



Published in final edited form as:

Stat Methods Med Res. 2019 July ; 28(7): 2112–2124. doi:10.1177/0962280217748675.

BSMHM2: Bayesian Semiparametric Mixed Hidden Markov Models for Delineating the Pathology of Alzheimer’s Disease

Kai Kang,

Department of Statistics, Chinese University of Hong Kong, Hong Kong, China

Jingheng Cai,

Department of Statistics, Sun Yat-sen University, Guangzhou, China

Xinyuan Song,

Department of Statistics, Chinese University of Hong Kong, Hong Kong, China

Hongtu Zhu, and

MD Anderson Cancer Center, University of Texas, Houston, USA

for the Alzheimer’s Disease Neuroimaging Initiative

Abstract

Alzheimer’s disease (AD) is a firmly incurable and progressive disease. The pathology of AD usually evolves from cognitive normal (CN), to mild cognitive impairment (MCI), to AD. The aim of this paper is to develop a Bayesian semiparametric mixed hidden Markov modeling (BSMHM2) framework to characterize disease pathology, identify hidden states corresponding to the diagnosed stages of cognitive decline, and examine the dynamic changes of potential risk factors associated with the CN-MCI-AD transition. The BSMHM2 framework consists of two major components. The first one is a state-dependent semiparametric regression for delineating the complex associations between clinical outcomes of interest and a set of prognostic biomarkers across neurodegenerative states. The second one is a parametric transition model, while accounting for potential covariate effects on the cross-state transition. The inter-individual and inter-process differences are taken into account via correlated random effects in both components. Based on the Alzheimer’s Disease Neuroimaging Initiative dataset, we are able to identify four states of AD pathology, corresponding to common diagnosed cognitive decline stages, including CN, early MCI, late MCI, and AD and examine the effects of hippocampus, age, gender, and APOE- $\epsilon 4$ on degeneration of cognitive function across the four cognitive states.

Keywords

Bayesian P-splines; Correlated random effects; Hidden Markov models; MCMC methods; Semiparametric models

1 Introduction

Alzheimer’s disease (AD) is a chronic neurodegenerative disease that usually starts slowly and worsens over time. The most common early symptom of AD is short-term memory loss, also referred to mild cognitive impairment (MCI). Patients at MCI state have high likelihood

to transit to dementia or AD within a few years (Albert et al., 2011). Despite an increasing attention to its growing public threat, the cause of AD remains poorly understood. Thus, it is great interest to discovering or validating prognostic biomarkers that may identify subjects at great risk for future cognitive decline and investigating the functional effects of various biomarkers on the conversion from NC to AD.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) study began in 2004 and it collected imaging, generic, clinical, and cognitive data from subjects under cognitive normal (CN) controls and subjects with MCI or AD in order to delineate the complex associations among various characteristics of the clinical spectrum of AD. The ADNI-1 recruited approximately 800 subjects according to its initial aim and has been extended by three follow-up studies, namely, ADNI-GO, ADNI-2, and ADNI-3. ADNI-1 subjects had an option to refuse follow-up monitoring in subsequent studies. More information on ADNI can be obtained in the official website (www.adni-info.org). Functional assessment questionnaire (FAQ), an assessment of abilities to function independently in daily life, is widely used to monitor the decline of cognitive ability over time. The FAQ scores of each subject were obtained at baseline and then every 6 months across 9 years multiple study phases. For this longitudinal study, several central questions are naturally raised as follows:

- (i) How many hidden pathophysiological states exist in the progression of AD?
- (ii) Which factors should contribute to the neuro-degenerative pathology from one state (e.g., MCI) to another (e.g., AD)?
- (iii) Whether the identified risk factors are equally good predictors of cognitive decline at each state?

Given these questions, there is a particular need for the development of statistical models that delineate cognitive decline in terms of the pathophysiological states of AD.

Hidden Markov models (HMM) are well suited to the characterization of longitudinal data in terms of a set of hidden states (Cappé et al., 2005; Maruotti, 2011; Bartolucci et al., 2013). HMMs consist of two components: a transition model to describe the dynamic transition of hidden states and a conditional regression model to examine state-specific covariate effects on responses. Owing to their ability to simultaneously reveal the longitudinal association structure and dynamic heterogeneity of the observed process, HMMs and their variants have attracted significant attention from the medical, behavioral, social, environmental, and psychological sciences (Vermunt et al., 1999; Scott et al., 2005; Schmittmann et al., 2005; Bartolucci and Farcomeni, 2009; Bartolucci et al., 2013; Chow et al., 2013). In particular, HMMs have previously been applied to investigate diseases progression to identify latent pathophysiological states. For instance, Albert et al. (1994) used HMMs to analyze multiple sclerosis disease across relapse and remission states (see also Altman and Petkau, 2005; Altman, 2007). Ip et al. (2013) identified ten disable states on the basis of a 10-year follow-up study of late-life disability in elder adults, and examined the patterns and risk factors for transition among disable states. Song et al. (2016) revealed the dynamic change of treatment effectiveness in preventing cocaine use across three cocaine addiction states.

Despite the rapid development and wide applications of HMMs, existing literature has mainly focused on parametric HMMs, in which the forms of covariate effects on responses and on transition probabilities are pre-specified. One problem of parametric models is that they may be too restrictive to reflect correctly the reality because the complex relationships among variables are seldom known a priori, and a pre-specified parametric form tends to overlook the subtle pattern of a function. A more comprehensive analysis can be performed by incorporating nonparametric functions into HMMs so that the functional effects of interest can be discovered. To the best of our knowledge, however, such nonparametric modeling has not been introduced into the HMM framework.

In this study, we propose a Bayesian mixed semiparametric hidden Markov modeling (BMSHM2) framework to analyze the ADNI-I dataset. Similar to conventional HMMs, the proposed model consists of two major components. The first component is a state-dependent semiparametric regression to investigate the linear and nonlinear effects of covariates, such as hippocampus, age, gender, and APOE- $\epsilon 4$, on the clinical outcome of cognitive decline (e.g., FAQ score). The second component is a mixed continuation-ratio logit transition model to examine various covariate effects on the probabilities of transitioning among neurodegenerative states. We introduce a random effect in both models in order to account for inter-individual differences and allow the random effects to be dependent by assigning a joint distribution for them. Such joint random effects enable the model to accommodate the situation where some omitted factors influence both the observed process and the hidden transition process (Wulfsohn and Tsiatis, 1997; Chi and Ibrahim, 2006). We develop a full Bayesian approach, along with Bayesian P-splines procedure and Markov chain Monte Carlo (MCMC) methods, for statistical inference. As far as we know, no previous study has ever been conducted either on the proposed BMSHM2 or on Bayesian HMMs. Also, this paper is the first to investigate the neurodegenerative pathology of AD.

Section 2 defines BMSHM2 and discusses the related identifiability issues. Section 3 introduces the Bayesian inference procedure. Section 4 illustrates the use of BMSHM2 in the analysis of the ADNI dataset. Section 5 demonstrates the empirical performance of the proposed methodology through a simulation study. Section 6 discusses the findings obtained from the analysis of the ADNI dataset. Technical details are provided in the Appendix.

2 Model description

2.1 Questions of Interest for ADNI-1

Data used in this article were obtained from the ADNI-1 database launched in 2003. A total of $n = 633$ patients at baseline, 6 months, 12 months, and 24 months ($t = 1, \dots, 4$) were considered in the analysis. We use the score of FAQ, denoted by y_{it} , to characterize the cognitive function of subject i at occasion t . Moreover, we observe a $r \times 1$ vector of discrete covariates, denoted by $\mathbf{b}_{it} = (b_{it,1}, \dots, b_{it,r})^T$, and a $q \times 1$ vector of continuous covariates, denoted by $\mathbf{x}_{it} = (x_{it,1}, \dots, x_{it,q})^T$. The covariates of interest include gender (1 = male; 0 = female), apolipoprotein E- $\epsilon 4$ (APOE- $\epsilon 4$), hippocampus, and age at baseline, in which APOE- $\epsilon 4$ is a known genetic risk factor for AD and is coded as 0, 1, and 2, denoting the number of APOE- $\epsilon 4$ alleles, and hippocampal volume is divided by whole brain volume to account for the confounding effect of brain size. Thus, APOE- $\epsilon 4$ (=1, $b_{it,1}$), APOE- $\epsilon 4$ (=2,

$b_{it,2}$), and gender ($b_{it,3}$) are discrete, whereas hippocampus ($x_{it,1}$) and age at baseline ($x_{it,2}$) are continuous.

Several kinds of dependencies/heterogeneities are worthy of investigation. The first one is the dynamic heterogeneity across different groups. Figure 1 plots the individual trajectories of FAQ scores for 20 randomly selected samples, who were initially diagnosed as CN, MCI, and AD, respectively, at baseline. The cognitive decline patterns are apparently distinct over the groups, suggesting at least three (and probably more) distinct neurodegenerative states existent underlying the observations of FAQ score. The second one is the dependency of FAQ score on potential covariates, such as hippocampus, age at baseline, APOE- $\epsilon 4$, and gender. The third one is the serial dependency of the longitudinal observations, owing to relative persistence of neurodegenerative states. The last one is the heterogeneity caused by the existence of some omitted clinical or genetic indicators that influence both cognitive decline and its underlying transition. The BMSHM2 described below perfectly accommodates all these features.

2.2 Model Setup

The BMSHM2 consists of two major components, including a conditional semiparametric regression model and a continuation-ratio logit transition model, as detailed below.

2.2.1 Conditional semiparametric regression model—Let y_{it} with subject $i = 1, \dots, n$ at $t = 1, \dots, T$ be the observation process. The hidden state process, Z_{it} , which takes values in $\{1, \dots, S\}$, is assumed to follow a first-order Markov chain. Given the hidden state $Z_{it} = s$, the conditional semiparametric regression model is defined as follows:

$$[y_{it}|Z_{it} = s] = \mu_s + \boldsymbol{\gamma}_s^T \mathbf{b}_{it} + \sum_{j=1}^q f_{sj}(x_{it,j}) + w_{i1} + \delta_{it} \quad (1)$$

where μ_s is a state-specific intercept, $\boldsymbol{\gamma}_s = (\gamma_1, \dots, \gamma_r)$ is a state-specific vector of fixed effect of discrete covariates, $f_{sj}(\cdot)$ s are state-specific unknown smoothing functions, $\mathbf{b}_{it} = (b_{it,1}, \dots, b_{it,r})^T$ and $\mathbf{x}_{it} = (x_{it,1}, \dots, x_{it,q})^T$ are $r \times 1$ vector of discrete covariates and $q \times 1$ vector of continuous covariates respectively, w_{i1} is a subject-specific random effect, δ_{it} is a random residual independent of y_{it} , and $[\delta_{it}|Z_{it} = s] \sim N[0, \boldsymbol{\psi}_s]$.

The conditional model defined by (1) extends the parametric regression to allow the additive nonparametric functions of covariates, so that the functional effects of interest can be discovered. Such nonparametric modeling provides great flexibility in fitting nonlinear effects whose forms need not be specified a priori. When used as an exploratory tool, the proposed model is able to help users to visually examine and interpret the functional effects of potential predictors on the response of interest. Moreover, the subject-specific random effect w_{i1} permits additional dependencies elicited from other sources and thus avoids a large number of hidden states caused by possible residual correlation among responses.

2.2.2 Continuation-ratio logit transition model—Let p_{itus} denote the transition probability from state $Z_{i,t-1} = u$ at occasion $t-1$ to state $Z_{it} = s$ at occasion t for individual i . Based on the assumption of the first-order Markov chain, we have

$$p_{itus} = P(Z_{it} = s | Z_{i1}, Z_{i2}, \dots, Z_{i,t-1} = u) = P(Z_{it} = s | Z_{i,t-1} = u). \quad (2)$$

The initial distribution of Z_{i1} is assumed to be a multinomial with probabilities $(\tau_1, \dots, \tau_S)^T$ such that $\tau_s \geq 0$ and $\sum_{s=1}^S \tau_s = 1$. The distribution of $\{Z_{it}\}_{t=1}^T$ is then fully determined by the transition probabilities and the distribution of the initial state.

In the study of disease progression, the hidden states can often be naturally ranked (e.g., CN, MCI, and AD can be ranked from the best to the worst cognitive condition). Thus, we assume that the hidden states $\{1, \dots, S\}$ are ordered and $\vartheta_{itus} = P(Z_{it} = s | Z_{it} = s, Z_{i,t-1} = u)$. Then, the transition across the ordered states can be described by continuation logits as follows: For $t = 2, \dots, T$, $s = 1, \dots, S-1$, and $u = 1, \dots, S$,

$$\log \left(\frac{P(Z_{it} = s | Z_{i,t-1} = u)}{P(Z_{it} > s | Z_{i,t-1} = u)} \right) = \log \left(\frac{p_{itus}}{p_{iu,s+1} + \dots + p_{iuS}} \right) = \text{logit}(\vartheta_{itus}). \quad (3)$$

The parameterization in (3) is intended to facilitate the interpretation of transition to a state rather than to a better one. To examine the effects of potential predictors on the transition probabilities, we consider a continuation-ratio logit transition model as follows:

$$\text{logit}(\vartheta_{itus}) = \zeta_{us} + \boldsymbol{\alpha}^T \mathbf{d}_{it} + w_{i2}, \quad (4)$$

where ζ_{us} is a state-specific intercept, $\mathbf{d}_{it} = (\mathbf{x}_{it}^T, \mathbf{b}_{it}^T)^T$ is the vector of covariates defined in (1), $\boldsymbol{\alpha}$ is a $(q+r) \times 1$ vector of regression coefficients that can be interpreted as conditional log odds ratios in a logistic regression, w_{i2} is a subject-specific random effect that is distinct from but correlated with w_{i1} , and $\mathbf{w}_i = (w_{i1}, w_{i2})^T$ is assumed to follow a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Phi})$. Similar to the proportional assumption in a cumulative logit model, $\boldsymbol{\alpha}$ in (4) is assumed to be independent of u and s in order to maintain the order of the hidden states and avoid a tedious transition model, in which every transition elicits a set of parameters for all possible states of origination and destination. This outcome, in turn, greatly reduces the complexity and enhances the interpretability of the transition model.

Notably, random effects w_{i1} and w_{i2} play different roles in the conditional and transition models. While w_{i1} in conditional model (1) relaxes the assumption that observations $\{y_{it}; i = 1, \dots, n, t = 1, \dots, T\}$ are conditionally independent given the hidden state $Z_{it} = s$, w_{i2} in transitional model (4) releases the assumption that hidden process Z_{it} is Markovian. Unlike the existing literature that usually treats w_{i1} and w_{i2} separately, we accommodate their possible correlation by assigning a joint distribution for $\mathbf{w}_i = (w_{i1}, w_{i2})^T$. Consequently, the

possible correlation between the heterogeneities existent within the two stochastic processes can be appropriately addressed and examined through the covariance matrix Φ .

2.3 Model identification

The proposed model is not identifiable because of the following two model indeterminacies. The first is caused by the additive nonparametric functions involved in (1), in which each unknown function is not identifiable up to a constant. To address this problem, we need to impose constraints on the unknown functions to enforce their integrations in the ranges of predictors to zero (Panagiotelis and Smith, 2008; Song and Lu, 2010) as follows:

$$\int_{\mathcal{X}_j} f_{sj}(x)dx = 0, \quad \text{for } s = 1, \dots, S, \quad j = 1, \dots, q, \quad (5)$$

where \mathcal{X}_j is the support of \mathbf{x}_j . The other model determinacy is the label switching problem elicited by the invariance of the likelihood function to a random permutation of the state labels, which results in a multi-modal posterior under a symmetric prior specification. We follow Frühwirth-Schnatter (2001) to conduct a permutation sampler to address this issue.

3 Bayesian Inference

3.1 Nonparametric modeling

The first critical issue in the Bayesian analysis of the proposed model is to estimate the nonparametric functions involved in (1). We consider the use of Bayesian P-splines (Berry et al., 2002; Lang and Brezger, 2004; Fahrmeir and Raach, 2007). The basic idea is to estimate the unknown smoothing functions through a sum of B-splines basis functions (De Boor, 2001) given a large number of knots in the domains of predictors. Specifically, $f_{sj}(x_{it,j})$, the functional effect of the j th covariate at state s for subject i at time t , can be approximated as follows:

$$f_{sj}(x_{it,j}) = \sum_{l=1}^L \beta_{sj,l} B_l(x_{it,j}) = \boldsymbol{\beta}_{sj}^T \mathbf{B}(x_{it,j}), \quad (6)$$

where L is the number of splines determined by the number of knots, $\boldsymbol{\beta}_{sj} = (\beta_{sj,1}, \dots, \beta_{sj,L})^T$ is the vector of the unknown parameters, $B(\cdot)$'s are cubic B-splines basis functions, and $\mathbf{B}(x_{it,j}) = (B_1(x_{it,j}), \dots, B_L(x_{it,j}))^T$. Usually, L taking a value from 10 to 30 provides sufficient flexibility in fitting most smooth functions.

One problem of applying (6) to approximate an unknown smooth function is the over-fitting caused by the use of a large number of knots. Eilers and Marx (1996) suggested the penalization of the coefficients of adjacent B-splines basis functions to prevent the overfitting. Such penalization can be implemented in the Bayesian framework by applying random walk priors to $\boldsymbol{\beta}_{sj}$ (Lang and Brezger, 2004; Fahrmeir and Raach, 2007; Song and Lu, 2010).

3.2 Prior distributions

We assign a truncated Gaussian priors for β_{sj} as follows:

$$p(\beta_{sj} | \nu_{sj}) = \left(\frac{1}{2\pi\nu_{sj}} \right)^{L_{sj}/2} \exp \left\{ -\frac{1}{2\nu_{sj}} \beta_{sj}^T \mathbf{K}_{sj} \beta_{sj} \right\} \mathbf{I}(\mathbf{1}_{n_s}^T \mathbf{B}_{sj} \beta_{sj} = 0), \quad (7)$$

where ν_{sj} is a smoothing parameter for controlling the amount of penalty, \mathbf{K}_{sj} is a penalty matrix derived according to the random walk penalties proposed, L_{sj} is the rank of \mathbf{K}_{sj} , $\mathbf{1}_{n_s}$ is an $n_s \times 1$ vector with all elements equal to 1, n_s is the sample size at state s , \mathbf{B}_{sj} is the sub-matrix of $\mathbf{B}_j = [B(x_{it}, j)]_{nT \times L}$ without the rows where $Z_{it} = s$, and the truncation term incorporates the identifiability constraint (5) into the splines approximation (6).

For the smoothing parameters ν_{sj} , we assign a highly dispersed but proper inverse gamma prior as follows:

$$p(\nu_{sj}^{-1}) \stackrel{D}{=} \text{Gamma}[\nu_1, \nu_2], \quad (8)$$

where ν_1 and ν_2 are hyperparameters whose values are pre-specified. A common choice for these hyperparameters is $\nu_1 = 1$ and ν_2 is small (Fahrmeir and Raach, 2007; Song and Lu, 2010). We set $\nu_1 = 1$ and $\nu_2 = 0.005$ in the present study.

For the parameters involved in conditional model (1), conjugate-type priors are assigned as follows: for $s = 1, \dots, S$,

$$p(\mu_s) \stackrel{D}{=} N[\mu_{s0}, \sigma_{\mu s0}^2], \quad p(\gamma_s) \stackrel{D}{=} N[\gamma_{s0}, \Sigma_{s0}], \quad (9)$$

$$p(\Phi^{-1}) \stackrel{D}{=} \text{Wishart}[\mathbf{R}_0, \rho_0], \quad p(\psi_s^{-1}) \stackrel{D}{=} \text{Gamma}[\tilde{\alpha}_{s0}, \tilde{\beta}_{s0}],$$

where μ_{s0} , $\sigma_{\mu s0}^2$, γ_{s0} , Σ_{s0} , $\tilde{\alpha}_{s0}$, $\tilde{\beta}_{s0}$, \mathbf{R}_0 , and ρ_0 are hyperparameters with preassigned values.

Finally, for the parameters involved in transition model (4), we consider the following Gaussian priors:

$$p(\zeta_{us}) \stackrel{D}{=} N[\zeta_{us0}, \sigma_{\zeta 0}^2], \quad p(\alpha) \stackrel{D}{=} N[\alpha_0, \mathbf{H}_{\alpha 0}], \quad p(\tau_s) \stackrel{D}{=} N[\tau_{s0}, \sigma_{\tau 0}^2], \quad (10)$$

where ζ_{us0} , $\sigma_{\zeta 0}^2$, α_0 , $\mathbf{H}_{\alpha 0}$, τ_{s0} , and $\sigma_{\tau 0}^2$ are hyperparameters with preassigned values.

3.3 Posterior computation

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})^T$, $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$, $\mathbf{D} = (\mathbf{d}_{11}, \dots, \mathbf{d}_{NT})$, $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iT})^T$, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_N)$, $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_N)$, and $\boldsymbol{\theta}$ be the vector that includes all the unknown parameters in the proposed model. The complete-data log-likelihood function that is used to derive the posterior distributions and compute the model selection criterion is given by

$$\begin{aligned} \log p(\mathbf{Y}, \mathbf{D}, \mathbf{W}, \mathbf{Z} | \boldsymbol{\theta}) &= \sum_{i=1}^n [\log p(\mathbf{y}_i | \mathbf{d}_i, \mathbf{w}_i, \mathbf{Z}_i, \boldsymbol{\theta}) + \log p(\mathbf{Z}_i | \mathbf{d}_i, \mathbf{w}_i, \boldsymbol{\theta}) + \log p(\mathbf{w}_i | \boldsymbol{\theta})] \quad (11) \\ &= \sum_{i=1}^n \sum_{t=1}^T \log p(y_{it} | \mathbf{d}_{it}, w_{i1}, Z_{it} = s, \boldsymbol{\theta}) + \sum_{i=1}^n \sum_{t=2}^T \log p(Z_{it} = s | Z_{i,t-1} = u, \mathbf{d}_{it}, w_{i2}, \boldsymbol{\theta}) \\ &+ \sum_{i=1}^n \log p(Z_{i1} = s | \boldsymbol{\theta}) + \sum_{i=1}^n \log p(\mathbf{w}_i | \boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^n \sum_{t=1}^T [\log(2\pi\psi_s) \\ &+ (y_{it} - \mu_s - \gamma_s^T \mathbf{b}_{it} - \sum_{j=1}^q \beta_{sj}^T \mathbf{B}(x_{it,j}) - w_{i1})^2 / \psi_s] + \sum_{i=1}^n \sum_{t=2}^T \log(p_{itus}) \\ &+ \sum_{i=1}^n \log(p_{i10s}) - \frac{1}{2} \sum_{i=1}^n [\log(4\pi^2 |\boldsymbol{\Phi}|) + \mathbf{w}_i^T \boldsymbol{\Phi}^{-1} \mathbf{w}_i], \end{aligned}$$

where

$$p_{itu1} = \frac{\exp\{a_{itu1}\}}{1 + \exp\{a_{itu1}\}}, \quad p_{i1uS} = \prod_{j=1}^{S-1} \frac{1}{1 + \exp\{a_{ituj}\}},$$

$$p_{itus} = \frac{\exp\{a_{itus}\}}{1 + \exp\{a_{itus}\}} \prod_{j=1}^{s-1} \frac{1}{1 + \exp\{a_{ituj}\}}, \quad s = 2, \dots, S-1, \quad (12)$$

$$p_{i10s} = \tau_s, \quad s = 1, \dots, S,$$

with $a_{itus} = \zeta_{us} + \boldsymbol{\alpha}^T \mathbf{d}_{it} + w_{i2}$.

The Bayesian estimate of $\boldsymbol{\theta}$ is obtained by drawing samples from $p(\boldsymbol{\theta} | \mathbf{Y})$, which is intractable because of the existence of latent states and random effects. We instead work on $p(\boldsymbol{\theta}, \mathbf{Z}, \mathbf{W} | \mathbf{Y})$ and use a Gibbs sampler to implement the posterior simulation. Owing to the nonlinearity of the continuation-logit transition model and the existence of the nonparametric functions in the conditional regression, some full conditional distributions, especially those related to the transition model, have complex forms. MCMC methods, such as the forward filtering and backward sampling algorithm (Cappé et al., 2005) and the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970), are employed to sample from them. The details are provided in the Appendix.

With the use of posterior samples, the hidden states can be estimated as follows:

$$\hat{Z}_{it} = \arg \max_{s \in \{1, \dots, S\}} P(Z_{it} = s | y_i, \theta) \approx \arg \max_{s \in \{1, \dots, S\}} \frac{1}{M} \sum_{m=1}^M I(Z_{it}^{(m)} = s), \quad (13)$$

where $Z_{it}^{(m)}$ denotes the latent allocation of y_{it} at the m th iteration, and $\frac{1}{M} \sum_{m=1}^M I(Z_{it}^{(m)} = s)$ is the posterior mean of the latent allocations of y_{it} drawn from the MCMC iterations.

3.4 Determination of the number of hidden states

In the applications of BMSHM2 to the ADNI dataset, the states of the Markov chain can often naturally be interpreted as proxies for the neurodegenerative states, although a one-to-one correspondence between nominal HMM states and the clinical cognitive stages diagnosed by doctors is unnecessary. In this regard, a relevant question is how to determine the number of hidden states in the analysis of ADNI data. We propose the use of a modified deviance information criteria (DIC) to determine the number of hidden states and choose a plausible model for the ADNI data analysis.

The modified DIC, which was developed by Celeux et al. (2006) for model comparison in the presence of incomplete data, is defined as follows:

$$\text{DIC} = -4E_{\theta, \mathbf{W}, \mathbf{Z}}\{\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z} | \theta) | \mathbf{Y}\} + 2E_{\mathbf{W}, \mathbf{Z}}\{\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z} | E_{\theta}[\theta | \mathbf{Y}, \mathbf{W}, \mathbf{Z}]) | \mathbf{Y}\}. \quad (14)$$

where $\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z} | \theta)$ is the complete-data log-likelihood function shown in (11). The expectations involved in (14) can be approximated using the posterior samples collected through MCMC methods (Celeux et al., 2006). In model selection, the model with the smallest value of DIC is selected.

4 Alzheimer's Disease Neuroimaging Initiative Data Analysis

4.1 Data description

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a \$60 million, 5-year public private partnership. The initial goal of ADNI was to recruit 800 adults, aged 55 to 90, to participate in the research – approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years.

We focused on 633 subjects who were all followed up at baseline, 6 months, 12 months, and 24 months. For each subject, we included his/her clinical, genetic, and imaging variables at the four time points. The clinical characteristics include gender (0 = male; 1 = female), age at baseline, and FAQ score. The FAQ score is an assessment of abilities to function independently in daily life and is widely used to monitor the decline of cognitive ability over

time. The genetic variables include APOE gene because mutations in APOE raise the risk of progression from amnesic MCI to AD (Petersen et al., 2005). The APOE SNPs, rs429358 and rs7412 were separately genotyped in ADNI-1. These two SNPs together define a 3-allele haplotype, namely, the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ variants. Among these variants, APOE- $\epsilon 4$ has been identified as a risk factor for early onset of AD (e.g., Okuizumi et al., 1994). Thus, we considered the presence of APOE- $\epsilon 4$ as a covariate in this analysis. APOE- $\epsilon 4$ is coded as 0, 1, and 2, denoting the number of APOE- $\epsilon 4$ alleles. Furthermore, the logarithm of the ratio of hippocampal volume over whole brain volume was included as a covariate because published reports (Kesslak, Nalcioglu and Cotman, 1991; Jack et al., 1992; Dickerson and Wolk, 2013) revealed that the atrophy of the hippocampal formation was a significant diagnostic marker of clinical dementia. Table 1 summarizes the basic characteristics of the aforementioned variables for the samples under consideration. Males account for about 56.2% in the samples. Mean values of patients' age (in year), adjusted hippocampal volume, and corresponding FAQ score are 73.0, -5.0 , and 4.1, respectively. 34.6% patients carry one APOE- $\epsilon 4$ allele while only 9.8% carry two APOE- $\epsilon 4$ alleles.

4.2 Data analysis

The aims of this ADNI data analysis are (I) to identify the hidden states of the neurodegenerative pathology on the basis of 633 patients enrolled in the ADNI-1, (II) to reveal a set of potential covariates that influence the between-states transition, and (III) to investigate the linear and/or functional covariate effects on cognitive decline across the hidden states of the AD progression.

We fitted BMSHM2 with the FAQ score as the response y_{it} , the clinical and genetic variables, gender and APOE- $\epsilon 4$, as covariates in \mathbf{b}_{it} , and hippocampus and age at baseline as covariates in \mathbf{x}_{it} . Three continuous variables, FAQ score, hippocampus, and age, were standardized prior to analysis. We first determined the number of hidden states. We considered five competing models M_k , $k = 1, \dots, 5$, where M_k represents a BMSHM2 with k states. A total of 24 equidistant knots were used to construct cubic P-splines, and the second-order random walk penalties were used for the Bayesian P-splines to estimate the unknown smooth functions. Given the lack of prior information, we assign the hyperparameters in (9) and (10) to reflect vague prior information as follows: $\mu_{s0} = \zeta_{us0} = \tau_{s0} = 0$, $\sigma_{\mu s0} = \sigma_{\zeta s0}^2 = \sigma_{\tau s0}^2 = 1$, $\tilde{\alpha}_{s0} = 9$, $\tilde{\beta}_{s0} = 4$, $\rho_0 = 7$, $\mathbf{R}_0 = 4\mathbf{I}_2$, \mathbf{a}_0 and $\boldsymbol{\gamma}_{s0}$ is a vector with all elements being zero, $\mathbf{H}_\alpha = \mathbf{I}_5$ and $\boldsymbol{\Sigma}_{s0} = \mathbf{I}_3$ where \mathbf{I}_r is a r -dimensional identity matrix. We used the random permutation sampler to search for a suitable identifiability constraint to solve the label switching problem. The MCMC algorithm converged within 2,000 iterations for all competing models. We collected a total of 10,000 observations after discarding 2,000 burn-in iterations to calculate DIC. The DIC values corresponding to M_1 to M_5 were 20,175, 1,823, 1,001, 950, and 1615, respectively. Thus, the four-state model M_4 was selected.

To examine the necessity of the random effects in the proposed model, we considered another competing model M_N : a four-state BMSHM2 without random effects. The DIC value for M_N is 1,122, which suggests an evident advantage of the proposed mixed effect model in the presence of high dependency/heterogeneity in longitudinal observations. Thus,

M_4 was selected for the subsequent analysis. The estimation results obtained under M_4 are reported in Table 2 (parametric part) and Figure 2 (nonparametric part).

We have the following observations. First, μ_1 , μ_2 , μ_3 , and μ_4 are ranked in an ascending order, indicating that patients in state 1 got the lowest score of FAQ, whereas those in state 4 got the highest. That is, patients' ability to function independently in daily life steadily deteriorated from state 1 to state 4. According to the existing literature (Kantarci et al., 2013), state 1 to state 4 can be explained as cognitive normal (CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD, respectively.

Second, the functional effect of hippocampus on FAQ exhibits a descending trend as hippocampus grows regardless of states. Specifically, in CN state, people with bigger hippocampus volume tend to have slightly better memory. This finding is in line with the common sense that hippocampus plays an important role in the consolidation of information from short-term memory to long-term memory. In EMCI and LMCI states, the descending trend of the functional effect of hippocampus on FAQ becomes much more pronounced. This result implies that atrophy in hippocampus increasingly impairs patients' cognitive ability. The published reports (e.g., Dickerson and Wolk, 2013; Kesslak, Nalcioglu and Cotman, 1991; Jack et al., 1992) also indicate that the volume loss of the hippocampus is greatly associated with clinical dementia. In AD state, the effect of hippocampal volume on FAQ is not significant because patients' cognitive ability and memory have already been damaged by serious hippocampus atrophy.

Third, the effect of age on FAQ is nonsignificant in CN and EMCI states, implying that for those who have cognitive normal function or undergo only EMCI, age may not be a decisive factor for the reduction of cognitive ability. On the contrary, age exhibits nonlinear effects on FAQ in LMCI and AD states. The age effect is nonsignificant for relatively young patients but becomes significantly positive for elderly patients (say, over 85 years old). The positive effect increases with age and gets an even sharper rise in AD state. Such age effect was also discovered by previous studies (e.g. Gao et al., 1998; Lindsay et al., 2002).

Fourth, as for the fixed effects of discrete variables, a significantly negative effect of gender on FAQ appears in state 4, which means that male AD patients are in a better condition than females in terms of independent abilities. The two APOE- $\epsilon 4$ alleles ($b_{it,1}$ and $b_{it,2}$) have significantly positive effects on FAQ in state 4 and $b_{it,1}$ also has a slightly positive effect on FAQ in other states. This finding agrees with the newly published clinical research output that the presence of $\epsilon 4$ alleles in the APOE gene is the only genetic variant broadly accepted as increasing risk for late-onset AD dementia (Albert et al., 2011).

Fifth, in the transition model, hippocampus positively affects the probability of transitioning from a state to a better one, indicating that controlling loss of hippocampal volume would be beneficial to prevent the deterioration of cognitive ability. Similar to previous studies (e.g., Lee et al., 2015), APOE- $\epsilon 4$ alleles have negative effects on the probability of transitioning from a state to a better one, reconfirming that APOE- $\epsilon 4$ alleles are important risk factors for the development of AD.

Sixth, the variances of the two random effects are significant, reconfirming the necessity of the random effects proposed. However, the covariance between the two random effects is nonsignificant, showing that some omitted clinical or genetic indicators influenced outcomes of the observation process or probabilities of the transition process but did not affect the two processes simultaneously.

Moreover, we estimated the hidden states of all patients at four time points based on Equation (13). Around 98% posterior transition patterns are from a state to a severer one, which is in line with the common knowledge of irreversibility of AD. Table 3 reports patients' estimated hidden states and their diagnosed status by doctors. For CN, LMCI, and AD states, a majority of the estimated states are consistent with those diagnosed by doctors. For EMCI state, however, 835 (67%) EMCI patients diagnosed by doctors were classified into CN state by our procedure. Such vague demarcation between CN and EMCI was also found and discussed in the literature (e.g., Petersen, 2004).

5 Simulation Study

We conduct Monte Carlo simulations to assess the empirical performance of the proposed method in estimation of the nonparametric functions and model parameters.

5.1 Model setup

We consider a BMSHM2 with four hidden states ($S = 4$), a continuous response y_{it} , three discrete covariates ($r = 3$), and two continuous covariates ($q = 2$) to mimic the scenario of the ADNI study. For $i = 1, \dots, 700$ and $t = 1, \dots, 9$, $b_{it,1}$, $b_{it,2}$, and $b_{it,3}$ are all generated from the Bernoulli distribution with the probability of success 0.5, and $x_{it,1}$ and $x_{it,2}$ are generated from $U(-1, 1)$ and $N(0, 1)$, respectively. The conditional model is defined as

$$[y_{it} | Z_{it} = s] = \mu_s + \gamma_{s1}b_{it,1} + \gamma_{s2}b_{it,2} + \gamma_{s3}b_{it,3} + f_{s1}(x_{it,1}) + f_{s2}(x_{it,2}) + w_{i1} + \delta_{it}, \quad (15)$$

where $f_{11}(x_{it,1}) = -1.305 + \exp(x_{it,1})$, $f_{12}(x_{it,2}) = 0.55 + \sin(1.5x_{it,2}) + x_{it,2}$, $f_{21}(x_{it,1}) = 0.06 - \log((1 + x_{it,1})/(1 - x_{it,1}))$, $f_{22}(x_{it,2}) = 0.125 + x_{it,2}^3$, $f_{31}(x_{it,1}) = -0.05 - 0.8x_{it,1}$, $f_{32}(x_{it,2}) = -0.275 + \cos(2x_{it,2}) + 0.5x_{it,2}$, $f_{41}(x_{it,1}) = -0.13 - x_{it,1}^3$ and $f_{42}(x_{it,2}) = -0.85 + 1.5x_{it,2}$.

The transition model is defined as

$$\text{logit}(\theta_{itus'}) = \zeta_{us} + \alpha_1 x_{it,1} + \alpha_2 x_{it,2} + \alpha_3 b_{it,1} + \alpha_4 b_{it,2} + \alpha_5 b_{it,3} + w_{i2}. \quad (16)$$

The true population values of the unknown parameters are set as $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4) = (-5, -1, 3, 7)$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4) = (0.27, 0.27, 0.23, 0.23)$, $\zeta_{11} = \zeta_{21} = \zeta_{31} = \zeta_{41} = -1$, $\zeta_{12} = \zeta_{22} = \zeta_{32} = \zeta_{42} = -1/2$, $\zeta_{13} = \zeta_{23} = \zeta_{33} = \zeta_{43} = 1/2$, $\boldsymbol{\gamma}_1 = (\gamma_{11}, \gamma_{12}, \gamma_{13}) = (-1, 0.5, 0.5)$, $\boldsymbol{\gamma}_2 = (\gamma_{21}, \gamma_{22}, \gamma_{23}) = (1, 1, 0.5)$, $\boldsymbol{\gamma}_3 = (\gamma_{31}, \gamma_{32}, \gamma_{33}) = (-0.5, -0.5, -0.5)$, $\boldsymbol{\gamma}_4 = (\gamma_{41}, \gamma_{42}, \gamma_{43}) = (0.5, -1, -1)$, $\boldsymbol{a} = (a_1, a_2, a_3, a_4, a_5)^T = (1, -1, -0.5, 0.5, 1)$, $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4) = (1,$

0.64, 0.36, 0.25), and Φ is a correlation matrix with off diagonal elements -0.5 . Based on the above setup, we simulate 100 datasets for analysis.

5.2 Simulation results

We used a total of 24 equidistant knots to construct the cubic B-splines of the covariates. Again, the second-order random walk penalties were used for the Bayesian P-splines to estimate the unknown smooth functions. The prior inputs in (9) and (10) were assigned as follows: $\mu_{s0} = \zeta_{us0} = \tau_{s0} = 0$, $\sigma_{\mu_{s0}} = \sigma_{\zeta_0}^2 = \sigma_{\tau_0}^2 = 1$, $\tilde{\alpha}_{s0} = 9$, $\tilde{\beta}_{s0} = 4$, $\rho_0 = 7$, $\mathbf{R}_0 = 4\mathbf{I}_2$, \mathbf{a}_0 and $\boldsymbol{\gamma}_{s0}$ are vectors with all elements being zero, $\mathbf{H}_\alpha = \mathbf{I}_5$ and $\boldsymbol{\Sigma}_{s0} = \mathbf{I}_3$, where \mathbf{I}_r is a $r \times r$ identity matrix. We conducted a few test runs to decide the number of burn-in iterations required for convergence and found that the MCMC algorithm converged within 2,000 iterations. Therefore, we obtain Bayesian results using 5,000 observations after discarding 2,000 burn-in iterations. The performance of the Bayesian estimates is assessed through the bias (BIAS) and the root mean square errors (RMSE) between the Bayesian estimates and the true population values of the parameters.

Table 4 summarizes the result of parameter estimation based on the 100 datasets. The BIAS and RMSE for most of the parameters are close to zero, indicating a satisfactory performance of Bayesian estimation regarding the parametric part. Figure 3 depicts the averages of the pointwise posterior means of the nonparametric functions, along with their 2.5%- and 97.5%- pointwise quantiles. The posterior means of the nonparametric functions are close to their true curves and all the ranges of the two pointwise quantiles are relatively small, indicating that the estimated functions can correctly recover the true functional relationships between the response and covariates. In this simulation, the average of the correct classification rates calculated through Equation (13) based on the 100 datasets is 91%. Considering the complexity of proposed model, this result is satisfactory.

To reveal the sensitivity of the Bayesian estimates to the input of priors, we disturbed the prior inputs as follows: $\mu_{s0} = \zeta_{us0} = \tau_{s0} = 2$, $\sigma_{\mu_{s0}} = \sigma_{\zeta_0}^2 = \sigma_{\tau_0}^2 = 2$, $\tilde{\alpha}_{s0} = 3$, $\tilde{\beta}_{s0} = 2$, $\rho_0 = 4$, $\mathbf{R}_0 = 2\mathbf{I}_2$, \mathbf{a}_0 and $\boldsymbol{\gamma}_{s0}$ are vectors with all elements being two, $\mathbf{H}_\alpha = 2\mathbf{I}_5$ and $\boldsymbol{\Sigma}_{s0} = 2\mathbf{I}_3$. The obtained results are similar and not reported.

6 Discussion

The BMSHM2 was developed and successfully applied to the ADNI data analysis. Although HMMs and their variants have already been extensively used for longitudinal data analysis, a majority of applications restrict analysis in a parametric framework. Nonetheless, examples of using HMMs to classify and characterise the neurodegenerative states of AD pathology are not prevalent, especially in a semiparametric context. In this study, we extended parametric HMMs to accommodate the functional effects of hippocampus and age on cognitive decline across four neurodegenerative states, namely, CN, EMCI, LMCI, and AD. The functional effect of hippocampus on cognitive function exhibited a descending trend as hippocampus grows regardless of states. This descending trend became more pronounced for EMCI and LMCI states than for CN and AD states, implying that atrophy in hippocampal volume had increasingly impaired patients' cognitive ability, especially during the

progression from EMCI to LMCI. On the contrary, age affected cognitive function mainly in LMCI and AD states. Elderly LMCI or AD patients suffered from more increasing neurodegeneration than relatively young patients.

Our model incorporates correlated random effects to account for individual and/or contextual differences in the progression of cognitive decline and in between-state transition. Large inter-individual variability is a prominent feature of the ADNI dataset and many other longitudinal datasets. As we demonstrated in the ADNI study, accounting for such differences can dramatically improve model fit, as evidenced by an apparent improvement in DIC value between models with and without random effects. In addition, the correlation between the random effects enhances the model capability of accommodating the situation where some omitted covariates influence both the state-dependent observation process and the hidden-state transition process. Another appealing feature of this study is that it implements a full Bayesian analysis along with efficient MCMC methods. The sampling-based Bayesian approach is not only applicable to the current parameter-rich BMSHM2 but also possesses potential to address highly complex problems with which huge challenges are confronted by ML-based procedures.

The present study can be extended in several directions: First, we considered the nonparametric modeling only in the conditional model. Generalizing the parametric transition model to a semiparametric or nonparametric one can further enhance model flexibility and analytic power. However, the statistical analysis of such comprehensive models can be challenging because the computational burden and sample size often limit the complexity of candidate models. Thus, the feasibility of this extension requires further investigation. Second, in the application to the ADNI dataset, a highly comprehensive characterization of cognitive function is to group the FAQ, Alzheimer's Disease Assessment Scale, and Mini-Mental State Examination into an integrated latent construct through multivariate techniques such as factor analysis (e.g. Song et al., 2016). Finally, this study did not consider missing data. Given that missingness is very common in longitudinal settings, accommodation of missing responses and/or missing covariates in the context of BMSHM2s is both of scientific interest and of practical value. These advances certainly require substantial efforts for further investigation.

References

- Albert PS, McFarland HF, Smith ME, Frank JA. 1994; Time series for modelling counts from a relapsingremitting disease: Application to modelling disease activity in multiple sclerosis. *Statistics in Medicine*. 13:453–466. [PubMed: 8023028]
- Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH. 2011; The diagnosis of mild cognitive impairment due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*. 7:270–279.
- Altman RM, Petkau AJ. 2005; Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine*. 24:2335–2344. [PubMed: 15909288]
- Altman RM. 2007; Mixed hidden Markov models. *Journal of the American Statistical Association*. 102:201–210.

- Ansari A, Jedidi K. 2000; Bayesian factor analysis for multilevel binary observations. *Psychometrika*. 65:475–498.
- Bartolucci F, Farcomeni A. 2009; A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*. 104:816–831.
- Bartolucci, F, Farcomeni, A, Pennoni, F. *Latent Markov Models for Longitudinal Data*. Florida: Chapman & Hall/CRC, Taylor and Francis Group; 2013.
- Baum LE, Petrie T, Soules G, Weiss N. 1970; A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*. 41:164–171.
- Behseta S, Kass RE, Wallstrom GL. 2005; Hierarchical models for assessing variability among functions. *Biometrika*. 92:419–434.
- Berry SM, Carroll RJ, Ruppert D. 2002; Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*. 97:160–169.
- Biller C, Fahrmeir L. 2001; Bayesian varying-coefficient models using adaptive regression splines. *Statistical Modelling*. 1:195–211.
- Cappé, O, Moulines, E, Rydén, T. *Inference in Hidden Markov Models*. Springer; New York: 2005.
- Celeux G, Forbes F, Robert CP, Titterton DM. 2006; Deviance information criteria for missing data models. *Bayesian Analysis*. 1:651–674.
- Chi YY, Ibrahim JG. 2006; Regime-switching bivariate dual change score model. *Biometrics*. 62:432–445. [PubMed: 16918907]
- Chow SM, Grimm KJ, Filteau G, Dolan CV, McArdle JJ. 2013; Regimes-switching bivariate dual change score model. *Multivariate Behavioral Research*. 48(4):463–502. [PubMed: 26742002]
- De Boor, C. *A Practical Guide to Splines*. Springer-Verlag; New York: 2001. revised edition
- Dickerson BC, Wolk D. 2013; Biomarker-based prediction of progression in MCI: comparison of AD-signature and hippocampal volume with spinal fluid amyloid- β and tau. *Frontiers in Aging Neuroscience*. 5:55. [PubMed: 24130528]
- DiMatteo I, Genovese CR, Kass RE. 2001; Bayesian curve fitting with free-knot splines. *Biometrika*. 88:1055–1071.
- Dunson DB. 2000; Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B*. 62:355–366.
- Dunson DB. 2003; Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*. 98:555–563.
- Eilers PHC, Marx BD. 1996; Flexible smoothing with B-splines and penalties. *Statistical Science*. 11:89–121.
- Fahrmeir L, Raach A. 2007; A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*. 72:327–346.
- Frühwirth-Schnatter S. 2001; Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*. 96:194–209.
- Gao S, Hendrie HC, Hall KS, Hui S. 1998; The relationships between age, sex, and the incidence of dementia and Alzheimer disease: a meta-analysis. *Archives of General Psychiatry*. 55:809–815. [PubMed: 9736007]
- Hastings WK. 1970; Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 57:97–109.
- Ip E, Zhang Q, Rejeski J, Harris T, Kritchevsky S. 2013; Partially ordered mixed hidden Markov model for the disablement process of older adults. *Journal of the American Statistical Association*. 108:370–384. [PubMed: 24058222]
- Jack CR, Petersen RC, O'Brien PC, Tangalos EG. 1992; MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology*. 42:183–183. [PubMed: 1734300]
- Kantarci K, Gunter JL, Tosakulwong N, Weigand SD, Senjem MS, Petersen RC, Aisen PS, Jagust WJ, Weiner MW, Jack CR. Alzheimer's Disease Neuroimaging Initiative. 2013; Focal hemosiderin deposits and -amyloid load in the ADNI cohort. *Alzheimer's & Dementia*. 9:S116–S123.

- Kass RE, Raftery AE. 1995; Bayes factors. *Journal of the American Statistical Association*. 90:773–795.
- Kesslak JP, Nalcioglu O, Cotman CW. 1991; Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. *Neurology*. 41:51–51. [PubMed: 1985296]
- Lang S, Brezger A. 2004; Bayesian P-splines. *Journal of Computational and Graphical Statistics*. 13:183–212.
- Lee E, Zhu H, Kong D, Wang Y, Giovanello KS, Ibrahim JG. 2015; BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *The Annals of Applied Statistics*. 9:2153–2178. [PubMed: 26900412]
- Lee SY, Song XY. 2004; Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*. 39:653–686. [PubMed: 26745462]
- Lindsay J, Laurin D, Verreault R, Hbert R, Helliwell B, Hill GB, McDowell I. 2002; Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. *American Journal of Epidemiology*. 156:445–453. [PubMed: 12196314]
- Maruotti A. 2011; Mixed hidden Markov models for longitudinal data: an overview. *International Statistical Review*. 79:427–454.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953; Equations of state calculations by fast computing machine. *The Journal of Chemical Physics*. 21:1087–1091.
- Myers SC, Majluf NS. 1984; Corporate financing and investment decision when firms have information that investor do not have. *Journal of Financial Economic*. 13:187–221.
- Okuizumi K, Onodera O, Tanaka H, Kobayashi H, Tsuji S, Takahashi H, Tsuji S, Takahashi H, Oyanagi K, Seki K, Tanaka M, Naruse S, Miyatake T, Mizusawa H, Kanazawa I. 1994; ApoE- ϵ 4 and earlyonset Alzheimer's. *Nature genetics*. *Journal of Financial Economic*. 7:10–11.
- Panagiotelis A, Smith M. 2008; Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *Journal of Econometrics*. 143:291–316.
- Petersen RC. 2004; Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine*. 256(3):183–194. [PubMed: 15324362]
- Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, Dyck CH, Thal LJ. 2005; Vitamin E and donepezil for the treatment of mild cognitive impairment. *New England Journal of Medicine*. 352:2379–2388. [PubMed: 15829527]
- Richardson S, Green PJ. 1997; On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B*. 59:731–792.
- Schmittmann VD, Dolan CV, van der Maas HL, Neale MC. 2005; Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research*. 40:461–488. [PubMed: 26788831]
- Scott SL, James GM, Sugar CA. 2005; Hidden Markov models for longitudinal comparisons. *Journal of the American Statistical Association*. 100:359–369.
- Song XY, Lu ZH. 2010; Semiparametric latent variable models with Bayesian P-splines. *Journal of Computational and Graphical Statistics*. 19:590–608.
- Song XY, Xia YM, Zhu HT. 2016; Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics*. 73:313–323. [PubMed: 27148857]
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002; Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*. 64:583–639.
- Spiegelhalter, DJ, Thomas, A, Best, NG, abd Lunn, D. WinBUGS User Manual. Version 1.4. MRC Biostatistics Unit; Cambridge, England: 2003.
- Tanner MA, Wong WH. 1987; The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*. 82:528–550.
- Wulfsohn MS, Tsiatis AA. 1997; A joint model for survival and longitudinal data measured with error. *Biometrics*. 53:330–339. [PubMed: 9147598]

Vermunt JK, Langeheine R, Böckenholt U. 1999; Latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*. 24:178–205.

Appendix

Full Conditional Distributions

(I) Full conditional distributions of Z_{it}

We follow Baum et al. (1970) to adopt a recursive method to sample Z_{it} from the full conditional distribution efficiently as follows:

Let $\mathbf{y}_i = \{y_{i1}, \dots, y_{iT}\}$ and $\mathbf{D}_i = \{\mathbf{d}_{i1}, \dots, \mathbf{d}_{iT}\}$, then we have

$$p(Z_{it} | \cdot) \propto p(\mathbf{y}_i, \mathbf{D}_i, \mathbf{w}_i, Z_{it} | \boldsymbol{\theta}) \quad (\text{A1})$$

$$= p(y_{i1}, \dots, y_{it}, \mathbf{d}_{i1}, \dots, \mathbf{d}_{it}, \mathbf{w}_i, Z_{it} | \boldsymbol{\theta}) \times p(y_{it+1}, \dots, y_{iT}, \mathbf{d}_{it+1}, \dots, \mathbf{d}_{iT} | \mathbf{w}_i, Z_{it}, \boldsymbol{\theta}) \quad (\text{A2})$$

$$\doteq q_{it}(\mathbf{y}_i, \mathbf{D}_i, \mathbf{w}_i, Z_{it} | \boldsymbol{\theta}) \times \bar{q}_{it}(\mathbf{y}_i, \mathbf{D}_i | \mathbf{w}_i, Z_{it}, \boldsymbol{\theta}). \quad (\text{A3})$$

We first initialize $q_{i1}(\mathbf{y}_i, \mathbf{D}_i, \mathbf{w}_i, Z_{i1} | \boldsymbol{\theta}) = p(y_{i1}, \mathbf{d}_{i1}, \mathbf{w}_i, Z_{i1} | \boldsymbol{\theta}) = p(y_{i1}, \mathbf{d}_{i1} | Z_{i1}, \boldsymbol{\theta}) p(Z_{i1} | \boldsymbol{\theta})$ and calculate $q_{it}(\mathbf{y}_i, \mathbf{D}_i, \mathbf{w}_i, Z_{it} | \boldsymbol{\theta})$ for $t = 2, \dots, T$, in a recursion manner as follows:

$$\begin{aligned} q_{it}(\mathbf{y}_i, \mathbf{D}_i, \mathbf{w}_i, Z_{it} | \boldsymbol{\theta}) &= q_{it}(y_{i1}, \dots, y_{it}, \mathbf{d}_{i1}, \dots, \mathbf{d}_{iT}, \mathbf{w}_i, Z_{it} | \boldsymbol{\theta}) \\ &= \sum_{u=1}^S p(y_{i1}, \dots, y_{it}, \mathbf{d}_{i1}, \dots, \mathbf{d}_{iT}, \mathbf{w}_i, Z_{it}, Z_{i,t-1} = u | \boldsymbol{\theta}) \\ &= \sum_{u=1}^S p(y_{i1}, \dots, y_{it}, \mathbf{d}_{i1}, \dots, \mathbf{d}_{iT}, \mathbf{w}_i, Z_{i,t-1} = u | \boldsymbol{\theta}) \times p(Z_{it} | Z_{i,t-1} = u, w_{i2}, \boldsymbol{\theta}) \times p(y_{it}, \mathbf{d}_{it} | Z_{it}, \\ &\quad w_{i1}, \boldsymbol{\theta}) \\ &= \sum_{u=1}^S \left[q_{it}(y_{i,t-1}, \mathbf{D}_i, \mathbf{w}_i, Z_{i,t-1} = u | \boldsymbol{\theta}) \times p(Z_{it} | Z_{i,t-1} = u, w_{i2}, \boldsymbol{\theta}) \times p(y_{it}, \mathbf{d}_{it} | Z_{it}, \right. \\ &\quad \left. w_{i1}, \boldsymbol{\theta}) \right], \end{aligned} \quad (\text{A4})$$

where $p(Z_{it} | Z_{i,t-1} = u, w_{i2}, \boldsymbol{\theta})$, $p(y_{it}, \mathbf{d}_{it} | Z_{it}, w_{i1}, \boldsymbol{\theta})$ and $p(\mathbf{w}_i | \boldsymbol{\theta})$ can be calculated through Equation (11).

Similarly, we initialize $\bar{q}_{iT}(\mathbf{y}_i, \mathbf{D}_i | \mathbf{w}_i, Z_{iT}, \boldsymbol{\theta}) = 1$ and calculate $\bar{q}_{it}(\mathbf{y}_i, \mathbf{D}_i | \mathbf{w}_i, Z_{it}, \boldsymbol{\theta})$ for $t = T - 1, \dots, 1$ as follows:

$$\begin{aligned} \bar{q}_{it}(\mathbf{y}_i, \mathbf{D}_i | \mathbf{w}_i, Z_{it}, \boldsymbol{\theta}) &= p(y_{it+1}, \dots, y_{iT}, \mathbf{x}_{it+1}, \dots, \mathbf{d}_{iT} | \mathbf{w}_i, Z_{it}, \boldsymbol{\theta}) \quad (\text{A5}) \\ &= \sum_{u=1}^S p(y_{it+1}, \dots, y_{iT}, \mathbf{d}_{it+1}, \dots, \mathbf{d}_{iT}, Z_{it+1} = u | \mathbf{w}_i, Z_{it}, \boldsymbol{\theta}) \\ &= \sum_{u=1}^S \left[p(y_{it+1}, \dots, y_{iT}, \mathbf{d}_{it+1}, \dots, \mathbf{d}_{iT} | Z_{it+1} = u, \mathbf{w}_i, \boldsymbol{\theta}) \times p(Z_{it+1} = u | Z_{it}, w_{i2}, \boldsymbol{\theta}) \quad (\text{A6}) \right. \\ &\quad \left. \times p(y_{it+1}, \mathbf{d}_{it+1} | Z_{it+1} = u, w_{i1}, \boldsymbol{\theta}) \right] \\ &= \sum_{u=1}^S \left[\bar{q}_{it}(\mathbf{y}_i, \mathbf{D}_i | \mathbf{w}_i, Z_{it+1} = u, \boldsymbol{\theta}) \times p(Z_{it+1} = u | Z_{it}, w_{i2}, \boldsymbol{\theta}) \times p(y_{it+1}, \mathbf{d}_{it+1} | Z_{it+1} = u, w_{i1}, \boldsymbol{\theta}) \right]. \end{aligned}$$

Thus, Z_{it} can be directly generated from (A1) when all $q_{it}(\cdot)$ s and $\bar{q}_{it}(\cdot)$ s defined in (A4) and (A5) are well calculated.

(II) Full conditional distributions of \mathbf{w}_i

$$\begin{aligned} p(\mathbf{w}_i | \cdot) &\propto p(\mathbf{y}_i, \mathbf{D}_i | \mathbf{w}_i, Z_{i1}, \dots, Z_{iT}, \boldsymbol{\theta}) \times p(Z_{i1}, \dots, Z_{iT} | \mathbf{w}_i, \boldsymbol{\theta}) \times p(\mathbf{w}_i | \boldsymbol{\theta}) \quad (\text{A7}) \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (y_{it} - \mu_s - \gamma_2^T \mathbf{b}_{it} - \sum_{j=1}^q \beta_{sj}^T \mathbf{B}_{it,j} - w_{i1})^2 \text{I}(Z_{it} = s) / \psi_s + \right. \\ &\quad \left. \sum_{t=2}^T \log(p_{itus}) \text{I}(Z_{i,t-1} = u, Z_{it} = s) - \frac{1}{2} \mathbf{w}_i^T \boldsymbol{\Phi}^{-1} \mathbf{w}_i \right\}, \end{aligned}$$

where p_{it0} and p_{itus} can be calculated via Equation (12).

(III) Full conditional distributions of μ_s , γ_s , ψ_s , and Φ

$$\begin{aligned} [\mu_s | \cdot] &\sim N[\mu_s^*, \sigma_{\mu_s}^*], & [\gamma_s | \cdot] &\sim N[\gamma_s^*, \boldsymbol{\Sigma}_s^*], \\ [\psi_s^{-1} | \cdot] &\sim \text{Gamma}[\tilde{\alpha}_s^*, \tilde{\beta}_s^*], & [\boldsymbol{\Phi}^{-1} | \cdot] &\sim \text{Wishart}[\mathbf{R}^*, N + \rho_0], \end{aligned} \quad (\text{A8})$$

where $\mathbf{R}^* = (\mathbf{R}_0 + \mathbf{W}\mathbf{W}^T)^{-1}$, $\mathbf{W} = (\mathbf{w}_1^T, \dots, \mathbf{w}_N^T)^T$ and

$$\sigma_{\mu s}^* = (n_s \psi_s^{-1} + \sigma_{\mu s 0}^{-1})^{-1}, \quad \Sigma_s^* = \left(\sum_{i=1}^N \sum_{t=1}^T \mathbf{b}_{it} \mathbf{b}_{it}^T \psi_s^{-1} + \Sigma_{s0}^{-1} \right)^{-1},$$

$$\mu_s^* = \sigma_{\mu s}^* \left[\psi_s^{-1} \sum_{i=1}^n \sum_{t=1}^T \mathbf{I}(Z_{it} = s) (y_{it} - \gamma_s^T \mathbf{b}_{it} - \sum_{j=1}^q \beta_{sj}^T \mathbf{B}_{it, j} - w_{i1}) + \sigma_{\mu s 0}^{-1} \mu_{s0} \right],$$

$$\gamma_s^* = \Sigma_s^* \left[\psi_s^{-1} \sum_{i=1}^n \sum_{t=1}^T \mathbf{I}(Z_{it} = s) \mathbf{b}_{it} (y_{it} - \mu_s - \sum_{j=1}^q \beta_{sj}^T \mathbf{B}_{it, j} - w_{i1}) + \Sigma_{s0}^{-1} \gamma_{s0} \right],$$

$$\tilde{\alpha}_s^* = n_s / 2 + \tilde{\alpha}_{s0}, \quad \tilde{\beta}_s^* = \tilde{\beta}_{s0} + \frac{1}{2} \left[\sum_{i=1}^n \sum_{t=1}^T \mathbf{I}(Z_{it} = s) (y_{it} - \mu_s - \gamma_s^T \mathbf{b}_{it} - \sum_{j=1}^q \beta_{sj}^T \mathbf{B}_{it, j} - w_{i1}) \right]^2.$$

(IV) Full conditional distributions of β_{sj} and θ_{sj}

$$[\beta_{sj} | \cdot] \sim N[\beta_{sj}^*, \mathbf{H}_{sj}] \mathbf{I}(\mathbf{1}_{n_s}^T \mathbf{B}_{sj} \beta_{sj} = 0), \quad (\text{A9})$$

where $\mathbf{H}_{sj} = (\psi_s^{-1} \mathbf{B}_{sj}^T \mathbf{B}_{sj} + \nu_{sj}^{-1} \mathbf{K}_{sj})^{-1}$, $\beta_{sj}^* = \psi_s^{-1} \mathbf{H}_{sj} \mathbf{B}_{sj}^T \mathbf{y}_s^*$, and $\mathbf{y}_s^* = \{y_{it, s}^*\}$ is an $n_s \times 1$ vector with

$$y_{it, s}^* = y_{it} - \mu_s - \gamma_s^T \mathbf{b}_{it} - \sum_{l \neq j, l=1}^q \beta_{sl}^T \mathbf{B}_{it, l} - w_{i1}, \quad \text{for } Z_{it} = s.$$

According to Panagiotelis and Smith (2008), sampling an observation β_{sj} from truncated normal (A9) is equivalent to sampling an observation $\beta_{sj}^{(temp)}$ from $N[\beta_{sj}^*, \mathbf{H}_{sj}]$ and then transforming $\beta_{sj}^{(temp)}$ to β_{sj} by

$$\beta_{sj} = \beta_{sj}^{(temp)} - \mathbf{H}_{sj} \mathbf{Q}_{sj}^T (\mathbf{Q}_{sj} \mathbf{H}_{sj} \mathbf{Q}_{sj}^T)^{-1} \mathbf{Q}_{sj} \beta_{sj}^{(temp)}, \quad (\text{A10})$$

where $\mathbf{Q}_{sj} = \mathbf{1}_{n_s}^T \mathbf{B}_{sj}$. Moreover,

$$[\nu_{sj}^{-1} | \cdot] \sim \text{Gamma} \left[\nu_1 + \frac{L}{2}, \nu_2 + \frac{1}{2} \boldsymbol{\beta}_{sj}^T \mathbf{K}_{sj} \boldsymbol{\beta}_{sj} \right]. \quad (\text{A11})$$

(V) Full conditional distributions of τ_s , ζ_{us} , and α

$$p(\tau_s | \cdot) \propto \exp \left\{ \sum_{u=s}^S \sum_{i=1}^n \log(p_{i10u}) \times I(Z_{i1} = u) - \frac{(\tau_s - \tau_{s0})^2}{2\sigma_{\tau 0}^2} \right\},$$

$$p(\zeta_{us} | \cdot) \propto \exp \left\{ \sum_{v=s}^S \sum_{i=1}^n \sum_{t=2}^T \log(p_{ituv}) \times I(Z_{it} = v, Z_{i,t-1} = u) - \frac{(\zeta_{us} - \zeta_{us0})^2}{2\sigma_{\zeta 0}^2} \right\}, \quad (\text{A12})$$

$$p(\alpha | \cdot) \propto \exp \left\{ \sum_{i=1}^n \sum_{t=2}^T \log(p_{itus}) \times I(Z_{it} = s, Z_{i,t-1} = u) - \frac{1}{2} (\alpha - \alpha_0)^T \mathbf{H}_{\alpha 0}^{-1} (\alpha - \alpha_0) \right\},$$

where p_{i10u} and p_{itus} can be calculated via Equation (12).

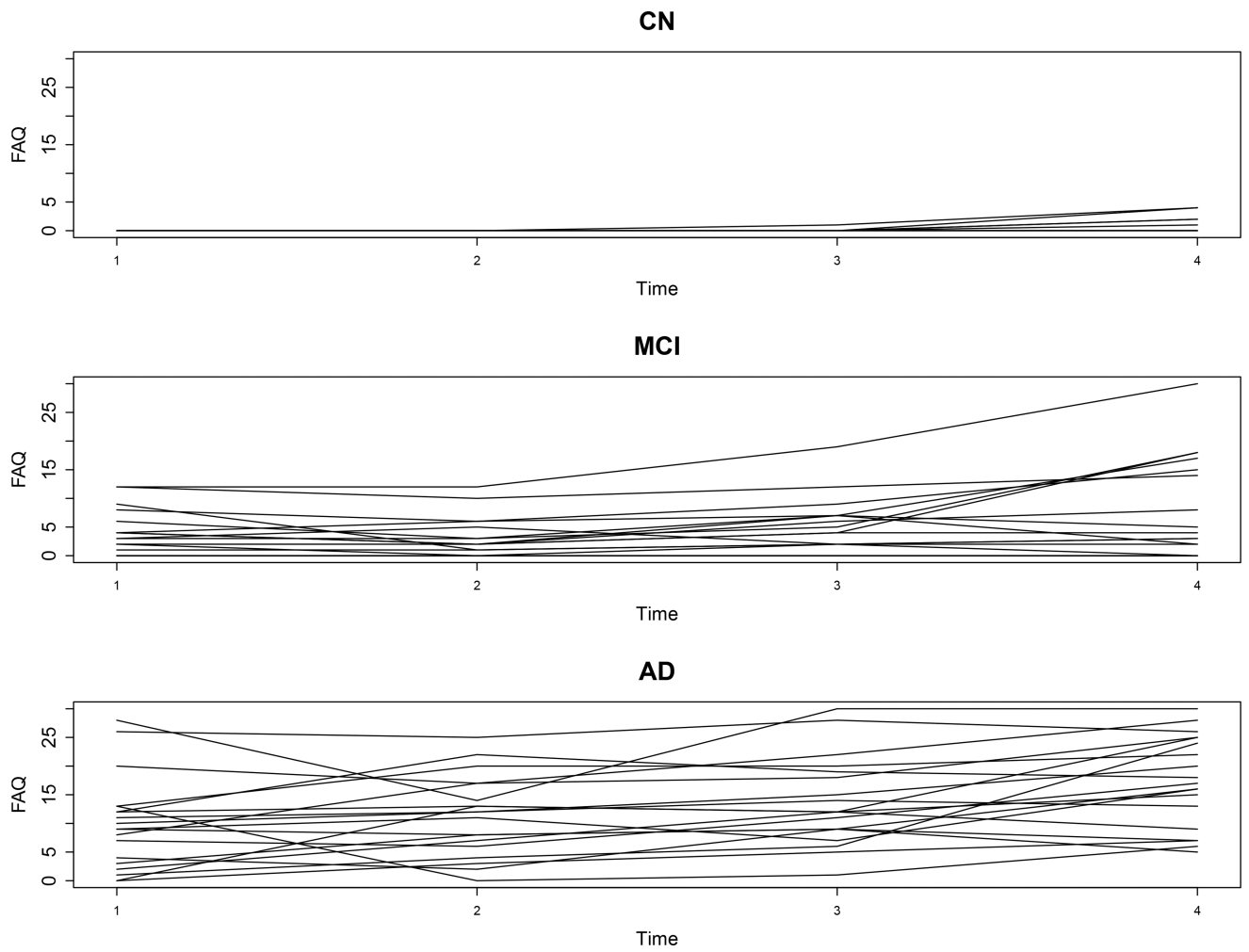


Figure 1.
ADNI-1 data analysis results: individual trajectories of FAQ scores for 20 randomly selective samples whose baseline states are CN, MCI, and AD, respectively.

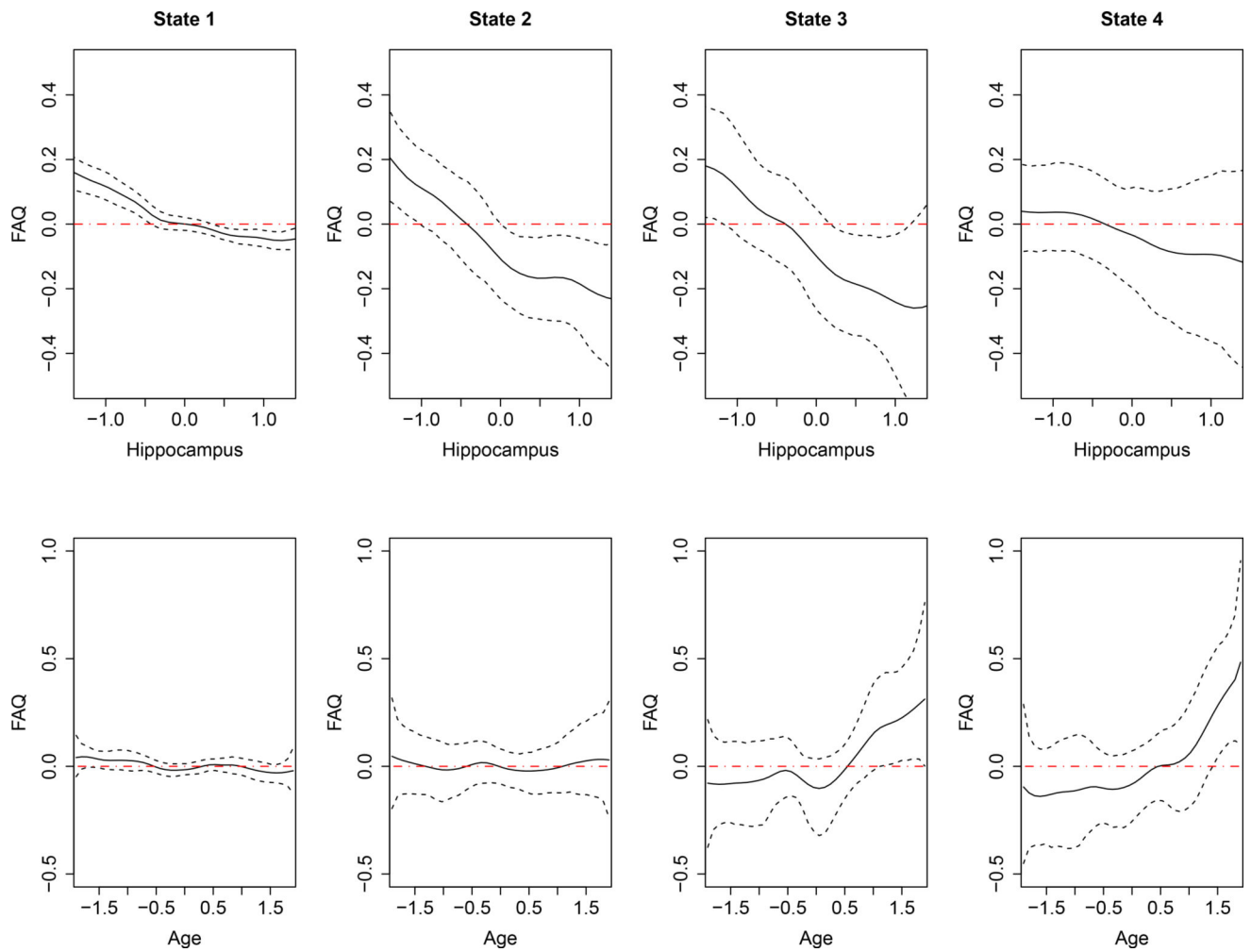


Figure 2. ADNI-1 data analysis results: estimates of the unknown smooth functions. The solid curves represent the pointwise mean curves, and the dashed curves represent the 2.5%- and 97.5%-pointwise quantiles. line $y = 0$ has been shown on each picture by red dot-dash to illustrate the range of significant effect for each variable.

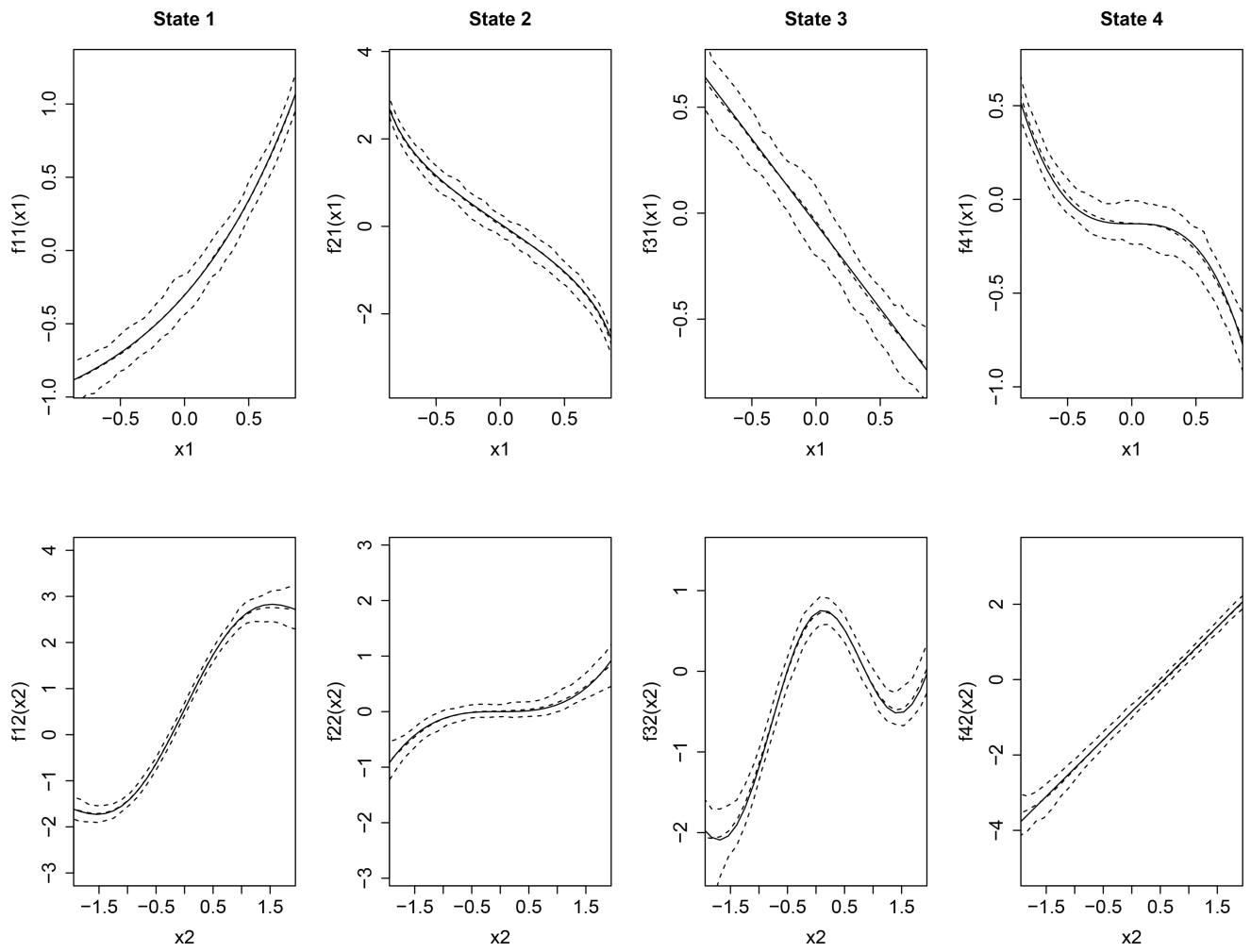


Figure 3.

Estimates of the unknown smooth functions in the simulation study: The solid curves represent the true curves, and the dashed curves represent the estimated posterior means and the 2.5%- and 97.5%- pointwise quantiles based on 100 replications, respectively.

Table 1

Characteristics of the study samples in the ADNI-1 dataset

Mean Age (in years)	73.0(6.9)
Gender (Male percentage)	56.2%
Mean log(hippocampus/whole brain volume)	-5.0(0.2)
Mean FAQ score	4.1(6.5)
One APOE- ϵ 4 allele carriers	34.6%
Two APOE- ϵ 4 alleles carriers	9.8%

The numbers in parentheses are standard deviations.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

ADNI-1 data analysis results; parameter estimation results.

<i>Parameters in conditional regression model</i>												
	State 1			State 2			State 3			State 4		
Par.	Est	SE	Par.	Est	SE	Par.	Est	SE	Par.	Est	SE	
μ_1	-0.556	0.013	μ_2	0.029	0.114	μ_3	1.108	0.134	μ_4	2.494	0.127	
ψ_1	0.013	0.001	ψ_2	0.135	0.019	ψ_3	0.187	0.025	ψ_4	0.397	0.057	
γ_{11}	0.016	0.018	γ_{21}	0.163	0.098	γ_{31}	0.102	0.131	γ_{41}	0.443	0.140	
γ_{12}	0.079	0.033	γ_{22}	0.260	0.132	γ_{32}	0.307	0.157	γ_{42}	0.407	0.193	
γ_{13}	0.011	0.015	γ_{23}	0.038	0.100	γ_{33}	-0.252	0.142	γ_{43}	-0.557	0.137	

<i>Parameters in probability transition model</i>												
Par.	Est	SE	Par.	Est	SE	Par.	Est	SE	Par.	Est	SE	
τ_1	0.938	0.110	τ_2	-0.303	0.373	τ_3	1.132	0.330	α_1	0.500	0.080	
α_2	0.035	0.067	α_3	-0.342	0.141	α_4	-0.728	0.230	α_5	0.085	0.162	
ζ_{11}	2.477	0.158	ζ_{21}	-1.714	0.342	ζ_{31}	-3.244	0.508	ζ_{41}	-3.252	0.512	
ζ_{12}	2.540	0.493	ζ_{22}	1.198	0.411	ζ_{32}	-1.661	0.479	ζ_{42}	-3.114	0.525	
ζ_{13}	1.877	0.779	ζ_{23}	2.618	0.445	ζ_{33}	1.298	0.307	ζ_{43}	-1.845	0.432	

<i>Covariance matrix of random effects</i>												
Par.	Est	SE	Par.	Est	SE	Par.	Est	SE				
φ_{11}	0.024	0.002	φ_{22}	0.223	0.051	φ_{12}	-0.007	0.006				

ADNI-1 data analysis results: comparison of estimated hidden states and diagnosis status

Table 3

Diagnosis	Estimates					Total
	CN	EMCI	LMCI	AD	Total	
CN	840	21	1	0	862	
EMCI	835	232	148	28	1243	
LMCI	9	21	39	21	90	
AD	23	49	106	159	337	
Total	1707	323	294	208	2532	

Table 4

Bayesian estimates of the parameters in the simulation study.

<i>Parameters in conditional regression model</i>												
State 1			State 2			State 3			State 4			
Par.	Bias	RMSE	Par.	Bias	RMSE	Par.	Bias	RMSE	Par.	Bias	RMSE	
μ_1	0.010	0.098	μ_2	0.024	0.115	μ_3	-0.046	0.195	μ_4	-0.006	0.106	
ψ_1	-0.006	0.042	ψ_2	0.011	0.047	ψ_3	0.003	0.044	ψ_4	0.016	0.028	
γ_{11}	-0.002	0.105	γ_{21}	-0.045	0.188	γ_{31}	0.025	0.147	γ_{41}	-0.020	0.112	
γ_{12}	0.002	0.091	γ_{22}	-0.022	0.126	γ_{32}	0.030	0.168	γ_{42}	0.018	0.089	
γ_{13}	-0.013	0.090	γ_{23}	-0.013	0.098	γ_{33}	0.026	0.126	γ_{43}	0.002	0.093	

<i>Parameters in probability transition model</i>												
Par.	Bias	RMSE	Par.	Bias	RMSE	Par.	Bias	RMSE	Par.	Bias	RMSE	
τ_1	0.008	0.137	τ_2	-0.030	0.187	τ_3	-0.048	0.236	α_1	0.007	0.068	
α_2	0.010	0.064	α_3	-0.024	0.103	α_4	-0.011	0.122	α_5	-0.029	0.112	
ζ_{11}	0.040	0.125	ζ_{21}	0.042	0.132	ζ_{31}	0.034	0.120	ζ_{41}	0.035	0.131	
ζ_{12}	0.050	0.168	ζ_{22}	0.031	0.156	ζ_{32}	0.037	0.162	ζ_{42}	0.034	0.167	
ζ_{13}	-0.034	0.197	ζ_{23}	-0.032	0.195	ζ_{33}	-0.027	0.186	ζ_{43}	-0.032	0.197	

<i>Covariance matrix of random effects</i>												
Par.	Bias	RMSE	Par.	Bias	RMSE	Par.	Bias	RMSE				
ϕ_{11}	0.004	0.066	ϕ_{22}	0.017	0.161	ϕ_{12}	0.001	0.067				