cemmap

centre for microdata methods and practice

Testing identifying assumptions in fuzzy regression discontinuity designs

Yoichi Arai
Yuchin Hsu
Toru Kitagawa
Ismael Mourifié
Yuanyuan Wan

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP10/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

TESTING IDENTIFYING ASSUMPTIONS IN FUZZY REGRESSION DISCONTINUITY DESIGNS*

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

ABSTRACT. We propose a new specification test for assessing the validity of fuzzy regression discontinuity designs (FRD-validity). We derive a new set of testable implications, characterized by a set of inequality restrictions on the joint distribution of observed outcomes and treatment status at the cut-off. We show that this new characterization exploits all the information in the data useful for detecting violations of FRD-validity. Our approach differs from, and complements existing approaches that test continuity of the distributions of running variables and baseline covariates at the cut-off since ours focuses on the distribution of the observed outcome and treatment status. We show that the proposed test has appealing statistical properties. It controls size in large sample uniformly over a large class of distributions, is consistent against all fixed alternatives, and has non-trivial power against some local alternatives. We apply our test to evaluate the validity of two FRD designs. The test does not reject the FRD-validity in the class size design studied by Angrist and Lavy (1999) and rejects in the insurance subsidy design for poor households in Colombia studied by Miller, Pinto, and Vera-Hernández (2013) for some outcome variables, while existing density tests suggest the opposite in each of the cases.
Keywords: Fuzzy regression discontinuity design, nonparametric test, inequality restriction, multiplier bootstrap.

Date: Thursday 14th March, 2019.

∗. This paper merges and replaces the unpublished independent works of Arai and Kitagawa (2016) and Hsu, Mourifié, and Wan (2016).

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: We would like to thank Josh Angrist, Matias Cattaneo, Hide Ichimura, Robert McMillan, Phil Oreopoulos, Christoph Rothe, Yuya Sasaki, and Marcos Vera-Hernandez for beneficial comments. We also thank the participants of the 2016 Asian, the 2017 North American and the European Meetings of the Econometric Society, the Cemmap Workshop on Regression Discontinuity and on Advances in Econometrics, the Workshop on Advances in Econometrics at Dogo, Tsinghua Workshop on Econometrics, and the department seminars at Duke University, PSU, UBC, and the Queen's University. 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.

1. INTRODUCTION

In recent years, the regression discontinuity (RD) design first introduced by Thistlethwaite and Campbell (1960) has become one of the most widely used quasi-experimental methods in program evaluation studies. In the RD design, the probability of being treated changes discontinuously at a known cut-off of an underlying assignment variable, i.e. the running variable. The RD design is called sharp if the probability jumps from zero to one, and is called fuzzy otherwise. This cut-off point is usually set by an administrative or legislative rule. For example, Angrist and Lavy (1999) use Maimonides's rule in Israel that forces a maximum class size of 40 to estimate the causal effect of class size on student performance, and Lee (2008) investigates the effect of the incumbency advantages on the next election in the United States House of Representatives.[1]

The RD design identifies the causal impact of the treatment at the cut-off by comparing the outcomes of treated and non-treated individuals who lie close to the cut-off. The validity of the RD design relies crucially on the assumption that the individuals right below and above the cut-off have similar distributions of the unobservables. Hahn, Todd, and Van der Klaauw (2001, HTV hereafter) first formalize this identification argument of causal effects at the cut-off using the framework of potential outcomes, and subsequently, Frandsen, Frölich, and Melly (2012, FFM hereafter), Dong and Lewbel (2015) and Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) consider a refined set of identifying conditions. In the fuzzy regression discontinuity (FRD) setting, the two key conditions for identification, which we refer to as *FRD-validity* in this paper, are (i) *local continuity*, the continuity of the distributions of the potential outcomes and treatment selection heterogeneity at the cut-off, and (ii) *local monotonicity*, the monotonicity of the treatment selection response to the running variable near the cutoff.

Credibility of FRD-validity is controversial in many empirical contexts. For instance, agents (or administrative staff) may manipulate the value of their running variable to be eligible for their preferred treatment. If their manipulation behavior depends on their underlying potential outcomes, this can lead to a violation of the local continuity condition. As another example, when multiple programs share the index of treatment assignment and its threshold (e.g. the poverty line), but an individual's treatment status is observed only for the treatment of interest, the potential outcome

[1] Imbens and Lemieux (2008), Lee and Lemieux (2010) for surveys, and Cattaneo and Escanciano (2017) for recent advances of the literature.

distributions indexed by the treatment of interest can be discontinuous at the cut-off (see Miller, Pinto, and Vera-Hernández (2013) and Carneiro and Ginja (2014) for examples and discussions of the issue). Motivated by a clearer economic interpretation and the availability of testable implications, Lee (2008) imposes a *stronger* set of identifying assumptions that implies continuity of the distributions of running variable and covariates at the cut-off. Following his approach, researchers routinely assess the continuity condition by applying the tests of McCrary (2008), Otsu, Xu, and Matsushita (2013), Cattaneo, Jansson, and Ma (2016), and Canay and Kamat (2018). When the running variable is manipulated, Gerard, Rokkanen, and Rothe (2018) provide a partial identification approach in the presence of "one-sided manipulation". As noted by McCrary (2008), however, in the absence of Lee's additional identifying assumption, the continuity of the distributions of running variable and baseline covariates at the cut-off is neither necessary nor sufficient for FRD-validity. In other words, rejection or acceptance of the existing tests is non-informative about FRD-validity or violation thereof.

This paper proposes a novel test for FRD-validity. We first derive a new set of testable implications, characterized by a set of inequality restrictions on the joint distribution of observed outcomes and treatment status at the cut-off. These testable implications are necessary conditions for FRD-validity, but we show that they cannot be strengthened without additional assumptions, i.e., they exploit all the information in the data useful for detecting the violation of FRD-validity. We propose a nonparametric test for these testable implications. The test controls size uniformly over a large class of distributions of observables, is consistent against all fixed alternatives violating the testable implications, and has non-trivial power against some local alternatives. Implementability and asymptotic validity of our test neither restricts the support of Y nor presumes the continuity of running variable's density at the cut-off. Monte Carlo simulations show that the test performs well in finite samples.

Our test differs from and complements existing continuity tests. As we will elaborate in more details in Section 2.1 and Section 5, there are important empirical contexts where results of existing tests are not informative about FRD-validity while our test is. For instance, if multiple programs share the same running variable and the same threshold, the FRD design ignoring the other programs can lead to violation of continuity of the potential outcome distributions (for the program of interest), which our test can detect even though the density of running variable is continuous. In contrast, when evidence of manipulation of the running variable exists but they are independent of the potential outcomes (possibly conditional on covariates), then the potential outcome distributions (controlling the covariates if any) are continuous. Then, our test does not reject FRD-validity even though the running variable density is discontinuous. A novelty of our test is to exploit the aspects of the data that are informative for assessing FRD-validity but have been neglected in the existing density continuity tests. We therefore recommend to implement our test, together with the existing tests for continuity of the running variable density regardless of the results thereof.

Since our test makes use of the observations of not only the running variable but also treatment status and observed outcome, our test has a unique feature of being outcome-specific. When multiple outcomes are studied within the same FRD design, the researchers can assess credibility of FRD-validity separately for each outcome variable. For instance, suppose that the running variable has a discontinuous density and the manipulation is correlated with outcome variable A but independent of outcome variable B, then instead of abandoning the data set, researchers can still use the FRD design to identify the causal effect on outcome variable B. In such scenario, the outcome-specific feature of our test is useful for pre-screening the outcome variables of which FRD-validity is not contaminated by the running variable manipulation.

To illustrate that our test can provide new insights in empirical applications, we apply it to the designs studied in Angrist and Lavy (1999) and Miller, Pinto, and Vera-Hernández (2013). Angrist and Lavy (1999) use the discontinuity of class size with respect to enrollment due to Maimonides' rule to identify the causal effect of class size on student performance. We find that the FRD validity in this example is not rejected by our test for all 4 outcome variables (Grade 4 Math and Verb, Grade 5 Math and Verb), even though an existing continuity test suggests evidence for discontinuity of the running variable's density at the cut-off (see Otsu, Xu, and Matsushita, 2013). Miller, Pinto, and Vera-Hernández (2013) evaluate the impact of the "Colombia's Régimen Subsidiado"–a publicly financed insurance program–on 33 outcomes, where program eligibility is determined by a poverty index. Although the continuity test supports continuity of the running variable density at the cutoff, our test rejects FRD validity for 3 outcome variables (Household Education Spending, Total Spending on Food, and Total Monthly Spending). This result suggests further investigation would be beneficial for identifying and estimating the causal effect on these outcomes.

The rest of the paper is organized as follows. In Section 2, we lay out the main identifying assumptions that our test aims to assess and derive their testable implications. Section 3 provides test statistics and shows how to obtain their critical values. Monte Carlo experiments shown in Section 4 examine the finite sample performance of our tests. Section 5 presents the empirical applications. We discuss possible extensions in Section 6, and conclude in Section 7. All proofs are collected in the Appendix.

2. IDENTIFYING ASSUMPTIONS AND SHARP TESTABLE IMPLICATIONS

We adopt the potential outcome framework introduced in Rubin (1974). Let $(\Omega, \mathcal{F}, P)$ be a probability space, where we interpret Ω as the population of interest and $\omega \in \Omega$ as a generic individual in the population.

Let R be an observed continuous random variable with support $\mathcal{R} \subset \mathbb{R}$.[2] We call R the *running variable*. Let $D(\cdot,\cdot) : \mathcal{R} \times \Omega \to \{0,1\}$ and $D(r,\omega)$ be the *potential treatment* that individual ω would have received, had her running variable been set to r. For $d \in \{0,1\}$, we define mappings $Y_d(\cdot,\cdot) : \mathcal{R} \times \Omega \to \mathcal{Y} \subset \mathbb{R}$ and let $Y_d(r,\omega)$ denote the *potential outcome* of individual ω had her treatment and running variable been set to d and r, respectively.

We view $(Y_1(r,\cdot), Y_0(r,\cdot), D(r,\cdot))_{r\in\mathcal{R}}$ as random elements indexed by r and write them as $(Y_1(r), Y_0(r), D(r))$ when it causes no confusion. By definition, $D(R) \in \{0,1\}$ is the observed treatment and we abbreviate it as D. Likewise, we denote the observed outcome by $Y = Y_1(R)D(R) + Y_0(R)(1 - D(R))$ throughout the paper. We use P to denote the joint distribution of $\big((Y_1(r), Y_0(r), D(r))_{r\in\mathcal{R}}, R\big)$, which induces the joint distribution of observables (Y, D, R).[3] We assume throughout that the conditional distribution of (Y, D) given $R = r$ is well-defined for all r in some neighborhood of r_0. Note that by letting the potential outcomes indexed also by r, we allow the running variable to have direct causal effect on outcomes. This could be relevant in some empirical applications as discussed in Dong and Lewbel (2015), and Dong (2018).

The main feature of the RD design is that the probability of being treated changes discontinuously at a known threshold $r_0 \in \mathcal{R}$. In FRD designs considered here, $D(r,\omega)$ can take different values for different ω but the proportion for which $D(r,\omega) = 1$ varies discontinuously at r_0. We also

[2] In this paper we consider continuous running variable. Kolesár and Rothe (2018) studied inference on ATE in sharp regression discontinuity designs with discrete running variable.

[3] For the purpose of exposition, we do not introduce other observable covariates X here. Section 6.1 incorporates X into the analysis.

follow the literature of FRD and assume that the two binary variables $D^+ \equiv \lim_{r \downarrow r_0} D(r)$ and $D^- \equiv \lim_{r \uparrow r_0} D(r)$ are well defined for all ω.

Analogous to the local average treatment effect (LATE) framework (Imbens and Angrist (1994)), we define the compliance status $T(r, \omega)$ of individual ω in a small neighborhood of the cut-off r_0 based on how the potential treatment varies with r. Similar to FFM, Bertanha and Imbens (2014), and Dong and Lewbel (2015), for $\epsilon > 0$, we classify the population members into one of the following five categories:

$$T_\epsilon(\omega) = \begin{cases} \mathbf{A}, & \text{if } D(r,\omega) = 1, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{C}, & \text{if } D(r,\omega) = 1\{r \geq r_0\}, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{N}, & \text{if } D(r,\omega) = 0, \text{ for all } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{DF}, & \text{if } D(r,\omega) = 1\{r < r_0\}, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{I}, & \text{otherwise} \end{cases}, \tag{1}$$

where **A, C, N, DF** and **I** represent "always takers", "compliers", "never takers", "defiers" and "indefinite", respectively. The above definition coincides with the definition of types in FFM as $\epsilon \to 0$. As pointed out by Dong and Lewbel (2015), for a given ϵ and a given individual ω, this definition implicitly assumes the group to which ω belongs does not vary with r. This way of defining the treatment selection heterogeneity does not restrict the shape of $P(D = 1|R = r)$ over $(r_0 - \epsilon, r_0 + \epsilon)$.

We are ready to present the main identifying assumptions of which we aim to test their implication. In the statements of the assumptions we assume that all the limiting objects exist.

Assumption 1 (Local monotonicity)**.** *For $t \in \{\mathbf{DF}, \mathbf{I}\}$, $\lim_{\epsilon \to 0} P(T_\epsilon = t|R = r_0 + \epsilon) = 0$ and $\lim_{\epsilon \to 0} P(T_\epsilon = t|R = r_0 - \epsilon) = 0$.*

Assumption 2 (Local continuity)**.** *For $d = 0, 1$, $t \in \{\mathbf{A}, \mathbf{C}, \mathbf{N}\}$, and an measurable subset $B \subseteq \mathcal{Y}$, we have*

$$\lim_{\epsilon \to 0} P(Y_d(r_0 + \epsilon) \in B, T_\epsilon = t|R = r_0 + \epsilon) = \lim_{\epsilon \to 0} P(Y_d(r_0 - \epsilon) \in B, T_\epsilon = t|R = r_0 - \epsilon).$$

Assumptions 1 and 2 play similar roles of the instrument exogeneity (exclusion and random assignment) and the instrument monotonicity assumptions in the LATE framework. Assumption 1 says that as the neighborhood of r_0 shrinks, the conditional proportion of defiers and indefinites converges to zero, implying that only "always takers", "compliers", and "never takers" may exist at the limit. The local continuity assumption says that the conditional joint distributions of potential outcomes and compliance types are continuous at the cut-off. Our local continuity condition concerns the distributional continuity rather than only the conditional mean (as in the HTV's spirit).

The main feature of the FRD designs is that the probability of receiving treatment is discontinuous at the cut-off.

Assumption 3 (Discontinuity)**.** $\pi^+ \equiv \lim_{r \downarrow r_0} P(D=1|R=r) \neq \pi^- \equiv \lim_{r \uparrow r_0} P(D=1|R=r)$.

Assumptions 1 to 3 together guarantee the existence of compliers at the cut-off so that the parameters of interest in the FRD design are well-defined. Under Assumptions 1 to 3, the complier's potential outcome distributions at the cut-off, defined by

$$\begin{aligned} F_{Y_1(r_0)|\mathbf{C},R=r_0}(y) &\equiv \lim_{r \to r_0} P\left(Y_1(r) \leq y | T_{|r-r_0|} = \mathbf{C}, R=r\right), \\ F_{Y_0(r_0)|\mathbf{C},R=r_0}(y) &\equiv \lim_{r \to r_0} P\left(Y_0(r) \leq y | T_{|r-r_0|} = \mathbf{C}, R=r\right), \end{aligned}$$

are identified by the following quantities:[4] for all $y \in \mathcal{Y}$,

$$\begin{aligned} F_{Y_1(r_0)|\mathbf{C},R=r_0}(y) &= \frac{\lim_{r \downarrow r_0} \mathbb{E}_P[1\{Y \leq y\}D|R=r] - \lim_{r \uparrow r_0} \mathbb{E}_P[1\{Y \leq y\}D|R=r]}{\pi^+ - \pi^-}, \\ F_{Y_0(r_0)|\mathbf{C},R=r_0}(y) &= \frac{\lim_{r \uparrow r_0} \mathbb{E}_P[1\{Y \leq y\}(1-D)|R=r] - \lim_{r \downarrow r_0} \mathbb{E}_P[1\{Y \leq y\}(1-D)R=r]}{\pi^+ - \pi^-}. \end{aligned}$$

This is analogous to the distributional identification result by Imbens and Rubin (1997) in the LATE model. The identification of complier's potential outcome distributions implies the identification of a wide class of causal parameters including the complier's average effect and local quantile treatment effects.[5] Our identification result modifies FFM's Lemma 1 to accommodate the fact that we do not exclude r from the potential outcomes.

Note that Assumption 3 can be tested using the inference methods proposed by Calonico, Cattaneo, and Titiunik (2014) and Canay and Kamat (2018). Our test, hence, focuses on testing Assumptions 1 and 2.

[4]For completeness, we show this identification result in Proposition 1 in Appendix C.2.

[5]Assumptions 1 and 2 play similar roles as FFM's Assumptions I3 and I2, respectively. The main difference from FFM's assumptions is that FFM define the compliance status just at the limit, and assume that the conditional distributions of potential outcomes given the limiting complying status and running variable are continuous at the cut-off.

The next theorem shows that *local monotonicity* and *local continuity* together imply a set of inequality restrictions on the distribution of data.

Theorem 1. *(i) Under Assumptions 1 and 2, the following inequalities hold:*

$$\lim_{r\uparrow r_0} \mathbb{E}_P[1\{y \leq Y \leq y'\}D|R=r] - \lim_{r\downarrow r_0} \mathbb{E}_P[1\{y \leq Y \leq y'\}D|R=r] \leq 0 \tag{2}$$

$$\lim_{r\downarrow r_0} \mathbb{E}_P[1\{y \leq Y \leq y'\}(1-D)|R=r] - \lim_{r\uparrow r_0} \mathbb{E}_P[1\{y \leq Y \leq y'\}(1-D)|R=r] \leq 0 \tag{3}$$

for all $y, y' \in \mathbb{R}$.

(ii) For a given distribution of observables (Y, D, R), assume that the conditional distribution of Y given (D, R) has the probability density function with respect to a dominating measure μ on $\mathcal{Y}$, has integrable envelope with respect to μ, and whose left-limit and right-limit with respect to the conditioning variable R are well defined at $R = r_0$, $\mu-$a.s. If inequalities (2) and (3) hold, there exists a joint distribution of $(\tilde{D}(r), \tilde{Y}_1(r), \tilde{Y}_0(r) : r \in \mathcal{R})$ such that Assumptions 1 and 2 hold, and the conditional distribution of $\tilde{Y} = \tilde{Y}_1(R)\tilde{D}(R) + \tilde{Y}_0(R)(1 - \tilde{D}(R))$ and $\tilde{D} = \tilde{D}(R)$ given $R = r$ induces the conditional distribution of (Y, D) given $R = r$ for all $r \in \mathcal{R}$.

Theorem 1 (i) shows a necessary condition that the distribution of observable variables has to satisfy under the FRD-validity conditions. In other words, a violation of inequalities (2) and (3) informs that at least one of the FRD-validity conditions is violated. Theorem 1 (ii) clarifies that inequalities (2) and (3) are the most informative way to detect all the observable violations of the FRD-validity assumptions and the testable implications cannot be strengthened without further assumptions. We, however, emphasize that FRD-validity is a refutable but not a *confirmable* assumption, i.e., finding inequalities (2) and (3) hold in data does not guarantee FRD-validity.

Remark 1. FRD-validity defined by Assumptions 1 and 2 does not constrain the marginal density of R to be continuous at the cut-off. This contrasts with the testable implication of continuity of the running variable density shown in Lee (2008) for the sharp RD setting and commonly assessed in empirical practice. Lee's testable implication comes with a restriction that assumes smooth potential outcome equations with respect to continuously distributed unobservables. The testable implications of Theorem 1 (i) are valid no matter whether one assumes Lee's additional restrictions or not. In addition, the testable implications of Theorem 1 (i) concern the joint distribution of (Y, D) local to the cut-off, which the existing continuity tests for the running variable and observable covariates never make use of. In this sense, our test developed below, which does not require continuity of running variable density for its asymptotic validity, complements the existing continuity tests and we recommend to implement it in any FRD studies whatever the results the existing continuity tests yield.

Remark 2. In analogy to the testable implications in the LATE model considered in Balke and Pearl (1997), Imbens and Rubin (1997), Heckman and Vytlacil (2005), Kitagawa (2015), and Mourifié and Wan (2017), the testable implications of Theorem 1 (i) can be interpreted as an FRD version of the *non-negativity of the potential outcome density functions for the compliers at the cut-off.*

Despite such analogy, the framework and features specific to RD designs gives rise to some important differences and challenges. First, FRD-validity we test is the continuity of the conditional distributions of potential outcomes and complying status local to the cut-off. We allow the potential outcome $Y_d(r)$ to be indexed by the running variable (even in the neighborhood of the cutoff), and being different from the standard LATE framework, we do not assume "exclusion" type restriction. Second, formally defining the complying status in the FRD setting is more involved than in the standard LATE setting, implying that the local monotonicity condition we test is not an immediate extension of the LATE monotonicity. Finally, since the testable implications concern the distributional inequalities local to the cut-off, the construction of the test statistic requires proper smoothing with respect to the conditioning running variable. This feature contrasts with the test of Kitagawa (2015) in the LATE setting, and delivers analytical challenges concerning the asymptotic Gaussian process approximation of the empirical processes local to the cut-off.

Remark 3. As the magnitude of propensity score jump $\pi^+ - \pi^-$ gets smaller, we expect that the inequalities of (2) and (3) get closer to binding. For instance, in the extreme case of $\pi^+ - \pi^- = 0$, for a distribution satisfying the testable implication, inequalities (2) and (3) must hold with equalities, i.e., the conditional distribution of $(Y, D)|R$ is continuous at the cut-off. This means a distribution of potential variables violating FRD-validity is more likely to violate the testable implications, as the magnitude of the jump in the propensity score gets smaller. In the opposite direction, the testable implication of Theorem 1 loses screening power when the FRD design is close to a sharp design.

Remark 4. 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 $\epsilon > 0$

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

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, the inequalities 2 and 3 remain to be unimprovable testable implications under this weaker version of the local monotonicity assumption.

Bertanha and Imbens (2014) consider an alternative local monotonicity assumption that is more restrictive than Assumption 1.

Assumption 4 (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:*

$$T = \begin{cases} \mathbf{A}, & \text{if } D(r) = 1, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{C}, & \text{if } D(r) = 1\{r \geq r_0\}, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon), \\ \mathbf{N}, & \text{if } D(r) = 0, \text{ for } r \in (r_0 - \epsilon, r_0 + \epsilon). \end{cases} \tag{4}$$

This monotonicity implies that in some neighborhood of the cut-off, the compliance status is invariant for any given individual. Indeed, the testable implications of Theorem 1 (i) remain valid. Furthermore, it can be shown that strengthening Assumption 1 to Assumption 4 does not provide more testable implications in that inequalities in Theorem 1 (i) remain to be unimprovable, i.e., Theorem 1 (ii) holds even if Assumption 1 is replaced by Assumption 4.[6]

Remark 5. The literature has considered the *local independence* assumption, which is a stronger identifying assumption than local continuity of Assumption 2.

[6]Given a distribution of observables, we prove Theorem 1 (ii) in Appendix C by constructing a distribution of potential outcomes and selection types that satisfies the identifying assumptions. The distribution of potential outcomes and selection types constructed there in fact satisfies Assumption 4.

Assumption 5 (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.*

This assumption is slightly weaker than the HTV local independence assumption, since the latter involves the local exclusion restriction that rules out causal dependence of Y_d on R in the neighborhood. Dong (2018) gives interesting comparison between the local continuity and HTV local independence, and shows that the HTV local independence is restrictive to accommodate various empirical RD specifications/models.[7] The statement of Theorem 1 (i) indeed holds even if Assumption 2 is replaced by Assumption 5.

2.1. How our test differs and complements existing tests?

In this subsection, we will elaborate how our test complements the existing ones. In the presence of covariates $X \in \mathcal{X} \subset \mathbb{R}^{d_x}$ and assuming that all the probability densities in the following equations are well defined, we can write

$$f_{Y_d(r), T_{|r-r_0|} | R, X}(y, t | r, x) = \frac{f_{R | Y_d(r), T_{|r-r_0|}, X}(r | y, t, x)}{f_{X|R}(x|r) f_R(r)} f_{Y_d(r), T_{|r-r_0|}, X}(y, t, x), \tag{5}$$

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 introduced in the sharp RD framework and later discussed in Dong (2018) in the FRD setting. Since the SLC assumption is not directly testable, the existing literature has derived tests for two sets of its implication: (i) the continuity of $f_R(r)$, see for instance McCrary (2008), Otsu, Xu, and Matsushita (2013), Cattaneo, Jansson, and Ma (2016), 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 (5) that the *local continuity* (LC) assumption that we aim to test can be considered as an implication of the SLC assumption. However, it should be emphasized that LC is not nested within either the continuity of $f_R(r)$ or the

[7] For example, a linear regression model with interactive terms $Y = a + b(R - r_0) + \tau D + \tau_1 (R - r_0) D + \varepsilon$ implies treatment effect $Y_1 - Y_0 = \tau + \tau_1 (r - r_0)$ and HTV's local independence essentially assumes that $\tau_1 = 0$, which can be restrictive in many applications (see also discussions about the "treatment effect derivative" parameter in Dong and Lewbel, 2015, for more details).

continuity of $F_{X|R}(x|r)$ in r. In fact, there are important empirical scenarios in which the conclusions of the existing tests are not necessarily informative about the continuity of $f_{Y_d(r),T_{|r-r_0|}|R,X}(y,t|r,x)$, as we illustrate below.

Scenario 1: Existing tests do not reject while our test does.

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 to determine their eligibility. Let $D, Z \in \{0,1\}$ denote the treatment statuses, respectively. The researcher is concerned with the casual effect of the first program D. For simplicity, let us assume that the assignment of the second program is sharp, $Z = \mathbf{1}[R \geq r_0]$. In such context, the potential outcome model can be written as

$$Y = \underbrace{\{Y_{11}Z + Y_{10}(1-Z)\}}_{Y_1} D + \underbrace{\{Y_{01}Z + Y_{00}(1-Z)\}}_{Y_0}(1-D), \tag{6}$$

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/he 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 $d=1$)

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

and 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, we see that in this scenario existing tests would not reject the FRD-validity, while our test can. The result of our second empirical application discussed in Section 5.2 is an illustration of this scenario.

Scenario 2: Existing tests reject while our test does not and the FRD-validity holds.

This scenario could happen for a data generating process in which the discontinuity of either $f_R(r)$ or $f_{X|R}(x|r)$ (or both) is exactly compensated 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 with the potential outcomes. For instance, in the study of the Maimonides's rule in Israel, Angrist, Lavy, Leder-Luis, and Shany (2017) argued that the presence of discontinuity in the running variable (school enrollment) is due mainly 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 justify $f_{R|Y_1(r),T_{|r-r_0|},X}(r|y,t,x) = f_{R|X}(r|x)$ in some local neighborhood of the cut-off. This reduces Equation (5) 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 density is discontinuous, FRD-validity can still hold and LATE type parameters can still be identified if the manipulation is unrelated to the underlying potential outcomes.

Figure 2.1 illustrates how the testable implication of Theorem 1 relates to the identifying assumptions of SLC and LC, and the DGPs of the two scenarios discussed above. We assume that all the DGPs represented on this figure satisfy Assumptions 1 and 3. The "blue" ellipse consists of the DGPs that satisfy the continuity of running variable density and the "red" ellipse consists of the DGPs that satisfy Assumption 2. The "green" ellipse collects all the DGPs that satisfy the testable implication of Theorem 1.

Finally, in addition to the aforementioned two differences, the existing tests are only concerned about the continuity conditions while our test jointly assess the validity of local continuity and also the local monotonicity assumption. Hence, our test is a useful complement of the existing tests also in this regard too, and we recommend to implement it in any FRD studies, together with the existing tests of the continuity of the running variable densities and regardless of the results thereof.

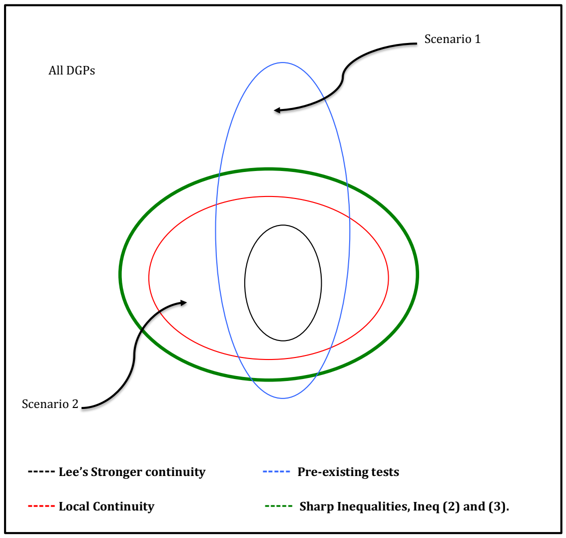


FIGURE 1. Testable Implications

3. Testing Procedure

This section proposes a testing procedure for the testable implications of Theorem 1 (i). We assume throughout that a sample consists of independent and identically distributed (i.i.d.) observations, $\{(Y_i, D_i, R_i)\}_{i=1}^{n}$. Noting that the inequality restrictions of Theorem 1 (i) can be seen as infinitely many unconditional moment inequalities local to the cut-off, we adopt and extend the inference procedure for conditional moment inequalities developed in Andrews and Shi (2013) by incorporating the local feature of the RD design. Implementability and asymptotic validity of our test neither restricts the support of Y nor presumes the continuity of running variable's density at the cut-off.[8] Consider a class of instrument functions $\mathcal{G}$ indexed by $\ell \in \mathcal{L}$:

$$\mathcal{G} = \{g_\ell(\cdot) = 1\{\cdot \in C_\ell\} : \ell \equiv (y, y') \in \mathcal{L}\}, \text{ where}$$
$$C_\ell = [y, y'] \cap \mathcal{Y},$$
$$\mathcal{L} = \{(y, y') : -\infty \le y \le y' \le \infty\}.$$

[8]See Appendix B for regularity conditions for asymptotic validity of our test.

$\mathcal{G}$ consists of indicator functions of closed and connected intervals on $\mathcal{Y}$. Expressing the inequalities (2) and (3) by

$$\begin{aligned}\nu_{P,1}(\ell) &\equiv \lim_{r\uparrow r_0} \mathbb{E}_P[g_\ell(Y)D|R=r] - \lim_{r\downarrow r_0} \mathbb{E}_P[g_\ell(Y)D|R=r] \le 0,\\ \nu_{P,0}(\ell) &\equiv \lim_{r\downarrow r_0} \mathbb{E}_P[g_\ell(Y)(1-D)|R=r] - \lim_{r\uparrow r_0} \mathbb{E}_P[g_\ell(Y)(1-D)|R=r] \le 0,\end{aligned} \tag{7}$$

for all $\ell \in \mathcal{L}$, we set up the null and alternative hypotheses as

$$\begin{aligned}&H_0: \ \nu_{P,1}(\ell) \le 0 \text{ and } \nu_{P,0}(\ell) \le 0 \text{ for all } \ell \in \mathcal{L},\\ &H_1: \ H_0 \text{ does not hold.}\end{aligned} \tag{8}$$

Since H_0 is equivalent to $S \equiv \sup_{d\in\{0,1\},\ell\in\mathcal{L}} \nu_{P,d}(\ell) \le 0$, we construct our test statistic as a standardized sample analog of S, which estimators of the unknown functions $\nu_{P,d}(\ell)$ weighted by their standard errors are plugged in.

We construct $\hat{\nu}_d(\ell)$ an estimator for $\nu_{P,d}(\ell)$ by a difference of the two local linear regressions estimated from below and above the cut-of. We do not vary the bandwidths over $\ell \in \mathcal{L}$, but we allow them to vary across the cut-offs; let $h_+ = c_+ h$ and $h_- = c_- h$ be the bandwidths above and below the cut-off, respectively. We assume that their convergence rates with respect to the sample size n are common as specified by h, e.g., $h = n^{-1/4.5}$. The difference of h_+ and h_- can be captured by possibly distinct constants c_+ and c_-.

Let $\sigma_{P,d}(\ell)$ be the asymptotic standard deviation of $\sqrt{nh}(\hat{\nu}_d(\ell) - \nu_{P,d}(\ell))$ and $\hat{\sigma}_d(\ell)$ be a uniformly consistent estimator for $\sigma_{P,d}(\ell)$. See Algorithm 1 below for its construction. To ensure uniform convergence of the variance weighted processes, we weigh $\hat{\nu}_d(\ell)$ by a trimmed version of the standard error estimators, $\hat{\sigma}_{d,\xi}(\ell) = \max\{\xi, \hat{\sigma}_d(\ell)\}$, where $\xi > 0$ is a trimming constant chosen by users. We then define a Kolmogorov-Smirnov (KS) type test statistic,

$$\widehat{S}_n = \sup_{d\in\{0,1\},\ \ell\in\mathcal{L}} \frac{\sqrt{nh}\cdot\hat{\nu}_d(\ell)}{\hat{\sigma}_{d,\xi}(\ell)}. \tag{9}$$

A large value of $\widehat{S}_n$ is statistical evidence against the null hypothesis. Since the cardinality of $\mathcal{L}$ is inifinite if Y is continuously distributed, this supremum over $\ell \in \mathcal{L}$ might appear infeasible to compute. With our construction of $\hat{\nu}_d(\ell)$ and $\hat{\sigma}_{d,\xi}(\ell)$ shown in Appendix A, however, $\hat{\nu}_d(\ell)/\hat{\sigma}_{d,\xi}(\ell)$ varies only with the set of observed Y_i's contained in C_ℓ. We can therefore coarsen $\mathcal{L}$ to the class of finite number of intervals spanned by the observed values of Y in the sample,

$$\hat{\mathcal{L}} \equiv \{[Y_i, Y_j] : Y_i \leq Y_j, i, j \in \{1, \ldots n\}\}, \tag{10}$$

without changing the value of the test statistic. Further coarsening might be desirable in large n situations, while it introduces an approximation error. In the Monte Carlo studies of Section 4 and the empirical applications of Section 5, we standardize (subtract sample average and divided by the sample standard error) and rescale the range of Y to unit interval (by applying transformation $\Phi(\cdot)$ with standard normal cdf),[9] and employ the following coarsening of the class of intervals.

$$\mathcal{L}_{coarse} = \left\{ (y, y+c) : \ c^{-1} = q, \text{ and } q \cdot y \in \{0, 1, 2, \cdots, (q-1)\} \text{ for } q = 1, 2, \cdots, Q \right\}. \tag{11}$$

As done in Hansen (1996) and Barrett and Donald (2003) in different contexts, we obtain asymptotically valid critical values by approximating the null distribution of the statistic using multiplier bootstrap. Algorithm 1 below summarizes the implementation of our test. Theorems 2-4 in Appendix B show that the proposed test controls the size at pre-specified significant levels uniformly, rejects fixed alternatives with probability approaching one, and has good power against a class of local alternatives.

Algorithm 1. (Implementation)

i. Specify a finite class of intervals $\mathcal{L}^*$. For instance, $\mathcal{L}^* = \hat{\mathcal{L}}$ of (10), or a coarsened version with the standardized outcome, $\mathcal{L}^* = L_{coarse}$ of (11) with a choice of finite integer Q (e.g., $Q = 15$).

ii. For each $\ell \in \mathcal{L}^*$, let $\hat{m}_{1,+}(\ell)$ and $\hat{m}_{1,-}(\ell)$ be local linear estimators for $\lim_{r \downarrow r_0} \mathbb{E}_P[g_\ell(Y)D|R = r]$ and $\lim_{r \uparrow r_0} \mathbb{E}_P[g_\ell(Y)D|R = r]$, respectively. Similarly, let $\hat{m}_{0,+}(\ell)$ and $\hat{m}_{0,-}(\ell)$ be local linear estimators for $\lim_{r \downarrow r_0} \mathbb{E}_P[g_\ell(Y)(1-D)|R = r]$ and $\lim_{r \uparrow r_0} \mathbb{E}_P[g_\ell(Y)(1-D)|R = r]$, respectively. See equation (20) in Appendix A for their closed-form expressions. Obtain $\hat{v}_1(\ell)$ and $\hat{v}_0(\ell)$ as follows:

$$\hat{v}_1(\ell) = \hat{m}_{1,-}(\ell) - \hat{m}_{1,+}(\ell), \quad \hat{v}_0(\ell) = \hat{m}_{0,+}(\ell) - \hat{m}_{0,-}(\ell). \tag{12}$$

[9] Since the null hypothesis and the test statistic are invariant to strictly monotonic transformations of Y, this standardization does not affect the theoretical guarantee and the empirical results of our test.

iii. For each $\ell \in \mathcal{L}^*$, calculate a sample analog of the influence function

$$\hat{\phi}_{\nu_1,i}(\ell) = \sqrt{nh}\left(w_{n,i}^{-} \cdot (g_\ell(Y_i)D_i - \hat{m}_{1,-}(\ell)) - w_{n,i}^{+} \cdot (g_\ell(Y_i)D_i - \hat{m}_{1,+}(\ell))\right),$$
$$\hat{\phi}_{\nu_0,i}(\ell) = \sqrt{nh}\left(w_{n,i}^{+} \cdot (g_\ell(Y_i)(1-D_i) - \hat{m}_{0,+}(\ell)) - w_{n,i}^{-} \cdot (g_\ell(Y_i)(1-D_i) - \hat{m}_{0,-}(\ell))\right),$$

where the definitions of the weighting terms $\{(w_{n,i}^{+}, w_{n,i}^{-}) : i = 1, \ldots, n\}$ are given in Appendix A. We then estimate the asymptotic standard deviation $\sigma_{P,d}(\ell)$ by $\hat{\sigma}_d(\ell) = \sqrt{\sum_{i=1}^{n} \hat{\phi}_{\nu_d,i}^2(\ell)}$ and obtain the trimmed estimators as $\hat{\sigma}_{d,\xi}(\ell) = \max\{\xi, \hat{\sigma}_d(\ell)\}$.[10]

iv. Calculate the test statistic $\widehat{S}_n$ defined in equation (9) with $\mathcal{L}$ replaced by $\mathcal{L}^*$.

v. Let a_n and B_n be sequences of non-negative numbers. For $d = 0, 1$ and $\ell \in \mathcal{L}$, define $\psi_{n,d}(\ell)$ as

$$\psi_{n,d}(\ell) = -B_n \cdot 1\left\{\frac{\sqrt{nh} \cdot \hat{\nu}_d(\ell)}{\hat{\sigma}_{d,\xi}(\ell)} < -a_n\right\}. \tag{13}$$

Following Andrews and Shi (2013, 2014), we use $a_n = (0.3\ln(n))^{1/2}$ and $B_n = (0.4\ln(n)/\ln\ln(n))^{1/2}$.

vi. Draw $U_1, U_2, \cdots U_n$ as i.i.d. standard normal random variables that are independent of the original sample. Compute the bootstrapped processes, $\widehat{\Phi}_{\nu_1}(\ell)$ and $\widehat{\Phi}_{\nu_0}(\ell)$, defined by

$$\widehat{\Phi}_{\nu_1}(\ell) = \sum_{i=1}^{n} U_i \cdot \hat{\phi}_{\nu_1,i}(\ell), \quad \widehat{\Phi}_{\nu_0}(\ell) = \sum_{i=1}^{n} U_i \cdot \hat{\phi}_{\nu_0,i}(\ell).$$

vii. Iterate Step (vi) $\bar{B}$ times ($\bar{B}$ is a large integer) and denote the realizations of the bootstrapped processes by $\left(\widehat{\Phi}_{\nu_1}^b(\cdot), \widehat{\Phi}_{\nu_0}^b(\cdot) : b = 1, \ldots, \bar{B}\right)$. Let $\hat{q}(\tau)$ be the τ-th empirical quantile of $\left\{\sup_{d\in\{0,1\},\, \ell\in\mathcal{L}^*}\left\{\frac{\widehat{\Phi}_{\nu_d}^b(\ell)}{\hat{\sigma}_{d,\xi}(\ell)} + \psi_{n,d}(\ell)\right\} : b = 1, \ldots, \bar{B}\right\}$. For significance level $\alpha < 1/2$, obtain a critical value of the test $\hat{c}_\eta(\alpha)$ by $\hat{c}_\eta(\alpha) = \hat{q}(1 - \alpha + \eta) + \eta$, where $\eta > 0$ is an arbitrarily small positive number, e.g., 10^{-6}.[11]

viii. Reject H_0 if $\widehat{S}_n > \hat{c}_\eta(\alpha)$.

Remark 6. Following the existing papers in the moment inequality literature, Step vii in Algorithm 1 uses the generalized moment selection (GMS) proposed by Andrews and Soares (2010) and Andrews

[10] In the simulations, we set $\xi = \sqrt{a(1-a)}$, where $a = 0.0001$. We also use $a \in \{0.001, 0.03, 0.5\}$. The results are insensitive to the choice of a. These tuning parameters are motivated by the observation that the denominator of the asymptotic variance takes the form of $p_\ell(1-p_\ell)$, where $p_\ell = \lim_{r\to r_0} \mathbb{P}(Y \in C_\ell, D = d | R = r)$.

[11] This η constant is called an infinitesimal uniformity factor and is introduced by Andrews and Shi (2013) to avoid the problems that arise due to the presence of the infinite-dimensional nuisance parameters $\nu_{P,1}(\ell)$ and $\nu_{P,0}(\ell)$.

and Shi (2013). It is similar to the recentering method of Hansen (2005) and Donald and Hsu (2016), and the contact set approach of Linton, Song, and Whang (2010). By estimating the null distribution from data, employing GMS can result in a higher finite sample power of the test compared with the test that sets the null distribution at a least favorable configuration.

Remark 7. Regarding the bandwidths for local linear estimators in step ii, our informal recommendation is to have the bandwidth of $\hat{m}_{d,+}(\ell)$, $d = 1, 0$, common for all $\ell \in \mathcal{L}^*$ and the bandwidth of $\hat{m}_{d,-}(\ell)$, $d = 1, 0$, common for all $\ell \in \mathcal{L}^*$. We denote them by h_+ and h_-, respectively, and allow $h_+ \neq h_-$. We consider there is a merit of using the bandwidths that are recommended for the point estimation of the complier's average effect at the cut-off, such as Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014), and Arai and Ichimura (2016), with undersmoothing. To explain why, suppose we estimate the complier's average causal effects using the local linear regressions with the bandwidths h_+ (above the cut-off) and h_- (below the cut-off). As in Imbens and Rubin (1997) for the standard LATE model, it can be shown that the FRD-Wald estimator is numerically equal to the difference of the means between the following distribution function estimates for compliers:

$$\hat{F}_{Y_1(r_0)|\mathbf{C},R=r_0}(y) = \frac{\hat{m}_{1,+}((-\infty, y)) - \hat{m}_{1,-}((-\infty, y))}{\hat{\pi}^+ - \hat{\pi}^-},$$
$$\hat{F}_{Y_0(r_0)|\mathbf{C},R=r_0}(y) = \frac{\hat{m}_{0,-}((-\infty, y)) - \hat{m}_{0,+}((-\infty, y))}{\hat{\pi}^+ - \hat{\pi}^-},$$

where $\hat{m}_{1,+}((-\infty, y))$ and $\hat{m}_{0,+}((-\infty, y))$ use h_+, $\hat{m}_{1,-}((-\infty, y))$ and $\hat{m}_{0,-}((-\infty, y))$ use h_-, and $\hat{\pi}^+$ and $\hat{\pi}^-$ are the local linear estimators for $\lim_{r \downarrow r_0} P(D = 1|R = r)$ and $\lim_{r \uparrow r_0} P(D = 1|R = r)$ with bandwidths h_+ and h_-, respectively. Accordingly, reusing these bandwidths to perform our test, we assess nonnegativity of the complier's potential outcome densities based on the same in-sample information as that the point estimate for complier's causal effect relies on. We believe this is sensible since our test aims to be used for specification check to assess credibility of the causal effect estimates.[12]

Algorithm 1 provides some default choices for other tuning parameters, ξ, a_n, and B_n, without arguing any optimality justification. According to our Monte Carlo studies and empirical applications

[12]Alternatively, we may want to choose bandwidths so as to optimizing a power criterion of the test. We leave power-optimizing choices of bandwidth for future research.

considered in Sections 4 and 5, the test results are not sensitive to mild departures from their default choices.

4. Simulation

This section investigates the finite sample performance of the proposed test by conducting extensive Monte Carlo experiments. We consider nine data generating processes (DGPs) including three DGPs, Size1 - Size3, for examining the size property and six DGPs, Power1 - Power6, for examining the power property. For all DGPs, we set the cutoff point at $r_0 = 0$. We specify these DGPs either by formulating the structural equations (Size1, Power1, and Power2) or by directly specifying the distribution of observables (the rest of DGPs).

4.1. Size properties.

We consider the following three designs to demonstrate the size property of our test.

Size1 Let

$$(\epsilon_1, \epsilon_0, R, V)' \sim N(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} 0.25 & 0 & 0. & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0.25 \end{pmatrix}.$$

So ϵ_1, ϵ_0, R and V are mutually independent. Define

$$\begin{cases} Y_1 = \epsilon_1, \\ Y_0 = -3 + Y_1 - \epsilon_0, \\ D = \mathbf{1}\{V \geq 0\}. \end{cases} \tag{14}$$

Size2 Let $R \sim N(0,1)$ truncated at -2 and 2. The propensity score $P(D = 1|R = r) = 0.5$ for all r. $Y|(D = 1, R = r) \sim N(1,1)$ for all r and $Y|(D = 0, R = r) \sim N(0,1)$ for all r.

Size3 Same as Size2 except that

$$P(D = 1|R = r) = \mathbf{1}\{-2 \leq r < 0\}\frac{(r+2)^2}{8} + \mathbf{1}\{0 \leq r \leq 2\}\left(1 - \frac{(r-2)^2}{8}\right).$$

In all these DGPs, the propensity scores are continuous at the cut-off (i.e., Assumption 3 does not hold). Combined with FRD-validity (Assumptions 1 and 2), the distributions of the observables are also continuous at the cut-off, implying that these DGPs correspond to least favorable nulls in our test. Size1 and Size2 have constant propensity scores, while in Size3, the left- and right-derivatives of the propensity scores differ at the cut-off. As a result, the first-order bias term in the local linear estimation in Size3 is nonzero.

For each DGP, we generate random samples of four sizes: 1000, 2000, 4000 and 8000. We consider three data-driven choices of bandwidths: Imbens and Kalyanaraman (2012, IK), Calonico, Cattaneo, and Titiunik (2014, CCT) and Arai and Ichimura (2016, AI), which are tuned to minimize the mean squared errors of the causal effect estimator. See Remark 7 for our rationale for using them in our test. For each bandwidth, we impose undersmoothing by multiplying $n^{\frac{1}{5}-\frac{1}{c}}$ with $c = 4.5$ to each bandwidth.[13] We specify $\mathcal{L}^* = \mathcal{L}_{coarse}$ with $Q = 15$. We note that for the DGPs we consider in this simulation exercise, our test produces very similar results when Q is greater than 10. For each simulation design, we conduct 1000 repetitions with bootstrap iterations $\bar{B} = 300$.

TABLE 1. Size Property

		AI			IK			CCT		
DGP	n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Size1	1000	0.008	0.032	0.076	0.006	0.043	0.081	0.018	0.056	0.128
	2000	0.009	0.041	0.078	0.011	0.039	0.095	0.010	0.078	0.122
	4000	0.007	0.034	0.073	0.006	0.044	0.090	0.012	0.054	0.102
	8000	0.008	0.045	0.073	0.010	0.048	0.088	0.006	0.062	0.110
Size2	1000	0.012	0.037	0.081	0.004	0.040	0.075	0.016	0.066	0.142
	2000	0.007	0.041	0.077	0.010	0.046	0.092	0.016	0.064	0.122
	4000	0.012	0.043	0.087	0.013	0.050	0.095	0.014	0.052	0.128
	8000	0.003	0.025	0.045	0.003	0.020	0.045	0.016	0.068	0.114
Size3	1000	0.006	0.032	0.067	0.007	0.033	0.069	0.006	0.044	0.078
	2000	0.004	0.031	0.066	0.005	0.043	0.077	0.008	0.038	0.110
	4000	0.005	0.028	0.063	0.009	0.039	0.072	0.014	0.042	0.100
	8000	0.003	0.013	0.043	0.005	0.015	0.058	0.006	0.036	0.068

Table 1 demonstrates that overall the proposed test controls size well over these designs, in particular when sample size is large.

4.2. **Power properties.** To investigate the power property, we consider the following six DGPs that violate the FRD-validity conditions. For example, The Power2 DGP violates the LM assumption but the LC condition is satisfied. In Power1, both the conditional distributions of Y_1 and Y_0 violates

[13]We run simulations for other choices of under-smoothing constant $c \in [3, 5)$; the results are similar.

the LC condition. In Power3-Power6, the conditional distribution of Y_1 violates the LC condition in different ways (location shift, scale change, or change on the whole distribution).

Power1 Violation of local continuity (Assumption 2). $(\epsilon_1, \epsilon_0, R, V)$ follows the same joint distribution as in Size1, but

$$\begin{cases} Y_1 = \mathbf{1}\{R > 0\} + 1.25R\mathbf{1}\{R \le 0\} + \epsilon_1, \\ Y_0 = -8 + 0.5R\mathbf{1}\{R > 0\} + 3\mathbf{1}\{R \le 0\} + \epsilon_0, \\ D = \mathbf{1}\{Y_1 - 0.1Y_0 + \mathbf{1}\{R \ge 0\} > 4V\}. \end{cases} \tag{15}$$

Power2 Violation of local monotonicity (Assumption 4). $(\epsilon_1, \epsilon_0, R, V)$ follows the same joint distribution as in Size1, but

$$\begin{cases} Y_1 = R + \epsilon_1, \\ Y_0 = -4 + Y_1 + \epsilon_0, \\ D = \mathbf{1}\{Y_1 - 0.1Y_0 + 2V\mathbf{1}\{R \ge 0\} > 0\}. \end{cases} \tag{16}$$

Power3 Let $R \sim N(0,1)$ truncated at -2 and 2. The propensity score is given by

$$P(D = 1|R = r) = \mathbf{1}\{-2 \le r < 0\} \max\{0, (r+2)^2/8 - 0.01\} \\ + \mathbf{1}\{0 \le r \le 2\} \min\{1, 1 - (r-2)^2/8 + 0.01\}$$

Let $Y|(D = 0, R = r) \sim N(0,1)$ for all $r \in [-2,2]$, and $Y|(D = 1, R = r) \sim N(0,1)$ for all $r \in [0,2]$. Let $Y|(D = 1, R = r) \sim N(-0.7,1)$ for all $r \in [-2,0)$.

Power4 Same as Power3 except that $Y|(D = 1, R = r) \sim N(0, 1.675^2)$ for all $r \in [-2,0)$.

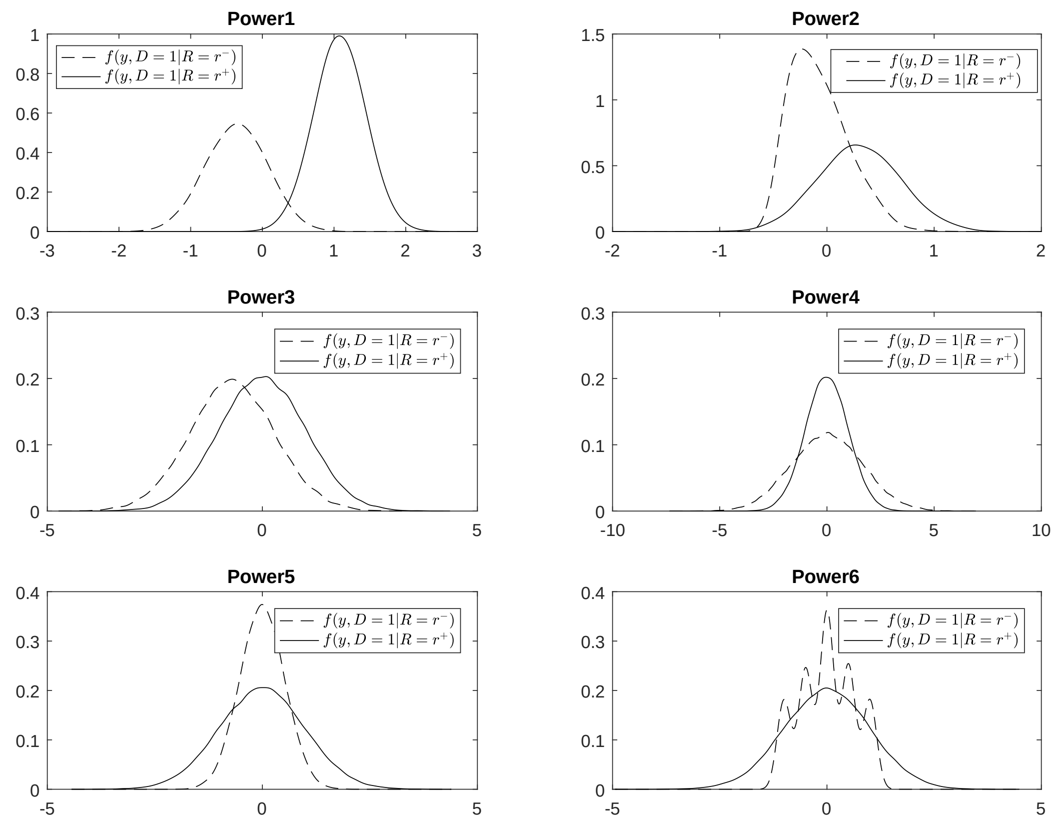
Power5 Same as Power3 except that $Y|(D = 1, R = r) \sim N(0, 0.515^2)$ for all $r \in [-2,0)$.

Power6 Same as Power3 except that $Y|(D = 1, R = r) \sim \sum_{j=1}^{5} \omega_j N(\mu_j, 0.125^2)$ for all $r \in [-2,0)$, where $\omega = (0.15, 0.2, 0.3, 0.2, 0.15)$ and $\mu = (-1, -0.5, 0, 0.5, 1)$.

Figure 2 plots the potential outcome density at the cut-off for each of Power1 - Power6. The testable implication of Theorem 1 (i) requires that the solid curves should lie above the dashed curve everywhere. We can also see that Power1 and Power2 are two designs which give "stronger" violations than Power3 - Power6, and hence we expect to see higher rejection rates for them.

Table 2 reports simulation results for the power property of our test. Overall, our test has good power in detecting deviations from the null under all the three bandwidth choices. As expected

FIGURE 2. Potential Outcome Densities



earlier, we reject more often in the first two designs. It is also interesting to see that it is harder for our test to reject in Power6. From Figure 2 we can see that the violation of null in Power6 occurs sharply with many peaks over narrow intervals, whereas in other designs (e.g. Power3 and Power4) mild violation occurs over relatively wide intervals. This phenomenon is consistent with what has been noted in the literature: the Bierens (1982) and Andrews and Shi (2013) type method that we adopt in this paper is efficient in detecting the latter type violations, see Chernozhukov, Lee, and Rosen (2013, footnote 10) for related discussions.

In Power1-Power6 DGPs, the running variable densities are all continuous at the cutoff (they are normal distributions). Although these DGPs violate FRD-validity testable implications, any

TABLE 2. Power Property

DGP	n	AI 1%	AI 5%	AI 10%	IK 1%	IK 5%	IK 10%	CCT 1%	CCT 5%	CCT 10%
Power1	1000	0.998	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
	2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	8000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Power2	1000	0.698	0.875	0.930	0.541	0.809	0.912	0.490	0.768	0.885
	2000	0.859	0.960	0.979	0.836	0.956	0.985	0.790	0.956	0.987
	4000	0.978	0.998	0.998	0.988	0.998	0.999	0.992	1.000	1.000
	8000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Power3	1000	0.125	0.265	0.381	0.076	0.182	0.289	0.049	0.135	0.211
	2000	0.289	0.510	0.641	0.218	0.436	0.550	0.126	0.278	0.396
	4000	0.606	0.784	0.851	0.611	0.766	0.829	0.374	0.608	0.735
	8000	0.924	0.977	0.985	0.932	0.979	0.990	0.791	0.901	0.933
Power4	1000	0.038	0.158	0.240	0.026	0.106	0.198	0.013	0.074	0.128
	2000	0.130	0.292	0.402	0.104	0.249	0.374	0.045	0.159	0.250
	4000	0.345	0.570	0.692	0.353	0.581	0.693	0.159	0.384	0.510
	8000	0.743	0.889	0.942	0.763	0.902	0.949	0.547	0.761	0.869
Power5	1000	0.071	0.182	0.291	0.034	0.110	0.214	0.027	0.091	0.181
	2000	0.124	0.302	0.431	0.109	0.270	0.382	0.069	0.183	0.286
	4000	0.357	0.581	0.702	0.313	0.569	0.689	0.178	0.374	0.516
	8000	0.752	0.895	0.935	0.760	0.896	0.934	0.541	0.751	0.826
Power6	1000	0.033	0.118	0. 222	0.014	0.072	0.146	0.003	0.044	0.096
	2000	0.083	0.186	0.281	0.038	0.150	0.248	0.020	0.083	0.172
	4000	0.133	0.329	0.448	0.121	0.285	0.433	0.069	0.178	0.302
	8000	0.367	0.586	0.713	0.326	0.569	0.694	0.173	0.378	0.509

running variable density tests cannot reject more often than the pre-specified significance level in large sample, corresponding to the first scenario in Equation (5).

5. APPLICATIONS

To illustrate that our test can provide new insights to empirical practice, we assess FRD-validity in the designs studied in Angrist and Lavy (1999, AL hereafter) and Miller, Pinto, and Vera-Hernández (2013, MPV hereafter). AL study test score data in Israeli public schools to estimate the causal effect of class size on student performance. MPV analyze the Colombian household data to evaluate the impact of the “Colombia’s Régimen Subsidiado", a publicly financed insurance program, on various outcome variables including measurements of household’s financial risk protection, use of medical care, and health outcomes.

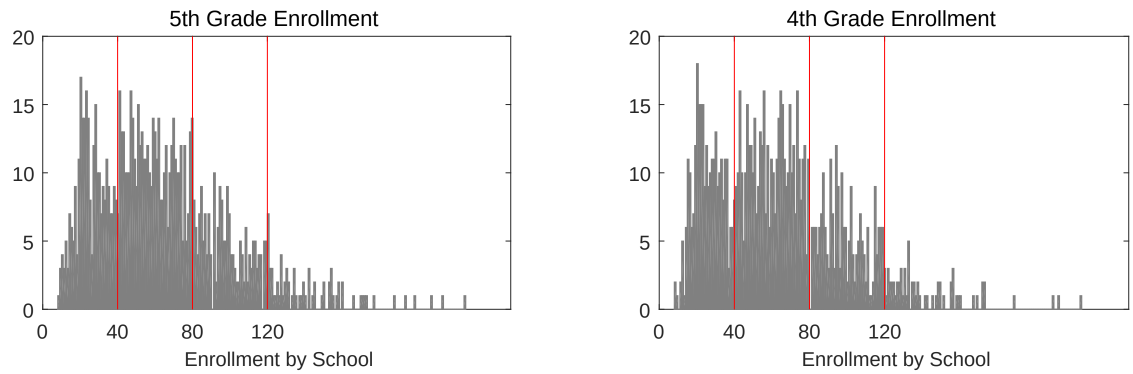
5.1. **Maimonides' Rule in Israel.** Israel has been implementing Maimonides' rule in public schools since 1969. The rule limits a class size up to 40 and therefore creates discontinuous changes in the average class size as the total enrollment exceeds multiples of 40. For example, a public school with 40 enrolled students in a grade can maintain one class, with (average) class size 40; another public school with 41 enrolled students has to offer two classes, hence the average class size drops discontinuously from 40 to 20.5. Maimonides' rule offers an example of FRD design since some schools in the data do not comply the treatment assignment rule.

Recent empirical evidence suggests that the density of the running variable (enrollment) is discontinuous near some cut-offs (Otsu, Xu, and Matsushita (2013) and Angrist, Lavy, Leder-Luis, and Shany (2017)). Along with the argument of Lee (2008) and McCrary (2008), this evidence raises concerns about FRD-validity in this application.

Class size manipulation by parents. As argued in AL, parents may selectively exploit Maimonides' rule by either (a) registering their children into schools with enrollments just above multiples of 40, hoping their children being placed in smaller size classes, or (b) withdrawing children from the public schools with enrollments just below multiples of 40. The possibility of (a) would lead to an increase in the density of enrollment counts just above a multiple of 40, while that of (b) would lead to a decrease of the enrollment density just below the multiples of 40. In either case, we expect to observe discontinuities of the density of the running variable at the cut-offs, as we can observe most notably at the enrollment count of 40 in Figure 3. The class size manipulation by parents can be a serious threat to the local continuity assumption. The parents who act according to (a) value more the small class-size education and tend to be wealthier and more concerned with the children's education. Children with such parents could perform better. If this is the case, the potential outcome distributions of the student's test scores conditional on the running variable may shift abruptly across the cut-off, leading to violation of local continuity.

Class size manipulation by the school board. On the other hand, AL defend FRD-validity by arguing that manipulation of the class sizes by parents is less likely. Concerning the possibility of (a), AL claim that: " *there is no way (for the parents) to know (exactly) whether a predicted enrollment of 41 will not decline to 38 by the time school starts, obviating the need for two small classes*". For the possibility of (b), private elementary schooling is rare in Israel so that withdrawing

FIGURE 3. Histograms for Enrollments by Schools: Panel A of Figure 6 in Angrist, Lavy, Leder-Luis, and Shany (2017)



is not a feasible option for most parents. Angrist, Lavy, Leder-Luis, and Shany (2017) recently re-investigate Maimonides' rule and argue that the manipulation is operated mainly on the school board side. Angrist, Lavy, Leder-Luis, and Shany (2017) state that:“ *A recent memo from Israeli Ministry of Education (MOE) officials to school leaders admonishes headmasters against attempts to increase staffing ratios through enrollment manipulation. In particular, schools are warned not to move students between grades or to enroll those who are overseas so as to produce an additional class.*" Although this type of manipulation can lead to the density jump observed in Figure 3, it is not necessarily a serious threat to FRD-validity depending on the school board's incentives to manipulate. As argued in Angrist, Lavy, Leder-Luis, and Shany (2017), if the main motivation of the manipulations is to increase their budgets — an increasing function of the number of classes, it may well be the case that the manipulations around the cut-off is done independently of the children's unobserved talents. Then, FRD-validity can hold even when the density of the running variable is discontinuous at the cut-offs.

Test Results. Our test focuses on the joint distribution of the observed outcomes and treatment status in contrast to the tests that focus only on the marginal distribution of the running variable. Hence, our test provides new empirical evidence that can contribute to the dispute about FRD-validity of Maimonides' rule reviewed above.

We apply the test proposed in Section 3 for each of the four outcome variables (grade 4 math and verbal test scores, and grade 5 math and verbal test scores) by treating the three cutoffs 40, 80, and

120, separately. We consider the bandwidths ($h_+ = h_- = 3$ and $h_+ = h_- = 5$) used in AL, as well as the three data-driven bandwidth choices (AI, IK and CCT).[14] We set the trimming constant at $\xi = 0.00999$ as described in Algorithm 1 of Section 3.[15]

Table 3 displays the p-values of the tests. For all the cases considered, we do not reject the null hypothesis at 10% significance level. The results are robust to the choice of bandwidths and the choice of trimming constants, see Tables 6 to 8 in Appendix D for details. Despite that the density of running variable appears discontinuous at the cut-off, "no rejection" of our test suggests empirical support for the argument of "manipulation by the school board"—the type of manipulation which is relatively innocuous for the AL's identification strategy.

TABLE 3. Testing Results for Israeli School Data: p-values, $\xi = 0.00999$

	3	5	AI	IK	CCT
g4math					
Cut-off 40	0.986	0.934	0.767	0.978	0.968
Cut-off 80	0.909	0.865	0.715	0.944	0.888
Cut-off 120	0.443	0.702	0.665	0.604	0.568
g4verb					
Cut-off 40	0.928	0.627	0.465	0.648	0.529
Cut-off 80	0.911	0.883	0.185	0.906	0.720
Cut-off 120	0.935	0.683	0.474	0.730	0.186
g5math					
Cut-off 40	0.876	0.282	0.488	0.631	0.609
Cut-off 80	0.516	0.446	0.930	0.482	0.765
Cut-off 120	0.939	0.827	0.626	0.883	0.838
g5verb					
Cut-off 40	0.594	0.893	0.953	0.906	0.938
Cut-off 80	0.510	0.692	0.504	0.525	0.929
Cut-off 120	0.696	0.811	0.601	0.699	0.781

5.2. **Colombia's Subsidized Regime.** MPV study the impact of the Subsidized Regime (SR, targeted to the poor households in Colombia) on financial risk protection, service use, and health

[14]See Table 15 in Appendix D for the obtained bandwidths and the number of observations therein.

[15]We try different choices for the trimming constant $\xi \in \{0.0316, 0.1706, 0.5\}$ and obtain similar results.

outcomes. The SR is a publicly financed health insurance program that subsidizes eligible Colombians to purchase insurance from private and government-approved insurers. The program eligibility is determined by a threshold rule based on a continuous index called Sistema de Identificacion de Beneficiarios (SISBEN) ranging from 0 to 100 (with 0 being the most impoverished). SISBEN is constructed by a proxy means-test using fourteen different measurements of the household's well-being. It is, however, well known that the original SISBEN index used to assign the actual program eligibility was manipulated by either the households or the administrative authorities (see MPV and the references therein for details). To circumvent this manipulation issue, MPV simulate their own SISBEN index for each household using a collection of survey data from independent sources. MPV then estimate a cut-off of the simulated SISBEN scores in each region by maximizing the in-sample prediction performance for the actual program take-up. Using the cut-offs thus estimated, MPV estimate the complier's effects of the SR on 33 outcome variables in four categories: (i) risk protection, consumption smoothing, and portfolio choice, (ii) medical care use, (iii) health status, and (iv) behavior distortions, see Table 9 below and Table 1 of MPV for details.

Although the density of the simulated SISBEN score passes the continuity test (see MPV's online Appendix C), it does not necessarily imply FRD-validity, e.g., the conditional distributions of the potential outcomes given the simulated SISBEN score may not be continuous at the cutoff.

For each of the 33 outcome variables, we implement our test using MPV's simulated SISBEN score as the running variable and the actual program enrollment as the treatment status. We consider the three bandwidths ($h_+ = h_- = 2, 3,$ and 4) used in MPV as well as the three data-driven bandwidth choices (AI, IK and CCT).[16] We use the same set of trimming constants ξ as in the AL application and again obtain similar results. We find robust evidence of rejecting the testable implications of FRD-validity for the following three outcome variables: "household education spending", "total spending on food", and "total monthly expenditure". Their p-values are reported in Table 4 (results for all other outcome variables and other choices of ξ are collected in Tables 9-12 in Appendix D to save space).

A few remarks are in order. First, as can be seen in Tables 9-12, the three outcome variables giving the robust rejections all belong to the first category: "risk protection, consumption smoothing, and portfolio choice". Our test does not provide evidences against the validity of other outcome variables.

[16]See Table 16 in Appendix D for the obtained bandwidths and the number of observations therein.

TABLE 4. Testing Results for Columbia's SR Data: p-values ($\xi = 0.00999$)

	MPV bandwidths			Other bandwidth choice		
Outcome variables	2	3	4	AI	IK	CCT
Household education spending	0.000	0.000	0.000	0.000	0.013	0.000
Total spending on food	0.000	0.000	0.000	0.000	0.000	0.000
Total monthly expenditure	0.000	0.000	0.000	0.000	0.000	0.000

The low p-values for these outcomes remain to be significant even when we take into account the multiple-testing of a group of outcome variables with the family-wise error rate (FWER) control. The results shown in Table 8 of Appendix D imply that, for the first category of 10 outcome variables, the multiple testing procedure of Holm (1979) concludes that H_0 is rejected at the control of FWER at 1%. With all the outcomes (33 hypotheses), H_0 is rejected at the control of FWER at 5%.

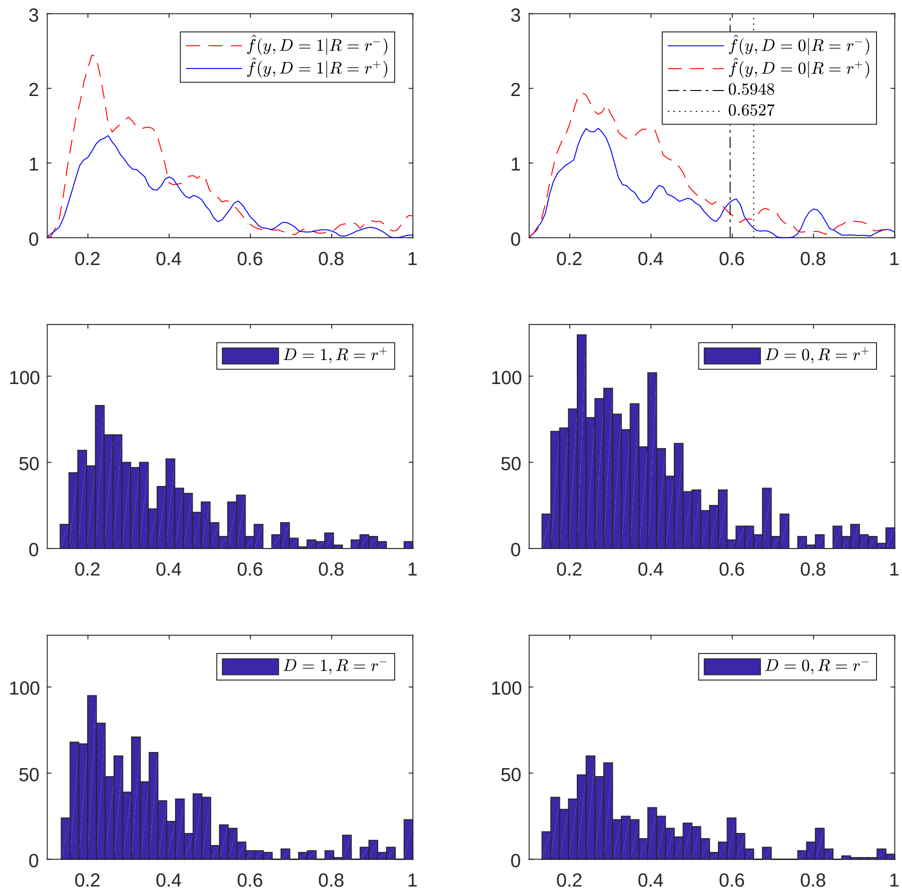
Second, it is possible to figure out what observations cause the rejection of the FRD-validity. Take the choice of $\mathcal{L}^* = \mathcal{L}_{coarse}$ with $Q = 15$ and bandwidth $h^+ = h^- = 2$ as an example (other choices similar). For the outcome variable "Total spending on food", the supremum in the test statistic is achieved at $d = 0$ and $[y, y'] = [0.5948, 0.6527]$. Figure 4 draws the kernel smoothed (pseudo) densities of normalized outcome variable, where the blue curve should be underneath the red curve under the FRD-validity. The two density curves in the right top panel indeed cross at this interval. The histograms and Table 5 show that there are 45 observations with $D = 0, R \in (r_0 - 2, r_0)$ and $Y \in [0.5948, 0.6527]$, which are about 6.73% of all observations with $R \in (r_0 - 2, r_0)$ and $D = 0$. On the other hand, there are 31 observations with $D = 0, R \in [r_0, r_0 + 2)$ and $Y \in [0.5948, 0.6527]$, which is about 2.06% of all observations with $R \in [r_0, r_0 + 2)$ and $D = 0$. We can extract the same type of information for the other two outcomes variables. For "Household education spending", the supremum of the test statistics is attained at $d = 0, R \in [r_0 - 2, r_0)$, and $Y \in [0.5604, 0.6693]$. For "Total monthly expenditure", the supremum is attained at $d = 1, R \in [r_0, r_0 + 2]$, and $Y \in [0.3042, 0.3737]$, that contains 61 observations. To save space, we collect figures (Figures 6 and 7) and tables (Tables 17 and 18) for "Household education spending" and "Total monthly expenditure" in Appendix D.

Third, we also implement the tests conditioning on each of the six regions in Colombia. The results are collected in Table 13 in Appendix D. We obtain strong rejections in the "Atlantica", "Oriental", "Central", and "Bogota" regions, and no rejection in "Pacifico" and "Territorios Nacionales". Taking

TABLE 5. Obs. in the maximizer interval ($h^+ = h^- = 2$): Total spending on food

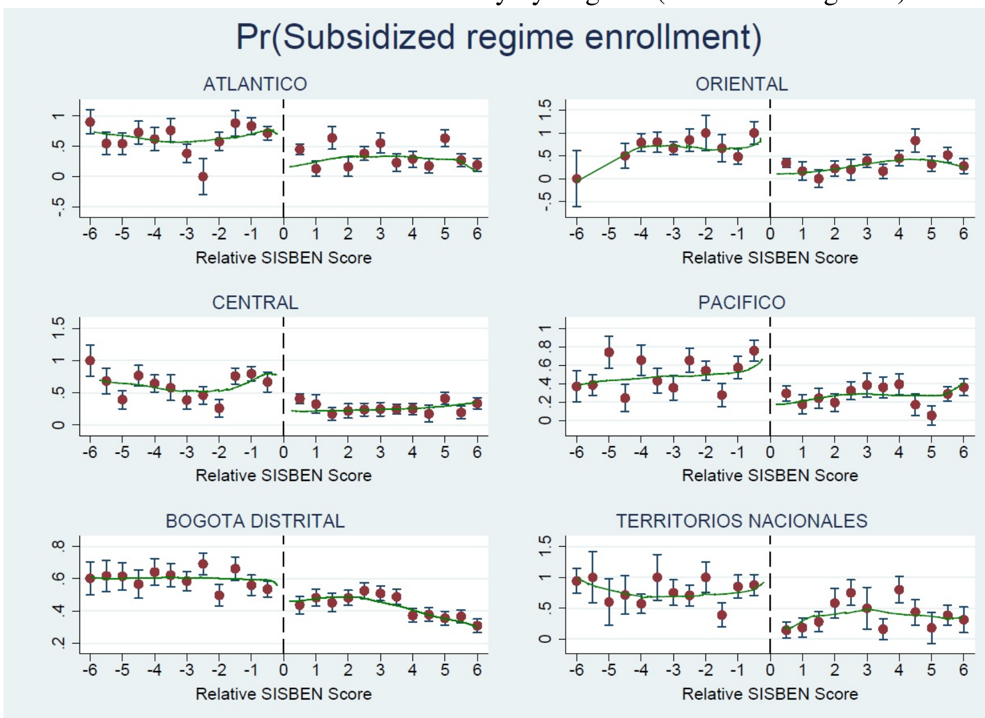
	# of observations		
Subsample of	All	$\{0.5948 \leq Y \leq 0.6527\}$	Ratio
$\{0 \leq R < h^+\} \cap \{D = 0\}$ ($\mathbf{N} \cup \mathbf{C}$)	1502	31	2.06%
$\{h^- < R < 0\} \cap \{D = 0\}$ ($\mathbf{N}$)	669	45	6.73%

FIGURE 4. Estimated complier's outcome density: Total spending on food



into account the relative sample sizes across the regions (Table 14), the Bogota sample seems to drive the test results of Table 4. Notice that the magnitude of the propensity score jump for the Bogota sample is relatively small compared with the samples in the regions giving no-rejections (see Figure 5). This observation is in line with Remark 3 of Section 2.

FIGURE 5. Enrollment Probability by Regions (from MPV Figure 2)



There can be multiple causes for why FRD-validity fails in this application. First, violation of local continuity may arise as a byproduct of estimating the cut-off based on the simulated SISBEN score. For instance, if there is some household characteristic that is *not* included in the construction of the simulated SISBEN score but has a strong predictive power for program enrollment, the estimated cut-off may pick up a value of the simulated SISBEN score across which the distributions of the excluded characteristics differ most. If the distribution of household consumption variables well depend on such excluded characteristic, it would result in the violation of local continuity. Second, there could be other *unobserved* programs using the same SISBEN index with similar cutoffs. If such programs affect the household budget significantly, we may expect the distribution of potential household consumption to be quite different between the two sides of the cutoff, again leading to the violation of local continuity.[17]

[17] MVP suggest that the second channel is less likely to be the cause of the rejection of FRD validity of the three outcome variables. See Table 2 in MPV for evidence that the enrollment rates for other programs do not change across the estimated cut-offs.

6. Extensions

In this section, we briefly discuss several extensions. Those include incorporating covariates and combining our test with some of the existing tests. We also discuss the testable implications of other FRD assumptions that have been used in the literature.

6.1. **Incorporating Covariates.** In the standard FRD design, it suffices to consider three variables: the outcome, treatment status, and running variables, as we have done so far. Although the identification of the treatment effect at the cut-off does not require covariates, they are often included in empirical studies to increase efficiency. See Imbens and Kalyanaraman (2012), Calonico, Cattaneo, Farrel, and Titiunik (2016), and Hsu and Shen (2019) for more detailed discussion. Another motivation for incorporating the conditioning covariates arises when their distributions are suspected to be discontinuous at the cut-off. If the potential outcome distributions depend on such covariates, RD analysis without conditioning on them leads to violation of the local continuity assumption. See Frölich and Huber (2018).

In what follows, we consider testing a version of FRD-validity where the 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 cutoff point conditioning on each realization of x. The conditional version of FRD-validity is stated formally as follows:

Assumption 6. *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 $x \in \mathcal{X}$.*

Assumption 7 (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 8 (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 6-8: for C a hypercube in $\mathcal{Z}$:

$$\lim_{r\uparrow r_0}\mathbb{E}_P[1\{Z\in C\}D|R=r]-\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Z\in C\}D|R=r]\leq 0, \tag{17}$$

$$\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Z\in C\}(1-D)|R=r]-\lim_{r\uparrow r_0}\mathbb{E}_P[1\{Z\in C\}(1-D)|R=r]\leq 0. \tag{18}$$

In comparison to inequalities (2) and (3), the only difference is that the indicator functions in (17) and (18) index boxes in $\mathcal{Z}$ instead of the intervals in $\mathcal{Y}$. Accordingly, by defining a class of instrument functions as

$$\begin{aligned}
&\mathcal{G}_z=\{g_\ell(\cdot)=1(\cdot\in C_\ell):\ell\equiv(z,c)\in\mathcal{L}\}\text{, where}\\
&C_\ell=\times_{j=1}^{d_x+1}[z_j,z_j+c]\cap\mathcal{Z}\text{ and}\\
&\mathcal{L}=\left\{(z,c):\ c^{-1}=q,\text{ and }q\cdot z_j\in\{0,1,2,\cdots,(q-1)\}^{d_x+1}\text{ for }q=1,2,\cdots\right\},
\end{aligned} \tag{19}$$

we can implement the testing procedure of Section 3 to test the conditional version of FRD-validity.

6.2. **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 cutoff is often considered to be a supporting evidence for no-selection around the cut-off, it would be 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*, a joint test can be obtained by modifying the test proposed in Section 3 to account for the additional set of equality constraints.

We hence consider testing inequalities (7) 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 similarly to $\mathcal{L}$ defined in Section 3.

In the same manner as $\hat{v}_{n,d}(\ell)$ is obtained in (12), we estimate $v^x(j)$ by $\hat{v}^x(j)$ the difference of two local linear estimators. Following the way Andrews and Shi (2013) incorporate moment equalities, we modify the KS test statistic of Section 3 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. 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 when the KS statistic is computed.

7. Conclusion

In this paper we propose a test for the key identifying conditions in the fuzzy regression design. We characterize the set of sharp testable implications for FRD-validity and propose an asymptotically valid test for it. Our test makes use of not only the information of running variable but also that of outcome and treatment status. As illustrated in our empirical applications, our test provides empirical evidence for or against FRD-validity, which would have been overlooked if one would have only assessed the continuity of the running variable's density at the cut-off.

A salient feature of our test is that by examining a density plot as done in Figure 4 of Section 5, we can learn about a subpopulation (defined in terms of the range of observed outcomes) that plays a role of refuting FRD-validity. Looking into the background of such subpopulation, we may be able to understand why FRD-validity is violated in the given application. Given the rejection of our test, one possible option to proceed is to relax FRD-validity and pursue the partial identification approach to bound the parameter of interest.

Appendix

We describe how to calculate the proposed test statistic in Appendix A. We formally state the asymptotic validity of our test in Appendix B. All proofs are collected in Appendix C. Additional empirical results of Section 5 are provided in Appendix D.

Appendix A. Calculating the Test Statistics

We first introduce notations. Let $m_{P,d}(\ell, r) = \mathbb{E}_P[g_\ell(Y)D^d(1-D)^{1-d}|R = r]$ and $m_{P,d,+}(\ell) = \lim_{r\downarrow r_0} m_{P,d}(\ell, r)$ and $m_{P,d,-}(\ell) = \lim_{r\uparrow r_0} m_{P,d}(\ell, r)$ for $d = 1, 0$, then we can estimate $\nu_{P,1}(\ell)$ and $\nu_{P,0}(\ell)$ respectively by Equation (12), which we restate below:

$$\hat{\nu}_1(\ell) = \hat{m}_{1,-}(\ell) - \hat{m}_{1,+}(\ell), \quad \hat{\nu}_0(\ell) = \hat{m}_{0,+}(\ell) - \hat{m}_{0,-}(\ell),$$

where the right hand side terms $\hat{m}_{d,\star}(\ell)$, for $d = 1, 0$ and $\star = +, -$, are local linear estimators, which can be constructed by the intercept estimates $\hat{a}_{d,+}(\ell)$ and $\hat{a}_{d,-}(\ell)$ in regressions of the form

$$(\hat{a}_{d,+}(\ell), \hat{b}_{d,+}(\ell)) = \underset{a,b}{\operatorname{argmin}} \frac{1}{nh_+} \sum_{i=1}^{n} 1\{R_i \geq r_0\} \cdot K\left(\frac{R_i - r_0}{h_+}\right) \left[g_\ell(Y_i)D_i^d(1-D_i)^{1-d} - a - b \cdot \left(\frac{R_i - r_0}{h_+}\right)\right]^2,$$

$$(\hat{a}_{d,-}(\ell), \hat{b}_{d,-}(\ell)) = \underset{a,b}{\operatorname{argmin}} \frac{1}{nh_-} \sum_{i=1}^{n} 1\{R_i < r_0\} \cdot K\left(\frac{R_i - r_0}{h_-}\right) \left[g_\ell(Y_i)D_i^d(1-D_i)^{1-d} - a - b \cdot \left(\frac{R_i - r_0}{h_-}\right)\right]^2,$$

where $K(\cdot)$ is a kernel function and (h_+, h_-) are the bandwidths for the running variable specified above and below the cut-off, respectively. In particular, we express $h_+ = c_+h$ and $h_- = c_-h$, with (c_+, c_-) be positive constants and h is a converging sequence indexed by sample size n. For simplicity of analysis and implementation, we specify the bandwidths h_+ and h_- to be the same over $\{g_\ell : \ell \in \mathcal{L}\}$.

We can write the local linear estimators in the following form: for $d = 1, 0$ and $\star = +, -$

$$\hat{m}_{d,\star}(\ell) = \sum_{i=1}^{n} w_{n,i}^{\star} \cdot g_\ell(Y_i)D_i^d(1-D_i)^{1-d}, \tag{20}$$

where the weights are defined as

$$w_{n,i}^{+} = \frac{1}{nh_{+}} \cdot \frac{1\{R_i \geq r_0\} \cdot K(\frac{R_i - r_0}{h_+}) \left[\hat{\vartheta}_2^{+} - \hat{\vartheta}_1^{+} \cdot \left(\frac{R_i - r_0}{h_+}\right)\right]}{\hat{\vartheta}_2^{+}\hat{\vartheta}_0^{+} - (\hat{\vartheta}_1^{+})^2},$$

$$w_{n,i}^{-} = \frac{1}{nh_{-}} \cdot \frac{1\{R_i < r_0\} \cdot K(\frac{R_i - r_0}{h_-}) \left[\hat{\vartheta}_2^{-} - \hat{\vartheta}_1^{-} \cdot \left(\frac{R_i - r_0}{h_-}\right)\right]}{\hat{\vartheta}_2^{-}\hat{\vartheta}_0^{-} - (\hat{\vartheta}_1^{-})^2}.$$

and for $j = 0, 1, 2$,

$$\hat{\vartheta}_j^{+} = \frac{1}{nh_{+}} \sum_{i=1}^{n} 1\{R_i \geq r_0\} \cdot K\left(\frac{R_i - r_0}{h_+}\right) \left(\frac{R_i - r_0}{h_+}\right)^j,$$

$$\hat{\vartheta}_j^{-} = \frac{1}{nh_{-}} \sum_{i=1}^{n} 1\{R_i < r_0\} \cdot K\left(\frac{R_i - r_0}{h_-}\right) \left(\frac{R_i - r_0}{h_-}\right)^j.$$

Appendix B. 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 C.3.

We normalize the support of observed outcome Y to $[0,1]$.[18] 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 $h_2^{(1)}$ and $h_2^{(2)}$, we define the distance between them by

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{\{\ell_1, \ell_2 \in \mathcal{L}\}} |h_2^{(1)}(\ell_1, \ell_2) - h_2^{(2)}(\ell_1, \ell_2)|. \tag{21}$$

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 $\sigma_{P,d,-}(\ell_1, \ell_2) = \lim_{r \uparrow r_0} \sigma_{P,d}(\ell_1, \ell_2|r)$. Existence of these limits is implied by the set of assumptions in Assumption 9 below.

[18]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 done in the first step of Algorithm 1.

For $j = 0,1,2,\dots$, 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

$$h_{2,P,d,\star}(\ell_1,\ell_2) = \frac{\int_0^\infty (\vartheta_2 - u\vartheta_1)^2 K^2(u)du \cdot \sigma_{P,d,\star}(\ell_1,\ell_2)}{c_\star f_R^\star(r_0)(\vartheta_2\vartheta_0 - \vartheta_1^2)^2}, \tag{22}$$

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 $\sqrt{nh}(\hat{v}_d(\ell) - v_{P,d}(\ell))$ is $h_{2,P,d} = h_{2,P,d,+} + h_{2,P,d,-}$.

We denote their v-th derivatives with respect to the running variable by $m_{P,d}^{(v)}$, $d = 1,0$. 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 9. *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 9 (*iii*)-(*v*) imply that with undersmoothing bandwidths, the bias terms of the $\hat{v}_1(\ell)$ and $\hat{v}_0(\ell)$ are asymptotically negligible uniformly over $\ell \in \mathcal{L}$. Note that Assumption 9 does not restrict the support of Y and allows Y to be discrete, continuous, or their mixtures. Note also that we allow $f_R(r)$ to be discontinuous at the cut-off, reflecting that the testable implications of FRD-validity that we are focusing on does not require continuity of $f_R(r)$ at the cut-off.

Assumption 10. *The kernel function $K(\cdot)$ and bandwidth h satisfy* *(i)* $K(\cdot)$ *is non-negative, symmetric, bounded by* $\bar{K} < \infty$*, and has a compact support in* R *(say* $[-1,1]$*).*

(ii) $\int K(u)du = 1$*, and* $\int u^2 K(u)du > 0$*,*

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

Assumption 10 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 11. *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 11 is standard for the multiplier bootstrap (see, e.g., Hsu (2016)). Note the standard normal distribution for U satisfies Assumption 11.

Assumption 12. a_n *is a sequence of non-negative 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 non-negative numbers satisfying that* B_n *is non-decreasing,* $\lim_{n\to\infty} B_n = \infty$ *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 $\mathcal{P}_0$ be the subset of $\mathcal{P}$ that satisfies Assumption 9 such that the null hypothesis in Equation (8) holds under P if $P \in \mathcal{P}_0$. The next assumption states that $\mathcal{P}_0$ contains a distribution of data that satisfies the moment inequalities $\{v_{P,d}(\ell) : d = 0, 1, \ell \in \mathcal{L}\}$ as equalities for some $\ell \in \mathcal{L}$.

Assumption 13. *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 (21).*

(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 2. *Suppose Assumptions 9-12 hold. Then, for every compact subset $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$, the following claims hold for the test 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 13 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 2 (a) shows that our test has asymptotically uniformly correct size over a compact set of covariance kernels. Theorem 2 (b) shows that our test is at most infinitesimally conservative asymptotically when the null contains at least one P_c defined in Assumption 13. Theorem 2 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 3. *Suppose Assumptions 9-12 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

$$\nu_{P_n,d}(\ell) = \nu_{P_c,d}(\ell) + \frac{\delta_d(\ell)}{\sqrt{nh}}, \tag{23}$$

for $d = 1, 0$ and some $P_c \in \mathcal{P}_0$ defined in Assumption 13. We consider local alternatives that satisfy the next set of assumptions:

Assumption 14. *A sequence of local alternatives $\{P_n \in \mathcal{P}\backslash\mathcal{P}_0 : n \geq 1\}$ satisfies the following conditions:*

(i) (23) 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 14 (i) requires that the local alternatives converge to a boundary null P_c at rate $(nh)^{-1/2}$. Assumption 14 (ii) ensures that our test is unbiased and Assumption 14 (iii) makes sure that P_n's are not in $\mathcal{P}_0$. Assumption 14 (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 14.

Theorem 4. *Suppose Assumptions 9 to 12 hold and $\alpha < 1/2$. If a sequence of local alternatives $\{P_n : n \geq 1\}$ satisfies Assumption 14, then $\lim_{\eta\to 0}\lim_{n\to\infty} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) \geq \alpha$.*

APPENDIX C. PROOFS

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

Lemma 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 $B \in \mathcal{B}(\mathcal{Y})$. □

C.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 by the definition of the compliance type. On the other hand, we have

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

where the third line follows by Assumption 1. Hence,

$$
\begin{aligned}
&\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y\in B\}D|R=r_0+\epsilon] \\
&=\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r)\in B, T_{|r-r_0|}=\mathbf{C}\}|R=r_0+\epsilon]+\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r)\in B, T_{|r-r_0|}=\mathbf{A}\}|R=r_0+\epsilon].
\end{aligned}
\tag{24}
$$

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

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

implying

$$
\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y\in B\}D|R=r_0-\epsilon]=\lim_{\epsilon\to 0}\mathbb{E}_P[1\{Y_1(r)\in B, T_{|r-r_0|}=\mathbf{A}\}|R=r_0-\epsilon]. \tag{25}
$$

Taking the difference of equation (24) and equation (25) and employing Assumption 2 lead to the desired inequality:

$$
\begin{aligned}
&\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y\in B\}D|R=r]-\lim_{r\uparrow r_0}\mathbb{E}_P[1\{Y\in B\}D|R=r] \\
&=\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y_1(r)\in B, T_{|r-r_0|}\in\{\mathbf{C}\}\}|R=r]\ge 0.
\end{aligned}
\tag{26}
$$

Similarly we can show that

$$\lim_{r\uparrow r_0}\mathbb{E}_P[1\{Y\in B\}(1-D)|R=r]-\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y\in B\}(1-D)|R=r]$$
$$=\lim_{r\downarrow r_0}\mathbb{E}_P[1\{Y_0(r)\in B, T_{|r-r_0|}\in\{\mathbf{C}\}\}|R=r]\geq 0. \tag{27}$$

Note that the proof is also valid when Assumption 1 is replaced by Assumption 4.

Claim (ii): Suppose that the distribution of observables (Y, D, R) satisfies inequalities (2) and (3). By Lemma 1, they hold for 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 notations of the potential outcomes and reduce them to $(\tilde{Y}_1, \tilde{Y}_0) \in \mathcal{Y}^2$. For the treatment selection response to 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, constructing a joint distribution of $(\tilde{Y}_1(r), \tilde{Y}_0(r), \tilde{D}(r) : r \in \mathcal{R})$ given R is done 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.

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

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

and

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

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 never-taker's potential outcome distributions, consider

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

and

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

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

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

and

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

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 $R = r$ thus constructed 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 complier, 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.

C.2. **Identification of the complier's potential outcome distributions.**

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

$$\begin{aligned} F_{Y_1(r_0)|\mathbf{C},R=r_0}(y) &\equiv \lim_{r\to r_0}P\left(Y_1(r)\le y|T_{|r-r_0|}=\mathbf{C},R=r\right),\\ F_{Y_0(r_0)|\mathbf{C},R=r_0}(y) &\equiv \lim_{r\to r_0}P\left(Y_0(r)\le y|T_{|r-r_0|}=\mathbf{C},R=r\right),\end{aligned}$$

are identified by

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

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

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 (26) 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 (27). We omit the details for brevity. □

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

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 9. This lemma extends the undersmoothing case of Lemma 1 in Chiang, Hsu, and Sasaki (2017) by having an approximation that is uniform also 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 variable 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 2. *Let $\mathcal{P}$ be a class of data generating processes satisfying Assumption 9, $d = 1, 0$, and $\star = +, -$. Under Assumption 10,*

$$\sup_{\{\ell\in\mathcal{L}\}}\left|\sqrt{nh}(\hat{m}_{d,\star}(\ell)-m_{P,d,\star}(\ell))-\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}w_i^{\star}\mathcal{E}_{d,i}(\ell)\right| = o_{\mathcal{P}}(1), \tag{28}$$

where

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

Proof. We provide a proof for the case of $d = 1$ and $\star = +$ only, as 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

$$\sqrt{nh}\left[\hat{m}_{1,+}(\ell) - m_{P,1,+}(\ell)\right]$$
$$= \sqrt{nh^3} \cdot c_+ \sum_{i=1}^n w_{n,i}^+ m_{P,1}^{(1)}(\ell, r_0)\left(\frac{R_i - r_0}{h_+}\right) + \sqrt{nh^5} \cdot \frac{c_+^2}{2} \sum_{i=1}^n w_{n,i}^+ m_{P,1}^{(2)}(\ell, \tilde{R}_i)\left(\frac{R_i - r_0}{h_+}\right)^2 \quad (29)$$
$$+ \sqrt{nh} \cdot \sum_{i=1}^n w_{n,i}^+ \mathcal{E}_{1,i}(\ell) \quad (30)$$

The first order conditions for the local linear regression implies the first term in (29) is zero. By the boundedness of $m_{P,1}^{(2)}$ (Assumption 9) (iv), the absolute value of the second term in (29) 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

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

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

$$\hat{\vartheta}_j^+ = f_R^+(r_0)\vartheta_j + o_{\mathcal{P}}(1) \quad (31)$$

holds where the $\mathcal{P}$-uniform convergence here follows by Assumption 9 (i.e., $\mathcal{P}$ shares the common marginal distribution of R). Hence, combined with the undersmoothing bandwidth (Assumption 10 (iii)), the second term in (29) 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

$$\begin{aligned}
&\sup_{\{\ell\in\mathcal{L}\}}\left|\sqrt{nh}\sum_{i=1}^n w_{n,i}^+\mathcal{E}_{1,i}(\ell) - \frac{1}{\sqrt{nh}}\sum_{i=1}^n w_i^+\mathcal{E}_{1,i}(\ell)\right| \\
&\leq c_+^{-1}\underbrace{\left|\frac{\hat{\vartheta}_2^+}{\hat{\vartheta}_2^+\hat{\vartheta}_0^+ - (\hat{\vartheta}_1^+)^2} - \frac{\vartheta_2}{f_R^+(r_0)(\vartheta_2\vartheta_0 - \vartheta_1^2)}\right|}_{\text{(i)}} \cdot \underbrace{\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{R_i-r_0}{h_+}\right)1\{R_i\geq r_0\}\mathcal{E}_{1,i}(\ell)\right|}_{\text{(ii)}} \\
&\quad + c_+^{-1}\underbrace{\left|\frac{\hat{\vartheta}_1^+}{\hat{\vartheta}_2^+\hat{\vartheta}_0^+ - (\hat{\vartheta}_1^+)^2} - \frac{\vartheta_1}{f_R^+(r_0)(\vartheta_2\vartheta_0 - \vartheta_1^2)}\right|}_{\text{(iii)}} \cdot \underbrace{\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{R_i-r_0}{h_+}\right)\left(\frac{R_i-r_0}{h_+}\right)1\{R_i\geq r_0\}\mathcal{E}_{1,i}(\ell)\right|}_{\text{(iv)}}.
\end{aligned} \tag{32}$$

Since (31) implies both terms (i) and (iii) in (32) are $o_{\mathcal{P}}(1)$, it suffices to show that the terms (ii) and (iv) in (32) are stochastically bounded uniformly in $P\in\mathcal{P}$. Let j be a nonnegative integer and

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

Consider obtaining a $\mathcal{P}$-uniform bounds 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,

$$\begin{aligned}
P\left(\sqrt{n}\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^n f_{n,i}^j(\ell)\right| > \epsilon\right) &\leq \epsilon^{-1}\sqrt{n}\mathbb{E}_P\left[\sup_{\{\ell\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^n f_{n,i}^j(\ell)\right|\right] \\
&= \epsilon^{-1}\sqrt{n}\left(\mathbb{E}_P\left[\max\left\{\sup_{\{\ell\in\mathcal{L}\}}\frac{1}{n}\sum_{i=1}^n f_{n,i}^j(\ell), \sup_{\{\ell\in\mathcal{L}\}}\frac{1}{n}\sum_{i=1}^n(-f_{n,i}^j(\ell))\right\}\right]\right) \\
&= \epsilon^{-1}\sqrt{n}\mathbb{E}_P\left[\sup_{\{f_i^j\in\mathcal{F}_n^+\cup\mathcal{F}_n^-\}}\frac{1}{n}\sum_{i=1}^n f_i^j\right],
\end{aligned} \tag{33}$$

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 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 $L_2(P)$ envelopes, a maximal inequality for VC-subgraph class of functions with bounded $L_2(P)$-envelope (Lemma A.5 in Kitagawa and Tetenov (2018)) applies and (33) 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 (32) are stochastically bounded $\mathcal{P}$-uniformly. Hence,

$$\sup_{\{\ell \in \mathcal{L}\}} \left| \sqrt{nh} \sum_{i=1}^n w_{n,i}^+ \mathcal{E}_{1,i}(\ell) - \frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i^+ \mathcal{E}_{1,i}(\ell) \right| = o_{\mathcal{P}}(1) \tag{34}$$

holds. □

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

Lemma 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 9 and assume that the kernel function and the bandwidth satisfy Assumption 10. 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 (22) 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 9, we can obtain a uniform upper bound of term (vi) as follows:

$$\left|\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)] - h_{2,P,1,+}(\ell_1,\ell_2)\right| \leq \frac{5M\bar{f}_R\int_0^\infty(\vartheta_2-\vartheta_1 u)^2 uK^2(u)du}{(f_R^+(r_0))^2(\vartheta_0\vartheta_2-\vartheta_1^2)^2}h, \tag{35}$$

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

$$\begin{aligned}
&\mathbb{E}_P\left[\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\hat{h}_{2,P,1,+}(\ell_1,\ell_2) - \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]\right|\right] \\
\leq &\frac{1}{[c_+ f_R^+(r_0)(\vartheta_0\vartheta_2-\vartheta_1^2)]^2}\mathbb{E}_P\left[\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^n f_{n,i}(\ell_1,\ell_2) - \mathbb{E}_P(f_{n,i}(\ell_1,\ell_2))\right|\right],
\end{aligned} \tag{36}$$

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

$$\mathbb{E}_P\left[\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^n f_{n,i}(\ell_1,\ell_2) - E_P(f_{n,i}(\ell_1,\ell_2))\right|\right] \leq C_1\left[c_+\bar{f}_R\int_0^\infty(\vartheta_2-\vartheta_1 u)^4 K^4(u)\,du\right]^{1/2}(nh)^{-1/2} \tag{37}$$

for all $nh \geq \frac{C_2(\vartheta_2+\vartheta_1)^4\bar{K}^4}{c_+\bar{f}_R\int_0^\infty(\vartheta_2-\vartheta_1 u)^4K^4(u)du}$, where C_1 and C_2 are positive constraints that do not depend on P and h. Combining (36), (37), and Markov's inequality, we conclude that the term (v) is $o_{\mathcal{P}}(1)$. □

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 4. *Suppose that Assumptions 9 and 10 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*

$$\sqrt{k_n h}(\hat{m}_{d,\star}(\cdot) - m_{P_{k_n},d,\star}(\cdot)) \Rightarrow \Phi_{h^*_{2,d,\star}}(\cdot), \tag{38}$$

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

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

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

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

By Lemma 2, (38) 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 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 (35), 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

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

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

$$\sum_{i=1}^{n} \mathbb{E}_{P_n}[F_{n,i}^2 \cdot 1\{F_{n,i} > \epsilon\}] \leq \sum_{i=1}^{n} \mathbb{E}_{P_n}\left[\frac{F_{n,i}^4}{\epsilon^2}\right] = \frac{1}{\epsilon^2 nh^2} E_{P_n}[(w_i^+)^4]$$

$$\leq (nh)^{-1} \frac{\bar{f}_R \int_0^\infty (\vartheta_2 - \vartheta_1 u)^4 K^4(u) du}{\epsilon^2 c_+^3 [f_R^+(r_0)(\vartheta_0\vartheta_2 - \vartheta_1^2)]^4} \to 0 \text{ as } n \to \infty,$$

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

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

where the second line follows by (35). Note that the convergence in the last line holds uniformly over $\ell_1, \ell_2 \in \mathcal{L}$ by Lemma 3, and this uniform convergence ensures condition (v) of Theorem 10.6 in Pollard (1990).

Hence, by 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 by the fact that $\hat{m}_{n,1,+}(\ell)$ and $\hat{m}_{n,1,-}(\ell)$ are estimated from separate samples, so 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 $\star = +, -$,

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

We denote the weak convergence conditional on a sample generated from a sample size dependent distribution of data P_n by $\overset{P_u}{\Rightarrow}$.[19] We denote the convergence in probability along the sequence $\{P_n\}$ by $\overset{P_n}{\to}$.

Lemma 5. *Suppose that Assumptions 9-11 hold, and let $\{P_n\}$ be a sequence of data generating processes in $\mathcal{P}$. For 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 $d = 0, 1$, $\widehat{\Phi}^u_{\nu_d,k_n} \overset{P_{k_n}}{\Rightarrow} \Phi_{h^*_{2,d}}$ with $h^*_{2,d} = h^*_{2,d,+} + h^*_{2,d,-}$ defined in (22).*

Proof. To simplify notation, we show this theorem for 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 the arguments for other cases are the same. We use the same arguments of 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

$$\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\hat{h}_{2,1,+}(\ell_1, \ell_2) - h^*_{2,1,+}(\ell_1, \ell_2)| \overset{P_n}{\to} 0, \tag{39}$$

and

$$nh \sum_{i=1}^n |w^+_{n,i}|^2 \overset{P_n}{\to} M_1, \tag{40}$$

$$n^3 h^3 \sum_{i=1}^n |w^+_{n,i}|^4 \overset{P_n}{\to} M_2, \tag{41}$$

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 between the two simulated processes, we have $\widehat{\Phi}^u_{\nu_1,n}(\ell) \overset{P_u}{\Rightarrow} \Phi_{h^*_{2,1}}(\ell)$.

[19] 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, $\widehat{\Phi}^u_n \overset{P_u}{\Rightarrow} \Phi$ 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.

Hence, the rest of the proof focuses on verifying (39) - (41). For positive integer $j < \infty$ and nonnegative integer $k < \infty$, a straightforward extension of Lemma 2 in Fan and Gijbels (1992) gives

$$(nh)^{(j-1)} \sum_{i=1}^{n} |w_{n,i}^{+}|^{j} \left(\frac{R_i - r_0}{h_+} \right)^{k} = \frac{\int_0^\infty K^j(u)(\vartheta_2 - \vartheta_1 u)^j u^k du}{c_+^{j-1}[\vartheta_0\vartheta_2 - \vartheta_1^2]^j} + o_{\mathcal{P}}(1), \tag{42}$$

where the first term in 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, (40) and (41) hold, as $\{P_n\} \in \mathcal{P}$.

To show (39), it suffices to show

$$\begin{aligned}
&\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\hat{h}_{2,1,+}(\ell_1,\ell_2) - h_{2,P,1,+}(\ell_1,\ell_2)| \\
\leq &\sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\hat{h}_{2,1,+}(\ell_1,\ell_2) - \mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]| + \sup_{\{\ell_1,\ell_2 \in \mathcal{L}\}} |\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)] - h_{2,P,1,+}(\ell_1,\ell_2)| \qquad (43) \\
= &o_{\mathcal{P}}(1).
\end{aligned}$$

The proof of Lemma 3 shows that the second term in (43) converges to zero uniformly in $\mathcal{P}$ in the proof of Lemma 3. We hence focus on showing that the first term in (43) is $o_{\mathcal{P}}(1)$.

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

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

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

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

By Lemma 4 and (42), term (i) is $o_{\mathcal{P}}(1)$ uniformly over $\ell_1, \ell_2 \in \mathcal{L}$. By Assumption 9 (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 (42) and $h_+ \to 0$. To examine term (vi), 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 4, equation (42), $nh \to \infty$, 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 for (34) in Lemma 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 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 with Lemma 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

$$\begin{aligned}
&\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\sum_{i=1}^n \hat{a}_{2,i}(\ell_1)\hat{a}_{3,i}(\ell_2)\right| \\
\leq & M(nh)^{-1/2}(h_+\vee h_+^2)\times\left\{\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}(w_{n,i}^+)^2\left(\frac{R_i-r_0}{h_+}\right)\mathcal{E}_{1,i}(\ell)\right|\right. \\
&\left.+\sup_{\{\ell\in\mathcal{L}\}}\left|(nh)^{3/2}\sum_{i=1}(w_{n,i}^+)^2\left(\frac{R_i-r_0}{h_+}\right)^2\mathcal{E}_{1,i}(\ell)\right|\right\}. \quad (44)
\end{aligned}$$

Similarly to the proof of (34) in Lemma 2, the two terms in the curly brackets of (44) can admit the following Bahadur representation: for positive integer $j<\infty$,

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

Similarly to the proof of Lemma 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 (44) 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. By modifying the proof of (34) by replacing $f_{n,i}^j(\ell)$ with $f_{n,i}(\ell_1,\ell_2)$ defined in the proof of Lemma 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

$$\begin{aligned}
&\mathbb{E}_P\left[\left|(nh)^{-1}\sum_{i=1}^n(w_i^+)^2\mathcal{E}_{1,i}(\ell_1)\mathcal{E}_{1,i}(\ell_2)-\mathbb{E}_P[\hat{h}_{2,P,1,+}(\ell_1,\ell_2)]\right|\right] \\
\leq&\frac{1}{[c_+f_R^+(r_0)(\vartheta_0\vartheta_2-\vartheta_1^2)]^2}\mathbb{E}_P\left[\sup_{\{\ell_1,\ell_2\in\mathcal{L}\}}\left|\frac{1}{n}\sum_{i=1}^n f_{n,i}(\ell_1,\ell_2)-\mathbb{E}_P(f_{n,i}(\ell_1,\ell_2))\right|\right], \quad (45)
\end{aligned}$$

where $f_{n,i}(\ell_1, \ell_2)$ is as defined in the proof of Lemma 3. Note that this upper bound coincides with (36). Hence, the proof of Lemma 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 6. *Suppose that Assumptions 9 and 10 hold. Let $\{P_n\}$ be a sequence of data generating processes in $\mathcal{P}$. For any subsequence of 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 notations defined in the proof of Lemma 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 (43) shown in the proof of Lemma 5 implies

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

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

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

Proof of Theorem 2: Having shown Lemmas 4 to 6, we apply the same arguments as in Hsu (2017). Let $\mathcal{H}_1$ denote the set of all functions from $\mathcal{L}$ to $[-\infty, 0]$. Let $h = (h_1, h_2)$, $h_1 = (h_{1,0}, h_{1,1})$ and $h_2 = (h_{2,0}, h_{2,1})$, where $h_{1,d} \in \mathcal{H}_1$ and $h_{2,d} = \mathcal{H}_2$ for $d = 0,1$. Define

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

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

$$\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ d\in\{0,1\},h_{2,P,d,+},h_{2,P,d,-}\in\mathcal{H}_{2,cpt}\}} P\left(\widehat{S}_n > c_0(h^P_{1,n}, h_{2,P}, 1-\alpha) + \xi\right) \leq \alpha, \quad (46)$$

where $h^P_{1,n} = (h^P_{0,d,n}, h^P_{1,d,n})$ such that for $d = 0, 1$, $h^P_{1,d,n} = \sqrt{n}h\nu_{P,d}(\cdot)$ and $h^P_{1,d,n}$ 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 $\alpha < 1/2$

$$\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ d\in\{0,1\}, h_{2,P,d,+}, h_{2,P,d,-}\in\mathcal{H}_{2,cpt}\}} P\Big(c_0(\psi_n, h_{2,P}, 1-\alpha) < c_0(h^P_{1,n}, h_{2,P}, 1-\alpha)\Big) = 0, \tag{47}$$

where $\psi_n = (\psi_{n,0}, \psi_{n,1})$ with $\psi_n(\ell) = (\psi_{n,0}(\ell), \psi_{n,1}(\ell))$. As a result, to complete the proof of Theorem 2, it suffices to show that for all $0 < \xi < \eta$

$$\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ d\in\{0,1\}, h_{2,P,d,+}, h_{2,P,d,-}\in\mathcal{H}_{2,cpt}\}} P\Big(\hat{c}_\eta(\alpha) < c_0(\psi_n, h_{2,P}, 1-\alpha) + \xi\Big) = 0. \tag{48}$$

Let $\{P_n \in \mathcal{P}_0 | n \geq 1\}$ be a sequence for which the probability in the statement of (48) evaluated at P_n differs from its supremum over $P \in \mathcal{P}_0$ by δ_n or less, where $\delta_n > 0$ and $\lim_{n\to\infty}\delta_n = 0$. By the definition of lim sup, such sequence always exists. Therefore, it is equivalent to show that for $0 < \xi < \eta$,

$$\lim_{n\to\infty} P\Big(\hat{c}_{n,\eta}(\alpha) < c_0(\psi_n, h_{2,P}, 1-\alpha) + \xi\Big) = 0, \tag{49}$$

where $\hat{c}_{n,\eta}(\alpha)$ denotes the critical value under P_n. To be more specific, it is true that the limit on the left hand side exists, but we want to show that it is 0. Given that we restrict 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 $h^*_{2,d,+}, h^*_{2,d,-} \in \mathcal{H}_{2,cpt}$. Also, $h^*_{2,d} = h^*_{2,d,+} + h^*_{2,d,-}$ for $d = 0, 1$. By Lemmas 5 and 6,

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

for $d = 0, 1$. By the definition of $\overset{P_{k_n}}{\Rightarrow}$ and $\overset{P_{k_n}}{\to}$, there exists a further subsequence m_n of k_n such that

$$\widehat{\Phi}^u_{\nu_d,m_n}(\cdot) \overset{a.s.}{\Rightarrow} \Phi_{h^*_{2,d}}(\cdot),$$
$$\sup_{\{\ell\in\mathcal{L}\}} |\hat{\sigma}^{-1}_{d,\xi}(\ell) - \sigma^{-1}_{d,P_{m_n},\xi}(\ell)| \overset{a.s.}{\to} 0,$$

for $d=0,1$. For any $\omega \in \Omega_1$ where

$$\Omega_1 \equiv \big\{\omega \in \Omega | \text{ for } d=0,1, \widehat{\Phi}^u_{\nu_d,m_n}(\cdot) \Rightarrow \Phi_{h^*_{2,d}}(\cdot), \sup_{\{\ell\in\mathcal{L}\}} |\hat{\sigma}^{-1}_{d,\xi}(\ell) - \sigma^{-1}_{d,P_{m_n},\xi}(\ell)| \to 0\big\},$$

by the same argument for Theorem 1 of Andrews and Shi (2013), we can show that for any constant $a_{m_n} \in \mathbb{R}$ which may depend on h_1 and P and for any $\xi_1 >$,

$$\limsup_{n\to\infty} \sup_{\{h_{1,0},h_{1,1}\in\mathcal{H}_1\}} P_u\Big(\sup_{\{d\in\{0,1\},\ell\in\mathcal{L}\}} \big(\frac{\widehat{\Phi}^u_{\nu_d,m_n}(\ell)(\omega)+h_{1,d}}{\hat{\sigma}_{d,\xi}(\ell)}\big) \le a_{m_n}\Big) - P\Big(\sup_{\{d\in\{0,1\},\ell\in\mathcal{L}\}} \big(\frac{\Phi_{h^*_{2,d}}(\ell)+h_{1,d}}{\sigma_{d,P_{m_n},\xi}(\ell)}\big) \le a_{m_n}+\xi_1\Big) \le 0. \tag{50}$$

(50) is similar to (12.28) in Andrews and Shi (2013). By (50) and by the similar argument for Lemma A5 of Andrews and Shi (2013), we have that for all $0<\xi<\xi_1<\eta$,

$$\liminf_{n\to\infty} \hat{c}_{m_n,\eta}(\alpha)(\omega) \ge c_0(\psi_{m_n}, h_{2,P_{m_n}}, 1-\alpha)+\xi_1. \tag{51}$$

Therefore, for any $\omega \in \Omega_1$, (51) holds. Given that $P(\Omega_1)=1$, we have that for all $0<\xi<\xi_1<\eta$

$$P\Big(\big\{\omega | \liminf_{n\to\infty} \hat{c}_{m_n,\eta}(\alpha)(\omega) \ge c_0(\psi_{m_n}, h_{2,P_{m_n}}, 1-\alpha)+\xi_1\big\}\Big) = 1,$$

which implies that

$$\lim_{n\to\infty} P(\hat{c}_{m_n,\eta}(\alpha) < c_0(\psi_{m_n}, h_{2,P_{m_n}}, 1-\alpha)+\delta) = 0. \tag{52}$$

Note that for any convergent sequence a_n, if there exists a subsequence a_{m_n} converging to a, then a_n converges to a as well. Therefore, (52) is sufficient for (49). Theorem 2(a) is shown by combining (46), (47) and (48).

We next show Theorem 2(b). Under Assumption 13, consider pointwise asymptotics under $P_c \in \mathcal{P}_0$. As in the proof of Proposition 1 of Barrett and Donald (2003) and Lemma 1 of Donald and Hsu (2016), we have $\widehat{S}_n \xrightarrow{d} \sup_{\{(d,\ell):\, \ell\in\mathcal{L}^o_{P_c,d}\}} \Phi_{h_{2,P_c,d}}(\ell)/\sigma_{d,P_c,\xi}(\ell)$ 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 $a \in (0,\infty)$ and $H(0) > 1/2$.

By the same proof for Theorem 2(b) of Andrews and Shi (2013), it is true that $\hat{c}_\eta(\alpha) \to c(1-\alpha+\eta)+\eta$ where $c(1-\alpha+\eta)$ 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 $c(1-\alpha)$, we have $\lim_{\eta\to 0} c(1-\alpha+\eta)+\eta = c(1-\alpha)$. This suffices to show that $\lim_{n\to\infty} P(\widehat{S}_n > \hat{c}_\eta(\alpha)) = \alpha$ under P_c. Combined with the claim of (a) in the current theorem, Theorem 2(b) holds. □

Proof of Theorem 3: 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 4: 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 $\hat{c}_\eta(\alpha) \overset{P_\eta}{\to} c_\eta + \eta$ 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 $\sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} |\Phi_{h^*_{2,d}}(\ell)| \overset{P_\eta}{\to} 0$ and $\widehat{S}_n \overset{P_\eta}{\to} \sup_{\{(d,\ell):\ell\in\mathcal{L}^o_{P_c,d}\}} \delta_d(\ell)/\sigma^*_{d,\xi}(\ell) > 0$. 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 non-negative 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 4. □

APPENDIX D. ADDITIONAL EMPIRICAL RESULTS FOR SECTION 5

TABLE 6. Testing Results for Israeli School Data: p-values, $\xi = 0.0316$

	3	5	AL	IK	CCT
g4math					
Cut-off 40	0.986	0.934	0.764	0.978	0.968
Cut-off 80	0.909	0.865	0.715	0.944	0.888
Cut-off 120	0.443	0.702	0.665	0.604	0.568
g4verb					
Cut-off 40	0.928	0.627	0.465	0.641	0.529
Cut-off 80	0.911	0.883	0.185	0.906	0.720
Cut-off 120	0.935	0.683	0.474	0.730	0.186
g5math					
Cut-off 40	0.876	0.282	0.482	0.631	0.609
Cut-off 80	0.516	0.446	0.930	0.482	0.765
Cut-off 120	0.939	0.827	0.626	0.883	0.838
g5verb					
Cut-off 40	0.594	0.893	0.953	0.900	0.938
Cut-off 80	0.510	0.692	0.504	0.519	0.929
Cut-off 120	0.696	0.811	0.601	0.699	0.774

TABLE 7. Testing Results for Israeli School Data: p-values, $\xi = 0.1706$

	3	5	AL	IK	CCT
g4math					
Cut-off 40	0.986	0.934	0.945	0.978	0.959
Cut-off 80	0.909	0.865	0.713	0.944	0.878
Cut-off 120	0.443	0.702	0.660	0.565	0.540
g4verb					
Cut-off 40	0.924	0.627	0.451	0.637	0.517
Cut-off 80	0.911	0.883	0.185	0.906	0.688
Cut-off 120	0.935	0.683	0.471	0.730	0.183
g5math					
Cut-off 40	0.861	0.275	0.481	0.623	0.600
Cut-off 80	0.516	0.429	0.916	0.479	0.762
Cut-off 120	0.939	0.827	0.624	0.883	0.836
g5verb					
Cut-off 40	0.594	0.893	0.953	0.934	0.938
Cut-off 80	0.510	0.671	0.496	0.513	0.946
Cut-off 120	0.696	0.811	0.594	0.699	0.757

TABLE 8. Testing Results for Israeli School Data: p-values, $\xi = 0.5$

	3	5	AL	IK	CCT
g4math					
Cut-off 40	0.984	0.934	0.940	0.978	0.950
Cut-off 80	0.907	0.853	0.832	0.936	0.893
Cut-off 120	0.443	0.683	0.633	0.557	0.519
g4verb					
Cut-off 40	0.907	0.599	0.450	0.637	0.499
Cut-off 80	0.907	0.880	0.165	0.906	0.760
Cut-off 120	0.935	0.668	0.449	0.719	0.164
g5math					
Cut-off 40	0.854	0.678	0.461	0.788	0.829
Cut-off 80	0.499	0.419	0.913	0.466	0.749
Cut-off 120	0.931	0.812	0.591	0.873	0.818
g5verb					
Cut-off 40	0.955	0.875	0.946	0.926	0.936
Cut-off 80	0.499	0.664	0.930	0.504	0.938
Cut-off 120	0.665	0.795	0.708	0.688	0.750

TABLE 9. Testing Results for Colombia's SR Data: p-values ($\xi = 0.00999$, full table)

	MPV Bandwidths			Other Bandwidth Choices		
Outcome variables	2	3	4	AI	IK	CCT
Risk protection, consumption smoothing and portfolio choice						
Individual inpatient medical spending	0.534	0.789	0.853	0.563	0.948	0.957
Individual outpatient medical spending	0.952	0.886	0.854	0.006	0.611	0.878
Variability of individual inpatient medical spending	0.502	0.792	0.870	0.655	0.937	0.959
Variability of individual outpatient medical spending	0.912	0.949	0.985	0.831	0.682	0.966
Individual education spending	0.150	0.191	0.174	0.016	0.896	0.093
Household education spending	0.001	0.000	0.000	0.000	0.007	0.000
Total spending on food	0.000	0.000	0.000	0.000	0.013	0.000
Total monthly expenditure	0.000	0.000	0.000	0.000	0.000	0.000
Has car	0.973	0.758	0.865	0.996	0.724	0.991
Has radio	1.000	1.000	1.000	1.000	1.000	1.000
Medical care use						
Preventive physician visit	0.615	0.990	1.000	1.000	0.368	0.978
Number of growth development checks last year	0.726	0.932	0.959	0.924	0.653	0.991
Curative care use	0.980	0.965	0.964	0.984	0.972	0.957
Primary care	0.919	0.920	0.946	0.994	0.966	0.952
Medical visit-specialist	0.979	0.936	0.734	0.927	0.897	0.634
Hospitalization	0.994	1.000	1.000	1.000	0.979	1.000
Medical visit for chronic disease	0.149	0.493	0.724	0.529	0.091	0.640
Curative care use among children	0.988	0.972	0.953	0.985	0.976	0.949
Health status						
Child days lost to illness	0.602	0.678	0.800	0.768	0.659	0.859
Cough, fever, diarrhea	0.989	1.000	1.000	1.000	0.991	1.000
Any health problem	0.996	0.999	0.983	0.998	0.998	0.990
Birthweight (KG)	0.901	0.999	1.000	0.995	0.904	0.999
Behavioral distortions						
Drank alcohol during pregnancy	0.425	0.743	0.870	0.930	0.190	0.937
Number of drinks per week during pregnancy	0.783	0.882	0.911	0.852	0.750	0.842
Months child breastfed	0.944	0.959	0.923	0.864	0.949	0.879
Folic acid during pregnancy	0.999	1.000	0.999	1.000	0.999	0.997
Number months folic acid during pregnancy	0.927	0.956	0.927	0.969	0.944	0.750
Contributory regime enrollment (ECV)	0.553	0.546	0.379	0.001	0.748	0.327
Contributory regime enrollment (DHS)	0.978	0.992	0.999	1.000	0.635	1.000
Other insurance (ECV)	0.890	0.932	0.928	0.876	0.818	0.904
Other insurance (DHS)	0.914	0.967	0.970	0.951	0.884	0.964
Uninsured (ECV)	0.675	0.688	0.450	0.079	0.751	0.682
Uninsured (DHS)	0.994	1.000	1.000	1.000	0.796	0.975

TABLE 10. 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
Risk protection, consumption smoothing and portfolio choice						
Individual inpatient medical spending	0.526	0.775	0.846	0.555	0.936	0.950
Individual outpatient medical spending	0.93	0.854	0.818	0.005	0.561	0.839
Variability of individual inpatient medical spending	0.497	0.773	0.860	0.645	0.927	0.951
Variability of individual outpatient medical spending	0.868	0.933	0.972	0.954	0.619	0.951
Individual education spending	0.148	0.183	0.172	0.016	0.893	0.093
Household education spending	0.001	0.000	0.000	0.000	0.007	0.000
Total spending on food	0.000	0.000	0.000	0.000	0.013	0.000
Total monthly expenditure	0.000	0.000	0.000	0.000	0.000	0.000
Has car	0.973	0.758	0.865	0.996	0.724	0.991
Has radio	1.000	1.000	1.000	1.000	1.0000	1.000
Medical care use						
Preventive physician visit	0.615	0.990	1.000	1.000	0.368	0.978
Number of growth development checks last year	0.716	0.932	0.959	0.924	0.641	0.991
Curative care use	0.980	0.965	0.964	0.984	0.972	0.957
Primary care	0.919	0.920	0.946	0.994	0.966	0.952
Medical visit-specialist	0.979	0.936	0.734	0.927	0.897	0.634
Hospitalization	0.994	1.000	1.000	1.000	0.979	1.000
Medical visit for chronic disease	0.149	0.493	0.724	0.529	0.091	0.64
Curative care use among children	0.988	0.972	0.953	0.985	0.976	0.949
Health status						
Child days lost to illness	0.602	0.678	0.800	0.768	0.659	0.859
Cough, fever, diarrhea	0.989	1.000	1.000	1.000	0.991	1.000
Any health problem	0.996	0.999	0.983	0.998	0.998	0.990
Birthweight (KG)	0.901	0.999	1.000	0.995	0.904	0.999
Behavioral distortions						
Drank alcohol during pregnancy	0.425	0.743	0.870	0.930	0.190	0.937
Number of drinks per week during pregnancy	0.783	0.882	0.904	0.852	0.719	0.840
Months child breastfed	0.944	0.959	0.923	0.864	0.949	0.879
Folic acid during pregnancy	0.999	1.000	0.999	1.000	0.999	0.997
Number months folic acid during pregnancy	0.927	0.956	0.927	0.969	0.944	0.750
Contributory regime enrollment (ECV)	0.553	0.546	0.379	0.001	0.748	0.327
Contributory regime enrollment (DHS)	0.978	0.992	0.999	1.000	0.635	1.000
Other insurance (ECV)	0.890	0.932	0.928	0.876	0.818	0.904
Other insurance (DHS)	0.914	0.967	0.970	0.951	0.884	0.964
Uninsured (ECV)	0.675	0.688	0.450	0.079	0.751	0.682
Uninsured (DHS)	0.994	1.000	1.000	1.000	0.796	0.975

TABLE 11. 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
Risk protection, consumption smoothing and portfolio choice						
Individual inpatient medical spending	0.427	0.684	0.764	0.990	0.875	0.900
Individual outpatient medical spending	0.846	0.840	0.782	0.637	0.411	0.799
Variability of individual inpatient medical spending	0.373	0.636	0.786	0.966	0.834	0.913
Variability of individual outpatient medical spending	0.767	0.887	0.947	0.927	0.347	0.913
Individual education spending	0.137	0.162	0.156	0.077	0.870	0.267
Household education spending	0.001	0.000	0.002	0.000	0.005	0.000
Total spending on food	0.000	0.000	0.000	0.000	0.013	0.000
Total monthly expenditure	0.000	0.000	0.000	0.000	0.000	0.000
Has car	0.973	0.758	0.865	0.996	0.724	0.991
Has radio	1.000	1.000	1.000	1.000	1.000	1.000
Medical care use						
Preventive physician visit	0.615	0.990	1.000	1.000	0.368	0.978
Number of growth development checks last year	0.767	0.888	0.953	0.991	0.823	0.986
Curative care use	0.980	0.965	0.964	0.984	0.972	0.957
Primary care	0.919	0.920	0.946	0.994	0.966	0.952
Medical visit-specialist	0.979	0.936	0.734	0.927	0.897	0.634
Hospitalization	0.994	1.000	1.000	1.000	0.979	1.000
Medical visit for chronic disease	0.149	0.493	0.724	0.529	0.091	0.640
Curative care use among children	0.988	0.972	0.953	0.985	0.976	0.949
Health status						
Child days lost to illness	0.602	0.678	0.800	0.768	0.659	0.859
Cough, fever, diarrhea	0.989	1.000	1.000	1.000	0.991	1.000
Any health problem	0.996	0.999	0.983	0.998	0.998	0.990
Birthweight (KG)	0.901	0.999	1.000	0.995	0.904	0.999
Behavioral distortions						
Drank alcohol during pregnancy	0.425	0.743	0.870	0.930	0.190	0.937
Number of drinks per week during pregnancy	0.731	0.853	0.888	0.834	0.694	0.805
Months child breastfed	0.944	0.959	0.923	0.864	0.949	0.879
Folic acid during pregnancy	0.999	1.000	0.999	1.000	0.999	0.997
Number months folic acid during pregnancy	0.927	0.956	0.927	0.969	0.944	0.750
Contributory regime enrollment (ECV)	0.553	0.546	0.379	0.001	0.748	0.327
Contributory regime enrollment (DHS)	0.978	0.992	0.999	1.000	0.635	1.000
Other insurance (ECV)	0.890	0.932	0.928	0.876	0.818	0.904
Other insurance (DHS)	0.914	0.967	0.970	0.951	0.884	0.964
Uninsured (ECV)	0.675	0.688	0.450	0.079	0.751	0.682
Uninsured (DHS)	0.994	1.000	1.000	1.000	0.796	0.975

TABLE 12. 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
Risk protection, consumption smoothing and portfolio choice						
Individual inpatient medical spending	0.519	0.681	0.708	0.892	0.756	0.706
Individual outpatient medical spending	0.718	0.840	0.928	0.215	0.232	0.898
Variability of individual inpatient medical spending	0.366	0.612	0.691	0.898	0.631	0.765
Variability of individual outpatient medical spending	0.400	0.751	0.768	0.544	0.176	0.778
Individual education spending	0.103	0.108	0.109	0.288	0.777	0.194
Household education spending	0.001	0.000	0.002	0.025	0.021	0.000
Total spending on food	0.000	0.000	0.000	0.000	0.011	0.000
Total monthly expenditure	0.000	0.000	0.000	0.000	0.000	0.000
Has car	0.973	0.758	0.865	0.996	0.724	0.991
Has radio	1.000	1.000	1.000	1.000	1.000	1.000
Medical care use						
Preventive physician visit	0.615	0.990	1.000	1.000	0.368	0.978
Number of growth development checks last year	0.823	0.915	0.970	0.973	0.866	0.991
Curative care use	0.980	0.965	0.964	0.984	0.972	0.957
Primary care	0.919	0.920	0.946	0.994	0.963	0.952
Medical visit-specialist	0.959	0.924	0.725	0.913	0.870	0.634
Hospitalization	0.994	1.000	1.000	1.000	0.979	1.000
Medical visit for chronic disease	0.149	0.493	0.724	0.529	0.091	0.640
Curative care use among children	0.988	0.972	0.953	0.985	0.976	0.949
Health status						
Child days lost to illness	0.602	0.678	0.800	0.768	0.659	0.859
Cough, fever, diarrhea	0.989	1.000	1.000	1.000	0.991	1.000
Any health problem	0.996	0.999	0.983	0.998	0.998	0.990
Birthweight (KG)	0.901	0.999	1.000	0.995	0.904	0.999
Behavioral distortions						
Drank alcohol during pregnancy	0.425	0.743	0.870	0.930	0.190	0.937
Number of drinks per week during pregnancy	0.666	0.801	0.851	0.739	0.650	0.751
Months child breastfed	0.941	0.958	0.918	0.864	0.942	0.876
Folic acid during pregnancy	0.999	1.000	0.999	1.000	0.999	0.997
Number months folic acid during pregnancy	0.927	0.956	0.927	0.969	0.944	0.750
Contributory regime enrollment (ECV)	0.553	0.546	0.379	0.001	0.748	0.327
Contributory regime enrollment (DHS)	0.978	0.992	0.999	1.000	0.635	1.000
Other insurance (ECV)	0.869	0.913	0.916	0.867	0.802	0.891
Other insurance (DHS)	0.907	0.967	0.970	0.951	0.871	0.964
Uninsured (ECV)	0.675	0.688	0.450	0.079	0.751	0.682
Uninsured (DHS)	0.994	1.000	1.000	1.000	0.796	0.975

TABLE 13. 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 14. 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 15. 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 bandwidths specified for the left and right of the cutoff, respectively. The data driven bandwidths presented in this table (AI, IK, and CCT) are under-smoothed 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 running variable in $(r_0 - h_-, r_0)$ and $[r_0, r_0 + h_+)$, respectively.

TABLE 16. Sample Sides and Bandwidths for Columbia'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 bandwidths specified for the left and right of the cutoff, respectively. The data driven bandwidths presented in this table (AI, IK, and CCT) are under-smoothed 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 running variable in $(r_0 - h_-, r_0)$ and $[r_0, r_0 + h_+)$, respectively.

FIGURE 6. Estimated complier's outcome density: Household education spending

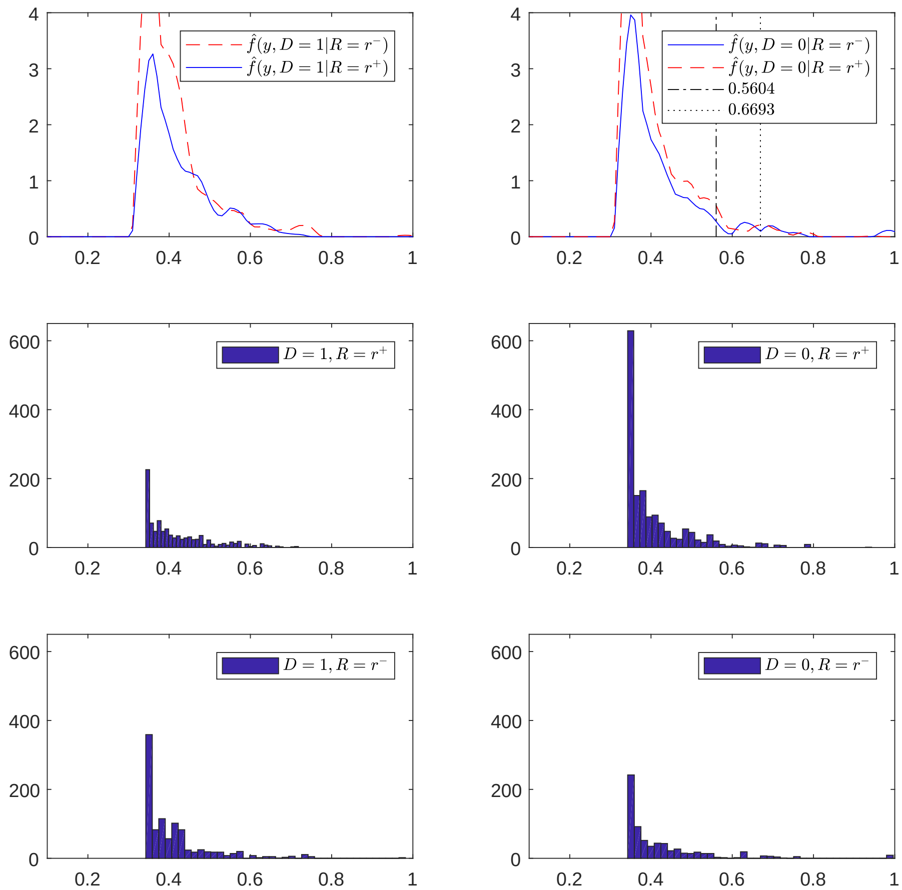


TABLE 17. Observations in the maximizer interval ($h^+ = h^- = 2$): Household edu. spending

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 7. Estimated complier's outcome density: Total monthly spending

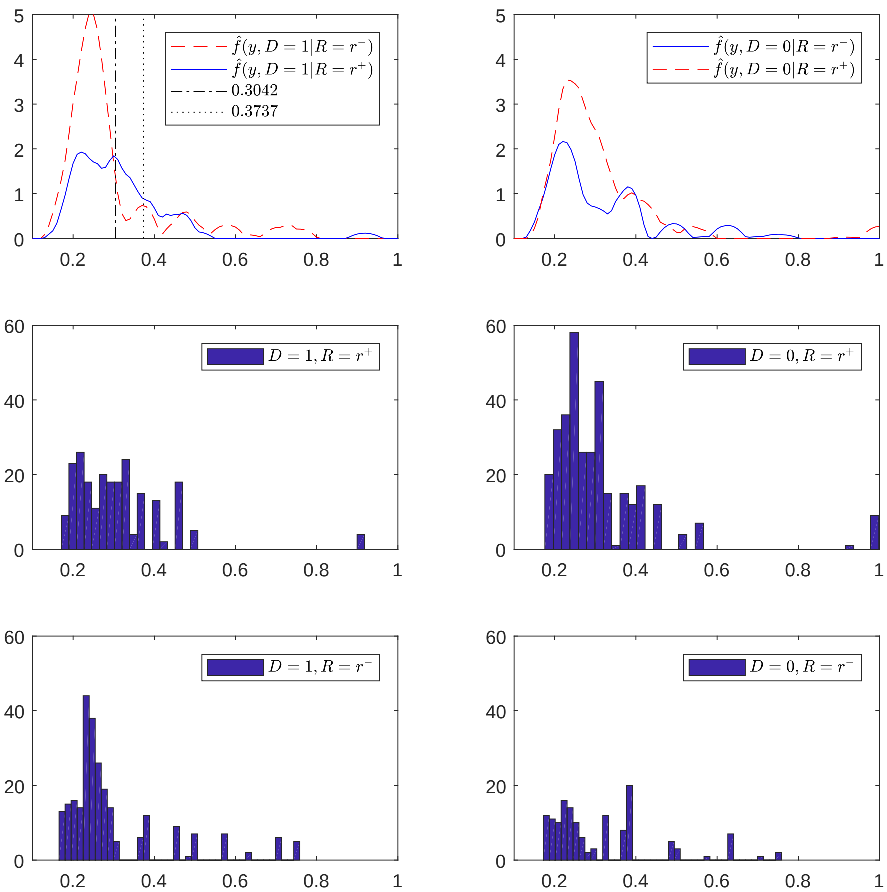


TABLE 18. Observations in the maximizer interval ($h^+ = h^- = 2$):Total monthly expenditure

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.

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

ANGRIST, J. D., AND V. LAVY (1999): "Using Maimonides' rule to estimate the effect of class size on scholastic achievement," *Quarterly Journal of Economics*, 114(2), 533–575.

ANGRIST, J. D., V. LAVY, J. LEDER-LUIS, AND A. SHANY (2017): "Maimonides rule redux," *NBER Working Paper, No. 23486*.

ARAI, Y., AND H. ICHIMURA (2016): "Optimal bandwidth selection for the fuzzy regression discontinuity estimator," *Economics Letters*, 141, 103–106.

ARAI, Y., AND T. KITAGAWA (2016): "A specification test in fuzzy regression discontinuity design," *unpublished manuscript*.

BALKE, A., AND J. PEARL (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92(439), 1171–1176.

BARRETT, G. F., AND S. G. DONALD (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

BERTANHA, M., AND G. W. IMBENS (2014): "External validity in fuzzy regression discontinuity designs," *NBER Working Paper, No. 20773*.

BIERENS, H. J. (1982): "Consistent model specification tests," *Journal of Econometrics*, 20(1), 105–134.

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 (2016): "Regression discontinuity designs using covariates," *Working Paper*.

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

CANAY, I. A., AND V. KAMAT (2018): "Approximate permutation tests and induced order statistics in the regression discontinuity design," *Review of Economic Studies, forthcoming*.

CARNEIRO, P., AND R. GINJA (2014): "Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start," *American Economic Journal: Economic Policy*, 6(4), 135–173.

CATTANEO, M. D., AND J. C. ESCANCIANO (eds.) (2017): *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing.

CATTANEO, M. D., M. JANSSON, AND X. MA (2016): "Simple local regression distribution estimators with an application to manipulation testing," *University of Michigan, Working Paper*.

CATTANEO, M. D., L. KEELE, R. TITIUNIK, AND G. VAZQUEZ-BARE (2016): "Interpreting regression discontinuity designs with multiple cutoffs," *Journal of Politics*, 78(4), 1229–1248.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81(2), 667–737.

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.

FRANDSEN, B. R., M. FRÖLICH, AND B. MELLY (2012): "Quantile treatment effects in the regression discontinuity design," *Journal of Econometrics*, 168(2), 382–395.

FRÖLICH, M., AND M. HUBER (2018): "Regression discontinuity design with covariates," *forthcoming in Journal of Business & Economic Statistics*.

GERARD, F., M. ROKKANEN, AND C. ROTHE (2018): "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable," *Working Paper*.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69(1), 201–209.

HANSEN, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64(2), 413–430.

HANSEN, P. R. (2005): "A test for superior predictive ability," *Journal of Business & Economic Statistics*, 23, 365–380.

HECKMAN, J. J., AND E. VYTLACIL (2005): "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 73(3), 669–738.

HOLM, S. (1979): "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, 6(2), 65–70.

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., I. MOURIFIÉ, AND Y. WAN (2016): "Testing identifying assumptions in regression discontinuity designs," *unpublished manuscript*.

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 J. D. ANGRIST (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62(2), 467–475.

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

IMBENS, G. W., AND T. LEMIEUX (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142(2), 615–635.

IMBENS, G. W., AND D. B. RUBIN (1997): "Estimating outcome distributions for compliers in instrumental variables models," *The Review of Economic Studies*, 64(4), 555–574.

KITAGAWA, T. (2015): "A test for instrument validity," *Econometrica*, 83(5), 2043–2063.

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

KOLESÁR, M., AND C. ROTHE (2018): "Inference in regression discontinuity designs with a discrete running variable," *American Economic Review*, 108(8), 2277–2304.

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

LEE, D. S., AND T. LEMIEUX (2010): "Regression discontinuity designs in economics," *Journal of Economic Literature*, 48(2), 281–355.

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

LINTON, O., K. SONG, AND Y.-J. WHANG (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154(2), 186–202.

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

MILLER, G., D. PINTO, AND M. VERA-HERNÁNDEZ (2013): "Risk protection, service use, and health outcomes under Colombia's health insurance program for the poor," *American Economic Journal: Applied Economics*, 5(4), 61–91.

MOURIFIÉ, I., AND Y. WAN (2017): "Testing local average treatment effect assumptions," *Review of Economics and Statistics*, 99(2), 305–313.

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.

RUBIN, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of Educational Psychology*, 66(5), 688.

THISTLETHWAITE, D. L., AND D. T. CAMPBELL (1960): "Regression-discontinuity analysis: An alternative to the ex post facto experiment.," *Journal of Educational psychology*, 51(6), 309.

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.