Monotonicity Test for Local Average Treatment Effects Under Regression Discontinuity

Yu-Chin Hsu[†] **Shu Shen**[‡]

This version: July 15, 2018

Abstract

Researchers are often interested in the relationship between treatment effects and observed individual heterogeneity. This paper proposes the first nonparametric monotonicity test under the popular regression discontinuity framework. The proposed test examines whether the average treatment effect or the local average treatment effect has a monotonic relationship with some of the observed individual characteristics. We show consistency and asymptotic uniform size control of the proposed test. We apply the test to study the heterogeneous effect of attending a more selective high school with respect to peer quality.

Keywords: average treatment effect, local average treatment effect, regression discontinuity, regression monotonicity

[†] Yu-Chin Hsu, Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University. E-mail: ychsu@econ.sinica.edu.tw. [‡] Shu Shen, Department of Economics, University of California, Davis. E-mail: shushen@ucdavis.edu.
Acknowledgement: Yu-Chin Hsu gratefully acknowledges the research support from Ministry of Science and Technology of Taiwan (MOST103-2628-H-001-001-MY4) and Career Development Award of Academia Sinica, Taiwan. Shu Shen gratefully acknowledges the research support from the Hellman Fellows Award. All errors are the authors'.

1 Introduction

In program evaluation, researchers are often interested in knowing the whole picture of a treatment effect that is beyond the overall population average. Estimators and tests of treatment effect heterogeneity (e.g., Heckman et al., 1998; Abadie, 2002; Hotz et al., 2005; Firpo, 2007; Crump et al., 2008; Wager and Athey, forthcoming, among many others) therefore play an important role in the literature. Researchers use these methods to quantify treatment effects for different groups of individuals, to look for relationships between the effects and observed factors, to understand how a policy intervention can affect tails of an outcome distribution, or to design extensions of the analyzed treatment to other populations.

Treatment effect heterogeneity analysis is also important in program evaluation studies that use the regression discontinuity (RD) design, which has become very popular in applied microeconomics, following pioneering works of Angrist and Lavy (1999), Black (1999), and van der Klaauw (2002). In this paper, we propose the first nonparametric RD monotonicity test for examining whether an average treatment effect (ATE) identified under a sharp RD design or a local average treatment effect (LATE) identified under a fuzzy RD design has a monotonic relationship with some of the observed individual characteristics.

The proposed test is important to the RD literature as applied economists are often interested in testing for such a relationship. For example, Ito (2015) studies the treatment effect of an electricity rebate program in California and is interested in knowing whether the effect on electricity consumption increases with average temperature or decreases with household income. Carneiro et al. (2015) investigate the impact of increased maternity leave on children's long-term outcome in Norway and examine how the effect changes with family characteristics such as mothers' education, distance to grandparents, and income. Barone et al. (2015) study the effect of being exposed to slanted information on decision making using a quasi-natural experiment in Italy. They find that switching to digital TV, which has tenfold more programs, decreased Italians' vote share of Berlusconi's coalition and the effect was stronger in towns with older and less educated voters.

Despite its popularity, tools used in the applied literature to test for monotonic relationships between treatment effects and observed individual heterogeneity are quite informal. Both Ito (2015) and Barone et al. (2015) use a naive interaction term method that adds interaction terms between additional controls of interest and the dummy variable indicating whether an observation passes the cut-off value of an underlying variable that affects treatment take-up decisions. The method would conclude that the treatment effect increases (decreases) with certain observable if the corresponding interaction term is positive (negative) and statistically significant. Carneiro et al. (2015), on the other hand, subsample the population by grouping individuals with quartiles of continuous random variables such as mothers' months of unpaid leave and log income and run separate subsample RD analysis. However, the interaction term method is parametric and subject to model misspecification, while the subsampling method is ad-hoc and often results in loss of information. In contrast, our proposed test is nonparametric and robust against all forms of deviations from the null.

Our test also contributes to the regression monotonicity literature in statistics and econometrics. Existing nonparametric regression monotonicity tests (e.g., Ghosal et al., 2000; Hall and Heckman, 2000; Chetverikov, 2013; Hsu et al., 2017), to the best of our knowledge, consider only the case of interior estimation. Since the RD treatment effect is estimated through nonparametric boundary estimation, none of the existing monotonicity tests could be applied to the RD setup. Our proposed test is hence the first regression monotonicity test that is compatible with nonparametric boundary estimation.

This paper is also related to the large literature on regression discontinuity, especially some recent developments that also look at treatment effect heterogeneity. Relevant papers include Frandsena et al. (2012) and Shen and Zhang (2015) for distributional RD analysis, Dong and Lewbel (2015) and Angrist and Rokkanen (2015) for extrapolating RD effects away from the cut-off, Bertanha (2016) and Cattaneo et al. (2016) for analyzing heterogeneous treatment effect when the RD design has multiple cut-offs, Bertanha and Imbens (2014) for examining the external validity of LATE under the fuzzy RD design, and Hsu and Shen (2016) for testing whether a treatment effect is heterogeneous among individuals with different observed characteristics.

To construct our test, we first formulate the null hypothesis of regression monotonicity as a conditional moment inequality that conditions on both the running variable of the

RD model and some other controls. We then use instrumental functions to transform the moment inequality into a series of conditional moment inequalities that condition only on the running variable and build our test statistic upon the latter. Critical values are constructed through multiplier bootstrap. The proposed nonparametric RD monotonicity test has several advantages. First, the test statistic does not involve nonparametric derivative estimation and is of order $(nh)^{-1/2}$, the same rate of convergence as the classic RD treatment effect estimators. Second, the proposed test has uniform size control over a broad set of data generating processes. Last but not least, the test is robust to weak identification of the conditional local average treatment effect (CLATE) under fuzzy RD designs. This is important because even when the identification of LATE is strong, the CLATE could be weakly identified for some subpopulations due to first stage heterogeneity.

The instrumental function approach adopted in the paper is related to Andrews and Shi (2013, 2014) and Hsu et al. (2017), who propose to reduce the dimension of the conditioning set in conditional moment equalities/inequalities by transforming the outcome variable with a series of countably many instrumental functions. They also show that such a transformation brings no loss of information. Our test is most related to Hsu et al. (2017), who extend the instrumental function approach in Andrews and Shi (2013, 2014) to test generalized regression monotonicity. However, as is discussed earlier, the general method developed in Hsu et al. (2017) cannot be applied to the RD framework because their test is not compatible with nonparametric boundary estimators. Our proposed test extends the testing idea in Hsu et al. (2017) to the RD setting.

Monte Carlo experiments show that our proposed test has great size and power properties. We apply the test to study the impact of attending a more selective high school in Romania following Pop-Eleches and Urquiola (2013). Mean analysis in Pop-Eleches and Urquiola (2013) finds that going to a better high school significantly improves grade averages in the Baccalaureate exam but does not seem to affect the probability of a student taking the Baccalaureate exam. In contrast, our monotonicity test reveals that the effect on the exam-taking rate increases monotonically with peer quality of the more selective school, indicating that the insignificant mean effect found in Pop-Eleches and Urquiola (2013) comes from the cancelation of positive and negative treatment effects among different schools.

The paper is organized as follows. Section 2 sets up the RD model and proposes the benchmark monotonicity test for the general fuzzy RD case. Section 3 discusses the asymptotic property of the proposed test. Section 4 extends the benchmark test to the special case of sharp RD. Section 5 examines the small sample performance of the proposed test and Section 6 carries out the empirical application. Proofs of all lemmas and theorems are provided in the Appendix.

2 Testing Treatment Effect Monotonicity Under the Fuzzy RD Design

2.1 Model Set-up and Null Hypothesis

Let Y denote the outcome of interest and T the treatment status of an individual. T is binary; $T = 0$ if an individual does not take the treatment and $T = 1$ if he/she does. Use $Y(0)$ and $Y(1)$ to denote potential outcomes when $T = 0$ and $T = 1$, respectively. The observed outcome is $Y = T \cdot Y(1) + (1 - T) \cdot Y(0)$. Whether an individual receives treatment or not depends at least partially on an underlying variable Z, called the running variable. A policy intervention encourages an individual to receive treatment if Z is larger than or equal to some known cut-off value c. Let $T(1)$ and $T(0)$ be the potential treatment decisions of an individual depending on whether he/she is encouraged or not. The observed treatment status is then $T = T(1)1(Z \geq c) + T(0)1(Z < c)$. Let X be a set of covariates with compact support $\mathcal{X} \subset R^{d_x}$. Without loss of generality, we assume that $\mathcal{X} = \times_{j=1}^{d_x}[0, 1]$. For notational simplicity, we assume that X includes only continuous variables. Our results could be easily extended to cases where X includes discrete variables.

The RD model follows a sharp design if the treatment decision T is a deterministic function of the running variable Z, or in other words $T = 1(Z \geq c)$. The model follows a fuzzy design if T is a probabilistic function of Z. In this section, we focus on the more general fuzzy RD case. We will discuss the sharp RD case in Section 4.

Let P be the underlying distribution of (Z, T, X, Y) and use E_P to denote expectation under P. Let $\mathcal{X}_z \subseteq \mathcal{X}$ be the support of X conditional on $Z = z$ and use $\mathcal{X}_c \subset \mathcal{X}$ to denote the compact support of X conditional on $Z = c$. For any $\delta > 0$, let $\mathcal{N}_{\delta,z}(c) = \{z : |z-c| \leq \delta\}$ denote a neighborhood of Z around $Z = c$. The following assumption collects the identifying conditions for the conditional local average treatment effect (CLATE), defined as

$$CLATE(x) = E_P[Y(1) - Y(0)|X = x, Z = c, T(1) - T(0) = 1],$$

for the general fuzzy RD set-up.

Assumption 2.1 *For a running variable Z continuously distributed in $\mathcal{N}_{\delta,z}(c)$ for some $\delta > 0$, assume that*

(i) $E_P[Y(t)|T(1)-T(0) = 1, X = x, Z = z]$ *and* $E_P[Y(t)|T(1) = T(0) = t', X = x, Z = z]$ *are continuous in x and z on $\mathcal{X}_c \times \mathcal{N}_{\delta,z}(c)$ for $t, t' \in \{0, 1\}$;*

(ii) $E_P[T(1) - T(0) = 1|X = x, Z = z]$ *and* $E_P[T(1) = T(0) = t|X = x, Z = z]$ *are continuous in x and z on $\mathcal{X}_c \times \mathcal{N}_{\delta,z}(c)$ for $t \in \{0, 1\}$;*

(iii) $T(1) \geq T(0)$*;*

(iv) $E_P[T(1) - T(0)|X = x, Z = c] > 0$ *for all $x \in \mathcal{X}_c$.*

Assumption 2.1(i) requires the continuity of average potential outcomes for always-takers (individuals with $T(0) = T(1) = 1$), compliers (individuals with $T(1) - T(0) = 1$), and never-takers (individuals with $T(0) = T(1) = 0$) along both the running variable Z and the additional control X near the cutoff value $Z = c$. Assumption 2.1(ii) requires that the proportion of each group be continuous along both Z and X near $Z = c$. Assumption 2.1(iii) and (iv) require no defiers and non-trivial presence of compliers, respectively. Assumption 2.1(i), (ii) and (iv) are stronger than their counterparts that are unconditional on X (c.f. Dong and Lewbel, 2015) as we are interested in identifying the CLATE conditional on the additional covariate X. Under Assumption 2.1, it is easy to show that CLATE is identified by

$$CLATE(x) = \frac{\lim_{z \searrow c} E_P[Y|X = x, Z = z] - \lim_{z \nearrow c} E_P[Y|X = x, Z = z]}{\lim_{z \searrow c} E_P[T|X = x, Z = z] - \lim_{z \nearrow c} E_P[T|X = x, Z = z]}.$$

See Hsu and Shen (2016) for details. The dependence of LATE and CLATE(x) on P is suppressed for notational simplicity.

Researchers are often interested in testing whether the CLATE monotonically increases with some elements in X while keeping the rest fixed. To formalize the test, partition X such that $X = (W, S)$ with $W \in \mathcal{W} = [0,1]^{d_w}$ and $S \in \mathcal{S} = [0,1]^{d_s}$; $d_w \geq 1$, $d_s \geq 0$ and $d_s + d_w = d_x$, and define the null and alternative hypotheses as

$$
\begin{aligned}
&H_{0,FRD} : CLATE(x) \text{ is non-decreasing in } w \text{ on } \mathcal{W} \text{ for all } s \in \mathcal{S}; \qquad (2.1)\\
&H_{1,FRD} : \ H_{0,FRD} \text{ does not hold.}
\end{aligned}
$$

To test the null, a direct approach is to take partial derivative of $CLATE(x)$ with respect to w and examine the sign of the derivative function for all values of w and s. This direct approach requires both the numerator and the denominator of $CLATE(x)$ to be differentiable with respect to w. It also requires the use of nonparametric derivative estimation, which has a slow convergence rate. In this paper, we take an alternative route. Following Hsu et al. (2017), we transform the null hypothesis $H_{0,FRD}$ to conditional moment inequalities that do not involve derivatives.

2.2 Transformation of the Null Hypothesis

For any $w_1, w_2 \in \mathcal{W}$, $s \in \mathcal{S}$, and $q \in \mathcal{Z}_+$, where $\mathcal{Z}_+$ denotes the set of positive intergers, define

$$
\begin{aligned}
&C_{w_1,q} \equiv \prod_{j=1}^{d_w} \left[w_{1j}, w_{1j} + \frac{1}{q+1} \right], \quad C_{w_2,q} \equiv \prod_{j=1}^{d_w} \left[w_{2j}, w_{2j} + \frac{1}{q+1} \right],\\
&C_{s,q} \equiv \prod_{j=1}^{d_s} \left[s_j, s_j + \frac{1}{q} \right],
\end{aligned}
$$

where w_{1j}, w_{2j}, and s_j are the j-th dimension of w_1, w_2 and s. Also, when $d_w \geq 2$, denote $w_1 \geq w_2$ iff $w_{1j} \geq w_{2j}$ for all $j = 1, \ldots, d_w$; denote $w_1 > w_2$ iff $w_{1j} \geq w_{2j}$ for all $j = 1, \ldots, d_w$, and $w_{1k} > w_{2k}$ for some $k \in \{1, \ldots, d_w\}$; denote $w_1 \gg w_2$ iff $w_{1j} > w_{2j}$ for all $j = 1, \ldots, d_w$.

Lemma 2.1 *Let $\lambda(w,s)$ be a continuous function in (w,s) on $\mathcal{W} \times \mathcal{S}$, and $0 < h(w,s) \leq M < \infty$ be a weight function. The following two statements are equivalent:*

(i) $\lambda(w_1,s)\geq\lambda(w_2,s)$ *whenever* $w_1\geq w_2$ *for any* $w_1,w_2\in\mathcal{W}$ *and* $s\in\mathcal{S}$*;*

(ii) for any $q\in\mathcal{Z}_+$*, and* $w_1\geq w_2$ *such that* $(q+1)\cdot w_1,(q+1)\cdot w_2\in\{0,1,2,\cdots,q\}^{d_w}$ *and* $q\cdot s\in\{0,1,2,\ldots,q-1\}^{d_s}$*,*

$$\frac{\int_{C_{w_1,q}\times C_{s,q}}\lambda(\tilde{w},\tilde{s})\cdot h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}{\int_{C_{w_1,q}\times C_{s,q}}h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}\geq\frac{\int_{C_{w_2,q}\times C_{s,q}}\lambda(\tilde{w},\tilde{s})\cdot h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}{\int_{C_{w_2,q}\times C_{s,q}}h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}.$$

Lemma 2.1 states that the monotonicity condition of any continuous function could be re-formulated as countably many moment inequalities. Lemma 2.1 could be applied to simplify the null hypothesis $H_{0,FRD}$.

To be specific, set $\lambda(w,s)$ to $CLATE(w,s)$, then the statement (i) in Lemma 2.1 reduces to the null hypothesis $H_{0,FRD}$. Let the weight function in (ii) be $h(w,s)=E_P[T(1)-T(0)|W=w,S=s,Z=c]\cdot f_{W,S|Z=c}(w,s)$, the inequality in Lemma 2.1.(ii) reduces to

$$\begin{aligned}&\frac{E_P[g_{w_1,\ell}(W)g_{s,\ell}(S)(Y(1)-Y(0))(T(1)-T(0))|Z=c]}{E_P[g_{w_1,\ell}(W)g_{s,\ell}(S)(T(1)-T(0))|Z=c]}\\ \geq&\frac{E_P[g_{w_2,\ell}(W)g_{s,\ell}(S)(Y(1)-Y(0))(T(1)-T(0))|Z=c]}{E_P[g_{w_2,\ell}(W)g_{s,\ell}(S)(T(1)-T(0))|Z=c]},\end{aligned}\tag{2.2}$$

where $g_{w_\kappa,\ell}(.)=1(.\in C_{w_\kappa,q})$ for $\kappa=1,2$, $g_{s,\ell}(.)=1(.\in C_{s,q})$, and $\ell=(w_1,w_2,s,q)\in\mathcal{W}^2\times\mathcal{S}\times\mathcal{Z}_+$.

To further identify the conditional expectations in above inequality, we make the following continuity assumption on the conditional density of X given Z.

Assumption 2.2 $f_{X|Z}(x|z)$*, the probability density function of* $X|Z=z$*, is continuous and uniformly bounded in* z *and* x *on* $\mathcal{N}_{\delta,z}(c)\times\mathcal{X}_c$*.*

Under both Assumptions 2.1 and 2.2, the conditional mean $E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)(Y(1)-Y(0))(T(1)-T(0))|Z=c]$ in the above inequality (2.2) for $\kappa=1,2$ could be identified by $\lim_{z\searrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)Y|Z=z]-\lim_{z\nearrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)Y|Z=z]$, and the conditional mean $E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)(T(1)-T(0))|Z=c]$ could be identified by $\lim_{z\searrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)T|Z=z]-\lim_{z\nearrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)T|Z=z]$.

For $\ell\in\mathcal{W}^2\times\mathcal{S}\times\mathcal{Z}_+$ and $\kappa=1,2$, define $m^{(\kappa)}_{P,+}(\ell)=\lim_{z\searrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)Y|Z=z]$, $m^{(\kappa)}_{P,-}(\ell)=\lim_{z\nearrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)Y|Z=z]$, $q^{(\kappa)}_{P,+}(\ell)=\lim_{z\searrow c}E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)T|Z=$

$z]$, $q_{P,-}^{(\kappa)}(\ell) = \lim_{z \nearrow c} E_P[g_{w_\kappa,\ell}(W) g_{s,\ell}(S) T | Z = z]$, $\rho_P^{(\kappa)}(\ell) = m_{P,+}^{(\kappa)}(\ell) - m_{P,-}^{(\kappa)}(\ell)$, and $\varrho_P^{(\kappa)}(\ell) = q_{P,+}^{(\kappa)}(\ell) - q_{P,-}^{(\kappa)}(\ell)$. The following lemma summarizes the hypothesis transformation result discussed above.

Lemma 2.2 *Under Assumption 2.1 and 2.2, the null hypothesis $H_{0,FRD}$ is equivalent to*

$$H'_{0,FRD} : \mu_P(\ell) \equiv \rho_P^{(2)}(\ell)\varrho_P^{(1)}(\ell) - \rho_P^{(1)}(\ell)\varrho_P^{(2)}(\ell) \leq 0, \quad \text{for all } g_\ell \in \mathcal{G}, \tag{2.3}$$

where $\mathcal{G} = \{g_\ell = (g_{w_1,\ell}, g_{w_2,\ell}, g_{s,\ell}) : \ell \in \mathcal{L}\}$ is a set of the indicator functions of countable hypercubes with

$$\mathcal{L} = \Big\{(w_1, w_2, s, q) : \ (q+1)\cdot(w_1, w_2) \in \{0, 1, 2, \cdots, q\}^{2d_w}, w_1 \geq w_2, \\ q \cdot s \in \{0, 1, 2, \cdots, q-1\}^{d_s}, \ \text{for } q = 1, 2, 3 \cdots \Big\}. \tag{2.4}$$

Proof of Lemma 2.2 is given in the appendix. Note that the inequality in (2.3) also avoids the use of fraction terms in inequality (2.2). The monotonicity test built upon (2.3) is therefore robust to weak identification. This is important for the proposed test to have good small sample performance because first stage heterogeneity can easily result in weak identification of $CLATE(x)$ for some values of x even when the identification of $LATE$ is strong.

Before we move on to the next section to construct test statistic for the null hypothesis $H'_{0,FRD}$, we would like to point out that Assumption 2.2 required in Lemma 2.2 is not a strong condition. In fact, Assumption 2.2 is a direct implication of the "no precise control over the running variable" rule introduced by Lee and Lemieux (2010) and well-accepted in the applied RD literature. See discussions in Hsu and Shen (2016) for details.

2.3 Test Statistic

Let $\{Z_i, T_i, X_i, Y_i\}_{i=1}^n$ be a sample of size n randomly drawn from the underlying distribution of (Z, T, X, Y). In the following, we introduce the nonparametric local linear estimator of the moment function $\mu_P(\ell)$ defined in equation (2.3), and then the test statistic.

For $\ell \in \mathcal{L}$ and $\kappa = 1, 2$, let $\hat{m}_{n,+}^{(\kappa)}(\ell)$, $\hat{m}_{n,-}^{(\kappa)}(\ell)$, $\hat{p}_{n,+}^{(\kappa)}(\ell)$ and $\hat{p}_{n,-}^{(\kappa)}(\ell)$ be the local linear estimators of $m_{P,+}^{(\kappa)}(\ell)$, $m_{P,-}^{(\kappa)}(\ell)$, $p_{P,+}^{(\kappa)}(\ell)$ and $p_{P,-}^{(\kappa)}(\ell)$. Let $K(.)$ be the symmetric kernel function and h the bandwidth; estimators $\hat{m}^{(\kappa)}_{n,+}(\ell)$, $\hat{m}^{(\kappa)}_{n,-}(\ell)$, $\hat{p}^{(\kappa)}_{n,+}(\ell)$ and $\hat{p}^{(\kappa)}_{n,-}(\ell)$ are the constant terms of the following minimization problems, respectively.

$$(\hat{m}^{(\kappa)}_{n,+}(\ell), \hat{b}^{(\kappa)}_{n,m+}(\ell)) = \arg\min_{a,b} \sum_{Z_i \geq c}^{n} K\left(\frac{Z_i - c}{h}\right)\left[g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) Y_i - a - b(Z_i - c)\right]^2,$$
$$(\hat{m}^{(\kappa)}_{n,-}(\ell), \hat{b}^{(\kappa)}_{n,m-}(\ell)) = \arg\min_{a,b} \sum_{Z_i < c}^{n} K\left(\frac{Z_i - c}{h}\right)\left[g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) Y_i - a - b(Z_i - c)\right]^2,$$
$$(\hat{q}^{(\kappa)}_{n,+}(\ell), \hat{b}^{(\kappa)}_{n,q+}(\ell)) = \arg\min_{a,b} \sum_{Z_i \geq c}^{n} K\left(\frac{Z_i - c}{h}\right)\left[g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) T_i - a - b(Z_i - c)\right]^2,$$
$$(\hat{q}^{(\kappa)}_{n,-}(\ell), \hat{b}^{(\kappa)}_{n,q-}(\ell)) = \arg\min_{a,b} \sum_{Z_i < c}^{n} K\left(\frac{Z_i - c}{h}\right)\left[g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) T_i - a - b(Z_i - c)\right]^2.$$

Following Fan and Gijbels (1992), for $l = 0, 1, 2$, define

$$S^+_{n,l} = \sum_{i=1}^{n} 1(Z_i \geq c) K\left(\frac{Z_i - c}{h}\right)(Z_i - c)^l, \;\; S^-_{n,l} = \sum_{i=1}^{n} 1(Z_i < c) K\left(\frac{Z_i - c}{h}\right)(Z_i - c)^l,$$
$$\mathsf{w}^+_{ni} = \frac{1(Z_i \geq c) K(\frac{Z_i - c}{h})[S^+_{n,2} - S^+_{n,1}(Z_i - c)]}{S^+_{n,0} S^+_{n,2} - S^+_{n,1} S^+_{n,1}}, \;\; \mathsf{w}^-_{ni} = \frac{1(Z_i < c) K(\frac{Z_i - c}{h})[S^-_{n,2} - S^-_{n,1}(Z_i - c)]}{S^-_{n,0} S^-_{n,2} - S^-_{n,1} S^-_{n,1}}.$$

Then estimators $\hat{m}^{(\kappa)}_{n,+}(\ell)$, $\hat{m}^{(\kappa)}_{n,-}(\ell)$, $\hat{q}^{(\kappa)}_{n,+}(\ell)$ and $\hat{q}^{(\kappa)}_{n,-}(\ell)$ could be re-written as

$$\hat{m}^{(\kappa)}_{n,+}(\ell) = \sum_{i=1}^{n} \mathsf{w}^+_{ni} \cdot m^{(\kappa)}(Y_i, W_i, S_i, \ell), \quad \hat{m}^{(\kappa)}_{n,-}(\ell) = \sum_{i=1}^{n} \mathsf{w}^-_{ni} \cdot m^{(\kappa)}(Y_i, W_i, S_i, \ell),$$
$$\hat{q}^{(\kappa)}_{n,+}(\ell) = \sum_{i=1}^{n} \mathsf{w}^+_{ni} \cdot q^{(\kappa)}(T_i, W_i, S_i, \ell), \quad \hat{q}^{(\kappa)}_{n,-}(\ell) = \sum_{i=1}^{n} \mathsf{w}^-_{ni} \cdot q^{(\kappa)}(T_i, W_i, S_i, \ell).$$

where $m^{(\kappa)}(Y_i, W_i, S_i, \ell) = g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) Y_i$, and $q^{(\kappa)}(T_i, W_i, S_i, \ell) = g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) T_i$, for $\ell \in \mathcal{L}$ and $\kappa = 1, 2$.

Let $\hat{\rho}^{(\kappa)}_n(\ell) = \hat{m}^{(\kappa)}_{n,+}(\ell) - \hat{m}^{(\kappa)}_{n,-}(\ell)$, $\hat{\varrho}^{(\kappa)}_n(\ell) = \hat{q}^{(\kappa)}_{n,+}(\ell) - \hat{q}^{(\kappa)}_{n,-}(\ell)$, and $\hat{\mu}_n(\ell) = \hat{\rho}^{(2)}_n(\ell)\hat{\varrho}^{(1)}_n(\ell) - \hat{\rho}^{(1)}_n(\ell)\hat{\varrho}^{(2)}_n(\ell)$ be the estimators of $\rho^{(\kappa)}_P(\ell)$, $\varrho^{(\kappa)}_P(\ell)$, and $\nu_P(\ell)$, respectively. These estimators are uniformly consistent over $\ell \in \mathcal{L}$ under proper regularity conditions. Furthermore, under suitable conditions, we can show that uniformly over $\ell \in \mathcal{L}$,

$$\begin{aligned}
&\sqrt{nh}(\hat{\mu}_n(\ell) - \mu_P(\ell)) \\
=& \sum_{i=1}^{n} \varrho^{(1)}_P(\ell) \cdot \phi^{(2)}_{\rho,ni}(\ell) + \rho^{(2)}_P(\ell) \cdot \phi^{(1)}_{\varrho,ni}(\ell) - \varrho^{(2)}_P(\ell) \cdot \phi^{(1)}_{\rho,ni}(\ell) - \rho^{(1)}_P(\ell) \cdot \phi^{(2)}_{\varrho,ni}(\ell) + o_p(1) \\
\equiv& \sum_{i=1}^{n} \phi_{\mu,ni}(\ell) + o_p(1),
\end{aligned}$$

with $\phi^{(\kappa)}_{\rho,ni}(.) = \sqrt{nh}(\mathsf{w}^+_{ni}(m^{(\kappa)}(Y_i,W_i,S_i,.) - m^{(\kappa)}_{P,+}(.)) - \mathsf{w}^-_{ni}(m^{(\kappa)}(Y_i,W_i,S_i,.) - m^{(\kappa)}_{P,-}(.)))$ and $\phi^{(\kappa)}_{\varrho,ni}(.) = \sqrt{nh}(\mathsf{w}^+_{ni}(q^{(\kappa)}(T_i,W_i,S_i,.) - q^{(\kappa)}_{P,+}(.)) - \mathsf{w}^-_{ni}(q^{(\kappa)}(T_i,W_i,S_i,.) - q^{(\kappa)}_{P,-}(.)))$. This influence function representation will be used to derive the weak limit of $\hat{\mu}_n(.)$ in Section 3.

For any $\ell \in \mathcal{L}$, let $\hat{\sigma}^2_{\mu,n}(\ell) = \sum_{i=1}^n \hat{\phi}^2_{\mu,ni}(\ell)$, where $\hat{\phi}_{\mu,ni}(\ell)$ is the estimated influence function with $\rho^{(\kappa)}_P(\ell)$ and $\varrho^{(\kappa)}_P(\ell)$ in $\phi_{\mu,ni}(\ell)$ replaced by $\hat{\rho}^{(\kappa)}_n(\ell)$ and $\hat{\varrho}^{(\kappa)}_n(\ell)$, respectively. In Lemma B.4 in the appendix, we show that $\hat{\sigma}^2_{\mu,n}(\ell)$ is a consistent estimator for the asymptotic variance of $\sqrt{nh}(\hat{\mu}_n(\ell) - \mu_P(\ell))$ under proper regularity conditions. Let ϵ be some small positive number and $\ell_0 = (\mathbf{1/2}, \mathbf{0}, \mathbf{0}, 1)$. Let $\hat{\sigma}^2_{\mu,\epsilon}(\ell) = \max\left\{\hat{\sigma}^2_{\mu,n}(\ell), \epsilon \cdot \hat{\sigma}^2_{\mu,n}(\ell_0)\right\}$, which manually bounds the variance estimator away from zero. To test the null hypothesis $H'_{0,FRD}$ described in the previous section, we use the Kolmogorov-Smirnov type statistic

$$\widehat{T}_{n,FRD} = \sup_{\ell \in \mathcal{L}} \sqrt{nh} \frac{\hat{\mu}_n(\ell)}{\hat{\sigma}_{\mu,\epsilon}(\ell)}.$$

Under regularity conditions, the test statistic converges to a known limiting distribution when the monotonicity condition in $H_{0,FRD}$ is true and diverges if the monotonicity condition is violated.

2.4 LFC and GMS Based Critical Values and Decision Rules

In this section, we introduce two simulated critical values for the proposed tests. The first is based on the least favorable condition (LFC), which is simple and popular but potentially conservative for moment inequality tests. The second is based on the generalized moment selection (GMS) method introduced by Andrews and Soares (2010) and Andrews and Shi (2013, 2014, 2017), which is often employed to improve the power of inequality tests over the LFC method. [1]

To construct the simulated critical values, we introduce a multiplier bootstrap method that can simulate a process that converges to the same limit as $\sqrt{nh}(\hat{\mu}_n(\ell) - \mu_P(\ell))$.

[1] The recentering method proposed by Hansen (2005), Donald and Hsu (2016), as well as the contact set approach proposed in Linton et al. (2010) could also be adopted to improve the power of our monotonicity tests.

Let $\{U_i : 1 \le i \le n\}$ be a sequence of i.i.d. random variables that satisfy the following conditions.

Assumption 2.3 *$\{U_i : 1 \le i \le n\}$ is a sequence of i.i.d. random variables that is independent of the sample path of $\{(Y_i, X_i, Z_i, T_i) : 1 \le i \le n\}$ such that $E[U_i] = 0$, $E[U_i^2] = 1$, and $E[|U_i|^4] < M$ for some $M > 0$.*

Construct the simulated process as

$$\widehat{\Phi}^u_{\mu,n}(\ell) = \sum_{i=1}^{n} U_i \cdot \hat{\phi}_{\mu,ni}(\ell), \tag{2.5}$$

where $\hat{\phi}_{\mu,ni}(\ell)$ are the estimated influence functions defined earlier. For significance level $\alpha < 1/2$, define the LFC critical value as

$$\hat{c}^{\eta,LFC}_{n,FRD}(\alpha) = \sup\left\{ q \middle| P^u\Big(\sup_{\ell \in \mathcal{L}} \frac{\widehat{\Phi}^u_{\mu,n}(\ell)}{\hat{\sigma}_{\mu,\epsilon}(\ell)} \le q \Big) \le 1 - \alpha + \eta \right\} + \eta,$$

where P^u is the multiplier probability measure. The LFC critical value is therefore the $(1-\alpha+\eta)$-th quantile of the simulated distribution of $\sup_{\ell \in \mathcal{L}} \frac{\widehat{\Phi}^u_{\mu,n}(\ell)}{\hat{\sigma}_{\mu,\epsilon}(\ell)}$ plus a small positive constant η. We will reject the null if the test statistic $\widehat{T}_{n,FRD}$ is larger than the critical value $\hat{c}^{\eta,LFC}_{n,FRD}(\alpha)$. The small positive constant η is required for our testing approach to have uniform size control over a broad range of data generating processes (DGPs). If pointwise size control is desired, η could be set to zero.

Alternatively, one could adopt a GMS method to construct the simulated critical value.

Assumption 2.4 *Assume that*

(i) Let a_n be a sequence of non-negative numbers satisfying $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} a_n/\sqrt{nh} = 0$;

(ii) Let B_n be a sequence of non-negative numbers satisfying that B_n is non-decreasing, $\lim_{n\to\infty} B_n = \infty$ and $\lim_{n\to\infty} B_n/a_n = 0$.

Define the GMS critical value as

$$\hat{c}^{\eta,GMS}_{n,FRD}(\alpha) = \sup\left\{ q \middle| P^u\left(\sup_{\ell \in \mathcal{L}} \left(\frac{\widehat{\Phi}^u_{\mu,n}(\ell)}{\hat{\sigma}_{\mu,\epsilon}(\ell)} + \hat{\psi}_\mu(\ell) \right) \le q \right) \le 1 - \alpha + \eta \right\} + \eta,$$

$$\hat{\psi}_\mu(\ell) = -B_n \cdot 1\left(\sqrt{nh} \cdot \frac{\hat{\mu}_n(\ell)}{\hat{\sigma}_{\mu,\epsilon}(\ell)} < -a_n \right).$$

Compared with the LFC critical value, the GMS critical value uses the $\hat{\psi}_\mu(\ell)$ term to suppress the influence of negative moment functions on the simulated critical value and improve the power of the proposed test. The tuning non-negative sequences a_n and B_n used to define the $\hat{\psi}_\mu(\ell)$ term are required to satisfy the conditions in Assumption 2.4. In practice, we follow Andrews and Shi (2013, 2014) and use $a_n = (0.3\ln(n))^{1/2}$, $B_n = (0.4\ln(n)/\ln\ln(n))^{1/2}$, and $\eta = 10^{-6}$. Again, the decision rule is to reject the null when the test statistic is larger than the critical value.

3 Asymptotics of Proposed Tests

In this section, we study the asymptotic properties of the proposed tests based on both decision rules discussed in the previous section.

3.1 Regularity Conditions and Asymptotics of the Local Linear Estimators

Let $f_z(z)$ and $f_{xz}(x,z)$ denote the marginal density function of Z, and the joint density of X and Z, respectively. Let $\zeta_{P,+}(x,z) = E_P[Y|X=x, Z=z]$, $\sigma^2_{P,+}(x,z) = V_P(Y|X=x, Z=z)$ and $\varsigma_{P,+}(x,z) = E_P[T|X=x, Z=z]$ for $z \geq c$ and $\zeta_{P,-}(x,z) = E_P[Y|X=x, Z=z]$, $\sigma^2_{P,-}(x,z) = V_P(Y|X=x, Z=z)$ and $\varsigma_{P,-}(x,z) = E_P[T|X=x, Z=z]$ for $z < c$. Let $\mathcal{N}^+_{\delta,z}(c) = \{z : \ 0 \leq z - c \leq \delta\}$ be a neighborhood of Z from the cut-off value c to the right and $\mathcal{N}^-_{\delta,z}(c) = \{z : \ 0 < c - z \leq \delta\}$ be a neighborhood from c to the left. Let P_z denote the distribution of Z under P, and let $\mathcal{P}$ denote the collection of distributions P. We make the following assumptions.

Assumption 3.1 *Assume that for some $\delta > 0$ and all $P \in \mathcal{P}$, the following conditions are satisfied.*

(i) *$\mathcal{X}_z = \mathcal{X}_c$ for all $z \in \mathcal{N}_{\delta,z}(c)$.*

(ii) *The random vaiable Z has the same distribution across all $P \in \mathcal{P}$ and its density $f_z(z)$ is uniformly bounded away from zero and twice continuously differentiable in z on $\mathcal{N}_{\delta,z}(c)$.*

(iii) $f_{xz}(x,z)$ *is twice continuously differentiable in* z *on* $\mathcal{N}_{\delta,z}(c)$ *for all* $x \in \mathcal{X}_c$*, and* $\partial^2 f_{xz}(x,z)/\partial x \partial z$ *is uniformly bounded on* $\in \mathcal{X}_c \times \mathcal{N}_{\delta,z}(c)$.

(iv) for all $x \in \mathcal{X}_c$*,* $\zeta_{P,+}(x,z)$*,* $\varsigma_{P,+}(x,z)$*,* $\zeta_{P,-}(x,z)$*, and* $\varsigma_{P,-}(x,z)$ *are all twice continuously differentiable in* z *on* $\mathcal{N}^+_{\delta,z}(c)$.

(v) $\partial \zeta_{P,+}(x,z)/\partial z$, $\partial^2 \zeta_{P,+}(x,z)/\partial x \partial z$, $\partial \varsigma_{P,+}(x,z)/\partial z$, $\partial^2 \varsigma_{P,+}(x,z)/\partial x \partial z$, $\partial z \zeta_{P,-}(x,z)/\partial z$, $\partial^2 \zeta_{P,-}(x,z)/\partial x \partial z$, $\partial \varsigma_{P,-}(x,z)/\partial z$, *and* $\partial^2 \varsigma_{P,-}(x,z)/\partial x \partial z$ *are all uniformly bounded on* $\mathcal{X}_c \times \mathcal{N}^+_{\delta,z}(c)$.

(vi) Both $\sigma^2_{P,+}(x,z)$ *and* $\sigma^2_{P,-}(x,z)$ *are uniformly bounded and uniformly bounded away from zero on* $\mathcal{X}_c \times \mathcal{N}^+_{\delta,z}(c)$.

(vii) $E_P[Y^4|Z=z]$ *is uniformly bounded for all* $z \in \mathcal{N}_{\delta,z}(c)$.

(viii) $E_P[T(1) - T(0)|Z=c]$ *is uniformly bounded away from zero.*

Assumption 3.1(i) is assumed for notational simplicity. We can allow $\mathcal{X}_z$ to depend on z, and the theory will be the same, but it is more tedious in terms of notation. Assumption 3.1(ii)-(iv) are standard in nonparametric estimation. Assumption 3.1(v) is needed to show that the bias terms of the $\hat{\nu}(\ell)$ are asymptotically negligible uniformly over $\ell \in \mathcal{L}$ and $P \in \mathcal{P}$. Assumption 3.1(vi) and (vii) are required for the covariance estimator of the limiting process to be uniformly consistent, which is in turn used for the showing the validity of the multiplier bootstrap. Similar conditions are also assumed in Andrews and Shi (2014), Hsu (2016) and Hsu and Shen (2016). Assumption 3.1($viii$) is assumed such that the group of compliers which is the subpopulation of interest under fuzzy design is not of mass zero. Assumption 3.1(vi) and ($viii$) imply that the asymptotic limit of $\hat{\sigma}^2_{\mu,\epsilon}(\ell)$ defined in Section 2 is bounded away from zero for all $\ell \in \mathcal{L}$.

Assumption 3.2 *Assume that*

(i) $K(\cdot)$ *is a non-negative symmetric bounded kernel with a compact support in* R*, and* $\int K(u)du = 1$*;*

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

Assumption 3.2(i) is a standard assumption on the kernel function. The triangular kernel ($K(u) = (1 - |u|) \cdot 1(|u| \leq 1)$), which is the most frequently used kernel function in RD estimation and testing, satisfies the stated conditions. Assumption 3.2(ii) is the standard undersmoothing condition for local linear estimation. It helps to eliminate the nuisance bias term and obtain centered asymptotic distributions of the local linear estimators.

Let $\ddot{m}(\ell) = (m^{(1)}(Y, W, S, \ell), m^{(2)}(Y, W, S, \ell), q^{(1)}(T, W, S, \ell), q^{(2)}(T, W, S, \ell))'$ be a 4×1 random vector. Let $h^{+}_{2,P}(\ell_1, \ell_2) = \lim_{z \searrow c} Cov_P(\ddot{m}(\ell_1), \ddot{m}(\ell_2)|Z = z)$ and $h^{-}_{2,P}(\ell_1, \ell_2) = \lim_{z \nearrow c} Cov_P(\ddot{m}(\ell_1), \ddot{m}(\ell_2)|Z = z)$ be the left and right limit of its conditional variance-covariance matrix of $\ddot{m}(\ell)$ at $Z = c$. For $j = 0, 1, 2$, let $\vartheta_j = \int_0^\infty u^j K(u) du$. Define $C_k = \frac{\int_0^\infty (\vartheta_2 - u\vartheta_1)^2 K^2(u) du}{(\vartheta_2 \vartheta_0 - \vartheta_1^2)^2 \cdot f_z(c)}$. Let $h_{1,P}(\ell) = (-\varrho^{(2)}_P(\ell), \varrho^{(1)}_P(\ell), \rho^{(2)}_P(\ell), -\rho^{(1)}_P(\ell))$ be a 1×4 matrix. The following lemma states the asymptotic limit of the local linear estimator of $\hat{\nu}_n$ for a fixed underlying distribution P_n.

Lemma 3.1 *Under Assumptions 2.1, 2.2, 3.1, and 3.2, we have*

$$\sqrt{nh}(\hat{\mu}_n - \mu_P) \Rightarrow \Phi_{C_k h_{2,\mu,P}}$$

where $h_{2,\mu,P}(\ell_1, \ell_2) = h_{1,P}(\ell_1) \left(h^{+}_{2,P}(\ell_1, \ell_2) + h^{-}_{2,P}(\ell_1, \ell_2) \right) h_{1,P}(\ell_2)'$ *and* $\Phi_{C_k h_{2,\mu,P}}$ *is a mean zero Gaussian processes with covariance kernel* $C_k h_{2,\mu,P}$.

As is discussed in Section 2, the limiting process $C_k h_{2,\mu,P}$ could be approximated by the proposed multiplier bootstrap method. The following lemma summarizes the asymptotic property of the simulated process $\widehat{\Phi}^u_{\mu,n}$ for a fixed underlying distribution P_n.

Lemma 3.2 *Under Assumptions 3.1–3.2,* $\widehat{\Phi}^u_{\mu,n} \Rightarrow \Phi_{C_k h_{2,\mu,P}}$ *with probability approaching 1.*

Note that the weak convergence results in both Lemma 3.1 and 3.2 are for a fixed underlying distribution P, which is sufficient to show the pointwise size control and consistency results of the proposed test. However, to show that the test also has uniform size control and consistency properties over a broad set of underlying DGPs, we need the above discussed weak convergence results to hold for any P_{k_n}, with subsequence k_n of n. The corresponding results are stated and proven in Lemmas B.1 and B.3 in the appendix.

3.2 Uniform Size Control and Consistency of the Proposed Test

Define $\mathcal{H}_1 = \{h_{1,P}(\cdot) : P \in \mathcal{P}\}$, $\mathcal{H}_2^+ = \{h_{2,P}^+(\cdot,\cdot) : P \in \mathcal{P}\}$, and $\mathcal{H}_2^- = \{h_{2,P}^-(\cdot,\cdot) : P \in \mathcal{P}\}$. Let $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2^+ \times \mathcal{H}_2^-$. For any two functions $h = (h_1, h_2^+, h_2^-)$ and $\tilde{h} = (\tilde{h}_1, \tilde{h}_2^+, \tilde{h}_2^-)$ in the space of $\mathcal{H}$, define metric d as

$$d(h,\tilde{h}) = \max\left\{d_1(h_1,\tilde{h}_1), d_2(h_2^+,\tilde{h}_2^+), d_2(h_2^-,\tilde{h}_2^-)\right\},$$

where $d_1(h_1,\tilde{h}_1) = \sup_{\ell\in\mathcal{L}}\left\|h_1(\ell) - \tilde{h}_1(\ell)\right\|$, $d_2(h_2,\tilde{h}_2) = \sup_{\ell_1,\ell_2\in\mathcal{L}}\left\|h_2(\ell_1,\ell_2) - \tilde{h}_2(\ell_1,\ell_2)\right\|$ and $\|\cdot\|$ is the Euclidean norm.

Assumption 3.3 *Let $\mathcal{P}_0$ be the subset of $\mathcal{P}$ that satisfies Assumption 3.1 such that the null hypothesis in (2.1) holds under P if $P \in \mathcal{P}_0$.*

Define $\mathcal{L}^o(P) = \{\ell : \mu_P(\ell) = 0\}$ which is the collection of indices satisfying $\mu_P(\ell) = 0$ under P. Then we have the following results of the proposed monotonicity test.

Theorem 3.1 *Suppose that Assumptions 3.1-3.3 hold. Then, for every compact subset of $\mathcal{H}_{cpt}$ of $\mathcal{H}$, we have*

(a) $\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ h_P\in\mathcal{H}_{cpt}\}} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,LFC}_{n,FRD}(\alpha)) \leq \alpha$;

(b) $\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ h_P\in\mathcal{H}_{cpt}\}} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,GMS}_{n,FRD}(\alpha)) \leq \alpha$;

(c) if there exists $P_c^{LFC} \in \mathcal{P}_0$ such that $\mathcal{L}^o(P_c^{LFC}) = \mathcal{L}$ and $h_{2,\mu,P_c^{LFC}}$ is not a zero function, then

$$\lim_{\eta\to 0}\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ h_P\in\mathcal{H}_{cpt}\}} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,LFC}_{n,FRD}(\alpha)) = \alpha;$$

(d) if there exists $P_c \in \mathcal{P}_0$ such that $\mathcal{L}^o(P_c)$ is not empty and h_{2,μ,P_c} restricted to $\mathcal{L}^o(P_c)\times \mathcal{L}^o(P_c)$ is not a zero function, then

$$\lim_{\eta\to 0}\limsup_{n\to\infty} \sup_{\{P\in\mathcal{P}_0:\ h_P\in\mathcal{H}_{cpt}\}} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,GMS}_{n,FRD}(\alpha)) = \alpha.$$

Parts (a) and (b) of Theorem 3.1 show that our tests based on both LFC and GMS critical values have uniform asymptotic size control over a compact subset of covariance kernels which is similar to Theorem 2(a) of Andrews and Shi (2013). Theorem 3.1(c) shows that our test based on the LFC critical value is at most infinitesimally conservative asymptotically when there exists at least one P_c^{LFC} satisfying the LFC condition: $\mu_P(\ell) = 0$ for all $\ell \in \mathcal{L}$. Theorem 3.1(d) shows that our test based on the GMS critical value is at most infinitesimally conservative asymptotically when there exists at least one P_c such that $\mathcal{L}^o(P_c)$ is not empty and h_{2,μ,P_c} restricted to $\mathcal{L}^o(P_c) \times \mathcal{L}^o(P_c)$ is not a zero function.

Last, we show that the proposed tests based on both decision rules are consistent, as is summarized in the following theorem.

Theorem 3.2 *Suppose that Assumptions 3.1-2.4 hold and that under $P^* \in \mathcal{P}$, there exist $w_1^*, w_2^* \in \mathcal{W}$ with $w_2^* > w_1^*$ and $s^* \in \mathcal{S}$ such that $CLATE(w_2^*, s^*) < CLATE(w_1^*, s^*)$. Then*

(a) $\lim_{n\to\infty} P^*(\widehat{T}_{n,FRD} > \hat{c}_{n,FRD}^{\eta,LFC}(\alpha)) = 1$.

(b) $\lim_{n\to\infty} P^*(\widehat{T}_{n,FRD} > \hat{c}_{n,FRD}^{\eta,GMS}(\alpha)) = 1$.

4 Special Case: Sharp RD

When the treatment status is a deterministic function of the running variable such that $T = 1(Z \geq c)$, the RD model follows the sharp RD design. In this case, every individual is a complier, and the identification restrictions for the fuzzy RD case in Assumption 2.1 reduce to the following conditions.

Assumption 4.1 *For a running variable Z continuously distributed in $\mathcal{N}_{\delta,z}(c)$ for some $\delta > 0$, $E_P[Y(t)|X = x, Z = z]$ is continuous in x and z on $\mathcal{X}_c \times \mathcal{N}_{\delta,z}(c)$ for $t = 0, 1$.*

Under Assumption 4.1, the conditional average treatment effect (CATE) $E_P[Y(1) - Y(0)|X = x, Z = c]$ is identified as

$$CATE(x) = \lim_{z \searrow c} E_P[Y|X = x, Z = z] - \lim_{z \nearrow c} E_P[Y|X = x, Z = z].$$

Recall that $X = (W, S)$. The test of interest then examines whether the $CATE(x)$ is monotonically increasing in w for all values of s. Mathematically, the null and alternative hypotheses could be written as

$H_{0,SRD} : CATE(x)$ is non-decreasing in w on $\mathcal{W}$ for all $s \in \mathcal{S}$;

$H_{1,SRD} :\ H_{0,SRD}$ does not hold.

Under Assumption 4.1, it is easy to see that $CATE(x)$ is continuous in $x \in \mathcal{X}$. Similar to the discussion in Section 2.2, testing the null $H_{0,SRD}$ is equivalent to testing

$$H'_{0,SRD} : \nu_P(\ell) \equiv \rho_P^{(2)}(\ell) p_P^{(1)}(\ell) - \rho_P^{(1)}(\ell) p_P^{(2)}(\ell) \leq 0, \quad \text{for all } \ell \in \mathcal{L},$$

where $p_P^{(\kappa)} = E_P[g_{w_\kappa,\ell}(W) g_{s,\ell}(S) | Z = c]$ for $\kappa = 1, 2$, and $\mathcal{L}$ is defined in Lemma 2.2. First, we note that $p_P^{(\kappa)}$ is a special case of $\varrho_P^{(\kappa)}$ with $T = 1(Z \geq c)$. So the monotonicity test for $H'_{0,SRD}$ could be carried out using the same test statistic and decision rules for testing $H'_{0,FRD}$.

More efficiently, one could estimate $p_P^{(\kappa)}(\ell)$, $\kappa = 1, 2$. by nonparametrically regressing $p^{(\kappa)}(W_i, S_i, \ell) = g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i)$ on Z_i using data from both sides of the cut-off c. Let $\hat{p}_n^{(\kappa)}(\ell)$ be the constant term solving of the following minimization problem:

$$\left(\hat{p}_n^{(\kappa)}(\ell), \hat{b}_{n,p}^{(\kappa)}(\ell)\right) = \arg\min_{a,b} \sum_{i=1}^{n} K\left(\frac{Z_i - c}{h}\right) \left[g_{w_\kappa,\ell}(W_i) g_{s,\ell}(S_i) - a - b(Z_i - c)\right]^2.$$

For $l = 0, 1, 2$, define $S_{n,l} = \sum_{i=1}^n K(\frac{Z_i - c}{h})(Z_i - c)^l$ and $\mathsf{w}_{ni} = \frac{K(\frac{Z_i - c}{h})[S_{n,2} - S_{n,1}(Z_i - c)]}{S_{0,2} S_{n,2} - S_{n,1}^2}$. It then follows straightforwardly that $\hat{p}_n^{(\kappa)}(\ell) = \sum_{i=1}^n \mathsf{w}_{ni} \cdot p^{(\kappa)}(W_i, S_i, \ell)$ for both $\kappa = 1, 2$.

Let $\hat{\nu}_n(\ell) = \hat{\rho}_n^{(1)}(\ell) \hat{p}_n^{(2)}(\ell) - \hat{\rho}_n^{(2)}(\ell) \hat{p}_n^{(1)}(\ell)$ be the estimator of $\nu_P(\ell)$. Then the influence function representation of $\hat{\nu}_n(\ell)$ could be formulated as

$$\sqrt{nh}(\hat{\nu}_n(\ell) - \nu_P(\ell)) = \sum_{i=1}^{n} \phi_{\nu,ni}(\ell) + o_p(1),$$

where $\phi_{\nu,ni}(\ell)$ is defined similar to $\phi_{\mu,ni}(\ell)$ in Section 2 except that $\varrho_P^{(\kappa)}(\ell)$ and $\phi_{\varrho,ni}^{(\kappa)}(\ell)$ in $\phi_{\mu,ni}(\ell)$ are replaced by $p_P^{(\kappa)}(\ell)$ and $\phi_{p,ni}^{(\kappa)}(\ell) = w_{ni}(p^{(\kappa)}(W_i, S_i, \ell) - p_P^{(\kappa)}(\ell))$ for $\ell \in \mathcal{L}$ and $\kappa = 1, 2$.

Let $\hat{\phi}_{\nu,ni}(\ell)$ be the estimated influence function and $\hat{\sigma}_{\nu,n}^2(\ell) = \sum_{i=1}^n \hat{\phi}_{\nu,ni}^2(\ell)$ be the variance estimator of $\hat{\nu}_n(\ell)$. Let $\hat{\sigma}_{\nu,\epsilon}^2(\ell) = \max\left\{\hat{\sigma}_{\nu,n}^2(\ell), \epsilon \cdot \hat{\sigma}_{\nu,n}^2(\ell_0)\right\}$. Then we can define the test statistic for the sharp RD case as

$$\widehat{T}_{n,SRD} = \sup_{\ell \in \mathcal{L}} \sqrt{nh} \frac{\hat{\nu}_n(\ell)}{\hat{\sigma}_{\nu,\epsilon}(\ell)}.$$

The LFC and GMS simulated critical values are defined as

$$\hat{c}_{n,SRD}^{\eta,LFC}(\alpha) = \sup\left\{q \middle| P^u\Big(\sup_{\ell \in \mathcal{L}} \frac{\widehat{\Phi}_{\nu,n}^u(\ell)}{\hat{\sigma}_{\nu,\epsilon}(\ell)} \leq q\Big) \leq 1 - \alpha + \eta\right\} + \eta, \text{ and}$$

$$\hat{c}_{n,SRD}^{\eta,GMS}(\alpha) = \sup\left\{q \middle| P^u\left(\sup_{\ell \in \mathcal{L}} \left(\frac{\widehat{\Phi}_{\nu,n}^u(\ell)}{\hat{\sigma}_{\nu,\epsilon}(\ell)} + \hat{\psi}_\nu(\ell)\right) \leq q\right) \leq 1 - \alpha + \eta\right\} + \eta$$

respectively, with $\widehat{\Phi}^u_{\nu,n}(\ell) = \sum_{i=1}^n U_i \cdot \hat{\phi}_{\nu,ni}(\ell)$ being the process simulating the limiting process of $\sqrt{nh}(\hat{\nu}_n(\ell) - \nu_P(\ell))$ and $\hat{\psi}_\nu(\ell) = -B_n \cdot 1\left(\sqrt{nh}\frac{\hat{\nu}_n(\ell)}{\hat{\sigma}_{\nu,\epsilon}(\ell)} < -a_n\right)$. If we reject the null hypothesis $H_{0,SRD}$ when $\widehat{T}_{n,SRD} > \hat{c}^{\eta,LFC}_{n,SRD}(\alpha)$ or when $\widehat{T}_{n,SRD} > \hat{c}^{\eta,GMS}_{n,SRD}(\alpha)$, the resulting tests are consistent and have uniform size control in the limit. The asymptotic properties are similar to those given in Section 3 for the fuzzy RD case and we omit the details for brevity.

5 Simulations

In this section, we investigate the small sample performance of the proposed tests. For all data generating processes (DGPs), the running variable Z, the additional control X, and the error term u in the outcome equation are generated as following

$$Z \sim 2Beta(2,2) - 1;\ X \sim U[0,1];\ u \sim N(0,1).$$

The outcome Y and the treatment decision T are DGP specific. All DGPs are either estimated from the empirical example or modified from the data-driven DGPs to demonstrate specific properties of the proposed tests.

We use DGPs 1-3 to illustrate the small sample performance of the proposed monotonicity tests under the sharp RD design. DGPs 1-2 are estimated from the empirical dataset. DGP 3 is altered from DGP 2 to have a U-shaped CLATE to demonstrate the potential power gain of the GMS method. The DGPs are plotted in Figure 1 with detailed DGP generating procedures described in the footnote.

DGP 1: Sharp RD, Homogeneous Zero Effect

$$Y = \begin{cases} -0.373 + 0.545Z - 0.056Z^2 + 0.1u & \text{if } Z \geq 0 \\ -0.531 + 0.556Z + 0.192Z^2 + 0.1u & \text{if } Z > 0 \end{cases}$$

DGP 2: Sharp RD, Monotonically Increasing Treatment Effect

$$Y = \begin{cases} -0.755 - 0.254W + 0.742Z - 0.219WZ - 0.063Z^2 + 1.175W^2 + 0.1u & \text{if } Z \geq 0 \\ -0.607 - 0.220W + 0.386Z + 0.228WZ + 0.204Z^2 + 0.469W^2 + 0.1u & \text{if } Z < 0 \end{cases}.$$

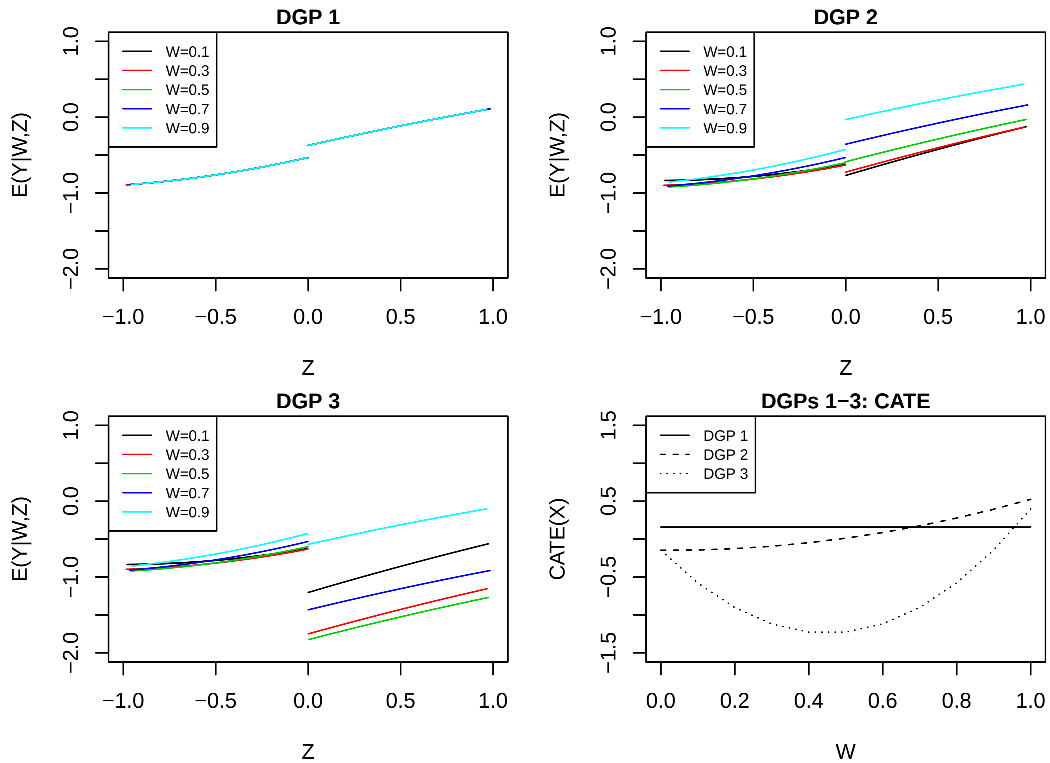
DGP 3: Sharp RD, Inverse-U Shape Treatment Effect

$$Y = \begin{cases} -0.755 - \mathbf{5.080}W + 0.742Z - 0.219WZ - 0.063Z^2 + \mathbf{5.875}W^2 + 0.1u & \text{if } Z \geq 0 \\ -0.607 - 0.220W + 0.386Z + 0.228WZ + 0.204Z^2 + 0.469W^2 + 0.1u & \text{if } Z < 0 \end{cases}.$$

The null hypotheses of interest are "H_0 : $CATE(w)$ is non-decreasing in w for all $w \in [0,1]$" and "H_0 : $CATE(w)$ is non-increasing in w for all $w \in [0,1]$". For each DGP, three different sample sizes $n = 2,000$, $n = 4,000$ and $n = 8,000$ are used. For each DGP and sample size combination, 2,000 simulation samples are drawn and the proposed monotonicity test is conducted in each Monte Carlo simulation. In each test, the bootstrap critical value is calculated using 1,000 bootstrap simulations. All tests carried out in this section use the triangular kernel and bandwidths selected according to the formula $h_{CCT} \times n^{1/5-1/k}$, where h_{CCT} is the robust bandwidth following Calonico et al. (2014) (CCT) and k is the under-smoothing parameter. In all simulation tables, we report results with $k = 4.25, 4.5$ and 4.75. The cubes defined in equation (2.4) have side-lengths $1/q$ for $q = 1, ..., Q$. We use benchmark $Q = 10$ which includes 165 combinations of overlapping $C_{w_1,10}$ and $C_{w_2,10}$ intervals. We also report robustness checks with $Q = 15$, which includes 560 combinations of overlapping $C_{w_1,15}$ and $C_{w_2,15}$ intervals. When $n = 2,000$ the bandwidth of DGPs 1-3 is around 0.18-0.25, which means that the expected effective sample size of the smallest $C_{w_1,15}$ and $C_{w_2,15}$ intervals ranges from 27 to 36 when $Q = 10$ and 18 to 24 when $Q = 15$ (for each local linear regression on one side of the RD cut-off).

Table 1 summarizes the rejection proportions of the proposed tests for the sharp RD case, with $p_P^{(\kappa)}(\ell)$ estimated using the full sample method as is discussed in the second half of Section 4. We see from the table that, regardless of whether the simulated critical value uses the LFC or the GMS method, the proposed monotonicity test controls size well in small samples and have power going to one as the sample size increases. The GMS method brings the size of the proposed test closer to the nominal level (5%) when the null of monotonic effect holds but not with equality (columns (1)-(3) and (7)-(9) in DGP 2) and increases the power when the null of monotonic effect is violated (DGP 3 and columns (4)-(6) and (10)-(12) in DGP 2). The power gain is especially large for DGP 3. This is because, with the inverse U-Shape model in DGP3, the GMS method can get rid of the influence of about half of the hypercubes on critical value calculation when testing either direction of monotonicity. Last but not least, we notice that the simulation performance of the proposed test is not very sensitive to either the choice of bandwidth

Figure 1: DGP 1-3



Note: The DGPs are estimated from the data of the empirical section. To obtain the models, we first rescale the running variable (i.e., transition score) in the data set to $[-1, 1]$ to match the support of the generated X variable. Then for the outcome equation in DGP 1, we regress the outcome (i.e., score in Baccalaureate exam) on the running variable and its second order polynomial term separately for the subsample to the left and the right of the cutoff value (i.e., 0). To get the outcome equation in DGP 2, we add the additional control of interest (i.e., average peer admission score), its second order polynomial term, and its interaction with the running variable to the set of regressors. To get the outcome equation in DGP 3 (bottom left graph), we take the model for DGP 2 and multiply the slope coefficient of the additional control variable by 20 and the slope coefficient of its second order polynomial term by 5.

Table 1: Small Sample Performance of Proposed Monotonicity Tests, Sharp RD

	LFC Critical Value						GMS Critical Value					
	H_0: Non-decreasing			H_0: Non-increasing			H_0: Non-decreasing			H_0: Non-increasing		
	c=4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						$Q = 10$						
	DGP 1: Homogeneous Zero Effect											
n=2000	0.019	0.020	0.022	0.017	0.020	0.020	0.024	0.026	0.026	0.019	0.023	0.022
n=4000	0.025	0.026	0.032	0.023	0.025	0.030	0.028	0.030	0.036	0.027	0.030	0.035
n=8000	0.042	0.042	0.043	0.039	0.039	0.041	0.048	0.045	0.048	0.044	0.044	0.044
	DGP 2: Monotonically Increasing Treatment Effect											
n=2000	0.005	0.005	0.006	0.329	0.373	0.418	0.009	0.008	0.009	0.334	0.378	0.429
n=4000	0.008	0.007	0.007	0.721	0.783	0.838	0.010	0.012	0.011	0.725	0.785	0.840
n=8000	0.006	0.005	0.005	0.970	0.981	0.990	0.011	0.012	0.013	0.970	0.981	0.990
	DGP 3: Inverse U-Shaped Treatment Effect											
n=2000	0.049	0.058	0.066	0.337	0.401	0.472	0.069	0.075	0.086	0.364	0.425	0.495
n=4000	0.180	0.205	0.236	0.833	0.895	0.927	0.214	0.244	0.277	0.844	0.900	0.929
n=8000	0.428	0.490	0.546	0.995	0.998	0.999	0.498	0.562	0.611	0.995	0.999	0.999
						$Q = 15$						
	DGP 1: Homogeneous Zero Effect											
n=2000	0.006	0.008	0.006	0.006	0.007	0.008	0.006	0.009	0.008	0.008	0.008	0.008
n=4000	0.010	0.012	0.013	0.010	0.012	0.014	0.012	0.014	0.014	0.010	0.012	0.015
n=8000	0.027	0.033	0.030	0.030	0.027	0.031	0.030	0.038	0.033	0.032	0.031	0.034
	DGP 2: Monotonically Increasing Treatment Effect											
n=2000	0.002	0.002	0.001	0.180	0.220	0.260	0.004	0.003	0.002	0.189	0.226	0.268
n=4000	0.004	0.002	0.004	0.556	0.636	0.710	0.006	0.006	0.006	0.564	0.640	0.712
n=8000	0.004	0.004	0.005	0.933	0.956	0.976	0.007	0.008	0.008	0.933	0.956	0.976
	DGP 3: Inverse U-Shaped Treatment Effect											
n=2000	0.017	0.022	0.030	0.170	0.219	0.272	0.023	0.029	0.036	0.184	0.234	0.296
n=4000	0.090	0.110	0.132	0.701	0.772	0.841	0.106	0.135	0.161	0.721	0.788	0.852
n=8000	0.320	0.390	0.453	0.984	0.992	0.997	0.374	0.449	0.500	0.986	0.992	0.997

Note: Reported are rejection proportions among 2,000 simulations, where all tests are carried out using the 5% significance level. For each test, the simulated critical value is calucalted with 1,000 bootstrap repetitions.

choice or that of Q.

DGPs 4-6 illustrate the small sample performance of the proposed monotonicity tests under the fuzzy RD design. The outcome equations in these DGPs are the same as those in DGPs 1-3, respectively. The selection equation is modeled by

$$T = \begin{cases} 1(0.331 + 0.277Z + 0.049Z^2 + u > 0) & \text{if } Z \geq 0 \\ 0 & \text{if } Z > 0 \end{cases},$$

which is estimated from the data using probit regression. Table 2 summarizes the rejection proportions of the proposed tests. We observe similar small sample performances as those reported in Table 1, although the tests generally have lower power under the fuzzy RD design due to the extra noise in the first stage.

6 Empirical Example: The Effect of Going to a Better High School

As is discussed in Pop-Eleches and Urquiola (2013), in Romania, a typical elementary school student takes a nationwide test in the last year of elementary school and applies to a list of high schools (and tracks). The admission decision is entirely dependent on the student's transition score, an average of the student's performance on the national test and grade point average, as well as a student's preference for schools. A student is eligible to a high school if his/her transition score passes the school's cut-off.

Using an administrative dataset, Pop-Eleches and Urquiola (2013) find that attending a better school on average improves a student's performance on the Baccalaureate exam but does not significantly affect his/her probability of taking the exam. In this section, we apply the proposed monotonicity test to study how the effects of attending a more selective high school interact with peer quality of the school. As in Shen and Zhang (2015) and Hsu and Shen (2016), we focus on two-school towns because score cutoffs within a town are often quite close to each other and having more than one discontinuity point within the estimation window can introduce severe estimation bias.

In this application, the running variable (Z) is a student's standardized transition score subtracting individual school cut-off and the cut-off value (c) for having an offer from the more selective high school is zero for all students. The treatment variable (T) indicates whether a student attends the more selective high school in town. The outcome variable (Y) is a student's decision of whether to take the Baccalaureate exam

Table 2: Small Sample Performance of Proposed Monotonicity Tests, Fuzzy RD

	LFC Critical Value						GMS Critical Value					
	H_0: Non-decreasing			H_0: Non-increasing			H_0: Non-decreasing			H_0: Non-increasing		
	c=4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	$Q = 10$											
	DGP 4: Homogeneous Zero Effect											
n=2000	0.001	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.001	0.001
n=4000	0.002	0.002	0.002	0.004	0.005	0.005	0.003	0.003	0.004	0.004	0.005	0.006
n=8000	0.010	0.011	0.014	0.009	0.010	0.012	0.014	0.012	0.016	0.010	0.014	0.014
	DGP 5: Monotonically Increasing Treatment Effect											
n=2000	0.000	0.000	0.000	0.117	0.138	0.161	0.000	0.001	0.000	0.118	0.139	0.165
n=4000	0.000	0.000	0.001	0.404	0.468	0.546	0.001	0.001	0.001	0.407	0.472	0.550
n=8000	0.001	0.001	0.002	0.840	0.894	0.932	0.002	0.002	0.002	0.840	0.895	0.932
	DGP 6: Inverse U-Shaped Treatment Effect											
n=2000	0.021	0.024	0.030	0.127	0.170	0.214	0.028	0.032	0.038	0.139	0.185	0.234
n=4000	0.112	0.144	0.186	0.645	0.740	0.809	0.154	0.198	0.246	0.666	0.760	0.827
n=8000	0.479	0.568	0.658	0.994	0.998	1.000	0.574	0.664	0.742	0.994	1.000	1.000
	$Q = 15$											
	DGP 4: Homogeneous Zero Effect											
n=2000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
n=4000	0.001	0.000	0.000	0.001	0.002	0.002	0.0005	0.000	0.000	0.001	0.002	0.002
n=8000	0.008	0.009	0.011	0.008	0.009	0.009	0.011	0.012	0.014	0.010	0.013	0.013
	DGP 5: Monotonically Increasing Treatment Effect											
n=2000	0.000	0.000	0.000	0.056	0.072	0.088	0.000	0.000	0.000	0.058	0.076	0.090
n=4000	0.000	0.000	0.000	0.236	0.301	0.358	0.000	0.000	0.000	0.239	0.304	0.361
n=8000	0.001	0.001	0.001	0.635	0.689	0.762	0.002	0.001	0.002	0.640	0.699	0.768
	DGP 6: Inverse U-Shaped Treatment Effect											
n=2000	0.008	0.009	0.014	0.060	0.078	0.103	0.010	0.013	0.017	0.064	0.084	0.114
n=4000	0.046	0.066	0.086	0.448	0.552	0.646	0.066	0.085	0.112	0.466	0.574	0.666
n=8000	0.319	0.380	0.428	0.747	0.789	0.820	0.354	0.424	0.461	0.754	0.809	0.835

Note: Reported are rejection proportions among 2,000 simulations, where all tests are carried out using the 5% significance level. For each test, the simulated critical value is calucalted with 1,000 bootstrap repetitions.

(exam take) and his/her Baccalaureate exam grade (exam grade) conditional on taking the exam. The additional control variable (X) is the leave-one-out average transition score in the more selective high school in town, which is used to proxy the peer quality of the school. As is discussed in the simulation section, the proposed monotonicity test is carried out using the triangular kernel, the undersmoothed CCT bandwidth and the cubes defined in equation (2.4) with $Q = 50$ and 75. When $Q = 75$, for example, there is a total of 70,300 combinations of overlapping $C_{w_1,75}$ and $C_{w_2,75}$ intervals. The effective sample size of the smallest $C_{w_1,75}$ and $C_{w_2,75}$ intervals ranges from 50 to 62 for each local linear regression on one side of the RD cut-off.[2]

Table 3 reports the results of the monotonicity tests. First, we discuss the effect of attending a more selective high school on a student's probability of taking the Baccalaureate exam. As we see from the table, regardless of the choice of the under-smoothing parameter, the number of cubes, or whether to use the LFC or GMS critical value, we fail to reject the null of monotonically non-decreasing effect with very high p-values and reject the null of monotonically non-increasing effect with p-values ranging between 1-3%. These results suggest that the effect of attending a better high school on a marginal individual's probability of taking the Baccalaureate exam increases monotonically with peer quality. The finding indicates that the insignificant mean effect found in Pop-Eleches and Urquiola (2013) comes from the cancelation of positive and negative treatment effects among different schools. The result is also in line with the results of uniform sign tests conducted in Hsu and Shen (2016), which find that the effect on the Baccalaureate exam taking rate is positive for some subpopulation of schools and negative for the others.

In contrast, the effect on the exam grade outcome is likely homogeneous among schools with different peer quality as we fail to reject both nulls of monotonically non-decreasing and monotonically non-increasing effects at the 10% significance level. Note that the p-values associated with the GMS method are always the same or smaller than those

[2]Notice that in the simulation section the effective sample size of smallest cubes is kept around 20. For the empirical sample, due to its immense sample size, this rule-of-thumb will result in a Q value as large as around 200 and a total of 1,333,300 combinations of overlapping $C_{w_1,200}$ and $C_{w_2,200}$ intervals, which is computationally infeasible. Instead, we report results with both $Q = 50$ and 75. We find our empirical results insensitive to the Q choice.

Table 3: P-values of Monotonicity Tests

	LFC			GMS		
	c=4.5	c=4.75	c=5	c=4.5	c=4.75	c=5
			$Q = 50$			
H_0: the effect is non-decreasing						
First Stage	0.001	0.000	0.000	0.001	0.000	0.000
Exam Take	0.801	0.794	0.747	0.776	0.774	0.723
Exam Grade	0.360	0.335	0.353	0.347	0.322	0.336
H_0: the effect is non-increasing						
First Stage	0.032	0.024	0.016	0.025	0.019	0.012
Exam Take	0.027	0.016	0.009	0.027	0.016	0.009
Exam Grade	0.196	0.144	0.140	0.196	0.144	0.139
			$Q = 75$			
H_0: the effect is non-decreasing						
First Stage	0.001	0.000	0.000	0.001	0.000	0.000
Exam Take	0.807	0.802	0.757	0.786	0.787	0.734
Exam Grade	0.374	0.349	0.345	0.362	0.333	0.326
H_0: the effect is non-increasing						
First Stage	0.032	0.024	0.016	0.025	0.019	0.012
Exam Take	0.027	0.016	0.009	0.027	0.016	0.009
Exam Grade	0.219	0.163	0.156	0.215	0.159	0.152

Notes: Data are from Pop-Eleches and Urquiola (2013). Nonparametric local linear estimations are conducted using the triangular kernel, the undersmoothed CCT bandwidth defined in the simulation section, and the cubes defined in equation (2.4). All simulated critical values are calculated with 1,000 bootstrap repetitions.

associated with the LFC method. This is because the GMS method potentially improves the power of the proposed monotonicity test, as is discussed in the theoretical section of the paper.

In Table 3, we also report the results of the monotonicity tests for the first stage enrollment decision. The tests are carried out using the sharp RD test discussed in Section 4. We reject both nulls of monotonically non-decreasing and monotonically non-increasing first stage effects with very small p-values. The testing results suggest that the first-stage enrollment decision is strongly heterogeneous, but its relationship with peer quality in the more selective high school is not simple and not monotonic.

APPENDIX

A Proofs of Lemmas in Section 2

This section proves the Lemmas in Section 2 that illustrate the equivalence between the original null hypothesis in (2.1) to the transformed null hypothesis in (2.3).

Proof of Lemma 2.1:

For notational simplicity and without loss of generality, we prove the lemma for the case with $d_w = d_s = 1$. Proof for the higher dimension case follows the same idea but requires more complicated notations.

First, we prove that (i) implies (ii). For any $w_1 = w_2$, the inequality in (ii) holds trivially with equality. For any $w_1 > w_2$ and any $q = 1, 2, 3, \ldots$, such that $(q+1) \cdot w_1, (q+1) \cdot w_2 \in \{0, 1, 2, \cdots, q\}$, we have $w_1 \geq w_2 + 1/(q+1)$. By (i), this implies that $\lambda(w, s) \geq \lambda(w', s)$ for all $w \in [w_1, w_1 + 1/(q+1)]$ and $w' \in [w_2, w_2 + 1/(q+1)]$ and $s \in \mathcal{S}$. Therefore, the weighted average of $\lambda(w, s)$ over $w \in [w_1, w_1 + /(q+1)]$ and $s \in C_{s,q}$ has to be greater or equal to that of $\lambda(w', s)$ over $w' \in [w_2, w_2 + 1/(q+1)]$ and $s \in C_{s,q}$. Equivalently,

$$\frac{\int_{C_{w_1,q} \times C_{s,q}} \lambda(\tilde{w}, \tilde{s}) \cdot h(\tilde{w}, \tilde{s}) d\tilde{w} d\tilde{s}}{\int_{C_{w_1,q} \times C_{s,q}} h(\tilde{w}, \tilde{s}) d\tilde{w} d\tilde{s}} \geq \frac{\int_{C_{w_2,q} \times C_{s,q}} \lambda(\tilde{w}, \tilde{s}) \cdot h(\tilde{w}, \tilde{s}) d\tilde{w} d\tilde{s}}{\int_{C_{w_2,q} \times C_{s,q}} h(\tilde{w}, \tilde{s}) d\tilde{w} d\tilde{s}}.$$

We prove the inequality in (ii).

Second, we prove that (ii) implies (i) by contradiction. Suppose that $\lambda(w, s') < \lambda(w', s')$ for some s' and $w \geq w'$. By continuity of $\lambda(w, s)$, there exist $w_u > w_l > w'_u > w'_l$ and $s_u > s_l$ so that $\lambda(w, s') < \lambda(w', s')$ for all $w \in [w_l, w_u]$, $w' \in [w'_l, w'_u]$ and $s' \in [s_l, s_u]$. In turn, we can find a q large enough such that for some w_1, w_2, and s, $(q+1) \cdot w_1, (q+1) \cdot w_2 \in \{0, 1, 2, \cdots, q\}$, and $q \cdot s \in \{0, 1, 2, \cdots, q-1\}$ so that $[w_1, w_1 + 1/(q+1)] \subseteq [w_l, w_u]$, $[w_2, w_2 + 1/(q+1)] \subseteq [w'_l, w'_u]$ and $[s, s + 1/q] \subseteq [s_l, s_u]$. Then the weighted average of $\lambda(w, s)$ over $w \in [w_1, w_1 + 1/(q+1)]$ and $s \in [s, s + 1/q]$ has to be strictly less than that of $\lambda(w', s)$ over $w' \in [w_2, w_2 + 1/(q+1)]$ and $s \in [s, s + 1/q]$.

That is,

$$\frac{\int_{C_{w_1,q}\times C_{s,q}} \lambda(\tilde{w},\tilde{s})\cdot h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}{\int_{C_{w_1,q}\times C_{s,q}} h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}} < \frac{\int_{C_{w_1,q}\times C_{s,q}} \lambda(\tilde{w},\tilde{s})\cdot h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}{\int_{C_{w_1,q}\times C_{s,q}} h(\tilde{w},\tilde{s})d\tilde{w}d\tilde{s}}.$$

This completes our proof. $\square$

Proof of Lemma 2.2:

Since $CLATE(w,s)$ is continuous on both w and s, we can set $\lambda(w,s)$ to $CLATE(w,s)$ and apply the results of Lemma 2.1. Let $\lambda(w,s) = CLATE(w,s) = E_P[(Y(1)-Y(0))(T(1)-T(0))|W=w,S=s,Z=c]/E_P[T(1)-T(0)|W=w,S=s,Z=c]$ and $h(w,s) = E_P[T(1)-T(0)|W=w,S=s,Z=c]\cdot f_{W,S|Z=c}(w,s)$. Then, for both $\kappa = 1,2$,

$$\begin{aligned}
&\int_{C_{w_\kappa,q}\times C_{s,q}} \lambda(w,s)\cdot h(w,s)dwds\\
=&E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[(Y(1)-Y(0))(T(1)-T(0))|W=w,S=s,Z=c]|Z=c]\\
=&E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(1)+Y(0)(1-T(1))|W=w,S=s,Z=c]|Z=c]\\
&-E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(0)+Y(0)(1-T(0))|W=w,S=s,Z=c]|Z=c]\\
=&\lim_{z\searrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(1)+Y(0)(1-T(1))|W=w,S=s,Z=c]|Z=z]\\
&-\lim_{z\nearrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(0)+Y(0)(1-T(0))|W=w,S=s,Z=c]|Z=z]\\
=&\lim_{z\searrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y|W=w,S=s,Z=z]|Z=z]\\
&-\lim_{z\nearrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y|W=w,S=s,Z=z]|Z=z]\\
=&\lim_{z\searrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)Y|Z=z]-\lim_{z\nearrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)Y|Z=z],\\
\equiv&\rho_P^{(\kappa)}(\ell).
\end{aligned}$$

The third equality holds from Assumptions 2.1 and 2.2. Specifically, Assumption 2.2 implies that for all $\epsilon>0$, there exists $\delta>0$ such that $\sup_{w,s}|f_{W,S|Z}(w,s|z)-f_{W,S|Z}(w,s|c)|<\epsilon$ for all $|z-c|<\delta$. Then given that $g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(1)+Y(0)(1-T(1))|W=w,S=s,Z=c]$ is uniformly bounded (due to Assumption 2.1 and the fact that a continuous function on a compact set is uniformly bounded), we know that there exists some $C>0$, such that $|E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(1)+Y(0)(1-T(1))|W=$

$w, S = s, Z = c]|Z = z] - E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(1) + Y(0)(1 - T(1))|W = w, S = s, Z = c]|Z = c]| \leq C \cdot \epsilon$ for all $|z - c| < \delta$. A similar result also hold for $E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)E_P[Y(1)T(0) + Y(0)(1 - T(0))|W = w, S = s, Z = c]|Z = z]$. Therefore, the limiting result in the third equality holds. The fourth equality holds by the fuzzy RD set-up and the continuity conditions in Assumption 2.1. The fifth equality holds by the law of iterated expectation.

Similarly, we can show that

$$\begin{aligned}
&\int_{C_{w_\kappa,q}\times C_{s,q}} h(w,s)dwds \\
=&\lim_{z\searrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)T|Z = z] - \lim_{z\nearrow c} E_P[g_{w_\kappa,\ell}(W)g_{s,\ell}(S)T|Z = z] \\
\equiv&\varrho_P^{(\kappa)}(\ell).
\end{aligned}$$

Applying the results in Lemma 2.1, we know that testing the null hypothesis $H_{0,FRD}$ is equivalent to testing

$$\rho_P^{(2)}(\ell)\varrho_P^{(1)}(\ell) - \rho_P^{(1)}(\ell)\varrho_P^{(2)}(\ell) \leq 0$$

for all $q = 1, 2, 3, \cdots$, and $w_1 \geq w_2$ such that $(q + 1) \cdot (w_1, w_2) \in \{0, 1, 2, \cdots, q\}^{2d_w}$, and $q \cdot s \in \{0, 1, 2, \cdots, q - 1\}^{d_s}$.

B Proofs of Theorems in Section 3

To prove the theorems for the asymptotic properties of the proposed tests, we first give auxiliary lemmas that will be used in the proof of the main results. The auxiliary lemmas also include results in Lemma 3.1 and 3.2 as special cases. Then we give the proofs of the main theorems in Section 3.

B.1 Auxiliary lemmas

Let E_Z denote the expectation conditional on sample path $\{Z_1, Z_2, \ldots\}$.

Lemma B.1 *Let $\{P_n : n \geq 1\}$ be a sequence of distributions. $P_n \in \mathcal{P}$ for all n so each of the distribution satisfies Assumption 3.1. Suppose 3.2 also holds. Then for any* *subsequence* $\{P_{k_n}\}$ *of* $\{P_n\}$ *such that* $\lim_{n\to\infty} d_2(h^+_{2,P_{k_n}}, h^{*+}_2) = 0$ *for some* $h^{*+}_2 \in \mathcal{H}^+_2$ *and* $\lim_{n\to\infty} d_2(h^-_{2,P_{k_n}}, h^{*-}_2) = 0$ *for some* $h^{*-}_2 \in \mathcal{H}^-_2$*, we have*

$$\sqrt{k_n h}\begin{pmatrix} \hat{m}^{(1)}_{k_n,+}(\cdot) - m^{(1)}_{P_{k_n},+}(\cdot) \\ \hat{m}^{(2)}_{k_n,+}(\cdot) - m^{(2)}_{P_{k_n},+}(\cdot) \\ \hat{q}^{(1)}_{k_n,+}(\cdot) - q^{(1)}_{P_{k_n},+}(\cdot) \\ \hat{q}^{(2)}_{k_n,+}(\cdot) - q^{(2)}_{P_{k_n},+}(\cdot) \end{pmatrix} \Rightarrow \Phi_{C_k h^{*+}_2}(\cdot) \ ,$$

$$\sqrt{k_n h}\begin{pmatrix} \hat{m}^{(1)}_{k_n,-}(\cdot) - m^{(1)}_{P_{k_n},-}(\cdot) \\ \hat{m}^{(2)}_{k_n,-}(\cdot) - m^{(2)}_{P_{k_n},-}(\cdot) \\ \hat{q}^{(1)}_{k_n,-}(\cdot) - q^{(1)}_{P_{k_n},-}(\cdot) \\ \hat{q}^{(2)}_{k_n,-}(\cdot) - q^{(2)}_{P_{k_n},-}(\cdot) \end{pmatrix} \Rightarrow \Phi_{C_k h^{*-}_2}(\cdot) \ ,$$

where $\Phi_{C_k h^{*+}_2}$ *and* $\Phi_{C_k h^{*-}_2}$ *are independent mean zero Gaussian processes with covariance kernel* $C_k h^{*+}_2$ *and* $C_k h^{*-}_2$*.*

Proof of Lemma B.1:

Here we prove the first weak convergence result. The second one follows by essentially the same proof but with the $+$ notation replaced by $-$. The resulting two limiting Gaussian processes are independent because the local linear estimators involved in the two convergence results use different subsamples and the data are independent.

By the Cramér-Wold Theorem, it is sufficient to show the $m^{(1)}_{P_{k_n},+}(\ell)$ case. Note that

$$\begin{aligned} &\sqrt{k_n h}(\hat{m}^{(1)}_{k_n,+}(\ell) - m^{(1)}_{P_{k_n},+}(\ell)) \\ =&\sum_{i=1}^{k_n} \sqrt{k_n h}(\mathsf{w}^+_{k_n i}(m^{(1)}(Y_i, W_i, S_i, \ell) - m^{(1)}_{P_{k_n},+}(\ell))) \\ =&\sum_{i=1}^{k_n} \sqrt{k_n h}(\mathsf{w}^+_{k_n i}(m^{(1)}(Y_i, W_i, S_i, \ell) - E_Z[m^{(1)}(Y_i, W_i, S_i, \ell)])) \\ &+ \sum_{i=1}^{k_n} \sqrt{k_n h}(\mathsf{w}^+_{k_n i}(E_Z[m^{(1)}(Y_i, W_i, S_i, \ell)] - m^{(1)}_{P_{k_n},+}(\ell))). \end{aligned}$$

We first consider the second term which is the bias term. By Theorem 4 of Fan and Gijbels (1992), we know that

$$\sum_{i=1}^{k_n} \sqrt{k_n h}(\mathsf{w}^+_{k_n i}(E_Z[m^{(1)}(Y_i, W_i, S_i, \ell)] - m^{(1)}_{P_{k_n},+}(\ell))) = O_p(\sqrt{k_n h^5}) = o_p(1).$$

Note that the first equality holds because the magnitude of the bias is proportional to the second derivative of $m^{(1)}_{P_{k_n},+}(\ell)$ with respect to z. By Assumption 3.1, we know that for all P_{k_n}, $\partial^2\zeta_{P_{k_n},+}(x,z)/\partial z\partial z$ is uniformly bounded on $x \in \mathcal{X}_c$ and $z \in \mathcal{N}^+_{\delta,z}(c)$. Since $m^{(1)}_{P_{k_n},+}(\ell) = \lim_{z\searrow c} E_{P_{k_n}}[g_{w_1,\ell}(W)g_{s,\ell}(S)Y|Z=z]$, $m^{(1)}_{P_{k_n},+}(\ell)$ is uniformly bounded as well. Then given the additional assumption that $k_n h^5 \to 0$, we know that the above $o_p(1)$ result holds uniformly over $\ell \in \mathcal{L}$.

Therefore, uniformly over $\ell \in \mathcal{L}$, we have

$$
\begin{aligned}
&\sqrt{k_n h}(\hat{m}^{(1)}_{k_n,+}(\ell) - m^{(1)}_{P_{k_n},+}(\ell)) \\
=&\sum_{i=1}^{k_n} \sqrt{k_n h}(\mathsf{w}^+_{k_n i}(m^{(1)}(Y_i,W_i,S_i,\ell) - E_Z[m^{(1)}(Y_i,W_i,S_i,\ell)])) + o_p(1).
\end{aligned}
$$

It is then easy to show that $\{(m^{(1)}(Y_i,W_i,S_i,\ell):\ 1 \le i \le k_n,\ n \ge 1\}$ satisfies the manageability condition in the functional central limit theorem (FCLT), or Theorem 10.6 of Pollard (1990). The arguments are similar to those in the proof of Lemma 3.2 of Hsu and Shen (2016) and hold along the sequence $\{P_{k_n}:\ n \ge 1\}$. $\square$

Lemma B.2 *Let $\{P_n\}$ be a sequence of distributions such that each of them satisfies Assumption 3.1. Suppose 3.2 also holds. Then for any subsequence $\{P_{k_n}\}$ of $\{P_n\}$ such that such that for $\lim_{n\to\infty} d(h_{P_{k_n}}, h^*) = 0$ for some $h^* \in \mathcal{H}$, we have*

$$
\begin{aligned}
&\sup_{\ell\in\mathcal{L}} \Big|\sqrt{k_n h}(\hat{\mu}_{k_n}(\ell) - \mu_{P_{k_n}}(\ell)) - \sum_{i=1}^{k_n} \phi_{\mu,k_n i}(\ell)| = o_p(1), \ \textit{and} \\
&\sqrt{k_n h}(\hat{\mu}_{k_n}(\cdot) - \mu_{P_{k_n}}(\cdot)) \Rightarrow \Phi_{C_k h^*_{2,\mu}}
\end{aligned}
$$

where $h^*_{2,\mu} = h^*_1(h^{*+}_2 + h^{*-}_2)h^{*\prime}_1$.

Proof of Lemma B.2: First, note that Lemma B.1 implies that for $\kappa = 1, 2$,

$$
\begin{aligned}
&\sup_{\ell\in\mathcal{L}} |\hat{\rho}^{(\kappa)}_{k_n}(\ell) - \rho^{(\kappa)}_{P_{k_n}}(\ell)| = O_p((k_n h)^{-1/2}), \\
&\sup_{\ell\in\mathcal{L}} |\hat{\varrho}^{(\kappa)}_{k_n}(\ell) - \varrho^{(\kappa)}_{P_{k_n}}(\ell)| = O_p((k_n h)^{-1/2}). \qquad \text{(B.1)}
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
&\sqrt{k_n h}(\hat{\rho}^{(2)}_{k_n}(\ell)\hat{\varrho}^{(1)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\varrho^{(1)}_{P_{k_n}}(\ell)) \\
=&\hat{\varrho}^{(1)}_{k_n}(\ell)\sqrt{k_n h}\big(\hat{\rho}^{(2)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\big) - \rho^{(2)}_{P_{k_n}}(\ell)\sqrt{k_n h}\big(\hat{\varrho}^{(1)}_{k_n}(\ell) - \varrho^{(1)}_{P_{k_n}}(\ell)\big) \\
=&\varrho^{(1)}_{P_{k_n}}(\ell)\sqrt{k_n h}\big(\hat{\rho}^{(2)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\big) - \rho^{(2)}_{P_{k_n}}(\ell)\sqrt{k_n h}\big(\hat{\varrho}^{(1)}_{k_n}(\ell) - \varrho^{(1)}_{P_{k_n}}(\ell))\big) \\
&+ \sqrt{k_n h}\big(\hat{\rho}^{(2)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\big)\big(\hat{\varrho}^{(1)}_{k_n}(\ell) - \varrho^{(1)}_{P_{k_n}}(\ell))\big) \\
=&\varrho^{(1)}_{P_{k_n}}(\ell)\sqrt{k_n h}\big(\hat{\rho}^{(2)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\big) - \rho^{(2)}_{P_{k_n}}(\ell)\sqrt{k_n h}\big(\hat{\varrho}^{(1)}_{k_n}(\ell) - \varrho^{(1)}_{P_{k_n}}(\ell))\big) + o_p(1)
\end{aligned}
$$

where the $o_p(1)$ result holds uniformly over $\ell \in \mathcal{L}$ due to Equation (B.1). Further, since

$$
\sqrt{k_n h}(\hat{\rho}^{(\kappa)}_{k_n}(\ell) - \rho^{(\kappa)}_{P_{k_n}}(\ell)) = \sum_{i=1}^{k_n} \phi^{(\kappa)}_{\rho,k_n i}(\ell), \qquad \sqrt{k_n h}(\hat{\varrho}^{(\kappa)}_{k_n}(\ell) - \varrho^{(\kappa)}_{P_{k_n}}(\ell)) = \sum_{i=1}^{k_n} \phi^{(\kappa)}_{\varrho,k_n i}(\ell),
$$

for all $\ell \in \mathcal{L}$, we have that

$$
\begin{aligned}
&\sqrt{k_n h}(\hat{\rho}^{(2)}_{k_n}(\ell)\hat{\varrho}^{(1)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\varrho^{(1)}_{P_{k_n}}(\ell)) \\
=&\sum_{i=1}^{k_n} \varrho^{(1)}_{P_{k_n}}(\ell)\phi^{(2)}_{\rho,k_n i}(\ell) - \sum_{i=1}^{k_n} \rho^{(2)}_{P_{k_n}}(\ell)\phi^{(1)}_{\varrho,k_n i}(\ell) + o_p(1).
\end{aligned}
$$

and the $o_p(1)$ result holds uniformly over $\ell \in \mathcal{L}$.

Similarly, we can write

$$
\begin{aligned}
&\sqrt{k_n h}(\hat{\rho}^{(1)}_{k_n}(\ell)\hat{\varrho}^{(2)}_{k_n}(\ell) - \rho^{(1)}_{P_{k_n}}(\ell)\varrho^{(2)}_{P_{k_n}}(\ell)) \\
=&\sum_{i=1}^{k_n} \varrho^{(2)}_{P_{k_n}}(\ell)\phi^{(1)}_{\rho,k_n i}(\ell) - \sum_{i=1}^{k_n} \rho^{(1)}_{P_{k_n}}(\ell)\phi^{(2)}_{\varrho,k_n i}(\ell) + o_p(1).
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
&\sqrt{k_n h}(\hat{\mu}_{k_n}(\ell) - \mu_{P_{k_n}}(\ell)) \\
=&\sqrt{k_n h}(\hat{\rho}^{(2)}_{k_n}(\ell)\hat{\varrho}^{(1)}_{k_n}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\varrho^{(1)}_{P_{k_n}}(\ell) - \hat{\rho}^{(1)}_{k_n}(\ell)\hat{\varrho}^{(2)}_{k_n}(\ell) + \rho^{(1)}_{P_{k_n}}(\ell)\varrho^{(2)}_{P_{k_n}}(\ell)) \\
=&\sum_{i=1}^{k_n} \varrho^{(1)}_{P_{k_n}}(\ell)\phi^{(2)}_{\rho,k_n i}(\ell) - \rho^{(2)}_{P_{k_n}}(\ell)\phi^{(1)}_{\varrho,k_n i}(\ell) - \varrho^{(2)}_{P_{k_n}}(\ell)\phi^{(1)}_{\rho,k_n i}(\ell) + \rho^{(1)}_{P_{k_n}}(\ell)\phi^{(2)}_{\varrho,k_n i}(\ell) + o_p(1) \\
=&\sum_{i=1}^{k_n} \phi_{\mu,k_n i}(\ell) + o_p(1).
\end{aligned}
$$

with the $o_p(1)$ result holding uniformly over $\ell \in \mathcal{L}$.

The above equation shows the first part of Lemma B.2. To show the second part, we will apply the Theorem 10.6 of Pollard (1990). Define our triangular array as

$\{\phi_{\mu,k_n i}(\ell) : \ \ell \in \mathcal{L}, \ i \leq k_n, \ 1 \leq n\}$. Note the by the same argument as in Lemma A1 of Hsu et al. (2017), we can show that the triangular is manageable so the part (i) of Theorem 10.2 of Pollard (1990) holds. We can apply similar arguments of Lemma 3.2 of Hsu and Shen (2016) to show that Parts (ii)-(v) hold too. These would complete our proof and we omit the details for brevity. $\square$

Lemma B.3 *Assume that Assumptions 3.1, 3.2, and 2.3 hold. For any subsequence of* k_n *of* n *such that* $\lim_{n\to\infty} d(h_{P_{k_n}}, h^*) = 0$ *for some* $h^* \in \mathcal{H}$*, we have the simulated process* $\widehat{\Phi}^u_{\mu,k_n}(\cdot) \Rightarrow \Phi_{C_k h^*_{2,\mu}}(\cdot)$ *conditional on sample path with probability approaching 1.*

Proof of Lemma B.3: Recall that $\widehat{\Phi}^u_{\mu,k_n}(\cdot) = \sum_{i=1}^{k_n} U_i \cdot \hat{\phi}_{\mu,k_n i}(\cdot)$. It is straightforward to see that $\{U_i \cdot \hat{\phi}_{\mu,k_n i}(\ell) : \ \ell \in \mathcal{L}, \ i \leq k_n, \ 1 \leq n\}$ is manageable. Define $\ddot{h}_{2,k_n,\mu}(\ell_1, \ell_2) = \sum_{i=1}^{k_n} \hat{\phi}_{\mu,k_n i}(\ell_1)\hat{\phi}_{\mu,k_n i}(\ell_2)$. We know that $\sup_{\ell_1,\ell_2 \in \mathcal{L}} |\ddot{h}_{2,k_n,\mu}(\ell_1, \ell_2) - C_k h^*_{2,\mu}(\ell_1, \ell_2)| \xrightarrow{p} 0$. The rest of the proof is similar to that for Lemma 3.3 of Hsu and Shen (2016) and we omit the details. $\square$

Lemma B.4 *Let* $\hat{\sigma}^2_{\mu,k_n,\epsilon}(\ell) = \max\{\hat{\sigma}^2_{\mu,k_n}(\ell), \epsilon \cdot \hat{\sigma}^2_{\mu,k_n}(\ell_0)\}$ *. Assume that Assumptions 3.1 and 3.2 hold. For any subsequence of* k_n *of* n *such that* $\lim_{n\to\infty} d(h_{P_{k_n}}, h^*) = 0$ *for some* $h^* \in \mathcal{H}$*, we know that* $\sup_{\ell\in\mathcal{L}} |\hat{\sigma}_{\mu,k_n,\epsilon}(\ell)^{-1} - \sigma^*_{\mu,\epsilon}(\ell)^{-1}| \xrightarrow{p} 0$*, where* $\sigma^*_{\mu,\epsilon}(\ell) = \max\{\big(C_k h^*_{2,\mu}(\ell,\ell)\big)^{1/2}, \big(\epsilon \cdot C_k h^*_{2,\mu}(\ell_0,\ell_0)\big)^{1/2}\}$.

Proof of Lemma B.4: Since $\hat{\sigma}_{\mu,k_n}(\ell) = \ddot{h}_{2,k_n,\mu}(\ell,\ell)$, where $\ddot{h}_{2,k_n,\mu}(\ell,\ell)$ is defined in the proof of Lemma B.3, we know that $\sup_{\ell\in\mathcal{L}} |\hat{\sigma}_{\mu,k_n}(\ell) - \big(C_k h^*_{2,\mu}(\ell,\ell)\big)^{1/2}| \xrightarrow{p} 0$. Next, by the fact that the maximum operator is a continuous functional, we have $\sup_{\ell\in\mathcal{L}} |\hat{\sigma}_{\mu,k_n,\epsilon}(\ell) - \sigma^*_{\mu,\epsilon}(\ell)| \xrightarrow{p} 0$. Further, since $\sigma^*_{\mu}(\ell_0)$ is bounded away from zero under Assumption 3.1, it follows that $\sup_{\ell\in\mathcal{L}} |\hat{\sigma}_{\mu,\epsilon}(\ell)^{-1} - \sigma^*_{\mu,\epsilon}(\ell)^{-1}| \xrightarrow{p} 0$. This completes the proof of Lemma B.4. $\square$

B.2 Proofs of Theorems

Proof of Lemma 3.1: Lemma 3.1 is a special case of Lemma B.2 proven in the last section.

Proof of Lemma 3.2: Lemma 3.2 is a special case of Lemma B.3 proven in the last section.

Proof of Theorem 3.1: Note that by construction, $\hat{c}^{\eta,LFC}_{n,FRD}(\alpha) \geq \hat{c}^{\eta,GMS}_{n,FRD}(\alpha)$, so the size of the test based on the LFC critical value is always smaller than that based on $\hat{c}^{\eta,GMS}_{n,FRD}(\alpha)$. Therefore, it is sufficient to show that the test based on $\hat{c}^{\eta,GMS}_{n,FRD}(\alpha)$ has uniform size control. To show this, we can apply the same arguments as in the proof of Theorem 4.1 of Hsu et al. (2017) given Lemmas B.2, B.3 and B.4 and we omit the details.

To show part (d) of the theorem, note that if there exists $P_c \in \mathcal{P}_0$ such that $\mathcal{L}^o(P_c)$ is not empty and h_{2,μ,P_c} restricted to $\mathcal{L}^o(P_c) \times \mathcal{L}^o(P_c)$ is not a zero function, then by the same proof based on the pointwise asymptotics as in Lemma 1 of Donald and Hsu (2016) and by Tsirel'Son (1976), we have that under P_c, the CDF function, $G(.)$, of the limiting null distribution of $\widehat{T}_{n,FRD}$ is continuous and is strictly increasing on $(0,\infty)$, and $G(0) > 1/2$. Then, by the same proof for Theorem 2(b) of Andrews and Shi (2013), it is true that under P_c, $\lim_{\eta\to 0} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,GMS}_{n,FRD}(\alpha)) = \alpha$ which implies that $\lim_{\eta\to 0}\limsup_{n\to\infty}\sup_{\{P\in\mathcal{P}_0:\ h_P\in\mathcal{H}_{cpt}\}} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,GMS}_{n,FRD}(\alpha)) \geq \alpha$. Then by combining the result in part (b) of the Theorem, we obtain the uniform size control result in part (d).

To show part (c) of the theorem, the asymptotic results in Lemmas B.2, B.3 and B.4 directly imply that $\lim_{\eta\to 0} P^{LFC}_c(\widehat{T}_{n,FRD} > \hat{c}^{\eta,LFC}_{n,FRD}(\alpha)) = \alpha$ under P^{LFC}_c. Then we know $\lim_{\eta\to 0}\limsup_{n\to\infty}\sup_{\{P\in\mathcal{P}_0:\ h_P\in\mathcal{H}_{cpt}\}} P(\widehat{T}_{n,FRD} > \hat{c}^{\eta,LFC}_{n,FRD}(\alpha)) \geq \alpha$. Combining the result in part (a) of the theorem, we obtain the result in part (c). This completes our proof. $\square$

Proof of Theorem 3.2:

Suppose $CLATE(w_2^*, s) > CLATE(w_1^*, s)$ for some $w_2^* < w_1^*$ and some s. Then by continuity of $CLATE(w, s)$, we can find $w_2' \ll w_1'$ such that $CLATE(w_2', s) > CLATE(w_1', s)$. Again, by continuity of $CLATE(w)$, we can find a small δ such that for all $w_2'' \in \mathcal{N}_{\delta,w}(w_2')$, $w_1'' \in \mathcal{N}_{\delta,w}(w_1')$, and $s'' \in \mathcal{N}_{\delta,s}(s)$, we have $w_2'' \ll w_1''$ and $CLATE(w_2'', s'') > CLATE(w_1'', s'')$. Then we can find a q large enough and $\ddot{\ell} = (\ddot{w}_1, \ddot{w}_2, \ddot{s}, q) \in \mathcal{L}$ such that $\Pi_{j=1}^{d_w}[\ddot{w}_{j1}, \ddot{w}_{j1} + 1/(q+1)] \subseteq \mathcal{N}_{\delta,w}(w_1')$, $\Pi_{j=1}^{d_w}[\ddot{w}_{j2}, \ddot{w}_{j2} + 1/(q+1)] \subseteq \mathcal{N}_{\delta,w}(w_2')$, and $\Pi_{j=1}^{d_s}[\ddot{s}_j, \ddot{s}_j + 1/q] \subseteq \mathcal{N}_{\delta,s}(s)$. It is then straightforward to see that $\mu_{P^*}(\ddot{\ell}) > 0$.

By the definition of $\widehat{T}_{n,FRD}$, we know that $\widehat{T}_{n,FRD} \geq \sqrt{nh}\hat{\mu}_n(\ddot{\ell})/\hat{\sigma}_{\mu,\epsilon}(\ddot{\ell})$, and $\widehat{T}_{n,FRD}$

will diverge to positive infinity when $n \to \infty$, because $\sqrt{nh}\hat{\mu}_n(\ddot{\ell})$ will diverge to positive infinity and $\hat{\sigma}_{\mu,\epsilon}(\ddot{\ell})$ is bounded in probability. Also, both simulated critical values $\hat{c}^{\eta,LFC}_{n,FRD}(\alpha)$ and $\hat{c}^{\eta,GMS}_{n,FRD}(\alpha)$ are bounded in probability. The consistency result of the proposed monotonicity tests then follows. $\square$

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American statistical Association*, 97(457), 284–292.

Andrews, D. and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

Andrews, D. W. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

——— (2014): "Nonparametric Inference Based on Conditional Moment Inequalities," *Journal of Econometrics*, 179, 31–45.

——— (2017): "Inference Based on Many Conditional Moment Inequalities," *Journal of Econometrics*, 196(2), 275–287.

Angrist, J. D. and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 114, 533–575.

Angrist, J. D. and M. Rokkanen (2015): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff," *Journal of the American Statistical Association*, 110(512), 1331–1344.

Barone, G., F. DAcunto, and G. Narciso (2015): "Telecracy: Testing for Channels of Persuasion," *American Economic Journal: Economic Policy*, 7(2), 3060.

Bertanha, M. (2016): "Regression Discontinuity Design with Many Thresholds," *working paper.*

Bertanha, M. and G. W. Imbens (2014): "External validity in fuzzy regression discontinuity designs," *working paper, National Bureau of Economic Research*, No. w20773.

Black, S. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2), 577–599.

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

Carneiro, P., K. V. Loken, and K. G. Salvanes (2015): "A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children," *Journal of Political Economy*, 123(2), 365–412.

Cattaneo, M. D., R. Titiunik, G. Vazquez-Bare, and L. Keele (2016): "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *The Journal of Politics*, 78(4), 1229–1248.

Chetverikov, D. (2013): "Testing Regression Monotonicity in Econometric Models," *working paper*.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008): "Nonparametric Tests for Treatment Effect Heterogeneity," *The Review of Economics and Statistics*, 90(3), 389405.

Donald, S. G. and Y.-C. Hsu (2016): "Improving the Power of Tests of Stochastic Dominance," *Econometric Review*, 35, 553–585.

Dong, Y. and A. Lewbel (2015): "Identifying the effect of changing the policy threshold in regression discontinuity models," *Review of Economics and Statistics*, 97(5), 1081–1092.

Fan, J. and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20(4), 2008–2036.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Frandsena, B. R., M. Frölich, and B. Melly (2012): "Quantile Treatment Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168, 382–395.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000): "Testing Monotonicity of Regression," *Annals of Statistics*, 28(4), 1054–1082.

Hall, P. and N. E. Heckman (2000): "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions," *Annals of Statistics*, 28(1), 20–39.

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

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies*, 65(2), 261–294.

Hotz, V., G. W.Imben, and J. H.Mortimerc (2005): "Predicting the efficacy of future training programs using past experiences at other locations," *Journal of Econometrics*, 125, 241–270.

Hsu, Y.-C. (2016): "Multiplier Bootstrap," *working paper*.

Hsu, Y.-C., C.-A. Liu, and X. Shi (2017): "Testing Generalized Regression Monotonicity," *working paper*.

Hsu, Y.-C. and S. Shen (2016): "Testing Treatment Effect Heterogeneity in Regression Discontinuity Designs," *working paper*.

Ito, K. (2015): "Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program," *American Economic Journal: Economic Policy*, 7(3), 209–237.

Lee, D. S. and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Linton, O., K. Song, and Y.-J. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154, 186–202.

Pollard, D. (1990): "Empirical processes: theory and applications," in *NSF-CBMS regional conference series in probability and statistics.*

Pop-Eleches, C. and M. Urquiola (2013): "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 103, 1289–1324.

Shen, S. and X. Zhang (2015): "Distributional Test for Regression Discontinuity: Theory and Applications," *Review of Economics and Statistics, forthcoming.*

Tsirel'Son, V. S. (1976): "The density of the distribution of the maximum of a Gaussian process," *Theory of Probability & Its Applications*, 20(4), 847–856.

van der Klaauw, W. (2002): "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach," *International Economic Review*, 43(4), 1249–1287.

Wager, S. and S. Athey (forthcoming): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association.*