Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2023, Vol. 32(2) 353–372

DOI: 10.1177/09622802221137746
journals.sagepub.com/home/smm
SAGE

Additive hazards model with time-varying coefficients and imaging predictors

Qi Yang[1,*], Chuchu Wang[2,*], Haijin He[3], Xiaoxiao Zhou[2], and Xinyuan Song[2]

Abstract

Conventional hazard regression analyses frequently assume constant regression coefficients and scalar covariates. However, some covariate effects may vary with time. Moreover, medical imaging has become an increasingly important tool in screening, diagnosis, and prognosis of various diseases, given its information visualization and quantitative assessment. This study considers an additive hazards model with time-varying coefficients and imaging predictors to examine the dynamic effects of potential scalar and imaging risk factors for the failure of interest. We develop a two-stage approach that comprises the high-dimensional functional principal component analysis technique in the first stage and the counting process-based estimating equation approach in the second stage. In addition, we construct the pointwise confidence intervals for the proposed estimators and provide a significance test for the effects of scalar and imaging covariates. Simulation studies demonstrate the satisfactory performance of the proposed method. An application to the Alzheimer's disease neuroimaging initiative study further illustrates the utility of the methodology.

Keywords

Functional principal component analysis, estimating equation, imaging data, survival analysis, time-varying coefficients

1 Introduction

One popular survival model in analyzing censored survival data is the additive hazards (AH) model.[1–3] Like other survival models, the AH model has wide applications in various fields, such as biomedical science,[4–7] environmental study,[8] materials research,[9] and finance.[10,11] Moreover, compared with its common alternative, namely, the proportional hazards (PH) or Cox model,[12–15] the AH model characterizes the effects of potential covariates on the hazards of interest differently and has remarkable features. For example, the linear form makes the AH model not so vulnerable to the consistency problems that the PH model frequently confronts.[1] In addition, it interprets covariate effects more intuitively than its PH counterpart and pertains to the risk difference, which is particularly relevant to and informative in epidemiological and clinical studies.[16] However, conventional AH models typically assume constant covariate effects.

The constant-coefficient AH models may be unrealistic in some circumstances. For instance, in developing a new drug targeting the treatment of acquired immunodeficiency syndrome (AIDS), the drug effect can be significant in the beginning stage but gradually disappear later on due to virus mutation or increasing drug resistance. If a drug eventually loses efficacy, we are particularly interested in detecting when and how fast the drug becomes ineffective. A time-varying coefficient model can uncover such dynamic treatment efficacy and aid medical scientists in designing future experiments and

[1]School of Management, Shandong University, Jinan, China
[2]Department of Statistics, Chinese University of Hong Kong, Hong Kong, China
[3]College of Mathematics and Statistics, Shenzhen University, Shenzhen, China

*Qi Yang and Chuchu Wang are the joint first author of this article.

Corresponding author:
Xinyuan Song, Department of Statistics, Chinese University of Hong Kong, Hong Kong, China.
Email: xysong@sta.cuhk.edu.hk.

effective treatment strategies for AIDS prevention. Many existing works have investigated time-varying coefficient AH models under various application contexts. For example, after Aalen[1] first proposed the time-varying coefficient AH model, Mckeague[18] adopted the sieve method to estimate Aalen's additive risk model, with certain assumptions on L^2-norm to guarantee consistency. Huffer and Mckeague[2] developed a weighted estimator for the time-varying coefficient functions and established the corresponding asymptotic theory. Mckeague and Sasieni[19] considered a highly flexible and partly parametric AH model that simultaneously accommodates time-invariant and time-varying covariate effects. Li et al.[20] adopted the local linear approximation technique proposed in Cai and Sun[21] to investigate Aalen's additive risk model. They estimated the functional coefficients with the expectation–maximization (EM) algorithm and developed the asymptotic properties with explicit expressions for the variance and bias. Yang et al.[22] recently considered the time-varying coefficient AH model in the presence of latent variables and developed a hybrid procedure that combined the EM algorithm and corrected estimating equation method to obtain consistent estimators. Nevertheless, the preceding works only considered scalar covariates.

Imaging data are a kind of functional data, with the whole voxel space and intensity at each voxel as the domain and observation, respectively. The existing literature has extensively investigated the functional linear model and its variants.[23–29] In particular, Yao[25] considered longitudinal (functional) data under the framework of survival analysis. He adopted B-splines to approximate the functional covariates and used the functional principal component analysis (FPCA) for dimension reduction of the model. Zhu et al.[26] considered a spatially varying coefficient model for neuroimaging data in the presence of discontinuous jumps. They accounted for the spatial correlation structure of raw images and proposed a three-stage inference procedure to estimate the nonparametric coefficients and approximate the asymptotic covariance matrix. Feng et al.[29] conducted a scalar-on-image regression with nonignorable missing responses. Kang and Song[30] developed a joint modeling approach for longitudinal imaging and survival data.

Despite the broad applications of imaging data in diagnosis, prognosis, and other clinical practices, incorporating high-dimensional images into regression analysis faces challenges. The first and foremost challenge is the ultrahigh dimensionality of the imaging data. For example, the magnetic resonance imaging (MRI) data considered in the current study involve a matrix of over one million voxels, and its dimension is much higher than the sample size. Therefore, treating each image voxel as an independent covariate in regression analysis is computationally infeasible. In addition, the time-varying coefficients of the imaging covariate are also voxelwise. However, there is no available method to handle such time-varying coefficient images in the survival analysis literature. This study proposes a novel AH model to analyze time-to-event responses with imaging predictors. The model allows for scalar and imaging covariates, time-varying coefficients, and an unknown baseline hazard function. Existing methods for analyzing AH models are inapplicable to the present study that incorporates imaging predictors and nonparametric coefficients. Therefore, we develop a hybrid procedure by combining the FPCA method and counting process-based estimating equations for regression analysis. Our method is a two-stage approach. The first stage extracts the desired imaging features through FPCA. The second stage regards the extracted eigenscores as ordinary covariates and estimates the nonparametric coefficient functions using the estimating equation approach. The estimated time-varying coefficients for eigenscores are then utilized to reconstruct the coefficient images at different time points. Moreover, theoretical results, including the pointwise confidence intervals (CIs) of the proposed estimators and a significance test for scalar and imaging covariates, are provided.

The current research is motivated by a dataset derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI). This project was launched in 2004, aiming to collect various features, such as imaging and clinical variables, from cognitively normal controls and subjects with mild cognitive impairment or Alzheimer's disease (AD). ADNI recruited participants aged 55 to 90 and provided medical services such as physical and cognitive examinations. In particular, the MRI data, represented by a three-dimensional (3D) intensity matrix, were scanned from each participant. Participants were diagnosed with AD in the cohort study provided that they met the specific inclusion criteria shown in the general procedure manual. More information about the ADNI data is available at its official website (www.adni-info.org). The primary interest of this study is to investigate whether the observed clinical and imaging variables affect the occurrence of AD, which is the event of interest. Similar to the decreasing drug effects on AIDS prevention illustrated above, the potential scalar and imaging risk factors of AD may have pronounced effects at the initial stage of AD progression. However, these effects may fade out with time. Therefore, the proposed time-varying coefficient model with imaging data can uncover this dynamic feature.

The remainder of this article is organized as follows. Section 2, describes the proposed time-varying coefficient AH model with imaging covariates and the inference procedures, including FPCA and the counting process-based estimation equation approach. Section 3, constructs pointwise CIs for the proposed estimators and provides a significance test for the effects of scalar and imaging covariates. Section 4, presents simulation studies to assess the finite sample performance of the proposed method. Section 5, illustrates the application to the ADNI dataset to reveal the temporal effects of clinical variables and brain images. Section 6, concludes the paper and discusses several promising extensions of the current research.

2 Model and estimation

2.1 Model description

Let $\mathbf{X}_i$ denote an image sample for the ith ($i = 1, \ldots, n$) subject, where n is the total number of subjects. The MRI data for subjects are collected in a 3D matrix structure of dimension $V = v_x \times v_y \times v_z$ with voxels being the same among all the subjects. These images are unfolded into $V \times 1$ vectors in a specific order that can be preserved across all the subjects, namely, $\mathbf{X}_i = (X_i(1), \ldots, X_i(V))^{\mathrm{T}}$. Following the existing literature,[31] we assume that $X_i(v)$, $v \in 1, \ldots, V$ are zero-mean, square-integrable, and mutually independent stochastic processes on a compact space $\mathcal{V}$, where v denotes the vth voxel.

We use T_i, C_i, $\widetilde{T}_i = \min(T_i, C_i)$, and $\Delta_i = I(T_i \leq C_i)$ to denote the failure time of interest, censoring time, observed time, and censoring indicator, respectively. The time-varying coefficient AH model with imaging predictors can be formulated as follows:

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) = \beta(t)^{\mathrm{T}}\mathbf{Z}_i + \int_{\mathcal{V}} \gamma(v, t)^{\mathrm{T}} X_i(v)\mathrm{d}v, \tag{1}$$

where $\lambda(t|\cdot)$ is the hazard function of T_i, $\mathbf{Z}_i = (1, Z_{i1}, \ldots, Z_{is})$ is the $(s+1) \times 1$ vector of observed covariates. $\beta(t)$ is the $(s+1) \times 1$ vector of unknown time-varying coefficients, and $\gamma(v, t)$ is the time-varying functional coefficient at the vth voxel of the image. An intuitive interpretation for $\gamma(v, t)$ is that the regions of the image with large $|\gamma(v, t)|$ have strong effects on the hazard function. By fixing the first element of $\mathbf{Z}_i$ to 1, the unspecified baseline hazard function is included as the first element of $\beta(t)$. Thus, separately estimating the baseline hazard function is unnecessary for the proposed model.

Let $C_X(v_1, v_2) = E\{X_i(v_1)X_i(v_2)\}$ be the covariance operator of $X_i(v)$. To handle the imaging integration item $\int_{\mathcal{V}} \gamma(v, t)^{\mathrm{T}} X_i(v)\mathrm{d}v$ in Model (1), we adopt the Karhunen–Loeve (KL) expansion of the stochastic process based on the eigendecomposition of the covariance operator, which gives

$$X_i(v) = \sum_{m=1}^{\infty} \xi_{im}\psi_m(v), \tag{2}$$

where $\{\psi_m(v), m = 1, 2, \ldots\}$ denotes the eigenimage basis of $C_X(v_1, v_2)$ for the image data, ξ_{im}'s represent mutually independent eigenscores for the ith subject with decreasing variances σ_m^2, and σ_m^2's are the eigenvalues of $C_X(v_1, v_2)$. Notably, the eigenimages $\psi_m(v)$'s are orthogonal to each other, and thus $\int_v \psi_m(v)\psi_l(v)\mathrm{d}v = 1$ if $m = l$ and 0 if $m \neq l$. In practice, retaining excessive eigenimages in regression analysis may result in model overfitting. Previous studies[32,26,29] suggested that the first few eigenimages are sufficient to account for the major functional variability of $X_i(v)$. Therefore, a criterion-based approach is typically used to determine an integer M, such that the first M eigenimage can explain most of $X_i(v)$'s variation.

We assume that $\gamma(v, t)$ can also be expanded on the same eigenimage basis with time-varying coefficients $\{\gamma_m(t), m = 1, 2, \ldots\}$ as follows:

$$\gamma(v, t) = \sum_{m=1}^{\infty} \gamma_m(t)\psi_m(v). \tag{3}$$

Suppose that the eigenimages and eigenscores for the imaging data are available (see the derivation in Section 2.2.). The truncated KL expansions of $X_i(v)$ and $\gamma(v, t)$ are $X_i(v) = \sum_{m=1}^{M} \xi_{im}\psi_m(v)$ and $\gamma(v, t) = \sum_{m=1}^{M} \gamma_m(t)\psi_m(v)$, respectively. Then, model (1) can be rewritten as

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) \approx \beta(t)^{\mathrm{T}}\mathbf{Z}_i + \gamma(t)^{\mathrm{T}}\xi_i, \tag{4}$$

where $\gamma(t) = (\gamma_1(t), \ldots, \gamma_M(t))^{\mathrm{T}}$ is a vector of time-varying eigenbasis coefficients, and $\xi_i = (\xi_{i1}, \ldots, \xi_{iM})^{\mathrm{T}}$ is a vector of eigenscores of $X_i(v)$.

2.2 FPCA for dimension reduction of imaging data

It is challenging to directly eigendecompose the imaging data $X_i(v)$ owing to its ultrahigh dimensionality. For instance, 3D imaging data with 128 grids on each dimension results in a vectorized image $\mathbf{X}_i$ with $V = 128^3 = 2,097,152$. The corresponding covariance operator $C_X(v_1, v_2)$ will be of dimension $V \times V$. A direct eigendecomposition of this operator requires $O(V^3)$ operations, which is practically infeasible. Zipunnikov et al.[32] proposed the FPCA technique for high-dimensional data based on the singular value decomposition (SVD) to tackle this difficulty. The computational burden of this method

is much smaller compared with a direct eigendecomposition. Suppose we consider an imaging design matrix $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of dimension $V \times n$ with a rank of at most $\min(n, V)$. The SVD of the imaging design matrix is as follows:

$$\mathbf{X} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathrm{T}}, \tag{5}$$

where $\mathbf{U}$ is the $V \times n$ matrix of the orthonormal eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\mathbf{V}$ is the $n \times n$ matrix containing the orthonormal eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, and $\mathbf{W}$ is the $n \times n$ diagonal matrix of the singular values which are the square roots of the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Note that the diagonal elements of $\mathbf{W}$ are nonincreasing and nonnegative, indicating that the functional variability of these images can be represented by a few leading eigenimages. The computational cost to perform SVD is approximately $O(Vn^2 + n^3)$. The outcomes of interest, such as the eigenimages $\psi_m(v)$, eigenvalues c_m of C_X, and eigenscores ξ_{im} of the subjects, can be obtained as follows. We first conduct spectral decomposition of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ as $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{V}\mathbf{W}^2\mathbf{V}$. Then, the orthonormal matrix $\mathbf{U}$ can be calculated as $\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{W}^{-1}$, and the eigenimage $\psi_m = \{\psi_m(v), v \in \mathcal{V}\}$ is the mth column of $\mathbf{U}$. The eigenvalue $c_m = w_m^2$, where w_m is the mth diagonal element of $\mathbf{W}$. Finally, the eigenscores ξ_{im}s are specified by the columns of $\mathbf{W}\mathbf{V}^{\mathrm{T}}$ truncated to the first M coordinates.

Another challenging issue related to the ultrahigh dimension of MRI images is its loading into the computer memory. To overcome this difficulty, the MRI images can be sliced into many small pieces as $\mathbf{X} = [(\mathbf{X}^{(1)})^{\mathrm{T}}|(\mathbf{X}^{(2)})^{\mathrm{T}}|\cdots|(\mathbf{X}^{(J)})^{\mathrm{T}}]^{\mathrm{T}}$, where the number of pieces, J, can be determined to make each piece adapt to the computer memory. Then, we can calculate $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{U}$ as $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \sum_{j=1}^{J}(\mathbf{X}^{(j)})^{\mathrm{T}}(\mathbf{X}^{(j)})$, $\mathbf{U}^{(j)} = \mathbf{X}^{(j)}\mathbf{V}\mathbf{W}^{-1}$, and $\mathbf{U} = [(\mathbf{U}^{(1)})^{\mathrm{T}}|(\mathbf{U}^{(2)})^{\mathrm{T}}|\cdots|(\mathbf{U}^{(J)})^{\mathrm{T}}]^{\mathrm{T}}$.

2.3 Estimation procedure

We use $N_i(t) = 1(\widetilde{T}_i \le t, \Delta_i = 1)$ to denote the counting process, and $Y_i(t) = 1(\widetilde{T}_i \ge t)$, $i = 1, \ldots, n$ to denote the at-risk process. For brevity, the following notations are introduced for estimation. We define

$$\alpha(t) = \begin{pmatrix} \beta(t) \\ \gamma(t) \end{pmatrix}, \quad \mathbf{d}_i = \begin{pmatrix} \mathbf{Z}_i \\ \xi_i \end{pmatrix}, \quad \mathbf{A}(t) = \int_0^t \alpha(s)\mathrm{d}s,$$

where $\mathbf{A}(t)$ denotes the cumulative functional coefficient vector. With these notations, model (4) can be simplified as

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) = \alpha(t)^{\mathrm{T}}\mathbf{d}_i. \tag{6}$$

Then, $\alpha(t)$ can be estimated using the estimating equation method[1,3] as follows:

$$\sum_{i=1}^{n}\mathbf{d}_i\mathrm{d}N_i(t) - \sum_{i=1}^{n}Y_i(t)\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\alpha(t)\mathrm{d}t = 0. \tag{7}$$

Accordingly, simple algebraic calculation gives the estimator $\hat{\alpha}(t)$ as

$$\hat{\alpha}(t)\mathrm{d}t = \left[\sum_{i=1}^{n}Y_i(t)\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\right]^{-1}\left(\sum_{j=1}^{n}\mathbf{d}_j\mathrm{d}N_j(t)\right),$$

which is equivalent to

$$\hat{\mathbf{A}}(t) = \int_0^t\left[\sum_{i=1}^{n}Y_i(s)\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\right]^{-1}\left(\sum_{j=1}^{n}\mathbf{d}_j\mathrm{d}N_j(s)\right) = \sum_{j=1}^{n}1\{\widetilde{T}_j \le t\}\left[\sum_{i=1}^{n}Y_i(\widetilde{T}_j)\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\right]^{-1}\mathbf{d}_j\Delta_j. \tag{8}$$

As we have included 1 as the first item of $\mathbf{d}_i$, the cumulative baseline hazard function is simultaneously estimated as the first element of $\hat{\mathbf{A}}(t)$. For any $j \ge 1$, $A_j(t)$ can be regarded as an empirical function revealing the influence of covariate j on the hazard function, wherein the jth covariate can be a scalar variable or an eigenscore of the image data. When plotting $A_j(t)$ against time, its slope provides information about the jth covariate effect, $\alpha_j(t)$. For example, if $\hat{A}_j(t)$ is nearly a horizontal line in a period, then $\hat{\alpha}_j(t)$ is approximately zero in the same period.

Based on the above estimation, one may evaluate the cumulative hazard and survival functions for given covariates. The estimators of the cumulative hazard function $\Lambda(t)$ and survival function $S(t)$ can be formulated as

$$\hat{\Lambda}(t) = \hat{\mathbf{A}}^{\mathrm{T}}(t)\mathbf{Z} \quad \text{and} \quad \hat{S}(t) = \exp(-\hat{\Lambda}(t)).$$

As suggested by the existing literature,[33,2,21,4] the kernel smoothing method can be adopted to estimate the effects of

covariates, $\alpha(t)$. The corresponding estimator, denoted as $\hat{\alpha}_{Ker}(t)$, is calculated as follows:

$$\hat{\alpha}_{Ker}(t) = \frac{1}{h}\int_0^{\mathrm{T}} Ker\Big(\frac{t-s}{h}\Big)\mathrm{d}\hat{\mathbf{A}}(s), \tag{9}$$

where $Ker(\cdot)$ is the kernel function with bounded variation, integral 1, and support $(\epsilon, 1]$ for some $0 < \epsilon < 1$, and h is the bandwidth that may require empirical selection. Theoretically, a few conditions are required on h to guarantee the consistency of $\hat{\alpha}_{Ker}(t)$, which states that $n \to \infty$, $h \to 0$, and $nh \to \infty$. More details are available in Chapter 4 of Martinussen and Scheike[4]. Practically, the kernel estimator $\hat{\alpha}_{Ker}(t)$ is usually not sensitive to the functional form of $Ker(\cdot)$ but depends heavily on the bandwidth h. For simplicity, we impose an elementary form, namely, a constant, for the kernel function $Ker(\cdot)$ over $(\epsilon, 1]$.[2] The elementary kernel function takes the form of

$$Ker(x) = \frac{1}{1-\epsilon} I\{x \in (\epsilon, 1]\}, \tag{10}$$

where $I\{\cdot\}$ is an indicator function to illustrate the support for $Ker(x)$, and ϵ is a small positive value close to 0 (e.g. 0.001). This elementary form can simplify the associated inference process and facilitate efficient computation while maintaining satisfactory performance in uncovering the coefficient functions. Therefore, we use (10) for computationally intensive bandwidth selection in the application to the ADNI study (see Section 5.).

3 Theoretical results

3.1 Pointwise CI

For the truncated time-varying coefficient model (4) with a fixed truncation level M, the estimation of the covariance matrices for the elements of $\hat{\mathbf{A}}(t)$ can be easily derived with the theory of martingales and the counting process.[1,19] The theory shows that $\hat{\mathbf{A}}(t)$ is consistent with its true value, denoted by $\mathbf{A}_\star(t)$, and $n^{1/2}[\hat{\mathbf{A}}(t) - \mathbf{A}_\star(t)]$ converges weakly to a Gaussian process with zero mean and covariance function $E[\varphi(t)\varphi(t)^{\mathrm{T}}]$ at time t, which can be consistently estimated by

$$\hat{\mathbf{O}}(t) = \frac{1}{n}\sum_{i=1}^{n} \hat{\varphi}_i(t)\hat{\varphi}_i^{\mathrm{T}}(t),$$

where

$$\hat{\varphi}_i(t) = 1\{\widetilde{T}_i \le t\}\left[\sum_{j=1}^{n} Y_j(\widetilde{T}_i)\mathbf{d}_j\mathbf{d}_j^{\mathrm{T}}\right]^{-1}\mathbf{d}_i\Delta_i.$$

The variance estimation for $\hat{\Lambda}(t)$ and $\hat{S}(t)$ can also be obtained straightforwardly. Given that $\hat{\mathbf{A}}(t)$ is asymptotically Gaussian distributed with a known covariance process, the distribution of $f(\hat{\mathbf{A}}(t))$ can be easily derived through some basic calculations in probability theory for a specific function $f(\cdot)$. As pointed out by Aalen,[1] this covariance (or Gaussian) process has independent increments under certain assumptions, and the corresponding asymptotic distribution is uniquely determined by the covariance function $\hat{\mathbf{O}}(t)$ at all times.

With the above asymptotic result for $\hat{\mathbf{A}}(t)$, the pointwise CIs for the cumulative regression coefficients $\{A_j(t), j = 1, \ldots, 1+s+q\}$ with the confidence level of $100(1-p)\%$ can be constructed as follows:

$$\hat{A}_j(t) \pm z_{p/2}\hat{O}_{jj}(t)^{1/2}, \quad t \in [0, T], \tag{11}$$

where $z_{p/2}$ is the upper $p/2$ quantile point of the standard normal distribution, $\hat{O}_{jj}(t)$ is the jth diagonal element of $\hat{\mathbf{O}}(t)$, and T is the maximum observed failure time. Notably, the construction of CIs in (11) is in line with that of the least square type estimator proposed in Section 2. Hence, the CIs constructed above may not be the most efficient compared with the weighted least square type estimator-based CIs.[2]

Notably, the above results are developed when the truncation number M is fixed. For the case where M depends on n, Kong et al.[34] proposed three sets of conditions in their supplementary material to establish the asymptotic results under the framework of the functional linear Cox model. However, difficulties exist in extending their proof to the current model framework. First, although their conditions (C1) to (C3) for deriving the diverging speed of the truncation number M_n can be similarly used in our model, their conditions (A1) to (A4) and (B2) are inapplicable due to the formulation of time-varying coefficients of the proposed model. Suitable conditions merit future exploration. Second, while the proof of Kong et al.[34] is quite

technically involved, the time-varying coefficients in the proposed model pose additional challenges in deriving results similar to those of Kong et al.[34] (e.g. Lemmas 3 and 5). Therefore, the feasibility of establishing the theoretical results when $M_n \to \infty$ under the proposed model framework is uncertain. We hope to leave this important but challenging problem to future research.

3.2 Hypothesis testing

We propose an empirical method of hypothesis testing procedure to determine whether a specific covariate has any significant effect on the hazards of interest. We consider the following null hypothesis:

$$H_j: \quad \alpha_j(t) = 0, \quad \text{for all} \quad \mathrm{t} \in [0, \mathrm{T}].$$

Aalen[1] proposed a test statistic by calculating a weighted sum of the cumulative coefficient increments with a predictable weight function. This statistic is defined as follows:

$$\mathbf{U} = \sum_{j=1}^{n} \mathbf{K}(\widetilde{T}_j) \left[\sum_{i=1}^{n} Y_i(\widetilde{T}_j) \mathbf{d}_i \mathbf{d}_i^{\mathrm{T}} \right] \mathbf{d}_j \Delta_j,$$

where $\mathbf{K}(t)$ is a weight function in the form of

$$\mathbf{K}(t) = \left\{ \mathrm{diag} \left[\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Y_i(t)} \mathbf{d_i} \mathbf{d_i^{\mathrm{T}}} \right)^{-1} \right] \right\}^{-1},$$

in which diag($\mathbf{B}$) is a diagonal matrix that keeps the same diagonal elements of $\mathbf{B}$. The jth element of $\mathbf{U}$, U_j, is associated with the null hypothesis H_j.

The covariance matrix for the test statistic $\mathbf{U}$ can be obtained through some simple calculations, which yield

$$\mathbf{L} = \sum_{j=1}^{n} \left\{ \mathbf{K}(\tilde{T}_j) \left[\sum_{i=1}^{n} Y_i(\widetilde{T}_j) \mathbf{d}_i \mathbf{d}_i^{\mathrm{T}} \right] \mathbf{d}_j \Delta_j \right\}^{\otimes 2},$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ for any vector or matrix $\mathbf{a}$.

Let $\mathbf{k}$ be a subset of $\{1, \ldots, 1 + s + q\}$ containing r elements, $\mathbf{U_k}$ be the associated subvector of $\mathbf{U}$ and $\mathbf{L_k}$ be the submatrix of $\mathbf{L}$ associated with the estimated covariance matrix of $\mathbf{U_k}$. After an operation of regularization, $\mathbf{U_k^{\mathrm{T}}} \mathbf{L_k^{-1}} \mathbf{U_k}$ follows a chi-square distribution with r degrees of freedom, conditional on that H_j holds for all $j \in \{1, \ldots, 1 + s + q\}$. Consequently, if we aim to test the significance of a specific covariate effect, then $U_j L_{jj}^{-\frac{1}{2}}$ can be used as the test statistic and compared with the specified quantile of the standard normal distribution. Alternatively, the significance of a covariate effect can be concluded based on the pointwise CI, which suggests that a covariate exhibit a significant effect if its pointwise CI does not include a horizontal line (e.g. Figures 1 and 2).

4 Simulation study

4.1 Simulation I

We consider a two-dimensional (2D) $V_1 \times V_2$ monochrome image with $V_1 = V_2 = 300$, resulting in $\mathbf{X}_i$ of length $V = 90,000$. The simulated data are generated from the following model with $s = 1$ and $M = 3$:

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) = \beta_1(t) Z_{i1} + \beta_2(t) Z_{i2} + \int_{\mathcal{V}} \gamma(v, t) X_i(v) \mathrm{d}v, \tag{12}$$

with $X_i(v) = \sum_{m=1}^{3} \xi_{im} \psi_m(v)$ and $\gamma(v, t) = \sum_{m=1}^{3} \gamma_m(t) \psi_m(v)$, where $\xi_{im} \sim N(0, 0.5^{m-1})$, and $\alpha(t) = (\beta(t)^{\mathrm{T}}, \gamma(t)^{\mathrm{T}})^{\mathrm{T}} = (\beta_1(t), \beta_2(t), \gamma_1(t), \gamma_2(t), \gamma_3(t))^{\mathrm{T}} = (2t + 1, 0.75\sqrt{t}, 0.5, -t, 0.5/(1 + t))^{\mathrm{T}}$. Accordingly, $\mathbf{A}(t) = (t^2 + t, 0.5t^{1.5}, 0.5t, -0.5t^2, 0.5\log(1 + t))^{\mathrm{T}}$. Covariate $Z_{i1} = 1$, and Z_{i2} is independently generated from the binomial distribution *Binomial*(3, 0.5). The failure time T_i is generated from (12). The censoring time C_i is independently generated from the uniform distribution $U(0, c)$, where c is set to 3 and 1.5 to obtain censoring rates (CRs) of 15% and 30%, respectively. Two sample sizes $n = 500$ and $n = 1000$ are considered to investigate the finite sample performance of the proposed estimators. The image data generating method follows the previous studies,[32,31] which treated eigenimages as 2D grayscale images with pixel intensities on the [0, 1] scale, where the white and black pixels are set to 0 and 1, respectively.

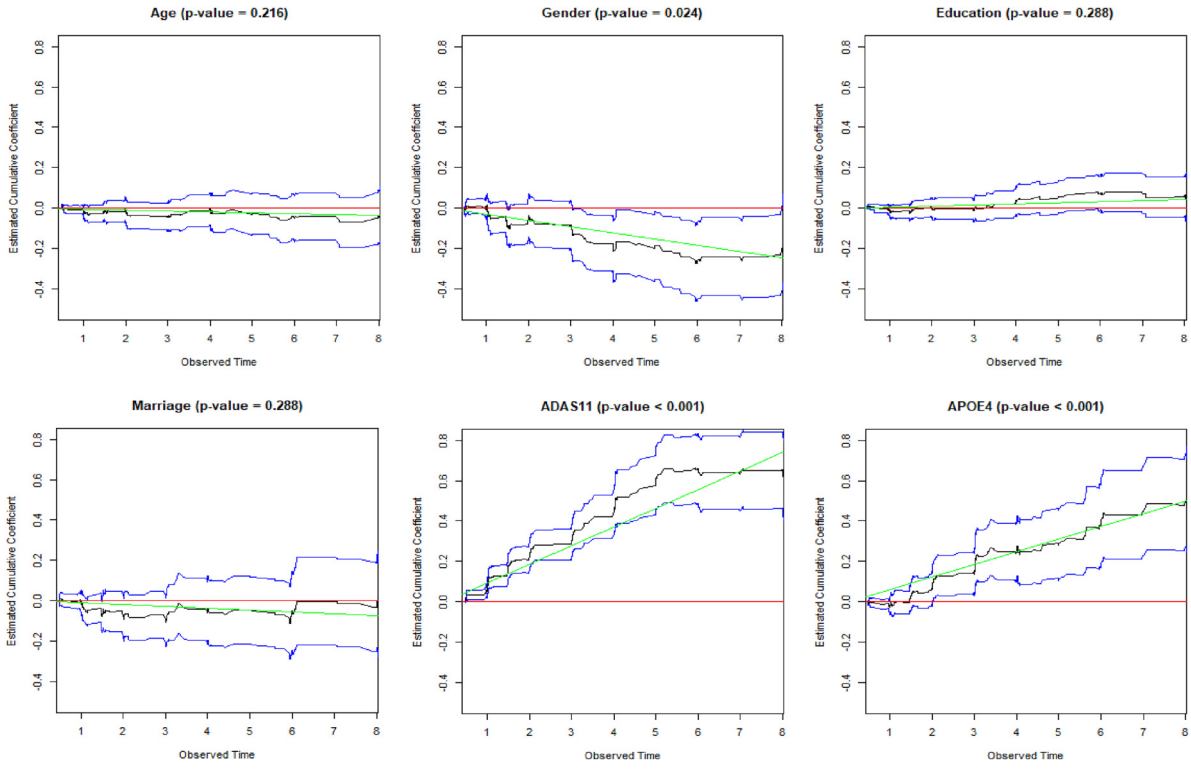

Figure 1. The black and green lines denote the estimated cumulative functional coefficients $\hat{\mathbf{A}}(t)$ and $\hat{\mathbf{A}}_{ly}(t)$ for each covariate, respectively. The blue lines show the 95% pointwise estimated confidence intervals for $\hat{\mathbf{A}}(t)$. The red lines denote the horizontal line of value 0. The *p*-value based on the test statistic value defined in Section 3.2. is subsumed in parenthesis following each risk factor. See the online version for exact depiction of the graph.

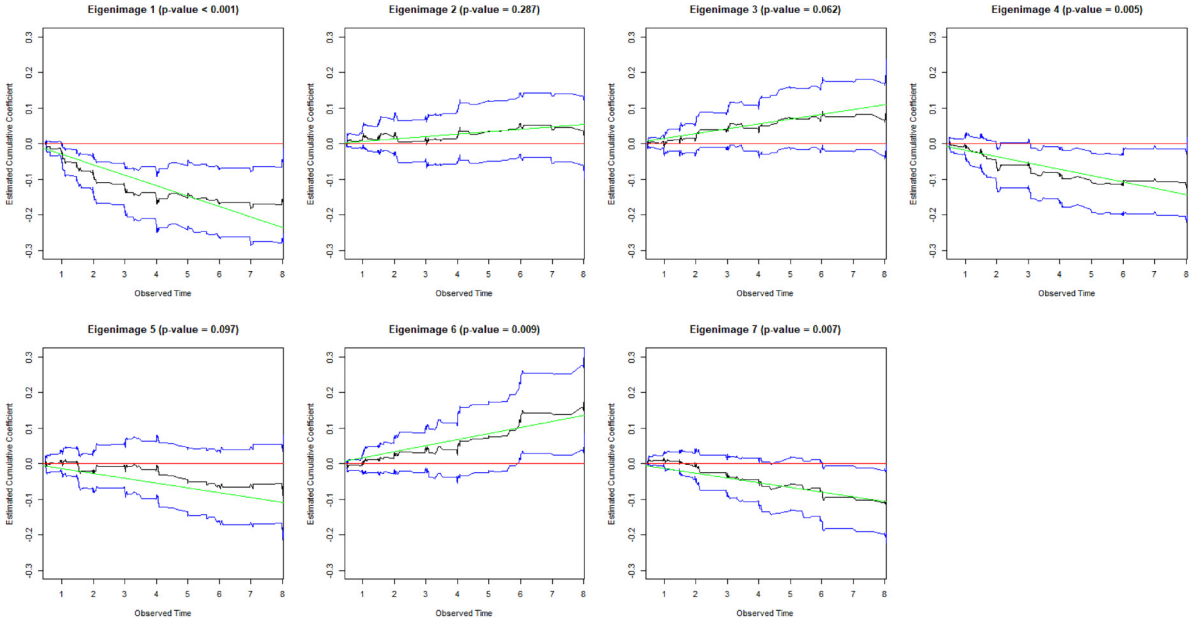

Figure 2. The black and green lines denote the estimated cumulative functional coefficients $\hat{\mathbf{A}}(t)$ and $\hat{\mathbf{A}}_{ly}(t)$, respectively. The blue lines show the 95% pointwise estimated confidence intervals for $\hat{\mathbf{A}}(t)$. The red lines denote the horizontal line of value 0. The *p*-value based on the test statistic value defined in Section 3.2. is subsumed in parenthesis following each risk factor.

In this simulation, a total of 1000 replicated datasets are generated under each combination of (CR, n). We first examine the performance of the FPCA method in recovering the imaging data. We use the R package "corpor" to obtain the SVD of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. In Figure 3, the first row presents the true eigenimages $\psi_m(v)$, $m = 1, 2, 3$, and the second row depicts the means of the estimated eigenimages $\hat{\psi}_m(v)$, $m = 1, 2, 3$, which are normalized through the formula $(\hat{\psi}_m(v) - \min_v \hat{\psi}_m(v))/(\max_v \hat{\psi}_m(v) - \min_v \hat{\psi}_m(v))$ to obtain the grayscale images with pixel values on the intensity interval [0, 1]. Figure 3 shows that the spatial

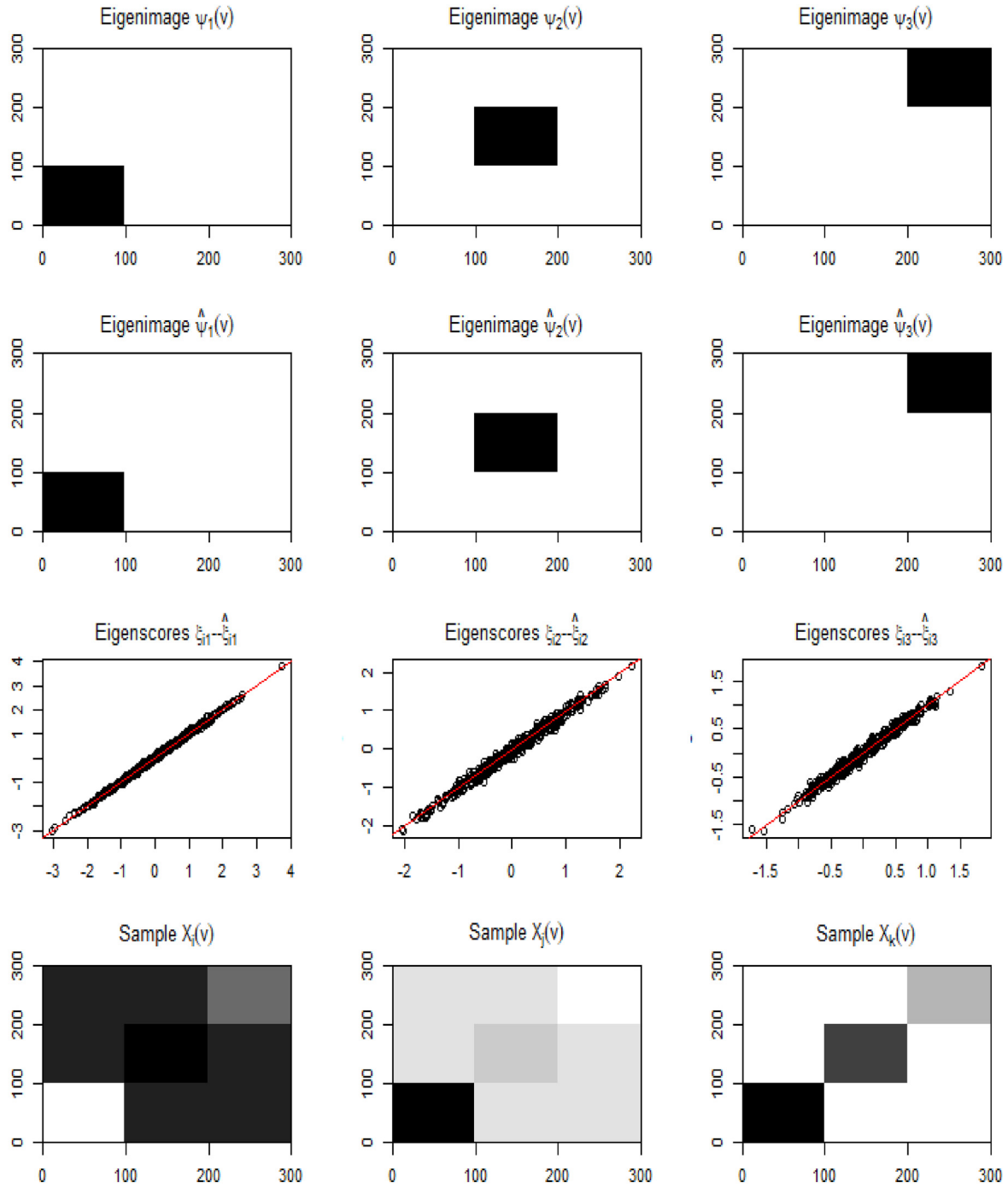

Figure 3. True eigenimages (top row), means of estimated eigenimages (second row), scatter diagrams of true versus estimated eigenscores in a randomly picked replication with the red line being $y = x$ (third row), and three sample images $\{X_i(v), X_j(v), X_k(v)\}$ (bottom row) in grayscale. For eigenimages and sample images, white and black pixels stand for 0 and 1, respectively, while gray pixels correspond to values in (0,1).

layout of true eigenimages can be perfectly reconstructed by the averages of the estimated eigenimages. Furthermore, we randomly pick a replication and plot the scatter diagram of the true versus estimated eigenscores in the third row of Figure 3, which suggests a good agreement between the estimators and the truth. In addition, we present three sample images $\{X_i(v), X_j(v), X_k(v)\}$ in the bottom row of Figure 3 with $(\xi_{i1}, \xi_{i2}, \xi_{i3}) = (-0.63, 0.09, -0.21)$, $(\xi_{j1}, \xi_{j2}, \xi_{j3}) = (1.59, 0.16, -0.21)$, and $(\xi_{k1}, \xi_{k2}, \xi_{k3}) = (0.48, 0.36, 0.14)$. Specifically, we calculate $X_i(v) = \xi_{i1}\psi_1(v) + \xi_{i2}\psi_2(v) + \xi_{i3}\psi_3(v)$ at each pixel to form a 300×300 matrix. Then, we normalize $X_i(v)$ through $(X_i(v) - min_v X_i(v))/(\max_v X_i(v) - min_v X_i(v))$ to obtain a grayscale image with pixel values in [0, 1] and plot the monochrome sample image $X_i(v)$. Likewise, we calculate and plot the monochrome sample images $X_j(v)$ and $X_k(v)$.

Table 1 summarizes the finite sample performance of the estimated nonparametric coefficients $\hat{\mathbf{A}}(t) = (\hat{A}_1(t), \ldots, \hat{A}_5(t))^{\mathrm{T}}$ at three different time points $t_1 = 0.1$, $t_2 = 0.5$, and $t_3 = 0.9$. Bias is the sampling mean of the estimate minus its true value, SE is the sampling standard error of the estimate, SEE is the sampling mean of the standard error estimate, and CP is the empirical coverage probability of 95% CI based on the normal approximation. Under all the situations considered, the proposed approach performs satisfactorily. Specifically, the estimators are basically unbiased, the estimated and empirical standard errors are in good agreement, and the 95% empirical CIs are close to the nominal level. As expected, the estimation performance improves when the sample size increases from 500 to 1000. Figure 4 depicts $\hat{A}_1(t), \ldots, \hat{A}_5(t)$ under various settings. The estimated cumulative time-varying coefficients are similar to their true functional curves, and the empirical and estimated pointwise CIs are close to each other, demonstrating the satisfactory performance of the proposed method in uncovering the time-varying coefficients.

Table 1. Estimation results of varying coefficients in Simulation 1.

		CR = 15%				CR = 30%			
n	$\hat{A}(t)$	Bias	SE	SEE	CP	Bias	SE	SEE	CP
500	$A_1(t_1)$	0.000	0.031	0.033	0.939	0.000	0.031	0.033	0.939
	$A_2(t_1)$	−0.001	0.018	0.020	0.941	−0.001	0.019	0.020	0.938
	$A_3(t_1)$	−0.002	0.016	0.016	0.937	−0.002	0.016	0.016	0.939
	$A_4(t_1)$	−0.001	0.023	0.024	0.944	−0.001	0.023	0.024	0.946
	$A_5(t_1)$	−0.003	0.033	0.033	0.940	−0.003	0.033	0.033	0.941
	$A_1(t_2)$	0.000	0.110	0.108	0.958	0.003	0.118	0.117	0.950
	$A_2(t_2)$	0.000	0.070	0.069	0.952	−0.001	0.075	0.075	0.949
	$A_3(t_2)$	−0.003	0.060	0.064	0.932	−0.004	0.064	0.068	0.930
	$A_4(t_2)$	−0.001	0.085	0.088	0.943	−0.002	0.091	0.093	0.941
	$A_5(t_2)$	0.002	0.120	0.118	0.957	0.002	0.129	0.127	0.964
	$A_1(t_3)$	0.009	0.254	0.254	0.952	0.016	0.315	0.316	0.950
	$A_2(t_3)$	−0.003	0.168	0.167	0.958	−0.004	0.210	0.208	0.953
	$A_3(t_3)$	0.000	0.143	0.153	0.935	0.001	0.178	0.192	0.933
	$A_4(t_3)$	−0.001	0.202	0.202	0.953	−0.005	0.252	0.251	0.961
	$A_5(t_3)$	0.003	0.286	0.284	0.960	0.009	0.357	0.360	0.958
1000	$A_1(t_1)$	0.000	0.022	0.022	0.944	0.000	0.022	0.023	0.945
	$A_2(t_1)$	−0.001	0.013	0.014	0.945	−0.001	0.013	0.014	0.945
	$A_3(t_1)$	−0.001	0.011	0.011	0.946	−0.001	0.011	0.011	0.943
	$A_4(t_1)$	−0.001	0.016	0.016	0.956	−0.001	0.016	0.016	0.958
	$A_5(t_1)$	0.000	0.023	0.023	0.948	0.000	0.023	0.024	0.949
	$A_1(t_2)$	0.002	0.077	0.074	0.952	0.002	0.083	0.081	0.951
	$A_2(t_2)$	−0.001	0.049	0.048	0.949	0.000	0.052	0.053	0.945
	$A_3(t_2)$	0.001	0.042	0.042	0.952	0.001	0.045	0.045	0.943
	$A_4(t_2)$	−0.002	0.059	0.060	0.951	−0.002	0.064	0.065	0.949
	$A_5(t_2)$	−0.005	0.085	0.085	0.957	−0.005	0.091	0.091	0.948
	$A_1(t_3)$	0.009	0.176	0.172	0.956	0.003	0.213	0.214	0.954
	$A_2(t_3)$	−0.003	0.117	0.115	0.952	0.000	0.143	0.145	0.954
	$A_3(t_3)$	0.004	0.099	0.100	0.942	0.000	0.120	0.124	0.954
	$A_4(t_3)$	0.001	0.139	0.145	0.947	−0.002	0.170	0.176	0.938
	$A_5(t_3)$	−0.009	0.199	0.204	0.947	−0.006	0.242	0.245	0.944

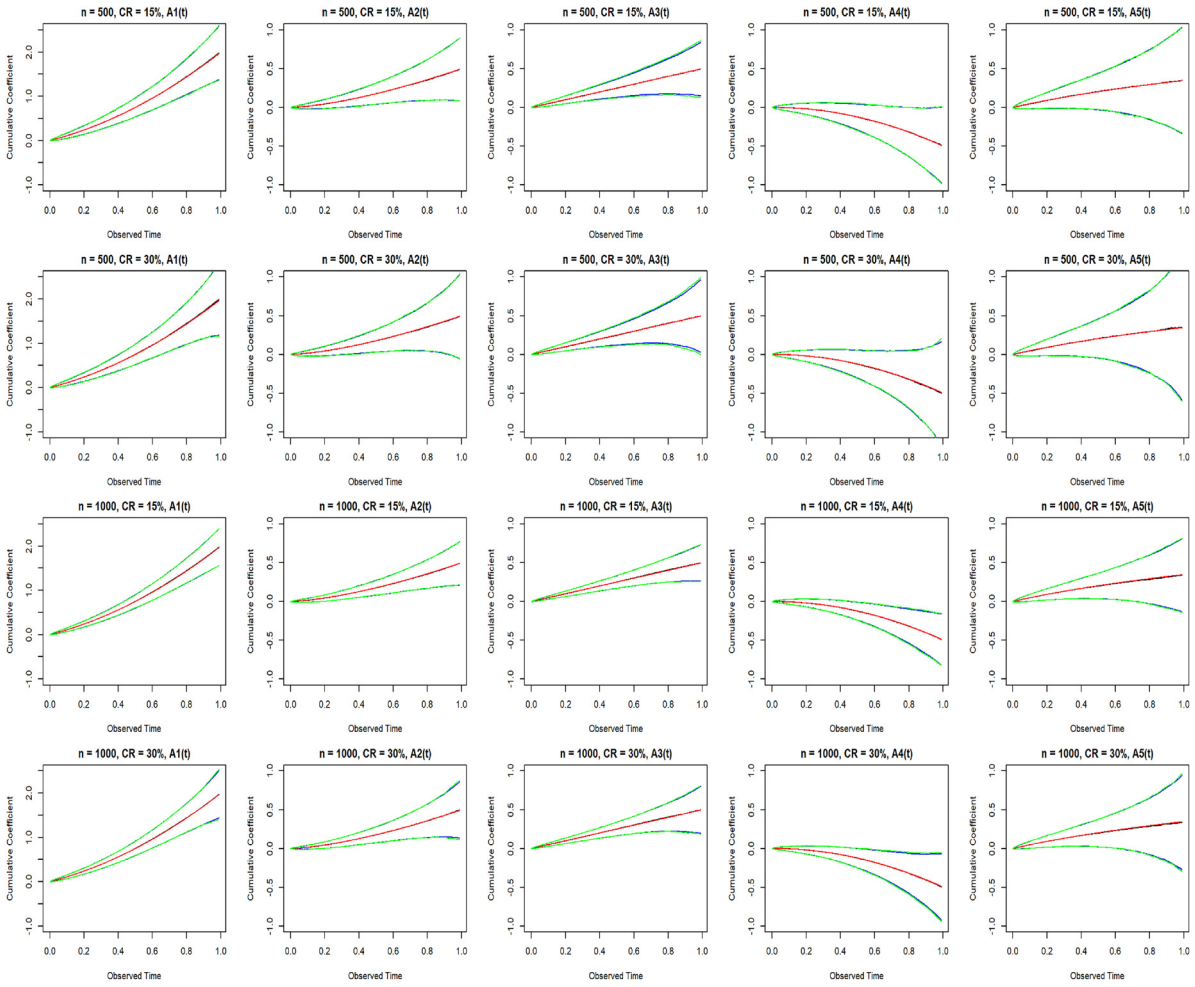

Figure 4. The black and red lines denote the estimated and true cumulative functional coefficients $\{\hat{A}_j(t), j = 1, \ldots, 1 + s + q\}$ for each covariate. The blue and green lines show the corresponding empirical and estimated 95% pointwise confidence intervals.

Moreover, we use the estimated eigenimages and time-varying coefficients $\{\hat{\Gamma}_m(t), m = 1, 2, 3\}$, which correspond to $\hat{A}_{m+2}(t)$, $m = 1, 2, 3$, respectively, to obtain $\hat{\Gamma}(v, t)$, namely $\sum_{m=1}^{3} \hat{\Gamma}_m(t)\hat{\psi}_m(v)$. Similarly, we normalize $\Gamma(v, t)$ and $\hat{\Gamma}(v, t)$ to obtain grayscale images with intensities on interval [0, 1]. Figure 5 presents the true (top row) and the mean of the estimated (bottom row) imaging coefficient at the three time points. The top and bottom rows are nearly identical, reconfirming the excellent performance of the proposed method in recovering the varying coefficient image effect. In addition, Figure 5 shows that the color of the imaging coefficient is light in an early stage (t_1) but becomes increasingly darker in certain areas in later stages (t_2 and t_3), indicating that the significance of a specific area may increase with time. This finding also implies that the significant area in the image can vary with time. Such dynamic changes deserve special attention but cannot be discovered through a constant-coefficient model, thereby demonstrating the utility of the proposed methodology.

The computer code for conducting the above analysis is written in R and is freely available at https://github.com/yqml/AHtimage.

4.2 Simulation 2

An essential assumption in this study is that the coefficient image $\gamma(\cdot, t)$ lies in the space expanded by the eigenimages of imaging covariates. This section investigates the performance of the proposed method when this assumption is violated. We extract 2D slices from the 3D MRI images in the ADNI data set as the imaging input $\mathbf{X}_i$. Without loss of generality, we select the axial slices with a size of 133×170. After deleting voxels of value 0 in the contour, we have $V = 17, 337$. The sample size is 585, the same as that in the real data analysis.

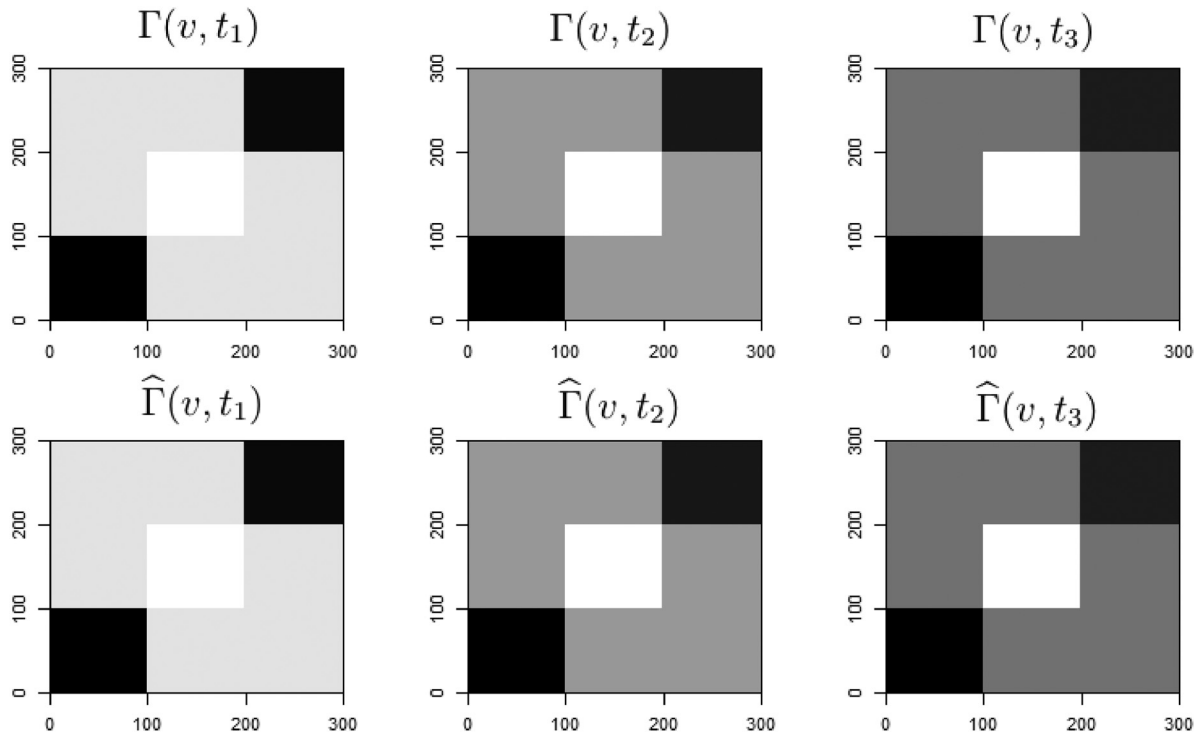

Figure 5. True cumulative imaging coefficient $\Gamma(v, t)$ (top row) and the mean of estimated imaging coefficient $\hat{\Gamma}(v, t)$ (bottom row) at $t_1 = 0.1$, $t_2 = 0.5$, and $t_3 = 0.9$.

We consider model (12) with $\beta(t) = (\beta_1(t), \beta_2(t))^{\mathrm{T}} = (2t, 0.75\sqrt{t})^{\mathrm{T}}$, $Z_{i1} = 1$, and $Z_{i2} \sim Binomial(3, 0.5)$. Let $\gamma(v, t) = \gamma_1(t)\tilde{X}_1(v) + \gamma_2(t)\tilde{X}_2(v)$, where $\gamma_1(t) = 2t$, $\gamma_2(t) = \frac{1}{2}t^{-\frac{1}{2}}$, and $\tilde{X}_1(v)$ and $\tilde{X}_2(v)$, $v = 1, \ldots, V$, are generated using R and shown in Figure 6 (first row). The true voxel values of $\tilde{X}_1(v)$ and $\tilde{X}_2(v)$ are all set to 0.1 in the influential (darker) regions and 0 in others. Denote the cumulative coefficient function of $\gamma(v, t)$ by $\Gamma(v, t)$. A simple calculation yields $\Gamma(v, t) = \int_0^t \gamma(v, u)du = t^2\tilde{X}_1(v) + t^{\frac{1}{2}}\tilde{X}_2(v)$. Figure 6 (second row) depicts $\Gamma(v, t)$ at several time points. With the above setting, we guarantee that the hazard function is positive and a proper distribution of the observed time $\widetilde{T}_i$. Furthermore, the maximum illustration time is restricted at $t = 0.7$ as data points are sparse beyond this boundary. The censoring time C_i is generated from $U(0, 2)$, leading to an overall censoring rate of approximately 20%. Notably, the coefficient image $\gamma(v, t)$ or $\Gamma(v, t)$ is generated using $\tilde{X}_1(v)$ and $\tilde{X}_2(v)$ based on the above procedure rather than expanded through the eigenimages of $\mathbf{X}_i$.

We first conduct FPCA based on $\mathbf{X}_i$ to obtain the eigenimages and eigenscores. Then, we include M eigenscores ξ_{im} in (12) to estimate the coefficient image. We consider $M = 30$, 50, and 70 to conduct the subsequent analysis based on a replicated data set and find that they perform similarly in recovering $\Gamma(v, t)$. Therefore, we fix $M = 50$ and repeat the simulation 1,000 times. The estimation results for the cumulative functions of $\beta(t)$ are similar to Simulation 1 and not reported. Figure 6 depicts the estimated $\Gamma(\cdot, t)$, $\widehat{\Gamma}(\cdot, t)$, at different time points. Even though the abovementioned assumption does not hold, $\widehat{\Gamma}(\cdot, t)$ can roughly recover the significant regions (in deeper color) and the time-varying trend of $\Gamma(\cdot, t)$.

4.3 Simulation 3

In this section, we conduct an additional simulation to mimic the application setup. We consider the following model:

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) = \beta_1(t)Z_{i1} + \beta_2(t)Z_{i2} + \beta_3(t)Z_{i3} + \int_{\mathcal{V}} \gamma(v, t)X_i(v)\mathrm{d}v, \tag{13}$$

where $\beta_1(t)$, $\beta_2(t)$, Z_{i1}, Z_{i2}, $\gamma(v, t)$, and $X_i(v)$ are the same as in Simulation 1, $\beta_3(t) = 0.5$, $Z_{i3} \sim U(0, 1)$, and $\mathbf{A}(t) = (t^2 + t, 0.5t^{1.5}, 0.5t, 0.5t, -0.5t^2, 0.5\log(1 + t))^{\mathrm{T}}$. The censoring time is independently generated from a beta distribution $Beta(1, 2.5)$ to obtain a censoring rate of approximately 60%. The sample size is set to $n = 600$. We perform estimation similarly to Simulation 1 based on 1000 replicated datasets.

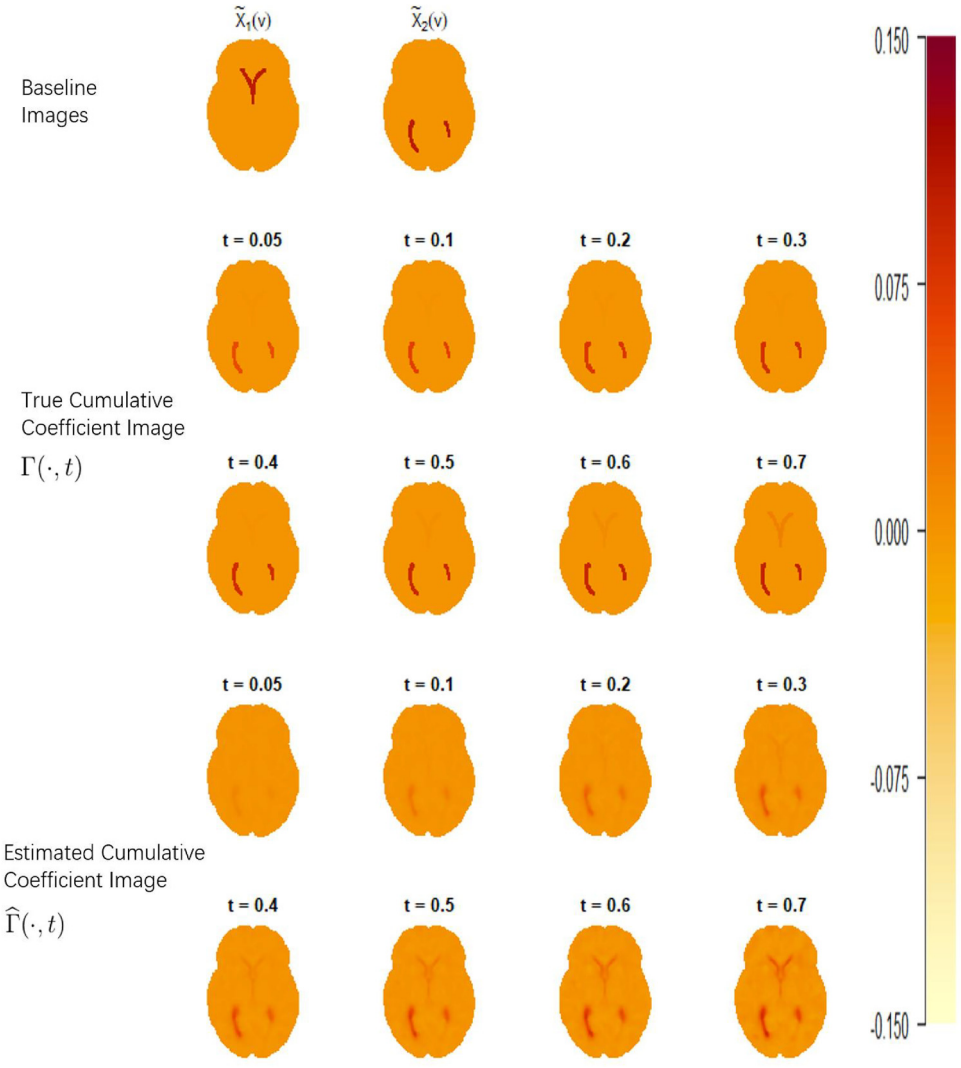

Figure 6. The first row presents $\widetilde{X}_1(\cdot)$ and $\widetilde{X}_2(\cdot)$. The second and third rows represent the true and estimated cumulative coefficient images, $\Gamma(\cdot, t)$ and $\hat{\Gamma}(\cdot, t)$, at different time points in Simulation 2.

Table 2. Estimation results of varying coefficients in Simulation 3 ($n = 600$, CR = 60%).

$t_1 = 0.1$	$A_1(t_1)$	$A_2(t_1)$	$A_3(t_1)$	$A_4(t_1)$	$A_5(t_1)$	$A_6(t_1)$
Bias	0.000	−0.001	−0.002	−0.001	−0.002	−0.003
SE	0.044	0.020	0.061	0.017	0.024	0.035
SEE	0.046	0.020	0.062	0.018	0.024	0.036
CP	0.937	0.941	0.949	0.934	0.947	0.941
$t_2 = 0.3$	$A_1(t_2)$	$A_2(t_2)$	$A_3(t_2)$	$A_4(t_2)$	$A_5(t_2)$	$A_6(t_2)$
Bias	0.004	−0.001	−0.004	−0.003	−0.007	−0.003
SE	0.112	0.051	0.157	0.044	0.063	0.090
SEE	0.110	0.053	0.157	0.043	0.067	0.093
CP	0.950	0.946	0.947	0.961	0.940	0.938
$t_3 = 0.6$	$A_1(t_3)$	$A_2(t_3)$	$A_3(t_3)$	$A_4(t_3)$	$A_5(t_3)$	$A_6(t_3)$
Bias	0.018	−0.011	−0.020	−0.013	−0.005	0.002
SE	0.353	0.168	0.504	0.143	0.204	0.289
SEE	0.352	0.168	0.505	0.150	0.210	0.307
CP	0.957	0.954	0.955	0.947	0.953	0.945

Table 2 presents the estimated nonparametric coefficients $\hat{\mathbf{A}}(t) = (\hat{A}_1(t), \ldots, \hat{A}_6(t))^{\mathrm{T}}$ at three different time points: $t_1 = 0.1$, $t_2 = 0.3$, and $t_3 = 0.6$ as data points are sparse beyond t_3. The estimation performance under this setting is still acceptable, although slightly reduced because of a higher censoring rate. Furthermore, Figure 7 shows the estimated cumulative coefficients $\hat{A}_1(t), \ldots, \hat{A}_6(t)$, also suggesting the satisfactory estimation performance of the proposed method.

5 Application to ADNI data

We applied the proposed model to the ADNI study as described in the Introduction. Our primary interest is to investigate possible genetic, imaging, and clinical risk factors for AD hazards and reveal their dynamic effects. AD is a progressive disease that typically starts slowly with mild cognitive impairment, such as short-term memory loss, and worsens over time, eventually leading to loss of the ability to carry on a conversation, respond to the environment, and conduct activities of daily living. In clinical practice, subjects are diagnosed with AD provided that they meet the following inclusion criteria: (1) the abnormal memory loss recorded by scoring is lower than the education-adjusted cutoff on the Logical Memory II subscale from the Wechsler Memory Scale, (2) scores of Mini-Mental State Exam are between 20 and 26, (3) compliant of memory from the subject or study partner (SP), which the SP confirmed, (4) a score of 0.5 on the Clinical Dementia Rating test, and (5) National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association criteria for AD diagnosis. Initially, approximately 800 subjects were recruited for the study. However, many participants' clinical biomarkers or image information were not completely recorded, leading to a certain percentage of missing data. In this study, we excluded subjects with incomplete information and considered the remaining 585 participants, among whom 231 of them met the inclusion criteria and were diagnosed with AD during the cohort study, whereas 354 remained at the stage of cognitive normal or mild cognitive impairment. The observed time was counted from the enrollment of subjects to the date of the diagnosis of AD if $\delta_i = 1$ or the date of the last visit if $\delta_i = 0$. The censoring

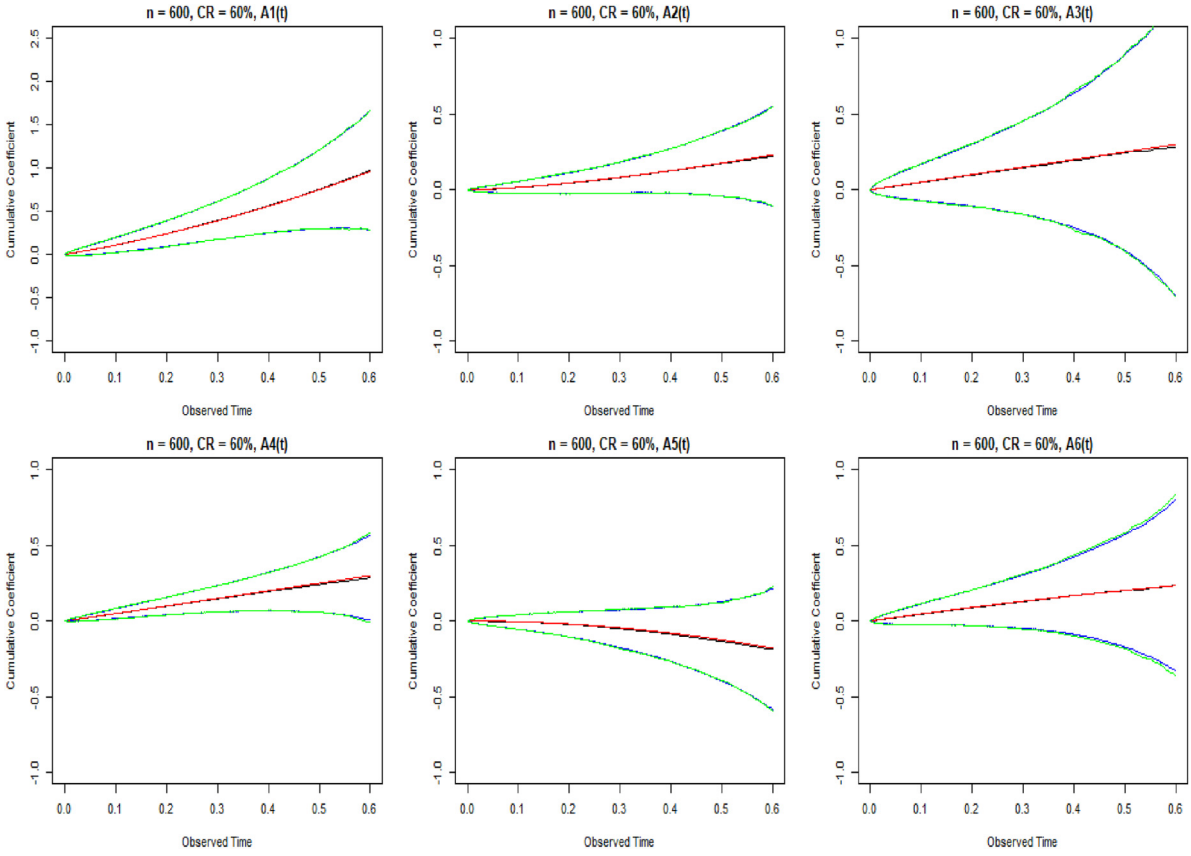

Figure 7. The first row presents $\hat{A}_1(t)$, $\hat{A}_2(t)$, and $\hat{A}_3(t)$. The second row represents $\hat{A}_4(t)$, $\hat{A}_5(t)$, and $\hat{A}_6(t)$ in Simulation 3. The black and red lines denote the estimated and true cumulative functional coefficients $\{\hat{A}_j(t), j = 1, \ldots, 6\}$. The blue and green lines show the corresponding empirical and estimated 95% pointwise confidence intervals.

rate is approximately 61%. The possible clinical risk factors include age (z_1), gender (male = 1, female = 0, z_2), educational level (years of school attainment, z_3), marriage status (unmarried = 1, married = 0, z_4), AD Assessment Scale-Cognitive Subscale 11 (ADAS11, high scores indicating poor cognitive ability, z_5), and genetic biomarkers Apolipoprotein E epsilon 4 (APOE-ϵ4, carrier = 1, noncarrier = 0, z_6).

For statistical inference, we first conducted FPCA for the imaging data to extract the leading eigenvalues, eigenimages, and eigenscores and then included the extracted eigenscores as known covariates in the time-varying coefficient AH model. We used Bayesian information criterion (BIC) to determine the number of retained eigenimages. The computation of BIC involves the estimation of the time-varying coefficient $\hat{\alpha}(t)$ because the likelihood function of model (4) includes $\sum_{i=1}^{n} \hat{\lambda}(\widetilde{T}_i) \exp(-\hat{\Lambda}(\widetilde{T}_i))$, where $\hat{\lambda}(\widetilde{T}_i) = \hat{\alpha}^{\mathrm{T}}(\widetilde{T}_i)\mathbf{Z}_i$. We adopted the kernel estimator $\hat{\alpha}_{Ker}(t)$ (9) and the elementary kernel function (10) with $\epsilon = 0.001$ to estimate $\hat{\alpha}(t)$. We set the number of leading eigenimages M ranging from 1 to 50 and the bandwidth h ranging from 0.01 to 2 in a step size of 0.01 to calculate BIC. Then, we adopted the same normalization technique as for the eigenimages and coefficient images to convert the negative BIC values to the range of [0, 1]. Figure 8 graphically displays a submatrix of the converted BIC values as a grayscale image, with the darkest part denoting the lowest BIC value. The BIC value attained minimal when $M = 7$ and $h = 0.13$. Figures 1 and 2 depict the estimated cumulative coefficients $\hat{\mathbf{A}}(t)$ in model (4), along with their 95% pointwise CIs. The p-values presented in the parentheses following the scalar or imaging covariates show the statistical significance test results of the time-varying coefficients. The slopes of the estimated cumulative functional coefficients reflect the effects of the corresponding covariates on the hazard

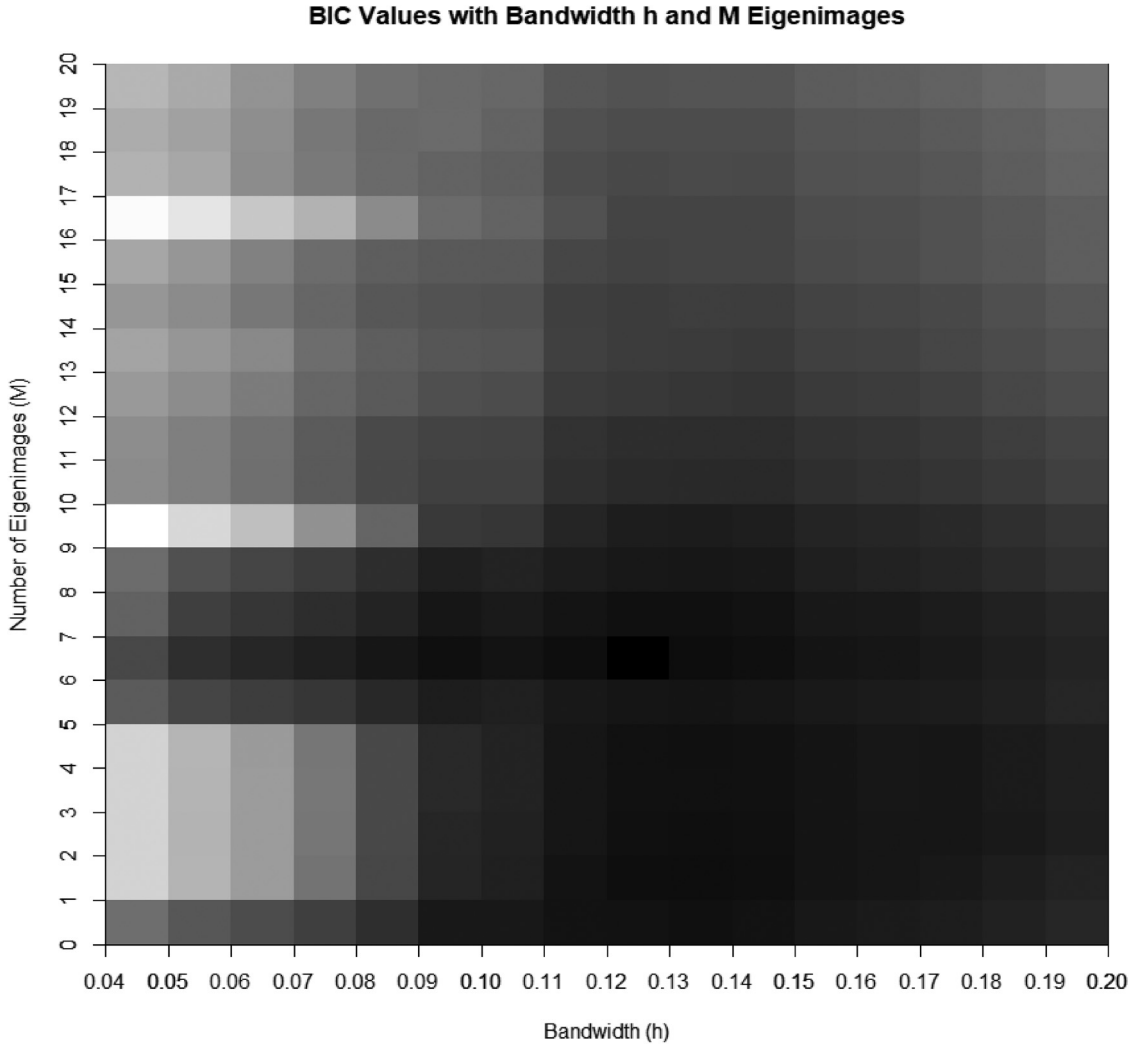

Figure 8. The graphical illustration of the Bayesian information criterion (BIC) value submatrix in a grayscale image, with the darkest part indicating the lowest BIC value.

function. We find in Figure 1 the following results. First, gender negatively affects AD hazards, indicating that male patients are less likely to progress to AD than female patients. Moreover, this effect gradually disappears after approximately six years because the estimated cumulative coefficient curve becomes nearly a horizontal line. Second, ADAS11 and APOE-ϵ4 exhibit significant positive effects on AD hazards. Patients with higher ADAS11 scores or APOE-ϵ4 carriers have a higher risk of developing AD than those with lower ADAS11 scores or APOE-ϵ4 noncarriers. However, the effect of ADAS11 seems to fade out after five to six years due to the horizontal trend of its estimated cumulative coefficient curve. In contrast, the influence of APOE is relatively persistent. These findings agree with the existing literature.[35,36,29] Moreover, we observe that age, an influential risk factor identified in previous studies,[37,38] does not significantly affect AD hazards. A possible reason is that we simultaneously included ADAS11, a commonly used measurement to assess cognitive impairment, as a covariate in the AH model, and the age effect might have become negligible when controlling ADAS11. Regarding the imaging predictors, the first, fourth, sixth, and seventh eigenimages significantly affect AD hazards. However, similar to the scalar covariates, the impacts of these eigenimages gradually fade out after five to six years. It is also worth noting that the significance results reflected by the p-values are in line with the 95% pointwise CIs, which indicate the significance of a covariate by excluding the horizontal line of value 0 (red lines in Figures 1 and 2).

For comparison, we also analyzed the constant-coefficient AH model[3] with imaging predictors, which could be formulated as follows:

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) = \lambda_0(t) + \beta_{ly}^{\mathrm{T}}\mathbf{Z}_i + \int_{\mathcal{V}} \gamma_{ly}^{\mathrm{T}}(v)X_i(v)\mathrm{d}v, \tag{14}$$

where $\lambda_0(t)$ and $\beta_{ly} = (\beta_{ly,1}, \ldots, \beta_{ly,s+1})^{\mathrm{T}}$ denote the unspecified baseline hazard function and the constant coefficient of covariate $\mathbf{Z}_i$, respectively, and $\gamma_{ly}(v)$ denotes the coefficient image of the imaging data $X_i(v)$. To guarantee comparability, we adopted the same number of eigenimages, namely, $M = 7$, in model (14) without conducting model selection through BIC. Accordingly, model (14) can be reformulated as follows:

$$\lambda(t|\mathbf{Z}_i, \mathbf{X}_i) = \lambda_0(t) + \beta_{ly}^{\mathrm{T}}\mathbf{Z}_i + \gamma_{ly}^{\mathrm{T}}\xi_i, \tag{15}$$

where $\gamma_{ly} = (\gamma_{ly,1}, \ldots, \gamma_{ly,7})^{\mathrm{T}}$ denotes the constant coefficient of the imaging covariate ξ_i. We integrated the estimated constant coefficients as $\hat{\alpha}_{ly} = (\hat{\beta}_{ly}^{\mathrm{T}}, \hat{\gamma}_{ly}^{\mathrm{T}})^{\mathrm{T}}$. Then, the estimated cumulative functional coefficients $\hat{\mathbf{A}}_{ly}(t) = \hat{\alpha}_{ly}t$, which are also included (green lines) in Figures 1 and 2 for easy comparison. $\hat{\mathbf{A}}_{ly}(t)$ and $\hat{\mathbf{A}}(t)$ are in good agreement during the first couple of years, reflected by their similar slopes. However, the diminishing effects of the observed and imaging predictors, as indicated by the change of slopes in $\hat{\mathbf{A}}(t)$, cannot be captured by $\hat{\mathbf{A}}_{ly}(t)$. For the estimated parameters shown in Table 3, the estimated constant coefficients in $\hat{\alpha}_{ly}$ have the same directions as those in $\hat{\alpha}_{Ker}(t)$, the averaged temporal covariate effects calculated using the mean of $\hat{\alpha}_{Ker}(t)$ over the observed time interval [0, 8]. However, the averaged time-varying effects in $\hat{\alpha}_{Ker}(t)$ are more pronounced than the constant effects in $\hat{\alpha}_{ly}$. Moreover, the significance test results indicated by the p-values are in line for both models, except for the 5th eigenimage.

Table 3. Numerical comparison between $\hat{\alpha}_{ly}$ and $\hat{\alpha}_{Ker}(t)$.

Covariates	Est ($\hat{\alpha}_{ly}$)	p-value	Est ($\hat{\alpha}_{Ker}$)	p-value
Age	−0.005	0.272	0.003	0.216
Gender	−0.031	0.011	−0.066	0.024
Education	0.005	0.179	0.007	0.288
Marriage	−0.010	0.256	−0.008	0.288
ADAS11	0.062	<0.001	0.240	<0.001
APOE4	0.092	<0.001	0.238	<0.001
Eigenimage 1	−0.030	<0.001	−0.084	<0.001
Eigenimage 2	0.006	0.157	0.021	0.287
Eigenimage 3	0.013	0.014	0.045	0.022
Eigenimage 4	−0.018	0.005	−0.066	0.005
Eigenimage 5	−0.014	0.041	−0.036	0.097
Eigenimage 6	0.016	0.008	0.053	0.009
Eigenimage 7	−0.014	0.013	−0.045	0.007

ADAS11: AD Assessment Scale-Cognitive Subscale 11.

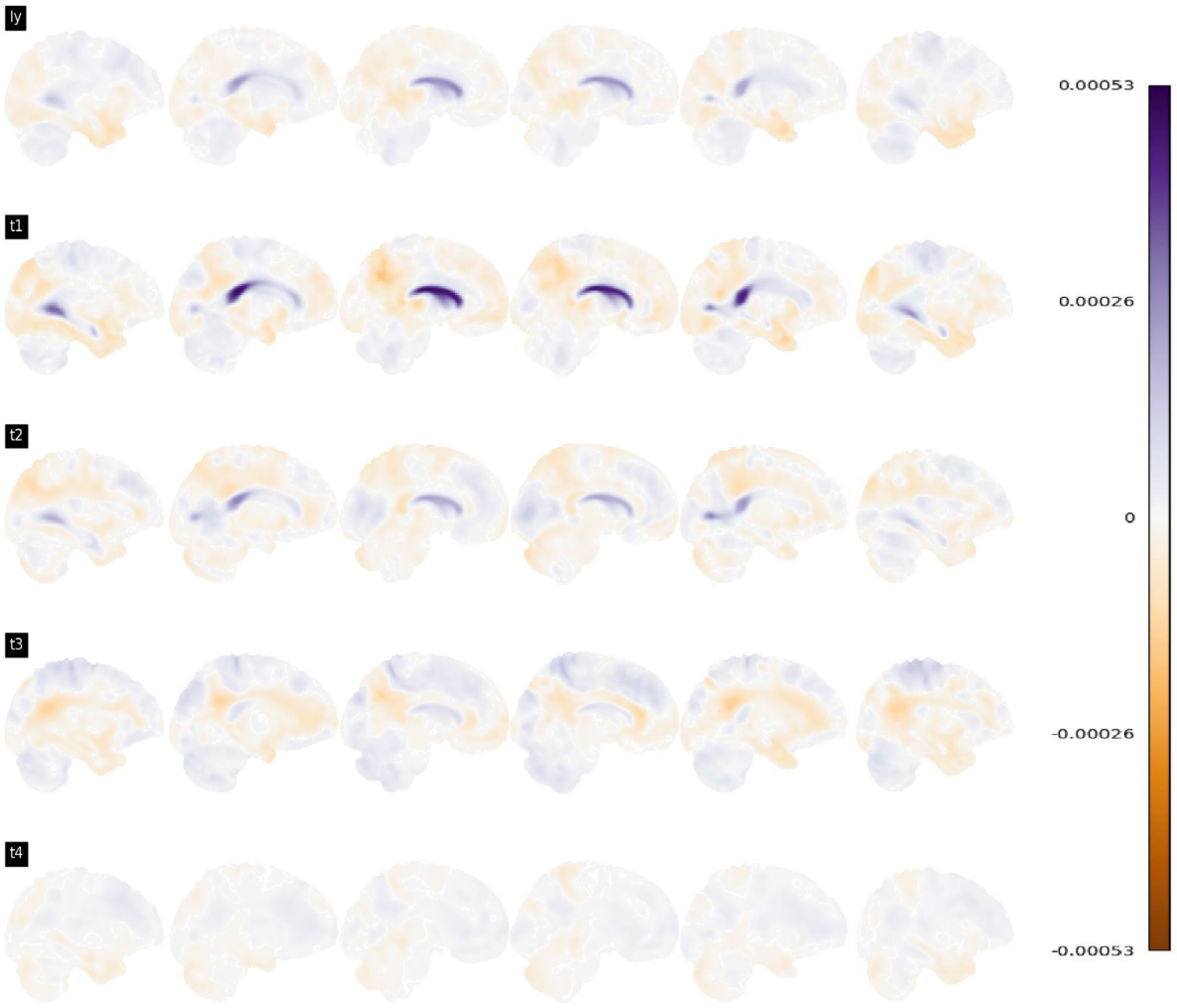

Figure 9. The cornal planes of estimated time-varying coefficient image $\gamma(v, t)$ at four time points $t_1 = 1$, $t_2 = 3$, $t_3 = 5$, $t_4 = 7$ (in year).

Furthermore, we utilized the estimated functional coefficients $\{\hat{\gamma}_m(t), m = 1, \ldots, M\}$ obtained through the kernel estimator $\hat{\alpha}_{Ker}(t)$ and the estimated eigenimages to reconstruct the coefficient image $\hat{\gamma}(v, t)$. The estimated coefficient image was calculated through the truncated KL expansion with $M = 7$ leading estimated time-varying coefficients and eigenimages, namely, $\hat{\gamma}(v, t) = \sum_{m=1}^{7} \hat{\gamma}_m(t)\hat{\psi}_m(v)$. Similarly, we reconstructed the estimated constant coefficient image $\hat{\gamma}(v)$ through $\hat{\gamma}(v) = \sum_{m=1}^{M} \hat{\gamma}_{ly,m}\hat{\psi}_m(v)$. Figures 9 to 11 depict the estimated constant imaging coefficient $\hat{\gamma}(\cdot)$ and time-varying imaging coefficient $\hat{\gamma}(\cdot, t)$ at four different time points. The significant regions (deep colors) in $\hat{\gamma}(\cdot)$ are time-invariant (first row). In contrast, those in $\hat{\gamma}(\cdot, t)$ (second to fifth rows) are the most apparent at t_1 but become steadily weak from t_2 to t_4, exhibiting a declining trend. Such diminishing covariate effects are perhaps partially due to their collection at baseline. Nevertheless, a constant-coefficient AH model cannot reveal this time-varying feature.

6 Conclusion

This study considered a time-varying coefficient AH model with imaging predictors. Our model extended the work of Aalen[1] by incorporating image data as covariates. A hybrid procedure that combines the FPCA technique and estimating equation approach was developed to conduct statistical inference. Applying the proposed method to the ADNI dataset provided new insights into the time-varying effects of the scalar and imaging risk factors on AD progression. We found that several potential risk factors affect AD hazards significantly in the first few years but gradually fade out. Such dynamic covariate effects must exert researchers' attention, but conventional AH models with time-invariant coefficients cannot uncover them.

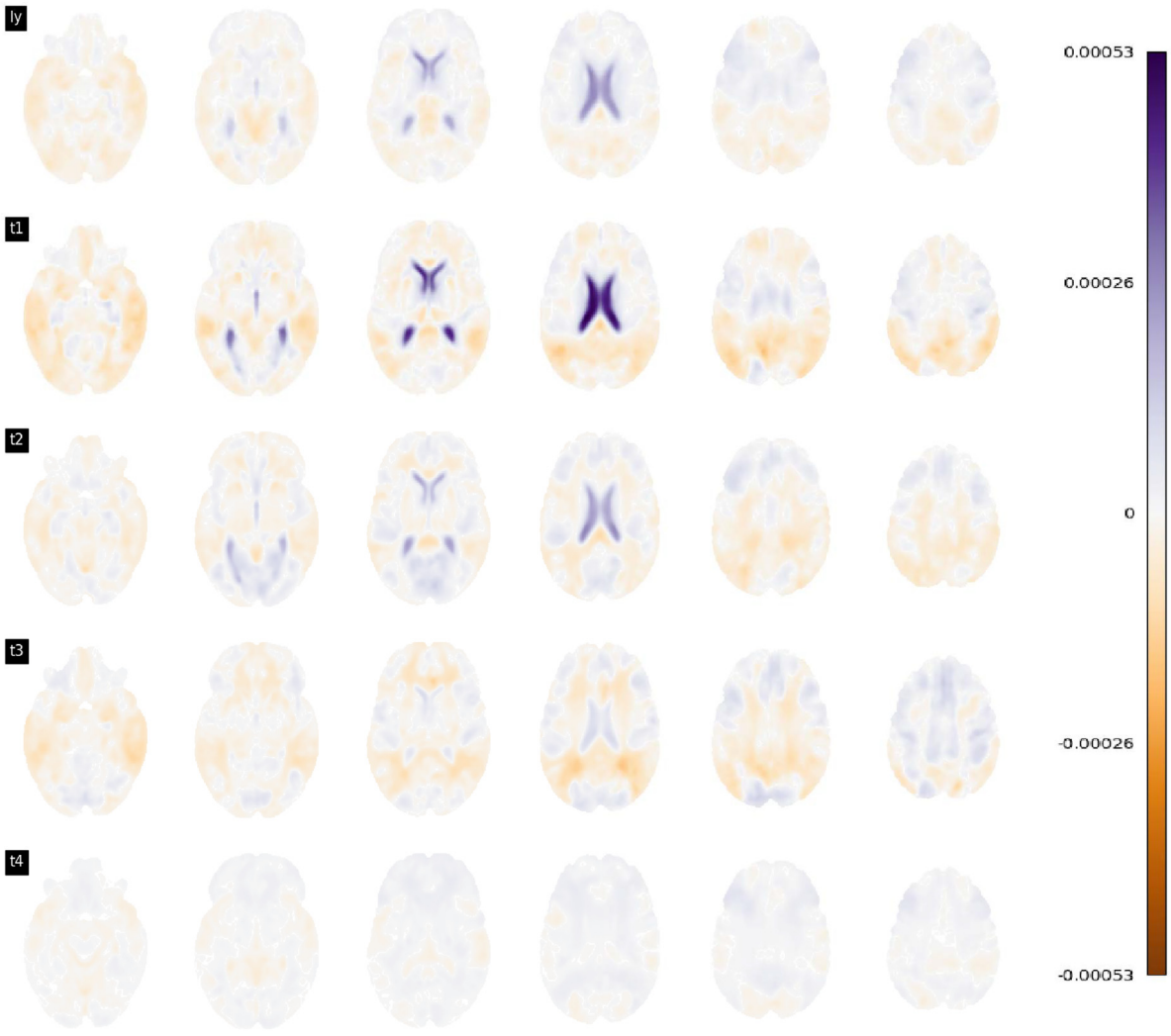

Figure 10. The axial planes of estimated time-varying coefficient image $\gamma(v, t)$ at four time points $t_1 = 1$, $t_2 = 3$, $t_3 = 5$, $t_4 = 7$ (in year).

The present study has several limitations and potential extensions. First, the model assumed that the scalar and imaging covariates were time-invariant. Considering longitudinal scalar and imaging covariates can further enhance model applicability. However, an identification issue arises when the coefficients and covariates in the AH model are both time-varying, leading to difficulties in obtaining consistent estimators. Exploring constraints to ensure identifiability in this situation is worth exploring in the future. Second, we vectorized the imaging data in the proposed approach, which ignored the spatial structures or dependences of the imaging data. The method of Goldsmith et al.[39] can accommodate the spatial correlations in analyzing multidimensional images. However, incorporating the above procedure into the proposed model framework may be computationally demanding, and its feasibility requires further investigation. Third, we can consider other survival models such as time-varying proportional hazards,[21,40–42] mean residual life,[43,44] and accelerated failure time models.[45,46] The inclusion of imaging predictors in various models can provide deep insight into the associations between images and survival outcomes of interest. Fourth, we assumed that the images and the associated time-varying coefficients could be expanded on the same eigenbases. This assumption may not hold in practice. Developing a general method that does not require such an assumption is of considerable interest. Finally, we can extend the current research to account for missing data commonly encountered in realistic research settings. The existing literature on missing data under the survival analysis framework[47–51] has never considered time-varying coefficient AH models with missing scalar or imaging data. Thus, this extension represents a potential further research topic.

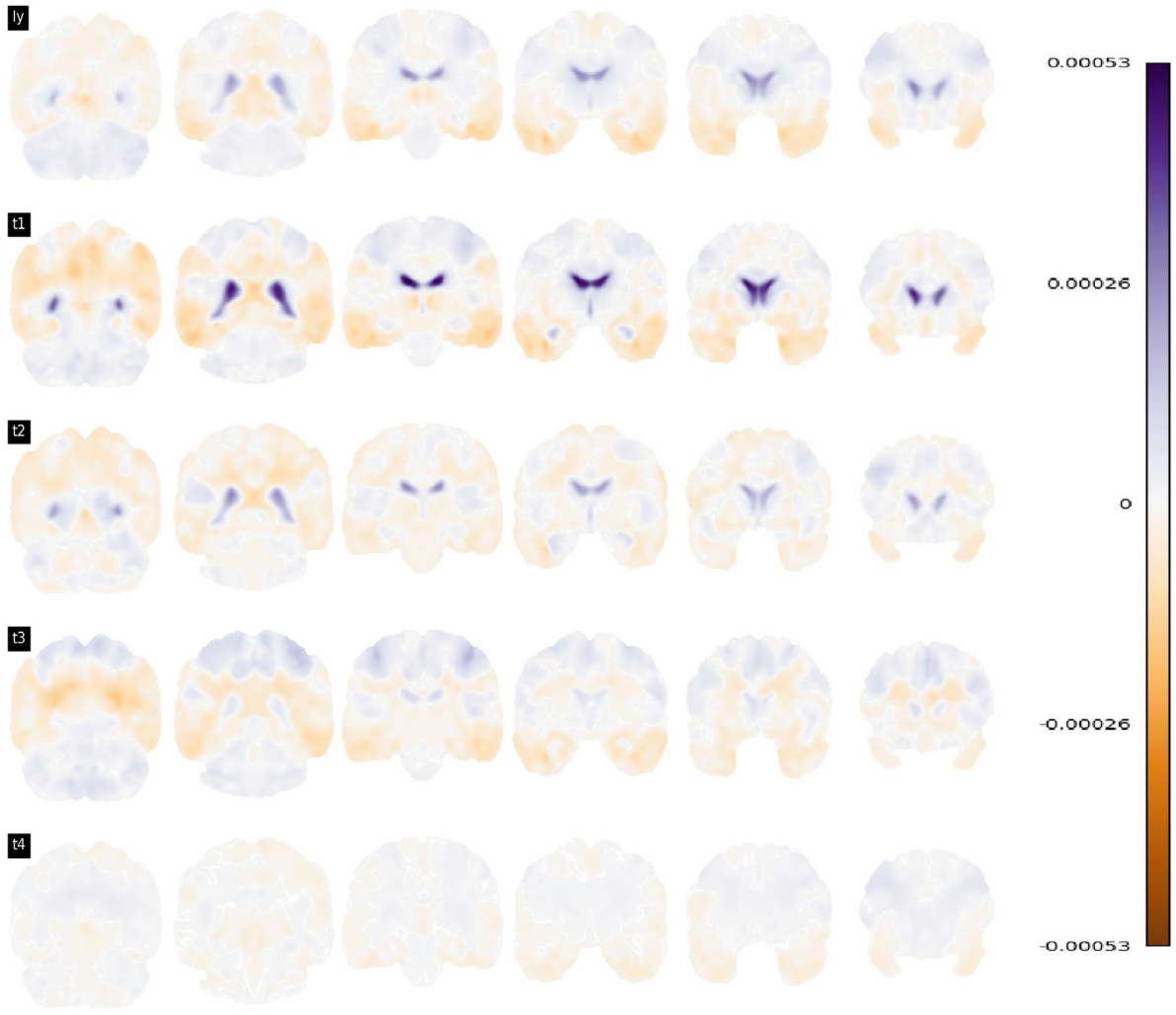

Figure 11. The sagittal planes of estimated time-varying coefficient image $\gamma(v, t)$ at four time points $t_1 = 1$, $t_2 = 3$, $t_3 = 5$, $t_4 = 7$ (in year).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was fully supported by GRF grants 14301918 and 14302220 from the Research Grant Council of the Hong Kong Special Administrative Region.

ORCID iD

Xinyuan Song https://orcid.org/0000-0002-4877-3200

References

1. Aalen OO. A linear regression model for the analysis of life times. *Stat Med* 1989; **8**: 907–925.
2. Huffer FW and McKeague IW. Weighted least squares estimation for Aalen's additive risk model. *J Am Stat Assoc* 1991; **86**: 114–129.
3. Lin D and Ying Z. Semiparametric analysis of the additive risk model. *Biometrika* 1994; **81**: 61–71.
4. Martinussen T and Scheike TH. *Dynamic regression models for survival data*. New York: Springer, 2006.
5. Jiang J and Haibo Z. Additive hazard regression with auxiliary covariates. *Biometrika* 2007; **94**: 359–369.

6. Pan D, He H, Song X, et al. Regression analysis of additive hazards model with latent variables. *J Am Stat Assoc* 2015; **110**: 1148–1159.
7. Qu L, Song X and Sun L. Identification of local sparsity and variable selection for varying coefficient additive hazards models. *Comput Stat Data Anal* 2018; **125**: 119–135.
8. Bard JF. A reliability analysis of occupational exposure data using a family of proportional and additive hazard functions. *J Environ Syst* 1986; **15**: 293–306.
9. Colorado HA, Velásquez EIG and Monteiro SN. Sustainability of additive manufacturing: the circular economy of materials and environmental perspectives. *J Mater Res Technol* 2020; **9**: 8221–8234.
10. Kremer A, Weißbach R and Liese F. Maximum likelihood estimation for left-censored survival times in an additive hazard model. *J Stat Plan Inference* 2014; **149**: 33–45.
11. Yang Q, He H, Lu B, et al. Mixture additive hazards cure model with latent variables: Application to corporate default data. *Comput Stat Data Anal* 2022; **167**: 107365.
12. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol* 1972; **34**: 187–202.
13. Cox DR. Partial likelihood. *Biometrika* 1975; **62**: 269–276.
14. Fleming TR and Harrington DP. *Counting processes and survival analysis*. New York: Wiley, 2011.
15. Pan D, Kang K, Wang C, et al. Bayesian proportional hazards model with latent variables. *Stat Methods Med Res* 2019; **28**: 986–1002.
16. Buckley J. Additive and multiplicative models for relative survival rates. *Biometrics* 1984; **40**: 51–62.
17. McKeague IW. Estimation for a semimartingale regression model using the method of sieves. *Ann Stat* 1986; **4**: 579–589.
18. McKeague IW and Sasieni PD. A partly parametric additive risk model. *Biometrika* 1994; **81**: 501–514.
19. Li H, Yin G and Zhou Y. Local likelihood with time-varying additive hazards model. *Can J Stat* 2007; **35**: 321–337.
20 . Cai Z and Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand J Stat* 2003; **30**: 93–111.
21. Yang Q, He H and Song X. Time-varying coefficient additive hazards model with latent variables. *Stat Methods Med Res* 2022; **31**(5): 928–946.
22. Ramsay JO and Silverman BW. *Functional data analysis*. New York: Springer-Verlag, 2005.
23. Ferraty F and Vieu P. *Nonparametric functional data analysis: Theory and practice*. New York: Springer, 2006.
24. Yao F. Functional principal component analysis for longitudinal and survival data. *Stat Sin* 2007; **17**: 965–983.
25. Zhu H, Fan J and Kong L. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J Am Stat Assoc* 2014; **109**: 1084–1098.
26. Morris JS. Functional regression. *Annu Rev Stat Appl* 2015; **2**: 321–359.
27. Wang X, Zhu H and Initiative ADN. Generalized scalar-on-image regression models via total variation. *J Am Stat Assoc* 2017; **112**: 1156–1168.
28. Feng X, Li T, Song X, et al. Bayesian scalar on image regression with nonignorable nonresponse. *J Am Stat Assoc* 2020; **115**: 1574–1597.
29. Kang K and Song XY. Joint modeling of longitudinal imaging and survival data. *J Comput Graph Stat* 2022; **0**: 1–11.
30. Zipunnikov V, Greven S, Shou H, et al. Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Ann Appl Stat* 2014; **8**: 2175.
31. Zipunnikov V, Caffo B, Yousem DM, et al. Functional principal component model for high-dimensional brain imaging. *NeuroImage* 2011; **58**: 772–784.
32. Ramlau-Hansen H. Smoothing counting process intensities by means of kernel functions. *Ann Stat* 1983; **11**: 453–466.
33. Kong D, Ibrahim JG, Lee E, et al. FLCRM: Functional linear Cox regression model. *Biometrics* 2018; **74**: 109–117.
34. Huang Y. Aβ-independent roles of apolipoprotein E4 in the pathogenesis of Alzheimer's disease. *Trends Mol Med* 2010; **16**: 287–294.
35. Petersen RC, Roberts RO, Knopman DS et al. Prevalence of mild cognitive impairment is higher in men: the Mayo clinic study of aging. *Neurology* 2010; **75**: 889–897.
36. Herrup K. Reimagining Alzheimer's disease—an age-based hypothesis. *J Neurosci* 2010; **30**: 16755–16762.
37. Guerreiro R and Bras J. The age factor in Alzheimer's disease. *Genome Med* 2015; **7**: 1–3.
38. Goldsmith J, Bobb J, Crainiceanu CM, et al. Penalized functional regression. *J Comput Graph Stat* 2011; **20**: 830–851.
39. Tian L, Zucker D and Wei L. On the Cox model with time-varying regression coefficients. *J Am Stat Assoc* 2005; **100**: 172–183.
40. Zhou X and Song X. Mediation analysis for mixture Cox proportional hazards cure models. *Stat Methods Med Res* 2021; **30**: 1554–1572.
41. Sun R, Zhou X and Song X. Bayesian causal mediation analysis with latent mediators and survival outcome. *Struct Equ Modeling* 2021; **28**: 778–790.
42. He H, Cai J, Song X, et al. Analysis of proportional mean residual life model with latent variables. *Stat Med* 2017; **36**: 813–826.
43. He H, Pan D, Song X, et al. Additive mean residual life model with latent variables under right censoring. *Stat Sin* 2019; **29**: 47–66.
44. Ouyang M, Wang X, Wang C, et al. Bayesian semiparametric failure time models for multivariate censored data with latent variables. *Stat Med* 2018; **37**: 4279–4297.
45. Tang Y, Song X and Yi GY. Bayesian analysis under accelerated failure time models with error-prone time-to-event outcomes. *Lifetime Data Anal* 2022; **28(1)**: 139–168.
46. Lipsitz SR and Ibrahim JG. Estimating equations with incomplete categorical covariates in the Cox model. *Biometrics* 1998; **54**: 1002–1013.
47. Chen HY and Little RJ. Proportional hazards regression with missing covariates. *J Am Stat Assoc* 1999; **94**: 896–908.

48. Herring AH and Ibrahim JG. Likelihood-based methods for missing covariates in the Cox proportional hazards model. *J Am Stat Assoc* 2001; **96**: 292–302.
49. Herring AH, Ibrahim JG and Lipsitz SR. Non-ignorable missing covariate data in survival analysis: A case-study of an International Breast Cancer Study Group trial. *J R Stat Soc Ser C Appl Stat* 2004; **53**: 293–310.
50. Song X and Wang CY. Time-varying coefficient proportional hazards model with missing covariates. *Stat Med* 2013; **32**: 2013–2030.