79	0.97	0.83	0.91

TABLE G.7. Testing Results for Colombia's SR Data: p-values ($\xi = 0.1706$, full table)

Outcome variables	MPV Bandwidths 2	3	4	Other Bandwidth Choices AI	IK	CCT	MSE RBC	CER RBC
Risk protection, consumption smoothing and portfolio choice								
Individual inpatient medical spending	0.43	0.68	0.76	0.99	0.88	0.90	0.5	0.82
Individual outpatient medical spending	0.85	0.84	0.78	0.64	0.41	0.80	0.84	0.83
Variability of individual inpatient medical spending	0.37	0.64	0.79	0.97	0.83	0.91	0.61	0.85
Variability of individual outpatient medical spending	0.77	0.89	0.95	0.93	0.35	0.91	0.90	0.90
Individual education spending	0.14	0.16	0.16	0.08	0.87	0.26	0.17	0.14
Household education spending	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total spending on food	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total monthly expenditure	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Has car	0.97	0.76	0.86	0.99	0.72	0.99	0.87	0.95
Has radio	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medical care use								
Preventive physician visit	0.62	0.99	1.00	1.00	0.37	0.98	0.50	0.75
Number of growth development checks last year	0.77	0.89	0.95	0.99	0.82	0.98	0.90	.95
Curative care use	0.98	0.96	0.96	0.98	0.97	0.95	0.99	0.95
Primary care	0.92	0.92	0.94	0.99	0.96	0.95	0.95	0.88
Medical visit-specialist	0.98	0.93	0.73	0.93	0.89	0.63	0.79	0.78
Hospitalization	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00
Medical visit for chronic disease	0.15	0.49	0.72	0.53	0.09	0.64	0.65	0.73
Curative care use among children	0.95	0.98	0.98	1.00	0.95	0.99	0.99	0.99
Health status								
Child days lost to illness	0.60	0.67	0.80	0.76	0.66	0.86	0.73	0.83
Cough, fever, diarrhea	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00
Any health problem	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.99
Birthweight (KG)	0.90	100	1.00	0.99	0.90	1.00	1.00	1.00
Behavioral distortions								
Drank alcohol during pregnancy	0.42	0.74	0.87	0.93	0.19	0.93	0.88	0.92
Number of drinks per week during pregnancy	0.73	0.85	0.88	0.83	0.69	0.80	0.76	0.80
Months child breastfed	0.94	0.95	0.92	0.86	0.94	0.87	0.96	0.90
Folic acid during pregnancy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Number months folic acid during pregnancy	0.92	0.95	0.92	0.96	0.94	0.75	0.86	0.80
Contributory regime enrollment (ECV)	0.55	0.54	0.37	0.00	0.74	0.32	0.51	0.57
Contributory regime enrollment (DHS)	0.97	0.99	0.99	1.00	0.63	1.00	1.00	1.00
Other insurance (ECV)	0.89	0.93	0.92	0.87	0.81	0.90	0.93	0.93
Other insurance (DHS)	0.91	0.96	0.97	0.95	0.88	0.96	0.96	0.98
Uninsured (ECV)	0.67	0.68	0.45	0.07	0.75	0.68	0.69	0.68
Uninsured (DHS)	0.99	1.00	1.00	1.00	0.79	0.97	0.84	0.92

TABLE G.8. Testing Results for Colombia's SR Data: p-values ($\xi = 0.5$, full table)

Outcome variables	MPV Bandwidths 2	3	4	Other Bandwidth Choices AI	IK	CCT	MSE RBC	CER RBC
Risk protection, consumption smoothing and portfolio choice								
Individual inpatient medical spending	0.52	0.68	0.71	0.89	0.76	0.71	0.58	0.76
Individual outpatient medical spending	0.72	0.84	0.93	0.22	0.23	0.90	0.71	0.90
Variability of individual inpatient medical spending	0.37	0.61	0.69	0.89	0.63	0.76	0.60	0.76
Variability of individual outpatient medical spending	0.40	0.75	0.76	0.54	0.18	0.78	0.66	0.74
Individual education spending	0.10	0.11	0.11	0.29	0.78	0.19	0.13	0.10
Household education spending	0.00	0.00	0.00	0.03	0.02	0.00	0.00	0.00
Total spending on food	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total monthly expenditure	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Has car	0.97	0.76	0.86	0.99	0.72	0.99	0.86	0.95
Has radio	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medical care use								
Preventive physician visit	0.62	0.99	1.00	1.00	0.37	0.98	0.51	0.76
Number of growth development checks last year	0.81	0.89	0.97	0.98	0.88	0.99	0.90	0.96
Curative care use	0.98	0.96	0.96	0.98	0.97	0.95	0.98	0.94
Primary care	0.92	0.92	0.94	0.99	0.96	0.95	0.96	0.90
Medical visit-specialist	0.96	0.92	0.72	0.91	0.87	0.63	0.80	0.79
Hospitalization	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00
Medical visit for chronic disease	0.15	0.49	0.72	0.53	0.09	0.64	0.63	0.74
Curative care use among children	0.96	0.99	0.98	1.00	0.94	0.99	0.99	0.99
Health status								
Child days lost to illness	0.60	0.67	0.80	0.76	0.66	0.86	0.75	0.82
Cough, fever, diarrhea	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00
Any health problem	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.99
Birthweight (KG)	0.90	100	1.00	0.99	0.90	1.00	0.91	0.94
Behavioral distortions								
Drank alcohol during pregnancy	0.42	0.74	0.87	0.93	0.19	0.93	0.89	0.94
Number of drinks per week during pregnancy	0.66	0.80	0.85	0.73	0.65	0.75	0.72	0.78
Months child breastfed	0.94	0.95	0.91	0.86	0.94	0.87	0.95	0.92
Folic acid during pregnancy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Number months folic acid during pregnancy	0.92	0.95	0.92	0.96	0.94	0.75	0.88	0.80
Contributory regime enrollment (ECV)	0.55	0.54	0.37	0.00	0.74	0.32	0.50	0.63
Contributory regime enrollment (DHS)	0.97	0.99	0.99	1.00	0.63	1.00	1.00	1.00
Other insurance (ECV)	0.86	0.91	0.91	0.86	0.80	0.89	0.94	0.92
Other insurance (DHS)	0.90	0.96	0.97	0.95	0.87	0.96	0.96	0.97
Uninsured (ECV)	0.67	0.68	0.45	0.07	0.75	0.68	0.69	0.70
Uninsured (DHS)	0.99	1.00	1.00	1.00	0.79	0.97	0.83	0.90

TABLE G.9. Testing Results for Colombia's SR Data by Regions ($\xi = 0.00999$)

	MPV bandwidths			Other bandwidth choice		
	2	3	4	AI	IK	CCT
Atlantica						
Household education spending	0.001	0.001	0.001	0.000	0.000	0.001
Total spending on food	0.009	0.008	0.026	0.000	0.015	0.020
Total monthly expenditure	0.000	0.001	0.000	0.000	0.000	0.000
Oriental						
Household education spending	0.000	0.000	0.000	0.000	0.000	0.002
Total spending on food	0.000	0.001	0.000	0.000	0.001	0.002
Total monthly expenditure	n.a.*	n.a.	n.a.	n.a.	n.a.	n.a.
Central						
Household education spending	0.000	0.098	0.058	0.000	0.000	0.000
Total spending on food	0.000	0.002	0.001	0.001	0.000	0.021
Total monthly expenditure	0.000	0.007	0.008	0.000	0.000	0.001
Pacifica						
Household education spending	0.001	0.147	0.073	0.000	0.043	0.003
Total spending on food	0.150	0.237	0.236	0.013	0.107	0.385
Total monthly expenditure	0.091	0.347	0.231	0.002	0.071	0.125
Bogota						
Household education spending	0.000	0.000	0.000	0.000	0.014	0.000
Total spending on food	0.000	0.000	0.001	0.003	0.002	0.000
Total monthly expenditure	0.000	0.000	0.000	0.000	0.000	0.000
Territorios Nacionales						
Household education spending	0.085	0.247	0.063	0.000	0.037	0.090
Total spending on food	0.029	0.310	0.032	0.000	0.057	0.281
Total monthly expenditure	0.227	0.271	0.349	0.001	0.364	0.752

∗: not available due to small sample size.

TABLE G.10. Subsample Sizes by Regions

	Household Edu. Spending	Total Spending on Food	Total Monthly Exp.
Atlantica	3969	3969	1480
Oriental	1496	1496	452
Central	5341	5318	2728
Pacifica	6370	6370	3203
Bogota	43656	41108	14634
Territorios Nacionales	1137	1137	643

TABLE G.11. Sample Sizes and Bandwidths for the Israeli School Data

		3		5		AI		IK		CCT	
g4math											
Cut-off 40 ($n = 984$)	(n_-, n_+)	17	67	26	93	102	302	23	84	89	227
	(h_-, h_+)	3	3	5	5	11.1	15.0	3.8	3.9	10.6	10.4
Cut-off 80 ($n = 1376$)	(n_-, n_+)	29	45	76	71	292	142	29	45	206	107
	(h_-, h_+)	3	3	5	5	15.0	9.3	2.8	2.8	10.5	10.6
Cut-off 120 ($n = 976$)	(n_-, n_+)	27	20	66	34	189	66	47	34	117	60
	(h_-, h_+)	3	3	5	5	15.0	10.4	4.0	4.2	8.7	9.0
g4verb											
Cut-off 40 ($n = 984$)	(n_-, n_+)	17	67	26	93	57	302	23	84	89	227
	(h_-, h_+)	3	3	5	5	7.7	15.0	4.0	4.0	11.0	10.8
Cut-off 80 ($n = 1376$)	(n_-, n_+)	29	45	76	71	270	142	55	54	206	107
	(h_-, h_+)	3	3	5	5	13.7	9.7	3.2	3.2	10.2	10.4
Cut-off 120 ($n = 976$)	(n_-, n_+)	27	20	66	34	189	93	66	34	138	66
	(h_-, h_+)	3	3	5	5	15.0	13.3	4.3	4.4	10.3	10.7
g5math											
Cut-off 40 ($n = 983$)	(n_-, n_+)	19	77	38	112	143	328	29	94	47	130
	(h_-, h_+)	3	3	5	5	15.0	15.0	4.0	4.0	5.6	5.5
Cut-off 80 ($n = 1359$)	(n_-, n_+)	59	44	80	86	285	223	72	65	201	150
	(h_-, h_+)	3	3	5	5	15.0	15.0	3.9	4.0	10.4	10.6
Cut-off 120 ($n = 905$)	(n_-, n_+)	36	22	61	31	166	56	49	25	109	56
	(h_-, h_+)	3	3	5	5	15.0	8.1	3.7	3.9	8.1	8.4
g5verb											
Cut-off 40 ($n = 983$)	(n_-, n_+)	19	77	38	112	58	268	38	112	70	184
	(h_-, h_+)	3	3	5	5	6.4	11.5	4.2	4.1	7.2	7.0
Cut-off 80 ($n = 1359$)	(n_-, n_+)	59	44	80	86	285	223	72	65	201	154
	(h_-, h_+)	3	3	5	5	15.0	15.0	3.7	3.8	10.5	10.7
Cut-off 120 ($n = 905$)	(n_-, n_+)	36	22	61	31	166	45	49	25	79	45
	(h_-, h_+)	3	3	5	5	15.0	6.8	3.2	3.3	6.7	7.0

Note: h_- and h_+ denote the specified bandwidths to the left and right of the cut-off, respectively. The data driven bandwidths presented in this table (AI, IK, and CCT) are undersmoothed by multiplying $(\sum_{i=1}^{n} 1\{R_i < r_0\})^{1/5-1/4.5}$ and $(\sum_{i=1}^{n} 1\{R_i \geq r_0\})^{1/5-1/4.5}$, respectively. We set the upper-bound of the data driven bandwidths at 15. n_- and n_+ denote the number of observations with values of the running variable in $(r_0 - h_-, r_0)$ and $[r_0, r_0 + h_+)$, respectively.

TABLE G.12. Sample Sizes and Bandwidths for Colombia's SR Data

Outcomes		2		3		4		AI		IK		CCT	
HES	(n_-, n_+)	1701	2521	2474	3783	3034	5204	3484	18798	1082	1423	3586	6611
	(h_-, h_+)	2	2	3	3	4	4	54.99	11.8	1.11	1.05	5.23	4.96
TSF	(n_-, n_+)	1664	2432	2420	3655	2979	5034	3410	18247	1050	1384	3512	6385
	(h_-, h_+)	2	2	3	3	4	4	3.78	23.5	1.36	1.29	3.70	3.51
TME	(n_-, n_+)	402	564	567	828	643	1136	732	4867	285	314	754	1398
	(h_-, h_+)	2	2	3	3	4	4	6.08	8.32	0.99	0.92	2.12	1.98

HES: Household Education Spending; TSF: Total Spending on Food; TME: Total Monthly Expenditure.
Note: h_- and h_+ denote the specified bandwidths to the left and right of the cut-off, respectively. The data driven bandwidths presented in this table (AI, IK, and CCT) are undersmoothed by multiplying $(\sum_{i=1}^{n} 1\{R_i \leq r_0\})^{1/5-1/4.5}$ and $(\sum_{i=1}^{n} 1\{R_i > r_0\})^{1/5-1/4.5}$, respectively. We set the upper-bound of the data driven bandwidths at 15. n_- and n_+ denote the number of observations with values of the running variable in $(r_0 - h_-, r_0)$ and $[r_0, r_0 + h_+)$, respectively.

FIGURE G.1. Estimated compliers' outcome density: Household education spending

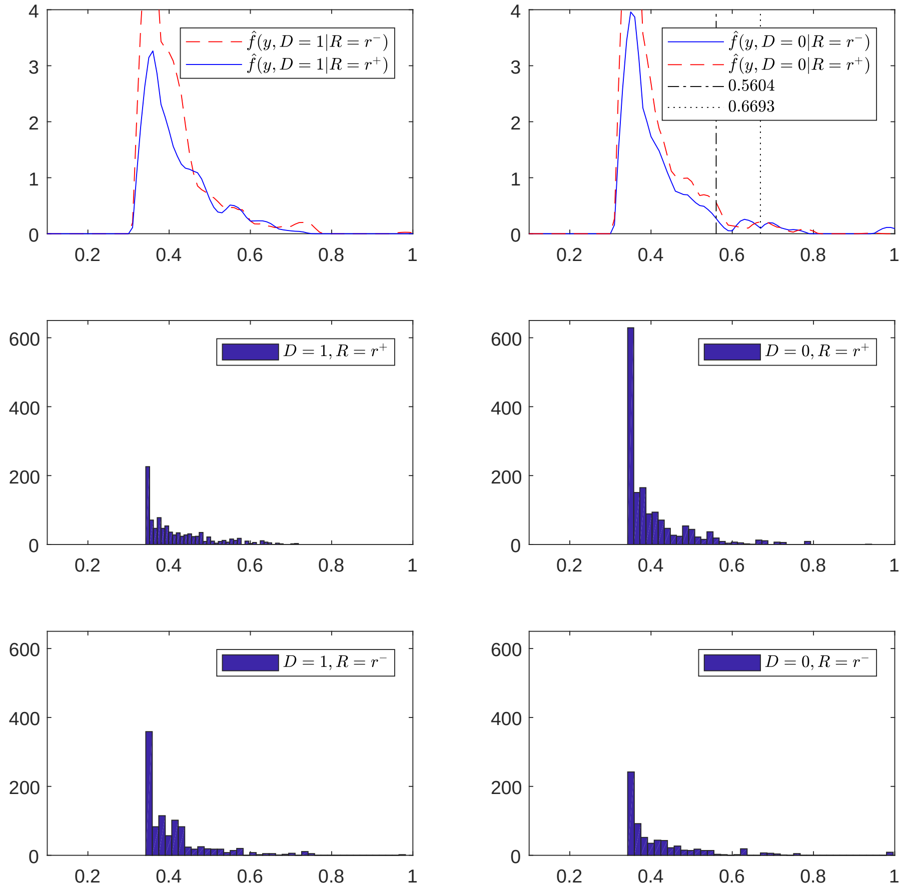

TABLE G.13. Observations in the maximizing interval ($h^+ = h^- = 2$): Household edu. spending

Household education spending	# of observations		
Subsample of	All	$\cap\{0.5604 \leq Y \leq 0.6693\}$	Ratio
$\{0 \leq R < h^+\} \cap \{D=0\} \Leftrightarrow \mathbf{N} \cup \mathbf{C}$	1563	43	2.75%
$\{h^- < R < 0\} \cap \{D=0\} \Leftrightarrow \mathbf{N}$	690	25	3.62%

FIGURE G.2. Estimated compliers' outcome density: Total monthly spending

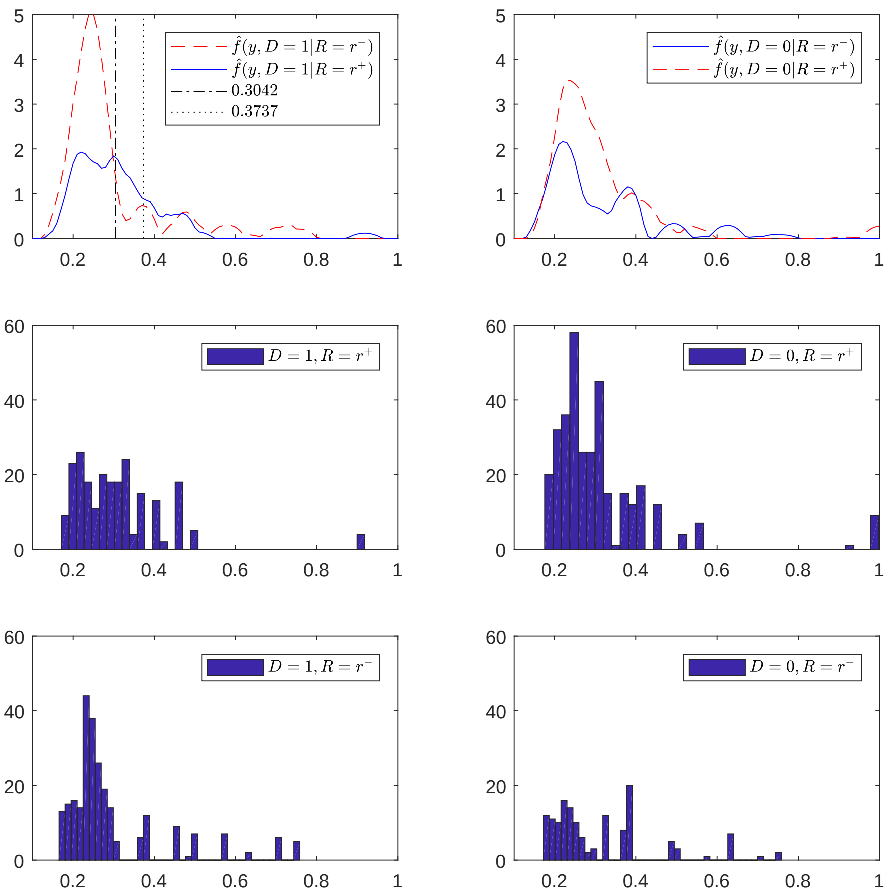

TABLE G.14. Observations in the maximizing interval ($h^+ = h^- = 2$):Total monthly expenditure

Total monthly expenditure	# of observations		
Subsample of	All	$\cap\{0.3042 \leq Y \leq 0.3737\}$	Ratio
$\{0 \leq R < h^+\} \cap \{D=1\} \Leftrightarrow \mathbf{A}$	228	61	26.7%
$\{h^- < R < 0\} \cap \{D=1\} \Leftrightarrow \mathbf{A} \cup \mathbf{C}$	259	6	2.32%

References

Andrews, D. W. K., and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81(2), 609–666.

Andrews, D. W. K., and X. Shi (2014): "Nonparametric inference based on conditional moment inequalities," *Journal of Econometrics*, 179(1), 31–45.

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

Barrett, G. F., and S. G. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

Bertanha, M., and G. W. Imbens (2020): "External Validity in Fuzzy Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 38, 593–612.

Bugni, F. A., and I. A. Canay (2018): "Testing Continuity of a Density via g order statistics in the Regression Discontinuity Design," *Working Paper*.

Calonico, S., M. D. Cattaneo, M. H. Farrel, and R. Titiunik (2019): "Regression discontinuity designs using covariates," *Review of Economics and Statistics*, 101, 442–451.

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

Canay, I. A., and V. Kamat (2018): "Approximate permutation tests and induced order statistics in the regression discontinuity design," *Review of Economic Studies*, 85, 1577–1608.

Cattaneo, M. D., M. Jansson, and X. Ma (2020): "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 115(531), 1449–1455.

Chiang, H. D., Y.-C. Hsu, and Y. Sasaki (2017): "Robust Uniform Inference for Quantile Treatment Effects in Regression Discontinuity Designs," *arXiv*.

de Chaisemartin, C. (2017): "Tolerating defiance? Local average treatment effects without monotonicity," *Quantitative Economics*, 8(2), 367–396.

Donald, S. G., and Y.-C. Hsu (2016): "Improving the power of tests of stochastic dominance," *Econometric Reviews*, 35(4), 553–585.

Dong, Y. (2018): "Alternative assumptions to identify LATE in fuzzy regression discontinuity designs," *Oxford Bulletin of Economics and Statistics*, 80(5), 1020–1027.

DONG, Y., AND A. LEWBEL (2015): “Identifying the effect of changing the policy threshold in regression discontinuity models,” *Review of Economics and Statistics*, 97(5), 1081–1092.

DUDLEY, R. (1999): *Uniform Central Limit Theorems*. Cambridge University Press.

FAN, J., AND I. GIJBELS (1992): “Variable bandwidth and local linear regression smoothers,” *The Annals of Statistics*, 20(4), 2008–2036.

FRÖLICH, M., AND M. HUBER (2019): “Regression discontinuity design with covariates,” *Journal of Business & Economic Statistics*, 37(4), 736–748.

HSU, Y.-C. (2016): “Multiplier bootstrap for empirical processes,” *Working Paper*.

——— (2017): “Consistent tests for conditional treatment effects,” *The Econometrics Journal*, 20(1), 1–22.

HSU, Y.-C., AND S. SHEN (2019): “Testing treatment effect heterogeneity in regression discontinuity designs,” *Journal of Econometrics*, 208(2), 468–486.

IMBENS, G. W., AND K. KALYANARAMAN (2012): “Optimal bandwidth choice for the regression discontinuity estimator,” *The Review of Economic Studies*, 79(3), 933–959.

KITAGAWA, T., AND A. TETENOV (2018): “Supplement to “Who should be treated? Empirical welfare maximization methods for treatment choice",” *Econometrica Supplementary Material*, 86(2).

LEE, D. S. (2008): “Randomized experiments from non-random selection in US House elections,” *Journal of Econometrics*, 142(2), 675–697.

LEE, S., K. SONG, AND Y.-J. WHANG (2015): “Uniform asymptotics for nonparametric quantile regression with an application to testing monotonicity,” *arXiv*.

MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142(2), 698–714.

OTSU, T., K.-L. XU, AND Y. MATSUSHITA (2013): “Estimation and inference of discontinuity in density,” *Journal of Business & Economic Statistics*, 31(4), 507–524.

POLLARD, D. (1990): “Empirical processes: theory and applications,” in *NSF-CBMS regional conference series in probability and statistics*, pp. i–86. JSTOR.

TSIREL’SON, V. S. (1975): “The Density of the Distribution of the Maximum of a Gaussian Process,” *Theory of Probability and its Application*, 16, 847–856.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.