

Regression Discontinuity Design with Distribution-Valued Outcomes

David Van Dijcke*

Department of Economics, University of Michigan

October 6, 2025

[Click here for the latest version](#)

Abstract

This article introduces a regression discontinuity design (RDD) for distribution-valued outcomes (R3D), extending the standard RDD framework to settings where outcomes are entire distributions rather than single values. This arises when treatment is assigned at the group level (e.g., firms, schools) but the objects of interest are within-group distributions (e.g., employee wages, student test scores). Standard RDDs are not designed for this two-level structure, as they assume scalar outcomes observed at the same level as treatment assignment. To address this, I develop a novel approach based on random distributions and show that, under a mild continuity condition on the average quantile function, the jump at the cutoff identifies a *local average quantile treatment effect*. To estimate it, I propose a distribution-valued local polynomial estimator, which fits the full quantile curve with a single bandwidth, avoids quantile crossing, and yields a meaningful “average distribution”. I derive uniform asymptotic normality, valid multiplier bootstrap confidence bands, and a data-driven bandwidth selection method. Simulations demonstrate strong performance and reveal that standard quantile RDD is biased and inconsistent in this setting. An application to U.S. gubernatorial close elections (1984–2010) uncovers an equality–efficiency trade-off under Democratic control, driven by income reductions 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, optimal transport

*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, Elena Manresa, Hans-Georg Müller, Alexander Petersen, Jesper Riis-Vestergaard Sørensen, Yuya Sasaki, Rasmus Søndergaard Pedersen, Konstantin Sonin, Benjamin Scuderi, Kaspar Wuthrich, Yidong Zhou, and participants at the University of Copenhagen Econometrics Workshop, the EEA 2025 Congress, 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 designs (RDD) are among the most widely used non-experimental strategies for causal inference. In their canonical form, they compare observations just above and just below a known cutoff in an assignment variable—test scores, income thresholds, electoral margins—and recover the jump in the conditional mean of a scalar outcome, interpreted as a local average treatment effect. This setup presumes that each unit has a single running variable and a single outcome measured 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 students’ test scores within each district; a firm’s revenue determines its eligibility for a subsidy, yet the object 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: Table 1 shows that 38% of all RDD papers published in the top five economics journals, and 32% of those in the top three political science journals feature some form of such a “group-level” RDD setting between 2014 and 2024 – either disaggregated (e.g. employee-level data) or aggregated (e.g. average employee outcomes within a firm).¹ Moreover, the design is becoming increasingly 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 the setting. These observations motivate the development of a novel 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 underlying *average* quantile function around the cutoff, where the average is with respect to the *distribution of distributions*. Identification follows from an intuitive condition: the conditional mean

¹The top five 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 three 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		

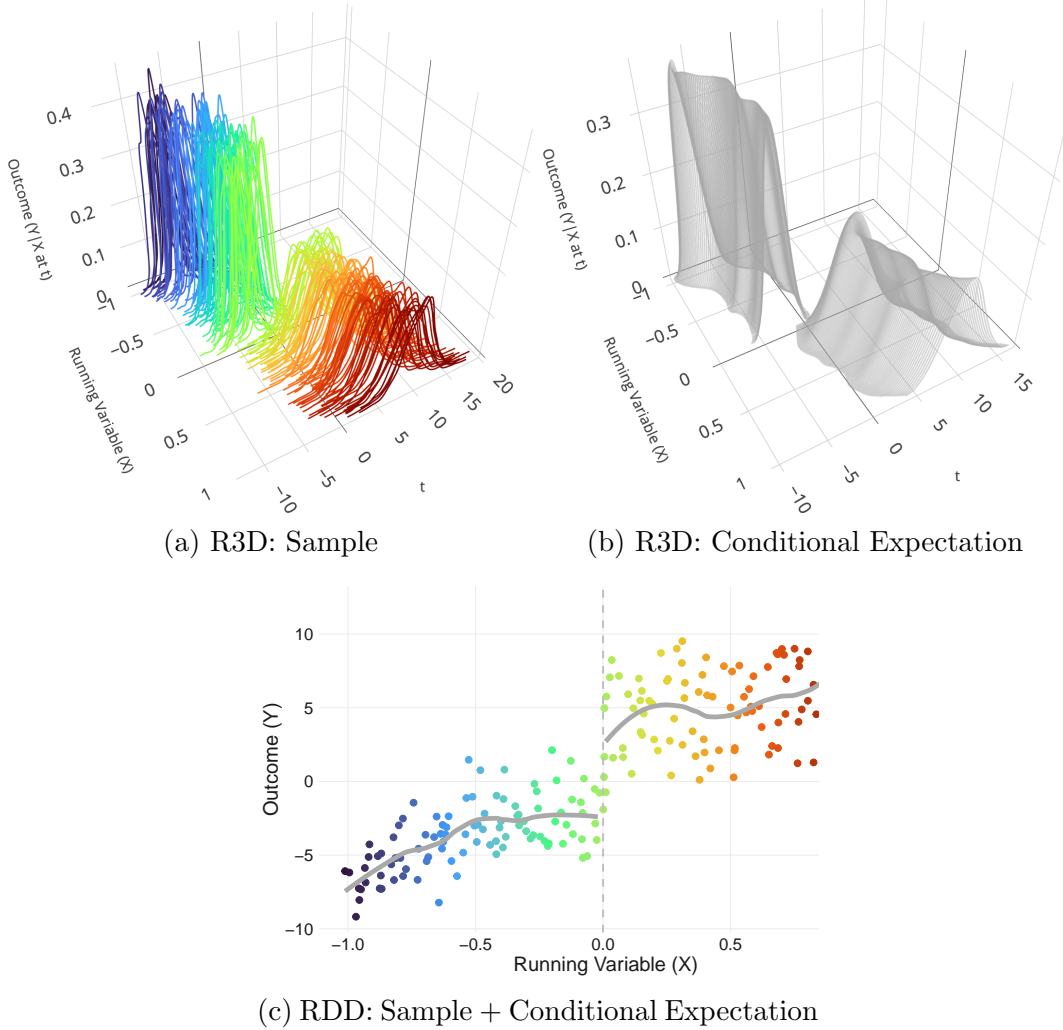
Note: this table shows percentage of papers in top five Economics and top three Political Science journals with RD designs that qualified as an R3D. “Any R3D” indicates any form of R3D setting, while “Aggregated” and “Disaggregated” indicate whether the data was aggregated to the level of the treatment 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.

of the quantile functions must be continuous in the running variable, but not the quantile functions themselves. This is the direct analogue, for distribution-valued 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 *distribution-valued 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 for each quantile. The distribution-valued estimator can be viewed as the L^2 projection of these pointwise fits onto the cone of valid quantile functions. Leveraging this link, I extend the results of Chiang et al. (2019) to derive uniform, debiased confidence bands for the pointwise estimator, and then extend these results to distribution space by exploiting the smoothness of the projection operator. This approach of casting the distribution-valued estimator as a

Figure 1: Example of a Distribution-Valued RDD



Note: (a) shows sample of distributions drawn as normal distributions with uniform means, $N(5x^3 + \delta - 5, 1)$ when $x \leq 0$ and $N(5x^3 + \delta - 2, 4)$ when $x > 0$ with $\delta \sim \text{Unif}(0, 5)$. (b) shows estimated conditional mean of these random distributions; (c) shows canonical RD design drawn as scalars instead of distributions, from the same data-generating process. Rainbow-colored objects indicate the sampled data in both designs, gray-colored objects indicate conditional means.

projection of existing frameworks allows me to sidestep the lack of central limit theorems for general metric spaces (like the space of distributions) 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 point-

wise local polynomial estimator. I also derive a novel integrated MSE-optimal bandwidth selection procedure that works for the whole distribution at once.

Further, I demonstrate the estimator 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/Republican 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 estimator to this setting, I estimate that Democratic governors prevail over a decrease in income inequality in their states, at the expense of a reduction in income for the top 10% of earners. These results point to a classical equality–efficiency trade-off (Okun, 1975), where a decrease in income inequality can only be achieved at the cost of an overall loss of income.

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, as it ignores randomness across aggregate units. Indeed, in what follows, the quantile RD estimator is shown to be biased and inconsistent in the R3D setting, both theoretically and in simulations. Section 2.4 below discusses this in more detail.

In summary, this article introduces a novel RD setting with distribution-valued outcomes. I define and identify a novel local average quantile treatment effect concept appropriate for this setting; develop a distribution-valued version of local polynomial regression, for which I derive uniform inference with a multiplier bootstrap and data-driven bandwidth selection; validate these results through simulations; and conclude with an empirical application to gubernatorial close elections.

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 RD designs with distribution-valued outcomes, which are not supported by existing quantile RD methods. 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, Titiunik, Vazquez-Bare and Keele (2016); Cattaneo et al. (2021); Bertanha (2020) considered aggregation schemes for RD with multiple cutoffs. Relatedly, Gunsilius and Van Dijke (2025); Papay et al. (2011); Cheng (2023) considered RD designs with multi-dimensional or multiple assignment variables. A related line of work develops clustered RDD methods for settings where the running variable is constant within groups (e.g., districts or states), inducing clustering in individual-level outcomes (Lee and Card, 2008; Lee and Lemieux, 2010; Barreca et al., 2016; Bartalotti and Brummet, 2017; Cattaneo, Titiunik and Vazquez-Bare, 2016; Kolesár and Rothe, 2018; Cattaneo and Titiunik, 2022). These methods derive cluster-robust asymptotics, bandwidths, and inference. While the approach of clustering standard errors at the group level works for average treatment effect estimation, this does not carry through to quantile treatment effects due to their nonlinearity, motivating the new approach in this article.

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; Groeneboom and Jongbloed, 2010; Lin et al., 2019) 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 Dijke 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 (Kurusu 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; Callaway and Li, 2019; Ghanem et al., 2023; Zhou et al., 2025) and synthetic control estimators (Kurusu et al., 2025a; Gunsilius et al., 2024).

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 $X \in \mathbb{R}$ denote the scalar-valued running variable. Let \mathcal{Y} be the space of cumulative distribution functions (cdfs) G on \mathbb{R} with finite variance, $\int_{\mathbb{R}} x^2 dG(x) < \infty$. Let $(X, Y) \sim F$ be a random element with joint distribution F on $\mathbb{R} \times \mathcal{Y}$. Here, Y is the distribution-valued outcome variable, and I emphasize that it is a random *distribution function* from \mathbb{R} to $[0, 1]$ so that we can write $Y(y)$ for some $y \in \mathbb{R}$. 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, unlike the canonical RD design. Then, denote by $T \in \{0, 1\}$ the treatment status. I assume that T is a 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 ??). Further, I will write $Q_Y(q)$ for the function mapping the cdf Y to quantiles,

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

Note that, in most of the paper, I will assume that Y_i and hence $Q_{Y_i}(q)$ are fully observed for a given unit (e.g., firm) i . This eases notation and reflects the natural setting where data is available on all within-group units; e.g. all employees within a firm (which I will use as the running example). The setting where only a sample of within-group units is observed is a straightforward extension that is covered in Section 3.3.1.

Finally, denote the marginal distributions of X and Y as F_X, F_Y , with $f_X := \frac{\partial F_X(x)}{\partial x}$ the pdf of X which will be well-defined near the cutoff c under the stated assumptions. I equip \mathcal{Y} with the Borel σ -algebra induced by the W_2 (2-Wasserstein) metric. Let $F_{Y|X=x}$ denote the conditional distribution of Y given $X = x$, which exists because (\mathcal{Y}, W_2) is Polish (Villani et al., 2008) and hence the disintegration theorem applies (Chang and Pollard, 1997). In addition, let $\mu = E[X]$ and $\Sigma = \text{var}(X)$ with Σ positive definite.

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 (running variable X) received grant writing assistance from the federal government (treatment T) 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 (outcomes Y) 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 fewer 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, where parent vote share is the running variable (X). 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 (T) on the entire distribution of student test scores within a school (Y). 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, many 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 (T) in New Jersey on wages, employment, and prices in fast food restaurants, comparing establishments on either side of the border with Pennsylvania. The running variable here is distance to the border (X). Since establishments typically sell many items and employ tens to hundreds of employees, one could, with the right data, observe entire distributions (Y) 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 (vote share, poverty level, distance to the border), one observes an entire distribution of the outcome (test scores, child mortality, store prices), and these outcome distributions vary across any two units (schools, counties, or restaurants). This implies that one needs to model the outcome as a *random distribution* instead of a random variable. 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 once more that Y is a cdf so one 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 $Z^t \in \mathbb{R}$, $t = 0, 1$ 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 individual-level causal effect at the threshold.

In the R3D setting, Y^t is a full distribution function. An intuitive generalization of the classical average treatment effect to this setting is given in the following definition.

Definition 1 (Local Average Quantile Treatment Effects (LAQTE)). *The local average*

quantile treatment effects for the R3D design are,

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

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

$$m_t(q) = E[Q_{Y^t}(q) \mid X = 0] = \int_y Q_y(q) dF_{Y^t|X=0}(0, y), \quad t = 0, 1,$$

where Proposition A-3 in Appendix proves the uniqueness and existence of these objects. These local average quantile treatment effects are a compelling way to summarize random distributional treatment effects. 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, the resulting “average quantile function” is meaningfully an “average of distributions” in that it respects 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. Note that one can also easily incorporate weights for each unit here (e.g., firm size weights), which would turn the population target into a weighted average quantile function and does not change the results that follow.

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

2.3.2 Identification

To identify τ^{R3D} from the data, I impose two assumptions that generalize the canonical RDD versions. 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. This is unlike the quantile

RD literature, which assumes the observed distributions themselves are smooth and hence precludes any heterogeneity in the distributions across aggregate units (firms) (Frandsen et al., 2012).

The following example may help to clarify this point. Suppose the counterfactual distribution functions Y^t are drawn from a class of normal distributions with uniformly 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 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 shows that even this simple collection of random Gaussian distributions satisfies the weaker continuity assumption I1 but still fails the continuity in quantiles. The following proposition formally establishes this for a large class of quantile models that subsumes classical quantile IV models (Chernozhukov and Hansen, 2005).

Proposition 1 (Smooth average vs. discontinuous observed quantile functions). *Fix $q \in (0, 1)$. Suppose*

$$(2) \quad Q_Y(q) = \mu(X, q) + \varepsilon(q) \quad a.s.,$$

where $\mu(\cdot, q)$ is continuous in x near 0, and $\varepsilon(q)$ is independent of X with a continuous density and $E[\varepsilon(q)]$ finite (w.l.o.g. take $E[\varepsilon(q)] = 0$, else absorb it into μ). Let $\{(X_i, Y_i)\}_{i=1}^n$ be i.i.d., and write $\varepsilon_i(q)$ for the copy attached to unit i .

(i) Smooth average quantiles. *The conditional average quantile*

$$E[Q_Y(q) \mid X = x] = \mu(x, q) + E[\varepsilon(q)]$$

is continuous in x .

(ii) Observed discontinuities for nearby units. *For any $\delta > 0, q \in [0, 1]$,*

$$P\left(\exists i \neq j : |X_i - X_j| < \delta \text{ and } Q_{Y_i}(q) \neq Q_{Y_j}(q) \mid X_1, \dots, X_n\right) = 0 \quad a.s.$$

Proof. (i) Follows from independence and continuity of $\mu(\cdot, q)$. (ii) For any pair (i, j) ,

$$Q_{Y_i}(q) - Q_{Y_j}(q) = [\mu(X_i, q) - \mu(X_j, q)] + (\varepsilon_i(q) - \varepsilon_j(q)).$$

Conditional on (X_i, X_j) , the difference $\varepsilon_i(q) - \varepsilon_j(q)$ has a continuous density, so it equals the fixed value $-\{\mu(X_i, q) - \mu(X_j, q)\}$ with probability 0. A finite union over pairs yields the claim. \square

Remark 1. The model in (2), though it is already quite general – non-parametric and with infinite-dimensional individual heterogeneity – is simplified to illustrate the smoothness assumptions at play. The assumptions I impose later in Section 3 allow for an even more general class of models that are nonseparable in ε , i.e. $Q_Y(q) = f(X, q, \varepsilon(q))$.

Remark 2. For a more concrete parametric example, consider the normal distribution with normal means in Figure 1. Let $W = g(X) + U$ with $U \sim N(0, 1)$ independent of X , and $Y = \mathcal{N}(W, 1)$. Then

$$Q_Y(q) = \underbrace{g(X) + \Phi^{-1}(q)}_{\mu(X, q)} + \underbrace{U}_{\varepsilon(q)},$$

so $m(x, q) = g(x) + \Phi^{-1}(q)$ is smooth in x , while (ii) implies a.s. distinct observed q -quantiles for any two nearby units.

The example solidifies the intuition behind Figure 1. While the *average* of the random distributions evolves smoothly in X , the actually observed 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. The model in (2) also clarifies where this randomness in the distributions arises from: the error term $\varepsilon(q)$ captures the across-group heterogeneity in the quantile at level q (firm-specific differences in the distributions).

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 $0 < \lim_{x \rightarrow c} f_X(x) < \infty$.

Then, I obtain the following identification result.

Lemma 1 (Identification). Under Assumptions I1 and I2, for a given $q \in [0, 1]$, the unobserved $\tau^{\text{R3D}}(q)$ is identified from the joint distribution of the observed (X, Y) as,

$$\begin{aligned} (3) \quad \tau^{\text{R3D}}(q) &= \lim_{x \rightarrow 0^+} E[Q_Y(q) \mid X = x] - \lim_{x \rightarrow 0^-} E[Q_Y(q) \mid X = x] \\ &=: \lim_{x \rightarrow 0^+} m(q) - \lim_{x \rightarrow 0^-} m(q) \\ &=: m_+(q) - m_-(q). \end{aligned}$$

Note that the lemma implicitly defines $m_{\pm}(q)$ and $m(q)$. Thus, the R3D estimand is the jump in the conditional average quantile at $x = 0$.

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 $E[Q_Y(q) | X = c]$ at the threshold $X = 0$. Thus, I can define a discontinuity in our setting to occur when, for some $q \in [0, 1]$

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

The uniform confidence bands I derive below allow one to test for the presence of such discontinuities uniformly over all q while accounting for the randomness in the quantiles (across groups). 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 Existing Approaches

2.4.1 Quantile RD

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. In practice, Q-RDD estimates a jump in the *observed* quantile function, while R3D estimates a jump in the underlying *average* quantile function. The former occurs almost surely in the R3D setting, the latter occurs only when the treatment has an effect.

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,

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

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,

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

As can be seen, the R3D approach *first* estimates quantiles and only then runs a local

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 across units. In the presence of such randomness, the observed distributions will almost surely not vary 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 look the same *on average*, 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: there would be no random variation left in the averages. Third, and similarly, the standard assumption that treatment effects are heterogeneous 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 workforce at Company A differently than at 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.

2.4.2 Group-Level Average Treatment Effects

Though the focus in this article is on quantile treatment effects, a comparison to average treatment effect (ATE) estimation helps to further illuminate the setting. The literature has taken two approaches for estimating average treatment effects in the grouped data setting considered in this paper: 1) aggregate to the group (firm) level and fit a local polynomial regression on the group averages; 2) fit a local polynomial regression directly on the disaggregated (employee) data, while clustering the standard errors at the group (firm) level (Bartalotti and Brummet, 2017). The former approach estimates a weighted average treatment effect, weighted by group size, because it combines the probability that an observation falls within a certain group (whether an employee works at a given firm) with the outcome the observation has within that group (the wage the employee receives at a given firm). Treatment affects both dimensions at the disaggregated level (employees may see their wage change at a given firm but may also change firms), so estimating at that level targets a group size-weighted average. Clustering helps adjust standard errors, but does not affect the estimand. The strategy of first aggregating to the group level and then estimating treatment effects, on the other hand, will estimate the unweighted average treatment effect and inherently accounts for clustering (Bartalotti and Brummet, 2017). In that sense, it is analogous to the idea behind R3D, except that in the R3D design one “aggregates” to the group level by working directly with the groups’ distributions as the target objects. Moreover, the implied average treatment effect is identical, since, under Fubini-Tonelli,

$$E[Q_Y(q)] = \int_{[0,1]} \int_Y Q_y(q) dF_Y(y) dq = \int_Y \int_{[0,1]} Q_y(q) dq dF_Y(y) = \int_Y E[y] dF_Y(y)$$

so that first estimating R3D and then averaging estimates the same ATE as first averaging and then estimating a standard local polynomial regression. Unlike in the ATE setting, however, directly estimating a quantile RD on the disaggregated data does not yield a meaningful mixture of group-level distributions (i.e., an average of group-level quantile treatment effects), but rather the quantile of a pooled mixture, which has no coherent causal interpretation at the group level. The reason why the equivalence does not carry through there is that quantiles are not linear, and hence the resulting “weighted quantiles” do not correspond to the quantiles of any coherent weighted-average distribution, unlike in the linear ATE case – except in the edge case where there is no heterogeneity in the distribution-valued outcomes, as discussed in the previous section.

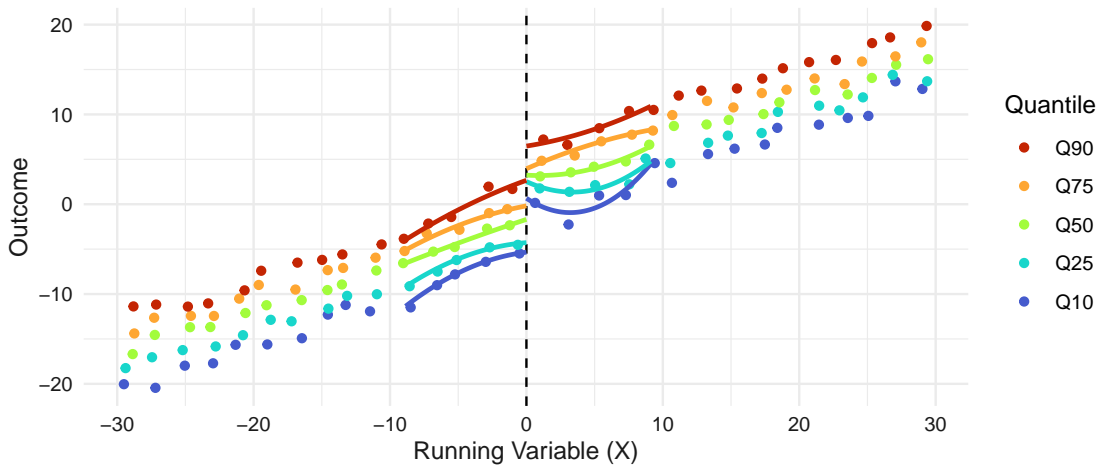
2.5 Estimators

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 builds on this by projecting the local polynomial estimator back onto the space of quantile functions. As shown in Proposition A-3, 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 an “average” distribution in finite samples.

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.

Figure 2: Local Polynomial Estimator: Illustration



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,

$$(4) \quad \hat{m}_{\pm,p}(q) = e_0^\top \hat{\alpha}_{\pm,p} = \left(\text{polynomial fit at } x = 0^\pm \right) \Big|_{\text{order}=p}$$

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

where e_0 is the first standard basis vector, $K(x)$ a kernel function, $\delta_i^\pm := 1\{X_i \gtrless c\}$, and $r_p(x) := (1, x, x^2, \dots, x^p)$. 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,

$$(5) \quad \hat{m}_{\pm,p}(q) = \sum_{i=1}^n s_{\pm,in}^{(p)}(h) Q_{Y_i}(q)$$

where $s_{\pm,in}^{(p)}(h)$ 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 $s_{\pm,in}^{(p)}(h)$ can be negative, $\hat{m}_{\pm,p}$ need not be a quantile function. To resolve this, I use the monotone rearrangement from Chernozhukov et al. (2010), though the Fréchet estimator proposed below is preferred and circumvents the need for rearrangement altogether.

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

$$(6) \quad \hat{\tau}_p^{\text{R3D}}(q) := \hat{m}_{+,p}(q) - \hat{m}_{-,p}(q).$$

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 (see Section 2.7 below).

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 $s_{\pm,in}^{(p)}(h)$ can be negative and thus introduce non-monotonicity

(quantile crossing). Second, the pointwise estimation approach ignores global function information, which degrades the estimator’s finite-sample performance, 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 estimator,

$$(7) \quad \hat{m}_{\pm, \oplus, p} := \Pi_{\mathcal{Q}}(\hat{m}_{\pm, p}) := \operatorname{argmin}_{m_{\pm} \in Q(\mathcal{Y})} \int_a^b (\hat{m}_{\pm, p}(q) - m_{\pm}(q))^2 dq,$$

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-3, I show that $\hat{m}_{\pm, \oplus, p}$ is unique and exists under the stated assumptions. The estimated treatment effects are then defined as

$$(8) \quad \hat{\tau}_{\oplus, p}^{\text{R3D}}(q) := \hat{m}_{+, \oplus, p}(q) - \hat{m}_{-, \oplus, p}(q).$$

Intuitively, the projection step ensures that the estimated functions are valid quantile functions, while also borrowing information across quantiles and avoiding repeated bandwidth selection. As a result, the Fréchet regression estimator typically outperforms the pointwise local polynomial estimator in finite samples, while still converging to the same population object. Indeed, the projection $\Pi_{\mathcal{Q}}$ is characterized by a no-regret property: the resulting projected estimator will always lie closer to the true quantile function than the (unprojected) local polynomial estimator; see Lemma 2 below for a formal statement.

Readers mainly interested in applied use can take away that the Fréchet estimator is the preferred choice in practice, as it is a properly *distributional* estimator that regresses directly on the distributions as a whole. For readers interested in the deeper theoretical connection, the following paragraph provides some background on its equivalence with local Fréchet regression in Wasserstein space.

Theoretical Motivation. The projection estimator in (7) has a rich theoretical foundation connecting optimal transport, isotonic regression, and local polynomial methods. At its core, this estimator is equivalent to the local Fréchet regression estimator of Petersen and Müller (2019) restricted to the space of probability distributions \mathcal{Y} equipped with the 2-Wasserstein distance d_{W_2} . This equivalence, established in Proposition A-3, arises from a fundamental isometry: the L^2 distance between quantile functions equals the 2-Wasserstein

²Working 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.

distance between their corresponding distributions.

In particular, $\hat{m}_{\pm, \oplus, p}$ solves a natural optimization problem in distribution space. It finds the unique distribution whose quantile function minimizes the weighted sum of squared 2-Wasserstein distances to the observed distributions $\{Y_i\}_{i=1}^n$, with local polynomial weights $s_{\pm, in}^{(p)}(h)$. In this sense, the estimator computes the weighted “central tendency” of distributions in probability space—the direct analogue of how the Euclidean mean minimizes squared distances in \mathbb{R}^n . This Wasserstein barycenter interpretation (Agueh and Carlier, 2011; Fan and Müller, 2024) provides a principled notion of an “average distribution” at the cutoff.

The approach can be viewed as a “double regression”: local polynomial regression on pointwise quantiles, followed by a functional regression on the full quantile functions. This projection, a form of isotonic regression (Robertson et al., 1988), ensures monotonicity while preserving asymptotic properties. The key insight from Theorem 4 below is that because the population quantity $m_{\pm}(q)$ is already a valid quantile function, the projection acts as an identity operator asymptotically. In finite samples, however, it does enforce the shape constraint while also enjoying a “no-regret” property (Lemma 2): the projected estimate is always closer to the truth than the unprojected one in any L^p norm.

Lemma 2 (Improvement in Estimation Property). *Suppose that \hat{Q} is an estimator of some true quantile curve Q_0 . Then the projected curve $\hat{Q}^* = \Pi_Q(\hat{Q})$ is closer to the true curve in the sense that, for each $x \in \mathbb{R}$,*

$$\left\| \hat{Q}^* - Q_0 \right\|_{L^p} \leq \left\| \hat{Q} - Q_0 \right\|_{L^p} \quad \text{for all } p \in [1, \infty].$$

Proof. See Corollary 2.4 in Lin et al. (2019), which explicitly proves the functional version of the classical result in Robertson et al. (1988, Theorem 1.6.1), see also Groeneboom and Jongbloed (2014, Theorem 2.1). \square

These theoretical properties establish clear advantages over alternative approaches. Like monotone rearrangement (Chernozhukov et al., 2010), the Fréchet estimator preserves asymptotic limits and satisfies the no-regret property in all L^p norms. However, where rearrangement mechanically sorts values to enforce monotonicity, the Fréchet approach finds the *closest* valid quantile function in the 2-Wasserstein metric—the natural distance for probability distributions, leading to the desirable “average distribution” interpretation. This optimization-based approach also has practical benefits: by treating quantile functions as integrated objects rather than collections of isolated points, it exploits cross-quantile information to improve finite-sample performance. Combined with its computational advantage of requiring only a single bandwidth selection, the Fréchet estimator offers both theoretical

rigor and practical efficiency, making it the preferred choice for distribution-valued RDD estimation.

2.5.3 Fuzzy R3D Estimator

So far, I have focused on the sharp RD estimators for ease of reading. They can be easily extended to the fuzzy RD setting where treatment is assigned probabilistically (Hahn et al., 2001), as follows. Define the local polynomial fuzzy R3D estimator as,

$$(9) \quad \hat{\tau}_p^{\text{F3D}}(q) := \frac{\hat{\tau}_p^{\text{R3D}}(q)}{\hat{m}_{+,T,p} - \hat{m}_{-,T,p}}$$

where

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

The corresponding Fréchet estimator is,

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

Identification. To define the corresponding population fuzzy estimator and establish its identification, define T_i^0, T_i^1 , the local potential treatment states as $\lim_{x \rightarrow 0^\pm} T_i(x)$, 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,*

$$(11) \quad \tau^{\text{F3D}}(q) := E[Q_{Y^1}(q) - Q_{Y^0}(q) \mid X = 0, C] \quad q \in [0, 1].$$

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

I3 (Fuzzy RD). $\lim_{x \rightarrow 0^+} P(T \mid X = x) > \lim_{x \rightarrow 0^-} P(T \mid X = x).$

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

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

This gives,

Lemma 3 (Fuzzy Identification). *Under Assumptions I1– I5, the unobserved τ^{F3D} is identified from the joint distribution of the observed (X, Y, T) as,*

$$(12) \quad \begin{aligned} \tau^{F3D}(q) &= \frac{\lim_{x \rightarrow 0^+} E[Q_Y(q) | X = x] - \lim_{x \rightarrow 0^-} E[Q_Y(q) | X = x]}{\lim_{x \rightarrow 0^+} E[T | X = x] - \lim_{x \rightarrow 0^-} E[T | X = x]} \\ &:= \frac{m_+(q) - m_-(q)}{m_{+,T} - m_{-,T}}. \end{aligned}$$

This Wald estimator takes the same form as the standard fuzzy RDD estimator (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.

2.6 Overview of Inference

Algorithm 1 Multiplier Bootstrap for Uniform Confidence Bands

Require: Sample $\{(X_i, Y_i, T_i)\}_{i=1}^n$, quantile grid $\mathcal{T}^* = \{q_1, \dots, q_M\} \subset [a, b]$, polynomial order p , kernel K , bandwidth h , bootstrap repetitions B , significance level $\lambda \in (0, 1)$.

- 1: **for** each $q_j \in \mathcal{T}^*$ **do**
 - 2: Compute conditional mean estimates $\hat{m}_{\pm,p}(q_j)$ (average quantiles above/below cutoff) using local polynomials.
 - 3: (Optional: Fréchet) Project $\hat{m}_{\pm,p}$ onto the space of valid quantile functions using isotonic regression to get $\hat{m}_{\pm,(\oplus),p}$.
 - 4: Estimate residuals $\hat{\mathcal{E}}(Y_i, X_i, q) = Q_{Y_i}(q) - \tilde{E}[Q_{Y_i}(q) | X_i]$, where $\tilde{E}[\cdot | X_i]$ reuses the t -th order local polynomials estimated in Step 1-2, with $t \leq p$.
 - 5: **end for**
 - 6: **for** $b = 1$ to B **do**
 - 7: Draw i.i.d. $N(0, 1)$ multipliers.
 - 8: **for** each $q_j \in \mathcal{T}^*$, $k \in \{1, 2\}$ **do**
 - 9: Compute bootstrap processes as multiplier-weighted average of residuals (plus some additional weighting terms).
 - 10: **end for**
 - 11: Form estimate of limiting process $\hat{\mathbb{G}}^{\text{R3D},b}(q_j) = \hat{m}_{+,(\oplus),p}(q_j) - \hat{m}_{-,(\oplus),p}(q_j)$.
 - 12: **end for**
 - 13: Compute critical value $\hat{c}_n^B(a, b; \lambda)$ as $(1 - \lambda)$ -quantile of $\max_{q \in \mathcal{T}^*} |\hat{\mathbb{G}}^{\text{R3D},b}(q)|$ over b .
 - 14: Construct uniform bands: $[\hat{\mathbb{G}}^{\text{R3D},b}(q) \pm (\sqrt{nh})^{-1} \hat{c}_n^B(a, b; \lambda)]$ for $q \in \mathcal{T}^*$
-

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.

A sketch of how these uniform confidence bands are computed in the sharp RD case is provided in Algorithm 1. The fuzzy RD case proceeds analogously. Note that some of the notation is slightly simplified compared to what follows below, for ease of reading. The algorithm also assumes the bandwidth h is given, see Section 2.7 below for more detail on how it is chosen optimally.

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 $[\underline{q}, \bar{q}] \subset (0, 1)$ through the following tests (Chiang and Sasaki, 2019):

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

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

2.7 Bandwidth Selection

In practice, the estimators require choosing a bandwidth h . The standard approach for doing so in the local polynomial regression and RDD literatures is to pick a h that minimizes the mean-squared error (MSE), and hence balances bias and variance. Here, I give an overview of the proposed approach for our random distribution setting, with full details in Appendix A-4.2.

For the local polynomial estimator in Section 2.5.1, I follow these canonical approaches. In particular, the goal is to pick a separate bandwidth $h(q)$ for every quantile $q \in [a, b]$. Since the estimator is simply a local polynomial regression for each quantile, we can rely on the canonical result of Fan and Gijbels (1992), who derived a closed-form expression for the asymptotic MSE. Minimizing that expression with respect to h gives the optimal

bandwidths,

$$(13) \quad h^*(q) = \left(\frac{1}{2(s+1)} \frac{\text{Var}(\hat{\tau}_s^{\text{R3D}}(q))}{\text{Bias}(\hat{\tau}_s^{\text{R3D}}(q))^2} \right)^{1/(2s+3)} n^{-1/(2s+3)},$$

with the full bias and variance terms given in Appendix A-4.2.1. Here, s is the desired order of the local polynomial regression (typically $s = 1$ for local linear), which is used for bandwidth selection, while estimation is done with a p -th order polynomial where $p > s$. As shown in Calonico et al. (2014), this approach is equivalent to estimating an s -th order local polynomial with explicit bias correction.

For the Fréchet estimator in Section 2.5.2, no such closed-form expression for the asymptotic MSE exists, because the Taylor expansions it relies on is not well-defined in the space of distributions. This would seem to preclude data-driven bandwidth selection for the Fréchet estimator and, indeed, existing approaches have relied on cross-validation instead (Petersen and Müller, 2019). However, in the statistical results below, I show that both estimators converge to the same asymptotic limit. Leveraging this fact, one can show that, for a given h , $\|\hat{\tau}_{\oplus,p}^{\text{R3D}} - \hat{\tau}_p^{\text{R3D}}\|_{l^\infty([a,b])} = o_p(1)$ (see Lemma A-9), that is, the difference between the two estimators vanishes asymptotically. As a result, their mean squared errors are asymptotically equivalent. Thus, to obtain a single bandwidth h for the full distribution of treatment effects $\hat{\tau}_{\oplus,p}^{\text{R3D}}$, one can compute the integrated mean-squared error (IMSE) as,

$$\text{IMSE}[\hat{\tau}_{\oplus,s}] = \int_a^b \text{MSE}(\hat{\tau}_{\oplus,s}^{\text{R3D}}(q)) \, dq = \int_a^b \text{MSE}(\hat{\tau}_s^{\text{R3D}}(q)) \, dq + o_p(1).$$

Minimizing this with respect to h gives the IMSE-optimal distributional bandwidth,

$$h_{\oplus}^* = \left(\frac{1}{2(s+1)} \frac{\int_a^b \text{Var}(\hat{\tau}_{\oplus,s}^{\text{R3D}}(q)) \, dq}{\int_a^b \text{Bias}(\hat{\tau}_{\oplus,s}^{\text{R3D}}(q))^2 \, dq} \right)^{1/(2s+3)} n^{-1/(2s+3)}.$$

The full algorithm for estimating it is given in Appendix A-4.2.3.

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 the 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 by showing that the projection mapping is Hadamard differentiable in the limit. 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, with Ω^x the sample space. 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 $c_1(q) : [a, b] \rightarrow C \subset \mathbb{R}$ a bounded Lipschitz function and $c_2(q) = \bar{c}_2 > 0$. The baseline bandwidth $h = h_n$ satisfies $h \rightarrow 0$, $nh^2 \rightarrow \infty$, $nh^{2p+3} \rightarrow 0$.*

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 $i = 1, \dots, n$.

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

(i) The maps $(x, q) \rightarrow E[Q_Y(q)|X = x]$ and $x \rightarrow E[T|X = x]$ 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 all $(q_1, q_2) \in [a, b]^2$, the map $x \mapsto E[Q_Y(q_1)Q_Y(q_2) | X = x]$ lies in $C^1(\mathcal{N})$, with the partial derivatives with respect to x bounded uniformly in (q_1, q_2) .

L3 (Random Quantile Spread). *The maps $Y \rightarrow \sup_{q \in [a, b]} |Q_Y(q)|$ and $x \rightarrow \sup_{q \in [a, b]} |E[Q_Y(q) | X = x]|$ are in $L^{2+\epsilon}(P^x)$ on $[\underline{c}, \bar{c}] \times \mathcal{Y}_c$, $\epsilon > 0$.*

L4 (Average Strict Monotonicity). *For each $x \in \mathcal{N} := (-\epsilon, \epsilon) \setminus \{0\}$ for $\epsilon > 0$ and $q \in [a, b] \subset (0, 1)$, there exist constants $0 < \kappa \leq K < \infty$ such that $\kappa \leq \frac{\partial}{\partial q} E[Q_Y(q) | X = x] \leq K$ for a.e. $q \in [a, b]$.*

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 $h = n^{-1/5}$ (Calónico 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 pointwise 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 over the class of random distributions. Assumption **L4** ensures the limiting quantile function does not have flat parts with some buffer, which in turn guarantees that the projection operator Π_Q is fully Hadamard differentiability in the limit (see the proof of Theorem 4). Note that it is only imposed on the *average* quantile function and hence still allows the underlying quantile functions for each unit i to have flat parts, as long as at every quantile q there is *some* positive mass of units that have a strictly increasing quantile function. It also allows for atoms in the underlying quantile functions, as long as they almost surely occur at different points of the support for different units (firms). In that sense, the assumption is much weaker than what is imposed in quantile RD, where every single unit’s quantile function is assumed to be strictly increasing and continuous (Frandsen et al., 2012, Assumption Q). 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 Quantile Functions). *For each unit $i \leq n$ let \widehat{Q}_{Y_i} be any estimator of the quantile function Q_{Y_i} on $[a, b]$. There exist deterministic rates $r_{n_i} \downarrow 0$ such that*

$$\sup_{q \in [a,b]} |\widehat{Q}_{Y_i}(q) - Q_{Y_i}(q)| = O_p(r_{n_i}) \quad \text{for all } i \leq n,$$

and the following holds,

$$\sqrt{nh} \max_{1 \leq i \leq n} r_{n_i} \longrightarrow 0.$$

Q1 is a high-level condition that accommodates any empirical quantile estimator that converges “fast enough”. In particular, it balances the speed of the empirical quantile estimator \hat{Q}_{Y_i} against the ratio of the within-group sample size n_i to the across-group sample size n , requiring that together, they imply a within-group convergence rate that is faster than \sqrt{nh} , the across-group rate. This is a light assumption in the sense that, as long as the within-group rate is of a similar order as \sqrt{nh} , it accommodates most standard empirical quantile estimators (e.g. [van der Vaart \(2000, Corollary 21.5\)](#)), including ones that allow for within-group dependence.

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. \(2019\)](#). Assume without loss of generality that the kernel K is supported on $[-1, 1]$. Define $g_1 : (Y, T, q) \subset (\mathcal{Y}, \{0, 1\}, [a, b]) \rightarrow \mathcal{Q}_Y(q)$, $g_2 : (Y, T, q) \subset (\mathcal{Y}, \{0, 1\}, [a, b]) \rightarrow T$. Further, define the population residual $\mathcal{E}_k(y, t, x, q) := g_k(y, t, q) - E[g_k(y, t, q) | X_i = x]$, $k = 1, 2$ and let

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

with $k, l \in \{1, 2\}$, $q, q' \in [a, b]$, and $\sigma_{kl}(q, q' | 0^\pm) = \lim_{x \rightarrow 0^\pm} \sigma_{kl}(q, q' | x)$. Moreover, let e_0 denote the 0th standard basis vector of \mathbb{R}^p , $(1, 0, \dots, 0)$, and write $\Gamma_{\pm, p} := \int_{\mathbb{R}_\pm} K(u) r_p(u) r_p'(u) du$ where I remind the reader that $r_p(u) := (1, u, \dots, u^p)$. Also, let $X_n \rightsquigarrow X$ denote weak convergence for some sequence of random variables X_n and a random variable X , while $X_n \xrightarrow[\xi]{p} X$ denotes conditional weak convergence. The latter is defined as $\sup_{h \in BL_1} |E_{\xi|x}[h(X_n) - E[h(X)]]| \xrightarrow[x]{p} 0$ where BL_1 the set of bounded Lipschitz functions with supremum norm bounded by 1 and $\xrightarrow[x]{p}$ 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 $\mathbb{G}_{H\pm} : \Omega^x \rightarrow l^\infty([a, b] \times \{1, 2\})$ is a zero-mean Gaussian process with covariance function,

$$H_{\pm,p}((q, k), (q', l)) = \frac{\sigma_{kl}(q, q'|0^\pm) e'_0(\Gamma_{\pm,p})^{-1} \Psi_{\pm,p}((q, k), (q', l)) (\Gamma_{\pm,p}^{-1}) e_0}{\sqrt{c_k(q) c_l(q')} f_X(0)},$$

where,

$$\Psi_{\pm,p}((q, k), (q', l)) := \int_{\mathbb{R}} r_p(u/c_k(q)) r'_p(u/c_l(q')) K(u/c_k(q)) K(u/c_l(q')) du$$

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,*

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

and under the additional Assumptions *I3–I5*

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

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

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

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,

$$(14) \quad \hat{\nu}_{\xi,n}^\pm(q, k) = \sum_{i=1}^n \xi_i \frac{e'_0(\Gamma_{\pm,p})^{-1} \hat{\mathcal{E}}_k(Y_i, T_i, X_i, q) r_p\left(\frac{X_i}{h_k(q)}\right) K\left(\frac{X_i}{h_k(q)}\right) \delta_i^\pm}{\sqrt{nh_k(q)} \hat{f}_X(0)},$$

where $\hat{f}_X(0)$ is any uniformly consistent estimator of $f_X(0)$, and $\hat{\mathcal{E}}_k(Y_i, T_i, X_i, q)$ is any uni-

formly 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 $\hat{\nu}_{\xi,n}^\pm(q, k)$ 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 $\hat{\nu}_{\xi,n}^\pm \xrightarrow[p]{\xi} \mathbb{G}_{H^\pm}$ and thus,*

$$\hat{\mathbb{G}}^{\text{R3D}}(\cdot) := c_1(\cdot)^{-1/2} \hat{\nu}_{\Delta,n}(\cdot, 1) \xrightarrow[p]{\xi} \mathbb{G}^{\text{R3D}}$$

and under the additional Assumptions [I4](#), [I5](#),

$$\begin{aligned} \hat{\mathbb{G}}^{\text{F3D}}(\hat{m}_{+,p}, \hat{m}_{-,p}, \hat{m}_{+,T,p}, \hat{m}_{-,T,p})(\cdot) := \\ \frac{c_1(\cdot)^{-1/2} (\hat{m}_{+,T,p}(\cdot) - \hat{m}_{-,T,p}(\cdot)) \hat{\nu}_{\Delta,n}(\cdot, 1) - c_2(\cdot)^{-1/2} (\hat{m}_{+,p}(\cdot) - \hat{m}_{-,p}(\cdot)) \hat{\nu}_{\Delta,n}(\cdot, 2)}{(\hat{m}_{+,T,p}(\cdot) - \hat{m}_{-,T,p}(\cdot))^2} \\ \xrightarrow[p]{\xi} \mathbb{G}^{\text{F3D}} \end{aligned}$$

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

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

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](#) and [Assumption L4](#),*

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

where \mathbb{G}_{H^\pm} is the same zero-mean Gaussian process as in [Theorem 1](#).

Proof sketch. The idea behind the proof is simple. Recall that $\hat{m}_{\pm,\oplus,p} = \Pi_{\mathcal{Q}}(\hat{m}_{\pm,p})$, the L^2 projection onto the space of quantile functions. It is well-known that $\Pi_{\mathcal{Q}}$ is Hadamard

directionally differentiable (Zarantonello, 1971). Moreover, its Hadamard derivative at the limiting point m_{\pm} is equal to the identity operator, i.e. $\Pi_{\mathcal{Q}}[m_{\pm}](h) = h$ for any $h \in C([a, b])$. This follows because the limit m_{\pm} is a proper quantile function, which under Assumption L4 is also strictly increasing. That guarantees that small, smooth perturbations of m_{\pm} remain valid quantile functions, and hence the projection remains “inactive” no matter what direction we perturb m_{\pm} in. As a result, we can apply the functional delta method (van der Vaart, 2000, Theorem 20.8) to show that the Fréchet estimator has the same asymptotic limit as the unprojected estimator. The full proof in Appendix A-5 works out the details to make sure we work in the correct function spaces to obtain uniform convergence based on this argument. In that sense, the proof extends the uniform convergence arguments for “discrete” isotonic regression in Yang and Barber (2019) to the “continuous” version in Groeneboom and Jongbloed (2010). \square

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 and Assumption L4 it follows that,*

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

and under the additional Assumptions I3–I5,

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

Corollary 2 (Bootstrap: Fréchet). *Under the assumptions of Theorem 3, the estimated bootstrap processes $\hat{\mathbb{G}}^{\text{R3D}}$ and $\hat{\mathbb{G}}^{\text{F3D}}(\hat{m}_{+,\oplus,p}, \hat{m}_{-,\oplus,p}, \hat{m}_{+,T,p}, \hat{m}_{-,T,p})$ deliver asymptotically valid confidence intervals for the Fréchet estimators $\hat{\tau}_{\oplus,p}^{\text{R3D}}, \hat{\tau}_{\oplus,p}^{\text{F3D}}$.*

3.3 Extensions

3.3.1 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 matched employer-employee data that contains all employees within a firm. In practice, however, there is often another sampling layer, where one only observes further i.i.d. samples $Z_{ij}, j = 1, \dots, n_i$ from these

distribution functions, with $Z_{ij} \in \mathbb{R}$ distributed according to Y_i .³ This corresponds to the case where one only observes a subsample of all employees within any given firm. Here, 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,

$$(15) \quad \bar{\mu}_{\pm,p}(q) := \frac{1}{n} \sum_{i=1}^n s_{\pm,in}^{(p)}(h) \hat{Q}_{Y_i}(q),$$

where

$$(16) \quad \hat{Q}_{Y_i}(q) := \inf \left\{ z : \hat{Y}_i(z) \geq q \right\}$$

with

$$\hat{Y}_i(z) := \frac{1}{n_i} \sum_{j=1}^{n_i} 1(Z_{ij} \leq z),$$

the empirical distribution function of unit i . The other estimators are similarly modified by plugging in \hat{Q}_{Y_i} , and denoted with a bar instead of a hat, e.g. $\bar{\tau}_p^{\text{R3D}}$. Sampling weights can be incorporated by constructing \hat{Q}_{Y_i} as weighted quantile functions.

Then, I obtain the following results,

Proposition 2 (Empirical Quantiles). *Under the same respective Assumptions of Theorems 1 and 2, as well as Assumption Q1, the estimators with empirical quantile functions, $\bar{m}_{\pm,p}$, $\bar{m}_{\pm,\oplus,p}$, $\bar{\tau}_p^{\text{R3D}}$, $\bar{\tau}_p^{\text{F3D}}$, $\bar{\tau}_{\oplus,p}^{\text{R3D}}$, $\bar{\tau}_{\oplus,p}^{\text{F3D}}$ 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 Assumption Q1, the estimated bootstrap processes $\hat{\mathbb{G}}^{\text{R3D}}$ and $\hat{\mathbb{G}}^{\text{F3D}}(\cdot, \cdot, \cdot, \cdot)$ (with the appropriate conditional mean estimators plugged in) deliver asymptotically valid confidence bands for the treatment effect estimators with empirical quantile functions, $\bar{\tau}_p^{\text{R3D}}$, $\bar{\tau}_p^{\text{F3D}}$, $\bar{\tau}_{\oplus,p}^{\text{R3D}}$, $\bar{\tau}_{\oplus,p}^{\text{F3D}}$.*

³See Chen et al. (2023) for an analogous setting in the context of distribution-on-distribution regression, as well as Zhou and Müller (2024).

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 of [Calonico et al. \(2014\)](#), this is equivalent to using explicitly bias-corrected linear estimators.

In the simulations, I estimate the quantile treatment effects τ^{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 original estimator in [Frandsen et al. \(2012\)](#).

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

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

$$(17) \quad \begin{aligned} \mu_i &\sim N(5 + 5X_i + \delta^+ \Delta, 1), \\ \sigma_i &\sim |N(1 + X_i, 1)|, \end{aligned}$$

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

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

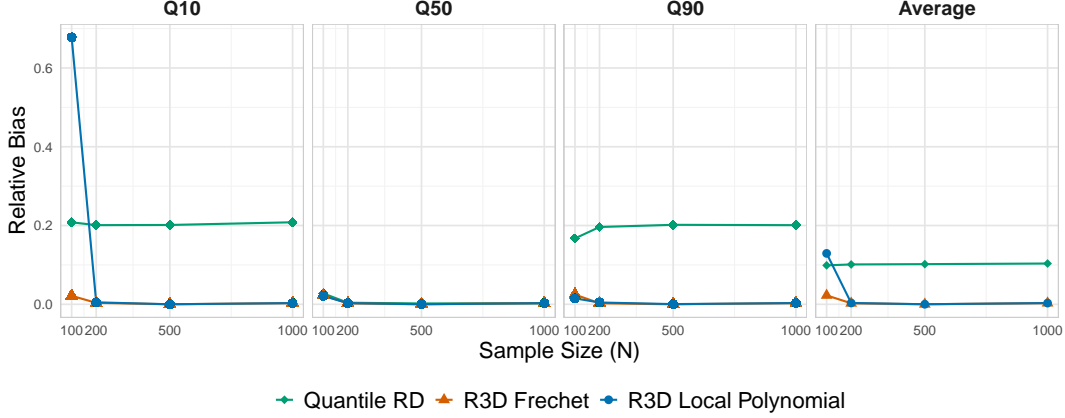
$$(18) \quad Y_i = N(\mu_i + \delta^+ \Delta, 1) + 2 \text{Exp}(\lambda_i + \delta^+ \Delta_\lambda).$$

In both setups, I let Δ 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 are estimated by averaging across a large number of simulated quantile functions.

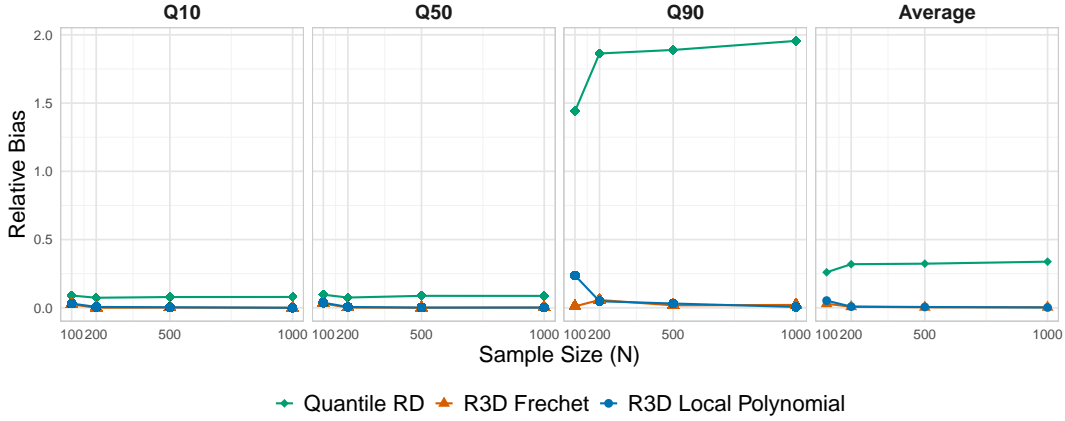
Figure 3 shows the estimators' performance in terms of relative bias, which is the magnitude of the estimated bias at a given quantile as a proportion of the treatment effect at that quantile. I set $\Delta = 2$ but the results are similar for other values. The green line (diamonds) shows the quantile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

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

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



(a) Normal Distributions with Normal Means



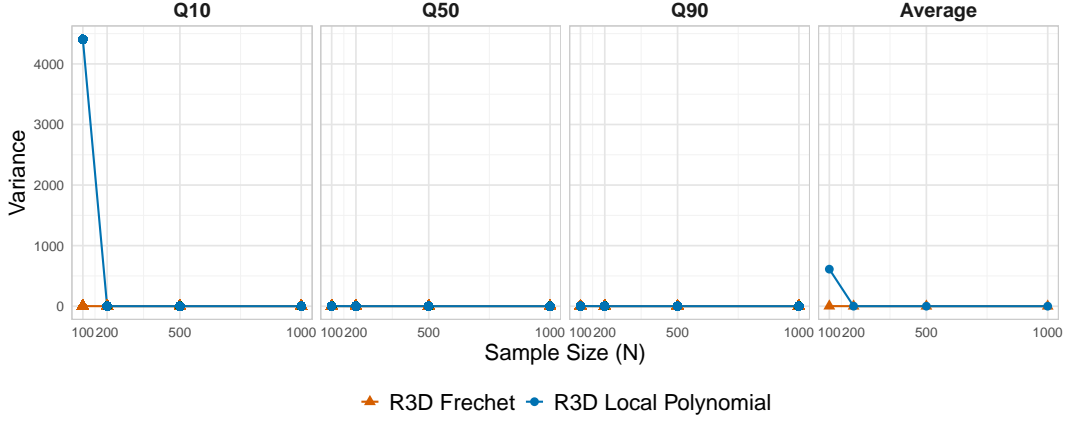
(b) Normal-Exponential Distributions with Normal-Exponential Means

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.

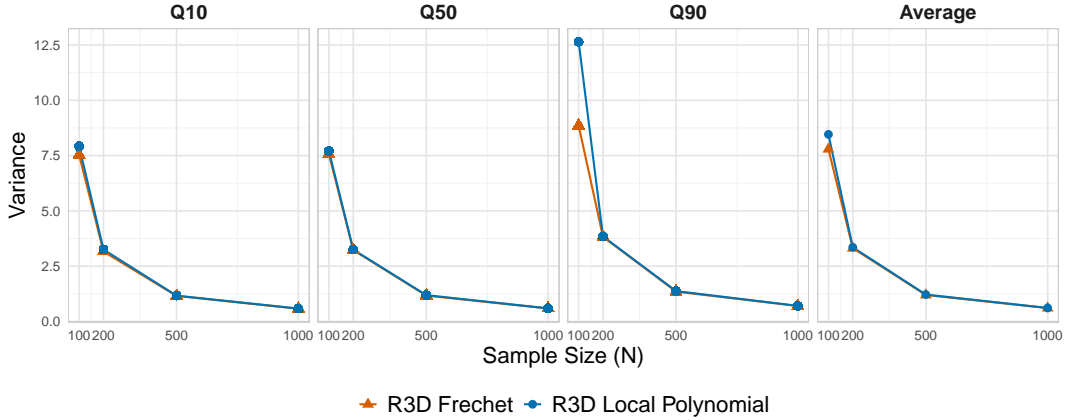
and the blue line (circles) the local polynomial one. In line with theoretical expectations, the quantile RD estimator appears to be inconsistent and suffers from large finite sample bias, with a relative bias that is at least an order of magnitude higher than the R3D estimators', for some quantiles. As mentioned, this is not a defect of Q-RD per se; it targets a different estimand under different sampling (scalar outcomes), and, as a result, is misspecified in the R3D setting. 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: Simulated Variance of R3D Estimators



(a) Normal Distributions with Normal Means



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

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.

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.

Table 2: Acceptance Probabilities of R3D Estimators

Method:	Fréchet				Local Polynomial			
DGP 1								
Unif. CIs	<i>n</i>	$\Delta=0$	$\Delta=1.14$	$\Delta=2.27$	<i>n</i>	$\Delta=0$	$\Delta=1.14$	$\Delta=2.27$
	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
Homogen.	<i>n</i>	$\Delta=0$	$\Delta=1.14$	$\Delta=2.27$	<i>n</i>	$\Delta=0$	$\Delta=1.14$	$\Delta=2.27$
	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
Nullity	<i>n</i>	$\Delta=0$	$\Delta=1.14$	$\Delta=2.27$	<i>n</i>	$\Delta_\mu=0$	$\Delta=1.14$	$\Delta=2.27$
	500	–	9.36	0	500	–	10.12	0
	1000	–	0.28	0	1000	–	0.28	0
	2000	–	0	0	2000	–	0	0
DGP 2								
Unif. CIs	<i>n</i>	$\Delta=0$	$\Delta=1.86$	$\Delta=3.72$	<i>n</i>	$\Delta=0$	$\Delta=1.86$	$\Delta=3.72$
	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
Homogen.	<i>n</i>	$\Delta=0$	$\Delta=1.86$	$\Delta=3.72$	<i>n</i>	$\Delta=0$	$\Delta=1.86$	$\Delta=3.72$
	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
Nullity	<i>n</i>	$\Delta=0$	$\Delta=1.86$	$\Delta=3.72$	<i>n</i>	$\Delta=0$	$\Delta=1.86$	$\Delta=3.72$
	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 (17) and (18). All simulations used 2,500 repetitions and 5,000 bootstrap replications and estimated quantile treatment effects at the 9 deciles. Values of Δ reflects Cohen’s d of 0, 0.5, and 1.

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 Δ in Table 2. The values of Δ 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 converges 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 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. The timeframe was chosen to ensure 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 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 con-

trol. Indeed, while the effects remain similar when including post-2010 data, their precision and magnitude decrease (see Figure A-6).

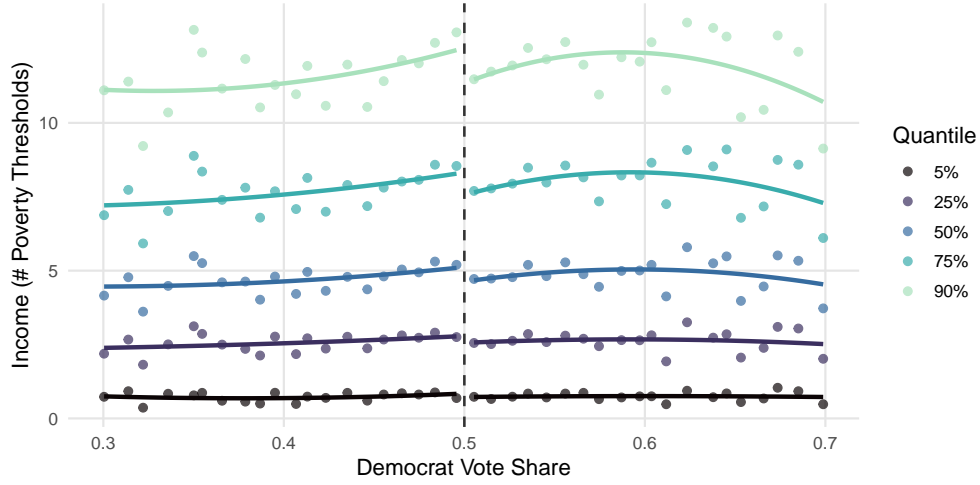
I combine these data with family-level income data from the UNICON extract of the March Current Population Survey (CPS) for the final year of the state governor’s tenure, in order to capture the cumulative effect of that tenure on the income distribution. Practically, this means the election data is 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 instead of a Republican one. The outcome variable $Z_{ijt'}$ is real income of family j in state i in year $t' = t + t_i$, where t_i 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 3.3.1. 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 $j = 1, \dots, n_i$. Based on that, I construct the empirical income quantile functions $\hat{Q}_{Y_{it}}$, 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 1×10^{-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 (Calonico et al., 2015a) 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, $\frac{1}{|B_j|} \sum_{j \in B_i} \hat{Q}_{Y_j}(q)$, with

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.

$B_j = \{i : X_i \in [x_{j,\min}, x_{j,\max}]\}$ 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 $\hat{Q}_{Y_{it}}$. For these, 90% uniform confidence bands are constructed using the bootstrap algorithm sketch in 2.6, 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 Related Literature

This application 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 results 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.

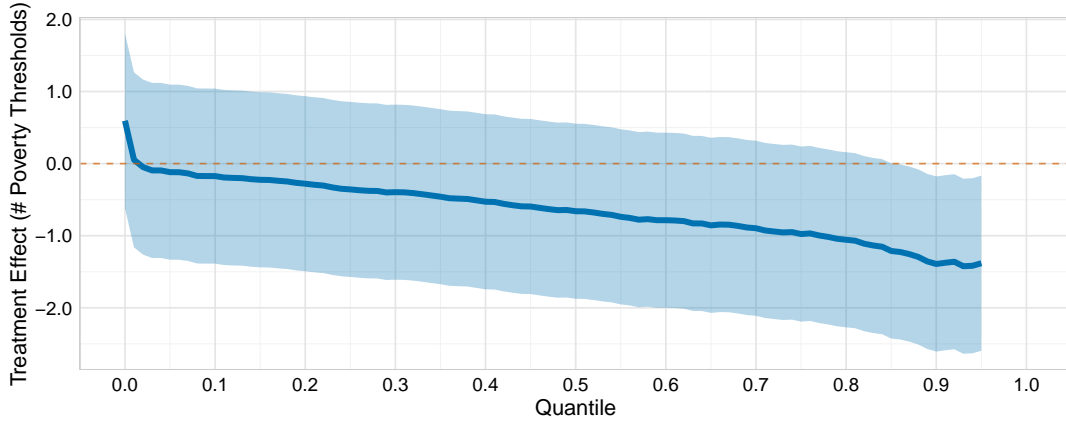
4.2.3 Results

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.

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 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 0.043 and 0.069, 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. That aligns with the small effects of local economic conditions on voting behavior estimated in the literature on retrospective voting (Healy and Malhotra, 2013). Furthermore, I check whether

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: 0.043, treatment homogeneity p-value: 0.069, IMSE-optimal bandwidth: 0.22.

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 -0.64 and significant at the 90% level, while the family-level estimate is -0.891 and marginally significant at the 5% level. As predicted theoretically, the state-level estimates are in line with the average of the LAQTEs produced by the R3D estimators, which is -0.646 , while the family-level estimates are 40% stronger, reflecting a state-size weighted average effect. 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. In line with the simulations above, the estimated effects exhibit substantial bias. Specifically, the estimated quantile treatment effects are, on average, 20% stronger than the R3D estimates.⁵ Moreover, the estimated confidence bands are extremely

⁵A comparison of the 4 (implied) average treatment effects is given in Figure A-10.

narrow, giving the impression that there are negative income effects across the distribution. By contrast, the R3D confidence bands are generally not significant except at the top 10% of the distribution. This stark difference reflects the fact that the quantile RD estimator does not account for across-group heterogeneity, ignoring the variation induced by the random distributions and hence estimating artificially tight confidence bands.

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, a distributional version of the local polynomial regression that regresses on the entire quantile function at once. I establish the asymptotic normality of the estimator and develop uniform, debiased confidence bands that can be estimated with a multiplier bootstrap. Additionally, I introduce a data-driven bandwidth selection procedure for distribution-valued outcomes that is optimal for an integrated version of the mean-squared error. Simulations confirm the robustness of these theoretical properties, demonstrating good finite-sample performance and correct nominal 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. This evidence underscores the method’s ability to uncover nuanced distributional impacts beyond average treatment effects.

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, Titiunik, Vazquez-Bare and Keele, 2016), or multivariate outcome distributions (Chen and Müller, 2023). In Van Dijke (2025), I extend the idea of testing for discontinuities in non-parametric regression functions to a general metric-space setting, allowing for outcomes beyond distributions such as networks or covariance matrices, while Kurisu et al. (2025b) propose a way to define treatment effects in that setting. 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 distribution-valued 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.
- Barreca, A. I., Lindo, J. M. and Waddell, G. R. (2016), ‘Heaping-induced bias in regression-discontinuity designs’, *Economic Inquiry* **54**(1), 268–293.
- Bartalotti, O. and Brummet, Q. (2017), Regression discontinuity designs with clustered data, in M. D. Cattaneo and J. C. Escanciano, eds, ‘Regression Discontinuity Designs: Theory and Applications’, Vol. 38, Emerald Publishing Limited, pp. 383–420.
- Bauschke, H. H., Combettes, P. L., Bauschke, H. H. and Combettes, P. L. (2017), *Convex analysis and monotone operator theory in Hilbert spaces*, Springer.
- Bertanha, M. (2020), ‘Regression discontinuity design with many thresholds’, *Journal of Econometrics* **218**(1), 216–241.
- Borusyak, K. and Kolesman-Shemer, M. (2024), ‘Regression discontinuity aggregation, with an application to the union effects on inequality’, *arXiv preprint arXiv:2501.00428*.
- Callaway, B. and Li, T. (2019), ‘Quantile treatment effects in difference in differences models with panel data’, *Quantitative Economics* **10**(4), 1579–1618.
- 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. (2015a), ‘Optimal data-driven regression discontinuity plots’, *Journal of the American Statistical Association* **110**(512), 1753–1769.
- Calonico, S., Cattaneo, M. D. and Titiunik, R. (2015b), ‘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. and Vazquez-Bare, G. (2016), ‘Inference in regression discontinuity designs under local randomization’, *The Stata Journal* **16**(2), 331–367.
- 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.
- Chang, J. T. and Pollard, D. (1997), ‘Conditioning as disintegration’, *Statistica Neerlandica* **51**(3), 287–317.
- 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 treatment 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 Santos, A. (2019), ‘Inference on directionally differentiable functions’, *The Review of Economic Studies* **86**(1), 377–412.

- 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 design’, *Journal of Business & Economic Statistics* **37**(4), 736–748.
- Ghanem, D., Kédagni, D. and Mourifié, I. (2023), ‘Evaluating the impact of regulatory policies on social welfare in difference-in-difference settings’, *arXiv preprint arXiv:2306.04494*.
- 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. (2010), ‘Generalized continuous isotonic regression’, *Statistics & probability letters* **80**(3-4), 248–253.
- 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 Dijke, 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.
- Kolesár, M. and Rothe, C. (2018), ‘Inference in regression discontinuity designs with a discrete running variable’, *American Economic Review* **108**(8), 2277–2304.
- Kroshnin, A., Spokoiny, V. and Suvorikova, A. (2021), ‘Statistical inference for bures–wasserstein barycenters’, *The Annals of Applied Probability* **31**(3), 1264–1298.
- Kurusu, D., Zhou, Y., Otsu, T. and Müller, H.-G. (2024), ‘Geodesic causal inference’, *arXiv preprint arXiv:2406.19604*.
- Kurusu, D., Zhou, Y., Otsu, T. and Müller, H.-G. (2025a), ‘Geodesic synthetic control methods for random objects and functional data’, *arXiv preprint arXiv:2505.00331*.
- Kurusu, D., Zhou, Y., Otsu, T. and Müller, H.-G. (2025b), ‘Regression discontinuity designs for functional data and random objects in geodesic spaces’, *arXiv preprint arXiv:2506.18136*.
- 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 Card, D. (2008), ‘Regression discontinuity inference with specification error’, *Journal of Econometrics* **142**(2), 655–674.
- 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, L., Thomas, B. S., Piegorsch, W. W., Scott, J. and Carvalho, C. (2019), ‘A projection approach for multiple monotone regression’, *arXiv preprint arXiv:1911.07553*.

- 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.
- 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.
- 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 Dijke, D. (2025), ‘A test for jumps in metric-space conditional means’, *arXiv preprint arXiv:2507.04560*.
- Van Dijke, 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.
- Villani, C. et al. (2008), *Optimal transport: old and new*, Vol. 338, Springer.
- Yang, F. and Barber, R. F. (2019), ‘Contraction and uniform convergence of isotonic regression’.
- Zarantonello, E. H. (1971), Projections on convex sets in hilbert space and spectral theory: Part i. projections on convex sets: Part ii. spectral theory, in ‘Contributions to nonlinear functional analysis’, Elsevier, pp. 237–424.
- 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 $(\Omega^x, \mathcal{F}^x)$*

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

A-2 Derivation of Local Polynomial Regression Weights

The aim is to estimate

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

via the following one-sided weighted least squares:

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

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

Define,

$$\mathbf{X}_{\pm} = \begin{pmatrix} r_p\left(\frac{X_1 - c}{h}\right)^\top \\ r_p\left(\frac{X_2 - c}{h}\right)^\top \\ \vdots \\ r_p\left(\frac{X_n - c}{h}\right)^\top \end{pmatrix}, \quad \mathbf{W}_{\pm} = \text{diag}\left\{ \delta_i^\pm K\left(\frac{X_i - c}{h}\right) : i = 1, \dots, n \right\},$$

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

The solution to the above least-squares problem is:

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

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

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

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:

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

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

$$\begin{aligned} \hat{m}_{\pm,p}(q) &= \sum_{i=1}^n \left[s_{\pm,in}^{(p)}(h) \right] Q_{Y_i}(q), \quad \text{where} \\ s_{\pm,in}^{(p)}(h) &= \delta_i^\pm e_0^\top (\mathbf{X}_\pm^\top \mathbf{W}_\pm \mathbf{X}_\pm)^{-1} r_p\left(\frac{X_i - c}{h}\right) K\left(\frac{X_i - c}{h}\right). \end{aligned}$$

A-3 Overview of Local Fréchet Regression

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

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,

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

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

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

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

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

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),

$$(A-2) \quad d_{W_2}^2(Y_1, Y_2) = \int_0^1 (Q_{Y_1}(q) - Q_{Y_2}(q))^2 dq,$$

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),

$$Q_{F_\oplus}(q) = EQ_Y(q) = \int_{\mathcal{Y}} Q_Y(q) dP(Y).$$

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 Ω 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,

$$(\beta_0^*, \beta_1^*) = \operatorname{argmin}_{\beta_0, \beta_1} \int K_h(x' - x) \times \left[\int z dF_{Z|X}(x', z) - (\beta_0 + \beta_1(x' - x)) \right]^2 dF_X(x'),$$

with $K_h(\cdot) = h^{-1}K(\cdot/h)$ with K a smoothing kernel and h a bandwidth. Defining $\mu_j =$

$E[K_h(X-x)(X-x)^j]$ and $\sigma_0^2 = \mu_0\mu_2 - \mu_1^2$, the solution β_0^* can be written as,

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

with weight function,

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

which corresponds to the local Fréchet mean,

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

Just as with the definition of the classical Fréchet mean, this can be generalized to $Y \in \Omega$ on a general metric space as,

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

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

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

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

with the empirical weights,

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

where

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

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 Ω .

Local Fréchet regression “from the left and right”, as considered in the main text, simply requires adding a δ_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-3. *The projected local polynomial regression estimator in (4) 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}) ,*

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

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,

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

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

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

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

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,

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

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

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

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. \square

A-4 Implementation Details

A-4.1 First-Stage Estimators

Let $t \leq p$, $t \in \mathbb{N}_+$ and denote $\delta_x^\pm := 1\{x \gtrless 0\}$. 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,

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

and

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

where

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

and

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

with $\hat{\alpha}_{+,t}$ defined in (4) and similarly

$$\hat{\alpha}_{\pm,T,p} = \arg \min_{\alpha \in \mathbb{R}^{p+1}} \sum_{i=1}^n \delta_i^\pm K\left(\frac{X_i}{h}\right) \left[T_i - \alpha^\top r_p\left(\frac{X_i}{h}\right) \right]^2.$$

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 $\mathcal{E}_1(y, t, x, q) 1(|x/h_1(q)| \leq 1)$ and $\mathcal{E}_2(y, t, x, q) 1(|x/h_2(q)| \leq 1)$ on $[\underline{c}, \bar{c}] \times \mathcal{Y} \times \{0, 1\}$. 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 $\hat{f}_X(0) := \frac{1}{nb} \sum_{i=1}^n K(X_i/b)$ with $b = b_n \rightarrow 0$ and $nb \rightarrow \infty$.

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, $r_s(u) = (1, u, \dots, u^s)$, $\Gamma_s^\pm = \int_{\mathbb{R}_\pm} K(u) r_s(u) r_s(u)' du$, $\Lambda_{s,s+1} = \int_{\mathbb{R}} u^{s+1} r_s(u) K(u) du$, $\Lambda_{s,s+1}^\pm = \int_{\mathbb{R}_\pm} u^{s+1} r_s(u) K(u) du$, 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:

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

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

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

Similar derivations for the variance expressions imply that

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

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

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

Based on the standard bias-variance expression $\text{MSE} = \text{Bias}^2 + \text{Var}$, 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,

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

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 $\sigma_{2,2}(q, q|0^{\pm}) = \lim_{x \rightarrow 0^{\pm}} \text{Var}(T|X = x)$ in the formulas above and noting that the optimal bandwidth will be the same for all q .

A-4.2.2 Fréchet Estimator

Define

$$\text{IMSE}[\hat{\tau}_s] = \int_a^b \text{MSE}[\hat{\tau}_s(q)] \, dq = \int_a^b \left[\text{Bias}[\hat{\tau}_s(q)]^2 + \text{Var}[\hat{\tau}_s(q)] \right] \, dq.$$

Putting the above expansions together, it follows that

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

where

$$A_s = \int_a^b (B_+(q) - B_-(q))^2 \, dq$$

and

$$B_s = \int_a^b (V_+(q) + V_-(q)) \, dq.$$

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

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

By Lemma A-9,

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

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

Then write,

$$\text{IMSE}[\hat{\tau}_{\oplus,s}] - \text{IMSE}[\hat{\tau}_s] = \int_a^b \left[\text{MSE}(\hat{\tau}_{\oplus,s}(q)) - \text{MSE}(\hat{\tau}_s(q)) \right] \, dq.$$

Pointwise in q ,

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

By Cauchy–Bunyakovsky-Schwarz,

$$|E[(\hat{\tau}_s - \tau) \Delta_{\oplus,s}]| \leq \left(E[(\hat{\tau}_s - \tau)^2] \right)^{1/2} \left(E[\Delta_{\oplus,s}^2] \right)^{1/2}.$$

Under Theorem 1 and Lemma A-9,

$$\sup_{q \in [a,b]} E[(\hat{\tau}_s(q) - \tau(q))^2] = O\left(\frac{1}{nh}\right), \quad \sup_{q \in [a,b]} E[\Delta_{\oplus,s}(q)^2] = o\left(\frac{1}{nh}\right),$$

so the integrand satisfies

$$\sup_{q \in [a,b]} |\text{MSE}[\hat{\tau}_{\oplus,s}(q)] - \text{MSE}[\hat{\tau}_s(q)]| = o\left(\frac{1}{nh}\right).$$

Hence

$$\text{IMSE}[\hat{\tau}_{\oplus,s}] = \text{IMSE}[\hat{\tau}_s] + o\left(\frac{1}{nh}\right).$$

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

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

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 $h_{k,n}^0$ for local polynomial fits of order p using the bias-variance formulas derived above,

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

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

$$\begin{aligned} C_{1,0}(q) &= e'_0 \left[(\Gamma_s^+)^{-1} \Lambda_{s,s+1}^+ \frac{\partial^{s+1} \bar{m}_+(q)}{\partial x^{s+1}} - (\Gamma_s^-)^{-1} \Lambda_{s,s+1}^- \frac{\partial^{s+1} \bar{m}_-(q)}{\partial x^{s+1}} \right] / (s+1)!, \\ C'_{1,0}(q) &= \frac{1}{\hat{f}_X(0)} e'_0 \left[\bar{\sigma}_{1,+}^2(q) (\Gamma_s^+)^{-1} \Psi_s^+ (\Gamma_s^+)^{-1} + \bar{\sigma}_{1,-}^2(q) (\Gamma_s^-)^{-1} \Psi_s^- (\Gamma_s^-)^{-1} \right] e_0. \end{aligned}$$

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

Step 2: First-Stage Local Polynomial Fits.

Using the pilot bandwidths $\{h_{1,n}^0(q), h_{2,n}^0(q)\}$ from Step 1, run local polynomial regres-

sions of order s at each quantile q ,

$$\check{\alpha}_{\pm,s}(q) = \arg \min_{\alpha \in \mathbb{R}^{s+1}} \sum_{i=1}^n \delta_i^{\pm} K\left(\frac{X_i}{h_{k,n}^0}\right) \left[Q_{Y_i}(q) - \alpha^\top r_s\left(\frac{X_i}{h_{k,n}^0}\right)\right]^2,$$

which gives the first-stage estimates

$$[\check{m}_{\pm}(q), \dots, \frac{\partial^s \check{m}_{\pm}(q)}{\partial x^s}] = \check{\alpha}'_{\pm,s} \text{diag} [1, 1!/h_{1,n}^0, \dots, s!/(h_{1,n}^0)^s]$$

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

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

as well as the corresponding variance estimates,

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

and analogously for $\check{E}[T | X = x]$, $\check{\sigma}_{22}(q, q | 0^\pm)$. Then the uniform consistency of $\check{E}[Q_Y(q) | X = x]1\{|x| \leq h_{1,n}^0(q)\}$ and $\check{E}[T | X = x]1\{|x| \leq h_{2,n}^0\}$ 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 $\hat{h}_{k,n}(q)$ for each quantile q .
- *Fréchet estimator*: $h_{\oplus,1}^*$: IMSE across $q \in [a, b]$ can be obtained by averaging the bias² 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,

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

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

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 , $\mathcal{T}^* := (q_1, \dots, q_M) \subset [a, b] \subset (0, 1)$.
- 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 (16). If the entire population is observed, these estimates coincide with the true distribution and quantile function, otherwise the results in Section ?? 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 $X_i \gtrless 0$.

- (ii) (Sharp RDD) Set

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

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

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

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

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

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

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 ,

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

where $g_1(Y_i, q) = Q_{Y_i}(q)$, $g_2(Y_i, T_i) = T_i$, and $\tilde{E}\{\dots \mid X_i\}$ 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 $\{\{\xi_i^b\}_{i=1}^n\}_{b=1}^B$ i.i.d. from $N(0, 1)$, independent of the data, for $b = 1, \dots, B$. For $k \in \{1, 2\}$, compute

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

where $\delta_i^\pm = \mathbf{1}\{X_i \gtrless 0\}$, $r_p(\cdot)$ 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 :

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

(ii) (Fuzzy RDD) For each b :

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

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

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

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

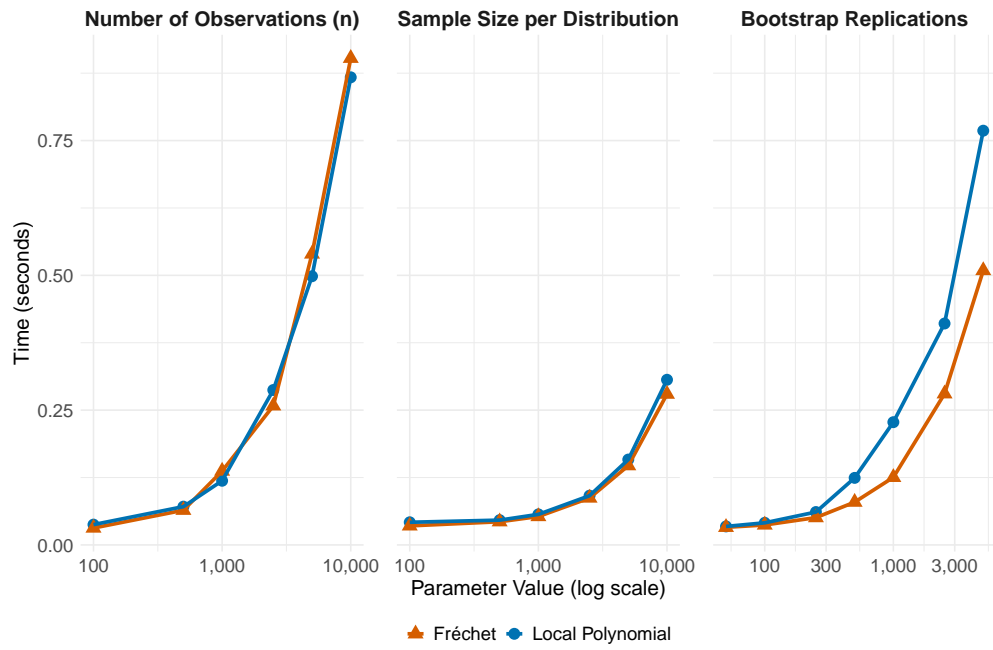
Then, an asymptotically valid uniform $(1 - \lambda)100\%$ confidence band for $\tau^{\text{R3D}}(q)$ (sharp) or $\tau^{\text{F3D}}(q)$ (fuzzy) on $q \in [a, b]$ is given by:

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

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,

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

and similarly for $x \rightarrow 0^-$. 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 expectation operator. \square

Proof of Lemma 3.

Proof. It holds that,

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

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](#),

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

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

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

Combining these two derivations with Assumption [I3](#) gives the result. \square

A-5.2 Asymptotic Results

Lemma A-4 (Quantile functionals are VC type). *Let $\Theta = [a, b] \subset (0, 1)$ (or $\Theta = [0, 1]$ if all $Y \in \mathcal{Y}$ have compact support). For each $q \in \Theta$ and $Y \in \mathcal{Y}$, define $f_q(Y) = Q_Y(q) = \inf\{y : Y(y) \geq q\}$. Then the class $\mathcal{F} = \{Y \mapsto Q_Y(q) : q \in \Theta\}$ is VC-type.*

Proof. Let $\Theta = [a, b]$ be a compact subset of $(0, 1)$ (or $\Theta = [0, 1]$ in the case where all cdfs in \mathcal{Y} have compact support). For each $q \in \Theta$ and $Y \in \mathcal{Y}$, define

$$f_q(Y) = Q_Y(q) = \inf\{y : Y(y) \geq q\}.$$

Then the family $\mathcal{F} = \{Y \mapsto Q_Y(q) : q \in \Theta\}$ is a VC-subgraph class ((in the sense of [van der Vaart and Wellner \(1996, §2.6.2\)](#))) with index $V(\mathcal{F}) \leq 2$. Indeed, the (non-strict) subgraph of f_q satisfies

$$G_{f_q} = \{(Y, z) : z \leq Q_Y(q)\} = \{(Y, z) : q > Y(z^-)\},$$

where $Y(z^-) = \lim_{t \uparrow z} Y(t)$. Thus membership in G_{f_q} depends only on the scalar $c := Y(z^-)$ via the threshold rule $c < q$. A single point (Y, z) can be shattered: include it by setting $q > Y(z^-)$, exclude it by setting $q \leq Y(z^-)$. 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 $q < Y_1(z_1^-)$, $q \in (Y_1(z_1^-), Y_2(z_2^-)]$, and $q > Y_2(z_2^-)$, respectively. However, the subset $\{(Y_1, z_1)\}$ cannot be realized, due to the monotonicity of cumulative distribution functions and the resulting nested threshold structure. Hence, the largest shattered set has size 1, which implies $V(\mathcal{F}) \leq 2$.

Furthermore, let $F(Y) = \sup_{q \in \Theta} |Q_Y(q)|$. Then F is a measurable envelope for \mathcal{F} ; and by Assumption [L3](#) we have $F \in L^r(P)$ for the required r (when $\Theta \subset (0, 1)$, $F(Y) = \max\{|Q_Y(a)|, |Q_Y(b)|\}$). By Theorem 2.6.7 of [van der Vaart and Wellner \(1996\)](#), there is a universal $K > 0$ such that for all $0 < \varepsilon < 1$ and all probability measures Q on $(\mathcal{Y}, \mathcal{G})$,

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

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

Lemma A-5 (Conditional Expected Quantile Functions are VC type). *The class $\mathcal{F} = \{x \mapsto g_q(x) : q \in [0, 1]\}$ with $g_q(x) = E[Q_Y(q) \mid X = x]$ is 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

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

Denote this family by

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

I claim \mathcal{F} is a VC-subgraph class of finite index. Indeed, by the results in Proposition A-3, 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

$$G_{g_q} = \{(x, z) : z \leq Q_{m_{\oplus}(x)}(q)\} = \{(x, z) : q > m_{\oplus}(x)(z^-)\},$$

where $m_{\oplus}(x)(z^-) := \lim_{t \uparrow z} m_{\oplus}(x)(t)$. Since each $x \in \mathbb{R}$ defines a distinct, unique cdf $m_{\oplus}(x) \in \mathcal{Y}$ by Proposition A-3, the conclusion follows by an identical argument as in Lemma A-4. \square

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 (6) and the F3D estimator (9) so that in both cases, their functions Υ, ψ, ϕ are all equal to the identity operator. The rest of their notation is closely followed for ease of comparison.

Proof of Theorem 1

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

Assumption 1, Chiang et al., 2019. Let $\underline{c} < 0 < \bar{c}$. (i) (a) This part is equivalent to Assumption L1-(i). (b) This part is equivalent to Assumption I2. (ii) (a) The collections of real-valued functions $\{x \rightarrow E[Q_Y(q) \mid X = x] : q \in [a, b]\}$, $\{Y \rightarrow Q_Y(q) : q \in [a, b]\}$ are of VC type with common integrable envelope $F_{\mathcal{E}}$ such that $\int_{\mathcal{Y} \times [\underline{c}, \bar{c}]} |F_{\mathcal{E}}(y, x)|^{2+\epsilon} dP^x(y, x) < \infty$ 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 $\sigma_{kl}(q, q' \mid \cdot) \in C^1([\underline{c}, \bar{c}] \setminus \{0\})$ with bounded derivatives in x and $\sigma_{kl}(q, q' \mid 0^{\pm}) < \infty$. (d) For each $Y \in \mathcal{Y}$, $Q_Y(q)$ is left- or right-continuous in q . (iii) This

part is equivalent to Assumption **K2**. (iv) (a) $K : [-1, 1] \rightarrow \mathbb{R}^+$ 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-4** and **A-5**. A common integrable envelope can be constructed as follows. Define $F_1(y, x) = \sup_{q \in [a, b]} |Q_Y(q)|$ and $F_2(y, x) = \sup_{q \in [a, b]} |E[Q_Y(q) | X = x]|$ and define $F_\varepsilon(y, x) := F_1(y, x) + F_2(y, x)$. Then clearly $\sup_{q \in [a, b]} |Q_Y(q)| \leq F_\varepsilon(y, x)$ and $\sup_{q \in [a, b]} |E[Q_Y(q) | X = x]| \leq F_\varepsilon(y, x)$. Moreover, by Assumption **L3**,

$$\begin{aligned} & \int_{[\underline{c}, \bar{c}] \times \mathcal{Y}} (F_\varepsilon(x, y))^{2+\varepsilon} dP^x(y, x) \\ & \leq 2^{1+\varepsilon} \int_{[\underline{c}, \bar{c}] \times \mathcal{Y}} (F_1(y, x)^{2+\varepsilon} + F_2(y, x)^{2+\varepsilon}) dP^x(y, x) < \infty. \end{aligned}$$

- (ii) (c) The covariance

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

where the second equality follows from the law of total expectation. The variance term $\sigma_{22}(q, q' | X = x) = \text{var}(T | X = x)$ is in $C^1([\underline{c}, \bar{c}] \setminus \{0\})$ by Assumption **L2** (i). Finally, the cross-variance term $\sigma_{11}(q, q' | X = x)$

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

Expand the brackets and note that $E[Q_Y(q)Q_Y(q') | X = x]$ 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,

$$\begin{aligned} & \sqrt{nh_1(q)} \left(\hat{m}_{\pm,p}(q) - m_{\pm}(q) - h_1(q)^{p+1}(q) \frac{e'_0(\Gamma_{\pm,p})^{-1} \Lambda_{p,p+1}^{\pm}}{(p+1)!} \lim_{x \rightarrow 0^{\pm}} \frac{\partial m(q)^{p+1}}{\partial x^{p+1}} \right) \\ &= \sum_{i=1}^n \frac{e'_0(\Gamma_p^{\pm})^{-1} \mathcal{E}_1(Y_i, t_i, X_i, q) r_p\left(\frac{X_i}{h_1(q)}\right) K\left(\frac{X_i}{h_1(q)}\right) \delta_i^{\pm}}{\sqrt{nh_1(q)} f_X(0)} + o_p^x(1) := \sum_{i=1}^n f_{ni}(q, k) + o_p^x(1) \end{aligned}$$

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. \(2019\)](#), $\nu_n^+(q, k) = \sum_{i=1}^n [f_{ni}(q, k) - E f_{ni}(q, k)]$ converges weakly to a tight zero-mean Gaussian process \mathbb{G}_{H^+} with covariance function H^+ defined in the main theorem, where,

$$f_{ni}(q, k) = \frac{e'_0(\Gamma_{+,p})^{-1} r_p\left(\frac{X_i}{h_k(q)}\right)}{\sqrt{nh_k(q)} f_X(0)} \mathcal{E}_k(Y_i, T_i, X_i, q) K\left(\frac{X_i}{h_k(q)}\right) \delta_i^+$$

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

Proof of Theorem 2

Proof. The result for the sharp RD estimator in (6) simply follows from Theorem 1 and the continuous mapping theorem, and similarly for the denominator of the fuzzy RDD estimator. Then, the ratio map

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

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). \square

Proof of Theorem 3

Proof. The result follows from Theorem 1 and Theorem 2 in [Chiang et al. \(2019\)](#). The latter applies because their Assumptions 1–4 are satisfied in my setting. Their Assumption 1 was shown to hold in the proof of Theorem 1. Moreover, their Assumption 2 is satisfied since their operators ψ, ϕ, Υ are trivially Hadamard differentiable in my setting, and by Assumptions [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 $m_{\pm}(\cdot), m_{T,\pm}(\cdot)$. \square

Lemma A-6 (Properties of the projection). *Let $\mathcal{Q} \subset L^2([a, b])$ be the closed convex cone of (equivalence classes of) a.e. nondecreasing functions. For $f \in L^2([a, b])$, let $\Pi_{\mathcal{Q}}(f)$ denote the (unique) L^2 -metric projection onto \mathcal{Q} . For definiteness, for every $u \in \mathcal{Q}$, fix the right-continuous representative on $[a, b)$ and set $u(b) := \lim_{q \uparrow b} u(q)$.⁶ This turns $\Pi_{\mathcal{Q}}$ into a map $\Pi_{\mathcal{Q}} : \ell^\infty([a, b]) \rightarrow \ell^\infty([a, b])$.*

Then for all $f, g \in \ell^\infty([a, b])$, we have,

(i) (Order preserving) *If $f \leq g$ pointwise on $[a, b]$, then $\Pi_{\mathcal{Q}}(f) \leq \Pi_{\mathcal{Q}}(g)$ pointwise on $[a, b]$.*

(ii) (Translation equivariant) *For every constant $c \in \mathbb{R}$, $\Pi_{\mathcal{Q}}(f + c) = \Pi_{\mathcal{Q}}(f) + c$.*

(iii) (1-Lipschitz in $\|\cdot\|_\infty$)

$$\|\Pi_{\mathcal{Q}}(f) - \Pi_{\mathcal{Q}}(g)\|_\infty \leq \|f - g\|_\infty.$$

Proof. Write $\langle \cdot, \cdot \rangle$ for the L^2 inner product and $\|\cdot\|_2$ for the L^2 norm. Recall the following property of projections in Hilbert spaces. For $u = \Pi_{\mathcal{Q}}(f)$,

$$(A-10) \quad \langle f - u, h - u \rangle \leq 0 \quad \forall h \in \mathcal{Q}.$$

Moreover, note that if u, v are nondecreasing, then $u \wedge v$ and $u \vee v$ are nondecreasing; since $|u \vee v| \leq |u| + |v|$ and $|u \wedge v| \leq |u| + |v|$, both belong to L^2 . Hence $u \wedge v, u \vee v \in \mathcal{Q}$.

(i) *Order preservation.* Assume $f \leq g$ pointwise. Let $u = \Pi_{\mathcal{Q}}(f)$ and $v = \Pi_{\mathcal{Q}}(g)$, and set $A := \{q \in [a, b] : u(q) > v(q)\}$. Apply (A-10) with (f, u) and $h = u \wedge v$, and with (g, v) and $h = u \vee v$:

$$\int_{[a, b]} (f - u)(u \wedge v - u) dq \leq 0, \quad \int_{[a, b]} (g - v)(u \vee v - v) dq \leq 0.$$

On A we have $u \wedge v = v$ and $u \vee v = u$, whereas away from A the differences $u \wedge v - u$ and

⁶Every a.e. nondecreasing function admits a version which is nondecreasing and right-continuous on $[a, b)$ with a finite left limit at b . I always work with this canonical version.

$u \vee v - v$ vanish. Therefore

$$\int_A (f - u)(v - u) dq \leq 0, \quad \int_A (g - v)(u - v) dq \leq 0.$$

Summing both inequalities gives

$$\begin{aligned} 0 &\geq \int_A [(f - u)(v - u) + (g - v)(u - v)] dq \\ &= \int_A (u - v) [(g - v) - (f - u)] dq \\ &= \int_A (u - v) [(g - f) + (u - v)] dq. \end{aligned}$$

Since $f \leq g$ and $u > v$ on A , the integrand satisfies

$$(u - v) [(g - f) + (u - v)] \geq (u - v)^2 \geq 0.$$

Hence

$$0 \geq \int_A (u - v) [(g - f) + (u - v)] dq \geq \int_A (u - v)^2 dq,$$

which forces $\int_A (u - v)^2 dq = 0$, i.e. $u \leq v$ a.e. on $[a, b]$.

Then, to upgrade a.e. to pointwise by the choice of representatives, note that if there were $x_0 \in [a, b]$ with $u(x_0) > v(x_0)$, then by right-continuity there exist $\varepsilon, \delta > 0$ such that $u(x) \geq v(x) + \varepsilon$ for all $x \in (x_0, x_0 + \delta)$, contradicting $u \leq v$ a.e. If $u(b) > v(b)$, choose a null set N outside of which $u \leq v$ holds; since N contains no interval, one can pick an increasing sequence $q_n \uparrow b$ with $q_n \notin N$. Then $u(q_n) \leq v(q_n)$ for all n , and taking limits along $n \rightarrow \infty$ yields $u(b) \leq v(b)$, a contradiction. Thus $\Pi_{\mathcal{Q}}(f) = u \leq v = \Pi_{\mathcal{Q}}(g)$ pointwise, proving (i).

(ii) *Translation equivariance.* Let $c \in \mathbb{R}$ and $u = \Pi_{\mathcal{Q}}(f)$. Since $h \in \mathcal{Q} \Rightarrow h + c \in \mathcal{Q}$, we have

$$\|(f + c) - (u + c)\|_2 = \|f - u\|_2 \leq \|f - h\|_2 = \|(f + c) - (h + c)\|_2 \quad \forall h \in \mathcal{Q}.$$

Because \mathcal{Q} is closed and convex, the minimizer is unique, so $\Pi_{\mathcal{Q}}(f + c) = u + c$ in L^2 . Passing to our fixed right-continuous representatives yields the pointwise identity $\Pi_{\mathcal{Q}}(f + c) = \Pi_{\mathcal{Q}}(f) + c$.

(iii) $\|\cdot\|_{\infty}$ -contraction. Let $\delta := \|f - g\|_{\infty}$. Then $f \leq g + \delta$ and $g \leq f + \delta$ pointwise. By (i) and (ii),

$$\Pi_{\mathcal{Q}}(f) \leq \Pi_{\mathcal{Q}}(g + \delta) = \Pi_{\mathcal{Q}}(g) + \delta, \quad \Pi_{\mathcal{Q}}(g) \leq \Pi_{\mathcal{Q}}(f + \delta) = \Pi_{\mathcal{Q}}(f) + \delta.$$

Hence $|\Pi_{\mathcal{Q}}(f) - \Pi_{\mathcal{Q}}(g)| \leq \delta$ pointwise, and taking suprema gives $\|\Pi_{\mathcal{Q}}(f) - \Pi_{\mathcal{Q}}(g)\|_{\infty} \leq$

$\|f - g\|_\infty$. In particular, taking $g \equiv 0$ shows $\|\Pi_{\mathcal{Q}}(f)\|_\infty \leq \|f\|_\infty$, so indeed $\Pi_{\mathcal{Q}} : \ell^\infty([a, b]) \rightarrow \ell^\infty([a, b])$. \square

Lemma A-7 (Hadamard differentiability of the isotonic projection in ℓ^∞). *Let $\mathcal{Q} \subset L^2([a, b])$ be the closed convex cone of a.e. nondecreasing functions and let $\Pi_{\mathcal{Q}}$ be the L^2 metric projection onto \mathcal{Q} . Fix the right-continuous representative so that $\Pi_{\mathcal{Q}} : \ell^\infty([a, b]) \rightarrow \ell^\infty([a, b])$. Suppose $m \in \mathcal{Q}$ is absolutely continuous with $m'(q) \geq \kappa > 0$ for a.e. $q \in [a, b]$. Then $\Pi_{\mathcal{Q}}$ is Hadamard differentiable at m tangentially to $C([a, b])$ with the derivative in the direction h equal to*

$$D\Pi_{\mathcal{Q}}(m)[h] = h \quad \forall h \in C([a, b]).$$

Proof. By [Zarantonello \(1971\)](#), in the Hilbert space L^2 one has that the directional Hadamard derivative is $D\Pi_{\mathcal{Q}}(m)[h] = \Pi_{T_{\mathcal{Q}}(m)}(h)$, where $T_{\mathcal{Q}}(m)$ is the Bouligand tangent cone (this holds generally for metric projections onto convex sets in Hilbert spaces). If $h \in C^\infty([a, b])$, then $\|h'\|_\infty < \infty$ and, for all $0 < t \leq \kappa/(2\|h'\|_\infty)$, $(m + th)' = m' + th' \geq \kappa - t\|h'\|_\infty \geq \kappa/2 > 0$ a.e.; hence $m + th \in \mathcal{Q}$. Thus $h \in T_{\mathcal{Q}}(m)$. Since $C^\infty([a, b])$ is dense in $L^2([a, b])$ and $T_{\mathcal{Q}}(m)$ is a closed cone, $T_{\mathcal{Q}}(m) = L^2([a, b])$, and therefore $D\Pi_{\mathcal{Q}}(m)[h] = h$ for all $h \in L^2$. Since this is clearly linear and continuous, the directional Hadamard differentiability strengthens to full Hadamard differentiability (see e.g. the definitions in [Fang and Santos \(2019\)](#)).

To pass to ℓ^∞ tangentially to $C([a, b])$, let $t_n \downarrow 0$ and $h_n \rightarrow h$ uniformly with $h \in C([a, b])$. Fix $\varepsilon > 0$ and choose $\ell \in C^1([a, b])$ with $\|\ell - h\|_\infty < \varepsilon$ and $\text{Lip}(\ell) = L < \infty$. For n large, $L \leq K_n := \kappa/(2t_n)$, hence $(m + t_n\ell)' \geq \kappa/2$ a.e., so $m + t_n\ell \in \mathcal{Q}$ and $\Pi_{\mathcal{Q}}(m + t_n\ell) = m + t_n\ell$. Using that together with the fact that $\Pi_{\mathcal{Q}}$ is 1-Lipschitz in $\|\cdot\|_\infty$ from [Lemma A-6](#), we have,

$$\begin{aligned} & \left\| \frac{\Pi_{\mathcal{Q}}(m + t_n h_n) - \Pi_{\mathcal{Q}}(m)}{t_n} - h \right\|_\infty \\ & \leq \left\| \frac{\Pi_{\mathcal{Q}}(m + t_n h_n) - \Pi_{\mathcal{Q}}(m + t_n \ell)}{t_n} \right\|_\infty + \left\| \frac{\Pi_{\mathcal{Q}}(m + t_n \ell) - \Pi_{\mathcal{Q}}(m)}{t_n} - h \right\|_\infty \\ & \leq \|h_n - \ell\|_\infty + \|\ell - h\|_\infty \leq \|h_n - h\|_\infty + 2\varepsilon, \end{aligned}$$

where the first inequality follows from the triangle inequality in l^∞ . Letting $n \rightarrow \infty$ and then $\varepsilon \downarrow 0$ shows that $\Pi_{\mathcal{Q}}$ is Hadamard differentiable at m tangentially to $C([a, b])$ with derivative $D\Pi_{\mathcal{Q},m}(h) = h$. \square

Lemma A-8 (Continuity of the Semimetric ρ). *Under Assumptions [L2](#), [L3](#), and [L4](#), the semimetric ρ_\pm induced by the covariance kernel $H_{\pm,p}$ of the Gaussian process $\mathbb{G}_{H\pm}$ in [Theorem 3.1](#) is continuous on $[a, b] \times [a, b]$. Consequently, $\rho_\pm(q, q') \rightarrow 0$ as $|q - q'| \rightarrow 0$.*

Proof. The semimetric is given by

$$\rho_{\pm}(q, q')^2 = \mathbb{E}[(\mathbb{G}_{H\pm}(q, 1) - \mathbb{G}_{H\pm}(q', 1))^2] = H_{\pm,p}(q, q) + H_{\pm,p}(q', q') - 2H_{\pm,p}(q, q').$$

To show that ρ_{\pm} is continuous, it suffices to show that $H_{\pm,p}((q, 1), (q', 1))$ is continuous in (q, q') . All pieces except $\sigma_{11}(q, q' \mid 0^{\pm})$ are continuous under the stated assumptions: $c_1(\cdot)$ is bounded and Lipschitz by Assumption **K2**; $\Gamma_{\pm,p}$ is constant in (q, q') ; $\Psi_{\pm,p}((q, 1), (q', 1)) = \int r_p(u/c_1(q))r_p^{\top}(u/c_1(q'))K(u/c_1(q))K(u/c_1(q'))du$ is continuous in (q, q') by dominated convergence (polynomials and K are continuous and K has compact support). Hence, I focus on proving $\sigma_{11}(q, q' \mid 0^{\pm})$ is continuous in (q, q') .

Step 1: No “synchronous” jumps of random quantiles at fixed levels. Let $\Delta_Y^+(q) := Q_Y(q^+) - Q_Y(q) \geq 0$ denote the right-jump of the (left-continuous) quantile function at q . Define the conditional mean $m_{\pm}(q) := \lim_{x \rightarrow 0^{\pm}} E[Q_Y(q) \mid X = x]$. By Assumption **L4** the map $q \mapsto m_{\pm}(q)$ is absolutely continuous on $[a, b]$ (it has a derivative $m'_{\pm}(q)$ a.e., bounded between $\kappa > 0$ and $K < \infty$). In particular m_{\pm} is continuous, hence it has no jumps: $\lim_{h \downarrow 0} \{m_{\pm}(q+h) - m_{\pm}(q)\} = 0$ for all $q \in [a, b]$. For every $h > 0$, $m_{\pm}(q+h) - m_{\pm}(q) = \lim_{x \rightarrow 0^{\pm}} E[Q_Y(q+h) - Q_Y(q) \mid X = x]$. As $h \downarrow 0$, for each fixed Y the inner difference increases to $\Delta_Y^+(q)$ (monotone convergence), and by Assumption **L3** the class $\{Q_Y(q) : q \in [a, b]\}$ has an $L^{2+\epsilon}$ envelope, so monotone (or dominated) convergence yields $\lim_{h \downarrow 0} \{m_{\pm}(q+h) - m_{\pm}(q)\} = \lim_{x \rightarrow 0^{\pm}} E[\Delta_Y^+(q) \mid X = x]$. The left-hand side is 0 for all q . Because $\Delta_Y^+(q) \geq 0$, I conclude $\lim_{x \rightarrow 0^{\pm}} E[\Delta_Y^+(q) \mid X = x] = 0 \Rightarrow P(\Delta_Y^+(q) = 0 \mid X = 0^{\pm}) = 1$ for each fixed $q \in [a, b]$. Hence, under the conditional law at $X = 0^{\pm}$, the random quantile $Q_Y(\cdot)$ is almost surely right-continuous at every fixed level q .

Step 2: Joint continuity of $(q, q') \mapsto E[Q_Y(q)Q_Y(q') \mid X = x]$. Fix $x \in \mathcal{N} := (-\varepsilon, \varepsilon) \setminus \{0\}$. Let $(q_n, q'_n) \rightarrow (q, q')$ in $[a, b]^2$. By Step 1 together with the left-continuity of quantile functions, for each fixed q we have $P(Q_Y(\cdot)$ continuous at $q \mid X = x) = 1$ and similarly for q' . Hence for $P(\cdot \mid X = x)$ -a.e. Y , $Q_Y(q_n) \rightarrow Q_Y(q)$, $Q_Y(q'_n) \rightarrow Q_Y(q')$. By Assumption **L3** the envelope $F(Y) := \sup_{q \in [a, b]} |Q_Y(q)|$ is in $L^{2+\epsilon}(P)$, so $F(Y)^2$ is integrable conditionally on $X = x$. Dominated convergence then yields $E[Q_Y(q_n)Q_Y(q'_n) \mid X = x] \rightarrow E[Q_Y(q)Q_Y(q') \mid X = x]$. Thus $(q, q') \mapsto E[Q_Y(q)Q_Y(q') \mid X = x]$ is jointly continuous on $[a, b]^2$, for each fixed $x \in \mathcal{N}$.

Step 3: Passing to the one-sided limits $x \rightarrow 0^{\pm}$ uniformly in (q, q') . By Assumption **L2(ii)**, for every (q, q') the map $x \mapsto E[Q_Y(q)Q_Y(q') \mid X = x]$ is C^1 on \mathcal{N} with bounded derivative uniformly over $(q, q') \in [a, b]^2$. In particular, it is uniformly Lipschitz in x near 0: $\sup_{(q, q') \in [a, b]^2} |E[Q_Y(q)Q_Y(q') \mid X = x] - E[Q_Y(q)Q_Y(q') \mid X = x']| \leq L|x - x'|$ for all x, x' sufficiently close to 0. Consequently, the one-sided limits $S_{\pm}(q, q') :=$

$\lim_{x \rightarrow 0^\pm} E[Q_Y(q)Q_Y(q') \mid X = x]$ exist and the convergence is uniform in $(q, q') \in [a, b]^2$. Since uniform limits of continuous functions are continuous, $S_\pm(\cdot, \cdot)$ is jointly continuous on $[a, b]^2$. Finally, $\sigma_{11}(q, q' \mid 0^\pm) = S_\pm(q, q') - m_\pm(q)m_\pm(q')$, and $m_\pm(\cdot)$ is continuous by Assumption L4. Hence $\sigma_{11}(\cdot, \cdot \mid 0^\pm)$ is jointly continuous on $[a, b]^2$.

Since every factor in $H_{\pm,p}((q, 1), (q', 1))$ is continuous in (q, q') , the covariance kernel $H_{\pm,p}$ is continuous on the compact domain $[a, b]^2$. Therefore ρ_\pm is continuous on $[a, b]^2$. \square

Proof of Theorem 4.

Proof. By Theorem 1,

$$\sqrt{nh}(\hat{m}_{\pm,p} - m_\pm) \rightsquigarrow \tilde{\mathbb{G}}_\pm \quad \text{in } \ell^\infty([a, b]),$$

where $\tilde{\mathbb{G}}_\pm := \mathbb{G}_{H_\pm}(\cdot, 1)$ is a mean-zero tight Gaussian process on $[a, b]$ with covariance kernel $H_{\pm,p}$.

The empirical process (and its Gaussian limit) is indexed by a VC-type class and is tight in $\ell^\infty([a, b])$ with sample paths uniformly continuous relative to the canonical semimetric

$$\rho_\pm(q, q')^2 = E[(\mathbb{G}_{H_\pm}(q, 1) - \mathbb{G}_{H_\pm}(q', 1))^2] = H_{\pm,p}(q, q) + H_{\pm,p}(q', q') - 2H_{\pm,p}(q, q'),$$

as in the proof of Lemma 4 of Chiang et al. (2019). Lemma A-8 shows $\rho_\pm(q, q') \rightarrow 0$ when $|q - q'| \rightarrow 0$, hence ρ_\pm is uniformly continuous on the compact $[a, b]$. By Addendum 1.5.8 in van der Vaart and Wellner (1996), \mathbb{G}_{H_\pm} admits a version in $C([a, b])$ a.s., and the same holds for $\mathbb{G}^{\text{R3D}} = \mathbb{G}_{H_+} - \mathbb{G}_{H_-}$. Now, all Fréchet objectives and W_2 distances are L^2 -integrals of quantile functions, so changing representatives on null sets (e.g. from left- to right-continuous) does not alter the objective nor its argmin. I can therefore compute the L^2 -projection in $L^2([a, b])$ and then fix the right-continuous representative merely to view the result as an element of $\ell^\infty([a, b])$; this choice is immaterial for the objective and for limits stated in ℓ^∞ .

Then, by Lemma A-7 and Assumption L4 ($m'_\pm(q) \geq \kappa > 0$ a.e.), Π_Q is 1-Lipschitz in sup-norm and Hadamard directionally differentiable at m_\pm tangentially to $C([a, b])$ with derivative

$$D\Pi_Q[m_\pm](h) = h \quad \forall h \in C([a, b]).$$

The Lipschitz property immediately yields uniform consistency,

$$\|\hat{m}_{\oplus,\pm,p} - m_\pm\|_\infty = \|\Pi_Q(\hat{m}_{\pm,p}) - \Pi_Q(m_\pm)\|_\infty \leq \|\hat{m}_{\pm,p} - m_\pm\|_\infty = o_p(1).$$

Moreover, since $\sqrt{nh}(\hat{m}_{\pm,p} - m_{\pm}) \Rightarrow \mathbb{G}_{H\pm} \in C([a, b])$ a.s. by the above the functional delta method in ℓ^∞ (van der Vaart, 2000, Thm. 20.8) gives

$$\sqrt{nh}(\hat{m}_{\oplus,\pm,p} - m_{\pm}) \rightarrow D\Pi_Q[m_{\pm}](\mathbb{G}_{H\pm}(\cdot, 1)) = \mathbb{G}_{H\pm}(\cdot, 1) \quad \text{in } \ell^\infty([a, b]).$$

This proves the stated convergence for the conditional Fréchet means. The treatment-effect statements follow by linearity for the sharp case and by the standard delta method for ratios for the fuzzy case (as in Theorem 2). \square

Proof of Corollary 1.

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

Proof of Proposition 2.

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

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

I have that,

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

uniformly over $q \in [a, b]$. The last equality follows from Assumption Q1.

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

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

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

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

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 $\bar{m}_{\pm, p}(\cdot) - m_{\pm}(\cdot)$. Finally, the results for the treatment effects with empirical quantile functions, $\bar{\tau}_p^{\text{R3D}}(\cdot)$, $\bar{\tau}_p^{\text{F3D}}(\cdot)$, $\bar{\tau}_{\oplus, p}^{\text{R3D}}(\cdot)$, $\bar{\tau}_{\oplus, p}^{\text{F3D}}(\cdot)$, then follow from identical arguments as in the proof of Theorem 2. \square

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-9. *Under the Assumptions of Theorem 1 and Assumption L4,*

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

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

Let $\epsilon_n := \hat{m}_{\pm, p} - m_{\pm}$. By Theorem 1 (and the continuity of the canonical semimetric; see the proof of Theorem 4),

$$\sqrt{nh} \epsilon_n \rightsquigarrow \mathbb{G}_{H\pm}(\cdot, 1) \quad \text{in } \ell^{\infty}([a, b]),$$

where the limit admits a version with a.s. continuous sample paths. Hence $\|\epsilon_n\|_{\infty} = O_p((nh)^{-1/2})$ and ϵ_n is asymptotically tight in $C([a, b])$.

As argued in the proof of Theorem 4, $m_{\pm}(\cdot)$ is a point in the convex set \mathcal{Q} of quantile functions. Under Assumption L4, it is also strictly increasing. Therefore, the L^2 -metric projection $\Pi_{\mathcal{Q}}: \ell^{\infty} \rightarrow \ell^{\infty}$ is Hadamard differentiable at m_{\pm} tangentially to $C([a, b])$ with derivative equal to the identity operator (Lemma A-7). Concretely, there is a remainder map $r(\cdot)$ such that, uniformly on compact subsets of $C([a, b])$,

$$\Pi_{\mathcal{Q}}(m_{\pm} + \epsilon) = m_{\pm} + \epsilon + r(\epsilon), \quad \|r(\epsilon)\|_{\infty} = o(\|\epsilon\|_{\infty}) \quad \text{as } \|\epsilon\|_{\infty} \rightarrow 0.$$

Furthermore, $\Pi_{\mathcal{Q}}(m_{\pm}) = m_{\pm}$. Setting $\epsilon = \epsilon_n$ in the above expansion, and noting that $\|\epsilon_n\|_{\infty} = O_p((nh)^{-1/2}) \rightarrow 0$, we obtain

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

Hence

$$\hat{m}_{\pm, \oplus, p} - \hat{m}_{\pm, p} = r(\epsilon_n), \quad \text{so} \quad \|\hat{m}_{\pm, \oplus, p} - \hat{m}_{\pm, p}\|_{\infty} = \|r(\epsilon_n)\|_{\infty} = o(\|\epsilon_n\|_{\infty}) = o_p((nh)^{-1/2}).$$

□

A-6 Additional Results

A-6.1 Tables

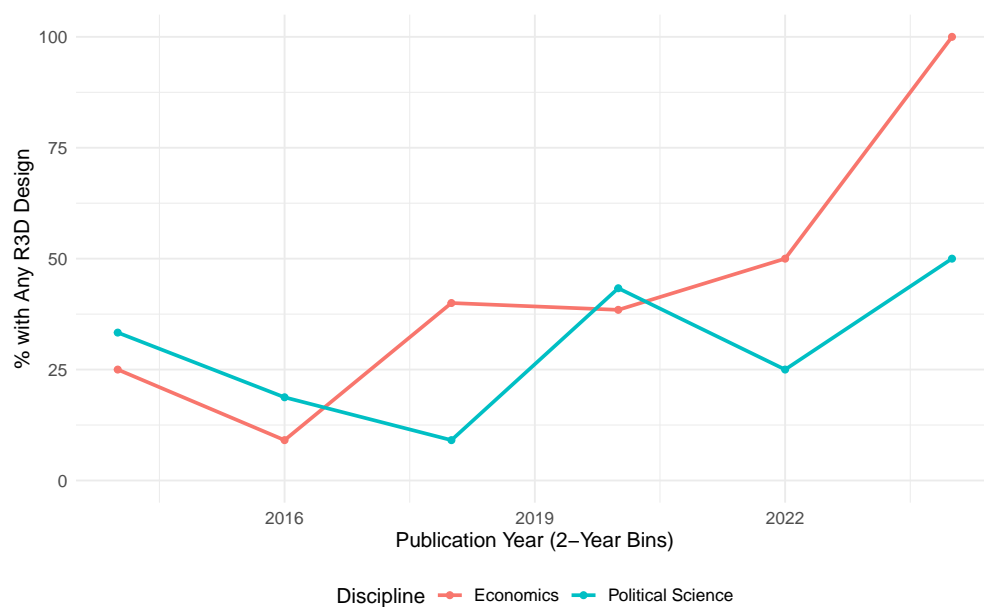
Table A-1: Canonical RD Estimates

Level:	State	Family
Treatment Effect	-0.64 (0.436)	-0.891 (0.52)
Robust 95% CI	[-1.597, 0.111]	[-2.037, 0]
p-value	0.088	0.05
Bandwidth	0.084	0.072
Effective Observations	204	230340

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., 2015b](#)). 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.

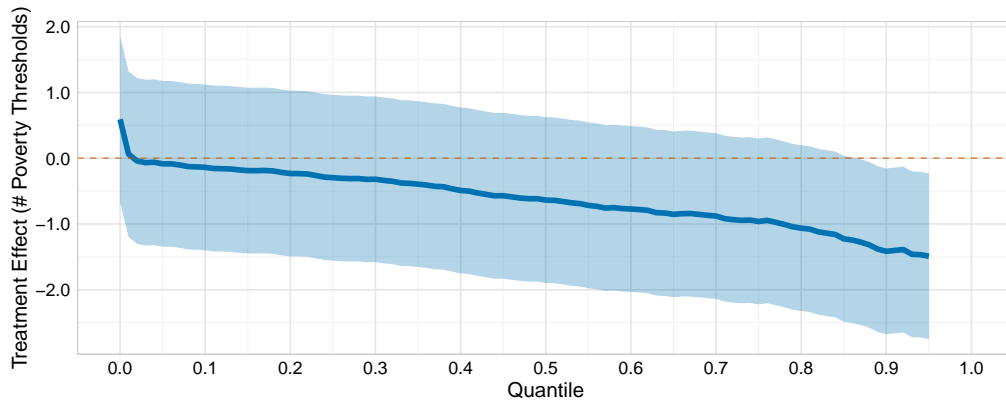
A-6.2 Figures

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



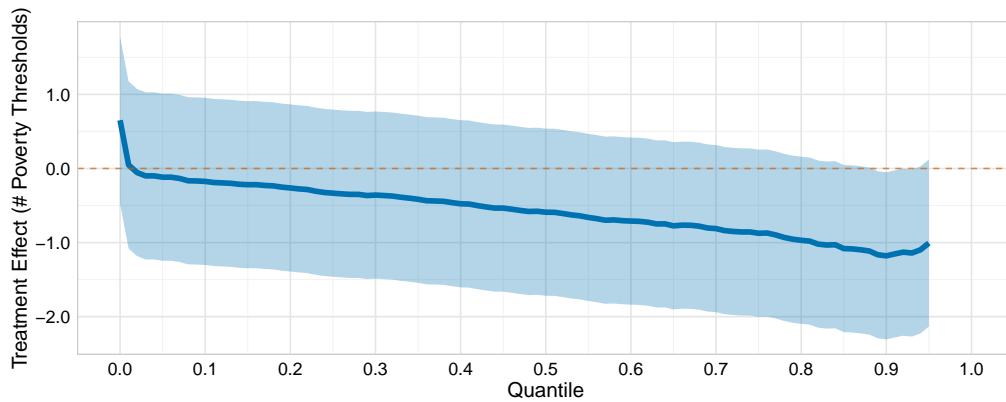
Note: this figure shows percentage of top five Economics and top three Political Science journals with RD designs that fall into an R3D setting in the last 10 years, by 2-year periods. “Any R3D” indicates any form of R3D setting, including settings where the outcome of interest concerns sub-aggregate 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



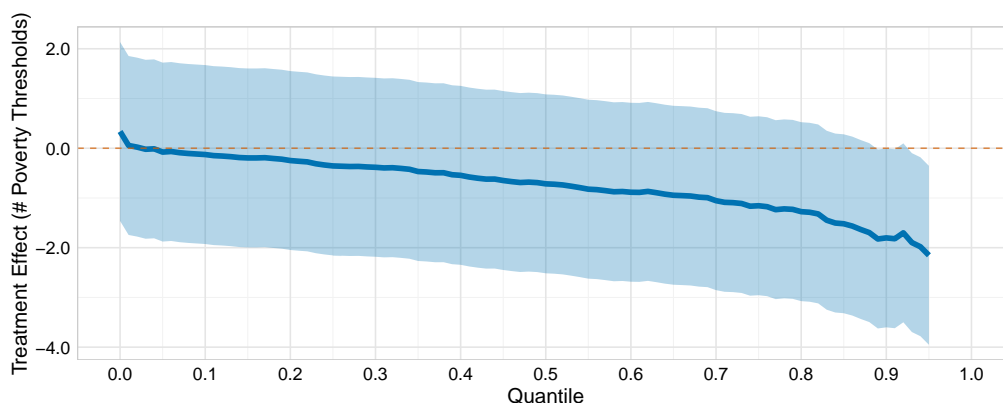
Note: this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average 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: 0.043, treatment homogeneity p-value: 0.061. Average MSE-optimal bandwidths: 0.267.

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



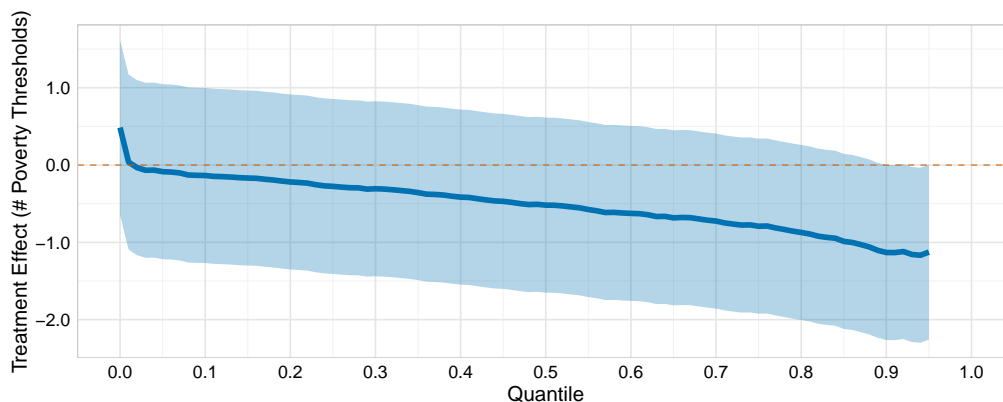
Note: this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average 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: 0.070, treatment homogeneity p-value: 0.117, IMSE-optimal bandwidth: 0.223.

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



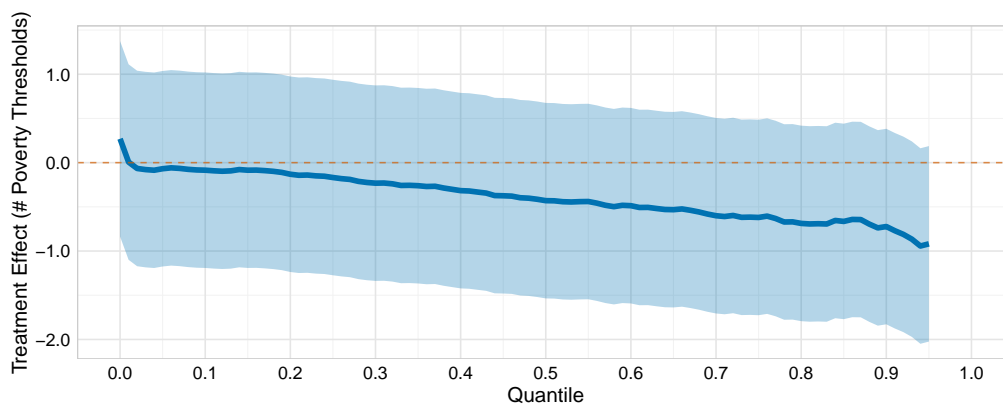
Note: local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average state-level income 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 $1/2 \times$ 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: 0.055, treatment homogeneity p-value: 0.050, IMSE-optimal bandwidth: 0.11.

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



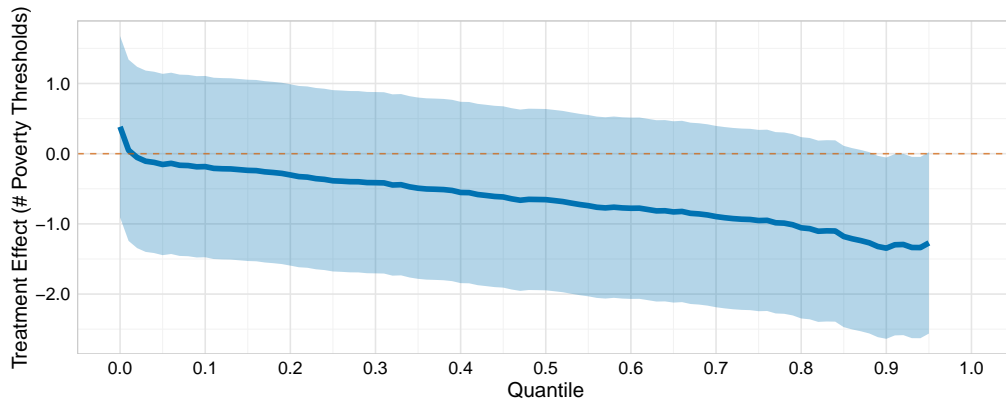
Note: this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average 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: 0.068, treatment homogeneity p-value: 0.093, IMSE-optimal bandwidth: 0.255.

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



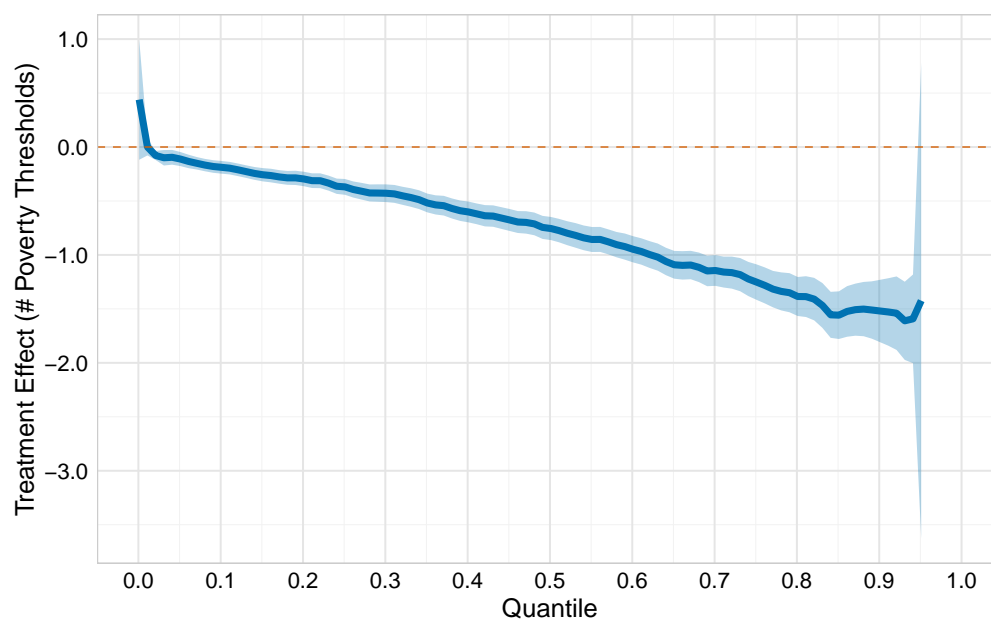
Note: this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average 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: 0.142, treatment homogeneity p-value: 0.164, IMSE-optimal bandwidth: 0.241.

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



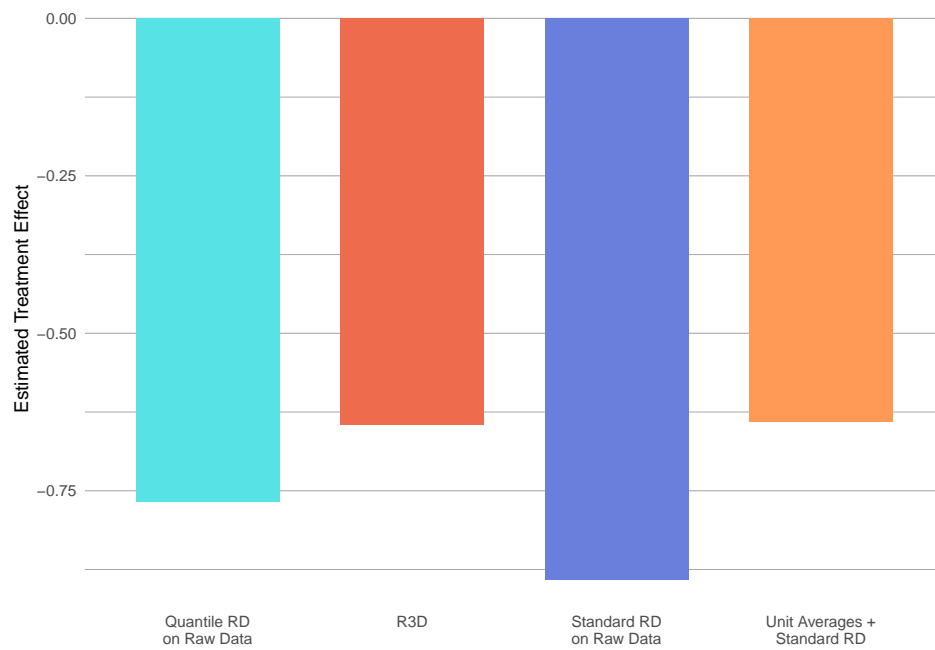
Note: this figure shows local average quantile treatment effects estimates and uniform 90% confidence bands for R3D of effect of Democratic governor control on within-state income distribution. X-axis indicates quantile of the (average) income distribution while Y-axis indicates the difference in average 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: 0.067, treatment homogeneity p-value: 0.120, 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](#).

Figure A-10: Effects of Democratic Governor Control: ATE Comparison



Note: this plot shows the average treatment effects of Democratic governorship on within-state family income estimated either directly or indirectly by 4 estimation approaches, from left to right: 1) quantile RD on the raw data and taking the mean of the quantile treatment effect estimates; 2) R3D on the raw data and taking the mean of the quantile treatment effect estimates; 3) standard RD (local polynomial regression) on the raw data; 4) standard RD on the group-level averages.

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.