Implementing the Proposed Method With R

All the methods proposed in the present article have been implemented in R (R Development Core Team, 2013). This appendix briefly discusses how to use those R codes to analyze the emotion data. The input values that are included in this code may be changed by the users according to their own data.

List of R files and their function:

- main.R: Main program to run.
- def_con.R: Define the sample size, number of observed variables, number of factors, number of burn-in, number of MCMC samples for inference, etc.
- def_rec.R: Allocation of computation variables.
- Gibbs.R: Details of Gibbs sampler steps, including Gibbs_Omega.R, Gibbs_LY.R, Gibbs_MU.R, Gibbs_PHI.R, Gibbs_PSX.R, Gibbs_MISY.R and Postp.R, where
  - Gibbs_Omega.R: Generate $\mathbf{\Omega}$ from $p(\mathbf{\Omega}|\mathbf{Y}, \boldsymbol{\mu}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi})$;
  - Gibbs_LY.R: Generate $\mathbf{\Lambda}$ from $p(\mathbf{\Lambda}|\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Psi})$;
  - Gibbs_MU.R: Generate $\boldsymbol{\mu}$ from $p(\boldsymbol{\mu}|\mathbf{Y}, \mathbf{\Omega}, \mathbf{\Phi}, \mathbf{\Lambda}, \mathbf{\Psi})$;
  - Gibbs_PHI.R: Generate $\mathbf{\Phi}$ from $p(\mathbf{\Phi}|\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Lambda}, \mathbf{\Psi})$;
  - Gibbs_PSX.R: Generate $\mathbf{\Sigma}$ from $p(\mathbf{\Sigma}|\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Lambda}, \mathbf{\Phi}, \boldsymbol{\tau}, \lambda)$ and compute $\mathbf{\Psi} = \mathbf{\Sigma}^{-1}$.
  - Gibbs_MISY.R: Generate the missing response from its conditional distributions to deal with missing responses (if any).
  - Postp.R: Calculate the posterior predictive (PP) p-value.

Note: The derivation of the above-mentioned conditional distributions will be presented in the next section.

- ind.R: Define the position of fixed and unknown parameters in factor loading matrix.

- init1.R: Set initial values for the unknown parameters.

- Prior.R: Set hyperparameters of prior distribution.

- read_observed.R: Read the data to be analyzed.

- write_model fit.R: Write the posterior predictive p-value into a text file and save in "Result\Est\PP.txt".

- write_result.R: Write the parameter estimates into separate text files and save in "Result\Est\". Write the standard error estimates into separate text files and save in "Result\SE\".

- HPD.R: Calculate the HPD intervals, write them into separate text files and save in "Result\Est\".

Once R has been downloaded and installed, the data can be analyzed step by step as follows:

1. Put all the files of R codes and data (e.g. emotion.txt) in an appropriate directory (e.g., E:\R code\) and set this directory in main.R by using setwd("E:\R code\"). Some packages are required to be installed into R: msm, MCMCpack, statmod, and psychometric. If any of these packages doesn't exist, it will be installed automatically by the commands written in main.R.

2. Prepare the data file and replace the missing responses with NA. In read_observed.R, change the name of the data file (e.g. emotion.txt).

3. In def_con.R, change the settings according to model and data. For example, in the emotion data analysis, sample size was 175, number of items was 28, number of factors was

5. So we can set N←175, NY←28, NZ←5. Similarly, we can set the numbers according to the MCMC procedure, e.g., number of burn-in and so on.

4. In ind.R, set the position of fixed and unknown parameters according to model specification and factor structure. IDY is a NY×NZ indicator matrix where 0 represents the fixed parameters and 1 represents the unknown parameters needed to be estimated in $\mathbf{\Lambda}$.

5. In Prior.R, set values of hyperparameters according to prior input.

6. In init1.R, set initial values for the unknown parameters.

7. Run the file main.R.

9. All the results are saved in "Result\Est\" and "Result\SE\". Names of parameter estimates, standard error estimates and HPD intervals begin with Em, SE and HPD respectively. Note: LY, MU, PHI and PSX present the unknown parameters in $\mathbf{\Lambda}$, $\boldsymbol{\mu}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ respectively.

Derivation of Conditional Distributions

The specific prior distributions for inverse of the covariance matrix can be expressed as the product of double exponential distributions for the off-diagonal elements and exponential distributions for the diagonal elements:

$$p(\mathbf{\Sigma}|\lambda) \propto \prod_{i<j} \{\mathrm{DE}(\sigma_{ij}|\lambda)\} \prod_{i=1}^{p} \left\{\mathrm{EXP}(\sigma_{ii}|\frac{\lambda}{2})\right\} \mathbf{I}(\mathbf{\Sigma} \succ 0), \tag{1}$$

where $\mathrm{DE}(\sigma_{ij}|\lambda)$ is the double exponential density function of the form $\frac{\lambda}{2}\exp(-\lambda|\sigma_{ij}|)$, and $\mathrm{EXP}(\sigma_{ii}|\frac{\lambda}{2})$ is the exponential density function of the form $\frac{\lambda}{2}\exp(-\frac{\lambda}{2}\sigma_{ii})\mathbf{I}(\sigma_{ii} > 0)$ where $\mathbf{I}(\sigma_{ii} > 0)$ is an indicator function of the event that $\sigma_{ii} > 0$, and $\mathbf{I}(\mathbf{\Sigma} \succ 0)$ is an indicator function of the event that $\mathbf{\Sigma}$ is positive definite.

The double exponential distribution can be represented as a scale mixture of normals (Hans, 2009; Park & Casella, 2008; Wang, 2012). Let $\boldsymbol{\sigma} = (\sigma_{ij})_{i\leq j}$ be the vector of the upper off-diagonal and diagonal elements of $\mathbf{\Sigma}$, and $\boldsymbol{\tau} = (\tau_{ij})_{i<j}$ be the vector of latent scale parameters. The density for $\boldsymbol{\sigma}$ given $\boldsymbol{\tau}$ and λ is given by:

$$p(\boldsymbol{\sigma}|\boldsymbol{\tau}, \lambda) = \frac{1}{C_{\boldsymbol{\tau}}} \prod_{i<j} \left\{ \frac{1}{\sqrt{2\pi\tau_{ij}}} \exp(-\frac{\sigma_{ij}^2}{2\tau_{ij}}) \right\} \prod_{i=1}^{p} \left\{ \frac{\lambda}{2} \exp(-\frac{\lambda}{2}\sigma_{ii}) \right\} \mathbf{I}(\mathbf{\Sigma} \succ 0), \qquad (2)$$

where the normalizing term $C_{\boldsymbol{\tau}}$ depends on $\boldsymbol{\tau}$. The following prior distribution is proposed for $\boldsymbol{\tau}$,

$$p(\boldsymbol{\tau}|\lambda) \propto C_{\boldsymbol{\tau}} \prod_{i<j} \frac{\lambda^2}{2} \exp(-\frac{\lambda^2}{2}\tau_{ij}). \qquad (3)$$

Wang (2012) proved that $p(\boldsymbol{\sigma}|\boldsymbol{\tau}, \lambda)$ and $p(\boldsymbol{\tau}|\lambda)$ are proper priors, and the analytically intractable terms $C_{\boldsymbol{\tau}}$ in Eqs. 2 and 3 can be canceled out so that the marginal distribution of the σ_{ij}'s follows Eq. 1. However, because of the positive definite constraint, the normal distributions for the σ_{ij}'s are dependent given the scale parameters τ_{ij}. To solve this problem, a data-augmented block Gibbs sampling scheme proposed by Wang (2012) is considered for posterior computations under the hierarchical frameworks represented by Eqs. 2 and 3.

For the conditional distribution $p(\mathbf{\Sigma}|\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Lambda}, \mathbf{\Phi}, \boldsymbol{\tau}, \lambda)$, we describe the operationalization of the required Gibbs sampler under the Bayesian Lasso prior. To simplify notation, denote the vector that contains all unknown distinct parameters in $\boldsymbol{\mu}$, $\mathbf{\Lambda}$, and $\mathbf{\Phi}$ by $\boldsymbol{\theta}$.

The conditional distribution $p(\mathbf{\Sigma}|\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \boldsymbol{\tau}, \lambda)$ can be decomposed as follows:

$$\begin{aligned} p(\mathbf{\Sigma}|\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \boldsymbol{\tau}, \lambda) &\propto p(\mathbf{Y}|\mathbf{\Omega}, \boldsymbol{\theta}, \mathbf{\Sigma}) p(\mathbf{\Sigma}|\boldsymbol{\tau}, \lambda) \\ &\propto |\mathbf{\Sigma}|^{\frac{n}{2}} \exp\left[-\frac{1}{2} tr(\tilde{\mathbf{S}}\mathbf{\Sigma})\right] \prod_{i<j} \left\{ \exp(-\frac{\sigma_{ij}^2}{2\tau_{ij}}) \right\} \\ &\quad \times \prod_{i=1}^{p} \left\{ \exp(-\frac{\lambda}{2}\sigma_{ii}) \right\} \mathbf{I}(\mathbf{\Sigma} \succ 0), \end{aligned} \qquad (4)$$

where $\tilde{\mathbf{S}} = \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{\Lambda}\boldsymbol{\omega}_i)(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{\Lambda}\boldsymbol{\omega}_i)^T$. Following Wang (2012), we consider an efficient block Gibbs sampling to update $\mathbf{\Sigma}$ one column at a time after appropriate reparametrization.

Denote the running index for observed measurements by k. For $k = 1, 2, \cdots, p$, partition and rearrange the columns of $\mathbf{\Sigma}$ and $\tilde{\mathbf{S}}$ as follows:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{-kk} & \boldsymbol{\sigma}_k \\ \boldsymbol{\sigma}_k^T & \sigma_{kk} \end{pmatrix}, \quad \tilde{\mathbf{S}} = \begin{pmatrix} \tilde{\mathbf{S}}_{-kk} & \tilde{\mathbf{s}}_k \\ \tilde{\mathbf{s}}_k^T & \tilde{s}_{kk} \end{pmatrix},$$

where σ_{kk} is the kth diagonal element of $\mathbf{\Sigma}$, $\boldsymbol{\sigma}_k = (\sigma_{k1}, \cdots, \sigma_{k,k-1}, \sigma_{k,k+1}, \cdots, \sigma_{kp})^T$ is the vector of all off-diagonal elements of the kth column, and $\mathbf{\Sigma}_{-kk}$ is the $(p-1) \times (p-1)$ matrix resulting from deleting the kth row and kth column from $\mathbf{\Sigma}$. Similarly, $\tilde{s}_{kk}$ is the kth diagonal element of $\tilde{\mathbf{S}}$, $\tilde{\mathbf{s}}_k$ is the vector of all off-diagonal elements of the kth column of $\tilde{\mathbf{S}}$, and $\tilde{\mathbf{S}}_{-kk}$ is the matrix with the kth row and kth column of $\tilde{\mathbf{S}}$ deleted. Therefore we have:

$$\begin{aligned} p(\boldsymbol{\sigma}_k, \sigma_{kk} | \mathbf{\Sigma}_{-kk}, \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \boldsymbol{\tau}, \lambda) \quad &\propto \quad (\sigma_{kk} - \boldsymbol{\sigma}_k^T \mathbf{\Sigma}_{-kk}^{-1} \boldsymbol{\sigma}_k)^{\frac{n}{2}} \\ &\quad \times \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\sigma}_k^T \mathbf{M}_{\boldsymbol{\tau}}^{-1} \boldsymbol{\sigma}_k + 2\tilde{\mathbf{s}}_k^T \boldsymbol{\sigma}_k + (\tilde{s}_{kk} + \lambda)\sigma_{kk} \right] \right\}, \end{aligned} \tag{5}$$

where $\mathbf{M}_{\boldsymbol{\tau}}$ is the diagonal matrix with diagonal elements $\tau_{k1}, \cdots, \tau_{k,k-1}, \tau_{k,k+1}, \cdots, \tau_{kp}$. Let $\boldsymbol{\beta} = \boldsymbol{\sigma}_k$ and $\gamma = \sigma_{kk} - \boldsymbol{\sigma}_k^T \mathbf{\Sigma}_{-kk}^{-1} \boldsymbol{\sigma}_k$. It can be shown that:

$$\boldsymbol{\beta} | \mathbf{\Sigma}_{-kk}, \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \boldsymbol{\tau}, \lambda \quad \sim \quad N(-\left[(\tilde{s}_{kk} + \lambda)\mathbf{\Sigma}_{-kk}^{-1} + \mathbf{M}_{\boldsymbol{\tau}}^{-1}\right]^{-1} \tilde{\mathbf{s}}_k, \left[(\tilde{s}_{kk} + \lambda)\mathbf{\Sigma}_{-kk}^{-1} + \mathbf{M}_{\boldsymbol{\tau}}^{-1}\right]^{-1}), \tag{6}$$

$$\gamma | \mathbf{\Sigma}_{-kk}, \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \boldsymbol{\tau}, \lambda \quad \sim \quad Gamma(\frac{n}{2} + 1, \frac{\tilde{s}_{kk} + \lambda}{2}). \tag{7}$$

After simulating observations from the above conditional distributions, we can obtain $\boldsymbol{\sigma}_k = \boldsymbol{\beta}$, $\boldsymbol{\sigma}_k^T = \boldsymbol{\beta}^T$ and $\sigma_{kk} = \gamma + \boldsymbol{\sigma}_k^T \mathbf{\Sigma}_{-kk}^{-1} \boldsymbol{\sigma}_k$, then the last column and row of $\mathbf{\Sigma}$ are updated at a time. At the end, $\mathbf{\Psi} = \mathbf{\Sigma}^{-1}$ is computed.

The conditional distribution $p(\boldsymbol{\tau} | \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \mathbf{\Sigma}, \lambda)$ can be expressed as follows:

$$\begin{aligned} p(\boldsymbol{\tau} | \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta}, \mathbf{\Sigma}, \lambda) \quad &\propto \quad p(\mathbf{\Sigma} | \boldsymbol{\tau}, \lambda) p(\boldsymbol{\tau} | \lambda) \\ &\propto \quad \prod_{i<j} \tau_{ij}^{-\frac{1}{2}} \exp\left\{ -\frac{\sigma_{ij}^2 + \lambda^2 \tau_{ij}^2}{2\tau_{ij}} \right\}. \end{aligned} \tag{8}$$

It can be shown that for $i < j$,

$$\frac{1}{\tau_{ij}}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \lambda \sim IG(\sqrt{\frac{\lambda^2}{\sigma_{ij}^2}}, \lambda^2), \tag{9}$$

where $IG(a, b)$ indicates the inverse-Gaussian distribution with mean a and shape parameter b.

Additionally, it can be shown that the conditional distribution $p(\lambda|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\tau})$ follows:

$$\lambda|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\tau} \sim Gamma(\alpha_{\lambda 0} + \frac{p(p+1)}{2}, \beta_{\lambda 0} + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}|\sigma_{ij}|). \tag{10}$$

Let $\boldsymbol{\Lambda}_k^T$ be the kth row of $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}_{-k}$ be submatrix of $\boldsymbol{\Lambda}$ with the kth row deleted, $\mathbf{Y}_k$ be the submatrix of $\mathbf{Y}$ only with the kth row, and $\mathbf{Y}_{-k}$ be the submatrix of $\mathbf{Y}$ with the kth row deleted. As $\boldsymbol{\Psi}$ is not diagonal, the conditional distribution of $\boldsymbol{\Lambda}_k$ can be expressed as:

$$p(\boldsymbol{\Lambda}_k|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}_{-k}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \propto p(\mathbf{Y}_k|\mathbf{Y}_{-k}, \boldsymbol{\Lambda}_k, \boldsymbol{\Lambda}_{-k}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi})p(\boldsymbol{\Lambda}_k), \tag{11}$$

for $k = 1, \cdots, p$. Without a loss of generality, we partition and rearrange the columns of $\boldsymbol{\Psi}$ as follows:

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{-kk} & \boldsymbol{\psi}_k \\ \boldsymbol{\psi}_k^T & \psi_{kk} \end{pmatrix},$$

where ψ_{kk} is the kth diagonal element of $\boldsymbol{\Psi}$, $\boldsymbol{\psi}_k = (\psi_{k1}, \cdots, \psi_{k,k-1}, \psi_{k,k+1}, \cdots, \psi_{kp})^T$ is the vector of all off-diagonal elements of the kth column, and $\boldsymbol{\Psi}_{-kk}$ is the $(p-1) \times (p-1)$ matrix resulting from deleting the kth row and kth column from $\boldsymbol{\Psi}$. It can be shown that:

$$\begin{aligned} \boldsymbol{\Lambda}_k|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}_{-k}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi} \quad \sim \quad & N[(\boldsymbol{\Psi}_k^{*-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \mathbf{H}_{0k}^{-1})^{-1}(\boldsymbol{\Psi}_k^{*-1}\boldsymbol{\Omega}\mathbf{Y}_k^{*T} + \mathbf{H}_{0k}^{-1}\boldsymbol{\Lambda}_{0k}), \\ & (\boldsymbol{\Psi}_k^{*-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \mathbf{H}_{0k}^{-1})^{-1}], \end{aligned} \tag{12}$$

where $\boldsymbol{\Psi}_k^* = \psi_{kk} - \boldsymbol{\psi}_k^T\boldsymbol{\Psi}_{-kk}^{-1}\boldsymbol{\psi}_k$ and $\mathbf{Y}_k^*$ is the matrix with element $y_{ik}^* = y_{ik} - \mu_k - \boldsymbol{\psi}_k^T\boldsymbol{\Psi}_{-kk}^{-1}(\mathbf{y}_{-ik} - \boldsymbol{\mu}_{-k} - \boldsymbol{\Lambda}_{-k}\boldsymbol{\omega}_i)$, $\mathbf{y}_{-ik}$ is the vector of $\mathbf{y}_i$ with the kth element deleted, and $\boldsymbol{\mu}_{-k}$ is the vector of $\boldsymbol{\mu}$ with the kth element deleted.

The conditional distributions $p(\boldsymbol{\Omega}|\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$, $p(\boldsymbol{\Phi}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$, and $p(\boldsymbol{\mu}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$ are similar to those presented in Lee (2007):

$$\boldsymbol{\omega}_i|\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi} \sim N((\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}), (\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}), \quad (13)$$

$$\boldsymbol{\Phi}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi} \sim IW(\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \mathbf{R}_0^{-1}, n + \rho_0), \quad (14)$$

$$\boldsymbol{\mu}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi} \sim N((\mathbf{H}_{\mu 0}^{-1} + n\boldsymbol{\Psi}^{-1})^{-1}(\boldsymbol{\Psi}^{-1}\mathbf{V} + \mathbf{H}_{\mu 0}^{-1}\boldsymbol{\mu}_0), (\mathbf{H}_{\mu 0}^{-1} + n\boldsymbol{\Psi}^{-1})^{-1}), \quad (15)$$

where $IW(\cdot, \cdot)$ denotes the inverse-Wishart distribution, and $\mathbf{V} = \sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)$.

For $i = 1, 2, \cdots, n$, let $\mathbf{y}_i = (\mathbf{y}_{i,mis}, \mathbf{y}_{i,obs})$, where $\mathbf{y}_{i,mis}$ represents the missing data while $\mathbf{y}_{i,obs}$ represents the observed data. For a fully observed $\mathbf{y}_i$ data point, $\mathbf{y}_{i,mis}$ is empty. Here, we require the additional notation due to the presence of missing data. Let p_i be the dimension of $\mathbf{y}_{i,mis}$, as $\boldsymbol{\Psi}$ is not diagonal, the conditional distribution $p(\mathbf{y}_{i,mis}|\mathbf{y}_{i,obs}, \boldsymbol{\omega}_i, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$ can be expressed as follows:

$$\begin{aligned}\mathbf{y}_{i,mis}|\mathbf{y}_{i,obs}, \boldsymbol{\omega}_i, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi} \sim\ & N[(\boldsymbol{\mu}_{i,mis} + \boldsymbol{\Lambda}_{i,mis}\boldsymbol{\omega}_i + \boldsymbol{\Psi}_{mis,obs}\boldsymbol{\Psi}_{obs}^{-1}(\mathbf{y}_{i,obs} - \boldsymbol{\mu}_{i,obs} - \boldsymbol{\Lambda}_{i,obs}\boldsymbol{\omega}_i), \\ & \boldsymbol{\Psi}_{mis} - \boldsymbol{\Psi}_{mis,obs}\boldsymbol{\Psi}_{obs}^{-1}\boldsymbol{\Psi}_{obs,mis}], \quad (16)\end{aligned}$$

where $\boldsymbol{\mu}_{i,mis}$ is a $p_i \times 1$ subvector of $\boldsymbol{\mu}$ with elements corresponding to missing components, $\boldsymbol{\Lambda}_{i,mis}$ is a $p_i \times q$ submatrix of $\boldsymbol{\Lambda}$ with rows corresponding to missing components, and $\boldsymbol{\Psi}_{mis}$ is a $p_i \times p_i$ submatrix of $\boldsymbol{\Psi}$ corresponding to the residual covariance matrix of missing components. Similarly, $\boldsymbol{\mu}_{i,obs}$ is a $(p - p_i) \times 1$ subvector of $\boldsymbol{\mu}$ with elements corresponding to observed components, $\boldsymbol{\Lambda}_{i,obs}$ is a $(p - p_i) \times q$ submatrix of $\boldsymbol{\Lambda}$ with rows corresponding to observed components, and $\boldsymbol{\Psi}_{obs}$ is a $(p - p_i) \times (p - p_i)$ submatrix of $\boldsymbol{\Psi}$ corresponding to the residual covariance matrix of observed components. Moreover, $\boldsymbol{\Psi}_{mis,obs}$ is a $p_i \times (p - p_i)$ submatrix of $\boldsymbol{\Psi}$ corresponding to the residual covariance matrix of missing and observed components, and $\boldsymbol{\Psi}_{obs,mis} = \boldsymbol{\Psi}_{mis,obs}^T$.

The following table summarizes the related conditional distributions to simulate observations of the unknown parameters of interest, the missing responses and the corresponding R files:

Parameters	Distribution	R File
$\boldsymbol{\Psi}$	Normal, Gamma	Gibbs_PSX.R
$\boldsymbol{\Lambda}$	Normal	Gibbs_LY.R
$\boldsymbol{\Omega}$	Normal	Gibbs_Omega.R
$\boldsymbol{\mu}$	Normal	Gibbs_MU.R
$\boldsymbol{\Phi}$	Inverse-Wishart	Gibbs_PHI.R
$\mathbf{y}_{i,mis}$	Normal	Gibbs_MISY.R