A Appendix

A.1 Posterior distributions in the MDA algorithm

A.1.1 Posterior distribution of $(d_i, \mathcal{W}_i, \ \boldsymbol{y}_{im})$

Let $\widetilde{\boldsymbol{y}}_{im} = (\mathcal{W}_i, \boldsymbol{y}_{im})$ and $X_i = (x_{i1}, \ldots, x_{iq}, \boldsymbol{y}_{io}{}')'$. Let U_{ij_o} and U_{ij_m} be a partition of the $(Q+s_i)\times 1$ vector $(\boldsymbol{\theta}_j, -1, \boldsymbol{0}_{s_i-j}{}')'$ according to the elements in X_i and $\widetilde{\boldsymbol{y}}_{im}$. Let $y_{ij}^* = -U_{ij_o}{}' X_i$ for model (10). For model (15), $y_{ij}^* = -U_{ij_o}{}' X_i - \sum_{k=1}^{R} \eta_k \underline{z}_{ikj}$. The posterior distribution of $(d_i, \mathcal{W}_i, \boldsymbol{y}_{im})$ is given by

$$\pi(d_i, \mathcal{W}_i, \boldsymbol{y}_{im} \mid \boldsymbol{y}_{io}, \phi, \nu) \propto f(d_i, \mathcal{W}_i)\left[\prod_{j=1}^{s_i}\sqrt{d_i\gamma_j}\right]\exp\left[-\sum_{j=1}^{s_i}\frac{d_i\gamma_j(y_{ij}^* - U_{ij_m}' \widetilde{\boldsymbol{y}}_{im})^2}{2}\right]$$

$$\propto d_i^{\frac{\nu+s_i+1}{2}-1}\exp\left[-d_i\,\frac{b_d + (\widetilde{\boldsymbol{y}}_{im} - \mu_{w_i})' A_{w_i}(\widetilde{\boldsymbol{y}}_{im} - \mu_{w_i})}{2}\right]$$

$$\propto \left[1 + \frac{\dfrac{(\mathcal{W}_i - \mu_{wi1})^2}{U_{w11}^2 b_d / b_a}}{b_a}\right]^{-\frac{b_a+1}{2}} \mathcal{G}\left(d_i \mid \frac{b_a+1}{2}, \frac{b_d + (\mathcal{W}_i - \mu_{wi1})^2 L_{w11}^2}{2}\right)$$

$$N\left(\boldsymbol{y}_{im} \mid \mu_{im}, \frac{U_{w22}U_{w22}{}'}{d_i}\right)$$

(19)

subject to $\mathcal{W}_i > 0$, where $\tilde{m} = m_i + 1$, A_0 is a $\tilde{m}_i \times \tilde{m}_i$ matrix with 1 at its (1, 1) entry and 0 elsehwere, $A_{w_i} = A_0 + \sum_{j=1}^{s_i}\gamma_j U_{ij_m}{}' U_{ij_m}$, the lower triangle matrix $L_{w_i} = \begin{bmatrix} L_{w11} & \boldsymbol{0} \\ L_{w21} & L_{w22}\end{bmatrix}$ satisfies $A_{w_i} = L_{w_i}{}' L_{w_i}$, L_{w11} is a scalar, $U_{w_i} = \begin{bmatrix} U_{w11} & \boldsymbol{0} \\ U_{w21} & U_{w22}\end{bmatrix} = L_{w_i}^{-1}$, $B_{w_i} = \sum_{j=1}^{s_i}\gamma_j U_{ij_m}{}' y_{ij}^*$, $C_{w_i} = U_{w_i}{}' B_{w_i}$, and $\mu_{w_i} = A_{w_i}^{-1}B_{w_i} = U_{w_i}C_{w_i} = \begin{bmatrix}\mu_{wi1} \\ \mu_{wi2}\end{bmatrix}$, $\mu_{im} = \mu_{wi2} + U_{w21}L_{w11}(\mathcal{W}_i - \mu_{wi1})$, $b_a = \nu + o_i$, $B_{w_i}{}' A_{w_i}^{-1} B_{w_i} = C_{w_i}{}' C_{w_i}$ and

$b_d = \nu + \sum_{j=1}^{s_i} \gamma_j y_{ij}^{*2} - C_{w_i}{}' C_{w_i}$. In SAS IML, L_{w_i} can be computed can use the following syntax

```
index=mtilde:1;
Lwi = (root(Awi[index,index]))[index,index];
```

Equation (19) implies that we can draw $(\mathcal{W}_i, d_i, \boldsymbol{y}_{im})$ sequentially from

$$d_i^* \sim \mathcal{G}\left(\frac{b_a}{2}, \frac{b_d}{2}\right), \quad \mathcal{W}_i \mid d_i^* \sim N^+\left(\mu_{wi1}, \frac{U_{w11}^2}{d_i}\right),$$

$$d_i \mid \mathcal{W}_i, d_i^* \sim \mathcal{G}\left(\frac{b_a+1}{2}, \frac{b_d + (\mathcal{W}_i - \mu_{wi1})^2 L_{w11}^2}{2}\right), \tag{20}$$

$$\boldsymbol{y}_{im} \mid d_i, \mathcal{W}_i, d_i^* \sim N\left(\mu_{im}, \frac{U_{w22} U_{w22}{}'}{d_i}\right).$$

The marginal distribution of $\mathcal{W}_i$ is $t^+\left(\mu_{wi1}, U_{w11}^2 \frac{b_d}{b_a}, b_a\right)$, but the marginal distribution of d_i is not gamma due to the restriction $\mathcal{W}_i > 0$.

The posterior distribution of $\boldsymbol{y}_{im}$ can be equivalently written as

$$\boldsymbol{y}_{im} \mid d_i, \mathcal{W}_i, \boldsymbol{y}_{io}, \phi, \nu \sim N\left(\mu_{im}, \frac{V_{y_{\mathrm{m}}}}{d_i}\right), \tag{21}$$

where *h* is the index of the first missing observation for subject *i* in pattern *s*, $\tilde{U}_{ij_m}$ is a sub-vector of $(\boldsymbol{\theta}_j, -1, \boldsymbol{\theta}_{s_i - j}{}')'$ corresponding to the elements in $\boldsymbol{y}_{im}$, $y_{ij}^{**} = y_{ij}^* - \psi_j \mathcal{W}_i$, $V_{y_{\mathrm{m}}} = (\sum_{j=h}^{s} \gamma_j \tilde{U}_{ij_m} \tilde{U}_{ij_m}{}')^{-1}$ and $\mu_{im} = \hat{V}_{y_{\mathrm{m}}} \sum_{j=h}^{s} \gamma_j \tilde{U}_{ij_m} y_{ij}^{**}$.

A.1.2 Posterior distribution of *g* in step PX1

The posterior distribution of *g* with a Harr prior g^{-1} and Jacobian g^{n-p} is

$$
\begin{aligned}
\mathrm{pos}(g) &\propto g^{n-p} g^{-1} \pi\left(Y_m, gd_1, \ldots, gd_n, W, \boldsymbol{\theta}_1, \frac{\gamma_1}{g}, d_{1_\psi}, \ldots, \boldsymbol{\theta}_p, \frac{\gamma_p}{g}, d_{p_\psi}, \nu \mid Y_o \right) \\
&\propto g^{n-p} g^{-1} \prod_{i=1}^{n} \left\{ (gd_i)^{\frac{\nu+1}{2}-1} \exp\left[-g \frac{d_i(\nu + \mathcal{W}_i^2)}{2} \right] \right\} \\
&\quad \prod_{j=1}^{p} \left\{ \left(\frac{\gamma_j}{g} \right)^{\frac{n_w+2j+r-p-3}{2}} \exp\left[-\frac{\gamma_j \tilde{\boldsymbol{\theta}}_j' E_j \tilde{\boldsymbol{\theta}}_j}{2g} \right] \right\} \\
&\propto g^{\frac{n(1+\nu)-p(n_w+r)}{2}-1} \exp\left[-g \frac{\sum_{i=1}^{n} d_i(\nu + \mathcal{W}_i^2)}{2} \right] \exp\left[-\frac{\sum_{j=1}^{p} \gamma_j \tilde{\boldsymbol{\theta}}_j' E_j \tilde{\boldsymbol{\theta}}_j}{2g} \right].
\end{aligned}
$$

A.1.3 Posterior distribution of *h* in step PX2

The posterior distribution of *h* with a Harr prior h^{-1} and Jacobian h^{n-p} is

$$
\mathrm{pos}(h) \propto h^{n-p} h^{-1} \exp\left(-h^2 \frac{\sum_{i=1}^{n} d_i \mathcal{W}_i^2}{2} \right) \exp\left(-\frac{\sum_{j=1}^{p} \gamma_j \underline{\psi}_j^2 \frac{4d_{j_\psi}}{\pi^2}}{2h^2} \right).
$$

The posterior distribution of $H = h^2$ is

$$
\mathrm{pos}(H) \propto \mathrm{pos}(h) \left| \frac{\partial h}{\partial H} \right| \propto H^{\frac{n-p}{2}-1} \exp\left(-H \frac{\sum_{i=1}^{n} d_i \mathcal{W}_i^2}{2} \right) \exp\left(-\frac{\sum_{j=1}^{p} \gamma_j \underline{\psi}_j^2 \frac{4d_{j_\psi}}{\pi^2}}{2H} \right).
$$

A.1.4 Prior and posterior distributions of *v*

We firstly derive the PC prior for *v*. The Kullback-Leibler (KL) distance between the multivariate t distribution $t(\boldsymbol{\mu}, \frac{\nu-2}{\nu}\Sigma, \nu)$ and the normal distribution $N(\boldsymbol{\mu}, \Sigma)$ is

$$
\begin{aligned}
\mathrm{KL}(\nu) &= \int t\left(\boldsymbol{x} \mid \boldsymbol{\mu}, \frac{\nu-2}{\nu}\Sigma, \nu\right) \log t\left(\boldsymbol{x} \mid \boldsymbol{\mu}, \frac{\nu-2}{\nu}\Sigma, \nu\right) d\boldsymbol{x} \\
&\quad - \int t\left(\boldsymbol{x} \mid \boldsymbol{\mu}, \frac{\nu-2}{\nu}\Sigma, \nu\right) \log \phi(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma) d\boldsymbol{x} \\
&= \frac{p}{2}\left[1 + \log\left(\frac{2}{\nu-2}\right)\right] + \log\Gamma\left(\frac{\nu+p}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) \\
&\quad - \frac{\nu+p}{2}\left[\Psi\left(\frac{\nu+p}{2}\right) - \Psi\left(\frac{\nu}{2}\right)\right].
\end{aligned}
$$

since the first integration equals

$$
\log\Gamma\left(\frac{\nu+p}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{\nu+p}{2}\left[\Psi\left(\frac{\nu+p}{2}\right) - \Psi\left(\frac{\nu}{2}\right)\right] - \frac{1}{2}\log|\Sigma| - \frac{p}{2}\log(\nu-2) - \frac{p}{2}\log(\pi)
$$

by Kotz and Nadarajah [76], and the second integration equals

$$
-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{p}{2}.
$$

By the definition of the PC prior [52], the density is $\lambda \exp(-\lambda d(\nu))\left|\frac{\partial d(\nu)}{\partial \nu}\right|$, where $d(\nu) = \sqrt{2\mathrm{KL}(\nu)}$.

The posterior distribution of ν is given by

$$
\begin{aligned}
&\pi(\nu \mid \Sigma, \boldsymbol{\psi}, \theta_j, \gamma_j) \propto \pi(\nu) \prod_{i=1}^{n} t_p(\boldsymbol{y}_{io} \mid \boldsymbol{\mu}_{io}, \Omega_{io}, \nu) \\
&\quad T_{\nu+o_i}\left[\boldsymbol{\lambda}_{io}^{*\prime}(\boldsymbol{y}_{io} - \boldsymbol{\mu}_{io})\sqrt{\frac{\nu+o_i}{\nu + (\boldsymbol{y}_{io} - \boldsymbol{\mu}_{io})'\Omega_{io}^{-1}(\boldsymbol{y}_{io} - \boldsymbol{\mu}_{io})}}\right] I(\nu > \nu_l),
\end{aligned}
$$

where $t_p(\boldsymbol{y}_{io} \mid \boldsymbol{\mu}_{io}, \Omega_{io}, \nu) \propto \dfrac{\Gamma(\frac{\nu+o_i}{2})}{\Gamma(\frac{\nu}{2})\nu^{o_i/2}}[1 + (\boldsymbol{y}_{io} - \boldsymbol{\mu}_{io})'\Omega_{io}^{-1}(\boldsymbol{y}_{io} - \boldsymbol{\mu}_{io})/\nu]^{-\frac{\nu+o_i}{2}}$, and $\boldsymbol{\lambda}_{io}^{*} = \Sigma_{io}^{-1}\boldsymbol{\psi}_{io} / \sqrt{1 + \boldsymbol{\psi}_{io}'\Sigma_{io}^{-1}\boldsymbol{\psi}_{io}}$.

For subjects with no intermittent missing data, the skew-t density function can be computed without matrix inversion using the following relationship

$\boldsymbol{\lambda}_{io}^{*\prime}(\boldsymbol{y}_{io} - \boldsymbol{\mu}_{io}) = \sum_{t=1}^{s} \lambda_t R_t \underline{\psi}_t$ and

$(\boldsymbol{y}_{io}-\boldsymbol{\mu}_{io})'\Omega_{io}^{-1}(\boldsymbol{y}_{io}-\mu_{io})=\sum_{t=1}^{s}\lambda_t r_t^2-[\sum_{t=1}^{s}\lambda_t r_t \underline{\psi}_t]^2/[1+\sum_{t=1}^{s}\lambda_t \underline{\psi}_t^2]$, where

$r_j = y_{ij}-\sum_{t=1}^{j-1}\beta_{jt}y_{it}-\sum_{k=1}^{Q}\underline{\alpha}_{kj}x_{ik}$ for model (10) and $r_j = y_{ij}-\sum_{t=1}^{j-1}\beta_{jt}y_{it}-\sum_{k=1}^{R}\eta_k \underline{z}_{ikj}-\sum_{k=1}^{Q}\underline{\alpha}_{kj}x_{ik}$ for model (15).

The candidate ν^* is generated from $\log(\nu^*-\nu_l)\sim N[\log(\nu-\nu_l),c^2]$. It will be accepted with probability $\alpha_\nu=\min\left\{1,\frac{(\nu^*-\nu_l)\pi(\nu^*)\prod_{i=1}^{n} f(\boldsymbol{y}_{io}\,|\,\boldsymbol{\theta}_i,\gamma_i,\nu^*)}{(\nu-\nu_l)\pi(\nu)\prod_{i=1}^{n} f(\boldsymbol{y}_{io}\,|\,\boldsymbol{\theta}_i,\gamma_i,\nu)}\right\}$.

A.2 Adaption of the MDA algorithm for MMRM-sn

For MMRM-sn (i.e. $\nu\equiv\infty$), the MDA algorithm *A* can easily adapted by ignoring steps P2 and PX1, and setting $d_i\equiv 1$ and $d_i^*\equiv 1$ in drawing $(\boldsymbol{\theta}_j,\gamma_j,d_{\psi_j},\rho_j)$'s and $(\mathcal{W}_i,\boldsymbol{y}_{im})$'s. For example, $(\mathcal{W}_i,\boldsymbol{y}_{im})$ in step I can be imputed by modifying the posterior distribution (20) as

$$\mathcal{W}_i\sim N^+(\mu_{wi1},U_{w11}^2)\text{ and }\boldsymbol{y}_{im}\,|\,\mathcal{W}_i\sim N(\mu_{im},U_{w22}U_{w22}').$$

A.3 Adaption of the MDA algorithm for MMRM-t

For MMRM-t (i.e. $\underline{\psi}_j\equiv 0$ for $j=1,\dots,p,\mathcal{W}_i\equiv 0$), the MDA algorithm *A* needs the following modifications:

1. Step PX2 is no longer needed.
2. Step P1: Remove $\mathcal{W}_i$ from the model. Set $d_{\psi_j}\equiv 0$ and $\boldsymbol{\theta}_j=(\underline{\alpha}_{1j},\dots,\underline{\alpha}_{qj},\beta_{j1},\dots,\beta_{jj-1})'$. Sample $(\boldsymbol{\theta}_j,\gamma_j)$'s from

$$\pi(\boldsymbol{\theta}_j,\gamma_j\,|\,d_i'\text{s},\nu,\mathrm{Y_o},\mathrm{Y_m})\propto\gamma_j^{\frac{n_w+2j+r^*-p-1}{2}-1}\exp\left[-\frac{\gamma_j}{2}\tilde{\boldsymbol{\theta}}_j'(E_j+\sum_{i\le n_j}d_i\tilde{\boldsymbol{x}}_{ij}\tilde{\boldsymbol{x}}_{ij}')\tilde{\boldsymbol{\theta}}_j\right]$$

for $j=1,\dots,p$, where E_j is the $(q+j)\times(q+j)$ leading principle submatrix of the $(q+p)\times(q+p)$ matrix $E=\begin{bmatrix} M & M\boldsymbol{\alpha}_0' \\ \boldsymbol{\alpha}_0 M & \boldsymbol{\alpha}_0 M\boldsymbol{\alpha}_0'+A_w\end{bmatrix}$ and r^* is the rank of *M*. If the inverse Wishart or Jeffrey's prior (with fixed A_w and n_w) instead of the hierarchical prior of Huang and Wand [49] is used, step P0 shall be ignored.

3. Step I: draw $d_i \sim \mathcal{G}\left(\frac{b_a}{2}, \frac{b_d}{2}\right)$ and $\boldsymbol{y}_{im} \sim N(\mu_{w_i}, (d_i A_{w_i})^{-1})$ since

$$\pi(d_i, \quad \boldsymbol{y}_{im} \mid \boldsymbol{y}_{io}, \nu, \phi) \propto f(d_i) d_i^{\frac{s_i}{2}} \exp\left[-\frac{\sum_{j=1}^{s_i} d_i \gamma_j (y_{ij}^* - \tilde{U}_{ij_m}{}' \boldsymbol{y}_{im})^2}{2}\right]$$

$$\propto \left\{ d_i^{\frac{\nu+o_i}{2}-1} \exp\left[-\frac{d_i b_d}{2}\right]\right\}\left\{ d_i^{\frac{m_i}{2}} \exp[-\frac{d_i (\boldsymbol{y}_{im} - \mu_{w_i})' A_{w_i} (\boldsymbol{y}_{im} - \mu_{w_i})}{2}]\right\},$$

where $f(d_i) \propto d_i^{\nu/2-1} \exp(-d_i \nu)$, $A_{w_i} = \sum_{j=1}^{s_i} \gamma_j \tilde{U}_{ij_m}{}' \tilde{U}_{ij_m}$, $B_{w_i} = \sum_{j=1}^{s_i} \gamma_j \tilde{U}_{ij_m}{}' y_{ij}^*$, $\mu_{w_i} = A_{w_i}^{-1} B_{w_i}$,

$b_a = \nu + o_i$ and $b_d = \nu + \sum_{j=1}^{s_i} \gamma_j y_{ij}^{*2} - B_{w_i}{}' A_{w_i}^{-1} B_{w_i}$.

4. Step PX1: g is randomly drawn from

$$g^{\frac{n\nu - p(n_w + r^*)}{2}-1} \exp\left[-g\frac{\nu \sum_{i=1}^{n} d_i}{2}\right] \exp\left[-\frac{\sum_{j=1}^{p} \gamma_j \tilde{\boldsymbol{\theta}}_j{}' E_j \tilde{\boldsymbol{\theta}}_j}{2g}\right].$$