STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2020, VOL. 12, NO. 2, 244–252
https://doi.org/10.1080/19466315.2019.1677492

Taylor & Francis
Taylor & Francis Group

High-Dimensional Smoothing Splines and Application in Alzheimer's Disease Prediction Using Magnetic Resonance Imaging

Xiaowu Dai for the Alzheimer's Disease Neuroimaging Initiative

Consortium for Data Analytics in Risk, University of California Berkeley, Berkeley, CA

ABSTRACT

Recent evidence has shown that structural magnetic resonance imaging (MRI) is an effective tool for Alzheimer's disease (AD) prediction. While traditional MRI-based prediction uses images acquired at a single time point, a longitudinal study is more sensitive and accurate in detecting early pathological changes of the AD. Two main statistical difficulties arise in the longitudinal MRI-based analysis: (i) the inconsistent longitudinal scans among subjects (i.e., the different scanning time and the different total number of scans); (ii) the heterogeneous progressions of high-dimensional regions of interest (ROIs) in MRI. In this work, we propose a new feature selection and estimation method which can be applied to extract AD-related features from the heterogeneous longitudinal MRI. A key ingredient of our approach is a hybrid of the smoothing splines and the l_1-penalty. Smoothing splines can integrate information from heterogeneous progressions of ROIs and adapt to inconsistent scans of MRIs. The selection property of the l_1-penalty helps to select important ROIs related to AD. We introduce an efficient algorithm to perform the proposed method. Real data experiments on the Alzheimer's Disease Neuroimaging Initiative database are provided to corroborate some advantages of the proposed method for AD prediction in longitudinal studies.

ARTICLE HISTORY
Received October 2018
Accepted September 2019

KEYWORDS
Longitudinal data analysis; Smoothing splines; Variable selection; Varying-coefficient model

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia in the aged population (Prince et al. 2013). It is vital to identify AD-related pathological biomarkers and diagnose early-stage AD to prevent disease progression and take treatment in the earliest stage. A considerable amount of research has been devoted to the use of structured magnetic resonance imaging (MRI) for early-stage AD diagnosis (e.g., Jack et al. 2010, 2013). The structural MRI provides measures of cerebral atrophy, and it is shown to be closely coupled with clinical symptoms in AD (Jack et al. 2009).

Most works consider the cross-sectional study with MRI acquired at one single time point (see, e.g., Tzourio-Mazoyer et al. 2002; Aguilar et al. 2013; Liu, Zhang, and Shen 2016). However, the cross-sectional study could be insensitive to early pathological changes. As an alternative, longitudinal analysis of structural abnormalities has recently attracted attention (Zhang and Shen 2012; Yau et al. 2015; Chincarini et al. 2016). Existing longitudinal studies focus on the atrophy of a few well-known regions of interest (ROIs) such as the hippocampus, entorhinal cortex, and ventricular cortex. However, the prespecified ROIs are insufficient to capture the full morphological abnormality pattern of the brain MRI. A few other issues also remain as challenges in the longitudinal analysis. First, longitudinal scans across subjects are usually inconsistent. For example, subjects could have different scanning time and the different total number of scans. Second, the total number of ROIs in the brain is enormous compared with the number of subjects, which poses a challenge to select AD-related longitudinal biomarkers from the whole brain. Third, the rates of longitudinal change of ROIs are different, and this heterogeneity has not been considered in the modeling of AD progression.

Our goal is to identify AD-related ROIs in the whole brain with longitudinal MRI and perform AD prediction using selected ROIs. Specifically, we consider the varying coefficient model (Hastie and Tibshirani 1993) to characterize the heterogeneous changes of ROIs in the structural MRI. This model allows the nonlinear functional modeling between MRI and clinical cognition functions. We propose a novel feature selection method for high-dimensional varying coefficient models, where the key idea is combining the smoothing splines (Wahba 1990) and an l_1-penalty (Tibshirani 1996). Our method can simultaneously select and estimate AD-related ROIs. We also provide an efficient algorithm to implement the proposed feature selection method. Then the prediction is performed based on the selected longitudinal features and the estimated varying coefficients. Our approach is robust to the inconsistency among longitudinal scans and is adaptive to the heterogeneity of changes in different ROIs. The hypothetical AD dynamic biomarker curves (Jack et al. 2010, 2013) motivate the use of varying coefficient models in our approach. In particular, Jack et al. (2010) suggest that the rates of change over time for MRI and clinical cognition functions are in a temporally ordered manner, which implies the functional relationship between the atrophy of MRI and the change of cognition functions is nonlinear with time.

To evaluate our method, we conduct experiments using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We predict future clinical changes of mild cognitive impairment (MCI) subjects with longitudinal MRI data. The MCI is a prodromal stage of AD and the prediction of clinical changes helps

CONTACT Xiaowu Dai *xwdai@berkeley.edu* Consortium for Data Analytics in Risk, University of California Berkeley, 410 Evans Hall, Berkeley, CA 94720.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/SBR*.

to determine whether an MCI subject will convert into AD at a future time point, which is vital for early diagnosis of AD.

Main differences between this article and existing longitudinal studies in Zhang and Shen (2012), Yau et al. (2015), and Chincarini et al. (2016) are listed as follows.

- Different feature representations. We consider the varying coefficient model to characterize the nonlinear and smooth progression of longitudinal features, which is motivated by clinical findings and dynamic biomarker curves in Jack et al. (2010, 2013). On the contrary, Zhang and Shen (2012), Yau et al. (2015), and Chincarini et al. (2016) considered only linear representations for features.
- Different scalability to heterogeneous longitudinal scans. Our method does not require identical scanning times or an equal number of scans across samples. However, these conditions are necessary for Zhang and Shen (2012), Yau et al. (2015), and Chincarini et al. (2016).
- Different feature selection methods. We proposed a new feature selection method by combining the smoothing splines with an l_1-penalty, which enables simultaneous feature selection and varying coefficient estimation. The method is different from Zhang and Shen (2012), which performs the feature selection and estimation separately in a two-step procedure, and it is also different from Yau et al. (2015) and Chincarini et al. (2016), which only use prespecified features.

The rest of the article is organized as follows. We introduce our method in Section 2. We give the experiment results in Section 3. The concluding remarks and discussions are provided in Section 4. Additional material and proofs are relegated to Appendices A–E.

2. Methodology

The varying coefficient model describes time-dependent covariate effects on the response (Hastie and Tibshirani 1993). Given the scaled time $t \in [0, 1]$, the response functional $Y(\cdot)$ is related to covariates $X_1(\cdot), \ldots, X_p(\cdot)$ through

$$Y(t) = b + \sum_{j=1}^{p} \beta_j(t) X_j(t) + \varepsilon(t), \quad b \in \mathbb{R}, \tag{1}$$

where the centered noise process $\varepsilon(\cdot)$ is independent of covariates $X_j(\cdot)$'s. The model (1) allows a nonlinear functional relationship between $X_j(\cdot)$'s and $Y(\cdot)$ as the coefficients $\beta_j(\cdot)$'s vary over $t \in [0, 1]$. Take AD prediction in Section 3 as an example. The response represents the clinical cognitive test score, and the covariates include demographic information (such as age, gender, and education years) and ROIs in the structural brain MRI. The dependence of $\beta_j(\cdot)$'s on t implies the time-varying effects of covariates on the response. On the other hand, model (1) has an additive structure on covariates $X_j(\cdot)$'s to allow efficient estimation of the coefficients $\beta_j(\cdot)$'s.

In practice, we observe data for subjects $i = 1, \ldots, n$ at time t_{iv}, where $v = 1, 2, \ldots, m_i$ and $0 \le t_{i1} \le t_{i2} \le \cdots \le t_{im_i} \le 1$. Here, m_i and t_{iv}'s can be different for different subjects. Let $x_{ij}(\cdot)$ be the observation of covariate $X_j(\cdot)$ for subject i. Let y_{iv} be the response for the subject i at time t_{iv}. Then model (1) suggests

$$y_{iv} = b + \sum_{j=1}^{p} \beta_j(t_{iv}) x_{ij}(t_{iv}) + \varepsilon(t_{iv}), \quad b \in \mathbb{R}. \tag{2}$$

The structure of heterogeneous longitudinal data is illustrated in Figure 1, where some subjects could have missing feature values at certain time points. The number of covariates p in (2) can be larger than the sample size n, where (2) becomes a high-dimensional model. Since some covariates can be irrelevant with the response, we want to select relevant covariates $X_j(\cdot)$'s based on data (2). Then we use the chosen covariates for prediction.

We propose a new method to simultaneously select covariates and estimate their varying coefficients as follows. Assume that varying coefficients $\beta_1(\cdot), \beta_2(\cdot), \ldots, \beta_p(\cdot)$ reside in a reproducing kernel Hilbert space (RKHS), $(\mathcal{H}_K, \|\cdot\|_{\mathcal{H}_K})$, where the reproducing kernel is denoted by $K(\cdot, \cdot)$ (Wahba 1990). We want to find $\beta_1(\cdot), \beta_2(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K$ and $b \in \mathbb{R}$ to minimize

$$\frac{1}{N} \sum_{i=1}^{n} \sum_{v=1}^{m_i} \left[y_{iv} - b - \sum_{j=1}^{p} \beta_j(t_{iv}) x_{ij}(t_{iv}) \right]^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|_{\mathcal{H}_K}, \tag{3}$$

where $N = \sum_{i=1}^{n} m_i$ and $\|\cdot\|_{\mathcal{H}_K}$ denotes the RKHS norm. Note that the measurements for the same covariate at different time points: $x_{ij}(t_{i1}), \ldots, x_{ij}(t_{im_i})$, are highly correlated since they are

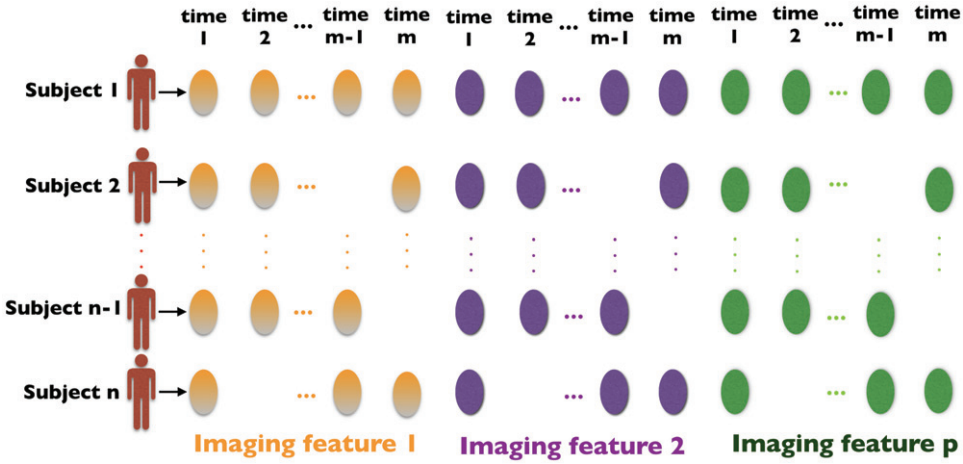

Figure 1. Illustration of heterogeneous longitudinal data with n subjects, p features, and m time points.

observed from the same random function $x_{ij}(\cdot)$, $j = 1, \ldots, p$. The proposed method (3) allows such correlation among measurements for the same covariate. The first term in (3) measures the goodness of data fitting and the second term merits the selection property by the l_1-like penalty $\sum_{j=1}^{p} \|\beta_j\|_{\mathcal{H}_K}$. We first provide the following theorem to justify the existence of minimizer for (3).

Theorem 2.1. There exists a minimizer of (3) that is in the domain $\beta_1(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K$ and $b \in \mathbb{R}$.

We give the proof of this theorem in Appendix C. The variable selection method (3) is new in the literature, and it is efficient for optimization due to the convexity in $\beta_j(\cdot)$'s and having only one tuning parameter λ. We provide an algorithm for giving a minimizer of (3) in Appendix D.

The following theorem gives further insights into (3), which is actually a combination of the smoothing splines (Wahba 1990) and the Lasso (Tibshirani 1996).

Theorem 2.2. Consider the following optimization problem. Find $\beta_1(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K$ and $\theta_1, \ldots, \theta_p, b \in \mathbb{R}$ to minimize

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{v=1}^{m_i}\left[y_{iv} - b - \sum_{j=1}^{p}\beta_j(t_{iv})x_{ij}(t_{iv})\right]^2 + \tau_0\sum_{j=0}^{p}\theta_j^{-1}\|\beta_j\|_{\mathcal{H}_K}^2 + \tau_1\sum_{j=0}^{p}\theta_j,$$
$$\text{s.t. } \theta_j \geq 0, j = 0, 1, \ldots, p, \tag{4}$$

where τ_0 is a constant and τ_1 is a tuning parameter. Let $\tau_1 = \lambda^4/(4\tau_0)$. The following equivalence holds.

1. If $\left(\widehat{\beta}_0, \widehat{\beta}_1(\cdot), \ldots, \widehat{\beta}_p(\cdot)\right)$ minimizes (3), by letting $\widehat{\theta}_j = \tau_0^{1/2}\tau_1^{-1/2}\|\widehat{\beta}_j\|_{\mathcal{H}_K}$, we have that $(\widehat{\theta}_1, \ldots, \widehat{\theta}_p; \widehat{\beta}_0, \widehat{\beta}_1(\cdot), \ldots, \widehat{\beta}_p(\cdot))$ minimizes (4).
2. If there exists $(\widehat{\theta}_1, \ldots, \widehat{\theta}_p; \widehat{\beta}_0, \widehat{\beta}_1(\cdot), \ldots, \widehat{\beta}_p(\cdot))$ minimizes (4), then $(\widehat{\beta}_0, \widehat{\beta}_1(\cdot), \ldots, \widehat{\beta}_p(\cdot))$ minimizes (3).

We give the proof of this theorem in Appendix E. The first two terms in (4):

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{v=1}^{m_i}[y_{iv} - b - \sum_{j=1}^{p}\beta_j(t_{iv})x_{ij}(t_{iv})]^2 + \tau_0\sum_{j=0}^{p}\theta_j^{-1}\|\beta_j\|_{\mathcal{H}_K}^2$$

amount to the smoothing splines in nonparametric statistics (Wahba 1990), which integrate information from heterogeneous progressions of ROIs through the varying coefficients $\beta_j(\cdot)$'s and adapt to inconsistent scans of MRIs with different m_i's. The last term in (4): $\tau_1\sum_{j=0}^{p}\theta_j$ is the same as the l_1 Lasso penalty (Tibshirani 1996) with weights θ_j's, which helps to select features.

Now we consider to use the selected covariates for prediction. Let $X_{j_1}, X_{j_2}, \ldots, X_{j_s}$ be s of selected features by (3), where $1 \leq j_1 \leq j_2 \leq \cdots \leq j_s \leq p$. Let $\widehat{\beta}_{j_1}(t), \widehat{\beta}_{j_2}(t), \ldots, \widehat{\beta}_{j_s}(t)$ be the corresponding varying coefficients estimated by (3). The prediction for a new subject with features $X_{j_1}^*(t), X_{j_2}^*(t), \ldots, X_{j_s}^*(t)$ at time t is given by

$$\widehat{Y}^*(t) = \widehat{\beta}_{j_1}(t)X_{j_1}^*(t) + \widehat{\beta}_{j_2}(t)X_{j_2}^*(t) + \cdots + \widehat{\beta}_{j_s}(t)X_{j_s}^*(t).$$

Note that the prediction for subject $i = 1, \ldots, n$ of the model (2) at a future time point t can also be performed. Here, the selected features $x_{ij_1}(t), x_{ij_2}(t), \ldots, x_{ij_s}(t)$ of subject i in the future time t are usually unknown. There are two methods to estimate the values of $x_{ij_1}(t), x_{ij_2}(t), \ldots, x_{ij_s}(t)$. The first method is to apply nonparametric statistical approaches such as the smoothing splines (Wahba 1990). Assume that the covariates $X_j(\cdot)$'s are smooth and belong to the RKHS $(\mathcal{H}_K, \|\cdot\|_{\mathcal{H}_K})$. We find $z(\cdot) \in \mathcal{H}_K$ to minimize

$$\frac{1}{m_i}\sum_{v=1}^{m_i}\left[z(t_{iv}) - x_{ij_1}(t_{iv})\right]^2 + \lambda_1\|z\|_{\mathcal{H}_K}^2,$$

where the smoothing parameter λ can be chosen by generalized cross-validation (GCV) (Wahba 1990). The feature $x_{ij_1}(t)$ at time t is then estimated by evaluating $z(\cdot)$ at t. That is, $\widehat{x}_{ij_1}(t) = z(t)$. Similarly, the smoothing splines can also be used to estimate other features: $x_{ij_2}(t), \ldots, x_{ij_s}(t)$. The second method to estimate the values of $x_{ij_1}(t), x_{ij_2}(t), \ldots, x_{ij_s}(t)$ is using the selected features of subject i at an observed time $\tilde{t}$ which is closest to time t: $\tilde{t} \in \{t_{i1}, t_{i2}, \ldots, t_{im_i}\}$ and $|\tilde{t} - t| = \min\{|t_{i_1} - t|, \ldots, |t_{im_i} - t|\}$. That is, $\widehat{x}_{ij_1}(t) = x_{ij_1}(\tilde{t}), \widehat{x}_{ij_2}(t) = x_{ij_2}(\tilde{t}), \ldots, \widehat{x}_{ij_s}(t) = x_{ij_s}(\tilde{t})$. With the estimated covariates values $\widehat{x}_{ij_1}(t), \widehat{x}_{ij_2}(t), \ldots, \widehat{x}_{ij_s}(t)$, we give the prediction for subject i in the future time t by

$$\widehat{y}_i(t) = \widehat{\beta}_{j_1}(t)\widehat{x}_{ij_1}(t) + \widehat{\beta}_{j_2}(t)\widehat{x}_{ij_2}(t) + \cdots + \widehat{\beta}_{j_s}(t)\widehat{x}_{ij_s}(t).$$

The experiments in Section 3 have known features in the future time t such as age, gender, and education years. However, the brain MRI data are unknown in future time. We consider the second method above to estimate the values of covariates $x_{ij_1}(t), x_{ij_2}(t), \ldots, x_{ij_s}(t)$, that is, using the feature values of subject i at an observed time $\tilde{t}$ which is closest to time t. The reasons are that the changes of MRI features are generally monotone and the variations of MRI features within subjects (i.e., longitudinal variations) are significantly smaller compared to that across subjects.

3. Experiment Results

In this section, we predict future clinical changes of MCI subjects with real data from the ADNI database. A detailed description of the ADNI database is relegated to Appendix A. The MCI is a prodromal stage of AD. Generally, some MCI subjects will convert into AD after a certain time (i.e., MCI converters, MCI-C for short), while others will not convert (i.e., MCI non-converters, MCI-NC for short) (Zhang and Shen 2012). The prediction of clinical change for an MCI subject helps to determine whether the subject will convert into AD at a future time, and it is a central task for the early diagnosis of AD. We summarize the baseline demographic information of ADNI subjects studied here in Table 1.

The preprocessing steps of brain MR imaging are described in Appendix B. Specifically, we have a total of 324 ROIs for each imaging. For MCI subjects, MRI scans were taken at baseline (bl), 6 months (M06), 1 year (M12), 18 months (M18), 2 years (M24), 3 years (M36), and 4 years (M48). However, some subjects may miss a few visits, and they do not have MRI scans at these missed time points. We choose $n = 172$ MCI subjects

Table 1. Demographics of ADNI subjects studied here.

	MCI-C ($n = 74$)	MCI-NC ($n = 98$)
Male/female	44/30	61/37
Age (years)	73.03 ± 6.65	74.35 ± 7.47
Edu. (years)	15.51 ± 3.05	15.59 ± 3.07

Table 2. Distribution of visit times for ADNI subjects studied here.

	MCI-C ($n = 74$)	MCI-NC ($n = 98$)
$\leq$3 scans	6	6
4 scans	8	14
5 scans	15	33
6 scans	45	45

Table 3. ADAS-Cog scores of ADNI subjects studied here.

	MCI-C ($n = 74$)	MCI-NC ($n = 98$)
ADAS-Cog (baseline)	20.12 ± 3.79	16.01 ± 3.91
ADAS-Cog (M48)	25.73 ± 4.22	17.14 ± 4.16
p-value	0.0031	0.2188

who have M48 imaging data. Table 2 lists the distributions of visit times for these 172 MCI subjects, where, for example, six of MCI-C subjects make at most three visits among the scheduled six times (bl, M06, M12, M18, M24, M36) such that they have at most three longitudinal MRI scans.

Our goal is to use longitudinal information (from bl up to M36) to predict the clinical changes of MCI subjects at M48. Empirical evidence suggests that the rates of change over time for structural MRI and clinical cognition functions are in a temporally ordered manner (Jack et al. 2010, 2013). Hence, we consider the varying coefficient model (1) for the nonlinear modeling of the functional relationship between the atrophy of MRI and the change in clinical cognition functions. The Alzheimer's Disease Assessment Scale—Cognitive Subscale (ADAS-Cog) is used as the response clinical cognitive test score $Y(\cdot)$. The ADAS-Cog score ranges from 70 (severe cognitive impairment) to 0 (no cognitive impairment) and it measures disturbances of memory, language, and other cognitive abilities. The prediction of future clinical scores based on the information at previous time points helps monitor the disease progression. We provide in Table 3 the average ADAS-Cog scores at baseline and M48 time point together with the p-value for the difference. There exists a significant difference between ADAS-Cog scores of baseline and M48 for MCI-C group while no significant difference for MCI-NC group, which indicates that ADAS-Cog scores for MCI-NC subjects increase much slower than that of the MCI-C subjects. In other words, Table 3 suggests that for an MCI subject if there is a significant increase in the prediction of ADAS-Cog score at M48 compared to ADAS-Cog score at baseline, the MCI subject is likely to be an MCI converter. Furthermore, a statistical classification model such as the support vector machines can also be built to classify MCI-C and MCI-NC subjects based on the predicted ADAS-Cog scores and selected features.

The covariates $X_j(\cdot)$'s in the varying coefficient model (1) consist of 324 MR imaging ROIs and three demographic covariates: age, gender, and education years. We let t in (1) be the scaled time relative to subjects entering the ADNI study. Hence, t is identifiable. Figure 2 gives the flowchart of our approach, where the selected features (e.g., cortical thickness of the right parahippocampal cortex and cortical thickness of the right entorhinal cortex) shown in the middle panel exhibit the abnormal shrinkages with time for the MCI-C. We build six models based on six different levels of longitudinal information in the training.

- Model 1 ($p = 327, \max_i m_i = 1$): All subjects that have observations at bl.
- Model 2 ($p = 327, \max_i m_i = 2$): All subjects that have observations at bl or M06.
- Model 3 ($p = 327, \max_i m_i = 3$): All subjects that have observations at bl or M06 or M12.
- Model 4 ($p = 327, \max_i m_i = 4$): All subjects that have observations at bl or M06 or M12 or M18.
- Model 5 ($p = 327, \max_i m_i = 5$): All subjects that have observations at bl or M06 or M12 or M18 or M24.
- Model 6 ($p = 327, \max_i m_i = 6$): All subjects that have observations at bl or M06 or M12 or M18 or M24 or M36.

In each of these six models, we allow the training subjects to have missing. For example, Model 3 includes subjects that only have observations at bl and M12 but missed M06.

Following the flowchart in Figure 2, we perform the feature selection method (3) and prediction as discussed in Section 2 for each of the six models. Take the training and testing for Model 3 as an example. The MCI-C and MCI-NC subjects are trained and tested separately. First, we randomly leave out half samples of both MCI-C and MCI-NC subjects, respectively, for testing in each experiment. For the training of Model 3, we choose subjects that have data at bl or M06 or M12 and use the longitudinal data in (3) and (4) to select features and estimate the varying coefficients. Here, $n = 172$, $p = 327$, and $\max_i m_i = 3$. Tuning parameters including the λ in (3) and the τ_1 in (4) are selected by 10-fold cross-validation with the criteria that minimizes the predictive root MSE of ADAS-Cog scores of subjects at bl, M06, and M12. With the selected features and estimated coefficients, we estimate the values of covariates at M48 as discussed in Section 2 and predict the ADAS-Cog scores at M48 for both MCI-C and MCI-NC subjects. The experiment of training and testing for Model 3 is replicated 100 times. We summarized the averaged predictive root MSE for ADAS-Cog scores at M48 in Figure 3. Similarly, we train and test all other five models besides Model 3. It is evident in Figure 3 that the longitudinal information can significantly improve the prediction results compared with only using baseline information. And the more longitudinal data included, the better prediction results are obtained. We also observe that prediction results for MCI-NC are better than those for MCI-C. This observation is expected since MCI-NC subjects have more stable clinical status and less varied clinical scores.

We show in Figure 4 the four features that are consistently selected in 100 experiments in training Model 6. These learned features agree with the existing medical discoveries on AD-related features; see Lin et al. (2015) for gender, Tognin et al. (2014) for the cortical thickness of the right parahippocampal cortex, Panizzon et al. (2009) for the surface area of the left

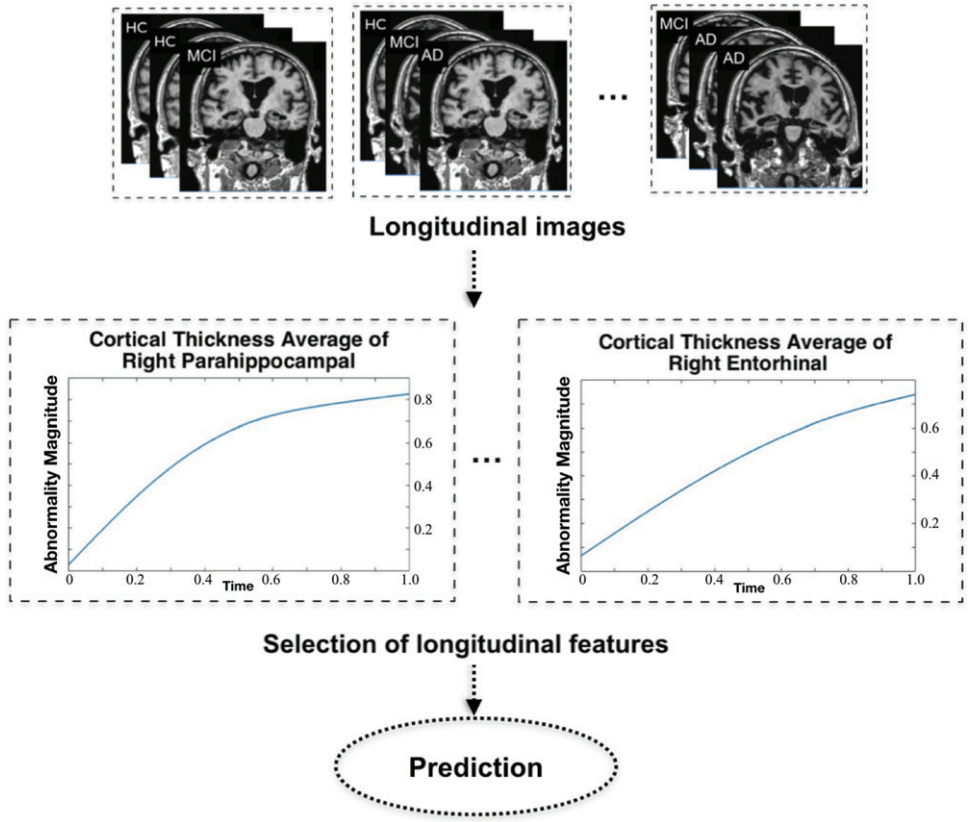

Figure 2. Flowchart of the proposed method. In the middle panel, the *x*-axis ("Time") is scaled to $[0, 1]$, where $t = 0$ represents bl, $t = 1$ represents M48. The *y*-axis ("Abnormality Magnitude") is the quantity that measures the abnormal shrinkage of a feature, that is, 1 − (average shrinkage of a feature of MCI-NC)/(average shrinkage of a feature of MCI-C).

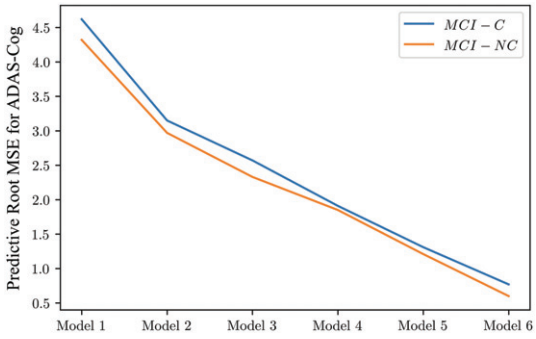

Figure 3. The prediction comparisons of our method using six levels of longitudinal data. The predictive root MSEs of ADAS-Cog scores for both MCI-C and MCI-NC subjects at M48 are reported.

parahippocampal cortex, and Velayudhan et al. (2013) for the cortical thickness of the right entorhinal cortex. Figure 4 also illustrates the varying coefficients of these selected features. The maximum effect of each MRI ROI varies throughout disease progression, which suggests that different MRI ROIs have different functional relations with the clinical ADAS-Cog score. This observation confirms the evidence and hypothesis in Sabuncu et al. (2011) and Schuff et al. (2012) that atrophy does not affect all regions of the brain simultaneously, but in a sequential manner.

Now we compare our method (3) with other two state-of-the-art methods:

- The longitudinal analysis in Chincarini et al. (2016), which only uses the hippocampal volume shrinkage rate as the feature and assumes the longitudinal trend being a linear map. This method is different from our proposed method (3) that uses all MRI ROIs, including the hippocampal volume shrinkage as features and assumes the nonlinear longitudinal trend.
- The longitudinal analysis in Zhang and Shen (2012), which depends on linear feature representations and a group Lasso for variable selection (Yuan and Lin 2006). This is different from our proposed method (3) that uses nonlinear feature representations in the varying coefficient model (1).

Since the methods in Chincarini et al. (2016) and Zhang and Shen (2012) require identical scanning times and an equal number of scans across samples, we perform Models 1–6 for AD prediction with data of subjects having no missing visits. In each experiment, we randomly leave out half of the samples in both MCI-C and MCI-NC for prediction. For the training of each model, we use the 10-fold cross-validation to select tuning parameters in (3) and also for the methods in Chincarini et al. (2016) and Zhang and Shen (2012). We replicate each experiment for 100 times. The prediction comparison results for MCI-C are shown in Figure 5 and the prediction comparison results for MCI-NC are summarized in Figure 6, where the predictive root MSEs for ADAS-Cog at M48 are reported. It is clear that our method consistently achieves better prediction performances for both MCI-C and MCI-NC. The reasons are that our method

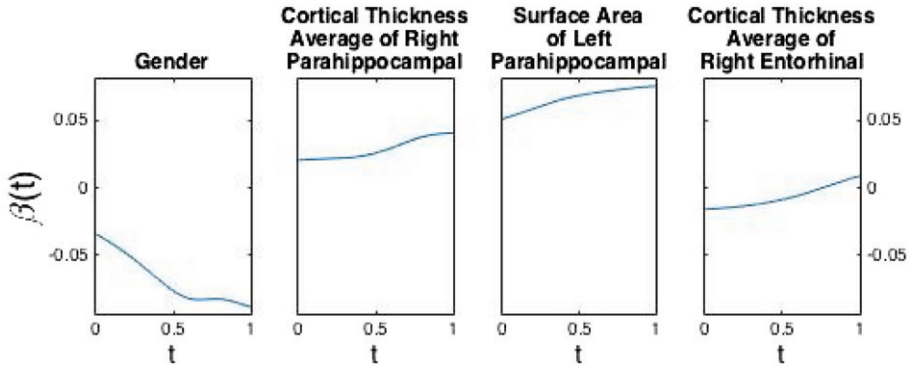

Figure 4. Examples of selected features in training Model 6. The time t is scaled to [0, 1], where $t = 0$ represents bl and $t = 1$ represents M48.

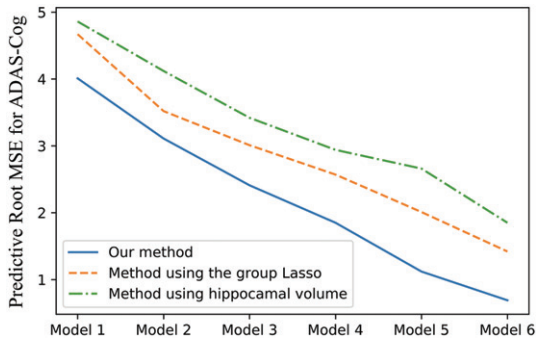

Figure 5. The prediction comparisons of three methods for MCI-C. The predictive root MSEs of ADAS-Cog scores for MCI-C subjects at M48 are reported.

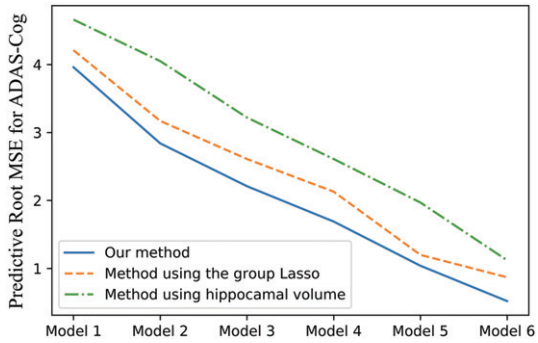

Figure 6. The prediction comparisons of three methods for MCI-NC. The predictive root MSEs of ADAS-Cog scores for MCI-NC subjects at M48 are reported.

models the nonlinear progression of longitudinal features and it selects AD-related features from the whole brain MRI instead of using only prespecified features for prediction.

4. Discussion

We study a framework to integrate longitudinal features from the structural MR images for AD prediction based on varying coefficient models. We propose a novel variable selection method by combining smoothing splines and Lasso, which enables simultaneous selection and estimation and is adaptive to heterogeneous longitudinal data. Our work is the first in the literature to model nonlinear progressions of longitudinal features in the high-dimensional setting. For validating our method, we conduct experiments with the ADNI dataset and show that the proposed method outperforms the state-of-the-art longitudinal analysis methods. It is promising and easy to implement the proposed method in other longitudinal data analysis examples.

It would be interesting to use the predicted ADAS-Cog scores in a statistical classification model to predict whether a new MCI subject will become a converter or not. Here, the MCI-C and MCI-NC subjects need to be trained together, which is slightly different from experiments in Section 3. Furthermore, we only use MR images for AD prediction in this article. It would be of great interest to apply the proposed method to integrate multimodal data, including MRI, PET, and functional MRI. We expect the integration of multimodal information would further improve the accuracy of the AD prediction. Another important future direction is to understand whether incorporating the known or estimated correlation structure between measurements at the same covariate, $x_{ij}(t_{i1}), \ldots, x_{ij}(t_{im_i})$, into our method (3) and (4) can improve the selection or estimation results for varying coefficients $\beta_1(\cdot), \ldots, \beta_p(\cdot)$.

Appendix A: ADNI Database Description

The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.

Clinical diagnosed AD patients must have had mild AD and had to meet the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD in McKhann et al. (1984). The mild cognitive impairment subjects should have largely intact general cognition as well as functional performance. Study subjects should have been given written informed consent at the time of enrollment for imaging and genetic sample collection and completed questionnaires approved by each participating sites Institutional Review Board (IRB).

Appendix B: Preprocessing of the Brain MRI Used Here

The structural MRI used in this study are cortical gray matter volumes processed using FreeSurfer software version 4.4 longitudinal image processing framework (*https://surfer.nmr.mgh.harvard.edu/*) ("ucsffsl" file). This dataset has been used in, for example, Tosun et al. (2011), Toledo et al. (2014), and Dai (2018). Specifically, subjects with a 1.5-T MRI were included in the dataset where the scans were preprocessed by certain correction methods including gradwarp, B1 calibration, N3 correction,

and skull-stripping (see, e.g., Jack et al. 2008 for detail), and the FreeSurfer 4.4 implements the symmetric registration (Reuter, Rosas, and Fischl 2010) and unbiased robust template estimation (Reuter et al. 2012). Only MRIs which passed the quality control for all the areas were included in our study. There are total 393 ROIs of brain MRI created by FreeSurfer 4.4 and they consist of volumes of brain regions obtained after cortical parcellation and white matter parcellation, surface area of the brain regions and cortical thickness of the brain regions. However, some ROIs are missing more than 90% across all samples due to the preprocessing. In Section 3, we use 324 ROIs with at most 20% missing values across the preprocessed samples.

Appendix C: Proof of Theorem 2.1

Denote by $A(b, \beta_1(\cdot) \ldots, \beta_p(\cdot))$ the functional to be minimized in (3). It is clear that $A(b, \beta_1(\cdot), \ldots, \beta_p(\cdot))$ is convex and continuous in $\beta_j(\cdot)$'s. Denote by $J(\beta_1(\cdot), \ldots, \beta_p(\cdot)) = \lambda \sum_{j=1}^p \|\beta_j\|_{\mathcal{H}_K}$, and without loss of generality, we assume $\lambda = 1$. Denote by $c_K = \max_{i,\nu} K^{1/2}(t_{i\nu}, t_{i\nu})$ and $c_x = \max_{j,i,\nu} |x_{ij}(t_{i\nu})|$. By Cauchy–Schwarz inequality, for any $i = 1, \ldots, n$, $\nu = 1, \ldots, m_i$,

$$\begin{aligned}\left|\sum_{j=1}^p \beta_j(t_{i\nu})x_{ij}(t_{i\nu})\right| &= \left|\left\langle \sum_{j=1}^p \beta_j(\cdot)x_{ij}(t_{i\nu}), K(t_{i\nu}, \cdot)\right\rangle_{\mathcal{H}_K}\right| \\ &\le \left\|\sum_{j=1}^p \beta_j(\cdot)x_{ij}(t_{i\nu})\right\|_{\mathcal{H}_K} K(t_{i\nu}, t_{i\nu}) \le c_K \left\|\sum_{j=1}^p \beta_j(\cdot)x_{ij}(t_{i\nu})\right\|_{\mathcal{H}_K} \\ &\le c_K c_x J(b, \ldots, \beta_p).\end{aligned} \tag{C.1}$$

Denote $\rho = \max_{i,\nu}\{y_{i\nu}^2 + |y_{i\nu}| + 1\}$. Consider the set

$$\begin{aligned}\Omega = \{&\beta_1(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K, b \in \mathbb{R} : J(\beta_1(\cdot), \ldots, \beta_p(\cdot)) \le \rho, \\ &|b| \le \rho^{1/2} + (c_K c_x + 1)\rho\}.\end{aligned}$$

Since Ω is closed, convex, and bounded set, there exists a minimizer for (3) in Ω. Denote the minimizer by $\tilde\beta_0, \tilde\beta_1(\cdot), \ldots, \tilde\beta_p(\cdot)$. Then, $A(\tilde\beta_0, \tilde\beta_1(\cdot), \ldots, \tilde\beta_p(\cdot)) \le A(0, 0, \ldots, 0) < \rho$. On the other hand, for any $\beta_1(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K$ satisfying $J(\beta_1(\cdot), \ldots, \beta_p(\cdot)) > \rho$. It is clear that $A(b, \beta_1(\cdot) \ldots, \beta_p(\cdot)) \ge J(\beta_1(\cdot), \ldots, \beta_p(\cdot)) > \rho$. For any $\beta_1(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K$ with $J(\beta_1(\cdot), \ldots, \beta_p(\cdot)) \le \rho$ and $|b| > \rho^{1/2} + (c_K c_x + 1)\rho$, (C.1) implies that for any $i = 1, \ldots, n$, $\nu = 1, \ldots, m_i$,

$$\begin{aligned}&\left|b + \sum_{j=1}^p \beta_j(t_{i\nu})x_{ij}(t_{i\nu}) - y_{i\nu}\right| \\ &\quad > \rho^{1/2} + (c_K c_x + 1)\rho - c_K c_x \rho - \rho = \rho^{1/2}.\end{aligned}$$

Hence, $A(b, \beta_1(\cdot), \ldots, \beta_p(\cdot)) > \rho$. Therefore, for any $b, \beta_1(\cdot), \ldots, \beta_p(\cdot) \notin \Omega$, we have that $A(b, \beta_1(\cdot), \ldots, \beta_p(\cdot)) > A(\tilde\beta_0, \tilde\beta_1(\cdot), \ldots, \tilde\beta_p(\cdot))$, where $\tilde\beta_0, \tilde\beta_1(\cdot), \ldots, \tilde\beta_p(\cdot)$ is the minimizer of (3). We complete the proof.

Appendix D: Algorithm for Solving the Minimizer of (3)

We provide an algorithm for solving the minimizer of (3). The algorithm is based on Theorem 2.2, whose proof will be given later in Appendix 4. Consider for any fixed $\theta_1, \ldots, \theta_p \ge 0$. If $\theta_j = 0$ for some j, then $\beta_j = 0$ in the optimization (4). Without less of generality, let $\theta_1, \ldots, \theta_p > 0$. Then (4) is equivalent to the smoothing spline type problem: find $b \in \mathbb{R}, \beta_1(\cdot), \ldots, \beta_p(\cdot) \in \mathcal{H}_K$ to minimize

$$\frac{1}{N}\sum_{i=1}^n \sum_{\nu=1}^{m_i} [y_{i\nu} - b - \sum_{j=1}^p \beta_j(t_{i\nu})x_{ij}(t_{i\nu})]^2 + \sum_{j=1}^p (\tau_0\theta_j^{-1})\|\beta_j\|_{\mathcal{H}_K}^2. \tag{D.1}$$

By the representer lemma (Wahba 1990), $\beta_1(\cdot), \ldots, \beta_p(\cdot)$ have a closed form expression:

$$\beta_j(t) = \sum_{i=1}^n \sum_{\nu=1}^{m_i} c_{i\nu}^j K(t_{i\nu}, t), \quad \forall j = 1, \ldots, p.$$

Define a $m_{i_1} \times m_{i_2}$ matrix $\Sigma_j^{(i_1,i_2)}$ by

$$\Sigma_j^{(i_1,i_2)} = \begin{pmatrix} x_{i_1 j}(t_{i_1 1})K(t_{i_2 1}, t_{i_1 1}) & \cdots & x_{i_1 j}(t_{i_1 1})K(t_{i_2 m_{i_2}}, t_{i_1 1}) \\ \vdots & & \vdots \\ x_{i_1 j}(t_{i_1 m_{i_1}})K(t_{i_2 1}, t_{i_1 m_{i_1}}) & \cdots & x_{i_1 j}(t_{i_1 m_{i_1}})K(t_{i_2 m_{i_2}}, t_{i_1 m_{i_1}}) \end{pmatrix}$$

and let Σ_j be a $N \times N$ ($N = \sum_{i=1}^n m_i$) matrix where the (i_1, i_2)th $m_{i_1} \times m_{i_2}$ matrix is $\Sigma_j^{(i_1,i_2)}$. Define kernel matrix Σ by

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_2 & \cdots & \Sigma_p \end{pmatrix} \in \mathbb{R}^{N \times N \cdot p}.$$

Let the unknown coefficient vector c^j be

$$c^j = \begin{pmatrix} c_{11}^j & \cdots & c_{1m_1}^j & \cdots & c_{n1}^j & \cdots & c_{nm_n}^j \end{pmatrix}^\top \in \mathbb{R}^N,$$

and

$$c = \begin{pmatrix} \{c^1\}^\top & \{c^2\}^\top & \cdots & \{c^p\}^\top \end{pmatrix}^\top \in \mathbb{R}^{Np}.$$

Write the response vector y as

$$y = \begin{pmatrix} y_{11} & \cdots & y_{1m_1} & \cdots & y_{n1} & \cdots & y_{nm_n} \end{pmatrix}^\top \in \mathbb{R}^N.$$

Let $\mathbf{1}_N$ be the column vector consisting of N 1's. Then (D.1) becomes

$$\frac{1}{N}\left(y - \Sigma c - b\mathbf{1}_N\right)^\top \left(y - \Sigma c - b\mathbf{1}_N\right) + \sum_{j=1}^p (\tau_0\theta_j^{-1})\{c^j\}^\top \Sigma_j c^j,$$

which has the unique solution given as follows:

$$\begin{aligned}\hat b &= [\mathbf{1}_N^\top(\mathbf{1}_{N\times N} - \Sigma\tilde\Sigma^{-1}\Sigma^\top)\mathbf{1}_N]^{-1} \cdot \mathbf{1}_N^\top(\mathbf{1}_{N\times N} - \Sigma\tilde\Sigma^{-1}\Sigma^\top)y, \\ \hat c &= \tilde\Sigma^{-1}\Sigma^\top(y - \mathbf{1}_N\hat b),\end{aligned} \tag{D.2}$$

where $\tilde\Sigma = \Sigma^\top\Sigma + N\mathrm{diag}\{(\tau_0\theta_1^{-1})\Sigma_1, \ldots, (\tau_0\theta_p^{-1})\Sigma_p\}$.

Note that when $\theta_1, \ldots, \theta_p$ are fixed, (4) is equivalent to find $b \in \mathbb{R}, c \in \mathbb{R}^{Np}$ to minimize

$$\begin{aligned}&\frac{1}{N}(y - b\mathbf{1}_N - \sum_{j=0}^p \theta_j\Sigma_j c^j)^\top \cdot (y - b\mathbf{1}_N - \sum_{j=0}^p \theta_j\Sigma_j c^j) \\ &\quad + \sum_{j=0}^p (\tau_0\theta_j)\{c^j\}^\top \Sigma_j c^j.\end{aligned} \tag{D.3}$$

The minimizer of (D.3) is

$$b = \hat b \quad \text{and} \quad c^j = \theta_j^{-1}\hat c^j, \quad j = 0, 1, \ldots, p,$$

where $\hat b$ and $\hat c$ are given by (D.2).

On the other hand, consider when c is fixed, then the minimization of (4) is equivalent to

$$\min_{\theta, b} \|y - \sum_{j=0}^p \theta_j\Sigma_j c^j - b\mathbf{1}_N\|^2 + N\tau_0 \sum_{j=0}^p \theta_j\{c^j\}^\top\Sigma_j c^j + N\tau_1 \sum_{j=0}^p \theta_j,$$
$$\text{s.t. } \theta_j \ge 0, j = 0, 1, \ldots, p,$$

which can be written as

$$\min_{\theta, b} \|y - \sum_{j=0}^p \theta_j\Sigma_j c^j - b\mathbf{1}_N\|^2 + N\tau_0 \sum_{j=0}^p \theta_j\{c^j\}^\top\Sigma_j c^j,$$
$$\text{s.t. } \theta_j \ge 0, j = 0, 1, \ldots, p; \sum_{j=0}^p \theta_j \le M, \tag{D.4}$$

for some $M \ge 0$.

Therefore, we propose an algorithm to iterate (D.3) and (D.4) to give the minimizer of (4). We observe in experiments that the objective function in optimization (4) decreases quickly in the first iteration and after the first iteration the objective function is close to the objective function at convergence. It motivates us to consider the following one-step update algorithm:

1. Initialization: fix $\theta_j = 1$ for $j = 0, 1, \ldots, p$.
2. Solve for c and b in (D.3) and tune τ_0 according to the generalized cross-validation (GCV). Fix τ_0 at the chosen value in all later steps.

3. For c and b obtained in Step 2, solve for θ in (D.4) with a fixed M.
4. With θ obtained in Step 3, solve for c and b in (D.3).

We choose the best M in Step 3 according to the 5-fold cross-validation in the experiments.

Appendix E: Proof of Theorem 2.2

Recall that we introduce $A(b, \beta_1(\cdot), \ldots, \beta_p(\cdot))$ in Appendix C to denote the functional to be minimized in (3). Let $B(\theta_1, \ldots, \theta_p; b, \beta_1(\cdot), \ldots, \beta_p(\cdot))$ be the functional in (4). Note

$$\tau_0\theta_j^{-1}\|\beta_j\|_{\mathcal{H}_K}^2 + \tau_1\theta_j \geq 2\tau_0^{1/2}\tau_1^{1/2}\|\beta_j\|_{\mathcal{H}_K} = \lambda^2\|\beta_j\|_{\mathcal{H}_K},\ \forall\theta_j \geq 0,$$

and the equality in the above formula holds if and only if $\theta_j = \tau_0^{1/2}\tau_1^{-1/2}\|\beta_j\|_{\mathcal{H}_K}$. Therefore,

$$B(\theta_1, \ldots, \theta_p; b, \beta_1(\cdot), \ldots, \beta_p(\cdot)) \geq A(b, \beta_1(\cdot), \ldots, \beta_p(\cdot)),\ \ \forall\theta_j \geq 0,$$

and the equality holds if and only if $\theta_j = \tau_0^{1/2}\tau_1^{-1/2}\|\beta_j\|_{\mathcal{H}_K}$ for all $j = 1, \ldots, p$. We complete the proof.

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (*www.adni.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*. The author would like to thank the editor, the associate editor, and two referees for their helpful and constructive comments. The author would also like to express warm thanks to Grace Wahba for insightful comments.

Funding

ADNI (National Institutes of Health Grant U01 AG024904 and Department of Defense award number W81XWH-12-2-0012) is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai, Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development, LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (*www.fnih.org*). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California. The research was supported in part by NSF grant DMS-1308877.

References

Aguilar, C., Westman, E., Muehlboeck, J. S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., and Simmons, A. (2013), "Different Multivariate Techniques for Automated Classification of MRI Data in Alzheimer's Disease and Mild Cognitive Impairment," *Psychiatry Research: Neuroimaging*, 212, 89–98. [244]

Chincarini, A., Sensi, F., Rei, L., Gemme, G., Squarcia, S., Longo, R., Brun, F., Tangaro, S., Bellotti, R., Amoroso, N., and Bocchetta, M. (2016), "Integrating Longitudinal Information in Hippocampal Volume Measurements for the Early Detection of Alzheimer's Disease," *NeuroImage*, 125, 834–847. [244,245,248]

Dai, X. (2018), "High-Dimensional Varying Coefficient Models for Alzheimer's Disease Diagnosis With Longitudinal and Heterogeneous Structural MR Images," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 14, 1446. [249]

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Series B, 55, 757–779. [244,245]

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., and Dale, A. M. (2008), "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," *Journal of Magnetic Resonance Imaging*, 27, 685–691. [250]

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., and Lesnick, T. G. (2013), "Tracking Pathophysiological Processes in Alzheimer's Disease: An Updated Hypothetical Model of Dynamic Biomarkers," *The Lancet Neurology*, 12, 207–216. [244,245,247]

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., and Trojanowski, J. Q. (2010), "Hypothetical Model of Dynamic Biomarkers of the Alzheimer's Pathological Cascade," *The Lancet Neurology*, 9, 119–128. [244,245,247]

Jack, C. R., Lowe, V. J., Weigand, S. D., Wiste, H. J., Senjem, M. L., Knopman, D. S., Shiung, M. M., Gunter, J. L., Boeve, B. F., Kemp, B. J., and Weiner, M. (2009), "Serial PIB and MRI in Normal, Mild Cognitive Impairment and Alzheimer's Disease: Implications for Sequence of Pathological Events in Alzheimer's Disease," *Brain*, 132, 1355–1365. [244]

Lin, K. A., Choudhury, K. R., Rathakrishnan, B. G., Marks, D. M., Petrella, J. R., and Doraiswamy, P. M. (2015), "Marked Gender Differences in Progression of Mild Cognitive Impairment Over 8 Years," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 1, 103–110. [247]

Liu, M., Zhang, D., and Shen, D. (2016), "Relationship Induced Multi-Template Learning for Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment," *IEEE Transactions on Medical Imaging*, 35, 1463–1474. [244]

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984), "Clinical Diagnosis of Alzheimer's Disease Report of the NINCDS-ADRDA Work Group Under the Auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, 34, 939–939. [249]

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., and Xian, H. (2009), "Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness," *Cerebral Cortex*, 19, 2728–2735. [247]

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., and Ferri, C. P. (2013), "The Global Prevalence of Dementia: A Systematic Review and Metaanalysis," *Alzheimer's & Dementia*, 9, 63–75. [244]

Reuter, M., Rosas, H. D., and Fischl, B. (2010), "Highly Accurate Inverse Consistent Registration: A Robust Approach," *Neuroimage*, 53, 1181–1196. [250]

Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012), "Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis," *Neuroimage*, 61, 1402–1418. [250]

Sabuncu, M. R., Desikan, R. S., Sepulcre, J., Yeo, B. T. T., Liu, H., Schmansky, N. J., Reuter, M., Weiner, M. W., Buckner, R. L., Sperling, R. A., and Fischl, B. (2011), "The Dynamics of Cortical and Hippocampal Atrophy in Alzheimer Disease," *Archives of Neurology*, 68, 1040–1048. [248]

Schuff, N., Tosun, D., Insel, P. S., Chiang, G. C., Truran, D., Aisen, P. S., Jack, C. R., and Weiner, M. W. (2012), "Nonlinear Time Course of Brain Volume Loss in Cognitively Normal and Impaired Elders," *Neurobiology of Aging*, 33, 845–855. [248]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [244,246]

Tognin, S., Riecher-Rössler, A., Meisenzahl, E. M., Wood, S. J., Hutton, C., Borgwardt, S. J., Koutsouleris, N., Yung, A. R., Allen, P., Phillips, L. J., and McGorry, P. D. (2014), "Reduced Parahippocampal Cortical Thickness in Subjects at Ultra-High Risk for Psychosis," *Psychological Medicine*, 44, 489–498. [247]

Toledo, J. B., Da, X., Weiner, M. W., Wolk, D. A., Xie, S. X., Arnold, S. E., Davatzikos, C., Shaw, L. M., and Trojanowski, J. Q. (2014), "CSF Apo-E Levels Associate With Cognitive Decline and MRI Changes," *Acta Neuropathologica*, 127, 621–632. [249]

Tosun, D., Schuff, N., Shaw, L. M., Trojanowski, J. Q., and Weiner, M. W. (2011), "Relationship Between CSF Biomarkers of Alzheimer's Disease and Rates of Regional Cortical Thinning in ADNI Data," *Journal of Alzheimer's Disease*, 26, 77–90. [249]

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002), "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *Neuroimage*, 15, 273–289. [244]

Velayudhan, L., Proitsi, P., Westman, E., Muehlboeck, J., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Spenger, C., and Hodges, A. (2013), "Entorhinal Cortex Thickness Predicts Cognitive Decline in Alzheimer's Disease," *Journal of Alzheimer's Disease*, 33, 755–766. [248]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia, PA: SIAM. [244,245,246,250]

Yau, W.-Y. W., Tudorascu, D. L., McDade, E. M., Ikonomovic, S., James, J. A., Minhas, D., Mowrey, W., Sheu, L. K., Snitz, B. E., Weissfeld, L., and Gianaros, P. J. (2015), "Longitudinal Assessment of Neuroimaging and Clinical Markers in Autosomal Dominant Alzheimer's Disease: A Prospective Cohort Study," *The Lancet Neurology*, 14, 804–813. [244,245]

Yuan, M., and Lin, Y. (2006), "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society*, Series B, 68, 49–67. [248]

Zhang, D., and Shen, D. (2012), "Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers," *PLoS One*, 7, e33182. [244,245,246,248]