Estimation in Multivariate t Nonlinear Mixed-effects Models with Missing Outcomes

Wan-Lun Wang

Department of Statistics, Graduate Institute of Statistics & Actuarial Science
Feng Chia University, Taiwan

(Joint work with Professor Tsung-I Lin)

August 24-28, 2014

Outline

Multivariate Longitudinal Data

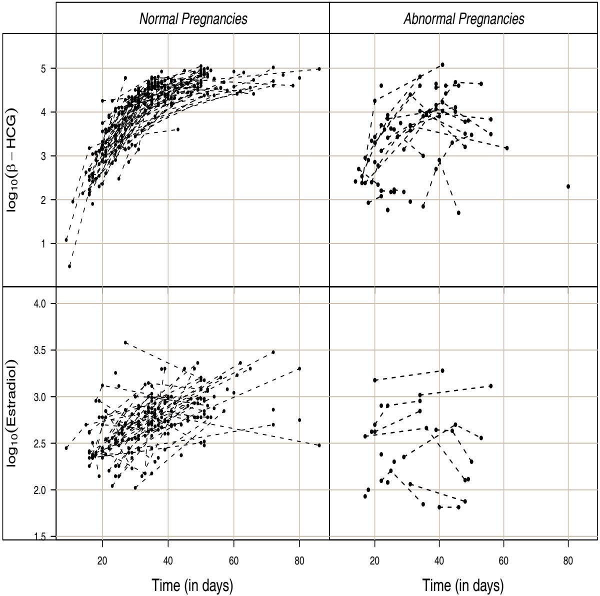
Subject i	Occasions t	Responses j				Covariates		
1	1	y_{111}	y_{121}	$\cdots$	y_{1r1}	x_{111}	$\cdots$	x_{1q1}
1	2	y_{112}	y_{122}	$\cdots$	y_{1r2}	x_{112}	$\cdots$	x_{1q2}
1	3	y_{113}	y_{123}	$\cdots$	y_{1r3}	x_{113}	$\cdots$	x_{1q3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
1	s_1	y_{11s_1}	y_{12s_1}	$\cdots$	y_{1rs_1}	x_{11s_1}	$\cdots$	x_{1qs_1}
2	1	y_{211}	y_{221}	$\cdots$	y_{2r1}	x_{211}	$\cdots$	x_{2q1}
2	2	y_{212}	y_{222}	$\cdots$	y_{2r2}	x_{212}	$\cdots$	x_{2q2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
2	s_2	y_{21s_2}	y_{22s_2}	$\cdots$	y_{2rs_2}	x_{21s_2}	$\cdots$	x_{2qs_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	1	y_{N11}	y_{N21}	$\cdots$	y_{Nr1}	x_{N11}	$\cdots$	x_{Nq1}
N	2	y_{N12}	y_{N22}	$\cdots$	y_{Nr2}	x_{N12}	$\cdots$	x_{Nq2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	s_N	y_{N1s_N}	y_{N2s_N}	$\cdots$	y_{Nrs_N}	x_{N1s_N}	$\cdots$	x_{Nqs_N}

Motivating Example: Pregnant Women Data

- The study consists of 124 women diagnosed normal pregnancies and 37 women with abnormal pregnancies over a period of two years in a private fertilization obstetrics clinic in Santiago, Chile (Marshall et al. 2006).
- For $N = 161$ young women, the beta-subunit human chorionic gonadotropin (β-HCG) and estradiol concentrations were repeatedly measured during the first trimester of pregnancy.
- Estradiol and β-HCG concentrations were measured in order to detect complications or a high risk of losing the foetus.
- The threshold after 50 days of pregnancy change appears to have a non-linear and linear relationship with the mean of $\log_{10}$ β-HCG and $\log_{10}$ estradiol, respectively.

Motivation

Return



Preliminary Analysis of Pregnancy Women Data

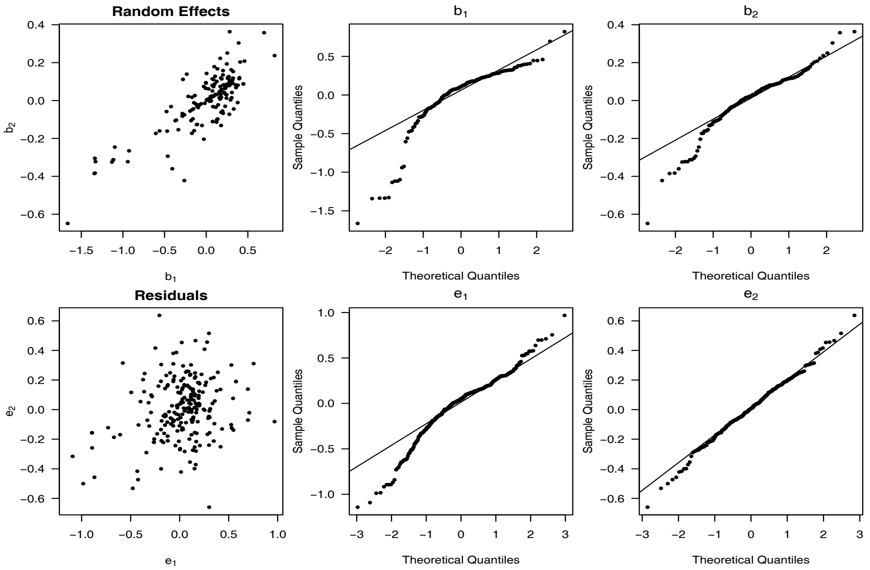
- Let $y_{i1,k}$ and $y_{i2,k}$ be the β-HCG and estradiol responses in $\log_{10}$ for woman i measured at time (days) t_{ik} ($i = 1, \ldots, 161$, $k = 1, \ldots, s_i$).

MNLMM (Marshall *et al.* 2006)

We fit the MNLMM by logistic and linear regression to $y_{i1,k}$ and $y_{i2,k}$:

$$\begin{aligned} y_{i1,k} &= \frac{\beta_1 + b_{i1}}{1 + \exp\left\{(\beta_2 - t_{ik})/\beta_3\right\}} + e_{i1,k}, \\ y_{i2,k} &= \beta_4 + \beta_5 t_{ik} + b_{i2} + e_{i2,k}. \end{aligned}$$

- Fixed effects $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^{\mathrm{T}}$ describe the mean profiles of the bivariate responses.
- Random effects $(b_{i1}, b_{i2})^{\mathrm{T}} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{D})$ describe how the profile of the ith woman deviates from the mean profiles.
- Within-subject errors $(e_{i1,1}, \cdots, e_{i1,s_i}, e_{i2,1}, \cdots, e_{i2,s_i})^{\mathrm{T}} \sim \mathcal{N}_{2s_i}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_{s_i})$ are residuals and uncorrelated with the random effects.


Random Effects
b1
b2
0.4
0.2
0.0
−0.2
−0.4
−0.6
−1.5 −1.0 −0.5 0.0 0.5
Sample Quantiles
0.5
0.0
−0.5
−1.0
−1.5
−2 −1 0 1 2
Theoretical Quantiles
Residuals
e1
e2
0.6
−1.0 −0.5 0.0 0.5 1.0
1.0
−3 −2 −1 0 1 2 3

Multivariate t Nonlinear Mixed-effects Model

Notation ($i = 1, \ldots, N$; $j = 1, \ldots, r$; $t = 1, \ldots, s_i$)

- $\boldsymbol{Y}_i = [\boldsymbol{y}_{i1} : \cdots : \boldsymbol{y}_{ir}]$: $s_i \times r$ outcome matrix of subject i, $\boldsymbol{y}_i = \text{vec}(\boldsymbol{Y}_i)$
- $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijs_i})^{\mathrm{T}}$: response variable j from subject i over time t
- $\boldsymbol{E}_i = [\boldsymbol{e}_{i1} : \cdots : \boldsymbol{e}_{ir}]$: $s_i \times r$ within-subject errors matrix, $\boldsymbol{e}_i = \text{vec}(\boldsymbol{E}_i)$
- $\boldsymbol{X}_i$: covariates variables
- Write $n_i = s_i r$, for $i = 1, \ldots, N$, $p = \sum_{j=1}^{r} p_j$ and $q = \sum_{j=1}^{r} q_j$.

MtNLMM for the ith subject

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i(\boldsymbol{\eta}_i, \boldsymbol{X}_i) + \boldsymbol{e}_i, \quad \text{with} \begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{e}_i \end{bmatrix} \sim t_{q+n_i}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R}_i \end{bmatrix}, \nu \right) \tag{1}$$

- $\boldsymbol{\mu}_i$ is a nonlinear vector-valued and differentiable function.
- The fixed effects $\boldsymbol{\beta}$ and the random effects $\boldsymbol{b}_i$ can be incorporated into the model through $\boldsymbol{\eta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$ such that $\boldsymbol{\mu}_i(\boldsymbol{\eta}_i, \boldsymbol{X}_i) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)$.
- $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are design matrices of size $g \times p$ and $g \times q$ for fixed effects and random effects.

Multivariate t Nonlinear Mixed-effects Model

Notation ($i = 1, \ldots, N$; $j = 1, \ldots, r$; $t = 1, \ldots, s_i$)

- $\boldsymbol{Y}_i = [\boldsymbol{y}_{i1} : \cdots : \boldsymbol{y}_{ir}]$: $s_i \times r$ outcome matrix of subject i, $\boldsymbol{y}_i = \text{vec}(\boldsymbol{Y}_i)$
- $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijs_i})^{\mathrm{T}}$: response variable j from subject i over time t
- $\boldsymbol{E}_i = [\boldsymbol{e}_{i1} : \cdots : \boldsymbol{e}_{ir}]$: $s_i \times r$ within-subject errors matrix, $\boldsymbol{e}_i = \text{vec}(\boldsymbol{E}_i)$
- $\boldsymbol{X}_i$: covariates variables
- Write $n_i = s_i r$, for $i = 1, \ldots, N$, $p = \sum_{j=1}^{r} p_j$ and $q = \sum_{j=1}^{r} q_j$.

MtNLMM for the ith subject

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i(\boldsymbol{\eta}_i, \boldsymbol{X}_i) + \boldsymbol{e}_i, \quad \text{with} \begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{e}_i \end{bmatrix} \sim t_{q+n_i}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_i \end{bmatrix}, \nu \right) \tag{1}$$

- $\boldsymbol{\mu}_i$ is a nonlinear vector-valued and differentiable function.
- The fixed effects $\boldsymbol{\beta}$ and the random effects $\boldsymbol{b}_i$ can be incorporated into the model through $\boldsymbol{\eta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$ such that $\boldsymbol{\mu}_i(\boldsymbol{\eta}_i, \boldsymbol{X}_i) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)$.
- $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are design matrices of size $g \times p$ and $g \times q$ for fixed effects and random effects.

The model for the jth column (outcome) of $\boldsymbol{Y}_i$

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu}_{ij}(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij}) + \boldsymbol{e}_{ij}$$

- $\boldsymbol{\mu}_{ij}(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij}) = (\mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,1}), \ldots, \mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,s_i}))^{\mathrm{T}}$ is the vector of a link function relating the jth outcome $\boldsymbol{y}_{ij}$ over s_i time-points to the covariates $\boldsymbol{x}_{ij}$ by the mixed effects β and $\boldsymbol{b}_i$.
- $\boldsymbol{e}_{ij} \sim t_{s_i}(\boldsymbol{0}, \sigma_{jj}\boldsymbol{C}_i, \nu)$

The model for the kth row (occasion) of $\boldsymbol{Y}_i$

$$\boldsymbol{y}_{i,k} = \boldsymbol{\mu}_i^k(\boldsymbol{\eta}_i, \boldsymbol{x}_{ik}) + \boldsymbol{e}_{i,k}$$

- $\boldsymbol{\mu}_i^k(\boldsymbol{\eta}_i, \boldsymbol{x}_{ik}) = (\mu_1(\boldsymbol{\eta}_i, \boldsymbol{x}_{i1,k}), \ldots, \mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,k}), \ldots, \mu_r(\boldsymbol{\eta}_i, \boldsymbol{x}_{ir,k}))$ is a vector of r link functions with each function relating each outcome variable at the same time to the covariates $\boldsymbol{x}_{i,k}$ by the mixed effects β and $\boldsymbol{b}_i$.
- $\boldsymbol{e}_{i,k} \sim t_r(\boldsymbol{0}, \boldsymbol{\Sigma}, \nu)$

The model for the jth column (outcome) of $\boldsymbol{Y}_i$

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu}_{ij}(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij}) + \boldsymbol{e}_{ij}$$

- $\boldsymbol{\mu}_{ij}(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij}) = (\mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,1}), \ldots, \mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,s_i}))^{\mathrm{T}}$ is the vector of a link function relating the jth outcome $\boldsymbol{y}_{ij}$ over s_i time-points to the covariates $\boldsymbol{x}_{ij}$ by the mixed effects $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$.
- $\boldsymbol{e}_{ij} \sim t_{s_i}(\boldsymbol{0}, \sigma_{jj}\boldsymbol{C}_i, \nu)$

The model for the kth row (occasion) of $\boldsymbol{Y}_i$

$$\boldsymbol{y}_{i,k} = \boldsymbol{\mu}_i^k(\boldsymbol{\eta}_i, \boldsymbol{x}_{ik}) + \boldsymbol{e}_{i,k}$$

- $\boldsymbol{\mu}_i^k(\boldsymbol{\eta}_i, \boldsymbol{x}_{ik}) = (\mu_1(\boldsymbol{\eta}_i, \boldsymbol{x}_{i1,k}), \ldots, \mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,k}), \ldots, \mu_r(\boldsymbol{\eta}_i, \boldsymbol{x}_{ir,k}))$ is a vector of r link functions with each function relating each outcome variable at the same time to the covariates $\boldsymbol{x}_{i,k}$ by the mixed effects $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$.
- $\boldsymbol{e}_{i,k} \sim t_r(\boldsymbol{0}, \boldsymbol{\Sigma}, \nu)$

- Under the above assumption, we have

$$\text{cov}(\boldsymbol{E}_i) = \boldsymbol{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i$$

where a damped exponential correlation (DEC; Muñoz *et al.* 1992) structure is considered:

$$\boldsymbol{C}_i = \boldsymbol{C}_i(\phi, \gamma; \boldsymbol{t}_i) = \left[\phi^{|t_{ik}-t_{ik'}|^{\gamma}}\right], \quad 0 \le \phi < 1, \quad 0 \le \gamma.$$

- Let $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{\Sigma}, \phi, \gamma, \nu\}$ be the entire model parameters.

Two-level hierarchy of model (1)

Introducing a set of scaling weight variables $\tau_i \sim \text{Gamma}(\nu/2, \nu/2)$ leads to

$$\begin{aligned} \boldsymbol{y}_i|(\boldsymbol{b}_i, \tau_i) &\sim \mathcal{N}_{n_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{b}_i), \tau_i^{-1}\boldsymbol{R}_i), \\ \boldsymbol{b}_i|\tau_i &\sim \mathcal{N}_q(\boldsymbol{0}, \tau_i^{-1}\boldsymbol{D}). \end{aligned}$$

MtNLMM with Pseudo Data

Using a Taylor series expansion for model (1) around $\hat{\boldsymbol{\eta}}_i^{(h)} = \boldsymbol{A}_i\hat{\boldsymbol{\beta}}^{(h)} + \boldsymbol{B}_i\hat{\boldsymbol{b}}_i^{(h)}$ and letting $\dot{\mu}_j(\hat{\boldsymbol{\eta}}_i^{(h)}, \boldsymbol{x}_{ij,k})$ be the first partial derivative of $\mu_j(\hat{\boldsymbol{\eta}}_i^{(h)}, \boldsymbol{x}_{ij,k})$ with respect to $\boldsymbol{\eta}_i$, model (1) can be rewritten as

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i + \boldsymbol{e}_i \qquad (2)$$

- $\tilde{\boldsymbol{y}}_i$ is an $n_i \times 1$ vector composed of r *pseudo-response* vectors $\tilde{\boldsymbol{y}}_{ij} = (\tilde{y}_{ij,1}, \cdots, \tilde{y}_{ij,s_i})^{\mathrm{T}}$ in which

$$\tilde{y}_{ij,k} = y_{ij,k} - \mu_j(\hat{\boldsymbol{\eta}}_i^{(h)}, \boldsymbol{x}_{ij,k}) + \tilde{\boldsymbol{x}}_{ij,k}\hat{\boldsymbol{\beta}}^{(h)} + \tilde{\boldsymbol{z}}_{ij,k}\hat{\boldsymbol{b}}_i^{(h)}$$

- $\tilde{\boldsymbol{X}}_i$ is an $n_i \times p$ matrix with rows made up of $p \times 1$ vector $\tilde{\boldsymbol{x}}_{ij,k} = \dot{\mu}_j(\hat{\boldsymbol{\eta}}_i^{(h)}, \boldsymbol{x}_{ij,k})^{\mathrm{T}}\boldsymbol{A}_i$
- $\tilde{\boldsymbol{Z}}_i$ is an $n_i \times q$ matrix with rows made up of $q \times 1$ vector $\tilde{\boldsymbol{z}}_{ij,k} = \dot{\mu}_j(\hat{\boldsymbol{\eta}}_i^{(h)}, \boldsymbol{x}_{ij,k})^{\mathrm{T}}\boldsymbol{B}_i$

It follows that

$$\tilde{\boldsymbol{y}}_i \sim t_{n_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}, \tilde{\boldsymbol{\Lambda}}_i, \nu)$$

where $\tilde{\boldsymbol{\Lambda}}_i = \tilde{\boldsymbol{Z}}_i\boldsymbol{D}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}} + \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i$.

GLS-SC

Three-level Hierarchy for MtNLMM with Pseudo Data

Treating the random effects $\boldsymbol{b} = \{\boldsymbol{b}_i\}_{i=1}^N$ and scaling weights $\boldsymbol{\tau} = \{\tau_i\}_{i=1}^N$ as latent data, the complete-data log-likelihood function is obtained based on

$$\begin{aligned} \tilde{\boldsymbol{y}}_i|(\boldsymbol{b}_i,\tau_i) &\sim \mathcal{N}_{n_i}\big(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i, \tau_i^{-1}\boldsymbol{R}_i\big), \\ \boldsymbol{b}_i|\tau_i &\sim \mathcal{N}_q(\mathbf{0}, \tau_i^{-1}\boldsymbol{D}), \\ \tau_i &\sim \text{Gamma}(\nu/2, \nu/2). \end{aligned}$$

Proposition

Using the Bayes' theorem, simple matrix algebra gives

$$\begin{aligned} \boldsymbol{b}_i|\tilde{\boldsymbol{y}}_i &\sim t_q\Big(\boldsymbol{D}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_i^{-1}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta}), \big(\frac{\nu + \Delta_{\tilde{\boldsymbol{y}}_i}}{\nu + n_i}\big)(\boldsymbol{D}^{-1} + \tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}\tilde{\boldsymbol{Z}}_i)^{-1}, \nu + n_i\Big), \\ \tau_i|\tilde{\boldsymbol{y}}_i &\sim \text{Gamma}\Big(\frac{n_i+\nu}{2}, \frac{\nu + (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_i^{-1}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})}{2}\Big). \end{aligned}$$

Pseudo-data ECM Algorithm (E-step)

Let $\hat{\boldsymbol{\theta}}^{(h)} = \{\hat{\boldsymbol{\beta}}^{(h)}, \hat{\boldsymbol{D}}^{(h)}, \hat{\boldsymbol{\Sigma}}^{(h)}, \hat{\phi}^{(h)}, \hat{\gamma}^{(h)}, \hat{\nu}^{(h)}\}$. Evaluate the Q-function:

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)}) = -\frac{1}{2}\sum_{i=1}^{N}\Big\{\log|\boldsymbol{R}_i| + \log|\boldsymbol{D}| + \mathrm{tr}(\boldsymbol{D}^{-1}\hat{\boldsymbol{B}}_i^{(h)}) + \mathrm{tr}(\boldsymbol{R}_i^{-1}\hat{\boldsymbol{\Psi}}_i^{(h)}(\boldsymbol{\beta})) -\nu\Big(\log(\frac{\nu}{2}) + \hat{\kappa}_i^{(h)} - \hat{\tau}_i^{(h)}\Big) + 2\log\Gamma(\frac{\nu}{2})\Big\} \quad (3)$$

where

$$\hat{\tau}_i^{(h)} = E[\tau_i|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(h)}] = (\hat{\nu}^{(h)} + n_i)/(\hat{\nu}^{(h)} + \hat{\Delta}_{\tilde{\boldsymbol{y}}_i}^{(h)}),$$

$$\hat{\kappa}_i^{(h)} = E[\log\tau_i|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(h)}] = \mathcal{D}_g\Big(\frac{\hat{\nu}^{(h)} + n_i}{2}\Big) - \log\Big(\frac{\hat{\nu}^{(h)} + \hat{\Delta}_{\tilde{\boldsymbol{y}}_i}^{(h)}}{2}\Big),$$

$$\hat{\boldsymbol{B}}_i^{(h)} = E\Big[\tau_i\boldsymbol{b}_i\boldsymbol{b}_i^{\mathrm{T}}|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(h)}\Big] = \hat{\tau}_i^{(h)}\hat{\boldsymbol{b}}_i^{(h)}\hat{\boldsymbol{b}}_i^{(h)\mathrm{T}} + \hat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(h)},$$

$$\hat{\boldsymbol{\Psi}}_i^{(h)} = E\Big[\tau_i\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(h)}\Big] = \hat{\tau}_i^{(h)}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i^{(h)})(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i^{(h)})^{\mathrm{T}} + \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(h)}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}$$

with $\hat{\boldsymbol{b}}_i^{(h)} = \hat{\boldsymbol{D}}^{(h)}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_i^{(h)-1}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}}^{(h)})$ and $\hat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(h)} = (\hat{\boldsymbol{D}}^{(h)-1} + \tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\hat{\boldsymbol{R}}_i^{(h)-1}\tilde{\boldsymbol{Z}}_i)^{-1}$.

Pseudo-data ECM Algorithm (CM-steps)

$\hat{\tau}_i$

1. Update the current estimates $\hat{\beta}^{(h)}$, $\hat{D}^{(h)}$, and $\hat{\Sigma}^{(h)}$ by

$$\hat{\beta}^{(h+1)} = \Big(\sum_{i=1}^{N} \hat{\tau}_i^{(h)} \tilde{X}_i^{\mathrm{T}} \hat{R}_i^{(h)-1} \tilde{X}_i\Big)^{-1} \sum_{i=1}^{N} \hat{\tau}_i^{(h)} \tilde{X}_i^{\mathrm{T}} \hat{R}_i^{(h)-1} (\tilde{y}_i - \tilde{Z}_i \hat{b}_i^{(h)}),$$

$$\hat{D}^{(h+1)} = N^{-1} \sum_{i=1}^{N} \hat{B}_i^{(h)},$$

$$\hat{\sigma}_{jl}^{(h+1)} = \begin{cases} \Big(\sum_{i=1}^{N} s_i\Big)^{-1} \sum_{i=1}^{N} \mathrm{tr}\Big(\hat{C}_i^{(h)} \psi_{ijl}^{(h)}(\hat{\beta}^{(h+1)})\Big), & \text{for } j = l; \\ \Big(2\sum_{j=1}^{N} s_i\Big)^{-1} \sum_{i=1}^{N} \mathrm{tr}\Big(\hat{C}_i^{(h)} \big[\psi_{ijl}^{(h)}(\hat{\beta}^{(h+1)}) + \psi_{ilj}^{(h)}(\hat{\beta}^{(h+1)})\big]\Big), & \text{for } j \neq l \end{cases}$$

2. Use the nlminb routine to update the $(\hat{\phi}^{(h)}, \hat{\gamma}^{(h)})$ and $\hat{\nu}^{(h)}$ sequentially.

$$(\hat{\phi}^{(h+1)}, \hat{\gamma}^{(h+1)}) = \arg\max_{(\phi,\gamma)} \Big\{\frac{r}{2}\sum_{i=1}^{N} \log|C_i^{-1}| - \frac{1}{2}\mathrm{tr}\Big((\hat{\Sigma}^{-1(h+1)} \otimes C_i^{-1})\hat{\Psi}_i^{(h+1/2)}(\hat{\beta}^{(h+1)})\Big)\Big\},$$

and

$$\hat{\nu}^{(h+1)} = \arg\max_{\nu} \Big\{\frac{\nu}{2}\sum_{i=1}^{N}(\log(\frac{\nu}{2}) + \hat{\kappa}_i^{(h)} - \hat{\tau}_i^{(h)}) - N\log\Gamma(\frac{\nu}{2})\Big\}.$$

Multivariate Longitudinal Data with Missing Outcomes

Subject i	Occasions t	Responses j				Covariates		
1	1	y_{111}	NA	$\cdots$	y_{1r1}	x_{111}	$\cdots$	x_{1q1}
1	2	y_{112}	y_{122}	$\cdots$	y_{1r2}	x_{112}	$\cdots$	x_{1q2}
1	3	y_{113}	y_{123}	$\cdots$	NA	x_{113}	$\cdots$	x_{1q3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
1	s_1	NA	NA	$\cdots$	y_{1rs_1}	x_{11s_1}	$\cdots$	x_{1qs_1}
2	1	NA	y_{221}	$\cdots$	y_{2r1}	x_{211}	$\cdots$	x_{2q1}
2	2	y_{212}	y_{222}	$\cdots$	y_{2r2}	x_{212}	$\cdots$	x_{2q2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
2	s_2	NA	y_{22s_2}	$\cdots$	NA	x_{21s_2}	$\cdots$	x_{2qs_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	1	y_{N11}	y_{N21}	$\cdots$	y_{Nr1}	x_{N11}	$\cdots$	x_{Nq1}
N	2	y_{N12}	NA	$\cdots$	NA	x_{N12}	$\cdots$	x_{Nq2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	s_N	y_{N1s_N}	y_{N2s_N}	$\cdots$	NA	x_{N1s_N}	$\cdots$	x_{Nqs_N}

Incomplete-data Framework

- We partitioned $\tilde{\boldsymbol{y}}_i$ $(n_i \times 1)$ into two components $(\tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \tilde{\boldsymbol{y}}_i^{\mathrm{m}})$ accordingly.
  - $\tilde{\boldsymbol{y}}_i^{\mathrm{o}}$ $(n_i^{\mathrm{o}} \times 1)$: the observed component
  - $\tilde{\boldsymbol{y}}_i^{\mathrm{m}}$ $((n_i - n_i^{\mathrm{o}}) \times 1)$: the missing component
- Auxiliary permutation matrices

Example

  - $\boldsymbol{O}_i$ $(n_i^{\mathrm{o}} \times n_i)$: $\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{y}}_i$; $\tilde{\boldsymbol{X}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{X}}_i$; $\tilde{\boldsymbol{Z}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{Z}}_i$
  - $\boldsymbol{M}_i$ $((n_i - n_i^{\mathrm{o}}) \times n_i)$: $\tilde{\boldsymbol{y}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{y}}_i$; $\tilde{\boldsymbol{X}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{X}}_i$; $\tilde{\boldsymbol{Z}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{Z}}_i$
- Model (2) can be rewritten as

$$\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \tilde{\boldsymbol{X}}_i^{\mathrm{o}} \boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i^{\mathrm{o}} \boldsymbol{b}_i + \boldsymbol{e}_i^{\mathrm{o}}.$$

- Missing at Random (MAR; Rubin 1976)

Proposition

$$P(\boldsymbol{r} | \boldsymbol{y}^{\mathrm{o}}, \boldsymbol{y}^{\mathrm{m}}, \boldsymbol{x}, \boldsymbol{\theta}) = P(\boldsymbol{r} | \boldsymbol{y}^{\mathrm{o}}, \boldsymbol{x}, \boldsymbol{\theta})$$

Condition

Incomplete-data Framework

- We partitioned $\tilde{\boldsymbol{y}}_i$ $(n_i \times 1)$ into two components $(\tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \tilde{\boldsymbol{y}}_i^{\mathrm{m}})$ accordingly.
  - $\tilde{\boldsymbol{y}}_i^{\mathrm{o}}$ $(n_i^{\mathrm{o}} \times 1)$: the observed component
  - $\tilde{\boldsymbol{y}}_i^{\mathrm{m}}$ $((n_i - n_i^{\mathrm{o}}) \times 1)$: the missing component
- Auxiliary permutation matrices (Example)
  - $\boldsymbol{O}_i$ $(n_i^{\mathrm{o}} \times n_i)$: $\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{y}}_i$; $\tilde{\boldsymbol{X}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{X}}_i$; $\tilde{\boldsymbol{Z}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{Z}}_i$
  - $\boldsymbol{M}_i$ $((n_i - n_i^{\mathrm{o}}) \times n_i)$: $\tilde{\boldsymbol{y}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{y}}_i$; $\tilde{\boldsymbol{X}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{X}}_i$; $\tilde{\boldsymbol{Z}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{Z}}_i$
- Model (2) can be rewritten as

$$\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \tilde{\boldsymbol{X}}_i^{\mathrm{o}} \boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i^{\mathrm{o}} \boldsymbol{b}_i + \boldsymbol{e}_i^{\mathrm{o}}.$$

- Missing at Random (MAR; Rubin 1976) (Proposition)

$$P(\boldsymbol{r} | \boldsymbol{y}^{\mathrm{o}}, \boldsymbol{y}^{\mathrm{m}}, \boldsymbol{x}, \boldsymbol{\theta}) = P(\boldsymbol{r} | \boldsymbol{y}^{\mathrm{o}}, \boldsymbol{x}, \boldsymbol{\theta})$$

(Condition)

Incomplete-data Framework

- We partitioned $\tilde{\boldsymbol{y}}_i$ $(n_i \times 1)$ into two components $(\tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \tilde{\boldsymbol{y}}_i^{\mathrm{m}})$ accordingly.
  - $\tilde{\boldsymbol{y}}_i^{\mathrm{o}}$ $(n_i^{\mathrm{o}} \times 1)$: the observed component
  - $\tilde{\boldsymbol{y}}_i^{\mathrm{m}}$ $((n_i - n_i^{\mathrm{o}}) \times 1)$: the missing component
- Auxiliary permutation matrices

  Example
  - $\boldsymbol{O}_i$ $(n_i^{\mathrm{o}} \times n_i)$: $\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{y}}_i$; $\tilde{\boldsymbol{X}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{X}}_i$; $\tilde{\boldsymbol{Z}}_i^{\mathrm{o}} = \boldsymbol{O}_i \tilde{\boldsymbol{Z}}_i$
  - $\boldsymbol{M}_i$ $((n_i - n_i^{\mathrm{o}}) \times n_i)$: $\tilde{\boldsymbol{y}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{y}}_i$; $\tilde{\boldsymbol{X}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{X}}_i$; $\tilde{\boldsymbol{Z}}_i^{\mathrm{m}} = \boldsymbol{M}_i \tilde{\boldsymbol{Z}}_i$
- Model (2) can be rewritten as

$$\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \tilde{\boldsymbol{X}}_i^{\mathrm{o}} \boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i^{\mathrm{o}} \boldsymbol{b}_i + \boldsymbol{e}_i^{\mathrm{o}}.$$

- Missing at Random (MAR; Rubin 1976)

  Proposition

$$P(\boldsymbol{r} | \boldsymbol{y}^{\mathrm{o}}, \boldsymbol{y}^{\mathrm{m}}, \boldsymbol{x}, \boldsymbol{\theta}) = P(\boldsymbol{r} | \boldsymbol{y}^{\mathrm{o}}, \boldsymbol{x}, \boldsymbol{\theta})$$

Condition

Modified Pseudo-data ECM Algorithm

Impute the missing (pseudo) responses at each iteration of ECM by

$$\hat{\tilde{\boldsymbol{y}}}_i^{\mathrm{m}(h)} = E[\tilde{\boldsymbol{y}}_i^{\mathrm{m}} | \tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(h)}] = \tilde{\boldsymbol{X}}_i^{\mathrm{m}} \hat{\boldsymbol{\beta}}^{(h)} + \boldsymbol{M}_i \hat{\tilde{\boldsymbol{\Lambda}}}_i^{(h)} \hat{\tilde{\boldsymbol{S}}}_i^{\mathrm{oo}(h)} (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}^{(h)}).$$

It follows that

$$\hat{\tilde{\boldsymbol{y}}}_i^{(h)} = E[\tilde{\boldsymbol{y}}_i | \tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(h)}] = \tilde{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}^{(h)} + \hat{\tilde{\boldsymbol{\Lambda}}}_i^{(h)} \hat{\tilde{\boldsymbol{S}}}_i^{\mathrm{oo}(h)-1} (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}^{(h)}).$$

The Q-function (3) of ECM are modified by changing

$$\begin{aligned}
\hat{\tau}_i^{(h)} &= E\left[\tau_i | \tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(h)}\right] = (\hat{\nu}^{(h)} + n_i^{\mathrm{o}})/(\hat{\nu}^{(h)} + \hat{\Delta}_{\tilde{\boldsymbol{y}}_i^{\mathrm{o}}}^{(h)}), \\
\hat{\kappa}_i^{(h)} &= E\left[\log \tau_i | \tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(h)}\right] = \mathcal{D}_{\mathcal{G}}\left(\frac{\hat{\nu}^{(h)} + n_i^{\mathrm{o}}}{2}\right) - \log\left(\frac{\hat{\nu}^{(h)} + \hat{\Delta}_{\tilde{\boldsymbol{y}}_i^{\mathrm{o}}}^{(h)}}{2}\right), \\
\hat{\boldsymbol{B}}_i^{(h)} &= E\left[\tau_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} | \tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(h)}\right] = \hat{\tau}_i^{(h)} \hat{\boldsymbol{b}}_i^{(h)} \hat{\boldsymbol{b}}_i^{(h)\mathrm{T}} + \left(\hat{\boldsymbol{D}}^{(h)^{-1}} + \tilde{\boldsymbol{Z}}_i^{\mathrm{o}\,\mathrm{T}} \hat{\boldsymbol{R}}_i^{\mathrm{oo}(h)^{-1}} \tilde{\boldsymbol{Z}}_i^{\mathrm{o}}\right)^{-1}, \\
\hat{\boldsymbol{\Psi}}_i^{(h)} &= E\left[\tau_i \tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_i^{\mathrm{T}} | \tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(h)}\right] = \hat{\tau}_i^{(h)} \hat{\tilde{\boldsymbol{e}}}_i^{(h)} \hat{\tilde{\boldsymbol{e}}}_i^{(h)\mathrm{T}} + (\boldsymbol{I}_{n_i} - \hat{\boldsymbol{R}}_i^{(h)} \hat{\tilde{\boldsymbol{S}}}_i^{\mathrm{oo}(h)}) \hat{\boldsymbol{R}}_i^{(h)},
\end{aligned}$$

where $\hat{\boldsymbol{b}}_i^{(h)} = \hat{\boldsymbol{D}}^{(h)} \tilde{\boldsymbol{Z}}_i^{\mathrm{T}} \hat{\tilde{\boldsymbol{S}}}_i^{\mathrm{oo}(h)} (\hat{\tilde{\boldsymbol{y}}}_i^{(h)} - \tilde{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}^{(h)})$ and $\hat{\tilde{\boldsymbol{e}}}_i^{(h)} = \hat{\tilde{\boldsymbol{y}}}_i^{(h)} - \tilde{\boldsymbol{X}}_i \boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i \hat{\boldsymbol{b}}_i^{(h)}$.

Imputation of Missing Values

The predictor of raw missing values is

$$\hat{\boldsymbol{y}}_i^{\mathrm{m}} = \hat{\tilde{\boldsymbol{y}}}_i^{\mathrm{m}} + \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i) - \tilde{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i.$$

Estimation of Random Effects

Substituting $\hat{\theta}$ into

$$\boldsymbol{b}_i(\boldsymbol{\theta}) = \boldsymbol{D}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{S}}_i^{\mathrm{oo}}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})$$

yields empirical Bayes estimates of random effects, denoted by $\hat{\boldsymbol{b}}_i = \boldsymbol{b}_i(\hat{\boldsymbol{\theta}})$.

Fitted Values of Responses

Substituting the estimates of $\hat{\beta}$ and $\hat{b}_i$ into the nonlinear function μ_i yields

$$\hat{\boldsymbol{y}}_i = \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i).$$

Imputation of Missing Values

The predictor of raw missing values is

$$\hat{\boldsymbol{y}}_i^{\mathrm{m}} = \hat{\tilde{\boldsymbol{y}}}_i^{\mathrm{m}} + \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i) - \tilde{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i.$$

Estimation of Random Effects

Substituting $\hat{\boldsymbol{\theta}}$ into

$$\boldsymbol{b}_i(\boldsymbol{\theta}) = \boldsymbol{D}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{S}}_i^{\mathrm{oo}}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})$$

yields empirical Bayes estimates of random effects, denoted by $\hat{\boldsymbol{b}}_i = \boldsymbol{b}_i(\hat{\boldsymbol{\theta}})$.

Fitted Values of Responses

Substituting the estimates of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}_i$ into the nonlinear function $\boldsymbol{\mu}_i$ yields

$$\hat{\boldsymbol{y}}_i = \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i).$$

Imputation of Missing Values

The predictor of raw missing values is

$$\hat{\boldsymbol{y}}_i^{\mathrm{m}} = \hat{\tilde{\boldsymbol{y}}}_i^{\mathrm{m}} + \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i) - \tilde{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i.$$

Estimation of Random Effects

Substituting $\hat{\boldsymbol{\theta}}$ into

$$\boldsymbol{b}_i(\boldsymbol{\theta}) = \boldsymbol{D}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{S}}_i^{\mathrm{oo}}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})$$

yields empirical Bayes estimates of random effects, denoted by $\hat{\boldsymbol{b}}_i = \boldsymbol{b}_i(\hat{\boldsymbol{\theta}})$.

Fitted Values of Responses

Substituting the estimates of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}_i$ into the nonlinear function $\boldsymbol{\mu}_i$ yields

$$\hat{\boldsymbol{y}}_i = \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i).$$

Analysis of Pregnant Women Data

Figure

- Let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^{\mathrm{T}}, \boldsymbol{y}_{i2}^{\mathrm{T}})^{\mathrm{T}}$ for patient i, where $\boldsymbol{y}_{i1} = \log_{10}\beta$-HCG and $\boldsymbol{y}_{i2} = \log_{10}$ estradiol.
- Employ two distinct curves for the ith woman at time $t_{ik} = \text{day}_{ik}/7$ (weeks) in group l ($l = 1$ for normal; $l = 2$ for abnormal):

$$\begin{aligned} y_{i1,k}^{(l)} &= \frac{\beta_{1l} + b_{i1}^{(l)}}{1 + \exp\left\{(\beta_{2l} - t_{ik})/\beta_{3l}\right\}} + e_{i1,k}^{(l)}; \\ y_{i2,k}^{(l)} &= \beta_{4l} + \beta_{5l} t_{ik} + b_{i2}^{(l)} + e_{i2,k}^{(l)}. \end{aligned}$$

- MNLMM: $(b_{i1}^{(l)}, b_{i2}^{(l)}) \sim N_2(\mathbf{0}, \boldsymbol{D}_l) \perp (\boldsymbol{e}_{i1}^{(l)\mathrm{T}}, \boldsymbol{e}_{i2}^{(l)\mathrm{T}})^{\mathrm{T}} \sim N_{2s_i}(\mathbf{0}, \boldsymbol{R}_{il})$
- MtNLMM: $(b_{i1}^{(l)}, b_{i2}^{(l)}) \sim t_2(\mathbf{0}, \boldsymbol{D}_l, \nu_l) \perp (\boldsymbol{e}_{i1}^{(l)\mathrm{T}}, \boldsymbol{e}_{i2}^{(l)\mathrm{T}})^{\mathrm{T}} \sim t_{2s_i}(\mathbf{0}, \boldsymbol{R}_{il}, \nu_l)$
- Since $\boldsymbol{R}_{il} = \boldsymbol{\Sigma}_l \otimes \boldsymbol{C}_{il}$, we adopt the UNC, AR(1), and DEC for $\boldsymbol{C}_{il}$.
- The mean- and variance-homogeneity model is also considered: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, $\boldsymbol{D}_1 = \boldsymbol{D}_2$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, $\phi_1 = \phi_2$, $\gamma_1 = \gamma_2$ and $\nu_1 = \nu_2$

Analysis of Pregnant Women Data

Figure

- Let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^{\mathrm{T}}, \boldsymbol{y}_{i2}^{\mathrm{T}})^{\mathrm{T}}$ for patient i, where $\boldsymbol{y}_{i1} = \log_{10}$ β-HCG and $\boldsymbol{y}_{i2} = \log_{10}$ estradiol.
- Employ two distinct curves for the ith woman at time $t_{ik} = \mathrm{day}_{ik}/7$ (weeks) in group l ($l = 1$ for normal; $l = 2$ for abnormal):

$$
\begin{aligned}
y_{i1,k}^{(l)} &= \frac{\beta_{1l} + b_{i1}^{(l)}}{1 + \exp\left\{(\beta_{2l} - t_{ik})/\beta_{3l}\right\}} + e_{i1,k}^{(l)}; \\
y_{i2,k}^{(l)} &= \beta_{4l} + \beta_{5l} t_{ik} + b_{i2}^{(l)} + e_{i2,k}^{(l)}.
\end{aligned}
$$

- MNLMM: $(b_{i1}^{(l)}, b_{i2}^{(l)}) \sim N_2(\mathbf{0}, \boldsymbol{D}_l) \perp (\boldsymbol{e}_{i1}^{(l)\mathrm{T}}, \boldsymbol{e}_{i2}^{(l)\mathrm{T}})^{\mathrm{T}} \sim N_{2s_i}(\mathbf{0}, \boldsymbol{R}_{il})$
- MtNLMM: $(b_{i1}^{(l)}, b_{i2}^{(l)}) \sim t_2(\mathbf{0}, \boldsymbol{D}_l, \nu_l) \perp (\boldsymbol{e}_{i1}^{(l)\mathrm{T}}, \boldsymbol{e}_{i2}^{(l)\mathrm{T}})^{\mathrm{T}} \sim t_{2s_i}(\mathbf{0}, \boldsymbol{R}_{il}, \nu_l)$
- Since $\boldsymbol{R}_{il} = \boldsymbol{\Sigma}_l \otimes \boldsymbol{C}_{il}$, we adopt the UNC, AR(1), and DEC for $\boldsymbol{C}_{il}$.
- The mean- and variance-homogeneity model is also considered: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, $\boldsymbol{D}_1 = \boldsymbol{D}_2$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, $\phi_1 = \phi_2$, $\gamma_1 = \gamma_2$ and $\nu_1 = \nu_2$

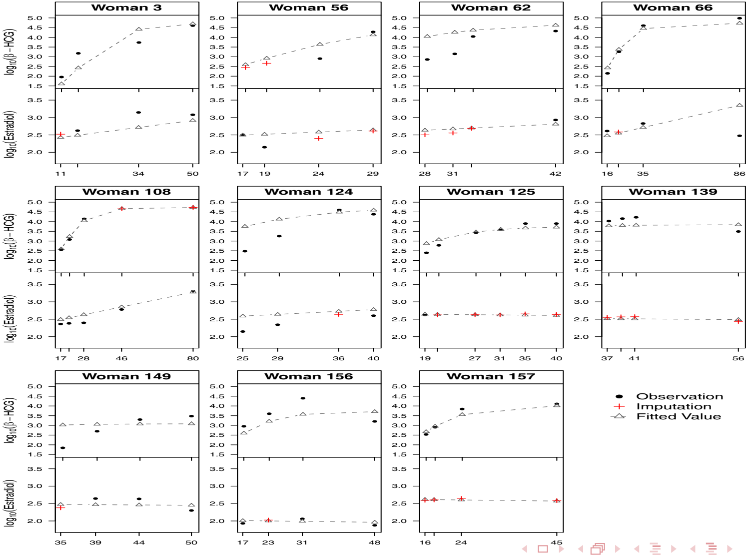
Curves	Model	C_i	No. of parameters	$-2\ell_{\max}$	AIC	BIC
Homogeneous	MNLMM	UNC	11	549.26	571.26	605.15
		AR(1)	12	534.06	558.06	595.04
		DEC	13	524.61	550.61	590.66
	MtNLMM	UNC	12	473.95	497.95	534.92
		AR(1)	13	457.58	483.58	523.64
		DEC	14	451.63	479.63	522.77
Heteroscedastic	MNLMM	(UNC, UNC)	22	368.26	412.26	480.05
		(UNC, AR(1))	23	366.10	412.10	482.97
		(UNC, DEC)	24	365.32	413.32	487.28
		(AR(1), UNC)	23	361.32	407.32	478.19
		(AR(1), AR(1))	24	359.16	407.16	481.11
		(AR(1), DEC)	25	358.38	408.38	485.42
		(DEC, UNC)	24	353.06	401.06	475.02
		(DEC, AR(1))*	25	350.90	400.90	477.94
		(DEC, DEC)	26	350.12	402.12	482.24
	MtNLMM	(UNC, UNC)	24	347.70	395.70	469.66
		(UNC, AR(1))	25	345.48	395.48	472.52
		(UNC, DEC)	26	344.71	396.71	476.82
		(AR(1), UNC)	25	342.11	392.11	**469.15**
		(AR(1), AR(1))	26	339.89	391.89	472.01
		(AR(1), DEC)	27	339.12	393.12	476.31
		(DEC, UNC)	26	335.75	387.75	467.87
		(DEC, AR(1))	27	333.53	**387.53**	470.73
		(DEC, DEC)	28	332.75	388.75	475.03

AIC $= 2m - 2\ell_{\max}$, and BIC $= m\log(N) - 2\ell_{\max}$, where m is the number of parameters.

Model Fitting for Pregnant Women Data

Parameter		Normal Group ($l = 1$) EST	Normal Group ($l = 1$) SE	Abnormal Group ($l = 2$) EST	Abnormal Group ($l = 2$) SE
$\boldsymbol{\beta}_l$	β_{1l}	4.7263	0.0392	3.6164	0.1822
	β_{2l}	15.6440	0.3181	11.9448	2.3528
	β_{3l}	6.9311	0.3808	5.9718	2.2860
	β_{4l}	2.2842	0.0507	2.4904	0.1329
	β_{5l}	0.0125	0.0013	-0.0015	0.0033
$\boldsymbol{D}_l$	d_{11l}	0.0001	0.0261	0.4875	0.1948
	d_{21l}	0.0001	0.0068	0.1779	0.0825
	d_{22l}	0.0006	0.0146	0.1269	0.0510
$\boldsymbol{\Sigma}_l$	σ_{11l}	0.0793	0.0221	0.2957	0.0919
	σ_{21l}	0.0106	0.0073	0.0444	0.0252
	σ_{22l}	0.0522	0.0158	0.0310	0.0121
$\boldsymbol{C}_{il}$	ϕ_l	0.5944	0.1500	0.7784	0.1155
	γ_l	0.4682	0.2096	1	–
	ν	6.6467	1.9069	165.3484	1475.3042

Imputed and Fitted Values under the MtNLMM



Summary and Future Research

- We have proposed a robust extension of the MNLMM by using the multivariate-t distributed random effects and within-subject errors.
- We create the pseudo data by using the Taylor approximation and then implement the ECM algorithm for carrying out ML estimation.
- Techniques for imputation of missing values and estimation of random effects are provided for ease of use.
- The methodology is motivated by, and applied to the data from a study of 161 pregnant women in Santiago, Chile.
-
-

Summary and Future Research

- We have proposed a robust extension of the MNLMM by using the multivariate-t distributed random effects and within-subject errors.
- We create the pseudo data by using the Taylor approximation and then implement the ECM algorithm for carrying out ML estimation.
- Techniques for imputation of missing values and estimation of random effects are provided for ease of use.
- The methodology is motivated by, and applied to the data from a study of 161 pregnant women in Santiago, Chile.
- Develop discriminant analysis under the MtNLMM.
- Apply a fully Bayesian approach to inferring the MtNLMM.

Summary and Future Research

- We have proposed a robust extension of the MNLMM by using the multivariate-t distributed random effects and within-subject errors.
- We create the pseudo data by using the Taylor approximation and then implement the ECM algorithm for carrying out ML estimation.
- Techniques for imputation of missing values and estimation of random effects are provided for ease of use.
- The methodology is motivated by, and applied to the data from a study of 161 pregnant women in Santiago, Chile.
- Develop discriminant analysis under the MtNLMM.
- Apply a fully Bayesian approach to inferring the MtNLMM.

Thanks For Your Attention!

Reference

- Akaike H. Information theory and an extension of the maximum likelihood principle. In 2nd Int. Symp. on Information Theory, (Edited by B. N. Petrov and F. Csaki). 1973; 267–281. Akademiai Kiado, Budapest.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B* 1977; **39:**1–38.
- Fitzmaurice GM, Laird NM, Ware J. *Applied Longitudinal Analysis.* Wiley: New York, 2004.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970; **57:** 97–109.
- Hedeker D, Gibbons RD. *Longitudinal Data Analysis.* John Wiley & Sons: New Jersey, 2006.
- Kotz S, Nadarajah S. Multivariate t Distributions and Their Applications. Cambridge University Press. 2004.
- Laird NM. Missing data in longitudinal studies. *Statistics in Medicine* 1988; **7:**305–315.
- Laird NM, Ware JH. Random effects models for longitudinal data. *Biometrics* 1982; **38:**963–974.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data (2nd ed.).* Wiley: New York, 2002.
- Little RJ, Rubin DB. *Statistical Analysis with Missing Data.* Wiley: New York, 1987.

- Marshall G, De la Cruz-Mesía R, Barón AE, Rutledge JH, Zerbe GO. Non-linear random effects model for multivariate responses with missing data. *Statistics in Medicine* 2006; **25:**2817–2830.
- Marshall G, De la Cruz-Mesía R, Quintana FA, Barón AE. Discriminant analysis for longitudinal data with multiple continuous responses and possibly missing data. *Biometrics* 2009; **65:**69–80.
- Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 1993; **80:**267–278.
- Muñoz A, Carey V, Schouten JP, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 1992; **48:**733–742.
- Pinheiro J, Bates D, DebRoy S, Sarkar D and the R Development Core Team (2012). R nlme package: Linear and Nonlinear Mixed Effects Models. R package version 3.1-104.
- R Development Core Team. *R: A Language and Environment for Statistical Computing, the R Foundation for Statistical Computing.* Vienna, Austria. 2011.
- Roy A. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biometrical Journal* 2006; **48:**286–301.
- Roy J, Lin X. Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. *Journal of the American Statistical Association* 2002; **97:**40–52.
- Rubin DB. Characterizing the estimation of parameters in incomplete-data problems. *Journal of the American Statistical Association* 1974; **69:**474–476.
- Rubin DB. Inference and missing data. *Biometrika* 1976; **63:**581–592.

- Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* 1997; **92:**775–779.
- Schafer JL. *Analysis of Incomplete Multivariate Data.* Chapman and Hall: London, 1997.
- Schluchter MD. Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Statistics in Medicine* 1988; **7:**317–324.
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; **6:** 461–464.
- Song X, Davidian M, Tsiatis AA. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* 2002; **3:**511–528.
- Wang WL. Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal* 2013; **55:**554–571
- Wang WL, Fan TH. ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Computational Statistics and Data Analysis* 2010; **54:**1328–1341.
- Wang WL, Fan TH. Estimation in multivariate t linear mixed models for multiple longitudinal data. *Statistica Sinica* 2011; **21:**1857–1880.
- Wang WL, Fan TH. Bayesian analysis of multivariate t linear mixed models using a combination of IBF and Gibbs samplers. *Journal of Multivariate Analysis* 2012; **105:**300–310.
- Yamashita T, Okamoto S, Thomas A, MacLachlan V, Healy DL. Predicting pregnancy outcome after in vitro fertilization and embryo transfer using estradiol, progesterone and human chorionic gonadotrophin β-subunit. *Fertility and Sterility* 1989; **51:**304–309.
- Zacks S. *The Theory of Statistical Inference.* Wiley: New York, 1971.

Issues to be investigated

1. If the random effects and errors indeed exhibit heavy tails, then how bad can the MNLMM behave?
2. When the random effects and errors are generated from the multivariate normal distribution, whether the MtNLMM is over-fitted?

Simulation Study

- Generate bivariate longitudinal data from the MtNLMM with the same mean profiles for the pregnant women data.
- We make the following assumption

$$(b_{i1}, b_{i2}, \boldsymbol{e}_{i1}^{\mathrm{T}}, \boldsymbol{e}_{i2}^{\mathrm{T}})^{\mathrm{T}} \sim t_{22}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \otimes \boldsymbol{I}_{10} \end{bmatrix}, \nu \right).$$

- The time t_k range from 10 to 100 changing by an increment of 10 units.
- The presumed parameters are given as

$$\boldsymbol{\beta} = (5, 17, 7, 2, 0.05)^{\mathrm{T}}, \quad \boldsymbol{D} = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.75 \\ 0.75 & 1 \end{bmatrix}.$$

- Degrees of freedom: $\nu = 5$ and $\nu = 50$
- Sample sizes: $N = 25$ and $N = 100$

Simulation results based on 100 replications under each combination of considered N and ν.

Parameter		$N = 25$				$N = 100$			
		$\nu = 5$		$\nu = 50$		$\nu = 5$		$\nu = 50$	
(True)		MNLMM	MtNLMM	MNLMM	MtNLMM	MNLMM	MtNLMM	MNLMM	MtNLMM
β_1 (5)	EST	5.012	4.994	5.002	5.004	4.987	4.991	5.007	5.007
	STD	0.273	0.220	0.206	0.207	0.158	0.131	0.112	0.111
	MSE	0.074	0.048	0.042	0.042	0.025	0.017	0.013	0.012
β_2 (17)	EST	17.068	17.073	16.881	16.886	16.997	17.021	17.000	17.007
	STD	0.898	0.798	0.796	0.804	0.600	0.419	0.359	0.364
	MSE	0.803	0.636	0.642	0.653	0.357	0.174	0.128	0.131
β_3 (7)	EST	6.732	6.842	6.863	6.869	6.934	6.938	6.960	6.962
	STD	0.797	0.622	0.568	0.554	0.398	0.303	0.293	0.293
	MSE	0.700	0.408	0.338	0.321	0.161	0.095	0.086	0.087
β_4 (2)	EST	2.034	1.993	2.010	2.014	1.997	1.990	2.015	2.011
	STD	0.320	0.231	0.227	0.233	0.170	0.121	0.119	0.115
	MSE	0.102	0.053	0.051	0.054	0.028	0.015	0.014	0.013
β_5 (0.05)	EST(10^{-2})	5.001	5.028	4.984	4.982	5.004	5.010	5.005	5.006
	STD(10^{-3})	2.377	1.792	2.013	2.035	1.425	1.072	0.948	0.948
	MSE(10^{-6})	5.594	3.255	4.035	4.131	2.012	1.147	0.982	0.893

Parameter (True)		N = 25, $\nu = 5$ MNLMM	N = 25, $\nu = 5$ MtNLMM	N = 25, $\nu = 50$ MNLMM	N = 25, $\nu = 50$ MtNLMM	N = 100, $\nu = 5$ MNLMM	N = 100, $\nu = 5$ MtNLMM	N = 100, $\nu = 50$ MNLMM	N = 100, $\nu = 50$ MtNLMM
d_{11} (1)	EST	1.705	1.013	0.959	0.932	1.698	1.006	1.035	0.996
	STD	0.874	0.383	0.313	0.308	0.550	0.174	0.163	0.158
	MSE	1.254	0.145	0.099	0.099	0.786	0.030	0.028	0.025
d_{21} (0.25)	EST	0.517	0.255	0.236	0.232	0.411	0.244	0.283	0.273
	STD	0.725	0.259	0.269	0.263	0.414	0.127	0.122	0.115
	MSE	0.592	0.066	0.072	0.069	0.195	0.016	0.016	0.014
d_{22} (1)	EST	1.792	1.050	0.986	0.953	1.655	1.003	1.010	0.969
	STD	0.946	0.353	0.320	0.308	0.478	0.179	0.145	0.135
	MSE	1.513	0.126	0.102	0.096	0.655	0.032	0.021	0.019
σ_{11} (1)	EST	1.699	1.027	1.039	0.997	1.680	1.012	1.034	0.992
	STD	0.595	0.191	0.099	0.094	0.226	0.082	0.047	0.045
	MSE	0.840	0.037	0.011	0.009	0.513	0.007	0.003	0.002
σ_{21} (0.75)	EST	1.274	0.769	0.783	0.752	1.268	0.761	0.773	0.741
	STD	0.440	0.143	0.095	0.091	0.178	0.064	0.039	0.037
	MSE	0.466	0.021	0.010	0.008	0.299	0.004	0.002	0.001
σ_{22} (1)	EST	1.686	1.011	1.047	1.008	1.682	1.015	1.037	0.995
	STD	0.581	0.165	0.112	0.108	0.227	0.083	0.051	0.049
	MSE	0.805	0.027	0.015	0.012	0.516	0.007	0.004	0.002
ν	EST	–	5.245	–	75.095	–	5.154	–	52.623
	STD	–	2.712	–	31.026	–	0.782	–	23.019
	MSE	–	8.167	–	1361.317	–	0.623	–	583.639
AIC	Mean	1601.52	1511.04	1380.13	1380.50	6409.85	6020.21	5482.89	5477.13
	Freq	0	100	71	29	0	100	13	87
BIC	Mean	1614.93	1525.66	1393.53	1395.12	6438.51	6051.47	5511.55	5508.39
	Freq	0	100	76	24	0	100	35	65

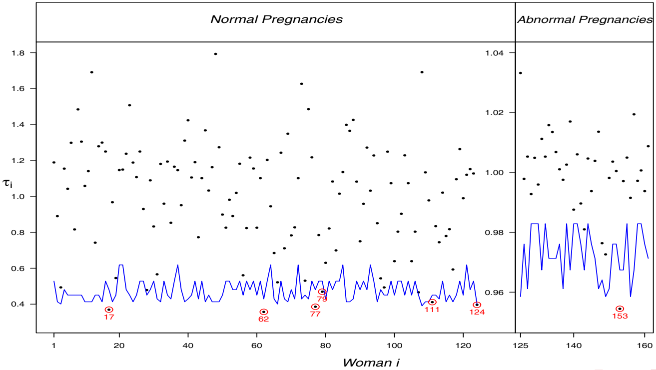
Outlier Detection for Pregnancy Women Data

ECM

Because $\hat{\tau}_i$ follows $(1+n_i^o/\nu)\mathcal{B}eta(\nu/2, n_i^o/2)$, under a significance level α, if

$$\hat{\tau}_i < (1+n_i^o/\nu)\mathcal{B}_\alpha(\nu/2, n_i^o/2)$$

then the corresponding subject would be identified as an outlier, where $\mathcal{B}_\alpha(\cdot,\cdot)$ denotes the α percentile of the Beta distribution such that $P(B \geq \mathcal{B}_\alpha) = 1-\alpha$.



Multivariate t distribution

Let $\boldsymbol{y} \sim t_d(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$, then the density of $\boldsymbol{y}$ is

$$f(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)|\boldsymbol{\Omega}|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{d/2}}\left(1 + \frac{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\nu}\right)^{-(\nu+d)/2}, \quad \boldsymbol{y} \in \mathcal{R}^d.$$

- If $\nu > 1$, $E(\boldsymbol{y}) = \boldsymbol{\mu}$.
- If $\nu > 2$, cov$(\boldsymbol{y}) = \nu(\nu-2)^{-1}\boldsymbol{\Omega}$.
- As $\nu \to \infty$, $\boldsymbol{y} \stackrel{\mathrm{D}}{\to} \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

Return

Example

- Take $\boldsymbol{y}_i = [1, 2, 3, 4]^{\mathrm{T}}$. Regard the elements 2 and 4 as the missing information of $\boldsymbol{y}_j$.
-
$$\boldsymbol{y}_i^{\mathrm{o}} = \boldsymbol{O}_i \boldsymbol{y}_i\,; \qquad \boldsymbol{y}_i^{\mathrm{m}} = \boldsymbol{M}_i \boldsymbol{y}_i$$
$$\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1\\2\\3\\4\end{bmatrix}; \qquad \begin{bmatrix}2\\4\end{bmatrix} = \begin{bmatrix}0 & 1 & 0 & 0\\0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\2\\3\\4\end{bmatrix}$$
-
$$\boldsymbol{y}_i = \boldsymbol{O}_i^{\mathrm{T}} \boldsymbol{y}_i^{\mathrm{o}} + \boldsymbol{M}_i^{\mathrm{T}} \boldsymbol{y}_i^{\mathrm{m}}$$
$$\begin{bmatrix}1\\2\\3\\4\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\\0 & 1\\0 & 0\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix} + \begin{bmatrix}0 & 0\\1 & 0\\0 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$$

Return

Identifiability under Missing Data

If $\boldsymbol{Y}$ has three variables $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$ with the following missing data pattern ('NA' represents missing values)

y_{11},	$\cdots$,	y_{1m_1},	NA,	$\cdots$,	NA,	NA,	$\cdots$,	NA
NA,	$\cdots$,	NA,	y_{2,m_1+1},	$\cdots$,	y_{2,m_1+m_2},	NA,	$\cdots$,	NA
NA,	$\cdots$,	NA,	NA,	$\cdots$,	NA,	y_{3,m_1+m_2+1}	$\cdots$,	$y_{3,n}$

1. The $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$ are never jointly observed.
2. The parameters in off-diagonal entries of covariance matrix $\boldsymbol{R} = \text{Cov}(\boldsymbol{Y})$ are inestimable.
3. Some model parameters are inestimable.

Return

Missing Data Mechanism (Rubin, 1976)

Let $\boldsymbol{y}$ be the full-data response vector (observed $\boldsymbol{y}^{\mathrm{o}}$ and missing $\boldsymbol{y}^{\mathrm{m}}$ parts), $\boldsymbol{r}$ be the missingness indicators, and $\boldsymbol{x}$ be covariates of interest.

Missing Completely at Random (MCAR)

$$P(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}) = P(\boldsymbol{r}|\boldsymbol{x}, \boldsymbol{\theta})$$

Missing at Random (MAR)

$$P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}}, \boldsymbol{y}^{\mathrm{m}}, \boldsymbol{x}, \boldsymbol{\theta}) = P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}}, \boldsymbol{x}, \boldsymbol{\theta})$$

Missing Not at Random (MNAR)

For some $\boldsymbol{y}^{\mathrm{m}} \neq \boldsymbol{y}'^{\mathrm{m}}$,

$$P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}}, \boldsymbol{y}^{\mathrm{m}}, \boldsymbol{x}, \boldsymbol{\theta}) \neq P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}}, \boldsymbol{y}'^{\mathrm{m}}, \boldsymbol{x}, \boldsymbol{\theta})$$

Return

Missing Data Mechanism (Rubin, 1976)

Let $\boldsymbol{y}$ be the full-data response vector (observed $\boldsymbol{y}^{\mathrm{o}}$ and missing $\boldsymbol{y}^{\mathrm{m}}$ parts), $\boldsymbol{r}$ be the missingness indicators, and $\boldsymbol{x}$ be covariates of interest.

Missing Completely at Random (MCAR)

$$P(\boldsymbol{r}|\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta}) = P(\boldsymbol{r}|\boldsymbol{x},\boldsymbol{\theta})$$

Missing at Random (MAR)

$$P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{y}^{\mathrm{m}},\boldsymbol{x},\boldsymbol{\theta}) = P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{x},\boldsymbol{\theta})$$

Missing Not at Random (MNAR)

For some $\boldsymbol{y}^{\mathrm{m}} \neq \boldsymbol{y}'^{\mathrm{m}}$,

$$P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{y}^{\mathrm{m}},\boldsymbol{x},\boldsymbol{\theta}) \neq P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{y}'^{\mathrm{m}},\boldsymbol{x},\boldsymbol{\theta})$$

Return

Missing Data Mechanism (Rubin, 1976)

Let $\boldsymbol{y}$ be the full-data response vector (observed $\boldsymbol{y}^{\mathrm{o}}$ and missing $\boldsymbol{y}^{\mathrm{m}}$ parts), $\boldsymbol{r}$ be the missingness indicators, and $\boldsymbol{x}$ be covariates of interest.

Missing Completely at Random (MCAR)

$$P(\boldsymbol{r}|\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta}) = P(\boldsymbol{r}|\boldsymbol{x},\boldsymbol{\theta})$$

Missing at Random (MAR)

$$P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{y}^{\mathrm{m}},\boldsymbol{x},\boldsymbol{\theta}) = P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{x},\boldsymbol{\theta})$$

Missing Not at Random (MNAR)

For some $\boldsymbol{y}^{\mathrm{m}} \neq \boldsymbol{y}'^{\mathrm{m}}$,

$$P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{y}^{\mathrm{m}},\boldsymbol{x},\boldsymbol{\theta}) \neq P(\boldsymbol{r}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{y}'^{\mathrm{m}},\boldsymbol{x},\boldsymbol{\theta})$$

Return

Kronecker Product $\otimes$

Return

If A is an $m \times n$ matrix and B is a $p \times q$ matrix then the kronecker product

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \dots & \dots & \vdots & \dots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

is a $mp \times nq$ matrix.

$$\begin{aligned} \boldsymbol{R}_i &= \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}_{(2\times 2)} \otimes \boldsymbol{C}_{i(s_i \times s_i)} \\ &= \begin{bmatrix} \sigma_{11}\boldsymbol{C}_i & \sigma_{12}\boldsymbol{C}_i \\ \sigma_{21}\boldsymbol{C}_i & \sigma_{22}\boldsymbol{C}_i \end{bmatrix}_{(2s_i \times 2s_i)} \end{aligned}$$

GLS-Scoring Iterative Procedure

Return

1. Set an initial guess of $\{\tau_i\}_{i=1}^N$ as

$$\hat{\tau}_i^{(h)} = \arg\min_{\tau_i} \Big\{ \tau_i \Delta_{\tilde{\boldsymbol{y}}_i} - \nu(\log \tau_i - \tau_i) - n_i \log(\tau_i) \Big\}, \quad i = 1, \ldots, N.$$

2. Perform a generalized least squares step:

$$\hat{\boldsymbol{\beta}}^{(h+1)} = \Big(\sum_{i=1}^N \hat{\tau}_i^{(h)} \tilde{\boldsymbol{X}}_i^{\mathrm{T}} \tilde{\boldsymbol{\Lambda}}_i^{(h)} \tilde{\boldsymbol{X}}_i^{\mathrm{T}} \Big)^{-1} \sum_{i=1}^N \hat{\tau}_i^{(h)} \tilde{\boldsymbol{X}}_i^{\mathrm{T}} \tilde{\boldsymbol{\Lambda}}_i^{(h)^{-1}} \tilde{\boldsymbol{y}}_i,$$

where $\tilde{\boldsymbol{\Lambda}}_i^{(h)}$ is $\tilde{\boldsymbol{\Lambda}}_i$ evaluated at the h iteration.

3. Let $\boldsymbol{\alpha} = \big(\text{vech}(\boldsymbol{D}), \text{vech}(\boldsymbol{\Sigma}), \phi, \gamma, \nu\big)$, and then update $\hat{\boldsymbol{\alpha}}^{(h)}$ by one iteration of scoring procedure:

$$\hat{\boldsymbol{\alpha}}^{(h+1)} = \hat{\boldsymbol{\alpha}}^{(h)} + \hat{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(h+1/2)-1} \hat{\mathbf{s}}_{\boldsymbol{\alpha}}^{(h+1/2)},$$

where $\hat{\mathbf{s}}_{\boldsymbol{\alpha}}^{(h+1/2)}$ and $\hat{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(h+1/2)}$ are score vector and Fisher information matrix of $\boldsymbol{\alpha}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(h+1)}$ and $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}^{(h)}$.