




Regression Analysis of Asynchronous Longitudinal Functional and Scalar Data

Ting Li , Tengfei Li , Zhongyi Zhu & Hongtu Zhu


To cite this article: Ting Li , Tengfei Li , Zhongyi Zhu & Hongtu Zhu (2020): Regression Analysis of Asynchronous Longitudinal Functional and Scalar Data, Journal of the American Statistical Association, DOI: [10.1080/01621459.2020.1844211](https://doi.org/10.1080/01621459.2020.1844211)

To link to this article: <https://doi.org/10.1080/01621459.2020.1844211>

 View supplementary material 

 Published online: 08 Dec 2020.

 Submit your article to this journal 

 Article views: 368

 View related articles 

 View Crossmark data 



Regression Analysis of Asynchronous Longitudinal Functional and Scalar Data

Ting Li^a, Tengfei Li^b, Zhongyi Zhu^c, and Hongtu Zhu^d

^aSchool of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China; ^bDepartment of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina at Chapel Hill, Chapel Hill, NC; ^cDepartment of Statistics, Fudan University, Shanghai, China; ^dDepartment of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC

ABSTRACT

Many modern large-scale longitudinal neuroimaging studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, have collected/are collecting asynchronous scalar and functional variables that are measured at distinct time points. The analyses of temporally asynchronous functional and scalar variables pose major technical challenges to many existing statistical approaches. We propose a class of generalized functional partial-linear varying-coefficient models to appropriately deal with these challenges through introducing both scalar and functional coefficients of interest and using kernel weighting methods. We design penalized kernel-weighted estimating equations to estimate scalar and functional coefficients, in which we represent functional coefficients by using a rich truncated tensor product penalized B-spline basis. We establish the theoretical properties of scalar and functional coefficient estimators including consistency, convergence rate, prediction accuracy, and limiting distributions. We also propose a bootstrap method to test the nullity of both parametric and functional coefficients, while establishing the bootstrap consistency. Simulation studies and the analysis of the ADNI study are used to assess the finite sample performance of our proposed approach. Our real data analysis reveals significant relationship between fractional anisotropy density curves and cognitive function with education, baseline disease status and APOE4 gene as major contributing factors. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received July 2019
Accepted October 2020

KEYWORDS

Asynchronous longitudinal functional data; Functional regression; Generalized functional partial linear model; Kernel-weighted estimation equations; Penalized B-spline

1. Introduction

This article is motivated by exploring the association between neuroimaging measures and cognitive function in patients with Alzheimer's disease (AD) through the analysis of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (<http://www.adni-info.org/>). The ADNI is a large-scale multi-site neuroimaging study that has collected clinical, imaging, genetic and cognitive data at multiple time points from cognitive normal (CN) subjects, subjects with mild cognitive impairment (MCI), and AD patients. In ADNI, Mini-Mental State Examination (MMSE) score has been used to assess cognitive mental status over time, with lower scores indicating impairment. The primary goal of ADNI is to test whether genetic, structural and functional neuroimaging, and clinical data can be integrated to assess the progression of MCI and early AD.

We consider a subset of ADNI with $n = 256$ subjects over a 5-year follow-up to examine the association between the change of diffusion weighted imaging (DWI) measures and cognitive decline. Fractional anisotropy (FA) is one of the most popular DWI measures that reflects fiber density, axonal diameter, and myelination in white matter. Previous studies based on FA have shown marked damage and dysfunction in the white matter in AD (Bozzali et al. 2002; Zhang et al. 2009; Nir et al. 2013). For each subject, we consider the log-hazard curves of FA in the

whole brain and 20 regions of interest (ROIs) observed at 1–8 time points and the MMSE scores examined at 1–7 time points. The observed FA curves and MMSE scores are mismatched over time within individuals, leading to asynchronous longitudinal functional and scalar data. In Figure 1, we display observations of log hazard curves of FA along the whole brain and the MMSE scores for two subjects over time. It is clear that the observations times for the functional covariate and the outcome are mismatched within and between subjects. Given this new data structure, there is a particular need for the development of longitudinal regression models for establishing the association between asynchronous functional covariates and scalar responses.

We formally introduce asynchronous dataset as follows. Suppose that there are n subjects and we observe response and a set of functional and scalar covariates over different time points for each subject. Specifically, for $i = 1, \dots, n$ and $d = 1, \dots, D$, we observe

$$Y_i(T_{ij}), \quad j = 1, \dots, L_i, \quad (1)$$

$$\{Z_i(S_{ik}), X_{di}(S_{ik}, u)\}, \quad k = 1, \dots, M_i,$$

where L_i and M_i are, respectively, the number of time points for response and that for covariates, $Z_i(S_{ik})$ is a p -dimensional

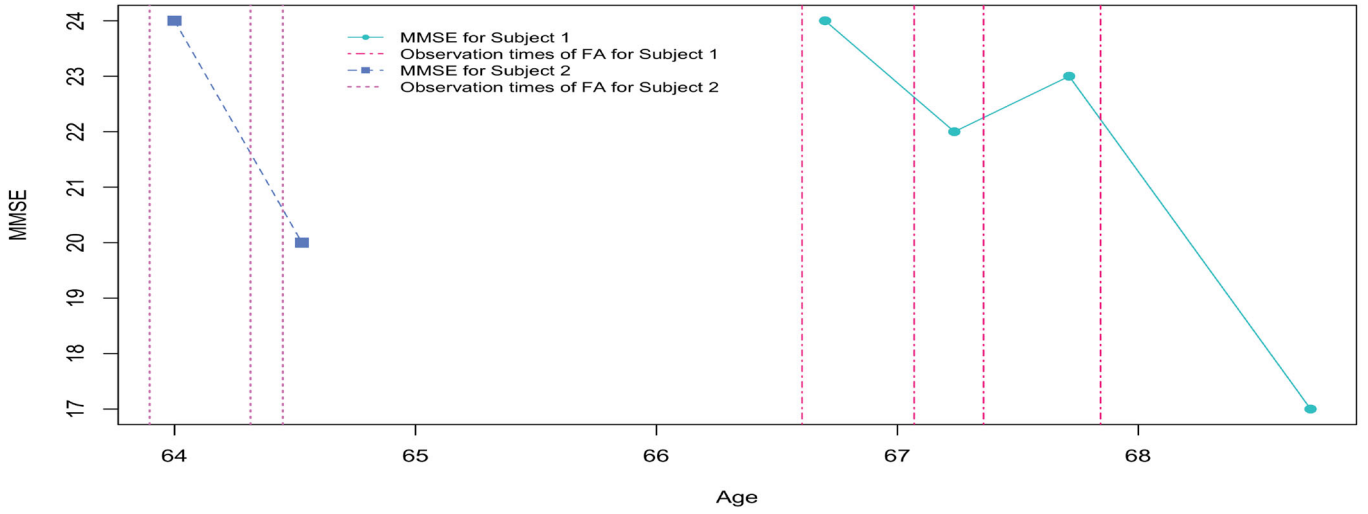


Figure 1. Structure of the ADNI dataset: MMSE scores for two subjects over several visits, the dotted vertical lines correspond to the observation times for log hazard functions of FA.

vector of time varying covariates and $X_{di}(S_{ik}, u)$'s are D functional covariates and can be observed on a set of dense grids for a given S_{ik} . Moreover, T_{ij} 's are the observation times of responses $Y_i(T_{ij})$ and S_{ik} 's are those of the functional and scalar covariates $\{Z_i(S_{ik}), X_{di}(S_{ik}, u)\}$. In practice, T_{ij} 's are different from S_{ik} 's. In our ADNI dataset, d denotes a specific ROI and D is the number of ROIs, $Y_i(T_{ij})$'s are the MMSE scores, $Z_i(S_{ik})$ includes various clinical variables, such as age and education, and $X_{di}(S_{ik}, u)$'s are the log hazard curves of FA measured repeatedly over time and ROI, leading to longitudinal functional covariates. Although much has been done on the analyses of the ADNI data (Diggle 2002; Leow et al. 2009; Weiner et al. 2013; Li et al. 2017; Wang, Zhu, and ADNI 2017; Zhu et al. 2017), almost all existing works consider synchronous data setting and the related statistical methods cannot appropriately handle asynchronous dataset (1). Research on asynchronous data is very scarce with a few exceptions including Xiong and Dubin (2010), Şentürk et al. (2013), Cao, Zeng, and Fine (2015), Cao, Li, and Fine (2016), and Chen and Cao (2017). Moreover, to the best of our knowledge, longitudinal functional predictors have not been considered for asynchronous data setting.

The observed dataset (1) consists of longitudinal functional data, which are observed repeatedly for each subject at multiple visits. Analysis of repeatedly measured functional outcome has received much attention recently due to an increasing number of such data. There are vast majority of approaches based on functional principal component analysis (FPCA), see, for example, multilevel FPCA by Di et al. (2009); longitudinal FPCA by Greven et al. (2011), Chen and Müller (2012), Park and Staicu (2015), and Chen, Delicado, and Müller (2017); FPCA of spatiotemporal point process by Li and Guan (2014); multi-dimensional FPCA by Hasenstab et al. (2017); and hybrid FPCA by Scheffler et al. (2020). Meanwhile, Morris et al. (2003), Morris and Carroll (2006), Yang et al. (2017), and Lee et al. (2019) combined Bayesian methods with longitudinal functional mixed models. They defined mixed effects models at the level of the basis coefficients and had to specify priors about the random effects and their covariance structure.

In contrast, Staicu, Lahiri, and Carroll (2015), Park et al. (2018), and Zhu et al. (2019), among others, studied inference for the fixed effects in longitudinal functional mixed effects models.

The goal of this article is to develop a set of new longitudinal regression models for the association analysis of asynchronous functional covariates and scalar response and their related estimations and inferences. We consider a set of generalized functional partial linear varying-coefficient (GFPV) models as follows:

$$E\{Y(t)|X(t, u), Z(t)\} = g\left(Z(t)^\top \gamma_0 + \sum_{d=1}^D \int X_d(t, u) \beta_{0d}(t, u) du\right), \quad (2)$$

where γ_0 is a $p \times 1$ vector of regression coefficients, $\beta_{0d}(\cdot, \cdot)$'s are functional regression coefficients, $g(\cdot)$ is a known, strictly increasing and continuously twice-differentiable link function, $Z(t)$ is a $p \times 1$ vector of time varying covariates with each element resides in $L^2([0, 1])$, $X_d(t, u) \in L^2([0, 1]^{\otimes 2})$, $d = 1, \dots, D$ are bivariate stochastic processes, and $Y(t)$ is a temporal response process. At a fixed time t , $X_d(t, u)$ is a random process of u for each $d = 1, \dots, D$. Following Cao, Zeng, and Fine (2015) and Lin and Ying (2001), the observation times of $X_{di}(s, u)$, $Z_i(s)$, and $Y_i(t)$ can be viewed from a bivariate counting process defined as

$$N_i(t, s) = \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} I(T_{ij} \leq t, S_{ik} \leq s),$$

where $I(A)$ is the indicator function of an event A . We are interested in using asynchronous dataset (1) to estimate both parametric regression coefficients in γ_0 and functional regression coefficients $\{\beta_{0d}(\cdot, \cdot) : d = 1, \dots, D\}$ in model (2). Such asynchronous setting, however, greatly complicates model estimations and inferences for model (2).

Model (2) falls within a general functional linear regression modeling framework, which can characterize the association between scalar (or functional) response and functional and scalar covariates. See Ramsay and Silverman (2005) and

references therein for an extensive overview of functional linear models. Recently, there is an increasing interest in the development of longitudinal functional regression models for the analyses of synchronous functional/scalar covariates and responses. Goldsmith et al. (2012) proposed a longitudinal penalized functional regression model, and explored the relationship between cerebral white matter tracts and cognitive impairment over time. Gertheiss et al. (2013) extended the study of Goldsmith et al. (2012) and applied their model to a brain diffusion tensor imaging tractography study. Kundu, Harezlak, and Randolph (2016) considered time-varying coefficient functions in longitudinal functional regression models to analyze the neurocognitive impairment of human immunodeficiency virus patients and its association with structural imaging curves. Staicu et al. (2020) developed a longitudinal dynamic functional regression to study time-varying association between Gaussian or non-Gaussian responses and functional covariates. Scheipl, Staicu, and Greven (2015), Scheipl, Gertheiss, and Greven (2016), and Greven and Scheipl (2017) also considered bivariate coefficients in (generalized) functional additive mixed models for functional responses and expanded the bivariate functional coefficient by using the tensor product of spline bases with a smooth penalty. Although all existing methods provide a convenient vehicle for synchronous longitudinal data with functional covariates, they are inapplicable to asynchronous dataset (1).

In this article, we propose the GFPV model (2) to characterize the relationship between asynchronous longitudinal functional covariates and scalar responses. Compared with the existing literature discussed above, we make several major contributions as follows. (i) It is the first time that we conduct association analysis between the asynchronous longitudinal functional covariates and scalar responses. We construct penalized kernel-weighted estimating equations by using penalized B-splines to represent functional coefficients and develop a relatively fast algorithm for solving such estimating equations. (ii) We systematically carry out the theoretical analyses of all scalar and functional estimators including consistency, convergence rate, prediction accuracy, and their limiting distributions. We also establish the bootstrap consistency of our proposed test statistics. (iii) The novel application of GFPV to the ADNI dataset reveals that the deterioration of intelligence is associated with the well-known factors, such as education and APOE4 gene, as well as the decreasing of the top quantiles of FA at the whole brain and across 5 specific ROIs. (iv) The package for GFPV along with its documentation is freely accessible from our lab's github website.

The rest of the article is organized as follows. Section 2.1 gives the detailed procedure for our estimation method. Section 2.2 discusses the computational complexity of our estimation method. In Section 2.3, we present a hypothesis testing procedure for testing the nullity of both parametric and functional coefficients. In Section 3, we show the consistency and convergence rates of parameter and functional estimators under some mild assumptions. Section 4 investigates the finite sample performance of the proposed method in simulations. We apply our methods to a real dataset obtained from the ADNI study in Section 5. All proofs and additional figures can be found in the supplementary materials.

2. GFPV Models

2.1. Estimation

We propose penalized kernel-weighted estimating equations to obtain the estimators of γ_0 and $\beta_{0d}(t, u)$ as follows. Kernel-weighted estimating equations have been widely used in handling longitudinal synchronous data (Carroll et al. 1997; Lin and Carroll 2001). Cao, Zeng, and Fine (2015) and Chen and Cao (2017) have shown that the kernel-weighted estimating equation method can be efficient for handling asynchronous data by borrowing information from all possible pairings of response and covariate observations.

Without loss of generality and for notation simplicity, we assume $D = 1$ and drop the subscript d . Penalized spline approach offers a number of computational advantages and can produce consistent estimators with a potentially large number of knots (Yu and Ruppert 2002). Let m be the order of the B-splines. Without loss of generality, it is assumed that the numbers of knots are the same in the t and u directions. It can be easily extended to different number of knots for the two directions. To be more specific, let $0 = t_0 < t_1 < \dots < t_M < t_{M+1} = 1$ and $0 = u_0 < u_1 < \dots < u_M < u_{M+1} = 1$ be the sequences of knots on intervals $[0, 1]$ and $[0, 1]$, respectively. This gives $K = M + m$ basis functions for each direction.

We represent $\beta(t, u)$ by using a tensor product penalized spline basis as follows:

$$\beta(t, u) \approx \sum_{k=1}^K \sum_{l=1}^K b_{kl} \eta_k(t) \theta_l(u) = \boldsymbol{\eta}(t)^\top \mathbf{B} \boldsymbol{\theta}(u), \quad (3)$$

where $\boldsymbol{\eta}(t)$ and $\boldsymbol{\theta}(u)$ are two $K \times 1$ basis vectors corresponding to the K basis functions evaluated in the t and u directions, and \mathbf{B} is a spline coefficient matrix. Denote $\tilde{X}(t) = (\tilde{X}_{11}(t), \dots, \tilde{X}_{K1}(t), \dots, \tilde{X}_{1K}(t), \dots, \tilde{X}_{KK}(t))^\top$ and $\mathbf{b} = \text{Vec}(\mathbf{B})$ as the vectorization of the matrix \mathbf{B} , where $\tilde{X}_{kl}(t) = \eta_k(t) \int_0^1 X(t, u) \theta_l(u) du$. Model (2) can be approximated by

$$E\{Y(t)|X(t, u), Z(t)\} = g\left(Z(t)^\top \gamma_0 + \tilde{X}(t)^\top \mathbf{b}\right). \quad (4)$$

We introduce a roughness penalty on $\eta_k(t)$ and $\theta_l(u)$. Similar to Wood (2006), we choose the roughness penalty as follows:

$$\mathbf{P} \equiv \mathbf{P}(\lambda_t, \lambda_u) = \lambda_t (\mathbf{P}_t \otimes \mathbf{I}_K) + \lambda_u (\mathbf{P}_u \otimes \mathbf{I}_K),$$

where λ_t and λ_u are smoothing parameters and \mathbf{P}_t and \mathbf{P}_u are the fixed and known marginal penalty matrices for the t and u directions, respectively. One example of the marginal penalty matrices is

$$\begin{aligned} \mathbf{P}_t(\beta) &= \int \int [L_t \beta(t, u)]^2 dt du \quad \text{and} \\ \mathbf{P}_u(\beta) &= \int \int [L_u \beta(t, u)]^2 dt du \end{aligned}$$

with L_t and L_u being linear differential operators. One of the commonly used linear differential operators is the curvature operator with $L_t = \partial^2 / \partial t^2$ and $L_u = \partial^2 / \partial u^2$.

We use the idea of penalized estimating equations (Fu 2003) to construct penalized kernel-weighted estimating equations

below. Specifically, the penalty added to the estimation equations takes the form $\partial P / \partial \mathbf{b}$ given by

$$\begin{aligned} \frac{\partial P}{\partial \mathbf{b}} &= \lambda_t \frac{\partial (\mathbf{P}_t \otimes \mathbf{I}_K)}{\partial \mathbf{b}} + \lambda_u \frac{\partial (\mathbf{P}_u \otimes \mathbf{I}_K)}{\partial \mathbf{b}} \\ &= (\lambda_t J_{\theta\theta} \otimes \mathbf{R} + \lambda_u S \otimes J_{\eta\eta}) \mathbf{b}, \end{aligned}$$

where $R = \int [L_t \boldsymbol{\eta}(t)][L_s \boldsymbol{\eta}^\top(t)] dt$, $S = \int [L_u \boldsymbol{\theta}(u)][L_s \boldsymbol{\theta}^\top(u)] du$, $J_{\theta\theta} = \int \boldsymbol{\theta}(u) \boldsymbol{\theta}^\top(u) du$, and $J_{\eta\eta} = \int \boldsymbol{\eta}(u) \boldsymbol{\eta}^\top(u) du$. One may refer to Ramsay and Silverman (2005) for more details about various penalty functions. Denote $P_{\lambda_t, \lambda_u} = \lambda_t J_{\theta\theta} \otimes \mathbf{R} + \lambda_u S \otimes J_{\eta\eta}$ and $\zeta = (\gamma^\top, M^{-1} \mathbf{b}^\top)^\top$, our penalized estimating equations for \mathbf{b} and γ , denoted as $\Psi_n(\zeta)$, are given by

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} K_h(T_{ij} - S_{ik}) \tilde{Z}_i(S_{ik}) \\ [Y_i(T_{ij}) - g\{Z_i(S_{ik})^\top \gamma + \tilde{X}_i(S_{ik})^\top \mathbf{b}\}] - \tilde{P}, \end{aligned} \quad (5)$$

where $\tilde{Z}_i(t) = (Z_i(t)^\top, \tilde{X}_i(t)^\top)^\top$ and $\tilde{P}^\top = (\mathbf{0}^\top, (P_{\lambda_t, \lambda_u} \mathbf{b})^\top)$. Moreover, $K_h(t) = K(t/h)/h$ is the kernel weighting function, where $K(t)$ is a symmetric kernel function and h is the bandwidth. Following Fan, Yao, and Cai (2003), we choose $K(t)$ to be the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$. Subsequently, applying the Newton-Raphson algorithm to (5) leads to the estimator of ζ , denoted as $\hat{\zeta} = (\hat{\gamma}^\top, M^{-1} \hat{\mathbf{b}}^\top)^\top$.

Although response and covariates are mismatched, the kernel weighting function enables the use of all covariate observations for each observed response by down-weighting those observations far away in time. Observation times T_{ij} and S_{ik} for $i = 1, \dots, n$ are required to be close for some but not all subjects. We use all covariates for each observed response through the use of the kernel weighting function $K_h(t) = K(t/h)/h$. As long as there exist time points T_{ij} and S_{ik} with distance smaller than h , we can estimate the unknown parameters. Theoretically, we explicitly impose the condition $h > \min_i \min_{j,k} |T_{ij} - S_{ik}|$. Thus, all possible pairings of response and covariate observations with the time differences smaller than h contribute to our estimating equations, making the estimation possible.

2.2. Computational Complexity

The computational complexity of our estimation procedure is extremely important for asynchronous longitudinal functional data, which usually contain a large number of points in the time domain and the functional domain. Denote $N_t = \max\{L_i, M_i, i = 1, \dots, n\}$ and N_u as the maximal number of points in the t direction and u direction for all subjects.

We discuss the computational complexity of our estimating procedure as follows. First, we calculate $\tilde{X}_{kl}(s)$ for each k, l and each s , which is computationally straightforward. The computational complexity is $O(N_u)$ for each time point s , so overall it is $O(N_u N_t K^2)$ for calculating all $\tilde{X}_{kl}(s)$. Second, the computational complexity of calculating the matrix P_{λ_t, λ_u} is $O(K^4)$. Third, the computational complexity of calculating the estimators in (5) is $O(nh N_t^2 K^4 + K^6)$. Hence, taking the computational complexity of the above three steps together, the computational complexity of our estimating procedure in Section 2.1 is $O(nh N_t^2 K^4 + K^6 + N_u N_t K^2)$. If we choose the tuning parameters $\{\lambda_t, \lambda_u, h\}$

by a grid search method, then the computational complexity increases by a factor of the number of the set of tuning parameters.

2.3. Hypothesis Testing

Although estimation and prediction are interesting, it is also important to test the nullity of β and γ . For instance, in the ADNI data analysis, we are interested in examining whether the FA values across the ROIs and other covariates have significant effects on the MMSE decline. Therefore, we propose to test two sets of null and alternative hypotheses as follows:

$$H_{0,\beta} : \beta_0(t, u) = 0, \text{ for any } (t, u) \quad \text{vs.}$$

$$H_{1,\beta} : \beta_0(t, u) \neq 0, \text{ for some } (t, u), \quad (6)$$

$$H_{0,\gamma} : \gamma_0 = 0, \quad \text{vs.} \quad H_{1,\gamma} : \gamma_0 \neq 0. \quad (7)$$

To test (6) and (7), we propose two statistics as follows:

$$T_\beta = \frac{nh}{M^2} \|\hat{\beta}(\cdot, \cdot)\|^2 = \frac{nh}{M^2} \int \int \hat{\beta}^2(t, u) dt du \quad \text{and}$$

$$T_\gamma = (nh) \|\hat{\gamma}\|^2 = (nh) \hat{\gamma}^\top \gamma. \quad (8)$$

The statistic T_β is to measure the L^2 distance between $\hat{\beta}(\cdot, \cdot)$ and 0, whereas T_γ is to measure the Euclidean distance between $\hat{\gamma}$ and 0. Since $\beta(t, u)$ is approximated by a B-spline tensor product in (3), T_β can be rewritten as

$$T_\beta = (nh/M^2) \sum_{k=1}^K \sum_{l=1}^K \hat{b}_{kl}^2. \quad (9)$$

Under the null hypotheses, the two statistics are expected to be close to zero. One can reject $H_{0,\beta}$ for large T_β and reject $H_{0,\gamma}$ for large T_γ .

Although we can derive the null limiting distributions of T_β and T_γ based on the limiting distributions with respect to $\hat{\gamma}$ and $\hat{\mathbf{b}}$ in Section 3, it is challenging to estimate their critical points involved. Thus, a resampling procedure is proposed to approximate the null limiting distributions of the two statistics. Chatterjee and Bose (2005) proposed a generalized bootstrap for estimators by solving estimating equations. Cheng and Huang (2010) showed that the bootstrap is asymptotically consistent in estimating the distribution of the M -estimate of Euclidean parameter. We modified the bootstrap methods by Chatterjee and Bose (2005) and Cheng and Huang (2010) to account for functional variables and asynchronous data setting. Specifically, we bootstrap our estimating equations as follows.

- Step 1. Solve estimating Equation (5) to obtain the estimators $\hat{\mathbf{b}}$ and $\hat{\gamma}$.
- Step 2. For each a , generate n iid random variables $W_n^{(a)} = (W_{n1}^{(a)}, \dots, W_{nn}^{(a)})$ with $E(W_{ni}^{(a)}) = 1$, $E(W_{ni}^{(a)} - 1)^2 \rightarrow 1$, and $E(W_{ni}^{(a)})^8 < \infty$ for $i = 1, \dots, n$, where \rightarrow denotes the limit as $n \rightarrow \infty$.
- Step 3. Solve the following $W_n^{(a)}$ -weighted estimating equations and denote their solutions as $\hat{\mathbf{b}}^{(a)}$ and $\hat{\gamma}^{(a)}$,

$$\begin{aligned} \sum_{i=1}^n W_{ni}^{(a)} \left(\sum_{j=1}^{L_i} \sum_{k=1}^{M_i} K_h(T_{ij} - S_{ik}) \tilde{Z}_i(S_{ik}) \right. \\ \left. [Y_i(T_{ij}) - g\{Z_i(S_{ik})^\top \gamma + \tilde{X}_i(S_{ik})^\top \mathbf{b}\}] - \tilde{P} \right) = 0. \end{aligned}$$

The tuning parameters are the same as those selected in Step 1.

- Step 4. Repeat Steps 2 and 3 for A times. Calculate

$$T_{\beta}^{(a)} = (nh/M^2) \sum_{k=1}^K \sum_{l=1}^K (\hat{b}_{kl}^{(a)} - \hat{b}_{kl})^2 \quad \text{and}$$

$$T_{\gamma}^{(a)} = (nh)(\hat{\gamma}^{(a)} - \hat{\gamma})^{\top} (\hat{\gamma}^{(a)} - \hat{\gamma}).$$

- Compute $\hat{p}_{\beta} = A^{-1} \sum_{a=1}^A I(T_{\beta}^{(a)} > T_{\beta})$ and $\hat{p}_{\gamma} = A^{-1} \sum_{a=1}^A I(T_{\gamma}^{(a)} > T_{\gamma})$. The null hypothesis $H_{0,\beta}$ is rejected if \hat{p}_{β} is smaller than a prefixed significance level α , and $H_{0,\gamma}$ is rejected if \hat{p}_{γ} is smaller than α .

The estimation and testing procedures depend on the selection of the smoothing parameters λ_s and λ_u and the bandwidth h . Since response and covariate observations are mismatched, the widely adopted tuning parameter selection methods such as generalized cross-validation are inapplicable here. Following the tuning parameter selection procedure in Cao, Zeng, and Fine (2015) and Chen and Cao (2017), we propose to use a data-driven tuning parameter selection method by minimizing the roughly estimated mean-squared error for $\hat{\zeta}$. We calculate the squared bias and variance of $\hat{\zeta}$ separately. Denote $\lambda = \max(\lambda_s, \lambda_u)$. For variance, we randomly split the data into two sub samples and derive the estimates $\hat{\zeta}_1$ and $\hat{\zeta}_2$, and then the sum of the variance of each estimate in $\hat{\zeta}$ is given by $\hat{V}(\lambda, h) = \sum_{i=1}^{p+K^2} (\hat{\zeta}_{1i} - \hat{\zeta}_{2i})^2 / 4$. According to Theorem 1 shown later, the squared bias of $\hat{\zeta}$ is of the order $O_p(\lambda + M^2 h^4 + M^{-2r})$ as $\lambda M^4 = O_p(1)$. In practice, although the order of smoothness r is often unknown, we may use the same M for different tuning parameters, indicating that M^{-2r} is the same over different tuning parameters. Then λ and h are selected to minimize

$$\text{Criterion} = M^2 h^4 + \lambda + \hat{V}(\lambda, h). \quad (10)$$

3. Theoretical Results

In this section, we study the asymptotic properties of all estimators. Prior to presenting the theoretical results, we need to introduce some notations. For any matrix A , let $\|A\|$ denote the modulus of the largest singular value of A . Let $\|a\|$ be the Euclidean norm for any vector a , and $\|\beta\|$ be the L^2 norm for any functional process β .

Define $C^r([0, 1]^{\otimes 2})$ as the space of all bivariate functions $h(t, u)$ on $[0, 1]^{\otimes 2}$ such that $D^{(r_1, r_2)} h = \partial^{r_1+r_2} h / (\partial^{r_1} t \partial^{r_2} u)$ is continuous and Lipschitz of order q , that is, for any $T_1, T_2 \in [0, 1]^{\otimes 2}$ and $r \geq r_1 + r_2 + q$,

$$|D^{(r_1, r_2)}(T_1) - D^{(r_1, r_2)}(T_2)| \leq W_0 \|T_1 - T_2\|^q,$$

where W_0 is a finite constant.

We assume that $\beta \in C^r([0, 1]^{\otimes 2})$ holds for some $r \geq 1$. According to Theorem 12.7 of Schumaker (1981), there exists a constant $W > 0$ such that

$$\sup_{s, u} |\beta_0(s, u) - \eta(s)^{\top} \mathbf{B}_0 \theta(u)| \leq WM^{-r}, \quad (11)$$

where \mathbf{B}_0 is a $K \times K$ -dimensional matrix depending on β_0 . Let the vectorization of \mathbf{B}_0 be \mathbf{b}_0 , and the true value of γ be γ_0 .

Assumptions and proofs of all theoretical results can be found in the supplementary materials. The following theorem studies the estimation error of the estimators $\hat{\zeta} = (\hat{\gamma}^{\top}, M^{-1} \hat{\mathbf{b}}^{\top})^{\top}$ and the estimation error of the functional estimator.

Theorem 1. Suppose that Assumptions 1–9 in the supplementary materials hold. If $M = o(n)$, $\lambda M^4 = O(1)$, $\lambda \rightarrow 0$, $Mh^2 \rightarrow 0$, and $M^2(nh)^{-1/2} \rightarrow 0$ hold as $n \rightarrow \infty$, then

$$\begin{aligned} \int \int \{\hat{\beta}(s, u) - \beta_0(s, u)\}^2 ds du &= O_p(M^2/(nh)) \\ &+ O_p(\lambda + \lambda M^{4-2r} + M^{-2r} + M^2 h^4), \end{aligned} \quad (12)$$

and $\|\hat{\zeta} - \zeta_0\|^2 = O_p(M^2/(nh)) + O_p(\lambda + \lambda M^{4-2r} + M^{-2r} + M^2 h^4)$.

The proof of the convergence rate of $\|\hat{\zeta} - \zeta_0\|^2$, which depends on the theory of Z-estimation tailored to asynchronous data, is much more complex than that of Cao, Zeng, and Fine (2015) and Chen and Cao (2017). The estimation error of $\hat{\beta}(s, u)$ consists of three parts including the bias of the estimators $\lambda + \lambda M^{4-2r} + M^{-2r} + M^2 h^4$, the variance of the B-spline coefficient estimators $M^2(nh)^{-1}$, and the approximation error M^{-2r} . The bias, which is caused by the representation and the penalty, of $\hat{\beta}$ is similar to that for functional regression models in Cardot, Ferraty, and Sarda (2003) and Cardot and Sarda (2005). If we set $h = O(n^{-1/5})$, $M = O(n^{2/(5(r+1))})$, and $\lambda = O(n^{-4r/(5(r+1))})$, which leads to a minimal convergence rate in (12), then we can achieve a convergence rate of $O_p(n^{-4r/(5(r+1))})$ or $O_p((nh)^{-r/(r+1)})$, which is slower than the optimal nonparametric rate of convergence $O_p(n^{-r/(r+1)})$ in Stone (1982). The loss of efficiency is due to the asynchronous data setting.

Denote $\eta_0(t, \beta) = Z(t)^{\top} \gamma_0 + \int X(t, u) \beta(t, u) du$ and $\hat{\eta}_0(t, \beta) = Z(t)^{\top} \hat{\gamma} + \int X(t, u) \hat{\beta}(t, u) du$. For a given t , let $g'(\eta_0(t, \beta))$ be the first-order derivative of g with respect to $\eta_0(t, \beta)$. The following theorem details the prediction error of the estimator of $g\{\eta_0(t, \beta)\}$.

Theorem 2. Under the conditions of Theorem 1, for a given time point t and β in a neighborhood of β_0 , if $E[g'^2(\eta_0(t, \beta))Z(t)^{\top}Z(t)] < \infty$ and $E[g'^2(\eta_0(t, \beta)) \int X^2(t, u) du] < \infty$ hold, then we have

$$\begin{aligned} &|g\{\hat{\eta}_0(t, \beta)\} - g\{\eta_0(t, \beta)\}|^2 \\ &= O_p(M^2/(nh)) + O_p(\lambda + \lambda M^{4-2r} + M^{-2r} + M^2 h^4). \end{aligned}$$

Theorem 2 shows that under some mild conditions, the prediction error of $g\{\hat{\eta}_0(t, \beta)\}$ is of the same order as the estimation error. Inspecting Theorem 1 might reveal that the rate of convergence for $\hat{\gamma}$ would be the same as that for $\hat{\beta}$. However, the convergence rate for $\hat{\gamma}$ is supposed to be a faster one. For example, in Cheng, Zhang, and Chen (2009), the parametric component enjoys a parametric convergence rate whereas the estimator of the nonparametric component has a nonparametric one. So a different convergence rate of $\hat{\gamma}$ is investigated in Theorem 3.

We need to assume that for $s, t \in [0, 1]$, $\text{var}\{Y(t)|X(t, u), Z(t)\} = \sigma(t, X(t, u), Z(t))^2$ and there

exists a twice-continuous differentiable function $\lambda(t, s)$ such that $E\{dN_i(t, s)\} = \lambda(t, s)dt ds$. Let $A_Z = \int E\{\tilde{Z}_i(s)g'\{\eta_0(s, \beta_0)\}\tilde{Z}_i(s)^\top\}\lambda(s, s)ds$ and

$$\Sigma_Z = \int K(z)^2 dz \int E\{\tilde{Z}_i(s)\tilde{Z}_i(s)^\top\}\sigma(s, X(s, u), Z(s))^2\lambda(s, s)ds.$$

Also, denote Π_n and Ω_n as

$$\Pi_n = Mn^{-1} \sum_{i=1}^n \int \int K_h(t-s)\tilde{X}_i(s)g'\{\eta_0(s, \beta_0)\}Z_i(s)^\top dN_i(t, s) \text{ and} \quad (13)$$

$$\Omega_n = M^2 n^{-1} \sum_{i=1}^n \int \int K_h(t-s)\tilde{X}_i(s)g'\{\eta_0(s, \beta_0)\}\tilde{X}_i(s)^\top dN_i(t, s), \quad (14)$$

respectively. Moreover, $\tilde{Z}_i(s) = Z_i(s) - M\Pi_n^\top \Omega_n^{-1} \tilde{X}_i(s)$, where $M\Pi_n^\top \Omega_n^{-1} \tilde{X}_i(s)$ can be viewed as a projection of the scalar variables to the functional space.

Theorem 3. Suppose that the conditions of [Theorem 1](#) are satisfied. Also, $nh^5 = o(1)$, $nh\lambda M^{-2} \rightarrow 0$, $nhM^{-(2r+2)} \rightarrow 0$, and A_Z and Σ_Z are positive definite. Then as $n \rightarrow \infty$, we have

$$\sqrt{nh}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, \Sigma_{\gamma_0}),$$

where $\Sigma_{\gamma_0} = \lim_{n \rightarrow \infty} A_Z^{-1} \Sigma_Z (A_Z^{-1})^\top$ and \xrightarrow{d} denotes the convergence in distribution.

[Theorem 3](#) presents a different rate of convergence of $\hat{\gamma}$ compared with that in [Theorem 1](#). We also extend the asymptotic normality result for the parametric estimator in Cao, Zeng, and Fine (2015) and Chen and Cao (2017) to the GFPV models. The rate of convergence $(nh)^{1/2}$ is the same, but the asymptotic variance of the parametric estimator is different and more complicated than that in Cao, Zeng, and Fine (2015) and Chen and Cao (2017). Specifically, we have to adjust for the dependence between the scalar variables and the functional variables, which is an additional complication arising in functional partial linear models and is more challenging under the longitudinally asynchronous data setting.

We will present the asymptotic distribution with respect to $\hat{\mathbf{b}}$ below.

Theorem 4. Assume that the conditions of [Theorem 3](#) hold, and $nh^5 M = o(1)$. Moreover, the eigenvalues of

$$\Sigma_{\beta_0} = \text{var}\left(\int \int Mh^{1/2} K_h(t-s)\tilde{X}_i(s)[Y_i(t) - g(\eta_0(s, \beta_0))]dN_i(t, s)\right)$$

are bounded away from 0 and infinity. As $n \rightarrow \infty$, we have

$$\frac{(nh/M^2)(\hat{\mathbf{b}} - \mathbf{b}_0)^\top \Omega_n^2 (\hat{\mathbf{b}} - \mathbf{b}_0) - \text{tr}(\Sigma_{\beta_0})}{\sqrt{2\text{tr}(\Sigma_{\beta_0}^2)}} \xrightarrow{d} N(0, 1).$$

To the best of our knowledge, it is the first time that the asymptotic normality has been proved for the quadratic form of $\hat{\mathbf{b}} - \mathbf{b}_0$ under generalized functional partial linear models. A similar asymptotic distribution result about a Wald-type statistic

was obtained in Kong, Staicu, and Maity (2016). However, it is targeted for functional partial linear models with normal distribution error structure, which is too restrictive and hard to be satisfied in practice. Meanwhile, the mismatched observation times and a more general structure bring additional difficulties in deriving the distribution. We also present the point-wise limiting distribution of $\hat{\beta}$ and extension to $D > 1$ functional variables of [Theorem 4](#) in the supplementary materials. Although we can derive the asymptotic distributions of T_β and T_γ in [Theorems 3](#) and [4](#), we resort to the bootstrapping procedure to approximate them. In [Theorem 5](#), we will show that the bootstrap distributions of the proposed two statistics, conditional on the observations, asymptotically imitate their corresponding null limiting distributions.

Theorem 5. Suppose that the conditions of [Theorems 3](#) and [4](#) are satisfied. Also, for $i = 1, \dots, n$ and $a = 1, \dots, A$, the bootstrap weights satisfy

$$E(W_{ni}^{(a)}) = 1, E(W_{ni}^{(a)} - 1)^2 \rightarrow 1 \text{ and } E(W_{ni}^{(a)})^8 < \infty.$$

The conditional distribution of $T_\beta^{(a)} = (nh/M^2)\|\hat{\mathbf{b}}^{(a)} - \hat{\mathbf{b}}\|^2$ given the observed data converges almost surely to the null limiting distribution of $T_\beta = (nh/M^2)\|\hat{\mathbf{b}} - \mathbf{b}_0\|^2$. Similar result also holds for $T_\gamma^{(a)} = (nh)\|\hat{\gamma}^{(a)} - \hat{\gamma}\|^2$ and $T_\gamma = (nh)\|\hat{\gamma} - \gamma_0\|^2$.

[Theorem 5](#) validates the bootstrapped processes of $T_\beta^{(a)}$ and $T_\gamma^{(a)}$.

4. Simulation Studies

In this section, we investigate the finite sample performance of the proposed method in two most commonly encountered cases including the linear regression case and the logistic regression case. We present the empirical estimation and prediction errors in [Example 1](#), and explore the finite sample performance of the proposed testing procedure in [Examples 2](#) and [3](#).

Example 1. To obtain different observation times for the response and the covariates, the number of observation times of $Y(t)$ and $Z(t)$ was generated from a Poisson distribution with the intensity rate λ_{Poisson} . Once we had the two numbers of observation times, the observation times for the response as well as the covariates were generated from the uniform distribution $U(0, 1)$, independently.

For the i th subject, the process $Z_i(t)$ was generated from a Gaussian process with values at fixed time points being multivariate normal distribution with mean 0, variance 1, and correlation structure $\exp(-|t_{ij} - t_{ik}|)$, where t_{ij} is the j th measurement time and t_{ik} is the k th measurement time within the same subject. The observations times for $X(t, u)$ are the same as those for $Z(t)$ in the t direction, while 200 observation times from the uniform distribution were included throughout the simulation in the u direction. We mixed the observation times of the response with those of the covariates when generating the

covariates, but we computed $Y(t)$ only at the specific measurement times of $Y(t)$. The functional process $X(t, u)$ is of the form

$$X(t, u) = \sum_{k=1}^{50} (-1)^{k+1} k^{-1} M_k(t) \vartheta(u, k),$$

where each $M_k(t)$ is the same Gaussian process as $Z(t)$, and $\vartheta(u, k) = 1$ if $k = 1$ and $\sqrt{2} \cos((k-1)\pi u)$ otherwise. For the coefficient function $\beta_0(t, u)$, we set

$$\beta_0(t, u) = 4 * \sum_{k=1}^{50} (-1)^{k+1} k^{-2} \vartheta(t, k) \vartheta(u, k).$$

In the linear regression case, responses were generated from

$$Y(t) = Z(t)\gamma_0 + \int_0^1 X(t, u)\beta_0(t, u)du + \epsilon(t),$$

where $\epsilon(t)$ is a Gaussian process, with mean 0, variance 1 and $\text{cov}\{\epsilon(s), \epsilon(t)\} = 2^{-|t-s|}$. In the logistic regression case, the responses were generated from

$$P(Y(t) = 1 | Z(t), X(t, u)) = \frac{\exp\left(Z(t)\gamma_0 + \int_0^1 X(t, u)\beta_0(t, u)du\right)}{1 + \exp\left(Z(t)\gamma_0 + \int_0^1 X(t, u)\beta_0(t, u)du\right)}.$$

We set $\gamma_0 = 1$ and chose the Poisson parameter λ_{Poisson} to be 10, 15, and 20, respectively. All simulation results were based on 200 replications and the sample size n was in $\{100, 200, 400\}$ by using R (version 3.3.2) on a linux server (equipped with Intel(R) Xeon(R) CPU E5-2643 v4 @ 3.40GHz, 10GB RAM).

We used the penalized cubic spline to represent the coefficient function $\beta(t, u)$ and chose equally based knots. Then, we chose the number of the basis functions K according to the suggestions in Ruppert (2002). Specifically, the number of knots is not a crucial parameter since smoothing is controlled by the penalty parameter. So, throughout the simulation studies, we chose 15 knots for both the t and u directions, respectively, leading to $K = 18$ and 256 basis coefficients. To simplify the procedure, we set $\lambda_s = \lambda_u = \lambda$. The tuning parameters λ and h were selected to minimize the roughly estimated mean-squared error in (10). We also included the results when choosing 5 knots ($K = 8$) without penalties (denoted as “No penalty”) to investigate the effects of the penalties.

Following the suggestions from two reviewers, we compared our method with three additional methods in linear cases in terms of both estimation and prediction accuracy. The first method, denoted as LDFR_y, consists of pre-smoothing the sparsely observed $Y(t)$ by *fpca.sc()* function in the *refund* R package, which leads to an estimated functional response, and fitting a “synchronous” model using the functional dynamic functional regression proposed by Staicu et al. (2020). The second one, denoted as LDFR_{xz}, consists of pre-smoothing $Z(t)$ by *fpca.sc()* function, pre-smoothing $X(t, u)$ by longitudinal FPCA (Park and Staicu 2015), and fitting a “synchronous” model. The third method, denoted as LFPCA, consists of fitting the longitudinal functional principal component (FPCA), that represents the bivariate coefficient function by features extracted

from the functional predictors, and fitting a standard asynchronous longitudinal regression (Cao, Zeng, and Fine 2015). Moreover, the numbers of FPCA scores in the t and u directions were truncated by the 90% and 95% overall variation, respectively.

We computed the empirical MSEs of $\hat{\gamma}$ and $\hat{\beta}$ as well as the empirical relative MSE of $\hat{\beta}$, denoted as RMSE, as follows:

$$\text{MSE}(\hat{\gamma}) = E\{||\hat{\gamma} - \gamma_0||^2\} \quad \text{and}$$

$$\text{MSE}(\hat{\beta}) = E\left\{\int \int (\hat{\beta}(t, u) - \beta_0(t, u))^2 dt du\right\},$$

$$\text{RMSE}(\hat{\beta}) = E\left\{\int \int (\hat{\beta}(t, u) - \beta_0(t, u))^2 dt du / \int \int (\beta_0(t, u))^2 dt du\right\}.$$

The prediction mean squared error (PMSE) is also reported based on 200 new test samples. In contrast, in the logistic regression case, we predict the probability of $Y(t) = 1$ and present the classification error as the misclassification rate of the testing data. Furthermore, we calculated the computation time (in minutes).

Table 1 presents the estimation accuracy and prediction results and their computation times under the linear regression setting. Our method outperforms all the other methods in terms of MSEs of β and γ and PMSE, even though the computation times of all methods except the “No penalty” method are comparable with each other. As expected, MSEs and PMSEs decrease as the sample size and/or the number of observation times increase, confirming our theoretical results. For our method, it costs more time to match all covariates with each response and search grid points of the tuning parameters. In contrast, LDFR_y, LDFR_{xz}, and LFPCA are not computationally stable since they may require interpolating a large number of time points. Specifically, when $n = 400$, the computation programs of LDFR_y, LDFR_{xz}, and LFPCA may crash due to out of memory. Figure S2 in the supplementary materials presents images of the true β_0 and the estimated β_0 from a randomly selected dataset with $n = 100$ and $\lambda_{\text{Poisson}} = 20$. Inspecting Figure S2 reveals that our estimate is closer to the true coefficient function than all the other three methods. Additional simulation results about a further discussion on LDFR_y and LDFR_{xz} are included in Section 2 of the supplementary materials.

The estimation and prediction results in the logistic cases are summarized in Table 2, showing similar patterns to those in the linear case. Although our method without penalty gives smaller computation time than that using the roughness penalty, it gives inaccurate estimation accuracy and prediction results. The roughness penalty not only controls the smoothness of the functional estimator, but also enables stable computation.

Example 2. In this example, we evaluated the Type I and II error rates of the proposed testing statistics. Data settings were the same as those in Example 1 except that we set $\gamma \in \{0, 0.1, 0.3, 0.5\}$ and $\beta_0(t, u) = B * \sum_{k=1}^{50} (-1)^{k+1} k^{-2} \vartheta(t, k) \vartheta(u, k)$, where $B \in \{0, 0.1, 0.3, 0.5\}$. For each case, we used 1000 simulated datasets with 1000 bootstrap samples. The bootstrap weights were simulated from the

Table 1. Results from a simulation study of the linear case in Example 1 for five methods, 3 λ_{Poisson} values {10, 15, 20}, and 3 sample sizes $n \in \{100, 200, 400\}$.

n	λ_{Poisson}		MSE_{β}	RMSE_{β}	MSE_{γ}	PMSE	Time (min)
100	10	Proposed	0.632(0.210)	0.037(0.012)	0.022(0.029)	1.270(0.187)	4.730(0.386)
		No penalty	13.100(5.830)	0.753(0.333)	0.017(0.022)	1.520(0.200)	0.481(0.087)
		LDFR _y	5.890(6.260)	0.340(0.362)	0.157(0.155)	3.850(1.160)	0.990(0.299)
		LDFR _{xz}	1.790(1.310)	0.103(0.075)	0.065(0.151)	1.820(1.240)	4.550(4.990)
		LFPCA	5.620(2.390)	0.324(0.138)	0.031(0.042)	2.120(0.438)	1.590(0.655)
	15	Proposed	0.480(0.130)	0.027(0.007)	0.014(0.021)	1.160(0.124)	7.590(0.773)
		No penalty	7.910(2.590)	0.456(0.148)	0.012(0.017)	1.320(0.143)	0.682(0.054)
		LDFR _y	3.830(1.540)	0.221(0.089)	0.138(0.103)	3.360(0.964)	1.260(0.134)
		LDFR _{xz}	1.070(0.532)	0.062(0.031)	0.031(0.055)	1.350(0.205)	12.500(5.780)
		LFPCA	4.030(0.709)	0.232(0.041)	0.003(0.003)	1.820(0.281)	17.500(18.100)
	20	Proposed	0.401(0.101)	0.023(0.006)	0.012(0.018)	1.110(0.117)	11.60(1.330)
		No penalty	6.090(2.250)	0.351(0.129)	0.013(0.021)	1.240(0.115)	0.872(0.085)
		LDFR _y	3.260(0.993)	0.188(0.057)	0.106(0.074)	3.080(0.719)	1.950(0.199)
		LDFR _{xz}	0.862(0.390)	0.050(0.023)	0.012(0.017)	1.260(0.122)	17.000(6.220)
		LFPCA	3.200(1.230)	0.184(0.071)	0.035(0.026)	1.900(0.199)	27.400(26.200)
200	10	Proposed	0.517(0.136)	0.030(0.008)	0.014(0.017)	1.200(0.149)	11.400(1.280)
		No penalty	5.740(2.230)	0.331(0.127)	0.009(0.011)	1.270(0.133)	0.835(0.098)
		LDFR _y	3.260(0.998)	0.188(0.058)	0.139(0.082)	3.380(0.974)	2.060(0.582)
		LDFR _{xz}	0.896(0.777)	0.052(0.045)	0.061(0.141)	1.440(0.587)	8.920(9.320)
		LFPCA	3.420(1.580)	0.197(0.0915)	0.013(0.016)	1.740(0.183)	7.760(3.070)
	15	Proposed	0.388(0.075)	0.022(0.004)	0.009(0.012)	1.100(0.101)	22.800(3.100)
		No penalty	3.720(1.330)	0.215(0.076)	0.007(0.011)	1.170(0.106)	1.200(0.181)
		LDFR _y	2.590(0.735)	0.150(0.042)	0.100(0.058)	2.850(0.643)	15.700(16.700)
		LDFR _{xz}	0.599(0.384)	0.035(0.022)	0.014(0.034)	1.210(0.298)	25.400(7.710)
		LFPCA	2.930(0.877)	0.169(0.0506)	0.010(0.013)	1.640(0.176)	23.300(1.880)
	20	Proposed	0.319(0.060)	0.018(0.003)	0.006(0.009)	1.070(0.094)	33.600(6.480)
		No penalty	2.960(0.991)	0.171(0.057)	0.006(0.007)	1.120(0.102)	1.630(0.294)
		LDFR _y	2.230(0.552)	0.129(0.032)	0.091(0.053)	2.650(0.515)	23.500(22.300)
		LDFR _{xz}	0.446(0.195)	0.026(0.011)	0.006(0.010)	1.130(0.105)	34.900(9.440)
		LFPCA	2.670(1.010)	0.154(0.0581)	0.008(0.009)	1.580(0.184)	45.700(3.700)
400	10	Proposed	0.435(0.098)	0.025(0.007)	0.009(0.008)	1.150(0.121)	36.000(1.370)
		No penalty	2.920(1.090)	0.168(0.062)	0.006(0.008)	1.160(0.115)	1.930(0.767)
	15	Proposed	0.308(0.068)	0.018(0.004)	0.006(0.007)	1.090(0.103)	71.500(6.490)
		No penalty	1.910(0.615)	0.110(0.035)	0.005(0.007)	1.110(0.096)	2.580(0.497)
	20	Proposed	0.255(0.049)	0.015(0.003)	0.004(0.006)	1.050(0.095)	147.000(33.100)
		No penalty	1.430(0.475)	0.083(0.027)	0.003(0.004)	1.070(0.091)	3.960(0.687)

NOTE: We include the empirical means of $\text{MSE}(\hat{\gamma})$'s, $\text{MSE}(\hat{\beta})$'s, $\text{RMSE}(\hat{\beta})$'s, PMSEs, and computation times in minutes with their standard errors in the parentheses. Results for LDFR_y, LDFR_{xz} and LFPCA with $n = 400$ are not available due to out of memory. For each case, 200 simulated datasets were used.

Table 2. Results from a simulation study of the logistic case in Example 1 for two methods, 3 $\lambda_{\text{Poisson}} \in \{10, 15, 20\}$, and 3 sample sizes $n \in \{100, 200, 400\}$.

n	λ_{Poisson}		MSE_{β}	RMSE_{β}	MSE_{γ}	PMSE _g	Classification error	Time (min)
100	10	Proposed	3.230(0.852)	0.186(0.049)	0.169(0.101)	0.007(0.002)	0.228(0.018)	3.530(2.070)
		No penalty	18.700(7.530)	1.080(0.433)	0.302(0.123)	0.028(0.006)	0.269(0.021)	0.761(0.381)
	15	Proposed	2.370(0.626)	0.137(0.036)	0.123(0.069)	0.005(0.002)	0.216(0.016)	9.900(1.100)
		No penalty	13.700(3.730)	0.789(0.212)	0.293(0.112)	0.023(0.005)	0.265(0.019)	0.984(0.432)
	20	Proposed	1.750(0.526)	0.101(0.030)	0.082(0.054)	0.003(0.001)	0.208(0.013)	14.800(4.780)
		No penalty	11.100(2.430)	0.642(0.139)	0.280(0.112)	0.020(0.005)	0.259(0.020)	1.410(0.577)
200	10	Proposed	2.860(0.650)	0.165(0.038)	0.154(0.065)	0.006(0.001)	0.224(0.017)	15.400(9.460)
		No penalty	10.300(2.060)	0.595(0.117)	0.296(0.093)	0.020(0.004)	0.263(0.019)	1.020(0.377)
	15	Proposed	2.020(0.478)	0.116(0.028)	0.103(0.049)	0.004(0.001)	0.211(0.013)	34.300(12.500)
		No penalty	8.390(1.410)	0.484(0.081)	0.281(0.078)	0.017(0.004)	0.256(0.018)	1.490(0.302)
	20	Proposed	1.600(0.360)	0.092(0.020)	0.073(0.033)	0.003(0.001)	0.205(0.011)	55.100(13.600)
		No penalty	7.460(1.490)	0.431(0.086)	0.250(0.081)	0.014(0.004)	0.249(0.019)	1.570(0.698)
400	10	Proposed	2.600(0.484)	0.150(0.028)	0.130(0.044)	0.005(0.001)	0.218(0.013)	37.300(26.700)
		No penalty	7.470(1.120)	0.431(0.063)	0.279(0.066)	0.016(0.003)	0.257(0.017)	2.560(0.603)
	15	Proposed	1.720(0.428)	0.099(0.025)	0.085(0.037)	0.003(0.001)	0.207(0.012)	98.000(39.200)
		No penalty	6.400(0.967)	0.369(0.055)	0.260(0.060)	0.014(0.003)	0.251(0.016)	4.080(0.423)
	20	Proposed	1.370(0.287)	0.079(0.017)	0.060(0.023)	0.002(0.001)	0.200(0.012)	182.000(43.900)
		No penalty	5.650(0.995)	0.326(0.057)	0.233(0.060)	0.012(0.004)	0.244(0.017)	4.800(2.550)

NOTE: We include the empirical means of $\text{MSE}(\hat{\gamma})$'s, $\text{MSE}(\hat{\beta})$'s, $\text{RMSE}(\hat{\beta})$'s, PMSEs, and computation times in minutes with their standard errors in the parentheses. For each case, 200 simulated datasets were used.

Table 3. Simulation study in Example 2 for T_β : estimates of rejection rates were reported at 3 different λ_{Poisson} 's (10, 15, 20), 2 different cases (linear and logistic), 4 B values ($B = 0, 0.1, 0.3, 0.5$), and 3 different sample sizes ($n = 100, 200, 400$) at $\alpha = 5\%$.

n	λ_{Poisson}	Linear				Logistic			
		$B = 0$	$B = 0.1$	$B = 0.3$	$B = 0.5$	$B = 0$	$B = 0.1$	$B = 0.3$	$B = 0.5$
100	10	6.8	8.2	29.1	76.7	5.2	6.8	21.0	70.0
	15	6.5	7.8	34.2	87.9	5.5	9.0	45.4	96.8
	20	6.7	8.2	39.1	91.6	5.2	9.0	59.8	99.4
200	10	6.2	8.8	50.5	98.7	5.5	5.0	39.0	96.2
	15	6.5	10.1	69.5	100.0	5.4	9.8	68.4	100.0
	20	5.5	9.8	76.5	100.0	5.0	11.4	83.8	100.0
400	10	5.0	10.0	92.9	100.0	5.5	7.8	61.8	100.0
	15	5.3	13.3	98.3	100.0	5.0	10.4	82.6	100.0
	20	5.5	15.1	99.9	100.0	5.8	9.42	93.4	100.0

NOTE: For each case, 1000 simulated datasets were used, while 1000 bootstrap samples were used for each simulated dataset.

Table 4. Simulation study for T_γ in Example 2: estimates of rejection rates were reported at 3 different λ_{Poisson} 's (10, 15, 20), 2 different cases (linear and logistic), 4 γ values ($\gamma = 0, 0.1, 0.3, 0.5$), and 3 different sample sizes ($n = 100, 200, 400$) at $\alpha = 5\%$.

n	λ_{Poisson}	Linear				Logistic			
		$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.5$
100	10	6.3	21.6	88.2	100.0	5.8	26.2	97.4	100.0
	15	6.2	24.0	93.8	100.0	6.3	37.8	99.4	100.0
	20	5.9	23.8	95.7	100.0	6.4	52.6	100.0	100.0
200	10	6.5	34.2	99.2	100.0	5.8	40.6	99.8	100.0
	15	5.9	40.7	99.8	100.0	5.5	60.8	100.0	100.0
	20	5.3	42.0	99.9	100.0	4.8	75.4	100.0	100.0
400	20	5.2	54.6	100.0	100.0	5.0	51.0	100.0	100.0
	15	4.8	62.1	100.0	100.0	5.5	78.8	100.0	100.0
	20	5.3	68.6	100.0	100.0	5.8	94.2	100.0	100.0

NOTE: For each case, 1000 simulated datasets were used, while 1000 bootstrap samples were used for each simulated dataset.

Multinomial distribution $\text{Multi}(n, (n^{-1}, \dots, n^{-1}))$. The significant level is set at 5% throughout the simulation.

Tables 3 and 4 present the Type I and II error rates of testing $\beta_0(t, u) = 0$ and those of testing $\gamma = 0$, respectively. Inspecting Tables 3 and 4 reveals that the proposed bootstrap procedure performs well. For both test statistics, the Type I error rates were relatively accurate, while the statistical power for rejecting the null hypotheses was significantly increased with sample size and signal. We also present the histograms of the p -values corresponding to $\lambda_{\text{Poisson}} = 10$ for testing $\beta_0(t, u) = 0$ and $\gamma = 0$ in the supplementary materials. These histograms are close to the uniform distribution, demonstrating the validity of the proposed test procedure.

Example 3. We are interested in assessing the finite sample performance of the proposed test procedure for testing the nullity of one functional coefficient with the presence of other functional predictors. The datasets were generated from

$$Y(t) = Z(t)\gamma_0 + \int_0^1 X_1(t, u)\beta_{01}(t, u)du + \int_0^1 X_2(t, u)\beta_{02}(t, u)du + \epsilon(t),$$

where $Z(t)$ and $\epsilon(t)$ were generated in the same way as those in Example 1, and $X_1(t, u)$ and $X_2(t, u)$ were generated in the same way as $X(t, u)$. To test the nullity of β_{01} , we fix $\gamma_0 = 1$ and $\beta_{02}(t, u) = \sqrt{tu}$, and set $\beta_{01}(t, u) = B * \sum_{k=1}^{50} (-1)^{k+1} k^{-2} \vartheta(t, k) \vartheta(u, k)$, where $B \in \{0, 0.1, 0.3, 0.5\}$.

Table 5. Simulation study in Example 3 for using T_β to test the nullity of $\beta_{01}(t, u)$: estimates of rejection rates were reported at 3 different λ_{Poisson} 's (10, 15, 20), 4 B values ($B = 0, 0.1, 0.3, 0.5$), and 3 different sample sizes ($n = 100, 200, 400$) at $\alpha = 5\%$.

n	λ_{Poisson}	$B = 0$	$B = 0.1$	$B = 0.3$	$B = 0.5$
100	10	5.5	6.8	25.8	78.2
	15	6.8	8.8	32.3	86.5
	20	6.1	9.5	36.3	90.0
200	10	6.4	9.2	51.6	99.1
	15	5.5	9.0	70.3	100.0
	20	5.8	9.5	76.5	100.0
400	10	5.5	11.1	92.7	100.0
	15	5.1	12.9	98.3	100.0
	20	5.0	14.4	99.7	100.0

NOTE: For each case, 1000 simulated datasets were used, while 1000 bootstrap samples were used for each simulated dataset.

Similar to the results in Example 2, Table 5 shows that the proposed test maintains the Type I error rate reasonably well, while achieving high power in detecting alternative hypotheses.

5. Application to the ADNI Data

Alzheimer's disease as a chronic neurodegenerative disease is the most common cause of dementia. AD usually starts slowly and worsens over time with the degeneration and death of brain cells and decline in thinking, behavioral and social skills that disrupts a person's ability to function independently. Although much progress has been made on understanding the etiology

of AD, the exact cause of Alzheimer's disease remains not fully understood.

We applied model (2) to the DWI dataset collected from ADNI GO and ADNI2 in the ADNI study. One goal of the ADNI study is to test whether genetic, structural, functional neuroimaging and clinical data can be integrated to measure the progression of MCI and early AD. The MMSE score ranging from 0 to 30 measures global cognitive performance and is often used as a screening test for dementia.

The DWI data were processed by using an FSL TBSS pipeline (McMahon and Thompson 2017). We obtained maps of FA by a diffusion tensor model after eddy current correction and automatic brain extraction (Basser, Mattiello, and LeBihan 1994; Smith 2002; Andersson and Sotiropoulos 2016). Then, we did linear registration by FSL to register the individual FA images to the ENIGMA FA template with $1 \times 1 \times 1 \text{ mm}^3$ resolution. We aligned the linearly registered FA images to the ENIGMA FA template through nonlinear registration and masked the registered FA with the template mask. Finally, we obtained the FA density values within each of 20 ROIs contoured on white matter of the FA template by the experienced clinical expert. Index of the 20 ROIs are given in Table 7. Figure 2 shows the spatial locations of the 20 ROIs in the brain. Since density functions do not live in a linear space, the commonly used methods for functional data analysis are not applicable. To address this problem, we applied the log hazard transformation proposed in Petersen and Müller (2016) and took the log hazard function of FA density curves as functional variables.

The dataset includes $n = 256$ subjects over a 5-year follow-up. For each subject, we consider log hazard function of FA along the whole brain and log hazard functions of FA along 20 ROIs at 100 grid points observed at 1–8 time points, the MMSE scores examined from 1 to 7 time points, diagnostic status at baseline (NC, MCI, or AD), age ranging from 55 to 92 (years), the number of APOE4 gene copies, and education level ranging from 11 to 20 (years). The measured time points of the log hazard

functions of FA and the MMSE scores were different between and within subjects. The aim of this study is to establish the association between the MMSE scores and log hazard functions of FA across the whole brain and individual ROIs measured at different ages, while accounting for demographic and clinical covariates.

We first examine the effects of log hazard function of FA along the whole brain, age, education level, and APOE4 on the MMSE score by considering the following model

$$\begin{aligned} Y_{\text{MMSE}}(t) = & \gamma_0 + Z_{\text{MCI}}\gamma_{\text{MCI}} + Z_{\text{AD}}\gamma_{\text{AD}} + Z_{\text{age}}(t)\gamma_{\text{age}} \\ & + Z_{\text{edu}}\gamma_{\text{edu}} + Z_{\text{APOE4}(1)}\gamma_{\text{APOE4}(1)} \\ & + Z_{\text{APOE4}(2)}\gamma_{\text{APOE4}(2)} + \int_0^1 \text{FA}_{\text{WB}}(t, u)\beta(t, u)du \\ & + \epsilon(t), \end{aligned} \quad (15)$$

where t denotes age rescaled to $[0, 1]$ and $\text{FA}_{\text{WB}}(t_0, u)$ denotes the log hazard function along the whole brain at a given age t_0 . Moreover, Z_{MCI} , Z_{AD} , $Z_{\text{APOE4}(1)}$, and $Z_{\text{APOE4}(2)}$ are, respectively, the indicator variables for MCI, AD, one copy of APOE4, and two copies of APOE4, respectively. We implemented the estimation procedure in Section 2.1 to compute the estimates of all coefficients and applied the testing procedure in Section 2.3 to assess the effects of possible covariates on the MMSE scores. We also applied LDFR_y, LDFR_{xz}, and LFPCA to the same dataset under model (15). Table 6 summarizes the coefficient estimates and p -values of the scalar parameters for all the four estimation methods. The p -values were obtained by using 1000 bootstrap samples. Except for LDFR_y, LDFR_{xz}, LFPCA, and our method produce similar results for the scalar parameters. Such differences between LDFR_y and all the three other methods may be caused by pre-smoothing the sparsely observed MMSE scores, which may not be appropriate for this dataset. The estimation results show that the disease MCI at baseline has a significant negative effect on MMSE, and AD has a more serious effect

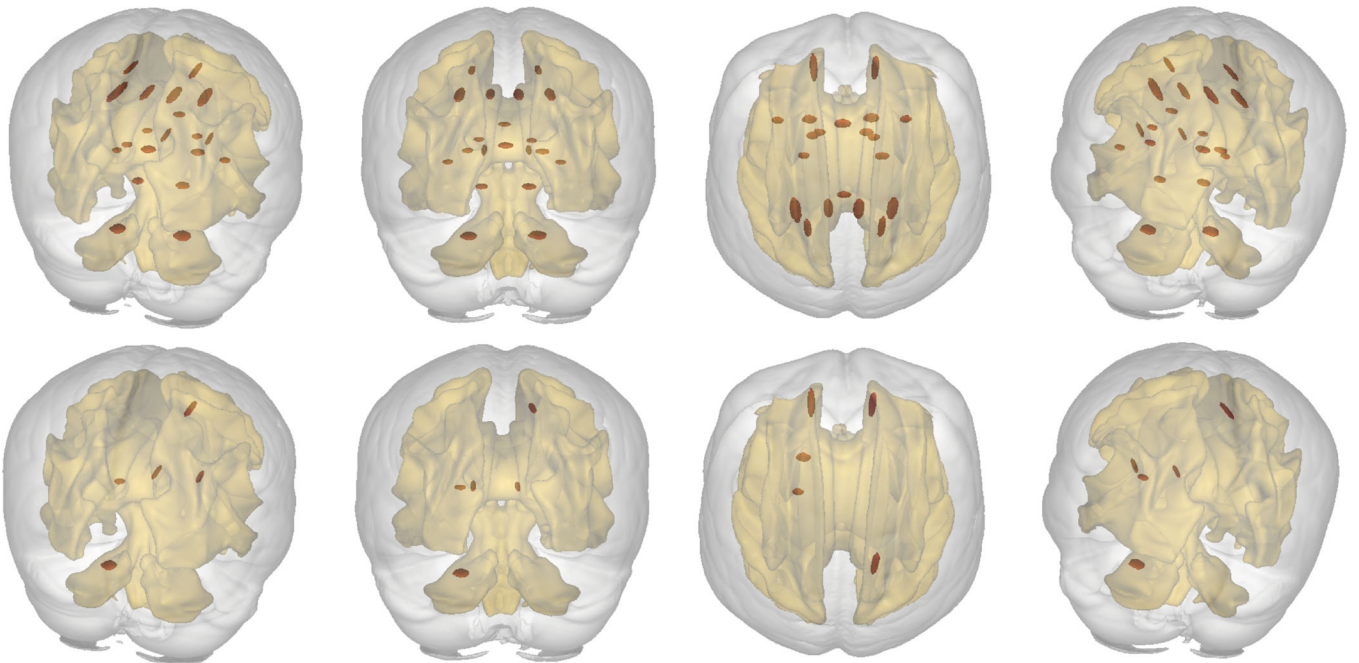


Figure 2. In the first row are the spatial locations of the 20 ROIs, in the second row are spatial locations of the 5 significant ROIs.

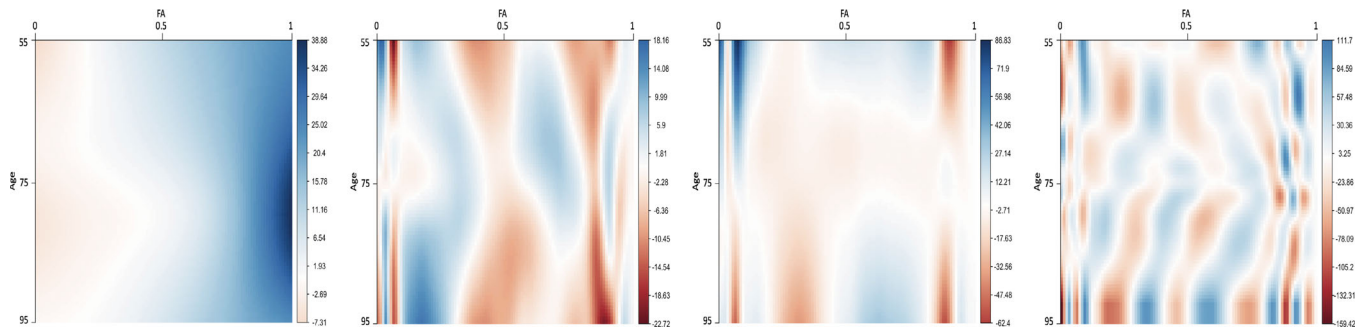
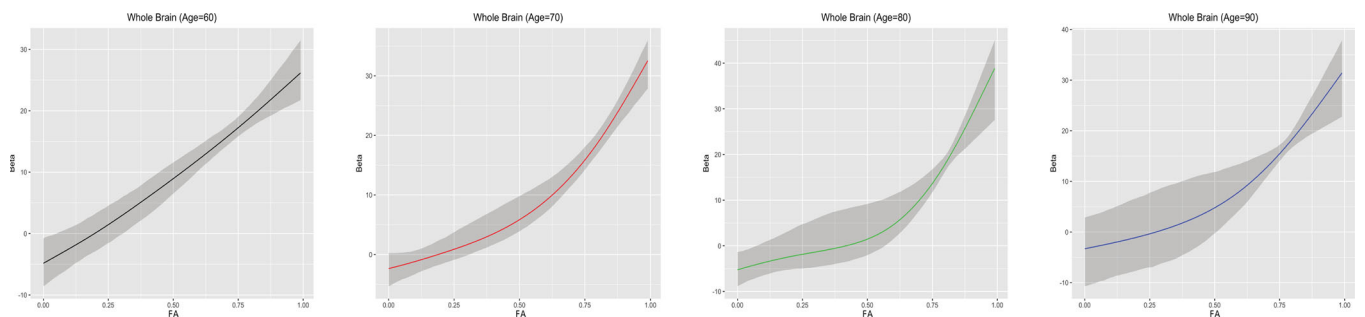
Table 6. ADNI real data analysis results: estimates of scalar parameters and their p -values obtained from the four different methods under model (15).

		γ_0	γ_{MCI}	γ_{AD}	γ_{age}	γ_{edu}	$\gamma_{APOE4(1)}$	$\gamma_{APOE4(2)}$
Proposed	Estimate	1.467	-1.430	-5.915	0.303	0.130	-0.965	-1.306
	p -value	0.944	0.001	0.001	0.268	0.012	0.001	0.023
LDFR _y	Estimate	-3e+8	-0.448	-0.072	4e+6	-0.036	-0.156	1.084
	p -value	0.001	0.001	0.667	0.333	0.333	0.001	0.667
LDFR _{xz}	Estimate	27.380	-1.477	-5.905	0.012	0.057	-1.289	-1.307
	p -value	0.001	0.001	0.001	0.272	0.241	0.036	0.004
LFPCA	Estimate	32.196	-1.365	-5.891	-0.072	0.134	-0.859	-1.122
	p -value	0.001	0.001	0.001	0.001	0.008	0.034	0.002

Table 7. ADNI real data analysis results: the p -values and corrected p -values of the 20 ROIs.

	p -value	Corrected		p -value	Corrected
Left anterior limb internal capsule	0.298	0.362	Right anterior limb internal capsule	0.466	0.491
Genu corpus callosum	0.134	0.234	Left cerebellar peduncle	0.043	0.172
Left cerebral peduncle	0.202	0.289	Left cingulate gyrus	0.326	0.362
Left convexity	0.105	0.210	Left corona radiata	0.143	0.234
Left external capsule	0.065	0.200	Left orbital gyrus	0.04	0.172
Left posterior limb of the internal capsule	0.016	0.172	Right posterior limb of the internal capsule	0.07	0.200
Right cerebellar peduncle	0.104	0.210	Right cerebral peduncle	0.091	0.210
Right cingulate gyrus	0.800	0.800	Right convexity	0.019	0.172
Right corona radiata	0.152	0.234	Right external capsule	0.311	0.362
Right orbital gyrus	0.042	0.172	Splenium of the corpus callosum	0.310	0.360

NOTE: The regions with corrected p -values smaller than 0.2 (marked as bold values) in Table 7 are significant.

**Figure 3.** Estimates of $\beta(t, u)$ along the whole brain in model(15) for the proposed method, LDFR_y, LDFR_{xz}, and LFPCA (from left to right).**Figure 4.** Estimates and pointwise confidence intervals for $\beta(t, u)$ along the whole brain at ages 60, 70, 80, and 90 in model (15).

than MCI, whereas education exhibits positive effect (Bekris et al. 2010). The APOE4 allele has a negative effect on cognitive ability, and two copies of APOE4 has a more serious effect than one copy of APOE4. Age is not detected as a significant variable based on our proposed method.

We estimated $\beta(t, u)$, denoted as $\hat{\beta}(t, u)$, in model (15) and tested its nullity by using all the four estimation methods. For our method, the p -value is 0.001, which is significant at the 5% significance level. In contrast, by using a similar bootstrap

testing procedure, LDFR_y, LDFR_{xz}, and LFPCA give the p -values of 0.999, 0.950, and 0.480, respectively. Figure 3 presents the estimated functional coefficients of the log hazard function of FA for all the four methods. The proposed method produces more smoother functional coefficient than the three other methods. Inspecting $\hat{\beta}(t, u)$ reveals a positive relationship between log hazard function and MMSE score at large FA values and a negative relationship at small FA values. Furthermore, Figure 4 presents $\hat{\beta}(t, u)$'s at four age values including $t = 60, 70, 80$, and

90 with their 95% pointwise bootstrap confidence intervals. The log hazard function of FA shows a positive effect on MMSE. This finding is consistent with that in the existing literature (Bozzali et al. 2002) that lower FA values, which typically indicate greater white matter deficits, are associated with lower MMSE scores, and can be viewed as an indicator of great impairment. Moreover, the effect of the log hazard function on the MMSE scores varies across time, but with similar trends. Specifically, when patients are at ages 60 and 70, the log hazard function at FA values larger than 0.250 has a significant positive effect on the MMSE scores, and such effect increases with FA. These results reveal that only log hazard function at large FA values exhibits a significant positive effect on cognitive ability, and the significant range of FA values becomes smaller for older people.

To relate changes in cognitive status to the corresponding changes in each specific ROI, we extended model (15) by replacing the log hazard function of FA along the whole brain with that along each specific ROI and then applied our method to fit the model. Thus, we had 20 models corresponding to the 20 ROIs. Estimates of the functional coefficients of the 20 ROIs are quite similar to the functional estimate of the whole brain. This is reasonable because there exist correlations between the log hazard functions along the 20 ROIs and that along the whole brain. For example, correlations between the 25%, 50%, and 75% quantiles of the log hazard function along the whole brain and right corona radiata are 0.265, 0.323, and 0.432, respectively.

We also carried out additional analyses by investigating the effects of the log hazard function of FA along each ROI on MMSE, while controlling that along the whole brain. For each $d = 1, \dots, 20$, we consider the following model

$$Y_{\text{MMSE}}(t) = \gamma_0 + Z_{\text{MCI}}\gamma_{\text{MCI}} + Z_{\text{AD}}\gamma_{\text{AD}} + Z_{\text{age}}(t)\gamma_{\text{age}} \\ + Z_{\text{edu}}\gamma_{\text{edu}} + Z_{\text{APOE4}(1)}\gamma_{\text{APOE4}(1)}$$

$$+ Z_{\text{APOE4}(2)}\gamma_{\text{APOE4}(2)} + \int_0^1 FA_d(t, u)\beta_d(t, u)du \\ + \int_0^1 FA_{\text{WB}}(t, u)\beta(t, u)du + \epsilon(t), \quad (16)$$

where $FA_d(t, u)$ denotes the log hazard function along the d th ROI at a given age t . By using the testing procedure in Section 2.3 with 1000 bootstrap samples, we obtained 20 p -values of testing the nullity of the functional coefficients across the 20 ROIs. To identify important ROIs from the 20 ROIs tested, we adopted the false discovery rate correction procedure of Benjamini and Hochberg (1995) with the commonly used level 0.2 (Han et al. 2010; Glickman, Rao, and Schultz 2014; Weinstein and Yekutieli 2020), and calculated their corresponding corrected p -values by using the $p.adjust(\text{method}=\text{"BH"})$ function in the *stats* R package. ROIs with corrected p -values smaller than 0.2 are declared to be significant. Table 7 presents the p -values and corrected p -values of the 20 ROIs.

We also examined the 5 significant ROIs identified with their corrected p -values smaller than 0.2 including left cerebellar peduncle, left orbital gyrus, left posterior limb of the internal capsule, right convexity and right orbital gyrus. Figure 2 presents the spatial locations of the 5 ROIs, whereas Figure 5 presents the estimated functional coefficients corresponding to the 5 significant ROIs. Log hazard functions along the 5 ROIs exhibit positive effects on cognitive status at large FA values, and their effects vary across age. Figure 6 gives the estimates with 95% pointwise bootstrap confidence intervals at ages 60, 70, 80, and 90 for left cerebellar peduncle. We observe a positive effect of the log hazard function at FA values larger than 0.800 on the MMSE score for people between ages 80 and 90. We observe similar patterns for all the other four regions and include their corresponding results in the supplementary materials.

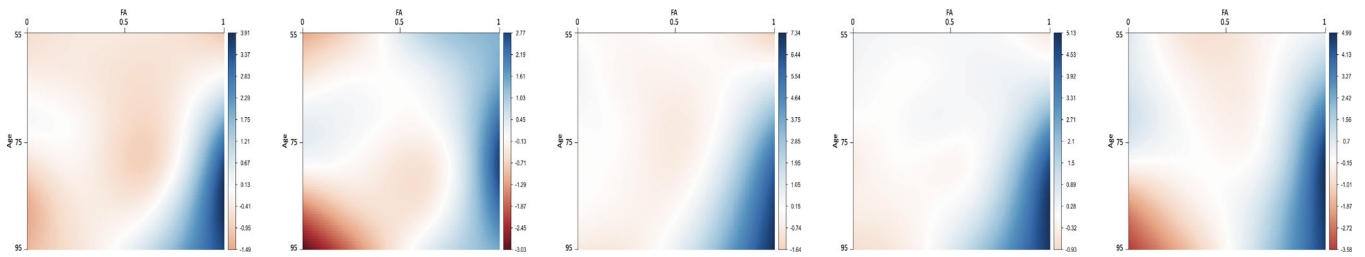


Figure 5. Estimates of $\beta(t, u)$ in model (16) along left cerebellar peduncle, left orbital gyrus, left posterior limb of the internal capsule right convexity and right orbital gyrus (from left to right).

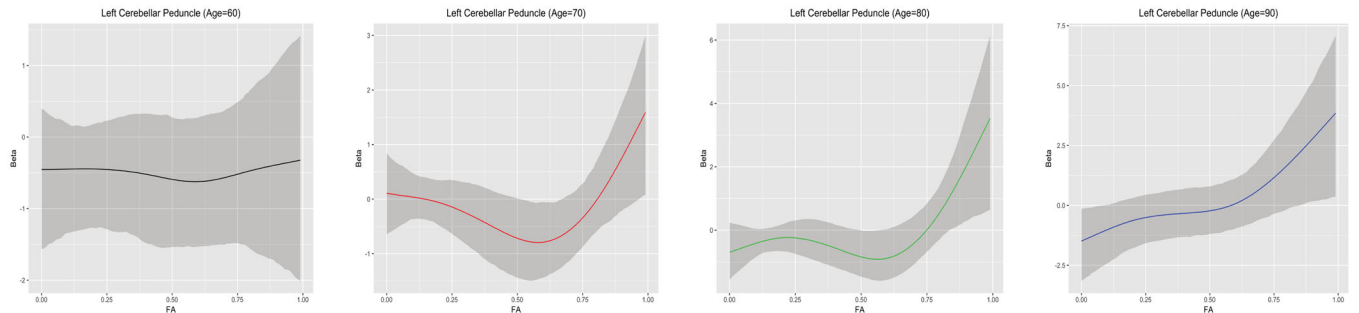


Figure 6. Estimates and pointwise confidence intervals for $\beta(t, u)$ along left cerebellar peduncle at ages 60, 70, 80, and 90 in model (16).

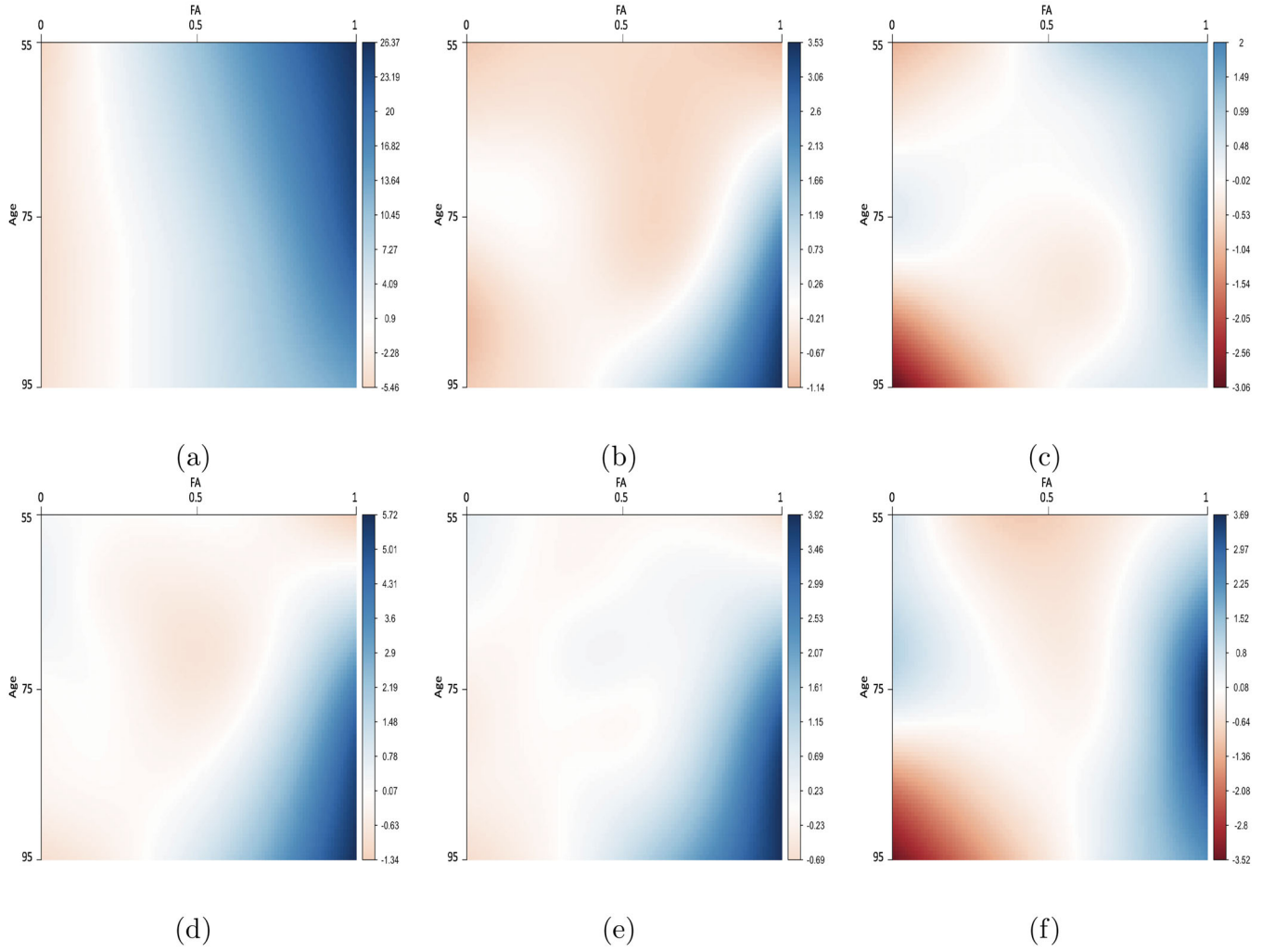


Figure 7. Estimated functional coefficients in model (17) along whole brain (a), left cerebellar peduncle (b), left orbital gyrus (c), left posterior limb of the internal capsule (d), right convexity (e), and right orbital gyrus (f).

Table 8. ADNI real data analysis results: estimates of scalar parameters and their p -values obtained from the proposed method under model (17).

	γ_0	γ_{MCI}	γ_{AD}	γ_{age}	γ_{edu}	$\gamma_{APOE4(1)}$	$\gamma_{APOE4(2)}$
Estimate	9.054	-1.240	-5.786	0.201	0.113	-0.768	-1.028
p -value	0.668	0.001	0.001	0.488	0.021	0.039	0.002

Finally, we integrated all the log hazard functions of FA along the 5 significant ROIs and along the whole brain into the following model given by

$$\begin{aligned}
 Y_{MMSE}(t) = & \gamma_0 + Z_{AD}\gamma_{AD} + Z_{age}(t)\gamma_{age} + Z_{edu}\gamma_{edu} \\
 & + Z_{APOE4(1)}\gamma_{APOE4(1)} + Z_{APOE4(2)}\gamma_{APOE4(2)} \\
 & + \sum_{d=1}^5 \int_0^1 FA_d(t, u)\beta_d(t, u)du \\
 & + \int_0^1 FA_{WB}(t, u)\beta(t, u)du + \epsilon(t). \quad (17)
 \end{aligned}$$

Table 8 presents the estimates and p -values of all scalar estimates. The p -values were also obtained by using the testing procedure in Section 2.3 with 1000 bootstrap samples. The

results are similar to those for model (15). The p -values of left cerebellar peduncle, left orbital gyrus, left posterior limb of the internal capsule, right convexity, right orbital gyrus, and the whole brain are, respectively, given by 0.031, 0.036, 0.020, 0.041, 0.019, and 0.001 and are significant at the 5% significance level.

Figure 7 presents the estimated functional coefficients for the 5 ROIs and the whole brain in model (17). Additional results on all estimates with 95% pointwise bootstrap confidence intervals at ages 60, 70, 80, and 90 for the whole brain and the 5 ROIs can be found in the supplementary materials. Inspecting Figure 7 reveals that the trends of the estimated functional coefficients for the whole brain in model (17) are similar to those in model (15), except that the estimate in model (15) has a stronger effect on the MMSE score than that in model (17). The estimated coefficient functions for the 5 ROIs and accompanying confidence intervals at ages 60, 70, 80, and 90 are largely unchanged compared with those in model (16). These findings may reveal that the deterioration of intelligence is associated with the decreasing of the top quantiles of FA at the whole brain, while for older people (e.g., age > 80), the decreasing of the top quantiles of FA at 5 specific ROIs exhibits additional effects on the intelligence deterioration.

6. Discussion

Motivated by the clinical studies where functional predictors and the outcome are longitudinally observed at different times, we proposed a GFPV model for asynchronous longitudinally observed functional data. The time-varying functional coefficient $\beta(t, u)$ was represented by a rich truncated tensor product penalized B-spline basis. To address the asynchronous data setting and the increasing dimension after representation, we proposed local kernel-weighted estimating equations with penalty to estimate all the coefficients. The estimators were shown to be consistent and we derived the convergence rates of the estimation error as well as the prediction error. Limiting distributions were also established. The rate of convergence of the functional estimator was slower than the optimal nonparametric rate of convergence owing to the loss of efficiency caused by the mismatched observation times. Meanwhile, driven by practical interest, we proposed a bootstrap testing procedure to test the nullity of the parameters and established the bootstrap consistency. Through the analyses of the ADNI study, we have shown that the proposed method is a valuable statistical tool for quantifying the complex relationship between the FA density curves and the cognitive function under asynchronous data setting. Our results show that the deterioration of intelligence is associated with the decreasing of the top quantiles of FA at the whole brain, while for older people, the decreasing of the top quantiles of FA at 5 specific ROIs exhibits additional effects on the intelligence deterioration.

There are several directions for future study. A useful area for improvement would be to accommodate high-dimensional scalar covariates, because many clinical studies record high-dimensional scalar variables. Another interesting consideration for future research would be to accommodate the commonly encountered situation where the observation times for the response, covariate and functional data are all different. Such extensions are worthy of further investigation.

Supplementary Materials

Assumptions, more explanations on the data structure, all the technical proofs, and additional figures are provided in the supplementary materials.

Funding

Zhongyi Zhu's research was partially supported by National Science Foundation of China grants 11671096, 112071087, and 11731011. Ting Li's work was partially supported by the Fundamental Research Funds for the Central Universities.

References

- Andersson, J. L., and Sotiropoulos, S. N. (2016), "An Integrated Approach to Correction for Off-Resonance Effects and Subject Movement in Diffusion MR Imaging," *Neuroimage*, 125, 1063–1078. [10]
- Basser, P. J., Mattiello, J., and LeBihan, D. (1994), "Estimation of the Effective Self-Diffusion Tensor From the NMR Spin Echo," *Journal of Magnetic Resonance*, Series B, 103, 247–254. [10]
- Bekris, L. M., Yu, C.-E., Bird, T. D., and Tsuang, D. W. (2010), "Genetics of Alzheimer Disease," *Journal of Geriatric Psychiatry and Neurology*, 23, 213–227. [11]
- Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B*, 57, 289–300. [12]
- Bozzali, M., Falini, A., Franceschi, M., Cercignani, M., Zuffi, M., Scotti, G., Comi, G., and Filippi, M. (2002), "White Matter Damage in Alzheimer's Disease Assessed In Vivo Using Diffusion Tensor Magnetic Resonance Imaging," *Journal of Neurology, Neurosurgery & Psychiatry*, 72, 742–746. [1,12]
- Cao, H., Li, J., and Fine, J. P. (2016), "On Last Observation Carried Forward and Asynchronous Longitudinal Regression Analysis," *Electronic Journal of Statistics*, 10, 1155–1180. [2]
- Cao, H., Zeng, D., and Fine, J. P. (2015), "Regression Analysis of Sparse Asynchronous Longitudinal Data," *Journal of the Royal Statistical Society, Series B*, 77, 755–776. [2,3,5,6,7]
- Cardot, H., Ferraty, F., and Sarda, P. (2003), "Spline Estimators for the Functional Linear Model," *Statistica Sinica*, 13, 571–591. [5]
- Cardot, H., and Sarda, P. (2005), "Estimation in Generalized Linear Models for Functional Data via Penalized Likelihood," *Journal of Multivariate Analysis*, 92, 24–41. [5]
- Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 92, 477–489. [3]
- Chatterjee, S., and Bose, A. (2005), "Generalized Bootstrap for Estimating Equations," *The Annals of Statistics*, 33, 414–436. [4]
- Chen, K., Delicado, P., and Müller, H.-G. (2017), "Modelling Function-Valued Stochastic Processes, With Applications to Fertility Dynamics," *Journal of the Royal Statistical Society, Series B*, 79, 177–196. [2]
- Chen, K., and Müller, H.-G. (2012), "Modeling Repeated Functional Observations," *Journal of the American Statistical Association*, 107, 1599–1609. [2]
- Chen, L., and Cao, H. (2017), "Analysis of Asynchronous Longitudinal Data With Partially Linear Models," *Electronic Journal of Statistics*, 11, 1549–1569. [2,3,5,6]
- Cheng, G., and Huang, J. Z. (2010), "Bootstrap Consistency for General Semiparametric M -Estimation," *The Annals of Statistics*, 38, 2884–2915. [4]
- Cheng, M.-Y., Zhang, W., and Chen, L.-H. (2009), "Statistical Estimation in Generalized Multiparameter Likelihood Models," *Journal of the American Statistical Association*, 104, 1179–1191. [5]
- Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009), "Multilevel Functional Principal Component Analysis," *The Annals of Applied Statistics*, 3, 458. [2]
- Diggle, P. (2002), *Analysis of Longitudinal Data*, Oxford: Oxford University Press. [2]
- Fan, J., Yao, Q., and Cai, Z. (2003), "Adaptive Varying-Coefficient Linear Models," *Journal of the Royal Statistical Society, Series B*, 65, 57–80. [4]
- Fu, W. J. (2003), "Penalized Estimating Equations," *Biometrics*, 59, 126–132. [3]
- Gertheiss, J., Goldsmith, J., Crainiceanu, C., and Greven, S. (2013), "Longitudinal Scalar-on-Functions Regression With Application to Tractography Data," *Biostatistics*, 14, 447–461. [3]
- Glickman, M. E., Rao, S. R., and Schultz, M. R. (2014), "False Discovery Rate Control Is a Recommended Alternative to Bonferroni-Type Adjustments in Health Studies," *Journal of Clinical Epidemiology*, 67, 850–857. [12]
- Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012), "Longitudinal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements," *Journal of the Royal Statistical Society, Series C*, 61, 453–469. [3]
- Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2011), "Longitudinal Functional Principal Component Analysis," in *Recent Advances in Functional Data Analysis and Related Topics*, Springer, pp. 149–154. [2]
- Greven, S., and Scheipl, F. (2017), "A General Framework for Functional Regression Modelling," *Statistical Modelling*, 17, 1–35. [3]
- Han, S., Lee, K.-M., Park, S. K., Lee, J. E., Ahn, H. S., Shin, H. Y., Kang, H. J., Koo, H. H., Seo, J. J., Choi, J. E., and Ahn, Y. O. (2010), "Genome-Wide Association Study of Childhood Acute Lymphoblastic Leukemia in Korea," *Leukemia Research*, 34, 1271–1274. [12]

- Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C. A., Jeste, S., DiStefano, C., and Şentürk, D. (2017), "A Multi-Dimensional Functional Principal Components Analysis of EEG Data," *Biometrics*, 73, 999–1009. [2]
- Kong, D., Staicu, A.-M., and Maity, A. (2016), "Classical Testing in Functional Linear Models," *Journal of Nonparametric Statistics*, 28, 813–838. [6]
- Kundu, M. G., Harezlak, J., and Randolph, T. W. (2016), "Longitudinal Functional Models With Structured Penalties," *Statistical Modelling*, 16, 114–139. [3]
- Lee, W., Miranda, M. F., Rausch, P., Baladandayuthapani, V., Fazio, M., Downs, J. C., and Morris, J. S. (2019), "Bayesian Semiparametric Functional Mixed Models for Serially Correlated Functional Data, With Application to Glaucoma Data," *Journal of the American Statistical Association*, 114, 495–513. [2]
- Leow, A. D., Yanovsky, I., Parikshak, N., Hua, X., Lee, S., Toga, A. W., Jack, C. R., Jr., Bernstein, M. A., Britson, P. J., and Gunter, J. L. (2009), "Alzheimer's Disease Neuroimaging Initiative: A One-Year Follow Up Study Using Tensor-Based Morphometry Correlating Degenerative Rates, Biomarkers and Cognition," *Neuroimage*, 45, 645–655. [2]
- Li, J., Huang, C., Zhu, H., and ADNI (2017), "A Functional Varying-Coefficient Single-Index Model for Functional Response Data," *Journal of the American Statistical Association*, 112, 1169–1181. [2]
- Li, Y., and Guan, Y. (2014), "Functional Principal Component Analysis of Spatiotemporal Point Processes With Applications in Disease Surveillance," *Journal of the American Statistical Association*, 109, 1205–1215. [2]
- Lin, D., and Ying, Z. (2001), "Semiparametric and Nonparametric Regression Analysis of Longitudinal Data," *Journal of the American Statistical Association*, 96, 103–126. [2]
- Lin, X., and Carroll, R. J. (2001), "Semiparametric Regression for Clustered Data Using Generalized Estimating Equations," *Journal of the American Statistical Association*, 96, 1045–1056. [3]
- McMahon, M., and Thompson, P. (2017), "Enhancing Neuro Imaging Genetics Through Meta Analysis: Global Collaborations in Psychiatry by the ENIGMA Consortium," *European Neuropsychopharmacology*, 27, S715. [10]
- Morris, J. S., and Carroll, R. J. (2006), "Wavelet-Based Functional Mixed Models," *Journal of the Royal Statistical Society, Series B*, 68, 179–199. [2]
- Morris, J. S., Vannucci, M., Brown, P. J., and Carroll, R. J. (2003), "Wavelet-Based Nonparametric Modeling of Hierarchical Functions in Colon Carcinogenesis," *Journal of the American Statistical Association*, 98, 573–583. [2]
- Nir, T. M., Jahanshad, N., Villalon-Reina, J. E., Toga, A. W., Jack, C. R., Weiner, M. W., and Thompson, P. M. (2013), "Effectiveness of Regional DTI Measures in Distinguishing Alzheimer's Disease, MCI, and Normal Aging," *NeuroImage: Clinical*, 3, 180–195. [1]
- Park, S. Y., and Staicu, A.-M. (2015), "Longitudinal Functional Data Analysis," *Stat*, 4, 212–226. [2,7]
- Park, S. Y., Staicu, A.-M., Xiao, L., and Crainiceanu, C. M. (2018), "Simple Fixed-Effects Inference for Complex Functional Models," *Biostatistics*, 19, 137–152. [2]
- Petersen, A., and Müller, H.-G. (2016), "Functional Data Analysis for Density Functions by Transformation to a Hilbert Space," *The Annals of Statistics*, 44, 183–218. [10]
- Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis* (Vol. 3, 2nd ed.), New York: Springer. [2,4]
- Ruppert, D. (2002), "Selecting the Number of Knots for Penalized Splines," *Journal of Computational and Graphical Statistics*, 11, 735–757. [7]
- Scheffler, A., Telesca, D., Li, Q., Sugar, C. A., Distefano, C., Jeste, S., and Şentürk, D. (2020), "Hybrid Principal Components Analysis for Region-Referenced Longitudinal Functional EEG Data," *Biostatistics*, 21, 139–157. [2]
- Scheipl, F., Gertheiss, J., and Greven, S. (2016), "Generalized Functional Additive Mixed Models," *Electronic Journal of Statistics*, 10, 1455–1492. [3]
- Scheipl, F., Staicu, A.-M., and Greven, S. (2015), "Functional Additive Mixed Models," *Journal of Computational and Graphical Statistics*, 24, 477–501. [3]
- Schumaker, L. L. (1981), *Spline Functions: Basic Theory*, New York: Wiley. [5]
- Şentürk, D., Dalrymple, L. S., Mohammed, S. M., Kaysen, G. A., and Nguyen, D. V. (2013), "Modeling Time-Varying Effects With Generalized and Unsynchronized Longitudinal Data," *Statistics in Medicine*, 32, 2971–2987. [2]
- Smith, S. M. (2002), "Fast Robust Automated Brain Extraction," *Human Brain Mapping*, 17, 143–155. [10]
- Staicu, A.-M., Islam, M. N., Dumitru, R., and Heugten, E. v. (2020), "Longitudinal Dynamic Functional Regression," *Journal of the Royal Statistical Society, Series C*, 69, 25–46. [3,7]
- Staicu, A.-M., Lahiri, S. N., and Carroll, R. J. (2015), "Significance Tests for Functional Data With Complex Dependence Structure," *Journal of Statistical Planning and Inference*, 156, 1–13. [2]
- Stone, C. J. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10, 1040–1053. [5]
- Wang, X., Zhu, H., and ADNI (2017), "Generalized Scalar-on-Image Regression Models via Total Variation," *Journal of the American Statistical Association*, 112, 1156–1168. [2]
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., and Liu, E. (2013), "The Alzheimer's Disease Neuroimaging Initiative: A Review of Papers Published Since Its Inception," *Alzheimer's & Dementia*, 9, e111–e194. [2]
- Weinstein, A., and Yekutieli, D. (2020), "Selective Sign-Determining Multiple Confidence Intervals With FCR Control," *Statistica Sinica*, 30, 531–555. [12]
- Wood, S. N. (2006), "Low-Rank Scale-Invariant Tensor Product Smoother for Generalized Additive Mixed Models," *Biometrics*, 62, 1025–1036. [3]
- Xiong, X., and Dubin, J. A. (2010), "A Binning Method for Analyzing Mixed Longitudinal Data Measured at Distinct Time Points," *Statistics in Medicine*, 29, 1919–1931. [2]
- Yang, J., Cox, D. D., Lee, J. S., Ren, P., and Choi, T. (2017), "Efficient Bayesian Hierarchical Functional Data Analysis With Basis Function Approximations Using Gaussian–Wishart Processes," *Biometrics*, 73, 1082–1091. [2]
- Yu, Y., and Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 97, 1042–1054. [3]
- Zhang, Y., Schuff, N., Du, A.-T., Rosen, H. J., Kramer, J. H., Gorno-Tempini, M. L., Miller, B. L., and Weiner, M. W. (2009), "White Matter Damage in Frontotemporal Dementia and Alzheimer's Disease Measured by Diffusion MRI," *Brain*, 132, 2579–2592. [1]
- Zhu, H., Chen, K., Luo, X., Yuan, Y., and Wang, J.-L. (2019), "Fmem: Functional Mixed Effects Models for Longitudinal Functional Responses," *Statistica Sinica*, 29, 2007. [2]
- Zhu, W., Yuan, Y., Zhang, J., Zhou, F., Knickmeyer, R. C., Zhu, H., and ADNI (2017), "Genome-Wide Association Analysis of Secondary Imaging Phenotypes From the Alzheimer's Disease Neuroimaging Initiative Study," *NeuroImage*, 146, 983–1002. [2]