Supplemental material

Figure S1 presents trace plots of selected parameters from the simulation study. Panel a) and b) show the plots for subpopulations 1 and 2, respectively. The true parameter value is indicated by a solid grey line.

Table S1 provides BIC values and their accuracy across different number of mixture components (or subpopulations) in different sample sizes (N = 300 and 1000), using a single replicate. The BIC value was lower when H = 2, and the percentages of the average accuracy were kept stable with values greater (at least) than 98%. Table S2 presents the results of the simulation study analyzed with a single replicate under N = 300 with H = 1 and 2. Table S3 presents the results when the number of hidden subpopulations with the sequential two-step ERA model was fixed as one, two and three (H = 1, 2, 3) for the simulated data under N = 300. Table S3 presents the parameter estimates obtained from the sequential two-step ERA (applying K-means and ERA sequentially). While 18 clusters were chosen from the K-means, the results up to eight clusters are displayed in the table.

The sampling scheme for our proposed model is provided in this Supplemental material. Also, an R script for the proposed method is provided. The script presented here generates data of $N = 300$ from the simulation setting described in the main text and fits a mixture with H = 2 to the generated data.

Table S1. BIC values and their accuracy across different number of mixture components (or subpopulations) in different sample sizes ($N = 300$ and 1000)

	$N = 300$		$N = 1000$	
	BIC	Accuracy*	BIC	Accuracy*
$H = 1$	2775.2	-	9096.7	-
$H = 2$	1930.7	98.7%	5808.9	99.0%
$H = 3$	1888.6	98.7%	5878.8	98.9%

* When $H = 1$, no classification was performed.

Table S2. Simulation results when the number of hidden subpopulations with the proposed model was fixed as one, two and three ($H = 1, 2, 3$) with $N = 300$

parameter		truth	$H = 1$			$H = 2$			$H = 3$		
			mean	low	up	mean	low	up	mean	low	up
$\mathbf{W}_1$	w_{11}^1	0.6	0.52	0.37	0.66	0.60	0.59	0.62	0.60	0.59	0.62
	w_{21}^1	0.4	0.51	0.35	0.65	0.40	0.38	0.42	0.40	0.38	0.42
	w_{32}^1	1	0.79	0.67	0.90	0.96	0.94	0.98	0.96	0.94	0.98
	w_{42}^1	0	0.25	0.11	0.40	0.01	-0.02	0.03	0.01	-0.02	0.03
$\mathbf{W}_2$	w_{11}^2	0.6	-	-	-	0.61	0.56	0.65	0.61	0.56	0.65
	w_{21}^2	0.4	-	-	-	0.45	0.41	0.50	0.45	0.41	0.50
	w_{32}^2	0.2	-	-	-	0.22	0.08	0.35	0.22	0.08	0.35
	w_{42}^2	0.8	-	-	-	0.80	0.69	0.90	0.80	0.69	0.90
$\mathbf{A}_1$	a_{11}^1	2	1.44	1.09	1.80	1.96	1.87	2.05	1.96	1.88	2.04
	a_{21}^1	3	2.16	1.87	2.44	3.18	3.09	3.26	3.18	3.09	3.26
	a_{12}^1	-2	0.15	-0.18	0.48	-1.95	-2.13	-1.77	-1.95	-2.13	-1.77
	a_{22}^1	0	0.66	0.38	0.94	0.13	-0.03	0.28	0.13	-0.02	0.28
$\mathbf{A}_2$	a_{11}^2	1	-	-	-	1.04	0.82	1.26	1.04	0.82	1.26
	a_{21}^2	1.5	-	-	-	1.24	1.06	1.43	1.24	1.06	1.42
	a_{12}^2	3	-	-	-	2.88	2.66	3.11	2.88	2.65	3.11
	a_{22}^2	1	-	-	-	1.09	0.90	1.27	1.09	0.90	1.28
$\boldsymbol{\Sigma}_1$	σ_{11}^1	0.2	5.89	5.03	6.94	0.19	0.16	0.24	0.19	0.15	0.24
	σ_{12}^1	0.18	-1.36	-2.07	-0.70	0.24	0.17	0.32	0.23	0.16	0.32
	σ_{22}^1	1	5.52	4.69	6.49	0.93	0.75	1.16	0.93	0.74	1.16
Σ_2	σ_{11}^2	1	-	-	-	0.98	0.76	1.25	0.97	0.76	1.24
	σ_{12}^2	0.1	-	-	-	-0.01	-0.19	0.17	-0.01	-0.18	0.17
	σ_{22}^2	1	-	-	-	1.02	0.79	1.29	1.02	0.79	1.31
$\mathbf{a}_{01}$	$\mathbf{a}_{01[1]}$	-1	0.84	0.56	1.11	-0.96	-1.03	-0.89	-0.96	-1.03	-0.89
	$\mathbf{a}_{01[2]}$	0	-0.24	-0.51	0.02	-0.10	-0.24	0.06	-0.10	-0.24	0.05
$\mathbf{a}_{02}$	$\mathbf{a}_{02[1]}$	3	-	-	-	2.93	2.76	3.10	2.93	2.76	3.10
	$\mathbf{a}_{02[2]}$	0	-	-	-	-0.16	-0.34	0.02	-0.16	-0.34	0.01
$\boldsymbol{\lambda}$	λ_1	0.6	1.00	1.00	1.00	0.56	0.51	0.62	0.56	0.51	0.62
	λ_2	0.4	-	-	-	0.44	0.38	0.49	0.44	0.38	0.49

Note: When $H = 3$, the results up to the second subpopulations that were close to the true parameter values were presented in the table, and λ_3 was estimated to be zero correctly.

Table S3. Simulation results when the number of hidden subpopulations with the sequential two-step ERA model was fixed as one, two and three (H = 1, 2, 3) with N = 300

		H = 1			H = 2			H = 3		
parameter		mean	low	up	mean	low	up	mean	low	up
$\mathbf{W}_1$	w_{11}^1	0.54	0.41	0.68	0.73	-0.35	0.89	0.85	-0.67	0.94
	w_{21}^1	0.49	0.35	0.65	0.30	-0.52	0.76	0.02	-0.53	0.64
	w_{32}^1	0.82	0.69	0.96	1.00	-1.12	1.19	0.91	0.52	1.2
	w_{42}^1	0.21	0.05	0.36	-0.13	-0.52	0.58	0.30	0.00	0.62
$\mathbf{W}_2$	w_{11}^2	-	-	-	0.66	-0.21	1.25	-0.34	-0.7	0.32
	w_{21}^2	-	-	-	1.39	-1.56	1.93	0.43	0.21	0.66
	w_{32}^2	-	-	-	-0.41	-1.23	0.5	0.63	-0.14	1.02
	w_{42}^2	-	-	-	0.83	-0.24	1.72	0.91	0.63	1.24
$\mathbf{W}_3$	w_{11}^3	-	-	-	-	-	-	0.68	0.46	0.94
	w_{21}^3	-	-	-	-	-	-	-1.58	-1.84	-1.33
	w_{32}^3	-	-	-	-	-	-	0.49	0.33	0.64
	w_{42}^3	-	-	-	-	-	-	0.53	0.34	0.74
$\mathbf{A}_1$	a_{11}^1	1.45	1.07	1.81	-1.42	-1.9	-1.13	0.58	0.26	0.98
	a_{21}^1	2.16	1.79	2.47	-0.25	-0.95	0.34	0.39	-0.06	0.76
	a_{12}^1	0.15	-0.35	0.73	0.42	0.26	0.66	1.61	1.26	1.97
	a_{22}^1	0.65	0.31	0.93	0.59	0.39	0.75	1.26	0.64	1.76
$\mathbf{A}_2$	a_{11}^2	-	-	-	0.36	0.08	0.61	0.2	-0.87	1.25
	a_{21}^2	-	-	-	0.77	0.46	1.02	-0.55	-1.25	0.21
	a_{12}^2	-	-	-	1.03	0.73	1.36	2.59	2.28	3.02
	a_{22}^2	-	-	-	1.02	0.62	1.44	1.34	0.75	1.96
$\mathbf{A}_3$	a_{11}^3	-	-	-	-	-	-	0.62	0.54	0.76
	a_{21}^3	-	-	-	-	-	-	0.39	0.31	0.45
	a_{12}^3	-	-	-	-	-	-	0.88	0.78	1.01
	a_{22}^3	-	-	-	-	-	-	0.01	-0.04	0.07
$\mathbf{a}_{01}$	$\mathbf{a}_{01[1]}$	0.84	0.58	1.13	0.61	-0.09	1.47	1.83	1.34	2.13
	$\mathbf{a}_{01[2]}$	-0.24	-0.51	0.04	0.57	0.24	1.03	3.35	2.85	3.72
$\mathbf{a}_{02}$	$\mathbf{a}_{02[1]}$	-	-	-	2.52	2.17	2.77	-2.01	-2.57	-1.5
	$\mathbf{a}_{02[2]}$	-	-	-	-0.38	-0.75	-0.03	-0.07	-0.68	0.46
$\mathbf{a}_{03}$	$\mathbf{a}_{03[1]}$	-	-	-	-	-	-	-1.12	-1.39	-0.91
	$\mathbf{a}_{03[2]}$	-	-	-	-	-	-	-0.25	-0.89	0.36

		$H = 1$			$H = 2$			$H = 3$		
parameter		mean	low	up	mean	low	up	mean	low	up
$\boldsymbol{\lambda}$	λ_1	1.00	-	-	0.42	-	-	0.43	-	-
	λ_2	-	-	-	0.58	-	-	0.37	-	-
	λ_3	-	-	-	-	-	-	0.20	-	-

Note: $\boldsymbol{\lambda}$ was estimated as proportions of observations assigned to the hidden subpopulations as a result of *K*-means clustering. The original ERA model did not estimate residual covariance matrix $\boldsymbol{\Sigma}$, and thus no estimates were provided for each subpopulation. The true parameter values for the two subpopulations can be found in Table S2.

Table S4. Parameter estimates obtained from the sequential two-step ERA model. Note that the results up to eight clusters (out of 18 clusters) are shown.

Parameters		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
		est	95% CI	est	95% CI	est	95% CI	est	95% CI
Perpetration	a_{11}	0.08	(-0.09, 0.23)	-0.07	(-0.13, -0.03)	-0.02	(-0.06, 0.01)	-0.06	(-0.25, 0.07)
	a_{21}	-0.15	(-0.32, 0.18)	0.08	(-0.27, 0.15)	-0.09	(-0.17, 0.13)	0.14	(-0.17, 0.3)
	a_{31}	0.2	(-0.38, 0.32)	0.02	(-0.11, 0.17)	0.02	(-0.07, 0.12)	0.06	(-0.26, 0.28)
	a_{41}	0.1	(-0.14, 0.36)	0	(-0.06, 0.07)	0.05	(-0.03, 0.1)	0.16	(-0.28, 0.36)
	a_{12}	0.14	(0.02, 0.26)	0	(-0.03, 0.02)	0	(0, 0.01)	0.01	(-0.06, 0.09)
	a_{22}	-0.13	(-0.24, 0.15)	0.01	(-0.08, 0.07)	-0.02	(-0.04, 0.04)	0.02	(-0.06, 0.08)
	a_{32}	0.13	(-0.22, 0.27)	0.02	(-0.03, 0.08)	0.01	(-0.02, 0.05)	0.02	(-0.05, 0.07)
	a_{42}	0.06	(-0.13, 0.22)	0.02	(-0.02, 0.05)	0.02	(-0.01, 0.04)	0.02	(-0.09, 0.09)
Victimization	a_{13}	-0.3	(-0.75, -0.02)	-0.1	(-0.22, -0.04)	-0.05	(-0.14, 0.05)	0.17	(-0.03, 0.33)
	a_{23}	0.47	(-0.61, 0.85)	0.04	(-0.16, 0.29)	-0.12	(-0.23, 0.2)	-0.03	(-0.33, 0.32)
	a_{33}	-0.57	(-0.96, 0.81)	0.17	(-0.08, 0.28)	0.13	(-0.29, 0.37)	-0.33	(-0.53, 0.51)
	a_{43}	0.08	(-0.71, 0.14)	0.05	(-0.11, 0.16)	0.04	(-0.11, 0.14)	-0.05	(-0.27, 0.27)
	a_{14}	-0.23	(-0.52, -0.06)	-0.04	(-0.08, -0.01)	0.01	(-0.01, 0.02)	0.10	(0.02, 0.16)
	a_{24}	0.15	(-0.45, 0.5)	0.01	(-0.08, 0.05)	-0.02	(-0.06, 0.04)	0.02	(-0.04, 0.08)
	a_{34}	-0.11	(-0.53, 0.48)	0.04	(-0.01, 0.11)	0	(-0.04, 0.05)	0.02	(-0.07, 0.11)
	a_{44}	0.34	(-0.48, 0.5)	0.02	(-0.03, 0.06)	-0.01	(-0.03, 0.02)	-0.04	(-0.08, 0.09)
Parameters		Cluster 5		Cluster 6		Cluster 7		Cluster 8	
		est	95% CI	est	95% CI	est	95% CI	est	95% CI
Perpetration	a_{11}	-0.02	(-0.05, 0.02)	-0.04	(-0.08, 0)	-0.03	(-0.06, 0.03)	-0.02	(-0.05, 0)
	a_{21}	-0.03	(-0.12, 0.13)	0.03	(-0.07, 0.08)	0.11	(-0.13, 0.17)	0.09	(0.02, 0.14)
	a_{31}	-0.09	(-0.18, 0.09)	0.04	(-0.05, 0.09)	0.08	(-0.21, 0.19)	-0.07	(-0.14, 0.08)
	a_{41}	-0.06	(-0.12, 0.07)	0.04	(-0.01, 0.09)	0.03	(-0.08, 0.07)	-0.01	(-0.05, 0.07)
	a_{12}	0.00	(-0.01, 0)	-0.01	(-0.02, 0.01)	0.00	(-0.02, 0.01)	0.00	(0, 0)
	a_{22}	0.00	(0, 0.01)	0.01	(-0.03, 0.04)	0.02	(-0.04, 0.07)	0.01	(0, 0.03)
	a_{32}	0.00	(-0.01, 0.01)	0.02	(-0.03, 0.06)	-0.01	(-0.04, 0.05)	-0.01	(-0.07, 0.07)
	a_{42}	-0.01	(-0.02, 0.02)	0.01	(-0.01, 0.04)	-0.01	(-0.01, 0.02)	0.00	(-0.01, 0.01)
Victimization	a_{13}	-0.02	(-0.06, 0.03)	-0.01	(-0.08, 0.05)	-0.01	(-0.09, 0.07)	-0.06	(-0.1, -0.03)
	a_{23}	0.13	(-0.06, 0.21)	0.10	(-0.18, 0.16)	-0.04	(-0.21, 0.08)	0.15	(0.07, 0.23)
	a_{33}	-0.10	(-0.19, 0.16)	0.20	(-0.33, 0.33)	0.06	(-0.22, 0.34)	-0.03	(-0.1, 0.12)
	a_{43}	0.04	(-0.09, 0.18)	-0.05	(-0.14, 0.09)	-0.03	(-0.11, 0.12)	-0.12	(-0.19, 0.17)
	a_{14}	0.00	(-0.01, 0.01)	0.01	(-0.01, 0.04)	0.02	(0, 0.05)	0.00	(-0.01, 0.01)
	a_{24}	-0.01	(-0.02, 0.02)	0.01	(-0.04, 0.05)	-0.01	(-0.04, 0.02)	0.01	(-0.01, 0.03)
	a_{34}	0.01	(-0.01, 0.02)	0.03	(-0.07, 0.08)	0.01	(-0.03, 0.05)	0.00	(-0.02, 0.02)
	a_{44}	0.00	(-0.01, 0.03)	0.02	(-0.03, 0.05)	0.01	(-0.02, 0.04)	-0.01	(-0.04, 0.03)

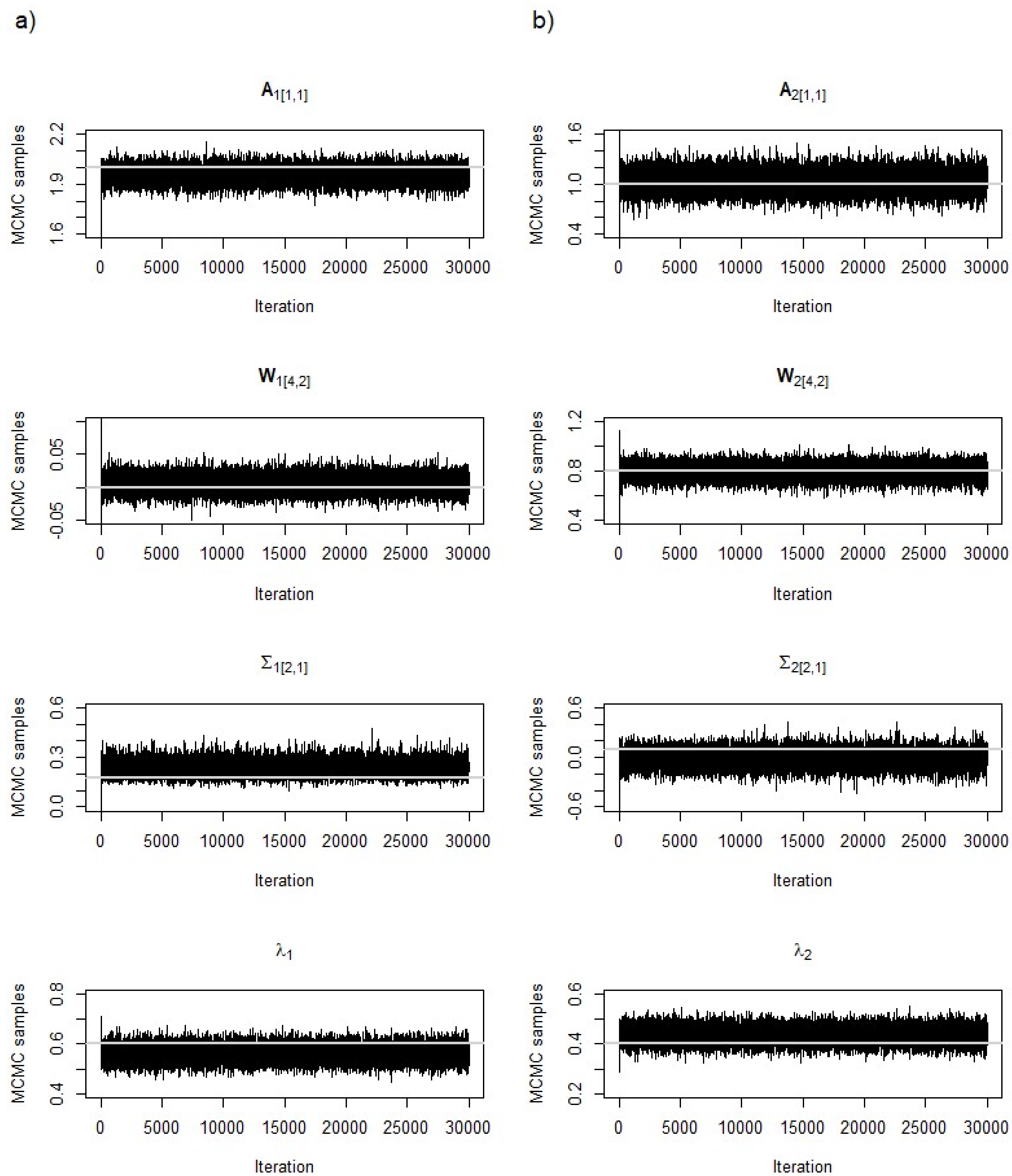

Figure S1. Trace plots of selected parameters from the simulation study. Panel a) and b) show the plots for subpopulations 1 and 2, respectively. A solid grey line indicates the true parameter value.

Sampling scheme of the parameters for finite mixture ERA models

With the conjugate prior specification described in the section of Bayesian inference on finite mixture ERA models, the full-data joint posterior distribution, from which the full conditional distributions are derived for a Gibbs sampler, has the form of

$$
\begin{aligned}
&\pi\left(\boldsymbol{\lambda}, \mathbf{C}, \mathbf{W}, \mathbf{A}, \boldsymbol{\Sigma}, \mathbf{a}_0 \mid H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right) \\
&\propto \prod_{i=1}^{N}\prod_{h=1}^{H}\left[\lambda_h^{I(C_i=h)}\left|\boldsymbol{\Sigma}_{C_i}\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{Y}_{[i,]}'-\mathbf{a}_{oC_i}-\mathbf{A}_{C_i}'\mathbf{W}_{C_i}'\mathbf{X}_{[i,]}'\right)'\boldsymbol{\Sigma}_{C_i}^{-1}\left(\mathbf{Y}_{[i,]}'-\mathbf{a}_{oC_i}-\mathbf{A}_{C_i}'\mathbf{W}_{C_i}'\mathbf{X}_{[i,]}'\right)\right\}\right] \\
&\times\prod_{h=1}^{H}\left[\left(\lambda_h\right)^{\alpha_h-1}\exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{a}_{oh}-\boldsymbol{\alpha}_0\right)'\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\left(\mathbf{a}_{oh}-\boldsymbol{\alpha}_0\right)\right\}\left|\boldsymbol{\Sigma}_h^{-1}\right|^{\frac{1}{2}(\nu-Q-1)}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_h^{-1}\right)\right\}\right. \\
&\qquad\times\prod_{k=1}^{K}\left[\exp\left\{-\frac{1}{2\tau^2}\left(\mathbf{W}_{h[,k]}-\mathbf{W}_{[,k]}^0\right)'\Sigma_{\mathbf{W}_k}^{0\ -1}\left(\mathbf{W}_{h[,k]}-\mathbf{W}_{[,k]}^0\right)\right\}\right] \\
&\qquad\left.\times\prod_{k=1}^{K}\left[\exp\left\{-\frac{1}{2\eta^2}\left(\mathbf{A}_{h[k,]}'-\mathbf{A}_{[k,]}^{0\ \prime}\right)'\Sigma_{\mathbf{A}_k}^{0\ -1}\left(\mathbf{A}_{h[k,]}'-\mathbf{A}_{[k,]}^{0\ \prime}\right)\right\}\right]\right]
\end{aligned}
$$

where $\alpha_h > 0$ for all $h = 1, \cdots, H$ and $\nu \geq Q$.

We obtain the posterior samples of the parameters by alternating the following steps:

- conditional posterior distribution of C_i: for $h = 1, \cdots, H$ and $i = 1, \cdots, N$,

$$C_i \mid \boldsymbol{\lambda}, \mathbf{X}_{[i,]}, \boldsymbol{\Sigma}, \mathbf{W}, \mathbf{A}, \mathbf{a}_0 \sim P\left(C_i = h \mid \boldsymbol{\lambda}, \mathbf{X}_{[i,]}, \boldsymbol{\Sigma}, \mathbf{W}, \mathbf{A}, \mathbf{a}_0\right)$$

$$
\begin{aligned}
&\propto \lambda_h^{I(C_i=h)}\left|\boldsymbol{\Sigma}_{C_i}\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{Y}_{[i,]}'-\mathbf{a}_{oC_i}-\mathbf{A}_{C_i}'\mathbf{W}_{C_i}'\mathbf{X}_{[i,]}'\right)'\boldsymbol{\Sigma}_{C_i}^{-1}\left(\mathbf{Y}_{[i,]}'-\mathbf{a}_{oC_i}-\mathbf{A}_{C_i}'\mathbf{W}_{C_i}'\mathbf{X}_{[i,]}'\right)\right\} \\
&= \lambda_h\left|\boldsymbol{\Sigma}_h\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{Y}_{[i,]}'-\mathbf{a}_{oh}-\mathbf{A}_h'\mathbf{W}_h'\mathbf{X}_{[i,]}'\right)'\boldsymbol{\Sigma}_h^{-1}\left(\mathbf{Y}_{[i,]}'-\mathbf{a}_{oh}-\mathbf{A}_h'\mathbf{W}_h'\mathbf{X}_{[i,]}'\right)\right\}
\end{aligned}
$$

Then update C_i from its multinomial conditional posterior with probability

$$P\left(C_i = h \mid \boldsymbol{\lambda}, \mathbf{X}_{[i,]}, \boldsymbol{\Sigma}, \mathbf{W}, \mathbf{A}, \mathbf{a}_0\right) = \frac{\lambda_h N_Q\left(\mathbf{Y}_{[i,]} \mid \mathbf{a}_{0h} + \mathbf{A}_h'\mathbf{W}_h'\mathbf{X}_{[i,]}', \boldsymbol{\Sigma}_h\right)}{\sum_{l=1}^{H}\lambda_l N_Q\left(\mathbf{Y}_{[i,]} \mid \mathbf{a}_{0l} + \mathbf{A}_l'\mathbf{W}_l'\mathbf{X}_{[i,]}', \boldsymbol{\Sigma}_l\right)}.$$

- conditional posterior distribution of λ_h : for $h = 1, \cdots, H$

$$\pi\left(\lambda_h \mid \mathbf{C}, \mathbf{W}, \mathbf{A}, \boldsymbol{\Sigma}, \mathbf{a}_0, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right)$$
$$\propto \prod_{i=1}^{N} \lambda_h^{I(C_i=h)} \times \left(\lambda_h\right)^{\alpha_h - 1} = \left(\lambda_h\right)^{\sum_{i=1}^{N} I(C_i=h)} \times \left(\lambda_h\right)^{\alpha_h - 1} = \left(\lambda_h\right)^{N_h + \alpha_h - 1}$$

where $\sum_{i=1}^{N} I\left(C_i = h\right) = N_h$. Thus,

$$\boldsymbol{\lambda} \mid \mathbf{C}, \mathbf{W}, \mathbf{A}, \boldsymbol{\Sigma}, \mathbf{a}_0, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2 \sim \text{Dirichlet}\left(\alpha_1^*, \cdots, \alpha_H^*\right) \text{ where } \alpha_h^* = \alpha + N_h .$$

- conditional posterior distribution of $\mathbf{W}_{h[\,,k]}$: for $h = 1, \cdots, H$ and $k = 1, \cdots, K$

$$\pi\left(\mathbf{W}_{h[\,,k]} \mid \mathbf{W}_{h[,-k]}, \lambda_h, \mathbf{C}, \mathbf{A}_h, \mathbf{a}_{\mathbf{0}h}, \boldsymbol{\Sigma}_h, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right)$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]}{}' - \mathbf{a}_{oh} - \mathbf{A}_h{}'\mathbf{W}_h{}'\mathbf{X}_{[I(C_i=h),]}{}'\right)' \boldsymbol{\Sigma}_h^{-1}\left(\mathbf{Y}_{[I(C_i=h),]}{}' - \mathbf{a}_{oh} - \mathbf{A}_h{}'\mathbf{W}_h{}'\mathbf{X}_{[I(C_i=h),]}{}'\right)\right\}$$
$$\times \exp\left\{-\frac{1}{2\tau^2}\left(\mathbf{W}_{h[,k]} - \mathbf{W}_{[,k]}^0\right)' \Sigma_{\mathbf{W}_k}^{0\ -1}\left(\mathbf{W}_{h[,k]} - \mathbf{W}_{[,k]}^0\right)\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]} - \mathbf{a}_{0h}{}' - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,-k]}\mathbf{A}_{h[-k,]} - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\mathbf{A}_{h[k,]}\right)\right.$$
$$\left.\boldsymbol{\Sigma}_h^{-1}\left(\mathbf{Y}_{[I(C_i=h),]} - \mathbf{a}_{0h}{}' - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,-k]}\mathbf{A}_{h[-k,]} - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\mathbf{A}_{h[k,]}\right)'\right\}$$
$$\times \exp\left\{-\frac{1}{2\tau^2}\left(\mathbf{W}_{h[,k]} - \mathbf{W}_{[,k]}^0\right)' \Sigma_{\mathbf{W}_k}^{0\ -1}\left(\mathbf{W}_{h[,k]} - \mathbf{W}_{[,k]}^0\right)\right\}$$

Let $\mathbf{Y}_{[I(C_i=h),]}^{-k} = \mathbf{Y}_{[I(C_i=h),]} - \mathbf{a}_{0h}{}' - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,-k]}\mathbf{A}_{h[-k,]}$ and here $\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}$ is a scalar, then $\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]} = \mathbf{W}_{h[,k]}{}'\mathbf{X}_{[I(C_i=h),]}{}'$. Thus,

$$\pi\left(\mathbf{W}_{h[\,,k]} \mid \mathbf{W}_{h[,-k]}, \lambda_h, \mathbf{C}, \mathbf{A}_h, \mathbf{a}_{\mathbf{0}h}, \boldsymbol{\Sigma}_h, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right)$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]}^{-k} - \mathbf{W}_{h[,k]}{}'\mathbf{X}_{[I(C_i=h),]}{}'\mathbf{A}_{h[k,]}\right)\boldsymbol{\Sigma}_h^{-1}\left(\mathbf{Y}_{[I(C_i=h),]}^{-k} - \mathbf{W}_{h[,k]}{}'\mathbf{X}_{[I(C_i=h),]}{}'\mathbf{A}_{h[k,]}\right)'\right\}$$
$$\times \exp\left\{-\frac{1}{2\tau^2}\left(\mathbf{W}_{h[,k]} - \mathbf{W}_{[,k]}^0\right)' \Sigma_{\mathbf{W}_k}^{0\ -1}\left(\mathbf{W}_{h[,k]} - \mathbf{W}_{[,k]}^0\right)\right\}$$

$$\propto \exp\left[-\frac{1}{2}\left\{\mathbf{W}_{h[,k]}{}'\left(\sum_{i=1}^{N}\mathbf{X}'_{[I(C_i=h),]}\mathbf{A}_{h[k,]}\mathbf{\Sigma}_h{}^{-1}\mathbf{A}_{h[k,]}{}'\mathbf{X}_{[I(C_i=h),]}+\frac{1}{\tau^2}\mathbf{\Sigma}^{0}_{\mathbf{W}_k}{}^{-1}\right)\mathbf{W}_{h[,k]}\right.\right.$$

$$\left.\left.-2\mathbf{W}_{h[,k]}{}'\left(\sum_{i=1}^{N}\mathbf{X}'_{[I(C_i=h),]}\mathbf{A}_{h[k,]}\mathbf{\Sigma}_h{}^{-1}\mathbf{Y}^{-k}_{[I(C_i=h),]}{}'+\frac{1}{\tau^2}\mathbf{\Sigma}^{0}_{\mathbf{W}_k}{}^{-1}\mathbf{W}^{0}_{[,k]}\right)\right\}\right]$$

Therefore, $\mathbf{W}_{h[,k]} \mid \mathbf{W}_{h[,-k]},\lambda_h,\mathbf{C},\mathbf{A}_h,\mathbf{a}_{\mathbf{0}h},\mathbf{\Sigma}_h,H,\mathbf{X},\mathbf{Y},\tau^2,\eta^2,\sigma^2 \sim N_{p_k}\left(\widetilde{\mathbf{W}}_{h[,k]},\ \mathbf{\Sigma}_{\mathbf{W}_{hk}}\right)$

where $\mathbf{\Sigma}_{\mathbf{W}_{hk}}=\left(\sum_{i=1}^{N}\mathbf{X}'_{[I(C_i=h),]}\mathbf{A}_{h[k,]}\mathbf{\Sigma}_h{}^{-1}\mathbf{A}_{h[k,]}{}'\mathbf{X}_{[I(C_i=h),]}+\frac{1}{\tau^2}\mathbf{\Sigma}^{0}_{\mathbf{W}_k}{}^{-1}\right)^{-1}$ and

$$\widetilde{\mathbf{W}}_{h[,k]} = \mathbf{\Sigma}_{\mathbf{W}_{hk}}\left(\sum_{i=1}^{N}\mathbf{X}'_{[I(C_i=h),]}\mathbf{A}_{h[k,]}\mathbf{\Sigma}_h{}^{-1}\mathbf{Y}^{-k}_{[I(C_i=h),]}{}'+\frac{1}{\tau^2}\mathbf{\Sigma}^{0}_{\mathbf{W}_k}{}^{-1}\mathbf{W}^{0}_{[,k]}\right).$$

- conditional posterior distribution of $\mathbf{A}_{h[k,]}$: for $h=1,\cdots,H$ and $k=1,\cdots,K$

$$\pi\left(\mathbf{A}_{h[k,]} \mid \mathbf{A}_{h[-k,]},\lambda_h,\mathbf{C},\mathbf{W}_h,\mathbf{a}_{\mathbf{0}h},\mathbf{\Sigma}_h,H,\mathbf{X},\mathbf{Y},\tau^2,\eta^2,\sigma^2\right)$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]}{}'-\mathbf{a}_{oh}-\mathbf{A}_h{}'\mathbf{W}_h{}'\mathbf{X}_{[I(C_i=h),]}{}'\right)'\mathbf{\Sigma}_h{}^{-1}\left(\mathbf{Y}_{[I(C_i=h),]}{}'-\mathbf{a}_{oh}-\mathbf{A}_h{}'\mathbf{W}_h{}'\mathbf{X}_{[I(C_i=h),]}{}'\right)\right\}$$

$$\times\exp\left\{-\frac{1}{2\eta^2}\left(\mathbf{A}_{h[k,]}{}'-\mathbf{A}^{0}_{[k,]}{}'\right)'\Sigma^{0}_{\mathbf{A}_k}{}^{-1}\left(\mathbf{A}_{h[k,]}{}'-\mathbf{A}^{0}_{[k,]}{}'\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]}-\mathbf{a}_{0h}{}'-\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,-k]}\mathbf{A}_{h[-k,]}-\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\mathbf{A}_{h[k,]}\right)\right.$$

$$\left.\mathbf{\Sigma}_h{}^{-1}\left(\mathbf{Y}_{[I(C_i=h),]}-\mathbf{a}_{0h}{}'-\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,-k]}\mathbf{A}_{h[-k,]}-\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\mathbf{A}_{h[k,]}\right)'\right\}$$

$$\times\exp\left\{-\frac{1}{2\eta^2}\left(\mathbf{A}_{h[k,]}{}'-\mathbf{A}^{0}_{[k,]}{}'\right)'\Sigma^{0}_{\mathbf{A}_k}{}^{-1}\left(\mathbf{A}_{h[k,]}{}'-\mathbf{A}^{0}_{[k,]}{}'\right)\right\}$$

Thus,

$$\pi\left(\mathbf{A}_{h[k,]} \mid \mathbf{A}_{h[-k,]},\lambda_h,\mathbf{C},\mathbf{W}_h,\mathbf{a}_{\mathbf{0}h},\mathbf{\Sigma}_h,H,\mathbf{X},\mathbf{Y},\tau^2,\eta^2,\sigma^2\right)$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}^{-k}_{[I(C_i=h),]}-\mathbf{W}_{h[,k]}{}'\mathbf{X}_{[I(C_i=h),]}{}'\mathbf{A}_{h[k,]}\right)\mathbf{\Sigma}_h{}^{-1}\left(\mathbf{Y}^{-k}_{[I(C_i=h),]}-\mathbf{W}_{h[,k]}{}'\mathbf{X}_{[I(C_i=h),]}{}'\mathbf{A}_{h[k,]}\right)'\right\}$$

$$\times\exp\left\{-\frac{1}{2\eta^2}\left(\mathbf{A}_{h[k,]}{}'-\mathbf{A}^{0}_{[k,]}{}'\right)'\Sigma^{0}_{\mathbf{A}_k}{}^{-1}\left(\mathbf{A}_{h[k,]}{}'-\mathbf{A}^{0}_{[k,]}{}'\right)\right\}$$

$$\propto \exp\left[-\frac{1}{2}\left\{\mathbf{A}_{h[k,]}\left(\sum_{i=1}^{N}\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\boldsymbol{\Sigma}_h^{-1}\mathbf{W}_{h[,k]}'\mathbf{X}_{[I(C_i=h),]}' + \frac{1}{\eta^2}\boldsymbol{\Sigma}_{\mathbf{A}_k}^{0\ -1}\right)\mathbf{A}_{h[k,]}' \right.\right.$$
$$\left.\left. -2\mathbf{A}_{h[k,]}\left(\sum_{l=1}^{N}\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\boldsymbol{\Sigma}_h^{-1}\mathbf{Y}_{[I(C_i=h),]}^{-k\ \prime} + \frac{1}{\eta^2}\boldsymbol{\Sigma}_{\mathbf{A}_k}^{0\ -1}\mathbf{A}_{[k,]}^{0}\right)\right\}\right]$$

Therefore,

$$\mathbf{A}_{h[k,]}' \mid \mathbf{A}_{h[-k,]}, \lambda_h, \mathbf{C}, \mathbf{W}_h, \mathbf{a}_{\mathbf{0}h}, \boldsymbol{\Sigma}_h, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2 \sim N_Q\left(\widetilde{\mathbf{A}}_{h[k,]}',\ \boldsymbol{\Sigma}_{\mathbf{A}_{hk}}\right)$$

where $\boldsymbol{\Sigma}_{\mathbf{A}_{hk}} = \left(\sum_{i=1}^{N}\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\boldsymbol{\Sigma}_h^{-1}\mathbf{W}_{h[,k]}'\mathbf{X}_{[I(C_i=h),]}' + \frac{1}{\eta^2}\boldsymbol{\Sigma}_{\mathbf{A}_k}^{0\ -1}\right)^{-1}$ and

$$\widetilde{\mathbf{A}}_{h[k,]}' = \boldsymbol{\Sigma}_{\mathbf{A}_{hk}}\left(\sum_{l=1}^{N}\mathbf{X}_{[I(C_i=h),]}\mathbf{W}_{h[,k]}\boldsymbol{\Sigma}_h^{-1}\mathbf{Y}_{[I(C_i=h),]}^{-k\ \prime} + \frac{1}{\eta^2}\boldsymbol{\Sigma}_{\mathbf{A}_k}^{0\ -1}\mathbf{A}_{[k,]}^{0}\right).$$

- conditional posterior distribution of $\mathbf{a}_{\mathbf{0}h}$: for $h = 1, \cdots, H$

$$\pi\left(\mathbf{a}_{\mathbf{0}h} \mid \mathbf{A}_h, \lambda_h, \mathbf{C}, \mathbf{W}_h, \boldsymbol{\Sigma}_h, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right)$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]}' - \mathbf{a}_{oh} - \mathbf{A}_h'\mathbf{W}_h'\mathbf{X}_{[I(C_i=h),]}'\right)'\boldsymbol{\Sigma}_h^{-1}\left(\mathbf{Y}_{[I(C_i=h),]}' - \mathbf{a}_{oh} - \mathbf{A}_h'\mathbf{W}_h'\mathbf{X}_{[I(C_i=h),]}'\right)\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{a}_{oh} - \boldsymbol{\alpha}_0\right)'\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\left(\mathbf{a}_{oh} - \boldsymbol{\alpha}_0\right)\right\}$$
$$\propto \exp\left[-\frac{1}{2}\left\{\mathbf{a}_{oh}'\left(N_h\boldsymbol{\Sigma}_h^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\right)\mathbf{a}_{oh}\right.\right.$$
$$\left.\left. +2\mathbf{a}_{oh}'\left(\boldsymbol{\Sigma}_h^{-1}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]} - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_h\mathbf{A}_h\right)' + \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\boldsymbol{\alpha}_0\right)\right\}\right]$$

Therefore,

$$\mathbf{a}_{\mathbf{0}h} \mid \mathbf{A}_h, \lambda_h, \mathbf{C}, \mathbf{W}_h, \boldsymbol{\Sigma}_h, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2 \sim N_Q\left(\tilde{\mathbf{a}}_{0h},\ \left(N_h\boldsymbol{\Sigma}_h^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\right)^{-1}\right)$$

where $\tilde{\mathbf{a}}_0 = \left(N_h\boldsymbol{\Sigma}_h^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_h^{-1}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]} - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_h\mathbf{A}_h\right)' + \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\mathbf{a}_0}^{-1}\boldsymbol{\alpha}_0\right).$

- conditional posterior distribution of $\boldsymbol{\Sigma}_h^{-1}$: for $h = 1, \cdots, H$

$$\pi\left(\boldsymbol{\Sigma}_h \mid \mathbf{A}_h, \lambda_h, \mathbf{C}, \mathbf{W}_h, \mathbf{a}_{\mathbf{0}h}, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right)$$
$$\propto \left|\boldsymbol{\Sigma}_h\right|^{-N_h/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_{[I(C_i=h),]}{}' - \mathbf{a}_{oh} - \mathbf{A}_h{}'\mathbf{W}_h{}'\mathbf{X}_{[I(C_i=h),]}{}'\right)' \boldsymbol{\Sigma}_h^{-1}\left(\mathbf{Y}_{[I(C_i=h),]}{}' - \mathbf{a}_{oh} - \mathbf{A}_h{}'\mathbf{W}_h{}'\mathbf{X}_{[I(C_i=h),]}{}'\right)\right\}$$
$$\times \left|\boldsymbol{\Sigma}_h^{-1}\right|^{\frac{1}{2}(\nu-Q-1)} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_h^{-1}\right)\right\}$$

Let $\mathbf{E}_{h[I(C_i=h),]} = \mathbf{Y}_{[I(C_i=h),]} - \mathbf{a}_{oh}{}' - \mathbf{X}_{[I(C_i=h),]}\mathbf{W}_h\mathbf{A}_h$, then

$$\pi\left(\boldsymbol{\Sigma}_h \mid \mathbf{A}_h, \lambda_h, \mathbf{C}, \mathbf{W}_h, \mathbf{a}_{\mathbf{0}h}, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2\right)$$
$$\propto \left|\boldsymbol{\Sigma}_h^{-1}\right|^{\frac{1}{2}(\nu+N\_\mathrm{h}-Q-1)} \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{N}\mathbf{E}_{h[I(C_i=h),]}\boldsymbol{\Sigma}_h^{-1}\mathbf{E}_{h[I(C_i=h),]}{}' + \operatorname{tr}\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_h^{-1}\right)\right\}\right]$$
$$= \left|\boldsymbol{\Sigma}_h^{-1}\right|^{\frac{1}{2}(\nu+N\_\mathrm{h}-Q-1)} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\left\{\sum_{i=1}^{N}\mathbf{E}_{h[I(C_i=h),]}{}'\mathbf{E}_{h[I(C_i=h),]} + \boldsymbol{\Sigma}_0\right\}\boldsymbol{\Sigma}_h^{-1}\right)\right]$$

Therefore,

$$\boldsymbol{\Sigma}_h \mid \mathbf{A}_h, \lambda_h, \mathbf{C}, \mathbf{W}_h, \mathbf{a}_{\mathbf{0}h}, H, \mathbf{X}, \mathbf{Y}, \tau^2, \eta^2, \sigma^2 \sim Wishart_Q\left(\nu + N_h,\ \boldsymbol{\Sigma}_0 + \sum_{i=1}^{N}\mathbf{E}_{h[I(C_i=h),]}{}'\mathbf{E}_{h[I(C_i=h),]}\right).$$

R code for the simulation study

```
rm(list=ls())

library(mvtnorm)

library(tmvtnorm)

library(MCMCpack)

library(msm)


# sample size

ns<-300


## true parameters for two subpopulations

# subpopulation 1

tr.W1<-matrix(c(0.6, 0.4, 0, 0, 0, 0, 1, 0), 4, 2)

tr.A1<-matrix(c(2, 3, -2, 0), 2, 2)

tr.Sig1<-matrix(c(0.2,sqrt(0.2)*0.4,sqrt(0.2)*0.4,1), 2,2)

tr.int1<-c(-1,0)


# subpopulation 2

tr.W2<-matrix(c(0.6, 0.4, 0, 0, 0, 0, 0.2, 0.8), 4, 2)

tr.A2<-matrix(c(1, 1.5, 3, 1), 2, 2)

tr.Sig2<-matrix(c(1,sqrt(1)*0.1,sqrt(1)*0.1,1), 2,2)

tr.int2<-c(3,0)


# Covariance matrix for four predictors

tr.Sig.Xs<-matrix(c(1, 0.3, 0.1, 0.1, 0.3, 1, 0.1, 0.1, 0.1, 0.1, 1, 0.6, 0.1, 0.1,
            0.6, 1), 4, 4)



## generate the predictors and outcome variable.

Xs<-rmvnorm(ns, mean=rep(0, 4), sigma=tr.Sig.Xs)

## column-wise standardization

Xs<-(Xs-matrix(apply(Xs,2,mean),ns,4,byrow=T))/matrix(apply(Xs,2,sd),ns,4,byrow=T)


ys<-matrix(0, ns, 2)
```

```
tr.sub<-rep(NA, ns)
for (i in 1:ns){
tr.sub[i]<-Z.tr<-rbinom(1,1,prob=0.4)
if (Z.tr==0){
ys[i,]<-tr.int1+Xs[i,]%*%tr.W1%*%tr.A1+mvrnorm(1, mu=rep(0, nrow(tr.Sig1)),
        Sigma=tr.Sig1)
}else{
ys[i,]<-tr.int2+Xs[i,]%*%tr.W2%*%tr.A2+mvrnorm(1, mu=rep(0, nrow(tr.Sig2)),
        Sigma=tr.Sig2)
}
}

N<-dim(ys)[1]       # no. of observations
Q<-dim(ys)[2]       # no. of responses
P<-dim(Xs)[2]       # no. of independent variables
K<-2                # no. latent variables
H<-2                # no. mixture components

## a matrix indicating which predictors contribute latent components
indx.Ws<-matrix(c(1,1,0,0,0,0,1,1),P, K)



### hyperparameters
cW<-10              # multiplicative constant for the prior variance of matrix W
cA<-100                    # multiplicative constant for the prior variance of matrix A
E.As<-rep(0, K+1)   # prior mean for regression coefficents A include the intercept
V.As<-cA*diag(K+1)  # prior variance for regression coefficents A include
                      # the intercept
E.Ws<-rep(0, P)     # prior mean for weights W
V.Ws<-cW*diag(P)    # prior mean for weights W

df.S<-Q+1           # degrees of freedom for a Wishart distribution for the inverse
                      # of a residual covariance matrix S
V.S<-10/df.S*diag(Q)   # scale matrix for a Wishart distribution for the inverse of
```

```
                        # a residual covariance matrix S

alps<-rep(1/2, H)    # concentration parameters for a Dirichlet distribution
                     # for mixture weights ps

### initial values for the Gibbs algorithm
Ss<-diag(Q)*1000
Ss.inv<-solve(Ss)
Ss<-array(Ss, dim=c(Q, Q, H))
Ss.inv<-array(Ss.inv, dim=c(Q, Q, H))

As<-array(t(rmvnorm(Q*H, mean=E.As, sigma=V.As)), dim=c(K+1, Q, H))
Ws<-array(t(rmvnorm(K*H, mean=E.Ws, sigma=V.Ws)), dim=c(P, K, H))
Ws[indx.Ws==0]<-0

bt0s<-matrix(0, Q, H)
bt0s.mat<-array(rep(bt0s, each=N), dim=c(N, Q, H))
As2<-As[-1,,]

loc.H<-rbinom(N, H-1, prob=c(0.4, 0.3))+1
nh<-table(loc.H)
ps<-rep(1/H, H)

### number of iterations for the Gibbs algorithm
nitr<-30000

post.Ws<-array(0, dim=c(P,K,H,nitr))
post.As<-array(0, dim=c(K,Q,H,nitr))
post.Bs<-array(0, dim=c(P,Q,H,nitr))
post.bt0<-array(0, dim=c(Q,H,nitr))
post.R<-array(0, dim=c(Q*(Q+1)/2, H, nitr))
post.ps<-array(0, dim=c(H, nitr))
post.k<-rep(0, nitr)
post.loc<-array(0, dim=c(N, nitr))
```

```
### Gibbs sampling begins here.

for (ni in 1:nitr){



for (h in 1:H){

## update matrix W

if (nh[h]>0){

for (k in 1:K){

if (sum(indx.Ws[,k]==1)>1){

post.var.Wk<-
solve(t(Xs[loc.H==h,indx.Ws[,k]==1])%*%Xs[loc.H==h,indx.Ws[,k]==1]*as.numeric(As2[k,,h]%*
%Ss.inv[,,h]%*%As2[k,,h])+solve(V.Ws[indx.Ws[,k]==1,indx.Ws[,k]==1]))

if (K>2){

post.mean.Wk<-post.var.Wk%*%t(Xs[loc.H==h,indx.Ws[,k]==1])%*%((ys[loc.H==h,]-
matrix(bt0s[,h], nh[h], Q, byrow=T)-Xs[loc.H==h,]%*%(Ws[,-k,h]%*%As2[-k,,h])))%*%

(Ss.inv[,,h])%*%As2[k,,h]

}else{

if (nh[h]>1){

post.mean.Wk<-post.var.Wk%*%t(Xs[loc.H==h,indx.Ws[,k]==1])%*%((ys[loc.H==h,]-
matrix(bt0s[,h], nh[h], Q, byrow=T)-Xs[loc.H==h,]%*%(Ws[,-k,h]%o%As2[-k,,h])))%*%

(Ss.inv[,,h])%*%As2[k,,h]

}else{

post.mean.Wk<-post.var.Wk%*%Xs[loc.H==h,indx.Ws[,k]==1]%*%((ys[loc.H==h,]-
matrix(bt0s[,h], nh[h], Q, byrow=T)-Xs[loc.H==h,]%*%(Ws[,-k,h]%o%As2[-k,,h])))%*%

(Ss.inv[,,h])%*%As2[k,,h]

}

}

Ws[indx.Ws[,k]==1,k,h]<-t(rmvnorm(1, mean=as.vector(post.mean.Wk), sigma=post.var.Wk

))

}else{Ws[indx.Ws[,k]==1,k,h]<-1}                # if there is only one predictor for a
latent variable.

# normalization to satisfy the constrant that t(Xs%*%Ws[,k])%*%(Xs%*%Ws[,k])=N.

Ws[,k,h]<-
Ws[,k,h]*as.numeric(sqrt(nh[h])/sqrt(t(Xs[loc.H==h,]%*%Ws[,k,h])%*%(Xs[loc.H==h,]%*%Ws[,k
,h])))
```

```
}

## update matrix A including the intercept

if (nh[h]>1){

Fs<-cbind(1, Xs[loc.H==h,]%*%Ws[,,h])

tmp.XX<-matrix(apply(sapply(1:nh[h],
function(x){t(kronecker(diag(Q),matrix(Fs[x,],1)))%*%Ss.inv[,,h]%*%(kronecker(diag(Q),mat
rix(Fs[x,],1)))}), 1, sum), dim(Fs)[2]*Q)

tmp.XZ<-apply(sapply(1:nh[h],
function(x){t(kronecker(diag(Q),matrix(Fs[x,],1)))%*%Ss.inv[,,h]%*%ys[loc.H==h,][x,]}),1,
sum)

}else{

Fs<-cbind(1, Xs[loc.H==h,]%*%Ws[,,h])

tmp.XX<-
t(kronecker(diag(Q),matrix(Fs,1)))%*%Ss.inv[,,h]%*%(kronecker(diag(Q),matrix(Fs,1)))

tmp.XZ<-apply(t(kronecker(diag(Q),matrix(Fs,1)))%*%Ss.inv[,,h]%*%ys[loc.H==h,],1,sum)

}

As[,,h]<-matrix(rmvnorm(1,
mean=as.numeric(solve(tmp.XX+solve(kronecker(diag(Q),V.As)))%*%(tmp.XZ+solve(kronecker(di
ag(Q),V.As))%*%rep(E.As, Q))), sigma=solve(tmp.XX+solve(kronecker(diag(Q),

V.As)))), K+1)


## update covariance matrix S

if (nh[h]>1){

tmp.ee<-matrix(apply(sapply(1:nh[h], function(x){t(ys[loc.H==h,][x,]-Fs[x,]%*%

As[,,h])%*%(ys[loc.H==h,][x,]-Fs[x,]%*%As[,,h])}), 1, sum), Q)

}else{

tmp.ee<-t(ys[loc.H==h,]-Fs%*%As[,,h])%*%(ys[loc.H==h,]-Fs%*%As[,,h])

}

Ss.inv[,,h]<-rWishart(1, df.S+nh[h], solve(solve(V.S)+tmp.ee))[,,1]

Ss[,,h]<-chol2inv(chol(Ss.inv[,,h]))

}else{

for (k in 1:K){

Ws[indx.Ws[,k]==1,k,h]<-t(rmvnorm(1, mean=E.Ws[indx.Ws[,k]==1],
sigma=V.Ws[indx.Ws[,k]==1,indx.Ws[,k]==1]))

}
```

```
As[,,h]<-t(rmvnorm(Q, mean=E.As, sigma=V.As))
Ss.inv[,,h]<-rWishart(1, df.S, V.S)[,,1]
Ss[,,h]<-chol2inv(chol(Ss.inv[,,h]))
}
}

## divide As into two parts: bt0s and As
bt0s<-As[1,,]
As2<-As[-1,,]

## update mixture weights ps
loc.H<-sapply(1:N, function(x){
tmp.d<-sapply(1:H, function(y){dmvnorm(ys[x,], mean=bt0s[,y]+as.vector(Xs[x,]%*%
Ws[,,y]%*%As2[,,y]), sigma=Ss[,,y], log=T)})
tmp.d<-tmp.d-max(tmp.d)
as.vector(rmultinom(1,1, prob=(exp(tmp.d)*ps)/sum(exp(tmp.d)*ps)))%*%(1:H)})
nh<-apply(matrix(loc.H, N, H)==matrix(1:H, N, H, byrow=T), 2, sum)
ps<-rdirichlet(1,alps+nh)

## save MCMC samples
post.Ws[,,,ni]<-Ws
post.As[,,,ni]<-As2
post.Bs[,,,ni]<-sapply(1:H, function(z){Ws[,,z]%*%As2[,,z]})
post.bt0[,,ni]<-bt0s
post.R[,,ni]<-sapply(1:H, function(z){Ss[,,z][lower.tri(Ss[,,z], diag=T)]})
post.ps[,ni]<-ps
post.k[ni]<-length(unique(loc.H))
post.loc[,ni]<-loc.H
}
```