Received: 22 November 2019 | Revised: 9 January 2020 | Accepted: 31 January 2020

DOI: 10.1002/sim.8513

RESEARCH ARTICLE

Statistics in Medicine WILEY

Two-part hidden Markov models for semicontinuous longitudinal data with nonignorable missing covariates

Xiaoxiao Zhou | **Kai Kang** | **Xinyuan Song**

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

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

This study develops a two-part hidden Markov model (HMM) for analyzing semicontinuous longitudinal data in the presence of missing covariates. The proposed model manages a semicontinuous variable by splitting it into two random variables: a binary indicator for determining the occurrence of excess zeros at all occasions and a continuous random variable for examining its actual level. For the continuous longitudinal response, an HMM is proposed to describe the relationship between the observation and unobservable finite-state transition processes. The HMM consists of two major components. The first component is a transition model for investigating how potential covariates influence the probabilities of transitioning from one hidden state to another. The second component is a conditional regression model for examining the state-specific effects of covariates on the response. A shared random effect is introduced to each part of the model to accommodate possible unobservable heterogeneity among observation processes and the nonignorability of missing covariates. A Bayesian adaptive least absolute shrinkage and selection operator (lasso) procedure is developed to conduct simultaneous variable selection and estimation. The proposed methodology is applied to a study on the Alzheimer's Disease Neuroimaging Initiative dataset. New insights into the pathology of Alzheimer's disease and its potential risk factors are obtained.

KEYWORDS

Bayesian adaptive lasso, hidden Markov models, nonignorable missing covariates, semicontinuous data, two-part models

1 | INTRODUCTION

Hidden Markov models (HMMs) are widely used to address the serial heterogeneity in the analysis of longitudinal data. HMMs comprise two components: a transition model for describing the dynamic transitions of hidden states and a regression-type conditional model for examining state-specific covariate effects on the response of interest. Owing to their superiority in simultaneously revealing the longitudinal dependency structure and dynamic heterogeneity of the observed process, HMMs have received considerable attention in medical, behavioral, sociopsychological, and economic research.[1-6]

Despite the rapid development of HMMs and their variants, nearly all existing literature has assumed that the longitudinal response variable is either continuous or discrete. However, semicontinuous data, which are characterized by a

mixture of a probability mass at zero and a right-skewed continuous distribution for values greater than zero, are commonly encountered in substantive research. One typical example in medical studies is a scenario with a point mass at zero representing a subpopulation of "non-users" who do not receive medical care in a given time interval and a continuous distribution representing the level of expenditures among another subpopulation of "users." This semicontinuous structure may also exist in a longitudinal setting, in which subjects have two subpopulations, namely, "stayers," who have no probability of a nonzero observation in all occasions, and "movers," who may have a nonzero observation at one or more time points. Toward the semicontinuous data, a two-part model is often adopted to address the preponderance of zeros that cannot adequately be reflected in another manner. This model generally includes a logistic regression model for characterizing the membership probability of "non-users" (or "stayers") and a regression-type model for examining the relationship between continuous responses and potential covariates among "users" (or "movers"). Several two-part models for longitudinal semicontinuous data have been investigated in recent years. Olsen and Schafer[7] first introduced correlated random effects in a two-part model and applied the model to analyze longitudinal semicontinuous data. Smith et al[8] developed a marginalized two-part model that allows investigation on the effect of covariates on the overall population mean. Nonetheless, most contemporary works, including the preceding studies, have modeled the longitudinal nonzero responses by using a simple mixed-effects model rather than HMMs and thus cannot reveal the dynamic heterogeneity of the nonzero observations.

Another important issue in the analysis of HMMs is the presence of missing data. Simply deleting missing data, especially those that are missing not at random (MNAR), is problematic and may lead to a considerable bias in parameter estimation. Lee[9] further demonstrated that a case-wise deletion of missing data in a mixture-type model will result in misleading conclusions in determining the number of mixture components. A common and convenient approach of managing missing data is to treat them as missing at random (MAR), which assumes that the probability of missingness depends only on observed data but not on missing data.[10] Goldstein et al[11] used the multiple imputation approach to deal with missing data problem in the context of multilevel models. Their approach managed missing data in the response and covariates based on the MAR assumption. However, in a longitudinal setting, such MAR assumption is often violated when the probability of missingness depends dynamically on subject-specific characteristics. Hence, developing sound statistical methods to cope with MNAR data in the context of HMMs is of scientific interest and practical value. Several existing studies have investigated HMMs with nonignorable missing data. For example, Bartolucci and Farcomeni[12] developed an event-history approach to analyze mixed HMMs with informative dropout. Cai et al[13] considered a hidden Markov structural equation model with nonignorable missing responses. Nevertheless, the preceding analyses assumed that the response variable is continuous and only the response variable is subject to missingness. Motivated by the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset analyzed in the present study (Section 5), in which the response variable is semicontinuous and the covariates of interest are subject to missingness, we develop a new joint modeling approach to accommodate these multiple features.

We propose a novel two-part HMM for longitudinal semicontinuous data with nonignorable missing covariates. The proposed model comprises three parts. First, we consider a subject-level binary indicator to determine the occurrence of excess zeros at all occasions ("stayers"). A logistic regression model is utilized to examine the effects of potential covariates on the membership of "stayers." Second, for the longitudinal observations with one or more nonzero values ("movers"), we adopt an HMM to characterize the relationship between the transition process of hidden states and the observation process. We propose a transition model to investigate how potential covariates influence the transition probabilities from one hidden state to another and a conditional regression model to examine the state-specific effects of covariates on the response. Third, we use a logistic regression model to formulate the mechanism of missing covariates. We introduce a shared random effect to the binary indicator, conditional, and missing data models for jointly addressing the heterogeneity caused by the existence of omitted covariates that simultaneously affect these three parts of interest. The introduction of such random effect also accounts for the nonignorability of missingness. We develop a full Bayesian approach coupled with the adaptive least absolute shrinkage and selection operator method (lasso) and an efficient Markov chain Monte Carlo (MCMC) algorithm to perform variable selection and parameter estimation. A modified deviance information criterion (DIC) is used to determine the number of hidden states and an appropriate missing mechanism for the proposed model. DeSantis et al[14] considered a similar two-part HMM for zero-inflated Poisson counts. However, the excessive zeros in their model were assumed as count data arisen from the binary part of the model and a Poisson distribution. Their model framework neither accommodated semicontinuous data nor managed nonignorable missingness. To our knowledge, the present study is the first to propose a two-part HMM for jointly analyzing longitudinal semicontinuous data with nonignorable missing covariates.

Our proposed method is motivated by a real study on the ADNI dataset, which recruited subjects aging from 55 to 90 years old and collected a set of characteristics of the clinical spectrum of Alzheimer's disease (AD) over time. Functional Assessment Questionnaire (FAQ), a widely used test with a score from 0 to 30 (high scores reflect poor cognitive ability), is utilized in the ADNI study to monitor the cognitive impairment of patients. The longitudinal FAQ scores of majority of the participants exhibit a steadily increasing trend, which is in line with the AD pathology that often evolves a neurodegenerative progression from cognitive normal (CN) to mild cognitive impairment (MCI) or even to AD. However, approximately 30% of participants obtained zero FAQ scores over all the time points, suggesting that their cognitive functions remain unimpaired during the entire study period. Such special pattern is called normal aging process by cognitive studies.[15] Distinguishing AD from normal aging has been a recurring nosologic and diagnostic problem to enhance AD prognosis and early treatment.[16] In addition, some covariates, such as hippocampal volumes of participants, have considerable amount of missing entries in the ADNI dataset. Appendix S1 shows that subjects in the "missing group" generally hold smaller hippocampal volumes than those in the "observed group", which implies that the missingness of the hippocampal volume is not at random. Simply deleting subjects with missing covariates or naively treating them as MAR is problematic. This study aims (i) to differentiate normal aging from AD progression and examine the potential factors that affect the membership of these two different processes, (ii) to identify the hidden states that correspond to the diagnosed stages of cognitive decline and investigate the possibly variant effects of biomarkers on the cognitive impairment across states, (iii) to detect the factors that contribute to the neurodegenerative pathology from one state to another, and (iv) to propose an appropriate mechanism for modeling missing covariates. Although many existing studies have considered the relationship between various biomarkers and cognitive impairment across the AD progression,[17,18] they have regarded the normal aging process as a part of continuum in AD progression, thereby blurring the different features of the two processes. Moreover, they have disregarded the possibly informative missingness in covariates and thus may produce biased results. The proposed modeling approach enables us to investigate all the aforementioned features jointly and provides new insights into the precision medicine in AD treatment.

The remainder of the article is organized as follows. Section 2 defines the two-part HMMs with nonignorable missing covariates and discusses the associated model identifiability issue. Section 3 presents a Bayesian approach for the analysis of the proposed model. An adaptive lasso procedure in conjunction with a hybrid MCMC algorithm, which combines the Gibbs sampler, Metropolis-Hastings (MH) algorithm, and the forward filtering and backward sampling (FFBS) algorithm, is developed for simultaneous variable selection and estimation. Moreover, a modified DIC is used to determine the number of hidden states and an appropriate missing mechanism. Section 4 investigates the empirical performance of the Bayesian estimation and model selection through several simulation studies. Section 5 applies the proposed methodology to the aforementioned ADNI study. Section 6 concludes the paper. The technical details are provided in Appendix S1.

2 | TWO-PART HIDDEN MARKOV MODEL

For subject $i = 1, \dots, n$ at $t = 1, \dots, T$, let $y_{it} \geq 0$ be a semicontinuous response variable. We introduce a binary indicator variable V_i to accommodate the excess zeros in y_{it} as follows: $V_i = 1$ if $y_{it} = 0$ for $\forall t$ and 0 otherwise. Then, the probability of $V_i = 1$ is modeled as follows:

$$\text{logit}\{p(V_i = 1)\} = \boldsymbol{\beta}^T \mathbf{c}_i + w_i, \tag{1}$$

where $\mathbf{c}_i = (1, c_{i1}, \dots, c_{ir})^T$ is an $(r+1) \times 1$ vector of baseline covariates, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_r)^T$ is an $(r+1) \times 1$ vector of unknown coefficient, and w_i is a subject-specific random effect following a normal distribution $N(0, \sigma^2)$.

For subjects with $V_i = 0$, the longitudinal responses $y_{it}, t = 1, \dots, T$ are modeled through an HMM. The hidden state process, Z_{it}, is assumed to follow a first-order Markov chain and takes values in a finite set $\{1, \dots, S\}$. Given $V_i = 0$ and hidden state $Z_{it} = s$, a conditional regression model for the continuous response is defined as follows:

$$[y_{it} | V_i = 0, Z_{it} = s] = \mu_s + \boldsymbol{\gamma}_s^T \mathbf{x}_{it} + w_i + \epsilon_{it}, \tag{2}$$

where μ_s is a state-specific intercept, $\boldsymbol{\gamma_s} = (\gamma_{s1}, \dots, \gamma_{sp})^T$ is a $p \times 1$ state-specific coefficient, $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^T$ is a $p \times 1$ vector of covariates, w_i is the subject-specific random effect that is also incorporated in (1), ϵ_{it} is a random residual term independent of y_{it}, and $[\epsilon_{it} | V_i = 0, Z_{it} = s] \sim N(0, \psi_s)$.

Let p_{itus} denote the transition probability from state $Z_{it} = u$ at time $t-1$ to state $Z_{it} = s$ at time t for subject i. On the basis of the assumption of the first-order Markov chain, we have $p_{itus} = P(Z_{it} = s|Z_{i1}, Z_{i2}, \dots, Z_{it-1} = u) = P(Z_{it} = s|Z_{it-1} = u)$. In substantive research, hidden states often have natural ranking information. Thus, we assume that hidden states $\{1, \dots, S\}$ are ordered and $\vartheta_{itus} = P(Z_{it} = s|Z_{it} \geq s, Z_{it-1} = u)$. A continuation-ratio logit model is considered as follows:

$$\log\left(\frac{P(Z_{it} = s|Z_{it-1} = u)}{P(Z_{it} > s|Z_{it-1} = u)}\right) = \log\left(\frac{p_{itus}}{p_{itu,s+1} + \cdots + p_{ituS}}\right) = \text{logit}(\vartheta_{itus}) = \zeta_{us} + \boldsymbol{\alpha}^T \mathbf{x}_{it}, \tag{3}$$

where the left-hand side is the log odds of transition to state s rather than to a state that is higher than s given that $Z_{it-1} = u$, ζ_{us} is a transition-specific intercept, $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^T$ is the covariate vector defined in (2), and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$ is a $p \times 1$ vector of unknown parameters. Similar to the proportional odds assumption in a cumulative logit model, $\boldsymbol{\alpha}$ in (3) is assumed to be independent of u and s to maintain the order of the hidden states and obtain a parsimonious model. The initial probability of the hidden state at occasion $t = 1$ for subject i is defined as follows: For $s = 1, \dots, S-1$, $\log\left(\frac{P(Z_{i1}=s)}{P(Z_{i1}>s)}\right) = \log\left(\frac{p_{i10s}}{p_{i10,s+1}+\cdots+p_{i10,S}}\right) = \tau_s$.

Managing missingness in covariates $\mathbf{x}_{it}$ and $\mathbf{c}_i$ is another important issue. In this study, we focus on missingness in the time-variant covariate $\mathbf{x}_{it}$ because the baseline covariate $\mathbf{c}_i$ is usually fully observed in most substantive studies. Yet the extension to consider missingness of both $\mathbf{x}_{it}$ and $\mathbf{c}_i$ is trivial in the proposed framework. Let r_{itk} be the missing indicator for x_{itk}, such that $r_{itk} = 1$ if x_{itk} is missing and 0 otherwise. Denote $\mathbf{r}_{it} = (r_{it1}, \dots, r_{itp})^T$, $\mathbf{R}_i = \{\mathbf{r}_{it}; t = 1, \dots, T\}$, and $\mathbf{R} = \{\mathbf{R}_i; i = 1, \dots, n\}$, and partition $\mathbf{X}$ into $\{\mathbf{X}_o, \mathbf{X}_m\}$, where $\mathbf{X}_o$ and $\mathbf{X}_m$ are the observed and missing data sets of $\mathbf{X}$, respectively. Given that MAR assumption is not always a guarantee in reality, we consider a shared random effect model to accommodate the possible nonignorability of missing covariates in this study. Conditional on state s, r_{itk} is assumed to follow independent Bernoulli distribution, and for $t = 2, \dots, T$,

$$\text{logit}\{p(r_{itk} = 1|Z_{it} = s, w_i, \mathbf{f}_{it}, \boldsymbol{\varphi}_{sk})\} = \boldsymbol{\varphi}_{sk}^T \mathbf{f}_{it} + w_i, \tag{4}$$

where $\mathbf{f}_{it} = (1, f_{it1}, \dots, f_{itl})^T$ is the $(l+1) \times 1$ vector of covariates, and $\boldsymbol{\varphi}_{sk} = (\varphi_{sk0}, \dots, \varphi_{skl})^T$ is a $(l+1) \times 1$ state-specific regression parameter. The elements in $\mathbf{f}_{it}$ are usually chosen from the available covariates that may affect the missing probability of x_{itk}. The inclusion of response y_{it} and missing covariates in $\mathbf{f}_{it}$ is allowed but unnecessary because we include a shared random effect w_i into the conditional regression and missing data models to account for the nonignorability of missingness. Nevertheless, a model comparison procedure is required to assess the sensitivity of Bayesian results to misspecification of the missing data mechanism because the true missing mechanism is unknown. Notably, the covariates $\mathbf{c}_i$, $\mathbf{x}_{it}$, and $\mathbf{f}_{it}$ can have overlapping elements.

On the basis of the two-part HMMs defined by (1)-(4), several heterogeneities that are worthy of investigation in longitudinal semicontinuous data are well addressed. First, the binary indicator model defined by (1) examines the heterogeneity characterized by excess zeros over all occasions and continuous longitudinal responses. Second, the HMM defined by (2)-(3) describes the dynamic heterogeneity of the longitudinal observations across various hidden states. Finally, the random effect shared by (1), (2), and (4) accounts for the heterogeneity caused by the omitted covariates that simultaneously affect the binary indicator, longitudinal response, and missing probability. Although distinct but correlated random effects can be considered in (1), (2), and (4), introducing the same random effect into different models can avoid additional nuisance parameters and obtain a parsimonious joint model.

The proposed model is not identifiable without imposing identifiability constraints on model parameters. The model indeterminacy stems from the label switching problem caused by the likelihood function of an HMM being invariant to a random permutation of the state labels. Therefore, the resulting posterior distribution becomes multimodal under the symmetric priors of the parameters in different states. Basically, any constraint that defines the ranking of the state-specific parameter is sufficient to address the label switching problem. Among the literature, imposing an ordering restriction on component means is an efficient method for preventing label switching and ensuring model identifiability.[19-21] In the ADNI study, previous medical reports about AD indicated that patient's cognitive impairment steadily increases from mild to severe stage. Thus, we choose the state-specific mean and impose the constraint $\mu_1 < \dots < \mu_S$ in posterior sampling.

3 | BAYESIAN INFERENCE

Let $\boldsymbol{\theta}$ be the vector that contains all the unknown parameters; $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})^T$, $\mathbf{Y} = \{\mathbf{y}_i; i = 1, \ldots, n\}$, $\mathbf{F}_i = \{\mathbf{f}_{it}; t = 1, \ldots, T\}$, $\mathbf{F} = \{\mathbf{F}_i; i = 1, \ldots, n\}$, $\mathbf{C} = \{\mathbf{c}_i; i = 1, \ldots, n\}$, $\mathbf{Z} = \{\mathbf{Z}_i; i = 1, \ldots, n\}$, $\mathbf{w} = (w_1, \ldots, w_n)^T$, and $\mathbf{H}_o = \{\mathbf{Y}, \mathbf{X}_o, \mathbf{F}, \mathbf{C}, \mathbf{R}\}$. On the basis of proposed model defined in (1)-(4), the adaptive lasso estimator can be formulated as

$$\arg\min_\theta \left\{ -\sum_{i=1}^{n} \log(p(V_i)) + \sum_{i=1}^{n_s}\sum_{t=1}^{T}\left[(y_{it} - \eta_{it})^2 - \sum_{k=1}^{p} \log p(r_{itk})\right] - \sum_{i=1}^{n_s}\sum_{t=2}^{T} \log(p_{itus}) - P(\theta) \right\},$$

where $p(V_i)$ is the probability of excess zeros defined in (1), η_{it} is the mean of y_{it}, $p(r_{itk})$ is the probability of missingness defined in (4), p_{itus} is the transition probability defined in (3), and $P(\theta) = \sum_{j=1}^{r} \lambda_{\beta j}|\beta_j| + \sum_{s=1}^{S}\sum_{k=1}^{p} \lambda_{\gamma sk}|\gamma_{sk}| + \sum_{k=1}^{p} \lambda_{\alpha k}|\alpha_k| + \sum_{s=1}^{S}\sum_{k=1}^{p}\sum_{h=1}^{l} \lambda_{\varphi skh}|\varphi_{skh}|$, in which $\beta_j, \gamma_{sk}, \alpha_k$, and φ_{skh} are the coefficients of fixed effects, and $\lambda_{\beta j}, \lambda_{\gamma sk}, \lambda_{\alpha k}$, and $\lambda_{\varphi skh}$ are the corresponding tuning parameters.

Under the Bayesian framework, the adaptive lasso procedure can be implemented by introducing a multivariate conditional Laplace prior to the regression coefficient in $\boldsymbol{\theta}^* = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\varphi})$ as follows:

$$p(\boldsymbol{\theta}^*|\boldsymbol{\psi}, \boldsymbol{\sigma}^2) \propto \exp\left\{ -\sum_{j=1}^{r} \frac{\lambda_{\beta j}}{\sqrt{\sigma_\beta^2}}|\beta_j| - \sum_{k=1}^{p}\left(\frac{\lambda_{\gamma sk}}{\sqrt{\psi_s}}|\gamma_{sk}| - \frac{\lambda_{\alpha k}}{\sqrt{\sigma_\alpha^2}}|\alpha_k| - \sum_{h=1}^{l} \frac{\lambda_{\varphi_{skh}}}{\sqrt{\sigma_{\varphi sk}^2}}|\varphi_{skh}| \right) \right\}, \tag{5}$$

where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_S)$ and $\boldsymbol{\sigma}^2 = (\sigma_\beta^2, \sigma_\alpha^2, \sigma_{\varphi 1}^2, \ldots, \sigma_{\varphi S}^2)$. This conditional Laplace prior can be represented as a scale mixture of normals with an exponential mixing density, leading to a hierarchical representation of the full model as follows: for $i = 1, \ldots, n_s, t = 1, \ldots, T, s = 1, \ldots, S, j = 1, \ldots, r, k = 1, \ldots, p$, and $h = 1, \ldots, l$, we have

$$\begin{aligned}
&y_{it}|Z_{it} = s, \mu_s, \boldsymbol{\gamma}_s, w_i, \psi_s, \mathbf{x}_{it} \sim N(\eta_{it}, \psi_s) \\
&\boldsymbol{\gamma}_s|\psi_s, \tau_{\gamma s1}^2, \ldots, \tau_{\gamma sp}^2 \overset{ind}{\sim} N_p\left(\mathbf{0}, \psi_s \boldsymbol{\Sigma}_{\gamma s}\right), \quad \boldsymbol{\Sigma}_{\gamma s} = \mathrm{diag}\left(\tau_{\gamma s1}^2, \ldots, \tau_{\gamma sp}^2\right) \\
&\boldsymbol{\beta}|\sigma_\beta^2, \tau_{\beta 1}^2, \ldots, \tau_{\beta r}^2 \overset{ind}{\sim} N_p\left(\mathbf{0}, \sigma_\beta^2 \boldsymbol{\Sigma}_\beta\right), \quad \boldsymbol{\Sigma}_\beta = \mathrm{diag}\left(\tau_{\beta 1}^2, \ldots, \tau_{\beta r}^2\right) \\
&\boldsymbol{\alpha}|\sigma_\alpha^2, \tau_{\alpha 1}^2, \ldots, \tau_{\alpha p}^2 \overset{ind}{\sim} N_p\left(\mathbf{0}, \sigma_\alpha^2 \boldsymbol{\Sigma}_\alpha\right), \quad \boldsymbol{\Sigma}_\alpha = \mathrm{diag}\left(\tau_{\alpha 1}^2, \ldots, \tau_{\alpha p}^2\right) \\
&\boldsymbol{\varphi}_{sk}|\sigma_{\varphi sk}^2, \tau_{\varphi sk1}^2, \ldots, \tau_{\varphi skl_k}^2 \overset{ind}{\sim} N_p\left(\mathbf{0}, \sigma_{\varphi sk}^2 \boldsymbol{\Sigma}_{\varphi sk}\right), \quad \boldsymbol{\Sigma}_{\varphi sk} = \mathrm{diag}\left(\tau_{\varphi sk1}^2, \ldots, \tau_{\varphi skl}^2\right),
\end{aligned} \tag{6}$$

where $\overset{ind}{\sim}$ represents "independently distributed according to." For the tuning parameters $\lambda_{\beta j}, \lambda_{\gamma sk}, \lambda_{\alpha k}$, and $\lambda_{\varphi skh}$, we assign gamma priors as follows:

$$\begin{aligned}
&p(\lambda_{\beta j}) \overset{ind}{\sim} \mathrm{Gamma}(\alpha_{\beta j0}, \beta_{\beta j0}), \qquad p(\lambda_{\gamma sk}) \overset{ind}{\sim} \mathrm{Gamma}(\alpha_{\gamma sk0}, \beta_{\gamma sk0}), \\
&p(\lambda_{\alpha k}) \overset{ind}{\sim} \mathrm{Gamma}(\alpha_{\alpha k0}, \beta_{\alpha k0}), \qquad p(\lambda_{\varphi skh}) \overset{ind}{\sim} \mathrm{Gamma}(\alpha_{\varphi skh0}, \beta_{\varphi skh0}).
\end{aligned} \tag{7}$$

The prior distributions for the other unknown parameters of $\boldsymbol{\theta}$ are specified as follows:

$$p(\mu_s) \overset{ind}{\sim} N(\mu_{s0}, \sigma_{\mu s0}^2), \qquad p(\zeta_{us}) \overset{ind}{\sim} N(\zeta_{us0}, \sigma_{\zeta us0}^2), \qquad p(\tau_s) \overset{ind}{\sim} N(\tau_{s0}, \sigma_{\tau s0}^2), \tag{8}$$

where $\alpha_{\beta j0}, \alpha_{\gamma sk0}, \alpha_{\alpha k0}, \alpha_{\varphi skh0}, \beta_{\beta j0}, \beta_{\gamma sk0}, \beta_{\alpha k0}, \beta_{\varphi skh0}, \mu_{s0}, \zeta_{us0}, \tau_{s0}, \sigma_{\mu s0}^2, \sigma_{\zeta us0}^2$, and $\sigma_{\tau s0}^2$ are hyperparameters with preassigned values. Notably, Bayesian lasso does not shrink coefficients corresponding to unimportant predictors exactly to 0. We follow the existing literature[22-24] to determine the importance of predictors. If the absolute value of a coefficient is less than or equal to 0.1, we conclude the corresponding predictor is unimportant and should be removed from the model.

The Bayesian estimate of $\boldsymbol{\theta}$ can be obtained through the mean of the posterior samples drawn from $p(\boldsymbol{\theta}|\mathbf{H}_o)$. However, this posterior distribution involves high-dimensional integral with respect to latent quantities, such as hidden states, random effects, and missing data. Thus, directly sampling from $p(\boldsymbol{\theta}|\mathbf{H}_o)$ is intractable. We use the idea of data augmentation to augment the observed data $\mathbf{H}_o$ with the latent quantities $\mathbf{H}_m = \{\mathbf{Z}, \mathbf{w}, \mathbf{X}_m\}$. The Gibbs sampler is used to

sample from $p(\boldsymbol{\theta}, \mathbf{H}_m|\mathbf{H}_o)$ iteratively by generating (i) $\mathbf{X}_m$ from $P(\mathbf{X}_m|\mathbf{Z}, \mathbf{w}, \boldsymbol{\theta}, \mathbf{H}_o)$, (ii) $\mathbf{Z}$ from $P(\mathbf{Z}|\mathbf{X}_m, \mathbf{w}, \boldsymbol{\theta}, \mathbf{H}_o)$, (iii) $\mathbf{w}$ from $P(\mathbf{w}|\mathbf{X}_m, \mathbf{Z}, \boldsymbol{\theta}, \mathbf{H}_o)$, and (iv) $\boldsymbol{\theta}$ from $P(\boldsymbol{\theta}|\mathbf{X}_m, \mathbf{w}, \mathbf{Z}, \mathbf{H}_o)$. The details of the full conditional distributions involved in the MCMC algorithm are described in Appendix S1.

For the present two-part HMMs, other relevant inference issues include (i) determination of the number of hidden states, (ii) selection of an appropriate missing data mechanism for missing covariates, and (iii) examination of necessity of the shared random effect. We propose to use DIC to perform model selection. Given that the conventional DIC is inapplicable in the presence of missing data,[25] we adopt a modified DIC[26] to determine the number of hidden states, select a plausible missing mechanism, and examine the necessity of the random effect. The computation of the modified DIC is presented in Appendix S1.

4 | SIMULATION STUDY

4.1 | Simulation 1

We generate datasets from a two-part HMM defined by (1)-(4) with two states ($S = 2$). In conditional and transition models (2) and (3), we consider four time-variant covariates, $\mathbf{x}_{it} = (x_{it1}, x_{it2}, x_{it3}, x_{it4})^T$, in which $x_{it1}, x_{it2}, x_{it3}$, and x_{it4} are independently generated from *Bernoulli*(0.5), $U(-1, 1)$, $N(\sqrt{t-1}, 1)$, and $N(m_x, \sigma_x^2)$, respectively. In model (1), we set time-invariant covariate $\mathbf{c}_i$ to be the baseline covariate of $\mathbf{x}_{it}$, that is, $\mathbf{c}_i = \mathbf{x}_{i1}$. The shared random effect w_i is generated from $N(0, \sigma^2)$. For subjects with $V_i = 1$, y_{it} is set to 0 for $t = 1, \ldots, T$. We generate y_{it} for subjects with $V_i = 0$ based on a two-state HMM. The proposed joint model is defined as follows:

$$\text{logit}\{p(V_i = 1)\} = \beta_0 + \beta_1 x_{i11} + \beta_2 x_{i12} + \beta_3 x_{i13} + \beta_4 x_{i14} + w_i, \tag{9}$$

$$\text{logit}(\vartheta_{itus}) = \zeta_{us} + \alpha_1 x_{it1} + \alpha_2 x_{it2} + \alpha_3 x_{it3} + \alpha_4 x_{it4}, \tag{10}$$

$$[y_{it}|V_i = 0, Z_{it} = s] = \mu_s + \gamma_{s1} x_{it1} + \gamma_{s2} x_{it2} + \gamma_{s3} x_{it3} + \gamma_{s4} x_{it4} + w_i + \epsilon_{it}, \tag{11}$$

where $[\epsilon_{it}|Z_{it} = s] \sim N(0, \psi_s)$. Moreover, we assume that covariate x_{it4} is subject to nonignorable missingness for $t = 2, \ldots, T$ and generate the missing indicator r_{it4} conditional on $\mathbf{f}_{it} = (x_{it1}, x_{it2}, x_{it3})^T$ based on a logistic regression model as follows:

$$\text{logit}\{p(r_{it4} = 1|Z_{it} = s, w_i, \mathbf{f}_{it}, \boldsymbol{\varphi}_s)\} = \varphi_{s0} + \varphi_{s1} x_{it1} + \varphi_{s2} x_{it2} + \varphi_{s3} x_{it3} + w_i. \tag{12}$$

The true population values of the unknown parameters are set as $m_x = 0$, $\sigma_x^2 = 1$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T = (-1, 0, 1, 0, 0)^T$, $\sigma^2 = 1$, $\boldsymbol{\psi} = (\psi_1, \psi_2)^T = (0.25, 0.25)$, $\tau_1 = 0$, $\zeta_{11} = -1$, $\zeta_{21} = 1$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T = (2, 0, -1, 0)^T$, $\boldsymbol{\gamma}_1 = (\gamma_{11}, \gamma_{21})^T = (0, 1)^T, \boldsymbol{\gamma}_2 = (\gamma_{12}, \gamma_{22})^T = (-2, 0)^T, \boldsymbol{\gamma}_3 = (\gamma_{13}, \gamma_{23})^T = (-2, 1)^T, \boldsymbol{\gamma}_4 = (\gamma_{14}, \gamma_{24})^T = (-2, 1)^T$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^T = (-2, 1)^T$, $\boldsymbol{\varphi}_0 = (\varphi_{10}, \varphi_{20})^T = (-1.5, 2)^T$, $\boldsymbol{\varphi}_1 = (\varphi_{11}, \varphi_{21})^T = (0, 0)^T$, $\boldsymbol{\varphi}_2 = (\varphi_{12}, \varphi_{22})^T = (2, 0)^T$, and $\boldsymbol{\varphi}_3 = (\varphi_{13}, \varphi_{23})^T = (0, -2)^T$.

On the basis of the abovementioned setting, we generate $n = 1000$ observations with $T = 6$, of which approximately 25% of subjects are coded as $V_i = 1$ with observation $y_{it} = 0$ at each occasion, whereas the rest of the subjects are coded as $V_i = 0$ with longitudinal observations y_{it} generated from the specified HMM. Among the observations of subjects with $V_i = 0$, 45% and 55% of them come from the first and second states, respectively. The missing proportion is approximately 25% in each state, which mimics the scenario of the ADNI study.

In the simulation study, the prior inputs in (7)-(8) are assigned as follows (Prior I): $\mu_{s0} = \tau_{s0} = \zeta_{us0} = 0$ and $\sigma_{\mu s0}^2 = \sigma_{\tau s0}^2 = \sigma_{\zeta us0}^2 = 4$, and we follow a common practice[24] in the literature to set $\alpha_{\beta j0} = \alpha_{\gamma sk0} = \alpha_{\alpha k0} = \alpha_{\varphi skh0} = 1$, $\beta_{\beta j0} = \beta_{\gamma sk0} = \beta_{\alpha k0} = \beta_{\varphi skh0} = 0.1$. We conduct a few test runs to decide the number of burn-in iterations at convergence and find that 1000 burn-in iterations are sufficient. Therefore, we collect 5000 simulated observations after 1,000 burn-in iterations to obtain the Bayesian estimates of model parameters.

Table 1 presents the results summarized based on 100 replications. The bias, root mean square error (RMSE) between the Bayesian estimates and true population values of the parameters, and the ratio of standard error (SE)/standard deviation (SD) are used to assess the performance of Bayesian estimation. The bias and RMSE for most of the parameters are

close to zero, and the SE/SD ratios are basically close to one, indicating a satisfactory performance of Bayesian estimation. Moreover, the hidden states can be estimated through the posterior samples as follows:

$$\hat{Z}_{it} = \arg\max_{s\in\{1,\ldots,S\}} P(Z_{it} = s|\mathbf{y}_i, \boldsymbol{\theta}) \approx \arg\max_{s\in\{1,\ldots,S\}} \frac{1}{J}\sum_{j=1}^{J} I\left(Z_{it}^{(j)} = s\right), \tag{13}$$

where $Z_{it}^{(j)}$ denotes the hidden state of y_{it} at the jth iteration, and $\arg\max_{s\in\{1,\ldots,S\}} \frac{1}{J}\sum_{j=1}^{J} I\left(Z_{it}^{(j)} = s\right)$ is the posterior mean of the hidden state of y_{it} drawn from the MCMC iterations. In this simulation, the correct classification rate calculated using (13) is over 94% in each replication, and the average accuracy based on 100 datasets is 94.68%, thereby indicating that the proposed model can correctly identify the hidden states of observations.

To assess the sensitivity of Bayesian estimation to prior inputs, we disturb the hyperparameters as follows (Prior II): $\mu_{s0} = \tau_{s0} = \zeta_{us0} = 0$, $\sigma^2_{\mu s0} = \sigma^2_{\tau s0} = \sigma^2_{\zeta us0} = 100$. Table 1 presents the Bayesian results obtained under Prior II for comparison. The results under the two prior inputs are similar.

We also reanalyze the simulated 100 datasets by using a naive model M_C, which performs case-wise deletion and only uses available data with complete covariates to fit the model. The results are presented in Appendix S1, along with the results of M_2 for comparison. Most of the estimates, especially those involved in the transition model, show significantly larger biases for M_C than for M_2. Therefore, ignoring the missing data problem and adopting the case-wise deletion procedure leads to serious information loss and produces biased estimation.

4.2 | Simulation 2

To investigate the necessity of the shared random effect in the proposed model, we consider another competing model, M_{S_1}, which excludes the random effect in the binary indicator, conditional, and missing data models. Appendix S1 reports the estimates of the unknown parameters under M_{S_1} based on the 100 datasets generated in Simulation 1. Compared with the results under the true model M_2, nearly all Bayesian estimates obtained from M_{S_1} show considerably larger Bias and RMSE. Moreover, the average accuracy of state identification under M_{S_1} is 91.18%, which is lower than that under the true model.

To check the performance of the proposed method in a homogenous scenario, we generate datasets with the same setting as in Simulation 1 but without the random effect w_i. We consider two competing models, namely, M_{S_2} and M_{S_3}, which represent the two-part HMMs defined by (1)-(4) with and without the shared random effect, respectively. Appendix S1 presents the Bayesian estimates of unknown parameters obtained from M_{S_2} and M_{S_3} based on 100 replications. Majority of the estimates are close to the population true values under both models. The correct classifications of Z_{it} based on M_{S_2} and M_{S_3} are 95.22% and 95.35%.

4.3 | Simulation 3

To examine the performance of the modified DIC in model selection, we re-analyze the 100 datasets generated in Simulation 1 by using the following seven competing models:

(1) Determination of the number of hidden states S:
   M_S: $S = 1, \ldots, 4$, missing covariates are treated as MNAR.
(2) Selection of missing mechanisms:
   M_5: $S = 2$, the missing mechanism (12) is wrongly specified as MAR as follows: $\text{logit}\{p(r_{it4} = 1|Z_{it} = s, \mathbf{f}_{it} = (x_{it1}, x_{it2}, x_{it3})^T, \boldsymbol{\varphi}_s)\} = \varphi_{s0} + \varphi_{s1}x_{it1} + \varphi_{s2}x_{it2} + \varphi_{s3}x_{it3}$.
   M_6: $S = 2$, the missing mechanism (12) is wrongly specified as $\text{logit}\{p(r_{it4} = 1|Z_{it} = s, w_i, \mathbf{f}_{it} = (x_{it1}, x_{it2}, x_{it3})^T, \boldsymbol{\varphi}_s)\} = \varphi_0 + \varphi_1 x_{it1} + \varphi_2 x_{it2} + \varphi_3 x_{it3} + w_i$, where φ_0, φ_1, φ_2, and φ_3 are set as state-invariant.
(3) Examination of necessity of the shared random effect:
   M_{S1}: $S = 2$, the random effect in models (1), (2), and (4) are omitted.

TABLE 1 Summary of the Bayesian estimates in Simulation 1

	Prior I			Prior II		
	Model (9)					
Par	**Bias**	**RMSE**	**SE/SD**	**Bias**	**RMSE**	**SE/SD**
β_0	−0.011	0.146	1.000	−0.011	0.146	1.000
β_1	0.030	0.182	1.027	0.031	0.184	1.031
β_2	−0.071	0.200	1.127	−0.073	0.201	1.131
β_3	−0.001	0.107	0.995	0.002	0.108	0.995
β_4	−0.000	0.108	0.995	0.000	0.109	0.995
σ^2	0.009	0.082	1.009	0.010	0.084	1.010
	Model (10)					
Par	**Bias**	**RMSE**	**SE/SD**	**Bias**	**RMSE**	**SE/SD**
τ	0.003	0.113	0.996	0.026	0.116	0.995
α_1	−0.100	0.160	1.108	−0.104	0.164	1.161
α_2	−0.008	0.094	0.999	−0.006	0.102	0.999
α_3	0.016	0.069	1.063	0.018	0.071	1.057
α_4	−0.011	0.071	1.010	−0.011	0.069	1.014
ζ_{11}	0.074	0.161	1.167	0.078	0.168	1.170
ζ_{21}	0.023	0.150	1.009	0.020	0.154	1.008
	Model (11)					
	State 1			**State 2**		
Par	**Bias**	**RMSE**	**SE/SD**	**Bias**	**RMSE**	**SE/SD**
$\boldsymbol{\mu}$	0.007	0.061	1.016	−0.005	0.063	0.998
$\boldsymbol{\gamma_1}$	−0.006	0.043	1.024	−0.001	0.034	0.997
$\boldsymbol{\gamma_2}$	0.004	0.037	1.011	0.002	0.031	0.998
$\boldsymbol{\gamma_3}$	0.001	0.019	0.998	0.003	0.014	1.024
$\boldsymbol{\gamma_4}$	−0.012	0.025	1.198	0.002	0.018	1.005
$\boldsymbol{\psi}$	0.053	0.055	1.159	0.014	0.019	1.128
	State 1			**State 2**		
Par	**Bias**	**RMSE**	**SE/SD**	**Bias**	**RMSE**	**SE/SD**
$\boldsymbol{\mu}$	0.007	0.064	1.007	−0.003	0.065	0.997
$\boldsymbol{\gamma_1}$	−0.005	0.044	1.010	−0.001	0.035	0.998
$\boldsymbol{\gamma_2}$	0.005	0.037	1.015	0.001	0.031	0.997
$\boldsymbol{\gamma_3}$	0.002	0.019	1.001	0.002	0.014	1.020
$\boldsymbol{\gamma_4}$	−0.012	0.025	1.120	0.003	0.018	1.011
$\boldsymbol{\psi}$	0.050	0.053	1.222	0.013	0.022	1.120

(Continues)

TABLE 1 (Continued)

	Prior I						Prior II					
	Model (12)											
	State 1			State 2			State 1			State 2		
Par	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD
φ_0	0.064	0.195	1.081	−0.011	0.209	0.999	0.077	0.220	1.083	−0.012	0.227	0.997
φ_1	−0.052	0.187	1.072	−0.001	0.189	0.995	−0.059	0.221	1.060	0.002	0.191	0.995
φ_2	−0.090	0.198	1.158	0.005	0.145	0.997	−0.088	0.191	1.165	0.003	0.140	0.995
φ_3	0.008	0.095	0.998	−0.007	0.131	0.996	0.003	0.091	0.996	−0.005	0.133	0.996

Abbreviations: Par, parameter; RMSE, root mean square errors; SD, standard deviation: SE, standard error.

Among these competing models, M_2 is the true model; M_1, M_3, and M_4 are models with an incorrect number of hidden states; and M_5, M_6, and M_{S1} are the same as M_2, except that M_5 assumes a MAR mechanism; M_6 assigns state-invariant-coefficient missing data model, and M_{S1} overlooks the linkage across the three models. Our main goal is to evaluate the performance of the modified DIC in identifying the true number of hidden states, selecting an appropriate missingness mechanism, and examining the necessity of the random effect. The boxplots of DIC values under all competing models are reported in Appendix S1. The true model M_2 is consistently selected with the smallest DIC value in each of the 100 replications.

The code for implementing the simulation study is written in Python and can be freely downloaded at https://orcid.org/0000-0002-4877-3200.

5 | ADNI DATA ANALYSIS

We applied the proposed method to analyze a dataset extracted from the ADNI study. The initial goal of ADNI was to recruit 800 participants aging between 55 and 90 years old, including 200 subjects for elderly control, 400 subjects with MCI, and 200 subjects with early AD. After obtaining informed consent, participants underwent a series of tests, and their neuroimaging, genetic, and biochemical markers as well as clinical and cognitive data were collected over time. Additional information about ADNI can be found in the official website (www.adni-info.org). FAQ, a cognitive measurement collected in ADNI, is a functional and behavioral assessment with 10 items corresponding to function independently in daily life and is widely used to measure cognitive impairment over time. FAQ score ranges from 0 to 30, with high scores indicating poor cognitive ability. In the ADNI study, the longitudinal FAQ scores of most participants exhibit a gradually increasing trend, indicating a disease progression from CN to MCI or further to AD.[27] However, approximately 30% of subjects have unchanged zero FAQ score in each follow-up visit. Such a pattern of longitudinal cognitive ability, in which the cognitive function remains unimpaired over the life span, is often referred to as normal aging by cognitive studies.[15] Previous neuropathologic studies[28] found specific neuronal loss in the entorhinal cortex in patients with very mild AD but no change in the same region for a cognitively intact elderly. These observations imply that the CN–MCI–AD progression and normal aging process are dichotomous and that normal aging should be treated separately rather than be regarded as the earliest stage of AD progression.

Given the presence of longitudinal semicontinuous response FAQ and the existence of hidden pathophysiological states in AD pathology, the two-part HMMs defined in (1)-(4) is proposed to fit the ADNI dataset. This ADNI data analysis aims (i) to investigate the potential covariates that influence the probability of a subject belonging to the normal aging group, (ii) to identify the hidden states of the neurodegenerative pathology on the basis of subjects who are in an AD progression, (iii) to examine a set of covariates that affect the between-state transition, and (iv) to reveal the effects of potential risk factors on cognitive impairment across the hidden states of the AD progression.

We focused on 651 subjects from ADNI-1 study and collected their FAQ score (y_{it}) at baseline, 6 months, 12 months, 24 months, and 36 months. Among the subjects, 182 had zero FAQ scores in all five visits, whereas 469 had at least one nonzero FAQ score during the five follow-up visits. We used $V_i = 1$ to denote these 182 subjects with normal aging pattern and used $V_i = 0$ for the other 469 subjects in the AD progression. The apolipoprotein E-ϵ4 (APOE-ϵ4), a well-known risk factor for progression from MCI to AD,[29] was coded using two dummy variables: "one-APOE-ϵ4-allele carrier (x_{it1})" and "two-APOE-ϵ4-allele carrier (x_{it2})" in the current study. Two discrete demographic characteristics, gender (x_{it3}, 1 = female) and marital status (x_{it4}, 1 = has never married) were also included. Moreover, three continuous covariates, namely, years of education (x_{it5}), age (x_{it6}), and the logarithm of the ratio of hippocampal volume over whole brain (x_{it7}), were considered. These seven covariates were incorporated into the conditional model (2) and transition model (3) for investigating their associations with FAQ and the transition pattern of AD. The baseline covariates, namely, $\mathbf{c}_i = (x_{i11}, \ldots, x_{i17})^T$, were included in the binary indicator model (1) to explore their potential effects on the classification of normal aging or AD progression. In this study, the covariate "hippocampal volume (x_{it7})" contains approximately 25% missing entries in the follow-up study. Given that the hippocampus plays an important role in consolidating information from short- to long-term memory and that subjects with lower hippocampal volume are more likely to have missingness in the following visits (see Appendix S1), simply treating the missing entries of "hippocampal volume" as MAR was certainly inappropriate. Thus, we considered model (4) with covariates $\mathbf{f}_{it} = (x_{it1}, \ldots, x_{it6})^T$ to accommodate the nonignorable missingness of "hippocampal volume."

TABLE 2 DIC values in the analysis of the ADNI dataset

Models	M_{Ns}	M_{Is}	M_{Ws}
$S = 1$	13,990	13,904	16,828
$S = 2$	11,373	11,492	14,343
$S = 3$	11,130	11,248	14,071
$S = 4$	11,343	11,434	14,233
$S = 5$	11,955	12,010	14,552

Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; DIC, deviance information criterion.

We started with determining the number of hidden states, selecting an appropriate missing mechanism, and examining the necessity of the shared random effect by using the modified DIC. We considered the following competing models:

M_{Ns}: an s-state HMM defined by (1)-(4) with a nonignorable missing mechanism.

M_{Is}: an s-state HMM defined by (1)-(4) with the random effect being excluded in (4).

M_{Ws}: an s-state HMM defined by (1)-(4) with the shared random effect being excluded in all the submodels.

The above three settings were considered under five different numbers of states, $s = 1, 2, 3, 4, 5$, leading to 15 competing models. The identifiability constraint $\mu_1 < \ldots < \mu_S$ was used to avoid label switching. We generated several MCMC chains from different initial values to check the convergence of the MCMC algorithm. The trace plot (not reported) indicated that the MCMC algorithm converged within 5000 iterations. Thus, we collected 5000 MCMC samples after discarding 5000 burn-in iterations to calculate the modified DIC value of each competing model. Based on the results in Table 2, M_{N3}, the three-state HMM with nonignorable missing mechanism, has the smallest DIC value and was therefore chosen for the subsequent analysis. Table 3 shows the Bayesian estimates (Est), standard error estimates (SE), and the 95% credible interval (CI) of all unknown parameters in M_{N3}. For each parameter, the 95% CI is constructed by using the 2.5% and 97.5% percentiles of its posterior samples collected from the MCMC iterations. A covariate effect is significant if its estimate exceeds twice of the standard error estimate or the corresponding 95% CI does not include zero.

We have the following observations. In model (1), APOE-ϵ4 allele, hippocampal volume, and gender have significant effects on the probability of classification between normal aging and AD-progression. Patients with APOE-ϵ4 alelles have significant negative effect [$\hat{\beta}_1 = -0.390(0.188), \hat{\beta}_2 = -0.610(0.301)$] on being classified to normal aging group. This finding agrees with the published medical report[30] and implies that the mutation in APOE-ϵ4 is an important risk factor for AD development. By contrast, hippocampal volume shows significant positive effect [$\hat{\beta}_7 = 0.524(0.062)$] on the probability of experiencing normal aging, which is also verified by other medical researches.[31] Female [$\hat{\beta}_3 = 0.491(0.182)$] are more likely to be classified into normal aging group. The strong association between these baseline covariates and the membership probability of normal aging provides a possible way for doctors to distinguish AD-progression from normal aging in an early period, which therefore reduces the medical cost and enhances the targeted treatment.

In conditional model (2), $\hat{\mu}_1 = -0.491(0.030)$, $\hat{\mu}_2 = 0.320(0.082)$, and $\hat{\mu}_3 = 1.883(0.109)$ are ranked in an ascending order, indicating that the subjects in state 1 have the lowest FAQ scores, whereas the subjects in state 3 have the highest. Hence, the subjects' cognitive impairment become steadily severer from state 1 to state 3. Existing literature[32] has indicated that states 1 to 3 can be explained as CN, MCI, and AD, which describe the three major neurodegenerative stages in AD progression. The APOE-ϵ4 alleles [$\hat{\gamma}_{11} = 0.089(0.039), \hat{\gamma}_{12} = 0.204(0.065), \hat{\gamma}_{21} = 0.210(0.089), \hat{\gamma}_{22} = 0.529(0.145)$] have positive effects on FAQ in CN and MCI states and the effect of two APOE-ϵ4 alleles is even larger than that of one APOE-ϵ4 allele. Such detrimental role of APOE-ϵ4 alleles played in the cognitive decline has also been confirmed by other studies.[33] On the contrary, the hippocampal volume is negatively associated with FAQ in CN and MCI state [$\hat{\gamma}_{17} = -0.024(0.012), \hat{\gamma}_{27} = -0.117(0.054)$], implying that atrophy of hippocampus increasingly impairs patients' cognitive ability on the progression from CN to MCI.[34] However, in the AD state, all the indicators become insignificant except that two APOE-ϵ4 alleles still have positive effects [$\hat{\gamma}_{32} = 0.306(0.162)$] on FAQ. Such salient effect demonstrates that two APOE-ϵ4 alleles are the decisive factor for the cognitive impairment in AD state.[35]

In transition model (3), two-APOE-ϵ4 alleles have negative effect on the probability of transitioning from a state to a better one [$\hat{\alpha}_2 = -0.702(0.204)$]. This result is in line with previous studies.[36] By contrast, preventing the loss of hippocampal volume [$\hat{\alpha}_7 = 0.377(0.090)$] is beneficial to postpone transitioning to a severer state. Such finding is supported by the medical research, which shows that hippocampus is critically involved in the encoding, storage, and retrieval of

TABLE 3 Bayesian estimates of the parameters in the analysis of ADNI dataset

Parameters in the binary indicator model											
Par	**Est**	**SE**	**95%CI**	**Par**	**Est**	**SE**	**95 %CI**	**Par**	**Est**	**SE**	**95%CI**
β_0	1.584	0.283	(0.858, 1.945)	β_1	−0.390	0.188	(−0.758, −0.023)	β_2	−0.610	0.301	(−1.241, −0.053)
β_3	0.461	0.182	(0.154, 0.851)	β_4	0.416	0.328	(−0.207, 1.047)	β_5	0.115	0.087	(−0.059, 0.291)
β_6	0.094	0.090	(−0.074, 0.282)	β_7	0.524	0.062	(0.374, 0.619)	σ^2	0.116	0.010	(0.095, 0.140)
Parameters in the transition model											
τ_1	1.531	0.153	(1.224, 1.831)	τ_2	1.893	0.521	(1.042, 3.105)	α_1	−0.280	0.155	(−0.558, 0.037)
α_2	−0.702	0.204	(−1.150, −0.277)	α_3	−0.094	0.151	(−0.400, 0.140)	α_4	0.380	0.347	(−0.327, 1.018)
α_5	0.040	0.078	(−0.112, 0.184)	α_6	−0.119	0.079	(−0.265, 0.040)	α_7	0.377	0.090	(0.205, 0.547)
ζ_{11}	2.117	0.135	(1.855, 2.386)	ζ_{21}	−1.956	0.276	(−2.522, −1.454)	ζ_{31}	−4.617	0.883	(−6.639, −3.202)
ζ_{12}	3.486	0.541	(2.578, 4.706)	ζ_{22}	1.064	0.219	(0.634, 1.472)	ζ_{32}	−2.654	0.516	(−3.786, −1.756)
Parameters in the conditional regression model											
State 1				**State 2**				**State 3**			
Par	**Est**	**SE**	**95%CI**	**Par**	**Est**	**SE**	**95%CI**	**Par**	**Est**	**SE**	**95%CI**
μ_1	−0.491	0.030	(−0.545, −0.430)	μ_2	0.320	0.082	(0.156, 0.456)	μ_3	1.883	0.186	(1.449, 2.183)
γ_{11}	0.089	0.039	(0.007, 0.161)	γ_{21}	0.210	0.089	(0.037, 0.395)	γ_{31}	−0.021	0.123	(−0.291, 0.219)
γ_{12}	0.204	0.065	(0.074, 0.325)	γ_{22}	0.529	0.145	(0.245, 0.807)	γ_{32}	0.306	0.162	(0.002, 0.638)
γ_{13}	−0.040	0.040	(−0.115, 0.040)	γ_{23}	−0.065	0.081	(−0.225, 0.096)	γ_{33}	0.074	0.117	(−0.150, 0.312)
γ_{14}	−0.015	0.100	(−0.216, 0.175)	γ_{24}	−0.329	0.386	(−1.011, 0.498)	γ_{34}	−0.002	0.479	(−0.989, 0.981)
γ_{15}	−0.023	0.017	(−0.056, 0.009)	γ_{25}	−0.063	0.041	(−0.142, 0.017)	γ_{35}	−0.039	0.066	(−0.172, 0.090)
γ_{16}	0.031	0.017	(−0.003, 0.067)	γ_{26}	0.024	0.051	(−0.071, 0.131)	γ_{36}	0.115	0.061	(−0.003, 0.234)
γ_{17}	−0.024	0.012	(−0.048, −0.003)	γ_{27}	−0.117	0.054	(−0.219, −0.010)	γ_{37}	0.006	0.078	(−0.161, 0.145)
ψ_1	0.039	0.002	(0.035, 0.043)	ψ_2	0.215	0.022	(0.175, 0.263)	ψ_3	0.4745	0.055	(0.376, 0.590)

(Continues)

TABLE 3 (Continued)

Parameters in the missing data model										
State 1				**State 2**				**State 3**		
φ_{10}	−1.494	0.159	(−1.802, −1.170)	φ_{20}	−1.108	0.269	(−1.667, −0.433)	φ_{30}	−0.243	0.288	(−0.910, 0.695)
φ_{11}	−0.201	0.260	(−0.504, 0.052)	φ_{21}	0.360	0.427	(−0.185, 0.940)	φ_{31}	−0.363	0.358	(−0.993, 0.189)
φ_{12}	−0.130	0.118	(−0.757, 0.363)	φ_{22}	−0.212	0.246	(−1.090, 0.581)	φ_{32}	0.199	0.271	(−0.454, 1.000)
φ_{13}	0.146	0.074	(0.021, 0.550)	φ_{23}	0.170	0.297	(−0.404, 0.780)	φ_{33}	−0.228	0.294	(−0.878, 0.297)
φ_{14}	0.152	0.165	(−0.157, 0.463)	φ_{24}	−0.860	0.296	(−1.420, −0.325)	φ_{34}	−0.089	0.321	(−0.959, 0.461)
φ_{15}	0.225	0.071	(0.092, 0.362)	φ_{25}	0.397	0.183	(0.056, 0.765)	φ_{35}	−0.181	0.138	(−0.435, 0.121)
φ_{16}	0.260	0.080	(0.119, 0.409)	φ_{26}	0.435	0.182	(0.084, 0.803)	φ_{36}	0.190	0.152	(−0.111, 0.482)

Abbreviations: Est, estimate; Par, parameter; SE, standard error estimate.

long-term memories[37] and that smaller hippocampus is associated with increased risk for conversion to AD state.[38] Marital status and education years have insignificant effect on the transition probability when APOE allele and hippocampus are controlled.

In missing data model (4), the effect of age on the probability of missingness in the hippocampal volume is significantly positive in CN and MCI state [$\hat{\varphi}_{16} = 0.225(0.071)$, $\hat{\varphi}_{26} = 0.397(0.183)$], implying that older people are likely to have missing follow-up measurements of the hippocampal volume in the early stages of AD progression. This finding is in line with the common sense that the elderly tends to drop out or be unresponsive in survey due to medical problems. Another finding is that females are prone to have missing measurement of hippocampal volume in CN state [$\hat{\varphi}_{13} = 0.146(0.074)$]. Moreover, the variance of the shared random effect is significant [$\hat{\sigma}^2 = 0.116(0.011)$], which reconfirms the necessity of the nonignorable missing mechanism for "hippocampal volume" and implies that some omitted clinical or genetic indicators may simultaneously influence the outcomes of the observation process, transition probabilities, the membership of normal aging, and the missing probability of hippocampal volume.

Furthermore, we estimated the hidden states of subjects who are in AD progression on the basis of (13). Over 95% posterior transition patterns exhibited an early state to a severe one, which agrees with the irreversibility of AD. The total number of participants in this analysis is 651. Each participant was diagnosed as CN, MCI, or AD at five occasions. Appendix S1 compares subjects' estimated hidden states with their diagnosed status given by doctors at all the occasions. For CN and AD states, majority of the estimated states were consistent with those diagnosed by doctors. However, for MCI state, 1311 (78.8%) of symptomatic stages were diagnosed as MCI by doctors but classified into CN by our proposed method. This inconsistency may result from the vague demarcation between CN and MCI, which was also found and discussed in existing literature.[30]

To investigate the sensitivity of Bayesian estimation and model selection to the prior input, the preceding ADNI analysis was repeated with ad hoc perturbations of the current prior input. The obtained results were similar and not reported.

6 | DISCUSSION

We proposed a two-part HMM to analyze longitudinal semicontinuous data in the presence of nonignorable missing covariates. The longitudinal semicontinuous response was represented by two random variables: a binary indicator variable for determining the occurrence of excess zeros in all occasions and a continuous variable for examining its actual value. A logistic regression model was used to link the binary indicator with potential covariates. Furthermore, we considered an HMM to analyze the longitudinal continuous responses and introduced a shared random effect to the missing data model to address the nonignorability of missingness in covariates. We developed a hybrid algorithm to conduct efficient statistical inference.

The proposed model assumed that covariates effects on outcomes have parametric forms. However, the relationships between covariates and outcomes are seldom known a priori. Generalizing the existing model framework to a nonparametric context for further enhancing the model flexibility and analytic power is of substantial interest. Moreover, we simply used FAQ to represent patients' cognitive ability. A highly comprehensive measurement of cognitive ability should incorporate other relevant assessments, such as the Alzheimer's Disease Assessment Scale and Mini-Mental State Examination. Adopting a factor analysis model to group such highly correlated but different cognitive tests into an integrated latent variable can reduce information loss and improve model interpretability. In this case, the binary indicator variable for excess zeros should be redefined accordingly because the observed longitudinal response variable is no longer a scalar y_{it} but a vector $\mathbf{y}_{it}$ with several indicators. Furthermore, the proposed model involves a logit link in the binary indicator and missing data models. The MH algorithm, which does not have the same speed as the Gibbs samplers, is currently used to manage such a logistic likelihood. A possible future interest is to use the Pólya-Gamma augmentation to conduct a full Bayesian inference for the proposed model. This type of efficient sampling method possesses a high potential for highly efficient Gibbs updates in the presence of logistic likelihoods, but requires further investigation in the context of the current complex model framework. Finally, the binary indicator, conditional regression, and missing data models were assumed to share the same random effect w_i. This is a strong assumption for multiple outcomes. Extending the current framework to allow different random effects in the three submodels is of great interest in the future research. We may consider a multivariate normal distribution for these three random effects[39] or adopt copula representation to formulate the association structure among different random effects.[40] This extension enables us to accommodate various

sources of heterogeneity within different outcomes and simultaneously investigate their relationships, thereby enhancing model capability and flexibility. However, developing such statistical approaches raises new challenges and requires further investigation.

ACKNOWLEDGEMENTS

The authors thank the co-editor, an associate editor, and two referees for their insightful comments and suggestions that greatly improved the article. This research was fully supported by GRF grants 14303017 and 14301918 from the Research Grant Council of Hong Kong Special Administraive Region.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

ORCID


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

REFERENCES

1. Scott SL, James GM, Sugar CA. Hidden Markov models for longitudinal comparisons. *J Am Stat Assoc*. 2005;100(470):359-369.
2. Altman RMK. Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *J Am Stat Assoc*. 2007;102(477):201-210.
3. Bartolucci F, Farcomeni A. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *J Am Stat Assoc*. 2009;104(486):816-831.
4. Cappé O, Moulines E, Rydén T. *Inference in Hidden Markov Models*. New York, NY: Springer; 2005.
5. Maruotti A. Mixed hidden Markov models for longitudinal data: an overview. *Int Stat Rev*. 2011;79(3):427-454.
6. Song XY, Xia YM, Zhu HT. Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics*. 2017;73(1):313-323.
7. Olsen MK, Schafer JL. A two-part random-effects model for semicontinuous longitudinal data. *J Am Stat Assoc*. 2001;96(454):730-745.
8. Smith VA, Neelon B, Preisser JS, Maciejewski ML. A marginalized two-part model for longitudinal semicontinuous data. *Stat Methods Med Res*. 2017;26(4):1949-1968.
9. Lee SY. *Handbook of Latent Variable and Related Models*. Amsterdam, The Netherlands: Elsevier; 2011.
10. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592.
11. Goldstein H, Carpenter J, Kenward MG, Levin KA. Multilevel models with multivariate mixed response types. *Stat Model*. 2009;9(3):173-197.
12. Bartolucci F, Farcomeni A. A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. *Biometrics*. 2015;71(1):80-89.
13. Cai JH, Ouyang M, Kang K, Song XY. Bayesian diagnostics of hidden Markov structural equation models with missing data. *Multivar Behav Res*. 2018;53(2):151-171.
14. De Santis SM, Bandyopadhyay D. Hidden Markov models for zero-inflated Poisson counts with an application to substance use. *Stat Med*. 2011;30(14):1678-1694.
15. Morris JC, McKeel DW, Martha S, et al. Very mild Alzheimer's disease: informant-based clinical, psychometric, and pathologic distinction from normal aging. *Neurology*. 1991;41(4):469-469.
16. Ohnishi T, Matsuda H, Tabira T, Asada T, Uno M. Changes in brain morphology in Alzheimer disease and normal aging: Is Alzheimer disease an exaggerated aging process? *AJNR Am J Neuroradiol*. 2001;22(9):1680-1685.
17. Li D, Iddi S, Thompson WK, Donohue MC, Alzheimer's Disease Neuroimaging Initiative. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Stat Methods Med Res*. 2019;28(3):835-845.
18. Zhang L, Lim CY, Maiti T, et al. Analysis of conversion of Alzheimer's disease using a multi-state Markov model. *Stat Methods Med Res*. 2019;28(9):2801-2819.
19. Frühwirth-Schnatter S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J Am Stat Assoc*. 2001;96(453):194-209.
20. Lee SY, Song XY. Bayesian model selection for mixture of structure equation models with an unknown number of components. *Brit J Math Stat Psy*. 2003;56:145-165.
21. Jasra A, Holmes CC, Stephens DA. Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat Sci*. 2005;20:50-67.
22. Feng XN, Wang GC, Wang YF, Song XY. Structure detection of semiparametric structural equation models with Bayesian adaptive group lasso. *Stat Med*. 2015;34(9):1527-1547.
23. Li J, Wang Z, LR, Wu R. Bayesian group LASSO for nonparametric varying-coefficient models with application to functional genome-wide association studies. *Ann Appl Stat*. 2015;9(2):640.
24. Guo R, Zhu H, Chow S-M, Ibrahim JG. Bayesian lasso for semiparametric structural equation models. *Biometrics*. 2012;68(2):567-577.

25. Mason A, Richardson S, Best N. Two-pronged strategy for using DIC to compare selection models with non-ignorable missing responses. *Bayesian Anal*. 2012;7(1):109-146.
26. Spiegelhalter DJ, Best NG, Carlin BP, Vander Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):583-639.
27. Jr Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Cardiovasc Magn Reson*. 2008;27(4):685-691.
28. Gómez-Isla T, Price JL, McKeel Jr DW, Morris JC, Growdon JH, Hyman BT. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J Neurosci*. 1996;16(14):4491-4500.
29. Houlden H, Crook R, Backhovens H, et al. ApoE genotype is a risk factor in nonpresenilin early-onset Alzheimer's disease families. *Am J Med Genet*. 1998;81(1):117-121.
30. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3):183-194.
31. Mosconi L, De Santi S, Li J, et al. Hippocampal hypometabolism predicts cognitive decline from normal aging. *Neurobiol Aging*. 2008;29(5):676-692.
32. Kantarci K, Gunter JL, Tosakulwong N, et al. Focal hemosiderin deposits and β-amyloid load in the ADNI cohort. *Alzheimers Dement*. 2013;9(5):S116-S123.
33. Cosentino S, Scarmeas N, Helzner E, et al. APOE ϵ4 allele predicts faster cognitive decline in mild Alzheimer disease. *Neurology*. 2008;70(19 Part 2):1842-1849.
34. De Santi S, Leon MJ, Rusinek H, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. *Neurobiol Aging*. 2001;22(4):529-539.
35. Podewils LJ, Guallar E, Kuller LH, et al. Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. *Am J Epidemiol*. 2005;161(7):639-651.
36. Mosconi L, Perani D, Sorbi S, et al. MCI conversion to dementia and the APOE genotype: a prediction study with FDG-PET. *Neurology*. 2004;63(12):2332-2340.
37. Simons JS, Spiers HJ. Prefrontal and medial temporal lobe interactions in long-term memory. *Nat Rev Neurosci*. 2003;4(8):637.
38. Costafreda SG, Dinov ID, Tu Z, et al. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *NeuroImage*. 2011;56(1):212-219.
39. Yau KKW. Multilevel models for survival analysis with random effects. *Biometrics*. 2001;57(1):96-102.
40. Rizopoulos D, Verbeke G, Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika*. 2008;95(1):63-74.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhou X, Kang K, Song X. Two-part hidden Markov models for semicontinuous longitudinal data with nonignorable missing covariates. *Statistics in Medicine*. 2020;39:1801–1816. https://doi.org/10.1002/sim.8513