Web-based Supplementary Materials for "A Note on MAR, Identifying Restrictions, Model Comparison, and Sensitivity Analysis in Pattern Mixture Models With and Without Covariates for Incomplete Data" by Chenguang Wang and Michael Daniels

1 Web Appendix A

Proofs of Theorems/Corollaries from Sections 2 and 3

Theorem 1

Proof: By Lemma 1, we only need to show that MAR constraints exist if and only if for all $1 < j < J$, the conditional distributions $p_s(y_j|\overline{y}_{j-1})$ are identical for $s \geq j$.

Molenberghs et al. (1998) proved that MAR holds if and only if

$$p_k(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1}) = \sum_{s=j}^{J} \frac{P(S=s)}{\sum_{s=j}^{J} P(S=s)} p_s(y_j|\overline{y}_{j-1}) \tag{1}$$

for all $j \geq 2$ and $k < j$. These conditionals are normal distributions since we assume $\boldsymbol{Y}|S$ is multivariate normal.

Suppose that there exists j such that $p_s(y_j|\overline{y}_{j-1})$ is not the same for all $s \geq j$. Then from (1), $p_{\geq j}(y_j|\overline{y}_{j-1})$ will be a mixture of normals whereas $p_k(y_j|\overline{y}_{j-1})$ will be a normal distribution. Thus, Molenbergh's condition will not be satisfied, i.e. the MAR constraints do not exist.

On the other hand, if for all $1 < j < J$, the conditional distributions $p_s(y_j|\overline{y}_{j-1})$ are identical for $s \geq j$, then $p_k(y_j|\overline{y}_{j-1})$ and $p_{\geq j}(y_j|\overline{y}_{j-1})$ are both normally distributed and the identification restrictions $p_k(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1})$ will result in MAR.

Corollary 1

Proof: Since Y_1 is always observed (by assumption), $S|\boldsymbol{Y} \simeq S|Y_1$ implies that $S|\boldsymbol{Y}_{\text{mis}}, \boldsymbol{Y}_{\text{obs}} \simeq S|\boldsymbol{Y}_{\text{obs}}$, where $\boldsymbol{Y}_{\text{mis}}$ and $\boldsymbol{Y}_{\text{obs}}$ denote the missing and observed data respectively. This shows that MAR holds.

On the other hand, MAR implies that

$$p(S=s|\boldsymbol{Y}) = p(S=s|\boldsymbol{Y}_{\text{obs}}) = p(S=s|\overline{Y}_s).$$

By Theorem 1, we have that MAR holds only if for all $1 < j < J$, the conditional distributions $p_s(y_j|\overline{y}_{j-1})$ are identical for $s \geq j$. Thus, under MAR

$$p_k(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1}) = p_s(y_j|\overline{y}_{j-1})$$

for all $j \geq 2$, $k < j$ and $s \geq j$. This implies that for all $j \geq 2$

$$p(y_j|\overline{y}_{j-1}) = \sum_{s=1}^{J} p_s(y_j|\overline{y}_{j-1})p(S = s) = p_s(y_j|\overline{y}_{j-1})$$

for all s.

Therefore,

$$\begin{aligned}
p(S = s|\boldsymbol{Y}) = p(S = s|\overline{y}_s) &= \frac{p_s(\overline{y}_s)}{p(\overline{y}_s)}p(S = s) \\
&= \frac{p_s(y_s|\overline{y}_{s-1})\dots p_s(y_2|y_1)p_s(y_1)}{p(y_s|\overline{y}_{s-1})\dots p(y_2|y_1)p(y_1)}p(S = s) \\
&= \frac{p_s(y_1)}{p(y_1)}p(S = s) = p(S = s|y_1).
\end{aligned}$$

Corollary 2

Proof: First, MCAR implies MAR. Second, in the proof of Corollary 1, we showed that MAR holds if

$$p(S = s|\boldsymbol{Y}) = \frac{p_s(y_1)}{p(y_1)}p(S = s).$$

Thus under the assumption that $p_s(y_1) = p(y_1)$, MAR implies that $p(S = s|\boldsymbol{Y}) = p(S = s)$, i.e. MCAR.

Corollary 3

Proof By Theorem 1, the MAR constraints imply

$$p_j(y_j|\overline{y}_{j-1}) = p_J(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1}).$$

Therefore for all $k < j$, the MAR constraints

$$p_k(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1})$$

are identical to CCMV restrictions

$$p_k(y_j|\overline{y}_{j-1}) = p_J(y_j|\overline{y}_{j-1})$$

and to NCMV restrictions

$$p_k(y_j|\overline{y}_{j-1}) = p_j(y_j|\overline{y}_{j-1}).$$

Corollary 4

Proof: Theorem 1 shows that identification via MAR constraints exists if and only if conditional distributions $p_s(y_j|\overline{y}_{j-1})$ are identical for $s \geq j$ and $j \geq 2$. That is, for observed data, we have

$$p_s(y_j|\overline{y}_{j-1}) \sim \mathrm{N}(\mu^{(\geq j)}_{j|j^-}, \sigma^{(\geq j)}_{j|j^-}).$$

2 Web Appendix B

Missing Data Mechanism under MNAR and Multivariate Normality (Section 3)

To see the impact of the Δ parameters on the missing data mechanism (MDM), we introduce notation $\Delta^{(j)}_{j|j^-} = \Delta^{(j)}_0 + \sum_{l=1}^{j-1} \Delta^{(j)}_l Y_l$ and then for $k < j$ we have

$$Y_j|\overline{Y}_{j-1}, S = k \sim \mathrm{N}\left(\mu^{(\geq j)}_{j|j^-} + \Delta^{(j)}_{j|j^-}, e^{\Delta^{(j)}_\sigma}\sigma^{(\geq j)}_{j|j^-}\right).$$

The conditional probability (hazard) of observing the first s observations given at least s observations is derived as follows:

$$\begin{aligned}
&\log \frac{P(S=s|\boldsymbol{Y})}{P(S\geq s|\boldsymbol{Y})} = \log \frac{P(S=s)p_s(\boldsymbol{Y})}{P(\boldsymbol{Y}, S\geq s)} \\
&= \log \frac{P(S=s)p_s(Y_1)\prod_{l=2}^J p_s(Y_l|\overline{Y}_{l-1})}{\sum_{k=s}^J \left\{P(S=k)p_k(Y_1)\prod_{l=2}^J p_k(Y_l|\overline{Y}_{l-1})\right\}} \\
&= \log \frac{\prod_{l=2}^s p_{\geq l}(Y_l|\overline{Y}_{l-1})P(S=s)p_s(Y_1)\prod_{l=s+1}^J p_s(Y_l|\overline{Y}_{l-1})}{\prod_{l=2}^s p_{\geq l}(Y_l|\overline{Y}_{l-1})\sum_{k=s}^J \left\{P(S=k)p_k(Y_1)\prod_{l=s+1}^J p_k(Y_l|\overline{Y}_{l-1})\right\}} \\
&= \log \frac{P(S=s)p_s(Y_1)\prod_{l=s+1}^J p_s(Y_l|\overline{Y}_{l-1})}{\sum_{k=s}^J \left\{P(S=k)p_k(Y_1)\prod_{l=s+1}^k p_k(Y_l|\overline{Y}_{l-1})\prod_{l=k+1}^J p_k(Y_l|\overline{Y}_{l-1})\right\}} \\
&= \log P(S=s) - \frac{\sigma^{(s)}_1}{2} + \frac{(Y_1-\mu^{(s)}_1)^2}{2\sigma^{(s)}_1} + \sum_{l=s+1}^J \left\{-\frac{e^{\Delta^{(l)}_\sigma}}{2} + \frac{\left(Y_l - \mu^{(\geq l)}_{l|l^-} - \Delta^{(l)}_{l|l^-}\right)^2}{2e^{\Delta^{(l)}_\sigma}\sigma^{(\geq l)}_{l|l-1}}\right\} \\
&\quad - \log\sum_{k=s}^J \left\{P(S=k)(\sigma^k_1)^{-\frac{1}{2}}\exp\left\{\frac{(Y_1-\mu^{(k)}_1)^2}{2\sigma^{(k)}_1}\right\}\prod_{l=s+1}^k \exp\left\{\frac{(Y_l-\mu^{(\geq l)}_{l|l^-})^2}{2\sigma^{(\geq l)}_{l|l-1}}\right\}\right. \\
&\qquad\qquad \left.\times\prod_{l=k+1}^J (e^{\Delta^{(l)}_\sigma})^{-\frac{1}{2}}\exp\left\{\frac{(Y_l-\mu^{(\geq l)}_{l|l^-}-\Delta^{(l)}_{l|l^-})^2}{2e^{\Delta^{(l)}_\sigma}\sigma^{(\geq l)}_{l|l-1}}\right\}\right\}.
\end{aligned}$$

In general the MDM depends on $\overline{Y}_J$, i.e. MNAR.

3 Web Appendix C

Mean and Variance of $\left[Y_j|\overline{Y}_{j-1}, S=s\right]$ *(Section 5)*

The mean and variance of $[Y_j|\overline{Y}_{j-1}, S=s]$ under MNAR assumption are derived as follows:

$$\begin{aligned}
\mu_{j|j-}^{(s),\mathrm{MNAR}} &= \mathrm{E}(Y_j|\overline{Y}_{j-1}, S=s) = e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} \int y_j p_k\left(\frac{y_j - \Delta_\mu^{(j)} - (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}}|\overline{y}_{j-1}\right) dy_j \\
&= e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} \int \left(e^{\Delta_\sigma^{(j)}} y_j^* + \Delta_\mu^{(j)} + (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}\right) p_k(y_j^*|\overline{y}_{j-1}) e^{\Delta_\sigma^{(j)}} dy_j^* \\
&= \Delta_\mu^{(j)} + \sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(s)}
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{j|j-}^{(s),\mathrm{MNAR}} &= \mathrm{Var}(Y_j|\overline{Y}_{j-1}, S=s) \\
&= e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} \int y_j^2 p_k\left(\frac{y_j - \Delta_\mu^{(j)} - (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}}|\overline{y}_{j-1}\right) dy_j - \mathrm{E}^2(Y_j|\overline{Y}_{j-1}, S=k) \\
&= e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} \int \left(e^{\Delta_\sigma^{(j)}} y_j^* + \Delta_\mu^{(j)} + (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}\right)^2 p_k(y_j^*|\overline{y}_{j-1}) e^{\Delta_\sigma^{(j)}} dy_j^* \\
&\quad - \left(\Delta_\mu^{(j)} + \sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)}\right)^2 \\
&= e^{2\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} E((y_j^*)^2|\overline{y}_{j-1}, S=k) + \sum_{k=j}^{J} \varpi_{j,k} (\Delta_\mu^{(j)} + (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)})^2 \\
&\quad + 2e^{\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} (\Delta_\mu^{(j)} + (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}) E(y_j^*|\overline{y}_{j-1}, S=k) - \left(\Delta_\mu^{(j)} + \sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)}\right)^2 \\
&= e^{2\Delta_\sigma^{(j)}} \left\{ \sum_{k=j}^{J} \varpi_{j,k} \left(\sigma_{j|j-1}^{(k)} + (\mu_{j|j-1}^{(k)})^2\right) - \left(\sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)}\right)^2 \right\} \\
&\quad + (1-e^{2\Delta_\sigma^{(j)}}) \left\{ \sum_{k=j}^{J} \varpi_{j,k} (\mu_{j|j-}^{(k)})^2 - \left(\sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)}\right)^2 \right\}
\end{aligned}$$

where $y_j^* = \frac{y_j - \Delta_\mu^{(j)} - (1-e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}}$.

4 Web Appendix D

Full-data Model for the Growth Hormone Example (Section 6)

We specify a pattern mixture model with sensitivity parameters for the two treatment arms. For compactness, we suppress subscript treatment indicator z from all the parameters in the following models.

Missing Pattern S

$$S \sim \text{Mult}(\boldsymbol{\phi})$$

with the multinomial parameter $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$, $\phi_s = P(S = s)$ for $s \in \{1, 2, 3\}$, and $\sum_{s=1}^{3} \phi_s = 1$.

Observed Response Data $\boldsymbol{Y}_{\textbf{obs}}$ given S

We specify the same MVN and OMVN model for $[Y_1|S]$ as follows:

$$\begin{aligned} Y_1|S = 1 &\sim \text{N}(\mu_1^{(1)}, \sigma_1^{(1)}) \\ Y_1|S = 2 &\sim \text{N}(\mu_1^{(2)}, \sigma_1^{(2)}) \\ Y_1|S = 3 &\sim \text{N}(\mu_1^{(3)}, \sigma_1^{(3)}). \end{aligned}$$

For MVN model, we specify

$$\begin{aligned} \left.\begin{matrix} Y_2|Y_1, S = 2 \\ Y_2|Y_1, S = 3 \end{matrix}\right\} &\sim \text{N}\left(\beta_0^{(\geq 2)} + \beta_1^{(\geq 2)} Y_1,\ \sigma_{2|2^-}^{(\geq 2)}\right) \\ Y_3|Y_2, Y_1, S = 3 &\sim \text{N}\left(\beta_0^{(\geq 3)} + \beta_1^{(\geq 3)} Y_1 + \beta_2^{(\geq 3)} Y_2,\ \sigma_{3|3^-}^{(\geq 3)}\right). \end{aligned}$$

For OMVN model, we specify

$$\begin{aligned} Y_2|Y_1, S = 2 &\sim \text{N}\left(\beta_{2,0}^{(2)} + \beta_{2,1}^{(2)} Y_1,\ \sigma_{2|2^-}^{(2)}\right) \\ Y_2|Y_1, S = 3 &\sim \text{N}\left(\beta_{2,0}^{(3)} + \beta_{2,1}^{(3)} Y_1,\ \sigma_{2|2^-}^{(3)}\right) \\ Y_3|Y_2, Y_1, S = 3 &\sim \text{N}\left(\beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2,\ \sigma_{3|3^-}^{(3)}\right). \end{aligned}$$

Missing Response Data $\boldsymbol{Y}_{\textbf{mis}}$ given $\boldsymbol{Y}_{\textbf{obs}}$, S

For MVN model, we specify

$$\begin{aligned} Y_2|Y_1, S = 1 &\sim \text{N}\left(\beta_0^{(\geq 2)} + \Delta_0^{(2)} + (\beta_1^{(\geq 2)} + \Delta_1^{(2)}) Y_1,\ e^{\Delta_\sigma^{(2)}} \sigma_{2|2^-}^{(\geq 2)}\right) \\ Y_3|Y_2, Y_1, S = 2 &\sim \text{N}\left(\beta_0^{(\geq 3)} + \Delta_0^{(3)} + (\beta_1^{(\geq 3)} + \Delta_1^{(3)}) Y_1 + (\beta_2^{(\geq 3)} + \Delta_2^{(3)}) Y_2,\ e^{\Delta_\sigma^{(3)}} \sigma_{3|3^-}^{(\geq 3)}\right) \\ Y_3|Y_2, Y_1, S = 1 &\sim \frac{\phi_3}{\phi_2 + \phi_3} \text{N}\left(\beta_0^{(\geq 3)} + \beta_1^{(\geq 3)} Y_1 + \beta_2^{(\geq 3)} Y_2,\ \sigma_{3|3^-}^{(\geq 3)}\right) \\ &\quad + \frac{\phi_2}{\phi_2 + \phi_3} \text{N}\left(\beta_0^{(\geq 3)} + \Delta_0^{(3)} + (\beta_1^{(\geq 3)} + \Delta_1^{(3)}) Y_1 + (\beta_2^{(\geq 3)} + \Delta_2^{(3)}) Y_2,\ e^{\Delta_\sigma^{(3)}} \sigma_{3|3^-}^{(\geq 3)}\right). \end{aligned}$$

For OMVN model, we specify

$$\begin{aligned} Y_2|Y_1, S = 1 &\sim \frac{\phi_3}{\phi_2 + \phi_3} \text{N}\left(\Delta_u^{(2)} + \beta_{2,0}^{(3)} + \beta_{2,1}^{(3)} Y_1,\ e^{\Delta_\sigma^{(2)}} \sigma_{2|2^-}^{(3)}\right) \\ &\quad + \frac{\phi_2}{\phi_2 + \phi_3} \text{N}\left(\Delta_u^{(2)} + \beta_{2,0}^{(2)} + \beta_{2,1}^{(2)} Y_1,\ e^{\Delta_\sigma^{(2)}} \sigma_{2|2^-}^{(2)}\right) \\ Y_3|Y_2, Y_1, S = 2 &\sim \text{N}\left(\Delta_u^{(3)} + \beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2,\ e^{\Delta_\sigma^{(3)}} \sigma_{3|3^-}^{(3)}\right) \\ Y_3|Y_2, Y_1, S = 1 &\sim \frac{\phi_3}{\phi_2 + \phi_3} \text{N}\left(\beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2,\ \sigma_{3|3^-}^{(3)}\right) \\ &\quad + \frac{\phi_2}{\phi_2 + \phi_3} \text{N}\left(\Delta_u^{(3)} + \beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2,\ e^{\Delta_\sigma^{(3)}} \sigma_{3|3^-}^{(3)}\right). \end{aligned}$$

5 Web Appendix E

Simulation for Multivariate t and Multivariate Skewed Normal (Section 7)

For multivariate t (MVT) case, we chose $df = 3$. For multivariate skewed normal (MSKN), we let $\omega = 3$ to make the distribution right-skewed with marginal mean denoted as $\boldsymbol{\mu}' = \{\mu_1', \mu_2', \mu_3'\}$. The parameters $\boldsymbol{\mu}$, $\boldsymbol{\mu}'$ and $\boldsymbol{\Sigma}$, estimated from the observed EG arm data and adjusted to reflect general setting, are reported in Table 1.

The missing data mechanism model for both the MVT and MSKN cases is constructed as follows:

$$\text{logit}\, P(S = 1 | S \geq 1, \boldsymbol{Y}) = \psi_{1,0} + \psi_{1,1} Y_1 + \delta_1 Y_2$$
$$\text{logit}\, P(S = 2 | S \geq 2, \boldsymbol{Y}) = \psi_{2,0} + \psi_{2,1} Y_2 + \delta_2 Y_3.$$

To generate data according to MAR restraint, we let $\psi_{1,1} = \psi_{2,1} = -0.1$ and $\delta_1 = \delta_2 = 0$. To generate data according to MNAR restraint, we let $\psi_{1,1} = \psi_{2,1} = 0$, $\delta_1 = -0.09$, and $\delta_2 = -0.1$. For both MAR and MNAR, we let $\psi_{1,0} = \psi_{2,0} = 5.5$. The MDM parameters are chosen to have $P(S = 1)$ and $P(S = 2)$ to be roughly 0.2 (dropout rate of 40%).

To choose priors for the sensitivity parameters for MNAR analysis with MVN and OMVN model, we use the same approach as in Section 6. Based on the observed variety, we set $\mathscr{D}(\boldsymbol{\tau}) = [-4.7, 0] \times [-2.9, 0]$ for MVT simulation and $\mathscr{D}(\boldsymbol{\tau}) = [-2.7, 0] \times [-1.8, 0]$ for MSKN simulation.

Table 1: Simulation Scenarios

MVT $\boldsymbol{\mu}$		MSKN $\boldsymbol{\mu}'$	
μ_1	69	μ_1'	66.6
μ_2	81	μ_1'	77.8
μ_3	78	μ_1'	74.6
$\boldsymbol{\Sigma}$ for MVT and MSKN			
σ_{11}	11.3	σ_{22}	19
σ_{12}	11.0	σ_{23}	17.4
σ_{13}	12.3	σ_{33}	20.1

6 Web Appendix F

ACMV (MAR) on the residuals for multivariate case (Section 8)

To incorporate baseline covariates in the multivariate case and apply similar MAR restrictions, we specify the model for the observed data as follows:

$$\begin{aligned} p_s(y_1|X) &\sim \text{N}(\mu_1^{(s)} + X\alpha^{(s)}, \sigma_1^{(s)}) && 1 \leq s \leq J \\ p_s(y_j|\overline{y}_{j-1}, X) &\sim \text{N}(\mu_{j|j^-}^{(s)}, \sigma_{j|j^-}^{(\geq j)}) && 2 \leq j \leq s \leq J, \end{aligned}$$

where

$$\mu_{j|j^-}^{(s)} = \mu_j^{(\geq j)} + X\alpha^{(s)} + \sum_{l=1}^{j-1} \beta_l^{(\geq j)} (Y_l - \mu_l^{(\geq j)} - X\alpha^{(s)}). \tag{2}$$

For the missing data, the conditional distributions are specified as

$$p_s(y_j|\overline{y}_{j-1}) \sim \mathrm{N}(\mu^{(s)}_{j|j^-}, \sigma^{(j)}_{j|j^-}) \qquad 1 \le s < j \le J$$

where

$$\mu^{(s)}_{j|j^-} = \mu^{(s)}_j + X\alpha^{(s)} + \sum_{l=1}^{j-1} \beta^{(s)}_{j,l}(Y_l - \mu^{(s)}_l - X\alpha^{(s)}). \tag{3}$$

The conditional mean structures in (2) and (3) induce the following form for the marginal mean response

$$E(Y_j|S=s) = U^{(s)}_j + X\alpha^{(s)},$$

where $U^{(s)}_j$ is a function of intercept (e.g. $\mu^{(\ge j)}_j$) and slope (e.g. $\beta^{(\ge j)}_l$) parameters from $\mu^{(s)}_{j|j^-}$, but not X or α. This marginal mean response reflects the fact that X is the baseline covariates and α is its time-invariant coefficient. This form is also necessary for resolving over-identification of α via the MAR on the residuals restrictions as shown later.

Note that since Y_1 is always observed, $\alpha^{(s)}$ $(1 \le s \le J)$ are identified by the observed data. However, in the model given by (2) and (3), there is a two-fold over-identification of $\alpha^{(s)}$ under MAR:

1. For MAR constraints to exist under the model given in (1), $\mu^{(s)}_{j|j^-}$ as defined in (2) must be equal for $2 \le j \le s \le J$ and for all X. This requires that $\alpha^{(s)} = \alpha^*$ for $2 \le j \le s \le J$.
2. MAR constraints also imply that $\mu^{(s)}_{j|j^-}$ as defined in (3) must be equal to $\mu^{(\ge j)}_{j|j^-}$ for $1 \le s < j$. This places another restriction on $\alpha^{(s)}$.

Similar over-identification exists under CCMV and NCMV.

Similar to the bivariate case, to avoid the over-identification, we again use the MAR on the residuals restriction,

$$\begin{aligned} &p_k(y_j - X\alpha^{(k)}|y_1 - X\alpha^{(k)}, \ldots, y_{j-1} - X\alpha^{(k)}, X) = \\ &\qquad \sum_{s=j}^{J} \frac{P(S=s)}{P(S \ge j)} p_s(y_j - X\alpha^{(S)}|y_1 - X\alpha^{(S)}, \ldots, y_{j-1} - X\alpha^{(S)}, X) \qquad k < j. \end{aligned} \tag{4}$$

With the conditional mean structures specified as (2) and (3), the MAR on the residuals restriction places no assumptions on $\alpha^{(s)}$. To see this, let $[Z_j|S] \simeq \left[Y_j - X\alpha^{(s)}\right]$. The MAR on the residuals constraints are

$$p_k(z_j|\overline{z}_{j-1}, X) = \sum_{s=j}^{J} \frac{P(S=s)}{P(S \ge j)} p_s(z_j|\overline{z}_{j-1}, X).$$

Note that

$$\begin{aligned} p_s(y_j, \ldots, y_1) &= p_s(y_1) \prod_{l=2}^{j} p_s(y_l|\overline{y}_{l-1}) \\ &= p_s(y_1) \prod_{l=2}^{j} \frac{\exp\left\{ \frac{1}{2\sigma^{(s)}_{l|l-}} \left(y_l - \mu^{(\ge l)}_l - X\alpha^{(s)} - \sum_{t=1}^{j-1} \beta^{(\ge j)}_t (y_t - \mu^{(\ge l)}_t - X\alpha^{(s)}) \right)^2 \right\}}{\sqrt{2\pi\sigma^{(s)}_{l|l-}}}. \end{aligned}$$

Thus,

$$p_s(z_j, \dots, z_1) = p_s(z_1) \prod_{l=2}^{j} p_s(z_l|\overline{z}_{l-1})$$
$$= p_s(z_1) \prod_{l=2}^{j} \frac{\exp\left\{ \frac{1}{2\sigma^{(s)}_{l|l-}} \left(z_l - \mu_l^{(\geq l)} - \sum_{t=1}^{j-1} \beta_t^{(\geq j)} (z_t - \mu_t^{(\geq l)}\right)^2 \right\}}{\sqrt{2\pi\sigma^{(s)}_{l|l-}}}.$$

We can further show that

$$[Z_j|\overline{Z}_{j-1}, S = s, X] \sim \mathrm{N}\left(\mu_j^{(\geq j)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(\geq j)} (Z_l - \mu_l^{(\geq j)}), \sigma^{(\geq j)}_{j|j^-} \right),$$

which is independent of s. Therefore,

$$\sum_{s=j}^{J} \frac{P(S = s)}{P(S \geq j)} p_s(z_j|\overline{z}_{j-1}, X) = \mathrm{N}\left(\mu_j^{(\geq j)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(\geq j)} (Z_l - \mu_l^{(\geq j)}), \sigma^{(\geq j)}_{j|j^-} \right).$$

Similarly, we may derive that

$$p_k(z_j|\overline{z}_{j-1}, X) = \mathrm{N}\left(\mu_j^{(s)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(s)} (Z_l - \mu_l^{(s)}), \sigma^{(j)}_{j|j^-} \right).$$

The constraints (4) thus imply

$$\begin{aligned} \beta_{j,l}^{(s)} &= \beta_{j,l}^{(\geq j)} \\ \mu_j^{(s)} &= \mu_j^{(\geq j)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(\geq j)} (\mu_l^{(s)} - \mu_l^{(\geq j)}) \\ \sigma^{(j)}_{j|j^-} &= \sigma^{(\geq j)}_{j|j^-}, \end{aligned}$$

which places no restrictions on $\alpha^{(s)}$.

The corresponding MDM is

$$\log \frac{P(S = s|\boldsymbol{Y}, X)}{P(S \geq s|\boldsymbol{Y}, X)} = \log \frac{P(S = s)p(\boldsymbol{Y}|S = s, X)}{P(\boldsymbol{Y}, S \geq s|X)}$$
$$= \log \frac{P(S = s)p_s(Y_J|\overline{Y}_{J-1}, X)p_s(Y_{J-1}|\overline{Y}_{J-2}, X)\dots p_s(Y_1|X)}{\sum_{l=s}^{J} p_l(Y_J|\overline{Y}_{J-1}, X)p_l(Y_{J-1}|\overline{Y}_{J-2}, X)\dots p_l(Y_1|X)P(S = l)}.$$

It does not have a simple form in general. However, if $\alpha^{(s)} = \alpha^*$ for all s, then

$$\log \frac{P(S = s|\boldsymbol{Y}, X)}{P(S \geq s|\boldsymbol{Y}, X)} = \log \frac{p_s(Y_1|X)P(S = s)}{\sum_{l=s}^{J} p_l(Y_1|X)P(S = l)},$$

i.e. the MDM only depends on Y_1 and X. Otherwise, the missingness is MNAR.

References

Molenberghs, G., Michiels, B., Kenward, M., and Diggle, P. (1998). Monotone Missing Data and Pattern-Mixture Models. *Statistica Neerlandica* **52,** 153–161.