Statistica Sinica Preprint No: SS-2025-0349	
Title	Varying-coefficient Fréchet Regression
Manuscript ID	SS-2025-0349
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0349
Complete List of Authors	Yanzhao Wang, Jianqiang Zhang and Wangli Xu
Corresponding Authors	Wangli Xu
E-mails	wlxu@ruc.edu.cn
Notice: Accepted author version.	

Statistica Sinica: Newly accepted Paper

Varying-Coefficient Fréchet Regression

Yanzhao Wang*, Jianqiang Zhang† and Wangli Xu*

** Center for Applied Statistics and School of Statistics,Renmin University of China*

† School of Statistics and Data Science, Zhejiang Gongshang University

Abstract: As a growing number of problems involve variables that are random objects, the development of models for such data has become increasingly important. This paper introduces a novel varying-coefficient Fréchet regression model that extends the classical varying-coefficient framework to accommodate random objects as responses. The proposed model provides a unified methodology for analyzing both Euclidean and non-Euclidean response variables. We develop a comprehensive estimation procedure that accommodates diverse predictor settings. Specifically, the model allows the effect-modifier variable U to be either Euclidean or non-Euclidean, while the predictors $\boldsymbol{X}$ are assumed to be Euclidean. Tailored estimation methods are provided for each scenario. To examine the asymptotic properties of the estimators, we introduce a smoothed version of the model and establish convergence rates through separate theoretical analyses of the bias and stochastic terms. The effectiveness and practical utility of the proposed methodology are demonstrated through extensive simulation studies and a real-data application.

Keywords: Fréchet regression, metric space, random objects, varying-coefficient model.

Statistica Sinica: Newly accepted Paper

1. Introduction

Statistical analysis of non-Euclidean data has gained significant attention due to the increasing availability of complex data structures. Relevant examples include probability distributions, covariance matrices, phylogenetic trees, and graph networks. For instance, Zhang et al. (2025) studied the distribution of stock market returns to monitor structural changes over time. Dryden et al. (2009) analyzed brain images as covariance matrices to examine the diffusion of water molecules in the brain. Nye et al. (2017) explored genomic data by investigating phylogenetic trees for dimensionality reduction and structure discovery. Dubey and Müller (2022) examined the dynamics of time-varying networks, focusing on the evolving connectivity within brain networks and other complex systems. Given these developments, investigating the relationship between predictors and responses becomes crucial when the responses are metric space-valued random objects rather than Euclidean variables.

In the literature, various models have been developed to analyze cases where the response consists of random objects while the predictors are Euclidean variables. Among these, Petersen and Müller (2019) introduced the Fréchet regression method, which accommodates both linear and nonparametric relationships between the response and predictor variables. Building

Statistica Sinica: Newly accepted Paper

upon this foundation, Bhattacharjee and Müller (2023) proposed the single index Fréchet regression, which enables effective dimension reduction by projecting a multivariate predictor onto a single direction vector. Tucker and Wu (2025) developed the partially-global Fréchet regression to combine the local and global Fréchet regression, which generalizes the partially linear regression model. To mitigate the curse of dimensionality, Qiu et al. (2024) introduced a random forest weighted local Fréchet regression, which relies on a locally adaptive kernel generated by random forests. More recently, Iao et al. (2025) employed deep neural networks to develop deep Fréchet regression, which captures relationships between non-Euclidean responses and high-dimensional Euclidean predictors without imposing parametric assumptions. In summary, the development of Fréchet regression has progressed from parametric models towards flexible nonparametric and semiparametric frameworks, and further into high-dimensional and machine learning domains.

Although the existing literature has extensively studied various Fréchet regression models, these models naturally reduce to their Euclidean counterparts when the response variables lie in Euclidean space. For instance, global Fréchet regression simplifies to standard multiple linear regression (Petersen and Müller, 2019), while partially-global Fréchet regression re-

Statistica Sinica: Newly accepted Paper

duces to classical partially linear regression model in the special case where random objects lie in Euclidean space (Tucker and Wu, 2025). Nevertheless, when dealing with Euclidean response variables, varying-coefficient models represent another important class of regression approaches that has not been similarly generalized to the Fréchet regression. These models capture dynamic relationships by allowing regression parameters to vary with covariates or over time, rather than being fixed constants. For example, Pei et al. (2022) analyzed the relationship between income and pollution, showing that the impacts varied with energy consumption. Since their initial proposal by Hastie and Tibshirani (1993), these models have generated a substantial body of follow-up research. On the estimation side, kernel based local polynomial methods have been widely adopted for estimating functional coefficients by Fan and Zhang (1999). With further developments on efficient estimation, Huang et al. (2002) introduced basis function and spline based techniques to provide flexible global approximations suited for the repeated measurements. For inference on coefficient variability, existing tests are based on the discrepancy between the restricted and unrestricted sum of squared residuals using smooth estimates of the varying coefficients. Fan et al. (2001) developed the generalized likelihood ratio tests to examine whether coefficient functions are constant or vary over the domain, with

desirable wilks phenomenon properties facilitating inference. In addition, Huang et al. (2002) proposed resampling-based subject bootstrap procedures to construct confidence regions and to perform hypothesis testing.

Varying-coefficient models represent a vital class of regression methods for characterizing relationships among variables in Euclidean spaces. For non-Euclidean responses, Bhattacharjee and Müller (2022) developed concurrent object regression to characterize time-varying relationships with Euclidean predictors. In particular, their partially global concurrent object regression framework provides an important foundation for varying-coefficient modeling with non-Euclidean responses and Euclidean predictors. To accommodate both Euclidean and non-Euclidean predictors, this study establishes an estimation framework for varying coefficient Fréchet models, with the following principal innovations:

1. We develop a varying-coefficient Fréchet regression model that extends the classical varying-coefficient framework to accommodate random objects as both responses and effect modifiers. This extension provides a unified methodology for analyzing response variables that are either Euclidean or non-Euclidean, without relying on parametric assumptions.

2. We develop an estimation procedure for the proposed varying-coefficient

Statistica Sinica: Newly accepted Paper

Fréchet regression model. Unlike existing methods for random object responses, which are typically limited to Euclidean predictors, our approach accommodates predictors that are themselves random objects residing in metric spaces.

3. To investigate the asymptotic convergence rate of the estimators, we introduce a smoothed version of the varying-coefficient Fréchet regression model. By separately analyzing the bias term and the stochastic term, we theoretically establish the convergence rate of the estimator. The effectiveness of the proposed method is validated through simulation studies and a real-data application.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive review of varying-coefficient models and corresponding estimation methods. Section 3 introduces the proposed varying-coefficient Fréchet regression, including its definition and estimators. Section 4 establishes the consistency and convergence rates of the proposed estimators. Section 5 presents simulation studies with responses including scalars, probability distributions, and symmetric positive-definite matrices. Section 6 illustrates the empirical utility of the method through an application to human mortality data. Finally, Section 7 concludes with a discussion of methodological implications and potential directions for future research.

2. Varying-coefficient model

To better introduce the varying-coefficient Fréchet model, this section first presents the properties of varying-coefficient models in Euclidean space. A varying-coefficient model with Euclidean responses takes the form

$$Y = \beta_0(U) + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}(U) + \epsilon, \tag{2.1}$$

where Y is the response variable, the functional coefficient intercept $\beta_0(U)$ varies with the scalar $U \in \mathbb{R}$, $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$ is a p-dimensional vector of predictor variable, $\boldsymbol{\beta}(U) = (\beta_1(U), \beta_2(U), \ldots, \beta_p(U))^{\mathrm{T}}$ is a vector of coefficient function, and random error ϵ satisfying $\mathbb{E}(\epsilon|\boldsymbol{X}, U) = 0$. For every $u \in \mathbb{R}$, we have

$$\mathbb{E}(Y \mid U = u) = \beta_0(u) + \mathbb{E}(\boldsymbol{X} \mid U = u)^{\mathrm{T}}\boldsymbol{\beta}(u). \tag{2.2}$$

From equations (2.1) and (2.2), we obtain $Y - \mathbb{E}(Y \mid U = u) = (\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))^{\mathrm{T}}\boldsymbol{\beta}(u) + \epsilon$. Thus, $\boldsymbol{\beta}(u)$ can be obtained as follows

$$\begin{aligned}\boldsymbol{\beta}(u) \;=\; & (\mathbb{E}((\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))(\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))^{\mathrm{T}} \mid U = u))^{-1} \\ & \times \mathbb{E}\left((\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))(Y - \mathbb{E}(Y \mid U = u)) \mid U = u\right). \end{aligned} \tag{2.3}$$

To estimate $\boldsymbol{\beta}(u)$ in equation (2.3), we first estimate $\mathbb{E}(\boldsymbol{X} \mid U = u)$ and $\mathbb{E}(Y \mid U = u)$, denoted by $\widehat{\mathbb{E}}(\boldsymbol{X} \mid U = u)$ and $\widehat{\mathbb{E}}(Y \mid U = u)$, respectively. Suppose the random sample is $\{(Y_i, \boldsymbol{X}_i, U_i) : i = 1, \ldots, n\}$. The

estimators $\widehat{\mathbb{E}}(\boldsymbol{X} \mid U = u)$ and $\widehat{\mathbb{E}}(Y \mid U = u)$ are given by $\widehat{\mathbb{E}}(\boldsymbol{X} \mid U = u) = n^{-1}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{X}_i$ and $\widehat{\mathbb{E}}(Y \mid U = u) = n^{-1}\sum_{i=1}^{n} s_{in}(u)Y_i$, where the weighting function $s_{in}(u)$ can be either the local constant weighting $s^c_{in}(u)$ (Nadaraya, 1964; Watson, 1964) or the local linear weighting $s^l_{in}(u)$ (Fan and Gijbels, 1996) with their explicit forms given by

$$s^c_{in}(u) = \frac{K_h\left(U_i - u\right)}{n^{-1}\sum_{j=1}^{n} K_h(U_j - u)}, \quad s^l_{in}(u) = K_h(U_i - u)\frac{\widehat{\mu}_2(u) - \widehat{\mu}_1(u)(U_i - u)}{\widehat{\mu}_0(u)\widehat{\mu}_2(u) - \widehat{\mu}_1^2(u)}. \tag{2.4}$$

Here $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a smoothing kernel function, h is a bandwidth, and $\widehat{\mu}_j(u) = n^{-1}\sum_{i=1}^{n} K_h(U_i - u)(U_i - u)^j$ for $j = 0, 1, 2$. Hereafter, unless otherwise specified, the weight function $s_{in}(u)$ may denote either the local constant weight $s^c_{in}(u)$ or the local linear weight $s^l_{in}(u)$, and will be abbreviated as s_{in} for simplicity. Using these notations, the estimator $\widehat{\boldsymbol{\beta}}(u)$ takes the form

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(u) \;=\; & \Big(\frac{1}{n}\sum_{i=1}^{n} s_{in}(\boldsymbol{X}_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}\boldsymbol{X}_j)(\boldsymbol{X}_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}\boldsymbol{X}_j)^{\mathrm{T}}\Big)^{-1} \\ & \times\frac{1}{n}\sum_{i=1}^{n} s_{in}(\boldsymbol{X}_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}\boldsymbol{X}_j)(Y_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}Y_j).\end{aligned} \tag{2.5}$$

Based on equation (2.2), the intercept term $\beta_0(u)$ can be estimated by

$$\widehat{\beta}_0(u) = \widehat{\mathbb{E}}(Y \mid U = u) - \widehat{\mathbb{E}}(\boldsymbol{X} \mid U = u)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u). \tag{2.6}$$

Based on the estimated coefficients $\widehat{\boldsymbol{\beta}}(u)$ and $\widehat{\beta}_0(u)$, the predicted re-

sponse at any given covariate vector $\boldsymbol{x}$ and location u is given by

$$\widehat{Y} = \widehat{\beta}_0(u)+\boldsymbol{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u) = \frac{1}{n}\sum_{j=1}^{n} s_{jn}Y_j+(\boldsymbol{x}-\frac{1}{n}\sum_{k=1}^{n} s_{kn}\boldsymbol{X}_k)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u) = \sum_{j=1}^{n} c_j(\boldsymbol{x},u)Y_j, \tag{2.7}$$

where

$$c_j(\boldsymbol{x},u) = \frac{s_{jn}}{n}+(\boldsymbol{x}-\frac{1}{n}\sum_{k=1}^{n} s_{kn}\boldsymbol{X}_k)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{X}|\boldsymbol{u}}\sum_{i=1}^{n}\frac{s_{in}}{n}(\boldsymbol{X}_i-\frac{1}{n}\sum_{k=1}^{n} s_{kn}\boldsymbol{X}_k)(\delta_{ij}-\frac{s_{jn}}{n}),$$

with the conditional covariance matrix estimator defined as

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}|\boldsymbol{u}} = \frac{1}{n}\sum_{i=1}^{n} s_{in}(\boldsymbol{X}_i - \frac{1}{n}\sum_{k=1}^{n} s_{kn}\boldsymbol{X}_k)(\boldsymbol{X}_i - \frac{1}{n}\sum_{k=1}^{n} s_{kn}\boldsymbol{X}_k)^{\mathrm{T}}.$$

Here, the Kronecker delta δ_{ij} is defined by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. According to the above analysis and noting that $\sum_{j=1}^{n} c_j(\boldsymbol{x},u) = 1$, we can express the prediction in (2.7) as the solution to the following weighted least squares minimization problem

$$\widehat{Y} = \widehat{\beta}_0(u) + \boldsymbol{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u) = \arg\min_{y\in\mathcal{Y}} \sum_{j=1}^{n} c_j(\boldsymbol{x},u)\,(Y_j - y)^2\,. \tag{2.8}$$

3. Varying-coefficient Fréchet regression

Before presenting the definition of varying-coefficient Fréchet regression and its estimation methods, we introduce some preliminary notation and assumptions. Specifically, let $(\mathcal{Y}, d)$ and $(\mathcal{U}, \delta)$ be metric spaces, with metrics d and δ defined on the sets $\mathcal{Y}$ and $\mathcal{U}$, respectively. We consider

a random triple of objects $(\boldsymbol{X}, U, Y) \sim F$ defined in the product space $\mathbb{R}^p \times \mathcal{U} \times \mathcal{Y}$. Here, $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$ represents a p-dimensional real-valued predictor and $U \in \mathcal{U}$ is another predictor. The response variable $Y \in \mathcal{Y}$ is a random object associated with both $\boldsymbol{X}$ and U. F denotes the joint distribution of $(\boldsymbol{X}, U, Y)$, with marginal distributions $F_{\boldsymbol{X}}$, F_U and F_Y. The conditional distributions $F_{Y|(\boldsymbol{X},U)}$, $F_{\boldsymbol{X}|U}$ and $F_{Y|U}$ are assumed to exist. The conditional covariance matrix of $\boldsymbol{X}$ given $U = u$ is $\mathrm{Cov}(\boldsymbol{X} \mid U = u) = \mathbb{E}((\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))(\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))^{\mathrm{T}} \mid U = u)$. For a fixed $y \in \mathcal{Y}$, the conditional cross-covariance matrix between $\boldsymbol{X}$ and $d^2(Y, y)$ given $U = u$ is $\mathrm{Cov}(\boldsymbol{X}, d^2(Y, y) \mid U = u) = \mathbb{E}((\boldsymbol{X} - \mathbb{E}(\boldsymbol{X} \mid U = u))(d^2(Y, y) - \mathbb{E}(d^2(Y, y) \mid U = u)) \mid U = u)$.

Building upon the above notation, we extend the varying-coefficient model $Y = \beta_0(U) + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}(U) + \epsilon$, where the response Y is Euclidean, to the case where Y is a random object. Motivated by Petersen and Müller (2019), which generalizes linear regression with Euclidean responses to settings where the response is a random object, we reformulate the varying-coefficient regression model by replacing the Euclidean distance with the intrinsic metric d on $\mathcal{Y}$. This leads to the following definition of the varying-coefficient Fréchet regression model.

Definition 1 (Varying-coefficient Fréchet regression)**.** We denote the con-

ditional Fréchet regression function of Y given $\boldsymbol{X} = \boldsymbol{x}$ and $U = u$ as $m_{\oplus}(\boldsymbol{x}, u) = \underset{y \in \mathcal{Y}}{\operatorname{argmin}} \mathbb{E}(d^2(Y, y) \mid \boldsymbol{X} = \boldsymbol{x}, U = u)$. The varying-coefficient Fréchet regression model is said to hold if $m_{\oplus}(\boldsymbol{x}, u) = s_{\oplus}(\boldsymbol{x}, u)$ for any $\boldsymbol{x} \in \mathbb{R}^p$ and $u \in \mathcal{U}$, where $s_{\oplus}(\boldsymbol{x}, u)$ is defined by

$$s_{\oplus}(\boldsymbol{x}, u) = \arg\min_{y \in \mathcal{Y}} S_{\oplus}(y; \boldsymbol{x}, u), \tag{3.1}$$

with $S_{\oplus}(y; \boldsymbol{x}, u) = w_0(y; u) + w_1(\boldsymbol{x}, u) w_2^{-1}(u) w_3(y; u)$. Here, $w_0(y; u)$, $w_1(\boldsymbol{x}, u)$, $w_2(u)$, and $w_3(y; u)$ are defined as $w_0(y; u) = \mathbb{E}(d^2(Y, y) \mid U = u)$, $w_1(\boldsymbol{x}, u) = \boldsymbol{x}^{\mathrm{T}} - \mathbb{E}(\boldsymbol{X} \mid U = u)^{\mathrm{T}}$, $w_2(u) = \operatorname{Cov}(\boldsymbol{X} \mid U = u)$, and $w_3(y; u) = \operatorname{Cov}(\boldsymbol{X}, d^2(Y, y) \mid U = u)$.

To estimate the varying-coefficient Fréchet regression model, it is necessary to estimate $S_{\oplus}(y; \boldsymbol{x}, u)$, which is denoted as $\widehat{S}_{\oplus}(y; \boldsymbol{x}, u)$, and correspondingly obtain the estimate of $s_{\oplus}(\boldsymbol{x}, u)$, denoted as $\widehat{s}_{\oplus}(\boldsymbol{x}, u)$. Based on an independent and identically distributed sample $\{(\boldsymbol{X}_i, U_i, Y_i) : i = 1, 2, \ldots n\}$ from the joint distribution F, the estimate for the varying-coefficient Fréchet regression model is defined as

$$\widehat{s}_{\oplus}(\boldsymbol{x}, u) = \arg\min_{y \in \mathcal{Y}} \widehat{S}_n(y; \boldsymbol{x}, u), \tag{3.2}$$

where $\widehat{S}_n(y; \boldsymbol{x}, u) = \widehat{w}_0(y; u) + \widehat{w}_1(\boldsymbol{x}, u) \widehat{w}_2^{-1}(u) \widehat{w}_3(y; u)$. The estimators $\widehat{w}_0(y; u)$, $\widehat{w}_1(\boldsymbol{x}, u)$, $\widehat{w}_2(u)$, and $\widehat{w}_3(y; u)$ are the respective estimates of $w_0(y; u)$,

$w_1(\boldsymbol{x}, u)$, $w_2(u)$, and $w_3(y; u)$. We now consider two cases based on the domain of the predictor u: Euclidean space and non-Euclidean space.

For the case where u lies in a Euclidean space, the expressions for $\widehat{w}_0(y; u)$, $\widehat{w}_1(\boldsymbol{x}, u)$, $\widehat{w}_2(u)$, and $\widehat{w}_3(y; u)$ are given as follows

$$\begin{aligned}
&\widehat{w}_0(y; u) = \frac{1}{n}\sum_{i=1}^{n} s_{in}(u) d^2(Y_i, y), \qquad \widehat{w}_1(\boldsymbol{x}, u) = \boldsymbol{x}^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{X}_i^{\mathrm{T}}, \\
&\widehat{w}_2(u) = \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)(\boldsymbol{X}_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}(u)\boldsymbol{X}_j)(\boldsymbol{X}_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}(u)\boldsymbol{X}_j)^{\mathrm{T}}, \qquad (3.3) \\
&\widehat{w}_3(y; u) = \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)(\boldsymbol{X}_i - \frac{1}{n}\sum_{j=1}^{n} s_{jn}(u)\boldsymbol{X}_j)(d^2(Y_i, y) - \frac{1}{n}\sum_{j=1}^{n} s_{jn}(u) d^2(Y_j, y)).
\end{aligned}$$

Here the weight function $s_{in}(u)$ may denote either the local constant weight $s_{in}^c(u)$ or the local linear weight $s_{in}^l(u)$ in (2.4). Corresponding to the varying-coefficient model with a Euclidean response variable, we have

$$\widehat{s}_{\oplus}(\boldsymbol{x}, u) = \arg\min_{y \in \mathcal{Y}} \sum_{j=1}^{n} c_j(\boldsymbol{x}, u) d^2(Y_j, y),$$

where $c_j(\boldsymbol{x}, u)$ is defined in (2.8).

For the case where u is a random object, the expressions for $\widehat{w}_0(y; u)$, $\widehat{w}_1(\boldsymbol{x}, u)$, $\widehat{w}_2(u)$, and $\widehat{w}_3(y; u)$ are the same as those in (3.3), except for the weight function $s_{in}(u)$, which takes the following form

$$s_{in}^r(u) = \frac{K_h(\delta(U_i, u))}{n^{-1}\sum_{j=1}^{n} K_h\left(\delta(U_j, u)\right)}, \qquad (3.4)$$

where $\delta(\cdot, \cdot)$ denotes the metric on the space $\mathcal{U}$.

Table 1 summarizes the weighting schemes used throughout the paper. For clarity and reproducibility, Algorithm 1 summarizes the estimation procedure for the proposed varying-coefficient Fréchet regression. The final minimization step in Algorithm 1 depends on the response space. Computational complexity and numerical optimization details are provided in Supplementary Material S1.1.

Table 1: Summary of weighting schemes and their formulations.

Scheme & Notation	Formula	Type of U
Local Constant $s_{in}^{c}(u)$	$\dfrac{K_h(U_i-u)}{n^{-1}\sum_{j=1}^{n}K_h(U_j-u)}$	Euclidean
Local Linear $s_{in}^{l}(u)$	$\dfrac{K_h(U_i-u)\left[\widehat{\mu}_2(u)-\widehat{\mu}_1(u)(U_i-u)\right]}{\widehat{\mu}_0(u)\widehat{\mu}_2(u)-\widehat{\mu}_1^2(u)}$	Euclidean
Non-Euclidean $s_{in}^{r}(u)$	$\dfrac{K_h(\delta(U_i,u))}{n^{-1}\sum_{j=1}^{n}K_h(\delta(U_j,u))}$	Non-Euclidean

An important case arises when the random objects lie in a Hilbert space $\mathcal{Y}$ equipped with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|_{\mathcal{Y}}$. Under mild assumptions, each of minimization problem (3.1) and (3.2) has an explicit solution in this setting, thanks to the linearity of the inner product and Riesz representation theorem. To facilitate a detailed exposition of this case, we introduce the following notation. For $p>1$, let $\mathcal{Y}^p$ denote the p-fold Cartesian product of $\mathcal{Y}$, equipped with inner product $\langle \boldsymbol{y},\boldsymbol{y}'\rangle_p = \sum_{i=1}^{p}\langle y_i,y_i'\rangle$, $\boldsymbol{y},\boldsymbol{y}'\in\mathcal{Y}^p$. Then $\mathcal{Y}^p$ is itself a Hilbert space. For a $p\times p$

Algorithm 1 Estimation Procedure for Varying-Coefficient Fréchet Regression

Input: Data $\{(\boldsymbol{X}_i, U_i, Y_i)\}_{i=1}^n$, target point $(\boldsymbol{x}, u)$, bandwidth h, kernel function $K_h(\cdot)$

Output: Estimator $\widehat{s}_\oplus(\boldsymbol{x}, u) = \arg\min_{y\in\mathcal{Y}} \sum_{i=1}^n c_i(\boldsymbol{x}, u) d^2(Y_i, y)$

1: Compute the weight function $s_{in}(u)$ for $i = 1, \ldots, n$. If U is Euclidean, $s_{in}(u)$ denotes either the local constant weight $s_{in}^c(u)$ or the local linear weight $s_{in}^l(u)$. If U is non-Euclidean, set $s_{in}(u) = s_{in}^r(u)$ (see Table 1 for the explicit forms).

2: Compute the local conditional expectation of $\boldsymbol{X}$ at u by
$$\widehat{\mu}_{\boldsymbol{X}|u} = n^{-1} \sum_{i=1}^n s_{in}(u) \boldsymbol{X}_i.$$

3: Compute the local conditional covariance matrix of $\boldsymbol{X}$ at u by
$$\widehat{\Sigma}_{\boldsymbol{X}|u} = n^{-1} \sum_{i=1}^n s_{in}(u)(\boldsymbol{X}_i - \widehat{\mu}_{\boldsymbol{X}|u})(\boldsymbol{X}_i - \widehat{\mu}_{\boldsymbol{X}|u})^{\mathrm{T}}.$$

4: Compute the varying-coefficient Fréchet regression weights by
$$c_i(\boldsymbol{x}, u) = n^{-1} s_{in}(u)(1 + (\boldsymbol{x} - \widehat{\mu}_{\boldsymbol{X}|u})^{\mathrm{T}} \widehat{\Sigma}_{\boldsymbol{X}|u}^{-1} (\boldsymbol{X}_i - \widehat{\mu}_{\boldsymbol{X}|u})).$$

5: Obtain the estimator $\widehat{s}_\oplus(\boldsymbol{x}, u) = \arg\min_{y\in\mathcal{Y}} \sum_{i=1}^n c_i(\boldsymbol{x}, u)\, d^2(Y_i, y)$.

6: **return** $\widehat{s}_\oplus(\boldsymbol{x}, u)$

matrix A, $\boldsymbol{x} \in \mathbb{R}^p$, $y \in \mathcal{Y}$ and $\boldsymbol{y} \in \mathcal{Y}^p$, we define $A\boldsymbol{y} \in \mathcal{Y}^p$ with elements $(A\boldsymbol{y})_i = \sum_{j=1}^p A_{ij} y_j$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^p x_i y_i \in \mathcal{Y}$ and $\boldsymbol{x}y \in \mathcal{Y}^p$ with elements $(\boldsymbol{x}y)_i = x_i y$.

Lemma 1 (Varying–coefficient Fréchet regression in Hilbert space)**.** *For a given $u \in \mathcal{U}$, set $\mu(u) := \mathbb{E}(\boldsymbol{X} \mid U = u) \in \mathbb{R}^p$, and assume $\mathbb{E}(\|Y\|_{\mathcal{Y}}^2 \mid U = u) < \infty$. Define $s_u(\boldsymbol{x}, \boldsymbol{X}) := 1 + (\boldsymbol{x} - \mu(u))^{\mathrm{T}} w_2^{-1}(u)(\boldsymbol{X} - \mu(u))$ and thus $s_\oplus(\boldsymbol{x}, u) = \arg\min_{y\in\mathcal{Y}} \mathbb{E}(s_u(\boldsymbol{x}, \boldsymbol{X}) \|Y - y\|_{\mathcal{Y}}^2 \mid U = u)$. Then there exist unique elements $\gamma_0 \in \mathcal{Y}$ and $\gamma_1 \in \mathcal{Y}^p$, such that for all $y \in \mathcal{Y}$ and $\boldsymbol{y} \in \mathcal{Y}^p$, $\mathbb{E}(\langle Y, y\rangle \mid U = u) = \langle \gamma_0(u), y\rangle$ and $\mathbb{E}(\langle (\boldsymbol{X} - \mu(u))Y, \boldsymbol{y}\rangle_p \mid U = u) = \langle \gamma_1(u), \boldsymbol{y}\rangle_p$. Define $\boldsymbol{\beta}(u) := w_2^{-1}(u)\gamma_1(u)$ and $\beta_0(u) := \gamma_0(u) -$*

$\mu(u)^{\mathrm{T}} w_2^{-1}(u)\gamma_1(u)$, the solution to (3.1) *is* $s_{\oplus}(\boldsymbol{x}, u) = \beta_0(u) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}(u)$. *In addition, define* $\widehat{\mu}(u) := \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{X}_i$, $\widehat{\gamma}_0(u) := \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)Y_i$, $\widehat{\gamma}_1(u) := \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)\left(\boldsymbol{X}_i - \widehat{\mu}(u)\right)Y_i$, $\widehat{\boldsymbol{\beta}}(u) := \widehat{w}_2^{-1}(u)\widehat{\gamma}_1(u)$ *and* $\widehat{\beta}_0(u) := \widehat{\gamma}_0(u) - \widehat{\mu}(u)^{\mathrm{T}}\widehat{w}_2^{-1}(u)\widehat{\gamma}_1(u)$. *The solution to* (3.2) *is* $\widehat{s}_{\oplus}(\boldsymbol{x}, u) = \widehat{\beta}_0(u) + x^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u)$.

For responses that are random objects in a Hilbert space, Lemma 1 provides explicit solutions to the minimization problems (3.1) and (3.1) that define the varying-coefficient Fréchet regression. In particular, when $\mathcal{Y}$ is Euclidean space, the varying-coefficient Fréchet regression model can simplify to the classical varying-coefficient model (2.1).

4. Theoretical properties

This section establishes the theoretical properties of the proposed varying-coefficient Fréchet regression estimator $\widehat{s}_{\oplus}(\boldsymbol{x}, u)$ for both Euclidean and non-Euclidean U. For the theoretical results below, we assume that $\boldsymbol{X}$ has bounded support and $\mathcal{Y}$ is a totally bounded metric space. To derive the convergence rate of the distance $d(s_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$, we first introduce the smoothed version of the varying-coefficient Fréchet regression $s_{\oplus}(\boldsymbol{x}, u)$ as

$$\widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right) = \underset{y \in \mathcal{Y}}{\operatorname{argmin}}\, \widetilde{S}_n\left(y; \boldsymbol{x}, u\right), \tag{4.1}$$

Statistica Sinica: Newly accepted Paper

where $\widetilde{S}_n\left(y;\boldsymbol{x},u\right)=\widetilde{w}_0\left(y;u\right)+\widetilde{w}_1\left(\boldsymbol{x},u\right)\widetilde{w}_2^{-1}\left(u\right)\widetilde{w}_3\left(y;u\right)$. The terms $\widetilde{w}_0(y;u)$, $\widetilde{w}_1(\boldsymbol{x},u)$, $\widetilde{w}_2(u)$, and $\widetilde{w}_3(y;u)$ are defined by the following expressions,

$$\widetilde{w}_0\left(y;u\right)=\mathbb{E}(\zeta_h\left(U,u\right)d^2\left(Y,y\right)),\qquad \widetilde{w}_1\left(\boldsymbol{x},u\right)=\boldsymbol{x}^{\mathrm{T}}-\mathbb{E}\left(\zeta_h\left(U,u\right)\boldsymbol{X}\right)^{\mathrm{T}},$$

$$\widetilde{w}_2\left(u\right)=\mathbb{E}(\zeta_h(U,u)(\boldsymbol{X}-\mathbb{E}(\zeta_h(U,u)\boldsymbol{X}))(\boldsymbol{X}-\mathbb{E}(\zeta_h(U,u)\boldsymbol{X}))^{\mathrm{T}}),$$

$$\widetilde{w}_3\left(y;u\right)=\mathbb{E}(\zeta_h(U,u)(\boldsymbol{X}-\mathbb{E}(\zeta_h(U,u)\boldsymbol{X}))(d^2(Y,y)-\mathbb{E}(\zeta_h(U,u)d^2(Y,y)))).$$

The form of weight function $\zeta_h(U,u)$ depends on the space of the predictor U. For Euclidean U, the weight function may denote either $\zeta_h^l(U,u)$ or $\zeta_h^c(U,u)$, given respectively by

$$\zeta_h^l(U,u)=K_h(U-u)\frac{\widetilde{\mu}_2(u)-\widetilde{\mu}_1(u)(U-u)}{\widetilde{\sigma}_0^2(u)},\quad \zeta_h^c(U,u)=\frac{K_h(U-u)}{\mathbb{E}\left(K_h(U-u)\right)},\tag{4.2}$$

where $\widetilde{\mu}_j(u)=\mathbb{E}\left(K_h(U-u)(U-u)^j\right)$ for $j=0,1,2$, and $\widetilde{\sigma}_0^2(u)=\widetilde{\mu}_0(u)\widetilde{\mu}_2(u)-\widetilde{\mu}_1^2(u)$. For non-Euclidean U, the weight function is denoted by $\zeta_h^r(U,u)$, which is defined as

$$\zeta_h^r(U,u)=\frac{K_h\left(\delta\left(U,u\right)\right)}{\mathbb{E}\left(K_h\left(\delta\left(U,u\right)\right)\right)}.\tag{4.3}$$

4.1 Predictor U in Euclidean space

In this subsection, we study the properties of the varying-coefficient Fréchet regression estimator with a Euclidean predictor U. We focus on points $\boldsymbol{x}\in\mathbb{R}^p$ and $u\in\mathbb{R}$ for which the marginal densities satisfy $f_{\boldsymbol{X}}(\boldsymbol{x})>0$ and

Statistica Sinica: Newly accepted Paper

$f_U(u) > 0$. The empirical estimator corresponding to the weight function $\zeta_h(U,u)$ in $\widetilde{s}_\oplus(\boldsymbol{x},u)$ is $s_{in}(u)$ in $\widehat{s}_\oplus(\boldsymbol{x},u)$. The weight function $\zeta_h(U,u)$ can be chosen as either $\zeta_h^l(U,u)$ or $\zeta_h^c(U,u)$, with the corresponding estimators denoted by $s_{in}^l(u)$ and $s_{in}^c(u)$, respectively. In this paper, our theoretical analysis focuses on the more complex case in which the weight function takes the form $\zeta_h^l(U,u)$.

With the notation established, we aim to obtain the convergence rate for the distance $d\left(s_\oplus\left(\boldsymbol{x},u\right),\widehat{s}_\oplus\left(\boldsymbol{x},u\right)\right)$. The analysis is based on a bias variance decomposition: we first derive the convergence rate of bias term $d(s_\oplus(\boldsymbol{x},u),\widetilde{s}_\oplus(\boldsymbol{x},u))$, and then analyze the variance term $d(\widetilde{s}_\oplus(\boldsymbol{x},u),\widehat{s}_\oplus(\boldsymbol{x},u))$. The necessary assumptions (A1)–(A6) are stated below.

(A1) The object $s_\oplus(\boldsymbol{x},u)$ exists and is unique. There exist a positive integer n_0 such that for all $n \geq n_0$, $\widetilde{s}_\oplus(\boldsymbol{x},u)$ and $\widehat{s}_\oplus(\boldsymbol{x},u)$ exist and are unique, the latter almost surely. Additionally, for any $\epsilon > 0$,

$$\inf_{dt(y,s_\oplus(\boldsymbol{x},u))>\epsilon} \left\{ S_\oplus(y;\boldsymbol{x},u) - S_\oplus\left(s_\oplus(\boldsymbol{x},u);\boldsymbol{x},u\right) \right\} > 0,$$

$$\text{and} \quad \liminf_{n\to\infty} \inf_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))>\epsilon} \left\{ \widetilde{S}_n(y;\boldsymbol{x},u) - \widetilde{S}_n\left(\widetilde{s}_\oplus(\boldsymbol{x},u);\boldsymbol{x},u\right) \right\} > 0.$$

(A2) Let $B_\mathcal{Y}\left(s_\oplus(\boldsymbol{x},u),\delta\right) \subseteq \mathcal{Y}$ be the ball of radius δ centered at $s_\oplus(\boldsymbol{x},u)$ and $N(\epsilon,\ B_\mathcal{Y}(s_\oplus(\boldsymbol{x},u),\delta),d)$ be its ϵ-covering number under the metric d. Then $\int_0^1 (1+\log N\left(\delta\epsilon, B_\mathcal{Y}\left(s_\oplus(\boldsymbol{x},u),\delta\right),d\right))^{1/2} d\epsilon = O(1)$ as $\delta \to 0^+$.

Statistica Sinica: Newly accepted Paper

(A3) The kernel $K(\cdot)$ is a probability density function, symmetric around zero. Furthermore, defining $K_{kj} = \int_{\mathbb{R}} K^k(u)u^j du$, K_{12}, K_{14}, K_{22}, K_{24} and K_{26} are both finite.

(A4) The densities $f_U(\cdot)$, $f_{U|Y=y}(\cdot)$, $f_{U|X_i=x_i,X_j=x_j}(\cdot)$ and $f_{U|X_i=x_i,Y=y}(\cdot)$ exist and are twice continuously differentiable. The $\sup_{y,u}|f''_{U|Y=y}(u)|$ and $\sup_{x_i,u}|f''_{U|X_i=x_i}(u)|$ are finite. For any open set $A \subseteq \mathcal{Y}$, $B \subseteq \mathbb{R}$, $C \subseteq \mathbb{R}^2$, and $D \subseteq \mathcal{Y} \times \mathbb{R}$, the integrals $\int_A d\mathbb{P}_{Y|U=u}$, $\int_B d\mathbb{P}_{X_i|U=u}$, $\int_C d\mathbb{P}_{X_i,X_j|U=u}$ and $\int_D d\mathbb{P}_{Y,X_i|U=u}$ are continuous as a function of u, with $i, j = 1, 2, \ldots, p$.

(A5) There exist $\eta_1 > 0, C_1 > 0$, and $\beta_1 > 1$ such that if $d(y, s_\oplus(\boldsymbol{x}, u)) < \eta_1$, $S_\oplus(y; \boldsymbol{x}, u) - S_\oplus\left(s_\oplus(\boldsymbol{x}, u); \boldsymbol{x}, u\right) \geq C_1 d\left(y, s_\oplus(\boldsymbol{x}, u)\right)^{\beta_1}$.

(A6) There exist $\eta_2 > 0, C_2 > 0$, and $\beta_2 > 1$ such that if $d\left(y, \widetilde{s}_\oplus(\boldsymbol{x}, u)\right) < \eta_2$, $\liminf_{n\to\infty}\left\{\widetilde{S}_n(y; \boldsymbol{x}, u) - \widetilde{S}_n\left(\widetilde{s}_\oplus(\boldsymbol{x}, u); \boldsymbol{x}, u\right)\right\} \geq C_2 d\left(y, \widetilde{s}_\oplus(\boldsymbol{x}, u)\right)^{\beta_2}$.

Assumptions (A1)–(A4) provide conditions for consistency and convergence-rate analysis of M-estimators. Assumptions (A5) and (A6) impose local curvature conditions on the population and smoothed Fréchet objectives, with β_1 and β_2 quantifying local growth near the minimizers. For Euclidean responses, $S_\oplus(y; \boldsymbol{x}, u) - S_\oplus(s_\oplus; \boldsymbol{x}, u) = |y - s_\oplus|^2$ and $\widetilde{S}_n(y; \boldsymbol{x}, u) - \widetilde{S}_n(\widetilde{s}_\oplus; \boldsymbol{x}, u) = |y - \widetilde{s}_\oplus|^2$, so (A5) and (A6) hold with $\beta_1 = \beta_2 = 2$, $C_1 = C_2 = 1$, and $\eta_1, \eta_2 > 0$ arbitrary. Detailed discussion and verifications are provided in Supplementary Material S1.7.

Statistica Sinica: Newly accepted Paper

Theorem 1. *Under Assumptions (A1)–(A5), the bias component satisfies*

$$d\left(s_{\oplus}\left(\boldsymbol{x},u\right),\widetilde{s}_{\oplus}\left(\boldsymbol{x},u\right)\right)=O(h^{2/(\beta_1-1)}),$$

as the bandwidth $h\to 0$, where the β_1 is specified in Assumption (A5).

This result quantifies the convergence rate of the bias, which depends on the local geometric structure of the metric space around the minimizer through β_1. The parameter β_1 characterizes the order of smoothness of $S_{\oplus}(y;\boldsymbol{x},u)$ and thereby determines the bias decay rate $O(h^{2/(\beta_1-1)})$. In the Euclidean case, $\beta_1=2$ corresponds to the standard quadratic curvature condition, yielding the familiar bias rate of order $O(h^2)$ in classical nonparametric regression, which is known to be optimal.

Theorem 2. *Under Assumptions (A1)–(A3), and (A6), if $h\to 0$ and $nh\to\infty$ as $n\to\infty$, the stochastic component exhibits the probabilistic convergence rate*

$$d\left(\widetilde{s}_{\oplus}\left(\boldsymbol{x},u\right),\widehat{s}_{\oplus}\left(\boldsymbol{x},u\right)\right)=O_p\big((nh)^{-1/(2(\beta_2-1))}\big),$$

where β_2 is specified in Assumption (A6).

Theorem 2 establishes the rate of variance term $d(\widetilde{s}_{\oplus}(\boldsymbol{x},u),\widehat{s}_{\oplus}(\boldsymbol{x},u))$. The rate depends on β_2, which reflects the local curvature of the $\widetilde{S}_{\oplus}(y;\boldsymbol{x},u)$ around its minimizer. This result is consistent with the rate established for

Statistica Sinica: Newly accepted Paper

local Fréchet regression in Petersen and Müller (2019). In the Euclidean case, setting $\beta_2 = 2$ recovers the optimal variance rate $O_p((nh)^{-1/2})$.

Corollary 1. *Under the assumptions of Theorem 1 and Theorem 2, the distance between $s_\oplus(\boldsymbol{x}, u)$ and $\widehat{s}_\oplus(\boldsymbol{x}, u)$ satisfies the following rate*

$$d\left(s_\oplus(\boldsymbol{x}, u), \widehat{s}_\oplus(\boldsymbol{x}, u)\right) = O_p\left(h^{2/(\beta_1-1)} + (nh)^{-1/(2(\beta_2-1))}\right).$$

In general, the rate of $d(s_\oplus(\boldsymbol{x}, u), \widehat{s}_\oplus(\boldsymbol{x}, u))$ is determined by the local geometry near the minimum as quantified in (A5) and (A6). This result presents the bias-variance trade-off analogous to that in classical nonparametric regression, and the optimal h is of order $n^{-\frac{\beta_1-1}{4\beta_2+\beta_1-5}}$. Additionally, as $h \to 0$ and $nh \to \infty$, the result coincides with the case $\mathcal{Y} = \mathbb{R}$ under the Euclidean distance. Specifically, Corollary 1 yields $d(s_\oplus(\boldsymbol{x}, u), \widehat{s}_\oplus(\boldsymbol{x}, u)) = O_p(h^2 + (nh)^{-1/2})$, recovering the rate in Cai et al. (2000) for local varying-coefficient models with Euclidean responses.

In addition to these convergence rates, we further develop asymptotic distribution results when the response space $\mathcal{Y}$ is a Hilbert space. The conditional covariance operator of the error term ϵ given $(U, \boldsymbol{X})$, denoted by $\Sigma_\epsilon(U, \boldsymbol{X}) : \mathcal{Y} \to \mathcal{Y}$, is defined by $\Sigma_\epsilon(U, \boldsymbol{X})y = \mathbb{E}[\langle y, \epsilon\rangle_\mathcal{Y}\epsilon \mid U, \boldsymbol{X}]$ for all $y \in \mathcal{Y}$. We first establish the asymptotic distribution of the local linear estimators of the coefficient functions.

Statistica Sinica: Newly accepted Paper

Theorem 3. *Under Assumptions (A3)–(A4), if* $nh^5 \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *we have*

$$\sqrt{nh}\left[(\widehat{\beta}_0(u), \widehat{\boldsymbol{\beta}}^{\mathrm{T}}(u))^{\mathrm{T}} - (\beta_0(u), \boldsymbol{\beta}^{\mathrm{T}}(u))^{\mathrm{T}}\right] \xrightarrow{w} \mathcal{G},$$

where "$\xrightarrow{w}$*" denotes weak convergence and* $\mathcal{G}$ *is a zero-mean Gaussian random object in* $\mathcal{Y}^{p+1}$ *with covariance operator* $\Sigma_{\mathcal{G}}(u)$ *given by*

$$\frac{K_{20}}{f_U(u)} \left(\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \mid U = u\right]\right)^{-1} \mathbb{E}\left[(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}) \otimes \Sigma_\epsilon(U, \boldsymbol{X}) \mid U = u\right] \left(\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \mid U = u\right]\right)^{-1}.$$

Here, $\boldsymbol{Z} = (1, \boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \otimes \Sigma_\epsilon(U, \boldsymbol{X})$ *denotes a block operator matrix on* $\mathcal{Y}^{p+1}$ *with* (i, j)*th block* $Z_i Z_j \Sigma_\epsilon(U, \boldsymbol{X})$, *where* Z_i *is the* i*th component of* $\boldsymbol{Z}$.

We next establish the asymptotic distribution of the stochastic process

$$\{\mathcal{M}_n(\boldsymbol{x}) = \sqrt{nh}(\widehat{s}_\oplus(\boldsymbol{x}, u) - s_\oplus(\boldsymbol{x}, u)) : \boldsymbol{x} \in \mathcal{X}\}$$

for a fixed $u \in \mathcal{U}$ in the space $\ell_{\mathcal{Y}}^{\infty}(\mathcal{X})$ of bounded functions from $\mathcal{X}$ to $\mathcal{Y}$, endowed with the sup-norm $\|f\|_{\mathcal{X}} = \sup_{\boldsymbol{x}\in\mathcal{X}} \|f(\boldsymbol{x})\|_{\mathcal{Y}}$.

Corollary 2. *Under the assumptions of Theorem 3,*

$$\mathcal{M}_n \xrightarrow{w} \mathcal{M} \quad \text{in } \ell_{\mathcal{Y}}^{\infty}(\mathcal{X}),$$

where $\mathcal{M}$ *is a zero-mean Gaussian process. For any* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ *and any* $a, b \in \mathcal{Y}$, *the covariance operator of* $\mathcal{M}$ *is characterized by*

$$\mathrm{Cov}\big(\langle a, \mathcal{M}(\boldsymbol{x})\rangle_{\mathcal{Y}}, \langle b, \mathcal{M}(\boldsymbol{y})\rangle_{\mathcal{Y}}\big) = \langle (a, \boldsymbol{x}^{\mathrm{T}} a),\ \Sigma_{\mathcal{G}}(u)\ (b, \boldsymbol{y}^{\mathrm{T}} b)\rangle_{\mathcal{Y}^{p+1}},$$

Statistica Sinica: Newly accepted Paper

where $(a, \boldsymbol{x}^{\mathrm{T}}a) = (a, x_1a, \ldots, x_pa)^{\mathrm{T}} \in \mathcal{Y}^{p+1}$. *Additionally,* $\sup_{x\in\mathcal{X}} \|\widehat{s}_{\oplus}(\boldsymbol{x}, u) - s_{\oplus}(\boldsymbol{x}, u)\|_{\mathcal{Y}} = O_p((nh)^{-1/2})$.

4.2 Predictor U in non-Euclidean space

This subsection studies the properties of the varying-coefficient Fréchet regression estimator with a non-Euclidean predictor U. The weight function $\zeta_h(U, u)$ is given by $\zeta_h^r(U, u)$ in (4.3). The corresponding empirical weight in $\widehat{s}_{\oplus}(\boldsymbol{x}, u)$ is $s_{in}^r(u) = K_h(\delta(U_i, u))/(n^{-1}\sum_{j=1}^n K_h(\delta(U_j, u)))$ as in (3.4).

Because the probability of non-Euclidean U depends on the metric space $(\mathcal{U}, \delta)$, we define the small ball probability of random object $U \in \mathcal{U}$ as $\varphi_{\mathcal{U},u}(h) = \mathbb{P}(U \in B_{\mathcal{U}}(u, h))$, where $B_{\mathcal{U}}(u, h) = \{u' \in \mathcal{U} : \delta(u', u) \leq h\}$. With the small ball probability definitions, we require the following assumptions to handle the predictor U in the metric space.

(A7) For any $\epsilon > 0$, $\mathbb{P}(U \in B_{\mathcal{U}}(u, \epsilon)) = \varphi_{\mathcal{U},u}(\epsilon) > 0$.

(A8) As $n \to \infty$, we have $h \to 0$, $\log n/(n\varphi_{\mathcal{U},u}(h)) \to 0$, and $nh \to \infty$.

(A9) There exist constants $0 < c_1 \leq c_2 < \infty$ and C, such that $c_1\mathbb{1}_{[0,1]}(\cdot) \leq K(\cdot) \leq c_2\mathbb{1}_{[0,1]}(\cdot)$ and $\int K_h(\delta(u', u))\delta(u', u)d\nu(u') \leq Ch\int K_h(\delta(u', u))d\nu(u')$, where $\mathbb{1}_{[0,1]}(\cdot)$ denotes the indicator function.

(A10) The marginal density $f_U(u)$ and the conditional density $f_{U|Y=y}(u)$, $f_{U|X_i=x_i}(u)$, $f_{U|X_i=x_i,X_j=x_j}(u)$, $f_{U|X_i=x_i,Y=y}(u)$ for $i, j = 1, 2\ldots, p$, exist

Statistica Sinica: Newly accepted Paper

(with respect to a reference measure ν) and are Lipschitz continuous. That is, there exists a constant $C > 0$ such that for all $u' \in B_{\mathcal{U}}(u, h)$ and any conditional density $f(\cdot)$ in the above, $|f(u') - f(u)| \leq C\delta(u', u)$.

Assumptions (A7)–(A10) ensure sufficient local probability mass around u, the bandwidth conditions for asymptotic bias and variance vanishing, and the required smoothness of the kernel and conditional densities. Further discussion is provided in Supplementary Material S1.7.

Next, we present the main theoretical results for the estimator when U is treated as a random object. Similar to the case where u is a Euclidean variable, we first analyze the bias term $d(s_{\oplus}(\boldsymbol{x}, u), \widetilde{s}_{\oplus}(\boldsymbol{x}, u))$, then derive the rate of variance term $d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$. A corollary combines these results to establish the overall convergence rate of $d(s_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$.

Theorem 4. *If Assumptions (A1), (A2), (A5) hold, and (A8)–(A10), then*

$$d\left(s_{\oplus}\left(\boldsymbol{x}, u\right), \widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right)\right) = O\big(h^{\frac{1}{\beta_1 - 1}}\big),$$

where β_1 is specified in Assumption (A5).

This result establishes the convergence rate of the distance bias term $d(s_{\oplus}(\boldsymbol{x}, u), \widetilde{s}_{\oplus}(\boldsymbol{x}, u))$. The rate is determined by the Lipschitz continuity in (A10) and the local geometry of the metric space through β_1 in (A5). When $\mathcal{Y} = \mathbb{R}$, we have $\beta_1 = 2$, and the convergence rate becomes $O(h)$.

Statistica Sinica: Newly accepted Paper

One might naturally ask why the result here is $O(h)$, given that local constant smoothing with a symmetric kernel is well known to achieve a bias of order $O(h^2)$ in Euclidean settings. The key lies in the specific form of the kernel function used in our estimator. Our smoothing term is constructed as $K_h(\delta(u', u)) = K_h(|u' - u|)$, rather than the classical form $K_h(u' - u)$. This difference is crucial: the integration of the first-order term $\int K_h(|u' - u|)\,|u' - u|\,du'$ does not vanish. Therefore, even under a Euclidean response space and apply a symmetric kernel, the first-order bias term in the Taylor expansion persists and dominates the convergence rate, leading to the $O(h)$ result. Therefore, the convergence rate established in Theorem 4 is reasonable for a general metric space $\mathcal{Y}$.

Theorem 5. *If Assumptions (A1), (A2), (A6) and (A7)–(A9) hold, then*

$$d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u)) = O_p\big((n\varphi_{\mathcal{U},u}(h))^{\frac{-1}{2(\beta_2-1)}}\big),$$

and

$$d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u)) = O_{a.s.}\big((n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{-1}{2(\beta_2-1)}}\big),$$

where β_2 defined in Assumption (A6). Here, the almost sure convergence rate means that there exists a constant $M > 0$, such that

$$\mathbb{P}\Big(\limsup_{n\to\infty} d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))(n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{1}{2(\beta_2-1)}}) \leq M\Big) = 1.$$

This theorem establishes the convergence rates of the variance term $d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$. These rates depend on both the local curvature pa-

Statistica Sinica: Newly accepted Paper

rameter β_2 in Assumption (A6) and the small ball probability $\varphi_{\mathcal{U},u}(h)$. It is worth noting that if U lies in Euclidean space equipped with the Euclidean distance, then the small ball probability $\varphi_{\mathcal{U},u}(h) = \int_{u-h}^{u+h} dF_{\mathcal{U}}$ is of order h. Consequently, the stochastic convergence rate in Theorem 5 reduces to $O_p\left((nh)^{-1/2}\right)$, which aligns with the rate established in Theorem 2. Furthermore, the almost sure convergence rate is nearly optimal, differing from the convergence rate in probability only by a logarithmic factor.

Corollary 3. *If the assumptions of Theorems 4 and 5 hold, then*

$$d(s_{\oplus}(\boldsymbol{x},u),\widehat{s}_{\oplus}(\boldsymbol{x},u)) = O_p\big(h^{\frac{1}{\beta_1-1}} + (n\varphi_{\mathcal{U},u}(h))^{\frac{-1}{2(\beta_2-1)}}\big),$$

$$\textit{and} \qquad d(s_{\oplus}(\boldsymbol{x},u),\widehat{s}_{\oplus}(\boldsymbol{x},u)) = O_{a.s.}\big(h^{\frac{1}{\beta_1-1}} + (n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{-1}{2(\beta_2-1)}}\big).$$

This corollary combines the bias term in Theorem 4 and the variance term in Theorem 5. The convergence rate of the $d(s_{\oplus}(\boldsymbol{x},u),\widehat{s}_{\oplus}(\boldsymbol{x},u))$ depends on the Lipschitz continuity of the conditional densities and the marginal density in Assumption (A10), and local geometry properties as quantified in (A5) and (A6) through β_1, β_2. Moreover, the variance term is governed by the effective local sample size $n\varphi_{\mathcal{U},u}(h)$. Thus, smaller $\varphi_{\mathcal{U},u}(h)$ generally requires a larger bandwidth. In our simulations, SPD matrix predictors exhibit smaller small ball probabilities and consequently larger selected bandwidths than Euclidean and distribution predictors. Empirical

Statistica Sinica: Newly accepted Paper

illustrations are provided in Supplementary Material S1.2.

5. Simulation studies

In the simulation experiments, two types of responses are considered to demonstrate the performance of the varying-coefficient Fréchet regression model. The one is probability distributions and another one is symmetric positive definitive matrices. For every type of response, we present three examples where the predictor U be chosen as the scalar, the density and the symmetric positive definite matrix. These examples cover both Euclidean predictor and non-Euclidean predictor.

We denote the proposed varying-coefficient Fréchet regression as VFR. For comparison purposes, we also consider the following three competitors: (1) global Fréchet regression (GFR) proposed by Petersen and Müller (2019); (2) the local Fréchet regression (LFR) introduced by Petersen and Müller (2019); (3) the partially-global Fréchet regression (PFR) proposed by Tucker and Wu (2025). The computations for the GFR and the LFR can be carried out by R package `frechet` (Chen et al., 2020). For the competing methods, U is included in GFR and LFR only when it is Euclidean. When U is non-Euclidean, GFR and LFR are not applicable. PFR uses $\boldsymbol{X}$ as the global predictor and U as the local predictor. For bandwidth selection,

we use 10-fold cross-validation over the grid $h_j = c_j \widehat{\sigma}_U n^{-\gamma}$, $j = 1, \ldots, K$, where $\widehat{\sigma}_U = \{n^{-1} \sum_{i=1}^{n} \delta^2(U_i, \widehat{u}_\oplus)\}^{1/2}$, $\widehat{u}_\oplus$ is the sample Fréchet mean of U, $K = 20$, and $\{c_j\}_{j=1}^{K}$ is an equally spaced grid on $[0.10, 3.00]$. If the cross-validation minimum occurs at the boundary, the search range is enlarged. Grid sensitivity analysis and the empirical scaling of selected bandwidths with n are reported in Supplementary Material S1.3.

Based on the independent testing data $\{\widetilde{\mathbf{X}}_i, \widetilde{U}_i, \widetilde{Y}_i\}_{i=1}^{\widetilde{n}}$ and a specific metric $d(\cdot, \cdot)$, we calculate a generalized mean squared error (GMSE) defined as $\widetilde{n}^{-1} \sum_{i=1}^{\widetilde{n}} d^2(\widehat{s}_\oplus(\widetilde{\mathbf{X}}_i, \widetilde{U}_i), \widetilde{Y}_i)$, where $\widehat{s}_\oplus(\cdot, \cdot)$ is an estimated Fréchet regression function based on the training data using GFR, LFR, PFR and VFR, respectively. These criteria are used to evaluate the prediction performance of each method. In the following simulations, all the results are obtained by computing the averaged values of GMSE with 100 replications. In addition, training samples of size $n \in \{50, 100, 200\}$ and testing samples of size $\widetilde{n} = 1000$ were used for all examples. The raw predictors were generated in two steps: (1) $(S_1, S_2, S_3, S_4)^{\mathrm{T}}$ is multivariate Gaussian with $\mathbb{E}(S_j) = 0$ and $\mathrm{Cov}(S_j, S_{j'}) = 0.5$ for $i \neq j$; (2) Set $T = \Phi(S_4)$, where $\Phi(\cdot)$ is the standard normal distribution function, so that $T \sim Unif[0, 1]$.

Statistica Sinica: Newly accepted Paper

5.1 Probability distribution as response

Let Ω_1 be the set of probability distributions. The 2-Wasserstein metric distance between two distributions with cumulative distribution functions $H(\cdot)$ and $G(\cdot)$ is defined as $d_W(H,G) = \left(\int_0^1 \left(H^{-1}(t) - G^{-1}(t)\right)^2 dt \right)^{1/2}$. We denote (Ω_1, d_W) as the metric space of probability distributions equipped with the Wasserstein distance. Consider $\mathcal{Y} \subseteq \Omega_1$, the responses Y represent the distribution function with the corresponding quantile function $Q(Y)$. Following Petersen and Müller (2019), for the sake of notational simplicity, we also denote the quantile function corresponding to Y as Y. For computational tractability, we approximate the distance $d_W^2(H,G)$ using the discrete form $m^{-1}\sum_{i=1}^{m}\left(H^{-1}(t_i) - G^{-1}(t_i)\right)^2$, where $\{t_1, \ldots, t_m\}$ denotes an equally grid on the interval $[0,1]$ with $m = 20$ as in Tucker et al. (2023).

Example 1. (U is a scalar) In this example, we consider $\mathcal{U} \subseteq \mathbb{R}$. Set $X_1 = S_1, X_2 = S_2, X_3 = S_3$, and $U = T$. The Fréchet regression function is $m_{\oplus}(\boldsymbol{x}, u) = \mathbb{E}(Y(\cdot) \mid \boldsymbol{X} = \boldsymbol{x}, U = u) = \mu_0 + \gamma_1 u x_1 + \gamma_2 u^2 x_2 + (\sigma_0 + \gamma_3 \sin(\pi u) x_3)\Phi^{-1}(\cdot)$. Conditional on $\boldsymbol{X}$ and U, we generate $Y(\cdot) = \mu + \sigma\Phi^{-1}(\cdot)$, where $\mu \mid (\boldsymbol{X}, U) \sim N(\mu_0 + \gamma_1 U X_1 + \gamma_2 U^2 X_2, \nu_1)$ and $\sigma \mid U \sim \text{Gamma}(\sigma_U^2/\nu_2, \nu_2/\sigma_U)$, with $\sigma_U = \sigma_0 + \gamma_3 \sin(\pi U) X_3$. Set $\mu_0 = 1, \gamma_1 = 2, \sigma_0 = 2, \nu_1 = 0.1$, $\nu_2 = 2$ and $(\gamma_2, \gamma_3) \in \{(1,1), (3,3)\}$ for two different Fréchet regression functions $m_{\oplus}(\boldsymbol{x}, u)$.

Statistica Sinica: Newly accepted Paper

Example 2. (U is a density) In this example, we consider $\mathcal{U} \subseteq \Omega_1$. Set $X_1 = S_1, X_2 = S_2, X_3 = S_3$. Further, we set $\mathbb{E}\left(U\left(\cdot\right) \mid T = t\right) = \mu_{0u} + \left(\sigma_{0u} + \gamma_u t\right)\Phi^{-1}\left(\cdot\right)$. Conditional on T, the predictor $U(\cdot)$ is generated by adding noise as follows: $U(\cdot) = \mu_{0u} + \sigma_u \Phi^{-1}(\cdot)$, where $\sigma_u \mid T \sim Gamma((\sigma_{0u} + \gamma_u T)^2/\nu_u, \nu_u/(\sigma_{0u} + \gamma_u T))$is independently sampled. The Fréchet regression function is given by $m_{\oplus}\left(\boldsymbol{x}, t\right) = \mathbb{E}\left(Y\left(\cdot\right) \mid \boldsymbol{X} = \boldsymbol{x}, T = t\right) = \mu_0 + \gamma_1 t x_1 + \gamma_2 t^2 x_2 + \left(\sigma_0 + \gamma_3 \sin\left(\pi t\right) x_3\right)\Phi^{-1}\left(\cdot\right)$. The response $Y(\cdot)$ is generated as in Example 1, with U replaced by T. Set $\mu_{0u} = 0, \nu_u = 1, \gamma_u = 3, \sigma_{0u} = 2$.

Example 3. (U is a symmetric positive definite matrix) Consider $\mathcal{U} \subseteq \Omega_2$, which is the set of symmetric positive definite matrices. Let $X_1 = S_1, X_2 = S_2$ and $X_3 = S_3$. Further, set $\mathbb{E}(U \mid T = t) = \mathbb{E}(A \mid T = t)^{\mathrm{T}}\mathbb{E}(A \mid T = t)$, where $\mathbb{E}(A \mid T = t) = (\mu_{0u} + \beta_u t + \sigma_{0u} + \gamma_u t)I + (\sigma_{0u} + \gamma_u t)V$, I denote an $M \times M$ identity matrix and $V = (V_{i,j})$ denote an $M \times M$ matrix with $V_{i,j} = I_{(i<j)}$. Conditional on T, the predictor U is generated by adding noise as follows: $U = A^{\mathrm{T}}(N^{-1}\sum_{i=1}^{N} Z_i Z_i^{\mathrm{T}})A$, where $(Z_1, \ldots, Z_N)$ are independently generated from $N_M(0, I_M)$, $A = \left(\mu_u + \sigma_u\right) I + \sigma_u V$, $\mu_u = \mu_{0u} + \beta_u T$, and $\sigma_u \mid T$ is independently sampled from $Gamma((\sigma_{0u} + \gamma_u T)^2/\nu_u, \nu_u/(\sigma_{0u} + \gamma_u T))$. The Fréchet regression function and response generation are the same as in Example 2. Set $M = 2, N = 10, \beta_u = 1, \mu_{0u} = 2$.

In the above examples, the parameter γ_1 governs the linear component

Statistica Sinica: Newly accepted Paper

of $m_{\oplus}(\boldsymbol{x}, u)$, whereas (γ_2, γ_3) determine the nonlinear component. Accordingly, we consider the setting $(\gamma_2, \gamma_3) \in \{(1,1),(3,3)\}$ to evaluate each method's prediction performance under different degrees of nonlinearity. As shown in Table 2, for Example 1, the proposed VFR is significantly superior to GFR, LFR, and PFR when U is scalar, especially under higher nonlinearity $(\gamma_2, \gamma_3) = (3,3)$. The VFR also performs well in both Example 2 and Example 3 when the predictor U is non-Euclidean. We also observe that the GMSE of VFR decreases as the training sample size n increases.

5.2 Symmetric positive definite matrix as response

Let Ω_2 be the set of symmetric positive definite matrices. The Cholesky decomposition distance between two matrices P_1 and P_2 is defined as $d_C(P_1, P_2) = \left(\text{trace}((P_1^{1/2} - P_2^{1/2})^{\mathrm{T}}(P_1^{1/2} - P_2^{1/2}))\right)^{1/2}$. We denote (Ω_2, d_C) as the metric space of symmetric positive definite matrices equipped with the Cholesky decomposition distance. Consider $\mathcal{Y} \subseteq \Omega_2$. Let I denotes an $M \times M$ identity matrix and $V = (V_{i,j})$ denotes an $M \times M$ matrix where $V_{i,j} = I_{(i<j)}$.

Example 4. (U is a scalar) Consider $\mathcal{U} \subseteq \mathbb{R}$. We set $X_1 = S_1, X_2 = S_2, X_3 = S_3$, and $U = T$. The Fréchet regression function is given by $m_{\oplus}(\boldsymbol{x}, u) = \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}, U = u) = \mathbb{E}(B \mid \boldsymbol{X} = \boldsymbol{x}, U = u)^{\mathrm{T}}\mathbb{E}(B \mid \boldsymbol{X} = \boldsymbol{x}, U = u)$, where $\mathbb{E}(B \mid \boldsymbol{X} = \boldsymbol{x}, U = u) = (\mu_0 + \gamma_1 u x_1 + \gamma_2 u^2 x_2 +$

Statistica Sinica: Newly accepted Paper

$(\sigma_0 + \gamma_3 \sin(\pi u)x_3))I + (\sigma_0 + \gamma_3 \sin(\pi u)x_3)V$. Given $\boldsymbol{X}$ and U, generate $Y = B^{\mathrm{T}}(N^{-1}\sum_{i=1}^{N} Z_i Z_i^{\mathrm{T}})B$, where $B = (\mu + \sigma)I + \sigma V$, $\mu \mid (\boldsymbol{X}, U) \sim N(\mu_0 + \gamma_1 U X_1 + \gamma_2 U^2 X_2, \nu_1)$, and $\sigma \mid (\boldsymbol{X}, U) \sim \text{Gamma}(\sigma_U^2/\nu_2, \nu_2/\sigma_U)$.

Example 5. (U is a density) Here we consider $\mathcal{U} \subseteq \Omega_1$. Set $X_1 = S_1, X_2 = S_2$ and $X_3 = S_3$. Further, $\mathbb{E}\left(U\left(\cdot\right) \mid T = t\right) = \mu_{0u} + \left(\sigma_{0u} + \gamma_u t\right)\Phi^{-1}(\cdot)$. Conditional on T, the predictor $U(\cdot)$ is generated by adding noise as $U(\cdot) = \mu_{0u} + \sigma_u \Phi^{-1}(\cdot)$ with $\sigma_u \mid T \sim Gamma((\sigma_{0u} + \gamma_u T)^2/\nu_u, \nu_u/(\sigma_{0u} + \gamma_u T))$ being independently sampled. The Fréchet regression function and response generation are as in Example 4, with U replaced by T.

Example 6. (U is a symmetric positive definite matrix) We consider $\mathcal{U} \subseteq \Omega_2$. Set $X_1 = S_1, X_2 = S_2$ and $X_3 = S_3$. Further, we set $\mathbb{E}(U \mid T = t) = \mathbb{E}(A \mid T = t)^{\mathrm{T}}\,\mathbb{E}(A \mid T = t)$, where $\mathbb{E}(A \mid T = t) = (\mu_{0u} + \beta_u t + \sigma_{0u} + \gamma_u t)I_{M_u} + (\sigma_{0u} + \gamma_u t)U_{M_u}$. Conditional on T, the predictor U is generated as $U = A^{\mathrm{T}}(N^{-1}\sum_{i=1}^{N} Z_i Z_i^{\mathrm{T}})A$, where $(Z_1, \ldots, Z_N)$ are generated from $N_M(0, I_M)$, $A = (\mu_u + \sigma_u)I_{M_u} + \sigma_u U_{M_u}$, $\mu_u \mid T \sim N(\mu_{0u} + \beta_u T, \nu_{u1})$, and $\sigma_u \mid T \sim Gamma((\sigma_{0u} + \gamma_u T)^2/\nu_{u2}, \nu_{u2}/(\sigma_{0u} + \gamma_u T))$. The Fréchet regression function and response generation are the same as in Example 5.

Response space specific implementation details, stopping criteria, and runtime summaries are given in Supplementary Material S1.6. The results for Example 4–6 in Table 2 correspond to symmetric positive definite matrix

responses under different nonlinearity settings. The superiority of proposed VFR over the three competing models becomes more pronounced under higher nonlinearity case $(\gamma_2, \gamma_3) = (3, 3)$. Moreover, VFR outperforms other competitive Fréchet regression methods when U is the Euclidean predictor in Example 4, and achieves lower GMSE than PFR when the predictor U comes from the non-Euclidean space in Example 5 and 6.

Additional comparisons with random-forest-weighted local Fréchet regression (RFWLFR; Qiu et al. (2024)) and deep Fréchet regression (DFR; Iao et al. (2025)) are given in Supplementary Material S1.4. In these comparisons, the proposed VFR yields smaller prediction errors while retaining interpretable varying-coefficient structure. We also report small sample simulations with $n \in \{30, 35, 40\}$ in Supplementary Material S1.5. These sample sizes are comparable to a real data example in Section 6 with $n = 39$, and the results are broadly consistent with the main simulation findings.

6. Real data analysis

In this section, we apply the proposed varying-coefficient Fréchet regression and other Fréchet regression methods to a human mortality dataset. The goal is to model age-at-death distributions in 2013 as random objects using country-specific covariates. Compared with traditional summary measures

Table 2: The averaged GMSE of various methods and the associated standard errors (in parenthesis).

		$(\gamma_2, \gamma_3) = (1, 1)$				$(\gamma_2, \gamma_3) = (3, 3)$			
		GFR	LFR	PFR	VFR	GFR	LFR	PFR	VFR
Example 1	$n = 50$	2.551 (0.019)	2.250 (0.031)	2.480 (0.023)	**1.997** (0.029)	4.207 (0.071)	3.001 (0.065)	3.794 (0.071)	**2.345** (0.071)
	$n = 100$	2.430 (0.022)	2.002 (0.019)	2.296 (0.023)	**1.845** (0.023)	3.968 (0.052)	2.604 (0.088)	3.441 (0.052)	**2.050** (0.056)
	$n = 200$	2.354 (0.012)	1.841 (0.019)	2.181 (0.012)	**1.723** (0.013)	3.944 (0.079)	2.323 (0.077)	3.357 (0.080)	**1.941** (0.082)
Example 2	$n = 50$	–	–	2.532 (0.023)	**2.141** (0.039)	–	–	3.948 (0.055)	**3.647** (0.036)
	$n = 100$	–	–	2.344 (0.016)	**1.910** (0.028)	–	–	3.580 (0.036)	**3.323** (0.029)
	$n = 200$	–	–	2.245 (0.014)	**1.789** (0.013)	–	–	3.331 (0.026)	**1.953** (0.023)
Example 3	$n = 50$	–	–	2.551 (0.024)	**2.295** (0.019)	–	–	4.473 (0.124)	**3.720** (0.120)
	$n = 100$	–	–	2.448 (0.018)	**2.169** (0.018)	–	–	4.136 (0.065)	**3.341** (0.063)
	$n = 200$	–	–	2.393 (0.017)	**2.107** (0.016)	–	–	3.968 (0.058)	**3.169** (0.056)
Example 4	$n = 50$	10.414 (0.098)	15.397 (0.204)	10.426 (0.104)	**10.036** (0.139)	14.121 (0.143)	31.670 (1.344)	13.859 (0.166)	**12.727** (0.187)
	$n = 100$	9.952 (0.077)	16.095 (0.312)	9.834 (0.081)	**9.277** (0.094)	13.296 (0.111)	29.344 (0.798)	12.913 (0.103)	**11.452** (0.109)
	$n = 200$	9.606 (0.059)	15.994 (0.245)	9.420 (0.060)	**8.777** (0.061)	12.771 (0.092)	25.441 (0.703)	12.293 (0.086)	**10.531** (0.077)
Example 5	$n = 50$	–	–	11.688 (0.107)	**11.288** (0.104)	–	–	16.455 (0.236)	**16.088** (0.326)
	$n = 100$	–	–	11.044 (0.091)	**10.423** (0.093)	–	–	15.748 (0.340)	**14.502** (0.343)
	$n = 200$	–	–	10.865 (0.098)	**10.273** (0.099)	–	–	14.968 (0.150)	**13.528** (0.146)
Example 6	$n = 50$	–	–	10.346 (0.096)	**10.142** (0.097)	–	–	15.009 (0.358)	**14.456** (0.366)
	$n = 100$	–	–	9.952 (0.069)	**9.635** (0.066)	–	–	14.453 (0.262)	**13.766** (0.257)
	$n = 200$	–	–	9.659 (0.066)	**9.273** (0.065)	–	–	13.759 (0.450)	**13.008** (0.448)

Note: "–" denotes that the corresponding method cannot be applied in these examples.

such as the death rate, modeling the age-at-death distribution as a random object provides deeper insights into human longevity.

For this purposes, we consider the random object derived from HMD (2024) (HMD, `http://www.mortality.org`), which contains life tables for 39 countries in 2013. The response object is the country-specific age-at-death distribution in 2013, restricted to the age range $[20, 110]$. For predictors, we consider carbon dioxide emissions in metric tons per capita (CO2E) serve as the effect-modifier U, and that year-on-year percentage change in GDP (GDPC), current health care expenditure as a percentage of GDP (HCE), the human development index (HDI), and infant mortality per 1000 live births (IM) form the Euclidean predictor vector $\boldsymbol{X} = (X_1, \ldots, X_4)^{\mathrm{T}}$. Thus, $\boldsymbol{X}_i \in \mathbb{R}^4$ and $U_i \in \mathbb{R}$ constitute the predictors for the i-th country, $i = 1, \ldots, 39$. The data are collected by Bhattacharjee and Müller (2023).

To evaluate the prediction performance of the proposed method, we randomly select n_{train} observations as the training data and use the remaining n_{test} observations as the test data. Then, GFR, LFR, PFR and VFR are accessed by the mean squared prediction errors: $\text{MSPE} = n_{\text{test}}^{-1} \sum_{i=1}^{n_{\text{test}}} d_W^2(Y_i, \widehat{Y}_i)$. Here, Y_i is the i-th testing observation and $\widehat{Y}_i$ represents the prediction for each method based on the training data. We repeat the above procedure 50 times and set the training sample size as $n_{\text{train}} \in \{20, 25, 30\}$.

Table 3: The MSPE values (with standard errors in parentheses) of each methods.

Methods	$n_{\text{train}} = 20$	$n_{\text{train}} = 25$	$n_{\text{train}} = 30$
GFR	6.380 (0.230)	6.538 (0.358)	5.959 (0.305)
LFR	47.080 (7.832)	89.679 (59.037)	25.915 (1.571)
PFR	10.223 (0.561)	11.559 (0.670)	9.565 (0.717)
VFR	**6.204** (0.253)	**6.489** (0.374)	**5.677** (0.294)

Table 3 reports the mean MSPEs and standard errors for various Fréchet regression. Among the five candidate predictors considered as the local predictor U in VFR, CO2E yielded the best performance in terms of predictive accuracy MSPE. This result is reasonable because CO2E emissions exhibit a clear long-term trend and are closely tied to environmental and health risks that shape the age-at-death distribution (Azimi and Rahman, 2024). Moreover, VFR always achieves smaller mean MSPE values than GFR, LFR, and PFR across all scenarios, indicating that our model provides superior predictive performance for the human mortality data. From an interpretive perspective, unlike global or partially global Fréchet regression models which assume constant predictor effects across countries, the varying-coefficient Fréchet regression allows the effects of the predictors $\boldsymbol{X}$ on the age-at-death distribution to vary with the level of CO2E.

7. Discussion

In this paper, we introduce a varying-coefficient Fréchet regression model and its corresponding estimators to handle both Euclidean and Non-Euclidean predictors. Additionally, we derive the convergence rates of the proposed estimators. Several examples with random objects are used to show the performance of varying-coefficient Fréchet regression model. Taken together, our results provide a foundation for further methodological and theoretical developments in modeling non-Euclidean data.

Deriving the convergence rate for the bias term $d(s_{\oplus}(\boldsymbol{x}, u), \widetilde{s}_{\oplus}(\boldsymbol{x}, u))$ involves two fundamental technical obstacles that are not directly handled by existing theoretical tools for varying-coefficient models or global Fréchet regression. The first technical obstacle is to establish that $\mathbb{E}\left[\zeta_h(U, u) f(\boldsymbol{X}, Y)\right]$ converges to $\mathbb{E}[f(\boldsymbol{X}, Y) \mid U = u]$ at a certain rate as $h \to 0$. Establishing such a convergence result is nontrivial when both Y and U take values in general metric spaces. The second technical obstacle arises from the implicit definition of the varying-coefficient Fréchet regression. Neither $s_{\oplus}(\boldsymbol{x}, u)$ nor $\tilde{s}_{\oplus}(\boldsymbol{x}, u)$ admits a closed form solution, which poses nontrivial challenges for the subsequent theoretical analysis. To overcome this, we develop a novel approach that generalizes the empirical process theory for M-estimators.

Despite the progress made in this work, a number of challenges and op-

portunities for further development remain. The incorporation of variable selection methods into Fréchet regression analysis would greatly enhance its applicability. However, in the absence of explicit coefficients in the VFR model, it remains unclear how existing methods can be adapted to this more complex setting. Another important direction is the development of inference tools to understand the relationships between random object response and predictor. For the broader applicability of Fréchet regression, it is essential to clarify the significance of predictor effects. Dubey and Müller (2019) introduced a Fréchet analysis of variance for random objects, including a test statistic and its asymptotic distribution. A natural direction for future work is to investigate whether likelihood ratio tests can be developed for general metric spaces, thereby offering a unified framework for assessing and testing variable significance in Fréchet regression.

Supplementary Materials

This supplementary material contains additional implementation and simulation details, explanations and verifications of assumptions, and proofs of the main results.

Statistica Sinica: Newly accepted Paper

Acknowledgements

This work was supported by the National Key R&D Program of China (Grant No. 2023YFA1008702) and the National Natural Science Foundation of China (Grant No. 12571300, 12601543).

References

Azimi, M. N. and M. M. Rahman (2024, April). Unveiling the health consequences of air pollution in the world's most polluted nations. *Scientific Reports 14*(1), 9856.

Bhattacharjee, S. and H.-G. Müller (2022). Concurrent object regression. *Electronic Journal of Statistics 16*(2), 4031–4089.

Bhattacharjee, S. and H.-G. Müller (2023, August). Single index Fréchet regression. *Annals of Statistics 51*(4), 1770–1798.

Cai, Z., J. Fan, and R. Li (2000, September). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association 95*(451), 888–902.

Chen, Y., A. Gajardo, J. Fan, Q. Zhong, P. Dubey, K. Han, S. Bhattacharjee, and H.-G. Müller (2020). *frechet: Statistical Analysis for Random Objects and Non-Euclidean Data*. R package version 0.2.0.

Dryden, I. L., A. Koloydenko, and D. Zhou (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics 3*(3),

Statistica Sinica: Newly accepted Paper

1102–1123.

Dubey, P. and H.-G. Müller (2019, December). Fréchet analysis of variance for random objects. *Biometrika 106*(4), 803–821.

Dubey, P. and H.-G. Müller (2022, October). Modeling time-varying random objects and dynamic networks. *Journal of the American Statistical Association 117*(540), 2252–2267.

Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications* (1 ed.), Volume 66. London: Chapman & Hall.

Fan, J., C. Zhang, and J. Zhang (2001, February). Generalized likelihood ratio statistics and wilks phenomenon. *Annals of Statistics 29*(1), 153–193.

Fan, J. and W. Zhang (1999, October). Statistical estimation in varying coefficient models. *Annals of Statistics 27*(5), 1491–1518.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B (Statistical Methodology) 55*(4), 757–796.

HMD (2024). Human mortality database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at `http://www.mortality.org` (data downloaded on January 19, 2025).

Huang, J. Z., C. O. Wu, and L. Zhou (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika 89*(1), 111–128.

Statistica Sinica: Newly accepted Paper

REFERENCES

Iao, S. I., Y. Zhou, and H.-G. Müller (2025). Deep Fréchet regression. *Journal of the American Statistical Association 0*(0), 1–12.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications 9*, 141–142.

Nye, T. M. W., X. Tang, G. Weyenberg, and R. Yoshida (2017). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika 104*(4), 901–922.

Pei, Y., T. Huang, H. Peng, and J. You (2022). Network-based clustering for varying coefficient panel data models. *Journal of Business* & Economic Statistics *40*(2), 578–594.

Petersen, A. and H.-G. Müller (2019, April). Fréchet regression for random objects with euclidean predictors. *Annals of Statistics 47*(2), 691–719.

Qiu, R., Z. Yu, and R. Zhu (2024). Random forest weighted local Fréchet regression with random objects. *Journal of Machine Learning Research 25*(107), 1–69.

Tucker, D. C. and Y. Wu (2025). Partially-global Fréchet regression. *Statistica Sinica 35*(1), 1–24.

Tucker, D. C., Y. Wu, and H.-G. Müller (2023). Variable selection for global Fréchet regression. *Journal of the American Statistical Association 118*(542), 1023–1037.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A 26*, 359–372.

Statistica Sinica: Newly accepted Paper

REFERENCES

Zhang, Y., C. Zhu, and X. Shao (2025). Change-point detection for object-valued time series. *Journal of Business* & Economic Statistics *0*(0), 1–25.

Authors information:

Yanzhao Wang

Center for Applied Statistics and School of Statistics, Renmin University of China

Co-first author. E-mail: wangyanzhao@ruc.edu.cn

Jianqiang Zhang

School of Statistics and Data Science, Zhejiang Gongshang University

Co-first author. E-mail: zhangjqs@163.com

Wangli Xu

Center for Applied Statistics and School of Statistics, Renmin University of China

Corresponding author. E-mail: wlxu@ruc.edu.cn, ORCID: 0000-0003-4983-9950