Submitted to the Annals of Statistics

SINGLE INDEX FRÉCHET REGRESSION

BY SATARUPA BHATTACHARJEE[1,a] AND HANS-GEORG MÜLLER[2,b]

[1]*Department of Statistics, Pennsylvania State University,* [a]*sfb5992@psu.edu*

[2]*Department of Statistics, University of California, Davis,* [b]*hgmueller@ucdavis.edu*

Single index models provide an effective dimension reduction tool in regression, especially for high dimensional data, by projecting a general multivariate predictor onto a direction vector. We propose a novel single-index model for regression models where metric space-valued random object responses are coupled with multivariate Euclidean predictors. The responses in this regression model include complex, non-Euclidean data, including covariance matrices, graph Laplacians of networks, and univariate probability distribution functions, among other complex objects that lie in abstract metric spaces. While Fréchet regression has proved useful for modeling the conditional mean of such random objects given multivariate Euclidean vectors, it does not provide for regression parameters such as slopes or intercepts, since the metric space-valued responses are not amenable to linear operations. As a consequence, distributional results for Fréchet regression have been elusive. We show here that for the case of multivariate Euclidean predictors, the parameters that define a single index and projection vector can be used to substitute for the inherent absence of parameters in Fréchet regression. Specifically, we derive the asymptotic distribution of suitable estimates of these parameters, which then can be utilized to test linear hypotheses for the parameters, subject to an identifiability condition. Consistent estimation of the link function of the single index Fréchet regression model is obtained through local linear Fréchet regression. We demonstrate the finite sample performance of estimation and inference for the proposed single index Fréchet regression model through simulation studies, including the special cases where responses are probability distributions and graph adjacency matrices. The method is illustrated for resting-state functional Magnetic Resonance Imaging (fMRI) data from the ADNI study.

1. Introduction. Modeling the regression relationship between a real-valued response and a multivariate Euclidean predictor vector $\mathbf{X}$ corresponds to specifying the form of the conditional means $(\mathbf{x})$ $\mathbb{E}(|\mathbf{X} \mathbf{x})$ Higher dimensionality of $\mathbf{X}$ can be problematic when one is interested to go beyond the standard multiple linear models and aims for a nonparametric estimation of $(\mathbf{x})$ This provides strong motivation to consider regression models that provide dimension reduction. Single index models are one of the most popular approaches to achieve this under the assumption that the influence of the predictors on the response can be collapsed to a single index, i.e., a projection on a specific direction, complemented by a nonparametric link function. This reduces the predictors to a univariate index while still capturing relevant features and since the nonparametric link function acts only on a one-dimensional index, these models are not subject to the curse of dimensionality. The single index model generalizes linear regression, where the link function is the identity. For a real-valued response, and a -dimensional predictor $\mathbf{X}$, the semiparametric single index regression model is given by

$$\mathbb{E}(|\mathbf{X} \mathbf{x}) \quad \mathbb{E}(|\mathbf{X}^\top) \quad ( ) \tag{1.1}$$

MSC2020 subject classifications: Primary 62R20, 62G10; secondary 62G05, 62G20.
Keywords and phrases: Non-Euclidean data, Dimension reduction, Random objects, M-estimator, Inference.

In model (1.1), the dependence between and $\mathbf{X}$ characterized by the conditional mean, is summarized by the parameter vector and the link function .

The function is nonparametric and thus includes location and level changes, and therefore the vector $\mathbf{X}$ cannot include a constant that would serve as an intercept. For identifiability reasons, is often assumed to be a unit vector with a positive first coordinate. A second approach is to require one component to equal one. This presupposes that the component that is set to equal 1 indeed has a non-zero coefficient [44, 15]. Model (1.1) is only meaningful if the Euclidean predictor vector $\mathbf{X}$ is of dimension or larger. If $\mathbf{X}$ is one-dimensional, the corresponding special case of the model is the one-dimensional nonparametric regression $\mathbb{E}(\mid) \quad ( )$, which does not feature any parametric component.

The classical single index regression model with Euclidean responses has attracted attention from the scientific community for a long time due to its flexibility and the interpretability of the (linear) coefficients and flexibility, owing to the nonparametric link function, as well as due to its wide applicability in many scientific fields. The coefficient that defines the single index $\mathbf{x}^\top$ along with the shape of the nonparametric component characterizes the relationship between the response and the predictor. The parametric component is of primary interest for inference in this model. The problem of recovering the true direction can be viewed as a subclass of sufficient dimension reduction (SDR) techniques, where identifying the central subspace of $\mathbf{X}$ that explains most of the variation in has been a prime target [42, 14, 40].

In addition to sufficient dimension reduction techniques, various related approaches to estimate in (1.1) have been studied. These include projection pursuit regression (PPR) [24, 28], average derivatives [30, 62], sliced inverse regression (SIR) [41], conditional minimum average variance estimation (MAVE) [69] and various other methods [68, 67]. These approaches have focused on the nonparametric estimation of the link function to recover the index parameter in (1.1) [29, 32, 31], partially linear versions [10, 71] and various noise models [11, 66]. Inference for the index parameters has also been well studied [20, 43, 25] for the classical single index model.

Various extensions of single index regression have been considered more recently [75, 36], including models with multiple indices or high-dimensional predictors [78, 76, 38], censored data [46], and longitudinal and functional data as predictors [34, 12, 21, 50]. However, none of these extensions has covered situations where responses are not in a Euclidean vector space, even though this case is increasingly important for data analysis. Two very recent exceptions are [70] and [73], who considered extending sufficient dimension reduction approaches for the case of random objects. The overall lack of available methodology for single-index models with random object responses motivates our approach. Non-Euclidean complex data structures arising in areas such as biological or social sciences are becoming increasingly common, due to technological advances that have made it possible to record and efficiently store sensor data and images [56], shapes [61] or networks [64]. For example, one might be interested in functional connectivity, quantified in the form of correlation matrices obtained from neuroimaging studies, to study the effect of predictors on brain connectivity, an application that we explore in Section 5.1.

Other examples of general metric space objects include probability distributions [18], such as age-at-death distributions as observed in demography or network objects, such as internet traffic networks. Such "object-oriented data" [47] or "random objects" [48] can be viewed as random variables taking values in a separable metric space that is devoid of a vector space structure and where only pairwise distances between the observed data are available. Almost all existing methodology for single-index models as briefly reviewed above assumes that one has Euclidean responses, and these methods rely in a fundamental way on the vector space structure of the space where the responses reside. When there is no linear structure, a new methodology is needed and this paper contributes to this development.

A natural measure of location for random elements of a metric space is the Fréchet mean [23], which is a direct generalization of the standard mean and is defined as the element of the metric space for which the expected squared distance to all other elements, known as the Fréchet function, is minimized. Depending on the space and metric, Fréchet means may or may not exist as unique minimizers of the Fréchet function. Fréchet regression is an extension of Fréchet means to the notion of conditional Fréchet means, and local as well as global versions have been recently studied in several papers [55, 53, 57, 58, 4].

Global Fréchet regression is a generalization of linear regression for random object responses. In analogy to classical linear regression, it features a restrictive structural model assumption. While the local linear version of Fréchet regression is more flexible, it suffers from the curse of dimensionality as the dimension of the predictors increases. Further, neither version of the Fréchet regression incorporates an interpretable inference regime. In this paper, we introduce (single) Index Fréchet Regression (IFR) to facilitate inference in the context of Fréchet regression when the response variable is a random object lying in general metric space and the predictor is a -dimensional Euclidean vector $\mathbf{X}$ with Our goal is to develop an extension of the conventional estimation and inference paradigm for single-index models for this challenging case. It is assumed that the conditional expectation (Fréchet regression) of depends on the predictor vector $\mathbf{X}$ only through the projection or index $\mathbf{X}^\top$ for a parameter vector $\in \in \mathbb{R}$ Since there is no notion of direction or sign in a general metric space, we interpret the index parameter in the proposed index Fréchet regression model (IFR) as the direction in the predictor space along which the variability of the response is maximized. The semiparametric framework provided by the proposed single index model facilitates stable estimation and interpretable inference.

It turns out to be useful to cast the direction estimation problem in the framework of M-estimation for an appropriate objective function and to use empirical process theory to show consistency of the proposed estimate. We derive an asymptotic normality result for these estimators under mild assumptions on the metric space and the unknown link function by utilizing an appropriate version of recent results of [13] concerning local linear Fréchet regression estimators. Under suitable regularity assumptions, the asymptotic distribution of the estimated index parameter can then be harnessed to construct a Wald-type statistic to conduct inference. Combining this with an auxiliary result on the asymptotic convergence of the estimated covariance matrix makes it possible to employ a bootstrap method to obtain inference in finite sample situations.

When we finalized this work, we became aware that independently and simultaneously another group also developed an approach for single index Fréchet regression [26]. We wish to emphasize that this paper was not in any way influenced by this parallel development (with preprints becoming available within days of each other).

The paper is organized as follows: The basic setup is defined in Section 2 and the theory on the asymptotic behavior of the index parameter is provided in Section 3, with a focus on results for inference. The index vector is assumed to lie on a hyper-sphere, with a non-negative first element to facilitate identifiability. Then it is natural to quantify the performance of the proposed estimators by the geodesic distances between the estimated and true directions. The results of simulation studies with various types of random objects as responses are reported in Section 4 with additional results in the Supplement [5]. In Section 5 we apply the methods to infer and analyze the effect of age, sex, total Alzheimer's brain score and the stage of Alzheimer's Disease on the brain connectivity of patients with dementia. Brain connectivity is derived from fMRI signals of brain regions of interest [63] and quantified in the form of correlation matrix objects. We present additional illustrations for human mortality data as distributional objects and mood data of unemployed workers as compositional objects, with details in the Supplement [5]. A brief discussion follows in Section 6.

2. Model and Estimation Methods. In all of the following, ( ) is a totally bounded metric space with metric and a probability measure The random objects take values in . This is coupled with a -dimensional real-valued predictor $\mathbf{X}$ Throughout we will use bold letters to denote multivariate real vectors. The conditional Fréchet mean of given $\mathbf{X}$ is a generalization of $\mathbb{E}(|\mathbf{X} \mathbf{x})$ to metric spaces, defined as the argmin of $\mathbb{E}(( )|\mathbf{X} \mathbf{x}) \in$ [55], i.e.,

$$\mathbb{E}_+(|\mathbf{X} \mathbf{x}) \in \mathbb{E}(( )|\mathbf{X} \mathbf{x}) \tag{2.1}$$

Evaluated at the minimizer, the objective function in (2.1) is the corresponding generalized measure of dispersion around the conditional Fréchet mean and can be viewed as a conditional Fréchet function.

As discussed earlier, obtaining inference for Fréchet regression is an elusive goal, for both the more restrictive global as well as the more flexible but the curse of dimensionality afflicted local version of Fréchet regression. To move towards inference, we propose here a more structured model, inspired by its Euclidean single index equivalent in (1.1), given by

$$\mathbb{E}_+(|\mathbf{X} \mathbf{x}) \quad {}_+(\mathbf{x}^\top) \tag{2.2}$$

where is the true direction parameter of interest. Model (1.1) emerges as a special case of model (2.2) for a Euclidean response, as the conditional Fréchet mean coincides with the conditional expectation $\mathbb{E}(|\mathbf{X})$ for the choice of the absolute Euclidean distance metric for the case $\mathbb{R}$ In other words, the conditional Fréchet mean is assumed to be a function of in such a way that the distribution of only depends on $\mathbf{X}$ only through the index $\mathbf{X}^\top$ that is, $\perp \mathbb{E}_+(|\mathbf{X})|(\mathbf{X}^\top)$ Thus

$$\mathbb{E}_+(|\mathbf{X} \mathbf{x}) \quad \mathbb{E}_+(|\mathbf{X}^\top) \quad {}_+( )$$

and invoking local linear nonparametric Fréchet regression for the one-dimensional index promises to overcome the curse of dimensionality problem. For projections $\mathbf{X}^\top \in \mathcal{T}_{\boldsymbol{\theta}_0} \mathbb{R}$, which depend on , we consider predictors $\mathbf{X}$ with bounded norm such that $\mathcal{T}_{\boldsymbol{\theta}_0} \mathcal{T}$, where $\mathcal{T}$ is a compact interval on $\mathbb{R}$ We note that the link function, for given $\in$ $_+ \mathcal{T}_{\boldsymbol{\theta}_0} \mapsto ( )$ in the true model depends on the multivariate predictor $\mathbf{X}$ $\mathbf{x}$ only through the single-index $\mathbf{x}^\top$, as well as on the direction vector implicitly. Thus, explicitly characterizing this dependence, we define the Index Fréchet Regression (IFR) model for random object response and Euclidean predictor $\mathbf{X}$ as

$${}_+( ) \in \mathbb{E}(( )|\mathbf{X}^\top) \tag{2.3}$$

The coefficient $\in \mathbb{R}$ is the quantity of interest for the single index Fréchet model owing to its interpretability by quantifying the contribution of each predictor component. More generally, the quantity in model (2.3) can be evaluated for any direction vector $\in$ by

$${}_+(\mathbf{x}^\top) \in \mathbb{E}(( )|\mathbf{X}^\top \mathbf{x}^\top) \tag{2.4}$$

In the Euclidean case, identifiability conditions for the direction parameter have been widely discussed in the literature [10, 44, 15, 77]. We assume the parameter space to be constrained in order to ensure that in the representation (2.4) is uniquely defined, where

$$\{ ( )^\top \| \| \in \mathbb{R} \} \tag{2.5}$$

We first choose an identifiable parametrization that transforms the boundary of a unit ball in $\mathbb{R}$ to the interior of a unit ball in $\mathbb{R}^( )$. By eliminating the parameter space can be rearranged to $\{(( ) /)^\top \}$ This re-parametrization is

the key to analyzing the asymptotic properties of the estimates for and also facilitating efficient computation. The true parameter is then partitioned into ( )$^\top$ where ( )$^\top$ We estimate the ( ) dimensional vector in the single-index model and then use ( ) / to obtain

PROPOSITION 1 (Identifiability of model (2.3)). *Suppose* $_+(\boldsymbol{x})$ $\mathbb{E}_+($ $|\boldsymbol{X}$ $\boldsymbol{x})$*, that the support of* $_+( )$ *is a convex bounded set with at least one interior point and that* $_+( )$ *is a non-constant continuous function on If*

$$_+(\boldsymbol{x}) \quad _+(\ ^\top\boldsymbol{x} \ ) \quad _+(\ ^\top\boldsymbol{x} \ ) \ \textit{for all } \boldsymbol{x} \in$$

for some continuous object-valued link functions $_+$ *and* $_+$*, and some* $\in$ *where is as described in* (2.5)*. Then and* $_+$ $_+$ *on* $\{ \ ^\top\boldsymbol{x} | \boldsymbol{x} \in \ \}$

The above result can be proved using a similar argument as given in the proof of Theorem of [44].

Scrutinizing the special case of a Euclidean response in model (1.1), the variation in is seen to result from the variation in $\mathbf{X}^\top$ as well as from the variation in the error term in the model, denoted by [33]. On the contour line $\mathbf{X}^\top$, the variability in only results from the variability in . Along contour lines $\mathbf{X}^\top$ for , $\mathbf{X}^\top$ is not constant and therefore the variability in along the contour lines $\mathbf{X}^\top$ is due to both the variation in $\mathbf{X}^\top$ and in . Since $|\mathbf{X}^\top$ measures the variability in on a contour line $\mathbf{X}^\top$, one can characterize as the minimizer of the objective function ( ), where ( ) $\mathbb{E}($ ($|\mathbf{X}^\top$)) and $_{\boldsymbol{\theta}\in}$ ( ) The constraint $^\top$ with the first element of the index , ensures the identifiability of the objective function. Defining an equivalence class of the parameter vector $_{\boldsymbol{\theta}_0}$ $\{ \in$ $(\mathbf{x}^\top$) $(\mathbf{x}^\top$) a.e. in $\mathbf{x}$ for some $\}$ for $\notin$ $_{\boldsymbol{\theta}_0}$, one has ( ) ( )

To recover the true direction of the single index from model (2.3), the conditional variance of given $\mathbf{X}$ $\mathbf{x}$ for a real-valued response can be replaced by the conditional Fréchet variance ($_+(\mathbf{x}^\top$)) for any given unit orientation vector Thus, for a general object response $\in ( )$ can alternatively be expressed as

$$\begin{aligned} &( ) \ \text{where} \ ( ) \ \mathbb{E} \ _+(\mathbf{X}^\top \ ) \\ &\boldsymbol{\theta} \in \\ &_+(\ ) \ \in \ (\ ) \ \text{with} \ (\ ) \ \mathbb{E} \ (\ )|\mathbf{X}^\top \end{aligned} \tag{2.6}$$

This corresponds to finding the true parameter through the optimal direction that maximizes the total variability of the responses, an idea developed in [33] for the case of Euclidean responses. Instead of choosing the parameter minimizing the expected variance explained by the single index $\mathbf{X}^\top$, for object responses the new goal is to choose the parameter minimizing the expected Fréchet variance.

To recover from the representation (2.6), one needs to also estimate the conditional Fréchet mean, as in the IFR model (2.3), for which we employ the local linear Fréchet regression estimate [55]. The idea is as specified below. We approximate the conditional Fréchet mean $_+$ in (2.6) by a locally weighted Fréchet mean that we refer to as intermediate weighted Fréchet mean. The weights for this intermediate Fréchet mean are derived from a weight function ( ) that characterizes the effect on the predictors via a chosen kernel function ( ) and a bandwidth parameter such that ( ) (/) (/) For any given unit direction index this intermediate localized weighted Fréchet mean is

$$_+(\ ) \quad _{\in} \quad (\ ) \ \text{with} \ (\ ) \ \mathbb{E} \ (\mathbf{X}^\top \ ) \ (\ ) \tag{2.7}$$

where

(2.8)

$$(\mathbf{X}^\top \quad) \quad \frac{\quad}{(\quad)} \quad (\mathbf{X}^\top \quad)[\quad(\quad) \quad (\quad)(\mathbf{X}^\top \quad)]$$

$$(\quad) \quad \mathbb{E}(\quad(\mathbf{X}^\top \quad)(\mathbf{X}^\top \quad)) \qquad (\quad) \quad (\quad) \quad (\quad) \quad (\quad)$$

and $(\quad) \quad (\quad) \quad (\,)$ for all $\quad$ and $\quad$; note that $\quad_+(\quad)$ is a non-random population quantity.

Suppose we observe a random sample of paired observations $(\mathbf{X} \quad)$, where $\mathbf{X}$ is a dimensional Euclidean predictor and is an object response situated in a metric space $(\quad)$ Using the form of the intermediate target in (2.7) and replacing the auxiliary parameters by their corresponding empirical estimates, the local Fréchet regression estimator at a given value of the single index for a given direction parameter $\in$ is defined as

(2.9) $\quad_+(\quad) \qquad_{\in} \quad (\quad)$ with $\quad(\quad) \quad - \quad \widehat{\ }(\mathbf{X}^\top \quad) \quad (\quad)$

where

(2.10)

$$\widehat{\ }(\mathbf{X}^\top \quad) \quad \frac{\quad}{(\quad)} \quad (\mathbf{X}^\top \quad)[\quad(\quad) \quad (\quad)(\mathbf{X}^\top \quad)]$$

$$(\quad) \quad - \quad (\mathbf{X}^\top \quad)(\mathbf{X}^\top \quad) \qquad (\quad) \quad (\quad) \quad (\quad) \quad (\quad)$$

The following assumption pertains to the existence and uniqueness of the Fréchet means in (2.6) and (2.9).

(A0) The conditional and weighted Fréchet means in (2.6), (2.7), and (2.9) are well defined, i.e., they exist and are unique, the latter one almost surely. Further, for all $\in$ such that , $(\quad \in \mathbb{R} \quad \_+(\mathbf{X}^\top \quad) \quad \_+(\mathbf{X}^\top \quad))$

Existence and uniqueness of Fréchet means depend on the nature of the metric space and the underlying probability measure and will be discussed further after (A4) in section 3. For example, in the case of Euclidean responses, Fréchet means coincide with the usual means for random vectors with finite second moments. In the case of Riemannian manifolds, the existence, uniqueness, and convexity of the center of mass are guaranteed under certain conditions [1, 52]. In a space with a negative or zero curvature, or in a Hadamard space unique Fréchet means always exist [6, 7, 51, 37]. The existence of unique Fréchet means in assumption (A0) is satisfied for the space $(\quad)$ of univariate probability distributions with the 2-Wasserstein metric and also for the space $(\quad)$ of covariance matrices with the Frobenius metric [55].

Assume that for all unit direction vectors the support $\mathcal{T}_{\boldsymbol{\theta}}$ of $\mathbf{X}^\top$ is compact, where all $\mathcal{T}_{\boldsymbol{\theta}}$ are subsets of a fixed interval. For the derivation of distributional limit results, one needs to establish sufficiently fast convergence of the estimated means. This challenge can be overcome by partitioning the interval where the linear predictor is situated. Specifically, we partition $\mathcal{T}_{\boldsymbol{\theta}}$ into equal-width non-overlapping bins $\{\quad\}$ where data falling in different bins are independent and identically distributed. We denote by $\mathbf{X}$ and the representative data points in the th bin, The number of bins depends on the sample size , where the choice of the sequence $(\,)$ is discussed in (A4) in section 3 below. The proposed estimator for the true direction in (2.6) is then given by

(2.11) $\quad \hat{=} \qquad_{\boldsymbol{\theta} \in} \quad (\,)$ where $\quad(\,) \quad —\quad \_+(\mathbf{X}^\top \quad)$

Here $_{+}(\mathbf{X}^{\top}\quad)$ is the local linear Fréchet regression estimator, constructed based on the sample $(\mathbf{X}\quad)$, and evaluated at each sample point of the binned sample $(\mathbf{X}\quad)$ as described in (2.9) and (2.10). We also require an intermediate quantity that corresponds to the empirical version of $(\ )$ in (2.6), defined as

$$(2.12)\qquad \bar{\quad}\qquad\qquad (\ ) \text{ where } (\ ) \text{ —} \qquad\qquad {}_{+}(\mathbf{X}^{\top}\quad)$$
$$\boldsymbol{\theta}\in$$

The bandwidth $(\ )$ is a tuning parameter and features in the rate of convergence of $_{+}$ to $_{+}$. We note that another possible estimator for $_{+}$ could be obtained by applying global Fréchet regression. This alternative estimator for the unknown link function in the IFR model (2.3) does not depend on a tuning parameter as is needed for locally linear Fréchet regression but is considerably less flexible.

3. Theory. The unknown quantities that constitute the Index Fréchet Regression (IFR) model consist of the nonparametric link function and the index parameter, and thus the asymptotic properties of the estimate of the true unit direction rely on those of the estimates of the link function (based on local linear Fréchet regression) and the index parameter (through an M-estimator of the criterion function in (2.6)). The metric space $(\quad)$ is assumed to be totally bounded with diameter , hence separable. In order to obtain the right bound on the metric entropy of the space , the boundedness assumption is crucial. While boundedness imposes a restriction that is not needed in the Euclidean case, it is a quite feasible assumption in general metric spaces, since, for commonly observed non-Euclidean objects, the underlying metric space satisfies the total boundedness property. Examples include the Wasserstein-2 space of one-dimensional distributions with compact support and the space of spheres with the geodesic metric and positive semi-definite matrices with Frobenius or power metric.

We make the following assumption on the objective function $(\ )$ in (2.6).

(A1) There exist and such that whenever $\|\quad\|$ for $\in$, we have $(\ )\quad(\ )\quad\|\quad\|$

The above condition on the curvature of the objective function is standard in the empirical process theory literature and controls the behavior of near the minimum in order to obtain rates of convergence. In addition, with regard to the quantities in (2.6), (2.9), and (2.11) we require the following assumptions.

(A2) The link function $_{+}$ is Lipschitz continuous, that is, there exists a real constant such that, for all $\mathbf{x}$ with a bounded norm, and for all $\in$

$$_{+}(\mathbf{x}^{\top}\quad)\quad{}_{+}(\mathbf{x}^{\top}\quad)\quad\|\quad\|$$

(A3) For any given direction the univariate index variable $\mathbf{X}^{\top}$ is assumed to have a density $_{\boldsymbol{\theta}}(\ )$ with a compact support $\mathcal{T}_{\boldsymbol{\theta}}\quad\mathcal{T}$ for some bounded $\mathcal{T}\quad\mathbb{R}$ We denote the space of predictors for which this holds by $\mathcal{X}\quad\mathbb{R}$

(A4) For that satisfy assumption (U3) in the Supplement [5] and any , let

$$(3.1)\qquad \{\ ^{/(\ _{1}\ )}(\quad)\ ^{/(\ (\ _{2}\ )\ )}(\quad(\quad\quad)\ )^{/(\ (\ _{2}\ ))}\}$$

The number of non-overlapping bins $(\ )$ as defined in Section 2, is such that $(\ )\to\infty$ and $\to$ as $\to\infty$

We note that for which is the most common situation, reduces to

$$\{\quad(\quad)\ ^{/(\quad)}(\quad(\quad\quad)\ )^{/}\}$$

Assumption (A2) is a strong form of uniform continuity for the link function. Intuitively, it limits how fast the object $_+$ can change, introducing a concept of smoothness in the link function for the IFR model (2.3). Lipschitz continuity is a natural choice of morphisms between metric spaces. This assumption is slightly stronger than the assumption of a strictly monotone link function that is commonly used in classical single index literature to ensure identifiability. Since the domain of the link function is compact, in the Euclidean response case, our assumption would translate to having a strictly monotone continuous link function with a bounded derivative. Essentially, assumption (A2) is weaker than a derivative condition and stronger than assuming only the strict monotonicity of the link function. Assumption (A3) is basic. The predictors needed for the nonparametric Fréchet regression are required to be randomly distributed over the domain where the function is to be estimated, and on average, to become denser as more data are collected. Sufficient for this to be satisfied is that there is at least one continuous predictor and the predictors are bounded. Assumption (A4) is required for the rate of convergence and limit distribution results, for which we involve the binning device, and it connects the uniform rate of convergence for the local linear Fréchet regression estimator as given in (3.1) with the number of bins .

For most types of random objects, such as those in the Wasserstein-2 space (the space of probability distributions equipped with the 2-Wasserstein distance) or the space of symmetric positive semidefinite matrices endowed with the Frobenius or power metric, one has in the definition of in assumption (A4) (see assumptions (U1)-(U3) in Section S.2. of the Supplement [5]). If one chooses the bandwidth sequence for the local linear Fréchet regression such that, for a given $(_1)/(_1 _2)$ then is of the order $\frac{1}{(\beta_1+2\beta_2-3+\varepsilon)}$ [13]. For this becomes $\frac{1}{3+\varepsilon}$ Any sequence $( )$ with – will then satisfy assumption (A4).

As an alternative characterization for the true direction parameter , an important property of the objective function $( )$ in (2.6) is as follows.

PROPOSITION 2. *Under assumptions (A0) and (A2),* $( )$ *in model* (2.6) *is a continuous function of* $\in$ *and* $\boldsymbol{\theta} \in$ $( )$

Additional assumptions (U1)-(U3) and (R1)-(R2) have been used previously in [55] and [13], though in a slightly weaker form, and can be found in Section S.2. of the Supplement [5]. These are regarding They concern the existence, uniqueness, and well separateness of the minimizers, the metric entropy condition in terms of the covering number, and the curvature of the metric space near the minimizers and are commonly used for the asymptotic analysis of M estimators utilizing empirical process theory [65], here specifically to establish consistency and uniform rate of convergence for the local Fréchet regression estimator in (2.11), uniform across the single-index values and the direction parameter. Uniformity over the single index value was already required in [13] to achieve uniform convergence of local linear Fréchet regression. In the single index model framework, there is a new parameter vector , the presence of which requires an additional uniformity requirement over . Assumptions (R1)-(R2) are commonly used in the local regression literature [60, 19].

We will make use of the following lemma, which is an appropriately modified version of a known result (Theorem of [13]), to deal with the link function when investigating the asymptotic convergence rates of the proposed IFR estimator.

LEMMA 1. *Under assumptions (U1)-(U3), (R1)-(R2) (see Supplement [5]) and if* $\to$ *such that* $( ) \to \infty$ *as* $\to \infty$ *for any*

$$\sup_{\boldsymbol{\theta}\in \ \in\mathcal{T}_{\bar{\boldsymbol{\theta}}}} (_+( ) \ _+( )) \qquad ( ) \tag{3.2}$$

where is as given in equation (3.1) *in assumption (A4).*

It is worth mentioning here that the binning approach is not required for basic consistency results without rates (Theorem 3.1 and Corollary 1). One can indeed re-define the criteria functions in (2.11) based on the whole sample ($\mathbf{X}$) as

$$\hat{\boldsymbol{\theta}} \quad \underset{\boldsymbol{\theta}\in}{} \quad (\ ) \quad \text{where} \quad (\ ) \quad - \quad {}_{+}(\ ^{\top} \quad)$$

and carry on with the same proof techniques to show consistency of $\hat{\boldsymbol{\theta}}$ to the true unit direction vector However, to prove rates of convergence and investigate the asymptotic behavior of the estimated parameter, we need to make use of the uniform convergence rate for local linear Fréchet regression, as given in Lemma 1. The binning step is necessary to reduce the effective sample size from to ( ), the latter being intrinsically tied by assumption (A4) to the uniform convergence rate The rate is effectively slower than $^{/}$, again by virtue of the uniform convergence rate for the local linear Fréchet regression estimator. One may alternatively consider global Fréchet regression to estimate the unknown link function ${}_{+}$, resulting in a near parametric rate of $^{/}$. However, the global Fréchet model may suffer from model-induced bias, since as a direct generalization of linear regression, it may be overly restrictive for random object responses. For a consistent unambiguous representation, we refer to the minimizers in (2.11) and (2.12) based on the binned samples as our quantities of interest throughout the rest of the manuscript.

For all of the following results, the basic assumptions (A0)-(A3) are assumed to be satisfied. We first demonstrate the consistency of the proposed estimator for the true index direction. All proofs can be found in Section S of the Supplement [5].

THEOREM 3.1. *Under the basic assumptions (A0)-(A3), and the technical assumptions (U1)-(U3), and (R1)-(R2) listed in Section S.2. of the Supplement [5],*

$$\hat{\boldsymbol{\theta}} \quad \longrightarrow \quad \textit{on}$$

where is as defined in (2.5).

Combining the consistency result for the direction vector in Theorem 3.1 with the uniform convergence of the local linear Fréchet regression estimator in Lemma 1 leads to the asymptotic consistency of the estimated single index regression (IFR) model.

COROLLARY 1. *Under the conditions required for Theorem 3.1, for any* $\boldsymbol{x} \in \mathcal{X}$ $\mathbb{R}$

$${}_{+}(\boldsymbol{x}^{\top}\hat{\boldsymbol{\theta}}\ \hat{\ }) \quad {}_{+}(\boldsymbol{x}^{\top} \quad) \quad (\ )$$

Since any $\in$ can be decomposed into $(\quad)^{\top}$ where and $\| \ \|$ due to the identifiability requirement, is a function of This makes it possible to write the criteria function and the corresponding minimizers in terms of the sub-vector only,

$$\underset{\boldsymbol{\theta}\in}{} \quad (\ ) \qquad \underset{\boldsymbol{\theta}\in}{} \quad (\ ) \qquad \underset{\boldsymbol{\theta}\in}{} \quad (\ ) \quad \text{where} \tag{3.3}$$

$$\{ \quad \in \mathbb{R} \quad {}^{\top} \ {}^{\top} \quad \} \tag{3.4}$$

We note that and are the unconstrained minimizers for the criteria functions ( ) ( ) and ( ) respectively, which are continuous functions of the latter two almost

surely. Similarly the link function $_+(\mathbf{x}^\top \quad)$ can be rewritten as $_+(\mathbf{x}^\top(\ )\ (\ ))$, where $(\ )\ (\ \sqrt{\|\ \|}\ )^\top$.

To study limit distributions, we impose an additional requirement on the interplay between the metric $(\ )$ in the metric space of responses and the true regression function $_+$, namely that the second order difference of the function $(\ _+(\ ))$ is bounded away from zero, for any $\in\mathcal{T}$ where $\mathcal{T}\ \mathbb{R}$ denotes the domain of $_+$ Specifically, for $^\top(\ )$ for some $\in\mathbb{R}$ and $\in$ we denote $_+(\ ^\top(\ )\ (\ ))\ _+(\ )$ by $_+(\ )$ We assume

(A5) For any $\in\mathcal{T}\ \mathbb{R}$ and $\in$ there exists some and such that for any sufficiently small and $\in\mathcal{T}$

$$— \quad (\ _+(\ )) \quad (\ _+(\ )) \quad (\ _+(\ ))$$

In the Euclidean case, assumption (A5) means that $_+$ can be locally approximated by straight lines and is satisfied for twice differentiable functions $_+$, a common assumption for classical single index modeling. Beyond the Euclidean special case, assumption (A5) can be shown to be satisfied for fairly general metric spaces. An example for this are CAT(0) spaces (see [9]), where the regression function between two distinct points $_+(\ )$ and $_+(\ )$ for some small can be approximated arbitrarily closely by the geodesic path connecting them. Further details on this are provided in Appendix A and B.

The geometric assumption (A5) is crucial to show that the intermediate objective function $(\ )$ has non-negative curvature near its minimizer with high probability. This is necessary to bound the rate of the convergence of the discrepancy between the intermediate index parameter and the estimated version We proceed to define partial derivatives of the criteria functions with respect to the components of For any $\mathbf{x}\in\mathbb{R}$ with bounded norm and $\in(\ )$, define the function $_\mathbf{x}\ \mathbb{R}\ \mapsto\mathbb{R}$ such that

(3.5)
$$_\mathbf{x}(\ )\ _\mathbf{x}(\ )\ _+(\mathbf{x}^\top(\ ))$$

The first and second ordered forward finite differences of $_\mathbf{x}$ are given as follows

(3.6)
$$\nabla(\mathbf{x}\ )\ _\mathbf{x}(\ )\ _\mathbf{x}(\ )$$
$$\nabla(\mathbf{x}\ )\ _\mathbf{x}(\ )\ _\mathbf{x}(\ )$$
$$_\mathbf{x}(\ )\ _\mathbf{x}(\ )$$

Define

$$\begin{cases}\boldsymbol{\Delta}\ (\ ) & \frac{\partial\ (\boldsymbol{\theta})}{\partial\ _2}\ \frac{\partial\ (\boldsymbol{\theta})}{\partial\ _p}^\top\ \frac{\partial\ (\boldsymbol{\theta})}{\partial\ _r} \to -\mathbb{E}(\nabla\ (\mathbf{X}\ ))\\ \boldsymbol{\Delta}^2\ (\ ) & \frac{\partial^2\ (\boldsymbol{\theta})}{\partial\ _r\partial\ _s}\ \frac{\partial^2\ (\boldsymbol{\theta})}{\partial\ _r\partial\ _s} \to -_2\mathbb{E}(\nabla\ (\mathbf{X}\ ))\end{cases}$$

We note that $(\ )$ $(\ )$ and $(\ )$ are all real-valued functions with domain in a constrained subset of $\mathbb{R}$ The appropriate limits for defining the partial derivatives can be shown to exist under (A2) and the assumed total boundedness of the metric space The estimated versions of the finite difference derivatives are, for ,

$$\begin{cases}\boldsymbol{\Delta}\ (\ ) & \frac{\partial\ _n(\boldsymbol{\theta})}{\partial\ _2}\ \frac{\partial\ _n(\boldsymbol{\theta})}{\partial\ _p}^\top\ \frac{\partial\ _n(\boldsymbol{\theta})}{\partial\ _r} — \widehat{\nabla}\ (\mathbf{X}\ )\\ \boldsymbol{\Delta}^2\ (\ ) & \frac{\partial^2\ _n(\boldsymbol{\theta})}{\partial\ _r\partial\ _s}\ \frac{\partial^2\ _n(\boldsymbol{\theta})}{\partial\ _r\partial\ _s}\ \frac{}{_2}\ \widehat{\nabla}\ (\mathbf{X}\ )\end{cases}$$

(3.7)

with

(3.8)

Here ( ) is a tuning parameter depending on for which we assume that

(A6) ( ) → and ( ) → ∞ as → ∞

Assumptions (A4) and (A6) together imply that furthermore / → as → ∞

Observe that the true and estimated index directions can be framed as M-estimators of their respective criteria functions. This suggests utilizing empirical process-based approaches to obtain distributional convergence of specifically to adopt a linearization approach [65]. Specifically, we show that √( ) ( ) and √( ) → where is a Gaussian random variable. Combining these results, it follows that

THEOREM 3.2. *Under assumptions (A1)-(A6), and assumptions (U1)-(U3), and (R1)-(R2) listed in the Supplement Section S.2.,*

$$\sqrt{(}\qquad)\rightarrow\qquad(\quad(\quad))$$

where and are as defined in assumption (A4), ( ) $\boldsymbol{\Delta^2}$ ( ) ( ) $\boldsymbol{\Delta^2}$ ( ) *and* ( ) ((( ))) *with*

$$(\quad)\quad\begin{cases}\rightarrow & -\nabla\ (\boldsymbol{X}\qquad)\quad \textit{if}\qquad\in\{\qquad\}\\ \rightarrow & -\nabla\ (\boldsymbol{X}\qquad)\ -\nabla\ (\boldsymbol{X}\qquad)\quad \textit{if}\qquad\in\{\qquad\}\end{cases}$$

The asymptotic normality of ( ) follows from Theorem 3.2 with a simple application of the multivariate delta method as √ ‖ ‖ implying (/)

COROLLARY 2. *Under the conditions required for Theorem 3.2,*

$$\sqrt{(}\qquad)\rightarrow\qquad(\quad)^{\top}$$

where $\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\theta}}$ $\boldsymbol{\theta}\ \boldsymbol{\theta}_0$ $^{\top}/\sqrt{\quad\|\ \|}$ $\boldsymbol{\theta}\ \boldsymbol{\theta}_0$ *is the Jacobian matrix of size* ( ).

Define the intuitive estimator ( ) for ( ) given by ( ) ((( ))) , with

$$\left\{\begin{array}{ll}— & \nabla\ (\mathbf{X}\qquad)\quad —\quad \nabla\ (\mathbf{X}\qquad)\quad \text{if}\\ — & \nabla\ (\mathbf{X}\qquad)\nabla\ (\mathbf{X}\qquad)\\ — & \nabla\ (\mathbf{X}\qquad)\quad —\quad \nabla\ (\mathbf{X}\qquad)\quad \text{if}\end{array}\right.$$

The following two propositions imply consistent estimation of the covariance matrix.

PROPOSITION 3. *Under assumptions (A1)-(A6),* √ (( )) (( )) *converges to a* ( ) *dimensional normal distribution with mean vector and a finite covariance matrix.*

Details about the limiting covariance matrix can be found in Section S of the Supplement [5]. A natural estimate for the asymptotic covariance matrix in Theorem 3.2 is
$\hat{}( ) \quad \boldsymbol{\Delta}^{2} \quad ( ) \quad \hat{}( ) \; \boldsymbol{\Delta}^{2} \quad ( )$

PROPOSITION 4. *Under assumptions (A1)-(A6), and assumptions (U1)-(U3), and (R1)-(R2) listed in the Supplement Section S.2.,*

$$\hat{}( ) \quad (\quad) \rightarrow$$

With Slutsky's theorem, combining the above propositions with Theorem 3.2,

COROLLARY 3. *Under assumptions (A1)-(A6), and assumptions (U1)-(U3), and (R1)-(R2) listed in the Supplement Section S.2.,*

$$\sqrt{\quad}(\hat{}( )) \quad / (\qquad) \rightarrow \quad (\qquad)$$

where and are as defined in assumption (A4).

Again it is straightforward to extend the above result to obtain the limit distribution for the full parameter vector $\hat{\hat{}} \quad (\hat{} \quad)$ as due to the constraints the full parameter vector is a function of the reduced one. Define the estimate for the Jacobian matrix of size $(\quad)$ as $\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} \qquad ^{\top}/\sqrt{\quad \| \quad \|}$. Then using Corollary 2 and Proposition 4 one has $\sqrt{\quad}(\quad \hat{}(\quad)^{\top}) \quad / (\hat{\hat{}} \qquad) \rightarrow \quad (\quad)$ and furthermore

COROLLARY 4. *Under the conditions required for Corollary 3, for any* $\boldsymbol{x} \in \mathcal{X} \quad \mathbb{R}$

$$_{+}(\boldsymbol{x}^{\top}\hat{\hat{}} \; \hat{\hat{}}) \quad _{+}(\boldsymbol{x}^{\top} \qquad) \qquad (\quad ^{/})$$

In applications of regression models, it is often important to test the statistical significance of added predictors. Based on the above normality results, one can obtain Wald-type statistics to test the significance of certain variables in the linear index. Since is on the surface of the unit sphere, the constraint $\| \quad \|$ removes one dimension. The actual dimension of the surface of the unit sphere is and the values of $(\quad)$ components of determine when without loss of generality, the value of the first component of is assumed to be positive. Therefore one can obtain confidence regions for by constructing confidence regions for the last $(\quad)$ components of only.

A common testing problem concerns the null hypothesis H : , for any . More general tests of a linear null hypothesis H : $^{( )}$ for a known matrix of full row rank and $^{( )} \quad (\quad)^{\top}$ are also of interest and are implied by the following result, which also provides (elliptical) asymptotic confidence regions for the components of interest and whereas before $(\;)$ is the number of bins in the binning step.

COROLLARY 5. *Under the null hypothesis* $^{( )}$ *for some* $(\quad)$ *matrix with of rank , denoting the estimated asymptotic covariance matrix for* $^{( )}$ *by* $\hat{}(\; ^{( )})$ *then under the conditions required for Corollary 3,*

$$(\quad ^{( )} \quad)^{\top}(\quad (\hat{}(\; ^{( )})/ \quad) \quad ^{\top}) \quad (\quad ^{( )} \quad) \rightarrow$$

Specifying the last ( ) *components of the true direction index as* $^{(\,)}$ ($)^{\top}$ *where* , *a* ( ) *confidence region for* $^{(\,)}$ *is*

$$\{\ \in \mathbb{R} \quad (\ ^{(\,)}\ )^{\top}(\widehat{\ }(\ ^{(\,)})/\ )\ (\ ^{(\,)}\ )\ \ \|\ \|\ \ \}$$

with ( ) . *Here* $\widehat{\ }(\ ^{(\,)})$ *is the* ( ) *dimensional sub-matrix of the asymptotic covariance matrix* $\widehat{\ }(\ )$.

Observe that for , $^{(\,)}$ Then Corollary 5 yields the confidence region for the parameter as $\{\ \in \mathbb{R}\ \ (\ \ )^{\top}(\widehat{\ }(\ )/\ )\ (\ \ )\ \ \|\ \|\ \ \}$ with ( ) . Then the confidence region for can be obtained immediately through the relationship ($)^{\top}$ with $\sqrt{\ \ \|\ \|}$

For practical implementation, direct estimation of the asymptotic covariance matrix is tedious since it involves a tuning parameter to approximate the partial derivative for multiple variables by finite difference quotients. Instead, we use a nonparametric bootstrap approach to provide a consistent estimator of the asymptotic covariance matrix [17, 59]. Consistency of the bootstrap moment estimators for a general M-estimator is a well-studied problem. [35] established uniform integrability of the bootstrap M-estimator, thereby giving sufficient conditions for the consistency of the bootstrap moment estimators. Following similar arguments as Theorem 2.2 in [35], we obtain consistency of the proposed bootstrap covariance matrix estimator.

Let ($\mathbf{X}$) ($\mathbf{X}$) denote a bootstrap sample, i.e., an independent sample from the empirical distribution of the observed sample ($\mathbf{X}$) ($\mathbf{X}$). The bootstrap M-estimator of is $\underset{\boldsymbol{\theta}\in}{\ \ -\ \ \ \ }{}_{+}((\mathbf{X}^{\intercal}))$ Here and $\mathbf{X}$ are the response and predictor values for observations falling in the th bin, A bootstrap estimator of the asymptotic covariance matrix is given by [35, 49, 8, 27]

$$\mathbb{E}\ \ (\ \ )(\ \ )^{\intercal}|(\mathbf{X}\ \ )\ \ (\mathbf{X}\ \ )$$

PROPOSITION 5. *Under assumptions (A1)-(A6),* *is consistent for the true asymptotic covariance matrix* ( )

Combining the above proposition with Theorem 3.2 using the bootstrap covariance estimator, an analog of Corollary 3 immediately follows, as $\sqrt{\ }(\ \ )\ ^{/}(\ \ )\rightarrow\ (\ \ )$ justifying the bootstrap construction of confidence regions and ensuing inference, where we approximate the bootstrap covariance matrix by Monte Carlo estimation. The observed sample ($\mathbf{X}$) ($\mathbf{X}$) is resampled with replacement times and the estimate for the index parameter computed for each bootstrap sample. Based on the $^{\text{th}}$ bootstrap sample the index parameter is estimated as The bootstrap estimate of the covariance matrix is then $-\ \ (\ \ )(\ \ )^{\intercal}$ which is also consistent for ( )

As an example, if one applies the statistic in Corollary 5 to test the null hypothesis

(3.9) where

one can examine the power of the test for alternatives indexed by a parameter ,

(3.10)

Under $^{\top}(\widehat{\ }\ )$ $_{(\ \ )}$ asymptotically. Noting that $\widehat{\ }$ is consistent for ( ) under both and , the asymptotic distribution of under is the non-central

chi-square distribution $_{(\ \ )}(\ \ )$ with $(\ \ \ \ )$ degrees of freedom and non-centrality parameter $^{\top}_{\delta}(\ (\ \delta))\ \ \delta$, where $_{\delta}\ (\ \ \ \ )$. The asymptotic power of the level test under is $(\ \ \ \ _{(\ \ )}(\ \ \ ))$, where $\ \ \ \ _{(\ \ )}(\ \ \ )$, which demonstrates that for all the asymptotic power converges to 1 with the rate .

4. Implementation and simulation studies. Implementation of the single index Fréchet regression (IFR) model in (2.6) requires the choice of two tuning parameters, the bandwidth $(\ \ )$ used for the local linear Fréchet regression as per (2.3) and the number of bins $(\ \ )$ (see assumption (A4)). In applications, the tuning parameters $(\ \ \ \ )$ can be chosen by leave-one-out cross-validation. The first step is to select the optimal bandwidth parameter by minimizing the mean discrepancy between the local linear Fréchet regression estimates and the observed object responses, i.e.,

$$-\ \ \ \ (\ \ \ \ _{(\ \ )}\ \mathbf{X}^{\top}\ \ \ \ )$$

where $_{(\ \ )}\ \mathbf{X}^{\top}\ \ \ \ $ is the local linear Fréchet regression estimate at $\mathbf{X}^{\top}$ obtained with bandwidth based on the sample excluding the th pair $(\mathbf{X}\ \ \ )$, i.e.,

$$_{(\ \ )}\ \mathbf{X}^{\top}\ \ \ \ \ \ \in\ \ \frac{\ \ \ \ \ \ }{(\ \ \ \ )}\ \ \ \widehat{\ }(\mathbf{X}^{\top}\ \ \mathbf{X}^{\top}\ \ \ \ )\ \ (\ \ \ \ )$$

In practice, we replace leave-one-out cross-validation by fold cross-validation when . Once is chosen a second leave-one-out cross-validation step is applied to select the number of non-overlapping bins where the objective function to minimize is the empirical Fréchet variance for the binned data,

$$—\ \ \ \ (\ \ \ \ ^{*}_{+(\ \ )}(\mathbf{X}^{\top}\ \ \ \ ))$$

Here $^{*}_{+(\ \ )}(\mathbf{X}^{\top}\ \ \ \ )$ is the local linear Fréchet regression estimate at $\mathbf{X}^{\top}$ obtained with bandwidth based on the sample excluding the observation at the th bin $(\mathbf{X}\ \ \ \ )$, i.e.,

$$^{*}_{+(\ \ )}(\mathbf{X}^{\top}\ \ \ \ )\ \ \ \ \ \ \in\ \ -\ \ \ \widehat{\ }(\mathbf{X}^{\top}\ \ \mathbf{X}^{\top}\ \ \ \ \ \ )\ \ (\ \ \ \ )$$

Thus, for each given unit direction we first select the optimal tuning parameters $(\ \ \ \ )$ which will generally vary with and then employ them when computing the loss function $(\ \ )$ Finally, the index parameter is estimated as $\widehat{\ }$ the unit direction minimizing $(\ \ )$ over such that $^{\top}\ \ \ \ $ This leads to an iterative scheme, where for a given unit direction the tuning parameters $(\ \ \ \ )$ are initially selected by cross-validation and then iteratively updated, in turn with updating to minimize the loss function. We numerically optimize the empirical loss $(\ \ )$ under the constraint $^{\top}\ \ \ \ $ via the following algorithm.

1. Take a grid of unit vectors such that $^{\top}\ \ \ \ $. This is achieved by generating dimensional standard Gaussian random vectors with positive first elements and standardizing them, utilizing the spherical symmetricity of -dimensional standard Gaussian vectors.
2. For each select optimal tuning parameters $(\ \ \ \ )$ (for bandwidth and number of non-overlapping bins, respectively) by cross-validation.
3. Using $(\ \ \ \ )$ compute the loss function $(\ \ )\ \ —\ \ \ \ (\ \ \ \ _{+}(\mathbf{X}^{\top}\ \ ))$

4. Find the minimizer of ( ) such that ⊤ by searching over all directions generated in step 1. In our implementation, we considered 500 directions.

The computational challenges to obtain Fréchet means vary by metric space. In many cases, the key idea to compute the weighted Fréchet means reduces to solving a constrained quasi-quadratic optimization problem and projecting back into the solution space. For random objects such as distributions, positive semi-definite matrices, networks, and Riemannian manifolds among others, obtaining the unique solution is computationally straightforward. For our simulations we considered random objects corresponding to samples of univariate distributions equipped with the Wasserstein metric and samples of multivariate data with the usual Euclidean metric.

We generated Monte Carlo runs for each setting, and for each run obtained a direction estimate ≃( ) The intrinsic Fréchet mean of these estimates on the unit sphere was then computed as ^ Given that both the ≃( ) and their target lie on the unit sphere in ℝ , bias and deviance of the estimator can be obtained as

$$\text{bias}(\widehat{\ }) \quad \langle \widehat{\ } \quad \rangle \qquad \text{dev}(\widehat{\ }) \quad \widehat{\ } \left(\quad \langle \widehat{\simeq}(\ ) \quad \widehat{\ } \rangle \right) \tag{4.1}$$

To illustrate the performance of the Wald-type statistic for testing a linear hypothesis, we again created Monte Carlo runs as described above except that components of the index were generated to follow the null hypothesis in (3.9). To obtain the power function of the test against the sequence of alternatives given in (3.10), we calculated the test statistic for simulation runs and determined the fraction of tests that rejected the null hypothesis at the nominal level

4.1. *Distributional responses.* The space of univariate distributions with the Wasserstein metric provides an ideal setting for illustrating the efficacy of the proposed methods. For any two distribution objects ∈ ( ), the Wasserstein-2 distance is given by

$$(\quad) \quad \lceil \left(\quad (\ ) \quad (\ ) \right) \tag{4.2}$$

where and are the quantile functions corresponding to and respectively. We consider distributions on a bounded domain as responses ( ) that we represent by their respective quantile functions ( )( ) and that are paired with a dimensional Euclidean predictor vector **X**. The true single index projections $\mathbf{x}^\top$ were obtained by first generating ( )ᵀ from a multivariate Multivariate Gaussian distribution with $\mathbb{E}(\ )$ and (′) . Then the components of **X** ( )ᵀ were computed as ( ) where is the standard normal distribution function. Finally, we generated a dimensional unit vector such that ‖ ‖ and , and computed the projection $\mathbf{X}^\top$ We selected and random responses were generated conditional on **X**, by adding noise to the true regression quantile function

$$(\ _+(\mathbf{x}))(\ ) \quad \mathbb{E} \quad (\ )(\ )|\mathbf{X}^\top \quad \mathbf{x}^\top \tag{4.3}$$

For generating the distributional responses, two simulation settings were examined (see Table 1). For both scenarios, three different link functions were considered for the data-generating mechanism, namely ( ) , ( ) , and ( ) ( ) In the first setting, the true response was generated as a normal distribution with parameters depending on $\mathbf{X}^\top$ For $\mathbf{X}^\top$ $\mathbf{x}^\top$, the distribution parameters (**x**) (($\mathbf{x}^\top$)) and (**x**) (/ ($\mathbf{x}^\top$)) were independently sampled, where ( ) $\frac{(\ )}{(\ )}$. The corresponding

distribution-valued regression function is given by $_+(\mathbf{x}^\top \quad) \quad \mathbb{E}(\quad(\quad)(\,)|\mathbf{X}^\top \qquad \mathbf{x}^\top \quad)$ $(\mathbf{x}^\top \quad) \qquad (\mathbf{x}^\top \quad) \qquad (\,)$ where $(\,)$ is the standard normal distribution function.

For the second setting, the distributional parameter $(\mathbf{x})$ was sampled as before, while the standard deviation parameter was fixed at The resulting distributions were then subjected to a random transport map in Wasserstein space that is uniformly sampled from the collection of transport maps $(\,) \qquad (\quad)/|\,|$ for $\in \{\qquad\}$. The observed distributions are not Gaussian anymore due to the added random transports Nevertheless, the Fréchet mean can be shown to equal $(\mathbf{x}^\top \quad) \qquad (\,)$.

In Table 1, is a push-forward measure such that $(\quad) \qquad (\{\quad(\,) \in \quad\})$, for any measurable function $\mathbb{R} \to \mathbb{R}$ distribution $\in \mathcal{W}$, and set $\mathbb{R}$ Here is a Gaussian distribution with parameters and as described above, and $\mathcal{W}$ is the metric space of distributions on a compact support equipped with the 2-Wasserstein metric.

TABLE 1
Two different simulation settings for distributional objects.

Setting I	Setting II
$Q(Y)(\cdot) = \mu + \sigma\Phi \quad(\cdot)$, where $\mu \sim N(\zeta(\mathbf{x}^\top\bar{\boldsymbol{\theta}} \;), 0.25)$, $\sigma \sim Exp \quad \frac{(\mathbf{x}^\top\boldsymbol{\theta_0})}{\mathbf{x}^\top\boldsymbol{\theta_0}}$.	$Q(Y)(\cdot) = T\#(\mu + \sigma\Phi \quad(\cdot))$, where $\mu \sim N(\zeta(\mathbf{x}^\top\bar{\boldsymbol{\theta}} \;), 0.25), \sigma = 0.1$, $T\;(a) = a - \sin(ka)/\lvert a\rvert, k \in \{\pm1, \pm2, \pm3\}$.

Following these specifications, for each Monte Carlo run we generated density objects and multivariate Euclidean predictors from the true model. The bias and deviance of the estimated direction vectors for varying sample sizes and resulting from 500 Monte Carlo runs are displayed in Table 2. The bias due to the local linear Fréchet estimation is generally low and the variance of the estimates is seen to diminish with increasing sample size.

TABLE 2
Two different simulation settings for distributional objects. Bias and deviance (within parenthesis) of $\hat{\hat{\boldsymbol{\theta}}}$ *(measured in radians as per (4.1)) obtained from* 500 *Monte Carlo runs, where the predictor dimension is* $p = 4$*, and the tuning parameters* (b, M) *were chosen by* $5-$*fold cross-validation.*

	Setting I						Setting II					
	link1 $(x \mapsto x)$		link2 $(x \mapsto x \;)$		link3 $(x \mapsto \exp(x))$		link1 $(x \mapsto x)$		link2 $(x \mapsto x \;)$		link3 $(x \mapsto \exp(x))$	
	bias	dev	bias	dev	bias	dev	bias	dev	bias	dev	bias	dev
n = 100	0.041	0.029	0.053	0.039	0.045	0.061	0.029	0.027	0.022	0.037	0.028	0.044
n = 1000	0.023	0.013	0.027	0.012	0.029	0.012	0.010	0.012	0.011	0.014	0.017	0.021

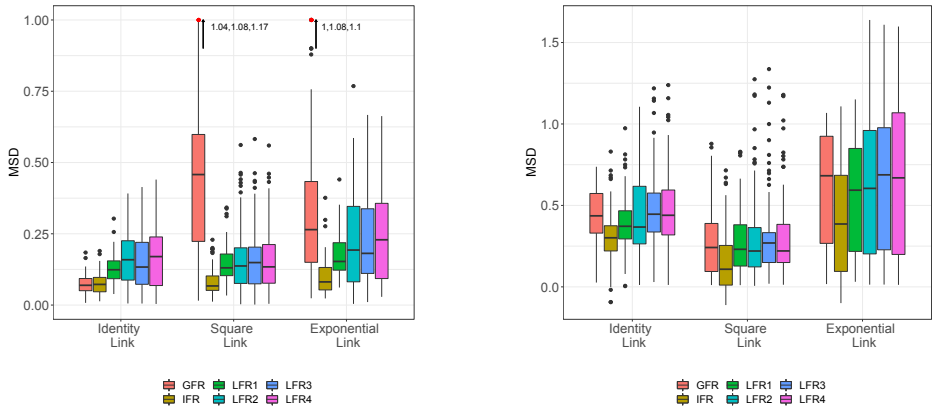
The performance of the proposed method was further evaluated by computing the mean squared deviation (MSD) between the observed and the fitted distributions. Denoting the simulated true and estimated distribution objects by $_+(\mathbf{X}^\top \quad)$ and $_+(\mathbf{X}^{\top\widehat{\;}})$, respectively, for the utility of the estimation was measured quantitatively by

$$\text{(4.4)} \qquad \text{—} \qquad (\quad{}_+(\mathbf{X}^\top \qquad) \quad {}_+(\mathbf{X}^{\top\widehat{\;}\;\widehat{\;}}))$$

where $(\quad)$ is the Wasserstein-2 distance between two distributions.

We compared the estimation performance of the proposed single index Fréchet regression (IFR) method with global Fréchet regression (GFR), which directly handles multivariate predictors as it is a generalization of global least squares regression [55]. Since local linear

Fréchet regression (LFR) is subject to the curse of dimensionality and not suitable for predictors, we fitted four separate LFR models in turn for each of the univariate component predictors and computed the Mean Squared Deviation (MSD) for each of these four fits. No binning is required for either GFR or LFR model fits. In Figure 1 we denote the MSDs for the four local linear Fréchet regression fits as LFR1, LFR2, LFR3, and LFR4, respectively. Figure 1, displaying boxplots of the MSDs over simulation runs for a sample size of The left and right panels correspond to simulation settings I and II, respectively, and in each panel, three cases are considered corresponding to the different link functions used to generate the distributional data. Overall six Fréchet regression methods are compared, for two simulation settings and three data generation mechanisms. We observe that the IFR method outperforms the baseline GFR and all four of the LFR methods in all scenarios. The smallest difference between the IFR and GFR occurs when an identity link function is used in the data generation mechanism. This is as expected since in this case the true model essentially reduces to GFR, the equivalent of a linear model. The individual LFR models have higher MSDs, which can be attributed to the fact that we are ignoring the effect of the other predictors when fitting the local model with one predictor at a time.



(a) Simulation Setting I.

(b) Simulation Setting II.

Fig 1: Boxplot of the mean squared deviation (MSD) of the fits using the single index Fréchet regression model (IFR), the Global Fréchet regression (GFR) model, and four Local Fréchet Regression (LFR) models using the univariate predictor components, for sample size Left and right panels correspond to simulation settings I and II, respectively. The left, middle, and right columns in each of the panels correspond to the three different link functions used in the data generation mechanism, namely, identity, square, and exponential link functions, respectively; in all scenarios, the link functions are estimated from the data. In the left panel, the outliers having MSD greater than are marked in red with an upward arrow and the corresponding MSD values are overlaid.

Figure 2 demonstrates the effect of the index values on the distributional objects under simulation setting I for the different link functions when responses are represented in the form od densities. The three data generation mechanisms are shown in the left, middle, and right panels of Figure 2 respectively. For each case, the IFR model was fitted at the mean and mean sigma levels of the index values, displayed in red, blue, and green lines respectively, while the observed/simulated densities are overlaid in orange in each panel. In each case, for a higher value of the index level, the fitted densities shift towards the top-right, indicating a positive association of the single-index values on the mode of the distributions.

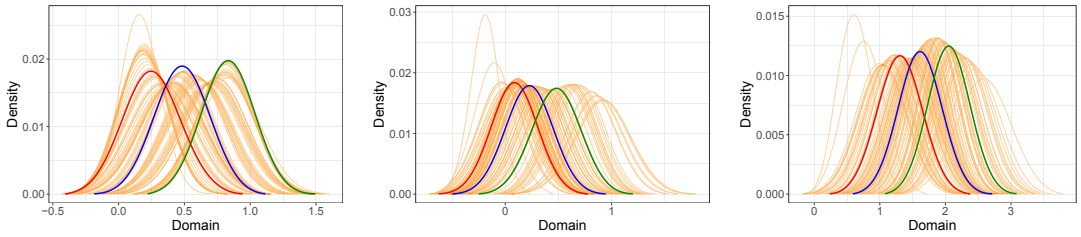


Fig 2: Simulated (orange) and fitted (red, blue, and green) distributional objects represented as densities for simulation setting I for sample size The left, middle, and right panels correspond to three link functions (identity, square, and exponential link) used in the data generation process. In each case, the IFR model fits are obtained at three different levels of the estimated index values, namely, at mean($\mathbf{x}^\top\widehat{\ }$) sd($\mathbf{x}^\top\widehat{\ }$) (red), mean($\mathbf{x}^\top\widehat{\ }$) (blue) and mean($\mathbf{x}^\top\widehat{\ }$) sd($\mathbf{x}^\top\widehat{\ }$) (green).

To illustrate the out-of-sample prediction performance of the proposed IFR model, the dataset was randomly split into a training set with sample size $_{\text{train}}$ $\lfloor$ / $\rfloor$ and a test set with the remaining $_{\text{test}}$ $\lfloor$ / $\rfloor$ subjects. The IFR method was implemented as follows: for any given unit direction $\in$ we partition the domain of the projections into equal-width non-overlapping bins and consider the representative observations $\mathbf{X}$ and for the data points belonging to the th bin. The "true" index parameter was estimated as $\widehat{\ }$ as per (2.11). We then took the fitted index obtained from the training set and predicted the responses in the test set using the covariates present in the test set. As a measure of the efficacy of the fitted model, we computed the root mean squared prediction error (RMPE) as

$$\text{RMPE} \quad \frac{}{_{\text{test}}} \quad ^{n_{\text{test}}} \quad _{\text{test}} \quad _{+}\left(\quad ^{\text{test}\mathsf{T}\widehat{\ }} \ \widehat{\ }\right) \quad / \tag{4.5}$$

where $^{\text{test}}$ and $_{+}\left(\quad ^{\text{test}\mathsf{T}\widehat{\ }} \ \widehat{\ }\right)$ denote, respectively, the $^{\text{th}}$ observed and predicted responses in the test set, evaluated at the binned observation $^{\text{test}}$ and denotes the Wasserstein-2 metric in (4.2). We repeated this process times and computed RMPE for each split for the subjects separately. The mean and sd of the RMPE over the repetitions are shown in Table 3. The IFR model is seen to fare best across the different models and scenarios.

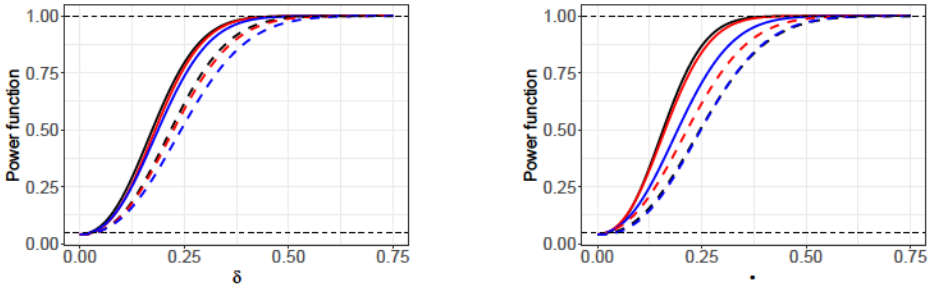
For the case of distributional objects, the linear hypothesis test of in (3.9) against the sequence of alternatives in (3.10) was also carried out. The power functions corresponding to the two simulation settings are shown in Figure 3a and 3b, respectively. As increases, the power is seen to increase rapidly. This shows that the proposed test has non-trivial power (see Figure 3). When is close to the test sizes are approximately equal to the nominal significance level of As expected, power increases with increasing sample size, most notably under the identity link. In the second simulation setting when the distributional objects are obtained by transporting a normal distribution, the power function increases at a slower rate, especially under the highly nonlinear (exponential) link function.

4.2. *Euclidean Responses.* We applied the new approach targeting general random objects as responses for the special case of Euclidean responses. It is not specifically designed for this case, where targeted, well-studied and well-honed single index models have a long history. The numerical results show that the proposed method yields results that are somewhat inferior but overall still comparable to those obtained with specially tailored traditional single index approaches; see Subsection S.4.4 of the Supplement [5].

TABLE 3

Mean and sd (in parenthesis) of the RMPE as given in (4.5) comparing the performance of various Fréchet regression models: Index Fréchet Regression (IFR), Global Fréchet Regression (GFR), Local Fréchet Regression (LFR). The LFR fits are obtained for four individual predictor components separately.

	Setting I			Setting II		
	Identity link	Square link	Exponential link	Identity link	Square link	Exponential link
IFR	0.0023 (0.0012)	0.0092 (0.0276)	0.0302 (0.0979)	0.0490 (0.0330)	0.1452 (0.0286)	0.1666 (0.0988)
GFR	0.0136 (0.0002)	0.1668 (0.0085)	0.1599 (0.0176)	0.0661 (0.0189)	0.2531 (0.0095)	0.3413 (0.0186)
LFR1	0.0478 (0.0014)	0.1671 (0.0084)	0.3516 (0.0299)	0.0679 (0.0191)	0.1317 (0.0096)	0.2371 (0.0310)
LFR2	0.0479 (0.0015)	0.1667 (0.0081)	0.3507 (0.0294)	0.0563 (0.0190)	0.1666 (0.0091)	0.1881 (0.0302)
LFR3	0.0476 (0.0020)	0.1684 (0.0133)	0.3468 (0.0296)	0.1218 (0.0191)	0.1992 (0.0142)	0.1812 (0.0304)
LFR4	0.0454 (0.0062)	0.1659 (0.0101)	0.3346 (0.0284)	0.0880 (0.0189)	0.2177 (0.0110)	0.2033 (0.0293)



(a) Simulation Setting I.

(b) Simulation Setting II.

Fig 3: Empirical power as function of δ for density object responses. The black, red, and blue curves correspond to the identity, square, and exponential link functions used in the data-generating mechanism, respectively, while the dashed and solid lines correspond to sample sizes $n = 100$ and $n = 1000$ respectively. The level of the tests is $\alpha = 0.05$ and is indicated by the dashed line parallel to the x-axis.

5. Data analysis.

5.1. *Resting state functional Magnetic Resonance Imaging: ADNI data.*

Resting-state functional Magnetic Resonance Imaging (fMRI) methodology makes it possible to study brain activation and to identify brain regions or cortical hubs that exhibit similar activity when subjects are in the resting state [2, 22]. In resting state fMRI, time series of Blood Oxygen Level Dependent (BOLD) signals are observed in regions of interest (ROI), where each ROI is represented by the signal of a seed voxel, which is the voxel in an ROI that has the highest correlation with the signals of nearby voxels. Alzheimer's Disease has been found to be associated with anomalies in the functional integration of ROIs [16, 72].

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). BOLD signals for $V = 11$ brain seed voxels for each subject were extracted for the following ROIs: MPFC (Anterior medial prefrontal cortex), PCC (Posterior cingulate cortex), dMFPC (Dorsal medial prefrontal cortex), TPJ (Temporal parietal junction), LTC (Lateral temporal cortex), TempP

(Temporal pole), vMFPC (Ventral medial prefrontal cortex), pIPL (Posterior inferior parietal lobule), Rsp (Retrosplenial cortex), PHC (Parahippocampal cortex) and HF (Hippocampal formation) [3]. The pre-processing of the BOLD signals was implemented by adopting standard procedures of slice-timing correction, head motion correction and other standard steps. The signals for each subject were recorded over the interval [ ] (in seconds), with measurements available at two-second intervals. From this the temporal correlations were computed to construct the connectivity correlation matrix, also referred to as the Pearson correlation matrix in the neuroimaging community.

The data set in our analysis consists of subjects at four stages of the disease: CN (cognitively normal), EMCI (early mild cognitive impairment), LMCI (late mild cognitive impairment), and AD (Alzheimer's) subjects. The inter-hub connectivity Pearson correlation matrix for the subject with elements

$$(5.1) \qquad (\quad) \qquad \frac{(\qquad)(\qquad)}{(\qquad)\qquad(\qquad)}^{/}$$

is the response object for each subject, where is the ( )$^{\text{th}}$ element of the signal matrix for the $^{\text{th}}$ subject and — is the mean signal strength for the $^{\text{th}}$ voxel. For Alzheimer's disease studies, the ADAS-Cog-13 score (henceforth referred to as C score) is a widely-used measure of cognitive performance. It quantifies impairments across cognitive domains that are affected by Alzheimer's disease [39]; higher scores indicate more serious cognitive deficiency.

We considered predictors, namely, stage for the disease (coded as 0-3, indicating Cognitive normal (CN), Early and Late Mild cognitive impairment (EMCI and LMCI), or Alzheimer's Disease (AD), respectively), age of the subject (in years), is the subject is female and if male), C score for the subject at the time of the first scan, and additionally all pairwise interaction terms between the above predictors, i.e., the products

In a first step, we test the null hypothesis of no regression effect, i.e., with ,

$$(\quad) \qquad \text{not all} \quad \text{are}$$

where $(\qquad)^{\top}$ and $(\qquad)^{\top}$ with $\sqrt{\quad\|\quad\|}$ The null model has included with since it is known that the stage of cognitive impairment has an effect on brain connectivity/ We obtain an estimate of the $(\quad)$ dimensional vector as the minimizer of $(\ )$ as per (3.3) and $\sqrt{\quad\|\quad\|}$ Under the null hypothesis, $^{\top}(\widehat{\ }\ )$ $_{(\quad)}$ We find that , corresponding to a value of providing evidence that there is indeed a regression relationship. We also implemented sequential predictor selection, where we specified an "alpha-to-enter"

TABLE 4
Details on step-wise model selection.

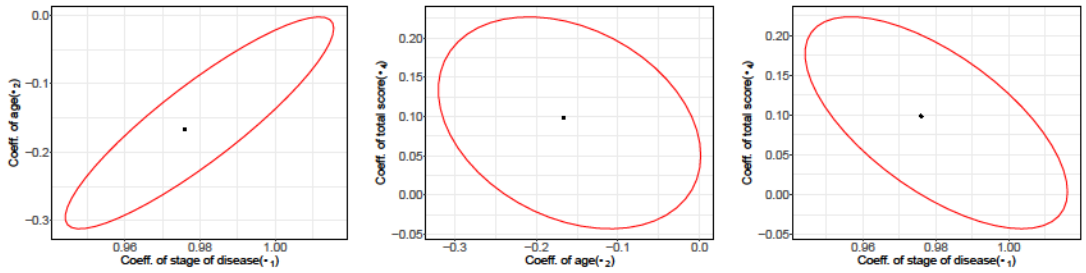
	Step 1		Step 2		Step 3	
	Coeff.	p-value	Coeff.	p-value	Coeff	p-value
Age	-0.364	0.005	-0.394	-	-0.401	-
Gender	0.198	0.122	0.558	0.161	0.173	0.113
C Score	0.371	0.094	0.207	0.010	0.279	-

level and considered to be in the model and included each of and

in the model separately along with X_1 then testing the null hypotheses $\theta_j = 0,\ j = 2,3,4$ separately. Table 4 illustrates the resulting step-wise model selection.

For example, for testing $\theta_2 = 0$, we first obtained $\hat{\theta}_2 = -0.364$, $\hat{\theta}_1 = \sqrt{1-(-.364)^2} = 0.931$) and $\tilde{T}_n = 7.88$ with a p-value of 0.005. Thus X_2 (age) was added to the model in step 1, followed by adding X_4 (C score) in step 2, while X_3 (gender) was not significant. With X_1, X_2, and X_4 in the model, we tested for the significance of the pairwise interaction terms. The null hypothesis for this test is $H_0 : \theta_5 = \theta_6 = \cdots = \theta_{10} = 0$. The p-value was 0.106, and we did not include interactions in the final model. The estimated average Fréchet error $\frac{1}{n}\sum_{i=1}^{n} d^2(Y_i, \hat{m}_{\oplus}(X_{1i}\hat{\theta}_1 + X_{2i}\hat{\theta}_2 + X_{4i}\hat{\theta}_4))$ was quite small (0.239).

To construct the confidence regions for the coefficients $(\theta_1, \theta_2, \theta_4)$, we implemented the local linear Fréchet regression with the Epanechnikov kernel and used 5-fold cross-validation to select the bandwidth b. Using the bootstrap method to obtain the estimated covariance matrix of the limiting distribution we obtained the 95% pairwise confidence ellipses for the coefficients $(\theta_1, \theta_2, \theta_4)$ of the predictors- disease stage, age, and C score, which are displayed in Figure 4. We observe that none of the pairwise confidence ellipses includes the origin and therefore the p-values are < 0.05, implying the significance of the predictors.



(a) Confidence ellipse for (θ_1, θ_2).

(b) Confidence ellipse for (θ_2, θ_4).

(c) Confidence ellipse for (θ_1, θ_4).

Fig 4: The 95% confidence ellipses for pairs of coefficients for predictors stage of the disease (X_1), age (X_2), and C score (X_4).

To illustrate the effect of the single index on the response, we computed the estimated index of the fitted model for each subject and then obtained the 25%, 50%, and 75% quantiles across all subjects, with values $q_1 = 15.048$, $q_2 = 16.430$ and $q_3 = 18.250$, respectively. The values of the four covariates for the subjects with estimated index values closest to q_1, q_2, and q_3 are in Table 5, and their observed and fitted functional connectivity correlation matrices are illustrated in Figure 5. The fitted correlation matrices correspond to the values of the estimated object link function at the three index values and are contrasted with the observed correlation matrices for the three subjects. This gives an idea of how the fitted correlation matrix changes as the index move through the three quantile levels.

We observe that the fits match the general pattern of the observed matrices quite well. The Frobenius distances between the observed and the estimated matrices at q_1, q_2, and q_3 are 1.68, 1.10, and 0.79, respectively. The fitted model reflects the trends seen in the observed correlation matrices and illustrates the nonlinear dependence of functional connectivity on the index value.

TABLE 5

Covariate values for the subjects with estimated index values closest to the first three quantiles of the estimated index when considered across all subjects, $q_1(15.048)$, $q_2(16.430)$, and $q_3(18.250)$, respectively. Subject 726 has an estimated index value that is closest to q_1, subject 695 closest to q_2, and subject 556 closest to q_3.

Subject number	Estd. index value	Stage of the disease	Age	Gender	C score
726	15.045	2	66.10 y	M	20.33
695	16.430	2	78.12 y	M	14
556	18.252	1	72.55 y	M	51.67

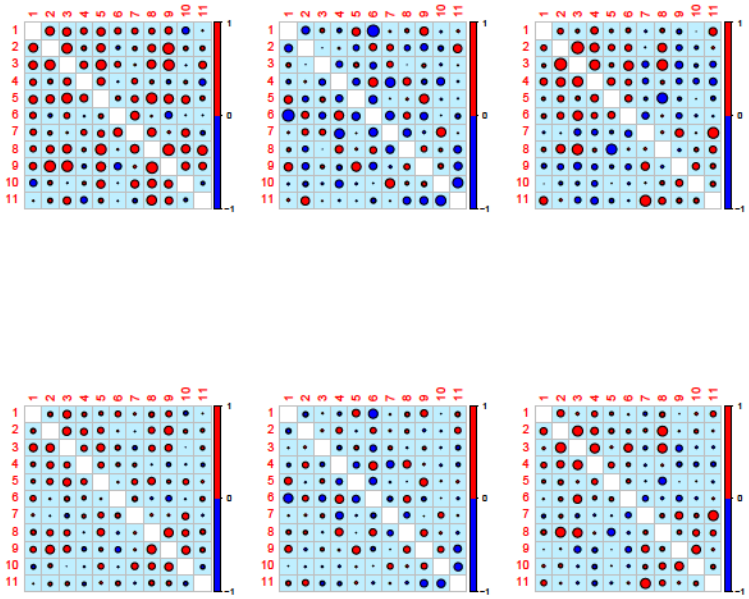


Fig 5: Observed and fitted functional connectivity correlation matrices for different values of the single index. The panels in the top row, from left to right, depict the observed functional connectivity correlation matrices for those subjects for whom the estimated index values are closest to the $25\%, 50\%$, and 75% quantile of all indices across subjects, respectively. The bottom row shows the fitted functional connectivity correlation matrices for the same subjects, (from left to right). Positive (negative) values for correlations are drawn in red (blue), where larger circles correspond to larger absolute values.

We also studied the out-of-sample prediction performance of the proposed IFR model, for which we used the root mean squared prediction error

$$\text{RMPE} = \left[\frac{1}{M_{n_{\text{test}}}} \sum_{i=1}^{M_{n_{\text{test}}}} d_F^2 \left(\tilde{Y}_l^{\text{test}}, \hat{m}_{\oplus}(\tilde{\mathbf{X}}_l^{\top} \hat{\widehat{\theta}}, \hat{\widehat{\theta}}) \right) \right]^{1/2}, \tag{5.2}$$

where $\tilde{Y}_l^{\text{test}}$ and $\hat{m}_{\oplus}(\tilde{\mathbf{X}}_l^{\top} \hat{\widehat{\theta}})$ denote, respectively, the l^{th} observed and predicted responses in the test set, evaluated at the binned observation $\tilde{\mathbf{X}}_l$. Here, n_{train} and n_{test} denote the sample

sizes of the training and testing sets formed by randomly splitting the data. We repeated this process times, and computed RMPE for each split for the subjects separately. The tuning parameters ( ) were chosen by a fold cross-validation method for each replication of the process. The prediction performance of the IFR model was compared with other applicable Fréchet regression models, namely, the global Fréchet regression (GFR) model with the three-dimensional predictor ( ) and two separate local linear Fréchet regression (LFR) models, one with the single predictor (age) and the other with the single predictor (C score). When comparing the performance of these models (Table 6), we find that

TABLE 6

Mean and sd (in parenthesis) of the root mean prediction error (RMPE) over 200 *Monte Carlo simulation runs for various object regression methods. The methods compared are index Fréchet regression (IFR); global Fréchet Regression (GFR) with the three predictors stage of the disease, age, and ADSA score; and two local linear Fréchet regression (LFR) models with separate one-dimensional predictors.*

IFR	GFR	LFR1 (Predictor Age)	LFR2 (Predictor C Score)
0.3066 (0.012)	0.5083 (0.011)	0.5076 (0.012)	0.5326 (0.013)

the out-of-sample prediction error is low for the IFR model, as compared to the global and local Fréchet regression approaches. In fact, it is not far from the in-sample prediction error ( ), calculated as the average distance between the observed training sample and the predicted objects based on the covariates in the training sets. This motivates the proposed IFR models.

5.2. *Human mortality data: Age-at-death distributions as responses.* Lifetables reflecting human mortality across 40 countries correspond to distributional responses, coupled with various country-specific covariates. We implement an overall test for the regression effect for these data. Details about this analysis are in the Supplement [5], subsection S.4.1.

5.3. *Emotional well-being of unemployed workers: Compositional data as responses.*. We further demonstrate the proposed IFR method for the analysis of mood compositional data. Here the object-valued responses lie on a manifold (sphere) with positive curvature. Thus the sufficient (but not necessary) condition for assumption (A5) that the underlying metric space behaves like a CAT(0) space is not satisfied, however, the numerical performance of the IFR method remains quite good; see Supplement [5], subsection S.4.2. This suggests a certain degree of model robustness.

6. Discussion. Binning the data to reduce the effective sample size is not necessary for the basic consistency results without rates. As discussed at the end of Section 2, the binning method is introduced in order to invoke the uniform consistency rate for the local Fréchet regression and the effective sample size ( ) is tied to this rate by virtue of assumption (A4). To avoid confusion, we discuss the binning approach throughout. The rate of convergence for $\widehat{\ }$ is / . Since our rate results and proofs rely on the uniform convergence rate of local Fréchet regression, this rate cannot be improved within the current framework and overcoming these limits would require a fundamentally different approach.

The assumptions required to obtain the technical results are essentially the same as those used before in the Fréchet regression literature, specifically in [13]. We require curvature and entropy conditions to hold uniformly across all index values and direction parameters. The curvature and entropy conditions can be verified for commonly observed objects such as univariate probability distributions, positive definite matrices, or data on the surface of a

sphere, as well as other random objects under suitable metrics. The Lipschitz condition (A2) on the link function is standard in single-index models, while assumption (A5) reflects the interplay between the properties of the metric and the link function. Assumption (A5) is implied by the easier-to-interpret assumption (K1)-(K3) (see Appendix B).

The classical single index model for Euclidean responses has been recently extended to a single index coefficient model for quantile regression [74]. This is a desirable extension for the object case of index Fréchet regression as well. One problem to resolve in this case is to define quantiles in the metric space where the object responses lie since there is no order. The problem of defining quantiles is already difficult and ambiguous for multivariate Euclidean objects. This is a potentially interesting topic for future research.

Finally, inference results for object regression are scarce. For example, the Wasserstein -tests proposed by [54] are exclusively aimed at univariate distribution quantiles within the specific setting of global Fréchet regression. We provide here a general framework to obtain inference for the case of vector predictors coupled with object responses, which includes generalized versions of inference for model comparisons and for assessing the significance of individual predictors.

APPENDIX A: GEODESICS AND CURVATURE

The length of a curve $[\ ] \to$ connecting two distinct points $\in$ can be measured by taking partitions $\{\ \}$ $[\ ]$ and finding the supremum polygonal length

$$\sum_{\in \mathcal{P}} \| \ (\ ) \ (\ ) \|$$

where $\mathcal{P}$ is any collection of subsets of $[\ ]$ with finite cardinality. The metric space $(\ )$ is a length space if $(\ ) \ | \ |$, where the infimum ranges over all curves $[\ ] \to$ connecting two distinct points and , that is, i.e., $(\ )$ and $(\ )$. A geodesic on connecting two distinct points and is the shortest path connecting the two points. Geodesics in a metric space are analogous to straight lines in a Euclidean space.

Unlike Euclidean spaces, a general metric space may not be flat, and curvature is used to measure the amount of deviation from being flat. The curvature of a given geodesic space is classified by comparing the geodesic triangles on the metric space to those on the corresponding reference spaces When , $\mathbb{R}$ with the standard Euclidean distance $(\ ) \ \| \ \|$ for any $\in \mathbb{R}$ A geodesic triangle with vertices in a geodesic space , denoted by $\triangle(\ )$, consists of three geodesic segments that connect to , to and to , respectively. A comparison triangle $\triangle(\ )$ in the reference space $\mathbb{R}$ is a geodesic triangle in $\mathbb{R}$ formed by the vertices and such that,

$$(\ ) \ \| \ \| \ (\ ) \ \| \ \| \ (\ ) \ \| \ \| \tag{A.1}$$

is said to have a non-positive curvature if there exists a comparison triangle $\triangle(\ )$ in the reference space $\mathbb{R}$ such that $(\ ) \ \| \ \|$ for all $\in \overline{\ }$ and $\in \overline{\ }$ and their comparison points and on $\triangle(\ )$. A geodesic space with curvature upper bounded by in which every geodesic triangle $\triangle(\ )$ satisfies the following $(\ )$ inequality is a CAT(0) space,

$$(\ ) \ \| \ \| \text{ for all } \in \overline{\ } \text{ and } \in \overline{\ } \text{ and their comparison points } \in \mathbb{R} \tag{A.2}$$

Every CAT(0) space is uniquely geodesic. Examples of CAT(0) spaces include Euclidean space, the space of symmetric positive definite matrices, Wasserstein-2 spaces, or phylogenetic tree spaces. For a detailed introduction to metric geometry, we refer to [9]. A compilation of the most relevant facts can be found in [45].

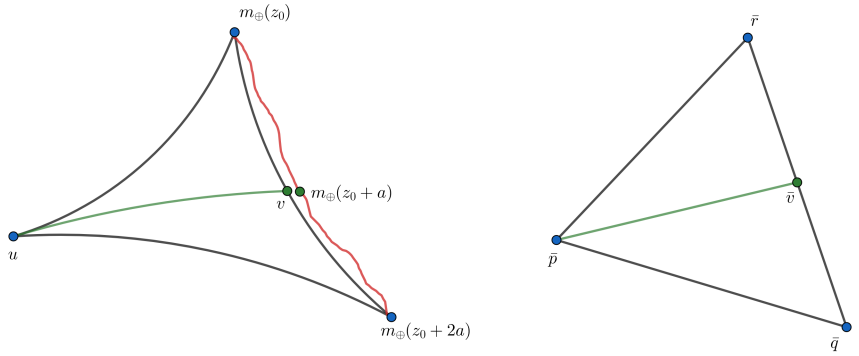


Fig 6: Left figure: Geodesic triangle formed by the three points $_+(\ )$ $_+(\ )$, where is the midpoint of the geodesic connecting the points $_+(\ )$ and $_+(\ )$ The red line depicts the true regression function $_+$ $_+(\ )$ is closely approximated by lying on a geodesic that connects $_+(\ )$ with $_+(\ )$ Right figure: Reference triangle in $\mathbb{R}$ as an illustration of the CAT(0) inequality.

APPENDIX B: SUFFICIENT CONDITIONS FOR ASSUMPTION (A5)

We discuss here sufficient conditions under which assumption (A5) holds. For this we consider the following assumptions:

(K1) $(\ )$ is a CAT(0) space, that is every geodesic triangle satisfies the CAT(0) inequality in (A.2).

For any $\in \mathbb{R}$ and $\in$ there exists some such that for small enough $\in (\ ]$ we may consider the geodesic triangle formed by $_+(\ )$ $_+(\ )$ for $\in \mathcal{T}$ for which we assume the following.

(K2) Defining the midpoint of the geodesic path connecting $_+(\ )$ and $_+(\ )$ such that

(B.1) $(\ _+(\ )\ )\ \ (\ _+(\ )\ )\ -\ (\ _+(\ )\ \ _+(\ ))$

we require

(B.2) $(\ _+(\ )\ )$

where does not depend on and is such that, and being the lower Lipschitz constant for $_+$ from assumption (A2), and the diameter of the metric space respectively.

(K3) There exist real constants such that, for all $\mathbf{x}$ with norm bounded both above and below, and for all $\in$

$$_+(\mathbf{x}^\top\ )\ \ _+(\mathbf{x}^\top\ )\ \ \|\ \ \|$$

Figure 6 illustrates the geometry of the geodesic triangles in and its reference space $\mathbb{R}$ Assumption (K2) can be verified when the link function $_+$ is smooth enough for the case of conventional Euclidean single index models. It thus provides an extension of the usual smoothness assumption in the case of random object responses. In section S of the Supplement [5]. We discuss this further in the context of Euclidean responses and in the case where the responses lie in the space of distributions equipped with Wasserstein-2 metric, and derive assumption (A5) under the sufficient conditions (K1), (K2), and (K3).

Assumption (K3) in conjunction with assumption (A2) implies that the link function $_{+}$ is bi-Lipschitz. This limits the rate at which the object $_{+}$ can change, essentially it cannot change too fast or too slowly. A bi-Lipschitz function is an injective Lipschitz function whose inverse function is also Lipschitz. The bi-Lipschitz condition is stronger than the common assumption of a monotone link function in classical single index modeling with Euclidean responses. In the special case of $\mathbb{R}$ this reduces to requiring a monotone differentiable function with strictly positive derivative almost everywhere and restricts the monotonicity to a smaller subclass of strictly monotone functions. In the special case of Euclidean responses, this simplifies to the assumption that the link function $_{+}$ is monotone and differentiable such that $'( )$ is strictly monotone with continuous derivative bounded away from zero. Such technical assumptions are commonly used for deriving distributional results in the existing single index literature, by virtue of a Taylor expansion of the link function in the Euclidean case.

Acknowledgments. We thank four referees and an associate editor for helpful comments that led to many improvements. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this paper. A complete listing of ADNI investigators and databases can be found at http://adni.loni.usc.edu.

Funding. The research is supported in part by NSF grant DMS-2014626 and a NIH ECHO grant. Funding sources for ADNI are as listed at http://adni.loni.usc.edu.

SUPPLEMENTARY MATERIAL

Supplement to Single Index Fréchet Regression
Section S.1. in the Supplement [5] includes the proofs of main and auxiliary results. Various technical assumptions are compiled and discussed in Section S.2., while Section S.3. introduces alternative sufficient conditions for assumption (A5). Section S.4. includes additional data illustrations and simulation results.

REFERENCES

[1] AFSARI, B. (2011). Riemannian L center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* **139** 655–673.
[2] ALLEN, E., DAMARAJU, E., PLIS, S., ERHARDT, E., EICHELE, T. and CALHOUN, V. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex* **24** 663–676.
[3] ANDREWS-HANNA, J. R., REIDLER, J. S., SEPULCRE, J., POULIN, R. and BUCKNER, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron* **65** 550–562.
[4] BHATTACHARJEE, S. and MÜLLER, H.-G. (2022). Concurrent object regression. *Electronic Journal of Statistics* **16** 4031–4089.
[5] BHATTACHARJEE, S. and MÜLLER, H.-G. (2022). Supplement to "Single index Fréchet regression".
[6] BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *Annals of Statistics* **31**.
[7] BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds:II. *The Annals of Statistics* **33** 1225–1259.
[8] BUCHINSKY, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *Journal of Econometrics* **68** 303–338.
[9] BURAGO, D., BURAGO, I. D., BURAGO, Y., IVANOV, S., IVANOV, S. V. and IVANOV, S. A. (2001). *A Course in Metric Geometry* **33**. American Mathematical Society.
[10] CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* **92** 477–489.
[11] CHANG, Z., XUE, L. and ZHU, L. (2010). On an asymptotically more efficient estimation of the single-index model. *Journal of Multivariate Analysis* **101** 1898–1901.

[12] CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics* **39** 1720–1747.
[13] CHEN, Y. and MÜLLER, H.-G. (2022). Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories. *The Annals of Statistics* **50** 1573–1592.
[14] COOK, R. D. (1994). Using dimension-reduction subspaces to identify important inputs in models of physical systems. In *Proceedings of the Section on Physical and Engineering Sciences* 18–25.
[15] CUI, X., HÄRDLE, W. K. and ZHU, L. (2011). The EFM approach for single-index models. *The Annals of Statistics* **39** 1658–1688.
[16] DAMOISEAUX, J. S., PRATER, K. E., MILLER, B. L. and GREICIUS, M. D. (2012). Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiology of Aging* **33** 828–e19.
[17] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and their Applications*. Cambridge University Press.
[18] DELICADO, P. and VIEU, P. (2017). Choosing the Most Relevant Level Sets for Depicting a Sample of Densities. *Computational Statistics* **32** 1083-1113.
[19] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.
[20] FAN, J. and HUANG, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057.
[21] FERRATY, F., PARK, J. and VIEU, P. (2011). Estimation of a functional single index model. In *Recent Advances in Functional Data Analysis and Related Topics* 111–116. Springer.
[22] FERREIRA, L. R. K. and BUSATTO, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neuroscience & Biobehavioral Reviews* **37** 384-400.
[23] FRÉCHET, M. R. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré* **10** 215-310.
[24] FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association* **76** 817–823.
[25] GAO, J. and LIANG, H. (1997). Statistical inference in single-index and partially nonlinear models. *Annals of the Institute of Statistical Mathematics* **49** 493–517.
[26] GHOSAL, A., MEIRING, W. and PETERSEN, A. (2021). Fréchet single index models for object response regression. *arXiv preprint arXiv:2108.06058*.
[27] GONÇALVES, S. and WHITE, H. (2005). Bootstrap standard error estimates for linear regression. *Journal of the American Statistical Association* **100** 970–979.
[28] HALL, P. (1989). On projection pursuit regression. *The Annals of Statistics* **17** 573–588.
[29] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics* **28** 157–178.
[30] HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* **84** 986–995.
[31] HRISTACHE, M., JUDITSKY, A. and SPOKOINY, V. (2001). Direct estimation of the index coefficient in a single-index model. *The Annals of Statistics* **29** 595–623.
[32] HUH, J. and PARK, B. (2002). Likelihood-based local polynomial fitting for single-index models. *Journal of Multivariate Analysis* **80** 302–321.
[33] ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58** 71–120.
[34] JIANG, C.-R. and WANG, J.-L. (2011). Functional single index models for longitudinal data. *The Annals of Statistics* **39** 362–388.
[35] KATO, K. (2011). A note on moment convergence of bootstrap M-estimators. *Statistics & Decisions* **28** 51–61.
[36] KERETA, Ž., KLOCK, T. and NAUMOVA, V. (2020). Nonlinear generalization of the monotone single index model. *Information and Inference: A Journal of the IMA*.
[37] KLOECKNER, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze* **9** 297–323.
[38] KUCHIBHOTLA, A. K. and PATRA, R. K. (2020). Efficient estimation in single index models through smoothing splines. *Bernoulli* **26** 1587 – 1618. https://doi.org/10.3150/19-BEJ1183
[39] KUEPER, J. K., SPEECHLEY, M. and MONTERO-ODASSO, M. (2018). The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. a narrative review. *Journal of Alzheimer's Disease* **63** 423–444.
[40] LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association* **102** 997–1008.
[41] LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* **86** 316–327.

[42] LI, K.-C. and DUAN, N. (1989). Regression analysis under link violation. *The Annals of Statistics* **17** 1009–1052.
[43] LIANG, H., LIU, X., LI, R. and TSAI, C.-L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics* **38** 3811.
[44] LIN, W. and KULASEKERA, K. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501.
[45] LIN, Z. and MÜLLER, H.-G. (2019). Total Variation Regularized Fréchet Regression for Metric-Space Valued Data. *arXiv preprint arXiv:1904.09647*.
[46] LOPEZ, O., PATILEA, V. and VAN KEILEGOM, I. (2013). Single index regression models in the presence of censoring depending on the covariates. *Bernoulli* **19** 721–747.
[47] MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal* **56** 732-753. https://doi.org/10.1002/bimj.201300072
[48] MÜLLER, H.-G. (2016). Peter Hall, functional data analysis and random objects. *The Annals of Statistics* **44** 1867–1887. https://doi.org/10.1214/16-AOS1492
[49] NISHIYAMA, Y. (2010). Moment convergence of M-estimators. *Statistica Neerlandica* **64** 505–507.
[50] NOVO, S., ANEIROS, G. and VIEU, P. (2019). Automatic and location-adaptive estimation in functional single-index regression. *Journal of Nonparametric Statistics* **31** 364–392.
[51] PATRANGENARU, V. and ELLINGSON, L. (2015). *Nonparametric Statistics on Manifolds and their Applications to Object Data Analysis*. CRC Press.
[52] PENNEC, X. (2018). Barycentric subspace analysis on manifolds. *Annals of Statistics* **46** 2711–2746.
[53] PETERSEN, A., DEONI, S. and MÜLLER, H.-G. (2019). Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *The Annals of Applied Statistics* **13** 393–419.
[54] PETERSEN, A., LIU, X. and DIVANI, A. A. (2021). Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves. *The Annals of Statistics* **49** 590–611.
[55] PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics* **47** 691–719. https://doi.org/10.1214/17-AOS1624
[56] PEYRÉ, G. (2009). Manifold models for signals and images. *Computer Vision and Image Understanding* **113** 249–260.
[57] SCHÖTZ, C. (2019). Convergence rates for the generalized Frćhet mean via the quadruple inequality. *Electronic Journal of Statistics* **13** 4280–4345.
[58] SCHÖTZ, C. (2020). Regression in Nonstandard Spaces with Fréchet and Geodesic Approaches. *arXiv preprint arXiv:2012.13332*.
[59] SHAO, J. and TU, D. (2012). *The Jackknife and Bootstrap*. Springer Science & Business Media.
[60] SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics* **6** 177 – 184. https://doi.org/10.1214/aos/1176344076
[61] SMALL, C. G. (2012). *The Statistical Theory of Shape*. Springer Science & Business Media.
[62] STOKER, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica: Journal of the Econometric Society* **54** 1461-1481.
[63] THOMAS YEO, B., KRIENEN, F. M., SEPULCRE, J., SABUNCU, M. R., LASHKARI, D., HOLLINSHEAD, M., ROFFMAN, J. L., SMOLLER, J. W., ZÖLLEI, L. and POLIMENI, J. R. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106** 1125–1165.
[64] TSOCHANTARIDIS, I., HOFMANN, T., JOACHIMS, T. and ALTUN, Y. (2004). Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the Twenty-first International Conference on Machine Learning* 104.
[65] VAN DER VAART, A. and WELLNER, J. (2000). *Weak Convergence and Empirical Processes: with Applications to Statistics (Springer Series in Statistics)*. Springer.
[66] WANG, J.-L., XUE, L., ZHU, L. and CHONG, Y. S. (2010). Estimation for a partial-linear single-index model. *The Annals of Statistics* **38** 246–274.
[67] XIA, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *The Annals of Statistics* **35** 2654–2690.
[68] XIA, Y. and HÄRDLE, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* **97** 1162–1184.
[69] XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2009). An adaptive estimation of dimension reduction space. In *Exploration Of A Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics* 299–346. World Scientific.
[70] YING, C. and YU, Z. (2020). Fréchet Sufficient Dimension Reduction for Random Objects. *arXiv preprint arXiv:2007.00292*.

[71] YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association* **97** 1042–1054.

[72] ZHANG, H.-Y., WANG, S.-J., LIU, B., MA, Z.-L., YANG, M., ZHANG, Z.-J. and TENG, G.-J. (2010). Resting brain connectivity: changes during the progress of Alzheimer disease. *Radiology* **256** 598–606.

[73] ZHANG, Q., XUE, L. and LI, B. (2021). Dimension reduction and data visualization for Fréchet regression. *arXiv preprint arXiv:2110.00467*.

[74] ZHAO, W., LIAN, H. and LIANG, H. (2017). Quantile regression for the single-index coefficient model. *Bernoulli* **23** 1997 – 2027. https://doi.org/10.3150/16-BEJ802

[75] ZHAO, W., ZHANG, F., LI, R. and LIAN, H. (2020). Principal single-index varying-coefficient models for dimension reduction in quantile regression. *Journal of Statistical Computation and Simulation* **90** 800–818.

[76] ZHOU, J. and HE, X. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *The Annals of Statistics* **36** 1649–1668.

[77] ZHU, L. and XUE, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 549–570.

[78] ZHU, L.-P. and ZHU, L.-X. (2009). On distribution-weighted partial least squares with diverging number of highly correlated predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71** 525–548.