Supplementary Information of Bayesian Inferences of Latent Class Models with an Unknown Number of Classes

S.1 Technical Details in Deriving Acceptance Probabilities for Birth-Death and Split-Merge Moves

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

$$\begin{aligned} A_B &= \frac{p\,(J+1)}{p\,(J)} \times \frac{p(\mathbf{Y}|J+1, \boldsymbol{\gamma}^{birth}, \boldsymbol{\alpha}, \mathbf{S}^{birth}; \mathbf{z}) \times p(\mathbf{S}^{birth}|J+1, \boldsymbol{\beta}^{birth}; \mathbf{x})}{p(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}) \times p(\mathbf{S}|J, \boldsymbol{\beta}; \mathbf{x})} \\ &\quad \times L_{+1} \times \frac{p\left(\boldsymbol{\beta}^{birth}|J+1\right) \times p\left(\boldsymbol{\gamma}^{birth}|J+1\right)}{p\,(\boldsymbol{\beta}|J) \times p\,(\boldsymbol{\gamma}|J)} \\ &\quad \times \frac{1}{h_1(\boldsymbol{\beta}^{[j^*]}) \times h_2(\boldsymbol{\gamma}^{[j^*]})} \times \frac{1}{J} \times \frac{P^{BC}_{alloc}}{P^{CB}_{alloc}} \times Jacobian. \end{aligned} \tag{S.1}$$

In (S.1), J is the number of classes before birth. In the second line of (S.1), L_{+1} arises from the order statistics for the unique class labelling. When the true class number $J \geq 3$, the posterior distribution of $\boldsymbol{\beta}$ is not identifiable for various permutations of the first $J-1$ class labels. Class relabelling involving the reference class, say swap between Class 1 and Class J, can lead to changes in $\boldsymbol{\beta}$'s values from $\{\boldsymbol{\beta}^{[1]}, \boldsymbol{\beta}^{[2]}, \ldots, \boldsymbol{\beta}^{[J-1]}\}$ to $\{-\boldsymbol{\beta}^{[1]}, (\boldsymbol{\beta}^{[2]} - \boldsymbol{\beta}^{[1]}), \ldots, (\boldsymbol{\beta}^{[J-1]} - \boldsymbol{\beta}^{[1]})\}$, which does not change our RLCA likelihood (2), but does change the prior distribution of $\boldsymbol{\beta}$. As a result, the posterior distribution of $\boldsymbol{\beta}$ can alter when switching labels between the reference class and others. However, if the true class number $J = 2$, label switch between Class 1 and Class 2 changes $\boldsymbol{\beta}$ values from $\{\boldsymbol{\beta}^{[1]}\}$ to $\{-\boldsymbol{\beta}^{[1]}\}$, but this time neither the RLCA likelihood (2) nor the prior distribution of $\boldsymbol{\beta}$ becomes different because we adopt a prior distribution symmetric about zero. Therefore,

if $J \geq 3$, our posterior distribution is no identifiable for various permutations of the first $J-1$ class labels. For $J=2$, the posterior distribution is invariant when switching all J class labels. When we adopt a unique labelling for the class membership, there are $(J-1)!$ and $J!$ equivalent ways of writing the joint distribution of the parameters for $J \geq 3$ and $J=2$, respectively. Once a unique labelling is determined, such a procedure is equivalent to ordering the joint prior distribution. So, the joint prior distribution of the parameters is $(J-1)!$ and $J!$ times the product of the individual densities for $J \geq 3$ and $J=2$, respectively. For these reasons,

$$L_{+1} = \begin{cases} \frac{J!}{(J-1)!} & \text{if } J \geq 3 \\ 1 & \text{if } J = 2 \\ 2 & \text{if } J = 1 \end{cases} .$$

In the third line, $h_1(\boldsymbol{\beta}^{[j^*]}) = \prod_{p=0}^{P} h(\beta_{pj^*})$ 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 $N(0, \sigma_{BD}^2)$ density. When a random vector is generated for birth, the proposal from the current state to the birth state is deterministic, that is, inserting the generated vector into the first column of the original matrices. Conversely, there are J possibilities from the birth state to the current state. Moreover, P_{alloc}^{CB} is the probability of making the class reallocation $\mathbf{S}^{birth}$ when moving from the current state to the birth state, and P_{alloc}^{BC} is the reallocation probability for $\mathbf{S}$ when moving from the birth state to the current state. Since reallocation probabilities are calculated as Gibbs allocation of Equation (8) in the manuscript,

$$\frac{P_{alloc}^{BC}}{P_{alloc}^{CB}} = \frac{p(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}) \times p(\mathbf{S}|J, \boldsymbol{\beta}; \mathbf{x})}{p(\mathbf{Y}|J+1, \boldsymbol{\gamma}^{birth}, \boldsymbol{\alpha}, \mathbf{S}^{birth}; \mathbf{z}) \times p(\mathbf{S}^{birth}|J+1, \boldsymbol{\beta}^{birth}; \mathbf{x})},$$

which is the reciprocal of the second term in the first line of expression (S.1). *Jacobian* is the Jacobian arising from the change from $(\boldsymbol{\beta}, \boldsymbol{\beta}^{[j^*]}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^{[j^*]})$ to $(\boldsymbol{\beta}^{birth}, \boldsymbol{\gamma}^{birth})$. Under our birth procedure, *Jacobian* is equal to one. For the corresponding death move, the acceptance probability is $\min\{1, A_D\}$, where

$$\begin{aligned} A_D \;=\; & \frac{p(J-1)}{p(J)} \times L_{-1} \times \frac{p(\boldsymbol{\beta}^{death} \mid J-1) \times p(\boldsymbol{\gamma}^{death} \mid J-1)}{p(\boldsymbol{\beta} \mid J) \times p(\boldsymbol{\gamma} \mid J)} \\ & \times h_1(\boldsymbol{\beta}^{[j^*]}) \times h_2(\boldsymbol{\gamma}^{[j^*]}) \times J - 1, \end{aligned} \tag{S.2}$$

with

$$L_{-1} = \begin{cases} \frac{1}{J-1} & \text{if } J \geq 4 \\ 1 & \text{if } J = 3 \\ 0.5 & \text{if } J = 2 \end{cases}.$$

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

$$\begin{aligned} A_S &= \frac{p\,(J+1)}{p\,(J)} \times \frac{p(\mathbf{Y}|J+1, \boldsymbol{\gamma}^{split}, \boldsymbol{\alpha}, \mathbf{S}^{split}; \mathbf{z}) \times p(\mathbf{S}^{split}|J+1, \boldsymbol{\beta}^{split}; \mathbf{x})}{p(\mathbf{Y}|J, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{S}; \mathbf{z}) \times p(\mathbf{S}|J, \boldsymbol{\beta}; \mathbf{x})} \\ &\times L_{+1} \times \frac{p\left(\boldsymbol{\beta}^{split}|J+1\right) \times p\left(\boldsymbol{\gamma}^{split}|J+1\right)}{p\,(\boldsymbol{\beta}|J) \times p\,(\boldsymbol{\gamma}|J)} \\ &\times \frac{1}{h_3(\mathbf{u}) \times h_4(\mathbf{v})} \times \frac{\frac{1}{J+1}}{\frac{1}{J}} \times \frac{1}{w_J(j^*)} \times \frac{P^{SC}_{alloc}}{P^{CS}_{alloc}} \times Jacobian. \end{aligned} \tag{S.3}$$

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 fractions $1/(J+1)$ and $1/J$ respectively represent that each class in the split state has equal probability $1/(J+1)$ to be selected for merging and conversely each class in the current state has probability $1/J$ to be selected for splitting. The weight $w_J(j^*)$ is defined as

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

When $j^* \neq J$ is selected for splitting, $\mathbf{v}$ and $-\mathbf{v}$ that follow the same distribution will create an identical split pair; thus, the probability of moving from the current state to the split state needs to be multiplied by 2. P^{CS}_{alloc} is the reallocation probability for $\mathbf{S}^{split}$ when moving from the current state to the split state, and P^{SC}_{alloc} is the reallocation probability for $\mathbf{S}$ when moving from the split state to the current state. Since reallocation probabilities are calculated as Gibbs allocation of Equation (8) in the manuscript, $P^{SC}_{alloc}/P^{CS}_{alloc}$ is equal to the reciprocal of the second term in the first line of expression (S.3). *Jacobian* of the split move is $(1/2)^{(P+1)+\left(\sum_{m=1}^{M}(K_m-1)\right)}$. Notice that, in the split move, it is necessary to check whether the adjacency condition (14) in the manuscript 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

$$\begin{aligned} A_M &= \frac{p(J-1)}{p(J)} \times L_{-1} \times \frac{p(\boldsymbol{\beta}^{merge} \mid J-1) \times p(\boldsymbol{\gamma}^{merge} \mid J-1)}{p(\boldsymbol{\beta} \mid J) \times p(\boldsymbol{\gamma} \mid J)} \\ &\quad \times h_3(\mathbf{u}) \times h_4(\mathbf{v}) \times \frac{J}{J-1} \times w_{J-1}(j^*) \times 2^{(P+1)+(\sum_{m=1}^{M}(K_m-1))}. \end{aligned} \tag{S.4}$$

S.2 Additional Simulation Study

S.2.1 Bayes Factor to the True Number of Classes

To see how strong our simulation outputs support the true number of latent classes, one can examine the Bayes factors. The Bayes factor

$$B_{10} = \Pr(\mathbf{Y}|J = \text{the true number of classes}) / \Pr(\mathbf{Y}|J = j)$$

is for evaluating $\mathrm{H}_1 : J =$ the true number of classes against $\mathrm{H}_0 : J = j$. Since RJMCMC produces samples from the joint posterior distribution, as suggested by Carlin & Chib (1995), the posterior probability $\Pr(J = j|\mathbf{Y})$ can be estimated as

$$\hat{\Pr}(J = j|\mathbf{Y}) = \frac{\text{number of sweeps with posterior } J = j}{\text{total number of sweeps}},$$

$j = 1, \ldots, J_{max}$. We assume that J is uniformly distributed among $\{1, ..., J_{max}\}$, and thus $\Pr(J = j) = 1/J_{max}$ for all $j = 1, ..., J_{max}$. As a result, our Bayes factor is calculated as $\hat{B}_{10} = \hat{\Pr}(J = \text{the true number of class}|\mathbf{Y})/\hat{\Pr}(J = j|\mathbf{Y})$. Table S.1 shows the distribution of 2×log Bayes factors from 100 replications. The magnitude of $2 \times \log(\hat{B}_{10})$ is discretized according to Kass & Raftery (1995), in which values between 0 and 2 represent that the evidence against H_0 is "not worth more than a bare mention"; between 2 and 6 represent "positive"; between 6 and 10 represent "strong"; and larger than 10 represent "very strong". The evidence from most replications generated under the three-class model suggest positively ($j = 5$), strongly ($j = 6$) or very strongly ($j = 2, 7, 8$ and 9) that the underlying number of classes is three, except for testing $\mathrm{H}_0 : j = 4$. The

evidence from most replications generated under the six-class model suggest positively ($j = 4, 5, 7$ and 8), strongly ($j = 2$ and 3) or very strongly ($j = 9, 10, 11$ and 12) that the underlying number of classes is six, except for the testing $\mathrm{H}_0 : j = 5$.

To use Bayes factors to select the best number of classes, we need to perform a series of hypothesis testings of $H_0 : J = j - 1$ versus $H_1 : J = j$ for $j = 3, 4, 5, \cdots$, and decide the the best number of classes through their corresponding estimated Bayes factors $\hat{B}_{10}^{(j)}, j = 3, 4, 5, \cdots$. As shown above,

$$\hat{B}_{10}^{(j)} = \frac{\text{number of sweeps with posterior } J = j}{\text{number of sweeps with posterior } J = j - 1}.$$

Therefore, the best number of classes suggested by Bayes factors is also the mode of the posterior distribution of J that we use for class number selection in RJMCMC.

S.2.2 Detailed Explanation on Some Extreme Parameter Estimation

Some parameters are significantly overestimated with substantially larger standard deviations in Tables 1 and 2 of the paper. Two most apparent examples are α_{142} in Table 1 and $(\beta_{12}, \beta_{04}, \beta_{14})$ in Table 2. The parameter α_{142} in Table 1 of the paper is for the relationship between the 4th indicator Y_{i4} and the 1st conditional probability covariate z_{i1}. The contingency table of Y_{i4} versus z_{i1} is shown in the following Table S.2. There was no individual that had $Y_{i4} = 2$ and $z_{i1} = 1$, which can cause an unstable parameter estimate of α_{142}. As a result, the bias and standard deviation of $\hat{\alpha}_{142}$ is large. Such sparseness between response indicators and incorporated covariates is often encountered in a given data set.

In Table 2 of the paper, β_{12}, β_{14} and β_{04} are all for the relationship between the class membership S_i and the 1st latent prevalence covariate x_{i1}. We thus display the contingency table of S_i versus x_{i1} in the following Table S.3. Notice that, in Table S.3, the value of S_i is obtained through randomly assigning the ith individual $\in \{1, \cdots, J\}$ with probabilities $\eta_1(\mathbf{x}_i), \cdots, \eta_J(\mathbf{x}_i)$, calculated from the true $\boldsymbol{\beta}$ and $\mathbf{x}_i$. The $\boldsymbol{\beta}$ coefficient

in the parentheses is the parameter directly affected by its corresponding cell count. The overestimation and large standard deviation for β_{12} are caused by the zero cell count. For β_{04} and β_{14}, their corresponding cell counts are not small, and, thus, their estimates are not unstable (i.e., the standard deviations are not large). Their large biases are probably caused by the bias for β_{12}. Very low coverage rates for β_{04} and β_{14} are due to large bias but small standard deviation.

S.2.3 Performance of the Proposed RJMCMC Algorithm on a Relatively Small Sample

To understand the performance of the proposed RJMCMC algorithm on a sample size similar to that in our real data example, we adopted a three-class RLCA model similar the one used in Section 4.1 of the paper, but with the sample size $N = 50$. The sample size (50) was less than the number of parameters of RLCA (56), which had the size similar to that in our real data example.

Among 100 replications, 41 replicates gave correct class number estimates (i.e., had the largest proportion for $J = 3$), 25 replicates supported 4 classes, and 34 replicates supported 2 classes. Given that only 41% of the replications produced correct class number estimates, we examined the distribution of empty classes in RJMCMC (Table S.4) to further justify the class number estimation. We calculated the total number of sweeps with posterior $J = j$ and the proportion of sweeps that contained empty classes among them for $j = 1, \cdots, 9$, averaging over replications whose modes of posterior J's were the same. The patterns of empty class proportions over j are similar among different modes of posterior J's. They are roughly 0.05 when posterior J equals the true class number ($J = 3$), and become increasingly large when $J \geq 4$. For replications with the estimated class number equal to 4, on average 17% of the sweeps have empty classes in the posterior distribution of $J = 4$, which indicates that the real class number is less than 4. For replications with the estimated class number equal to 2, only 6% of the sweeps

contain empty classes in the posterior distribution of $J = 3$, which indicates that the real class number might be larger than 2.

Table S.5 lists results of parameter estimation under the true number of classes, averaging over 100 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. The results indicate that the RJMCMC gives sensible inferences in this case. Most of the estimated posterior means $\bar{\hat{\theta}}$ are reasonably close to true values of parameters, except for some 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 posterior distribution of θ when the sample size is small. Excluding parameters with true values > 3 or < -3, there are 2 CRs that are less than 80%; nevertheless, the other CRs are pretty close to 95%. The estimated posterior standard deviations $\bar{s}_{\hat{\theta}}$ are within the interval from 1.04 to 1.94.

In summary, the small sample size can greatly reduce the likelihood of selecting a correct number of latent classes. However, by examining the patterns of empty class proportions for posterior J equal to various j's, one can reconfirm the class number estimate as the value j above which empty class proportions become large. The small sample size can also increase biases in extreme-valued parameter estimation when adopting a noninformative prior mean zero.

S.2.4 Disagreement between Bayes Factors and BIC in Simulation $(J, N) = (3, 500)$ and $(6, 3, 000)$

Table 3 of the paper shows that the selected numbers of classes by RJMCMC (Bayes factor) and BIC are very consistent when $(J, N) = (3, 1500)$ and (6, 10,000). This verifies the well known fact that Bayes factor and BIC yield similar results under large sample size. However, RJMCMC and BIC disagree when $(J, N) = (3, 500)$ and (6, 3000). Table

S.8 summaries the values of $-2\times$loglikelihood ($-2\log L$), the number of parameters, AIC and BIC for one randomly selected simulated dataset with $(J, N) = (6, 3000)$. We found that $-2\log L$ decreases slowly as $J \geq 5$. This may be due to having some extremely small classes. The expected class sizes for this simulated dataset are 62, 200, 230, 450, 814 and 1244. Classes with small size may not have significant contribution to the likelihood. As a result, BIC is not sensitive enough to reflect the change due to inclusion/deletion of these small classes. When the sample size is increased to $N = 10{,}000$, the sizes for classes are larger and the conclusions of RJMCMC and BIC are similar.

S.3 Sensitivity Analysis

In this section, we highlight some specifications of hyperparameters and proposal parameters, and give a detailed discussion of their influence made by tuning. We also examine the proposed procedure's sensitivity to initial values, the de-outlier step and rejection sampling that will be discussed in Sections S.3.4, S.3.5 and S.3.6, respectively. In the simulation presented in Section 4 of the manuscript, we begin with the default setting $\sigma_P = 3.0$, $\sigma_{BD} = 0.3$ and $\sigma_{SM} = 0.3$. 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 of the manuscript.

S.3.1 Sensitivity of Posterior Distribution of J

Four different settings were specified: scenario (1) $\sigma_P = 4.0$ and other σs being equal to the default, scenario (2) $\sigma_P = 2.0$ and other σs being equal to the default, scenario (3) $\sigma_{BD} = \sigma_{SM} = 0.2$ and other σs equal to the default, and scenario (4) $\sigma_{BD} = \sigma_{SM} = 0.4$ and other σs equal to the default. Figure S.1 displays histograms of J obtained from each sweep for one randomly selected simulated dataset, and suggests 6 classes for the default and scenarios (1) and (3) and 5 classes for scenarios (2) and (4). The distribution of the modes of posterior J's over 100 replications from each scenario is shown in Table S.9.

Over 100 replications, the default and scenarios (1) and (3) generally suggest 6 classes, and scenarios (2) and (4) generally suggest 5 classes. The trace of J over sweeps for each scenario in one randomly selected simulated dataset is shown in Figure S.2. In scenarios (1) and (3), jumps to simpler models, especially for $J < 6$, are unlikely. In scenarios (2) and (4), moves increasing J to be larger than 6 are rare. These discrepancies in acceptance probabilities of jumping moves result in different posterior distributions of the number of classes J that we see in Figure S.1 and Table S.9. To examine the speed of convergence to the stationary distribution of J, we draw a plot in which the x-axis indicates the sweep index, and the y-axis is the occupancy fraction for a specific value of J. Figure S.3 shows the occupancy fractions for the number of classes selected by each scenario based on Figure S.1 (i.e., the mode of J's posterior distribution in each scenario). One can observe that the lines for scenarios (1) and (3) still go downward at the end of sweep, which indicates that these two chains are not going long enough to reach their J's static state. From traces of J over sweeps in Figure S.2, scenarios (2) and (4) and the default setting apparently have higher proportions of accepting the jumping moves. This may explain the instability of the chains for scenarios (1) and (3) in Figure S.3, where the lower acceptance probability causes J's states to frequently stay at the same value and more RJMCMC sweeps are needed to have a well mixing J. In fact, we have re-run scenarios (1) and (3) for more iterations. Their occupancy fractions remained unchanged after 400,000 iterations.

To understand the influences when larger σ_P is used, we set $\sigma_P = 10$ and $\sigma_{BD} = \sigma_{SM}$ varying between 0.1 and 0.7. Figure S.4 shows the occupancy fractions for the number of classes selected by each setting. One can observe that large σ_P causes the estimation of J to be highly sensitive to the value of σ_{BD} and σ_{SM}, and also causes the chains unstable under the selected number of classes, which means that more RJMCMC sweeps are needed. Even though the chain has converged when $\sigma_{BD} = \sigma_{SM} = 0.7$, the estimated number of classes is far smaller than the true class number. Therefore,

selecting large variance σ_P represents vague information for priors, but this can result in high computational cost for obtaining correct posterior estimation.

Summarizing the results from Figures S.1–S.4 and Table S.9, we have found that it is the value of ratios σ_{BD}/σ_P and σ_{SM}/σ_P that can influence the posterior distribution of J. The smaller the value of σ_{BD} and σ_{SM} relative to σ_P, the lower the acceptance probability of jumping to smaller J and the longer run the RJMCMC sweeps to achieve stationarity in J. However, too big values of σ_{BD} and σ_{SM} relative to σ_P tend to make the moves to larger J difficult, and make the algorithm converge too early and, thus, underestimate the class number. Brooks, Giudici, & Roberts (2003) and Al-Awadhi, Hurn, & Jennison (2004) also obtained similar results of the impacts of proposal parameters on acceptance probabilities. These two characteristics are generally observed in other simulated datasets, too. Our default thus adopts $\sigma_{BD}/\sigma_P = \sigma_{SM}/\sigma_P \approx 0.1$.

S.3.2 Sensitivity of Posterior Distribution of Class Allocation

Figure S.5 displays the counts of subjects belonging to every class over sweeps with a selected number of classes in one randomly selected simulated dataset. Under each scenario, the mode of J in Figure S.1 is selected for displaying. Classes are colored according to their ordering in the number of contained subjects in every sweep (i.e., in each sweep, the black line is the class with the largest member count, and so on). The figure shows that the stability of class allocation for a fixed J depends on the hyperparameter and the proposal parameter values, with smaller values of σ_{BD}/σ_P and σ_{SM}/σ_P being more stable. This phenomenon is reasonable because higher SD values of proposals (relative to the SD value of parameter priors) can invoke a higher acceptance rate for the jump move, and jumping between different values of J can cause high variation in parameter estimates and, thus, class allocation on the first several sweeps right after the move, which creates break points in the color bands of the figure.

Our algorithm does not preclude empty classes, and they will be included in our count

of J. This may cause concerns if many empty classes persist in our model. Table S.10 shows the average (over 100 replications) proportion of sweeps that contain empty classes for J varying from 3 to 9 under each scenario. We have found persistently existing empty classes in sweeps with $J \geq 7$. Also, a large proportion of sweeps that end up with $J = 7$ or 8 under scenarios (1) and (3) have empty classes. Therefore, while adopting a small value for σ_{BD}/σ_P and σ_{SM}/σ_P helps achieve stability in class allocation for a fixed J, it can cause difficulty in expelling empty classes and, thus, inflates the estimated number of classes due to the count of empty classes. Our default setting ($\sigma_{BD}/\sigma_P = \sigma_{SM}/\sigma_P \approx 0.1$) seems to be a balanced choice.

S.3.3 Sensitivity of Posterior Distributions of Parameters

Posterior distributions of model parameters under scenarios (1)–(4) do not look different (data are not shown). We thus focus on three scenarios that satisfy the specification of $\sigma_{BD}/\sigma_P = \sigma_{SM}/\sigma_P \approx 0.1$ to illustrate how the hyperparameter σ_P influences posterior distributions of parameters. Three settings are: scenario (5) $\sigma_P = 1.5$ and $\sigma_{BD/SM} = 0.18$, scenario (6) $\sigma_P = 4.5$ and $\sigma_{BD/SM} = 0.41$, and the default setting $\sigma_P = 3.0$ and $\sigma_{BD/SM} = 0.3$. The best estimate of the class number is 6 for all three scenarios. Averaging over the replications with the mode of J being 6, Figure S.6 contains the average 97.5th percentile (inverted triangle), the average 2.5th percentile (triangle) and the average mean (cross line) of posterior samples of each parameter. The vertical lines represent the 95% credible intervals of each parameter, and intervals are colored differently for different scenarios. From the figure, we can observe that the spread of the credible intervals for scenario (6) is generally wider than the spread from the default setting; the default setting is generally wider than for scenario (5). The means of parameters under scenario (5) are closer to 0 than the ones under the default and scenario (6). It is not difficult to comprehend these observations. We have set the means of prior distributions to be 0. If the SD values of priors (i.e., σ_P) are small, priors are facilitated

to dominate the posterior likelihood. The smaller the σ_P's value, the narrower the credible intervals of posterior distribution, and the stronger the influence of the prior means on the posterior means (i.e., the posterior means are closer to 0). We also discover that, in our RJMCMC algorithm, the above observation is most apparent in $\boldsymbol{\beta}$, then in $\boldsymbol{\gamma}$ and the least in $\boldsymbol{\alpha}$. Furthermore, the 2nd and the 5th βs' credible intervals are very wide. Closer examination finds that under scenario (6), these two parameters have multimodal posterior distributions. Interestingly, scenarios (5) and (6) do not exhibit different class allocation patterns. These reveal that increasing the value for σ_P may lead to posterior distributions of parameter estimates (especially for $\boldsymbol{\beta}$) with multimodality; and it is the ratio of σ_P to σ_{BD} (or σ_{SM}), not the individual parameter, that is related to class allocation.

S.3.4 Sensitivity to Initial Values

In Figure S.7, we show the occupancy fraction plots from 6 different sets of initial values under the default setting in one randomly selected simulated dataset. They are similar and converge to the neighborhood of the same value. In posterior distributions of other parameters ($\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$), results from 6 different sets of initial values are all similar as well (results are not shown). Notice that, due to the insensitivity to initial values, we can examine occupancy fractions from different initial values to judge whether or not we have a stable solution. If the RJMCMC algorithm attains its stable state, chains from different initial values will have a similar convergent trend in their trace plots. Kinney & Dunson (2007) used similar approaches to make sure their Markov chains reached a stable state.

S.3.5 Sensitivity to De-outlier Steps

Figure S.5 shows high variation in class allocation on certain sweeps. Careful examination reveals that most of these high-variation sweeps show up right after the jumping

moves. Notice that these disturbances in parameter estimates and class allocation created by the jumping moves can greatly affect the estimate of the posterior distribution. To reduce their impact, we suggest a "de-outlier" step to exclude extreme RJMCMC samples from the calculation of the posterior distribution. For each parameter of interest, the de-outlier step removes its corresponding RJMCMC samples that fall out the interval between the 25th percentile − 3×IQR (interquartile range) and the 75th percentile + 3×IQR of all samples. This de-outlier step is implemented in our simulation study presented in Section 4 and in the real data example in Section 6 of the manuscript.

In Figure S.8, we show the posterior distributions (with default setting of proposal parameters and hyperparameters and $J = 6$) of β_{11}, β_{13} and β_{14} before and after applying the de-outlier procedure in one randomly selected simulated dataset. It is apparent that standard deviations, 2.5th percentiles and 97.5th percentiles of posterior distributions are affected by outliers. We also recorded the average percentage of posterior samples deleted in the de-outlier step over the replications with the mode of J being 6 (Table S.11). The percentage for $\boldsymbol{\beta}$ is higher than the one for $\boldsymbol{\gamma}$, and the lowest is for $\boldsymbol{\alpha}$. This is because both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are affected by the dimensional jump but $\boldsymbol{\alpha}$ is not. In addition to the dimensional jump, values of $\boldsymbol{\beta}$ are also affected by the necessary modification when the reference class is selected for splitting. The $\boldsymbol{\alpha}$ can be treated as pure MCMC samples (i.e., not affected by dimensional jump and changing reference). Nevertheless, a small percentage of posterior samples is deleted through the proposed de-outlier procedure.

The proposed de-outlier procedure can remove high-variation sweeps due to jumping moves and yet retain valuable information in constructing the true posterior distribution. To illustrate this, we perform MCMC (without reversible jump steps) with fixed $J = 6$ to generate 100,000 sweeps (after burn-in) from the randomly selected simulated data used above, where the true number of classes is six. Here, we use rejection methods to sample data from full conditional distributions and do not implement the de-outlier procedure. The posterior distributions of β_{11}, β_{13} and β_{14} without reversible jump steps

can be found in bottom panels of Figure S.8. These posterior distributions based on sweeps of no reversible jumps and fixed $J = 6$ are very similar to those adopting the de-outlier procedure, which indicates that our de-outlier procedure successfully excludes outliers resulting from reversible jumps.

S.3.6 Sensitivity to Rejection Sampling

We have used rejection sampling to generate samples from full conditional distributions of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$. The proposal densities are chosen to be multivariate Gaussian distributions with variance being the inverse of the observed fisher information of (9) multiplied by a constant c. The choice of c involves a trade-off between accuracy and computational effort. To study the influence of c, we let c vary between 0.7 and 1.8 when performing RJMCMC in the randomly selected simulated data used in Section S.3.5. Table S.12 presents, under the true number of classes $J = 6$, the average number of rejection sampling iterations needed until the new value of the parameter is accepted, and the percentage of invalid posterior samples (i.e., those posterior, $\boldsymbol{\beta}^*$, e.g., violating $p(\mathbf{S}|J, \boldsymbol{\beta}^*, \boldsymbol{\beta}^{[-j]}; \mathbf{x})p(\boldsymbol{\beta}^*|J) \leq (c^* \cdot g_1(\boldsymbol{\beta}^*))$). Large values of c can ensure valid posterior samples but increase the computational time (i.e., require more iterations in rejection sampling). When $c \leq 1.1$ in $\boldsymbol{\beta}$, $c \leq 1.0$ in $\boldsymbol{\alpha}$ and $c \leq 1.2$ in $\boldsymbol{\gamma}$, we observe at least 1% of invalid posterior samples, and we thus set $c = 1.2$ for $\boldsymbol{\beta}$, $c = 1.2$ for $\boldsymbol{\alpha}$ and $c = 1.3$ for $\boldsymbol{\gamma}$ as our defaults. To save computational time, we did not select larger values for c. In fact, inferences on posterior samples seem not very sensitive to different c's. To visualize this, we plot the densities of posterior β_{11}, β_{13} and β_{14} with settings $c = 1.2$ and $c = 1.8$ (Figure S.9). The summary statistics (mean, standard deviation, 2.5th percentile and 97.5th percentile) have no significant differences.

In our rejection sampling, we use Newton-Raphson (NR) methods to numerically solve MLEs that are used for updating the means and variances of envelope distributions. The average numbers of iterations needed for NR methods to obtain MLE's of $\boldsymbol{\beta}^{[j]}$, $\boldsymbol{\alpha}_m^{[k]}$ and

$\boldsymbol{\gamma}^{[j]}$ are 2.99, 2.71 and 4.14, respectively. The computational cost for updating such means and variances does not seem high.

S.3.7 Recommendation

In view of these results, we suggest starting with hyperparameter settings $\sigma_P = 3$ and $\sigma_{BD}/\sigma_P = \sigma_{SM}/\sigma_P = 0.1$. 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 σ_{BD}/σ_P and σ_{SM}/σ_P to ensure stable and well mixing Figures S.2–S.5. We also suggest running several Markov chains in the selected setting but engaging in different initial 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 and to obtain outlier-excluded estimates. 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.

References

Al-Awadhi, F., Hurn, M., & Jennison, C. (2004). Improving the acceptance rate of reversible jump MCMC proposals. *Statistics & Probability Letters*, *69*, 189–198.

Brooks, S. P., Giudici, P., & Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *65*, 3–55.

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.

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.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kinney, S. K., & Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics*, *63*, 690–698.

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.

Table S.1: The distribution of Bayes factors from 100 replications. The upper part of the table shows the Bayes factors $2\log(\hat{B}_{10})$ for evaluating $\mathrm{H}_1 : J = 3$ against $\mathrm{H}_0 : J = j$. The lower part of the table shows the Bayes factors $2\log(\hat{B}_{10})$ for evaluating $\mathrm{H}_1 : J = 6$ against $\mathrm{H}_0 : J = j$.

	j	< 0	$[0, 2)$	$[2, 6)$	$[6, 10)$	$[10, \infty]$
	2	0	0	0	3	97
	4	5	51	42	2	0
true no. of	5	3	6	69	21	1
classes=3	6	0	3	32	47	18
	7	0	0	7	35	58
	8	0	0	2	14	84
	9	0	0	1	7	92
	2	0	1	10	42	47
	3	1	1	22	55	21
	4	3	3	51	36	7
	5	16	31	42	11	0
true no. of	7	1	32	67	0	0
classes=6	8	0	2	66	29	3
	9	0	0	3	48	49
	10	0	0	0	9	91
	11	0	0	0	3	97
	12	0	0	0	0	100

Table S.2: The contingency table of Y_{i4} versus z_{i1} from simulation dataset with $N = 500$ and $J = 3$ in Section 4.1 of the paper.

contingency table		values of z_{i1}	
		0	1
levels of Y_{i4}	1	62	179
	2	60	0
	3	122	77

Table S.3: The contingency table of x_{i1} versus S_i from simulation dataset with $N = 1500$ and $J = 6$ in Section 4.1 of the paper.

contingency table		values of S_i					
		1	2	3	4	5	6(ref. group)
values of x_{i1}	0	251(β_{01})	157(β_{02})	97(β_{03})	93(β_{04})	77(β_{05})	80
	1	18(β_{11})	0(β_{12})	266(β_{13})	138(β_{14})	236(β_{15})	87

Table S.4: Total numbers of sweeps and proportions of sweeps that contain empty classes, averaging over replications with the same mode of posterior J.

	mode $J = 2$ (# of replicates = 34)		mode $J = 3$ (# of replicates = 41)		mode $J = 4$ (# of replicates = 25)	
J	total sweeps	proportion of empty classes	total sweeps	proportion of empty classes	total sweeps	proportion of empty classes
1	5473.65	0.00	1605.12	0.00	419.24	0.00
2	41326.15	0.00	20827.93	0.01	11249.00	0.00
3	26349.06	0.06	33293.93	0.05	25589.44	0.04
4	15902.56	0.20	24613.00	0.23	30145.36	0.17
5	6841.21	0.41	12101.54	0.45	18942.56	0.39
6	2426.82	0.64	4640.12	0.68	8744.12	0.61
7	746.21	0.81	1533.10	0.82	3008.40	0.79
8	225.24	0.91	565.00	0.93	966.48	0.90
9	130.62	0.97	244.51	0.98	345.68	0.96

Table S.5: The posterior distributions of model parameters from simulated data with $J = 3$ and $N = 50$ over 100 replications.

	β_{01}	β_{11}	β_{21}	β_{02}	β_{12}	β_{22}	α_{111}	α_{211}	α_{112}	α_{212}
θ	-1.65	4.36	-4.87	0.45	-2.71	-4.18	-2.61	-0.32	-1.27	-3.65
$\overline{\hat{\theta}}$	-0.82	1.18	-1.12	0.40	-1.20	-1.48	-1.59	0.06	-0.09	-0.50
$\overline{s}_{\hat{\theta}}$	1.79	1.84	1.62	1.51	1.94	1.63	1.51	1.04	1.28	1.63
CR	97.30	75.68	40.54	100.00	97.30	64.86	94.59	94.59	89.19	62.16

	α_{121}	α_{221}	α_{122}	α_{222}	α_{131}	α_{231}	α_{132}	α_{232}
θ	1.35	-3.58	2.40	0.75	3.80	4.38	4.74	2.24
$\overline{\hat{\theta}}$	1.05	-1.34	1.95	-0.93	0.45	2.28	1.30	1.11
$\overline{s}_{\hat{\theta}}$	1.45	1.60	1.35	1.14	1.96	1.51	1.90	1.43
CR	97.30	72.97	100.00	72.97	56.76	75.68	56.76	94.59

	α_{141}	α_{241}	α_{142}	α_{242}	α_{151}	α_{251}	α_{152}	α_{252}
θ	-2.30	-1.92	-0.66	0.34	-2.39	0.67	4.85	4.16
$\overline{\hat{\theta}}$	-1.43	-2.30	-0.92	0.45	-1.34	0.00	2.14	2.86
$\overline{s}_{\hat{\theta}}$	1.68	1.09	1.19	0.81	1.75	0.89	1.55	1.13
CR	100.00	100.00	97.30	97.30	97.30	89.19	70.27	89.19

	γ_{111}	γ_{112}	γ_{121}	γ_{122}	γ_{131}	γ_{132}	γ_{141}	γ_{142}	γ_{151}	γ_{152}
θ	2.50	2.71	-0.37	-2.99	0.21	0.44	-1.55	-0.61	0.31	-4.54
$\overline{\hat{\theta}}$	1.65	1.52	0.16	-2.18	0.14	0.13	-1.19	-0.13	0.87	-2.43
$\overline{s}_{\hat{\theta}}$	1.64	1.78	1.58	1.84	1.76	1.69	1.74	1.39	1.54	1.77
CR	94.59	97.30	100.00	97.30	97.30	100.00	100.00	91.89	94.59	81.08

	γ_{211}	γ_{212}	γ_{221}	γ_{222}	γ_{231}	γ_{232}	γ_{241}	γ_{242}	γ_{251}	γ_{252}
θ	0.25	2.73	0.43	3.25	2.49	0.44	-0.65	-3.16	2.84	4.62
$\overline{\hat{\theta}}$	-0.74	2.40	-0.74	2.22	2.02	-0.87	-1.13	-2.37	1.15	2.10
$\overline{s}_{\hat{\theta}}$	1.90	1.61	1.94	1.73	1.72	1.91	1.55	1.68	1.91	1.82
CR	100.00	97.30	100.00	94.59	97.30	94.59	97.30	94.59	97.30	91.89

	γ_{311}	γ_{312}	γ_{321}	γ_{322}	γ_{331}	γ_{332}	γ_{341}	γ_{342}	γ_{351}	γ_{352}
θ	-0.04	1.72	1.97	-4.23	-4.09	0.14	-2.05	-3.63	-3.99	0.42
$\overline{\hat{\theta}}$	-0.81	0.43	1.85	-1.78	-1.77	0.53	-2.12	-2.57	-1.00	-0.12
$\overline{s}_{\hat{\theta}}$	1.73	1.49	1.66	1.92	1.96	1.55	1.68	1.72	1.85	1.59
CR	100.00	97.30	100.00	91.89	91.89	100.00	100.00	91.89	75.68	97.30

[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.

Table S.6: The average of Bayes posterior means and MLEs of parameters from simulated data with $J = 6$ and $N = 3000$ over 100 replications.

	β_{01}	β_{11}	β_{02}	β_{12}	β_{03}	β_{13}	β_{04}	β_{14}	β_{05}	β_{15}
θ	4.66	-3.29	-0.93	2.97	0.72	1.95	-0.97	2.79	3.46	-0.80
$\overline{\hat{\theta}}_{\mathrm{PM}}$	3.52	-1.90	-0.65	1.62	0.64	1.41	1.45	0.20	0.01	1.31
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.78	3.49	1.70	1.86	1.17	1.22	1.22	1.36	1.56	1.80
$\mathrm{CR}_{\mathrm{PM}}$	48.00	99.00	98.00	90.00	98.00	93.00	49.00	45.00	33.00	59.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	1.60	-3.24	1.09	-2.42	-2.29	2.64	-1.29	1.21	0.57	-1.01
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.09	0.15	0.10	0.15	0.28	0.29	0.20	0.22	0.10	0.13
$\mathrm{CR}_{\mathrm{MLE}}$	0.00	0.00	4.00	0.00	2.00	0.00	20.00	0.00	0.00	0.00

	α_{111}	α_{121}	α_{131}	α_{141}	α_{151}
θ	-4.53	4.63	0.67	1.56	-0.03
$\overline{\hat{\theta}}_{\mathrm{PM}}$	-5.54	4.61	1.10	1.84	-0.06
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.80	0.50	0.23	0.61	0.24
$\mathrm{CR}_{\mathrm{PM}}$	41.00	97.00	33.00	97.00	98.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	-4.69	4.91	0.72	1.39	-0.02
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.18	0.21	0.09	0.21	0.09
$\mathrm{CR}_{\mathrm{MLE}}$	73.00	60.00	72.00	73.00	92.00

	γ_{111}	γ_{121}	γ_{131}	γ_{141}	γ_{151}	γ_{211}	γ_{221}	γ_{231}	γ_{241}	γ_{251}
θ	2.36	-3.04	-1.64	3.94	-0.86	1.17	-0.41	1.87	2.54	2.25
$\overline{\hat{\theta}}_{\mathrm{PM}}$	3.92	-2.50	-1.46	5.37	-0.58	0.89	0.70	0.95	3.37	3.35
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.65	0.24	0.18	0.88	0.36	1.43	1.44	1.62	1.45	1.55
$\mathrm{CR}_{\mathrm{PM}}$	2.00	38.00	86.00	40.00	48.00	51.00	35.00	88.00	63.00	62.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	2.28	-3.11	-1.97	7.79	-0.94	1.98	-2.19	-1.12	4.51	-0.18
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.15	0.19	0.11	145.98	0.09	0.15	0.18	0.11	5.30	0.10
$\mathrm{CR}_{\mathrm{MLE}}$	56.00	35.00	37.00	78.00	37.00	8.00	4.00	5.00	29.00	6.00

	γ_{311}	γ_{321}	γ_{331}	γ_{341}	γ_{351}	γ_{411}	γ_{421}	γ_{431}	γ_{441}	γ_{451}
θ	-4.96	4.05	-1.40	4.90	-2.58	-4.56	2.53	-4.11	-0.88	2.97
$\overline{\hat{\theta}}_{\mathrm{PM}}$	-3.62	3.86	-1.52	5.18	-1.58	-3.09	2.79	-2.66	1.79	3.97
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.91	0.97	0.55	1.34	1.95	1.27	1.38	1.30	1.19	1.63
$\mathrm{CR}_{\mathrm{PM}}$	74.00	93.00	46.00	93.00	83.00	84.00	53.00	76.00	29.00	63.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	-5.67	4.85	-1.55	12.27	-1.93	-4.92	3.42	-3.34	-0.09	2.76
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	2.13	0.97	0.11	255.47	0.13	162.58	0.56	0.23	0.14	0.20
$\mathrm{CR}_{\mathrm{MLE}}$	74.00	71.00	70.00	97.00	20.00	49.00	56.00	27.00	15.00	50.00

	γ_{511}	γ_{521}	γ_{531}	γ_{541}	γ_{551}	γ_{611}	γ_{621}	γ_{631}	γ_{641}	γ_{651}
θ	-1.42	4.17	-0.65	3.68	3.42	-3.53	-1.95	2.01	2.03	2.69
$\overline{\hat{\theta}}_{\mathrm{PM}}$	-0.45	1.54	-0.50	2.93	3.80	-2.50	-2.53	2.30	1.48	-0.55
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	1.22	1.42	1.38	1.50	1.61	1.51	1.59	1.27	0.83	1.30
$\mathrm{CR}_{\mathrm{PM}}$	30.00	56.00	42.00	79.00	89.00	88.00	72.00	89.00	47.00	21.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	-1.85	3.66	-0.81	10.26	1.90	0.24	0.30	0.81	3.76	2.34
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.19	0.51	0.10	147.22	0.14	0.12	0.14	0.11	9.63	0.15
$\mathrm{CR}_{\mathrm{MLE}}$	26.00	50.00	23.00	71.00	15.00	0.00	0.00	4.00	18.00	28.00

[a] θ: the true parameter.

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

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

$\mathrm{CR}_{\mathrm{PM}}$: the coverage rate of the 95% credible intervals to contain the true parameter value over 100 replicates.

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

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

$\mathrm{CR}_{\mathrm{MLE}}$: the coverage rate of the 95% confidence intervals to contain the true parameter value over 100 replicates.

Table S.7: The average of Bayes posterior means and MLEs of parameters from simulated data with $J = 6$ and $N = 10,000$ over 100 replications.

	β_{01}	β_{11}	β_{02}	β_{12}	β_{03}	β_{13}	β_{04}	β_{14}	β_{05}	β_{15}
θ	4.66	-3.29	-0.93	2.97	0.72	1.95	-0.97	2.79	3.46	-0.80
$\overline{\hat{\theta}}_{\mathrm{PM}}$	3.37	-1.77	-1.60	2.67	0.84	1.65	1.32	0.86	-1.36	2.57
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.42	3.66	2.12	2.35	1.00	1.33	1.24	0.77	1.85	1.97
$\mathrm{CR}_{\mathrm{PM}}$	4.00	100.00	100.00	100.00	100.00	100.00	75.00	29.00	10.00	55.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	4.76	-3.45	-0.81	2.76	0.83	1.79	-1.13	2.89	3.56	-0.95
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.14	0.16	0.24	0.25	0.17	0.18	0.30	0.31	0.14	0.16
$\mathrm{CR}_{\mathrm{MLE}}$	37.00	0.00	59.00	0.00	40.00	0.00	44.00	0.00	33.00	0.00

	α_{111}	α_{121}	α_{131}	α_{141}	α_{151}
θ	-4.53	4.63	0.67	1.56	-0.03
$\overline{\hat{\theta}}_{\mathrm{PM}}$	-5.71	4.65	1.02	1.86	0.02
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.52	0.35	0.14	0.74	0.14
$\mathrm{CR}_{\mathrm{PM}}$	1.00	98.00	12.00	99.00	99.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	-4.55	4.69	0.67	1.57	-0.05
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.09	0.10	0.05	0.12	0.05
$\mathrm{CR}_{\mathrm{MLE}}$	89.00	81.00	94.00	87.00	92.00

	γ_{111}	γ_{121}	γ_{131}	γ_{141}	γ_{151}	γ_{211}	γ_{221}	γ_{231}	γ_{241}	γ_{251}
θ	2.36	-3.04	-1.64	3.94	-0.86	1.17	-0.41	1.87	2.54	2.25
$\overline{\hat{\theta}}_{\mathrm{PM}}$	3.90	-2.53	-1.48	5.30	-0.49	1.77	0.28	0.82	4.13	3.85
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.59	0.17	0.14	0.85	0.22	1.60	1.68	1.75	1.63	1.82
$\mathrm{CR}_{\mathrm{PM}}$	1.00	7.00	86.00	7.00	36.00	60.00	52.00	98.00	66.00	78.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	2.36	-3.09	-1.65	3.99	-0.86	1.24	-0.40	1.92	2.50	2.30
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.07	0.08	0.05	0.14	0.04	0.09	0.08	0.09	0.15	0.10
$\mathrm{CR}_{\mathrm{MLE}}$	83.00	72.00	93.00	91.00	89.00	58.00	65.00	51.00	76.00	56.00

	γ_{311}	γ_{321}	γ_{331}	γ_{341}	γ_{351}	γ_{411}	γ_{421}	γ_{431}	γ_{441}	γ_{451}
θ	-4.96	4.05	-1.40	4.90	-2.58	-4.56	2.53	-4.11	-0.88	2.97
$\overline{\hat{\theta}}_{\mathrm{PM}}$	-4.20	4.25	-1.95	6.41	-2.63	-3.66	4.60	-3.05	0.27	5.64
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	0.61	0.73	0.49	1.44	2.46	0.75	1.52	0.87	1.39	1.75
$\mathrm{CR}_{\mathrm{PM}}$	87.00	100.00	90.00	100.00	98.00	89.00	61.00	90.00	67.00	88.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	-5.43	4.25	-1.40	5.35	-2.57	-5.29	2.63	-4.43	-0.90	3.09
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.52	0.25	0.06	0.71	0.08	292.87	0.19	0.26	0.09	0.16
$\mathrm{CR}_{\mathrm{MLE}}$	47.00	46.00	87.00	52.00	86.00	71.00	75.00	35.00	59.00	62.00

	γ_{511}	γ_{521}	γ_{531}	γ_{541}	γ_{551}	γ_{611}	γ_{621}	γ_{631}	γ_{641}	γ_{651}
θ	-1.42	4.17	-0.65	3.68	3.42	-3.53	-1.95	2.01	2.03	2.69
$\overline{\hat{\theta}}_{\mathrm{PM}}$	1.32	1.14	1.03	3.81	3.96	-3.61	-3.56	2.78	0.95	-1.89
$\overline{s}_{\hat{\theta}_{\mathrm{PM}}}$	1.44	1.61	1.89	1.65	1.74	1.50	1.90	0.94	0.41	1.33
$\mathrm{CR}_{\mathrm{PM}}$	42.00	49.00	60.00	99.00	100.00	99.00	100.00	95.00	21.00	1.00
$\overline{\hat{\theta}}_{\mathrm{MLE}}$	-1.42	4.25	-0.64	3.76	3.47	-3.67	-2.04	2.08	2.51	2.93
$\overline{s}_{\hat{\theta}_{\mathrm{MLE}}}$	0.06	0.18	0.04	0.14	0.09	0.56	0.21	0.23	0.38	0.34
$\mathrm{CR}_{\mathrm{MLE}}$	70.00	44.00	82.00	52.00	55.00	87.00	49.00	40.00	42.00	34.00

[a] θ: the true parameter.

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

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

$\mathrm{CR}_{\mathrm{PM}}$: the coverage rate of the 95% credible intervals to contain the true parameter value over 100 replicates.

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

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

$\mathrm{CR}_{\mathrm{MLE}}$: the coverage rate of the 95% confidence intervals to contain the true parameter value over 100 replicates.

Table S.8: $-2\times$loglikelihood ($-2\log L$), the number of parameters (T), AIC and BIC for one randomly selected simulated dataset with $(J, N) = (6, 3{,}000)$.

J	-2log L	T	AIC	BIC
2	13656.70	17	13690.70	13792.81
3	13420.91	24	13468.91	13613.06
4	13320.94	31	13382.94	13569.13
5	13279.51	38	13355.51	13583.75
6	13252.55	45	13342.55	13612.83
7	13251.58	52	13355.58	13667.91
8	13250.03	59	13368.03	13722.41

Table S.9: The distribution of the modes of posterior J's over 100 replications from each secnario.

	modes of posterior samples of J						
scenarios	2	3	4	5	6	7	8
(1): $\sigma_P = 4, \sigma_{BD/SM} = 0.3$	0	1	0	2	75	22	0
(2): $\sigma_P = 2, \sigma_{BD/SM} = 0.3$	0	1	8	90	1	0	0
default: $\sigma_P = 3, \sigma_{BD/SM} = 0.3$	0	0	2	14	84	0	0
(3): $\sigma_P = 3, \sigma_{BD/SM} = 0.2$	0	0	1	0	83	15	1
(4): $\sigma_P = 3, \sigma_{BD/SM} = 0.4$	4	14	16	58	8	0	0

Table S.10: The average proportion of sweeps that contain empty classes for each scenario from $J = 3$ to 6 over 100 replications.

	scenario (1): $\sigma_P = 4$, $\sigma_{BD/SM} = 0.3$		scenario (2): $\sigma_P = 2$, $\sigma_{BD/SM} = 0.3$		default: $\sigma_P = 3$, $\sigma_{BD/SM} = 0.3$		scenario (3): $\sigma_P = 3$, $\sigma_{BD/SM} = 0.2$		scenario (4): $\sigma_P = 3$, $\sigma_{BD/SM} = 0.4$	
J	total sweeps	proportion of empty class	total sweeps	proportion of empty class	total sweeps	proportion of empty class	total sweeps	proportion of empty class	total sweeps	proportion of empty class
3	1082.17	0.02	8880.05	0.00	1854.45	0.00	236.75	0.00	16273.56	0.00
4	2355.99	0.03	19555.43	0.00	5318.76	0.00	755.53	0.00	23519.37	0.00
5	5983.20	0.05	48783.12	0.00	19828.31	0.00	1605.11	0.01	36840.58	0.00
6	42961.86	0.03	14807.98	0.00	49985.26	0.00	50493.88	0.00	11779.67	0.01
7	29923.13	0.82	3417.17	0.04	17647.28	0.12	32363.20	0.38	2389.35	0.09
8	12836.88	0.96	570.63	0.09	3784.89	0.24	11085.64	0.63	369.53	0.18
9	3273.18	0.99	80.58	0.13	610.90	0.36	2907.85	0.78	50.04	0.25

Table S.11: The average of percentage of posterior samples deleted in the de-outlier steps over the replications with the mode of J being 6.

parameter	percentage (%)
$\boldsymbol{\beta}$	2.6687
$\boldsymbol{\gamma}$	0.2621
$\boldsymbol{\alpha}_1$	0.0047
$\boldsymbol{\alpha}_2$	0.02184
$\boldsymbol{\alpha}_3$	0.005
$\boldsymbol{\alpha}_4$	0.0268
$\boldsymbol{\alpha}_5$	0.0053

Table S.12: The sensitivity to rejection sampling with c varying between 0.7 and 1.8. The columns contain the information about the average number of rejection sampling iterations needed until the new value of the parameter is accepted (# iters) and the percentage of invalid posterior samples (invalid %) for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ under the true number of classes $J = 6$.

	$\boldsymbol{\beta}$		$\boldsymbol{\alpha}$		$\boldsymbol{\gamma}$	
c	# iters	invalid %	# iters	invalid %	# iters	invalid %
0.7	1.00	100.00	1.00	100.00	1.00	99.46
0.8	1.00	100.00	1.00	100.00	1.01	96.97
0.9	1.00	97.36	1.00	99.99	1.06	52.85
1.0	1.04	52.39	1.02	50.81	1.69	11.30
1.1	1.15	5.27	1.10	0.06	2.46	3.68
1.2	1.31	0.87	1.20	0.00	3.63	1.54
1.3	1.48	0.00	1.30	0.00	5.22	0.67
1.4	1.65	0.00	1.40	0.00	7.32	0.36
1.5	1.82	0.00	1.50	0.00	10.09	0.17
1.6	2.01	0.02	1.60	0.00	13.53	0.12
1.7	2.20	0.00	1.70	0.00	18.03	0.06
1.8	2.41	0.00	1.80	0.00	24.05	0.00

Figure Captions

Figure S.1: Histograms of J's posterior distribution under different scenarios.

Figure S.2: The trace of J over 100,000 sweeps for each scenario.

Figure S.3: Convergence plot in which the x-axis indicates the sweep index and the y-axis is the occupancy fraction for the number of classes selected by each scenario. Different colors represent occupancy fractions from different scenarios.

Figure S.4: Convergence plot in which the x-axis indicates the sweep index and the y-axis is the occupancy fraction for the number of classes selected by each setting of σ's. Different colors represent occupancy fractions from different settings.

Figure S.5: The numbers of subjects belonging to every class over sweeps with a select value of J. Under each scenario, the mode of J in Figure S.1 is selected for displaying. Classes are colored according to their ordering in the number of contained subjects in every sweep.

Figure S.6: The average 97.5th percentile (inverted triangle), the average 2.5th percentile (triangle) and the mean (cross line) of posterior samples of each parameter with $J = 6$. The vertical lines represent the average 95% credible intervals of each parameter. The black lines represent the average intervals from scenario (5); the red lines represent the average intervals from the default setting; the green lines represent the intervals from the scenario (6). Here the average is calculated over the replications with the mode of J being 6.

Figure S.7: Occupancy fractions for $\sigma_P = 3$, $\sigma_{BD/SM} = 0.3$ and $J = 6$ (the default scenario of Figure S.3). Here, different colors represent occupancy fractions resulting from different initial values. The red line is the one with the default scenario in Figure S.3.

Figure S.8: The posterior distributions of β_{11}, β_{13} and β_{14} before and after applying the de-outlier procedure. Information in the plot contains the number of values (N), median (M), standard deviation (SD), minimum value (min) and maximum value (max). The two red dash lines represent the 2.5th percentile and 97.5th percentile, respectively, of the posterior distribution of each β.

Figure S.9: The posterior distributions of β_{11}, β_{13} and β_{14} with $c = 1.2$ and $c = 1.8$ in rejection sampling. Information in the plot contains the number of values (N), median (M), standard deviation (SD), minimum value (min) and maximum value (max). The red dash line represents the 2.5th percentile and 97.5th percentile of the posterior distribution of each β.

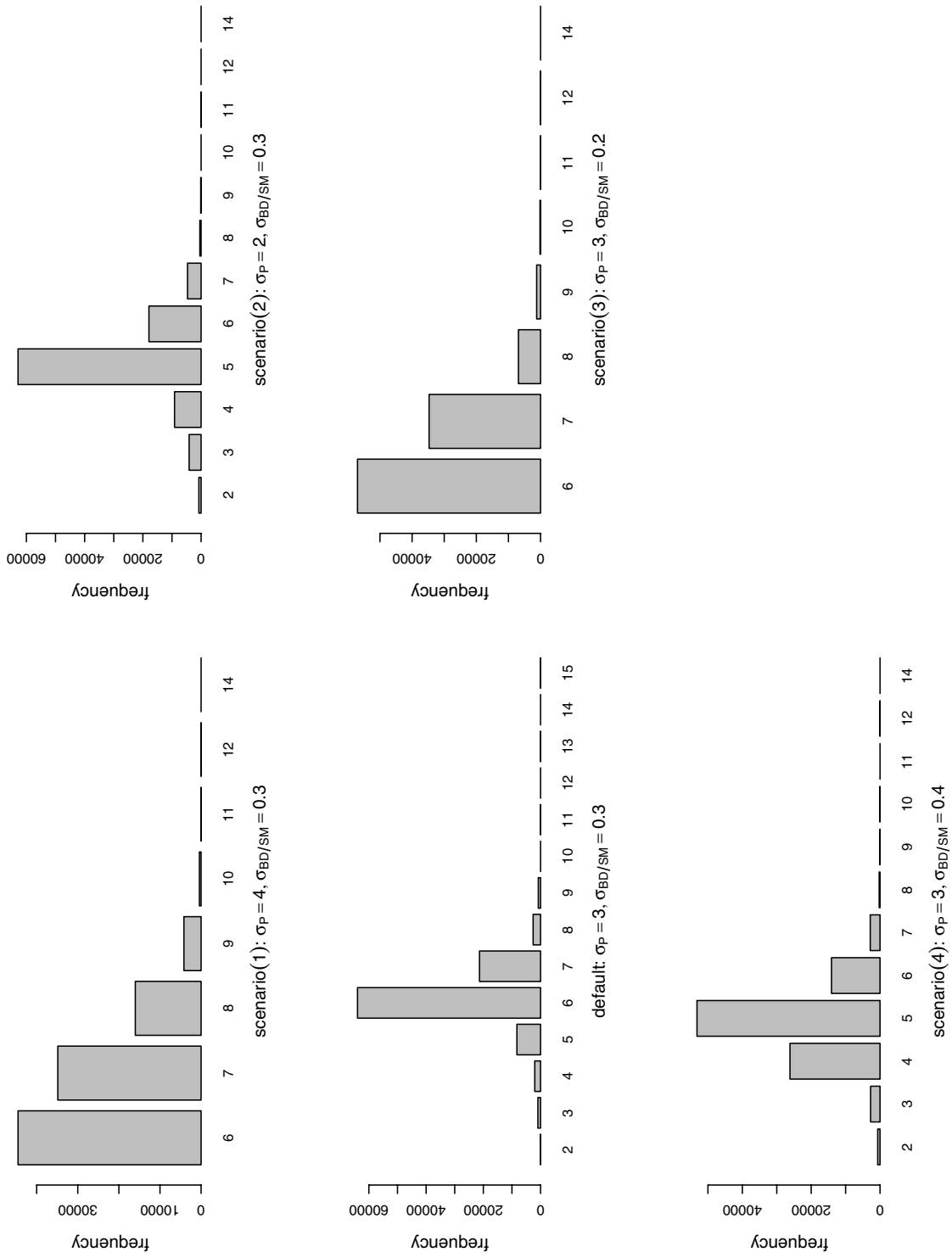
scenario(2): σP = 2, σBD/SM = 0.3
scenario(3): σP = 3, σBD/SM = 0.2
scenario(1): σP = 4, σBD/SM = 0.3
default: σP = 3, σBD/SM = 0.3
scenario(4): σP = 3, σBD/SM = 0.4
frequency

Figure S.1:

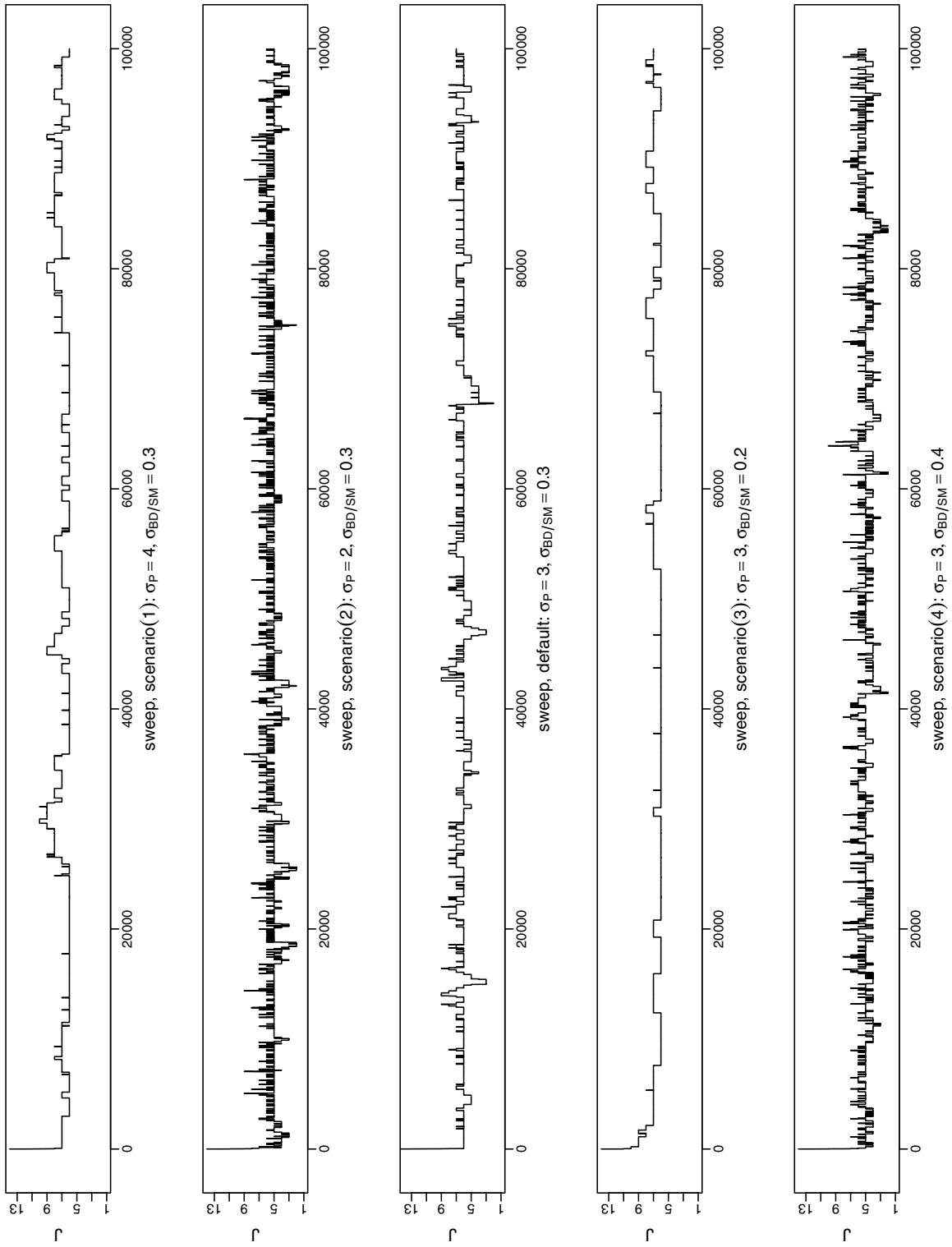
J
1
5
9
13
0
20000
40000
60000
80000
100000
sweep, scenario(1): $\sigma_P = 4$, $\sigma_{BD/SM} = 0.3$
sweep, scenario(2): $\sigma_P = 2$, $\sigma_{BD/SM} = 0.3$
sweep, default: $\sigma_P = 3$, $\sigma_{BD/SM} = 0.3$
sweep, scenario(3): $\sigma_P = 3$, $\sigma_{BD/SM} = 0.2$
sweep, scenario(4): $\sigma_P = 3$, $\sigma_{BD/SM} = 0.4$

Figure S.2:

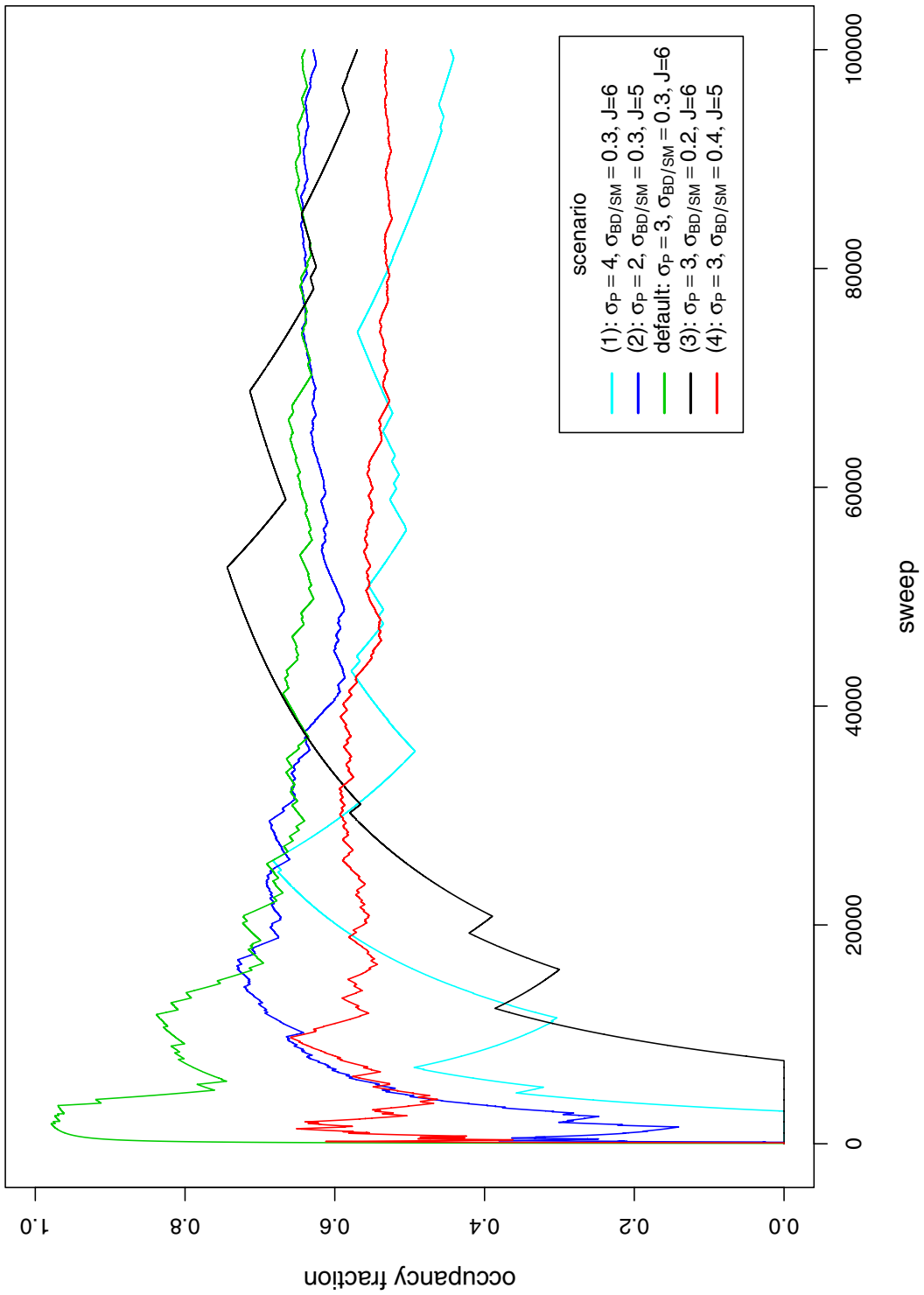
occupancy fraction
0.0
0.2
0.4
0.6
0.8
1.0
sweep
0
20000
40000
60000
80000
100000
scenario
(1): σP = 4, σBD/SM = 0.3, J=6
(2): σP = 2, σBD/SM = 0.3, J=5
default: σP = 3, σBD/SM = 0.3, J=6
(3): σP = 3, σBD/SM = 0.2, J=6
(4): σP = 3, σBD/SM = 0.4, J=5

Figure S.3:

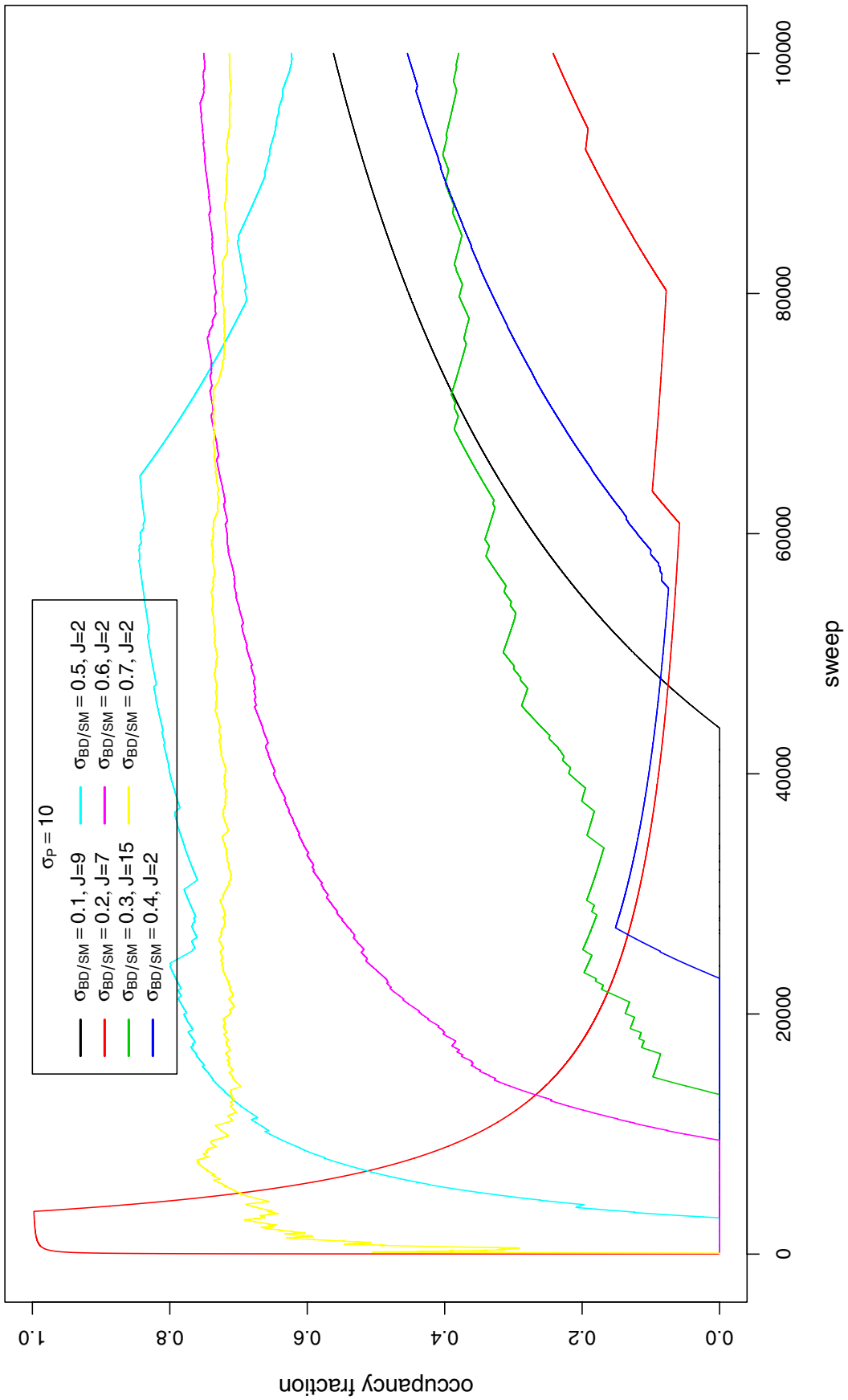
occupancy fraction
0.0
0.2
0.4
0.6
0.8
1.0
sweep
0
20000
40000
60000
80000
100000
σP = 10
σBD/SM = 0.1, J=9
σBD/SM = 0.2, J=7
σBD/SM = 0.3, J=15
σBD/SM = 0.4, J=2
σBD/SM = 0.5, J=2
σBD/SM = 0.6, J=2
σBD/SM = 0.7, J=2

Figure S.4:

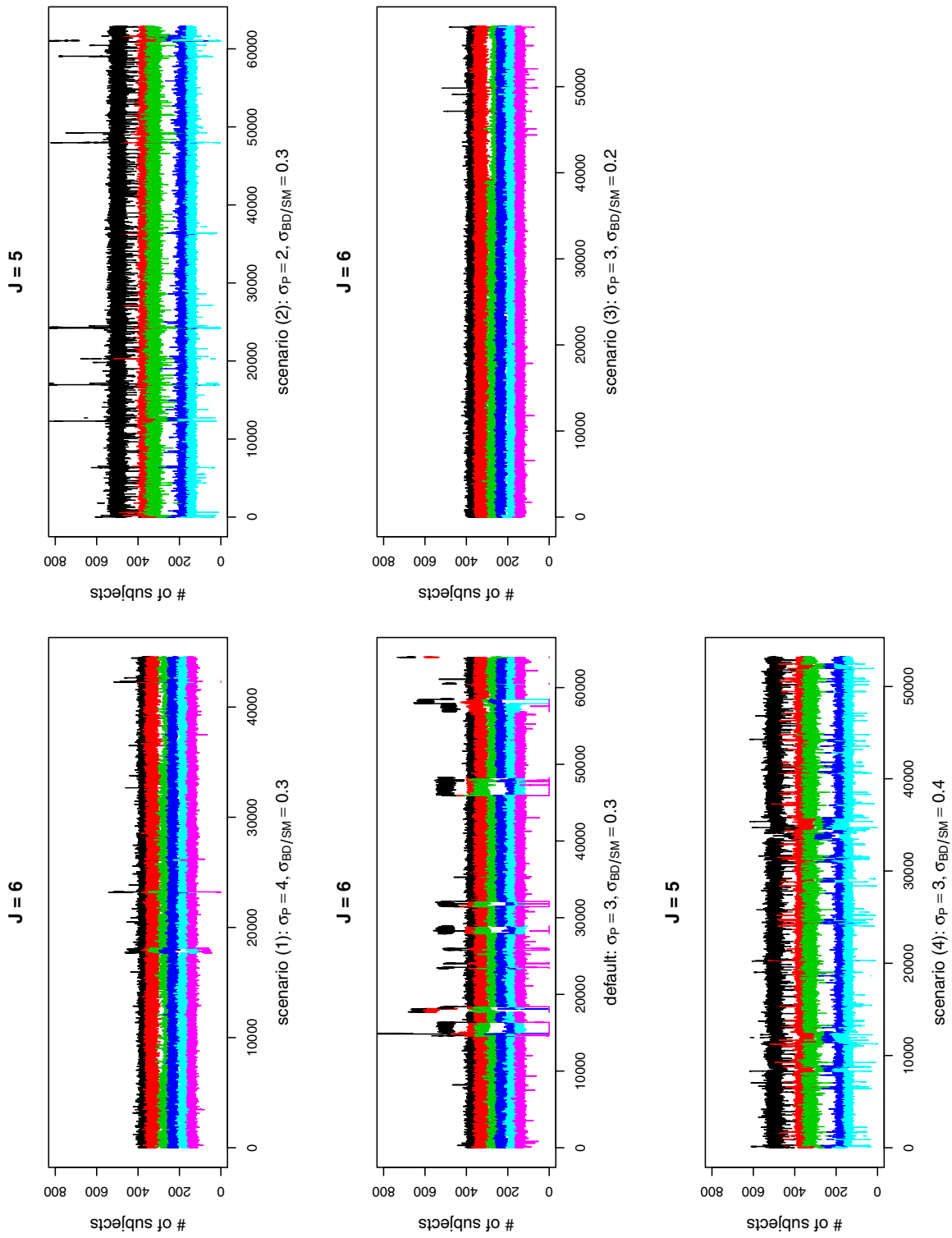
J = 5
J = 6
J = 6
J = 6
J = 5
# of subjects
0
200
400
600
800
0
10000
20000
30000
40000
50000
60000
scenario (1): σP = 4, σBD/SM = 0.3
default: σP = 3, σBD/SM = 0.3
scenario (2): σP = 2, σBD/SM = 0.3
scenario (3): σP = 3, σBD/SM = 0.2
scenario (4): σP = 3, σBD/SM = 0.4

Figure S.5:

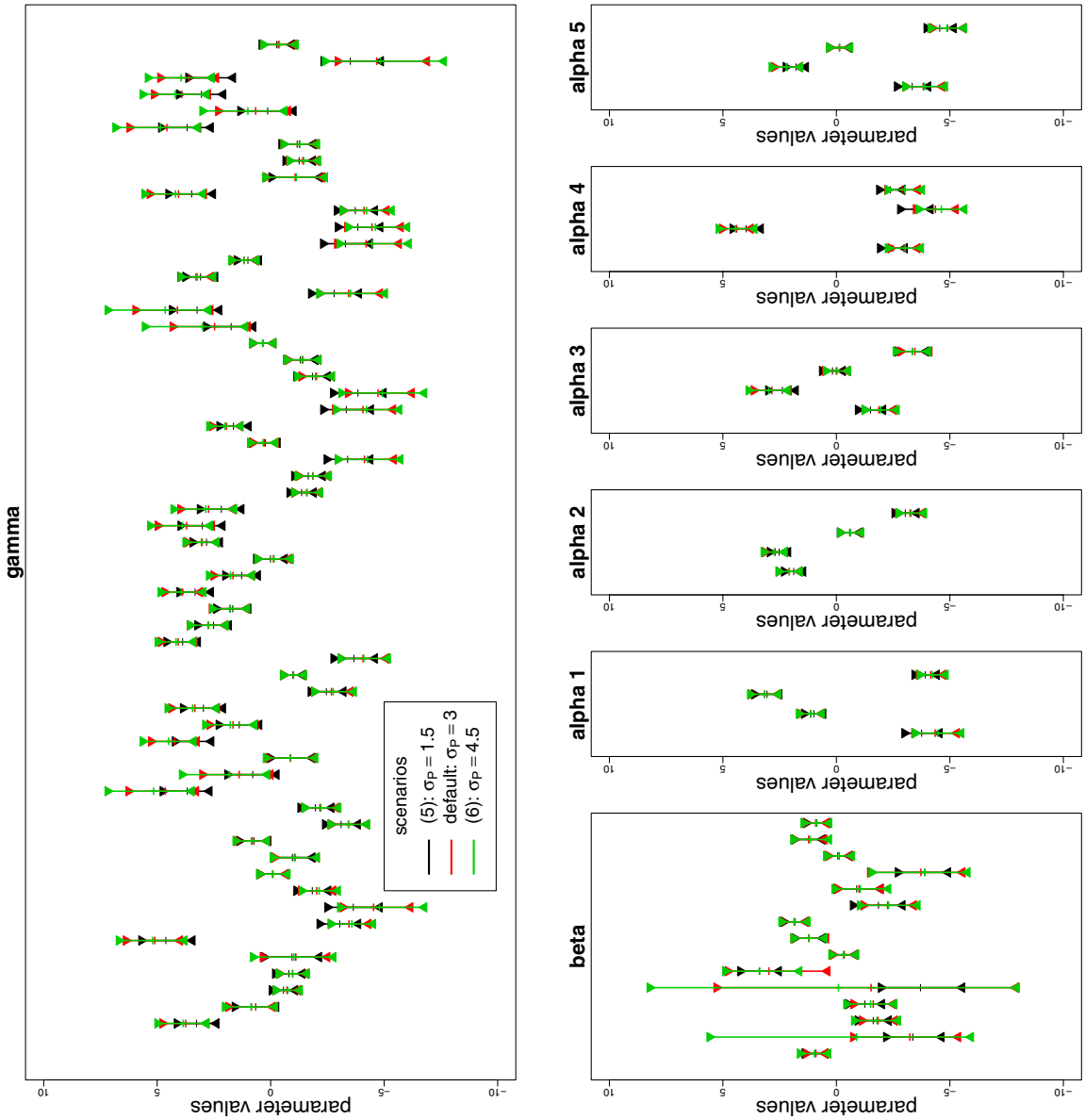
gamma
beta
alpha 1
alpha 2
alpha 3
alpha 4
alpha 5
parameter values
scenarios
(5): σP = 1.5
default: σP = 3
(6): σP = 4.5

Figure S.6:

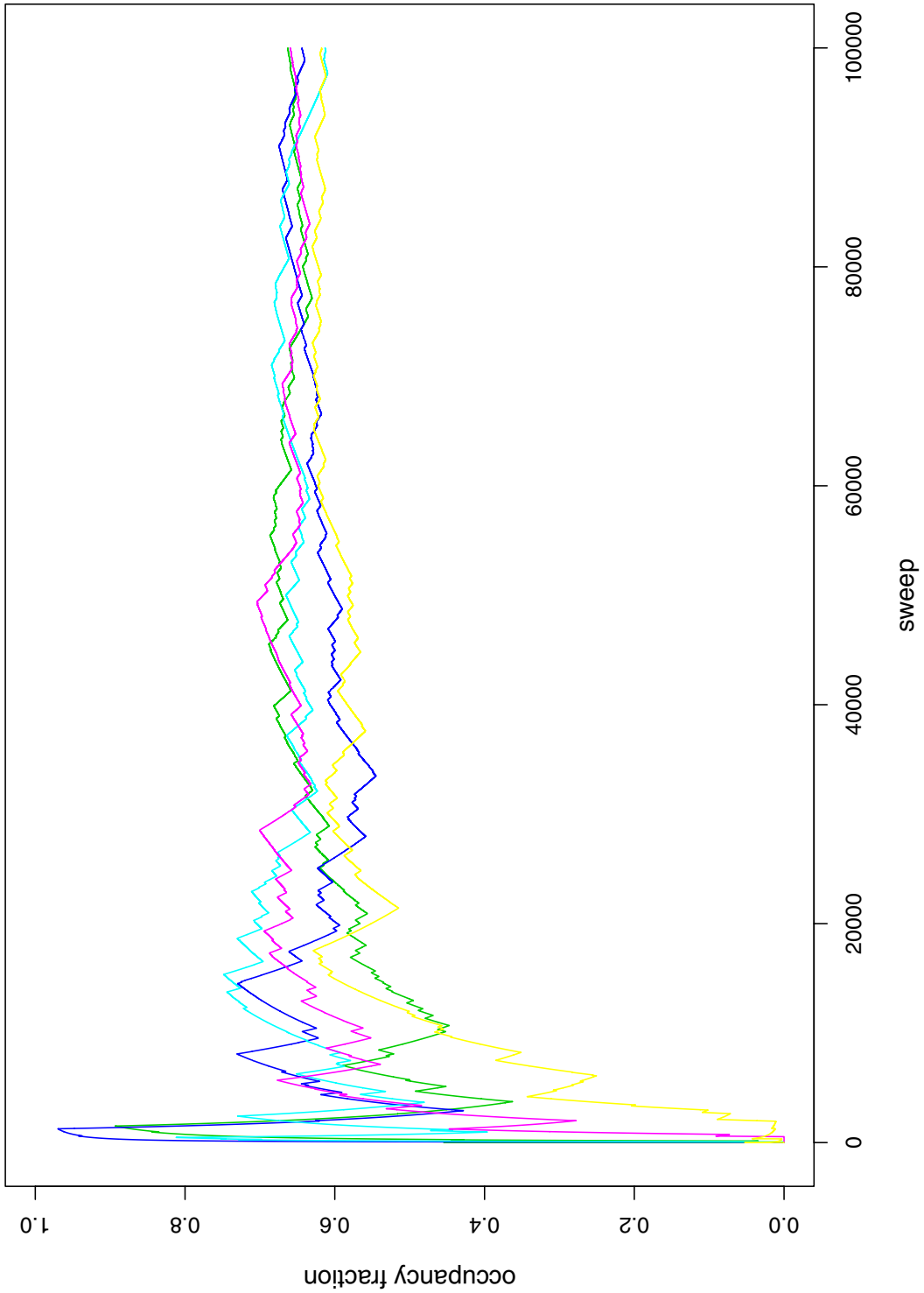
occupancy fraction
0.0
0.2
0.4
0.6
0.8
1.0
sweep
0
20000
40000
60000
80000
100000

Figure S.7:

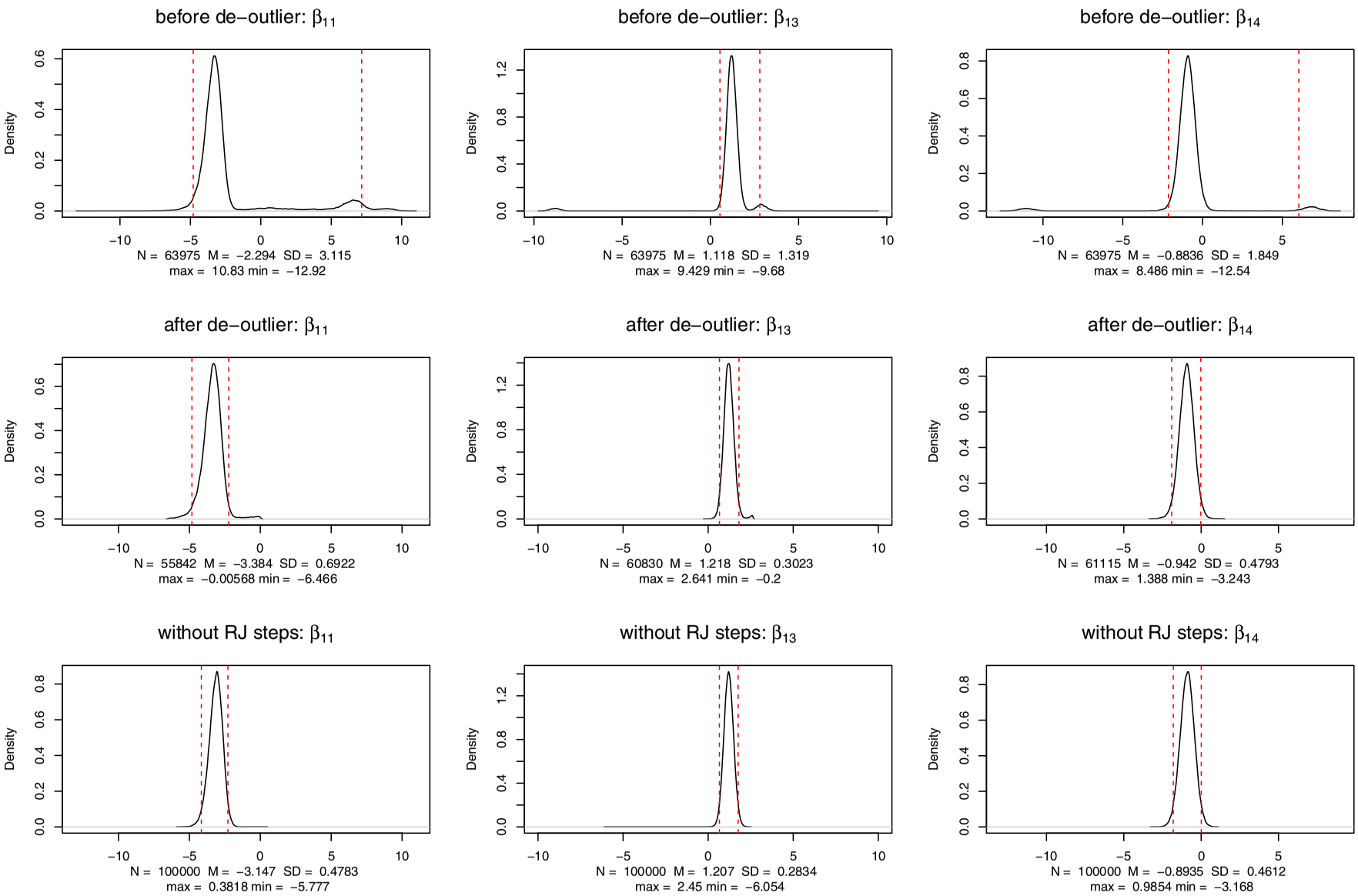
before de−outlier: β_{11}
Density
N = 63975 M = −2.294 SD = 3.115
max = 10.83 min = −12.92
before de−outlier: β_{13}
Density
N = 63975 M = 1.118 SD = 1.319
max = 9.429 min = −9.68
before de−outlier: β_{14}
Density
N = 63975 M = −0.8836 SD = 1.849
max = 8.486 min = −12.54
after de−outlier: β_{11}
Density
N = 55842 M = −3.384 SD = 0.6922
max = −0.00568 min = −6.466
after de−outlier: β_{13}
Density
N = 60830 M = 1.218 SD = 0.3023
max = 2.641 min = −0.2
after de−outlier: β_{14}
Density
N = 61115 M = −0.942 SD = 0.4793
max = 1.388 min = −3.243
without RJ steps: β_{11}
Density
N = 100000 M = −3.147 SD = 0.4783
max = 0.3818 min = −5.777
without RJ steps: β_{13}
Density
N = 100000 M = 1.207 SD = 0.2834
max = 2.45 min = −6.054
without RJ steps: β_{14}
Density
N = 100000 M = −0.8935 SD = 0.4612
max = 0.9854 min = −3.168

Figure S.8:

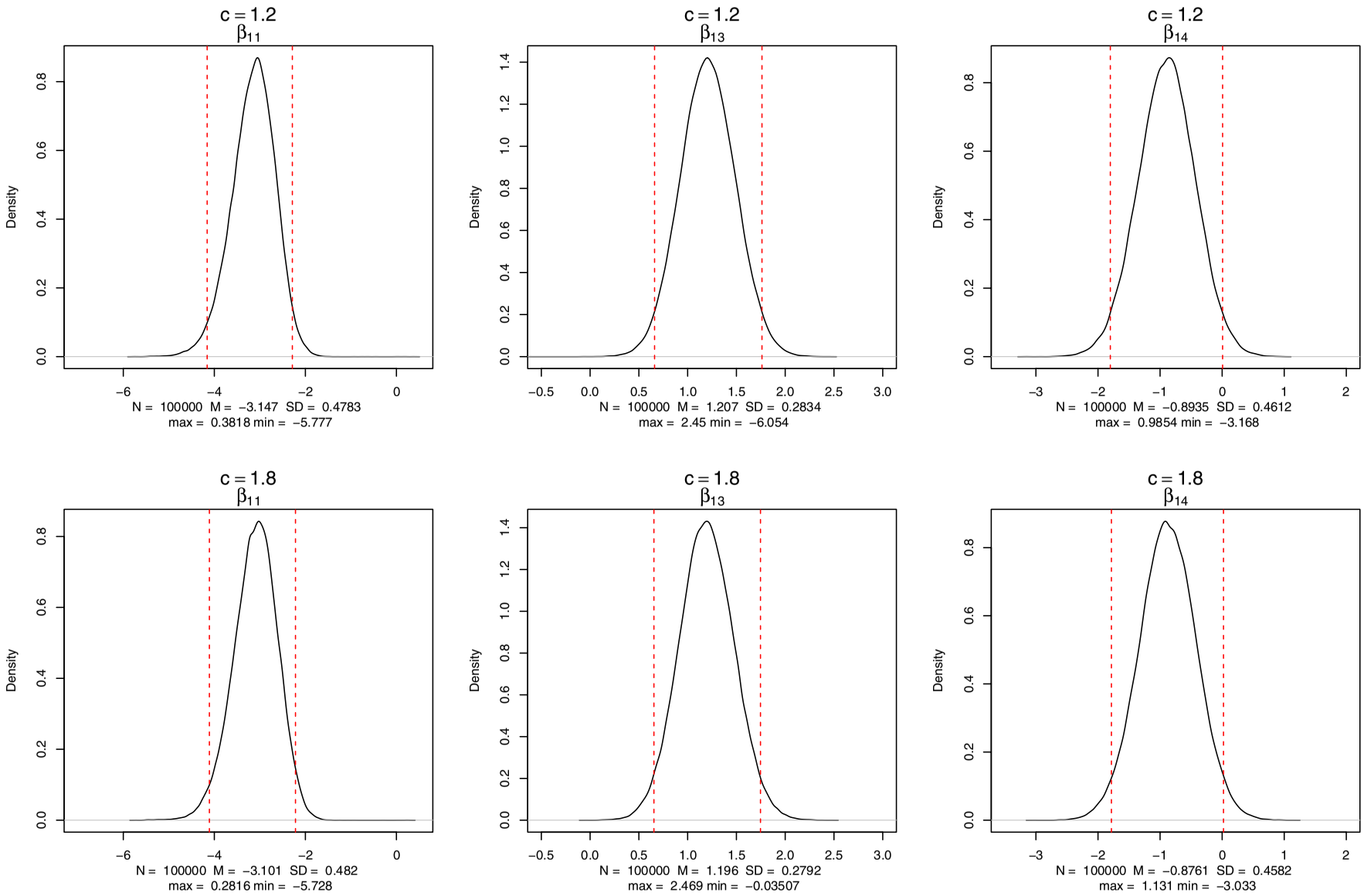
c = 1.2
β11
Density
N = 100000 M = −3.147 SD = 0.4783
max = 0.3818 min = −5.777
c = 1.2
β13
N = 100000 M = 1.207 SD = 0.2834
max = 2.45 min = −6.054
c = 1.2
β14
N = 100000 M = −0.8935 SD = 0.4612
max = 0.9854 min = −3.168
c = 1.8
β11
N = 100000 M = −3.101 SD = 0.482
max = 0.2816 min = −5.728
c = 1.8
β13
N = 100000 M = 1.196 SD = 0.2792
max = 2.469 min = −0.03507
c = 1.8
β14
N = 100000 M = −0.8761 SD = 0.4582
max = 1.131 min = −3.033

Figure S.9: