Testing Identification Conditions of LATE in Fuzzy Regression Discontinuity Designs

Yu-Chin Hsu* Ji-Liang Shiu† Yuanyuan Wan‡

November 13, 2022

Abstract

This paper derives testable implications of the identifying conditions that include the local monotonicity assumption and the continuity in means assumption for the local average treatment effect in fuzzy regression discontinuity (FRD) designs. Building upon the seminal work of Horowitz and Manski (1995), we show that the testable implications of these identifying conditions are a finite number of inequality restrictions on the observed data distribution. We then propose a specification test for the testable implications and show that the proposed test controls the size and is asymptotically consistent. We apply our test to the FRD designs used in Miller, Pinto, and Vera-Hernández (2013) for Columbia's insurance subsidy program, in Angrist and Lavy (1999) for Israel's class size effect, in Pop-Eleches and Urquiola (2013) for Romanian school effect, and in Battistin, Brugiavini, Rettore, and Weber (2009) for the retirement effect on consumption.

Keywords: Fuzzy regression discontinuity design; Moment inequalities; Local continuity in means; Weighted bootstrap

JEL classification: C12, C14, C15

*Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University; CRETA, National Taiwan University. E-mail: ychsu@econ.sinica.edu.tw. 128 Academia Rd., Section 2, Nankang, Taipei, 115, Taiwan.

†Institute for Economic and Social Research, Jinan University, Email: jishiu.econ@gmail.com

‡Department of Economics, University of Toronto. E-mail: yuanyuan.wan@utoronto.ca. 232, Max Gluskin House, 150 St. George St., Toronto, Ontario, M5S 3G7, Canada.

1 Introduction

This paper aims to derive the testable implications of the identification assumptions for the local average treatment effect in fuzzy regression discontinuity (FRD) designs and develops a specification test for such testable implications. Since the seminal work of Thistlethwaite and Campbell (1960), the regression discontinuity (RD) design has gained popularity in applied research to identify causal effects (see Lee and Lemieux, 2010; Cattaneo and Escanciano, 2017, for surveys). In a sharp RD design, the treatment assignment is deterministically determined by whether a running variable exceeds a known cutoff. On the other hand, the probability of receiving the treatment changes discontinuously at the cutoff in an FRD design but not necessarily from 0 to 1. In both designs, if units of the study located just above or below the cutoff are "comparable", then the RD design creates a "pseudo-random experiment" near the cutoff and thus enables us to identify the causal effect of the treatment.

The identification idea is formalized by Hahn, Todd, and Van der Klaauw (2001) in a potential outcome framework, where they provide conditions to identify the average treatment effect (ATE) and the local average treatment effect (LATE) at the cutoff, respectively. These conditions are revisited later by Lee (2008), Imbens and Lemieux (2008), Frandsen, Frölich, and Melly (2012), Dong (2018), Bertanha and Moreira (2020), and Frandsen, Frölich, and Melly (2012), among many others.

While the identification problem has been well studied, the credibility of identification assumptions can be controversial in practice, which has motivated many specification tests in the RD framework. There are two strands of tests. The first strand focuses on testing the identifying assumptions for ATE-type parameters, among which Lee (2008) proposes testable implications for sharp designs (i) the continuity of the density of a running variable at the cutoff, and (ii) the continuity of the conditional distributions of predetermined variables given the running variable at the cutoff. The testable implications in Lee (2008) are the foundation for many tests and can be generalized to fuzzy designs; see, for example, McCrary (2008), Otsu, Xu, and Matsushita (2013), Cattaneo, Jansson, and Ma (2020), and Bugni and Canay (2021) for testing the continuity of the running variable density, and Canay and Kamat (2018) for testing the continuity of the conditional distributions of predetermined variables given the running variable. A common feature of these tests is that they utilize running variables (and other baseline variables) but not the outcome or treatment variables.

The second strand focuses on testing the identifying assumptions of LATE-type parameters in fuzzy regression discontinuity (FRD) designs. Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) show that if the parameter of interest is the LATE or the local quantile treatment effects, then the continuity of the running variable density and the continuity of the predetermined variable distributions are neither sufficient nor necessary; see discussions in McCrary (2008) too. They test sharp implications of identifying assumptions that are similar to those used in Frandsen, Frölich, and Melly (2012), including (i) the monotonicity of the treatment response to the running variable at the cutoff (local monotonicity assumption), and (ii) the continuity of the conditional distributions of the potential outcomes and complying status given the running variable at the cutoff (local continuity in distributions assumption). These conditions, which we refer to as the "FRD distributional assumptions" hereafter, are used to identify quantile or distributional treatment effects for compliers.

Our paper contributes to the second strand by proposing a specification test for identifying conditions for LATE. The identification assumptions required for LATE replace the local continuity in distributions assumption with local continuity in means assumption, i.e. the expectations of potential outcomes given that the running variable is continuous near the cutoff. The local monotonicity assumption and the local continuity in means assumption together are referred as the "FRD mean assumptions". We consider testing the FRD mean assumptions useful for the following reasons. First, while it is true that the FRD distributional assumptions imply the FRD mean assumptions, we will show in Section 2 that an arbitrary proper subset of testable implications for the FRD distributional assumptions is neither sufficient nor necessary for the FRD mean assumptions. Therefore, if we only test a pre-chosen subset of testable implications of the FRD distributional assumptions, we may falsely reject the FRD mean assumptions with a higher probability than the pre-specified level in practice when the FRD mean assumptions actually hold in the data. Second, in many empirical applications, the mean effect is of primary interest, and its identification requires only the weaker FRD mean assumptions instead of the stronger FRD distributional assumptions. Therefore, as we will illustrate in an empirical example in Section 5, it is possible that a test on FRD distributional assumptions rejects, but our test on FRD mean assumptions accepts. In such cases, an easy-to-implement specification test focusing directly on the FRD mean assumptions would be helpful. Our paper fills this gap.

Given the FRD mean assumptions, we derive sharp (observable) bounds for the expectation of the potential outcome $Y(1)$ for always-takers when the running variable approaches the cutoff from below by applying the results in Horowitz and Manski (1995) and Lee (2009). Therefore, its identifiable estimand must lie within the bounds too. This creates two inequality constraints on the observed data distribution. We can obtain another two constraints by applying a similar argument for the expectation of the potential outcome $Y(0)$ for never-takers when the running variable approaches the cutoff from above. We also show that if the observable data distribution satisfies these four constraints, there exists a joint distribution of potential outcomes and complying status that satisfies the FRD mean assumptions and is observationally equivalent to the observable data distribution.

The proposed specification test is based on these inequality constraints. Our test statistic is significantly different from the test in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) because the inequality constraints in our case involve nuisance parameters that have to be estimated first. We need to account for the estimation effect when deriving the null distribution of the test statistic. The critical value is constructed based on a weighted bootstrap and the generalized moment selection (GMS) procedure that we use to approximate the null distribution. We show that our test controls the size well under the null and is consistent against any fixed alternative.

Our paper also makes empirical contributions. We apply our test to four FRD designs in the literature. The first is in Miller, Pinto, and Vera-Hernández (2013), who estimate the mean effect of a publicly subsidized insurance program on Columbian households' welfare, measured by various outcome variables. Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) find that the joint assumptions of local monotonicity and local continuity in distributions are rejected for three outcome variables: household educational spending, total spending on food, and total monthly expenditure. Since the monotonicity assumption is likely to be satisfied by the institutional rules, the testing result suggests that the distribution of these variables is discontinuous near the cutoff. We revisit this empirical application and find that our mean test does not reject the implication of local continuity in means. Our result, together with the findings in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), suggest that the mean causal effect estimates in Miller, Pinto, and Vera-Hernández (2013) are credible. Still, one needs to be cautious when estimating the quantile LATE for these outcome variables. In our second empirical application, we consider Israel's schooling data used in Angrist and Lavy (1999) to study the effect of class size on students' performance. In the application, Israel's Maimonides' rule creates a discontinuity of class size with respect to enrollment. Our result re-confirms those in Angrist, Lavy, Leder-Luis, and Shany (2019) and Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), and shows no evidence of failing FRD identification assumptions. In our third application, we revisit Romanian secondary school data, which Pop-Eleches and Urquiola (2013) use to identify the effect of school quality on students' academic performance. The probability of enrollment into better schools changes discontinuously in transition scores because the centralized allocation process first meets the need for "better students". We find no evidence to reject the FRD mean assumptions. The final application uses data from Battistin, Brugiavini, Rettore, and Weber (2009), who study the effect of retirement on Italian seniors' consumption using the pension eligibility policy as the identification device. In their data set, the retirement probability changes discontinuously at the eligibility cutoff for a pension because it provides an additional incentive to retire. Again, our test result re-confirms the validity of FRD design in this empirical study.

In addition to the RD literature, our paper also contributes to the growing literature on specification tests in causal inference frameworks. For example, Kitagawa (2015) and Mourifié and Wan (2017) test the statistical independence assumption and the monotonicity assumption in the framework of a binary instrument and a binary treatment; Sun (2020) considers models with a discrete but multi-valued instrument and treatment. Huber and Mellace (2015) also consider the binary IV and binary treatment framework and propose a specification test for the mean-independence and monotonicity assumptions. Kédagni and Mourifié (2020) derive a set of generalized inequalities from Pearl (1995) to test the IV-independence assumption with discrete treatment with unrestricted outcomes and instruments. Acerenza, Bartalotti, and Kédagni (2020) test identifying assumptions in bivariate Probit models. Among these works, our paper is more relevant to and gains motivation from Huber and Mellace (2015), who also construct testable implications based on the bounds derived in Horowitz and Manski (1995). While our testable implications share the same spirit as those in Huber and Mellace (2015), the two papers focus on different models. Furthermore, our testable implications are local to the cutoff instead of global restrictions due to the feature of regression discontinuity. For this reason, our test also differs significantly from that proposed in Huber and Mellace (2015) because smoothing for the dimension of the running variable is required in our case.

The rest of the paper is organized as follows. We discuss the identifying assumptions and derive the testable implications in Section 2. In Section 3, we describe the testing procedure and establish the asymptotic proprieties of our test. In Section 4, we conduct several sets of Monte Carlo experiments to show the finite sample performance of our test and report empirical application results in Section 5. Section 6 discusses possible extensions of our test. We conclude the paper in Section 7. All the proofs, additional simulations and empirical results are collected in the appendix.

2 Assumptions and Testable Implications

Let $(\Omega, \mathcal{F}, P)$ be the probability space, where Ω is the sample space with a generic element denoted by ω, $\mathcal{F}$ is the sigma-algebra, and the P is the probability distribution that generates all the random variables. Among all the variables, $D(\cdot) : \Omega \to \{0, 1\}$ is the observed binary treatment assignment, $Y(\cdot) : \Omega \to \mathcal{Y}$ is the observed outcome of interest, and $Z(\cdot) : \Omega \to \mathcal{Z}$ is a continuous running variable with a known cut-off c. A given individual ω in the population is endowed with a potential treatment function $D(\cdot, \omega) : \mathcal{Z} \to \{0, 1\}$. $D(z, \omega)$ represents the treatment that the individual ω would have taken had his/her running variable been externally set to z. Likewise, let $Y(d, \omega)$ be his/her potential outcome had the treatment been externally set to d. The observed treatment and outcome are connected as $D(\omega) \equiv D(Z, \omega)$ and $Y(\omega) \equiv D(\omega)Y(1, \omega) + (1 - D(\omega))Y(0, \omega)$, respectively.

Based on the shape of the potential treatment function in a small neighbourhood $B_\epsilon = \{z \in \mathcal{Z} : |z - c| \leq \epsilon\}$ of the cutoff, we define compliance status T of an individual ω as:

$$T(\omega) = \begin{cases} \mathbf{A}, & \text{if } D(z,\omega) = 1, \text{ for } z \in B_\epsilon, \\ \mathbf{N}, & \text{if } D(z,\omega) = 0, \text{ for } z \in B_\epsilon, \\ \mathbf{C}, & \text{if } D(z,\omega) = 1\{z \geq c\}, \text{ for } z \in B_\epsilon, \\ \mathbf{DF}, & \text{otherwise} \end{cases} \tag{2.1}$$

where **A, C, N** and **DF** represent "always-takers", "compliers", "never-takers", and "defiers", respectively. Hereafter, we will suppress the argument ω whenever it causes no confusion. We make the following assumptions as in Imbens and Lemieux (2008).

Assumption 2.1 (Local monotonicity) *There exists a small $\epsilon > 0$ such that $T \in \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{N}\}$ almost surely.*

Assumption 2.1 requires that the potential treatment status be weakly increasing in the running variable near the cutoff for all individuals in the population.[1] It rules out defiers.

For $d \in \{0,1\}$ and $t \in \{\mathbf{A}, \mathbf{N}, \mathbf{C}\}$, let $f_{Y(d)|T,Z}(y|t,z)$ be the probability density function of $Y(d)$ given type $T = t$ and $Z = z$ (when $Y(d)$ is discrete, these densities are understood as probability mass function). We will focus on the cases in which $\lim_{z \downarrow c} f_{Y(d)|T,Z}(y|t,z)$ and $\lim_{z \uparrow c} f_{Y(d)|T,Z}(y|t,z)$ are proper densities and the corresponding conditional expectations, denoted by $E[Y(d)|T = t, Z = c^+]$ and $E[Y(d)|T = t, Z = c^-]$, are finite.[2]

Assumption 2.2 (Local continuity in means) *$E[Y(d)|T = t, Z = z]$ and $P(T = t|Z = z)$ are continuous in z in the neighborhood B_ϵ of c for any $t \in \{\mathbf{A}, \mathbf{N}, \mathbf{C}\}$.*

Assumption 2.2 requires the continuity of the conditional mean of potential outcomes as a function of the running variable in the neighbourhood of the cutoff for each type of individual, as well as for the type probabilities. This assumption is weaker than the local continuity assumption in distributions, which is tested in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022). We restate this assumption below.

Assumption 2.3 (Local continuity in distributions) *For $d = 0, 1$, $t \in \{\mathbf{A}, \mathbf{C}, \mathbf{N}\}$, and an measurable subset $V \subseteq \mathcal{Y}$, we have*

$$\lim_{z \uparrow c} P(Y_d \in V, T = t|Z = z) = \lim_{z \downarrow c} P(Y_d \in V, T = t|Z = z).$$

The following proposition re-states the results of Hahn, Todd, and Van der Klaauw (2001), Frandsen, Frölich, and Melly (2012), and Arai, Hsu, Kitagawa, Mourifié, and Wan (2022). It shows that the LATE at the cutoff is identified under the monotonicity assumption and continuity in means assumption, and the distributional LATE is identified if the continuity in means assumption is strengthened to continuity in distributions assumption. For the purpose of exposition, the proof is omitted. For generic random variables (R_1, R_2, R_3), let $E[R_1|R_2, R_3 = c^+]$ be the limit of a conditional mean $\lim_{b \downarrow c} E[R_1|, R_2, R_3 = b]$ and $E[R_1|R_2, R_3 = c^-]$ be the

[1]This monotonicity assumption is equivalent to those imposed in Hahn, Todd, and Van der Klaauw (2001) and Imbens and Lemieux (2008), and is slightly stronger than that in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022). All the versions have the same intuition and are equivalent in the limit of ϵ approaches to zero. We use this version for its simplicity.

[2]It only requires the conditional density of potential outcomes to have a well-defined limit from above and below the cutoff, respectively, but not necessarily equal to each other.

limit of a conditional mean $\lim_{b\uparrow c} E[R_1|, R_2, R_3 = b]$, whenever these quantities are properly defined.

Proposition 2.1 *Suppose Assumptions 2.1 and 2.2 are satisfied, and* $E[D|Z = c^+] > E[D|Z = c^-]$*, then LATE at the cutoff is identified by the fuzzy regression discontinuity estimand:*

$$LATE \equiv E[Y(1) - Y(0)|\boldsymbol{C}, Z = c] = \frac{E[Y|Z = c^+] - E[Y|Z = c^-]}{E[D|Z = c^+] - E[D|Z = c^-]}. \tag{2.2}$$

If Assumption 2.3 holds in place of Assumption 2.2, then the complier's potential outcome distributions at the cutoff are identified by the following quantities:

$$F_{Y_1|\mathbf{C},Z=c}(y) = \frac{E[1\{Y \leq y\}D|Z = c^+] - E[1\{Y \leq y\}D|Z = c^-]}{E[D|Z = c^+] - E[D|Z = c^-]}, \tag{2.3}$$

$$F_{Y_0|\mathbf{C},Z=c}(y) = \frac{E[1\{Y \leq y\}(1 - D)|Z = c^+] - E[1\{Y \leq y\}(1 - D)|Z = c^-]}{E[D|Z = c^+] - E[D|Z = c^-]}. \tag{2.4}$$

Furthermore, the sharp testable implications for Assumptions 2.1 and 2.3 are characterized by the following set of inequality constraints:

$$E[g(Y)D|Z = c^-] - E[g(Y)D|Z = c^+] \leq 0 \tag{2.5}$$

$$E[g(Y)(1 - D)|Z = c^+] - E[g(Y)(1 - D)|Z = c^-] \leq 0, \tag{2.6}$$

for any g belonging to the class of close intervals: $\mathcal{G} = \{g : g(Y) = 1[y \leq Y \leq y'], y, y' \in \mathcal{Y}\}$.

The inequality constraints (2.5) and (2.6) can be interpreted as the "nonnegativity of the potential outcome density functions for the compliers at the cutoff". As shown in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), if inequalities (2.5) and (2.6) are satisfied for all $g \in \mathcal{G}$, then we can construct a joint distribution for the potential outcomes and the running variable which is observationally equivalent to the observed data distribution and satisfies the FRD distributional assumptions, hence the FRD mean assumptions. However, (2.5) and (2.6) are not necessary implications of FRD mean assumptions. That is, one can find a distribution that satisfies FRD mean assumptions, but its implied distribution would violate inequality (2.5) or (2.6) for a function $g \in \mathcal{G}$. Therefore, inequalities (2.5) and (2.6) are "over aggressive" for testing the FRD mean assumptions. A natural follow-up question is, does there exist a subclass $\mathcal{G}_V \subseteq \mathcal{G}$ that is dedicated to assessing the validity of FRD mean assumptions? Unfortunately, the answer is negative. For a given sub-class of closed intervals $\mathcal{G}_V$, there always exists a DGP that satisfies the FRD mean assumptions but violates one of the inequalities at $g^0 \in \mathcal{G}_V$. This is demonstrated in the Lemma 2.1 below.

Lemma 2.1 *Let $V = [y^*, y^{**}]$ be an arbitrary interval such that $P(Y_d \in V) > 0$ for $d = 0, 1$. Let $\mathcal{G}_V \subseteq \mathcal{G}$ be a class of intervals that generated from V: $\mathcal{G}_V = \{g : g(Y) = 1[y \leq Y \leq y'], y, y' \in V\}$. Let $\mathcal{P}_0$ be the set of joint distributions of $\{Y(0), Y(1), D(z), Z\}_{z\in\mathcal{Z}}$ that satisfy Assumptions 2.1 and 2.2 and one of the following two conditions,*

(a) the distribution of $Y(d)|T = t, Z = z$ admits a positive density $f_{Y(d)|T=t,Z=z}$ over V for all t and z in the small neighborhood of cutoff c, or

(b) if the distribution of $Y(d)$ is discrete, then V contains at least three elements $\{y_1, y_2, y_3\}$ such that $P(Y(d) = y_k|T = t, Z = z) > 0$ for all $k = 1, 2, 3$, all t, and all z in a small neighborhood of cutoff c,[3]

then for any pre-chosen V, there exists a distribution $P \in \mathcal{P}_0$ and a function $g^0 \in \mathcal{G}_{\mathcal{V}}$ such that the implied distribution of (Y, D, Z) violates inequality (2.5) or (2.6) at g^0.

Remark 2.1 *Another temptation to proceed is to change the function class, for example, to replace $g(Y)$ by Y and check whether the following inequalities hold,*

$$E[YD|Z = c^-] - E[YD|Z = c^+] \leq 0 \tag{2.7}$$

$$E[Y(1-D)|Z = c^+] - E[Y(1-D)|Z = c^-] \leq 0, \tag{2.8}$$

This is, however, not a valid approach either. Even though the FRD distributional assumptions are satisfied (and hence the mean assumptions), the inequalities (2.7) and (2.8) can fail to hold. For example, simple calculation shows that the left-hand side of inequality (2.7) is $E[Y(1)|\boldsymbol{A}, Z = c]\pi_{\boldsymbol{A}} - E[Y(1)|\boldsymbol{A}\cup\boldsymbol{C}, Z = c]\pi_{\boldsymbol{A}\cup\boldsymbol{C}}$, where π_t is the size of subpopulation t at the cutoff. Although $\pi_{\boldsymbol{A}} < \pi_{\boldsymbol{A}\cup\boldsymbol{C}}$, the sign of $E[Y(1)|\boldsymbol{A}, Z = c]\pi_{\boldsymbol{A}} - E[Y(1)|\boldsymbol{A}\cup\boldsymbol{C}, Z = c]\pi_{\boldsymbol{A}\cup\boldsymbol{C}}$ can be in either direction, depending on the relative magnitude of $E[Y(1)|\boldsymbol{A}]$ and $E[Y(1)|\boldsymbol{C}]$.

To state a proper set of testable implications for Assumptions 2.1 and 2.2, let us define some notation. Let $G_1(y) = \lim_{z\downarrow c} P(Y \leq y|D = 1, Z = z)$ be the conditional distribution of Y given

[3] In the discrete case we require that there be a least three support points because if there are only two, then testing the mean is equivalent to testing the distribution.

$D = 1$ and $Z = z$ when z converges to c from above. Note that $G_1(y) = \lim_{z \downarrow c} P(Y(1) \leq y | \boldsymbol{A} \cup \boldsymbol{C}, Z = z)$, and is well defined under Assumption 2.2-(i). Likewise, define $G_0(y) = \lim_{z \uparrow c} P(Y \leq y | D = 0, Z = z)$. We let $q = P_{1|0}/P_{1|1}$ be the relative size of always-takers with respect to the combination of always-takers and compliers, where $P_{1|0} = P(D = 1|Z = c^-)$ and $P_{1|1} = P(D = 1|Z = c^+)$.[4] Likewise, we define $r = P_{0|1}/P_{0|0}$, where $P_{0|1} = P(D = 0|Z = c^+)$ and $P_{0|0} = P(D = 0|Z = c^-)$. Note that G_1, G_0, q, and r are all directly identifiable from the data. Finally, for a generic cumulative distribution function $\tilde{F}$ and a $\tau \in (0, 1)$, let

$$\tilde{F}_L^{-1}(\tau) = \inf\{y \in \mathcal{Y} : \tilde{F}(y) \geq \tau\}, \quad \tilde{F}_U^{-1}(\tau) = \sup\{y \in \mathcal{Y} : \tilde{F}(y) \leq \tau\}. \tag{2.9}$$

When $\tilde{F}$ is continuous at its τ-th quantile, $\tilde{F}_L^{-1}(\tau) = \tilde{F}_U^{-1}(\tau)$, and they all coincide with the usual definition of the quantile functions.

Now we are ready to present the testable implications of the FRD mean assumptions. Applying the results in Horowitz and Manski (1995), we derive the bounds for $E[Y(1)|T = \boldsymbol{A}, Z = c^-]$ and $E[Y(0)|T = \boldsymbol{N}, Z = c^+]$, respectively. Their identifiable estimands must satisfy the bounds too, and form restrictions on the distribution of observed data. We summarize them in Proposition 2.2.

Proposition 2.2 *Suppose that Assumptions 2.1 and 2.2 are satisfied, $q \in (0, 1)$, and $r \in (0, 1)$, then the following inequality constraints hold:*

$$E[Y|D = 1, Y < G_{1L}^{-1}(q), Z = c^+] \leq E[Y|D = 1, Z = c^-], \tag{2.10}$$

$$E[Y|D = 1, Z = c^-] \leq E[Y|D = 1, Y > G_{1U}^{-1}(1 - q), Z = c^+], \tag{2.11}$$

$$E[Y|D = 0, Y < G_{0L}^{-1}(r), Z = c^-] \leq E[Y|D = 0, Z = c^+], \tag{2.12}$$

$$E[Y|D = 0, Z = c^+] \leq E[Y|D = 0, Y > G_{0U}^{-1}(1 - r), Z = c^-]. \tag{2.13}$$

When the distribution of Y given $(D = 1, Z = c^+)$ is continuous at its q-th and $(1 - q)$-th quantiles, inequalties (2.10) and (2.11) are sharp. When the distribution of Y given $(D = 0, Z = c^-)$ is continuous at its r-th and $(1 - r)$-th quantiles, inequalties (2.12) and (2.13) are sharp.

[4] If the local monotonicity condition holds with an "increasing" direction, then $P_{1|0}/P_{1|1} \geq P_{1|1}/P_{1|0}$ and thus $q = P_{1|0}/P_{1|1}$ measures the ratio of always-takers against the combination of always-takers and compliers. If it holds with a "decreasing" direction, then we define $q = P_{1|1}/P_{1|0}$ for this ratio. The same argument applies to the definition for r below.

Remark 2.2 *The bounds in Proposition 2.2 are not necessarily sharp when Y is discrete; however, we can provide the formula for the sharp bounds for such cases, but at the cost of more complicated notation. The results are given in Section 6.1. We only focus on (2.10) to (2.13) for the following reason. In some practices, the ways we treat a particular outcome variable as continuous or discrete are mixed. For instance, if Y is "education level by years", then some treat it as a continuous variable, but others might treat it as a discrete variable. Using the inequalities (2.10) to (2.13), we would have a unified expression regardless. And it also facilitates the development of the asymptotic theory of our test. If deciding to treat an outcome variable as discrete, one can always use the sharp bounds reported in Section 6.1.*

Remark 2.3 *From the inequalities in Proposition 2.2, we know that our test is expected to have better power when q and r are relatively large. This is because, for example, the lower bound in (2.10) is decreasing in q giving everything else equal. This is the case when the size of compliers is close to one, and the size of the propensity score jump at the cutoff is large. This feature is also shared in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) for testing the distributional assumptions, where they show that their testable implication always holds in sharp design ($q = r = 0$). For the mean test we study in this paper, when $q = 0$, the inequalities (2.10) and (2.11) reduce to $Y_{1,min} \leq Y_{1,max}$, where $Y_{1,min}$ and $Y_{1,max}$ are the lower and upper bounds of the conditional distribution of $Y|D = 1, Z = c^+$. This holds trivially.*

Note that

$$
\begin{aligned}
& E[Y|D = 1, Y < G_{1L}^{-1}(q), Z = c^+] - E[Y|D = 1, Z = c^-] \leq 0 \\
\Leftrightarrow\ & \frac{E[DY1(Y < G_{1L}^{-1}(q))|Z = c^+]}{E[D1(Y < G_{1L}^{-1}(q))|Z = c^+]} - \frac{E[DY|Z = c^-]}{E[D|Z = c^-]} \leq 0 \\
\Leftrightarrow\ & \theta_1 \equiv E[DY1(Y < G_{1L}^{-1}(q))|Z = c^+] \cdot E[D|Z = c^-] \\
& \qquad - E[DY|Z = c^-] \cdot E[D1(Y < G_{1L}^{-1}(q))|Z = c^+] \leq 0. \qquad (2.14)
\end{aligned}
$$

Similarly, the rest of the three inequalities (2.11)-(2.13) are equivalent to

$$\begin{aligned}
\theta_2 \equiv\ & E[DY|Z=c^-]\cdot E[D1(Y>G_{1U}^{-1}(1-q))|Z=c^+] \\
& -E[DY1(Y>G_{1U}^{-1}(1-q))|Z=c^+]\cdot E[D|Z=c^-]\le 0, && (2.15)\\
\theta_3 \equiv\ & E[(1-D)Y1(Y<G_{0L}^{-1}(r))|Z=c^-]\cdot E[1-D|Z=c^+] \\
& -E[(1-D)Y|Z=c^+]\cdot E[(1-D)1(Y<G_{0L}^{-1}(r))|Z=c^-]\le 0, && (2.16)\\
\theta_4 \equiv\ & E[(1-D)Y|Z=c^+]\cdot E[(1-D)1(Y>G_{0U}^{-1}(1-r))|Z=c^-] \\
& -E[(1-D)Y1(Y>G_{0U}^{-1}(1-r))|Z=c^-]\cdot E[(1-D)|Z=c^+]\le 0. && (2.17)
\end{aligned}$$

Therefore, we can formulate our null hypothesis H_0 as

$$H_0:\theta_j\le 0 \text{ for } j=1,\,2,\,3 \text{ and } 4. \tag{2.18}$$

3 Proposed Test

3.1 Estimation of θ_j's

We first consider the estimation of θ_j for $j=1,\dots,4$ and derive the asymptotics of corresponding estimators before proposing a test for H_0 defined in (2.18).

Let $K(\cdot)$ be a kernel function and h a bandwidth. For a general random variable A, let $\widehat{E}[A|Z=c^+]$ and $\widehat{E}[A|Z=c^-]$ be the local linear estimators for $E[A|Z=c^+]$ and $E[A|Z=c^-]$, respectively. To be specific,

$$\begin{aligned}
(\widehat{E}[A|Z=c^+],\hat{b}_+) &= \arg\min_{a,b}\sum_{i=1}^n 1(Z_i\ge c)\cdot K\left(\frac{Z_i-c}{h}\right)\Big[A_i-a-b\cdot(Z_i-c)\Big]^2,\\
(\widehat{E}[A|Z=c^-],\hat{b}_-) &= \arg\min_{a,b}\sum_{i=1}^n 1(Z_i< c)\cdot K\left(\frac{Z_i-c}{h}\right)\Big[A_i-a-b\cdot(Z_i-c)\Big]^2.
\end{aligned}$$

We consider multi-step estimators for θ_j's. Let $G_1(y)=E[1(Y\le y)|D=1,Z=c^+]=E[D1(Y\le y)|Z=c^+]/E[D|Z=c^+]$ and $G_0(y)=E[1(Y\le y)|D=0,Z=c^-]=E[(1-D)1(Y\le y)|Z=c^-]/E[1-D|Z=c^-]$ be estimated by

$$\widehat{G}_1(y)=\frac{\widehat{E}[D1(Y\le y)|Z=c^+]}{\widehat{E}[D|Z=c^+]},\quad \widehat{G}_0(y)=\frac{\widehat{E}[(1-D)1(Y\le y)|Z=c^-]}{\widehat{E}[1-D|Z=c^-]}.$$

Let $q = P_{1|0}/P_{1|1} = E[D|Z = c^-]/E[D|Z = c^+]$ and $r = P_{0|1}/P_{0|0} = E[1 - D|Z = c^+]/E[1 - D|Z = c^-]$ be estimated by

$$\hat{q} = \frac{\widehat{E}[D|Z = c^-]}{\widehat{E}[D|Z = c^+]}, \quad \hat{r} = \frac{\widehat{E}[1 - D|Z = c^+]}{\widehat{E}[1 - D|Z = c^-]}.$$

Let $G_{1L}^{-1}(q)$, $G_{1U}^{-1}(1 - q)$, $G_{0L}^{-1}(r)$, $G_{0U}^{-1}(1 - r)$ be estimated by

$$\begin{aligned}
\widehat{G}_{1L}^{-1}(\hat{q}) &= \inf\{y \in \mathcal{Y} : \widehat{G}_1(y) \geq \hat{q}\}, \quad \widehat{G}_{1U}^{-1}(1 - \hat{q}) = \sup\{y \in \mathcal{Y} : \widehat{G}_1(y) \leq 1 - \hat{q}\}, \\
\widehat{G}_{0L}^{-1}(\hat{r}) &= \inf\{y \in \mathcal{Y} : \widehat{G}_0(y) \geq \hat{r}\}, \quad \widehat{G}_{0U}^{-1}(1 - \hat{r}) = \sup\{y \in \mathcal{Y} : \widehat{G}_0(y) \leq 1 - \hat{r}\}.
\end{aligned}$$

Then θ_1, θ_2, θ_3 and θ_4 are estimated by

$$\begin{aligned}
\hat{\theta}_1 = {} & \widehat{E}[DY1(Y < \widehat{G}_{1L}^{-1}(\hat{q}))|Z = c^+] \cdot \widehat{E}[D|Z = c^-] \\
& - \widehat{E}[DY|Z = c^-] \cdot \widehat{E}[D1(Y < \widehat{G}_{1L}^{-1}(\hat{q}))|Z = c^+], \\
\hat{\theta}_2 = {} & \widehat{E}[DY|Z = c^-] \cdot \widehat{E}[D1(Y > \widehat{G}_{1U}^{-1}(1 - \hat{q}))|Z = c^+] \\
& - \widehat{E}[DY1(Y > \widehat{G}_{1U}^{-1}(1 - \hat{q}))|Z = c^+] \cdot \widehat{E}[D|Z = c^-], \\
\hat{\theta}_3 = {} & \widehat{E}[(1 - D)Y1(Y < \widehat{G}_{0L}^{-1}(\hat{r}))|Z = c^-] \cdot \widehat{E}[1 - D|Z = c^+] \\
& - \widehat{E}[(1 - D)Y|Z = c^+] \cdot \widehat{E}[(1 - D)1(Y < \widehat{G}_{0L}^{-1}(\hat{r}))|Z = c^-], \\
\hat{\theta}_4 = {} & \widehat{E}[(1 - D)Y|Z = c^+] \cdot \widehat{E}[(1 - D)1(Y > \widehat{G}_{0U}^{-1}(1 - \hat{r}))|Z = c^-] \\
& - \widehat{E}[(1 - D)Y1(Y > \widehat{G}_{0U}^{-1}(1 - \hat{r}))|Z = c^-] \cdot \widehat{E}[(1 - D)|Z = c^+].
\end{aligned}$$

In the appendix, under suitable regularity conditions, we derive the asymptotic linear representations of the estimators $\sqrt{nh}(\hat{\theta} - \theta)$ that take into account the estimation effects of the estimated nuisance parameters where $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)'$. We also show the joint asymptotic normality of the estimators in that $\sqrt{nh}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \Omega)$ and Ω is a 4×4 asymptotic covariance matrix.

3.2 Weighted Bootstrap

The analytical variance estimator of the proposed estimators could be tedious to calculate. As in Hsu and Shen (2022), we propose to use a weighted bootstrap procedure first introduced in Ma and Kosorok (2005) to simulate the limiting distribution of the proposed estimators.

Let $\{W_i\}_{i=1}^n$ be a sequence of pseudo-random variables that is independent of the sample path with both mean and variance equal to one. For a generic random variable A, let $\widehat{E}^w[A|Z=c^+]$ and $\widehat{E}^w[A|Z=c^-]$ be the weighted bootstrap local linear estimators for $E[A|Z=c^+]$ and $E[A|Z=c^-]$, respectively. To be specific,

$$(\widehat{E}^w[A|Z=c^+], \hat{b}_+^w) = \arg\min_{a,b} \sum_{i=1}^n W_i \cdot 1(Z_i \geq c) \cdot K\left(\frac{Z_i - c}{h}\right)\Big[A_i - a - b \cdot (Z_i - c)\Big]^2,$$
$$(\widehat{E}^w[A|Z=c^-], \hat{b}_-^w) = \arg\min_{a,b} \sum_{i=1}^n W_i \cdot 1(Z_i < c) \cdot K\left(\frac{Z_i - c}{h}\right)\Big[A_i - a - b \cdot (Z_i - c)\Big]^2.$$

Let the weighted bootstrap estimators for $G_1(y)$ and $G_0(y)$ be

$$\widehat{G}_1^w(y) = \frac{\widehat{E}^w[D1(Y \leq y)|Z=c^+]}{\widehat{E}^w[D|Z=c^+]}, \quad \widehat{G}_0^w(y) = \frac{\widehat{E}^w[(1-D)1(Y \leq y)|Z=c^-]}{\widehat{E}^w[1-D|Z=c^+]}.$$

Let the weighted bootstrap estimators for q and r be

$$\hat{q}^w = \frac{\widehat{E}^w[D|Z=c^-]}{\widehat{E}^w[D|Z=c^+]}, \quad \hat{r}^w = \frac{\widehat{E}^w[1-D|Z=c^+]}{\widehat{E}^w[1-D|Z=c^-]}.$$

Let the weighted bootstrap estimators for $G_{1L}^{-1}(q)$, $G_{1U}^{-1}(1-q)$, $G_{0L}^{-1}(r)$, $G_{0U}^{-1}(1-r)$ be

$$\widehat{G}_{1L}^{-1,w}(\hat{q}) = \inf\{y \in \mathcal{Y} : \widehat{G}_1^w(y) \geq \hat{q}^w\}, \quad \widehat{G}_{1U}^{-1,w}(1-\hat{q}) = \sup\{y \in \mathcal{Y} : \widehat{G}_1^w(y) \leq 1-\hat{q}^w\},$$
$$\widehat{G}_{0L}^{-1,w}(\hat{r}) = \inf\{y \in \mathcal{Y} : \widehat{G}_0^w(y) \geq \hat{r}^w\}, \quad \widehat{G}_{0U}^{-1,w}(1-\hat{r}) = \sup\{y \in \mathcal{Y} : \widehat{G}_0^w(y) \leq 1-\hat{r}^w\}.$$

Then the weighted bootstrap estimators for θ_1, θ_2, θ_3 and θ_4 are

$$\begin{aligned}
\hat{\theta}_1^w &= \widehat{E}^w[DY1(Y < \widehat{G}_{1L}^{-1}(\hat{q}^w))|Z=c^+] \cdot \widehat{E}^w[D|Z=c^-] \\
&\quad - \widehat{E}^w[DY|Z=c^-] \cdot \widehat{E}^w[D1(Y < \widehat{G}_{1L}^{-1,w}(\hat{q}^w))|Z=c^+], \\
\hat{\theta}_2^w &= \widehat{E}^w[DY|Z=c^-] \cdot \widehat{E}^w[D1(Y > \widehat{G}_{1U}^{-1,w}(1-\hat{q}^w))|Z=c^+] \\
&\quad - \widehat{E}^w[DY1(Y > \widehat{G}_{1U}^{-1,w}(1-\hat{q}^w))|Z=c^+] \cdot \widehat{E}^w[D|Z=c^-], \\
\hat{\theta}_3^w &= \widehat{E}^w[(1-D)Y1(Y < \widehat{G}_{0L}^{-1,w}(\hat{r}^w))|Z=c^-] \cdot \widehat{E}^w[1-D|Z=c^+] \\
&\quad - \widehat{E}^w[(1-D)Y|Z=c^+] \cdot \widehat{E}^w[(1-D)1(Y < \widehat{G}_{0L}^{-1,w}(\hat{r}^w))|Z=c^-], \\
\hat{\theta}_4^w &= \widehat{E}^w[(1-D)Y|Z=c^+] \cdot \widehat{E}^w[(1-D)1(Y > \widehat{G}_{0U}^{-1,w}(1-\hat{r}^w))|Z=c^-]
\end{aligned}$$

$$- \widehat{E}^w[(1-D)Y1(Y > \widehat{G}_{0U}^{-1,w}(1-\hat{r}^w))|Z=c^-] \cdot \widehat{E}^w[(1-D)|Z=c^+].$$

Under the same set of regularity conditions (see details in Appendix B), we can derive the asymptotic linear representation of the weighted bootstrap estimators $\sqrt{nh}(\hat{\theta}^w - \hat{\theta})$ and show that $\widehat{\Phi}^w = \sqrt{nh}(\hat{\theta}^w - \hat{\theta})$ converges to the same limiting distribution as $\sqrt{nh}(\hat{\theta} - \theta)$ conditional on the same path with probability approaching one. That is, $\widehat{\Phi}$ can approximate the limiting distribution of $\sqrt{nh}(\hat{\theta} - \theta)$ well.

3.3 Test statistics and the decision rule

We define the test statistic as

$$\widehat{T}_n = \sqrt{nh} \max_{j=1,\ldots,4} \frac{\hat{\theta}_j}{\hat{\sigma}_j}, \tag{3.1}$$

where $\hat{\sigma}_j^2$ is a consistent estimator for σ_j^2, the asymptotic variance of $\sqrt{nh}(\hat{\theta}_j - \theta_j)$ for $j \in \{1,2,3,4\}$. For $\hat{\sigma}_j^2$, we suggest using the weighted bootstrap estimators. To be specific, let $b = 1, \ldots, B$ and say $B = 1000$. Then for each b, we get $\hat{\theta}^{w,b}$ and let $\hat{\sigma}_j^2 = nhB^{-1}\sum_{b=1}^{B}(\hat{\theta}_j^{w,b} - \hat{\theta}_j)^2$.

Define $\hat{\mu}_j = \hat{\theta}_j 1(\sqrt{nh}\hat{\theta}_j \leq -a_n\hat{\sigma}_j)$ where a_n is sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} a_n\sqrt{n^{-1}h^{-1}} = 0$. For significance level $\alpha < 1/2$, define the critical value as $\hat{c}_n(\alpha) = \max\{\tilde{c}_n(\alpha), 0\}$ where $\tilde{c}_n(\alpha)$ is defined as

$$\tilde{c}_n(\alpha) = \inf_c \Big\{c:\ P\Big(\max_{j=1,\ldots,4}\Big\{\frac{\hat{\phi}_j^w + \sqrt{nh}\hat{\mu}_j}{\hat{\sigma}_j}\Big\} \leq c\Big) \geq 1-\alpha\Big\}.$$

The **decision rule** will be "Reject H_0 if $\widehat{T}_n > \hat{c}_n(\alpha)$."

Proposition 3.3 *Suppose that Assumptions B.1 to B.7 in Appendix B hold and let* $0 < \alpha < 1/2$. *Then under* H_0 *in (??),* $\lim_{n\to\infty} P(\widehat{T}_n > \hat{c}_n(\alpha)) \leq \alpha$*; under* H_1, $\lim_{n\to\infty} P(\widehat{T}_n > \hat{c}_n(\alpha)) = 1$.

For implementation of our test, one can set W_i as normal distributions with mean and variance both equal to 1, but we suggest setting W_i as a binary variable taking values on 0 and 2 with equal probability, so all the realized weights W_i will be non-negative. We also suggest setting $a_n = 0.1\sqrt{\log\log n}$ as in Donald and Hsu (2016) or $a_n = \sqrt{0.3\log n}$ as in Andrews and Shi (2013).

4 Simulation

In this section, we consider a few numerical examples to illustrate the performance of our procedure. For comparison purpose, our first set of DGPs are the same as the power designs in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), which we listed below:

DGP1 Let $Z \sim N(0,1)$ truncated at -2 and 2. The propensity score is given by

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

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

DGP2 Same as DGP1 except that $Y|(D=1,Z=z) \sim N(0, 1.675^2)$ for all $z \in [-2,0)$.

DGP3 Same as DGP1 except that $Y|(D=1,Z=z) \sim N(0, 0.515^2)$ for all $z \in [-2,0)$.

DGP4 Same as DGP1 except that $Y|(D=1,Z=z) \sim \sum_{j=1}^{5} \omega_j N(\mu_j, 0.125^2)$ for all $z \in [-2,0)$, where $\omega = (0.15, 0.2, 0.3, 0.2, 0.15)$ and $\mu = (-1, -0.5, 0, 0.5, 1)$.

Note that all four DGPs violate the null hypothesis in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) because the conditional distributions of the potential outcome are not continuous near the cutoff. DGP1 has a location shift, so it violates the FRD mean assumption and therefore violates our null hypothesis. On the other hand, DGP2, DGP3, and DGP4 satisfy our null hypothesis. In DGP2 and DGP3, only the conditional variance changes but not the conditional expectation. For DGP4, while the shape of the distribution changes from normal to a mixture of normals, the conditional expectation is still zero on both sides of the cutoff.

For all the designs, we consider five sample sizes $n \in \{500, 1000, 2000, 4000, 8000\}$, 1000 bootstrap draws, 800 replications, and three significance levels $\alpha \in \{1\%, 5\%, 10\%\}$. We set $a_n = 0.1\sqrt{\log \log n}$ following Donald and Hsu (2016). For the bandwidth, 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). To deal with the bias in the local regression, we consider undersmoothing (US) by multiplying each bandwidth by $n^{\frac{1}{5}-\frac{1}{c}}$ with $c = 4.5$. We also consider MSE-optimal robust bias correction (MSE-RBC) implementation (see

Calonico, Cattaneo, and Farrell, 2018) and the rule-of-thumb coverage error rate (CER-RBC) optimal implementation (see Calonico, Cattaneo, and Farrell, 2020).

Table 1 reports the results for $\alpha = 5\%$. Full sets of results for other significance levels are collected in Tables 7 and 8 of Appendix C and give the same results qualitatively. The first panel (DGP1) is the power design, where the local continuity in means assumption is violated. We can see the rejection rate is low when the sample size is small ($n = 500$), which is not surprising because there are fewer observations near the cutoff to provide screening power. However, the rejection rate increases as the sample size increases for all choices of bandwidths.

Panels 2-4 of Table 1 are size designs in which both Assumptions 2.1 and 2.2 are satisfied. For these designs, all the rejection rates are below the nominal level of 5% for all sample sizes, suggesting that the size is well controlled. As the sample size increases, the rejection rates get closer to the nominal rate. Overall, the MSE-RBC and CER-RBC implementation work slightly better than undersmoothing, although they are not designed for model specification tests. On the other hand, Arai, Hsu, Kitagawa, Mourifié, and Wan (2022, Table 2) rejects DGP2-4 because either the variance or the shape of the potential outcome distribution is not continuous near the cutoff.

We also conduct another set of experiments to examine how the rejection rate varies with the "magnitude" of violation.

DGP5 The same as DGP1 except that $Y|(D = 1, Z = z) \sim N(-d, 1)$ for $z \in [-2, 0)$, where $d \in \{0.1, 0.2, \cdots, 1.0\}$.

Figure 1 plots the rejection rate of using IK-US bandwidth at different values of d and sample size n. The results of using other bandwidths are similar. When $d = 0$, the local continuity in the mean condition is satisfied, and we would expect the rejection rate to be no larger than the nominal rate. As d increases, the magnitude of violation is larger, and we expect to see the rejection rate increase. This is confirmed by Figure 1.

Finally, we consider DGP6, where we allow the jump size π of the propensity score to change.

DGP6 The same as DGP1 except that

$$P(D = 1|Z = z) = \mathbf{1}\{-2 \leq z < 0\} \max\left\{0, (z + 2)^2/8 - \frac{\pi}{2}\right\} + \mathbf{1}\{0 \leq z \leq 2\} \min\left\{1, 1 - (z - 2)^2/8 + \frac{\pi}{2}\right\}.$$

Table 1: Rejection Rate at 5% Level

	n	Undersmoothing IK	Undersmoothing CCT	Undersmoothing AI	MSE-RBC IK	MSE-RBC CCT	MSE-RBC AI	CER-RBC IK	CER-RBC CCT	CER-RBC AI
	500	0.125	0.071	0.179	0.039	0.005	0.089	0.081	0.044	0.184
DGP1	1000	0.371	0.275	0.340	0.198	0.080	0.221	0.310	0.210	0.344
(Power)	2000	0.558	0.506	0.564	0.414	0.288	0.424	0.510	0.443	0.565
	4000	0.800	0.679	0.824	0.665	0.511	0.660	0.745	0.628	0.823
	8000	0.970	0.916	0.964	0.893	0.763	0.868	0.940	0.871	0.965
	500	0.013	0.009	0.028	0.003	0.001	0.018	0.011	0.004	0.021
DGP2	1000	0.054	0.024	0.020	0.026	0.005	0.035	0.043	0.021	0.019
(Size)	2000	0.043	0.044	0.035	0.031	0.030	0.059	0.041	0.039	0.038
	4000	0.035	0.049	0.016	0.044	0.038	0.033	0.035	0.048	0.015
	8000	0.035	0.039	0.033	0.051	0.048	0.046	0.048	0.045	0.033
	500	0.010	0.009	0.018	0.005	0.001	0.028	0.008	0.003	0.018
DGP3	1000	0.036	0.018	0.028	0.011	0.010	0.050	0.028	0.011	0.026
(Size)	2000	0.043	0.030	0.021	0.026	0.024	0.044	0.034	0.030	0.023
	4000	0.026	0.034	0.035	0.033	0.025	0.055	0.033	0.041	0.040
	8000	0.034	0.041	0.025	0.045	0.040	0.043	0.041	0.035	0.024
	500	0.013	0.010	0.019	0.004	0.000	0.031	0.010	0.009	0.014
DGP4	1000	0.038	0.013	0.026	0.019	0.008	0.048	0.029	0.013	0.030
(Size)	2000	0.040	0.045	0.031	0.039	0.028	0.055	0.040	0.044	0.030
	4000	0.049	0.029	0.029	0.040	0.041	0.051	0.053	0.033	0.038
	8000	0.021	0.041	0.031	0.030	0.033	0.045	0.029	0.035	0.033

Here, $\pi \in \{0, 0.05, 0.1, 0.15, \cdots, 0.6\}$ and $Y|(D = 1, Z = z) \sim N(-d, 1)$ for $z \in [-2, 0)$ for $d \in \{0.7, 1.0, 1.5\}$. As we discussed earlier in Remark 2.3, when π increases, q (or r) will decrease (given everything else equal). Therefore, the bounds in Proposition 2.2 will be wider and thus we will expect a lower rejection rate. Figure 2 verifies this point for bandwidth IK-US and sample size $n = 8000$. Again, the results for other sample sizes and bandwidths are qualitatively similar.

5 Empirical Application

In this section, we illustrate the use of our method in a few empirical applications.[5]

[5]We thank the authors of Miller, Pinto, and Vera-Hernández (2013), Angrist and Lavy (1999), Pop-Eleches and Urquiola (2013), and Battistin, Brugiavini, Rettore, and Weber (2009) for sharing the data or making the data publicly available on journal websites. All errors in the empirical illustration are ours.

Figure 1: Rejection Rate at 5%, IK-US Bandwdith (DGP5)

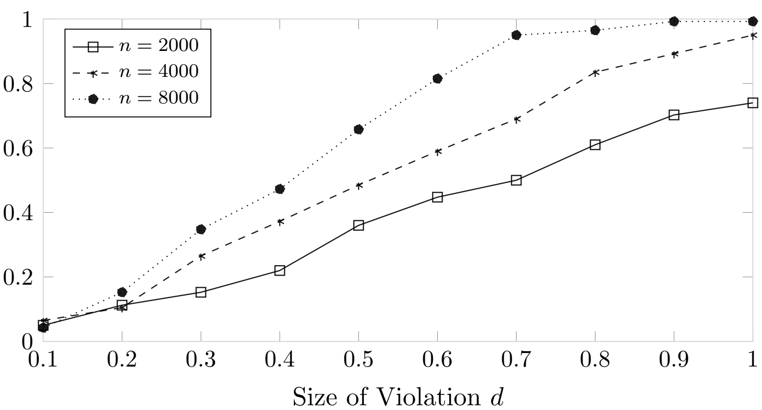
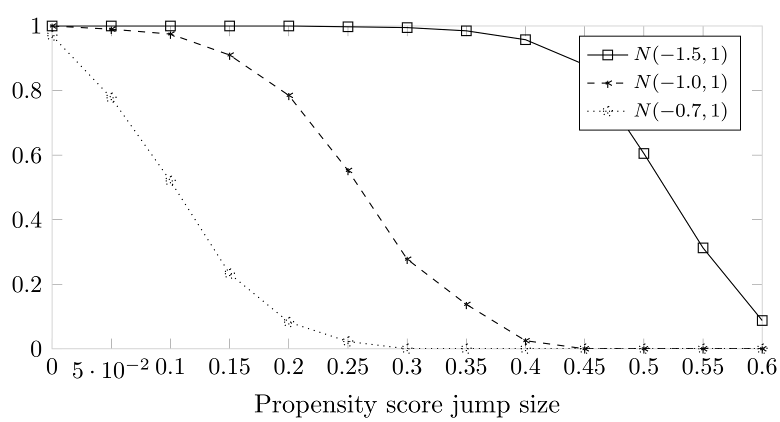


Figure 2: Rejection Rate at 5%, IK-US Bandwdith (DGP6)



5.1 Effect of Enrolling in a Subsidized Insurance Program

We first access the validity of Assumptions 2.1 and 2.2 in the empirical context studied by Miller, Pinto, and Vera-Hernández (2013), who use the FRD design to identify the causal effect of enrolling in a publicly funded insurance program (Subsidized Regime, SR) on many household-level outcome variables in Columbia. In Columbia, a household is eligible to enroll in SR if their SISBEN score (Sistema de Identificación de Beneficiarios, a continuous index taking values from 0 to 100, with 0 being the poorest) is below a cutoff. The SISBEN score thus serves as the running variable. In their empirical implementation, Miller, Pinto, and Vera-Hernández (2013) use a simulated SISBEN score to alleviate the threat of possible manipulation on the score, and the resulting density passes the density test and appears to be continuous at the cutoff.

Motivated by the observation that the continuity of running variable density is neither sufficient nor necessary to identify the local average treatment effect (LATE), Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) test the set of (distributional) identifying assumptions for LATE-type parameters. They found that the FRD distributional assumptions are rejected for three dependent variables: "household educational spending", "total spending on food", and "total monthly expenditure". In this application, the monotonicity assumption appears to be reasonable. Therefore, the rejection can be interpreted as the discontinuity of the conditional distribution of potential outcomes of these three dependent variables given the running variable (SISBEN score) near the cutoff. One needs to be careful when making inference on parameters that requires distributional identification assumptions, such as quantile treatment effect. However, discontinuity in the conditional distribution does not necessarily imply a discontinuity in the conditional expectation. If the local continuity in means (Assumption 2.2) holds, we can still credibly identify the mean effect.

Table 2 reports the p-values of our test on the three dependent variables under different bandwidth choices (including the three fixed bandwidths used in Miller, Pinto, and Vera-Hernández, 2013). The bandwidth values are reported in Table 9 in Appendix C. We observe no rejection across the board even at the 10% level. While our test is designed for the necessary implications of the local continuity in means assumptions and local monotonicity instead of their sufficient conditions, the results in Table 2 do suggest that the violation of continuity in distributions is more likely caused by discontinuity of higher moments (such as variance) or tail shapes.

Table 2: Testing Results for Columbia's SR Data: p-values

Bandwidth	Household edu.exp.	Total exp.on food	Total monthly exp.
2	0.591	0.613	0.605
3	0.522	0.550	0.717
4	0.582	0.535	0.865
IK-US	0.608	0.506	0.688
IK-MSE-RBC	0.594	0.539	0.662
IK-CER-RBC	0.521	0.564	0.620
CCT-US	0.539	0.513	0.737
CCT-MSE-RBC	0.616	0.573	0.659
CCT-CER-RBC	0.520	0.524	0.686
AI-US	0.770	0.500	0.734
AI-MSE-RBC	0.831	0.720	0.642
AI-CER-RBC	0.653	0.591	0.751

5.2 Effect of Class Size

Our second empirical application is the one studied by Angrist and Lavy (1999) and Angrist, Lavy, Leder-Luis, and Shany (2019), where Israel's Maimonides' rule creates an FRD design and can be used to identify the effect of class size on students' performance. Maimonides' rule in Israel's public school system requires that the class size be no larger than 40 students. Whenever the enrollment exceeds 40, the school must offer at least two classes. Under this policy, therefore, the average class size of a grade as a function of enrollment is discontinuous at the multiples of the upper limit (40, 80, 120 etc.). In practice, some schools choose smaller class sizes than 40. This creates an FRD design because the probability of dividing classes is larger than zero before reaching the cutoff. In a seminal paper, Angrist and Lavy (1999) use this FRD design to identify the causal effect of class size on students' performance.

There are concerns about the validity of the identification strategy due to possible manipulation of the enrollment (running variable). For example, Otsu, Xu, and Matsushita (2013) found that the enrollment density is not continuous at some of the cutoffs. However, as discussed in Angrist, Lavy, Leder-Luis, and Shany (2019), the discontinuity of running variable density is likely caused by schools' budgetary consideration and is independent of students' potential performance, and therefore need not violate the identifying assumptions for the LATE parameters. This discussion is supported by Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), who test the (distributional) identifying assumptions for four dependent variables (grade 4 and 5's math and vocabulary) and did not find evidence for rejection.

Table 3: Testing Results for Israeli School Data (Grade 4): p-values

	g4math			g4verb		
Bandwidth	40	80	120	40	80	120
3	0.814	0.614	1.000	0.569	0.38	0.787
5	0.979	0.638	0.988	0.44	0.842	0.963
IK-US	0.996	0.93	0.988	0.419	0.897	0.668
IK-MSE-RBC	0.611	0.551	0.978	0.416	0.53	0.956
IK-CER-RBC	0.982	0.822	0.942	0.481	0.907	0.971
CCT-US	0.997	0.829	0.973	0.786	0.891	0.853
CCT-MSE-RBC	0.981	0.879	0.544	0.462	0.895	0.932
CCT-CER-RBC	0.998	0.902	0.985	0.584	0.905	0.861
AI-US	1.000	0.79	0.901	0.438	0.823	0.715
AI-MSE-RBC	0.979	0.884	0.958	0.38	0.825	0.938
AI-CER-RBC	0.997	0.803	0.899	0.468	0.814	0.859

Table 4: Testing Results for Israeli School Data (Grade 5): p-values

	g5math			g5verb		
Bandwidth	40	80	120	40	80	120
3	0.8.00	0.482	0.656	0.734	0.545	0.601
5	0.854	0.344	0.603	0.845	0.470	0.359
IK-US	0.617	0.614	0.888	0.653	0.821	0.811
IK-MSE-RBC	0.959	0.469	0.369	0.791	0.452	0.315
IK-CER-RBC	0.742	0.464	0.620	0.687	0.713	0.735
CCT-US	0.790	0.902	0.501	0.620	0.687	0.790
CCT-MSE-RBC	0.902	0.415	0.394	0.996	0.763	0.240
CCT-CER-RBC	0.869	0.766	0.737	0.760	0.686	0.741
AI-US	0.533	0.865	0.865	0.649	0.472	0.883
AI-MSE-RBC	0.631	0.819	0.933	0.956	0.818	0.905
AI-CER-RBC	0.479	0.917	0.974	0.739	0.466	0.927

In this subsection, we revisit this empirical question. Since the distributional test does not reject the continuity of conditional distributions, we expect our test on the continuity of conditional expectations will conclude with no rejection either. It is indeed the case. As reported in Tables 3 and 4 (bandwidth values reported in Tables 10 and 11), the p-values for the cutoff 40 are greater than 5% for all bandwidths choices and all four dependent variables, and they are greater than 10% for nearly all combinations.

5.3 Effect of Attending Better Schools

Estimating the effect of school quality on student performance is one of the most important research questions in labour/education economics. The difficulty lies in that students are heterogeneous in their ability and how much they can benefit from a higher-achievement school, and they are not randomly allocated to different schools. Pop-Eleches and Urquiola (2013) apply the FRD design to Romanian secondary school data and find that students who enroll in better schools tend to perform better in the Baccalaureate exams, among other findings.

In Romania, students' chances of enrolling in higher-ranked schools solely depend on a performance measure in schools, which depends on their nationwide test outcome and their GPA. The centralized allocation process satisfies the needs of students with higher scores first, thus creating cutoff scores at which the enrollment probability (in better schools) changes discontinuously. Please see Pop-Eleches and Urquiola (2013) for detailed institutional background. If the students who are just above the cutoff on average benefit from the higher-achievement school the same way as those who are just below the cutoff, such jumps in enrollment probability can provide identification power for the causal effect near the cutoff.

In our empirical illustration, the outcome variable is the continuous Baccalaureate exam score. The running variable is the transition score, and the treatment variable is if a student enrolls in a "better school." Here we consider two cutoffs: enrolling in the best school in town or avoiding the worst school in town. The validity of an FRD design using test scores as a cutoff is not self-ensured and depends on specific empirical contexts. For example, if a school teacher has a targeted group of students that he/she always prefers to put on the treatment (or on the right-hand side of the cutoff), then the teacher may manipulate the cutoff to guarantee this. If this group of students is different from other students in an unobserved way, then the local continuity condition can be violated. See discussions about running variable manipulation in Gerard, Rokkanen, and Rothe (2020), too. The FRD design using the Romanian secondary school transition test, however, is likely to be valid since the test is at the national level and the cutoffs are quite difficult to manipulate.

The testing results are reported in Table 5, with bandwidth values reported in Table 12 of Appendix C. We see that the validity of Assumptions 2.1 and 2.2 are not rejected at 10%

throughout different choices of bandwidths. As a comparison, we also conduct the distributional test of Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), and obtain the same result qualitatively.[6]

Table 5: Testing Results for Romania High School Data: p-values

Bandwidth	Attending best school: Mean Test	Attending best school: Distr. Test	Avoiding worst school: Mean Test	Avoiding worst school: Distr. Test
0.100	0.904	0.900	0.625	0.271
0.200	0.377	0.999	0.818	0.838
0.300	0.278	1.000	0.879	0.967
IK-US	0.467	0.635	0.719	0.815
IK-MSE-RBC	0.545	0.588	0.748	0.786
IK-CER-RBC	0.445	0.590	0.754	0.848
CCT-US	0.450	0.561	0.689	0.666
CCT-MSE-RBC	0.505	0.871	0.356	0.514
CCT-CER-RBC	0.475	1.000	0.593	0.350
AI-US	0.469	0.670	0.785	0.100
AI-MSE-RBC	0.533	0.553	0.787	0.213
AI-CER-RBC	0.443	0.570	0.793	0.199

5.4 Effect of Retirement on Consumption

As population aging accelerates in developed countries, there is an increasing number of studies on the impact of retirement on personal physical health, psychological health, cognitive competence, and family income and consumption. The key issue for identifying the causal effect is the endogeneity of the retirement decision. One common solution is using RD designs based on retirement-related policies or incentives. For example, many countries implement "official retirement ages," and such legislation provides exogenous variations for retirement decisions; see Müller and Shaikh (2018) for a summary of OECD country retirement ages.

Our empirical illustration uses the data from Battistin, Brugiavini, Rettore, and Weber (2009), which identify the effect of Italy seniors' retirement on consumption drop. The idea is that becoming eligible for a pension provides an additional incentive for retirement; thus, as empirically observed, the retirement probability changes discontinuously at the eligibility cutoff. Suppose the seniors who are marginally younger than the cutoff age are comparable to those who are marginally older in their average potential consumption behavior. In that case, such

[6] Pop-Eleches and Urquiola (2013) reports that McCrary (2008)'s density test does not reject the continuity of the running variable density at the cutoffs; we do not repeat the test here.

an FRD design can identify the causal effect of retirement on consumption.

In our implementation, we follow Battistin, Brugiavini, Rettore, and Weber (2009) and choose the running variable as the difference between the ”family head's” age and the eligibility age. The treatment variable is retirement. We consider three outcome variables. They are log values of total expenditure, total non-durable goods consumption, and food consumption. Because the running variable is discrete (age by year), we do not implement data-dependent bandwidths choices. Instead, we consider a wide range of choices from 3 to 10. The test does not reject Assumptions 2.1 and 2.2 for all three outcome variables across all bandwidth choices at 10%: the p-values are quite high. We observe nearly no rejection for the distributional test at 10% either (except that the p-value for food consumption is around 10% when the bandwidth is small, but they are all above 5%). Overall, we do not see strong evidence against the validity of the FRD design (either for the mean assumptions or for distributional assumptions).

Table 6: Testing Results for Italian Retirement Consumption Data: p-values

	Mean Test			Distribution Test		
Bandwidth	Total Exp	Non Durable	Food	Total Exp	Non Durable	Food
3	0.555	0.613	0.241	0.709	0.831	0.121
4	0.847	0.812	0.696	0.941	0.883	0.087
5	0.913	0.845	0.500	0.970	0.940	0.106
6	0.758	0.857	0.579	0.998	0.491	0.067
7	0.727	0.939	0.486	0.990	0.382	0.358
8	0.857	0.873	0.468	0.911	0.495	0.884
9	0.739	0.671	0.488	0.904	0.875	0.988
10	0.777	0.676	0.527	0.731	0.795	0.985

6 Extensions and Discussions

6.1 Sharp bounds when Y is not continuous

The bounds reported in Proposition 2.2 are not necessarily sharp when Y is discrete. When the support of Y contains a large number of points, it is not unreasonable to treat it as continuous. For example, the exam score, although it only takes integer values, is often treated as continuous. In such cases, one can just apply our result in Proposition 2.2. Alternatively, one can also derive the sharp bounds in a similar way, as shown in the following corollary. The proof is collected in the proof of Proposition 2.2 and omitted.

Corollary 6.1 *Suppose the conditions for Proposition 2.2 are satisfied. Suppose $Y(d)$, $d \in \{0,1\}$, are discrete and takes value from a countable set $\mathcal{Y} = \{y_1, y_2, \cdots, y_J\}$, where $y_j < y_{j+1}$ for any $1 \le j < J$. Then the bounds in Proposition 2.2 can be tightened to*

$$LB^+ \le E[Y|D=1, Z=c^-] \le UB^+, \tag{6.1}$$

$$LB^- \le E[Y|D=0, Z=c^+] \le UB^-. \tag{6.2}$$

where

$$LB^+ \equiv \sum_{j=1}^{j^*} y_j \frac{P(Y=y_j|D=1, Z=c^+)}{q} + y_{j^*+1}\left\{1 - \frac{\sum_{j=1}^{j^*} P(Y=y_j|D=1, Z=c^+)}{q}\right\},$$

$$UB^+ \equiv \sum_{j=j^\dagger}^{J} y_j \frac{P(Y=y_j|D=1, Z=c^+)}{q} + y_{j^\dagger-1}\left\{1 - \frac{\sum_{j=j^\dagger}^{J} P(Y=y_j|D=1, Z=c^+)}{q}\right\},$$

$$LB^- \equiv \sum_{j=1}^{j^*} y_j \frac{P(Y=y_j|D=0, Z=c^-)}{r} + y_{j^*+1}\left\{1 - \frac{\sum_{j=1}^{j^*} P(Y=y_j|D=1, Z=c^-)}{r}\right\},$$

$$UB^- \equiv \sum_{j=j^\dagger}^{J} y_j \frac{P(Y=y_j|D=0, Z=c^-)}{q} + y_{j^\dagger-1}\left\{1 - \frac{\sum_{j=j^\dagger}^{J} P(Y=y_j|D=0, Z=c^-)}{r}\right\},$$

and the definitions for j^* and $j^\dagger$ are given in Equation (A.4) and Equation (A.7).

Example 6.1 *Consider the lower bound for always-takers' expectation in the special case of a binary outcome variable: $Y \in \{y_1, y_2\}$. In this case, the bounds in the statement of Proposition 2.2 will be trivial because inequalities 2.10 and 2.11 would simply imply:*

$$y_1 \le E[Y|D=1, Z=c^-] \le y_2.$$

However, the bounds derived on the left-hand sides of Equations (A.6) and (A.8) still have empirical content. To see this, if $P(Y=y_1|Z=c^+) > qP(D=1|Z=c^+)$, then $j^*=0$. This is the case where (conditioning on $Z=c^+$) the total size of always-takers is smaller than the size of the subpopulation for which $Y=y_j$. The smallest possible value of $\mathbb{E}[Y|\boldsymbol{A}, Z=c^+]$ would be generated by the distribution such that all the always-takers are concentrated on the *subpopulation of $Y = y_1$, which is just y_1 by Equation (A.6). On the other hand, if $P(Y = y_1|Z = c^+) \leq qP(D = 1|Z = c^+)$, then $j^* = 1$ and there are more always-takers than the size of the subpopulation $Y = y_1$. Hence, the smallest value of $\mathbb{E}[Y|\boldsymbol{A}, Z = c^+]$ would be generated by the distribution where we allocate always-takers first to the cell $Y = y_1$, and then the rest to the cell $Y = y_2$, and it gives bound as*

$$y_1 \frac{P(Y = y_1|D = 1, Z = c^+)}{q} + y_2 \left(1 - \frac{P(Y = y_1|D = 1, Z = c^+)}{q}\right).$$

To summarize, in the binary outcome case, when $P(Y = y_1|Z = c^+) \leq qP(D = 1|Z = c^+)$, the lower bound of $E[Y|D = 1, Z = c^-]$ has empirical content and equals

$$y_1 \frac{P(Y = y_1|D = 1, Z = c^+)}{q} + y_2 \left(1 - \frac{P(Y = y_1|D = 1, Z = c^+)}{q}\right).$$

6.2 Including covariates X

Our testable implication can be extended if the local monotonicity and local continuity in means assumptions hold when conditioning on covariates X. In particular, consider:

Assumption 6.1 (Conditional local monotonicity) $\lim_{z\downarrow c} P(T = \mathbf{DF}|Z = z, X = x) = 0$ *and* $\lim_{z\uparrow c} P(T = \mathbf{DF}|Z = z, X = x) = 0$ *for all $x \in \mathcal{X}$.*

Let $f_d(y|t, z, x)$ be the conditioning density of $Y(d)$ given $T = t$, $Z = z$, and $X = x$.

Assumption 6.2 (Conditional local continuity in means) *For all $x \in \mathcal{X}$, assume that (i) $\lim_{z\downarrow c} f_d(y|t, z, x)$ and $\lim_{z\uparrow c} f_d(y|t, z, x)$ are proper densities and bounded away from zero for all $y \in \mathcal{Y}$; (ii) $E[Y(d)|T = t, Z = z, X = x]$ exists and is continuous in z in the neighborhood B_ϵ of c; (iii) $P(T = t|Z = z, X = x)$ is continuous in z in the neighborhood B_ϵ of c for any $t \in \{\mathbf{A}, \mathbf{N}, \mathbf{C}\}$.*

Let $G_{1x}(y) = \lim_{z\downarrow c} P(Y \leq y|D = 1, Z = z, X = x)$ be the conditional distribution of Y given $D = 1$, $Z = z$, $X = x$ when z converges to c from above. Similarly, define $G_{0x}(y) = \lim_{z\uparrow c} P(Y \leq y|D = 0, Z = z)$. We let $q_x = P_{1|0}(x)/P_{1|1(x)}$ where $P_{1|0}(x) = P(D = 1|Z = c^-, X = x)$ and $P_{1|1}(x) = P(D = 1|Z = c^+, X = x)$. Likewise, we define $r_x = P_{0|1}(x)/P_{0|0}(x)$. Again, G_{1x}, G_{0x}, q_x, and r_x are all directly identifiable from the data. Then we have the following results.

Corollary 6.2 *Suppose that Assumptions 6.1 and 6.2 are satisfied, and suppose for all* $x \in \mathcal{X}$*,* $q_x \in (0,1)$*, and* $r_x \in (0,1)$*, then the following inequality constraints hold:*

$$E[Y|D=1, Y < G_{1xL}^{-1}(q_x), Z=c^+, X=x] \le E[Y|D=1, Z=c^-, X=x], \tag{6.3}$$

$$E[Y|D=1, Z=c^-, X=x] \le E[Y|D=1, Y > G_{1xU}^{-1}(1-q_x), Z=c^+, X=x], \tag{6.4}$$

$$E[Y|D=0, Y < G_{0xL}^{-1}(r_x), Z=c^-, X=x] \le E[Y|D=0, Z=c^+, X=x], \tag{6.5}$$

$$E[Y|D=0, Z=c^+, X=x] \le E[Y|D=0, Y > G_{0xU}^{-1}(1-r_x), Z=c^-, X=x]. \tag{6.6}$$

To implement the testable implication, we can transform these inequalities as in Section 2. Take Corollary 6.2 as an example. It implies that

$$\begin{aligned}
& E[Y|D=1, Y < G_{1L}^{-1}(q), Z=c^+, X=x] - E[Y|D=1, Z=c^-, X=x] \le 0 \\
\Leftrightarrow\ & \frac{E[DY1(Y < G_{1L}^{-1}(q))|Z=c^+, X=x]}{E[D1(Y < G_{1L}^{-1}(q))|Z=c^+, X=x]} - \frac{E[DY|Z=c^-, X=x]}{E[D|Z=c^-, X=x]} \le 0 \\
\Leftrightarrow\ & \theta_1(x) \equiv E[DY1(Y < G_{1L}^{-1}(q))|Z=c^+, X=x] \cdot E[D|Z=c^-, X=x] \\
& \qquad - E[DY|Z=c^-, X=x] \cdot E[D1(Y < G_{1L}^{-1}(q))|Z=c^+, X=x] \le 0. \qquad (6.7)
\end{aligned}$$

Similarly calculating $\theta_j(x)$ for $j = 2, 3, 4$, we can then transform the null hypothesis as

$$H_0 : \sup_{x \in \mathcal{X}, j \in \{1,2,3,4\}} \theta_j(x) \le 0.$$

When X is discrete, our testing procedure in Section 3 can be easily extended by implementing the test on each subsample defined by the value of covariates. When X is continuous, it is possible to extend our results to this case by restricting x's to a compact subset of interior points of $\mathcal{X}$, but it is more technically challenging. Therefore, we leave this as a future extension.

7 Conclusion

This paper offers a new specification test for researchers interested in estimating the mean causal effect for compliers at the cutoff in FRD designs. The test is easy to implement, has the asymptotic size control under the null and is consistent against all fixed alternatives. We illustrate the use of this new test in several empirical examples and show how it complements the existing tests that target testing the continuity of potential outcome distributions, running variable densities, and baseline variable distributions. The Monte Carlo simulation shows our test performs well in finite samples with moderate sample sizes.

APPENDIX

A Proofs of the Main Results

A.1 Proof of Lemma 2.1

Without loss of generality, we consider the violation of inequality (2.5). We first consider the case $Y(1)$ is continuous over V. The proof is by construction. Let $y' \in V$ such that $y^* < y' < y^{**}$. Let $g^0(Y) = 1[y^* \leq Y \leq y']$. Clearly, $g^0 \in \mathcal{G}_{\mathcal{V}}$. For simplicity, consider a set of DGPs in which $Y(d)$ and type T are independent conditioning on Z and Assumption 2.1 is satisfied. Among this set, choose $Y(1)|Z = c^-$ be uniformly distributed over $[y^*, y^{**}]$, that is, $f_{Y(1)|Z=c^-}(y) = \delta$ for all $y \in [y^*, y^{**}]$. Let $Y(1)|Z = c^+$ have the same density as $Y(1)|Z = c^-$ outside of $[y^*, y^{**}]$, but

$$f_{Y(1)|Z=c^+}(y) = \begin{cases} 0, & \text{if } y \in [y^*, y'], \\ \frac{\delta P(Y(1) \in [y^*, y^{**}])}{P(Y(1) \in [y^*, y'])}, & \text{if } y \in (y', y^{**}] \end{cases}$$

It is easy to see that $E[Y(1)|Z = z^-] = E[Y(1)|Z = z^+]$ by construction. On the other hand,

$$\begin{aligned} &E[g^0(Y)D|Z = c^-] - E[g^0(Y)D|Z = c^+] \\ &= P(Y(1) \in [y^*, y'], T = \boldsymbol{A}|Z = c^-) - P(Y(1) \in [y^*, y'], T = \boldsymbol{A} \cup \boldsymbol{C}|Z = c^+) \\ &= P(Y(1) \in [y^*, y']|Z = c^-)P(T = \boldsymbol{A}|Z = c^-) - P(Y(1) \in [y^*, y']|Z = c^+)P(T = \boldsymbol{A} \cup \boldsymbol{C}|Z = c^+) \\ &= P(Y(1) \in [y^*, y']|Z = c^-)P(T = \boldsymbol{A}|Z = c^-) > 0, \end{aligned}$$

where the first equality holds by definition, the second equality holds because the candidate distribution has $Y(d)$ and T independent conditioning on Z, and the third equality and the last inequality hold by construction. Therefore, inequality (2.5) is violated.

Next, we consider the case in which $Y(1)$ is discrete but V contains at least three support points $\{y_1, y_2, y_3\}$ when conditioning on T and Z. Again, let $Y(1)$ and T be independent conditioning on Z, and let the density or probability mass of $Y(1)|Z = c^+$ and $Y(1)|Z = c^-$ be the exactly the same over $\mathcal{Y}/\{y_1, y_2, y_3\}$. Over $\{y_1, y_2, y_3\}$, let $P(Y(1) = y_k|Z = c^-) = \delta$ for all k; let

$$P(Y(1) = y_k|Z = c^+) = \begin{cases} 0, & \text{if } k = 2, \\ \delta + \delta\frac{y_2 - y_1}{y_3 - y_1}, & \text{if } k = 1, \\ \delta + \delta\frac{y_3 - y_2}{y_3 - y_1}, & \text{if } k = 3 \end{cases}$$

It is clear to see that $E[Y(1)|Z = z^-] = E[Y(1)|Z = z^+]$ by construction. However, if we choose $g^0(Y) = 1[y_2 \leq Y \leq y_2] = 1[Y = y_2]$, then $E[g^0(Y)D|Z = c^-] - E[g^0(Y)D|Z = c^+] > 0$ by the same reasoning as in the continuous case.

A.2 Proof of Proposition 2.2

We prove the first pair of inequalities (2.10) and (2.11); the other two hold analogously. Inequalities (2.10) and (2.11) provide bounds for $E[Y(1)|\boldsymbol{A}, Z = c^{+}]$. Under the identifying assumptions, $E[Y|D = 1, Z = c^{-}]$ can be written as:

$$\begin{aligned} E[Y|D=1,Z=c^{-}] &= \lim_{z\uparrow c} E[Y|D=1,Z=z] = \lim_{z\uparrow c} E[Y(1)|D=1,Z=z] \\ &= \lim_{z\uparrow c} E[Y(1)|\boldsymbol{A},Z=z] = \lim_{z\downarrow c} E[Y(1)|\boldsymbol{A},Z=z] = E[Y(1)|\boldsymbol{A},Z=c^{+}], \end{aligned} \tag{A.1}$$

where the first and last equalities hold by definition, the second equality holds by the definition of $Y(1)$, the third one holds by Assumption 2.1 (local montononicity) that when z approaches c from below, the event $\{D = 1, Z = z\}$ is equivalent to the event $\{T = \boldsymbol{A}, Z = z\}$, and the fourth equality holds by Assumption 2.2 (local continuity in means). Hence, the bounds for $E[Y(1)|\boldsymbol{A}, Z = c^{+}]$ is equivalent to the bounds for $E[Y|D = 1, Z = c^{-}]$. Likewise, under Assumptions 2.1 and 2.2,

$$q \equiv \frac{P(D=1|Z=c^{-})}{P(D=1|Z=c^{+})} = \frac{P(\boldsymbol{A}|Z=c^{-})}{P(\boldsymbol{A}\cup\boldsymbol{C}|Z=c^{+})} = \frac{P(\boldsymbol{A}|Z=c^{+})}{P(\boldsymbol{A}\cup\boldsymbol{C}|Z=c^{+})}. \tag{A.2}$$

In the following, we will derive sharp bounds for $E[Y(1)|\boldsymbol{A}, Z = c^{+}]$. Since we assume the conditioning density of $Y(d)$ is well-defined when a running variable converges to the cutoff from either side, we have

$$\begin{aligned} G_1(y) &\equiv \lim_{z\downarrow c} P(Y \le y|D=1,Z=z) = \lim_{z\downarrow c} P(Y(1) \le y|\boldsymbol{A}\cup\boldsymbol{C},Z=z) \\ &= \lim_{z\downarrow c}\{P(Y(1)\le y|\boldsymbol{A},Z=z)P(\boldsymbol{A}|\boldsymbol{A}\cup\boldsymbol{C},Z=z) + P(Y(1)\le y|\boldsymbol{C},Z=z)P(\boldsymbol{C}|\boldsymbol{A}\cup\boldsymbol{C},Z=z)\} \\ &= \lim_{z\downarrow c} P(Y(1)\le y|\boldsymbol{A},Z=z)q + \lim_{z\downarrow c} P(Y(1)\le y|\boldsymbol{C},Z=z)(1-q) \end{aligned}$$

where the first equality is by definition, the second equality is by definition of the potential outcome and the fact that when z approches to c from above, the event $\{D = 1, Z = z\}$ is equivalent to the event $\{T = \boldsymbol{A} \cup \boldsymbol{C}, Z = z\}$, the third equality is by the law of total probabilities, and the fourth equality is because all the probabilities are well-defined by assumption. Therefore, the observed distribution G_1 is the mixture of conditional distributions of $Y(1)$ for always-takers and compliers, with mixing weight equalling to q and $1 - q$, respectively. Our goal is to find the bounds of expectation of the mixing component $\lim_{z\downarrow c} P(Y(1) \le y|\boldsymbol{A}, Z = z)$.

(i) Suppose first the conditional distribution of $Y(1)$ given $D = 1$ and Z in a small neighborhood of c is continuous in y at its q-th quantile. Applying Horowitz and Manski (1995, Corollary 4.1), we know that the sharp bounds for $E[Y(1)|\boldsymbol{A}, Z = c^+]$ are given by

$$LB^+ \equiv E[Y|D = 1, Y < G_{1L}^{-1}(q), Z = c^+] \leq E[Y(1)|\boldsymbol{A}, Z = c^+],$$
$$E[Y(1)|\boldsymbol{A}, Z = c^+] \leq E[Y|D = 1, Y > G_{1U}^{-1}(1-q), Z = c^+] \equiv UB^+,$$

where the lower bound is generated by a DGP in which always-takers concentrate at the lower tail $\{y : G_1(y) \leq q\}$, and the upper bound is achieved when always-takers are concentrated at its upper tail. Using Equation (A.1) to replace $E[Y(1)|\boldsymbol{A}, Z = c^+]$ by $E[Y|D = 1, Z = z^-]$, we obtain inequalities (2.10) and (2.11).

(ii) Suppose $Y(d)$, $d \in \{0, 1\}$, are discrete and take values from a countable set $\mathcal{Y} = \{y_1, y_2, \cdots, y_J\}$, where $y_j < y_{j+1}$ for any $1 \leq j < J$. When the set $\mathcal{Y}$ takes infinitely many values, J is understood as ∞. Consider the lower bound for $E[Y(1)|\boldsymbol{A}, Z = c^+]$ (or equivalently the lower bound of $E[Y(1)|\boldsymbol{A}, Z = c^-] = E[Y|D = 1, Z = c^-]$). Again, the identifiable quantity $E[Y|D = 1, Z = c^+]$ can be expressed as:

$$\begin{aligned}
&E[Y|D = 1, Z = c^+] \\
&\quad = \sum_{j=1}^{J} y_j P(Y = y_j|D = 1, Z = c^+) = \sum_{j=1}^{J} y_j P(Y = y_j|\boldsymbol{A} \cup \boldsymbol{C}, Z = c^+) \\
&\quad = \sum_{j=1}^{J} y_j \frac{P(\boldsymbol{A} \cup \boldsymbol{C}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A} \cup \boldsymbol{C}|Z = c^+)} \\
&\quad = \sum_{j=1}^{J} y_j \left\{ \frac{P(\boldsymbol{A}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A} \cup \boldsymbol{C}|Z = c^+)} \right. \\
&\qquad \left. + \frac{P(\boldsymbol{C}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A} \cup \boldsymbol{C}|Z = c^+)} \right\}, \qquad \text{(A.3)}
\end{aligned}$$

where the first equality holds by the definition of conditional expectation, the second holds because as z approaches c from above, the event $\{D = 1, Z = z\}$ is equivalent to $\{\boldsymbol{A} \cup \boldsymbol{C}, Z = c\}$, the third holds by Bayes' rule, and the fourth holds by the law of total probabilities. The lower bound for $E[Y(1)|\boldsymbol{A}, Z = c^+]$, or equivalently the lower bound of $E[Y|\boldsymbol{A}, Z = c^+]$, is obtained by choosing $P(\boldsymbol{A}|Y = y_j, Z = c^+) \in [0, 1]$ and $P(\boldsymbol{C}|Y = y_j, Z = c^+) \in [0, 1]$ for $j = 1, \cdots, J$, to minimize

$$\sum_{j=1}^{J} y_j P(Y = y_j|\boldsymbol{A}, Z = z^+) = \sum_{j=1}^{J} y_j \frac{P(\boldsymbol{A}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A}|Z = c^+)}$$

subject to Equation (A.3), and by Equation (A.2) and Assumption 2.2, also subject to

$$\sum_{j=1}^{J} P(\boldsymbol{A}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+) = qP(D = 1|Z = c^+).$$

$$\sum_{j=1}^{J} P(\boldsymbol{C}|Y=y_j, Z=c^+)P(Y=y_j|Z=c^+) = (1-q)P(D=1|Z=c^+).$$

The solution to the problem depends on an index $j^* \geq 0$ such that

$$\sum_{j=1}^{j^*} P(Y=y_j|Z=c^+) \leq qP(D=1|Z=c^+), \tag{A.4}$$

but

$$\sum_{j=1}^{j^*+1} P(Y=y_j|Z=c^+) > qP(D=1|Z=c^+)$$

Note that j^* is identifiable from the data. Here we abuse the notation to define $\sum_{j=1}^{0}(\cdot)_j = 0$ to accommodate the case where $j^* = 0$. Then to minimize $E[Y|\boldsymbol{A}, Z=c^+]$, it is clear that we need to set $P(\boldsymbol{A}|Y=y_j, Z=c^+) = 1$ for all $j \leq j^*$, and then set

$$P(\boldsymbol{A}|Y=y_{j^*+1}, Z=c^+) = \frac{qP(D=1|Z=c^+) - \sum_{j=1}^{j^*} P(Y=y_j|Z=c^+)}{P(Y=y_{j^*+1}|Z=c^+)}.$$

Finally, set $P(\boldsymbol{A}|Y=y_j, Z=c^+) = 0$ for all $j > j^*+1$.

In this case, the lower bound for $E[Y|\boldsymbol{A}, Z=c^+]$ is achieved by

$$\begin{aligned} LB^+ &\equiv \sum_{j=1}^{j^*} y_j \frac{P(Y=y_j|Z=c^+)}{P(\boldsymbol{A}|Z=c^+)} + y_{j^*+1} \frac{qP(D=1|Z=c^+) - \sum_{j=1}^{j^*} P(Y=y_j|Z=c^+)}{P(\boldsymbol{A}|Z=c^+)} \\ &= \sum_{j=1}^{j^*} y_j \frac{P(Y=y_j|D=1, Z=c^+)}{q} + y_{j^*+1}\left\{1 - \frac{\sum_{j=1}^{j^*} P(Y=y_j|D=1, Z=c^+)}{q}\right\}, \end{aligned} \tag{A.5}$$

where the second equality holds because

$$\frac{P(Y=y_j|Z=c^+)}{P(\boldsymbol{A}|Z=c^+)} = \frac{P(Y=y_j|Z=c^+)}{qP(D=1|Z=c^+)} = \frac{P(Y=y_j|D=1, Z=c^+)}{q},$$

And this lower bound can be relaxed to fit the same notation as the continuous case. To see this, note

$$LB^+ \geq \frac{\sum_{j=1}^{j^*} y_j P(Y=y_j|D=1, Z=c^+)}{\sum_{j=1}^{j^*} P(Y=y_j|D=1, Z=c^+)} = E[Y|D=1, Y < G_{1L}^{-1}(q), Z=c^+] \tag{A.6}$$

where the inequality holds because $y_{j^*+1} > y_{j^*}$, and the equality holds by the definition of G_{1L}^{-1}. Therefore, $E[Y|D=1, Y \leq G_{1L}^{-1}(q), Z=c^+]$ is a valid lower bound for $E[Y(1)|\boldsymbol{A}, Z=c^+]$.

Likewise, the upper bound for $E[Y(1)|\boldsymbol{A}, Z=c^+]$ or equivalently the upper bound of $E[Y|\boldsymbol{A}, Z=c^+]$, is basically obtained by choosing $P(\boldsymbol{A}|Y=y_j, Z=c^+) \in [0,1]$ and $P(\boldsymbol{C}|Y=y_j, Z=c^+) \in [0,1]$ for

$j = 1, \cdots, J$, to maximize

$$\sum_{j=1}^{J} y_j P(Y = y_j | \boldsymbol{A}, Z = z^+) = \sum_{j=1}^{J} y_j \frac{P(\boldsymbol{A}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A}|Z = c^+)}$$

subject to Equation (A.3) and

$$\sum_{j=1}^{J} P(\boldsymbol{A}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+) = qP(D = 1|Z = c^+).$$

$$\sum_{j=1}^{J} P(\boldsymbol{C}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+) = (1 - q)P(D = 1|Z = c^+).$$

let $j^\dagger \geq 0$ be such that

$$\sum_{j=j^\dagger}^{J} P(Y = y_j|Z = c^+) \leq qP(D = 1|Z = c^+) \tag{A.7}$$

but

$$\sum_{j=j^\dagger-1}^{J} P(Y = y_j|Z = c^+) > qP(D = 1|Z = c^+)$$

To maximize $E[Y|\boldsymbol{A}, Z = c^+]$, it is clear that we need to set $P(\boldsymbol{A}|Y = y_j, Z = c^+) = 1$ for all $j \geq j^\dagger$, and set

$$P(\boldsymbol{A}|Y = y_{j^\dagger-1}, Z = c^+) = \frac{qP(D = 1|Z = c^+) - \sum_{j=j^\dagger}^{J} P(Y = y_j|Z = c^+)}{P(Y = y_{j^\dagger-1}|Z = c^+)}$$

and set $P(\boldsymbol{A}|Y = y_j, Z = c^+) = 0$ for all $j < j^\dagger - 1$.

In this case, the upper bound for $E[Y|\boldsymbol{A}, Z = c^+]$ is achieved by

$$\begin{aligned} UB^+ &\equiv \sum_{j=j^\dagger}^{J} y_j \frac{P(Y = y_j|Z = c^+)}{P(\boldsymbol{A}|Z = c^+)} + y_{j^\dagger-1} \frac{qP(D = 1|Z = c^+) - \sum_{j=j^\dagger}^{J} P(Y = y_j|Z = c^+)}{P(Y = y_{j^\dagger-1}|Z = c^+)} \\ &= \sum_{j=j^\dagger}^{J} y_j \frac{P(Y = y_j|D = 1, Z = c^+)}{q} + y_{j^\dagger-1} \left\{ 1 - \frac{\sum_{j=j^\dagger}^{J} P(Y = y_j|D = 1, Z = c^+)}{q} \right\}. \end{aligned} \tag{A.8}$$

This bound can also be relaxed:

$$UB^+ \leq \frac{\sum_{j=j^\dagger}^{J} y_j P(Y = y_j|D = 1, Z = c^+)}{\sum_{j=j^\dagger}^{J} P(Y = y_j|D = 1, Z = c^+)} = E[Y|D = 1, Y > G_{1U}^{-1}(1 - q), Z = c^+],$$

where the inequality holds because $y_{j^\dagger} > y_{j^\dagger-1}$, and the last equality holds by the definition of G_{1U}^{-1}. Therefore, $E[Y|D = 1, Y > G_{1U}^{-1}(1 - q), Z = c^+]$ is a valid upper bound for $E[Y(1)|\boldsymbol{A}, Z = c^+]$ or equivalently $E[Y|\boldsymbol{A}, Z = c^+]$.

Following the same reasoning, we can derive the sharp bounds for $E[Y|\boldsymbol{N}, Z = c^-]$ as:

$$LB^- = \sum_{j=1}^{j^*} y_j \frac{P(Y = y_j|D = 0, Z = c^-)}{r} + y_{j^*+1}\left\{1 - \frac{\sum_{j=1}^{j^*} P(Y = y_j|D = 1, Z = c^-)}{r}\right\}.$$

$$UB^- = \sum_{j=j^\dagger}^{J} y_j \frac{P(Y = y_j|D = 0, Z = c^-)}{q} + y_{j^\dagger-1}\left\{1 - \frac{\sum_{j=j^\dagger}^{J} P(Y = y_j|D = 0, Z = c^-)}{r}\right\}.$$

where the definitions for j^* and $j^\dagger$ are analogous to those in LB^+ and UB^+.

(iii) Finally, suppose $Y(d)$ is continuous but has possibly mass points. If the q-th quantile is a continuous point, then bounds can be derived following the same argument as in part (i); if a mass point y^* is such that

$$P(Y < y^*|Z = c^+) \leq qP(D = 1|Z = c^+),$$

but

$$P(Y \leq y^*|Z = c^+) > qP(D = 1|Z = c^+)$$

then the lower bound can be derived following the same argument as in part (ii) with y^* playing the role of y_{j^*} as

$$\begin{aligned} LB^+ = E[Y|D = 1, Y < y^*, Z = c^+]\frac{P(Y < y^*|D = 1, Z = c^+)}{q} + y^*\left(1 - \frac{P(Y < y^*|D = 1, Z = c^+)}{q}\right) \\ \geq E[Y|D = 1, Y < y^*, Z = c^+] = E[Y|D = 1, Y < G_{1L}^{-1}(q), Z = c^+]. \end{aligned}$$

If a mass point $y^\dagger$ is such that

$$P(Y > y^\dagger|Z = c^+) \leq qP(D = 1|Z = c^+),$$

but

$$P(Y \geq y^\dagger|Z = c^+) > qP(D = 1|Z = c^+),$$

then the upper bound can be derived following the same argument as in part (ii) with $y^\dagger$ playing the role of $y_{j^\dagger}$, and it is given by

$$\begin{aligned} UB^+ = E[Y|D = 1, Y > y^\dagger, Z = c^+]\frac{P(Y > y^\dagger|D = 1, Z = c^+)}{q} + y^\dagger\left(1 - \frac{P(Y < y^*|D = 1, Z = c^+)}{q}\right) \\ \geq E[Y|D = 1, Y > y^\dagger, Z = c^+] = E[Y|D = 1, Y > G_{1U}^{-1}(1 - q), Z = c^+]. \end{aligned}$$

□

Example 1.2 *Suppose* $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$ *and* $P(Y = y_j | D = 1, Z = c^+) = 0.25$ *for all* j. *Case 1. Suppose* $q = 0.26$. *In this case,* $j^* = 1$ *and the sharp lower bound is given by*

$$LB^+ = \frac{0.25}{0.26} y_1 + \frac{0.01}{0.26} y_2;$$

$j^\dagger = 4$ *and the sharp upper bound is*

$$UB^+ = \frac{0.25}{0.26} y_4 + \frac{0.01}{0.26} y_3.$$

On the other hand, $G_{1L}^{-1}(0.26) = \inf\{y \in \mathcal{Y} : G_1(y) \geq 0.26\} = y_2$*; hence, a valid but non-sharp lower bound is given by:*

$$E[Y|D = 1, Y < G_{1L}^{-1}(0.26), Z = c^+] = E[Y|D = 1, Y < y_2, Z = c^+] = y_1 < LB^+.$$

$G_{1U}^{-1}(1 - 0.26) = \sup\{y \in \mathcal{Y} : G_1(y) \leq 0.74\} = y_3$*; hence, a valid but non-sharp upper bound is given by:*

$$E[Y|D = 1, Y > G_{1U}^{-1}(0.74), Z = c^+] = E[Y|D = 1, Y > y_2, Z = c^+] = y_4 > UB^+.$$

In this case, the relaxed bounds are fairly close to the sharp bounds.

Case 2. Now if $q = 0.5$, *then* $j^* = 2$ *and the sharp lower bound is*

$$LB^+ = \frac{0.25}{0.5} y_1 + \frac{0.25}{0.5} y_2 = \frac{y_1 + y_2}{2};$$

In this case, $G_{1L}^{-1}(0.5) = \inf\{y \in \mathcal{Y} : G_1(y) \geq 0.5\} = y_2$, *and the valid but non-sharp lower bound is given by:*

$$E[Y|D = 1, Y < G_{1L}^{-1}(0.5), Z = c^+] = E[Y|D = 1, Y < y_3, Z = c^+] = y_1 < LB^+.$$

For the upper bound, we see $j^\dagger = 3$ *and the sharp upper bound is*

$$UB^+ = \frac{y_3 + y_4}{2}.$$

In this case, $G_{1U}^{-1}(1 - 0.5) = \sup\{y \in \mathcal{Y} : G_1(y) \leq 0.5\} = y_3$*; hence, a valid but non-sharp upper bound is given by:*

$$E[Y|D = 1, Y > G_{1U}^{-1}(0.5), Z = c^+] = E[Y|D = 1, Y > y_3, Z = c^+] = y_4 > UB^+.$$

Case 3. *In the third case, suppose* $q = 0.24$*, then* $j^* = 0$ *and* $j^\dagger = 4$*.The sharp bounds are given by*

$$LB^+ = y_1; \quad UB^+ = y_4,$$

In this case, the sharp bounds are not informative. On the other hand, $G_{1L}^{-1}(0.24) = \inf\{y \in \mathcal{Y} : G_1(y) \geq 0.24\} = y_1$*. The valid but non-sharp lower bound* $E[Y|D = 1, Y < G_{1L}^{-1}(0.24), Z = c^+] = E[Y|D = 1, Y < y_1, Z = c^+]$ *is not well-defined and hence is understood as* $-\infty$*.* $G_{1U}^{-1}(1 - 0.24) = \sup\{y \in \mathcal{Y} : G_1(y) \leq 0.76\} = y_4$*; hence, the valid but non-sharp upper bound is given by:* $E[Y|D = 1, Y > G_{1U}^{-1}(0.5), Z = c^+] = E[Y|D = 1, Y > y_4, Z = c^+]$*. It is also not well-defined and is understood as* $+\infty$*.*

A.3 Proof of Proposition 3.3

By the results in Appendix B, we have

$$\sqrt{nh}\begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \\ \hat{\theta}_3 - \theta_3 \\ \hat{\theta}_4 - \theta_4 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where $\Omega_{j,k} = \lim_{n\to\infty} h^{-1}E[\phi_{\theta_j,i}\phi_{\theta_k,i}]$ for $j, k = 1, 2, 3, 4$. Also, for $j = 1, 2, 3, 4$, $\phi_{\theta_j,i}$ is either $\phi^c_{\theta_j,i}$ or $\phi^d_{\theta_j,i}$ depending on whether θ_j is a continuous case or discrete case. We also have

$$\sqrt{nh}\begin{pmatrix} \hat{\theta}^w_1 - \hat{\theta}_1 \\ \hat{\theta}^w_2 - \hat{\theta}_2 \\ \hat{\theta}^w_3 - \hat{\theta}_3 \\ \hat{\theta}^w_4 - \hat{\theta}_4 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega),$$

conditional on sample path with probability approaching one. Note that for $j = 1, 2, 3, 4$,

$$\lim_{B\to\infty} \hat{\sigma}^2_j = \lim_{B\to\infty} nhB^{-1}\sum_{b=1}^{B}(\hat{\theta}^{w,b}_j - \hat{\theta}_j)^2 = (nh)^{-1}\sum_{i=1}^{n}\phi^2_{\theta_j,i} + o_p(1) \xrightarrow{p} \sigma^2_j.$$

Then we can apply the results in Donald and Hsu (2011) to show Proposition A.3 and we omit the details.
□

B Useful Lemmas

In this section, we provide regularity conditions, and show the asymptotic normality of the proposed estimator $\hat{\theta}$ and the validity of the weighted bootstrap. We focus on the θ_1 case and will briefly summarize the results for θ_2, θ_3 and θ_4.

Assumption B.1 *Assume that* $0 < q < 1$.

Assumption B.2 *Assume that density* $f_z(z)$ *is twice continuously differentiable in* z *on* $(c-\epsilon, c+\epsilon)$ *and* $\delta \leq f_z(z) \leq M$ *on* $(c-\epsilon, c+\epsilon)$ *for some* $\epsilon > 0$ *and* $0 < \delta < M$.

Assumption B.3 *Assume that for the same* ϵ *and* M *in Assumption B.2,*

1. $E[D|Z=z]$ *is twice continuously differentiable on* $z \in (c-\epsilon, c)$ *with absolute values of corresponding derivatives bounded by* M*;*
2. $E[D|Z=z]$ *is twice continuously differentiable on* $z \in [c, c+\epsilon)$ *with absolute values of corresponding derivatives bounded by* M.

Assumption B.4 *Assume that for the same* ϵ *and* M *in Assumption B.2, for* $d = 0$ *and* 1,

1. $E[Y|D=d, Z=z]$ *is twice continuously differentiable on* $z \in (c-\epsilon, c)$ *with absolute values of corresponding derivatives bounded by* M*;*
2. $E[Y|D=d, Z=z]$ *is twice continuously differentiable on* $z \in (c, c+\epsilon)$ *with absolute values of corresponding derivatives bounded by* M.
3. $E[|Y|^3|D=d, Z=z] \leq M$ *for* $z \in (c-\epsilon, c+\epsilon)$.

Assumption B.5 *Assume that*

1. *The kernel function* $K(\cdot)$ *is a non-negative symmetric bounded kernel with support* $[-1, 1]$*;* $\int K(u)du = 1$.
2. *The bandwidth* h *satisfies that* $h \to 0$, $nh^3 \to \infty$, *and* $nh^5 \to 0$ *as* $n \to \infty$.

Assumption B.6 *(Continuous Case) Assume that* $G_1(y)$ *is continuous on* $(G_{1L}^{-1}(q)-\delta, G_{1L}^{-1}(q)+\delta)$ *with* $G_1(G_{1L}^{-1}(q)) = q$ *and the derivative of* $G_1(y)$ *is greater than* δ *for the same delta in Assumption B.2. In addition, assume that for the same* ϵ, ϵ *and* M *in Assumption B.2, for all* $y \in (G_{1L}^{-1}(q)-\delta, G_{1L}^{-1}(q)+\delta)$, $E[DY1(Y \leq y)|Z=z]$ *and* $E[D1(Y \leq y)|Z=z]$ *are twice continuously differentiable on* $z \in (c, c+\epsilon)$ *with absolute values of corresponding derivatives bounded by* M.

Assumption B.6' (Discrete Case) Assume that $y_{1L,\ell} < y_{1L,u}$ with $G_1(y_{1L,\ell}) < q < G_1(y_{1L,u})$ and $\lim_{z \downarrow c} P(Y \in (y_{1L,\ell}, y_{1L,u})|D=1, Z=z) = 0$.

Assumption B.7 *Assume that* $\{W_i\}_{i=1}^n$ *is a sequence of i.i.d. pseudo random variables independent of the sample path with* $E[W_i] = Var[W_i] = 1$ *for all* i.

Lemma B.1 *Suppose that Assumptions B.1-B.6 hold. Then*

$$\sqrt{nh}(\hat{\theta}_1 - \theta_1) \equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi^c_{\theta_1,i} + o_p(1), \tag{B.1}$$

where $\phi^c_{\theta_1,i}$ *is given in (B.5). Also,* $\sqrt{nh}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V^c_{\theta_1})$*, where* $V^c_{\theta_1} = \lim_{n\to} \frac{1}{h} E[(\phi^c_{\theta_1,i})^2]$.

Lemma B.2 *Suppose that Assumptions B.1-B.6 and Assumption B.7 hold. Then*

$$\sqrt{nh}(\hat{\theta}^w_1 - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (W_i - 1)\phi^c_{\theta_1,i} + o_p(1), \tag{B.2}$$

where $\phi^c_{\theta_1,i}$ *is given in (B.5). Also,* $\sqrt{nh}(\hat{\theta}^w_1 - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V^c_{\theta_1})$ *conditional on the sample path with probability approaching 1.*

Lemma B.1' Suppose that Assumption B.6' is in place of Assumption B.6 in Lemma B.1. Then

$$\sqrt{nh}(\hat{\theta}_1 - \theta_1) \equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi^d_{\theta_1,i} + o_p(1), \tag{B.3}$$

where $\phi^d_{\theta_1,i}$ is given in (B.6). Also, $\sqrt{nh}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V^d_{\theta_1})$, where $V^d_{\theta_1} = \lim_{n\to} \frac{1}{h} E[(\phi^d_{\theta_1,i})^2]$.

Lemma B.2' Suppose that Assumption B.6' is in place of Assumption B.6 in Lemma B.2. Then

$$\sqrt{nh}(\hat{\theta}^w_1 - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (W_i - 1)\phi^d_{\theta_1,i} + o_p(1), \tag{B.4}$$

where $\phi^d_{\theta_1,i}$ is given in (B.5). Also, $\sqrt{nh}(\hat{\theta}^w_1 - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V^d_{\theta_1})$ conditional on the sample path with probability approaching 1.

Let

$$\Delta_z = f_z(0) \cdot \begin{pmatrix} \mu_{z,0} & \mu_{z,1} \\ \mu_{z,1} & \mu_{z,2} \end{pmatrix} \text{ with } \mu_{z,j} = \int_{u \geq 0} u^j K(u) du, \text{ for } j = 0, 1, 2.$$

For a general random variable X_i, let

$$(\widehat{E}[X|Z = c^+], \hat{\beta}^+_x) = \arg\min_{a,b} \sum_{i=1}^{n} 1(Z_i \geq c) K\left(\frac{Z_i - c}{h}\right) \Big[X_i - a - bZ_i\Big]^2,$$
$$(\widehat{E}[X|Z = c^-], \hat{\beta}^-_x) = \arg\min_{a,b} \sum_{i=1}^{n} 1(Z_i < c) K\left(\frac{Z_i - c}{h}\right) \Big[X_i - a - bZ_i\Big]^2.$$

Suppose that $E[X|Z=z]$ is twice continuously differentiable on $z \in (c-\epsilon, c)$ and $z \in [c, c+\epsilon)$ with absolute values of corresponding derivatives bounded by M. Also, $E[|X|^3|D=d, Z=z] \leq M$ for $z \in (c-\epsilon, c+\epsilon)$.

Then by Chiang, Hsu, and Sasaki (2019) and Hsu and Shen (2022), we have

$$
\begin{aligned}
&\sqrt{nh}\left(\widehat{E}[X|Z=c^+] - E[X|Z=c^+]\right) \\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(1\ 0)\Delta_z^{-1}1(Z_i \geq c)K\left(\frac{Z_i-c}{h}\right)(X_i - E[X_i|Z_i])\begin{pmatrix} 1 \\ \frac{Z_i-c}{h}\end{pmatrix} + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{X,i}^{+} + o_p(1), \\
&\sqrt{nh}\left(\widehat{E}[X|Z=c^-] - E[X|Z=c^-]\right) \\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(1\ 0)\Delta_z^{-1}1(Z_i < c)K\left(\frac{Z_i-c}{h}\right)(X_i - E[X_i|Z_i])\begin{pmatrix} 1 \\ \frac{Z_i-c}{h}\end{pmatrix} + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{X,i}^{-} + o_p(1).
\end{aligned}
$$

Also, define $H_{1,\leq}(y) = E[DY1(Y \leq y)|Z=c^+]$, $H_{1,\geq}(y) = E[DY1(Y \geq y)|Z=c^+]$, $H_{0,\leq}(y) = E[(1-D)Y1(Y \leq y)|Z=c^-]$ and $H_0(y,\geq) = E[(1-D)Y1(Y \geq y)|Z=c^-]$. Let $I_{1,\leq}(y) = E[D1(Y \leq y)|Z=c^+]$, $I_1(y,\geq) = E[D1(Y \geq y)|Z=c^+]$, $I_{0,\leq}(y) = E[(1-D)1(Y \leq y)|Z=c^-]$ and $I_{0,\geq}(y) = E[(1-D)1(Y \geq y)|Z=c^-]$.

Proof of Lemma B.1: Recall that

$$
\hat{q} = \frac{\widehat{E}[D|Z=c^-]}{\widehat{E}[D|Z=c^+]}
$$

Then by delta method, we have

$$
\begin{aligned}
\sqrt{nh}\,(\hat{q}-q) &= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{1}{E[D|Z=c^+]}\phi_{d,i}^{-} - \frac{q}{E[D|Z=c^+]}\phi_{d,i}^{+} + o_p(1) \\
&\equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{q,i} + o_p(1).
\end{aligned}
$$

Similarly,

$$
\widehat{G}_1(y) = \frac{\widehat{E}[D1(Y \leq y)|Z=c^+]}{\widehat{E}[D|Z=c^+]}.
$$

Because $\{1(Y \leq y) : y \in R\}$ is a Vapnik-Chervonenkis (VC) class of functions, we have uniformly over

$y \in R$,

$$\sqrt{nh}\left(\widehat{G}_1(y) - G_1(y)\right) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{1}{E[D|Z=c^+]}\phi^+_{D1(Y\leq y),i} - \frac{G_1(y)}{E[D|Z=c^+]}\phi^+_{D,i} + o_p(1)$$
$$\equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{G_1(y),i} + o_p(1).$$

In this case, $G^{-1}_{1L}(q)$ is differentiable and its derivative with respective to q is $g_1(G^{-1}_{1L}(q))$. Then by functional delta method, we have

$$\begin{aligned}
&\sqrt{nh}\left(\widehat{G}^{-1}_{1L}(\hat{q}) - G^{-1}_{1L}(q)\right) \\
=&\sqrt{nh}\left(\widehat{G}^{-1}_{1L}(\hat{q}) - G^{-1}_{1L}(\hat{q})\right) + \sqrt{nh}\left(G^{-1}_{1L}(\hat{q}) - G^{-1}_{1L}(q)\right) \\
=&\sqrt{nh}\left(\widehat{G}^{-1}_{1L}(q) - G^{-1}_{1L}(q)\right) + o_p(1) + g_1(G^{-1}_{1L}(q))\sqrt{nh}\left(\hat{q} - q\right) + o_p(1) \\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{-1}{g_1(G^{-1}_{1L}(q))}\phi_{G_1(G^{-1}_{1L}(q)),i} + g_1(G^{-1}_{1L}(q))\phi_{q,i} + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{G^{-1}_{1L}(q),i} + o_p(1).
\end{aligned}$$

Note that under Assumption B.6, $G_1(y)$ is continuous in a neighborhood of $G^{-1}_{1L}(q)$, so that $E[DY1(Y < G^{-1}_{1L}(q))|Z = c^+] = E[DY1(Y \leq G^{-1}_{1L}(q))|Z = c^+] =$ and $\widehat{E}[DY1(Y < \widehat{G}^{-1}_{1L}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}[DY1(Y \leq \widehat{G}^{-1}_{1L}(\hat{q}))|Z = c^+]$. Also, $E[D1(Y < G^{-1}_{1L}(q))|Z = c^+] = E[D1(Y \leq G^{-1}_{1L}(q))|Z = c^+] =$ and $\widehat{E}[D1(Y < \widehat{G}^{-1}_{1L}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}[D1(Y \leq \widehat{G}^{-1}_{1L}(\hat{q}))|Z = c^+]$.

We have $dH_{1,\leq}(y)/dy = P_{1|1}\cdot y\cdot g_1(y)$ and $dI_{1,\leq}(y)/dy = P_{1|1}\cdot g_1(y)$. Then by the delta method, we have

$$\begin{aligned}
&\sqrt{nh}(\widehat{H}_{1,\leq}(\widehat{G}^{-1}_{1L}(\hat{q})) - H_{1,\leq}(G^{-1}_{1L}(q))) \\
=&\sqrt{nh}(\widehat{H}_{1,\leq}(\widehat{G}^{-1}_{1L}(\hat{q})) - H_{1,\leq}(\widehat{G}^{-1}_{1L}(\hat{q}))) + \sqrt{nh}(H_{1,\leq}(\widehat{G}^{-1}_{1L}(\hat{q})) - H_{1,\leq}(G^{-1}_{1L}(q))) \\
=&\sqrt{nh}(\widehat{H}_{1,\leq}(G^{-1}_{1L}(q)) - H_{1,\leq}(G^{-1}_{1L}(q))) + o_p(1) + \sqrt{nh}(H_{1,\leq}(\widehat{G}^{-1}_{1L}(\hat{q})) - H_{1,\leq}(G^{-1}_{1L}(\hat{q}))) \\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^+_{DY1(Y\leq G^{-1}_{1L}(q)),i} + P_{1|1}\cdot G^{-1}_{1L}(q)\cdot g_1(G^{-1}_{1L}(q))\phi_{G^{-1}_{1L}(q),i} + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{1,\leq}(G^{-1}_{1L}(q)),i} + o_p(1), \\
&\sqrt{nh}(\widehat{I}_{1,\leq}(\widehat{G}^{-1}_{1L}(\hat{q})) - I_{1,\leq}(G^{-1}_{1L}(q))) \\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^+_{D1(Y\leq G^{-1}_{1L}(q)),i} + P_{1|1}\cdot g_1(G^{-1}_{1L}(q))\phi_{G^{-1}_{1L}(q),i} + o_p(1)
\end{aligned}$$

$$\equiv\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{I_{1,\leq(G_{1L}^{-1}(q)),i}}+o_p(1).$$

To derive the asymptotics of $\hat{\theta}_1$, note that

$$\begin{aligned}
&\sqrt{nh}(\hat{\theta}_1-\theta_1)\\
=&\sqrt{nh}(\widehat{H}_{1,\leq}(\widehat{G}_{1L}^{-1}(\hat{q}))\cdot\widehat{E}[D|Z=c^-]-\widehat{I}_{1,\leq}(\widehat{G}_{1L}^{-1}(\hat{q}))\cdot\widehat{E}[DY|Z=c^-]\\
&\quad-H_{1,\leq}(G_{1L}^{-1}(q))\cdot E[D|Z=c^-]+I_{1,\leq}(G_{1L}^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\sqrt{nh}(\widehat{H}_{1,\leq}(\widehat{G}_{1L}^{-1}(\hat{q}))\cdot\widehat{E}[D|Z=c^-]-H_{1,\leq}(G_{1L}^{-1}(q))\cdot E[D|Z=c^-])\\
&\quad-\sqrt{nh}(\widehat{I}_{1,\leq}(\widehat{G}_{1L}^{-1}(\hat{q}))\cdot\widehat{E}[DY|Z=c^-]-I_{1,\leq}(G_{1L}^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}E[D|Z=c^-]\phi_{H_{1,\leq(G_{1L}^{-1}(q)),i}}+H_{1,\leq}(G_{1L}^{-1}(q))\phi_{D,i}^-\\
&\quad-E[DY|Z=c^-]\phi_{I_{1,\leq(G_{1L}^{-1}(q)),i}}-I_{1,\leq}(G_{1L}^{-1}(q))\phi_{DY,i}^-+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{\theta_1,i}^{c}+o_p(1).
\end{aligned}\tag{B.5}$$

Then by central limit theorem, we have $\sqrt{nh}(\hat{\theta}_1-\theta_1)\xrightarrow{d}\mathcal{N}(0,V_{\theta_1}^{c})$. This completes the proof.□

Proof of Lemma B.2: Note that by the same arguments for Theorem 5.2 of Hsu and Shen (2022) and the arguments for Lemma B.1, we can show that

$$\sqrt{nh}(\hat{\theta}_1^{w}-\theta_1)\equiv\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}W_i\phi_{\theta_1,i}^{c}+o_p(1),$$

and it follows that

$$\sqrt{nh}(\hat{\theta}_1^{w}-\hat{\theta}_1)\equiv\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(W_i-1)\phi_{\theta_1,i}^{c}+o_p(1).$$

In the last step, note that W_i-1 has a mean of zero and variance of one, so we can apply the multiplier bootstrap arguments in Chiang, Hsu, and Sasaki (2019) and obtain that $\sqrt{nh}(\hat{\theta}_1^{w}-\hat{\theta}_1)\xrightarrow{d}\mathcal{N}(0,V_{\theta_1}^{d})$ conditional on the sample path with probability approaching 1. This completes the proof.□

Proof of Lemma B.1': Assumption B.6' assumes that $y_{1L,\ell}<y_{1L,u}$ with $G_1(y_{1L,\ell})<q<G_1(y_{1L,u})$ and $\lim_{z\downarrow c}P(Y\in(y_{1L,\ell},y_{1L,u})|D=1,Z=z)=0$. In this case, we have $G_{1L}^{-1}(q))=y_{1L,u}$. Therefore, it is true that $\widehat{G}_1(y_{1L,\ell})<\hat{q}<\widehat{G}_1(y_{1L,u})$ with probability approaching one and this implies that $\widehat{G}_{1L}^{-1}(\hat{q}))=y_{1L,u}$ with probability approaching one. That is, we have $\sqrt{nh}(\widehat{G}_{1L}^{-1}(\hat{q}))-y_{1L,u})=o_p(1)$. In addition, $E[DY1(Y<y_{1L,u})|Z=c^+]=E[DY1(Y\leq y_{1L,\ell})|Z=c^+]$, and $\widehat{E}[DY1(Y<\widehat{G}_{1L}^{-1}(\hat{q}))|Z=c^+]$ is asymptotically equivalent to $\widehat{E}[DY1(Y<y_{1L,u})|Z=c^+]=\widehat{E}[DY1(Y\leq y_{1L,\ell})|Z=c^+]$. Sim-

ilarly, $E[D1(Y < y_{1L,u})|Z = c^+] = E[D1(Y \leq y_{1L,\ell})|Z = c^+]$, and $\widehat{E}[D1(Y < \widehat{G}_{1L}^{-1}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}[D1(Y < y_{1L,u})|Z = c^+] = \widehat{E}[D1(Y \leq y_{1L,\ell})|Z = c^+]$. Because $\sqrt{nh}(\widehat{G}_{1L}^{-1}(\hat{q})) - y_{1L,u}) = o_p(1)$, the estimation effect of $\widehat{G}_{1L}^{-1}(\hat{q}))$ will be asymptotically negligible. Therefore,

$$\sqrt{nh}(\widehat{H}_{1,\leq}(y_{1L,\ell}) - H_{1,\leq}(y_{1L,\ell})) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{+}_{DY1(Y\leq y_{1L,\ell}),i} + o_p(1) \equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{1,\leq}(y_{1L,\ell}),i} + o_p(1),$$

$$\sqrt{nh}(\widehat{I}_{1,\leq}(y_{1L,\ell}) - I_{1,\leq}(y_{1L,\ell})) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{+}_{D1(Y\leq y_{1L,\ell}),i} + o_p(1) \equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{I_{1,\leq}(y_{1L,\ell}),i} + o_p(1).$$

As a result, in this case,

$$\begin{aligned}
&\sqrt{nh}(\hat{\theta}_1 - \theta_1)\\
=&\sqrt{nh}(\widehat{H}_{1,\leq}(y_{1L,\ell}) \cdot \widehat{E}[D|Z = c^-] - \widehat{I}_{1,\leq}(y_{1L,\ell}) \cdot \widehat{E}[DY|Z = c^-]\\
&\quad - H_{1,\leq}(y_{1L,\ell}) \cdot E[D|Z = c^-] + I_{1,\leq}(y_{1L,\ell}) \cdot E[DY|Z = c^-])\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} E[D|Z = c^-]\phi_{H_{1,\leq}(y_{1L,\ell}),i} + H_{1,\leq}(y_{1L,\ell})\phi^{-}_{D,i}\\
&\quad - E[DY|Z = c^-]\phi_{I_{1,\leq}(y_{1L,\ell}),i} - I_{1,\leq}(y_{1L,\ell})\phi^{-}_{DY,i} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{d}_{\theta_1,i} + o_p(1).
\end{aligned} \tag{B.6}$$

Then by the central limit theorem, we have $\sqrt{nh}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V^d_{\theta_1})$. This completes the proof.□

Proof of Lemma B.2': The proof is similar to that for Lemma B.2, so we omit the details.□

To conclude this section, we provide the influence function representations for $\hat{\theta}_2$, $\hat{\theta}_3$ and $\hat{\theta}_4$. For brevity, we do not write down the regularity conditions for these estimators because they are similar to those for θ_1 case.

Let $H_{1,\geq}(y) = E[DY1(Y \geq y)|Z = c^+]$, $H_{0,\leq}(y) = E[(1-D)Y1(Y \leq y)|Z = c^-]$ and $H_0(y, \geq) = E[(1-D)Y1(Y \geq y)|Z = c^-]$. Let $I_1(y, \geq) = E[D1(Y \geq y)|Z = c^+]$, $I_{0,\leq}(y) = E[(1-D)1(Y \leq y)|Z = c^-]$ and $I_{0,\geq}(y) = E[(1-D)1(Y \geq y)|Z = c^-]$. In addition, for the continuous case, We have $dH_{1,\geq}(y)/dy = -P_{1|1} \cdot y \cdot g_1(y)$, $dH_{0,\leq}(y)/dy = P_{0|0} \cdot y \cdot g_0(y)$, $dH_{0,\geq}(y)/dy = -P_{0|0} \cdot y \cdot g_0(y)$, $dI_{1,\geq}(y)/dy = -P_{1|1} \cdot g_1(y)$, $dI_{0,\leq}(y)/dy = P_{0|0} \cdot g_0(y)$, and $dI_{0,\geq}(y)/dy = -P_{0|0} \cdot g_0(y)$.

Recall that

$$\hat{r} = \frac{\widehat{E}[1 - D|Z = c^+]}{\widehat{E}[1 - D|Z = c^-]}.$$

Then by delta method, we have

$$\begin{aligned}\sqrt{nh}\,(\hat{r}-r) &= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{1}{E[1-D|Z=c^{-}]}\phi^{+}_{1-d,i}-\frac{r}{E[1-D|Z=c^{-}]}\phi^{-}_{1-d,i}+o_p(1)\\ &\equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{r,i}+o_p(1).\end{aligned}$$

Similarly,

$$\widehat{G}_0(y)=\frac{\widehat{E}[(1-D)1(Y\leq y)|Z=c^{-}]}{\widehat{E}[1-D|Z=c^{-}]}$$

and

$$\begin{aligned}\sqrt{nh}\left(\widehat{G}_0(y)-G_0(y)\right) &= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{1}{E[1-D|Z=c^{-}]}\phi^{-}_{(1-D)1(Y\leq y),i}-\frac{G_0(y)}{E[1-D|Z=c^{-}]}\phi^{-}_{1-D,i}+o_p(1)\\ &\equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{G_0(y),i}+o_p(1).\end{aligned}$$

For the continuous case, we have

$$\begin{aligned}&\sqrt{nh}\left(\widehat{G}^{-1}_{1U}(1-\hat{q})-G^{-1}_{1U}(1-q)\right)\\ =&\sqrt{nh}\left(\widehat{G}^{-1}_{1U}(1-q)-G^{-1}_{1U}(1-q)\right)+o_p(1)-g_1(G^{-1}_{1U}(1-q))\sqrt{nh}\,(\hat{q}-q)+o_p(1)\\ =&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{-1}{g_1(G^{-1}_{1U}(1-q))}\phi_{G_1(G^{-1}_{1U}(1-q)),i}-g_1(G^{-1}_{1U}(1-q))\phi_{q,i}+o_p(1)\\ \equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{G^{-1}_{1U}(1-q),i}+o_p(1),\\ &\sqrt{nh}\left(\widehat{G}^{-1}_{0L}(\hat{r})-G^{-1}_{0L}(r)\right)\\ =&\sqrt{nh}\left(\widehat{G}^{-1}_{0L}(r)-G^{-1}_{0L}(r)\right)+o_p(1)+g_0(G^{-1}_{0L}(r))\sqrt{nh}\,(\hat{r}-r)+o_p(1)\\ =&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{-1}{g_0(G^{-1}_{0L}(r))}\phi_{G_0(G^{-1}_{0L}(r)),i}+g_0(G^{-1}_{0L}(r))\phi_{r,i}+o_p(1)\\ \equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{G^{-1}_{0L}(r),i}+o_p(1),\\ &\sqrt{nh}\left(\widehat{G}^{-1}_{0U}(1-\hat{r})-G^{-1}_{0U}(1-r)\right)\\ =&\sqrt{nh}\left(\widehat{G}^{-1}_{0U}(1-r)-G^{-1}_{0U}(1-r)\right)+o_p(1)-g_0(G^{-1}_{0U}(1-r))\sqrt{nh}\,(\hat{r}-r)+o_p(1)\\ =&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{-1}{g_0(G^{-1}_{0U}(1-r))}\phi_{G_0(G^{-1}_{0U}(1-r)),i}-g_0(G^{-1}_{0U}(1-r))\phi_{r,i}+o_p(1)\\ \equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{G^{-1}_{0U}(1-r),i}+o_p(1).\end{aligned}$$

Next,

$$
\begin{aligned}
&\sqrt{nh}(\widehat{H}_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q}))-H_{1,\geq}(G_{1U}^{-1}(1-q)))\\
=&\sqrt{nh}(\widehat{H}_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q}))-H_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q})))+\sqrt{nh}(H_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q}))-H_{1,\geq}(G_{1U}^{-1}(1-q)))\\
=&\sqrt{nh}(\widehat{H}_{1,\geq}(G_{1U}^{-1}(1-q))-H_{1,\geq}(G_{1U}^{-1}(1-q)))+o_p(1)+\sqrt{nh}(H_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q}))-H_{1,\geq}(G_{1U}^{-1}(1-q)))\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{+}_{DY1(Y\leq G_{1U}^{-1}(1-q)),i}+P_{1|1}\cdot G_{1U}^{-1}(q)\cdot g_1(G_{1U}^{-1}(1-q))\phi_{G_{1U}^{-1}(1-q),i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{1,\geq}(G_{1U}^{-1}(1-q)),i}+o_p(1)\\
&\sqrt{nh}(\widehat{H}_{0,\leq}(\widehat{G}_{0L}^{-1}(\hat{r}))-H_{0,\leq}(G_{0L}^{-1}(r)))\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{-}_{(1-D)Y1(Y\leq G_{0L}^{-1}(r)),i}+P_{0|0}\cdot G_{0L}^{-1}(r)\cdot g_0(G_{0L}^{-1}(r))\phi_{G_{0L}^{-1}(r),i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{0,\leq}(G_{0L}^{-1}(r)),i}+o_p(1),\\
&\sqrt{nh}(\widehat{H}_{0,\geq}(\widehat{G}_{0U}^{-1}(1-\hat{r}))-H_{1,\geq}(G_{0U}^{-1}(1-r)))\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{-}_{(1-D)Y1(Y\leq G_{1U}^{-1}(1-r)),i}+P_{0|0}\cdot G_{0U}^{-1}(r)\cdot g_0(G_{1U}^{-1}(1-r))\phi_{G_{0U}^{-1}(1-r),i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{0,\geq}(G_{0U}^{-1}(1-r)),i}+o_p(1),
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
&\sqrt{nh}(\widehat{I}_{1,\leq}(\widehat{G}_{1L}^{-1}(\hat{q}))-I_{1,\leq}(G_{1L}^{-1}(q)))\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{+}_{D1(Y\leq G_{1L}^{-1}(q)),i}+P_{1|1}\cdot g_1(G_{1L}^{-1}(q))\phi_{G_{1L}^{-1}(q),i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{I_{1,\leq}(G_{1L}^{-1}(q)),i}+o_p(1),\\
&\sqrt{nh}(\widehat{I}_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q}))-I_{1,\geq}(G_{1U}^{-1}(1-q)))\\
=&\sqrt{nh}(\widehat{I}_{1,\geq}(G_{1U}^{-1}(1-q))-I_{1,\geq}(G_{1U}^{-1}(1-q)))+o_p(1)+\sqrt{nh}(I_{1,\geq}(\widehat{G}_{1U}^{-1}(1-\hat{q}))-I_{1,\geq}(G_{1U}^{-1}(1-q)))\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{+}_{D1(Y\leq G_{1U}^{-1}(1-q)),i}+P_{1|1}\cdot g_1(G_{1U}^{-1}(1-q))\phi_{G_{1U}^{-1}(1-q),i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{I_{1,\geq}(G_{1U}^{-1}(1-q)),i}+o_p(1),\\
&\sqrt{nh}(\widehat{I}_{0,\leq}(\widehat{G}_{0L}^{-1}(\hat{r}))-I_{0,\leq}(G_{0L}^{-1}(r)))\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{-}_{(1-D)1(Y\leq G_{0L}^{-1}(r)),i}+P_{0|0}\cdot g_0(G_{0L}^{-1}(r))\phi_{G_{0L}^{-1}(r),i}+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&\equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{I_{0,\le}(G_{0L}^{-1}(r)),i} + o_p(1), \\
&\sqrt{nh}(\widehat{I}_{0,\ge}(\widehat{G}_{0U}^{-1}(1-\hat{r})) - I_{1,\ge}(G_{0U}^{-1}(1-r))) \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi^{-}_{(1-D)1(Y\le G_{1U}^{-1}(1-r)),i} + P_{0|0} \cdot g_0(G_{1U}^{-1}(1-r))\phi_{G_{0U}^{-1}(1-r),i} + o_p(1) \\
&\equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{I_{0,\ge}(G_{0U}^{-1}(1-r)),i} + o_p(1).
\end{aligned}
$$

Finally, for the continuous case, we have

$$
\begin{aligned}
&\sqrt{nh}(\hat{\theta}_2 - \theta_2) \\
&= \sqrt{nh}(\widehat{I}_{1,\ge}(\widehat{G}_{1U}^{-1}(1-\hat{q})) \cdot \widehat{E}[DY|Z=c^-] - \widehat{H}_{1,\ge}(\widehat{G}_{1U}^{-1}(1-\hat{q})) \cdot \widehat{E}[D|Z=c^-] \\
&\qquad - I_{1,\ge}(G_{1U}^{-1}(1-q)) \cdot E[DY|Z=c^-] + H_{1,\ge}(G_{1U}^{-1}(1-q)) \cdot E[D|Z=c^-]) \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} E[DY|Z=c^-]\phi_{I_{1,\ge}(G_{1U}^{-1}(1-q)),i} - I_{1,\ge}(G_{1U}^{-1}(1-q))\phi^{-}_{DY,i} \\
&\qquad - E[D|Z=c^-]\phi_{H_{1,\ge}(G_{1U}^{-1}(1-q)),i} + H_{1,\ge}(G_{1U}^{-1}(1-q))\phi^{-}_{D,i} + o_p(1) \\
&\equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi^{c}_{\theta_2,i} + o_p(1), \\
&\sqrt{nh}(\hat{\theta}_3 - \theta_3) \\
&= \sqrt{nh}(\widehat{H}_{0,\le}(\widehat{G}_{0L}^{-1}(\hat{r})) \cdot \widehat{E}[1-D|Z=c^+] - \widehat{I}_{0,\le}(\widehat{G}_{0L}^{-1}(\hat{r})) \cdot \widehat{E}[(1-D)Y|Z=c^+] \\
&\qquad - H_{0,\le}(G_{0L}^{-1}(r)) \cdot E[1-D|Z=c^+] + I_{1,\le}(G_{1L}^{-1}(q)) \cdot E[(1-D)Y|Z=c^+]) \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} E[(1-D)|Z=c^+]\phi_{H_{0,\le}(G_{0L}^{-1}(r)),i} + H_{0,\le}(G_{0L}^{-1}(r))\phi^{+}_{1-D,i} \\
&\qquad - E[(1-D)Y|Z=c^+]\phi_{I_{0,\le}(G_{0L}^{-1}(r)),i} - I_{0,\le}(G_{0L}^{-1}(r))\phi^{+}_{(1-D)Y,i} + o_p(1) \\
&\equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi^{c}_{\theta_3,i} + o_p(1), \\
&\sqrt{nh}(\hat{\theta}_4 - \theta_4) \\
&= \sqrt{nh}(\widehat{I}_{0,\ge}(\widehat{G}_{0U}^{-1}(1-\hat{r})) \cdot \widehat{E}[(1-D)Y|Z=c^+] - \widehat{H}_{0,\ge}(\widehat{G}_{0U}^{-1}(1-\hat{r})) \cdot \widehat{E}[1-D|Z=c^+] \\
&\qquad - I_{0,\ge}(G_{0U}^{-1}(1-r)) \cdot E[(1-D)Y|Z=c^+] + H_{0,\ge}(G_{0U}^{-1}(1-r)) \cdot E[1-D|Z=c^+]) \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} E[DY|Z=c^+]\phi_{I_{0,\ge}(G_{0U}^{-1}(1-r)),i} - I_{0,\ge}(G_{0U}^{-1}(1-r))\phi^{+}_{(1-D)Y,i} \\
&\qquad - E[D|Z=c^+]\phi_{H_{0,\ge}(G_{0U}^{-1}(1-r)),i} + H_{0,\ge}(G_{0U}^{-1}(1-r))\phi^{+}_{1-D,i} + o_p(1) \\
&\equiv \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi^{c}_{\theta_4,i} + o_p(1).
\end{aligned}
$$

For the discrete case, we have

$$
\begin{aligned}
&\sqrt{nh}(\hat{\theta}_2-\theta_2)\\
=&\sqrt{nh}(\widehat{I}_{1,\geq}(y_{1U,u})\cdot\widehat{E}[DY|Z=c^-]-\widehat{H}_{1,\geq}(y_{1U,u})\cdot\widehat{E}[D|Z=c^-]\\
&\qquad -I_{1,\geq}(y_{1U,u})\cdot E[DY|Z=c^-]+H_{1,\geq}(y_{1U,u})\cdot E[D|Z=c^-])\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}E[DY|Z=c^-]\phi_{I_{1,\geq}(y_{1U,u}),i}-I_{1,\geq}(y_{1U,u})\phi^-_{DY,i}\\
&\qquad -E[D|Z=c^-]\phi_{H_{1,\geq}(y_{1U,u}),i}+H_{1,\geq}(y_{1U,u})\phi^-_{D,i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^d_{\theta_2,i}+o_p(1),\\
&\sqrt{nh}(\hat{\theta}_3-\theta_3)\\
=&\sqrt{nh}(\widehat{H}_{0,\leq}(y_{0L,\ell})\cdot\widehat{E}[1-D|Z=c^+]-\widehat{I}_{0,\leq}(y_{0L,\ell})\cdot\widehat{E}[(1-D)Y|Z=c^+]\\
&\qquad -H_{0,\leq}(y_{0L,\ell})\cdot E[1-D|Z=c^+]+I_{1,\leq}(y_{0L,\ell})\cdot E[(1-D)Y|Z=c^+])\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}E[(1-D)|Z=c^+]\phi_{H_{0,\leq}(y_{0L,\ell}),i}+H_{0,\leq}(y_{0L,\ell})\phi^+_{1-D,i}\\
&\qquad -E[(1-D)Y|Z=c^+]\phi_{I_{0,\leq}(y_{0L,\ell}),i}-I_{0,\leq}(y_{0L,\ell})\phi^+_{(1-D)Y,i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^d_{\theta_3,i}+o_p(1),\\
&\sqrt{nh}(\hat{\theta}_4-\theta_4)\\
=&\sqrt{nh}(\widehat{I}_{0,\geq}(y_{0U,u})\cdot\widehat{E}[(1-D)Y|Z=c^+]-\widehat{H}_{0,\geq}(y_{0U,u})\cdot\widehat{E}[1-D|Z=c^+]\\
&\qquad -I_{0,\geq}(y_{0U,u})\cdot E[(1-D)Y|Z=c^+]+H_{0,\geq}(y_{0U,u})\cdot E[1-D|Z=c^+])\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}E[DY|Z=c^+]\phi_{I_{0,\geq}(y_{0U,u}),i}-I_{0,\geq}(y_{0U,u})\phi^+_{(1-D)Y,i}\\
&\qquad -E[D|Z=c^+]\phi_{H_{0,\geq}(y_{0U,u}),i}+H_{0,\geq}(y_{0U,u})\phi^+_{1-D,i}+o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^d_{\theta_4,i}+o_p(1).
\end{aligned}
$$

C Additional Simulation and Empirical Results

Table 7: Rejection Frequency at 1% Level

		Undersmoothing			MSE-RBC			CER-RBC		
	n	IK	CCT	AI	IK	CCT	AI	IK	CCT	AI
	500	0.014	0.001	0.040	0.001	0.000	0.013	0.008	0.001	0.035
DGP1	1000	0.153	0.059	0.135	0.030	0.008	0.071	0.116	0.038	0.144
(Power)	2000	0.375	0.259	0.379	0.181	0.068	0.203	0.325	0.204	0.376
	4000	0.583	0.499	0.608	0.465	0.328	0.463	0.534	0.446	0.610
	8000	0.873	0.745	0.870	0.740	0.573	0.703	0.816	0.674	0.864
	500	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
DGP2	1000	0.001	0.000	0.004	0.000	0.000	0.005	0.000	0.000	0.001
(Size)	2000	0.004	0.005	0.003	0.004	0.003	0.011	0.004	0.003	0.005
	4000	0.014	0.009	0.001	0.009	0.001	0.003	0.011	0.005	0.003
	8000	0.010	0.008	0.005	0.010	0.010	0.010	0.008	0.009	0.006
	500	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000
DGP3	1000	0.000	0.000	0.001	0.000	0.000	0.010	0.000	0.000	0.003
(Power)	2000	0.005	0.000	0.004	0.001	0.001	0.008	0.003	0.001	0.001
	4000	0.003	0.004	0.006	0.003	0.001	0.018	0.004	0.003	0.008
	8000	0.004	0.005	0.003	0.005	0.006	0.011	0.004	0.008	0.003
	500	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
DGP4	1000	0.001	0.000	0.001	0.000	0.000	0.006	0.001	0.000	0.001
(Size)	2000	0.003	0.006	0.006	0.001	0.000	0.011	0.003	0.004	0.006
	4000	0.011	0.005	0.006	0.004	0.003	0.019	0.006	0.005	0.005
	8000	0.006	0.005	0.004	0.008	0.004	0.010	0.005	0.005	0.005

Table 8: Rejection Frequency at 10% Level

	n	Undersmoothing IK	Undersmoothing CCT	Undersmoothing AI	MSE-RBC IK	MSE-RBC CCT	MSE-RBC AI	CER-RBC IK	CER-RBC CCT	CER-RBC AI
DGP1 (Power)	500	0.253	0.203	0.309	0.110	0.046	0.185	0.220	0.150	0.303
	1000	0.494	0.389	0.469	0.319	0.201	0.341	0.444	0.361	0.464
	2000	0.683	0.604	0.675	0.540	0.413	0.544	0.629	0.559	0.679
	4000	0.890	0.798	0.891	0.773	0.624	0.764	0.858	0.738	0.893
	8000	0.988	0.963	0.994	0.946	0.865	0.920	0.973	0.931	0.994
DGP2 (Size)	500	0.071	0.056	0.079	0.034	0.014	0.086	0.056	0.041	0.073
	1000	0.103	0.079	0.060	0.074	0.050	0.089	0.089	0.078	0.058
	2000	0.100	0.093	0.071	0.086	0.089	0.114	0.105	0.090	0.073
	4000	0.064	0.100	0.060	0.080	0.085	0.081	0.073	0.094	0.059
	8000	0.066	0.078	0.076	0.095	0.091	0.094	0.085	0.075	0.075
DGP3 (Power)	500	0.060	0.051	0.056	0.031	0.023	0.063	0.054	0.048	0.056
	1000	0.084	0.075	0.065	0.068	0.051	0.098	0.084	0.066	0.061
	2000	0.081	0.084	0.059	0.070	0.074	0.095	0.081	0.073	0.055
	4000	0.060	0.074	0.070	0.078	0.065	0.100	0.065	0.071	0.071
	8000	0.060	0.079	0.055	0.094	0.086	0.078	0.079	0.078	0.055
DGP4 (Size)	500	0.055	0.048	0.060	0.035	0.025	0.068	0.049	0.036	0.060
	1000	0.076	0.063	0.081	0.059	0.034	0.103	0.073	0.061	0.083
	2000	0.078	0.076	0.075	0.081	0.083	0.091	0.085	0.078	0.074
	4000	0.081	0.069	0.071	0.091	0.099	0.094	0.095	0.079	0.068
	8000	0.054	0.076	0.056	0.065	0.086	0.084	0.056	0.081	0.056

Table 9: Bandwidth Values for Columbian SR Data

Bandwidth	Household edu. spending $n = 61969$	Total spending on food $n = 59398$	Total monthly exp. $n = 23140$
2	(2,2)	(2,2)	(2,2)
3	(3,3)	(3,3)	(3,3)
4	(4,4)	(4,4)	(4,4)
IK-US	(6.5,6.1)	(6.7,6.4)	(12.4,11.6)
IK-MSE-RBC	(7.9, 7.9)	(8.1,8.1)	(14.5,14.5)
IK-CER-RBC	(5.1,4.5)	(5.3,4.7)	(10.2,8.8)
CCT-US	(2.9,2.8)	(3.5,3.4)	(3.2,3.0)
CCT-MSE-RBC	(3.5,3.5)	(4.2, 4.2)	(3.8,3.8)
CCT-CER-RBC	(2.3,2.0)	(2.8,2.5)	(2.7,2.3)
AI-US	(5.0,11.9)	(3.5,7.6)	(6.8,11.7)
AI-MSE-RBC	(6.1,15.1)	(4.2,9.7)	(7.9,14.6)
AI-CER-RBC	(4.0,8.8)	(2.8,5.6)	(5.6,8.8)

Table 10: Bandwidth Values Israeli School Data (Grade 4)

	g4math			g4verb		
Bandwidth	40(n=984)	80(n=1376)	120(n=976)	40(n=984)	80(n=1376)	120(n=976)
Fixed at 3	(3,3)	(3,3)	(3,3)	(3,3)	(3,3)	(3,3)
Fixed at 5	(5,5)	(5,5)	(5,5)	(5,5)	(5,5)	(5,5)
IK-US	(3.9,3.9)	(2.7,2.8)	(4.0,4.1)	(4.0,3.9)	(3.2,3.2)	(4.2,4.4)
IK-MSE-RBC	(4.4,4.4)	(3.2,3.2)	(4.6,4.6)	(4.5,4.5)	(3.7,3.7)	(4.9,4.9)
IK-CER-RBC	(3.3,3.3)	(2.3,2.3)	(3.3,3.6)	(3.4,3.3)	(2.6,2.7)	(3.5,3.8)
CCT-US	(10.6,10.4)	(10.4,10.5)	(8.6,9.0)	(11,10)	(10.2,10.2)	(10.3,10.7)
CCT-MSE-RBC	(12.0,12.0)	(12.1,12.1)	(10,10)	(12,12)	(11.8,11.8)	(11.9,11.9)
CCT-CER-RBC	(9.0,8.7)	(8.6,8.9)	(7.2,7.8)	(9.4,9.0)	(8.4,8.7)	(8.5,9.2)
AI-US	(11.1,15.0)	(15,9.2)	(15,10)	(7.6,15)	(13.6,9.7)	(15,13.3)
AI-MSE-RBC	(12.6,17.8)	(23,10.6)	(44.7,11.6)	(8.6,17.9)	(15.9,11.1)	(74.9,14.9)
AI-CER-RBC	(9.5,12.8)	(16.6,7.8)	(32,9)	(6.5,12.9)	(11.3,8.1)	(53.6,11.5)

Table 11: Bandwidth Values Israeli School Data (Grade 5)

	g5math			g5verb		
Bandwidth	40(n=984)	80(n=1359)	120(n=905)	40(n=984)	80(n=1359)	120(n=905)
Fixed at 3	(3,3)	(3,3)	(3,3)	(3,3)	(3,3)	(3,3)
Fixed at 5	(5,5)	(5,5)	(5,5)	(5,5)	(5,5)	(5,5)
IK-US	(4.0,3.9)	(3.8,3.9)	(3.7,3.8)	(4.2,4.1)	(3.7,3.7)	(3.2,3.3)
IK-MSE-RBC	(4.5,4.5)	(4.5,4.5)	(4.3,4.3)	(4.7,4.7)	(4.3,4.3)	(3.7,3.7)
IK-CER-RBC	(3.4,3.3)	(3.2,3.3)	(3.1,3.3)	(3.6,3.4)	(3.1,3.2)	(2.7,2.9)
CCT-US	(5.6,5.5)	(10.4,10.5)	(8.0,8.3)	(7.1,7.0)	(10.5,10.7)	(6.7,7.0)
CCT-MSE-RBC	(6.3,6.3)	(12.1,12.1)	(9.3,9.3)	(8.1,8.1)	(12.2,12.2)	(7.8,7.8)
CCT-CER-RBC	(4.7,4.5)	(8.6,8.9)	(6.7,7.2)	(6.1,5.8)	(8.7,9.0)	(5.6,6.1)
AI-US	(14.9,15.0)	(14.9,15)	(15,8.1)	(6.4,11.5)	(15,15)	(15,6.9)
AI-MSE-RBC	(16.9,52)	(17.4,22)	(31.7,9.1)	(7.3,13.3)	(17.8,18.3)	(28,7.7)
AI-CER-RBC	(12.7,38.2)	(12.9,16.2)	(22.7,7.1)	(5.5,9.6)	(12.6,13.5)	(20.3,6.0)

Table 12: Bandwidth for Romanian High School Data

	Attending best school		Avoiding worst school	
Bandwidth	Left	Right	Left	Right
IK-US	0.741	0.750	0.720	0.667
IK-MSE-RBC	0.940	0.940	0.855	0.855
IK-CER-RBC	0.550	0.566	0.582	0.490
CCT-US	0.572	0.579	0.144	0.133
CCT-MSE-RBC	0.726	0.726	0.171	0.171
CCT-CER-RBC	0.424	0.436	0.116	0.098
AI-US	0.777	0.763	0.105	0.375
AI-MSE-RBC	0.986	0.957	0.124	0.481
AI-CER-RBC	0.577	0.575	0.085	0.276

References

Acerenza, S., O. Bartalotti, and D. Kédagni (2020): "Testing Identifying Assumptions in Bivariate Probit Models," *Available at SSRN 3753058*.

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

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 (2019): "Maimonides Rule Redux," *American Economic Review Insights*, 1, 309–324.

Arai, Y., Y.-C. Hsu, T. Kitagawa, I. Mourifié, and Y. Wan (2022): "Testing identifying assumptions in fuzzy regression discontinuity designs," *Quantitative Economics*, 13(1), 1–28.

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

Battistin, E., A. Brugiavini, E. Rettore, and G. Weber (2009): "The retirement consumption puzzle: evidence from a regression discontinuity approach," *American Economic Review*, 99(5), 2209–26.

Bertanha, M., and M. J. Moreira (2020): "Impossible inference in econometrics: Theory and applications," *Journal of Econometrics*, 218(2), 247–270.

Bugni, F. A., and I. A. Canay (2021): "Testing continuity of a density via g-order statistics in the regression discontinuity design," *Journal of Econometrics*, 221(1), 138–159.

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

——— (2020): "Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs," *The Econometrics Journal*, 23(2), 192–210.

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," *The Review of Economic Studies*, 85(3), 1577–1608.

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 (2020): "Simple local polynomial density estimators," *Journal of the American Statistical Association*, 115(531), 1449–1455.

CHIANG, H. D., Y.-C. HSU, AND Y. SASAKI (2019): "Robust Uniform Inference for Quantile Treatment Effects in Regression Discontinuity Designs," *Journal of Econometrics*, 211, 589–618.

DONALD, S. G., AND Y.-C. HSU (2011): "A New Test for Linear Inequality Constraints When the Variance Covariance Matrix Depends on the Unknown Parameters," *Economics Letters*, 113, 241–243.

——— (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.

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

GERARD, F., M. ROKKANEN, AND C. ROTHE (2020): "Bounds on treatment effects in regression discontinuity designs with a manipulated running variable," *Quantitative Economics*, 11(3), 839–870.

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.

HOROWITZ, J. L., AND C. F. MANSKI (1995): "Identification and robustness with contaminated and corrupted data," *Econometrica: Journal of the Econometric Society*, pp. 281–302.

HSU, Y.-C., AND S. SHEN (2022): "Dynamic Regression Discontinuity under Treatment Effect Heterogeneity," *working paper*.

HUBER, M., AND G. MELLACE (2015): "Testing instrument validity for LATE identification based on inequality moment constraints," *Review of Economics and Statistics*, 97(2), 398–411.

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.

Kédagni, D., and I. Mourifié (2020): "Generalized instrumental inequalities: testing the instrumental variable independence assumption," *Biometrika*, 107(3), 661–675.

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

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

——— (2009): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 76(3), 1071–1102.

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

Ma, S., and M. R. Kosorok (2005): "Robust semiparametric M-estimation and the weighted bootstrap," *Journal of Multivariate Analysis*, 96(1), 190–217.

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.

Müller, T., and M. Shaikh (2018): "Your retirement and my health behavior: Evidence on retirement externalities from a fuzzy regression discontinuity design," *Journal of health economics*, 57, 45–59.

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.

Pearl, J. (1995): "Causal inference from indirect experiments," *Artificial intelligence in medicine*, 7(6), 561–582.

Pop-Eleches, C., and M. Urquiola (2013): "Going to a better school: Effects and behavioral responses," *American Economic Review*, 103(4), 1289–1324.

Sun, Z. (2020): "Instrument Validity for Heterogeneous Causal Effects," *arXiv preprint arXiv:2009.01995*.

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.