Order Selection for Regression-based Hidden Markov Model(RHMM)

LIN, Yi-Qi & SONG, Xin-Yuan

Department of Statistics
The Chinese University of Hong Kong

June 26, 2021

Outline

1. Order Selection for Regression-based Hidden Markov Model(RHMM)
   - Motivation
   - Proposed Model
     - Regression Based Hidden Markov Model
     - Estimating the order
   - Order Selection via Extended-Group-Sort-Fuse (EGSF) Procedure
   - Double Penalization
     - Asymptotic Result
   - Estimation Via ECM-ITD memthod
     - ECM-ITD algorithm
   - Simulation
   - Application
     - ADNI data analysis
     - Result
   - Discussion

Motivation

- Hidden Markov model (HMM) is a practical statistical tool to simultaneously analysing the longitudinal observation process and its dynamic transition process.
- the most existing HMMs and its extensions in the literature suffer from the determined number of states (order of HMM) that usually are unknown in applications.
- a data-driven procedure to choose the number of hidden states still remains a challenge problem.
- The most common method in the literature for model selection is information based criterion, such as AIC (Akaike (1974)) and BIC (Schwarz et al. (1978)).
- Even though this prevailing information-based criterion succeed in applications (see e.g. Song et al. (2017) and Ip et al. (2013)), it still lack of theoretical justi cation for HMMs and its extensions (MacDonald and Zucchini (1997)).

Common Settings

$\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_n)$, where $\boldsymbol{Y}_i = \{y_{it}\}_{t=1}^{T}$ and y_{it} be the response of subject i at time t; $\boldsymbol{S} = (\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_n)$, where $\boldsymbol{S}_i = \{S_{it}\}_{t=1}^{T}$ is a set of hidden states associated with y_{it}, and S_{it} is assumed to be a finite-state stationary Markov chain taking values in $\{1, \cdots, K\}$.

1. transition between different states can be described by a homogeneous transition matrix $P = [P_{rs}]_{K\times K}$ with $P_{rs} = P(S_{it} = s|S_{i,t-1} = r)$ for $\forall i$ and $t = 2, \ldots, T$ and stationary probability π_r, where $r, s \in \{1, \cdots, K\}$.
2. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)$ be the set of covariates, where $\boldsymbol{X}_i = \{\boldsymbol{x}_{it}\}_{t=1}^{T}$ and $\boldsymbol{x}_{it}$ is a $(q+1)\times 1$ vector of covariates for subject i at time t.
3. Conditional on hidden state $S_{it} = k$ and covariates $\boldsymbol{x}_{it}$, a generalized linear model for response y_{it} is considered as

$$\begin{aligned} f(y_{it}|S_{it} = k, \boldsymbol{x}_{it}, \boldsymbol{\beta}_k) &= \exp\left\{\left(y_{it}\theta_{itk} - b\left(\theta_{itk}\right)\right)/a(\phi) + c\left(y_{it}, \phi\right)\right\}, \\ \theta_{itk} &= \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_k, \end{aligned} \tag{1}$$

4. For ease of exposition, K and K_0 are denoted as the upper bound and true value of the order, respectively. Let $\Psi = (\pi_1, \pi_2, \cdots, \pi_K; P_{11}, \cdots, P_{KK}; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_K)$. Then, the probability mass/density function of $\boldsymbol{Y}_i$ can be written as

$$F(\boldsymbol{Y}_i; \boldsymbol{X}_i, \Psi) = \sum_{S_{i1}=1}^{K} \cdots \sum_{S_{iT}=1}^{K} \left[\prod_{t=1}^{T} \left[f\left(y_{it}; \boldsymbol{x}_{it}, \boldsymbol{\beta}_{S_{it}}\right)\right] \pi_{S_{i1}} P_{S_{i1}S_{i2}} \cdots P_{S_{i,T-1}S_{iT}}\right]. \tag{2}$$

Order Estimation

A natural way to estimate the order of RHMM is the maximum likelihood estimate (MLE) of overfitted log-likelihood with the upper bound of order K ($K \geq K_0$):

$$l_n(\Psi) = \sum_{i=1}^{n} \log F(\boldsymbol{Y}_i; \boldsymbol{X}_i, \Psi)\,. \tag{3}$$

However, the overfitted MLE leads to an inconsistent estimate of K_0 (Chen and Khalili, 2009; Hung et al., 2013; Manole and Khalili, 2020). The overfitting of MLE is of two types:

Overfitting Types

1. near-zero values of mixing probability.
2. densities of some components are close to each other.

Penalty

The double penalized log-likelihood can be written as follows:

$$\tilde{l}_n(\Psi) = l_n(\Psi) + C_K \sum_{k=1}^{K} \log \pi_k - n \sum_{k=2}^{K} p_{\lambda_n}\left(\|\boldsymbol{\eta}_{\boldsymbol{k}}\|_2\right), \tag{4}$$

where C_K is a tuning parameter, $p_{\lambda_n}(\cdot)$ is a penalty function, and $\|\boldsymbol{\eta}_k\|_2 = \|\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k-1}\|_2$, which will be clarified in the subsequent section. For simplicity, we only consider the SCAD penalty for our model, and it is not essentially hard to implement the MCP and Adaptive LASSO penalties for this setting.

- we proposed a Group-Sort-Fuse procedure to sort the multidimensional parameters of the finite mixture model.

$$\boldsymbol{\beta}_{(k)} = \underset{j \notin \left\{\boldsymbol{\beta}_{(i)}: 1 \le i \le k-1\right\}}{\operatorname{argmin}} \left\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{(k-1)}\right\|_2, \quad k = 2, 3, \ldots, K, \tag{5}$$

  and $\boldsymbol{\beta}_{(1)} = \underset{k=1,2,\ldots,K}{\operatorname{argmax}} \|\boldsymbol{\beta}_k\|_2$.

- $\hat{\Psi}_n = \operatorname{argmax} \tilde{l}_n(\Psi)$ as the MPLE of Ψ. Then,

$$\hat{K}_n = \text{number of distinct values of } \left\{\hat{\boldsymbol{\beta}}_{(k)}, k = 1, \cdots, K\right\} \tag{6}$$

  is an estimator of true order K_0, and we show that $\hat{K}_n$ converges to K_0 in probability in the subsequent section.

Asymptotic Result

Theorem 1

Suppose that RHMM is identifiable and $F(\boldsymbol{Y}; \boldsymbol{X}, \Psi)$ satisfies the mild regular conditions stated in Appendix A. If $\lambda_n = cn^{-\frac{1}{4}} \log n$ for SCAD penalty and some $c > 0$. Then, we have the following:

(1) For any continuous point of $\boldsymbol{\beta}^S$ of G_0, we have $\hat{G}_n(\boldsymbol{\beta}^S) \xrightarrow{p} G_0(\boldsymbol{\beta}^S)$.

(2) $\sum_{k=1}^{K} \log \hat{\pi}_k = O_p(1)$ and $\hat{\alpha}_k = \pi_{0k} + o_p(1)$ for all $k = 1, 2, \ldots, K_0$. Furthermore, for each $l = 1, 2, \ldots, K$, a unique $k = 1, 2, \ldots, K_0$ exists, such that $\|\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_{0k}\|_2 = o_p(1)$. Thus, $\left\{\hat{\mathcal{V}}_k : k = 1, 2, \ldots, K_0\right\}$ is a cluster partition of $\left\{\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots, \hat{\boldsymbol{\beta}}_K\right\}$ in probability.

Theorem 2

We assume that the same conditions in Theorem 1 hold. Under the true dynamic finite mixture density $F(\boldsymbol{Y}; G_0)$, if $\hat{G}_n$ falls into an $O(n^{-\frac{1}{4}})$ neighborhood of G_0, then $P\left(\hat{K}_n = K_0\right) \to 1$ as $n \to \infty$.

ECM

In the presence of hidden states, the expectation–maximization (EM) algorithm together with the Baum–Welch algorithm (Baum et al., 1970) is known as an efficient statistical estimation method to obtain the maximum likelihood estimate of Ψ. In this section, we propose an ECM–ITD algorithm to obtain the MPLE $\hat{\Psi}_n$ of Ψ in RHMM.

1. CM-step 1:

$$\pi_k^{(p+1)} = \frac{\sum_{i=1}^n h^p(S_{i,1} = k) + C_k}{n + KC_k}, \tag{7}$$

$$P_{r,s}^{(p+1)} = \frac{\sum_{i=1}^n \sum_{j=2}^T h^p(S_{i,j-1} = r, S_{i,j} = s)}{\sum_{i=1}^n \sum_{j=2}^T h^p(S_{i,j-1} = r)}. \tag{8}$$

2. CM-step 2:

$$\boldsymbol{\beta}^{(p+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{k=1}^K \left[\sum_{i=1}^n \sum_{t=1}^T \log f(y_{it}|S_{it} = k, \boldsymbol{x}_{it}, \boldsymbol{\beta}) h^{(p+1)}(S_{it} = k) \right] - n \sum_{k=2}^K p_{\lambda_n}(\|\boldsymbol{\eta_k}\|_2) \tag{9}$$

where $k, r, s = 1, 2, \ldots, K$ and $\boldsymbol{\eta_k} = \boldsymbol{\beta}_k - \boldsymbol{\beta}_{k-1}$.

ITD

In order to solve the optimization problem in CM-step 2, we develop an extended ITD algorithm in our multidimensional setting.

1. Impose a constraint $\boldsymbol{\eta}_1 = \boldsymbol{\beta}_1$ to form a one-to-one mapping between $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_K)$ and $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta}_k = \sum_{l=1}^{k} \boldsymbol{\eta}_l$ for $k = 1, 2, \ldots, K$.
2. we convert the optimization of updating $\boldsymbol{\beta}^{(p+1)}$ as

$$\boldsymbol{\eta}^{(p+1)} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \left\{ G(\boldsymbol{\eta}) = -\sum_{k=1}^{K} \varphi_k\Big(\sum_{l=1}^{k} \boldsymbol{\eta}_l\Big) + n \sum_{k=2}^{K} p_{\lambda_n}\left(\|\boldsymbol{\eta}_k\|_2\right) \right\}, \tag{10}$$

where $\varphi_k(\boldsymbol{\beta}_k) = \sum_{i=1}^{n} \sum_{t=1}^{T} \{y_{it}\theta_{itk} - b(\theta_{itk})\}\, h^{(p+1)}(S_{it} = k)$ and $\theta_{itk} = \mathbf{x}_{it}^T \boldsymbol{\beta}_k$.

3. Inspired by the prevailing iterative shrinkage-thresholding algorithm (ISTA) for regulated convex optimization problem, we optimize a surrogate function $\tilde{Q}(\boldsymbol{\xi}; \boldsymbol{\eta}^{(m)})$:

$$\begin{aligned} \tilde{Q}(\boldsymbol{\xi}; \boldsymbol{\eta}) = \rho G(\boldsymbol{\xi}) + \frac{1}{2} \sum_{j=1}^{K} \|\boldsymbol{\xi}_j - \boldsymbol{\eta}_j\|_2^2 \qquad (11) \\ -\rho \left[\sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ b\big(\mathbf{x}_{it}^T \boldsymbol{\zeta}_k\big) - b\big(\mathbf{x}_{it}^T \boldsymbol{\beta}_k\big) - b'\left(\mathbf{x}_{it}^T \boldsymbol{\beta}_k\right) \left[\mathbf{x}_{it}^T \left(\boldsymbol{\beta}_k - \boldsymbol{\beta}_k\right)\right] \right\} h_{itk} \right], \end{aligned}$$

ITD

CM-step 2 could be reformulated using multivariate thresholding operator $\vec{\boldsymbol{S}}(\cdot\,; 2n, a, \lambda_n)$ as

$$\boldsymbol{\eta}_1^{(m+1)} = \boldsymbol{\eta}_1^{(m)} + \rho \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{T} h_{ij,k}\left(y_{ij} - b'\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k^{(m)}\right)\right)\boldsymbol{x}_{ij} \tag{12}$$

$$\boldsymbol{\eta}_j^{(m+1)} = \vec{\boldsymbol{S}}\left(\boldsymbol{\eta}_j^{(m)} + \rho \sum_{k=j}^{K}\sum_{i=1}^{n}\sum_{j=1}^{T} h_{ij,k}\left(y_{ij} - b'\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k^{(m)}\right)\right)\boldsymbol{x}_{ij}; 2n\rho, a, \lambda_n\right), \tag{13}$$

for $j = 2, 3, \dots, K$, and then we continue iterating $\boldsymbol{\eta}^{(m)}$ as the above until it converges.

Theorem 3

Assume that sequence $\boldsymbol{\eta}^{(m)}$ is generated from (12) and $\boldsymbol{\beta}_k^{(m)} = \sum_{l=1}^{k}\boldsymbol{\eta}_l^{(m)}$. Let τ_1 be the maximum eigenvalue of $\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}^T$ and $\tau_2^{(m)}$ be assigned to

$$\tau_2^{(m)} = \max_{i,t,k}\sup_{0<\alpha<1} b''\left\{\boldsymbol{x}_{it}^T\left(\alpha\boldsymbol{\beta}^{(m+1)} + (1-\alpha)\boldsymbol{\beta}^{(m)}\right)\right\}. \tag{14}$$

If $\rho^{-1} \geq K\tau_2^{(m)}\tau_1$, then $G\left(\boldsymbol{\eta}^{(m+1)}\right) \leq G(\boldsymbol{\eta}^{(m)})$. Furthermore, if space $\{\boldsymbol{\eta} : G(\boldsymbol{\eta}) \leq G\left(\boldsymbol{\eta}^{(0)}\right)\}$ is compact, then sequences $\{\boldsymbol{\eta}^{(m)}\}$ and $\{\boldsymbol{\beta}^{(m)}\}$ converge to a stationary point of $G(\boldsymbol{\eta})$.

Simulation

Consider RHMMs with $K_0 = 2, 3, 4$. For each setting of K_0, y_{it} in state k is generated from a normal distribution with mean $\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_k$ and standard deviation $\sigma_k = 0.25$ Covariates $\boldsymbol{x}_{it} = (x_{it1}, x_{it2}, x_{it3})$, where $x_{it1} = 1$, and x_{it2} and x_{it3} are independently generated from $N(0, 1)$ and $U(0, 1)$, respectively, where $U(0, 1)$ stands for the uniform distribution on (0,1). Two sample sizes, $n = 50$ and 100 for normal and a transition matrix with elements $P_{rs} = \frac{1}{K_0}$, $r, s = 1, \cdots, K_0$ are considered. The state-specific regression coefficients for case (1) are set as follows:

- when $K_0 = 2$, $T = 4$, $\boldsymbol{\beta}_1 = (0, -0.5, 0.2)^T$, and $\boldsymbol{\beta}_2 = (0.5, 0, -0.2)^T$;
- when $K_0 = 3$, $T = 4$, $\boldsymbol{\beta}_1 = (0, 0.5, 0.2)^T$, $\boldsymbol{\beta}_2 = (0.5, 0.5, -0.2)^T$ and $\boldsymbol{\beta}_3 = (1, -0.5, 0.2)^T$;
- when $K_0 = 4$, $T = 6$, $\boldsymbol{\beta}_1 = (0, 1, 1.25)^T$, $\boldsymbol{\beta}_2 = (1, 2, 1)^T$, $\boldsymbol{\beta}_3 = (1.5, 1.25, 0.75)^T$, and $\boldsymbol{\beta}_4 = (2, 1, 1.5)^T$.

Simulation Result

Table: Proportion of order selection for Case (1) in Simulation 1

K_0	$\hat{K}_n$	$n=50$			$n=100$		
		AIC	BIC	ECM–ITD	AIC	BIC	ECM–ITD
2	**2**	**0.488**	**0.998**	**1**	**0.636**	**1**	**1**
	3	0.278	0	0	0.232	0	0
	4	0.234	0.002	0	0.132	0	0
3	2	0.010	**0.578**	0.344	0	0.072	0.026
	3	**0.408**	0.422	**0.654**	**0.562**	**0.924**	**0.974**
	4	0.376	0	0.002	0.304	0.004	0
	5	0.206	0	0	0.134	0	0
4	3	0.002	0.474	0.190	0	0.002	0.014
	4	**0.614**	**0.524**	**0.808**	**0.690**	**0.980**	**0.984**
	5	0.384	0.002	0.002	0.310	0	0.02

Simulation Result

To mimic the scenario of the ADNI study in real data analysis, we consider a larger RHMM with $K_0 = 5$, $T = 6$, six covariates, and two different transition matrices: (1) a general transition matrix $\boldsymbol{P}_1$ and (2) a band transition matrix $\boldsymbol{P}_2$ that only allows transitions between adjacent states.

$$\boldsymbol{P}_1 = \begin{pmatrix} 0.380 & 0.120 & 0.295 & 0.066 & 0.140 \\ 0.580 & 0.237 & 0.057 & 0.066 & 0.060 \\ 0.310 & 0.090 & 0.378 & 0.083 & 0.140 \\ 0.221 & 0.056 & 0.141 & 0.533 & 0.049 \\ 0.134 & 0.121 & 0.137 & 0.065 & 0.543 \end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0.7 & 0.2 & 0 \\ 0 & 0 & 0.4 & 0.4 & 0.2 \\ 0 & 0 & 0 & 0.7 & 0.3 \end{pmatrix}$$

- Here, y_{it} in state k is generated from a normal distribution with mean $\boldsymbol{x}_{it}^\top \boldsymbol{\beta}_k$ and standard deviation $\sigma_k = 0.25$.
- Covariates $\boldsymbol{x}_{it} = (x_{it1}, \cdots, x_{it6})^T$, where $x_{it1} = 1$, x_{it2} are independently generated from $N(0, 1)$, and x_{it3} to x_{it6} are independently generated from $U(0, 1)$.
- Three sample sizes, $n = 100$, 200, and 400, are considered.
- The state-specific regression coefficients are assigned as $\boldsymbol{\beta}_1 = (0, 0, 0, 0, 0, 0)^T$, $\boldsymbol{\beta}_2 = (-1.5, 2.25, -1, 0, 0.5, 0.75)^T$, $\boldsymbol{\beta}_3 = (0.25, 1.5, 0.75, 0.25, -0.5, -1)^T$, $\boldsymbol{\beta}_4 = (-0.25, 0.5, -2.5, 1.25, 0.75, 1.5)^T$, and $\boldsymbol{\beta}_5 = (-1, -1.5, -0.25, 1.75, -0.5, 2)^T$.

Simulation Result

Table: Proportion of order selection for normal case in Simulation 2

N	$\hat{K}_n$	$\boldsymbol{P}_1$ (general transition matrix)			$\boldsymbol{P}_2$ (band transition matrix)		
		AIC	BIC	ECM–ITD	AIC	BIC	ECM–ITD
100	4	0.015	0.025	0	0.025	0.03	0.005
	5	**0.400**	**0.515**	**0.740**	**0.445**	**0.540**	**0.720**
	6	0.320	0.320	0.240	0.270	0.330	0.250
	7	0.265	0.140	0.020	0.260	0.100	0.025
200	4	0.0.005	0.005	0	0	0	0
	5	**0.595**	**0.695**	**0.890**	**0.510**	**0.660**	**0.950**
	6	0.250	0.2350	0.110	0.280	0.305	0.045
	7	0.150	0.065	0	0.210	0.035	0.005
400	**5**	**0.675**	**0.785**	**0.940**	**0.635**	**0.720**	**0.975**
	6	0.230	0.190	0.060	0.225	0.225	0.025
	7	0.095	0.025	0	0.140	0.020	0

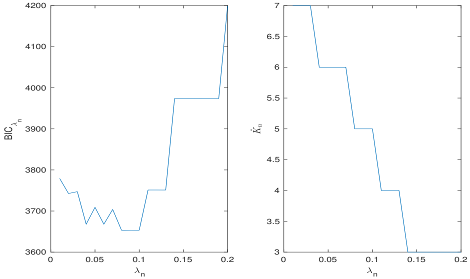
ADNI

we analyze the a data set extracted from ADNI study by employing the proposed ESGF Procedure to detect the number of hidden of phases of the neuro degenerative pathology.

- We focused on 616 subjects collected from the ADNI-I, ADNI-II, and ADNI-Go study with four follow-up visits at baseline, 6 months, 12 months, and 24 months.
- The Alzheimer's Disease Assessment Scale (ADAS) was devised to evaluate cognitive impairment in the assessment of AD. we treat 13-terms ADAS (ADAS.13) as the responses y_{it} in proposed RHMM.
- For each subject, the some clinical and genetic variables at four time points were introduced as covariates as listed: the clinical variables consist of $\boldsymbol{X}_{i2}$ gender (0 stands for male and 1 represents female) and $\boldsymbol{X}_{i3}$ age at baseline.
- The $\boldsymbol{X}_{i4}$ apolipoprotein E (APOE)-ε4 have been identified as high risk factor with respect to early onset of AD. Here we encode APOE-ε4 as 0,1 and 2 to represent the number of APOE-ε4 alleles.
- Likewise, hippocampal volume and atrophy of hippocampal formation were verified as a diagnostic biomarker of AD in literature. Thus $\boldsymbol{X}_{i5}$ the logarithm of the ratio of hippocampal volume over the whole brain volume were introduced into the RHMM.
- Hence, the covariates for each subject will be $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{i5})$, where $\boldsymbol{X}_{i1} = \mathbf{1}$ denoting the intercept.

Result

- We assumed that y_{it} in given state k followed a normal distribution with mean $\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_k$ and an unknown standard deviation σ_k. The continuous variables including y_{it}, x_{i2}, and x_{it4} were standardized prior to analysis.
- Based on the published reports in the AD literature, we set $K = 7$ as the upper bound for the number of hidden states to implement the proposed procedure.
- Figure 1 provides the details of determining the tuning parameter λ_n through BIC, where the left panel shows the plot of BIC values versus tuning parameter λ_n, and the right panel illustrates the estimated hidden states corresponding to tuning parameter λ_n. The minimum value of BIC was attained at $\lambda_n = 0.08$, and the corresponding estimated order $\hat{K}_n = 5$ was then selected.



Result

Table: Estimated coefficients (bootstrap variability estimates) for ADNI study

	State				
Par.	1 (CN)	2 (SMC)	3 (EMCI)	4 (LMCI)	5(AD)
β_{k1}	-1.007(0.023)	-0.628(0.031)	-0.033(0.029)	0.876(0.035)	2.225(0.167)
β_{k2}	0.092(0.017)	0.096(0.019)	0.069(0.023)	0.076(0.027)	0.034(0.068)
β_{k3}	0.037(0.022)	0.111(0.026)	0.023(0.036)	-0.066(0.052)	-0.464(0.168)
β_{k4}	-0.146(0.013)	-0.257(0.021)	-0.378(0.017)	-0.430(0.025)	-0.347(0.105)
β_{k5}	0.057(0.036)	0.203(0.042)	0.192(0.037)	0.089(0.046)	0.519(0.215)
β_{k6}	0.278(0.214)	0.445(0.236)	0.407(0.145)	0.184(0.214)	0.240(0.409)
σ_k	0.227(0.010)	0.266(0.009)	0.315(0.011)	0.377(0.015)	0.770(0.049)

Result

- state-specific intercept β_{k1} exhibits an ascending trend; patients had the lowest ADAS13 score in state 1, and the highest score in state 5.
- As ADAS13 measures cognitive impairment with a high score indicating low cognitive ability, states 1 to 5 can be explained as CN, significant memory concern (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD accordingly.
- This classification has been reported in the public literature and ADNI study from ADNI-II to the latest phase (Jessen et al., 2014)
- Some existing studies identify four (CN, EMCI, LMCI, AD) instead of five states. The ADNI-II study suggests that SMC is highly relevant to the AD progression and introducing an additional SMC state minimizes the stratification of cognitive ability and fills the gap between CN and EMCI. Published reports also argued that the introduction of SMC could address the vague demarcation between CN and EMCI (Risacher et al., 2015).

Discussion

The present work has limitations.

1. we assume that the transition is homogeneous. This assumption may be restrictive in practice because between-state transitions are frequently influenced by certain covariates, thereby leading to heterogenous transitions.
2. A continuation-ratio logit model (Agresti, 2003; Song et al., 2017) can be considered to model heterogeneous transitions.

Future investigation direction:

1. In many substantive studies, selection of potential predictors is also interesting, especially when the dimension of predictors is high. Thus, the proposed approach may be generalized to include additional penalties on regression parameters to simultaneously conduct order and variable selection.
2. The conditional regression model in the proposed RHMM only accommodates scalar predictors. Extending our method to incorporate functional or image predictors can considerably improve model flexibility and utility.

Reference

Agresti, A. (2003). *Categorical data analysis*, volume 482. John Wiley & Sons.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171.

Chen, J. and Khalili, A. (2009). Order selection in finite mixture models with a nonsmooth penalty. *Journal of the American Statistical Association*, 104(485):187–196.

Hung, Y., Wang, Y., Zarnitsyna, V., Zhu, C., and Wu, C. J. (2013). Hidden markov models with applications in cell adhesion experiments. *Journal of the American Statistical Association*, 108(504):1469–1479.

Jessen, F., Wolfsgruber, S., Wiese, B., Bickel, H., Mösch, E., Kaduszkiewicz, H., Pentzek, M., Riedel-Heller, S. G., Luck, T., Fuchs, A., et al. (2014). Ad dementia risk in late mci, in early mci, and in subjective memory impairment. *Alzheimer's & Dementia*, 10(1):76–83.

Manole, T. and Khalili, A. (2020). Estimating the number of components in finite mixture models via the group-sort-fuse procedure. *arXiv preprint arXiv:2005.11641*.

Risacher, S. L., Kim, S., Nho, K., Foroud, T., Shen, L., Petersen, R. C., Jack Jr, C. R., Beckett, L. A., Aisen, P. S., Koeppe, R. A., et al. (2015). Apoe effect on alzheimer's disease biomarkers in older adults with significant memory concern. *Alzheimer's & Dementia*, 11(12):1417–1429.

Song, X., Xia, Y., and Zhu, H. (2017). Hidden markov latent variable models with multivariate longitudinal data. *Biometrics*, 73(1):313–323.