# Regression Discontinuity Design with Distribution-Valued Outcomes

David Van Dijcke*

Department of Economics, University of Michigan

August 12, 2025
Click here for the latest version

**Abstract**

This article introduces Regression Discontinuity Design (RDD) with Distribution-Valued Outcomes (R3D), extending the standard RDD framework to settings where the outcome is a distribution rather than a scalar. Such settings arise where treatment is assigned at the level of an aggregate unit (e.g. firms, school districts) but the outcome of interest is an entire within-unit distribution (e.g. employee wages, student test scores). Since standard RDD methods cannot accommodate such two-level randomness, I propose a novel approach based on random distributions. Under a mild continuity assumption on the *average* of the random quantile functions, the jump in this average at the cutoff identifies a "local average quantile treatment effect". To estimate this target, I develop a functional local polynomial estimator based on local Fréchet regression. The estimator simultaneously fits the whole quantile curve, requires only one bandwidth, avoids quantile crossing, and—through optimal-transport geometry—yields a meaningful "average distribution". Exploiting the fact that the Fréchet fit is a projection of pointwise local polynomials onto the space of quantile functions, I establish asymptotic normality, construct robust, bias-corrected uniform confidence bands via a multiplier bootstrap, and propose a data-driven bandwidth selection procedure tailored to distribution-valued outcomes. Simulations confirm the accuracy and rapid convergence of this estimator and demonstrate that standard quantile RDD methods are biased and inconsistent in this setting. I illustrate R3D by estimating the effect of gubernatorial party control on within-state income distributions in the US, using a close-election design. The results suggest a classic equality–efficiency tradeoff under Democratic governorship, driven by reductions in income at the top of the distribution.

*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 Zach Brown, Alessandro Casini, Matias Cattaneo, Harold Chiang, Jacob Dorn, Ying Fan, Zheng Fang, Florian Gunsilius, Stefan Hoderlein, Eva Janssens, Harry Kleyer, Alexander Petersen, Yuya Sasaki, Konstantin Sonin, Benjamin Scuderi, Kaspar Wuthrich, and participants at the Emory Econometrics Seminar, the University of Michigan IO-Econometrics workshop and Labor Lunch, and the New Connections in the Study of Political Economy workshop for helpful discussion and comments. An R package accompanying this paper is available at https://davidvandijcke.com/R3D.

# 1 Introduction

Regression–discontinuity (RD) designs are among the most widely used non-experimental strategies for causal inference. By comparing observations just above and just below a known cutoff in an assignment variable—test scores, income thresholds, electoral margins—standard RDDs recover the jump in the conditional mean of a scalar outcome and interpret it as a local average treatment effect. The canonical set-up therefore presumes that each observational unit carries a single running variable and a single outcome measurement recorded at the same level of aggregation.

Many policy settings violate this premise. The running variable is often defined for an aggregate entity, while the outcome that matters is the entire distribution of micro-level observations contained in that entity. For example, a district-wide poverty rate determines eligibility for an education grant, yet the object of interest is the distribution of pupils' test scores within each district; a firm's revenue determines its eligibility for a subsidy, yet the objects of interest is the distribution of employee wages within each firm. These settings are marked by two layers of randomness: one *across* units (districts, establishments), and one *within*, which reflects heterogeneity in the different units' outcome distributions (students within districts, goods sold in an establishment). The scalar RDD framework is silent about this additional structure. Nonetheless, it is ubiquitous, with 38% of all RDD papers published in the top 5 economics journal, and 32% of those in the top 3 political science journals featuring some form of R3D setting – either disaggregated (e.g. employee-level data) or aggregated (e.g. average employee outcomes within a firm).[1] Moreover, the design is becoming only more common, as shown in Figure A-2, with every single RDD paper in the top economics journals in the last two years involving some form of R3D setting. These observations motivate the development of a more general RDD framework that can accommodate distribution-valued outcomes.

In this article, I extend the classical RD framework to a functional-data setting that accommodates distribution-valued outcomes. I refer to the resulting design as a *Regression Discontinuity Design with Distributions* (R3D). Conceptually, each observational unit provides not a single outcome but an entire distribution, so the outcome distribution is itself a random object. This naturally leads to a novel concept of distribution-valued treatment effects, the *local-average quantile treatment effect* (LAQTE), which captures the underyling *average* quantile function around the cutoff, where the average is with respect to the *dis-*

[1]The top 5 economics journals are considered to be, in no particular order: Quarterly Journal of Economics, Econometrica, American Economic Review, Review of Economic Studies, Journal of Political Economy. The top 3 political Science journals are considered to be: American Political Science Review, American Journal of Political Science, Journal of Politics.

Table 1: R3D-Like Settings in Top Journals (2014–2024)

| | Economics | Political Science |
|---|---|---|
| Any R3D (%) | 37.9 | 32.3 |
| Disaggregated (%) | 25.8 | 15.1 |
| Aggregated (%) | 19.7 | 19.4 |
| Total Number of RD Papers | 78 | 104 |
| Total Number of Papers | | |

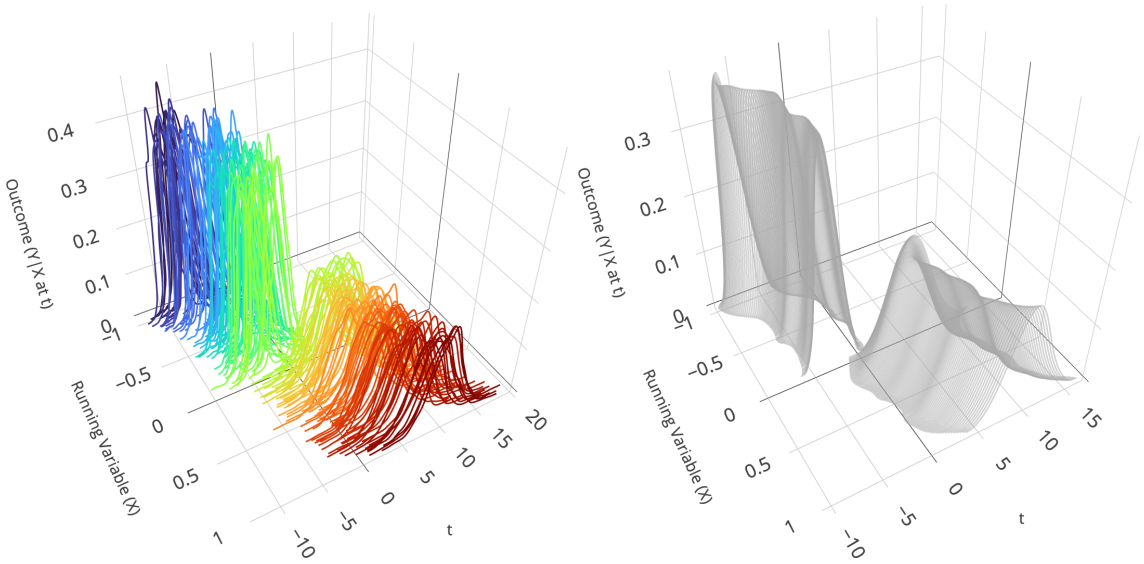
*Note:* this table shows percentage of top 5 Economics and top 3 Political Science journals with RD designs that qualified as an R3D. "Any R3D" indicates any form of R3D setting, while "Aggregated" and "Disaggregated" indicate whether the data was aggregated to the level of the treatment assignment for estimation, or whether the disaggregated (within-unit) data was used. Sample consists of any paper in those journals that had "regression discontinuity" or "RDD" in any of its fields.

*tribution of distributions*. Identification follows from an intuitive condition: the conditional mean of the quantile functions must be continuous in the running variable, but not the quantile functions themselves. This is the direct analogue, for functional outcomes, of the smoothness assumption that underpins scalar RDDs. Figure 1 illustrates the contrast: in a standard design the data are a random point cloud (rainbow colors) and their conditional expectation (gray color) is a smooth scalar-valued function; in R3D, the data points are random distributions (rainbow colors), and their conditional expectations (gray color) are a smooth path of distributions.

To estimate these average quantile treatment effects in practice, I propose a *functional local–polynomial estimator*, which regresses the entire quantile functions of the units to the left and right of the cutoff on the scalar-valued running variable. The procedure is based on local Fréchet regression in 2-Wasserstein space (Petersen and Müller, 2019), tailored to the RD setting. It minimizes a local least-squares criterion defined on the space of probability distributions. By virtue of its functional nature, it exploits all information in each unit's distribution, needs only one bandwidth and estimation step, avoids quantile crossing, and produces a directly interpretable object—-the "conditional average distribution" at the cutoff (Agueh and Carlier, 2011; Fan and Müller, 2024).

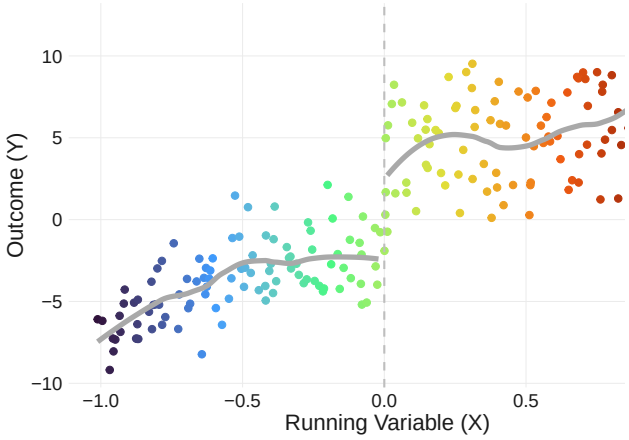
For comparison, I also develop a pointwise alternative that extends the canonical local–polynomial RDD to *random quantiles*: each unit's empirical quantile function is evaluated on a grid of probabilities and the resulting values are smoothed separately. The functional estimator can be viewed as the $L^2$ projection of these pointwise fits onto the cone of valid quantile functions. Leveraging this link, I adapt the general theoretical results of Chiang et al. (2019) to the distribution-valued setting to derive uniform, debiased confidence bands for the pointwise estimator, and then extend these results to function space by ex-

Figure 1: Example of a Distribution-Valued RDD




(a) R3D: Sample

(b) R3D: Conditional Expectation




(c) RDD: Sample + Conditional Expectation

ploiting the smoothnness of the projection operator. This approach of casting the functional estimator as a projection of existing frameworks allows me to sidestep the lack of central limit theorems for general metric spaces and derive the first uniform inference for local Wasserstein–Fréchet regression (Dubey and Müller, 2019). As such, my results complement those of Petersen et al. (2021), who derived confidence bands for global Wasserstein-Fréchet regression by leveraging optimal transport geometry and the linearity of the global regression model. I similarly leverage that space's optimal transport geometry, but without requiring a linear response model, instead exploiting the connection to the pointwise local polynomial estimator.

Further, I demonstrate the estimator's use in an empirical application. The question

studied is, "what is the effect of partisan control of the state governor's office on the within-state income distribution?". To answer it, I leverage a close-election R3D design, which compares states where the Democratic candidate narrowly won their election to states where they narrowly lost. Because each state has only a single election outcome but an entire distribution of family incomes, this is a prototypical R3D setting. Applying the proposed estimators to this setting, I estimate reductions in income for above-median earners that get stronger with income and become statistically significant for the top 10 percentiles, but no such effects for lower-income families. These results point to a classical equality–efficiency tradeoff (Okun, 1975), where a decrease in income inequality can only be achieved at the cost of an overall loss of income.

To conclude the introduction, I note that the setting considered here is distinct from that of the quantile RD (Q-RD) setting first developed in Frandsen et al. (2012). That approach estimates quantile treatment effects for *scalar-valued* outcomes, and thus does not apply to the distribution-valued setting considered here. Indeed, in what follows, the quantile RD estimator is shown to be biased and inconsistent in the R3D setting, both theoretically and in simulations. This bias results from the fact that its scalar-valued sampling framework is inappropriate for the R3D setting, and its identifying smoothness assumption highly restrictive. This can be seen in Figure 1. Because of random sampling, the *observed* distributions (rainbow) exhibit *discontinuous* changes, violating the Q-RD assumption. The *average* distributions (gray) do evolve smoothly, however. Of course, when treatment and outcome are measured at the same level – i.e., we find ourselves in the classical RD setting – the quantile RD estimator is preferred over the R3D estimator.

## *Literature*

This article contributes to several strands of literature, the primary one being the literature on regression discontinuity design (Thistlethwaite and Campbell, 1960; Hahn et al., 2001), see Lee and Lemieux (2010) and Cattaneo and Titiunik (2022) for an older and more recent overview. I contribute to this large area of research in three ways.

First, I extend the literature on quantile treatment effects in RDD to allow for distribution-valued outcomes. Frandsen et al. (2012) first developed the framework for quantile RD and derived uniform convergence results, though they did not derive uniform confidence bands. These were developed later for different types of quantile RD estimators in Qu and Yoon (2019); Qu et al. (2024); Chiang et al. (2019). Further variations of the classical quantile RD were studied in Jin et al. (2025); Chiang and Sasaki (2019); Qu et al. (2024); Chen et al. (2020). I build on this literature, in particular the general framework of Chiang et al. (2019),

to derive uniform confidence bands for distribution-valued RD designs. This also connects to the larger literature on distributional inference (Chernozhukov et al., 2013) and quantile regression and treatment effects (Koenker and Bassett Jr, 1978; Firpo et al., 2009; Firpo, 2007; Chernozhukov and Hansen, 2005).

Second, I contribute to the strands of literature that have developed robust, debiased confidence bands for local polynomial estimators with mean-squared error (MSE) based bandwidth selection procedures (Calonico et al., 2014, 2018, 2020, 2022; Armstrong and Kolesár, 2018; Imbens and Kalyanaraman, 2012), by extending these tools to distribution-valued settings. That places this paper in a rich literature built on the foundational contributions in local polynomial regression, particularly related to bias reduction and bandwidth selection, made by Fan and Gijbels (1992); Fan (1993); Fan and Gijbels (1995); Linton and Nielsen (1994).

Third, this article relates to several other papers that have considered RD designs with varying levels of aggregation. Borusyak and Kolerman-Shemer (2024) considered the opposite design, where the treatment assignment is at a *lower* instead of a higher level of aggregation than the outcome. Cattaneo et al. (2016, 2021); Bertanha (2020) considered aggregation schemes for RD with multiple cutoffs. Relatedly, Gunsilius and Van Dijcke (2025); Papay et al. (2011); Cheng (2023) considered RD designs with multi-dimensional or multiple assignment variables.

The other main strand this paper contributes to is the literature on Fréchet (1948) regression, which was originally developed by Petersen and Müller (2019) for general metric spaces, with several further contributions for distribution regression in Wasserstein space (Chen et al., 2023; Fan and Müller, 2022; Chen and Müller, 2023; Ghodrati and Panaretos, 2022; Zhou and Müller, 2024) and for local Fréchet regression (Chen and Müller, 2022; Iao et al., 2024; Qiu et al., 2024). As noted above, I contribute to this literature by deriving uniform confidence bands for local Fréchet regression in Wasserstein space, complementing related results for global Fréchet regression in Petersen et al. (2021) and for Wasserstein barycenters in Carlier et al. (2021); Agueh and Carlier (2017); Kroshnin et al. (2021). My results also hold for general polynomial orders while the literature has mostly focused on local linear regression, with the exception of Schötz (2022). More broadly, this article contributes to the large literature on functional data analysis (Ramsay and Silverman, 2005).

Relatedly, my results leverage the fact that Fréchet regression in 2-Wasserstein space is an $L^2$ projection of the local polynomial estimator onto the space of quantile functions. This relates closely to isotonic regression (Barlow et al., 1972) and monotone rearrangement methods (Chernozhukov et al., 2010), as well as shape-constrained inference with convex projection operators (Chetverikov et al., 2018; Groeneboom and Jongbloed, 2014; Fang and

Seo, 2021; Dümbgen, 2024).

This article also contributes to the literature applying optimal transport tools to causal inference – see Gunsilius (2025) for a recent overview. In particular, Gunsilius (2023) considered a similar setting to mine, where treatment is at a higher level than the outcome, in the context of synthetic controls (see also Van Dijcke et al. (2024) for an application to firm tenure distributions). More broadly, the paper relates to a small but rapidly growing literature developing causal inference methods for outcomes in geodesic spaces (Kurisu et al., 2024; Lin et al., 2023; Katta et al., 2024), including the well-known difference-in-differences (Athey and Imbens, 2006; Torous et al., 2024; Callaway et al., 2018; Zhou et al., 2025) and synthetic control estimators (Kurisu et al., 2025; Gunsilius et al., 2024).

Finally, this paper's empirical application-—to gubernatorial party control and income distributions—fits into a rich literature linking partisan control of US state governments to inequality and other economic outcomes. Building on Hibbs' partisan theory (Hibbs, 1977) and Kelly's *market conditioning* (Kelly, 2009), research has generally argued that Democrats, allied with lower income groups, adopt policies that narrow income gaps, whereas Republicans, favoring upper and business income constituencies, may widen them. Panel studies show that Democratic legislatures raise taxes and spending (Reed, 2006), implying stronger redistribution. Though recent evidence from difference-in-difference designs and close-election RDDs found no evidence that party control significantly affects most state-level economic outcomes within a governor's tenure (Dynes and Holbein, 2020), other close-election RDs have shown that Democratic control of state and local offices often increases minimum wages and welfare caseloads, compressing the post-tax income distribution (Leigh, 2008), and leads to more liberal policies (Caughey et al., 2017), including increased government spending and taxes (de Benedictis-Kessner and Warshaw, 2016). I contribute to this literature by providing credible causal estimates of the effect of gubernatorial party control on the income distribution, using rich individual-level data within each state instead of just state-level aggregates. That way, I estimate significant declines in pre-tax income for upper-income families, which compress the income distribution.

## 2 Regression Discontinuity with Distribution-Valued Outcomes

In this section, I present the distribution-valued version of the canonical regression discontinuity design. First, I formally introduce the setting, before providing several concrete examples from the literature. Then, I introduce a new definition of "local average quantile treatment effects" (LAQTE) appropriate for this setting, where the average is over *random* quantile functions. Before presenting two consistent estimators for these LAQTEs, I briefly

discuss the distinction between my R3D setting and the classical quantile RD setting of Frandsen et al. (2012). I conclude providing an overview of the statistical inference tools developed in Section 3, including extensions to fuzzy RDD and empirical quantile functions.

## 2.1 Setting

First, I define and discuss the R3D setting. Let $\mathcal{Y}$ be the space of cumulative distribution functions (cdfs) $G$ on $\mathbb{R}$ with finite variance, $\int_{\mathbb{R}} x^2 \, \mathrm{d}G(x) < \infty$. Let $(X, Y) \sim F$ be a random element with joint distribution $F$ on $\mathbb{R} \times \mathcal{Y}$. I call $X$ the running variable and $Y$ the outcome variable. Unlike the canonical RD design, here $Y$ is a random *distribution* rather than a random variable. Hence, each draw $(X_i, Y_i)$ from $(X, Y)$ provides a full distribution $Y_i$ at the running variable value $X_i$, rather than a single real number. Then, denote $T \in \{0, 1\}$ the treatment status. I assume that $T$ is some monotonic function of $X$ such that,

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

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 $c = 0$. 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 2.7.1).

In addition, denote the marginal distributions of $X$ and $Y$ as $F_X, F_Y$. I assume that $\mu = E[X], \Sigma = \mathrm{var}(X)$ and the conditional distributions $F_{X|Y}, F_{Y|X}$ exist with $\Sigma$ positive definite. Here, $F_{Y|X}$ is a probability measure supported on the set of cdfs $\mathcal{Y}$, $F_{Y|X=x}(A) \coloneqq P(Y \in A \mid X = x)$, $A \subseteq \mathcal{Y}$ with $A$ measurable.

## 2.2 Motivating Examples

To make the setting more concrete, I now provide several prominent examples from the literature that can be viewed as R3D designs. They are instances of broader classes of settings where treatment is assigned to units at a higher level of aggregation than the outcomes.

**Example 1** (Administrative units)**.** *In an influential article, Ludwig and Miller (2007) study 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 with less years of schooling? More broadly, the R3D design applies whenever a treatment is jointly assigned to all members of an administrative unit, such as counties, school 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 vote 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 increase in test scores, or do only the lowest-scoring 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, since vote-based allocation systems typically aggregate decisions of many individuals into higher-level outcomes, nearly all instances of the ubiquitous "close-election" RD design fall under the R3D framework.*

**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. Since establishments typically sell many items and employ tens to hundreds of employees, one could, with the right data, observe entire distributions for each establishment. Wages and tenure (length of employment) could be measured at the employee level, and prices at the product level. Then one could answer questions such as: did the minimum wage increase mainly spur new hires, or did employment increase across the tenure 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 one wants to study 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, poverty level), I observe an entire distribution of the outcome (store prices, test scores, child mortality), and these outcome distributions vary 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 discussed above. Consequently, new concepts of average treatment effects and discontinuities that are appropriate for random distributions are required, which I introduce in the next section.

## *2.3 Local Average Quantile Treatment Effects*

### 2.3.1 Definition

To begin, I define a new treatment effects concept for distribution-valued outcomes appropriate for the setting introduced above. Following Neyman-Fisher-Rubin notation, denote $Y^0 \in \mathcal{Y}$ the counterfactual outcome distribution in the absence of treatment and $Y^1 \in \mathcal{Y}$ the outcome distribution under treatment. Define the observed outcome

$$Y = \begin{cases} Y^0 & \text{if } T = 0 \\ Y^1 & \text{if } T = 1 \end{cases},$$

noting that $Y$ is a cdf so I can write $Y(y)$, $y \in \mathbb{R}$ to evaluate the function at a given point $y$.

Consider, for a moment, the canonical RD setting, such that $Z^T \in \mathbb{R}$ the classical scalar-valued counterfactual outcome. Then, assuming that the treatment effects vary between units, the classical local treatment effect is (Hahn et al., 2001)

$$E[Z^1 - Z^0 \mid X = 0],$$

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

In the R3D setting, $Y^T$ is a full distribution function. An intuitive generalization of the classical average treatment effect to settings with distribution-valued outcomes is given in the below definition. Write $Q_Y(q)$ for the function mapping the cdf $Y$ to quantiles,

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

Then I get,

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

$$\begin{aligned} \tau^{R3D}(q) &\coloneqq E\left[Q_{Y^1}(q) - Q_{Y^0}(q) \mid X = 0\right] \\ &\coloneqq m_1(q) - m_0(q), \quad q \in [0,1]. \end{aligned} \tag{1}$$

Observe that the expectation is taken with respect to the conditional *distribution of distributions*, $F_{Y^T|X=0}$,

$$m_T(q) = E[Q_{Y^T}(q) \mid X = 0] = \int_{\mathcal{Y}} Q_y(q)\, \mathrm{d}F_{Y^T|X=0}(0, y), \quad T = 0, 1.$$

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 conditional Wasserstein barycenters, which respect the intrinsic geometry of the underlying probability measures being averaged over. In particular, the distribution defined by the LAQTEs has the intuitive interpretation of being the unique distribution with the lowest possible cumulative "least-squares" cost of transporting its probability mass 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, i.e. the unique quantity that has the lowest expected least-squares distance to all points.

Next, I show that these unobserved LAQTEs can be identified from observed $(X, Y)$.

### 2.3.2 Identification

To identify $\tau^{R3D}$ from the data, I impose two assumptions that generalize the canonical RDD requirements. First, I assume that the average quantile function is continuous in the running variable around the threshold.

**I1** (Continuity)**.** $E[Q_{Y^T}(q) \mid X = x]$ *is continuous in* $x$ *for all* $q \in [0,1]$*, for* $x \in ]-\varepsilon, \varepsilon[$*,* $T \in \{0,1\}$ *and* $\varepsilon > 0$*.*

Importantly, this assumption allows for the *observed* random distributions $Y$ to evolve discontinuously with $x$, like in the top left panel of Figure 1.

The following example may help to clarify this point. Suppose $F_{Y^T|X=x} \sim N(N(g(x) + \tau T, 1), 1)$ for $T = 0, 1$, $\tau > 0$ and $supp(X) = [-1, 1]$. In words, the counterfactual distribution functions $Y^T$ are drawn from a class of normal distributions with normally distributed means that depend on $X$ and shift with treatment $T$. The distributions in Figure 1 are an instance of this class. 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 conditional average distributions estimated on either side of the cutoff using the local polynomial approach set out in Section 2.5. These average distributions are clearly continuous in the running variable. This demonstrates that even this simple collection of random Gaussian distributions satisfies the weaker continuity assumption in I1 but still fails

continuity in quantiles. The following example establishes this formally. I include the proof here for intuition.

**Example 4.** *Suppose* $(X, Y) \in (\mathbb{R}, \mathcal{Y})$ *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 functions* $Y_X(t, x)$ *themselves are not continuous in* $x$*.*

*Proof.* For a given $t$, define a new random variable $Z = \Phi(t - W)$ where $\Phi$ 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 $x$ since $\Phi^{-1}(z)$ is constant for fixed $z \in (0, 1), t - g(x)$ is continuous by continuity of $g(x)$, and $\Phi$ is continuous everywhere. Moreover, since $0 < \Phi(x) < 1$, $Y|X = x$ is almost surely not continuous in $x$. □

The example solidifies the intuition behind Figure 1. While the *probability* of drawing a certain distribution varies smoothly in $X$ the actual distributions at any two points $x, x'$ close to each other will always be different with probability 1. This follows from the distributions being *random objects* themselves. In Section 2.4 how this setting precludes the smoothness assumption used in the classical quantile RD (Frandsen et al., 2012).

The second assumption I 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, I obtain the following identification result.

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

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

where the lemma defines $m_\pm(q), m(q)$.

### 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, I can define a discontinuity in our setting to occur when, for some $q \in [0, 1]$

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

The uniform confidence bands I derive below allow one to test for the presence of such discontinuities for a given quantile $q$. Alternatively, one can conduct inference on entire segments of the distribution at once. An overview of inference is given in Section 2.6.

## *2.4 Comparison to Quantile RDD*

Before developing the estimators for the LAQTEs, I briefly discuss the difference between the R3D setting and the quantile RD estimator of Frandsen et al. (2012). The key insight is that quantile RDs are appropriate for estimating quantile treatment effects (QTE) for scalar-valued outcomes, while the R3D estimator can estimate (average) QTEs for distribution-valued outcomes, and there is no overlap in use cases.

A comparison of the population quantities targeted by each estimator makes this point clearer. Let $Y \in \mathcal{Y}$ and $Z \in \mathbb{R}$ as before. The two population objects targeted are,

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

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,

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

As can be seen, the R3D approach *first* estimates quantiles and only then runs a local linear regression. This properly accounts for the two-level randomness intrinsic to the R3D setting. Distribution estimation at a given $X = x$ precedes smoothing. By contrast, the Q-RDD estimator intrinsically estimates the distribution *by smoothing*, ignoring the randomness *within* units. In the presence of such randomness, the observed distributions will almost surely not very smoothly, and the Q-RD approach will be biased and inconsistent.

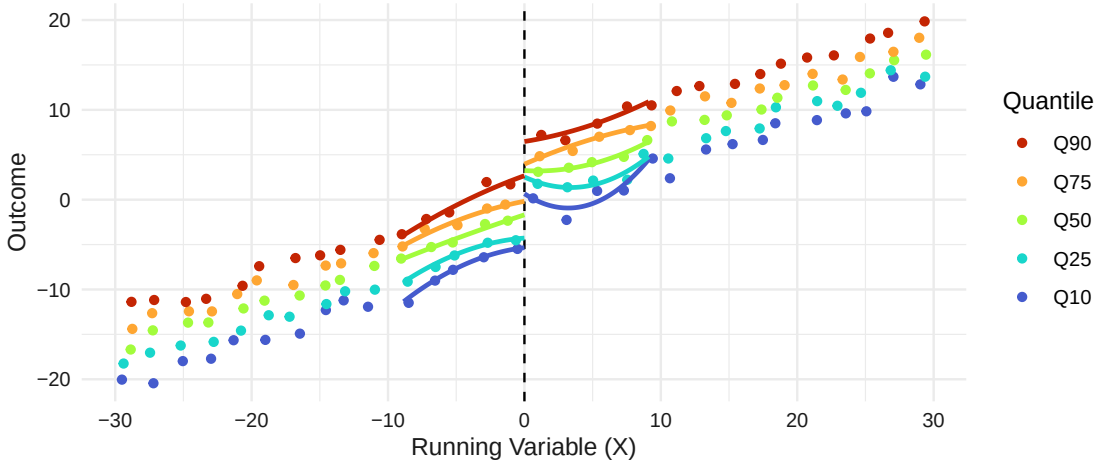
Underlying these arguments are three distinct differences between the R3D and the Q-RDD setting. First, as mentioned, the sampling model imposed by the Q-RDD setting does not correctly represent the underlying data-generating process. In particular, it assumes i.i.d. sampling of scalar-valued outcomes instead of distribution-valued ones, which ignores

the within-unit sampling that characterizes the R3D setting. As such, the sampling framework of the Q-RD design could never result in multiple data points having the same value of the (continuous) running variable. Second, as mentioned, the quantile continuity assumption required for the identification of the estimator in Frandsen et al. (2012) is highly restrictive in the R3D setting, requiring that two units that are both close to the threshold have essentially 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 requires a much weaker continuity assumption in I1. In particular, it only demands, 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 distribution-valued 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. Indeed, while I1 is consistent with the common approach of averaging the outcome variable at the level of the aggregate unit and then estimating a standard RD, the continuity assumption in Frandsen et al. (2012) is not, because there would be no random variation left in the averages, which are assumed to evolve smoothly. 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 distributions in the absence of treatment (an unrealistically strong assumption), the outcome distributions of those companies under treatment will still differ.

## *2.5 Estimators*

To estimate the local average quantile treatment effects introduced above, I now propose two intuitive estimators that generalize local polynomial regression to the R3D setting with distribution-valued 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

Figure 2: Local Polynomial Estimator: Illustration




builds on this by projecting the local polynomial estimator back onto the space of quantile functions. As shown in Proposition A-2, the resulting estimator coincides with the local Fréchet regression estimator of 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). In Section 3, I derive valid uniform confidence intervals for both approaches, though the Fréchet estimator is preferable due to its computational advantages, superior finite-sample performance, and its more meaningful interpretation as the "average" distribution.

### 2.5.1 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.

The local polynomial R3D estimators $\hat{m}_{\pm,p}(q)$ of order $p$ for each quantile $q$ can then be written in their standard form

$$\hat{m}_{\pm,p}(q) = \Big(\text{polynomial fit at } x = 0^{\pm}\Big)\Big|_{\text{order}=p} \tag{3}$$

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

where $K(x)$ a kernel function, $\delta_i^{\pm} := 1\{X_i \gtrless c\}$, and $r_p(x) := (1, x, x^2, \ldots, x^p)$. The

only difference with the standard local polynomial RDD estimator is that I now have i.i.d. samples $(X_i, Q_{Y_i}(q))$ instead of $(X_i, Y_i)$. Standard derivations give the following solution for the conditional mean estimator,

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

where $s^{(p)}_{+,in}(h)$ are the usual 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 cutoff point $X = 0$. Further, observe that 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).

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

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

In Section 3 below, I show that, under some assumptions, this estimator converges uniformly to an asymptotic normal distribution centered at the true treatment effect, for $p \geq 1$. Following Chiang et al. (2019), I build bias correction into the estimator by leveraging Remark 7 in (Calonico et al., 2014), which establishes an equivalence between explicitly bias-corrected estimators and estimators where the MSE-optimal bandwidth is chosen based on a pilot estimator of lower order – which is the approach I will take.

### 2.5.2 Local Fréchet Regression

Three intuitive improvements can be made to the local polynomial regression on quantiles introduced above. First, as noted, the resulting function is not guaranteed to be a quantile function because the weights $s^{(p)}_{\pm,in}(h)$ 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, as confirmed in the simulations below. Third, the pointwise estimation approach also requires repeated bandwidth selection and estimation for each quantile, leading to computational overhead. To resolve these three issues at once, I consider the following extension of the estimator,

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

where $Q(\mathcal{Y})$ is the space of quantile functions of the cdfs in $\mathcal{Y}$, restricted to $[a, b] \subseteq [0, 1]$. I define $\Pi_{\mathcal{Q}}$ as the $L^2$ projection onto that space of restricted quantile functions.[2] In Proposition A-2, I show that $\hat{m}_{\pm,\oplus,p}$ is unique and exists under the stated assumptions. The estimated treatment effects are then defined as,

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

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). Indeed, the approach can be viewed as a "double regression": a local linear regression on pointwise quantile functions, followed by a functional 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 $\mathcal{Y}$ equipped with the 2-Wasserstein distance, $d_{W_2}$ (i.e. 2-Wasserstein space). 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 same population quantile function $m_{\pm}(q)$ as 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 connection is that the local Fréchet estimator converges to the conditional Fréchet mean on $(\mathcal{Y}, d_{W_2})$, 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 at the cutoff, i.e., $m_{\pm}$ (see the proof in Proposition A-2). In short, the Fréchet estimator 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). Similar to the population object $m_{\pm}(q)$, 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 sample quantile functions $\{Q_{Y_i}\}_{i=1}^n$ 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

[2] Working on $[a, b]$ instead of $[0, 1]$ requires much weaker assumptions on the support of the distributions and is nearly equivalent in practice, see Section 3.

latter generally does not have this desirable interpretation. Another advantage is that local Fréchet regression more naturally leverages global function information by smoothing large deviations across quantiles 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 express themselves in better finite-sample performance in the simulations in Section 4.1.

## *2.6 Overview of Inference*

For these two R3D estimators (one local polynomial, one Fréchet), 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 $\hat{\mathbb{G}}^{\text{R3D}}$, $\hat{\mathbb{G}}^{\text{F3D}}$ for the sharp and fuzzy design, respectively, that are shown to converge to the uniform limiting law and hence can be used to construct uniform confidence bands. This allows one to determine what quantiles have a statistically significant treatment effect while accounting for multiple testing due to the functional nature of the estimands.

Moreover, the bootstrapped distributions can also be used to construct critical values for various distributional hypothesis tests. In particular, treatment nullity and homogeneity can be tested in a particular part of the distribution $[\underline{q}, \overline{q}] \subset (0, 1)$ through the following tests (Chiang and Sasaki, 2019):

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

where the critical values can be constructed by taking the $(1 - \lambda)$-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 $\left\{ \max_{q \in [\underline{q}, \overline{q}]} \left\lvert \hat{\mathbb{G}}^{\text{R/F3D}'}(q) - \frac{1}{\overline{q} - \underline{q}} \int_{[\underline{q}, \overline{q}]} \hat{\mathbb{G}}^{\text{R/F3D}'}(q')\, dq' \right\rvert \right\}_{b=1}^{B}$ with $\lambda$ the desired level of statistical significance and $B$ the number of bootstrap repetitions.

## *2.7 Extensions*

### 2.7.1 Fuzzy R3D

So far, I have focused on the sharp regression discontinuity design, where treatment assignment is a deterministic function of the cutoff. Now, I show how to define a fuzzy R3D design

in which 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 $T_i^0$, $T_i^1$, the local potential treatment states as $\lim_{x\to 0^\pm} T_i(x)$, where $T_i(x)$ is the potential treatment status of unit $i$ 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** (Fuzzy RD)**.** $\lim_{x\to 0^+} P(T \mid X = x) > \lim_{x\to 0^-} P(T \mid X = x)$.

**I4** (Treatment Continuity)**.** $E[T|X = x]$ *is continuous in* $x$ *over* $]-\varepsilon, \varepsilon[$, $\varepsilon > 0$.

**I5** (Monotonicity)**.** $\lim_{x\to 0} P(T^1 > T^0 \mid X = x) = 1$ *and* $P(\text{Indefinites}) = 0$.

This gives,

**Lemma 2** (Fuzzy Identification)**.** *Under Assumptions I1– I5, the unobserved* $\tau^{F3D}$ *is identified from the joint distribution of the observed* $(X, Y, T)$ *as,*

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

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

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

The corresponding Fréchet estimator is,

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

### 2.7.2 Empirical Quantile Functions

So far, I have assumed that the researcher observes entire quantile functions $Q_{Y_i}$. This is realistic in settings where an entire population of sub-units within a given aggregate unit is observed – for example, when a researcher has access to the census so that all firms within a US county are in the data. In practice, however, there is often another sampling layer, where one only observes further i.i.d. *samples* $Z_{ij}, j = 1, \ldots, n_i$ from these distribution functions, with $Z_{ij} \in \mathbb{R}$ distributed according to $Y_i$.[3] In Section 3, I show that under standard assumptions, the empirical quantile functions converge to the true quantile functions faster than the R3D estimators and hence do not affect the asymptotic results. The corresponding sharp RD estimator is defined as,

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

where

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

with

$$\widehat{Y}_i(z) \coloneqq \frac{1}{n_i} \sum_{j=1}^{n_i} 1\left(Z_{ij} \leq z\right),$$

the empirical distribution function. The other estimators are similarly modified by plugging in $\widehat{Q}_{Y_i}$, and denoted with a bar instead of a hat, e.g. $\bar{\tau}_p^{\mathrm{R3D}}$. Sampling weights can be incorporated by constructing $\widehat{Q}_{Y_i}$ as weighted quantile functions. The asymptotic results for this setting are established in Section 3.3.1.

[3] 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.

# 3 Statistical Results

In this section, I derive the asymptotic distributions of the local polynomial and Fréchet regression estimators. I do so in full generality for $p$-th order local polynomials, accommodating 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 distribution-valued outcomes. The corresponding results for the local Fréchet regression follow from the functional delta method and a projection argument that uses a version of Rademacher's theorem for Banach spaces (Preiss, 2014). I conclude by extending these asymptotic results to the empirical quantile setting.

## *3.1 Assumptions*

Throughout, I work on the restricted set of quantiles $[a, b]$, a compact subset of $(0, 1)$, and let $\underline{c} < 0 < \bar{c}$. Also, define $\mathcal{Y}_c := \{Y(\omega) : \omega \in \Omega^x, X(\omega) \in [\underline{c}, \bar{c}]\}$ as the set of random cdfs that are realized in a small neighborhood around the cutoff. Also, let $h_1(q)$ be the bandwidth for the numerator in the fuzzy RD at quantile $q$, and $h_2(q)$ the same for the denominator.

**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 $c_1(q) : [a, b] \to C \subset \mathbb{R}$ a bounded Lipschitz function and $c_2(q) = \bar{c}_2 > 0$. The baseline bandwidth $h = h_n$ satisfies $h \to 0$, $nh^2 \to \infty$, $nh^{2p+3} \to 0$.*

**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 $Y_i$ for each $i = 1, \ldots, n$.*

**L2** (Average Quantile Continuity). *For each $x \in \mathcal{N} := (-\epsilon, \epsilon)$ for $\epsilon > 0$ and $q \in [a, b] \subset (0, 1)$, the following conditions hold:*

*(i) The maps $(x, q) \to E[Q_Y(q)|X = x]$ 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 $x \to E\big[Q_Y(q_1)Q_Y(q_2) \mid X = x\big] \in C^1(\mathcal{N} \setminus \{0\})$ 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* $[\underline{c}, \overline{c}] \times \mathcal{Y}_c$, $\epsilon > 0$.

**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 important benefit that for 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 assumption I1 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 $E[Q_Y(q)|X = x]$ through the covariance of the quantiles. 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) and is imposed in Petersen and Müller (2019). 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 point-wise fact into a statement that these caps cannot 'blow up' too often in all possible realizations $Y$. In practice, this means that while each $Y$ can have unbounded support, the family $\mathcal{Y}$ must not produce extremely large quantiles too often 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.

For the extension with empirical quantile functions, I further impose the following,

**Q1** (Empirical Quantiles). *Any distribution function* $Y \in \mathcal{Y}_c$ *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* $[Q_Y(q_1) - \varepsilon, Q_Y(q_2) + \varepsilon]$ *with strictly positive density.*

**Q2** (Sample Balance). *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). Also note that if the entire distribution *is* observed, then these assumptions are not required and the quantile functions are allowed to have discontinuities. 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 can derive the asymptotic distribution of the local polynomial estimator. For that, I need a few additional pieces of notation, borrowing from Chiang et al. (2019). For the formal results, I assume without loss of generality that the kernel $K$ is supported on $[-1,1]$. Define $g_1 : (Y,T,q) \subset (\mathcal{Y},\{0,1\},[a,b]) \to Q_Y(q)$, $g_2 : (Y,T,q) \subset (\mathcal{Y},\{0,1\},[a,b]) \to T$. Further, define the population residual $\mathcal{E}_k(y,t,x,q) \coloneqq g_k(y,t,q) - E[g_k(y,t,q)|X_i = x]$, $k = 1,2$ 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 $k,l \in \{1,2\}$, $q,q' \in [a,b]$, and $\sigma_{kl}(q,q' \mid 0^{\pm}) = \lim_{x\to 0^{\pm}} \sigma_{kl}(q,q' \mid x)$. 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 $X$, while $X_n \underset{\xi}{\overset{p}{\rightsquigarrow}} X$ denotes conditional weak convergence. The latter is defined as $\sup_{h\in BL_1}\left|E_{\xi|x}\left[h(X_n) - E[h(X)]\right]\right| \underset{x}{\overset{p}{\to}} 0$ 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, I first get the following preliminary result for the conditional means.

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

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

*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,*

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

*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,*

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

*and under the additional Assumptions I3–I5*

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

*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 $\{\xi_i\}_{i=1}^n$ defined in M1 to define the estimated multiplier process,

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

where $\hat{f}_X(0)$ is any uniformly consistent estimator of $f_X(0)$, and $\hat{\mathcal{E}}_k\left(Y_i,T_i,X_i,q\right)$ 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 bias-corrected 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* $\hat{\nu}^{\pm}_{\xi,n} \underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}_{H^{\pm}}$ *and thus,*

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

*and under the additional Assumptions I4, I5,*

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

*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 2.6 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,*

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

*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, barring some slight intricay on the boundaries of $\mathcal{Q}$. 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. □

Note that in this theorem the $c_1(\cdot)$ term does not appear because the Fréchet estimator uses a single bandwidth for all quantiles. 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,*

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

*and under the additional Assumptions I3–I5,*

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

**Corollary 2** (Bootstrap: Fréchet)**.** *Under the assumptions of Theorem 3, the estimated bootstrap processes* $\hat{\mathbb{G}}^{\mathrm{R3D}}$ *and* $\hat{\mathbb{G}}^{\mathrm{F3D}}\left(\hat{m}_{+,\oplus,p}, \hat{m}_{-,\oplus,p}, \hat{m}_{+,T,p}, \hat{m}_{-,T,p}\right)$ *deliver asymptotically valid confidence intervals for 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 1** (Empirical Quantiles)**.** *Under the same respective Assumptions of Theorems 1 and 2, as well as Assumptions Q1, Q2, the estimators with empirical quantile functions,* $\bar{m}_{\pm,p}$, $\bar{m}_{\pm,\oplus,p}$, $\bar{\tau}^{\mathrm{R3D}}_{p}$, $\bar{\tau}^{\mathrm{F3D}}_{p}$, $\bar{\tau}^{\mathrm{R3D}}_{\oplus,p}$, $\bar{\tau}^{\mathrm{F3D}}_{\oplus,p}$ *converge to the same uniform limiting processes as their respective population analogs.*

**Corollary 3** (Bootstrap: Empirical Quantiles)**.** *Under the assumptions of Theorem 3, along with 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 proposed estimators' performance, I conduct Monte Carlo simulations under several data-generating processes. Throughout this and the next section, I use R3D estimators of quadratic order but with bandwidths that are MSE-optimal for the linear estimators. As argued in Remark 7 Calonico et al. (2014), this is equivalent to using explicitly bias-corrected linear estimators.

In the simulations, I estimate the quantile treatment effects $\tau^{\mathrm{R3D}}$ at 10 quantiles using three estimators: 1) a local polynomial estimator for classical quantile RDDs (Qu and Yoon, 2019);[4] 2) the local polynomial R3D estimator in Section 2.5.1; 3) the (Fréchet) R3D estimator in Section 2.5.2. The Q-RD estimator is corrected for bias using the approach in (Qu et al., 2024). The reason for using the Q-RD estimator of Qu and Yoon (2019) is to give Q-RD the best possible chance, since this estimator allows for bias-corrected, uniform inference, improving on the estimator in Frandsen et al. (2012).

I consider two data-generating processes, where $X_i \sim \mathrm{Uniform}(-1, 1)$.

*DGP 1: Normal with Normal Means.* For each $i$, draw

$$\begin{aligned} \mu_i &\sim N\Big(5 + 5X_i + \delta^+ \Delta,\ 1\Big), \\ \sigma_i &\sim \big|\, N\big(1 + X_i,\ 1\big)\big|, \end{aligned} \tag{15}$$

and define $Y_i = N(\mu_i,\, \sigma_i^2)$.

*DGP 2: Normal–Exponential Mixture with Normal–Exponential means.* Set $\mu_i = \mathrm{Uniform}(-5, 5) + 2\,X_i$ and $\lambda_i = \mathrm{Uniform}(0.5, 1.5)$. Then, generate

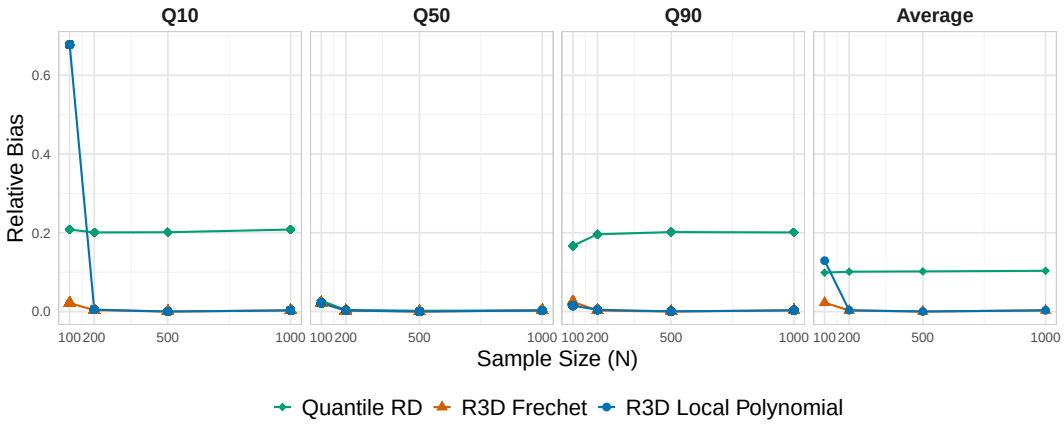
$$Y_i = N\big(\mu_i + \delta^+ \Delta,\ 1\big) + 2\,\mathrm{Exp}\big(\lambda_i + \delta^+ \Delta_\lambda\big). \tag{16}$$

In both setups, I let $\Delta$ vary across different simulations to test different treatment effect magnitudes. For the first DGP, the true treatment effects have the closed-form solution $N(\Delta, 2)$, implying constant treatment effects. The heterogeneous treatment effects in the second DGP

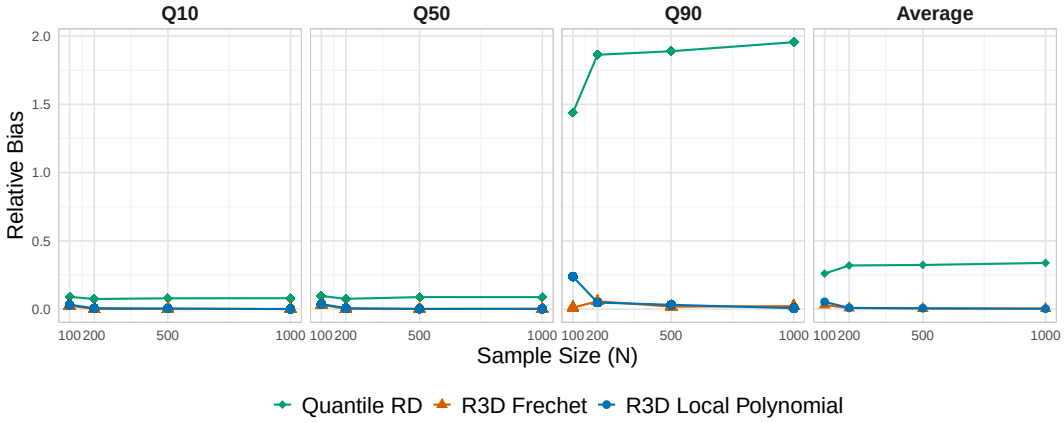
[4] Computed using the `rd.qte` command in R (Qu and Yoon, 2024).

are estimated by averaging across a large number of simulated quantile functions.

Figure 3: Simulated Bias of R3D and Q-RD Estimators




(a) Normal Distributions with Normal Means




(b) Normal–Exponential Distributions with Normal–Exponential Means

*Note*: this figure compares relative bias (absolute bias as a percent of the true treatment effect size) of R3D and Q-RD estimators. Each measure is reported for $n = n_i = 200, 500, 1000$, and 2000 (x axis) at quantiles 10, 50, 90, and the average over all quantiles (facets). Results are averaged over $2,500$ simulations for each sample size. 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 local polynomial R3D estimator in Section 2.5.1; 3) the Fréchet R3D estimator in Section 2.5.2. Bandwidths are selected using (I)MSE-optimal procedure in Section A-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.

Figure 3 shows the estimators' performance in terms of relative bias, which is the magnitude of the estimated bias at a given quantile as a proportion of the treatment effect at that quantile. I set $\Delta = 2$ but the results are similar for other values. The green line (diamonds) shows the quantile RD estimator, the orange line (triangles) the Fréchet estimator, and the blue line (circles) the local polynomial one. In line with theoretical expectations, the

quantile RD estimator appears to be inconsistent and suffers from large finite sample bias, with a relative bias that is at least an order of magnitude higher than the R3D estimators', for some quantiles. As expected, the quantile RD estimator performs well at the median in DGP 1, because a mixture of normals approximates the average normal at the median. Similarly, due to the heavy tails of the exponential distribution, it performs much worse at the upper quantiles in DGP 2. Between the two R3D estimators, the Fréchet estimator has much lower bias than the local polynomial one for small sample sizes, but both converge quickly to near-zero, supporting the asymptotic theory.

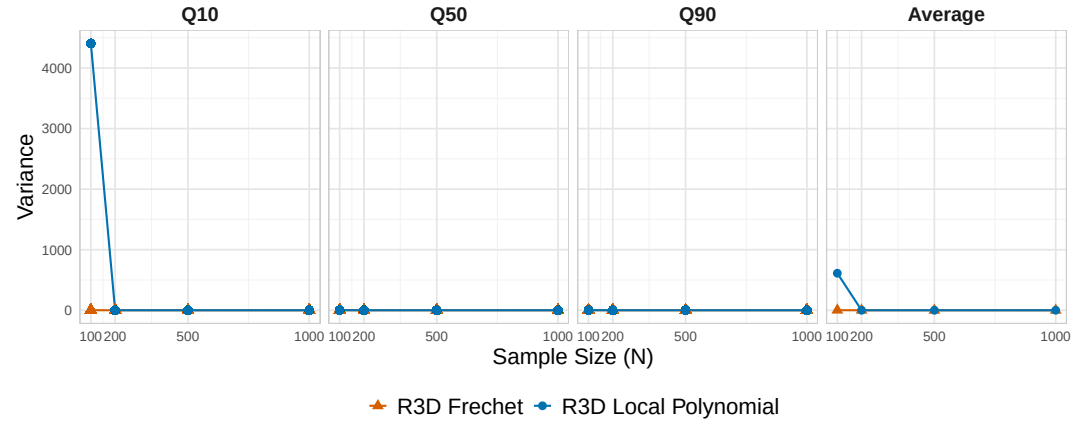
Figure 4 further supports the theoretical arguments that the Fréchet estimator is preferred over the local polynomial one, as the latter has much larger variance in small samples, though again both estimators quickly converge to a similarly low variance. I do not report results for the quantile RD as its inferential properties are irrelevant due to its inconsistency and bias in the R3D settings.

To study the coverage properties of the confidence bands and tests proposed in 2.6, I report their acceptance probabilities for both DGPs with varying values of $\Delta$ in Table 2. The values of $\Delta$ are chosen to reflect an average Cohen's d (treatment effect size relative to standard error) of $0, 0.5$, and 1 which correspond roughly to no, medium, and large treatment effects. The coverage and acceptance probabilities of the uniform confidence intervals and the homogeneity test in the first two rows are not affected by the magnitude of the treatment effect. Moreover, both the Fréchet and the local polynomial estimator rapidly converge to the correct nominal coverage level, with the Fréchet estimator exhibiting slightly better coverage. The slight undercoverage in small samples is expected insofar as the estimators are only asymptotically unbiased, as also illustrated in Figure 3. For DGP 2, which has heterogeneous treatment effects, the homogeneity test's coverage rapidly coverges to 0, illustrating the test's consistency and sharp power in finite sample. Finally, for both DGPs, the treatment nullity test also exhibits consistency and significant finite-sample power for rejecting the null hypothesis of no effect.

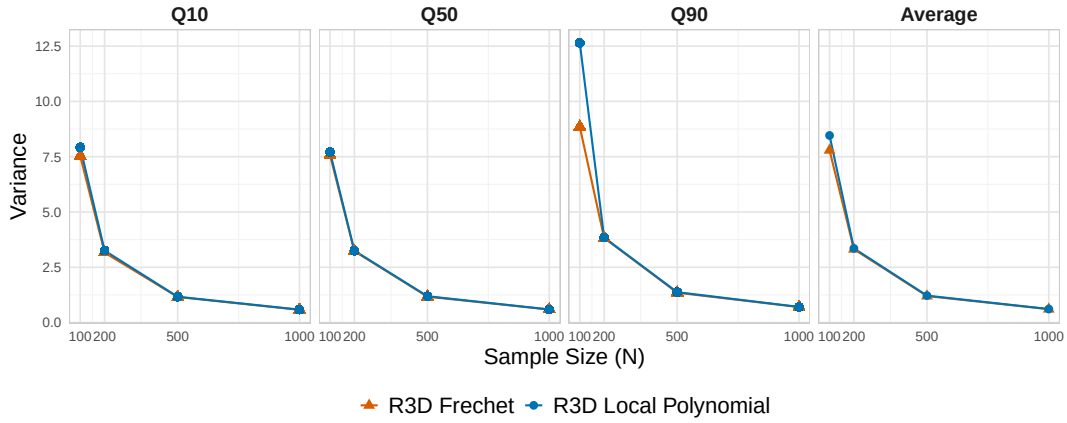
### *4.2 Empirical Illustration: State Governors and the Income Distribution*

To further illustrate the method, I estimate the effect of partisan governorship on the income distribution within US states. To that end, I deploy a classical and widely used RD design in economics and political science: the close-election design (Lee, 2008). This design compares constituencies where a political party barely won an election to those where it barely lost in order to estimate the effect of that party's win on some outcome of interest. The identification assumption is that the outcome of interest evolves smoothly with the party's vote share in

Figure 4: Simulated Variance of R3D Estimators




(a) Normal Distributions with Normal Means




(b) Normal–Exponential Distributions with Normal–Exponential Means

d

*Note*: this figure compares performance of the two R3D estimation methods in terms of the variance. Each measure is reported for $N = 100, 200, 500, 1000$ (x axis) at quantiles 10, 50, 90, and the average over all quantiles (facets). Results are averaged over $2,500$ simulations for each sample size. The methods are: 1) the local polynomial R3D estimator in Section 2.5.1; 2) the Fréchet R3D estimator in Section 2.5.2. Bandwidths are selected using (I)MSE-optimal procedure in Section A-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.

a small window around 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 close election outcomes. Such a close-election design naturally leads to an R3D setting (see also Motivating Example 2), since many outcomes of interest are 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. This produces a dataset of 356 state-year combinations where a gubernatorial election took place. Restricting the sample to data before 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 Shor et al. (2022)) 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

Table 2: Acceptance Probabilities of R3D Estimators

| Method: | **Fréchet** | | | | **Local Polynomial** | | | |
|---|---|---|---|---|---|---|---|---|
| | **DGP 1** | | | | | | | |
| | $n$ | $\Delta=0$ | $\Delta=1.14$ | $\Delta=2.27$ | $n$ | $\Delta=0$ | $\Delta=1.14$ | $\Delta=2.27$ |
| Unif. CIs | 500 | 92.56 | 92.56 | 92.56 | 500 | 92.12 | 92.12 | 92.12 |
| | 1000 | 93.28 | 93.28 | 93.28 | 1000 | 93.36 | 93.36 | 93.36 |
| | 2000 | 94.20 | 94.20 | 94.20 | 2000 | 94.40 | 94.40 | 94.40 |
| | $n$ | $\Delta=0$ | $\Delta=1.14$ | $\Delta=2.27$ | $n$ | $\Delta=0$ | $\Delta=1.14$ | $\Delta=2.27$ |
| Homogen. | 500 | 92.64 | 92.64 | 92.64 | 500 | 92.52 | 92.52 | 92.52 |
| | 1000 | 93.32 | 93.32 | 93.32 | 1000 | 92.60 | 92.60 | 92.60 |
| | 2000 | 94.20 | 94.20 | 94.20 | 2000 | 93.64 | 93.64 | 93.64 |
| | $n$ | $\Delta=0$ | $\Delta=1.14$ | $\Delta=2.27$ | $n$ | $\Delta_\mu=0$ | $\Delta=1.14$ | $\Delta=2.27$ |
| Nullity | 500 | – | 9.36 | 0 | 500 | – | 10.12 | 0 |
| | 1000 | – | 0.28 | 0 | 1000 | – | 0.28 | 0 |
| | 2000 | – | 0 | 0 | 2000 | – | 0 | 0 |
| | **DGP 2** | | | | | | | |
| | $n$ | $\Delta=0$ | $\Delta=1.86$ | $\Delta=3.72$ | $n$ | $\Delta=0$ | $\Delta=1.86$ | $\Delta=3.72$ |
| Unif. CIs | 500 | 93.20 | 93.20 | 93.20 | 500 | 92.60 | 92.60 | 92.60 |
| | 1000 | 93.12 | 93.12 | 93.12 | 1000 | 93.20 | 93.20 | 93.20 |
| | 2000 | 94.36 | 94.36 | 94.36 | 2000 | 94 | 94 | 94 |
| | $n$ | $\Delta=0$ | $\Delta=1.86$ | $\Delta=3.72$ | $n$ | $\Delta=0$ | $\Delta=1.86$ | $\Delta=3.72$ |
| Homogen. | 500 | 2.40 | 2.40 | 2.40 | 500 | 13.40 | 13.40 | 13.40 |
| | 1000 | 0 | 0 | 0 | 1000 | 3.40 | 3.40 | 3.40 |
| | 2000 | 0 | 0 | 0 | 2000 | 0.72 | 0.72 | 0.72 |
| | $n$ | $\Delta=0$ | $\Delta=1.86$ | $\Delta=3.72$ | $n$ | $\Delta=0$ | $\Delta=1.86$ | $\Delta=3.72$ |
| Nullity | 500 | – | 66.96 | 12.92 | 500 | – | 69.24 | 14.76 |
| | 1000 | – | 43.68 | 1.56 | 1000 | – | 46.96 | 1.72 |
| | 2000 | – | 16.12 | 0.04 | 2000 | – | 18.24 | 0 |

*Note*: this table shows simulated acceptance probabilities for the 95% uniform confidence bands ("Unif. CIs", probability of coverage), uniform homogeneity test ("Homogen."), and uniform treatment nullity test ("Nullity") presented in Section 2.6 for various sample sizes, where $n = n_i$ for all simulations, with $n_i$ the sample size for the empirical quantile function. Data-generating processes are described in Equations (15) and (16). All simulations used 2,500 repetitions and 5,000 bootstrap replications and estimated quantile treatment effects at the 9 deciles. Values of $\Delta$ reflects Cohen's d of 0, 0.5, and 1.

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 and magnitude decrease (see Figure A-6).

I combine these data with family-level income data from the UNICON extract of the March Current Population Survey (CPS) for the final year of the state governor's tenure, in order to capture the cumulative effect of that tenure on the income distribution. Practically, this means the election data is lagged 3 years, except in New Hampshire and Vermont, which hold gubernatorial elections every 2 years.

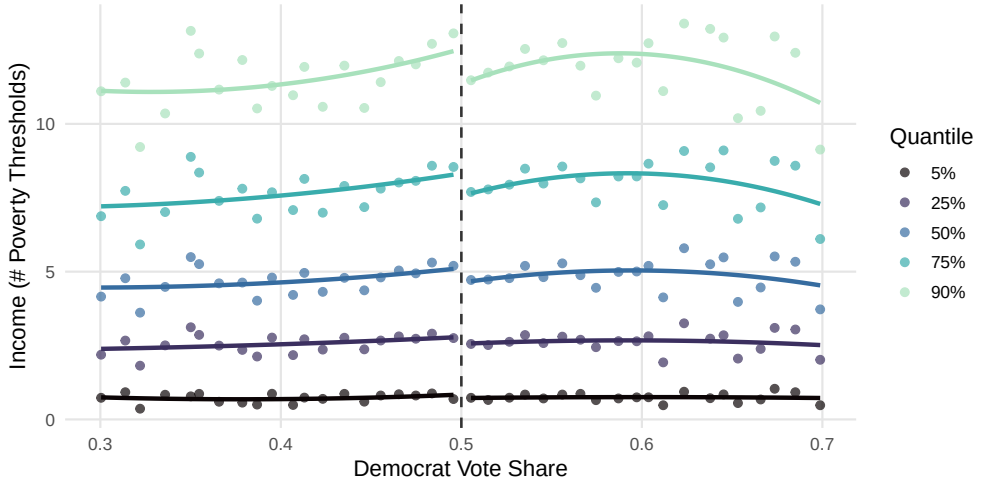
The variables in the sample are defined as follows. The running variable $X_{it}$ is the Democratic candidate's votes in state $i$ in year $t$ as a share of the combined Democratic and Republican votes. When this threshold exceeds 50%, the Democratic candidate is elected. As such, the treatment $T_{it}$ indicates whether state $i$ elected a Democratic governor in year $t$ compared to a Republican one.

The outcome variable $Z_{ijt'}$ is real income of family $j$ in state $i$ in year $t' = t + t_j$, where $t_j$ is a state-specific offset to match the income distribution in the final year of a governor's tenure to their electoral results. Real family income is constructed as the ratio of family income in year $t'$ to the federal poverty threshold in that year. Family income is defined in the standard fashion as 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. Normalizing income by the year-specific poverty threshold makes the units of the outcome variable comparable across years, thus accounting for growth in real income levels over time and making the i.i.d. assumption required for the R3D estimator more likely to hold.

The CPS data are a sample of the full census data, thus placing this application in the empirical quantile setting discussed in Section 2.7.2. In particular, instead of observing the full population income distribution, in each state $i$ in year $t$, I observe a sample of $n_i$ families $j = 1, \ldots, n_i$. Based on that, I construct the empirical income quantile functions $\widehat{Q}_{Y_{it}}$, where $Y_{it}$ is the distribution function of family income in state $i$ at time $t$ such that $Z_{ijt} \sim Y_{it}$. I use the family probability weights provided in the CPS to construct these as weighted quantile functions. Further, I winsorize the distribution at the 95th percentile to account for top-coding in the CPS. In practice, I estimate the quantile function on an equally spaced grid of 95 points between $[1 \times 10^{-6}, 0.95 + 1 \times 10^{-6}]$, where the $1 \times 10^{-6}$ offset ensures I work on a compact subset of $[0, 1]$ as required by the theoretical results.

The data are depicted in Figure 5, which shows a version of the classical RD plot

Figure 5: R3D Plot: Average Income Quantiles vs. Democrat Vote Share




*Note:* this figure shows scatterplot (blue palette) of various average quantiles of within-state income (in multiples of the federal poverty threshold), calculated within bins of width 0.01. Average quantiles were constructed by computing the weighted quantile functions of family income within each state and year, and then taking the average of the estimated quantile values for a given quantile (0.05, 0.25, etc) within the corresponding bin. These data points were then used to fit separate second-order polynomial regressions for each quantile, shown in the solid lines.

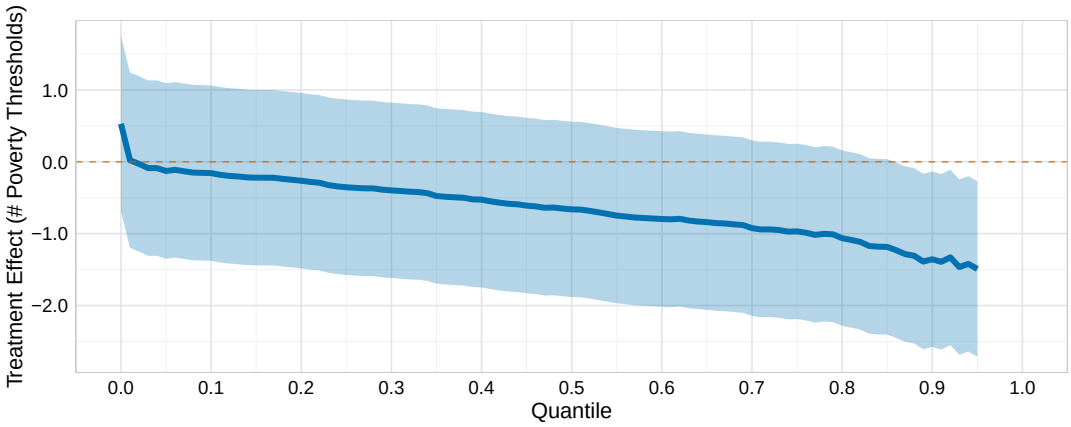
(Calonico et al., 2015*a*) appropriate for the R3D setting, similar to Figure 2. In particular, it shows a scatterplot of the "data", which are the quantile functions at various quantiles $q$, averaged within equal-width bins $B_j$ of the running variable, $\frac{1}{|B_j|}\sum_{j\in B_i}\widehat{Q}_{Y_j}(q)$, with $B_j = \{i : X_i \in [x_{j,\min}, x_{j,\max})\}$ the $j$ bins. For 5 illustrative quantiles $q$, I then fit a second-order polynomial regression line to these data. This simple descriptive plot already suggests that there is a drop in income at the higher (average) quantiles that becomes stronger as it moves up the income distribution.

Based on these data, I use the Fréchet estimator (Section 2.5.2) to estimate the local average quantile treatment effects in Definition 1, plugging in the estimated empirical quantile functions $\widehat{Q}_{Y_{it}}$. For these, 90% uniform confidence bands are constructed using the bootstrap algorithm described in A-4.3, where I use the 90% nominal level to follow the standard in the literature (Frandsen et al., 2012; Qu and Yoon, 2019; Chiang and Sasaki, 2019). To address some of the small-sample undercoverage reported in the simulations above, I apply the rule-of-thumb coverage correction of Calonico et al. (2018) to the IMSE-optimal bandwidth (see Appendix A-4.2). In addition, I formally test for uniform treatment nullity and homogeneity using the tests described in Section 2.6.

### 4.2.2 Results

The main results are shown in Figure 6. The graph depicts the LAQTE estimates, with the Y-axis indicating the effect as a multiple of the federal income threshold for the quantile of the distribution indicated by the X-axis. The light blue band depicts the 90% uniform confidence band.

Figure 6: Distributional Effects of Democratic Governor Control, 1984-2010




*Note:* this figure shows local average quantile treatment effects estimates and uniform 90% 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, in the final year of the governor's tenure, 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 2.5 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. Treatment nullity p-value: 0.038, treatment homogeneity p-value: 0.061, IMSE-optimal bandwidth: 0.22.

As shown, treatment effects are slightly positive at the lowest quantiles and become increasingly negative farther up the income distribution, with the top 10 percentiles seeing a decline in income of 1.5 times the federal poverty threshold. By contrast, the very bottom quantiles see their income increase by nearly half the poverty threshold. Only the effects for the top 10 percentiles (85th–95th) are uniformly significant at the 90% level. The p-values for the uniform treatment nullity test and the treatment heterogeneity test are 0.0376 and 0.061, respectively, suggesting the observed negative relation between income quantile and effect size is significant.

The estimated results are very similar for alternative specifications with the local polynomial estimator of Section 2.5.1, when using a uniform instead of a triangular kernel, or when using half the IMSE-optimal bandwidth in Figures A-3, A-4, and A-5. In contrast, when estimating the baseline specification with the income distribution of the same year as the

election as outcome variable, none of the quantile treatment effects are significant, nor are the nullity and homogeneity tests. This suggests the results are not driven by reverse causality, where the pre-existing income distribution drives the election outcomes. This aligns with the small effects of local economic conditions on voting behavior estimated in the literature on retrospective voting (Healy and Malhotra, 2013). Additionally, I check whether the results are not driven by families "voting with their feet" by moving across states. To that end, Figure A-8 demonstrates that the results are near-identical when excluding families that moved across state borders in the previous year, barring some expected loss of precision.

Finally, Table A-1 reports estimates of the local average treatment effect using the standard RD estimator with robust confidence bands (Calonico et al., 2014), using both the state-level weighted average family income and the raw family-level outcome data. The state-level treatment effect estimate is $-0.631$ and significant at the 90% level, while the family-level estimate is $-0.525$ and is very precisely estimated. As predicted theoretically, the state-level estimates are in line with the average of the LAQTEs produced by the R3D estimators, which is -0.647, while the family-level estimates are 15% less strong, because they do not account for the two-level sampling when using the disaggregated family-level data. Both standard RD estimates cloak the underlying heterogeneity, in particular the redistribution that is achieved at the cost of the estimated drop in overall income. I also report the quantile RD estimator of Qu and Yoon (2019) in Figure A-9. Unfortunately, the confidence bands in the companion `R` package are not currently implemented. However, in line with the simulations above, the estimated effects exhibit substantial bias, effectively precluding the need for inference. Specifically, the estimated quantile treatment effects are more than twice as small as the R3D estimates, and the corresponding average effect is only -0.158, 4 times smaller than the standard RD estimates using the aggregated data.

Taken together, these results suggest a classical equality–efficiency trade–off under Democratic governorship, with some redistribution of income achieved at the cost of a loss of income for upper-income earners (Okun, 1975). They also highlight the practical utility of the R3D estimator in uncovering distributional heterogeneity in treatment effects, compared to standard RD methods, while producing estimates that are consistent with those standard RD estimates in the aggregate.

## 5 Conclusion

This paper introduces the Regression Discontinuity Design with Distributions (R3D), a novel extension of the standard Regression Discontinuity Design (RDD) framework tailored to settings where the outcome of interest is a distribution rather than a single scalar value. This

generalization is motivated by the common real-world setting where treatment is assigned at a higher level of aggregation than the outcome of interest, such as firm-level policies that affect employees, county-level policies that affect inhabitants, or school-level policies that affect students. Standard RD methods do not apply to such settings since they do not account for the two-level randomness involved in these settings, which introduces sampling at the level of distributions. To address this, I define the local average quantile treatment effect (LAQTE) as the primary estimand, which quantifies the difference in *average* quantile functions, instead of observed ones, just above and below a treatment cutoff. This measure offers a natural and intuitive extension of the traditional RDD treatment effect to distribution-valued outcomes.

To estimate the LAQTE, I propose two complementary estimators: one based on local polynomial regression applied to random quantiles and another leveraging local Fréchet regression in 2-Wasserstein space. The local polynomial approach adapts familiar RDD techniques to handle distribution-valued data pointwise, while the Fréchet regression method treats the quantile function as a cohesive functional object, improving efficiency and finite-sample performance. Both estimators are developed for the sharp as well as the fuzzy R3D setting. I establish the asymptotic normality of both estimators and develop uniform, debiased confidence bands that can be estimated with a multiplier bootstrap. Additionally, I introduce a data-driven bandwidth selection procedure for functional outcomes. Simulations confirm the robustness of these theoretical properties, demonstrating good finite-sample performance and reliable coverage of the confidence bands.

The practical utility of the R3D framework is illustrated through an empirical application examining the effect of gubernatorial party control on within-state income distributions in the United States, using a close-election RDD. The findings reveal a classical equality–efficiency trade-off under Democratic governorship, with some redistribution achieved at the cost of an overall loss of income. In particular, incomes at the top of the distribution decline, while slight but not statistically significant improvements are observed at the lower end of the distribution. This evidence underscores the method's ability to uncover nuanced distributional impacts that scalar-based approaches might overlook. Moreover, the implied average effect is in line with standard RD estimates at the aggregate state level, unlike quantile RD methods, which estimate effects that are up to 4 times smaller in magnitude.

There are several avenues for future research. The R3D framework could be extended to allow for covariates (Jin et al., 2025; Frölich and Huber, 2019), multiple running variables or cutoffs (Bertanha, 2020; Gunsilius, 2023; Cheng, 2023; Cattaneo et al., 2016), or multivariate outcome distributions (Chen and Müller, 2023). Further, applying these methods to empirical domains where the R3D setting occurs frequently, such as education, labor policy, or politics, promises to yield new insights into the distributional consequences of policy

interventions.

In summary, the R3D framework offers a powerful and versatile new tool for causal inference with functional outcomes, making the estimation of distributional treatment effects practical in a novel but commonly occurring setting. By providing both theoretical foundations and practical estimation strategies, this article equips researchers with a new way to address pressing questions about how policies shape distributions.

# References


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

Agueh, M. and Carlier, G. (2017), 'Vers un théorème de la limite centrale dans l'espace de wasserstein?', *Comptes Rendus. Mathématique* **355**(7), 812–818.

Armstrong, T. B. and Kolesár, M. (2018), 'Optimal inference in a class of regression models', *Econometrica* **86**(2), 655–683.

Athey, S. and Imbens, G. W. (2006), 'Identification and inference in nonlinear difference-in-differences models', *Econometrica* **74**(2), 431–497.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972), *Statistical Inference Under Order Restrictions*, Wiley, New York.

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

Benyamini, Y. and Lindenstrauss, J. (1998), *Geometric nonlinear functional analysis*, Vol. 48, American Mathematical Soc.

Bertanha, M. (2020), 'Regression discontinuity design with many thresholds', *Journal of Econometrics* **218**(1), 216–241.

Borusyak, K. and Kolerman-Shemer, M. (2024), 'Regression discontinuity aggregation, with an application to the union effects on inequality', *arXiv preprint arXiv:2501.00428* .

Callaway, B., Li, T. and Oka, T. (2018), 'Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods', *Journal of Econometrics* **206**(2), 395–413.

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 Farrell, M. H. (2022), 'Coverage error optimal confidence intervals for local polynomial regression', *Bernoulli* **28**(4), 2998–3022.

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

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2015*a*), 'Optimal data-driven regression discontinuity plots', *Journal of the American Statistical Association* **110**(512), 1753–1769.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2015*b*), 'rdrobust: An r package for robust

nonparametric inference in regression-discontinuity designs'.

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.

Carlier, G., Eichinger, K. and Kroshnin, A. (2021), 'Entropic-wasserstein barycenters: Pde characterization, regularity, and clt', *SIAM Journal on Mathematical Analysis* **53**(5), 5880–5914.

Cattaneo, M. D., Keele, L., Titiunik, R. and Vazquez-Bare, G. (2021), 'Extrapolating treatment effects in multi-cutoff regression discontinuity designs', *Journal of the American Statistical Association* **116**(536), 1941–1952.

Cattaneo, M. D. and Titiunik, R. (2022), 'Regression discontinuity designs', *Annual Review of Economics* **14**(1), 821–851.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G. and Keele, L. (2016), 'Interpreting regression discontinuity designs with multiple cutoffs', *The Journal of Politics* **78**(4), 1229–1248.

Caughey, D., Xu, Y. and Warshaw, C. (2017), 'Incremental democracy: The policy effects of partisan control of state government', *The Journal of Politics* **79**(4), 1342–1358.

Chen, H., Chiang, H. D. and Sasaki, Y. (2020), 'Quantile treatment effects in regression kink designs', *Econometric Theory* **36**(6), 1167–1191.

Chen, H. and Müller, H.-G. (2023), 'Sliced wasserstein regression', *arXiv preprint arXiv:2306.10601* .

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

Chen, Y. and Müller, H.-G. (2022), 'Uniform convergence of local fréchet regression with applications to locating extrema and time warping for metric space valued trajectories', *The Annals of Statistics* **50**(3), 1573–1592.

Cheng, A. (2023), *Estimation of Regression Discontinuity and Kink Designs with Multiple Running Variables*, SSRN.

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

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2013), 'Inference on counterfactual distributions', *Econometrica* **81**(6), 2205–2268.

Chernozhukov, V. and Hansen, C. (2005), 'An iv model of quantile treatment effects', *Econometrica* **73**(1), 245–261.

Chetverikov, D., Santos, A. and Shaikh, A. M. (2018), 'The econometrics of shape restrictions', *Annual Review of Economics* **10**(1), 31–63.

Chiang, H. D., Hsu, Y.-C. and Sasaki, Y. (2019), 'Robust uniform inference for quantile treat-

ment effects in regression discontinuity designs', *Journal of Econometrics* **211**(2), 589–618.

Chiang, H. D. and Sasaki, Y. (2019), 'Causal inference by quantile regression kink designs', *Journal of Econometrics* **210**(2), 405–433.

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

de Benedictis-Kessner, J. and Warshaw, C. (2016), 'Mayoral partisanship and municipal fiscal policy', *The Journal of Politics* **78**(4), 1124–1138.

Dubey, P. and Müller, H.-G. (2019), 'Fréchet analysis of variance for random objects', *Biometrika* **106**(4), 803–821.

Dümbgen, L. (2024), 'Shape-constrained statistical inference', *Annual Review of Statistics and Its Application* **11**.

Dynes, A. M. and Holbein, J. B. (2020), 'Noisy retrospection: The effect of party control on policy outcomes', *American Political Science Review* **114**(1), 237–257.

Fan, J. (1993), 'Local linear regression smoothers and their minimax efficiencies', *The annals of Statistics* pp. 196–216.

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. (1995), 'Adaptive order polynomial fitting: bandwidth robustification and bias reduction', *Journal of Computational and Graphical Statistics* **4**(3), 213–227.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Fan, J. and Müller, H.-G. (2022), 'Conditional distribution regression for functional responses', *Scandinavian Journal of Statistics* **49**(2), 502–524.

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

Fang, Z. and Seo, J. (2021), 'A projection framework for testing shape restrictions that form convex cones', *Econometrica* **89**(5), 2439–2458.

Firpo, S. (2007), 'Efficient semiparametric estimation of quantile treatment effects', *Econometrica* **75**(1), 259–276.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009), 'Unconditional quantile regressions', *Econometrica* **77**(3), 953–973.

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

Fréchet, M. (1948), Les éléments aléatoires de nature quelconque dans un espace distancié, *in* 'Annales de l'institut Henri Poincaré', Vol. 10, pp. 215–310.

Frölich, M. and Huber, M. (2019), 'Including covariates in the regression discontinuity de-

sign', *Journal of Business & Economic Statistics* **37**(4), 736–748.

Ghodrati, L. and Panaretos, V. M. (2022), 'Distribution-on-distribution regression via optimal transport maps', *Biometrika* **109**(4), 957–974.

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

Groeneboom, P. and Jongbloed, G. (2014), *Nonparametric estimation under shape constraints*, number 38, Cambridge University Press.

Gunsilius, F. F. (2023), 'Distributional synthetic controls', *Econometrica* **91**(3), 1105–1117.

Gunsilius, F. F. (2025), 'A primer on optimal transport for causal inference with observational data', *arXiv preprint arXiv:2503.07811* .

Gunsilius, F., Hsieh, M. H. and Lee, M. J. (2024), 'Tangential wasserstein projections', *Journal of Machine Learning Research* **25**(69), 1–41.

Gunsilius, F. and Van Dijcke, D. (2025), 'Free discontinuity regression: With an application to the economic effects of internet shutdowns', *arXiv preprint arXiv:2309.14630* .

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.

Healy, A. and Malhotra, N. (2013), 'Retrospective voting reconsidered', *Annual review of political science* **16**(1), 285–306.

Hibbs, D. A. (1977), 'Political parties and macroeconomic policy', *American Political Science Review* **71**(4), 1467–1487.

Iao, S. I., Zhou, Y. and Müller, H.-G. (2024), 'Deep fréchet regression', *arXiv preprint arXiv:2407.21407* .

Imbens, G. and Kalyanaraman, K. (2012), 'Optimal bandwidth choice for the regression discontinuity estimator', *The Review of economic studies* **79**(3), 933–959.

Jin, Z., Zhang, Y., Zhang, Z. and Zhou, Y. (2025), 'Identification and inference in a quantile regression discontinuity design under rank similarity with covariates', *Econometric Theory* **41**(1), 172–217.

Katta, S., Parikh, H., Rudin, C. and Volfovsky, A. (2024), Interpretable causal inference for analyzing wearable, sensor, and distributional data, *in* 'International Conference on Artificial Intelligence and Statistics', PMLR, pp. 3340–3348.

Kelly, N. J. (2009), *The Politics of Income Inequality in the United States*, Cambridge University Press, Cambridge, UK.

Koenker, R. and Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica: journal of the Econometric Society* pp. 33–50.

Kroshnin, A., Spokoiny, V. and Suvorikova, A. (2021), 'Statistical inference for bures–wasserstein barycenters', *The Annals of Applied Probability* **31**(3), 1264–1298.

Kurisu, D., Zhou, Y., Otsu, T. and Müller, H.-G. (2024), 'Geodesic causal inference', *arXiv preprint arXiv:2406.19604* .

Kurisu, D., Zhou, Y., Otsu, T. and Müller, H.-G. (2025), 'Geodesic synthetic control methods for random objects and functional data', *arXiv preprint arXiv:2505.00331* .

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

Lee, D. S. and Lemieux, T. (2010), 'Regression discontinuity designs in economics', *Journal of economic literature* **48**(2), 281–355.

Leigh, A. (2008), 'Estimating the impact of gubernatorial partisanship on policy settings and economic outcomes: A regression discontinuity approach', *European Journal of Political Economy* **24**(1), 256–268.

Lin, Z., Kong, D. and Wang, L. (2023), 'Causal inference on distribution functions', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **85**(2), 378–398.

Linton, O. and Nielsen, J. P. (1994), 'A multiplicative bias reduction method for nonparametric regression', *Statistics & Probability Letters* **19**(3), 181–187.

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.

Nishimura, H. and Ok, E. A. (2012), 'Solvability of variational inequalities on hilbert lattices', *Mathematics of Operations Research* **37**(4), 608–625.

Okun, A. M. (1975), *Equality and Efficiency: The Big Tradeoff*, Brookings Institution Press, Washington, D.C.

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

Papay, J. P., Willett, J. B. and Murnane, R. J. (2011), 'Extending the regression-discontinuity approach to multiple assignment variables', *Journal of Econometrics* **161**(2), 203–207.

Petersen, A., Liu, X. and Divani, A. A. (2021), 'Wasserstein f-tests and confidence bands for the fréchet regression of density response curves', *The Annals of Statistics* **49**(1), 590–611.

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

Phelps, R. (1978), 'Gaussian null sets and differentiability of lipschitz map on banach spaces', *Pacific Journal of Mathematics* **77**(2), 523–531.

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

Qiu, R., Yu, Z. and Zhu, R. (2024), 'Random forest weighted local fréchet regression with

random objects', *Journal of Machine Learning Research* **25**(107), 1–69.

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. and Yoon, J. (2024), 'Qte. rd: An r package for quantile treatment effects in regression-discontinuity designs with/without covariates'.

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.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer, New York.

Reed, W. R. (2006), 'Democrats, republicans, and taxes: Evidence that political parties matter', *Journal of Public Economics* **90**(4–5), 725–750.

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

Schötz, C. (2022), 'Nonparametric regression in nonstandard spaces', *Electronic Journal of Statistics* **16**(2), 4679–4741.

Shor, B., McCarty, N. et al. (2022), 'Two decades of polarization in american state legislatures', *Journal of Political Institutions and Political Economy* **3**(3–4), 343–370.

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

Thistlethwaite, D. L. and Campbell, D. T. (1960), 'Regression-discontinuity analysis: An alternative to the ex post facto experiment.', *Journal of Educational psychology* **51**(6), 309.

Torous, W., Gunsilius, F. and Rigollet, P. (2024), 'An optimal transport approach to estimating causal effects via nonlinear difference-in-differences', *Journal of Causal Inference* **12**(1), 20230004.

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*

Van Dijcke, D., Gunsilius, F. and Wright, A. (2024), 'Return to office and the tenure distribution', *arXiv preprint arXiv:2405.04352* .

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

Zajíček, L. (1983), 'Differentiability of the distance function and points of multi-valuedness of the metric projection in banach space', *Czechoslovak Mathematical Journal* **33**(2), 292–308.

Zhou, Y., Kurisu, D., Otsu, T. and Müller, H.-G. (2025), 'Geodesic difference-in-differences', *arXiv preprint arXiv:2501.17436* .

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

The aim is 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:

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

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 the regression function at $X = 0$ is the *intercept* component, let $e_0 = (1,\ 0,\ 0, \ldots, 0)^\top$ as before. Then

$$\hat{m}_{\pm,p}(q) = e_0^\top \hat{\boldsymbol{\alpha}}_{\pm,p} = e_0^\top \big(\mathbf{X}_{\pm}^\top \mathbf{W}_{\pm} \mathbf{X}_{\pm}\big)^{-1} \big(\mathbf{X}_{\pm}^\top \mathbf{W}_{\pm} \mathbf{Q}\big).$$

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., it follows:

$$\hat{m}_{\pm,p}(q) = \sum_{i=1}^{n} \underbrace{e_0^\top \left(\mathbf{X}_\pm^\top \mathbf{W}_\pm \mathbf{X}_\pm\right)^{-1} \left(\mathbf{X}_\pm^\top \mathbf{W}_\pm\right)_{:,i}}_{=:\, s^{(p)}_{\pm,in}(h)} Q_{Y_i}(q).$$

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

$$\hat{m}_{\pm,p}(q) = \sum_{i=1}^{n} \Big[ s^{(p)}_{\pm,in}(h) \Big] Q_{Y_i}(q), \quad \text{where}$$
$$s^{(p)}_{\pm,in}(h) = \delta_i^{\pm}\, e_0^\top \Big(\mathbf{X}_\pm^\top \mathbf{W}_\pm \mathbf{X}_\pm\Big)^{-1} r_p\Big(\tfrac{X_i - c}{h}\Big)\, K\Big(\tfrac{X_i - c}{h},\Big).$$

## A-3 Overview of Local Fréchet Regression

This supplement provides a brief overview of local Fréchet regression as proposed in Petersen and Müller (2019).

### *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 $Z \in \mathbb{R}$, then the conditional expectation $E[Z \mid X = x]$ at $x$ can be defined as the unique minimizer $f$ of the mean squared error,

$$E[Z \mid X = x] \coloneqq \underset{f \in \mathbb{R}}{\operatorname{argmin}}\, E[d_E\,(Z, f)^2 \mid X = x],$$

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 $(\Omega, d)$ 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),

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

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*

Consider using the 2-Wasserstein distance $d_{W_2}(Y_1, Y_2)$ to measure the distance between two distribution functions $Y_1, Y_2 \in \mathcal{Y}$. For one-dimensional distribution functions, this metric can be shown to equal (Villani, 2021, Theorem 2.18),

$$d^2_{W_2}(Y_1, Y_2) = \int_0^1 \left(Q_{Y_1}(q) - Q_{Y_2}(q)\right)^2 \mathrm{d}q, \tag{A-2}$$

where remember that $Q_{Y_1}$ and $Q_{Y_2}$ are the quantile functions corresponding to $Y_1$ and $Y_2$, respectively.

The reason for the asymptotic equivalence between Fréchet regression in 2-Wasserstein space $(\mathcal{Y}, d_{w_2})$ and local polynomial regression on quantiles is that the Fréchet mean of any random distribution $Y \in \mathcal{Y}$ equipped with $d_{W_2}$ is the unique cdf $F_\oplus$ with the quantile function (Panaretos and Zemel, 2020, Theorem 3.2.11),

$$Q_{F_\oplus}(q) = EQ_Y(q) = \int_{\mathcal{Y}} Q_Y(q)\, \mathrm{d}P(Y).$$

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.

### *A-3.3 Local Fréchet Regression*

Fréchet regression was introduced in Petersen and Müller (2019) as a generalization of linear regression (Fan and Gijbels, 1996) when the outcome $Y$ takes values in a general metric space $\Omega$ beyond just the Euclidean space $\mathbb{R}$.

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

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

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 $\beta_0^*$ can be written as,

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

with weight function,

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

which corresponds to the local Fréchet mean,

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

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

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

These weights are identical to those for the classical local polynomial regression (Fan and Gijbels, 1996). The generalization lies in the use of the distance metric and the projection onto $\Omega$.

Local Fréchet regression "from the left and right", as considered in the main text, simply requires adding a $\delta_i^{\pm}$ term to the appropriate equations.

### *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 $(\mathcal{Y}, d_{W_2})$.

**Proposition A-2.** *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 $(\mathcal{Y}, d_{W_2})$,*

$$\operatorname*{argmin}_{\omega \in \mathcal{Y}} \frac{1}{n} \sum_{i=1}^{n} s^{(p)}_{\pm,in}(h) d^2_{W_2}(\omega, Y_i),$$

*which exists and is unique. Similarly, the projected conditional average quantile function $\Pi_Q(m_\pm)$ is equivalent to the quantile function of the conditional Fréchet mean on $(\mathcal{Y}, d_{W_2})$ 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 $\hat{m}_{+,p}$, I have,

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

where the second equality follows from $\frac{1}{n} \sum_{i=1}^{n} s^{(p)}_{+,in}(h) = 1$ and the definition of $\hat{m}_{+,p}$ and the third equality follow from $\frac{1}{n} \sum_{i=1}^{n} s^{(p)}_{+,in}(h) = 1$. As a result, the local Fréchet regression

estimator "from the right", $\hat{l}_{+,\oplus}(c)$ on $(\mathcal{Y}, d_{W_2})$ equals

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

where $Q^{-1}$ is the quantile function's generalized inverse, which maps it back to its corresponding CDF. An identical argument holds for the Fréchet estimator from the left, $\hat{l}_{-,\oplus}$. Thus, the quantile function of the local Fréchet regression estimator is the $L^2$ projection of the 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 $Q(\mathcal{Y})$.

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 $(\mathcal{Y}, d_{W_2})$ as,

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

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 $Q_\omega$, and the second equality follows from Fubini-Tonelli by the fact that all distributions in $\mathcal{Y}$ have finite variance. As a result, the conditional Fréchet mean $m_\oplus$ on $(\mathcal{Y}, d_{W_2})$ equals

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

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*

Let $t \leq p$, $t \in \mathbb{N}_+$ and denote $\delta_x^{\pm} \coloneqq 1\{x \gtrless 0\}$. 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 $[\underline{c},\overline{c}] \times \mathcal{Y} \times \{0,1\}$. By the coerciveness of the projection onto quantile functions (Bauschke et al., 2017), 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 $\Psi_s^{\pm} = \int_{\mathbb{R}_\pm} r_s(u) r_s'(u) K^2(u) du$. Below, I drop the R3D superscript on the treatment effect estimators $\hat{\tau}^{\text{R3D}}$ to ease notation.

### 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,

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

Similar derivations for the variance expressions imply that

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

where I remind the reader that $\sigma_{1,1}(q, q \mid 0^+) = \lim_{x \to 0^{\pm}} \text{Var}\big(Q_Y(q) \mid X = x\big)$. Summing these expressions for both sides for $\hat{\tau}_s(q)$ (the difference) yields

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

Based on the standard bias-variance expression MSE = Bias$^2$ + Var, these expressions allow one 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 $\sigma_{2,2}(q,q|0^\pm) = \lim_{x \to 0^\pm} \mathrm{Var}(T|X=x)$ in the formulas above and noting that the optimal bandwidth will be the same for all $q$.

### A-4.2.2 Fréchet Estimator

Define

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

Putting the above expansions together, it follows that

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

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 of the local polynomial estimator onto quantile functions, and the corresponding treatment effect estimator $\hat{\tau}_\oplus(q)$. Write the difference between the treatment effect estimators as

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

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 $q$,

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

By the Cauchy–Bunyakovsky–Schwarz inequality and the previously established convergence results, this difference is $o_p(1/nh)$. Hence

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

As a result, the leading terms of the IMSEs of the local polynomial and the Fréchet estimator are the same, and thus,

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

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 Fréchet 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):

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

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

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

$$h^0_{1,n}(q) = \left( \frac{1}{2(p+1)} \frac{C_{1,0}(q)'}{C_{1,0}(q)^2} \right)^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}},$$

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 $\overline{\sigma}^2_{1,\pm}(q)$ 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, $h^0_{2,n}(q)$ 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 $\{\, h^0_{1,n}(q),\, h^0_{2,n}(q)\}$ from Step 1, run local polynomial regressions of order $s$ at each quantile $q$,

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

which gives the first-stage estimates

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

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,

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

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, one can optionally apply the rule-of-thumb bandwidth algorithm from Calonico et al. (2018, 2020) for optimal coverage error to 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 $Y_i \in \mathcal{Y}$ (distributional outcome), $T_i \in \{0,1\}$, and running variable $X_i \in \mathbb{R}$ with cutoff normalized to 0.
- A finite grid of $M$ quantiles $q_j$, $\mathcal{T}^* \coloneqq (q_1,\ldots,q_M) \subset [a,b] \subset (0,1)$.
- A chosen local polynomial order $p$.
- A kernel function $K$ and bandwidth $h>0$. 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 samples $\{Z_{ij}\}_{j=1}^{n_i} \sim Y_i$ based on (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

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

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

(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 $\hat{m}_{\pm,T,p} = \sum_{i=1}^{n} s_{\pm,in}^{(p)}(h)\, T_i$, 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 $\hat{m}_{\pm,p}$ onto the space of monotone functions through the isotonic regression:

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

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 $g_1(Y_i, q) = Q_{Y_i}(q_j)$, $g_2(Y_i, T_i) = T_i$, 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 $k \in \{1, 2\}$, compute

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

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 $b$:

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

(ii) (Fuzzy RDD) For each $b$:

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

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.

After carrying out step 2–4 for each $q_j \in \mathcal{T}^*$, do:

**Step 5: Compute the critical value and construct bands.**

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

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

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

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

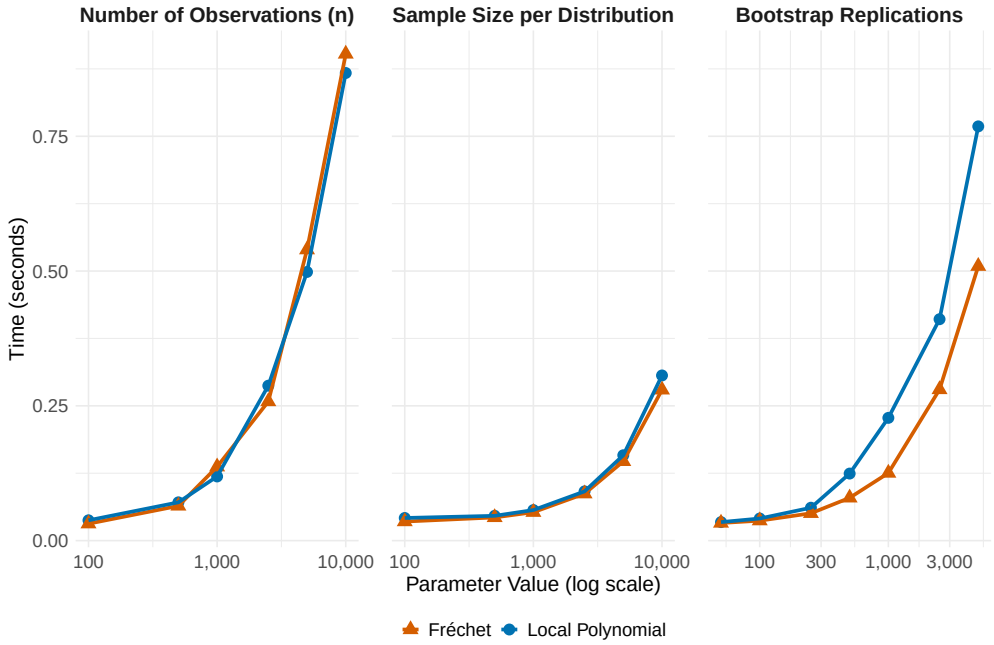
$$\Big[\hat{\tau}_p(q) \; \pm \; \tfrac{1}{\sqrt{nh}} \, \hat{c}_n^B(a, b; \lambda)\Big], \quad \text{for } q \in \mathcal{T}^*.$$

### *A-4.4 Computational Details*

An R implementation of the package can be found at https://davidvandijcke.com/R3D. The main polynomial weights estimation was implemented with a Fortran backend, leading

to highly performant code, as illustrated in Figure A-1. For example, the model with 5 million total observations evaluated at 20 quantiles and 100 bootstrap repetitions solves in less than a second. 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 code can use optimized vector operations with one single bandwidth in the Fréchet case. 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.* It holds that,

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

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.* It holds that,

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

where the first equality follows from the definition of $T^1$ and $T^0$, the continuity assumption I4 and the zero-measure indefinites assumption in I5. The third equality follows from the law of total expectation and Assumption I5. Again, by Assumptions I1 and I5,

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

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

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

Combining these two derivations with Assumption I3 gives the result. □

## *A-5.2 Asymptotic Results*

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

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

Then the family $\mathcal{F} = \{\, Y \to Q_Y(q) \colon q \in \Theta\}$ 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

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

so membership in $G_{f_q}$ depends only on the scalar value $Y(z)$. A single point $(Y, z)$ can be shattered: include it by setting $q \leq Y(z)$, exclude it by setting $q > Y(z)$. However, 2 points cannot be shattered. To see this, let $(Y_1, z_1)$ and $(Y_2, z_2)$ have $Y_1(z_1) \leq Y_2(z_2)$. The subsets $\emptyset, \{(Y_2, z_2)\}$, and $\{(Y_1, z_1), (Y_2, z_2)\}$ can be realized by choosing $q > Y_2(z_2), q \in [Y_1(x_1), Y_2(z_2)]$, and $q \leq Y_1(z_1)$, respectively. However, the subset $\{(Y_1, z_1)\}$ 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 $F(Y) = \sup_{q\in\Theta} |Q_Y(q)|$. Then $F$ is a measurable envelope for the class of functions $f_q(Y)$ in $L^r(P)$ for any probability measure $P$ on $(\Omega^x, \mathcal{F}^x)$ and any $r \geq 1$, since each $Q_Y$ 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

$$g_q(x) = E\big[\, Q_Y(q) \;\mid\; X = x\big].$$

Denote this family by

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

I claim $\mathcal{F}$ is a VC-subgraph class of finite index. Indeed, by the results in Proposition A-2, each $g_q(\cdot)$ can be identified with a one-dimensional quantile function: specifically, there is a

conditional Fréchet mean $m_\oplus(x) \in \mathcal{Y}$, as defined in (A-1), such that

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

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

$$G_{g_q} = \big\{ (x,z) : z \leq Q_{m_\oplus(x)}(q) \big\} = \big\{ (x,z) : m_\oplus(x)(z) \geq q \big\}.$$

Since each $x \in \mathbb{R}$ defines a distinct, unique cdf $m_\oplus(x) \in \mathcal{Y}$ by Proposition A-2, 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 the sharp R3D setting, and $\mu_2(x,\theta_2) = E[T \mid X = x]$ in the fuzzy R3D 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 $\Upsilon, \psi, \phi$ 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 $\underline{c} < 0 < \overline{c}$. (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 $\int_{\mathcal{Y}\times[\underline{c},\overline{c}]} |F_{\mathcal{E}}(y,x)|^{2+\epsilon} \, d\mathrm{P}^x(y,x) < \infty$ for some $\epsilon > 0$. (b) This part is equivalent to Assumption L2 (i). (c) For any $(q,k),(q',l) \in [a,b] \times \{1,2\}$, it holds that $\sigma_{kl}(q,q'|\cdot) \in C^1([\underline{c},\overline{c}] \setminus \{0\})$ with bounded derivatives in $x$ and $\sigma_{kl}(q,q' \mid 0^\pm) < \infty$. (d) For each $Y \in \mathcal{Y}$, $Q_Y(q)$ 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 $\sup_{q\in[a,b]} |Q_Y(q)| \leq F_{\mathcal{E}}(y,x)$ and $\sup_{q\in[a,b]} |E[Q_Y(q) \mid X = x]| \leq F_{\mathcal{E}}(y,x)$. Moreover, by Assumption L3,

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

- *(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 in $C^1([\underline{c},\overline{c}] \setminus \{0\})$ 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 I 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,

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

uniformly for all $q \in [a,b]$. 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), $\nu_n^+(q,k) = \sum_{i=1}^n \left[f_{ni}(q,k) - Ef_{ni}(q,k)\right]$ converges weakly to a tight zero-mean Gaussian process $\mathbb{G}_{H^+}$ with covariance function $H^+$ defined in the main theorem, where,

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

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

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

is Hadamard differentiable tangentially to $\ell^\infty[a,b]$ on the subset where $g_+ - g_- \neq 0$, which holds by I5 (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 $\psi, \phi, \Upsilon$ are trivially Hadamard differentiable in my setting, and by Assumptions I5 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. Thus, their Theorem 2 follows. The final result obtains by combining it with Theorem 1 by plugging in $\hat{m}_{\pm,p}(\cdot), \hat{m}_{T,\pm,p}(\cdot)$ for $m_\pm(\cdot), m_{T,\pm}(\cdot)$. □

**Proof of Theorem 4**

*Proof.* The proof relies on two main arguments. First, I establish that the unprojected estimator converges uniformly to a limit process with continuous sample paths. Second, I use this first result to establish the Hadamard differentiability of the projection operator tangentially to $C([a,b])$, implying asympotic normality of the projected estimator by the functional delta method.

From Theorem 1, we know the unprojected estimator converges weakly to a tight Gaussian process, $\mathbb{G}_\pm$, in $\ell^\infty([a,b])$. I now strengthen this result to show that the limit resides in $C([a,b])$. By van der Vaart (2000, Thm. 18.14 and Lemma 18.15), a tight Gaussian limit on a compact set has a version with almost surely continuous sample paths with respect to its intrinsic metric, $\rho(s,t) = \text{sd}(\mathbb{G}_\pm(s) - \mathbb{G}_\pm(t))$. To show this implies continuity in the standard topology, I verify stochastic continuity. The covariance function $\Gamma(q,q') = E[Q_Y(q)Q_Y(q') \mid X = x] - m_\pm(q)m_\pm(q')$ is continuous in $(q,q')$: for $(q_n,q_n') \to (q,q')$, $Q_Y(q_n)Q_Y(q_n') \to Q_Y(q)Q_Y(q')$ pointwise a.s. (by monotonicity and càdlàg property of quantiles) and is bounded by $(\sup_q |Q_Y(q)|)^2$ which is integrable as a consequence of Assumption L3. Then, by the Dominated Convergence Theorem, the cross-moment is continuous, and hence $\Gamma$ is too. This implies mean-square continuity of $\mathbb{G}$: $\mathbb{E}[(\mathbb{G}(q) - \mathbb{G}(q'))^2] = \Gamma(q,q) + \Gamma(q',q') - 2\Gamma(q,q') \to 0$ as $q' \to q$. Mean-square continuity then yields stochastic continuity by Chebyshev's inequality: $\mathbb{P}(|\mathbb{G}(q) - \mathbb{G}(q')| > \epsilon) \leq \mathbb{E}[(\mathbb{G}(q) - \mathbb{G}(q'))^2]/\epsilon^2 \to 0$. Thus, convergence of the unprojected estimator is in $C([a,b])$,

$$\sqrt{nh}(\hat{m}_{\pm,p} - m_\pm) \rightsquigarrow \mathbb{G}_\pm \quad \text{in the space } C([a,b]). \tag{A-10}$$

Then, form the projected estimator,

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

where $\Pi_{\mathcal{Q}}$ is now the metric projection from functions in $C([a,b])$ 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, I want to check that $\Pi_{\mathcal{Q}}$ has a suitable Hadamard derivative at $m_{\pm}$ which acts as the identity map. To that end, I distinguish two cases.

**Case 1: $m_{\pm}$ is strictly monotone.**

*Step 1. Lipschitz Continuity of the Projection Operator.* I analyze the projection operator $\Pi_{\mathcal{Q}}$ as a map between the space where the limit process resides, $C([a,b])$, and the space where the projection is defined, $L^2([a,b])$. That is, I consider $\Pi_{\mathcal{Q}} : C([a,b]) \to L^2([a,b])$. To establish the Hadamard differentiability of this map, I first establish that it is Lipschitz.

To see this, note that, as a map on the Hilbert space $L^2([a,b])$, the operator is non-expansive by standard properties of metric projections onto convex sets (Bauschke et al., 2017, Proposition 4.16),

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

Second, for any function $f$ on the compact interval $[a,b]$, the $L^2$ norm is bounded by the uniform norm,

$$\|f\|_{L^2} \leq \sqrt{b-a} \cdot \|f\|_{\infty}.$$

For any $g_1, g_2 \in C([a,b])$, we can combine these inequalities to get,

$$\|\Pi_{\mathcal{Q}}(g_1) - \Pi_{\mathcal{Q}}(g_2)\|_{L^2} \leq \|g_1 - g_2\|_{L^2} \leq \sqrt{b-a} \cdot \|g_1 - g_2\|_{\infty}.$$

This shows that $\Pi_{\mathcal{Q}}$ is globally Lipschitz as a map from $(C([a,b]), \|\cdot\|_{\infty})$ to $(L^2([a,b]), \|\cdot\|_{L^2})$.

*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 spaces $C([a,b])$, $L^2([a,b])$, there is no Lebesgue measure. Instead, Preiss (2014) showed that any Lipschitz map $T\colon X \to Y$ between a separable Banach space $X$ and a Banach space $Y$ with the Radon-Nikodym property admits Hadamard derivatives at all but an exceptional set $N$, where $N$ belongs to the $\sigma$-ideal of directionally porous sets (as introduced in Zajíček (1983)).

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

The final step is to connect this topological notion of smallness to a probabilistic one. By Theorem 6.42 and the discussion on p.164 in Benyamini and Lindenstrauss (1998) any $\sigma$-directionally porous set in a separable Banach space is Gauss-null (as defined in Phelps (1978)). By definition, a Gauss-null set has measure zero under any non-degenerate Gaussian measure. As a result, the probability that the process $\mathbb{G}_{\pm}$ realizes a function that lies within the non-differentiable $\mathcal{N}$ set is zero. Hence, the operator $\Pi_{\mathcal{Q}}$ is Hadamard differentiable at $m_{\pm,p}$ tangentially to $C([a,b])$ so that the functional delta method applies (van der Vaart, 2000, Theorem 20.8).

*Step 3. The derivative at $m_{\pm}$ is the identity map.* Since $m_{\pm} \in \mathcal{Q}$ (Proposition A-2), I 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 $C([a,b])$. Concretely, if $m_{\pm} + \varepsilon u \in \mathcal{Q}$ then $\Pi_{\mathcal{Q}}(m_{\pm} + \varepsilon u) - \Pi_{\mathcal{Q}}(m_{\pm})$ is just $\varepsilon u$, 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 $\hat{m}_{\pm,p} - m_{\pm} \in C([a,b]) \setminus \mathcal{N}$ 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 $D\Pi_{\mathcal{Q}}[m_{\pm}] = \mathrm{Id}$, I get

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

Then, since $\|\hat{m}_{\pm,p} - m_{\pm}\|_{L^2} = O_p\big((nh)^{-1/2}\big)$,

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

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

**Case 2: $m_\pm$ is weakly monotone.**

Suppose $m_\pm$ is only weakly increasing and hence may lie on the boundary of the cone $\mathcal{Q}$. To move it inside the interior of $\mathcal{Q}$, fix a small "hill" function $H(\cdot)$ on $[0,1]$ (e.g. strictly positive on $(0,1)$ and zero near the endpoints). For each $n$, pick a scalar $\eta_n > 0$ that shrinks to zero as $n \to \infty$ at rate $\eta_n = o\big((nh)^{-1/2}\big)$. Define the *$\eta$-perturbed* functions

$$m_\pm^\eta(\cdot) = m_\pm(\cdot) + \eta_n\, H(\cdot), \quad \hat{m}_{\pm,p}^\eta(\cdot) = \hat{m}_{\pm,p}(\cdot) + \eta_n\, H(\cdot).$$

Each $m_\pm^\eta$ is now *strictly* increasing on $(0,1)$, so $m_\pm^\eta$ is an interior point of $\mathcal{Q}$. In similar vein, consider the projected estimator

$$\hat{m}_{\pm,\oplus,p}^\eta = \Pi_{\mathcal{Q}}\Big[\hat{m}_{\pm,p}^\eta\Big],$$

which falls under the "strictly monotone" scenario of Case 1. In particular, the same Hadamard-differentiability and functional delta method argument implies that

$$\sqrt{nh}\big[\hat{m}_{\pm,\oplus,p}^\eta - m_\pm^\eta\big] \rightsquigarrow \mathbb{G}_\pm \quad \text{in } C([a,b]),$$

the same uniform limiting law as before. Then, note that the isotonic projection operator $\Pi_{\mathcal{Q}}$ is non-expansive in the uniform ($L^\infty$) norm. To see this, setting $M = \|f - g\|_\infty$ implies $g - M \le f \le g + M$, and the order-preserving property of $\Pi_{\mathcal{Q}}$ (Nishimura and Ok, 2012, Lemma 2.4) together with its translation-equivariance yield $\|\Pi_{\mathcal{Q}}(f) - \Pi_{\mathcal{Q}}(g)\|_\infty \le M$. As a result, we have

$$\big\|\hat{m}_{\pm,\oplus,p}^\eta - \hat{m}_{\pm,\oplus,p}\big\|_\infty \le \big\|\hat{m}_{\pm,p}^\eta - \hat{m}_{\pm,p}\big\|_\infty = \eta_n\, \|H\|_\infty,$$

and $\big\|m_\pm^\eta - m_\pm\big\|_\infty = \eta_n \|H\|_\infty$. Hence, the difference between the perturbed and unperturbed scaled processes vanishes uniformly:

$$\Big\|\sqrt{nh}\big(\hat{m}_{\pm,\oplus,p}^\eta - m_\pm^\eta\big) - \sqrt{nh}\big(\hat{m}_{\pm,\oplus,p} - m_\pm\big)\Big\|_\infty = \sqrt{nh} \cdot O_p(\eta_n),$$

which vanishes in probability since $\eta_n = o\big((nh)^{-1/2}\big)$. By Slutsky's lemma, it follows that

$$\sqrt{nh}\,\big[\hat{m}_{\pm,\oplus,p} - m_\pm\big] \rightsquigarrow \mathbb{G}_\pm \quad \text{in } C([a,b])$$

also holds in the weakly-monotone case. This completes the proof. □

**Proof of Corollary 1.**

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

**Proof of Proposition 1.**

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

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

I have that,

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

uniformly over $q \in [a,b]$. The last equality follows from Assumption Q1 and Corollary 21.5 in van der Vaart (2000) which guarantees the uniform convergence of each empirical quantile function $\widehat{Q}_{Y_i}$ 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,

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

An identical argument to the one in Proposition A-2 shows that the quantile function of this estimator is,

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

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 $\bar{m}_{\pm,p}(\cdot) - m_{\pm}(\cdot)$, with the only difference that $m_{\pm}$ is always strictly increasing under Assumption Q1 so that we only need to consider Case 1 from that proof. 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,*

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

*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 a point in the convex set $\mathcal{Q}$ of quantile functions. Assume first that it is an interior point (strictly increasing). 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 $\|r_n(\epsilon)\|_{L^2}/\|\epsilon\|_{L^2} \to 0$ whenever $\|\epsilon\|_{L^2} \to 0$, such that

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

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

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

Hence

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

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

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

The argument for the case where $m_\pm$ is a boundary point of $\mathcal{Q}$ follows analogously as in the proof of Theorem 4 by perturbing the quantile functions to make them strictly monotonic and then letting the perturbation go to 0 at rate $o_p((nh)^{-1/2})$.

□

# A-6 Additional Results

## *A-6.1 Tables*

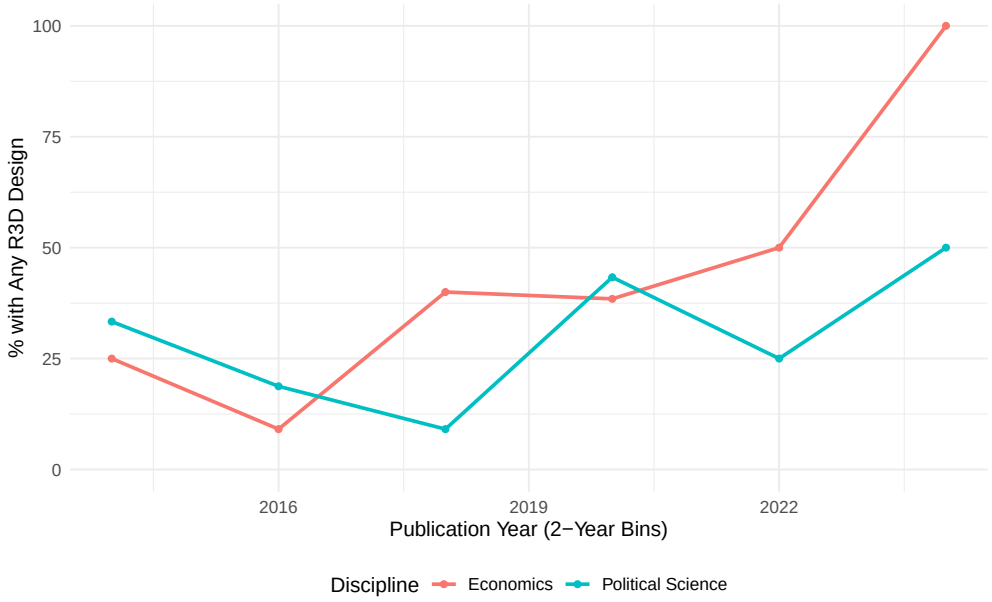
Table A-1: Canonical RD Estimates

| Level: | State | Family |
|---|---|---|
| Treatment Effect | -0.631 | -0.525 |
| | (0.419) | (0.062) |
| Robust 95% CI | [-1.537, 0.105] | [-0.681, -0.438] |
| P-value | 0.087 | 0 |
| Bandwidth | 0.083 | 0.036 |
| Effective Observations | 203 | 138692 |

*Note*: this table presents canonical RD estimates using both state-level average family income (weighted by the family-level probability weights) and family-level income as outcome variable, computed using the `rdrobust` command in R (Calonico et al., 2015*b*). MSE-optimal bandwidth was selected using the method in Calonico et al. (2020) and robust confidence intervals were calculated as in Calonico et al. (2014), clustered at the state level for the state-level data and the state-year level for the family-level data.

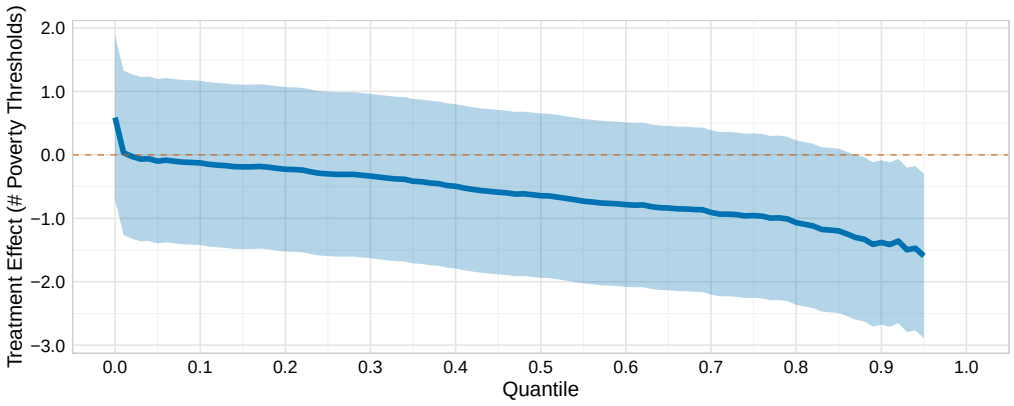
## *A-6.2 Figures*

Figure A-2: Percent of Top RDD Publications with R3D Setting, 2014–2024



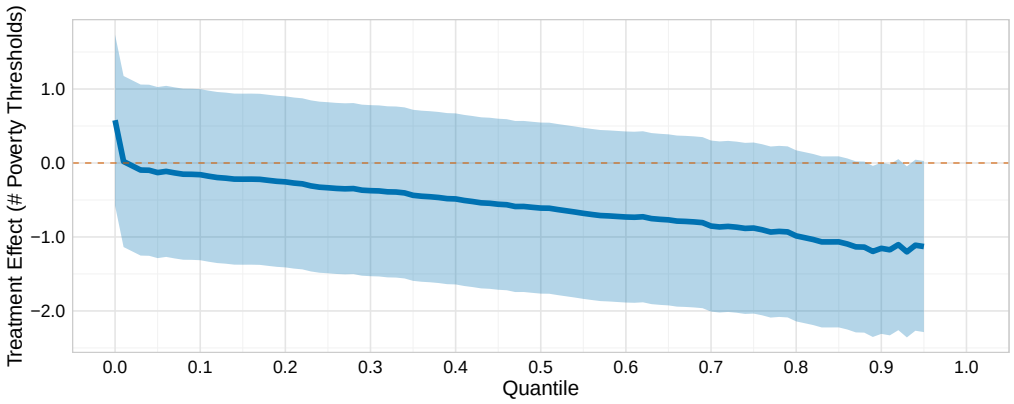

*Note*: this figure shows percentage of top 5 Economics and top 3 Political Science journals with RD designs that qualified as an R3D in the last 10 years, by 2-year periods. "Any R3D" indicates any form of R3D setting, including settings where the outcome of interest concerns sub-aggregate units but was aggregated to the same level as the treatment variable. Sample consists of any paper in those journals that had "regression discontinuity" or "RDD" in any of its fields.

Figure A-3: Distributional Effects of Democratic Governor Control, 1984-2010: Local Polynomial



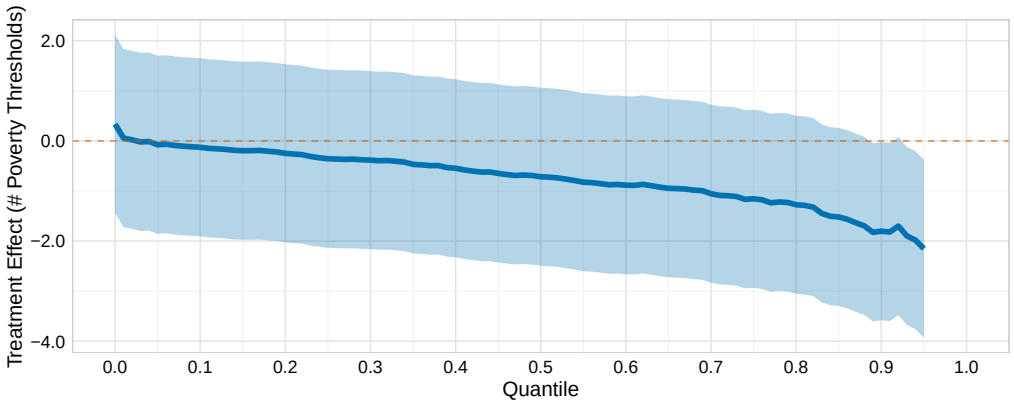

*Note:* this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average state-level income distributions, in the final year of the governor's tenure, 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 local polynomial estimator in Section 2.5.1 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. Treatment nullity p-value: 0.04, treatment homogeneity p-value: 0.067. Average MSE-optinal bandwidths: 0.262.

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



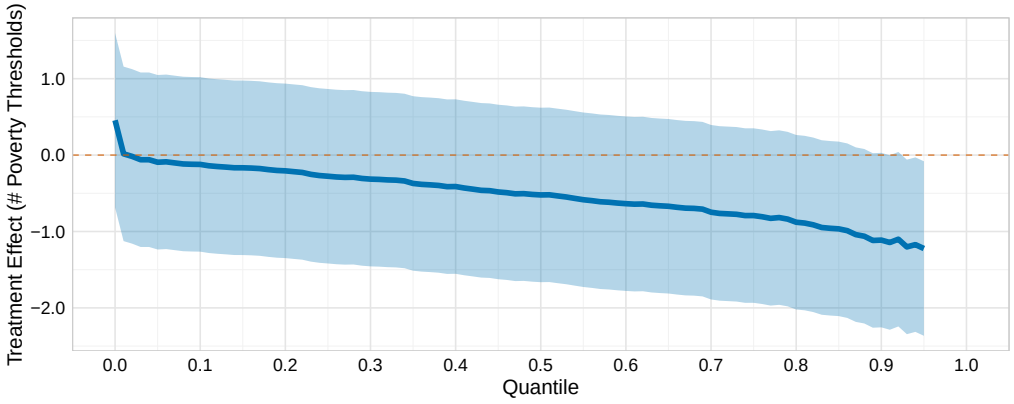

*Note:* this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average state-level income distributions, in the final year of the governor's tenure, 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 2.5 with first-order IMSE-optimal bandwidth and uniform kernel as in Section A-4.2, and uniform bands are constructed using Algorithm A-4.3 with 5,000 bootstrap repetitions. Treatment nullity p-value: 0.071, treatment homogeneity p-value: 0.140, IMSE-optimal bandwidth: 0.223.

Figure A-5: Distributional Effects of Democratic Governor Control, 1984-2010: 1/2 Bandwidth



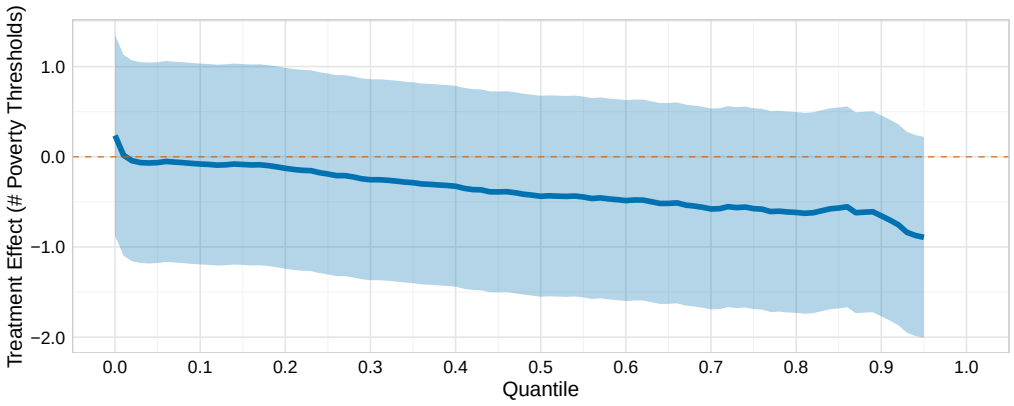

*Note:* local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average state-level income distributions, in the final year of the governor's tenure, 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 local polynomial estimator in Section 2.5.1 with 1/2×the first-order IMSE-optimal bandwidth (0.16) and triangular kernel as in Section A-4.2, and uniform bands are constructed using Algorithm A-4.3 with 5,000 bootstrap repetitions. Treatment nullity p-value: 0.045, treatment homogeneity p-value: 0.047, IMSE-optimal bandwidth: 0.11.

Figure A-6: Distributional Effects of Democratic Governor Control, 1984-2018



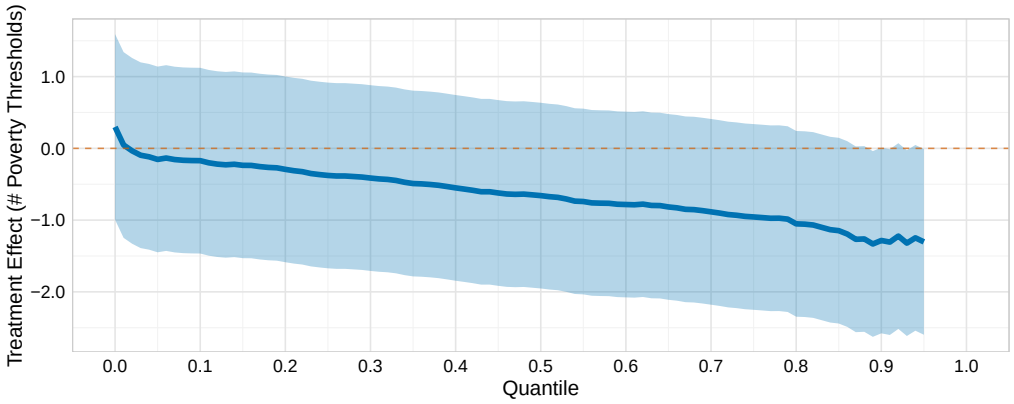

*Note:* this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average state-level income distributions, in the final year of the governor's tenure, near the 50% vote share threshold. Income is measured as real equivalized family income in multiples of the federal poverty threshold. Sample runs from 1984–2018, estimates are obtained using the second-order Fréchet estimator in Section 2.5 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. Treatment nullity p-value: 0.067, treatment homogeneity p-value: 0.087, IMSE-optimal bandwidth: 0.255.

Figure A-7: Distributional Effects of Democratic Governor Control, Robustness: Election-Year Incomes



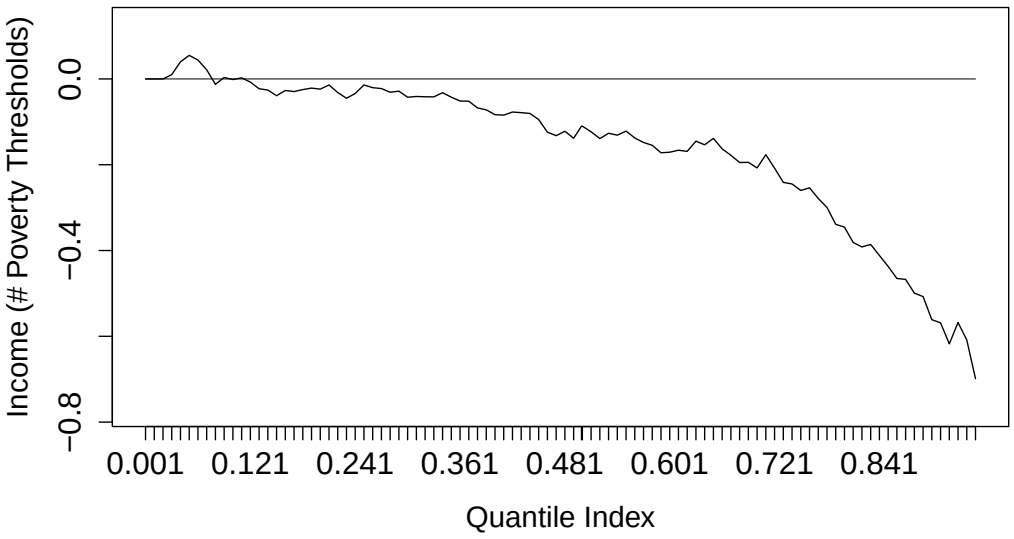

*Note:* this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average state-level income distributions, in the election year, 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 local polynomial estimator in Section 2.5.1 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. Treatment nullity p-value: 0.183, treatment homogeneity p-value: 0.224, IMSE-optimal bandwidth: 0.241.

Figure A-8: Distributional Effects of Democratic Governor Control, Robustness: No Cross-State Migration




*Note:* this figure shows local average quantile treatment effects estimates and uniform 90% 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, in the final year of the governor's tenure, near the 50% vote share threshold. Only families that did not migrate across state borders in the previous year are included. 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 local polynomial estimator in Section 2.5.1 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. Treatment nullity p-value: 0.072, treatment homogeneity p-value: 0.137, IMSE-optimal bandwidth: 0.229.

Figure A-9: Distributional Effects of Democratic Governor Control: Quantile RD Estimates




*Note*: this plot shows quantile RD estimates 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 state-level income distributions, in the final year of the governor's tenure, 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 quantile RD estimator of Qu and Yoon (2019) with bias correction (Qu et al., 2024), with the same bandwidth as Figure 6 and triangular kernel A-4.2.

## A-7 Software Appendix

All results in this paper were produced in `R` using `RStudio`. A complete reference list of packages used is provided below.

## References


Analytics, R. and Weston, S. (2022). *iterators: Provides Iterator Construct*. R package version 1.0.14.

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., Schwendinger, B., and Krylov, I. (2025). *data.table: Extension of 'data.frame'*. R package version 1.17.0.

Borchers, H. W. (2023). *pracma: Practical Numerical Math Functions*. R package version 2.4.4.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2023). *rdrobust: Robust Data-Driven Statistical Inference in Regression-Discontinuity Designs*. R package version 2.2.

Chen, Y., Zhou, Y., Chen, H., Gajardo, A., Fan, J., Zhong, Q., Dubey, P., Han, K., Bhattacharjee, S., Alexander, P., and Müller, H.-G. (2023). *frechet: Statistical Analysis for Random Objects and Non-Euclidean Data*. R package version 0.3.0.

Corporation, M. and Weston, S. (2022). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.17.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Kulichova, T. and Kratochvil, M. (2023). *scattermore: Scatterplots with More Points*. R package version 1.2.

Microsoft and Weston, S. (2022). *foreach: Provides Foreach Looping Construct*. R package version 1.5.2.

Müller, K. (2020). *here: A Simpler Way to Find Your Files*. R package version 1.0.1.

Qu, Z. and Yoon, J. (2024). *QTE.RD: Quantile Treatment Effects under the Regression Discontinuity Design*. R package version 1.1.0.

Rinker, T. W. and Kurkiewicz, D. (2018). *pacman: Package Management for R*. Buffalo, New York. version 0.5.0.

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*.

Ushey, K., Allaire, J., and Tang, Y. (2024). *reticulate: Interface to 'Python'*. R package version 1.37.0.

Van Dijcke, D. (2025). *R3D: Regression Discontinuity with Distributional Outcomes*. R

package version 0.1.0.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York, fourth edition. ISBN 0-387-95457-0.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*.

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4.

Wickham, H. and Henry, L. (2023). *purrr: Functional Programming Tools*. R package version 1.0.2.

Wickham, H., Vaughan, D., and Girlich, M. (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1.

Çetinkaya Rundel, M., Diez, D., and Dorazio, L. (2024). *usdata: Data on the States and Counties of the United States*. R package version 0.3.1.