

# REGRESSION DISCONTINUITY DESIGNS FOR FUNCTIONAL DATA AND RANDOM OBJECTS IN GEODESIC SPACES

DAISUKE KURISU\*, YIDONG ZHOU\*, TAISUKE OTSU, AND HANS-GEORG MÜLLER

**ABSTRACT.** Regression discontinuity designs, quasi-experimental designs for observational studies to estimate causal effects of an intervention or treatment at a cutoff point, have been widely applied in various fields of data science, such as economics, education, environmental studies, epidemiology, and political science. We propose an extension of regression discontinuity designs, where we extend the customary scenario of scalar responses and aim at causal inference for a general class of complex non-Euclidean outcomes. Specifically, the outcomes considered extend beyond scalars or vectors and include networks, compositional data and functional data, as well as other types of random objects situated in geodesic metric spaces that may be subject to causal effects. For this extension a major challenge is how to quantify a treatment effect, 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 point as a geodesic from the local Fréchet mean of the untreated outcome to the treated outcome, which reduces to the ordinary average treatment effect and thus to the well-established approach for regression discontinuity designs in the special case of scalar or vector outcomes. Our estimation method is based on local Fréchet regression, a regression method for non-Euclidean responses that corresponds to local linear regression for the special case of scalar responses. We equip local Fréchet regression with a novel bandwidth selection that is specifically aimed at regression discontinuity designs. The proposed bandwidth selector is shown to be competitive even in the well-established case of regression discontinuity designs for scalar outcomes. The proposed approach is supported by theory and we establish the convergence rate of the geodesic regression discontinuity design estimator. We demonstrate the relevance of this general approach for specific practical applications, where one seeks to assess causal effects in the framework of policy interventions and natural experiments that give rise to regression discontinuity designs. These applications include an assessment of the changes caused in daily CO concentration curves by the introduction of a metro system in Taipei and in UK voting behavior as quantified by compositional data after wins by the conservative party. We also develop an extension for fuzzy regression discontinuity designs with non-Euclidean outcomes, broadening the scope of causal inference to settings that allow for imperfect compliance with the assignment rule.

## 1. INTRODUCTION

In causal inference, discontinuities in regression functions induced by an assignment variable can provide useful information to identify certain causal effects. As a quasi-experimental design, the regression discontinuity design (RDD) has been widely applied in observational studies in several fields, such as economics, education, environmental studies, epidemiology, and political science, to identify causal effects at the discontinuity point. For the conventional RDD with a scalar outcome,

---

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

*Key words and phrases.* causal inference, Fréchet regression, functional data, program evaluation, random object, regression discontinuity design

*MSC2020 subject classifications:* 62D20, 62R20.

The work of D. K. was partially supported by JSPS KAKENHI Grant Numbers 23K12456 and 25K00624, and the work of H.G.M. was partially supported by NSF grant DMS-2310450.

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 proposes an extension of RDD analysis, where we extend the customary scenario of scalar outcomes and aim at causal inference for a general class of complex non-Euclidean outcomes. Specifically, this paper extends the scope of RDD analysis to accommodate not only scalar or vector outcomes but also networks, distributions, compositional data, and functional data, as well as other types of random objects situated 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 point as a geodesic from the local Fréchet mean of the untreated outcome to the treated one, which reduces to the ordinary average treatment effect and thus to the well-established approach for RDDs in the special case of scalar or vector outcomes.

Once we characterize the causal object of interest, it is estimated by local Fréchet regression, a regression method for non-Euclidean responses that corresponds to local linear regression for the special case of scalar responses ([Petersen and Müller, 2019](#)), and we establish the convergence rate of the proposed estimator for the causal effect. To improve practical implementation, we introduce a novel bandwidth selection method tailored to the structure of regression discontinuity problems. Unlike existing cross-validation or MSE-optimal criteria designed for Euclidean outcomes ([Imbens and Kalyanaraman, 2012](#); [Calonico et al., 2014, 2020](#)), our approach exploits the expected smoothness of the regression function away from the cutoff, selecting bandwidths that minimize local discrepancies between left and right Fréchet regression estimates where no discontinuity is expected.

After developing the above basic results for sharp RDDs assuming full compliance with the assignment rule, our methodology is extended to the fuzzy RDD setup allowing imperfect compliance, where for the conventional case of scalar outcomes one can identify the average treatment effect for compliers at the cutoff point by amplifying the discontinuous change in the outcome variable at the cutoff by the reciprocal of the discontinuous change in the probability of treatment at the cutoff ([Hahn et al., 2001](#)). In particular, 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, derive its identification formula, 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 [Kurusu et al. \(2024\)](#) for a general class of geodesic metric spaces. For distributional outcomes, [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 [Kurusu 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 RDD analyses for univariate distributional outcomes. In particular, [Van Dijke \(2025\)](#) focused on the local average quantile treatment effect, which averages across random quantiles, and proposed suitable non-parametric estimators. This approach relies on the fact that for univariate distributions one can employ restricted linear operations in the space of quantile functions and is specific for univariate distributional outcomes. It is not directly extendable to the case of multivariate distributions or other random objects, such as networks, symmetric positive-definite matrices, and compositional data. Indeed, our identification results for outcomes in geodesic metric spaces imply identification of [Van Dijke \(2025\)](#)'s local average quantile treatment effect as a special case of univariate distributional outcomes in the Wasserstein space.

We illustrate the proposed methodology with two real data applications. In the first application, we investigate the impact of the Taipei Metro expansion on daily carbon monoxide pollution curves. In this setting, the outcome is functional and irregularly observed, which renders standard RDD methods designed for scalar outcomes unsuitable. The second application analyzes election outcomes in the United Kingdom, where the outcome takes the form of compositional data that represent the distribution of votes across political parties. In this case, the treatment is determined by the prior victories of the Conservative Party. Both applications illustrate how the proposed methodology provides a robust and interpretable framework for estimating causal effects in settings involving complex structured outcomes.

This paper is organized as follows. Section 2 presents the basic concepts for geodesic metric spaces and motivating examples. In Section 3, we introduce geodesic RDDs for the sharp design case, establish identification of the geodesic local average treatment effect (Section 3.1) and study the asymptotic properties of the estimator using local Fréchet regression (Section 3.2). Section 4 presents the proposed bandwidth selector and assesses its finite sample performance using Monte Carlo simulations. In Section 5, we illustrate the application of GRDD to real-world datasets. Section 6 extends the framework to fuzzy designs with imperfect compliance. Concluding remarks are provided in Section 7.

## 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*.

**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 2.1** (Functional Data). Functional data arise when observations are best represented as curves or functions, typically over time or space. Such data are prevalent in longitudinal studies, environmental monitoring, and biomedical imaging (Ramsay and Silverman, 2005; Hsing and Eubank, 2015; Wang et al., 2016). 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  $d_{L^2}(f, g) = (\int_{\mathcal{T}} \{f(t) - g(t)\}^2 dt)^{1/2}$ . 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 2.2** (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  $\mathcal{S}_+^{d-1}$  (Sealey 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 2.3** (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 2.4** (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  $\text{Sym}_m^+$  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  $\text{Sym}_m^+$ . Other metrics such as the power metric (Dryden et al., 2009) and Log-Cholesky metric (Lin, 2019) also endow  $\text{Sym}_m^+$  with a uniquely geodesic structure.

**Example 2.5** (One-Dimensional 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),  $d_{\mathcal{W}}(\mu, \nu) = \left( \int_0^1 \{F_{\mu}^{-1}(s) - F_{\nu}^{-1}(s)\}^2 ds \right)^{1/2}$ , where  $F_{\mu}^{-1}$  and  $F_{\nu}^{-1}$  are the quantile functions of  $\mu$  and  $\nu$ , respectively. The unique geodesic connecting  $\mu$  and  $\nu$  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  $\text{id}$  denotes the identity map,  $F_{\mu}$  is the cumulative distribution function of  $\mu$ , and  $\tau_{\#}\mu$  denotes the pushforward of  $\mu$  under the map  $\tau$ . 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  $f_{\mu}$  and  $f_{\nu}$ , 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 6.

**3.1. Identification.** We first introduce some notations. Suppose that we observe an independent and identically distributed sample  $\{Y_i, R_i\}_{i=1}^n$  of  $(Y, R) \in \mathcal{M} \times \mathbb{R}$ , where  $\mathcal{M}$  equipped with a metric  $d$  is a compact uniquely geodesic space. The treatment variable is defined as  $T = 1\{R \geq c\}$ , where  $c$  is a known threshold, referred to as the cutoff, and  $1\{\cdot\}$  denotes the indicator function. Hence, the treatment  $T$  is fully determined by the running variable  $R$ . 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  $T = t$ .

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. 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 now define the causal estimand of interest.

**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 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

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

The GLATE can be interpreted as an extension of  $\tau_{\text{RDD}}$  for Euclidean outcomes to non-Euclidean outcomes. Indeed, the geodesic  $\tau_{\text{GRDD}}$  incorporates the directional information from  $\mathbb{E}_{\oplus}[Y(0)|R = c]$  to  $\mathbb{E}_{\oplus}[Y(1)|R = c]$ , which corresponds to the sign of  $\tau_{\text{RDD}}$  and one can quantify the magnitude of the treatment effect by the length of  $\tau_{\text{GRDD}}$ , i.e.,  $d(\mathbb{E}_{\oplus}[Y(0)|R = c], \mathbb{E}_{\oplus}[Y(1)|R = c])$ , which corresponds to the absolute value of  $\tau_{\text{RDD}}$ .

When the outcomes are univariate probability distributions equipped with the 2-Wasserstein metric (see Example 2.5), 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),

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

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.** *Suppose that  $\mathbb{E}_{\oplus}[Y(0)|R = r]$  and  $\mathbb{E}_{\oplus}[Y(1)|R = r]$  exist uniquely for each  $r$  in a neighborhood of  $c$ , and that they are continuous in this neighborhood. Then*

$$\tau_{\text{GRDD}} = \gamma_{\beta_0, \beta_1}, \tag{3.1}$$

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

We note that this identification results also apply 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  $\tau_{\text{GRDD}}$ , 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.

We begin by imposing a mild condition on the kernel function used to define the estimator.

**Assumption 3.1.** *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 local Fréchet regressions, 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 local Fréchet regression 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 reviewed with more details in Appendix A.

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  $t \in \{0, 1\}$ ,

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

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

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

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

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

### Assumption 3.2.

- (i) *The objects  $\mu_{0,\oplus}$  and  $\mu_{1,\oplus}$  exist and are unique. For all  $n$ ,  $\nu_{0,\oplus}, \nu_{1,\oplus}, \widehat{\nu}_{0,\oplus}$ , and  $\widehat{\nu}_{1,\oplus}$  exist and are unique, the latter two almost surely. Additionally, for  $t \in \{0, 1\}$  and for any  $\varepsilon > 0$ ,*

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

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

(ii) For  $t \in \{0, 1\}$ , let  $B_\delta(\mu_{t,\oplus}) \subset \mathcal{M}$  be the ball of radius  $\delta$  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) The density  $f_R$  of  $R_i$ , as well as the conditional densities  $g_{y,t}$  of  $R_i|Y_i(t) = y$  for  $t \in \{0, 1\}$ , exist and are twice continuously differentiable, the latter for all  $y \in \mathcal{M}$ , and  $\sup_{r,y} (|g''_{y,0}(r)| + |g''_{y,1}(r)|) < \infty$ . Additionally,  $f_R(c) > 0$  and for any open set  $U \subset \mathcal{M}$  and  $t \in \{0, 1\}$ ,  $\int_U dF_{Y(t)|R}(r, y)$  is continuous as a function of  $r$ .

(iv) For  $t \in \{0, 1\}$ , 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  $t \in \{0, 1\}$ , 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  $d(\nu, \nu_{t,\oplus}) < \tilde{\eta}$ .

Similar assumptions are made in [Petersen and Müller \(2019\)](#). Condition (i) is a standard separation condition to achieve the consistency of the M-estimator (see, e.g., Chapter 3.2 in [van der Vaart and Wellner \(2023\)](#)). Condition (ii) postulates an entropy condition for  $\mu_{t,\oplus}$ , while (iii) is a standard distributional assumption for local linear regression. Conditions (iv) and (v) provide the rate for the bias and stochastic terms, respectively. Together, conditions (ii), (iv) and (v) leverage tools from empirical process theory to control deviations of  $M_{t,n}(\cdot) - M_{t,\oplus}(\cdot, c)$  and  $\widehat{M}_{t,n}(\cdot) - M_{t,n}(\cdot)$  around their respective minimizers  $\mu_{t,\oplus}$  and  $\nu_{t,\oplus}$ .

To quantify the contrast among the geodesics,  $\tau_{\text{GRDD}}$ ,  $\tau_{\text{GRDD}}^*$  and  $\widehat{\tau}_{\text{GRDD}}$ , 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 3.3.** *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 2.1–2.5 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  $[\gamma_{\alpha_1,\beta_1}], [\gamma_{\alpha_2,\beta_2}] \in \mathcal{G}(\mathcal{M})/\sim$ ,

$$d_{\mathcal{G}}([\gamma_{\alpha_1,\beta_1}], [\gamma_{\alpha_2,\beta_2}]) = d_{\mathcal{G},\omega_{\oplus}}([\gamma_{\alpha_1,\beta_1}], [\gamma_{\alpha_2,\beta_2}]) = d(\Gamma_{\alpha_1,\beta_1}(\omega_{\oplus}), \Gamma_{\alpha_2,\beta_2}(\omega_{\oplus})), \quad (3.2)$$

where  $\omega_{\oplus} \in \mathcal{M}$  is a fixed reference point, such as the Fréchet mean of the outcome. 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 3.4.** *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 in Zhou et al. (2025) for the verification of Assumption 3.4 for Examples 2.1–2.5 provided in Section 2. 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, 3.3, 3.4 and the assumptions in Theorem 3.1 are satisfied.*

(i) *Under Assumptions 3.2 (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 3.2 (i), (ii) and (v) and if  $h_{\max} \rightarrow 0$  and  $nh_{\min} \rightarrow \infty$  as  $n \rightarrow \infty$ , then*

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

By an argument similar to the proof of Propositions 1–3 in Petersen and Müller (2019), one can see that Assumption 3.2 holds with  $\beta_1 = \beta_2 = 2$  for Examples 2.1–2.4 in Section 2, while for Example 2.5 a different argument provides the same rate. Therefore, Theorem 3.2 yields that if  $h_0 = h_1 = h$ , then

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

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.

## 4. IMPLEMENTATION AND SIMULATIONS

**4.1. Bandwidth selection.** Selecting an appropriate bandwidth is critical for reliable estimation in RDDs, especially when the outcome resides in a general metric space. The conventional MSE-optimal bandwidth rule (Imbens and Kalyanaraman, 2012; Calonico et al., 2014, 2020), designed for scalar outcomes, relies on derivative-based approximations and is not applicable in our setting of non-Euclidean outcomes. The GRDD estimator instead requires a method that respects both the geometry of the outcome space and the local nature of discontinuity estimation.

We propose a data-adaptive bandwidth selection strategy that leverages the assumption of smoothness away from the cutoff point  $c$ . Rather than targeting global fit, our method selects the bandwidth that minimizes the integrated squared discrepancy between left and right regression estimates at points where the function is assumed to be continuous. This local criterion helps guard against bias from discontinuities and boundary effects.

The candidate bandwidth range is defined using data-driven heuristics:  $b_{\min}$  is the maximum of (i) the largest gap between adjacent  $R_i$  values, (ii) the distance to the 20th closest  $R_i$  below  $c$ , and (iii) the distance to the 20th closest  $R_i$  above  $c$ ;  $b_{\max}$  is set to half the distance from  $c$  to the edge of the support. A grid of candidate bandwidths is defined over  $[b_{\min}, b_{\max}]$ , and the evaluation region excludes  $[c - b_{\min}, c + b_{\min}]$  and the tails of the support. For each bandwidth, left and right regression estimates are computed at regularly spaced points, and the optimal bandwidth minimizes the integrated squared discrepancy. The full procedure is given in Algorithm 1.

**4.2. Comparison of bandwidth selection methods for the special case of scalar outcomes.** Unlike the conventional MSE-optimal bandwidth rule, which is tailored to Euclidean outcomes, the proposed bandwidth selection strategy is applicable to outcomes residing in general metric spaces. To evaluate its empirical performance, we consider the special case of scalar outcomes, allowing for direct comparison with the MSE-optimal method implemented in the `rdrobust`

---

**Algorithm 1** Data-Adaptive Bandwidth Selection for GRDD estimator

---

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

**Output:** Optimal 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: Compute  $b_{\min}$  as the maximum of:
    - the largest gap between adjacent  $R_i$ ,
    - the distance from  $c$  to the 20th closest  $R_i < c$ ,
    - the distance from  $c$  to the 20th closest  $R_i \geq c$
  - 3: Compute  $b_{\max}$  as half the minimum of  $c - R_{\min}$  and  $R_{\max} - c$
  - 4: Construct a grid of bandwidths  $\mathcal{B} \subset [b_{\min}, b_{\max}]$
  - 5: Define the evaluation region  $\mathcal{R}$  as the support of  $R_i$  excluding  $[c - b_{\min}, c + b_{\min}]$  and both boundary intervals  $[R_{\min}, R_{\min} + b_{\min}]$  and  $[R_{\max} - b_{\min}, R_{\max}]$
  - 6: **for** each  $b \in \mathcal{B}$  **do**
  - 7:   **for** each  $r \in \mathcal{R}$  **do**
  - 8:     Compute left regression estimate  $\hat{m}_-^b(r)$  using observations with  $R_i \in [\max\{r - 2b, R_{\min}\}, r]$  if  $r < c$  or  $R_i \in [\max\{r - 2b, c\}, r]$  if  $r \geq c$
  - 9:     Compute right regression estimate  $\hat{m}_+^b(r)$  using observations with  $R_i \in [r, \min\{r + 2b, c\}]$  if  $r < c$  or  $R_i \in [r, \min\{r + 2b, R_{\max}\}]$  if  $r \geq c$
  - 10:   **end for**
  - 11:   Compute  $L(b) = \int_{\mathcal{R}} d^2(\hat{m}_-^b(r), \hat{m}_+^b(r)) dr$
  - 12: **end for**
  - 13:  $b^* = \arg \min_b L(b)$
- 

package (Calonico et al., 2017). The goal is to examine the accuracy and robustness of each method in recovering treatment effects under varying levels of complexity of the underlying regression functions.

The simulation design features four distinct data-generating processes, each characterized by a different degree of oscillation in the regression functions. In all cases, we generate  $n = 1000$  independent and identically distributed samples  $\{Y_i, R_i\}_{i=1}^n$ , where  $R_i \sim \text{Unif}(-1, 1)$  denotes the running variable. The outcome variable is defined as

$$Y_i = \begin{cases} m_-(R_i) + \varepsilon_i, & R_i < 0 \\ m_+(R_i) + \varepsilon_i, & R_i \geq 0 \end{cases} \quad \text{with } \varepsilon_i \sim N(0, 0.5^2).$$

The four simulation settings are specified as follows:

- (I)  $m_-(r) = r, \quad m_+(r) = r + \tau;$
- (II)  $m_-(r) = r + \sin(3\pi r), \quad m_+(r) = r + \sin(3\pi r) + \tau;$
- (III)  $m_-(r) = r + \sin(8\pi r) + \cos(6\pi r), \quad m_+(r) = r + \sin(6\pi r) + \cos(8\pi r) + \tau;$
- (IV)  $m_-(r) = r + \sin(6\pi r), \quad m_+(r) = r + \sin(6\pi r) + \tau.$

We set the treatment effect to  $\tau = 1$  and perform 500 Monte Carlo replications for each setting. In each replication, the treatment effect is estimated using both the proposed GRDD-based bandwidth selector and the conventional MSE-optimal selector. Figure 1 summarizes the results across replications for all four simulation settings. The left column shows the true regression functions,

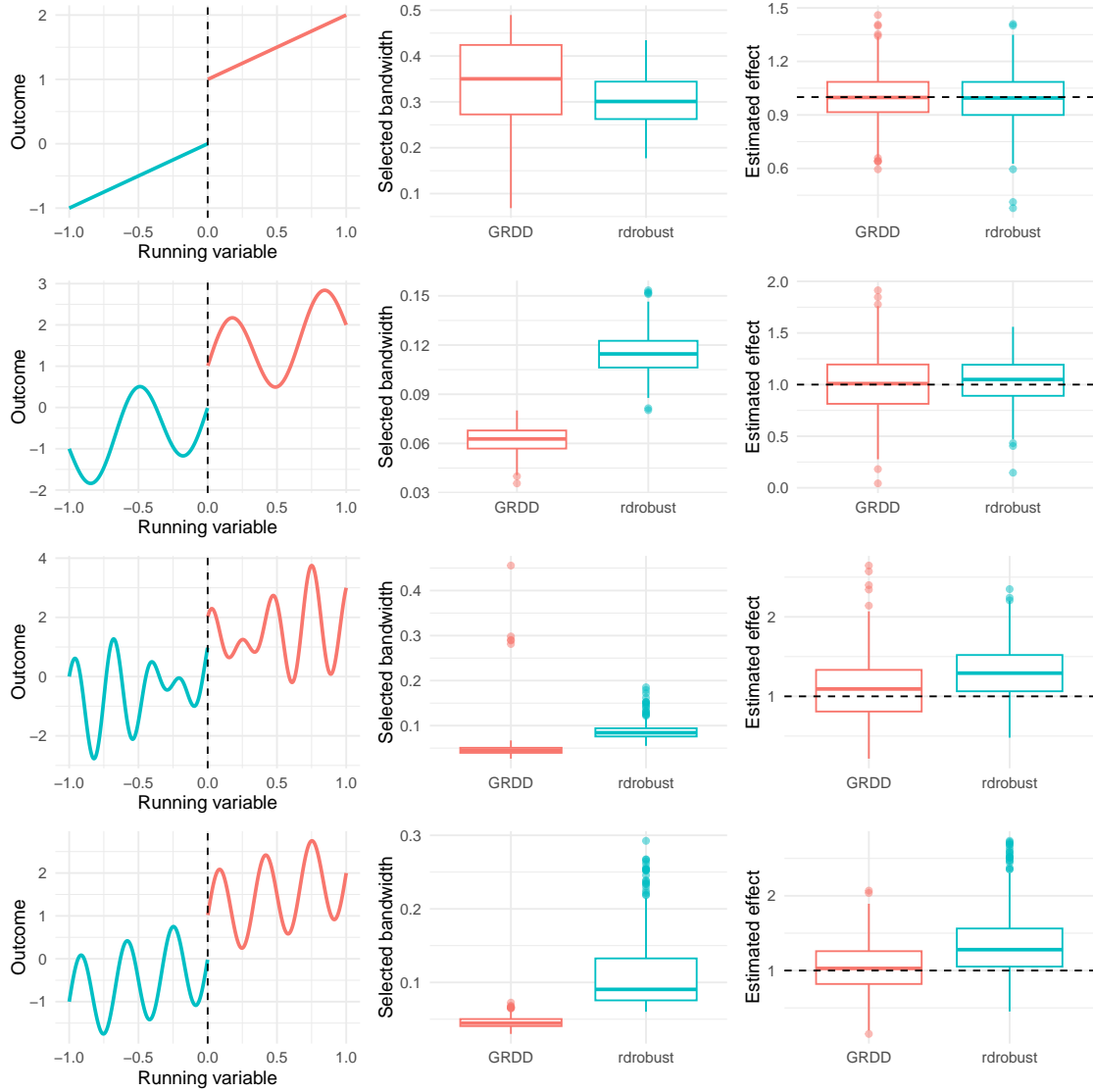


FIGURE 1. Comparison of bandwidth selection methods. Left column displays the true underlying functions (blue for control, red for treatment), while middle and right columns show boxplots of selected bandwidth and estimated treatment effects using rdrobust (blue) and GRDD (red) methods. The dashed line in the boxplots indicates the true treatment effect ( $\tau = 1$ ). The four rows correspond to four data-generating processes with different levels of function complexity.

the middle column presents boxplots of the selected bandwidths, and the right column displays boxplots of the corresponding estimated treatment effects.

In Settings I and II, both methods yield estimates that are tightly centered around the true treatment effect. In the linear case (Setting I), the two methods perform nearly identically. In Setting II, the GRDD-based method yields more accurate estimates with slightly higher variance, while the MSE-optimal approach achieves lower variance but exhibits mild bias. This pattern is consistent with the tendency of the MSE-optimal rule to select relatively large bandwidths, which

performs well when the regression function is globally smooth and exhibits limited local variation near the cutoff.

In contrast, the performance of the MSE-optimal method degrades in Settings III and IV, where the regression functions contain high-frequency components. The larger bandwidths selected by this method lead to oversmoothing, which masks the local discontinuity at the cutoff and results in noticeable upward bias. The GRDD-based method, by contrast, demonstrates superior robustness in these more challenging scenarios. By leveraging local structure and enforcing smoothness away from the cutoff, it better preserves the treatment discontinuity and avoids contamination from irrelevant oscillations. Across Settings III and IV, the GRDD estimates exhibit substantially lower bias and comparable variance relative to the MSE-optimal method.

Overall, these results underscore the strengths of the proposed bandwidth selection strategy in complex settings. Its localized design and adaptive behavior enable it to mitigate the risk of oversmoothing and maintain fidelity to the treatment effect, even in the presence of highly structured noise. This makes the method particularly well suited for RDDs involving metric space-valued outcomes or scalar outcomes with heterogeneous features.

**4.3. Simulation for network-valued outcomes.** To assess the applicability of GRDD in more complex settings, we consider outcomes in the form of network structures. Specifically, each outcome is a graph Laplacian belonging to the space  $\mathcal{L}$ , equipped with the Frobenius metric (see Example 2.3). This simulation setting reflects common use cases where the observed objects are structured networks and the analysis seeks to identify discontinuous changes in their geometry.

Networks are generated under a weighted stochastic block model with 10 nodes divided equally into two communities. The binary adjacency matrix is determined by a block-structured edge probability matrix  $\begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{pmatrix}$ , where the entries denote the probabilities of edge formation within and between communities. Conditional on the adjacency matrix, edge weights are generated using the running variable  $R \sim \text{Unif}(-1, 1)$ . Specifically, for each present edge, the weight is given by

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

where  $\epsilon_{ij} \sim \text{Unif}(0, 1)$  introduces independent noise. Each weighted adjacency matrix is then converted into a graph Laplacian, which serves as the outcome  $Y_i$ .

To evaluate estimation performance, we perform Monte Carlo simulations across four sample sizes:  $n = 100, 200, 500,$  and  $1000$ . For each  $n$ , we generate 500 independent replications of the dataset  $\{Y_i, R_i\}_{i=1}^n$  and apply GRDD with the data-adaptive bandwidth selection procedure (see Section 4.1). In each replication, the performance of GRDD is assessed using the bias of the estimated treatment effect  $d_G(\hat{\tau}_{\text{GRDD}}, \tau_{\text{GRDD}})$ , where  $d_G$  is the metric in the space of geodesics.

Figure 2 presents the boxplot of the resulting bias for each sample size. We observe that the bias steadily decreases with increasing sample size, reflecting improved estimation accuracy. This behavior supports the consistency of the GRDD estimator when applied to network-valued outcomes. To assess the convergence rate, we fit a least-squares regression line to the log-transformed average bias against  $\log n$ . The resulting slope of  $-0.41$  closely matches the theoretical rate of  $-0.4$  and confirms the expected asymptotic behavior.

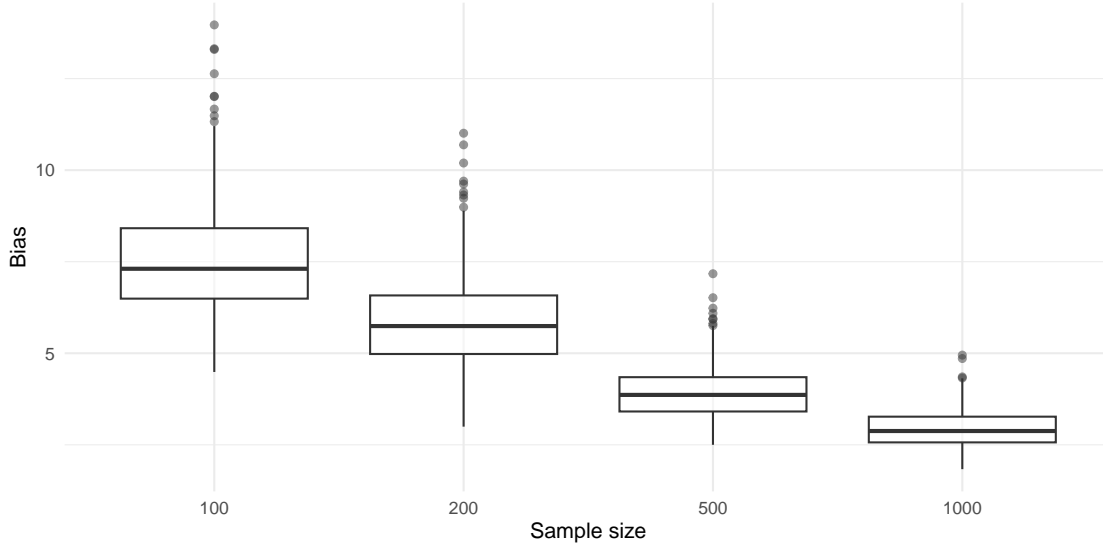


FIGURE 2. Boxplots of bias in estimating the treatment effect for network-valued outcomes across 500 replications for each sample size  $n \in \{100, 200, 500, 1000\}$ .

## 5. REAL WORLD APPLICATIONS

**5.1. Air pollution and urban rail in Taipei.** We apply the GRDD estimator to evaluate the environmental impact of the Taipei Metro on carbon monoxide (CO) concentrations in urban Taipei. The analysis is based on hourly air quality data from the Taiwanese Environmental Protection Administration, covering the period from 1994 to 2007 (Chen and Whalley, 2012). We focus on ten monitoring stations in the Taipei metropolitan area that operated continuously during this period. Among them, seven stations are located adjacent to major highways and three are situated further away, allowing us to assess differential impacts of the metro system across areas with varying exposure to traffic-related pollution.

The Taipei Metro underwent a major expansion in March 1996, providing a natural experimental setting to examine whether this urban transit investment caused discontinuous changes in air quality. We define the official policy intervention date (March 28, 1996) as the cutoff for the proposed RDD. The running variable is calendar time, measured in days relative to the cutoff, and the outcome is a functional observation, consisting of the daily curve of CO concentrations. For each station-day, we smooth the available hourly CO measurements using local linear smoothing and then aggregate the resulting curves across stations within each group (near and far from highway) to obtain a group-level smoothed daily CO curve. This yields a sequence of functional data, one per day per group, that can be analyzed using the GRDD method.

The proposed GRDD framework is particularly well-suited for this application for two main reasons. First, the raw data exhibit missingness: not every station reports CO measurements for every hour of every day. This irregularity renders a traditional hour-by-hour scalar RDD analysis infeasible or highly sensitive to incomplete data. GRDD circumvents this by allowing us to reconstruct full functional curves per station and then modeling the discontinuity in the Hilbert space of functional data. Second, the functional data perspective respects the inherent temporal structure and smoothness of the daily CO trajectories, capturing interactions across

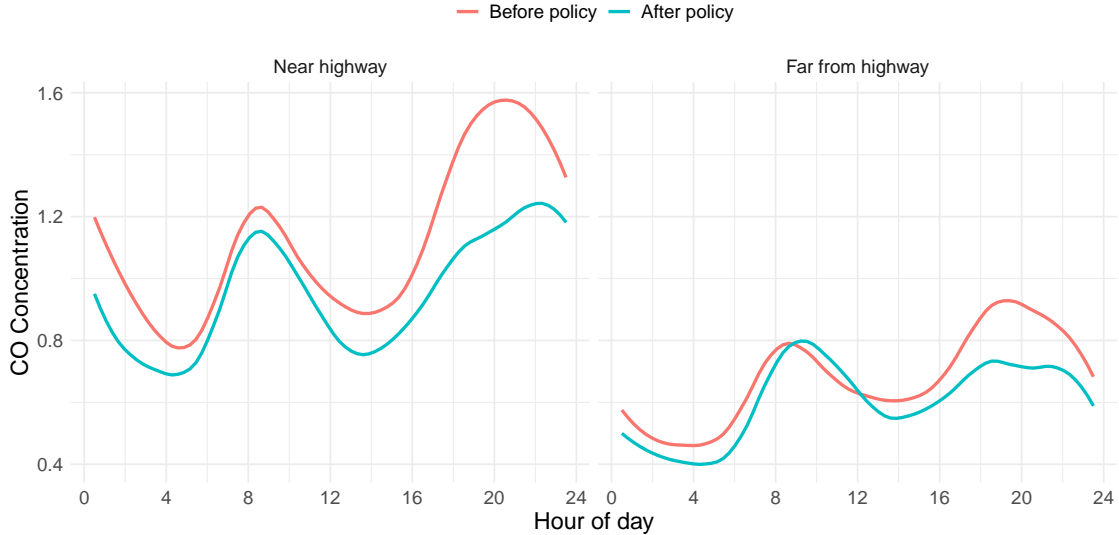


FIGURE 3. Estimated daily CO concentration curves before and after the Taipei Metro expansion. Each panel shows the average 24-hour CO concentration curves estimated for stations near highways (left) and far from highways (right) using geodesic sharp RDD. The blue and red curves represent the estimated CO concentration trajectories just before and just after the March 28, 1996 policy cutoff.

time (e.g., morning and evening peaks) that scalar analyses would ignore. By working with entire curves, GRDD also facilitates more nuanced analyses on how the shape of pollution profiles changes in response to the policy intervention, providing a more comprehensive perspective compared to focusing on shifts in level at a single time point, which would correspond to a scalar response.

To estimate the causal effect of the metro expansion, we apply GRDD separately to the near-highway and far-from-highway station groups. For each group, we estimate the left and right regression limits at the cutoff using LFR with the proposed bandwidth selector, using the data-adaptive strategy described in Section 4.1, yielding bandwidths of 66 days for the near-highway group and 53 days for the far-from-highway group. Figure 3 displays the estimated average 24-hour CO concentration curves immediately before and after the policy cutoff. The results indicate a marked reduction in CO levels for stations near highways, particularly during peak traffic hours. In contrast, the far-from-highway group exhibits a more modest and diffuse change, suggesting a smaller direct effect of the policy in these areas.

To further illustrate the dynamics of the transition around the policy date, we perform LFR over a 200-day window centered at the cutoff (100 days before and after). Figure 4 shows the 3D surface plot of daily CO concentration curves. The plots reveal a pronounced reduction in CO levels for the near-highway group after the policy, whereas changes in the far-from-highway group are more gradual and less pronounced. This evidence suggests that the intervention of building the Taipei Metro led to an improvement in air quality near major roads, consistent with reductions in traffic-related emissions due to increased public transit usage (Chen and Whalley, 2012; Jayasooriya et al., 2017; Gendron-Carrier et al., 2022).

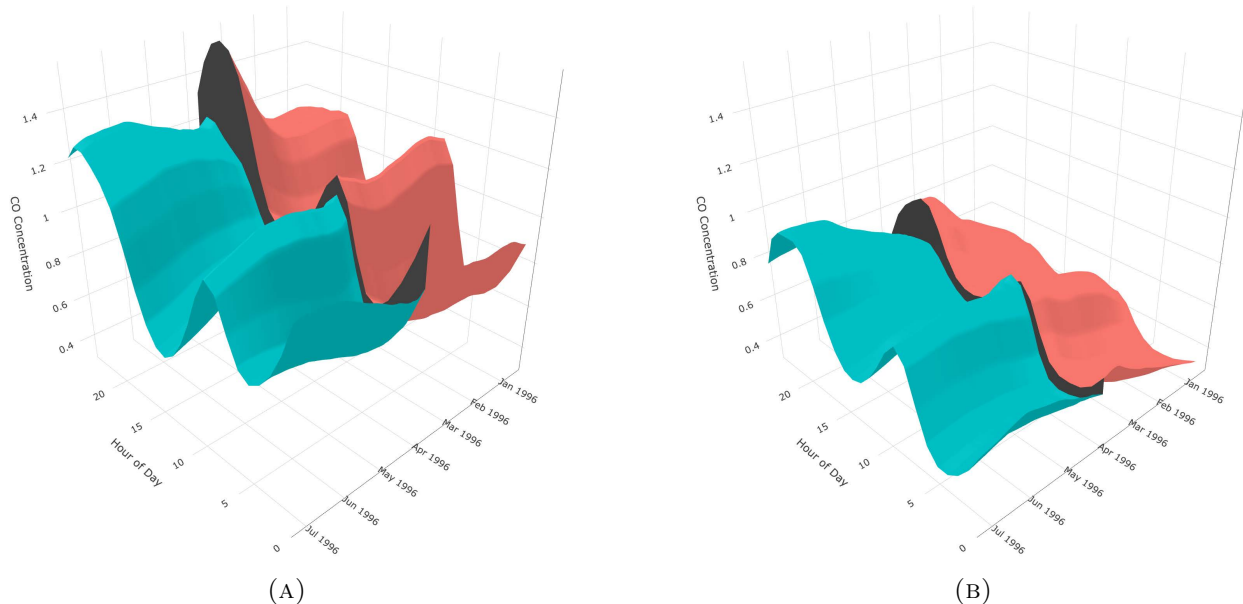


FIGURE 4. Daily CO concentration surfaces around the metro policy cutoff for monitoring stations located (A) near and (B) far from highways. Dates before the policy cutoff are shown in red, and those after the cutoff are shown in blue. The policy cutoff date (March 28, 1996) is marked by a gray vertical wall. Both panels use a shared vertical axis scale to facilitate comparison.

**5.2. Incumbency effects in multiparty parliamentary 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 environment 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: the difference in vote share between the Conservative candidate and their top rival. The outcome is the compositional vector of vote shares in the subsequent election, restricted to the three major parties. Standard RDDs typically estimate treatment effects on each outcome component separately. However, such marginal analyses neglect the compositional constraint—that all vote shares must sum to one—and fail to capture the interdependent structure of the outcome space. In contrast, GRDD treats the full composition as a point on the probability simplex and estimates the treatment effect within a metric space framework. This joint modeling approach respects the intrinsic geometry of the data and makes it possible to detect global changes in the distribution of voter support across parties.

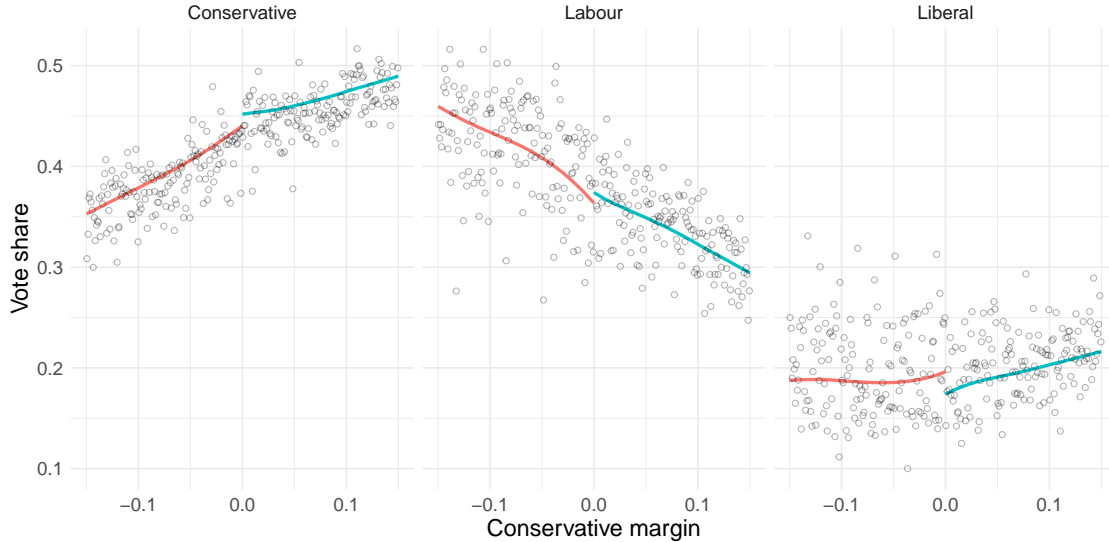


FIGURE 5. Estimated GRDD fits and bin-averaged outcomes by Conservative margin using UK parliamentary election data from 1950 to 2010.

The bandwidth is selected using the data-adaptive strategy described in Section 4.1, which yields a value of 0.09. Using this bandwidth, we estimate the vote share vector to the left of the cutoff (where the Conservative candidate narrowly lost) as 44.0% Conservative, 36.4% Labour, and 19.6% Liberal. To the right of the cutoff (where the Conservative candidate narrowly won), the estimated vector shifts to 45.2% Conservative, 37.4% Labour, and 17.4% Liberal. These results suggest a subtle but meaningful incumbency effect. The Conservative share increases modestly, while the Liberal share decreases more noticeably, indicating a potential reallocation of support from third-party voters toward the incumbent. The Labour share also rises slightly, suggesting more complex realignment dynamics that go beyond a simple two-party shift.

Figure 5 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. A visible discontinuity appears at the cutoff, which is most notable in the Liberal panel.

## 6. 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 proposed GRDD method to fuzzy RDDs. A basic approach is to estimate the average treatment effect by amplifying the discontinuous change in the outcome variable at the cutoff by the reciprocal of the discontinuous change in the probability of treatment at the cutoff.

**6.1. Setup.** Since the fuzzy RDD case is quite distinct from the sharp RDD case studied in Section 3, we first reintroduce notations. 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  $(\mathcal{M}, d)$  is a compact uniquely geodesic space,  $T$  is a binary treatment indicator variable,  $Y \in \mathcal{M}$  is an outcome object, 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 average treatment effect of the compliers, given by  $E[Y(T(1))|T(1) > T(0), R = c] - E[Y(T(0))|T(1) > T(0), R = c]$ . 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  $\Psi$  of  $\mu_{z,\oplus}$  to a Hilbert space, where the linear structure makes it possible to evaluate a contrast for  $\Psi(\mu_{z,\oplus})$ .

**6.2. Identification.** To formally introduce the mapping  $\Psi$ , let  $\mathcal{Z}$  be a nonempty set and  $\rho : \mathcal{Z} \times \mathcal{Z} \rightarrow [0, \infty)$  be a function such that for any  $z_1, z_2 \in \mathcal{Z}$ ,  $\rho(z_1, z_2) = 0$  if and only if  $z_1 = z_2$ , and  $\rho(z_1, z_2) = \rho(z_2, z_1)$ , i.e.,  $\rho$  is a semimetric on  $\mathcal{Z}$  and  $(\mathcal{Z}, \rho)$  is a semimetric space. The semimetric space  $(\mathcal{Z}, \rho)$  with a semimetric  $\rho$  is of negative type if, for all  $n \geq 2$ ,  $z_1, z_2, \dots, z_n \in \mathcal{Z}$  and  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$  with  $\sum_{i=1}^n \alpha_i = 0$ , we have  $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho(z_i, z_j) \leq 0$ .

Throughout this subsection, we assume that  $(\mathcal{M}, d)$  is a uniquely geodesic space for which there exists a Hilbert space  $\mathcal{H}$  equipped with an inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  and an injective map  $\Psi : \mathcal{M} \rightarrow \mathcal{H}$  such that for any  $\alpha, \beta \in \mathcal{M}$ ,

$$d^2(\alpha, \beta) = d_{\mathcal{H}}^2(\Psi(\alpha), \Psi(\beta)). \quad (6.1)$$

Here  $d_{\mathcal{H}}$  is the metric induced by the inner product so that any object  $\alpha \in \mathcal{M}$  can be mapped into a Hilbert space  $\mathcal{H}$ , and the map  $\Psi$  is an isometric mapping. A sufficient condition for this property is provided in Proposition 3 in [Sejdinovic et al. \(2013\)](#) for example, which implies that if  $d^2$  is of negative type, then this condition is satisfied; see also [Schoenberg \(1938\)](#). Let  $\Psi(\mathcal{M}) = \{\Psi(\mu) : \mu \in \mathcal{M}\}$  be the image space of  $\mathcal{M}$  under  $\Psi$  and assume that  $\Psi(\mathcal{M})$  is a convex subset of  $\mathcal{H}$ . Commonly encountered uniquely geodesic spaces for which a mapping  $\Psi$  that satisfies (6.1) exists include the space of functional data with the  $L^2$  metric (Example 2.1), the space of networks represented as graph Laplacians with the Frobenius metric (Example 2.3), the space of SPD matrices with the Frobenius metric, the power metric, the Log-Euclidean metric, or the Log-Cholesky metric (Example 2.4) and also the space of one-dimensional probability distributions on a common compact support with the 2-Wasserstein metric (Example 2.5), among others. See Appendix C.1 for explicit maps  $\Psi$  for these spaces.

Based on the above notation, our causal estimand of interest is defined as follows.

**Definition 6.1** (Compliers' LATE at the cutoff). For outcomes in a metric space that admits a map  $\Psi$  satisfying (6.1), we define the compliers' LATE at the cutoff as

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

When  $\mathcal{M}$  is (a subset of) the Euclidean space  $\mathbb{R}$ , one can take  $\Psi$  as the identity map, and hence we have

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

Thus,  $\tau_{\text{FRDD}}$  is an extension of the conventional notion of the compliers' LATE at the cutoff for Euclidean outcomes. Note that the assumption on the metric space  $(\mathcal{M}, d)$  in this subsection excludes the space of compositional data with the arc-length metric (Example 2.2) since we do not have an isometric embedding map  $\Psi$  for outcomes in a (subset of) Riemannian manifolds with a positive curvature. However, one can still obtain a sensible definition of the compliers' LATE in this setting; see Section 6.5.

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

**Assumption 6.1.**

- (i)  $T(0) \leq T(1)$ .
- (ii)  $P(T(1) > T(0)|R = c) \in (\eta_0, 1 - \eta_0)$  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  $\tau_{\text{FRDD}}$ .

**Theorem 6.1.** *Consider the setup of Section 6.1. Under Assumption 6.1,*

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

Note that for the space of one-dimensional distributions with the 2-Wasserstein metric (Example 2.5), 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,  $\tau_{\text{FRDD}}$  corresponds to the fuzzy local average quantile treatment effect  $F_{\mu_{1, \oplus}^{\dagger}}^{-1}(\cdot) - F_{\mu_{0, \oplus}^{\dagger}}^{-1}(\cdot)$  that was considered in Van Dijke (2025). Theorem 6.1 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.

**6.3. Estimation.** We now construct an estimator of  $\tau_{\text{FRDD}}$  and provide its asymptotic properties. A natural estimator is

$$\widehat{\tau}_{\text{FRDD}} = \frac{\Psi(\widehat{\nu}_{1, \oplus}) - \Psi(\widehat{\nu}_{0, \oplus})}{\widehat{m}_1 - \widehat{m}_0}, \tag{6.2}$$

where  $\widehat{\nu}_{0, \oplus}$  and  $\widehat{\nu}_{1, \oplus}$  are the LFR estimators defined in Section 3.2, and  $\widehat{m}_0$  and  $\widehat{m}_1$  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  $\widehat{m}_z = \widehat{m}_{z,0}$ , leading to

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

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

**Theorem 6.2.** *Consider the setup of Sections 6.1 and 6.2. Assume that  $\widehat{m}_z - m_z = O_p(r_n)$  for  $z \in \{0, 1\}$ , where  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . Under Assumptions 3.1 and 6.1 and under Assumption 3.2, 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*

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

Many common metric spaces satisfy Assumption 3.2 with  $\beta_1 = \beta_2 = 2$ . If one uses the local linear estimator (6.2) for estimating  $m_z$ , its typical convergence rate is  $\widetilde{r}_n = O(h_{\max}^2) + O_p((nh_{\min})^{-1/2})$  (cf. Theorems 19.1 and 19.2 in Hansen (2022)). Then letting  $\beta_1 = \beta_2 = 2$ ,  $h_0 = h_1 = h$  and  $r_n = \widetilde{r}_n$ , we have

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

The right-hand side is optimized with  $h = n^{-1/5}$  and the resulting convergence rate is the nonparametric rate  $d_{\mathcal{H}}(\tau_{\text{FRDD}}, \widehat{\tau}_{\text{FRDD}}) = O_p(n^{-2/5})$ .

**6.4. 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 instead of  $\tau_{\text{FRDD}}$ .

**Definition 6.2** (Compliers' GLATE at the cutoff). The complier's 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 6.1, compliers' GLATE  $\tau_{\text{GFRD}}$  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,  $\mu_{\oplus} = \mathbb{E}_{\oplus}[Y(T) | T(1) = T(0), R = c]$ . When the metric space  $(\mathcal{M}, d)$  admits a map  $\Psi$  satisfying (6.1), one can see that

$$\Psi(\mu_{z,\oplus}) = \Psi(\mu_{\oplus}) + \frac{\Psi(\mu_{z,\oplus}^\dagger) - \Psi(\mu_{\oplus})}{\mathbb{E}[T(1) | R = c] - \mathbb{E}[T(0) | R = c]} \quad \text{for } z \in \{0, 1\}, \quad (6.3)$$

where  $\mu_{z,\oplus}^\dagger = \mathbb{E}_{\oplus}[Y(T(z)) | R = c]$  (see (C.4) in Appendix C.2 for the derivation of (6.3)). Since  $\Psi : \mathcal{M} \rightarrow \Psi(\mathcal{M})$  is bijective, we have

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

where  $\Psi^{-1} : \Psi(\mathcal{M}) \rightarrow \mathcal{M}$  is the inverse map of  $\Psi$ . In the following, we will assume that the argument of  $\Psi^{-1}$  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 6 illustrates the identification of  $\tau_{\text{FRDD}}$  and  $\tau_{\text{GFRD}}$ . Note that we can estimate  $\mu_{z,\oplus}^\dagger$  and  $m_z$  by  $\widehat{\nu}_{z,\oplus}$  and  $\widehat{m}_z$ , 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 noncompliance:

**Assumption 6.2.** *For any unit, either one of the following holds.*

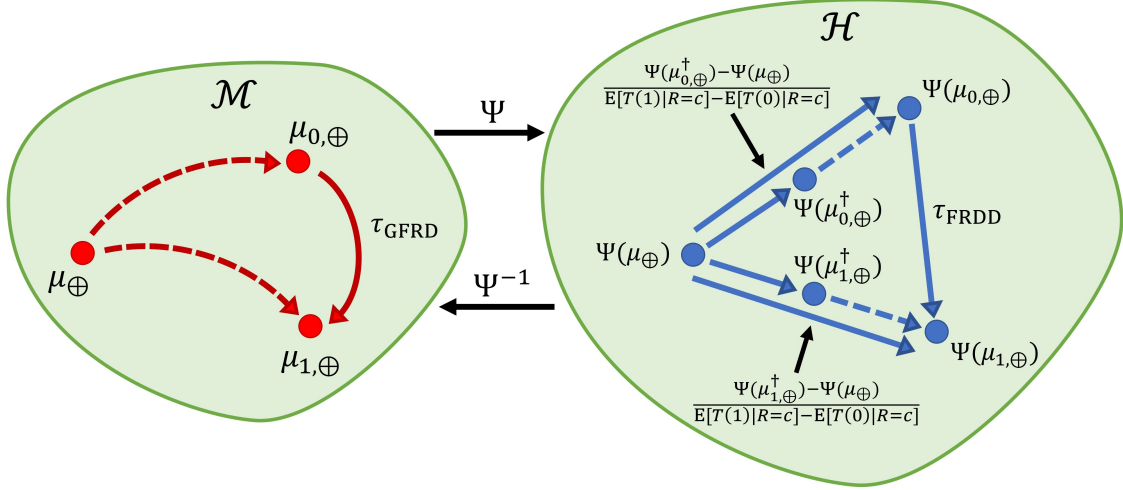


FIGURE 6. Illustration of the identification of  $\tau_{\text{FRDD}}$  and  $\tau_{\text{GFRD}}$ . The circles symbolize objects in the geodesic space  $\mathcal{M}$ , and the arrows emanating from these circles represent the geodesic paths connecting them. The solid blue arrow from  $\Psi(\mu_{0,\oplus})$  to  $\Psi(\mu_{1,\oplus})$  represents the compliers' LATE. The solid red arrow represents the compliers' GLATE.

- (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 6.1 with this condition as Assumption 6.1'. Under one-sided noncompliance, we can identify  $\mu_{\oplus}$  as follows.

**Proposition 6.1.** *Consider the setup of Sections 6.1 and 6.2 and suppose that Assumption 6.1' holds true. Then*

$$\mu_{\oplus} = \begin{cases} \lim_{r \uparrow c} E_{\oplus}[Y|T = 1, Z = 0, R = r] & \text{under Assumption 6.2 (i)} \\ \lim_{r \downarrow c} E_{\oplus}[Y|T = 0, Z = 1, R = r] & \text{under Assumption 6.2 (ii)} \end{cases}.$$

From this proposition, we can estimate  $\tau_{\text{GFRD}}$  as follows.

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

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

$$\widehat{\mu}_{\oplus} = \begin{cases} \widehat{\mu}_{A,\oplus} := \arg \min_{\nu \in \mathcal{M}} \frac{1}{n_A} \sum_{i \in I_A} \widehat{s}_0(c; R_i, h_0) d^2(\nu, Y_i) & \text{under Assumption 6.2 (i)} \\ \widehat{\mu}_{N,\oplus} := \arg \min_{\nu \in \mathcal{M}} \frac{1}{n_N} \sum_{i \in I_N} \widehat{s}_1(c; R_i, h_1) d^2(\nu, Y_i) & \text{under Assumption 6.2 (ii)} \end{cases},$$

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  $\widehat{s}_z(c; r, h_z)$  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

$$\begin{aligned} M_{A, n}(\nu) &= E[s_0(c; R, h_0) d^2(\nu, Y(T)) | T = 1, Z = 0], \\ M_{N, n}(\nu) &= E[s_1(c; R, h_1) d^2(\nu, Y(T)) | T = 0, Z = 1]. \end{aligned}$$

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

**Proposition 6.2.** *Consider the setup of Sections 6.1 and 6.2. Under Assumptions 3.3, 3.4, 6.1', 6.2 and the assumptions in Theorem 6.2, except for Assumption 6.1, as well as Assumption 3.2 when replacing  $\mu_{t, \oplus}$ ,  $\nu_{t, \oplus}$ ,  $\widehat{\nu}_{t, \oplus}$ ,  $M_{t, n}(\nu)$ ,  $M_{t, \oplus}(\nu, c)$ ,  $Y(t)$  for  $t \in \{0, 1\}$  with  $\mu_{\oplus}$ ,  $\mu_{S, \oplus}$ ,  $\widehat{\mu}_{S, \oplus}$ ,  $M_{S, n}(\nu)$ ,  $M_{\oplus}(\nu, c)$ ,  $Y(T)$  for  $S \in \{A, N\}$ , it holds that*

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

**6.5. 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 one has well-defined Riemannian logarithmic and exponential maps. 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  $\text{Log}_{\omega}(\mathcal{M})$ , the image space of  $\mathcal{M}$  under  $\text{Log}_{\omega}$ , 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 2.2), which is the positive orthant of a finite-dimensional sphere.

For  $z \in \{0, 1\}$ , define  $\mu_{z, \omega}^{\sharp} = E[\text{Log}_{\omega}(Y(T(z))) | T(1) > T(0), R = c]$  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 6.3** (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  $\omega$  as

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

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

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

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

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

**Assumption 6.3.**

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

By a similar argument to Theorem 6.1, the causal effect  $\tau_{\text{FRDD}}^\sharp$  can be identified as follows.

**Theorem 6.3.** *Consider the setup of Sections 6.1 and 6.5 and suppose that Assumption 6.3 holds true. Then*

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

From Theorem 6.3, we can construct an estimator of  $\tau_{\text{FRDD}}^\sharp$  as follows.

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

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

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

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

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

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

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

**Theorem 6.4.** *Consider the setup of Sections 6.1 and 6.5. 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 3.1 and 6.3 and Assumption 3.2 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  $(T_\omega \mathcal{M}, d_E)$ ,  $\nu_{z,\omega}^\sharp$ ,  $\tilde{\nu}_{z,\omega}^\sharp$ ,  $\hat{\nu}_{z,\omega}^\sharp$ ,  $M_{z,n}^\sharp(\nu)$ ,  $M_{z,\oplus}^\sharp(\nu, c)$ ,  $\text{Log}_\omega(Y(T(z)))$  for  $z \in \{0, 1\}$ , if  $h_{\max} \rightarrow 0$  and  $nh_{\min} \rightarrow \infty$  as  $n \rightarrow \infty$ , then*

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

If one uses the sample Fréchet mean  $\hat{m}_\oplus$  of the outcomes  $\{Y_i\}_{i=1}^n$  as the reference point  $\omega$ , then the techniques used to derive Theorem 6.4 cannot be directly applied since the resulting estimator involves the logarithmic map at the estimated point  $\hat{m}_\oplus$ . Chen et al. (2023) investigates 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.

**6.6. Geodesic fuzzy RDD for outcomes in Riemannian manifolds.** As discussed in Section 6.4, one may prefer to consider the compliers' LATE at the cutoff as a quantity defined on the manifold  $\mathcal{M}$ , i.e., the complier's GLATE.

**Definition 6.4** (Compliers' GLATE at the cutoff for outcomes in a Riemannian manifold). For the outcomes in a Riemannian manifold considered in Section 6.5, compliers' GLATE is defined as the geodesic from  $\mu_{0,\omega,\oplus}^\sharp$  to  $\mu_{1,\omega,\oplus}^\sharp$ , that is,

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

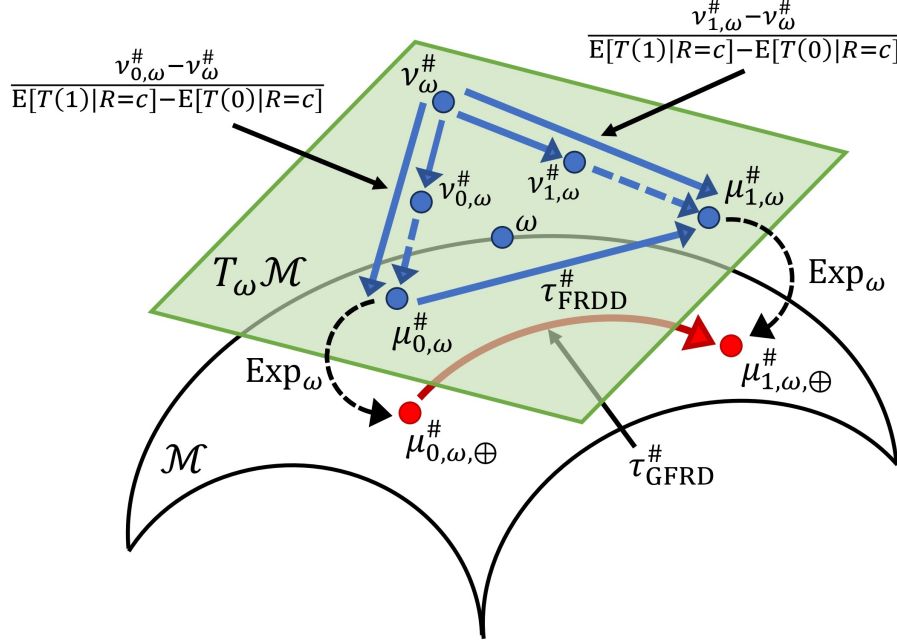


FIGURE 7. Illustration of the identification of  $\tau_{\text{FRDD}}^\#$  and  $\tau_{\text{GFRD}}^\#$ . The circles symbolize objects in the geodesic space  $\mathcal{M}$ , and the arrows emanating from these circles represent the geodesic paths connecting them. The solid blue arrow from  $\mu_{0,\omega}$  to  $\mu_{1,\omega}$  represents the compliers' LATE. The solid red arrow represents the compliers' GLATE.

where  $\mu_{z,\omega,\oplus}^\# = \text{Exp}_\omega(\mu_{z,\omega}^\#)$ ,  $\mu_{z,\omega}^\# = \text{E}[\text{Log}_\omega(Y(T(z))) | T(1) > T(0), R = c]$ , 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  $\nu_{z,\omega}^\# = \text{E}[\text{Log}_\omega(Y(T(z))) | R = c]$  and let  $\nu_\omega^\#$  be the expectation of  $\text{Log}_\omega(Y)$  for always/never-takers at the cutoff, that is,  $\nu_\omega^\# = \text{E}[\text{Log}_\omega(Y) | T(1) = T(0), R = c]$ . For the Riemannian manifold considered in Section 6.5,

$$\mu_{z,\omega,\oplus}^\# = \text{Exp}_\omega \left( \nu_\omega^\# + \frac{\nu_{z,\omega}^\# - \nu_\omega^\#}{m_1 - m_0} \right) \quad \text{for } z \in \{0, 1\}, \quad (6.4)$$

where  $m_z = \text{E}[T(z) | R = c]$ , assuming this is well-defined, i.e., the argument of  $\text{Exp}_\omega$  is in the image space  $\text{Log}_\omega$ ; see (C.10) in Appendix C.6 for the derivation of (6.4). An illustration of the identification of  $\tau_{\text{FRDD}}^\#$  and  $\tau_{\text{GFRD}}^\#$  is in Figure 7. Note that  $\nu_{z,\omega}^\#$  and  $m_z$  can be estimated by  $\widehat{\nu}_{z,\omega}^\#$  and  $\widehat{m}_z$ , respectively. Therefore, if  $\nu_\omega^\#$  can be estimated, we can construct an estimator of  $\mu_{z,\omega,\oplus}^\#$ .

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 6.3 with this condition as Assumption 6.3'. Under one-sided noncompliance, we can identify  $\mu_\oplus$  as follows.

**Proposition 6.3.** *Consider the setup of Sections 6.1 and 6.5. Under Assumption 6.3'*

$$\nu_\omega^\# = \begin{cases} \lim_{r \uparrow c} \text{E}[\text{Log}_\omega(Y) | T = 1, Z = 0, R = r] & \text{under Assumption 6.2 (i)} \\ \lim_{r \downarrow c} \text{E}[\text{Log}_\omega(Y) | T = 0, Z = 1, R = r] & \text{under Assumption 6.2 (ii)} \end{cases}.$$

Proposition 6.3 motivates estimators

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

where  $\widehat{\nu}_{z,\omega}^\#$  and  $\widehat{m}_z$  are estimators considered in Sections 6.5 and 3.2, respectively, and  $\widehat{\nu}_\omega^\#$  is the LFR estimator defined as

$$\widehat{\nu}_\omega^\# = \begin{cases} \widehat{\nu}_{\omega,A}^\# := \arg \min_{\nu \in T_\omega \mathcal{M}} \frac{1}{n_A} \sum_{i \in I_A} \widehat{s}_0(c; R_i, h_0) d_E^2(\nu, \text{Log}_\omega(Y_i)) & \text{under Assumption 6.2 (i)} \\ \widehat{\nu}_{\omega,N}^\# := \arg \min_{\nu \in T_\omega \mathcal{M}} \frac{1}{n_N} \sum_{i \in I_N} \widehat{s}_1(c; R_i, h_1) d_E^2(\nu, \text{Log}_\omega(Y_i)) & \text{under Assumption 6.2 (ii)} \end{cases}.$$

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

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

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

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

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

**Proposition 6.4.** *Consider the setup of Sections 6.1 and 6.5. Suppose that Assumptions 3.3, 3.4, 6.2, 6.3', and assumptions in Theorem 6.4 except for Assumption 6.3 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 3.2 is satisfied when replacing  $(\mathcal{M}, d)$ ,  $\mu_{t,\oplus}$ ,  $\nu_{t,\oplus}$ ,  $\widehat{\nu}_{t,\oplus}$ ,  $M_{t,n}(\nu)$ ,  $M_{t,\oplus}(\nu, c)$ ,  $Y(t)$  for  $t \in \{0, 1\}$  with  $(T_\omega \mathcal{M}, d_E)$ ,  $\nu_\omega^\#, \nu_{\omega,S}^\#, \widehat{\nu}_{\omega,S}^\#, M_{S,n}^\#(\nu)$ ,  $M^\#(\nu, c)$ ,  $\text{Log}_\omega(Y(T))$  for  $S \in \{A, N\}$ . Then*

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

## 7. CONCLUDING REMARKS

We have introduced a general framework for RDD that accommodates complex, non-Euclidean outcomes such as functional data, compositional data, and networks. The proposed GRDD quantifies causal effects via geodesics between LFR estimates, extending the classical RDD framework to metric space-valued responses. To enable practical estimation, we develop a novel bandwidth selection strategy that adapts to local structure near the cutoff while remaining agnostic to the specific geometry of the outcome space. The proposed method is supported by theoretical guarantees and is demonstrated through a series of simulations and empirical applications in environmental and political domains.

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 relevant developments in the scalar case. Second, generalizing GRDD to handle multivariate running variables would make the method applicable to more complex treatment assignment mechanisms, as discussed in Black (1999), Keele and Titiunik (2015), Sawada et al. (2024), and Cattaneo et al. (2025). This extension introduces challenges related to neighborhood construction and bandwidth selection in higher-dimensional spaces. Finally, while not pursued in this paper, the proposed bandwidth selection framework naturally facilitates the development of pseudo-permutation tests to assess the

significance of treatment effects. Designing such tests with strong finite-sample and asymptotic properties remains an open area of interest, especially for testing directional or localized effects in structured outcome spaces; see [Calonico et al. \(2018\)](#) and [Calonico et al. \(2019\)](#) for the special case of scalar outcomes.

## APPENDIX A. 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.

**A.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  $\mathbb{E}[Y|R = r]$  by solving a localized least squares problem centered at  $r$ :

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

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

The solution for  $\beta_0$  admits a closed-form expression. Let  $\mu_k = \mathbb{E}[K_h(R - r)(R - r)^k]$  for  $k = 0, 1, 2$  and define the weighted cross-moments:

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

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:

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

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.

**A.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  $d^2(Y, \nu)$ .

**A.3. Sample Estimator.** Given a sample  $\{(Y_i, R_i)\}_{i=1}^n$ , we define the empirical moments:

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

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

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

The local Fréchet regression estimator is then:

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

## APPENDIX B. PROOFS FOR SECTION 3

### B.1. Proof of Theorem 3.1.

For the proof of (3.1), it suffices to show

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

$$\mathbb{E}_{\oplus}[Y(0)|R = c] = \lim_{r \uparrow c} \mathbb{E}_{\oplus}[Y|R = r]. \quad (\text{B.2})$$

Here we only prove (B.1) since the proof of (B.2) is almost the same. Observe that

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

which completes the proof.

### B.2. Proof of Theorem 3.2.

**Proof of (i):** From Assumptions 3.3 and 3.4, we have

$$\begin{aligned} d_{\mathcal{G}}(\tau_{\text{GRDD}}, \tau_{\text{GRDD}}^*) &= d_{\mathcal{G}}(\gamma_{\mu_{0,\oplus}, \mu_{1,\oplus}}, \gamma_{\nu_{0,\oplus}, \nu_{1,\oplus}}) \\ &= d(\Gamma_{\mu_{0,\oplus}, \mu_{1,\oplus}}(\mu_{0,\oplus}), \Gamma_{\nu_{0,\oplus}, \nu_{1,\oplus}}(\mu_{0,\oplus})) \\ &\leq C\{d(\mu_{0,\oplus}, \nu_{0,\oplus}) + d(\mu_{1,\oplus}, \nu_{1,\oplus})\}. \end{aligned} \quad (\text{B.3})$$

Applying a similar argument for proving Theorem 3 in Petersen and Müller (2019), one can show that  $d(\mu_{t,\oplus}, \nu_{t,\oplus}) = O(h_t^{2/(\beta_1-1)})$ ,  $t \in \{0, 1\}$  under Assumptions 3.1, 3.2 (i)-(iv). Combining this with (B.3), we obtain  $d_{\mathcal{G}}(\tau_{\text{GRDD}}, \tau_{\text{GRDD}}^*) = O(h_{\max}^{2/(\beta_1-1)})$ .

**Proof of (ii):** Under Assumptions 3.3 and 3.4, a similar argument as in the proof of (B.3) yields

$$d_{\mathcal{G}}(\tau_{\text{GRDD}}^*, \widehat{\tau}_{\text{GRDD}}) \leq C\{d(\nu_{0,\oplus}, \widehat{\nu}_{0,\oplus}) + d(\nu_{1,\oplus}, \widehat{\nu}_{1,\oplus})\}. \quad (\text{B.4})$$

Applying a similar argument as in the proof of Theorem 4 in [Petersen and Müller \(2019\)](#), one can show that  $d(\nu_{t,\oplus}, \widehat{\nu}_{t,\oplus}) = O_p((nh_t)^{-\frac{1}{2(\beta_2-1)}})$ ,  $t \in \{0, 1\}$  under Assumptions [3.1](#), [3.2](#) (i), (ii) and (v). Combining this with [\(B.4\)](#), we obtain  $d_G(\tau_{\text{GRDD}}^*, \widehat{\tau}_{\text{GRDD}}) = O_p((nh_{\min})^{-\frac{1}{2(\beta_2-1)}})$ .

## APPENDIX C. PROOFS FOR SECTION 6

### C.1. Verification of [\(6.1\)](#).

#### C.1.1. SPD matrices.

First, we consider the space of SPD matrices  $\text{Sym}_m^+$  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  $\text{id}$  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.

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

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

where  $U\Lambda U'$  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 [\(6.1\)](#) with  $\Psi(A) = F_p(A)$ .

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 [\(6.1\)](#) with  $\Psi(A) = \log(A)$ .

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  $S = LL'$  and a diffeomorphism  $\mathcal{L}$  between  $\text{Sym}_m^+$  and  $\mathcal{L}_m^+$ . For any  $L \in \mathcal{L}_m^+$ , let  $[L]$  be the strictly lower triangular part and  $\mathbb{D}(L)$  the diagonal part of  $L$ . The Log-Cholesky metric is defined as

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

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 [\(6.1\)](#) with  $\Psi(L) = [L] + \log(\mathbb{D}(L))$ .

#### C.1.2. 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 [\(6.1\)](#) with  $\Psi = \text{id}$ . Likewise, one can see that  $\Psi(A) = A^p$  for the space of networks represented by graph Laplacians with the power metric.

### C.1.3. 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  $d_{L^2}(f, g) = (\int_{\mathcal{T}} \{f(t) - g(t)\}^2 dt)^{1/2}$ .

### C.1.4. One-dimensional distributions with the 2-Wasserstein metric.

Let  $\mathcal{W}$  be the space of one-dimensional 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 (6.1) with  $\Psi(\mu) = F_{\mu}^{-1}(\cdot)$ .

## C.2. Proof of Theorem 6.1.

(Step 1) In this step, we show

$$\mu_{1, \oplus}^{\dagger} = \mathbb{E}_{\oplus}[Y(T(1))|R = c] = \lim_{r \downarrow c} \mathbb{E}_{\oplus}[Y|R = r], \quad (\text{C.1})$$

$$\mu_{0, \oplus}^{\dagger} = \mathbb{E}_{\oplus}[Y(T(0))|R = c] = \lim_{r \uparrow c} \mathbb{E}_{\oplus}[Y|R = r], \quad (\text{C.2})$$

$$\mathbb{E}[T(1)|R = c] - \mathbb{E}[T(0)|R = c] = \lim_{r \downarrow c} \mathbb{E}[T|R = r] - \lim_{r \uparrow c} \mathbb{E}[T|R = r]. \quad (\text{C.3})$$

**Proof of (C.1):** Observe that for  $r \geq c$ ,

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

Then we have  $\lim_{r \downarrow c} \mathbb{E}_{\oplus}[Y|R = r] = \lim_{r \downarrow c} \mathbb{E}_{\oplus}[Y(T(1))|R = r] = \mathbb{E}_{\oplus}[T(1)|R = c]$ . Likewise, one can show (C.2).

**Proof of (C.3)** 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 convexity of the image space  $\Psi(\mathcal{M})$  for the third equality. 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

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

Then we have

$$\begin{aligned} \Psi(\mu_{z,\oplus}) &= \Psi(\mathbb{E}_{\oplus}[Y(T(z))|T(1) > T(0), R = c]) \\ &= \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c]) \\ &\quad + \frac{\Psi(\mathbb{E}_{\oplus}[Y(T(z))|R = c]) - \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c])}{\mathbb{P}(T(1) > T(0)|R = c)} \\ &= \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c]) \\ &\quad + \frac{\Psi(\mu_{z,\oplus}^{\dagger}) - \Psi(\mathbb{E}_{\oplus}[Y(T)|T(1) = T(0), R = c])}{\mathbb{P}(T(1) > T(0)|R = c)}. \end{aligned} \tag{C.4}$$

Therefore,

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

Combining this with (C.1)-(C.3) completes the proof.

### C.3. Proof of Theorem 6.2.

Define

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

Observe that

$$\begin{aligned} d_{\mathcal{H}}(\tau_{\text{FRDD}}, \widehat{\tau}_{\text{FRDD}}) &\leq d_{\mathcal{H}}(\tau_{\text{FRDD}}, \tau_{\text{FRDD}}^*) + d_{\mathcal{H}}(\tau_{\text{FRDD}}^*, \widehat{\tau}_{\text{FRDD}}) \\ &\leq \frac{\left\{ d_{\mathcal{H}}(\Psi(\mu_{1,\oplus}^{\dagger}), \Psi(\nu_{1,\oplus})) + d_{\mathcal{H}}(\Psi(\mu_{0,\oplus}^{\dagger}), \Psi(\nu_{0,\oplus})) \right\}}{|m_1 - m_0|} \\ &\quad + \frac{\left\{ d_{\mathcal{H}}(\Psi(\nu_{1,\oplus}), \Psi(\widehat{\nu}_{1,\oplus})) + d_{\mathcal{H}}(\Psi(\nu_{0,\oplus}), \Psi(\widehat{\nu}_{0,\oplus})) \right\}}{|\widehat{m}_1 - \widehat{m}_0|} \\ &\quad + \left| \frac{1}{\widehat{m}_1 - \widehat{m}_0} - \frac{1}{m_1 - m_0} \right| d_{\mathcal{H}}(\Psi(\nu_{0,\oplus}), \Psi(\nu_{1,\oplus})) \\ &= \frac{\left\{ d(\mu_{1,\oplus}^{\dagger}, \nu_{1,\oplus}) + d(\mu_{0,\oplus}^{\dagger}, \nu_{0,\oplus}) \right\}}{|m_1 - m_0|} \\ &\quad + \frac{\left\{ d(\nu_{1,\oplus}, \widehat{\nu}_{1,\oplus}) + d(\nu_{0,\oplus}, \widehat{\nu}_{0,\oplus}) \right\}}{|\widehat{m}_1 - \widehat{m}_0|} + \left| \frac{1}{\widehat{m}_1 - \widehat{m}_0} - \frac{1}{m_1 - m_0} \right| d(\nu_{0,\oplus}, \nu_{1,\oplus}). \end{aligned} \tag{C.5}$$

Furthermore,

$$\begin{aligned} \frac{1}{\widehat{m}_1 - \widehat{m}_0} - \frac{1}{m_1 - m_0} &= \frac{(m_1 - \widehat{m}_1) - (m_0 - \widehat{m}_0)}{(\widehat{m}_1 - \widehat{m}_0)(m_1 - m_0)} \\ &= \frac{O_p(r_n)}{(m_1 - m_0 + o_p(1))(m_1 - m_0)} = O_p(r_n). \end{aligned} \tag{C.6}$$

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

$$d(\mu_{z,\oplus}^\dagger, \nu_{z,\oplus}) = O(h_z^{2/(\beta_1-1)}), \quad d(\widehat{\nu}_{z,\oplus}, \nu_{z,\oplus}) = O_p((nh_z)^{-\frac{1}{2(\beta_2-1)}}), \quad z \in \{0, 1\}. \quad (\text{C.7})$$

Combining (C.5), (C.6), and (C.7), we obtain

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

#### C.4. Proof of Proposition 6.1.

Under Assumption 6.2 (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 6.2 (ii), we can show

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

#### C.5. Proof of Proposition 6.2.

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

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

Therefore,

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

Applying a similar argument as in the proof of Theorem 6.2 yields the desired result.

### C.6. Proof of Theorem 6.3.

Applying the same argument as used to obtain (C.1) and (C.2),

$$\nu_{1,\omega}^\sharp = \mathbb{E}[\text{Log}_\omega(Y(T(1)))|R=c] = \lim_{r \downarrow c} \mathbb{E}[\text{Log}_\omega(Y)|R=r], \quad (\text{C.8})$$

$$\nu_{0,\omega}^\sharp = \mathbb{E}[\text{Log}_\omega(Y(T(0)))|R=c] = \lim_{r \uparrow c} \mathbb{E}[\text{Log}_\omega(Y)|R=r]. \quad (\text{C.9})$$

Observe

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

and

$$\begin{aligned} \mu_{z,\omega}^\sharp &= \mathbb{E}[\text{Log}_\omega(Y(T(z)))|T(1) > T(0), R=c] \\ &= \mathbb{E}[\text{Log}_\omega(Y(T(z)))|T(1) = T(0), R=c] + \frac{\nu_{z,\omega}^\sharp - \mathbb{E}[\text{Log}_\omega(Y(T(z)))|T(1) = T(0), R=c]}{\mathbb{P}(T(1) > T(0)|R=c)}. \end{aligned} \quad (\text{C.10})$$

This yields

$$\tau_{\text{FRDD}}^\sharp = \mu_{1,\omega}^\sharp - \mu_{0,\omega}^\sharp = \frac{\nu_{1,\omega}^\sharp - \nu_{0,\omega}^\sharp}{\mathbb{P}(T(1) > T(0)|R=c)} = \frac{\nu_{1,\omega}^\sharp - \nu_{0,\omega}^\sharp}{\mathbb{E}[T(1)|R=c] - \mathbb{E}[T(0)|R=c]}.$$

Combining this with (C.3), (C.8), and (C.9) leads to the desired result.

### C.7. Proof of Theorem 6.4.

The proof is omitted as it is similar to that of Theorem 6.2.

### C.8. Proof of Proposition 6.3.

The proof is omitted as it is similar to that of Proposition 6.1.

### C.9. Proof of Proposition 6.4.

Observe that  $d_G(\tau_{\text{GFRD}}^\sharp, \widehat{\tau}_{\text{GFRD}}^\sharp) \leq C\{d(\mu_{0,\omega,\oplus}^\sharp, \widehat{\mu}_{0,\omega,\oplus}^\sharp) + d(\mu_{1,\omega,\oplus}^\sharp, \widehat{\mu}_{1,\omega,\oplus}^\sharp)\}$  and

$$\begin{aligned} &d(\mu_{z,\omega,\oplus}^\sharp, \widehat{\mu}_{z,\omega,\oplus}^\sharp) \\ &= d\left(\text{Exp}_\omega\left(\nu_\omega^\sharp + \frac{\nu_{z,\omega}^\sharp - \nu_\omega^\sharp}{m_1 - m_0}\right), \text{Exp}_\omega\left(\widehat{\nu}_\omega^\sharp + \frac{\widehat{\nu}_{z,\omega}^\sharp - \widehat{\nu}_\omega^\sharp}{\widehat{m}_1 - \widehat{m}_0}\right)\right) \\ &\leq C_\omega d_E\left(\nu_\omega^\sharp + \frac{\nu_{z,\omega}^\sharp - \nu_\omega^\sharp}{m_1 - m_0}, \widehat{\nu}_\omega^\sharp + \frac{\widehat{\nu}_{z,\omega}^\sharp - \widehat{\nu}_\omega^\sharp}{\widehat{m}_1 - \widehat{m}_0}\right) \\ &\leq C_\omega \left\{ \left(1 + \frac{1}{|\widehat{m}_1 - \widehat{m}_0|}\right) d_E(\nu_\omega^\sharp, \widehat{\nu}_\omega^\sharp) + \frac{1}{|\widehat{m}_1 - \widehat{m}_0|} d_E(\nu_{z,\omega}^\sharp, \widehat{\nu}_{z,\omega}^\sharp) \right\} \\ &\quad + C_\omega \left| \frac{1}{m_1 - m_0} - \frac{1}{\widehat{m}_1 - \widehat{m}_0} \right| d_E(\nu_{z,\omega}^\sharp, \nu_\omega^\sharp), \quad z \in \{0, 1\}. \end{aligned}$$

Then

$$\begin{aligned} &d_G(\tau_{\text{GFRD}}^\sharp, \widehat{\tau}_{\text{GFRD}}^\sharp) \\ &\leq CC_\omega \left(1 + \frac{1}{|\widehat{m}_1 - \widehat{m}_0|}\right) \{2d_E(\nu_\omega^\sharp, \widehat{\nu}_\omega^\sharp) + d_E(\nu_{0,\omega}^\sharp, \widehat{\nu}_{0,\omega}^\sharp) + d_E(\nu_{1,\omega}^\sharp, \widehat{\nu}_{1,\omega}^\sharp)\} \end{aligned}$$

$$+ CC_\omega \left| \frac{1}{m_1 - m_0} - \frac{1}{\widehat{m}_1 - \widehat{m}_0} \right| \{d_E(\nu_{1,\omega}^\#, \nu_\omega^\#) + d_E(\nu_{0,\omega}^\#, \nu_\omega^\#)\}.$$

Applying a similar argument in the proof of Theorem 6.2 then implies the desired result.

## 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 edition.
- 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(4):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(1):328–347.
- Bertanha, M. (2020). Regression discontinuity design with many thresholds. *Journal of Econometrics*, 218(1):216–241.
- Black, S. E. (1999). Do better schools matter? parental valuation of elementary education. *The Quarterly Journal of Economics*, 114(2):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. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779.
- Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Cattaneo, M. D. and Escanciano, J. C. (2017). *Regression Discontinuity Designs: Theory and Applications*, volume 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(536):1941–1952.
- Cattaneo, M. D. and Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, 14(1):821–851.
- Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2020). The Regression Discontinuity Design. In *The SAGE Handbook of Research Methods in Political Science and International Relations*, pages 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*.
- Chen, Y., Lin, Z., and Müller, H.-G. (2023). Wasserstein regression. *Journal of the American Statistical Association*, 118(542):869–882.

- Chen, Y. and Müller, H.-G. (2022). Uniform convergence of local Fréchet regression, with applications to locating extrema and time warping for metric-space valued trajectories. *Annals of Statistics*, 50(3):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(1):58–97.
- Cherian, A. and Sra, S. (2016). *Positive Definite Matrices: Data Representation and Applications to Computer Vision*, pages 93–114. Springer, Cham.
- Dai, X. (2022). Statistical inference on the Hilbert sphere with application to random densities. *Electronic Journal of Statistics*, 16(1):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(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(2):757–792.
- 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(1):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(3):903–920.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.
- Gendron-Carrier, N., Gonzalez-Navarro, M., Polloni, S., and Turner, M. A. (2022). Subways and urban air pollution. *American Economic Journal: Applied Economics*, 14(1):164–96.
- Gunsilius, F., Hsieh, M. H., and Lee, M. J. (2024). Tangential Wasserstein projections. *Journal of Machine Learning Research*, 25(69):1–41.
- Gunsilius, F. F. (2023). Distributional synthetic controls. *Econometrica*, 91(3):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(1):201–209.
- Hansen, B. (2022). *Econometrics*. Princeton University Press.
- 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*, volume 997. John Wiley & Sons.
- Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.
- Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.
- 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.
- Keele, L. J. and Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155.
- Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical analysis of network data with R*, volume 65. Springer New York.

- Kurusu, D., Zhou, Y., Otsu, T., and Müller, H.-G. (2024). Geodesic causal inference. *arXiv:2406.19604*.
- Kurusu, D., Zhou, Y., Otsu, T., and Müller, H.-G. (2025). Geodesic synthetic control methods for random objects and functional data. *arXiv:2505.00331*.
- 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(4):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(2):378–398.
- McCann, R. J. (1997). A convexity principle for interacting gases. *Advances in Mathematics*, 128(1):153–179.
- Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics*, 47(2):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*, volume 426 of *Springer Series in Statistics*. Springer New York, NY, 2nd edition.
- 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(3):351–375.
- Scealy, J. and Welsh, A. (2014). Colours and cocktails: Compositional data analysis. *Australian & New Zealand Journal of Statistics*, 56(2):145–169.
- Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):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(5):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(1):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*, pages 1–8.
- 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(1):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 edition.

- 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.
- 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(320):1–41.
- Zhou, Y. and Müller, H.-G. (2024). Wasserstein regression with empirical measures and density estimation for sparse data. *Biometrics*, 80(4):ujae127.

(D. Kurisu) CENTER FOR SPATIAL INFORMATION SCIENCE, THE UNIVERSITY OF TOKYO, 5-1-5, KASHIWANOHA, KASHIWA-SHI, CHIBA 277-8568, JAPAN.  
*Email address:* `daisukekurisu@csis.u-tokyo.ac.jp`

(Y. Zhou) DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, DAVIS, ONE SHIELDS AVENUE, DAVIS, CA, 95616, USA.  
*Email address:* `ydzhou@ucdavis.edu`

(T. Otsu) DEPARTMENT OF ECONOMICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON, WC2A 2AE, UK.  
*Email address:* `t.otsu@lse.ac.uk`

(H.-G. Müller) DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, DAVIS, ONE SHIELDS AVENUE, DAVIS, CA, 95616, USA.  
*Email address:* `hgmueLLer@ucdavis.edu`