

Regression Discontinuity Designs for Functional Data and Random Objects in Geodesic Spaces

Daisuke Kurisu^{1,*}, Yidong Zhou^{2,*}, Taisuke Otsu³, and Hans-Georg Müller^{2,†}

¹Faculty of Economics, The University of Tokyo, Japan

²Department of Statistics, University of California, Davis, USA

³Department of Economics, London School of Economics, UK

September 1, 2026

Abstract

Regression discontinuity designs (RDDs) are widely used for causal inference in observational studies with cutoff-based treatment assignment, primarily for Euclidean outcomes. We propose the geodesic regression discontinuity design (GRDD), which extends RDDs to complex non-Euclidean outcomes, including networks, compositional data, functional data, and other random objects in geodesic metric spaces. Since algebraic operations are unavailable in such spaces, we define the causal effect at the cutoff as the geodesic connecting the local Fréchet means of untreated and treated outcomes, recovering the classical local average treatment effect in the scalar case. Estimation is conducted intrinsically via local Fréchet regression to preserve geometric validity and interpretability. For inference, we adopt an extrinsic approach by embedding the metric space into a Hilbert space, enabling tractable asymptotic analysis. We establish asymptotic normality and develop bootstrap-based procedures for hypothesis testing and confidence intervals for the treatment effect magnitude. We also propose a data-adaptive bandwidth selection method tailored to RDDs in metric spaces and study its empirical performance. Applications include compositional voting outcomes in UK elections and daily CO concentration curves after the Taipei metro introduction, and we extend the framework to fuzzy designs with imperfect compliance.

Keywords: causal inference, Fréchet regression, functional data, program evaluation, random object, regression discontinuity design

MSC2020 subject classifications: 62D20, 62R20

*The first two authors contributed equally to this work and are listed alphabetically.

†Corresponding author: hgmueLLer@ucdavis.edu.

1 Introduction

In causal inference, discontinuities in regression functions induced by an assignment variable can provide useful information for identifying causal effects. As a quasi-experimental design, the regression discontinuity design (RDD) has been widely applied in observational studies across fields such as economics, education, environmental studies, epidemiology, and political science to estimate causal effects at a cutoff point. In an RDD, treatment assignment is determined by whether a running variable crosses a known threshold, so that units on either side of the cutoff receive different treatments. Although the data are observational, this rule creates treatment and control groups that are locally comparable near the cutoff, making the design analogous to a randomized experiment in a neighborhood of the threshold. Causal effects are therefore identified by comparing outcomes for units just above and below the cutoff, with those on one side serving as counterfactuals for those on the other. For the conventional RDD with a scalar outcome, the causal parameters of interest are identified by differences in the left and right limits of the conditional mean functions, and various estimation and inference methods are available in the literature; see, e.g., [Imbens and Lemieux \(2008\)](#), [Cattaneo et al. \(2020\)](#), [Cattaneo and Titiunik \(2022\)](#), an edited volume by [Cattaneo and Escanciano \(2017\)](#), and references therein.

This paper extends RDD analysis beyond the conventional setting of scalar outcomes to a broad class of complex non-Euclidean outcomes. Specifically, we develop a framework for causal inference that accommodates not only scalar or vector-valued responses but also networks, distributions, compositional data, functional data, and other types of random objects that reside in geodesic metric spaces. To achieve this goal, a major challenge is how to quantify a causal effect of interest in this generalized RDD setting, since algebraic operations are not available and taking differences is thus not feasible. To overcome this challenge, we express the causal effect at the cutoff as the geodesic connecting the local Fréchet means of the untreated and treated potential outcomes. This object reduces to

the ordinary local average treatment effect in the scalar or vector case and thus provides a natural geometric extension of the classical RDD estimand.

A key feature of our framework is the distinction between intrinsic estimation and extrinsic inference. For estimation, we adopt an intrinsic approach based on local Fréchet regression, ensuring that all estimated quantities remain in the original metric space $\mathcal{M}$ and thus retain full interpretability as valid elements of $\mathcal{M}$. For statistical inference, however, the situation is fundamentally different: in general metric spaces, the lack of linear structure precludes the formulation of suitable central limit theorems and tractable limiting distributions. We therefore employ an extrinsic approach by embedding $\mathcal{M}$ into a Hilbert space (Schoenberg, 1938; Sejdinovic et al., 2013; Dubey and Müller, 2020), where the linear structure enables asymptotic analysis and hypothesis testing. We note that a key limitation of extrinsic approaches for estimation is their lack of reversibility: such embeddings are typically injective but not surjective, so that an inverse map from the Hilbert space to the original space generally does not exist. As a result, representing the treatment effect in the linear embedding space may involve elements that cannot be mapped back to $\mathcal{M}$ and therefore do not correspond to valid objects. This can lead to distortions even in relatively simple settings such as Wasserstein spaces (Bigot et al., 2017; Cazelles et al., 2018; Pegoraro and Beraha, 2022; Chen et al., 2023) with an ensuing loss of interpretability. Our framework resolves this tension by using intrinsic methods for estimation to preserve validity and interpretability, while conducting inference in the extrinsic framework, since inference does not in any way rely on the inverse map back to $\mathcal{M}$.

Building on this formulation, we estimate the causal effect using local Fréchet regression, a regression method for metric space-valued responses that generalizes local linear regression (Petersen and Müller, 2019), and establish convergence rates for the proposed estimator. For practical implementation, we introduce a data-adaptive bandwidth selection procedure tailored to the boundary nature of regression discontinuity problems, based on

one-sided cross-validation near the cutoff. Furthermore, we develop an inferential framework by deriving asymptotic normality for embedded estimators and proposing multiplier-bootstrap procedures for hypothesis testing and confidence intervals for the magnitude of the treatment effect.

After developing the above results for sharp RDDs under full compliance with the assignment rule, we extend our methodology to the fuzzy RDD setting with imperfect compliance. In the classical case of scalar outcomes, the average treatment effect for compliers at the cutoff can be identified by scaling the discontinuity in the outcome by the discontinuity in the treatment probability (Hahn et al., 2001). Motivated by this framework, we propose a generalized notion of compliers’ average treatment effect at the cutoff as a contrast of mappings of the local Fréchet means of the potential outcomes, establish its identification, and develop an estimation method for compliers’ average treatment effect. Furthermore, we discuss identification and estimation of the geodesic from the local Fréchet mean of the untreated outcome to the treated one for compliers.

Metric statistics for complex data situated in metric spaces is an emerging field of statistics (Dubey et al., 2024), and the literature on causal inference for outcomes in metric spaces is rapidly growing. Lin et al. (2023) and Ecker et al. (2024) studied estimation of the average treatment effect under the unconfoundedness assumption for distributional and functional outcomes, respectively, and their results were further extended by Kurisu et al. (2024) for a general class of geodesic metric spaces. For distributional outcomes, Athey and Imbens (2006), Gunsilius (2023), Gunsilius et al. (2024), and Torous et al. (2024) proposed extensions of the synthetic control and difference-in-differences methods. Using the notions of geodesic metric spaces, Zhou et al. (2025) and Kurisu et al. (2025) developed general methodologies for the difference-in-differences and synthetic control methods, respectively, to accommodate not only functional or distributional outcomes but also other non-Euclidean random objects.

Very recently and independently of our work, [Van Dijke \(2025\)](#) investigated regression discontinuity designs for univariate distributional outcomes; see also [Frandsen et al. \(2012\)](#), who proposed a quantile RDD estimator for scalar outcomes. In particular, [Van Dijke \(2025\)](#) studies the local average quantile treatment effect and develops corresponding nonparametric estimation and inference procedures, including asymptotic theory and bootstrap-based methods. Their approach exploits the representation of univariate distributions via quantile functions, which form a subset of a Hilbert space, and this structure is specific to univariate distributions. As a result, it does not readily extend to multivariate distributions or more general random objects such as networks, symmetric positive-definite matrices, or compositional data. In contrast, our framework formulates RDD for outcomes in general geodesic metric spaces and establishes identification of treatment effects via Fréchet means. As a special case, when the outcome is a univariate distribution equipped with the Wasserstein metric, our identification results recover the local average quantile treatment effect considered in [Van Dijke \(2025\)](#).

We illustrate the proposed methodology primarily through an application to multiparty election outcomes in the United Kingdom, where the response is a vote-share composition across political parties and treatment is determined by prior Conservative victories. This application highlights both the intrinsic handling of compositional outcomes and the new inferential tools developed in this paper, including comparison with componentwise scalar RDD analyses. As a secondary illustration, we also consider functional air-pollution outcomes associated with the Taipei Metro expansion; this application is presented in the Supplement and is intended mainly to demonstrate how GRDD can be implemented for irregularly observed functional data.

This paper is organized as follows. Section 2 presents the basic concepts for geodesic metric spaces and motivating examples. Section 3 introduces geodesic RDDs for the sharp design case, establishes identification of the geodesic local average treatment effect, and

studies estimation via local Fréchet regression. Section 4 develops inference for the magnitude of the treatment effect using Hilbert space embedding and a multiplier bootstrap. Section 5 presents the proposed bandwidth selector and evaluates finite-sample performance in simulations for network-valued outcomes. Section 6 applies the methodology to UK parliamentary election data. Section 7 extends the framework to fuzzy designs with imperfect compliance. Concluding remarks are provided in Section 8.

2 Preliminaries

In this section, we lay the geometric foundations for modeling and analyzing non-Euclidean outcomes, introduce uniquely geodesic spaces, and discuss motivating examples.

2.1 Metric and geodesic spaces

Let $(\mathcal{M}, d)$ be a metric space. A continuous mapping $\gamma : [0, 1] \rightarrow \mathcal{M}$ is called a *curve*, and its length is defined by $\ell(\gamma) = \sup \sum_{i=1}^k d(\gamma(t_{i-1}), \gamma(t_i))$, where the supremum is taken over all finite partitions $0 = t_0 < t_1 < \dots < t_k = 1$ of $[0, 1]$. A curve is said to be *rectifiable* if $\ell(\gamma) < \infty$.

The space $(\mathcal{M}, d)$ is a *length space* if the distance between any two points $\alpha, \beta \in \mathcal{M}$ equals the infimum of the lengths of all rectifiable curves joining them: $d(\alpha, \beta) = \inf_{\gamma} \ell(\gamma)$, where the infimum ranges over all curves with $\gamma(0) = \alpha$ and $\gamma(1) = \beta$. If this infimum is attained for all such pairs, the space is called a *geodesic space* (Bridson and Haefliger, 1999).

A *geodesic* between two points is a curve $\gamma : [0, 1] \rightarrow \mathcal{M}$ satisfying

$$d(\gamma(s), \gamma(t)) = |t - s|d(\alpha, \beta) \quad \text{for all } s, t \in [0, 1],$$

with $\gamma(0) = \alpha$ and $\gamma(1) = \beta$. When such a geodesic exists and is unique for every pair $\alpha, \beta \in \mathcal{M}$, the space is said to be *uniquely geodesic*. Hereafter, we denote the geodesic with

the starting point $\gamma(0) = \alpha$ and the ending point $\gamma(1) = \beta$ as $\gamma_{\alpha,\beta}$.

2.2 Examples of uniquely geodesic spaces

We now introduce five representative examples of uniquely geodesic spaces commonly encountered in applications. We will revisit some of these example spaces to demonstrate the practical performance of the proposed approach in both simulation studies and real data analyses.

Example 1 (Functional Data). *Functional data arise when observations are naturally represented as curves or functions, typically indexed by time or space. Such data are common in longitudinal studies, environmental monitoring, and biomedical imaging (Ramsay and Silverman, 2005; Hsing and Eubank, 2015; Wang et al., 2016). There is also a substantial literature on regression with functional responses; see, for example, Faraway (1997), Chiou et al. (2004) and Morris (2015). Let $\mathcal{T}$ be a compact interval and consider the space $L^2(\mathcal{T})$ of square-integrable functions on $\mathcal{T}$. This Hilbert space is equipped with the inner product $\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t) dt$ and the associated L^2 metric $d_{L^2}(f, g) = \left(\int_{\mathcal{T}} \{f(t) - g(t)\}^2 dt \right)^{1/2}$. The space $L^2(\mathcal{T})$ is uniquely geodesic, with geodesics given by the path $\gamma_{f,g}(t) = (1-t)f + tg$ for $t \in [0, 1]$.*

Example 2 (Compositional Data). *Compositional data describe quantities that represent parts of a whole, such as proportions or percentages. Common examples include vote shares across political parties in election data, relative gene expression levels in transcriptomics, and dietary intake compositions in nutrition studies (Li, 2015). Let $\Delta^{d-1} = \{\mathbf{y} \in \mathbb{R}^d : y_j \geq 0, \sum_{j=1}^d y_j = 1\}$ denote the unit simplex in $\mathbb{R}^d$. Through the square-root transformation $\mathbf{z} = \sqrt{\mathbf{y}}$, compositional data may be mapped to the positive orthant of the unit sphere $\mathcal{S}_+^{d-1}$ (Scealy and Welsh, 2011, 2014). The induced arc-length distance is given by $d_g(\mathbf{z}_1, \mathbf{z}_2) = \arccos(\mathbf{z}_1^\top \mathbf{z}_2)$ for $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}_+^{d-1}$. The geodesic joining $\mathbf{z}_1$ and $\mathbf{z}_2$ follows the shortest great-circle path on the sphere and is uniquely determined when the points are not antipodal.*

Example 3 (Symmetric Positive-Definite Matrices). *Symmetric positive-definite (SPD) matrices are widely used to represent covariance structures, diffusion tensors, and kernel matrices in applications such as neuroimaging (Dryden et al., 2009), signal processing (Arnaudon et al., 2013), and machine learning (Cherian and Sra, 2016). Let Sym_m^+ denote the set of $m \times m$ SPD matrices. This space can be endowed with multiple geodesic metrics. A natural starting point is the Frobenius metric, $d_F(A, B) = \|A - B\|_F$, under which the geodesic between A and B is simply the linear interpolation $\gamma_{A,B}(t) = (1 - t)A + tB$ for $t \in [0, 1]$. Alternatively, under the Log-Euclidean metric (Arsigny et al., 2007) $d_{\text{LE}}(A, B) = \|\log A - \log B\|_F$, the geodesic is given by $\gamma_{A,B}(t) = \exp((1 - t)\log A + t\log B)$. Both metrics induce a uniquely geodesic structure on Sym_m^+ . Other metrics such as the power metric (Dryden et al., 2009) and Log-Cholesky metric (Lin, 2019) also endow Sym_m^+ with a uniquely geodesic structure.*

Example 4 (Networks). *Network data capture relationships or interactions between entities, such as social connections, transportation links, or gene regulatory networks. A common representation is via graph Laplacians (Kolaczyk and Csárdi, 2014; Zhou and Müller, 2022; Severn et al., 2022). Consider the space of undirected, weighted graphs with m nodes and bounded edge weights, where each graph is encoded by its Laplacian matrix. Let $\mathcal{L}$ denote the collection of such Laplacians, and equip it with the Frobenius metric $d_F(L_1, L_2) = \|L_1 - L_2\|_F$. Since $\mathcal{L}$ forms a convex and closed subset of $\mathbb{R}^{m^2}$, the Frobenius metric yields a Euclidean geometry. Geodesics are given by straight-line paths: $\gamma_{L_1, L_2}(t) = (1 - t)L_1 + tL_2$.*

Example 5 (Univariate Probability Distributions). *Probability distributions arise as natural data objects in diverse applications such as econometrics (Petersen et al., 2022), population pyramids (Hron et al., 2016), and multi-cohort studies (Zhou and Müller, 2024). Let $\mathcal{W}$ denote the space of Borel probability measures on a closed interval $\mathcal{I} \subset \mathbb{R}$ with finite second moments. This space becomes a geodesic metric space when equipped with the*

2-Wasserstein metric ([Ambrosio et al., 2008](#)), $d_W(\mu, \nu) = \left(\int_0^1 \{F_\mu^{-1}(s) - F_\nu^{-1}(s)\}^2 ds \right)^{1/2}$, where F_μ^{-1} and F_ν^{-1} are the quantile functions of μ and ν , respectively. The unique geodesic connecting μ and ν under this metric is given by McCann’s interpolant ([McCann, 1997](#)), $\gamma_{\mu, \nu}(t) = \{\text{id} + t(F_\nu^{-1} \circ F_\mu - \text{id})\}_\# \mu$ for $t \in [0, 1]$, where id denotes the identity map, F_μ is the cumulative distribution function of μ , and $\tau_\# \mu$ denotes the pushforward of μ under the map τ . An alternative geometry is induced by the Fisher-Rao metric ([Rao, 1945](#); [Srivastava et al., 2007](#); [Dai, 2022](#)), which is particularly useful when probability densities exist and are smooth. For distributions with the Lebesgue densities f_μ and f_ν , the Fisher-Rao distance is defined as $d_{\text{FR}}(\mu, \nu) = \arccos \left(\int_{\mathcal{I}} \sqrt{f_\mu(x) f_\nu(x)} dx \right)$.

3 Geodesic regression discontinuity designs

In this section, we develop the proposed methodology for geodesic regression discontinuity designs (GRDD), define the causal estimand of interest and establish its identification in uniquely geodesic metric spaces. We then present an estimation strategy based on local Fréchet regression. For clarity of exposition, we focus on the sharp design, in which treatment is deterministically assigned based on whether the running variable crosses a known cutoff. An extension to fuzzy designs that allows for imperfect compliance is discussed in [Section 7](#).

3.1 Identification

We begin with the canonical sharp regression discontinuity design. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $R : \Omega \rightarrow \mathbb{R}$ denote the running variable, and let $c \in \mathbb{R}$ be a fixed cutoff. The cutoff assignment is $Z = 1\{R \geq c\}$. In the sharp design, treatment receipt coincides with the cutoff assignment, so that $T = Z = 1\{R \geq c\}$ almost surely.

For each treatment state $t \in \{0, 1\}$, let $Y(t) : \Omega \rightarrow \mathcal{M}$ denote the corresponding

potential outcome, where $(\mathcal{M}, d)$ is a separable uniquely geodesic metric space such that $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ is a standard Borel space, with $\mathcal{B}(\mathcal{M})$ denoting the Borel σ -field generated by the metric topology. The observed outcome is $Y = Y(T)$. Thus, units with running-variable values below the cutoff receive control, whereas those with values at or above the cutoff receive treatment. We assume that the essential range of $Y(t)$ is bounded, meaning that $\mathbb{P}\{d(\nu_0, Y(t)) \leq D\} = 1$ for some $\nu_0 \in \mathcal{M}$ and $D < \infty$. The fuzzy design considered in Section 7 relaxes the perfect-compliance condition $T = Z$, while retaining the deterministic cutoff assignment $Z = 1\{R \geq c\}$.

For a random object $Y \in \mathcal{M}$, the Fréchet mean of Y , which generalizes the notion of expectation to metric spaces, is defined as

$$E_{\oplus}[Y] = \arg \min_{\nu \in \mathcal{M}} E[d^2(\nu, Y)].$$

The existence and uniqueness of the minimizer is guaranteed for Hadamard spaces (Sturm, 2003) and the example spaces described in Section 2; see Petersen and Müller (2019) for further background. The conditional Fréchet mean of Y given a covariate $X \in \mathbb{R}^p$ (Petersen and Müller, 2019) is analogously defined as

$$E_{\oplus}[Y|X] = \arg \min_{\nu \in \mathcal{M}} E[d^2(\nu, Y)|X],$$

which reduces to the usual conditional expectation when $\mathcal{M} = \mathbb{R}$ and d is the Euclidean distance.

We consider the following assumptions.

Assumption 3.1.

- (i) *Under the stable unit treatment value assumption (SUTVA), the potential outcomes $Y(0), Y(1) \in \mathcal{M}$ are well-defined, and the observed outcome satisfies $Y = Y(0)$ if $T = 0$ and $Y = Y(1)$ if $T = 1$.*
- (ii) *For each $t \in \{0, 1\}$, there exists $\nu_0 \in \mathcal{M}$ such that $E[d^2(\nu_0, Y(t))|R = r] < \infty$ for every*

r in a neighborhood of c . Moreover, the conditional Fréchet mean $E_{\oplus}[Y(t)|R = r]$ exists uniquely for every such r and is continuous at $r = c$.

(iii) We observe an independent and identically distributed sample $\{Y_i, R_i\}_{i=1}^n$ drawn from $(Y, R) \in \mathcal{M} \times \mathbb{R}$.

These assumptions parallel those in the classical RDD framework (e.g., [Imbens and Lemieux, 2008](#)), with the key difference that outcomes now lie in a general metric space $\mathcal{M}$. Because treatment is a deterministic function of the running variable, unconfoundedness $(Y(0), Y(1)) \perp T|R$ holds automatically, while the overlap condition is violated by design. Assumption 3.1(i) imposes SUTVA and therefore rules out interference and hidden variations of treatment, as is standard in RDD settings (see, e.g., [Imbens and Rubin, 2005](#)). Assumption 3.1(ii) is the metric-space analogue of the conventional continuity condition for conditional mean functions in RDD. It permits the counterfactual conditional Fréchet means at the cutoff to be recovered from their respective one-sided limits.

To define the causal estimand of interest, recall that the geodesic defined in Section 2.1 with the starting point $\gamma(0) = \alpha$ and the ending point $\gamma(1) = \beta$ is denoted by $\gamma_{\alpha, \beta}$.

Definition 3.1 (Geodesic local average treatment effect at the cutoff). *The geodesic local average treatment effect (GLATE) at the cutoff c is defined as the geodesic connecting the conditional Fréchet means of the untreated and treated potential outcomes,*

$$\tau_{\text{GRDD}} := \gamma_{E_{\oplus}[Y(0)|R=c], E_{\oplus}[Y(1)|R=c]}.$$

In the classical Euclidean setting where $Y \in \mathbb{R}$, the sharp RDD estimand is given by the jump in the conditional mean function

$$\tau_{\text{RDD}} := E[Y(1)|R = c] - E[Y(0)|R = c].$$

The proposed GLATE can be interpreted as a natural extension of τ_{RDD} to non-Euclidean outcomes, obtained by replacing Euclidean expectations with conditional Fréchet means

and linear differences with geodesics. The resulting geodesic τ_{GRDD} captures both the direction and magnitude of the treatment effect: its direction corresponds to the transition from $E_{\oplus}[Y(0)|R = c]$ to $E_{\oplus}[Y(1)|R = c]$, analogous to the sign of τ_{RDD} , while its length, $d(E_{\oplus}[Y(0)|R = c], E_{\oplus}[Y(1)|R = c])$, corresponds to the magnitude of the effect, analogous to the absolute value of τ_{RDD} .

Remark 3.1 (Interpretation of GLATE). *In general metric spaces, the geodesic connecting the counterfactual Fréchet means is generally not equal to the Fréchet mean of the unit-level geodesics joining $Y_i(0)$ and $Y_i(1)$. Consequently, unlike the Euclidean setting, several distinct notions of treatment effects are possible. One possibility is the proposed GLATE, which is defined through the counterfactual Fréchet means and the geodesic connecting them. Another possibility would be to define treatment effects at the unit level through the geodesics joining $Y_i(0)$ and $Y_i(1)$ and then aggregate these geodesics across units. Such an approach would require additional structure to define an average of geodesics, for example by specifying a metric on the space of geodesics, and different choices may lead to different population-level estimands. The proposed GLATE therefore represents one particular notion of a population-level causal effect induced by the metric structure of the outcome space. It should be interpreted as a causal summary defined through counterfactual Fréchet means rather than as an aggregation of individual treatment-effect geodesics. While these two notions coincide in Euclidean spaces, they generally differ for metric space-valued outcomes, reflecting the richer geometric structure of non-Euclidean spaces.*

When the outcomes are univariate probability distributions equipped with the 2-Wasserstein metric (see Example 5), the GLATE admits a natural interpretation via quantile functions. In this setting, geodesics correspond to linear interpolations between quantile functions, and the causal estimand introduced by Van Dijke (2025),

$$\tau^{\text{R3D}} := F_{E_{\oplus}[Y(1)|R=c]}^{-1} - F_{E_{\oplus}[Y(0)|R=c]}^{-1},$$

can be viewed as the secant of the Wasserstein geodesic connecting the two distributions.

This difference is well-defined due to the pseudo-linear structure of the univariate Wasserstein space and offers a convenient representation in quantile coordinates.

The geometric structure of our approach allows for broader generality. While conventional RDD methods may be adapted to distributional outcomes by summarizing them through Euclidean functionals (e.g., means or quantiles), such simplifications are often infeasible for more complex outcomes such as networks, compositional data, or manifold-valued objects. In contrast, the GLATE is intrinsically defined within the metric space, obviating the need for Euclidean (scalar or vector) representations. The following result establishes the identification of the GLATE.

Theorem 3.1. *Under Assumption 3.1(i)–(ii), the causal estimand τ_{GRDD} admits the representation*

$$\tau_{\text{GRDD}} = \gamma_{\beta_0, \beta_1}, \tag{1}$$

where $\beta_0 = \lim_{r \uparrow c} \mathbb{E}_{\oplus}[Y|R = r]$ and $\beta_1 = \lim_{r \downarrow c} \mathbb{E}_{\oplus}[Y|R = r]$.

This theorem says that the causal estimand, GLATE τ_{GRDD} in Definition 3.1, is identified by the statistical estimand, i.e. the geodesic $\gamma_{\beta_0, \beta_1}$ from β_0 to β_1 . We note that this identification result also applies to multivariate distributions endowed with the Fisher–Rao metric, whereas the extension of quantile-based RDD methods such as Van Dijke (2025) to the case of multivariate distributions remains a challenging open problem.

3.2 Estimation

To estimate the GLATE τ_{GRDD} , we employ the local Fréchet regression (LFR) framework (Petersen and Müller, 2019; Chen and Müller, 2022), which extends local linear regression (Fan and Gijbels, 1996) to accommodate responses that lie in general metric spaces. Here we impose a mild condition on the kernel function used to define the estimator.

Assumption 3.2. *The kernel function $k(\cdot)$ is supported on $[-1, 1]$, zero outside its support, bounded, nonnegative, and positive and continuous on $(-1, 1)$.*

This assumption, which is satisfied by standard choices such as the triangular kernel $k(x) = (1+x)1\{x \in [-1, 0]\} + (1-x)1\{x \in [0, 1]\}$ or uniform kernel $k(x) = 1\{x \in [-1, 1]\}$, mirrors commonly assumed conditions in the scalar RDD literature (e.g., Assumption 2 in [Calonico et al. \(2014\)](#)).

In the sharp regression discontinuity setup, we construct two LFRs, one on each side of the cutoff c , to estimate the conditional Fréchet means of Y given $R = c^-$ and $R = c^+$. Define $K_0(x) = k(x)1\{x < 0\}$ and $K_1(x) = k(x)1\{x \geq 0\}$ for use on the left and right-hand side of the cutoff point, respectively. By $t \in \{0, 1\}$ we indicate the side of the cutoff and consider rescaled kernels $K_{t,h_t}(x) = h_t^{-1}K_t(x/h_t)$ utilizing bandwidths $h_t > 0$ such that $h_t \rightarrow 0$ as $n \rightarrow \infty$. The LFR target on side t is defined as the minimizer of the population weighted Fréchet loss

$$\nu_{t,\oplus} = \arg \min_{\nu \in \mathcal{M}} M_{t,n}(\nu), \quad \text{where } M_{t,n}(\nu) = \mathbb{E}[s_t(c; R, h_t)d^2(\nu, Y)].$$

Here the weight function $s_t(c; R, h_t)$ is given by

$$s_t(c; R, h_t) = \frac{1}{\sigma_t^2} K_{t,h_t}(R - c) \{\mu_{t,2} - \mu_{t,1}(R - c)\},$$

with moment terms $\mu_{t,k} = \mathbb{E}[K_{t,h_t}(R - c)(R - c)^k]$ for $k = 0, 1, 2$ and normalization constant $\sigma_t^2 = \mu_{t,0}\mu_{t,2} - \mu_{t,1}^2$. This construction mirrors the equivalent weighted formulation of local linear regression in Euclidean settings and ensures that the weight function integrates to one. The construction and intuition behind LFR, including its connection to local linear regression in Euclidean settings, are further discussed in Section S.1 of the Supplement.

We define the intermediate target $\tau_{\text{GRDD}}^* = \gamma_{\nu_{0,\oplus}, \nu_{1,\oplus}}$, which approximates the GLATE based on smoothed local Fréchet estimates. The empirical GRDD estimator is defined as $\hat{\tau}_{\text{GRDD}} = \gamma_{\hat{\nu}_{0,\oplus}, \hat{\nu}_{1,\oplus}}$, where for $t \in \{0, 1\}$,

$$\hat{\nu}_{t,\oplus} = \arg \min_{\nu \in \mathcal{M}} \hat{M}_{t,n}(\nu), \quad \text{with } \hat{M}_{t,n}(\nu) = \frac{1}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h_t)d^2(\nu, Y_i).$$

Here, $n_t = \sum_{i=1}^n 1\{T_i = t\}$ is the number of observations on side t , and the empirical weights are defined as

$$\hat{s}_t(c; R, h_t) = \frac{1}{\hat{\sigma}_t^2} K_{t,h_t}(R - c) \{\hat{\mu}_{t,2} - \hat{\mu}_{t,1}(R - c)\},$$

where $\hat{\mu}_{t,k} = n_t^{-1} \sum_{i=1}^n K_{t,h_t}(R_i - c)(R_i - c)^k$, and $\hat{\sigma}_t^2 = \hat{\mu}_{t,0}\hat{\mu}_{t,2} - \hat{\mu}_{t,1}^2$.

In the special case where $\mathcal{M} = \mathbb{R}$ and d is the Euclidean distance, the LFR estimator $\hat{\nu}_{t,\oplus}$ corresponds to the intercept of the local linear regression fit, recovering the standard RDD estimator. For general metric space-valued outcomes, this construction provides a natural generalization of classical sharp RDD estimation, enabling causal inference even when subtraction is undefined and vector space structure is absent.

For $t \in \{0, 1\}$, define the population target

$$\mu_{t,\oplus} := E_{\oplus}[Y(t)|R = c] = \arg \min_{\nu \in \mathcal{M}} M_{t,\oplus}(\nu, c),$$

where $M_{t,\oplus}(\nu, c) = E[d^2(\nu, Y(t))|R = c]$. To establish the convergence rate of the GRDD estimator, we impose the following assumptions.

Assumption 3.3.

(i) For each n , $\nu_{0,\oplus}$, $\nu_{1,\oplus}$, $\hat{\nu}_{0,\oplus}$, and $\hat{\nu}_{1,\oplus}$ exist and are unique, the latter two almost surely.

Additionally, for $t \in \{0, 1\}$ and every $\varepsilon > 0$,

$$\inf_{d(\nu, \mu_{t,\oplus}) > \varepsilon} \{M_{t,\oplus}(\nu, c) - M_{t,\oplus}(\mu_{t,\oplus}, c)\} > 0,$$

$$\liminf_{n \rightarrow \infty} \inf_{d(\nu, \nu_{t,\oplus}) > \varepsilon} \{M_{t,n}(\nu) - M_{t,n}(\nu_{t,\oplus})\} > 0.$$

(ii) For $t \in \{0, 1\}$, let $B_\delta(\mu_{t,\oplus}) \subset \mathcal{M}$ be the ball of radius δ centered at $\mu_{t,\oplus}$ and $N(\varepsilon, B_\delta(\mu_{t,\oplus}), d)$ be its covering number using balls of size $\varepsilon > 0$. Then

$$\int_0^1 \sqrt{1 + \log N(\delta\varepsilon, B_\delta(\mu_{t,\oplus}), d)} d\varepsilon = O(1) \text{ as } \delta \rightarrow 0.$$

(iii) *There exists a neighborhood $\mathcal{I}$ of c such that the density f_R of R is twice continuously differentiable on $\mathcal{I}$ with $f_R(c) > 0$. Moreover, for each $t \in \{0, 1\}$, the reverse conditional densities $g_{y,t}$ of $R|Y(t) = y$ exist and are twice continuously differentiable for all $y \in \mathcal{M}$ and all $r \in \mathcal{I}$, with $\sup_{r \in \mathcal{I}, y \in \mathcal{M}} (|g''_{y,0}(r)| + |g''_{y,1}(r)|) < \infty$. Additionally, for any open set $\mathcal{U} \subset \mathcal{M}$ and $t \in \{0, 1\}$, the map $r \mapsto \int_{\mathcal{U}} dF_{Y(t)|R=r}(r, y)$ is continuous for $r \in \mathcal{I}$.*

(iv) *For $t \in \{0, 1\}$ and $\nu \in \mathcal{M}$, there exist $\bar{\eta} > 0$, $\bar{C} > 0$, and $\beta_1 > 1$ such that*

$$M_{t,\oplus}(\nu, c) - M_{t,\oplus}(\mu_{t,\oplus}, c) \geq \bar{C}d(\nu, \mu_{t,\oplus})^{\beta_1},$$

provided $d(\nu, \mu_{t,\oplus}) < \bar{\eta}$.

(v) *For $t \in \{0, 1\}$ and $\nu \in \mathcal{M}$, there exist $\tilde{\eta} > 0$, $\tilde{C} > 0$, and $\beta_2 > 1$ such that*

$$\liminf_{n \rightarrow \infty} \{M_{t,n}(\nu) - M_{t,n}(\nu_{t,\oplus}) - \tilde{C}d(\nu, \nu_{t,\oplus})^{\beta_2}\} \geq 0,$$

provided $d(\nu, \nu_{t,\oplus}) < \tilde{\eta}$.

Assumption 3.3 is standard in the LFR literature; see, for example, [Petersen and Müller \(2019\)](#); [Zhou and Müller \(2022\)](#), where it is verified for Examples 1–5. Condition (i) is a standard separation condition that ensures consistency of the M-estimator (see, e.g., [van der Vaart and Wellner \(2023, Chapter 3.2\)](#)). Condition (ii) imposes a local entropy bound on metric balls centered at the population Fréchet means and is used to control the associated empirical processes. This condition is not required for functional outcomes in $L^2(\mathcal{T})$ or for univariate probability distributions equipped with the 2-Wasserstein metric, as in Examples 1 and 5. In these settings, the outcomes admit natural Hilbert space representations through the identity map for functional data and through quantile functions for Wasserstein data. The required stochastic bounds can therefore be derived using Hilbert space methods and, when necessary, projection onto the corresponding closed convex image space; see [Petersen and Müller \(2019\)](#).

Condition (iii) specifies distributional smoothness requirements that ensure sufficient regularity of the population Fréchet objective in a neighborhood of the cutoff. The conditional distributions $F_{Y(t)|R}$ are well-defined because $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ is a standard Borel space, which guarantees the existence of regular conditional distributions. Conditions (iv) and (v) characterize the bias and stochastic error rates, respectively. Together, conditions (ii), (iv), and (v) use tools from empirical process theory to control deviations of $M_{t,n}(\cdot) - M_{t,\oplus}(\cdot, c)$ and $\hat{M}_{t,n}(\cdot) - M_{t,n}(\cdot)$ around their respective minimizers $\mu_{t,\oplus}$ and $\nu_{t,\oplus}$.

To quantify the contrast among the geodesics, τ_{GRDD} , τ_{GRDD}^* and $\hat{\tau}_{\text{GRDD}}$, we define a distance between two geodesics. Let $\mathcal{G}(\mathcal{M}) = \{\gamma_{\alpha,\beta} : \alpha, \beta \in \mathcal{M}\}$ be the space of geodesics on the uniquely geodesic space $(\mathcal{M}, d)$. We first introduce the notion of a geodesic transport map.

Assumption 3.4. *Let $(\mathcal{M}, d)$ be a uniquely geodesic space. For any two points $\alpha, \beta \in \mathcal{M}$, there exists a geodesic transport map $\Gamma_{\alpha,\beta} : \mathcal{M} \mapsto \mathcal{M}$ with the following property: $\Gamma_{\alpha,\beta}(\alpha) = \beta$ and for any $\omega \in \mathcal{M}$, there exists a unique point $\zeta \in \mathcal{M}$ such that $\Gamma_{\alpha,\beta}(\omega) = \zeta$.*

Specific constructions of geodesic transport maps for Examples 1–5 can be found in Section 3 of Kurisu et al. (2025). We define a binary relation $\sim$ on $\mathcal{G}(\mathcal{M})$ through the geodesic transport map, where $\gamma_{\alpha_1,\beta_1} \sim \gamma_{\alpha_2,\beta_2}$ if and only if $\Gamma_{\alpha_1,\beta_1}(\omega) = \Gamma_{\alpha_2,\beta_2}(\omega)$ for all $\omega \in \mathcal{M}$. This relation is an equivalence relation on $\mathcal{G}(\mathcal{M})$ (Proposition 4.1, Zhou et al. (2025)), and we denote the equivalence class of a geodesic $\gamma_{\alpha,\beta}$ by $[\gamma_{\alpha,\beta}]$. The set of all equivalence classes $\mathcal{G}(\mathcal{M})/\sim := \{[\gamma_{\alpha,\beta}] : \gamma_{\alpha,\beta} \in \mathcal{G}(\mathcal{M})\}$ is the quotient space of $\mathcal{G}(\mathcal{M})$, for which we define a metric as follows. For any two equivalence classes $[\gamma_{\alpha_1,\beta_1}], [\gamma_{\alpha_2,\beta_2}] \in \mathcal{G}(\mathcal{M})/\sim$,

$$d_{\mathcal{G}}([\gamma_{\alpha_1,\beta_1}], [\gamma_{\alpha_2,\beta_2}]) = d_{\mathcal{G},\omega_{\oplus}}([\gamma_{\alpha_1,\beta_1}], [\gamma_{\alpha_2,\beta_2}]) = d(\Gamma_{\alpha_1,\beta_1}(\omega_{\oplus}), \Gamma_{\alpha_2,\beta_2}(\omega_{\oplus})), \quad (2)$$

where $\omega_{\oplus} \in \mathcal{M}$ is a fixed reference point, such as the Fréchet mean of the outcome. Intuitively, $d_{\mathcal{G}}$ measures the distance between two geodesics by comparing where they map

a common reference point $\omega_{\oplus}$ under their associated transport maps. In this sense, the distance reflects how differently the two geodesics act on the space, rather than comparing their endpoints directly. Note that $(\mathcal{G}(\mathcal{M})/\sim, d_{\mathcal{G}})$ is a metric space (Proposition 4.2 in Zhou et al. (2025)). Additionally, we require the following regularity assumption.

Assumption 3.5. *There exists a constant $C > 0$ such that for any $\alpha_1, \alpha_2, \beta_1, \beta_2, \omega \in \mathcal{M}$,*

$$d(\Gamma_{\alpha_1, \beta_1}(\omega), \Gamma_{\alpha_2, \beta_2}(\omega)) \leq C\{d(\alpha_1, \alpha_2) + d(\beta_1, \beta_2)\}.$$

We refer to Section S.2 of the Supplement in Zhou et al. (2025) for a verification of Assumption 3.5 for Examples 1–5. Letting $h_{\max} = \max\{h_0, h_1\}$ and $h_{\min} = \min\{h_0, h_1\}$, the convergence rate of the GRDD estimator is obtained as follows.

Theorem 3.2. *Consider the setup of Section 3.1 and suppose that Assumptions 3.1, 3.2, 3.4, and 3.5 are satisfied.*

(i) *Under Assumptions 3.3 (i)–(iv) and if $h_{\max} \rightarrow 0$ as $n \rightarrow \infty$, then*

$$d_{\mathcal{G}}(\tau_{\text{GRDD}}, \tau_{\text{GRDD}}^*) = O(h_{\max}^{2/(\beta_1-1)}).$$

(ii) *Under Assumptions 3.3(i), (ii) and (v) and if $h_{\max} \rightarrow 0$ and $nh_{\min} \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$d_{\mathcal{G}}(\tau_{\text{GRDD}}^*, \hat{\tau}_{\text{GRDD}}) = O_p((nh_{\min})^{-1/(2(\beta_2-1))}).$$

By an argument similar to the proof of Propositions 1–3 in Petersen and Müller (2019), one can see that Assumption 3.3 holds with $\beta_1 = \beta_2 = 2$ for Examples 1–5 in Section 2. Therefore, Theorem 3.2 yields that if $h_0 = h_1 = h$, then

$$d_{\mathcal{G}}(\tau_{\text{GRDD}}, \hat{\tau}_{\text{GRDD}}) \leq d_{\mathcal{G}}(\tau_{\text{GRDD}}, \tau_{\text{GRDD}}^*) + d_{\mathcal{G}}(\tau_{\text{GRDD}}^*, \hat{\tau}_{\text{GRDD}}) = O(h^2) + O_p((nh)^{-1/2}).$$

The right-hand side is optimized with $h = n^{-1/5}$ and the resulting convergence rate is $O_p(n^{-2/5})$, which corresponds to the optimal rate for local linear regression with scalar responses (Stone, 1980).

4 Inference for the magnitude of the treatment effect

We develop here statistical inference for the magnitude of the treatment effect, defined as $d(\mu_{1,\oplus}, \mu_{0,\oplus})$, which corresponds to the length of the geodesic treatment effect τ_{GRDD} . While estimation of the endpoints $\mu_{t,\oplus}$ is carried out intrinsically via LFR, an intrinsic approach to inference in general metric spaces is challenging. Therefore, we adopt an extrinsic approach by embedding the metric space $(\mathcal{M}, d)$ into a Hilbert space. This makes it possible to exploit linearity to derive asymptotic distributions as a prerequisite to derive inference. We establish asymptotic properties of the LFR estimators and develop bootstrap-based methods for testing and inference on $d(\mu_{1,\oplus}, \mu_{0,\oplus})$.

4.1 Hilbert space embedding and LFR representation

We impose the following assumption.

Assumption 4.1.

(i) *There is a separable Hilbert space $\mathcal{H}$ equipped with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and a continuous injection $\Psi : \mathcal{M} \rightarrow \mathcal{H}$ such that $\Psi : \mathcal{M} \rightarrow \Psi(\mathcal{M})$ is an isometry, that is, for any $\alpha, \beta \in \mathcal{M}$, $d(\alpha, \beta) = \|\Psi(\alpha) - \Psi(\beta)\|_{\mathcal{H}}$ where $\|\cdot\|_{\mathcal{H}}$ is the norm induced by the inner product.*

(ii) *The image $\Psi(\mathcal{M})$ is a nonempty closed convex subset of $\mathcal{H}$.*

Under Assumption 4.1, one obtains a linear representation of the metric space that is crucial for inference. In particular, it implies that conditional Fréchet means admit the representation $E_{\oplus}[Y(t)|R = r] = \Psi^{-1}(E[\Psi(Y(t))|R = r])$, and that the conditional expectation $E[\Psi(Y(t))|R = r]$ remains in $\Psi(\mathcal{M})$. A useful sufficient condition for Assumption 4.1(i) is that d^2 is of negative type. To formalize this, recall that a semimetric space $(\mathcal{Z}, \rho)$ is a set equipped with a symmetric, nonnegative function ρ satisfying $\rho(z_1, z_2) = 0$ if and only if $z_1 = z_2$. The semimetric ρ is of negative type if, for any $n \geq 2$, points $z_1, \dots, z_n \in \mathcal{Z}$ and

coefficients $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ with $\sum_{i=1}^n \alpha_i = 0$, one has $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho(z_i, z_j) \leq 0$. If d^2 is of negative type, then an isometric embedding into a Hilbert space exists; see [Schoenberg \(1938\)](#) and Proposition 3 of [Sejdinovic et al. \(2013\)](#).

Assumption 4.1 is satisfied by many commonly encountered spaces, including functional data under the L^2 metric (Example 1), SPD matrices under the Frobenius, power, Log-Euclidean or Log-Cholesky metrics (Example 3), networks represented by graph Laplacians with the Frobenius metric (Example 4), and univariate probability distributions on a compact support equipped with the 2-Wasserstein metric (Example 5). Explicit constructions of the embedding Ψ are provided in Section S.7 of the Supplement. For compositional data (Example 2), we embed observations on the positive orthant of the sphere into the tangent space using the logarithmic map with base point given by the estimated Fréchet mean $\hat{\nu}_{0,\oplus}$. Although the logarithmic map is not globally isometric, it preserves distances from the base point ([Lee, 2018](#)), which is the relevant quantity for testing the null hypothesis of no treatment effect. This allows us to carry out the testing procedure described in Section 4.3 in this setting. A rigorous justification of this argument is provided in Section S.7.2 of the Supplement.

Under this embedding, we consider the regression model in $\mathcal{H}$ for an i.i.d. sample $\{Y_i, T_i, R_i\}_{i=1}^n$:

$$\Psi(Y_i(t)) = m_t^\Psi(R_i) + \varepsilon_{t,i}, \quad t \in \{0, 1\},$$

where $m_t^\Psi(r) : \mathbb{R} \rightarrow \Psi(\mathcal{M})$ is the regression function and the error satisfies $\mathbb{E}[\langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} | R_i] = 0$ for every $g \in \mathcal{H}$. Under Assumptions 4.1 and 4.2 below, $m_t^\Psi(r) = \mathbb{E}[\Psi(Y(t)) | R = r]$ and the embedded conditional mean $\mathbb{E}[\Psi(Y(t)) | R = r]$ coincides with the image of the conditional Fréchet mean, so that $m_t^\Psi(r) = \Psi(\mathbb{E}_\oplus[Y(t) | R = r])$. In particular, at the cutoff, $m_t^\Psi(c) = \Psi(\mu_{t,\oplus})$ or $\mu_{t,\oplus} = \Psi^{-1}(m_t^\Psi(c))$. We estimate $m_t^\Psi(c)$ by applying LFR in $\mathcal{H}$. Specifically,

$$\hat{\nu}_t^\Psi = \arg \min_{g \in \mathcal{H}} \frac{1}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h_t) \|g - \Psi(Y_i)\|_{\mathcal{H}}^2 = \frac{1}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h_t) \Psi(Y_i).$$

4.2 Asymptotic properties of the LFR estimators

Under the Hilbert space embedding, we study the asymptotic behavior of the LFR estimators in $\mathcal{H}$. Recall that $\mu_{t,\oplus} = \mathbb{E}_{\oplus}[Y(t)|R = c] = \Psi^{-1}(m_t^{\Psi}(c))$. We denote its image in the Hilbert space by $\mu_{t,\oplus}^{\Psi} = \Psi(\mu_{t,\oplus}) = m_t^{\Psi}(c)$. Therefore, inference on $\mu_{t,\oplus}$ can be carried out by studying $\mu_{t,\oplus}^{\Psi}$ in $\mathcal{H}$. Let $\{e_j\}_{j \geq 1}$ be an orthonormal basis of $\mathcal{H}$. We impose the following regularity conditions.

Assumption 4.2. *For each $t \in \{0, 1\}$, the following conditions hold.*

- (i) *The density function f_R of the running variable R is continuous and positive on a neighborhood $\mathcal{I}$ of c .*
- (ii) *For any $g \in \mathcal{H}$, the function $m_t^{\Psi}(r; g) = \langle m_t^{\Psi}(r), g \rangle_{\mathcal{H}}$ is twice continuously differentiable on $\mathcal{I}$. Additionally, $\sup_{r \in \mathcal{I}} \{\|m_t^{\Psi}(r)\|_{\mathcal{H}} + \sum_{s=1}^2 \sum_{j \geq 1} |\partial^s m_t^{\Psi}(r; e_j)|^2\} < \infty$ where $\partial^s m_t^{\Psi}(r; g) = \frac{d^s}{dr^s} \langle m_t^{\Psi}(r), g \rangle_{\mathcal{H}}$, $s = 1, 2$, $g \in \mathcal{H}$.*
- (iii) *For any $g \in \mathcal{H} \setminus \{0\}$, $\mathbb{E}[\langle \varepsilon_{t,1}, g \rangle_{\mathcal{H}}^2 | R_1 = r]$ is continuous on $\mathcal{I}$, $\inf_{r \in \mathcal{I}} \mathbb{E}[\langle \varepsilon_{t,1}, g \rangle_{\mathcal{H}}^2 | R_1 = r] > 0$, and $\sup_{r \in \mathcal{I}} \mathbb{E}[\|\varepsilon_{t,1}\|_{\mathcal{H}}^4 | R_1 = r] < \infty$. Additionally, there exists a sequence of positive constants $\{b_j\}_{j \geq 1}$ such that $\sup_{r \in \mathcal{I}} \mathbb{E}[\langle \varepsilon_{t,1}, e_j \rangle_{\mathcal{H}}^2 | R_1 = r] \leq b_j$, $\sum_{j \geq 1} b_j < \infty$.*

Assumption 4.2 parallels standard conditions for local linear regression with scalar responses (Fan and Gijbels, 1996), applied here to the Hilbert space through projections onto one-dimensional directions. Condition (ii) ensures sufficient smoothness of the regression function in $\mathcal{H}$, while condition (iii) controls the stochastic variability of the error process. We use $\rightsquigarrow$ to denote weak convergence. The following result establishes the joint asymptotic normality of the LFR estimators in the Hilbert space.

Theorem 4.1. *Suppose that Assumptions 3.1, 3.2, 4.1, and 4.2 hold. Assume that $h_t = h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^5 \rightarrow 0$ as $n \rightarrow \infty$. Then we have $\sqrt{nh}(\hat{\nu}_0^{\Psi} - \mu_{0,\oplus}^{\Psi}, \hat{\nu}_1^{\Psi} - \mu_{1,\oplus}^{\Psi}) \rightsquigarrow$*

$(\mathbb{G}_0, \mathbb{G}_1)$, where $\mathbb{G}_0$ and $\mathbb{G}_1$ are independent centered Gaussian elements in $\mathcal{H}$. Their covariance structure is characterized through projections:

$$\text{Var}(\langle \mathbb{G}_t, g \rangle_{\mathcal{H}}) = \frac{\int (\kappa_{t,2} - \kappa_{t,1}u)^2 K_t^2(u) du}{(\kappa_{t,2}\kappa_{t,0} - \kappa_{t,1}^2)^2} \frac{\mathbb{E}[\langle \varepsilon_{t,1}, g \rangle_{\mathcal{H}}^2 | R_1 = c]}{f_R(c)}, \quad t \in \{0, 1\}, \quad g \in \mathcal{H},$$

where $\kappa_{t,p} = \int u^p K_t(u) du$.

Theorem 4.1 shows that the LFR estimators behave analogously to local linear estimators with Euclidean responses, but now for responses taking values in a possibly infinite-dimensional Hilbert space. The convergence rate $\sqrt{nh}$ corresponds to the standard boundary rate in RDD, and the limiting Gaussian elements are independent across the two sides of the cutoff due to the use of one-sided kernels. This result provides the foundation for conducting statistical inference using the extrinsic approach described in Section 4.1.

4.3 Bootstrap-based inference for the treatment effect magnitude

Building on the asymptotic normality established in Theorem 4.1, we develop bootstrap-based procedures for statistical inference on the magnitude of the treatment effect $d(\mu_{1,\oplus}, \mu_{0,\oplus})$. Since the Gaussian limit in $\mathcal{H}$ involves unknown covariance structures, we employ a multiplier bootstrap to approximate its distribution; similar resampling strategies have been considered, for example, in Chiang et al. (2019) and Van Dijke (2025).

Let $\{\xi_i\}_{i=1}^n$ be i.i.d. standard normal random variables defined on a probability space $(\Omega_\xi, \mathcal{A}_\xi, \mathbb{P}_\xi)$, independent of the data $\{Y_i, T_i, R_i\}_{i=1}^n$. Define the bootstrap processes

$$\hat{\mathbb{G}}_t = \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \xi_i \hat{s}_t(c; R_i, h_t) \{\Psi(Y_i) - \hat{\nu}_t^\Psi\}, \quad t \in \{0, 1\}.$$

We use $\overset{\mathbb{P}_\xi}{\rightsquigarrow}$ to denote conditional weak convergence given the data; see Section 1.13 of van der Vaart and Wellner (2023).

Proposition 4.1. *Suppose that the assumptions of Theorem 4.1 hold. Then, conditionally on the data $\{Y_i, T_i, R_i\}_{i=1}^n$ and with probability approaching one, the process $(\hat{\mathbb{G}}_0, \hat{\mathbb{G}}_1)$ weakly converges to the limit process $(\mathbb{G}_0, \mathbb{G}_1)$, that is, $(\hat{\mathbb{G}}_0, \hat{\mathbb{G}}_1) \xrightarrow{P_\xi} (\mathbb{G}_0, \mathbb{G}_1)$.*

Proposition 4.1 shows that the multiplier bootstrap consistently approximates the joint Gaussian limit in $\mathcal{H}$, thereby enabling inference without explicit estimation of the covariance operator.

We first consider testing for the presence of a treatment effect at the cutoff, i.e., testing $\mathbb{H}_0 : d(\mu_{1,\oplus}, \mu_{0,\oplus}) = 0$ versus $\mathbb{H}_1 : d(\mu_{1,\oplus}, \mu_{0,\oplus}) > 0$. Define the test statistic $T_n = nh\|\hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi\|_{\mathcal{H}}^2$, and its bootstrap analogue $\hat{T}_n = \|\hat{\mathbb{G}}_1 - \hat{\mathbb{G}}_0\|_{\mathcal{H}}^2$. Let $q_{n,\alpha}$ denote the $(1-\alpha)$ -quantile of the conditional distribution of $\hat{T}_n$ given the data, and define the test function $\hat{\phi}_n = \mathbf{1}\{T_n > q_{n,\alpha}\}$. The following result establishes the validity of this bootstrap test.

Corollary 4.1. *Suppose that the assumptions of Theorem 4.1 hold. Then the test $\hat{\phi}_n$ is asymptotically level α , i.e., $E[\hat{\phi}_n] \rightarrow \alpha$ under $\mathbb{H}_0$ as $n \rightarrow \infty$. Moreover, $\hat{\phi}_n$ is consistent, i.e. $E[\hat{\phi}_n] \rightarrow 1$ under $\mathbb{H}_1$ as $n \rightarrow \infty$.*

When $d(\mu_{1,\oplus}, \mu_{0,\oplus}) > 0$, we construct confidence intervals for the magnitude of the treatment effect. Based on the linear expansion implied by Theorem 4.1, let $c_{n,\alpha}$ denote the $(1-\alpha)$ -quantile of the conditional distribution of $2|\langle \hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi, \hat{\mathbb{G}}_1 - \hat{\mathbb{G}}_0 \rangle_{\mathcal{H}}|$. A $(1-\alpha)$ confidence interval for $d(\mu_{1,\oplus}, \mu_{0,\oplus})$ is given by

$$I_{n,\alpha} = \left[\left(\max \left\{ \|\hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi\|_{\mathcal{H}}^2 - \frac{c_{n,\alpha}}{\sqrt{nh}}, 0 \right\} \right)^{1/2}, \left(\|\hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi\|_{\mathcal{H}}^2 + \frac{c_{n,\alpha}}{\sqrt{nh}} \right)^{1/2} \right].$$

Corollary 4.2. *Under the assumptions of Theorem 4.1, if $d(\mu_{1,\oplus}, \mu_{0,\oplus}) > 0$, Then*

$$\Pr(d(\mu_{1,\oplus}, \mu_{0,\oplus}) \in I_{n,\alpha}) \rightarrow 1 - \alpha \quad \text{as } n \rightarrow \infty.$$

Remark 4.1. *Theorem 4.1 and Proposition 4.1 also enable us to construct a confidence region of $\mu_{t,\oplus}$ for $t \in \{0, 1\}$. Let $q_{n,\alpha}^{(t)}$ denote the $(1-\alpha)$ -quantile of the conditional distribution*

of $\|\hat{G}_t\|_{\mathcal{H}}^2$ given the data. A $(1 - \alpha)$ confidence region for $\mu_{t,\oplus}$ is

$$R_{n,\alpha}^{(t)} = \left\{ \nu \in \mathcal{M} : \|\hat{\nu}_t^\Psi - \Psi(\nu)\|_{\mathcal{H}} \leq \sqrt{\frac{q_{n,\alpha}^{(t)}}{nh_t}} \right\}.$$

Indeed, under the assumptions of Theorem 4.1, one has $\Pr(\mu_{t,\oplus} \in R_{n,\alpha}^{(t)}) \rightarrow 1 - \alpha$ as $n \rightarrow \infty$.

5 Implementation and simulations

5.1 Bandwidth selection

Selecting an appropriate bandwidth is critical for reliable estimation in RDDs, especially when the outcome resides in a general metric space. For scalar outcomes, conventional RDD bandwidth selectors are typically based on asymptotic MSE expansions for local polynomial estimators at the cutoff (Imbens and Kalyanaraman, 2012; Calonico et al., 2014, 2020). Such plug-in rules rely on Euclidean structure and derivative-based approximations, and are therefore not directly available in our general non-Euclidean setting.

We adopt a boundary-focused one-sided cross-validation approach, motivated by Ludwig and Miller (2007), to select the bandwidth in the GRDD framework. For a fraction $\delta \in (0, 1)$, the cross-validation criterion is evaluated over a neighborhood of the cutoff, defined by the $(1 - \delta)$ -quantile of the running variable below c and the δ -quantile above c . At each evaluation point, LFR is fitted using only observations farther from the cutoff on the same side. The bandwidth is then selected by minimizing the resulting sum of squared metric prediction errors over a candidate grid; see Algorithm 1. This combination of one-sided fitting and cutoff-focused evaluation yields a practical bandwidth selection procedure adapted to the boundary structure of GRDD in general metric spaces.

The parameter δ controls the width of the evaluation region, with larger values including more observations. In the default implementation, δ is selected automatically so that the number of evaluation observations is close to a fixed computational target. Specifically, the

search is restricted to $\delta \in [0.05, 0.5]$, and the target number of evaluation observations is set to 100 throughout all simulations and empirical applications. These fixed defaults allow the evaluation region to adapt to the sample size while keeping the cross-validation procedure computationally manageable. Users may instead specify δ directly when substantive considerations favor a particular neighborhood around the cutoff.

While the cross-validation procedure provides a practical data-driven bandwidth selection rule, the asymptotic validity of the bootstrap procedure in Theorem 4.1 is established under the undersmoothing condition $nh^5 \rightarrow 0$. Our simulation studies suggest that the proposed bootstrap procedure nevertheless provides reliable empirical size control with cross-validation-based bandwidth selection. Developing bandwidth selection methods specifically tailored for inference in GRDD is an interesting direction for future research.

5.2 Simulation for network-valued outcomes

To evaluate both estimation and inference in a non-Euclidean setting, we consider outcomes given by network structures. Specifically, each observation is a graph Laplacian taking values in the space $\mathcal{L}$ equipped with the Frobenius metric (Example 4). This setting reflects applications where the outcome is a weighted network and the objective is to detect discontinuities in its geometry at the cutoff.

Networks are generated under a weighted stochastic block model with 10 nodes divided equally into two communities, C_1 and C_2 . The binary adjacency matrix is determined by a block-structured edge probability matrix $\begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$, where the entries denote the probabilities of edge formation within and between communities. The running variable R is generated from a truncated normal distribution on $[-1, 1]$ with mean 0 and standard

Algorithm 1 One-sided cross-validation bandwidth selection for GRDD

Input: Sample $\{(Y_i, R_i)\}_{i=1}^n$, cutoff c , fraction $\delta \in (0, 1)$

Output: Selected bandwidth b^*

- 1: Sort the running variables $\{R_i\}_{i=1}^n$ in increasing order and let $R_{\min} = \min_i R_i$ and $R_{\max} = \max_i R_i$
 - 2: Let $q_{-, 1-\delta}$ be the $(1 - \delta)$ -quantile of $\{R_i : R_i < c\}$ and $q_{+, \delta}$ be the δ -quantile of $\{R_i : R_i \geq c\}$
 - 3: Define the evaluation region $\mathcal{R} = [q_{-, 1-\delta}, q_{+, \delta}]$
 - 4: Compute $b_{\max}$ as half the minimum of $c - R_{\min}$ and $R_{\max} - c$ and $b_{\min}$ as a data-adaptive lower bound based on the local spacing of the running variable near the cutoff
 - 5: Construct a grid of bandwidths $\mathcal{B} \subset [b_{\min}, b_{\max}]$
 - 6: **for** each $b \in \mathcal{B}$ **do**
 - 7: Set $\text{CV}(b) = 0$
 - 8: **for** each $R_i \in \mathcal{R}$ **do**
 - 9: **if** $R_i < c$ **then**
 - 10: Compute the one-sided LFR estimate $\hat{m}_b(R_i)$ at R_i using only observations with $R_j < R_i$
 - 11: **else**
 - 12: Compute the one-sided LFR estimate $\hat{m}_b(R_i)$ at R_i using only observations with $R_j > R_i$
 - 13: **end if**
 - 14: Update
$$\text{CV}(b) = \text{CV}(b) + d^2(Y_i, \hat{m}_b(R_i))$$
 - 15: **end for**
 - 16: **end for**
 - 17: $b^* = \arg \min_{b \in \mathcal{B}} \text{CV}(b)$
-

deviation 0.1. Given R , each present edge is assigned weight

$$w_{ij}(R) = \begin{cases} 1 + \cos\left(\frac{\pi}{2}R\right) + \epsilon_{ij}, & R < 0, \\ 1 + \cos\left(\frac{\pi}{2}R\right) + \Delta\tau_{ij} + \epsilon_{ij}, & R \geq 0, \end{cases}$$

where $\epsilon_{ij} \sim \text{Unif}(0, 0.1)$ independently. The term τ_{ij} induces heterogeneous changes across edge types: $\tau_{ij} = 1$ for edges within C_1 , $\tau_{ij} = -1$ for edges within C_2 , and $\tau_{ij} = 0$ for edges between communities. The parameter $\Delta \geq 0$ controls the magnitude of the discontinuity at the cutoff at the edge level, where $\Delta = 0$ corresponds to no treatment effect. Each weighted adjacency matrix is converted into a graph Laplacian, which serves as the outcome Y_i .

We first assess estimation performance by fixing $\Delta = 1$ and considering sample sizes $n \in \{100, 200, 500, 1000\}$. For each n , we generate 500 independent replications $\{Y_i, R_i\}_{i=1}^n$

and apply GRDD with the data-adaptive bandwidth selector described in Section 5.1. In each replication, performance is evaluated using the relative bias $d_{\mathcal{G}}(\hat{\tau}_{\text{GRDD}}, \tau_{\text{GRDD}})/|\tau_{\text{GRDD}}|$, where $d_{\mathcal{G}}$ denotes the metric on the space of geodesics defined in (2) and $|\tau_{\text{GRDD}}|$ is the magnitude of the true effect. As this metric aggregates discrepancies across all edge weights in the network, it reflects a global measure of estimation error.

To evaluate inference, we use the same data-generating mechanism but vary Δ over the grid $\{0, 0.1, \dots, 1\}$. For each pair (n, Δ) , we again generate 500 independent replications and apply the bootstrap-based inference procedure developed in Section 4.3. Specifically, we test $\mathbb{H}_0 : d(\mu_{1,\oplus}, \mu_{0,\oplus}) = 0$ versus $\mathbb{H}_1 : d(\mu_{1,\oplus}, \mu_{0,\oplus}) > 0$, using the multiplier bootstrap with $B = 1000$ replications and nominal level $\alpha = 0.05$. The empirical rejection frequency over replications estimates the size of the test when $\Delta = 0$ and its power when $\Delta > 0$.

Figure 1 summarizes the results. Panel (a) displays boxplots of the relative bias for $\Delta = 1$. The bias decreases steadily with increasing sample size, indicating consistency of the GRDD estimator for network-valued outcomes. Panel (b) displays the empirical rejection probabilities as a function of Δ for each sample size, with the dashed horizontal line indicating the nominal level $\alpha = 0.05$. At $\Delta = 0$, the rejection rates across all sample sizes are close to the nominal level, suggesting that the bootstrap procedure provides reliable empirical size control in the simulation settings considered. As Δ increases, the rejection probability rises rapidly, with larger sample sizes exhibiting substantially higher power. In particular, for moderate values of Δ (e.g., $\Delta = 0.2$ or 0.3), the power improves markedly as n increases, and for $\Delta \geq 0.5$, the test attains near-perfect power even at moderate sample sizes. These results suggest that the proposed inference procedure yields empirical rejection probabilities close to the nominal level under the null while remaining sensitive to increasingly strong treatment effects.

An additional null simulation with unequal curvature of the regression functions on the two sides of the cutoff is provided in Section S.13 of the Supplement. This setting further

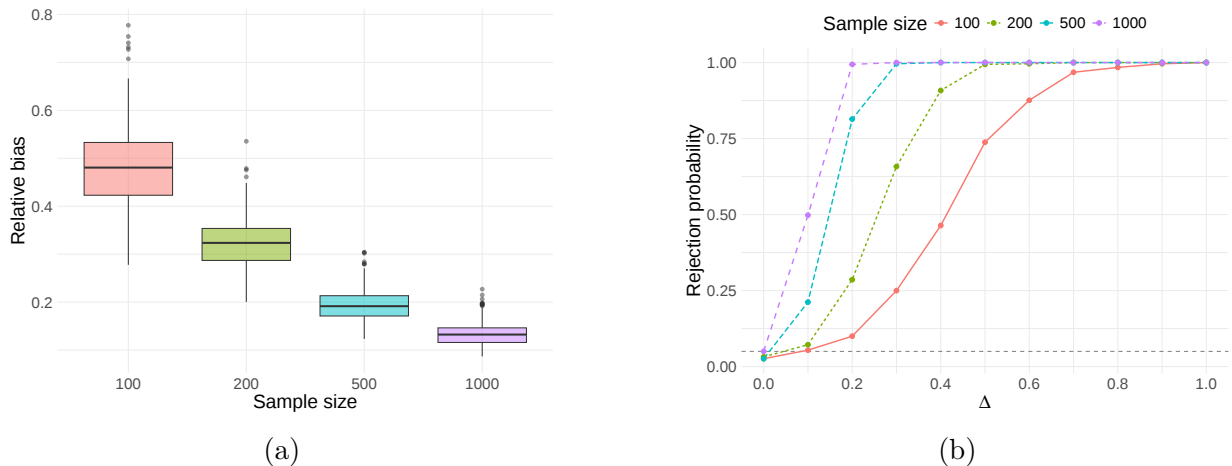


Figure 1: Simulation results for network-valued outcomes. (a) Boxplots of the relative bias across 500 replications for each sample size $n \in \{100, 200, 500, 1000\}$. (b) Empirical rejection probabilities of the bootstrap test for $\mathbb{H}_0 : d(\mu_{1,\oplus}, \mu_{0,\oplus}) = 0$ at level $\alpha = 0.05$.

Alt text: The left panel shows decreasing estimation error as the sample size grows. The right panel shows rejection probability increasing with the discontinuity magnitude, with higher power for larger samples.

examines empirical size under cross-validation-based bandwidth selection in a more general configuration.

6 Incumbency effects in multiparty elections

We use the proposed GRDD method to investigate the impact of incumbency in a multiparty electoral system, focusing on data from UK parliamentary elections compiled by [Eggers and Spirling \(2017\)](#). The dataset covers all general elections in the United Kingdom from 1950 to 2010, including constituency-level results for each of the three major parties: Conservative, Labour, and Liberal (or its historical equivalents, such as the Liberal Democrats or SDP-Liberal Alliance). The full sample contains 8,170 constituency-level contests, with information on vote totals, margins of victory, and party affiliations for every candidate in each election. The UK employs a single-member district plurality system, in which candidates compete in geographically defined constituencies and the winner is determined by a simple plurality rule. This institutional setting provides a compelling en-

vironment for studying incumbency effects, as each contest produces one winner and allows clear identification of party continuity or change.

For each constituency, we define the running variable as the Conservative margin in the previous election, i.e., the difference in vote share between the Conservative candidate and the top rival. The outcome is the compositional vector of vote shares in the subsequent election, restricted to the three major parties. Following Example 2, these compositions are mapped to the positive orthant of the unit sphere via the square-root transformation and analyzed using the arc-length distance. In this representation, the geodesic between two compositions corresponds to great-circle interpolation on the unit sphere. As a benchmark, we also estimate separate scalar RDDs for each vote-share component using the `rdrobust` package (Calonico et al., 2017).

The proposed data-adaptive procedure selects $\delta = 0.05$ for the boundary evaluation region and yields a bandwidth of 0.245. Using this bandwidth, the estimated vote-share composition at the cutoff on the left side (where the Conservative candidate narrowly lost) is (43.5%, 37.7%, 18.8%) for the Conservative, Labour, and Liberal parties, respectively. On the right side (where the Conservative candidate narrowly won), the estimated composition shifts to (44.6%, 37.6%, 17.8%). Thus, the estimated GRDD discontinuity corresponds to a gain of about 1.1 percentage points for the Conservatives, essentially no change for Labour, and a decline of about 1.0 percentage point for the Liberals. This pattern is consistent with a modest incumbency advantage for the Conservatives, operating primarily through a reallocation of support away from the Liberal party.

The componentwise scalar RDD analysis yields a broadly similar pattern. The bias-corrected `rdrobust` estimates (Calonico et al., 2020) indicate a discontinuity of about 1.6 percentage points for the Conservatives, essentially no change for Labour, and a decline of about 1.5 percentage points for the Liberals. These results are consistent with the GRDD estimates, suggesting that the primary shifts occur in the Conservative and Liberal

shares, with little change for Labour. However, the scalar approach treats the components independently, does not provide a unified assessment of the compositional effect, and fails to preserve the sum-to-zero constraint, which arises because the treatment effect is defined as the difference between two compositions.

For statistical inference, GRDD analyzes the outcome jointly as a composition and provides a global inferential statement. The estimated effect magnitude is 0.0147. Using the bootstrap-based inference procedure described in Section 4.3, we obtain a one-sided p -value of 0.190. Thus, while the estimated shift is directionally consistent with an incumbency effect, we do not reject the null of no overall compositional treatment effect at the 5% level. In contrast, the scalar approach conducts inference separately for each component. The componentwise `rdrobust` analysis yields p -values of 0.001 for the Conservatives, 0.928 for Labour, and 0.081 for the Liberals, indicating statistical significance only for the Conservative component at the 5% level. This illustrates that, while componentwise analyses may suggest localized changes, the joint analysis provides a different perspective on whether there is a statistically meaningful shift in the composition as a whole.

Figure 2 illustrates the estimated results. Each panel corresponds to one of the three major parties. The horizontal axis displays the Conservative margin in the previous election, and the vertical axis shows the party’s current vote share. Points represent bin-averaged outcomes (in 0.1 percentage point bins), while the smooth curves show the GRDD estimates on either side of the cutoff. The estimated discontinuity at the cutoff is modest and is most apparent in the Conservative and Liberal panels.

An additional application to functional air pollution data is provided in Section S.2 of the Supplement.

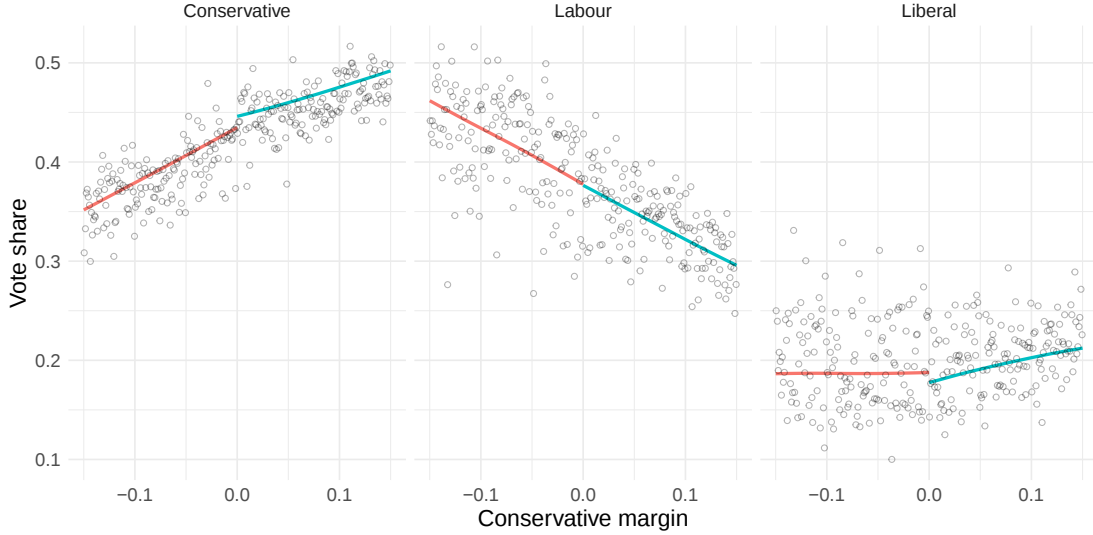


Figure 2: Estimated GRDD fits and bin-averaged outcomes by Conservative margin using UK parliamentary election data from 1950 to 2010.

Alt text: Three panels display Conservative, Labour, and Liberal vote shares against the previous Conservative margin. The fitted curves show a small upward discontinuity for Conservatives, almost no discontinuity for Labour, and a small downward discontinuity for Liberals.

7 Fuzzy regression discontinuity design for random objects

In the sharp RDD considered so far, the treatment is assigned based on whether a threshold is met, assuming full compliance where everyone adheres to the assigned treatment. However, in reality, some individuals might not follow the assigned treatment. The fuzzy RDD addresses such situations of imperfect compliance, which are practically relevant. This motivates extending the GRDD framework to the fuzzy design, where the average treatment effect at the cutoff is obtained by amplifying the discontinuity in the outcome by the reciprocal discontinuity in the treatment probability.

We first develop this extension for outcome spaces that admit an isometric embedding into a Hilbert space, leading to an extrinsic definition of the causal effect with corresponding identification and estimation procedures. An intrinsic formulation, where the causal effect is defined as the geodesic between the relevant Fréchet means in the outcome space, is pre-

sented in Section S.3 of the Supplement. The special case of Riemannian manifold-valued outcomes is discussed in Section S.4, where logarithmic and exponential maps are used to represent the causal effect in the tangent space. An intrinsic, geodesic-based formulation for manifold-valued outcomes is provided in Section S.5 of the Supplement.

7.1 Setup

Since the fuzzy RDD case is quite distinct from the sharp RDD case studied in Section 3, we first reintroduce notation. Suppose we observe an independent and identically distributed sample $\{Y_i, T_i, R_i\}_{i=1}^n$ of $(Y, T, R) \in \mathcal{M} \times \{0, 1\} \times \mathbb{R}$, where $Y \in \mathcal{M}$ denotes the outcome, T is the binary treatment indicator, and R is a running variable that yields the assignment variable $Z = 1\{R \geq c\}$ with a known constant c . Then the observed treatment is determined by $T = ZT(1) + (1 - Z)T(0)$, where $T(z)$ is the indicator of the potential treatment status for $z \in \{0, 1\}$. The outcome variable satisfies $Y = Y(0)$ if $T = 0$ and $Y = Y(1)$ if $T = 1$, where $Y(t) \in \mathcal{M}$ is a potential outcome for treatment $T = t$. Here we assume that there is no direct effect of Z on the potential outcomes (exclusion restriction).

In fuzzy RDDs with scalar outcomes, a causal effect of interest is the local average treatment effect (LATE) of the compliers, given by $E[Y(T(1))|T(1) > T(0), R = c] - E[Y(T(0))|T(1) > T(0), R = c]$. In the case of scalar outcomes, compliers' LATE is identified by the difference of the expected outcomes at the cutoff divided by the difference of the treatment probabilities at the cutoff (Hahn et al., 2001). However, when outcomes take values in a general metric space, not only differencing but also division operators may not be applicable. Thus, the fuzzy RDD analysis for random objects is fundamentally different from the sharp RDD analysis considered in the previous sections.

Denote by $\mu_{z,\oplus} = E_{\oplus}[Y(T(z))|T(1) > T(0), R = c]$ for $z \in \{0, 1\}$ the unique Fréchet mean of the potential outcome $Y(T(z))$ of compliers at the cutoff. To extend the notion of compliers' causal effect to outcomes in geodesic metric spaces, we introduce a mapping

Ψ of $\mu_{z,\oplus}$ to a Hilbert space, where the linear structure makes it possible to evaluate a contrast for $\Psi(\mu_{z,\oplus})$.

7.2 Identification

To define the causal estimand in the fuzzy RDD setting, we build on the Hilbert space embedding framework introduced in Section 4.1. Specifically, we assume that the metric space $(\mathcal{M}, d)$ admits an isometric embedding into a separable Hilbert space $\mathcal{H}$ via a map $\Psi : \mathcal{M} \rightarrow \mathcal{H}$ satisfying Assumption 4.1.

In particular, Ψ is injective and distance-preserving, and its image $\Psi(\mathcal{M})$ is a closed convex subset of $\mathcal{H}$. This structure allows us to represent Fréchet means and causal effects in terms of linear operations in $\mathcal{H}$. Based on this representation, we define the causal estimand as follows.

Definition 7.1 (Compliers' LATE at the cutoff). *Suppose that $(\mathcal{M}, d)$ satisfies Assumption 4.1. Then the compliers' local average treatment effect at the cutoff is defined as*

$$\tau_{\text{FRDD}} := \Psi(\mu_{1,\oplus}) - \Psi(\mu_{0,\oplus}).$$

When $\mathcal{M} = \mathbb{R}$, one can take Ψ as the identity map, and the above definition reduces to the classical complier LATE:

$$\tau_{\text{FRDD}} = \mu_{1,\oplus} - \mu_{0,\oplus} = \mathbb{E}[Y(T(1)) | T(1) > T(0), R = c] - \mathbb{E}[Y(T(0)) | T(1) > T(0), R = c].$$

The above construction requires the existence of an isometric embedding Ψ , which excludes certain spaces such as compositional data equipped with the arc-length metric (Example 2), due to their positive curvature. For such spaces, alternative intrinsic formulations based on tangent space representations are provided in Section S.5 of the Supplement.

For the identification of τ_{FRDD} , we assume the following conditions.

Assumption 7.1.

- (i) $\Pr(T(1) \geq T(0)|R = r) = 1$ for all r in a neighborhood of c .
- (ii) $\Pr(T(1) > T(0)|R = c) \in (\eta_0, 1 - \eta_0)$ for some $\eta_0 \in (0, 1/2)$.
- (iii) $E[T(1)|R = r]$ and $E[T(0)|R = r]$ are continuous over a neighborhood of c .
- (iv) $E_{\oplus}[Y(T(0))|R = r]$ and $E_{\oplus}[Y(T(1))|R = r]$ exist uniquely for each r over a neighborhood of c , and they are continuous on that neighborhood.

Condition (i) is a monotonicity condition on the potential treatment variable, which rules out defiers having $(T(0), T(1)) = (1, 0)$. Condition (ii) is a non-overlapping condition for compliers and other types (i.e., always/never-takers). Conditions (iii) and (iv) are continuity requirements for the conditional Fréchet means. Define $\mu_{z,\oplus}^{\dagger} = E_{\oplus}[Y(T(z))|R = c]$ for $z \in \{0, 1\}$ and assume existence and uniqueness. The next result provides identification of τ_{FRDD} .

Theorem 7.1. *Consider the setup of Section 7.1. Under Assumption 7.1,*

$$\tau_{\text{FRDD}} = \frac{\Psi(\lim_{r \downarrow c} E_{\oplus}[Y|R = r]) - \Psi(\lim_{r \uparrow c} E_{\oplus}[Y|R = r])}{\lim_{r \downarrow c} E[T|R = r] - \lim_{r \uparrow c} E[T|R = r]}.$$

Note that for the space of univariate distributions with the 2-Wasserstein metric (Example 5), we can take $\Psi(\alpha) = F_{\alpha}^{-1}(\cdot)$ where $F_{\alpha}^{-1}(\cdot)$ is the quantile function of $\alpha \in \mathcal{W}$. In this case, τ_{FRDD} corresponds to the fuzzy local average quantile treatment effect $F_{\mu_{1,\oplus}}^{-1}(\cdot) - F_{\mu_{0,\oplus}}^{-1}(\cdot)$ that was considered in Van Dijke (2025). Theorem 7.1 gives

$$\tau_{\text{FRDD}} = \frac{F_{\mu_{1,\oplus}^{\dagger}}^{-1}(\cdot) - F_{\mu_{0,\oplus}^{\dagger}}^{-1}(\cdot)}{E[T(1)|R = c] - E[T(0)|R = c]},$$

which corresponds to Lemma 2 in Van Dijke (2025). Therefore, our framework for fuzzy RDD with non-Euclidean outcomes can be seen as a generalization of the fuzzy design in Van Dijke (2025) to general geodesic metric spaces.

7.3 Estimation

We now construct an estimator of τ_{FRDD} and provide its asymptotic properties. A natural estimator is

$$\hat{\tau}_{\text{FRDD}} = \frac{\Psi(\hat{\nu}_{1,\oplus}) - \Psi(\hat{\nu}_{0,\oplus})}{\hat{m}_1 - \hat{m}_0}, \quad (3)$$

where $\hat{\nu}_{0,\oplus}$ and $\hat{\nu}_{1,\oplus}$ are the LFR estimators defined in Section 3.2, and $\hat{m}_0$ and $\hat{m}_1$ are (nonparametric) estimators for $m_0 = E[T(0)|R = c]$ and $m_1 = E[T(1)|R = c]$, respectively. For example, one can use local linear regression for the construction of $\hat{m}_z = \hat{m}_{z,0}$, leading to

$$(\hat{m}_{z,0}, \hat{m}_{z,1})^\top = \arg \min_{\beta_0, \beta_1 \in \mathbb{R}} \frac{1}{n_z} \sum_{i=1}^n (T_i - \beta_0 - \beta_1(R_i - c))^2 K_{z,h_z}(R_i - c) \quad \text{for } z \in \{0, 1\}. \quad (4)$$

Define $M_{z,\oplus}^\dagger(\nu, c) = E[d^2(\nu, Y(T(z)))|R = c]$ for $z \in \{0, 1\}$. Letting $h_{\max} = \max\{h_0, h_1\}$ and $h_{\min} = \min\{h_0, h_1\}$, the convergence rate of $\hat{\tau}_{\text{FRDD}}$ is obtained as follows.

Theorem 7.2. *Consider the setup of Sections 7.1 and 7.2. Assume that $\hat{m}_z - m_z = O_p(r_n)$ for $z \in \{0, 1\}$, where $r_n \rightarrow 0$ as $n \rightarrow \infty$. Under Assumptions 3.2 and 7.1 and under Assumption 3.3, replacing $\mu_{t,\oplus}$, $M_{t,\oplus}(\nu, c)$, $Y(t)$ for $t \in \{0, 1\}$ with $\mu_{z,\oplus}^\dagger$, $M_{z,\oplus}^\dagger(\nu, c)$, $Y(T(z))$ for $z \in \{0, 1\}$, if $h_{\max} \rightarrow 0$ and $nh_{\min} \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$d_{\mathcal{H}}(\tau_{\text{FRDD}}, \hat{\tau}_{\text{FRDD}}) = O(h_{\max}^{2/(\beta_1-1)}) + O_p(r_n + (nh_{\min})^{-\frac{1}{2(\beta_2-1)}}).$$

Many common metric spaces satisfy Assumption 3.3 with $\beta_1 = \beta_2 = 2$. If one uses the local linear estimator (4) for estimating m_z , its typical convergence rate is $\tilde{r}_n = O(h_{\max}^2) + O_p((nh_{\min})^{-1/2})$ (cf. Theorems 19.1 and 19.2 in Hansen (2022)). Then letting $\beta_1 = \beta_2 = 2$, $h_0 = h_1 = h$ and $r_n = \tilde{r}_n$, we have

$$d_{\mathcal{H}}(\tau_{\text{FRDD}}, \hat{\tau}_{\text{FRDD}}) = O(h^2) + O_p((nh)^{-1/2}).$$

The right-hand side is optimized with $h = n^{-1/5}$ and the resulting convergence rate is the nonparametric rate $d_{\mathcal{H}}(\tau_{\text{FRDD}}, \hat{\tau}_{\text{FRDD}}) = O_p(n^{-2/5})$. In the fuzzy RDD setting, we can also perform a test for no treatment effect by implementing the test

under the sharp RDD discussed in Section 4.3. Indeed, under the null hypothesis $\mathbb{H}_0 : d_{\mathcal{H}}(\Psi(\mu_{1,\oplus}), \Psi(\mu_{0,\oplus})) = d(\mu_{1,\oplus}, \mu_{0,\oplus}) = 0$, we have $d(\mu_{1,\oplus}, \mu_{0,\oplus}) = 0$ if and only if $d(\mathbb{E}_{\oplus}[Y(T(1))|R = c], \mathbb{E}_{\oplus}[Y(T(0))|R = c]) = 0$.

8 Concluding remarks

We introduced a general framework for RDD for complex non-Euclidean outcomes, including functional data, compositional data and networks, where causal effects are quantified via geodesics between local Fréchet means. A key feature of the proposed framework is the distinction between intrinsic estimation and extrinsic inference: estimation is carried out within the original metric space using LFR, ensuring interpretability, while inference is conducted via Hilbert space embedding, which enables asymptotic analysis and hypothesis testing. For implementation, we develop a data-adaptive bandwidth selection method based on one-sided cross-validation near the cutoff. The resulting methodology is supported by theoretical guarantees and is demonstrated through simulations and an application.

Several promising directions remain for future research. First, extending GRDD to accommodate multiple cutoff points would enhance its utility in settings involving staggered interventions or tiered eligibility rules; see Bertanha (2020) and Cattaneo et al. (2021) for the scalar case. Second, generalizing GRDD to handle multivariate running variables would make the method applicable to more complex treatment assignment mechanisms (Black, 1999; Keele and Titiunik, 2015; Sawada et al., 2024; Cattaneo et al., 2025). This extension introduces challenges related to neighborhood construction and bandwidth selection in higher-dimensional spaces.

Acknowledgements

We wish to thank the editor, the associate editor, and referees for their valuable comments, which have significantly improved the quality of the paper.

Conflicts of interest: None declared.

Funding

D.K. was supported in part by JSPS KAKENHI Grant Numbers 23K12456, 25K00624, and 26K00323. H.G.M. was partially supported by NSF grant DMS-2310450. We thank Takuya Ishihara and Yasumasa Matsuda for their helpful comments and discussion.

Data availability

All source code required to implement the proposed methods and reproduce the simulation and empirical results in this paper is publicly available at <https://github.com/yidongzhou/GRDD>. The UK parliamentary election data analyzed in Section 6 were compiled by [Eggers and Spirling \(2017\)](#) and are available through the JOP Dataverse at <https://doi.org/10.7910/DVN/7QLCV1>.

References

- Ambrosio, L., Gigli, N. and Savaré, G. (2008) *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2nd edn.
- Arnaudon, M., Barbaresco, F. and Yang, L. (2013) Riemannian medians and means with applications to radar signal processing. *IEEE Journal of Selected Topics in Signal Processing*, **7**, 595–604.
- Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2007) Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, **29**, 328–347.

- Athey, S. and Imbens, G. W. (2006) Identification and inference in nonlinear difference-in-differences models. *Econometrica*, **74**, 431–497.
- Bertanha, M. (2020) Regression discontinuity design with many thresholds. *Journal of Econometrics*, **218**, 216–241.
- Bigot, J., Gouet, R., Klein, T. and López, A. (2017) Geodesic PCA in the Wasserstein space by convex PCA. *Annales de l’Institut Henri Poincaré B: Probability and Statistics*, **53**, 1–26.
- Black, S. E. (1999) Do better schools matter? parental valuation of elementary education. *The Quarterly Journal of Economics*, **114**, 577–599.
- Bridson, M. R. and Haefliger, A. (1999) *Metric Spaces of Non-Positive Curvature*. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg.
- 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**, 192–210.
- Calonico, S., Cattaneo, M. D., Farrell, M. H. and Titiunik, R. (2017) rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, **17**, 372–404. URL: <https://doi.org/10.1177/1536867X1701700208>.
- Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82**, 2295–2326.
- Cattaneo, M. D. and Escanciano, J. C. (2017) *Regression Discontinuity Designs: Theory and Applications*, vol. 38 of *Advances in Econometrics*. Emerald Group Publishing.
- 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**, 1941–1952.
- Cattaneo, M. D. and Titiunik, R. (2022) Regression discontinuity designs. *Annual Review of Economics*, **14**, 821–851.
- Cattaneo, M. D., Titiunik, R. and Vazquez-Bare, G. (2020) The Regression Discontinuity Design. In *The SAGE Handbook of Research Methods in Political Science and International Relations*, 835–857. SAGE Publications Ltd.
- Cattaneo, M. D., Titiunik, R. and Yu, R. R. (2025) Estimation and inference in boundary discontinuity designs. *arXiv preprint arXiv:2505.05670*.
- Cazelles, E., Seguy, V., Bigot, J., Cuturi, M. and Papadakis, N. (2018) Geodesic PCA versus log-PCA of histograms in the Wasserstein space. *SIAM Journal on Scientific Computing*, **40**, B429–B456.
- Chen, Y., Lin, Z. and Müller, H.-G. (2023) Wasserstein regression. *Journal of the American Statistical Association*, **118**, 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. *Annals of Statistics*, **50**, 1573–1592.
- Chen, Y. and Whalley, A. (2012) Green infrastructure: The effects of urban rail transit on air quality. *American Economic Journal: Economic Policy*, **4**, 58–97.
- Cherian, A. and Sra, S. (2016) *Positive Definite Matrices: Data Representation and Applications to Computer Vision*, 93–114. Cham: Springer. URL: https://doi.org/10.1007/978-3-319-45026-1_4.
- 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**, 589–618.
- Chiou, J.-M., Müller, H.-G. and Wang, J.-L. (2004) Functional response models. *Statistica Sinica*, **14**, 675–693.
- Dai, X. (2022) Statistical inference on the Hilbert sphere with application to random densities. *Electronic Journal of Statistics*, **16**, 700–736.
- Dryden, I. L., Koloydenko, A. and Zhou, D. (2009) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, **3**, 1102–1123.
- Dubey, P., Chen, Y. and Müller, H.-G. (2024) Metric statistics: Exploration and inference for random objects with distance profiles. *Annals of Statistics*, **52**, 757–792.
- Dubey, P. and Müller, H.-G. (2020) Functional models for time-varying random objects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **82**, 275–327.
- Ecker, K., de Luna, X. and Schelin, L. (2024) Causal inference with a functional outcome. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **73**, 221–240.
- Eggers, A. C. and Spirling, A. (2017) Incumbency effects and the strength of party preferences: Evidence from multiparty elections in the united kingdom. *The Journal of Politics*, **79**, 903–920.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.
- Faraway, J. J. (1997) Regression analysis for a functional response. *Technometrics*, **39**, 254–261.
- Frandsen, B. R., Frölich, M. and Melly, B. (2012) Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics*, **168**, 382–395.
- Gendron-Carrier, N., Gonzalez-Navarro, M., Polloni, S. and Turner, M. A. (2022) Subways and urban air pollution. *American Economic Journal: Applied Economics*, **14**, 164–96. URL: <https://www.aeaweb.org/articles?id=10.1257/app.20180168>.

- Gunsilius, F., Hsieh, M. H. and Lee, M. J. (2024) Tangential Wasserstein projections. *Journal of Machine Learning Research*, **25**, 1–41.
- Gunsilius, F. F. (2023) Distributional synthetic controls. *Econometrica*, **91**, 1105–1117.
- Hahn, J., Todd, P. and Van der Klaauw, W. (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, **69**, 201–209.
- Hansen, B. (2022) *Econometrics*. Princeton University Press.
- Hausman, C. and Rapson, D. S. (2018) Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, **10**, 533–552.
- Hron, K., Menafoglio, A., Templ, M., Hruzova, K. and Filzmoser, P. (2016) Simplicial principal component analysis for density functions in Bayes spaces. *Computational Statistics & Data Analysis*, **94**, 330–350.
- Hsing, T. and Eubank, R. (2015) *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, vol. 997. John Wiley & Sons.
- Imbens, G. and Kalyanaraman, K. (2012) Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, **79**, 933–959. URL: <https://doi.org/10.1093/restud/rdr043>.
- Imbens, G. W. and Lemieux, T. (2008) Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142**, 615–635.
- Imbens, G. W. and Rubin, D. B. (2005) *Causal Inference*. Cambridge University Press.
- Jayasooriya, V., Ng, A., Muthukumaran, S. and Perera, B. (2017) Green infrastructure practices for improvement of urban air quality. *Urban Forestry & Urban Greening*, **21**, 34–47. URL: <https://www.sciencedirect.com/science/article/pii/S1618866715301539>.
- Keele, L. J. and Titunik, R. (2015) Geographic boundaries as regression discontinuities. *Political Analysis*, **23**, 127–155.
- Kolaczyk, E. D. and Csárdi, G. (2014) *Statistical analysis of network data with R*, vol. 65. Springer New York.
- Kosorok, M. R. (2008) *Introduction to empirical processes and semiparametric inference*. Springer.
- Kurusu, D., Zhou, Y., Otsu, T. and Müller, H.-G. (2024) Geodesic causal inference. *arXiv:2406.19604*.
- (2025) Geodesic synthetic control methods for random objects and functional data. *arXiv:2505.00331*.
- Lee, J. M. (2018) *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer Cham, 2nd edn.

- Li, H. (2015) Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annual Review of Statistics and Its Application*, **2**, 73–94.
- Lin, Z. (2019) Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications*, **40**, 1353–1370.
- Lin, Z., Kong, D. and Wang, L. (2023) Causal inference on distribution functions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **85**, 378–398.
- 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**, 159–208.
- McCann, R. J. (1997) A convexity principle for interacting gases. *Advances in Mathematics*, **128**, 153–179.
- Morris, J. S. (2015) Functional regression. *Annual Review of Statistics and Its Application*, **2**, 321–359.
- Pauly, M. (2011) Weighted resampling of martingale difference arrays with applications. *Electronic Journal of Statistics*, **5**, 41–52.
- Pegoraro, M. and Beraha, M. (2022) Projected statistical methods for distributional data on the real line with the Wasserstein metric. *Journal of Machine Learning Research*, **23**, 1686–1744.
- Petersen, A. and Müller, H.-G. (2019) Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics*, **47**, 691–719.
- Petersen, A., Zhang, C. and Kokoszka, P. (2022) Modeling probability density functions as data objects. *Econometrics and Statistics*, **21**, 159–178.
- Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, vol. 426 of *Springer Series in Statistics*. Springer New York, NY, 2nd edn.
- Rao, C. R. (1945) Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of Calcutta Mathematical Society*, **37**, 81–91.
- Sawada, M., Ishihara, T., Kurisu, D. and Matsuda, Y. (2024) Local-polynomial estimation for multivariate regression discontinuity designs. *arXiv preprint arXiv:2402.08941*.
- Scealy, J. and Welsh, A. (2011) Regression for compositional data by using distributions defined on the hypersphere. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **73**, 351–375.
- (2014) Colours and cocktails: Compositional data analysis. *Australian & New Zealand Journal of Statistics*, **56**, 145–169.
- Schoenberg, I. J. (1938) Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, **44**, 522–536.

- Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013) Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, **41**, 2263–2291.
- Severn, K. E., Dryden, I. L. and Preston, S. P. (2022) Manifold valued data analysis of samples of networks, with applications in corpus linguistics. *Annals of Applied Statistics*, **16**, 368–390.
- Srivastava, A., Jermyn, I. and Joshi, S. (2007) Riemannian analysis of probability density functions with applications in vision. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.
- Stone, C. J. (1980) Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**, 1348–1360.
- Sturm, K.-T. (2003) Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002)*. *Contemp. Math.*, 338. *Amer. Math. Soc., Providence, RI*, **338**, 357–390.
- 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**, 20230004.
- van der Vaart, A. and Wellner, J. (2023) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer Cham, 2nd edn.
- Van Dijke, D. (2025) Regression discontinuity design with distribution-valued outcomes. *arXiv preprint arXiv:2504.03992*.
- Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016) Functional Data Analysis. *Annual Review of Statistics and its Application*, **3**, 257–295.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.
- Zhou, Y., Kurisu, D., Otsu, T. and Müller, H.-G. (2025) Geodesic difference-in-differences. *arXiv:2501.17436*.
- Zhou, Y. and Müller, H.-G. (2022) Network regression with graph Laplacians. *Journal of Machine Learning Research*, **23**, 1–41.
- Zhou, Y. and Müller, H.-G. (2024) Wasserstein regression with empirical measures and density estimation for sparse data. *Biometrics*, **80**, ujae127.

SUPPLEMENTARY MATERIAL

S.1 Local Fréchet regression

Local Fréchet regression (LFR) (Petersen and Müller, 2019; Chen and Müller, 2022) generalizes local linear regression (Fan and Gijbels, 1996) to settings where the outcome lies in a metric space. Here we outline the intuition and construction of LFR by first revisiting local linear regression, then drawing parallels that motivate the metric-space extension.

S.1.1 Local linear regression as a weighted Fréchet mean

Let $Y \in \mathbb{R}$ be a scalar outcome and $R \in \mathbb{R}$ a scalar predictor. The local linear regression function $m(r)$ estimates the conditional mean $E[Y|R = r]$ by solving a localized least squares problem centered at r :

$$(\beta_0, \beta_1) = \arg \min_{\beta_0, \beta_1} E \left[K_h(R - r) \{Y - \beta_0 - \beta_1(R - r)\}^2 \right],$$

where $K_h(u) = h^{-1}K(u/h)$ is a kernel function with bandwidth $h > 0$.

The solution for β_0 admits a closed-form expression. Let $\mu_k = E[K_h(R - r)(R - r)^k]$ for $k = 0, 1, 2$ and define the weighted cross-moments:

$$\xi_k = E[K_h(R - r)(R - r)^k Y], \quad k = 0, 1,$$

and set $\sigma^2 = \mu_0\mu_2 - \mu_1^2$. Then:

$$\beta_0 = \frac{\mu_2\xi_0 - \mu_1\xi_1}{\sigma^2}, \quad \beta_1 = \frac{\mu_0\xi_1 - \mu_1\xi_0}{\sigma^2}.$$

This yields the regression function:

$$m(r) = \beta_0 = E[s(r; R, h) Y],$$

where the weight function is defined by:

$$s(r; R, h) = \frac{1}{\sigma^2} K_h(R - r) (\mu_2 - \mu_1(R - r)).$$

Moreover, $m(r)$ can also be characterized as a weighted mean:

$$m(r) = \arg \min_{y \in \mathbb{R}} \mathbb{E} [s(r; R, h)(Y - y)^2] .$$

That is, local linear regression solves a local weighted Fréchet mean problem in the Euclidean space, where squared Euclidean distance is used and the weights adapt to the proximity of R to r via a kernel function.

S.1.2 Local Fréchet regression

The insight from the above formulation extends naturally to non-Euclidean settings. Suppose now that the outcome Y resides in a metric space $(\mathcal{M}, d)$, and $R \in \mathbb{R}$ remains the scalar running variable. The local Fréchet regression function $m(r)$ is defined as:

$$m(r) = \arg \min_{\nu \in \mathcal{M}} \mathbb{E} [s(r; R, h) d^2(Y, \nu)] ,$$

where the weight function $s(r; R, h)$ is the same as in the Euclidean case:

$$s(r; R, h) = \frac{1}{\sigma^2} K_h(R - r) (\mu_2 - \mu_1(R - r)) ,$$

with $\mu_k = \mathbb{E}[K_h(R - r)(R - r)^k]$ and $\sigma^2 = \mu_0\mu_2 - \mu_1^2$ as before.

Thus, LFR generalizes local linear regression by replacing the Euclidean outcome $Y \in \mathbb{R}$ and the squared error loss $(Y - y)^2$ with a metric space-valued outcome $Y \in \mathcal{M}$ and the squared metric loss $d^2(Y, \nu)$.

S.1.3 Sample estimator

Given a sample $\{(Y_i, R_i)\}_{i=1}^n$, we define the empirical moments:

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n K_h(R_i - r)(R_i - r)^k, \quad k = 0, 1, 2,$$

and $\hat{\sigma}^2 = \hat{\mu}_0\hat{\mu}_2 - \hat{\mu}_1^2$. The sample weight function becomes:

$$\hat{s}(r; R_i, h) = \frac{1}{\hat{\sigma}^2} K_h(R_i - r) (\hat{\mu}_2 - \hat{\mu}_1(R_i - r)) .$$

The local Fréchet regression estimator is then:

$$\hat{m}(r) = \arg \min_{\nu \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \hat{s}(r; R_i, h) d^2(Y_i, \nu).$$

S.2 Air pollution and urban rail in Taipei

As an illustration of the proposed GRDD methodology for functional outcomes, we revisit the Taipei air pollution setting studied by [Chen and Whalley \(2012\)](#) and consider a functional extension of their analysis. The data consist of hourly carbon monoxide (CO) measurements from ten monitoring stations in the Taipei metropolitan area, obtained from the Taiwanese Environmental Protection Administration for the period 1994–2007. Among these stations, seven are located near major highways and three are located farther away, allowing a comparison of areas with different exposure to traffic-related pollution.

The Taipei Metro underwent a major expansion in March 1996. Following [Chen and Whalley \(2012\)](#), we take March 28, 1996, the official opening date as the cutoff and use calendar time, measured in days relative to this date, as the running variable. The outcome for each observational unit is a daily CO concentration curve, which we view as an element of a Hilbert space of functions (Example 1). In this setting, the geodesic between two curves corresponds to pointwise linear interpolation. Since the measurements are recorded hourly and exhibit substantial missingness and noise, the functional observations are not directly observed in continuous form but must be reconstructed from discrete data. We therefore first smooth the hourly measurements to obtain station-day curves. This preprocessing step falls within the broader problem of empirical function estimation from discrete and potentially incomplete data, which is common in functional data analysis ([Ramsay and Silverman, 2005](#); [Yao et al., 2005](#); [Wang et al., 2016](#)).

These reconstructed station-level curves are then aggregated within the near-highway and far-from-highway groups to obtain group-level daily curves, yielding a sequence of functional observations suitable for GRDD analysis. Aggregation is motivated by the presence

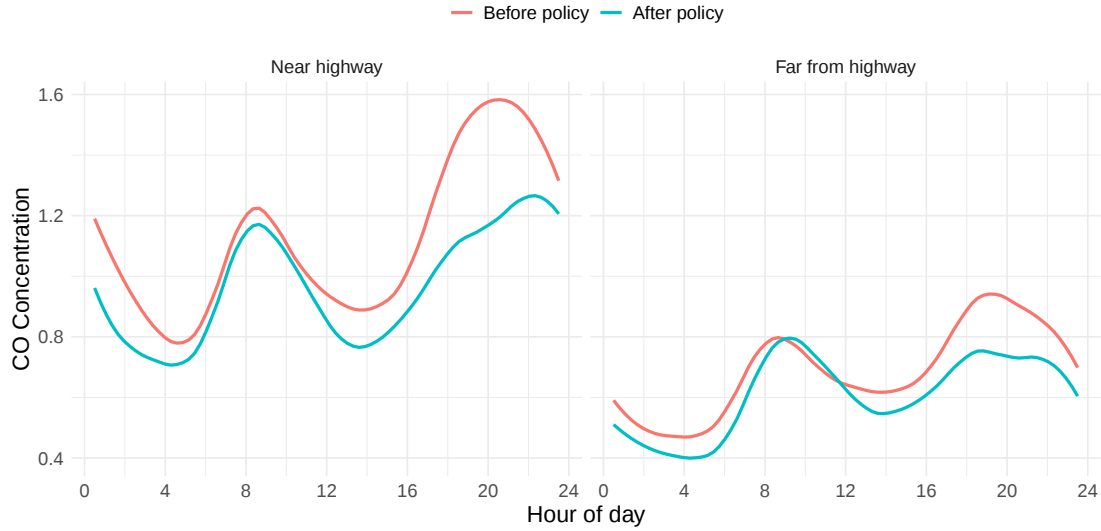


Figure S1: Estimated daily CO concentration curves before and after the Taipei Metro expansion. Each panel shows the average 24-hour CO concentration curves estimated for stations near highways (left) and far from highways (right) using geodesic sharp RDD. The blue and red curves represent the estimated CO concentration trajectories just before and just after the March 28, 1996 policy cutoff.

of missingness and measurement noise in the station-level data, which can vary across stations and over time and may affect the stability of the reconstructed curves. Constructing group-level curves provides a more stable representation of temporal pollution patterns near the cutoff, although it may smooth over some station-level heterogeneity. The results should therefore be interpreted with this trade-off in mind ([Gendron-Carrier et al., 2022](#)).

This application serves primarily to demonstrate how GRDD can be implemented in a functional data setting. The empirical design follows a regression discontinuity in time framework, with daily pollution profiles treated as curve-valued outcomes. As in related time-based RDD settings ([Hausman and Rapson, 2018](#)), interpretation of the estimated discontinuity may be affected by features such as gradual policy implementation and temporal dependence in the data. Accordingly, the results are best viewed as summarizing local changes in pollution profiles around the cutoff, rather than as definitive causal evidence. Extending GRDD to dependent data settings (e.g., under mixing conditions) is an important direction for future work.

While one could instead estimate station-level curves and apply pointwise scalar RDD methods at each time point, such approaches treat each hour independently and do not account for the inherent smoothness and functional structure of the data. In contrast, GRDD models the discontinuity at the level of entire curves, capturing cross-time dependence in daily pollution patterns and facilitating interpretation of changes in overall curve shape. At the same time, GRDD should be viewed as complementary to, rather than a replacement for, pointwise approaches in this setting.

We apply GRDD separately to the near-highway and far-from-highway groups using local Fréchet regression with the data-adaptive bandwidth selector described in Section 5.1. The selected bandwidths are 60 days for the near-highway and far-from-highway groups. Figure S1 displays the estimated average 24-hour CO concentration curves immediately before and after the cutoff. For stations near highways, the estimated post-cutoff curve lies below the pre-cutoff curve over much of the day, with the largest differences occurring during peak traffic hours. For stations farther from highways, the differences are smaller and more diffuse.

To further visualize the evolution of the fitted curves around the policy date, Figure S2 displays the local Fréchet regression surfaces over a 200-day window centered at the cutoff. The results suggest a more pronounced post-cutoff decline in the near-highway group relative to the far-from-highway group (Chen and Whalley, 2012; Jayasooriya et al., 2017). Overall, this application illustrates how GRDD can be used to summarize local changes in functional outcomes around a cutoff.

S.3 Geodesic fuzzy RDD

One might prefer to treat compliers' LATE at the cutoff as a quantity defined within the metric space $\mathcal{M}$, rather than involving an extrinsic Hilbert space. For such an intrinsic approach, the geodesic from $\mu_{0,\oplus}$ to $\mu_{1,\oplus}$ can be considered as the causal effect of interest

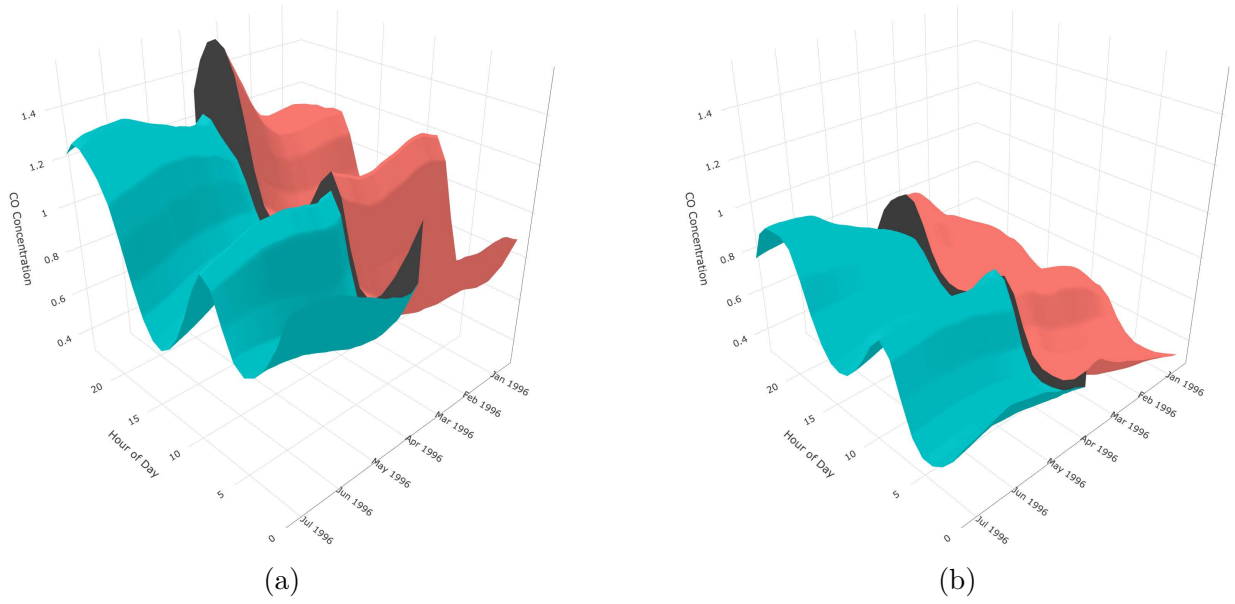


Figure S2: Daily CO concentration surfaces around the metro policy cutoff for monitoring stations located (a) near and (b) far from highways. Dates before the policy cutoff are shown in red, and those after the cutoff are shown in blue. The policy cutoff date (March 28, 1996) is marked by a gray vertical wall. Both panels use a shared vertical axis scale to facilitate comparison.

instead of τ_{FRDD} .

Definition S1 (Compliers' GLATE at the cutoff). *The compliers' GLATE is defined as the geodesic from $\mu_{0,\oplus}$ to $\mu_{1,\oplus}$, that is,*

$$\tau_{\text{GFRD}} = \gamma_{\mu_{0,\oplus}, \mu_{1,\oplus}}.$$

In contrast to compliers' LATE in Definition 7.1, compliers' GLATE τ_{GFRD} requires identification of the starting and ending points of the geodesic, i.e., $\mu_{z,\oplus}$. To this end, define $\mu_{\oplus}$ as the Fréchet mean of the outcome for always/never-takers at the cutoff, that is, $\mu_{\oplus} = E_{\oplus}[Y(T)|T(1) = T(0), R = c]$. When the metric space $(\mathcal{M}, d)$ admits a map Ψ satisfying Assumption 4.1, one can see that

$$\Psi(\mu_{z,\oplus}) = \Psi(\mu_{\oplus}) + \frac{\Psi(\mu_{z,\oplus}^{\dagger}) - \Psi(\mu_{\oplus})}{E[T(1)|R = c] - E[T(0)|R = c]} \quad \text{for } z \in \{0, 1\}, \quad (\text{S1})$$

where $\mu_{z,\oplus}^{\dagger} = E_{\oplus}[Y(T(z))|R = c]$ (see (S20) for the derivation of (S1)). Since $\Psi : \mathcal{M} \rightarrow$

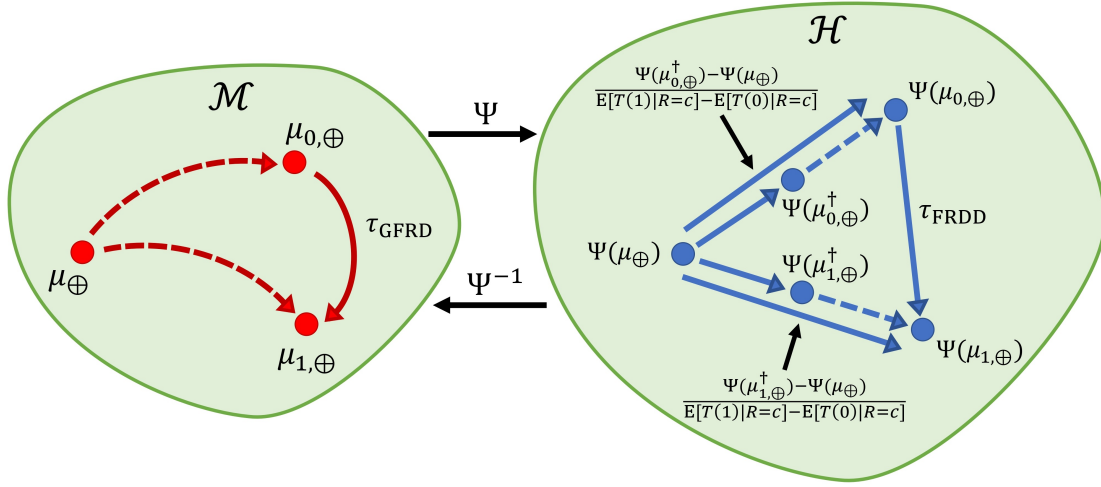


Figure S3: Illustration of the identification of τ_{FRDD} and τ_{GFRD} . The circles symbolize objects in the geodesic space $\mathcal{M}$, and the arrows emanating from these circles represent the geodesic paths connecting them. The solid blue arrow from $\Psi(\mu_{0,\oplus})$ to $\Psi(\mu_{1,\oplus})$ represents the compliers' LATE. The solid red arrow represents the compliers' GLATE.

$\Psi(\mathcal{M})$ is bijective, we have

$$\mu_{z,\oplus} = \Psi^{-1} \left(\Psi(\mu_{\oplus}) + \frac{\Psi(\mu_{z,\oplus}^{\dagger}) - \Psi(\mu_{\oplus})}{E[T(1)|R=c] - E[T(0)|R=c]} \right) \quad \text{for } z \in \{0, 1\},$$

where $\Psi^{-1} : \Psi(\mathcal{M}) \rightarrow \mathcal{M}$ is the inverse map of Ψ . In the following, we will assume that the argument of Ψ^{-1} is in the image space $\Psi(\mathcal{M})$ and the results are conditional on the event that the corresponding estimates introduced below also satisfy this condition. Figure S3 illustrates the identification of τ_{FRDD} and τ_{GFRD} . Note that we can estimate $\mu_{z,\oplus}^{\dagger}$ and m_z by $\hat{\nu}_{z,\oplus}$ and $\hat{m}_z$, respectively. Therefore, once $\mu_{\oplus}$ is identified, we can also identify $\mu_{z,\oplus}$.

However, in the present setup, $\mu_{\oplus}$ cannot be identified in general. As an important special case where $\mu_{\oplus}$ can be identified, we hereafter focus on the case of one-sided non-compliance:

Assumption S1. *For any unit, either one of the following holds.*

(i) $(Z, T) \in \{(1, 1), (0, 1), (0, 0)\}.$

(ii) $(Z, T) \in \{(1, 1), (1, 0), (0, 0)\}.$

Condition (i) implies that only compliers and always-takers exist. In this case, $\mu_{\oplus}$ becomes the Fréchet mean of the outcome for always-takers at the cutoff. On the other hand, Condition (ii) implies that only compliers and never-takers exist. In this case, $\mu_{\oplus}$ becomes the Fréchet mean of the outcome for never-takers at the cutoff.

Assume that $E_{\oplus}[Y(T)|T(1) = T(0), R = r]$ exists uniquely for any $R = r$ over a neighborhood of c , and that it is continuous over that neighborhood. We refer to Assumption 7.1 with this condition as Assumption 7.1'. Under one-sided noncompliance, we can identify $\mu_{\oplus}$ as follows.

Proposition S1. *Consider the setup of Sections 7.1 and 7.2 and suppose that Assumption 7.1' holds true. Then*

$$\mu_{\oplus} = \begin{cases} \lim_{r \uparrow c} E_{\oplus}[Y|T = 1, Z = 0, R = r] & \text{under Assumption S1(i)} \\ \lim_{r \downarrow c} E_{\oplus}[Y|T = 0, Z = 1, R = r] & \text{under Assumption S1(ii)} \end{cases}.$$

From this proposition, we can estimate τ_{GFRD} as follows.

$$\hat{\tau}_{\text{GFRD}} = \gamma_{\hat{\mu}_{0,\oplus}, \hat{\mu}_{1,\oplus}}, \quad \hat{\mu}_{z,\oplus} = \Psi^{-1} \left(\Psi(\hat{\mu}_{\oplus}) + \frac{\Psi(\hat{\nu}_{z,\oplus}) - \Psi(\hat{\mu}_{\oplus})}{\hat{m}_1 - \hat{m}_0} \right) \quad \text{for } z \in \{0, 1\},$$

where $\hat{\nu}_{z,\oplus}$ and $\hat{m}_z$ are the estimators considered in Section 3.2, and $\hat{\mu}_{\oplus}$ is the LFR estimator defined as

$$\hat{\mu}_{\oplus} = \begin{cases} \hat{\mu}_{A,\oplus} := \arg \min_{\nu \in \mathcal{M}} \frac{1}{n_A} \sum_{i \in I_A} \hat{s}_0(c; R_i, h_0) d^2(\nu, Y_i) & \text{under Assumption S1(i)} \\ \hat{\mu}_{N,\oplus} := \arg \min_{\nu \in \mathcal{M}} \frac{1}{n_N} \sum_{i \in I_N} \hat{s}_1(c; R_i, h_1) d^2(\nu, Y_i) & \text{under Assumption S1(ii)} \end{cases},$$

where $I_A = \{1 \leq i \leq n : T_i = 1, Z_i = 0\}$, $I_N = \{1 \leq i \leq n : T_i = 0, Z_i = 1\}$, n_S is the cardinality of I_S for $S \in \{A, N\}$, and $\hat{s}_z(c; r, h_z)$ is the estimated weight function of the LFR defined in Section 3.2.

For $z \in \{0, 1\}$, define $M_{\oplus}(\nu, c) = E[d^2(\nu, Y(T))|T(1) = T(0), R = c]$ and $\mu_{S,\oplus} = \arg \min_{\nu \in \mathcal{M}} M_{S,n}(\nu)$, $S \in \{A, N\}$ where

$$M_{A,n}(\nu) = E[s_0(c; R, h_0) d^2(\nu, Y(T))|T = 1, Z = 0],$$

$$M_{N,n}(\nu) = \mathbb{E}[s_1(c; R, h_1)d^2(\nu, Y(T))|T = 0, Z = 1].$$

Letting $h_{\max} = \max\{h_0, h_1\}$ and $h_{\min} = \min\{h_0, h_1\}$, the convergence rate of $\hat{\tau}_{\text{GFRD}}$ is obtained as follows.

Proposition S2. *Consider the setup of Sections 7.1 and 7.2. Under Assumptions 3.4, 3.5, 7.1', S1 and the assumptions in Theorem 7.2, except for Assumption 7.1, as well as Assumption 3.3 when replacing $\mu_{t,\oplus}$, $\nu_{t,\oplus}$, $\hat{\nu}_{t,\oplus}$, $M_{t,n}(\nu)$, $M_{t,\oplus}(\nu, c)$, $Y(t)$ for $t \in \{0, 1\}$ with $\mu_{\oplus}$, $\mu_{S,\oplus}$, $\hat{\mu}_{S,\oplus}$, $M_{S,n}(\nu)$, $M_{\oplus}(\nu, c)$, $Y(T)$ for $S \in \{A, N\}$, it holds that*

$$d_G(\tau_{\text{GFRD}}, \hat{\tau}_{\text{GFRD}}) = O(h_{\max}^{2/(\beta_1-1)}) + O_p(r_n + (nh_{\min})^{-\frac{1}{2(\beta_2-1)}}).$$

S.4 Fuzzy RDD for outcomes in Riemannian manifolds

In this subsection, we consider fuzzy RDD for outcomes in (a subset of) a Riemannian manifold $\mathcal{M} \subset \mathbb{R}^p$, for which the Riemannian logarithmic and exponential maps are well-defined. Let $T_{\mu}\mathcal{M}$ be the $(p-1)$ -dimensional tangent space at $\mu \in \mathcal{M}$. We assume that for any $\mu \in \mathcal{M}$, the logarithmic map $\text{Log}_{\mu}(\cdot) : \mathcal{M} \rightarrow T_{\mu}\mathcal{M}$ is bijective and $\text{Log}_{\mu}(\mathcal{M})$, the image space of $\mathcal{M}$ under Log_{μ} , is a convex subset of $\mathbb{R}^{p-1}$. A typical example for $\mathcal{M}$ is the space of compositional data with the arc-length distance (Example 2), which is the positive orthant of a finite-dimensional sphere.

For $z \in \{0, 1\}$, define $\mu_{z,\omega}^{\sharp} = \mathbb{E}[\text{Log}_{\omega}(Y(T(z))) | T(1) > T(0), R = c]$ where $\omega \in \mathcal{M}$ is a fixed reference point that can be chosen arbitrarily; a simple choice is the sample Fréchet mean of the outcomes $\{Y_i\}_{i=1}^n$. Due to the linear structure of the tangent space, one may consider a contrast of the means $\mu_{z,\omega}^{\sharp}$ of potential outcomes (transformed by the logarithmic map). Specifically, our causal estimand of interest is defined as follows.

Definition S2 (Compliers' LATE at the cutoff for outcomes in a Riemannian manifold).

For outcomes taking values in a Riemannian manifold considered in this subsection, we define the compliers' LATE at the cutoff with a reference point ω as

$$\tau_{\text{FRDD}}^{\#} := \mu_{1,\omega}^{\#} - \mu_{0,\omega}^{\#}.$$

When $\mathcal{M}$ is (a subset of) the Euclidean space $\mathbb{R}$, $\tau_{\text{FRDD}}^{\#}$ is independent of ω and is reduced to

$$\tau_{\text{FRDD}}^{\#} = \mathbb{E}[Y(T(1))|T(1) > T(0), R = c] - \mathbb{E}[Y(T(0))|T(1) > T(0), R = c],$$

which is the conventional compliers' LATE at the cutoff for Euclidean outcomes.

For the identification of $\tau_{\text{FRDD}}^{\#}$, we add the following conditions.

Assumption S2.

- (i) $T(0) \leq T(1)$.
- (ii) $\Pr(T(1) > T(0)|R = c) \in (\eta_0, 1 - \eta_0)$ for some $\eta_0 \in (0, 1/2)$.
- (iii) $\mathbb{E}[T(1)|R = r]$, $\mathbb{E}[T(0)|R = r]$, $\mathbb{E}[\text{Log}_{\omega}(Y(T(0)))|R = r]$, and $\mathbb{E}[\text{Log}_{\omega}(Y(T(1)))|R = r]$ are continuous over a neighborhood of c .

By a similar argument to Theorem 7.1, the causal effect $\tau_{\text{FRDD}}^{\#}$ can be identified as follows.

Theorem S1. *Consider the setup of Sections 7.1 and S.4 and suppose that Assumption S2 holds true. Then*

$$\tau_{\text{FRDD}}^{\#} = \frac{\lim_{r \downarrow c} \mathbb{E}[\text{Log}_{\omega}(Y)|R = r] - \lim_{r \uparrow c} \mathbb{E}[\text{Log}_{\omega}(Y)|R = r]}{\lim_{r \downarrow c} \mathbb{E}[T|R = r] - \lim_{r \uparrow c} \mathbb{E}[T|R = r]}.$$

From Theorem S1, we can construct an estimator of $\tau_{\text{FRDD}}^{\#}$ as follows.

$$\hat{\tau}_{\text{FRDD}}^{\#} = \frac{\hat{\nu}_{1,\omega}^{\#} - \hat{\nu}_{0,\omega}^{\#}}{\hat{m}_1 - \hat{m}_0},$$

where $\hat{m}_z$ is the estimator considered in Section 3.2 and $\hat{\nu}_{z,\omega}^\#$ is the LFR estimator defined as

$$\hat{\nu}_{z,\omega}^\# = \arg \min_{\nu \in T_\omega \mathcal{M}} \hat{M}_{z,n}(\nu), \quad \hat{M}_{z,n}(\nu) = \frac{1}{n_z} \sum_{i=1}^n \hat{s}_z(c; R_i, h_z) d_E^2(\nu, \text{Log}_\omega(Y_i)) \quad \text{for } z \in \{0, 1\},$$

where d_E is the standard Euclidean metric and $\hat{s}_z(c; r, h_z)$ is the estimated weight function of the LFR defined in Section 3.2.

For $z \in \{0, 1\}$, define $\nu_{z,\omega}^\# = \arg \min_{\nu \in T_\omega \mathcal{M}} M_{z,\oplus}^\#(\nu, c)$ and $\tilde{\nu}_{z,\omega}^\# = \arg \min_{\nu \in T_\omega \mathcal{M}} M_{z,n}^\#(\nu)$ where

$$\begin{aligned} M_{z,\oplus}^\#(\nu, c) &= \mathbb{E}[d_E^2(\nu, \text{Log}_\omega(Y(T(z)))) | R = c], \\ M_{z,n}^\#(\nu) &= \mathbb{E}[s_z(c; R, h_z) d_E^2(\nu, \text{Log}_\omega(Y(T(z))))]. \end{aligned}$$

Letting $h_{\max} = \max\{h_0, h_1\}$ and $h_{\min} = \min\{h_0, h_1\}$, the convergence rate of $\hat{\tau}_{\text{FRDD}}^\#$ is obtained as follows.

Theorem S2. *Consider the setup of Sections 7.1 and S.4. Assume that $\hat{m}_z - m_z = O_p(r_n)$, $z \in \{0, 1\}$, where r_n is a sequence of constants such that $r_n \rightarrow 0$ as $n \rightarrow \infty$. Under Assumptions 3.2 and S2 and Assumption 3.3 when replacing $(\mathcal{M}, d)$, $\mu_{t,\oplus}$, $\nu_{t,\oplus}$, $\hat{\nu}_{t,\oplus}$, $M_{t,n}(\nu)$, $M_{t,\oplus}(\nu, c)$, $Y(t)$ for $t \in \{0, 1\}$ with $(T_\omega \mathcal{M}, d_E)$, $\nu_{z,\omega}^\#$, $\tilde{\nu}_{z,\omega}^\#$, $\hat{\nu}_{z,\omega}^\#$, $M_{z,n}^\#(\nu)$, $M_{z,\oplus}^\#(\nu, c)$, $\text{Log}_\omega(Y(T(z)))$ for $z \in \{0, 1\}$, if $h_{\max} \rightarrow 0$ and $nh_{\min} \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$d_E(\tau_{\text{FRDD}}^\#, \hat{\tau}_{\text{FRDD}}^\#) = O(h_{\max}^{2/(\beta_1-1)}) + O_p(r_n + (nh_{\min})^{-\frac{1}{2(\beta_2-1)}}).$$

If one uses the sample Fréchet mean $\hat{m}_\oplus$ of the outcomes $\{Y_i\}_{i=1}^n$ as the reference point ω , then the techniques used to derive Theorem S2 cannot be directly applied since the resulting estimator involves the logarithmic map at the estimated point $\hat{m}_\oplus$. Chen et al. (2023) investigate the theoretical properties of estimators in a similar setting. By leveraging the techniques developed in that paper, it is expected that the asymptotic properties of the estimator can be established. However, additional substantial work is required and hence we leave this as an open problem for future research.

S.5 Geodesic fuzzy RDD for outcomes in Riemannian manifolds

As discussed in Section S.3, one may prefer to consider the compliers' LATE at the cutoff as a quantity defined on the manifold $\mathcal{M}$, i.e., the compliers' GLATE.

Definition S3 (Compliers' GLATE at the cutoff for outcomes in a Riemannian manifold).

For the outcomes in a Riemannian manifold considered in Section S.4, compliers' GLATE is defined as the geodesic from $\mu_{0,\omega,\oplus}^\sharp$ to $\mu_{1,\omega,\oplus}^\sharp$, that is,

$$\tau_{\text{GFRD}}^\sharp := \gamma_{\mu_{0,\omega,\oplus}^\sharp, \mu_{1,\omega,\oplus}^\sharp},$$

where $\mu_{z,\omega,\oplus}^\sharp = \text{Exp}_\omega(\mu_{z,\omega}^\sharp)$, $\mu_{z,\omega}^\sharp = \text{E}[\text{Log}_\omega(Y(T(z))) | T(1) > T(0), R = c]$, and $\text{Exp}_\omega(\cdot) : T_\omega\mathcal{M} \rightarrow \mathcal{M}$ is the exponential map at a fixed reference point $\omega \in \mathcal{M}$.

Let $\nu_{z,\omega}^\sharp = \text{E}[\text{Log}_\omega(Y(T(z))) | R = c]$ and let $\nu_\omega^\sharp$ be the expectation of $\text{Log}_\omega(Y)$ for always/never-takers at the cutoff, that is, $\nu_\omega^\sharp = \text{E}[\text{Log}_\omega(Y) | T(1) = T(0), R = c]$. For the Riemannian manifold considered in Section S.4,

$$\mu_{z,\omega,\oplus}^\sharp = \text{Exp}_\omega \left(\nu_\omega^\sharp + \frac{\nu_{z,\omega}^\sharp - \nu_\omega^\sharp}{m_1 - m_0} \right) \quad \text{for } z \in \{0, 1\}, \quad (\text{S2})$$

where $m_z = \text{E}[T(z) | R = c]$, assuming this is well-defined, i.e., the argument of Exp_ω is in the image space Log_ω ; see (S26) for the derivation of (S2). An illustration of the identification of $\tau_{\text{FRDD}}^\sharp$ and $\tau_{\text{GFRD}}^\sharp$ is shown in Figure S4. Note that $\nu_{z,\omega}^\sharp$ and m_z can be estimated by $\hat{\nu}_{z,\omega}^\sharp$ and $\hat{m}_z$, respectively. Therefore, if $\nu_\omega^\sharp$ can be estimated, we can construct an estimator of $\mu_{z,\omega,\oplus}^\sharp$.

Assume that $\text{E}[\text{Log}_\omega(Y) | T(1) = T(0), R = r]$ is continuous over a neighborhood of c and we refer to Assumption S2 with this condition as Assumption S2'. Under one-sided noncompliance, we can identify $\mu_\oplus$ as follows.

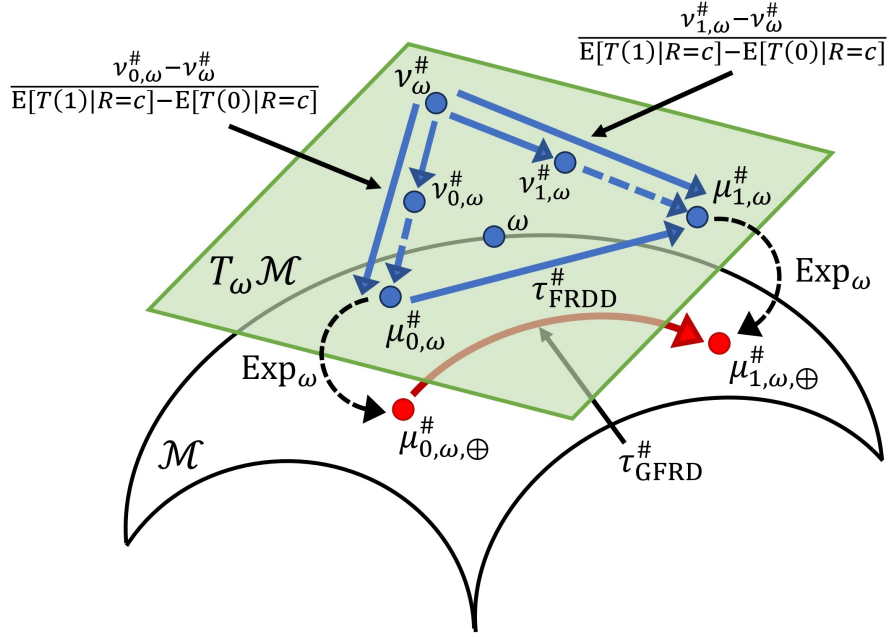


Figure S4: Illustration of the identification of $\tau_{\text{FRDD}}^\#$ and $\tau_{\text{GFRD}}^\#$. The circles symbolize objects in the geodesic space $\mathcal{M}$, and the arrows emanating from these circles represent the geodesic paths connecting them. The solid blue arrow from $\mu_{0,\omega}$ to $\mu_{1,\omega}$ represents the compliers' LATE. The solid red arrow represents the compliers' GLATE.

Proposition S3. *Consider the setup of Sections 7.1 and S.4. Under Assumption S2',*

$$\nu_\omega^\# = \begin{cases} \lim_{r \uparrow c} \mathbb{E}[\text{Log}_\omega(Y) | T = 1, Z = 0, R = r] & \text{under Assumption S1(i)} \\ \lim_{r \downarrow c} \mathbb{E}[\text{Log}_\omega(Y) | T = 0, Z = 1, R = r] & \text{under Assumption S1(ii)} \end{cases}.$$

Proposition S3 motivates estimators

$$\hat{\tau}_{\text{GFRD}}^\# = \gamma_{\hat{\mu}_{0,\omega,\oplus}^\#, \hat{\mu}_{1,\omega,\oplus}^\#}, \quad \hat{\mu}_{z,\omega,\oplus}^\# = \text{Exp}_\omega \left(\hat{\nu}_\omega^\# + \frac{\hat{\nu}_{z,\omega}^\# - \hat{\nu}_\omega^\#}{\hat{m}_1 - \hat{m}_0} \right) \quad \text{for } z \in \{0, 1\},$$

where $\hat{\nu}_{z,\omega}^\#$ and $\hat{m}_z$ are estimators considered in Sections S.4 and 3.2, respectively, and $\hat{\nu}_\omega^\#$ is the LFR estimator defined as

$$\hat{\nu}_\omega^\# = \begin{cases} \hat{\nu}_{\omega,A}^\# := \arg \min_{\nu \in T_\omega \mathcal{M}} \frac{1}{n_A} \sum_{i \in I_A} \hat{s}_0(c; R_i, h_0) d_E^2(\nu, \text{Log}_\omega(Y_i)) \\ \quad \text{under Assumption S1(i),} \\ \hat{\nu}_{\omega,N}^\# := \arg \min_{\nu \in T_\omega \mathcal{M}} \frac{1}{n_N} \sum_{i \in I_N} \hat{s}_1(c; R_i, h_1) d_E^2(\nu, \text{Log}_\omega(Y_i)) \\ \quad \text{under Assumption S1(ii)} \end{cases}.$$

For $z \in \{0, 1\}$, define $M^\sharp(\nu, c) = \mathbb{E}[d_E^2(\nu, \text{Log}_\omega(Y(T))) | T(1) = T(0), R = c]$,

$$\nu_{\omega, S}^\sharp = \arg \min_{\nu \in T_\omega \mathcal{M}} M_{S, n}^\sharp(\nu), \quad S \in \{A, N\},$$

$$M_{A, n}^\sharp(\nu) = \mathbb{E}[s_0(c; R, h_0) d_E^2(\nu, \text{Log}_\omega(Y(T))) | T = 1, Z = 0],$$

$$M_{N, n}^\sharp(\nu) = \mathbb{E}[s_1(c; R, h_1) d_E^2(\nu, \text{Log}_\omega(Y(T))) | T = 0, Z = 1].$$

Letting $h_{\max} = \max\{h_0, h_1\}$ and $h_{\min} = \min\{h_0, h_1\}$, the convergence rate of $\hat{\tau}_{\text{GFRD}}^\sharp$ is obtained as follows.

Proposition S4. *Consider the setup of Sections 7.1 and S.4. Suppose that Assumptions 3.4, 3.5, S1, S2', and assumptions in Theorem S2 except for Assumption S2 are satisfied and that there exists a constant $C_\omega > 0$ such that for any $\mu_1, \mu_2 \in T_\omega \mathcal{M}$, $d(\text{Exp}_\omega(\mu_1), \text{Exp}_\omega(\mu_2)) \leq C_\omega d_E(\mu_1, \mu_2)$. Additionally, suppose that Assumption 3.3 is satisfied when replacing $(\mathcal{M}, d)$, $\mu_{t, \oplus}$, $\nu_{t, \oplus}$, $\hat{\nu}_{t, \oplus}$, $M_{t, n}(\nu)$, $M_{t, \oplus}(\nu, c)$, $Y(t)$ for $t \in \{0, 1\}$ with $(T_\omega \mathcal{M}, d_E)$, $\nu_\omega^\sharp$, $\nu_{\omega, S}^\sharp$, $\hat{\nu}_{\omega, S}^\sharp$, $M_{S, n}^\sharp(\nu)$, $M^\sharp(\nu, c)$, $\text{Log}_\omega(Y(T))$ for $S \in \{A, N\}$. Then*

$$d_G(\tau_{\text{GFRD}}^\sharp, \hat{\tau}_{\text{GFRD}}^\sharp) = O(h_{\max}^{2/(\beta_1-1)}) + O_p(r_n + (nh_{\min})^{-\frac{1}{2(\beta_2-1)}}).$$

S.6 Proofs for Section 3

S.6.1 Proof of Theorem 3.1

For the proof of (1), it suffices to show

$$\mathbb{E}_\oplus[Y(1) | R = c] = \lim_{r \downarrow c} \mathbb{E}_\oplus[Y | R = r], \quad (\text{S3})$$

$$\mathbb{E}_\oplus[Y(0) | R = c] = \lim_{r \uparrow c} \mathbb{E}_\oplus[Y | R = r]. \quad (\text{S4})$$

Here we only prove (S3) since the proof of (S4) is almost the same. Observe that

$$\begin{aligned} \mathbb{E}_\oplus[Y(1) | R = c] &= \lim_{r \downarrow c} \mathbb{E}_\oplus[Y(1) | R = r] \\ &= \lim_{r \downarrow c} \mathbb{E}_\oplus[Y(1) | T = 1, R = r] \end{aligned}$$

$$\begin{aligned}
&= \lim_{r \downarrow c} E_{\oplus}[Y|T = 1, R = r] \\
&= \lim_{r \downarrow c} E_{\oplus}[Y|R = r],
\end{aligned}$$

where the first equality follows from Assumption 3.1(ii) and the third equality follows from Assumption 3.1(i). Therefore, we obtain the conclusion.

S.6.2 Proof of Theorem 3.2

For each side $t \in \{0, 1\}$ of the cutoff, the population target $\nu_{t,\oplus}$ and the estimator $\hat{\nu}_{t,\oplus}$ are defined through one-sided local Fréchet regression. Since these are exactly local Fréchet regression objects with Euclidean predictor $R \in \mathbb{R}$, the rates established in Theorems 3 and 4 of [Petersen and Müller \(2019\)](#) apply directly under our Assumptions 3.2 and 3.3. For completeness, we restate these results in our notation.

Theorem S3 (Theorem 3 of [Petersen and Müller \(2019\)](#)). *If Assumptions 3.2 and 3.3(i)–(iv) hold, then for each $t \in \{0, 1\}$,*

$$d(\mu_{t,\oplus}, \nu_{t,\oplus}) = O(h_t^{2/(\beta_1-1)}) \quad \text{as } h_t \rightarrow 0.$$

Theorem S4 (Theorem 4 of [Petersen and Müller \(2019\)](#)). *If Assumptions 3.2 and 3.3(i), (ii), and (v) hold, then for each $t \in \{0, 1\}$,*

$$d(\nu_{t,\oplus}, \hat{\nu}_{t,\oplus}) = O_p((nh_t)^{-1/(2(\beta_2-1))}) \quad \text{as } h_t \rightarrow 0 \text{ and } nh_t \rightarrow \infty.$$

Proof of (i). By definition of $d_{\mathcal{G}}$ and Assumption 3.5,

$$\begin{aligned}
d_{\mathcal{G}}(\tau_{\text{GRDD}}, \tau_{\text{GRDD}}^*) &= d_{\mathcal{G}}(\gamma_{\mu_{0,\oplus}, \mu_{1,\oplus}}, \gamma_{\nu_{0,\oplus}, \nu_{1,\oplus}}) \\
&= d(\Gamma_{\mu_{0,\oplus}, \mu_{1,\oplus}}(\omega_{\oplus}), \Gamma_{\nu_{0,\oplus}, \nu_{1,\oplus}}(\omega_{\oplus})) \\
&\leq C\{d(\mu_{0,\oplus}, \nu_{0,\oplus}) + d(\mu_{1,\oplus}, \nu_{1,\oplus})\}.
\end{aligned} \tag{S5}$$

Under Assumptions 3.2 and 3.3(i)–(iv), Theorem [S3](#) yields

$$d(\mu_{t,\oplus}, \nu_{t,\oplus}) = O(h_t^{2/(\beta_1-1)}), \quad t \in \{0, 1\}.$$

Substituting these bounds into (S5), we obtain

$$d_{\mathcal{G}}(\tau_{\text{GRDD}}, \tau_{\text{GRDD}}^*) = O(h_{\max}^{2/(\beta_1-1)}).$$

Proof of (ii). Again by Assumption 3.5,

$$\begin{aligned} d_{\mathcal{G}}(\tau_{\text{GRDD}}^*, \hat{\tau}_{\text{GRDD}}) &= d_{\mathcal{G}}(\gamma_{\nu_{0,\oplus}, \nu_{1,\oplus}}, \gamma_{\hat{\nu}_{0,\oplus}, \hat{\nu}_{1,\oplus}}) \\ &\leq C \{d(\nu_{0,\oplus}, \hat{\nu}_{0,\oplus}) + d(\nu_{1,\oplus}, \hat{\nu}_{1,\oplus})\}. \end{aligned} \quad (\text{S6})$$

Under Assumptions 3.2 and 3.3(i), (ii), and (v), Theorem S4 gives

$$d(\nu_{t,\oplus}, \hat{\nu}_{t,\oplus}) = O_p((nh_t)^{-1/(2(\beta_2-1))}), \quad t \in \{0, 1\}.$$

Combining these bounds with (S6), we obtain

$$d_{\mathcal{G}}(\tau_{\text{GRDD}}^*, \hat{\tau}_{\text{GRDD}}) = O_p((nh_{\min})^{-1/(2(\beta_2-1))}).$$

This completes the proof.

S.7 Hilbert space embedding

S.7.1 Functional data

Let $L^2(\mathcal{T})$ be the Hilbert space of square integrable functions where $\mathcal{T}$ is a compact interval.

In this example, we can take $\Psi = \text{id}$ since $\mathcal{H} := L^2(\mathcal{T})$ is the Hilbert space with the inner product $\langle f, g \rangle_{\mathcal{H}} = \int_{\mathcal{T}} f(t)g(t)dt$ which induces the L^2 metric $d_{L^2}(f, g) = (\int_{\mathcal{T}} \{f(t) - g(t)\}^2 dt)^{1/2}$.

S.7.2 Compositional data

For compositional data, we work with the square-root transformation so that the observations take values in the positive orthant of the unit sphere $\mathcal{S}_+^{d-1}$. To simplify notation, we write Y_i and $Y_i(t)$ for the square-root transformed compositions. Let $\mathcal{S}^{d-1} =$

$\{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| = 1\}$ be the unit sphere and let $T_{\mathbf{z}_0}\mathcal{S}^{d-1} = \{\mathbf{z} \in \mathbb{R}^d : \mathbf{z}_0^\top \mathbf{z} = 0\}$ denote the tangent space of $\mathcal{S}^{d-1}$ at the fixed point $\mathbf{z}_0 \in \mathcal{S}^{d-1}$. Using the logarithmic map $\text{Log}_{\mathbf{z}_0} : \mathcal{S}^{d-1} \setminus \{-\mathbf{z}_0\} \rightarrow T_{\mathbf{z}_0}\mathcal{S}^{d-1}$ defined as

$$\text{Log}_{\mathbf{z}_0}(\mathbf{z}) = \frac{\theta}{\sin(\theta)}(\mathbf{z} - \cos(\theta)\mathbf{z}_0), \quad \theta = d_g(\mathbf{z}_0, \mathbf{z}),$$

compositional data can be embedded into $\text{Log}_{\mathbf{z}_0}(\mathcal{S}_+^{d-1})$, a subset of the linear space $T_{\mathbf{z}_0}\mathcal{S}^{d-1}$.

Under the arc-length metric d_g , the null hypothesis of no treatment effect is

$$\mathbb{H}_0 : d_g(\mu_{0,\oplus}, \mu_{1,\oplus}) = 0.$$

Since $d_g(\mu_{0,\oplus}, \mu_{1,\oplus}) = 0$ is equivalent to $\mu_{1,\oplus} = \mu_{0,\oplus}$, and the Fréchet mean on the positive orthant of the unit sphere is characterized by the first-order optimality condition $\mathbb{E}[\text{Log}_{\mu_{1,\oplus}}(Y(1))|R = c] = 0$ (Lee, 2018), it follows that $\mu_{1,\oplus} = \mu_{0,\oplus}$ if and only if $\mathbb{E}[\text{Log}_{\mu_{0,\oplus}}(Y(1))|R = c] = 0$. Therefore, for compositional outcomes, testing for no treatment effect is equivalent to testing

$$\mathbb{H}'_0 : \mathbb{E}[\text{Log}_{\mu_{0,\oplus}}(Y(1))|R = c] = 0.$$

In practice, the base point $\mu_{0,\oplus}$ is unknown and is replaced by its local Fréchet regression estimator $\hat{\nu}_{0,\oplus}$. Consider the estimator

$$\hat{\nu}_t^\Psi = \frac{1}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h_t) \text{Log}_{\hat{\nu}_{0,\oplus}}(Y_i), \quad t = 0, 1.$$

To analyze $\hat{\nu}_1^\Psi$, write

$$\begin{aligned} \hat{\nu}_1^\Psi &= \frac{1}{n_1} \sum_{i=1}^n \hat{s}_1(c; R_i, h_1) \text{Log}_{\hat{\nu}_{0,\oplus}}(Y_i(1)) \\ &= \frac{1}{n_1} \sum_{i=1}^n \hat{s}_1(c; R_i, h_1) \text{Log}_{\mu_{0,\oplus}}(Y_i(1)) \\ &\quad + \frac{1}{n_1} \sum_{i=1}^n \hat{s}_1(c; R_i, h_1) \{\text{Log}_{\hat{\nu}_{0,\oplus}}(Y_i(1)) - \text{Log}_{\mu_{0,\oplus}}(Y_i(1))\} \\ &=: D_1 + D_2. \end{aligned}$$

The first term D_1 is the ideal statistic constructed with the true base point. Under the null hypothesis $\mathbb{H}'_0$ and $n_1 h_1^5 \rightarrow 0$, $\sqrt{n_1 h_1} D_1$ is asymptotically normal.

For the second term, we use the Lipschitz continuity of the logarithmic map on the positive orthant of the unit sphere to obtain

$$\begin{aligned} \|D_2\| &\leq \frac{1}{n_1} \sum_{i=1}^n |\hat{s}_1(c; R_i, h_1)| \left\| \text{Log}_{\hat{\nu}_{0,\oplus}}(Y_i(1)) - \text{Log}_{\mu_{0,\oplus}}(Y_i(1)) \right\| \\ &\lesssim d_g(\hat{\nu}_{0,\oplus}, \mu_{0,\oplus}) \frac{1}{n_1} \sum_{i=1}^n |\hat{s}_1(c; R_i, h_1)|. \end{aligned}$$

Since the local Fréchet regression estimator satisfies

$$d_g(\hat{\nu}_{0,\oplus}, \mu_{0,\oplus}) = O_p(h_0^2 + (n_0 h_0)^{-1/2}),$$

see [Petersen and Müller \(2019\)](#), and

$$\frac{1}{n_1} \sum_{i=1}^n |\hat{s}_1(c; R_i, h_1)| = O_p(1),$$

it follows that

$$\|D_2\| = O_p(h_0^2 + (n_0 h_0)^{-1/2}).$$

Now assume that $n_1/n_0 \rightarrow c \in (0, \infty)$, $h_0 \asymp n^{-1/5}$ and $h_1 \ll n^{-1/5}$. Then

$$\begin{aligned} \sqrt{n_1 h_1} \|D_2\| &= O_p\left(\sqrt{n_1 h_1} h_0^2\right) + O_p\left(\sqrt{\frac{n_1 h_1}{n_0 h_0}}\right) \\ &= O_p\left(\sqrt{n_1 h_1} n^{-2/5}\right) + O_p\left(\sqrt{\frac{h_1}{h_0}}\right). \end{aligned}$$

Both terms are $o_p(1)$, and hence

$$\sqrt{n_1 h_1} \|D_2\| = o_p(1).$$

Consequently,

$$\sqrt{n_1 h_1} \hat{\nu}_1^\Psi = \frac{\sqrt{n_1 h_1}}{n_1} \sum_{i=1}^n \hat{s}_1(c; R_i, h_1) \text{Log}_{\mu_{0,\oplus}}(Y_i(1)) + o_p(1),$$

Therefore, the effect of estimating the base point is asymptotically negligible, so that the test statistic behaves as if the true embedding $\text{Log}_{\mu_{0,\oplus}}$ were used. It follows that the bootstrap-based testing procedure described in Section 4.3, when implemented with

$$\Psi = \text{Log}_{\hat{\mu}_{0,\oplus}},$$

is asymptotically equivalent to the procedure based on the oracle embedding $\text{Log}_{\mu_{0,\oplus}}$. Consequently, it yields a valid test of

$$\mathbb{H}'_0 : \mathbb{E}[\text{Log}_{\mu_{0,\oplus}}(Y(1)) | R = c] = 0.$$

As established above, this condition is equivalent to

$$\mathbb{H}_0 : d_g(\mu_{0,\oplus}, \mu_{1,\oplus}) = 0,$$

which corresponds to the null hypothesis of no treatment effect for compositional outcomes under the arc-length metric.

S.7.3 SPD matrices

First, we consider the space of SPD matrices Sym_m^+ with the Frobenius metric d_F defined as $d_F(A, B) = \|A - B\|_F$ where $\|A\|_F$ is the Frobenius norm of a real matrix A . In this example, we can take $\Psi = \text{id}$ where id is the identity map since $A \in \text{Sym}_m^+$ is also an element of the Hilbert space $\mathcal{H}$ with the inner product $\langle A, B \rangle_{\mathcal{H}} = \text{tr}(A^\top B)$ where $\text{tr}(A)$ is the trace of a real matrix A and this inner product induces the Frobenius metric. Let $\lambda_{\min}(A)$ be the minimum eigenvalue of $A \in \text{Sym}_m^+$. For $\varepsilon > 0$, define the space $\text{Sym}_{m,\varepsilon}^+ = \{A \in \text{Sym}_m^+ : \lambda_{\min}(A) \geq \varepsilon\}$. We can see that $\Psi(\text{Sym}_{m,\varepsilon}^+) = \text{Sym}_{m,\varepsilon}^+$ is a closed convex subset of Sym_m^+ .

Second, we consider the space of SPD matrices with the power metric family defined as

$$d_{F,p}(A, B) = d_F(F_p(A), F_p(B)),$$

where $p > 0$ is a constant and F_p is a matrix power map defined as

$$F_p(A) = A^p = U\Lambda^p U^\top : \text{Sym}_m^+ \rightarrow \text{Sym}_m^+,$$

where $U\Lambda U^\top$ is the usual spectral decomposition of $A \in \text{Sym}_m^+$. Applying almost the same arguments for the space of SPD matrices with the Frobenius metric, one can verify Assumption 4.1 with $\Psi(A) = F_p(A)$, $\mathcal{M} = \text{Sym}_{m,\varepsilon}^+$.

Third, we consider the space of SPD matrices with the Log-Euclidean metric defined as

$$d_{\text{LE}}(A, B) = d_F(\log(A), \log(B)),$$

where $\log(A)$ is the matrix logarithmic map for a real matrix A . Again applying similar arguments as for the space of SPD matrices with the Frobenius metric, one can verify Assumption 4.1 with $\Psi(A) = \log(A)$, $\mathcal{M} = \text{Sym}_m^+$.

Fourth, we consider the space of SPD matrices with the Log-Cholesky metric. Let $\mathcal{L}_m^+$ be the set of lower triangular matrices with positive diagonal entries. For any $S \in \text{Sym}_m^+$, there exists a unique $L \in \mathcal{L}_m^+$ such that $S = LL^\top$ and a diffeomorphism $\mathcal{L}$ between Sym_m^+ and $\mathcal{L}_m^+$. For any $L \in \mathcal{L}_m^+$, let $\lfloor L \rfloor$ be the strictly lower triangular part and $\mathbb{D}(L)$ the diagonal part of L . The Log-Cholesky metric is defined as

$$\begin{aligned} d_{\text{LC}}^2(L_1, L_2) &= d_F^2(\lfloor L_1 \rfloor, \lfloor L_2 \rfloor) + d_F^2(\log(\mathbb{D}(L_1)), \log(\mathbb{D}(L_2))) \\ &= d_F^2(\lfloor L_1 \rfloor + \log(\mathbb{D}(L_1)), \lfloor L_2 \rfloor + \log(\mathbb{D}(L_2))). \end{aligned}$$

Note that the space $(\mathcal{L}_m^+, d_{\text{LC}})$ is a geodesic metric space (Lin, 2019), and consider it as the space of SPD matrices under the Log-Cholesky metric. One can then verify Assumption 4.1 with $\Psi(L) = \lfloor L \rfloor + \log(\mathbb{D}(L))$, $\mathcal{M} = \mathcal{L}_m^+$.

S.7.4 Networks

Let $\mathcal{L}$ be the space of networks represented by graph Laplacians with the Frobenius metric d_F . Applying almost the same arguments for the space of SPD matrices with the Frobenius metric, one can verify Assumption 4.1 with $\Psi = \text{id}$. Likewise, one can verify Assumption 4.1 with $\Psi(A) = A^p$ for the space of networks represented by graph Laplacians with the power metric.

S.7.5 Univariate probability distributions

Let $\mathcal{W}$ be the space of univariate probability distributions on a closed interval $\mathcal{I} \subset \mathbb{R}$ with finite second moments equipped with the Wasserstein metric $d_{\mathcal{W}}$ defined as $d_{\mathcal{W}}(\mu, \nu) = d_{L^2}(F_{\mu}^{-1}(\cdot), F_{\nu}^{-1}(\cdot))$ where $F_{\mu}^{-1}(\cdot)$ is the quantile function of $\mu \in \mathcal{W}$. It is easy to verify Assumption 4.1 with $\Psi(\mu) = F_{\mu}^{-1}(\cdot)$.

S.8 Proofs for Section 4

S.8.1 Proof of Theorem 4.1

Recall that $\{e_j\}_{j \geq 1}$ is an orthonormal basis of $\mathcal{H}$. We will show

$$\sqrt{nh}(\hat{\nu}_0^{\Psi} - \mu_{0,\oplus}^{\Psi}, \hat{\nu}_1^{\Psi} - \mu_{1,\oplus}^{\Psi}) \rightsquigarrow (\mathbb{G}_0, \mathbb{G}_1).$$

From Theorem 1.8.4 in [van der Vaart and Wellner \(2023\)](#), it suffices to verify the following conditions.

(i) Finite-dimensional weak convergence: For all $g_0, g_1 \in \mathcal{H} \setminus \{0\}$,

$$\sqrt{nh}(\langle \hat{\nu}_0^{\Psi} - \mu_{0,\oplus}^{\Psi}, g_0 \rangle_{\mathcal{H}}, \langle \hat{\nu}_1^{\Psi} - \mu_{1,\oplus}^{\Psi}, g_1 \rangle_{\mathcal{H}}) \rightsquigarrow (\langle \mathbb{G}_0, g_0 \rangle_{\mathcal{H}}, \langle \mathbb{G}_1, g_1 \rangle_{\mathcal{H}}).$$

(ii) $\sqrt{nh}(\hat{\nu}_0^{\Psi} - \mu_{0,\oplus}^{\Psi})$ and $\sqrt{nh}(\hat{\nu}_1^{\Psi} - \mu_{1,\oplus}^{\Psi})$ are asymptotically finite dimensional.

Note that Condition (i) characterizes the asymptotic distribution of $\sqrt{nh}(\hat{\nu}_0^{\Psi} - \mu_{0,\oplus}^{\Psi}, \hat{\nu}_1^{\Psi} - \mu_{1,\oplus}^{\Psi})$, and Conditions (i) and (ii) imply asymptotic tightness of $\sqrt{nh}(\hat{\nu}_0^{\Psi} - \mu_{0,\oplus}^{\Psi}, \hat{\nu}_1^{\Psi} - \mu_{1,\oplus}^{\Psi})$ (Lemmas 1.8.1 and 1.8.3, [van der Vaart and Wellner, 2023](#)).

Verification of Condition (i): Observe that

$$\begin{aligned} & \sqrt{nh} \langle \hat{\nu}_t^{\Psi} - \mu_{t,\oplus}^{\Psi}, g \rangle_{\mathcal{H}} \\ &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \langle \Psi(Y_i), g \rangle_{\mathcal{H}} - \sqrt{nh} \langle m_t^{\Psi}(c), g \rangle_{\mathcal{H}} \end{aligned}$$

$$\begin{aligned}
&= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} + \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \langle m_t^\Psi(R_i) - m_t^\Psi(c), g \rangle_{\mathcal{H}} \\
&\quad + \sqrt{nh} \langle m_t^\Psi(c), g \rangle_{\mathcal{H}} \underbrace{\left(\frac{1}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) - 1 \right)}_{=0} \\
&= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} + \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \langle m_t^\Psi(R_i) - m_t^\Psi(c), g \rangle_{\mathcal{H}} \\
&=: A_{t,n}(g) + B_{t,n}(g). \tag{S7}
\end{aligned}$$

For $B_{t,n}(g)$, we have

$$\begin{aligned}
B_{t,n}(g) &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \left\{ \partial m_t^\Psi(c; g)(R_i - c) + \frac{1}{2} \partial^2 m_t^\Psi(\tilde{c}_i; g)(R_i - c)^2 \right\} \\
&= \partial m_t^\Psi(c; g) \underbrace{\frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h)(R_i - c)}_{=0} + \frac{\sqrt{nh}}{2n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \partial^2 m_t^\Psi(\tilde{c}_i; g)(R_i - c)^2 \\
&= \frac{\sqrt{nh}}{2n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \partial^2 m_t^\Psi(\tilde{c}_i; g)(R_i - c)^2,
\end{aligned}$$

where $\tilde{c}_i$ lies between c and R_i . Then we have

$$\begin{aligned}
|B_{t,n}(g)| &\leq \sup_{r \in \mathcal{I}} |\partial^2 m_t^\Psi(r; g)| \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n |\hat{s}_t(c; R_i, h)| |R_i - c|^2 \\
&\leq \|g\|_{\mathcal{H}} \left(\sup_{r \in \mathcal{I}} \sum_{j \geq 1} |\partial^2 m_t^\Psi(r; e_j)|^2 \right)^{1/2} \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n |\hat{s}_t(c; R_i, h)| |R_i - c|^2 \\
&\lesssim O(h^2 \sqrt{nh}) \underbrace{\frac{1}{n_t} \sum_{i=1}^n |\hat{s}_t(c; R_i, h)|}_{=O_p(1)} = O_p(\sqrt{nh^5}) = o_p(1). \tag{S8}
\end{aligned}$$

For $A_{t,n}(g)$, observe that

$$\begin{aligned}
A_{t,n}(g) &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} \\
&= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n s_t(c; R_i, h) \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} + \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \{\hat{s}_t(c; R_i, h) - s_t(c; R_i, h)\} \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} \\
&=: A_{t,n,1}(g) + A_{t,n,2}(g). \tag{S9}
\end{aligned}$$

For $A_{t,n,2}(g)$, we have

$$\begin{aligned}
\mathbb{E}[|A_{t,n,2}(g)|^2 | R_1, \dots, R_n] &\lesssim \|g\|_{\mathcal{H}}^2 \frac{nh}{n_t^2} \sum_{i=1}^n \{\hat{s}_t(c; R_i, h) - s_t(c; R_i, h)\}^2 \\
&\quad \times \mathbb{E}[\|\varepsilon_{t,i}\|_{\mathcal{H}}^2 | R_i] \\
&\lesssim \sup_{r \in \mathcal{I}} \mathbb{E}[\|\varepsilon_{t,i}\|_{\mathcal{H}}^2 | R_i = r] \\
&\quad \times \underbrace{\frac{nh}{n_t^2} \sum_{i=1}^n \{\hat{s}_t(c; R_i, h) - s_t(c; R_i, h)\}^2}_{=o_p(1)} = o_p(1). \tag{S10}
\end{aligned}$$

This yields

$$A_{t,n,2}(g) = o_p(1). \tag{S11}$$

Combining (S7)–(S9), (S11), one can see

$$\begin{aligned}
\sqrt{nh} \langle \hat{\nu}_t^\Psi - \mu_{t,\oplus}^\Psi, g \rangle_{\mathcal{H}} &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n s_t(c; R_i, h) \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} + o_p(1) \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left\{ \frac{\kappa_{t,2} - \kappa_{t,1}(R_i - c)/h}{(\kappa_{t,2}\kappa_{t,0} - \kappa_{t,1}^2)f_R(c)} \right\} \\
&\quad \times K_t \left(\frac{R_i - c}{h} \right) \langle \varepsilon_{t,i}, g \rangle_{\mathcal{H}} + o_p(1).
\end{aligned}$$

Applying the standard asymptotic theory for local linear estimators with real-valued outcomes, as developed in Theorems 3.1 and 6.1 of [Fan and Gijbels \(1996\)](#), to the preceding linear representation verifies Condition (i).

Verification of Condition (ii): Applying the same argument to obtain (S11), we have

$$\begin{aligned}
\sqrt{nh}(\hat{\nu}_t^\Psi - \mu_{t,\oplus}^\Psi) &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \varepsilon_{t,i} + \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \hat{s}_t(c; R_i, h) \{m_t^\Psi(R_i) - m_t^\Psi(c)\} \\
&=: A_{t,n} + B_{t,n}.
\end{aligned}$$

By an argument analogous to obtaining (S8) and (S10), we have $\|B_{t,n}\|_{\mathcal{H}} = O_p(\sqrt{nh^5}) = o_p(1)$ and $\|A_{t,n} - \tilde{A}_{t,n}\|_{\mathcal{H}} = o_p(1)$ where

$$\tilde{A}_{t,n} = \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left\{ \frac{\kappa_{t,2} - \kappa_{t,1}(R_i - c)/h}{(\kappa_{t,2}\kappa_{t,0} - \kappa_{t,1}^2)f_R(c)} \right\} K_t \left(\frac{R_i - c}{h} \right) \varepsilon_{t,i} =: \frac{1}{\sqrt{nh}} \sum_{i=1}^n W_{n,i} \varepsilon_{t,i}.$$

Then it suffices to verify that $\tilde{A}_{t,n}$ is asymptotically finite dimensional, that is,

$$\lim_{J \rightarrow \infty} \limsup_n \Pr \left(\sum_{j>J} \langle \tilde{A}_{t,n}, e_j \rangle_{\mathcal{H}}^2 > \delta \right) = 0 \quad (\text{S12})$$

for any $\delta > 0$. Observe that

$$\begin{aligned} \mathbb{E} \left[\sum_{j>J} \langle \tilde{A}_{t,n}, e_j \rangle_{\mathcal{H}}^2 \right] &= \sum_{j>J} \mathbb{E} \left[\langle \tilde{A}_{t,n}, e_j \rangle_{\mathcal{H}}^2 \right] = \sum_{j>J} \frac{1}{nh} \sum_{i=1}^n \mathbb{E} [W_{n,i}^2 \langle \varepsilon_{t,i}, e_j \rangle_{\mathcal{H}}^2] \\ &= \sum_{j>J} \frac{1}{h} \mathbb{E} [W_{n,i}^2 \mathbb{E} [\langle \varepsilon_{t,i}, e_j \rangle_{\mathcal{H}}^2 | R_i]] \leq \frac{1}{h} \mathbb{E} [W_{n,i}^2] \sum_{j>J} \sup_{r \in \mathcal{I}} \mathbb{E} [\langle \varepsilon_{t,i}, e_j \rangle_{\mathcal{H}}^2 | R_i = r] \\ &\lesssim \sum_{j>J} \sup_{r \in \mathcal{I}} \mathbb{E} [\langle \varepsilon_{t,i}, e_j \rangle_{\mathcal{H}}^2 | R_i = r]. \end{aligned}$$

Since $\lim_{J \rightarrow \infty} \sum_{j>J} \sup_{r \in \mathcal{I}} \mathbb{E} [\langle \varepsilon_{t,i}, e_j \rangle_{\mathcal{H}}^2 | R_i = r] = 0$, we have

$$\lim_{J \rightarrow \infty} \limsup_n \mathbb{E} \left[\sum_{j>J} \langle \tilde{A}_{t,n}, e_j \rangle_{\mathcal{H}}^2 \right] = 0.$$

Therefore, using Markov's inequality, we obtain (S12).

S.8.2 Proof of Proposition 4.1

Following the characterization of weak convergence in Theorem 1.8.4 of [van der Vaart and Wellner \(2023\)](#), it suffices to verify its conditional analogues for the conditional weak convergence of the multiplier bootstrap process $(\hat{\mathbb{G}}_0, \hat{\mathbb{G}}_1)$.

From the argument in the proof of Theorem 4.1, one can see

$$\sqrt{nh}(\hat{\nu}_t^\Psi - \mu_{t,\oplus}^\Psi) = \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left\{ \frac{\kappa_{t,2} - \kappa_{t,1}(R_i - c)/h}{(\kappa_{t,2}\kappa_{t,0} - \kappa_{t,1}^2)f_R(c)} \right\} K_t \left(\frac{R_i - c}{h} \right) \varepsilon_{t,i} + o_p(1).$$

Further, applying a similar argument, one can also see

$$\begin{aligned} \hat{\mathbb{G}}_t &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \xi_i \hat{s}_t(c; R_i, h) \varepsilon_{t,i} \\ &\quad + \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \xi_i \hat{s}_t(c; R_i, h) \{m_t^\Psi(R_i) - \mu_{t,\oplus}^\Psi\} - \left(\frac{1}{n_t} \sum_{i=1}^n \xi_i \hat{s}_t(c; R_i, h) \right) \sqrt{nh}(\hat{\nu}_t^\Psi - \mu_{t,\oplus}^\Psi) \\ &= \frac{\sqrt{nh}}{n_t} \sum_{i=1}^n \xi_i \hat{s}_t(c; R_i, h) \varepsilon_{t,i} + o_{\mathbb{P}_\xi}(1) \end{aligned}$$

$$= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \xi_i \left\{ \frac{\kappa_{t,2} - \kappa_{t,1}(R_i - c)/h}{(\kappa_{t,2}\kappa_{t,0} - \kappa_{t,1}^2)f_R(c)} \right\} K_t \left(\frac{R_i - c}{h} \right) \varepsilon_{t,i} + o_{P_\xi}(1). \quad (\text{S13})$$

Combining (S13) and the conditional multiplier central limit theorem for triangular arrays (Theorem 2.1, Pauly, 2011), we have the conditional finite-dimensional weak convergence: For all $g_0, g_1 \in \mathcal{H} \setminus \{0\}$, $(\langle \hat{\mathbb{G}}_0, g_0 \rangle_{\mathcal{H}}, \langle \hat{\mathbb{G}}_1, g_1 \rangle_{\mathcal{H}}) \xrightarrow{P_\xi} (\langle \mathbb{G}_0, g_0 \rangle_{\mathcal{H}}, \langle \mathbb{G}_1, g_1 \rangle_{\mathcal{H}})$. To obtain the conclusion, it remains to show the conditional asymptotic tightness of $(\hat{\mathbb{G}}_0, \hat{\mathbb{G}}_1)$, that is, to prove that $\hat{\mathbb{G}}_0$ and $\hat{\mathbb{G}}_1$ are conditionally asymptotically finite-dimensional: For any $\delta, \eta > 0$,

$$\lim_{J \rightarrow \infty} \limsup_n \Pr \left(\Pr_\xi \left(\sum_{j>J} \langle \hat{\mathbb{G}}_t, e_j \rangle_{\mathcal{H}}^2 > \delta \right) > \eta \right) = 0, \quad (\text{S14})$$

where $\Pr_\xi$ denotes the probability with respect to the bootstrap multipliers $\{\xi_i\}_{i=1}^n$. Let E_ξ denote the expectation with respect to $\{\xi_i\}_{i=1}^n$. Observe that

$$\begin{aligned} E_\xi \left[\sum_{j>J} \langle \hat{\mathbb{G}}_t, e_j \rangle_{\mathcal{H}}^2 \right] &= \frac{nh}{n_t^2} \sum_{i=1}^n \hat{s}_t^2(c; R_i, h) \sum_{j>J} \langle \Psi(Y_i) - \hat{\nu}_t^\Psi, e_j \rangle_{\mathcal{H}}^2 \\ &\leq 2 \frac{nh}{n_t^2} \sum_{i=1}^n \hat{s}_t^2(c; R_i, h) \sum_{j>J} \langle \Psi(Y_i) - m_t^\Psi(c), e_j \rangle_{\mathcal{H}}^2 \\ &\quad + 2 \frac{nh}{n_t^2} \sum_{i=1}^n \hat{s}_t^2(c; R_i, h) \sum_{j>J} \langle m_t^\Psi(c) - \hat{\nu}_t^\Psi, e_j \rangle_{\mathcal{H}}^2 =: 2A_{n,J}^* + 2B_{n,J}^*. \end{aligned}$$

For $B_{n,J}^*$, we have

$$B_{n,J}^* \leq \|m_t^\Psi(c) - \hat{\nu}_t^\Psi\|_{\mathcal{H}}^2 \frac{nh}{n_t^2} \sum_{i=1}^n \hat{s}_t^2(c; R_i, h) = o_p(1) \times O_p(1) = o_p(1). \quad (\text{S15})$$

For $A_{n,J}^*$, we have

$$\begin{aligned} A_{n,J}^* &= \frac{nh}{n_t^2} \sum_{i=1}^n s_t^2(c; R_i, h) \sum_{j>J} \langle \Psi(Y_i) - m_t^\Psi(c), e_j \rangle_{\mathcal{H}}^2 + o_p(1) \\ &\lesssim \sum_{j>J} \sup_{r \in \mathcal{I}} E \left[\langle \Psi(Y_i) - m_t^\Psi(c), e_j \rangle_{\mathcal{H}}^2 \mid R_i = r \right] + o_p(1) \\ &= \sum_{j>J} \sup_{r \in \mathcal{I}} E[\langle \varepsilon_{t,1}, e_j \rangle_{\mathcal{H}}^2 \mid R_1 = r] + o_p(1). \end{aligned} \quad (\text{S16})$$

Since $\lim_{J \rightarrow \infty} \sum_{j>J} \sup_{r \in \mathcal{I}} E[\langle \varepsilon_{t,1}, e_j \rangle_{\mathcal{H}}^2 \mid R_1 = r] = 0$, from Markov's inequality together with (S15) and (S16), we obtain (S14).

S.8.3 Proof of Corollary 4.1

By applying the (conditional) continuous mapping theorem (Theorems 7.7 and 10.8, [Kosorok, 2008](#)), we have $nh\|\hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi\|_{\mathcal{H}}^2 \rightsquigarrow \|\mathbb{G}_1 - \mathbb{G}_0\|_{\mathcal{H}}^2$ and $\hat{T}_n \overset{P_\xi}{\rightsquigarrow} \|\mathbb{G}_1 - \mathbb{G}_0\|_{\mathcal{H}}^2$. Then, the conclusion follows by applying almost the same argument in the proof of Theorem 3.1 in [Pauly \(2011\)](#).

S.8.4 Proof of Corollary 4.2

By applying the (conditional) continuous mapping theorem, we have

$$\begin{aligned} \sqrt{nh}\{\|\hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi\|_{\mathcal{H}}^2 - d^2(\mu_{1,\oplus}, \mu_{0,\oplus})\} &\rightsquigarrow 2\langle \mu_{1,\oplus}^\Psi - \mu_{0,\oplus}^\Psi, \mathbb{G}_1 - \mathbb{G}_0 \rangle_{\mathcal{H}}, \\ 2\langle \hat{\nu}_1^\Psi - \hat{\nu}_0^\Psi, \hat{\mathbb{G}}_1 - \hat{\mathbb{G}}_0 \rangle_{\mathcal{H}} &\overset{P_\xi}{\rightsquigarrow} 2\langle \mu_{1,\oplus}^\Psi - \mu_{0,\oplus}^\Psi, \mathbb{G}_1 - \mathbb{G}_0 \rangle_{\mathcal{H}}. \end{aligned}$$

The conclusion follows by applying almost the same argument in the proof of Theorem 3.1 in [Pauly \(2011\)](#).

S.9 Proofs for Section 7

S.9.1 Proof of Theorem 7.1

(Step 1) In this step, we show

$$\mu_{1,\oplus}^\dagger = E_\oplus[Y(T(1))|R = c] = \lim_{r \downarrow c} E_\oplus[Y|R = r], \quad (\text{S17})$$

$$\mu_{0,\oplus}^\dagger = E_\oplus[Y(T(0))|R = c] = \lim_{r \uparrow c} E_\oplus[Y|R = r], \quad (\text{S18})$$

$$E[T(1)|R = c] - E[T(0)|R = c] = \lim_{r \downarrow c} E[T|R = r] - \lim_{r \uparrow c} E[T|R = r]. \quad (\text{S19})$$

Proof of (S17): Observe that for $r \geq c$,

$$\begin{aligned} E_\oplus[Y|R = r] &= E_\oplus[Y|Z = 1, R = r] \\ &= E_\oplus[Y(T(1))|Z = 1, R = r] \end{aligned}$$

$$= \mathbb{E}_\oplus[Y(T(1))|R = r].$$

Then we have $\lim_{r \downarrow c} \mathbb{E}_\oplus[Y|R = r] = \lim_{r \downarrow c} \mathbb{E}_\oplus[Y(T(1))|R = r] = \mathbb{E}_\oplus[Y(T(1))|R = c]$.

Likewise, one can show (S18).

Proof of (S19) Observe that

$$\begin{aligned} \lim_{r \downarrow c} \mathbb{E}[T|R = r] - \lim_{r \uparrow c} \mathbb{E}[T|R = r] &= \lim_{r \downarrow c} \mathbb{E}[T|Z = 1, R = r] - \lim_{r \uparrow c} \mathbb{E}[T|Z = 0, R = r] \\ &= \lim_{r \downarrow c} \mathbb{E}[T(1)|Z = 1, R = r] - \lim_{r \uparrow c} \mathbb{E}[T(0)|Z = 0, R = r] \\ &= \lim_{r \downarrow c} \mathbb{E}[T(1)|R = r] - \lim_{r \uparrow c} \mathbb{E}[T(0)|R = r] \\ &= \mathbb{E}[T(1)|R = c] - \mathbb{E}[T(0)|R = c]. \end{aligned}$$

(Step 2) Observe that for $z \in \{0, 1\}$,

$$\begin{aligned} \Psi(\mathbb{E}_\oplus[Y(T(z))|R = c]) &= \Psi(\arg \min_{\nu \in \mathcal{M}} \mathbb{E}[d^2(\nu, Y(T(z)))|R = c]) \\ &= \Psi(\arg \min_{\nu \in \mathcal{M}} \mathbb{E}[d_{\mathcal{H}}^2(\Psi(\nu), \Psi(Y(T(z))))|R = c]) \\ &= \arg \min_{\omega \in \Psi(\mathcal{M})} \mathbb{E}[d_{\mathcal{H}}^2(\omega, \Psi(Y(T(z))))|R = c] \\ &= \mathbb{E}[\Psi(Y(T(z)))|R = c], \end{aligned}$$

where we used the closedness and convexity of the image space $\Psi(\mathcal{M})$ for the third and fourth equalities. Likewise, one can see

$$\begin{aligned} \Psi(\mathbb{E}_\oplus[Y(T(z))|T(1) > T(0), R = c]) &= \mathbb{E}[\Psi(Y(T(z)))|T(1) > T(0), R = c], \\ \Psi(\mathbb{E}_\oplus[Y(T(z))|T(1) = T(0), R = c]) &= \mathbb{E}[\Psi(Y(T(z)))|T(1) = T(0), R = c]. \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{E}[\Psi(Y(T(z)))|R = c] &= \mathbb{E}[\Psi(Y(T(z)))|T(1) > T(0), R = c] \times \Pr(T(1) > T(0)|R = c) \\ &\quad + \mathbb{E}[\Psi(Y(T(z)))|T(1) = T(0), R = c] \times \Pr(T(1) = T(0)|R = c). \end{aligned}$$

Combining these identities with the fact that $\Pr(T(1) = T(0)|R = c) = 1 - \Pr(T(1) > T(0)|R = c)$ (by Assumption 7.1(i)), we have

$$\begin{aligned}
\Psi(\mu_{z,\oplus}) &= \Psi(\mathbb{E}_{\oplus}[Y(T(z))|T(1) > T(0), R = c]) \\
&= \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c]) \\
&\quad + \frac{\Psi(\mathbb{E}_{\oplus}[Y(T(z))|R = c]) - \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c])}{\Pr(T(1) > T(0)|R = c)} \\
&= \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c]) \\
&\quad + \frac{\Psi(\mu_{z,\oplus}^{\dagger}) - \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c])}{\Pr(T(1) > T(0)|R = c)}. \tag{S20}
\end{aligned}$$

Therefore,

$$\tau_{\text{FRDD}} = \Psi(\mu_{1,\oplus}) - \Psi(\mu_{0,\oplus}) = \frac{\Psi(\mu_{1,\oplus}^{\dagger}) - \Psi(\mu_{0,\oplus}^{\dagger})}{\Pr(T(1) > T(0)|R = c)} = \frac{\Psi(\mu_{1,\oplus}^{\dagger}) - \Psi(\mu_{0,\oplus}^{\dagger})}{\mathbb{E}[T(1)|R = c] - \mathbb{E}[T(0)|R = c]}.$$

Combining this with (S17)–(S19) completes the proof.

S.9.2 Proof of Theorem 7.2

Define

$$\tau_{\text{FRDD}}^* = \frac{\Psi(\nu_{1,\oplus}) - \Psi(\nu_{0,\oplus})}{m_1 - m_0}.$$

Observe that

$$\begin{aligned}
d_{\mathcal{H}}(\tau_{\text{FRDD}}, \hat{\tau}_{\text{FRDD}}) &\leq d_{\mathcal{H}}(\tau_{\text{FRDD}}, \tau_{\text{FRDD}}^*) + d_{\mathcal{H}}(\tau_{\text{FRDD}}^*, \hat{\tau}_{\text{FRDD}}) \\
&\leq \frac{\left\{ d_{\mathcal{H}}(\Psi(\mu_{1,\oplus}^{\dagger}), \Psi(\nu_{1,\oplus})) + d_{\mathcal{H}}(\Psi(\mu_{0,\oplus}^{\dagger}), \Psi(\nu_{0,\oplus})) \right\}}{|m_1 - m_0|} \\
&\quad + \frac{\left\{ d_{\mathcal{H}}(\Psi(\nu_{1,\oplus}), \Psi(\hat{\nu}_{1,\oplus})) + d_{\mathcal{H}}(\Psi(\nu_{0,\oplus}), \Psi(\hat{\nu}_{0,\oplus})) \right\}}{|\hat{m}_1 - \hat{m}_0|} \\
&\quad + \left| \frac{1}{\hat{m}_1 - \hat{m}_0} - \frac{1}{m_1 - m_0} \right| d_{\mathcal{H}}(\Psi(\nu_{0,\oplus}), \Psi(\nu_{1,\oplus})) \\
&= \frac{\left\{ d(\mu_{1,\oplus}^{\dagger}, \nu_{1,\oplus}) + d(\mu_{0,\oplus}^{\dagger}, \nu_{0,\oplus}) \right\}}{|m_1 - m_0|} \\
&\quad + \frac{\left\{ d(\nu_{1,\oplus}, \hat{\nu}_{1,\oplus}) + d(\nu_{0,\oplus}, \hat{\nu}_{0,\oplus}) \right\}}{|\hat{m}_1 - \hat{m}_0|} + \left| \frac{1}{\hat{m}_1 - \hat{m}_0} - \frac{1}{m_1 - m_0} \right| d(\nu_{0,\oplus}, \nu_{1,\oplus}). \tag{S21}
\end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{1}{\hat{m}_1 - \hat{m}_0} - \frac{1}{m_1 - m_0} &= \frac{(m_1 - \hat{m}_1) - (m_0 - \hat{m}_0)}{(\hat{m}_1 - \hat{m}_0)(m_1 - m_0)} \\ &= \frac{O_p(r_n)}{(m_1 - m_0 + o_p(1))(m_1 - m_0)} = O_p(r_n). \end{aligned} \quad (\text{S22})$$

Applying a similar argument as in the proof of Theorem 3.2,

$$d(\mu_{z,\oplus}^\dagger, \nu_{z,\oplus}) = O(h_z^{2/(\beta_1-1)}), \quad d(\hat{\nu}_{z,\oplus}, \nu_{z,\oplus}) = O_p((nh_z)^{-\frac{1}{2(\beta_2-1)}}), \quad z \in \{0, 1\}. \quad (\text{S23})$$

Combining (S21), (S22), and (S23), we obtain

$$d_{\mathcal{H}}(\tau_{\text{FRDD}}, \hat{\tau}_{\text{FRDD}}) = O(h_{\max}^{2/(\beta_1-1)}) + O_p(r_n + (nh_{\min})^{-\frac{1}{2(\beta_2-1)}}).$$

S.10 Proofs for Section S.3

S.10.1 Proof of Proposition S1

Under Assumption S1(i),

$$\mu_{\oplus} = \mathbb{E}_{\oplus}[Y(T)|T(1) = T(0) = 1, R = c] = \mathbb{E}_{\oplus}[Y(1)|T(0) = 1, R = c].$$

For $r < c$,

$$\begin{aligned} \mathbb{E}_{\oplus}[Y(1)|T(0) = 1, R = r] &= \mathbb{E}_{\oplus}[Y(T(0))|T(0) = 1, R = r] \\ &= \mathbb{E}_{\oplus}[Y(T(0))|T(0) = 1, Z = 0, R = r] \\ &= \mathbb{E}_{\oplus}[Y|T = 1, Z = 0, R = r]. \end{aligned}$$

Hence,

$$\begin{aligned} \lim_{r \uparrow c} \mathbb{E}_{\oplus}[Y|T = 1, Z = 0, R = r] &= \lim_{r \uparrow c} \mathbb{E}_{\oplus}[Y(1)|T(0) = 1, R = r] \\ &= \mathbb{E}_{\oplus}[Y(1)|T(0) = 1, R = c] \\ &= \mu_{\oplus}. \end{aligned}$$

Likewise, under Assumption S1(ii), we can show

$$\lim_{r \downarrow c} \mathbb{E}_{\oplus}[Y|T = 0, Z = 1, R = r] = \mathbb{E}_{\oplus}[Y(0)|T(1) = 0, R = c] = \mu_{\oplus}.$$

S.10.2 Proof of Proposition S2

Observe that $d_{\mathcal{G}}(\tau_{\text{GFRD}}, \hat{\tau}_{\text{GFRD}}) \leq C\{d(\mu_{0,\oplus}, \hat{\mu}_{0,\oplus}) + d(\mu_{1,\oplus}, \hat{\mu}_{1,\oplus})\}$ and

$$\begin{aligned}
d(\mu_{z,\oplus}, \hat{\mu}_{z,\oplus}) &= d_{\mathcal{H}}\left(\Psi(\mu_{\oplus}) + \frac{\Psi(\nu_{z,\oplus}) - \Psi(\mu_{\oplus})}{m_1 - m_0}, \Psi(\hat{\mu}_{\oplus}) + \frac{\Psi(\hat{\nu}_{z,\oplus}) - \Psi(\hat{\mu}_{\oplus})}{\hat{m}_1 - \hat{m}_0}\right) \\
&\leq \left(1 + \frac{1}{|\hat{m}_1 - \hat{m}_0|}\right) d_{\mathcal{H}}(\Psi(\mu_{\oplus}), \Psi(\hat{\mu}_{\oplus})) + \frac{1}{|\hat{m}_1 - \hat{m}_0|} d_{\mathcal{H}}(\Psi(\nu_{z,\oplus}), \Psi(\hat{\nu}_{z,\oplus})) \\
&\quad + \left|\frac{1}{\hat{m}_1 - \hat{m}_0} - \frac{1}{m_1 - m_0}\right| d_{\mathcal{H}}(\Psi(\nu_{z,\oplus}), \Psi(\mu_{\oplus})) \\
&\leq \left(1 + \frac{1}{|\hat{m}_1 - \hat{m}_0|}\right) d(\mu_{\oplus}, \hat{\mu}_{\oplus}) + \frac{1}{|\hat{m}_1 - \hat{m}_0|} d(\nu_{z,\oplus}, \hat{\nu}_{z,\oplus}) \\
&\quad + \left|\frac{1}{\hat{m}_1 - \hat{m}_0} - \frac{1}{m_1 - m_0}\right| d(\nu_{z,\oplus}, \mu_{\oplus}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
d_{\mathcal{G}}(\tau_{\text{GFRD}}, \hat{\tau}_{\text{GFRD}}) &\leq C \left(1 + \frac{1}{|\hat{m}_1 - \hat{m}_0|}\right) \{2d(\mu_{\oplus}, \hat{\mu}_{\oplus}) + d(\nu_{0,\oplus}, \hat{\nu}_{0,\oplus}) + d(\nu_{1,\oplus}, \hat{\nu}_{1,\oplus})\} \\
&\quad + C \left|\frac{1}{\hat{m}_1 - \hat{m}_0} - \frac{1}{m_1 - m_0}\right| \{d(\nu_{1,\oplus}, \mu_{\oplus}) + d(\nu_{0,\oplus}, \mu_{\oplus})\}.
\end{aligned}$$

Applying a similar argument as in the proof of Theorem 7.2 yields the desired result.

S.11 Proofs for Section S.4

S.11.1 Proof of Theorem S1

Applying the same argument as used to obtain (S17) and (S18),

$$\nu_{1,\omega}^{\#} = \mathbb{E}[\text{Log}_{\omega}(Y(T(1)))|R = c] = \lim_{r \downarrow c} \mathbb{E}[\text{Log}_{\omega}(Y)|R = r], \quad (\text{S24})$$

$$\nu_{0,\omega}^{\#} = \mathbb{E}[\text{Log}_{\omega}(Y(T(0)))|R = c] = \lim_{r \uparrow c} \mathbb{E}[\text{Log}_{\omega}(Y)|R = r]. \quad (\text{S25})$$

Observe

$$\begin{aligned}
\nu_{z,\omega}^{\#} &= \mathbb{E}[\text{Log}_{\omega}(Y(T(z)))|T(1) > T(0), R = c] \Pr(T(1) > T(0)|R = c) \\
&\quad + \mathbb{E}[\text{Log}_{\omega}(Y(T(z)))|T(1) = T(0), R = c] \Pr(T(1) = T(0)|R = c)
\end{aligned}$$

and

$$\begin{aligned}
\mu_{z,\omega}^\# &= \mathbb{E}[\text{Log}_\omega(Y(T(z)))|T(1) > T(0), R = c] \\
&= \mathbb{E}[\text{Log}_\omega(Y(T(z)))|T(1) = T(0), R = c] \\
&\quad + \frac{\nu_{z,\omega}^\# - \mathbb{E}[\text{Log}_\omega(Y(T(z)))|T(1) = T(0), R = c]}{\Pr(T(1) > T(0)|R = c)}.
\end{aligned} \tag{S26}$$

This yields

$$\begin{aligned}
\tau_{\text{FRDD}}^\# &= \mu_{1,\omega}^\# - \mu_{0,\omega}^\# \\
&= \frac{\nu_{1,\omega}^\# - \nu_{0,\omega}^\#}{\Pr(T(1) > T(0)|R = c)} \\
&= \frac{\nu_{1,\omega}^\# - \nu_{0,\omega}^\#}{\mathbb{E}[T(1)|R = c] - \mathbb{E}[T(0)|R = c]}.
\end{aligned}$$

Combining this with (S19), (S24), and (S25) leads to the desired result.

S.11.2 Proof of Theorem S2

The proof is omitted as it is similar to that of Theorem 7.2.

S.12 Proofs for Section S.5

S.12.1 Proof of Proposition S3

The proof is omitted as it is similar to that of Proposition S1.

S.12.2 Proof of Proposition S4

Observe that $d_{\mathcal{G}}(\tau_{\text{GFRD}}^\#, \hat{\tau}_{\text{GFRD}}^\#) \leq C\{d(\mu_{0,\omega,\oplus}^\#, \hat{\mu}_{0,\omega,\oplus}^\#) + d(\mu_{1,\omega,\oplus}^\#, \hat{\mu}_{1,\omega,\oplus}^\#)\}$ and

$$\begin{aligned}
&d(\mu_{z,\omega,\oplus}^\#, \hat{\mu}_{z,\omega,\oplus}^\#) \\
&= d\left(\text{Exp}_\omega\left(\nu_\omega^\# + \frac{\nu_{z,\omega}^\# - \nu_\omega^\#}{m_1 - m_0}\right), \text{Exp}_\omega\left(\hat{\nu}_\omega^\# + \frac{\hat{\nu}_{z,\omega}^\# - \hat{\nu}_\omega^\#}{\hat{m}_1 - \hat{m}_0}\right)\right)
\end{aligned}$$

$$\begin{aligned}
&\leq C_\omega d_E \left(\nu_\omega^\# + \frac{\nu_{z,\omega}^\# - \nu_\omega^\#}{m_1 - m_0}, \hat{\nu}_\omega^\# + \frac{\hat{\nu}_{z,\omega}^\# - \hat{\nu}_\omega^\#}{\hat{m}_1 - \hat{m}_0} \right) \\
&\leq C_\omega \left\{ \left(1 + \frac{1}{|\hat{m}_1 - \hat{m}_0|} \right) d_E(\nu_\omega^\#, \hat{\nu}_\omega^\#) + \frac{1}{|\hat{m}_1 - \hat{m}_0|} d_E(\nu_{z,\omega}^\#, \hat{\nu}_{z,\omega}^\#) \right\} \\
&\quad + C_\omega \left| \frac{1}{m_1 - m_0} - \frac{1}{\hat{m}_1 - \hat{m}_0} \right| d_E(\nu_{z,\omega}^\#, \nu_\omega^\#), \quad z \in \{0, 1\}.
\end{aligned}$$

Then

$$\begin{aligned}
&d_{\mathcal{G}}(\tau_{\text{GFRD}}^\#, \hat{\tau}_{\text{GFRD}}^\#) \\
&\leq CC_\omega \left(1 + \frac{1}{|\hat{m}_1 - \hat{m}_0|} \right) \{ 2d_E(\nu_\omega^\#, \hat{\nu}_\omega^\#) + d_E(\nu_{0,\omega}^\#, \hat{\nu}_{0,\omega}^\#) + d_E(\nu_{1,\omega}^\#, \hat{\nu}_{1,\omega}^\#) \} \\
&\quad + CC_\omega \left| \frac{1}{m_1 - m_0} - \frac{1}{\hat{m}_1 - \hat{m}_0} \right| \{ d_E(\nu_{1,\omega}^\#, \nu_\omega^\#) + d_E(\nu_{0,\omega}^\#, \nu_\omega^\#) \}.
\end{aligned}$$

Applying a similar argument in the proof of Theorem 7.2 then implies the desired result.

S.13 Additional simulation with unequal curvature

In the simulation study presented in the main text, the regression functions on the two sides of the cutoff have identical curvature. Under the null hypothesis, this implies that the leading-order local linear bias terms cancel. To assess the finite-sample behavior of the proposed inference procedure when this property does not hold, we consider an additional simulation in which the regression functions have different curvature on the two sides of the cutoff.

The simulation setup is identical to that in Section 5.2 of the main text except for the conditional mean edge weights. Specifically, for each edge (i, j) , we generate

$$w_{ij}(R) = \begin{cases} 1 + 0.5R^2 + \epsilon_{ij}, & R < 0, \\ 1 + R^2 + \Delta\tau_{ij} + \epsilon_{ij}, & R \geq 0, \end{cases}$$

where $\epsilon_{ij} \sim \text{Unif}(0, 0.1)$ independently across edges and τ_{ij} is defined as in Section 5.2.

The remainder of the simulation setup, including the cross-validation bandwidth selection

procedure, is unchanged. Throughout this experiment, we set $\Delta = 0$, so that the null hypothesis holds. Under the null, the conditional mean function is continuous at the cutoff with equal first derivatives from the left and right, but different second derivatives. Consequently, unlike the simulation considered in the main text, the leading-order local linear bias terms do not cancel.

Table S1 reports the empirical rejection probabilities of the proposed bootstrap test based on 500 Monte Carlo replications.

Table S1: Empirical rejection probabilities under the null hypothesis for the additional simulation with unequal curvature across the cutoff. The nominal significance level is 0.05.

Sample size	100	200	500	1000
Empirical rejection probability	0.026	0.032	0.052	0.058

The empirical rejection probabilities are close to the nominal significance level for moderate and large sample sizes ($n = 500$ and 1000), while the procedure is somewhat conservative for smaller sample sizes. Overall, this additional simulation provides further empirical evidence that the proposed bootstrap inference procedure maintains accurate empirical size under a setting where the leading-order bias terms no longer cancel.