University of Toronto
Department of Economics

Working Paper 761

Testing Identification Conditions of LATE in Fuzzy Regression Discontinuity Designs

By Yu-Chin Hsu, Ji-Liang Shiu and Yuanyuan Wan

October 19, 2023

Testing Identification Conditions of LATE in Fuzzy Regression Discontinuity Designs*

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

October 7, 2023

Abstract

This paper derives testable implications of the identifying conditions for the local average treatment effect (LATE) 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

*We thank helpful comments from Yingying Dong, Ismael Mourifie, and participants of the Asian meeting of the Econometric Society. All errors are ours.

†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. Yu-Chin Hsu gratefully acknowledges research support from the National Science and Technology Council of Taiwan (NSTC111-2628-H-001-001), the Academia Sinica Investigator Award of the Academia Sinica, Taiwan (AS-IA-110-H01), and the Center for Research in Econometric Theory and Applications (107L9002) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education of 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. Yuanyuan Wan thanks the support from the SSHRC Insight Grant #435190500.

1 Introduction

This paper derives testable implications of the identification assumptions for the local average treatment effect (LATE) in fuzzy regression discontinuity (FRD) designs and develops a specification test for the 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 the potential outcome framework, where they provide conditions to identify the average treatment effect (ATE) and LATE at the cutoff, respectively. These conditions are revisited later by Lee (2008), Imbens and Lemieux (2008), Frandsen, Frölich, and Melly (2012), Dong (2018), and Bertanha and Moreira (2020), 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. For sharp designs, Lee (2008) proposes testable implications: (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 was 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 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 (also see McCrary, 2008). 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, referred to as the "FRD distributional assumptions" hereafter, are used to identify quantile or distributional treatment effects for compliers. Our paper contributes to this second strand by proposing a specification test for identifying conditions for the mean effect. The identification assumptions required for LATE replace the local continuity in distributions with local continuity in means, i.e. the expectations of potential outcomes given that the running variable is continuous near the cutoff. The local monotonicity and the local continuity in means assumptions are referred to as the "FRD mean assumptions" hereafter.

We consider a test for the FRD mean assumptions a useful addition to the existing tests on the distributional assumptions for the following reasons. First, the mean effect is of primary interest in many empirical applications requiring weaker FRD mean assumptions. On the other hand, it is possible that the FRD mean assumptions hold but not the distributional assumptions. For example, if there is an unobserved program using the same running variable and if the unobserved program changes the risk faced by individuals, then even if the true potential outcomes satisfy the FRD distributional assumptions, the perceived potential outcomes (due to ignorance of the unobserved program) by the researcher can violate the continuity in distributions while satisfying the continuity in means. In Section 2, we postulate a simple model to demonstrate this point further. Second, the disagreement between the mean and distribution tests can happen in real RD data. For example, we applied our test to Columbia's subsidized insurance program data (analyzed by Miller, Pinto, and Vera-Hernández, 2013) in Section 5.1, and we find that the mean test rejects but not the distribution test. Recently, Canay and Kamat (2018) find that in the US House election data analyzed in the seminal paper of Lee (2008), the distribution of the past vote margin (baseline variable) given current vote margin (running variable) is not continuous at the cutoff. On the other hand, the continuity of the conditional mean is not rejected and is considered a reasonable assumption in the literature. While these results do not directly involve the outcome variable (future vote margin), it suggests that we need to be cautious to assume the continuity in distribution. Again, we will discuss this point in greater detail in Section 2. Our test can provide an additional perspective and modelling reference for empirical researchers in these scenarios. Third, statistically, there does not exist a subset of testable implications for FRD distribution assumptions that are "dedicated" to the FRD mean assumptions. When testing the FRD mean assumptions is desirable, naively applying the distributional test may falsely reject the mean continuity. Therefore, having an easy-to-implement specification test focusing directly on the FRD mean assumptions would be useful. Our paper fills this gap.

Specifically, we derive sharp testable implications for the FRD mean assumptions that share the same spirit as Huber and Mellace (2015) for the binary IV setting. We construct sharp (observable) bounds for the expectation of the potential outcome $Y(1)$ for always-takers when the running variable approaches the cutoff by applying the results in Horowitz and Manski (1995), and Lee (2009). Therefore, its identifiable estimand must lie within the bounds as well. 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. These four inequalities constitute necessary (but not sufficient) conditions for FRD mean assumptions. Huber and Mellace (2015) also use the idea of Horowitz and Manski (1995) and Lee (2009) to test the validity of instruments (mean independence) in the binary treatment and binary IV setting. While drawing motivation from Huber and Mellace (2015), we focus on the FRD framework. Therefore, our testable implications are characterized by the limits of conditional expectations when the running variable approaches the cutoff. For this reason, the test statistics, their asymptotic properties, and the calculation of critical value are significantly different.[1] Furthermore, built upon the results of Laffers and Mellace (2017), we also show that if the observable data distribution satisfies our testable implications, then there exists a joint distribution of potential outcomes and complying status which (i) satisfies the FRD mean assumptions, and (ii) is observationally local-equivalent to the observable data distribution. In this sense, the testable implication is sharp or the best possible to detect the violation of FRD mean assumptions.

The proposed specification test is based on these inequality constraints. Our test statistic

[1] Horowitz and Manski (1995) and Lee (2009)'s idea is also applied to bound the treatment effect in RDD designs with manipulation, see for instance Gerard, Rokkanen, and Rothe (2020). Our paper focus on specification test.

is significantly different from that in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) because the inequality constraints in our case involve nuisance parameters that need to be estimated in advance. 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 alternatives.

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 FRD distributional assumptions 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 implies 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 results and findings in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) suggest that one needs to be especially 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 this application, Israel's Maimonides' rule creates a discontinuity of class size with respect to enrollment. Our results align with those in Angrist, Lavy, Leder-Luis, and Shany (2019) and Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), and show no evidence to reject FRD mean assumptions and FRD distributional 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 does not reject the validity of FRD design in this empirical study. It is not surprising that our test does not reject the FRD mean assumptions in these data sets, as they are generated from classical FRD designs. However, as we will illustrate in simulation (Section 4) using Battistin, Brugiavini, Rettore, and Weber (2009)'s retirement consumption data, a modest "artificial distortion" of the real data can be detected by our test if the distortion results in the discontinuity of mean.

In addition to Huber and Mellace (2015) and Laffers and Mellace (2017) mentioned above, 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. 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 (2023) test identifying assumptions in bivariate Probit models.

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 Model 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 σ-algebra, and the P is the probability distribution of all random variables. Among all the variables, $D(\cdot) : \Omega \to \{0, 1\}$ is the observed binary treatment, $Y(\cdot) : \Omega \to \mathcal{Y}$ is the observed outcome of interest, and $Z(\cdot) : \Omega \to \mathcal{Z}$ is a continuous running variable with a known cut-off c.[2] A given individual ω in the population is endowed with a potential treatment

[2]The treatment variable may take multiple values in empirical applications with multiple cutoffs for the running variable. In such cases, we can apply our test to the specific cutoff of interest or test the identifying conditions

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 set to z. Likewise, let $Y(d,\omega)$ be his/her potential outcome had the treatment been set to d. The observed treatment and outcome are connected as $D(\omega) \equiv D(Z(\omega),\omega)$ and $Y(\omega) \equiv D(\omega)Y(1,\omega) + (1 - D(\omega))Y(0,\omega)$, respectively.

Let $\varepsilon > 0$ be a small positive number and define $B_\varepsilon = \{z \in \mathcal{Z} : |z - c| \le \varepsilon\}$ be an interval centred at the cutoff. Following Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), we define compliance status T_ε of an individual ω based on the shape of the potential treatment function over B_ε:

$$T_\varepsilon(\omega) = \begin{cases} \mathbf{A}, & \text{if } D(z,\omega) = 1, \text{ for } z \in B_\epsilon, \\ \mathbf{N}, & \text{if } D(z,\omega) = 0, \text{ for } z \in B_\epsilon, \\ \mathbf{C}, & \text{if } D(z,\omega) = 1\{z \ge c\}, \text{ for } z \in B_\epsilon, \\ \mathbf{DF}, & \text{if } D(z,\omega) = 1\{z < c\}, \text{ for } z \in B_\epsilon, \\ \mathbf{I}, & \text{if } D(z,\omega) \text{ takes any other forms,} \end{cases} \tag{2.1}$$

where **A, C, N** and **DF** represent always-takers, compliers, never-takers, and defiers, respectively. Here, **I(ndeterminant)** represents individuals whose potential treatment is a non-constant function of z over $(-\varepsilon - c, c)$ or $(c, c+\varepsilon)$. Introducing this type allows the probability of treatment to be a non-constant function near the cutoff.[3] Hereafter, we will suppress the argument ω whenever it causes no confusion. Following Imbens and Lemieux (2008), we make the following assumptions.

Assumption 2.1 (Local monotonicity) *For $t = \boldsymbol{DF}$ or $\boldsymbol{I}$,*

$$\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = t | Z = c - \varepsilon) = \lim_{\varepsilon\downarrow 0} P(T_\varepsilon = t | Z = c + \varepsilon) = 0.$$

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. Therefore, there are only compliers, always-takers and never-takers when the running variable approaches the cutoff.

jointly for all cutoffs.

[3]There are other ways to define complying status while allowing the probability of treatment to be non-constant near the cutoff. Fox example, Dong (2018) separated the role of "IV" from the running variable by defining $W = 1\{Z \ge c\}$ and the observed treatment $D = WD_1 + (1 - W)D_0$, where $D_w = h_w(Z, U_w)$ is the potential treatment and U_w is an unobserved latent variable. Finally, the complying status is defined based on the combination of the pair of potential treatments (D_1, D_0). In our paper, we follow Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) because both papers focus on testing, and we do so for better comparison.

Example 2.1 *Consider the following single-threshold crossing specification of potential treatment,*

$$D(z) = 1\{z + 1\{z \geq c\} + V \geq 0\}, \quad z \in [c - \varepsilon, c + \varepsilon]$$

where $V \sim N(0,1)$. *Fixing* $\varepsilon > 0$ *and setting* $c = 0$ *(without loss of generality), the support of* V *is divided into four groups, as shown in the following figure. Consider an individual*

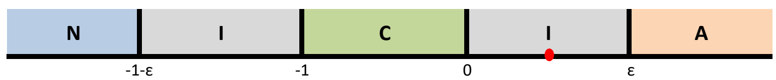


Figure 1: Complying Types

whose $v = \frac{\varepsilon}{2}$. *This person is not a compiler or an always-taker or a never-taker because his/her potential treatment as a function of* z *is not constant over* $(-\varepsilon, 0)$: $D\left(-\frac{3\varepsilon}{4}\right) = 0$ *but* $D\left(-\frac{\varepsilon}{4}\right) = 1$. *Based on our definition in Equation (2.1), he/she belongs to* $\boldsymbol{I}$. *In this example, Assumption 2.1 is satisfied because the probability of* $\boldsymbol{I}$, *or equivalently the probability of* $V \in (-1-\varepsilon, -1) \cup (0, \varepsilon)$ *converges to zero as* $\varepsilon \downarrow 0$. *Also,* $P(D = 1|Z = z)$ *is not constant near the cutoff because of the existence of type* $\boldsymbol{I}$.

For given ε, d and t, let $f_{Y(d)|T_\varepsilon,Z}(y|t,z)$ be the probability density function of $Y(d)$ given type $T_\varepsilon = t$ and $Z = z$ (when $Y(d)$ is discrete, these densities are understood as probability mass functions).

Assumption 2.2 (Local continuity in means) *For all* t *and* d, *we have*

(i) $\lim_{\varepsilon \downarrow 0} f_{Y(d)|T_\varepsilon,Z}(y|t, c-\varepsilon)$ *and* $\lim_{\varepsilon \downarrow 0} f_{Y(d)|T_\varepsilon,Z}(y|t, c+\varepsilon)$ *are proper densities. Furthermore,* $E[|Y(d)||T_\varepsilon = t, Z = z] < \infty$ *for all* $z \in B_\delta$ *for some* $\delta > 0$.

(ii) $\lim_{\varepsilon \downarrow 0} E[Y(d)|T_\varepsilon = t, Z = c - \varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y(d)|T_\varepsilon = t, Z = c + \varepsilon]$ *and* $\lim_{\varepsilon \downarrow 0} P(T_\varepsilon = t|Z = c - \varepsilon) = \lim_{\varepsilon \downarrow 0} P(T_\varepsilon = t|Z = c + \varepsilon)$.

Assumption 2.2-(i) contains regularity conditions. It requires the conditional densities of potential outcomes to have well-defined limits from above and below the cutoff, respectively, but not necessarily equal to each other. It also requires the conditional (truncated or untruncated) expectations of potential outcomes of each type to be finite in the limit. We make these assumptions so that the quantities in our testable implication are well-defined. These conditions can be supported by reasonable models. For instance, they are satisfied in Example 2.1

if $(Y(1), Y(0), V)$ follows a joint normal distribution. In this example, the density function of $Y(d)$ given $T_\varepsilon = \boldsymbol{I}$ exists for any ε and also exists when $\varepsilon \to 0$, despite that the population size of type $\boldsymbol{I}$ converges to zero as $\varepsilon \to 0$. Assumption 2.2-(ii) is the key assumption. It 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 and for the type probabilities.[4] It shares the same spirit of the Assumption LS1 of Dong (2018). Given $\lim_{\varepsilon \downarrow 0} f_{Y(d)|T,Z}(y|t, c-\varepsilon)$ and $\lim_{\varepsilon \downarrow 0} f_{Y(d)|T_\varepsilon,Z}(y|t, c+\varepsilon)$ are well-defined, Assumption 2.2-(ii) is weaker than the local continuity assumption in distributions, which is tested in Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) and restated below.

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

$$\lim_{\varepsilon \downarrow 0} P(Y(d) \in V, T_\varepsilon = t | Z = c - \varepsilon) = \lim_{\varepsilon \downarrow 0} P(Y(d) \in V, T_\varepsilon = t | Z = c + \varepsilon).$$

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 the distributions assumption. For the purpose of exposition, the proof is omitted. For generic random variables (R_1, R_2, R_3), let $E[R_1|R_2, R_3 = c^+] = \lim_{\varepsilon \downarrow 0} E[R_1|R_2, R_3 = c+\varepsilon]$ and $E[R_1|R_2, R_3 = c^-] = \lim_{\varepsilon \downarrow 0} E[R_1|R_2, R_3 = c-\varepsilon]$, 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 dis-

[4] Please note that if the continuity of the conditional mean is imposed on a known transformation $\kappa(Y(d))$, our test can be adapted for it by using $\kappa(Y)$ in the place of Y. In practice, the choice of κ should be guided by the specific needs of empirical research.

tributions at the cutoff are identified by the following quantities:

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

$$F_{Y(0)|\mathbf{C},Z=c}(y) = \frac{E[1\{Y \leq y\}(1-D)|Z=c^+] - E[1\{Y \leq y\}(1-D)|Z=c^-]}{E[D|Z=c^+] - E[D|Z=c^-]}. \quad (2.4)$$

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

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

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

for any g belonging to the class of closed 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, and hence the FRD mean assumptions.

However, (2.5) and (2.6) are not necessarily implied by FRD mean assumptions.[5] In practice, continuity in mean and discontinuity in distribution can coexist for various economic reasons. For example, suppose that there are multiple programs (say D and D') that are implemented based on the same running variable Z. To fix the idea, take D as a job training program, $D' = 1[Z \geq c]$ as a government-subsidized insurance program, and the running variable as the poverty index. Suppose that a researcher hopes to estimate the effect of the job training program on labor income. In this case, the potential outcome should be defined as $\tilde{Y}(d, d')$ for $d \in \{0, 1\}$

[5] For example, suppose $Y \geq 0$. Note that $Y(d) = \int_0^{Y(d)} dy = \int_0^{\infty} 1\{Y(d) \geq y\}dy$ and under suitable regularity conditions, we have

$$E[Y(d)|Z=z] = \int_0^{\infty} E[1\{Y(d) \geq y\}|Z=z]dy = \int_0^{\infty} P(Y(d) \geq y|Z=z)dy.$$

It is possible that $P(Y(d) \geq y|Z=z)$ is discontinuous in z for some y, but the "averaged" version $\int_0^{\infty} P(Y(d) \geq y|Z=z)dy$ is continuous in z. We thank an anonymous referee for offering this perspective.

and $d' \in \{0,1\}$, so that the observed outcome is

$$\begin{aligned} Y &= \tilde{Y}(1,1)DD' + \tilde{Y}(1,0)D(1-D') + \tilde{Y}(0,1)(1-D)D' + \tilde{Y}(0,0)(1-D)(1-D') \\ &= \underbrace{(\tilde{Y}(1,1)D' + \tilde{Y}(1,0)(1-D'))}_{\equiv Y(1)} D + \underbrace{(\tilde{Y}(0,1)D' + \tilde{Y}(0,0)(1-D'))}_{\equiv Y(0)}(1-D). \end{aligned}$$

If the other program D' were ignored, then the researcher would consider $Y(d) = \tilde{Y}(d,1)D' + \tilde{Y}(d,0)(1-D')$ as his/her potential outcome. In this case, the conditional distribution of $Y(d)$ given $Z = z$ will not be continuous in z even if the distributions of $\tilde{Y}(d,d')$ are. To see this, observe that $P(Y(1) \leq y|Z = c+\varepsilon) = P(\tilde{Y}(1,1) \leq y|Z = c+\varepsilon)$, whereas $P(Y(1) \leq y|Z = z-\varepsilon) = P(\tilde{Y}(1,0) \leq y|Z = c-\varepsilon)$. Therefore, $P(Y(1) \leq y|Z = z)$ is generally not continuous at c as long as $\tilde{Y}(1,1)$ and $\tilde{Y}(1,0)$ have different distributions. However, if the insurance program targets zero profit and thus an individual pays zero expected premium (since the government subsidizes it), then $\tilde{Y}(1,0)$ will be a mean-preserving spread of $\tilde{Y}(1,1)$. The distributional assumption is violated in this case, but the conditional expectation of $Y(1)$ can still be continuous at c.

The departure between continuity in mean and in distribution is also documented in other scenarios of RD designs. In a seminal paper, Lee (2008) investigates the "incumbency advantage of US House election", where the treatment variable is the indicator of "Democratic party wins at time t", the potential outcome variable $Y(d)$ is the Democratic party's counterfactual winning margin at time $t+1$, the running variable Z is the winning margin at t, and a baseline variable X is the winning margin at $t-1$. Lee (2008, Condition 1b and 2b) assumes that (i): $(Y(1), Y(0), X)$ are measurable functions of underlying latent variable W,

$$Y(1) = m_1(W), \quad Y(0) = m_0(W), \quad X = m_x(W),$$

and (ii): $F_{Z|W}$ is continuously differentiable at the cutoff. A necessary implication of these assumptions is that the distribution of X given $Z = z$ is continuous in z at the cutoff (Lee, 2008, Proposition 2). Despite that the continuity of $E[X|Z = z]$ is not rejected, a recent study by Canay and Kamat (2018) shows that the continuity of $F_{X|Z}(\cdot|z)$ in z is actually rejected.[6]

[6]Examining the result of Canay and Kamat (2018, Figure 2(a)), one can note that $f_{X|Z}(\cdot|Z = c^-)$ puts significantly more mass over the interval just below the cuooff value (50%) than $f_{X|Z}(\cdot|Z = c^+)$. Because the future can not change the past, one possible explanation for the discontinuity is that persistent unobserved factors

Such rejection implies that conditions (1b) and (2b) of Lee (2008) fail to hold for the conditional distribution of period $t-1$ vote margin. While it is not a piece of direct evidence against the continuity of the conditional distribution of period $t+1$ (potential) vote margin, it does remind us to be cautious in making such an assumption. On the other hand, the continuity in mean appears to be a reasonable assumption (see Lee, 2008, Table 1).[7]

To state the proper set of testable implications for Assumptions 2.1 and 2.2, we first define some notation. Let $G_1(y) = \lim_{\varepsilon \downarrow 0} P(Y \leq y|D = 1, Z = c+\varepsilon)$ be the conditional distribution of Y given $D = 1$ and $Z = z$ when z converges to c from above. Note that $G_1(y) = \lim_{\varepsilon \downarrow 0} P(Y(1) \leq y|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}\}, Z = c+\varepsilon)$, and is well defined under Assumption 2.2-(i). Likewise, define $G_0(y) = \lim_{\varepsilon \downarrow 0} P(Y \leq y|D = 0, Z = c - \varepsilon)$. 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 $P_{1|0} = \lim_{\varepsilon \downarrow 0} P(D = 1|Z = c - \varepsilon)$ and $P_{1|1} = \lim_{\varepsilon \downarrow 0} P(D = 1|Z = c + \varepsilon)$.[8] Likewise, we define $r = P_{0|1}/P_{0|0}$, where $P_{0|1} = \lim_{\varepsilon \downarrow 0} P(D = 0|Z = c + \varepsilon)$ and $P_{0|0} = \lim_{\varepsilon \downarrow 0} P(D = 0|Z = z - \varepsilon)$. 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 $\tau \in (0, 1)$, define its τ-quantile as $\tilde{F}^{-1}(\tau) = \inf\{y \in \mathcal{Y} : \tilde{F}(y) \geq \tau\}$.

Now, we are ready to present the testable implications of the FRD mean assumptions. Applying the results in Horowitz and Manski (1995) and Lee (2009), we derive the bounds for $\lim_{\varepsilon \downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c - \varepsilon]$ and $\lim_{\varepsilon \downarrow 0} E[Y(0)|T_\varepsilon = \boldsymbol{N}, Z = c + \varepsilon]$, respectively. Their identifiable estimands must satisfy the bounds as well, and form restrictions on the distribution of observed data. We summarize them in Proposition 2.2.

Proposition 2.2 *Suppose that $q \in (0, 1)$, $r \in (0, 1)$, and the distributions of Y given $(D = 1, Z = z)$ and Y given $(D = 0, Z = z)$ are continuous.*

affect voting margins over time. A thorough study of all possible reasons is beyond the scope of our paper.

[7]Because this empirical application is a sharp design, we can not apply our method to test it.

[8]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.

(i) If Assumptions 2.1 and 2.2 are satisfied, then the following inequality constraints hold:

$$E[Y|D=1, Y < G_1^{-1}(q), Z=c^+] \leq E[Y|D=1, Z=c^-], \quad (2.7)$$

$$E[Y|D=1, Z=c^-] \leq E[Y|D=1, Y > G_1^{-1}(1-q), Z=c^+], \quad (2.8)$$

$$E[Y|D=0, Y < G_0^{-1}(r), Z=c^-] \leq E[Y|D=0, Z=c^+], \quad (2.9)$$

$$E[Y|D=0, Z=c^+] \leq E[Y|D=0, Y > G_0^{-1}(1-r), Z=c^-]. \quad (2.10)$$

(ii) If the joint distribution of (Y, D, Z) satisfies inequalities (2.7) to (2.10), then there exists a collection of random variables $(\tilde{Y}(1), \tilde{Y}(0), Z, \tilde{D}(z), z \in \mathcal{Z})$ whose distribution satisfies Assumptions 2.1 and 2.2. Furthermore, let $\tilde{Y} \equiv \tilde{Y}(1)\tilde{D} + \tilde{Y}(0)(1-\tilde{D})$ and $\tilde{D} \equiv \tilde{D}(Z)$. Then the distribution of $(\tilde{Y}, \tilde{D})|Z=z$ is the same as that of $(Y, D)|Z=z$.

The first part of the proposition shows that the FRD mean assumptions have empirical content and generate the four inequalities as necessary (but not sufficient) conditions. The inequalities constraints are intuitive. For example, under the FRD mean assumptions, the right-hand side of inequality (2.7) identifies the expectation of $Y(1)$ for always takers (approaching the cutoff), while $E[Y|D=1, z=c^+]$ is the expectation of $Y(1)$ for both always-takers and compliers ((approaching to the cutoff)). In one extreme case, all the always takers concentrate on the left tail of the mixing distribution: this gives inequality (2.7). In another extreme case, all the always takers concentrate on the left tail of the mixing distribution, which gives inequality (2.8).

The first part of the proposition has a similar structure as the results of Huber and Mellace (2015), who derive the bounds for the expectations of $Y(1)$ for always takers and $Y(0)$ for never takers in the binary IV setting. In our case, we bounded these quantities at the cutoff. The second part of the proposition states that the implications are sharp to detect the violations in means. It is analogous to Laffers and Mellace (2017, Theorem 1-(i)), which shows that the testable implication of Huber and Mellace (2015) is the best possible to detect violations of IV assumptions in the binary IV settings.

Remark 2.1 *In Proposition 2.2, we assume the distribution of the outcome is continuous. As we will show in Corollary 6.1 in Section 6.1, the bounds in Proposition 2.2 are still valid but not necessarily sharp when Y is discrete or a mixture of continuous and discrete parts. Corollary 6.1* *reports the sharp bounds at the cost of additional notation.*

Remark 2.2 *One temptation to test FRD mean assumptions is to change the function class of Arai, Hsu, Kitagawa, Mourifié, and Wan (2022), for example, to replace $g(Y)$ by Y and check whether the following inequalities hold,*

$$E[YD|Z = c^-] - E[YD|Z = c^+] \leq 0 \tag{2.11}$$

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

This is, however, not a valid approach. The inequalities (2.11) and (2.12) can fail to hold even when the FRD distributional assumptions are satisfied. Let π_t be the size of subpopulation t at the cutoff. A simple calculation shows that the left-hand side of inequality (2.11) is $-E[Y(1)|\boldsymbol{C}, Z = c]\pi_{\boldsymbol{C}}$. Its sign can not be determined unless $Y(1) \geq 0$ or $Y(1) \leq 0$ almost surely. When $Y(d)$ is nonnegative (e.g. test score) for $d = 0, 1$, the restriction in (2.11) can be implied by our testable implication (2.8). To see this, note that (2.8) implies that

$$
\begin{aligned}
&E[YD|Z = c^-] = \pi_{\boldsymbol{A}} E[Y|D = 1, Z = c^-] \\
\leq &\pi_{\boldsymbol{A}} E[Y|D = 1, Y > G_1^{-1}(1-q), Z = c^+] \\
= &\pi_{\boldsymbol{A}} \frac{E[Y|D = 1, Z = c^+] - E[Y|D = 1, Y \leq G_1^{-1}(1-q), Z = c^+]P(Y \leq G_1^{-1}(1-q)|D = 1, Z = c^+)}{P(Y > G_1^{-1}(1-q)|D = 1, Z = c^+)} \\
\leq &\pi_{\boldsymbol{A}} \frac{E[Y|D = 1, Z = c^+]}{P(Y > G_1^{-1}(1-q)|D = 1, Z = c^+)} \\
= &\pi_{\boldsymbol{A}} \frac{E[Y|D = 1, Z = c^+]}{q} \\
= &\pi_{\boldsymbol{A} \cup \boldsymbol{C}} E[Y|D = 1, Z = c^+] \\
= &E[YD|Z = c^+],
\end{aligned}
$$

where the first equality holds by Assumption 2.1, the first inequality holds by applying (2.8), the second inequality holds because $Y > 0$, and the final two equalities hold because $q = \frac{\pi_{\boldsymbol{A}}}{\pi_{\boldsymbol{A} \cup \boldsymbol{C}}}$. It is exactly restriction (2.11). Hence, even when the sign of $Y(d)$ is known to be positive or negative, it is always more informative to use restriction (2.8).

Remark 2.3 *From the inequalities in Proposition 2.2, we expect that our test be more powerful when q and r are relatively large under the alternatives. The main reason is that, for example,* *the lower bound in (2.7) decreases when q decreases to zero (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 (2.7) and (2.8) 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 $Y|D = 1, Z = c^+$. Then (2.7) and (2.8) hold automatically.*

3 Proposed Test

In this section, we propose a test for the implications in Proposition 2.2. It turns out to be useful to reformulate the inequalities. For inequality (2.7),

$$\begin{aligned}
& E[Y|D=1, Y < G_1^{-1}(q), Z = c^+] - E[Y|D=1, Z=c^-] \leq 0 \\
\Leftrightarrow\ & \frac{E[DY1(Y < G_1^{-1}(q))|Z=c^+]}{E[D1(Y < G_1^{-1}(q))|Z=c^+]} - \frac{E[DY|Z=c^-]}{E[D|Z=c^-]} \leq 0 \\
\Leftrightarrow\ & \theta_1 \equiv E[DY1(Y < G_1^{-1}(q))|Z=c^+] \cdot E[D|Z=c^-] \\
& \quad - E[DY|Z=c^-] \cdot E[D1(Y < G_1^{-1}(q))|Z=c^+] \leq 0. \qquad (3.1)
\end{aligned}$$

Similarly, the rest of the three inequalities (2.8)-(2.10) are equivalent to

$$\begin{aligned}
\theta_2 &\equiv E[DY|Z=c^-] \cdot E[D1(Y > G_1^{-1}(1-q))|Z=c^+] \\
&\quad - E[DY1(Y > G_1^{-1}(1-q))|Z=c^+] \cdot E[D|Z=c^-] \leq 0, \qquad (3.2) \\
\theta_3 &\equiv E[(1-D)Y1(Y < G_1^{-1}(r))|Z=c^-] \cdot E[1-D|Z=c^+] \\
&\quad - E[(1-D)Y|Z=c^+] \cdot E[(1-D)1(Y < G_1^{-1}(r))|Z=c^-] \leq 0, \qquad (3.3) \\
\theta_4 &\equiv E[(1-D)Y|Z=c^+] \cdot E[(1-D)1(Y > G_1^{-1}(1-r))|Z=c^-] \\
&\quad - E[(1-D)Y1(Y > G_1^{-1}(1-r))|Z=c^-] \cdot E[(1-D)|Z=c^+] \leq 0. \qquad (3.4)
\end{aligned}$$

Note that the inequalities in the reformulation are well-defined regardless the design is sharp or has one-sided compliance. To this end, we can formulate our null hypothesis H_0 as

$$H_0 : \theta_j \leq 0 \text{ for } j = 1, 2, 3 \text{ and } 4. \tag{3.5}$$

The rest of the section gives the details of our testing precedure.

3.1 Estimation of θ_j's

We first consider the estimation of θ_j for $j = 1, \ldots, 4$ and derive the asymptotics of corresponding estimators. For a generic random variable A, let $\widehat{E}_h[A|Z = c^+]$ and $\widehat{E}_h[A|Z = c^-]$ be the local quadratic regression estimators for $E[A|Z = c^+]$ and $E[A|Z = c^-]$, respectively, with bandwidth h and kernel function $K(\cdot)$. To be specific,

$$(\widehat{E}_h[A|Z = c^+], \hat{\beta}_+, \hat{\kappa}_+) = \operatorname*{argmin}_{a,b,k} \sum_{i=1}^{n} 1(Z_i \geq c) \cdot K\left(\frac{Z_i - c}{h}\right) \left[A_i - a - b \cdot (Z_i - c) - k \cdot (Z_i - c)^2\right]^2,$$
$$(\widehat{E}_h[A|Z = c^-], \hat{\beta}_-, \hat{\kappa}_-) = \operatorname*{argmin}_{a,b,k} \sum_{i=1}^{n} 1(Z_i < c) \cdot K\left(\frac{Z_i - c}{h}\right) \left[A_i - a - b \cdot (Z_i - c) - k \cdot (Z_i - c)^2\right]^2.$$

We consider two-step estimators for θ_j's. For the first step, we respectively estimate $G_1(y) = E[1(Y \leq y)|D = 1, Z = c^+] \equiv E[D1(Y \leq y)|Z = c^+]/E[D|Z = c^+]$ and $G_0(y) \equiv E[1(Y \leq y)|D = 0, Z = c^-] = E[(1 - D)1(Y \leq y)|Z = c^-]/E[1 - D|Z = c^-]$ by

$$\widehat{G}_1(y) = \frac{\widehat{E}_{h_1}[D1(Y \leq y)|Z = c^+]}{\widehat{E}_{h_1}[D|Z = c^+]}, \quad \widehat{G}_0(y) = \frac{\widehat{E}_{h_1}[(1 - D)1(Y \leq y)|Z = c^-]}{\widehat{E}_{h_1}[1 - D|Z = c^-]}.$$

Let $q \equiv E[D|Z = c^-]/E[D|Z = c^+]$ and $r \equiv E[1 - D|Z = c^+]/E[1 - D|Z = c^-]$ be estimated by

$$\hat{q} = \frac{\widehat{E}_{h_1}[D|Z = c^-]}{\widehat{E}_{h_1}[D|Z = c^+]}, \quad \hat{r} = \frac{\widehat{E}_{h_1}[1 - D|Z = c^+]}{\widehat{E}_{h_1}[1 - D|Z = c^-]}.$$

Then $G_1^{-1}(q)$ and $G_0^{-1}(r)$ are estimated by

$$\widehat{G}_1^{-1}(\hat{q}) = \inf\{y \in \mathcal{Y} : \widehat{G}_1(y) \geq \hat{q}\}, \qquad \widehat{G}_0^{-1}(\hat{r}) = \inf\{y \in \mathcal{Y} : \widehat{G}_0(y) \geq \hat{r}\}.$$

In the second step, θ_1, θ_2, θ_3 and θ_4 are estimated by

$$\hat{\theta}_1 = \widehat{E}_{h_2}[DY1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+] \cdot \widehat{E}_{h_2}[D|Z = c^-]$$

$$
\begin{aligned}
&\quad - \widehat{E}_{h_2}[DY|Z=c^-]\cdot\widehat{E}_{h_2}[D1(Y<\widehat{G}_1^{-1}(\hat{q}))|Z=c^+],\\
\hat{\theta}_2 &= \widehat{E}_{h_2}[DY|Z=c^-]\cdot\widehat{E}_{h_2}[D1(Y>\widehat{G}_1^{-1}(1-\hat{q}))|Z=c^+]\\
&\quad - \widehat{E}_{h_2}[DY1(Y>\widehat{G}_1^{-1}(1-\hat{q}))|Z=c^+]\cdot\widehat{E}_{h_2}[D|Z=c^-],\\
\hat{\theta}_3 &= \widehat{E}_{h_2}[(1-D)Y1(Y<\widehat{G}_0^{-1}(\hat{r}))|Z=c^-]\cdot\widehat{E}_{h_2}[1-D|Z=c^+]\\
&\quad - \widehat{E}_{h_2}[(1-D)Y|Z=c^+]\cdot\widehat{E}_{h_2}[(1-D)1(Y<\widehat{G}_0^{-1}(\hat{r}))|Z=c^-],\\
\hat{\theta}_4 &= \widehat{E}_{h_2}[(1-D)Y|Z=c^+]\cdot\widehat{E}_{h_2}[(1-D)1(Y>\widehat{G}_0^{-1}(1-\hat{r}))|Z=c^-]\\
&\quad - \widehat{E}_{h_2}[(1-D)Y1(Y>\widehat{G}_0^{-1}(1-\hat{r}))|Z=c^-]\cdot\widehat{E}_{h_2}[(1-D)|Z=c^+].
\end{aligned}
$$

Note that the two-step estimators for θ_j's involve two different bandwidths, h_1 and h_2. We use h_1 to estimate r, q, $G_0^{-1}(r)$, $G_0^{-1}(1-r)$, $G_1^{-1}(q)$ and $G_1^{-1}(1-q)$, and use h_2 to estimate the conditional means in the expressions of $\hat{\theta}_j$'s. As shown in details in Appendix B, we take h_1 to be a larger bandwidth than h_2 in that $h_2/h_1 \to 0$, so the estimation effects in the first step are asymptotically negligible when we derive the asymptotics of $\hat{\theta}_j$'s. In other words, we can treat $\widehat{G}_0^{-1}(\hat{r})$, $\widehat{G}_0^{-1}(1-\hat{r})$, $\widehat{G}_1^{-1}(\hat{q})$ and $\widehat{G}_1^{-1}(1-\hat{q})$ as true values when we estimate θ_j's. In the second step, we use a local quadratic regression to estimate the components in $\hat{\theta}_j$'s. By Remark 7 of Calonico, Cattaneo, and Titiunik (2014, CCT), a local quadratic regression estimator for a conditional mean is numerically equivalent to a local linear bias-corrected estimator when the pilot bandwidth is the same as h_2. The implication of these results is that the resulting $\hat{\theta}_j$'s are also bias-corrected estimators and we can set $h_2 = O(n^{-1/5})$. Note that the popular data-driven bandwidths proposed in the literature, such as Imbens and Kalyanaraman (2012, IK), CCT and Arai and Ichimura (2016, AI), all are of order $n^{-1/5}$ and satisfy this rate condition. For h_1, we suggest to set $h_1 = h_2 \cdot n^{1/5} \cdot n^{-1/6}$ and this ensures that $h_2/h_1 \to 0$. These requirements are summarized in Assumption B.5.

While we provide formal results for the rates of h_1 and h_2, we do not develop the optimal bandwidths for our case. There are four θ_j's, and each θ_j contains multiple conditional expectations. The linearized version of $\sqrt{nh_2}(\hat{\theta}_j - \theta_j)$ is not of the form $\sqrt{nh_2}(\widehat{E}_{h_2}[A|Z = c^+] - \widehat{E}_{h_2}[A|Z = c^-] - E[A|Z = c^+] - E[A|Z = c^-]) + o_p(1)$, so the results in CCT can not be applied directly. Furthermore, we are unaware of any existing results for optimal bandwidth choice for the specification test in such settings, and fully investigating it is not the focus of our paper. Therefore, we leave this for future research. In our simulation and empirical study, we try three different bandwidths (IK, CCT, AI) of order $n^{-1/5}$ and the testing results based on these choices are quite similar, so we expect that the results based on an optimal bandwidth for our case will be similar in practice.

In Appendix B, under suitable regularity conditions, we formally derive the asymptotic linear representations of the estimators $\sqrt{nh_2}(\hat{\theta}-\theta)$ where $\theta=(\theta_1,\theta_2,\theta_3,\theta_4)'$ and $\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4)'$. We also show the joint asymptotic normality of the estimators in that $\sqrt{nh_2}(\hat{\theta}-\theta)\xrightarrow{d}\mathcal{N}(0,\Omega)$ and Ω is a 4×4 asymptotic covariance matrix.

3.2 Weighted Bootstrap

The analytical form of the variance estimator of the proposed estimators is complicated to calculate even though we can ignore the estimation effect from the first step. To facilitate implementation, 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 under the assumptions in Appendix B. We consider two different approaches. In the first approach, we consider a traditional two-step bootstrap procedure in which we repeat the bootstrap estimation procedures in both steps. In the second approach, we consider a one-step bootstrap procedure in which we take the first-step estimators as given and only consider the bootstrap in the second-step estimation. The advantage of the second approach is that it is less time-consuming because we do not need to calculate the bootstrapped first-stage estimators in the bootstrap repetitions.

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 $\widehat{E}_h^w[A|Z=c^+]$ and $\widehat{E}_h^w[A|Z=c^-]$ be the weighted bootstrap local quadratic estimators with bandwidth h for $E[A|Z=c^+]$ and $E[A|Z=c^-]$, respectively. To be specific,

$$(\widehat{E}_h^w[A|Z=c^+],\hat{\beta}_+^w,\hat{\gamma}_+^w)=\underset{a,b,k}{\operatorname{argmin}}\sum_{i=1}^n W_i\cdot 1(Z_i\geq c)\cdot K\left(\frac{Z_i-c}{h}\right)\Big[A_i-a-b\cdot(Z_i-c)-k\cdot(Z_i-c)^2\Big]^2,$$

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

For the first bootstrap approach, let the weighted bootstrap estimators for $G_1(y)$ and $G_0(y)$ be

$$\widehat{G}_1^w(y) = \frac{\widehat{E}_{h_1}^w[D1(Y \leq y)|Z = c^+]}{\widehat{E}_{h_1}^w[D|Z = c^+]}, \quad \widehat{G}_0^w(y) = \frac{\widehat{E}_{h_1}^w[(1-D)1(Y \leq y)|Z = c^-]}{\widehat{E}_{h_1}^w[1-D|Z = c^+]}.$$

Let the weighted bootstrap estimators for q and r be

$$\hat{q}^w = \frac{\widehat{E}_{h_1}^w[D|Z = c^-]}{\widehat{E}_{h_1}^w[D|Z = c^+]}, \quad \hat{r}^w = \frac{\widehat{E}_{h_1}^w[1-D|Z = c^+]}{\widehat{E}_{h_1}^w[1-D|Z = c^-]}.$$

Let the weighted bootstrap estimators for $G_1^{-1}(q)$, $G_1^{-1}(1-q)$, $G_1^{-1}(r)$, $G_1^{-1}(1-r)$ be

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

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

$$\begin{aligned}
\hat{\theta}_1^w &= \widehat{E}_{h_2}^w[DY1(Y < \widehat{G}_1^{-1,w}(\hat{q}^w))|Z = c^+] \cdot \widehat{E}_{h_2}^w[D|Z = c^-] \\
&\quad - \widehat{E}_{h_2}^w[DY|Z = c^-] \cdot \widehat{E}_{h_2}^w[D1(Y < \widehat{G}_1^{-1,w}(\hat{q}^w))|Z = c^+], \\
\hat{\theta}_2^w &= \widehat{E}_{h_2}^w[DY|Z = c^-] \cdot \widehat{E}_{h_2}^w[D1(Y > \widehat{G}_1^{-1,w}(1-\hat{q}^w))|Z = c^+] \\
&\quad - \widehat{E}_{h_2}^w[DY1(Y > \widehat{G}_1^{-1,w}(1-\hat{q}^w))|Z = c^+] \cdot \widehat{E}_{h_2}^w[D|Z = c^-], \\
\hat{\theta}_3^w &= \widehat{E}_{h_2}^w[(1-D)Y1(Y < \widehat{G}_0^{-1,w}(\hat{r}^w))|Z = c^-] \cdot \widehat{E}_{h_2}^w[1-D|Z = c^+] \\
&\quad - \widehat{E}_{h_2}^w[(1-D)Y|Z = c^+] \cdot \widehat{E}_{h_2}^w[(1-D)1(Y < \widehat{G}_0^{-1,w}(\hat{r}^w))|Z = c^-], \\
\hat{\theta}_4^w &= \widehat{E}_{h_2}^w[(1-D)Y|Z = c^+] \cdot \widehat{E}_{h_2}^w[(1-D)1(Y > \widehat{G}_0^{-1,w}(1-\hat{r}^w))|Z = c^-] \\
&\quad - \widehat{E}_{h_2}^w[(1-D)Y1(Y > \widehat{G}_0^{-1,w}(1-\hat{r}^w))|Z = c^-] \cdot \widehat{E}_{h_2}^w[(1-D)|Z = c^+].
\end{aligned}$$

Next, we consider second bootstrap in which we do not use $\widehat{G}_1^{-1,w}(\hat{q}^w)$, $\widehat{G}_1^{-1,w}(1-\hat{q}^w)$, $\widehat{G}_0^{-1,w}(\hat{r}^w)$ and $\widehat{G}_0^{-1,w}(1-\hat{r}^w)$ in the second step; instead, we use $\widehat{G}_1^{-1}(\hat{q})$, $\widehat{G}_1^{-1}(1-\hat{q})$, $\widehat{G}_0^{-1}(\hat{r})$ and $\widehat{G}_0^{-1}(1-\hat{r})$ from the first step of the estimation of $\hat{\theta}$. To be specific, let

$$\begin{aligned}
\hat{\theta}_1^b &= \widehat{E}_{h_2}^w[DY1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+] \cdot \widehat{E}_{h_2}^w[D|Z = c^-] \\
&\quad - \widehat{E}_{h_2}^w[DY|Z = c^-] \cdot \widehat{E}_{h_2}^w[D1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+], \\
\hat{\theta}_2^b &= \widehat{E}_{h_2}^w[DY|Z = c^-] \cdot \widehat{E}_{h_2}^w[D1(Y > \widehat{G}_1^{-1}(1-\hat{q}))|Z = c^+]
\end{aligned}$$

$$
\begin{aligned}
&\quad - \widehat{E}^w_{h_2}[DY1(Y > \widehat{G}_1^{-1}(1-\hat{q}))|Z=c^+] \cdot \widehat{E}^w_{h_2}[D|Z=c^-], \\
\hat{\theta}^b_3 &= \widehat{E}^w_{h_2}[(1-D)Y1(Y < \widehat{G}_0^{-1}(\hat{r}))|Z=c^-] \cdot \widehat{E}^w_{h_2}[1-D|Z=c^+] \\
&\quad - \widehat{E}^w_{h_2}[(1-D)Y|Z=c^+] \cdot \widehat{E}^w_{h_2}[(1-D)1(Y < \widehat{G}_0^{-1}(\hat{r}))|Z=c^-], \\
\hat{\theta}^b_4 &= \widehat{E}^w_{h_2}[(1-D)Y|Z=c^+] \cdot \widehat{E}^w_{h_2}[(1-D)1(Y > \widehat{G}_0^{-1}(1-\hat{r}))|Z=c^-] \\
&\quad - \widehat{E}^w_{h_2}[(1-D)Y1(Y > \widehat{G}_0^{-1}(1-\hat{r}))|Z=c^-] \cdot \widehat{E}^w_{h_2}[(1-D)|Z=c^+].
\end{aligned}
$$

Under the same set of regularity conditions (see details in Appendix B), we derive the asymptotic linear representations of both weighted bootstrap estimators $\sqrt{nh_2}(\hat{\theta}^w - \hat{\theta})$ and $\sqrt{nh_2}(\hat{\theta}^b - \hat{\theta})$. We can also show that both $\widehat{\Phi}^w = \sqrt{nh_2}(\hat{\theta}^w - \hat{\theta})$ and $\widehat{\Phi}^b = \sqrt{nh_2}(\hat{\theta}^b - \hat{\theta})$ converge to the same limiting distribution as $\sqrt{nh_2}(\hat{\theta} - \theta)$ conditional on the same path with probability approaching one. That is, $\widehat{\Phi}^w$ and $\widehat{\Phi}^b$ can approximate the limiting distribution of $\sqrt{nh_2}(\hat{\theta} - \theta)$ well.

3.3 Test statistic, critical value and decision rule

We define the test statistic as

$$
\widehat{T}_n = \sqrt{nh_2} \max_{j=1,\dots,4} \frac{\hat{\theta}_j}{\hat{\sigma}_j}, \tag{3.6}
$$

where $\hat{\sigma}_j$ is a consistent estimator for σ_j, the square root of the asymptotic variance of $\sqrt{nh_1}(\hat{\theta}_j - \theta_j)$ for $j = 1, 2, 3, 4$. For $\hat{\sigma}_j$, we suggest using the weighted bootstrap estimators. To be specific, let $t = 1, \dots, T$ and say $T = 1000$. Then for each t, we get $\hat{\theta}^{w,t}$ and $\hat{\theta}^{b,t}$. Let $\hat{\sigma}^w_j = \sqrt{nh_2}\big(T^{-1}\sum_{t=1}^T(\hat{\theta}^{w,t}_j - \hat{\theta}_j)^2\big)^{1/2}$ and $\hat{\sigma}^b_j = \sqrt{nh_2}\big(T^{-1}\sum_{t=1}^T(\hat{\theta}^{b,t}_j - \hat{\theta}_j)^2\big)^{1/2}$. Then $\hat{\sigma}_j$ can be $\hat{\sigma}^w_j$ or $\hat{\sigma}^b_j$.

Define the recentering parameters $\hat{\mu}_j$'s as $\hat{\mu}_j = \hat{\theta}_j \cdot 1(\sqrt{nh_2}\hat{\theta}_j \le -a_n\hat{\sigma}_j)$ where a_n is sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} a_n/\sqrt{nh_2} = 0$. For significance level $\alpha < 1/2$, define the critical value as $\hat{c}^w_n(\alpha) = \max\{\tilde{c}^w_n(\alpha), 0\}$ and $\hat{c}^b_n(\alpha) = \max\{\tilde{c}^b_n(\alpha), 0\}$ where $\tilde{c}^w_n(\alpha)$ and $\tilde{c}^b_n(\alpha)$ are defined as

$$
\begin{aligned}
\tilde{c}^w_n(\alpha) &= \inf_c \Big\{c:\ P\Big(\max_{j=1,\dots,4}\Big\{\frac{\hat{\phi}^w_j + \sqrt{nh_2}\hat{\mu}_j}{\hat{\sigma}_j}\Big\} \le c\Big) \ge 1-\alpha\Big\}, \\
\tilde{c}^b_n(\alpha) &= \inf_c \Big\{c:\ P\Big(\max_{j=1,\dots,4}\Big\{\frac{\hat{\phi}^b_j + \sqrt{nh_2}\hat{\mu}_j}{\hat{\sigma}_j}\Big\} \le c\Big) \ge 1-\alpha\Big\}.
\end{aligned}
$$

The **decision rule** will be "Reject H_0 if $\widehat{T}_n > \hat{c}_n(\alpha)$," where $\hat{c}_n(\alpha)$ can be $\hat{c}_n^w(\alpha)$ or $\hat{c}_n^b(\alpha)$.

Proposition 3.3 *Suppose that Assumptions B.1 to B.8 in Appendix B hold and let* $0 < \alpha < 1/2$. *Then under* H_0 *in (3.5),* $\lim_{n\to\infty} P(\widehat{T}_n > \hat{c}_n(\alpha)) \leq \alpha$*; under* H_1*,* $\lim_{n\to\infty} P(\widehat{T}_n > \hat{c}_n(\alpha)) = 1$.

For the 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 $a_n = \sqrt{2 \log \log n}$ as in Donald and Hsu (2011) or $a_n = \sqrt{0.3 \log n}$ as in Andrews and Shi (2013).

Remark 3.4 *Another possible way to construct our test is to estimate the quantities in the inequalities in (2.7), (2.8), (2.9) and (2.10) by the truncated conditional mean estimators proposed by Olma (2021), which is based on the Neyman-orthogonal moment. However, we consider the advantage of our estimator is that if the quantile index in the truncated conditional mean is close to zero, then the estimation can be less stable due to the fact that the estimated quantile index is in the denominator. In our estimation, there is no fractional expression in the* θ_j*'s.*

4 Simulation

4.1 Baseline Simulation

In this section, we consider a few numerical examples to illustrate the performance of our procedure. For comparison purposes, 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 assumptions 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 report results based on three sample sizes $n \in \{1000, 2000, 4000\}$, 1000 bootstrap draws, 800 replications, and significance level $\alpha = 5\%$. We set $a_n = \sqrt{2 \log \log n}$ following Donald and Hsu (2016). For the bandwidth, as mentioned in Section 3, we choose $h_1 = kn^{-\frac{1}{6}}$ and $h_2 = kn^{-\frac{1}{5}}$, where the constant k are taken as the constants from three data-driven choices of bandwidths: IK, CCT and AI.

Table 1 reports the results using $\hat{\theta}^b$. The results of using $\hat{\theta}^w$ are qualitatively similar and therefore omitted to save space. For the same reason, we only report the results at 5% level. The northwest 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 ($n = 1000$), 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. The remaining three panels 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. The rejection rate being smaller than the nominal level is due to some inequality are far away from binding, but the moment selection procedure does not completely discard them in finite sample. This is a common phenomenon for inference in inequality models. We do observe that, as the sample size increases, the rejection rates get closer to the nominal rate. In contrast, the test of 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.

Table 1: Rejection Rate at 5% Level (based on $\hat{\theta}^b$)

n	DGP1 (power) IK	DGP1 (power) CCT	DGP1 (power) AI	DGP2 (size) IK	DGP2 (size) CCT	DGP2 (size) AI
1000	0.253	0.1050	0.197	0.018	0.008	0.010
2000	0.572	0.398	0.475	0.027	0.015	0.015
4000	0.873	0.718	0.797	0.023	0.035	0.030

n	DGP3 (size) IK	DGP3 (size) CCT	DGP3 (size) AI	DGP4 (size) IK	DGP4 (size) CCT	DGP4 (size) AI
1000	0.000	0.000	0.013	0.018	0.003	0.020
2000	0.022	0.018	0.027	0.025	0.020	0.015
4000	0.038	0.023	0.028	0.033	0.020	0.028

4.2 Further Power Analysis

When implementing our test, the non-rejection can result from multiple factors. At the population level, it can be the case that the FRD mean assumptions hold, or the FRD mean assumptions do not hold, but the sharp testable implications hold. At the finite sample level, it could be the case that the sample size is small relative to the magnitude of the violation so that the null can not be rejected. In this subsection, we further examine a few factors that affect the finite sample power of our test.

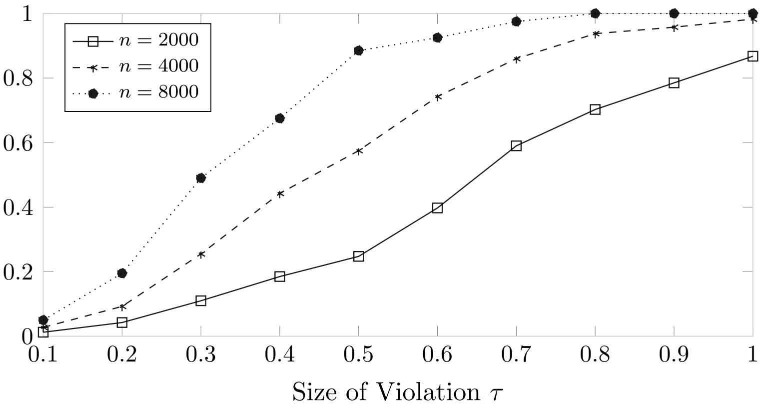
4.2.1 Magnitude of violation

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

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

In this design, τ measures the ratio of the violation size over the standard deviation of the potential outcomes. Figure 2 plots the rejection rate of using IK bandwidth constant at different values of τ and sample size n. The results of using other bandwidths are similar. When $\tau=0$, 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 τ increases, the magnitude of violation is larger, and we expect to see the rejection rate increase. For example, when the size of the violation is about

Figure 2: Rejection frequency and the magnitude of violation (IK, $\alpha = 5\%$)

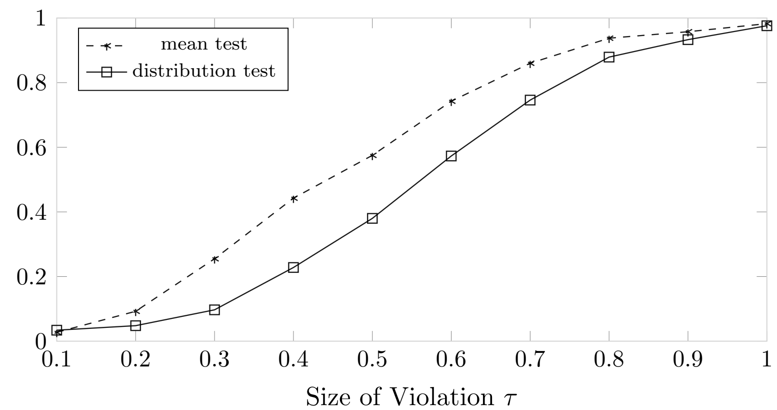


half of the standard deviation, we already see a quite large rejection frequency at reasonable sample sizes.

We also conduct the distribution test of Arai, Hsu, Kitagawa, Mourifié, and Wan (2022) using DGP5. Figure 3 plots the rejection frequency of the mean and distribution tests at the sample size $n = 4000$. For the distribution test, we use the under-smoothed IK bandwidth. The mean test curve is copied from Figure 2 for the convenience of comparison. It lies above the rejection curve of the distribution test for all the violation magnitude. It is interesting to note that in DGP5, it is the location shift that drives the discontinuity in the distribution, and the test designed for testing the mean discontinuity works better. Of course, this is just one simulation study and there is not enough information to draw a more general conclusion. However, it does suggest that the mean test can perform better in the finite sample over a reasonable class of DGPs, particularly when the violation is on the means. If the violation is on other moments like DGP2-4, the mean test will not have power, but the distribution test is still consistent against those alternatives.

As our next simulation study, we examine how well our test performs in more realistic data scenarios. As the benchmark, we use the data from Battistin, Brugiavini, Rettore, and Weber (2009), who study the effect of retirement on consumption. As we will show later in Section 5, neither our test nor the distributional test rejects the validity of the FRD design in Battistin, Brugiavini, Rettore, and Weber (2009). Hence, we proceed to estimate the distribution of $Y(1)$ for always-takers and compliers, and the distribution of $Y(0)$ for never-takers and compliers

Figure 3: Rejection frequency and the magnitude of violation (IK, $\alpha = 5\%$)

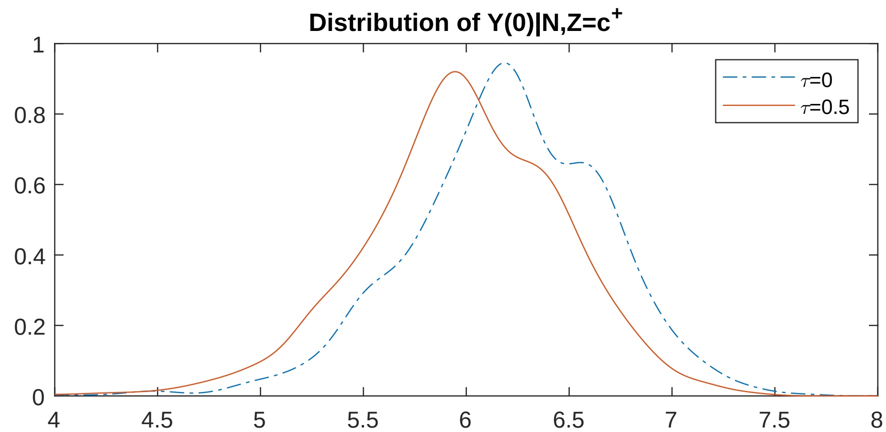


under the FRD distribution assumptions. Next, we artificially distort the mean of the estimated distribution so that it violates the FRD mean assumptions and check the magnitude of the distortion so that our test can reject. We chose this data set and the outcome variable of "food consumption" (logs) to implement the exercise because it contains many observations, and the distribution of outcome variables is approximately normal so that we can compare it with our simulation design DGP5. The same message is obtained from other outcome variables.

To be more specific, we shifted never-taker's $Y(0)$ distribution given $Z = c^+$ to the left by the amount of $\tau\sigma_Y$, where σ_Y is the standard deviation of Y in data and $\tau \in \{0.1, 0.2, \cdots, 1.0\}$. So τ measures the distortion's magnitude relative to the observed outcome's standard deviation. Please see an illustration from Figure 4 with $\tau = 0.5$, where the dotted line is the pdf of the estimated distribution of $Y(0)|\boldsymbol{N}, Z = c^+$, and the solid line is the pdf of distorted distribution.

Figures 5a and 5b report the p-values for different distorted DGPs under different bandwidth choices. Note that the bandwidth from the original Battistin, Brugiavini, Rettore, and Weber (2009) paper was $h = 10$. There are a total of 30,689 observations, and 9,804 of them are within $[c-10, c+10]$ interval. For robustness concerns, we also consider bandwidths $h \in \{8, 9, 10, 11, 12\}$. Not surprisingly, the p-values are decreasing in the distortion magnitude for all bandwidth choices. When the distortion reaches roughly 0.5 times the standard deviation of Y, our test starts to reject at 5% level. Again, these exercises show that because our test only tests necessary conditions, it can not screen out all possible violations; however, it does have power when the magnitude of the violation is reasonably large. This result matches our observation from the

Figure 4: Distorting the potential outcome distribution at $\tau = 0.5$



simulation results based on DGP5 above.

4.2.2 Size of the propensity jump

Next, we consider DGP6, where the propensity score's jump size π takes different values.

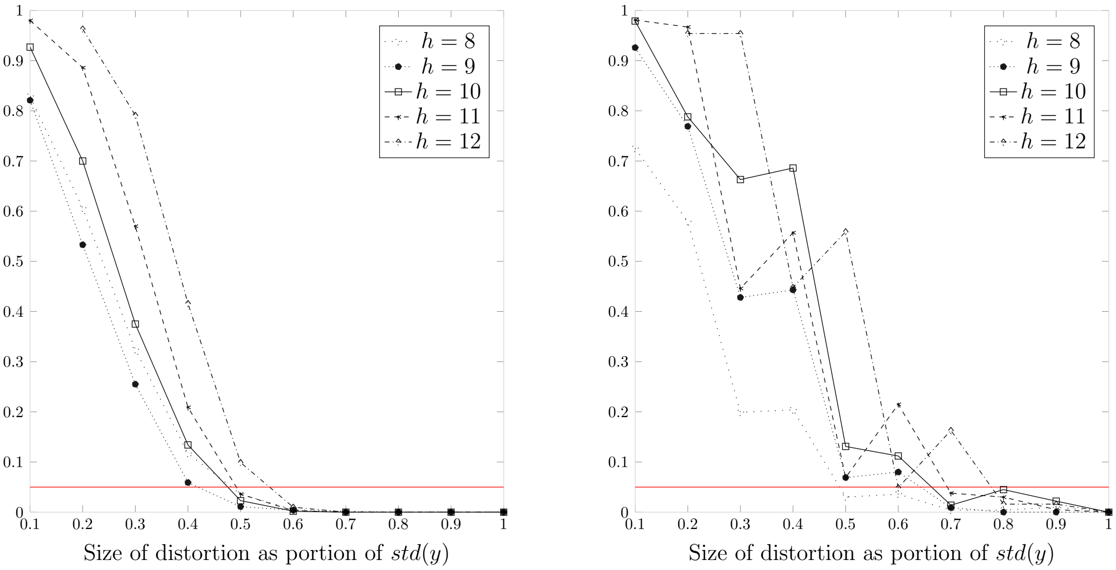
DGP6 The same as DGP1 except that

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

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 6 verifies this point for bandwidth IK-US and sample size $n = 8000$. Again, the results for other sample sizes and bandwidths are qualitatively similar.

4.2.3 Non-constant regression function

In our baseline simulation, for both size and power designs, the regression function of the potential outcomes on the running variable is constant on both sides of the cutoff. This subsection

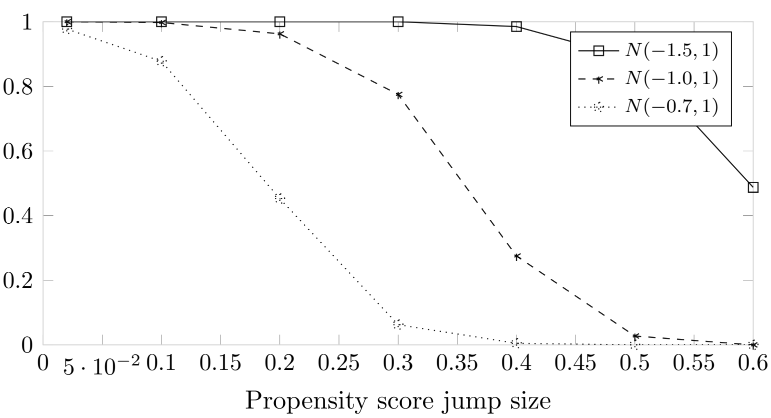


(a) Mean test

(b) Distribution test

Figure 5: p-values and data distortion

Figure 6: Rejection frequency and propensity score jump size (IK, $\alpha = 5\%$)



considers DGPs in which $\mathbb{E}[Y(d)|Z=z]$ are polynomials of z.

DGP7 Same as DGP1 expect that $Y|(D=1, Z=z) \sim N(\delta_1 z + \delta_2 z^2, 1)$ for all $z \in [0,2]$, and $Y|(D=1, Z=z) \sim N(-0.7 + \pi_1 z + \pi_2 z^2, 1)$ for all $z \in [-2, 0)$

DGP8 Same as DGP7 except that $Y|(D=1, Z=z) \sim N(\pi_1 z + \pi_2 z^2, 1.675^2)$ for all $z \in [-2, 0)$.

DGP9 Same as DGP7 except that $Y|(D=1, Z=z) \sim N(\pi_1 z + \pi_2 z^2, 0.515^2)$ for all $z \in [-2, 0)$.

DGP10 Same as DGP7 except that $Y|(D=1, Z=z) \sim \sum_{j=1}^{5} \omega_j N(\mu_j + \pi_1 z + \pi_2 z^2, 0.125^2)$ for all $z \in [-2, 0)$, where $\omega = (0.15, 0.2, 0.3, 0.2, 0.15)$ and $\mu = (-1, -0.5, 0, 0.5, 1)$.

In these designs, we set $\pi_1 = \pi_2 = \delta_1 = \delta_2 = 1$. DGP7 is the power design because the conditional expectation of $Y(1)$ given $Z = z$ is discontinuous at the cutoff. DGP8-DGP10 are size designs because the conditional expectation of $Y(1)$ is continuous despite not being a constant function of z.

Table 2: Rejection Rate with non-constant regression function ($\hat{\theta}^b$, 5%)

	DGP7 (power)			DGP8 (size)		
n	IK	CCT	AI	IK	CCT	AI
1000	0.126	0.066	0.202	0.008	0.006	0.024
2000	0.364	0.282	0.372	0.026	0.018	0.022
4000	0.666	0.658	0.776	0.022	0.036	0.032

	DGP9 (size)			DGP10 (size)		
n	IK	CCT	AI	IK	CCT	AI
1000	0.012	0.004	0.018	0.004	0.006	0.014
2000	0.016	0.020	0.024	0.012	0.016	0.022
4000	0.024	0.024	0.020	0.030	0.028	0.028

4.2.4 Discrete running variable

Last, we test the performance of our test when the running variable is discrete. For this, we reuse DGP1-DGP4 but change them such that the running variable is generated by

$$z = \frac{1}{10}\text{floor}(10z^*),$$

where z^* is generated from truncated standard normal on $[-2, 2]$ and floor($\cdot$) is the floor function that rounds a real number down to the maximum integer. So z has mass points on $\{-2, -1.9, \cdots, 1.9\}$. The rest of the design remains to be the same as DGP1-DGP4. We implement the test by ignoring the fact that the running variable is discrete (so that we still apply local polynomial regression and use bandwidths h_1 and h_2). The results are reported in Table 3. We can see that the conclusion is qualitatively similar to the baseline design.

Table 3: Rejection rate with discrete running variable ($\hat{\theta}^w$, 5%)

	DGP1-discrete (power)			DGP2-discrete (size)		
n	IK	CCT	AI	IK	CCT	AI
1000	0.2450	0.1175	0.2000	0.0225	0.0125	0.0325
2000	0.5475	0.4450	0.4225	0.0275	0.0225	0.0250
4000	0.8175	0.6425	0.6950	0.0325	0.0275	0.0225

	DGP3-discrete (size)			DGP4-discrete (size)		
n	IK	CCT	AI	IK	CCT	AI
1000	0.0150	0.0175	0.0075	0.0050	0.0225	0.0300
2000	0.0275	0.0150	0.0250	0.0275	0.0225	0.0300
4000	0.0250	0.0275	0.0225	0.0200	0.0100	0.0275

5 Empirical Application

In this section, we illustrate the use of our method in a few empirical applications.[9]

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

[9]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.

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 find 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.

Table 4: Testing Results for Columbia's SR Data: p-values

Bandwidth/constant h	Household edu.exp.	Total exp.on food	Total monthly exp.
2	0.838	0.696	0.178
3	0.988	0.779	0.592
4	0.465	0.555	0.766
IK	0.541	0.507	0.721
CCT	0.527	0.514	0.879
AI	0.696	0.553	0.685

However, discontinuity in the conditional distribution does not necessarily imply a discontinuity in the conditional expectation. Table 4 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). We observe no rejection across the board, even at the 10% level. While our test is designed for the necessary (but not sufficient) implications of the FRD mean assumptions and non-rejection may also be caused by the sample size being too small for the mean test to reject, this result can be considered supportive evidence for credibly estimating the mean effect.

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

	g4math			g4verb		
Bandwidth/constant	40	80	120	40	80	120
5	0.883	0.471	0.817	0.775	0.914	0.802
IK	0.459	0.557	0.209	0.384	0.845	0.763
CCT	0.505	0.988	0.988	0.852	0.985	0.997
AI	0.481	0.958	0.612	0.219	0.929	0.574

In this subsection, we revisit this empirical question. At the population level, if the data distribution satisfies the sharp testable implications of the distributional assumption, an observationally equivalent potential distribution exists that satisfies the FRD distributional assumptions

Table 6: Testing Results for Israeli School Data (Grade 5): p-values

	g5math			g5verb		
Bandwidth/constant	40	80	120	40	80	120
5	0.909	0.646	0.985	0.685	0.739	0.988
IK	0.764	0.349	0.781	1.000	0.958	0.666
CCT	0.997	0.792	0.601	0.993	0.975	0.658
AI	0.908	0.998	0.999	0.992	0.997	0.975

and hence satisfies FRD mean assumptions. Therefore, the data distribution must also satisfy the testable implications in this paper. However, the conclusions may differ in the finite sample due to sampling error. In this example, the two tests agree with each other. As reported in Tables 5 and 6, 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 also discussions about running variable manipulation in Gerard, Rokkanen, and Rothe (2020). 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 7. 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.[10]

Table 7: Testing Results for Romania High School Data: p-values

	Attending best school		Avoiding worst school	
Bandwidth	Mean Test	Distr. Test	Mean Test	Distr. Test
0.100	0.566	0.900	1.000	0.271
0.200	0.832	0.999	0.466	0.838
0.300	0.806	1.000	0.907	0.967
IK	0.518	0.635	0.564	0.815
CCT	0.502	0.588	0.612	0.786
AI	0.512	0.590	0.983	0.848

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

[10] 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.

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 identifies 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 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 bandwidth 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 evidence against the validity of the FRD design (either for the mean assumptions or for distributional assumptions).

Table 8: 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	0.705	0.768	0.982	0.709	0.831	0.121
4	1.000	1.000	0.960	0.941	0.883	0.087
5	1.000	1.000	0.884	0.970	0.940	0.106
6	0.988	0.819	0.928	0.998	0.491	0.067
7	0.748	0.813	0.932	0.990	0.382	0.358
8	0.789	0.771	0.938	0.911	0.495	0.884
9	0.926	0.944	0.936	0.904	0.875	0.988
10	0.967	0.933	0.973	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 or a mixture of continuous and discrete parts. In practice, it is not unreasonable to treat Y as continuous if its support contains a large number of points. For example, although the exam score only takes integer values, it is often treated as continuous. In such cases, one can apply our results in Proposition 2.2. Alternatively, one can also similarly derive the sharp bounds, as shown in the following corollary.

Corollary 6.1 *Suppose Assumptions 2.1 and 2.2 are satisfied. (i) 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.5) and Equation (A.8).

(ii) If $Y(d)$'s distribution contains both a continuous part and mass points, then depending on the location of the mass points, the sharp bounds take either the form of the continuous case or the discrete case, and are reported in (A.10) and (A.11) in the appendix.

Example 6.2 *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 2.7 and 2.8 would simply imply:*

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

However, the bounds derived in Equations (6.1) and (6.2) 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*. 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* $E[Y|D = 1, Z = c^-]$ *is nontrivial 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) $\lim_{\varepsilon \downarrow 0} P(T_\varepsilon \in \{\boldsymbol{DF}, \boldsymbol{I}\}|Z = c + \varepsilon, X = x) = 0$ *and* $\lim_{\varepsilon \downarrow 0} P(T_\varepsilon \in \{\boldsymbol{DF}, \boldsymbol{I}\}|Z = c - \varepsilon, X = x) = 0$ *for all* $x \in \mathcal{X}$*.*

Let $f_{Y(d)|T_\varepsilon, Z, X}(y|t, z, x)$ be the conditioning density of $Y(d)$ given $T_\varepsilon = t$, $Z = z$, and $X = x$.

Assumption 6.2 (Conditional local continuity in means) *For all* $x \in \mathcal{X}$*,* *(i)* $\lim_{\varepsilon\downarrow 0} f_{Y(d)|T_\varepsilon,Z,X}(y|t, c-\varepsilon, x)$ *and* $\lim_{\varepsilon\downarrow 0} f_{Y(d)|T_\varepsilon,Z,X}(y|t, c+\varepsilon, x)$ *are proper densities and admits finite expectations for* $|Y(d)|$.

(ii) $\lim_{\varepsilon\downarrow 0} E[Y(d)|T_\varepsilon = t, Z = c-\varepsilon, X = x] = \lim_{\varepsilon\downarrow 0} E[Y(d)|T_\varepsilon = t, Z = c+\varepsilon, X = x]$ *and* $\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = t|Z = c-\varepsilon, X = x) = \lim_{\varepsilon\downarrow 0} P(T_\varepsilon = t|Z = c+\varepsilon, X = x)$

Let $G_{1x}(y) = \lim_{z\downarrow c} P(Y \le y|D = 1, Z = z, X = x)$ be the conditional distribution of Y given $D = 1$, $Z = z$, $X = x$ when z converges to c from above. Similarly, define $G_{0x}(y) = \lim_{z\uparrow c} P(Y \le y|D = 0, Z = z, X = x)$. 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, for all* $x \in \mathcal{X}$, $q_x \in (0, 1)$, *and* $r_x \in (0, 1)$, *and the distributions of* Y *given* $(D = 1, Z = c^+, X = x)$, *and* Y *given* $(D = 0, Z = c^-, X = x)$ *are continuous, then the following inequality constraints hold:*

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

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

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

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

And the bounds for $(E[Y|D = 1, Z = c^-], E[Y|D = 0, Z = c^+])$ *in Proposition 2.2 can be tightened as*

$$\overline{LB}^+ \le E[Y|D = 1, Z = c^-] \le \overline{UB}^+ \tag{6.7}$$

$$\overline{LB}^- \le E[Y|D = 0, Z = c^+] \le \overline{UB}^- \tag{6.8}$$

where

$$\begin{aligned}
\overline{LB}^+ &= \int E[Y|D = 1, Y < G_{1x}^{-1}(q_x), Z = c^+, X = x] dH_{X|D,Z}(x|1, z^-) \\
\overline{UB}^+ &= \int E[Y|D = 1, Y > G_{1x}^{-1}(1 - q_x), Z = c^+, X = x] dH_{X|D,Z}(x|1, z^-), \\
\overline{LB}^- &= \int E[Y|D = 1, Y < G_{0x}^{-1}(r_x), Z = c^-, X = x] dH_{X|D,Z}(x|0, z^+) \\
\overline{UB}^- &= \int E[Y|D = 1, Y > G_{0x}^{-1}(1 - r_x), Z = c^-, X = x] dH_{X|D,Z}(x|0, z^+).
\end{aligned}$$

and $H_{X|D,Z}(x|d,z^+)$ and $H_{X|D,Z}(x|d,z^-)$, respectively, are the limits of the conditioning distribution of X given D and $Z=z$ when z approaches to c from above and below, respectively.

The bounds in (6.7) and (6.8) are tighter than those in Proposition 2.2 because the truncated mean of the lower tail of the conditioning distribution (on X) is necessarily larger than the truncated mean of the unconditioning distribution. This result shares the same spirit of Lee (2009, Proposition 1b) and Gerard, Rokkanen, and Rothe (2020, Corollary 6), who also construct bounds by conditioning on covariates and then integrated out X to obtain tighter bounds for (reconditioning) treatment effect.

We can transform these inequalities as in Section 2 to implement the testable implication. Take inequality (6.3) as an example. It implies that

$$\begin{aligned}
&E[Y|D=1, Y<G_1^{-1}(q), Z=c^+, X=x] - E[Y|D=1, Z=c^-, X=x] \leq 0 \\
\Leftrightarrow\ & \frac{E[DY1(Y<G_1^{-1}(q))|Z=c^+, X=x]}{E[D1(Y<G_1^{-1}(q))|Z=c^+, X=x]} - \frac{E[DY|Z=c^-, X=x]}{E[D|Z=c^-, X=x]} \leq 0 \\
\Leftrightarrow\ & \theta_1(x) \equiv E[DY1(Y<G_1^{-1}(q))|Z=c^+, X=x]\cdot E[D|Z=c^-, X=x] \\
&\qquad - E[DY|Z=c^-, X=x]\cdot E[D1(Y<G_1^{-1}(q))|Z=c^+, X=x] \leq 0. \qquad (6.9)
\end{aligned}$$

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

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

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. Alternatively, one can work with inequalities (6.7) and (6.8).

Lastly, if the covariates X are predetermined and has conditional distribution given $Z=z$ that is continuous at the cutoff, then it must be the case that the expectation of X for always-takers is bounded in the same manner as $Y(1)$ in the Proposition 2.2, that is,

$$\begin{aligned}
&E[X|D=1, X<G_{X1}^{-1}(q), Z=c^+] \leq E[X|D=1, Z=c^-], \\
&E[X|D=1, Z=c^-] \leq E[X|D=1, Y>G_{X1}^{-1}(1-q), Z=c^+].
\end{aligned}$$

where G_{1X} is defined in the same as G_1 but with X replacing Y. Similar bounds can be created for the expectation of X for never-takers. These inequalities can enhance the screening power in the presence of pre-determined covariates X. We leave these discussions for future studies.

7 Conclusion

This paper proposes a specification test for researchers interested in estimating the mean causal effect for compliers in FRD designs and complements the existing tests on distributions. The test is easy to implement, has the asymptotic size control under the null and is consistent against all fixed alternatives that violate the testable implication. 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 Proposition 2.2

The proof follows the approach of Horowitz and Manski (1995) and Lee (2009). Part (i) shows that inequalities (2.7) to (2.10) are necessary conditions, and part (ii) shows that they are best to detect the violation of FRD mean assumptions.

Part (i). We prove the first pair of inequalities (2.7) and (2.8); the other two hold analogously. Inequalities (2.7) and (2.8) provide bounds for $E[Y(1)|D=1, Z=c^-]$.

Let $\boldsymbol{DF}^\dagger$ denote the combination of $\boldsymbol{DF}$ and $\boldsymbol{I}$. Let $\Omega_\varepsilon^\dagger = \{\omega : T_\varepsilon(\omega) = \boldsymbol{DF}^\dagger\}$. By definition of $\boldsymbol{DF}^\dagger$, it must be case that $\Omega_{\varepsilon_1}^\dagger \subseteq \Omega_{\varepsilon_2}^\dagger$ for any $\varepsilon_1 \leq \varepsilon_2$. To see this, let $\omega^* \in \Omega_{\varepsilon_1}^\dagger$. If $T_{\varepsilon_1}(\omega^*) = \boldsymbol{I}$, then $D(z, \omega^*)$ is neither constant over $[c-\varepsilon_1, c)$ nor over $[c, c+\varepsilon_1]$, so it must be nonconstant over $[c-\varepsilon_2, c)$ or $[c, c+\varepsilon_2]$ as well, hence $T_{\varepsilon_2}(\omega^*) = \boldsymbol{I}$ and $\omega^* \in \Omega_{\varepsilon_2}^\dagger$. If $T_{\varepsilon_1}(\omega^*) = \boldsymbol{DF}$, then $D(z, \omega^*) = 1\{z < c\}$ for all $z \in B_{\varepsilon_1}$. If $D(z, \omega^*) = 1\{z < c\}$ for all $z \in B_{\varepsilon_2}$ as well, then $T_{\varepsilon_2}(\omega^*) = \boldsymbol{DF}$, hence $\omega^* \in \Omega_{\varepsilon_2}^\dagger$. If not, then $D(z, \omega^*)$ is a nonconstant function on either $[c-\varepsilon_2, c)$ or $[c, c+\varepsilon_2]$, so $T_{\varepsilon_2}(\omega^*) = \boldsymbol{I}$, hence $\omega^* \in \Omega_{\varepsilon_2}^\dagger$. Regardless which case, if we take an arbitrary $\omega^* \in \Omega_{\varepsilon_1}^\dagger$, it must be that $\omega^* \in \Omega_{\varepsilon_2}^\dagger$.

Note that the event of $(D = 1, Z = c - \varepsilon)$ is equivalent to $(T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF}^\dagger\}, Z = c - \varepsilon)$. We will show that $E[Y|D = 1, Z = c^-]$ can be written as

$$
\begin{aligned}
E[Y|D=1, Z=c^-] &= \lim_{z \uparrow c} E[Y|D=1, Z=z] \\
&= \lim_{\varepsilon \downarrow 0} E[Y(1)|D=1, Z=c-\varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y(1)|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF}^\dagger\}, Z=c-\varepsilon] \\
&= \lim_{\varepsilon \downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z=c+\varepsilon]. \quad \text{(A.1)}
\end{aligned}
$$

The first three equalities hold by definition. We will verify the last one. In the first case, if there exists an $\bar{\varepsilon}$ such that $\Omega_{\bar{\varepsilon}}^\dagger = \emptyset$, then $\Omega_\varepsilon^\dagger = \emptyset$ for all $\varepsilon < \bar{\varepsilon}$. In this case, there is no type $\boldsymbol{DF}^\dagger$ near the cutoff, so Equation (A.1) holds because

$$
\lim_{\varepsilon \downarrow 0} E[Y(1)|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF}^\dagger\}, Z=c-\varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z=c-\varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z=c+\varepsilon]
$$

by Assumption 2.2. Here $\boldsymbol{DF}^\dagger$ drops out because $\Omega_\varepsilon^\dagger = \emptyset$ for $\varepsilon < \bar{\varepsilon}$, and the conditional expectation is well defined because we assume $q > 0$ so always takers exist.

In the second case, if there does not exist an $\bar{\varepsilon}$ such that $\Omega^{\dagger}_{\bar{\varepsilon}} = \emptyset$, then

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF^\dagger}\}, Z = c-\varepsilon] \\
&\qquad = \lim_{\varepsilon\downarrow 0} \big\{E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c-\varepsilon]P(T_\varepsilon = \boldsymbol{A}|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF^\dagger}\}, Z = c-\varepsilon) \\
&\qquad + E[Y(1)|T_\varepsilon = \boldsymbol{DF^\dagger}, Z = c-\varepsilon]P(T_\varepsilon = \boldsymbol{DF^\dagger}|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF^\dagger}\}, Z = c-\varepsilon)\big\} \\
&\qquad = \lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c-\varepsilon] = \lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon], \quad \text{(A.2)}
\end{aligned}$$

where the second equality holds because $\lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{DF^\dagger}, Z = c-\varepsilon]$ exists and is well-defined by Assumption 2.2-(i), $\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = \boldsymbol{DF^\dagger}|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{DF^\dagger}\}, Z = c-\varepsilon) = 0$ under Assumption 2.1 (local montononicity) and that always takers exist ($q > 0$). The third equality holds by Assumption 2.2-(ii) (local continuity in means). Hence, the bounds for $\lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon]$ is equivalent to the bounds for $E[Y|D = 1, Z = c^-]$.

Likewise, under Assumptions 2.1 and 2.2,

$$q \equiv \frac{P(D=1|Z=c^-)}{P(D=1|Z=c^+)} = \frac{\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = \boldsymbol{A}|Z = c-\varepsilon)}{\lim_{\varepsilon\downarrow 0} P(T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}\}|Z = c+\varepsilon)} = \frac{\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = \boldsymbol{A}|Z = c+\varepsilon)}{\lim_{\varepsilon\downarrow 0} P(T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}\}|Z = c+\varepsilon)}. \quad \text{(A.3)}$$

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

$$\begin{aligned}
G_1(y) &\equiv \lim_{z\downarrow c} P(Y \le y|D = 1, Z = z) = \lim_{\varepsilon\downarrow 0} P(Y(1) \le y|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{I}\}, Z = c+\varepsilon) \\
&= \lim_{\varepsilon\downarrow 0} \big\{P(Y(1) \le y|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon)P(T_\varepsilon = \boldsymbol{A}|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{I}\}, Z = c+\varepsilon) \\
&\quad + P(Y(1) \le y|T_\varepsilon = \boldsymbol{C}, Z = z+\varepsilon)P(T_\varepsilon = \boldsymbol{C}|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{I}\}, Z = c+\varepsilon) \\
&\quad + P(Y(1) \le y|T_\varepsilon = \boldsymbol{I}, Z = z+\varepsilon)P(T_\varepsilon = \boldsymbol{I}|T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{I}\}, Z = c+\varepsilon)\big\} \\
&\quad = \lim_{\varepsilon\downarrow 0} P(Y(1) \le y|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon)q + \lim_{\varepsilon\downarrow 0} P(Y(1) \le y|T_\varepsilon = \boldsymbol{C}, Z = c+\varepsilon)(1-q)
\end{aligned}$$

where the first equality is by definition, the second equality is by definition of the potential outcome and the fact that when z approaches to c from above, the event $\{D = 1, Z = c+\varepsilon\}$ is equivalent to the event $\{T_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{I}\}, Z = c+\varepsilon\}$, the third equality is by the law of total probabilities, Assumption 2.1, and that all the probabilities are well-defined. 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. Note if there exists $\bar{\varepsilon}$ such that the set $\{\omega : T_{\bar{\varepsilon}}(\omega) = \boldsymbol{I}\} = \emptyset$, the same conclusion follows, as we what we discussed for the right-hand side of Equation (A.1).

Now we characterize the bounds of expectation of the mixing component $\lim_{\varepsilon\downarrow 0} P(Y(1) \le y|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon)$. Since the conditional distribution of $Y(1)$ given $D = 1$ and $Z = z$ is continuous in y at its q-th quantile, we can apply Horowitz and Manski (1995, Corollary 4.1), and it follows that the sharp bounds for $\lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon]$ are given by

$$LB^+ \equiv E[Y|D=1, Y < G_1^{-1}(q), Z = c^+] \leq \lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon],$$
$$\lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon] \leq E[Y|D=1, Y > G_1^{-1}(1-q), Z = c^+] \equiv UB^+,$$

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 $\lim_{\varepsilon\downarrow 0} E[Y(1)|T_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon]$ by $E[Y|D=1, Z=c^-]$, we obtain inequalities (2.7) and (2.8).

Part (ii). Now suppose that the observed data distribution satisfies inequalities (2.7) to (2.10). We will show that there exists a distribution of latent variables such that the FRD mean assumptions hold, and is observationally equivalent to the observed data distribution. Our construction extends the argument of Laffers and Mellace (2017) to the FRD setup.

Given the inequalities (2.7) to (2.10), let λ_1 and λ_0 be such that

$$E[Y|D=1, Z=c^-] = \lambda_1 E[Y|D=1, Y < G_1^{-1}(q), Z=c^+] + (1-\lambda_1) E[Y|D=1, Y > G_1^{-1}(1-q), Z=c^+].$$

$$E[Y|D=0, Z=c^+] = \lambda_0 E[Y|D=0, Y < G_0^{-1}(r), Z=c^-] + (1-\lambda_0) E[Y|D=0, Y > G_0^{-1}(1-r), Z=c^-].$$

By construction, $\lambda_d \in [0,1]$ for $d = 0, 1$, and they are uniquely determined and identifiable directly from the data.

Let $\varepsilon > 0$ be an arbitrary positive number. To construct the distributions of potential variables $(\tilde{Y}(1), \tilde{Y}(0), \tilde{T}_\varepsilon)|Z$. We first construct the distribution of $\tilde{T}_\varepsilon|Z$ as follows.

$$P(\tilde{T}_\varepsilon = \boldsymbol{A}|Z = c \pm \varepsilon) = P(D=1|Z=c-\varepsilon),$$

$$P(\tilde{T}_\varepsilon = \boldsymbol{N}|Z = c \pm \varepsilon) = P(D=0|Z=c+\varepsilon),$$

$$P(\tilde{T}_\varepsilon = \boldsymbol{C}|Z = c \pm \varepsilon) = P(D=1|Z=c+\varepsilon) - P(D=1|Z=c-\varepsilon),$$

and

$$P(\tilde{T}_\varepsilon \in \{\boldsymbol{DF}, \boldsymbol{I}\}|Z = c \pm \varepsilon) = 0.$$

By construction, Assumption 2.1 is satisfied and Assumption 2.2 is satisfied for type probabilities.

Next, we construct the distribution for $\tilde{Y}(1)$ and $\tilde{Y}(0)$ conditioning on $\tilde{T}_\varepsilon$ and Z. Let

$$P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c-\varepsilon) = P(Y \leq y|D=1, Z=c-\varepsilon)$$

$$P(\tilde{Y}(1) \leq y | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = c + \varepsilon) = \lambda_1 P(Y \leq y | D = 1, Y < G_1^{-1}(q), Z = c + \varepsilon) \\ + (1 - \lambda_1) P(Y \leq y | D = 1, Y > G_1^{-1}(1 - q), Z = c + \varepsilon)$$

$$P(\tilde{Y}(0) \leq y | \tilde{T}_\varepsilon = \boldsymbol{N}, Z = c + \varepsilon) = P(Y \leq y | D = 0, Z = c + \varepsilon)$$

$$P(\tilde{Y}(0) \leq y | \tilde{T}_\varepsilon = \boldsymbol{N}, Z = c - \varepsilon) = \lambda_0 P(Y \leq y | D = 0, Y < G_1^{-1}(r), Z = c - \varepsilon) \\ + (1 - \lambda_0) P(Y \leq y | D = 0, Y > G_1^{-1}(1 - r), Z = c - \varepsilon)$$

Note that the distributions of $\tilde{Y}(1)$ for always takers and the distributions of $\tilde{Y}(0)$ for never takers are not necessarily continuous at the cutoff. This differs significantly from the construction of Gerard, Rokkanen, and Rothe (2020) and Arai, Hsu, Kitagawa, Mourifié, and Wan (2022). Next, let

$$P(\tilde{Y}(1) \leq y | \tilde{T}_\varepsilon = \boldsymbol{C}, Z = c \pm \varepsilon) = \frac{P(Y \leq y | D = 1, Z = c + \varepsilon) - q_\varepsilon P(\tilde{Y}(1) \leq y | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = c + \varepsilon)}{1 - q_\varepsilon}$$

where

$$q_\varepsilon = \frac{P(D = 1 | Z = c - \varepsilon)}{P(D = 1 | Z = c + \varepsilon)}$$

$$P(\tilde{Y}(0) \leq y | \tilde{T}_\varepsilon = \boldsymbol{C}, Z = c \pm \varepsilon) = \frac{P(Y \leq y | D = 0, Z = c - \varepsilon) - r_\varepsilon P(\tilde{Y}(0) \leq y | \tilde{T}_\varepsilon = \boldsymbol{N}, Z = c - \varepsilon)}{1 - r_\varepsilon}$$

where

$$r_\varepsilon = \frac{P(D = 0 | Z = c + \varepsilon)}{P(D = 0 | Z = c - \varepsilon)}$$

By construction, the potential outcome distribution for compliers satisfies Assumption 2.2.

The conditioning distributions of $\tilde{Y}(1)$ given $(\tilde{T}_\varepsilon = \boldsymbol{N}, Z = z)$, and the conditioning distribution of $\tilde{Y}(0)$ given $(\tilde{T}_\varepsilon = \boldsymbol{A}, Z = z)$ are left to be arbitrary distributions (chosen to satisfy Assumption 2.2). Also, we leave $\tilde{Y}(1)$ and $\tilde{Y}(0)$ to be independent with each other conditioning on $\tilde{T}_\varepsilon$ and Z.

To this end, it remains to verify continuity holds for $E[\tilde{Y}(1) | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = z]$ and $E[Y(0) | \tilde{T}_\varepsilon = \boldsymbol{N}, Z = z]$. Consider $E[\tilde{Y}(1) | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = z]$ first. Note that

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} E[\tilde{Y}(1) | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = c + \varepsilon] &= \lim_{\varepsilon \downarrow 0} \int y dF_{\tilde{Y}(1) | (\tilde{T}_\varepsilon, Z)}(y | \boldsymbol{A}, c + \varepsilon) \\ &= \lim_{\varepsilon \downarrow 0} \int y d \left\{ \lambda_1 F_{Y | (D, Y < G_1^{-1}(q), Z)}(y | 1, c + \varepsilon) + (1 - \lambda_1) dF_{Y | (D, Y > G_1^{-1}(1-q), Z)}(y | 1, c + \varepsilon) \right\} \\ &= \lambda_1 E[Y | D = 1, Y < G_1^{-1}(q), Z = c^+] + (1 - \lambda_1) E[Y | D = 1, Y > G_1^{-1}(1 - q), Z = c^+] \\ &= E[Y | D = 1, Z = c^-] = \lim_{\varepsilon \downarrow 0} E[Y(1) | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = c - \varepsilon], \end{aligned}$$

where the second equality is by the construction of $P(\tilde{Y}(1) \leq y | \tilde{T}_\varepsilon = \boldsymbol{A}, Z = c + \varepsilon)$, the third one holds because we assume all the expectations and their limits are well-defined, and the fourth is by the definition of λ_1, and the fifth one holds by construction $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c-\varepsilon) = P(Y \leq y|D = 1, Z = c-\varepsilon)$. Similarly, the continuity holds for $E[\tilde{Y}(0)|\tilde{T}_\varepsilon = \boldsymbol{N}, Z = z]$.

Proper Distributions. $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c-\varepsilon)$ and $P(\tilde{Y}(0) \leq y|\tilde{T}_\varepsilon = \boldsymbol{N}, Z = c+\varepsilon)$ and their limits (of $\varepsilon \downarrow 0$) are proper distributions by construction. $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon)$ is a mixture of two proper distributions; hence it is a proper distribution as well. Same for $P(\tilde{Y}(0) \leq y|T_\varepsilon = \boldsymbol{N}, Z = c-\varepsilon)$.

For $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c \pm \varepsilon)$, note that for any ε, when $y \to \sup\{\mathcal{Y}\}$, $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c \pm \varepsilon) \to 1$; when when $y \to \inf\{\mathcal{Y}\}$, $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c \pm \varepsilon) \to 0$. It remains to verify the non-decreasing property. Consider two cases. In the first case, we have $G_1^{-1}(q) < G_1^{-1}(1-q)$ or $q < \frac{1}{2}$, we have

$$
\begin{aligned}
(1-q)\lim_{\varepsilon\downarrow 0} P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c\pm\varepsilon) &\overset{(1)}{=} P(Y \leq y|D = 1, Z = c^+) - q\lim_{\varepsilon\downarrow 0} P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon) \\
&\overset{(2)}{=} qP(Y \leq y|D = 1, Y < G_1^{-1}(q), Z = c^+) + qP(Y \leq y|D = 1, Y > G_1^{-1}(1-q), Z = c^+) \\
&+ (1-2q)P(Y \leq y|D = 1, G_1^{-1}(q) \leq Y \leq G_1^{-1}(1-q), Z = c^+) - q\lim_{\varepsilon\downarrow 0} P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon) \\
&\overset{(3)}{=} (q - q\lambda_1)P(Y \leq y|D = 1, Y < G_1^{-1}(q), Z = c^+) + q\lambda_1 P(Y \leq y|D = 1, Y > G_1^{-1}(1-q), Z = c^+) \\
&+ (1-2q)P(Y \leq y|D = 1, G_1^{-1}(q) \leq Y \leq G_1^{-1}(q), Z = c^+),
\end{aligned}
$$

where the first equality is by the construction of $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c \pm \varepsilon)$, the second one is by the operation of conditional probabilities, and the third one is by substitution of $P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon)$. Therefore, the right-hand side is non-decreasing because it is the sum of three non-decreasing functions (recall that $q - q\lambda_1 \geq 0$, $q\lambda_1 \geq 0$ and $1 - 2q \geq 0$).

In the second case, $G_1^{-1}(q) \geq G_1^{-1}(1-q)$ or $q \geq \frac{1}{2}$, we have

$$
\begin{aligned}
&(1-q)\lim_{\varepsilon\downarrow 0} P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c\pm\varepsilon) \overset{(1)}{=} P(Y \leq y|D=1, Z=c^+) - \lim_{\varepsilon\downarrow 0} qP(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon) \\
&\overset{(2)}{=} (1-q)P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+) + (1-q)P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+) \\
&+ (2q-1)P(Y \leq y|D=1, G_1^{-1}(1-q) \leq Y \leq G_1^{-1}(q), Z=c^+) - q\lim_{\varepsilon\downarrow 0} P(\tilde{Y}(1) \leq y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon) \\
&\overset{(3)}{=} (1-q)P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+) + (1-q)P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+) \\
&+ (2q-1)P(Y \leq y|D=1, G_1^{-1}(1-q) \leq Y \leq G_1^{-1}(q), Z=c^+) - q\lambda_1 P(Y \leq y|D=1, Y < G_1^{-1}(q), Z=c^+) \\
&\qquad - q(1-\lambda_1)P(Y \leq y|D=1, Y > G_1^{-1}(1-q), Z=c^+) \\
&\overset{(4)}{=} (1-q)P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+) + (1-q)P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+) \\
&+ (2q-1)P(Y \leq y|D=1, G_1^{-1}(1-q) \leq Y \leq G_1^{-1}(q), Z=c^+) - (1-q)\lambda_1 P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+) \\
&- (2q-1)\lambda_1 P(Y \leq y|D=1, G_1^{-1}(1-q) \leq Y \leq G_1^{-1}(q), Z=c^+) - (1-q)(1-\lambda_1)P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+) \\
&\qquad - (2q-1)(1-\lambda_1)P(Y \leq y|D=1, G_1^{-1}(1-q) \leq Y \leq G_1^{-1}(q), Z=c^+) \\
&\overset{(5)}{=} (1-q)(1-\lambda_1)P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+) + (1-q)\lambda_1 P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+),
\end{aligned}
$$

where the first three equalities hold as in the previous case, the fifth is by simple calculation, and the fourth equality holds by noting that

$$
\begin{aligned}
&q\lambda_1(Y \leq y|D=1, Y < G_1^{-1}(q), Z=c^+) \\
&= q\lambda_1 P(Y \leq y|D=1, G_1^{-1}(1-q) < Y < G_1^{-1}(q), Z=c^+)P(G_1^{-1}(1-q) < Y < G_1^{-1}(q)|D=1, Y < G^{-1}(q), C=c^+) \\
&\quad + q\lambda_1 P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+)P(Y < G_1^{-1}(1-q)|D=1, Y < G_1^{-1}(q), Z=c^+) \\
&= (2q-1)\lambda_1 P(Y \leq y|D=1, G_1^{-1}(1-q) < Y < G_1^{-1}(q), Z=c^+) + (1-q)\lambda_1 P(Y \leq y|D=1, Y < G_1^{-1}(1-q), Z=c^+)
\end{aligned}
$$

and

$$
\begin{aligned}
&q(1-\lambda_1)P(Y \leq y|D=1, Y > G_1^{-1}(1-q), Z=c^+) \\
&= q(1-\lambda_1)P(Y \leq y|D=1, G_1^{-1}(1-q) < Y < G_1^{-1}(q), Z=c^+)P(G_1^{-1}(1-q) < Y < G_1^{-1}(q)|D=1, Y > G^{-1}(1-q), C=c^+) \\
&\quad + q(1-\lambda_1)P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+)P(Y > G_1^{-1}(q)|D=1, Y > G_1^{-1}(1-q), Z=c^+) \\
&= (2q-1)(1-\lambda_1)P(Y \leq y|D=1, G_1^{-1}(1-q) < Y < G_1^{-1}(q), Z=c^+) + (1-q)(1-\lambda_1)P(Y \leq y|D=1, Y > G_1^{-1}(q), Z=c^+).
\end{aligned}
$$

In the end, the right-hand side is non-decreasing because it is the sum of two non-decreasing functions.

Matches the observed distribution. For the final step, we verify that the constructed latent variables distributions generate the same observed data distribution. First,

$$P(\tilde{Y} \le y|\tilde{D} = 1, Z = c+\varepsilon) = P(\tilde{Y}(1) \le y|\tilde{T}_\varepsilon \in \{\boldsymbol{A}, \boldsymbol{C}\}, Z = c+\varepsilon)$$
$$= P(\tilde{Y}(1) \le y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c+\varepsilon)\frac{P(\boldsymbol{A}|Z = c+\varepsilon)}{P(\boldsymbol{A}\cup\boldsymbol{C}|Z = c+\varepsilon)} + P(\tilde{Y}(1) \le y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c+\varepsilon)\frac{P(\boldsymbol{C}|Z = c+\varepsilon)}{P(\boldsymbol{A}\cup\boldsymbol{C}|Z = c+\varepsilon)}$$
$$= P(Y \le y|D = 1, Z = c+\varepsilon),$$

where the first equality is by definition, the second is by the total law of probabilities, and the third is by construction.

Next, for $P(\tilde{Y} \le y|\tilde{D} = 1, Z = c-\varepsilon)$, we have

$$P(\tilde{Y} \le y|\tilde{D} = 1, Z = c-\varepsilon) = P(\tilde{Y}(1) \le y|\tilde{T}_\varepsilon = \boldsymbol{A}, Z = c-\varepsilon) = P(Y \le y|D = 1, Z = c-\varepsilon),$$

where the first equality is by definition, and the second is by construction.

For $P(\tilde{Y} \le y|\tilde{D} = 0, Z = c+\varepsilon)$, we have

$$P(\tilde{Y} \le y|\tilde{D} = 0, Z = c+\varepsilon) = P(\tilde{Y}(0) \le y|\tilde{T}_\varepsilon = \boldsymbol{N}, Z = c+\varepsilon) = P(Y \le y|D = 0, Z = c+\varepsilon),$$

where the last equality holds because we set $P(\tilde{Y}(0) \le y|\tilde{T}_\varepsilon = \boldsymbol{N}, Z = c+\varepsilon) = P(Y \le y|D = 0, Z = c+\varepsilon)$.

$$P(\tilde{Y} \le y|\tilde{D} = 0, Z = c-\varepsilon) = P(\tilde{Y}(0) \le y|\tilde{T}_\varepsilon \in \{\boldsymbol{N}, \boldsymbol{C}\}, Z = c-\varepsilon)$$
$$= P(\tilde{Y}(0) \le y|\tilde{T}_\varepsilon = \boldsymbol{N}, Z = c-\varepsilon)\frac{P(\boldsymbol{N}|Z = c-\varepsilon)}{P(\boldsymbol{N}\cup\boldsymbol{C}|Z = c-\varepsilon)} + P(\tilde{Y}(0) \le y|\tilde{T}_\varepsilon = \boldsymbol{C}, Z = c-\varepsilon)\frac{P(\boldsymbol{C}|Z = c-\varepsilon)}{P(\boldsymbol{N}\cup\boldsymbol{C}|Z = c-\varepsilon)}$$
$$= P(Y \le y|D = 0, Z = c-\varepsilon),$$

where the first equality is by definition, the second is by the total law of probabilities, and the third is by construction.

A.2 Proof of Corollary 6.1

In this proof, to save space, we abuse the notation to write $E[Y(d)|\boldsymbol{t}, Z = c^+]$ for $\lim_{\varepsilon\downarrow 0} E[Y(d)|T_\varepsilon = \boldsymbol{t}, Z = c+\varepsilon]$, and write $E[Y(d)|\boldsymbol{t}, Z = c^-] \equiv \lim_{\varepsilon\downarrow 0} E[Y(d)|T_\varepsilon = \boldsymbol{t}, Z = c-\varepsilon]$. Similarly, we write $P(\boldsymbol{t}|Z = c^+)$ for $\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = \boldsymbol{t}|Z = c+\varepsilon)$ and write $P(\boldsymbol{t}|Z = c^-)$ for $\lim_{\varepsilon\downarrow 0} P(T_\varepsilon = \boldsymbol{t}|Z = c-\varepsilon)$.

(i) Suppose $Y(d)$, $d \in \{0,1\}$, is 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 $1 \le j < J$. 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 observed quantity $E[Y|D = 1, Z = c^+]$ can be expressed as:

$$
\begin{aligned}
&E[Y|D = 1, Z = c^+] \\
&\quad = \sum_{j=1}^{J} y_j P(Y = y_j|D = 1, Z = c^+) = \sum_{j=1}^{J} y_j P(Y = y_j|\boldsymbol{A} \cup \boldsymbol{C}, Z = c^+) \\
&\quad = \sum_{j=1}^{J} y_j \frac{P(\boldsymbol{A} \cup \boldsymbol{C}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A} \cup \boldsymbol{C}|Z = c^+)} \\
&\quad = \sum_{j=1}^{J} y_j \left\{ \frac{P(\boldsymbol{A}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A} \cup \boldsymbol{C}|Z = c^+)} \right. \\
&\qquad \left. + \frac{P(\boldsymbol{C}|Y = y_j, Z = c^+)P(Y = y_j|Z = c^+)}{P(\boldsymbol{A} \cup \boldsymbol{C}|Z = c^+)} \right\}, \qquad \text{(A.4)}
\end{aligned}
$$

where the first equality holds by the definition of conditional expectation, the second holds because as z approaches c from above, the event $\{D = 1, Z = c^+\}$ 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.4), and by definition of q 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

$$\sum_{j=1}^{j^*} P(Y = y_j|Z = c^+) \leq qP(D = 1|Z = c^+), \qquad \text{(A.5)}$$

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 $j \le j^*$, 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

$$\begin{aligned} LB^+ &\equiv \sum_{j=1}^{j^*} y_j \frac{P(Y=y_j|Z=c^+)}{P(\boldsymbol{A}|Z=c^+)} + y_{j^*+1} \frac{qP(D=1|Z=c^+) - \sum_{j=1}^{j^*} P(Y=y_j|Z=c^+)}{P(\boldsymbol{A}|Z=c^+)} \\ &= \sum_{j=1}^{j^*} y_j \frac{P(Y=y_j|D=1, Z=c^+)}{q} + y_{j^*+1} \left\{ 1 - \frac{\sum_{j=1}^{j^*} P(Y=y_j|D=1, Z=c^+)}{q} \right\}, \end{aligned} \tag{A.6}$$

where the second equality holds because

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

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

$$LB^+ \ge \frac{\sum_{j=1}^{j^*} y_j P(Y=y_j|D=1, Z=c^+)}{\sum_{j=1}^{j^*} P(Y=y_j|D=1, Z=c^+)} = E[Y|D=1, Y < G_1^{-1}(q), Z=c^+] \tag{A.7}$$

where the inequality holds because $y_{j^*+1} > y_{j^*}$, and the equality holds by the definition of G_1^{-1}. Therefore, $E[Y|D=1, Y \le G_1^{-1}(q), Z=c^+]$ 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.4) 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

$$\sum_{j=j^\dagger}^{J} P(Y = y_j|Z = c^+) \leq qP(D = 1|Z = c^+) \tag{A.8}$$

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

$$\begin{aligned} UB^+ &\equiv \sum_{j=j^\dagger}^{J} y_j \frac{P(Y = y_j|Z = c^+)}{P(\boldsymbol{A}|Z = c^+)} + y_{j^\dagger-1} \frac{qP(D = 1|Z = c^+) - \sum_{j=j^\dagger}^{J} P(Y = y_j|Z = c^+)}{P(Y = y_{j^\dagger-1}|Z = c^+)} \\ &= \sum_{j=j^\dagger}^{J} y_j \frac{P(Y = y_j|D = 1, Z = c^+)}{q} + y_{j^\dagger-1} \left\{1 - \frac{\sum_{j=j^\dagger}^{J} P(Y = y_j|D = 1, Z = c^+)}{q}\right\}. \end{aligned} \tag{A.9}$$

This bound can also be relaxed:

$$UB^+ \leq \frac{\sum_{j=j^\dagger}^{J} y_j P(Y = y_j|D = 1, Z = c^+)}{\sum_{j=j^\dagger}^{J} P(Y = y_j|D = 1, Z = c^+)} = E[Y|D = 1, Y > G_1^{-1}(1 - q), Z = c^+],$$

where the inequality holds because $y_{j^\dagger} > y_{j^\dagger-1}$, and the last equality holds by the definition of G_1^{-1}. Therefore, $E[Y|D = 1, Y > G_1^{-1}(1 - q), Z = c^+]$ 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^+.

(ii) Suppose $Y(1)$ is continuous but also has possibly mass points. If the q-th quantile is a continuous point, then bounds can be derived following the same argument as in Proposition 2.2; 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 (i) with y^* playing the role of y_{j^*} as

$$\begin{aligned} LB^+ = E[Y|D = 1, Y < y^*, Z = c^+]\frac{P(Y < y^*|D = 1, Z = c^+)}{q} + y^* \left(1 - \frac{P(Y < y^*|D = 1, Z = c^+)}{q}\right) \\ \geq E[Y|D = 1, Y < y^*, Z = c^+] = E[Y|D = 1, Y < G_1^{-1}(q), Z = c^+]. \quad \text{(A.10)} \end{aligned}$$

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

$$\begin{aligned} UB^+ = E[Y|D = 1, Y > y^\dagger, Z = c^+]\frac{P(Y > y^\dagger|D = 1, Z = c^+)}{q} + y^\dagger \left(1 - \frac{P(Y < y^*|D = 1, Z = c^+)}{q}\right) \\ \geq E[Y|D = 1, Y > y^\dagger, Z = c^+] = E[Y|D = 1, Y > G_1^{-1}(1 - q), Z = c^+]. \quad \text{(A.11)} \end{aligned}$$

□

Example 1.3 *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, $G_1^{-1}(0.26) = \inf\{y \in \mathcal{Y} : G_1(y) \geq 0.26\} = y_2$; *hence, a valid but non-sharp lower* *bound is given by:*

$$E[Y|D=1, Y < G_1^{-1}(0.26), Z=c^+] = E[Y|D=1, Y < y_2, Z=c^+] = y_1 < LB^+.$$

$G_1^{-1}(1-0.26) = \sup\{y \in \mathcal{Y} : G_1(y) \leq 0.74\} = y_2$*; hence, a valid but non-sharp upper bound is given by:*

$$E[Y|D=1, Y > G_1^{-1}(0.74), Z=c^+] = E[Y|D=1, Y > y_2, Z=c^+] = \frac{y_3+y_4}{2} > UB^+.$$

<u>Case 2</u>. 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, $G_1^{-1}(0.5) = \inf\{y \in \mathcal{Y} : G_1(y) \geq 0.5\} = y_2$*, and the valid but non-sharp lower bound is given by:*

$$E[Y|D=1, Y < G_1^{-1}(0.5), Z=c^+] = E[Y|D=1, Y < y_2, Z=c^+] = y_1 < LB^+.$$

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

$$E[Y|D=1, Y > G_1^{-1}(0.5), Z=c^+] = E[Y|D=1, Y > y_3, Z=c^+] = y_4 > UB^+.$$

<u>Case 3.</u> 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, $G_1^{-1}(0.24) = \inf\{y \in \mathcal{Y} : G_1(y) \geq 0.24\} = y_1$*. The valid but non-sharp lower bound* $E[Y|D=1, Y < G_1^{-1}(0.24), Z=c^+] = E[Y|D=1, Y < y_1, Z=c^+]$ *is not well-defined and hence is understood as* $-\infty$*.* $G_1^{-1}(1-0.24) = \sup\{y \in \mathcal{Y} : G_1(y) \leq 0.76\} = y_4$*; hence, the valid but non-sharp upper bound is given by:* $E[Y|D=1, Y > G_1^{-1}(0.5), Z=c^+] = E[Y|D=1, Y > y_4, Z=c^+]$*. 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

$$\sqrt{nh_2}\begin{pmatrix}\hat{\theta}_1-\theta_1\\ \hat{\theta}_2-\theta_2\\ \hat{\theta}_3-\theta_3\\ \hat{\theta}_4-\theta_4\end{pmatrix}\xrightarrow{d}\mathcal{N}(0,\Omega),$$

where $\Omega_{j,k}=\lim_{n\to\infty}h_2^{-1}E[\phi_{\theta_j,h_2,i}\phi_{\theta_k,h_2,i}]$ for $j,k=1,2,3,4$. We also have

$$\sqrt{nh_2}\begin{pmatrix}\hat{\theta}_1^w-\hat{\theta}_1\\ \hat{\theta}_2^w-\hat{\theta}_2\\ \hat{\theta}_3^w-\hat{\theta}_3\\ \hat{\theta}_4^w-\hat{\theta}_4\end{pmatrix}\xrightarrow{d}\mathcal{N}(0,\Omega),$$

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

$$\lim_{T\to\infty}\hat{\sigma}_j^w=\lim_{T\to\infty}\Big(nh_2T^{-1}\sum_{t=1}^{T}(\hat{\theta}_j^{w,t}-\hat{\theta}_j)^2\Big)^{1/2}=\Big((nh_2)^{-1}\sum_{i=1}^{n}\phi_{\theta_j,h_2,i}^2+o_p(1)\Big)^{1/2}\xrightarrow{p}\sigma_j.$$

Then we can apply the results in Donald and Hsu (2011) to show Proposition 3.3 and we omit the details. The proof for the second bootstrap method is the same 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* $\delta\le f_z(z)\le M$ *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 three-time continuously differentiable on* $z\in(c-\epsilon,c)$ *with absolute values of corresponding derivatives bounded by* M*;*

2. $E[D|Z=z]$ *is three-time 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 three-time 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 three-time 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_1 *satisfies that* $h_1 \to 0$, $nh_1^2 \to \infty$, *and* $nh_1^7 \to 0$ *as* $n \to \infty$.

3. *The bandwidth* h_2 *satisfies that* $h_2 \to 0$, $nh_2^2 \to \infty$, *and* $nh_2^7 \to 0$ *as* $n \to \infty$.

4. $h_2/h_1 \to 0$.

Assumption B.6 *(Continuous Case) Assume that* $G_1(y)$ *is continuous on* $(G_1^{-1}(q) - \delta, G_1^{-1}(q) + \delta)$ *with* $G_1(G_1^{-1}(q)) = q$ *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* $y \in (G_1^{-1}(q) - \delta, G_1^{-1}(q) + \delta)$, $E[DY1(Y \leq y)|Z = z]$ *and* $E[D1(Y \leq y)|Z = z]$ *are three-time 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.

Assumption B.8 *Assume that* a_n *is a sequence of positive number with* $a_n \to \infty$ *and* $a_n/\sqrt{nh_2} \to 0$.

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

$$\sqrt{nh_2}(\hat{\theta}_1 - \theta_1) \equiv \frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} \phi_{\theta_1,h_2,i} + o_p(1), \quad \text{(B.1)}$$

where $\phi_{\theta_1,h_2,i}$ *is given in Equation (B.7). Also,* $\sqrt{nh_2}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$, *where* $V_{\theta_1} = \lim_{n\to\infty} h_2^{-1} E[\phi^2_{\theta_1,h_2,i}]$.

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

$$\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} (W_i - 1)\phi_{\theta_1,h_2,i} + o_p(1), \tag{B.2}$$

$$\sqrt{nh_2}(\hat{\theta}_1^b - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} (W_i - 1)\phi_{\theta_1,h_2,i} + o_p(1) \tag{B.3}$$

where $\phi_{\theta_1,h_2,i}$ is given in (B.7). Also, $\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$ and $\sqrt{nh_2}(\hat{\theta}_1^b - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$ 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

$$\sqrt{nh_2}(\hat{\theta}_1 - \theta_1) \equiv \frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} \phi_{\theta_1,h_2,i} + o_p(1), \tag{B.4}$$

where $\phi_{\theta_1,h_2,i}$ is given in (B.7). Also, $\sqrt{nh_2}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$, where $V_{\theta_1} = \lim_{n\to\infty} h_2^{-1} E[\phi^2_{\theta_1,h_2,i}]$.

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

$$\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} (W_i - 1)\phi_{\theta_1,h_2,i} + o_p(1), \tag{B.5}$$

$$\sqrt{nh_2}(\hat{\theta}_1^b - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} (W_i - 1)\phi_{\theta_1,h_2,i} + o_p(1) \tag{B.6}$$

where $\phi_{\theta_1,h_2,i}$ is given in (B.7). Also, $\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$ and $\sqrt{nh_2}(\hat{\theta}_1^b - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$ conditional on the sample path with probability approaching 1.

Let

$$\Delta_z = f_z(0) \cdot \begin{pmatrix} \mu_{z,0} & \mu_{z,1} & \mu_{z,2} \\ \mu_{z,1} & \mu_{z,2} & \mu_{z,3} \\ \mu_{z,2} & \mu_{z,3} & \mu_{z,4} \end{pmatrix} \text{ with } \mu_{z,j} = \int_{u\geq 0} u^j K(u)du, \text{ for } j = 0, 1, 2, \ldots.$$

Recall that for a general random variable X_i,

$$(\widehat{E}_h[X|Z = c^+], \hat{\beta}^+, \hat{\gamma}^+) = \underset{a,b,r}{\operatorname{argmin}} \sum_{i=1}^{n} 1(Z_i \geq c) K\left(\frac{Z_i - c}{h}\right) \left[X_i - a - b\cdot(Z_i - c) - r\cdot(Z_i - c)^2\right]^2,$$

$$(\widehat{E}_h[X|Z = c^-], \hat{\beta}^-, \hat{\gamma}^-) = \underset{a,b,r}{\operatorname{argmin}} \sum_{i=1}^{n} 1(Z_i < c) K\left(\frac{Z_i - c}{h}\right) \left[X_i - a - b\cdot(Z_i - c) - r\cdot(Z_i - c)^2\right]^2.$$

Suppose that $E[X|Z = z]$ is three-time 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

$$
\begin{aligned}
&\sqrt{nh}\left(\widehat{E}_h[X|Z=c^+] - E[X|Z=c^+]\right)\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(1\ 0\ 0)\Delta_z^{-1}1(Z_i \geq c)K\left(\frac{Z_i-c}{h}\right)(X_i - E[X_i|Z_i])\begin{pmatrix}1\\ \frac{Z_i-c}{h}\\ \left(\frac{Z_i-c}{h}\right)^2\end{pmatrix} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^+_{X,h,i} + o_p(1),\\
&\sqrt{nh}\left(\widehat{E}[X|Z=c^-] - E[X|Z=c^-]\right)\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(1\ 0\ 0)\Delta_z^{-1}1(Z_i < c)K\left(\frac{Z_i-c}{h}\right)(X_i - E[X_i|Z_i])\begin{pmatrix}1\\ \frac{Z_i-c}{h}\\ \left(\frac{Z_i-c}{h}\right)^2\end{pmatrix} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^-_{X,h,i} + o_p(1).
\end{aligned}
$$

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

$$
\hat{q} = \frac{\widehat{E}_{h_1}[D|Z=c^-]}{\widehat{E}_{h_1}[D|Z=c^+]}
$$

Then by delta method, we have

$$
\begin{aligned}
\sqrt{nh_1}\,(\hat{q}-q) &= \frac{1}{\sqrt{nh_1}}\sum_{i=1}^{n}\left\{\frac{1}{E[D|Z=c^+]}\phi^-_{d,h_1,i} - \frac{q}{E[D|Z=c^+]}\phi^+_{d,h_1,i}\right\} + o_p(1)\\
&\equiv \frac{1}{\sqrt{nh_1}}\sum_{i=1}^{n}\phi_{q,h_1,i} + o_p(1).
\end{aligned}
$$

It is true that $\sqrt{nh_1}(\hat{q}-q) = O_p(1)$. Similarly,

$$
\widehat{G}_1(y) = \frac{\widehat{E}_{h_1}[D1(Y \leq y)|Z=c^+]}{\widehat{E}_{h_1}[D|Z=c^+]}.
$$

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

$y \in R$,

$$\sqrt{nh_1}\left(\widehat{G}_1(y) - G_1(y)\right) = \frac{1}{\sqrt{nh_1}} \sum_{i=1}^{n} \frac{1}{E[D|Z=c^+]}\phi^+_{D1(Y\le y),h_1,i} - \frac{G_1(y)}{E[D|Z=c^+]}\phi^+_{D,h_1,i} + o_p(1)$$
$$\equiv \frac{1}{\sqrt{nh_1}} \sum_{i=1}^{n} \phi_{G_1(y),h_1,i} + o_p(1).$$

In this case, $G_1^{-1}(q)$ is differentiable and its derivative with respective to q is $g_1(G_1^{-1}(q))$. Then by functional delta method, we have

$$\begin{aligned}
&\sqrt{nh_1}\left(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(q)\right)\\
=&\sqrt{nh_1}\left(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(\hat{q})\right) + \sqrt{nh_1}\left(G_1^{-1}(\hat{q}) - G_1^{-1}(q)\right)\\
=&\sqrt{nh_1}\left(\widehat{G}_1^{-1}(q) - G_1^{-1}(q)\right) + o_p(1) + g_1(G_1^{-1}(q))\sqrt{nh_1}\left(\hat{q} - q\right) + o_p(1)\\
=&\frac{1}{\sqrt{nh_1}} \sum_{i=1}^{n} \frac{-1}{g_1(G_1^{-1}(q))}\phi_{G_1(G_1^{-1}(q)),h_1,i} + g_1(G_1^{-1}(q))\phi_{q,h_1,i} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh_1}} \sum_{i=1}^{n} \phi_{G_1^{-1}(q),h_1,i} + o_p(1).
\end{aligned}$$

Under the assumptions in Lemma B.1, we have $\sqrt{nh_1}\left(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(q)\right) = O_p(1)$. Note that under Assumption B.6, $G_1(y)$ is continuous in a neighborhood of $G_1^{-1}(q)$, and this implies that $E[DY1(Y < G_1^{-1}(q))|Z = c^+] = E[DY1(Y \le G_1^{-1}(q))|Z = c^+]$ and $\widehat{E}_{h_2}[DY1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}_{h_2}[DY1(Y \le \widehat{G}_1^{-1}(\hat{q}))|Z = c^+]$. Also, $E[D1(Y < G_1^{-1}(q))|Z = c^+] = E[D1(Y \le G_1^{-1}(q))|Z = c^+]$ and $\widehat{E}_{h_2}[D1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}_{h_2}[D1(Y \le \widehat{G}_1^{-1}(\hat{q}))|Z = c^+]$.

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

$$\begin{aligned}
&\sqrt{nh_2}(\widehat{H}_{1,\le}(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\le}(G_1^{-1}(q)))\\
=&\sqrt{nh_2}(\widehat{H}_{1,\le}(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\le}(\widehat{G}_1^{-1}(\hat{q}))) + \sqrt{nh_2}(H_{1,\le}(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\le}(G_1^{-1}(q)))\\
=&\sqrt{nh_2}(\widehat{H}_{1,\le}(G_1^{-1}(q)) - H_{1,\le}(G_1^{-1}(q))) + o_p(1) + \sqrt{nh_2}(H_{1,\le}(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\le}(G_1^{-1}(q)))\\
=&\frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} \phi^+_{DY1(Y\le G_1^{-1}(q)),h_2,i} + o_p(1) + P_{1|1} \cdot \ddot{y} \cdot g_1(\ddot{y}) \cdot \sqrt{nh_2}(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(q))\\
\equiv&\frac{1}{\sqrt{nh_2}} \sum_{i=1}^{n} \phi_{H_{1,\le}(G_1^{-1}(q)),h_2,i} + o_p(1),
\end{aligned}$$

where $\ddot{y}$ is between $\widehat{G}_1^{-1}(\hat{q})$ and $G_1^{-1}(q)$, and $\ddot{y}$ is bounded above in probability. The last line holds by the fact that

$$\sqrt{nh_2}(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(q)) = \sqrt{\frac{h_2}{h_1}}\sqrt{nh_1}(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(q)) = o(1)O_p(1) = o_p(1).$$

In other words, the estimation effect from the first stage can be ignored asymptotically. Similarly, we have

$$\sqrt{nh_2}(\widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q})) - I_{1,\leq}(G_1^{-1}(q))) = \frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{+}_{D1(Y\leq G_1^{-1}(q)),h_2,i} + o_p(1)$$
$$\equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} + o_p(1).$$

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

$$\begin{aligned}
&\sqrt{nh_2}(\hat{\theta}_1 - \theta_1)\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[D|Z=c^-] - \widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[DY|Z=c^-]\\
&\qquad - H_{1,\leq}(G_1^{-1}(q))\cdot E[D|Z=c^-] + I_{1,\leq}(G_1^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[D|Z=c^-] - H_{1,\leq}(G_1^{-1}(q))\cdot E[D|Z=c^-])\\
&\qquad - \sqrt{nh_2}(\widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[DY|Z=c^-] - I_{1,\leq}(G_1^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\frac{1}{\sqrt{nh_2},}\sum_{i=1}^{n}\Big\{E[D|Z=c^-]\cdot\phi_{H_{1,\leq}(G_1^{-1}(q)),h_2,i} + H_{1,\leq}(G_1^{-1}(q))\cdot\phi^-_{D,h_2,i}\\
&\qquad - E[DY|Z=c^-]\cdot\phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} - I_{1,\leq}(G_1^{-1}(q))\cdot\phi^-_{DY,h_2,i}\Big\} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh_2},}\sum_{i=1}^{n}\phi_{\theta_1,h_2,i} + o_p(1).
\end{aligned}\tag{B.7}$$

Then, by the central limit theorem, we have $\sqrt{nh_2}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$. This completes the proof.□

Proof of Lemma B.2: Recall that

$$\hat{q}^w = \frac{\widehat{E}^w_{h_1}[D|Z=c^-]}{\widehat{E}^w_{h_1}[D|Z=c^+]}.$$

Then by the same arguments for Theorem 5.2 of Hsu and Shen (2022) and by the arguments for Lemma B.1, we have

$$\sqrt{nh_1}\,(\hat{q}^w - q) = \frac{1}{\sqrt{nh_1}}\sum_{i=1}^{n}W_i\Big(\frac{1}{E[D|Z=c^+]}\phi^-_{d,h_1,i} - \frac{q}{E[D|Z=c^+]}\phi^+_{d,h_1,i}\Big) + o_p(1)$$
$$\equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}W_i\cdot\phi_{q,h_1,i} + o_p(1).$$

This implies that $\sqrt{nh_1}(\hat{q}^w - q) = O_p(1)$. Similarly,

$$\widehat{G}_1^w(y) = \frac{\widehat{E}_{h_1}^w[D1(Y \leq y)|Z = c^+]}{\widehat{E}_{h_1}^w[D|Z = c^+]}.$$

and we have uniformly over $y \in R$,

$$\begin{aligned}
\sqrt{nh_1}\left(\widehat{G}_1^w(y) - G_1(y)\right) &= \frac{1}{\sqrt{nh_1}}\sum_{i=1}^n W_i\Big(\frac{1}{E[D|Z=c^+]}\phi^+_{D1(Y\leq y),h_1,i} - \frac{G_1(y)}{E[D|Z=c^+]}\phi^+_{D,h_1,i}\Big) + o_p(1) \\
&\equiv \frac{1}{\sqrt{nh_1}}\sum_{i=1}^n W_i \cdot \phi_{G_1(y),h_1,i} + o_p(1).
\end{aligned}$$

Then by functional delta method, we have

$$\begin{aligned}
&\sqrt{nh_1}\left(\widehat{G}_1^{-1,w}(\hat{q}^w) - G_1^{-1}(q)\right) \\
=&\sqrt{nh_1}\left(\widehat{G}_1^{-1,w}(\hat{q}^w) - G_1^{-1}(\hat{q}^w)\right) + \sqrt{nh_1}\left(G_1^{-1}(\hat{q}^w) - G_1^{-1}(q)\right) \\
=&\sqrt{nh_1}\left(\widehat{G}_1^{-1,w}(q) - G_1^{-1}(q)\right) + o_p(1) + g_1(G_1^{-1}(q))\sqrt{nh_1}\left(\hat{q}^w - q\right) + o_p(1) \\
=&\frac{1}{\sqrt{nh_1}}\sum_{i=1}^n W_i\Big(\frac{-1}{g_1(G_1^{-1}(q))}\phi_{G_1(G_1^{-1}(q)),h_1,i} + g_1(G_1^{-1}(q))\phi_{q,h_1,i}\Big) + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh_1}}\sum_{i=1}^n W_i \cdot \phi_{G_1^{-1}(q),h_1,i} + o_p(1).
\end{aligned}$$

Under the assumptions in Lemma B.1, we have $\sqrt{nh_1}\left(\widehat{G}_1^{-1,w}(\hat{q}^w) - G_1^{-1}(q)\right) = O_p(1)$. Also, $\widehat{E}_{h_2}^w[DY1(Y < \widehat{G}_1^{-1,w}(\hat{q}^w))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}_{h_2}^w[DY1(Y \leq \widehat{G}_1^{-1,w}(\hat{q}^w))|Z = c^+]$ and $\widehat{E}_{h_2}^w[D1(Y < \widehat{G}_1^{-1,w}(\hat{q}^w))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}_{h_2}^w[D1(Y \leq \widehat{G}_1^{-1,w}(\hat{q}^w))|Z = c^+]$. Therefore, we have

$$\begin{aligned}
&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(\widehat{G}_1^{-1,w}(\hat{q}^w)) - H_{1,\leq}(G_1^{-1}(q))) \\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(\widehat{G}_1^{-1,w}(\hat{q}^w)) - H_{1,\leq}(\widehat{G}_1^{-1,w}(\hat{q}^w))) + \sqrt{nh_2}(H_{1,\leq}(\widehat{G}_1^{-1,w}(\hat{q}^w)) - H_{1,\leq}(G_1^{-1}(q))) \\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(G_1^{-1}(q)) - H_{1,\leq}(G_1^{-1}(q))) + o_p(1) + \sqrt{nh_2}(H_{1,\leq}(\widehat{G}_1^{-1,w}(\hat{q}^w)) - H_{1,\leq}(G_1^{-1}(q))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi^+_{DY1(Y\leq G_1^{-1}(q)),h_2,i} + o_p(1) + P_{1|1} \cdot \ddot{y} \cdot g_1(\ddot{y}^w) \cdot \sqrt{nh_2}(\widehat{G}_1^{-1,w}(\hat{q}^w) - G_1^{-1}(q)) \\
\equiv&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi_{H_{1,\leq}(G_1^{-1}(q)),h_2,i} + o_p(1),
\end{aligned}$$

where $\ddot{y}^w$ is between $\widehat{G}_1^{-1,w}(\hat{q}^w)$ and $G_1^{-1}(q)$, and $\ddot{y}^w$ is bounded above in probability. The last line holds by the fact that

$$\sqrt{nh_2}(\widehat{G}_1^{-1,w}(\hat{q}^w) - G_1^{-1}(q)) = \sqrt{\frac{h_2}{h_1}}\sqrt{nh_1}(\widehat{G}_1^{-1,w}(\hat{q}^w) - G_1^{-1}(q)) = o(1)O_p(1) = o_p(1).$$

In other words, in the weighted bootstrap estimation, the estimation effect from the first stage can be ignored asymptotically as well. Similarly, we have

$$\sqrt{nh_2}(\widehat{I}_{1,\leq}^w(\widehat{G}_1^{-1,w}(\hat{q}^w)) - I_{1,\leq}(G_1^{-1}(q))) = \frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi^+_{D1(Y\leq G_1^{-1}(q)),h_2,i} + o_p(1)$$
$$\equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} + o_p(1).$$

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

$$\begin{aligned}
&\sqrt{nh_2}(\hat{\theta}_1^w - \theta_1)\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(\widehat{G}_1^{-1,w}(\hat{q}^w)) \cdot \widehat{E}_{h_2}^w[D|Z=c^-] - \widehat{I}_{1,\leq}(\widehat{G}_1^{-1,w}(\hat{q}^w)) \cdot \widehat{E}_{h_2}^w[DY|Z=c^-]\\
&\quad - H_{1,\leq}(G_1^{-1}(q)) \cdot E[D|Z=c^-] + I_{1,\leq}(G_1^{-1}(q)) \cdot E[DY|Z=c^-])\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(\widehat{G}_1^{-1,w}(\hat{q}^w)) \cdot \widehat{E}_{h_2}^w[D|Z=c^-] - H_{1,\leq}(G_1^{-1}(q)) \cdot E[D|Z=c^-])\\
&\quad - \sqrt{nh_2}(\widehat{I}_{1,\leq}^w(\widehat{G}_1^{-1,w}(\hat{q}^w)) \cdot \widehat{E}_{h_2}^w[DY|Z=c^-] - I_{1,\leq}(G_1^{-1}(q)) \cdot E[DY|Z=c^-])\\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i\Big\{E[D|Z=c^-] \cdot \phi_{H_{1,\leq}(G_1^{-1}(q)),h_2,i} + H_{1,\leq}(G_1^{-1}(q)) \cdot \phi^-_{D,h_2,i}\\
&\quad - E[DY|Z=c^-] \cdot \phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} - I_{1,\leq}(G_1^{-1}(q)) \cdot \phi^-_{DY,h_2,i}\Big\} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh_2},}\sum_{i=1}^n W_i \cdot \phi_{\theta_1,h_2,i} + o_p(1).
\end{aligned}$$

It follows that

$$\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^n (W_i - 1) \cdot \phi_{\theta_1,h_2,i} + o_p(1).$$

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 $\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1}^d)$ conditional on the sample path with probability approaching 1.

Next, we consider the asymptotics of $\hat{\theta}_1^b$. We first consider $\widehat{E}_{h_2}^w[DY1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z=c^+]$. Similarly, we have

$$\begin{aligned}
&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\leq}(G_1^{-1}(q)))\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}^w(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))) + \sqrt{nh_2}(H_{1,\leq}(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\leq}(G_1^{-1}(q)))
\end{aligned}$$

$$
\begin{aligned}
=&\sqrt{nh_2}(\widehat{H}^w_{1,\leq}(G_1^{-1}(q)) - H_{1,\leq}(G_1^{-1}(q))) + o_p(1) + \sqrt{nh_2}(H_{1,\leq}(\widehat{G}_1^{-1}(\hat{q})) - H_{1,\leq}(G_1^{-1}(q)))\\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi^+_{DY1(Y\leq G_1^{-1}(q)),h_2,i} + o_p(1) + P_{1|1}\cdot \ddot{y}\cdot g_1(\ddot{y})\cdot \sqrt{nh_2}(\widehat{G}_1^{-1}(\hat{q}^w) - G_1^{-1}(q))\\
\equiv&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi_{H_{1,\leq}(G_1^{-1}(q)),h_2,i} + o_p(1),
\end{aligned}
$$

where $\ddot{y}$ is between $\widehat{G}_1^{-1}(\hat{q})$ and $G_1^{-1}(q)$ which is the same in the proof of Lemma B.1. The last line holds by the fact that $\sqrt{nh_2}(\widehat{G}_1^{-1}(\hat{q}) - G_1^{-1}(q)) = o_p(1)$. Similarly, we have

$$
\begin{aligned}
&\sqrt{nh_2}(\widehat{I}^w_{1,\leq}(\widehat{G}_1^{-1}(\hat{q})) - I_{1,\leq}(G_1^{-1}(q))) = \frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i\cdot \phi^+_{D1(Y\leq G_1^{-1}(q)),h_2,i} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i \cdot \phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} + o_p(1).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\sqrt{nh_2}(\hat{\theta}_1^b - \theta_1)\\
=&\sqrt{nh_2}(\widehat{H}^w_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot \widehat{E}^w_{h_2}[D|Z=c^-] - \widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot \widehat{E}^w_{h_2}[DY|Z=c^-]\\
&\quad - H_{1,\leq}(G_1^{-1}(q))\cdot E[D|Z=c^-] + I_{1,\leq}(G_1^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\sqrt{nh_2}(\widehat{H}^w_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot \widehat{E}^w_{h_2}[D|Z=c^-] - H_{1,\leq}(G_1^{-1}(q))\cdot E[D|Z=c^-])\\
&\quad - \sqrt{nh_2}(\widehat{I}^w_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot \widehat{E}^w_{h_2}[DY|Z=c^-] - I_{1,\leq}(G_1^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n W_i\Big\{E[D|Z=c^-]\cdot \phi_{H_{1,\leq}(G_1^{-1}(q)),h_2,i} + H_{1,\leq}(G_1^{-1}(q))\cdot \phi^-_{D,h_2,i}\\
&\quad - E[DY|Z=c^-]\cdot \phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} - I_{1,\leq}(G_1^{-1}(q))\cdot \phi^-_{DY,h_2,i}\Big\} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh_2},}\sum_{i=1}^n W_i\cdot \phi_{\theta_1,h_2,i} + o_p(1).
\end{aligned}
$$

It follows that

$$
\sqrt{nh_2}(\hat{\theta}_1^b - \hat{\theta}_1) \equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^n (W_i - 1)\cdot \phi_{\theta_1,h_2,i} + o_p(1),
$$

and $\sqrt{nh_2}(\hat{\theta}_1^w - \hat{\theta}_1) \xrightarrow{d} \mathcal{N}(0, V^d_{\theta_1})$ 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 $G_1^{-1}(q) = y_{1L,u}$. 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 $\widehat{G}_1^{-1}(\hat{q})) = y_{1L,u}$

with probability approaching one. That is, we have $\sqrt{nh_2}(\widehat{G}_1^{-1}(\hat{q})) - y_{1L,u}) = o_p(1)$. In addition, $E[DY1(Y < y_{1L,u})|Z = c^+] = E[DY1(Y \leq y_{1L,\ell})|Z = c^+]$, and $\widehat{E}_{h_2}[DY1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}_{h_2}[DY1(Y < y_{1L,u})|Z = c^+] = \widehat{E}[DY1(Y \leq y_{1L,\ell})|Z = c^+]$. Similarly, $E[D1(Y < y_{1L,u})|Z = c^+] = E[D1(Y \leq y_{1L,\ell})|Z = c^+]$, and $\widehat{E}_{h_2}[D1(Y < \widehat{G}_1^{-1}(\hat{q}))|Z = c^+]$ is asymptotically equivalent to $\widehat{E}_{h_2}[D1(Y < y_{1L,u})|Z = c^+] = \widehat{E}[D1(Y \leq y_{1L,\ell})|Z = c^+]$. Because $\sqrt{nh_2}(\widehat{G}_1^{-1}(\hat{q})) - y_{1L,u}) = o_p(1)$, the estimation effect of $\widehat{G}_1^{-1}(\hat{q}))$ will be asymptotically negligible. Therefore,

$$
\begin{aligned}
&\sqrt{nh_2}(\widehat{H}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))) - H_{1,\leq}(y_{1L,\ell})) = \sqrt{nh_2}(\widehat{H}_{1,\leq}(y_{1L,\ell}) - H_{1,\leq}(y_{1L,\ell})) + o_p(1)\\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{+}_{DY1(Y\leq y_{1L,\ell}),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi_{H_{1,\leq}(y_{1L,\ell}),h_2,i} + o_p(1),\\
&\sqrt{nh_2}(\widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q})) - I_{1,\leq}(y_{1L,\ell})) = \sqrt{nh_2}(\widehat{I}_{1,\leq}(y_{1L,\ell}) - I_{1,\leq}(y_{1L,\ell})) + o_p(1)\\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi^{+}_{D1(Y\leq y_{1L,\ell}),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{I_{1,\leq}(y_{1L,\ell}),h_2,i} + o_p(1).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
&\sqrt{nh_2}(\hat{\theta}_1 - \theta_1)\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[D|Z=c^-] - \widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[DY|Z=c^-]\\
&\qquad - H_{1,\leq}(G_1^{-1}(q))\cdot E[D|Z=c^-] + I_{1,\leq}(G_1^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\sqrt{nh_2}(\widehat{H}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[D|Z=c^-] - H_{1,\leq}(G_1^{-1}(q))\cdot E[D|Z=c^-])\\
&\qquad - \sqrt{nh_2}(\widehat{I}_{1,\leq}(\widehat{G}_1^{-1}(\hat{q}))\cdot\widehat{E}_{h_2}[DY|Z=c^-] - I_{1,\leq}(G_1^{-1}(q))\cdot E[DY|Z=c^-])\\
=&\frac{1}{\sqrt{nh_2},}\sum_{i=1}^{n}\Big\{E[D|Z=c^-]\cdot\phi_{H_{1,\leq}(G_1^{-1}(q)),h_2,i} + H_{1,\leq}(G_1^{-1}(q))\cdot\phi^{-}_{D,h_2,i}\\
&\qquad - E[DY|Z=c^-]\cdot\phi_{I_{1,\leq}(G_1^{-1}(q)),h_2,i} - I_{1,\leq}(G_1^{-1}(q))\cdot\phi^{-}_{DY,h_2,i}\Big\} + o_p(1)\\
\equiv&\frac{1}{\sqrt{nh_2},}\sum_{i=1}^{n}\phi_{\theta_1,h_2,i} + o_p(1),
\end{aligned}
$$

and the influence function is identical to the continuous case. Then by the central limit theorem, we have $\sqrt{nh_2}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, V_{\theta_1})$. 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 $dI_{0,\geq}(y)/dy = -P_{0|0} \cdot g_0(y)$. Note that

$$
\begin{aligned}
&\sqrt{nh_2}(\widehat{H}_{1,\geq}(\widehat{G}_1^{-1}(1-\hat{q})) - H_{1,\geq}(G_1^{-1}(1-q))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{+}_{DY1(Y\leq G_1^{-1}(1-q)),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{1,\geq}(G_1^{-1}(1-q)),h_2,i} + o_p(1) \\
&\sqrt{nh_2}(\widehat{H}_{0,\leq}(\widehat{G}_0^{-1}(\hat{r})) - H_{0,\leq}(G_0^{-1}(r))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{-}_{(1-D)Y1(Y\leq G_0^{-1}(r)),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{0,\leq}(G_0^{-1}(r)),h_2,i} + o_p(1), \\
&\sqrt{nh_2}(\widehat{H}_{0,\geq}(\widehat{G}_0^{-1}(1-\hat{r})) - H_{1,\geq}(G_0^{-1}(1-r))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{-}_{(1-D)Y1(Y\leq G_1^{-1}(1-r)),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\phi_{H_{0,\geq}(G_0^{-1}(1-r)),h_2,i} + o_p(1).
\end{aligned}
$$

We also have

$$
\begin{aligned}
&\sqrt{nh_2}(\widehat{I}_{1,\geq}(\widehat{G}_1^{-1}(1-\hat{q})) - I_{1,\geq}(G_1^{-1}(1-q))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{+}_{D1(Y\leq G_1^{-1}(1-q)),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi_{I_{1,\geq}(G_1^{-1}(1-q)),h_2,i} + o_p(1), \\
&\sqrt{nh_2}(\widehat{I}_{0,\leq}(\widehat{G}_0^{-1}(\hat{r})) - I_{0,\leq}(G_0^{-1}(r))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{-}_{(1-D)1(Y\leq G_0^{-1}(r)),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi_{I_{0,\leq}(G_0^{-1}(r)),h_2,i} + o_p(1), \\
&\sqrt{nh_2}(\widehat{I}_{0,\geq}(\widehat{G}_0^{-1}(1-\hat{r})) - I_{1,\geq}(G_0^{-1}(1-r))) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi^{-}_{(1-D)1(Y\leq G_1^{-1}(1-r)),h_2,i} + o_p(1) \equiv \frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi_{I_{0,\geq}(G_0^{-1}(1-r)),h_2,i} + o_p(1).
\end{aligned}
$$

Finally, for the continuous case, we have

$$
\begin{aligned}
&\sqrt{nh_2}(\hat{\theta}_2 - \theta_2) = \sqrt{nh_2}(\widehat{I}_{1,\geq}(\widehat{G}_1^{-1}(1-\hat{q})) \cdot \widehat{E}_{h_2}[DY|Z = c^-] - \widehat{H}_{1,\geq}(\widehat{G}_1^{-1}(1-\hat{q})) \cdot \widehat{E}_{h_2}[D|Z = c^-] \\
&\qquad - I_{1,\geq}(G_1^{-1}(1-q)) \cdot E[DY|Z = c^-] + H_{1,\geq}(G_1^{-1}(1-q)) \cdot E[D|Z = c^-]) \\
=&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\Big\{E[DY|Z = c^-]\phi_{I_{1,\geq}(G_1^{-1}(1-q)),h_2,i} - I_{1,\geq}(G_1^{-1}(1-q))\phi^{-}_{DY,h_2,i} \\
&\qquad - E[D|Z = c^-]\phi_{H_{1,\geq}(G_1^{-1}(1-q)),h_2,i} + H_{1,\geq}(G_1^{-1}(1-q))\phi^{-}_{D,h_2,i}\Big\} + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh_2}}\sum_{i=1}^{n}\phi_{\theta_2,h_2,i} + o_p(1),
\end{aligned}
$$

$$\sqrt{nh}(\hat{\theta}_3-\theta_3)=\sqrt{nh_2}(\widehat{H}_{0,\leq}(\widehat{G}_0^{-1}(\hat{r}))\cdot\widehat{E}_{h_2}[1-D|Z=c^+]-\widehat{I}_{0,\leq}(\widehat{G}_0^{-1}(\hat{r}))\cdot\widehat{E}_{h_2}[(1-D)Y|Z=c^+]$$
$$-H_{0,\leq}(G_0^{-1}(r))\cdot E[1-D|Z=c^+]+I_{1,\leq}(G_1^{-1}(q))\cdot E[(1-D)Y|Z=c^+])$$
$$=\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n\Big\{E[(1-D)|Z=c^+]\phi_{H_{0,\leq}(G_0^{-1}(r)),h_2,i}+H_{0,\leq}(G_0^{-1}(r))\phi^+_{1-D,h_2,i}$$
$$-E[(1-D)Y|Z=c^+]\phi_{I_{0,\leq}(G_0^{-1}(r)),h_2,i}-I_{0,\leq}(G_0^{-1}(r))\phi^+_{(1-D)Y,h_2,i}\Big\}+o_p(1)$$
$$\equiv\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n\phi_{\theta_3,h_2,i}+o_p(1),$$
$$\sqrt{nh_2}(\hat{\theta}_4-\theta_4)=\sqrt{nh_2}(\widehat{I}_{0,\geq}(\widehat{G}_0^{-1}(1-\hat{r}))\cdot\widehat{E}_{h_2}[(1-D)Y|Z=c^+]-\widehat{H}_{0,\geq}(\widehat{G}_0^{-1}(1-\hat{r}))\cdot\widehat{E}_{h_2}[1-D|Z=c^+]$$
$$-I_{0,\geq}(G_0^{-1}(1-r))\cdot E[(1-D)Y|Z=c^+]+H_{0,\geq}(G_0^{-1}(1-r))\cdot E[1-D|Z=c^+])$$
$$=\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n\Big\{E[DY|Z=c^+]\phi_{I_{0,\geq}(G_0^{-1}(1-r)),h_2,i}-I_{0,\geq}(G_0^{-1}(1-r))\phi^+_{(1-D)Y,h_2,i}$$
$$-E[D|Z=c^+]\phi_{H_{0,\geq}(G_0^{-1}(1-r)),h_2,i}+H_{0,\geq}(G_0^{-1}(1-r))\phi^+_{1-D,h_2,i}\Big\}+o_p(1)$$
$$\equiv\frac{1}{\sqrt{nh_2}}\sum_{i=1}^n\phi_{\theta_4,h_2,i}+o_p(1).$$

For discrete case, the expressions are identical to those for continuous cases based on Lemma B.1'.

References

Acerenza, S., O. Bartalotti, and D. Kédagni (2023): "Testing identifying assumptions in bivariate probit models," *Journal of Applied Econometrics*, 38(3), 407–422.

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 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.

Laffers, L., and G. Mellace (2017): "A note on testing instrument validity for the identification of LATE," *Empirical Economics*, 53, 1281–1286.

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.

Olma, T. (2021): "Nonparametric estimation of truncated conditional expectation functions," *arXiv preprint arXiv:2109.06150*.

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.