Supplementary Materials for Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances

Y. Wang
yanpin@ufl.edu
Department of Biostatistics
University of Florida

M. J. Daniels
mdaniels@stat.ufl.edu
Department of Statistics
University of Florida

1 Supplementary Materials

1.1 Derivations from Section 3.3

$$\begin{aligned}
E(\gamma) &= \begin{pmatrix} w^\star & w^\perp \end{pmatrix}^{-1} E(z(\pi)) \\
&= \begin{pmatrix} ((w^\star)^T w^\star)^{-1}(w^\star)^T \\ (w^\perp)^T \end{pmatrix} \mu 1_{T\times 1} \\
&= \mu \begin{pmatrix} ((w^\star)^T w^\star)^{-1}(w^\star)^T \\ (w^\perp)^T \end{pmatrix} 1_{T\times 1}
\end{aligned}$$

$$\begin{aligned}
Var(\gamma) &= \begin{pmatrix} w^\star & w^\perp \end{pmatrix}^{-1} Var(z(\pi))(\begin{pmatrix} w^\star & w^\perp \end{pmatrix}^{-1})^T \\
&= \begin{pmatrix} ((w^\star)^T w^\star)^{-1}(w^\star)^T \\ (w^\perp)^T \end{pmatrix} \sigma^2 I_{T\times T} \begin{pmatrix} ((w^\star)^T w^\star)^{-1}(w^\star)^T)^T & w^\perp \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} ((w^\star)^T w^\star)^{-1}(w^\star)^T \\ (w^\perp)^T \end{pmatrix} \begin{pmatrix} ((w^\star)^T w^\star)^{-1}(w^\star)^T)^T & w^\perp \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} ((w^\star)^T w^\star)^{-1} & 0 \\ 0 & I_{(T-q)\times(T-q)} \end{pmatrix}.
\end{aligned}$$

$$\begin{aligned}
var(z_{i(\gamma)}) &= var(w_{(i,:)}\gamma) \\
&= var(w^\star_{(i,:)}\gamma^\star) + var(w^\perp_{(i,:)}\gamma^\perp) \\
&= \sigma^2(w^\star_{(i,:)}((w^\star)^T w^\star)^{-1}(w^\star_{(i,:)})^T + w^\perp_{(i,:)}(w^\perp_{(i,:)})^T).
\end{aligned}$$

$$
\begin{aligned}
\overline{var(\mathbf{z})} &= \frac{1}{T}\sum_{i=1}^{T} var(z_i) \\
&= \sigma^2 \frac{1}{T} tr\{w^\star((w^\star)^T w^\star)^{-1}(w^\star)^T\} \\
&= \sigma^2 \frac{1}{T} tr\{((w^\star)^T w^\star)^{-1}(w^\star)^T w^\star\} \\
&= \sigma^2 \frac{1}{T} tr\{I_{q\times q}\} \\
&= \frac{q}{T}\sigma^2,
\end{aligned}
$$

1.2 Sampling algorithm

We sample (β, η, γ) using a (block) Gibbs sampler along with data augmentation for the missing responses (straightforward given the full data response is multivariate normal).

We sample the full conditional for the parameters as follows. At iteration k0,

1. Sample $Y_{imiss}|(\beta^{(k0-1)}, \gamma^{(k0-1)}, \eta^{(k0-1)}, y_{iobs})$ by data augmentation.

   $Y_{imiss} \sim N(x_{imiss}^T\beta^{(k0-1)} + \Sigma_{i12}^{(k0-1)}(\Sigma_{iobs}^{(k0-1)})^{-1}(y_{iobs} - x_{iobs}^T\beta^{(k0-1)}), \Sigma_{imiss}^{(k0-1)} - \Sigma_{i21}^{(k0-1)}(\Sigma_{iobs}^{(k0-1)})^{-1}\Sigma_{i12}^{(k0-1)})$

   where $x_i^T = (x_{iobs}^T, x_{imiss}^T)^T$, and $\Sigma_i^{(k0-1)} = \begin{pmatrix} \Sigma_{iobs}^{(k0-1)} & \Sigma_{i12}^{(k0-1)} \\ \Sigma_{i21}^{(k0-1)} & \Sigma_{imiss}^{(k0-1)} \end{pmatrix}$.

2. Sample $\beta|(\eta^{(k0-1)}, \gamma^{(k0-1)}, y^{(k0)})$ from a normal distribution with mean

   $$\mu^{(k_0-1)} = \Sigma_0^{(k0-1)} \sum_{i=1}^{n} x_i D^{-1}_{i(\eta^{(k0-1)})} R^{-1}_{i(\gamma^{(k0-1)})} D^{-1}_{i(\eta^{(k0-1)})} y_i^{(k0)}$$

   and variance,

   $$\Sigma_0^{(k_0-1)} = (\sum_{i=1}^{n} x_i D^{-1}_{i(\eta^{(k0-1)})} R^{-1}_{i(\gamma^{(k0-1)})} D^{-1}_{i(\eta^{(k0-1)})} x_i^T)^{-1}.$$

3. Sample $\eta|(\beta^{(k0)}, \gamma^{(k0-1)}, y^{(k0)})$ using a random walk Metropolis-Hastings algorithm. The full conditional is proportional to

$$
\begin{aligned}
\pi(\eta|\beta,\gamma,Y,X) \quad \propto \quad & \{(\prod_{i=1}^{n}|D_{i(\eta)}|^{-1})|\sum_{i=1}^{n} x_i D^{-1}_{i(\eta)} R^{-1}_{i(\gamma^{(k0-1)})} D^{-1}_{i(\eta)} x_i^T|^{-\frac{1}{2}}\} \\
& \exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i^{(k0)} - x_i^T\beta^{(k0)})^T D^{-1}_{i(\eta)} R^{-1}_{i(\gamma^{(k0-1)})} D^{-1}_{i(\eta)}(y_i^{(k0)} - x_i^T\beta^{(k0)})\} \\
& \exp(-\frac{1}{2}\eta^T\Sigma_\eta\eta).
\end{aligned}
$$

4. Sample $\gamma|(\beta^{(k0)}, \eta^{(k0)}, y^{(k0)})$ using a Quasi-Newton Metropolis-Hasting Algorithm (details in the web appendix). The full conditional is proportional to

   $\pi(\gamma|\beta, \eta, Y, X)$

   $\propto (\prod_{i=1}^{n} |R_{i(\gamma)}|^{-\frac{1}{2}})| \sum_{i=1}^{n} x_i D_{i(\eta^{(k0)})}^{-1} R_{i(\gamma)}^{-1} D_{i(\eta^{(k0)})}^{-1} x_i^T|^{-\frac{1}{2}}$

   $exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i^{(k0)} - x_i^T\beta^{(k0)})^T D_{i(\eta^{(k0)})}^{-1} R_{i(\gamma)}^{-1} D_{i(\eta^{(k0)})}^{-1}(y_i^{(k0)} - x_i^T\beta^{(k0)})\} exp\{-\frac{1}{2}(\gamma - \mu_\gamma)^T \Sigma_\gamma^{-1}(\gamma - \mu_\gamma)\}$.

1.3 Simulating from the full Conditional for γ

For convenience, we denote log full conditional $\pi(\gamma|\beta, \eta, Y, X)$ by $\ell^\star(\gamma)$. Let γ denote the mode of $\ell^\star(\gamma)$ and assume γ is in a neighborhood of γ_0. Then,

$$\frac{\partial \ell^\star(\gamma)}{\partial \gamma} = \frac{\partial \ell^\star(\gamma)}{\partial \gamma}|_{\gamma=\gamma_0} + \frac{\partial^2 \ell^\star(\gamma)}{\partial \gamma \partial \gamma^T}|_{\gamma=\gamma_0}(\gamma - \gamma_0) + o(|\gamma - \gamma_0|^2)\mathbf{1_{q\times 1}}.$$

Therefore,

$$\gamma \approx \gamma_0 - (\frac{\partial^2 \ell^\star(\gamma)}{\partial \gamma \partial \gamma^T}|_{\gamma=\gamma_0})^{-1} \frac{\partial \ell^\star(\gamma)}{\partial \gamma}|_{\gamma=\gamma_0}.$$

We can approximate the observed information matrix $-\frac{\partial^2 \ell^\star(\gamma)}{\partial \gamma \partial \gamma^T}|_{\gamma=\gamma_0}$ by the expected fisher information matrix at γ_0. Using a quasi-Newton method to update γ, $\gamma = \gamma_0 + \alpha I(\gamma_0)^{-1} \bigtriangledown \frac{\partial \ell^\star(\gamma)}{\partial \gamma}|_{\gamma=\gamma_0}$, we choose α to satisfy Wolf's condition (Wolf, 1969) . The form of the expected information, $I(\gamma)$ is given in the following section.

At iteration k_0, for Step 4 in our algorithm, we first approximate the model of $l^\star(\cdot)$, $(\gamma_*^{(k_0)})$ by $\gamma_*^{(k_0)} = \gamma^{(k_0-1)} + \alpha I(\gamma^{(k_0-1)})^{-1} \bigtriangledown \frac{\partial \ell^\star(\gamma)}{\partial \gamma}|_{\gamma=\gamma^{(k_0-1)}}$. Then we sample

$$\gamma^{*(k_0)} \sim N(\gamma_*^{(k_0)}, I^{-1}(\gamma_*^{(k_0)})), \tag{1}$$

and accept $\gamma^{(k_0)} = \gamma^{*(k_0)}$ with probability

$$p = min\{1, \frac{\pi(\gamma^{*(k_0)})|\beta^{(k_0)}, \eta^{(k_0)}, Y^{(k_0)}, X)}{\pi(\gamma^{*(k_0-1)})|\beta^{(k_0)}, \eta^{(k_0)}, Y^{(k_0)}, X)} \times \frac{h(\gamma^{*(k_0-1)})|\beta^{(k_0)}, \eta^{(k_0)}, Y^{(k_0)}, X)}{h(\gamma^{*(k_0)})|\beta^{(k_0)}, \eta^{(k_0)}, Y^{(k_0)}, X)}\}, \tag{2}$$

where $h(\bullet)$ is the pdf of (1).

1.4 Deriving the Expected Information Matrix for γ

We derive the expected information matrix for γ (and for (γ, σ_0)), where $\sigma_0 = (\sigma_{11}, ..., \sigma_{pp})^T$. To do this, we first define some needed quantities. Let A be a $n \times n$ symmetric matrix with elements

$\{a_{ij}\}$, B be a $m \times n$ matrix with elements $\{b_{ij}\}$, and C be a $s \times s$ matrix with elements $\{c_{ij}\}$ matrices. We define

$$vec(A) = \begin{pmatrix} a_{11} & a_{21} & \cdot & \cdot & \cdot & a_{n1} & a_{12} & \cdot & \cdot & \cdot & a_{n2} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix}^T$$

to be a vector including all the elements in the matrix sorted by column. We define

$$v(A) = \begin{pmatrix} a_{11} & a_{21} & \cdot & \cdot & \cdot & a_{n1} & a_{22} & \cdot & \cdot & \cdot & a_{n2} & \cdot & \cdot & \cdot & \alpha_{nn} \end{pmatrix}^T$$

to be a vector including all the elements in the lower triangular part of a square matrix (sorted by column). Finally, we define

$$vh(A) = \begin{pmatrix} a_{21} & a_{31} & \cdot & \cdot & \cdot & a_{n1} & a_{32} & \cdot & \cdot & \cdot & a_{n2} & \cdot & \cdot & \cdot & \alpha_{nn-1} \end{pmatrix}^T,$$

to be a vector with all the elements in the lower triangular part of a square matrix (without the main diagonal). Define the Kronecker product of two square matrices as $B \bigotimes C = (b_{ij}C)$. Now, define a matrix D_n such that $D_n v(A) = vec(A)$. So, $v(A) = D_n^{\dagger} vec(A)$, where $D_n^{\dagger}$ is the Moore-Penrose inverse of D_n, and

$$D_n^{\dagger} = (D_n' D_n)^{-1} D_n'.$$

Let $Y_1, Y_2, ..., Y_n$ be independently and identically distributed $p \times 1$ random vector such that

$$Y_i \sim N_p(\mu, \Sigma)$$

where $i = 1, 2, ..., n$, $\Sigma = (\sigma_{jk})$ is positive definite, and let $n \geq p + 1$. The expected information matrix for $v(\Sigma)$ is

$$F_n = \frac{n}{2} D_p' (\Sigma^{-1} \otimes \Sigma^{-1}) D_p$$

(Magnus and Neudecker, 1984).
To derive the expected Fisher information of (γ, σ_0), we specify the following series of transformations,

$$\sigma \to \sigma^* \to \rho \to \rho^* \to \pi \to z \to (\gamma, \gamma^\star, \sigma_0)$$

The transformations are defined as $\sigma = (v(\Sigma))$, $\sigma^* = (vh(\Sigma)^T, \sigma_0^T)^T$, $\rho = (vh(R)^T, \sigma_0^T)^T$, $\rho* = (\rho_{12}, \rho_{23}, \rho_{13}, ..., \rho_{p-1p}, ..., \rho_{1p}, \sigma_0^T)^T$, $\pi = (\pi_{12}, \pi_{23}, \pi_{13}, ..., \pi_{p-1p}, ..., \pi_{1p}, \sigma_0^T)^T$, and $z = (z_{12}, z_{23}, z_{13}, ..., z_{p-1p}, ..., z_{1p}, \sigma_0)^T$.
Details on the Jacobian of each transformation follow.

1. $g_1(\sigma; \sigma^*) : \sigma \to \sigma^*$ separates the diagonal and off-diagonal elements. The Jacobin matrix J^{σ^*} of $g_1(\sigma; \sigma^*)$ is obtained by re-ordering the $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ dimensional identity matrix corresponding to this re-ordering transformation.

2. $g_2(\sigma^*; \rho) : \sigma^* \rightarrow \rho$ is a 1-1 transformation from the covariance parameters to variance/correlation parameters. The Jacobian, J^ρ is

$$J^\rho = \begin{pmatrix} J^\rho_{11} & J^\rho_{12} \\ \underline{0} & I_{p\times p} \end{pmatrix}$$

where $J^\rho_{11} = diag(vh(\sigma_0 \bigotimes \sigma_0^T)^{\frac{1}{2}})$ and

$$J^\rho_{12} = \begin{pmatrix} \frac{\rho_{12}\sqrt{\sigma_{11}\sigma_{22}}}{2\sigma_{11}} & \frac{\rho_{12}\sqrt{\sigma_{11}\sigma_{22}}}{2\sigma_{22}} & 0 & \dots & 0 & 0 \\ \frac{\rho_{13}\sqrt{\sigma_{11}\sigma_{33}}}{2\sigma_{11}} & 0 & \frac{\rho_{13}\sqrt{\sigma_{11}\sigma_{33}}}{2\sigma_{33}} & \dots & 0 & 0 \\ \dots\dots & & & & & \\ \frac{\rho_{1p}\sqrt{\sigma_{11}\sigma_{pp}}}{2\sigma_{11}} & 0 & 0 & \dots & 0 & \frac{\rho_{1p}\sqrt{\sigma_{11}\sigma_{pp}}}{2\sigma_{pp}} \\ 0 & \frac{\rho_{23}\sqrt{\sigma_{22}\sigma_{33}}}{2\sigma_{22}} & \frac{\rho_{23}\sqrt{\sigma_{22}\sigma_{33}}}{2\sigma_{33}} & \dots & 0 & 0 \\ \dots\dots & & & & & \\ 0 & \frac{\rho_{2p}\sqrt{\sigma_{22}\sigma_{pp}}}{2\sigma_{22}} & 0 & \dots & 0 & \frac{\rho_{2p}\sqrt{\sigma_{22}\sigma_{pp}}}{2\sigma_{pp}} \\ \dots\dots & & & & & \\ 0 & 0 & 0 & \dots & \frac{\rho_{p-1p}\sqrt{\sigma_{p-1p-1}\sigma_{pp}}}{2\sigma_{p-1p-1}} & \frac{\rho_{p-1p}\sqrt{\sigma_{p-1p-1}\sigma_{pp}}}{2\sigma_{pp}} \end{pmatrix}$$

3. $g_3(\rho; \rho^*) : \rho \rightarrow \rho^*$ is a 1-1 transformation which changes the order of the parameters ρ to ρ^* with Jacobian,

$$J^{\rho *} = \begin{pmatrix} J^{\rho *}_{11} & \underline{0} \\ \underline{0} & I_{p\times p} \end{pmatrix},$$

where $J^{\rho *}_{11}$ is a matrix obtained by reordering identity matrix $I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}$ corresponding to the reordering from $(\rho_{12}, ..., \rho_{1p}, \rho_{23}, ..., \rho_{2p}, ..., \rho_{p-1p})$, to $(\rho_{12}, \rho_{23}, \rho_{13}, ..., \rho_{p-1p}, ..., \rho_{1p})$.

4. $g_4(\rho*; \pi) : \rho* \rightarrow \pi$ is a 1-1 transformation defined in (1) with σ_{jj} unchanged. The Jacobian is

$$J^\pi = \begin{pmatrix} J^\pi_{11} & \underline{0} \\ \underline{0} & I_{p\times p} \end{pmatrix}$$

where J^π_{11} is Jacobian matrix of transformation from $(\rho_{12}, \rho_{23}, \rho_{13}, ..., \rho_{p-1p}, ..., \rho_{1p})$ to $(\pi_{12}, \pi_{23}, \pi_{13}, ..., \pi_{p-1p}, ..., \pi_{1p})$, which is a lower triangular matrix with elements $\left\{\frac{\partial\rho_{jk}}{\partial\pi_{lm}}\right\}$ in position $(\frac{(j-1)(2p-j)}{2} + k, \frac{(l-1)(2p-l)}{2} + m)$.

5. $g_5(\pi; z) : \pi \rightarrow z$ is a 1-1 transformation which transforms the π to $z(\pi)$. The Jacobian is

$$J^z = \begin{pmatrix} J^z_{11} & \underline{0} \\ \underline{0} & I_{p\times p} \end{pmatrix}$$

where

$$J^z_{11} = \begin{pmatrix} 1-\pi^2_{12} & 0 & 0 & \dots & 0 \\ 0 & 1-\pi^2_{23} & 0 & \dots & 0 \\ 0 & 0 & 1-\pi^2_{13} & \dots & 0 \\ \dots\dots & & & & \\ 0 & 0 & 0 & \dots & 1-\pi^2_{1p} \end{pmatrix}.$$

6. $g_6(z;(\gamma,\gamma^\perp,\sigma_0)) : z \to (\gamma,\gamma^\perp,\sigma_0)$ is defined in (10) with σ_{jj} unchanged, and the Jacobian

$$J^{(\gamma,\gamma^\perp,\sigma_0)} = \begin{pmatrix} J_{11}^{(\gamma,\gamma^\perp)} & 0 \\ 0 & I_{p\times p} \end{pmatrix},$$

where

$$J_{11}^{(\gamma,\gamma^\perp)} = \begin{pmatrix} w^\star & w^\perp \end{pmatrix}.$$

Now, let $I_{(\sigma^*)}$ denote the expected information of σ^*. Since the information matrix is invariant under transformation,

$$I_{(\sigma*)} = J^{\sigma^{\star T}}\{\frac{n}{2}[D_T^T(\Sigma^{-1}\bigotimes\Sigma^{-1})D_T]\}J^{\sigma^\star} = \begin{pmatrix} I_{11}^{\sigma*} & I_{12}^{\sigma*} \\ I_{21}^{\sigma*} & I_{22}^{\sigma*} \end{pmatrix}. \tag{3}$$

Then,

$$\begin{aligned} I_{(\rho)} &= J^{\rho T} I_{(\sigma*)} J^{\rho} \\ &= \begin{pmatrix} J_{11}^{\rho} & J_{12}^{\rho} \\ \underline{0} & I_{p\times p} \end{pmatrix}^T \begin{pmatrix} I_{11}^{\sigma*} & I_{12}^{\sigma*} \\ I_{21}^{\sigma*} & I_{22}^{\sigma*} \end{pmatrix} \begin{pmatrix} J_{11}^{\rho} & J_{12}^{\rho} \\ \underline{0} & I_{p\times p} \end{pmatrix} \\ &= \begin{pmatrix} J_{11}^{\rho T} I_{11}^{\sigma*} J_{11}^{\rho} & J_{11}^{\rho T} I_{11}^{\sigma*} J_{12}^{\rho} + J_{11}^{\rho} I_{12}^{\sigma*} \\ J_{12}^{\rho T} I_{11}^{\sigma*} J_{11}^{\rho} + I_{21}^{\sigma*} J_{11}^{\rho} & J_{12}^{\rho T} I_{11}^{\sigma*} J_{12}^{\rho} + (I_{21}^{\sigma*} J_{12}^{\rho} + J_{12}^{\rho T} I_{12}^{\sigma*}) + I_{22}^{\sigma*} \end{pmatrix} \\ &= \begin{pmatrix} I_{(\rho_0)} & I_{(\rho_0\sigma_0)} \\ I_{(\rho_0\sigma_0)}^T & I_{(\sigma_0)} \end{pmatrix}. \end{aligned}$$

where $\rho_0 = (\rho_{12}, ..., \rho_{1p}, \rho_{23}, ...\rho_{p-1p})$.

It then follows that
$I^{(\gamma,\gamma^\perp,\sigma_0)} = (J^{(\gamma,\gamma^\perp,\sigma_0)T} J^{zT} J^{\pi T} J^{\rho*T}) I_{(\rho)} (J^{\rho*} J^{\pi} J^{z} J^{(\gamma,\gamma^\perp,\sigma_0)})$

$$= \begin{pmatrix} (J_{11}^{(\gamma,\gamma^\perp)T} J_{11}^{zT} J_{11}^{\pi T} J_{11}^{\rho*T}) I_{(\rho_0)} (J_{11}^{\rho*} J_{11}^{\pi} J_{11}^{z} J_{11}^{(\gamma,\gamma^\perp)}) & J_{11}^{(\gamma,\gamma^\perp)T} J_{11}^{zT} J_{11}^{\pi T} J_{11}^{\rho*T} I_{(\rho_0\sigma_0)} \\ (J_{11}^{(\gamma,\gamma^\perp)T} J_{11}^{zT} J_{11}^{\pi T} J_{11}^{\rho*T} I_{(\rho_0\sigma_0)})^T & I_{(\sigma_0)} \end{pmatrix}. \tag{4}$$

So the three blocks of the information matrix for (γ, σ_0) are

$$I(\gamma) = (w^{\star T} J_{11}^{zT} J_{11}^{\pi T} J_{11}^{\rho*T}) I_{(\rho_0)} (J_{11}^{\rho*} J_{11}^{\pi} J_{11}^{z} w^\star) \tag{5}$$

$$I(\gamma,\sigma_0) = w^{\star T} J_{11}^{zT} J_{11}^{\pi T} J_{11}^{\rho*T} I_{(\rho_0\sigma_0)} \tag{6}$$

$$I(\sigma_0) = J_{12}^{\rho T} I_{11}^{\sigma*} J_{12}^{\rho} + (I_{21}^{\sigma*} J_{12}^{\rho} + J_{12}^{\rho T} I_{12}^{\sigma*}) + I_{22}^{\sigma*}. \tag{7}$$

1.5 Proof of Theorem 1

Proof:

Let $Y_i^{k_i} = \{Y_{ij}, j = 1, \ldots, k_i$ where $k_i : Q_{ik_i} = 1, Q_{i,k_i+1} = 0\}$, $S_k = \{i, Q_{ik_i} = 1$ and $Q_{ik_i+1} = 0$, and $k_i = k$, where $1 \leq k \leq p-1; i = 1, \ldots, n\}$ and $\Im_\beta, \Im_\gamma, \Im_\eta$ be sample spaces of β, γ, η, respectively. Thus the observed data distribution of ith subject is $Y_i^{k_i} \sim N_k(x_i^{k_i T}\beta, \Sigma_i^{k_i})$, where $x_i^{k_i} = x_{i[:,1:k_i]}$ is a $p_\beta \times k_i$ submatrix of x_i and $\Sigma_i^{k_i} = \Sigma_{i[1:k_i,1:k_i]}$ is a $k_i \times k_i$ principal submatrix of Σ_i. Define the observed data, $Y_{obs} = (Y_1^{k_1}, \ldots, Y_n^{k_n})$.

Therefore,

$$f(y_i^{k_i}|\beta, \gamma, \eta, x_i^{k_i}) \propto |\Sigma_i^{k_i}|^{-\frac{1}{2}} \exp[-\frac{1}{2}(y_i^{k_i} - x_i^{k_i T}\beta)^T(\Sigma_i^{k_i})^{-1}(y_i^{k_i} - x_i^{k_i T}\beta)],$$

and

$$\begin{aligned} & f(y_{obs}|\gamma, \beta, \eta, x) \\ & \propto \prod_{i=1}^{n} |\Sigma_i^{k_i}|^{-\frac{1}{2}} exp\{-\frac{1}{2}[\sum_{i=1}^{n}(y_i^{k_i} - x_i^{k_i T}\beta)^T(\Sigma_i^{k_i})^{-1}(y_i^{k_i} - x_i^{k_i T}\beta)\} \\ & = \prod_{k=1}^{p-1}\prod_{i\in S_k} |\Sigma_i^{k_i}|^{-\frac{1}{2}} exp\{-\frac{1}{2}\sum_{k=1}^{p-1}[\sum_{i\in S_k}(y_{ik} - x_i^{k_i T}\beta)^T(\Sigma_i^{k_i})^{-1}(y_i^{k_i} - x_i^{k_i T}\beta)]\} \\ & = \prod_{k=1}^{p-1}\prod_{i\in S_k} |\Sigma_i^{k_i}|^{-\frac{1}{2}} exp\{-\frac{1}{2}[\sum_{k=1}^{p-1}\sum_{i\in S_k}(y_i^{k_i T}(\Sigma_i^{k_i})^{-1}y_i^{k_i}) \\ & \quad -2(\sum_{k=1}^{p-1}\sum_{i\in S_k} y_i^{k_i T}(\Sigma_i^{k_i})^{-1}x_i^{k_i T})\beta + \beta^T(\sum_{k=1}^{p-1}\sum_{i\in S_k} x_i^{k_i}(\Sigma_i^{k_i})^{-1}x_i^{k_i T})\beta]\} \end{aligned}$$

Therefore, posterior distribution of (β, γ, η) is

$$\pi(\beta, \gamma, \eta|y_{obs}, x) = \frac{m(\beta, \gamma, \eta|y_{obs}, x)}{\int_{\beta\in\Im_\beta}\int_{\gamma\in\Im_\gamma}\int_{\eta\in\Im_\eta} m(\beta, \gamma, \eta|y_{obs}, x)d\beta d\gamma d\eta},$$

where

$$\begin{aligned} & m(\beta, \gamma, \eta|y_{obs}, x) \\ & = \prod_{k=1}^{p-1}\prod_{i\in S_k} |(\Sigma_i^{k_i})|^{-\frac{1}{2}} \\ & \quad \exp\{-\frac{1}{2}[\sum_{k=1}^{p-1}\sum_{i\in S_k}(y_i^{k_i T}(\Sigma_i^{k_i})^{-1}y_i^{k_i}) - 2(\sum_{k=1}^{p-1}\sum_{i\in S_k} y_i^{k_i T}(\Sigma_i^{k_i})^{-1}x_i^{k_i T})\beta \\ & \quad +\beta^T(\sum_{k=1}^{p-1}\sum_{i\in S_k} x_i^{k_i}(\Sigma_i^{k_i})^{-1}x_i^{k_i T})\beta]\}\pi_{(\gamma)}\pi_{(\eta)}. \end{aligned}$$

Since $\sum x_i^{k_i} x_i^{k_i T}$ is full rank, $\sum_{k=1}^{p-1} \sum_{i \in S_k} x_i^{k_i} \Sigma_{ik}^{-1} x_i^{k_i T}$ is a positive-definite matrix. Therefore, its smallest eigenvalue λ_p is larger than zero and $|\sum_{k=1}^{p-1} \sum_{i \in S_k} x_i^{k_i} (\Sigma_i^{k_i})^{-1} x_i^{k_i T}|^{-\frac{1}{2}} < \lambda_p^{-\frac{p}{2}}$.
Define $\hat{\beta} = (\sum_{k=1}^{p-1} \sum_{i \in S_k} x_i^{k_i} (\Sigma_i^{k_i})^{-1} x_i^{k_i T})^{-1} \sum_{k=1}^{p-1} \sum_{i \in S_k} x_i^{k_i} \Sigma_k^{-1} y_i^{k_i}$. We obtain,

$$\int_{\Im_\beta} m(\beta, \gamma, \sigma | y_{obs}, x) d\beta$$

$$\propto |\sum_{k=1}^{p-1} \sum_{i \in S_k} x_i^{k_i} (\Sigma_i^{k_i})^{-1} x_i^{k_i T}|^{-\frac{1}{2}} (\prod_{k=1}^{p-1} \prod_{i \in S_k} |(\Sigma_i^{k_i})|^{-\frac{1}{2}})$$

$$\exp\{-\frac{1}{2} \sum_{k=1}^{p-1} \sum_{i \in S_k} (y_i^{k_i} - x_i^{k_i T} \hat{\beta})^T (\Sigma_i^{k_i})^{-1} (y_i^{k_i} - x_i^{k_i T} \hat{\beta})\} \pi_{(\gamma)} \pi_{(\eta)}$$

$$\leq \lambda_p^{-\frac{p}{2}} (\prod_{k=1}^{p-1} \prod_{i \in S_k} |(\Sigma_i^{k_i})|^{-\frac{1}{2}}) \exp\{-\frac{1}{2} \sum_{k=1}^{p-1} \sum_{i \in S_k} (y_i^{k_i} - x_i^{k_i T} \hat{\beta})^T (\Sigma_i^{k_i})^{-1} (y_i^{k_i} - x_i^{k_i T} \hat{\beta})\} \pi_{(\gamma)} \pi_{(\eta)}.$$

Now, define $M_i^{k_i}(\hat{\beta}) = (y_i^{k_i} - x_i^{k_i T} \hat{\beta})(y_i^{k_i} - x_i^{k_i T} \hat{\beta})^T$. The only positive eigenvalue of $M_i^{k_i}(\hat{\beta})$ is $\lambda_1(M_i^{k_i}(\hat{\beta})) = (y_i^{k_i} - x_i^{k_i T} \hat{\beta})^T (y_i^{k_i} - x_i^{k_i T} \hat{\beta}) > 0$ (Marshall and Olkin, 1979). Similar to Daniels (2006), we remove the dependence of $M_i^{k_i}(\hat{\beta})$ on $\Sigma_i^{k_i}$ by bounding the exponential term. Rewrite the exponential term in the above expression as:

$$\begin{aligned} & (y_i^{k_i} - x_i^{k_i T} \hat{\beta})^T (\Sigma_i^{k_i})^{-1} (y_i^{k_i} - x_i^{k_i T} \hat{\beta}) \\ = \; & trace[(\Sigma_i^{k_i})^{-1} M_i^{k_i}(\hat{\beta})] \\ \geq \; & \sum_{t=1}^{k} \lambda_t(\Sigma_i^{-k_i}) \lambda_{k-t+1}(M_i^{k_i}(\hat{\beta})) \\ = \; & \lambda_k(\Sigma_i^{-k_i}) \lambda_1(M_i^{k_i}(\hat{\beta})) \end{aligned}$$

where $\lambda_t(\bullet)$, defined as $\lambda_1(A), \lambda_2(A), ..., \lambda_p(A)$ are the ordered eigenvalues of a $p \times p$ matrix A. The first inequality is from Marshall and Olkin (1979).
Let $\lambda_{min,k} = min_{i \in S_k}\{\lambda_k(\Sigma_i^{-k_i})\} > 0$ and $\lambda_{min,k}^{(M)} = min_{i \in S_k}\{\lambda_1(M_i^{-k_i})\} > 0$. Then,

$$\begin{aligned} & \sum_{i \in S_k} (y_i^{k_i} - x_i^{k_i T} \hat{\beta})^T (\Sigma_i^{k_i})^{-1} (y_i^{k_i} - x_i^{k_i T} \hat{\beta}) \\ \geq \; & \sum_{i \in S_k} \lambda_k(\Sigma_i^{-k_i}) \lambda_1(M_i^{-k_i}) \\ \geq \; & s_k \lambda_{min,k} \lambda_{min,k}^{(M)} \\ = \; & trace\{\frac{s_k \lambda_{min,k}}{k} I_k \times \lambda_{min,k}^{(M)} I_k\} \end{aligned}$$

where s_k denotes cardinality of set S_k and I_k is a $k \times k$ identity matrix. Finally, for each i, simulate a 'new' set of data, $y_i^{\star k_i}$ from a multivariate normal distribution such that $\sum_{i \in S_k} y_i^{\star k_i} y_i^{\star k_i T} = \lambda_{min,k}^{(M)} I_k$.

We then obtain

$$
\begin{aligned}
& (\prod_{k=1}^{p-1}\prod_{i\in S_k}|(\Sigma^{-k_i})|^{\frac{1}{2}})\exp\{-\frac{1}{2}\sum_{k=1}^{p-1}\sum_{i\in S_k}(y_i^{k_i}-x_i^{k_iT}\hat{\beta})^T(\Sigma_i^{k_i})^{-1}(y_i^{k_i}-x_i^{k_iT}\hat{\beta})\}\pi_{(\gamma)}\pi_{(\eta)} \\
\leq\ & (\prod_{k=1}^{p-1}\prod_{i\in S_k}[\Pi_{t=1}^{k}\lambda_t(\Sigma_i^{-k_i})]^{\frac{1}{2}})\exp\{-\frac{1}{2}\sum_{k=1}^{p-1}\sum_{i\in S_k}y_i^{\star k_iT}\frac{s_k\lambda_{min,k}}{k}I_k y_i^{\star k_i}\}\pi_{(\gamma)}\pi_{(\eta)} \\
=\ & (\prod_{k=1}^{p-1}\prod_{i\in S_k}\Pi_{t=1}^{k}[\frac{k\lambda_t(\Sigma_i^{-k_i})}{s_k\lambda_{min,k}}]^{\frac{1}{2}}[\frac{s_k\lambda_{min,k}}{k}]^{\frac{1}{2}})\exp\{-\frac{1}{2}\sum_{k=1}^{p-1}\sum_{i\in S_k}y_i^{\star k_iT}\frac{s_k\lambda_{min,k}}{k}I_k y_i^{\star k_i}\}\pi_{(\gamma)}\pi_{(\eta)} \\
=\ & (\prod_{k=1}^{p-1}\prod_{i\in S_k}\Pi_{t=1}^{k}[\frac{k\lambda_t(\Sigma_i^{-k_i})}{s_k\lambda_{min,k}}]^{\frac{1}{2}})\prod_{k=1}^{p-1}\prod_{i\in S_k}|\frac{s_k\lambda_{min,k}}{k}I_k|^{\frac{1}{2}}\exp\{-\frac{1}{2}\sum_{k=1}^{p-1}\sum_{i\in S_k}y_i^{\star k_iT}\frac{s_k\lambda_{min,k}}{k}I_k y_i^{\star k_i}\}\pi_{(\gamma)}\pi_{(\eta)} \\
\leq\ & M_0
\end{aligned}
$$

where M_0 is a finite constant since all three terms above are bounded. Therefore, $\int_{\Im_\beta} m(\beta,\gamma,\sigma|y_{obs},x)d\beta$ is finite.
Since the priors on γ,η are proper under the assumption that $\sum A_iA_i^T$ and $\sum w_iw_i^T$ are full rank, we have

$$
\int_{\beta\in\Im_\beta}\int_{\gamma\in\Im_\gamma}\int_{\eta\in\Im_\eta} m(\beta,\gamma,\sigma|y_{obs},x)d\beta d\gamma d\eta \leq \int_{\gamma\in\Im_\gamma}\int_{\eta\in\Im_\eta} M_0\pi_{(\gamma)}\pi_{(\eta)}d\gamma d\eta < \infty.
$$

Hence, the posterior of (β,γ,η) is proper.