Regression Discontinuity Design with Distributional Outcomes

David Van Dijcke
Preliminary and incomplete, please do not share

Department of Economics
University of Michigan, Ann Arbor

March 21, 2025

Abstract

This paper introduces Regression Discontinuity Design with Distributional Outcomes (R3D), which generalizes the RDD framework by proposing a novel concept of discontinuities in the conditional average of random quantiles. Two estimators are developed: one adapting local polynomial regression to random quantiles, and another based on local Fréchet regression. Both come with uniform, debiased confidence bands and a multiplier bootstrap procedure. Simulation studies demonstrate that, unlike existing methods, these estimators remain unbiased in the R3D setting. Finally, I apply the proposed methods to estimate the effects of gubernatorial party control on the within-state income distribution. The results suggest a classical equity-efficiency tradeoff under Democratic governorship.

JEL Codes: C14, C21, C13, C12.
Keywords: causal inference, random distributions, quantile treatment effects, Fréchet regression, Wasserstein barycenter, local polynomial regression, RDD, functional data, program evaluation

dvdijcke@umich.edu. All errors are mine. I thank Matias Cattaneo, Ying Fan, Florian Gunsilius, Harry Kleyer, Benjamin Scuderi, Kaspar Wuthrich, and participants at the University of Michigan IO-Econometrics workshop, the University of Michigan Labor lunch, and the New Connections in the Study of Political Economy workshop for helpful discussion and comments.

1 Introduction

The regression discontinuity design (RDD) is a popular non-experimental method for causal inference and program evaluation. It exploits a threshold rule in the assignment variable—often a running variable such as a test score or an income threshold—to identify sharp changes in treatment status among units just above and just below the cutoff. In recent years, it has seen widespread adoption in economics and political science.

Despite its success, the conventional RDD setup presumes that the running variable and outcome are aligned at the same level of aggregation—each unit has its own running variable and a single outcome measurement. In many policy and program contexts, however, the outcome of interest manifests as an entire *distribution* rather than a single scalar. For example, when a school district implements an educational policy based on a district-wide threshold (like a cutoff in the district's poverty rate), we may observe the distribution of student test scores in each district. Similarly, when a minimum wage is implemented along a state border, the outcome of interest could be the distribution of prices across goods in each establishment, rather than a single average price. This motivates the development of a more general framework in which the local average treatment effect is defined over distributions rather than scalars.

In this paper, I extend the standard RDD framework to accommodate distributional outcomes, allowing me to capture how an intervention shifts entire distributions rather than just their means or selected quantiles. I call this the *Regression Discontinuity Design with Distributions (R3D)*. I develop an estimator based on an extension of local polynomial regression to distributional outcomes, called local Fréchet regression Petersen and Müller (2019), for which I derive a novel ANOVA-style test statistic. I also establish the uniform asymptotic normality of a "simple" estimator based on local polynomial regression on random quantiles. Based on that, I further establish the validity of the bootstrap for the Fréchet regression estimator by leveraging the insight that it coincides with a projection of the simple estimator onto the space of quantile functions.

I illustrate the setting and estimators through simulations and an application. Both proposed estimators significantly outperform the quantile RDD estimator proposed in Frandsen et al. (2012), which does not accommodate settings with distribu-

tional outcomes. An empirical application illustrates the estimators' use.

2 Regression Discontinuity with Distributional Outcomes

In this section, I introduce a "distributional" version of the canonical regression discontinuity design. It is distributional in the sense that, conditional on the running variable, the researcher observes entire distributions of the outcome of interest. As such, when the running variable passes the treatment threshold, the entire distribution of this outcome is affected by the treatment. First, I formally introduce the setting, before providing several concrete examples from the literature. Then, I introduce a new definition of distributional local treatment effects appropriate for this setting, as well as a corresponding empirical estimator. I conclude with a brief discussion of why the quantile estimator of Frandsen et al. (2012) cannot be used for this distributional setting.

2.1 Setting

First, I define and discuss the distributional RDD setting. Let $\mathcal{G}$ be the space of probability distributions G on $\mathbb{R}$ such that $\int_{\mathbb{R}} x^2 \, \mathrm{d}G(x) < \infty$. Let $(X, Y) \sim F$ be a random process with joint distribution F on $\mathbb{R} \times \mathcal{G}$. We call X the running variable and Y the outcome variable. Unlike the canonical RD design, Y is no longer a random variable, but instead a random *distribution*. Thus, the ith draw from (X, Y) delivers a complete distribution Y_i at a given value of the running variable X_i, rather than a single real number. Then, denote $D \in \{0, 1\}$ the treatment status. Further, I assume that D is some monotonic function of X such that,

$$D = \begin{cases} 0 & \text{if } X < c \\ 1 & \text{if } X \geq c \end{cases}$$

for some threshold c. That is, treatment is assigned deterministically when the running variable X crosses the threshold c, where I assume without loss of generality that $c = 0$ in the rest of the paper. This is the so-called "sharp" RD design, on which I focus in the main text for expositional clarity, though I derive statistical results for the fuzzy RDD case as well (see Section 2.6.1.

In addition, denote the marginal distributions of X and Y as F_X, F_Y. I assume

that $\mu = E[X], \Sigma = \text{var}(X)$ and the conditional distributions $F_{X|Y}, F_{Y|X}$ exist with Σ positive definite.

2.2 Motivating Examples

To make the setting more concrete, I now provide several examples from the literature that can be viewed as distributional RD designs. They are instances of broader classes of settings where treatment is assigned to units at higher levels of aggregation than the outcomes.

Example 1 (Administrative units)**.** *In an influential article, Ludwig and Miller (2007) studies the impact of Head Start, an early childhood education and development program, on child mortality and educational attainment. Counties above a threshold poverty rate received grant writing assistance from the federal government to develop Head Start proposals, causing a discontinuity in Head Start funding rates at the cut-off point. In an R3D setting, this discontinuity could be exploited to estimate the effect of the program on the life expectancy and test score distributions of children growing up in counties just above the cut-off point. At what ages did child mortality drop the most? Did the program's positive impact on years of schooling help all students equally, or mostly those who struggled the hardest? More generally, the R3D design covers any setting in which a treatment is assigned jointly to all members of an administrative unit, such as counties, districts, or government agencies.*

Example 2 (Institutions)**.** *Clark (2009) considers a British reform allowing public high schools to become autonomous (directly funded by the central instead of the local government) if a majority of parents voted in favor. The paper finds large increases in examination pass rates at schools that narrowly won the vote, compared to those that narrowly lost. This can be cast as an R3D design, by considering the effect of school autonomy on the entire distribution of student test scores within a school. Does school autonomy lead to a broad-based increased in test scores, or do only the worst or best students benefit? More generally, the R3D design comprises any setting where an entire institution is exposed to a treatment, but the outcome of interest affects its members. Furthermore, insofar as vote-based allocation systems typically aggregate decisions of many individuals into higher-level outcomes, many instances of the ubiquitous "close race" RD design can be recast into the R3D framework (Eggers et al., 2015).*

Example 3 (Establishments)**.** *In another seminal article, Card and Krueger (2000) studied the effect of a minimum wage increase in New Jersey on wages, employment, and prices in fast food restaurants, comparing establishments on either side of the border with Pennsylvania. All the outcomes considered can be cast as distributions, with wages and tenure (length of employment) measured at the employee level, and prices at the product level. Did the minimum wage increase mainly spur new hires, or did employment increase across the seniority distribution? Did the pass-through of the wage increase to consumers affect all products equally or mostly premium ones? More generally, the R3D design applies to any setting where the establishments are treated as a whole, but researchers are interested in changes to transactions within the establishment.*

Common to all these examples is that for any value of the running variable (distance to the border, vote share, or poverty level), we observe an entire distribution of the outcome (store prices, test scores, or child mortality), and these outcome distributions are almost certainly different across any two units (across restaurants, schools, or counties). This implies that one needs to model the outcome as a *random distribution* instead of a random variable, as introduced above. Correspondingly, I require new concepts of average treatment effects and discontinuities that are appropriate for random distributions, which I introduce in the next section.

2.3 Local Average Quantile Treatment Effects

2.3.1 Definition

Now, I define a new concept of distributional treatment effects appropriate for the setting introduced above. In the Neyman-Fisher-Rubin notation, denote $Y(0) \in \mathcal{G}$ the counterfactual outcome distribution in the absence of treatment $T = 0$ and $Y(1) \in \mathcal{G}$ the outcome distribution under treatment. Define the observed outcome

$$Y = \begin{cases} Y(0) & \text{if } T = 0 \\ Y(1) & \text{if } T = 1 \end{cases}.$$

Consider, for a moment, the canonical RD setting, such that $Y_E(T)$ the "Euclidean" outcome $\in \mathbb{R}$ instead of $\mathcal{G}$. Then, assuming that treatment effects vary

across units, the classical local treatment effect is,

$$E[Y_E(1) - Y_E(0) \mid X = 0],$$

the conditional expectation of the jump in the outcome variable at the threshold.

In the D-RD setting, $Y(T)$ is a distribution function. An intuitive generalization of the classical average treatment effect to settings with distributional outcomes is given in the below definition. Write $Q_Y(q)$ for the function mapping the cdf Y to quantiles,

$$Q_Y(q) \coloneqq \inf\{y \in \mathbb{R} : p \leq Y(y)\}.$$

Then we get,

Definition 1 (Average Quantile Treatment Effects (AQTE))**.** *The local average quantile treatment effects for the R3D design are,*

$$\begin{aligned} \tau^{DRD}(q) &\coloneqq E\left[Q_{Y(1)}(q) - Q_{Y(0)}(q) \mid X = 0\right] \\ &\coloneqq m_1(q) - m_0(q), \quad q \in [0,1] \end{aligned} \tag{1}$$

Note that the expectation is over the conditional *distribution of distributions*, $P_{Y(T)|X=0}$,

$$m_T(q) = E[Q_{Y(T)}(q) \mid X = 0] = \int Q_{Y(T)}(q)\, \mathrm{d}F_{Y(T)|X=0},$$

for $T = 0, 1$. Similarly, define the conditional variance of the quantile functions at quantile q as,

$$\sigma_T^2(q) \coloneqq \mathrm{var}(Q_{Y(T)}(q) \mid X = 0) = \int \left(Q_{Y(T)}(q) - m_T(q)\right)^2 \mathrm{d}F_{Y(T)|X=0}.$$

These average quantile treatment effects (AQTE) are a compelling way to summarize random distributional treatment effects. First, they offer an intuitive generalization of average treatment effects in the Euclidean setting. In particular, they allow one to study what happens to the outcome distribution of the "average" unit when it crosses the cutoff and receives treatment. Moreover, as I discuss in more detail below, they are equivalent to a difference of Wasserstein barycenters, which respect the intrinsic geometry of the underlying probability measures being averaged over. In

particular, the conditional average quantiles effect as defined above has the intuitive interpretation of being the unique distribution with the lowest possible cumulative cost of being transformed into each of the underlying distributions of the individual units. This is exactly analogous to the interpretation of the mean as the "central tendency" in the standard Euclidean setting.

In what follows, I will show that these average quantile treatment effects (AQTE) can be identified from observed (X, Y) and propose two closely related estimators to estimate them that generalize the canonical local polynomial regression estimators.

2.3.2 Identification

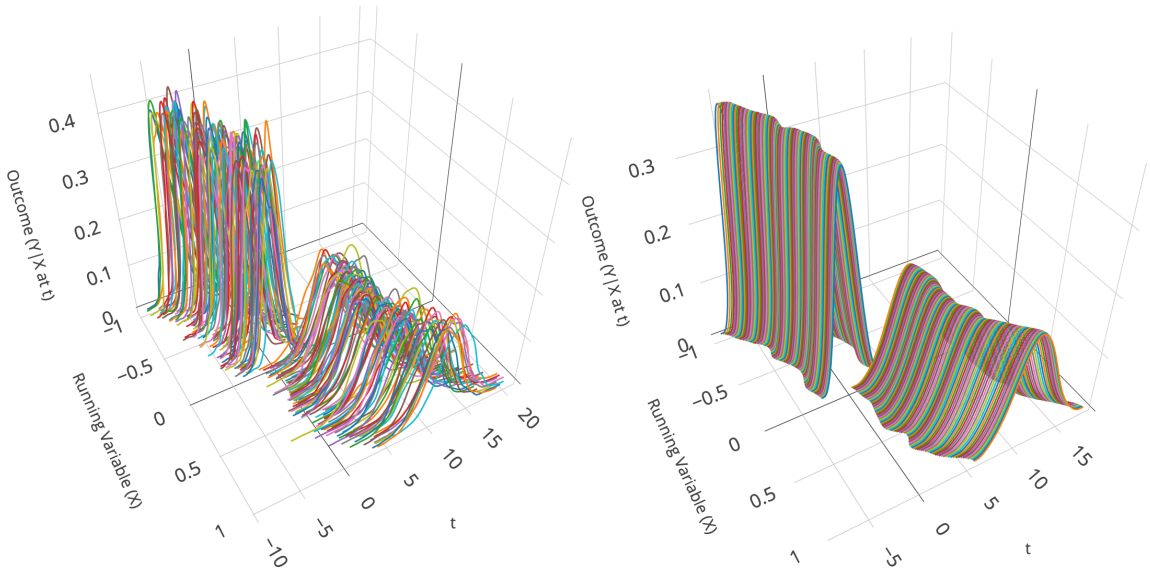
To identify τ^{DRD} from the data, we require two assumptions, which are distributional generalization of the canonical RDD assumptions. First, I assume that the distribution of Y is continuous in the running variable around the threshold.

I1 (Continuity)**.** $\int_U \mathrm{d}F_{Y(T)|X}(x, y)$ *is continuous in* x *for all open sets* $U \subset \Omega$ *and* $x \in]-\varepsilon, \varepsilon[$*, for* $d \in \{0, 1\}$ *and* $\varepsilon > 0$*.*

Importantly, this assumption allows for pointwise discontinuities in the random distribution functions $Y_{|X}(x, y)$, as the assumption pertains to the *distribution of distribution functions* rather than the distribution functions themselves. By contrast, Frandsen et al. (2012) assume pointwise continuity of the observed conditional distribution function $Y_{|X}(t, x)$ around the threshold for all $t \in supp(Y)$. As I discuss in Section 2.7 below, this is an extremely restrictive assumption.

The following example, which I revisit in the simulations below, may help to clarify this important point. Suppose $F_{Y(T)|X=x} \sim N(N((x)^3 + 0.5d, 1), 1)$ for $d = 0, 1$ and $supp(X) = [-1, 1]$. In words, the distribution functions Y are drawn from a class of normal distributions with normally distributed means that depend on X. The corresponding samples of the random objects Y are depicted in Figure 1a. The figure clearly demonstrates what it means for distributions to be drawn randomly: the densities at a given value of the running variable fluctuate, leading to a lack of pointwise continuity with respect to X. This directly generalizes the Euclidean setting, where samples form a random point cloud that generally also lacks continuity. By contrast, 1b shows the Fréchet means estimated on either side of the cutoff using the local polynomial approach set out in Section 2.4. Again, it is immediately apparent that these distributional averages *are* continuous in the running variable, illustrating that

even this simple collection of random Gaussian distributions fails to satisfy continuity in the quantiles even as it does satisfy the weaker continuity assumption in I1. The following proposition establishes this. I include the proof here for intuition.



(a) Sample (b) Fréchet Means

Figure 1: Distributional RDD: Gaussian Example

Proposition 1. *Suppose* $(X, Y) \in (\mathbb{R}, \mathcal{G})$ *and* $Y|X = x \sim N(N(g(x), 1), 1)$ *for some continuous function* $g(x)$*. Then the conditional distribution of distributions* $F_{Y|X}(x, y)$ *satisfies assumption I1, but the conditional distribution function* $Y_X(t, x)$ *is not continuous in* X*.*

Proof. For a given t, define a new random variable $Z = \Phi(t - W)$ where Φ the cdf of the standard normal and $W \sim N(g(x), 1)$. Then $P(Z \leq z) = P(\Phi(t - W) \leq z) = P(t - W \leq \Phi^{-1}(z))$. Since $t - W \sim N(t - g(x), 1)$, we have $P(Z \leq z) = \Phi(\Phi^{-1}(z) - (t - g(x)))$ which is continuous in x. Moreover, since $0 < \Phi(x) < 1$ $Y|X = x$ is almost surely not continuous in x. □

The proposition solidifies the intuition behind Figure 1. While the *probability* of drawing a certain distribution does vary smoothly with X, by virtue of it being a probability, the actual distributions at any two points x, x' close to each other will always be different with probability 1. This arises from the distributions being *random objects* themselves.

The second assumption we need for identification is a standard RDD assumption which posits no manipulation and a non-zero mass of observations around the threshold.

I2 (Density at threshold)**.** *$F_X(x)$ is differentiable at c and $0 < \lim_{x \to c} f_X(x) < \infty$.*

Then, we obtain the following identification result.

Lemma 1 (Identification)**.** *Under Assumptions I1 and I2, the unobserved τ^{R3D} is identified from the joint distribution of the observed (X, Y) as,*

$$\begin{aligned}(2) \qquad \tau^{\text{R3D}}(q) &= \lim_{x \to 0^+} E[Q_Y(q) \mid X = x] - \lim_{x \to 0^-} E[Q_Y(q) \mid X = x] \\ &\coloneqq \lim_{x \to 0^+} m(q) - \lim_{x \to 0^-} m(q) \\ &\coloneqq m_+(q) - m_-(q).\end{aligned}$$

The limits of the observed conditional variances are defined similarly,

$$\sigma^2_{\pm}(q) \coloneqq \lim_{x \to 0^{\pm}} \text{var}(Q_Y(q) \mid X = 0),$$

and the identification of the counterfactual variances $\sigma^2_T(q)$ for $T = 0, 1$ follows from an identical argument.

2.3.3 Discontinuities in Average Distributions

The weak distributional continuity assumption I1 introduced above implies that the treatment has an effect when there is a discontinuity in the observed *average* distribution $E[Q_Y(q) \mid X = c]$ at the threshold $X = 0$. Thus, we can define a discontinuity in our setting to correspond as, for some $q \in [0, 1]$

$$\lim_{x \to 0^+} E[Q_Y(q) \mid X = x] \neq \lim_{x \to 0^-} E[Q_Y(q) \mid X = x].$$

Below, I derive the asymptotic distributions of estimators of the conditional average quantiles that allow one to test for the presence of such discontinuities for a given quantile q. Alternatively, one can conduct inference on the entire distribution at once. To that end, I additionally develop a distributional discontinuity test for the

null hypothesis,

$$\int_0^1 \left(\lim_{x \to 0^+} E[Q_Y(q) \mid X = x] E[Q_Y(q) \mid X = c] - \lim_{x \to 0^-} E[Q_Y(q) \mid X = x] \right)^2 \mathrm{d}q = 0.$$

By the properties of the 2-Wasserstein space, this is equivalent to testing for equality of the average distributions to the left and right from the cutoff, assuming that they have finite variances (Panaretos and Zemel, 2020).

2.4 Estimators

To estimate the local average quantile treatment effects introduced above, I now propose two intuitive estimators that generalize local polynomial regression to our R3D setting with distributional outcomes. The first is based on the simple idea of running local polynomial regressions on the quantile functions, separately at each quantile. The second estimator builds on this idea by projecting the resulting estimator back onto the space of quantile functions. This coincides with the local Fréchet regression estimator proposed in Petersen and Müller (2019), restricted to the space of cumulative distribution functions equipped with the 2-Wasserstein distance (see Appendix A-3 for an overview of the general estimator). In Section 3, I derive valid uniform confidence intervals for both approaches. Nonetheless, extensive simulations in Section 4.1 show that the Fréchet regression estimator outperforms the simple local polynomial approach in finite sample.

2.4.1 Local Polynomial Regression on Quantiles

A simple and intuitive first approach to estimating the average distributions near the threshold is to treat quantiles as the fundamental unit of observation, and estimate their conditional expectations using the local polynomial regression approach that has become canon in RDD (Hahn et al., 2001).

The intuition behind the approach is illustrated in Figure 2: regression lines are fitted through data points that represent randomly scattered quantiles. Note that this is different from the quantile RDD estimator developed in Frandsen et al. (2012), since the latter estimates the quantile functions of *random variables*, rather than the average of *random quantiles* – see the discussion in Section 2.7.

The local polynomial R3D estimators $\hat{m}_{\pm,p}(q)$ of order p for each quantile q can

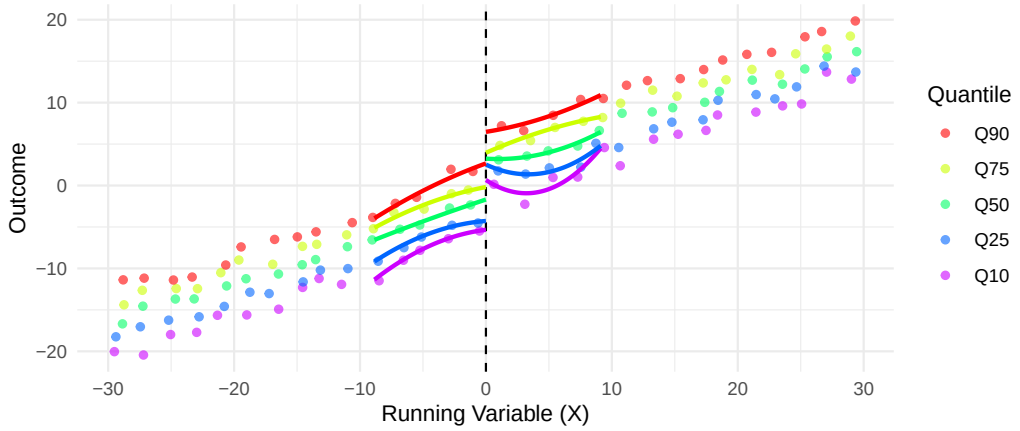


Figure 2: Simple Estimator, Example

thus be written in their standard form

$$\begin{aligned} (3) \qquad \hat{m}_{\pm,p}(q) &= \Big(\text{polynomial fit at } x = 0^{\pm}\Big)\Big|_{\text{order}=p} \\ \hat{\boldsymbol{\alpha}}_{\pm,p} &= \underset{\boldsymbol{\alpha}\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} \delta_i^{\pm}\ K\Big(\tfrac{X_i}{h}\Big)\Big[Q_{Y_i}(q) - \boldsymbol{\alpha}^{\top} r_p\Big(\tfrac{X_i}{h}\Big)\Big]^2, \end{aligned}$$

where $K_h(x) \coloneqq \frac{1}{h}K(x/h)$, $\delta_i^{\pm} \coloneqq 1\big\{X_i \underset{<}{\geq} c\big\}$. One can add higher order polynomial terms, but linear approximations are standard in the RDD literature (Pei et al., 2022). The only difference with the standard local linear RDD estimator is that we now have i.i.d. samples $(Q_Y(q), X_i)$ instead of (Y_i, X_i). Standard derivations give the following solution,

$$(4) \qquad \hat{m}_{\pm,p}(q) = \sum_{i=1}^{n} s^{(p)}_{\pm,in}(h) Q_{Y_i}(q)$$

where $s^{(p)}_{+,in}(h)$ are the standard empirical weights for a local polynomial regression of order p (Fan and Gijbels, 1996), which I derive explicitly in Appendix A-2.

Note that the estimator $\hat{m}_{\pm,p}(q)$ is technically a function of x, but I suppress this for all estimators to ease notation, since I only consider the estimator at $X = 0$. Further, since the weights $s_{\pm,in}(h)$ can be negative, $\hat{m}_{\pm,p}$ need not be a quantile function. To resolve this, I use the standard monotone rearrangement from Chernozhukov et al. (2010). In the next section, I also propose a more principled projection approach, which allows me to interpret the resulting estimator as a local Fréchet regression in

the 2-Wasserstein space.

The corresponding R3D estimator then is, for each $q \in [0,1]$,

$$\hat{\tau}_p^{\mathrm{R3D}}(q) \coloneqq \hat{m}_{+,p}(q) - \hat{m}_{-,p}(q) \tag{5}$$

In Section 3 below, I show that this estimator converges uniformly to an asymptotic normal distribution centered at the true treatment effect, for $p \geq 2$. The lack of asymptotic bias follows from Remark 7 in (Calonico et al., 2014), which established that their robust and unbiased local linear estimator is equivalent to a local quadratic estimator with bandwidth chosen based on the local linear fit (Chiang et al., 2019).

2.4.2 Local Fréchet Regression

Three intuitive improvements can be made to the local linear regression on quantiles introduced above. First, the resulting function is not guaranteed to be a quantile function because the weights $s_{\pm,in}^{(p)}(h)$ can be negative and thus introduce non-monotonicity (quantile crossing). Second, the pointwise estimation approach ignores global function information, which degrades the estimator's finite-sample performance. Third, the pointwise estimation approach also requires repeated bandwidth selection and estimation for each quantile, potentially leading to computational overhead. To resolve these three issues at once, I consider the following extension of the estimator,

$$\hat{m}_{\pm,\oplus,p} \coloneqq \Pi_{\mathcal{Q}}\left(\hat{m}_{\pm,p}\right) \coloneqq \underset{m_\pm \in Q(\mathcal{G})}{\operatorname{argmin}} \int_0^1 \left(\hat{m}_{\pm,p}(q) - m_\pm(q)\right)^2 \mathrm{d}q, \tag{6}$$

where $\mathcal{G}$ is the space of probability distributions with finite variance and $Q(\mathcal{G})$ is the corresponding space of quantile functions, and I define $\Pi_{\mathcal{Q}}$ as the L^2 projection onto the space of quantile functions. In Proposition A-3, I show that $\hat{m}_{\pm,\oplus,p}$ is unique and exists under the stated assumptions. The estimated treatment effects are then defined as,

$$\hat{\tau}_{\oplus,p}^{\mathrm{R3D}}(q) \coloneqq \hat{m}_{+,\oplus,p}(q) - \hat{m}_{-,\oplus,p}(q). \tag{7}$$

The augmented estimator in (6) is an L^2 projection of the local polynomial estimator introduced above, with the entire function projected onto the space of quantile functions. As such, it is a form of isotonic regression (Robertson et al., 1988), and the

approach can be viewed as a "double regression": a local linear regression on pointwise quantile functions, followed by a regression on quantile functions. More importantly, due to the deep connection between L^2 space and the 2-Wasserstein space, this extended estimator is equivalent to the local Fréchet regression estimator of Petersen and Müller (2019), restricted to the space of finite-variance probability distributions $\mathcal{G}$ equipped with the 2-Wasserstein distance, d_{W_2}. In Appendix A-3, I define these objects and provide an overview of local Fréchet regression. Here, the main thing to note is that $\hat{m}_{\pm,\oplus,p}$ converges to the population object $\lim_{x\to 0^{\pm}} E[Q_Y \mid X = x]$, which is the quantile function with quantiles equal to the average quantile $m_{\pm}(q)$, the population limit of the local polynomial estimator. This is established in Theorem 2 through the insight that the projection of $m_{\pm}(q)$ onto the space of quantile functions is just an identity operator, as $m_{\pm}(q)$ is a valid quantile function. Another way to view this beautiful property is that the local Fréchet estimator converges to the conditional Fréchet mean on $(\mathcal{G}, d_{W_2})$, which is the "conditional Wasserstein barycenter" (Agueh and Carlier, 2011; Fan and Müller, 2024), the unique distribution that has a quantile function equal to the average of the quantile functions with distribution $F_{Y(T)|X}$. In short, the Fréchet estimator in (6) offers a principled functional approach to estimating the LAQTE in Definition 1, while converging to the same object in population.

This connection to local Fréchet regression in Wasserstein space explains why the "double regression" approach in (6) is preferred over monotonizing the simple local linear estimator in (3). The estimator in (6) can be interpreted as the unique quantile function that minimizes the "quantile least squares distance" (the 2-Wasserstein distance) to each of the quantile functions Q_{Y_i} in the space of probability distributions, weighted by the local regression weights $s^{(p)}_{\pm,in}(h)$. In other words, it is the weighted central tendency of the quantile functions in probability space. This interpretation makes the projection approach in (6) preferable over the mononotization approach of Chernozhukov et al. (2010), as the quantile function resulting from the latter generally does not have this desirable interpretation. Another advantage is that the local Fréchet regression approach more naturally utilizes global function information by smoothing large deviations across quantiles in order to minimize the objective function. In comparison, the monotone rearrangement approach just sorts the quantiles but does not otherwise use the global function information to do so in any optimal manner. These superior theoretical qualities of the Fréchet regression approach ex-

press themselves in better finite-sample performance in the simulations in Section 4.1.

2.5 Overview of Inference

For these two R3D estimators, based on local polynomial and Fréchet regression, I derive the asymptotic distribution, uniformly over $q \in [a, b]$ a compact subset of $[0, 1]$, in Section 3 below. Further, I propose estimated multiplier bootstrap processes $\hat{\mathbb{G}}^{\text{R3D}}$, $\hat{\mathbb{G}}^{\text{F3D}}$ that are shown to converge to the uniform limiting law and hence can be used to construct uniform confidence bands. This allows one to answer determine what quantiles have a statistically significant treatment effect while accounting for multiple testing due to the functional nature of the estimands.

Further, the bootstrapped distributions can also be used to construct critical values for various distributional hypothesis tests. In particular, one can test for treatment nullity and treatment homogeneity on a particular part of the distribution $[\underline{q}, \overline{q}] \subset (0, 1)$ through the following tests, where the critical values for these tests can

	Test Statistic
Uniform Treatment Nullity	$\max_{q \in [\underline{q}, \overline{q}]} \sqrt{n\, h_n} \left\lvert \hat{\tau}^{\text{R/F3D}}(q) \right\rvert$
Treatment Homogeneity	$\max_{q \in [\underline{q}, \overline{q}]} \sqrt{n\, h_n} \left\lvert \hat{\tau}^{\text{R/F3D}}(q) - \dfrac{1}{\overline{q} - \underline{q}} \displaystyle\int_{[\underline{q}, \overline{q}]} \hat{\tau}^{\text{R/F3D}}(q')\, dq' \right\rvert$

be constructed by taking the $(1-\alpha)$-th quantiles of $\left\{ \max_{q \in [\underline{q}, \overline{q}]} \left\lvert \hat{\mathbb{G}}^{\text{R/F3D}'}(q) \right\rvert \right\}_{b=1}^{B}$ and $\left\{ \max_{q \in [\underline{q}, \overline{q}]} \left\lvert \hat{\mathbb{G}}^{\text{R/F3D}'}(q) - \frac{1}{\overline{q} - \underline{q}} \int_{[\underline{q}, \overline{q}]} \hat{\mathbb{G}}^{\text{R/F3D}'}(q')\, \mathrm{d}q' \right\rvert \right\}_{b=1}^{B}$, where α is the desired level of statistical significance and B the number of bootstrap repetitions.

2.6 Extensions

2.6.1 Fuzzy R3D

So far, I have focused on the sharp regression discontinuity design, where treatment assignment is a deterministic function of the cutoff, to simplify notation. Here, I show how one can define a fuzzy R3D design, where treatment assignment is a random function of the cutoff so only a fraction of units are treated on either side of it (Hahn et al., 2001).

Define T^0, T^1, the local potential treatment states, as $\lim_{x\to 0^\pm} T(x)$ where $T(x)$ is the potential treatment status as a function of the running variable. Further, define the events,

- Compliers: $C = \{\omega : T^1(\omega) > T^0(\omega)\}$
- Indefinites: $I = \{\omega : T^1(\omega) = T^0(\omega)\} \setminus \{\omega : T^1(\omega) = T^0(\omega) \in \{0, 1\}\}$

The treatment effects of interest are,

Definition 2 (Fuzzy LAQTE)**.** *The local average quantile treatment effects for the fuzzy R3D design are,*

$$\tau^{\text{F3D}}(q) := E\left[Q_{Y(1)}(q) - Q_{Y(0)}(q) \mid X = 0,\, C\right] \quad q \in [0, 1] \tag{8}$$

To identify these, I need the following standard additional assumptions,

I3 (Treatment Continuity)**.** *$E[T|X = x]$ is continuous in x over $]-\varepsilon, \varepsilon[$, $\varepsilon > 0$.*

I4 (Monotonicity)**.** *$|\lim_{x\to 0^+} P(D = 1 \mid X = x) - \lim_{x\to 0^-} P(D = 1 \mid X = x)| > 0$ and $Pr(\text{Indefinites}) = 0$.*

Under these assumptions, I obtain,

Lemma 2 (Fuzzy Identification)**.** *Under Assumptions I3 and I4, the unobserved τ^{F3D} is identified from the joint distribution of the observed (X, Y, T) as,*

$$\begin{aligned}\tau^{\text{F3D}}(q) &= \frac{\lim_{x\to 0^+} E\left[Q_Y(q) \mid X = x\right] - \lim_{x\to 0^-} E\left[Q_Y(q) \mid X = x\right]}{\lim_{x\to 0^+} E\left[T \mid X = x\right] - \lim_{x\to 0^-} E\left[T \mid X = x\right]} \\ &:= \frac{m_+(q) - m_-(q)}{m_{+,T}(q) - m_{-,T}(q)}.\end{aligned} \tag{9}$$

This Wald estimator takes the same form as the standard fuzzy RDD one (Hahn et al., 2001), except that the outcomes are random quantiles. Note that I can work with this simpler form compared to Frandsen et al. (2012, Eq. 2-3) because I work directly with the random quantiles and hence do not need to invert the CDFs on each side.

The corresponding treatment effect estimator, using local polynomial regressions of order p, then becomes,

$$\hat{\tau}_p^{\text{F3D}}(q) := \frac{\hat{\tau}_p^{R3D}(q)}{\hat{m}_{+,T,p} - \hat{m}_{-,T,p}} \tag{10}$$

where

$$\hat{m}_{\pm,T,p} \coloneqq \sum_{i=1}^{n} s^{(p)}_{\pm,in}(h) T_i.$$

The corresponding Fréchet estimator is simply,

$$\hat{\tau}^{\mathrm{F3D}}_{\oplus,p}(q) \coloneqq \frac{\hat{\tau}^{R3D}_{\oplus,p}(q)}{\hat{m}_{+,T,p} - \hat{m}_{-,T,p}} \tag{11}$$

2.6.2 Empirical Quantile Functions

So far, I have assumed that the researcher observes entire distribution functions Y_i. In practice, however, there is often another layer of randomness where one only observes further i.i.d. *samples* $Z_{ij}, j = 1, \ldots, n_i$ from these distribution functions, such that $Z_{ij} \in \mathbb{R} \sim Y_i$.[1] In Section 3, I show that under standard assumptions for the convergence of the empirical quantile functions, they converge to the true quantile functions faster than the proposed R3D estimators and hence do not affect the asymptotic results. The corresponding sharp RD estimator is defined as,

$$\bar{\mu}_{\pm,p}(q) \coloneqq \frac{1}{n} \sum_{i=1}^{n} s^{(p)}_{\pm,in}(h) \hat{Q}_{Y_i}(q), \tag{12}$$

where

$$\hat{Q}_{Y_i}(q) \coloneqq \inf \left\{ x : \hat{Y}_i(x) \geq q \right\} \tag{13}$$

with

$$\hat{Y}_i(x) \coloneqq \frac{1}{n} \sum_{j=1}^{n} 1\left(Z_{ij} \leq x\right).$$

The other estimators are similarly modified by plugging in $\hat{Q}_{Y_i}$ and denoted with a bar instead of a hat, e.g. $\bar{\tau}^{\mathrm{R3D}}_p$.

2.7 Comparison to Quantile RDD

Before diving into the statistical properties of these estimators, I briefly discuss the difference between the R3D setting and the quantile RD estimator of Frandsen et al.

[1] See Chen et al. (2023) for an analogous setting in the context of distribution-on-distribution regression, and Zhou and Müller (2024) in a similar setting.

(2012). The key insight is that quantile RDs are appropriate for estimating QTEs for scalar-valued outcomes, while the R3D estimator can estimate (average) QTEs for distributional outcomes, and there is no overlap in use cases.

Comparing the population quantities targeted by each estimator illustrates this point clearly. The estimators target the quantities

$$\text{R3D} : \lim_{x \to 0^{\pm}} E[Q_Y(q) \mid X = x] \qquad \text{Q-RDD} : \lim_{x \to 0^{\pm}} E[1(Y \leq y) \mid X = x],$$

Thus, the R3D aims to estimate a conditional average quantile. The Q-RDD on the other hand, aims to estimate a fixed distribution function. Practically, they do so with the following local linear estimators,

$$\text{R3D} : \frac{1}{n} \sum_{i=1}^{n} s_{\pm,i}(h) Q_{Y_i}(q) \qquad \text{Q-RDD} : \frac{1}{n} \sum_{i=1}^{n} s_{\pm,i}(h) 1(Y_i \leq y).$$

As can be seen, the R3D approach *first* estimates quantiles and only then runs a local linear regression. Thus, distribution estimation at a given $X = x$ precedes smoothing. By contrast, the Q-RDD estimator intrinsically estimates the distribution *by smoothing*. If the observed distributions do not vary smoothly, as is the case in the R3D setting, then this approach will be biased and inconsistent. Moreover, even if the distributions do vary smoothly (i.e. they are not random) the quantile RD estimator fails to account for the two layers of sampling variability, within and across units, that characterize the R3D setting. Though the estimator may be consistent in this edge case, it is still characterized by large finite-sample bias and low precision due to its incorrect model of the sampling distribution.

Underlying these arguments are three distinct differences between the R3D and the Q-RDD setting. First, the sampling model imposed by the Q-RDD setting does not correctly represent the underlying data-generating process, assuming i.i.d. sampling of scalar-valued outcomes instead of distributional ones, which ignores the within-unit sampling that characterizes the R3D setting, and could never result in multiple data points having the same value of the (continuous) running variable. Second, the quantile continuity assumption required for the identification of the estimator in Frandsen et al. (2012) is extremely restrictive, imposing that two units that are both close to the threshold have identical distributions. In the examples in Section 2.2, this would imply that, conditional on having the same value of the running variable, two

different restaurants would have the exact same distributions of product prices, two different schools the same distribution of tests, and two different counties the same distribution of child mortality. Of course, there is no reason why the cheapest product in one restaurant should have the same price as in another, or the best student in one school the same score as in another, even if their running variables did happen to take on the same value. The estimator I propose allows one to impose a much weaker continuity assumption in I1. In particular, it only requires, for example, that the test score distributions of schools near the cutoff *on average* look the same, while allowing the distributions of *specific* schools to differ. In this way, I1 is the direct distributional analogue of the conditional mean continuity assumption originally imposed in Hahn et al. (2001, A2), which only requires the expectation of the random outcome variable to be continuous but leaves its distribution otherwise unrestricted. Third, and similarly, the standard assumption that treatment effects vary across units automatically implies that the counterfactual distributions must be *random* objects themselves: the outcome is a distribution, and receiving treatment affects this distribution differently for different units. More concretely: if a policy affects the entire workforce of a company, but does so differently at Company A compared to Company B, then *even if* all untreated companies have identical outcome distributions (an unrealistically strong assumption), the counterfactual outcome distributions of those companies will still differ.

3 Statistical Results

Now, I derive the asymptotic distribution of both the local polynomial and the Fréchet regression estimator. I do so in full generality for p-th order local polynomials, allowing for both the sharp and fuzzy RDD setting. The results for the local polynomial estimator follow from an application of the general results in Chiang et al. (2019), extended to random distributional outcomes, and those for the local Fréchet regression follow consequently from the functional delta method. Then, I propose a general test for discontinuities in conditional expectations in general metric spaces, and establish its asymptotic validity and consistency.

3.1 Assumptions

Throughout, I work on the restricted set of quantiles $[a, b]$, a compact subset of $(0, 1)$, and let $\underline{c} < 0 < \bar{c}$.

K1 (Kernel). *The kernel K is a continuous probability density function, symmetric around zero, and non-negative valued with compact support.*

K2 (Bandwidth). *The bandwidths satisfy $h_1(q) = c_1(q)h$ and $h_2(q) = c_2(q)h$ for $c_1(q) : [a, b] \to [\underline{c}, \bar{c}] \subset \mathbb{R}$ a bounded Lipschitz function, $c_2(q) = \bar{c}_2 > 0$, for baseline bandwidth $h = h_n$ satisfying $h \to 0$, $nh^2 \to \infty$, $nh^{2p+3} \to 0$.*

L1 (Sampling). *(i) $\{(Y_i, T_i, X_i)\}_{i=1}^n$ are i.i.d. copies of a random element (Y, T, X) defined on a probability space $(\Omega^x, \mathcal{F}^x, P^x)$.*

(ii) $\{Z_{ij}\}_{j=1}^{n_i}$ are i.i.d. draws from the random distribution Y_i.

L2 (Average Quantile Continuity). *Let $\mathcal{N} \subset \mathbb{R}$ be an open interval containing the threshold c. For each $x \in \mathcal{N} \coloneqq (-\epsilon, \epsilon)$ for $\epsilon > 0$ and $q \in [a, b] \subset (0, 1)$, the following conditions hold:*

(i) The maps $(x, q) \mapsto E[Q_Y(q)|X = x]$ and $x \to E[T|X = x]$ are p-times continuously differentiable in x on $\mathcal{N}$, with all partial derivatives (up to order p) Lipschitz in x on $\mathcal{N} \times [a, b]$.

(ii) For any fixed $q_1, q_2 \in [a, b]^2$, the map $x \mapsto \mathbb{E}\big[\, Q_Y(q_1)\, Q_Y(q_2) \mid X = x\big] \in C^1(\mathcal{N} \setminus \{c\})$ with bounded derivatives in x and bounded limits as $x \to 0^{\pm}$.

L3 (Random Quantile Spread). *The maps $Y \to \sup_{q\in[a,b]} |Q_Y(q)|\}$ and $x \to \sup_{q\in[a,b]} |E[Q_Y(q) \mid X = x]|$ are in $L^{2+\epsilon}(P^x)$ on $[\underline{c}, \bar{c}] \times \mathcal{G}$.*

M1 (Multiplier). *$\{\xi_i\}_{i=1}^n$ is an independent standard normal random sample defined on a probability space $(\Omega^\xi, \mathcal{F}^\xi, P^\xi)$ independent of $(\Omega^x, \mathcal{F}^x, P^x)$.*

K1 is a standard kernel assumption and is satisfied by the commonly used triangular and uniform kernels. K2 is a standard bandwidth assumption, with the caveat that for the local polynomial order $p > 1$, it accommodates the bandwidth rates implied by common bandwidth selection procedures, which are typically slower than $h = n^{-1/5}$ (Calonico et al., 2014). Moreover, the assumption accommodates

quantile-specific bandwidths. L2 (i) is a stronger version of the standard continuity assumptions I1 and I3 that ensures the Taylor expansions required for local polynomial regression of order p are well-defined. L2 (ii) further provides some minimal control over the functional objects $E[Q_Y(q)|X = x]$. Note that both (i) and (ii) are implied by the much stronger assumption that the random distribution $F_{Y|X}$ evolves smoothly, which would be the random-distribution equivalent of Assumption E1 in Frandsen et al. (2012). For Assumption L3, first note that clearly, for every Y, there exists an M_Y such that $\sup_{q\in[a,b]} Q_Y(q) < M_Y$. However, the assumption strengthens this pointwise fact into a statement that these caps cannot "blow up" too often across all possible realizations Y. In practice, this means "although each Y can have unbounded support, the family $\mathcal{G}$ must not produce extremely large quantiles too frequently around the cutoff. As such, the assumption controls the across-distribution variance, enabling uniform statistical arguments. Finally, Assumption M1 is a standard assumption for multiplier bootstraps that can easily be satisfied in practice.

Finally, for the extension with empirical quantile functions, I further impose the following two assumptions,

Q1 (Empirical Quantiles)**.** *Any distribution function $Y \in \mathcal{G}$ either*

(i) has compact support and is C^1 with strictly positive density, or

(ii) has infinite support and for every $0 < q_1 < q_2 < 1$ there exists an $\varepsilon > 0$ such that Y is continuously differentiable on the interval

$$[Q_{Y_i}(q_1) - \varepsilon, Q_{Y_i}(q_2) + \varepsilon]$$

with positive density.

Q2. *There exists a sequence $m = n^\gamma, \gamma \geq 1$ such that $\min\{n_i : i = 1, \ldots, n\} \geq m$. Moreover, the sample sizes for each Z_{ij} are asymptotically balanced, i.e. $\frac{n_i}{n_j} \to \eta_{ij}$ with $0 < \eta_{ij} < \infty$ for $i, j \in \{1, \ldots, n\}$.*

Q1 implies that the empirical quantile functions converge uniformly (van der Vaart, 2000, Corollary 21.5). The assumption can be relaxed for the case with discrete distributions and finite support, since then standard pointwise convergence results for local linear regression imply uniform convergence (Fan and Gijbels, 1996). Q2 is a weak assumption on the number of measurements per distribution that guarantees the empirical quantile functions will converge faster than the estimators (3) and (6).

3.2 Asymptotic Distribution

3.2.1 Local Polynomial Estimator

Under these assumptions, I obtain the asymptotic distribution for the local polynomial estimator. For the formal results, I assume without loss of generality that $c = 0$ and that the kernel K is supported on $[-1, 1]$, to simplify notation. Define $g_1 : (Y, T, q) \subset (\mathcal{G}, [a, b]) \to Q_Y(q)$, $g_2 : (Y, T, q) \subset 0, 1 \to T$. Further, define the population residual $\mathcal{E}(Y_i, T_i, X_i, q) \coloneqq g_k(Y_i, T_i, Q) - E[g_k(Y_i, T_i, q)|X_i = x]$, $k = 1, 2$ and let

$$\sigma_{kl}(q, q' \mid x) = E\left[\mathcal{E}_k\left(Y_i, T_i, X_i, q\right) \cdot \mathcal{E}_l\left(Y_i, T_i, X_i, q'\right) \mid X_i = x\right]$$

with $\sigma_{kl}(q, q' \mid 0^{\pm}) = \lim_{x \to 0^{\pm}} \sigma_{kl}(q, q' \mid x)$. Moreover, let e_0 denote the 0th standard basis vector of $\mathbb{R}^p$, $(1, 0, \ldots, 0)$, and write $\Gamma_{\pm,p} \coloneqq \int_{\mathbb{R}_{\pm}} K(u) r_p(u) r_p'(u)\, \mathrm{d}u$. Also, let $X_n \rightsquigarrow X$ denote weak convergence for some sequence of random variables X_n and a random variable X, and $X_n \underset{\xi}{\overset{p}{\rightsquigarrow}} X$ denote conditional weak convergence. The latter is defined as $\sup_{h \in BL_1} \left|E_{\xi|x}[h(X_n) - E[h(X)]]\right| \underset{x}{\overset{p}{\to}} 0$ where BL_1 the set of bounded Lipschitz functions with supremum norm bounded by 1 and $\underset{x}{\overset{p}{\to}}$ denotes convergence in probability with respect to probability measure P^x (van der Vaart and Wellner, 1996, §1.13). Then we first get the following preliminary result for the conditional means,

Theorem 1 (Convergence: Conditional Means)**.** *Under Assumptions I2, K1, K2, L1-(i), L2, L3,*

$$\sqrt{nh} \begin{bmatrix} \hat{m}_{\pm,p}(\cdot) - m_{\pm}(\cdot) \\ \hat{m}_{\pm,T,p}(\cdot) - m_{\pm,T}(\cdot) \end{bmatrix} \rightsquigarrow \begin{bmatrix} c_1(\cdot)^{-1/2} \mathbb{G}_{H\pm}(\cdot, 1) \\ c_2(\cdot)^{-1/2} \mathbb{G}_{H\pm}(\cdot, 2) \end{bmatrix}$$

where $\mathbb{G}_{H\pm} : \Omega^x \to l^{\infty}\left([a, b] \times \{1, 2\}\right)$ *is a zero-mean Gaussian process with covariance function,*

$$H_{\pm,p}((q, k), (q', l)) = \frac{\sigma_{kl}(q, q' | 0^{\pm}) e_0' \left(\Gamma_{\pm,p}\right)^{-1} \Psi_{\pm,p}\left((q, k), (q', l)\right) \left(\Gamma_{\pm,p}^{-1}\right) e_0}{\sqrt{c_k(q) c_l(q')} f_X(0)},$$

where,

$$\Psi_{\pm,p}\left((q,k),(q',l)\right) \coloneqq \int_{\mathbb{R}} r_p(u/c_k(q))r_p'(u/c_l(q'))K(u/c_k(q)K(u/c_l(q'))\,\mathrm{d}u$$

for each $q, q' \in [a,b]$.

Then, a simple application of the functional delta method yields the following result.

Theorem 2 (Convergence: Treatment Effect)**.** *Under the assumptions of Theorem 1 it follows that,*

$$\sqrt{nh}\big(\hat{\tau}_p^{\mathrm{R3D}}(q) - \tau^{\mathrm{R3D}}\big) \rightsquigarrow c_1(\cdot)^{-1/2}\,\mathbb{G}_\Delta(\cdot,1) \coloneqq \mathbb{G}^{\mathrm{R3D}},$$

and under the additional Assumptions I3, I4

$$\begin{aligned}
&\sqrt{nh}\big(\hat{\tau}_p^{\mathrm{F3D}}(q) - \tau^{\mathrm{F3D}}\big) \\
&\quad\rightsquigarrow \frac{c_1(\cdot)^{-1/2}(m_{+,D}(\cdot) - m_{-,D}(\cdot))\,\mathbb{G}_\Delta(\cdot,1) - c_2(\cdot)^{-1/2}(m_+(\cdot) - m_-(\cdot))\,\mathbb{G}_\Delta(\cdot,2)}{(m_{+,D}(\cdot) - m_{-,D}(\cdot))^2}, \\
&\quad\coloneqq \mathbb{G}^{\mathrm{F3D}}
\end{aligned}$$

where, for $k \in \{1,2\}$,

$$\mathbb{G}_\Delta(\cdot,k) \coloneqq \mathbb{G}_{H+}(\cdot,k) - \mathbb{G}_{H-}(\cdot,k),$$

and $\mathbb{G}_{H\pm}(\cdot,k)$ *are as defined in Theorem 1.*

In practice, it is easier to approximate the limiting processes in Theorem 2 with a multiplier bootstrap, which preserves the local structure without full resampling. To that end, I use the pseudo-random samples $\{\xi\}_{i=1}^n$ defined in L1 to define the estimated multiplier process,

$$\hat{\nu}_{\xi,n}^{\pm}(q,k) = \sum_{i=1}^{n} \xi_i \frac{e_0'\left(\Gamma_{\pm,p}\right)^{-1}\hat{\mathcal{E}}_k\left(Y_i,T_i,X_i,\theta\right) r_p\left(\frac{X_i}{h_k(q)}\right) K\left(\frac{X_i}{h_k(q)}\right)\delta_i^{\pm}}{\sqrt{nh_k\left(q\right)}\hat{f}_X(0)}, \tag{14}$$

where $\hat{f}_X(0)$ is any uniformly consistent estimator of $f_X(0)$, and $\hat{\mathcal{E}}_k\left(Y_i,T_i,X_i,\theta\right)$ is any uniformly consistent first-stage estimator of the residual $\mathcal{E}_k$. In practice, I will

use the first-stage estimator proposed in Chiang et al. (2019, A.6), described in detail in Appendix A-4.1. The process $\hat{\nu}^{\pm}_{\xi,n}(q,k)$ is an estimator for the uniform Bahadur representation of the debiased processes $\hat{m}_{\pm,p}(q) - m_{\pm}(q)$, $\hat{m}_{\pm,T,p}(q) - m_{\pm,T}(q)$ (Chiang et al., 2019), see the proof of Theorem 1 for more details. Then, I obtain the uniform validity of the multiplier bootstrap,

Theorem 3 (Bootstrap)**.** *Under the Assumptions of Theorem 1 and Assumption M1 it follows that* $\hat{\nu}^{\pm}_{\xi,n} \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{H^{\pm}}$ *and thus,*

$$\hat{\mathbb{G}}^{\mathrm{R3D}}(\cdot) \coloneqq c_1(\cdot)^{-1/2}\hat{\nu}_{\Delta,n}(\cdot,1) \rightsquigarrow \mathbb{G}^{\mathrm{R3D}}$$

and under the additional Assumptions I3, I4,

$$\hat{\mathbb{G}}^{\mathrm{F3D}}\left(\hat{m}_{+,p},\hat{m}_{-,p},\hat{m}_{+,D,p},\hat{m}_{-,D,p}\right)(\cdot) \coloneqq \frac{c_1(\cdot)^{-1/2}\left(\hat{m}_{+,D,p}(\cdot)-\hat{m}_{-,D,p}(\cdot)\right)\hat{\nu}_{\Delta,n}(\cdot,1) - c_2(\cdot)^{-1/2}\left(\hat{m}_{+,p}(\cdot)-\hat{m}_{-,p}(\cdot)\right)\hat{\nu}_{\Delta,n}(\cdot,2)}{\left(\hat{m}_{+,p}(\cdot)-\hat{m}_{-,p}(\cdot)\right)^2} \rightsquigarrow \mathbb{G}^{\mathrm{F3D}}$$

where, for $k \in \{1,2\}$*,*

$$\hat{\nu}_{\Delta,n}(\cdot,k) \coloneqq \hat{\nu}^{+}_{\xi,n}(\cdot,k) - \hat{\nu}^{-}_{\xi,n}(\cdot,k)$$

A practical algorithm for computing the empirical bootstrap is provided in Appendix A-4.3. The asymptotic validity and consistency of the tests proposed in Section 2.5 follow immediately from Theorem 3.

3.2.2 Fréchet Estimator

Turning to the Fréchet estimator, I now show that it has the same asymptotic distribution as the local polynomial estimator. I include a proof sketch to explain the intuition behind this striking result.

Theorem 4 (Convergence: Conditional Fréchet Means)**.** *Under the Assumptions of Theorem 1,*

$$\sqrt{nh}\left(\hat{m}_{\pm,\oplus,p}(\cdot) - m_{\pm}(\cdot)\right) \rightsquigarrow c_1(q)^{-1/2}\mathbb{G}_{H\pm}(\cdot,1),$$

where $\mathbb{G}_{H\pm}$ *is the* same *zero-mean Gaussian process as in Theorem 1.*

Proof sketch. The result obtains by the fact that the Fréchet estimator is the projection of the local polynomial estimator onto the space of quantile functions $\Pi_{\mathcal{Q}}$, but this projection is only active in finite sample. The limit of the local polynomial estimator is the conditional average quantile function, which is a valid quantile function of its own right, and hence the projection simply becomes the identity function in the limit. The projection operator $\Pi_{\mathcal{Q}}$ is a metric projection onto convex sets and hence well-known to be globally Lipschitz (Bauschke et al., 2017, Prop. 4.16). Then, by a generalization of Rademacher's theorem to Banach spaces, $\Pi_{\mathcal{Q}}$ is Hadamard differentiable except on a special set that is directionally porous, hence negligible in the sense of Preiss (2014). Since the local polynomial estimator $\hat{m}_{\pm,p}$ converges to the true quantile function $m_{\pm}$ in L^2 norm (an implication of Theorem 1), it almost surely avoids this exceptional set in large samples. Thus, the Hadamard derivative of $\Pi_{\mathcal{Q}}$ evaluated at the true limit $m_{\pm}$ coincides with the identity map. Consequently, by the functional delta method, the projected estimator inherits the same asymptotic distribution as the original local polynomial estimator. □

Then, the following result again follows by a simple application of the functional delta method.

Corollary 1 (Convergence: Fréchet Treatment Effects)**.** *Under the assumptions of Theorem 1 it follows that,*

$$\sqrt{nh}\big(\hat{\tau}^{\mathrm{R3D}}_{\oplus,p}(q) - \tau^{\mathrm{R3D}}\big) \rightsquigarrow \mathbb{G}^{\mathrm{R3D}},$$

and under the additional Assumptions I3, I4

$$\sqrt{nh}\big(\hat{\tau}^{\mathrm{F3D}}_{\oplus,p}(q) - \tau^{\mathrm{F3D}}\big) \rightsquigarrow \mathbb{G}^{\mathrm{F3D}}$$

Corollary 2 (Bootstrap: Fréchet)**.** *Under the same Assumptions as Theorem 3, the estimated bootstrap processes* $\hat{\mathbb{G}}^{\mathrm{R3D}}$ *and* $\hat{\mathbb{G}}^{\mathrm{F3D}}\left(\hat{m}_{+,\oplus,p}, \hat{m}_{-,\oplus,p}, \hat{m}_{+,D,p}, \hat{m}_{-,D,p}\right)$ *deliver asymptotically valid confidence intervals for the the Fréchet estimators* $\hat{\tau}^{\mathrm{R3D}}_{\oplus,p}$, $\hat{\tau}^{\mathrm{F3D}}_{\oplus,p}$.

3.3 Extensions

3.3.1 Empirical Distribution Functions

Proposition 2 (Empirical Quantiles)**.** *Under the same respective Assumptions of Theorems 1 and 2, as well as Assumptions Q1, Q2, the estimators with empirical quantile functions,* $\bar{m}_{\pm,p}(\cdot)$*,* $\bar{m}_{\pm,\oplus,p}(\cdot)$*,* $\bar{\tau}_p^{\mathrm{R3D}}(\cdot)$*,* $\bar{\tau}_p^{\mathrm{F3D}}(\cdot)$ *converge to the same uniform limiting processes as their respective counterparts.*

Corollary 3 (Bootstrap: Empirical Quantiles)**.** *Under the same Assumptions as Theorem 3, as well as Assumptions Q1, Q2, the estimated bootstrap processes* $\hat{\mathbb{G}}^{\mathrm{R3D}}$ *and* $\hat{\mathbb{G}}^{\mathrm{F3D}}(\cdot,\cdot,\cdot,\cdot)$ *(with the appropriate conditional mean estimators plugged in) deliver asymptotically valid confidence bands for the treatment effect estimators with empirical quantile functions,* $\bar{\tau}_p^{\mathrm{R3D}}$*,* $\bar{\tau}_p^{\mathrm{F3D}}$*,* $\bar{\tau}_{\oplus,p}^{\mathrm{R3D}}$*,* $\bar{\tau}_{\oplus,p}^{\mathrm{F3D}}$*.*

4 Empirical Applications

4.1 Simulations

To evaluate the performance of the proposed local Fréchet regression estimator within a regression discontinuity design (RDD) framework, I conduct a series of Monte Carlo simulations under various data-generating processes (DGP). For a detailed description, see Section A-6. I estimate the quantile treatment effects τ^{DRD} at 100 quantiles for three estimators: 1) a local polynomial estimator for classical quantile RDDs (Qu and Yoon, 2019); 2) the simple R3D estimator in Section 2.4.1; 3) the (Fréchet) R3D estimator in Section 2.4.2. The classical estimator is corrected for bias using the approach in (Qu et al., 2024) to give it the best possible chance against the D-RD estimators. For the sample estimator, I ensure the monotonicity of the estimated quantile functions using the monotone rearrangement approach in Chernozhukov et al. (2010). Figure 3 shows the estimators' performance for a DGP that features normal distributions with normally distributed means and variances that depend on X and jump across the threshold. The estimators' bias, standard error, and mean squared error are reported for quantiles 25, 50, and 75 as well as the average of all quantiles. The simulations are run for $n = 100, 200$, and 500 with $n_i = n$ for all units i. Even in this very simple setting with normal distributions, where the distribution of distributions is itself normal (as shown in Proposition 1), the R3D estimators far outperform the

classical quantile RD estimators on all measures. In fact, for low sample size n, the bias, MSE, and standard errors of the classical estimator are so large compared to the D-RD estimator that they could not be included on the y-axis without making the figure illegible. Moreover, the bias of the classical estimator appears to diverge rather than converge. This aligns with theoretical intuition, as the classical estimator should converge to the *observed* distribution at the threshold rather than the average one. By contrast, both R3D estimator have a consistently low and convergent bias. The Fréchet estimator, however, is markedly more efficient than the simple estimator at "small" sample sizes – though $n = 100$ is certainly not unreasonably small in the applied economics literature. Similar results obtain when introducing an exponentially distributed shift in the means in Figure A-2 or exponential distributions with exponential means in Figure A-3. Interestingly, even for a data-generating process with degenerate instead of random distributions in Figure A-4, the R3D estimators outperform the classical estimator on most metrics for smaller sample sizes. This is because the R3D properly accounts for the occurence of an entire distribution of outcomes at a single value of the running variable, which violates the assumption of i.i.d. random sampling two random variables in the quantile RDD setting, even in the absence of random distributions. Contrary to the other settings, however, the bias of the classical estimator does appear to converge in the degenerate setting, as is expected theoretically.

Though the use of classical quantile estimators is justified theoretically in these settings, it appears that the R3D estimators offer a more efficient alternative for finite samples. The continuity assumption I1 of course coincides with the assumption of continuous observed quantiles in this setting.

4.2 Empirical Illustration: State Governors and the Income Distribution

To further illustrate the method, I apply it to estimate the effect of having a Republican vs. Democratic governor on the income distribution within one of the US states. This is done by way of a classical and widely used RD design in political science and political economy: the close-election design (Lee, 2008). This design compares geographic units where a political party barely won an election to to those where it barely lost. The identifying assumption is that the outcome of interest evolves

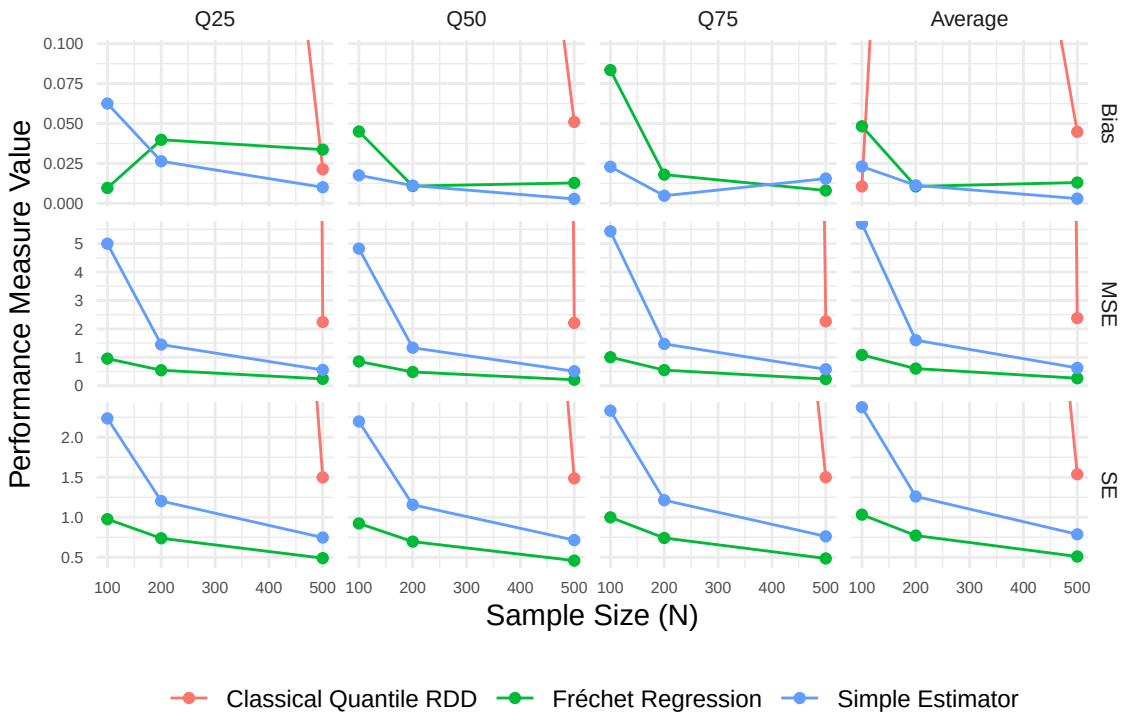


Figure 3: Monte Carlo Simulations, Normal Distributions with Normal Means

Note: figure compares performance of three quantile estimation methods in terms of the absolute bias, mean squared error (MSE), and standard error (SE). Each measure is reported for $N = 100, 200,$ and 500 (x axis) at quantiles 25, 50, 75, and the average over all quantiles (facets). The methods are: 1) a local polynomial estimator for classical quantile RDDs (Qu and Yoon, 2019) with bias correction (Qu et al., 2024); 2) the simple R3D estimator in Section 2.4.1; 3) the (Fréchet) R3D estimator in Section 2.4.2. Data-generating process: outcome variable Y is a normal distribution with normally distributed means and variances that depend on running variable X and jump across the threshold. For details, see Section A-6.

smoothly with the party's vote share, until it crosses the 50% electoral threshold that puts the party in power. Under that assumption, any jump observed in the outcome at the threshold is induced by the party's electoral win, and thus identifies its causal effect, locally for states with near 50% party vote share. Here, I illustrate how such a close-election design naturally leads to an R3D setting, since the outcome of interest is usually measured at the constituent level, leading to an entire distribution of outcomes within each constituency.

4.2.1 Data and Method

I use data on gubernatorial election outcomes from Congressional Quarterly's Voting and Elections Collection, collating election data from 1984 to 2010. Cutting the sample off in 2010 ensures a stable and clearly defined environment for estimating gubernatorial impacts on state-level income distributions. The year 2010 marked a structural breakpoint in state politics (see e.g. the sharp increase in state-level polarization documented in) due to the significant Republican gains from the Tea Party wave and the subsequent implementation of the Affordable Care Act (ACA). The ACA introduced confounding by influencing state policy choices through federal incentives, while increased partisan polarization changed the nature and meaning of gubernatorial party control itself. Restricting the analysis to pre-2010 thus guarantees a stable treatment definition, ensuring clearer identification of causal effects attributable specifically to Democratic versus Republican gubernatorial control. Indeed, while the magnitude of the effects remains similar when including post-2010 data, their precision decreases.

I combine these data with family-level income data from the UNICON extract of the March Current Population Survey (CPS) for the same time period.

The outcome variable (Y_{it}) that I consider is the *distribution* of equivalized real family income in state i in election year t. Real family income is constructed as the ratio of family income in year t divided by the federal poverty threshold in that year, where family income is the combined pre-tax cash income of the family, including earnings and cash transfers but excluding non-cash benefits or tax credits. The federal poverty threshold is adjusted yearly and depends on family size and the number of children.

The running variable X_{it-} is the Democratic candidate's votes as a share of the combined Democratic and Republican candidates' total votes. When this threshold

exceeds 50%, the Democratic candidate is elected. As such, the treatment T_{it} is whether a

4.2.2 Results

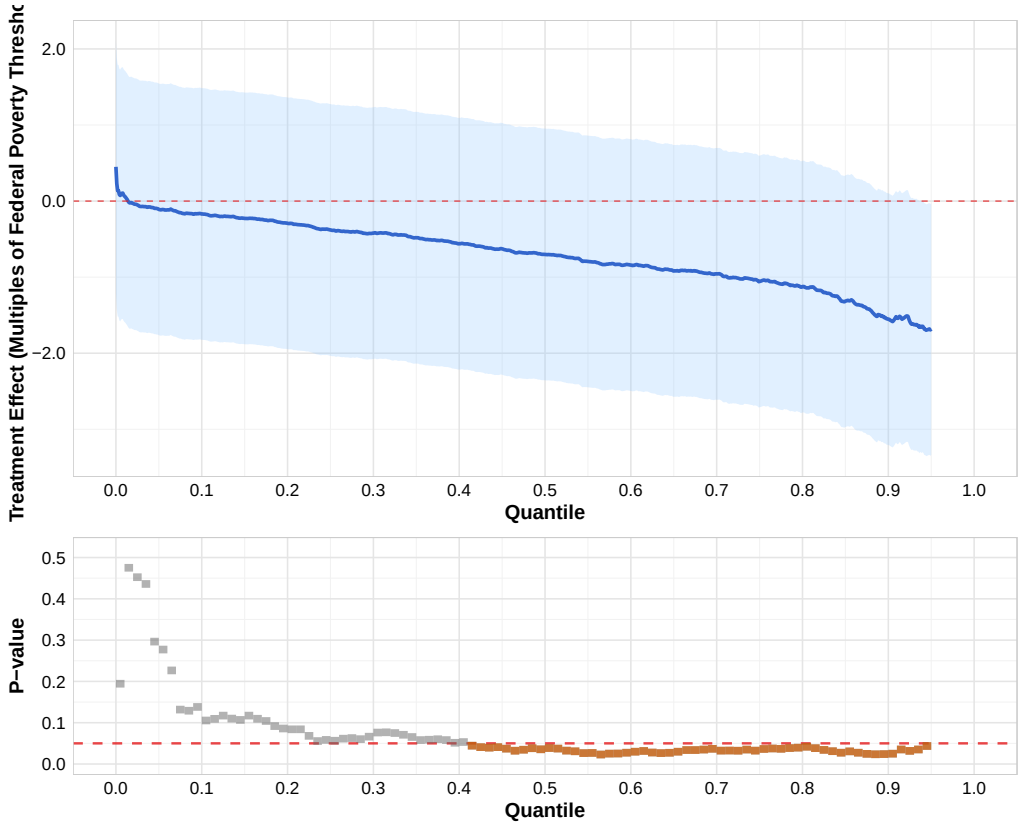


Figure 4: Distributional Effects of Democratic Governor Control, 1984-2010

Note: **Top**: local average quantile treatment effects estimates and uniform 95% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X axis indicates quantile of the (average) income distribution while Y axis indicates the difference in average state-level income distributions near the 50% vote share threshold. Income is measured as real equivalized family income in multiples of the federal poverty threshold. Sample runs from 1984–2010, estimates are obtained using the second-order Fréchet estimator in Section 2.4 with first-order IMSE-optimal bandwidth and triangular kernel as in Section A-4.2, and uniform bands are constructed using Algorithm A-4.3 with 5,000 bootstrap repetitions. **Bottom**: p-values for treatment nullity tests at each of 95 equally spaced segments of the distribution between the 0th and 95th quantile.

5 Conclusion

References

Agueh, M. and Carlier, G. (2011), 'Barycenters in the wasserstein space', *SIAM Journal on Mathematical Analysis* **43**(2), 904–924.

Bauschke, H. H., Combettes, P. L., Bauschke, H. H. and Combettes, P. L. (2017), *Convex analysis and monotone operator theory in Hilbert spaces*, Springer.

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

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

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014), 'Robust nonparametric confidence intervals for regression-discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Card, D. and Krueger, A. B. (2000), 'Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: reply', *American Economic Review* **90**(5), 1397–1420.

Chen, Y., Lin, Z. and Müller, H.-G. (2023), 'Wasserstein regression', *Journal of the American Statistical Association* **118**(542), 869–882.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010), 'Quantile and probability curves without crossing', *Econometrica* **78**(3), 1093–1125.

Chiang, H. D., Hsu, Y.-C. and Sasaki, Y. (2019), 'Robust uniform inference for quantile treatment effects in regression discontinuity designs', *Journal of Econometrics* **211**(2), 589–618.

Clark, D. (2009), 'The performance and competitive effects of school autonomy', *Journal of political Economy* **117**(4), 745–783.

Eggers, A. C., Fowler, A., Hainmueller, J., Hall, A. B. and Snyder Jr, J. M. (2015), 'On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races', *American Journal of Political Science* **59**(1), 259–274.

Fan, J. and Gijbels, I. (1992), 'Variable bandwidth and local linear regression smoothers', *The Annals of Statistics* pp. 2008–2036.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chap-

man & Hall, London.

Fan, J. and Müller, H.-G. (2024), 'Conditional wasserstein barycenters and interpolation/extrapolation of distributions', *IEEE Transactions on Information Theory* .

Frandsen, B. R., Frölich, M. and Melly, B. (2012), 'Quantile treatment effects in the regression discontinuity design', *Journal of Econometrics* **168**(2), 382–395.

Gelman, A. and Imbens, G. (2019), 'Why high-order polynomials should not be used in regression discontinuity designs', *Journal of Business & Economic Statistics* **37**(3), 447–456.

Giné, E. and Nickl, R. (2021), *Mathematical foundations of infinite-dimensional statistical models*, Cambridge university press.

Hahn, J., Todd, P. and Van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression-discontinuity design', *Econometrica* **69**(1), 201–209.

Lee, D. S. (2008), 'Randomized experiments from non-random selection in us house elections', *Journal of Econometrics* **142**(2), 675–697.

Ludwig, J. and Miller, D. L. (2007), 'Does head start improve children's life chances? evidence from a regression discontinuity design', *The Quarterly journal of economics* **122**(1), 159–208.

Panaretos, V. M. and Zemel, Y. (2020), *An invitation to statistics in Wasserstein space*, Springer Nature.

Pei, Z., Lee, D. S., Card, D. and Weber, A. (2022), 'Local polynomial order in regression discontinuity designs', *Journal of Business & Economic Statistics* **40**(3), 1259–1267.

Petersen, A. and Müller, H.-G. (2019), 'Fréchet regression for random objects with euclidean predictors', *The Annals of Statistics* **47**(2), 691–719.

Preiss, D. (2014), 'Gâteaux differentiability of cone-monotone and pointwise lipschitz functions', *Israel Journal of Mathematics* **203**(1), 501–534.

Qu, Z. and Yoon, J. (2019), 'Uniform inference on quantile effects under sharp regression discontinuity designs', *Journal of Business & Economic Statistics* **37**(4), 625–647.

Qu, Z., Yoon, J. and Perron, P. (2024), 'Inference on conditional quantile processes in partially linear models with applications to the impact of unemployment benefits', *Review of Economics and Statistics* **106**(2), 521–541.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, John Wiley & Sons, New York.

Silverman, B. W. (2018), *Density estimation for statistics and data analysis*, Routledge.

van der Vaart, A. W. (2000), *Asymptotic statistics*, Vol. 3, Cambridge university press.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York.
URL: *https://link.springer.com/book/10.1007/978-1-4757-2545-2*

Villani, C. (2021), *Topics in optimal transportation*, Vol. 58, American Mathematical Soc.

Zhou, Y. and Müller, H.-G. (2024), 'Wasserstein regression with empirical measures and density estimation for sparse data', *Biometrics* **80**(4), ujae127.

A-1 Mathematical Notation and Definitions

Definition A-3 (VC Type (Def. 3.6.10, Giné and Nickl (2021))). *A class of measurable functions $\mathcal{F}$ is of VC type with respect to a measurable envelope F of $\mathcal{F}$ if there exist finite constants A, v such that for all probability measures Q on $(\Omega^x, \mathcal{F}^x)$*

$$N\left(\mathcal{F}, L^2(Q), \varepsilon \|F\|_{L^2(Q)}\right) \leq (A/\varepsilon)^v.$$

A-2 Derivation of Local Polynomial Regression Weights

We want to estimate

$$\hat{m}_{\pm,p}(q) \;=\; \Big(\text{polynomial fit at } x = c\Big)\Big|_{\text{order}=p}$$

via the following one-sided weighted least squares:

$$\hat{\boldsymbol{\alpha}}_{\pm,p} \;=\; \underset{\boldsymbol{\alpha}\,\in\,\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} 1\{X_i \;\geq\; c\} \;\; K\Big(\tfrac{X_i - c}{h}\Big) \Big[Q_{Y_i}(q) \;-\; \boldsymbol{\alpha}^\top \, r_p\Big(\tfrac{X_i - c}{h}\Big)\Big]^2,$$

where $r_p(u) := (\,1,\; u,\; u^2, \ldots, u^p\,)^\top$.

Define:

$$\mathbf{X}_\pm \;=\; \begin{pmatrix} r_p\big(\frac{X_1 - c}{h}\big)^\top \\ r_p\big(\frac{X_2 - c}{h}\big)^\top \\ \vdots \\ r_p\big(\frac{X_n - c}{h}\big)^\top \end{pmatrix}, \quad \mathbf{W}_\pm \;=\; \operatorname{diag}\Big\{\delta_i^\pm \, K\big(\tfrac{X_i - c}{h}\big) : i = 1, \ldots, n\Big\},$$

$$\mathbf{Q} \;=\; \big(Q_{Y_1}(q),\; Q_{Y_2}(q), \ldots, Q_{Y_n}(q)\big)^\top.$$

The solution to the above least-squares problem is:

$$\hat{\boldsymbol{\alpha}}_{\pm,p} \;=\; \Big(\mathbf{X}_\pm^\top \, \mathbf{W}_\pm \, \mathbf{X}_\pm\Big)^{-1} \, \big(\mathbf{X}_\pm^\top \, \mathbf{W}_\pm \, \mathbf{Q}\big).$$

Since our regression function at $X = 0$ is the *intercept* component, let $\mathbf{e}_0 = (1,\, 0,\, 0, \ldots, 0)^\top$ as before. Then

$$\hat{m}_{\pm,p}(q) \;=\; \mathbf{e}_0^\top \, \hat{\boldsymbol{\alpha}}_{\pm,p} \;=\; \mathbf{e}_0^\top \, \big(\mathbf{X}_\pm^\top \, \mathbf{W}_\pm \, \mathbf{X}_\pm\big)^{-1} \, \big(\mathbf{X}_\pm^\top \, \mathbf{W}_\pm \, \mathbf{Q}\big).$$

Noting that everything in front of $\mathbf{Q}$ is independent of $Q_{Y_i}(q)$ and depends only on $\{X_i\}$, $K_h(\cdot)$, h, etc., we see:

$$\hat{m}_{\pm,p}(q) \;=\; \sum_{i=1}^{n} \underbrace{\mathbf{e}_0^\top \left(\mathbf{X}_\pm^\top \mathbf{W}_\pm \mathbf{X}_\pm\right)^{-1} \left(\mathbf{X}_\pm^\top \mathbf{W}_\pm\right)_{:,\,i}}_{=:\; s_{\pm,\,in,p}(h)} Q_{Y_i}(q).$$

Therefore, the one-sided local-polynomial estimator of order p can be written as a simple weighted average:

$$\begin{aligned}
\hat{m}_{\pm,p}(q) \;&=\; \sum_{i=1}^{n} \Big[\, s^{(p)}_{\pm,\,in}(h) \Big]\, Q_{Y_i}(q), \quad \text{where} \\
s^{(p)}_{\pm,\,in}(h) \;&=\; \delta_i^{\pm}\, \mathbf{e}_0^\top \Big(\mathbf{X}_\pm^\top \mathbf{W}_\pm \mathbf{X}_\pm\Big)^{-1} r_p\Big(\tfrac{X_i - c}{h}\Big)\, K\Big(\tfrac{X_i - c}{h}\Big).
\end{aligned}$$

A-3 Overview of Local Fréchet Regression

This supplement provides a brief overview of local Fréchet regression as proposed in Petersen and Müller (2019) and extends their convergence results to the setting at boundary points of the support.

A-3.1 Generalized Conditional Expectations

The concept of the Fréchet mean arises as a natural generalization of the Euclidean mean. To see this, let $Y \in \mathbb{R}$, then the conditional expectation $E[Y \mid X = x]$ can be defined as the unique minimizer $f(x)$ of the mean squared error,

$$E[d_E\left(Y, f(x)\right)^2 \mid X = x],$$

where $d_E(x, y) \coloneqq \|x - y\|$ the standard Euclidean metric. The conditional Fréchet mean $m_\oplus(x)$ generalizes this to any metric space (Ω, d) equipped with a distance metric d by replacing the squared Euclidean distance with the generalized squared distance $d(Y, \cdot), Y \in \Omega$ (Petersen and Müller, 2019),

$$\omega_\oplus = \underset{\omega \in \Omega}{\operatorname{argmin}} E\left[d^2(Y, \omega)\right].$$

The corresponding Fréchet variance is defined as $V_\oplus = E\left[d^2(Y, \omega_\oplus)\right]$. In analogy to this, Petersen and Müller (2019) define the *conditional* Fréchet mean $m_\oplus(x)$ on a general metric space (Ω, d) as,

$$m_\oplus(x) \coloneqq \underset{\omega \in \Omega}{\operatorname{argmin}} M_\oplus(\omega, x), \quad M_\oplus(\cdot, x) \coloneqq E\left[d^2(Y, \cdot) \mid X = x\right]. \tag{A-1}$$

The corresponding conditional Fréchet variance $V_\oplus(x)$ is defined analogously to the classical variance operator as the squared distance from the mean,

$$V_\oplus(x) \coloneqq E\left[d^2(Y, m_\oplus(x)) \mid X = x\right].$$

The "unconditional" Fréchet mean and variance are defined analogously.

A-3.2 The 2-Wasserstein Metric

In our distributional setting, it is natural to work with the 2-Wasserstein distance $d_{W_2}(G_1, G_2)$ to measure the distance between two distribution functions $G_1, G_2 \in \mathcal{G}$. Since we are working with one-dimensional distribution functions, this metric can be shown to equal (Villani, 2021, Theorem 2.18),

$$d^2_{W_2}(G_1, G_2) = \int_0^1 \left(G_1^{-1}(t) - G_2^{-1}(t)\right)^2 \mathrm{d}t, \tag{A-2}$$

where G_1^{-1} and G_2^{-1} are the quantile functions corresponding to G_1 and G_2, respectively.

The main motivation for working with the 2-Wasserstein distance is the fact that the Fréchet mean of any random distribution $G \in \mathcal{G}$ equipped with d_{W_2} is the unique cdf $G_\oplus$ with the quantile function (Panaretos and Zemel, 2020, Theorem 3.2.11),

$$G_\oplus^{-1}(t) = EG^{-1}(t) = \int_{\mathcal{G}} G^{-1}(t)\, \mathrm{d}P(G).$$

Informally, the "average" distribution computed by means of the Fréchet mean under the 2-Wasserstein distance is the only distribution that has a quantile function equal to the expected quantile function at each quantile t. In that sense, it is the "correct" metric for computing average quantile functions.

Moreover, d_{W_2} is well-suited for testing for discontinuities in distributional pro-

cesses. In particular, for any probability measures μ, ν with finite variance, $\mu = \nu \iff d_{W_2}(G_\mu, G_\nu) = 0$ (Panaretos and Zemel, 2020, p.38).

A-3.3 Local Fréchet Regression

[David: TODO: make this from left and right] Fréchet regression was introduced in Petersen and Müller (2019) as a generalization of linear regression when the outcome Y takes values in a general metric space Ω rather than in Euclidean space $\mathbb{R}$. The prevalent approach to estimate RDDs is based on local linear or quadratic regression (Gelman and Imbens, 2019; Hahn et al., 2001). In that light, I focus on local Fréchet regression, which was proposed as a generalization of local linear regression in Euclidean spaces (Fan and Gijbels, 1996).

In the definition of the conditional Fréchet mean introduced above, consider the case $Y \in \Omega = \mathbb{R}$ and write $m = m_\oplus$ for brevity. Then the (population) local linear estimate of $m(x)$ is $\tilde{l}(x) = \beta_0^*$, where,

$$\begin{aligned}(\beta_0^*, \beta_1^*) = &\underset{\beta_0,\beta_1}{\operatorname{argmin}} \int K_h(z-x) \\ &\times \left[\int y \, \mathrm{d}F_{Y|X}(z,y) - (\beta_0 + \beta_1(z-x))\right]^2 \mathrm{d}F_X(z),\end{aligned}$$

with $K_h(\cdot) = h^{-1}K(\cdot/h)$ with K a smoothing kernel and h a bandwidth. Defining $\mu_j = E\left[K_h(X-x)(X-x)^j\right]$ and $\sigma_0^2 = \mu_0\mu_2 - \mu_1^2$, the solution β_0^* can be written as,

$$\tilde{l}(x) = \beta_0^* = E[s(X,x,h)Y],$$

with weight function,

$$s(z,x,h) = \frac{1}{\sigma_0^2}\left\{K_h(z-x)\left[\mu_2 - \mu_1(z-x)\right]\right\} \tag{A-3}$$

which corresponds to the local Fréchet mean,

$$\tilde{l}(x) = \underset{y\in\mathcal{R}}{\operatorname{argmin}} E\left[s(X,x,h)(Y-y)^2\right]. \tag{A-4}$$

Just as with the definition of the classical Fréchet mean, this can be generalized

to $Y \in \Omega$ on a general metric space as,

$$\tilde{l}_{\oplus}(x) = \underset{\omega \in \Omega}{\operatorname{argmin}} \left\{ \tilde{L}_n(\omega) \coloneqq E\left[s(X, x, h) d^2(Y, \omega) \right] \right\}$$

where the dependence on n is through the bandwidth sequence $h = h_n$.

Then, assume that $(X_i, Y_i) \sim F, i = 1, \ldots, n$ are independent. The corresponding sample estimator is,

$$\hat{l}_{\oplus}(x) = \underset{\omega \in \Omega}{\operatorname{argmin}} \left\{ \hat{L}_n(\omega, x) \coloneqq n^{-1} \sum_{i=1}^{n} s_{in}(x, h) d^2\left(Y_i, \omega\right) \right\}, \tag{A-5}$$

with the empirical weights,

$$s_{in}(x, h) = \frac{1}{\hat{\sigma}_0^2} K_h\left(X_i - x\right) \left[\hat{\mu}_2 - \hat{\mu}_1 \left(X_i - x\right) \right], \tag{A-6}$$

where

$$\hat{\mu}_j = n^{-1} \sum_{i=1}^{n} K_h\left(X_i - x\right) \left(X_i - x\right)^j, \quad \hat{\sigma}_0^2 = \hat{\mu}_0 \hat{\mu}_2 - \hat{\mu}_1^2.$$

A-3.4 Existence, Uniqueness, and Local Polynomial Equivalence

The following result establishes the equivalence of the projected local polynomial regression estimator and the local Fréchet regression estimator from Petersen and Müller (2019) in the metric space $(\mathcal{G}, d_{W_2})$.

Proposition A-3. *The projected local polynomial regression estimator in (3) is equivalent to the quantile function of the local polynomial Fréchet regression estimator of order p on the metric space $(\mathcal{G}, d_{W_2})$,*

$$\underset{\omega \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} s_{\pm, in}^{(p)}(h) d_{W_2}^2(\omega, Y_i),$$

which exists and is unique. Similarly, the projected conditional average quantile function in [David: TODO] is equivalent to the quantile function of the conditional Fréchet mean on $(\mathcal{G}, d_{W_2})$ in (A-1), which exists and is unique.

Proof. Denote $\langle \cdot, \cdot \rangle_{L^2}, \| \cdot \|_{L^2}$ and $d_{L^2}(\cdot, \cdot)$ the L^2 inner product, norm, and distance on $[0, 1]$, respectively.

From the definition of $\hat{m}_{+,p}$, we have,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) d_{L^2}(Q_{Y_i}, \hat{m}_{+,p})^2 + d_{L^2}(Q_\omega, \hat{m}_{+,p})^2 \\
&= \langle \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h), Q_{Y_i}^2 \rangle_{L^2} - 2\langle \hat{m}_{+,p}, \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) Q_{Y_i} \rangle_{L^2} + \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) \langle \hat{m}_{+,p}, \hat{m}_{+,p} \rangle_{L^2} \\
&+ \langle Q_\omega, Q_\omega \rangle_{L^2} - 2\langle \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) Q_{Y_i}, Q_\omega \rangle L^2 + \langle \hat{m}_{+,p}, \hat{m}_{+,p} \rangle_{L^2} \\
&= \langle \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h), Q_{Y_i}^2 \rangle_{L^2} - \langle Q_\omega, Q_\omega \rangle_{L^2} - 2\langle \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) Q_{Y_i}, Q_\omega \rangle_{L^2} \\
&= \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) d_{L^2}(Q_{Y_i}, Q_\omega)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) d_{W_2}(Y_i, \omega) = \text{(A-5)}
\end{aligned}
$$

where the third equality follows from $\frac{1}{n}\sum_{i=1}^{n} s_{+,in}^{(p)}(h) = 1$. As a result, the local Fréchet regression estimator $\hat{l}_{\oplus}(c)$ on $(\mathcal{G}, d_{W_2})$ equals

$$
\hat{l}_{+,\oplus}(c) = Q^{-1}\left(\operatorname*{argmin}_{h \in Q(\mathcal{G})} d_{L^2}(h, \hat{m}_{+,p})^2 \right), \tag{A-7}
$$

where Q^{-1} is the quantile function's generalized inverse, which maps it back to its corresponding CDF. Thus, the quantile function of the local Fréchet regression estimator is the L^2 projection of my local polynomial estimator onto the space of quantile functions. Further, note that the solution to (A-7) always exists and is unique by the convexity of the objective function and of the space of quantile functions $Q(\mathcal{G})$.

To see that the quantile function of the conditional Fréchet mean is equivalent to the projected conditional average quantile, rewrite the conditional Fréchet functional in (A-1) on $(\mathcal{G}, d_{W_2})$ as,

$$
\begin{aligned}
M_{\oplus}(\Omega, x) &= E[d_{W_2}^2(Y_i, \omega) \mid X = x] \\
&= \int_{\mathcal{X}} \int_0^1 \left(Q_{Y_i}(q) - Q_\omega(q)\right)^2 \mathrm{d}q \, \mathrm{d}F_{Y|X=x} \\
&= \int_0^1 \int_{\mathcal{X}} \left(Q_{Y_i}(q)^2 - 2Q_{Y_i}(q)Q_\omega(q) + Q_\omega(q)^2\right) \mathrm{d}F_{Y|X=x} \, \mathrm{d}q
\end{aligned}
$$

$$= C' + \int_0^1 \left(m(q) - Q_\omega(q)\right)^2 \mathrm{d}q$$

where $C' = \int_0^1 \int_{\mathcal{X}} Q_{Y_i}(q)^2 \,\mathrm{d}F_{Y|X=x}\,\mathrm{d}q - \int_0^1 m(q)^2\,\mathrm{d}q$ is a constant that does not depend on Q_ω, and the second equality follows from Fubini-Tonelli by the fact that all distributions in $\mathcal{G}$ have finite variance. As a result, the conditional Fréchet mean $m_\oplus$ on $(\mathcal{G}, d_{W_2})$ equals

$$m_\oplus(x) = Q^{-1}\left(\operatorname*{argmin}_{h \in Q(\mathcal{G})} d_{L^2}(h, m)\right),$$

the L^2 projection of $E[Q_Y(\cdot) \mid X = x]$ onto the space of quantile functions. But since $E[Q_Y(\cdot) \mid X = x]$ is a valid quantile function, the two functions are in fact equivalent. This is the well-known result that the Fréchet mean in 2-Wasserstein space has a quantile function equal to the average quantile function. Existence and uniqueness follow by standard properties of conditional expectations. □

A-4 Implementation Details

A-4.1 First-Stage Estimators

[David: derive uniform consistency of Fréchet first-stage estimator]

Let $t \leq p$, $t \in \mathbb{N}_+$ and denote $\delta_x^\pm \coloneqq 1\{x \underset{<}{\geq} 0\}$. The uniformly consistent first-stage estimators $\hat{\mathcal{E}}_1(y,t,x,q), \hat{\mathcal{E}}_2(y,t,x,q)$ proposed by Chiang et al. (2019, A.6) are,

$$\hat{\mathcal{E}}_1(y,t,x,q) = \Big(Q_Y(q) - \tilde{E}[Q_Y(q) \mid X = x]\Big)\, 1\left(|x/h_1(q)| \leq 1\right)$$

and

$$\hat{\mathcal{E}}_2(y,t,x,q) = \Big(T - \tilde{E}[T \mid X = x]\Big)\, 1\left(|x/h_2(q)| \leq 1\right)$$

where

$$\tilde{E}[Q_Y(q) \mid X = x] \coloneqq r_t(x/h_1(q))'\hat{\boldsymbol{\alpha}}_{+,t}\,\delta_x^+ + r_t(x/h_2(q))'\hat{\boldsymbol{\alpha}}_{-,t}\delta_x^-$$

and

$$\tilde{E}[T \mid X = x] \coloneqq r_t(x/h_2(q))'\hat{\boldsymbol{\alpha}}_{+,T,t}\,\delta_x^+ + r_t(x/h_2(q))'\hat{\boldsymbol{\alpha}}_{-,T,t}\,\delta_x^-$$

with $\hat{\boldsymbol{\alpha}}_{+,t}$ defined in (3) and similarly

$$\hat{\boldsymbol{\alpha}}_{\pm,T,p} \;=\; \underset{\boldsymbol{\alpha}\,\in\,\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} \delta_i^{\pm}\, K\Big(\tfrac{X_i}{h}\Big) \Big[T_i \;-\; \boldsymbol{\alpha}^{\top}\, r_p\Big(\tfrac{X_i}{h}\Big)\Big]^2, .$$

For the corresponding Fréchet first-stage estimates, one simply projects $\tilde{E}[Q_Y(q) \mid X = x]$ onto the space of quantile functions before evaluating it at a given q in the expression for $\hat{\mathcal{E}}_1(y,t,x,q)$.

Then, by Lemma 7 in Chiang et al. (2019), the first-stage local polynomial estimators are uniformly consistent for $\mathcal{E}_1(y,t,x,q)1\left(|x/h_1(q)| \leq 1\right)$ and $\mathcal{E}_2(y,t,x,q)1\left(|x/h_2(q)| \leq 1\right)$ on $[\underline{c},\bar{c}]\times\mathcal{G}\times\{0,1\}$. By the coerciveness of the projection onto quantile functions, the local Fréchet version of these first-stage estimates is also uniformly consistent. The benefit of these estimators is that the same p-th order local polynomial estimators can be reused for both the first and second stage, reducing computational load.

Furthermore, a standard consistent estimator for $f_X(0)$ is the kernel density estimator $\hat{f}_X(0) \coloneqq \frac{1}{nb}\sum_{i=1}^{n} K\left(X_i/b\right)$ with $b = b_n \to 0$ and $nb \to \infty$.

A-4.2 Bandwidth Selection

While Assumption K2 prescribes asymptotic bandwidth rates, in practice, researchers need to choose a bandwidth in finite sample. Here, I derive MSE-optimal bandwidths for the local polynomial estimator (analogous to (Chiang et al., 2019, Supplement F)), and integrated MSE-optimal (IMSE) bandwidths for the Fréchet regression estimator. Using the IMSE for the latter delivers a single bandwidth for the entire average quantile function, as assumed by the Fréchet estimator. To operationalize the one-step robust bias correction from Calonico et al. (2014), I need to compute the bandwidth that is optimal for the desired order of local polynomial estimation s, but then use that s-th order optimal bandwidth to estimate a p-th order local regression, with $p > s$.

I remind the reader of the following notation, $r_s(u) = \big(1, u, \ldots, u^s\big)$, $\Gamma_s^{\pm} = \int_{\mathbb{R}_\pm} K(u)\, r_s(u)\, r_s(u)'\, du$, $\Lambda_{s,s+1} = \int_{\mathbb{R}} u^{s+1}\, r_s(u)\, K(u)\, du$, $\Lambda_{s,s+1}^{\pm} = \int_{\mathbb{R}_\pm} u^{s+1}\, r_s(u)\, K(u)\, du$, and $\Psi_s^{\pm} = \int_{\mathbb{R}_\pm} r_s(u) r_s'(u) K^2(u) du$.

A-4.2.1 Local Polynomial Estimator

It is well known (cf. (Fan and Gijbels, 1992), Calonico et al. (2014)) that for a *p-th order local polynomial at a boundary*, the leading bias is on the order of h^{s+1}. Specifically:

$$\text{Bias}\big[\hat{m}_{\pm,s}(q)\big] \;=\; h_1(q)^{s+1} B_{\pm}(q) := \frac{h(q)^{s+1}}{(s+1)!}\, e_0' \left(\Gamma_s^{\pm}\right)^{-1} \Lambda_{s,s+1}^{\pm}\, \frac{\partial^{s+1} m_{\pm}(q)}{\partial x^{s+1}} + O\big(h^{s+2}\big),$$

where $e_0 = (1, 0, \ldots, 0)' \in \mathbb{R}^{s+1}$ is the row vector picking out the intercept term. Hence the bias of the difference $\hat{\tau}_s(q)$ at each q is,

$$\text{Bias}\big[\hat{\tau}_s(q)\big] \;\approx\; h_1(q)^{s+1} \left(B_+(q) - B_-(q)\right)$$

Similar derivations for the variance expressions imply that

$$\text{Var}\big[\hat{m}_{\pm,s}(q)\big] \;\approx\; \frac{1}{nh} V_{\pm}(q) := \frac{1}{(nh) f_X(0)} e_0' \big(\Gamma_s^{\pm}\big)^{-1} \Big[\sigma_{1,1}(q, q \mid 0^{\pm})\Big]\; \Psi_s^{\pm}\big((q,1),(q,1)\big) \left(\Gamma_s^{\pm}\right)^{-1} e_0,$$

where I remind the reader that $\sigma_{1,1}(q, q \mid 0^+) = \lim_{x \to 0^{\pm}} \text{Var}\big(Q_Y(q) \mid X = x\big)$. On the left side we get an analogous expression with $(\Gamma_s^-)^{-1}$ and Ψ_s^-. Summing them for $\hat{\tau}_s(q)$ (the difference) yields

$$\text{Var}\big[\hat{\tau}_s(q)\big] \;\approx\; \frac{1}{n\,h} \Big[V_+(q) + V_-(q)\Big].$$

Based on the standard bias-variance expression MSE = Bias^2 + Var, these expressions allow me to derive an MSE-optimal bandwidth $h_1^*(q)$ at each q for the local polynomial estimator by optimizing with respect to h_1, which gives,

$$h_1^*(q) = \left(\frac{1}{2(s+1)} \frac{V_+(q) - V_-(q)}{(B_+(q) - B_-(q))^2}\right)^{1/(2s+3)} n^{-1/(2s+3)}. \tag{A-8}$$

The derivation for the denominator's bandwidth in the fuzzy RDD, h_2^*, follows identically by replacing $m_{\pm}(q)$ with $m_{\pm,T}$ and $\sigma_{1,1}(q, q|0^{\pm})$ with $\sigma_{2,2}(q, q|0^{\pm}) = \lim_{x \to 0^{\pm}} \text{Var}(T|X = x)$ in the formulas above and noting that the optimal bandwidth will be the same for all q,

A-4.2.2 Fréchet Estimator

Define

$$\mathrm{IMSE}\big[\hat{\tau}_s\big] \;=\; \int_a^b \mathrm{MSE}\big[\hat{\tau}_s(q)\big]\,dq \;=\; \int_a^b \Big[\mathrm{Bias}\big[\hat{\tau}_s(q)\big]^2 \;+\; \mathrm{Var}\big[\hat{\tau}_s(q)\big]\Big]\,dq.$$

Putting the expansions above together, we see

$$\mathrm{IMSE}\big[\hat{\tau}_s\big] \;=\; h^{2(s+1)}\,A_s \;+\; \frac{1}{n\,h}\,B_s \;+\; o\big(h^{2(s+1)} + \tfrac{1}{n\,h}\big),$$

where

$$A_s \;=\; \int_a^b (B_+(q) - B_-(q))^2\,\mathrm{d}q$$

and

$$B_s \;=\; \int_a^b (V_+(q) - V_-(q))\,\mathrm{d}q.$$

Then remember the Fréchet conditional mean estimator $\hat{m}_{\pm,\oplus}$ is $\Pi_{\mathcal{Q}}[\hat{m}_{\pm,1}(\cdot)]$, the L^2-projection onto quantile functions, and the corresponding treatment effect estimator $\hat{\tau}_{\oplus}(q)$. Write the difference

$$\Delta_{\oplus,s}(q) \;=\; \hat{\tau}_{\oplus,s}(q) \;-\; \hat{\tau}_s(q) \;=\; \big[\hat{m}_{+,\oplus,s} - \hat{m}_{+,s}\big](q) \;-\; \big[\hat{m}_{-,\oplus,s} - \hat{m}_{-,s}\big](q).$$

I show in Lemma A-5 that

$$\big\|\hat{m}_{\pm,\oplus,s} - \hat{m}_{\pm,s}\big\|_{L^2([a,b])} \;=\; o_p\big((nh)^{-1/2}\big).$$

Hence the difference $\Delta_{\oplus,s}(\cdot)$ is also $o_p\big((nh)^{-1/2}\big)$ in $L^2([a,b])$.

Then write,

$$\mathrm{IMSE}\big[\hat{\tau}_{\oplus,s}\big] \;-\; \mathrm{IMSE}\big[\hat{\tau}_s\big] \;=\; \int_a^b \Big[\mathrm{MSE}\big(\hat{\tau}_{\oplus,s}(q)\big) \;-\; \mathrm{MSE}\big(\hat{\tau}_s(q)\big)\Big]\,dq.$$

But pointwise in q,

$$\mathrm{MSE}\big[\hat{\tau}_{\oplus,s}(q)\big] \;-\; \mathrm{MSE}\big[\hat{\tau}_s(q)\big] \;=\; 2\,\mathbb{E}\big[(\hat{\tau}_s - \tau)\,\Delta_{\oplus,s}\big] \;+\; \mathbb{E}\big[\Delta_{\oplus,s}^2\big].$$

By the Cauchy–Bunyakovsky–Schwarz inequality and the previously established con-

vergence results, this difference is $o_p(1/nh)$. Hence

$$\text{IMSE}\big[\hat{\tau}_{\oplus,s}\big] \;=\; \text{IMSE}\big[\hat{\tau}_s\big] \;+\; o_p((1/nh).$$

Hence, the leading terms of the IMSEs of the local polynomial and the Fréchet estimator are the same, and thus,

$$\text{IMSE}\big[\hat{\tau}_{\oplus,s}\big] \;\approx\; A_s\, h^{2(p+1)} \;+\; \frac{B_s}{n\,h}. \tag{A-9}$$

Taking a derivative in h and setting it to 0 gives the IMSE-optimal bandwidth for the sharp Fréchet RD setting,

$$h^*_{\oplus,1} \;=\; \left(\frac{B_s}{2(s+1)\,A_s}\right)^{1/(2s+3)} n^{-\frac{1}{2s+3}}.$$

For the fuzzy Fréchet RD setting, one simply uses this rate for the numerator and h^*_2, derived above, for the denominator.

Note that to obtain the estimates of the A and B terms, as explained below, I rely on standard local polynomial estimates rather than the projected Fréchet estimates. The reason is that the bias term involves the second derivative of the conditional expectation. Derivatives of quantile functions are not quantile functions themselves, and hence projecting them onto the space of quantile functions lacks meaning. This approach is justified by the above derivations, since the IMSE optimal bandwidth expression is an asymptotically valid expression and the Fr'echet and local polynomial estimators coincide asymptotically.

A-4.2.3 Practical Estimation

To estimate the "oracle" bandwidths derived above in practice, I propose the following three-step procedure:

Step 1: Preliminary Bandwidths.

(i) Estimate the density of X at zero by a kernel-density estimator using the rule of thumb of Silverman (2018):

$$\hat{f}_X(0) \;=\; \frac{1}{nc_n}\sum_{i=1}^{n} K\Big(\tfrac{X_i}{c_n}\Big) \quad \text{where} \quad c_n \;=\; 1.06\,\hat{\sigma}_X\, n^{-1/5},$$

and $\hat{\sigma}_X$ is the sample standard deviation of $\{X_i\}_{i=1}^n$.

(ii) Compute the *pilot* bandwidths $h_{k,n}^0$ for local polynomial fits of order p using the bias-variance formulas derived above,

$$h_{1,n}^0(q) \;=\; \left(\tfrac{1}{2(p+1)} \, \frac{C_{1,0}(q)'}{C_{1,0}(q)^2} \right)^{\frac{1}{2p+3}} \; n^{-\frac{1}{2p+3}}, \quad k \in \{1,2\}.$$

Where $C_{1,0}$ and $C'_{1,0}$ are the bias and variance expressions derived above with first-stage estimates plugged in,

$$C_{1,0}(q) = e_0' \left[\left(\Gamma_s^+\right)^{-1} \Lambda_{s,\,s+1}^+ \, \frac{\partial^{s+1} \overline{m}_+(q)}{\partial x^{s+1}} \;-\; \left(\Gamma_s^-\right)^{-1} \Lambda_{s,\,s+1}^- \, \frac{\partial^{s+1} \overline{m}_-(q)}{\partial x^{s+1}} \right] \Big/ \, (s+1)!,$$

$$C'_{1,0}(q) = \frac{1}{\hat{f}_X(0)} \; e_0' \left[\, \overline{\sigma}^2_{1,+} \left(\Gamma_s^+\right)^{-1} \Psi_s^+ \left(\Gamma_s^+\right)^{-1} \;+\; \overline{\sigma}^2_{1,-}(q) \left(\Gamma_s^-\right)^{-1} \Psi_s^- \left(\Gamma_s^-\right)^{-1} \right] e_0.$$

where $\frac{\partial^{s+1} \overline{m}_\pm(q)}{\partial x^{s+1}}$ and $\overline{\sigma}^2_{k,\pm}(q)$ are preliminary guesses of the $(s+1)$-th derivative term and the variance, respectively. In practice, one can obtain them by fitting a global polynomial of degree $\geq s+1$ and computing the sample variance of the first term. As suggested in Chiang et al. (2019, Supp. F), simply setting them to 1 can also deliver satisfactory performance. The pilot bandwidth for the denominator in the fuzzy RDD, $h_{2,n}^0(q)$ can be obtained entirely analogously by substituting a guess for $\frac{\partial^{s+1} \overline{m}_{\pm,T}(q)}{\partial x^{s+1}}$ and its corresponding variance.

Step 2: First-Stage Local Polynomial Fits.

Using the pilot bandwidths $\{\, h_{1,n}^0(q),\, h_{2,n}^0(q)\}$ from Step 1, run local polynomial regressions of order s at each quantile q,

$$\breve{\alpha}_{\pm,s}(q) \;=\; \underset{\alpha \in \mathbb{R}^{s+1}}{\arg\min} \; \sum_{i=1}^n \delta_i^\pm \; K\Big(\tfrac{X_i}{h_{k,n}^0} \Big) \Big[Q_{Y_i}(q) \;-\; \alpha^\top \, r_s\big(\tfrac{X_i}{h_{k,n}^0} \big) \Big]^2,$$

which gives the first-stage estimates

$$\big[\breve{m}_\pm(q), \ldots, \frac{\partial^s \breve{m}_\pm(q))}{\partial x^s} \big] = \breve{\alpha}'_{\pm,s} \, \mathrm{diag}\left[1, 1!/h_{1,n}^0, \ldots, s!/\left(h_{1,n}^0\right)^s \right]$$

and the corresponding first-stage s-th order expansion,

$$\begin{aligned}\breve{E}[Q_Y(q) \mid X = x] = &\left[\breve{m}_+(q) + \breve{m}_+^{(1)}(q)\, x + \cdots + \frac{\partial^s \breve{m}_+(q)}{\partial x^s}\,\frac{x^s}{s!}\right] \delta_x^+ + \\ &\left[\breve{m}_-(q) + \breve{m}_-^{(1)}(q)\, x + \cdots + \frac{\partial^s \breve{m}_-(q)}{\partial x^s}\,\frac{x^s}{s!}\right] \delta_x^-.,\end{aligned}$$

as well as the corresponding variance estimates,

$$\breve{\sigma}_{11}\left(q, q \mid 0^{\pm}\right) = \left(\frac{\sum_{i=1}^n \left(Q_{Y_i}(q) - \breve{E}[Q_Y(q) \mid X_i]\right)^2 K\left(\frac{X_i}{h_{1,n}^0}\right) \delta_i^{\pm}}{\sum_{i=1}^n K\left(\frac{X_i}{h_{1,n}^0}\right) \delta_i^{\pm}}\right)^{1/2},$$

and analogously for $\breve{E}[T \mid X = x]$, $\breve{\sigma}_{22}\left(q, q \mid 0^{\pm}\right)$. Then the uniform consistency of $\breve{E}[Q_Y(q) \mid X = x] 1\left\{|x| \leq h_{1,n}^0(q)\right\}$ and $\breve{E}[T \mid X = x] 1\left\{|x| \leq h_{2,n}^0\right\}$ is implied by Lemma 7 in Chiang et al. (2019), see the discussion in Appendix A-4.1.

Step 3: Final Bandwidth via MSE (or IMSE).

Finally, plug these first-stage expansions into the MSE- and IMSE-optimal bandwidth formulas derived in (A-8) and (A-9).

- *Local polynomial estimator*, $h_1^*(q)$: MSE requires a separate $\widehat{h}_{k,n}(q)$ for each quantile q.
- *Fréchet estimator*: $h_{\oplus,1}^*$: IMSE across $q \in [a, b]$ can be obtained by averaging the bias^2 and variance from Step 2 over $q \in [a, b]$ to get a single bandwidth for all q.

Finally, I apply the rule of thumb bandwidth algorithm from Calonico et al. (2018, 2020) for optimal coverage error these (I)MSE-optimal estimated bandwidths,

$$h_1^{\mathrm{ROT}}(q) = h_1^*(q) n^{-s/(2s+3)(s+3)}$$

and similarly for $h_2^{\mathrm{ROT}}(q)$.

A-4.3 Multiplier Bootstrap: Algorithm

Input:

- A sample $\{(X_i, Y_i, T_i)\}_{i=1}^n$, where $Y_i \in \mathcal{G}$ (distributional outcome), $T_i \in \{0,1\}$, and running variable $X_i \in \mathbb{R}$ with cutoff normalized to 0.
- A finite grid of quantiles q_j, $\mathcal{T}^* := (q_1, \ldots, q_M) \subset [a,b] \subset (0,1)$.
- A chosen local polynomial order p.
- A kernel function K and bandwidth $h = h_n \to 0$. For simplicity, assume a single bandwidth here, but see A-4.2 for more details on bandwidth selection.
- Number of bootstrap repetitions B and significance level $\lambda \in (0,1)$.

Remark. In practice, Y_i, Q_{Y_i} are computed using (13). If the entire population is observed, these estimates coincide with the true distribution and quantile function, otherwise the results in Section 3.3.1 apply.

Step 1: Estimate conditional means on a grid of quantiles.
For each $q_j \in \mathcal{T}^*$:

(i) Form the local polynomial estimator as

$$\hat{m}_{\pm,p}(q_j) \;=\; \sum_{i=1}^{n} s^{(p)}_{\pm,in}(h)\, Q_{Y_i}(q_j),$$

where $Q_{Y_i}(q_j)$ is the q_j-quantile of Y_i, and $s^{(p)}_{\pm,in}(h)$ are the usual local-polynomial weights for $X_i \gtrless 0$.

(ii) (Sharp RDD) Set

$$\hat{\tau}_p^{\mathrm{R3D}}(q_j) \;=\; \hat{m}_{+,p}(q_j) \;-\; \hat{m}_{-,p}(q_j).$$

(iii) (Fuzzy RDD only) Also compute $\hat{m}_{\pm,T,p} = \sum_{i=1}^n s^{(p)}_{\pm,in}(h)\, T_i$, and form

$$\hat{\tau}_p^{\mathrm{F3D}}(q_j) \;=\; \frac{\hat{m}_{+,p}(q_j) \;-\; \hat{m}_{-,p}(q_j)}{\hat{m}_{+,T,p} \;-\; \hat{m}_{-,T,p}}.$$

Optional: Fréchet estimator: Project $\hat{m}_{\pm,p}$ into the space of monotone functions through the isotonic regression:

$$\hat{m}_{\oplus,+,p} = \operatorname*{argmin}_{u_1,\ldots,u_M \in \mathbb{R}^M} \sum_{j=1}^{M} \left(\hat{m}_{\pm,p}(q_j) - u_j\right)^2$$

subject to the constraint $u_1 \leq \ldots \leq u_M$.

Then, for each $q_j \in \mathcal{T}^*$, carry out the following steps.

Step 2: Estimate residuals for first-stage weighting.

Obtain uniformly consistent first-stage estimators of the residual functions. For instance, for $k \in \{1, 2\}$ and each i,

$$\hat{\mathcal{E}}_k\big(Y_i, T_i, X_i, q_j\big) \;=\; \Big[\, g_k(Y_i, T_i, q_j) \;-\; \tilde{E}\big\{ g_k(Y, T, q_j) \;\big|\; X_i \big\}\Big]\; \mathbf{1}\Big\{ \big| X_i / h_k(q_j) \big| \leq 1 \Big\},$$

where $g_1(Y_i, \theta) = Q_{Y_i}(q_j)$, $g_2(Y_i, T_i) = T_i$, and $\tilde{E}\{\cdots | X_i\}$ is a local-polynomial fit of order $t \leq p$ that reuses the p-th order estimates computed in Step 1 (see A-4.1).

Step 3: Generate bootstrap draws.

Draw $\big\{ \{\xi_i^b\}_{i=1}^n \big\}_{b=1}^B$ i.i.d. from $N(0,1)$, independent of the data, for $b = 1, \ldots, B$. For each $\theta \in \mathcal{T}^*$ and $k \in \{1, 2\}$, compute

$$\hat{\nu}_{\xi,n}^{\pm,b}(q_j, k) \;=\; \sum_{i=1}^{n} \xi_i^b \, \frac{e_0^\top \big[\Gamma_{\pm,p}\big]^{-1} \, \hat{\mathcal{E}}_k\big(Y_i, T_i, X_i, \theta\big)\, r_p\big(X_i / h_k(q_j)\big)\, K\big(X_i / h_k(q_j)\big)\, \delta_i^{\pm}}{\sqrt{n h_k(q_j)}\; \hat{f}_X(0)},$$

where $\delta_i^{\pm} = 1\{X_i \gtrless 0\}$, $r_p(\cdot)$ is the local-polynomial basis, and $\Gamma_{\pm,p} = \int_{\mathbb{R}_\pm} K(u)\, r_p(u)\, r_p(u)^\top \, du$.

Step 4: Form the bootstrap processes.

(i) (Sharp RDD) For each b and θ:

$$\hat{\mathbb{G}}^{\mathrm{R3D},b}(q_j) \;=\; c_1(q_j)^{-\frac{1}{2}} \Big[\hat{\nu}_{\xi,n}^{+,b}(q_j, 1) \;-\; \hat{\nu}_{\xi,n}^{-,b}(q_j, 1) \Big].$$

(ii) (Fuzzy RDD) For each b and θ:

$$\hat{\mathbb{G}}^{\mathrm{F3D},b}(q_j) \;=\; \frac{\big[\hat{m}_{+,T,p} - \hat{m}_{-,T,p}\big]\, \hat{\nu}_{\xi,n}^{\Delta,b}(q_j, 1) \;-\; \big[\hat{m}_{+,p}(q_j) - \hat{m}_{-,p}(q_j)\big]\, \hat{\nu}_{\xi,n}^{\Delta,b}(q_j, 2)}{\big[\hat{m}_{+,T,p} - \hat{m}_{-,T,p}\big]^2},$$

where $\hat{\nu}_{\xi,n}^{\Delta,b}(q_j, k) \;=\; \hat{\nu}_{\xi,n}^{+,b}(q_j, k) \;-\; \hat{\nu}_{\xi,n}^{-,b}(q_j, k)$.

(iii) (Optional local-Fréchet version) In the above equations, replace $\hat{m}_{\pm,p}$ by the Fréchet estimator $\hat{m}_{\pm,\oplus,p}$ if needed.

Step 5: Compute the critical value and construct bands.

For a given significance level $\lambda \in (0,1)$, define

$$\hat{c}_n^B(a,b;\lambda) \;=\; (1-\lambda)\text{-quantile of } \left\{ \max_{\theta \in \mathcal{T}^*} \left| \hat{\mathbb{G}}^b(\theta) \right| \;:\; b = 1, \ldots, B \right\},$$

where $\hat{\mathbb{G}}^b(\theta)$ stands for either $\hat{\mathbb{G}}^{\mathrm{R3D},b}(\theta)$ or $\hat{\mathbb{G}}^{\mathrm{F3D},b}(\theta)$ depending on the design.

Then, an asymptotically valid uniform $(1-\lambda)100\%$ confidence band for $\tau^{\mathrm{R3D}}(\theta)$ (sharp) or $\tau^{\mathrm{F3D}}(\theta)$ (fuzzy) on $\theta \in [a,b]$ is given by:

$$\left[\hat{\tau}_p(\theta) \;\pm\; \tfrac{1}{\sqrt{nh}}\, \hat{c}_n^B(a,b;\lambda) \right], \quad \text{for } \theta \in \mathcal{T}^*.$$

A-4.4 Computational Details

An R implementation of the package can be found at https://github.com/Davidvandijcke/r3d. The main polynomial weights estimation was implemented with a Fortran backend, leading to highly performant code, as illustrated in Figure A-1. The computational complexity scales linearly with the number of observations, the number of draws for the empirical distributions, and the number of bootstrap replications. The Fréchet estimator solves faster for increasing bootstrap repetitions than the local polynomial one, likely because the former only uses one single bandwidth for all quantiles. The package also includes the option to parallelize the bootstrap for further speed improvements with large datasets.

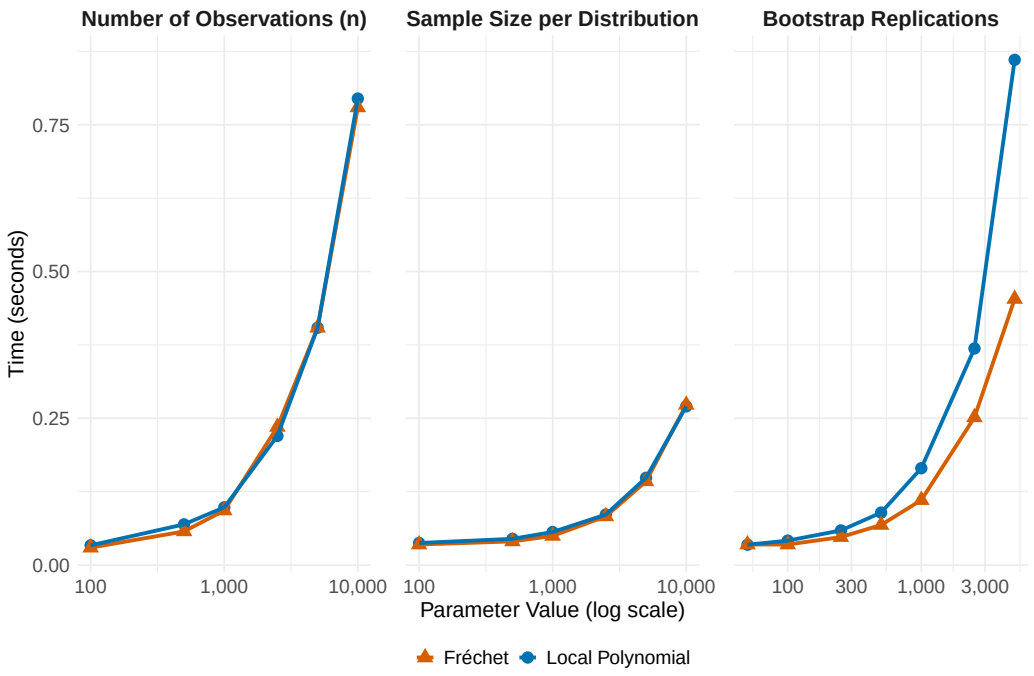


Figure A-1: Speed Benchmarks for R3D Package

Note: plots indicate seconds taken to estimate the model on an Apple M1 Pro computer with 16GB RAM, for various data and bootstrap sizes. The base model used for all computations, unless indicated otherwise, had $n = 500$ with 500 samples per distribution, 100 bootstrap repetitions, and the quantile function evaluated at 20 quantiles.

A-5 Proofs

A-5.1 Identification Results

Proof of Lemma 1.

Proof. We have,

$$\lim_{x\to 0^+} E[Q_Y(q) \mid X = x] = \lim_{x\to 0^+} E[Q_{Y(1)}(q) \mid X = x] = E[Q_{Y(1)}(q) \mid X = 0]$$

and similarly for $x \to 0^-$. The first equality follows from the definition of Y in terms of potential outcomes and the second from I1 and I2. The result then follows from taking differences and using the linearity of the expectations operator. □

Proof of Lemma 2.

Proof. We have,

$$\begin{aligned}
&\lim_{x\to 0^+} E[T \mid X = 0] - \lim_{x\to 0^-} E[T \mid X = 0] \\
&\quad = E\left[T^1 \mid X = 0\right] - E\left[T^0 \mid X = 0\right] \\
&\quad = E\left[T^1 - T^0 \mid X = 0\right] \\
&\quad = \Pr\left(T^1 > T^0 \mid X = 0\right) = \Pr\left(\mathrm{C} \mid X = 0\right)
\end{aligned}$$

where the first equality follows from the definition of T^1 and T^0, the continuity assumption I3 and Assumption I4. The third equality follows from the law of total expectation and Assumption I4. Again, by Assumptions I1 and I4,

$$\lim_{x\to 0^+} \mathbb{E}\left[Q_Y(q) \mid X = x\right] = \mathbb{E}\left[Q_{Y(1)}(q)\mathbf{1}\{\mathrm{C}\} + Q_{Y(\omega)}(q)\mathbf{1}\{\text{not C}\} \mid X = 0\right].$$

and similarly for $x \to 0^-$. As a result,

$$\begin{aligned}
&\lim_{x\to 0^+} \mathbb{E}\left[Q_Y(q) \mid X = x\right] - \lim_{x\to 0^-} \mathbb{E}\left[Q_Y(q) \mid X = x\right] \\
&= \mathbb{E}\left[Q_{Y(1)}(q) - Q_{Y(0)}(q) \mid \mathrm{C},\, X = 0\right] \times \Pr\left(\mathrm{C} \mid X = 0\right).
\end{aligned}$$

Combining these two derivations gives the result. □

A-5.2 Asymptotic Results

Lemma A-3 (Quantile functionals are VC type)**.** *Proof.* Let $T = [a, b]$ a compact subset of $[0, 1]$ or $[0, 1]$ in the case where all cdfs in $\mathcal{G}$ have compact support. For each $q \in T$ and $Y \in \mathcal{G}$, define

$$f_q(Y) \;=\; Q_Y(q) \;=\; \inf\{\, x : Y(x) \geq q\},$$

Then the family $\mathcal{F} = \{\, Y \mapsto Q_Y(q) \colon q \in T\}$ is a VC-subgraph class with index $V(\mathcal{F}) = 2$ (in the sense of (van der Vaart and Wellner, 1996, §2.6.2)). Indeed, the subgraph of f_q is

$$G_{f_q} \;=\; \big\{\, (Y, x) \colon x \leq Q_Y(q)\big\} \;=\; \big\{\, (Y, x) \colon Y(x) \;\geq\; q\big\},$$

so membership in G_{f_q} depends only on the scalar value $Y(x)$. A single point (Y, x) can be shattered: include it by setting $q \leq Y(x)$, exclude it by setting $q > Y(x)$. However, 2 points cannot be shattered. To see this, let (Y_1, x_1) and (Y_2, x_2) have $Y_1\,(x_1) \leq Y_2\,(x_2)$. The subsets $\emptyset, \{(Y_2, x_2)\}$, and $\{(Y_1, x_1)\,, (Y_2, x_2)\}$ can be realized by choosing $q > Y_2\,(x_2)\,, q \in (Y_1\,(x_1)\,, Y_2\,(x_2)]$, and $q \leq Y_1\,(x_1)$, respectively. However, the subset $\{(Y_1, x_1)\}$ cannot be realized, due to the monotonicity of cumulative distribution functions. Hence, the largest shattered set has size 1, which implies $V(\mathcal{F}) = 2$.

Furthermore, let $F(Y) = \sup_{q \in T} |Q_Y(q)|$. Then F is a measurable envelope in $L^r(Q)$ for any probability measure Q on $(\Omega^x, \mathcal{F}^x)$ and any $r \geq 1$, since each $Q_Y(t)$ is bounded by the support of Y. By Theorem 2.6.7 of van der Vaart and Wellner (1996), there is a universal $K > 0$ such that for all $0 < \varepsilon < 1$,

$$N\big(\varepsilon \|F\|_{Q,r},\, \mathcal{F},\, L_r(Q)\big) \;\leq\; K\, V(\mathcal{F})\, (16e)^{V(\mathcal{F})} \left(\tfrac{1}{\varepsilon}\right)^{r\,[V(\mathcal{F})-1]},$$

and plugging in $V(\mathcal{F}) = 2$ gives a polynomial bound in $(1/\varepsilon)^r$. □

Lemma A-4 (Conditional Expected Quantile Functions are VC type)**.** *Proof.* Let Y be a random distribution (with finite second moment), and for each $q \in [0, 1]$ define the real-valued function

$$f_q(x) \;=\; \mathbb{E}\big[\, Q_Y(q) \;\mid\; X = x\big].$$

Denote this family by

$$\mathcal{F} \;=\; \big\{\, x \mapsto f_q(x) : q \in [0,1] \big\}.$$

I claim $\mathcal{F}$ is a VC-subgraph class of finite index. Indeed, by the results in Proposition A-3, each $f_q(\cdot)$ can be identified with a one-dimensional quantile function: specifically, there is a conditional Fréchet mean $m_\oplus(x) \in \mathcal{G}$, as defined in (A-1), such that

$$f_q(x) \;=\; Q_{\,m_\oplus(x)}(q),$$

where $Q_{\,m_\oplus(x)}$ is the quantile function of $m_\oplus(x)$. In other words, for each q, the subgraph of f_q can be written as

$$G_{f_q} \;=\; \big\{\,(x,z) : z \le Q_{m_\oplus(x)}(q)\big\} \;=\; \big\{\,(x,z) : m_\oplus(x)(z) \ge q\big\}$$

Since each $x \in \mathbb{R}$ defines a distinct, unique cdf $m_\oplus(x) \in \mathcal{G}$ by Proposition A-3, the conclusion follows by an identical argument as in Lemma A-3. □

Remark In the proofs that follow, I apply the results from Chiang et al. (2019) to my R3D setting. For ease of comparison, note that their $\mu_1(x,\theta_1) = E[Q_Y(\theta_1) \mid X = x]$, $\mu_2(x,\theta_2) = 1$ in my sharp RD setting, and $\mu_2(x,\theta_2) = E[T \mid X = x]$ in the fuzzy RD setting. Further, the specific instances of their class of Wald estimands (Chiang et al., 2019, Eq. 4.1) I consider are the sharp R3D (5) and the F3D estimator (10) so that in both cases, their functions Υ, ψ, ϕ are all equal to the identity operator. The rest of their notation is closely followed for ease of comparison.

Proof of Theorem 1

Proof. The result follows by an application of Theorem 1 in Chiang et al. (2019), which holds for any random object Y as long as their assumptions are satisfied (despite the fact that the authors call the random element (Y,T,X) a "random vector"). To that end, I need to verify Assumptions 1 and 2 in that paper. I restate them in my notation for clarity.

Assumption 1, Chiang et al., 2019. *Let $\underline{c} < 0 < \bar{c}$. (i) (a) This part is equivalent to Assumption L1-(i) (b) This part is equivalent to Assumption I2. (ii) (a) the collections of real-valued functions $\{x \to E[Q_Y(q) \mid X = x] : q \in [a,b]\}$, $\{Y \to Q_Y(q) : q \in [a,b]\}$ are of VC type with common integrable envelope $F_{\mathcal{E}}$ such*

that $\int_{\mathcal{G}\times[\underline{c},\bar{c}]} |F_{\mathcal{E}}(y,x)|^{2+\epsilon}\, d\mathrm{P}^x(y,x) < \infty$ *for some* $\epsilon > 0$. *(b) This part is equivalent to Assumption L2 (i). (c) For any* $(q,k),(q',l) \in [a,b]\times\{1,2\}$, *we have* $\sigma_{kl}(q,q'|\cdot) \in C^1([\underline{c},\bar{c}]\setminus\{0\})$ *with bounded derivatives in* x *and* $\sigma_{kl}(q,q' \mid 0^{\pm}) < \infty$; *(d) for each* $y \in \mathcal{G}$, $Q_Y(q)$ *is left- or right-continuous in* q. *(iii) This part is equivalent to Assumption K2. (iv) (a)* $K : [-1,1] \to \mathbb{R}^+$ *is bounded and continuous, (b)* $\{K(\cdot/h) : h > 0\}$ *is of VC type, (c)* $\Gamma_{\pm,p}$ *is positive definite.*

- *(ii) (a)* The fact that the function classes are of VC type is proved in Lemmas A-3 and A-4. A common integrable envelope can be constructed as follows. Define $F_1(y,x) = \sup_{q\in[a,b]} |Q_Y(q)|$ and $F_2(y,x) = \sup_{q\in[a,b]} |E[Q_Y(q) \mid X = x]|$ and define $F_\varepsilon(y,x) \coloneqq F_1(y,x) + F_2(y,x)$. Then clearly $\sup_{q\in[a,b]} |Q_Y(q)| \le F_{\mathcal{E}}(y,x)$ and $\sup_{q\in[a,b]} |E[Q_Y(q) \mid X = x]| \le F_{\mathcal{E}}(y,x)$. Moreover, by Assumption L3,

$$\begin{aligned}&\int_{[\underline{c},\bar{c}]\times\mathcal{G}} (F_{\mathcal{E}}(x,y))^{2+\varepsilon}\, \mathrm{d}P^x(y,x)\\ &\le 2^{1+\varepsilon} \int_{[\underline{c},\bar{c}]\times\mathcal{G}} \left(F_1(y,x)^{2+\varepsilon} + F_2(y,x)^{2+\varepsilon}\right) \mathrm{d}P^x(y,x) < \infty.\end{aligned}$$

- *(ii) (c)* The covariance

$$\begin{aligned}&\sigma_{12}(q,q' \mid X = x) = E[(Q_Y(q) - E[Q_Y(q) \mid X = x])(T - E[T \mid X = x]) \mid X = x]\\ &= P(T = 1 \mid X = x)E[Q_Y(q) \mid X = x] - P(T = 1 \mid X = x)E[Q_Y(q) \mid X = x]\\ &+ E[T \mid X = x]E[Q_Y(q) \mid X = x] - E[T \mid X = x]E[Q_Y(q) \mid X = x] = 0\end{aligned}$$

  where the second equality follows from the law of total expectation. The variance term $\sigma_{22}(q,q' \mid X = x) = \mathrm{var}(T|X = x)$ is C^1 by Assumption L2 (i). Finally, the cross-variance term $\sigma_{11}(q,q' \mid X = x)$

$$= E[(Q_Y(q) - E[Q_Y(q) \mid X = x])(Q_Y(q') - E[Q_Y(q') \mid X = x]) \mid X = x].$$

  Expand the brackets and note that $E[Q_Y(q)Q_Y(q') \mid X = x]$ satisfies the assumption by Assumption L2 (ii) and the three other terms do so by Assumption L2 (i).

- *(ii) (d)* follows by the left-continuity of quantile functions.

- *(iv) (a)* Follows from Assumption K1 where we can always normalize K to have

bounded support on $[-1,1]$ without loss of generality.

- *(iv) (b)* To show that the function class $\{K(\cdot,/h) : h > 0\}$ is of VC type, consider the class of level sets $\{\{x : K(x/h) > t\} : h > 0, t \in \mathbb{R}\}$. For any $h > 0$ and $t \in \mathbb{R}$, the set $\{x : K(x/h) > t\}$ is an interval in $\mathbb{R}$. The class of intervals in $\mathbb{R}$ has a VC dimension of 2, which is finite. Hence, the function class $\{K(\cdot,/h) : h > 0\}$ is of VC type.

- *(iv) (c)* Follows by the non-negativeness of K in Assumption K1.

Under this set of assumptions, Chiang et al. (2019) showed in their Lemma 1 that there exists a uniform Bahadur representation,

$$\begin{aligned}
&\sqrt{nh_1(q)}\left(\hat{m}_{\pm,p}(q) - m_{\pm}(q) - h_1(q)^{p+1}(q)\,\frac{e_0'\left(\Gamma_{\pm,p}\right)^{-1}\Lambda^{\pm}_{p,p+1}}{(p+1)!}\lim_{x\to 0^{\pm}}\frac{\partial m(q)^{p+1}}{\partial x^{p+1}}\right)\\
&=\sum_{i=1}^{n}\frac{e_0'\left(\Gamma_p^{\pm}\right)^{-1}\mathcal{E}_1(Y_i,t_i,X_i,\theta)\,r_p\left(\frac{X_i}{h_1(q)}\right)K\left(\frac{X_i}{h_1(q)}\right)\delta_i^{\pm}}{\sqrt{nh_1(q)}f_X(0)}+o_p^x(1)
\end{aligned}$$

uniformly for all $q \in [a,b]$ and each $k \in \{1,2\}$. An analogous expression obtains for $\hat{m}_{+,T,p}(q) - m_{+,T}(q)$. Note that this Bahadur representation is for the debiased estimator where the bias is of order $O(h^{p+1})$. Then, by the proof of Theorem 1 in Chiang et al. (2019), $\nu_n^+(q,k) = \sum_{i=1}^n [f_{ni}(q,k) - Ef_{ni}(q,k)]$ converges weakly to a tight zero-mean Gaussian process $\mathbb{G}_{H^+}$ with covariance function H^+ defined in the main theorem, where,

$$f_{ni}(q,k) = \frac{e_0'\left(\Gamma_{+,p}\right)^{-1}r_p\left(\frac{X_i}{h_k(q)}\right)}{\sqrt{nh_k(q)}f_X(0)}\mathcal{E}_k(Y_i,T_i,X_i,q)\,K\left(\frac{X_i}{h_k(q)}\right)\delta_i^{+}$$

and similarly for $\nu_n^-(q,k)$. Then Slutksy's theorem and Assumption 1, Chiang et al., 2019 (iv) give the result. □

Proof of Theorem 2

Proof. The result for the sharp RD estimator in (5) simply follows from Theorem 1 and the continuous mapping theorem, and similarly for the denominator of the fuzzy

RDD estimator. Then, the ratio map

$$(f_+, g_+) \mapsto [f_+ - f_-] / [g_+ - g_-]$$

is Hadamard differentiable tangentially to $\ell^\infty[a,b]$ on the subset where $g_+ - g_- \neq 0$, which holds by I4 (Chiang et al., 2019, Lemma 3). Then the functional delta method yields the result (van der Vaart and Wellner, 1996, Lemma 3.9.3). □

Proof of Theorem 3

Proof. The result follows from Theorem 1 and Theorem 2 in Chiang et al. (2019). The latter applies because their Assumptions 1–4 are satisfied in my setting. Their Assumption 1 was shown to hold in the proof of Theorem 1. Moreover, their Assumption 2 is satisfied since their operators ψ, ϕ, Υ are trivially Hadamard differentiable in my setting, and by Assumptions I4 and K2. Further, their Assumption 3 is equivalent to Assumption M1. Finally, their Lemma 7 implies that the first-stage estimators $\hat{\mathcal{E}}_k(y,t,x,q)$ they propose are uniformly consistent for the population quantities $\mathcal{E}_k(y,t,x,q)$ on the kernel support $|X_i/h_k(q)| \leq 1$. Furthermore, I have assumed that $\hat{f}_X(0)$ is a consistent estimator of $f_X(0)$. As a result their Assumption 4 is also satisfied, so their Theorem 2 follows, and the result obtains by combining it with Theorem 1 by plugging in $\hat{m}_{\pm,p}(\cdot), \hat{m}_{D,\pm,p}(\cdot)$ for $m_\pm(\cdot), m_{D,\pm}(\cdot)$. □

Proof of Theorem 4

Proof. From Theorem 1 we already know that

$$\sqrt{nh}\big[\hat{m}_{\pm,p}(q) - m_\pm(q)\big] \rightsquigarrow \mathbb{G}_\pm(q) \quad \text{in } \ell^\infty([0,1]), \tag{A-10}$$

where $\hat{m}_{\pm,p}(\cdot)$ are the unprojected local-polynomial estimators and $\mathbb{G}_{H^\pm}(\cdot)$ is a tight zero-mean Gaussian limit process. Moreover, the standard ordering of L^p spaces implies that $\hat{m}_{\pm,p} - m_\pm$ also converges in L^2 norm on $[0,1]$. In particular,

$$\|\hat{m}_{\pm,p} - m_\pm\|_{L^2} = O_p\big((n\,h)^{-\frac{1}{2}}\big).$$

I then form the *projected estimators*:

$$\hat{m}_{\pm,\oplus,p}(q) := \big[\Pi_{\mathcal{Q}}(\hat{m}_{\pm,p})\big](q), \qquad m_\pm(q) := \big[\Pi_{\mathcal{Q}}(m_\pm)\big](q),$$

where $\Pi_{\mathcal{Q}}$ is the *metric projection* onto the closed, convex cone $\mathcal{Q} \subset L^2$ of quantile functions. Since $m_{\pm}(q)$ themselves are already valid quantile functions, they are unchanged by the projection:

$$m_{\pm}^{\oplus} = \Pi_{\mathcal{Q}}(m_{\pm}) = m_{\pm}.$$

Hence

$$\hat{m}_{\pm,\oplus,p} - m_{\pm} = \Pi_{\mathcal{Q}}(\hat{m}_{\pm,p}) - \Pi_{\mathcal{Q}}(m_{\pm}).$$

The goal is to show that

$$\sqrt{nh}\left[\hat{m}_{\pm,\oplus,p} - m_{\pm}\right] \rightsquigarrow \text{the } \textit{same} \text{ limit laws as in (A-10).}$$

Equivalently, we want to check that $\Pi_{\mathcal{Q}}$ has a suitable Hadamard derivative at $m_{\pm}$ which acts as the identity map.

Step 1. $\Pi_{\mathcal{Q}}$ *is 1-Lipschitz in* L^2. Since $\mathcal{Q}$ is convex and closed in a Hilbert space, the map

$$\Pi_{\mathcal{Q}} \colon L^2 \to \mathcal{Q} \subset L^2$$

is the usual orthogonal (metric) projection. By Proposition 4.16 in Bauschke et al. (2017), it is *nonexpansive*, i.e.

$$\|\Pi_{\mathcal{Q}}(g_1) - \Pi_{\mathcal{Q}}(g_2)\|_{L^2} \leq \|g_1 - g_2\|_{L^2}, \quad \forall\, g_1, g_2 \in L^2.$$

Hence $\Pi_{\mathcal{Q}}$ is globally Lipschitz with constant 1.

Step 2. Almost-everywhere Hadamard differentiability (Preiss, 2014).

In finite dimensions, Rademacher's theorem implies that a Lipschitz function is almost-everywhere differentiable (w.r.t. Lebesgue measure). In the infinite-dimensional space L^2, there is no Lebesgue measure. Instead, Preiss (2014) shows that any Lipschitz map $T \colon X \to Y$ between a separable Banach spaces X and a Banach space Y with the Radon-Nikodym property admits Hadamard derivatives at all but a "small" set N, where N belongs to the σ-ideal of directionally porous sets. (Equivalently, $X \setminus N$ is a *full measure* set in the sense that N is σ-directionally porous.)

Hence, since $\Pi_{\mathcal{Q}} \colon L^2 \to L^2$ is Lipschitz and L^2 is separable while L^2 also has the Radon–Nikodym property (being a Hilbert space), there exists an exceptional set $\mathcal{N} \subset L^2$ such that for each $x \in L^2 \setminus \mathcal{N}$, $\Pi_{\mathcal{Q}}$ is Hadamard differentiable at x.

Since $\hat{m}_{\pm,p} \to m_\pm$ in L^2 norm with probability 1, eventually $(\hat{m}_{\pm,p} - m_\pm)$ remains in $L^2 \backslash \mathcal{N}$. Hence with probability 1, for large n, $\Pi_{\mathcal{Q}}$ is Hadamard differentiable almost surely at $\hat{m}_{\pm,p}$ (and in the relevant directions $\hat{m}_{\pm,p} - m_\pm$).

Step 3. The derivative at $m_\pm$ is the identity map. Since $m_\pm \in \mathcal{Q}$ (Proposition A-3), we have

$$\Pi_{\mathcal{Q}}\big(m_\pm\big) = m_\pm.$$

In a convex set, the projection acts as the identity on any $m_\pm$. Thus $D\,\Pi_{\mathcal{Q}}[m_\pm] = \mathrm{Id}$ on the relevant tangent space in L^2. Concretely, if $m_\pm + \varepsilon u \in \mathcal{Q}$ then $\Pi_{\mathcal{Q}}(m_\pm + \varepsilon u) - \Pi_{\mathcal{Q}}(m_\pm)$ is just εu, forcing the derivative to be the identity operator.

Step 4. Hadamard derivative expansion. Then, write

$$\hat{m}_{\pm,\oplus,p} - m_\pm = \Pi_{\mathcal{Q}}\big(\hat{m}_{\pm,p}\big) - \Pi_{\mathcal{Q}}\big(m_\pm\big) = \Pi_{\mathcal{Q}}\big(m_\pm + (\hat{m}_{\pm,p} - m_\pm)\big) - \Pi_{\mathcal{Q}}(m_\pm).$$

Since $\hat{m}_{\pm,p} - m_\pm \in L^2 \backslash \mathcal{N}$ eventually a.s., by the definition of the Hadamard derivative of $\Pi_{\mathcal{Q}}$ at $m_\pm$ (van der Vaart, 2000, §20.2),

$$\Pi_{\mathcal{Q}}(m_\pm + u) - \Pi_{\mathcal{Q}}(m_\pm) = D\Pi_{\mathcal{Q}}[m_\pm](u) + o(\|u\|_{L^2}),$$

where $\|o(\|u\|_{L^2})\|_{L^2}/\|u\|_{L^2} \to 0$ as $\|u\| \to 0$. Substituting $u = \hat{m}_{\pm,p} - m_\pm$ and recalling Step 4 implies $D\Pi_{\mathcal{Q}}[m_\pm] = \mathrm{Id}$, we get

$$\hat{m}_{\pm,\oplus,p} - m_\pm = (\hat{m}_{\pm,p} - m_\pm) + o\big(\|\hat{m}_{\pm,p} - m_\pm\|_{L^2}\big).$$

Since $\|\hat{m}_{\pm,p} - m_\pm\|_{L^2} = O_p\big((nh)^{-1/2}\big)$, the two estimators $\hat{m}_{\pm,\oplus,p}$ and $\hat{m}_{\pm,p}$ coincide to first order with respect to $\sqrt{nh}$ scaling. In other words,

$$\sqrt{nh}\big[\hat{m}_{\pm,\oplus,p} - m_\pm\big] = \sqrt{nh}\big[\hat{m}_{\pm,p} - m_\pm\big] + o_p(1),$$

so $\hat{m}_{\pm,\oplus,p}$ inherits precisely the same uniform limit law as (A-10). □

Proof of Corollary 1.

Proof. The result follows by an identical argument as the proof of Theorem 2. □

Proof of Proposition 2.

Proof. Remember the definition of the local polynomial estimator with empirical quantile functions,

$$\bar{m}_{\pm,p}(q) = \frac{1}{n}\sum_{i=1}^{n} s^{(p)}_{\pm,in}(h)\hat{Q}_{Y_i}(q).$$

We have that,

$$\begin{aligned}
&\sqrt{nh}\left(\bar{m}_{\pm,p}(q) - m_{\pm}(q)\right) \\
&= \sqrt{nh}\left((\bar{m}_{\pm,p}(q) - \hat{m}_{\pm,p}(q))\right) + \sqrt{nh}\left((\hat{m}_{\pm,p}(q) - m_{\pm}(q))\right) \\
&= \sqrt{nh}\left(\frac{1}{n}\sum_{i=1}^{n} s^{(p)}_{\pm,in}(h)\left(\hat{Q}_{Y_i}(q) - Q_{Y_i}(q)\right)\right) + \sqrt{nh}\left((\hat{m}_{\pm,p}(q) - m_{\pm}(q))\right) \\
&= o_p(1) + \sqrt{nh}\left((\hat{m}_{\pm,p}(q) - m_{\pm}(q))\right),
\end{aligned}$$

uniformly over $q \in [a, b]$. The last equality follows from Assumptions Q1 Corollary 21.5 in van der Vaart (2000) which guarantees the uniform convergence of each empirical quantile function $\hat{Q}_{Y_i}$ at rate $n_i^{-1/2}$ for $i = 1, \ldots, n$. Then Assumption Q2 translates this to a convergence rate of at least $O_p(n^{-1/2})$ for all $i = 1, \ldots, n$, which is faster than the rate $\sqrt{nh}$ for any $h \to 0$ as required by Assumption K2.

Then, define the Fréchet estimator with empirical distribution functions as,

$$\underset{\omega \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} s^{(p)}_{\pm,in}(h) d^2_{W_2}\left(\omega, \hat{Y}_i\right).$$

An identical argument to the one in Proposition A-3 shows that the quantile function of this estimator is,

$$\bar{m}_{\pm,\oplus,p} = \underset{h \in Q(\mathcal{G})}{\operatorname{argmin}}\, d_{L^2}\left(h, \bar{m}_{\pm,p}\right)^2,$$

that is, the projection of the local polynomial estimator with empirical quantile functions. Then, since I have established that the latter converges uniformly to the same limiting process as the standard local polynomial estimator, the same exact argument as in the proof of Theorem 4 implies that $\bar{m}_{\pm,\oplus,p}(\cdot) - m_{\pm,p}(\cdot)$ has the same limiting law as the $\hat{m}_{\pm,\oplus,p}(\cdot) - m_{\pm,p}(\cdot)$. Finally, the results for the treatment effects with empirical quantile functions, $\bar{\tau}^{\mathrm{R3D}}_p(\cdot)$, $\bar{\tau}^{\mathrm{F3D}}_p(\cdot)$, $\bar{\tau}^{\mathrm{R3D}}_{\oplus,p}(\cdot)$, $\bar{\tau}^{\mathrm{F3D}}_{\oplus,p}(\cdot)$, then follow from identical arguments as in the proof of Theorem 2. □

The following lemma establishes the intuitive result that the difference between the

Fréchet and local polynomial estimators converges faster than each of them converges to the population moment. In Section A-4.2, I use it to derive the IMSE-optimal bandwidth for the Fréchet estimator based on the standard MSE-optimal bandwidth for the local polynomial one,

Lemma A-5. *Under the Assumptions of Theorem 1,*

$$\|\hat{m}_{\pm,\oplus,p} \;-\; \hat{m}_{\pm,p}\|_{L^2} \;=\; o_p\big((n\,h)^{-1/2}\big).$$

Proof. I write $L^2 := L^2([a,b])$ for brevity, and let $\|\cdot\|_{L^2}$ denote the usual L^2 norm in $q \in [a,b] \subset (0,1)$.

As argued in the proof of Theorem 4, $m_\pm(\cdot)$ is an interior point of the convex set $\mathcal{Q}$ of quantile functions. Therefore, the metric projection $\Pi_{\mathcal{Q}}\colon L^2 \to \mathcal{Q}$ is Hadamard differentiable at $m_\pm$ with derivative equal to the identity operator. Concretely, this means there is a remainder function $r_n(h)$ with $\|r_n(h)\|_{L^2}/\|h\|_{L^2} \to 0$ whenever $\|h\|_{L^2} \to 0$, such that

$$\Pi_{\mathcal{Q}}\big(m_\pm + h\big) \;=\; m_\pm \;+\; h \;+\; r_n(h), \quad \text{where } \|r_n(h)\|_{L^2} = o\big(\|h\|_{L^2}\big).$$

Furthermore, $\Pi_{\mathcal{Q}}(m_\pm) = m_\pm$. Setting $h = h_n := \hat{m}_{\pm,p} - m_\pm$ in the above expansion, and noting that $\|h_n\|_{L^2} = O_p((n\,h)^{-1/2}) \to 0$ by Theorem 1, I obtain

$$\hat{m}_{\pm,\oplus,p} \;=\; \Pi_{\mathcal{Q}}\big(m_\pm + h_n\big) \;=\; m_\pm + h_n + r_n(h_n), \quad \|r_n(h_n)\|_{L^2} = o(\|h_n\|_{L^2}).$$

Hence

$$\hat{m}_{\pm,\oplus,p} \;-\; \hat{m}_{\pm,p} \;=\; \big[m_\pm + h_n + r_n(h_n)\big] \;-\; \big[m_\pm + h_n\big] \;=\; r_n(h_n).$$

By construction, $\|r_n(h_n)\|_{L^2} = o(\|h_n\|_{L^2})$ and from above, $\|h_n\|_{L^2} = O_p((n\,h)^{-1/2})$. Combining gives,

$$\|\hat{m}_{\pm,\oplus,p} - \hat{m}_{\pm,p}\|_{L^2} \;=\; \|r_n(h_n)\|_{L^2} \;=\; o\big(\|h_n\|_{L^2}\big) \;=\; o_p\big((n\,h)^{-1/2}\big).$$

□

A-6 Monte Carlo Simulations: Details

I conducted a simulation study to evaluate the performance of different estimators in an R3D setting under three distinct data-generating processes. In a given simulation with sample size n, the running variable X_i is drawn uniformly from $[-1, 1]$: $X_i \sim \text{Uniform}(-1, 1)$. Then, the outcome variable Z_{ij} is drawn n_i times for each i from an underlying random distribution Y_i. The data-generating process of the random distribution for each simulation is as follows,

1. Normal distributions with jumps in normal means and variances

   - $$Y_i \sim N\left(\mu(x), \sigma^2(x)\right),$$

     where

     $$\mu_(x) = \mu_0 + \beta_\mu x + \delta^{\pm}\Delta_\mu + \epsilon_{\mu,i}, \quad \epsilon_{\mu,i} \sim N(0, 1),$$

     $$\sigma(x) = \sigma_0 + \beta_\sigma x + 1(x \geq c)\Delta_\sigma + \epsilon_{\sigma,i}, \quad \epsilon_{\sigma,i} \sim \text{Abs}\left(N(0, 0.5)\right).$$

   - **Parameters**:
     - Base mean: $\mu_0 = 5$, slope: $\beta_\mu = 0.5$.
     - Base standard deviation: $\sigma_0 = 1$, slope: $\beta_\sigma = 0.2$.
     - Shifts at the cutoff: $\Delta_\mu = 2$, $\Delta_\sigma = 0.5$.
   - **True treatment effect**:

     $$\tau_1^{\text{DRD}}(q) = Q_{\text{above}}(q) - Q_{\text{below}}(q),$$

     where $Q_{\text{below}}(q)$ and $Q_{\text{above}}(q)$ are the q-th quantiles of the normal distributions below and above the cutoff, respectively.

2. Normal Distributions with Exponential Shifts in Normal Means

   - $x_i < c$:

     $$Y_i \sim N\left(\mu_{\text{below}}, \sigma^2_{\text{below}}\right),$$

     with μ_{below} and σ_{below} defined as in Scenario 1.
   - $x_i \geq c$:

     $$Y_i \sim N\left(\mu_{\text{above}}, \sigma^2_{\text{above}}\right),$$

where

$$\mu_{\text{above}} = \mu_{\text{below}} + S_i, \quad S_i \sim \text{Exp}(\lambda_{\text{shift}}),$$

$$\sigma_{\text{above}} = \sigma_c + \Delta_\sigma + \epsilon_{\sigma,i}, \quad \epsilon_{\sigma,i} \sim \text{Abs}\left(N(0, 0.5)\right).$$

- **Parameters**:
  - Exponential shift rate: $\lambda_{\text{shift}} = \Delta_\mu$.
- **True treatment effect**: Due to the complexity of the mixture distributions, the true treatment effect is computed via simulation:

$$\tau_2^{\text{DRD}}(q) = \mathbb{E}[Q_{\text{above}}(q)] - \mathbb{E}[Q_{\text{below}}(q)],$$

  where expectations are approximated using a large number of simulated samples.

3. Exponential Distributions with Shift in Exponential Means

- **Below the Cutoff** ($x_i < c$):

$$\mu_{\text{below}} \sim \text{Exp}\left(\lambda_\mu(x_i)\right), \quad \lambda_\mu(x_i) = \frac{1}{\mu_0 + \beta_\mu x_i},$$

$$Y_i \sim \text{Exp}\left(\frac{1}{\mu_{\text{below}}}\right).$$

- **Above the Cutoff** ($x_i \geq c$):

$$\mu_{\text{above}} = \mu_{\text{below}} + \Delta_\mu,$$

$$Y_i \sim \text{Exp}\left(\frac{1}{\mu_{\text{above}}}\right).$$

- **True treatment effect**:

Can be computed analytically as:

$$\tau_3^{\text{DRD}}(q) = \Delta_\mu \left(-\ln(1-q)\right).$$

The treatment effect at the cutoff is estimated using three methods:

1. A classical quantile RDD estimator (Qu and Yoon, 2019) with bias-corrected confidence intervals (Qu et al., 2024).
2. The simple R3D estimator in Section 2.4.1 with monotone rearrangement (Chernozhukov et al., 2010) of the estimated quantiles.
3. The (Fréchet) R3D estimator in Section 2.4.2.

Simulation Parameters

- **Sample Sizes**: $n = n_i = 100, 200, 500$.
- **Number of Replications**: $M = 1000$.
- **Quantile Levels**: q from 0.01 to 0.99 in increments of 0.01.
- **Bandwidth**: $0.1 \times \frac{N}{100}^{-1/5}$.

Performance Metrics

For each method, I compute:

- **Bias**: $\mathbb{E}[\hat{\tau}(q)] - \tau^{\text{DRD}}(q)$.
- **Standard Error (SE)**: $\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\tau}(q) - \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}(q)\right)^2$.
- **Mean Squared Error (MSE)**: $\mathbb{E}\left[(\hat{\tau}(q) - \tau^{\text{DRD}}(q))^2\right]$.

where $\hat{\tau}(q)$ is the treatment effect estimated using one of the three methods at quantile level q.

A-7 Additional Results

A-7.1 Tables

A-7.2 Figures

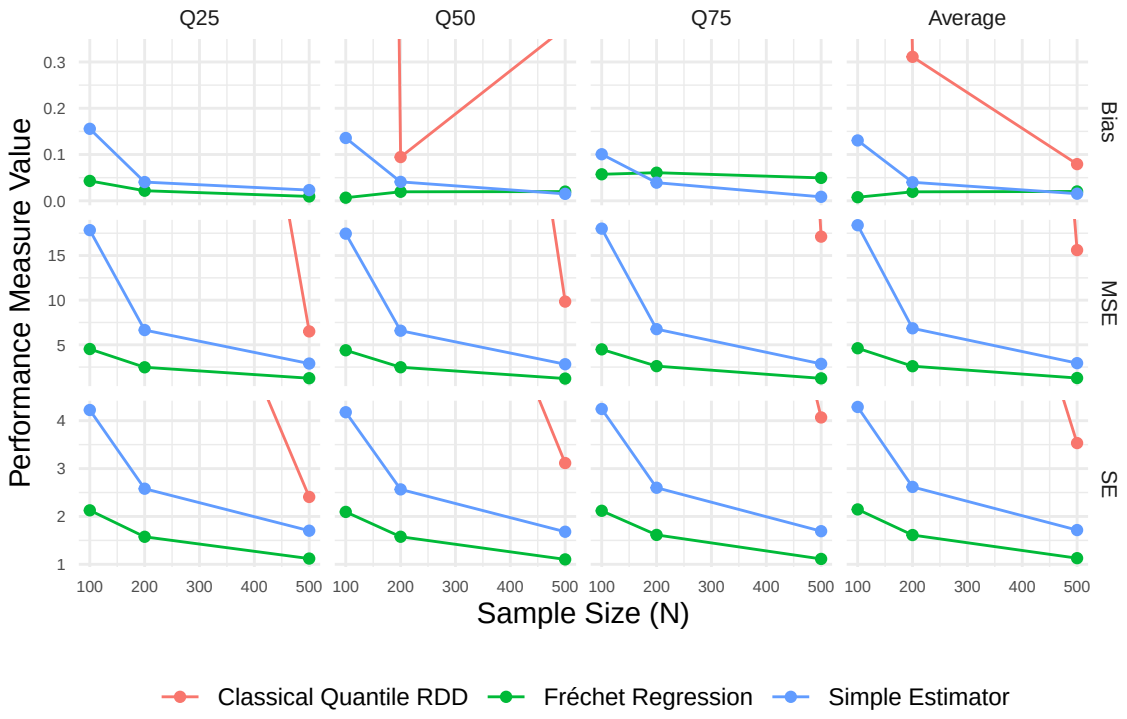


Figure A-2: Monte Carlo Simulations, Normal Distributions with Normal Means and Exponential Shift

Note: figure compares performance of three quantile estimation methods in terms of the relative bias (bias/treatment effect), mean squared error (MSE), and standard error (SE). Each measure is reported for $n = n_i = 200, 500, 1000$ and 5000 (x axis) at quantiles 25, 50, 75, and the average over all quantiles (facets). The methods are: 1) a local polynomial estimator for classical quantile RDDs (Qu and Yoon, 2019) with bias correction (Qu et al., 2024); 2) the simple R3D estimator in Section 2.4.1; 3) the (Fréchet) R3D estimator in Section 2.4.2. Data-generating process: outcome variable Y is a normal distribution with normally distributed means and variances that depend on running variable X and jump across the threshold according to an exponential distribution. For details, see Section A-6.

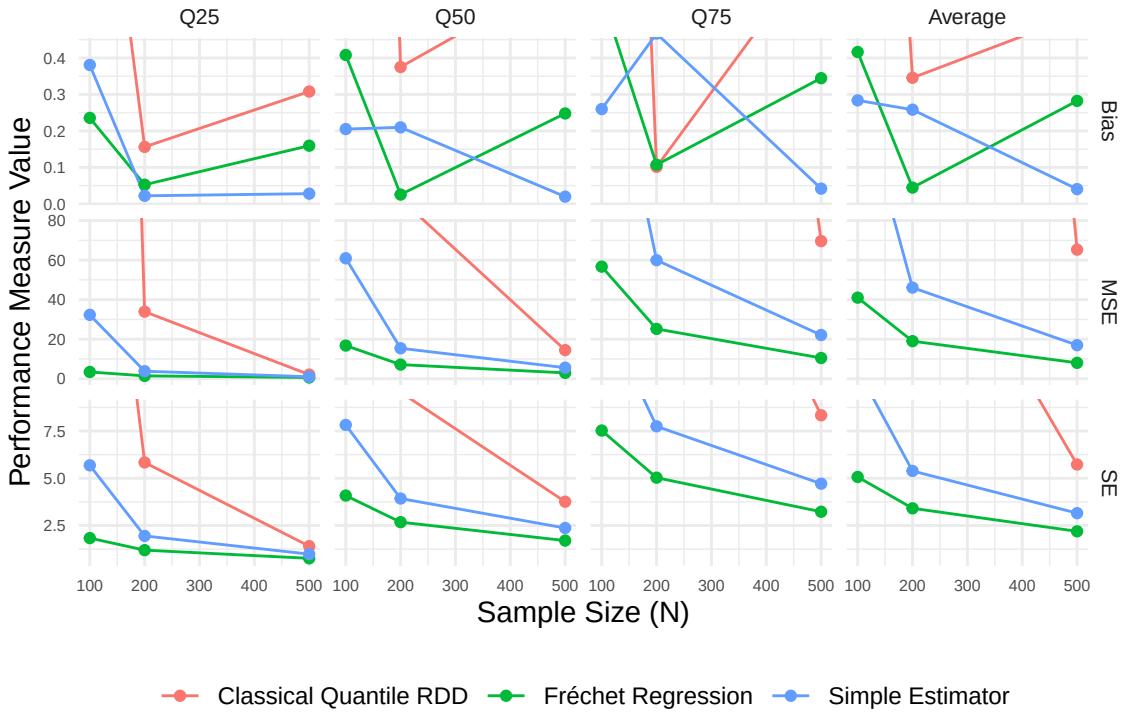


Figure A-3: Monte Carlo Simulations, Exponential Distributions with Exponential Means

Note: figure compares performance of three quantile estimation methods in terms of the absolute bias, mean squared error (MSE), and standard error (SE). Each measure is reported for $n = n_i = 100, 200$, and 500 (x axis) at quantiles 25, 50, 75, and the average over all quantiles (facets). The methods are: 1) a local polynomial estimator for classical quantile RDDs (Qu and Yoon, 2019) with bias correction (Qu et al., 2024); 2) the simple R3D estimator in Section 2.4.1; 3) the (Fréchet) R3D estimator in Section 2.4.2. Data-generating process: outcome variable Y is an exponential distribution with exponentially distributed means that depend on running variable X and jump across the threshold. For details, see Section A-6.

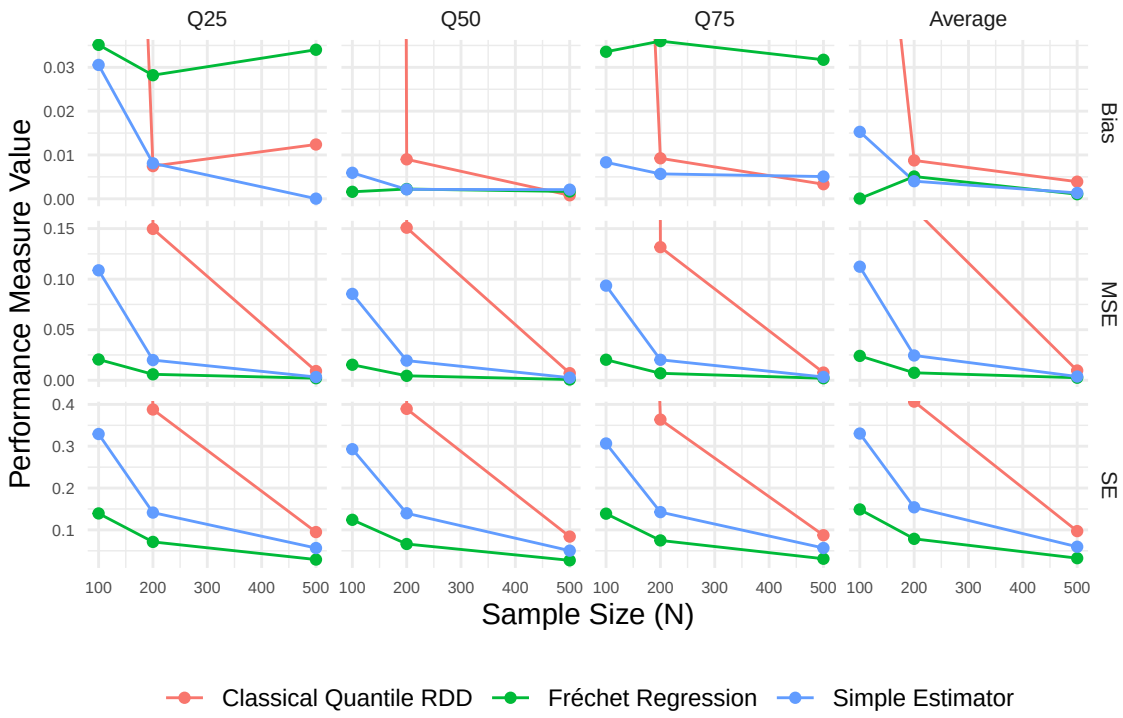


Figure A-4: Monte Carlo Simulations, Normal Distributions with Deterministic Means and Variances

Note: figure compares performance of three quantile estimation methods in terms of the absolute bias, mean squared error (MSE), and standard error (SE). Each measure is reported for $n = n_i = 100, 200$, and 500 (x axis) at quantiles 25, 50, 75, and the average over all quantiles (facets). The methods are: 1) a local polynomial estimator for classical quantile RDDs (Qu and Yoon, 2019) with bias correction (Qu et al., 2024); 2) the simple R3D estimator in Section 2.4.1; 3) the (Fréchet) R3D estimator in Section 2.4.2. Data-generating process: outcome variable Y is a normal distribution with deterministic means and variances that depend on running variable X and jump across the threshold. For details, see Section A-6.