Computational Statistics and Data Analysis 173 (2022) 107482

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Varying-coefficient hidden Markov models with zero-effect regions

Hefei Liu [a], Xinyuan Song [b,*,1], Baoxue Zhang [a,*,2]

[a] Capital University of Economics and Business, China
[b] The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 13 May 2021
Received in revised form 10 March 2022
Accepted 25 March 2022
Available online 8 April 2022

Keywords:
Bayesian method
Longitudinal data
Spline approximation
Varying-coefficient models
Zero-effect regions

ABSTRACT

In psychological, social, behavioral, and medical studies, hidden Markov models (HMMs) have been extensively applied to the simultaneous modeling of longitudinal observations and the underlying dynamic transition process. However, the existing HMMs mainly focus on constant-coefficient HMMs. This study considers a varying-coefficient HMM, which enables simultaneous investigation of the dynamic covariate effects and between-state transitions. Moreover, a soft-thresholding operator is introduced to detect zero-effect regions of the coefficient functions. A full Bayesian approach with a hybird Markov chain Monte Carlo algorithm that combines B-spline approximation and penalization technique is developed for statistical inference. The empirical performance of the propose method is evaluated through simulation studies. An application to a study on the Alzheimer's Disease Neuroimaging Initiative dataset is presented.

1. Introduction

Heterogeneous longitudinal data are frequently encountered in the medical, behavioral, social-economic, environmental, and psychological sciences. Hidden Markov model (HMM), which consists of a transition model to describe the dynamic transition of hidden states and a conditional regression model for examining state-specific covariate effects on the response of interest, provides a useful model framework to accommodate heterogeneous and longitudinal features (Cappé et al., 2005; Scott et al., 2005; Altman, 2007; Robert et al., 2000; Bartolucci and Farcomeni, 2009). HMM and its variants have attracted significant attention from various disciplines, owing to its ability to simultaneously reveal the longitudinal association structure and dynamic heterogeneity of the observed process (Gupta et al., 2007; Ip et al., 2013; Song et al., 2017; Marino et al., 2018; Yen and Chen, 2018; Kang et al., 2019; Liu et al., 2021). Despite the rapid developments and wide applications of HMMs, the existing literature has mainly focused on constant-coefficient HMMs, in which the conditional regression model assumes invariant covariate effects in each state. However, these effects may vary according to certain variables, leading to functional coefficients of interest. A highly comprehensive analysis should consider varying coefficients in the context of HMMs to discover the state-specific dynamic patterns of covariate effects on the response variable.

Varying-coefficient models (Hastie and Tibshirani, 1993) are important techniques for exploring the dynamic relationships between a response variable and potential predictors through coefficients that vary with certain modifiers (Hoover et

* Corresponding author.
E-mail addresses: xysong@sta.cuhk.edu.hk (X. Song), zhangbaoxue@cueb.edu.cn (B. Zhang).
[1] Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.
[2] School of Statistics, Capital University of Economics and Business, Beijing, China.

https://doi.org/10.1016/j.csda.2022.107482
0167-9473/

al., 1998; Fan and Zhang, 1999; Chiang et al., 2001; Eubank et al., 2004; Ma and Song, 2015). Early development on varying-coefficient models mainly focuses on estimation of coefficient functions. Hoover et al. (1998) estimated coefficient functions through smoothing splines. Fan and Zhang (1999) proposed a two-step procedure to estimate varying coefficients. Wu et al. (2000) considered a smoothing spline method for varying-coefficient models with repeated measurements, and Chiang et al. (2001) further extended their method to accommodate repeatedly measured dependent variables. Zhang et al. (2002) proposed local polynomial fitting in a semivarying coefficient model. Huang et al. (2002, 2004) used regression splines to analyze varying-coefficient models with longitudinal data. Ye et al. (2019) proposed a nested expectation–maximization algorithm with penalized likelihood to analyze finite mixture varying-coefficient models. Meanwhile, varying-coefficient models have also been developed in the Bayesian framework (e.g., Biller and Fahrmeir, 2001; Dunson et al., 2003; Haneuse et al., 2008). In the context of mixture varying-coefficient models, Lu and Song (2012) developed a Bayesian penalized spline method to estimate component-specific varying coefficients. However, to the best of our knowledge, no existing study has investigated varying-coefficient HMMs.

In the analysis of varying-coefficient models, an important issue is the detection of zero-effect regions, especially when the effect changes from positive to negative or vice versa. For instance, in the study on opioid use for pain treatment, opioid use is found to be ineffective in relieving pain in certain BMI ranges (Yang, 2020). A common method for the effect shrinkage is to introduce a soft-thresholded function, which was originally proposed by Donoho and Johnstone (1994) for wavelet shrinkage. Tibshirani (1996) discussed the relationship between the least absolute shrinkage and selection operator (lasso) and soft-thresholded estimators; he pointed out that the lasso estimator is reduced to a soft-thresholded estimator when the design matrix is orthonormal. Kang et al. (2018) proposed a soft-thresholded Gaussian process prior to spatial variable selection for scalar-on-image regression. Yang (2020) further introduced a soft-thresholding operator to varying-coefficient models and demonstrated that the soft-thresholded estimator can effectively uncover zero-effect regions of varying coefficients. Nevertheless, the preceding works did not consider HMMs, and thus cannot accommodate the dynamic heterogeneity, state-specific functional covariate effects, and state-specific zero-effect regions of varying coefficients.

In the present study, we consider a varying-coefficient HMM to analyze multivariate, heterogeneous, and longitudinal data. The model comprises two major components. The first component is a continuation-ratio logit transition model for investigating the influence of potential covariates on the probabilities of transition from one hidden state to another. The second component is a varying-coefficient regression model for exploring the dynamic relationships between the response and potential covariates. Moreover, a soft-thresholding operator is introduced to varying coefficients to uncover their zero-effect regions. This study contributes the literature in several aspects. First, it is the first to consider a varying-coefficient HMM. Thus, the dynamic patterns of heterogeneous covariate effects can be simultaneously investigated. Second, we introduce a soft-thresholding operator to the proposed model, thereby enabling us to effectively detect state-specific zero-effect regions and quantify the associated uncertainty. Lastly, we develop a full Bayesian approach and a hybrid algorithm that combines B-spline approximation and penalization techniques to conduct statistical inference. The proposed inference procedure is different from the existing Gaussian process-based approach (Kang et al., 2018) and frequentist method (Yang, 2020). The superiority of the proposed method is demonstrated through numerical studies.

The proposed method is motivated by an actual study conducted by the Alzheimer's Disease Neuroimaging Initiative (ADNI). With the increase in human life expectancy, Alzheimer's disease (AD) has attracted considerable attention from society and medical research. 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. Functional assessment questionnaire (FAQ), an assessment of abilities to function independently in daily life, has been widely used to monitor cognitive impairment due to short-term memory loss. The FAQ score and a set of biomarkers, such as gender, age, educational levels, hippocampal volume, and apolipoprotein E (APOE)-ϵ4, were collected across multiple time points in the ADNI dataset. The main goal of this study is to investigate the potential risk factors of AD from several perspectives, including (i) identifying various AD pathology states and the factors that affect the between-state transition, (ii) investigating the state-specific associations between FAQ and its potential risk factors, and (iii) revealing new insights into the pathological mechanism of AD. We are particularly interested in how participants' education years relate to their cognitive impairment interactively with other potential risk factors, such as hippocampal volume and age, considering that cognitive ability is highly associated with education level. Moreover, previous studies (e.g., Kang et al., 2019) revealed that the effects of hippocampal volume and age may exhibit zero-effect regions in early cognitive impairment stage. The proposed model accommodates these features and enables us to reveal the underlying hidden states of cognitive impairment and their transition mechanism, examine the change in the effects of risk factors over different education levels in each state, and uncover the zero-effect regions of the covariate effects.

The remainder of this article is organized as follows. Section 2 describes the proposed varying-coefficient HMM. Section 3 introduces a Bayesian approach, which incorporates B-splines for approximating unknown coefficient functions, the soft-thresholding operator for detecting zero-effect regions of the varying coefficients, the sieve spline space and penalized likelihood procedure for identifying a unique between-function mapping, and Markov chain Monte Carlo (MCMC) methods for posterior sampling. Section 4 conducts simulation studies to assess the empirical performance of the proposed method. Section 5 presents an application regarding the ADNI study. Section 6 presents the conclusion. Technical details are provided in the Appendix.

2. Model

For subject $i = 1, \cdots, N$ at $t = 1, \cdots, T$, we assume that $\{z_{it}\}_{t=1}^{T}$ is Markovian of order one, taking values in $\{1, \cdots, S\}$ and satisfying the following:

$$P(z_{it} = s|z_{i1}, \cdots, z_{i,t-1} = u) = P(z_{it} = s|z_{i,t-1} = u) \triangleq p_{itus},\ t \geq 2, \tag{1}$$

where p_{itus} is the transition probability of subject i from state $z_{i,t-1} = u$ at occasion $t-1$ to state $z_{it} = s$ at occasion t. The initial distribution of z_{i1} is assumed to be a multinomial distribution with probabilities $(p_1, p_2, \cdots, p_S)$, such that $p_s \geq 0$ and $\sum_{s=1}^{S} p_s = 1$. The distribution of $\{z_{it}\}_{t=1}^{T}$ is then fully determined by the transition probabilities and distribution of the initial state.

Following the existing literature (Ip et al., 2013; Song et al., 2017), we assume that hidden states $\{1, \cdots, S\}$ are ordered. Instead of directly modeling p_{itus}, we consider a continuation-ratio logistic model for $\eta_{itus} = P(z_{it} = s|z_{it} \geq s, z_{i,t-1} = u)$, $s = 1, \cdots, S-1$ as follows: for $i = 1, \cdots, N,\ t = 2, \cdots, T,\ u = 1, \cdots, S,\ s = 1, \cdots, S-1$,

$$\begin{aligned} \text{logit}(\eta_{itus}) &= \log \frac{P(z_{it} = s|z_{i,t-1} = u)}{P(z_{it} > s|z_{i,t-1} = u)} = \log \frac{p_{itus}}{p_{itu,s+1} + \cdots + p_{itus}} \\ &= \varphi_{us} + \boldsymbol{\alpha}' \mathbf{c}_{it}, \end{aligned} \tag{2}$$

where the left-hand side can be interpreted as the log odds of subject i transitioning to state s rather than to a higher state at occasion t given that his/her previous state at occasion $t-1$ is u, $\{\varphi_{us} : u = 1, \cdots, S,\ s = 1, \cdots, S-1\}$ are state-specific intercepts, $\mathbf{c}_{it}$ is an m-dimensional vector of covariates that may influence the transition probabilities, and $\boldsymbol{\alpha}$ is an m-dimensional vector of coefficients that can be interpreted as conditional log odds ratios in a logistic regression. On the basis of Equation (2), we can obtain the followings:

$$\begin{aligned} p_{itus} &= \frac{\exp(\varphi_{us} + \boldsymbol{\alpha}' \mathbf{c}_{it})}{\prod_{k=1}^{s} \{1 + \exp(\varphi_{uk} + \boldsymbol{\alpha}' \mathbf{c}_{it})\}},\ \ s = 1, \ldots, S-1, \\ p_{ituS} &= \frac{1}{\prod_{k=1}^{S-1} \{1 + \exp(\varphi_{uk} + \boldsymbol{\alpha}' \mathbf{c}_{it})\}}. \end{aligned} \tag{3}$$

For $i = 1, \cdots, N, t = 1, \cdots, T$, let y_{it} be an observation process. Given hidden state $z_{it} = s$, the conditional varying-coefficient regression model is defined as follows:

$$[y_{it}|z_{it} = s] = b_s + \sum_{j=1}^{r} g_{sj}(w_{itj}) x_{itj} + \epsilon_{it}, \tag{4}$$

where b_s is a state-specific intercept; x_{itj} represents observed covariates; w_{itj} is an independent variable; $g_{sj}(w_{itj})$ is a state-specific varying coefficient, which allows the effect of x_{itj} to be state-specific and vary smoothly over the group stratified by w_{itj}, and thus permits nonlinear interaction between w_{itj} and x_{itj}; r is the number of covariates; and ϵ_{it} is a random error independently distributed according to $N(0, \sigma_s^2)$. For function $g_{sj}(w_{itj})$, we make three assumptions, namely, (i) continuity, (ii) zero-effect regions exist, and (iii) piecewise smoothness. The proposed model includes three different types of covariates, c_{it}, w_{itj}, and x_{itj}, where c_{it} and w_{itj} or x_{itj} can be overlapped. We may determine these covariates in substantive research based on study objectives, subjects' knowledge, variables' meanings, and experts' suggestions (see Section 5). In addition, we can use a model selection criterion, such as the deviance information criterion (DIC, Celeux et al., 2006), to select an appropriate model if needed.

3. Bayesian inference

3.1. Soft-thresholded operator

To detect the zero-effect regions of a functional coefficient $g(w)$, we consider soft thresholding operator ζ_λ (Kang et al., 2018) to map $g(w)$ near zero to exact zero as follows:

$$\zeta_\lambda(h) = \begin{cases} 0, & |h| \leq \lambda, \\ \text{sgn}(h)(|h| - \lambda), & |h| > \lambda, \end{cases} \tag{5}$$

where $\text{sgn}(h) = 1$ if $h > 0$, $\text{sgn}(h) = -1$ if $h < 0$, and $\text{sgn}(0) = 0$. Equivalently, $\zeta_\lambda(h) = (h - \lambda) I\{h > \lambda\} + (h + \lambda) I\{h < -\lambda\}$. Thresholding parameter $\lambda > 0$ determines the degree of sparsity. As illustrated by Yang (2020), for any function $g(w)$ and any $\lambda > 0$, at least one $h(w)$ exists, such that $\zeta_\lambda(h(w)) = g(w)$.

By substituting $g_{sj}(w_{itj})$ with $\zeta_{\lambda_{sj}}(h_{sj}(w_{itj}))$ in model (4), we obtain the following:

$$[y_{it}|z_{it}=s]=b_s+\sum_{j=1}^{r}\zeta_{\lambda_{sj}}(h_{sj}(w_{itj}))x_{itj}+\epsilon_{it}. \tag{6}$$

Threshold parameter λ_{sj} plays a vital role in the soft-threshold method. Yang (2020) pointed out that any given positive real number can be used as the threshold. In this study, we regard λ_{sj} as a random variable and estimate it together with other parameters.

3.2. B-splines

B-splines are commonly used to approximate a nonparametric function. The basic idea of the spline approximation is to estimate each unknown smooth function through the sum of B-spline basis functions. Then, $h_{sj}(w)$ can be approximated through B-splines as follows:

$$h_{sj}(w)=\sum_{l=1}^{L}\beta_{sjl}B_l(w)=\boldsymbol{\beta}_{sj}^T\boldsymbol{B}(w), \tag{7}$$

where L is the number of splines, $\boldsymbol{\beta}_{sj}=(\beta_{sj1},\ldots,\beta_{sjL})^T$ is the vector of unknown coefficients, $B_l(\cdot)$s are B-spline basis functions of cubic order, and $\boldsymbol{B}(w)=(B_1(w),\ldots,B_L(w))^T$. For notation simplicity, the number of knots L and B-spline basis functions $B_l(\cdot)$s are assumed to be invariant across different s and j. Extension to the case, where L and $B_l(\cdot)$s are s- and j-specific, is straightforward. With the B-spline approximation, the "working model" can be written as follows:

$$[y_{it}|z_{it}=s]=b_s+\sum_{j=1}^{r}\zeta_{\lambda_{sj}}\Big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\Big)x_{itj}+\epsilon_{it}. \tag{8}$$

3.3. Model identification

The soft-thresholding operator $\zeta_\lambda(\cdot)$ defined in Equation (5) maps different smooth $h(\cdot)$s with different thresholding parameters to the same $g(\cdot)$. Even for a fixed λ, $\zeta_\lambda(h(\cdot))$ may not be uniquely defined. Thus, $\zeta_\lambda(h(w))$ is not estimable without further constraints. We follow Yang (2020) to use the methods of spline sieve space and penalized loss function to determine a unique $\zeta_\lambda(h(\cdot))$ for $g(\cdot)$. Spline sieve space method determines a reasonable number of knots in accordance with the sample size to prevent overfitting and reduce the scope of function $h(\cdot)$. Then, to identify a unique $h_{sj}(w)$ that corresponds to $g_{sj}(w)$, we introduce a penalty term to penalize "the length of $h_{sj}(w)$," namely, $\|h_{sj}(w)\|\overset{D}{=}\big(\boldsymbol{\beta}_{sj}^T\mathbf{B}_s(w)\big)^2$, such that a unique $h_{sj}(w)$ with minimal length within the range determined by the spline sieve space is identified. This process is equivalent to minimizing the following penalized likelihood function.

Let $\mathbf{Y}=\{y_{it},i=1,\cdots,N;t=1,\cdots,T\}$, $\mathbf{Z}=\{z_{it},i=1,\cdots,N;t=1,\cdots,T\}$, and $\boldsymbol{\theta}$ be the vector of all unknown parameters in the model. With the B-splines approximation, the complete-data likelihood function is as follows:

$$\begin{aligned}Lik&=P(\mathbf{Y},\mathbf{Z}|\boldsymbol{\theta})=P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\theta})P(\mathbf{Z}|\boldsymbol{\theta})\\&=\prod_{s=1}^{S}\prod_{i=1}^{N}\prod_{t=1}^{T}\left[\frac{1}{\sqrt{2\pi}\sigma_s}\exp\left\{-\frac{1}{2\sigma_s^2}\Big[y_{it}-b_s-\sum_{j=1}^{r}\zeta_{\lambda_{sj}}\Big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\Big)x_{itj}\Big]^2\right\}\right]^{I(z_{it}=s)}\\&\times\prod_{u=1}^{S}\prod_{s=1}^{S}\prod_{i=1}^{N}\prod_{t=2}^{T}p_{itus}^{I(z_{i,t-1}=u;z_{it}=s)}\prod_{s=1}^{S}p_s^{n_{s1}},\end{aligned}$$

where $I(\cdot)$ is the indicator function, and n_{s1} is the sample size of state s at initial time $t=1$. By introducing a penalty to $\|h_{sj}(w)\|=\big(\boldsymbol{\beta}_{sj}^T\boldsymbol{B}_s(w)\big)^2$, the penalized complete-data likelihood function can be expressed as follows:

$$Lik_{pen}=Lik-\sum_{s=1}^{S}\sum_{j=1}^{r}\rho_{sj}(\boldsymbol{\beta}_{sj}^T\boldsymbol{B}_{sj})^2, \tag{9}$$

where $\boldsymbol{B}_{sj}$ is the submatrix of $\boldsymbol{B}_j=[B_l(w_{itj})]_{NT\times L}$ with the rows corresponding to $Z_{it}\neq s$ deleted, and ρ_{sj} is a turning parameter that determines the magnitude of the adaptive penalty imposed on $\|h_{sj}(\cdot)\|$ for $s=1,\cdots,S$ and $j=1,\cdots,r$.

In this study, we follow Berry et al. (2002) and Kang et al. (2018) to set $\rho_{sj}=1/n_s^2$ to standardize $\|h_{sj}(w)\|$, where $n_s=\sum_{i=1}^{N}\sum_{t=1}^{T}I(z_{it}=s)$. This standardization can eliminate the effect of the state-specific sample size on the identification of a minimal-length $h_{sj}(w)$. Let $\mathbf{D}_{sj}=\mathbf{B}_{sj}\mathbf{B}_{sj}^T$, the penalty term in (9) can then be rewritten as $\sum_{s=1}^{S}\sum_{j=1}^{r}(\frac{1}{n_s}\boldsymbol{\beta}_{sj}^T\boldsymbol{B}_{sj})\big(\frac{1}{n_s}\boldsymbol{\beta}_{sj}^T\mathbf{B}_{sj}\big)^T=$

$\sum_{s=1}^{S}\sum_{j=1}^{r}\boldsymbol{\beta}_{sj}^{T}\left(n_s^2\mathbf{D}_{sj}^{-1}\right)^{-1}\boldsymbol{\beta}_{sj}$, which can be implemented under the Bayesian framework by assigning the normal prior distribution $N(\mathbf{0}, n_s^2\boldsymbol{D}_{sj}^{-1})$ to $\boldsymbol{\beta}_{sj}$.

In Bayesian analysis, label switching is another identifiability issue for mixture-type models. We use the permutation sampler (Frühwirth-Schnatter, 2001) to address the label switching problem. The permutation sample is implemented by inserting an additional step in each MCMC iteration to verify the proposed constraint and then permute labels accordingly.

3.4. Prior specification

To conduct a Bayesian analysis, prior distributions for the unknown parameters should be specified. Following the existing literature (Lu and Song, 2012; Liu et al., 2021; Ressell and Telesca, 2017), we consider the following prior distributions for the unknown parameters:

$$\begin{aligned}&(p_1,\cdots,p_S)\sim Dirichlet(a_0,\cdots,a_0),\\&\varphi_{us}\sim N(\varphi_{us0},\sigma_{us0}^2),\ \ \boldsymbol{\alpha}\sim N(\boldsymbol{\alpha}_0,\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}),\ \ b_s\sim N(\mu_{s0},\sigma_{s0}^2),\\&\sigma_s^2\sim IG(\xi_{s0},\tau_{s0}),\ \ \lambda_{sj}\sim U(0,d_{sj0}),\ \ \boldsymbol{\beta}_{sj}\sim N(\mathbf{0},n_s^2\boldsymbol{D}_{sj}^{-1}),\end{aligned}\tag{10}$$

where a_0, φ_{us0}, σ_{us0}^2, $\boldsymbol{\alpha}_0$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, μ_{s0}, σ_{s0}^2, ξ_{s0}, τ_{s0}, and d_{sj0} are hyperparameters with preassigned values, $Dirichlet(\cdot)$ denotes the Dirichlet distribution, $U(\cdot,\cdot)$ indicates the uniform distribution, and $IG(\cdot,\cdot)$ stands for the inverse Gamma distribution.

To prevent overfitting due to the use of a large number of knots, Eilers and Marx (1996) proposed using a difference penalty on coefficients of adjacent B-splines. In the Bayesian framework, the penalization can be implemented by assigning random walk priors to $\boldsymbol{\beta}_{sj}$ (Lang and Brezger, 2004; Lu and Song, 2012). Unlike the preceding works that introduce random walk priors to $\boldsymbol{\beta}_{sj}$ to avoid a large number of knots L, this study adopts the spline sieve space method to determine a relatively small L based on the sample size to prevent overfitting. Therefore, we only need to consider the normal prior of $\boldsymbol{\beta}_{sj}$ specified in (10) to implement the penalization as described in Equation (9) in Section 3.3. Our numerical studies show that this normal prior works satisfactorily.

Based on the above arguments, two main differences exist between the proposed approach and the work of Kang et al. (2019). First, we consider a varying-coefficient conditional model to investigate how the effects of explanatory variables vary smoothly over the group stratified by other covariates and thus allow the nonlinear interaction effects between potential covariates and specific effect modifiers. By contrast, Kang et al. (2019) considered a semiparametric conditional model to examine the nonlinear effects of covariates on the outcome of interest. Therefore, the focuses of these two conditional regression models are different. Second, our study aims to uncover the zero-effect regions of the varying coefficients. We introduce a soft-thresholding operator to coefficient functions and develop a novel algorithm that incorporates B-spline approximation, spline sieve space method, and a prior distribution with penalty to facilitate a simultaneous parameter estimation and local sparsity detection under the Bayesian framework. On the contrary, Kang et al. (2019) did not consider such a local sparsity, thereby failing to denoise the estimated functional effects of interest.

3.5. Posterior sampling

The main task of posterior inference is to sample from $P(\boldsymbol{\theta}|\mathbf{Y})$. However, directly sampling from this posterior distribution is difficult because it involves high-dimensional integral with respect to $\mathbf{Z}$. We instead work on $p(\boldsymbol{\theta},\mathbf{Z}|\mathbf{Y})$ and use MCMC techniques, such as the Gibbs sampler and Metropolis–Hastings (MH) algorithm, to implement the posterior sampling. The full conditional distributions involved in the MCMC algorithm are outlined as follows:

(1) Full conditional distribution of $(p_1,p_2,\cdots,p_S)$:

$$(p_1,p_2,\cdots,p_S|\cdot)\sim Dirichlet(a_0+n_{11},a_0+n_{21},\cdots,a_0+n_{S1}).$$

(2) Full conditional distribution of φ_{us}:

$$p(\varphi_{us}|\cdot)\propto\exp\left\{\sum_{v=s}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T}\log p_{ituv}^{I(Z_{i,t-1}=u,z_{it}=v)}-\frac{(\varphi_{us}-\varphi_{us0})^2}{2\sigma_{us0}^2}\right\}.$$

(3) Full conditional distribution of $\boldsymbol{\alpha}$:

$$p(\boldsymbol{\alpha}|\cdot)\propto\exp\left\{\sum_{u=1}^{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T}\log p_{itus}^{I(Z_{i,t-1}=u,z_{it}=s)}-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^T\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\right\}.$$

(4) Full conditional distribution of σ_s^2:

$$\sigma_s^2\sim IG\left(\xi_{s0}+\frac{n_s}{2},\tau_{s0}+\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}I(z_{it}=s)\Big[y_{it}-b_s-\sum_{j=1}^{r}\zeta_{\lambda_{sj}}\Big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\Big)x_{itj}\Big]^2\right).$$

(5) Full conditional distribution of b_s:

$$b_s \sim N\left(\frac{2n_s\bar{Y}_s\sigma_{s0}^2 + 2\sigma_s^2\mu_{s0}}{n_s\sigma_{s0}^2 + \sigma_s^2},\ \frac{\sigma_{s0}^2\sigma_s^2}{n_s\sigma_{s0}^2 + \sigma_s^2}\right),\ \text{where}\ \bar{Y}_s = \frac{1}{n_s}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it} I(z_{it} = s).$$

(6) Full conditional distribution of λ_{sj}:

$$p(\lambda_{sj}|\cdot) \propto I(0 < \lambda_{sj} < d_{sj0}) \prod_{i=1}^{N}\prod_{t=1}^{T}\left[\exp\left\{-\frac{\left[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\left(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\right)x_{itj}\right]^2}{2\sigma_s^2}\right\}\right]^{I(z_{it}=s)}.$$

(7) Full conditional distribution of $\boldsymbol{\beta}_{sj}$:

$$p(\boldsymbol{\beta}_{sj}|\cdot) \propto \exp\left\{-\frac{1}{2n_s^2}\boldsymbol{\beta}_{sj}\boldsymbol{D}_{sj}\boldsymbol{\beta}_{sj} - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} I(z_{it} = s)\left[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\left(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\right)x_{itj}\right]^2}{2\sigma_s^2}\right\}.$$

(8) Full conditional distribution of z_{it}:

$$p(z_{it}|\cdot) \propto p(\mathbf{y}_i, z_{it}|\boldsymbol{\theta}) = p(\mathbf{y}_{i,1:t}, z_{it}|\boldsymbol{\theta})\, p(\mathbf{y}_{i,t+1:T}|z_{it}, \boldsymbol{\theta}) \doteq q_{it}(\mathbf{y}_i, z_{it}|\boldsymbol{\theta})\bar{q}_{it}(\mathbf{y}_i|z_{it}, \boldsymbol{\theta}),$$

where $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT})^T$; $\mathbf{y}_{i,1:t}$ and $\mathbf{y}_{i,t+1:T}$ denote the observations up to time t and from $t+1$ to T, respectively, for subject i.

The derivations of the full conditional distributions are provided in the Appendix.

4. Simulation

4.1. Simulation 1

In this section, we assess the performance of the proposed method through simulation studies. We consider an HMM with $S = 2$ as follows:

$$\text{logit}(\eta_{itus}) = \varphi_{us} + \boldsymbol{\alpha}'\mathbf{c}_{it}, \tag{11}$$

$$[y_{it}|z_{it} = s] = b_s + g_{s1}(w_{it1})x_{it1} + g_{s2}(w_{it2})x_{it2} + \epsilon_{it}, \tag{12}$$

where $\mathbf{c}_{it} = (c_{it1}, c_{it2}, c_{it3})'$, c_{it1} and c_{it2} are independently generated from Bernoulli(0.5); and $c_{it3} = w_{it1} = w_{it2}$ is generated from $U(-1.5, 1.5)$; x_{it1} and x_{it2} are generated from $N(0,1)$ and $t(4)$, respectively, where $U(\cdot,\cdot)$ indicates the uniform distribution; $t(4)$ stands for the t distribution with four degrees of freedom; $\varphi_{11} = \varphi_{21} = 0.1$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)' = (-0.5, 0.5, -0.8)'$, $b_1 = -1.5$, $b_2 = 1.5$, $\sigma_1^2 = \sigma_2^2 = 0.3$; $g_{11}(w) = \{1.8w + 0.9\}I_{\{w<-0.5\}} + \{1.2w - 0.6\}I_{\{w>0.5\}}$, $g_{12}(w) = \{w^3 - 0.125\}I_{\{w<0.5\}}$, $g_{21}(w) = \{sin(\frac{\pi}{2}w + \frac{\pi}{2})\}I_{\{-1<w<1\}}$, $g_{22}(w) = \{1.65 - \exp(w)\}I_{\{w<ln(1.65)\}}$; the initial probabilities of the hidden states are $(p_1, p_2) = (0.4, 0.6)$.

The hyperparameters of the prior distributions in (10) are assigned as follows: Prior (I) $a_0 = 1$, $\varphi_{us0} = 0$, $\sigma_{us0}^2 = 1.0$, $\boldsymbol{\alpha}_0 = (0,0,0)'$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}0} = \mathbf{I}$, $\mu_{s0} = 0$, $\sigma_{s0}^2 = 1000$, $\xi_{s0} = 1$, $\tau_{s0} = 0.1$, and $d_{sj0} = 0.5$. We consider $T = 5$ and $N = 200, 400$, such that the total sample sizes are 1000 and 2000 with approximately 500 and 1000 samples in each state. The number of spline knots L is determined by the spline sieve space method (Yang, 2020) as follows: $L = K + d - 1$, where d is the order of B-splines, and K must satisfy three conditions: (i) $K = O(N \times T)^{\nu}$, (ii) $\frac{1}{K-1} = O(N \times T)^{-\nu}$, and (iii) $0 < \nu < 0.5$. For example, when $N = 200$, $T = 5$, and $\nu = 0.3$, we can calculate that $K \approx 7$; thus, $L \approx 7 + 3 - 1 = 9$. Similarly, when $N = 400$, $T = 5$, and $\nu = 0.3$, $L \approx 11$. Thus, we set $L = 9$ and 11 when $N = 200$ and 400, respectively. For comparison, we also consider the conventional method without applying soft-threshold operator. In the conventional method, we assign $N(\mathbf{0}, 100\mathbf{I})$ prior for $\boldsymbol{\beta}_{sj}$. The prior specification for other parameters is the same as the previous specification.

We conduct few test runs to decide the number of burn-in iterations at convergence and find that 4,000 burn-in iterations are sufficient. Therefore, we collect 10,000 posterior samples after 4,000 burn-in iterations to conduct Bayesian inference. The Bayesian estimates of the unknown parameters and coefficient functions can be obtained through the means of their posterior samples. Table 1 summarizes the estimation results obtained using the proposed (Soft-thre) and conventional (No-thre) methods based on 100 replicated datasets. The bias (Bias) and root mean square error (Rmse) between the Bayesian estimates and true population values of the parameters are used to assess the performance of Bayesian estimation. All the bias and Rmse values are close to zero under the "Soft-thre" method, but some of them, especially those related to variance parameters, are relatively large under the "No-thre" method. Thus, the proposed method outperforms the conventional one, especially in reducing the biases of variance parameters. As expected, the performance improves when the sample size increases. On the basis of 100 replications, the means and standard deviations of the posterior mean estimates

Table 1
Parameter estimation results under proposed and conventional methods.

Para	$N = 200, T = 5$							
	Soft-thre				No-thre			
	Prior I		Prior II		Prior I		Prior II	
	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
p_1	0.001	0.001	-0.001	0.001	-0.004	0.001	0.003	0.001
p_2	-0.001	0.001	0.001	0.001	0.004	0.001	-0.003	0.001
$\varphi_{1,1}$	0.009	0.014	0.011	0.015	-0.020	0.021	-0.018	0.019
$\varphi_{2,1}$	-0.020	0.015	-0.017	0.018	-0.026	0.024	-0.021	0.018
α_1	0.005	0.014	-0.006	0.017	0.018	0.026	0.012	0.023
α_2	0.011	0.012	0.013	0.016	0.011	0.024	0.016	0.027
α_3	-0.012	0.014	-0.011	0.013	-0.010	0.021	-0.017	0.022
b_1	-0.001	0.002	0.002	0.001	0.009	0.008	-0.012	0.006
b_2	0.006	0.002	0.009	0.002	0.011	0.010	0.007	0.003
σ_1^2	0.020	0.001	0.021	0.002	0.051	0.004	0.056	0.005
σ_2^2	0.019	0.001	0.023	0.002	0.047	0.004	0.052	0.004
Para	$N = 400, T = 5$							
	Soft-thre				No-thre			
	Prior I		Prior II		Prior I		Prior II	
	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
p_1	-0.000	0.001	0.001	0.001	-0.002	0.001	-0.003	0.001
p_2	0.000	0.001	-0.001	0.001	0.002	0.001	0.003	0.001
$\varphi_{1,1}$	-0.009	0.010	-0.007	0.009	-0.017	0.019	-0.015	0.016
$\varphi_{2,1}$	-0.018	0.013	-0.014	0.012	-0.021	0.022	-0.019	0.018
α_1	-0.004	0.010	0.005	0.011	0.006	0.015	-0.009	0.017
α_2	0.006	0.012	0.007	0.014	0.013	0.014	0.016	0.020
α_3	0.010	0.011	0.011	0.015	0.011	0.013	-0.013	0.018
b_1	0.002	0.001	-0.003	0.001	0.006	0.004	0.010	0.003
b_2	-0.005	0.001	0.004	0.001	-0.009	0.005	-0.013	0.004
σ_1^2	0.018	0.001	0.016	0.001	0.046	0.003	0.041	0.003
σ_2^2	0.015	0.001	0.020	0.002	0.035	0.002	0.044	0.003

of threshold parameters λ_{11}, λ_{12}, λ_{21}, and λ_{22} are 0.253(0.102), 0.311(0.117), 0.319(0.122), and 0.293(0.126) when $N = 200$ and 0.259(0.126), 0.314(0.115), 0.340(0.126), and 0.314(0.127) when $N = 400$.

To assess the estimation performance of the varying coefficients, we define the integrated squared error (ISE) and average ISE (AISE) as follows: for $g_{sj}(w)$, $j = 1, \cdots, r$,

$$\mathrm{ISE}(g_{sj}(w_k)) = \frac{1}{n_g}\sum_{i=1}^{n_g}[\hat{g}_{sj}(w_k) - g_{sj}(w_k)]^2, \quad \mathrm{AISE} = \frac{1}{S \times r}\sum_{s=1}^{S}\sum_{j=1}^{r}\mathrm{ISE}(g_{js}(w_k)),$$

where n_g is the number of grids, and w_ks are grids inserted into the range of w. Table 2 reports the means of ISE and AISE computed based on 100 replications. The Soft-thre method apparently outperforms the No-thre method in terms of ISE and AISE. From the 100 replications, we select the estimated curve with median ISE to further compare the performance of the proposed and conventional methods. Fig. 1 shows the median-performing estimates together with their 95% pointwise credible intervals of the coefficient functions under $N = 400$. The left and right panels correspond to the "Soft-thre" and "No-thre" methods, respectively. The "Soft-thre" method can effectively uncover zero-effect and tail regions, whereas the "No-thre" method exhibits noticeable fluctuations around zero. The median-performing estimated functions under $N = 200$ and the averages of the estimated functions under both sample sizes demonstrate the superiority of the "Soft-thre" method in a similar manner; thus they are omitted.

4.2. Simulation 2

In this section, we conduct sensitivity analyses to examine whether the proposed method is sensitive to the prior specification, the number of knots used in B-spline approximation, the normality assumption of the random error, and the specification of S.

We disturb the hyperparameters in Prior (I) as follows: Prior (II) $a_0 = 2$, $\varphi_{us0} = 1$, $\sigma_{us0}^2 = 4.0$, $\boldsymbol{\alpha}_0 = (0, 0)'$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}0} = 2\mathbf{I}$, $\mu_{s0} = 1.0$, $\sigma_{s0}^2 = 100$, $\xi_{s0} = 1$, $\tau_{s0} = 0.01$, and $d_{sj0} = 0.7$. The estimation results under Prior (II) are also presented in Tables 1 and 2, indicating that both prior inputs produce similar results. Hence, the proposed Bayesian estimation is insensitive to the prior specification under consideration.

Next, we consider $L = 7$ when $N = 200$ and $L = 9$ when $N = 400$ to examine the effect of the number of knots on the estimation results, in particular the estimates of the varying coefficients. The prior and other model settings are the same

Table 2
Coefficient function estimation results under proposed and conventional methods.

N	ISE/AISE	Soft-thre		No-thre	
		Prior I	Prior II	Prior I	Prior II
$N = 200$	ISE(g_{11})	3.26	3.11	4.87	4.79
	ISE(g_{12})	2.91	3.04	3.72	3.84
	ISE(g_{21})	0.71	0.79	1.46	1.61
	ISE(g_{22})	1.53	1.36	2.05	2.19
	AISE	2.10	2.08	3.02	3.11
$N = 400$	ISE(g_{11})	2.05	2.18	3.95	4.11
	ISE(g_{12})	2.53	2.42	3.31	3.48
	ISE(g_{21})	0.53	0.66	1.10	1.07
	ISE(g_{22})	1.22	1.37	1.89	1.64
	AISE	1.58	1.66	2.56	2.58

ISE: integrated squared errors; AISE: averaged integrated squared errors; values are the means of ISE and AISE from 100 replications and multiplied by 10^3.

Table 3
Parameter estimation results under nonnormal error distributions.

Para	Nonnormal error distributions							
	Case 1		Case 2		Case 3		Case 4	
	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
p_1	0.003	0.001	-0.002	0.001	-0.009	0.002	0.003	0.001
p_2	-0.003	0.001	0.002	0.001	0.009	0.002	-0.003	0.001
$\varphi_{1,1}$	0.012	0.013	-0.016	0.014	-0.022	0.018	0.015	0.014
$\varphi_{2,1}$	-0.010	0.011	0.011	0.012	-0.017	0.014	-0.019	0.021
α_1	-0.018	0.014	0.016	0.011	-0.010	0.008	0.009	0.007
α_2	0.016	0.010	0.024	0.018	-0.021	0.017	-0.016	0.012
α_3	0.009	0.011	-0.020	0.016	0.013	0.011	0.018	0.014
b_1	-0.005	0.004	0.012	0.003	-0.027	0.022	-0.036	0.024
b_2	0.011	0.009	-0.007	0.004	0.022	0.018	-0.019	0.016
σ_1^2	0.039	0.007	0.041	0.011	0.068	0.016	0.089	0.021
σ_2^2	0.044	0.010	0.046	0.012	0.031	0.012	0.074	0.013

as in Simulation 1. We also collect 10,000 iterations after discarding 4,000 burn-ins to conduct posterior inference. The estimation results under such smaller numbers of knots are similar to the previous ones and not reported.

We also access the sensitivity of the proposed method to the error distribution. A total of 100 datasets is generated from the proposed model with each of the following nonnormal random errors: (1) $\epsilon_{it} \sim U(-1, 1)$, (2) $\epsilon_{it} \sim t(4)$, (3) $\epsilon_{it} \sim Gamma(1, 2)$, and (4) $\epsilon_{it} \sim 0.6N(-0.3, 0.3) + 0.4N(0.5, 1)$, where $Gamma(\cdot, \cdot)$ denotes the gamma distribution. The random errors are then transformed to have the same mean and variance as in Simulation 1 for comparison. The prior and other model settings are also the same as in Simulation 1. We conduct estimation using the proposed method by disregarding the nonnormality of ϵ_{it}. Table 3 presents the parameter estimates under the nonnormal random errors. Majority of parameter estimates perform similarly to those in Table 1, except that the variance estimates are slightly biased. Moreover, we select the estimated curve with median ISE from the 100 replications to examine the performance of the estimated functions when the normality of the random error is violated. Fig. 2 presents the estimated coefficient functions together with their 95% pointwise credible intervals under the nonnormal random error of Cases (2) and (3). Despite violation of the normality assumption, the proposed method can still detect the zero-effect regions and the patterns of the coefficient functions.

Finally, we disturb the number of hidden states S by setting $S = 1$ and 3 and reanalyze the datasets generated in Simulation 1 with $S = 2$ based on the misspecified models. Table 4 presents the posterior means of the intercept and residual variance in the conditional model (12) when S is misspecified as $S = 3$. The results indicate that the new state (state 2) tends to absorb samples of the two other states, resulting in an intercept estimate that somehow compromises those of the two other states and a considerably large residual variance estimate in the nonexistent state. Likewise, the estimation results of the parameters in (11) and the varying-coefficient functions in (12) are misleading when S is misspecified.

4.3. Simulation 3

In this section, we assess the performance of DIC in selecting the number of hidden states. We considered five competing models, M_1 to M_5, where M_k is a model defined by (12) with $S = k$, $k = 1, \cdots, 5$. Here, M_2 is the true model, and M_1, M_3, M_4, and M_5 are models with misspecified numbers of hidden states. In addition, we consider M_c with $S = 2$ and constant coefficients to examine whether DIC can select the varying-coefficient HMM over its constant-coefficient counterpart. The DIC values of the competing models are computed based on the 100 datasets generated in Simulation 1 and reported in Fig. 3, which suggests that the true model M_2 is consistently selected in each replication.

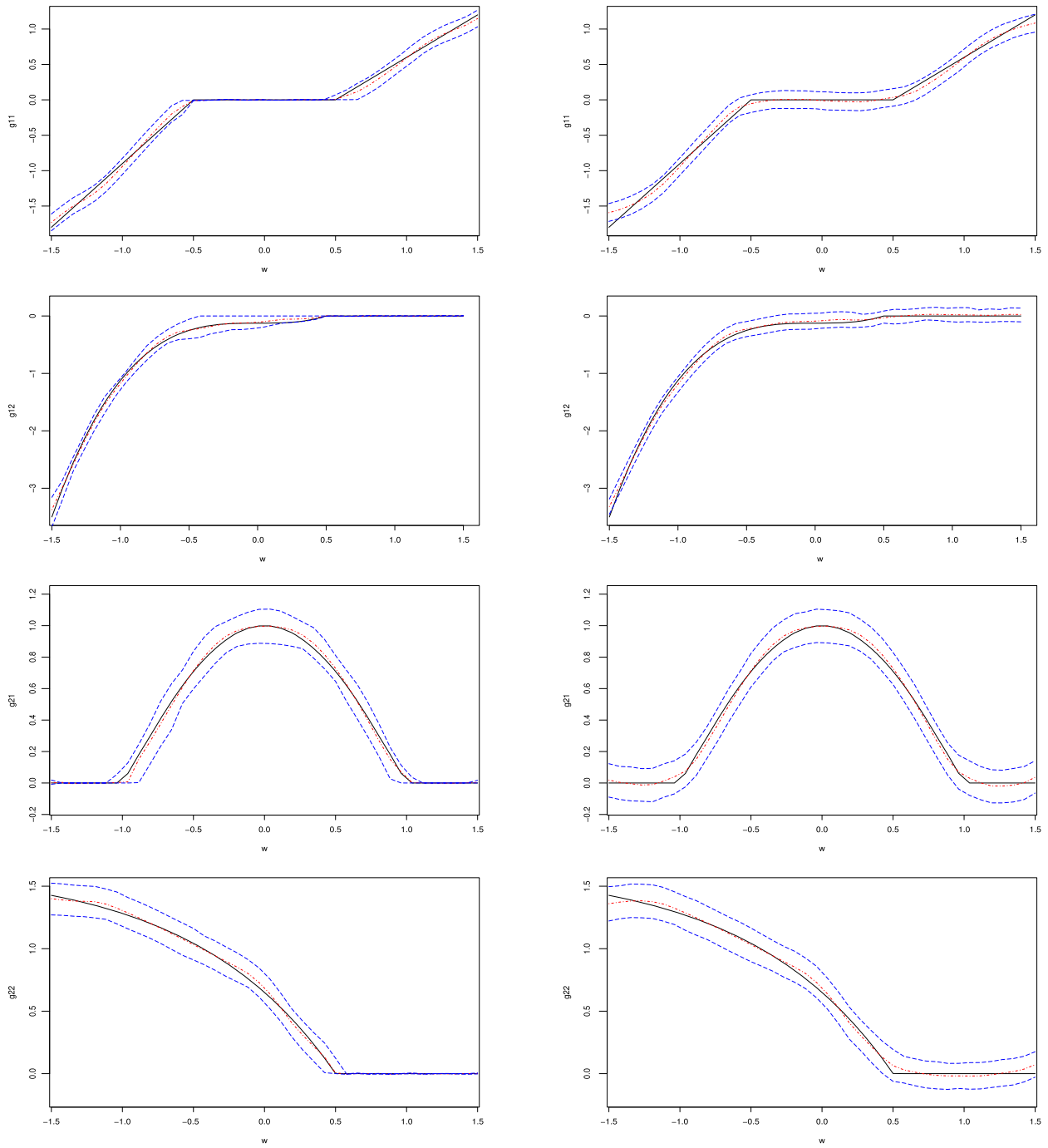


Fig. 1. Estimated coefficient functions with the median performance of 100 replications under $N = 400, T = 5$ and Prior (I). The left and right panels correspond to the "Soft-thre" and "No-thre" methods, respectively. Black solid line indicates the actual curve, red dotted line is the estimated curve, and blue dotted lines are the 2.5% and 97.5% pointwise quantiles. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

5. Application

We conducted an application to the ADNI study to demonstrate the utility of the proposed method. Detailed information about ADNI can be found in an official website (www.adni-info.org). We focused on 512 participants collected from ADNI-1 and included their clinical and genetic variables at five time points, including the baseline, sixth month, 12th month, 24th month, and 36th month. Thus, $N = 512$ and $T = 5$.

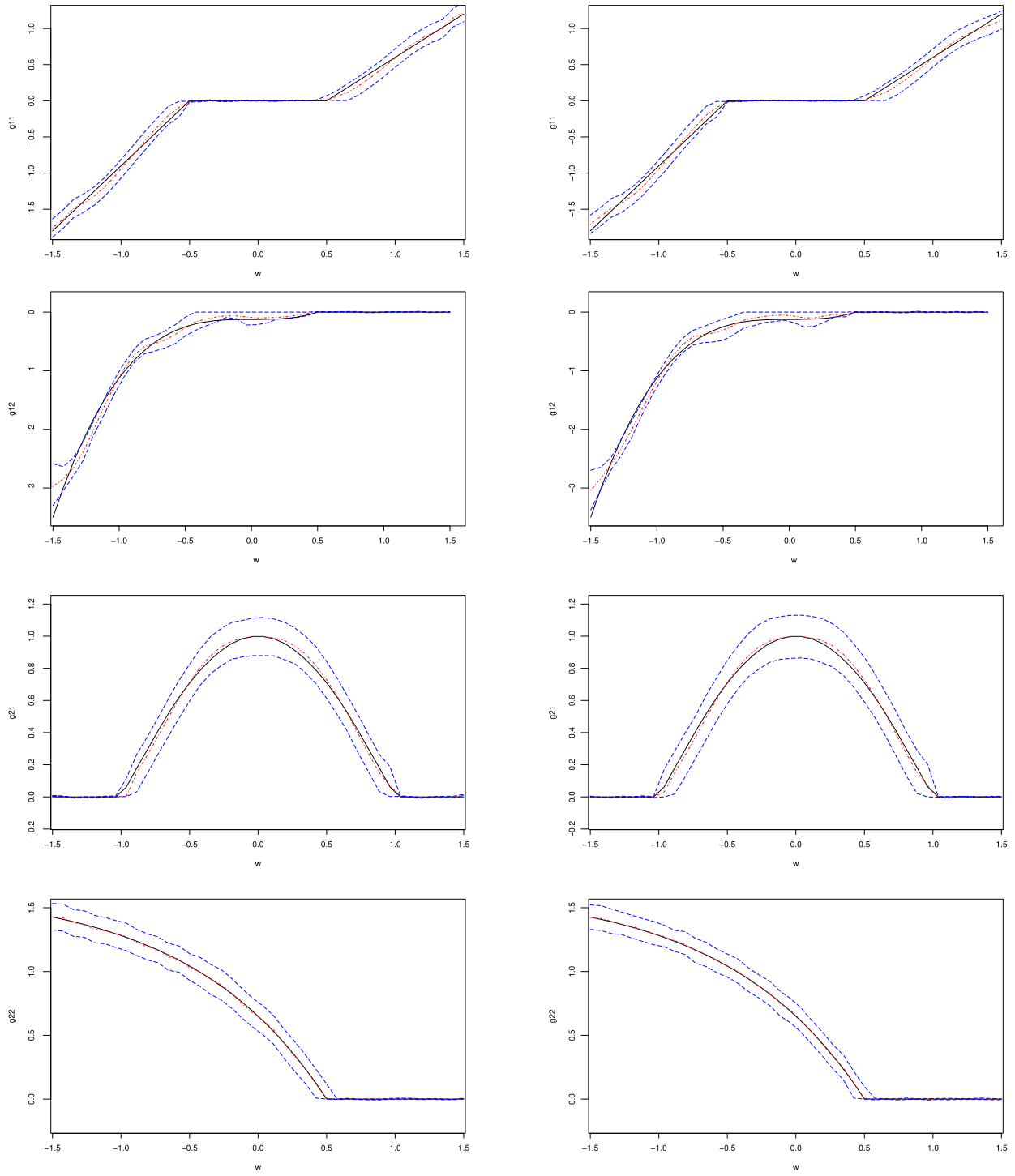


Fig. 2. Estimated coefficient functions with the median performance of 100 replications under nonnormal random error in Cases (2) and (3), $N = 400$, $T = 5$, and Prior (I). The left and right panels correspond to the nonnormal random errors of Cases (2) and (3), respectively. Black solid line indicates the actual curve, red dotted line is the estimated curve, and blue dotted lines are the 2.5% and 97.5% pointwise quantiles.

We used FAQ score, which ranges from 0 to 30 with high scores indicating poor cognitive ability, as response variable (y_{it}) to reflect cognitive impairment over time. Two continuous covariates, namely, the logarithm of the ratio of hippocampal volume over whole brain (x_{it1}) and age (x_{it2}), were considered in the conditional varying coefficient model. We included hippocampal volume and age as potential risk factors x_{itj}, $j = 1, 2$ in the conditional model because published reports (Kesslak et al., 1991; Jack et al., 1992; Dickerson and Wolk, 2013; Kang et al., 2019) revealed that the atrophy of the

Table 4
Estimation result with misspecified number of hidden states.

Parameters in conditional varying-coefficient regression model (12)											
State 1				State 2				State 3			
Para	True	Mean	SD	Para	True	Mean	SD	Para	True	Mean	SD
b_1	-1.5	-1.593	0.049	b_2	N.A.	-1.140	0.281	b_3	1.5	1.502	0.049
σ_1^2	0.3	0.515	0.078	σ_2^2	N.A.	1.384	0.936	σ_3^2	0.3	0.535	0.058

True: true value; Mean: posterior mean; SD: posterior standard deviation; N.A.: unavailable.

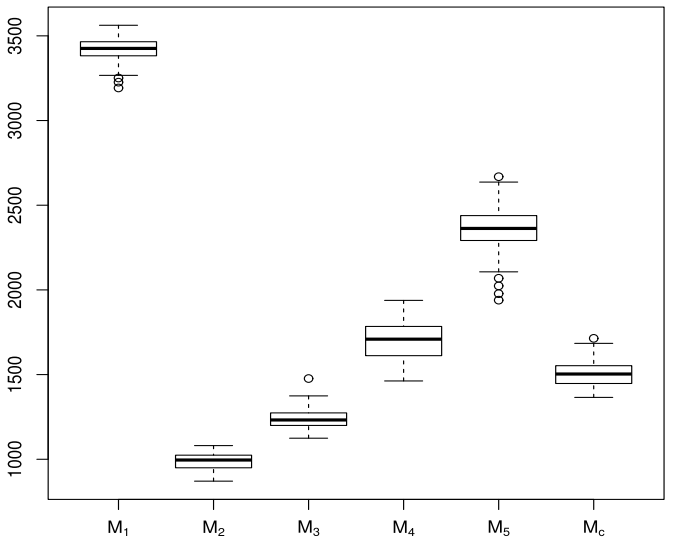


Fig. 3. Boxplot of the DIC values of the competing models in Simulation 3.

hippocampal formation and age were significant diagnostic markers of clinical dementia. Moreover, considering the evident positive effect of education level on cognitive impairment (e.g. Feng et al., 2020), we regarded continuous variable "year of education" as the independent variable ($w_{it1} = w_{it2} = w_{it}$) of the varying coefficients to examine its nonlinear interaction effects with hippocampal volume and age. The variable w_{it} took real numbers in the range of [6, 24] with decimal points. In addition, we considered two variables, gender (c_{it1}, 0 = female; 1 = male) and APOE-ε4 (c_{it2} and c_{it3}), coded as 0, 1, and 2, to denote the number of APOE-ε4 alleles in the transition model to examine their effects on the between-state transition because they were typical time-invariant individual characteristics and had been identified as potential risk factors for the AD progression (Okuizumi et al., 1994; Zhou et al., 2020). Four continuous variables, FAQ score, hippocampal volume, age, and years of education, were standardized prior to analysis. The aims of this analysis were (i) to identify the underlying hidden states of cognitive impairment and the factors that influenced the between-state transition, (ii) to investigate the effect of hippocampal volume and age on cognitive impairment in each hidden state and how these state-specific effects changed over education level, and (iii) to uncover the zero-effect regions of the varying coefficients.

In the implementation of the MCMC algorithm, we set the prior input to be the same as Prior (I) in Simulation 1 and $L = 11$ in cubic B-splines. The DIC was used to determine the number of hidden states. We considered varying- and constant-coefficient HMMs with the number of hidden states from 1 to 5. The corresponding DIC values were 16182, 12086, 10319, 11725, and 12743 for varying-coefficient HMMs and 17421, 15112, 14205, 14832, and 15327 for constant-coefficient HMMs. Thus, the three-state varying-coefficient HMM with the smallest DIC value was selected. Identifiability constraint $b_1 < b_2 < b_3$ was used to avoid label switching in estimation. In monitoring convergence, we draw three MCMC chains from different initial values and computed the estimated potential scale reduction (EPSR) value for all the parameters (Gelman, 1996). The EPSR plots shown in Fig. 4 indicated that the MCMC algorithm converged rapidly within 4,000 iterations. Therefore, we collected 10,000 observations after discarding 4,000 burn-in iterations. Table 5 and Fig. 5 present the estimation results. We have the following observations.

First, b_1, b_2, and b_3 are ranked in an ascending order, indicating that patients in state 1 obtained the lowest FAQ score, whereas those in state 3 got the highest. Patients' cognitive ability to function independently in daily life steadily deteriorated from state 1 to state 3. In accordance with the existing literature (e.g., Kantarci et al., 2013; Zhou et al., 2020), the three states can be explained as cognitive normal (CN) control, MCI, and AD, representing three major neurodegenerative stages in AD progression.

Second, in the transition model, $\hat{\alpha}_1 = -0.108(0.111)$ is the estimates of the posterior mean and posterior standard deviation (in parentheses) of gender (1 = male). Based on this result, we can conclude that gender only slightly (or in-

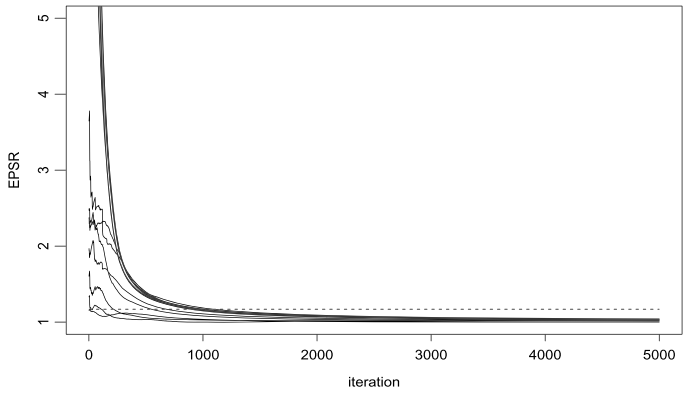


Fig. 4. Plot of estimated potential scale reduction (EPSR) values for the parameters in the ADNI-1 (Alzheimer's Disease Neuroimaging Initiative) data analysis. The horizontal dotted line is EPSR=1.2.

Table 5
Parameter estimation results in the ADNI study.

Parameters in conditional varying-coefficient regression model								
State 1			State 2			State 3		
Para	Est	SD	Para	Est	SD	Para	Est	SD
b_1	-0.614	0.001	b_2	-0.395	0.014	b_3	0.893	0.043
σ_1^2	0.014	0.002	σ_2^2	0.053	0.006	σ_3^2	1.132	0.059
Parameters in probability transition model								
p_1	0.334	0.022	p_2	0.471	0.026	p_3	0.195	0.021
$\varphi_{1,1}$	1.555	0.129	$\varphi_{2,1}$	1.019	0.289	$\varphi_{3,1}$	-1.667	0.147
$\varphi_{1,2}$	0.705	0.131	$\varphi_{2,2}$	-1.359	0.149	$\varphi_{3,2}$	-1.036	0.142
α_1	-0.108	0.111	α_2	-0.447	0.105	α_3	-0.723	0.165

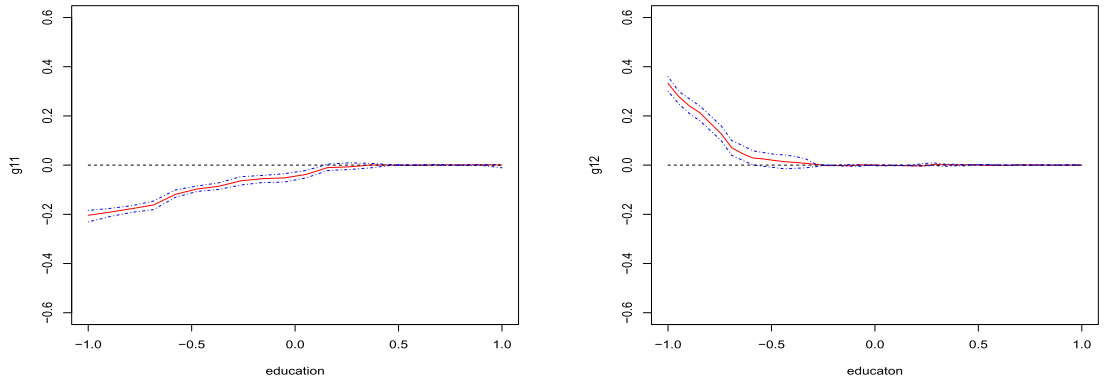
Para: parameter; Est: posterior mean estimate; SD: posterior standard deviation estimate.

significantly) affects the between-state transition. Moreover, $\hat{\alpha}_2 = -0.447(0.105)$ and $\hat{\alpha}_3 = -0.723(0.165)$ are the estimated effects of carrying one and two APOE-$\epsilon 4$ alleles, respectively, suggesting that carrying APOE-$\varepsilon 4$ alleles significantly and negatively affects logit$[P(z_{it} = s|z_{it} \geq s, z_{i,t-1} = u)]$ with a more pronounced effect when carrying two alleles. Thus, carrying APOE-$\varepsilon 4$ alleles increases the risk of transitioning to a worse state, and the risk is higher for two-allele carriers than for one-allele carries. This finding is in line with the existing literature (Lee et al., 2015).

Third, Fig. 5 shows that the effect of hippocampal volume on FAQ is negative in general, implying that participants with a greater hippocampal volume tend to have a lower FAQ score and thus less cognitive impairment. This finding is in good agreement with the previous medical reports that the hippocampus helps consolidate outside information from short- to long-term memory (Jack et al., 1992). Furthermore, the current study provides additional insights into the dynamic patterns and local sparsity of the effect. In CN and MCI states, hippocampal volume negatively affects cognitive impairment for participants with fairly low education, and such negative effect decreases when participants' education years increase, and then becomes zero when education years exceed 16 (university graduate). In AD state, the negative effect of hippocampal volume is much more pronounced. This pattern aligns with the finding of Kang et al. (2019), who also revealed that hippocampal volume affects FAQ more significantly in AD state than in CN state. Moreover, Feng et al. (2020) identified a positive effect of education level on cognitive impairment based on the ADNI dataset. This finding explains why the magnitude of the interactive effect of hippocampal volume and education level, as shown in Fig. 5, exhibits a decreasing trend in all the states. Such a decreasing trend implies that a high education level can eliminate the negative impact of hippocampal volume on cognitive impairment to a certain extent. However, neither the additive HMMs proposed by Kang et al. (2019) nor the linear model considered by Feng et al. (2020) cannot obtain such an interesting finding.

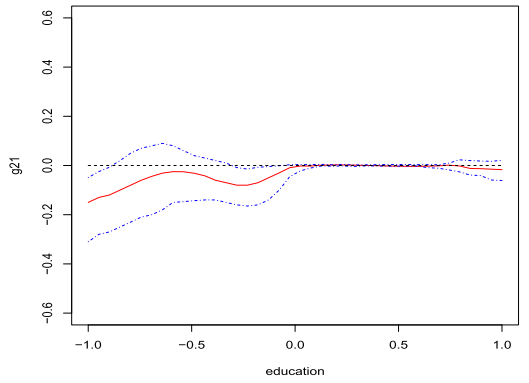
Fourth, in CN and MCI states, the effect of age on FAQ exhibits apparent zero-effect regions for participants whose education years exceed 12 (high school graduate). However, in AD state, such age effect becomes significantly positive regardless of education years, implying that older patients are likely to suffer from severer cognitive impairment than younger patients no matter their education level. This result also aligns with the finding of Kang et al. (2019), who reported that age affects cognitive function mainly in AD state.

In summary, this data analysis reveals that APOE-$\epsilon 4$ carriers have high risk of transitioning from a light to a severe cognitive impairment state. Moreover, hippocampal volume and age hardly influence cognitive impairment for highly educated participants in early to middle cognitive impairment stages (CN and MCI states). However, in a late progression stage (AD state), they become prominent dementia risk factors regardless of the length of education of the patients.

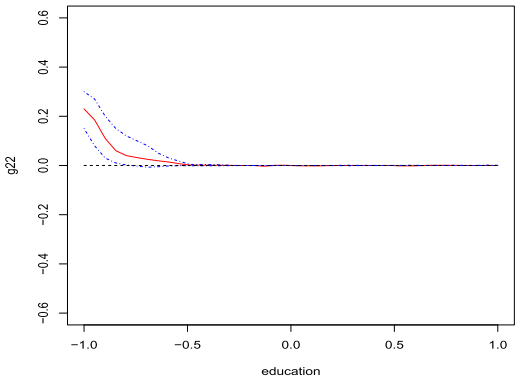


(a) Hippocampus(CN)

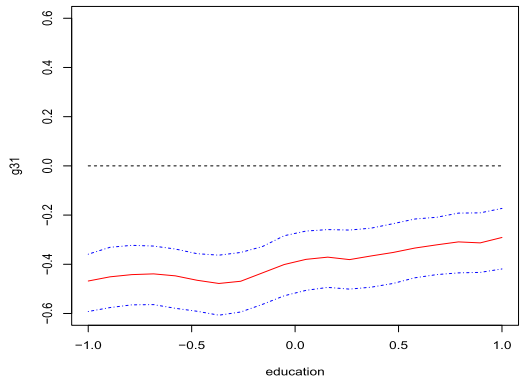
(b) Age(CN)



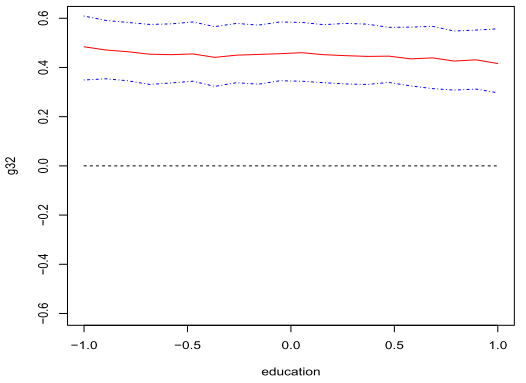
(c) Hippocampus(MCI)



(d) Age(MCI)



(e) Hippocampus(AD)



(f) Age(AD)

Fig. 5. Coefficient function estimate results of ADNI-1 data analysis. The red solid line is the mean estimated curve, the blue dotted lines are the 2.5 and 97.5 quantile estimated curve, and $y = 0$ is denoted in each picture by a black dotted line to show zero-effect.

6. Discussion

We develop a heterogeneous varying-coefficient HMM with zero-effect regions to analyze multivariate heterogeneous longitudinal data. The proposed model consists of a continuation-ratio logit transition model for examining the influence of potential covariates on the odds of transitioning from one hidden state to another and a conditional varying-coefficient regression model for investigating the functional covariate effects on the response of interest. Moreover, we introduce a soft-thresholding operator to detect the zero-effect regions of the state-specific coefficient functions. To the best of our knowledge, this study is the first to consider varying-coefficient HMMs and the related zero-effect region detection.

The proposed method can be extended in several directions. The current model assumes that random errors are normally distribution. This normality assumption may be violated in practice. Although our simulation study shows that the estimation results are relatively robust to several nonnormal random errors, relaxing the normality assumption of the random

error through sophisticated techniques, such as the Dirichlet process as a prior (Matthew and Alan, 2013), is greatly interesting and worthy of further investigation. In addition, variable selection is an important inference beyond estimation. The commonly used Bayesian variable selection method, such as reversible jump MCMC algorithm (Green, 1995), Bayesian Lasso (Park and Casella, 2008), and spike-and-slab prior procedure (Pan et al., 2021), can be developed to determine the number of hidden states and select important variables in the context of the proposed model. Missing data are frequently encountered in longitudinal data analysis. Accommodation of nonignorable missingness in longitudinal response and covariates is an important topic and requires substantial efforts in the future. Moreover, the functions of univariate covariates, $g_{sj}(\cdot)$s, in the proposed model can be extended to the functions of multivariate covariates. In this case, we can use tensor product B-splines (e.g., Song et al., 2014) to estimate the function of multivariate covariates. Finally, we assume that longitudinal observations are independent conditional on their hidden states. Thus, the dependence between the observations from a subject is explained by hidden states and their transition. This assumption can be relaxed by adding random effects into conditional and/or transition models.

Acknowledgement

The authors would like to thank two anonymous reviewers, an associate editor and the editor for constructive comments and helpful suggestions. The work is supported by National Natural Science Foundation of China grant (11671268) and GRF grant (14302519) from the Research Grant Council of Hong Kong Special Administration Region.

Appendix A

(1) Full conditional distribution of $(p_1, p_2, \cdots, p_S)$:

$$p(p_1, p_2, \cdots, p_S|\cdot) \propto \prod_{s=1}^{S} p_s^{a_0-1} \prod_{s=1}^{S} p_s^{n_{s1}} = \prod_{s=1}^{S} p_s^{a_0+n_{s1}-1},$$

where $n_{s1} = \sum_{i=1}^{N} I(Z_{i1} = s)$. Thus, the full conditional distribution of $(p_1, p_2, \cdots, p_S)$ is

$$(p_1, p_2, \cdots, p_S|\cdot) \sim Dirichlet(a_0 + n_{11}, a_0 + n_{21}, \cdots, a_0 + n_{S1})$$

(2) Full conditional distribution of φ_{us}:

$$\begin{aligned} p(\varphi_{us}|\cdot) &\propto \exp\left\{ -\frac{(\varphi_{us} - \varphi_{us0})^2}{2\sigma_{us0}^2} \right\} \exp\left\{ \log \prod_{v=s}^{S} \prod_{i=1}^{N} \prod_{t=2}^{T} p_{ituv}^{I(z_{i,t-1}=u, z_{it}=v)} \right\} \\ &= \exp\left\{ \sum_{v=s}^{S} \sum_{i=1}^{N} \sum_{t=2}^{T} \log p_{ituv}^{I(z_{i,t-1}=u, z_{it}=v)} - \frac{(\varphi_{us} - \varphi_{us0})^2}{2\sigma_{us0}^2} \right\}. \end{aligned}$$

(3) Full conditional distribution of $\boldsymbol{\alpha}$:

$$\begin{aligned} p(\boldsymbol{\alpha}|\cdot) &\propto \exp\left\{ -\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \right\} \exp\left\{ \log \prod_{u=1}^{S} \prod_{s=1}^{S} \prod_{i=1}^{N} \prod_{t=2}^{T} p_{itus}^{I(z_{i,t-1}=u, z_{it}=s)} \right\} \\ &= \exp\left\{ \sum_{u=1}^{S} \sum_{s=1}^{S} \sum_{i=1}^{N} \sum_{t=2}^{T} \log p_{itus}^{I(z_{i,t-1}=u, z_{it}=s)} - \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \right\}. \end{aligned}$$

(4) Full conditional distribution of σ_s^2:

$$\begin{aligned} p(\sigma_s^2|\cdot) &\propto (\sigma_s^2)^{-\xi_{s0}-1} e^{-\frac{\tau_{s0}}{\sigma_s^2}} \sigma_s^{-n_s} \\ &\quad \times \exp\left\{ -\frac{1}{2\sigma_s^2} \sum_{i=1}^{N} \sum_{t=1}^{T} I(z_{it} = s) \Big[y_{it} - b_s - \sum_{j=1}^{r} \zeta_{\lambda_{sj}} \Big(\sum_{l=1}^{L} \beta_{sjl} B_l(w_{itj}) \Big) x_{itj} \Big]^2 \right\} \\ &\propto (\sigma_s^2)^{-\xi_{s0}-\frac{n_s}{2}-1} \exp\left\{ -\frac{\tau_{s0} + \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} I(z_{it} = s) \Big[y_{it} - b_s - \sum_{j=1}^{r} \zeta_{\lambda_{sj}} \Big(\sum_{l=1}^{L} \beta_{sjl} B_l(w_{itj}) \Big) x_{itj} \Big]^2}{\sigma_s^2} \right\}, \end{aligned}$$

where $n_s = \sum_{i=1}^{N} \sum_{t=1}^{T} I(z_{it} = s)$. Thus, the full conditional distribution of σ_s^2 is

$$\sigma_s^2 \sim IG\Big(\xi_{s0} + \frac{n_s}{2}, \tau_{s0} + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T} I(z_{it}=s)\Big[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\big)x_{itj}\Big]^2\Big).$$

(5) Full conditional distribution of b_s:

$$p(b_s|\cdot) \propto \exp\Big\{\frac{(b_s-\mu_{s0})^2}{2\sigma_{s0}^2}\Big\} \times \prod_{i=1}^{N}\prod_{t=1}^{T}\Big[\frac{1}{\sqrt{2\pi}\sigma_s}\exp\Big\{-\frac{[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\big)x_{itj}]^2}{2\sigma_s^2}\Big\}\Big]^{I(z_{it}=s)}.$$

Let $\mathbf{Y}_s = \Big\{\big[y_{it} - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\big)x_{itj}\big]I(z_{it}=s) : i=1,\cdots,N; t=1,\cdots,T\Big\} \triangleq \{Y_{s1},\cdots,Y_{sn_s}\}$, where $n_s = \sum_{i=1}^{N}\sum_{t=1}^{T} I(z_{it}=s)$ and $\sum_{s=1}^{S} n_s = N\times T$. Then, $p(b_s|\cdot)$ can be rewritten as follows:

$$\begin{aligned}
p(b_s|\cdot) &\propto \exp\Big\{-\frac{(b_s-\mu_{s0})^2}{2\sigma_{s0}^2}\Big\}\exp\Big\{-\frac{\sum_{i=1}^{n_s}(Y_{si}-b_s)^2}{2\sigma_s^2}\Big\}\\
&= \exp\Big\{-\frac{1}{2}\Big[\frac{b_s^2 - 2\mu_{s0}+\mu_{s0}^2}{\sigma_{s0}^2} + \frac{b_s^2 - 2Y_{s1}b_s + Y_{s1}^2 + \cdots + b_s^2 - 2Y_{sn_s}b_s + Y_{sn_s}^2}{\sigma_s^2}\Big]\Big\}\\
&= \exp\Big\{-\frac{1}{2}\Big[\frac{(n_s\sigma_{s0}^2+\sigma_s^2)b_s^2 - (2n_s\bar{Y}_s\sigma_{s0}^2 + 2\sigma_s^2\mu_{s0})b_s + \sigma_{s0}^2\sum_{i=1}^{n_s}Y_{si}^2 + \sigma_s^2\mu_{s0}^2}{\sigma_{s0}^2\sigma_s^2}\Big]\Big\}\\
&= \exp\Big\{-\frac{1}{2}\Big[\frac{b_s^2 - \frac{2n_s\bar{Y}_s\sigma_{s0}^2+2\sigma_s^2\mu_{s0}}{n_s\sigma_{s0}^2+\sigma_s^2}b_s + \frac{\sigma_{s0}^2\sum_{i=1}^{n_s}Y_{si}^2+\sigma_s^2\mu_{s0}^2}{n_s\sigma_{s0}^2+\sigma_s^2}}{\frac{\sigma_{s0}^2\sigma_s^2}{n_s\sigma_{s0}^2+\sigma_s^2}}\Big]\Big\},
\end{aligned}$$

where $\bar{Y}_s = \frac{1}{n_s}\sum_{i=1}^{n_s} Y_{si}$. Thus, the full conditional distribution of b_s is as follows:

$$b_s \sim N\Big(\frac{2n_s\bar{Y}_s\sigma_{s0}^2 + 2\sigma_s^2\mu_{s0}}{n_s\sigma_{s0}^2+\sigma_s^2}, \frac{\sigma_{s0}^2\sigma_s^2}{n_s\sigma_{s0}^2+\sigma_s^2}\Big).$$

(6) Full conditional distribution of λ_{sj}:

$$p(\lambda_{sj}|\cdot) \propto I(0<\lambda_{sj}<d_{sj0})\prod_{i=1}^{N}\prod_{t=1}^{T}\Big[\exp\Big\{-\frac{\big[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\big)x_{itj}\big]^2}{2\sigma_s^2}\Big\}\Big]^{I(z_{it}=s)}.$$

(7) Full conditional distribution of $\boldsymbol{\beta}_{sj}$:

$$\begin{aligned}
p(\boldsymbol{\beta}_{sj}|\cdot) &\propto \exp\Big\{-\frac{1}{2n_s^2}\boldsymbol{\beta}_{sj}\boldsymbol{D}_{sj}\boldsymbol{\beta}_{sj}\Big\}\\
&\times \prod_{i=1}^{N}\prod_{t=1}^{T}\Big[\exp\Big\{-\frac{\big[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\big)x_{itj}\big]^2}{2\sigma_s^2}\Big\}\Big]^{I(z_{it}=s)}\\
&\propto \exp\Big\{-\frac{1}{2n_s^2}\boldsymbol{\beta}_{sj}\boldsymbol{D}_{sj}\boldsymbol{\beta}_{sj} - \frac{\sum_{z_{it}=s}\big[y_{it} - b_s - \sum_{j=1}^{r}\zeta_{\lambda_{sj}}\big(\sum_{l=1}^{L}\beta_{sjl}B_l(w_{itj})\big)x_{itj}\big]^2}{2\sigma_s^2}\Big\}.
\end{aligned}$$

(8) Full conditional distribution of z_{it}:

We adopt the recursive forward filtering backward sampling (FFBS) algorithm to sample z_{it} from $p(z_{it}|\cdot)$. The FFBS algorithm is implemented as follows:

$$p(z_{it}|\cdot) \propto p(\mathbf{y}_i, z_{it}|\boldsymbol{\theta}) = p(\mathbf{y}_{i,1:t}, z_{it}|\boldsymbol{\theta})p(\mathbf{y}_{i,t+1:T}|z_{it},\boldsymbol{\theta}) \doteq q_{it}(\mathbf{y}_i, z_{it}|\boldsymbol{\theta})\bar{q}_{it}(\mathbf{y}_i|z_{it},\boldsymbol{\theta}).$$

Moreover, we initialize

$$q_{i1}(\mathbf{y}_i, z_{i1}|\boldsymbol{\theta}) = p(y_{i1}, z_{i1}|\boldsymbol{\theta}) = p(y_{i1}|z_{i1},\boldsymbol{\theta})p(z_{i1}|\boldsymbol{\theta}),$$

and calculate $q_{it}(\mathbf{y}_i, z_{it}|\boldsymbol{\theta})$ for $t = 2, \cdots, T$ in a recursion manner as follows:

$$q_{it}(\mathbf{y}_i, z_{it}|\boldsymbol{\theta}) = p(\mathbf{y}_{i,1:t}, z_{it}|\boldsymbol{\theta}) = \sum_{u=1}^{S} p(\mathbf{y}_{i,1:t}, z_{it}, z_{i,t-1} = u|\boldsymbol{\theta})$$
$$= \sum_{u=1}^{S} p(\mathbf{y}_{i,1:t-1}, z_{i,t-1} = u|\boldsymbol{\theta}) p(z_{it}|z_{i,t-1} = u, \boldsymbol{\theta}) p(y_{it}|z_{it}, \boldsymbol{\theta})$$
$$= \sum_{u=1}^{S} q_{i,t-1}(\mathbf{Y}_i, z_{i,t-1} = u|\boldsymbol{\theta}) p(z_{it}|z_{i,t-1} = u, \boldsymbol{\theta}) p(y_{it}|z_{it}, \boldsymbol{\theta}),$$

where $p(z_{it}|z_{i,t-1} = u, \boldsymbol{\theta})$ is the transition probability.

We also initialize $\bar{q}_{iT}(\mathbf{y}_i|z_{it}, \boldsymbol{\theta}) = 1$ and calculate $\bar{q}_{it}(\mathbf{y}_i|z_{it}, \boldsymbol{\theta})$ for $t = T-1, \cdots, 1$ as

$$\bar{q}_{it}(\mathbf{y}_i|z_{it}, \boldsymbol{\theta}) = p(\mathbf{y}_{i,t+1:T}|z_{it}, \boldsymbol{\theta}) = \sum_{u=1}^{S} p(\mathbf{y}_{i,t+1:T}, z_{i,t+1} = u|z_{it}, \boldsymbol{\theta})$$
$$= \sum_{u=1}^{S} \{p(\mathbf{y}_{i,t+2:T})|z_{i,t+1} = u, \boldsymbol{\theta}) p(z_{i,t+1} = u|z_{it}, \boldsymbol{\theta}) p(y_{i,t+1}|z_{i,t+1} = u, \boldsymbol{\theta})\}$$
$$= \sum_{u=1}^{S} \{\bar{q}_{i,t+1}(\mathbf{y}_i|z_{i,t+1} = u, \boldsymbol{\theta}) p(z_{i,t+1} = u|z_{it}, \boldsymbol{\theta}) p(y_{i,t+1}|z_{i,t+1} = u, \boldsymbol{\theta})\},$$

where $p(z_{i,t+1} = u|z_{it}, \boldsymbol{\theta})$ is the transition probability and $p(y_{i,t+1}|z_{i,t+1} = u, \boldsymbol{\theta})$ is the likelihood in hidden state $z_{i,t+1} = u$.

References

Altman, R.M., 2007. Mixed hidden Markov models. J. Am. Stat. Assoc. 102, 201–210.

Bartolucci, F., Farcomeni, A., 2009. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. J. Am. Stat. Assoc. 104, 816–831.

Berry, S.M., Carroll, R.J., Ruppert, D., 2002. Bayesian smoothing and regression splines for measurement error problem. J. Am. Stat. Assoc. 97, 160–169.

Biller, C., Fahrmeir, L., 2001. Bayesian varying-coefficient models using adaptive regression splines. Stat. Model. 1, 195–211.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer, New York.

Celeux, G., Forbes, F., Robert, C.P., Titterington, D.M., 2006. Deviance information criteria for missing data models. Bayesian Anal. 1, 651–674.

Chiang, C.T., Rice, J.A., Wu, C.O., 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. J. Am. Stat. Assoc. 96, 605–619.

Dickerson, B.C., Wolk, D., 2013. Biomarker-based prediction of progression in MCI: comparison of AD-signature and hippocampal volume with spinal fluid amyloid-β and tau. Front. Aging Neurosci. 5, 55.

Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455.

Dunson, D., Chulada, P., Arbes, J.S.J., 2003. Bayesian modeling of time-varying and waning exposure effects. Biometrics 59, 83–91.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Stat. Sci. 11, 89–121.

Eubank, R.L., Huang, C., Maldonado, Y.M., Wang, N., Wang, S., Buchanan, R.J., 2004. Smoothing spline estimation in varying-coefficient models. J. R. Stat. Soc. B 66, 653–667.

Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. Ann. Stat. 27, 1491–1518.

Feng, X.N., Li, T.F., Song, X.Y., Zhu, H.T., 2020. Bayesian scalar on image regression with non-ignorable non-response. J. Am. Stat. Assoc. 115, 1574–1597.

Frühwirth-Schnatter, S., 2001. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. J. Am. Stat. Assoc. 96, 194–208.

Gelman, A., 1996. Inference and monitoring convergence. In: Gilks, W.R., et al. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, pp. 131–144.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Gupta, M., Qu, P.P., Ibrahim, J.G., 2007. A temporal hidden Markov regression model for analysis of gene regulatory networks. Biostatistics 4, 805–820.

Haneuse, S.J., Rudser, K.D., Gillen, D.L., 2008. The separation of timescales in Bayesian survival modeling of the time-varying effect of a time-dependent exposure. Biostatistics 9, 400–410.

Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. B 55, 757–796.

Hoover, D.R., Rice, J.A., Wu, C.O., Yang, L.P., 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85, 809–822.

Huang, J.Z., Wu, C.O., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89, 111–128.

Huang, J.Z., Wu, C.O., Zhou, L., 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Stat. Sin. 14, 763–788.

Ip, E., Zhang, Q., Rejeski, J., Harris, T., Kritchevsky, S., 2013. Partially ordered mixed hidden Markov model for the disablement process of older adults. J. Am. Stat. Assoc. 108, 370–384.

Jack, C.R., Petersen, R.C., Brien, P.C., 1992. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42, 183–188.

Kang, J., Reich, B.J., Staicu, A.M., 2018. Scalar-on-image regression via the soft-thresholded Gaussian process. Biometrika 105, 165–184.

Kang, K., Cai, J.H., Song, X.Y., Zhu, H.T., 2019. Bayesian hidden Markov models for delineating the pathology of Alzheimer's disease. Stat. Methods Med. Res. 28, 2112–2124.

Kantarci, K., Gunter, J.L., Tosakulwong, N., 2013. Focal hemosiderin deposits and β-amyloid load in the ADNI cohort. Alzheimer's Dement. 9, S116–S123.

Kesslak, J.P., Nalcioglu, O., Cotman, C.W., 1991. Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. Neurology 41, 51–54.

Lang, S., Brezger, A., 2004. Bayesian P-splines. J. Comput. Graph. Stat. 13, 183–212.

Lee, E., Zhu, H., Kong, D., Wang, Y., Giovanello, K.S., Ibrahim, J.G., 2015. BFLCRM: a Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. Ann. Appl. Stat. 9, 2153–2178.
Liu, H.F., Song, X.Y., Tang, Y.L., Zhang, B.X., 2021. Bayesian quantile nonhomogenous hidden Markov models. Stat. Methods Med. Res. 30, 112–128.
Lu, Z.H., Song, X.Y., 2012. Finite mixture varying coefficient models for analyzing longitudinal heterogenous data. Stat. Med. 31, 544–560.
Ma, S., Song, P.X.K., 2015. Varying index coefficient models. J. Am. Stat. Assoc. 110, 341–356.
Marino, M.F., Tzavidis, N., Alfo, M., 2018. Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequence. Stat. Methods Med. Res. 27, 2231–2246.
Matthew, J.J., Alan, S.W., 2013. Bayesian nonparametric hidden semi-Markov models. J. Mach. Learn. Res. 14, 673–701.
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-ϵ4 and early onset Alzheimer's. Nat. Genet. 7, 10–11.
Pan, D., Wei, Y.Y., Song, X.Y., 2021. Joint analysis of mixed types of outcomes with latent variables. Stat. Med. 40, 1272–1284.
Park, T., Casella, G., 2008. The Bayesian lasso. J. Am. Stat. Assoc. 103, 681–686.
Robert, C.P., Ryden, T., Titterington, D.M., 2000. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. J. R. Stat. Soc. B 62, 57–75.
Ressell, D., Telesca, D., 2017. Non-local priors for high-dimension estimation. J. Am. Stat. Assoc. 112, 254–265.
Scott, S.L., James, G.M., Sugar, C.A., 2005. Hidden Markov models for longitudinal comparisons. J. Am. Stat. Assoc. 100, 359–369.
Song, X.Y., Lu, Z.H., Feng, X.N., 2014. Latent variable models with nonparametric interaction effects of latent variables. Stat. Med. 33, 1723–1737.
Song, X.Y., Xia, Y.M., Zhu, H.T., 2017. Hidden Markov latent variable models with multivariate longitudinal data. Biometrics 73, 313–323.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc., Ser. B, Stat. Methodol. 58, 267–288.
Wu, C.O., Yu, K.F., Chiang, C.T., 2000. A two-step smoothing method for varying coefficient models with repeated measurement. Ann. Inst. Stat. Math. 52, 519–543.
Yang, Y., 2020. Novel Methods for Estimation and Inference in Varying Coefficient Models. PhD Dissertation. The University of Michigan.
Ye, M., Lu, Z.H., Li, Y.M., Song, X.Y., 2019. Finite mixture of varying coefficient model: estimation and component selection. J. Multivar. Anal. 171, 452–474.
Yen, M.F., Chen, H.H., 2018. Bayesian measurement-error-driven hidden Markov regression model for calibrating the effect of covariates on multistate outcomes: application to androgetic plopecia. Stat. Med. 21, 3125–3146.
Zhang, W.Y., Lee, S.Y., Song, X.Y., 2002. Local polynomial fitting in semivarying coefficient model. J. Multivar. Anal. 82, 166–188.
Zhou, X.X., Kang, K., Song, X.Y., 2020. Two-part hidden Markov models for semicontinuous longitudinal data with nonignorable missing covariates. Stat. Med. 39, 1801–1816.