Statistica Sinica Preprint No: SS-2023-0341	
Title	Robust Inverse Regression for Multivariate Elliptical Functional Data
Manuscript ID	SS-2023-0341
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0341
Complete List of Authors	Eftychia Solea, Eliana Christou and Jun Song
Corresponding Authors	Jun Song
E-mails	junsong@korea.ac.kr

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Statistica Sinica

Robust Inverse Regression for Multivariate Elliptical Functional Data

Eftychia Solea, Eliana Christou and Jun Song

Queen Mary University of London,

University of North Carolina at Charlotte, Korea University

Abstract: Functional data have received significant attention as they frequently appear in modern applications, such as functional magnetic resonance imaging (fMRI) and natural language processing. The infinite-dimensional nature of functional data makes it necessary to use dimension reduction techniques. Most existing techniques, however, rely on the covariance operator, which can be affected by heavy-tailed data and unusual observations. Therefore, in this paper, we consider a *robust sliced inverse regression for multivariate elliptical functional data*. For that reason, we introduce a new statistical linear operator, called the *conditional spatial sign Kendall's tau covariance operator*, which can be seen as an extension of the multivariate Kendall's tau to both the conditional and functional settings. The new operator is robust to heavy-tailed data and outliers, and hence can provide a robust estimate of the sufficient predictors. We also derive the convergence rates of the proposed estimators for both completely and partially observed data. Finally, we demonstrate the finite sample performance of our estimator using simulation examples and a real dataset based on fMRI.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Key words and phrases: dimension reduction, elliptical distribution, functional data, sliced inverse regression, spatial sign.

1. Introduction

The complexity of data structures has increased over the past few decades as data storage capacity and its usage demand have exploded. Following these phenomena, functional data analysis (FDA) has gained great attention, which treats an entire curve or a vector of curves as a single observation. Functional data are considered to be random elements in infinite-dimensional linear spaces and the extension of multivariate data analysis to functional data is highly non-trivial. However, the rapid development of FDA in the past few decades has enabled us to use a variety of techniques to analyze such infinite dimensional data. See Wang et al. (2016) for a comprehensive overview on FDA.

The infinite-dimensional nature of functional data leverages the use of dimension reduction techniques, i.e., techniques that replace the functional objects with finite ones while maintaining all the necessary information. At

Corresponding author: Jun Song, Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea. E-mail: junsong@korea.ac.kr

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

the unsupervised setting, a common dimension reduction technique is that of functional principal component analysis (FPCA); see Hall and Hosseini-Nasab (2006), and Happ and Greven (2018) for an overview. At the supervised setting, many authors considered extensions of dimension reduction techniques to functional data. For example, Ferré and Yao (2003) proposed functional sliced inverse regression (FSIR) by extending the sliced inverse regression (SIR) of Li (1991) to the case where the predictor is a function, while Li and Song (2022) further generalized sufficient dimension reduction to the nonlinear case and where both the response and the predictor are random elements in a Hilbert space.

The aforementioned methods are based on the covariance operator, which can be sensitive under heavy-tailed data and unusual observations. To address these challenges, at the unsupervised setting, efficient robust methods for FPCA have been recently developed. For example, Locantore et al. (1999) introduced the spherical covariance operator to replace the covariance operator. Gervini (2008) introduced the functional median for robust mean estimation and studied the properties of the principal components of the spherical covariance operator, under the assumption that the observed functions lie in a finite-dimensional Hilbert space. Bali et al. (2011) extended the projection-pursuit method of Li and Chen (1985) to

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

the functional setting and Kraus and Panaretos (2012) proposed to replace the covariance operator with a dispersion operator defined through a variational problem. Boente et al. (2019) and Wang et al. (2022) introduced the functional pairwise spatial sign operator that extends the multivariate Kendall's tau matrix of Choi and Marden (1998) to the functional setting. For Euclidean finite-dimensional data, the multivariate Kendall's tau matrix has also been studied by several other authors; see Marden (1999), Han and Liu (2018), Croux et al. (2002), Visuri et al. (2000), and Jackson and Chen (2004). In particular, Marden (1999) showed that the population multivariate Kendall's tau shares the same eigenspace with the covariance matrix under the coordinate-wise symmetric condition.

All above works tackle the problem at the unsupervised learning setting. In this article, we introduce *robust inverse regression for multivariate elliptical functional data*. We relax the Gaussian assumption in Ferré and Yao (2003), and we consider the situation where the multivariate functional predictors are not Gaussian and may be characterized by the presence of atypical observations. Our proposal thus extends two lines of existing and relevant research: from robust inverse regression of random variables to that of random functions, and from unsupervised robust dimension reduction for functional data to supervised robust dimension reduction for functional

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

data. We note that Chen et al. (2022) introduced elliptical sliced inverse regression for finite-dimensional data. However, our method involves vectors of random functions and hence, requires different techniques for both the method and the theory. We make several contributions as below:

- We define the elliptical distribution for a vector of random functions, extending the existing definition of Boente et al. (2014) to the multivariate setting.

- We introduce a new statistical linear operator, called the *conditional spatial sign Kendall's tau covariance operator*, which can be seen as an extension of the multivariate Kendall's tau to both the conditional and functional settings, and is capable to handle heavy-tailed functional data and outliers. We show that the conditional spatial sign Kendall's tau covariance operator has the same eigenfunctions with the conditional covariance operator, and hence we can formulate the generalized eigenvalue problem based on this new operator to achieve estimation robustness.

- We derive the convergence rates of the proposed estimators for both completely and partially observed data. In practice, we can only observe the functions at discrete time points, and the new theoretical

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

results support practical estimation procedure.

The rest of the paper is organized as follows. Section 2 introduces the multivariate elliptical random elements in Hilbert spaces and the conditional spatial sign Kendall's tau covariance operator. Section 3 describes the proposed method of robust functional SIR. Section 4 presents the asymptotic results of the method, while Section 5 demonstrates its sample estimation. Finally, the finite sample performance of the proposed method is illustrated through simulation studies in Section 6 and through a neuroimaging dataset in Section 7. All proofs and some additional results can be found in the supplementary file.

2. Multivariate elliptical random elements in Hilbert spaces

In this section we present basic notations and definitions that will be used throughout the paper. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathscr{H}$ be a separable Hilbert space of real-valued functions defined on T, where T is a closed interval in $\mathbb{R}$. Let $\langle \cdot, \cdot \rangle_{\mathscr{H}}$ represent the inner product in $\mathscr{H}$ with the induced norm $\| \cdot \|_{\mathscr{H}}$. A random element U in $\mathscr{H}$ is a mapping from Ω to $\mathscr{H}$ that is measurable with respect to the Borel σ-field generated by the open sets in $\mathscr{H}$. Assuming

Assumption 1. $\mathbb{E}\|U\|_{\mathscr{H}}^2 < \infty$,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

implies that $\mathbb{E}\|U\|_{\mathscr{H}} < \infty$, under which the mapping $\mathscr{H} \to \mathbb{R}, s \mapsto \mathbb{E}\langle s, U\rangle_{\mathscr{H}}$ is a bounded linear functional. Then, by Riesz representation there exists a unique $\mu_U \in \mathscr{H}$, such that

$$\langle s, \mu_U\rangle_{\mathscr{H}} = \mathbb{E}(\langle s, U\rangle_{\mathscr{H}}). \tag{2.1}$$

The function μ_U is called the mean of U and it can also be written as $\mathbb{E}(U)$. Under Assumption 1, we can define the covariance operator of U, $\Sigma_{UU} = \mathbb{E}[\{U - \mathbb{E}(U)\} \otimes \{U - \mathbb{E}(U)\}]$, where $\otimes$ represents the tensor product on $\mathscr{H}$. If, for example, $\mathscr{H} = L_2(T)$, where $L_2(T)$ denotes the space of square-integrable real-valued functions, then the mean of U is the function $t \mapsto \mathbb{E}U(t)$ and the covariance operator is defined as the integral operator $\Sigma_{UU}(f) = \int_T f(s)\sigma_{UU}(s,t)dt$ for $f \in \mathscr{H}$, where $\sigma_{UU}(s,t) = \mathrm{cov}\{U(s), U(t)\}$ is the covariance function of U.

Under Assumption 1, it can be shown that Σ_{UU} is a self-adjoint, non-negative definite and trace-class operator, and that it belongs to the Hilbert space of Hilbert-Schmidt operators over $\mathscr{H}$. Hence, it achieves a spectral decomposition $\sum_{r=1}^{\infty} \lambda_r \phi_r \otimes \phi_r$, where $\{\lambda_r\}_{r\geq 1}$ are the nonnegative eigenvalues satisfying $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$, and $\{\phi_r\}_{r\geq 1}$ are the eigenfunctions forming an orthonormal basis in $\mathscr{H}$. Moreover, $U - \mu_U$ can be expressed as

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

the Karhunen-Loève expansion (Bosq, 2000)

$$U - \mu_U = \sum_{r=1}^{\infty} \lambda_r^{1/2} \xi_r \, \phi_r, \tag{2.2}$$

where $\xi_r = \lambda_r^{-1/2} \langle U - \mu_U, \phi_r \rangle_{\mathscr{H}}$, $r = 1, 2, \ldots$, are zero mean, unit variance, and uncorrelated random variables. Next, we give the definition of the elliptical distribution for a univariate random element in $\mathscr{H}$, as introduced in Boente et al. (2014, 2019).

Definition 1. A random element U in $\mathscr{H}$ is said to follow a functional elliptical distribution with parameters $\mu \in \mathscr{H}$ and $\Sigma : \mathscr{H} \mapsto \mathscr{H}$, where Σ is a self-adjoint, nonnegative definite and compact operator, if and only if, for any $d \geq 1$ and for any linear and bounded operator $A : \mathscr{H} \mapsto \mathbb{R}^d$, we have that $\mathbb{E}[\exp\{it^\intercal (AU)\}] = \exp(it^\intercal A\mu)\varphi\{t^\intercal (A\Sigma A^*)t\}$, $t \in \mathbb{R}^d$, where φ is a valid characteristic function in $\mathbb{R}^d$ and $A^* : \mathbb{R}^d \mapsto \mathscr{H}$ denotes the adjoint operator of A. Equivalently, AU is a d-dimensional elliptical random vector, written as $AU \sim \mathscr{E}_d(A\mu, A\Sigma A^*, \varphi)$. We write $U \sim \mathcal{E}(\mu, \Sigma, \varphi)$, and we call μ the location parameter and Σ the scatter operator of U.

Note that, if $\mu_U = \mathbb{E}(U)$ exists, then $\mu_U = \mu$, and if $\mathbb{E}\|U\|^2_{\mathscr{H}} < \infty$, then $\Sigma_{UU} = a\Sigma$ for some constant a (Boente et al., 2014, 2019). Moreover, elliptical random elements in $\mathscr{H}$ are closed through linear and bounded transformations; see Lemma 2.1 in Boente et al. (2014, 2019). According

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

to Boente et al. (2014, 2019), the elliptical random elements in $\mathscr{H}$ can be constructed as follows: Let N be a Gaussian random element in $\mathscr{H}$ with zero mean and covariance operator Σ_{NN}, and let S be a nonnegative random variable that is independent of N. Given $\mu \in \mathscr{H}$, the random element $U \stackrel{d}{=} \mu + SN$ is an elliptical random element in $\mathscr{H}$, i.e., $U \sim \mathcal{E}(\mu, \Sigma, \varphi)$ with $\Sigma = \Sigma_{NN}$ and $\varphi(x) = \mathbb{E}\{\exp(-xS/2)\}$. Throughout the paper, for an elliptical random element U we assume that $\mathbb{E}(S^2) < \infty$, so that $\Sigma_{UU} = \mathbb{E}(S^2)\Sigma$. Further, for model identifiability, we assume that $\mathbb{E}(S^2) = 1$. The class of the elliptical distributions in $\mathscr{H}$ includes the Gaussian distribution by taking $\varphi(t) = \exp(-t/2)$ and $\Sigma = \Sigma_{UU}$.

We now turn to vector-valued random functions. For each $i = 1, \ldots, p$, let $\mathscr{H}_i$ be a separable Hilbert space of real-valued functions on T with inner product $\langle \cdot, \cdot \rangle_{\mathscr{H}_i}$. Let $\mathscr{H} = \oplus_{i=1}^p \mathscr{H}_i$ be the direct sum of $\mathscr{H}_1, \ldots, \mathscr{H}_p$. That is, $\oplus_{i=1}^p \mathscr{H}_i$ is the Cartesian product $\mathscr{H}_1 \times \cdots \times \mathscr{H}_p$ with its inner product defined by $\langle f, g \rangle_{\oplus \mathscr{H}} = \sum_{i=1}^p \langle f_i, g_i \rangle_{\mathscr{H}_i}$, where f and g are members of $\oplus_{i=1}^p \mathscr{H}_i$ and f_i and g_i are the ith components of f and g, respectively.

Let $X = (X^1, \ldots, X^p)$ be a random element in $\oplus_{i=1}^p \mathscr{H}_i$. For each $i = 1, \ldots, p$, we assume that $\mathbb{E}\|X^i\|_{\mathscr{H}}^2 < \infty$. The covariance operator between X^i and X^j is defined as $\Sigma_{X^i X^j} = \mathrm{cov}(X^i, X^j) = \mathbb{E}\{(X^i - \mu_{X^i}) \otimes (X^j - \mu_{X^j})\}$, where μ_{X^i} and μ_{X^j} are the means of X^i and X^j, respectively

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

as defined in (2.1), for $i, j = 1, \ldots, p$. Note that, $\Sigma_{X^iX^j} \in \mathscr{B}(\mathscr{H}_j, \mathscr{H}_i)$, where $\mathscr{B}(\mathscr{H}_j, \mathscr{H}_i)$ denotes the set of all bounded operators from $\mathscr{H}_j$ to $\mathscr{H}_i$. Define Σ_{XX} to be the operator $\oplus_{i=1}^p \mathscr{H}_i \to \oplus_{i=1}^p \mathscr{H}_i$, $(t_1, \ldots, t_p) \mapsto (\sum_{i=1}^p \Sigma_{X^1X^i} t_i, \ldots, \sum_{i=1}^p \Sigma_{X^pX^i} t_i)$. Intuitively, Σ_{XX} can be interpreted as the $p \times p$ matrix whose (i, j)th entry is $\Sigma_{X^iX^j}$. The covariance operator Σ_{XX} is a linear, self-adjoint, symmetric and nonnegative definite operator. Moreover, it is assumed to be a compact operator in $\oplus_{i=1}^p \mathscr{H}_i$; see Proposition 2 of Happ and Greven (2018) for more details. Then, there exists a complete orthonormal basis of eigenfunctions $\{\psi_r\}_{r\geq 1}$ in $\oplus_{i=1}^p \mathscr{H}_i$ such that $\Sigma_{XX} = \sum_{r=1}^\infty \gamma_r \psi_r \otimes \psi_r$, where the eigenvalues $\{\gamma_r\}_{r\geq 1}$ satisfy $\gamma_1 \geq \gamma_2 \geq \ldots \geq 0$. Then, X admits the multivariate Karhunen-Loève decomposition (see Proposition 4 of Happ and Greven (2018))

$$X - \mu_X = \sum_{r=1}^\infty \gamma_r^{1/2} \rho_r \, \psi_r, \tag{2.3}$$

where $\rho_r = \gamma_r^{-1/2} \langle X - \mu_X, \psi_r \rangle_{\oplus \mathscr{H}}$, $r = 1, 2, \ldots$, are zero mean, unit-variance, and uncorrelated random variables, and $\mu_X = (\mu_{X^1}, \ldots, \mu_{X^p}) \in \oplus_{i=1}^p \mathscr{H}_i$.

Suppose for any $d \geq 1$ and each $i = 1, \ldots, d, j = 1, \ldots, p$, A_{ij} is a linear and bounded operator from $\mathscr{H}_j$ to $\mathbb{R}$. We define the matrix of operators $A = \{A_{ij}\}_{i,j=1}^{d,p}$ as the mapping $A : \oplus_{i=1}^p \mathscr{H}_i \to \mathbb{R}^d$, $(t_1, \ldots, t_p) \mapsto (\sum_{\ell=1}^p A_{1\ell} t_\ell, \cdots, \sum_{\ell=1}^p A_{d\ell} t_\ell)$. Using this convention we can define the multivariate elliptical random element in $\oplus_{i=1}^p \mathscr{H}_i$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Definition 2. Suppose X is a random element in $\oplus_{i=1}^{p}\mathscr{H}_i$. We say that X follows the multivariate functional elliptical distribution with location parameter $\mu \in \oplus_{i=1}^{p}\mathscr{H}_i$ and scatter operator $\Sigma : \oplus_{i=1}^{p}\mathscr{H}_i \mapsto \oplus_{i=1}^{p}\mathscr{H}_i$, if and only if, for any $d \geq 1$, and for any linear and bounded $d \times p$ matrix of operators $A = \{A_{ij}\}_{i,j}^{d,p} : \oplus_{i=1}^{p}\mathscr{H}_i \mapsto \mathbb{R}^d$, we have that $AX \sim \mathscr{E}_d(A\mu, A\Sigma A^*, \varphi)$, where $A^* : \mathbb{R}^d \mapsto \oplus_{i=1}^{p}\mathscr{H}_i$ denotes the adjoint operator of A and φ is a valid characteristic function in $\mathbb{R}^d$. We write $X \sim \mathcal{E}_p(\mu, \Sigma, \varphi)$.

Remark 1. As in the classical setting, elliptical random elements in $\oplus_{i=1}^{p}\mathscr{H}_i$ can be characterised by their marginals. In particular, X is an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$ if and only if for any bounded and linear operator $B : \oplus_{i=1}^{p}\mathscr{H}_i \mapsto \oplus_{i=1}^{k}\mathscr{H}_i$, $k \leq p$, BX is an elliptical random element in $\oplus_{i=1}^{k}\mathscr{H}_i$ with parameters $B\mu$ and $B\Sigma B^*$. That is, if X is an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$, then X_j is an elliptical random element in $\mathscr{H}_j$. As in the case with $p = 1$, the following construction allows to obtain an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$. Let N be a Gaussian random element in $\oplus_{i=1}^{p}\mathscr{H}_i$ with zero mean and covariance operator Σ_{NN}, and let S be a non-negative random variable that is independent of N. Given $\mu \in \oplus_{i=1}^{p}\mathscr{H}_i$, the random element $X \overset{d}{=} \mu + SN$ is an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$, i.e., $X \sim \mathcal{E}_p(\mu, \Sigma, \varphi)$ with $\Sigma = \Sigma_{NN}$. We assume that the covariance operator of $X \in \oplus_{i=1}^{p}\mathscr{H}_i$ is the same as its scatter operator Σ.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

2.1 Spatial sign Kendall's tau covariance operator

We next define the spatial sign Kendall's tau covariance operator of a vector of functions $X \in \oplus_{i=1}^{p} \mathscr{H}_i$ to the multiple setting.

Definition 3. Assume X is a random element in $\oplus_{i=1}^{p} \mathscr{H}_i$. Let $\tilde{X}$ be an independent copy of X, the operator defined as

$$T_{XX} = \mathbb{E}\Big\{ \frac{(X - \tilde{X}) \otimes (X - \tilde{X})^{\intercal}}{\|X - \tilde{X}\|^2_{\oplus \mathscr{H}}} \Big\}, \tag{2.4}$$

is called the spatial sign Kendall's tau covariance operator.

The spatial sign Kendall's tau covariance operator T_{XX} can be seen as the covariance operator of the functional pairwise spatial signs $(X - \tilde{X})/\|X - \tilde{X}\|_{\oplus \mathscr{H}}$ and it exists without any moment assumptions. Moreover, (2.4) shows that T_{XX} is self-adjoint. An alternative robust estimator of the covariance operator, introduced by Locantore et al. (1999), is called the spherical covariance operator and is given by $R_{XX} = \mathbb{E}\Big\{ \frac{(X-\mu_X)\otimes(X-\mu_X)^{\intercal}}{\|X-\mu_X\|^2_{\oplus \mathscr{H}}} \Big\}$. The next lemma shows that the Kendall's tau covariance operator, T_{XX}, and the spherical covariance operator, R_{XX}, coincide when $X \sim \mathcal{E}_p(\mu_X, \Sigma, \varphi)$. However, the advantage of using T_{XX} is that it avoids the estimation of the location centre μ_X.

Lemma 1. *Let X be an elliptical random element in $\oplus_{i=1}^{p} \mathscr{H}_i$, satisfying*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Assumption 1. Then, $\mathbb{E}\left\{\frac{(X-\tilde{X})\otimes(X-\tilde{X})^{\intercal}}{\|X-\tilde{X}\|^2_{\oplus\mathscr{H}}}\right\} = \mathbb{E}\left\{\frac{(X-\mu_X)\otimes(X-\mu_X)^{\intercal}}{\|X-\mu_X\|^2_{\oplus\mathscr{H}}}\right\}$, *where* $\tilde{X}$ *is an independent copy of* X.

The following theorem states that, for elliptical random elements in $\oplus_{i=1}^{p}\mathscr{H}_i$, the spatial sign Kendall's tau covariance operator T_{XX} has the same set of eigenfunctions with the covariance operator Σ_{XX}. In addition, they have the same descending order of the eigenvalues. For a similar result for a univariate X see Kraus and Panaretos (2012) and Wang et al. (2022).

Theorem 1. *Let* X *be an elliptical random element in* $\oplus_{i=1}^{p}\mathscr{H}_i$, *satisfying Assumption 1, with mean function* $\mu_X \in \oplus_{i=1}^{p}\mathscr{H}_i$ *and covariance operator* Σ_{XX}. *Then, we have* $T_{XX} = \sum_{r=1}^{\infty}\delta_r\psi_r\otimes\psi_r$, *where the eigenvalues* $\{\delta_r\}_{r\geq1}$ *of* T_{XX} *satisfy* $\delta_r = \mathbb{E}\left(\gamma_r Y_r^2/\sum_{k=1}^{\infty}\gamma_k Y_k^2\right)$, $r = 1, 2, \ldots,$ $\{\gamma_r\}_{r\geq1}$ *and* $\{\psi_r\}_{r\geq1}$ *are the eigenvalues and the eigenfunctions of the covariance operator* Σ_{XX}, *respectively, and* $\{Y_r\}_{r\geq1}$, *are iid standard normal random variables.*

The proof of Theorem 1 for the multivariate case $X \in \oplus_{i=1}^{p}\mathscr{H}_i$ follows the same arguments as in the proof of Theorem 4 in Wang et al. (2022), and thus it is omitted. A similar result to Theorem 1 was obtained by Gervini (2008), under the assumption of exchangeability of the scores. However, Gervini (2008) does not require any moment assumptions as the author assumes that the Karhunen-Loève expansion is a finite sum, i.e., that function

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

X lies in finite-dimensional Hilbert spaces. In contrast, our results rely on infinite-dimensional functions X, and, without the Assumption 1 of finite second moments, the convergence of an infinite Karhunen-Loève series is not guaranteed (Kraus and Panaretos, 2012).

2.2 Conditional spatial sign covariance operator

We now define the conditional spatial sign covariance operator, a generalization of the Kendall's tau covariance matrix to the functional and conditional settings. Let Y be a random variable with support Ω_Y. For each fixed $y \in \Omega_Y$, let $\mu_{X|Y}(y)$ be the Riesz representation of the bounded linear functional $\oplus_{i=1}^{p}\mathscr{H}_i \ni s \mapsto \mathbb{E}(\langle s, X\rangle_{\oplus\mathscr{H}}|y)$. Then, the conditional expectation of X given Y is the mapping $y \mapsto \mu_{X|Y}(y)$. Moreover, we can define the conditional covariance operator of $\mu_{X|Y}(Y)$ by $\Sigma_{XX|Y} = \mathbb{E}[\{\mu_{X|Y}(Y) - \mu_X\} \otimes \{\mu_{X|Y}(Y) - \mu_X\}^{\intercal}]$. Then, define

$$T_{XX|Y} = \mathbb{E}\Big[\frac{\{\mu_{X|Y}(Y) - \mu_{X|Y}(\tilde{Y})\} \otimes \{\mu_{X|Y}(Y) - \mu_{X|Y}(\tilde{Y})\}^{\intercal}}{\|\mu_{X|Y}(Y) - \mu_{X|Y}(\tilde{Y})\|^2_{\oplus\mathscr{H}}}\Big], \tag{2.5}$$

where $\tilde{Y}$ is an independent copy of Y.

Definition 4. We call the operator $T_{XX|Y}$ defined in (2.5) the conditional spatial sign covariance operator of $\mu_{X|Y}(Y)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

3. Robust inverse regression

3.1 Population level

Let Y be a random variable and X be an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$ with mean function $\mu_X \in \oplus_{i=1}^{p}\mathscr{H}_i$ and covariance operator Σ_{XX}. Dimension reduction techniques aim at finding functions $\beta_1, \ldots, \beta_K$ in $\oplus_{i=1}^{p}\mathscr{H}_i$, $K \geq 1$, such that

$$Y = g(\langle \beta_1, X\rangle_{\oplus\mathscr{H}}, \ldots, \langle \beta_K, X\rangle_{\oplus\mathscr{H}}, \epsilon), \tag{3.1}$$

where g is an arbitrary unknown function on $\mathbb{R}^{K+1}$, and ϵ is independent of X. An equivalent definition of (3.1) is that $Y \perp\!\!\!\perp X \,|\, \langle \beta_1, X\rangle_{\oplus\mathscr{H}}, \ldots, \langle \beta_K, X\rangle_{\oplus\mathscr{H}}$. The functions $\beta_1, \ldots, \beta_K$ in $\oplus_{i=1}^{p}\mathscr{H}_i$ are called the *functional dimension reduction directions* and the subspace spanned by $\beta_1, \ldots, \beta_K$ is called the *functional dimension reduction subspace*. The smallest functional dimension reduction subspace is called the *functional central subspace* and is denoted by $\mathcal{S}_{Y|X}$. As in the classical setting, the functional dimension reduction directions $\beta_1, \ldots, \beta_K$ are not identifiable. However, the functional central subspace $\mathcal{S}_{Y|X}$ is identifiable, and is the goal of the estimation. A common assumption in the dimension reduction literature is that of the linearity condition, which is adopted to the functional case and is satisfied in our setting since X has an elliptical distribution.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Assumption 2. For any function $b \in \oplus_{i=1}^{p} \mathscr{H}_i$ there exist constants c_0, $c_1, \ldots, c_K$ such that $\mathbb{E}(\langle b, X\rangle_{\oplus\mathscr{H}} | \langle \beta_1, X\rangle_{\oplus\mathscr{H}}, \ldots, \langle \beta_K, X\rangle_{\oplus\mathscr{H}}) = c_0 + c_1\langle \beta_1, X\rangle_{\oplus\mathscr{H}} + \ldots + c_K\langle \beta_K, X\rangle_{\oplus\mathscr{H}}$.

Under Assumption 2, $\mu_{X|Y}(Y) - \mu_X$ belongs to the subspace spanned by $\Sigma_{XX}\beta_1, \ldots, \Sigma_{XX}\beta_K$ (Theorem 2.1, Ferré and Yao (2003)). The next Theorem provides a parallel result, where Σ_{XX} is replaced with the spatial sign Kendall's tau covariance operator T_{XX} defined in (2.4).

Theorem 2. *Let X be an elliptical random element in $\oplus_{i=1}^{p} \mathscr{H}_i$, satisfying Assumption 1. Then, $\mu_{X|Y}(Y) - \mu_X \in \text{span}\{T_{XX}\beta_1, \ldots, T_{XX}\beta_K\}$.*

It follows from Theorem 2 that $\Sigma_{XX|Y}$ is degenerate in any direction T_{XX}-orthogonal to the central subspace, implying that the range of $\Sigma_{XX|Y}$ is contained in $\text{span}\{T_{XX}\beta_1, \ldots, T_{XX}\beta_K\}$. Thus, a subspace of the central subspace can be recovered through the T_{XX}-orthonormal eigenfunctions of $\Sigma_{XX|Y}$ corresponding to its K largest eigenvalues. However, the covariance operator $\Sigma_{XX|Y}$ can be sensitive to outliers and heavy-tailed data. A robust alternative can be obtained using the next theorem, which states that for elliptical random elements, the range of $T_{XX|Y}$ is also contained in $\text{span}\{T_{XX}\beta_1, \ldots, T_{XX}\beta_K\}$. Thus, we can use $T_{XX|Y}$ to provide an efficient and robust estimate of the central subspace.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Theorem 3. *Let X be an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$, satisfying Assumption 1. Then, $T_{XX|Y}$ and $\Sigma_{XX|Y}$ have the same eigenfunctions.*

Under Assumption 1, Theorems 2 and 3 imply that the range of $T_{XX|Y}$ is contained in $\text{span}\{T_{XX}\beta_1, \ldots, T_{XX}\beta_K\}$. Hence, the K eigenfunctions of $T_{XX}^{\dagger}T_{XX|Y}$ corresponding to the K nonzero largest eigenvalues generate a subspace that is contained in the central subspace $\mathcal{S}_{Y|X}$. Here, $T_{XX}^{\dagger}$ denotes the Moore-Penrose inverse of T_{XX} which is defined in the next paragraph. Recall, the goal of sufficient dimension reduction is not about estimating the directions $\beta_1, \ldots, \beta_K$, but rather estimating the central subspace spanned by $\beta_1, \ldots, \beta_K$. Hence, with a little abuse of notation, we will use $\beta_1, \ldots, \beta_K$ to denote the T_{XX}-orthonormal eigenfunctions of $T_{XX|Y}$. To get those eigenfunctions, it is convenient to determine the eigenfunctions $\eta_1, \ldots, \eta_K$, of $T_{XX}^{\dagger\frac{1}{2}}T_{XX|Y}T_{XX}^{\dagger\frac{1}{2}}$ and to use $\beta_\ell = T_{XX}^{\dagger\frac{1}{2}}\eta_\ell$, $\ell = 1, \ldots, K$, leading to the following eigenvalue problem

$$\begin{aligned} &\text{maximize} \quad \langle \eta, T_{XX}^{\dagger\frac{1}{2}}T_{XX|Y}T_{XX}^{\dagger\frac{1}{2}}\eta\rangle_{\oplus\mathscr{H}} \\ &\text{subject to} \quad \eta \in \oplus_{i=1}^{p}\mathscr{H}_i,\ \langle \eta, \eta\rangle_{\oplus\mathscr{H}} = 1,\ \langle \eta, \eta_\ell\rangle_{\oplus\mathscr{H}} = 0, \ell = 1, \ldots, K-1. \end{aligned} \tag{3.2}$$

The functions $\eta_1, \ldots, \eta_K$ generate a subspace of the central subspace $\mathcal{S}_{Y|X}$. We refer to any sample estimator targeting the central subspace as R-FSIR. In the infinite-dimensional setting, T_{XX} is not necessarily invertible since it

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

is a compact operator. However, following He et al. (2003), we can define the inverse of the restricted operator of T_{XX}. Let $\mathcal{R}_{T_{XX}^{1/2}}$ be the range of $T_{XX}^{\frac{1}{2}}$, which is characterized by $\mathcal{R}_{T_{XX}^{1/2}} = \{f \in \oplus_{i=1}^{p}\mathscr{H}_i : \sum_{i=1}^{\infty}\delta_i^{-1}|\langle f, \psi_i\rangle_{\oplus\mathscr{H}}|^2 < \infty, f \in \ker(T_{XX})\}$, where $\ker(T_{XX})$ denotes the kernel of the operator T_{XX}. Define $\mathcal{R}^{-1}_{T_{XX}^{1/2}} = \{h \in \oplus_{i=1}^{p}\mathscr{H}_i : h = \sum_{r=1}^{\infty}\delta_i^{-1/2}\langle f, \psi_r\rangle_{\oplus\mathscr{H}}\psi_r, f \in \mathcal{R}_{T_{XX}^{1/2}}\} \subset \oplus_{i=1}^{p}\mathscr{H}_i$. Then, the restricted operator $\tilde{T}_{XX}^{\frac{1}{2}} = T_{XX}|_{\mathcal{R}^{-1}_{T_{XX}^{1/2}}}$ is a one-to-one mapping from $\mathcal{R}^{-1}_{T_{XX}^{1/2}}$ to $\mathcal{R}_{T_{XX}^{1/2}}$. We call the inverse of this restricted operator the Moore-Penrose inverse of $T_{XX}^{\frac{1}{2}}$ and denote it as $T_{XX}^{\dagger\frac{1}{2}}$. Thus, $T_{XX}^{\dagger\frac{1}{2}} : \mathcal{R}_{T_{XX}^{1/2}} \mapsto \mathcal{R}^{-1}_{T_{XX}^{1/2}}$ and $T_{XX}^{\dagger\frac{1}{2}} = \sum_{r=1}^{\infty}\delta_r^{-1/2}\psi_r \otimes \psi_r$. Moreover, $T_{XX}^{\dagger\frac{1}{2}}$ satisfies the usual properties of an inverse in the sense that $T_{XX}^{\frac{1}{2}}T_{XX}^{\dagger\frac{1}{2}}f = f$, for all $f \in \mathcal{R}_{T_{XX}^{1/2}}$ and $T_{XX}^{\dagger\frac{1}{2}}T_{XX}^{\frac{1}{2}}g = g$, for all $g \in \mathcal{R}^{-1}_{T_{XX}^{1/2}}$. Note that we do not assume that $T_{XX}^{\dagger\frac{1}{2}}$ is continuous, which would be a strong assumption since $T_{XX}^{\frac{1}{2}}$ is a trace-class operator whose eigenvalues tend to zero. The following assumption guarantees that the operator $T_{XX}^{\dagger\frac{1}{2}}T_{XX|Y}T_{XX}^{\dagger\frac{1}{2}}$ is a well-defined Hilbert-Schmidt operator as well as that the eigenfunctions of $T_{XX}^{\dagger\frac{1}{2}}T_{XX|Y}T_{XX}^{\dagger\frac{1}{2}}$ are well-defined in $\oplus_{i=1}^{p}\mathscr{H}_i$.

Assumption 3. $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_i^{-2}\delta_j^{-1}\mathbb{E}^2\left[\frac{\{\mathbb{E}(\rho_i^*|Y)-\mathbb{E}(\rho_i^*|\tilde{Y})\}\{\mathbb{E}(\rho_j^*|Y)-\mathbb{E}(\rho_j^*|\tilde{Y})\}}{\sum_{r=1}^{\infty}\{\mathbb{E}(\rho_r^*|Y)-\mathbb{E}(\rho_r^*|\tilde{Y})\}^2}\right] < \infty$, where $\tilde{Y}$ is an independent copy of Y, $\{\delta_r\}_{r\geq 1}$ and ψ_r are the eigenvalues and eigenfunctions of T_{XX}, respectively and $\rho_r^* = \gamma_r^{1/2}\rho_r = \langle X - \mu_X, \psi_r\rangle_{\oplus\mathscr{H}}$.

Assumption 3 implies $\sum_{i=1}^{\infty}\delta_i^{-1}|\langle T_{XX|Y}T_{XX}^{\dagger\frac{1}{2}}u, \psi_i\rangle_{\oplus\mathscr{H}}|^2 < \infty$ for all $u \in$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

$\mathcal{R}_{T_{XX}^{1/2}}$. Thus, $|\langle T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} u, \psi_i \rangle_{\oplus\mathscr{H}}|^2$ must decay to 0 faster than δ_i as $i \to \infty$. Essentially, it implies that $T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ is a Hilbert-Schmidt operator, and can be interpreted as a type of smoothness assumption. It requires the range space of $T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ to be sufficiently focused on the eigenspaces of the large eigenvalues of $T_{XX}^{\frac{1}{2}}$. The following proposition is analogue to Theorem 2.1 of Ferré and Yao (2005) and to Theorem 4.8 of He et al. (2003).

Proposition 1. *Let X be an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$, satisfying Assumption 1. Then, under Assumption 3, the eigenfunctions $\eta_1, \ldots, \eta_K$ associated with the K positive eigenvalues of $T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ are well-defined.*

3.2 Sample level

In this section, we derive the sample estimate of the conditional spatial sign covariance operator when the functions are fully observed. For $u = 1, \ldots, n$, let Y_u be an independent and identically distributed (iid) sample from Y, and let $X_1, \ldots, X_n$ be an iid sample from the random element $X = (X^1, \ldots, X^p)$, such that $X_u = (X_u^1, \ldots, X_u^p)^\intercal$. Consider partitioning Y into slices $J_1, \ldots, J_H$. Then, for each slice h, estimate $\mu_{X|Y}(h)$ using $\hat{\mu}_{X|Y}(h) = \frac{\mathbb{E}_n\{XI(Y\in J_h)\}}{\mathbb{E}_n\{I(Y\in J_h)\}}, h = 1, \ldots, H$, where $\mathbb{E}_n(\cdot)$ denotes the sample mean and $I(\cdot)$ denotes the indicator function. Then, by (2.4) we estimate T_{XX}

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

using

$$\hat{T}_{XX} = \frac{2}{n(n-1)} \sum_{1 \le u < u' \le n} \frac{(X_u - X_{u'}) \otimes (X_u - X_{u'})^\intercal}{\|X_u - X_{u'}\|^2_{\oplus \mathscr{H}}}. \tag{3.3}$$

Note that $\hat{T}_{XX}$ is a U-statistic of order 2 with the kernel,

$$k(X_u, X_{u'}) = \frac{(X_u - X_{u'}) \otimes (X_u - X_{u'})^\intercal}{\|X_u - X_{u'}\|^2_{\oplus \mathscr{H}}}.$$

Moreover, $k(\cdot, \cdot)$ is bounded operator, i.e., $\|k\|_{\text{op}} \le 1$. Similarly, we can estimate $T_{XX|Y}$ as defined in (2.5) by

$$\hat{T}_{XX|Y} = \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \Big[\frac{\{\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\} \otimes \{\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\}^\intercal}{\|\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\|^2_{\oplus \mathscr{H}}} \Big]. \tag{3.4}$$

4. Asymptotic theory

4.1 Convergence rates for fully observed random functions

We first establish the convergence rate of R-FSIR assuming the random functions X^i, $i = 1, \ldots, p$, are fully observed. The following Theorem is due to Zhong et al. (2022) (Theorem 2.1).

Theorem 4. *Let $\hat{T}_{XX}$ be as defined in (3.3). Then, $\|\hat{T}_{XX} - T_{XX}\|_{\text{op}} = O_p(n^{-1/2})$.*

For the iid sample $\{(Y_u, X_u) : u = 1, \ldots, n\}$, denote the order statistics $\{(Y_{(u)}, X_{(u)}) : u = 1, \ldots, n\}$ where $Y_{(1)} \le Y_{(2)} \ldots \le Y_{(n)}$, are called the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

concomitant of $Y_{(u)}$ (Yang, 1977). Then, following Chen et al. (2022), we introduce the following double subscripts, where the first subscript refers to the slice number and the second subscript refers to the order number of an observation in the given slice, that is, $X_{hi} = X_{(\ell(h-1)+i)}$ and $Y_{hi} = Y_{(\ell(h-1)+i)}$, $i = 1, \ldots, \ell, h = 1, \ldots, H$, where $\ell = [n/H]$ is the number of observations in a given slice. Define $m_k(Y) = \mathbb{E}(X^k|Y)$, $k = 1, \ldots, p$, and $\mathbf{m}(Y) = \mathbb{E}(X|Y) = (\mathbb{E}(X^1|Y), \ldots, \mathbb{E}(X^p|Y))^\intercal$. Next, to derive the convergence rate of $\hat{T}_{XX|Y}$, we need the following smoothness condition on the inverse regression curve, which was also used in Hsing and Carroll (1992) and Zhu and Ng (1995).

Assumption 4. Let $\prod_n(B)$ be the collection of all n-points partitions $-B \le y_{(1)} \le \ldots \le y_{(n)} \le B$ of the closed interval $[-B, B]$, where $B > 0$ and $n \ge 1$. Any function $\mathbf{m}(y) : \mathbb{R} \mapsto \oplus_{i=1}^p \mathscr{H}_i$ is said to have a total variation of order r if for any fixed $B > 0$, $\lim_{n\to\infty} n^{-r} \sup_{\prod_n(B)} \sum_{i=1}^{n-1} \|\mathbf{m}(y_{(i+1)}) - \mathbf{m}(y_{(i)})\|_{\oplus\mathscr{H}} = 0$.

Assumption 4 holds for any $r > 1$ if $\mathbf{m}$ is a continuous function.

Theorem 5. *Assume* $\mathbb{E}\|X\|^4_{\oplus\mathscr{H}} < \infty$, *Assumption 4 holds with* $r = 4$, *and there exists a nondecreasing real-valued function* $\tilde{M}$ *and* $B_0 > 0$ *such that for any two points* y_1 *and* y_2 *both in* $(-\infty, B_0]$ *or* $[B_0, \infty)$, $\|\mathbf{m}(y_1) - \mathbf{m}(y_2)\|_{\oplus\mathscr{H}} \le |\tilde{M}(y_1) - \tilde{M}(y_2)|$ *and* $\tilde{M}^4(u)\mathbb{P}(Y > u) \to 0$ *as* $t \to \infty, n \to \infty$. *Then,*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

$\|\hat{T}_{XX|Y} - T_{XX|Y}\|_{\text{op}} = O_p(n^{-1/2})$, *where* $\hat{T}_{XX|Y}$ *and* $T_{XX|Y}$ *are defined in (3.4) and (2.5), respectively.*

Assumption 5. There exists a bounded and self-adjoint operator $D_{XX} : \oplus_{i=1}^{p}\mathscr{H}_i \mapsto \oplus_{i=1}^{p}\mathscr{H}_i$ and $\beta > 0$, such that $T_{XX|Y} = T_{XX}^{1+\beta} D_{XX} T_{XX}^{1+\beta}$.

Assumption 5 requires $T_{XX}^{-1-\beta} T_{XX|Y}$ to be a bounded operator, and that

$$\text{ran}(T_{XX|Y}) \subseteq \text{ran}(T_{XX}^{1+\beta}), \tag{4.1}$$

where $\text{ran}(T_{XX}^{1+\beta})$ is the range of the operator $T_{XX}^{1+\beta}$, characterized by $\text{ran}(T_{XX}^{1+\beta}) = \left\{ f \in \oplus_{i=1}^{p}\mathscr{H}_i : \sum_{i=1}^{\infty} \frac{|\langle f, \psi_i\rangle_{\oplus\mathscr{H}}|^2}{\delta_i^{2(1+\beta)}} < \infty, f \perp \ker(T_{XX}) \right\}$, and $(\delta_1, \psi_1), (\delta_2, \psi_2), \ldots,$ are the eigenvalue-eigenfunctions pairs of T_{XX}. Note that, condition (4.1) requires that $\sum_{i=1}^{\infty} \frac{|\langle T_{XX|Y} f, \psi_i\rangle_{\oplus\mathscr{H}}|^2}{\delta_i^{2(1+\beta)}} < \infty$, for all $f \in \text{ran}(T_{XX}^{1+\beta})$. Therefore, in order for condition (4.1) to hold, $\langle T_{XX|Y} f, \psi_i\rangle_{\oplus\mathscr{H}}^2$ must decay faster than $\delta_i^{2(1+\beta)}$ as $i \to \infty$. This means that the operator $T_{XX|Y}$ sends any incoming function $f \in \text{ran}(T_{XX}^{1+\beta})$ to the eigen-spaces of T_{XX} corresponding to the large eigenvalues, or to the low-frequency components of T_{XX}. Moreover, the degree of concentration increases as β increases. For that reason, Assumption 5 can be interpreted as a type of smoothness, where β characterizes the degree of smoothness. Moreover, Assumption 5 allows to derive convergence rates of certain operators beyond consistency, when dealing with compact operators which are noninvertible in the infinite-dimensional case. Finally, in the sim-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

ulations setting, we approximate the functions through a few coefficients, and hence Assumption 5 holds since all operators involved are finite-rank operators. See also Li (2018), for more details. For simplicity, we use the notation $M = T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ and $\hat{M}^{(\epsilon_n)} = (\hat{T}_{XX} + \epsilon_n I)^{-\frac{1}{2}} \hat{T}_{XX|Y} (\hat{T}_{XX} + \epsilon_n I)^{-\frac{1}{2}}$, where $(\epsilon_n)_{n\in\mathbb{N}}$ is a sequence of positive numbers such that $\epsilon_n \to 0$ as $n \to \infty$.

Theorem 6. *Let X be an elliptical random element in $\oplus_{i=1}^{p} \mathscr{H}_i$, satisfying $\mathbb{E}\|X\|_{\oplus\mathscr{H}}^4 < \infty$. Then, under Assumptions 3, 4, and 5 and $n^{-2/5} \prec \epsilon_n \prec 1$, we have $\|\hat{M}^{(\epsilon_n)} - M\|_{\text{op}} = O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta)})$.*

Let $(\zeta_1, \eta_1), (\zeta_2, \eta_2) \ldots, (\zeta_K, \eta_K)$ and $(\hat{\zeta}_1, \hat{\eta}_1), (\hat{\zeta}_2, \hat{\eta}_2) \ldots, (\hat{\zeta}_K, \hat{\eta}_K)$ be the first K eigenvalues-eigenfunction pairs of M and $\hat{M}^{(\epsilon_n)}$, respectively. The following corollary provides the rates of convergence for the eigenvalues $\hat{\zeta}_1, \ldots, \hat{\zeta}_K$ and eigenfunctions $\hat{\eta}_1, \ldots, \hat{\eta}_K$, which are the same as the ones in Theorem 6. The corollary follows by applying perturbation theory for linear operators (Koltchinskii and Giné (2000), Kato (2013), Chapter VIII).

Corollary 1. *Under the assumptions of Theorem 6 and the assumption that the nonzero eigenvalues of M are distinct, we have for $k = 1, \ldots, K$,*

$$\|\hat{\eta}_k - \eta_k\|_{\oplus\mathscr{H}} = O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta)}),$$

$$|\hat{\zeta}_k - \zeta_k| = O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta)}).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Next, we derive the convergence rates of the transformed eigenfunctions $\hat{\beta}_k = (\hat{T}_{XX} + \epsilon I)^{-1/2}\hat{\eta}_k$ and the sufficient predictors $\langle \hat{\beta}_k, X\rangle_{\oplus\mathscr{H}}$ $k = 1, \ldots, K$.

Theorem 7. *Let X be an elliptical random element in $\oplus_{i=1}^{p}\mathscr{H}_i$, satisfying $\mathbb{E}\|X\|^4_{\oplus\mathscr{H}} < \infty$. Then, under Assumptions 3, 4, and 5, $\beta > 1$, and $n^{-1/3} \prec \epsilon_n \prec 1$, we have for $k = 1, \ldots, K$,*

$$\|\hat{\beta}_k - \beta_k\|_{\oplus\mathscr{H}} = O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}),$$

$$|\langle \hat{\beta}_k, X\rangle_{\oplus\mathscr{H}} - \langle \beta_k, X\rangle_{\oplus\mathscr{H}}| = O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}).$$

4.2 Convergence rates for partially observed random functions

We next derive the convergence rate under the scenario that each random function X^i, $i = 1, \ldots, p$, is only partially observed. Partially observed functions can only be observed at a *measurement schedule* and must be estimated from observed values at the sampled time points. Measurement schedules are classified as 'dense' if the covariance operator $\Sigma_{X^iX^i}$ can be estimated at $n^{-1/2}$ rate; otherwise they are classified as 'sparse'. See Yao et al. (2005) for more information about the measurement schedule.

For each $u = 1, \ldots, n$, let $\hat{X}_u(t) = (\hat{X}_u^1(t), \ldots, \hat{X}_u^p(t))^\intercal$ be the estimated

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

functions. We then estimate

$$\hat{\mu}_{X|Y}(h) = \frac{\mathbb{E}_n\{\hat{X} I(Y \in J_h)\}}{\mathbb{E}_n\{I(Y \in J_h)\}}, \quad h = 1, \ldots, H,$$

$$\hat{T}_{XX} = \frac{2}{n(n-1)} \sum_{1 \le u < u' \le n} \frac{(\hat{X}_u - \hat{X}_{u'}) \otimes (\hat{X}_u - \hat{X}_{u'})^{\mathsf{T}}}{\|\hat{X}_u - \hat{X}_{u'}\|^2_{\oplus \mathscr{H}}},$$

$$\hat{T}_{XX|Y} = \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \Big[\frac{\{\hat{\mu}_{\hat{X}|Y}(h) - \hat{\mu}_{\hat{X}|Y}(h')\} \otimes \{\hat{\mu}_{\hat{X}|Y}(h) - \hat{\mu}_{\hat{X}|Y}(h')\}^{\mathsf{T}}}{\|\hat{\mu}_{\hat{X}|Y}(h) - \hat{\mu}_{\hat{X}|Y}(h')\|^2_{\oplus \mathscr{H}}} \Big].$$

Theorems 5 and 6 show that the convergence rates depend on the rate of convergence of $\|\hat{T}_{XX} - T_{XX}\|_{\text{op}}$. This rate is equal to $n^{-1/2}$ when the functions are completely observed. However, when the functions are partially observed, we may assume that T_{XX} is estimated at an arbitrary rate $n^{-\gamma}$ such that $n^{-1/2} \preceq n^{-\gamma} \prec 1$ with $0 < \gamma \le 1/2$. The denser the measurement schedule, the closer γ is to $1/2$. The next theorem extends Theorem 6 and Theorem 7, and takes into account the effect of the measurement schedule.

Theorem 8. *Let X be an elliptical random element in $\oplus_{i=1}^{p} \mathscr{H}_i$, satisfying $\mathbb{E}\|X\|^4_{\oplus \mathscr{H}} < \infty$. Under Assumptions 3, 4, 5, the assumption that there exists $0 < \gamma \le 1/2$ such that $\|\hat{T}_{XX} - T_{XX}\|_{\text{op}} = O_p(n^{-\gamma})$, $\beta > 1$, and $n^{-2\gamma/3} \prec \epsilon_n \prec 1$, we have*

$$\|\hat{M}^{(\epsilon_n)} - M\|_{\text{op}} = O_p(n^{-2\gamma}\epsilon_n^{-5/2} + n^{-\gamma}\epsilon_n^{-1} + \epsilon_n),$$

$$\|\hat{\beta}_k - \beta_k\|_{\oplus \mathscr{H}} = O_p(n^{-2\gamma}\epsilon_n^{-5/2} + n^{-\gamma}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}),$$

$$|\langle \hat{\beta}_k, X \rangle_{\oplus \mathscr{H}} - \langle \beta_k, X \rangle_{\oplus \mathscr{H}}| = O_p(n^{-2\gamma}\epsilon_n^{-5/2} + n^{-\gamma}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

5. Implementation

In Section 3.2, we described the estimation procedures at the operator level. To implement the estimation procedure, we need to represent operators as matrices through coordinating mapping. To save space we leave the full details of the development of the coordinate mapping in the supplementary appendix, while only present the final results here.

5.1 Algorithm for R-FSIR

1. For each $i = 1, \ldots, p$, choose a finite set of functions $\{g_1^i, \ldots, g_{k_n}^i\}$ such that $\text{span}\{g_1^i, \ldots, g_{k_n}^i\} = \mathscr{H}_i$.
2. Obtain the Gram matrix $\mathbb{K}_i$ and its centered version $Q_{k_n}\mathbb{K}_i Q_{k_n} \equiv G_i$.
3. For each $X_u, u = 1, \ldots, n$, calculate the centered version $X_u - \mathbb{E}_n X_u$.
4. Compute the coordinate $[X_u]$ relative to the basis $\mathscr{B}$ of $\oplus_{i=1}^p \mathscr{H}_i$, and derive the gram matrix $G = Q_{k_n}\mathbb{K}Q_{k_n}$ of the basis $\mathscr{B}$.
5. Divide the range of Y into H equal slices, $J_1, \ldots, J_H$.
6. For each $h = 1, \ldots, H$, compute $[\hat{\mu}_{X|Y}(h)]$ according to
$$[\hat{\mu}_{X|Y}(h)] = \frac{[\mathbb{E}_n\{XI(Y \in J_h)\}]}{\mathbb{E}_n\{I(Y \in J_h)\}} = \frac{\mathbb{E}_n\{[X]I(Y \in J_h)\}}{\mathbb{E}_n\{I(Y \in J_h)\}}.$$
7. Compute the matrices Ω and Λ, defined as
$$\Lambda = \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \left\{ \frac{[\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')][\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')]^\intercal}{[\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')]^\intercal G [\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')]} \right\}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

$$\Omega = \frac{2}{n(n-1)} \sum_{1 \le u < u' \le n} \frac{[X_u - X_{u'}][X_u - X_{u'}]^\tau}{[X_u - X_{u'}]^\tau G [X_u - X_{u'}]}.$$

8. Compute the matrix $M\Lambda M^\tau$ and its K eigenvectors, $v_1, \ldots, v_K$, where $M = (G^{\frac{1}{2}} \Omega G^{\frac{1}{2}})^{\dagger \frac{1}{2}} G^{\frac{1}{2}}$.

9. Obtain the eigenfunctions $\hat{\beta}_\ell, \ell = 1, \ldots, K$ by $\hat{\beta}_\ell = v_\ell^\tau (G^{\frac{1}{2}} \Omega G^{\frac{1}{2}})^{\dagger \frac{1}{2}} G^{\dagger \frac{1}{2}} b$, $\ell = 1, \ldots, K$, where $b = (b^{1\top}, \ldots, b^{p\top})^\top$ and $b^i = (b^i_1, \ldots, b^i_{k_n})$.

We determine the dimension K of the central subspace using the CVBIC criterion introduced by Li et al. (2011), with full details developed in the supplementary material in Section S2.4.

6. Simulation Studies

The finite sample performance of the proposed methodology is illustrated through simulation examples. We use R-FSIR to denote the proposed rank-based FSIR method, and we compare it with FSIR (Ferré and Yao, 2003).

The simulation setting is an extension of Wang et al. (2022) to multivariate functional data and is as follows. For $j = 1, \ldots, p$, assume $\mu_{X^j} = 0$ and generate $X^j_u(t)$ from the univariate Karhunen-Loève expansion (2.2), that is, $X^j_u(t) = \sum_{q=1}^4 \xi^j_{uq} \phi^j_q(t)$, $t \in [0,1]$, $u = 1, \ldots, n$, where $\phi^j_1(t) = \sqrt{2}\sin(2\pi t)$, $\phi^j_2(t) = \sqrt{2}\cos(2\pi t)$, $\phi^j_3(t) = \sqrt{2}\sin(4\pi t)$ and $\phi^j_4(t) = \sqrt{2}\cos(4\pi t)$, and ξ^j_{uq} are mutually independent random variables with zero mean and variance

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

$\text{var}(\xi_{uq}^j) = \lambda_q$, $q = 1, \ldots, 4$ with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)^\top = (2, 1, 1/2, 1/4)^\top$.

To simulate the multivariate functional data $X_u(t) = (X_u^1(t), \ldots, X_u^p(t))$ from the Karhunen-Loève expansion (2.3), we obtain the multivariate FPCA eigenfunctions through an orthogonalization of the univariate eigenfunctions, as described in the Proposition 5 of Happ and Greven (2018). Specifically, let $\boldsymbol{\xi}_u = ((\boldsymbol{\xi}_u^1)^\top, \ldots, (\boldsymbol{\xi}_u^p)^\top)^\top \in \mathbb{R}^{4p}$, where $\boldsymbol{\xi}_u^j = (\xi_{u1}^j, \ldots, \xi_{u4}^j)^\top$, $j = 1, \ldots, p$, $u = 1, \ldots, n$, and let $Z \in \mathbb{R}^{4p \times 4p}$ be the covariance matrix of the univariate FPCA scores $\boldsymbol{\xi}_u$ whose (j,k)th entry is the matrix $Z^{jk} = \text{cov}(\boldsymbol{\xi}_u^j, \boldsymbol{\xi}_u^k) \in \mathbb{R}^{4\times 4}$. Then, the kth eigenfunction of Σ_{XX}, denoted by $\boldsymbol{\psi}_k(t) = (\psi_k^1(t), \ldots, \psi_k^p(t))^\top$, is given by $\psi_k^j(t) = \boldsymbol{\phi}^j(t)^\top \boldsymbol{v}_k^j$, $k = 1, \ldots, 12$, where $\boldsymbol{\phi}^j(t) = (\phi_1^j(t), \ldots, \phi_4^j(t))^\top$ and $\boldsymbol{v}_k^j = (v_{k1}^j, \ldots, v_{k4}^j)^\top \in \mathbb{R}^4$ denotes the jth block of the eigenvector $\boldsymbol{v}_k$ of Z. Then, the scores ρ_{uk} are obtained by $\rho_{uk} = \sum_{j=1}^p \sum_{q=1}^4 v_{kq}^j \xi_{uq}^j$, $k = 1, \ldots, 12$, $u = 1, \ldots, n$, where we assume that coordinate-wise scores ξ_{uq}^j are simulated from the following distributions: 1) Gaussian 2) multivariate Student-t distribution with two and three degrees of freedom and 3) Cauchy distribution. Moreover, to evaluate the robustness of the method, for each $j = 1, \ldots, p$ and each $q = 1, \ldots, 4$, we randomly select m out of the n simulated coordinate-wise scores ξ_{uq}^j and add a shift of +5 or -5 in an alternating way. We estimate each function X_u^j using 4 cubic piecewise polynomials.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

6.1 Results

We simulate n independent copies of the response Y from each of the following single- and double-index models

$$\text{Model I:} \quad Y = \frac{1}{0.5 + (\langle \beta_1, X \rangle_{\oplus \mathscr{H}} + 1)^2} + 0.2\epsilon,$$

$$\text{Model II:} \quad Y = \sin(\pi \langle \beta_1, X \rangle_{\oplus \mathscr{H}} / 4) + 0.5\epsilon,$$

$$\text{Model III:} \quad Y = \arctan(\pi \langle \beta_1, X \rangle_{\oplus \mathscr{H}} / 2) + \epsilon,$$

$$\text{Model IV:} \quad Y = \arctan(\pi \langle \beta_1, X \rangle_{\oplus \mathscr{H}}) + 0.5 \sin(\pi \langle \beta_2, X \rangle_{\oplus \mathscr{H}} / 6) + 0.1\epsilon,$$

$$\text{Model V:} \quad Y = \frac{\langle \beta_1, X \rangle_{\oplus \mathscr{H}}}{0.5 + (\langle \beta_2, X \rangle_{\oplus \mathscr{H}} + 1)^2} + 0.2\epsilon,$$

where $\boldsymbol{\beta}_1(t) = \boldsymbol{\psi}_1(t)$ and $\boldsymbol{\beta}_2(t) = \boldsymbol{\psi}_2(t)$ are the first and second eigenfunctions of Σ_{XX}, respectively, X is simulated as described in Section 6, assuming $\mu_{X^j}(t) = 0$ for all $j = 1, \ldots p$, and observed at 101 time points equally spaced in [0,1]. The error ϵ is generated according to a standard normal distribution. We use $n = 400$, $p = 5$, and $H = 10$ slices.

To evaluate the performance of each method, we use the multiple correlation between the true and estimated predictors, also considered in Li and Song (2022). Specifically, let U and V be random vectors of the same dimension and let C_{UU} , C_{UV} , and C_{VV} represent the sample covariance matrices. The multiple correlation between U and V is $\text{mcorr}(U, V) = \text{tr}(C_{VV}^{-1/2} C_{VU} C_{UU}^{-1} C_{UV} C_{VV}^{-1/2})$. This measure varies from 0 to K, and a value

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

close to K indicates better estimation accuracy.

Table 1 shows the observed, over the 100 simulation runs, means and standard deviations (in parenthesis) of the multiple correlation between the true and the estimated predictors when no outliers are present (upper part) and when outliers are added as described in Section 6 (lower part). We expect the multiple correlation to be close to $K = 1$ for the single-index Models I - III and close to $K = 2$ for the double-index Models IV - V. We observe that R-FSIR and FSIR have comparable performance for the Gaussian distribution with no outliers. However, R-FSIR outperforms FSIR for heavy-tailed data. Specifically, the efficiency of R-FSIR remains reasonably high, whereas the efficiency of FSIR decreases considerably. This is especially evident when outliers are added to the data.

In Sections S3 and S4 of the supplementary appendix, we further investigate the performance of R-FSIR for a variety of combinations of (n, p, H) and the performance of the CVBIC order-determination criterion (S2.10).

7. Neuroimaging data application

To illustrate the performance of the methodology we use an fMRI dataset, obtained from the ADHD-200 Consortium (`http://fcon_1000.projects.nitrc.org/indi/adhd200/index.html`), consisting of resting-state fMRI

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Table 1: Mean (and standard deviation) of the multiple correlation with no outliers (upper) and with outliers added (lower) for Study 1

number of outliers	Models	Gaussian		$t(3)$		$t(2)$		Cauchy	
		FSIR	R-FSIR	FSIR	R-FSIR	FSIR	R-FSIR	FSIR	R-FSIR
$m = 0$	I	0.96 (0.02)	0.96 (0.02)	0.60 (0.24)	0.82 (0.21)	0.37 (0.22)	0.85 (0.22)	0.22 (0.13)	0.84 (0.23)
	II	0.98 (0.001)	0.98 (0.001)	0.68 (0.17)	0.86 (0.17)	0.34 (0.23)	0.84 (0.20)	0.21 (0.14)	0.85 (0.23)
	III	0.98 (0.001)	0.98 (0.001)	0.88 (0.11)	0.99 (0.007)	0.69 (0.20)	0.99 (0.01)	0.30 (0.19)	0.98 (0.03)
	IV	1.91 (0.04)	1.93 (0.03)	1.45 (0.27)	1.81 (0.15)	0.91 (0.33)	1.69 (0.28)	0.33 (0.17)	1.65 (0.35)
	V	1.83 (0.07)	1.83 (0.09)	1.28 (0.28)	1.56 (0.26)	0.81 (0.26)	1.51 (0.31)	0.27 (0.14)	1.69 (0.30)
$m = 40$	I	0.24 (0.06)	0.75 (0.24)	0.26 (0.15)	0.50 (0.26)	0.26 (0.19)	0.48 (0.32)	0.23 (0.15)	0.78 (0.28)
	II	0.40 (0.1)	0.96 (0.02)	0.33 (0.14)	0.78 (0.15)	0.26 (0.16)	0.57 (0.30)	0.20 (0.14)	0.84 (0.23)
	III	0.82 (0.04)	0.99 (0.0005)	0.84 (0.07)	0.99 (0.02)	0.79 (0.17)	0.96 (0.05)	0.32 (0.18)	0.98 (0.03)
	IV	1.40 (0.09)	1.75 (0.11)	1.19 (0.20)	1.75 (0.17)	0.94 (0.27)	1.58 (0.29)	0.34 (0.15)	1.59 (0.34)
	V	1.57 (0.22)	1.55 (0.24)	1.18 (0.28)	1.52 (0.29)	0.84 (0.32)	1.47 (0.31)	0.26 (0.14)	1.62 (0.34)

and anatomical datasets of children with and without ADHD aggregated across 8 independent imaging sites. For our analysis, we consider the resting-state fMRI of the New York University Child Study Center. This dataset includes 222 subjects, of which 99 are the controls and the rest are diagnosed with ADHD. The ADHD group is further divided into the ADHD Combined group (77 subjects), the ADHD Inattentive group (44 subjects) and the ADHD Hyperactive group (2 subjects); we use the 77 subjects in the ADHD Combined group for our analysis. Moreover, 5 subjects were removed from the ADHD Combined group because of significant amount of missing observations, resulting in $n = 72$ subjects. Technical

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

details regarding the sample and the scanning parameters can be found at the ADHD-200 Consortium.

The dataset was preprocessed by the NeuroBureau community using the Athena pipeline. 116 brain regions-of-interest (ROI) were constructed for the preprocessed resting-state fMRI using the anatomical labelling atlas (AAL) developed by Craddock et al. (2012). fMRI time series were extracted for each of the 116 regions by averaging all voxels time series within each region at each time point, resulting in 172 time points for each of the 116 regions for each subject. Hence, for each subject we have 116 different regional fMRI time series, observed at 172 time points. The AAL atlas and the regional fMRI time series are publicly available at NITRC (www.nitrc.org). The aim is to determine the association between the ADHD index with the brain activities measured by the fMRI. To simplify the model, we use the results obtained in Mahzarnia and Song (2022) and choose the 42 regions out of the 116 most related with the ADHD index; the list of the these regions can be found in Mahzarnia and Song (2022). In Section S5 of the supplementary appendix, we provide the smoothed spline fMRI curves with outliers and boxplots of the first two principal components for two regions of interest, showing that the marginal distributions are heavy-tailed.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

Next, we apply R-FSIR using seven number of slices and 15 number of basis. Moreover, we use the CVBIC order-determination criterion defined in (S2.10) and conclude that $\hat{K} = 5$. Then, we apply R-FSIR and FSIR to obtain the five sufficient predictors $\langle X, \hat{\beta}_j \rangle_{\oplus \mathscr{H}}$. In order to compare the performance of the two methods, we apply a generalized additive model (GAM) using the new sufficient predictors and calculate the root mean squared errors. The results are 2.059 and 2.488 for R-FSIR and FSIR, respectively. We note that we try different values for the number of slices and the number of basis, but the results did not change much. There is a general tendency for R-FSIR and FSIR to perform better with smaller number of slices, but this pattern was not strong to require more consideration.

Figure 1 illustrates pair-plots of the five sufficient predictors $\langle X, \hat{\beta}_j \rangle_{\oplus \mathscr{H}}$, $j = 1, \ldots, 5$, where the different colors denote different values of the response variable Y. Specifically, we define a new variable, Y_Q, to take values $1, \ldots, 7$, depending on which quantile an observation falls. For example, for $u = 1, \ldots, n$, $Y_{Qu} = 1$ if Y_u falls in the smallest $(1/7)100\%$ of the data, $Y_{Qu} = 2$ if Y_u falls in the $(2/7)100\%$ of the data, and so on. For interpretation purposes, we refer to the ADHD index score as very low-level ($Y_Q = 1$), low-level ($Y_Q = 2$), mild low-level ($Y_Q = 3$), middle-level ($Y_Q = 4$), mild high-level ($Y_Q = 5$), high-level ($Y_Q = 6$), very high-level

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

($Y_Q = 7$). The diagonal plots illustrate the estimated density curves of $\langle X, \hat{\beta}_j \rangle_{\oplus \mathscr{H}}$, $j = 1, \ldots, 5$, for each Y_Q. Observe that each sufficient predictor represents different groups of Y and that the five sufficient predictors capture the important characteristics of the conditional distribution $Y|X$. For example, the first sufficient predictor differentiates middle-level ADHD score from other levels. The second sufficient predictor differentiates low-level, mild high-level, and very high-level ADHD scores. The third predictor can separate very high-level and high-level scores. The combination of second and third predictors can separate high-level ADHD index scores (mild-high, high, very high), while the combination of second and fourth predictors can separate low-level scores (very low, low, mild-low). Finally, the fifth predictor separates low-level and high-level ADHD index scores.

Supplementary Material

The Supplementary Appendix contains the algorithms for the proposed method, all the proofs of the theoretical results, and additional simulations.

Acknowledgments

Christou's research work was supported by the National Science Foundation under Grant No. DMS-2213140. Song's research was supported by the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

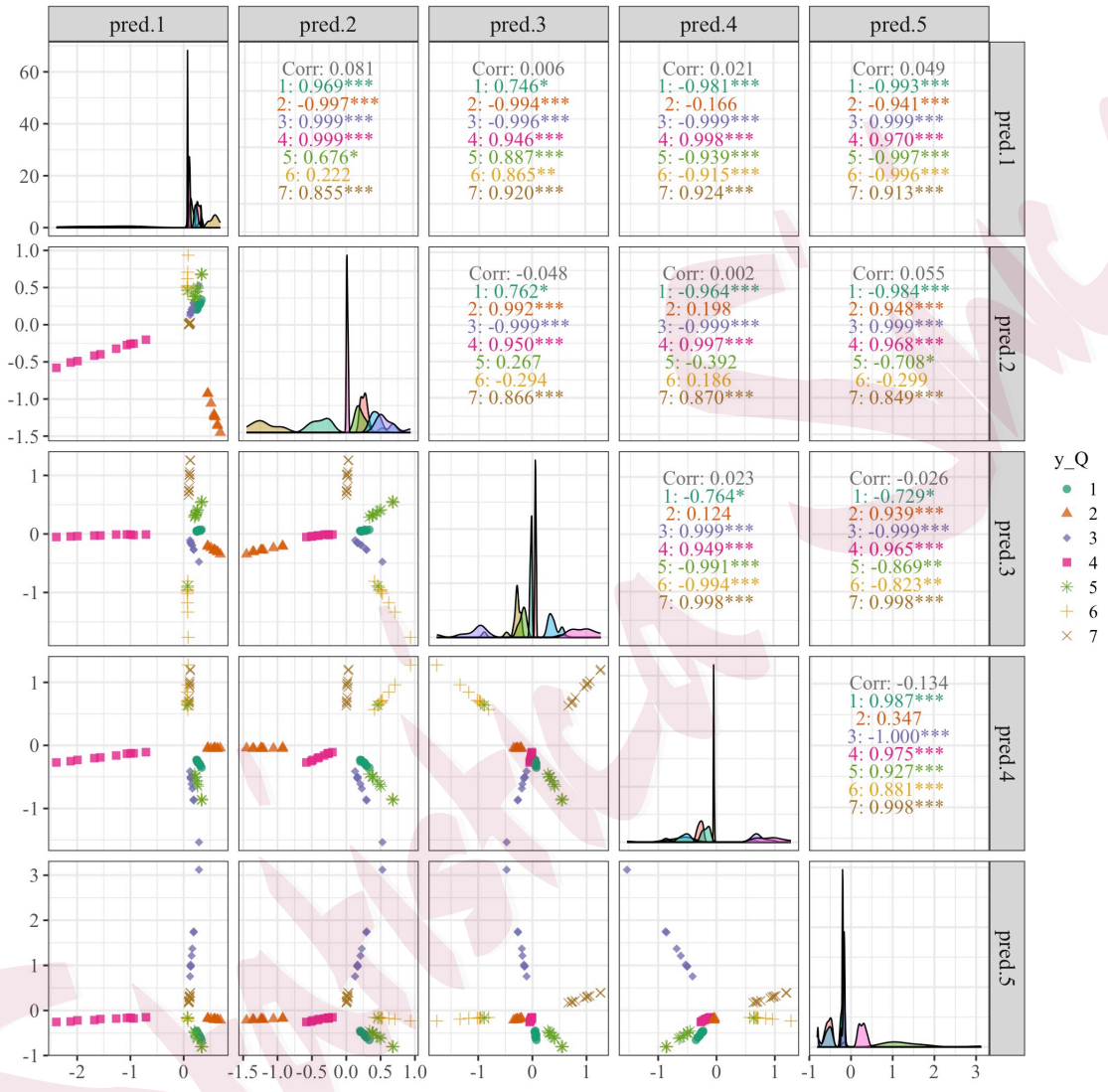

Figure 1: Pair plots of the R-FSIR sufficient predictors colored with the ADHD index score

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No. 2022R1C1C1003647, 2022M3J6A1063595, and No. RS-2023-00219212)

References

Bali, J. L., G. Boente, D. E. Tyler, and J.-L. Wang (2011). Robust functional principal components: A projection-pursuit approach. *The Annals of Statistics 39*(6), 2852 – 2882.

Boente, G., M. S. Barrera, and D. E. Tyler (2014). A characterization of elliptical distributions and some optimality properties of principal components for functional data. *Journal of Multivariate Analysis 131*, 254–264.

Boente, G., D. Rodriguez, and M. Sued (2019). The spatial sign covariance operator: Asymptotic results and applications. *Journal of Multivariate Analysis 170*, 115–128. Special Issue on Functional Data Analysis and Related Topics.

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*, Volume 149. Springer Science & Business Media.

Chen, X., J. Zhang, and W. Zhou (2022). High-dimensional elliptical sliced inverse regression in non-gaussian distributions. *Journal of Business & Economic Statistics 40*(3), 1204–1215.

Choi, K. and J. Marden (1998). A multivariate version of kendall's τ. *Journal of Nonparametric Statistics 9*(3), 261–293.

Craddock, R. C., G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg (2012). A

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping 33*(8), 1914–1928.

Croux, C., E. Ollila, and H. Oja (2002). Sign and rank covariance matrices: Statistical properties and application to principal components analysis. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Basel, pp. 257–269. Birkhäuser Basel.

Ferré, L. and A.-F. Yao (2003). Functional sliced inverse regression analysis. *Statistics 37*(6), 475–488.

Ferré, L. and A.-F. Yao (2005). Smoothed functional inverse regression. *Statistica Sinica 13*(3), 665–683.

Gervini, D. (2008, 09). Robust functional estimation using the median and spherical principal components. *Biometrika 95*(3), 587–600.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components. *Journal of the Royal Statistical Society, Series B 68*(1), 109–126.

Han, F. and H. Liu (2018). Eca: High-dimensional elliptical component analysis in non-gaussian distributions. *Journal of the American Statistical Association 113*(521), 252–268.

Happ, C. and S. Greven (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association 113*(522), 649–659.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

REFERENCES

He, G., H.-G. Müller, and J.-L. Wang (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis 85*(1), 54–77.

Hsing, T. and R. J. Carroll (1992). An asymptotic theory for sliced inverse regression. *The Annals of Statistics 20*(2), 1040–1061.

Jackson, D. A. and Y. Chen (2004). Robust principal component analysis and outlier detection with ecological data. *Environmetrics: The official journal of the International Environmetrics Society 15*(2), 129–139.

Kato, T. (2013). *Perturbation theory for linear operators*, Volume 132. Springer Science & Business Media.

Koltchinskii, V. and E. Giné (2000). Random matrix approximation of spectra of integral operators. *Bernoulli 6*(1), 113–167.

Kraus, D. and V. M. Panaretos (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika 99*(4), 813–832.

Li, B. (2018). Linear operator-based statistical analysis: A useful paradigm for big data. *Canadian Journal of Statistics 46*(1), 79–103.

Li, B., A. Artemiou, and L. Li (2011). Principal support vector machines for linear and nonlinear sufficient dimension reduction. *The Annals of Statistics 39*(6), 3182–3210.

Li, B. and J. Song (2022). Dimension reduction for functional data based on weak conditional moments. *The Annals of Statistics 50*(1), 107–128.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

REFERENCES

Li, G. and Z. Chen (1985). Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and monte carlo. *Journal of the American Statistical Association 80*(391), 759–766.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association 86*(414), 316–327.

Locantore, N., J. Marron, D. Simpson, N. Tripoli, J. Zhang, K. Cohen, G. Boente, R. Fraiman, B. Brumback, C. Croux, et al. (1999). Robust principal component analysis for functional data. *Test 8*(1), 1–73.

Mahzarnia, A. and J. Song (2022). Multivariate functional group sparse regression: Functional predictor selection. *PloS one 17*(4), e0265940.

Marden, J. I. (1999). Some robust estimates of principal components. *Statistics & probability letters 43*(4), 349–359.

Visuri, S., V. Koivunen, and H. Oja (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference 91*(2), 557–575.

Wang, G., S. Liu, F. Han, and C.-Z. Di (2022). Robust functional principal component analysis via a functional pairwise spatial sign operator. *Biometrics, https://doi.org/10.1111/biom.13695*.

Wang, J., J. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0341

REFERENCES

Yang, S.-S. (1977). General distribution theory of the concomitants of order statistics. *The Annals of Statistics 5*(5), 996–1002.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association 100*(470), 577–590.

Zhong, R., S. Liu, H. Li, and J. Zhang (2022). Functional principal component analysis estimator for non-gaussian data. *Journal of Statistical Computation and Simulation 92*(13), 1–14.

Zhu, L.-X. and K. W. Ng (1995). Asymptotics of sliced inverse regression. *Statistica Sinica 5*(2), 727–736.

Eftychia Solea

School of Mathematical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom

E-mail: e.solea@qmul.ac.uk

Eliana Christou

Department of Mathematics and Statistics, University of North Carolina at Charlotte 9201 University City Blvd, Charlotte, NC 28223, USA

E-mail: echris15@charlotte.edu

Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea

E-mail: junsong@korea.ac.kr