Statistica Sinica **35** (2025), 713-736
doi:https://doi.org/10.5705/ss.202021.0389

PARTIALLY-GLOBAL FRÉCHET REGRESSION

Danielle C. Tucker and Yichao Wu*

University of Illinois at Chicago

Abstract: We propose a partially-global Fréchet regression model by extending the profiling technique for the partially linear regression model (Severini and Wong, 1992). This extension allows for the response to come from a generic metric space and can incorporate a combination of Euclidean predictors and a predictor which comes from another generic metric space. By melding together the local and global Fréchet regression models proposed by Petersen and Müller (2019), we gain a model that is more flexible than global Fréchet regression and more accurate than local Fréchet regression when the data generating process relies on a non-Euclidean predictor or is truly "global (linear)" for some scalar predictors. In this paper, we provide theoretical support for partially-global Fréchet regression and demonstrate its competitive finite-sample performance when applied to both simulated data and to real data which is too complex for traditional statistical methods.

Key words and phrases: Fréchet regression, local polynomial smoothing, non-Euclidean predictor, non-Euclidean response, partially linear model.

1. Introduction

Regression has continued to be a central technique for analyzing data, especially to study how a response variable depends on one or more predictor variables. In classical regression, responses and covariates are limited to be scalars. However, more complex types of data, which are situated in a generic metric space, are becoming more common and readily available in this era of big data. There is now a necessity to develop statistical models which can extend classical regression to incorporate these complex data types, such as probability distributions, symmetric positive definite matrices or data on a Riemannian manifold, as is discussed in Wang and Marron (2007), Hein (2009), Marron and Alonso (2014), and Faraway (2014).

To this aim, recent work has been done to model the dependency of Riemannian manifolds on Euclidean predictors. This includes the development of local polynomial-type models (Pelletier, 2006; Davis et al., 2010; Hinkle et al., 2012; Yuan et al., 2012), geodesic regression (Fletcher, 2013; Cornea et al., 2017; Niethammer, Huang and Vialard, 2011; Ding et al., 2019), and even partially linear models (Gonzalez-Manteiga, Henry and Rodriguez, 2012). An additive functional regression model has also been developed for the case when densities

*Corresponding author. E-mail: yichaowu@uic.edu

are responses (Han, Müller and Park, 2020).

However, each of these methods is restricted to a specific type of response, just as classical regression is restricted to Euclidean data. Therefore, Petersen and Müller (2019) extended classical regression to Fréchet regression, which can handle responses coming from a generic metric space. Fréchet regression was cleanly formulated as a weighted Fréchet mean, with weights depending on Euclidean predictors. Further, Lin and Müller (2021) studied total variation regularized Fréchet regression.

Petersen and Müller (2019) proposed both a global Fréchet regression by extending classical linear regression as well as a local Fréchet regression by extending local polynomial non-parametric smoothing methods. In this paper, we propose partially-global Fréchet regression, a method which combines the two, extending already developed semi-parametric methods. This development allows for more flexible Fréchet regression models, ones which combine the strengths of global and local Fréchet regression. Further, it extends the capability of Fréchet regression to incorporate predictors which come from a generic metric space, not necessarily Euclidean space. This generic metric space need not match that of the response, enabling our models to capture quite complex data generating processes.

The remainder of this paper is organized as follows: Section 2 gives the basic set up for Fréchet regression. A quick review of semi-parametric regression and the profiling technique for estimation (Severini and Wong, 1992) is given in Section 3. Section 4 proposes our partially-global Fréchet regression model as well as a method for its estimation. Sections 5 and 6 develop the theoretical properties of our models, including pointwise convergence rates. Finally, Sections 7 and 8 share simulation studies and real data examples to justify the utilization of our method, and Section 9 gives concluding remarks as well as ideas for future work. All definitions, assumptions, and proofs are collected in the separate Supplementary Material document.

2. Set Up

First, we present the general set up for Fréchet regression. We consider a random process $(\mathbf{X}, Z, Y) \sim F$ on the product space $\mathcal{X} \times \mathcal{Z} \times \mathcal{Y}$, where $(\mathcal{Z}, \delta)$ and $(\mathcal{Y}, d)$ are two metric spaces, while $\mathcal{X} \subset \mathcal{R}^p$. Here, $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ denotes multivariate "global (linear)" predictors, $Z \in \mathcal{Z}$ denotes a univariate "local (nonlinear)" predictor, and $Y \in \mathcal{Y}$ denotes the response. We refer to Z and Y as random objects. We use F to denote the joint distribution of $(\mathbf{X}, Z, Y)$ on $\mathcal{X} \times \mathcal{Z} \times \mathcal{Y}$. Denote the marginal distributions of $\mathbf{X}$, Z, and Y by $F_{\mathbf{X}}$, F_Z, and F_Y, respectively. We assume that all conditional distributions exist and are well defined. This is similar to the set up in Petersen and Müller (2019).

In particular, we consider a univariate local predictor Z that takes value in a generic metric space to broaden our partially-global Fréchet regression model's applicability. This includes the scalar local predictor as a special case when $\mathcal{Z} \subset \mathcal{R}$.

Because we are considering random objects Y and Z in generic metric spaces, the conventional definitions of mean and variance for random variables from Euclidean space do not apply. Fréchet (1948) generalized the concepts of mean and variance by defining the Fréchet mean and Fréchet variance of a random object Y as

$$\omega_\oplus = \operatorname*{argmin}_{y \in \mathcal{Y}} E\{d^2(Y, y)\} \text{ and } V_\oplus = E\{d^2(Y, \omega_\oplus)\},$$

respectively.

To study and model the relationship between a random object response and multivariate random variable predictors, Petersen and Müller (2019) defined the Fréchet regression function of Y given $\mathbf{X} = \mathbf{x}$ with $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$ as

$$m_\oplus(\mathbf{x}) = \operatorname*{argmin}_{y \in \mathcal{Y}} E\{d^2(Y, y) | \mathbf{X} = \mathbf{x}\}.$$

Petersen and Müller (2019) developed both global and local Fréchet regression models. We combine the strength of these models and propose a partially-global Fréchet regression model to increase model flexibility.

3. Partially Linear Model

Our partially-global Fréchet regression model is an extension of the classical partially linear model (Härdle, Liang and Gao, 2000). Let's first review this. Assume that

$$Y = \mathbf{X}^T \boldsymbol{\beta} + f(Z) + \epsilon,$$

with $\mathbf{X} \in \mathcal{R}^p$, $Z \in \mathcal{R}$, and random error ϵ satisfying $E(\epsilon|\mathbf{X}) = 0$ and $E(\epsilon|Z) = 0$, with a slight abuse of notations. Based on a random sample $\{(\mathbf{X}_i, Z_i, Y_i) : i = 1, 2, \ldots, n\}$, we denote $\mathbb{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)^T$, $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$.

The profiling technique (Severini and Wong, 1992) is commonly used for the estimation of the above semi-parametric regression model. This is the method we will focus on. Suppose for now that $\boldsymbol{\beta}$ is known and we want to estimate the nonparametric component, $f(\cdot)$. We define "adjusted" responses as $Y_i - \mathbf{X}_i^T \boldsymbol{\beta}$, $i = 1, 2, \ldots, n$. Then we can apply local polynomial smoothing (Fan and Gijbels, 1996) to the "adjusted" data $\{(Z_i, Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) : i = 1, 2, \ldots, n\}$ to estimate $f(\cdot)$.

With a local qth order polynomial, we estimate $f(z)$ for any z by solving

$$(\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_q)^T = \underset{c_0,c_1,\ldots,c_q}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ (Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) - \sum_{j=0}^{q} c_j (Z_i - z)^j \right\}^2 K_h(Z_i - z) \quad (3.1)$$

and set $\widehat{f}(z) = \hat{c}_0$, where $K(\cdot)$ is a kernel function, $K_h(\cdot) = K(\cdot/h)/h$ and $h > 0$ is a smoothing bandwidth. Now, local polynomial smoothing is a linear smoother. That is, $\widehat{f}(z) = \mathbf{s}(z)^T(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})$ for a vector

$$\mathbf{s}(z) \equiv \mathbf{s}_{h,q}(z), \quad (3.2)$$

which depends on the choice of h and q in (3.1). Denote the smoothing matrix $\mathbf{S} \equiv \mathbf{S}_{h,q} = (\mathbf{s}_{h,q}(Z_1), \mathbf{s}_{h,q}(Z_2), \ldots, \mathbf{s}_{h,q}(Z_n))^T$. Then

$$(\hat{f}(Z_1), \hat{f}(Z_2), \ldots, \hat{f}(Z_n))^T = \mathbf{S}(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}).$$

Recall that the linear regression coefficients $\boldsymbol{\beta}$ were assumed to be known. $\boldsymbol{\beta}$ must still be estimated by solving $\min_{\boldsymbol{\beta}} \langle (\mathbf{I} - \mathbf{S})(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}), (\mathbf{I} - \mathbf{S})(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) \rangle$, with the closed form solution $\widehat{\boldsymbol{\beta}} = \{\mathbb{X}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbb{X}\}^{-1}\mathbb{X}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{Y}$. Thus, the estimate of $f(z)$ at any z is given by $\widehat{f}(z) = \mathbf{s}(z)^T(\mathbf{y} - \mathbb{X}\widehat{\boldsymbol{\beta}})$.

Putting it all together, the prediction at any future $\mathbf{x}$ and z is given by

$$\begin{aligned} &\mathbf{x}^T\widehat{\boldsymbol{\beta}} + \widehat{f}(z) \quad (3.3)\\ &= \mathbf{s}(z)^T\mathbf{Y} + \{\mathbf{x}^T - \mathbf{s}(z)^T\mathbb{X}\}\widehat{\boldsymbol{\beta}} \\ &= \left[\mathbf{s}(z)^T + \{\mathbf{x}^T - \mathbf{s}(z)^T\mathbb{X}\}\left\{\mathbb{X}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbb{X}\right\}^{-1}\mathbb{X}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\right]\mathbf{Y}. \end{aligned}$$

Just as Petersen and Müller (2019) reframed the classical linear regression model and local polynomial models as an optimization problem for a prediction at a future observation, here we can reframe the classical partially linear model as an optimization problem with (3.3) as the optimizer. Denote the weight $w_i(\mathbf{x}, z)$ to be the ith element of

$$\mathbf{s}(z)^T + \{\mathbf{x}^T - \mathbf{s}(z)^T\mathbb{X}\}\left\{\mathbb{X}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbb{X}\right\}^{-1}\mathbb{X}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S}).$$

Then we can write (3.3) as

$$\underset{y \in \mathcal{Y}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i(\mathbf{x}, z)(Y_i - y)^2. \quad (3.4)$$

This allows us to extend partially linear models from Euclidean space to a generic metric space $(\mathcal{Y}, d)$ by replacing the Euclidean distance with a general metric d. The details of this extension and our proposed model are given in Section 4. But first, we point out a further extension that can be made using the local constant smoother $\mathbf{s}(z) = \mathbf{s}_{h,0}(z)$.

3.1. Local constant smoothing to encompass a non-Euclidean Z

For the special case of $q = 0$, i.e. local constant smoothing, (3.2) has the closed form

$$\mathbf{s}_{h,0}(z) = \frac{1}{\sum_{i=1}^{n} K_h(Z_i - z)} \begin{pmatrix} K_h(Z_1 - z) \\ K_h(Z_2 - z) \\ \vdots \\ K_h(Z_n - z) \end{pmatrix}.$$

To facilitate non-Euclidean Z, we can generalize the local constant smoothing vector to

$$\frac{1}{\sum_{i=1}^{n} K_h(\delta(Z_i, z))} \begin{pmatrix} K_h(\delta(Z_1, z)) \\ K_h(\delta(Z_2, z)) \\ \vdots \\ K_h(\delta(Z_n, z)) \end{pmatrix}$$

by replacing the univariate Euclidean distance $|Z_i - z|$ with a metric distance $\delta(Z_i, z)$. This will allow us to develop a model that can incorporate both a non-Euclidean response and a non-Euclidean predictor.

4. Partially-Global Fréchet Regression

We now extend the partially linear semi-parametric model combined with the profiling technique for estimation, to the case when $Y \in \mathcal{Y}$ is a random object.

4.1. Partially-global Fréchet regression model

Let us first define the partially-global Fréchet regression model.

Definition 1. We denote the Fréchet regression function of Y given $\mathbf{X} = \mathbf{x}$ and $Z = z$

$$m_{\oplus}(\mathbf{x}, z) = \operatorname*{argmin}_{y \in \mathcal{Y}} E(d^2(Y, y)|\mathbf{X} = \mathbf{x}, Z = z). \tag{4.1}$$

The partially-global Fréchet regression model is said to hold if

$$\begin{aligned} m_{\oplus}(\mathbf{x}, z) &= s_{\oplus}(\mathbf{x}, z) \text{ for any } \mathbf{x} \in \mathcal{X}, z \in \mathcal{Z}, \text{ where} \\ s_{\oplus}(\mathbf{x}, z) &= \operatorname*{argmin}_{y \in \mathcal{Y}} S_{\oplus}(y; \mathbf{x}, z) \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} S_{\oplus}(y; \mathbf{x}, z) = E\{d^2(Y, y)|Z = z\} + \{\mathbf{x} - E(\mathbf{X}|Z = z)\}^T[\mathrm{cov}\{\mathbf{X} - E(\mathbf{X}|Z)\}]^{-1} \\ \mathrm{cov}[\mathbf{X} - E(\mathbf{X}|Z), d^2(Y, y) - E\{d^2(Y, y)|Z\}]. \end{aligned} \tag{4.3}$$

Notice that each of the terms in (4.2) corresponds to each of the terms in the estimator (3.4) when the metric space $(\mathcal{Y}, d)$ is the univariate Euclidean space with $\mathcal{Y} = \mathcal{R}$ and $d(y, y') = |y - y'|$. In fact, we have the following lemma to verify that the partially-global Fréchet regression model does indeed simplify the partially linear model.

Lemma 1. *For the special case of partially linear regression model, we have* $s_\oplus(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + f(z)$ *for any* $\mathbf{x}$ *and* z.

4.2. Estimation of the partially-global Fréchet regression model

Now that we have a model, we extend the profiling technique to estimate it. Let $s_{hi}(\cdot)$ denote the ith element of the smoothing vector $\mathbf{s}_{h,q}(\cdot)$, as defined in (3.2) coming from (3.1). Then we estimate (4.2) by

$$\hat{s}_\oplus(\mathbf{x}, z) = \underset{y \in \mathcal{Y}}{\operatorname{argmin}}\, \widehat{S}_n(y; \mathbf{x}, z), \tag{4.4}$$

where

$$\widehat{S}_n(y; \mathbf{x}, z) = \hat{w}_0(y; z) + \hat{w}_1(\mathbf{x}, z)\hat{w}_2^{-1}\hat{w}_3(y), \tag{4.5}$$

with

$$\hat{w}_0(y, z) = n^{-1}\sum_{i=1}^{n} s_{hi}(z) d^2(Y_i, y),$$

$$\hat{w}_1(\mathbf{x}, z) = \left\{\mathbf{x}^T - n^{-1}\sum_{i=1}^{n} s_{hi}(z)\mathbf{X}_i^T\right\},$$

$$\hat{w}_2 = n^{-1}\sum_{i=1}^{n}\left\{\mathbf{X}_i - n^{-1}\sum_{j=1}^{n} s_{hj}(Z_i)\mathbf{X}_j\right\}\left\{\mathbf{X}_i - n^{-1}\sum_{j=1}^{n} s_{hj}(Z_i)\mathbf{X}_j\right\}^T,$$

and

$$\hat{w}_3(y) = n^{-1}\sum_{i=1}^{n}\left\{\mathbf{X}_i - n^{-1}\sum_{j=1}^{n} s_{hj}(Z_i)\mathbf{X}_j\right\}\left\{d^2(Y_i, y) - n^{-1}\sum_{j=1}^{n} s_{hj}(Z_i) d^2(Y_j, y)\right\},$$

which is essentially an extension of the optimizer (3.4) introduced previously.

Now that we have our estimator, we derive its convergence rate to the target model under several different settings.

5. Theoretical Properties for Partially-Global Fréchet Regression with Local Linear Smoothing

We first consider the case when the partially-global model utilizes local linear smoothing. We assume that $\mathcal{X} \subset \mathcal{R}^p$, $\mathcal{Z} \subset \mathcal{R}$ and consider points $\mathbf{x} \in \mathcal{X}$ for which $f_{\mathbf{X}}(\mathbf{x}) > 0$ and points $z \in \mathcal{Z}$ for which $f_Z(z) > 0$. For ease of reading, we re-write (4.3) as

$$S_\oplus(y; \mathbf{x}, z) = w_0(y; z) + w_1(\mathbf{x}, z)w_2^{-1}w_3(y), \tag{5.1}$$

where $w_0(y;z) = E\{d^2(Y,y)|Z = z\}, w_1(\mathbf{x},z) = \{\mathbf{x} - E(\mathbf{X}|Z = z)\}, w_2 = \mathrm{Cov}\{\mathbf{X} - E(\mathbf{X}|Z)\}$, and $w_3(y) = \mathrm{Cov}[\mathbf{X} - E(\mathbf{X}|Z), d^2(Y,y) - E\{d^2(Y,y)|Z\}]$.

We then define the population version of (4.4) as

$$\tilde{s}_\oplus(\mathbf{x},z) = \operatorname*{argmin}_{y\in\mathcal{Y}} \widetilde{S}_n(y;\mathbf{x},z), \tag{5.2}$$

where

$$\widetilde{S}_n(y;\mathbf{x},z) = \tilde{w}_0(y;z) + \tilde{w}_1(\mathbf{x},z)w_2^{-1}w_3(y) \tag{5.3}$$

with $\tilde{w}_0(y,z) = E\{\zeta_h(Z,z)d^2(Y,y)\}$ and $\tilde{w}_1(\mathbf{x},z) = \mathbf{x} - E\{\zeta_h(Z,z)\mathbf{X}\}$. Because we are considering local linear smoothing, we have that

$$\zeta_h(Z,z) = K_h(Z-z)\frac{\tilde{\mu}_2(z) - \tilde{\mu}_1(z)(Z-z)}{\tilde{\sigma}_0^2(z)}, \tag{5.4}$$

where $\tilde{\mu}_j(z) = E\{K_h(Z-z)(Z-z)^j\}$ for $j = 0,1,2$, and $\tilde{\sigma}_0^2(z) = \tilde{\mu}_0(z)\tilde{\mu}_2(z) - \tilde{\mu}_1^2(z)$.

Further, under local linear smoothing the i-th element of (3.2) can be written as

$$s_{hi}(z) = K_h(Z_i - z)\frac{\hat{\mu}_2(z) - \hat{\mu}_1(z)(Z_i - z)}{\hat{\sigma}_0^2(z)}, \tag{5.5}$$

where $\hat{\mu}_j(z) = \sum_{i=1}^n K_h(Z_i - z)(Z_i - z)^j/n$ for $j = 0,1,2$, and $\hat{\sigma}_0^2(z) = \hat{\mu}_0(z)\hat{\mu}_2(z) - \hat{\mu}_1^2(z)$.

Now that our notations are set, the goal is to obtain the rate of convergence for the quantity $d(s_\oplus(z), \widehat{s}_\oplus(z))$. In order to do this, we need to quantify the convergence rate of the bias term $d(s_\oplus(z), \widetilde{s}_\oplus(z))$ and the stochastic term $d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z))$. The assumptions we require are analogous to those made in Petersen and Müller (2019) and are explicitly stated in the Supplementary Material.

Theorem 1. *If Assumptions* P1, K1, L1, L2, *and* L3 *hold, then*

$$d(s_\oplus(\mathbf{x},z), \widetilde{s}_\oplus(\mathbf{x},z)) = O(h^{2/(\beta_1-1)})$$

as $h \to 0$, *where* β_1 *comes from Assumption* L3.

Theorem 2. *If Assumptions* P1, K1, L1, *and* L4 *hold, and if* $h \to 0$ *and* $nh^2 \to \infty$, *then*

$$d(\widetilde{s}_\oplus(\mathbf{x},z), \widehat{s}_\oplus(\mathbf{x},z)) = O_p\{(nh)^{-1/2(\beta_2-1)}\},$$

where β_2 *comes from Assumption* L4.

Finally, utilizing the triangle inequality of metric spaces, we have the following result:

Corollary 1. *Under the assumptions of Theorem* 1 *and Theorem* 2*,*

$$d(s_\oplus(\mathbf{x}, z), \widehat{s}_\oplus(\mathbf{x}, z)) = O(h^{2/(\beta_1-1)}) + O_p\{(nh)^{-1/2(\beta_2-1)}\}.$$

These convergence rates match that of the local Fréchet regression model in Petersen and Müller (2019). This is similar to the well-known bias-variance tradeoff in classical nonparametric smoothing. Here β_1 controls the smoothing bias rate while β_2 relates to estimation variance rate. Also, by assuming $nh^2 \to \infty$, we can utilize the results of Theorem 1 from Speckman (1988); therefore, we derive a result which matches the special case when $\mathcal{Y} \subset \mathcal{R}$, d is the Euclidean distance, and the profiling technique is applied to estimate a partially linear model under local linear polynomial smoothing. For details, please refer to the proof of Theorem 2 in the Supplementary Material document, and note that $\beta_1 = \beta_2 = 2$ in this case.

6. Theoretical Properties for Partially-Global Fréchet Regression with Local Constant Smoothing

Next, we consider the case of partially-global Fréchet regression with local constant smoothing on Z which comes from a generic metric space $(\mathcal{Z}, \delta)$. We replace (5.4) with

$$\zeta_h(Z, z) = \frac{K_h(\delta(Z, z))}{E\{K_h(\delta(Z, z))\}}, \tag{6.1}$$

and we replace (5.5) with

$$s_{hi}(z) = \frac{K_h(\delta(Z_i, z))}{n^{-1}\sum_{j=1}^{n} K_h(\delta(Z_j, z))}. \tag{6.2}$$

Because the rates of convergence which we derive will rely on the metric spaces of interest, $(\mathcal{Z}, \delta)$ and $(\mathcal{Y}, d)$, we require the following definition as well.

Definition 2. As $n \to \infty$ and $h \to 0$, the small ball probability of random objects $Z \in \mathcal{Z}$ and $Y \in \mathcal{Y}$ are defined as

$$\varphi_{\mathcal{Z},z}(h) = P\{Z \in B_{\mathcal{Z}}(z, h)\} \text{ and } \varphi_{\mathcal{Y},y}(h) = P\{Y \in B_{\mathcal{Y}}(y, h)\}, \tag{6.3}$$

respectively, where

$$B_{\mathcal{Z}}(z, h) = \{z' \in \mathcal{Z}, \delta(z', z) \le h\} \text{ and } B_{\mathcal{Y}}(y, h) = \{z' \in \mathcal{Y}, d(y', y) \le h\}.$$

Note: When $\mathcal{Z} = \mathcal{R}$ and is equipped with the Euclidean distance d, $\varphi_{\mathcal{Z},z}(h) = \int_{z-h}^{z+h} dF_Z$. Thus, in this case, $\varphi_{\mathcal{Z},z} = O(h)$.

To handle local constant smoothing with a non-Euclidean predictor Z we must also incorporate the following assumptions, which are adapted from Ferraty and Vieu (2006).

Assumption 1. $\forall \epsilon > 0$, $P(Z \in B_{\mathcal{Z}}(z, \epsilon)) = \varphi_{\mathcal{Z},z}(\epsilon) > 0$.

This extends the assumption that the marginal density f of Z is strictly positive.

Assumption 2. $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} \log n / n\varphi_{\mathcal{Z},z}(h) = 0$, *and* $\lim_{n\to\infty} nh^2 = \infty$.

The following assumption allows us to still consider unbounded $\mathcal{Z}$.

Assumption 3. $\forall m \geq 1$, $\forall y \in \mathcal{Y}$, *and* $\forall z \in \mathcal{Z}$, $E\{d^{2m}(Y, y)|Z = z\} < \sigma_{Ym}(z) < \infty$ *and* $E(|X_j|^m|Z = z) < \sigma_{Xjm}(z) < \infty$ *for* $j = 1, \ldots, p$, *where* $\sigma_{Ym}, \sigma_{X1m}, \ldots, \sigma_{Xpm}$ *are continuous at* z.

To control the effect of δ in the rate of convergence of the bias term, $d(s_{\oplus}(z), \widetilde{s}_{\oplus}(z))$, we make the following Lipschitz-type assumption.

Assumption 4. *There exists* $\beta_{0\mathbf{X}} > 0$ *and* $\beta_{0Y} > 0$ *such that*

$$
\begin{aligned}
&E(X_j|Z) \in Lip_{\mathcal{Z},\beta_{0\mathbf{X}}} \text{ for } j = 1, \ldots, p, \text{ and}\\
&E\{d^2(Y, y)|Z\} \in Lip_{\mathcal{Z},\beta_{0Y}} \text{ for any } y \in \mathcal{Y}, \text{ where}\\
&Lip_{\mathcal{Z},\beta_{0\mathbf{X}}} = \{f : \mathcal{Z} \to \mathcal{R}, \exists C_0 > 0, \forall z, z' \in \mathcal{Z}, |f(z) - f(z')| < C_0 \delta(z, z')^{\beta_{0\mathbf{X}}}\}\\
&\text{and } Lip_{\mathcal{Z},\beta_{0Y}} =\\
&\{f : \mathcal{Z} \times \mathcal{Y} \to \mathcal{R}, \exists C_0 > 0, \forall z, z' \in \mathcal{Z}, |f(y, z) - f(y, z')| < C_0 \delta(z, z')^{\beta_{0Y}}\}.
\end{aligned}
$$

Given the definition of type I and type II kernels provided in the Supplementary Material, we have the following assumption.

Assumption 5. *K is a kernel of type I or K is a kernel of type II and satisfies*

$$
\exists C_5 > 0,\ \exists \epsilon_0,\ \forall \epsilon < \epsilon_0,\ \int_0^{\epsilon} \varphi_{\mathcal{Z},z}(u) du > C_5 \epsilon \varphi_{\mathcal{Z},z}(\epsilon).
$$

This differs from the kernel assumptions in Petersen and Müller (2019) in that the kernels are no longer symmetric around 0. However, Assumption 5 ensures that Lemma 4.4 from Ferraty and Vieu (2006) holds, which will serve as a useful tool.

Finally, we require Assumptions L1, L3, L4, and P1 as defined in the Supplementary Material and used in Section 5.

6.1. Local constant Fréchet regression with only a non-Euclidean predictor

First consider the case when $\mathbf{X}$ is not a random vector, but rather a vector of constants. This reduces the setting to local constant Fréchet regression with a non-Euclidean predictor, as (4.2) becomes $s_{\oplus}(z) = \text{argmin}_{y\in\mathcal{Y}} E\{d^2(Y, y)|Z = z\}$, (5.2) becomes $\widetilde{s}_{\oplus}(z) = \text{argmin}_{y\in\mathcal{Y}} E\left\{\zeta_h(Z, z) d^2(Y, y)\right\}$, and (4.4) becomes $\widehat{s}_{\oplus}(z) = \text{argmin}_{y\in\mathcal{Y}} \sum_{i=1}^n s_{hi}(z) d^2(Y_i, y)$.

Theorem 3. *Suppose Assumptions* P1, 5, L1, L3, *and* 2–4 *hold. If* $h \to 0$, *then*

$$d(s_{\oplus}(z), \widetilde{s}_{\oplus}(z)) = O\{h^{\beta_{0Y}/(\beta_1-1)}\}.$$

Theorem 4. *If assumptions* P1, 5, L1, L4, *and* 1–3 *hold, then*

$$d(\widetilde{s}_{\oplus}(z), \widehat{s}_{\oplus}(z)) = O_p[\{n\varphi_{\mathcal{Z},z}(h)\}^{-1/2(\beta_2-1)}].$$

Corollary 2. *Under the Assumptions of Theorem* 3 *and Theorem* 4, *we have*

$$d(s_{\oplus}(z), \widehat{s}_{\oplus}(z)) = O(h^{\beta_{0Y}/(\beta_1-1)}) + O_p[\{n\varphi_{\mathcal{Z},z}(h)\}^{-1/2(\beta_2-1)}].$$

6.2. Partially-global Fréchet regression with mixed predictors

Finally, we provide the theory for the partially-global Fréchet regression model with Z coming from a generic metric space. That is, we keep the definitions of (6.1) and (6.2), corresponding to local constant smoothing, but we include the global contribution of a random vector $\mathbf{X} \in \mathcal{R}^p$ along with the local contribution from the random object $Z \in \mathcal{Z}$.

Theorem 5. *Suppose Assumptions* P1, 5, L1, L3, *and* 2–4 *hold. Then*

$$d(s_{\oplus}(\mathbf{x}, z), \widetilde{s}_{\oplus}(\mathbf{x}, z)) = O(h^{\beta_0/(\beta_1-1)}),$$

where $\beta_0 = \min\{\beta_{0\mathbf{X}}, \beta_{0Y}\}$.

Theorem 6. *If Assumptions* P1, 5, L1, L4, *and* 1–3 *hold, then*

$$d(\widetilde{s}_{\oplus}(\mathbf{x}, z), \widehat{s}_{\oplus}(\mathbf{x}, z),) = O_p[\{n\varphi_{\mathcal{Z},z}(h)\}^{-1/2(\beta_2-1)}].$$

Corollary 3. *If the assumptions of Theorems* 5 *and* 6 *hold, then*

$$d(s_{\oplus}(\mathbf{x}, z), \widehat{s}_{\oplus}(\mathbf{x}, z),) = O(h^{\beta_0/(\beta_1-1)}) + O_p[\{n\varphi_{\mathcal{Z},z}(h)\}^{-1/2(\beta_2-1)}],$$

where $\beta_0 = \min\{\beta_{0\mathbf{X}}, \beta_{0Y}\}$.

Similarly as aforementioned, the right hand side mimics the bias-variance tradeoff in classicial smoothing.

7. Simulation Studies

We now demonstrate the capability of the partially-global Fréchet regression model and consider simulated datasets with the response coming from two different metric spaces.

Definition 3. Let Ω_1 be the set of probability distributions. The 2-Wasserstein metric distance between two distributions with CDFs $H(\cdot)$ and $G(\cdot)$ is defined as

$$d_W(H,G) = \sqrt{\int_0^1 (H^{-1}(t) - G^{-1}(t))^2 dt}.$$

We denote (Ω_1, d_W) as the metric space of probability distributions equipped with the Wasserstein distance.

Definition 4. Let Ω_2 be the set of symmetric, positive definite (SPD) matrices. Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two SPD matrices. Then, under the Cholesky decomposition, we can write $\mathbf{P}_1 = (\mathbf{P}_1^{1/2})^T \mathbf{P}_1^{1/2}$ and $\mathbf{P}_2 = (\mathbf{P}_2^{1/2})^T \mathbf{P}_2^{1/2}$, where $\mathbf{P}_1^{1/2}$ and $\mathbf{P}_2^{1/2}$ are upper triangle matrices with positive diagonal components. The Cholesky decomposition metric distance between two SPD matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$, is defined as

$$d_C(\mathbf{P}_1, \mathbf{P}_2) = \sqrt{\operatorname{trace}\left\{(\mathbf{P}_1^{1/2} - \mathbf{P}_2^{1/2})^T (\mathbf{P}_1^{1/2} - \mathbf{P}_2^{1/2})\right\}}.$$

We denote (Ω_2, d_C) as the metric space of SPD matrices equipped with the Cholesky decomposition distance.

7.1. Partially-global Fréchet regression for probability distributions equipped with the Wasserstein distance

Consider $\mathcal{Y} \subset \Omega_1$. Data are generated by adapting the simulation example in Petersen and Müller (2019). Let $\rho = 0.5$. Raw predictors are generated in two steps:

(1) $(S_1, S_2, S_3)^T$ is multivariate Gaussian with $E(S_j) = 0$ and $\operatorname{cov}(S_j, S_{j'}) = \rho^{|j-j'|}$ for $1 \le j, j' \le 3$;

(2) Set $T = \Phi(S_3)$, where Φ is the standard normal distribution function, so that $T \sim Unif[0,1]$.

Example 1. An example when Z is scalar.

In this example, we consider $\mathcal{Z} \subset \mathcal{R}$. Set $X_1 = S_1$, $X_2 = S_2$, and $Z = T$, generated as above. The Fréchet regression function is given by

$$m_\oplus(\mathbf{x}, z) = E\{Y(\cdot)|\mathbf{X} = \mathbf{x}, Z = z\} = \mu_0 + \beta(x_1 + x_2) + \{\sigma_0 + \gamma \sin(\pi z)\}\Phi^{-1}(\cdot).$$

Conditional on $\mathbf{X}$ and Z, the random response Y is generated by adding noise as follows: $Y = \mu + \sigma\Phi^{-1}$ with $\mu|\mathbf{X} \sim N(\mu_0 + \beta(X_1 + X_2), \nu_1)$ and $\sigma|Z \sim \text{Gamma}(\{\sigma_0 + \gamma \sin(\pi Z)\}^2/\nu_2, \nu_2/\{\sigma_0 + \gamma \sin(\pi Z)\})$ being independently sampled. The additional parameters are set to be $\mu_0 = 0$, $\sigma_0 = 3$, $\beta = 3/4$, $\gamma = 3$, $\nu_1 = 1$, and $\nu_2 = 2$.

Because we are only considering Euclidean predictors in Example 1, we can implement partially-global Fréchet regression with both local constant and local linear smoothing, as well as local and global Fréchet regression. The local Fréchet model we implement for comparison uses the local linear smoother, as in Petersen and Müller (2019).

Example 2. An example when Z is a density.

In this example, we consider $\mathcal{Z} \subset \Omega_1$. Set $X_1 = S_1$, $X_2 = S_2$. Further, we set

$$E\{Z(\cdot)|T = t\} = \mu_{0z} + (\sigma_{0z} + \gamma_z t)\Phi^{-1}(\cdot).$$

Conditional on T, the predictor Z is generated by adding noise as follows: $Z = \mu_{0z} + \sigma_z \Phi^{-1}$ with $\sigma_z|T \sim \text{Gamma}((\sigma_{0z} + \gamma_z T)^2/\nu_z, \nu_z/(\sigma_{0z} + \gamma_z T))$ being independently sampled.

The Fréchet regression function is given by

$$m_\oplus(\mathbf{x}, z) = E\{Y(\cdot)|\mathbf{X} = \mathbf{x}, T = t\} = \mu_0 + \beta(x_1 + x_2) + [\sigma_0 + \gamma\{\sin(\pi t)\}]\Phi^{-1}(\cdot).$$

Conditional on $\mathbf{X}$ and T, the random response Y is generated by adding noise as follows: $Y = \mu + \sigma\Phi^{-1}$ with $\mu|\mathbf{X} \sim N(\mu_0 + \beta(X_1 + X_2), \nu_1)$ and $\sigma|T \sim \text{Gamma}([\sigma_0 + \gamma\{\sin(\pi T)\}]^2/\nu_2, \nu_2/\{\sigma_0 + \gamma\sin(\pi T)\})$ being independently sampled. The additional parameters are set as $\mu_0, \mu_{0z} = 0$, $\sigma_0, \sigma_{0z} = 3$, $\beta = 3/4$, $\gamma = 3$, $\gamma_z = 6$, $\nu_1, \nu_z = 1$, and $\nu_2 = 2$.

In Example 2, we can see that Y depends on Z through the latent variable T. Further, because Z is no longer Euclidean, we can only implement our partially-global Fréchet regression model with local constant smoothing.

Now, let $\mathbf{I}$ denote an $M \times M$ identity matrix and $\mathbf{U} = (U_{i,j})$ denote an $M \times M$ matrix where $U_{i,j} = I_{(j>i)}$.

Example 3. An example when Z is an SPD matrix.

Consider $\mathcal{Z} \subset \Omega_2$, the set of SPD matrices. Set $X_1 = S_1$ and $X_2 = S_2$. Further, set

$$E(Z|T = t) = E(\mathbf{A}|T = t)^T E(\mathbf{A}|T = t),$$

where $E(\mathbf{A}|T = t) = \{\mu_z + \beta_z t\}\mathbf{I} + \{\sigma_{0z} + \gamma_z t\}\mathbf{U}$. Conditional on T, the predictor Z is generated by adding noise as follows: $Z = \mathbf{A}^T\mathbf{A}$ where $\mathbf{A} = \mu_z\mathbf{I} + \sigma_z\mathbf{U}$, $\mu_z = \mu_{0z} + \beta_z T$, and with $\sigma_z|T \sim \text{Gamma}((\sigma_{0z} + \gamma_z T)^2/\nu_z, \nu_z/(\sigma_{0z} + \gamma_z T))$ being independently sampled.

The Fréchet regression function is given by

$$m_\oplus(\mathbf{x}, t) = E\{Y(\cdot)|\mathbf{X} = \mathbf{x}, T = t\} = \mu_0 + \beta(x_1 + x_2) + \{\sigma_0 + \gamma\sin(\pi t)\}\Phi^{-1}(\cdot).$$

The additional parameters are set $M = 5$, $\mu_0, \mu_{0z}, \sigma_{0z}, \sigma_0 = 0$, $\beta, \beta_z, \gamma, \gamma_z = 5$, and $\nu_1, \nu_2, \nu_z = 1$.

Training samples of size $n = 50, 100$, and 200 were used for all examples. For all partially-global and local models, independent tuning sets of size n were generated in the same way as the training set to perform bandwidth selection. As mentioned previously, the local Fréchet model uses the local linear smoother, as in Petersen and Müller (2019).

Table 1. GMSE (standard error) across 100 simulations for partially-global Fréchet regression with local constant smoothing (PGc) and with local linear smoothing (PGl), and local (Ll) and global (G) Fréchet regression models when the response is density data. For examples 2 and 3, PGl, Ll, and G Fréchet regression cannot be applied.

Example	n	PGc	PGl	Ll	G
1	50	2.936 (0.0208)	**2.902** (0.022)	3.562 (0.015)	3.469 (0.019)
	100	**2.666** (0.009)	2.670 (0.008)	3.434 (0.011)	3.258 (0.009)
	200	2.558 (0.004)	**2.543** (0.004)	3.307 (0.006)	3.160 (0.004)
2	50	3.152 (0.021)			
	100	2.932 (0.011)			
	200	2.788 (0.005)			
3	50	2.463 (0.021)			
	100	2.118 (0.011)			
	200	1.890 (0.006)			

Independent test sets of size $\tilde{n} = 2000$, denoted by $\{(\tilde{\mathbf{X}}_i, \tilde{Z}_i, \tilde{Y}_i) : i = 1, 2, \ldots, \tilde{n}\}$, were then generated to evaluate the performance of the estimated Fréchet regression function. On the testing data sets, we calculate a generalized MSE (GMSE) defined as $\sum_{i=1}^{\tilde{n}} \{d_W(\widehat{s}_\oplus(\tilde{\mathbf{X}}_i, \tilde{Z}_i), \tilde{Y}_i)\}^2/\tilde{n}$. Table 1 shows the simulation results.

From Table 1, we see that our partially-global models outperform the local and global Fréchet models when Z is a scalar predictor. The partially-global model with local constant smoothing also performs well in both Example 2 and Example 3 when the local predictor Z is non-Euclidean. We are not aware of any other model that can handle this highly complex data, and therefore cannot compare our model to other methods for these examples.

To study the complexity of these models empirically, we also compute the average fitting time of each model in Table 2. Note: for partially-global Fréchet regression with local constant smoothing and with local linear smoothing, and for local Fréchet regression, the grids used for bandwidth selection all had a total of 200 points for consideration and were comparable width. However, these grids were shifted according to the tuning error curves of each model to ensure that the global minimum of each error curve was realized during the tuning process.

Figure 1 takes a look at three observations from the testing data of Example 3. This example is of particular interest because it mixes random objects from Euclidean space and two different metric spaces. In the figure, we represent the SPD matrix predictors as graphs to gain a visual understanding. We treat each SPD matrix as a weighted adjacency matrix of a weighted graph. That is, the number of nodes in the graph is the dimension of the matrix, and the weights between nodes correspond to the elements in the matrix.

We notice from Figure 1 that fluctuations in the structure of the SPD matrices are reflected in the structure of the density outputs. Further, we see

Table 2. Average computation time in seconds (standard error) across 100 simulations for partially-global Fréchet regression with local constant smoothing (PGc) and with local linear smoothing (PGl), and local (Ll) and global (G) Fréchet regression models when the response is density data.

Example	n	PGc	PGl	Ll	G
1	50	5.847 (0.062)	5.877 (0.014)	3.587 (0.012)	0.330 (0.003)
	100	17.148 (0.071)	17.243 (0.031)	9.623 (0.021)	0.361 (0.004)
	200	59.089 (0.040)	59.406 (0.058)	27.867 (0.009)	0.364 (0.002)
2	50	19.638 (0.058)			
	100	43.039 (0.144)			
	200	142.996 (0.310)			
3	50	1,200.293 (0.717)			
	100	2,464.333 (1.953)			
	200	4,887.486 (3.954)			

that our predicted densities visually match the observed densities well, affirming the results in Table 1.

7.2. Partially-global Fréchet regression for SPD matrices equipped with the Cholesky decomposition distance

Consider $\mathcal{Y} \subset \Omega_2$. For the following examples, raw predictors are generated in two steps:

(1) $(S_1, S_2, S_3, S_4)^T$ is multivariate Gaussian with $E(S_j) = 0$ and $\text{cov}(S_j, S_{j'}) = \rho^{|j-j'|}$ for $1 \leq j, j' \leq 4$;

(2) Set $T = \Phi(S_4)$, where Φ is the standard normal distribution function, so that $T \sim Unif[0, 1]$.

We set $\rho = 0.5$. Let $\mathbf{I}_M$ denote an $M \times M$ identity matrix and $\mathbf{U}_M = (U_{i,j})$ denote an $M \times M$ matrix where $U_{i,j} = I_{(j>i)}$, as before.

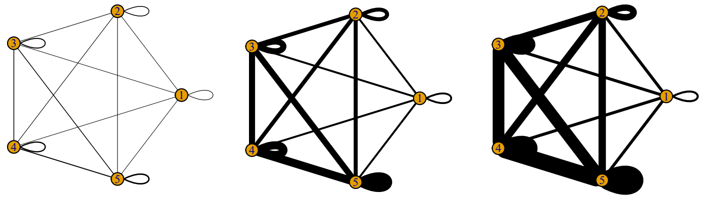
Example 4. An example when Z is scalar.

Consider $\mathcal{Z} \subset \mathcal{R}$. Further, we set $X_1 = S_1$, $X_2 = S_2$, $X_3 = S_3$, and $Z = T-1$. The Fréchet regression function is given by

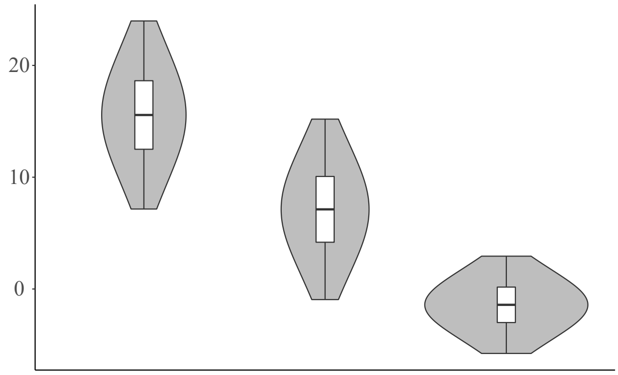
$$m_\oplus(\mathbf{x}, z) = E(Y|\mathbf{X} = \mathbf{x}, Z = z) = E(\mathbf{B}|\mathbf{X} = \mathbf{x}, Z = z)^T E(\mathbf{B}|\mathbf{X} = \mathbf{x}, Z = z)$$

where $E(\mathbf{B}|\mathbf{X} = \mathbf{x}, Z = z) = \{\mu_0 + \beta \sin(\pi z)\}\mathbf{I}_M + \{\mu_0 + \beta \sin(\pi z) + \sigma_0 + \gamma(x_1 + x_2 + x_3)\}\mathbf{U}_M$.

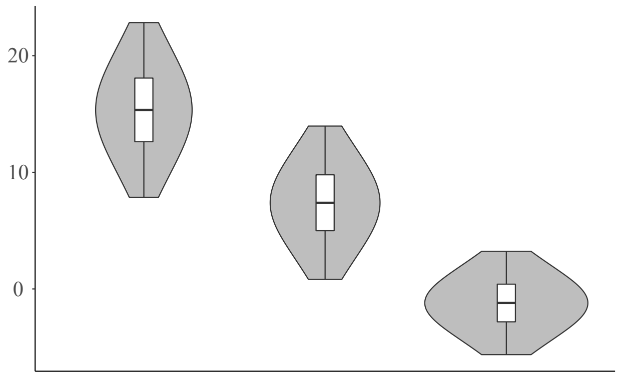
Conditional on $\mathbf{X}$ and Z, the random response Y is generated by adding noise as follows: $Y = \mathbf{B}^T\mathbf{B}$ where $\mathbf{B} = \mu\mathbf{I}_M + (\mu + \sigma)\mathbf{U}_M$ and with $\mu|Z \sim N(\mu_0 + \beta \sin(\pi Z), \nu_1)$ and $\sigma|\mathbf{X} \sim \text{Gamma}(\{\sigma_0 + \gamma(X_1 + X_2 + X_3)\}^2/\nu_2, \nu_2/\{\sigma_0 + \gamma(X_1 + X_2 + X_3)\})$ being independently sampled. The additional parameters are set as $M = 5$, $\mu_0, \sigma_0 = 3$, $\beta, \gamma = 4$, and $\nu_1, \nu_2 = 1/2$.



(a) Three observed SPD matrix predictors, represented as weighted graphs.



(b) Corresponding observed density responses.



(c) Corresponding predictions by PGl model.

Figure 1. Three observed SPD matrix predictors, observed density responses, and predicted density responses from Example 3.

Example 5. An example when Z is an SPD matrix.

We consider $\mathcal{Z} \subset \Omega_2$. Set $X_1 = S_1$, $X_2 = S_2$ and $X_3 = S_3$. Further, we set

$$E(Z|T = t) = E(\mathbf{A}|T = t)^T E(\mathbf{A}|T = t),$$

where $E(\mathbf{A}|T = t) = \{\mu_{0z} + \beta_z \sin(\pi t) + \sigma_{0z} + \gamma_z \sin(\pi t)\}\mathbf{I}_{M_z} + \{\sigma_{0z} + \gamma_z \sin(\pi t)\}\mathbf{U}_{M_z}$. Conditional on T, the predictor Z is generated by adding noise as follows: $Z = \mathbf{A}^T\mathbf{A}$ where $\mathbf{A} = (\mu_z + \sigma_z)\mathbf{I}_{M_z} + \sigma_z \mathbf{U}_{M_z}$ and with $\mu_z|T \sim$

$N(\mu_{0z} + \beta_z \sin(\pi T), \nu_{z1})$ and $\sigma_z | T \sim \text{Gamma}(\{\sigma_{0z} + \gamma_z \sin(\pi T)\}^2/\nu_{z2}, \nu_{z2}/\{\sigma_{0z} + \gamma_z \sin(\pi T)\})$ being independently sampled.

The Fréchet regression function is given by

$$m_\oplus(\mathbf{x}, z) = E(Y|\mathbf{X} = \mathbf{x}, T = t) = E(\mathbf{B}|\mathbf{X} = \mathbf{x}, T = t)^T E(\mathbf{B}|\mathbf{X} = \mathbf{x}, T = t),$$

where $E(\mathbf{B}|\mathbf{X} = \mathbf{x}, T = t) = \{\mu_0 + \beta(x_1 + x_2 + x_3) + \sigma_0 + \gamma \sin(\pi t)\}\mathbf{I}_M + \{\sigma_0 + \gamma \sin(\pi t)\}\mathbf{U}_M$. Conditional on $\mathbf{X}$ and T, the response Y is generated by adding noise as follows: $Y = \mathbf{B}^T\mathbf{B}$ where $\mathbf{B} = (\mu + \sigma)\mathbf{I}_M + \sigma\mathbf{U}_M$ and with $\mu|\mathbf{X}, T \sim N(\mu_0 + \beta(X_1 + X_2 + X_3), \nu_1)$ and $\sigma|T \sim \text{Gamma}(\sigma_0 + \gamma \sin(\pi T)^2/\nu_2, \nu_2/\{\sigma_0 + \gamma \sin(\pi T)\})$ being independently sampled. The additional parameters are set as $M_z = 3$, $M = 5$, $\mu_{0z}, \sigma_{0z}, \mu_0, \sigma_0 = 3$, $\nu_{z1}, \nu_1 = 1$, $\nu_{z2}, \nu_2 = 2$, $\beta_z, \beta = 2$, and $\gamma_z, \gamma = 3$.

Example 6. An example when Z is a density.

Here we consider $\mathcal{Z} \subset \Omega_1$. Set $X_1 = S_1$, $X_2 = S_2$ and $X_3 = S_3$. Further,

$$E\{Z(\cdot)|T = t\} = \mu_{0z} + \{\sigma_{0z} + \gamma_z \sin(\pi t)\}\Phi^{-1}(\cdot).$$

Conditional on T, the predictor Z is generated by adding noise as follows: $Z = \mu_{0z} + \sigma_z\Phi^{-1}$ with $\sigma_z|T \sim \text{Gamma}(\{\sigma_{0z} + \gamma_z \sin(\pi T)\}^2/\nu_z, \nu_z/\{\sigma_{0z} + \gamma_z \sin(\pi T)\})$ being independently sampled.

The Fréchet regression function is given by

$$m_\oplus(\mathbf{x}, t) = E(Y|\mathbf{X} = \mathbf{x}, T = t) = E(\mathbf{B}|\mathbf{X} = \mathbf{x}, T = t)^T E(\mathbf{B}|\mathbf{X} = \mathbf{x}, T = t),$$

where $E(\mathbf{B}|\mathbf{X} = \mathbf{x}, T = t) = \{\mu_0 + \beta(x_1 + x_2 + x_3) + \sigma_0 + \gamma \sin(\pi t)\}\mathbf{I}_M + \{\sigma_0 + \gamma \sin(\pi t)\}\mathbf{U}_M$. Conditional on $\mathbf{X}$ and T, the response Y is generated by adding noise as follows: $Y = \mathbf{B}^T\mathbf{B}$ where $\mathbf{B} = (\mu + \sigma)\mathbf{I}_M + \sigma\mathbf{U}_M$ and with $\mu|\mathbf{X}, T \sim N(\mu_0 + \beta(X_1 + X_2 + X_3), \nu_1)$ and $\sigma|T \sim \text{Gamma}(\sigma_0 + \gamma \sin(\pi T)^2/\nu_2, \nu_2/\{\sigma_0 + \gamma \sin(\pi T)\})$ being independently sampled. The additional parameters are set as $M = 8$, $\mu_{0z}, \sigma_{0z}, \mu_0, \sigma_0 = 3$, $\nu_z, \nu_1, \nu_2 = 1/2$, $\gamma_z, \gamma = 3$, and $\beta = 2$.

Training samples of size $n = 50, 100$, and 200 were used for all examples. For all partially-global and local models, independent tuning sets of size n were generated in the same way as the training set to perform bandwidth selection. Once again, we implement local linear smoothing for the local Fréchet regression model.

Independent test sets of size $\tilde{n} = 2000$, denoted by $\{(\tilde{\mathbf{X}}_i, \tilde{Z}_i, \tilde{Y}_i) : i = 1, 2, \ldots, \tilde{n}\}$, were then generated to evaluate the performance of the estimated Fréchet regression function. On the testing data sets, we calculate a generalized MSE (GMSE) defined as $\sum_{i=1}^{\tilde{n}}\{d_C(\widehat{s}_\oplus(\tilde{\mathbf{X}}_i, \tilde{Z}_i), \tilde{Y}_i)\}^2/\tilde{n}$. Table 3 presents the simulation results.

As expected, Table 3 shows that the partially-global Fréchet regression models outperform the local and global Fréchet regression models in Example

Table 3. GMSE (standard error) across 100 simulations for partially-global Fréchet regression with local constant smoothing (PGc) and with local linear smoothing (PGl), and local (Ll) and global (G) Fréchet regression models when the output is SPD matrix data. For examples 5 and 6, PGl, Ll, and G Fréchet regression cannot be applied.

Example	n	PGc	PGl	Ll	G
4	50	18.016 (1.356)	**15.297** (1.141)	29.287 (1.350)	34.524 (1.461)
	100	17.574 (1.417)	**14.907** (1.179)	25.469 (1.314)	33.269 (1.452)
	200	17.281 (1.425)	**14.869** (1.194)	21.310 (1.291)	32.719 (1.451)
5	50	21.027 (1.658)			
	100	19.240 (1.623)			
	200	18.453 (1.661)			
6	50	43.978 (2.976)			
	100	43.390 (3.089)			
	200	42.882 (3.105)			

Table 4. Average computation time in seconds (standard error) across 100 simulations for partially-global Fréchet regression with local constant smoothing (PGc) and with local linear smoothing (PGl), and local (Ll) and global (G) Fréchet regression models when the output is SPD matrix data.

Example	n	PGc	PGl	Ll	G
4	50	12.657 (0.036)	8.532 (0.008)	34.252 (0.024)	0.963 (0.004)
	100	43.353 (0.048)	27.338 (0.048)	146.206 (0.366)	1.923 (0.005)
	200	194.245 (0.399)	93.987 (0.042)	525.453 (0.888)	3.812 (0.011)
5	50	55.643 (0.046)			
	100	239.467 (0.387)			
	200	878.075 (0.716)			
6	50	13.175 (0.041)			
	100	45.824 (0.047)			
	200	180.338 (0.468)			

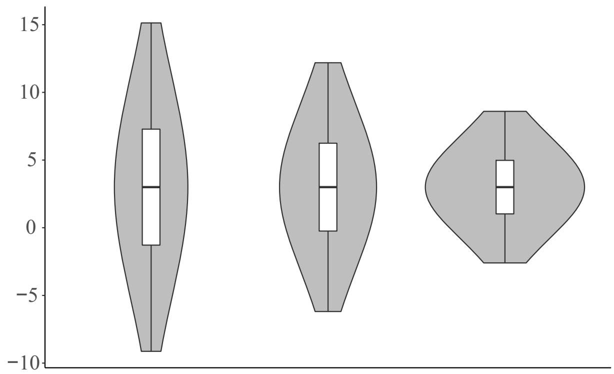
4 and also perform well when Z comes from a generic metric space. Empirical computation time is reported in Table 4.

Figure 2, similar to Figure 1, takes a look at three observations from the testing data of Example 6. We see that not only are the fluctuations in the density predictors reflected in the SPD matrix responses, but also our partially-global Fréchet regression model captures this relationship well.

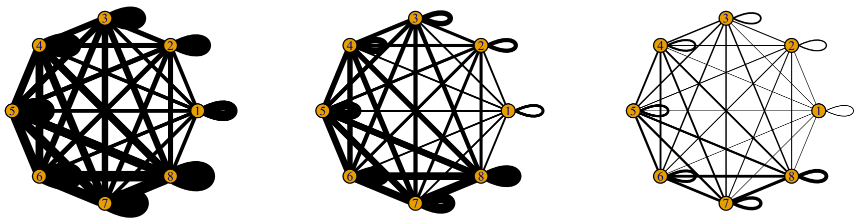
8. Real Data Applications

8.1. Bike rental distribution regression

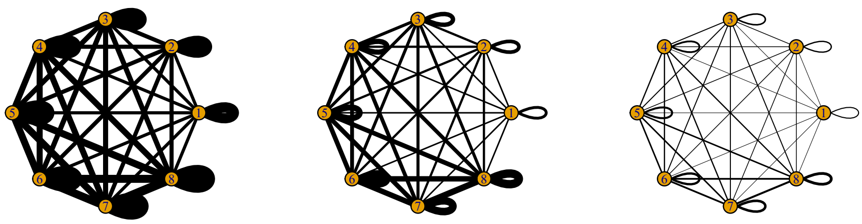
We first demonstrate the flexibility of partially-global Fréchet regression applied to bike rental data collected by Capital Bikeshare in Washington, D.C. This data set spans the years 2011 and 2012 for a total of 731 days. For each day,



(a) Three observed density predictors.



(b) Corresponding observed SPD matrix responses.



(c) Corresponding predictions by PGl model.

Figure 2. Three observed density predictors, observed SPD matrix responses, and predicted SPD matrix responses from Example 6.

there are 24 hourly observations of bike rental counts as well as the following 7 predictors (Fanaee-T and Gama, 2013):

- RBW: Indicator of bad weather (snowy and/or rainy), standardized
- Work: Indicator of neither the weekend nor a holiday, standardized
- Spring: Indicator of Spring season, standardized
- Summer: Indicator of Summer season, standardized
- Fall: Indicator of Fall season, standardized
- Year: Indicator of the year 2012, standardized

Table 5. GMSE (standard error) for partially-global Fréchet regression with local constant smoothing (PGc), partially-global Fréchet regression with local linear smoothing (PGl), and local (Ll) and global (G) Fréchet regression models when applied to the bike rental density data over 100 repetitions.

PGc	PGl	Ll	G
2,215.315 (13.191)	**2,191.892** (12.828)	2,404.486 (21.250)	2,549.441 (13.669)

- Temp: Daily mean temperature, standardized

We construct the response for each day to be the 24 observed quantiles for an underlying distribution of bike rental counts. We randomly split the data into a training set of size $n = 366$ and a testing set of size $\tilde{n} = 365$. We consider four models for the data: a partially-global model with local constant smoothing on the local predictor Temp (PGc), a partially-global model with local linear smoothing on the local predictor Temp (PGl), a local model utilizing local linear smoothing on all predictors (Ll), and a global model on all predictors (G). For the partially-global and the local models, we performed 10-fold cross validation on the training set to tune the bandwidth parameter, h. We considered $h \in \{0.2, 0.25, 0.3 \ldots, 0.9, 0.95, 1\}$. We then computed the GMSE of each model when applied to the testing set. We repeated this process 100 times. The results are recorded in Table 5.

From Table 5, we see that the partially-global models perform better than both the local and the global models in terms of prediction errors. The generalized $R^2_{\oplus} = 0.952$ for the partially-global model with linear smoothing (for details see Petersen and Müller, 2019). Figure 3 depicts the predicted distributions (after spline smoothing for prettier visuals) of all models for one observation in the testing set. Box plots are also superimposed on each distribution image. From this figure, we can see that our partially-global models match the true distribution's overall shape much better than the local and global models.

8.2. New York taxi network regression

Finally, we demonstrate the performance of our partially-global Fréchet regression model on real SPD matrix data. The New York City Taxi and Limousine Commission provides records on pick-up and drop-off dates and times, pick-up and drop-off locations, trip distances, and itemized fares for yellow taxis which are available from `https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page`. We transform this data into graph data, where neighborhoods are nodes and edges are weighted by the number of taxi rides which picked up in one neighborhood and dropped off in another within a single hour. After proper transformation, these graphs lie in a metric space of SPD matrices equipped with the Choleksy decomposition distance, as in Section 7.2.

To engineer SPD matrices from the taxi data, we do the following:

Figure 3. Predicted bike rental distributions for one observation in the testing data set from one random repetition. The quantiles of each distribution are smoothed using cubic splines to depict estimates of the exact underlying distributions.

1. We filter the data on the month of January 2016 due to resource restrictions.
2. We further filter on observations with both pick-up and drop-off occurring in Manhattan.
3. We then label the corresponding neighborhood for each pick-up and drop-off in the same manner as Dubey and Müller (2020).
4. For each hour, we collect the number of pairwise connections between nodes based on taxi pick-ups and drop-offs. These correspond to weights between nodes on a graph.

By doing this, we collect 723 weighted adjacency matrices of dimension 10×10 for our data set (removing a small handful of observations due to their sparsity). To ensure that these outputs are truly SPD matrices, we further square them.

From the taxi data, we also collect the following Euclidean predictors for each hour:

- Late Hour: Indicator for the hour being between 11pm and 5am, standardized
- Credit: Sum of credit indicator for type of payment, standardized

To showcase the ability of partially-global Fréchet regression to utilize non-Euclidean predictors, we consider a third predictor: the distribution of taxi ride distances. That is, for each hour we collect 20 quantiles from the observed taxi ride distances and call this non-Euclidean predictor Distance.

Table 6. GMSE (standard error) for partially-global Fréchet regression with local constant smoothing (PGc), and local (Ll) and global (G) Fréchet regression models when applied to taxi ride data with SPD matrix output over 100 repetitions.

PGc	Ll	G
320.754 (0.907)	323.820 (0.976)	329.651 (0.945)

We implement three models on the data: a partially-global model with Late Hour and Credit as global predictors and Distance as a local predictor, a local model which utilizes local linear smoothing with only the Euclidean predictors, and a global model with only the Euclidean predictors. To determine the efficacy of each model, we randomly split the data into a training set of size 361 and testing set of size 362. For the partially-global and the local models, we performed 5-fold cross validation on the training set to tune the bandwidth parameter, h, considering $h \in \{0.6, 0.8, 1 \ldots, 7.6, 7.8, 8\}$. We then computed the GMSE of each model when applied to the testing set. We repeated this process 100 times. The results are shown in Table 6.

From the results, we see that the partially-global model performs better than both the local and the global models in terms of prediction error. Therefore, including the non-Euclidean predictor, Distance, is apparently important. For one random repetition, the generalized $R^2_{\oplus} = 0.490$ for the partially-global model. This $R^2_{\oplus}$ value is not as impressive as that of the bike data, but is still respectable, considering that we did not perform variable selection.

Figure 4 depicts the predicted taxi networks from all models for one observation in the testing data set produced from one random repetition. From this figure, we see further evidence that our partially-global model best captures the nature of the complex taxi networks. In particular, both the local and global model depict a more connected taxi network, when in fact only a few connections should be present.

9. Concluding Remarks

In this work, we have developed the partially-global Fréchet regression model, a flexible model which not only capitalizes on the strengths of both global and local Fréchet regression but also allows for the incorporation of a non-Euclidean predictor. We have derived its pointwise convergence rates and have further provided simulation and real data examples to demonstrate its competitive finite sample performance on highly complex data types.

The realm of potential applications for the partially-global Fréchet regression model is vast. In this paper, we only considered density and SPD matrix data, only utilizing one choice of a metric for each. However, many more data types from varying metric spaces can now be modeled while maintaining the original structure of the data and retaining information that scalar summaries simply

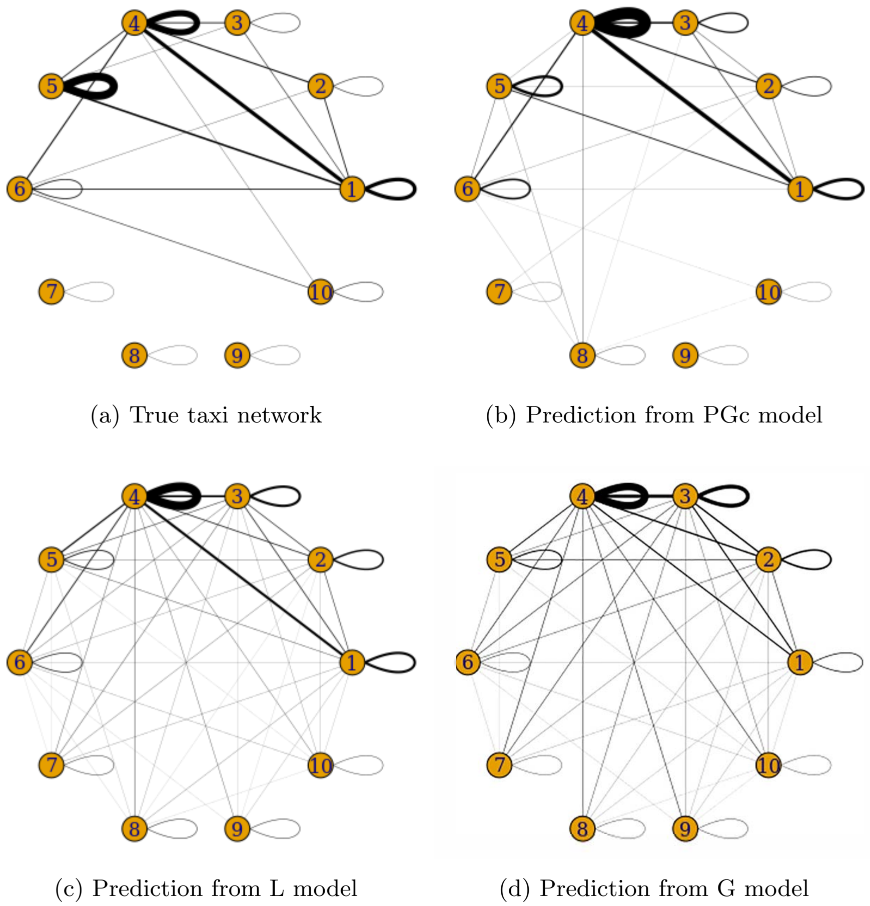


Figure 4. Predicted taxi networks for one observation in the testing data set from one random repetition. Each node represents a neighborhood in Manhattan. Each connection between nodes is weighted by the number of taxi rides between neighborhoods over a given hour time frame.

cannot provide. It is our opinion that much more research should be done in this area to meet the needs of every upcoming practical application. In short, there is still much to explore in this exciting new area of statistics.

Supplementary Material

All definitions, assumptions, and proofs are collected in the separate Supplementary Material document.

Acknowledgments

The authors thank the co-editor, an associate editor, and one reviewer for their constructive suggestions and comments that lead to substantial improvements in the paper. Tucker's research is partially supported by National Science Foundation grant DMS-1821171. Wu's research is partially supported by National Institutes of Health grant RF1MH125928 and National Science Foundation grants DMS-1821171, DMS-2152070 and CCF-2216912.

References

Cornea, E., Zhu, H., Kim, P. and Ibrahim, J. G. (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **79**, 463–482.

Davis, B. C., Fletcher, P. T., Bullitt, E. and Joshi, S. (2010). Population shape regression from random design data. *International Journal of Computer Vision* **90**, 255–266.

Ding, Z., Fleishman, G., Yang, X., Thompson, P., Kwitt, R. and Niethammer, M. (2019). Fast predictive simple geodesic regression. *Medical Image Analysis* **56**, 193–209.

Dubey, P. and Müller, H.-G. (2020). Functional models for time-varying random objects. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **82**, 275–327.

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

Fanaee-T, H. and Gama, J. (2013). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence* **2**, 113–127.

Faraway, J. J. (2014). Regression for non-Euclidean data using distance matrices. *Journal of Applied Statistics* **41**, 2342–2357.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York.

Fletcher, P. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds: Mathematical methods for medical imaging. *International Journal of Computer Vision* **105**, 171–185.

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

Gonzalez-Manteiga, W., Henry, G. and Rodriguez, D. (2012). Partly linear models on Riemannian manifolds. *Journal of Applied Statistics* **39**, 1797–1809.

Han, K., Müller, H.-G. and Park, B. U. (2020). Additive functional regression for densities as responses. *Journal of the American Statistical Association* **115**, 997–1010.

Härdle, W. K., Liang, H. and Gao, J. (2000). *Partially Linear Models*. Physica-Verlag.

Hein, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems* (Edited by Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta). Curran Associates, Inc.

Hinkle, J., Muralidharan, P., Fletcher, P. T. and Joshi, S. (2012). Polynomial regression on Riemannian manifolds. In *Computer Vision–ECCV 2012*, Lecture Notes in Computer Science, 1–14. Springer, Berlin, Heidelberg.

Lin, Z. and Müller, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *The Annals of Statistics* **49**, 3510–3533.

Marron, J. S. and Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal* **56**, 732–753.

Niethammer, M., Huang, Y. and Vialard, F. (2011). Geodesic regression for image time-series. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, Lecture Notes in Computer Science, 655–662. Springer, Berlin, Heidelberg.

Pelletier, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics* **18**, 57–67.

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

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of Statistics* **20**, 1768–1802.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)* **50**, 413–436.

Wang, H. and Marron, J. S. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics* **35**, 1849–1873.

Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012). Local polynomial regression for symmetric positive definite matrices: Local polynomial regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **74**, 697–719.

(Received November 2021; accepted May 2023)