Published in final edited form as:

Stat Sin. 2018 October ; 28(4): 1867–1886. doi:10.5705/ss.202016.0350.

Functional Linear Regression Models for Nonignorable Missing Scalar Responses

Tengfei Li¹, Fengchang Xie², Xiangnan Feng³, Joseph G. Ibrahim⁴, and Hongtu Zhu^{1,4} for the Alzheimers Disease Neuroimaging Initiative

¹University of Texas MD Anderson Cancer Center

²Nanjing Normal University

³The Chinese University of Hong Kong

⁴University of North Carolina at Chapel Hill

Abstract

As an important part of modern health care, medical imaging data, which can be regarded as densely sampled functional data, have been widely used for diagnosis, screening, treatment, and prognosis, such as finding breast cancer through mammograms. The aim of this paper is to propose a functional linear regression model for using functional (or imaging) predictors to predict clinical outcomes (e.g., disease status), while addressing missing clinical outcomes. We introduce an exponential tilting semiparametric model to account for the nonignorable missing data mechanism. We develop a set of estimating equations and its associated computational methods for both parameter estimation and the selection of the tuning parameters. We also propose a bootstrap resampling procedure for carrying out statistical inference. Under some regularity conditions, we systematically establish the asymptotic properties (e.g., consistency and convergence rate) of the estimates calculated from the proposed estimating equations. Simulation studies and a real data analysis are used to illustrate the finite sample performance of the proposed methods.

Keywords

Estimating equation; exponential tilting; functional data; imaging data; nonignorable missing data; tuning parameters

1. Introduction

Medical imaging data, such as Magnetic Resonance Imaging (MRI), have been widely used to extract useful biomarkers that could potentially aid detection, diagnosis, assessment of prognosis, and prediction of response to treatment, among many others, since imaging data

Tengfei Li and Hongtu Zhu, Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston TX 77030, USA
Ibrahim, J. G. and Hongtu Zhu, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
Fengchang Xie, School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, CHINA
Xiangnan Feng, Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China

Supplementary materials available in the attached file include the proofs of Lemmas 1–13, Theorems 1–3, and Corollary 1.

may contain important information associated with the pathophysiology of various diseases, such as breast cancer. A critical clinical question is to translate medical images into clinically useful information that can facilitate better clinical decision making (Gillies et al., 2016). Addressing this clinical question requires the development of statistical models that use medical imaging data to predict clinical scalar responses. Standard functional linear model belongs to this type of statistical models (Ramsay, 2006). There is an extensive literature on the development of various estimation and prediction methods for functional linear models. See, for example, Cardot et al. (2003), Yao et al. (2005), Hall and Horowitz (2007), Crambes et al. (2009), Cai and Yuan (2012), Crambes and André (2013), and Hall and Giles (2015), among many others. The aim of this paper is to propose a new functional linear regression model to deal with an important scenario in clinical practice, when some clinical responses are missing.

Missing data is common in surveys, clinical trials, and longitudinal studies, and statistical methods for handling it often depend on the mechanism that generated the missing values. Three types of missing-data mechanism including missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), have been extensively studied in the literature (Baker and Laird, 1988; Ibrahim et al., 1999; Wang and Chen, 2009; Zhou et al., 2008; Kang and Schafer, 2007; Rotnitzky et al., 2012; Little and Rubin, 2002; Shi et al., 2009; Ibrahim et al. 2005; Ibrahim and Molenberghs, 2009). Among these mechanisms, MNAR is not only more technically challenging, but also more sensitive to model misspecification. Under MNAR, it is well known that common practices such as a complete case analysis or ad-hoc imputation of missing data can lead to seriously biased results in both estimation and prediction (Molenberghs and Kenward, 2007; Ibrahim and Molenberghs, 2009). To deal with MNAR, Kim and Yu (2011) developed a novel exponential tilting semiparametric model for the missing data mechanism and proposed some nonparametric regression techniques to estimate the conditional expectation. Furthermore, Tang et al. (2014) developed general estimating equations by using the empirical likelihood, whereas Zhao and Shao (2014) studied the identifiability issue for generalized linear models with nonignorable missing responses and covariates. However, these methods are limited to joint modeling of scalar predictors and scalar responses under MNAR.

Little has been done on joint modeling of functional predictors and missing scalar variables. Recently, Preda et al. (2010) defined the missingness of functional data and proposed a method based on nonlinear iterative partial least squares (NIPALS). Ferraty et al. (2013) studied mean estimation for the functional predictors under MAR. Chiou et al. (2014) proposed a missing value imputation and an outlier detection approach for traffic monitoring data. All these methods are limited to functional linear models for MAR and one-dimensional functional predictors.

The aim of this paper is to propose a new functional linear regression framework by integrating the exponential tilting model for MNAR and the standard functional linear model. We call it as ETFLR hereafter. We derive estimating equations (EEs) for ETFLR by combining the nonparametric kernel approach and the Functional Principle Component Analysis (FPCA) approach. We further derive an explicit formula for the computational

solution to EEs and a method for choosing the tuning parameters. Theoretically, we investigate the consistency and convergence rate of the proposed estimates under some regularity conditions. We also propose a bootstrap procedure for carrying out statistical inference. We use simulated and real data sets to demonstrate the advantage of the proposed approach over competing methods under MCAR and MAR. Finally, we will develop companion software for ETFLR and release it to the public through <http://www.nitrc.org/> and <http://odin.mdacc.tmc.edu/bigs2/>.

The rest of the paper is organized as follows. Section 2 introduces the model setting for ETFLR and presents the estimation procedure. Section 3 establishes asymptotic properties of the proposed parameter estimates. Section 4 includes simulation studies to examine the finite sample performance of the proposed estimates. In Section 5, we apply ETFLR to investigate the predictability of brain images at baseline on learning ability scores at 18 months after baseline obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data.

2. Functional Linear Regression for Missing Responses

2.1 Model Setup

Let $(\delta_i, Z_i, W_i, Y_i)$, $i = 1, 2, \dots, n$, be n independently and identically distributed realizations of the random vector (δ, Z, W, Y) , where δ is an indicator, Z is a functional predictor (e.g., MRI data) belonging to a specific functional space $\mathbb{H}$ endowed with an inner product $\langle \cdot, \cdot \rangle$, W is a $p \times 1$ random vector, and Y is a random scalar and subject to missingness. Define $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ if Y_i is missing for $i = 1, \dots, n$. It is assumed that δ_i and δ_j are independent for any $i \neq j$, and δ_i only depends on Z_i, W_i and Y_i . Denote $O_i = (\delta_i, \delta_i Y_i, Z_i, W_i)$ as the i -th observation. For notational simplicity, we focus on one-dimensional functional data throughout the paper. Without loss of generality, we assume

$$\mathbb{H} = \{f: [0, 1] \rightarrow \mathbb{R} \mid f \text{ is continuous and } \langle f, f \rangle \triangleq \int_0^1 f^2(t) dt < \infty\}.$$

For identification, it is assumed that the Z_i 's are centered such that $E(Z) = 0$ (Crambes and André, 2013).

Our ETFLR consists of a functional linear model and an exponential tilting semiparametric model for the propensity score as follows:

$$Y = \langle \theta, Z \rangle + \beta_1^T W + \varepsilon, \varepsilon \sim N(0, \sigma^2), \quad (2.1)$$

$$\text{logit}[\pi(Z, W, Y)] = G(Z, W) + \phi Y, \quad (2.2)$$

where $\pi(Z, W, Y) \triangleq \Pr(\delta = 1 \mid Z, W, Y)$ is called the propensity score, $\phi \in \mathbb{R}$ is an unknown parameter that determines the amount of departure from the ignorability of the response

mechanism, $\theta(\cdot) \in \mathbb{H}$ is an unknown functional coefficient function, and $\beta_1 \in \mathbb{R}^p$ is a $p \times 1$ vector of unknown coefficients. Moreover, $G \in \mathbb{G}$ is a nonparametric function, where $\mathbb{G} = \{\text{all continuous functions } \mathbb{H} \times \mathbb{R}^p \mapsto \mathbb{R}\}$. To include an intercept in (2.1), the first element of W_i is set to be 1.

The inclusion of $G(Z, W)$ in the propensity score represents a major extension of the so-called exponential tilting (ET) model proposed by Kim and Yu (2011). Such assumption is quite reasonable, since patients with severe or weak disease symptoms are more likely to be missing and imaging data may be strongly correlated with clinical symptoms. If the domain of Z is limited to a set of d grid points, then Z is reduced to a d -dimensional vector and the logarithm of the propensity score in (2.2) reduces to the ET model. Therefore, ETFLR is a generalization of ET from a vector space to a functional space. Similar to ET, the nonparametric form $G(Z, W)$ in the model is expected to be more robust to possible model misspecification compared with some parametric forms, such as $G(Z, W) = \langle g, Z \rangle + W^T \beta_2$ with a functional coefficient function $g(\cdot) \in \mathbb{H}$ and a vector $\beta_2 \in \mathbb{R}^p$.

For method development, we introduce some operators as follows:

$$\Gamma u = E(\langle Z, u \rangle Z), \quad \hat{\Gamma}_n u = \frac{1}{n} \sum_{i=1}^n \langle Z_i, u \rangle Z_i,$$

$$\Delta u = E[\langle Z, u \rangle (Y - \beta_1^T W)] \quad \text{and} \quad \hat{\Delta}_n u = \frac{1}{n} \sum_{i=1}^n \langle Z_i, u \rangle (Y_i - \beta_1^T W_i)$$

for any $u(\cdot) \in \mathbb{H}$. Moreover, due to the Hilbert-Schmidt theorem, it is commonly assumed that Γ and $\hat{\Gamma}_n$ have the sequences of eigenvalues $\{\lambda_j\}_{j=1}^\infty$ and $\{\hat{\lambda}_j\}_{j=1}^\infty$, with corresponding sequences of eigenfunctions $\{v_j(\cdot)\}_{j=1}^\infty$ and $\{\hat{v}_j(\cdot)\}_{j=1}^\infty$ respectively. Such a condition has been widely used in the FPCA literature.

2.2 Estimation Method

2.2.1 Estimating Equations—First, we consider the case when all responses are fully observed. In this case, parameter estimation is equivalent to solving a least squares (LS) problem. For ETFLR, we minimize the following objective function given by

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n (Y_i - \langle \theta, Z_i \rangle - \beta_1^T W_i)^2 \\ &= \langle \hat{\Gamma}_n \theta, \theta \rangle - 2 \hat{\Delta}_n \theta + \frac{1}{n} \sum_{i=1}^n [(\beta_1^T W_i)^2 - 2 Y_i \beta_1^T W_i] + \text{constant}. \end{aligned} \quad (2.3)$$

Using FPCA, we can estimate $\hat{\theta}$ by minimizing (2.3) with respect to θ over the linear span of $\{\hat{v}_1(\cdot), \dots, \hat{v}_{k_n}(\cdot)\}$, where k_n is a positive integer. Therefore, by setting $\theta = \sum_{j=1}^{k_n} r_j \hat{v}_j$, $\hat{\theta}$ can be solved by minimizing

$$\langle \hat{\Gamma}_n \theta, \theta \rangle - 2\hat{\Delta}_n \sum_{j=1}^{k_n} r_j \hat{v}_j + \frac{1}{n} \sum_{i=1}^n [(\beta_1^T W_i)^2 - 2Y_i \beta_1^T W_i]$$

with respect to $r = (r_1, \dots, r_{k_n})^T$. Furthermore, it follows from the Hilbert-Schmidt theorem that we have

$$\langle \hat{\Gamma}_n \theta, \theta \rangle \approx \sum_{j=1}^{k_n} \hat{\lambda}_j [\langle \hat{v}_j, \theta \rangle]^2 = \sum_j \hat{\lambda}_j r_j^2 \triangleq r^T \hat{\Lambda} r.$$

Finally, minimizing (2.3) is equivalent to solving the following estimating equation (EE) given by

$$\begin{cases} -\frac{1}{n} \sum_{i=1}^n \langle Z_i, \hat{v}_j \rangle (Y_i - \beta_1^T W_i) + \hat{\lambda}_j r_j = 0 & \text{for } j = 1, \dots, k_n \\ -\frac{1}{n} \sum_{i=1}^n (Y_i - \beta_1^T W_i) W_i + \frac{1}{n} \sum_{j=1}^{k_n} r_j \sum_{i=1}^n \langle Z_i, \hat{v}_j \rangle W_i = 0. \end{cases} \quad (2.4)$$

Second, we consider the case when some responses are missing not at random. We define $\gamma = -\phi$ and for $j = 1, \dots, k_n$,

$$\begin{aligned} \psi_{1,j}(Y_i, Z_i, W_i, v_j, \lambda_j; r, \beta_1) &= n^{-1} \sum_{i=1}^n \langle Z_i, v_j \rangle (Y_i - \beta_1^T W_i) - \lambda_j r_j, \\ \psi(Y_i, Z_i, W_i, \{\hat{v}_j\}_{j \leq k_n}, \{\hat{\lambda}_j\}_{j \leq k_n}; r, \beta_1) &= (\psi_1^T(\dots), \psi_2^T(\dots))^T, \\ \psi_1(\dots) &= (\psi_{1,1}(Y_i, Z_i, W_i, v_1, \lambda_1; r, \beta_1), \dots, \psi_{1,k_n}(Y_i, Z_i, W_i, v_{k_n}, \lambda_{k_n}; r, \beta_1))^T, \\ \psi_2(\dots) &= [Y_i - \beta_1^T W_i - \sum_{j=1}^{k_n} r_j \langle Z_i, v_j \rangle] W_i. \end{aligned}$$

Then, (2.4) is equivalent to

$$n^{-1} \sum_{i=1}^n \psi(Y_i, Z_i, W_i, \{\hat{v}_j\}_{j \leq k_n}, \{\hat{\lambda}_j\}_{j \leq k_n}; r, \beta_1) = 0. \quad (2.5)$$

The law of large numbers ensures that the expectation of the left side of (2.5) converges to zero as $k_n \rightarrow \infty$, but this EE depends on missing data. By following the reasoning in Tang, et.al. (2014), we have

$$\begin{aligned}
& E[\psi(Y_i, Z_i, W_i, \dots)] \\
&= \Pr(\delta_i = 1)E[\psi(Y_i, Z_i, W_i, \dots)|\delta_i = 1] + \Pr(\delta_i = 0)E\{\psi(Y_i, Z_i, W_i, \dots)|\delta_i = 0\} \\
&= E\{\delta_i \psi(Y_i, Z_i, W_i, \dots) + (1 - \delta_i)E[\psi(Y_i, Z_i, W_i, \dots)|\delta_i = 0, Z_i, W_i]\} \\
&= \frac{E[(1 - \delta_i)\psi(Y_i, Z_i, W_i, \dots)|Z_i, W_i]}{E[(1 - \delta_i)|Z_i, W_i]} \\
&= \frac{E[\Pr(\delta_i = 0|Y_i, Z_i, W_i)\psi(Y_i, Z_i, W_i, \dots)|Z_i, W_i]}{E[\Pr(\delta_i = 0|Y_i, Z_i, W_i)|Z_i, W_i]} \\
&= \frac{E[\exp(\gamma Y_i)\Pr(\delta_i = 1|Y_i, Z_i, W_i)\psi(Y_i, Z_i, W_i, \dots)|Z_i, W_i]}{E[\exp(\gamma Y_i)\Pr(\delta_i = 1|Y_i, Z_i, W_i)|Z_i, W_i]} \\
&= \frac{E\{\delta_i \psi(Y_i, Z_i, W_i, \dots)\exp(\gamma Y_i)|Z_i, W_i\}}{E\{\delta_i \exp(\gamma Y_i)|Z_i, W_i\}}.
\end{aligned}$$

Therefore, the original EE $\psi(Y_i, Z_i, W_i, \dots)$ shares the same expectation with

$$\delta_i \psi(Y_i, Z_i, W_i, \dots) + (1 - \delta_i) \frac{E\{\delta_i \psi(Y_i, Z_i, W_i, \dots)\exp(\gamma Y_i)|Z_i, W_i\}}{E\{\delta_i \exp(\gamma Y_i)|Z_i, W_i\}}.$$

Finally, we propose to solve the following equation:

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n [\delta_i \psi(Y_i, Z_i, W_i, \{\hat{v}_j\}_{j \leq k_n}, \{\hat{\lambda}_j\}_{j \leq k_n}; \mathbf{r}, \boldsymbol{\beta}_1) \\
& + (1 - \delta_i) m_{\psi, i, \gamma}^0(Y_i, Z_i, W_i, \{\hat{v}_j\}_{j \leq k_n}, \{\hat{\lambda}_j\}_{j \leq k_n}; \mathbf{r}, \boldsymbol{\beta}_1)] = 0,
\end{aligned} \quad (2.6)$$

where for any $f(\cdot)$, i and γ , $m_{f, i, \gamma}^0(\cdot)$ is defined by

$$m_{f, i, \gamma}^0(\cdot) = \frac{E\{\delta_i f(\cdot) \exp(\gamma Y_i) | Z_i, W_i\}}{E\{\delta_i \exp(\gamma Y_i) | Z_i, W_i\}}. \quad (2.7)$$

To calculate the estimating equation (2.6), we need to know both ϕ and k_n and then approximate the conditional expectations in $m_{\psi, i, \gamma}^0$. In the following, we will discuss how to calculate them. We introduce a kernel function $K(\cdot)$ and define $K_h(u) = K(u/h)$, where h is a bandwidth. We use

$$\hat{m}_{\psi, i, \gamma}(\dots) = \sum_{l=1}^n w_{l, 0}(Z_i, W_i; \gamma) \psi(Y_l, Z_l, W_l, \{\hat{v}_j\}_{j \leq k_n}, \{\hat{\lambda}_j\}_{j \leq k_n}; \mathbf{r}, \boldsymbol{\beta}_1)$$

as a nonparametric estimate of $m_{\psi, i, \gamma}^0(\dots)$, where

$$w_{l,0}(Z_i, W_i; \gamma) = \frac{\delta_l^{\exp(\gamma Y_l) K_h(D_l(Z_i, W_i))}}{\sum_{k=1}^n \delta_k^{\exp(\gamma Y_k) K_h(D_k(Z_i, W_i))}},$$

in which $D(Z, W)$ is equal to the sum of $w_0 \sqrt{\sum_{j=1}^n \langle Z - Z_l, \hat{v}_j \rangle^2}$ and $(1 - w_0) \|W - W_l\|$. The notation $\|\cdot\|$ is used to denote the l_2 norm of a vector v or the L_2 norm of a function $\nu(\cdot)$. Moreover, w_0 is a scalar introduced to balance the functional and nonfunctional parts of $D(\cdot, \cdot)$

2.2.2 Computational Method—We develop a computational method for our proposed estimating equation as follows. Let $D = \text{diag}\{\delta_1, \delta_2, \dots, \delta_n\}$, $\mathbf{W} = [W_1 \dots W_n]^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{1}_n$ be an $n \times 1$ vector of ones, and $\Xi = (w_{i,j})$ with $w_{i,j} = w_{i,0}(Z_j, W_j; \gamma)$. We then discretize the observed function Z_i to a fine grid of K equally spaced values t_k that span the interval $[0, 1]$ (Ramsay and Silverman 2006). Denote the K equally-spaced discrete points by $\mathbf{t} = (t_0 = 0, t_1, \dots, t_K = 1)$, and then we approximate the inner product $\langle Z, \boldsymbol{\theta} \rangle$ by $\sum_{k=1}^K \boldsymbol{\theta}(t_k) Z(t_k)(t_k - t_{k-1})$. We introduce an $n \times K$ matrix $\bar{Z} = (\bar{Z}_{i,k})$, a $K \times k_n$ matrix $\bar{V}_{k_n} = (\hat{V}_1, \dots, \hat{V}_{k_n})$ and $\boldsymbol{\theta} = \bar{V}_{k_n}^T \mathbf{r}$ such that $Z^* = \bar{Z} \bar{V}_{k_n}$ is an $n \times k_n$ matrix, where $\bar{Z}_{i,k} = Z_i(t_k)(t_k - t_{k-1})$. Then solving (2.6) is equivalent to minimizing

$$(\mathbf{Y} - Z^* \mathbf{r} - \mathbf{W} \boldsymbol{\beta})^T \sum (\mathbf{Y} - Z^* \mathbf{r} - \mathbf{W} \boldsymbol{\beta}) + (Z^* \mathbf{r})^T (\mathbf{I}_n - \Sigma) Z^* \mathbf{r}, \quad (2.8)$$

where $\Sigma = \{D + \text{diag}[\Xi(\mathbf{I}_n - D)\mathbf{1}_n]\}$. If ϕ and k_n , w_0 and h are given, the solution to (2.8) has an explicit form given by

$$\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = [(\bar{Z}^*)^T \sum (\bar{Z}^*) + (\tilde{Z}^*)^T (\mathbf{I}_n - \Sigma)(\tilde{Z}^*)]^{-1} (\bar{Z}^*)^T \sum \mathbf{Y}, \quad (2.9)$$

where $\bar{Z}^* = (\bar{Z}, \mathbf{W})$ and $\tilde{Z}^* = (\bar{Z}, 0)$.

2.2.3 Selection of Smoothing and Tilting Parameters—When ϕ is given, similar to Crambes and Henchiri (2015), the smoothing tuning parameters can be achieved by using the generalized cross-validation (GCV) criterion given by

$$\text{GCV}(k_n) = \frac{1}{n} \frac{\|\mathbf{Y} - \Sigma^* \mathbf{Y}\|^2}{[\text{trace}((\mathbf{I} - \Sigma^*) \circ D)/n]^2}, \quad (2.10)$$

where $\Sigma^* = \bar{Z}^*[(\bar{Z}^*)^T \Sigma(\bar{Z}^*) + (\tilde{Z}^*)^T (I_n - \Sigma)(\tilde{Z}^*)]^{-1} (\bar{Z}^*)^T \Sigma$ depends on k_n and ‘o’ denotes the element-wise product. To select h and w_0 , we generate L random divisions and denote $\mathcal{T}_\ell$ as the test set of the ℓ -th

random division for $\ell = 1, \dots, L$. See the detailed algorithm in Section 4. Then, we use Repeated Random Sub-sampling Validation (RRSV) by minimizing

$$\text{RRSV}(h, w_0) = \frac{1}{L} \sum_{\ell=1}^L \text{Loss}(\mathbf{Y}_{\mathcal{T}_\ell}, \hat{\mathbf{Y}}_{\mathcal{T}_\ell}), \quad (2.11)$$

where $\text{Loss}(\cdot, \cdot)$ is the negative Pearson correlation between the true responses $\mathbf{Y}_{\mathcal{T}_\ell}$ and the predicted responses $\hat{\mathbf{Y}}_{\mathcal{T}_\ell}$.

Following Kim and Yu (2011), we use an external validation sample, a follow-up subset of nonrespondents, chosen for further investigation to retrieve missing responses. We propose two approaches as follows. The first approach is the Missing Not At Random for Nonparametric (MNARN) method. In this approach, the validation sample is assumed to be randomly selected. Similar to Kim and Yu (2011), $\phi = -\gamma$ could be determined by the following estimating equation

$$\sum_{i=1}^n (1 - \delta_i) \delta_i^* [Y_i - m_{e,i,\gamma}^0(Y_i)] = 0, \quad (2.12)$$

where $m_{e,i,\gamma}^0(Y_i)$ is defined in (2.7) for the identity function $e(y) = y$ and $\delta_i^* = 1$ if the i th subject belongs to the follow-up sample and 0 otherwise. It is easy to show that the expectation of the left-hand side of (2.12) is equal to zero. Specifically, it follows from $m_{e,i,\gamma}^0(Y_i) = E(Y_i | \delta_i = 0, Z_i, W_i)$ that

$$\begin{aligned} E\{(1 - \delta_i) \delta_i^* [Y_i - m_{e,i,\gamma}^0(Y_i)]\} &= E \delta_i^* E(1 - \delta_i)(Y_i - m_{e,i,\gamma}^0(Y_i)) \\ &= E \delta_i^* E[(Y_i - E(Y_i | \delta_i = 0, Z_i, W_i)) | \delta_i = 0] = 0. \end{aligned}$$

Computationally, we approximate $m_{e,i,\gamma}^0(Y_i)$ by $\hat{m}_{e,i,\gamma}(Y_i) = \sum_{l=1}^n w_{l,0}(Z_i, W_i; \gamma) Y_l$

The second approach is the Missing Not At Random for Parametric (MNARP) method. In this approach, if G is specified to be a linear function of Z and W by $G(Z, W) = \langle Z, g \rangle + \beta_2^T W$ for $g(\cdot) \in \mathbb{H}$ and $\beta_2 \in \mathbb{R}^p$, then we estimate ϕ by maximizing the likelihood function of the logistic model (2.2) given by

$$\prod_{i=1}^n \left[\frac{\exp \left(\sum_{j=1}^{k_n^*} \langle Z_i, \hat{v}_j \rangle s_j + W_i^T \beta_2 + \phi y_i \right)}{1 + \exp \left(\sum_{j=1}^{k_n^*} \langle Z_i, \hat{v}_j \rangle s_j + W_i^T \beta_2 + \phi y_i \right)} \right]^{\delta_i} \times \left[\frac{\exp \left(\sum_{j=1}^{k_n^*} \langle Z_i, \hat{v}_j \rangle s_j + W_i^T \beta_2 + \phi y_i \right)}{1 + \exp \left(\sum_{j=1}^{k_n^*} \langle Z_i, \hat{v}_j \rangle s_j + W_i^T \beta_2 + \phi y_i \right)} \right]^{(1-\delta_i)\delta_i^*} \quad (2.13)$$

with respect to $(\phi, \beta_2, s_j; j = 1, \dots, k_n^*)$. The tuning parameter k_n^* denotes the number of eigenfunctions used for estimating ϕ . When the validation set is not large, we also add a penalty term, such as the LASSO, to the likelihood function. For a fixed k_n^* , this optimization procedure can be directly implemented by the ‘glmnet’ R package (Friedman et al., 2009). The optimal k_n^* can be further determined by minimizing its corresponding cross-validation error.

3. Theoretical Results

To facilitate the theoretical development, some assumptions are needed. Prior to presenting assumptions, we list some notation. The true values of β_1, β_2 , and $\theta(\cdot)$ are denoted by $\beta_{1,0}, \beta_{2,0}$, and $\theta_0(\cdot)$, respectively. Let $\mathbf{a}^{\otimes 2} = \mathbf{a}^T \mathbf{a}$ for any vector or matrix $\mathbf{a}$ and $\|\Sigma\| = \sqrt{\text{tr}(\Sigma^T \Sigma)}$ be the Frobenius norm of a matrix Σ . Let C be a generic constant.

We need the following assumptions:

$$(A.1) \theta(\cdot) \in L^2([0, 1]) = \{f: [0, 1] \rightarrow \mathbb{R} \mid \int_0^1 f^2(t) dt < \infty\}.$$

(A.2) The function $Z \in \mathbb{H}$, is centered: $E(Z) = 0$ and has the decomposition

$$Z(\cdot) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j v_j(\cdot),$$

where the ξ_j 's are independent real random variables with zero mean and unit variance. For all $j, l \in \mathbb{N}$, there exists a constant b such that $E(\xi_j \xi_l) = \delta_{jl}$ and $E(\xi_j^2) = 1$.

$$(A.3) \lambda_j \sim C j^{-a-1} \text{ for } j \rightarrow \infty \text{ and } a > 1.$$

$$(A.4) \langle \theta_0, v_j \rangle \sim C j^{-b} \text{ for } j \rightarrow \infty \text{ and } b > 1 + a/2.$$

(A.5) For any $C > 0$, there exists a $\tau_0 > 0$ such that

$$\sup_{s \in [0, 1]} \{E(Z(t))^C\} < \infty \text{ and } \sup_{t_1, t_2 \in [0, 1]} E|t_1 - t_2|^{-\tau_0} |Z(t_1) - Z(t_2)|^C < \infty.$$

(A.6) $\max\{E(\|W\|^2), \sigma^2\} < \infty$, and ε is independent of Z and W .

(A.7) $k_n \rightarrow \infty$ and $k_n^{5a+3} n^{-1} \rightarrow 0$ as $n \rightarrow \infty$.

(A.8) $\mathcal{J} \triangleq E(W^{\otimes 2}) - \sum_{j=1}^{\infty} E(W_{\xi_j}^{\xi_j}) E(W_{\xi_j}^T) > 0$.

(A.9) The true value $\phi = \phi_0$ is known.

Remark 1

Assumptions (A.1)–(A.9) have been widely used in the literature. Specifically, we can find assumptions similar to (A.1) and (A.2) in Crambles and André (2013), those similar to (A.3) and (A.4) in Hall and Horowitz (2007), those similar to (A.5) in Hall and Hosseini-Nasab (2006), those similar to (A.7) in Kong et al. (2015), and those similar to Condition (A.9) in Tang et al. (2014). Assumptions (A.6) and (A.8) are very weak since they require some mild conditions on $E(\|W\|^2)$ and $E[W - \sum_{j=1}^{\infty} E(W_{\xi_j}^{\xi_j})]^{\otimes 2} = EW W^T - \sum_{j=1}^{\infty} E(W_{\xi_j}^{\xi_j}) E(W_{\xi_j}^T)$. In Assumption (A.7), $k_n \rightarrow \infty$ and $k_n^{5a+3} n^{-1} \rightarrow 0$ ensure that both the bias and variance of $\hat{\theta}$ asymptotically converge to 0. For the selection of k_n , a small k_n may incur substantial information loss and cause bias, whereas a large k_n can increase variance due to insufficient number of observations.

We first consider how to solve (2.6). Define $\tilde{M}(Y_i, W_i; \beta_1) = (Y_i - \beta_1^T W_i) W_i$,

$M_j(Y_i, Z_i, W_i, v_j; \beta_1) = \langle Z_i, v_j \rangle (Y_i - \beta_1^T W_i)$ and

$$r_j = \sum_{i=1}^n [\delta_i M_j(Y_i, Z_i, W_i, \hat{v}_j; \beta_1) + (1 - \delta_i) m_{M, i, \gamma}^0(Y_i, Z_i, W_i, \hat{v}_j; \beta_1)] / (n \hat{\lambda}_j)$$

Solving (2.6) is equivalent to solving $U(\beta_1) = 0$, where $U(\beta_1)$ is given by

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{K_n} r_j [\delta_i W_i \langle Z_i, \hat{v}_j \rangle + (1 - \delta_i) \frac{E\{\delta_i \langle Z_i, v_j \rangle W_i \exp(\gamma Y_i) | Z_i, W_i\}}{E\{\delta_i \exp(\gamma Y_i) | Z_i, W_i\}} | v_j = \hat{v}_j] \\ & - \frac{1}{n} \sum_{i=1}^n [\delta_i \tilde{M}(Y_i, W_i; \beta_1) + (1 - \delta_i) m_{M, i, \gamma}^0(Y_i, W_i; \beta_1)]. \end{aligned}$$

Then, we have the following theorem, whose proofs can be found in the supplementary document

Theorem 1—Suppose Assumptions (A.1)–(A.9) hold. Then, as $n \rightarrow \infty$ there exists a unique solution $\hat{\beta}_1$ of $U(\beta_1) = 0$, which converges to $\beta_{1,0}$ in probability, and

$\hat{\theta} = \sum_{j=1}^k r_j(\hat{\beta}_1) \hat{v}_j$ satisfies $\|\hat{\theta} - \theta_0\|_{L_2} \rightarrow 0$ in probability, where

$$r_j(\beta_1) \triangleq (n\hat{\lambda}_j)^{-1} \sum_{i=1}^n [\delta_i M_j(Y_i, Z_i, W_i, \hat{v}_j; \beta_1) + (1 - \delta_i) m_{M_j, i, \gamma}^0(Y_i, Z_i, W_i, \hat{v}_j; \beta_1)].$$

Moreover, we have $\|\hat{\beta}_1 - \beta_{1,0}\| = O_p(k_n^{2a+1} n^{-1/2} + k_n^{1/2-b})$ and

$$\|\hat{\theta} - \theta_0\|_{L_2} = O_p(k_n^{5/2a+3/2} n^{-1/2} + k_n^{1+a/2-b}).$$

Remark 2

Denote $\mathcal{M} = \{(y_i, Z_i, W_i), i = n+1, \dots, n+N_0\}$ as a test set with N_0 new observations. For $i = n+1, \dots, n+N_0$, by using $\hat{y}_i = \langle Z_i, \hat{\theta} \rangle + \hat{\beta}_1^T W_i$ as the predicted response of the i th observation, the squared prediction error can be bounded by

$$\begin{aligned} & \frac{1}{N_0} \sum_{i=n+1}^{n+N_0} |\hat{y}_i - y_i|^2 = \frac{1}{N_0} \sum_{i=n+1}^{n+N_0} |\langle \hat{\theta}, Z_i \rangle + \hat{\beta}_1^T W_i - \langle \theta_0, Z_i \rangle + \beta_{1,0}^T W_i + \varepsilon_i|^2 \\ & \leq \frac{1}{N_0} \sum_{i=n+1}^{N_0} [\|Z_i\|^2 \|\hat{\theta} - \theta_0\|^2 + W_i^T W_i \|\hat{\beta}_1 - \beta_{1,0}\|^2] + \sigma^2 + O_p(1/\sqrt{N_0}) \\ & = \sigma^2 + O_p(k_n^{a+2-2b} + k_n^{5a+3}/n + 1/\sqrt{N_0}). \end{aligned}$$

In this case, we can obtain the optimal convergence rate $O_p(n^{(a+2-2b)/(4a+1+2b)} + 1/\sqrt{N_0})$ by minimizing $k_n^{a+2-2b} + k_n^{5a+3}/n$, which leads to $k_n = O(n^{1/(4a+1+2b)})$ and $\|\hat{\theta} - \theta_0\| = O(n^{(a/2+1-b)/(4a+2b+1)})$. Although these convergence rates are slower than those in Tang, et.al. (2014) and Hall and Horowitz (2007), our ETFLR is much more complex due to the inclusion of $G(Z, W)$ in model (2.2).

Second, we consider some approximations to the terms in (2.6) discussed in Subsection 2.2.2 and then solve $\tilde{U}(\beta_1) = 0$, where $\tilde{U}(\beta_1)$ is given by

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k \hat{r}_j [\delta_i W_i \langle Z_i, \hat{v}_j \rangle + (1 - \delta_i) \sum_{l=1}^n w_{l,0} \langle Z_i, W_i; \gamma \rangle \delta_l \langle Z_i, \hat{v}_j \rangle W_l] \\ & - \frac{1}{n} \sum_{i=1}^n [\delta_i \tilde{M}(Y_i, W_i; \beta_1) + (1 - \delta_i) \tilde{m}_{M, i, \gamma}(Y_i, W_i; \beta_1)], \end{aligned}$$

in which $\hat{r}_j$ is equal to

$$(n\hat{\lambda}_j)^{-1} \sum_{i=1}^n [\delta_i M_j(Y_i, Z_i, W_i, \hat{v}_j; \beta_1) + (1 - \delta_i) \hat{m}_{M_j, i, \gamma}(Y_i, Z_i, W_i, \hat{v}_j; \beta_1)].$$

We need some additional assumptions.

(B.1) There exists a constant C_1 such that $\max(\|Z\|, \|W\|) \leq C_1$.

(B.2) There exists a monotone and continuous function $G_1 : \mathbb{R} \rightarrow \mathbb{R}$ such that $G_1(G(Z, W)) \leq \|Z\| + \|W\|$ and the Lipschitz's condition holds such that $|G(Z_1, W_1) - G(Z_2, W_2)| \leq C(\|Z_1 - Z_2\| + \|W_1 - W_2\|)$.

(B.3) K is a kernel of type I (Martinez, 2013) if the function $K : \mathbb{R} \rightarrow [0, \infty)$

satisfies $\int_0^\infty K(u) du = 1$ and there exist constants c_1 and $c_2 \in \mathbb{R}$ such that $c_1 1_{u \in [0, 1]}$

$K(u) \leq c_2 1_{u \in [0, 1]}$ holds for $0 < c_1 < c_2 < \infty$.

(B.4) There exist a function ψ and a constant $\delta > 0$ such that $\forall \tau \in (0, \delta)$ and $(z, x) \in H_0 \triangleq \{(z, x) | z \in H, x \in \mathbb{R}^p, \max\{\|z\|, \|x\|\} \leq C_1\}$, we have $\psi_{z, x}(\tau) \geq \psi(\tau) > 0$, where $\psi_{z, x}(\tau) \triangleq \Pr[(Z, W) \in \{(\tilde{z}, \tilde{x}) | \|\tilde{z} - z\| \leq \tau, \|\tilde{x} - x\| \leq \tau\}]$.

(B.5) $k_n^{a+1}[h + 1/\sqrt{n\psi(h)}] \rightarrow 0$ as $n \rightarrow \infty$.

(B.6) The weight parameter $w_0 \in (0, 1)$.

Remark 3

Assumption (B.2) holds if $G(Z, W)$ is a bounded linear operator of (Z, W) such that $\|G\| = \|G(Z, W)\|/(\|Z\| + \|W\|) \leq C$ holds for a constant C . Assumption (B.3) is similar to Condition (C.2.19) of Martinez (2013). Assumption (B.4) is similar to and weaker than Condition (C.2.23) of Martinez (2013), and is equivalent to that the infimum $\inf_{(z, x) \in H_0} \psi_{z, x}(\cdot) \triangleq \psi(\cdot)$

exists and is uniformly positive in its domain, where $\psi_{z, x}$ is usually called the small ball probability. More details about the small ball probability and $\psi_{x, z}(h)$ can be found in Li and Shao (2001) and Ferraty and Vieu (2006, 2011). Compared with $m_{\psi, i, \gamma}^0$, the nonparametric kernel estimate $\hat{m}_{\psi, i, \gamma}$ brings in additional bias and variance associated with the tuning parameter h . Assumption B.5 ensures that such additional bias and variance are asymptotically negligible.

Theorem 2—Suppose that Assumptions (A.1)–(A.9) and (B.1)–(B.6) hold. Then, as $n \rightarrow \infty$, there exists a unique solution $\tilde{\beta}_1$ of $\tilde{U}(\beta_1) = 0$, which converges to $\beta_{1,0}$ in probability and

$\tilde{\theta} = \sum_{j=1}^k \hat{r}_j(\hat{\beta}_1) \hat{v}_j$ satisfies $\|\tilde{\theta} - \theta_0\|_{L_2} \rightarrow 0$ in probability, where $\hat{r}_j(\beta_1)$ is equal to

$$\sum_{i=1}^n [\delta_i M_j(Y_i, Z_i, W_i, \hat{v}_j; \beta_1) + (1 - \delta_i) \hat{m}_{M_j, i, \gamma}(Y_i, Z_i, W_i, \hat{v}_j; \beta_1)] / (n\hat{\lambda}_j).$$

Moreover, we have $\|\tilde{\beta}_1 - \beta_{1,0}\| = O_p(k_n^{1/2-b} + k_n^{2a+1}n^{-1/2} + k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}])$ and $\|\tilde{\theta} - \theta_0\| = O_p(k_n^{1+a/2-b} + k_n^{5a/2+3/2}n^{-1/2} + k_n^{(a+1)}[h + 1/\sqrt{n\psi(h)}])$.

Corollary 1—Assume that either $(w_0, \beta_{1,0}) = (1, \mathbf{0})$ or $(w_0, \theta_0) = (0, \mathbf{0})$ holds, and $\psi(h)$ is equal to $\inf_{(z,x) \in H_0} \psi_{z,x}(h)$, where

$\psi_{z,x}(\tau) = \Pr[(Z, W) \in \{(\tilde{z}, \tilde{x}) | w_0\|\tilde{z} - z\| + (1 - w_0)\|\tilde{x} - x\| \leq \tau\}]$. Under Assumptions (A.1)–(A.9) and (B.1)–(B.5), the conclusions in Theorem 2 remain valid.

Finally, we present a theoretical result that justifies the computational method in Subsection 2.2.2.

Theorem 3—If the tuning parameters h , w_0 , and k_n and the tilting parameter ϕ are fixed, and for any f_1 and $f_2 \in H$, $\langle f_1, f_2 \rangle$ is defined as $\sum_{k=1}^K f_1(t_k)f_2(t_k)(t_k - t_{k-1})$, then the solution to (2.6) is equal to the minimizer that minimizes (2.8).

4. Simulation Studies

In this section, we conducted two sets of Monte Carlo simulations to evaluate the finite-sample performance of $\hat{\theta}$ by using four different evaluation measures.

In the first set, we simulated datasets from ETFLR given by

$$Y = 0.5 \int_0^1 |\sin(4\pi t)| Z(t) dt + \beta_1 W + \alpha_1 + \varepsilon,$$

$$\text{logit}[\Pr(\delta = 1)] = \phi Y - \int_0^1 \sin(4\pi t) Z(t) dt - W + \alpha_2,$$

where $W \sim N(0, 1)$, $\varepsilon \sim N(0, \sigma^2)$, $Z = Z(t)$ is the standard Brownian motion, and Z , W , and ε are mutually independent. Moreover, we set $\alpha_1 = 0$ and $\beta_1 = 0.5$, whereas σ , α_2 and ϕ were varied for comparison purposes. We used α_2 to determine the missingness rate and ϕ to measure how the model is different from MAR. Moreover, we set $\mathbf{t} = \{0, 0.01, \dots, 0.99, 1\}$ and $K = 100$, and approximated $\int f(t) dt \approx \sum_{k=0}^{100} f(0.01k) \times 0.01$. The values of α and ϕ are key factors for controlling the missingness rate in each scenario. Specifically, the missingness rates are 66%, 60%, 56%, and 52%, when (ϕ, α) takes the values $(1, -1)$, $(2, -1)$, $(1, -0.5)$, and $(2, -0.5)$, respectively. For each simulated dataset, we used GCV to select $k_n \in \{1, 2, \dots, 20\}$ and empirically fixed $h = n^{-1/3}\sigma$ and $w_0 = 1/2$. Moreover, the Gaussian Kernel $K(t) = \exp(-0.5t^2)/\sqrt{2\pi}$ is utilized in (2.9).

We considered four competing estimates as follows:

- i. MCAR. Use complete observations to estimate the parameters.
- ii. MNAR($\hat{k}_n$). Set ϕ as the true value ϕ_0 .

- iii. MNAR($\hat{\phi}, \hat{k}_n$). Calculate $\hat{\phi}$ according to the first approach in subsection 2.2.3.
- iv. MAR. Set $\phi = 0$.

We simulated $S = 5,000$ data sets for each combination of $(\alpha, \phi, n, N/n)$, in which n and N denote the total sample size and that of validation dataset, respectively. Each test set has the same size as the training set. Let $\hat{y}_i^{(s)}, \hat{\theta}^{(s)}, \hat{\beta}^{(s)}$, and $\hat{\alpha}_1$ represent the i -th predicted response, the estimated θ, β , and α_1 in the s -th simulation, respectively. We consider four evaluation criteria as follows:

- The prediction bias:

$$\text{Bias} = S^{-1} \sum_{s=1}^S \sum_{i=1}^n (\hat{y}_i^{(s)} - y_i^{(s)})/n;$$

- The nonfunctional bias:

$$\text{Bias}_n = S^{-1} \sum_{s=1}^S \sum_{i=1}^{n-1} [(\hat{\beta}_1^{(s)} - 0.5)W + \hat{\alpha}_1^{(s)}]1;$$

- Mean integrated squared error (MISE₁):

$$\text{MISE}_1 = \text{Median}[\{ \sum_{j=1}^J (\hat{\theta}^{(s)}(t_j) - \theta(t_j))^2 (t_j - t_{j-1}) \}_{s \leq S}] \approx \text{Median}[\int (\hat{\theta}(t) - \theta(t))^2 dt].$$

- Median integrated squared error (MISE₂):

$$\text{MISE}_2 = \text{Mean}[\{ \sum_{j=1}^J (\hat{\theta}^{(s)}(t_j) - \theta(t_j))^2 (t_j - t_{j-1}) \}_{s \leq S}] \approx \text{Mean}[\int (\hat{\theta}(t) - \theta(t))^2 dt].$$

- Mean squared error for nonfunctional:

$$\text{MSE} = \text{Mean}\{(\hat{\beta}^{(s)} - 0.5)^2 + (\hat{\alpha}_1^{(s)} - 0)^2\}_{s \leq S}.$$

Table 1 summarizes the simulation results in all scenarios. In terms of Bias, Bias_n, and MSE, MNAR($\hat{k}_n$) outperforms all other methods and MNAR($\hat{\phi}, \hat{k}_n$) is the second best, indicating the advantage of incorporating the non-functional part estimation and response prediction for the MNAR methods. When either σ or ϕ becomes larger, MNARs are much better than their competing methods. As both n and N/n increase, the performance of MNAR($\hat{\phi}, \hat{k}_n$) is more similar to that of MNAR($\hat{k}_n$). In terms of MISE₁ and MISE₂, MNAR and MAR have similar performance, whereas MCAR is not very stable and has large MISE₁ particularly when n is small.

In the second set, we generated simulation data sets from ETFLR given by

$$Y = 0.5\langle \theta, Z \rangle + \beta_1 W + \alpha_1 + \varepsilon,$$

$$\text{logit}[\Pr(\delta = 1)] = \phi Y - \langle g, Z \rangle + \beta_2 W + \alpha_2,$$

where $\varepsilon \sim N(0, \sigma^2)$. We fixed $\beta_1 = \beta_2 = 0$, $\alpha_1 = 0$, and $\alpha_2 = -1$. We consider an image pool consisting of 1457 two-dimensional images and randomly selected $Z_k(\cdot)$'s from the pool. Figure 1 presents several randomly selected images.

We consider two scenarios for (θ, g) . In both scenarios, we randomly selected an image from the image pool for each simulated data set. In the first scenario, we chose 7 images for θ from the image pool. For each θ image, we generated 1000 simulated data sets. In the second scenario, we generated 5,000 simulated data sets and randomly selected θ from the image pool in each simulated data set. We evaluated $\text{MNAR}(\hat{k}_n)$, MCAR, MAR and $\text{MNAR}(\hat{\phi}, \hat{k}_n)$ by using the prediction Bias and MISE_1 , since they are close to Bias_n and MISE_2 , respectively, in the second set.

Table 2 presents the simulation results for both scenarios. MNARs outperform MCAR and MAR, indicating that selecting the correct missing data mechanism can improve both prediction and estimation accuracy. Figure 2 presents some randomly selected estimated $\hat{\theta}$'s. These $\hat{\theta}$'s can be quite different from θ since the basis functions constructed from FPCA may not accurately capture the variation of θ . However, for missing data problem, it remains largely unclear how to choose a set of efficient basis functions to accurately recover both the missing data and the functional signal. We will address this issue in our future research.

5. Application to the ADNI dataset

Alzheimer's disease (AD) is the most common form of dementia and is an escalating national epidemic and a genetically complex, progressive, and fatal neurodegenerative disease. The ADNI study is a large scale multi-site study which has collected clinical, imaging, and laboratory data at multiple time points from cognitively normal controls (CN), individuals with significant memory concern (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and subjects with AD. One of the goals of ADNI is to develop prediction methods to predict the longitudinal course of clinical outcomes (e.g., learning ability) based on imaging and biomarker data. More information about data acquisition can be found at the ADNI website (www.loni.usc.edu/ADNI).

To illustrate the empirical utility of our proposed methods in imaging classification, we use a subset of the ADNI data which consists of 682 subjects (208 CN controls, 153 AD patients, and 321 LMCI patients), after removing missing or low quality imaging data. Among them, there are 395 males with average age 75.38 years old and standard deviation 6.48 years old, and 287 females with average age 74.81 years old and standard deviation 6.81 years old. The T1-weighted images for all subjects at baseline were preprocessed by standard steps including AC (anterior commissure) and PC (posterior commissure) correction, N2 bias field correction, skull-stripping, intensity inhomogeneity correction, cerebellum removal, segmentation, and registration (Wang et al., 2011). Afterwards, we generated RAVENS-

maps (Davatzikos et al., 2001) for the whole brain using the deformation field obtained during registration. We obtained the $256 \times 256 \times 256$ RAVENS-maps and then down-sampled them to $128 \times 128 \times 128$ for real data analysis.

The development of ETFLR is motivated by using imaging and clinical variables at baseline to predict clinical outcomes after baseline. The covariates of interest at baseline include age, gender, education, marriage status (married, divorced, or widowed), APOE4 (risk from variations of the APOE gene), DX-bl (CN, SMC, EMCI, LMCI, and AD), as well as the RAVENS map. The learning ability of each subject was scored (the so-called Rey Auditory Verbal Learning Test Score) at the 6, 12, 18, 24, and 30 months after baseline. The missingness rates of the test scores at the 18th, 24th and 30th month are very high, e.g., the missingness rate at the 18th month is 53.1%. We are interested in examining the learning ability at the 18th month, at which all LMCI and AD patients were tested for learning ability. We have 682 individuals in total, among which 362 individuals have missing data. It is necessary to model the missing responses given the high missingness rate.

We applied ETFLR to this ADNI data set as follows. First, to determine the tilting parameter ϕ , we obtained a validation set by investigating the responses at months other than the 18th month for those observations with missing responses. We interpolated the responses at the 18th month by using those at other months by a linear regression, and then we calculated the p -value associated with month. The interpolations with p -values less than .05 were considered as the validation set, and their corresponding interpolated responses were approximately taken as the missing true responses. By using both MNARN and MNARP (which are named in Subsection 2.2.3), we calculated two estimates of ϕ . Second, given $\hat{\phi}$, the first k_n components (columns) of the functional covariate Z , together with the non-functional covariates W (age, gender, etc.), we calculated the estimates of all coefficients by optimizing the quadratic form (2.8). We used GCV to choose k_n as in (2.10).

Third, we used RRSV to choose the optimal (h, w_0) as in (2.11). Given a grid of (h, w_0) values, we used the average prediction on the test set to choose the optimal (h, w_0) . The criterion used is the Pearson correlation between the true and predicted responses. Specifically, we divided the dataset into half the test set and half the training set 500 times randomly, ensuring that both the missingness rate and the proportion of the validation set between the training set and the test set are the same. At each division, for every given h and w_0 , we calculated, Cor_{tr} and Cor_{test} for all approaches, where Cor_{tr} is the correlation between the predicted responses and the true responses on the training dataset, and Cor_{test} is the correlation between the predicted responses and the true responses on the test dataset. After 500 divisions, we calculated their averages in comparison with MCAR and MAR approaches. See Table 3 for such results. We found that both MNARP and MNARN outperform in almost all tuning parameters (h, w_0) 's, and the best (h, w_0) is achieved at $(1.1^3 h_{\min}, 0.5)$ for MNARN. We also examine whether the functional covariate leads to better prediction. Specifically, by setting $\theta_0 = \mathbf{0}$, we repeated the same estimation procedure to calculate $(\text{Cor}_{\text{test}}, \text{Cor}_{\text{tr}})$ at $(1.1^3 h_{\min}, 0.5)$, leading to $\text{Cor}_{\text{tr}} = 0.257(0.054)$ and $\text{Cor}_{\text{test}} = 0.127(0.059)$. Comparing such results with those in Table 3 reveals that RAVEN images can substantially improve prediction accuracy.

After fixing $h = 1.1^3 h_{\min}$ and $w_0 = 0.5$, we calculated the estimates of the non-functional covariates in Table 4. The bootstrap resampling procedure was further utilized for inference. Specifically, we repeated the bootstrap resampling procedure 300 times. At each time, we calculated the parameter estimates and $\hat{\phi}$. Subsequently, we calculated the 90% confidence intervals for ϕ and all other parameters and their associated p -values by using the Fast Double Bootstrap (Davidson and James, 2007). Table 4 also presents the bootstrap confidence intervals and their corresponding p -values based on MNARN. Table 5 presents the coefficient estimates for the four significant nonfunctional covariates and those for the principle components associated with the RAVEN images. Figure 3 presents the selected slices of the first two principle component images. We repeated the RRSV procedure and calculated the prediction accuracy (standard deviation) based on the test set for MCAR, MAR, MNARP, and MNARN as 0.143(0.063), 0.154(0.062), 0.167(0.059), and 0.170(0.057), respectively. The results indicate that MNARN outperforms all other three methods.

We have the following findings. First, MNAR performs well in both training and test sets for most window widths h and k_n . Second, the four covariates, Education, Apoe4, whether the individual is divorced, and whether the DX-bl of the individual is the LMCI ($= 1$) or the AD ($= 0$), strongly influence the learning test score. Such findings are clinically significant in that AD has a more serious effect on the intelligence behavior than LMCI. Third, the negative value of $\hat{\phi}$ in Table 4 implies that people with high learning test scores have the tendency to drop out of the study as expected. Finally, inspecting the functional coefficient image based on MNARN and MNARP (Table 5, Figures 3 and 5) reveals that estimates in most voxels are negative and relatively large in the regions of “lateral ventricle left” and “lateral ventricle right”. Such regions may have negative effects on learning ability. These findings are consistent with the existing literature on the abnormal lateral ventricle (Nestor, et.al. 2008) of AD patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Dr. Ibrahim’s research was partially supported by NIH grants #GM 70335 and P01CA142538. The research of Dr. Zhu was supported by NSF grants SES-1357666 and DMS-1407655, NIH grants MH086633, and a grant from Cancer Prevention Research Institute of Texas. The research was partially supported by NSFC to Dr. Xie (11271193). This material was based upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We are grateful for the many valuable suggestions from referees, associated editor, and editor. Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<http://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/howtoapply/ADNIAcknowledgementList.pdf>.

References

Baker SG, Laird NM. Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*. 1988; 83:62–69.

- Cai TT, Yuan M. Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*. 2012; 107:1201–1216.
- Cardot H, Ferraty F, Sarda P. Spline estimators for the functional linear model. *Statistica Sinica*. 2003; 13:571–591.
- Cheng PE. Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*. 1994; 89:81–87.
- Chiou JM, Zhang YC, Chen WH, Chang CW. A functional data approach to missing value imputation and outlier detection for traffic flow data. *Transportmetrica B: Transport Dynamics*. 2014; 2:106–129.
- Crambes C, André M. Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli*. 2013; 19:2627–2651.
- Crambes C, Kneip A, Sarda P. Smoothing splines estimators for functional linear regression. *The Annals of Statistics*. 2009; 37:35–72.
- Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *NeuroImage*. 2001; 14:1361–1369. [PubMed: 11707092]
- Davidson R, James GM. Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*. 2007; 51:3259–3281.
- Ferraty F, Sued M, Vieu P. Mean estimation with data missing at random for functional covariates. *Statistics: A Journal of Theoretical and Applied Statistics*. 2013; 47:688–706.
- Ferraty F, Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer; New York: 2006. Springer Series in Statistics
- Ferraty F, Vieu P. Kernel regression estimation for functional data. In: Ferraty F, Romain Y, editors *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics. Oxford University Press; Oxford: 2011. 72–129.
- Friedman J, Hastie T, Tibshirani R. glmnet: Lasso and elastic-net regularized generalized linear models. 2009 R package version 1.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016; 278:563–577. [PubMed: 26579733]
- Hall P, Giles H. Truncated linear models for functional data. *Journal of the Royal Statistical Society: Series B*. 2015 To appear.
- Hall P, Horowitz JL. Methodology and convergence rates for functional linear regression. *The Annals of Statistics*. 2007; 35:70–91.
- Hall P, Hosseini-Nasab M. On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B*. 2006; 68:109–126.
- Hall P, Hosseini-Nasab M. Theory for high-order bounds in functional principal components analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*. 2009; 146:225–256.
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; 47:663–685.
- Ibrahim JG, Chen MH, Lipsitz SR, Herring A. Missing-data methods for generalized linear models: a comparative review. *Journal of the American Statistical Association*. 2005; 100:332–346.
- Ibrahim JG, Lipsitz SR, Chen M. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B*. 1999; 61:173–190.
- Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *Test*. 2009; 18:1–43. [PubMed: 21218187]
- Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*. 2007; 22:523–539.
- Kim JK, Yu CL. A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association*. 2011; 106:157–165.
- Kong D, Joseph GI, Lee E, Zhu H. Functional linear Cox regression models. 2015 in submission.

- Li WV, Shao QM. Gaussian processes: Inequalities, small ball probabilities and applications. In: Shanbhag D, Rao C, editors *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*. Elsevier; New York: 2001. 533–597.
- Little RJA. A note about models for selectivity bias. *Econometrica*. 1985; 53:1469–1474.
- Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. Wiley; New York: 2002.
- Martinez CA. Phd dissertation. Universidade de Santiago de Compostela; 2013. Estimates and bootstrap calibration for functional regression with scalar response.
- Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*. Chichester: Wiley; 2007.
- Nestor SM, Rupasingh R, Borrie M, Smith M, Accomazzi V, Wells JL, Fogarty J, Bartha R, Alzheimer's Disease Neuroimaging Initiative. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain*. 2008; 131(9):2443–2454. [PubMed: 18669512]
- Preda C, Saporta G, Hadj MMH. The NIPALS algorithm for functional data. *Revue Roumaine de Mathematique Pures et Appliquees*. 2010; 55:315–326.
- Ramsay JO, Silverman BW. *Functional Data Analysis*. Second. John Wiley and Sons, Inc; 2006.
- Robins M, Rotnitzky A, Zhao L. Estimation of regression coefficients when some regressors are not always observed. *Journal of American Statistical Association*. 1994; 89:846–866.
- Rotnitzky A, Lei Q, Sued M, Robins JM. Improved double robust estimation in missing data and causal inference models. *Biometrika*. 2012; 99:439–456. [PubMed: 23843666]
- Shi X, Zhu H, Ibrahim JG. Local influence for generalized linear models with missing covariates. *Biometrics*. 2009; 65:1164–1174. [PubMed: 19210738]
- Stewart GW. On the continuity of the generalized inverse. *SIAM Journal on Applied Mathematics*. 1969; 17:33–45.
- Tang N, Zhao P, Zhu H. Empirical likelihood for estimating equations with nonignorable missing data. *Statistica Sinica*. 2014; 24:723–747. [PubMed: 24976738]
- Van der Vaart AW. *Asymptotic Statistics*. Cambridge University Press; 2000.
- Van Der Vaart AW, Wellner JA. *Weak Convergence and Empirical Processes*. Springer; 2000.
- Wang D, Chen SX. Empirical likelihood for estimating equations with missing values. *The Annals of Statistics*. 2009; 37:490–517.
- Wang Y, Nie J, Yap P, Shi F, Guo L, Shen D. Robust deformable-surface- based skull-stripping for large-scale studies. In: Fichtinger G, Martel A, Peters T, editors *Medical Image Computing and Computer-Assisted Intervention*. Vol. 6893. Springer Berlin/Heidelberg; 2011. 635642
- Yao F, Müller HG, Wang JL. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*. 2005; 33:2873–2903.
- Zhao J, Shao J. Semiparametric pseudo likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association*. 2014; 110:1577–1590.
- Zhou Y, Wan AT, Wang X. Estimating equations inference with missing data. *Journal of the American Statistical Association*. 2008; 103:1187–1199.

Figure 1.
The image pool used in the second simulation setting.

Figure 2.

The selected 7 images in the second simulation setting—true value and their estimates. Images with IDs from 1 to 7 are shown in the 1st row to the 7th row, while the 5 columns represent the corresponding image estimates using the $\text{MNAR}(\hat{\phi}, \hat{k}_n)$, MCAR, MAR, and $\text{MNAR}(\hat{\phi}, \hat{k}_n)$, and the true image, respectively.

Figure 3.

The first Principle Component (PC) of RAVEN images of ADNI real data analysis: positive loadings with the coronal slice at $y = 62$ (upper left), the sagittal slice at $x = 71$ (middle left), and the axial slice at $z = 50$ (lower left); negative loadings with the coronal slice at $y = 66$ (upper left), the sagittal slice at $x = 71$ (middle left), and the axial slice at $z = 39$ (lower left).

Figure 4.

The second PC of RAVEN images of ADNI real data analysis: positive loadings with the coronal slice at $y = 62$ (upper left), the sagittal slice at $x = 71$ (middle left), and the axial slice at $z = 50$ (lower left); negative loadings with the coronal slice at $y = 62$ (upper right), the sagittal slice at $x = 71$ (middle right), and the axial slice at $z = 50$ (lower right).

Figure 5.

The functional coefficient images of MNARP of ADNI real data analysis: the positive part with the coronal slice at $y = 62$ (upper left), the sagittal slice at $x = 76$ (middle left), and the axial slice at $z = 43$ (lower left); the negative part with the coronal slice at $y = 62$ (upper right), the sagittal slice at $x = 71$ (middle right), and the axial slice at $z = 50$ (lower right).

Table 1

Simulation results based on 5000 replications for the first simulation setting.

test	(ϕ, α, σ^2)	$(n, N/n)$	MNAR($\hat{k}_H$)	MCAR	MAR	MNAR($\hat{\phi}, \hat{k}_H$)
Bias	(1,-1,.25)	(300,1/30)	0.090	0.159	0.158	0.144
Bias _n	(1,-1,.25)	(300,1/30)	0.129	0.188	0.192	0.179
MISE ₁	(1,-1,.25)	(300,1/30)	0.166	1.336	0.186	0.195
MISE ₂	(1,-1,.25)	(300,1/30)	0.065	0.070	0.065	0.066
MSE	(1,-1,.25)	(300,1/30)	0.011	0.017	0.018	0.017
Bias	(1,-1,.64)	(300,1/30)	0.137	0.376	0.378	0.242
Bias _n	(1,-1,.64)	(300,1/30)	0.189	0.411	0.393	0.278
MISE ₁	(1,-1,.64)	(300,1/30)	0.223	3.340	0.199	0.205
MISE ₂	(1,-1,.64)	(300,1/30)	0.077	0.099	0.089	0.082
MSE	(1,-1,.64)	(300,1/30)	0.024	0.070	0.067	0.042
Bias	(2,-1,.25)	(300,1/30)	0.200	0.534	0.536	0.267
Bias _n	(2,-1,.25)	(300,1/30)	0.262	0.608	0.563	0.324
MISE ₁	(2,-1,.25)	(300,1/30)	0.179	3.260	0.165	0.158
MISE ₂	(2,-1,.25)	(300,1/30)	0.074	0.101	0.077	0.074
MSE	(2,-1,.25)	(300,1/30)	0.033	0.132	0.117	0.047
Bias	(2,-1,.64)	(300,1/30)	0.138	0.260	0.264	0.187
Bias _n	(2,-1,.64)	(300,1/30)	0.174	0.301	0.272	0.213
MISE ₁	(2,-1,.64)	(300,1/30)	0.097	1.349	0.104	0.099
MISE ₂	(2,-1,.64)	(300,1/30)	0.062	0.069	0.066	0.062
MSE	(2,-1,.64)	(300,1/30)	0.014	0.034	0.030	0.020
Bias	(1,-1,.64)	(600,1/10)	0.116	0.375	0.377	0.166
Bias _n	(1,-1,.64)	(600,1/10)	0.160	0.406	0.397	0.204
MISE ₁	(1,-1,.64)	(600,1/10)	0.143	0.798	0.130	0.140
MISE ₂	(1,-1,.64)	(600,1/10)	0.072	0.089	0.077	0.075
MSE	(1,-1,.64)	(600,1/10)	0.065	0.070	0.069	0.065
Bias	(1,-1,.64)	(300,1/10)	0.137	0.376	0.378	0.200

test	$(\phi, \alpha, \sigma^2),$	$(n, N/n)$	MNAR($\hat{k}_n$)	MCAR	MAR	MNAR($\hat{\phi}, \hat{k}_n$)
Bias _n	(1,-1,.64)	(300,1/10)	0.189	0.406	0.393	0.240
MISE ₁	(1,-1,.64)	(300,1/10)	0.223	1.443	0.199	0.202
MISE ₂	(1,-1,.64)	(300,1/10)	0.077	0.088	0.089	0.079
MSE	(1,-1,.64)	(300,1/10)	0.024	0.068	0.067	0.032
Bias	(1,-0.5,.64)	(300,1/30)	0.105	0.315	0.318	0.203
Bias _n	(1,-0.5,.64)	(300,1/30)	0.153	0.352	0.340	0.237
MISE ₁	(1,-0.5,.64)	(300,1/30)	0.207	3.912	0.190	0.211
MISE ₂	(1,-0.5,.64)	(300,1/30)	0.072	0.089	0.077	0.075
MSE	(1,-0.5,.64)	(300,1/30)	0.016	0.052	0.050	0.031
Bias	(1,-0.5,.25)	(600,1/3)	0.059	0.134	0.134	0.078
Bias _n	(1,-0.5,.25)	(600,1/3)	0.094	0.159	0.162	0.112
MISE ₁	(1,-0.5,.25)	(600,1/3)	0.095	0.131	0.103	0.096
MISE ₂	(1,-0.5,.25)	(600,1/3)	0.057	0.057	0.058	0.057
MSE	(1,-0.5,.25)	(600,1/3)	0.005	0.011	0.011	0.006
Bias	(2,-0.5,.25)	(300,1/30)	0.103	0.216	0.220	0.150
Bias _n	(2,-0.5,.25)	(300,1/30)	0.140	0.257	0.239	0.180
MISE ₁	(2,-0.5,.25)	(300,1/30)	0.094	1.100	0.098	0.099
MISE ₂	(2,-0.5,.25)	(300,1/30)	0.059	0.067	0.061	0.060
MSE	(2,-0.5,.25)	(300,1/30)	0.010	0.025	0.023	0.015
Bias	(2,-0.5,.25)	(300,1/5)	0.103	0.217	0.219	0.131
Bias _n	(2,-0.5,.25)	(300,1/5)	0.140	0.251	0.239	0.166
MISE ₁	(2,-0.5,.25)	(300,1/5)	0.094	0.335	0.098	0.094
MISE ₂	(2,-0.5,.25)	(300,1/5)	0.059	0.062	0.061	0.059
MSE	(2,-0.5,.25)	(300,1/5)	0.010	0.024	0.023	0.012
Bias	(2,-0.5,.25)	(600,1/2)	0.092	0.217	0.219	0.110
Bias _n	(2,-0.5,.25)	(600,1/2)	0.131	0.252	0.248	0.149
MISE ₁	(2,-0.5,.25)	(600,1/2)	0.078	0.136	0.079	0.076
MISE ₂	(2,-0.5,.25)	(600,1/2)	0.055	0.056	0.056	0.055
MSE	(2,-0.5,.25)	(600,1/2)	0.007	0.023	0.022	0.009

Table 2

Simulation results based on 2-dimensional image covariates for the second simulation setting. Seven identification (ID) numbers were assigned to 7 randomly chosen images (see Figure 2) each producing 1000 replications of data sets. ID=‘-’ indicates data set were generated with randomly chosen image coefficient in each replication.

test	ID	(σ^2)	$(n, N/n)$	MNAR($\hat{\phi}, \hat{k}_n$)	MCAR	MAR	MNAR($\hat{\phi}, \hat{k}_n$)
Bias	1	(1.,25)	(600,1/10)	0.060	0.077	0.071	0.059
MISE ₁	1	(1.,25)	(600,1/10)	0.636	0.901	0.704	0.634
Bias	2	(1.,25)	(1000,1/10)	0.063	0.081	0.076	0.062
MISE ₁	2	(1.,25)	(1000,1/10)	0.689	0.902	0.751	0.696
Bias	3	(1.,25)	(600,1/10)	0.064	0.082	0.076	0.063
MISE ₁	3	(1.,25)	(600,1/10)	0.731	0.958	0.770	0.737
Bias	4	(1.,25)	(600,1/10)	0.060	0.078	0.072	0.059
MISE ₁	4	(1.,25)	(600,1/10)	0.660	0.910	0.726	0.669
Bias	5	(1.,25)	(600,1/10)	0.060	0.078	0.072	0.059
MISE ₁	5	(1.,25)	(600,1/10)	0.642	0.899	0.701	0.649
Bias	6	(1.,25)	(600,1/10)	0.064	0.083	0.077	0.064
MISE ₁	6	(1.,25)	(600,1/10)	0.673	0.893	0.758	0.674
Bias	7	(1.,25)	(600,1/10)	0.064	0.082	0.076	0.064
MISE ₁	7	(1.,25)	(600,1/10)	0.733	0.934	0.767	0.741
Bias	-	(1.,25)	(600,1/5)	0.065	0.082	0.077	0.064
MISE ₁	-	(1.,25)	(600,1/5)	0.671	0.814	0.746	0.670
Bias	-	(1.,64)	(600,1/5)	0.114	0.200	0.192	0.113
MISE ₁	-	(1.,64)	(600,1/5)	0.707	1.141	0.960	0.708
Bias	-	(2.,25)	(600,1/5)	0.074	0.097	0.093	0.074
MISE ₁	-	(2.,25)	(600,1/5)	0.629	0.778	0.686	0.630
Bias	-	(1.,64)	(1000,1/3)	0.114	0.199	0.194	0.113
MISE ₁	-	(1.,64)	(1000,1/3)	0.676	0.885	0.867	0.676

Table 3

ADNI data analysis results: prediction accuracy scores.

	(w_b, h)	MCAR	MAR	MINARP	MINARN
Cor _{tr}	(0, 1.1 ³ $h_{\min}$)	.3238(.0695)	.2984 (.0658)	.2994 (.0654)	.2970 (.0619)
Cor _{test}	(0, 1.1 ³ $h_{\min}$)	.1161 (.0605)	.1384 (.0609)	.1388 (.0623)	.1377 (.0629)
Cor _{tr}	(0, 1.1 ⁴ $h_{\min}$)	.3244(.0707)	.2993 (.0761)	.3000 (.0665)	.2967 (.0621)
Cor _{test}	(0, 1.1 ⁴ $h_{\min}$)	.1160 (.0607)	.1380 (.0608)	.1389 (.0615)	.1387 (.0624)
Cor _{tr}	(0, 1.1 ⁵ $h_{\min}$)	.3243(.0701)	.2997 (.0672)	.3013 (.0674)	.2984 (.0633)
Cor _{test}	(0, 1.1 ⁵ $h_{\min}$)	.1150 (.0616)	.1373 (.0603)	.1384 (.0616)	.1377 (.0619)
Cor _{tr}	(0, 1.1 ⁶ $h_{\min}$)	.3264(.0710)	.3010 (.0656)	.3031 (.0671)	.3014 (.0639)
Cor _{test}	(0, 1.1 ⁶ $h_{\min}$)	.1153 (.0613)	.1369 (.0592)	.1371 (.0604)	.1360 (.0611)
Cor _{tr}	(0, 1.1 ⁷ $h_{\min}$)	.3269(.0706)	.3021 (.0665)	.3039 (.0678)	.3012 (.0635)
Cor _{test}	(0, 1.1 ⁷ $h_{\min}$)	.1146 (.0601)	.1368 (.0586)	.1373 (.0588)	.1368 (.0600)
Cor _{tr}	(1, 1.1 ⁰ $h_{\min}$)	.3266(.0724)	.3023 (.0645)	.2949 (.0617)	.2900 (.0611)
Cor _{test}	(1, 1.1 ⁰ $h_{\min}$)	.1146 (.0620)	.1394 (.0586)	.1435 (.0593)	.1419 (.0598)
Cor _{tr}	(1, 1.1 ¹ $h_{\min}$)	.3255(.0695)	.3055 (.0672)	.2972 (.0627)	.2911 (.0633)
Cor _{test}	(1, 1.1 ¹ $h_{\min}$)	.1152 (.0625)	.1404 (.0580)	.1459 (.0593)	.1447 (.0594)
Cor _{tr}	(1, 1.1 ² $h_{\min}$)	.3265(.0720)	.3069 (.0713)	.3009 (.0675)	.2944 (.0648)
Cor _{test}	(1, 1.1 ² $h_{\min}$)	.1148 (.0609)	.1405 (.0577)	.1439 (.0587)	.1443 (.0589)
Cor _{tr}	(1, 1.1 ³ $h_{\min}$)	.3266(.0731)	.3063 (.0720)	.3020 (.0695)	.2947 (.0647)
Cor _{test}	(1, 1.1 ³ $h_{\min}$)	.1139 (.0614)	.1393 (.0589)	.1438 (.0593)	.1447 (.0593)
Cor _{tr}	(1, 1.1 ⁴ $h_{\min}$)	.3260(.0718)	.3064 (.0722)	.3032 (.0720)	.2972 (.0662)
Cor _{test}	(1, 1.1 ⁴ $h_{\min}$)	.1145 (.0613)	.1395 (.0588)	.1436 (.0598)	.1443 (.0594)
Cor _{tr}	(1, 1.1 ⁵ $h_{\min}$)	.3268(.0726)	.3060 (.0685)	.3030 (.0691)	.2990 (.0668)
Cor _{test}	(1, 1.1 ⁵ $h_{\min}$)	.1147 (.0621)	.1398 (.0593)	.1433 (.0597)	.1445 (.0597)
Cor _{tr}	(1, 1.1 ⁶ $h_{\min}$)	.3230(.0693)	.2997 (.0668)	.2998 (.0681)	.2957 (.0618)
Cor _{test}	(1, 1.1 ⁶ $h_{\min}$)	.1155 (.0609)	.1398 (.0609)	.1430 (.0612)	.1443 (.0616)
Cor _{tr}	(1, 1.1 ⁷ $h_{\min}$)	.3277(.0720)	.3055 (.0689)	.3024 (.0688)	.2991 (.0677)

	(w_0, h)	MCAR	MAR	MNARP	MNARN
Cor _{test}	(1, 1, 1 ⁷ $h_{\min}$)	.1139 (.0615)	.1363 (.0594)	.1404 (.0603)	.1405 (.0604)
Cor _{tr}	(1, 1, 1 ⁸ $h_{\min}$)	.3273(.0729)	.3052 (.0688)	.3026 (.0704)	.2996 (.0677)
Cor _{test}	(1, 1, 1 ⁸ $h_{\min}$)	.1135 (.0614)	.1364 (.0595)	.1398(.0597)	.1388 (.0600)
Cor _{tr}	(0.5, 1, 1 ⁰ $h_{\min}$)	.3236(.0692)	.2999 (.0614)	.2908 (.0568)	.2888 (.0555)
Cor _{test}	(0.5, 1, 1 ⁰ $h_{\min}$)	.1152 (.0615)	.1409 (.0598)	.1444 (.0597)	.1433(.0607)
Cor _{tr}	(0.5, 1, 1 ¹ $h_{\min}$)	.3243(.0705)	.3011 (.0653)	.2947 (.0619)	.2912 (.0594)
Cor _{test}	(0.5, 1, 1 ¹ $h_{\min}$)	.1150 (.0610)	.1413 (.0591)	.1443 (.0604)	.1436(.0621)
Cor _{tr}	(0.5, 1, 1 ² $h_{\min}$)	.3242(.0685)	.3055 (.0695)	.3000 (.0670)	.2947 (.0614)
Cor _{test}	(0.5, 1, 1 ² $h_{\min}$)	.1157 (.0602)	.1415 (.0599)	.1462 (.0597)	.1462(.0616)
Cor _{tr}	(0.5, 1, 1 ³ $h_{\min}$)	.3232(.0702)	.3007 (.0694)	.2997 (.0683)	.2935 (.0613)
Cor _{test}	(0.5, 1, 1 ³ $h_{\min}$)	.1163 (.0608)	.1418 (.0602)	.1461(.0599)	.1470 (.0613)
Cor _{tr}	(0.5, 1, 1 ⁴ $h_{\min}$)	.3234(.0700)	.3004 (.0704)	.3015 (.0692)	.2961 (.0630)
Cor _{test}	(0.5, 1, 1 ⁴ $h_{\min}$)	.1163 (.0609)	.1412 (.0614)	.1453 (.0605)	.1464 (.0610)
Cor _{tr}	(0.5, 1, 1 ⁵ $h_{\min}$)	.3238(.0689)	.3027 (.0692)	.3014 (.0696)	.2965 (.0631)
Cor _{test}	(0.5, 1, 1 ⁵ $h_{\min}$)	.1161 (.0601)	.1412 (.0607)	.1447 (.0606)	.1462 (.0613)
Cor _{tr}	(0.5, 1, 1 ⁶ $h_{\min}$)	.3237(.0702)	.2997 (.0666)	.3001 (.0679)	.2958 (.0616)
Cor _{test}	(0.5, 1, 1 ⁶ $h_{\min}$)	.1158 (.0608)	.1409 (.0603)	.1438 (.0603)	.1447 (.0620)
Cor _{tr}	(0.5, 1, 1 ⁷ $h_{\min}$)	.3239(.0698)	.2993 (.0648)	.2984 (.0651)	.2945 (.0596)
Cor _{test}	(0.5, 1, 1 ⁷ $h_{\min}$)	.1163 (.0607)	.1411 (.0604)	.1442 (.0601)	.1449 (.0599)
Cor _{tr}	(0.5, 1, 1 ⁸ $h_{\min}$)	.3248(.0702)	.2993 (.0653)	.2994 (.0655)	.2958 (.0591)
Cor _{test}	(0.5, 1, 1 ⁸ $h_{\min}$)	.1149 (.0606)	.1390 (.0605)	.1422 (.0606)	.1423 (.0605)
Cor _{tr}	(0.5, 1, 1 ⁹ $h_{\min}$)	.3236(.0693)	.3001 (.0665)	.2997 (.0665)	.2979 (.0618)
Cor _{test}	(0.5, 1, 1 ⁹ $h_{\min}$)	.1158 (.0608)	.1394 (.0605)	.1416 (.0612)	.1416 (.0616)

ADNI data analysis results: parameter estimates, 90% confidence intervals and *p*-values of regression coefficients.

Table 4

Covariates	MCAR	MAR	MNARP	MNARN	P-value
Age	-0.0157 [-0.020, 0.059]	-0.0059 [-0.028, 0.021]	0.0023 [-0.027, 0.035]	0.0112 [-0.029, 0.044]	0.56 -
Gender	0.0750 [-0.804, 0.602]	-0.1296 [-0.546, 0.559]	0.1014 [-0.516, 0.795]	0.2824 [-0.546, 0.969]	0.18 -
Education	0.0717 [-0.007, 0.145]	0.0714 [.0004, 0.140]	0.0980 [0.010, 0.201]	0.1276 [0.005, 0.209]	0.00 ***
Apoe4	-0.3857 [-0.551, -0.019]	-0.3567 [-0.594, -0.070]	-0.4180 [-0.728, -0.106]	-0.4563 [-0.817, -0.108]	0.02 **
if.widowed (marriage)	0.3891 [-0.303, 1.076]	0.3874 [-0.191, 1.035]	0.2955 [-0.265, 0.989]	0.2287 [-0.316, 0.980]	0.32 -
if.divorced (marriage)	1.2330 [0.405, 2.225]	1.2590 [0.321, 2.201]	1.4663 [0.418, 2.556]	1.6364 [0.452, 2.768]	0.04 **
if.lmci (DX-bl)	1.6478 [-0.135, 2.731]	1.7373 [0.945, 3.069]	2.7110 [1.309, 3.921]	3.4572 [1.531, 4.100]	0.02 **
$\hat{\phi}$	- -	0 0	-0.1367 [-0.292, -0.061]	-0.2224 [-0.392, -0.061]	

Table 5

ADNI: estimates for significant coefficients and their standard deviations in parentheses, and estimates for the top principle components.

	MCAR	MAR	MNARP	MNARN
Education	0.069 (0.046)	0.049 (0.044)	0.1101(0.053)	0.120 (0.055)
Apoe4	−0.379 (0.176)	−0.323 (0.169)	−0.3924(0.195)	−0.496 (0.216)
if.divorced (marriage)	1.158 (0.615)	1.118 (0.650)	1.3431(0.724)	1.500 (0.753)
if.lmci (DX-bl)	1.768 (0.958)	1.651 (0.556)	3.5929(0.613)	3.557(0.700)
1st.Pcomp	−0.227	−0.201	−0.314	−0.338
2ed.Pcomp	0	0.136	0.027	0
3ed.Pcomp	0	−0.090	−0.141	0
4th.Pcomp	0	0.178	0.0284	0
5th.Pcomp	0	0.202	0.060	0
6th.Pcomp	0	0.128	0	0
7th.Pcomp	0	0.041	0	0
8th.Pcomp	0	0.038	0	0
9th.Pcomp	0	−0.015	0	0
10th.Pcomp	0	−0.007	0	0