Supplementary Materials

Derivations of Full Conditional Distributions

Multivariate Skew-Normal Regression

Without loss of generality, we derive the full conditional distributions for the multivariate skew-normal regression model component under the assumption that all observations belong to a single cluster. To make the extension to the case where more than one cluster is specified, simply apply these distributional forms to cluster specific parameters and data. Finally, we assume for the moment that we have complete data for all outcomes for each subject. We extend consider the case of missing data in section (INSERT SECTION).

The multivariate skew-normal regression model can be written as follows in matrix form.

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{t}\boldsymbol{\psi}^T + \mathbf{E} = \mathbf{X}^*\mathbf{B}^* + \mathbf{E}$$

The matrix $\mathbf{Y}$ is of dimension $n \times J$. For convenience, we define $\mathbf{X}^*$ as a $n \times (P+1)$ matrix constructed by column binding $\mathbf{t}$ to $\mathbf{X}$, and $\mathbf{B}^*$ as a $(P+1) \times J$ matrix constructed by row binding $\boldsymbol{\psi}^T$ to $\mathbf{B}$. We assume that $t_i \overset{iid}{\sim} T_{[0,\infty)}(0,1)$ and that $\mathbf{E}$ is made of row vectors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{iJ})$ for $i = 1, ..., n$, where $\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_J(0, \boldsymbol{\Sigma})$.

The conditional likelihood for this model is given below.

$$p(\mathbf{Y}|\mathbf{X}^*, \mathbf{B}^*, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)^T(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)\boldsymbol{\Sigma}^{-1}\right\}$$

We choose conjugate priors for $\mathbf{B}^*$ and $\boldsymbol{\Sigma}$ as follows.

$$\boldsymbol{\Sigma} \sim \text{inverse-Wishart}(\mathbf{V}_0, \nu_0)$$

$$\mathbf{B}^*|\boldsymbol{\Sigma} \sim MatNorm_{(m+1),p}(\mathbf{B}_0^*, \mathbf{L}_0^{-1}, \boldsymbol{\Sigma})$$

We now derive the joint posterior distribution of the parameters $\mathbf{B}^*$ and $\boldsymbol{\Sigma}$.

$$\begin{aligned} p(\mathbf{B}^*, \boldsymbol{\Sigma}|\mathbf{X}^*, \mathbf{Y}) &\propto p(\mathbf{Y}|\mathbf{X}^*, \mathbf{B}^*, \boldsymbol{\Sigma})p(\mathbf{B}^*|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)^T(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)\boldsymbol{\Sigma}^{-1}\right]\right\} \\ &\times |\boldsymbol{\Sigma}|^{-(P+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{B}^* - \mathbf{B}_0^*)^T\mathbf{L}_0(\mathbf{B}^* - \mathbf{B}_0^*)\boldsymbol{\Sigma}^{-1}\right]\right\} \\ &\times |\boldsymbol{\Sigma}|^{(\nu_0+J+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{V}_0\boldsymbol{\Sigma}^{-1})\right\} \end{aligned}$$

Multinomial Logit Regression

Multivariate Normal Conditional Imputation

The multivariate normal conditional imputation derivations are given for a single cluster without loss of generality. In practice, the data and parameters in this section would be replaced by cluster specific estimates in the case of clustering.

For a given observation vector $\mathbf{y} \sim N_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we allow for missingness in at most $J-1$ of the multivariate outcomes through the use of a conditional imputation step embedded within our Gibbs sampler. Suppose $\mathbf{y}$ contains q missing observations and can be partitioned into two vectors $\mathbf{y_1}$ and $\mathbf{y_2}$ such that $\mathbf{y_1}$ is a $q \times 1$ vector of missing observations and $\mathbf{y_2}$ is a $(J-q) \times 1$ vector of complete observations. Similarly, partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as follows.

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

We will use these quantities to derive the conditional distribution $f(\mathbf{y_1}|\mathbf{y_2}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$\begin{aligned}
f(\mathbf{y_1}|\mathbf{y_2}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto f(\mathbf{y_1}, \mathbf{y_2}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \\
&= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^T \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}^*_{11} & \boldsymbol{\Sigma}^*_{12} \\ \boldsymbol{\Sigma}^*_{21} & \boldsymbol{\Sigma}^*_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}\left[(\mathbf{y}_1 - \boldsymbol{\mu}_{cond})^T \boldsymbol{\Sigma}_{cond}^{-1} (\mathbf{y}_1 - \boldsymbol{\mu}_{cond})\right]\right\}
\end{aligned}$$

$$\Rightarrow \mathbf{y_1}|\mathbf{y_2}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_q(\boldsymbol{\mu}_{cond}, \boldsymbol{\Sigma}_{cond})$$

$$\boldsymbol{\mu}_{cond} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \qquad \boldsymbol{\Sigma}_{cond} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

The block-wise inversion formula was used to invert $\boldsymbol{\Sigma}$ according to the following reparameterizations.

$$\begin{aligned}
\boldsymbol{\Sigma}^*_{11} &= \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} \\
\boldsymbol{\Sigma}^*_{12} &= -\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} \\
\boldsymbol{\Sigma}^*_{21} &= -(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} \\
\boldsymbol{\Sigma}^*_{22} &= (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}
\end{aligned}$$

Tables

Algorithm 1 Gibbs Sampler

Define n_{iter}; n_{burn}; K; $\boldsymbol{\theta}_{init}$; $\boldsymbol{\theta}_0$
$n_{sim} := n_{iter} - n_{burn}$
$\boldsymbol{\theta} := \boldsymbol{\theta}_{init}$ $\iota = 1, ..., n_{sim}$
I. Conditional Imputation $i = 1, ..., n$
Draw $\mathbf{y}_i^{miss}$ from $N_q(\boldsymbol{\mu}_i^{miss}, \boldsymbol{\Sigma}_i^{miss})$
$\mathbf{y}_i := \mathbf{y}_i^{miss} \cup \mathbf{y}_i^{obs}$
II. MSN Regression $k = 1, ..., K$
$n_k := \sum_{i=1}^{n} 1_{z_i=k}$ $i_k = 1, ..., n_k$
Draw t_i from $N_{[0,\infty)}(a_i, A)$
$\mathbf{X}^*{}_k := \texttt{cbind}(\mathbf{X}_k, \mathbf{t}_k)$
Draw $\mathbf{B}^*{}_k$ from $\text{MatNorm}(\mathbf{B}_k, \mathbf{L}_k^{-1}, \boldsymbol{\Sigma}_k)$
Draw $\boldsymbol{\Sigma}_k$ from $\text{InvWish}(\nu_k, \mathbf{V}_k)$
III. Multinomial Logit $i = 1, ..., n$ $k = 1, ..., K$
$\pi_{ik} := P(z_i = k | \mathbf{w}_i, \boldsymbol{\delta}_k)$
$p_{ik} := P(\mathbf{y}_i | \boldsymbol{\beta}_k^{*T} \mathbf{x}_i^*, \boldsymbol{\Sigma}_k)$
$\mathbf{p}_{z_i} := \frac{\mathbf{p}_i \circ \boldsymbol{\pi}_i}{\mathbf{p}_i \cdot \boldsymbol{\pi}_i}$
Draw z_i from $\text{Categorical}(\mathbf{p}_{z_i})$ $k = 1, ..., K-1$
Draw $\boldsymbol{\delta}_k$ from $N(\mathbf{M}, \mathbf{S})$
$\boldsymbol{\theta} := \{\mathbf{B}^*, \boldsymbol{\Sigma}, \mathbf{Z}, \boldsymbol{\delta}\}$
Store $\boldsymbol{\theta}$

Table 1: Model results for simulated data with n = 1,000, J = 4, p = 2, K = 3, r = 2. 1,000 iterations were run with a burn in of 100. Missingness mechanism was MAR and P(miss) = 0. Model results for the multivariate skew normal (MSN) and multivariate normal (MN) mixtures are presented.

		Class 1			Class 2			Class 3		
Component	Param.	True	MSN Est. (95% CrI)	MN Est. (95% CrI)	True	MSN Est. (95% CrI)	MN Est. (95% CrI)	True	MSN Est. (95% CrI)	MN Est. (95% CrI)
MVSN	β_{11}	11	11.07 (10.74, 11.39)	9.42 (8.91, 9.77)	-5	-4.95 (-5.2, -4.68)	-4.11 (-4.33, -3.86)	-10	-10.3 (-10.56, -10.01)	-8.37 (-8.86, -0.65)
Regression	β_{21}	12	12.02 (11.87, 12.17)	11.98 (11.77, 12.18)	-4	-4 (-4.1, -3.89)	-3.98 (-4.09, -3.87)	-11	-11 (-11.19, -10.82)	-10.85 (-11.1, 0.61)
	β_{31}	13	13.06 (12.75, 13.36)	11.39 (10.7, 11.78)	-3	-2.97 (-3.25, -2.68)	-3.74 (-3.99, -3.5)	-12	-11.9 (-12.22, -11.56)	-10.29 (-10.78, -0.14)
	β_{41}	14	14.06 (13.91, 14.22)	14.02 (13.78, 14.22)	-2	-1.96 (-2.07, -1.86)	-1.97 (-2.08, -1.84)	-13	-13.04 (-13.25, -12.87)	-12.89 (-13.13, 0.59)
	β_{12}	2	2.11 (1.82, 2.35)	0.42 (0.03, 0.83)	5	5.16 (4.88, 5.47)	5.83 (5.59, 6.07)	-2	-1.86 (-2.21, -1.52)	-0.35 (-0.77, 0.16)
	β_{22}	2	2.03 (1.88, 2.17)	2.02 (1.86, 2.22)	5	4.96 (4.86, 5.06)	4.96 (4.84, 5.07)	-2	-1.97 (-2.18, -1.79)	-1.89 (-2.11, 0.02)
	β_{32}	2	2.13 (1.8, 2.43)	0.49 (0.14, 0.86)	5	5.22 (4.96, 5.49)	4.23 (3.96, 4.5)	-2	-1.82 (-2.14, -1.5)	-0.37 (-0.77, 0.16)
	β_{42}	2	2.08 (1.93, 2.23)	2.08 (1.92, 2.28)	5	4.97 (4.86, 5.08)	4.96 (4.83, 5.08)	-2	-1.93 (-2.13, -1.77)	-1.84 (-2.04, 0.04)
	Ω_{11}	5	4.99 (3.84, 6.52)	3.08 (2.69, 4.16)	2	1.95 (1.52, 2.52)	1.38 (1.2, 1.59)	5	6.27 (4.84, 7.88)	3.49 (2.89, 186.69)
	Ω_{12}	4.5	4.55 (3.49, 5.88)	2.76 (2.39, 3.55)	-0.5	-0.51 (-0.83, -0.22)	0.2 (0.05, 0.36)	4.5	4.95 (3.78, 6.43)	3.02 (2.48, 215.36)
	Ω_{13}	4.25	4.53 (3.48, 5.85)	1.9 (1.58, 2.46)	1.25	1.08 (0.74, 1.53)	0.38 (0.24, 0.54)	4.25	4.7 (3.56, 6.04)	1.86 (1.42, 8.74)
	Ω_{14}	4.12	4.33 (3.31, 5.57)	1.78 (1.46, 2.24)	-0.88	-1.08 (-1.5, -0.75)	-0.41 (-0.58, -0.26)	4.12	4.51 (3.43, 5.69)	1.72 (1.29, 12.25)
	Ω_{22}	5	5.09 (3.98, 6.55)	3.43 (3.01, 4.23)	2	1.99 (1.53, 2.51)	1.43 (1.22, 1.67)	5	4.87 (3.75, 6.33)	3.71 (3.09, 253.61)
	Ω_{23}	4.5	4.77 (3.69, 6)	2.13 (1.78, 2.7)	-0.5	-0.46 (-0.75, -0.18)	0.05 (-0.1, 0.2)	4.5	4.29 (3.23, 5.61)	2.02 (1.55, 13.33)
	Ω_{24}	4.25	4.38 (3.39, 5.68)	1.86 (1.53, 2.39)	1.25	1.45 (1.04, 1.89)	0.64 (0.48, 0.83)	4.25	4.01 (3.04, 5.19)	1.78 (1.35, 16.8)
	Ω_{33}	5	5.47 (4.3, 6.67)	2.55 (2.23, 3.2)	2	1.73 (1.36, 2.3)	1.41 (1.22, 1.64)	5	4.79 (3.6, 6.24)	2.44 (1.99, 5.81)
	Ω_{34}	4.5	4.8 (3.7, 6.06)	1.99 (1.69, 2.68)	-0.5	-0.68 (-1.07, -0.37)	0.14 (-0.01, 0.3)	4.5	4.25 (3.2, 5.45)	1.94 (1.55, 5.3)
	Ω_{44}	5	5.17 (3.96, 6.69)	2.45 (2.14, 3.79)	2	2.34 (1.8, 2.97)	1.57 (1.37, 1.83)	5	4.64 (3.54, 5.92)	2.37 (1.93, 5.68)
	α_1	-0.99	-0.81 (-2.12, 0.05)	/	0.85	1.05 (0.37, 1.91)	/	0.99	2.82 (1.21, 4.3)	/
	α_2	-0.5	-0.22 (-1.3, 0.75)	/	-1.28	-1.29 (-2.22, -0.5)	/	0.5	-0.07 (-1.14, 1.13)	/
	α_3	-0.5	-0.96 (-2.14, 0.01)	/	1.28	1.16 (0.47, 2.06)	/	0.5	0.08 (-0.99, 1.46)	/
	α_4	-0.99	-1.18 (-2.44, -0.06)	/	-0.85	-0.91 (-1.76, -0.16)	/	0.99	1.1 (0.07, 2.33)	/
Multinom.	δ_{11}	-0.08	-0.07 (-0.27, 0.12)	-0.54 (-0.77, -0.32)	-0.08	-0.07 (-0.27, 0.12)	-0.54 (-0.77, -0.32)	-0.08	-0.07 (-0.27, 0.12)	-0.54 (-0.77, -0.32)
	δ_{12}	0.51	0.25 (-0.04, 0.53)	-0.26 (-0.6, 0.05)	0.51	0.25 (-0.04, 0.53)	-0.26 (-0.6, 0.05)	0.51	0.25 (-0.04, 0.53)	-0.26 (-0.6, 0.05)
	δ_{21}	-0.97	-0.91 (-1.15, -0.68)	-0.07 (-0.28, 0.14)	-0.97	-0.91 (-1.15, -0.68)	-0.07 (-0.28, 0.14)	-0.97	-0.91 (-1.15, -0.68)	-0.07 (-0.28, 0.14)
	δ_{22}	0.84	0.39 (0.09, 0.71)	0.24 (-0.04, 0.5)	0.84	0.39 (0.09, 0.71)	0.24 (-0.04, 0.5)	0.84	0.39 (0.09, 0.71)	0.24 (-0.04, 0.5)
Clustering	π_l	0.38	0.38 (0.38, 0.38)	0.38 (0.13, 0.41)	0.4	0.39 (0.39, 0.4)	0.39 (0.36, 0.43)	0.23	0.23 (0.22, 0.23)	0.23 (0.2, 0.44)

Figures

Figure 1: The distribution of residuals in repeated measures regression model of Bayley composite scores adjusted for race and sex.

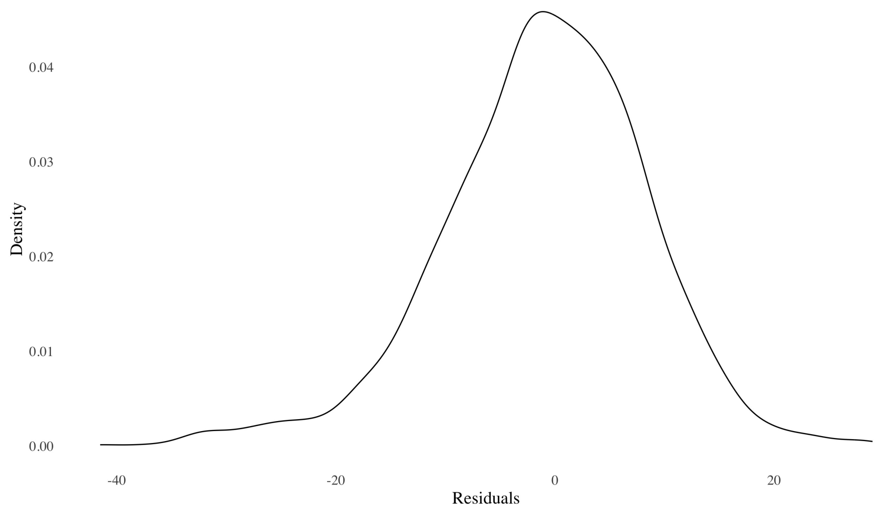

Figure 2: Marginal densities of simulated outcomes in each cluster at each measurement occasion.

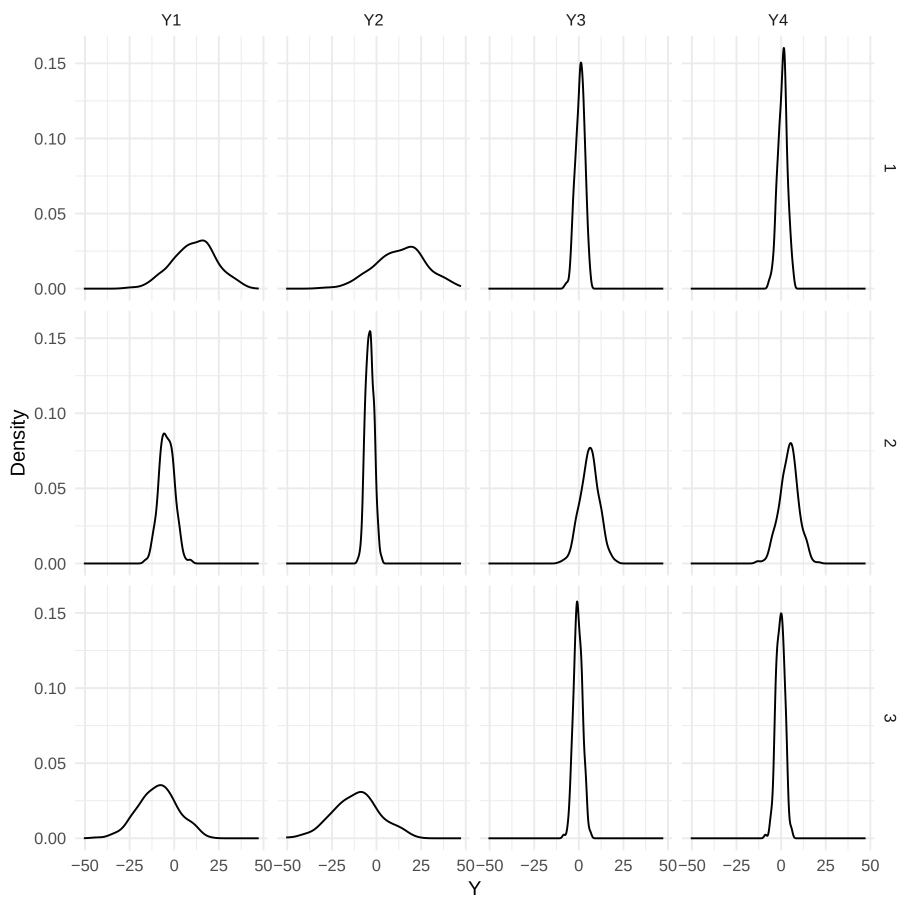

Figure 3: Mean plots of simulated outcomes in each cluster at each measurement occasion.

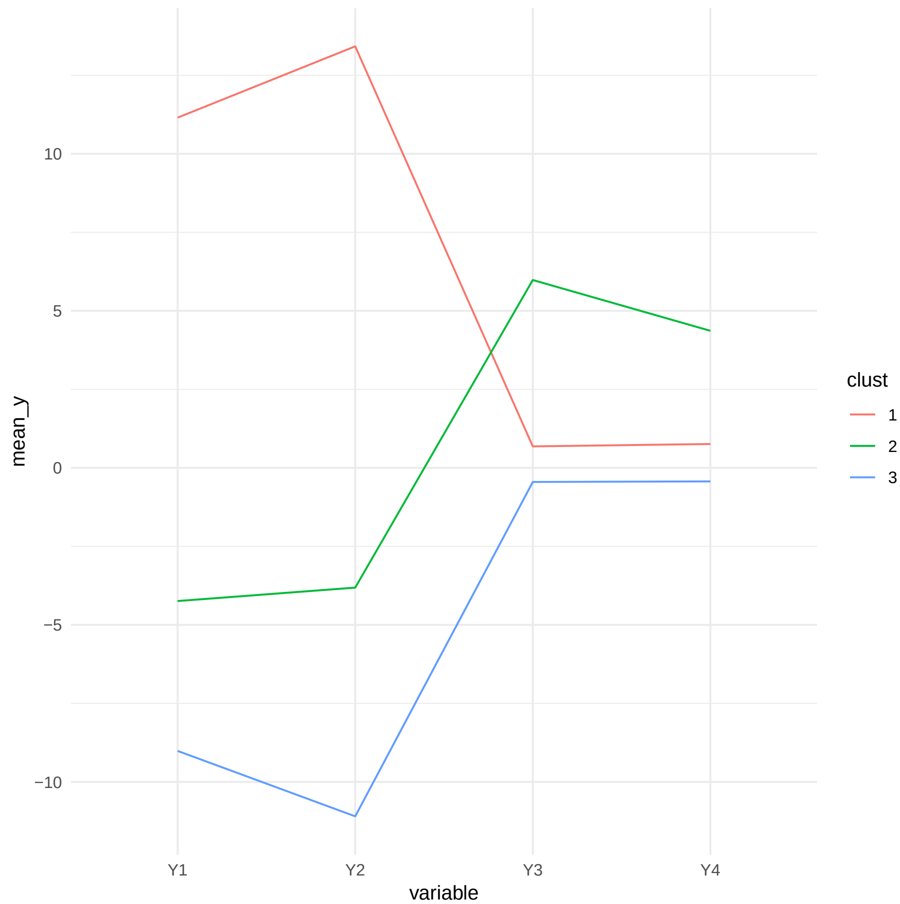

Figure 4: Trace plots of draws from posterior distribution of β parameters

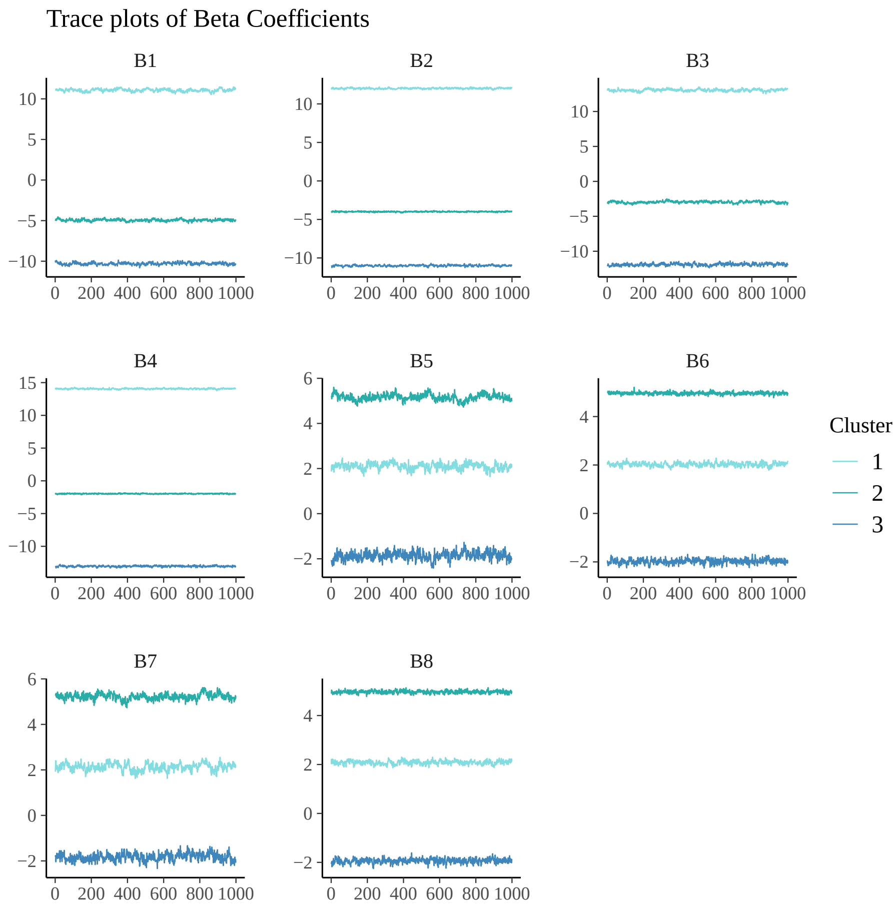

Figure 5: Trace plots of draws from posterior distribution of ψ parameters

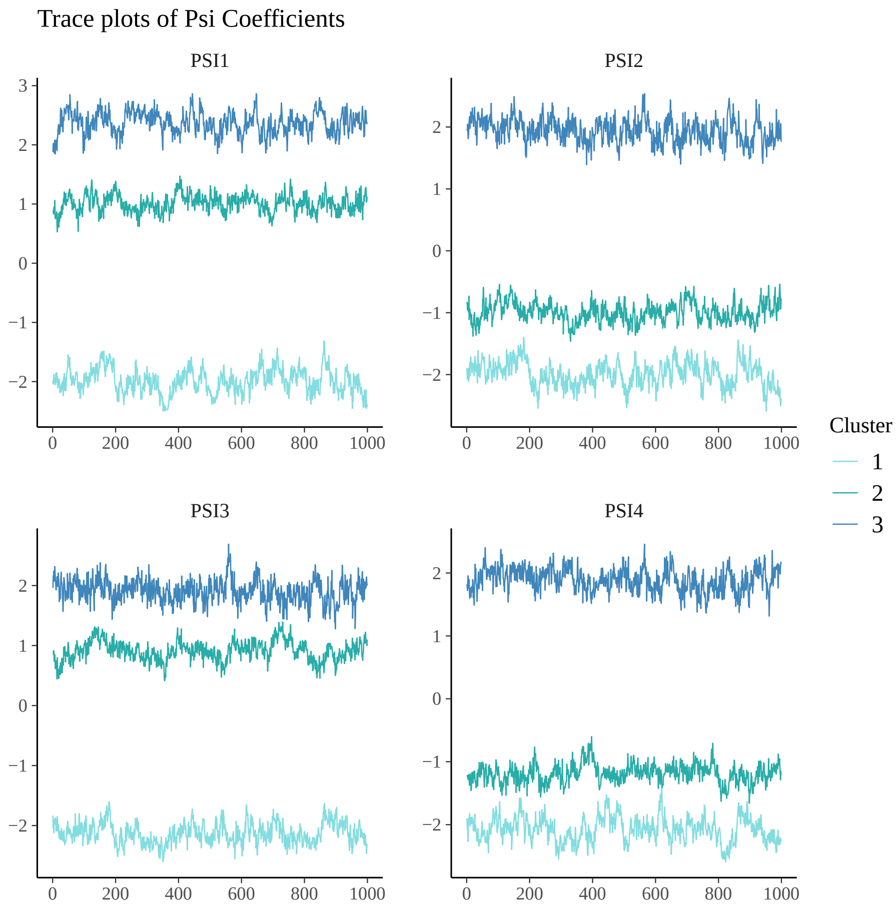

Figure 6: Trace plots of draws from posterior distribution of Σ parameters

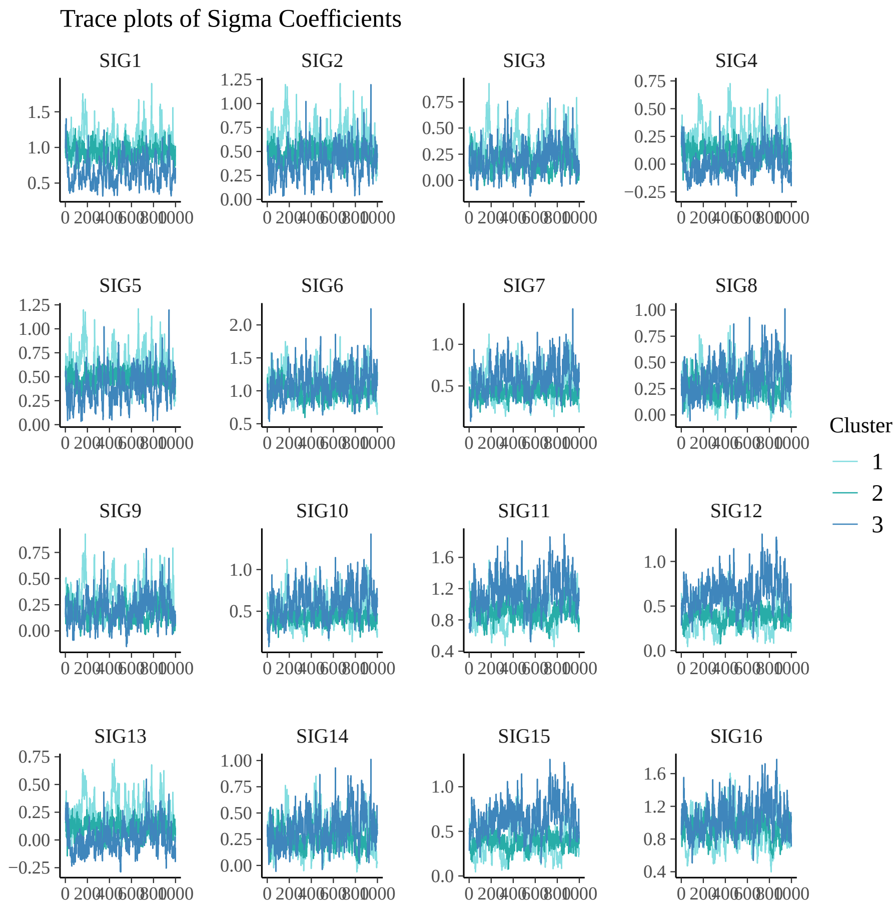

Figure 7: Trace plots of proportion within each class at each MCMC iteration