Biometrika (2018), **103**, 1, *pp*. 1–16
Printed in Great Britain Advance Access publication on 31 July 2018

Functional Hybrid Factor Regression Model for Handling Heterogeneity in Imaging Studies

BY C. HUANG

Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, Florida 32304, U.S.A.

chaohuang@stat.fsu.edu

AND H. ZHU

Department of Biostatistics, The University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, North Carolina 27599, U.S.A.

htzhu@email.unc.edu

SUMMARY

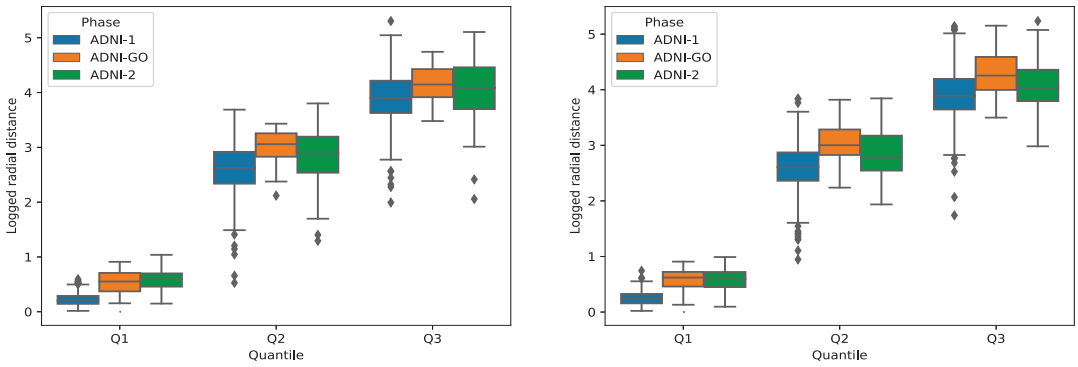
The aim of this paper is to develop a functional hybrid factor regression modeling framework to handle the heterogeneity of many large-scale imaging studies, such as Alzheimer's disease neuroimaging initiative (ADNI) study. Despite the numerous successes of those imaging studies, such heterogeneity may be caused by the differences in study environment, population, design, protocols, and some other hidden factors, and it has posed major challenges in integrative analysis of imaging data collected from multi-centers or multi-studies. We propose both estimation and inference procedures for estimating unknown parameters and detecting unknown factors under our new model. The asymptotic properties of both estimation and inference procedures are systematically investigated. The finite-sample performance of our proposed procedures is assessed by using Monte Carlo simulations and a real data example on hippocampal surface data obtained from the ADNI study.

Some key words: Alzheimer's disease; Functional hybrid factor regression model; Hippocampal surface; Imaging heterogeneity; Surrogate variable analysis.

1. INTRODUCTION

With the rapid growth of modern technology, many large-scale imaging studies, such as Alzheimer's disease neuroimaging initiative (ADNI) study (Mueller et al., 2005), Human Connectome Project (Van Essen et al., 2013), and UK Biobank study (Sudlow et al., 2015), have been conducted to collect massive datasets with large volumes of complex information from increasingly large cohorts for unraveling the etiology of different diseases, such as Alzheimer's Disease (AD). For example, the ADNI study is a multi-phase study including ADNI-1, ADNI-GO, ADNI-2, and ADNI-3 that aims to discover the progression of AD and improve clinical trials for the prevention and treatment of AD. However, such integrative data analysis is challenging largely due to the heterogeneity in those imaging studies, since those massive data sets are often collected from different centers and/or phases and needed to be rigorously integrated (Lock et al., 2013; Yu et al., 2017). The potential heterogeneity may be caused by the differences in study environment, population (e.g., race), design, protocols (e.g., imaging acquisition protocol), and some other (unknown) hidden factors in multiple centers and/or phases (Leek & Storey,

2007; Mirzaalian et al., 2016; Fortin et al., 2017). As an illustration, we consider a hippocampal surface dataset obtained from three different phases including ADNI-1, ADNI-GO, and ADNI-2 of the ADNI study. Fig. 1 presents the three quantiles including Q1, Q2, and Q3 of logged radial distances across all the vertices on the left and right hippocampal surfaces. More detailed on the calculation of radial distances will be discussed in Section 5. We observe different patterns in the quantile plots across the three phases, especially between ADNI-1 and the other two phases, indicating that the imaging heterogeneity does exist in the ADNI hippocampal surface data. Thus, appropriately handling the imaging heterogeneity can be critically important for understanding the role of imaging biomarkers in the etiological mechanism of AD. Another example on diffusion tensor imaging also illustrates the heterogeneity in different imaging data sets and can be found in Section 1 of Supplementary Material.



(a) (b)

Fig. 1: Heterogeneity in ADNI hippocampal surface dataset: (a) three quantiles of the logged radial distances across all the vertices on the left hippocampal surface; and (b) those on the right hippocampal surface for all subjects obtained from ADNI-1 [blue], ADNI-GO [orange], and ADNI-2 [green].

There are two approaches to tackle the imaging heterogeneity in the literature. The first one is the image-based meta analysis technique, in which the study-specific statistical analyses are performed (e.g., Fisher's combined probability test and Stouffer's z-transformation test) first and the results are combined afterwards (Salimi-Khorshidi et al., 2009). Although it has shown great promise for some studies with a large number of participants at each phase (or site) (Kochunov et al., 2014), this technique still suffers from at least two major limitations: (i) study-specific population might not be large enough to estimate the true biological variability in the entire population (Mirzaalian et al., 2016); and (ii) computing study-specific summary statistics can be affected by unbalanced data. For instance, the calculated variance in the z-score is highly affected by the ratio of cases over controls in each individual study, yielding inaccurate statistical inferences (Fortin et al., 2017). The second one is to apply either fixed-effect or mixed-effect models to capture the imaging heterogeneity. These methods estimate primary effects, while adjusting study (site or phase) related known covariates and (unknown) hidden factors. To identify those unknown factors, surrogate variable analysis has been developed in various genomic studies (Johnson et al., 2007; Leek & Storey, 2007, 2008; Sun et al., 2012; Wang et al., 2017; Lee et al., 2017), and recently adapted to imaging data analysis (Guillaume et al., 2018). Since sur-

rogate variable analysis assumes that massive univariate regression models share a common set of unknown factors, imaging measures are usually treated as multivariate phenotypes. However, image measures across different voxels (or grid points) are more likely to be treated as functional responses, so it is nature to use functional data analysis tools, which can explicitly account for the three key features of imaging data, including spatial smoothness, spatial correlation, and low-dimensional representation (Zhu et al., 2012). Furthermore, by applying some smoothing techniques, the noise component of image measures can be reduced and the estimates of primary effects outperform the ones under mass-univariate analysis in terms of estimation precision (Ramsay & Silverman, 2007). Therefore, it is greatly important to address the hidden factor issue in functional regression models by borrowing some ideas from surrogate variable analysis.

The aim of this paper is to develop a functional hybrid factor regression modeling framework to investigate the relationship between functional (or imaging) responses and (known) primary covariates, while adjusting for (unknown) hidden factors. Compared to existing surrogate variable analysis methods, our proposed method is the first one designed for functional data. In contrast, although some functional models also consider recovering hidden factors via functional principal component analysis (Zhu et al., 2012), they are inefficient for handling the imaging heterogeneity since the hidden factors and observed covariates are assumed to be uncorrelated. We develop a three-step estimation procedure to estimate unknown quantities in our proposed model. In addition to the estimation procedure, a global Wald-type test and a simultaneous confidence band are also constructed for coefficient functions. We also systematically investigate the asymptotic properties of estimated coefficient functions, detected hidden factors, and test statistics. Furthermore, both simulation studies and real data analysis show that our proposed method outperforms competing methods in terms of both estimation accuracy and robustness.

2. METHODS

2.1. *Functional hybrid factor regression model*

Suppose that we observe both imaging data and some covariates from n unrelated subjects. Assume that all the images have been registered to a common template, denoted as $\mathcal{S} \subset \mathbb{R}^d$. The template $\mathcal{S}$ includes n_v grid points, denoted as $s_1, \ldots, s_{n_v}$, which have a common density $p(s)$ with bounded support $supp(p) \subset \mathcal{S}$. For each registered image, it is assumed that J imaging measurements (or features) are derived at each point such that $y(s_k) = (y_{\cdot 1}(s_k), \ldots, y_{\cdot J}(s_k))$ is an $n \times J$ matrix of J features at s_k across n subjects. Let X be an $n \times p$ full column rank matrix of observed covariates including the intercept, and Z be an $n \times q$ full column rank matrix of hidden factors, where the number of latent factors, q, is unknown. Let $\mathbb{C}^2(\mathcal{S})$ denote a class of functions whose second order partial derivatives exist and are continuous everywhere in $\mathcal{S}$.

In this paper, to build up the relationship between imaging responses and both observed covariates and hidden factors, a functional hybrid factor regression model is described as below:

$$y_{\cdot j}(s) = X\beta_j(s) + Z\gamma_j(s) + \eta_{\cdot j}(s) + \epsilon_{\cdot j}(s),\ j = 1, \ldots, J, \tag{1}$$

where $\beta_j(s)$ is a $p \times 1$ vector with entries $\{\beta_{tj}(s) \in \mathbb{C}^2(\mathcal{S})\}_{t=1}^p$ representing the primary effect related to X on $y_{\cdot j}(s)$, and $\gamma_j(s)$ is a $q \times 1$ vector with entries $\{\gamma_{lj}(s) \in \mathbb{C}^2(\mathcal{S})\}_{l=1}^q$ representing the effect on $y_{\cdot j}(s)$ caused by the hidden factors Z. Moreover, let $\eta(s) = (\eta_{\cdot 1}(s), \ldots, \eta_{\cdot J}(s))$ be an $n \times J$ matrix which characterizes both subject-specific and location-specific spatial variability, and $\epsilon(s) = (\epsilon_{\cdot 1}(s), \ldots, \epsilon_{\cdot J}(s))$ be measurement errors. It is also assumed that each row in $\eta(s)$ and that in $\epsilon(s)$ are mutually independent and identical copies of $\mathrm{SP}(0, \Sigma_\eta)$ and $\mathrm{SP}(0, \Sigma_\epsilon)$, respectively, where $\mathrm{SP}(\mu, \Sigma)$ denotes a stochastic process vector with mean function $\mu(s)$ and covariance function $\Sigma(s, s')$. Moreover, $\Sigma_\epsilon(s, s')$ takes the form of $\Omega_\epsilon(s)\mathbf{1}(s = s')$, where $\Omega_\epsilon(s)$

is a diagonal matrix and $\mathbf{1}(\cdot)$ is the indicator function. As a comparison, we also consider multivariate varying coefficient models (Zhu et al., 2012) given by

$$y_{\cdot j}(s) = X\beta_j(s) + \eta_{\cdot j}(s) + \epsilon_{\cdot j}(s),\ j = 1, \ldots, J. \tag{2}$$

Here models (1) and (2) share several common features. First, both models account for the spatial smoothness, spatial correlation, and the low-dimensional representation of functional responses (Zhu et al., 2012). Second, both models are feasible to investigate the relationship between multivariate functional responses and some observed covariates of interest. Third, the individual function variations are considered through $\eta(s)$ in both models (Zhu et al., 2012). Fourth, the detection and adjustment of hidden factors are doable in both models.

However, models (1) and (2) use different strategies to handle the hidden factors. In Zhu et al. (2012), the hidden factors can be captured by the individual functions $\eta(s)$ based on the functional principal component analysis (Wang et al., 2016), where all the principal component scores can be used to recover the structure of hidden factors. A major issue associated with this strategy is that it cannot appropriately handle the case that observed covariates and hidden factors are correlated to each other. Specifically, in Zhu et al. (2012), the observed covariates X are assumed to be uncorrelated with the hidden factors in individual functions $\eta(s)$. However, such assumption may be questionable in some applications (Helmer et al., 1999; Sundström et al., 2016; Sommerlad et al., 2018) and thus, model (2) can be problematic for appropriately detecting and adjusting the hidden factors. In contrast, in model (1), the individual functions $\eta(s)$ are assumed to be uncorrelated with both observed covariates X and hidden factors Z, while no assumptions are made for the correlation between X and Z. Therefore, model (1) can handle hidden factors even when they are correlated with the observed covariates.

2.2. *Estimation procedure*

We present the estimation procedure for coefficient functions and hidden factors as follows.
Step 1: local linear kernel smoothing after reparameterization. By applying the orthogonal decomposition of the matrix Z onto the column space of X, we reparametrize model (1) as below

$$y_{\cdot j}(s) = X\beta_j^*(s) + Z^*\gamma_j(s) + \eta_{\cdot j}(s) + \epsilon_{\cdot j}(s),\ j = 1, \ldots, J, \tag{3}$$

where $\beta_j^*(s) = \beta_j(s) + (X^\top X)^{-1}X^\top Z\gamma_j(s)$, $Z^* = (I_n - P_X)Z$, and $P_X = X(X^\top X)^{-1}X^\top$. Obviously, the columns of X are orthogonal to those of Z^*. Then, given that $\{y_{\cdot j}(s)\}_{j=1}^J$ and X are observed, the multivariate local linear kernel smoothing technique (Ruppert & Wand, 1994; Fan & Gijbels, 1996) is then used to derive the weighted least squares estimator of $\beta_j^*(s)$ in (3). Let $e^{\otimes 2} = ee^\top$ for any vector e and $C \otimes D$ be the Kronecker product of two matrices C and D. In addition, denote $K_{H_\beta}(s) = |H_\beta|^{-1}K(H_\beta^{-1}s)$ and $z_{H_\beta}(s_k - s) = (1, (s_k - s)^\top H_\beta^{-1})^\top$, where $K(\cdot)$ is the kernel function, and H_β is the positive definite bandwidth matrix and $|H_\beta|$ is its determinant. For each j and fixed H_β, the estimator of $\beta_j^*(s)$ is derived as

$$\widehat{\beta}_j^*(s) = (X^\top X)^{-1}X^\top \sum_{k=1}^{n_v} \varrho_k(H_\beta, s)y_{\cdot j}(s_k), \tag{4}$$

where $\varrho_k(H_\beta, s) = (1, 0_{1\times d})\{\sum_{k=1}^{n_v} K_{H_\beta}(s_k - s)z_{H_\beta}^{\otimes 2}(s_k - s)\}^{-1}K_{H_\beta}(s_k - s)z_{H_\beta}(s_k - s)$.

Since there is no linearity assumption on the coefficient function $\beta_j^*(s)$, the local linear smoother in (4) is a biased estimator (Fan & Gijbels, 1996). To overcome this issue, a standard technique considered here is the bias correction. Following the pre-asymptotic substitution method in Fan & Gijbels (1996), the bias term can be obtained by using local cubic fit with a

pilot bandwidth selected in (4). Furthermore, according to the definition of $\beta_j^*(s)$, the aim of the following two steps is to seek an estimate of $Z\gamma_j(s)$. Then the estimate of $\beta_j(s)$ can be derived by subtracting the term $(X^\top X)^{-1}X^\top \widehat{Z\gamma_j}(s)$ from $\widehat{\beta}_j^*(s)$.

Step 2: singular value decomposition on extended residual matrix. The residual term in **Step 1** is defined as $R_{\cdot j}(s) = y_{\cdot j}(s) - X\widetilde{\beta}_j^*(s)$, where $\widetilde{\beta}_j^*(s)$ is the refined version of $\widehat{\beta}_j^*(s)$ after correcting the bias in local linear kernel smoothing technique. Next, we construct an $n \times Jn_v$ extended residual matrix written by $\bar{R} = (R_{\cdot 1}(s_1), \ldots, R_{\cdot 1}(s_{n_v}), \ldots, R_{\cdot J}(s_1), \ldots, R_{\cdot J}(s_{n_v}))$. Then, given $\mathcal{S}, X$, and Z, the conditional expectation of the extended residual matrix can be derived as (Ruppert & Wand, 1994):

$$E(\bar{R} \mid \mathcal{S}, X, Z) = Z^*\bar{\Gamma} + o_p(\mathrm{tr}(H_\beta^2)), \tag{5}$$

where $\bar{\Gamma} = (\gamma_{\cdot 1}(s_1), \ldots, \gamma_{\cdot 1}(s_{n_v}), \ldots, \gamma_{\cdot J}(s_1), \ldots, \gamma_{\cdot J}(s_{n_v}))$ and $\mathrm{tr}(\cdot)$ is the trace of a given matrix. To estimate the primary term Z^* in (5), the singular value decomposition technique is first performed on the $\bar{R}$, i.e., $\bar{R} = U\Lambda V^\top$, where the columns of U and V consist of the left and right singular vectors, respectively, and Λ is a diagonal matrix whose diagonal entries are the ordered singular values of $\bar{R}$. Specifically, the first q columns in U, denoted as $U_{1:q}$, can be treated as an estimator of linear combinations of the columns of Z^* (See Lemma 1 in Supplementary Material). Then, there exists a $q \times q$ orthonormal matrix Q such that $U_{1:q} = Z^*Q + o_p(1)$ and $Z^*\gamma_j(s) = U_{1:q}\alpha_j(s)$, where $\alpha_j(s) = Q^\top\gamma_j(s)$, $j = 1, \ldots, J$.

Step 3: correcting bias associated with the estimates in Step 1. To derive the estimate of $\alpha_j(s)$, the residual terms in **Step 2** are treated as functional responses. Then, a new varying coefficient model is constructed via substituting the singular value decomposition results:

$$R_{\cdot j}(s) = U_{1:q}\alpha_j(s) + \tilde{\eta}_{\cdot j}(s) + \tilde{\epsilon}_{\cdot j}(s),\ j = 1, \ldots, J, \tag{6}$$

where $\tilde{\eta}_{\cdot j}(s)$ and $\tilde{\epsilon}_{\cdot j}(s)$ are similarly defined as $\eta_{\cdot j}(s)$ and $\epsilon_{\cdot j}(s)$ respectively. For the fixed H_α, the estimator of $\alpha_j(s)$ can be derived as $U_{1:q}^\top \sum_{k=1}^{n_v} \varrho_k(H_\alpha, s)R_{\cdot j}(s_k)$, and $\widehat{\alpha}_j(s)$ is denoted as the corresponding bias corrected version. Then, an estimation equation can be constructed as follows:

$$X\widetilde{B}^*(s) + U_{1:q}\widehat{A}(s) = XB(s) + G\widehat{A}(s), \tag{7}$$

where $\widetilde{B}^*(s) = (\widetilde{\beta}_1^*(s), \ldots, \widetilde{\beta}_J^*(s))$, $G = ZQ$, and $\widehat{A}(s) = (\widehat{\alpha}_1(s), \ldots, \widehat{\alpha}_J(s))$. With an additional assumption that the row vectors of $B(s) = (\beta_1(s), \ldots, \beta_j(s))$ and the row vectors of $\Gamma(s) = (\gamma_1(s), \ldots, \gamma_J(s))$ are orthogonal with respect to $p(s)$ on $\mathcal{S}$ after mean centering, we can derive the estimator of G as

$$\widehat{G} = U_{1:q} + X\int_s \widetilde{B}^*(s)(I_J - P_J)\widehat{A}^\top(s)p(s)ds\Omega^{-1}, \tag{8}$$

where $\Omega = \int_s \widehat{A}(s)(I_J - P_J)\widehat{A}^\top(s)p(s)ds$ and $P_J = \mathbf{1}_J(\mathbf{1}_J^\top\mathbf{1}_J)^{-1}\mathbf{1}_J^\top$, in which $\mathbf{1}_J$ is a $J \times 1$ vector of ones. Since $Z\gamma_j(s) = G\alpha_j(s)$ for $j = 1, \ldots, J$, the estimator of $B(s)$ is given by

$$\widehat{B}(s) = \widetilde{B}^*(s) - (X^\top X)^{-1}X^\top\widehat{G}\widehat{A}(s). \tag{9}$$

2.3. *Other issues in estimation procedure*

Identification of G and $\{\alpha_j(s)\}$. According to the definitions of G and $\{\alpha_j(s)\}$, they are not identifiable due to the scaling issue. To address this issue, we impose the constraint $(nq)^{-1}\sum_{i=1}^n\sum_{j=1}^q G_{i,j}^2 = 1$, where $G_{i,j}$ is the (i,j)-th element of G and the estimated $\widehat{G}$ is adjusted to satisfy this constraint. Thus, G and $\{\alpha_j(s)\}$ are identifiable up to an orthonormal transformation only.

Smoothing individual functions. By using the smoothing method in Ruppert & Wand (1994), we smooth the individual functions of $\eta(s)$ based on the updated residual matrix as follows:

$$\widehat{\eta}(s) = \sum_{k=1}^{n_v} \varrho_k(H_\eta, s)\{y(s) - X\widehat{B}(s) - \widehat{G}\widehat{A}(s)\}, \tag{10}$$

where H_η is the fixed bandwidth matrix. Furthermore, we use the empirical covariance $\widehat{\Sigma}_\eta(s, s') = (n-p-q)^{-1}\sum_{i=1}^{n}\widehat{\eta}_{i\cdot}(s)\widehat{\eta}_{i\cdot}^T(s')$ to estimate $\Sigma_\eta(s, s')$.

Bandwidth Selection. To select the optimal bandwidth in $\widehat{B}(s)$ and $\widehat{A}(s)$, we use the leave-one-curve-out cross-validation, whereas for the optimal bandwidth in $\widehat{\eta}(s)$, we use the generalized cross validation score method (Zhang & Chen, 2007; Zhu et al., 2012). Moreover, we standardize all covariates to have mean zero and standard deviation one, as well as all functional features. Finally, we choose a common bandwidth for all covariates and features. More details can be found in Section 3 of Supplementary Material.

Determining the number of latent factors. Since the number of latent factors, q, is unknown, we consider four different methods including permutation version of the analytical-asymptotic approach (Johnstone, 2001), parallel analysis (Buja & Eyuboglu, 1992), eigenvalue difference method (Onatski, 2010), and bi-cross-validation method (Owen & Wang, 2016). We will compare all the four different methods in terms of both high detection accuracy and computation time in the simulation studies and select the one with the best performance in the rest of our data analyses.

2.4. *Inference procedure*

Hypothesis testing. We consider the linear hypotheses on $B(s)$ as below:

$$\mathrm{H}_0 : C\mathrm{vec}(B(s)) = b_0(s) \text{ for all } s \in \mathcal{S} \text{ vs. } \mathrm{H}_1 : C\mathrm{vec}(B(s)) \neq b_0(s) \text{ for some } s \in \mathcal{S}, \tag{11}$$

where C is a $r \times Jp$ matrix with rank r, $\mathrm{vec}(\cdot)$ denotes the vectorization of a given matrix, and $b_0(s)$ is a $r \times 1$ vector of functions. The global test statistic T_n is defined as:

$$T_n = \int_{s\in\mathcal{S}} T_n(s)p(s)ds \text{ with } T_n(s) = \zeta^\top(s)[C\{\widehat{\Sigma}_\eta(s,s) \otimes (\widehat{M}\widehat{M}^\top)\}C^\top]^{-1}\zeta(s), \tag{12}$$

where $\zeta(s) = C\mathrm{vec}(\widehat{B}(s)) - b_0(s)$, $\widehat{M} = (I_p, 0_{q\times q})(\widehat{W}^\top\widehat{W})^{-1}\widehat{W}^\top$, and $\widehat{W} = (X, \widehat{G})$.

As the asymptotic distribution of T_n under H_0 is quite complicated, it is difficult to derive the percentiles of T_n directly from the corresponding asymptotic results. To address this issue, the wild bootstrap method is developed (Zhu et al., 2012) consisting of the following four steps:

- Fit model (1) under H_0 on X and $\{y(s_k)\}_{k=1}^{n_v}$, yielding $\widehat{G}$, $\widehat{A}(s), \widehat{B}(s), \widehat{\eta}(s)$, $\widehat{\epsilon}(s)$, and the global test statistic T_n;
- Generate random vectors $\tau_i^{(m)}$ and $\tau_i^{(m)}(s_k)$ independently from the standard normal distribution $N(0, I_n)$ for $k = 1, \ldots, n_v$, and then construct

$$y^{(m)}(s_k) = X\widehat{B}(s_k) + \widehat{G}\widehat{A}(s_k) + \mathrm{diag}(\tau_i^{(m)})\widehat{\eta}(s_k) + \mathrm{diag}(\tau_i^{(m)}(s_k))\widehat{\epsilon}(s_k),$$

  where $\mathrm{diag}(\tau)$ denotes a diagonal matrix with the elements of τ lying along the diagonal;
- Based on X and $\{y^{(m)}(s_k)\}_{k=1}^{n_v}$, recalculate $\widehat{B}^{(m)}(s)$ and the global test statistic $T_n^{(m)}$;
- Repeat the previous two steps M times to obtain $\{T_n^{(1)}, \ldots, T_n^{(M)}\}$, which yields the empirical p-value as $p = \sum_{m=1}^{M}\mathbf{1}(T_n^{(m)} > T_n)/M$.

Simultaneous confidence bands. Construction of simultaneous confidence bands for coefficient functions is also of great interest in statistical inference for our proposed model. For a given confidence level ϑ, we construct the $1-\vartheta$ simultaneous confidence band for $\beta_{tj}(s)$

$$\left(\widehat{\beta}_{tj}(s) - n^{-1/2}C_{tj}(\vartheta),\ \widehat{\beta}_{tj}(s) + n^{-1/2}C_{tj}(\vartheta)\right), 1 \le t \le p, 1 \le j \le J, \tag{13}$$

where $C_{tj}(\vartheta)$ is a scalar, which is to be determined. Here an efficient resampling method (Kosorok, 2003; Zhu et al., 2007, 2012) is developed to approximate $C_{tj}(\vartheta)$ as follows:

- Fit model (1) on X and $\{y(s_k)\}_{k=1}^{n_v}$, yielding the residuals $\nu_{\cdot j}(s) = y(s) - X\widehat{\beta}(s) + \widehat{G}\widehat{\alpha}(s)$;
- Generate the random vector $\tau_i^{(m)}$ from the standard normal distribution $N(0, I_n)$, and then construct $\omega_{tj}^{(m)}(s) = n^{1/2}e_t^\top \widehat{M}\mathrm{diag}(\tau_i^{(m)})\sum_{k=1}^{n_v}\varrho_k(H,s)\nu_{\cdot j}(s_k)$, where e_t is a $p\times 1$ vector with the t-th element being 1 and 0 otherwise;
- Repeat the second step for M times to obtain $\{\sup_s|\omega_{tj}^{(1)}(s)|,\ldots,\sup_s|\omega_{tj}^{(M)}(s)|\}$, and use their $1-\vartheta$ empirical percentile to estimate $C_{tj}(\vartheta)$.

3. Asymptotic properties

We systematically investigate the asymptotic properties of all estimators proposed in Section 2.2 and inference procedures in Section 2.4. Assumptions used to facilitate the technical details can be found in the Supplementary Material.

Asymptotics of estimation procedure. The following theorem tackles the theoretical properties of $\widehat{B}(s)$ and $\widehat{G}$. The detailed proof can be found in the Supplementary Material.

Theorem 1. *Under Assumptions 1 – 7 , we have the following results:*
(i) The columns of $\widehat{G}$ span the same column space as the columns of Z in probability.
(ii) $n^{1/2}[\{I_J\otimes(\widehat{M}\widehat{M}^T)^{-1/2}\}\mathrm{vec}(\widehat{B}(s)-E[\widehat{B}(s)])\mid s\in\mathcal{S}]$ weakly converges to a centered Gaussian process with covariance matrix $\Sigma_\eta(s,s)\otimes I_p$.

Asymptotics of inference procedure. The following theorem derives the asymptotic distribution of global test statistic T_n under the null hypothesis and its asymptotic power under local alternative hypotheses. The detailed proof can be found in the Supplementary Material.

Theorem 2. *Under Assumptions 1 – 9, we have the following results:*
(i) $T_n \to \int_s \xi(s)^\top\xi(s)ds$ as $n\to\infty$, where $\xi(s)$ is a centered Gaussian process.
(ii) $P\{T_n > T_{n,\vartheta}\mid \mathrm{H}_{1n}\}\to 1$ as $n\to\infty$ for a sequence of local alternatives H_{1n}: $C\mathrm{vec}(B(s)) - b_0(s) = n^{-\tau/2}\zeta(s)$, where τ is any scalar in $[0,1)$, $T_{n,\vartheta}$ is the upper 100ϑ percentile of T_n under H_0, and $0<\int_s\|\zeta(s)\|^2ds<\infty$.

4. Simulation studies

To examine the proposed methods, we generated synthetic curves from the following model

$$y_{ij}(s_k) = x_i^\top\beta_j(s_k) + z_i\gamma_j(s_k) + \eta_{ij}(s_k) + \epsilon_{ij}(s_k),\ j=1,2, \tag{14}$$

where $s_1 = 0 \le s_2 \le \cdots \le s_{n_v} = 1$, in which we independently simulated $\tilde{s}_k \sim U(0,1)$ for $k = 2,\ldots,n_v-1$ and sorted them to obtain $\{s_k : k = 2,\ldots,n_v-1\}$. We set $x_i = (1, x_{i1}, x_{i2}, x_{i3})^\top$, in which we independently simulated $x_{i1}\sim$ Bernoulli(0.5), $x_{i2}\sim N(0,1)$, and $x_{i3}\sim N(0,1)$ for $i=1,\ldots,n$. We simulated z_i as follows:

$$z_i = x_i^\top\varphi + \omega_i,\ \omega_i\sim N(0,1),\ i=1,\ldots,n, \tag{15}$$

where $\varphi = (u_1(2b_1 - 1), u_2(2b_2 - 1), u_3(2b_3 - 1), u_4(2b_4 - 1))^\top$ with b_l being independently generated from Bernoulli(0.5). We independently simulated u_l for all l and consider four different simulation scenarios on u_l including (i) $u_l = 0$; (ii) $u_l \sim U(0, 0.2)$; (iii) $u_l \sim U(0.2, 0.5)$; and (iv) $u_l \sim U(0.5, 1)$. Those scenarios correspond to hidden factors Z being (i) independent with X, (ii) weakly correlated with X, (iii) moderately correlated with X, and (iv) highly correlated with X, respectively. The $\eta_{ij}(s)$ admits the Karhunen Loeve expansion as $\eta_{ij} = \xi_{ij1}\psi_{j1}(s) + \xi_{ij2}\psi_{j2}(s)$, where $\psi_{jl}(s)$ are the eigen functions and $\xi_{ijl} \sim N(0, 0.5)$ for $j = 1, 2$ and $l = 1, 2$. We simulated $(\epsilon_{i,1}, \epsilon_{i,2})^\top \sim N((0, 0)^\top, 0.5 * \text{diag}(\sigma_1^2, \sigma_2^2))$, where $\sigma_l^2 \sim$ Inverse-Gamma$(10, 9)$ for $l = 1, 2$. Also, we set the following functions:

$$\beta_1(s) = (3s^2, 3(1-s)^2, 6s(1-s), -s^2)^\top, \quad \gamma_1(s) = -\sqrt{2}\sin(\pi s),$$
$$\beta_2(s) = (12(s-0.5)^2, 1.5\sqrt{s}, 3s^2, -2s/3)^\top, \quad \gamma_2(s) = \sqrt{2}\cos(2\pi s),$$
$$\psi_{11}(s) = 0.5, \ \psi_{12}(s) = s - 0.5, \ \psi_{21}(s) = 2s - 1, \text{ and } \psi_{22}(s) = 1.$$

Without special saying, we set $n = 50$ and $n_v = 2000$. Finally, we generated 200 datasets for each simulation scenario.

We compare our method with two other methods including multivariate varying coefficient model in Zhu et al. (2012) and confounder adjustment method in Wang et al. (2017). For the method in Wang et al. (2017), the curved data is treated as multivariate responses. To evaluate the finite-sample performance of each method, we consider the integrated square error, i.e., $\sum_{l=1}^{2} \int_0^1 ||\widehat{\beta}_l(s) - \beta_l(s)||^2 ds$, where $\widehat{\beta}_j(s)$ is any estimator of $\beta_j(s)$. For both the method in Wang et al. (2017) and our method, the eigenvalue difference method (Onatski, 2010) is used to estimate the number of factors.

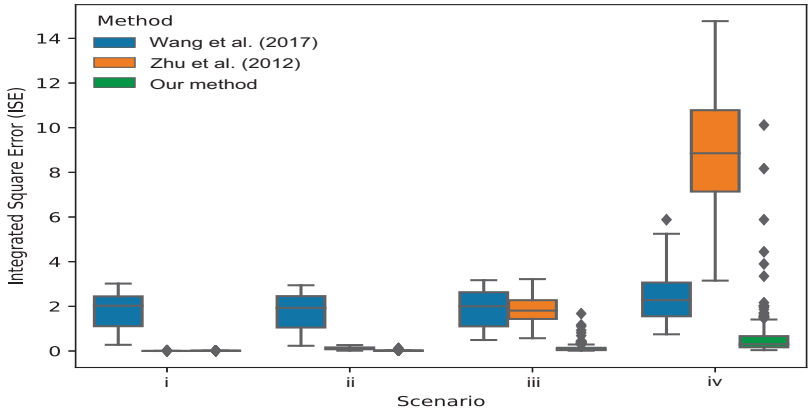


Fig. 2: Simulation results for comparisons of the proposed and competing methods on synthetic curve data in terms of integrated square error. Four scenarios were considered: the hidden factors Z are (i) independent with X, (ii) weakly correlated with X, (iii) moderately correlated with X, and (iv) highly correlated with X, respectively. The methods compared are: method in Wang et al. (2017) [blue]; method in Zhu et al. (2012) [orange]; and our method [green].

Fig. 2 presents the comparison results for all the three methods in all the four scenarios. Inspecting Fig. 2 reveals the following results. First, compared with the method in Zhu et al. (2012), both the method in Wang et al. (2017) and our method are very stable and robust to the correlation between X and Z. Second, our method outperforms the one in Wang et al. (2017) for all scenarios, indicating that it is critically important to use the functional data analysis tools. Third,

when Z and X are independent, the difference between our method and the one in Zhu et al. (2012) is very small. Fourth, when the correlation between X and Z is high in scenario (iv), the integrated square errors based on the method in Zhu et al. (2012) dramatically increase. In contrast, those of our method are much smaller even though there are a few outliers, which are caused by the uncertainty of estimating q as detailed below.

We compare four estimation methods for the number of hidden factors, including the analytical-asymptotic approach in Johnstone (2001), the permutation version of the parallel analysis in Buja & Eyuboglu (1992), the eigenvalue difference method in Onatski (2010), and the bi-cross-validation method in Owen & Wang (2016). Table 1 reports the estimation results for all the four methods. We observe that the last three methods can achieve almost 100 percent estimation accuracy, while outperforming the analytical-asymptotic approach with low estimation accuracy around 30%. In addition, in terms of average computation time, the eigenvalue difference method (Onatski, 2010) is much more efficient than the bi-cross-validation method (Owen & Wang, 2016) and the parallel analysis approach (Buja & Eyuboglu, 1992). Thus, the eigenvalue difference method is used in subsequent analyses.

Table 1: Comparison of four methods for estimating the number of hidden factor with $q = 1$. The average computation time for each method is reported as well. The four scenarios considered include (i) Z being independent with X, (ii) Z being weakly correlated with X, (iii) Z being moderately correlated with X, and (iv) Z being highly correlated with X

Method	Scenario				Average computation time
	(i)	(ii)	(iii)	(iv)	(seconds per data set)
Johnstone (2001)	62/200	65/200	64/200	64/200	0.1
Buja & Eyuboglu (1992)	190/200	191/200	192/200	191/200	70.2
Onatski (2010)	200/200	200/200	198/200	198/200	0.8
Owen & Wang (2016)	200/200	196/200	196/200	196/200	9.7

We investigate the sensitivity of our method with respect to the misspecification of q under the four scenarios since there are some outliers in Fig. 2 for our method when Z and X are highly correlated. We also consider three choices of q including $q = 0, 1,$ and 2, which represent the under-estimated, true, and over-estimated values, respectively. Fig. 3 presents the box plots of integrated square errors for all q values under the four scenarios. There are three major findings. First, when the hidden factor Z is independent or weakly correlated with X, the integrated square errors are relatively stable even when q is misspecified. Second, when Z is moderately or highly correlated with X, the integrated square errors dramatically increase for missepecified q values. Third, the under-estimated $q = 0$ has much larger effects on integrated square errors than the over-estimated $q = 2$.

We examine the correlation between the space spanned by the columns of detected latent factors with that spanned by the columns of true Z. Fig. 4 presents simulation results in all the four scenarios with the absolute values of Pearson correlation between $\widehat{G}$ and Z being greater than 0.90, indicating their consistency with each other. Moreover, when the correlation between Z and X gets higher, the absolute values of Pearson correlation coefficient are closer to 1.

We examine the type I and II error rates of T_n. For the sake of space, we only consider the third scenario (iii), in which $\varphi = (u_1, -u_2, u_3, -u_4)^T$ with independently simulating u_l from $U(0.2, 0.5)$ for all $l = 1, 2, 3, 4$. Moreover, we fix all other parameters at their values specified above except that we set $\beta_{14}(s) = -cs^2$ and $\beta_{24}(s) = -2cs/3$, where c is a scalar specified

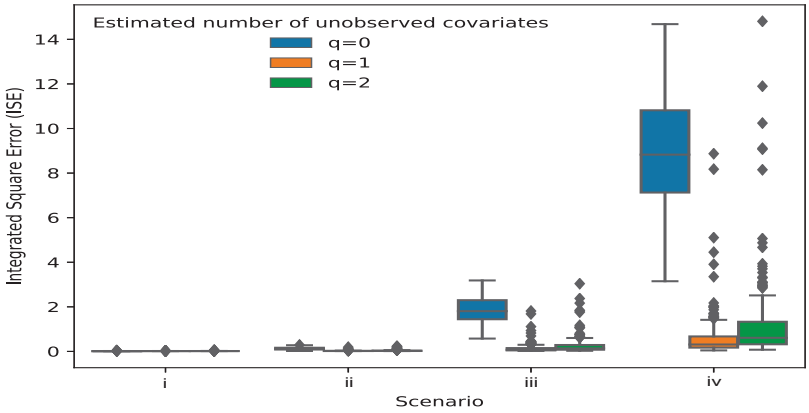


Fig. 3: Simulation results for the sensitivity analysis of our method under the three choices of q in the four scenarios including (i) Z being independent with X, (ii) Z being weakly correlated with X, (iii) Z being moderately correlated with X, and (iv) Z being highly correlated with X.

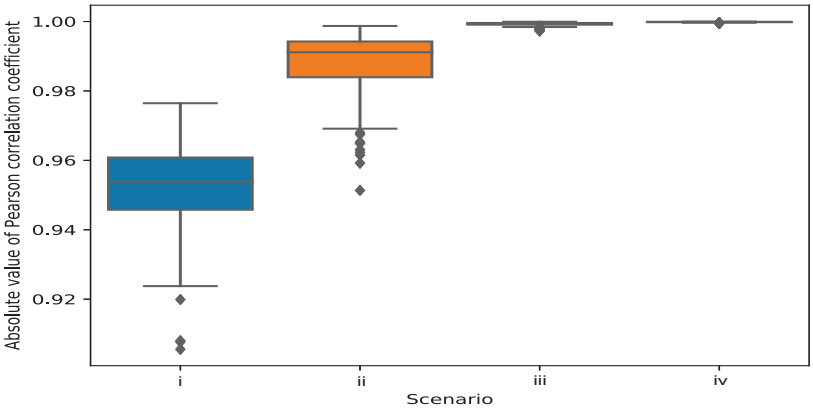


Fig. 4: Simulation results for the the absolute values of Pearson correlation between $\widehat{G}$ and Z. Four scenarios include (i) Z being independent with X, (ii) Z being weakly correlated with X, (iii) Z being moderately correlated with X, and (iv) Z being highly correlated with X.

below. We want to test the following hypotheses

$$\mathrm{H}_0 : \beta_{14}(s) = \beta_{24}(s) = 0 \text{ for all } s \text{ v.s. } \mathrm{H}_1 : \beta_{14}(s) \neq 0 \text{ or } \beta_{24}(s) \neq 0 \text{ for at lease one } s. \quad (16)$$

We set $c = 0$ to assess the type I error rates for T_n, and set $c = 0.1, 0.2, 0.3, 0.4$, and 0.5 to examine the power of T_n. We set the sample size as $n = 100$ and 200. For each case, 500 bootstrap replications were generated to construct the empirical distribution of T_n under H_0. Fig. 5 presents the power curves at the significance levels $\alpha = 0.05$ and 0.01. The rejection rates for T_n based on the wild bootstrap method are accurate for moderate sample sizes with $n = 100$ and 200 at both significance levels $\alpha = 0.01$ and 0.05. As expected, the power increases with the sample size.

Finally, we investigate the coverage probabilities of simultaneous confidence bands for the functional coefficients in $B(s)$ based on the resampling method. We only consider the third scenario (iii). We fix all parameters specified above except that we set $n = 200$ and the number

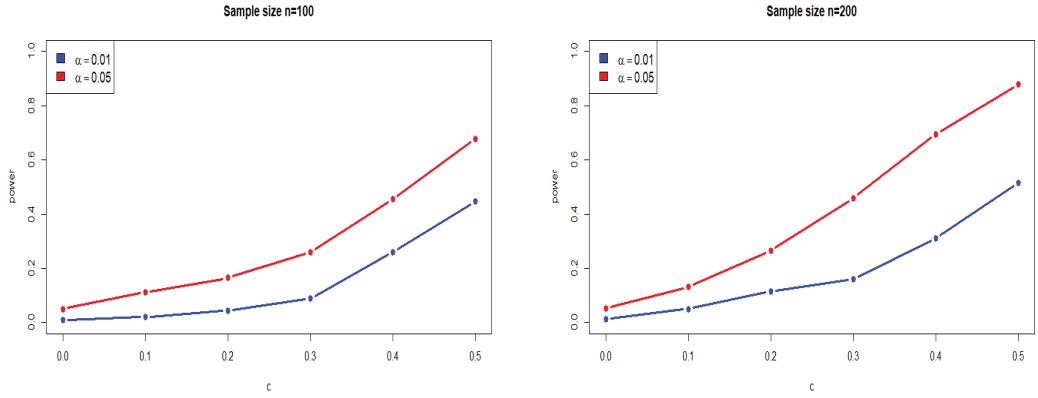


Fig. 5: Power curves for hypothesis testing problem (16) based on our method with different choice of c in $\beta_{14}(s)$ and $\beta_{24}(s)$.

of grid points $n_v = 200$ and 2000. We calculated the simultaneous confidence bands for each component in $B(s)$ based on 200 replications. Table 2 summarizes the empirical coverage probabilities at $\alpha = 0.05$ and 0.01. As expected, the coverage probabilities improve as the number of grid points n_v increases.

Table 2: Empirical coverage probabilities of $1 - \alpha$ simultaneous confidence bands

α	n_v	β_{11}	β_{12}	β_{13}	β_{14}	β_{21}	β_{22}	β_{23}	β_{24}
	200	0.935	0.920	0.925	0.920	0.915	0.915	0.930	0.940
0.05	2000	0.945	0.950	0.950	0.950	0.945	0.945	0.955	0.950
	200	0.985	0.990	0.995	0.980	0.980	0.995	0.990	0.990
0.01	2000	0.990	0.995	0.990	0.995	0.995	0.995	0.990	0.995

5. REAL DATA ANALYSIS

5.1. *Data processing*

In this data analysis, we consider 936 MRI scans from normal controls (NC) and individuals with mild cognitive impairment (MCI) or AD from three different phases including ADNI-1, ADNI-GO, and ADNI-2. Table 3 summarizes the demographic information of all the subjects.

Table 3: Hippocampal surface data: demographic information of 936 subjects

Phase	ADNI-1	ADNI-GO	ADNI-2	Total
Size	800	24	112	936
Gender (F/M)	465/335	13/11	61/51	539/397
Handedness (R/L)	738/62	20/4	9/103	861/75
Age range (years)	[58, 95]	[55, 84]	[53, 87]	[53, 95]
Edu. length range (years)	[4, 20]	[12, 20]	[8, 20]	[4, 20]
Disease (NC/MCI/AD)	224/389/187	0/24/0	29/58/25	253/471/212

We processed the MRI data by using standard steps and generated one-to-one hippocampal surface registration in (Shi et al., 2013). Then, we computed the various surface statistics on the

registered surface, such as multivariate tensor-based morphometry statistics, which retain the full tensor information of the deformation Jacobian matrix, together with the radial distance, which retains information on the deformation along the surface normal direction. More detailed image data processing procedures can be found in Supplementary Material.

5.2. *Data analysis*

The hippocampus is believed to be involved in memory, spatial navigation and memory, and behavioral inhibition. In AD, the hippocampus is one of the first regions of the brain to be affected, leading to the confusion and loss of memory so commonly seen in the early stages of the disease (Kong et al., 2019). The objective of this data analysis is to integrate the data from three different data phases (i.e., ADNI-1, ADNI-GO, and ADNI-2) and exam the effects of clinical variables and demographic variables on either the left or right hippocampus. Moreover, the hidden factors are expected to be recovered and discussed. Before conducting this analysis, we would like to check if there is any heterogeneity caused by phases. According to Fig. 1 and the related discussion in Introduction section, this study-level heterogeneity does exist in the ADNI hippocampal surface data. Therefore, the phase information should be included as predictors in the data analysis.

We applied our new method with either the left or right hippocampal surface data as the functional responses. Moreover, the method in Zhu et al. (2012) was used for comparison. Specifically, we consider four imaging measurements including the logged radial distance and three tensor-based morphometry statistics measured over 7,500 vertices on the hippocampal surface (3,750 on each side). In this case, we have $J = 4$. Moreover, we included an intercept, gender, handedness, education length, age, diagnostic information, and phase information as predictors in X. Subsequently, we test the effects of all the primary variables on all the four functional responses on hippocampal surfaces. We calculated the global test statistic for each predictor and used 500 replications in wild bootstrap approach. Table 4 summarizes the corresponding $p-$values with those p-values less than the significant level 5% highlighted in red. Given the significant level 0.05, both disease (AD vs. NC) and age are found to be significant on the left hippocampal surface based on the method in Zhu et al. (2012). In contrast, more predictors are found to be significant based on our method. For example, significant age effect is found on the left hippocampal surface, while both education length effect and disease effect (AD vs. NC) are significant on left and right hippocampal surfaces. Among all these variables, education length is found to be significant in our method, but not in the competing method. Education length is an important factor for the changes of hippocampus structure in the literature (Arenaza-Urquijo et al., 2013).

Table 4: Hippocampal surface data: comparison of $p-$values for primary variables

Variable	p-value			
	Left Hippocampus		Right Hippocampus	
	Zhu et al. (2012)	Our method	Zhu et al. (2012)	Our method
Gender	0.212	0.092	0.234	0.116
Handedness	0.652	0.102	0.704	0.082
Education length	0.132	0.036	0.244	0.048
Age	0.048	0.048	0.096	0.052
MCI vs. NC	0.156	0.066	0.082	0.064
AD vs. NC	0.046	0.034	0.054	0.040
ADNI-GO vs. ADNI-1	0.134	0.112	0.136	0.120
ADNI-2 vs. ADNI-1	0.118	0.106	0.112	0.114

Furthermore, we are also interested in detecting significant subregions by using the local test statistic and the false discovery rate (Benjamini & Yekutieli, 2001). Fig. 6 presents the false discovery rate adjusted $-\log_{10}(p)$-value maps. To better understand the significant subregions, we consider the cytoarchitectonic subregions mapped on blank MR-based models at 3T of the hippocampal formation (Frisoni et al., 2008) presented in the right-hand side of Fig. 6. All the significant subregions associated with age and disease circled in red are found in the CA1 subfield. Similar hippocampal subregions were found to be affected in AD (Frisoni et al., 2008), indicating that the findings based on our method are in agreement with those in the literature.

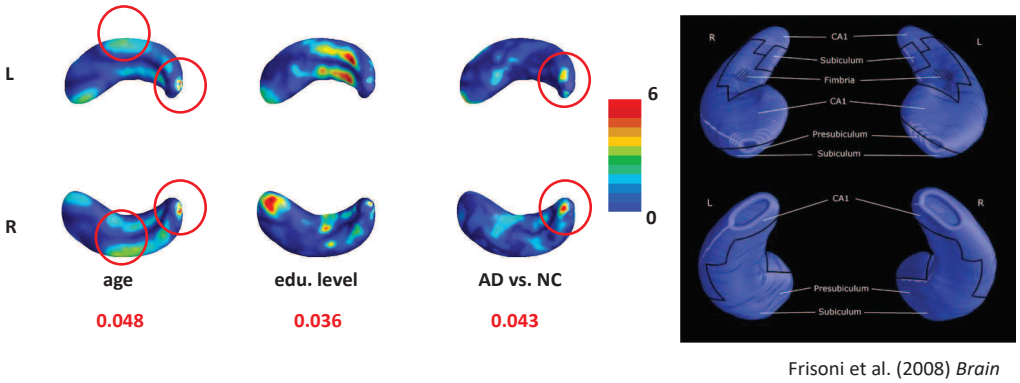


Fig. 6: Hippocampal surface data: adjusted $-\log_{10}(p)$-value maps (left) and the cytoarchitectonic subregions mapped on blank MR-based models at 3T of the hippocampal formation (right).

We investigate the potential hidden factors estimated by our method. Applying the eigenvalue difference method yields three hidden factors. Table 5 presents the correlation between primary variables and detected hidden factors. Specifically, we calculated the Pearson correlation between two continuous variables and the polyserial correlation between a continuous variable and a discrete one. Inspecting Table 5 reveals that on both left and right hippocampal surfaces, the detected factors are highly related to education length, age, disease status, and phase information. In contrast, for the method in Zhu et al. (2012), the key assumption that the hidden factors and primary variables are uncorrelated is inappropriate.

Finally, we investigate whether there are any other variables not included in our current analysis that may be strongly correlated with the latent factors. We consider 7 new variables in three categories including ethnic group information (three dummy variables were introduced to represented Asian, African American, and White), marital status (three dummy variables were introduced to represented widow, divorce and no-married), and retirement status. There are several reasons that we do not included the new regressors in the main model at the beginning. First, we only include a standard set of covariates, which have been widely considered in the existing literature (Kong et al., 2019), in the main model. Second, we apply our proposed method to detect some hidden factors that cannot be explained by the existing covariates. Third, we correlate the hidden factors with a set of new regressors and find that these regressors can partially explain these factors. This process also illustrates the importance of our FHFRM. Another reason is that there are many missing data in these new regressors. Specifically, the missing data rates for the new regressors in three categories are: 9.8% for ethnic group information, 10.9% for marital status, and 9.4% for retirement status. We observe that on the left hippocampal surface, the detected hidden factors are strongly correlated with all of them, whereas on the right hippocampal surface,

Table 5: Hippocampal surface data: correlations between primary variables and detected hidden factors and their associated $p-$values in the paratheses

Primary variable	Hidden factor					
	Left Hippocampus			Right Hippocampus		
	factor 1	factor 2	factor 3	factor 1	factor 2	factor 3
Gender	-0.038	0.015	-0.048	0.006	0.023	-0.045
	(0.358)	(0.724)	(0.239)	(0.883)	(0.582)	(0.278)
Handedness	-0.013	-0.041	0.076	0.041	-0.055	0.047
	(0.835)	(0.517)	(0.209)	(0.494)	(0.382)	(0.435)
Education length	-0.021	0.024	0.090	0.058	0.014	0.074
	(0.531)	(0.466)	(0.006)	(0.078)	(0.665)	(0.025)
Age	0.120	0.089	-0.079	-0.163	0.071	-0.131
	(<0.001)	(0.007)	(0.015)	(<0.001)	(0.030)	(<0.001)
MCI vs. NC	-0.045	0.061	0.020	0.064	0.003	0.062
	(0.272)	(0.144)	(0.617)	(0.119)	(0.944)	(0.131)
AD vs. NC	0.087	-0.058	0.061	-0.094	-0.029	-0.008
	(0.041)	(0.228)	(0.507)	(0.039)	(0.530)	(0.853)
ADNI-GO vs. ADNI-1	-0.305	0.392	0.215	0.440	-0.176	0.403
	(<0.001)	(<0.001)	(0.011)	(<0.001)	(0.064)	(<0.001)
ADNI-2 vs. ADNI-1	-0.221	-0.318	0.213	0.271	-0.469	0.466
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)

the detected hidden factors are only correlated with the marital status. More detailed results can be found in Supplementary Material.

6. DISCUSSION

The key assumption of our method is Assumption 6, which requires the row vectors of $B(s)$ and the row vectors of $\Gamma(s)$ are orthogonal with respect to the underlying density function $p(s)$ after mean centering. Similar assumptions for model identification can be found in some existing methods (Sun et al., 2012; Lee et al., 2017). Actually, this assumption is reasonable in many imaging studies. For example, in neuroimage data analysis, batch effects are usually caused by the heterogeneity in imaging acquisition protocols. Their effect sizes would not be correlated with those of population differences or diagnostic status (Lee et al., 2017). Also, our simulation studies show that our method is robust even when this assumption is violated. Specifically, in our simulation settings, when $\| \int_s B(s)(I_J - P_J)\Gamma^\top(s)p(s)ds\|_1 = 3.544$, indicating that this assumption does not hold, our method still outperforms the two competing methods.

Besides the assumption on functional coefficients, modeling of latent factors Z is also a key term in our method. In this paper, we treat the latent factors as fixed. However, to account for the imaging heterogeneity, it will be more flexible to assume that the latent factors are random. For example, Wang et al. (2017) modeled the latent factors Z through a linear model on primary variables X, i.e., $Z = X\alpha^\top + W$ and W is normally distributed. Therefore, it is important to extend our model in this paper to handle the random setting of latent factors, which will be our future work.

Another interesting topic is to extend our method to some unsupervised or semi-supervised learning, whose goal is to recover the sub-group structure within the functional data when the sub-group information is unknown or not completely observable. It is challenging because unwanted variations may be correlated with the sub-group information. For example, it is of great interest to conduct the clustering analysis in terms of brain atrophy variations among patients

with Alzheimer's disease (Poulakis et al., 2018), and there is increasing evidence that the patients' cluster information has strong association with some unknown factors like marital status (Sommerlad et al., 2018). Thus, it is interesting to extend our model to simultaneously investigate the latent sub-group structure, while accounting for unknown latent factors. We leave those extensions for our future research.

SUPPLEMENTARY MATERIAL

Supplementary material available at Biometrika online includes another example illustrating the heterogeneity in different imaging datasets, ADNI data description, image data processing, bandwidth selection, assumptions of theorems, proofs of the theoretical results, and additional simulation & real data analysis results.

REFERENCES

ARENAZA-URQUIJO, E. M., LANDEAU, B., LA JOIE, R., MEVEL, K., MÉZENGE, F., PERROTIN, A., DESGRANGES, B., BARTRÉS-FAZ, D., EUSTACHE, F. & CHÉTELAT, G. (2013). Relationships between years of education and gray matter volume, metabolism and functional connectivity in healthy elders. *NeuroImage* **83**, 450–457.

BENJAMINI, Y. & YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188.

BUJA, A. & EYUBOGLU, N. (1992). Remarks on parallel analysis. *Multivariate Behav. Res.* **27**, 509–540.

FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

FORTIN, J.-P., PARKER, D., TUNÇ, B., WATANABE, T., ELLIOTT, M. A., RUPAREL, K., ROALF, D. R., SATTERTHWAITE, T. D., GUR, R. C. & GUR, R. E. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage* **161**, 149–170.

FRISONI, G. B., GANZOLA, R., CANU, E., RÜB, U., PIZZINI, F. B., ALESSANDRINI, F., ZOCCATELLI, G., BELTRAMELLO, A., CALTAGIRONE, C. & THOMPSON, P. M. (2008). Mapping local hippocampal changes in alzheimer's disease and normal ageing with mri at 3 tesla. *Brain* **131**, 3266–3276.

GUILLAUME, B., WANG, C., POH, J., SHEN, M. J., ONG, M. L., TAN, P. F., KARNANI, N., MEANEY, M. & QIU, A. (2018). Improving mass-univariate analysis of neuroimaging data by modelling important unknown covariates: application to epigenome-wide association studies. *NeuroImage* **173**, 57–71.

HELMER, C., DAMON, D., LETENNEUR, L., FABRIGOULE, C., BARBERGER-GATEAU, P., LAFONT, S., FUHRER, R., ANTONUCCI, T., COMMENGES, D. & ORGOGOZO, J. (1999). Marital status and risk of alzheimer disease: a french population-based cohort study. *Neurology* **53**, 1953–1953.

JOHNSON, W. E., LI, C. & RABINOVIC, A. (2007). Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* **8**, 118–127.

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* **29**, 295–327.

KOCHUNOV, P., JAHANSHAD, N., SPROOTEN, E., NICHOLS, T. E., MANDL, R. C., ALMASY, L., BOOTH, T., BROUWER, R. M., CURRAN, J. E. & DE ZUBICARAY, G. I. (2014). Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and megaanalytical approaches for data pooling. *NeuroImage* **95**, 136–150.

KONG, D., AN, B., ZHANG, J. & ZHU, H. (2019). L2rm: Low-rank linear regression models for high-dimensional matrix responses. *J. Am. Stat. Assoc.* .

KOSOROK, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivar. Anal.* **84**, 299–318.

LEE, S., SUN, W., WRIGHT, F. A. & ZOU, F. (2017). An improved and explicit surrogate variable analysis procedure by coefficient adjustment. *Biometrika* **104**, 303–316.

LEEK, J. T. & STOREY, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, e161.

LEEK, J. T. & STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. U.S.A* **105**, 18718–18723.

LOCK, E. F., HOADLEY, K. A., MARRON, J. S. & NOBEL, A. B. (2013). Joint and individual variation explained (jive) for integrated analysis of multiple data types. *Ann. Appl. Stat.* **7**, 523.

MIRZAALIAN, H., NING, L., SAVADJIEV, P., PASTERNAK, O., BOUIX, S., MICHAILOVICH, O., GRANT, G., MARX, C., MOREY, R. A. & FLASHMAN, L. (2016). Inter-site and inter-scanner diffusion mri data harmonization. *NeuroImage* **135**, 311–323.

MUELLER, S. G., WEINER, M. W., THAL, L. J., PETERSEN, R. C., JACK, C., JAGUST, W., TROJANOWSKI, J. Q., TOGA, A. W. & BECKETT, L. (2005). The alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. N. Am.* **15**, 869–877.

ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92**, 1004–1016.

OWEN, A. B. & WANG, J. (2016). Bi-cross-validation for factor analysis. *Stat. Sci.* **31**, 119–139.

POULAKIS, K., PEREIRA, J. B., MECOCCI, P., VELLAS, B., TSOLAKI, M., KŁOSZEWSKA, I., SOININEN, H., LOVESTONE, S., SIMMONS, A., WAHLUND, L.-O. & WESTMAN, E. (2018). Heterogeneous patterns of brain atrophy in alzheimer's disease. *Neurobiol. Aging* **65**, 98–108.

RAMSAY, J. O. & SILVERMAN, B. W. (2007). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.

RUPPERT, D. & WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Stat.* **22**, 1346–1370.

SALIMI-KHORSHIDI, G., SMITH, S. M., KELTNER, J. R., WAGER, T. D. & NICHOLS, T. E. (2009). Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *NeuroImage* **45**, 810–823.

SHI, J., THOMPSON, P. M., GUTMAN, B. & WANG, Y. (2013). Surface fluid registration of conformal representation: application to detect disease burden and genetic influence on hippocampus. *NeuroImage* **78**, 111–134.

SOMMERLAD, A., RUEGGER, J., SINGH-MANOUX, A., LEWIS, G. & LIVINGSTON, G. (2018). Marriage and risk of dementia: systematic review and meta-analysis of observational studies. *J. Neurol. Neurosurg. Psychiatry* **89**, 231–238.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J. & LANDRAY, M. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779.

SUN, Y., ZHANG, N. R. & OWEN, A. B. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the agemap gene expression data. *Ann. Appl. Stat.* **6**, 1664–1688.

SUNDSTRÖM, A., WESTERLUND, O. & KOTYRLO, E. (2016). Marital status and risk of dementia: a nationwide population-based prospective study from sweden. *BMJ Open* **6**, e008565.

VAN ESSEN, D. C., SMITH, S. M., BARCH, D. M., BEHRENS, T. E., YACOUB, E., UGURBIL, K. & CONSORTIUM, W.-M. H. (2013). The wu-minn human connectome project: an overview. *NeuroImage* **80**, 62–79.

WANG, J., ZHAO, Q., HASTIE, T. & OWEN, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *Ann. Stat.* **45**, 1863–1894.

WANG, J.-L., CHIOU, J.-M. & MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat.* **3**, 257–295.

YU, Q., RISK, B. B., ZHANG, K. & MARRON, J. (2017). Jive integration of imaging and behavioral data. *NeuroImage* **152**, 38–49.

ZHANG, J. & CHEN, J. (2007). Statistical inference for functional data. *Ann. Stat.* **35**, 1052–1079.

ZHU, H., IBRAHIM, J. G., TANG, N., ROWE, D. B., HAO, X., BANSAL, R. & PETERSON, B. S. (2007). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans. Med. Imag.* **26**, 954–966.

ZHU, H., LI, R. & KONG, L. (2012). Multivariate varying coefficient model for functional responses. *Ann. Stat.* **40**, 2634–2666.

[*Received on* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]