PSYCHOMETRIKA—VOL. 79, NO. 4, 621–646
OCTOBER 2014
DOI: 10.1007/S11336-013-9368-7

BAYESIAN INFERENCES OF LATENT CLASS MODELS WITH AN UNKNOWN NUMBER OF CLASSES

JIA-CHIUN PAN

NATIONAL CHUNG CHENG UNIVERSITY

GUAN-HUA HUANG

NATIONAL CHIAO TUNG UNIVERSITY

This paper focuses on analyzing data collected in situations where investigators use multiple discrete indicators as surrogates, for example, a set of questionnaires. A very flexible latent class model is used for analysis. We propose a Bayesian framework to perform the joint estimation of the number of latent classes and model parameters. The proposed approach applies the reversible jump Markov chain Monte Carlo to analyze finite mixtures of multivariate multinomial distributions. In the paper, we also develop a procedure for the unique labeling of the classes. We have carried out a detailed sensitivity analysis for various hyperparameter specifications, which leads us to make standard default recommendations for the choice of priors. The usefulness of the proposed method is demonstrated through computer simulations and a study on subtypes of schizophrenia using the Positive and Negative Syndrome Scale (PANSS).

Key words: categorical data, finite mixture model, label switching, reversible jump Markov chain Monte Carlo, sensitivity analysis, surrogate endpoint.

1. Introduction

Latent class analysis (LCA), originally described by Green (1951) and systematically developed by Lazarsfeld and Henry (1968), as well as by Goodman (1974), has been found useful for classifying subjects based on their responses to a set of categorical items. The basic model postulates an underlying categorical latent variable with, say, J levels; and measured items are assumed independent of one another within any category of the latent variable. Recently, several authors extended the LCA model to describe the effects of measured covariates on the underlying categorical latent variable (Dayton & Macready, 1988; Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997), or on measured item distributions within latent levels (Melton, Liang, & Pulver, 2005). This paper studies a more general latent class model proposed by Huang and Bandeen-Roche (2004), which incorporates covariate effects both on the latent variable and the measured items themselves (henceforth, the regression extension of latent class analysis (RLCA) model).

To reduce complexity and enhance interpretability, one usually fixes the number of levels or "classes" in a given latent class model and does the parameter estimation under the fixed number of classes. When prior knowledge does not mandate the number of classes, selecting the number of classes to fit becomes an analytic challenge. Standard practice is to base selection on either the Pearson χ^2 or the likelihood ratio goodness-of-fit test, and to fix J at the lowest number of classes that yields acceptable fit (Goodman, 1974; Formann, 1992). Instead of testing the goodness of fit of a specified model, we might use a criterion for selecting among different numbers of classes. The AIC (Akaike, 1987) and BIC (Schwarz, 1978) criteria, which trade off

Electronic Supplementary Material The online version of this article (doi:10.1007/s11336-013-9368-7) contains supplementary material, which is available to authorized users.

Requests for reprints should be sent to Guan-Hua Huang, Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30010, Taiwan. E-mail: ghuang@stat.nctu.edu.tw

the value of the likelihood at the maximum likelihood solution and the number of estimated parameters, are commonly used approaches.

One common feature of the above methods is that they all must fit the latent class model repeatedly under different numbers of classes. Due to the slow convergence of commonly used fitting methods (e.g., the EM algorithm), these procedures may be infeasible in practical applications. Huang (2005) developed a new tool for identifying the number of latent classes, which requires no model fit based on the assumed number of classes and synthesizes ideas from factor analysis, latent variable theory and generalized linear model residuals. However, such two-stage approaches are inefficient and may create misleading results. A poor goodness of fit may be the result of poor parameter estimation or wrong model assumptions or a bad choice for the number of classes. Without knowing the exact sources for poor goodness of fit, the use of goodness-of-fit tests in the specification of the number of classes may result in over parameterization in the number of classes and create a meaningless interpretation.

Joint inferences on the number of classes and model parameters are preferable because it is convenient, accurate and flexible. Traditional frequentist likelihood-based approaches do not allow this joint analysis, but advances in Bayesian inferences provide possible solutions. Markov chain Monte Carlo (MCMC) has had a profound effect on Bayesian statistics. MCMC draws samples from the posterior distribution by running a cleverly constructed Markov chain (e.g., a Metropolis–Hastings algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970), and then forms sample averages to approximate posterior expectations of model parameters. The MCMC method is restricted to problems where the joint distribution has a density with respect to a fixed dimensional parameter space. Green (1995) proposed the reversible jump Markov chain Monte Carlo (RJMCMC) method, which offers a general framework for construction of reversible Markov chain samplers that jump between parameter subspaces of different dimensionality. Latent class models with different numbers of classes correspond to parameter subspaces of different dimensionality. RJMCMC thus provides a solution that can jointly estimate the number of classes and model parameters.

Richardson and Green (1997) made use of RJMCMC to estimate the number of components and the mixture component parameters jointly in finite mixtures of univariate normal distributions. Since then, RJMCMC has been applied to many other mixture distributions, such as mixtures of exponential distributions (Gruet, Philippe, & Robert, 1999), mixtures of Poisson distributions (Viallefont, Richardson, & Green, 2002), mixtures of multivariate normal distributions (Dellaportas & Papageorgiou, 2006), and mixtures of multivariate Poisson distributions (Meligkotsidou, 2007). Bartolucci, Mira, and Scaccia (2003) and Pandolfi, Bartolucci, and Friel (2010) discussed the RJMCMC implementation on latent class models with binary measured items. These can be viewed as finite mixtures of multivariate Bernoulli distributions. Bartolucci et al. (2003) used standard RJMCMC methods for updating estimates between different number of classes, and applied the delayed rejection strategy for increasing the acceptance rate of class number change. Pandolfi et al. (2010) generalized the multiple-try Metropolis algorithm for RJMCMC, and applied this generalization to latent class models to improve the mixing of estimates. In this paper, we propose to implement the RJMCMC method to perform the joint estimation of the number of classes and model parameters on a more general latent class model that has polytomous measured items and incorporates various covariate effects. We thus focus on the model of finite mixtures of multivariate multinomial distributions incorporating covariate effects.

We now provide an outline for the remainder of the paper. Section 2 gives a Bayesian version of RLCA. Section 3 illustrates the RJMCMC scheme we propose and how we deal with the label switching problem. In Section 4, simulation studies are conducted to evaluate the behavior of the proposed estimation method and the comparison with existing approaches. In Section 5, we perform a sensitivity analysis that discusses results from various hyperparameters and

proposal parameters, and provides default recommendations for the values of hyperparameters and proposal parameters. In Section 6, data on determining subtypes of schizophrenia using the Positive and Negative Syndrome Scale (PANSS) are used to illustrate the proposed methods. A discussion is provided in Section 7.

2. Bayesian Models for Regression Extension of Latent Class Analysis

2.1. Regression Extension Latent Class Model

To specify a latent class model with J classes, we define $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iM})^T$ to be a set of M polytomous response variables for the ith individual, $i = 1, \ldots, N$. The mth variable, Y_{im}, can take one of values $\{1, \ldots, K_m\}$, where $K_m \geq 2$; the latent variable, S_i, denotes the subpopulation in which the ith individual belongs to, and takes a value $\{1, \ldots, J\}$. The basic idea of latent class modelling is that the components of $\mathbf{Y}_i$ are associated because the overall population is comprised of a mixture of J subpopulations or classes, and these components are assumed to be statistically independent within latent classes. Therefore, the distribution of $\mathbf{Y}_i$ can be expressed as the finite mixture density:

$$\Pr(Y_{i1} = y_{i1}, \ldots, Y_{iM} = y_{iM}) = \sum_{j=1}^{J} \left\{ \Pr(S_i = j) \prod_{m=1}^{M} \prod_{k=1}^{K_m} \left[\Pr(Y_{im} = k | S_i = j)\right]^{y_{imk}} \right\}, \tag{1}$$

where $y_{imk} = I(y_{im} = k) = 1$ if $y_{im} = k$; 0 otherwise. The LCA model assumes that $\Pr(S_i = j) = \eta_j$ and $\Pr(Y_{im} = k | S_i = j) = \pi_{mkj}$ for all i. Thus, the model treats class membership probabilities, η_j, and item response probabilities conditional on class membership, π_{mkj}, as homogeneous over individuals. Huang and Bandeen-Roche (2004) incorporated covariates $(\mathbf{x}_i, \mathbf{z}_i)$ into LCA, where $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{iP})^T$ are predictors associated with latent variable S_i, and $\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iM})$; $\mathbf{z}_{im} = (z_{im1} \ldots, z_{imL})^T$ with $m = 1, \ldots, M$ are covariates built to have direct influence on response variables. Then the LCA model can be broadened to RLCA and be stated as

$$\Pr(Y_{i1} = y_{i1}, \ldots, Y_{iM} = y_{iM} | \mathbf{x}_i, \mathbf{z}_i) = \sum_{j=1}^{J} \left\{ \eta_j(\mathbf{x}_i) \prod_{m=1}^{M} \prod_{k=1}^{K_m} \pi_{mkj}^{y_{imk}}(\mathbf{z}_{im}) \right\}, \tag{2}$$

where $\eta_j(\mathbf{x}_i)$ and $\pi_{mkj}(\mathbf{z}_{im})$ are often implemented assuming the generalized logit link function under the generalized linear model framework (McCullagh & Nelder, 1989):

$$\log\left[\frac{\eta_j(\mathbf{x}_i)}{\eta_J(\mathbf{x}_i)}\right] = \beta_{0j} + \beta_{1j} x_{i1} + \cdots + \beta_{Pj} x_{iP} \tag{3}$$

and

$$\log\left[\frac{\pi_{mkj'}(\mathbf{z}_{im})}{\pi_{mK_mj'}(\mathbf{z}_{im})}\right] = \gamma_{mkj'} + \alpha_{1mk} z_{im1} + \cdots + \alpha_{Lmk} z_{imL} \tag{4}$$

for $i = 1, \ldots, N$; $m = 1, \ldots, M$; $k = 1, \ldots, (K_m - 1)$; $j = 1, \ldots, (J - 1)$; $j' = 1, \ldots, J$.

Notice that in the conditional probability model (4), we allow unrestricted intercepts and level- and item-specific covariate coefficients, but we do not allow the coefficients to vary across classes (i.e., α_{lmk} is dependent on m, k but independent of j'). This constraint is logical if the primary purpose of modeling conditional probabilities is to prevent possible misclassification by adjusting for characteristics associated with item measurements (Huang & Bandeen-Roche,

2004). It is also necessary to unambiguously distinguish covariate effects on measured response probabilities from covariate effects on class probabilities. Three assumptions are necessary to complete model (2):

(i) $\Pr(S_i = j|\mathbf{x}_i, \mathbf{z}_i) = \Pr(S_i = j|\mathbf{x}_i)$;
(ii) $\Pr(Y_{i1} = y_{i1}, \ldots, Y_{iM} = y_{iM}|S_i, \mathbf{x}_i, \mathbf{z}_i) = \Pr(Y_{i1} = y_{i1}, \ldots, Y_{iM} = y_{iM}|S_i, \mathbf{z}_i)$;
(iii) $\Pr(Y_{i1} = y_{i1}, \ldots, Y_{iM} = y_{iM}|S_i, \mathbf{z}_i) = \prod_{m=1}^{M} \Pr(Y_{im} = y_{im}|S_i, \mathbf{z}_i)$.

These assumptions give rise to implications that, from (i), class membership probabilities are associated with $\mathbf{x}_i$ only; from (ii), conditioning on class membership, responses are only associated with $\mathbf{z}_i$; and from (iii), the multiple measurements are independent given class membership S_i and $\mathbf{z}_i$.

2.2. Hierarchical Model and Priors

Under a Bayesian framework, we specify prior distributions for the parameters in the RLCA model. Parameters in RLCA can be summarized as the number of classes J, regression parameters for membership probabilities $\boldsymbol{\beta} = [\beta_{pj}]_{0\le p\le P, 1\le j\le J-1}$, regression intercepts for modelling conditional probabilities $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_M^T]^T$ with $\boldsymbol{\gamma}_m = [\gamma_{mkj}]_{1\le k\le K_m-1, 1\le j\le J}$, and covariate coefficients for modelling conditional probabilities $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M]^T$ with $\boldsymbol{\alpha}_m = [\alpha_{lmk}]_{1\le l\le L, 1\le k\le K_m-1}$.

We set all components of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ to be independently and normally distributed with mean 0 and pre-specified variance σ_P^2. Here, we adopt a noninformative prior mean zero. Such a setting is common when prior scientific knowledge does not provide appropriate values of parameters (Garrett & Zeger, 2000). For J, we assume it to follow an uniform distribution between 1 and a pre-specified integer $J_{\max}$. Class membership inference is often of our interest. We thus add the latent class variable into the joint distribution where Bayesian estimation is inferred from. Therefore, based on assumptions (i), (ii) and (iii) with selected prior distributions, the joint distribution of all variables for the RLCA model given covariates $(\mathbf{x}, \mathbf{z})$ can be represented as

$$\begin{aligned} p(J, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}, \mathbf{Y}; \mathbf{x}, \mathbf{z}) &= p(J)\times p(\boldsymbol{\beta}|J)\times p(\boldsymbol{\gamma}|J, \boldsymbol{\beta})\times p(\boldsymbol{\alpha}|J, \boldsymbol{\beta}, \boldsymbol{\gamma}) \\ &\quad \times p(\mathbf{S}|J, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{x}, \mathbf{z})\times p(\mathbf{Y}|J, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{x}, \mathbf{z}) \\ &= p(J)\times p(\boldsymbol{\beta}|J)\times p(\boldsymbol{\gamma}|J)\times p(\boldsymbol{\alpha}) \\ &\quad \times p(\mathbf{S}|J, \boldsymbol{\beta}; \mathbf{x})\times p(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}), \end{aligned} \tag{5}$$

where $\mathbf{Y} = [\mathbf{Y}_1^T, \ldots, \mathbf{Y}_N^T]^T$, $\mathbf{S} = (S_1, \ldots, S_N)^T$ and, here and throughout this paper, we use $p(\cdot)$ and $p(\cdot|\cdot)$ to denote joint and conditional distribution functions, respectively. Moreover, the distributions of $\mathbf{S}$ and $\mathbf{Y}$ given $\mathbf{S}$ can be expressed as

$$\begin{aligned} p(\mathbf{S}|J, \boldsymbol{\beta}; \mathbf{x}) &= \prod_{i=1}^{N} \Pr(S_i = s_i|J, \boldsymbol{\beta}; \mathbf{x}_i) \\ &= \prod_{i=1}^{N}\prod_{j=1}^{J} \eta_j^{s_{ij}}(\mathbf{x}_i) \end{aligned} \tag{6}$$

and

$$p(\mathbf{Y}|J,\boldsymbol{\gamma},\boldsymbol{\alpha},\mathbf{S};\mathbf{z}) = \prod_{i=1}^{N} \Pr(Y_{i1}=y_{i1},\ldots,Y_{iM}=y_{iM}|J,\boldsymbol{\gamma},\boldsymbol{\alpha},S_i=s_i;\mathbf{z}_i)$$
$$= \prod_{i=1}^{N}\prod_{j=1}^{J}\left(\prod_{m=1}^{M}\prod_{k=1}^{K_m}\pi_{mkj}^{y_{imk}}(\mathbf{z}_{im})\right)^{s_{ij}} \tag{7}$$

with $s_{ij} = I(s_i = j)$ and $y_{imk} = I(y_{im} = k)$.

It is important to talk about the issue of labeling the classes. Our RLCA likelihood is invariant when switching class labels. As a result, the posterior distribution is not identifiable for various permutations of class labels. We thus need to adopt a unique labeling for the class membership. Once a unique labeling is determined, such a procedure is equivalent to ordering the joint prior distribution. So, the joint prior distribution of the parameters is the number of invariant class label permutations times the product of the individual densities. Our labeling procedure is detailed in Section 3.4.

3. Bayesian Inferences with Variable Dimension Parameters

3.1. Reversible Jump Markov Chain Monte Carlo Algorithm

The Markov chain Monte Carlo (MCMC) method for the simulation of a distribution produces an ergodic Markov chain whose stationary distribution is the distribution of interest (Robert & Casella, 2004). There have been several earlier approaches on MCMC computation to deal with models with changing dimensionality, for instance, birth-and-death processes (Ripley, 1977; Geyer & Møller, 1994) or pseudo-priors (Carlin & Chib, 1995), but the general formalization named reversible jump MCMC (RJMCMC) had not been proposed until Green (1995). In brief, RJMCMC is a random sweep Metropolis–Hastings method (Metropolis et al., 1953; Hastings, 1970) adapted for general state spaces (Richardson & Green, 1997). In this paper, we use the RJMCMC algorithm for the situation where the number of classes is unfixed and to be determined.

For our hierarchical RLCA model, we shall make use of six move types:

(a) Updating the class memberships $\mathbf{S}$;
(b) Updating the regression parameters for membership probabilities $\boldsymbol{\beta}$;
(c) Updating the regression intercepts for conditional probabilities $\boldsymbol{\gamma}$;
(d) Updating the covariate coefficients for conditional probabilities $\boldsymbol{\alpha}$;
(e) Birth or death of a class;
(f) Splitting one class into two, or merging two classes into one.

Move types (a)–(d) are conventional and can apply the Gibbs sampling scheme (Geman & Geman, 1993; Zeger & Karim, 1991). Moves (e) and (f) involve updating the value of J, implying a change of dimensionality for the parameters in (b) and (c). One complete pass over these six moves will be called "a sweep" and is the basic time step of the algorithm.

3.2. Gibbs Sampling Scheme

For illustrating Gibbs sampling, we need to calculate the full conditional distribution of the interested variable given all others. In the case of S_i, the full conditional distribution is

$$p(S_i|J,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha},\mathbf{Y}_i;\mathbf{x}_i,\mathbf{z}_i) \propto p(\mathbf{Y}_i|J,\boldsymbol{\gamma},\boldsymbol{\alpha},S_i;\mathbf{z}_i) \times p(S_i|J,\boldsymbol{\beta};\mathbf{x}_i). \tag{8}$$

Since S_i is discrete, the sample of S_i given all other variables can be drawn directly from a multinomial distribution with probabilities in the right-hand side of (8), scaling to sum to 1.

The form of the full conditional distribution of $\boldsymbol{\beta}$ makes sampling from it difficult. Instead of sampling directly from the full conditional distribution, we use rejection sampling methods, where a sample from an envelope distribution where sampling is easier is first generated and this sample is then accepted with certain probability such that the accepted samples are precisely from full conditional distributions (Zeger & Karim, 1991). We had tried out the standard Metropolis–Hastings procedure for Steps (b), (c) and (d). However, the acceptance rate of Metropolis–Hastings sampling was very low and very long MCMC runs were needed. To generate samples of $\boldsymbol{\beta}$, $\boldsymbol{\beta}$ is divided into components $\{\boldsymbol{\beta}^{[1]}, \ldots, \boldsymbol{\beta}^{[J-1]}\}$ with $\boldsymbol{\beta}^{[j]}$ being the jth column of $\boldsymbol{\beta}$. Each component $\boldsymbol{\beta}^{[j]}$ is then updated conditioning on $\boldsymbol{\beta}^{[-j]}$, where $\boldsymbol{\beta}^{[-j]}$ is the $\boldsymbol{\beta}$ matrix deleting the jth column. The full conditional distribution of $\boldsymbol{\beta}^{[j]}$ given $\boldsymbol{\beta}^{[-j]}$ is in the form of

$$\begin{aligned} p\big(\boldsymbol{\beta}^{[j]}|J, \boldsymbol{\beta}^{[-j]}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}, \mathbf{Y}; \mathbf{x}, \mathbf{z}\big) &\propto p(\mathbf{S}|J, \boldsymbol{\beta}; \mathbf{x}) \times p\big(\boldsymbol{\beta}^{[j]}|J\big) \\ &= p\big(\mathbf{S}|J, \boldsymbol{\beta}^{[j]}, \boldsymbol{\beta}^{[-j]}; \mathbf{x}\big) p\big(\boldsymbol{\beta}^{[j]}|J\big). \end{aligned} \tag{9}$$

Following the approach of Zeger and Karim (1991), we choose the envelope distribution for $\boldsymbol{\beta}^{[j]}$, $g_1(\boldsymbol{\beta}^{[j]})$, to be a multivariate Gaussian distribution with mean $\hat{\boldsymbol{\beta}}^{[j]}$, the maximum likelihood estimator of $\boldsymbol{\beta}^{[j]}$ in (9), and variance $c\hat{\Sigma}_{\boldsymbol{\beta}^{[j]}}$, the inverse of the observed fisher information of (9) multiplied by a constant c. To perform rejection sampling, a constance c^* is chosen so that $p(\mathbf{S}|J, \boldsymbol{\beta}^{[j]}, \boldsymbol{\beta}^{[-j]}; \mathbf{x}) p(\boldsymbol{\beta}^{[j]}|J) \leq (c^* \cdot g_1(\boldsymbol{\beta}^{[j]}))$, over all $\boldsymbol{\beta}^{[j]} \in \mathbb{R}^{P+1}$. A vector sample of $\boldsymbol{\beta}^{[j]}$ is produced by the followings steps:

1. Generate a random variate $\boldsymbol{\beta}^*$ from $g_1(\boldsymbol{\beta}^{[j]})$;
2. Generate a random variate u from uniform (0, 1);
3. If $u \leq \frac{p(\mathbf{S}|J, \boldsymbol{\beta}^*, \boldsymbol{\beta}^{[-j]}; \mathbf{x}) p(\boldsymbol{\beta}^*|J)}{c^* \cdot g_1(\boldsymbol{\beta}^*)}$, accept $\boldsymbol{\beta}^*$, otherwise return to Step 1.

In $\boldsymbol{\beta}$'s rejection sampling scheme, we let $c = 1.2$ and set c^* to attain

$$p\big(\mathbf{S}|J, \hat{\boldsymbol{\beta}}^{[j]}, \boldsymbol{\beta}^{[-j]}; \mathbf{x}\big) p\big(\hat{\boldsymbol{\beta}}^{[j]}|J\big) = c^* \cdot g_1\big(\hat{\boldsymbol{\beta}}^{[j]}\big).$$

Our selection of c^* is adopted from Zeger and Karim (1991). These selections worked well in covering (9) in the simulations and real data analysis below.

Rejection method is applied to $\boldsymbol{\gamma}$ as well. First, $\boldsymbol{\gamma}$ is separated into column vectors $\{\boldsymbol{\gamma}^{[1]}, \ldots, \boldsymbol{\gamma}^{[J]}\}$. A column vector $\boldsymbol{\gamma}^{[j]}$ is then updated from the full conditional distribution

$$\begin{aligned} p\big(\boldsymbol{\gamma}^{[j]}|J, \boldsymbol{\beta}, \boldsymbol{\gamma}^{[-j]}, \boldsymbol{\alpha}, \mathbf{S}, \mathbf{Y}; \mathbf{x}, \mathbf{z}\big) &\propto p(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}) \times p\big(\boldsymbol{\gamma}^{[j]}|J\big) \\ &= p\big(\mathbf{Y}|J, \boldsymbol{\gamma}^{[j]}, \boldsymbol{\gamma}^{[-j]}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}\big) p\big(\boldsymbol{\gamma}^{[j]}|J\big), \end{aligned} \tag{10}$$

where $\boldsymbol{\gamma}^{[-j]}$ is the $\boldsymbol{\gamma}$ matrix deleting the jth column. The procedure of sampling $\boldsymbol{\gamma}^{[j]}$ is similar to that for $\boldsymbol{\beta}^{[j]}$, and $\boldsymbol{\gamma}$ is drawn by completing the loop for $\boldsymbol{\gamma}^{[j]}$, $j = 1, \ldots, J$. In $\boldsymbol{\gamma}$'s rejection sampling, we let $c = 1.3$ and set c^* to attain $p(\mathbf{Y}|J, \hat{\boldsymbol{\gamma}}^{[j]}, \boldsymbol{\gamma}^{[-j]}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}) p(\hat{\boldsymbol{\gamma}}^{[j]}|J) = c^* \cdot g_2(\hat{\boldsymbol{\gamma}}^{[j]})$, where $g_2(\cdot)$ denotes the density of a multivariate Gaussian distribution with mean $\hat{\boldsymbol{\gamma}}^{[j]}$ and variance $c\hat{\Sigma}_{\boldsymbol{\gamma}^{[j]}}$.

For generating $\boldsymbol{\alpha}$, we choose to update $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M\}$ sequentially. Each $\boldsymbol{\alpha}_m$ is divided into $\{\boldsymbol{\alpha}_m^{[1]}, \ldots, \boldsymbol{\alpha}_m^{[K_m-1]}\}$ with $\boldsymbol{\alpha}_m^{[k]}$ being the kth column of $\boldsymbol{\alpha}_m$. We first sample the values of $\boldsymbol{\alpha}_m^{[k]}$ from the full conditional distribution

$$\begin{aligned} p\big(\boldsymbol{\alpha}_m^{[k]}|J, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_m^{[-k]}, \boldsymbol{\alpha}_{-m}, \mathbf{S}, \mathbf{Y}; \mathbf{z}\big) &\propto p(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}) \times p\big(\boldsymbol{\alpha}_m^{[k]}\big) \\ &= p\big(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}_m^{[k]}, \boldsymbol{\alpha}_m^{[-k]}, \boldsymbol{\alpha}_{-m}, \mathbf{S}; \mathbf{z}\big) p\big(\boldsymbol{\alpha}_m^{[k]}\big) \end{aligned} \tag{11}$$

by the rejection sampling, where $\boldsymbol{\alpha}_m^{[-k]}$ is the $\boldsymbol{\alpha}_m$ matrix deleting the kth column and $\boldsymbol{\alpha}_{-m}$ is the $\boldsymbol{\alpha}$ collections without the component $\boldsymbol{\alpha}_m$. The rejection sampling steps are similar to those for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The value of $\boldsymbol{\alpha}$ was completely drawn by executing the loop $\{\boldsymbol{\alpha}_1^{[1]}, \ldots, \boldsymbol{\alpha}_1^{[K_1-1]}, \ldots, \boldsymbol{\alpha}_M^{[1]}, \ldots, \boldsymbol{\alpha}_M^{[K_M-1]}\}$. In $\boldsymbol{\alpha}$'s rejection sampling, we let $c = 1.2$ and set c^* to attain $p(\mathbf{Y}|J, \boldsymbol{\gamma}, \hat{\boldsymbol{\alpha}}_m^{[k]}, \boldsymbol{\alpha}_m^{[-k]}, \boldsymbol{\alpha}_{-m}, \mathbf{S}; \mathbf{z}) p(\hat{\boldsymbol{\alpha}}_m^{[k]}) = c^* \cdot g_3(\hat{\boldsymbol{\alpha}}_m^{[k]})$, where $g_3(\cdot)$ denotes the multivariate Gaussian density with mean $\hat{\boldsymbol{\alpha}}_m^{[k]}$ and variance $c\hat{\Sigma}_{\boldsymbol{\alpha}_m^{[k]}}$.

3.3. Reversible Jump Steps

For moves (e) and (f), the reversible jump mechanism is needed. The strategy of moves in RLCA is similar to Richardson and Green (1997). Increasing or decreasing of J fetches a dimensional change of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. In Step (e), a choice of birth or death is made given the equal probability 0.5. When a birth is selected, we produce a $(P+1) \times 1$ random vector $\boldsymbol{\beta}^{[j^*]}$ with each component generated from $N(0, \sigma_{\mathrm{BD}}^2)$. Next, $\boldsymbol{\beta}^{[j^*]}$ is added to the first column of $\boldsymbol{\beta}$ and the newly formed matrix is labeled as $\boldsymbol{\beta}^{\mathrm{birth}}$. The same procedure holds for $\boldsymbol{\gamma}$ as well, and the $(\sum_{m=1}^{M}(K_m - 1)) \times (J+1)$ matrix $\boldsymbol{\gamma}^{\mathrm{birth}}$ is the newly born matrix. It remains to propose the reallocation of $S_i \in \{1, \ldots, J+1\}$. This is done analogously to our Gibbs allocation (8) and creates a new class allocation $\mathbf{S}^{\mathrm{birth}}$. Reversely, a death move first randomly draws a number among $\{1, \ldots, J-1\}$, say j^*, and then the columns $\boldsymbol{\beta}^{[j^*]}$ and the $\boldsymbol{\gamma}^{[j^*]}$ are removed from $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively, to form new parameter matrices $\boldsymbol{\beta}^{\mathrm{death}}$ and $\boldsymbol{\gamma}^{\mathrm{death}}$. The reallocation of $S_i \in \{1, \ldots, J-1\}$ is done analogously to (8), and the new class allocation is labeled as $\mathbf{S}^{\mathrm{death}}$. Note that in the birth-and-death step, the generated group or deleted group does not directly birth in or delete out the reference group. We leave the opportunity of removing or adding the reference ($j^* = J$) to the work done by the split-and-merge step. Also, it is easy to see that the proposed birth-and-death move forms a reversible pair.

Following the RJMCMC recipe in Richardson and Green (1997), the acceptance probability for the birth move is $\min\{1, A_B\}$ with

$$A_B = \frac{p(J+1)}{p(J)} \times L_{+1} \times \frac{p(\boldsymbol{\beta}^{\mathrm{birth}}|J+1) \times p(\boldsymbol{\gamma}^{\mathrm{birth}}|J+1)}{p(\boldsymbol{\beta}|J) \times p(\boldsymbol{\gamma}|J)} \times \frac{1}{h_1(\boldsymbol{\beta}^{[j^*]}) \times h_2(\boldsymbol{\gamma}^{[j^*]})} \times \frac{1}{J}. \tag{12}$$

In (12), J is the number of classes before birth. L_{+1} arises from the order statistics for the unique class labeling, and L_{+1} is $J!/(J-1)!$ if $J \geq 3$; 1 if $J = 2$; 2 if $J = 1$. In the second line of (12), $h_1(\boldsymbol{\beta}^{[j^*]}) = \prod_{p=0}^{P} h(\beta_{pj^*})$ and $h_2(\boldsymbol{\gamma}^{[j^*]}) = \prod_{m=1}^{M} \prod_{k=1}^{K_m-1} h(\gamma_{mkj^*})$ are the proposal densities for the birth columns $\boldsymbol{\beta}^{[j^*]}$ and $\boldsymbol{\gamma}^{[j^*]}$ with $h(\cdot)$ being the $N(0, \sigma_{\mathrm{BD}}^2)$ density. Technical details in deriving this acceptance probability A_B for the birth move can be found in Section S.1 of the supplementary information. For the corresponding death move, the acceptance probability is $\min\{1, A_D\}$, where

$$A_D = \frac{p(J-1)}{p(J)} \times L_{-1} \times \frac{p(\boldsymbol{\beta}^{\mathrm{death}} \mid J-1) \times p(\boldsymbol{\gamma}^{\mathrm{death}} \mid J-1)}{p(\boldsymbol{\beta} \mid J) \times p(\boldsymbol{\gamma} \mid J)} \times h_1\big(\boldsymbol{\beta}^{[j^*]}\big) \times h_2\big(\boldsymbol{\gamma}^{[j^*]}\big) \times J - 1, \tag{13}$$

and L_{-1} is $1/(J-1)$ if $J \geq 4$; 1 if $J = 3$; 0.5 if $J = 2$.

For move (f), we make a random choice between attempting to split or merge with equal probability 0.5. Once the split procedure is adopted, randomly chosen columns $\boldsymbol{\beta}^{[j^*]}$ and $\boldsymbol{\gamma}^{[j^*]}$ from $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively, are each split into two columns, where $j^* \in \{1, \ldots, J\}$ and

$\boldsymbol{\beta}^{[J]} = \mathbf{0}$. Given a random vector $\mathbf{u}$ generated from a multivariate normal distribution with mean $(\overbrace{0, \ldots, 0}^{P+1})^T$ and variance $\text{diag}(\overbrace{\sigma^2_{\text{SM}}, \ldots, \sigma^2_{\text{SM}}}^{P+1})$, $\boldsymbol{\beta}^{[j^*]}$ is divided into two column vectors $\boldsymbol{\beta}^{[j_1]}$ and $\boldsymbol{\beta}^{[j_2]}$ with $\boldsymbol{\beta}^{[j_1]} = \boldsymbol{\beta}^{[j^*]} + \mathbf{u}$ and $\boldsymbol{\beta}^{[j_2]} = \boldsymbol{\beta}^{[j^*]} - \mathbf{u}$. Similarly, a random vector $\mathbf{v}$ is generated from a multivariate normal distribution with mean $(\overbrace{0, \ldots, 0}^{\sum_m (K_m - 1)})^T$ and variance $\text{diag}(\overbrace{\sigma^2_{\text{SM}}, \ldots, \sigma^2_{\text{SM}}}^{\sum_m (K_m - 1)})$ as well, and $\boldsymbol{\gamma}^{[j^*]}$ is split into $\boldsymbol{\gamma}^{[j_1]} = \boldsymbol{\gamma}^{[j^*]} + \mathbf{v}$ and $\boldsymbol{\gamma}^{[j_2]} = \boldsymbol{\gamma}^{[j^*]} - \mathbf{v}$. The new beta and gamma matrices, $\boldsymbol{\beta}^{\text{split}}$ and $\boldsymbol{\gamma}^{\text{split}}$, are formed by inserting $\boldsymbol{\beta}^{[j_1]}$ and $\boldsymbol{\gamma}^{[j_1]}$ before the first column of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and replacing $\boldsymbol{\beta}^{[j^*]}$ and $\boldsymbol{\gamma}^{[j^*]}$ with $\boldsymbol{\beta}^{[j_2]}$ and $\boldsymbol{\gamma}^{[j_2]}$, respectively. Reallocation of subjects is done with the Gibbs sampling move (8). The new allocation is then labeled as $\mathbf{S}^{\text{split}}$. When the reference class is chosen (i.e., $j^* = J$), splitting $\boldsymbol{\beta}$ needs additional attention. When $j^* = J$, the two split beta matrices are $\boldsymbol{\beta}^{[j_1]} = \mathbf{u}$ and $\boldsymbol{\beta}^{[j_2]} = -\mathbf{u}$. By the splitting rule we established above, $\mathbf{u}$ is with respect to the first class and $-\mathbf{u}$ is with respect to the new reference class. To maintain $\boldsymbol{\beta}^{[J]} = \mathbf{0}$ in latent class model (3), the beta matrix should be adjusted for $-\mathbf{u}$, and $\boldsymbol{\beta}^{\text{split}}$ becomes $[2\mathbf{u}, \boldsymbol{\beta}^{[1]} + \mathbf{u}, \ldots, \boldsymbol{\beta}^{[J-1]} + \mathbf{u}]$.

A merge move in $\boldsymbol{\gamma}$ starts with a randomly selected column vector $\boldsymbol{\gamma}^{[j_2]}$ with $j_2 \in \{1, \ldots, J\}$, and then $\boldsymbol{\gamma}^{[j_2]}$ and its "closest" column vector $\boldsymbol{\gamma}^{[j_1]}$ are merged, where

$$j_1 = \arg \min_{i \in \{1, \ldots, J-1\} \setminus \{j_2\}} \left(\boldsymbol{\gamma}^{[j_2]} - \boldsymbol{\gamma}^{[i]}\right)^T \left(\boldsymbol{\gamma}^{[j_2]} - \boldsymbol{\gamma}^{[i]}\right). \tag{14}$$

Note that we exclude J from the possible closest index set for the purpose of forming a reversible pair in the split-and-merge step. The merged column $\boldsymbol{\gamma}^{[j^*]} = (\boldsymbol{\gamma}^{[j_1]} + \boldsymbol{\gamma}^{[j_2]})/2$. With j_1 and j_2 obtained from $\boldsymbol{\gamma}$, $\boldsymbol{\beta}^{[j_1]}$ and $\boldsymbol{\beta}^{[j_2]}$ are also merged as $\boldsymbol{\beta}^{[j^*]} = (\boldsymbol{\beta}^{[j_1]} + \boldsymbol{\beta}^{[j_2]})/2$. The new allocation $\mathbf{S}^{\text{merge}}$ is formed by reallocating all subjects according to (8). The new merged matrices, $\boldsymbol{\gamma}^{\text{merge}}$ and $\boldsymbol{\beta}^{\text{merge}}$, are formed by replacing $\boldsymbol{\gamma}^{[j_2]}$ and $\boldsymbol{\beta}^{[j_2]}$ with $\boldsymbol{\gamma}^{[j^*]}$ and $\boldsymbol{\beta}^{[j^*]}$, respectively, and deleting $\boldsymbol{\gamma}^{[j_1]}$ and $\boldsymbol{\beta}^{[j_1]}$. When the reference group is selected to be merged (i.e., $j_2 = J$), $\boldsymbol{\beta}^{[j_1]}$ is first deleted from $\boldsymbol{\beta}$ and then every remaining column of $\boldsymbol{\beta}$ is subtracted by $\boldsymbol{\beta}^{[j_1]}/2$ to maintain zero $\boldsymbol{\beta}$ coefficients of the reference class. It can be readily shown that our merge and split proposals are reversible.

The acceptance probability for the split is $\min\{1, A_S\}$, where

$$\begin{aligned} A_S = {} & \frac{p(J+1)}{p(J)} \times L_{+1} \times \frac{p(\boldsymbol{\beta}^{\text{split}} | J+1) \times p(\boldsymbol{\gamma}^{\text{split}} | J+1)}{p(\boldsymbol{\beta} | J) \times p(\boldsymbol{\gamma} | J)} \\ & \times \frac{1}{h_3(\mathbf{u}) \times h_4(\mathbf{v})} \times \frac{J}{J+1} \times \frac{1}{w_J(j^*)} \times \left(\frac{1}{2}\right)^{(P+1)+(\sum_{m=1}^{M}(K_m-1))}. \end{aligned} \tag{15}$$

Here, J is the number of class before splitting. As before, L_{+1} arises from the order statistics permutations of the current and split states, and $h_3(\mathbf{u})$ and $h_4(\mathbf{v})$ are the densities for generating $\mathbf{u}$ and $\mathbf{v}$. The weight $w_J(j^*)$ is defined as

$$w_J(j^*) = \begin{cases} 2, & \text{if } j^* \neq J, \\ 1, & \text{if } j^* = J. \end{cases}$$

Technical details in deriving the acceptance probability A_S for the split move can be found in Section S.1 of the supplementary information. Notice that, in split move, it is necessary to check whether the adjacency condition (14) is satisfied. If not, the split move is rejected forthwith for the reason that the split-merge pair is not reversible (Richardson & Green, 1997). The acceptance

probability for the merge can be similarly obtained as $\min\{1, A_M\}$, where

$$A_M = \frac{p(J-1)}{p(J)} \times L_{-1} \times \frac{p(\boldsymbol{\beta}^{\text{merge}} \mid J-1) \times p(\boldsymbol{\gamma}^{\text{merge}} \mid J-1)}{p(\boldsymbol{\beta} \mid J) \times p(\boldsymbol{\gamma} \mid J)} \times h_3(\mathbf{u}) \times h_4(\mathbf{v}) \times \frac{J}{J-1} \times w_{J-1}(j^*) \times 2^{(P+1)+(\sum_{m=1}^{M}(K_m-1))}. \tag{16}$$

There are some special features in the proposed reversible jump steps. First, the reallocation of class membership in reversible jump steps is done for all subjects, not just for subjects who belong to the classes that are selected for change. This is because any birth–death or split-merge in the $\boldsymbol{\beta}$ coefficient will result in different class weights in all classes (Equation (3)); thus, subjects in the new state have different probabilities to be in any of the classes, and reallocation for all subjects is necessary. Second, unlike Richardson and Green (1997), where the birth-and-death step is only for empty classes, our birth–death step is applied to all classes. We experimented with the empty-class-only birth–death step for our model; however, the acceptance rate was so low that the empty classes existed all the time. Our approach allows birth or death on all classes and then reallocates all subjects, which can efficiently eliminate empty classes. Third, our $\boldsymbol{\beta}$ coefficient only corresponds to the first $J-1$ classes (the reference class has zero $\boldsymbol{\beta}$ coefficients); thus, some care is needed for renewing the $\boldsymbol{\beta}$ coefficient with respect to the reference class. We choose to update parameters of the reference class in the split-merge move but not in the birth–death to ease the model complication. We have found this simplification works well in our model.

3.4. Labeling Procedure

For identifiability concerns, we have adopted a unique labeling for class membership in each RJMCMC sweep. In our labeling, we use a modified version of the on-line processing method originally developed by Celeux, Hurn, and Robert (2000). The labeling procedure corresponds to an ordering of columns of $\boldsymbol{\gamma}$. It is performed to relabel classes after running each RJMCMC sweep, and sweeps resulting in the same number of classes are collected altogether for the relabeling process. Specifically, let $\boldsymbol{\gamma}^1 = [\gamma^1_{mkj}]$, $\boldsymbol{\gamma}^2 = [\gamma^2_{mkj}]$, ... be the sequence of RJMCMC samples for $\boldsymbol{\gamma}$ that contains J latent classes. The first q samples after burn-in $\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^q$ are used to initialize the procedure, where the choice of q has to be large enough to ensure initial estimates that crudely approximate posterior means, but is not so large that any label switch has occurred. For choosing a proper value of q, one can begin the RJMCMC with a pilot run. Then a trace plot of one particular parameter at every class can be drawn to look for where the first class switching phenomenon is taking place after burn-in. The choice of value of q should be smaller than the first sweep of switching problem. We take the dataset used in the sensitivity analysis for illustrating how to choose the value of q. The left-hand side of Figure 1 contains the trace plots of γ_{52j}, $j = 1, \ldots, 6$ at the first 5000 sweeps after burn-in before relabeling. From the figure, the first class switching occurred at about the 3000th sweep, and we can take the value of q not larger than 3000. Note that q is typically set to be 100 in this study.

Reference centers and variances are defined, respectively, as

$$\gamma^{(0)}_{mkj} = \frac{1}{q}\sum_{t=1}^{q} \gamma^t_{mkj}$$

and

$$s^{(0)}_{mkj} = \frac{1}{q}\sum_{t=1}^{q} \left(\gamma^t_{mkj} - \gamma^{(0)}_{mkj}\right)^2.$$

Let $(\boldsymbol{\gamma}^t)^{[i]}$ denote the ith column of $\boldsymbol{\gamma}^t$, and let $\boldsymbol{\gamma}^{(0)}$ be the matrix with components $\gamma_{mkj}^{(0)}$. The procedure runs to update the class labels for $\boldsymbol{\gamma}^{q+1}, \boldsymbol{\gamma}^{q+2}, \ldots$ sequentially with $\boldsymbol{\gamma}^{q+r}$ being processed as follows:

1. Let

$$\begin{aligned}
\mathbf{1.1} \quad (i_{j_1}, j_1) &= \arg \min_{\{(i,j)|i,j=1,\ldots,J\}} \sum_{m=1}^{M} \sum_{k=1}^{K_m-1} \frac{(\gamma_{mki}^{q+r} - \gamma_{mkj}^{(r-1)})^2}{s_{mkj}^{(r-1)}} \\
&= \arg \min_{\{(i,j)|i,j=1,\ldots,J\}} \left\| (\boldsymbol{\gamma}^{q+r})^{[i]} - (\boldsymbol{\gamma}^{(r-1)})^{[j]} \right\|^2; \\
\mathbf{1.2} \quad (i_{j_2}, j_2) &= \arg \min_{\{(i,j)|i,j=1,\ldots,J; i\neq i_{j_1}, j\neq j_1\}} \left\| (\boldsymbol{\gamma}^{q+r})^{[i]} - (\boldsymbol{\gamma}^{(r-1)})^{[j]} \right\|^2; \\
&\vdots \\
\mathbf{1.J} \quad (i_{j_J}, j_J) &= \arg \min_{\{(i,j)|i,j=1,\ldots,J; i\neq i_{j_n}, j\neq j_n, n=1,\ldots,J-1\}} \left\| (\boldsymbol{\gamma}^{q+r})^{[i]} - (\boldsymbol{\gamma}^{(r-1)})^{[j]} \right\|^2.
\end{aligned}$$

2. Reorder columns of $\boldsymbol{\gamma}^{q+r}$ as $\tilde{\boldsymbol{\gamma}}^{q+r} = [(\boldsymbol{\gamma}^{q+r})^{[i_1]}, \ldots, (\boldsymbol{\gamma}^{q+r})^{[i_{J-1}]}, (\boldsymbol{\gamma}^{q+r})^{[i_J]}]$ and $\boldsymbol{\beta}^{q+r}$ as $\tilde{\boldsymbol{\beta}}^{q+r} = [(\boldsymbol{\beta}^{q+r})^{[i_1]} - (\boldsymbol{\beta}^{q+r})^{[i_J]}, \ldots, (\boldsymbol{\beta}^{q+r})^{[i_{J-1}]} - (\boldsymbol{\beta}^{q+r})^{[i_J]}]$, where we define $(\boldsymbol{\beta}^{q+r})^{[J]} = \mathbf{0}$.
3. Update centers and variances:

$$\begin{aligned}
\gamma_{mkj}^{(r)} &= \frac{q+r-1}{q+r}\gamma_{mkj}^{(r-1)} + \frac{1}{q+r}\tilde{\gamma}_{mkj}^{q+r} \\
s_{mkj}^{(r)} &= \frac{q+r-1}{q+r}s_{mkj}^{(r-1)} + \frac{q+r-1}{q+r}\left(\gamma_{mkj}^{(r-1)} - \gamma_{mkj}^{(r)}\right)^2 + \frac{1}{q+r}\left(\tilde{\gamma}_{mkj}^{q+r} - \gamma_{mkj}^{(r)}\right)^2,
\end{aligned}$$

where $\tilde{\gamma}_{mkj}^{q+r}$ are components of the relabeled matrix $\tilde{\boldsymbol{\gamma}}^{q+r}$. In other words, we first compute $\boldsymbol{\gamma}^{(0)}$ to be regarded as the reference of labeling and use the standardized column distance to identify the ordering of columns of the gamma matrix. At the $(q+r)$th run of samples, columns of $\boldsymbol{\gamma}^{q+r}$ is permuted to the status that is the most marginally similar to $\boldsymbol{\gamma}^{(r-1)}$, and $\boldsymbol{\gamma}^{(r-1)}$ is updated to $\boldsymbol{\gamma}^{(r)}$ by incorporating relabeled $\boldsymbol{\gamma}^{q+r}$. Indices $(i_{j_1}, \ldots, i_{j_J})$ apply to $\boldsymbol{\beta}$ as well. The right-hand side of Figure 1 shows the relabeled trace plots. The proposed relabeling procedure appears to work well.

4. Simulation Study

The simulation study consists of two parts. The first part aims to examine the performance of our proposed RJMCMC method, and the second part focuses on the comparison with other existing approaches.

4.1. Performance of the Proposal Method

Two different RLCA (2) models were simulated. One was a three-class RLCA model with five measured indicators, each indicator with three levels (i.e., $J = 3$, $M = 5$, $K_1 = \cdots = K_5 = 3$). Four independent covariates were generated; two covariates ($z_{i1} \sim$ Bernoulli(0.5), $z_{i2} \sim$ Normal(0, 1)) were associated with Y_{im} given S_i for all m and the other two covariates ($x_{i1} \sim$ Bernoulli(0.5), $x_{i2} \sim$ Normal(0, 1)) were associated with S_i (i.e., $P = L = 2$). The other RLCA

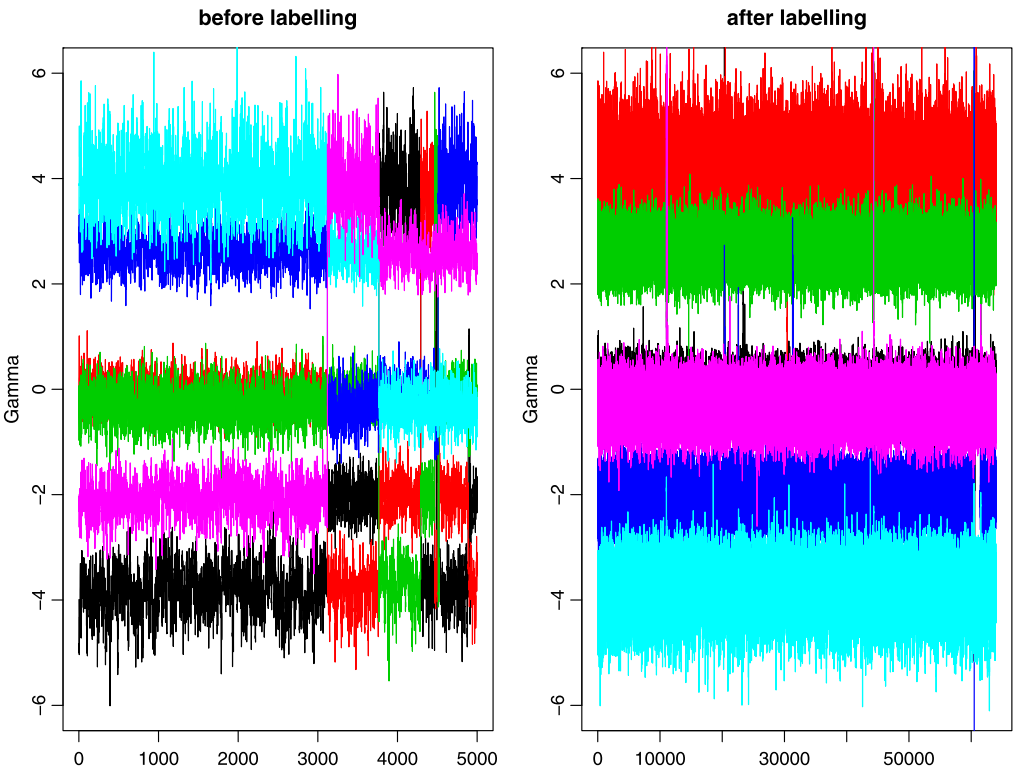

FIGURE 1.
Trace plots for selecting the first q samples to initial the relabelling procedure. The left-hand side of the plot is the trace plot of γ_{521}–γ_{526} at the first 5000 sweeps before labelling. In the figure, one color represents one parameter of γ_{521}–γ_{526}. After adopting the relabelling procedure, the relabeled trace plot is shown in the right-hand side of the plot.

model was a six-class model with all others settings the same as for the three-class model (i.e., $J = 6$, $M = 5$, $K_1 = \cdots = K_5 = 3$, $P = L = 2$). For each model, parameters $\boldsymbol{\beta}$ were selected with the purpose of allocating a similar number of individuals to each class. Each element of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ was determined by randomly choosing from Uniform$(-5, 5)$. For the three-class RLCA, we set the sample size $N = 500$, which gave roughly 10 individuals per parameter of RLCA (2). For the six-class model, $N = 1500$ was selected, which gave roughly 16 individuals per parameter. In the simulation, we set hyperparameters $\sigma_P = 3.0$, $\sigma_{\text{BD}} = \sigma_{\text{SM}} = 0.3$, and $J_{\max} = 30$. Observable measurements Y_i were generated from each of the two RLCA models with 100 replicates, and each replication was run for 100,000 RJMCMC sweeps. Because it was tedious to determine a best burn-in period for each of the 100 replications, to simply the process we examined first several replications and found that the occupancy fraction plots (as will be described in Section S.3.1 of the supplementary information and Section 6.2 of the following real data analysis) became stable roughly after the first 10,000 sweeps. We thus took the first 10,000 sweeps as the burn-in period for each replication. We did not use different initial values for RJMCMC estimation in the simulation study. However, the proposed procedure's sensitivity to initial values is evaluated in Section S.3.4 of the supplementary information. To obtain posterior distributions of parameters for the class number J, in each replication we calculated: the sample mean $\hat{\theta}(J)$, the sample standard deviation $s_{\hat{\theta}}(J)$ and the 2.5th and 97.5th percentiles of the one-dimensional marginal posterior distributions (i.e., a 95 % credible interval) from sweeps with the class number being equal to J.

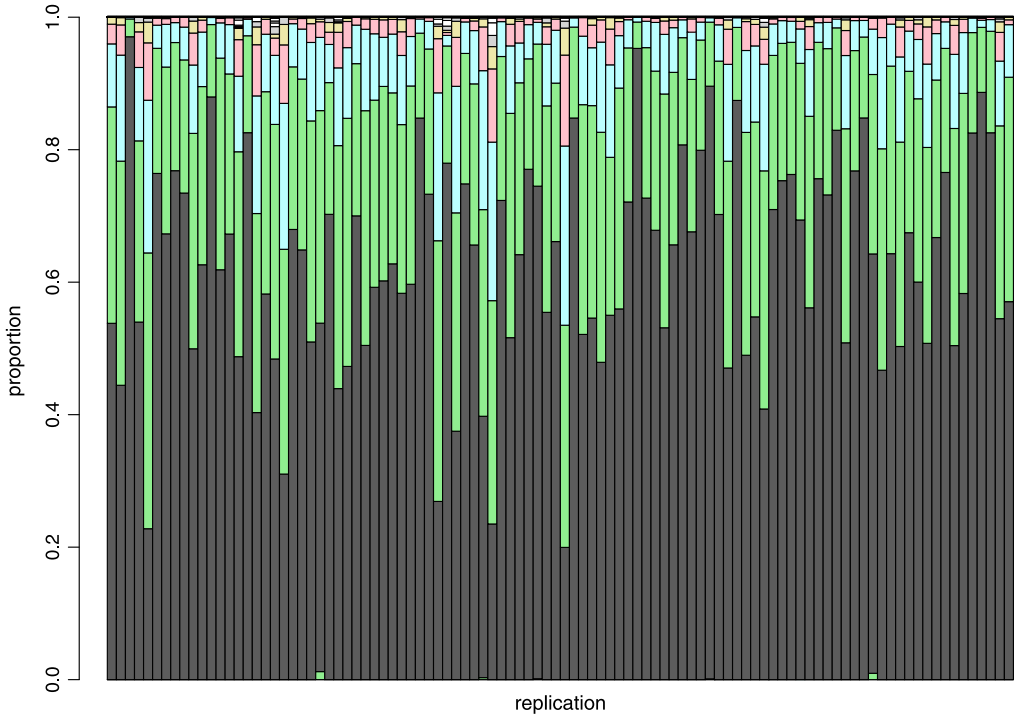

FIGURE 2.
Proportions of sweeps resulting in different values of J for 100 replications with data generated from the three-class RLCA model. The *darkest region* represents the proportion for the true number of class ($J = 3$ in this scenario), and *regions above the darkest* represent proportions of over-estimation ($J > 3$) and *regions below* for under-estimation ($J < 3$). Out of 100 replications, 95 replicates suggested the number of classes to be three.

Figure 2 shows J's marginal posterior distribution for the three-class RLCA model for all replications. Proportions of sweeps resulting in different values of J in each replication are coded in one bar, where the darkest region represents the proportion for the true number of classes ($J = 3$ in this scenario); regions above the darkest represent proportions of over-estimation ($J > 3$) and regions below for under-estimation ($J < 3$). Among 100 replications, 95 give correct class number estimates (i.e., have the largest proportion for $J = 3$), while the other five replications support four classes. Figure 3 displays J's marginal posterior distribution for the six-class RLCA model for 100 replications. There are 84 out of 100 replications correctly estimating the number of classes ($J = 6$). Fourteen replicates show the class number to be five and the remaining two suggest four classes. We also examined the Bayes factors to see how strongly the simulation outputs support the true number of latent classes. This information is presented in Section S.2.1 of the supplementary information. We found that the evidence from most replications supported strongly the true underlying number of classes. We also showed that the best number of classes suggested by Bayes factors was also the mode of the posterior distribution of J, which we used for class number selection in RJMCMC.

Tables 1 and 2 list results of parameter estimation under the true number of classes, averaging over replications whose estimated J (the mode of posterior J) was equal to the true value: the true parameter θ, the average of sample means $\bar{\hat{\theta}}$, the average of sample standard deviations $\bar{s}_{\hat{\theta}}$, and the coverage rate (CR) of the 95 % credible intervals to contain the true parameter value. Table 1, showing the results from $J = 3$, indicates that the RJMCMC gives sensible inferences in this case. The estimated posterior means $\bar{\hat{\theta}}$ are reasonably close to the true values of the

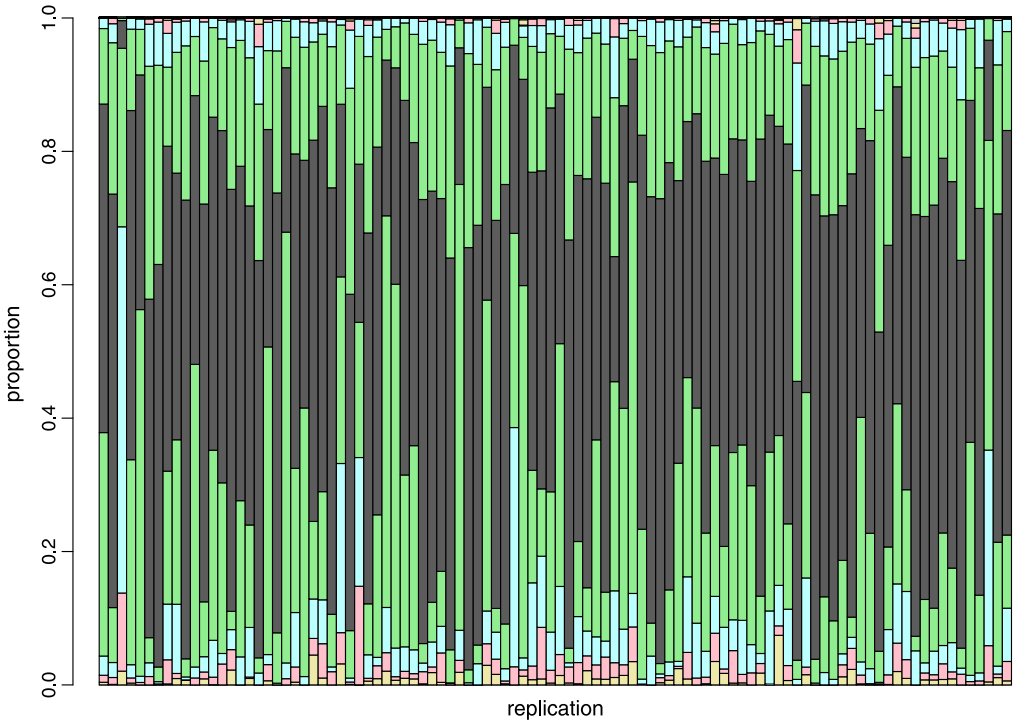

FIGURE 3.
Proportions of sweeps resulting in different values of J for 100 replications with data generated from the six-class RLCA model. The *darkest region* represents the proportion for the true number of class ($J = 6$ in this scenario), and *regions above the darkest* represent proportions of over-estimation ($J > 6$) and *regions below* for under-estimation ($J < 6$). Out of 100 replications, 84 replicates suggested the number of classes to be six, 14 replicates suggest five, and two replicates suggest four.

parameters, and the estimated posterior standard deviations $\bar{s}_{\hat{\theta}}$ are within the interval from 0.23 to 2.00. The coverage rates of 95 % credible intervals range from 88.42 % to 100 %. Table 2 gives the results from $J = 6$. Here, most of the estimated posterior means are close to the true values of the parameters, and the estimated posterior standard deviations range from 0.22 to 1.01, except one that attains 3.06. There are two (out of 95) CRs that are less than 80 %; nevertheless, other CRs are pretty close to 95 %. Although most parameter estimates have small biases and large CRs, some parameters are significantly overestimated with substantially larger standard deviations; for example, α_{142} in Table 1 and $(\beta_{12}, \beta_{04}, \beta_{14})$ in Table 2. After careful examination, we found that these may be caused by the sparseness between response indicators and incorporated covariates, which is often encountered in a given dataset. Detailed explanation can be found in Section S.2.2 of the supplementary information.

Moreover, we ran an additional simulation with a small sample size relative to the number of model parameters. We adopted a three-class RLCA model similar to the one used above but with the sample size $N = 50$. The sample size (50) was less than the number of parameters of RLCA (56), which was similar in size to that in our real data example. The results indicate that only 41 % of the replications give correct class number estimates, which is a much lower correct rate as compared with when the sample size is large (95 %). Given that, we further examined the distribution of empty classes in RJMCMC to justify the class number estimation. The parameter estimation under the true number of classes is reasonably accurate, except for parameters with true values larger than 3 or smaller than -3. These biases in extreme values are probably due to our selection of noninformative prior mean zero of θ, which can dominate the

TABLE 1.
The posterior distributions of model parameters from simulated data with $J = 3$ and $N = 500$ over 100 replications.

	β_{01}	β_{11}	β_{21}	β_{02}	β_{12}	β_{22}	α_{111}	α_{211}	α_{112}	α_{212}
θ	1.97	−2.36	1.38	1.43	−4.59	3.73	1.82	−1.26	4.53	4.29
$\overline{\hat{\theta}}$	1.98	−2.37	1.48	1.65	−4.54	3.64	1.60	−1.62	4.33	4.12
$\overline{s}_{\hat{\theta}}$	0.33	0.38	0.24	0.36	0.52	0.36	0.55	0.44	0.59	0.52
CR	100.00	100.00	100.00	100.00	100.00	97.89	89.47	88.42	97.89	94.74

	α_{121}	α_{221}	α_{122}	α_{222}	α_{131}	α_{231}	α_{132}	α_{232}
θ	−2.11	−3.92	−0.25	−2.49	−0.04	−3.61	−2.11	−1.05
$\overline{\hat{\theta}}$	−2.17	−3.87	−0.19	−2.51	0.04	−3.57	−2.22	−1.13
$\overline{s}_{\hat{\theta}}$	0.48	0.42	0.35	0.31	0.37	0.43	0.46	0.28
CR	97.89	97.89	96.84	97.89	92.63	96.84	90.53	95.79

	α_{141}	α_{241}	α_{142}	α_{242}	α_{151}	α_{251}	α_{152}	α_{252}
θ	3.57	3.08	−3.93	0.93	−2.62	4.90	2.50	2.76
$\overline{\hat{\theta}}$	3.53	2.93	−1.07	0.79	−2.59	4.81	1.99	2.82
$\overline{s}_{\hat{\theta}}$	0.46	0.33	2.00	0.31	0.40	0.44	0.77	0.43
CR	93.68	91.58	91.58	93.68	97.89	93.68	92.63	94.74

	γ_{111}	γ_{112}	γ_{121}	γ_{122}	γ_{131}	γ_{132}	γ_{141}	γ_{142}	γ_{151}	γ_{152}
θ	1.24	3.82	−1.77	−0.20	−3.87	−0.15	3.75	3.86	2.86	−2.41
$\overline{\hat{\theta}}$	1.14	3.89	−1.74	−0.20	−3.91	−0.21	3.84	3.91	2.97	−1.95
$\overline{s}_{\hat{\theta}}$	0.72	0.70	0.43	0.27	0.63	0.23	0.79	0.78	0.41	0.82
CR	93.68	96.84	96.84	95.79	94.74	96.84	95.79	95.79	93.68	94.74

	γ_{211}	γ_{212}	γ_{221}	γ_{222}	γ_{231}	γ_{232}	γ_{241}	γ_{242}	γ_{251}	γ_{252}
θ	1.51	2.13	−2.08	3.17	4.13	−1.76	−3.80	−1.30	2.51	−3.62
$\overline{\hat{\theta}}$	1.68	2.35	−2.59	3.33	4.40	−1.98	−3.83	−1.29	2.56	−3.40
$\overline{s}_{\hat{\theta}}$	0.65	0.66	1.38	0.45	0.59	1.47	0.50	0.32	0.40	0.88
CR	98.95	97.89	98.95	96.84	96.84	100.00	91.58	96.84	94.74	96.84

	γ_{311}	γ_{312}	γ_{321}	γ_{322}	γ_{331}	γ_{332}	γ_{341}	γ_{342}	γ_{351}	γ_{352}
θ	−3.98	−3.63	2.91	−2.96	0.00	−2.88	−3.00	−4.52	2.11	−2.66
$\overline{\hat{\theta}}$	−4.22	−3.45	2.88	−3.10	0.02	−3.09	−2.90	−4.44	2.20	−2.23
$\overline{s}_{\hat{\theta}}$	0.61	0.51	0.45	1.30	0.30	0.88	0.43	1.07	0.37	0.80
CR	94.74	95.79	97.89	98.95	89.47	96.84	91.58	98.95	91.58	94.74

[a] θ: the true parameter.

$\overline{\hat{\theta}}$: the average of posterior sample means over replications whose posterior mode of J was equal to the true value.

$\overline{s}_{\hat{\theta}}$: the average of posterior sample standard deviations over replications whose posterior mode of J was equal to the true value.

CR: the coverage rate of the 95 % credible intervals to contain the true parameter value over replications whose posterior mode of J was equal to the true value.

posterior distribution of θ when the sample size is small. Details of this simulation are shown in Section S.2.3 of the supplementary information.

TABLE 2.
The posterior distributions of model parameters from simulated data with $J = 6$ and $N = 1500$ for 100 replications.

	β_{01}	β_{11}	β_{21}	β_{02}	β_{12}	β_{22}	β_{03}	β_{13}	β_{23}
θ[a]	0.76	−2.91	−1.30	−1.71	−4.52	4.10	−0.42	1.19	2.12
$\overline{\hat{\theta}}$	0.90	−3.54	−1.77	−1.64	−2.53	3.26	−0.34	1.26	1.64
$\overline{s}_{\hat{\theta}}$	0.34	0.97	0.48	0.56	3.06	0.95	0.43	0.45	0.38
CR	100.00	100.00	80.00	100.00	100.00	98.18	100.00	100.00	94.55
	β_{04}	β_{14}	β_{24}	β_{05}	β_{15}	β_{25}	α_{111}	α_{211}	α_{112}	α_{212}
θ	−3.48	0.24	−4.60	−0.15	1.28	1.20	−4.54	1.10	2.98	−4.06
$\overline{\hat{\theta}}$	−2.26	−0.95	−3.87	−0.16	1.29	0.92	−4.35	1.11	3.12	−4.14
$\overline{s}_{\hat{\theta}}$	0.71	0.62	1.00	0.39	0.50	0.46	0.47	0.24	0.30	0.27
CR	54.55	23.64	96.36	100.00	100.00	94.55	94.55	96.36	90.91	89.09
	α_{121}	α_{221}	α_{122}	α_{222}	α_{131}	α_{231}	α_{132}	α_{232}		
θ	2.12	2.71	−0.60	−3.27	−1.99	2.98	−0.03	−3.30		
$\overline{\hat{\theta}}$	2.02	2.70	−0.66	−3.23	−1.90	2.81	0.02	−3.50		
$\overline{s}_{\hat{\theta}}$	0.23	0.22	0.23	0.27	0.33	0.39	0.25	0.32		
CR	94.55	96.36	89.09	98.18	94.55	94.55	90.91	94.55		
	α_{141}	α_{241}	α_{142}	α_{242}	α_{151}	α_{251}	α_{152}	α_{252}		
θ	−3.09	4.41	−4.65	−2.99	−3.90	2.25	−0.11	−4.75		
$\overline{\hat{\theta}}$	−2.98	4.42	−4.38	−2.89	−3.82	2.13	−0.14	−4.91		
$\overline{s}_{\hat{\theta}}$	0.30	0.32	0.44	0.34	0.41	0.28	0.23	0.35		
CR	94.55	96.36	89.09	94.55	94.55	94.55	96.36	92.73		
	γ_{111}	γ_{112}	γ_{121}	γ_{122}	γ_{131}	γ_{132}	γ_{141}	γ_{142}	γ_{151}	γ_{152}
θ	3.80	1.01	−0.79	−1.06	−1.45	4.72	−3.51	−4.48	−1.95	−0.15
$\overline{\hat{\theta}}$	3.69	0.84	−0.67	−0.94	−1.09	5.15	−3.42	−4.51	−2.02	−0.06
$\overline{s}_{\hat{\theta}}$	0.48	0.48	0.25	0.30	0.71	0.60	0.42	0.75	0.36	0.28
CR	94.55	92.73	92.73	92.73	92.73	89.09	98.18	98.18	94.55	96.36
	γ_{211}	γ_{212}	γ_{221}	γ_{222}	γ_{231}	γ_{232}	γ_{241}	γ_{242}	γ_{251}	γ_{252}
θ	−0.99	0.80	−3.44	−2.20	4.79	1.40	−0.71	4.56	1.57	3.18
$\overline{\hat{\theta}}$	−1.05	0.84	−3.36	−2.16	4.56	1.23	−0.90	4.24	1.64	3.35
$\overline{s}_{\hat{\theta}}$	0.47	0.34	0.39	0.36	0.72	0.73	0.52	0.51	0.51	0.51
CR	92.73	94.55	94.55	94.55	92.73	96.36	94.55	90.91	100.00	96.36
	γ_{311}	γ_{312}	γ_{321}	γ_{322}	γ_{331}	γ_{332}	γ_{341}	γ_{342}	γ_{351}	γ_{352}
θ	−2.55	−0.92	−4.25	3.96	2.74	1.76	3.92	1.79	−0.19	2.91
$\overline{\hat{\theta}}$	−2.63	−0.98	−4.05	4.05	2.84	1.85	3.90	1.66	−0.14	3.01
$\overline{s}_{\hat{\theta}}$	0.40	0.22	0.50	0.37	0.40	0.40	0.39	0.43	0.35	0.33
CR	92.73	96.36	94.55	98.18	96.36	98.18	98.18	96.36	96.36	96.36
	γ_{411}	γ_{412}	γ_{421}	γ_{422}	γ_{431}	γ_{432}	γ_{441}	γ_{442}	γ_{451}	γ_{452}
θ	3.83	2.88	−1.65	−1.90	−4.30	0.37	2.00	−4.16	−4.67	−2.10
$\overline{\hat{\theta}}$	3.69	2.69	−1.56	−1.84	−4.06	0.26	1.94	−4.13	−4.72	−2.02
$\overline{s}_{\hat{\theta}}$	0.60	0.58	0.27	0.32	0.62	0.26	0.31	0.65	0.71	0.34
CR	96.36	94.55	92.73	96.36	90.91	92.73	90.91	94.55	96.36	89.09

TABLE 2.
(*Continued*)

	γ_{511}	γ_{512}	γ_{521}	γ_{522}	γ_{531}	γ_{532}	γ_{541}	γ_{542}	γ_{551}	γ_{552}
θ	−1.37	0.34	2.64	4.31	−3.88	3.14	1.40	−4.60	−4.62	−4.02
$\overline{\hat{\theta}}$	−1.44	0.36	2.52	4.14	−3.42	3.29	1.19	−4.17	−4.40	−4.16
$\overline{s}_{\hat{\theta}}$	0.34	0.23	0.86	0.86	0.67	0.34	0.26	0.69	0.59	0.49
CR	98.18	100.00	100.00	100.00	85.45	94.55	85.45	96.36	100.00	94.55
	γ_{611}	γ_{612}	γ_{621}	γ_{622}	γ_{631}	γ_{632}	γ_{641}	γ_{642}	γ_{651}	γ_{652}
θ	4.08	−1.03	−1.47	−1.26	4.69	0.82	4.08	3.81	−4.87	−0.39
$\overline{\hat{\theta}}$	4.04	−1.06	−1.38	−1.25	4.56	0.58	4.06	3.70	−4.66	−0.32
$\overline{s}_{\hat{\theta}}$	0.60	0.62	0.31	0.36	0.76	0.77	0.61	0.64	1.01	0.35
CR	94.55	100.00	92.73	92.73	98.18	90.91	92.73	92.73	96.36	96.36

[a] θ: the true parameter.

$\overline{\hat{\theta}}$: the average of posterior sample means over replications whose posterior mode of J was equal to the true value.

$\overline{s}_{\hat{\theta}}$: the average of posterior sample standard deviations over replications whose posterior mode of J was equal to the true value.

CR: the coverage rate of the 95 % credible intervals to contain the true parameter value over replications whose posterior mode of J was equal to the true value.

4.2. Comparison with Alternative Approaches

Here we compare our proposed RJMCMC method with traditional frequentist two-stage approaches, where the number of classes is selected first, and the parameter estimation under the selected number of classes is then performed using Huang and Bandeen-Roche (2004) maximum likelihood estimation (MLE). Four approaches for selecting the number of classes are compared: (a) the goodness-of-fit (GOF) approach, where the RLCA model is fitted under different number of classes and the selected class number is the lowest number of classes that yields acceptable fit under the likelihood ratio goodness-of-fit test; (b) the AIC criterion, where the estimated number of classes is fixed at the class number J that minimizes $-2\log L + 2T$, where L is the likelihood and T is the total number of parameters in the RLCA model; (c) the BIC criterion, where the estimated number of classes is fixed at J that minimizes $-2\log L + T\log N$, where N is the number of observations; and (d) Huang's (2005) approach, where a factor-analysis analogical method is used for the class number estimation—a method that does not require repeated model fitting.

The simulated datasets were generated from two different RLCA (2) models. One was a three-class RLCA model with $M = 5, K_1 = \cdots = K_5 = 2, P = L = 1$; and the other was a six-class RLCA model with $M = 5, K_1 = \cdots = K_5 = 2, P = L = 1$. All parameters in both models were randomly generated from Uniform$(-5, 5)$. To avoid sparse response patterns that might invalidate the GOF approach, the covariates associated with conditional probabilities z_{im1}, $m = 1, \ldots, 5$ and the covariate associated with latent prevalences x_{i1} were binary from Bernoulli(0.5). To build up the correlation between z_{im1} and x_{i1}, one-fourth of the individuals had their conditional probability covariate values the same as latent prevalence covariate values, and the other individuals had independent values from two covariates. For the three-class RLCA, the selected sample sizes N were 500 and 1500, which gave 20 and 62 individuals per parameter, respectively. For the six-class RLCA, we set $N = 3000$ and 10,000, which gave 76 and 254 individuals per parameter, respectively. One hundred replications were performed for each generated RLCA model.

TABLE 3.
Comparison among five different methods in estimating the number of classes.

method	true no. of classes = 3 $N = 500$				true no. of classes = 3 $N = 1500$					true no. of classes = 6 $N = 3000$							true no. of classes = 6 $N = 10,000$						
estimated classes number	2	3	4	5	2	3	4	5	6	2	3	4	5	6	7	8	2	3	4	5	6	7	8
GOF	100	0	0	0	0	82	5	0	13	92	8	0	0	0	0	0	92	8	0	0	0	0	0
AIC	17	28	34	21	0	86	12	2	0	0	0	0	23	35	27	15	0	0	0	0	99	1	0
BIC	59	32	9	0	0	95	5	0	0	0	1	33	40	24	2	0	0	0	0	3	97	0	0
Huang	12	88	0	0	4	96	0	0	0	0	96	4	0	0	0	0	0	100	0	0	0	0	0
RJMCMC	0	95	4	1	0	97	2	1	0	0	0	1	9	82	8	0	0	0	0	6	94	0	0

The results for estimating the number of classes are shown in Table 3. Apparently, our RJMCMC approach outperforms other alternative approaches in selecting the correct number of classes, for both $J = 3$ and $J = 6$. It seems that increasing the sample size can improve the accuracy in class number estimation. When the sample size is large (i.e., $(J, N) = (3, 1500)$ and (6, 10,000)), our approach can correctly identify the class number over 97 % and 94 % of replications for $J = 3$ and 6, respectively, which indicates its ability in recovering the true number of classes. Also, under large sample size the selected numbers of classes by RJMCMC (Bayes factor) and BIC are very consistent, which verifies the well known fact that Bayes factor and BIC yield similar results (Kass & Raftery, 1995). RJMCMC and BIC disagree when $(J, N) =$ (3, 500) and (6, 3000). This may be due to having some extremely small classes (details can be found in Section S.2.4 of the supplementary information). Notice that the GOF and Huang (2005) approaches can seriously underestimate the class number when the true class number is 6. One possible explanation is due to the correlation between z_{im1} and x_{i1}, which violates the independence assumption made in Huang (2005).

For simulated data with $J = 3$, the comparison of Bayes posterior estimates from our RJMCMC procedure and MLEs using the Huang and Bandeen-Roche (2004) procedure for model parameters can be found in Tables 4 and 5 for $N = 500$ and $N = 1500$, respectively. Both Bayes posterior estimates and MLEs were calculated under the true class number for each replication and were then averaged over all 100 replications. The MLEs seem to be more biased than the Bayes estimates, especially when the absolute value of the true $\boldsymbol{\gamma}$ is larger than 3. For $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, the standard deviation estimates of the MLEs are typically smaller than the posterior sample standard deviations from RJMCMC. This is not surprising because the MLEs from the two-stage approach do not take the variation of class number selection into account, but the RJMCMC estimates do. Some $\boldsymbol{\gamma}$'s MLEs have extremely large standard deviations, which is probably due to the convergence to boundary solutions in $\boldsymbol{\gamma}$ estimates for some replicates. The comparison of Bayes estimates and MLEs for the cases with $J = 6$ is shown in Tables S.6 and S.7 of the supplementary information, which is similar to the results for $J = 3$. However, due to the much more complex model structure of the six-class model, problems of large biases and standard deviations of MLEs were more serious.

5. Sensitivity Analysis

In the sensitivity analysis, we highlight some specifications of hyperparameters and proposal parameters, and give a detailed discussion of their influence on the posterior distributions of J, class allocation and parameters. We also examine the proposed procedure's sensitivity to initial values, the de-outlier step, and rejection sampling. This de-outlier step was implemented in our simulation studies and in the real data example to exclude extreme RJMCMC samples from the calculation of the posterior distribution, which can greatly reduce the impact due to the disturbance in parameter estimates and class allocation created by the jumping moves. In the simulation presented in Section 4, we begin with the default setting $\sigma_P = 3.0$, $\sigma_{\mathrm{BD}} = 0.3$ and $\sigma_{\mathrm{SM}} = 0.3$. Different settings are specified to illustrate their impacts on the reversible jump procedure. These settings are applied to 100 simulated datasets with true $J = 6$ and $N = 1500$, generated in Section 4.1. Details of the sensitivity analysis and relevant results can be found in Section S.3 of the supplementary information.

In view of the results from sensitivity analysis, we suggest starting with hyperparameter settings $\sigma_P = 3$ and $\sigma_{\mathrm{BD}}/\sigma_P = \sigma_{\mathrm{SM}}/\sigma_P = 0.1$. Then, one can tune the value of σ_P to make sure the posterior distribution of $\boldsymbol{\beta}$ to be unimodal and tune the values of $\sigma_{\mathrm{BD}}/\sigma_P$ and $\sigma_{\mathrm{SM}}/\sigma_P$ to ensure stable and well mixing Figures S.2–S.5 of the supplementary information. We also suggest running several Markov chains in the selected setting but engaging in different initial

TABLE 4.
The average of Bayes posterior means and MLEs of model parameters from simulated data with $J = 3$ and $N = 500$ over 100 replications.

	β_{01}	β_{11}	β_{02}	β_{12}	α_{111}	α_{121}	α_{131}	α_{141}	α_{151}
θ[a]	−1.12	1.89	0.97	0.54	0.53	1.35	4.15	−1.30	2.73
$\overline{\hat{\theta}}_{\text{PM}}$	−1.12	2.03	0.98	0.60	0.48	1.19	4.21	−1.38	2.63
$\overline{s}_{\hat{\theta}_{\text{PM}}}$	0.48	0.49	0.17	0.31	0.44	0.61	0.39	0.39	0.29
CR_{PM}	98.00	93.00	98.00	94.00	95.00	95.00	99.00	95.00	92.00
$\overline{\hat{\theta}}_{\text{MLE}}$	−0.62	1.45	0.71	0.37	0.54	1.17	4.24	−1.32	2.60
$\overline{s}_{\hat{\theta}_{\text{MLE}}}$	0.18	0.26	0.14	0.23	0.33	0.35	0.38	0.29	0.24
CR_{MLE}	37.00	0.00	70.00	0.00	89.00	76.00	95.00	91.00	74.00

	γ_{111}	γ_{121}	γ_{131}	γ_{141}	γ_{151}	γ_{211}	γ_{221}	γ_{231}	γ_{241}	γ_{251}
θ	−1.17	−2.12	−2.27	0.25	4.22	−4.02	2.51	−3.90	3.43	−1.10
$\overline{\hat{\theta}}_{\text{PM}}$	−1.21	−2.14	−2.26	0.24	2.71	−4.28	2.86	−3.95	3.65	−1.02
$\overline{s}_{\hat{\theta}_{\text{PM}}}$	0.42	0.67	0.45	0.36	1.25	0.60	0.56	0.40	0.44	0.20
CR_{PM}	97.00	99.00	94.00	92.00	87.00	95.00	98.00	99.00	92.00	89.00
$\overline{\hat{\theta}}_{\text{MLE}}$	−2.15	−1.07	−2.71	0.95	1.70	−5.79	5.18	−4.06	4.49	−1.15
$\overline{s}_{\hat{\theta}_{\text{MLE}}}$	0.74	0.29	0.37	0.25	0.63	6.54	17.69	0.39	1.98	0.18
CR_{MLE}	52.00	35.00	72.00	48.00	10.00	89.00	54.00	91.00	65.00	81.00

	γ_{311}	γ_{321}	γ_{331}	γ_{341}	γ_{351}
θ	1.69	−4.58	−1.65	−1.40	−3.16
$\overline{\hat{\theta}}_{\text{PM}}$	2.03	−4.64	−1.63	−1.53	−3.13
$\overline{s}_{\hat{\theta}_{\text{PM}}}$	0.56	0.98	0.36	0.42	0.55
CR_{PM}	95.00	96.00	95.00	92.00	93.00
$\overline{\hat{\theta}}_{\text{MLE}}$	1.71	−6.17	−1.66	−1.45	−2.70
$\overline{s}_{\hat{\theta}_{\text{MLE}}}$	0.32	435.99	0.32	0.30	0.30
CR_{MLE}	62.00	79.00	91.00	83.00	54.00

[a]θ: the true parameter.

$\overline{\hat{\theta}}_{\text{PM}}$: the average of posterior sample means over 100 replicates.

$\overline{s}_{\hat{\theta}_{\text{PM}}}$: the average of posterior sample standard deviations over 100 replicates.

CR_{PM}: the coverage rate of the 95 % credible intervals to contain the true parameter value over 100 replicates.

$\overline{\hat{\theta}}_{\text{MLE}}$: the average of MLEs over 100 replicates.

$\overline{s}_{\hat{\theta}_{\text{MLE}}}$: the average of standard deviation estimates over 100 replicates.

CR_{MLE}: the coverage rate of the 95 % confidence intervals to contain the true parameter value over 100 replicates.

values. Keep running these chains until they all converge to the neighborhood of the same value. The proposed de-outlier step can be used to remove extreme posterior estimates of parameters due to the dimension jump. The selection of $c = 1.2$ for $\boldsymbol{\beta}$, $c = 1.2$ for $\boldsymbol{\alpha}$ and $c = 1.3$ for $\boldsymbol{\gamma}$ in rejection sampling provides reasonable balance between accuracy and computational effort.

TABLE 5.
The average of Bayes posterior means and MLEs of model parameters from simulated data with $J = 3$ and $N = 1500$ over 100 replications.

	β_{01}	β_{11}	β_{02}	β_{12}	α_{111}	α_{121}	α_{131}	α_{141}	α_{151}
θ[a]	−1.12	1.89	0.97	0.54	0.53	1.35	4.15	−1.30	2.73
$\overline{\hat{\theta}}_{\text{PM}}$	−1.11	1.96	0.97	0.58	0.53	1.26	4.18	−1.37	2.71
$\overline{s}_{\hat{\theta}_{\text{PM}}}$	0.26	0.27	0.10	0.17	0.24	0.37	0.23	0.22	0.17
CR_{PM}	96.00	94.00	96.00	93.00	97.00	100.00	95.00	94.00	89.00
$\overline{\hat{\theta}}_{\text{MLE}}$	−0.63	1.49	0.76	0.36	0.54	1.24	4.21	−1.25	2.62
$\overline{s}_{\hat{\theta}_{\text{MLE}}}$	0.11	0.15	0.08	0.13	0.18	0.19	0.22	0.16	0.14
CR_{MLE}	31.00	0.00	61.00	0.00	92.00	62.00	94.00	84.00	74.00

	γ_{111}	γ_{121}	γ_{131}	γ_{141}	γ_{151}	γ_{211}	γ_{221}	γ_{231}	γ_{241}	γ_{251}
θ	−1.17	−2.12	−2.27	0.25	4.22	−4.02	2.51	−3.90	3.43	−1.10
$\overline{\hat{\theta}}_{\text{PM}}$	−1.18	−2.14	−2.30	0.24	3.21	−4.08	2.71	−3.94	3.56	−1.07
$\overline{s}_{\hat{\theta}_{\text{PM}}}$	0.22	0.39	0.26	0.21	1.16	0.31	0.29	0.23	0.25	0.12
CR_{PM}	97.00	96.00	94.00	95.00	86.00	96.00	92.00	95.00	97.00	92.00
$\overline{\hat{\theta}}_{\text{MLE}}$	−1.69	−1.25	−2.67	0.79	1.88	−5.27	6.09	−4.04	4.12	−1.17
$\overline{s}_{\hat{\theta}_{\text{MLE}}}$	0.16	0.16	0.22	0.14	0.63	5.21	11.98	0.22	0.48	0.10
CR_{MLE}	48.00	31.00	58.00	44.00	10.00	77.00	62.00	89.00	61.00	78.00

	γ_{311}	γ_{321}	γ_{331}	γ_{341}	γ_{351}
θ	1.69	−4.58	−1.65	−1.40	−3.16
$\overline{\hat{\theta}}_{\text{PM}}$	1.82	−4.70	−1.65	−1.46	−3.22
$\overline{s}_{\hat{\theta}_{\text{PM}}}$	0.28	0.65	0.21	0.23	0.32
CR_{PM}	94.00	97.00	94.00	98.00	94.00
$\overline{\hat{\theta}}_{\text{MLE}}$	1.67	−5.22	−1.71	−1.37	−2.74
$\overline{s}_{\hat{\theta}_{\text{MLE}}}$	0.17	13564.74	0.20	0.16	0.17
CR_{MLE}	62.00	63.00	93.00	78.00	51.00

[a] θ: the true parameter.

$\overline{\hat{\theta}}_{\text{PM}}$: the average of posterior sample means over 100 replicates.

$\overline{s}_{\hat{\theta}_{\text{PM}}}$: the average of posterior sample standard deviations over 100 replicates.

CR_{PM}: the coverage rate of the 95 % credible intervals to contain the true parameter value over 100 replicates.

$\overline{\hat{\theta}}_{\text{MLE}}$: the average of MLEs over 100 replicates.

$\overline{s}_{\hat{\theta}_{\text{MLE}}}$: the average of standard deviation estimates over 100 replicates.

CR_{MLE}: the coverage rate of the 95 % confidence intervals to contain the true parameter value over 100 replicates.

6. Real Data Analysis

To illustrate the proposed RJMCMC method for RLCA, we use data from the Multidimensional Psychopathological Study on Schizophrenia (MPSS) project and the Study on Etiological Factors of Schizophrenia (SEFOS) project. The data are described in detail in Huang, Tsai, Hwu, Chen, Liu, Hua, and Chen (2011). Briefly, the MPSS and SEFOS projects recruited subsided patients of schizophrenia ($N = 225$) from three hospitals in Taiwan, based on the DSM-IV (American Psychiatric Association, 1994) criteria for schizophrenia. In this study, schizophrenia

is characterized by the Positive and Negative Syndrome Scale (PANSS) (Kay, Flszbein, & Opfer, 1987; Cheng, Ho, Chang, Lan, & Hwu, 1996). The purpose of our analysis here is to explore the underlying subgroups of schizophrenia based on the PANSS ratings, and study the relationship between external covariates and obtained patient subgroups.

6.1. Data

In both MPSS and SEFOS projects, schizophrenia symptoms were assessed by the PANSS. The PANSS has 30 items ($M = 30$) and is composed of three subscales: positive (seven symptoms, P1–P7), negative (seven symptoms, N1–N7) and general psychopathology (16 symptoms, G1–G16). Each item was originally rated on a 7-point scale (1 = absent, 7 = extreme), but we reduced the 7-point scale to a binary scale (no symptom and having symptom) to ease the sparseness problem ($K_1 = \cdots = K_{30} = 2$). External covariates considered in this study included demographic variables and one neuropsychological variable. Demographic variables were gender, age at recruitment, onset-age of psychotic symptoms, years of education, and occupation (having versus no occupation). The neuropsychological variable is the sensitivity score of attention based on the Continuous Performance Test (CPT) (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956; Chen, Liu, Chang, Lien, Chang, & Hwu, 1998). The CPT score was transformed into a z-score compared with a control group matched for age, gender and education years (Liu, Hsieh, Huang, Liu, Liu, Hua, Chen, & Hwu, 2006). The z-score was adjusted so that a higher score indicated better performance.

In our application, RLCA was applied to the 30 dichotomized PANSS items. Demographic variables and the z-standardized CPT score were the covariates associated with the underlying latent class through Equation (3). We identified gender and age as covariates incorporated in conditional probabilities in Equation (4). This analysis used the subsample of subjects that had no missing values ($N = 160$). The RLCA model was fitted through the proposed RJMCMC algorithm.

6.2. Analysis Results

In this data analysis, we set $\sigma_P = 1.5$ and $\sigma_{\mathrm{BD}} = \sigma_{\mathrm{SM}} = 0.2$. This setting was selected based on the recommended procedure for selecting hyperparameter and proposal parameter values provided in Section 5. We did not set $\sigma_{\mathrm{BD}}/\sigma_P = \sigma_{\mathrm{SM}}/\sigma_P = 0.1$ because this setting required many more RJMCMC sweeps to have a well mixing J (data are not shown). The posterior distribution of the number of classes is listed in Figure 4, which suggests the appropriate class number to be three. To further justify this class number estimation, we calculated the total number of sweeps with posterior $J = j$ and the proportion of sweeps that contain empty classes among them for $j = 1, \ldots, 9$ (Table 6). The proportion of sweeps with empty classes for $J = 3$ is 0.03 and is very large for $J \geq 4$. This indicates that $J = 3$ is the maximum and meaningful number of nonempty classes obtained in RJMCMC. To examine the stability of Markov chains, we drew a plot in which the x-axis indicated the sweep index, and the y-axis was the occupancy fraction for a specific value of J. Figure 5 is given to show the occupancy fractions for different values of J. We had found that the occupancy fractions for different J values went stable roughly after 150,000 sweeps. Thus, the burn-in period we took was 150,000 sweeps. The algorithm was run for additional 350,000 sweeps for estimating the posterior distribution.

Table 7 contains the association estimation between latent class membership and external covariates. The odds ratios (OR) are obtained by exponential transformation of regression coefficients $\boldsymbol{\beta}$. The same exponential transformation is also applied to the 2.5 % and 97.5 % quartiles of the posterior samples of $\boldsymbol{\beta}$ to obtain the 95 % credible interval (CI) of the corresponding OR. By comparing with the patients from Class 3, we can characterize the other two classes as follows. Patients of Class 1 tended to be younger at onset-age of psychotic symptoms. Patients of Class 2 were more likely to be younger at recruitment and have an occupation.

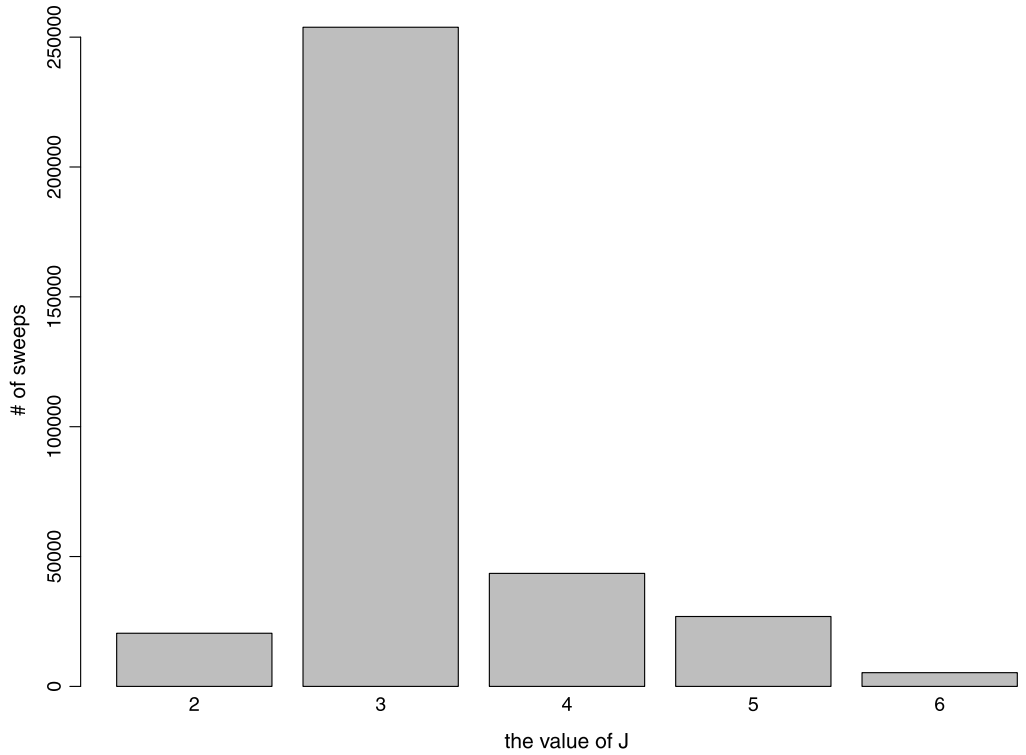

FIGURE 4.
Histogram of posterior distribution of J in analyzing the PANSS rating scale data.

TABLE 6.
The proportions of sweeps that contain empty classes in data from schizophrenia PANSS ratings.

J	total sweeps	proportion of empty classes
2	30917	0.00
3	360779	0.03
4	68907	0.78
5	28778	0.95
6	6093	0.99
7	3303	1.00
8	804	1.00
9	400	0.98

Table 8 displays the direct relationship between PANSS symptom items and incorporated covariates. The OR is the exponential transformation of regression coefficients $\boldsymbol{\alpha}$. Males were more likely to have the G12 symptom than females. The older the age, the higher the probability of having symptoms P4, G5, G6, G7, G8 and G14, but the lower the probability of having N4.

It is worthy to notice that the number of parameters of the RLCA model used to analyze this dataset (164) exceeds number of observations (160). Our RJMCMC suffered from this small sample size and needed a much longer run as compared with analyzing the data generated from the simulation study; nevertheless the algorithm can be stable and obtain satisfactory results.

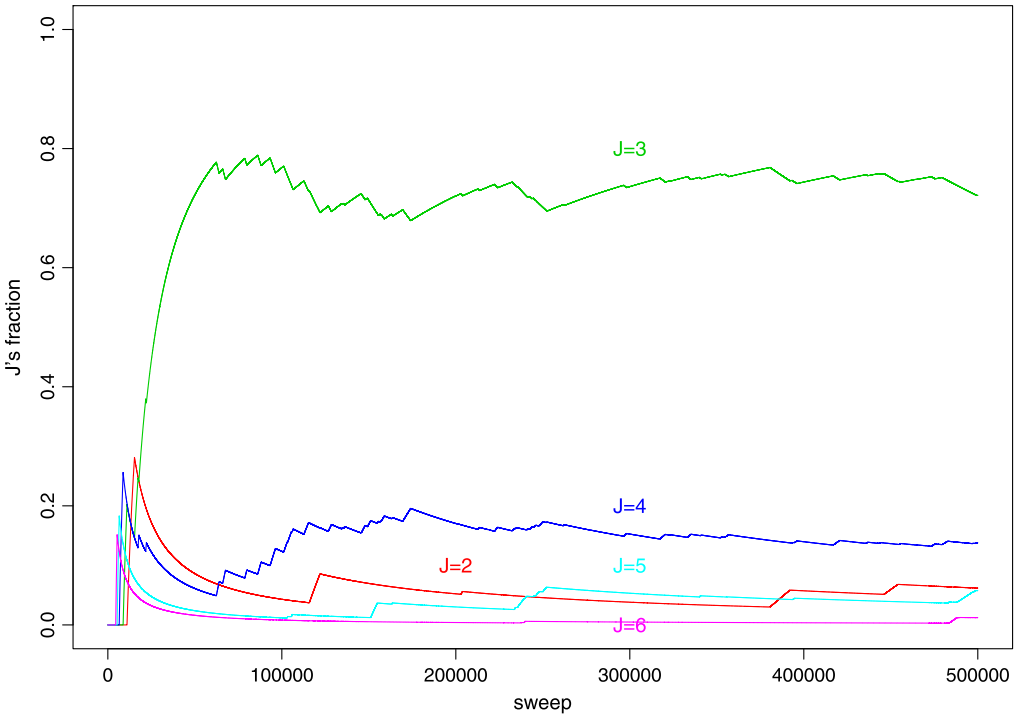

FIGURE 5.
Occupancy fractions for different values of J in analyzing the PANSS rating scale data.

TABLE 7.
Latent prevalence regression from RLCA for the relationship between subtypes underlying PANSS symptoms and external covariates.

Variable	Class 1 vs. Class 3		Class 2 vs. Class 3	
	OR[a]	CI[b]	OR	CI
Male gender	0.41	(0.15, 1.07)	0.97	(0.31, 3.03)
Age	0.95	(0.87, 1.04)	0.90*	(0.82, 0.98)
Age of onset	0.78*	(0.66, 0.92)	0.89	(0.73, 1.06)
Years of education	0.48	(0.16, 1.41)	1.55	(0.40, 5.88)
Having occupation	1.07	(0.96, 1.20)	1.12*	(1.00, 1.27)
Z-standardized CPT	0.78	(0.61, 1.00)	1.21	(0.94, 1.57)

*Values with asterisk are significantly different from 1, judged by CI not covering 1.
[a]OR: odds ratio.
[b]CI: 95 % credible interval of OR.

7. Discussion

We have proposed a Bayesian method for simultaneously estimating the unknown number of classes and parameters of the RLCA model. The proposed approach applies the reversible jump MCMC with Gibbs sampling scheme to analyze the multinomial finite mixture model. To deal with the issue of labeling the latent classes, we have developed a multivariate on-line labeling process. We successfully applied the methodology to simulated datasets and obtained satisfactory results. According to the performance of the proposed RJMCMC on various hyperparameter and

TABLE 8.
Conditional probability regression from RLCA for the direct relationship between PANSS symptoms and external covariates.

	Variable	Male gender		Age	
		OR[a]	CI[b]	OR	CI
P1	Delusis	1.24	(0.60, 2.58)	1.01	(0.97, 1.04)
P2	Conceptual disorganization	1.05	(0.46, 2.37)	1.03	(0.99, 1.07)
P3	Hallucinatory behavior	1.21	(0.58, 2.54)	1.02	(0.99, 1.06)
P4	Excitement	1.09	(0.46, 2.61)	1.04*	(1.00, 1.08)
P5	Grandiosity	1.37	(0.59, 3.25)	1.03	(0.99, 1.07)
P6	Suspiciousness/persecution	1.83	(0.86, 3.95)	1.01	(0.98, 1.05)
P7	Hostility	1.14	(0.48, 2.74)	1.03	(0.99, 1.07)
N1	Blunted affect	0.50	(0.19, 1.19)	0.98	(0.94, 1.02)
N2	Emotional withdrawal	1.19	(0.49, 2.80)	0.99	(0.96, 1.03)
N3	Poor rapport	0.67	(0.27, 1.59)	1.02	(0.98, 1.06)
N4	Passive/apathetic social withdrawal	1.28	(0.54, 3.01)	0.95*	(0.91, 0.99)
N5	Difficulty in abstract thinking	0.96	(0.44, 2.04)	1.00	(0.96, 1.04)
N6	Lack of spontaneity/flow of conversation	0.89	(0.40, 1.94)	1.02	(0.99, 1.06)
N7	Stereotyped thinking	1.85	(0.85, 4.04)	1.02	(0.99, 1.06)
G1	Somatic concern	0.91	(0.45, 1.81)	1.00	(0.97, 1.04)
G2	Anxiety	1.05	(0.52, 2.14)	1.02	(0.98, 1.05)
G3	Guilt fellings	0.42	(0.18, 0.96)	1.01	(0.97, 1.04)
G4	Tension	0.55	(0.25, 1.20)	1.00	(0.97, 1.04)
G5	Mannerisms and posturing	1.27	(0.46, 3.59)	1.08*	(1.03, 1.13)
G6	Depression	1.12	(0.53, 2.37)	1.05*	(1.02, 1.10)
G7	Motor retardation	0.73	(0.32, 1.67)	1.04*	(1.00, 1.08)
G8	Uncooperativeness	1.11	(0.45, 2.73)	1.04*	(1.00, 1.09)
G9	Unusual thought content	1.02	(0.47, 2.22)	1.02	(0.98, 1.06)
G10	Disorientation	0.41	(0.16, 1.01)	1.02	(0.98, 1.06)
G11	Poor attention	1.11	(0.48, 2.59)	1.02	(0.98, 1.06)
G12	Lack of judgment and insight	2.49*	(1.18, 5.38)	0.98	(0.94, 1.01)
G13	Disturbance of volition	1.07	(0.51, 2.21)	1.02	(0.98, 1.06)
G14	Poor impulse control	0.80	(0.35, 1.85)	1.04*	(1.00, 1.08)
G15	Preoccupation	0.76	(0.30, 1.86)	1.02	(0.98, 1.06)
G16	Active social avoidance	0.71	(0.34, 1.46)	1.02	(0.98, 1.05)

*Values with asterisk are significantly different from 1, judged by CI not covering 1.
[a] OR: odds ratio.
[b] CI: 95 % credible interval of OR.

proposal parameter settings, we provided standard default recommendations for the choice of priors. Finally, we analyzed a real dataset whose sample size was small relative to the number of parameters of the fitted RLCA model. Our RJMCMC algorithm demonstrated the ability to handle sparse data. Our algorithm is implemented using C++ and we have made the software available from our web page: http://ghuang.stat.nctu.edu.tw/software.htm.

Recently, many alternative methods have been proposed to deal with the label switching problem. These relabeling algorithms can be deterministic by minimizing a specified loss function (Stephens, 2000; Jasra, Holmes, & Stephens, 2005), be probabilistic by allowing for incorporation of the uncertainty in the relabeling process (Sperrin, Jaki, & Wit, 2010), or be based on the posterior modes and an ascending algorithm (Yao & Lindsay, 2009). These methods offer both on-line and post-processing approaches and are shown to have very good performance under

certain circumstances. Applying these relabeling algorithms to our Bayesian RLCA modelling might improve the goodness-of-fit results.

Models with many parameters sometimes are trapped in the sparseness problem when the sample size is relatively small. To overcome this spareness problem, some constraints on the parameters are required when implementing a traditional EM approach for parameter estimation. These constraints can be arbitrary and are specified by trial and error. In RJMCMC, model parameters are assumed to be from their corresponding prior distribution, which can be viewed as a systematic way for specifying the constraints. A Bayesian approach is much more flexible with the sample size and helps to alleviate this predicament.

The computation time of our RJMCMC procedure depends on the number of sweeps in RJMCMC, the number of parameters in RLCA, and the number of individuals in the dataset. For the three-class RLCA model with 56 parameters and 500 individuals and the six-class RLCA model with 95 parameters and 1500 individuals in the simulation study of Section 4.1, our RJMCMC procedure took on average 3.07 hours and 7.47 hours, respectively, to run 100,000 sweeps for each replication on an Intel Core 2 Quad 2.33 GHz laptop with 4 Gb memory. For the same three-class RLCA model but with a relatively small sample of 50 individuals, the computation time was 0.34 hours. For the three-class RLCA model with 24 parameters and 500 individuals and the six-class RLCA model with 45 parameters and 1500 individuals in the simulation study of Section 4.2, the computation times were 0.96 hours and 3.25 hours, respectively. In real data analysis, we ran a RLCA model with 164 parameters and 160 individuals for 500,000 sweeps, and the computation time was 67.93 hours. In comparison with the EM approach, our RJMCMC procedure took a much longer time to obtain satisfactory results, which was the price paid for our joint inferences on the number of classes and the model parameters.

Acknowledgements

This research was partially supported by grants from the National Science Council, Taiwan (NSC94-2118-M-009-008, NSC95-2118-M-009-002 and NSC101-2118-M-009-004-MY2). The authors wish to thank Dr. Hai-Gwo Hwu for kindly making the schizophrenia syndrome scale data available. We are also grateful to the National Center for High-performance Computing for computer time and facilities.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington: American Psychiatric Press.

Bandeen-Roche, K., Miglioretti, D.L., Zeger, S.L., & Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, *92*, 1375–1386.

Bartolucci, F., Mira, A., & Scaccia, L. (2003). Answering two biological questions with a latent class model via MCMC applied to capture-recapture data. In M.D. Bacco, G. D'Amore, & F. Scalfari (Eds.), *Applied Bayesian statistical studies in biology and medicine* (pp. 7–23). Dordrecht: Springer.

Carlin, B., & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *57*, 473–484.

Celeux, G., Hurn, H., & Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*, 957–970.

Chen, W.J., Liu, S.K., Chang, C.J., Lien, Y.J., Chang, Y.H., & Hwu, H.G. (1998). Sustained attention deficit and schizotypal personality features in nonpsychotic relatives of schizophrenic patients. *The American Journal of Psychiatry*, *155*, 1214–1220.

Cheng, J.J., Ho, H., Chang, C.J., Lan, S.Y., & Hwu, H.G. (1996). Positive and negative syndrome scale (panss): establishment and reliability study of a Mandarin Chinese language version. *Chin Psychiatry*, *10*, 251–258.

Dayton, C.M., & Macready, G.B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*, 173–178.

Dellaportas, P., & Papageorgiou, I. (2006). Multivariate mixtures of normals with unknown number of components. *Statistics and Computing*, *16*, 57–68.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, *87*, 476–486.
Garrett, E.S., & Zeger, S.L. (2000). Latent class model diagnosis. *Biometrics*, *56*, 1055–1067.
Geman, S., & Geman, D. (1993). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Journal of Applied Statistics*, *20*, 25–62.
Geyer, C.J., & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, *21*, 359–373.
Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*, 215–231.
Green, B.F. (1951). A general solution of the latent class model of latent structure analysis and latent profile analysis. *Psychometrika*, *16*, 151–166.
Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.
Gruet, M.A., Philippe, A., & Robert, C.P. (1999). MCMC control spreadsheets for exponential mixture estimation. *Journal of Computational and Graphical Statistics*, *8*, 298–317.
Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *59*, 97–109.
Huang, G.H. (2005). Selecting the number of classes under latent class regression: a factor analytic analogue. *Psychometrika*, *70*, 325–345.
Huang, G.H., & Bandeen-Roche, K. (2004). Building an identifiable latent variable model with covariate effects on underlying and measured variables. *Psychometrika*, *69*, 5–32.
Huang, G.H., Tsai, H.H., Hwu, H.G., Chen, C.H., Liu, C.C., Hua, M.S., & Chen, W.J. (2011). Patient subgroups of schizophrenia based on the positive and negative syndrome scale: composition and transition between acute and subsided disease states. *Comprehensive Psychiatry*, *52*, 469–478.
Jasra, A., Holmes, C.C., & Stephens, D.A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, *20*, 50–67.
Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
Kay, S.R., Flszbein, A., & Opfer, L.A. (1987). The positive and negative syndrome scale (panss) for schizophrenia. *Schizophrenia Bulletin*, *13*, 261–276.
Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. New York: Houghton-Mifflin.
Liu, S.K., Hsieh, M.H., Huang, T.J., Liu, C.M., Liu, C.C., Hua, M.S., Chen, W.J., & Hwu, G.H. (2006). Patterns and clinical correlates of neuropsychologic deficits in patients with schizophrenia. *Journal of the Formosan Medical Association*, *105*, 978–991.
McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.
Meligkotsidou, L. (2007). Bayesian multivariate Poisson mixtures with an unknown number of components. *Statistics and Computing*, *17*, 93–107.
Melton, B., Liang, K.Y., & Pulver, A.E. (2005). Extended latent class approach to the study of familial/sporadic forms of a disease: its application to the study of the heterogeneity of schizophrenia. *Genetic Epidemiology*, *11*, 311–327.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087–1091.
Pandolfi, S., Bartolucci, F., & Friel, N. (2010). A generalization of the multiple-try Metropolis algorithm for Bayesian estimation and model selection. In *Proceedings of the 13th international conference on artificial intelligence and statistics (AISTATS)*, Chia Laguna Resort, Sardinia, Italy (pp. 581–588).
Richardson, S., & Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *59*, 731–792.
Ripley, B.D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *39*, 172–212.
Robert, C.P., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer.
Rosvold, H.E., Mirsky, A.F., Sarason, I., Bransome, E.D. Jr., & Beck, L.H. (1956). A continuous performance test of brain damage. *American Psychological Association*, *20*, 343–350.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.
Sperrin, M., Jaki, T., & Wit, E. (2010). Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics and Computing*, *20*, 357–366.
Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *62*, 795–809.
Viallefont, V., Richardson, S., & Green, P.J. (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics*, *14*, 181–202.
Yao, W., & Lindsay, B.G. (2009). Bayesian mixture labeling by highest posterior density. *Journal of the American Statistical Association*, *104*, 758–767.
Zeger, S.L., & Karim, M.R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, *86*, 79–86.

Manuscript Received: 11 OCT 2011
Published Online Date: 11 DEC 2013