Journal of Multivariate Analysis 170 (2019) 263–274

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Spatial functional principal component analysis with applications to brain image data

Yingxing Li [a,*], Chen Huang [b], Wolfgang K. Härdle [c,d,a]

[a] The Wang Yanan Institute for Studies in Economics, Xiamen University, 422 Siming South Road, Xiamen 361005, China
[b] Fachbereich für Mathematik und Statistik, Universität St. Gallen, Bodanstraße 6, 9000 St. Gallen, Switzerland
[c] Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[d] Sim Kee Boon Institute for Financial Economics, Singapore Management University, 50 Stamford Road, Singapore 178899, Singapore

ARTICLE INFO

Article history:
Received 30 November 2017
Available online 19 November 2018

Keywords:
Asymptotics
Functional magnetic resonance imaging (fMRI)
Penalized smoothing
Principal component analysis

MSC 2010 subject classifications:
62H25
62P15

ABSTRACT

This paper considers a fast and effective algorithm for conducting functional principal component analysis with multivariate factors. Compared with the univariate case, our approach could be more powerful in revealing spatial connections or extracting important features in images. To facilitate fast computation, we connect singular value decomposition with penalized smoothing and avoid estimating a covariance operator in very high dimension. Under regularity assumptions, the results indicate that we may enjoy the optimal convergence rate by employing the smoothness assumption inherent to functional objects. We apply our method to the analysis of brain image data. Our extracted factors provide excellent recovery of the risk related regions of interest in the human brain and the estimated loadings are very informative in revealing individual risk attitude.

1. Introduction

As technology advances, the availability of complex and high-dimensional data has motivated an extensive growth of Object Oriented Data Analysis (OODA); see [21]. In the special case of functional data analysis (FDA), we regard each datum as a realization of a random object in an infinite-dimensional functional space and utilize the smoothness assumption to circumvent the difficulty in high-dimensional multivariate analysis. FDA has been popularized since the end of the 1990s, and leading references include [7,8,13,24,25]. As a major area in statistics, it includes a branch of related topics in either theoretical or empirical research, such as functional regression and classification, dimension reduction, global inference, spatial modeling, etc. We refer to [4,10] for surveys on recent development and trends. In this paper, we consider functional principal component analysis with particular application to spatial data or image.

Principal component analysis (PCA) and its functional version (FPCA) are widely used for dimension reduction. This method has been successfully applied in many fields including genetic studies, risk management, psychology, environmental studies, among others; see, e.g., [14,24]. The main idea is to summarize the data variation and information via some low-dimensional loadings, which are projections of the individual observations over some factors. In classical FPCA, factors are often used to capture the dynamics in time and are assumed to be smooth univariate functions of some time variable [6,28,32]. Nowadays, FPCA is very often used to analyze data sets that are measured at different locations or units. Menafoglio

* Corresponding author.
E-mail addresses: yxli@xmu.edu.cn (Y. Li), chen.huang@unisg.ch (C. Huang), haerdle@hu-berlin.de (W.K. Härdle).

https://doi.org/10.1016/j.jmva.2018.11.004
0047-259X/

and Petris [22] consider kriging for functional data through the notion of finite-dimensional approximation, but challenging questions remain on how to model spatial dependence nonparametrically, especially in the high-dimensional case.

Our motivating example is from a neuro-economics study using functional Magnetic Resonance Imaging (fMRI) data. In particular, multiple studies have concentrated on understanding the human perception to reward and risk [2,11]. Standard questions include how the brain responds and which region of the brain is activated in various decision-making tasks. For the former topic, there is an extensive literature about modeling the hemodynamic response function (HRF); see, e.g., [18,19,30]. Concerning the latter issue, one needs to identify the region of interest (ROI); see [27] and references therein. Although general linear models (GLM) are a popular tool to analyze Blood Oxygenation Level Dependent (BOLD) signals [9], they involve parametric assumptions. Moreover, they ignore the spatial patterns and interdependence of the BOLD signals as they model voxels separately. As an alternative, FPCA is becoming increasingly popular in the fMRI analysis as the extracted loadings could reflect individual variations while the factors could reveal spatial features.

The main challenge in conducting FPCA comes from handling the fMRI images that contain the BOLD signals captured on a 3-dimensional spatial brain map. In the study of van Bömmel et al. [1], each fMRI image comprises $91 \times 109 \times 91$ observations; different images were obtained at different times associated with different tasks or individuals. Such huge dimensions raise new challenges for statistical inference. A straightforward approach would be to vectorize the 3D image into a 1D vector of dimension 10^6, but this might implicitly break the spatial connections [12,33]. Chen et al. [3] propose to conduct FPCA in 3D, but their approach needs to estimate a 6D smooth covariance function, and the computation cost is quite high.

In this paper, we propose a general algorithm that could conduct FPCA for image data stored as arrays. Our contributions are summarized as follows. First, our approach could be viewed as a nonparametric multidimensional PCA. On one hand, we avoid vectorization and potential destruction of the spatial structure. On the other hand, the nonparametric approach guards us against severe model misspecification. Second, by utilizing penalized splines and singular value decomposition techniques, we propose a fast data-driven algorithm that could reduce computational burden without much scarification of estimation efficiency. Third, we develop asymptotic properties for our estimators. To the best of our knowledge, this is the first paper that discusses the theoretical properties of FPCA associated with a non-1D domain. Our settings embrace the scenarios with very large dimensional observations and a relatively small number of subjects. Under regularity assumptions, the results indicate that we may enjoy an optimal convergence rate by relying on a smoothness assumption. Our real data analysis also demonstrates the success of our approach.

The rest of the paper is organized in the following way. We introduce our estimation procedure and discuss its theoretical properties in Section 2. Section 3 conducts real data analysis and Section 4 concludes. All the proofs and technical details are deferred to the Appendix.

2. Methodology

2.1. Model and algorithm

Let $Y_{i,\boldsymbol{s}}$ denote the observed data for the ith subject at location $\boldsymbol{s} = (s_1, \ldots, s_d)$ in d-dimensional space. We focus on the regular dense setting, so repeated measurements are observed at some common location $\boldsymbol{s}$ that does not depend on i, and the total number of observations for each subject is $n = n_1 \times \cdots \times n_d$, which could be large. For example, in our fMRI data, $\boldsymbol{s} = (s_1, s_2, s_3)$ represents the 3D coordinates of each voxel in the brain with $s_1, s_3 \in \{1, \ldots, 91\}$ and $s_2 \in \{1, \ldots, 109\}$. All $Y_{i,\boldsymbol{s}} = Y_{i,(s_1,s_2,s_3)}$s form the observed fMRI image for the ith subject. The raw data $Y_{i,\boldsymbol{s}}$s are considered as a discrete sample, subject to measurement errors, of the continuous random process denoted as $X_i(\boldsymbol{s})$, defined in a bounded region $\mathcal{F} \subset \mathbb{R}^d$. We thus assume

$$Y_{i,\boldsymbol{s}} = X_i(\boldsymbol{s}) + \varepsilon_{i,\boldsymbol{s}},$$

where the error terms $\varepsilon_{i,\boldsymbol{s}}$ are random variables with mean 0 and variance $\sigma_i^2(\boldsymbol{s})$. They are also independent of the $X_i(\boldsymbol{s})$s.

Denote the mean function of all $X_i(\boldsymbol{s})$ as $\mu(\boldsymbol{s})$. According to the Karhunen–Loève (KL) decomposition [15,20], we can represent each $X_i(\boldsymbol{s})$ as

$$X_i(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \sum_{j=1}^{\infty} \psi_{ij}\phi_j(\boldsymbol{s}), \tag{1}$$

where $\phi_j(\boldsymbol{s})$ and ψ_{ij} are the jth factor and its loading, respectively; here, the ψ_{ij}s are uncorrelated random variables with mean 0 and variance γ_j, while $\int \phi_j(\boldsymbol{s})\phi_{j'}(\boldsymbol{s})d\boldsymbol{s} = 0$ for $j \neq j'$ and $\int \phi_j^2(\boldsymbol{s})d\boldsymbol{s} = 1$.

By analogy with the idea of PCA, it is natural to align all $\phi_j(\boldsymbol{s})$s into a long vector $\boldsymbol{\phi}_j$, all $\mu(\boldsymbol{s})$s into a long vector $\boldsymbol{\mu}$, all $X_i(\boldsymbol{s})$s and $Y_{i,\boldsymbol{s}}$s into $\mathbf{X}_i$ and $\mathbf{Y}_i$, respectively. Note that $\mathrm{cov}(\mathbf{Y}_i, \mathbf{Y}_i) \approx \Phi\Gamma\Phi^\top + V_\varepsilon$, where Γ is a diagonal matrix whose (j, j)th element is γ_j, the jth column of Φ is $\boldsymbol{\phi}_j$, and V_ε is a diagonal matrix whose diagonal elements form the same vector as the one that aligns all $\sum_{i=1}^{I} \sigma_i^2(\boldsymbol{s})/I$. Hence the traditional approach will first estimate the sample covariance of $\mathrm{cov}(\mathbf{Y}_i, \mathbf{Y}_i)$, then ignore the diagonal elements, smooth the covariance and then perform eigendecomposition to estimate Φ, or equivalently, the $\boldsymbol{\phi}_j$s. However, such an approach might not work very well under a dense setting, as it requires to smooth a covariance matrix of the dimension $n \times n$, where n might be of huge volume.

To circumvent the computation difficulty, we propose to estimate the loading first. Note that $V_x = \langle \mathbf{X}_i - \boldsymbol{\mu}, \mathbf{X}_{i'} - \boldsymbol{\mu} \rangle / n$ converges to $\Psi\Psi^\top$, where $\langle a, b \rangle = a_1 b_1 + \cdots + a_n b_n$ for $a, b \in \mathbb{R}^n$, and Ψ is the loadings matrix whose (i, j)th element is ψ_{ij}. Recall the fact that ψ_{ij} are uncorrelated distributed with mean 0 and variance γ_j. If we are able to estimate V_x and obtain its eigendecomposition as $\widehat{V}_x = \widehat{P}\widehat{\Gamma}\widehat{P}^\top$, then we could estimate the loadings matrix Ψ by $\widehat{P}\widehat{\Gamma}^{1/2}$.

Since we only observe $Y_{i,\boldsymbol{s}}$ instead of $X_i(\boldsymbol{s})$, we compute the inner product matrix $V_y = \langle \mathbf{Y}_i - \boldsymbol{\mu}, \mathbf{Y}_{i'} - \boldsymbol{\mu} \rangle / n$ instead. Given the assumed independence between the factors and the error terms, $V_y \approx \langle \mathbf{X}_i - \boldsymbol{\mu}, \mathbf{X}_{i'} - \boldsymbol{\mu} \rangle / n + V_\sigma = \Psi\Psi^\top + V_\sigma$. If we are able to estimate V_σ, we can subtract it from V_y, and then estimate the loadings matrix Ψ. Note that we can exploit the smoothness property of $X_i(\boldsymbol{s})$. Partition the whole domain into $m = m_1 \times \cdots \times m_d$ cubes and use $\widetilde{s} = (\widetilde{s}_1, \ldots, \widetilde{s}_d)$ as the spatial index of each cube, where $1 \le \widetilde{s}_k \le m_k$. Then estimate $\sigma_i^2(\boldsymbol{s})$ as the sample variance of all $Y_{i,\boldsymbol{s}}$ which falls into the same cube with index $\widetilde{\boldsymbol{s}}$. Take the 3D image object as an example and consider the simplest case with $m_k = n_k/2$. Then the first cube, with $\widetilde{s} = (1, 1, 1)$, has eight observations, viz. $Y_{i,(1,1,1)}$, $Y_{i,(1,1,2)}$, $Y_{i,(1,2,1)}$, $Y_{i,(1,2,2)}$, $Y_{i,(2,1,1)}$, $Y_{i,(2,1,2)}$, $Y_{i,(2,2,1)}$, $Y_{i,(2,2,2)}$. By computing the sample variance of these observations, we can obtain a raw estimate of $\sigma_i^2(s_1, s_2, s_3)$, which holds for all $\boldsymbol{s}$ satisfying $s_1, s_2, s_3 \in \{1, 2\}$. Similarly, we could estimate all $\sigma_i^2(\boldsymbol{s})$ and then obtain the (i, i)th element of V_σ by computing the integration $\int \sigma_i^2(\boldsymbol{s}) d\boldsymbol{s}$ via discretization. Therefore, we propose the following algorithm.

Algorithm A1.

Step 1: Partition the whole domain into $m = m_1 \times \cdots \times m_d$ cubes and compute the sample variance $\sigma_i^2(\boldsymbol{s})$.

Step 2: Compute the $I \times I$ inner product matrix V_y, whose (i, i')th element is given by $\langle \mathbf{Y}_i - \boldsymbol{\mu}, \mathbf{Y}_{i'} - \boldsymbol{\mu} \rangle / n$. Note that if $\boldsymbol{\mu}$ is unknown, we replace it by its estimate $\bar{\mathbf{Y}} = (\mathbf{Y}_1 + \cdots + \mathbf{Y}_I)/I$. Let V_σ be the diagonal matrix whose (i, i)th element is obtained from $\int \sigma_i^2(\boldsymbol{s}) d\boldsymbol{s}$. Perform eigendecomposition on $\widehat{V} = V_y - V_\sigma$ and attain the raw estimate of the loading $\widetilde{\psi}_{ij}$s.

Step 3: Regress $Y_{i,\boldsymbol{s}}$ on loadings $\widetilde{\psi}_{ij}$ over all i to get the raw estimate of the jth factor $\widetilde{\phi}_j(\boldsymbol{s})$. Then smooth $\widetilde{\phi}_j(\boldsymbol{s})$ over $\boldsymbol{s}$ to obtain $\widehat{\phi}_j(\boldsymbol{s})$.

Step 4: Regress $Y_{i,\boldsymbol{s}}$ over $\widehat{\phi}_j(\boldsymbol{s})$ to obtain the 1-step updated loadings $\widehat{\psi}_{ij}$s.

Remark 1. According to our theoretical investigations in Section 2.3, the choice of m is not so crucial, but it has to yield undersmoothing in order to obtain a consistent estimate of V_σ. Theoretical discussions on how to choose m is left to Section 2.3; see (C2). In practice, one could follow the recommendation by Ruppert [26] and let $m_k = \min(n_k/4, 35)$.

The above algorithm provides a computational advantage by combining the idea of singular value decomposition (SVD) with smoothing. Step 2 is essentially estimating the left singular vectors if we define the $I \times n$ matrix whose ith row is $\mathbf{X}_i^\top$, so it could reduce the computational burden from n to I compared with eigendecomposition on the covariance. However, a direct application of SVD on the original data or the smoothed data might still be less effective due to the existence of measurement errors. Therefore, we need to subtract V_σ for correction. Smoothing in Step 3 is crucial for satisfactory estimates $\widehat{\phi}_j(\boldsymbol{s})$s. To reduce the computational burden, we recommend using penalized spline techniques and provide more detailed discussion later. Step 4 updates the estimated loadings to improve efficiency. Our procedure shares some similarity with the univariate approach proposed by Kneip and Utikal [16], which is briefly described in Appendix B. However, our procedure significantly reduces the computational burden, and is valid for analyzing general d-dimensional dense functional objects.

2.2. Data-driven implementation

Now we provide more discussion on how our procedure above could be implemented in general for arbitrary fixed dimensional, dense functional objects. As in other nonparametric approaches, the choice of the smoothing parameters is crucial. Hence we discuss how to implement Step 3 using a fast data-driven algorithm. Suppose we have already obtained the raw estimate $\widetilde{\phi}_j(\boldsymbol{s})$. We will smooth each $\widetilde{\phi}_j(\boldsymbol{s})$ separately. We use Z to denote the array formed by $\widetilde{\phi}_j(\boldsymbol{s})$, where $1 \le s_k \le n_k$ for all $k \in \{1, \ldots, d\}$, and given $j \in \{1, \ldots, L\}$.

Extending the idea of Xiao et al. [31], we try to exploit the advantages of the tensor product for reducing the computation costs and smooth the covariates such that the fitted data $\widehat{Z}$ satisfying $\mathrm{vec}(\widehat{Z}) = (S_{\lambda_d} \otimes \cdots \otimes S_{\lambda_1})\,\mathrm{vec}(Z)$, where S_λ is the smoothing matrix with the penalty parameter λ defined as $S_\lambda = B(B^\top B + \lambda P)^{-1} B^\top$, with B and P being the B-spline design matrix and the penalty matrix, respectively.

As a simple illustration, first consider the bivariate case. When $d = 2$, Z is observed as a matrix and one could simply obtain $\widehat{Z} = S_{\lambda_1} Z S_{\lambda_2}$ due to properties of the Kronecker product. The bivariate smoothing could be interpreted as first smoothing each column of the matrix elements and then smoothing each row. Alternatively, the left and right multiplication of S_{λ_1} and S_{λ_2} are essentially conducting two conditional smoothing procedures along the first and the second axis, respectively. As an extension of this idea, we could interpret multivariate smoothing in the general case as a sequence of d conditional smoothing, where at each time, conditional on all the other axes except the ith one, we smooth the available n_i observations using S_{λ_i}. Such a finding is important because when $d > 2$, we do not have simple expressions as for the sandwich smoother. Moreover, it enables us to propose an algorithm that could avoid the rotated transformation for dimension flattening and reinstating as described in the Generalized Linear Array Model (GLAM) algorithm [5].

In particular, we propose the following generalized cross validation (GCV) criterion with a fast computation expression to select the appropriate smoothing parameter $\lambda = (\lambda_1, \ldots, \lambda_d)$.

Proposition 1. *Let Z be an array structure data of dimension $n = n_1 \times \cdots \times n_d$. Suppose the p_ith degree B-splines defined on K_i knots are used. Denote the spline design matrix and the penalty matrix for the ith axis as the matrix B_i and P_i respectively. Then the GCV value equals*

$$\frac{\langle Z - \widehat{Z}, Z - \widehat{Z}\rangle}{n - \mathrm{tr}(S_{\lambda_d} \otimes \cdots \otimes S_{\lambda_1})} = \frac{\langle W \circ W, \widetilde{Z} \circ \widetilde{Z}\rangle - 2\langle W, \widetilde{Z} \circ \widetilde{Z}\rangle + \langle Z, Z\rangle}{n - \prod_{j=1}^{d} tr_j}, \tag{2}$$

where $tr_j = \sum_{i=1}^{K_j+p_j} 1/(1+\lambda_j \eta_{j,i})$ with $\eta_{j,i}$ defined in Eq. (A.1), and W and $\widetilde{Z}$ are defined in Eqs. (A.2)–(A.3).

Remark 2. According to Proposition 1, we could precalculate $\eta_{j,i}$s, $\widetilde{Z}$ and $\langle Z, Z\rangle$. For each λ, we only need to update the calculation of tr_j as well as two inner products that involves W. As the dimension of W is $(K_1 + p_1) \times \cdots \times (K_d + p_d)$ instead of $n_1 \times \cdots \times n_d$, this could further reduce the computation cost.

2.3. Asymptotics

Now we consider the asymptotics of our estimates under the following conditions.

(C1) Suppose $E(Y_{i,s}^{2+\delta}) < \infty$ and $Y_{i,\boldsymbol{s}} = X_i(\boldsymbol{s}) + \varepsilon_{i,\boldsymbol{s}}$, where the error terms $\varepsilon_{i,\boldsymbol{s}}$ are independent across individuals is and the space ($\boldsymbol{s}$). They are also independent of $X_i(\boldsymbol{s})$ and have 0 mean and finite variance $\sigma_i^2(\boldsymbol{s})$. The random functions $X_i(\boldsymbol{s})$ admit the representation (1) with a bounded support. The factors $\phi_j(\boldsymbol{s})$ have continuous second order derivatives, and the loadings ψ_{ij} are uncorrelated with mean 0 and distinct positive variance γ_j for all $j \in \{1, \ldots, L\}$. The number of observations along each axis has the same order $n^{1/d}$, where $n = n_1 \times \cdots \times n_d$. The number of individuals satisfies $I = \mathcal{O}(n^{d/4})$.

(C2) Partition the whole domain into $m = m_1 \times \cdots \times m_d$ cubes, where m_k grows at the rate of n_k^{δ} with $3/4 < \delta < 1$.

(C3) With the raw estimates of the loadings $\widetilde{\psi}_{ij}$s and the factors $\widetilde{\phi}_j(\boldsymbol{s})$s, use a tensor product B-splines basis to estimate $\widehat{\phi}_j(\boldsymbol{s})$ with the following settings. Let $p_k > 0$ and K_k denote the spline degree and the number of knots along the kth axis. The knots are equally spaced on $\mathcal{F}$ and K_k grows faster than the rate of $n_k^{\alpha_k}$ for some fixed constant $\alpha_k > 1/5$. The second order penalty is imposed. Denote the penalty parameters as $\lambda_j = (\lambda_{j1}, \ldots, \lambda_{jd})$, where the relationship between the penalty parameter and the equivalent bandwidth, along the kth axis, satisfies

$$\lambda_{jk} = \{K_k h_{jk} (In)^{-1/(4+d)}\}^2, \tag{3}$$

for some positive constants h_{jk}s.

Theorem 1. *Assume Conditions (C1)–(C3). For any interior point $\boldsymbol{s}$, as $n \to \infty$,*

$$(In)^{2/(4+d)}\{\widehat{\phi}_j(\boldsymbol{s}) - \phi_j(\boldsymbol{s})\} \rightsquigarrow \mathcal{N}[\mathcal{B}_j(\boldsymbol{s}), \mathcal{V}_j(\boldsymbol{s})], \tag{4}$$

where the asymptotic bias $\mathcal{B}_j(\boldsymbol{s})$ and the asymptotic variance $\mathcal{V}_j(\boldsymbol{s})$ are given in Eq. (A.5).

Remark 3. (C1) imposes regularity conditions on the observations; (C2) guarantees a consistent estimate of V_σ. The detailed settings of the penalized spline smoothing, including the choices of the penalty and the splines are provided in (C3). As pointed out by Li and Ruppert [17], the placement of the knots is not crucial as long as the number of knots exceeds some threshold. Under this large penalty setting, the key smoothing parameter is the penalty term and the penalized smoothing approach using the penalty λ is equivalent to the kernel smoothing approach using an equivalent bandwidth h defined as in Eq. (3).

3. Applications

In this section, we implement our approach to the fMRI data to learn how the human brain responds to investment decision tasks. In particular, we extract spatial functional factors to recover the risk related ROIs and employ the factor loadings to predict the risk attitude on reward and risk.

3.1. Experiment design and data

Our data are collected from a risk perception investment decisions experiment. The data set consists of the high resolution fMRI brain images of 22 subjects (age 18–35 years, 11 females, native German speakers, right-handed and had no history of neurological or psychiatric disease). The same data set is analyzed by Mohr et al. [23], van Bömmel et al. [1], and Chen et al. [3]. Three participants had to be excluded due to extensive head motion ($>$ 5 mm absolute head movement) or modeling problems (always chose only one alternative). Another two were excluded due to a different scanning frequency. The experiment was conducted as follows: subjects were requested to answer investment related tasks. Each task consists of two phases. In the first phase a stream of 10 returns was sequentially presented for 2×10 s. The random return streams were independently drawn from Gaussian distributions with means of 6%, 9%, 12% and standard deviations of 1%, 5%, 9%;

there were thus nine different combinations in total. In the second phase, subjects performed one of the three task types within 7 s: either choose between 5% fixed safe and risky investment (as shown in the random returns stream) or tell their subjective judgment (expected return or perceived risk) on the random returns. There were 27 trails for each task type (3×27 tasks in total) and the order of the task types was determined at random.

The fMRI data were acquired every 2.5 s during the whole experiment. This resulted in 1400 observations of a 3-dimensional ($91 \times 109 \times 91$) array that represents the Blood Oxygenation Level Dependent (BOLD) signals. At the same time, the answer for each task from each subject was also collected. The data are available at the Research Data Centre (RDC), Humboldt-Universität zu Berlin.

3.2. Estimation

Recall that for each subject we have a series of 1400 three-dimensional images ($91 \times 109 \times 91$). We divide the whole time series into subseries based on the beginning of each task. Therefore, for the ℓth subject, during the qth task, the tth image could form a 3-dimensional array data $Z_{\ell,q,t}$. We shall also use $Z_{\ell,q,t,\boldsymbol{s}}$ to refer to its observation at the $\boldsymbol{s}$ location with coordinate $\boldsymbol{s} = (s_1, s_2, s_3)$. We assume that the subjects exhibit identical brain structure and commonly share the same active regions during the whole experiment procedure.

Since the subject was required to give an answer within 7 s and the fMRI was taken every 2.5 s, we shall consider the first three images ($t \in \{1, 2, 3\}$) taken for the ℓth subject during the qth task. Moreover, as we focus on exploring how the BOLD signals change when the subject makes decisions, we compute $\widetilde{Z}_{\ell,q,\boldsymbol{s}} = (Z_{\ell,q,2,\boldsymbol{s}} + Z_{\ell,q,3,\boldsymbol{s}})/2 - Z_{\ell,q,1,\boldsymbol{s}}$. To simplify the notation, we denote $Y_{i,\boldsymbol{s}} = \widetilde{Z}_{\ell,q,\boldsymbol{s}}$, where $i = 81 \times (\ell - 1) + q$, $\ell \in \{1, \ldots, 17\}$ and $q \in \{1, \ldots, 81\}$, or equivalently, $i \in \{1, \ldots, 1377\}$.

We follow the algorithm described in Section 2 to estimate the spatial functional factors and the corresponding loadings. In particular, we divide the $91 \times 109 \times 91$ voxels into $30 \times 35 \times 30$ cubes. When we do the 3D penalized smoothing on the factors in Step 3, we adopt the cubic B-spline basis with 30, 35, 30 knots along the x, y, z axes, respectively. The penalty matrices for all directions are of order 2. The tuning parameters in the penalized smoothing are selected by minimizing the GCV values computed from Eq. (2).

When implementing the above algorithm, we also need to determine the number of factors and we adopt a two-stage procedure for this purpose. First, we determine a rough choice on the number of factors based on the eigendecomposition of $\widetilde{V}$. Note that $\widetilde{V}$ might have some negative eigenvalues due to a numerical approximation in calculation. Denote by c the magnitude of the minimum eigenvalue. Eigenvalues with magnitude less than c are truncated to 0. For the rest of the eigenvalues, we compute the cumulative proportion and select the number of factors when the explained proportion first exceeds 80%. After we update the loadings in Step 4, we could compute the variance of the loadings associated with each factor. Then we update our estimate about the number of factors using the same principle, i.e., when the explained proportion first exceeds 80%. In our case, we first select 19 factors and then refine this choice into nine factors.

3.3. Empirical results

The objectives of our empirical analysis are to recover the risk related regions and also to predict subjects' risk attitude.

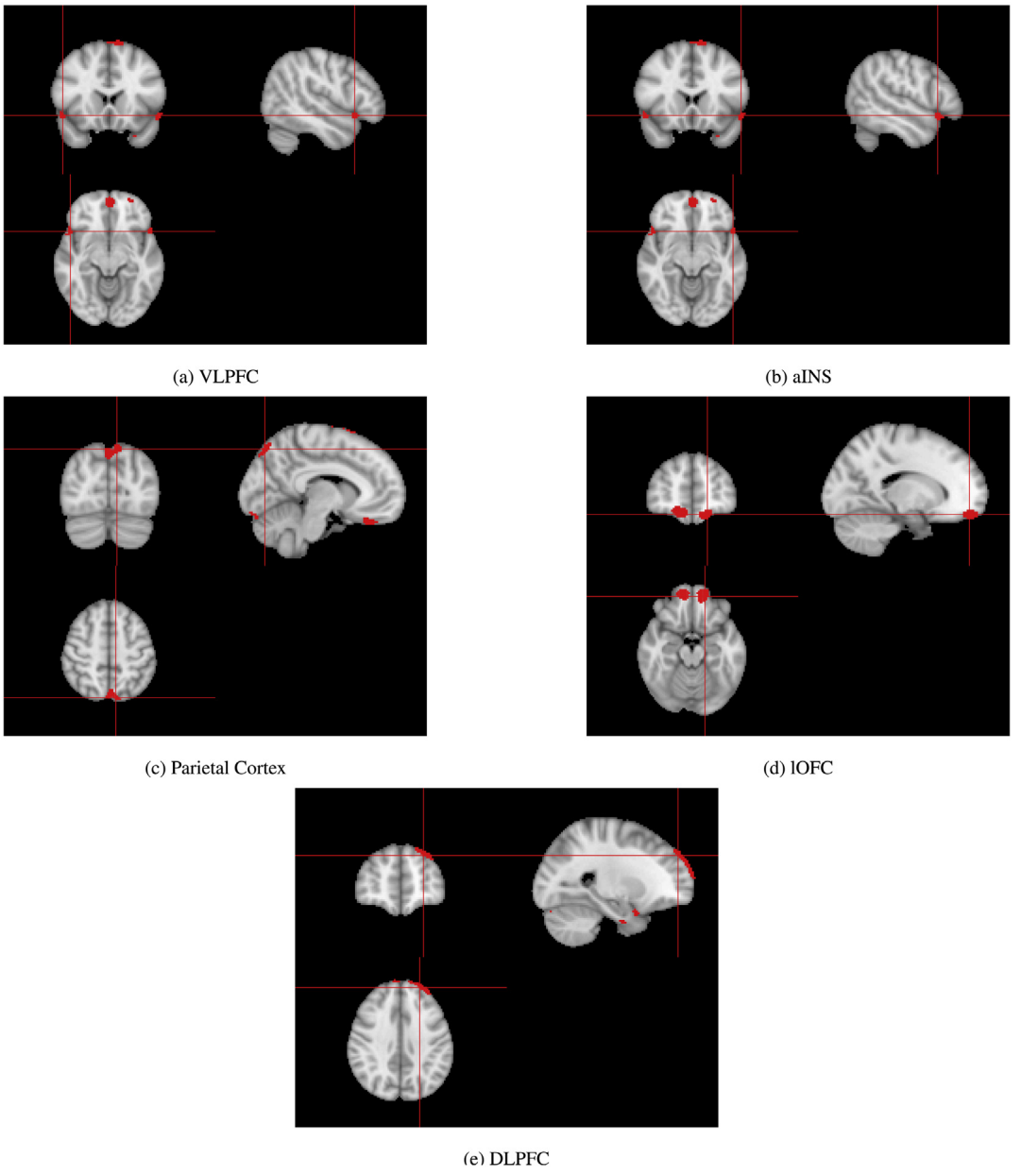
3.3.1. Recovery of the risk related ROIs

It is evident from [1,23,29] that parietal cortex, ventrolateral prefrontal cortex (VLPFC), lateral orbitofrontal cortex (lOFC), anterior insula (aINS), and dorsolateral prefrontal cortex (DLPFC) are the five regions which activate the brain activities in investment decision making. They are regarded as the active regions to be recovered in Chen et al. [3], where the same data set is analyzed. To evaluate the estimation performance of the factors, we also compare the spatial functional factors identified by our approach to those five target regions. In particular, the values of the factors are trimmed by the 0.1%− and 99.9%+ quantile levels for each j and the locations where $\widehat{\phi}_j(\boldsymbol{s})$ have nonzero values are marked as red area in Fig. 1, which presents the recovery of the five ROIs.

The results show that the five risk related regions can be identified by $\widehat{\phi}_j(\boldsymbol{s})$ for $j \in \{1, 4, 5, 8\}$ with our approach; 75.81% of the variance can be explained by the first eight factors, where the proportion of $j \in \{1, 4, 5, 8\}$ is 42.60%. In contrast, Chen et al. [3] consider the first 19 factors and find that the 4, 18, 3 (or 12), 5, 19th factors correspond to these five regions, respectively. Our method seems to be more efficient in extracting the crucial features when dealing with high-dimensional data. In addition, the computational time consumption of our approach is much less compared to Chen et al. [3], where the exact same data set is used. In our case, it took four hours to run the R codes for our analysis (2.5 h for loading and preprocessing the data and 1.5 h for the estimation procedure) on a desktop with Inter(R) Core(TM) i7-6700 CPU @ 3.40 GHz processor and 32.0 GB memory without calling parallelization.

3.3.2. Prediction of the risk attitude

We have estimated the subject specific factor loadings and we expect that the individual information carried by the loadings can predict the subjects' risk attitudes. Mohr et al. [23] estimate the risk attitude parameters R_ℓ for the 17 subjects in the same experiment by the psychological risk–return model; see Appendix C for more details. We will explore the relationship between their R_ℓ and our estimated loadings by simple regression and conduct predictions on the risk attitudes.



(a) VLPFC (b) aINS

(c) Parietal Cortex (d) lOFC

(e) DLPFC

Fig. 1. The recovery of VLPFC and aINS by $\widehat{\phi}_1(\boldsymbol{s})$, Parietal Cortex by $\widehat{\phi}_4(\boldsymbol{s})$, IOFC by $\widehat{\phi}_5(\boldsymbol{s})$, and DLPFC by $\widehat{\phi}_8(\boldsymbol{s})$, respectively. The cross of the red lines indicates the center of the true location of the target region.

From the brain images associated with the ℓth subject and the qth task, we obtain the corresponding loading $\widehat{\psi}_{i,j}$, $j \in \{1, \ldots, 9\}$, where $i = 81 \times (\ell - 1) + q$. In order to predict R_ℓ using the loadings, we first summarize the loadings associated with the ℓth subject and the jth factor by computing the sample mean and sample standard deviation among the 81 tasks, viz.

$$\nu_{\ell,j} = \frac{1}{81}\sum_{q=1}^{81} \widehat{\psi}_{\{81(\ell-1)+q\},j}, \quad \varsigma_{\ell,j} = \sqrt{\frac{1}{81}\sum_{q=1}^{81}(\widehat{\psi}_{\{81(\ell-1)+q\},j} - \nu_{\ell,j})^2}.$$

We will adopt three approaches to figure a suitable model for risk attitude prediction. To simplify the notation, we drop the subjects indicator ℓ from the subscript to represent the vector accordingly.

Table 1
The newly deleted variable and the corresponding AIC of the model in each step by backward regression with ν_j.

Model	1	2	3	4	5
Deleted variable	–	ν_6	ν_7	Intercept	ν_3
AIC	117.74	115.74	113.99	112.60	111.18
Model	6	7	8	9	10
Deleted variable	ν_8	ν_4	ν_2	ν_9	ν_5
AIC	110.90	110.60	108.81	**107.51**	112.13

Table 2
The newly deleted variable and the corresponding AIC of the model in each step by backward regression with incorporating ς_j.

Model	1	2	3	4	5
Deleted variable	–	Intercept	ς_2	ς_3	ς_8
AIC	112.84	110.88	109.17	107.69	106.22
Model	6	7	8		
Deleted variable	ς_6	ς_7	ς_1		
AIC	105.04	**103.50**	106.36		

In our first approach, we explore the relationship between risk attitude and the covariates of the sample mean ν_j in a standard linear regression model, viz.

$$R = \alpha_0 + \sum_{j=1}^{9} \alpha_j \nu_j + u.$$

As not all covariates might be important, we adopt backward regression and delete the variable that has the largest p-value iteratively until the model selection criterion — AIC no longer decreases. Table 1 summarizes the newly deleted variable in each step and the corresponding AIC. The final estimated regression is

$$\widehat{R} = \underset{(0.36)}{-1.50}\nu_1 \underset{(0.65)}{-1.75}\ \nu_5, \tag{5}$$

where the values in parentheses are the estimated standard deviation errors.

In contrast, if we incorporate the information that factors 1, 4, 5, 8 are associated with the risk related ROIs, then we may start from setting the regression model as

$$R = \alpha_0 + \sum_{j\in\{1,4,5,8\}} \alpha_j \nu_j + u,$$

By backward regression, we reach the same optimal result as (5).

We suspect that ν_j and ς_j might reflect different information of the overall performance and dispersion among tasks. In our second approach, we plan to include ς_j in the covariates for improvement. We start from the optimal model obtained in the first approach and first use all covariates defined by the sample standard deviation. We then adopt backward regression and delete the variable that has the largest p-value iteratively until the AIC no longer decreases. Table 2 summarizes the newly deleted variable in each step and the corresponding AIC. The final estimated regression is

$$\widehat{R} = \underset{(0.48)}{-0.65}\nu_1 \underset{(1.69)}{-2.68}\ \nu_5 + \underset{(0.70)}{1.34}\varsigma_1 + \underset{(0.89)}{2.60}\varsigma_4 \underset{(1.58)}{-3.4}\ \varsigma_5 \underset{(0.89)}{-2.14}\ \varsigma_9. \tag{6}$$

In contrast, if we incorporate the information that factors 1, 4, 5, 8 are associated with the risk related ROIs, then we may start from a base model as

$$R = \alpha_0 + \sum_{j\in\{1,4,5,8\}} \alpha_j \nu_j + \sum_{j\in\{1,4,5,8\}} \beta_j \varsigma_j + u. \tag{7}$$

By backward regression, we come to the following result:

$$\widehat{R} = \underset{(0.61)}{-0.98}\nu_1 \underset{(1.92)}{-1.66}\ \nu_5 \underset{(0.81)}{+1.37}\ \varsigma_1 \underset{(1.18)}{+2.46}\ \varsigma_4 \underset{(1.82)}{-3.62}\ \varsigma_5 \underset{(1.11)}{-1.55}\ \varsigma_8.$$

The result is quite similar to our proposed model (6), except that it includes ς_8 rather than ς_9. However, the AIC of this fitted model is 108.00, much higher than the AIC of 103.50 in model (6).

In our third approach, we try to further improve our model by adopting a bi-directional stepwise selection based on (6). To be more specific, alternately we add the variable which improves the AIC mostly due to the inclusion of itself and then delete the variable that has the highest p-value (only if the AIC can be further reduced by the exclusion, otherwise we do

Table 3
The actions taken in the bi-directional selection and the corresponding AIC of the model in each step, with (6) as the base model.

Model	1	2	3	4
Add(+)/Delete(−)	$+\nu_4$	$+\nu_9$	$+\varsigma_2$	$+\nu_7$
AIC	99.75	99.38	95.62	90.87

Note: In each step after a new variable is included into the model, ς_5, ν_9, ς_2, ν_7 are detected to have the highest p-values accordingly. But the exclusion was not implemented because the AIC was found to worsen.

Table 4
The actions taken in the bi-directional selection and the corresponding AIC of the model in each step, with (7) as the base model.

Model	1	2	3
Add(+)/Delete(−)	$+\varsigma_9$	$-\nu_8$	$+\nu_6$
AIC	101.81	99.82	95.53

Note: In the last step, ν_6 is detected to have the highest p-values. But the exclusion was not implemented because the AIC was found to worsen.

Table 5
Spearman's and Kendall's rank correlations for the out-of-sample prediction of the models with the selected variables.

Selected variables	Spearman's rho	Kendall's tau
ν_1, ν_5	0.15	0.13
$\nu_1, \nu_5, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_9$	0.46	0.35
$\nu_1, \nu_5, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_8$	0.37	0.26
$\nu_1, \nu_4, \nu_5, \nu_7, \nu_9, \varsigma_1, \varsigma_2, \varsigma_4, \varsigma_5, \varsigma_9$	0.52	0.36
$\nu_1, \nu_4, \nu_5, \nu_6, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_8, \varsigma_9$	0.57	0.45

not take the action). We stop the procedure once the AIC cannot be improved anymore by adding any new variable. All the actions taken in the procedure and the corresponding AIC are presented in Table 3.

The final estimation we get is

$$\widehat{R} = \underset{(0.43)}{-1.03}\nu_1 \underset{(1.15)}{+4.64}\,\nu_4 \underset{(1.34)}{-4.52}\,\nu_5 \underset{(0.83)}{+1.53}\,\nu_7 \underset{(2.05)}{+4.10}\,\nu_9 \underset{(0.79)}{+3.40}\,\varsigma_1 \underset{(0.70)}{+1.92}\,\varsigma_2 \underset{(0.78)}{+2.72}\,\varsigma_4 \underset{(1.93)}{-8.61}\,\varsigma_5 \underset{(0.77)}{-3.33}\,\varsigma_9. \tag{8}$$

In contrast, if we start from model (7) and adopt the bi-direction stepwise selection (see Table 4), we end up with the following model:

$$\widehat{R} = \underset{(0.43)}{-0.95}\nu_1 \underset{(1.50)}{+4.63}\,\nu_4 \underset{(1.42)}{-3.84}\,\nu_5 \underset{(1.22)}{-2.31}\,\nu_6 \underset{(0.55)}{+1.46}\,\varsigma_1 \underset{(0.84)}{+3.17}\,\varsigma_4 \underset{(1.30)}{-4.74}\,\varsigma_5 \underset{(1.81)}{+4.01}\,\varsigma_8 \underset{(1.30)}{-5.09}\,\varsigma_9.$$

For this fitted model, the coefficients associated with ν_4, ν_5, ς_1, ς_4, ς_5, ς_9 are significant at the level 0.05. Moreover, they are also significant in the estimated model (8).

Finally, we compare the out-of-sample performance of the three selection approaches described above. In particular, we conduct a leave-one-out procedure to predict $\widetilde{R}_1, \ldots, \widetilde{R}_{17}$. The steps are as below:

(1) For fixed $\ell \in \{1, \ldots, 17\}$, conduct ordinary linear regression with the sample of the remaining 16 subjects to estimate the coefficients of the variables in different models.
(2) Predict $\widetilde{R}_\ell$ by plugging in the coefficients estimated in Step (1).
(3) Repeat Steps (1) and (2) for each $\ell \in \{1, \ldots, 17\}$.
(4) Compute Spearman's and Kendall's rank correlations between $\widetilde{R}_1, \ldots, \widetilde{R}_{17}$ and $R_1, \ldots, R_{17}$ to check the prediction accuracy on the order of risk-aversion among the subjects.

We summarize the out-of-sample prediction performance of the models in Table 5. The results confirm that the factor loadings estimated in our analysis can provide credible prediction on subjects' risk attitude with appropriately determined models. In particular, we notice that factor 9 might also be of importance for risk perception (especially the dispersion measure of the corresponding loadings), even though it does not match with the five risk related regions. However, factor 8 might not be as necessary as the other factors associated with the ROIs (factors 1, 4, 5) in predicting subjects' risk attitude. In contrast to Chen et al. [3], we focus on the regression-based analysis, which aims at predicting the specific values of risk attitudes rather than classify the subjects into different groups.

4. Conclusions

In this paper, we propose the spatial FPCA method that is applicable to arbitrary fixed d-way array data. The approach could preserve the spatial structure and efficiently extract the important features via a model-free dimension reduction

approach. We design a fast algorithm that could reduce the computational burden without sacrificing the estimation efficiency. We also derive the asymptotic properties of our estimator. A potential future research direction could consider imposing time dynamic structure on the loadings to relax the uncorrelated assumption, thus modeling the temporal and spatial process simultaneously.

Acknowledgments

We thank the Editor-in-Chief, the Guest Managing Editor, and two anonymous referees for their constructive comments. Financial support from the Chinese Nature National Science Foundation #71571154, the Fujian Key Laboratory of Statistical Science, and the Deutsche Forschungsgemeinschaft via IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin, is gratefully acknowledged. The codes to implement the algorithms are publicly accessible at www.quantlet.de.

Appendix A. Detailed proofs

Proof of Proposition 1. Recall that the smoothing along the jth axis is $S_j = B_j(B_j^\top B_j + \lambda_j P_j)^{-1}B_j^\top$. If we apply singular value decomposition on $B_j^\top$ and obtain $B_j^\top = U_j D_j Q_j^\top$, then we could define the square root of $(B_j^\top B_j)^{-1}$ simply by $\Sigma_{B_j}^{-1/2} = U_j D_j{}^{-1} U_j^\top$. Correspondingly,

$$(B_j^\top B_j + \lambda_j P_j)^{-1} = \Sigma_{B_j}^{-1/2}(\mathbf{I}_{K_j+p_j} + \lambda_j \Sigma_{B_j}^{-1/2} P_j \Sigma_{B_j}^{-1/2})^{-1}\Sigma_{B_j}^{-1/2},$$

where $\mathbf{I}_{K_j+p_j}$ is the identity matrix of dimension $(K_j + p_j) \times (K_j + p_j)$. Perform the eigendecomposition to obtain

$$\Sigma_{B_j}^{-1/2} P_j \Sigma_{B_j}^{-1/2} = \widetilde{U}_j \widetilde{D}_j \widetilde{U}_j^\top, \tag{A.1}$$

where $\widetilde{D}_j$ is the diagonal matrix whose (i, i)th element is denoted as $\eta_{j,i}$. Then $B_j \Sigma_{B_j}^{-1/2} = Q_j D_j U_j^\top U_j D_j{}^{-1} U_j^\top = Q_j U_j^\top$, which is an orthogonal matrix. The eigendecomposition of S_j could therefore be expressed as

$$S_j = B_j \Sigma_{B_j}^{-1/2}(\mathbf{I}_{K_j+p_j} + \lambda_j \widetilde{U}_j \widetilde{D}_j \widetilde{U}_j^\top)^{-1}(B_j \Sigma_{B_j}^{-1/2})^\top = Q_j U_j^\top (\widetilde{U}_j \widetilde{U}_j^\top + \lambda_j \widetilde{U}_j \widetilde{D}_j \widetilde{U}_j^\top)^{-1}(Q_j U_j^\top \widetilde{U}_j)^\top = \widehat{U}_j W_j (\widehat{U}_j)^\top,$$

where $\widehat{U}_j = Q_j U_j^\top \widetilde{U}_j$ is an orthogonal matrix, and W_j is the diagonal matrix whose (i, i)th element is $1/(1 + \lambda_j \widetilde{D}_j)$.

Let W be an array with dimension $(K_1 + p_1, \ldots, K_d + p_d)$ constructed by applying the outer products on the vectors from the diagonal elements of W_j. For the element of W with coordinate $\boldsymbol{s}$ being $(s_1, \ldots, s_d)$, we have

$$W_{\boldsymbol{s}} = \prod_{k=1}^{d} \Big(\frac{1}{1 + \lambda_j \eta_{j,s_k}} \Big). \tag{A.2}$$

Let $\widetilde{Z}$ be the array obtained from multiplying all $\widehat{U}_j$s along the jth axis, respectively. Equivalently, we could also define $\widetilde{Z}$ in an iterative way. When $d = 2, \widetilde{Z} = \widehat{U}_1 Z \widehat{U}_2^\top$. When $d = k + 1$,

$$\mathcal{A}(\widetilde{Z}) = \big(\otimes_{j=1}^{k} \widehat{U}_j\big)\, \mathcal{A}(Z) \widehat{U}_{k+1}, \tag{A.3}$$

where $\mathcal{A}(Z)$ is a mapping that converts the $(n_1, \ldots, n_d)$ dimensional array Z into a $(\prod_{i=1}^{k} n_i) \times n_{k+1}$ matrix such that the kth column collects all observations whose last coordinate equals k. We now show that

$$\langle Z, \widehat{Z} \rangle = \langle W, \widetilde{Z} \circ \widetilde{Z} \rangle. \tag{A.4}$$

First consider $d = 2$. By the definition of $\widetilde{Z}$, we have

$$\widehat{Z} = S_1 Z S_2 = \widehat{U}_1 W_1 (\widehat{U}_1^\top Z \widehat{U}_2) W_2 \widehat{U}_2^\top = \widehat{U}_1 W_1 \widetilde{Z} W_2 \widehat{U}_2^\top.$$

By the definition of W and the property that W_1 and W_2 are both diagonal matrices, we have $W_1 \widetilde{Z} W_2 = W \circ \widetilde{Z}$. Let $n = n_1 \times \cdots \times n_d$. Then

$$\begin{aligned} \langle Z, \widehat{Z} \rangle &= \operatorname{vec}(Z^\top) \operatorname{vec}(\widehat{Z}) = \operatorname{vec}(Z^\top) \operatorname{vec}(\widehat{U}_1 (W \circ \widetilde{Z}) \widehat{U}_2^\top) \\ &= \operatorname{vec}(Z^\top)(\widehat{U}_2 \otimes \widehat{U}_1) \operatorname{vec}(W \circ \widetilde{Z}) \\ &= \operatorname{vec}(\widetilde{Z})^\top \operatorname{vec}(W \circ \widetilde{Z}) = \langle \widetilde{Z}, W \circ \widetilde{Z} \rangle = \langle W, \widetilde{Z} \circ \widetilde{Z} \rangle. \end{aligned}$$

Assume Eq. (A.4) holds when $d = k$. Now we prove that it also holds for $d = k + 1$. By the definition of $\widehat{Z}$ and Eq. (A.3), we have

$$\mathcal{A}(\widehat{Z}) = \big(\otimes_{j=1}^{k} \widehat{U}_j\big)\, \big(\otimes_{j=1}^{k} W_j\big)\, \mathcal{A}(\widetilde{Z}) W_{k+1} \widehat{U}_{k+1}^\top.$$

As $(\otimes_{j=1}^{k} W_j)\mathcal{A}(\widetilde{Z})W_{k+1} = \mathcal{A}(W \circ \widetilde{Z})$, we conclude that

$$\langle Z, \widehat{Z}\rangle = \mathrm{vec}(\widetilde{Z})^\top \mathrm{vec}(W \circ \widetilde{Z}) = \langle \widetilde{Z}, W \circ \widetilde{Z}\rangle = \langle W, \widetilde{Z} \circ \widetilde{Z}\rangle,$$

and thus Eq. (A.4) holds in general. Using similar techniques, we could also show that

$$\langle \widehat{Z}, \widehat{Z}\rangle = \mathrm{vec}(W \circ \widetilde{Z})^\top \mathrm{vec}(W \circ \widetilde{Z}) = \langle W \circ \widetilde{Z}, W \circ \widetilde{Z}\rangle = \langle W \circ W, \widetilde{Z} \circ \widetilde{Z}\rangle.$$

Finally, the smoother satisfies $\mathrm{tr}(S_{\lambda_d} \otimes \cdots \otimes S_{\lambda_1}) = \prod_{i=1}^{d} \mathrm{tr}(S_{\lambda_i})$. Moreover,

$$\mathrm{tr}(S_{\lambda_i}) = \mathrm{tr}(W_i) = \sum_{j=1}^{K_i+p_i} 1/(1+\lambda_i\eta_{i,j}) = tr_i,$$

where tr_i is defined in Proposition 1. Therefore, Eq. (2) holds. □

Proof of Theorem 1. Define V_x as the $I \times I$ matrix whose (i, j)th element is $\int X_i(\boldsymbol{s})X_j(\boldsymbol{s})d\boldsymbol{s}$. Note that γ_j, which is the variance of the ψ_{ij}s, is the jth largest eigenvalue of V_x. Let $\widetilde{\gamma}_j$ be the jth largest eigenvalue of $\widetilde{V}$ respectively. By our assumption (C2) on m_k, the bin width along each direction m_k/n_k is $o(n_k^{-1/4})$. Therefore, we could proceed as Kneip and Utikal [16] to show that

$$\|\widetilde{V} - V_x\|_2 = \sup_{\|v\|_2=1} \{v^\top(\widetilde{V} - V_x)^\top(\widetilde{V} - V_x)v\}^{1/2} = \mathcal{O}_p(In^{-1/2}).$$

Hence $\widetilde{\gamma}_j - \gamma_j = \mathcal{O}_p(I^{1/2}n^{-1/2})$, for all $j \in \{1, \ldots, L\}$.

Define $\psi_j = (\psi_{1j}, \ldots, \psi_{Ij})^\top$ and decompose it as γ_j multiplying a unit vector $p_j = (p_{1j}, \ldots, p_{Ij})^\top$. Similarly, we could define $\widetilde{\psi}_j$ and decompose it as $\widetilde{\gamma}_j\widetilde{p}_j$. Note that $|\widetilde{p}_{ij} - p_{ij}| = \mathcal{O}_p(I^{-1/2}n^{-1/2})$. Consequently, $\|\widetilde{p}_j - p_j\|_2 = \mathcal{O}_p(n^{-1/2})$ and hence we have $|\widetilde{\psi}_{ij} - \psi_{ij}| = \mathcal{O}_p(n^{-1/2})$, for all $j \in \{1, \ldots, L\}$.

Denote the regression coefficient of $Y_{i,\boldsymbol{s}}$ on ψ_{ij} as $z_{j,\boldsymbol{s}}$. For each j, we have $z_{j,\boldsymbol{s}} = \phi_{j,\boldsymbol{s}} + e_{j,\boldsymbol{s}}$, where the $e_{j,\boldsymbol{s}}$ are independent across $\boldsymbol{s}$ with mean 0 and variance $I^{-1}\bar{\sigma}_j^2(\boldsymbol{s})$, where $\bar{\sigma}_j^2(\boldsymbol{s}) = I^{-1}\sum_{i=1}^{I}\sigma_i^2(\boldsymbol{s})$. Denote $h_j = (In)^{-d/(4+d)}\prod_{k=1}^{d} h_{jk}$. For any interior point $\boldsymbol{s}_0 = (s_{01}, \ldots, s_{0d})$, denote

$$z_j^*(\boldsymbol{s}) = (nh)^{-1}\sum_{s_1=1}^{n_1}\cdots\sum_{s_d=1}^{n_d} z_{j,\boldsymbol{s}}\prod_{k=1}^{d} H\left(\frac{s_{0k} - s_k}{h_{jk}(In)^{-1/(4+d)}}\right),$$

where H is the second order kernel function $2^{-1}\exp(-|x|)$. Note that

$$E\{\widehat{z}_j(\boldsymbol{s}) - z_j^*(\boldsymbol{s})\} = \mathcal{O}\left[\max_{1\le k\le d}\{K_k h_{j,k}^{-2}(In)^{-2/(4+d)}\}\right] = o\left[\max_{1\le k\le d}\{h_{j,k}^2(In)^{-2/(4+d)}\}\right]$$

and $\mathrm{var}\{\widehat{z}_j(\boldsymbol{s}) - z_j^*(\boldsymbol{s})\} = o\{(Inh)^{-1}\}$. Given that $z_j^*(\boldsymbol{s})$ is the Nadaraya–Watson estimator with product kernel, we have

$$(In)^{2/(4+d)}[E\{\widehat{z}_j(\boldsymbol{s})\} - \phi_j(\boldsymbol{s})] \to \mathcal{B}_j(\boldsymbol{s}) \quad \text{and} \quad (In)^{4/(4+d)}\mathrm{var}\{\widehat{z}_j(\boldsymbol{s})\} \to \mathcal{V}_j(\boldsymbol{s}),$$

where

$$\mathcal{B}_j(\boldsymbol{s}) = \sum_{k=1}^{d} h_{j,k}^2\frac{\partial^2\phi_j(\boldsymbol{s})}{\partial^2 s_k}, \quad \mathcal{V}_j(\boldsymbol{s}) = \left(\prod_{k=1}^{d} h_{j,k}\right)^{-1}\bar{\sigma}_j(\boldsymbol{s})\left\{\int H^2(u)du\right\}^d. \tag{A.5}$$

Therefore, we could establish asymptotic normality as in Eq. (4).

When ψ_{ij} is unavailable, we will regress $Y_{i,\boldsymbol{s}}$ on $\widetilde{\psi}_{ij}$. Denote the regression coefficient as $\widetilde{z}_{j,\boldsymbol{s}}$. Note that the deviation between $\widetilde{\psi}_{ij}$ and ψ_{ij} is of higher order magnitude $\mathcal{O}_p(n^{-1/2})$. When $I = o(n^{d/4})$, such a deviation is negligible compared to $(In)^{-2/(4+d)}$. Therefore, we conclude that smoothing over $\widetilde{z}_{j,\boldsymbol{s}}$ will yield the same asymptotic results as Eq. (4). □

Appendix B. Estimation for density families by FPCA [16]

Consider the univariate i.i.d. data X_{it} with underlying density f_t for all $t \in \{1, \ldots, T\}$. The objective is to estimate the density families (f_t) based on the KL decomposition, viz.

$$f_t = f_\mu + \sum_{j=1}^{L}\theta_{tj}g_j, \tag{B.1}$$

where $f_\mu = (f_1 + \cdots + f_T)/T$.

The basic idea is to first estimate an efficient estimator $\widehat{M}$ of M, which is a $T \times T$ matrix M, whose (t, s)th element is given by $\langle f_t - f_\mu, f_s - f_\mu\rangle$. Note that the eigenvectors $\boldsymbol{p}_1 = (p_{11}, \ldots, p_{T1})^\top$, $\boldsymbol{p}_2 = (p_{12}, \ldots, p_{T2})^\top$, ... of M corresponding to the nonzero eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots$ are related to the θ_{tj} and g_j in (B.1) by

$$\theta_{tj} = \lambda_j^{1/2}p_{tj}, \quad g_j = \sum_{t=1}^{T}\theta_{tj}f_t \Big/ \sum_{t=1}^{T}\theta_{tj}^2.$$

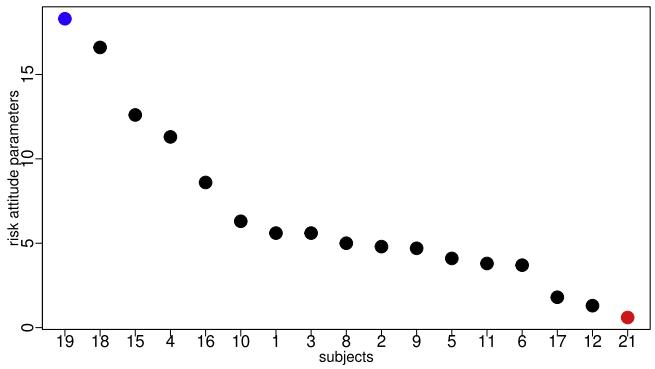


Fig. C.1. Risk attitude parameters for the 17 subjects.

So we can estimate $\widehat{\theta}_{tj}$ and $\widehat{g}_j$ of θ_{tj} and g_j by plugging in the eigenvalues and eigenvectors of $\widehat{M}$ as well as kernel estimators $\widehat{f}_t$ of f_t. However, this approach differs from ours in the following ways:

(1) The approach by Kneip and Utikal [16] focuses on the objects f_ts that are smooth densities. To obtain an efficient estimate $\widehat{M}$, they use kernel density estimators to first estimate $\widehat{f}_{t,h} = (n_t h)^{-1} \sum_i K\{(x - X_{it})/h\}$, where h denotes the bandwidth, and then construct $\widehat{M}$.
(2) Though they also consider removing the diagonal bias in $\widehat{M}$, they estimate $\widehat{g}_j$ by regressing the smoothed $\widehat{f}_{t,b}$ on $\widehat{\theta}_{tj}$. Here $\widehat{f}_{t,b}$ denotes a kernel estimator with bandwidth b not necessarily equal to h used above. Therefore, for each object f_t, smoothing twice with different bandwidths b and h is needed. They provide theoretical discussions on the bandwidths in the univariate case, but there is no data-driven algorithm to select them. It casts doubt, especially about the computational burden, when we want to extend it for multivariate dense functional objects.

In contrast, our objects are images rather than smooth densities. We directly regress the original data on the estimated loadings, and then use penalized splines to smooth each of the factors. Our algorithm is fit for general d-dimensional array objects with huge volume, and we also provide an objective criterion to determine the appropriate amount of smoothing via a data-driven GCV algorithm.

Appendix C. Risk attitude parameter

The risk attitude parameter R is estimated by the logistic model via maximum likelihood estimation (MLE), viz.

$$\Pr(\text{risky choice}|x) = \frac{1}{1 + \exp[-\{\bar{x} - S(x)R - 5\}]}, \quad \Pr(\text{sure choice}|x) = 1 - \frac{1}{1 + \exp[-\{\bar{x} - S(x)R - 5\}]},$$

where x is the displayed return stream, $\bar{x}$ and $S(x)$ are respectively the subjective expected return and perceived risk judged by the subjects.

The estimated risk attitude parameters for the 17 subjects in order are plotted in Fig. C.1. Lower parameters imply a risk-seeking behavior, while higher parameters indicate risk-averse patterns. The two extremes #19 and #21 are the most risk-averse and most risk-seeking person, respectively. Note that while five of the 22 subjects were excluded from the sample for reasons mentioned earlier, but we did not change the labels.

References

[1] A. van Bömmel, S. Song, P. Majer, P.N.C. Mohr, H.R. Heekeren, W.K. Härdle, Risk patterns and correlated brain activities. Multidimensional statistical analysis of fMRI data in economic decision making study, Psychometrika 79 (2014) 489–514.
[2] C.F. Camerer, Neuroeconomics: Using neuroscience to make economic predictions, Econ. J. 117 (2007) C26–C42.
[3] Y. Chen, W.K. Härdle, Q. He, P. Majer, Risk Related Brain Regions Detected with 3D Image FPCA, SFB 649 Discussion Paper 2015-022, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany, 2015, available at https://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2015-022.pdf.
[4] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[5] I.D. Currie, M. Durban, P.H.C. Eilers, Generalized linear array models with applications to multidimensional smoothing, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2006) 259–280.
[6] C.Z. Di, C.M. Crainiceanu, B.S. Caffo, N.M. Punjabi, Multilevel functional principal component analysis, Ann. Appl. Stat. 3 (2009) 458.
[7] F. Ferraty, Y. Romain, The Oxford Handbook of Functional Data Analysis, Oxford University Press, Oxford, 2011.
[8] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer Science & Business Media, New York, 2006.
[9] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, C.D. Frith, R.S.J. Frackowiak, Statistical parametric maps in functional imaging: A general linear approach, Hum. Brain Mapp. 2 (1995) 189–210.
[10] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[11] H.R. Heekeren, S. Marrett, L.G. Ungerleider, The neural systems that mediate human perceptual decision making, Nat. Rev. Neurosci. 9 (2008) 467–479.
[12] R. Heller, D. Stanley, D. Yekutieli, N. Rubin, Y. Benjamini, Cluster-based analysis of FMRI data, NeuroImage 33 (2006) 599–608.

[13] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer Science & Business Media, New, York, 2012.
[14] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, New York, 2002.
[15] K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Ann. Acad. Sci. Fenn. Ser. A. 1, Math.-Phys. 37 (1947) 1–79.
[16] A. Kneip, K.J. Utikal, Inference for density families using functional principal component analysis, J. Amer. Statist. Assoc. 96 (2001) 519–542.
[17] Y. Li, D. Ruppert, On the asymptotics of penalized splines, Biometrika 95 (2008) 415–436.
[18] M.A. Lindquist, J.M. Loh, L.Y. Atlas, T.D. Wager, Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling, NeuroImage 45 (2009) S187–S198.
[19] M.A. Lindquist, T.D. Wager, Validity and power in hemodynamic response modeling: A comparison study and a new approach, Hum. Brain Mapp. 28 (2007) 764–784.
[20] M. Loève, Sur les fonctions aléatoires stationnaires de second ordre, Rev. Sci. 83 (1945) 297–303.
[21] J.S. Marron, A.M. Alonso, Overview of object oriented data analysis, Biom. J. 56 (2014) 732–753.
[22] A. Menafoglio, G. Petris, Kriging for Hilbert-space valued random fields: The operatorial point of view, J. Multivariate Anal. 146 (2016) 84–94.
[23] P.N.C. Mohr, G. Biele, H.R. Heekeren, Neural processing of risk, J. Neurosci. 30 (2010) 6613–6619.
[24] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[25] J.O. Ramsay, B.W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, New York, 2007.
[26] D. Ruppert, Selecting the number of knots for penalized splines, J. Comput. Graph. Statist. 11 (2002) 735–757.
[27] W. Schultz, Neuronal reward and decision signals: From theories to data, Physiol. Rev. 95 (2015) 853–951.
[28] A.M. Staicu, Y. Li, C.M. Crainiceanu, D. Ruppert, Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis, Scand. J. Stat. 41 (2014) 932–949.
[29] P.N. Tobler, J.P. O'Doherty, R.J. Dolan, W. Schultz, Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems, J. Neurophysiol. 97 (2007) 1621–1632.
[30] J. Wang, H. Zhu, J. Fan, K. Giovanello, W. Lin, Multiscale adaptive smoothing models for the hemodynamic response function in fMRI, Ann. Appl. Stat. 7 (2013) 904.
[31] L. Xiao, Y. Li, D. Ruppert, Fast bivariate *P*-splines: The sandwich smoother, J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (2013) 577–599.
[32] F. Yao, H.G. Müller, A.J. Clifford, S.R. Dueker, J. Follett, Y. Lin, B.A. Buchholz, J.S. Vogel, Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate, Biometrics 59 (2003) 676–685.
[33] V. Zipunnikov, B. Caffo, D.M. Yousem, C. Davatzikos, B.S. Schwartz, C. Crainiceanu, Functional principal component model for high-dimensional brain imaging, NeuroImage 58 (2011) 772–784.