Received: 30 June 2021 | Accepted: 21 September 2022

DOI: 10.1111/biom.13767

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Asynchronous functional linear regression models for longitudinal data in reproducing kernel Hilbert space

Ting Li[1,2] | **Huichen Zhu**[3] | **Tengfei Li**[4] | **Hongtu Zhu**[5]

[1]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China

[2]Shanghai Institute of International Finance and Economics, Shanghai University of Finance and Economics, Shanghai, China

[3]Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong

[4]Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[5]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Correspondence
Hongtu Zhu, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
Email: htzhu@email.unc.edu

Funding information
National Science Foundation of China, Grant 12101388; Program for Innovative Research Team of Shanghai University of Finance and Economics

Abstract

Motivated by the analysis of longitudinal neuroimaging studies, we study the longitudinal functional linear regression model under asynchronous data setting for modeling the association between clinical outcomes and functional (or imaging) covariates. In the asynchronous data setting, both covariates and responses may be measured at irregular and mismatched time points, posing methodological challenges to existing statistical methods. We develop a kernel weighted loss function with roughness penalty to obtain the functional estimator and derive its representer theorem. The rate of convergence, a Bahadur representation, and the asymptotic pointwise distribution of the functional estimator are obtained under the reproducing kernel Hilbert space framework. We propose a penalized likelihood ratio test to test the nullity of the functional coefficient, derive its asymptotic distribution under the null hypothesis, and investigate the separation rate under the alternative hypotheses. Simulation studies are conducted to examine the finite-sample performance of the proposed procedure. We apply the proposed methods to the analysis of multitype data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, which reveals significant association between 21 regional brain volume density curves and the cognitive function. Data used in preparation of this paper were obtained from the ADNI database (adni.loni.usc.edu).

KEYWORDS

asynchronous longitudinal functional data, Bahadur representation, functional regression, kernel-weighted loss function, penalized likelihood ratio test, reproducing kernel Hilbert space

1 | INTRODUCTION

This paper is motivated by the analysis of a real neuroimaging data set collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (http://www.adni-info.org/). The ADNI has collected imaging, genetic, clinical, and cognitive data at multiple time points from three groups of subjects, including cognitive normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD). The mini-mental state examination (MMSE) score is used to measure cognitive function over time with lower scores indicating cognitive impairment. The magnetic resonance imaging (MRI) data set considered here consists of $n = 770$ subjects over a 5-year follow-up. For each subject, we consider brain local volumetric measures in the 87 regions of interest (ROIs), abbreviated as local volumetric curves from now on, for characterizing regional atrophy (Zhao et al., 2019). The local volumetric curves were observed at 1 to 6 time points, whereas the MMSE scores were examined at 1 to 13 time points.

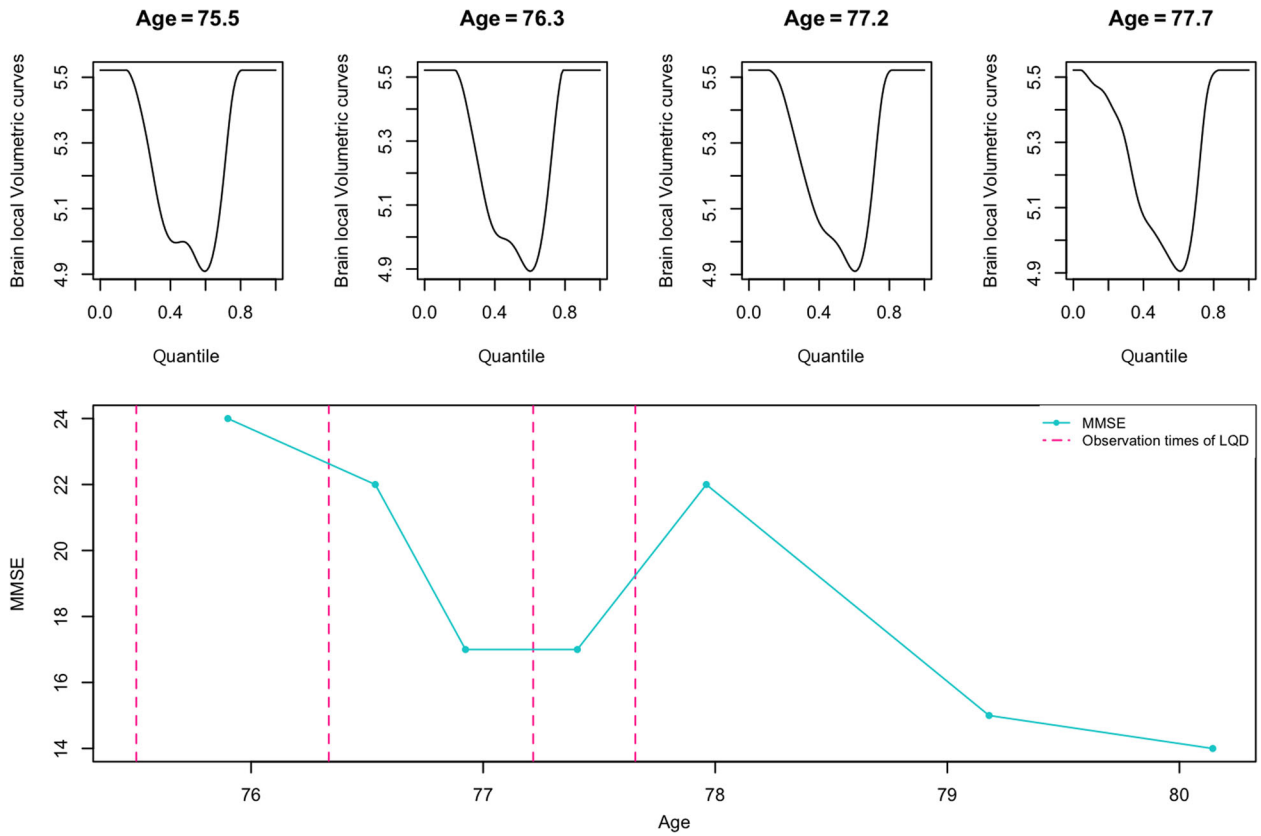

FIGURE 1 Brain local volumetric curves along left inferior temporal and the MMSE scores at different time points for one subject over time. The dotted vertical lines correspond to the observation times of Brain local volumetric curves. The solid line corresponds to the MMSE scores observed over time.

As an illustration, Figure 1 presents brain local volumetric curves along left inferior temporal and the MMSE scores at different time points for one randomly selected subject across time. The local volumetric curves and the MMSE scores were observed at mismatched time points within and across subjects, leading to asynchronous longitudinal functional and scalar data. We are interested in identifying risk factors (e.g., brain local volumetric curves) for the cognitive function across time (Burhanullah et al., 2020; Imtiaz et al., 2014; Reitz & Mayeux, 2014). However, the combination of the longitudinal functional variable and asynchronous measurement time points calls for new methods for longitudinal functional regression models.

The aim of this paper is to develop estimation and testing methods for functional linear regression models based on sparsely asynchronous longitudinal functional and scalar data. We consider the longitudinal functional linear regression model

$$Y(t) = \alpha_0 + \int_0^1 X(t,u)\beta_0(u)du + \epsilon(t), \qquad (1)$$

where $t \in [0,\tau]$ denotes the observation time, α_0 is the common intercept, and $\beta_0(u)$ is the functional coefficient as a function of $u \in \mathbb{U} = [0,1]$. Without loss of generality, we assume that $\alpha_0 = 0$. Given an observation time t, $X(t,u)$ is a square integrable random function recorded on the interval $\mathbb{U} = [0,1]$, and $Y(t)$ is a scalar variable at time t. Similar to Yuan and Cai (2010), Shang and Cheng (2015), and Li and Zhu (2020), the functional parameter β_0 is assumed to reside in a reproducing kernel Hilbert space (RKHS). We fit model (1) for the asynchronous data set, in which the observation times for the response and those for the functional covariate are different. Specifically, for subject $i = 1,\dots,n$, we observe $\{Y_i(T_{ij}) : j = 1,\dots,L_i\}$ and $\{X_i(S_{ik},u) : k = 1,\dots,M_i\}$, where S_{ik} and T_{ij} are observation time points, L_i and M_i are the number of time points for the response and that for the functional covariate, respectively, and $X_i(S_{ik},u)$'s can be observed on a set of dense grids for a given S_{ik}. In practice, T_{ij}'s may be different from S_{ik}'s, posing challenges to estimation and inference for model (1).

Despite a rich literature on regression models for synchronous data (Diggle, 2002; Jiang & Wang, 2011; Wang et al., 2019), the study on asynchronous data is scarce. It is often the case that synchronous data may not be available and existing methods are not directly applicable. To make traditional models applicable to asynchronous

data, Xiong and Dubin (2010) proposed a binning method to obtain equally spaced bins of time. Şentürk et al. (2013) adopted functional data analysis to pool information from all subjects. Cao et al. (2015, 2016) proposed kernel-weighted estimating equations for generalized linear models with either time-invariant or time-dependent coefficients. Chen and Cao (2017) extended the results of Cao et al. (2015) to the partially linear models with nonlinear time trend effects. Sun et al. (2021) considered the situation where asynchronous longitudinal data were observed and the corresponding observation process for response and covariate may depend on the covariate. However, all the above methods have never considered longitudinal functional data.

Model (1) can be regarded as an extension of the well-known functional linear model (FLM). The literature on FLM is too vast to summarize here. One may refer to the well-known monographs by Ramsay and Silverman (2005) for an extensive overview of FLM. There have been several recent attempts to tackle with longitudinal functional data. Goldsmith et al. (2012) proposed a longitudinal functional regression outcome model, estimating parameters through functional principal component analysis (FPCA) and penalized splines. Gertheiss et al. (2013) extended the study of Goldsmith et al. (2012) by using a longitudinal FPCA. Although the previous studies provide a convenient vehicle to analyze synchronous longitudinal data with functional variables, they are not applicable to asynchronous longitudinal data. Moreover, these methods rely heavily on the success of the FPCA approach, and may not be appropriate if the functional parameter cannot be represented effectively by the leading principals of the functional covariates (Yuan & Cai, 2010). Recently, Li et al. (2022) proposed generalized functional partial-linear varying-coefficient (GFPV) models to handle temporally asynchronous functional and scalar variables through using B-spline basis functions and kernel weighting methods.

In this paper, we propose estimation and testing methods for the longitudinal regression model (1) under the asynchronous data setting by using the RKHS framework. The major contributions of this paper are summarized as follows. First, we construct a kernel-weighted loss function with a roughness penalty to obtain the functional estimator. The kernel-weighted loss function downweights observations that are distinct in time and makes use of all functional covariate observations for each response under the asynchronous data setting. The roughness penalty regularizes the model complexity in a continuous manner (Shang & Cheng, 2015), while avoiding the drawbacks of FPCA. Following the spirit of Theorem 1.3.1 of Wahba (1990), a representer theorem is derived. It demonstrates that the estimator can be found in a finite-dimensional subspace of the infinite-dimensional space, enabling easy and efficient computation. Second, we derive the rate of convergence and pointwise distribution of our estimator based on its Bahadur representation. The rate of convergence derived here is slower than the optimal nonparametric rate of convergence in Section 4.5 of Wahba (1990) due to the loss of efficiency caused by the asynchronous data setting. Third, a penalized likelihood ratio test is proposed to test the nullity of the functional coefficient. We show that the null limit distribution of the proposed test statistic is a normal distribution and can be approximated by a χ^2 distribution. Moreover, we derive the separation rate that the proposed test can detect under the alternative hypotheses. To ensure easy implementation, a bootstrap method is proposed to estimate the unknown parameters in the limit distribution. Consistency of the bootstrap procedure is also established. Finally, the proposed estimation and testing methods are applied to the ADNI data set in order to identify significant associations between the local volumetric curves in 21 ROIs and the MMSE cognitive score. Furthermore, the R package "AsynchronousFLR" for this paper along with its documentation is freely accessible from our lab's GitHub website.

The rest of the paper is organized as follows. Section 2 gives the representer theorem and an estimation procedure. Section 3 presents the theoretical properties of the obtained functional estimator. The proposed test statistic and its null limit distribution are investigated in Section 4. Section 5 demonstrates the practicality of the proposed method through finite-sample simulation studies. We apply our methods to a real data set obtained from the ADNI study in Section 6. All the proofs can be found in the Supporting Information.

2 | RKHS-BASED ESTIMATION

In this section, we present some background information regarding to RKHS and then propose our estimation method.

2.1 | RKHS

A symmetric bivariate function $D(u_1, u_2)$ defined on $\mathbb{U} \times \mathbb{U}$ is said to be nonnegative definite, denoted by $D \geq 0$, if for all $g \in \mathbb{N}$, $a_1, \ldots, a_g \in \mathbb{R}$ and $u_1, \ldots, u_g \in \mathbb{U}$, we have $\sum_{j,j'=1}^{g} a_{j'} a_j D(u_j, u_{j'}) \geq 0$.

Definition 1. A Hilbert space $\mathcal{H}$ of functions $f : \mathbb{U} \to \mathbb{R}$ is said to be an RKHS if the elements of RKHS of $\mathcal{H}$ are

functions defined on some set $\mathbb{U}$, and there is a bivariate function D on $\mathbb{U} \times \mathbb{U}$ having the following two properties:

(1) For all $u \in \mathbb{U}$, $D_u = D(u, \cdot) \in \mathcal{H}$.

(2) For all $u \in \mathbb{U}$ and $f \in \mathcal{H}$, $f(u) = \langle f, D_u \rangle_{\mathcal{H}}$,

where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the associated inner product of $\mathcal{H}$. In this case, D is a reproducing kernel of $\mathcal{H}$.

The kernel $D(u_1, u_2)$ defines an integral operator such that $Df(u_1) = \int_{\mathbb{U}} D(u_1, u_2) f(u_2) du_2$ for $f \in L^2(\mathbb{U})$. It follows from Mercer's theorem that $D(u_1, u_2) = \sum_{k=1}^{\infty} \mu_k \phi_k(u_1) \phi_k(u_2)$, where $\{(\mu_k, \phi_k)\}_{k=1}^{\infty}$ are eigenvalue–eigenfunction pairs with $\mu_1 \geq \mu_2 \geq \ldots \geq 0$. Furthermore, we include Theorem 2.5 of Gu (2013) here for better understanding the decomposition of RKHS.

Proposition 1. *If the reproducing kernel D of a space $\mathcal{H}$ on domain $\mathbb{U}$ can be decomposed into $D = D_0 + D_1$, where both D_0 and D_1 are nonnegative definite, $D_0(u, \cdot), D_1(u, \cdot) \in \mathcal{H}, \forall u \in \mathbb{U}$ and $\langle D_0(u_1, \cdot), D_1(u_2, \cdot) \rangle_{\mathcal{H}} = 0, \forall u_1, u_2 \in \mathcal{H}$, then the spaces $\mathcal{H}_0$ and $\mathcal{H}_1$ corresponding to D_0 and D_1, respectively, form a tensor sum decomposition of $\mathcal{H}$. Conversely, if both D_0 and D_1 are nonnegative definite and $\mathcal{H}_0 \cap \mathcal{H}_1 = \{0\}$, then $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$ has a reproducing kernel $D = D_0 + D_1$.*

2.2 | Estimation

With the preparations given above, we present our estimation method for the unknown functional coefficient in model (1). Assume that $\beta(\cdot)$ resides in a Sobolev space of order m defined as

$$H^m(\mathbb{U}) = \{\beta : \mathbb{U} \mapsto \mathbb{R} | \ \beta^{(j)} \text{ is absolutely continuous for } j = 0, \ldots, m-1 \text{ and } \beta^{(m)} \in L^2(\mathbb{U})\}, \quad (2)$$

where $\beta^{(j)}$ denotes the jth order derivative of $\beta(u)$ with respect to u. We also assume $m > 1/2$ so that $H^m(\mathbb{U})$ is an RKHS.

Due to the infinite-dimensionality of β_0, we adopt the regularization method with roughness penalty to obtain its estimate. To handle mismatched observation times between the response and covariate within subjects, we introduce a kernel function into the loss function to borrow information from all possible pairs of responses and the covariate observations. This technique has been demonstrated to be an efficient way to manage asynchronous data with scalar covariates as shown in Cao et al. (2015) and Chen and Cao (2017). Hence, we consider a kernel-weighted penalized loss function $\ell_{n,\lambda}(\beta)$ and calculate an estimator of β, denoted as $\hat{\beta}_{n,\lambda}$, by maximizing

$$\ell_{n,\lambda}(\beta) = -\frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} K_b(t_{ij} - s_{ik}) \left\{ Y_i(t_{ij}) - \int X_i(s_{ik}, u) \beta(u) du \right\}^2 - (\lambda/2) J(\beta, \beta), \quad (3)$$

where $K_b(t) = K(t/b)/b$, $K(t)$ is a symmetric kernel function, b is the bandwidth for the kernel function, and $J(\beta_1, \beta_2) = \int_0^1 \beta_1^{(m)}(u) \beta_2^{(m)}(u) du$ for any $\beta_1, \beta_2 \in H^m(\mathbb{U})$ is the roughness penalty. Among various kernel functions, we use Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$ in this paper.

Following Lin and Carroll (2001), the observation times of $Y_i(t)$ and $X_i(s, \cdot)$ can be viewed from a bivariate counting process defined by $N_i(t, s) = \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} I(T_{ij} \leq t, S_{ik} \leq s)$ for $i = 1, \ldots, n$. The $N_i(t, s)$ counts the number of observation times up to t on the response and up to s on the covariates for subject i. We focus on the sparse asynchronous functional data such that both L_i and M_i are finite with probability 1. Although defined on the observation times, the counting process is a stochastic process due to the randomness of the observation times. It can be regarded as an extension of the representation of synchronous data in Martinussen and Scheike (2006). Since $dN_i(t, s) = 1$ when $t = T_{ij}$ and $s = S_{ik}$ and $dN_i(t, s) = 0$ otherwise, $\ell_{n,\lambda}(\beta)$ can be rewritten as

$$-\frac{1}{2n} \sum_{i=1}^{n} \int \int K_b(t - s) \left\{ Y_i(t) - \int X_i(s, u) \beta(u) du \right\}^2 dN_i(t, s) - (\lambda/2) J(\beta, \beta), \quad (4)$$

greatly facilitating investigating the theoretical properties of $\hat{\beta}_{n,\lambda}$.

We derive a representer theorem in Theorem 1 to show that the solution to (3) is in a finite-dimensional space even though β is in an infinite-dimensional space. The penalty functional $J(\beta, \beta)$ is a squared seminorm on $H^m(\mathbb{U})$ such that the null space $\mathcal{H}_0 = \{\beta : J(\beta, \beta) = 0\}$ is a finite-dimensional linear subspace of $H^m(\mathbb{U})$ with its orthonormal basis $\{\psi_1, \ldots, \psi_{N_0}\}$, where $N_0 = \dim(\mathcal{H}_0)$. Let $\mathcal{H}_1$ be the orthogonal complement of $\mathcal{H}_0$ in $H^m(\mathbb{U})$ such that $H^m(\mathbb{U}) = \mathcal{H}_0 \oplus \mathcal{H}_1$. It follows from Proposition 1 that for any function $\beta \in H^m(\mathbb{U})$, there exists a unique decomposition $\beta = \beta_0 + \beta_1$ such that $\beta_0 \in \mathcal{H}_0$ and $\beta_1 \in \mathcal{H}_1$. Such decomposition is popular in the smoothing spline literature (Gu, 2013; Wahba, 1990; Yuan & Cai, 2010). Let D_1 be the reproducing kernel of $\mathcal{H}_1$ such that $J(\beta, \beta) = \langle \beta, \beta \rangle_{D_1}$ for $\beta \in \mathcal{H}_1$. It follows from the properties of reproducing kernel that for any $u \in \mathbb{U}$ and $\beta_1 \in \mathcal{H}_1$, we have $D_1(u, \cdot) \in \mathcal{H}_1$ and $\langle \beta_1, D_1(u, \cdot) \rangle_{D_1} = \beta_1(u)$.

The dimension N_0 for the null space $\mathcal{H}_0$ is determined by the form of the penalty function $J(\beta,\beta)$. Specifically, when $J(\beta,\beta)=\int_0^1 \beta^{(m)}(u)\beta^{(m)}(u)du$ and $\beta\in H^m(\mathbb{U})$, $\mathcal{H}_0$ is an m-dimensional linear subspace of $H^m(\mathbb{U})$ with $\psi_1(u)=1$ and $\psi_r(u)=B_{r-1}(u)/(r-1)!$ for $r\in\{2,3,\dots,m\}$, where $B_{r-1}(u)$ is the $(r-1)$th Bernoulli polynomial. Moreover, the reproducing kernel D_1 of $\mathcal{H}_1$ has the form $D_1(u_1,u_2)=B_m(u_1)B_m(u_2)+(-1)^{m-1}B_{2m}(|u_1-u_2|)$. In this case, the dimension $N_0=m$. We refer to Chapter 1 of Wahba (1990) for more details about RKHS.

Theorem 1. *There exist* $\mathbf{d}=(d_1,\dots,d_{N_0})^\top$ *and* $\mathbf{c}_i=(c_{i1},\dots,c_{iM_i})^\top$ *such that the minimizer* $\hat\beta_{n,\lambda}$ *of (3) resides in the subspace of the functions of the form*

$$\hat\beta_{n,\lambda}(u)=\sum_{l=1}^{N_0}d_l\psi_l(u)+\sum_{i=1}^{n}\sum_{k=1}^{M_i}c_{ik}\int X_i(s_{ik},u_1)D_1(u_1,u)du_1$$
$$=\mathbf{d}^\top\boldsymbol{\psi}(u)+\sum_{i=1}^{n}\mathbf{c}_i^\top(D_1X_i)(u), \tag{5}$$

where $(D_1X_i)(u)=(\int D_1(u_1,u)X_i(s_{i1},u_1)du_1,\dots,\int D_1(u_1,u)X_i(s_{iM_i},u_1)du_1)^\top$ *and* $\boldsymbol{\psi}(u)=(\psi_1(u),\dots,\psi_{N_0}(u))^\top$.

Theorem 1 states that $\hat\beta_{n,\lambda}$ resides in a finite-dimensional subspace of $H^m(\mathbb{U})$. Thus, by making use of the information of observations at all the time points together, we generalize the representer theorem in Yuan and Cai (2010) to the functional linear regression model for asynchronous longitudinal data. It is noteworthy that this result reduces an infinite-dimensional optimization problem to a finite-dimensional one, which enables easy implementation. The details of how the representer theorem in Theorem 1 can be derived are included in Section S4 of the Supporting Information.

Denote $\mathbf{X}_i(u)=(X_i(s_{i1},u),\dots,X_i(s_{iM_i},u))^\top$, $\mathbf{X}(u)=(\mathbf{X}_1(u)^\top,\dots,\mathbf{X}_n(u)^\top)^\top$, $\mathbf{c}=(\mathbf{c}_1^\top,\dots,\mathbf{c}_n^\top)^\top$, and $\boldsymbol{\xi}_i(s_{ik})=\int X_i(s_{ik},u)\boldsymbol{\psi}(u)du$. By using the representation of β in (5), we can show that

$$\int X_i(s_{ik},u)\beta(u)du=\mathbf{d}^\top\int X_i(s_{ik},u)\boldsymbol{\psi}(u)du+\mathbf{c}^\top$$
$$\int X_i(s_{ik},u)(D_1\mathbf{X})(u)du \tag{6}$$

and $J(\beta)=\mathbf{c}^\top[\iint\mathbf{X}(u_1)D_1(u_1,u_2)\mathbf{X}(u_2)^\top du_1du_2]\mathbf{c}$. Then, the estimation of β becomes that of $\mathbf{d}$ and $\mathbf{c}$ through maximizing

$$-(2n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}K_b(t_{ij}-s_{ik})\{Y_i(t_{ij})-\mathbf{d}^\top\boldsymbol{\xi}_i(s_{ik})-\mathbf{c}^\top$$
$$\int X_i(s_{ik},u)(D_1\mathbf{X})(u)du\Big\}^2$$
$$-(\lambda/2)\mathbf{c}^\top\left[\int\int\mathbf{X}(u_1)D_1(u_1,u_2)\mathbf{X}(u_2)^\top du_1du_2\right]\mathbf{c}, \tag{7}$$

which is quadratic in $\mathbf{d}$ and $\mathbf{c}$. Thus, there exists a unique solution to the maximization problem (7) as follows. Denote $\theta=(\mathbf{d}^\top,\mathbf{c}^\top)^\top$ and $\tilde{\boldsymbol{\xi}}_i(s_{ik})=(\boldsymbol{\xi}_i(s_{ik})^\top,\int X_i(s_{ik},u)(D_1\mathbf{X})(u)du^\top)^\top$, W is a block-diagonal matrix with a zero matrix and $\int\int\mathbf{X}(u_1)D_1(u_1,u_2)\mathbf{X}(u_2)^\top du_1du_2$ lying along the diagonal. Therefore, (7) can be rewritten as

$$-(2n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}K_b(t_{ij}-s_{ik})\{Y_i(t_{ij})-\theta^\top\tilde{\boldsymbol{\xi}}_i(s_{ik})\}^2$$
$$-(\lambda/2)\theta^\top W\theta^\top, \tag{8}$$

leading to the solution $\hat\theta$ given by

$$\left\{\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}K_b(t_{ij}-s_{ik})\tilde{\boldsymbol{\xi}}_i(s_{ik})\tilde{\boldsymbol{\xi}}_i(s_{ik})^\top+n\lambda W\right\}^{-1}$$
$$\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}K_b(t_{ij}-s_{ik})Y_i(t_{ij})\tilde{\boldsymbol{\xi}}_i(s_{ik}). \tag{9}$$

Tuning parameter selection plays an important role in both estimation and testing. Following Li et al. (2020), we adopt K-fold cross-validation to simultaneously choose the tuning parameters including the penalty λ and the bandwidth b. The cross-validation criterion is the summation of the changes in the likelihood function after we leave onefold out, smaller changes are better.

3 | ASYMPTOTIC PROPERTIES

In this section, we introduce a new inner product in $H^m(\mathbb{U})$ and obtain the convergence rate and the Bahadur representation of the functional estimator in terms of the new inner product. We start with some notations. Let $a\asymp b$ if there exist positive constants $c_1,c_2>0$ such that $c_1\le a/b\le c_2$, and let $\|\cdot\|_{L^2}$ be the L^2 norm. For any $\beta_1,\beta_2\in H^m(\mathbb{U})$, we define

$$\langle\beta_1,\beta_2\rangle=V(\beta_1,\beta_2)+\lambda J(\beta_1,\beta_2)=\int_0^1\int_0^1C(u_1,u_2)$$
$$\beta_1(u_1)\beta_2(u_2)du_1du_2+\lambda J(\beta_1,\beta_2), \tag{10}$$

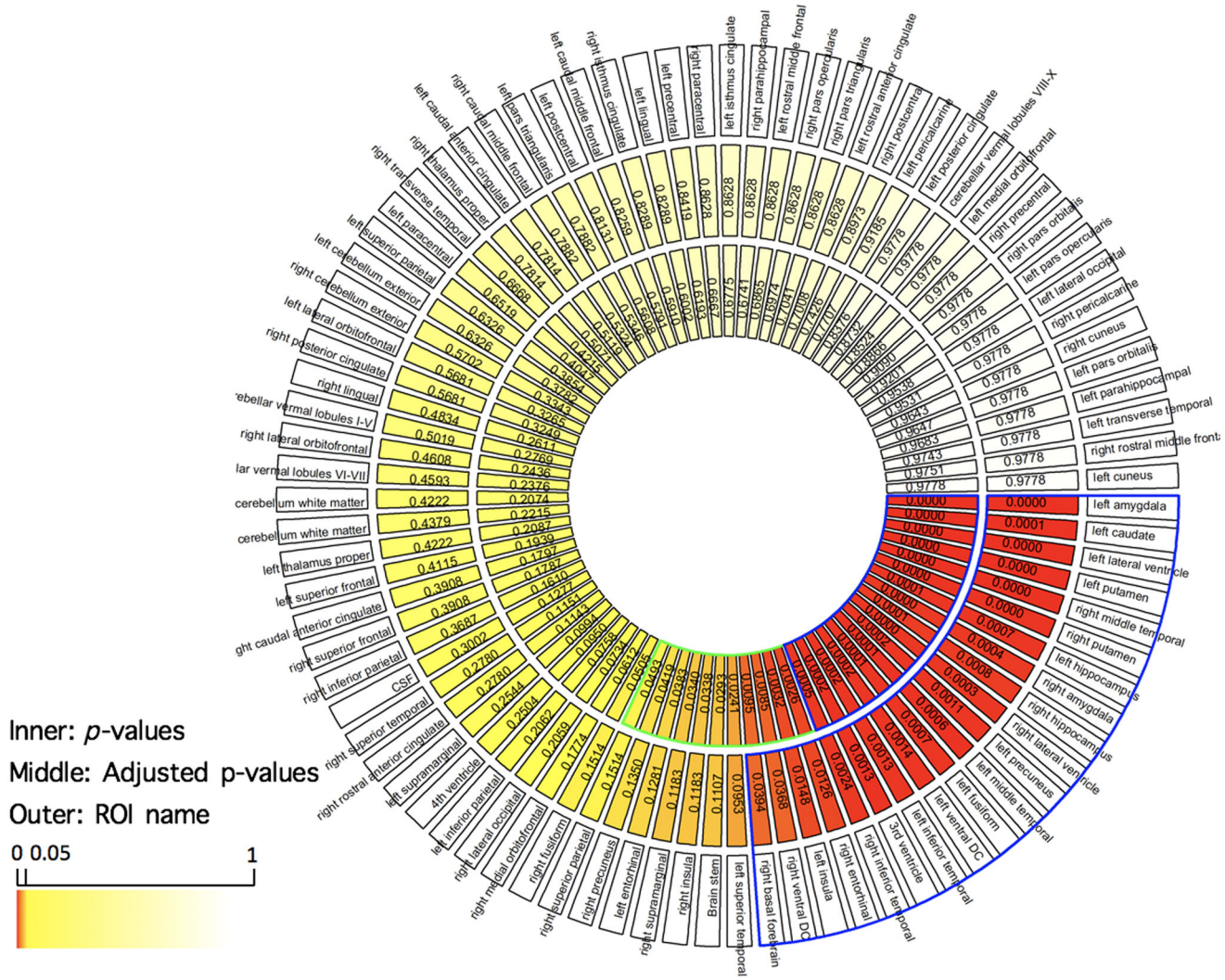

FIGURE 2 *p*-Values and adjusted *p*-values of the 87 ROIs in Scenario 2. ROIs in the inner blue sector are Bonferroni significant and ROIs in the outer blue sector are FDR significant at the 0.05 level. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

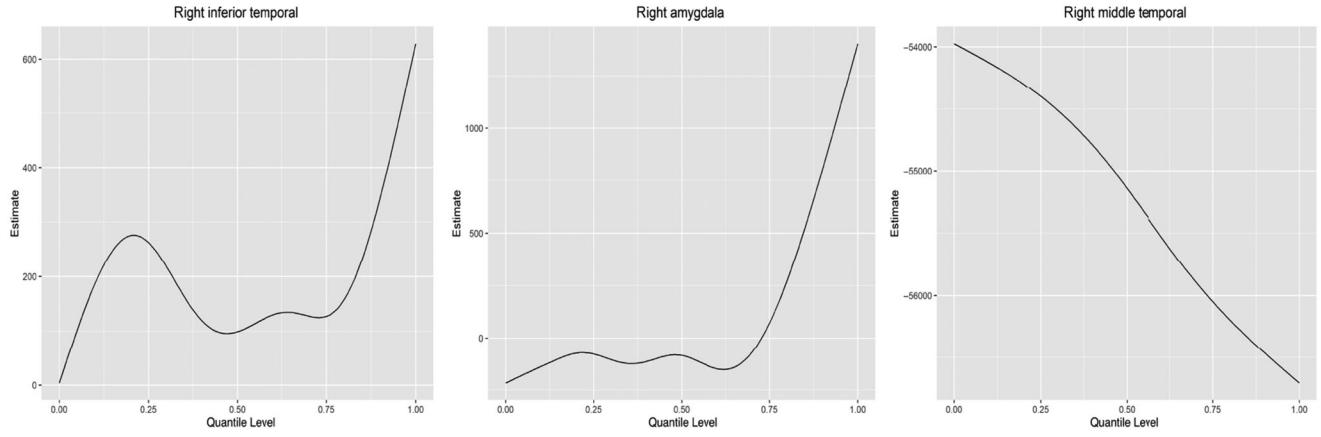

FIGURE 3 Functional estimates of the right inferior temporal, and right amygdala and middle temporal among the 21 FDR significant ROIs

where $C(u_1,u_2)=\int E\{X_i(s,u_1)X_i(s,u_2)\}\rho(s,s)ds$, in which $\rho(s,s)$ will be introduced in Assumption 1. Denote the induced norm by $\|\cdot\|$. The operator $C(u_1,u_2)$ can be viewed as a new type of covariance operator of X under the asynchronous data setting, since it differs from those operators proposed for functional data under cross-sectional setting (Li & Zhu, 2020; Shang & Cheng, 2015; Yuan & Cai, 2010), which take the form of $C_B(u_1,u_2)=E[B(X)X_i(u_1)X_i(u_2)]$, where $B(X)$ is a function of X. However, $C_B(u_1,u_2)$ is not directly applicable to longitudinal functional data.

In the following, we present some assumptions that help to derive the asymptotic results.

Assumption 1. The counting process $N_i(t,s)$ is independent of (X_i,ϵ_i) and $N_i(\tau,\tau)$ is bounded with probability 1. There exists a twice-continuous differentiable function $\rho(t,s)$ such that $E\{dN_i(t,s)\}=\rho(t,s)dtds$. The $\mathcal{G}=\{\rho(t,t)>0,t\in[0,1]\}$ has strictly positive Lebesgue measure. Also $P\{dN(t_1,t_2)=1|N(s_1,s_2)-N(s_1-,s_2-)=1\}=f(t_1,t_2,s_1,s_2)dt_1dt_2$ holds for any $t_1\neq s_1$ and $t_2\neq s_2$, where $f(t_1,t_2,s_1,s_2)$ is continuous. The left and right limits of $f(t_1,t_2,s_1,s_2)$ exist.

Assumption 2. The kernel function $K(\cdot)$ is a symmetric density function satisfying $\int z^2K(z)dz<\infty$, $\int K(z)^4dz<\infty$, $\int zK(z)^2dz<\infty$, and $\int zK(z)^4dz<\infty$.

Assumption 3. For $s,t\in[0,\tau]$, $E\{\epsilon(t)\}=0$, $var(\epsilon(t))=\sigma^2(t)$, and $cov(\epsilon(t),\epsilon(s))=r(s,t)$. Assume that $\sup_t\sigma^2(t)<\infty$ and $r(s,t)$ is twice continuously differentiable in $[0,\tau]^{\otimes2}$.

Following Cao et al. (2015) and Chen and Cao (2017), the counting process is required to be independent of the response and the functional covariate in Assumption 1. The assumption of bounded $N_i(t,s)$ is conventional in sparse longitudinal data (Diggle, 2002; Fan & Li, 2004; Lin et al., 2000; Sun et al., 2022). This differs from the dense setting where L_i and $M_i\to\infty$ for all i. We require that the intensity function $\rho(s,t)$ is positive when $s=t$ at some points, and $\rho(s,t)$ need not be greater than 0 when $s\neq t$. Similar assumptions have been widely used for synchronous data (Lin & Ying, 2001; Martinussen & Scheike, 2006; Yao et al., 2005). The $f(t_1,t_2,s_1,s_2)$ can be viewed as a marginal intensity function of (t_1,t_2). Regular conditions with respect to the kernel function are stated in Assumption 2. We also require the differentiability of the covariance of $\epsilon(t)$ in Assumption 3.

Assumption 4. The operator $C(u_1,u_2)$ is continuous in $[0,1]^{\otimes2}$, and for any $\beta\in L^2([0,1])$ satisfying $C\beta=0$, where $(C\beta)(u_1)=\int C(u_1,u_2)\beta(u_2)du_2$, we have $\beta=0$. Also, there exists $C_1>0$ such that for any $u_1,u_2\in\mathbb{U}$, we have $|\int E(X_i(s,u_1)X_i(s,u_2))\frac{\partial^2\rho(t,s)}{\partial t^2}|_{t=s}ds|<C_1$.

Assumption 5. There exists a sequence of functions $\{\varphi_\nu\}_{\nu\geq1}\subset H^m(\mathbb{U})$ such that $\|\varphi_\nu\|_{L^2}\leq C_\varphi\nu^a$ for each $\upsilon\geq1$, some constant $a\geq0$ and $C_\varphi>0$ and $V(\varphi_\nu,\varphi_\mu)=\delta_{\nu\mu}$ and $J(\varphi_\nu,\varphi_\mu)=\rho_\nu\delta_{\nu\mu}$ hold for any $\nu,\mu\geq1$, where $\delta_{\nu\mu}$ is the Kronecker delta and ρ_ν is a nondecreasing nonnegative sequence satisfying $\rho_\nu\asymp\nu^{2k}$ for $k>a+1/2$.

Assumption 6. For any φ_ν,φ_μ in Assumption 5, there exists some constant $C_2>0$ satisfying

$$\Big|\int\int\Big[\int E\{X_i(s,u_1)X_i(s,u_2)\}\frac{\partial^2\rho(t,s)}{\partial t^2}|_{t=s}ds\Big]\varphi_\nu(u_1)\varphi_\mu(u_2)du_1du_2\Big|<C_2. \quad (11)$$

Assumption 4 imposes conditions on the covariance operator. It enables (10) to be a well-defined inner product so that under $\langle\cdot,\cdot\rangle$, $H^m(\mathbb{U})$ is an RKHS. Assumption 5 indicates that any $\beta\in H^m(\mathbb{U})$ admits the Fourier expansion $\beta=\sum_{\nu=1}^\infty V(\beta,\varphi_\nu)\varphi_\nu$. This assumption is commonly made in smoothing spline literature and plays a critical role in controlling the local behaviors of the penalized estimate (Gu, 2013). In particular, the eigensystem can be obtained from the pseudo Sacks–Ylvisaker conditions in the Supplementary Material of Shang and Cheng (2015), leading to the explicit relations between m,a, and k such that $k=m+a$. Moreover, our inner product is different from those in Shang and Cheng (2015) and Li and Zhu (2020), leading to different eigenfunctions and eigenvalues. Assumption 6 is used to control the effect of the kernel function on the inner product (10). If the intensity function $\rho(t,s)$ is uniformly bounded, then (11) holds by the construction of the eigenfunctions.

Assumption 7. There exists a $w\in(0,1)$ such that $\sup_s E(\exp(w(\int X^2(s,u)du)^{1/2}))$ is bounded. Moreover, for $d\in\{4,6,8\}$ and any $\beta\in H^m(\mathbb{U})$, there exists a constant $M_d>0$ satisfying $\int E(\int X_i(s,u)\beta(u)du)^d\rho(s,s)ds\leq M_d[\int E(\int X_i(s,u)\beta(u)du)^2\rho(s,s)ds]^{d/2}$.

Assumption 8. The expectation $E[X(t,u_1)X(s,u_2)]$ is twice continuously differentiable in $[0,\tau]^{\otimes2}$ for all u_1,u_2, and there exist positive constants C_3 and C_4 such that

$$\Big|\int\frac{\partial E(X(t,u)X(s,u))}{\partial t}|_{t=s}\frac{\partial\rho(t,s)}{\partial t}|_{t=s}ds\Big|<C_3,\Big|$$
$$\int\frac{\partial^2E(X(t,u)X(s,u))}{\partial t^2}|_{t=s}\rho(s,s)ds\Big|<C_4. \quad (12)$$

Assumption 7 concerns the moments of a linear functional of X, which is similar to Assumption (b) in Yuan

and Cai (2010). Assumption 8 is similar to Condition 3 in Cao et al. (2015), which posits smoothness assumptions on $E[X(t,u_1)X(s,u_2)]$ and $\rho(t,s)$. Assumptions 7 and 8 hold for bounded intensity function and bounded covariance function. In clinical settings, it is common that each subject has a finite number of visits and many neuroimaging measures are uniformly bounded.

We derive the asymptotic convergence rate of the estimator $\hat{\beta}_{n,\lambda}$ under some mild conditions detailed above. Recall that b is the bandwidth for the kernel function and denote $h = \lambda^{1/(2k)}$, where k is defined in Assumption 5.

Theorem 2. *Suppose that Assumptions 1–8 are satisfied, as $n \to \infty$, if $h = o(1)$, $b = o(1)$, $b^4h^{-1} = o(1)$, and $n^{-1/2}b^{-1/2}h^{-(a+1)-(2k-2a-1)/4m}(\log n)(\log\log n)^{1/2} = o(1)$ hold, then we have*

$$\|\hat{\beta}_{n,\lambda} - \beta_0\| = O_p((nbh)^{-1/2} + h^k + b^2h^{-1/2}). \quad (13)$$

Theorem 2 show that the convergence rate in (13) is slower than the standard nonparametric rate $(nh)^{-1/2} + h^k$ for the smoothing spline in the literature (Gu, 2013). We achieve an optimal convergence estimation rate by minimizing $(nbh)^{-1/2} + h^k + b^2h^{-1/2}$, leading to $b = O(n^{-1/5})$ and $h = O(n^{-4/\{5(2k+1)\}})$. From Proposition 2.2 of Shang and Cheng (2015), which states that $k = m + a$ under some conditions, it follows that the optimal rate of convergence can be obtained for $m > 1/2$. With this choice of the bandwidth and the smoothing parameter, we achieve a rate of convergence $O(n^{-4k/\{5(2k+1)\}})$, which is slightly slower than the optimal nonparametric rate of convergence $O(n^{-k/(2k+1)})$ shown in Section 4.5 of Wahba (1990). The loss of efficiency is primarily caused by the asynchronous data setting.

We are able to establish the Bahadur representation for the functional estimator. We need to derive the Fréchet derivatives of the loss function. It follows from the definition of reproducing kernel that we have $D_u = \sum_{\nu\geq 1}\varphi_\nu(u)\varphi_\nu/(1+\lambda\rho_\nu)$. Define $\tau(x_s)$ satisfying $\langle\tau(x_s),\varphi_\nu\rangle = \int_0^1 x(s,u)\varphi_\nu(u)du \equiv x_{s\nu}$ for any $\nu \geq 1$. Thus, we have $\tau(x_s) = \sum_{\nu\geq 1} x_{s\nu}\varphi_\nu/(1+\lambda\rho_\nu)$. We also define two linear operators $R_x(s)$ for time point s and P_λ as $\langle R_x(s),\beta\rangle = \int_0^1 x(s,u)\beta(u)du$ for any time point s and $\langle P_\lambda\beta,\beta\rangle = \lambda J(\beta,\beta)$, respectively. Direct calculations give

$$R_x(s) = \tau(x_s) \quad \text{for any time pointsand } (P_\lambda\varphi_\nu)(\cdot) = \frac{\lambda\rho_\nu}{1+\lambda\rho_\nu}\varphi_\nu(\cdot). \quad (14)$$

Then, $\ell_{n,\lambda}(\beta)$ can be rewritten as

$$\ell_{n,\lambda}(\beta) = -\frac{1}{2n}\sum_{i=1}^n \int\int K_b(t-s)\{Y_i(t) - \langle R_{X_i}(s),\beta\rangle\}^2 dN_i(t,s) - \lambda/2\langle P_\lambda\beta,\beta\rangle. \quad (15)$$

The first-order Fréchet derivative of $\ell_{n,\lambda}(\beta)$ with respect to β is given by

$$S_{n,\lambda}(\beta) = \frac{1}{n}\sum_{i=1}^n \int\int K_b(t-s)\{Y_i(t) - \langle R_{X_i}(s),\beta\rangle\}R_{X_i}(s)dN_i(t,s) - P_\lambda\beta. \quad (16)$$

The following theorem establishes the Bahadur representation of the functional estimator.

Theorem 3. *If the conditions of Theorem 2 hold, then we have as $n \to \infty$,*

$$\|\hat{\beta}_{n,\lambda} - \beta_0 - S_{n,\lambda}(\beta_0)\| = O_p(a_n), \quad (17)$$

where $a_n = (n^{-1/2}b^{-1/2}h^{-(a+1)-(2k-2a-1)/4m}(\log n)(\log\log n)^{1/2} + b^2)\zeta_n$, in which $\zeta_n = ((nbh)^{-1/2} + h^k + h^{-1/2}b^2)$,

Theorem 3 establishes the Bahadur representation of $\hat{\beta}_{n,\lambda}$ for model (1) based on asynchronous longitudinal data. However, compared with the results in Shang and Cheng (2015), the mismatched time points between the response and the covariate and longitudinal functional data bring additional challenges to the theoretical investigation, leading to a different rate of convergence.

4 | HYPOTHESIS TESTING

In this section, we propose a test statistic to test the nullity of the functional parameter. Consider the null and alternative hypotheses given by $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$. Without loss of generality, we set $\beta_0 = 0$. We propose a penalized likelihood ratio test statistic as follows:

$$T_{n,b,\lambda} = -2(nb)\{\ell_{n,\lambda}(\beta_0) - \ell_{n,\lambda}(\hat{\beta}_{n,\lambda})\}. \quad (18)$$

Let $\tilde{\sigma}_l = h\sum_\nu[\int K(z)^2dz \int \sigma^2(s)E(\int X_i(s,u)\varphi_\nu(u))^2\rho(s,s)ds]^l/(1+\lambda\rho_\nu)^l$. The next theorem states the null limit distribution of $T_{n,b,\lambda}$.

Theorem 4. *Suppose the conditions of Theorem 3 and the null hypothesis H_0 hold, $(nb)a_n\zeta_n = o(h^{-1/2})$, $(nb)^{1/2}a_n = o(1)$, $(nb)h^{2k+1} = O(1)$, $bh^{-1} = o(1)$, and $n^{1/2}bh \to \infty$. Furthermore, if $\sup_t E(\epsilon^4(t)) < \infty$, then as $n \to \infty$, we have*

$$\frac{T_{n,b,\lambda} - h^{-1}\tilde{\sigma}_1 - (nb)\|P_\lambda\beta_0\|^2}{\sqrt{2\tilde{\sigma}_2h^{-1}}} \xrightarrow{d} N(0,1). \quad (19)$$

Denote $u_n = h^{-1}\tilde{\sigma}_1^2/\tilde{\sigma}_2$ and $\tilde{\sigma} = \tilde{\sigma}_1/\tilde{\sigma}_2$. It can be shown that $(nb)\tilde{\sigma}\|P_\lambda\beta_0\|^2 = o(u_n)$ holds and

$$(\tilde{\sigma}T_{n,b,\lambda} - u_n)/\sqrt{2u_n} \xrightarrow{d} N(0,1), \quad (20)$$

implying that $\tilde{\sigma}T_{n,b,\lambda}$ is approximately $\chi^2_{u_n}$.

To derive the separation rate of the proposed test, we consider a local alternative $H_1 : \beta = \beta_n$, where $\beta_n \in \Theta_l = \{\|\beta\|_{L^2} \le l, J(\beta,\beta) < l\}$ for some fixed constant $l > 0$.

Theorem 5. *If the conditions of Theorem 4 and $H_1 : \beta = \beta_n$ hold, then for any sequence $c_n \to \infty$, the power function of the proposed test is asymptotically one such that*

$$\inf_{\beta_n \in \Theta_l : \|\beta_n\| \ge c_n \eta_n} P_{\beta_n}\left(\frac{\tilde{\sigma}T_{n,b,\lambda} - u_n - (nb)\tilde{\sigma}\|P_\lambda \beta_0\|^2}{\sqrt{2u_n}} > z_\alpha \right) \to 1, \quad (21)$$

where $\eta_n \asymp \sqrt{(nbh^{1/2})^{-1} + h^{2k} + b^2}$ and z_α is the upper α quantile of $N(0,1)$.

From Theorem 5, it follows that the proposed test can detect any local alternative with the separation rate no faster than η_n. Specifically, the minimal separation rate, $n^{-2k/(6k+1)}$, is attained when $b = O(n^{-2k/(6k+1)})$ and $h = O(n^{-2/(6k+1)})$, which are different from the choice of (b,h) for the optimal convergence rate in Theorem 2.

To circumvent the difficulty of approximating the mean and variance in Theorem 4, we develop a random-weighting bootstrap procedure to approximate the asymptotic distribution of $T_{n,b,\lambda}$. According to (20), under the null hypothesis, $T_{n,b,\lambda}$ converges in distribution to a normal distribution with $E(T_{n,b,\lambda}) = u_n/\tilde{\sigma}$ and $\mathrm{var}(T_{n,b,\lambda}) = 2u_n/\tilde{\sigma}^2$, leading to $\tilde{\sigma} = 2E(T_{n,b,\lambda})/\mathrm{var}(T_{n,b,\lambda})$ and $u_n = 2E^2(T_{n,b,\lambda})/\mathrm{var}(T_{n,b,\lambda})$. Let $\ell^i_{n,\lambda}(\beta) = \sum_{j=1}^{L_i}\sum_{k=1}^{M_i} K_b(t_{ij} - s_{ik})\{Y_i(t_{ij}) - \int X_i(s_{ik},u)\beta(u)du\}^2 + \lambda J(\beta,\beta)$ for each i, we have to $\ell_{n,\lambda}(\beta) = -\sum_{i=1}^n \ell^i_{n,\lambda}(\beta)/(2n)$. Our bootstrap procedure consists of the following three steps:

- Step 1: Compute the estimator $\hat{\beta}_{n,\lambda}$ by maximizing (3) and then calculate the statistic $T_{n,b,\lambda}$.
- Step 2: For each bootstrapping replication $a = 1,\dots,A$, generate n i.i.d. random variables $\Pi^{(a)}_n = (\Pi^{(a)}_{n1},\dots,\Pi^{(a)}_{nn})$ with both mean and variance being 1 and $\max_i E(\Pi^{(a)}_{ni})^4 < \infty$. Then, we calculate $\hat{\beta}^{(a)} = \arg\sup_\beta \ell^{(a)}_{n,\lambda}(\beta)$ and compute $T^{(a)}_{n,b,\lambda} = -2(nb)[\ell^{(a)}_{n,\lambda}(\hat{\beta}_{n,\lambda}) - \ell^{(a)}_{n,\lambda}(\hat{\beta}^{(a)})]$, where $\ell^{(a)}_{n,\lambda}(\beta) = -\sum_{i=1}^n \Pi^{(a)}_{ni}\ell^i_{n,\lambda}(\beta)/(2n)$, in which the tuning parameters are the same as those selected in Step 1.
- Step 3: Calculate the sample mean and variance of $T^{(a)}_{n,b,\lambda}$ by $E = A^{-1}\sum_{a=1}^A T^{(a)}_{n,b,\lambda}$ and $V = A^{-1}\sum_{a=1}^A (T^{(a)}_{n,b,\lambda} - E)^2$ and estimate $\tilde{\sigma}$ and u_n by using $\hat{\tilde{\sigma}} = 2E/V$ and $\hat{u}_n = 2E^2/V$. Then, calculate the p-value of $T_{n,b,\lambda}$ according to the null limit distribution in (20) by replacing the unknown parameters with their estimates. If the p-value is smaller than a prespecified significance level, say 0.05, then one rejects the null hypothesis.

The above bootstrap procedure benefits from the time efficiency of sample averaging. Compared to the standard bootstrap method that uses the bootstrap samples, we only need to reestimate the unknown parameters. The following theorem shows that the bootstrap distribution, conditional on the observations, asymptotically imitates the null limit distribution.

Theorem 6. *Under the conditions of Theorem 4, and for $i = 1,\dots,n$ and $a = 1,\dots,A$, if the bootstrap weights satisfy $E(\Pi^{(a)}_{ni}) = 1$, $E(\Pi^{(a)}_{ni} - 1)^2 = 1$, and $\max_i E(\Pi^{(a)}_{ni})^4 < \infty$, then $T^{(a)}_{n,b,\lambda}$ converges weakly to the null limit distribution of $T_{n,b,\lambda}$ conditioning on the data.*

Theorem 6 validates the proposed bootstrap procedure. Because $T^{(a)}_{n,b,\lambda}$ and $T_{n,b,\lambda}$ share the same limit distribution under the null hypothesis, it is reasonable to estimate the unknown $\tilde{\sigma}$ and u_n by using the moments of $T^{(a)}_{n,b,\lambda}$.

5 | SIMULATION STUDIES

This section contains three simulation experiments for evaluating the finite-sample performance of the proposed testing procedure. Estimation and prediction errors are examined in Example 1. The size and power of the proposed test are approximated in Examples 2 and 3.

To obtain different observation times for the response and the covariate, the number of observation times of $Y(t)$ and $X(t,\cdot)$ comes from a Poisson distribution with the intensity rate 10. Once we have the two numbers of observation times, the observation times for the response and the covariates are independently generated from the uniform distribution $U(0,1)$.

Example 1. The $X_i(t,u)$ is of the form $X_i(t,u) = \sum_{k=1}^{100}\zeta_k M_k(t)\vartheta(u,k)$, where $\zeta_k = (-1)^{k+1}k^{-\omega/2}$ with $\omega \in \{1.1, 1.5, 2, 4\}$. The term ω controls the smoothness of the functional variables and they are expected to be less smooth for smaller ω. Moreover, $M_k(t)$ is 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} and t_{ik} are, respectively, the jth and kth observation times for the ith subject. Then, 100 observation times are generated from the uniform distribution in the u- direction throughout the simulation. Moreover, we set $\vartheta(u,k) = 1$ if $k = 1$ and $\vartheta(u,k) = \sqrt{2}\cos((k-1)\pi u)$ otherwise.

Instead of observing the functional process completely, we observe $W_i(t,u) = X_i(t,u) + e_i(t,u)$, where the measurement error $e_i(t,u)$ is independently generated from a normal distribution $N(0,\sigma_X^2)$. We set $\sigma_X^2 \in \{0, 0.25\}$. Similar to Yuan and Cai (2010), the functional coefficient $\beta_0(u)$ is chosen as $\beta_0(u) = 4\sum_{k=1}^{100}(-1)^{k+1}k^{-1}\vartheta(u,k)$. The error term $\epsilon_i(t)$ is from a Gaussian process with mean 0 and $\text{cov}\{\epsilon(s),\epsilon(t)\} = \sigma_Y^2 \cdot 4^{-|t-s|}$, in which we set $\sigma_Y^2 \in \{1, 2\}$. We use fivefold cross-validation to select the penalty parameter and the bandwidth. For the purpose of illustration, we take $J(\beta,\beta) = \int_0^1 \beta^{(2)}(u)\beta^{(2)}(u)du$ and hence $N_0 = 2$ for the proposed estimation method. All simulation results are based on 200 replicates with sample size $n \in \{100, 200\}$ by using R (version 3.6.0) on a Linux server (equipped with Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz, 125 GB RAM).

To the best of our knowledge, there is no existing work on the development of FLM for asynchronous data. We modify three ad hoc methods to ensure fair comparison. The first method, denoted as FPCA_s, is to find the pairs $(Y_i(t), X_i(s,u))$ that have the minimal distance between the time points of the response and those of the covariate, treat such pairs as "synchronous data," and apply the standard FLM based on FPCA (Hall & Horowitz, 2007). The number of FPCA components is selected such that the cumulative percentage of variance explained is above 95%. The second method, denoted as RKHS_s, is almost the same as the first one except that we consider FLM based on RKHS (Yuan & Cai, 2010). The third method, denoted as Raw, is to treat the data as "synchronous data" and apply the proposed method by taking the kernel function to be 1 for (t_{ij}, s_{ik}) pairs that have the minimal distance and 0 for other pairs.

To evaluate the finite-sample performance of the proposed estimation procedure, we compute the empirical mean squared error (MSE) and empirical relatively MSE (RMSE) of $\hat{\beta}_{n,\lambda}$. The prediction mean squared error (PMSE) is also reported based on 200 new test samples. Furthermore, we calculate the computation time (in seconds). Table 1 presents the estimation accuracy and prediction results and their computation time in Example 1. Our method outperforms the other three methods in terms of MSEs and RMSEs of β and PMSE. As expected, MSEs and PMSEs decrease as the sample size increases. For our method, it costs more time to match all covariates with response and search grid points of the tuning parameters. Furthermore, for our method and RKHS_s, one can find that at the same sample size, the estimation error tends to be smaller for smaller ω, whereas the prediction error tends to be smaller for larger ω. For the raw method, the estimation error also becomes smaller and the prediction error becomes larger as ω decreases. Similar phenomenon has also been reported in Yuan and Cai (2010). In contrast, it is the opposite for FPCA_s since it becomes harder to accurately estimate the functional scores when X becomes less smooth.

Example 2. Samples are generated in the same way as that in *Example* 1 except that we set $\beta_0(u) = B\sum_{k=1}^{100}(-1)^{k+1}k^{-1}\vartheta(u,k)$, where $B \in \{0, 0.1, 0.25, 0.5\}$ and ζ_k's are normalized such that $\zeta_k = \zeta_k/\sqrt{\sum_k^{100}\zeta_k^2}$. We choose $n \in \{100, 200\}$ and the significance level to be 5%.

For the null hypothesis, $H_0 : \beta = 0$, the sizes and powers of the proposed test procedure are summarized based on 1000 simulation runs with 500 bootstrap samples in Table 2. Inspecting Table 2 reveals that the empirical sizes are reasonably controlled around the nominal level, and the empirical power increases with the sample size n as well as the signal strength. Note that the decay rate of the eigenvalues for the covariance operator of the functional covariate is influenced by ν. We find that the proposed test becomes more powerful for larger νs'.

Example 3. The functional process is generated as $X_i(t,u) = \sum_{j=1}^{100}\sqrt{\lambda_j}M_j(t)V(u,j)$, where $M_j(t)$ is generated in the same way as that in *Example* 1, $V(u,j) = \sqrt{2}\sin((j-0.5)\pi u)$, and $\lambda_j = (j-0.5)^{-2}\pi^{-2}$. The true functional parameter is chosen in the same way as that in Hilgert et al. (2013) such that $\beta_0^{B,\xi}(t) = B\sum_{j=1}^{100} j^{-\xi-0.5}V_j(t)/\sqrt{\sum_{k=1}^{\infty}k^{-2\xi-1}}$, where B controls the signal strength and ξ reflects the smoothness of β_0. We choose $B \in \{0, 0.1, 0.5, 1\}$ and $\xi \in \{0.1, 0.5, 1\}$. The percentage of rejecting $H_0 : \beta = 0$ is based on 1000 simulation runs with 500 bootstrap samples for $n \in \{100, 400\}$. Table 3 summaries empirical rejection rates under the settings of *Example 3*. The proposed test procedure controls the empirical sizes at the nominal level 5%, and the power approaches to 1 as the sample size, the signal strength, and the smoothness of the functional parameter increase.

6 | APPLICATION TO THE ADNI DATA

AD, the most prevalent form of dementia among elderly people, is an irreversible, progressive brain disorder that slowly destroys memory and thinking skills, and, eventually, the ability to carry out the simplest tasks. The causes of dementia can vary, depending on the types of brain changes that may be taking place. A characteristic macroscopic consequence of AD pathology is brain atrophy, but most of the existing studies focused on restricted

TABLE 1 Estimation and prediction results in Example 1. We include the empirical means of $\text{MSE}_\beta = E\{\int_0^2(\hat{\beta}(u)-\beta_0(u))^2du\}$, $\text{RMSE}_\beta = E\{\int_0^2(\hat{\beta}(u)-\beta_0(u))^2du/\int_0^2\beta_0(u)^2du\}$, and the prediction errors with standard errors in the parentheses.

			$\sigma_X^2 = 0$					$\sigma_X^2 = 0.25$			
n		ω		MSE_β	RMSE_β	PMSE	Time (s)	MSE_β	RMSE_β	PMSE	Time (s)
100	$\sigma_Y^2 = 1$	4	Proposed	0.480(0.238)	0.028(0.014)	1.030(0.081)	7.670(0.677)	0.522(0.367)	0.030(0.021)	1.050(0.091)	7.650(0.892)
			FPCA_s	0.561(0.298)	0.032(0.017)	1.020(0.078)	0.642(0.149)	0.637(0.364)	0.036(0.021)	1.030(0.086)	0.695(0.155)
			RKHS_s	0.494(0.286)	0.028(0.016)	1.020(0.078)	1.050(0.028)	0.541(0.343)	0.031(0.020)	1.030(0.086)	1.060(0.040)
			Raw	0.497(0.239)	0.028(0.014)	1.054(0.094)	0.822(0.076)	0.545(0.420)	0.031(0.024)	1.063(0.091)	0.809(0.0701)
		2	Proposed	0.284(0.136)	0.016(0.008)	1.050(0.092)	6.280(0.526)	0.304(0.174)	0.017(0.010)	1.050(0.087)	7.240(0.757)
			FPCA_s	0.859(0.543)	0.049(0.031)	1.080(0.089)	0.648(0.145)	0.968(0.728)	0.055(0.042)	1.070(0.095)	0.649(0.140)
			RKHS_s	0.352(0.102)	0.020(0.006)	1.060(0.089)	1.060(0.039)	0.345(0.073)	0.020(0.004)	1.050(0.082)	1.060(0.035)
			Raw	0.353(0.082)	0.020(0.005)	1.070(0.091)	0.829(0.066)	0.361(0.065)	0.021(0.004)	1.070(0.092)	0.815(0.067)
		1.5	Proposed	0.226(0.096)	0.013(0.006)	1.060(0.093)	6.100(0.468)	0.233(0.115)	0.013(0.007)	1.060(0.088)	7.670(0.845)
			FPCA_s	1.000(0.685)	0.057(0.039)	1.110(0.115)	0.632(0.166)	1.040(0.722)	0.059(0.041)	1.120(0.107)	0.643(0.159)
			RKHS_s	0.339(0.096)	0.019(0.006)	1.090(0.105)	1.050(0.033)	0.327(0.061)	0.019(0.004)	1.080(0.088)	1.060(0.045)
			Raw	0.330(0.050)	0.019(0.003)	1.090(0.094)	0.824(0.077)	0.341(0.053)	0.020(0.003)	1.100(0.092)	0.826(0.077)
		1.1	Proposed	0.179(0.073)	0.010(0.004)	1.060(0.093)	6.260(0.574)	0.183(0.082)	0.010(0.005)	1.070(0.089)	7.600(0.959)
			FPCA_s	0.903(0.468)	0.052(0.027)	1.160(0.121)	0.583(0.140)	1.010(0.652)	0.058(0.037)	1.180(0.150)	0.583(0.133)
			RKHS_s	0.325(0.072)	0.019(0.004)	1.110(0.102)	1.040(0.039)	0.326(0.057)	0.019(0.003)	1.100(0.092)	1.040(0.036)
			Raw	0.327(0.049)	0.018(0.003)	1.110(0.092)	0.823(0.073)	0.333(0.056)	0.019(0.003)	1.130(0.102)	0.814(0.082)
	$\sigma_Y^2 = 2$	4	Proposed	0.596(0.358)	0.034(0.021)	2.050(0.160)	7.530(0.850)	0.634(0.401)	0.036(0.023)	2.050(0.157)	7.742(0.843)
			FPCA_s	0.731(0.547)	0.042(0.031)	2.040(0.154)	0.650(0.152)	0.731(0.657)	0.042(0.038)	2.040(0.169)	1.650(0.491)
			RKHS_s	0.663(0.535)	0.038(0.031)	2.040(0.154)	1.050(0.026)	0.660(0.635)	0.038(0.036)	2.040(0.169)	1.350(0.154)
			Raw	0.605(0.382)	0.035(0.022)	2.070(0.169)	0.813(0.072)	0.695(0.631)	0.040(0.036)	2.082(0.174)	0.816(0.078)
		2	Proposed	0.360(0.133)	0.021(0.008)	2.080(0.163)	7.650(0.813)	0.377(0.142)	0.022(0.008)	2.070(0.160)	7.652(0.805)
			FPCA_s	1.490(0.991)	0.085(0.057)	2.140(0.175)	0.648(0.146)	1.740(1.380)	0.099(0.079)	2.130(0.169)	1.640(0.514)
			RKHS_s	0.409(0.180)	0.023(0.010)	2.090(0.171)	1.060(0.038)	0.407(0.138)	0.023(0.008)	2.070(0.150)	1.350(0.142)
			Raw	0.381(0.110)	0.022(0.006)	2.080(0.153)	0.817(0.067)	0.388(0.104)	0.022(0.006)	2.120(0.189)	0.817(0.077)
		1.5	Proposed	0.321(0.182)	0.018(0.010)	2.080(0.179)	7.500(0.836)	0.332(0.230)	0.019(0.013)	2.090(0.169)	7.701(0.815)
			FPCA_s	1.780(1.250)	0.102(0.072)	2.200(0.201)	0.619(0.168)	1.950(1.350)	0.111(0.077)	2.210(0.216)	1.590(0.547)
			RKHS_s	0.386(0.140)	0.022(0.008)	2.120(0.182)	1.050(0.032)	0.378(0.119)	0.022(0.007)	2.100(0.185)	1.360(0.165)
			Raw	0.372(0.079)	0.021(0.005)	2.110(0.185)	0.820(0.073)	0.363(0.078)	0.021(0.005)	2.100(0.149)	0.809(0.073)
		1.1	Proposed	0.250(0.137)	0.014(0.008)	2.090(0.179)	7.620(0.903)	0.254(0.162)	0.145(0.009)	2.096(0.171)	7.643(0.817)
			FPCA_s	1.540(0.849)	0.088(0.049)	2.260(0.212)	0.583(0.136)	1.610(1.010)	0.092(0.058)	2.270(0.222)	1.460(0.712)
			RKHS_s	0.353(0.09)	0.020(0.006)	2.120(0.173)	1.030(0.025)	0.362(0.088)	0.021(0.005)	2.120(0.164)	1.330(0.261)
			Raw	0.342(0.059)	0.020(0.003)	2.120(0.174)	0.811(0.073)	0.366(0.072)	0.021(0.004)	2.120(0.171)	0.820(0.072)

(Continues)

TABLE 1 (Continued)

n		ω		$\sigma_X^2 = 0$				$\sigma_X^2 = 0.25$			
				MSE_β	$RMSE_\beta$	PMSE	Time (s)	MSE_β	$RMSE_\beta$	PMSE	Time (s)
200	$\sigma_Y^2 = 1$	4	Proposed	0.381(0.142)	0.022(0.008)	1.020(0.082)	22.600(1.580)	0.384(0.252)	0.022(0.014)	1.020(0.077)	27.100(2.940)
			$FPCA_s$	0.469(0.174)	0.027(0.010)	1.010(0.070)	0.688(0.168)	0.488(0.175)	0.028(0.010)	1.010(0.079)	0.781(0.174)
			$RKHS_s$	0.406(0.155)	0.023(0.009)	1.010(0.069)	4.260(0.088)	0.412(0.153)	0.024(0.009)	1.010(0.079)	4.340(0.131)
			Raw	0.389(0.133)	0.022(0.008)	1.050(0.077)	4.510(0.305)	0.401(0.126)	0.023(0.007)	1.050(0.089)	4.470(0.276)
		2	Proposed	0.216(0.085)	0.012(0.005)	1.020(0.077)	27.000(2.590)	0.220(0.065)	0.013(0.004)	1.030(0.078)	27.900(3.950)
			$FPCA_s$	0.505(0.424)	0.029(0.024)	1.040(0.092)	0.684(0.159)	0.552(0.449)	0.032(0.026)	1.030(0.086)	0.754(0.180)
			$RKHS_s$	0.308(0.034)	0.018(0.002)	1.050(0.088)	4.270(0.093)	0.317(0.045)	0.018(0.003)	1.040(0.085)	4.400(0.187)
			Raw	0.333(0.042)	0.019(0.002)	1.060(0.091)	4.470(0.270)	0.336(0.044)	0.019(0.003)	1.070(0.083)	4.410(0.269)
		1.5	Proposed	0.180(0.059)	0.010(0.003)	1.030(0.078)	26.900(2.380)	0.180(0.054)	0.010(0.003)	1.040(0.079)	26.700(2.990)
			$FPCA_s$	0.523(0.285)	0.030(0.016)	1.050(0.084)	0.709(0.150)	0.587(0.405)	0.034(0.023)	1.050(0.090)	0.772(0.177)
			$RKHS_s$	0.302(0.030)	0.017(0.002)	1.060(0.084)	4.260(0.091)	0.305(0.035)	0.018(0.002)	1.060(0.088)	4.360(0.147)
			Raw	0.317(0.029)	0.018(0.002)	1.090(0.083)	4.450(0.281)	0.320(0.031)	0.018(0.002)	1.090(0.082)	4.410(0.295)
		1.1	Proposed	0.150(0.047)	0.009(0.003)	1.040(0.078)	26.800(2.190)	0.146(0.044)	0.008(0.003)	1.050(0.080)	27.800(3.230)
			$FPCA_s$	0.499(0.272)	0.029(0.016)	1.090(0.102)	0.716(0.167)	0.501(0.300)	0.029(0.017)	1.090(0.124)	0.731(0.170)
			$RKHS_s$	0.299(0.027)	0.017(0.002)	1.090(0.100)	4.270(0.093)	0.305(0.081)	0.017(0.004)	1.090(0.119)	4.320(0.102)
			Raw	0.315(0.030)	0.018(0.002)	1.12(0.0866)	4.490(0.284)	0.322(0.033)	0.018(0.002)	1.110(0.102)	4.450(0.275)
	$\sigma_Y^2 = 2$	4	Proposed	0.372(0.196)	0.021(0.011)	2.002(0.151)	26.800(3.500)	0.371(0.117)	0.021(0.006)	2.020(0.150)	26.005(2.582)
			$FPCA_s$	0.552(0.255)	0.032(0.015)	2.020(0.157)	1.780(0.564)	0.517(0.192)	0.030(0.011)	2.030(0.154)	1.110(0.780)
			$RKHS_s$	0.485(0.256)	0.028(0.015)	2.020(0.157)	5.770(0.577)	0.437(0.164)	0.025(0.010)	2.030(0.154)	4.750(0.916)
			Raw	0.441(0.169)	0.025(0.010)	2.050(0.160)	4.400(0.301)	0.462(0.179)	0.026(0.010)	2.050(0.169)	4.440(0.294)
		2	Proposed	0.296(0.155)	0.017(0.009)	2.030(0.152)	26.400(3.020)	0.279(0.121)	0.016(0.007)	2.040(0.152)	27.489(3.583)
			$FPCA_s$	0.974(0.761)	0.056(0.044)	2.070(0.171)	1.730(0.533)	0.946(0.715)	0.054(0.041)	2.060(0.171)	0.609(0.125)
			$RKHS_s$	0.331(0.063)	0.019(0.004)	2.040(0.170)	5.890(0.479)	0.336(0.066)	0.019(0.004)	2.040(0.168)	4.180(0.073)
			Raw	0.348(0.064)	0.020(0.004)	2.070(0.150)	4.420(0.283)	0.352(0.066)	0.020(0.004)	2.060(0.165)	4.410(0.288)
		1.5	Proposed	0.225(0.104)	0.013(0.006)	2.040(0.153)	27.100(3.930)	0.220(0.085)	0.013(0.005)	2.049(0.153)	27.412(3.847)
			$FPCA_s$	0.886(0.581)	0.051(0.033)	2.100(0.175)	1.400(0.700)	0.914(0.587)	0.052(0.034)	2.100(0.165)	1.060(0.525)
			$RKHS_s$	0.321(0.049)	0.018(0.003)	2.080(0.170)	5.400(0.870)	0.319(0.044)	0.018(0.003)	2.070(0.154)	4.840(0.647)
			Raw	0.335(0.049)	0.019(0.003)	2.080(0.159)	4.460(0.293)	0.341(0.052)	0.020(0.003)	2.090(0.163)	4.405(0.270)
		1.1	Proposed	0.182(0.077)	0.010(0.004)	2.050(0.153)	25.800(3.170)	0.175(0.068)	0.101(0.004)	2.056(0.153)	27.412(3.754)
			$FPCA_s$	0.781(0.410)	0.045(0.024)	2.140(0.211)	1.030(0.549)	0.772(0.480)	0.044(0.028)	2.140(0.174)	0.740(0.250)
			$RKHS_s$	0.321(0.089)	0.018(0.005)	2.090(0.201)	4.820(0.768)	0.313(0.040)	0.018(0.002)	2.100(0.155)	4.390(0.376)
			Raw	0.325(0.046)	0.019(0.003)	2.120(0.174)	4.430(0.268)	0.329(0.049)	0.019(0.003)	2.130(0.165)	4.420(0.288)

TABLE 2 Sizes and powers (multiplied by 100) when testing $H_0 : \beta_0 = 0$ in Example 2

		$\sigma_X^2 = 0$					$\sigma_X^2 = 0.25$			
n		ω	$B = 0$	$B = 0.1$	$B = 0.25$	$B = 0.5$	$B = 0$	$B = 0.1$	$B = 0.25$	$B = 0.5$
100	$\sigma_Y^2 = 1$	1.1	5.2	7.2	17.6	57.0	5.6	8.0	18.2	54.0
		1.5	5.4	8.6	27.4	78.8	6.2	8.4	28.4	79.6
		2	5.6	9.0	40.0	91.8	6.2	10.4	40.2	91.2
		4	6.0	13.2	55.8	97.4	5.4	13.0	53.2	97.8
	$\sigma_Y^2 = 2$	1.1	5.6	6.8	9.6	30.2	6.2	6.6	12.2	31.0
		1.5	5.4	7.0	14.4	49.8	5.6	7.4	16.8	49.0
		2	5.6	6.8	19.6	67.2	5.4	7.8	21.4	65.8
		4	5.8	7.2	29.4	83.2	4.8	9.0	32.0	84.8
200	$\sigma_Y^2 = 1$	1.1	5.0	7.4	30.2	88.2	5.2	7.6	28.0	84.8
		1.5	4.6	10.2	50.6	98.4	5.2	11.4	48.8	96.4
		2	4.8	14.4	69.2	99.6	5.4	13.6	67.8	99.2
		4	5.4	19.4	86.4	100.0	6.4	18.2	84.2	100
	$\sigma_Y^2 = 2$	1.1	4.6	5.8	16.6	56.0	5.4	6.0	15.6	53.2
		1.5	4.6	7.0	27.0	81.2	5.2	6.8	25.4	81.2
		2	4.4	8.8	37.6	94.6	5.2	9.0	35.4	92.6
		4	4.4	10.8	56.6	99.0	6.2	11.4	54.8	98.6

TABLE 3 Sizes and powers (multiplied by 100) when testing $H_0 : \beta_0 = 0$ in Example 3

		$\sigma_X^2 = 0$					$\sigma_X^2 = 0.25$			
n		ξ	$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$
100	$\sigma_Y^2 = 1$	0.1	6.0	6.2	13.2	33.2	5.6	6.0	14.6	40.0
		0.5	6.0	6.4	29.8	83.0	5.6	7.2	35.6	81.4
		1	6.0	6.6	38.6	92.2	5.0	7.6	43.0	90.8
	$\sigma_Y^2 = 2$	0.1	5.4	6.6	7.0	16.4	5.4	5.0	8.4	20.4
		0.5	5.4	6.6	13.6	49.8	5.4	5.6	17.4	55.6
		1	5.4	8.0	21.6	65.6	5.4	5.8	22.6	67.8
400	$\sigma_Y^2 = 1$	0.1	5.6	5.6	31.80	86.0	4.8	6.6	37.0	92.4
		0.5	5.6	8.0	81.20	100.0	4.8	9.4	87.2	99.8
		1	5.6	8.8	91.20	100.0	4.8	12.0	94.8	100
	$\sigma_Y^2 = 2$	0.1	5.0	5.0	15.6	56.0	4.6	6.0	22.8	63.4
		0.5	5.0	5.2	50.0	98.6	4.6	8.4	57.8	99.4
		1	5.0	5.8	64.4	100.0	4.6	9.2	71.4	100

regions (Fjell et al., 2009). Hence, it is of great interest (i) to investigate the relationship between brain volumes of different brain regions and cognitive performance and (ii) to examine whether brain volumes in some specific ROIs have a significant impact on cognitive performance.

We apply model (1) to the MRI data set collected by the ADNI study. The MRI data are preprocessed using standard procedures via advanced normalization tools (Avants et al., 2011). The procedure consists of the N4 bias correction, registration-based brain extraction, and a prior-based N4-Atropos 6 tissue segmentation (oasis template). By performing multiatlas cortical parcellation, we obtain brain volumes of 101 ROIs defined by the manually edited labels of the publicly available MindBoggle-101 data set (Klein & Tourville, 2012). We exclude subjects for whom the imaging data do not pass the standard imaging quality controls, and remove 14 ROIs with many missing values. The log-Jacobian transformation map for each subject at the standard space is parceled into 87 ROIs. Finally, we obtain the brain volume density values within the 87 ROIs, which provide an intuitive and comprehensive representation of brain volume values for each region and contain potentially valuable information such as the mode, spread, and shape of these densities. To deal with the restrictions that density functions do not live in a linear space, and which are not convenient for functional regression

modeling, we adopt the log quantile density transformation proposed in Petersen and Müller (2016) and take the log quantile density function of brain volume density curves (abbreviated as brain local volumetric curves) along the 87 brain regions as functional variables. Specifically, the density f can equivalently be represented as the cumulative distribution function F and quantile functions $Q = F^{-1}$ with support $[0, 1]$. The log quantile density transformation is given by $\log(\frac{dQ(t)}{dt}) = \log(\frac{dF^{-1}(t)}{dt}) = \log(f^{-1}(Q(t))), t \in [0, 1]$. MMSE scores are treated as the response because it has been widely used to assess cognitive mental status with lower scores indicating impairment.

The aim of this data analysis is to establish the associations between the MMSE scores and brain local volumetric curves across the 87 brain regions measured at different ages, while investigating the statistical significance of these associations and identifying the significant brain regions. We consider two scenarios. In the first scenario (Scenario 1), we utilize model (1) to examine the significance of the functional coefficient for each brain region. In the second scenario (Scenario 2), we consider the same functional covariate in each model, while adding the clinical information, such as diagnostic status at baseline, age, education level, and the number of APOE4 gene copies. The first scenario plays the role of the most parsimonious approach, whereas the second one excludes the impacts of the clinical covariates on the response when examining the effects of the functional covariates.

Similar to Xie and Kang (2017), we study the effects of brain local volumetric curves region by region and utilize a multiple testing procedure to identify the significant brain regions. In Scenario 1, we explore the significance of the functional coefficient using the proposed test for each brain region and calculate its associated p-value based on 5000 bootstrap samples. In Scenario 2, we add all clinical variables of interest into model (1) and then explore the significance of the functional coefficient. The average computation time of the proposed method applied to the 87 ROIs is 19.62 min with the standard error 0.13 min. Hence, we obtain 87 p-values of testing the nullity of the functional coefficients across the 87 ROIs in each scenario. The Bonferroni correction method and the false discovery rate (FDR) method with the commonly used level 0.05 are adopted to identify important ROIs from the 87 ROIs tested. We provide the p-values and the corresponding adjusted p-values based on the FDR method of the 87 ROIs for Scenario 2 in Figure 2, and provide that for Scenario 1 in Figure S1 to save space. Inspecting Figures S1 and 2 reveals that the functional estimate in Scenario 1 has a stronger effect on the MMSE scores than that in Scenario 2, fewer functional covariates are detected to be significant in Scenario 2. Hereafter, we focus on the estimation results for Scenario 2.

There are 21 ROIs declared to be FDR significant and 17 ROIs declared to be Bonferroni significant after adjusting for the clinical covariates. Specifically, the top 10 ROIs with small p-values include left amygdala (9.81e−9), left lateral ventricle (4.32e−8), left putamen (5.76e−7), right putamen (7.73e−7), right middle temporal (2.09e−6), left caudate (4.66e−6), right lateral ventricle (2.65e−5), right amygdala (3.95e−5), left hippocampus (9.43e−5), and right hippocampus (1e−4). Preliminary queries indicate that the top 10 ROIs do indeed have clinical significance in cognitive functional and memory processes. For example, hippocampus plays a vital role in regulating learning, memory encoding, memory consolidation, and spatial navigation. Strong evidence of symmetry in the functional estimates of brain volumetric curves is observed in many brain regions. In Figure 2, many left/right pairs of regions are identified to be FDR significant, such as amygdala, lateral ventricle, putamen, hippocampus, middle temporal, ventral diencephalon (DC), and inferior temporal. It may suggest similar importance of these ROIs in both left and right brain hemispheres to the progression of Alzheimer disease. Besides the well-known crucial factors, such as hippocampus and amygdala (Du et al., 2001; Evans et al., 2010; Frisoni et al., 2010), the proposed model sheds new insight on AD study by successfully identifying brain regions, such as putamen and caudate, which receive less attention in the literature (de Jong et al., 2008).

Furthermore, we depict all the functional estimates of the 21 FDR significant ROIs in Figures S2–S4. In Figure 3, we give three examples of the functional estimates. Specifically, functional estimate of the right inferior temporal is positive, functional estimate of right amygdala has both positive and negative values, while functional estimate of right middle temporal is negative.

In summary, 21 ROIs are found to be FDR significant for the MMSE score with 11 ROIs in the left brain hemisphere and 10 ROIs in the right brain hemisphere. The estimated functional effects of the brain local volumetric curves of the 21 ROIs vary across the brain, and they may have different trends for symmetric brain regions in the left and right brain hemispheres. These findings may provide a new perspective for understanding the progression of AD and the 21 ROIs may be considered as risk factors for the onset of dementia.

7 | DISCUSSION

Many issues still merit further research. Taking multiple functional covariates into the model is an interesting and

important problem, which requires more comprehensive investigation. Another interesting problem is to consider functional quantile regression under the asynchronous data setting. We can also extend the results to a panel data setting with the large number of time points. We expect that the obtained rate of convergence would become faster for a panel data setting (Fan & Li, 2004; Hu et al., 2014). These extensions will be pursued in the future study.

ACKNOWLEDGMENTS

Data used in preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Ting Li's research was partially supported by National Science Foundation of China Grant 12101388 and the Program for Innovative Research Team of Shanghai University of Finance and Economics.

OPEN RESEARCH BADGES

This article has earned an Open Materials badge for making publicly available the components of the research methodology needed to reproduce the reported procedure and analysis. All materials are available at http://re3data.org/.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are openly available adni.loni.usc.edu.

ORCID

Huichen Zhu https://orcid.org/0000-0002-3530-1490
Hongtu Zhu https://orcid.org/0000-0002-6781-2690

REFERENCES

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A. & Gee, J.C. (2011) A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3), 2033–2044.

Burhanullah, M.H., Tschanz, J.T., Peters, M.E., Leoutsakos, J.M., Matyi, J., Lyketsos, C.G., et al. (2020) Neuropsychiatric symptoms as risk factors for cognitive decline in clinically normal older adults: the cache county study. *American Journal of Geriatric Psychiatry*, 28(1), 64–71.

Cao, H., Li, J. & Fine, J.P. (2016) On last observation carried forward and asynchronous longitudinal regression analysis. *Electronic Journal of Statistics*, 10(1), 1155–1180.

Cao, H., Zeng, D. & Fine, J.P. (2015) Regression analysis of sparse asynchronous longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4), 755–776.

Chen, L. & Cao, H. (2017) Analysis of asynchronous longitudinal data with partially linear models. *Electronic Journal of Statistics*, 11(1), 1549–1569.

de Jong, L.W., van der Hiele, K., Veer, I.M., Houwing, J., Westendorp, R., Bollen, E., et al. (2008) Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*, 131(12), 3277–3285.

Diggle, P. (2002) *Analysis of longitudinal data*. Oxford, England: Oxford University Press.

Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y., Jagust, W., et al. (2001) Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 71(4), 441–447.

Evans, M.C., Barnes, J., Nielsen, C., Kim, L.G., Clegg, S.L., Blair, M., et al. (2010) Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. *European Radiology*, 20(3), 674–682.

Fan, J. & Li, R. (2004) New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467), 710–723.

Fjell, A., Amlien, I., Westlye, L. & Walhovd, K. (2009) Mini-mental state examination is sensitive to brain atrophy in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 28(3), 252–258.

Frisoni, G.B., Fox, N.C., Jack, C.R., Philip, S. & Thompson, P.M. (2010) The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6(2), 67–77.

Gertheiss, J., Goldsmith, J., Crainiceanu, C. & Greven, S. (2013) Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, 14(3), 447–461.

Goldsmith, J., Crainiceanu, C.M., Caffo, B. & Reich, D. (2012) Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3), 453–469.

Gu, C. (2013) *Smoothing spline ANOVA models*. New York, NY: Springer Science & Business Media.

Hall, P. & Horowitz, J.L. (2007) Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35(1), 70–91.

Hilgert, N., Mas, A. & Verzelen, N. (2013) Minimax adaptive tests for the functional linear model. *Annals of Statistics*, 41(2), 838–869.

Hu, J., Liu, F. & You, J. (2014) Panel data partially linear model with fixed effects, spatial autoregressive error components and unspecified intertemporal correlation. *Journal of Multivariate Analysis*, 130, 64–89.

Imtiaz, B., Tolppanen, A.M., Kivipelto, M. & Soininen, H. (2014) Future directions in Alzheimer's disease from risk factors to prevention. *Biochemical Pharmacology*, 88(4), 661–670.

Jiang, C.R. & Wang, J.L. (2011) Functional single index models for longitudinal data. *Annals of Statistics*, 39(1), 362–388.

Klein, A. & Tourville, J. (2012) 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6(6), 171.

Li, T., Li, T., Zhu, Z. & Zhu, H. (2022) Regression analysis of asynchronous longitudinal functional and scalar data. *Journal of the American Statistical Association*, 117, 1228–1242.

Li, T. & Zhu, Z. (2020) Inference for generalized partial functional linear regression. *Statistica Sinica*, 30(3), 1379–1397.

Li, X., Li, Q., Zeng, D., Marder, K., Paulsen, J. & Wang, Y. (2020) Time-varying hazards model for incorporating irregularly measured, high-dimensional biomarkers. *Statistica Sinica*, 30(3), 1605–1632.

Lin, D. & Ying, Z. (2001) Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453), 103–126.

Lin, D.Y., Wei, L.J., Yang, I. & Ying, Z. (2000) Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 711–730.

Lin, X. & Carroll, R.J. (2001) Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96(455), 1045–1056.

Martinussen, T. & Scheike, T.H. (2006) *Dynamic regression models for survival data*. New York, NY: Springer.

Petersen, A. & Müller, H.G. (2016) Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, 44(1), 183–218.

Ramsay, J.O. & Silverman, B.W. (2005) *Functional data analysis*, 2nd edition. New York, NY: Springer.

Reitz, C. & Mayeux, R. (2014) Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical Pharmacology*, 88(4), 640–651.

Şentürk, D., Dalrymple, L.S., Mohammed, S.M., Kaysen, G.A. & Nguyen, D.V. (2013) Modeling time-varying effects with generalized and unsynchronized longitudinal data. *Statistics in Medicine*, 32(17), 2971–2987.

Shang, Z. & Cheng, G. (2015) Nonparametric inference in generalized functional linear models. *Annals of Statistics*, 43(4), 1742–1773.

Sun, D., Zhao, H. & Sun, J. (2021) Regression analysis of asynchronous longitudinal data with informative observation processes. *Computational Statistics & Data Analysis*, 157, 107161.

Sun, Z., Cao, H. & Chen, L. (2022) Regression analysis of additive hazards model with sparse longitudinal covariates. *Lifetime Data Analysis*, 28(2), 263–281.

Wahba, G. (1990) *Spline models for observational data*. Philadelphla, PA: Siam.

Wang, H.J., Feng, X. & Dong, C. (2019) Copula-based quantile regression for longitudinal data. *Statistica Sinica*, 29(1), 245–264.

Xie, J. & Kang, J. (2017) High-dimensional tests for functional networks of brain anatomic regions. *Journal of Multivariate Analysis*, 156, 70–88.

Xiong, X. & Dubin, J.A. (2010) A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statistics in Medicine*, 29(18), 1919–1931.

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

Yuan, M. & Cai, T.T. (2010) A reproducing kernel Hilbert space approach to functional linear regression. *Annals of Statistics*, 38(6), 3412–3444.

Zhao, B., Ibrahim, J.G., Li, Y., Li, T., Wang, Y., Shan, Y., et al. (2019) Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cerebral Cortex*, 29(7), 2904–2914.

SUPPORTING INFORMATION

Web Appendices, Figures referenced in Section 6 and an R package called "AsynchronousFLR" along with a real data demo code are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Li, T., Zhu, H., Li, T. & Zhu, H. (2023) Asynchronous functional linear regression models for longitudinal data in reproducing kernel Hilbert space. *Biometrics*, 79, 1880–1895. https://doi.org/10.1111/biom.13767