Contents

List of Tables

List of Figures

1. On supervised learning of multivariate skew normal mixture models with missing information

1.1. Introduction

Finite mixture models have become a flexible and powerful probabilistic learning tool for heterogeneous multivariate data and been used extensively in classification and clustering. During the last two decades, the usefulness of gaussian mixture (GMIX) (Pearson,1984) and Student's t mixture (TMIX) models, see Peel and McLachlan (2000), Shoham (2002), Shoham et al. (2003) and Lin et al. (2004), are being increasingly applied in various research fields such as pattern recognition, data mining, computer vision, signal and image processing, machine learning and bioinformatics, etc. For a comprehensive introduction to mixture models and their applications, see monographs by Titterington et al. (1985), McLachlan and Basford (1988), McLachlan and Peel (2000), Frühwirth-Schnatter (2006) and the references therein. Recently, mixtures of univariate skew normal and skew t distributions as natural extensions of univariate gaussian mixtures have been considered by Lin et al. (2007a; 2007b).

It is common in situations where data may exhibit highly asymmetric observations and thus statistical inferences drawn from the ordinary gaussian assumptions may frequently yield unreliable inferences. To reduce unduly skewness encountered in general practice, one commonly adopted approach is through the best known data-based power transformation proposed by Box and Cox (1964). Although such a treatment is very convenient to use, the achievement of joint normality is rarely satisfied and the transformed variables become more difficult to interpret. Instead of applying transformation methods, there has been a growing interest in proposing a wider class of distributions, the multivariate skew normal (MSN) distribution, which contains an extra vector of parameters in regulating skewness and includes the gaussian family as a special case.

The MSN distribution was originally studied by Azzalini and Dalla Valle (1996) and some further attractive features and applications are given in Azzalini and Capitaino (1999). Based on this class of distributions, a number of extensions or alternative proposals have appeared during the last decade. Arellano-Valle and Genton (2005) studied the family of fundamental skew normal (FUSN) distributions, giving a unified scheme to obtain MSN distributions starting from symmetric ones. Subsequently, Arrellano-Valle and Azzalini (2006) provide a survey on some of its extensions and variants. Sahu et al. (2003) defined a new class of MSN distributions and remarked that this sort of formulation is more flexible in terms of adjusting the correlation structure than the MSN of Azzalini and Dalla Valle (1996). Recently, Lin (2009) introduced a new mixture modeling framework with component densities using the MSN distribution of Sahu et al. (2003) and showed its great flexibility in modeling asymmetrically data.

Learning mixture models from incomplete data has become a powerful tool to handle real-world multivariate data sets with complex missing patterns. The work was pioneered by Ghahramani and Jordan (1994), who applied the Expectation Maximization (EM) algorithm (Dempster et al.,1997) to compute maximun likelihood (ML) estimates of the GMIX model with arbitrary patterns of missingness. Lin et al. (2006) extended their approach by introducing some efficient learning strategies from both ML and Bayesian perspectives. Wang et al. (2004) presented an ordinary EM algorithm for ML estimation of TMIX models with missing information. Related work on using the parameter expanded Expectation Maximization

(PX-EM) algorithm (Liu et al., 1998) for the supervised learning of TMIX models with incomplete data was done by Lin et al. (2008).

Throughout this dissertation, we assume that the mechanism of missingness is missing at random (MAR), which means the probability of outcome data being missing is conditionally independent of the values of missing data when given the observed data, see Rubin (1976), Schafer (1997), and Little and Rubin (2002) for a more detailed discussion. For computational aspects, we offer an analytically tractable EM algorithm coupled some useful model-based tools to handle data with general missing patterns in the class of multivariate skew normal mixture (MSNMIX) model. To reduce complications during the EM procedure, we introduce two binary auxiliary matrices for indexing the observed and missing components of each datum. Under this model, we also offer a conditional predictor to retrieve the missing components and a classifier for allocating partially observed vectors.

In the next two sections, we describe the MSN and the multivariate truncated normal distributions, define the notations and study some related properties. In Section 1.4, we present the MSNMIX model in an incomplete data framework and offer a computationally feasible EM algorithm to compute the ML estimates. The standard errors are derived from the information-based method instead of using resampling techniques. In Section 1.5, the proposed methodologies are applied to a real data set with varying proportions of synthetic missing values. Some concluding remarks are given in Section 1.6.

1.2. The multivariate skew normal distribution

A random vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ is said to have the p-dimensional skew normal distribution with a $p \times 1$ location vector $\boldsymbol{\xi}$, a $p \times p$ positive definite scale covariance matrix $\mathbf{\Sigma}$, and a $p \times p$ skewness matrix $\mathbf{\Lambda} = \text{Diag}\{\lambda_1, \ldots, \lambda_p\}$ if its joint probability density function (pdf) is given by

$$\psi(\mathbf{Y}|\ \boldsymbol{\xi}, \mathbf{\Sigma}, \mathbf{\Lambda}) = 2^p \phi_p(\mathbf{Y}|\ \boldsymbol{\xi}, \mathbf{\Omega}) \Phi_p(\mathbf{\Lambda}^\top \mathbf{\Omega}^{-1}(\mathbf{Y} - \boldsymbol{\xi}) \mid \mathbf{\Delta}), \tag{1.1}$$

with $\mathbf{\Omega} = \mathbf{\Sigma} + \mathbf{\Lambda}^2$ and $\mathbf{\Delta} = (\mathbf{I}_p + \mathbf{\Lambda}\mathbf{\Sigma}^{-1}\mathbf{\Lambda})^{-1} = \mathbf{I}_p - \mathbf{\Lambda}\mathbf{\Omega}^{-1}\mathbf{\Lambda}$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix. Moreover, $\phi_p(\cdot \mid \boldsymbol{\mu}, \mathbf{\Sigma})$ and $\Phi_p(\cdot \mid \mathbf{\Sigma})$ denote the pdf of $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ and cumulative density function (cdf) of $N_p(\mathbf{0}, \mathbf{\Sigma})$, respectively. If the p-dimensional random vector $\mathbf{Y}$ has the pdf in Eq. (1.1), it will be denoted by $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \mathbf{\Sigma}, \mathbf{\Lambda})$. Typically, if $\mathbf{\Lambda}$ is assumed to be a diagonal matrix, then the covariance structure of $\mathbf{Y}$ is not affected by the skewness. By the above reason, we assumed $\mathbf{\Lambda}$ is a diagonal matrix throughout this paper.

Assuming $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, it follows that $|\mathbf{Z}|$ is distributed as a p-dimensional standard half-normal distribution, denoted by $HN_p(\mathbf{0}, \mathbf{I}_p)$. A two-level hierarchical version of the linear mixed-effects model

$$\mathbf{Y} = \boldsymbol{\xi} + \mathbf{\Lambda}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1.2}$$

can be expressed by:

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\gamma} &\sim N_p(\boldsymbol{\xi} + \mathbf{\Lambda}\boldsymbol{\gamma}, \mathbf{\Sigma}), \\ \boldsymbol{\gamma} &\sim HN_p(\mathbf{0}, \mathbf{I}_p), \end{aligned}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are independently distributed as $HN_p(\mathbf{0}, \mathbf{I}_p)$ and $N_p(\mathbf{0}, \mathbf{\Sigma})$, respectively. Then the marginal distribution of $\mathbf{Y}$ is $SN_p(\boldsymbol{\xi}, \mathbf{\Sigma}, \mathbf{\Lambda})$. By Lin (2009), we have the following lemma.

Lemma 1.1 Let $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \mathbf{\Sigma}, \mathbf{\Lambda})$. Then

(i) $\text{E}(\mathbf{Y}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}}\mathbf{\Lambda}\mathbf{1}_p$,

(ii) $\mathrm{cov}(\mathbf{Y}) = \mathbf{\Sigma} + \left(1 - \frac{2}{\pi}\right) \mathbf{\Lambda}^2$,

where $\mathbf{1}_p$ is a p-dimensional vector of ones.

1.3. The multivariate truncated normal distribution

We use the notation $\Pi_{i=1}^{p} \int_{a_i}^{\infty} = \int_{a_1}^{\infty} \cdots \int_{a_p}^{\infty}$ for the abbreviation of multiple integrals. A p-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ is said to have the p-dimensional truncated normal distribution with a $p \times 1$ location vector $\boldsymbol{\mu}$, a $p \times p$ positive definite scale covariance matrix $\mathbf{\Sigma}$, and a truncated hyperplane region $\mathbb{A} = \{\mathbf{Y} = (Y_1, \ldots, Y_p)^\top | Y_1 > a_1, \ldots, Y_p > a_p\}$ with a_i being arbitary real numbers for all $i = 1, \ldots, p$ if its joint pdf is given by

$$f(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{\Sigma}; \mathbb{A}) = \frac{1}{\alpha} \phi_p(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{\Sigma}) I_{\mathbb{A}}(\mathbf{Y}), \tag{1.3}$$

where $\alpha = \Pi_{i=1}^{p} \int_{a_i}^{\infty} \phi_p(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{\Sigma}) d\mathbf{Y}$ and $I_{\mathbb{A}}(\mathbf{Y})$ is the indicator functions whose value equals one if $\mathbf{Y} \in \mathbb{A}$ and zero elsewhere. If the p-dimensional random vector $\mathbf{Y}$ has the pdf in Eq. (1.3), it will be denoted by $\mathbf{Y} \sim TN_p(\boldsymbol{\mu}, \mathbf{\Sigma}; \mathbb{A})$. By Lin (2009), we have the following lemma.

Lemma 1.2 Let $\mathbf{Y} \sim TN_p(\boldsymbol{\mu}, \mathbf{\Sigma}; \mathbb{A})$. Then

$$E(Y_i) = \mu_i + \alpha^{-1} \sum_{r=1}^{p} \sigma_{ri} f_r(a_r) G_{(r)}, \tag{1.4}$$

where σ_{ij} denotes the (i, j)th entry of $\mathbf{\Sigma}$, $f_r(a_r) = \phi(a_r|\mu_r, \sigma_{rr})$ denotes a normal density with mean μ_r and variance σ_{rr} for the rth variable evaluated at a_r and $G_{(r)} = \prod_{j \neq r} \int_{a_j}^{\infty} \phi_{p-1}(\mathbf{Y}_{(r)}|\boldsymbol{\mu}_{2\cdot 1}^{(r)}, \mathbf{\Sigma}_{22\cdot 1}^{(r)}) d\mathbf{Y}_{(r)}$ with $\phi_{p-1}(\mathbf{Y}_{(r)}|\boldsymbol{\mu}_{2\cdot 1}^{(r)}, \mathbf{\Sigma}_{22\cdot 1}^{(r)})$ being the conditional density of the remaining $p - 1$ variables given $Y_r = a_r$.

Moreover,

$$E(Y_iY_j) = \mu_i E(Y_j) + \mu_j E(Y_i) - \mu_i\mu_j + \sigma_{ij} + \alpha^{-1}\left\{\sum_{r=1}^{p}\frac{\sigma_{ri}\sigma_{rj}}{\sigma_{rr}}(a_r-\mu_r)f_r(a_r)G_{(r)} + \sum_{r=1}^{p}\sigma_{ir}\left[\sum_{s\neq r}^{p}\left(\sigma_{sj}-\frac{\sigma_{rs}\sigma_{rj}}{\sigma_{rr}}\right)f_{r,s}(a_r,a_s)G_{(rs)}\right]\right\}, \quad (1.5)$$

where $f_{r,s}(a_r, a_s)$ is a bivariate normal density of the (r,s)th variables of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ evaluated at (a_r, a_s) and $G_{(rs)} = \prod_{j\neq r,s}\int_{a_j}^{\infty}\phi_{p-2}(\mathbf{Y}_{(rs)}|\boldsymbol{\mu}_{2\cdot 1}^{(rs)}, \boldsymbol{\Sigma}_{22\cdot 1}^{(rs)})d\mathbf{Y}_{(rs)}$ with $\phi_{p-2}(\mathbf{Y}_{(rs)}|\boldsymbol{\mu}_{2\cdot 1}^{(rs)}, \boldsymbol{\Sigma}_{22\cdot 1}^{(rs)})$ being the conditional density of the remaining $p-2$ variables given $Y_r = a_r$ and $Y_s = a_s$.

Let $[\mathbf{A}]_{rs}$ denote the (r,s)th entry of $\mathbf{A}$. Expressions in Eqs (1.4) and (1.5) can be written in matrix form as follows:

$$E(\mathbf{Y}) = \boldsymbol{\mu} + \alpha^{-1}\boldsymbol{\Sigma}\mathbf{q} = \boldsymbol{\eta}, \quad (1.6)$$

where $\mathbf{q} = (q_1, \ldots, q_p)^\top$ is a $p \times 1$ vector with rth entry is $f_r(a_r)G_{(r)}$, and

$$E(\mathbf{Y}\mathbf{Y}^\top) = \boldsymbol{\mu}\boldsymbol{\eta}^\top + \boldsymbol{\eta}\boldsymbol{\mu}^\top - \boldsymbol{\mu}\boldsymbol{\mu}^\top + \boldsymbol{\Sigma} + \alpha^{-1}\boldsymbol{\Sigma}(\mathbf{H}+\mathbf{D})\boldsymbol{\Sigma}, \quad (1.7)$$

where $\mathbf{H}$ is a $p\times p$ matrix with all diagonal entries being zero and $f_{rs}(a_r, a_s)G_{(rs)}$ on the (r,s)th off-diagonal entry, and $\mathbf{D}$ is a $p \times p$ diagonal matrix with rth diagonal entry is $\sigma_{rr}^{-1}\big((a_r-\mu_r)f_r(a_r)G_{(r)} - [\boldsymbol{\Sigma}\mathbf{H}]_{rr}\big)$.

1.4. A multivariate skew normal mixture model with missing information

1.4.1. The model

In the MSNMIX model, we let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ be a set of p-dimensional random sample arising from a population with g subclasses $\mathcal{C}_1, \ldots, \mathcal{C}_g$. That is, each $\mathbf{Y}_j$ has the density

$$f(\mathbf{Y}_j|\boldsymbol{\Theta}) = \sum_{i=1}^{g} w_i\psi_p(\mathbf{Y}_j \mid \boldsymbol{\xi}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\Lambda}_i), \quad w_i \geq 0, \quad \sum_{i=1}^{g}\omega_i = 1, \quad (1.8)$$

where $\mathbf{\Lambda}_i = \text{Diag}(\boldsymbol{\lambda}_i)$ with $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{ip})^\top$ and the unknown parameter vector $\mathbf{\Theta}$ contains the mixing probabilities w_i $(i = 1, \ldots, g-1)$, the elements of component locations $\boldsymbol{\xi}_i$'s, the distinct elements of component scale covariance matrices $\mathbf{\Sigma}_i$'s and the skewness vectors $\boldsymbol{\lambda}_i$'s. Note that the notation $\psi_p(\cdot \mid \boldsymbol{\xi}_i, \mathbf{\Sigma}_i, \mathbf{\Lambda}_i)$ is the MSN density defined in (1.1) and Diag$(\cdot)$ denotes a diagonal matrix created by extracting the main diagonal elements of a square matrix or the diagonalization of a vector. The mean and covariance of $\mathbf{Y}_j$ are given by

$$\begin{aligned} E(\mathbf{Y}_j) &= \sum_{i=1}^{g} w_i \boldsymbol{\mu}_i, \\ \text{cov}(\mathbf{Y}_j) &= \sum_{i=1}^{g} \left\{ w_i (1 - w_i) \boldsymbol{\mu}_i \boldsymbol{\mu}_i^\top + w_i \mathbf{\Sigma}_i^* \right\} - \sum_{i \neq j}^{g} w_i w_j \boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top, \end{aligned}$$

where $\boldsymbol{\mu}_i = \boldsymbol{\xi}_i + \sqrt{2/\pi} \boldsymbol{\lambda}_i$ and $\mathbf{\Sigma}_i^* = \mathbf{\Sigma}_i + (1 - 2/\pi)\mathbf{\Lambda}_i^2$ are the mean vector and covariance matrix of $SN_p(\boldsymbol{\xi}_i, \mathbf{\Sigma}_i, \mathbf{\Lambda}_i)$, respectively.

To pose model (1.8) into an EM framework, we introduce allocation variables $\mathbf{Z}_j = (Z_{1j}, \ldots, Z_{gj})^\top$, one for each individual $\mathbf{Y}_j$, whose role is to encode which component has generated $\mathbf{Y}_j$. Specifically, the indicators $\mathbf{Z}_j$ $(j = 1, \ldots, n)$ are a $g \times 1$ vector of binary variables, whose elements are

$$Z_{sj} = \begin{cases} 1 & \text{if } \mathbf{Y}_j \text{ belongs to group } s, \\ 0 & \text{otherwise,} \end{cases}$$

and satisfy $\sum_{i=1}^{g} Z_{ij} = 1$. This implies $\mathbf{Z}_j$ follows a multinomial random vector with 1 trial and cell probabilities $w_1, \ldots, w_g$, denoted by $\mathbf{Z}_j \sim M(1; w_1, \ldots, w_g)$.

A three-level hierarchical representation of (1.8) can be expressed by

$$\begin{aligned} \mathbf{Y}_j \mid (\boldsymbol{\gamma}_j, Z_{ij} = 1) &\sim N_p(\boldsymbol{\xi}_i + \mathbf{\Lambda}_i \boldsymbol{\gamma}_j, \mathbf{\Sigma}_i), \\ \boldsymbol{\gamma}_j \mid (Z_{ij} = 1) &\sim HN_p(\mathbf{0}, \mathbf{I}_p), \\ \mathbf{Z}_j &\sim M(1; w_1, \ldots, w_g), \end{aligned} \tag{1.9}$$

for $i = 1, \ldots, g$ and $j = 1, \ldots, n$. From (1.9), we therefore declare the complete data vector to be $(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\gamma})$, where $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$, $\mathbf{Z} = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_n^\top)^\top$. From (1.9), the complete data likelihood function for $\boldsymbol{\Theta}$ is

$$L_c(\boldsymbol{\Theta} \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\gamma}) \propto \prod_{i=1}^{g} \prod_{j=1}^{n} \left(w_i \phi_p(\mathbf{Y}_j \mid \boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j, \boldsymbol{\Sigma}_i)\right)^{Z_{ij}}. \tag{1.10}$$

We are interested in ML estimation problem of model (1.8) when $\mathbf{Y}$ may be partially observed. The underlying missingness mechanism is assumed to be MAR. Simply speaking, the missingness of data is unrelated to missing values, but it might depend on the observed values.

Following Lin et al. (2006), we partition $\mathbf{Y}_j$ into two components $(\mathbf{Y}_j^{\mathrm{o}\top}, \mathbf{Y}_j^{\mathrm{m}\top})^\top$, where $\mathbf{Y}_j^{\mathrm{o}}$ $(p_j^{\mathrm{o}} \times 1)$ and $\mathbf{Y}_j^{\mathrm{m}}$ $((p - p_j^{\mathrm{o}}) \times 1)$ denote the observed and missing components of $\mathbf{Y}_j$, respectively. To facilitate the computation, we introduce two types of the binary indicator matrices, denoted by $\mathbf{O}_j$ $(p_j^{\mathrm{o}} \times p)$ and $\mathbf{M}_j$ $((p - p_j^{\mathrm{o}}) \times p)$, satisfying $\mathbf{Y}_j^{\mathrm{o}} = \mathbf{O}_j \mathbf{Y}_j$ and $\mathbf{Y}_j^{\mathrm{m}} = \mathbf{M}_j \mathbf{Y}_j$, which can be extracted from a p-dimensional identity matrix $\mathbf{I}_p$ corresponding to row positions of $\mathbf{Y}_j^{\mathrm{o}}$ and $\mathbf{Y}_j^{\mathrm{m}}$ in $\mathbf{Y}_j$, respectively. It is straightforward to verify that (a) $\mathbf{Y}_j = \mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}}$; (b) $\mathbf{O}_j^\top \mathbf{O}_j + \mathbf{M}_j^\top \mathbf{M}_j = \mathbf{I}_p$. Furthermore, we can establish the following results.

Theorem 1.1 *Let $\mathbf{Y}_j \sim \sum_{i=1}^{g} w_i \psi_p(\mathbf{Y}_j \mid \boldsymbol{\xi}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\Lambda}_i)$, and let $\mathbf{Y}_j^{\mathrm{o}}$ and $\mathbf{Y}_j^{\mathrm{m}}$ be the observed and the missing components corresponding to $\mathbf{Y}_j$, respectively. We have*

(a) *The marginal density of $\mathbf{Y}_j^{\mathrm{o}}$ is $\sum_{i=1}^{g} w_i \psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}, \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}})$, where $\boldsymbol{\xi}_{ij}^{\mathrm{o}} = \mathbf{O}_j \boldsymbol{\xi}_i$, $\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}} = \mathbf{O}_j \boldsymbol{\Sigma}_i \mathbf{O}_j^\top$ and $\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} = \mathbf{O}_j \boldsymbol{\Lambda}_i \mathbf{O}_j^\top$.*

(b) *The conditional density of $\mathbf{Y}_j^{\mathrm{m}}$ given $\mathbf{Y}_j^{\mathrm{o}}$ is*

$$f(\mathbf{Y}_j^{\mathrm{m}} \mid \mathbf{Y}_j^{\mathrm{o}}) = 2^p \sum_{i=1}^{g} \tilde{w}_{ij} \phi_{p-p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{m}} \mid \boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}) \Phi_p(\boldsymbol{\Lambda}_i \boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i) \mid \boldsymbol{\Delta}_i),$$

where $\tilde{w}_{ij} = w_i\phi_{p_j^{\rm o}}(\mathbf{Y}_j^{\rm o} \mid \boldsymbol{\xi}_{ij}^{\rm o}, \boldsymbol{\Omega}_{ij}^{\rm oo})/\sum_{h=1}^{g} w_h\psi_{p_j^{\rm o}}(\mathbf{Y}_j^{\rm o} \mid \boldsymbol{\xi}_{hj}^{\rm o}, \boldsymbol{\Sigma}_{hj}^{\rm oo}, \boldsymbol{\Lambda}_{hj}^{\rm oo})$, $\boldsymbol{\xi}_{ij}^{\rm m\cdot o} = \mathbf{M}_j\big(\boldsymbol{\xi}_i + \boldsymbol{\Omega}_i\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi}_i)\big)$ *and* $\boldsymbol{\Omega}_{ij}^{\rm mm\cdot o} = \mathbf{M}_j(\mathbf{I}_p - \boldsymbol{\Omega}_i\mathbf{C}_{ij}^{\rm oo})\boldsymbol{\Omega}_i\mathbf{M}_j^\top$ *with* $\boldsymbol{\Omega}_{ij}^{\rm oo} = \mathbf{O}_j\boldsymbol{\Omega}_i\mathbf{O}_j^\top$ *and* $\mathbf{C}_{ij}^{\rm oo} = \mathbf{O}_j^\top\boldsymbol{\Omega}_{ij}^{\rm oo^{-1}}\mathbf{O}_j$.

Proof. The proof is given in Appendix A.

Theorem 1.2 *From (1.9), we have the following conditional distributions:*

(a) *The conditional distribution of* $\mathbf{Y}_j^{\rm o}$ *given* $\boldsymbol{\gamma}_j$ *and* $Z_{ij} = 1$ *is*

$$\mathbf{Y}_j^{\rm o} \mid (\boldsymbol{\gamma}_j, Z_{ij} = 1) \sim N_{p_j^{\rm o}}(\boldsymbol{\mu}_{ij}^{\rm o}, \boldsymbol{\Sigma}_{ij}^{\rm oo}),$$

where $\boldsymbol{\mu}_{ij}^{\rm o} = \mathbf{O}_j(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)$ *and* $\boldsymbol{\Sigma}_{ij}^{\rm oo} = \mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top$.

(b) *The conditional distribution of* $\mathbf{Y}_j^{\rm m}$ *given* $\mathbf{Y}_j^{\rm o}$, $\boldsymbol{\gamma}_j$, *and* $Z_{ij} = 1$ *is*

$$\mathbf{Y}_j^{\rm m} \mid (\mathbf{Y}_j^{\rm o}, \boldsymbol{\gamma}_j, Z_{ij} = 1) \sim N_{p-p_j^{\rm o}}(\boldsymbol{\mu}_{ij}^{\rm m\cdot o}, \boldsymbol{\Sigma}_{ij}^{\rm mm\cdot o}),$$

where $\boldsymbol{\mu}_{ij}^{\rm m\cdot o} = \mathbf{M}_j(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j + \boldsymbol{\Sigma}_i\mathbf{S}_{ij}^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j))$, $\boldsymbol{\Sigma}_{ij}^{\rm mm\cdot o} = \mathbf{M}_j(\mathbf{I}_p - \boldsymbol{\Sigma}_i\mathbf{S}_{ij}^{\rm oo})\boldsymbol{\Sigma}_i\mathbf{M}_j^\top$, *and* $\mathbf{S}_{ij}^{\rm oo} = \mathbf{O}_j^\top(\mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top)^{-1}\mathbf{O}_j$.

(c) *The conditional distribution of* $\boldsymbol{\gamma}_j$ *given* $\mathbf{Y}_j^{\rm o}$ *and* $Z_{ij} = 1$ *is*

$$\boldsymbol{\gamma}_j \mid (\mathbf{Y}_j^{\rm o}, Z_{ij} = 1) \sim TN_p\big(\boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi}_i), \mathbf{I}_p - \boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\rm oo}\boldsymbol{\Lambda}_i; \mathbb{R}_+^p\big),$$

where $\mathbb{R}_+^p$ *is* $\mathbb{R}^p$ *with all elements being positive real numbers.*

Proof. The proofs of part (a) and part (b) are straightforward and hence are omitted. The proof of part (c) is given in Appendix B.

Let $E(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\rm o}, Z_{ij} = 1) = \boldsymbol{\eta}_{ij}$ and $E(\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\rm o}, Z_{ij} = 1) = \boldsymbol{\Psi}_{ij}$. Both of which are implicit functions of parameters $\boldsymbol{\xi}_i$, $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Lambda}_i$ and can be easily evaluated by using Eqs. (1.6) and (1.7). The following corollary is a direct implication of Theorem 1.2.

Corollary 1.1 *Recalling* $\mathbf{Y}_j = \mathbf{O}_j^\top \mathbf{Y}_j^{\rm o} + \mathbf{M}_j^\top \mathbf{Y}_j^{\rm m}$ *and* $\mathbf{O}_j^\top \mathbf{O}_j + \mathbf{M}_j^\top \mathbf{M}_j = \mathbf{I}_p$, *these give rise to* $\mathbf{O}_j^\top \mathbf{O}_j(\mathbf{I}_p - \mathbf{\Sigma}_i \mathbf{S}_{ij}^{\rm oo}) = \mathbf{0}$. *We can obtain*

(a) $E(\mathbf{Y}_j \mid \mathbf{Y}_j^{\rm o}, Z_{ij} = 1) = \boldsymbol{\xi}_i + \mathbf{\Lambda}_i \boldsymbol{\eta}_{ij} + \mathbf{\Sigma}_i \mathbf{S}_{ij}^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i \boldsymbol{\eta}_{ij})$.

(b) $\mathrm{cov}(\mathbf{Y}_j \mid \mathbf{Y}_j^{\rm o}, Z_{ij} = 1) = (\mathbf{I}_p - \mathbf{\Sigma}_i \mathbf{S}_{ij}^{\rm oo}) \left(\mathbf{\Sigma}_i + \mathbf{\Lambda}_i (\mathbf{\Psi}_{ij} - \boldsymbol{\eta}_{ij} \boldsymbol{\eta}_{ij}^\top) \mathbf{\Lambda}_i (\mathbf{I}_p - \mathbf{S}_{ij}^{\rm oo} \mathbf{\Sigma}_i) \right)$.

(c) $E(\mathbf{Y}_j \boldsymbol{\gamma}_j^\top \mid \mathbf{Y}_j^{\rm o}, Z_{ij} = 1) = (\mathbf{I}_p - \mathbf{\Sigma}_i \mathbf{S}_{ij}^{\rm oo})(\boldsymbol{\xi}_i \boldsymbol{\eta}_{ij}^\top + \mathbf{\Lambda}_i \mathbf{\Psi}_{ij}) + \mathbf{\Sigma}_i \mathbf{S}_{ij}^{\rm oo} \mathbf{Y}_j \boldsymbol{\eta}_{ij}^\top$.

For a p-dimensional observation with the probability of its missingness greater than or equal to zero for each attribute, there are $2^p - 1$ unique patterns of missingness. In general, completely missing pattern does not happen. This indicates that the missing rate should be smaller than $(p-1)/p$. To lessen the computational load, Lin et al. (2006) have described a simple procedure by rearranging $\mathbf{Y}$ according to unique missing patterns of data.

1.4.2. An efficient EM procedure for ML estimation

The EM algorithm of Dempster et al. (1977) has been widely used in the literature to carry out ML estimation in a variety of incomplete data problems. We offer an efficient EM algorithm for learning model (1.8) from incomplete data.

For notational simplicity, let $\Upsilon = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_g, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_g, \mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_g)$ and $\mathbf{w} = (w_1, \ldots, w_g)^\top$. Let $\mathbf{Y}^{\rm o} = (\mathbf{Y}_1^{\rm o}, \mathbf{Y}_2^{\rm o}, \cdots, \mathbf{Y}_n^{\rm o})$ and $\mathbf{Y}^{\rm m} = (\mathbf{Y}_1^{\rm m}, \mathbf{Y}_2^{\rm m}, \cdots, \mathbf{Y}_n^{\rm m})$ represent the observed portion and the missing portion of the data, respectively. From (1.10), the complete data log-likelihood function of $\mathbf{\Theta}$, aside from additive constant terms, can be written by

$$
\begin{aligned}
&\ell_c(\boldsymbol{\Theta}|\mathbf{Y},\mathbf{Z},\boldsymbol{\gamma})\\
= &\sum_{i=1}^{g}\sum_{j=1}^{n} Z_{ij}\log(w_i) - \frac{1}{2}\sum_{i=1}^{g}\left[\log|\boldsymbol{\Sigma}_i|\left(\sum_{j=1}^{n} Z_{ij}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}\right)\right],
\end{aligned}
\tag{1.11}
$$

where

$$
\boldsymbol{\Omega}_{ij} = Z_{ij}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top.
$$

Given the observed data $\mathbf{Y}^{\mathrm{o}}$ and the current parameter estimates $\hat{\boldsymbol{\Theta}}^{(k)}$, the E-step needs to compute the expected log-likelihood of the complete data.

Lemma 1.3 The conditional expectation of Eq. (1.11) is given by

$$
\begin{aligned}
Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}) &= \mathrm{E}\left(\ell_c(\boldsymbol{\Theta}|\mathbf{Y},\mathbf{Z},\boldsymbol{\gamma})|\mathbf{Y}^{\mathrm{o}},\hat{\boldsymbol{\Theta}}^{(k)}\right)\\
&= Q_1(\mathbf{w}|\hat{\boldsymbol{\Theta}}^{(k)}) + Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)}).
\end{aligned}
$$

It follows that

$$
Q_1(\mathbf{w}|\hat{\boldsymbol{\Theta}}^{(k)}) = \sum_{i=1}^{g}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\log(w_i), \tag{1.12}
$$

$$
Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)}) = -\frac{1}{2}\sum_{i=1}^{g}\left[\log|\boldsymbol{\Sigma}_i|\left(\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right)\right], \tag{1.13}
$$

where

$$
\hat{Z}_{ij}^{(k)} = \frac{\hat{w}_i^{(k)}\psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)},\hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}(k)},\hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}(k)})}{\sum_{i=1}^{g}\hat{w}_i^{(k)}\psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)},\hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}(k)},\hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}(k)})}, \tag{1.14}
$$

$$
\hat{\mathbf{Y}}_{ij}^{(k)} = \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}), \tag{1.15}
$$

$$
\begin{aligned}
\boldsymbol{\Omega}_{ij}^{(k)} = \hat{Z}_{ij}^{(k)}\Big[&(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} + (\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})^\top\\
&+\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)(\hat{\boldsymbol{\Psi}}_{ij}^{(k)} - \hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)\top})\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top\Big],
\end{aligned}
\tag{1.16}
$$

with $\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)} = \mathbf{O}_j\hat{\boldsymbol{\xi}}_i^{(k)}$, $\hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}(k)} = \mathbf{O}_j\hat{\boldsymbol{\Sigma}}_i^{(k)}\mathbf{O}_j^\top$, $\hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}(k)} = \mathbf{O}_j\hat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{O}_j^\top$, $\hat{\mathbf{A}}_{ij}^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)}$ and $\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)} = \mathbf{O}_j^\top(\mathbf{O}_j\hat{\boldsymbol{\Sigma}}_i^{(k)}\mathbf{O}_j^\top)^{-1}\mathbf{O}_j$ for $i = 1,\ldots,g$ and $j = 1,\ldots,n$.

Proof. The proof is given in Appendix C.

The EM algorithm is as follows:

E-step: Given $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(k)}$, compute $\hat{Z}_{ij}^{(k)}$, $\boldsymbol{\Omega}_{ij}^{(k)}$ and $\hat{\mathbf{Y}}_{ij}^{(k)}$ for $i = 1,\ldots,g$ and $j = 1,\ldots,n$, using Eqs (1.14), (1.16) and (1.15).

M-step:

1. Update $\hat{w}_i^{(k)}$ by maximizing Eq. (1.12) over w_i subject to their sum is unity, which gives

$$\hat{w}_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}.$$

2. Fix $\boldsymbol{\Lambda}_i$ at $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, update $\hat{\boldsymbol{\xi}}_i^{(k)}$ by maximizing Eq. (1.13) over $\boldsymbol{\xi}_i$, which leads to

$$\hat{\boldsymbol{\xi}}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\mathbf{Y}}_{ij}^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}}{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}}.$$

3. Fix $\boldsymbol{\xi}_i$ at $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i$ at $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, update $\hat{\boldsymbol{\Sigma}}_i^{(k)}$ by maximizing Eq. (1.13) over $\boldsymbol{\Sigma}_i$, which leads to

$$\hat{\boldsymbol{\Sigma}}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}}{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}},$$

where $\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}$ is $\boldsymbol{\Omega}_{ij}^{(k)}$ in Eq. (1.16) with $\boldsymbol{\xi}_i$ replaced by $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i$ replaced by $\hat{\boldsymbol{\Lambda}}_i^{(k)}$.

4. Here $\boldsymbol{\Lambda}_i$ assumed to be diagonal, say $\boldsymbol{\Lambda}_i = \mathrm{Diag}(\boldsymbol{\lambda}_i)$, where $\boldsymbol{\lambda}_i$ is a p-dimensional vector. Fix $\boldsymbol{\xi}_i$ at $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Sigma}_i$ at $\hat{\boldsymbol{\Sigma}}_i^{(k+1)}$, update $\hat{\boldsymbol{\lambda}}_i^{(k)}$ by maximizing Eq. (1.13)

over $\boldsymbol{\lambda}_i$, which leads to

$$\begin{aligned}
\hat{\boldsymbol{\lambda}}_i^{(k+1)} &= \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^{-1} \\
&\times \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \Big((\hat{\boldsymbol{\Psi}}_{ij}^{(k)} - \hat{\boldsymbol{\eta}}_{ij}^{(k)} \hat{\boldsymbol{\eta}}_{ij}^{(k)\top}) \hat{\mathbf{A}}_{ij}^{\top(k)} + \hat{\boldsymbol{\eta}}_{ij}^{(k)} (\hat{\mathbf{Y}}_{ij}^{(k)} - \hat{\boldsymbol{\xi}}_i^{(k+1)})^{\top} \Big)\right) \mathbf{1}_p,
\end{aligned}$$

where $\mathbf{1}_p$ denotes a p-dimensional vector of ones and the operator $\odot$ denotes the elementwise product of two matrices of the same dimension.

The detailed proof of the M-steps is shown in Appendix D.

Since the stability and monotone convergence of EM are maintained, the iterations are repeated until a suitable convergence rule is satisfied, e.g., $\|\hat{\boldsymbol{\Theta}}^{(k+1)} - \hat{\boldsymbol{\Theta}}^{(k)}\|$ is sufficiently small. When the convergence is achieved, the resulting estimates are denoted by $\hat{\boldsymbol{\Theta}} = (\hat{w}_1, \ldots, \hat{w}_{g-1}, \hat{\boldsymbol{\xi}}_1, \ldots, \hat{\boldsymbol{\xi}}_g, \hat{\boldsymbol{\Sigma}}_1, \ldots, \hat{\boldsymbol{\Sigma}}_g, \hat{\boldsymbol{\Lambda}}_1, \ldots, \hat{\boldsymbol{\Lambda}}_g)$. Therefore, the posterior probability of the $\mathbf{Y}_j$ belonging to group i can be estimated by

$$\hat{w}_{ij}^{*} = \mathrm{P}(Z_{ij} = 1 | \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}) = \frac{\hat{w}_i \psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} \mid \hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}}, \hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}}, \hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}})}{\sum_{i=1}^{g} \hat{w}_i \psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} \mid \hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}}, \hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}}, \hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}})}. \tag{1.17}$$

According to the ML classification theory of (Basford and McLachlan, 1985), $\mathbf{Y}_j$ is assigned to group s if $\hat{w}_{sj}^{*} > \hat{w}_{ij}^{*}$ for $i = 1, \ldots, g$ and $i \neq s$. Consequently, the ML predictor for the missing component $\mathbf{Y}_j^{\mathrm{m}}$ is given by

$$\begin{aligned}
\hat{\mathbf{Y}}_j^{\mathrm{m}} &= E(\mathbf{Y}_j^{\mathrm{m}} | \mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}) \\
&= \mathbf{M}_j \sum_{i=1}^{g} \hat{w}_{ij}^{*} \big(\hat{\boldsymbol{\xi}}_i + \hat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\eta}}_{ij} + \hat{\boldsymbol{\Sigma}}_i \hat{\mathbf{S}}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\eta}}_{ij}) \big).
\end{aligned} \tag{1.18}$$

1.4.3. Estimation of standard errors

Under the regularity conditions given in Zacks (1971, Chapter 5), the standard errors of the ML estimates, $\hat{\boldsymbol{\Theta}}$, can be obtained by inverting the observed or expected information matrix. Efron and Hinkely (1978) suggested using the observed information matrix instead of the expected information matrix for the evaluation of the standard errors. Meilijson (1989) showed a remark that the observed as well as the expected information matrix could be estimated consistently by the empirical covariance matrix of the individual scores.

Let $\ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j, \boldsymbol{\tau}_j, \mathbf{Z}_j)$ be the complete data log-likelihood formed from the single observation $\mathbf{Y}_j$. The individual score is defined as

$$\mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\Theta}) = E\bigg(\frac{\partial \ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \mathbf{Z}_j)}{\partial \boldsymbol{\Theta}}\bigg|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\Theta}\bigg).$$

The empirical information matrix, according to Meilijson's formula, is defined as

$$\mathbf{I}_e(\boldsymbol{\Theta}|\mathbf{Y}^{\mathrm{o}}) = \sum_{j=1}^{n} \mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\Theta})\mathbf{u}^{\top}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\Theta}) - n^{-1}\mathbf{U}(\mathbf{Y}^{\mathrm{o}}|\boldsymbol{\Theta})\mathbf{U}^{\top}(\mathbf{Y}^{\mathrm{o}}|\boldsymbol{\Theta}), \tag{1.19}$$

where $\mathbf{U}(\mathbf{Y}^{\mathrm{o}}|\boldsymbol{\Theta}) = \sum_{j=1}^{n} \mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\Theta})$.

Let vech$(\cdot)$ be the matrix operator which stacks only the distinct elements of a symmetric matrix into a single vector. The ML estimates $\hat{\boldsymbol{\Theta}}$ substituted for $\boldsymbol{\Theta}$ in the Eq. (1.19) and then it reduced to

$$\mathbf{I}_e(\hat{\boldsymbol{\Theta}} \mid \mathbf{Y}^{\mathrm{o}}) = \sum_{j=1}^{n} \hat{\mathbf{u}}_j^{\mathrm{o}} \hat{\mathbf{u}}_j^{\mathrm{o}\top},$$

where

$$\begin{aligned}
\hat{\mathbf{u}}_j^{\mathrm{o}} &= \mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\Theta}}) \\
&= (\hat{u}_{j,w_1}^{\mathrm{o}}, \ldots, \hat{u}_{j,w_{g-1}}^{\mathrm{o}}, \hat{\mathbf{u}}_{j,\boldsymbol{\xi}_1}^{\mathrm{o}\top}, \ldots, \hat{\mathbf{u}}_{j,\boldsymbol{\xi}_g}^{\mathrm{o}\top}, \hat{\mathbf{u}}_{j,\boldsymbol{\sigma}_1}^{\mathrm{o}\top}, \ldots, \hat{\mathbf{u}}_{j,\boldsymbol{\sigma}_g}^{\mathrm{o}\top}, \hat{\mathbf{u}}_{j,\boldsymbol{\lambda}_1}^{\mathrm{o}\top}, \ldots, \hat{\mathbf{u}}_{j,\boldsymbol{\lambda}_g}^{\mathrm{o}\top})^{\top},
\end{aligned}$$

with $\boldsymbol{\lambda}_i = \text{diag}(\boldsymbol{\Lambda}_i)$ and $\boldsymbol{\sigma}_i = \text{vech}(\boldsymbol{\Sigma}_i)$. Expressions for the elements of $\hat{\mathbf{u}}_j^{\text{o}}$ are given by

$$\begin{aligned}
\hat{u}^{\text{o}}_{j,w_r} &= \frac{\hat{Z}_{rj}}{\hat{w}_r} - \frac{\hat{Z}_{gj}}{\hat{w}_g}, \\
\hat{\mathbf{u}}^{\text{o}}_{j,\boldsymbol{\xi}_i} &= \hat{Z}_{ij}\hat{\boldsymbol{\Sigma}}_i^{-1}\big(\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij}\big), \\
\hat{\mathbf{u}}^{\text{o}}_{j,\boldsymbol{\sigma}_i} &= \text{vech}\left(\hat{Z}_{ij}\hat{\mathbf{B}}_{ij} - \frac{1}{2}\hat{Z}_{ij}\text{Diag}(\hat{\mathbf{B}}_{ij})\right), \\
\hat{\mathbf{u}}^{\text{o}}_{j,\boldsymbol{\lambda}_i} &= \text{Diag}\left(\hat{Z}_{ij}\hat{\mathbf{S}}^{\text{oo}}_{ij}((\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i)\hat{\boldsymbol{\eta}}_{ij}^\top - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij})\right),
\end{aligned}$$

where

$$\hat{\mathbf{B}}_{ij} = \hat{\mathbf{S}}^{\text{oo}}_{ij}\left(\hat{\boldsymbol{\Lambda}}_i(\hat{\boldsymbol{\Psi}}_{ij} - \hat{\boldsymbol{\eta}}_{ij}\hat{\boldsymbol{\eta}}_{ij}^\top)\hat{\boldsymbol{\Lambda}}_i + (\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij})(\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij})^\top\right)\hat{\mathbf{S}}^{\text{oo}}_{ij} - \hat{\mathbf{S}}^{\text{oo}}_{ij}.$$

The detailed proof is shown in Appendix E.

1.5. Experimental results

For illustration purposes, we apply the techniques presented so far to a subset of the Australian Institute of Sport (AIS) data, including 13 physical variables on 102 male and 100 female athletes, which are treated as two intrinsic classes. The data were originally reported by Cook and Weisberg (1994) and have been analyzed already by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999) and Azzalini (2005), among others. They pointed out the AIS data are better suited to the MSN distribution than gaussian, but neglected the situation where patterns of multimodality happen. In this example, we select three attributes: BMI, Bfat and LBM, which represent the body mass index, the percentage of body fat and lean body mass, respectively.

Fig. 1.1 depicts pairwise bivariate scatter plots of the data with superimposed contours of the fitted 2-component MSNMIX distribution. It can be observed from the figure that the scatter plots and fitted densities reveal an apparent bimodal asymmetric mixture pattern for each of the three pairs of variables. Note that the mixture components for BMI and LBM are not well separated because the two attributes are highly correlated with the Pearson's correlation coefficient being 0.71.

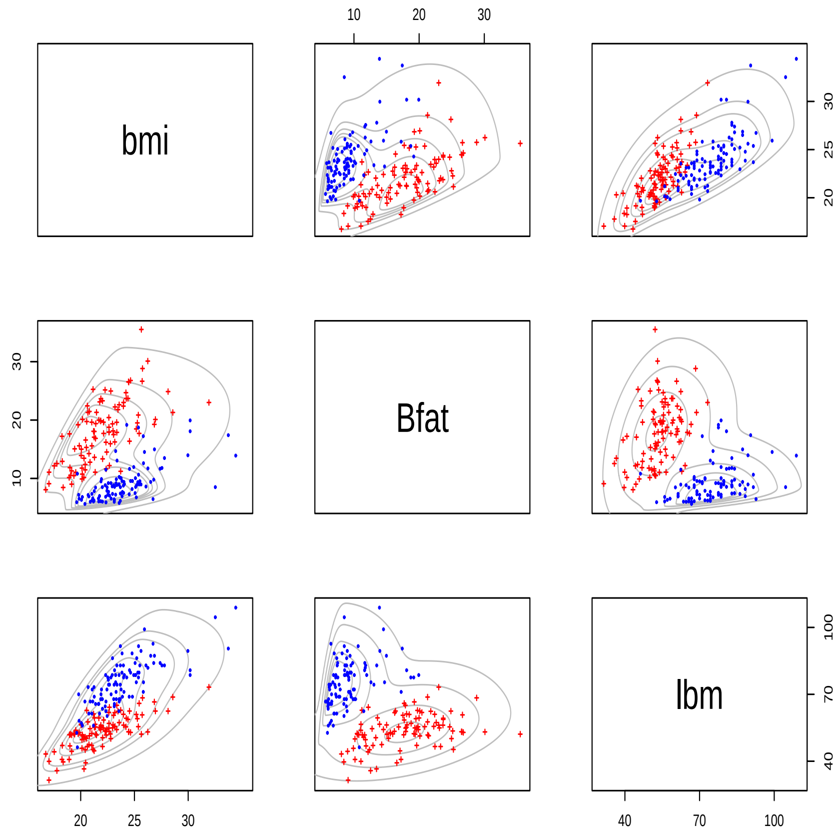

Figure 1.1: AIS data: bivariate scatter plots and fitted 2-component MSNMIX contours (+: Female; •: Male)

To conduct experimental studies, we first generate 500 synthetic missing data sets by deleting at random from the experimental data under various missing rates, $r\%$, where each datum retains at least one observed attribute. The missing rates of the synthetic data range from 10% up to 40% (increased by 10%). A relative difference of 10^{-5} in successive values of the log-likelihood is used as a stopping guideline for the EM algorithm.

We fit a MSNMIX model with density (1.8) to 500 synthetic missing data sets for $g = 1$ and $g = 2$, where $g = 1$ corresponds to the MSN model (a special case of MSNMIX model with a single component) of Sahu et al. (2003), which cannot capture the bimodality. Specifically, the 2-component MSNMIX model can be written as

$$f(\mathbf{Y}_j \mid \mathbf{\Theta}) = w\psi(\mathbf{Y}_j \mid \boldsymbol{\xi}_1, \mathbf{\Sigma}_1, \mathbf{\Lambda}_1) + (1-w)\psi(\mathbf{Y}_j \mid \boldsymbol{\xi}_2, \mathbf{\Sigma}_2, \mathbf{\Lambda}_2) \quad (j = 1, \ldots, 202),$$

where

$$\boldsymbol{\xi}_i = \begin{bmatrix} \xi_{i1} \\ \xi_{i2} \\ \xi_{i3} \end{bmatrix}, \quad \mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{i,11} & \sigma_{i,12} & \sigma_{i,13} \\ \sigma_{i,12} & \sigma_{i,22} & \sigma_{i,23} \\ \sigma_{i,13} & \sigma_{i,23} & \sigma_{i,33} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_i = \begin{bmatrix} \lambda_{i,11} & 0 & 0 \\ 0 & \lambda_{i,22} & 0 \\ 0 & 0 & \lambda_{i,33} \end{bmatrix}$$

for $i = 1, 2$.

For comparison, we test the null hypothesis $H_0 : g = 1$ (MSN) *versus* the alternative hypothesis $H_1 : g = 2$ (MSNMIX). The numbers of free parameters under H_0 and H_1 are 12 and 25, respectively. The likelihood ratio test (LRT) statistic, given by the difference in values of -2 times the log-likelihood between two nest models, is used to judge which of the two models is more suitable for this data set. Fig. 1.2 displays the histograms of converged log-likelihood values of the null and the alternative models along with a summarized box plot for their LRT statistics.

It is readily seen that the LRT statistics are highly significant compared with the χ^2_{13} distribution for all cases.

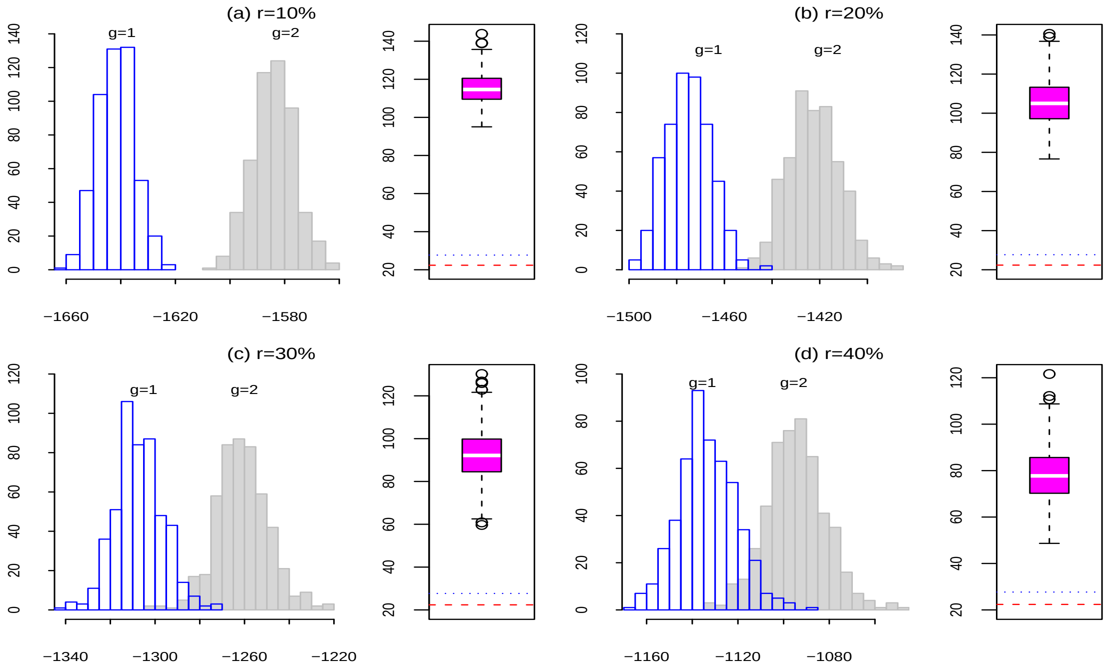

Figure 1.2: A comparison of converged log-likelihood values of the null ($g = 1$) and the alternative ($g = 2$) models and their LRT statistics, where dotted line $(\cdots)=\chi^2_{13}(0.99) = 27.69$ and dashed line $(- - -)=\chi^2_{13}(0.95) = 22.36$, for various proportions of missing values (Replications=500)

To exemplify the predictive accuracies on the imputation of missing values, we compare the MSN and MSNMIX predictors, see Eq. (1.18), together with the traditional randomization-based mean imputation (MI) predictor, known as a common heuristic by filling in a single value for each missing value with the observed sample mean of the associated attribute. As a measure of precision, the mean absolute error (MAE) and the mean absolute relative error (MARE) are used to evaluate the prediction discrepancy. Comparison results are listed in Table 1.1. The relative improvement percentage (RIP) in Table 1.1 is defined as the percentage decrease in the relative prediction error when comparing MSN and MSNMIX predictors. In this study, we found that both model-based predictors substantially outperform MI for all cases. Furthermore, the MSNMIX predictor exhibits considerable promising accuracy in the prediction of missing values when compared with those of MSN imputations over a wide range of missing rates.

Table 1.1: A comparison of averaged prediction accuracies and the associated standard deviations in parentheses for three imputation methods with varying proportions of missing values. The relative improvement percentage (RIP) is measured by (MSN-MSNMIX)/MSN×100%. (Replications=500)

Missing	MAE				MARE			
rate	MI	MSN	MSNMIX	RIP(%)	MI	MSN	MSNMIX	RIP(%)
10%	6.114 (0.733)	3.476 (0.495)	3.326 (0.499)	4.32	0.243 (0.032)	0.154 (0.024)	0.141 (0.023)	8.44
20%	6.124 (0.493)	3.777 (0.321)	3.637 (0.337)	3.71	0.246 (0.024)	0.166 (0.032)	0.155 (0.018)	6.63
30%	6.115 (0.380)	4.096 (0.300)	3.967 (0.319)	3.15	0.244 (0.018)	0.177 (0.015)	0.167 (0.015)	5.65
40%	6.115 (0.302)	4.382 (0.284)	4.275 (0.307)	2.44	0.245 (0.017)	0.190 (0.014)	0.182 (0.015)	4.21

As another illustration, we compare the supervised learning of classification accuracies between the GMIX and MSNMIX classifiers, see Eq. (1.17). Comparisons are made on the trivariate data and a reference bivariate sample (BMI, LBM). Table 1.2 shows the average misclassification rates from these models. As seen in the table, the misclassification rates of the MSNMIX classifier are all significantly smaller than those of GMIX classifier, especially for the bivariate sample with RIPs ranging between 43.3% and 60.1%. These observations signify that the MSNMIX model provides a sound statistical basis for clustering.

Table 1.2: A comparison of averaged misclassification rates (%) between GMIX and MSNMIX models with standard deviations in parentheses for three imputation methods with varying proportions of missing values. The relative improvement percentage (RIP) is measured by (GMIX-MSNMIX)/GMIX×100%. (Replications=500)

Missing	Trivariate data			Bivariate data (bmi, lbm)		
rate	GMIX	MSNMIX	RIP(%)	GMIX	MSNMIX	RIP(%)
10%	5.35 (0.015)	4.77 (0.010)	10.8	27.36 (0.129)	10.93 (0.022)	60.1
20%	6.99 (0.024)	6.24 (0.014)	12.0	31.71 (0.136)	14.60 (0.028)	54.0
30%	9.84 (0.034)	8.61 (0.018)	12.5	34.75 (0.123)	18.68 (0.036)	46.2
40%	13.11 (0.037)	11.63 (0.023)	11.3	36.20 (0.105)	20.52 (0.020)	43.3

1.6. Concluding remarks

We have established some properties related to the MSNMIX model in a missing information framework. The proposed model is very flexible in dealing with heterogeneous data that involve strong skewness and is persistent to the presence of missing observations. We discussed in detail how the EM algorithm coupled with auxiliary matrices can be applied on learning models from incomplete data in an efficient manner. Experimental results indicate that the MSNMIX model performs well for imputations as well as clustering when asymmetric multimodality and missing outcomes simultaneously occur in the input data. Finally, we highlight that, with the growing advances of modern stochastic computing technology and inexpensive high-speed computer power, it is worthwhile to pursue a fully Bayesian treatment (e.g., Hastings, 1970; Tanner and Wong, 1987; Diebolt and Robert, 1994; Escobar and West, 1995) in this context for enriching up-to-date account of the theory and applicability.

2. Robust Statistical Modeling Using The Multivariate Skew t Distribution With Incomplete Data

2.1. Introduction

In the statistical modeling of multivariate data, sometimes, not all designed measurements are fully collected. The occurrence of missing values in multivariate analysis is a common problem that might lead to biased estimates of parameters or inefficient inferences. Learning multivariate normal (MVN) models from incomplete data has been well-developed and systematically studied in the literature, see e.g., Anderson (1957), Hocking and Smith (1968), Rubin (1987) and Liu (1999). For analyzing data with incomplete observations, modern imputation methods such as the Expectation Maximization (EM; Dempster et al., 1977) and data augmentation (DA; Tanner and Wong, 1987) algorithms can be easily implemented by using the statistical packages: `proc MI` in SAS (SAS 2001) and the `Missing Data Analysis Library` in S-PLUS (Schimert et al., 2000).

Typically, the normality assumption for incomplete multivariate data is usually hard to be justified. Further, in some situations, the underlying distribution of input data may be asymmetrically distributed or it may contain some influential outliers. The multivariate t (MVT) distribution is a useful model for robust inference and covers a wide variety of applications (Kotz and Nadarajah, 2004). As pointed out by Liu (1995), to fit data having longer than normal tails, multiple imputations under MVT models allow to yield more valid statistical inferences than using the normal distribution. Lange et al. (1989) remarked the MVT model is not a panacea for modeling data with highly asymmetric observations.

Over the past decades, there has been a growing interest in proposing a parametric family of multivariate skew t (MST) distribution (Jones and Faddy, 2003; Azzalini and Capitaino, 2003), which is an extension of the MVT family with additional shape parameters to regulate skewness. In this paper, we consider the missing imputation problems under a new class of MST distribution, defined by Sahu et al. (2003), can accommodate a wider range of distributional features. Sahu et al. (2003) remarked that the correlation structure within this family is not affected by the introduction of skewness parameters.

In this class of MST models, we are devoted to developing tactical learning tools to handle missing data problems. Throughout, the mechanism of missingness is assumed to be MAR, pioneerly introduced by Rubin (1976), meaning that the missingness of input data depends only on the observed values. A computationally flexible Monte Carlo Expectation Conditional Maximization (MCECM) algorithm that incorporates two auxiliary matrices is provided to carry out ML estimation under a general pattern of missingness.

To pose the MST model in a Bayesian framework, a DA scheme for Bayesian sampling is developed to create multiple imputations of missing data. The priors are chosen to be weakly informative to avoid improper posterior distributions. Using a fully Bayesian approach to implement this model, the posterior and predictive inferences can be accurately extracted from a large converged Monte Carlo dependent samples.

In Sections 2.2 and 2.3, we describe the MST distribution and multivariate truncated t distribution with a simple pseudo random number generator via the Gibbs sampler, respectively. In Section 2.4, we describe a feasible MCECM procedure to conduct estimation and imputation of MST models from incomplete data. A simple way of calculating information-based standard errors of estimates is also presented.

In Section 2.5, an efficient DA scheme is employed to carry out a fully Bayesian inference as well as multiple imputations of missing values. In Section 2.6, the proposed methodologies are illustrated through two real examples with complete and synthetic missing values. Some concluding remarks are briefly summarized in Section 2.7.

2.2. The multivariate skew t distribution

For notational simplicity, we denote $t_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ to be the p-dimensional MVT distribution with location vector $\boldsymbol{\mu}$, scale covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν and $T_p(\cdot \mid \boldsymbol{\Sigma}; \nu)$ is the cdf of $t_p(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$.

A p-dimensional random vector $\mathbf{Y} = (Y_1, \dots, Y_p)^\top$ is said to have the p-dimensional MST distribution with a $p \times 1$ location vector $\boldsymbol{\xi}$, a $p \times p$ positive definite scale covariance matrix $\boldsymbol{\Sigma}$, a $p \times p$ skewness matrix $\boldsymbol{\Lambda} = \mathrm{Diag}\{\lambda_1, \dots, \lambda_p\}$, and degrees of freedom ν if its joint pdf is given by

$$f(\mathbf{Y} \mid \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu) = 2^p t_p(\mathbf{Y} \mid \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) T_p\left(\mathbf{q}\sqrt{\frac{\nu+p}{U+\nu}} \,\middle|\, \boldsymbol{\Delta}; \nu+p\right), \tag{2.1}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2$, $\boldsymbol{\Delta} = (\mathbf{I}_p + \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1} = \mathbf{I}_p - \boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}$, $\mathbf{q} = \boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \boldsymbol{\xi})$ and $U = (\mathbf{Y} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{Y} - \boldsymbol{\xi})$. If the p-dimensional random vector $\mathbf{Y}$ has the pdf in Eq. (2.1), it will be denoted by $\mathbf{Y} \sim ST_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$.

A three-level hierarchical version of the MST model can be expressed by:

$$\begin{aligned} \mathbf{Y} \mid (\boldsymbol{\gamma}, \tau) &\sim N_p(\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}, \boldsymbol{\Sigma}/\tau), \\ \boldsymbol{\gamma} \mid \tau &\sim HN_p(\mathbf{0}, \mathbf{I}_p/\tau), \\ \tau &\sim \Gamma(\nu/2, \nu/2) \quad . \end{aligned} \tag{2.2}$$

It follows from (2.2), the marginal distribution of $\mathbf{Y}$ is $ST_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$. By Sahu et al. (2003), we have the following Lemma.

Lemma 2.1 Let $\mathbf{Y} \sim ST_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$. Then

(a)

$$E(\mathbf{Y}) = \boldsymbol{\xi} + \sqrt{\frac{\nu}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \boldsymbol{\lambda},$$

(b)

$$\mathrm{cov}(\mathbf{Y}) = \frac{\nu}{\nu-2}\left(\boldsymbol{\Sigma} + \left(1 - \frac{2}{\pi}\right)\boldsymbol{\Lambda}^2\right) + \frac{2}{\pi}\left(\frac{\nu}{\nu-2} - \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right)^2 \frac{\nu}{2}\right) \boldsymbol{\lambda}\boldsymbol{\lambda}^\top,$$

where $\boldsymbol{\lambda} = \mathrm{Diag}(\boldsymbol{\Lambda})$.

From Lemma 2.1, the mean exists when $\nu > 1$ and covariance matrix exists when $\nu > 2$. Note that there is an error in the expression of $\mathrm{cov}(\mathbf{Y})$ in Sahu et al. (2003, pp. 137) that $\boldsymbol{\Lambda}^2$ is replaced by $\boldsymbol{\lambda}\boldsymbol{\lambda}^\top$ in Lemma 2.1.

2.3. The multivariate truncated t distribution

In this subsection, we briefly describe properties of the multivariate truncated t distribution whose truncations are left-positioned at arbitrary points. Further, we provide a simple means to generate random samples from this distribution.

Let $Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ denote a p-variate truncated t distribution for $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ lying within a truncated hyperplane region $\mathbb{A} = \{\mathbf{x} = (x_1, \ldots, x_p)^\top | x_1 > a_1, \ldots, x_p > a_p\}$ and use the notation $\prod_{i=1}^p \int_{a_i}^\infty = \int_{a_1}^\infty \cdots \int_{a_p}^\infty$ for the abbreviation of multiple integrals. Specifically, we say a p-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_p)^\top \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, if its density is given by

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A}) = \frac{1}{\alpha} t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) I_{\mathbb{A}}(\mathbf{x}), \tag{2.3}$$

where $\alpha = \prod_{i=1}^p \int_{a_i}^\infty t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{x}$ is the normalizing constant and $I_{\mathbb{A}}(\mathbf{x})$ is the indicator function whose value equals one if $\mathbf{x} \in \mathbb{A}$ and zero elsewhere.

Now, we present a flexible Gibbs procedure to draw random samples from (2.3).

Let $\mathbf{L}$ be the Cholesky factor such that $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top$. If $\mathbf{Z} = \mathbf{L}^{-1}\mathbf{X}$, then

$$\mathbf{Z} \sim Tt_p(\boldsymbol{\mu}^*, \mathbf{I}_p, \nu; \mathbb{B}), \quad \mathbb{B} = \{\mathbf{Lz} \geq \mathbf{a}\},$$

where $\boldsymbol{\mu}^* = \mathbf{L}^{-1}\boldsymbol{\mu} = (\mu_1^*, \ldots, \mu_p^*)^\top$. Based on the property of the MVT distribution concerning its conditional distribution, the full conditionals required for the Gibbs sampler are

$$Z_r|(\mathbf{Z}_{-r} = \mathbf{z}_{-r}) \sim Tt_1\Big(\mu_r^*, \frac{\nu + \delta_{-r}}{\nu + p - 1}, \nu + p - 1; \mathbb{B}_r\Big), \quad r = 1, \ldots, p, \tag{2.4}$$

and each which follows a univariate truncated t distribution with truncation $\mathbb{B}_r$. Here $\delta_{-r} = (\mathbf{z}_{-r} - \boldsymbol{\mu}^*_{-r})^\top(\mathbf{z}_{-r} - \boldsymbol{\mu}^*_{-r})$, $\mathbb{B}_r = \{Z_r \in \mathbb{R} \mid \mathbf{Lz} \geq \mathbf{a}\}$ and $\mathbf{z}_{-r}$ is a subvector of $\mathbf{z}$ by disposing of the rth entry. To generate random variates from (2.4), it can be easily done by using the 'R' programs provided by Nadarajah and Kotz (2007).

In summary, the Gibbs sampler proceeds as follows:

1. Get initial values $\mathbf{X}^{(0)}$ from the support region $\mathbb{A}$. Set $\mathbf{Z}^{(0)} = \mathbf{L}^{-1}\mathbf{X}^{(0)}$ and $k = 1$.

2. Generate $Z_r^{(k)}$, $r = 1, \ldots, p$, from $f(z_r|z_1^{(k)}, \ldots, z_{r-1}^{(k)}, z_{r+1}^{(k-1)}, \ldots, z_p^{(k-1)})$, which is a truncated t distribution as in (2.4).

3. Return the values $\mathbf{X}^{(k)} = \mathbf{L}\mathbf{Z}^{(k)}$. Set $k = k + 1$.

4. Repeat Steps 2 and 3.

2.4. The multivariate skew t distribution with missing information

2.4.1. Definition and some properties

Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ be a random sample of size n taken from $ST_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$ with $n > p$. According to (2.2), a three-level hierarchical representation of MST models can be expressed as

$$\begin{aligned}\mathbf{Y}_j \mid (\boldsymbol{\gamma}_j, \tau_j) &\sim N_p(\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}/\tau_j), \\ \boldsymbol{\gamma}_j \mid \tau_j &\sim HN_p(\mathbf{0}, \mathbf{I}_p/\tau_j), \\ \tau_j &\sim \Gamma(\nu/2, \nu/2) \quad (j = 1, \dots, n). \end{aligned} \tag{2.5}$$

To set up estimating equations for multivariate data allowing for missing values, we partition $\mathbf{Y}_j$ into two components $(\mathbf{Y}_j^{\mathrm{o}\top}, \mathbf{Y}_j^{\mathrm{m}\top})^\top$, where $\mathbf{Y}_j^{\mathrm{o}}$ $(p_j^{\mathrm{o}} \times 1)$ and $\mathbf{Y}_j^{\mathrm{m}}$ $((p - p_j^{\mathrm{o}}) \times 1)$ denote the observed and missing components of $\mathbf{Y}_j$, respectively. To facilitate computation, we introduce two types of the binary indicator matrices, denoted by $\mathbf{O}_j$ $(p_j^{\mathrm{o}} \times p)$ and $\mathbf{M}_j$ $((p - p_j^{\mathrm{o}}) \times p)$, satisfying $\mathbf{Y}_j^{\mathrm{o}} = \mathbf{O}_j\mathbf{Y}_j$ and $\mathbf{Y}_j^{\mathrm{m}} = \mathbf{M}_j\mathbf{Y}_j$, which can be extracted from a p-dimensional identity matrix $\mathbf{I}_p$ corresponding to row positions of $\mathbf{Y}_j^{\mathrm{o}}$ and $\mathbf{Y}_j^{\mathrm{m}}$ in $\mathbf{Y}_j$, respectively. It can be easily verified that (a) $\mathbf{Y}_j = \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{Y}_j^{\mathrm{m}}$; (b) $\mathbf{O}_j^\top\mathbf{O}_j + \mathbf{M}_j^\top\mathbf{M}_j = \mathbf{I}_p$. Accordingly, some important consequences are summarized in the following theorem.

Theorem 2.1 *Given the specification of (2.5), we have*

(a) *The conditional distribution of $\mathbf{Y}_j^{\mathrm{o}}$ given $\boldsymbol{\gamma}_j$ and τ_j is*

$$\mathbf{Y}_j^{\mathrm{o}} \mid (\boldsymbol{\gamma}_j, \tau_j) \sim N_{p_j^{\mathrm{o}}}(\boldsymbol{\zeta}_j^{\mathrm{o}}, \tau_j^{-1}\boldsymbol{\Sigma}_j^{\mathrm{oo}}),$$

where $\boldsymbol{\zeta}_j^{\mathrm{o}} = \mathbf{O}_j(\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)$ and $\boldsymbol{\Sigma}_j^{\mathrm{oo}} = \mathbf{O}_j\boldsymbol{\Sigma}\mathbf{O}_j^\top$.

(b) *The conditional distribution of $\mathbf{Y}_j^{\mathrm{m}}$ given $\mathbf{Y}_j^{\mathrm{o}}$, $\boldsymbol{\gamma}_j$, and τ_j is*

$$\mathbf{Y}_j^{\mathrm{m}} \mid (\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j) \sim N_{p - p_j^{\mathrm{o}}}(\boldsymbol{\zeta}_j^{\mathrm{m\cdot o}}, \tau_j^{-1}\boldsymbol{\Sigma}_j^{\mathrm{mm\cdot o}}),$$

where $\boldsymbol{\zeta}_j^{\mathrm{m\cdot o}} = \mathbf{M}_j\big(\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j + \boldsymbol{\Sigma}\mathbf{S}_j^{\mathrm{oo}}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\big)$, $\boldsymbol{\Sigma}_j^{\mathrm{mm\cdot o}} = \mathbf{M}_j(\mathbf{I}_p - \boldsymbol{\Sigma}\mathbf{S}_j^{\mathrm{oo}})\boldsymbol{\Sigma}\mathbf{M}_j^\top$ and $\mathbf{S}_j^{\mathrm{oo}} = \mathbf{O}_j^\top(\mathbf{O}_j\boldsymbol{\Sigma}\mathbf{O}_j^\top)^{-1}\mathbf{O}_j$.

(c) *The marginal distribution of* $\mathbf{Y}_j^{\mathrm{o}}$ *is* $ST_{p_j^{\mathrm{o}}}(\boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Sigma}_j^{\mathrm{oo}}, \boldsymbol{\Lambda}_j^{\mathrm{oo}}, \nu)$ *with density*

$$f(\mathbf{Y}_j^{\mathrm{o}}) \;=\; 2^{p_j^{\mathrm{o}}} t_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}, \nu) T_{p_j^{\mathrm{o}}}\Big(\mathbf{q}_j^{\mathrm{o}} \sqrt{\frac{\nu + p_j^{\mathrm{o}}}{\nu + U_j^{\mathrm{o}}}} \mid \boldsymbol{\Delta}_j^{\mathrm{oo}}; \nu + p_j^{\mathrm{o}}\Big),$$

where $\boldsymbol{\xi}_j^{\mathrm{o}} = \mathbf{O}_j \boldsymbol{\xi}$, $\boldsymbol{\Lambda}_j^{\mathrm{oo}} = \mathbf{O}_j \boldsymbol{\Lambda} \mathbf{O}_j^{\top}$, $\boldsymbol{\Omega}_j^{\mathrm{oo}} = \mathbf{O}_j \boldsymbol{\Omega} \mathbf{O}_j^{\top}$, $\mathbf{q}_j^{\mathrm{o}} = \boldsymbol{\Lambda}_j^{\mathrm{oo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} (\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})$, $U_j^{\mathrm{o}} = (\mathbf{Y}_j - \boldsymbol{\xi})^{\top} \mathbf{C}_j^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi})$, $\mathbf{C}_j^{\mathrm{oo}} = \mathbf{O}_j^{\top} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} \mathbf{O}_j$ *and* $\boldsymbol{\Delta}_j^{\mathrm{oo}} = \mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_j^{\mathrm{oo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} \boldsymbol{\Lambda}_j^{\mathrm{oo}}$.

(d) *The posterior distribution of* $\boldsymbol{\gamma}_j$ *given* $\mathbf{Y}_j^{\mathrm{o}}$ *follows a multivariate truncated* t *distribution. That is*

$$\boldsymbol{\gamma}_j \mid \mathbf{Y}_j^{\mathrm{o}} \sim Tt_p(\mathbf{q}_j^{*}, \frac{U_j^{\mathrm{o}} + \nu}{p_j^{\mathrm{o}} + \nu} \boldsymbol{\Delta}_j^{*}, \nu + p_j^{\mathrm{o}};\ \mathbb{R}_{+}^{p}),$$

where $\mathbf{q}_j^{*} = \boldsymbol{\Lambda} \mathbf{C}_j^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi})$ *and* $\boldsymbol{\Delta}_j^{*} = \mathbf{I}_p - \boldsymbol{\Lambda} \mathbf{C}_j^{\mathrm{oo}} \boldsymbol{\Lambda}$.

(e) *The conditional distribution of* τ_j *given* $\mathbf{Y}_j^{\mathrm{o}}$ *and* $\boldsymbol{\gamma}_j$ *is*

$$\tau_j \mid (\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j) \sim \Gamma\Big(\frac{p + p_j^{\mathrm{o}} + \nu}{2}, \frac{(\boldsymbol{\gamma}_j - \mathbf{q}_j^{*})^{\top} \boldsymbol{\Delta}_j^{*^{-1}} (\boldsymbol{\gamma}_j - \mathbf{q}_j^{*}) + U_j^{\mathrm{o}} + \nu}{2}\Big).$$

Proof. The proofs of part (a) and part (b) are straightforward and hence are omitted. The proof of part (c), (d), and (e) are given in Appendix F.

2.4.2. ML estimation via the MCECM procedure

To compute the ML estimates for the parameter vector $\boldsymbol{\theta}$ of MST models with partially observed data, we adopt a simple modification of the MCEM algorithm (Wei and Tanner, 1990), namely the MCECM algorithm. More precisely, it is an extension of the ECM algorithm (Meng and Rubin, 1993) in which the E-step is evaluated by approximating the conditional expectations through observations simulated by Markov chain Monte Carlo (MCMC) methods (Hastings, 1970), while the

M-step is simplified by performing a sequence of conditional maximization (CM) steps.

For notational simplicity, let $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$. Let $\mathbf{Y}^{\mathrm{o}} = (\mathbf{Y}_1^{\mathrm{o}}, \mathbf{Y}_2^{\mathrm{o}}, \cdots, \mathbf{Y}_n^{\mathrm{o}})$ and $\mathbf{Y}^{\mathrm{m}} = (\mathbf{Y}_1^{\mathrm{m}}, \mathbf{Y}_2^{\mathrm{m}}, \cdots, \mathbf{Y}_n^{\mathrm{m}})$ represent the observed portion and missing portion of the data, respectively. The complete data log-likelihood function of $\boldsymbol{\theta}$, ignoring additive constant terms, is

$$
\begin{aligned}
&\ell_c(\boldsymbol{\theta}|\mathbf{Y}^{\mathrm{o}}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\gamma}, \boldsymbol{\tau}) \\
= \;& \frac{n\nu}{2}\log\frac{\nu}{2} - n\log\Gamma\Big(\frac{\nu}{2}\Big) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{\nu}{2}\sum_{j=1}^{n}\tau_j + \Big(\frac{\nu}{2}+p-1\Big)\sum_{j=1}^{n}\log\tau_j \\
&-\frac{1}{2}\sum_{j=1}^{n}\tau_j\boldsymbol{\gamma}_j^{\top}\boldsymbol{\gamma}_j - \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_j\Big), \qquad (2.6)
\end{aligned}
$$

where $\boldsymbol{\Omega}_j = \tau_j(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^{\top}$.

Let $\hat{\tau}_j^{(k)} = E(\tau_j|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})$, $\hat{\kappa}_j^{(k)} = E(\log\tau_j|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})$, $\hat{\boldsymbol{\eta}}_j^{(k)} = E(\tau_j\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})$ and $\hat{\boldsymbol{\Psi}}_j^{(k)} = E(\tau_j\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^{\top}|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})$ be the necessary conditional expectations involved in (2.6). Let $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\xi}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}, \hat{\boldsymbol{\Lambda}}^{(k)}, \hat{\nu}^{(k)})$ denote the estimates of $\boldsymbol{\theta}$ at the kth iteration. Given the observed data $\mathbf{Y}^{\mathrm{o}}$ and the current parameter estimates $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, we can calculate

$$
\begin{aligned}
\boldsymbol{\Omega}_j^{(k)} &= E(\boldsymbol{\Omega}_j \mid \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)}) \\
&= (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}^{(k)} + (\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})\hat{\boldsymbol{\Psi}}_j^{(k)}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})^{\top} \\
&\quad + \hat{\tau}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})^{\top} + (\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})\hat{\boldsymbol{\eta}}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})^{\top} \\
&\quad + (\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})^{\top}, \qquad (2.7)
\end{aligned}
$$

where

$$
\hat{\mathbf{A}}_j^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}^{(k)} \quad \text{and} \quad \hat{\mathbf{b}}_j^{(k)} = \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\xi}}^{(k)}. \quad (2.8)
$$

Proof. The detailed proof is given in Appendix G.

Furthermore, a Monte Carlo estimate of the Q-function can be evaluated as

$$\hat{Q}(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = \frac{1}{M}\sum_{m=1}^{M} E\big(\ell_c(\boldsymbol{\theta} \mid \mathbf{Y}^{\rm o}, \mathbf{Y}^{\rm m}, \hat{\boldsymbol{\gamma}}_{[m]}^{(k)}, \hat{\boldsymbol{\tau}}_{[m]}^{(k)}, \hat{\boldsymbol{\theta}}^{(k)}) \mid \mathbf{Y}^{\rm o}, \hat{\boldsymbol{\theta}}^{(k)}\big), \tag{2.9}$$

where $\hat{\boldsymbol{\gamma}}_{[m]}^{*(k)} = \{\hat{\boldsymbol{\gamma}}_{j,m}^{*(k)}, j = 1, \ldots, n\}$ and $\hat{\boldsymbol{\tau}}_{[m]}^{*(k)} = \{\hat{\tau}_{j,m}^{*(k)}, j = 1, \ldots, n\}$ for $m = 1, \ldots, M$, are a set of independent random samples generated from each $f(\boldsymbol{\gamma}_j, \tau_j | \mathbf{Y}_j^{\rm o})$ given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$. The exact sampling of $\boldsymbol{\gamma}_j$ and τ_j can be conveniently implemented through the following generators:

$$\hat{\boldsymbol{\gamma}}_{j,m}^{(k+1)} \mid \mathbf{Y}_j^{\rm o} \sim Tt_p\Big(\hat{\mathbf{q}}_j^{*(k)}, \frac{\hat{U}_j^{{\rm o}(k)} + \hat{\nu}^{(k)}}{p_j^{\rm o} + \hat{\nu}^{(k)}} \hat{\boldsymbol{\Delta}}_j^{*(k)}, \hat{\nu}^{(k)} + p_j^{\rm o};\ \mathbb{R}_+^p\Big),$$

and

$$\hat{\tau}_{j,m}^{(k+1)} \mid (\hat{\boldsymbol{\gamma}}_{j,m}^{(k+1)}, \mathbf{Y}_j^{\rm o}) \sim \Gamma\Big(\frac{\hat{\nu}^{(k)} + p + p_j^{\rm o}}{2}, \frac{(\hat{\boldsymbol{\gamma}}_{j,m}^{(k+1)} - \hat{\mathbf{q}}_j^{*(k)})^\top \hat{\boldsymbol{\Delta}}_j^{*(k)^{-1}} (\hat{\boldsymbol{\gamma}}_{j,m}^{(k+1)} - \hat{\mathbf{q}}_j^{*(k)}) + \hat{U}_j^{{\rm o}(k)} + \hat{\nu}^{(k)}}{2}\Big),$$

where $\hat{\mathbf{q}}_j^{*(k)} = \hat{\boldsymbol{\Lambda}}^{(k)} \hat{\mathbf{C}}_j^{{\rm oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}^{(k)})$, $\hat{\boldsymbol{\Delta}}_j^{*(k)} = \mathbf{I}_p - \hat{\boldsymbol{\Lambda}}^{(k)} \hat{\mathbf{C}}_j^{{\rm oo}(k)} \hat{\boldsymbol{\Lambda}}^{(k)}$, $\hat{U}_j^{{\rm o}(k)} = (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}^{(k)})^\top \hat{\mathbf{C}}_j^{{\rm oo}(k)} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}^{(k)})$ and $\hat{\mathbf{C}}_j^{{\rm oo}(k)} = \mathbf{O}_j^\top \hat{\boldsymbol{\Omega}}_j^{{\rm oo}(k)^{-1}} \mathbf{O}_j$ with $\hat{\boldsymbol{\Omega}}_j^{{\rm oo}(k)} = \mathbf{O}_j(\hat{\boldsymbol{\Sigma}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)2})\mathbf{O}_j^\top$. Therefore, the conditional expectations defined above can be readily approximated as

$$\begin{aligned} \hat{\tau}_j^{(k)} &\simeq M^{-1}\sum_{m=1}^{M} \hat{\tau}_{j,m}^{(k)}, \quad \hat{\kappa}_j^{(k)} \simeq M^{-1}\sum_{m=1}^{M} \log \hat{\tau}_{j,m}^{(k)}, \\ \hat{\boldsymbol{\eta}}_j^{(k)} &\simeq M^{-1}\sum_{m=1}^{M} \hat{\tau}_{j,m}^{(k)} \hat{\boldsymbol{\gamma}}_{j,m}^{(k)}, \quad \hat{\boldsymbol{\Psi}}_j^{(k)} \simeq M^{-1}\sum_{m=1}^{M} \hat{\tau}_{j,m}^{(k)} \hat{\boldsymbol{\gamma}}_{j,m}^{(k)} \hat{\boldsymbol{\gamma}}_{j,m}^{(k)^\top}. \end{aligned} \tag{2.10}$$

Formally, the MCECM algorithm can be implemented as follows:

MCE-step: Given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, compute Monte Carlo expectations $\hat{\tau}_j^{(k)}$, $\hat{\kappa}_j^{(k)}$, $\hat{\boldsymbol{\eta}}_j^{(k)}$ and $\hat{\boldsymbol{\Psi}}_j^{(k)}$ by using (2.10) for $j = 1, \ldots, n$.

CM-steps:

CM-Step 1. Fix $\mathbf{\Lambda} = \hat{\mathbf{\Lambda}}^{(k)}$, update $\hat{\boldsymbol{\xi}}^{(k)}$ by maximizing (2.9) over $\boldsymbol{\xi}$, which leads to

$$\hat{\boldsymbol{\xi}}^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{\tau}_j^{(k)} \hat{\mathbf{b}}_j^{(k)} - \sum_{j=1}^{n} \hat{\mathbf{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\text{oo}(k)} \hat{\mathbf{\Lambda}}^{(k)} \hat{\boldsymbol{\eta}}_j^{(k)}}{\sum_{j=1}^{n} \hat{\tau}_j^{(k)}},$$

where $\hat{\mathbf{S}}_j^{\text{oo}(k)} = \mathbf{O}_j^\top (\mathbf{O}_j \hat{\mathbf{\Sigma}}^{(k)} \mathbf{O}_j^\top)^{-1} \mathbf{O}_j$.

CM-Step 2. Fix $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^{(k+1)}$ and $\mathbf{\Lambda} = \hat{\mathbf{\Lambda}}^{(k)}$, update $\hat{\mathbf{\Sigma}}^{(k)}$ by maximizing (2.9) over $\mathbf{\Sigma}$, which gives

$$\hat{\mathbf{\Sigma}}^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{\mathbf{\Omega}}_j^{(k+1/2)}}{n},$$

where $\hat{\mathbf{\Omega}}_j^{(k+1/2)}$ is $\mathbf{\Omega}_j^{(k)}$ in (2.7) with $\boldsymbol{\xi}$ and $\mathbf{\Lambda}$ replaced by $\hat{\boldsymbol{\xi}}^{(k+1)}$ and $\hat{\mathbf{\Lambda}}^{(k)}$, respectively.

CM-Step 3. Fix $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^{(k+1)}$ and $\mathbf{\Sigma} = \hat{\mathbf{\Sigma}}^{(k+1)}$, updating $\hat{\boldsymbol{\lambda}}^{(k)}$ by maximizing (2.9) over $\boldsymbol{\lambda}$ yields

$$\hat{\boldsymbol{\lambda}}^{(k+1)} = \left(\hat{\mathbf{\Sigma}}^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{\mathbf{\Psi}}_j^{(k)}\right)^{-1} \left(\hat{\mathbf{\Sigma}}^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \left((\hat{\mathbf{b}}_j^{(k)} - \hat{\boldsymbol{\xi}}^{(k+1)}) \hat{\boldsymbol{\eta}}_j^{(k)\top} + \hat{\mathbf{A}}_j^{(k)} \hat{\mathbf{\Psi}}_j^{(k)} \right)^\top \right) \mathbf{1}_p,$$

where $\hat{\mathbf{A}}_j^{(k)}$ and $\hat{\mathbf{b}}_j^{(k)}$ are defined in (2.8). It follows immediately that $\hat{\mathbf{\Lambda}}^{(k+1)} = \text{Diag}(\hat{\boldsymbol{\lambda}}^{(k+1)})$.

CM-Step 4. Obtain $\hat{\nu}^{(k+1)}$ as the solution of the following equation

$$\log\left(\frac{\nu}{2}\right) + 1 - \text{DG}\left(\frac{\nu}{2}\right) + \frac{1}{n}\sum_{j=1}^{n}\left(\hat{\kappa}_j^{(k)} - \hat{\tau}_j^{(k)}\right) = 0,$$

where $\text{DG}(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function.

The detailed of the CM-steps are shown in Appendix H.

Iterations of the above MCE- and CM- steps are alternated repeatedly until a suitable convergence rule is satisfied, e.g., the relative difference in successive values of the log-likelihood is less than a tolerance value, say 10^{-12}. As the log-likelihood function tends to have multiple modes, the algorithm needs to be initialized with a variety of several starting values. For the choice of initial values, Lin (2009) proposed a simple way of automatically generating a selection of initial values. The global optimum solution is obtained by comparing their relative converged log-likelihood values. Under suitable regularity conditions (Chan and Ledolter, 1995), the algorithm converges to local optimum of the likelihood function. The resulting ML estimates are denoted by $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\lambda}}, \hat{\nu})$. It follows from Theorem 2.1(b) that the ML predictor for the missing component $\mathbf{Y}_j^{\mathrm{m}}$ can be determined as

$$\hat{\mathbf{Y}}_j^{\mathrm{m}} = \mathbf{M}_j\big(\hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\gamma}}_j + \hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\mathrm{oo}}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\gamma}}_j)\big), \tag{2.11}$$

where $\hat{\boldsymbol{\gamma}}_j = E(\boldsymbol{\gamma}_j|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\theta}})$, which can be approximated by Monte Carlo average of samples simulated from $Tt_p\big(\hat{\mathbf{q}}_j^*, \dfrac{\hat{U}_j^{\mathrm{o}} + \hat{\nu}}{p_j^{\mathrm{o}} + \hat{\nu}}\hat{\boldsymbol{\Delta}}_j^*, \hat{\nu} + p_j^{\mathrm{o}};\ \mathbb{R}_+^p\big)$.

2.4.3. Standard errors estimates

Under certain regularity conditions, we apply the information-based method of Meilijson (1989) to compute the asymptotic covariance of the ML estimates. Let $\ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \tau_j)$ be the complete data log-likelihood formed from the single observation $\mathbf{Y}_j$. The individual score is defined as

$$\mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\theta}) = E\Big(\frac{\partial \ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \tau_j)}{\partial \boldsymbol{\theta}} \mid \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}\Big).$$

The empirical information matrix, according to Meilijson's formula, is defined as

$$\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{Y}^{\mathrm{o}}) = \sum_{j=1}^{n} \mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\theta})\mathbf{u}^{\top}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\theta}) - n^{-1}\mathbf{U}(\mathbf{Y}^{\mathrm{o}}|\boldsymbol{\theta})\mathbf{U}^{\top}(\mathbf{Y}^{\mathrm{o}}|\boldsymbol{\theta}), \tag{2.12}$$

where $\mathbf{U}(\mathbf{Y}^{\mathrm{o}}|\boldsymbol{\theta}) = \sum_{j=1}^{n} \mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\theta})$.

Let vech$(\cdot)$ be the matrix operator which stacks only the distinct elements of a symmetric matrix into a single vector. Substituting the ML estimates $\hat{\boldsymbol{\theta}}$ into $\boldsymbol{\theta}$, (2.12) is reduced to

$$\mathbf{I}_e(\hat{\boldsymbol{\theta}} \mid \mathbf{Y}^{\mathrm{o}}) = \sum_{j=1}^{n} \hat{\mathbf{u}}_j^{\mathrm{o}} \hat{\mathbf{u}}_j^{\mathrm{o}\top}, \tag{2.13}$$

where $\hat{\mathbf{u}}_j^{\mathrm{o}} = \mathbf{u}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\theta}}) = (\hat{\mathbf{u}}_{j,\boldsymbol{\xi}}^{\mathrm{o}\top}, \hat{\mathbf{u}}_{j,\boldsymbol{\sigma}}^{\mathrm{o}\top}, \hat{\mathbf{u}}_{j,\boldsymbol{\lambda}}^{\mathrm{o}\top}, \hat{u}_{j,\nu}^{\mathrm{o}})^\top$ is a $(p^2+5p+2)/2 \times 1$ score vector and $\boldsymbol{\sigma} = \mathrm{vech}(\boldsymbol{\Sigma})$. Then, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ can be approximated by the inverse of (2.13). Expressions for the elements of $\hat{\mathbf{u}}_j^{\mathrm{o}}$ can be obtained by standard matrix differentiation. Technical derivations are given as follows:

$$\begin{aligned}
\hat{\mathbf{u}}_{j,\boldsymbol{\xi}}^{\mathrm{o}} &= \hat{\tau}_j \hat{\mathbf{S}}_j^{\mathrm{oo}}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}) - \hat{\mathbf{S}}_j^{\mathrm{oo}} \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\eta}}_j, \quad \hat{\mathbf{u}}_{j,\boldsymbol{\sigma}}^{\mathrm{o}} = \mathrm{vech}\left(\hat{\mathbf{C}}_j - \frac{1}{2}\mathrm{Diag}(\hat{\mathbf{C}}_j)\right), \\
\hat{\mathbf{u}}_{j,\boldsymbol{\lambda}}^{\mathrm{o}} &= \mathrm{Diag}\left(\hat{\mathbf{S}}_j^{\mathrm{oo}}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}})\hat{\boldsymbol{\eta}}_j^\top - \hat{\mathbf{S}}_j^{\mathrm{oo}} \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Psi}}_j\right), \\
\hat{u}_{j,\nu}^{\mathrm{o}} &= \frac{1}{2}\left(\log\left(\frac{\hat{\nu}}{2}\right) + 1 - \mathrm{DG}\left(\frac{\hat{\nu}}{2}\right) + \hat{\kappa}_j - \hat{\tau}_j\right),
\end{aligned}$$

where $\hat{\mathbf{C}}_j = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}_j\hat{\boldsymbol{\Sigma}}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}$ with

$$\begin{aligned}
\hat{\mathbf{R}}_j &= \hat{\mathbf{D}}_j\hat{\boldsymbol{\Sigma}} + (\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})\hat{\boldsymbol{\Psi}}_j(\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})^\top + (\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})\hat{\boldsymbol{\eta}}_j(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\mathrm{oo}}\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})^\top \\
&\quad + (\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\mathrm{oo}}\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})\hat{\boldsymbol{\eta}}_j^\top(\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})^\top \\
&\quad + \hat{\tau}_j(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\mathrm{oo}}\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\mathrm{oo}}\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})^\top
\end{aligned}$$

and $\hat{\mathbf{D}}_j = \mathbf{I}_p - \hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\mathrm{oo}}$.

The detailed proof is shown in Appendix I.

2.5. A fully Bayesian approach

Due to the growing advances of modern computing technology, the Bayesian method is frequently considered as an alternative way to deal with missing data problems. Tanner and Wong (1987) proposed the DA algorithm, which has been shown to be an effective procedure for multiple imputations of missing data. In this section, we construct an efficient DA algorithm that combines the latent variables $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$, and unobserved data $\mathbf{Y}^{\mathrm{m}}$ to simulate the posterior distribution of $\boldsymbol{\theta}$.

The DA algorithm consists of the imputation step (I-step) and the posterior step (P-step). At the kth iteration, the I-step is defined by drawing imputations of $\boldsymbol{\gamma}_j^{(k)}$, $\tau_j^{(k)}$, and $\mathbf{Y}_j^{\mathrm{m}(k)}$ from the predictive distributions $p(\boldsymbol{\gamma}_j|\mathbf{Y}^{\mathrm{o}}, \boldsymbol{\theta}^{(k)})$, $p(\tau_j|\boldsymbol{\gamma}_j, \mathbf{Y}^{\mathrm{o}}, \boldsymbol{\theta}^{(k)})$, and $p(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\gamma}_j, \mathbf{Y}^{\mathrm{o}}, \tau_j, \boldsymbol{\theta}^{(k)})$, respectively, for $j = 1, \ldots, n$, and the P-step refers to generating $\boldsymbol{\theta}^{(k+1)}$ from $p(\boldsymbol{\theta}|\mathbf{Y}^{\mathrm{o}}, \mathbf{Y}^{\mathrm{m}(k+1)}, \boldsymbol{\tau}^{(k+1)}, \boldsymbol{\gamma}^{(k+1)})$. If iterations are performed by a sufficiently long time, then the simulations $\boldsymbol{\gamma}_j^{(k)}$, $\tau_j^{(k)}$, $\mathbf{Y}_j^{\mathrm{m}(k)}$, and $\boldsymbol{\theta}^{(k)}$ are distributed according to $p(\boldsymbol{\gamma}_j|\mathbf{Y}^{\mathrm{o}})$, $p(\tau_j|\mathbf{Y}^{\mathrm{o}})$, $p(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}^{\mathrm{o}})$, and $p(\boldsymbol{\theta}|\mathbf{Y}^{\mathrm{o}})$, respectively.

To avoid yielding the improper posterior distributions, we need to choose the proper prior distributions for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$. Let $\mathcal{W}_p(a, \boldsymbol{\Sigma})$ denote the p-dimensional Wishart distribution with degrees of freedom a and $p \times p$ scale matrix $\boldsymbol{\Sigma}$. When the prior information is not available, a convenient strategy of avoiding improper posterior distribution is to use diffuse proper priors. The prior distributions adopted are as follows:

$$
\begin{aligned}
\boldsymbol{\xi} &\sim N_p(\mathbf{a}, \boldsymbol{\kappa}^{-1}), \\
\boldsymbol{\Sigma}^{-1}|\mathbf{B} &\sim \mathcal{W}_p(2\alpha, (2\mathbf{B})^{-1}), \\
\mathbf{B} &\sim \mathcal{W}_p(2\gamma, (2\mathbf{H})^{-1}), \\
\boldsymbol{\lambda} &\sim N_p(\mathbf{0}, \boldsymbol{\Gamma}), \\
\log\left(\frac{1}{\nu}\right) &\sim U(-10, 10),
\end{aligned}
$$

where $(\mathbf{a}, \boldsymbol{\kappa}, \alpha, \gamma, \mathbf{H}, \boldsymbol{\Gamma})$ are fixed as appropriate quantities to yield the proper posterior distributions. Thus, the joint prior density function of $\boldsymbol{\theta}$ and $\mathbf{B}$ is

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{B}) \propto\ & |\mathbf{B}|^{\alpha+(2\gamma-p-1)/2} |\boldsymbol{\Sigma}|^{-(2\alpha-p-1)/2} \exp\Big\{-\frac{1}{2}(\boldsymbol{\xi}-\mathbf{a})^\top \boldsymbol{\kappa}(\boldsymbol{\xi}-\mathbf{a})\Big\} \\ & \times \exp\left\{-\mathrm{tr}\big((\boldsymbol{\Sigma}^{-1}+\mathbf{H})\mathbf{B}\big) - \frac{1}{2}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda}\right\} J_\nu, \qquad (2.14) \end{aligned}$$

where $J_\nu = \nu^{-1}$ $(0 < \nu < \infty)$ is the Jacobian of transforming $\log(1/\nu)$ to ν.

Multiplying (2.14) with the complete data likelihood function, the joint posterior density can be obtained by

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{B}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\tau}, \boldsymbol{\gamma} | \mathbf{Y}^{\mathrm{o}}) \propto\ & \pi(\boldsymbol{\theta}, \mathbf{B}) \prod_{j=1}^{n} f(\mathbf{Y}_j^{\mathrm{m}} | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) f(\tau_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) \\ & \times f(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}) f(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\theta}). \qquad (2.15) \end{aligned}$$

To implement the DA algorithm, we are now in a position to present the full conditional posterior densities.

Theorem 2.2 *The full conditional posteriors of $\boldsymbol{\theta}$, $\mathbf{B}$, $\boldsymbol{\gamma}_j$, τ_j and $\mathbf{Y}_j^{\mathrm{m}}$ are as follows (the symbol "$|\cdots$" denotes conditioning on all other variables):*

$$\begin{aligned} p(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}) &\sim Tt_p\Big(\mathbf{q}_j^*, \frac{\nu + U_j^{\mathrm{o}}}{\nu + p_j^{\mathrm{o}}}\boldsymbol{\Delta}_j^*, \nu + p_j^{\mathrm{o}};\ \mathbb{R}_+^p\Big), \\ p(\tau_j | \boldsymbol{\gamma}_j, \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}) &\sim \Gamma\Big(\frac{p + p_j^{\mathrm{o}} + \nu}{2}, \frac{(\boldsymbol{\gamma}_j - \mathbf{q}_j^*)^\top \boldsymbol{\Delta}_j^{*^{-1}} (\boldsymbol{\gamma}_j - \mathbf{q}_j^*) + U_j^{\mathrm{o}} + \nu}{2}\Big), \\ p(\mathbf{Y}_j^{\mathrm{m}} | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \boldsymbol{\theta}) &\sim N_{p-p_j^{\mathrm{o}}}(\boldsymbol{\zeta}_j^{\mathrm{m\cdot o}}, \tau_j^{-1}\boldsymbol{\Sigma}_j^{\mathrm{mm\cdot o}}), \\ p(\boldsymbol{\xi} | \cdots) &\sim N_p(\boldsymbol{\mu}^*, \boldsymbol{\kappa}^*), \\ p(\mathbf{B} | \cdots) &\sim \mathcal{W}_p(2\gamma^*, (2\mathbf{H}^*)^{-1}), \\ p(\boldsymbol{\Sigma}^{-1} | \cdots) &\sim \mathcal{W}_p(\alpha^*, \mathbf{B}^{*-1}), \\ p(\boldsymbol{\lambda} | \cdots) &\sim N_p(\boldsymbol{\delta}^*, \boldsymbol{\Gamma}^*), \end{aligned}$$

where

$$
\begin{aligned}
\boldsymbol{\kappa}^* &= (\sum_{j=1}^{n} \tau_j \boldsymbol{\Sigma}^{-1} + \boldsymbol{\kappa})^{-1}, \quad \boldsymbol{\mu}^* = \boldsymbol{\kappa}^* \Big(\boldsymbol{\Sigma}^{-1} (\sum_{j=1}^{n} \tau_j (\mathbf{Y}_j - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)) + \boldsymbol{\kappa}\mathbf{a} \Big), && (2.16)\\
\gamma^* &= \alpha + \gamma, \quad \mathbf{H}^* = \mathbf{H} + \boldsymbol{\Sigma}^{-1}, && (2.17)\\
\alpha^* &= 2\alpha + n, \quad \mathbf{B}^* = 2\mathbf{B} + \sum_{j=1}^{n} \tau_j (\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top, && (2.18)\\
\boldsymbol{\Gamma}^* &= \Big(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Sigma}^{-1} \odot \sum_{j=1}^{n} \tau_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^\top \Big)^{-1}, \boldsymbol{\delta}^* = \boldsymbol{\Gamma}^* \Big(\boldsymbol{\Sigma}^{-1} \odot \sum_{j=1}^{n} \tau_j \boldsymbol{\gamma}_j (\mathbf{Y}_j - \boldsymbol{\xi})^\top \Big) \mathbf{1}_p \\
&. && (2.19)
\end{aligned}
$$

The full conditional distribution of ν *is*

$$
p(\nu|\cdots) \propto \Big(\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \Big)^n \Big(\prod_{j=1}^{n} \tau_j^{\nu/2} \Big) \exp\Big(-\frac{\nu}{2} \sum_{j=1}^{n} \tau_j \Big) J_\nu, \qquad (2.20)
$$

which is not of a standard form.

Proof: The detailed proof is shown in Appendix J.

In the simulation process, samples for $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$, $\mathbf{Y}^m$, $\mathbf{B}$, and $\boldsymbol{\theta}$ are alternately generated. The DA algorithm using the "M-H within Gibbs" sampler (Chib and Greenberg, 1995) can be implemented as follows:

I-Step:

1. Generate $\boldsymbol{\gamma}_j$ from $Tt_p\Big(\mathbf{q}_j^*, \frac{\nu + U_j^{\mathrm{o}}}{\nu + p_j^{\mathrm{o}}} \boldsymbol{\Delta}_j^*, \nu + p_j^{\mathrm{o}};\ \mathbb{R}_+^p\Big)$, where $\mathbf{q}_j^*$, U_j^{o} and $\boldsymbol{\Delta}_j^*$ are given in Theorem 2.1.

2. Generate τ_j from

$$
\Gamma\Big(\frac{p + p_j^{\mathrm{o}} + \nu}{2}, \frac{(\boldsymbol{\gamma}_j - \mathbf{q}_j^*)^\top \boldsymbol{\Delta}_j^{*^{-1}} (\boldsymbol{\gamma}_j - \mathbf{q}_j^*) + U_j^{\mathrm{o}} + \nu}{2} \Big).
$$

3. Generate $\mathbf{Y}_j^{\mathrm{m}}$ from $N_{p-p_j^{\mathrm{o}}}(\boldsymbol{\zeta}_j^{\mathrm{m\cdot o}}, \tau_j^{-1} \boldsymbol{\Sigma}_j^{\mathrm{mm\cdot o}})$, where $\boldsymbol{\zeta}_j^{\mathrm{m\cdot o}}$ and $\boldsymbol{\Sigma}_j^{\mathrm{mm\cdot o}}$ are as in Theorem 2.1.

P-Step:

1. Generate $\boldsymbol{\xi}$ from $N_p(\boldsymbol{\mu}^*, \boldsymbol{\kappa}^*)$, where $\boldsymbol{\mu}^*$ and $\boldsymbol{\kappa}^*$ are given in (2.16).

2. Generate $\mathbf{B}$ from $\mathcal{W}_p(2\gamma^*, (2\mathbf{H}^*)^{-1})$, where γ^* and $\mathbf{H}^*$ are given in (2.17).

3. Generate $\boldsymbol{\Sigma}^{-1}$ from $\mathcal{W}_p(\alpha^*, \mathbf{B}^{*-1})$, where α^* and $\mathbf{B}^*$ are given in (2.18).

4. Generate $\boldsymbol{\lambda}$ from $N_p(\boldsymbol{\delta}^*, \boldsymbol{\Gamma}^*)$, where $\boldsymbol{\delta}^*$ and $\boldsymbol{\Gamma}^*$ are given in (2.19).

5. Generate ν from (2.20) via the M-H algorithm.

To elaborate on the P-Step 5 of the above algorithm, we first transform ν to $\nu^* = \log(1/\nu)$ and then apply the M-H algorithm to the function $g(\nu^*|\cdots) = p(\nu^*|\cdots)e^{-\nu^*}$. The *candidate jumping* kernel is chosen as $N(\nu^{*^{(k)}}, \sigma_{\nu^*}^{2^{(k)}})$, where the $\sigma_{\nu^*}^{2^{(k)}}$ is taken by $\nu^{(k)^{-2}} I^{-1}(\nu^{(k)})$. Denote by $\psi(x) = d^2/dx^2 \log \Gamma(x)$ the trigamma function. Here

$$I(\nu) = \frac{1}{4}\sum_{i=1}^{n}\left[\psi\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu + p_j^{\mathrm{o}}}{2}\right) - \frac{2(\nu+2)}{\nu(\nu+2+p_j^{\mathrm{o}})} - \frac{2}{\nu} + \frac{4}{\nu + p_j^{\mathrm{o}}}\right]$$

is the Fisher information of ν corresponding to the MVT distribution.

Edwards et al. (1963) proposed the "Principle of Stable Estimation" from the Bayesian perspective. By virtue of this point, we need to specify the parameters, $(\mathbf{a}, \boldsymbol{\kappa}, \alpha, \gamma, \mathbf{H}, \boldsymbol{\Gamma})$, so as to be insensitive to change of the prior. In this paper, we choose $\mathbf{a}$ to be the sample mean vector of the observed data and $\boldsymbol{\kappa}^{-1} = (1-r)^{-1}\mathrm{Diag}\{R_1^2, \ldots, R_p^2\}$, where R_i is the range of the observed values for the ith variable and r is the missing rate in the sense of the percentage of missing values of the data, which is used to adjust the flatness. The specification of the hyperparameters makes a weak prior information for $\boldsymbol{\xi}$. As a generalization of Richardson and Green (1997), we choose $\alpha = p+1$, $\gamma = (p+1)/10$, $\mathbf{H} = 10\boldsymbol{\kappa}$, and $\boldsymbol{\Gamma}$ is taken as a diagonal matrix with relative large variances, say 10^4 for each variable.

Given a set of converged Monte Carlo DA samples $\boldsymbol{\theta}^{(\ell)}$ ($\ell = 1, \ldots, L$), using the Rao-Blackwellization (Gelfand and Smith, 1990), the Bayesian predictor for the missing component $\mathbf{Y}_j^{\mathrm{m}}$ is given by

$$\begin{aligned}\tilde{\mathbf{Y}}_j^{\mathrm{m}} &= \frac{1}{L}\sum_{\ell=1}^{L} E(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}^{\mathrm{o}}, \boldsymbol{\theta}^{(\ell)}) \\ &= \mathbf{M}_j \frac{1}{L}\sum_{\ell=1}^{L}\Big(\boldsymbol{\xi}^{(\ell)} + \boldsymbol{\Lambda}^{(\ell)}\boldsymbol{\gamma}_j^{(\ell)} + \boldsymbol{\Sigma}^{(\ell)}\mathbf{S}_j^{\mathrm{oo}(\ell)}(\mathbf{Y}_j - \boldsymbol{\xi}^{(\ell)} - \boldsymbol{\Lambda}^{(\ell)}\boldsymbol{\gamma}_j^{(\ell)})\Big). \quad (2.21)\end{aligned}$$

2.6. Numerical Illustrations

2.6.1. The Interview Data

In this subsection, we apply the likelihood-based procedure to the interview data used by Sahu et al. (2003). This data set consists of bivariate measurements of science scores (Y_{i1}) and non-academic scores (Y_{i2}) on 731 applicants for the admission to a certain medical school. Each application was evaluated on its own academic and nonacademic performance. The score of nonacademic performance was recorded with the summation of the seven performance categories, including the work experience, sense of responsibility, commitment and caring, motivation, study skills, interest and referees' comments.

The score in science was obtained from the secondary examination for selected applicants with qualified non-academic totals. Sahu et al. (2003, p.144) concluded the MSN model is favored when compared with the MST model, especially it make great improvements over the usual MVN model. Our goal for this example is to investigate a good learning model under the complete data and different missing data scenarios.

Table 2.1: Comparison of ML estimation and modeling adequacy among three fitted models.

Parameters	MVN		MSN		MST	
estimation	mle	se	mle	se	mle	se
ξ_1	8.08	0.09	10.35	0.16	10.29	0.16
ξ_2	25.68	0.14	28.90	0.20	28.60	0.25
σ_{11}	5.08	0.29	2.10	0.31	1.82	0.28
σ_{12}	2.49	0.33	1.97	0.31	1.49	0.26
σ_{22}	10.85	0.40	4.07	0.55	3.74	0.55
λ_1	–	–	−2.86	0.19	−2.58	0.21
λ_2	–	–	−4.09	0.20	−3.36	0.35
ν	–	–	–	–	11.21	3.38
m	5		7		8	
$\ell(\hat{\boldsymbol{\theta}}^{(k)}\|\mathbf{Y}^o)$	−3496.30		−3451.38		−3435.93	
AIC	7002.60		6916.76		6887.86	
BIC	7025.57		6948.92		6924.62	

To conduct the complete data analysis, we fit the bivariate MVN, MSN and MST models separately to the data. The resulting ML solutions, including the parameter estimation and the associated standard errors calculated via the inverse of the observed information, together with the maximized log-likelihood values $\ell(\hat{\boldsymbol{\theta}}^{(k)}|\mathbf{Y}^o)$ and two widely used information criteria, Akaike information criterion (AIC) and Bayesian information criterion (BIC), are listed in Table 2.1. All estimates are significant, compared with two times their standard errors. Note that a smaller value of AIC and BIC is associated with a better fitted model. The results indicate that the MST model is the favorite choice because it has the smallest AIC and BIC values when compared with the other two competitive models. Moreover, the estimated

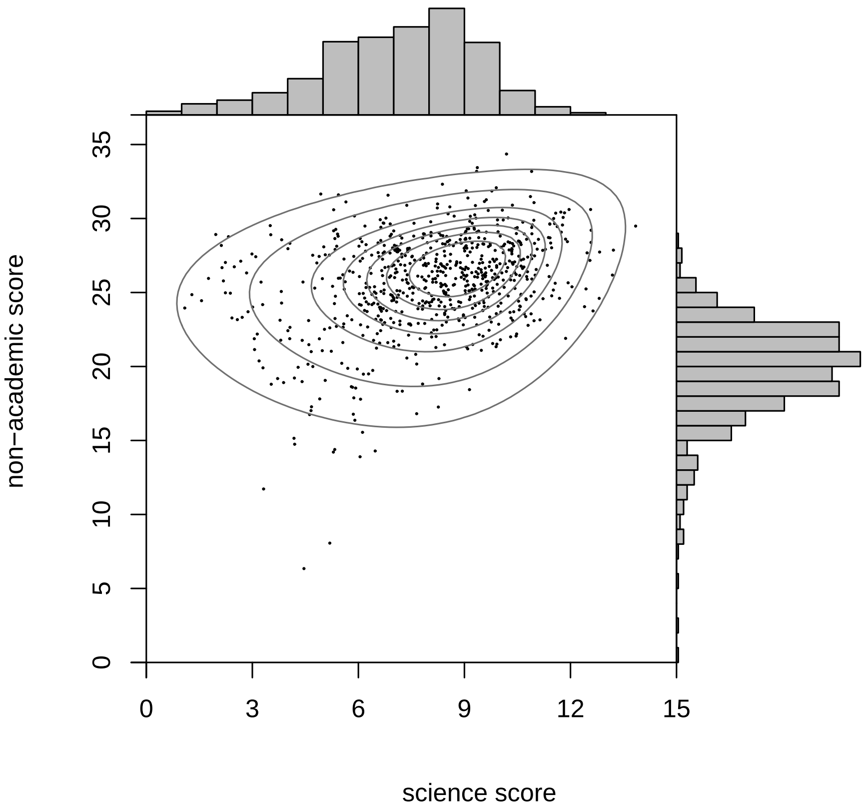

Figure 2.1: The scatter plot of the interview data, overlaid on several contour lines obtained from the fitted MST model.

degrees of freedom of the fitted skew t model is 11.21, confirming the presence of somewhat thick tails.

The fitted MST contours superimposed on a scatter plot along with two summary histograms concerning their marginal distribution are depicted in Fig. 2.1. It is clear to observe that the fitted contours adapt the shape of the scattering pattern ideally, indicating the appropriateness of the use of MST distribution.

To conduct experimental studies under the incomplete data scenario, we construct several artificially missing data sets by deleting at random from the interview data with various missing rates ($r = 0.1, 0.2, 0.3, 0.4$), and they are subsequently fitted with the MVN, MSN and MST models. The above simulation and fitting were repeated 500 times. The density plots of the converged log-likelihoods are depicted in Fig. 2.2. We found that the MST model also provides better model-fitting results even through the data were missing.

For this example with synthetic missing values, we are also interested in testing the null hypothesis H_0 : MSN model ($\nu = \infty$) *versus* the alternative hypothesis H_1 : MST model. The likelihood ratio test (LRT) statistic, given by the difference in values of -2 times the log-likelihood between two nest models, is used to judge which of the two models is more appropriate for this data set. Fig. 2.3 displays the box plots for each of 500 LRT statistics under four missing rates. In all 2,000 cases, the numbers of LRT statistics that are significant at the 5% significant level are 500, 500, 499 and 411 with respect to missing rates $r = 0, 1, 0, 2, 0.3, 0.4$. Therefore, the MST model works well for the fitting of multivariate continuous data in the presence of asymmetrically atypical observations and a number of missing values.

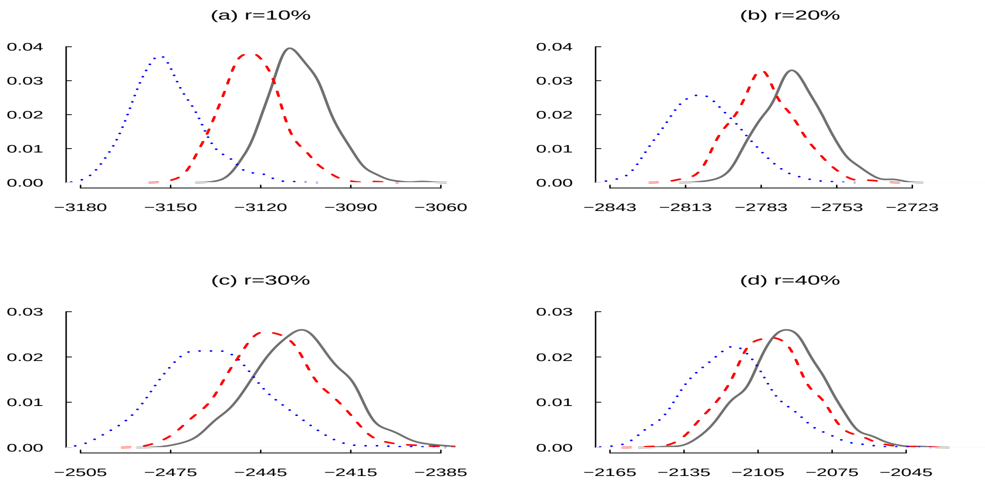

Figure 2.2: The densities plots of the convergent log-likelihood values under MST (solid line), MSN (dashed line) and MVN (dotted line) models for various proportions of missing values. (Replications=500)

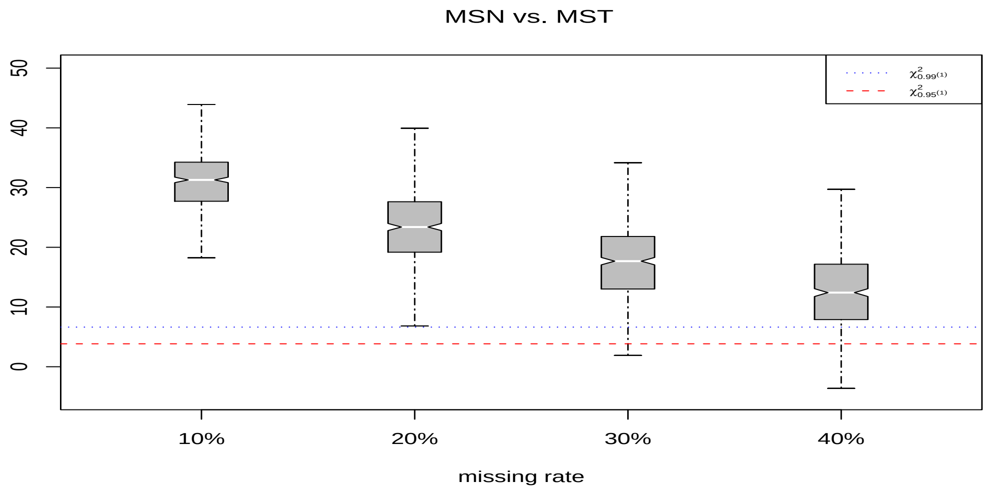

Figure 2.3: Boxplots of 500 LRT statistics (MSN vs. MST) for various missing rates.

2.6.2. The Wind Speed Data

We show a further comparison of Bayesian imputation method to the trivariate wind speed data, which were used by Gneiting et al. (2006) and subsequently analyzed by Azzalini and Genton (2008) for the study of spatial distribution of wind speed by means of another version of the MST distribution proposed by Azzalini and Capitanio (2003). This data set contains 278 hourly average speed assembled at three meteorological towers: Goodnoe Hills (gh), Kennewick (kw) and Vansycle (vs) from 23 February to 30 November 2003 recorded at midnight. For the measurements of this data set, the positive and negative signs stand for a westerly wind direction and an easterly wind direction, respectively.

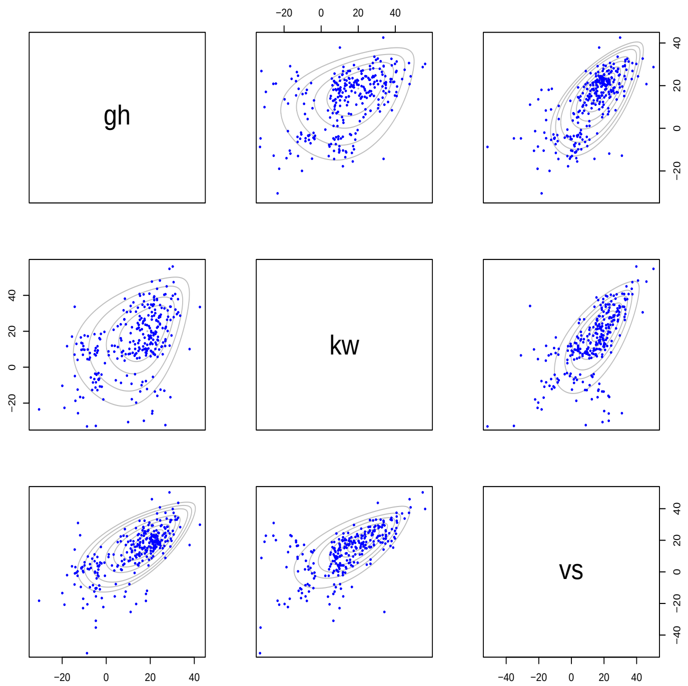

Figure 2.4: Wind speed data: bivariate scatter plots and their fitted MST contours.

To explore the distribution of wind speed data, Fig. 2.4 depicts pairwise bivariate scatter plots along with superimposed ML fitted MST contours. The patterns of fitted densities exhibit the presence of skewness and heavy tails, as displayed in Fig. 5 of Azzalini and Genton (2008) by using another MST distribution.

As in the previous example, we compare the MVN, MSN and MST models with the winspeed data, while they are fitted by using the MCMC sampling. We ran ten parallel chains with the starting values of each chain drawn independently from the prior distributions. The multivariate potential scale reduction factor (MPSRF) of Brooks and Gelman (1998), based on the parallel ten chains, was used to assess convergence. Posterior inferences for the three models, including the mean, standard deviation, and 2.5% and 97.5% quantiles, are shown in Table 2.2.

As seen in the table, it is found that the posterior intervals for the skewness parameters are all significantly different from zero, justifying the existence of skewness for the variables. Moreover, the Bayesian estimate for $\nu = 4.33$ is extremely small, signifying the presence of highly heavy-tailed distributions. For model comparison, the associated values of deviance information criterion (DIC; Spiegelhalter et al., 2002) of the models are included in the table. The result indicates that the MST model is the preferred choice because it has the smallest DIC value.

To examine the predictive abilities among three DA predictors, we utilize the cross validation approach to evaluate their performances. The missing data sets are generated artificially by deleting at random from the wind speed data under four missing rates $r = 0.1, 0.2, 0.3, 0.4$. To implement the DA algorithm, we run 10,000 iterations with the first 5,000 iterations as the burn-in and store the remaining 5,000 iterations as the inference samples. It is noted that our chosen burn-in number is much larger than needed. Simulations were run a total of 100 replications for each

Table 2.2: Summarized posterior inferences and the associated DIC values among three models.

Posterior	Parameter												
estimates	ξ_1	ξ_2	ξ_3	σ_{11}	σ_{12}	σ_{13}	σ_{22}	σ_{23}	σ_{33}	λ_1	λ_2	λ_3	ν
						MVN: (DIC=6527.20)							
Mean	12.71	14.02	16.97	178.98	110.40	127.32	298.89	148.53	186.54	—	—	—	
Std. Dev.	0.80	1.04	0.82	15.27	15.46	13.39	25.80	16.66	15.89	—	—	—	
Median	12.71	14.03	16.96	178.28	109.63	126.51	297.21	147.68	185.64	—	—	—	
$Q_{0.025}$	11.12	11.99	15.39	151.33	81.72	103.06	253.82	118.34	158.17	—	—	—	
$Q_{0.975}$	14.31	16.10	18.57	211.62	143.13	155.36	354.45	183.79	220.37	—	—	—	
						MSN: (DIC=6481.56)							
Mean	20.14	26.57	19.98	132.78	103.35	118.42	194.09	144.60	174.11	-8.93	-15.36	-3.81	—
Std. Dev.	2.90	1.49	1.09	25.00	13.85	13.15	24.63	15.56	15.63	3.34	1.36	0.93	—
Median	20.85	26.59	19.97	129.61	102.71	117.87	192.77	143.96	173.50	-9.88	-15.40	-3.80	—
$Q_{0.025}$	11.44	23.55	17.80	92.17	78.50	94.84	149.84	115.90	144.81	-12.58	-17.92	-5.69	—
$Q_{0.975}$	23.71	29.55	22.12	188.47	132.68	145.88	246.67	176.56	206.74	1.86	-12.60	-2.03	—
						MST: (DIC=6394.67)							
Mean	20.56	26.57	23.31	98.81	67.78	80.50	130.08	96.95	100.38	-7.39	-12.14	-5.46	4.33
Std. Dev.	1.41	1.45	1.01	17.11	11.99	11.54	20.13	12.92	13.04	1.69	1.53	1.08	0.78
Median	20.63	26.59	23.35	97.98	67.27	80.19	129.36	96.24	99.82	-7.46	-12.23	-5.51	4.29
$Q_{0.025}$	17.68	23.68	21.23	66.82	46.12	59.35	93.77	74.08	77.28	-10.50	-14.93	-7.43	2.99
$Q_{0.975}$	23.07	29.31	25.19	135.24	92.98	104.36	173.23	124.11	127.41	-3.92	-8.93	-3.20	6.11

Table 2.3: Coverage probabilities for missing values generated artificially from various missing rates. Values within parentheses are empirical standard errors. (Replications=100)

Models	Missing rates			
	10%	20%	30%	40%
MVN	0.169	0.195	0.203	0.225
	(0.042)	(0.027)	(0.024)	(0.022)
MSN	0.833	0.815	0.797	0.790
	(0.049)	(0.048)	(0.054)	(0.052)
MST	0.934	0.934	0.906	0.896
	(0.033)	(0.023)	(0.027)	(0.027)

missing rate r and each simulated missing data set was imputed using three DA predictors. Using the imputed samples, we calculate the 95% predictive interval for each missing datum and record whether or not the interval covers the true value. The average of coverage probabilities are given in Table 2.3. As it can be seen, the MST predictor yields much better coverage probabilities (close to the nominal 0.95 level) than the other two predictors.

2.7. Conclusion

From the likelihood-based and a Bayesian point of view, we have described analytically flexible estimation and imputation methods for MST models under a complete data framework. The proposed modeling approach can effectively accommodate the possible skewness as well as heavy tails for a general missingness pattern of the data. We present a convenient hierarchical representation that is useful for the implementation of computing algorithms. We offer a workable DA procedure, which can be used to simulate the entire posterior distributions of the parameters and perform multiple imputations in case of data incompleteness. The proposed algorithmic schemes through the incorporation of auxiliary matrices can significantly lessen the computing complexity. The experimental studies have highlighted the superiority of MST models on the provision of more adequate results when the available data are possibly incomplete. Finally, it is worthwhile to remark that the situation in which no missing values in data can be treated as a special case, namely with $\mathbf{O}_i$ taken by $\mathbf{I}_p$ for all i.

3. Robust model-based clustering using multivariate skew t mixtures with missing information

3.1. Introduction

Finite mixture models are known as powerful and flexible tools, which have been fully developed and used in various area and real problems. Moreover, the mixture models have been successfully applied in various kinds of area such as modelling the gene data, the failure rate data and unsupervised clustering problems. There are a number of fairly board monographs in this area, for example, Titterington et al. (1985), McLachlan and Basford (1988), McLachlan and Peel (2000), Frühwirth-Schnatter (2006) and the references therein.

Peel and McLachlan (2000) proposed the MVTMIX model as a robust extension of the MVNMIX model when the underlying data have heavy tails. In some situations, the t mixtures may not be suitable to handle data with highly asymmetric observations. Lin et al. (2007a) proposed a novel univariate skew t mixture model, which allows to regulate skewness and accommodate heavy tail simultaneously. Despite having sound experimental results using STMIX, its application is still limited to data with univariate outcomes. We are motivated to propose a multivariate version of skew t mixture (MSTMIX) model.

We assume that the mechanism of missingness is MAR, meaning that the missingness depends only on the observed values but not on the missing values. With the occurrence of missing values, we offer an efficient MCECM for the fitting of the MSTMIX model. The MCECM algorithm is a modification of the EM algorithm where the E step is computed numerically through Monte Carlo simulations. Since the E-step involved multiple integrations and the integrations are complicated, the

MCECM algorithm is used to solve this problem.

In Section 3.2, we describe the MSTMIX model, define the notations, and present some important statistical properties within the missing information framework. An efficient MCECM algorithm is used to compute the ML estimates and the associated standard errors are derived from the information-based technique. Issues on classification and prediction problems of incomplete features are also discussed. The proposed techniques are examined by using a simulation study and a real data set in Section 3.3. Some concluding remarks are given in Section 3.4.

3.2. The multivariate skew t mixture distribution with missing information

3.2.1. Definition and some properties

In the MSTMIX model, we assume that $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$ from a p-dimensional random sample from a population with g subclasses $G_1, \ldots, G_g$. The pdf of $\mathbf{Y}_j$ can be written as

$$f(\mathbf{Y}_j|\boldsymbol{\Theta}) = \sum_{i=1}^{g} w_i f_p(\mathbf{Y}_j|\boldsymbol{\xi}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\Lambda}_i, \nu_i), \quad w_i \geq 0, \quad \sum_{i=1}^{g} w_i = 1, \tag{3.1}$$

where w_i's are mixing probabilities, $f_p(\cdot|\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$ denotes a p-dimensional MST density with location vector $\boldsymbol{\xi}$, positive definite scale covariance matrix $\boldsymbol{\Sigma}$, skewness matrix $\boldsymbol{\Lambda} = \mathrm{Diag}(\boldsymbol{\lambda})$ with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^\top$, degrees of freedom ν, $\boldsymbol{\Theta} = (w_1, \ldots, w_g, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_g, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_g, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_g, \nu_1, \ldots, \nu_g)$ are the mixture model parameters subject to $\sum_{i=1}^{g} w_i = 1$.

The mean and covariance of $\mathbf{Y}_j$ are given by

$$E(\mathbf{Y}_j) = \sum_{i=1}^{g} w_i \boldsymbol{\mu}_i,$$
$$\text{cov}(\mathbf{Y}_j) = \sum_{i=1}^{g} \Big\{ w_i (1 - w_i) \boldsymbol{\mu}_i \boldsymbol{\mu}_i^\top + w_i \boldsymbol{\Sigma}_i^* \Big\} - \sum_{i \neq j}^{g} w_i w_j \boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top,$$

where

$$\boldsymbol{\mu}_i = \boldsymbol{\xi}_i + \sqrt{\frac{\nu_i}{\pi}} \frac{\Gamma(\frac{\nu_i - 1}{2})}{\Gamma(\frac{\nu_i}{2})} \boldsymbol{\lambda}_i$$

and

$$\boldsymbol{\Sigma}_i^* = \frac{\nu_i}{\nu_i - 2}\Big(\boldsymbol{\Sigma}_i + \Big(1 - \frac{2}{\pi}\Big)\boldsymbol{\Lambda}_i^2\Big) + \frac{2}{\pi}\bigg(\frac{\nu_i}{\nu_i - 2} - \bigg(\frac{\Gamma(\frac{\nu_i - 1}{2})}{\Gamma(\frac{\nu_i}{2})}\bigg)^2 \frac{\nu_i}{2}\bigg) \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^\top$$

are the mean vector and covariance matrix of $ST_p(\boldsymbol{\xi}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\Lambda}_i, \nu_i)$, respectively.

For each $\mathbf{Y}_j$, it is convenient to introduce a set of zero-one indicator variables $\mathbf{Z}_j = (Z_{1j}, \ldots, Z_{gj})^\top$ for $j = 1, \ldots, n$, which is a multinomial random vector with 1 trial and cell probabilities $w_1, \ldots, w_g$, denoted as $\mathbf{Z}_j \sim M(1; w_1, \ldots, w_g)$. Note that the rth element $Z_{rj} = 1$ if $\mathbf{Y}_j$ arises from component r. A four level hierarchical representation of Eq. (3.1) can be expressed by:

$$\begin{aligned}
\mathbf{Y}_j \mid (\boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1) &\sim N_p(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j, \boldsymbol{\Sigma}_i / \tau_j), \\
\boldsymbol{\gamma}_j \mid (\tau_j, Z_{ij} = 1) &\sim HN_p(\mathbf{0}, \mathbf{I}_p / \tau_j), \\
\tau_j \mid (Z_{ij} = 1) &\sim \Gamma(\nu_i/2, \nu_i/2), \\
\mathbf{Z}_j &\sim M(1; w_1, \ldots, w_g),
\end{aligned} \tag{3.2}$$

for $i = 1, \ldots, g$, $j = 1, \ldots, n$.

Let $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$, $\mathbf{Z} = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_n^\top)^\top$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)^\top$. From (3.2), the complete data log-likelihood function of $\boldsymbol{\Theta}$, ignoring additive constant terms, is

$$\begin{aligned}
&\ell_c(\boldsymbol{\Theta}|\mathbf{Y},\mathbf{Z},\boldsymbol{\gamma},\boldsymbol{\tau}) \\
= &\sum_{i=1}^{g}\sum_{j=1}^{n} Z_{ij}\bigg\{\log(w_i)+\frac{\nu_i}{2}\log\Big(\frac{\nu_i}{2}\Big)-\log\Gamma\Big(\frac{\nu_i}{2}\Big)-\frac{1}{2}\log|\boldsymbol{\Sigma}_i|+\frac{\nu_i}{2}\log\tau_j \\
&-\frac{\tau_j}{2}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)-\frac{\tau_j\nu_i}{2}\bigg\}. \qquad (3.3)
\end{aligned}$$

We consider the ML estimation problem of model (3.1) when $\mathbf{Y}$ may be partially observed. As in previous strategy, we partition the random vector $\mathbf{Y}_j$ into two components $(\mathbf{Y}_j^{\mathrm{o}\top},\mathbf{Y}_j^{\mathrm{m}\top})^\top$, with $p_j=p_j^{\mathrm{o}}+p_j^{\mathrm{m}}$, where $\mathbf{Y}_j^{\mathrm{o}}$ $(p_j^{\mathrm{o}}\times 1)$ and $\mathbf{Y}_j^{\mathrm{m}}$ $(p_j^{\mathrm{m}}\times 1)$ denote the observed and the missing components of $\mathbf{Y}_j$, respectively.

Theorem 3.1 *From (3.2), it can be shown that*

(a) *The distribution of* $\mathbf{Y}_j^{\mathrm{o}}$ *given* $\boldsymbol{\gamma}_j$, τ_j *and* $Z_{ij}=1$ *is*

$$\mathbf{Y}_j^{\mathrm{o}}|(\boldsymbol{\gamma}_j,\tau_j,Z_{ij}=1)\sim N_{p_j^{\mathrm{o}}}(\boldsymbol{\zeta}_{ij}^{\mathrm{o}},\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}/\tau_j)$$

where $\boldsymbol{\zeta}_{ij}^{\mathrm{o}}=\mathbf{O}_j(\boldsymbol{\xi}_i+\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)$ *and* $\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}=\mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top$.

(b) *The conditional distribution of* $\mathbf{Y}_j^{\mathrm{m}}$ *given* $\mathbf{Y}_j^{\mathrm{o}}$, $\boldsymbol{\gamma}_j$, τ_j *and* $Z_{ij}=1$ *is*

$$\mathbf{Y}_j^{\mathrm{m}}|(\mathbf{Y}_j^{\mathrm{o}},\boldsymbol{\gamma}_j,\tau_j,Z_{ij}=1)\sim N_{p-p_j^{\mathrm{o}}}(\boldsymbol{\zeta}_{ij}^{\mathrm{m\cdot o}},\boldsymbol{\Sigma}_{ij}^{\mathrm{mm\cdot o}}/\tau_j),$$

where $\boldsymbol{\zeta}_{ij}^{\mathrm{m\cdot o}}=\mathbf{M}_j(\boldsymbol{\xi}_i+\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j+\boldsymbol{\Sigma}_i\mathbf{S}_{ij}^{\mathrm{oo}}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j))$, $\boldsymbol{\Sigma}_{ij}^{\mathrm{mm\cdot o}}=\mathbf{M}_j(\mathbf{I}_p-\boldsymbol{\Sigma}_i\mathbf{S}_{ij}^{\mathrm{oo}})\boldsymbol{\Sigma}_i\mathbf{M}_j^\top$, *and* $\mathbf{S}_{ij}^{\mathrm{oo}}=\mathbf{O}_j^\top(\mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top)^{-1}\mathbf{O}_j$.

(c) *The marginal distribution of* $\mathbf{Y}_j^{\mathrm{o}}$ *is* $\sum_{i=1}^{g}w_i f_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}},\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}},\nu_i)$, *where* $\boldsymbol{\xi}_{ij}^{\mathrm{o}}=\mathbf{O}_j\boldsymbol{\xi}_i$ *and* $\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}=\mathbf{O}_j\boldsymbol{\Lambda}_i\mathbf{O}_j^\top$.

(d) *The posterior distribution of* $\boldsymbol{\gamma}_j$ *given* $\mathbf{Y}_j^{\mathrm{o}}$ *and* $Z_{ij}=1$ *is multivariate truncated* t *distribution. That is*

$$\boldsymbol{\gamma}_j|(\mathbf{Y}_j^{\mathrm{o}},Z_{ij}=1)\sim Tt_p\Big(\mathbf{q}_{ij}^*,\frac{(U_{ij}^{\mathrm{o}}+\nu)}{(p_j^{\mathrm{o}}+\nu)}\boldsymbol{\Delta}_{ij}^*,\nu+p_j^{\mathrm{o}};\mathbb{R}_+^p\Big),$$

where $U^{\mathrm{o}}_{ij} = (\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top \mathbf{C}^{\mathrm{oo}}_{ij}(\mathbf{Y}_j-\boldsymbol{\xi}_i)$, $\boldsymbol{\Delta}^*_{ij} = \mathbf{I}_p-\boldsymbol{\Lambda}_i\mathbf{C}^{\mathrm{oo}}_{ij}\boldsymbol{\Lambda}_i$, *and* $\mathbf{q}^*_{ij} = \boldsymbol{\Lambda}_i\mathbf{C}^{\mathrm{oo}}_{ij}(\mathbf{Y}_j-\boldsymbol{\xi}_i)$, $\mathbf{C}^{\mathrm{oo}}_{ij} = \mathbf{O}_j^\top {\boldsymbol{\Omega}^{\mathrm{oo}}_{ij}}^{-1}\mathbf{O}_j$, *and* $\boldsymbol{\Omega}^{\mathrm{oo}}_{ij} = \mathbf{O}_j(\boldsymbol{\Lambda}_i^2+\boldsymbol{\Sigma}_i)\mathbf{O}_j^\top$.

(e) *the posterior distribution of* τ_j *given* $\mathbf{Y}^{\mathrm{o}}_j$, $\boldsymbol{\gamma}_j$ *and* $Z_{ij}=1$ *is a gamma distribution. That is*

$$\tau_j|(\mathbf{Y}^{\mathrm{o}}_j,\boldsymbol{\gamma}_j,Z_{ij}=1) \sim \Gamma\left(\frac{p+p^{\mathrm{o}}_j+\nu_i}{2}, \frac{(\boldsymbol{\gamma}_j-\mathbf{q}^*_{ij})^\top {\boldsymbol{\Delta}^*_{ij}}^{-1}(\boldsymbol{\gamma}_j-\mathbf{q}^*_{ij})+U^{\mathrm{o}}_{ij}+\nu_i}{2}\right).$$

Proof. The proofs of part (a) and (b) are straightforward and hence are omitted. The proof of part (c), (d), and (e) are given in Appendix K.

Corollary 3.1 *From (3.2), we have the following:*

(a) *The conditional expectation of* τ_j *given* $\mathbf{Y}^{\mathrm{o}}_j$ *and* $Z_{ij}=1$ *is*

$$E(\tau_j \mid \mathbf{Y}^{\mathrm{o}}_j, Z_{ij}=1) = \left(\frac{\nu_i+p^{\mathrm{o}}_j}{U^{\mathrm{o}}_{ij}+\nu_i}\right)\frac{T_p\left(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\mathrm{o}}_j+2}{U^{\mathrm{o}}_{ij}+\nu_i}} \mid \boldsymbol{\Delta}^*_{ij};\nu_i+p^{\mathrm{o}}_j+2\right)}{T_p\left(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\mathrm{o}}_j}{U^{\mathrm{o}}_{ij}+\nu_i}} \mid \boldsymbol{\Delta}^*_{ij};\nu_i+p^{\mathrm{o}}_j\right)}.$$

(b) *The conditional expectation of* $\log(\tau_j)$ *given* $\mathbf{Y}^{\mathrm{o}}_j$ *and* $Z_{ij}=1$ *is*

$$\begin{aligned}
&E(\log\tau_j \mid \mathbf{Y}^{\mathrm{o}}_j, Z_{ij}=1)\\
=\ & \mathrm{DG}\left(\frac{\nu_i+p^{\mathrm{o}}_j}{2}\right)+\left(\frac{\nu_i+p^{\mathrm{o}}_j}{U^{\mathrm{o}}_{ij}+\nu_i}\right)\left(\frac{T_p\left(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\mathrm{o}}_j+2}{U^{\mathrm{o}}_{ij}+\nu_i}} \mid \boldsymbol{\Delta}^*_{ij};\nu_i+p^{\mathrm{o}}_j+2\right)}{T_p\left(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\mathrm{o}}_j}{U^{\mathrm{o}}_{ij}+\nu_i}} \mid \boldsymbol{\Delta}^*_{ij};\nu_i+p^{\mathrm{o}}_j\right)}-1\right)\\
&+T_p^{-1}\left(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\mathrm{o}}_j}{U^{\mathrm{o}}_{ij}+\nu_i}}\,\middle|\,\boldsymbol{\Delta}^*_{ij};\nu_i+p^{\mathrm{o}}_j\right)\\
&\times\prod_{r=1}^{p}\int_{-\infty}^{q^*_{ijr}} g_{\nu_i}(\mathbf{x})t_p\left(\mathbf{x}\,\middle|\,\frac{U^{\mathrm{o}}_{ij}+\nu_i}{\nu_i+p^{\mathrm{o}}_j}\times\boldsymbol{\Delta}^*_{ij};\nu_i+p^{\mathrm{o}}_j\right)d\mathbf{x}-\log\left(\frac{U^{\mathrm{o}}_{ij}+\nu_i}{2}\right),
\end{aligned}$$

and

$$\begin{aligned}
g_{\nu_i}(\mathbf{x}) =\ & \mathrm{DG}\left(\frac{\nu_i+p+p^{\mathrm{o}}_j}{2}\right)-\mathrm{DG}\left(\frac{\nu_i+p^{\mathrm{o}}_j}{2}\right)-\frac{p}{p^{\mathrm{o}}_j+\nu_i}+\frac{p(U^{\mathrm{o}}_{ij}-p^{\mathrm{o}}_j)}{(\nu_i+p^{\mathrm{o}}_j)(U^{\mathrm{o}}_{ij}+\nu_i)}\\
&-\log\left(1+\frac{\mathbf{x}^\top{\boldsymbol{\Delta}^*_{ij}}^{-1}\mathbf{x}}{U^{\mathrm{o}}_{ij}+\nu_i}\right)+\frac{(\nu_i+p+p^{\mathrm{o}}_j)(\mathbf{x}^\top{\boldsymbol{\Delta}^*_{ij}}^{-1}\mathbf{x})}{(U^{\mathrm{o}}_{ij}+\nu_i)(U^{\mathrm{o}}_{ij}+\nu_i+\mathbf{x}^\top{\boldsymbol{\Delta}^*_{ij}}^{-1}\mathbf{x})},
\end{aligned}$$

where q^*_{ijr} *is the* r*th element of* $\mathbf{q}^*_{ij} = (q^*_{ij1}, \ldots, q^*_{ijp})$.

Proof: The proof is given in Appendix L.

3.2.2. ML estimation via the MCECM procedure

For notational simplicity, let $\mathbf{\Theta} = (w_1, \ldots, w_g, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_g, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_g, \mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_g, \nu_1, \ldots, \nu_g)$, Let $\mathbf{Y}^{\rm o} = (\mathbf{Y}^{\rm o}_1, \mathbf{Y}^{\rm o}_2, \cdots, \mathbf{Y}^{\rm o}_n)$ and $\mathbf{Y}^{\rm m} = (\mathbf{Y}^{\rm m}_1, \mathbf{Y}^{\rm m}_2, \cdots, \mathbf{Y}^{\rm m}_n)$ represent the observed portion and the missing portion of the data, respectively. The complete data log-likelihood function of $\mathbf{\Theta}$, ignoring additive constant terms, is

$$\begin{aligned} & \ell_c(\mathbf{\Theta}|\mathbf{Y}^{\rm o}, \mathbf{Y}^{\rm m}, \mathbf{Z}, \boldsymbol{\gamma}, \boldsymbol{\tau}) \\ = & \sum_{i=1}^{g}\sum_{j=1}^{n} Z_{ij}\log(w_i) + \sum_{i=1}^{g}\sum_{j=1}^{n} Z_{ij}\left\{\frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right) - \log\Gamma\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2}(\log\tau_j - \tau_j)\right\} \\ & -\frac{1}{2}\sum_{i=1}^{g}\left[\log|\mathbf{\Sigma}_i|\left(\sum_{j=1}^{n} Z_{ij}\right) + \mathrm{tr}\left(\mathbf{\Sigma}_i^{-1}\sum_{j=1}^{n}\mathbf{\Omega}_{ij}\right)\right], \end{aligned} \tag{3.4}$$

where $\mathbf{\Omega}_{ij} = Z_{ij}\tau_j(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)^\top$.

Let $\hat{\tau}^{(k)}_{ij} = E(\tau_j|\mathbf{Y}^{\rm o}_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)})$, $\hat{\kappa}^{(k)}_{ij} = E(\log\tau_j|\mathbf{Y}^{\rm o}_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)})$, $\hat{\boldsymbol{\eta}}^{(k)}_{ij} = E(\tau_j\boldsymbol{\gamma}_j|\mathbf{Y}^{\rm o}_j Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)})$ and $\hat{\mathbf{\Psi}}^{(k)}_{ij} = E(\tau_j\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}^{\rm o}_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)})$ be the necessary conditional expectations involved in (3.4). Let $\hat{\mathbf{\Theta}}^{(k)} = (\hat{w}^{(k)}_1, \ldots, \hat{w}^{(k)}_g, \hat{\boldsymbol{\xi}}^{(k)}_1, \ldots, \hat{\boldsymbol{\xi}}^{(k)}_g, \hat{\mathbf{\Sigma}}^{(k)}_1, \ldots, \hat{\mathbf{\Sigma}}^{(k)}_g, \hat{\mathbf{\Lambda}}^{(k)}_1, \ldots, \hat{\mathbf{\Lambda}}^{(k)}_g, \hat{\nu}^{(k)}_1, \ldots, \hat{\nu}^{(k)}_g)$ denote the estimates of $\mathbf{\Theta}$ at the kth iteration. Given the observed data $\mathbf{Y}^{\rm o}$ and the current parameter estimates $\mathbf{\Theta} = \hat{\mathbf{\Theta}}^{(k)}$, we can calculate

$$\hat{Z}^{(k)}_{ij} = \Pr(Z_{ij} = 1 \mid \mathbf{Y}^{\rm o}_j, \hat{\mathbf{\Theta}}^{(k)}) = \frac{\hat{w}^{(k)}_i f_{p^{\rm o}_j}(\mathbf{Y}^{\rm o}_j|\hat{\boldsymbol{\xi}}^{{\rm o}(k)}_{ij}, \hat{\mathbf{\Sigma}}^{{\rm oo}(k)}_{ij}, \hat{\mathbf{\Lambda}}^{{\rm oo}(k)}_{ij}, \hat{\nu}^{(k)}_i)}{\sum_{i=1}^{g}\hat{w}^{(k)}_i f_{p^{\rm o}_j}(\mathbf{Y}^{\rm o}_j|\hat{\boldsymbol{\xi}}^{{\rm o}(k)}_{ij}, \hat{\mathbf{\Sigma}}^{{\rm oo}(k)}_{ij}, \hat{\mathbf{\Lambda}}^{{\rm oo}(k)}_{ij}, \hat{\nu}^{(k)}_i)}, \tag{3.5}$$

and

$$\begin{aligned}
\mathbf{\Omega}_{ij}^{(k)} &= E(\mathbf{\Omega}_{ij} \mid \mathbf{Y}^{\text{o}}, \hat{\mathbf{\Theta}}^{(k)}) \\
&= \hat{Z}_{ij}^{(k)} \Big[(\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\text{oo}(k)}) \hat{\mathbf{\Sigma}}_i^{(k)} + (\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i) \hat{\mathbf{\Psi}}_{ij}^{(k)} (\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i)^\top \\
&\quad \hat{\tau}_{ij}^{(k)} (\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top + (\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i) \hat{\boldsymbol{\eta}}_{ij}^{(k)} (\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top \\
&\quad (\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i) \hat{\boldsymbol{\eta}}_{ij}^{(k)^\top} (\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i)^\top \Big], \qquad (3.6)
\end{aligned}$$

where

$$\hat{\mathbf{A}}_{ij}^{(k)} = (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\text{oo}(k)}) \hat{\mathbf{\Lambda}}_i^{(k)} \quad \text{and} \quad \hat{\mathbf{b}}_{ij}^{(k)} = \hat{\mathbf{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\text{oo}(k)} \mathbf{Y}_j + (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\text{oo}(k)}) \hat{\boldsymbol{\xi}}_i^{(k)}. \quad (3.7)$$

Proof: The proof is given in Appendix M.

Furthermore, a Monte Carlo estimate of the Q-function can be evaluated as

$$\hat{Q}(\mathbf{\Theta} \mid \hat{\mathbf{\Theta}}^{(k)}) = \frac{1}{M} \sum_{m=1}^{M} E\big(\ell_c(\mathbf{\Theta} \mid \mathbf{Y}^{\text{o}}, \mathbf{Y}^{\text{m}}, \hat{\boldsymbol{\gamma}}_{[m]}^{(k)}, \hat{\boldsymbol{\tau}}_{[m]}^{(k)}, \mathbf{Z}, \hat{\mathbf{\Theta}}^{(k)}) \mid \mathbf{Y}^{\text{o}}, \hat{\mathbf{\Theta}}^{(k)}\big), \qquad (3.8)$$

where $\hat{\boldsymbol{\gamma}}_{[m]}^{*(k)} = \{\hat{\boldsymbol{\gamma}}_{ij,m}^{*(k)}, i = 1, \ldots, g; j = 1, \ldots, n\}$ and $\hat{\boldsymbol{\tau}}_{[m]}^{*(k)} = \{\hat{\tau}_{ij,m}^{*(k)}, i = 1, \ldots, g; j = 1, \ldots, n\}$ for $m = 1, \ldots, M$, are a set of independent random samples generated from each $f(\boldsymbol{\gamma}_j, \tau_j | \mathbf{Y}_j^{\text{o}}, Z_{ij} = 1)$ given $\mathbf{\Theta} = \hat{\mathbf{\Theta}}^{(k)}$. The exact sampling of $\boldsymbol{\gamma}_j$ and τ_j can be conveniently implemented through the following generators:

$$\hat{\boldsymbol{\gamma}}_{ij,m}^{(k+1)} \mid (\mathbf{Y}_j^{\text{o}}, Z_{ij} = 1) \sim Tt_p\Big(\hat{\mathbf{q}}_{ij}^{*(k)}, \frac{\hat{U}_{ij}^{\text{o}(k)} + \hat{\nu}_i^{(k)}}{p_j^{\text{o}} + \hat{\nu}_i^{(k)}} \hat{\mathbf{\Delta}}_{ij}^{*(k)}, \hat{\nu}_i^{(k)} + p_j^{\text{o}};\ \mathbb{R}_+^p\Big),$$

and

$$\begin{aligned}
&\hat{\tau}_{ij,m}^{(k+1)} \mid (\hat{\boldsymbol{\gamma}}_{j,m}^{(k+1)}, \mathbf{Y}_j^{\text{o}}, Z_{ij} = 1) \\
&\quad \sim \Gamma\Big(\frac{\hat{\nu}_i^{(k)} + p + p_j^{\text{o}}}{2}, \frac{(\hat{\boldsymbol{\gamma}}_{ij,m}^{(k+1)} - \hat{\mathbf{q}}_{ij}^{*(k)})^\top \hat{\mathbf{\Delta}}_{ij}^{*(k)^{-1}} (\hat{\boldsymbol{\gamma}}_{ij,m}^{(k+1)} - \hat{\mathbf{q}}_{ij}^{*(k)}) + \hat{U}_{ij}^{\text{o}(k)} + \hat{\nu}_i^{(k)}}{2}\Big),
\end{aligned}$$

where $\hat{\mathbf{q}}_{ij}^{*(k)} = \hat{\mathbf{\Lambda}}_i^{(k)} \hat{\mathbf{C}}_{ij}^{\text{oo}(k)} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)})$, $\hat{\mathbf{\Delta}}_{ij}^{*(k)} = \mathbf{I}_p - \hat{\mathbf{\Lambda}}_i^{(k)} \hat{\mathbf{C}}_{ij}^{\text{oo}(k)} \hat{\mathbf{\Lambda}}_i^{(k)}$, $\hat{U}_{ij}^{\text{o}(k)} = (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)})^\top \hat{\mathbf{C}}_{ij}^{\text{oo}(k)} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)})$ and $\hat{\mathbf{C}}_{ij}^{\text{oo}(k)} = \mathbf{O}_j^\top \hat{\mathbf{\Omega}}_{ij}^{\text{oo}(k)^{-1}} \mathbf{O}_j$ with $\hat{\mathbf{\Omega}}_{ij}^{\text{oo}(k)} = \mathbf{O}_j (\hat{\mathbf{\Sigma}}_i^{(k)} + \hat{\mathbf{\Lambda}}_i^{(k)2}) \mathbf{O}_j^\top$.

Therefore, the conditional expectations defined as above can be readily approximated as

$$\begin{aligned}\hat{\tau}_{ij}^{(k)} &\simeq M^{-1}\sum_{m=1}^{M}\hat{\tau}_{ij,m}^{(k)}, \quad \hat{\kappa}_{ij}^{(k)} \simeq M^{-1}\sum_{m=1}^{M}\log\hat{\tau}_{ij,m}^{(k)},\\ \hat{\boldsymbol{\eta}}_{ij}^{(k)} &\simeq M^{-1}\sum_{m=1}^{M}\hat{\tau}_{ij,m}^{(k)}\hat{\boldsymbol{\gamma}}_{j,m}^{(k)}, \quad \hat{\boldsymbol{\Psi}}_{ij}^{(k)} \simeq M^{-1}\sum_{m=1}^{M}\hat{\tau}_{ij,m}^{(k)}\hat{\boldsymbol{\gamma}}_{ij,m}^{(k)}\hat{\boldsymbol{\gamma}}_{ij,m}^{(k)\top}.\end{aligned} \tag{3.9}$$

Formally, the MCECM algorithm can be implemented as follows:

MCE-step: Given $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(k)}$, compute Monte Carlo expectations $\hat{\tau}_{ij}^{(k)}$, $\hat{\kappa}_{ij}^{(k)}$, $\hat{\boldsymbol{\eta}}_{ij}^{(k)}$ and $\hat{\boldsymbol{\Psi}}_{ij}^{(k)}$ by using (3.9) for $i = 1, \ldots, g$ and $j = 1, \ldots, n$.

CM-steps:

CM-Step 1. Update $\hat{w}_i^{(k)}$ by maximizing Equation (3.8) over w_i subject to their sum is unity, which gives

$$\hat{w}_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}.$$

CM-Step 2. Fix $\boldsymbol{\Lambda}_i = \hat{\boldsymbol{\Lambda}}_i^{(k)}$, update $\hat{\boldsymbol{\xi}}_i^{(k)}$ by maximizing (3.8) over $\boldsymbol{\xi}_i$, which leads to

$$\hat{\boldsymbol{\xi}}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\tau}_{ij}^{(k)}\mathbf{b}_{ij}^{(k)} - \sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\hat{\boldsymbol{\Lambda}}_i^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}}{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\tau}_{ij}^{(k)}},$$

where $\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)} = \mathbf{O}_j^\top(\mathbf{O}_j\hat{\boldsymbol{\Sigma}}_i^{(k)}\mathbf{O}_j^\top)^{-1}\mathbf{O}_j$.

CM-Step 3. Fix $\boldsymbol{\xi}_i = \hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i = \hat{\boldsymbol{\Lambda}}_i^{(k)}$, update $\hat{\boldsymbol{\Sigma}}_i^{(k)}$ by maximizing (3.8) over $\boldsymbol{\Sigma}_i$, which gives

$$\hat{\boldsymbol{\Sigma}}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}}{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}},$$

where $\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}$ is $\boldsymbol{\Omega}_{ij}^{(k)}$ in (3.6) with $\boldsymbol{\xi}_i$ and $\boldsymbol{\Lambda}_i$ replaced by $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, respectively.

CM-Step 4. Fix $\boldsymbol{\xi}_i = \hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Sigma}_i = \hat{\boldsymbol{\Sigma}}_i^{(k+1)}$, updating $\hat{\boldsymbol{\lambda}}_i^{(k)}$ by maximizing (3.8) over $\boldsymbol{\lambda}_i$ yields

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_i^{(k+1)} &= \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^{-1} \\ &\quad \times \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \left(\hat{Z}_{ij}^{(k)} (\hat{\mathbf{b}}_{ij}^{(k)} - \hat{\boldsymbol{\xi}}_i^{(k+1)}) \hat{\boldsymbol{\eta}}_{ij}^{(k)\top} + \hat{\mathbf{A}}_{ij}^{(k)} \hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^{\top}\right) \mathbf{1}_p,\end{aligned}$$

where $\hat{\mathbf{A}}_{ij}^{(k)}$ and $\hat{\mathbf{b}}_{ij}^{(k)}$ are defined in (3.7). It follows immediately that $\hat{\boldsymbol{\Lambda}}_i^{(k+1)} = \mathrm{Diag}(\hat{\boldsymbol{\lambda}}_i^{(k+1)})$.

CM-Step 5. Fix $\boldsymbol{\xi}_i = \hat{\boldsymbol{\xi}}_i^{(k+1)}$, $\boldsymbol{\Lambda}_i = \hat{\boldsymbol{\Lambda}}_i^{(k+1)}$, and $\boldsymbol{\Sigma}_i = \hat{\boldsymbol{\Sigma}}_i^{(k+1)}$, obtain $\hat{\nu}_i^{(k+1)}$ as the solution of the equation

$$\log\left(\frac{\nu_i}{2}\right) + 1 - \mathrm{DG}\left(\frac{\nu_i}{2}\right) + \frac{1}{\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)}} \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \left(\hat{\kappa}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)}\right) = 0.$$

The detailed proof of the CM-steps is shown in Appendix N.

If the degrees of freedom are assumed to be identical, i.e. $\nu_1 = \cdots = \nu_g = \nu$, the above CM-step 5 can be CML-step as follows:

CML-step: Fix $\boldsymbol{\xi}_i = \hat{\boldsymbol{\xi}}_i^{(k+1)}$, $\boldsymbol{\Lambda}_i = \hat{\boldsymbol{\Lambda}}_i^{(k+1)}$, and $\boldsymbol{\Sigma}_i = \hat{\boldsymbol{\Sigma}}_i^{(k+1)}$, obtain $\hat{\nu}^{(k)}$ as the solution of

$$\log\left(\frac{\nu}{2}\right) + 1 - \mathrm{DG}\left(\frac{\nu}{2}\right) + \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \left(\hat{\kappa}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)}\right) = 0.$$

The MCECM algorithm described in this section tends to be robust to the choice of starting values for the parameters. For the specification of initial values, Seidel et al. (2000) have demonstrated some strategies and stopping rules in the fitting of mixture model. Here, we choose the ML estimates for the complete data as the initial values for the parameters.

Applying Bayes' theorem, the posterior probability of the $\mathbf{Y}_j$ belonging to group i can be estimated by

$$\hat{w}^*_{ij} = \mathrm{P}(Z_{ij} = 1|\mathbf{Y}^{\mathrm{o}}, \hat{\mathbf{\Theta}}) = \frac{\hat{w}_i f_{p^{\mathrm{o}}_j}(\mathbf{Y}^{\mathrm{o}}_j|\hat{\boldsymbol{\xi}}^{\mathrm{o}}_{ij}, \hat{\mathbf{\Sigma}}^{\mathrm{oo}}_{ij}, \hat{\mathbf{\Lambda}}^{\mathrm{oo}}_{ij}, \hat{\nu}_i)}{\sum_{i=1}^{g} \hat{w}_i f_{p^{\mathrm{o}}_j}(\mathbf{Y}^{\mathrm{o}}_j|\hat{\boldsymbol{\xi}}^{\mathrm{o}}_{ij}, \hat{\mathbf{\Sigma}}^{\mathrm{oo}}_{ij}, \hat{\mathbf{\Lambda}}^{\mathrm{oo}}_{ij}, \hat{\nu}_i)}. \tag{3.10}$$

By the ML classification theory (Basford and McLachlan, 1985), $\mathbf{Y}_j$ is assigned to group s if $\hat{w}^*_{sj} > \hat{w}^*_{ij}$ for $i = 1, \ldots, g$ and $i \neq s$.

Consequently, the ML predictor for the missing component $\mathbf{Y}^{\mathrm{m}}_j$ is given by

$$\begin{aligned}\hat{\mathbf{Y}}^{\mathrm{m}}_j &= \mathrm{E}(\mathbf{Y}^{\mathrm{m}}_j|\mathbf{Y}^{\mathrm{o}}, \hat{\mathbf{\Theta}}) \\ &= \mathbf{M}_j \sum_{i=1}^{g} \hat{w}^*_{ij}\Big(\hat{\boldsymbol{\xi}}_i + \hat{\mathbf{\Lambda}}_i\hat{\boldsymbol{\gamma}}_{ij} + \hat{\mathbf{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i - \hat{\mathbf{\Lambda}}_i\hat{\boldsymbol{\gamma}}_{ij})\Big),\end{aligned}$$

where $\hat{\boldsymbol{\gamma}}_{ij} = \mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}^{\mathrm{o}}, Z_{ij} = 1, \hat{\mathbf{\Theta}})$, which can be approximated by Monte Carlo average of samples simulated from $Tt_p\big(\hat{\mathbf{q}}^*_{ij}, \frac{\hat{U}^{\mathrm{o}}_{ij} + \hat{\nu}_i}{p^{\mathrm{o}}_j + \hat{\nu}_i}\hat{\mathbf{\Delta}}^*_{ij}, \hat{\nu}_i + p^{\mathrm{o}}_j;\ \mathbb{R}^p_+\big)$.

3.2.3. Estimation of standard errors

Under some regularity conditions, we provide the information-based method used by Meilijson (1989) to compute the asymptotic covariance of the ML estimates of mixture model parameters.

Let $\ell_{cj}(\mathbf{\Theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \tau_j, \mathbf{Z}_j)$ be the complete data log-likelihood formed from the single observation $\mathbf{Y}_j$. The individual score is defined as

$$\mathbf{u}(\mathbf{Y}^{\mathrm{o}}_j|\mathbf{\Theta}) = E\bigg(\frac{\partial \ell_{cj}(\mathbf{\Theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \tau_j, \mathbf{Z}_j)}{\partial \mathbf{\Theta}}\Big|\mathbf{Y}^{\mathrm{o}}_j, \mathbf{\Theta}\bigg).$$

The empirical information matrix, according to Meilijson's formula, is defined as

$$\mathbf{I}_e(\mathbf{\Theta}|\mathbf{Y}^{\mathrm{o}}) = \sum_{j=1}^{n} \mathbf{u}(\mathbf{Y}^{\mathrm{o}}_j|\mathbf{\Theta})\mathbf{u}^\top(\mathbf{Y}^{\mathrm{o}}_j|\mathbf{\Theta}) - n^{-1}\mathbf{U}(\mathbf{Y}^{\mathrm{o}}|\mathbf{\Theta})\mathbf{U}^\top(\mathbf{Y}^{\mathrm{o}}|\mathbf{\Theta}), \tag{3.11}$$

where $\mathbf{U}(\mathbf{Y}^{\mathrm{o}}|\mathbf{\Theta}) = \sum_{j=1}^{n} \mathbf{u}(\mathbf{Y}^{\mathrm{o}}_j|\mathbf{\Theta})$.

Let vech$(\cdot)$ be the matrix operator which stacks only the distinct elements of a symmetric matrix into a single vector. The ML estimates $\hat{\boldsymbol{\Theta}}$ substituted for $\boldsymbol{\Theta}$ in the Eq. (3.11) and then it reduced to

$$\mathbf{I}_e(\hat{\boldsymbol{\Theta}} \mid \mathbf{Y}^{\rm o}) = \sum_{j=1}^{n} \hat{\mathbf{u}}_j^{\rm o} \hat{\mathbf{u}}_j^{\rm o\top},$$

where

$$\begin{aligned} \hat{\mathbf{u}}_j^{\rm o} &= \mathbf{u}(\mathbf{Y}_j^{\rm o}|\hat{\boldsymbol{\Theta}}) \\ &= (\hat{u}^{\rm o}_{j,w_1}, \ldots, \hat{u}^{\rm o}_{j,w_{g-1}}, \hat{\mathbf{u}}^{\rm o\top}_{j,\boldsymbol{\xi}_1}, \ldots, \hat{\mathbf{u}}^{\rm o\top}_{j,\boldsymbol{\xi}_g}, \hat{\mathbf{u}}^{\rm o\top}_{j,\boldsymbol{\sigma}_1}, \ldots, \hat{\mathbf{u}}^{\rm o\top}_{j,\boldsymbol{\sigma}_g}, \hat{\mathbf{u}}^{\rm o\top}_{j,\boldsymbol{\lambda}_1}, \ldots, \hat{\mathbf{u}}^{\rm o\top}_{j,\boldsymbol{\lambda}_g}, \hat{u}^{\rm o}_{j,\nu_1}, \ldots, \hat{u}^{\rm o}_{j,\nu_g})^\top \end{aligned}$$

with $\boldsymbol{\lambda}_i = \mathrm{diag}(\boldsymbol{\Lambda}_i)$ and $\boldsymbol{\sigma}_i = \mathrm{vech}(\boldsymbol{\Sigma}_i)$. Expressions for the elements of $\hat{\mathbf{u}}_j^{\rm o}$ are given by

$$\begin{aligned} \hat{u}^{\rm o}_{j,w_r} &= \frac{\hat{Z}_{rj}}{\hat{w}_r} - \frac{\hat{Z}_{gj}}{\hat{w}_g}, \\ \hat{\mathbf{u}}^{\rm o}_{j,\boldsymbol{\xi}_i} &= \hat{Z}_{ij}\left(\hat{\tau}_{ij}\hat{\mathbf{S}}^{\rm oo}_{ij}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i) - \hat{\mathbf{S}}^{\rm oo}_{ij}\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij}\right), \\ \hat{\mathbf{u}}^{\rm o}_{j,\boldsymbol{\sigma}_i} &= \mathrm{vech}\left(\hat{\mathbf{C}}_{ij} - \frac{1}{2}\mathrm{Diag}(\hat{\mathbf{C}}_{ij})\right), \\ \hat{\mathbf{u}}^{\rm o}_{j,\boldsymbol{\lambda}_i} &= \mathrm{Diag}\left[\hat{Z}_{ij}\left(\hat{\mathbf{S}}^{\rm oo}_{ij}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i)\hat{\boldsymbol{\eta}}_{ij}^\top - \hat{\mathbf{S}}^{\rm oo}_{ij}\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij}\right)\right], \\ \hat{u}^{\rm o}_{j,\nu_i} &= \frac{\hat{Z}_{ij}}{2}\left(\log\left(\frac{\hat{\nu}_i}{2}\right) + 1 - \mathrm{DG}\left(\frac{\hat{\nu}_i}{2}\right) + \hat{\kappa}_{ij} - \hat{\tau}_{ij}\right), \end{aligned}$$

where $\hat{\tau}_{ij}$, $\hat{\kappa}_{ij}$, $\hat{\boldsymbol{\eta}}_{ij}$ and $\hat{\boldsymbol{\Psi}}_{ij}$ are $\hat{\tau}_{ij}^{(k)}$, $\hat{\kappa}_{ij}^{(k)}$, $\hat{\boldsymbol{\eta}}_{ij}^{(k)}$ and $\hat{\boldsymbol{\Psi}}_{ij}^{(k)}$ in (3.9) with $\hat{\boldsymbol{\Theta}}^{(k)}$ replaced by $\hat{\boldsymbol{\Theta}}$ and $\hat{\mathbf{C}}_{ij} = \hat{Z}_{ij}\left(\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\mathbf{R}}_{ij}\hat{\boldsymbol{\Sigma}}_i^{-1} - \hat{\boldsymbol{\Sigma}}_i^{-1}\right)$ with

$$\begin{aligned} \hat{\mathbf{R}}_{ij} &= \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Sigma}}_i + (\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)\hat{\boldsymbol{\Psi}}_{ij}(\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)^\top \\ &\quad + (\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)\hat{\boldsymbol{\eta}}_{ij}(\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\rm oo}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)^\top \\ &\quad + (\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\rm oo}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)\hat{\boldsymbol{\eta}}_{ij}^\top(\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)^\top \\ &\quad + \hat{\tau}_{ij}(\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\rm oo}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)(\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\rm oo}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)^\top, \end{aligned}$$

and $\hat{\mathbf{D}}_{ij} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\rm oo}_{ij})$. If the degrees of freedom are assumed to be equal, say $\nu_1 = \cdots = \nu_g = \nu$, we have $\hat{u}^{\rm o}_{j,\nu} = \sum_{i=1}^{g} \hat{u}^{\rm o}_{j,\nu_i}$.

The detailed proof is shown in Appendix O.

3.3. Experimental results

3.3.1. Example 1: Simulated data

For learning MSTMIX models with missing observations, we conduct a simulation study to compare the misclassification rates under the MVNMIX, MVTMIX, MSNMIX and MSTMIX models. We first generated 500 observations from each of a three-component MSTMIX model at missing rates ranging from 10% to 40% with an increment of 10%. The presumed parameters of the MSTMIX model are given as

$$\begin{aligned}
\boldsymbol{\xi}_1 &= (0,3)^\top, \quad \boldsymbol{\xi}_2 = (3,0)^\top, \quad \boldsymbol{\xi}_3 = (-3,0)^\top,\\
\boldsymbol{\Sigma}_1 &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix},\\
\boldsymbol{\Lambda}_1 &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{\Lambda}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Lambda}_3 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},
\end{aligned}$$

with mixing probabilities $\omega_1 = \omega_2 = \omega_3 = 1/3$ and degrees of freedoms $\nu_1 = \nu_2 = \nu_3 = 4$.

The simulation aims to compare the performance of the four mixture models in terms of the average misclassification rates. For simplification of illustration, we fit a three component MSTMIX model with equal but unknown degrees of freedom. Now, we provide a simple way to obtain the initial values for the parameters, say $\hat{\boldsymbol{\Theta}}^{(0)} = \Big(\hat{w}_1^{(0)}, \ldots, \hat{w}_g^{(0)}, \hat{\boldsymbol{\xi}}_1^{(0)}, \ldots, \hat{\boldsymbol{\xi}}_g^{(0)}, \hat{\boldsymbol{\Sigma}}_1^{(0)}, \ldots, \hat{\boldsymbol{\Sigma}}_g^{(0)}, \hat{\boldsymbol{\Lambda}}_1^{(0)}, \ldots, \hat{\boldsymbol{\Lambda}}_g^{(0)}, \hat{\nu}_1^{(0)}, \ldots, \hat{\nu}_g^{(0)}\Big)$. The technique proceeds as follows: (i) Impute each missing value by averaging the sample mean of non-missing values of the corresponding variable. The imputed sample is denoted by $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_n)$. (ii) Perform a K-means clustering. (iii) Compute the zero-one component membership indicator $\hat{\mathbf{Z}}_j^{(0)} = \{\hat{Z}_{ij}^{(0)}\}_{i=1}^{g}$ according to the K-means clustering results. (iv) Generate a random number a from U(0,1). (v) The initial values of the parameters are chosen as

$$\begin{aligned}
\hat{\omega}_i^{(0)} &= \frac{\sum_{j=1}^n \hat{Z}_{ij}^{(0)}}{n}, \quad \hat{\mathbf{\Sigma}}_i^{(0)} = \mathbf{S}_i + (a-1)\mathrm{Diag}(\mathbf{S}_i), \\
\hat{\lambda}_{i,j}^{(0)} &= \pm\sqrt{(1-a)s_{i,jj}/(1-2/\pi)}, \quad \hat{\boldsymbol{\xi}}_i^{(0)} = \boldsymbol{\mu}_i - \sqrt{2/\pi}\hat{\mathbf{\Lambda}}_i^{(0)},
\end{aligned}$$

where $\hat{\lambda}_{i,j}^{(0)}$ can be either positive or negative, depending on the sign of the sample skewness of the ith variable with

$$\begin{aligned}
\boldsymbol{\mu}_i &= \frac{\sum_{j=1}^n \hat{Z}_{ij}^{(0)}\tilde{\mathbf{Y}}_j}{\sum_{j=1}^n \hat{Z}_{ij}^{(0)}}, \\
\mathbf{S}_i &= [s_{i,jk}] = \frac{\sum_{j=1}^n \hat{Z}_{ij}^{(0)}(\tilde{\mathbf{Y}}_j - \boldsymbol{\mu}_i)(\tilde{\mathbf{Y}}_j - \boldsymbol{\mu}_i)^\top}{\sum_{j=1}^n \hat{Z}_{ij}^{(0)}},
\end{aligned}$$

for $j,k = 1,\ldots,p$ and $i = 1,\ldots,g$. (vi) As for the degrees of freedoms, we set a relative large values, say $\hat{\nu}_1^{(0)} = \hat{\nu}_2^{(0)} = \hat{\nu}_3^{(0)} = 100$.

The MCECM algorithm is employed until the difference in successive values of the log-likelihood is less than a tolerance value, say 10^{-5}. For each iteration, we recorded the misclassification rates in Table 3.1. It is clear to see that the MSTMIX model outperforms the other three mixtures with varying proportions of missing values.

3.3.2. Example 2: The AIS data

The famous Australian Institute of Sport (AIS) data set was reported by Cook and Weisberg (1994). This data set has been analyzed by Azzalini and Dalla Valle (1996) and Azzalinin (2005), including 13 variables on 100 female and 102 male Australian athletes. We focused on the bivariate sample of two variables body mass index (BMI; kg/m^2) and body fat percentage (Bfat). The scatter plot of this

Table 3.1: A comparison of the average misclassification rates and the associated standard deviations in parentheses with various missing rates (Replications = 500)

r (%)	MVNMIX	MVTMIX	ν	MSNMIX	MSTMIX	ν
10	0.520	0.195	4.091	0.191	0.173	3.946
	(0.119)	(0.018)	(0.258)	(0.015)	(0.010)	(0.197)
20	0.491	0.241	4.108	0.259	0.219	3.900
	(0.108)	(0.027)	(0.520)	(0.046)	(0.015)	(0.255)
30	0.509	0.291	3.994	0.316	0.260	3.947
	(0.093)	(0.032)	(0.596)	(0.063)	(0.020)	(0.363)
40	0.515	0.337	3.824	0.372	0.309	3.909
	(0.070)	(0.038)	(0.679)	(0.065)	(0.018)	(0.298)

bivariate data set appears some outlying observations and a bimodal asymmetric mixture pattern.

This example is intended to illustrate the effect of robustness of the MSTMIX model after some perturbed values and missing values are introduced into the original data set simultaneously. The technique proceeds as follows: (i) generate one synthetic missing data by deleting at random from the bivariate data set under various missing rates (r = 10, 20, 30%). (ii) add one of the contaminated values $(\pm5, \pm10, \pm15, \pm20)$ to the first observation of the second variable. (iii) compute the ML estimates of each model and then classify each observation by Eq. (3.10).

Among these models, the average misclassification rates and the associated sample standard deviations for all possible combinations of missing rates and perturbations are listed in Table 3.2. As anticipated, the MSTMIX model provides better clustering results than the other three models for all cases.

Table 3.2: A comparison of misclassification rates when fitting models with a single perturbation for various proportions of missing values (Replications = 100). Values in parentheses are the associated standard deviations.

	Constants							
Missing rate	-20	-15	-10	-5	+5	+10	+15	+20
r=10%	MVNMIX							
	0.294	0.295	0.290	0.289	0.291	0.297	0.300	0.310
	(0.036)	(0.026)	(0.031)	(0.030)	(0.033)	(0.033)	(0.030)	(0.032)
	MVTMIX							
	0.221	0.225	0.220	0.220	0.220	0.221	0.220	0.220
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)
	MSNMIX							
	0.142	0.142	0.135	0.135	0.137	0.139	0.144	0.147
	(0.021)	(0.021)	(0.023)	(0.022)	(0.019)	(0.020)	(0.020)	(0.017)
	MSTMIX							
	0.108	0.116	0.109	0.109	0.113	0.114	0.112	0.111
	(0.015)	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)	(0.018)	(0.017)
r=20%	MVNMIX							
	0.369	0.361	0.356	0.350	0.354	0.359	0.368	0.371
	(0.036)	(0.034)	(0.032)	(0.034)	(0.038)	(0.034)	(0.036)	(0.036)
	MVTMIX							
	0.254	0.255	0.246	0.251	0.251	0.252	0.251	0.252
	(0.030)	(0.028)	(0.032)	(0.029)	(0.032)	(0.032)	(0.030)	(0.029)
	MSNMIX							
	0.172	0.167	0.168	0.164	0.165	0.169	0.174	0.182
	(0.027)	(0.025)	(0.025)	(0.023)	(0.027)	(0.028)	(0.027)	(0.030)
	MSTMIX							
	0.151	0.156	0.157	0.152	0.152	0.153	0.153	0.153
	(0.024)	(0.022)	(0.022)	(0.022)	(0.022)	(0.024)	(0.023)	(0.022)
r=30%	MVNMIX							
	0.461	0.449	0.444	0.453	0.444	0.454	0.456	0.464
	(0.050)	(0.053)	(0.053)	(0.045)	(0.050)	(0.053)	(0.055)	(0.052)
	MVTMIX							
	0.277	0.273	0.257	0.273	0.262	0.262	0.265	0.265
	(0.043)	(0.042)	(0.046)	(0.040)	(0.047)	(0.048)	(0.047)	(0.045)
	MSNMIX							
	0.211	0.219	0.223	0.211	0.212	0.214	0.229	0.220
	(0.026)	(0.035)	(0.042)	(0.036)	(0.035)	(0.035)	(0.041)	(0.033)
	MSTMIX							
	0.198	0.209	0.206	0.198	0.201	0.205	0.206	0.200
	(0.024)	(0.030)	(0.033)	(0.028)	(0.030)	(0.034)	(0.029)	(0.030)

3.4. Concluding remarks

We develop an efficient MCECM algorithm for the learning MSTMIX models under a missing information framework. Two binary auxiliary matrices are incorporated in the model that can substantially reduce the computational cost. Experimental results indicate that the MSTMIX model performs well for robust clustering when outlying observations and missing values both occur in the input data.

In the last two decades, many researchers (e.g. Diebolt and Robert 1994; Escobar and West 1995; Bensmail et al. 1997) have paid attention to the problem of Bayesian mixture modeling due to the popularity of MCMC techniques. More recently, the application of reversible jump MCMC (Green 1995; Brooks et al. 2003) has been shown to provide great power and flexibility by allowing simultaneous Bayesian estimation of both parameters and number of components (Richardson and Green 1997; Zhang et al. 2004; Dellaportas and Papageorgiou 2006). Therefore, it is worthwhile to investigate the applicability of a fully Bayesian treatment in this context.

Appendix

A. Proof of Theorem 1.1

The following lemma is used in proving the results.

Lemma A.1 If $\mathbf{\Lambda}_i$ is a diagonal matrix, then $\mathbf{O}_j\mathbf{\Lambda}_i\mathbf{O}_j^\top\mathbf{O}_j\mathbf{\Lambda}_i\mathbf{O}_j^\top = \mathbf{O}_j\mathbf{\Lambda}_i^2\mathbf{O}_j^\top$ and $\mathbf{M}_j\mathbf{\Lambda}_i\mathbf{M}_j^\top\mathbf{M}_j\mathbf{\Lambda}_i\mathbf{M}_j^\top = \mathbf{M}_j\mathbf{\Lambda}_i^2\mathbf{M}_j^\top$.

We partitioned the random vector $\mathbf{Y}_j$ and its parameters $\boldsymbol{\xi}_i$, $\mathbf{\Sigma}_i$ and $\mathbf{\Lambda}_i$ as follows:

$$\mathbf{Y}_j = \begin{bmatrix} \mathbf{Y}_j^{\mathrm{o}} \\ \mathbf{Y}_j^{\mathrm{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\mathbf{Y}_j \\ \mathbf{M}_j\mathbf{Y}_j \end{bmatrix}, \quad \boldsymbol{\xi}_i = \begin{bmatrix} \boldsymbol{\xi}_{ij}^{\mathrm{o}} \\ \boldsymbol{\xi}_{ij}^{\mathrm{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\xi}_i \\ \mathbf{M}_j\boldsymbol{\xi}_i \end{bmatrix},$$

$$\mathbf{\Sigma}_i = \begin{bmatrix} \mathbf{\Sigma}_{ij}^{\mathrm{oo}} & \mathbf{\Sigma}_{ij}^{\mathrm{om}} \\ \mathbf{\Sigma}_{ij}^{\mathrm{mo}} & \mathbf{\Sigma}_{ij}^{\mathrm{mm}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\mathbf{\Sigma}_i\mathbf{O}_j^\top & \mathbf{O}_j\mathbf{\Sigma}_i\mathbf{M}_j^\top \\ \mathbf{M}_j\mathbf{\Sigma}_i\mathbf{O}_j^\top & \mathbf{M}_j\mathbf{\Sigma}_i\mathbf{M}_j^\top \end{bmatrix},$$

and

$$\mathbf{\Lambda}_i = \begin{bmatrix} \mathbf{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\mathbf{\Lambda}_i\mathbf{O}_j^\top & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{M}_j\mathbf{\Lambda}_i\mathbf{M}_j^\top \end{bmatrix}.$$

By Lemma A.1

$$\mathbf{\Omega}_i = \mathbf{\Sigma}_i + \mathbf{\Lambda}_i^2 = \begin{bmatrix} \mathbf{O}_j(\mathbf{\Sigma}_i + \mathbf{\Lambda}_i^2)\mathbf{O}_j^\top & \mathbf{O}_j\mathbf{\Sigma}_i\mathbf{M}_j^\top \\ \mathbf{M}_j\mathbf{\Sigma}_i\mathbf{O}_j^\top & \mathbf{M}_j(\mathbf{\Sigma}_i + \mathbf{\Lambda}_i^2)\mathbf{M}_j^\top \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}_{ij}^{\mathrm{oo}} & \mathbf{\Omega}_{ij}^{\mathrm{om}} \\ \mathbf{\Omega}_{ij}^{\mathrm{mo}} & \mathbf{\Omega}_{ij}^{\mathrm{mm}} \end{bmatrix}.$$

Note that $\mathbf{\Sigma}_i$ and $\mathbf{\Lambda}_i$ are symmetric matrix. Thus, $\mathbf{\Omega}_i$, $\mathbf{\Omega}_{ij}^{\mathrm{oo}}$ and $\mathbf{\Omega}_{ij}^{\mathrm{mm}}$ are symmetric matrices and $\mathbf{\Omega}_{ij}^{\mathrm{om}^\top} = \mathbf{\Omega}_{ij}^{\mathrm{mo}}$.

Let $\mathbf{\Lambda}_i\mathbf{\Omega}_i^{-1} = \mathbf{B}_i = \begin{bmatrix}\mathbf{B}_{i1} & \mathbf{B}_{i2}\end{bmatrix}$. Thus, we have

$$\begin{aligned}
\mathbf{\Lambda}_i\mathbf{\Omega}_i^{-1} &= \begin{bmatrix}\mathbf{\Lambda}_{ij}^{\text{oo}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{ij}^{\text{mm}}\end{bmatrix}\begin{bmatrix}\mathbf{\Omega}_{ij}^{\text{oo}} & \mathbf{\Omega}_{ij}^{\text{om}} \\ \mathbf{\Omega}_{ij}^{\text{mo}} & \mathbf{\Omega}_{ij}^{\text{mm}}\end{bmatrix}^{-1} \\
&= \begin{bmatrix}\mathbf{\Lambda}_{ij}^{\text{oo}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{ij}^{\text{mm}}\end{bmatrix}\begin{bmatrix}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\left(\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}} + \mathbf{I}_{p_j^{\text{o}}}\right) & -\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}} \\ -\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}} & \mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{\Lambda}_{ij}^{\text{oo}}(\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}} + \mathbf{\Omega}_{ij}^{\text{oo}^{-1}}) & -\mathbf{\Lambda}_{ij}^{\text{oo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}} \\ -\mathbf{\Lambda}_{ij}^{\text{mm}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}} & \mathbf{\Lambda}_{ij}^{\text{mm}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{B}_{i1} & \mathbf{B}_{i2}\end{bmatrix},
\end{aligned}$$

where $\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}} = \mathbf{\Omega}_{ij}^{\text{mm}} - \mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}$. That is,

$$\begin{aligned}
\mathbf{B}_{i1} &= \begin{bmatrix}\mathbf{\Lambda}_{ij}^{\text{oo}}(\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}} + \mathbf{\Omega}_{ij}^{\text{oo}^{-1}}) \\ -\mathbf{\Lambda}_{ij}^{\text{mm}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{\Lambda}_{ij}^{\text{oo}}(\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}} + \mathbf{I}_{p_j^{\text{o}}}) \\ -\mathbf{\Lambda}_{ij}^{\text{mm}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\mathbf{\Omega}_{ij}^{\text{mo}}\end{bmatrix}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}
\end{aligned}$$

and

$$\mathbf{B}_{i2} = \begin{bmatrix}-\mathbf{\Lambda}_{ij}^{\text{oo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}} \\ \mathbf{\Lambda}_{ij}^{\text{mm}}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}\end{bmatrix} = \begin{bmatrix}-\mathbf{\Lambda}_{ij}^{\text{oo}}\mathbf{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{\Omega}_{ij}^{\text{om}} \\ \mathbf{\Lambda}_{ij}^{\text{mm}}\end{bmatrix}\mathbf{\Omega}_{ij}^{\text{mm}\cdot\text{o}^{-1}}.$$

From the above calculation, we have the following results

$$
\begin{aligned}
& \mathbf{B}_{i1} + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
= & \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}(\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \mathbf{I}_{p_j^{\mathrm{o}}}) \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
& + \begin{bmatrix} -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
= & \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} - \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
= & \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} \\ \mathbf{0}_{p_j^{\mathrm{m}} \times p_j^{\mathrm{o}}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}},
\end{aligned}
$$

$$
\begin{aligned}
& \boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Lambda}_i \\
= & \begin{bmatrix} \mathbf{B}_{i1} & \mathbf{B}_{i2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \\
= & \begin{bmatrix} \mathbf{B}_{i1}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{B}_{i2}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \\
= & \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\left(\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \mathbf{I}_{p_j^{\mathrm{o}}}\right)\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix},
\end{aligned}
\tag{A.1}
$$

and

$$
\begin{aligned}
&\mathbf{B}_{i2}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top} \\
= &\begin{bmatrix} -\mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{om}} \\ \mathbf{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}}(\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}})^{\top} \begin{bmatrix} -(\mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{om}})^{\top} & \mathbf{\Lambda}_{ij}^{\mathrm{mm}^{\top}} \end{bmatrix} \\
= &\begin{bmatrix} -\mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{om}} \\ \mathbf{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}} \begin{bmatrix} -(\mathbf{\Omega}_{ij}^{\mathrm{mo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{oo}}) & \mathbf{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \\
= &\begin{bmatrix} \mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{om}}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{mo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{oo}} & -\mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{om}}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{mm}} \\ -\mathbf{\Lambda}_{ij}^{\mathrm{mm}}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_{ij}^{\mathrm{mo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{\Lambda}_{ij}^{\mathrm{mm}}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix}.
\end{aligned}
\tag{A.2}
$$

Since

$$\mathbf{\Delta}_i + \mathbf{B}_{i2}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top} = \mathbf{I}_p - \mathbf{\Lambda}_i\mathbf{\Omega}_i^{-1}\mathbf{\Lambda}_i + \mathbf{B}_{i2}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top}\ , \tag{A.3}$$

we substituted $(A.1)$, $(A.2)$ to $(A.3)$ to obtain

$$
\begin{aligned}
\mathbf{\Delta}_i + \mathbf{B}_{i2}\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top} &= \mathbf{I}_p - \begin{bmatrix} \mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{m}}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_{p_j^{\mathrm{o}}} - \mathbf{\Lambda}_{ij}^{\mathrm{oo}}\mathbf{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{I}_{p_j^{\mathrm{m}}} \end{bmatrix}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&f(\mathbf{Y}_j^{\mathrm{o}}, \mathbf{Y}_j^{\mathrm{m}} | Z_{ij} = 1, \mathbf{\Theta}) \\
= &\ 2^p \phi_p(\mathbf{Y}_j \mid \boldsymbol{\xi}_i, \mathbf{\Omega}_i)\Phi_p\Big(\mathbf{\Lambda}_i\mathbf{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i) \mid \mathbf{\Delta}_i\Big) \\
= &\ 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_{ij}^{\mathrm{o}}, \mathbf{\Omega}_{ij}^{\mathrm{oo}})\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}} | \boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}})\Phi_p\Big(\begin{bmatrix} \mathbf{B}_{i1} & \mathbf{B}_{i2} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}} \\ \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}} \end{bmatrix} \Big| \mathbf{\Delta}_i\Big) \\
= &\ 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_{ij}^{\mathrm{o}}, \mathbf{\Omega}_{ij}^{\mathrm{oo}})\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}} | \boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}})\Phi_p\Big(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}}) \mid \mathbf{\Delta}_i\Big),
\end{aligned}
$$

where $\boldsymbol{\xi}_{ij}^{\mathrm{m}\cdot\mathrm{o}} = \boldsymbol{\xi}_{ij}^{\mathrm{m}} + \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})$.

The marginal density of $\mathbf{Y}_j^{\mathrm{o}}$ is given by

$$\begin{aligned} f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\boldsymbol{\Theta}) &= \int f(\mathbf{Y}_j^{\mathrm{o}},\mathbf{Y}_j^{\mathrm{m}}|Z_{ij}=1,\boldsymbol{\Theta})\ d\mathbf{Y}_j^{\mathrm{m}} \\ &= \int 2^p\phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}})\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m}\cdot\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{mm}\cdot\mathrm{o}}) \\ &\quad \times\Phi_p\Big(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}}-\boldsymbol{\xi}_{ij}^{\mathrm{m}})\mid\boldsymbol{\Delta}_i\Big)\ d\mathbf{Y}_j^{\mathrm{m}} \\ &= 2^p\phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}})\int\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m}\cdot\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{mm}\cdot\mathrm{o}}) \\ &\quad \times\Phi_p\Big(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}}-\boldsymbol{\xi}_{ij}^{\mathrm{m}})\mid\boldsymbol{\Delta}_i\Big)\ d\mathbf{Y}_j^{\mathrm{m}}. \end{aligned}$$

Let $\mathbf{z} = \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}\cdot\mathrm{o}}$, then $\mathbf{Y}_j^{\mathrm{m}} = \mathbf{z} + \boldsymbol{\xi}_{ij}^{\mathrm{m}\cdot\mathrm{o}} = \mathbf{z} + \boldsymbol{\xi}_{ij}^{\mathrm{m}} + \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})$. Thus, we have

$$\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m}\cdot\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{mm}\cdot\mathrm{o}}) = \phi_{p_j^{\mathrm{m}}}(\mathbf{z}|\mathbf{0},\boldsymbol{\Omega}_{ij}^{\mathrm{mm}\cdot\mathrm{o}})$$

and

$$\begin{aligned} &\Phi_p\Big(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}}-\boldsymbol{\xi}_{ij}^{\mathrm{m}})\mid\boldsymbol{\Delta}_i\Big) \\ =\ &\Phi_p\Big(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}(\mathbf{z}+\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\mid\boldsymbol{\Delta}_i\Big) \\ =\ &\Phi_p\Big((\mathbf{B}_{i1}+\mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}\mathbf{z}\mid\boldsymbol{\Delta}_i\Big). \end{aligned}$$

By Lemma 2.1 of Arellano-Valle and Genton (2005), we have

$$\begin{aligned} &f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\boldsymbol{\Theta}) \\ =\ &2^p\phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}) \\ &\times\int\phi_{p_j^{\mathrm{m}}}(\mathbf{z}|\mathbf{0},\boldsymbol{\Omega}_{ij}^{\mathrm{mm}\cdot\mathrm{o}})\Phi_p\big((\mathbf{B}_{i1}+\mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}\mathbf{z}\big|\boldsymbol{\Delta}_i\big)\ d\mathbf{z} \\ =\ &2^p\phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}})\ \mathrm{E}_{\mathbf{z}}\Big\{\Phi_p\big((\mathbf{B}_{i1}+\mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\mathbf{B}_{i2}\mathbf{Z}\big|\boldsymbol{\Delta}_i\big)\Big\} \\ =\ &2^p\phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}})\ \Phi_p\big((\mathbf{B}_{i1}+\mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\big|\boldsymbol{\Delta}_i+\mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mm}\cdot\mathrm{o}}\mathbf{B}_{i2}^{\top}\big) \\ =\ &2^{p_j^{\mathrm{o}}}\phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}})\ \Phi_{p_j^{\mathrm{o}}}\big(\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}}-\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\big). \end{aligned}$$

Thus, $\mathbf{Y}_j^{\mathrm{o}} \mid (Z_{ij}=1,\mathbf{\Theta}) \sim SN_{p_j^{\mathrm{o}}}\Big(\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \mathbf{\Sigma}_{ij}^{\mathrm{oo}}, \mathbf{\Lambda}_{ij}^{\mathrm{oo}}\Big)$. It implies that

$$f(\mathbf{Y}_j^{\mathrm{o}}|\mathbf{\Theta}) = \sum_{i=1}^{g} f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\mathbf{\Theta})p(Z_{ij}=1) = \sum_{i=1}^{g} w_i\psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \mathbf{\Sigma}_{ij}^{\mathrm{oo}}, \mathbf{\Lambda}_{ij}^{\mathrm{oo}}).$$

(b) By virtue of $\phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i,\mathbf{\Omega}_i) = \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\mathbf{\Omega}_{ij}^{\mathrm{oo}})\phi_{p-p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}},\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}})$, see Theorem 2.5.1 of Anderson (2003), we can deduce that

$$\begin{aligned} f(\mathbf{Y}_j^{\mathrm{m}} \mid \mathbf{Y}_j^{\mathrm{o}}) &= \frac{f(\mathbf{Y}_j^{\mathrm{m}},\mathbf{Y}_j^{\mathrm{o}})}{f(\mathbf{Y}_j^{\mathrm{o}})} \\ &= \frac{2^p\sum_{i=1}^{g} w_i\phi_p(\mathbf{Y}_j \mid \boldsymbol{\xi}_i,\mathbf{\Omega}_i)\Phi_p(\mathbf{\Lambda}_i\mathbf{\Omega}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i) \mid \mathbf{\Delta}_i)}{\sum_{i=1}^{g} w_i\psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} \mid \boldsymbol{\xi}_{ij}^{\mathrm{o}},\mathbf{\Sigma}_{ij}^{\mathrm{oo}},\mathbf{\Lambda}_{ij}^{\mathrm{oo}})} \\ &= 2^p\sum_{i=1}^{g}\tilde{w}_{ij}\phi_{p-p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{m}} \mid \boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}},\mathbf{\Omega}_{ij}^{\mathrm{mm\cdot o}})\Phi_p(\mathbf{\Lambda}_i\mathbf{\Omega}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i) \mid \mathbf{\Delta}_i). \end{aligned}$$

B. Proof of Theorem 1.2(c)

From equation (1.9), we have the following densities

$$f(\mathbf{Y}_j^{\rm o}|\boldsymbol{\gamma}_j, Z_{ij}=1, \boldsymbol{\Theta}) = (2\pi)^{\frac{-p_j^{\rm o}}{2}}|\boldsymbol{\Sigma}_{ij}^{\rm oo}|^{-\frac{1}{2}}\exp\Big(- (\mathbf{Y}_j^{\rm o} - \boldsymbol{\mu}_{ij}^{\rm o})^\top \boldsymbol{\Sigma}_{ij}^{\rm oo^{-1}} (\mathbf{Y}_j^{\rm o} - \boldsymbol{\mu}_{ij}^{\rm o})/2\Big)$$

and

$$f(\boldsymbol{\gamma}_j|Z_{ij}=1) \;=\; 2^p(2\pi)^{\frac{-p}{2}}\exp\big(-\boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j/2\big)I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j).$$

Thus, we have

$$\begin{aligned}
& f(\mathbf{Y}_j^{\rm o}, \boldsymbol{\gamma}_j|Z_{ij}=1, \boldsymbol{\Theta}) \\
= \;& f(\mathbf{Y}_j^{\rm o}|\boldsymbol{\gamma}_j, Z_{ij}=1, \boldsymbol{\Theta}) f(\boldsymbol{\gamma}_j|Z_{ij}=1) \\
\propto \;& |\boldsymbol{\Sigma}_{ij}^{\rm oo}|^{-\frac{1}{2}}\exp\Big\{ - \Big((\mathbf{Y}_j^{\rm o} - \boldsymbol{\mu}_{ij}^{\rm o})^\top \boldsymbol{\Sigma}_{ij}^{\rm oo^{-1}} (\mathbf{Y}_j^{\rm o} - \boldsymbol{\mu}_{ij}^{\rm o}) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j\Big)/2\Big\} I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j) \\
\propto \;& |\boldsymbol{\Sigma}_{ij}^{\rm oo}|^{-\frac{1}{2}}\exp\Big\{ - \Big((\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i))^\top \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\rm oo}\boldsymbol{\Lambda}_i(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j\Big)/2\Big\} \\
& \times I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j),
\end{aligned}$$

where

$$\begin{aligned}
& (\mathbf{Y}_j^{\rm o} - \boldsymbol{\mu}_{ij}^{\rm o})^\top \boldsymbol{\Sigma}_{ij}^{\rm oo^{-1}} (\mathbf{Y}_j^{\rm o} - \boldsymbol{\mu}_{ij}^{\rm o}) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j \\
= \;& (\mathbf{O}_j\mathbf{Y}_j - \mathbf{O}_j(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\gamma_j))^\top \boldsymbol{\Sigma}_{ij}^{\rm oo^{-1}} (\mathbf{O}_j\mathbf{Y}_j - \mathbf{O}_j(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\gamma_j)) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j \\
= \;& (\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\gamma_j)^\top \mathbf{O}_j^\top \boldsymbol{\Sigma}_{ij}^{\rm oo^{-1}} \mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\gamma_j) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j \\
= \;& (\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i))^\top \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\rm oo}\boldsymbol{\Lambda}_i(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j.
\end{aligned}$$

To prove the identity

$$\begin{aligned}
& (\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i))^\top \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\rm oo}\boldsymbol{\Lambda}_i(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)) + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j \\
= \;& \big(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi})\big)^\top \big(\mathbf{I}_p - \boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\rm oo}\boldsymbol{\Lambda}_i\big)^{-1} \big(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi})\big) \\
& + (\mathbf{Y}_j - \boldsymbol{\xi})^\top \mathbf{C}_j^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi}),
\end{aligned}$$

where $\mathbf{C}_j^{\text{oo}} = \mathbf{O}_j^\top \boldsymbol{\Omega}_j^{\text{oo}^{-1}} \mathbf{O}_j$.

We need the following lemma.

Lemma B.1 Let $\mathbf{x}$, $\mathbf{a}$ and $\mathbf{b}$ be $p \times 1$ vectors, and let $\mathbf{Q}_1$ and $\mathbf{Q}_2$ be $p \times p$ symmetric matrices such that $(\mathbf{Q}_1 + \mathbf{Q}_2)^{-1}$ exists. Then

$$\begin{aligned} & (\mathbf{x}-\mathbf{a})^\top \mathbf{Q}_1(\mathbf{x}-\mathbf{a}) + (\mathbf{x}-\mathbf{b})^\top \mathbf{Q}_2(\mathbf{x}-\mathbf{b}) \\ = \; & (\mathbf{x}-\hat{\mathbf{x}})^\top \mathbf{Q}(\mathbf{x}-\hat{\mathbf{x}}) + (\mathbf{a}-\mathbf{b})^\top \mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2(\mathbf{a}-\mathbf{b}), \end{aligned}$$

where $\hat{\mathbf{x}} = \mathbf{Q}^{-1}(\mathbf{Q}_1\mathbf{a} + \mathbf{Q}_2\mathbf{b})$ and $\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2$.

From Lemma B.1, we take $\mathbf{x} = \boldsymbol{\tau}_j$, $\mathbf{a} = \boldsymbol{\Lambda}_i^{-1}\mathbf{z}_{ij}$, $\mathbf{b} = \mathbf{0}$, $\mathbf{Q}_1 = \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i$ and $\mathbf{Q}_2 = \mathbf{I}_p$.

Therefore,

$$\begin{aligned} \mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2 &= \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i + \mathbf{I}_p \\ &= \boldsymbol{\Lambda}_i \mathbf{O}_j^\top \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} \mathbf{O}_j \boldsymbol{\Lambda}_i + \mathbf{I}_p \\ &= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} (\mathbf{O}_j\boldsymbol{\Lambda}_i) + \mathbf{I}_p, \end{aligned}$$

$$\begin{aligned} \mathbf{Q}^{-1} &= \Big(\mathbf{I}_p + (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} (\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big)^{-1} \\ &= \mathbf{I}_p - \mathbf{I}_p(\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \left(\boldsymbol{\Sigma}_{ij}^{\text{oo}} + (\mathbf{O}_j\boldsymbol{\Lambda}_i)\mathbf{I}_p(\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\right)^{-1} (\mathbf{O}_j\boldsymbol{\Lambda}_i)\mathbf{I}_p \\ &= \mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \left(\mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top + \mathbf{O}_j\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i\mathbf{O}_j^\top\right)^{-1} (\mathbf{O}_j\boldsymbol{\Lambda}_i) \\ &= \mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \left(\mathbf{O}_j(\boldsymbol{\Sigma}_i + \boldsymbol{\Lambda}_i^2)\mathbf{O}_j^\top\right)^{-1} (\mathbf{O}_j\boldsymbol{\Lambda}_i) \\ &= \mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}} (\mathbf{O}_j\boldsymbol{\Lambda}_i) \\ &= \mathbf{I}_p - \boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i, \end{aligned} \tag{B.1}$$

$$
\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{Q}^{-1}(\mathbf{Q}_1\mathbf{a} + \mathbf{Q}_2\mathbf{b}) \\
&= \Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big)\,\boldsymbol{\Lambda}_i\mathbf{S}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= \Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big)\,\boldsymbol{\Lambda}_i\mathbf{S}_{ij}^{\text{oo}}(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= \Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big)\,(\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}\mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= \Big((\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top(\mathbf{O}_j\boldsymbol{\Lambda}_i)\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}\Big)\,\mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\left(\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} - \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i^2\mathbf{O}_j^\top)\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}\right)\mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\left(\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} - \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\boldsymbol{\Omega}_{ij}^{\text{oo}} - \boldsymbol{\Sigma}_{ij}^{\text{oo}})\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}\right)\mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\left(\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} - \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} + \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}\right)\mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}\mathbf{O}_j(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= \boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\text{oo}}(\mathbf{Y}_j - \boldsymbol{\xi}_i), \qquad \text{(B.2)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2 &= \boldsymbol{\Lambda}_i\mathbf{S}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i\Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big)\mathbf{I}_p \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i)\Big) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\Big(\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} - \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i^2\mathbf{O}_j^\top)\boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}\Big)(\mathbf{O}_j\boldsymbol{\Lambda}_i) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\Big(\boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} - \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}}(\boldsymbol{\Omega}_{ij}^{\text{oo}} - \boldsymbol{\Sigma}_{ij}^{\text{oo}})\boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}\Big)(\mathbf{O}_j\boldsymbol{\Lambda}_i) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda}_i)^\top\boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}_i) \\
&= \boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
(\mathbf{a}-\mathbf{b})^\top\mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2(\mathbf{a}-\mathbf{b}) &= (\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i))^\top\boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i(\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)) \\
&= (\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i) \\
&= (\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\text{oo}}(\mathbf{Y}_j - \boldsymbol{\xi}_i). \qquad \text{(B.3)}
\end{aligned}
$$

Substituting Equations $(B.1)$, $(B.2)$, and $(B.3)$ into Lemma B.1, the identity is proved.

Moreover, the equation $|\boldsymbol{\Sigma}_{ij}^{\text{oo}}| = |\boldsymbol{\Omega}_{ij}^{\text{oo}}||\mathbf{I}_p - \boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i|$ holds.

The posterior distribution of $\boldsymbol{\gamma}_j$ given $\mathbf{Y}_j^{\text{o}}$ and $Z_{ij} = 1$ is given by

$$\begin{aligned}
& f(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\text{o}}, Z_{ij} = 1, \boldsymbol{\Theta}) \\
= \quad & \frac{f(\mathbf{Y}_j^{\text{o}}, \boldsymbol{\gamma}_j|Z_{ij} = 1, \boldsymbol{\Theta})}{f(\mathbf{Y}_j^{\text{o}}|Z_{ij} = 1, \boldsymbol{\Theta})} \propto f(\mathbf{Y}_j^{\text{o}}, \boldsymbol{\gamma}_j|Z_{ij} = 1, \boldsymbol{\Theta}) \\
\propto \quad & |\boldsymbol{\Sigma}_{ij}^{\text{oo}}|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}\Big((\mathbf{Y}_j^{\text{o}} - \boldsymbol{\mu}_{ij}^{\text{o}})^\top \boldsymbol{\Sigma}_{ij}^{\text{oo}^{-1}} (\mathbf{Y}_j^{\text{o}} - \boldsymbol{\mu}_{ij}^{\text{o}}) + \boldsymbol{\gamma}_j^\top \boldsymbol{\gamma}_j\Big)\Big\} I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j), \\
\propto \quad & |\boldsymbol{\Delta}_{ij}^*|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}\Big((\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*)^\top \boldsymbol{\Delta}_{ij}^{*^{-1}} (\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*)\Big)\Big\} I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j),
\end{aligned}$$

where $\boldsymbol{\Delta}_{ij}^* = \mathbf{I}_p - \boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i$, $\mathbf{q}_{ij}^* = \boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\text{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i)$, and $\mathbf{C}_{ij}^{\text{oo}} = \mathbf{O}_j^\top \boldsymbol{\Omega}_{ij}^{\text{oo}^{-1}} \mathbf{O}_j$. It implies that

$$\boldsymbol{\gamma}_j|(\mathbf{Y}_j^{\text{o}}, Z_{ij} = 1) \sim TN_p\big(\boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\text{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i), \mathbf{I}_p - \boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i, \mathbb{R}_+^p\big).$$

C. Proof of Lemma 1.3

Let $\hat{Z}_{ij}^{(k)} = \mathrm{E}(Z_{ij}|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)})$, $\hat{\mathbf{Y}}_{ij}^{(k)} = \mathrm{E}(\mathbf{Y}_j|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)})$ and $\boldsymbol{\Omega}_{ij}^{(k)} = \mathrm{E}(Z_{ij}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)})$. Then, it can be shown that

$$\hat{Z}_{ij}^{(k)} = \mathrm{P}(Z_{ij}=1|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) = \frac{\hat{w}_i^{(k)}\psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)}, \hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}(k)}, \hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}(k)})}{\sum_{i=1}^{g}\hat{w}_i^{(k)}\psi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)}, \hat{\boldsymbol{\Sigma}}_{ij}^{\mathrm{oo}(k)}, \hat{\boldsymbol{\Lambda}}_{ij}^{\mathrm{oo}(k)})}.$$

Since $\mathbf{Y}_j = \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{Y}_j^{\mathrm{m}}$ and $\mathbf{O}_j^\top\mathbf{O}_j(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) = \mathbf{0}$, we have

$$\begin{aligned}
\hat{\mathbf{Y}}_{ij}^{(k)} &= \mathrm{E}(\mathbf{Y}_j|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&= \mathrm{E}(\mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{Y}_j^{\mathrm{m}}|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&= \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathrm{E}(\mathbf{Y}_j^{\mathrm{m}}|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&= \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathrm{E}\left(\mathrm{E}(\mathbf{Y}_j^{\mathrm{m}}|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \boldsymbol{\gamma}_j, \hat{\boldsymbol{\Theta}}^{(k)})\big|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\right) \\
&= \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top \\
&\quad \times\mathrm{E}\left(\mathbf{M}_j\left(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j + \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j)\right)\big|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\right) \\
&= \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{M}_j\Big(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\mathrm{E}(\boldsymbol{\gamma}_j|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&\quad +\hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)}\mathrm{E}(\boldsymbol{\gamma}_j|Z_{ij}=1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}))\Big) \\
&= \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{M}_j\Big(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)} + \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)})\Big) \\
&= \mathbf{O}_j^\top\mathbf{O}_j\mathbf{Y}_j + (\mathbf{I}_p - \mathbf{O}_j^\top\mathbf{O}_j)(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}) \\
&\quad +(\mathbf{I}_p - \mathbf{O}_j^\top\mathbf{O}_j)\hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j \\
&= \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}),
\end{aligned}$$

and

$$
\begin{aligned}
& \operatorname{Cov}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \operatorname{Cov}(\mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \operatorname{Cov}(\mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \mathrm{E}\big(\operatorname{Cov}(\mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \boldsymbol{\gamma}_j, \hat{\boldsymbol{\Theta}}^{(k)}) \big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
& + \operatorname{Cov}\big(\mathrm{E}(\mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \boldsymbol{\gamma}_j, \hat{\boldsymbol{\Theta}}^{(k)}) \big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \; & \mathrm{E}\big(\mathbf{M}_j^\top \operatorname{Cov}(\mathbf{Y}_j^{\mathrm{m}} | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \boldsymbol{\gamma}_j, \hat{\boldsymbol{\Theta}}^{(k)}) \mathbf{M}_j \big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
& + \operatorname{Cov}\big(\mathbf{M}_j^\top \mathrm{E}(\mathbf{Y}_j^{\mathrm{m}} | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \boldsymbol{\gamma}_j, \hat{\boldsymbol{\Theta}}^{(k)}) - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j \big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \; & \mathbf{M}_j^\top \mathbf{M}_j (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}_i^{(k)} \mathbf{M}_j^\top \mathbf{M}_j + \operatorname{Cov}\big(\mathbf{M}_j^\top \mathbf{M}_j \big(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)} \boldsymbol{\gamma}_j \\
& + \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)} \boldsymbol{\gamma}_j)\big) - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j \big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \; & (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}_i^{(k)} + \operatorname{Cov}\big(\mathbf{M}_j^\top \mathbf{M}_j \big(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)} \boldsymbol{\gamma}_j \\
& + \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)} \boldsymbol{\gamma}_j)\big) - \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j \big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \; & (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}_i^{(k)} + \operatorname{Cov}\Big(\big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big) \boldsymbol{\gamma}_j \Big| Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}\Big) \\
= \; & (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}_i^{(k)} + \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big) \\
& \times \operatorname{Cov}(\boldsymbol{\gamma}_j | Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big)^\top \\
= \; & (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}_i^{(k)} \\
& + \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big) (\hat{\boldsymbol{\Psi}}_{ij}^{(k)} - \hat{\boldsymbol{\eta}}_{ij}^{(k)} \hat{\boldsymbol{\eta}}_{ij}^{(k)\top}) \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big)^\top.
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\boldsymbol{\Omega}_{ij}^{(k)} &= \mathrm{E}(Z_{ij}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&= \mathrm{E}(Z_{ij}|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)})\mathrm{E}((\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top|Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&= \mathrm{E}(Z_{ij}|\mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \Big[\mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&\quad \times\mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)})^\top \\
&\quad +\mathrm{Cov}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|Z_{ij} = 1, \mathbf{Y}^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}^{(k)})\Big] \\
&= \hat{Z}_{ij}^{(k)} \Big[(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} + (\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})^\top \\
&\quad +\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)(\hat{\boldsymbol{\Psi}}_{ij}^{(k)} - \hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top\Big],
\end{aligned}$$

where $\hat{\mathbf{A}}_{ij}^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)}$.

D. Proof of the M-steps

CM-step 1:

The mixing probabilities w_i's are subject to the constraint $\sum_{i=1}^{g} w_i = 1$. Define $L = \sum_{i=1}^{g} \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \log(w_i) - \lambda(\sum_{i=1}^{g} w_i - 1)$, where λ is the Lagrange multiplier. Let $dL/dw_i = 0$ and $dL/d\lambda = 0$, we have $w_i = \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)}/\lambda$ and $\sum_{i=1}^{g} w_i = 1$. It follows that $\lambda = n$ and the estimate of w_i at the $(k+1)$th iteration is

$$\hat{w}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)}}{n}.$$

CM-step 2:

Taking the partial derivative of $Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})$ with respect to $\boldsymbol{\xi}_i$ and setting it to zero yields

$$\begin{aligned}
\frac{\partial Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})}{\partial \boldsymbol{\xi}_i} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\text{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\sum_{j=1}^{n}\text{tr}\left(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Omega}_{ij}^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\sum_{j=1}^{n}\text{tr}\big(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})^\top\big) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\sum_{j=1}^{n}\big(\hat{Z}_{ij}^{(k)}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})^\top\boldsymbol{\Sigma}_i^{-1}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})\big) \\
&= \sum_{j=1}^{n}\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)}) \\
&= \boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)}) \\
&= 0.
\end{aligned}$$

With $\boldsymbol{\Lambda}_i$ fixed at $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, solving this equation gives

$$\hat{\boldsymbol{\xi}}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\mathbf{Y}}_{ij}^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}}{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}}.$$

CM-step 3:

Recalling the differential formulae with respect to a matrix, one has

$$\frac{\partial \log|\mathbf{A}|}{\partial \mathbf{A}} = 2\mathbf{A}^{-1} - \mathrm{Diag}(\mathbf{A}^{-1}) \quad \text{and} \quad \frac{\partial \mathrm{tr}(\mathbf{AB})}{\partial \mathbf{A}} = 2\mathbf{B} - \mathrm{Diag}(\mathbf{B}),$$

for any given symmetric matrix $\mathbf{A}$ and $\mathbf{B}$. Taking the partial derivative of $Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})$ with respect to $\boldsymbol{\Sigma}_i^{-1}$ and setting it to zero yields

$$\begin{aligned}
\frac{\partial Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})}{\partial \boldsymbol{\Sigma}_i^{-1}} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i^{-1}}\left[\log|\boldsymbol{\Sigma}_i|\left(\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right)\right] \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i^{-1}}\left[-\log|\boldsymbol{\Sigma}_i^{-1}|\left(\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right)\right] \\
&= -\frac{1}{2}\left[(-2\boldsymbol{\Sigma}_i + \mathrm{Diag}(\boldsymbol{\Sigma}_i))\left(\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\right) + 2\left(\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right)\right. \\
&\quad \left. -\mathrm{Diag}\left(\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right)\right] \\
&= 0.
\end{aligned}$$

Fix $\boldsymbol{\xi}_i$ at $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i$ at $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, solving this equation obtains

$$\hat{\boldsymbol{\Sigma}}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}}{\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}},$$

where $\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}$ is $\boldsymbol{\Omega}_{ij}^{(k)}$ in Equation (1.16) with $\boldsymbol{\xi}_i$ replaced by $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i$ replaced by $\hat{\boldsymbol{\Lambda}}_i^{(k)}$.

CM-step 4:

Let $\hat{\mathbf{A}}_{ij}^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)}$. The sum of the term in the $Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})$ function can be rearranged by

$$\begin{aligned}
&\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)})^\top\Big) \\
= \;&\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i)(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top\Big) - 2\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}(\hat{\mathbf{Y}}_{ij}^{(k)} - \boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\boldsymbol{\Lambda}_i^\top\Big) \\
&+\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\boldsymbol{\Lambda}_i^\top\Big).
\end{aligned}$$

Now, taking the partial derivative of $Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})$ with respect to $\boldsymbol{\lambda}_i$ and setting it to zero yields

$$
\begin{aligned}
&\frac{\partial Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})}{\partial \boldsymbol{\lambda}_i}\\
= &-\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right)\\
= &-\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\sum_{j=1}^{n}\mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Omega}_{ij}^{(k)}\right)\\
= &-\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\Big[-2\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}(\hat{\mathbf{Y}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\boldsymbol{\Lambda}_i^\top\Big)+\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\boldsymbol{\Lambda}_i^\top\Big)\\
&+\mathrm{tr}\Big(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Lambda}_i-\hat{\mathbf{A}}_{ij}^{(k)})(\hat{\boldsymbol{\Psi}}_{ij}^{(k)}-\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})(\boldsymbol{\Lambda}_i-\hat{\mathbf{A}}_{ij}^{(k)})^\top\Big)\Big]\\
= &-\frac{1}{2}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\Big[-2\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}(\hat{\mathbf{Y}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\Big)+2\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\Big)\\
&+2\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Lambda}_i-\hat{\mathbf{A}}_{ij}^{(k)})(\hat{\boldsymbol{\Psi}}_{ij}^{(k)}-\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})\Big)\Big]\\
= &\ \mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{Y}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\Big)-\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}\Big)\\
&-\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}(\hat{\boldsymbol{\Psi}}_{ij}^{(k)}-\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})\Big)\\
&+\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\mathbf{A}}_{ij}^{(k)}(\hat{\boldsymbol{\Psi}}_{ij}^{(k)}-\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})\Big)\\
= &\ \mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\Big[(\hat{\mathbf{Y}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}+\hat{\mathbf{A}}_{ij}^{(k)}(\hat{\boldsymbol{\Psi}}_{ij}^{(k)}-\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})\Big]\Big)\\
&-\mathrm{Diag}\Big(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\Big)\\
= &\ \Big(\boldsymbol{\Sigma}_i^{-1}\odot\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\Big[(\hat{\mathbf{Y}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}+\hat{\mathbf{A}}_{ij}^{(k)}(\hat{\boldsymbol{\Psi}}_{ij}^{(k)}-\hat{\boldsymbol{\eta}}_{ij}^{(k)}\hat{\boldsymbol{\eta}}_{ij}^{(k)^\top})\Big]^\top\Big)\mathbf{1}_p\\
&-\Big(\boldsymbol{\Sigma}_i^{-1}\odot\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\Big)\boldsymbol{\lambda}_i.
\end{aligned}
$$

Note that the last equality follows from the fact that $\mathrm{Diag}(\mathbf{ADB}) = (\mathbf{A}\odot\mathbf{B}^\top)\,\mathrm{Diag}(\mathbf{D})$ ($\mathbf{D}$ is a diagonal matrix) with $\mathrm{Diag}(\cdot)$ denoting the diagonal elements of a square matrix and '$\odot$' the elementwise product of two matrices. With $\boldsymbol{\xi}_i = \hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Sigma}_i = \hat{\boldsymbol{\Sigma}}_i^{(k+1)}$, setting $\partial Q_2(\Upsilon|\hat{\boldsymbol{\Theta}}^{(k)})/\partial \boldsymbol{\lambda}_i = 0$ yields the estimate of $\boldsymbol{\lambda}_i$

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_i^{(k+1)} &= \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^{-1} \\ &\times \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \Big((\hat{\boldsymbol{\Psi}}_{ij}^{(k)} - \hat{\boldsymbol{\eta}}_{ij}^{(k)} \hat{\boldsymbol{\eta}}_{ij}^{(k)^\top}) \hat{\mathbf{A}}_{ij}^{\top(k)} + \hat{\boldsymbol{\eta}}_{ij}^{(k)} (\hat{\mathbf{Y}}_{ij}^{(k)} - \hat{\boldsymbol{\xi}}_i^{(k+1)})^\top \Big) \right) \mathbf{1}_p.\end{aligned}$$

E. Proof of $\hat{\mathbf{s}}_j$ for the MSNMIX model

The single observation of the complete data log-likelihood, ignoring additive constant term, is

$$\begin{aligned}&\ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j,\mathbf{Z}_j,\boldsymbol{\gamma}_j)\\ =& \sum_{i=1}^{g} Z_{ij}\left[\log(w_i)-\frac{1}{2}\log|\boldsymbol{\Sigma}_i|-\frac{1}{2}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\right].\end{aligned}$$

The first derivatives of $\ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j,\mathbf{Z}_j,\boldsymbol{\gamma}_j)$ with respect to w_i is

$$\frac{\partial\ell_{cj}}{\partial w_i}=\frac{Z_{ij}}{w_i}-\frac{Z_{gj}}{w_g}.$$

Thus,

$$\hat{u}^{\mathrm{o}}_{j,w_i}=\mathrm{E}\Big(\frac{Z_{ij}}{w_i}-\frac{Z_{gj}}{w_g}\Big|\mathbf{Y}^{\mathrm{o}}_j,\hat{\boldsymbol{\Theta}}\Big)=\frac{\hat{Z}_{ij}}{\hat{w}_i}-\frac{\hat{Z}_{gj}}{\hat{w}_g}.$$

The first derivatives of $\ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j,\mathbf{Z}_j,\boldsymbol{\gamma}_j)$ with respect to $\boldsymbol{\xi}_i$ is

$$\frac{\partial\ell_{cj}}{\partial\boldsymbol{\xi}_i}=Z_{ij}\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j).$$

Thus,

$$\begin{aligned}\hat{\mathbf{u}}^{\mathrm{o}}_{j,\boldsymbol{\xi}_i} &= \mathrm{E}\Big(Z_{ij}\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\Big|\mathbf{Y}^{\mathrm{o}}_j,\hat{\boldsymbol{\Theta}}\Big)\\ &= \hat{Z}_{ij}\mathrm{E}\Big(\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\Big|\mathbf{Y}^{\mathrm{o}}_j,Z_{ij}=1,\hat{\boldsymbol{\Theta}}\Big)\\ &= \hat{Z}_{ij}\hat{\boldsymbol{\Sigma}}_i^{-1}(\hat{\mathbf{Y}}_{ij}-\hat{\boldsymbol{\xi}}_i-\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij}),\end{aligned}$$

where $\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}^{\mathrm{o}}_j,Z_{ij}=1,\hat{\boldsymbol{\Theta}})=\hat{\boldsymbol{\eta}}_{ij}$ and $\mathrm{E}(\mathbf{Y}_j|\mathbf{Y}^{\mathrm{o}}_j,Z_{ij}=1,\hat{\boldsymbol{\Theta}})=\hat{\mathbf{Y}}_{ij}$.

The first derivatives of $\ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j,\mathbf{Z}_j,\boldsymbol{\gamma}_j)$ with respect to $\boldsymbol{\Sigma}_i$ is

$$\begin{aligned}\frac{\partial\ell_{cj}}{\partial\boldsymbol{\Sigma}_i} &= -\frac{Z_{ij}}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}_i}\Big\{\log|\boldsymbol{\Sigma}_i|+\mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top\right)\Big\}\\ &= -\frac{Z_{ij}}{2}\Big\{2\boldsymbol{\Sigma}_i^{-1}-\mathrm{Diag}(\boldsymbol{\Sigma}_i^{-1})-2\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}_i^{-1}\\ &\quad+\mathrm{Diag}\left(\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}_i^{-1}\right)\Big\}\\ &= \frac{Z_{ij}}{2}\left(2\mathbf{B}_{ij}-\mathrm{Diag}(\mathbf{B}_{ij})\right),\end{aligned}$$

where $\mathbf{B}_{ij} = \boldsymbol{\Sigma}_i^{-1}\mathbf{R}_{ij}\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}$ and $\mathbf{R}_{ij} = \left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top$.

Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\mathrm{o}}_{j,\boldsymbol{\sigma}_i} &= \mathrm{vech}\left\{\mathrm{E}\Big(\frac{Z_{ij}}{2}\left(2\mathbf{B}_{ij} - \mathrm{Diag}(\mathbf{B}_{ij})\right)\Big|\mathbf{Y}^{\mathrm{o}}_j, \hat{\boldsymbol{\Theta}}\Big)\right\} \\
&= \mathrm{vech}\left\{\frac{\hat{Z}_{ij}}{2}\Big(2\hat{\mathbf{B}}_{ij} - \mathrm{Diag}(\hat{\mathbf{B}}_{ij})\Big)\right\},
\end{aligned}$$

where $\hat{\mathbf{B}}_{ij} = \hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\mathbf{R}}_{ij}\hat{\boldsymbol{\Sigma}}_i^{-1} - \hat{\boldsymbol{\Sigma}}_i^{-1}$ with

$$\begin{aligned}
\hat{\mathbf{R}}_{ij} &= \mathrm{E}(\mathbf{R}_{ij}|\mathbf{Y}^{\mathrm{o}}_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}) \\
&= (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij})\hat{\boldsymbol{\Sigma}}_i + (\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij})(\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij})^\top \\
&\quad + \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij})\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i\big)(\hat{\boldsymbol{\Psi}}_{ij} - \hat{\boldsymbol{\eta}}_{ij}\hat{\boldsymbol{\eta}}_{ij}^\top)\big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij})\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i\big)^\top.
\end{aligned}$$

Therefore,

$$\hat{\mathbf{B}}_{ij} = \hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\left(\hat{\boldsymbol{\Lambda}}_i(\hat{\boldsymbol{\Psi}}_{ij} - \hat{\boldsymbol{\eta}}_{ij}\hat{\boldsymbol{\eta}}_{ij}^\top)\hat{\boldsymbol{\Lambda}}_i + (\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij})(\hat{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij})^\top\right)\hat{\mathbf{S}}^{\mathrm{oo}}_{ij} - \hat{\mathbf{S}}^{\mathrm{oo}}_{ij}.$$

Furthermore, we can obtained

$$\mathrm{E}(\mathbf{Y}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}^{\mathrm{o}}_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}) = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij})(\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\eta}}_{ij}^\top + \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij}) + \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\mathbf{Y}_j\hat{\boldsymbol{\eta}}_{ij}^\top,$$

from the law of iterative expectations. The first derivatives of $\ell_{cj}(\boldsymbol{\Theta}|\mathbf{Y}_j, \mathbf{Z}_j, \boldsymbol{\gamma}_j)$ with respect to $\boldsymbol{\lambda}_i$ is

$$\frac{\partial \ell_{cj}}{\partial \boldsymbol{\lambda}_i} = \mathrm{Diag}\left(Z_{ij}\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\right).$$

Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\mathrm{o}}_{j,\boldsymbol{\lambda}_i} &= \mathrm{E}\Big(\mathrm{Diag}\left(Z_{ij}\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\right)\Big|\mathbf{Y}^{\mathrm{o}}_j, \hat{\boldsymbol{\Theta}}\Big) \\
&= \mathrm{Diag}\left(\hat{Z}_{ij}\mathrm{E}\Big(\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\boldsymbol{\tau}_j^\top\Big|\mathbf{Y}^{\mathrm{o}}_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}\Big)\right) \\
&= \mathrm{Diag}\left(\hat{Z}_{ij}\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}((\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i)\hat{\boldsymbol{\eta}}_{ij}^\top - \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij})\right),
\end{aligned}$$

where $\mathrm{E}(\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}^{\mathrm{o}}_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}) = \hat{\boldsymbol{\Psi}}_{ij}$.

F. Proof of Theorem 2.1(c), (d), and (e)

The following lemma is used in proving the results.

Lemma F.1 If $\tau \sim \Gamma(\alpha, \beta)$, then for any $\mathbf{a} \in \mathbb{R}^p$

$$E(\Phi_p(\mathbf{a}\sqrt{\tau})) = T_p\left(\mathbf{a}\sqrt{\frac{\alpha}{\beta}}; 2\alpha\right),$$

where $T_p(\ \cdot\ ;\ 2\alpha)$ denotes a p-dimensional cdf of the t distribution with degrees of freedom 2α.

Proof of Theorem 2.1(c):

We partitioned the random vector $\mathbf{Y}_j$ and its parameters $\boldsymbol{\xi}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ as follows:

$$\mathbf{Y}_j = \begin{bmatrix} \mathbf{Y}_j^{\mathrm{o}} \\ \mathbf{Y}_j^{\mathrm{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\mathbf{Y}_j \\ \mathbf{M}_j\mathbf{Y}_j \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_j^{\mathrm{o}} \\ \boldsymbol{\xi}_j^{\mathrm{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\xi} \\ \mathbf{M}_j\boldsymbol{\xi} \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_j^{\mathrm{oo}} & \boldsymbol{\Sigma}_j^{\mathrm{om}} \\ \boldsymbol{\Sigma}_j^{\mathrm{mo}} & \boldsymbol{\Sigma}_j^{\mathrm{mm}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\Sigma}\mathbf{O}_j^\top & \mathbf{O}_j\boldsymbol{\Sigma}\mathbf{M}_j^\top \\ \mathbf{M}_j\boldsymbol{\Sigma}\mathbf{O}_j^\top & \mathbf{M}_j\boldsymbol{\Sigma}\mathbf{M}_j^\top \end{bmatrix},$$

and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_j^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}} \times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}} \times p_j^{\mathrm{o}}} & \boldsymbol{\Lambda}_j^{\mathrm{mm}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\Lambda}\mathbf{O}_j^\top & \mathbf{0}_{p_j^{\mathrm{o}} \times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}} \times p_j^{\mathrm{o}}} & \mathbf{M}_j\boldsymbol{\Lambda}\mathbf{M}_j^\top \end{bmatrix}.$$

By Lemma A.1

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2 = \begin{bmatrix} \mathbf{O}_j(\boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2)\mathbf{O}_j^\top & \mathbf{O}_j\boldsymbol{\Sigma}\mathbf{M}_j^\top \\ \mathbf{M}_j\boldsymbol{\Sigma}\mathbf{O}_j^\top & \mathbf{M}_j(\boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2)\mathbf{M}_j^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_j^{\mathrm{oo}} & \boldsymbol{\Omega}_j^{\mathrm{om}} \\ \boldsymbol{\Omega}_j^{\mathrm{mo}} & \boldsymbol{\Omega}_j^{\mathrm{mm}} \end{bmatrix}.$$

Note that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are symmetric matrix. Thus, $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_j^{\mathrm{oo}}$ and $\boldsymbol{\Omega}_j^{\mathrm{mm}}$ are symmetric matrices and $\boldsymbol{\Omega}_j^{\mathrm{om}^\top} = \boldsymbol{\Omega}_j^{\mathrm{mo}}$.

Let $\mathbf{\Lambda}\mathbf{\Omega}^{-1} = \mathbf{B} = \begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_2\end{bmatrix}$. Thus, we have

$$\begin{aligned}
\mathbf{\Lambda}\mathbf{\Omega}^{-1} &= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}\begin{bmatrix}\mathbf{\Omega}_j^{\mathrm{oo}} & \mathbf{\Omega}_j^{\mathrm{om}} \\ \mathbf{\Omega}_j^{\mathrm{mo}} & \mathbf{\Omega}_j^{\mathrm{mm}}\end{bmatrix}^{-1} \\
&= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}\begin{bmatrix}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} + \mathbf{I}_{p_j^{\mathrm{o}}}) & -\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}} \\ -\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} & \mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}}(\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} + \mathbf{\Omega}_j^{\mathrm{oo}^{-1}}) & -\mathbf{\Lambda}_j^{\mathrm{oo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}} \\ -\mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} & \mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_2\end{bmatrix},
\end{aligned}$$

where $\mathbf{\Omega}_j^{\mathrm{mm\cdot o}} = \mathbf{\Omega}_j^{\mathrm{mm}} - \mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}$. Thus, we have

$$\begin{aligned}
\mathbf{B}_1 &= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}}(\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} + \mathbf{\Omega}_j^{\mathrm{oo}^{-1}}) \\ -\mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}}(\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}} + \mathbf{I}_{p_j^{\mathrm{o}}}) \\ -\mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\end{bmatrix}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}
\end{aligned}$$

and

$$\mathbf{B}_2 = \begin{bmatrix}-\mathbf{\Lambda}_j^{\mathrm{oo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}} \\ \mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\end{bmatrix} = \begin{bmatrix}-\mathbf{\Lambda}_j^{\mathrm{oo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}} \\ \mathbf{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}.$$

From the above calculation, we have the following results

$$\begin{aligned}
&\mathbf{B}_1 + \mathbf{B}_2\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} \\
&= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}}(\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}} + \mathbf{I}_{p_j^{\mathrm{o}}}) \\ -\mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\end{bmatrix}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} + \begin{bmatrix}-\mathbf{\Lambda}_j^{\mathrm{oo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}} \\ \mathbf{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} \\
&= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}} + \mathbf{\Lambda}_j^{\mathrm{oo}} - \mathbf{\Lambda}_j^{\mathrm{oo}}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}}\mathbf{\Omega}_j^{\mathrm{om}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}} \\ -\mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}} + \mathbf{\Lambda}_j^{\mathrm{mm}}\mathbf{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\mathbf{\Omega}_j^{\mathrm{mo}}\end{bmatrix}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}} \\
&= \begin{bmatrix}\mathbf{\Lambda}_j^{\mathrm{oo}} \\ \mathbf{0}_{p_j^{\mathrm{m}} \times p_j^{\mathrm{o}}}\end{bmatrix}\mathbf{\Omega}_j^{\mathrm{oo}^{-1}},
\end{aligned}$$

$$
\begin{aligned}
&\boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda} \\
=\;& \begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_j^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix} \\
=\;& \begin{bmatrix}\mathbf{B}_1\boldsymbol{\Lambda}_j^{\mathrm{oo}} & \mathbf{B}_2\boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix} \\
=\;& \begin{bmatrix}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\boldsymbol{\Omega}_j^{\mathrm{om}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}+\mathbf{I}_{p_j^{\mathrm{o}}})\boldsymbol{\Lambda}_j^{\mathrm{oo}} & -\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{om}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{mm}} \\ -\boldsymbol{\Lambda}_j^{\mathrm{mm}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}} & \boldsymbol{\Lambda}_j^{\mathrm{mm}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix},
\end{aligned} \tag{F.1}
$$

and

$$
\begin{aligned}
&\mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}\mathbf{B}_2^{\top} \\
=\;& \begin{bmatrix}-\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{om}} \\ \boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}(\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}})^{\top}\begin{bmatrix}-(\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{om}})^{\top} & \boldsymbol{\Lambda}_j^{\mathrm{mm}^{\top}}\end{bmatrix} \\
=\;& \begin{bmatrix}-\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{om}} \\ \boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\begin{bmatrix}-(\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}}) & \boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix} \\
=\;& \begin{bmatrix}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{om}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}} & -\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{om}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{mm}} \\ -\boldsymbol{\Lambda}_j^{\mathrm{mm}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}} & \boldsymbol{\Lambda}_j^{\mathrm{mm}}\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{mm}}\end{bmatrix}.
\end{aligned} \tag{F.2}
$$

Since

$$
\boldsymbol{\Delta}+\mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}\mathbf{B}_2^{\top} = \mathbf{I}_p - \boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda} + \mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}\mathbf{B}_2^{\top}\,, \tag{F.3}
$$

we substituted $(F.1)$, $(F.2)$ to $(F.3)$ to obtain

$$
\begin{aligned}
\boldsymbol{\Delta}+\mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}\mathbf{B}_2^{\top} &= \mathbf{I}_p - \begin{bmatrix}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{m}}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{I}_{p_j^{\mathrm{m}}}\end{bmatrix}.
\end{aligned}
$$

Since

$$
\mathbf{Y}_j \mid (\boldsymbol{\gamma}_j, \tau_j) \sim N_p(\boldsymbol{\xi}+\boldsymbol{\Lambda}\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}/\tau_j)
$$

and

$$\boldsymbol{\gamma}_j \mid \tau_j \sim HN_p(\mathbf{0}, \mathbf{I}_p/\tau_j),$$

we have

$$f(\mathbf{Y}_j \mid \boldsymbol{\gamma}_j, \tau_j) = \phi_p(\mathbf{Y}_j|\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}/\tau_j)$$

and

$$f(\boldsymbol{\gamma}_j \mid \tau_j) = 2^p \phi_p(\boldsymbol{\gamma}_j|\mathbf{0}, \mathbf{I}_p/\tau_j) I_{R_+^p}(\boldsymbol{\gamma}_j).$$

By Lemma B.1, we have

$$\begin{aligned} & (\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j) + \boldsymbol{\gamma}_j^\top \mathbf{I}_p^{-1}\boldsymbol{\gamma}_j \\ = \; & (\mathbf{Y}_j - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}) + (\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}))^\top \boldsymbol{\Delta}^{-1}(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)), \end{aligned}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2$ and $\boldsymbol{\Delta} = (\mathbf{I}_p + \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}$. Thus,

$$\begin{aligned} & \phi_p(\mathbf{Y}_j|\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\mathbf{0}, \mathbf{I}_p/\tau_j) I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j) \\ = \; & \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}, \boldsymbol{\Omega}/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}), \boldsymbol{\Delta}/\tau_j) I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j). \end{aligned}$$

Thus,

$$\begin{aligned} f(\mathbf{Y}_j|\tau_j, \boldsymbol{\Theta}) &= \int_{\mathbb{R}_+^p} f(\mathbf{Y}_j \mid \boldsymbol{\gamma}_j, \tau_j, \boldsymbol{\theta}) f(\boldsymbol{\gamma}_j \mid \tau_j, \boldsymbol{\theta}) d\boldsymbol{\gamma}_j \\ &= \int_{\mathbb{R}_+^p} 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\mathbf{0}, \mathbf{I}_p/\tau_j) d\boldsymbol{\gamma}_j \\ &= \int_{\mathbb{R}_+^p} 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}, \boldsymbol{\Omega}/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}), \boldsymbol{\Delta}/\tau_j) d\boldsymbol{\gamma}_j \\ &= 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}, \boldsymbol{\Omega}/\tau_j) \int_{\mathbb{R}_+^p} \phi_p(\boldsymbol{\gamma}_j|\boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}), \boldsymbol{\Delta}/\tau_j) d\boldsymbol{\gamma}_j \\ &= 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}, \boldsymbol{\Omega}/\tau_j) \Phi_p(\sqrt{\tau_j}\boldsymbol{\Lambda}\boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi})|\boldsymbol{\Delta}). \end{aligned}$$

Thus,

$$\begin{aligned}
& f(\mathbf{Y}_j^{\mathrm{o}}, \mathbf{Y}_j^{\mathrm{m}} | \tau_j, \boldsymbol{\theta}) \\
= \; & 2^p \phi_p(\mathbf{Y}_j | \boldsymbol{\xi}, \boldsymbol{\Omega}/\tau_j) \Phi_p(\sqrt{\tau_j} \boldsymbol{\Lambda} \boldsymbol{\Omega}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}) | \boldsymbol{\Delta}) \\
= \; & 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j) \phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}} | \boldsymbol{\xi}_j^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j) \Phi_p\Big(\sqrt{\tau_j} \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}} \\ \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_j^{\mathrm{m}} \end{bmatrix} \Big| \boldsymbol{\Delta}\Big) \\
= \; & 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j) \phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}} | \boldsymbol{\xi}_j^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j) \\
& \times \Phi_p\Big(\sqrt{\tau_j} \left(\mathbf{B}_1(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_j^{\mathrm{m}})\right) \mid \boldsymbol{\Delta}\Big),
\end{aligned}$$

where condition mean $\boldsymbol{\xi}_j^{\mathrm{m\cdot o}} = \boldsymbol{\xi}_j^{\mathrm{m}} + \boldsymbol{\Omega}_j^{\mathrm{mo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} (\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})$ and conditional covariance $\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j = (\boldsymbol{\Omega}_j^{\mathrm{mm}} - \boldsymbol{\Omega}_j^{\mathrm{mo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} \boldsymbol{\Omega}_j^{\mathrm{om}})/\tau_j$. The marginal density of $\mathbf{Y}_j^{\mathrm{o}}$ given τ_j and $\boldsymbol{\theta}$ is as follows

$$\begin{aligned}
f(\mathbf{Y}_j^{\mathrm{o}} | \tau_j, \boldsymbol{\theta}) &= \int f(\mathbf{Y}_j^{\mathrm{o}}, \mathbf{Y}_j^{\mathrm{m}} | \tau_j, \boldsymbol{\theta}) \, d\mathbf{Y}_j^{\mathrm{m}} \\
&= 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}} | \boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j) \int \phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}} | \boldsymbol{\xi}_j^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j) \\
& \quad \times \Phi_p\Big(\sqrt{\tau_j} \left(\mathbf{B}_1(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_j^{\mathrm{m}})\right) \mid \boldsymbol{\Delta}\Big) d\mathbf{Y}_j^{\mathrm{m}}.
\end{aligned}$$

Let $\mathbf{z} = \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_j^{\mathrm{m\cdot o}}$, then $\mathbf{Y}_j^{\mathrm{m}} = \mathbf{z} + \boldsymbol{\xi}_j^{\mathrm{m\cdot o}} = \mathbf{z} + \boldsymbol{\xi}_j^{\mathrm{m}} + \boldsymbol{\Omega}_j^{\mathrm{mo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} (\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})$. Thus, we have

$$\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}} | \boldsymbol{\xi}_j^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j) = \phi_{p_j^{\mathrm{m}}}(\mathbf{z} | \mathbf{0}, \boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j)$$

and

$$\begin{aligned}
& \Phi_p\Big(\sqrt{\tau_j} \left(\mathbf{B}_1(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_j^{\mathrm{m}})\right)\Big) \\
= \; & \Phi_p\Big(\sqrt{\tau_j} \Big(\mathbf{B}_1(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2(\mathbf{z} + \boldsymbol{\Omega}_j^{\mathrm{mo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}} (\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})\Big) \mid \boldsymbol{\Delta}\Big) \\
= \; & \Phi_p\Big(\sqrt{\tau_j} \Big((\mathbf{B}_1 + \mathbf{B}_2 \boldsymbol{\Omega}_j^{\mathrm{mo}} \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2 \mathbf{z}\Big) \mid \boldsymbol{\Delta}\Big).
\end{aligned}$$

By Lemma 2.1 of Arellano-Valle and Genton (2005), we have

$$\begin{aligned}
& f(\mathbf{Y}_j^{\mathrm{o}}|\tau_j, \boldsymbol{\theta}) \\
= \;& 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j) \\
& \times \int \phi_{p_j^{\mathrm{m}}}(\mathbf{z}|\mathbf{0}, \boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}/\tau_j)\Phi_p\big(\sqrt{\tau_j}\,\Big((\mathbf{B}_1 + \mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2\mathbf{z}\Big)\,|\boldsymbol{\Delta}\big)\, d\mathbf{z} \\
= \;& 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j)\mathrm{E}_{\mathbf{z}}\Big\{\Phi_p\big(\sqrt{\tau_j}\,\Big((\mathbf{B}_1 + \mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \mathbf{B}_2\mathbf{Z}\Big)\,|\boldsymbol{\Delta}\big)\Big\} \\
= \;& 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j)\Phi_p\big(\sqrt{\tau_j}(\mathbf{B}_1 + \mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})\big|\boldsymbol{\Delta} + \mathbf{B}_2\boldsymbol{\Omega}_j^{\mathrm{mm\cdot o}}\mathbf{B}_2^{\top}\big) \\
= \;& 2^{p_j^{\mathrm{o}}} \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_j^{\mathrm{o}}, \boldsymbol{\Omega}_j^{\mathrm{oo}}/\tau_j)\Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\big).
\end{aligned}$$

Since

$$\tau_j \sim \Gamma(\nu/2, \nu/2),$$

we have

$$\begin{aligned}
& f(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\theta}) \\
= \;& \int f(\mathbf{Y}_j^{\mathrm{o}}|\tau_j, \boldsymbol{\theta}) f(\tau_j) d\tau_j \\
= \;& \Big(\frac{\pi}{2}\Big)^{-\frac{p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} |\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}|^{-\frac{1}{2}} \\
& \times \int_0^\infty \tau_j^{\frac{\nu + p_j^{\mathrm{o}}}{2} - 1} \exp\Big(-\frac{\tau_j}{2}\big[(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})^\top \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \nu\big]\Big) \\
& \times \Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\big) d\tau_j \\
= \;& \Big(\frac{\pi}{2}\Big)^{-\frac{p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} |\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu + p_j^{\mathrm{o}}}{2})}{\Big\{\frac{1}{2}\big[(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})^\top \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \nu\big]\Big\}^{\frac{\nu + p_j^{\mathrm{o}}}{2}}} \\
& \times \int_0^\infty \Gamma\Big(\frac{\nu_i + p_j^{\mathrm{o}}}{2}, \frac{(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})^\top \boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}}) + \nu}{2}\Big) \\
& \times \Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_j^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_j^{\mathrm{oo}}\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_j^{\mathrm{oo}}\big) d\tau_j
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{\pi}{2}\right)^{-\frac{p_j^{\rm o}}{2}} \left(\frac{\nu_i}{2}\right)^{-\frac{p_j^{\rm o}}{2}} \left(\frac{\nu}{2}\right)^{\frac{\nu+p_j^{\rm o}}{2}} |\mathbf{\Omega}_j^{\rm oo^{-1}}|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+p_j^{\rm o}}{2})}{\Gamma(\frac{\nu}{2})} \\
&\quad \times \left[\frac{(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})^\top \mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})+\nu}{2}\right]^{-\frac{\nu+p_j^{\rm o}}{2}} \\
&\quad \times \int_0^\infty \Gamma\Big(\frac{\nu+p_j^{\rm o}}{2}, \frac{(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})^\top \mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})+\nu}{2}\Big) \\
&\quad \times \Phi_{p_j^{\rm o}}\big(\sqrt{\tau_j}\mathbf{\Lambda}_j^{\rm oo}\mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})\big|\mathbf{I}_{p_j^{\rm o}}-\mathbf{\Lambda}_j^{\rm oo}\mathbf{\Omega}_j^{\rm oo^{-1}}\mathbf{\Lambda}_j^{\rm oo}\big)d\tau_j \\
&= 2^{p_j^{\rm o}}(\pi\nu)^{-\frac{p_j^{\rm o}}{2}}|\mathbf{\Omega}_j^{\rm oo^{-1}}|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+p_j^{\rm o}}{2})}{\Gamma(\frac{\nu}{2})} \left[\frac{(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})^\top \mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})}{\nu}+1\right]^{-\frac{\nu+p_j^{\rm o}}{2}} \\
&\quad \times \int_0^\infty \Gamma\Big(\frac{\nu+p_j^{\rm o}}{2}, \frac{(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})^\top \mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})+\nu}{2}\Big) \\
&\quad \times \Phi_{p_j^{\rm o}}\big(\sqrt{\tau_j}\mathbf{\Lambda}_j^{\rm oo}\mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})\big|\mathbf{I}_{p_j^{\rm o}}-\mathbf{\Lambda}_j^{\rm oo}\mathbf{\Omega}_j^{\rm oo^{-1}}\mathbf{\Lambda}_j^{\rm oo}\big)d\tau_j \\
&= 2^{p_j^{\rm o}} \times t_{p_j^{\rm o}}(\boldsymbol{\xi}_j^{\rm o}, \mathbf{\Omega}_j^{\rm oo}, \nu) T_{p_j^{\rm o}}\left(\mathbf{\Lambda}_j^{\rm oo}\mathbf{\Omega}_j^{\rm oo^{-1}}(\mathbf{Y}_j^{\rm o}-\boldsymbol{\xi}_j^{\rm o})\Big(\frac{\nu+p_j^{\rm o}}{U_j^{\rm o}+\nu}\Big)^{\frac{1}{2}} \mid \mathbf{\Delta}_j^{\rm oo}; \nu+p_j^{\rm o}\right),
\end{aligned}
$$

where $U_j^{\rm o} = (\mathbf{Y}_j - \boldsymbol{\xi})^\top \mathbf{C}_j^{\rm oo}(\mathbf{Y}_j - \boldsymbol{\xi})$, $\mathbf{\Delta}_j^{\rm oo} = \mathbf{I}_{p_j^{\rm o}} - \mathbf{\Lambda}_j^{\rm oo}\mathbf{\Omega}_j^{\rm oo^{-1}}\mathbf{\Lambda}_j^{\rm oo}$ and $\mathbf{C}_j^{\rm oo} = \mathbf{O}_j^\top \mathbf{\Omega}_j^{\rm oo^{-1}}\mathbf{O}_j$ with the last equality is by Lemma F.1.

Proof of Theorem 2.1(d):

The joint density of $\mathbf{Y}_j^{\rm o}$ and $\boldsymbol{\gamma}_j$ is as follows

$$
\begin{aligned}
& f(\mathbf{Y}_j^{\rm o}, \boldsymbol{\gamma}_j) \\
&= \int_0^\infty f(\mathbf{Y}_j^{\rm o}, \boldsymbol{\gamma}_j, \tau_j) d\tau_j \\
&= 2^{\frac{p_j^{\rm m}}{2}} \pi^{-\frac{p+p_j^{\rm o}}{2}} \frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} |\mathbf{\Sigma}_j^{\rm oo}|^{-\frac{1}{2}} \\
&\quad \times \int_0^\infty \tau_j^{(\frac{\nu+p+p_j^{\rm o}}{2})-1} \exp\Bigg\{ -\tau_j\Big([\mathbf{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})]^\top \mathbf{S}_j^{\rm oo}[\mathbf{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})] \\
&\quad + \boldsymbol{\gamma}_j^\top \boldsymbol{\gamma}_j + \nu\Big)/2\Bigg\} d\tau_j
\end{aligned}
$$

$$= \Gamma\left(\frac{\nu+p+p_j^{\text{o}}}{2}\right)\left(\frac{[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi})]^\top\mathbf{S}_j^{\text{oo}}[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi})]+\boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j+\nu}{2}\right)^{-\frac{\nu+p+p_j^{\text{o}}}{2}}$$
$$\times 2^{\frac{p_j^{\text{m}}}{2}}\pi^{-\frac{p+p_j^{\text{o}}}{2}}\frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}|\boldsymbol{\Sigma}_j^{\text{oo}}|^{-\frac{1}{2}}.$$
$$= \Gamma\left(\frac{\nu+p+p_j^{\text{o}}}{2}\right)\left(\frac{\big(\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi})\big)^\top\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}\boldsymbol{\Lambda}\big(\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi})\big)+\boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j+\nu}{2}\right)^{-\frac{\nu+p+p_j^{\text{o}}}{2}}$$
$$\times 2^{\frac{p_j^{\text{m}}}{2}}\pi^{-\frac{p+p_j^{\text{o}}}{2}}\frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}|\boldsymbol{\Sigma}_j^{\text{oo}}|^{-\frac{1}{2}}.$$

By Lemma B.1, we let $\mathbf{x}=\boldsymbol{\gamma}_j$, $\mathbf{a}=\boldsymbol{\Lambda}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi})$, $\mathbf{b}=\mathbf{0}$, $\mathbf{Q}_1=\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}\boldsymbol{\Lambda}$, $\mathbf{Q}_2=\mathbf{I}_p$, $\mathbf{Q}=\mathbf{Q}_1+\mathbf{Q}_2=\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}\boldsymbol{\Lambda}+\mathbf{I}_p$, and $\hat{\mathbf{x}}=\mathbf{Q}^{-1}(\mathbf{Q}_1\mathbf{a}+\mathbf{Q}_2\mathbf{b})=\boldsymbol{\Delta}_j^*\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi})=\boldsymbol{\Delta}_j^*\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi})$.

Furthermore, we have

(a)$\boldsymbol{\Delta}_j^*\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi})=\boldsymbol{\Lambda}\mathbf{C}_j^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi})$, where $\mathbf{C}_j^{\text{oo}}=\mathbf{O}_j^\top\boldsymbol{\Omega}_j^{\text{oo}^{-1}}\mathbf{O}_j$ and (b)$\mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2=\boldsymbol{\Lambda}\mathbf{C}_j^{\text{oo}}\boldsymbol{\Lambda}$.

Proof:

(a)

$$\begin{aligned}
&\boldsymbol{\Delta}_j^*\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi})\\
=\;&\Big(\mathbf{I}_p-(\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda})\Big)\boldsymbol{\Lambda}\mathbf{S}_j^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi})\\
=\;&\Big(\mathbf{I}_p-(\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda})\Big)(\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}\mathbf{O}_j(\mathbf{Y}_j-\boldsymbol{\xi})\\
=\;&\Big((\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}-(\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda})(\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}\Big)(\mathbf{Y}_j^{\text{o}}-\boldsymbol{\xi}_j^{\text{o}})\\
=\;&(\mathbf{O}_j\boldsymbol{\Lambda})^\top\Big(\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}-\boldsymbol{\Omega}_j^{\text{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}^2\mathbf{O}_j^\top)\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}\Big)(\mathbf{Y}_j^{\text{o}}-\boldsymbol{\xi}_j^{\text{o}})\\
=\;&(\mathbf{O}_j\boldsymbol{\Lambda})^\top\Big(\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}-\boldsymbol{\Omega}_j^{\text{oo}^{-1}}(\boldsymbol{\Omega}_j^{\text{oo}^{-1}}-\boldsymbol{\Sigma}_j^{\text{oo}})\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}\Big)(\mathbf{Y}_j^{\text{o}}-\boldsymbol{\xi}_j^{\text{o}})\\
=\;&(\mathbf{O}_j\boldsymbol{\Lambda})^\top\Big(\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}-\boldsymbol{\Sigma}_j^{\text{oo}^{-1}}+\boldsymbol{\Omega}_j^{\text{oo}^{-1}}\Big)(\mathbf{Y}_j^{\text{o}}-\boldsymbol{\xi}_j^{\text{o}})\\
=\;&(\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\text{oo}^{-1}}(\mathbf{Y}_j^{\text{o}}-\boldsymbol{\xi}_j^{\text{o}})\\
=\;&\boldsymbol{\Lambda}\mathbf{C}_j^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi}).
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2 &= \boldsymbol{\Lambda}\mathbf{S}_j^{\mathrm{oo}}\boldsymbol{\Lambda}\Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda})\Big)\mathbf{I}_p \\
&= (\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Sigma}_j^{\mathrm{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda})\Big(\mathbf{I}_p - (\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda})\Big) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda})^\top\Big(\boldsymbol{\Sigma}_j^{\mathrm{oo}^{-1}} - \boldsymbol{\Sigma}_j^{\mathrm{oo}-1}(\mathbf{O}_j\boldsymbol{\Lambda}^2\mathbf{O}_j^\top)\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\Big)(\mathbf{O}_j\boldsymbol{\Lambda}) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda})^\top\Big(\boldsymbol{\Sigma}_j^{\mathrm{oo}^{-1}} - \boldsymbol{\Sigma}_j^{\mathrm{oo}-1}(\boldsymbol{\Omega}_j^{\mathrm{oo}} - \boldsymbol{\Sigma}_j^{\mathrm{oo}})\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}\Big)(\mathbf{O}_j\boldsymbol{\Lambda}) \\
&= (\mathbf{O}_j\boldsymbol{\Lambda})^\top\boldsymbol{\Omega}_j^{\mathrm{oo}^{-1}}(\mathbf{O}_j\boldsymbol{\Lambda}) \\
&= \boldsymbol{\Lambda}\mathbf{C}_j^{\mathrm{oo}}\boldsymbol{\Lambda}.
\end{aligned}$$

Thus,

$$\begin{aligned}
& f(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}) \propto f(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j) \\
\propto\;& \big([\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi})]^\top\boldsymbol{\Lambda}^\top\mathbf{S}_j^{\mathrm{oo}}\boldsymbol{\Lambda}[\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi})] + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j + \nu\big)^{-\frac{\nu+p+p_j^{\mathrm{o}}}{2}} \\
=\;& \Big([\boldsymbol{\gamma}_j - \mathbf{q}_j^*]^\top\boldsymbol{\Delta}_j^{*^{-1}}[\boldsymbol{\gamma}_j - \mathbf{q}_j^*] + (U_j^{\mathrm{o}} + \nu)\Big)^{-\frac{\nu+p+p_j^{\mathrm{o}}}{2}} \\
\propto\;& \Big(\frac{[\boldsymbol{\gamma}_j - \mathbf{q}_j^*]^\top\boldsymbol{\Delta}_j^{*^{-1}}[\boldsymbol{\gamma}_j - \mathbf{q}_j^*]}{U_j^{\mathrm{o}} + \nu} + 1\Big)^{-\frac{\nu+p+p_j^{\mathrm{o}}}{2}} \\
=\;& \Big(\frac{[\boldsymbol{\gamma}_j - \mathbf{q}_j^*]^\top\boldsymbol{\Delta}_j^{*^{-1}}[\boldsymbol{\gamma}_j - \mathbf{q}_j^*]}{p_j^{\mathrm{o}} + \nu}\frac{p_j^{\mathrm{o}} + \nu}{U_j^{\mathrm{o}} + \nu} + 1\Big)^{-\frac{\nu+p+p_j^{\mathrm{o}}}{2}},
\end{aligned}$$

where $\mathbf{q}_j^* = \boldsymbol{\Lambda}\mathbf{C}_j^{\mathrm{oo}}(\mathbf{Y}_j - \boldsymbol{\xi})$, and $\boldsymbol{\Delta}_j^* = \mathbf{I}_p - \boldsymbol{\Lambda}\mathbf{C}_j^{\mathrm{oo}}\boldsymbol{\Lambda}$.

It implies $\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}} \sim Tt_p(\mathbf{q}_j^*, \frac{U_j^{\mathrm{o}}+\nu}{p_j^{\mathrm{o}}+\nu}\boldsymbol{\Delta}_j^*, \nu + p_j^{\mathrm{o}}; \mathbb{R}_p^+)$.

Proof of Theorem 2.1(e):

The conditional density of τ_j given $\mathbf{Y}_j^{\mathrm{o}}$ and $\boldsymbol{\gamma}_j$ is as follows

$$\begin{aligned}
& f(\tau_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j) = \frac{f(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j)}{f(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j)} \\
=\;& \frac{\tau_j^{(\frac{\nu+p+p_j^{\mathrm{o}}}{2})-1}}{\Gamma\Big(\frac{\nu+p+p_j^{\mathrm{o}}}{2}\Big)}\Bigg(\frac{[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})]^\top\mathbf{S}_j^{\mathrm{oo}}[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})] + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j + \nu}{2}\Bigg)^{\frac{\nu+p+p_j^{\mathrm{o}}}{2}} \\
& \times\exp\Bigg\{-\tau_j\Big([\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})]^\top\mathbf{S}_j^{\mathrm{oo}}[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})] + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j + \nu\Big)/2\Bigg\}.
\end{aligned}$$

It implies

$$\tau_j|(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j) \sim \Gamma\left(\frac{\nu + p + p_j^{\mathrm{o}}}{2}, \frac{[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})]^\top \mathbf{S}_j^{\mathrm{oo}}[\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi})] + \boldsymbol{\gamma}_j^\top \boldsymbol{\gamma}_j + \nu}{2}\right).$$

By Lemma B.1, we have

$$\tau_j|(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j) \sim \Gamma\left(\frac{\nu + p + p_j^{\mathrm{o}}}{2}, \frac{(\boldsymbol{\gamma}_j - \mathbf{q}_j^*)^\top {\boldsymbol{\Delta}_j^*}^{-1}(\boldsymbol{\gamma}_j - \mathbf{q}_j^*) + U_j^{\mathrm{o}} + \nu}{2}\right).$$

G. Proof of the $\boldsymbol{\Omega}_j^{(k)}$

Let $\boldsymbol{\Omega}_j^{(k)} = \mathrm{E}(\tau_j(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top | \mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})$. Since $\mathbf{Y}_j = \mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}}$ and $\mathbf{O}_j^\top \mathbf{O}_j(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) = 0$, we have

$$
\begin{aligned}
& \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathrm{E}(\mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j + \mathbf{M}_j^\top \mathrm{E}(\mathbf{Y}_j^{\mathrm{m}} | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j + \mathbf{M}_j^\top \mathbf{M}_j(\hat{\boldsymbol{\xi}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)}\boldsymbol{\gamma}_j + \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}^{(k)} - \hat{\boldsymbol{\Lambda}}^{(k)}\boldsymbol{\gamma}_j)) \\
= & \mathbf{O}_j^\top \mathbf{O}_j \mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j + \mathbf{M}_j^\top \mathbf{M}_j(\hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} \mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)}\boldsymbol{\gamma}_j)) \\
= & \mathbf{O}_j^\top \mathbf{O}_j \mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j + (\mathbf{I}_p - \mathbf{O}_j^\top \mathbf{O}_j)\hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} \mathbf{Y}_j \\
& + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)}\boldsymbol{\gamma}_j) \\
= & \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} \mathbf{Y}_j - (\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j) + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)}\boldsymbol{\gamma}_j), \\
& \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathrm{E}(\mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathrm{E}(\hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} \mathbf{Y}_j - (\boldsymbol{\xi} + \boldsymbol{\Lambda}\boldsymbol{\gamma}_j) + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)}\boldsymbol{\gamma}_j) | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} \mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\xi}}^{(k)} - \boldsymbol{\xi} \\
& + \left((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}^{(k)} - \boldsymbol{\Lambda}\right) \mathrm{E}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi} + \left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right) \mathrm{E}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathrm{Cov}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathrm{Cov}(\mathbf{O}_j^\top \mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
= & \mathrm{Cov}(\mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{E}\big(\mathrm{Cov}(\mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\big) \\
&\quad + \mathrm{Cov}\big(\mathrm{E}(\mathbf{M}_j^\top \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\big) \\
&= \mathrm{E}\big(\mathbf{M}_j^\top \mathrm{Cov}(\mathbf{Y}_j^{\mathrm{m}} | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \mathbf{M}_j \big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\big) \\
&\quad + \mathrm{Cov}\big(\mathbf{M}_j^\top \mathrm{E}(\mathbf{Y}_j^{\mathrm{m}} | \mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j \big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\big) \\
&= \mathrm{E}\big(\frac{1}{\tau_j} \mathbf{M}_j^\top \mathbf{M}_j (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}^{(k)} \mathbf{M}_j^\top \mathbf{M}_j \big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\big) \\
&\quad + \mathrm{Cov}\big(\mathbf{M}_j^\top \mathbf{M}_j (\hat{\boldsymbol{\xi}}^{(k)} + \hat{\boldsymbol{\Lambda}}^{(k)} \boldsymbol{\gamma}_j + \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} (\mathbf{Y}_j - \hat{\boldsymbol{\xi}}^{(k)} - \hat{\boldsymbol{\Lambda}}^{(k)} \boldsymbol{\gamma}_j)) - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j \big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\big) \\
&= \frac{1}{\tau_j} (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}^{(k)} + \mathrm{Cov}\Big(\big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}^{(k)} - \boldsymbol{\Lambda}\big) \boldsymbol{\gamma}_j \Big| \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}\Big) \\
&= \frac{1}{\tau_j} (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}^{(k)} \\
&\quad + \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}^{(k)} - \boldsymbol{\Lambda}\big) \mathrm{Cov}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}^{(k)} - \boldsymbol{\Lambda}\big)^\top \\
&= \frac{1}{\tau_j} (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}^{(k)} + \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big) \mathrm{Cov}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)^\top,
\end{aligned}
$$

where $\hat{\mathbf{A}}_j^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Lambda}}^{(k)}$ and $\hat{\mathbf{b}}_j^{(k)} = \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)} \mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\xi}}^{(k)}$.

Thus,

$$
\begin{aligned}
& \mathrm{E}((\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
&= \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})^\top \\
&\quad + \mathrm{Cov}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \\
&= \Big[\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi} + \Big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\Big) \mathrm{E}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})\Big] \\
&\quad \times \Big[\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi} + \Big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\Big) \mathrm{E}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})\Big]^\top \\
&\quad + \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big) \mathrm{Cov}(\boldsymbol{\gamma}_j | \mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)}) \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)^\top + \frac{1}{\tau_j} (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)} \hat{\mathbf{S}}_j^{\mathrm{oo}(k)}) \hat{\boldsymbol{\Sigma}}^{(k)}
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)^\top + \left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})^\top \left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right)^\top \\
&\quad + \left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)^\top + \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}^{(k)} \\
&\quad + \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})^\top\big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)^\top \\
&\quad + \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)\mathrm{Cov}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})\big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)^\top \\
&= \left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)^\top + \left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})^\top \left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right)^\top \\
&\quad + \left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)^\top + \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}^{(k)} \\
&\quad + \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)\mathrm{E}(\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})^\top\big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)^\top.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\boldsymbol{\Omega}_j^{(k)} &= \mathrm{E}(\tau_j(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)}) \\
&= \mathrm{E}(\tau_j\mathrm{E}((\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \hat{\boldsymbol{\theta}}^{(k)})|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)}) \\
&= \big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)\mathrm{E}(\tau_j\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})\big(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\big)^\top \\
&\quad + \left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)\mathrm{E}(\tau_j\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})^\top\left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right)^\top \\
&\quad + \left(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda}\right)\mathrm{E}(\tau_j\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)^\top \\
&\quad + \mathrm{E}(\tau_j|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\theta}}^{(k)})\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)\left(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi}\right)^\top + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}^{(k)}.
\end{aligned}
$$

H. Proof the of CM-steps

Let $\hat{\mathbf{A}}_j^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}^{(k)}$ and $\hat{\mathbf{b}}_j^{(k)} = \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)})\hat{\boldsymbol{\xi}}^{(k)}$.

CM-step 1:

Taking the partial derivative of Q with respect to $\boldsymbol{\xi}$ and setting it to zero yields

$$
\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\xi}} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_j^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}}\sum_{j=1}^{n}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}_j^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}}\sum_{j=1}^{n}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\hat{\tau}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi})(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi})^\top\big) \\
&\quad +\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)}-\boldsymbol{\Lambda})\hat{\boldsymbol{\eta}}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi})^\top\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top}(\hat{\mathbf{A}}_j^{(k)}-\boldsymbol{\Lambda})^\top\big)\Big] \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}}\sum_{j=1}^{n}\Big[\hat{\tau}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi})^\top\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi}) + 2(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi})^\top\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)}-\boldsymbol{\Lambda})\hat{\boldsymbol{\eta}}_j^{(k)}\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{n}\Big[-2\boldsymbol{\Sigma}^{-1}\hat{\tau}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi}) - 2\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)}-\boldsymbol{\Lambda})\hat{\boldsymbol{\eta}}_j^{(k)}\Big] \\
&= \boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\Big[\hat{\tau}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)}-\boldsymbol{\xi}) + (\hat{\mathbf{A}}_j^{(k)}-\boldsymbol{\Lambda})\hat{\boldsymbol{\eta}}_j^{(k)}\Big] \\
&= 0.
\end{aligned}
$$

With $\boldsymbol{\Lambda}$ fixed at $\hat{\boldsymbol{\Lambda}}^{(k)}$, solving this equation gives

$$
\hat{\boldsymbol{\xi}}^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{\tau}_j^{(k)}\hat{\mathbf{b}}_j^{(k)} - \sum_{j=1}^{n}\hat{\boldsymbol{\Sigma}}^{(k)}\hat{\mathbf{S}}_j^{\mathrm{oo}(k)}\hat{\boldsymbol{\Lambda}}^{(k)}\hat{\boldsymbol{\eta}}_j^{(k)}}{\sum_{j=1}^{n}\hat{\tau}_j^{(k)}}.
$$

CM-step 2:

Taking the partial derivative of Q with respect to $\boldsymbol{\Sigma}^{-1}$ and setting it to zero yields

$$
\frac{\partial Q}{\partial \boldsymbol{\Sigma}^{-1}} = \frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}}\left[\frac{n}{2}\log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_j^{(k)}\right)\right]
$$

$$
\begin{aligned}
&= \left[\frac{n}{2}\left(2\boldsymbol{\Sigma} - \mathrm{Diag}(\boldsymbol{\Sigma})\right) - \frac{1}{2}\left(2\sum_{j=1}^{n}\boldsymbol{\Omega}_j^{(k)} - \mathrm{Diag}\left(\sum_{j=1}^{n}\boldsymbol{\Omega}_j^{(k)}\right)\right)\right] \\
&= 0.
\end{aligned}
$$

Fix $\boldsymbol{\xi}$ at $\hat{\boldsymbol{\xi}}^{(k+1)}$ and $\boldsymbol{\Lambda}$ at $\hat{\boldsymbol{\Lambda}}^{(k)}$, solving this equation obtains

$$
\hat{\boldsymbol{\Sigma}}^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{\boldsymbol{\Omega}}_j^{(k+1/2)}}{n},
$$

where $\hat{\boldsymbol{\Omega}}_j^{(k+1/2)}$ is $\boldsymbol{\Omega}_j^{(k)}$ in Equation (2.7) with $\boldsymbol{\xi}$ replaced by $\hat{\boldsymbol{\xi}}^{(k+1)}$ and $\boldsymbol{\Lambda}$ replaced by $\hat{\boldsymbol{\Lambda}}^{(k)}$.

CM-step 3:

Here $\boldsymbol{\Lambda}$ assumed to be diagonal, say $\boldsymbol{\Lambda} = \mathrm{Diag}(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a p-dimensional vector. Taking the partial derivative of Q with respect to $\boldsymbol{\lambda}$ and setting it to zero yields

$$
\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\lambda}} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_j^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}}\sum_{j=1}^{n}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}_j^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}}\sum_{j=1}^{n}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})\hat{\boldsymbol{\Psi}}_j^{(k)}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})^\top\big) \\
&\quad +\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})\hat{\boldsymbol{\eta}}_j^{(k)}(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})^\top\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})^\top\big)\Big] \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}}\sum_{j=1}^{n}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})\hat{\boldsymbol{\Psi}}_j^{(k)}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})^\top\big) \\
&\quad +2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})^\top\big)\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{n}\Big[-2\mathrm{Diag}\left(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{A}}_j^{(k)} - \boldsymbol{\Lambda})\hat{\boldsymbol{\Psi}}_j^{(k)}\right) - 2\mathrm{Diag}\left(\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top}\right)\Big] \\
&= -\mathrm{Diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\sum_{j=1}^{n}\hat{\boldsymbol{\Psi}}_j^{(k)}\right) + \mathrm{Diag}\left(\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left((\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top} + \hat{\mathbf{A}}_j^{(k)}\hat{\boldsymbol{\Psi}}_j^{(k)}\right)\right)
\end{aligned}
$$

$$= -\Big(\boldsymbol{\Sigma}^{-1} \odot \sum_{j=1}^{n} \hat{\boldsymbol{\Psi}}_j^{(k)}\Big)\boldsymbol{\lambda} + \Big(\boldsymbol{\Sigma}^{-1} \odot \sum_{j=1}^{n} \Big((\hat{\mathbf{b}}_j^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_j^{(k)\top} + \hat{\mathbf{A}}_j^{(k)}\hat{\boldsymbol{\Psi}}_j^{(k)}\Big)^\top\Big)\mathbf{1}_p.$$

With $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^{(k+1)}$ and $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}^{(k+1)}$, setting $\partial Q/\partial\boldsymbol{\lambda} = \mathbf{0}$ yields the estimate of $\boldsymbol{\lambda}$

$$\hat{\boldsymbol{\lambda}}^{(k+1)} = \Big(\hat{\boldsymbol{\Sigma}}^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \hat{\boldsymbol{\Psi}}_j^{(k)}\Big)^{-1}\Big(\hat{\boldsymbol{\Sigma}}^{(k+1)^{-1}} \odot \sum_{j=1}^{n} \Big((\hat{\mathbf{b}}_j^{(k)} - \hat{\boldsymbol{\xi}}^{(k+1)})\hat{\boldsymbol{\eta}}_j^{(k)\top} + \hat{\mathbf{A}}_j^{(k)}\hat{\boldsymbol{\Psi}}_j^{(k)}\Big)^\top\Big)\mathbf{1}_p.$$

I. Proof of $\hat{\mathbf{s}}_j$ for the MST model

The single observation of the complete data log-likelihood, ignoring additive constant term, is

$$\begin{aligned}\ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j,\boldsymbol{\gamma}_j,\tau_j) &= \frac{\nu}{2}\log\frac{\nu}{2}-\log\Gamma\left(\frac{\nu}{2}\right)+\frac{1}{2}\log\left|\boldsymbol{\Sigma}^{-1}\right|+\frac{\nu}{2}\log\tau_j\\ &\quad -\frac{\tau_j}{2}\left\{\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)+\nu\right\}.\end{aligned}$$

Let $\hat{\mathbf{D}}_j=(\mathbf{I}_p-\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}_j^{\text{oo}})$.

Furthermore, we can obtained

$$\mathrm{E}(\tau_j\mathbf{Y}_j|\mathbf{Y}_j^{\text{o}},\hat{\boldsymbol{\theta}})=\hat{\tau}_j\hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}}+\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\eta}}_j+\hat{\tau}_j\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}_j^{\text{oo}}\mathbf{Y}_j,$$

from the law of iterative expectations.

The first derivatives of $\ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j,\boldsymbol{\gamma}_j,\tau_j)$ with respect to $\boldsymbol{\xi}$ is

$$\frac{\partial\ell_{cj}}{\partial\boldsymbol{\xi}}=\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j).$$

Thus,

$$\hat{\mathbf{u}}_{j,\boldsymbol{\xi}}^{\text{o}}=\mathrm{E}\Big(\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\Big|\mathbf{Y}_j^{\text{o}},\hat{\boldsymbol{\theta}}\Big)=\hat{\tau}_j\hat{\mathbf{S}}_j^{\text{oo}}(\mathbf{Y}_j-\hat{\boldsymbol{\xi}})-\hat{\mathbf{S}}_j^{\text{oo}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\eta}}_j.$$

The first derivatives of $\ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j,\boldsymbol{\gamma}_j,\tau_j)$ with respect to $\boldsymbol{\Sigma}$ is

$$\begin{aligned}\frac{\partial\ell_{cj}}{\partial\boldsymbol{\Sigma}} &= -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}}\Big\{\log|\boldsymbol{\Sigma}|+\mathrm{tr}\Big(\boldsymbol{\Sigma}^{-1}\tau_j\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)^\top\Big)\Big\}\\ &= -\frac{1}{2}\Big\{2\boldsymbol{\Sigma}^{-1}-\mathrm{Diag}(\boldsymbol{\Sigma}^{-1})-2\boldsymbol{\Sigma}^{-1}\tau_j\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}^{-1}\\ &\quad +\mathrm{Diag}\Big(\boldsymbol{\Sigma}^{-1}\tau_j\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}^{-1}\Big)\Big\}\\ &= \frac{1}{2}\left(2\mathbf{C}_j-\mathrm{Diag}(\mathbf{C}_j)\right),\end{aligned}$$

where $\mathbf{C}_j = \boldsymbol{\Sigma}^{-1}\mathbf{R}_j\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}$ and $\mathbf{R}_j = \tau_j\left(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j\right)^\top$. Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\text{o}}_{j,\boldsymbol{\sigma}} &= \text{vech}\left\{\text{E}\Big(\frac{1}{2}\left(2\mathbf{C}_j - \text{Diag}(\mathbf{C}_j)\right)\Big|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}\Big)\right\} \\
&= \text{vech}\left\{\frac{1}{2}\text{E}\Big(2\mathbf{C}_j - \text{Diag}(\mathbf{C}_j)\Big|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}\Big)\right\} \\
&= \text{vech}\left\{\frac{1}{2}\Big(2\hat{\mathbf{C}}_j - \text{Diag}(\hat{\mathbf{C}}_j)\Big)\right\},
\end{aligned}$$

where $\hat{\mathbf{C}}_j = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}_j\hat{\boldsymbol{\Sigma}}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}$ with

$$\begin{aligned}
\hat{\mathbf{R}}_j &= \text{E}\Big(\mathbf{R}_j\Big|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}\Big) \\
&= \hat{\mathbf{D}}_j\hat{\boldsymbol{\Sigma}} + (\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})\hat{\boldsymbol{\Psi}}_j(\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})^\top + (\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})\hat{\boldsymbol{\eta}}_j(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}^{\text{oo}}_j\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})^\top \\
&\quad +(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}^{\text{oo}}_j\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})\hat{\boldsymbol{\eta}}_j^\top(\hat{\mathbf{D}}_j\hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{\Lambda}})^\top \\
&\quad +\hat{\tau}_j(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}^{\text{oo}}_j\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})(\hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}^{\text{oo}}_j\mathbf{Y}_j + \hat{\mathbf{D}}_j\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}})^\top.
\end{aligned}$$

Furthermore, we can obtained

$$\text{E}(\tau_j\mathbf{Y}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}) = \hat{\mathbf{D}}_j(\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\eta}}_j^\top + \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}_j) + \hat{\boldsymbol{\Sigma}}\hat{\mathbf{S}}^{\text{oo}}_j\mathbf{Y}_j\hat{\boldsymbol{\eta}}_j^\top$$

from the law of iterative expectations. The first derivatives of $\ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \tau_j)$ with respect to $\boldsymbol{\lambda}$ is

$$\frac{\partial \ell_{cj}}{\partial \boldsymbol{\lambda}} = \text{Diag}\left(\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\right)$$

Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\text{o}}_{j,\boldsymbol{\lambda}} &= \text{E}\Big(\text{Diag}\left(\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\right)\Big|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}\Big) \\
&= \text{Diag}\left(\text{E}\Big(\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\Big|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}\Big)\right) \\
&= \text{Diag}\left(\hat{\boldsymbol{\Sigma}}^{-1}E(\tau_j\mathbf{Y}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}^{\text{o}}_j, \hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\eta}}_j^\top + \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}_j)\right) \\
&= \text{Diag}\left(\hat{\mathbf{S}}^{\text{oo}}_j\left((\mathbf{Y}_j - \hat{\boldsymbol{\xi}})\hat{\boldsymbol{\eta}}_j^\top - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}_j\right)\right).
\end{aligned}$$

The first derivatives of $\ell_{cj}(\boldsymbol{\theta}|\mathbf{Y}_j, \boldsymbol{\gamma}_j, \tau_j)$ with respect to ν is

$$\frac{\partial \ell_{cj}}{\partial \nu} = \frac{1}{2}\left\{\log\left(\frac{\nu}{2}\right) + 1 - \mathrm{DG}\left(\frac{\nu}{2}\right) + \log(\tau_j) - \tau_j\right\}.$$

Thus,

$$\begin{aligned}
\hat{u}^{\mathrm{o}}_{j,\nu} &= \mathrm{E}\Big(\frac{1}{2}\left\{\log\left(\frac{\nu}{2}\right) + 1 - \mathrm{DG}\left(\frac{\nu}{2}\right) + \log(\tau_j) - \tau_j\right\}\Big|\mathbf{Y}^{\mathrm{o}}_j, \hat{\boldsymbol{\theta}}\Big) \\
&= \frac{1}{2}\left[\log\left(\frac{\hat{\nu}}{2}\right) + 1 - \mathrm{DG}\left(\frac{\hat{\nu}}{2}\right) + \hat{\kappa}_j - \hat{\tau}_j\right].
\end{aligned}$$

J. Proof of Theorem 2.2

The proposed model is as follows:

$$\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n \sim ST_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu),$$

where $\boldsymbol{\Lambda} = \text{Diag}(\boldsymbol{\lambda})$.

The prior distributions are as follows:

$$\begin{aligned}
\boldsymbol{\xi} &\sim N_p(\mathbf{a}, \boldsymbol{\kappa}^{-1}) \\
\boldsymbol{\lambda} &\sim N_p(\mathbf{0}, \boldsymbol{\Gamma}) \\
\boldsymbol{\Sigma}^{-1}|\mathbf{B} &\sim \mathcal{W}_p(2\alpha, (2\mathbf{B})^{-1}) \\
\mathbf{B} &\sim \mathcal{W}_p(2\gamma, (2\mathbf{H})^{-1}) \\
\log\left(\frac{1}{\nu}\right) &\sim U(-10, 10),
\end{aligned}$$

where $\mathbf{a}$ and $\mathbf{0}$ are $p \times 1$ vector, $\boldsymbol{\kappa}$, $\mathbf{B}$, $\mathbf{H}$, and $\boldsymbol{\Gamma}$ are positive definite matrix, α and γ are scalars.

The prior densities are as follows:

$$\begin{aligned}
\pi(\boldsymbol{\xi}) &= (2\pi)^{-\frac{p}{2}}|\boldsymbol{\kappa}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\xi} - \mathbf{a})^\top \boldsymbol{\kappa}(\boldsymbol{\xi} - \mathbf{a})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\xi} - \mathbf{a})^\top \boldsymbol{\kappa}(\boldsymbol{\xi} - \mathbf{a})\right\} \\
\pi(\boldsymbol{\Sigma}^{-1}|\mathbf{B}) &= \frac{|\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(2\alpha-p-1)}\exp\left\{-\frac{1}{2}\text{tr}\left(((2\mathbf{B})^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\right)\right\}}{2^{\frac{1}{2}p(2\alpha)}|(2\mathbf{B})^{-1}|^{\frac{1}{2}(2\alpha)}\Gamma_p(\frac{1}{2}(2\alpha))} \\
&\propto |\mathbf{B}|^{\alpha}|\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(2\alpha-p-1)}\exp\{-\text{tr}(\mathbf{B}\boldsymbol{\Sigma}^{-1})\} \\
\pi(\mathbf{B}) &= \frac{|\mathbf{B}|^{\frac{1}{2}(2\gamma-p-1)}\exp\left\{-\frac{1}{2}\text{tr}\left(\left((2\mathbf{H})^{-1}\right)^{-1}\mathbf{B}\right)\right\}}{2^{\frac{1}{2}p(2\gamma)}|(2\mathbf{H})^{-1}|^{\frac{1}{2}(2\gamma)}\Gamma_p(\frac{1}{2}(2\gamma))} \\
&\propto |\mathbf{B}|^{\frac{1}{2}(2\gamma-p-1)}\exp\{-\text{tr}(\mathbf{H}\mathbf{B})\}
\end{aligned}$$

$$\begin{aligned}
\pi(\boldsymbol{\lambda}) &= (2\pi)^{-\frac{p}{2}}|\boldsymbol{\Gamma}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda}\right\} \\
\pi(\nu) &= \frac{1}{20\nu} \\
&\propto \frac{1}{\nu},
\end{aligned}$$

where $\Gamma_p(\frac{1}{2}(2\mathrm{x})) = \prod_{i=1}^{p}\Gamma_p(\frac{1}{2}(2\mathrm{x}+1-i))$. Let $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$. The joint prior density for $(\boldsymbol{\theta}, \mathbf{B})$ is

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \mathbf{B}) \propto\ & |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(2\alpha-p-1)} \times |\mathbf{B}|^{\alpha+\frac{1}{2}(2\gamma-p-1)} \times \frac{1}{\nu} \\
& \times\exp\left\{-\frac{1}{2}(\boldsymbol{\xi}-\mathbf{a})^\top\boldsymbol{\kappa}(\boldsymbol{\xi}-\mathbf{a}) - \mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}^{-1}) - \mathrm{tr}(\mathbf{H}\mathbf{B}) - \frac{1}{2}\boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda}\right\}.
\end{aligned}$$

The complete-data likelihood of $(\mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\gamma})$ is

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\gamma}) &= \ell_c(\boldsymbol{\theta}|\mathbf{Y}^{\mathrm{o}}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\tau}, \boldsymbol{\gamma}) \\
&= \prod_{j=1}^{n} f(\mathbf{Y}_j^{\mathrm{o}}, \mathbf{Y}_j^{\mathrm{m}}, \tau_j, \boldsymbol{\gamma}_j|\boldsymbol{\theta}) \\
&= \prod_{j=1}^{n} f(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) f(\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \boldsymbol{\gamma}_j|\boldsymbol{\theta}) \\
&= \prod_{j=1}^{n} f(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) f(\tau_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) f(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}) f(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\theta}) \\
&\propto \prod_{j=1}^{n} \left(\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1} |\boldsymbol{\Sigma}^{-1}|^{1/2} \tau_j^{\left(\frac{\nu}{2}\right)} \exp\left(-\frac{\tau_j}{2}(\Delta_j + \nu)\right)\right),
\end{aligned}$$

where $\Delta_j = (\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)$.

The joint posterior density function is as follows

$$p(\boldsymbol{\theta}, \mathbf{B}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\tau}, \boldsymbol{\gamma}|\mathbf{Y}^{\mathrm{o}}) \propto \pi(\boldsymbol{\theta}, \mathbf{B})\ell_c(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\gamma}). \tag{J.1}$$

By using Equation (J.1), we have the following posterior distributions.

The conditional distribution of $\boldsymbol{\gamma}_j$ given $\mathbf{Y}_j^{\mathrm{o}}$ and $\boldsymbol{\theta}$ is as follows

$$p(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}) \propto p(\boldsymbol{\theta}, \mathbf{B}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\tau}, \boldsymbol{\gamma}|\mathbf{Y}^{\mathrm{o}}) \propto f(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\theta}).$$

It implies that

$$\boldsymbol{\gamma}_j|(\mathbf{Y}_j^{\rm o},\boldsymbol{\theta}) \sim Tt_p\Big(\mathbf{q}_j^*,\frac{\nu+U_j^{\rm o}}{\nu+p_j^{\rm o}}\boldsymbol{\Delta}_j^*,\nu+p_j^{\rm o};\ \mathbb{R}_+^p\Big),$$

by Theorem 2.1(d). The conditional distribution of τ_j given $\mathbf{Y}_j^{\rm o}$, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\theta}$ is as follows

$$p(\tau_j|\mathbf{Y}_j^{\rm o},\boldsymbol{\gamma}_j,\boldsymbol{\theta}) \propto p(\boldsymbol{\theta},\mathbf{B},\mathbf{Y}^{\rm m},\boldsymbol{\tau},\boldsymbol{\gamma}|\mathbf{Y}^{\rm o}) \propto f(\tau_j|\mathbf{Y}_j^{\rm o},\boldsymbol{\gamma}_j,\boldsymbol{\theta}).$$

It implies that

$$\tau_j|(\mathbf{Y}_j^{\rm o},\boldsymbol{\gamma}_j,\boldsymbol{\theta}) \sim \Gamma\Big(\frac{p+p_j^{\rm o}+\nu}{2},\frac{(\boldsymbol{\gamma}_j-\mathbf{q}_j^*)^\top{\boldsymbol{\Delta}_j^*}^{-1}(\boldsymbol{\gamma}_j-\mathbf{q}_j^*)+U_j^{\rm o}+\nu}{2}\Big),$$

by Theorem 2.1(e). The conditional distribution of $\mathbf{Y}_j^{\rm m}$ given $\mathbf{Y}_j^{\rm o}$, τ_j, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\theta}$ is as follows

$$p(\mathbf{Y}_j^{\rm m}|\mathbf{Y}_j^{\rm o},\boldsymbol{\tau}_j,\boldsymbol{\gamma}_j,\boldsymbol{\theta}) \propto p(\boldsymbol{\theta},\mathbf{B},\mathbf{Y}^{\rm m},\boldsymbol{\tau},\boldsymbol{\gamma}|\mathbf{Y}^{\rm o}) \propto f(\mathbf{Y}_j^{\rm m}|\mathbf{Y}_j^{\rm o},\tau_j,\boldsymbol{\gamma}_j,\boldsymbol{\theta}).$$

It implies that

$$\mathbf{Y}_j^{\rm m}|(\tau_j,\boldsymbol{\gamma}_j,\mathbf{Y}^{\rm o},\boldsymbol{\theta}) \sim N_{p_j^{\rm m}}(\boldsymbol{\zeta}_j^{\rm m\cdot o},\tau_j^{-1}\boldsymbol{\Sigma}_j^{\rm mm\cdot o}),$$

by Theorem 2.1(b).

The conditional distribution of $\mathbf{B}$ given $\mathbf{Y}_j^{\rm o}$, $\mathbf{Y}_j^{\rm m}$, τ_j, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\theta}$ is as follows

$$\begin{aligned} p(\mathbf{B}|\tau_j,\mathbf{Y}_j^{\rm o},\mathbf{Y}_j^{\rm m},\boldsymbol{\gamma}_j,\boldsymbol{\theta}) &\propto |\mathbf{B}|^{\alpha+\frac{1}{2}(2\gamma-p-1)}\exp\left\{-\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}^{-1})-\mathrm{tr}(\mathbf{H}\mathbf{B})\right\} \\ &= |\mathbf{B}|^{\frac{1}{2}(2\alpha+2\gamma-p-1)}\exp\left\{-\mathrm{tr}\left((\boldsymbol{\Sigma}^{-1}+\mathbf{H})\mathbf{B}\right)\right\} \\ &= |\mathbf{B}|^{\frac{1}{2}(2(\alpha+\gamma)-p-1)}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(2(\boldsymbol{\Sigma}^{-1}+\mathbf{H})\mathbf{B}\right)\right\} \\ &= |\mathbf{B}|^{\frac{1}{2}(2\gamma^*-p-1)}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(2\mathbf{H}^*\mathbf{B}\right)\right\}, \end{aligned}$$

where $\mathbf{H}^*=\boldsymbol{\Sigma}^{-1}+\mathbf{H}$ and $\gamma^*=\alpha+\gamma$.

Thus,

$$\mathbf{B}|(\tau_j,\mathbf{Y}_j^{\rm o},\mathbf{Y}_j^{\rm m},\boldsymbol{\gamma}_j,\boldsymbol{\theta}) \sim \mathcal{W}_p\left(2\gamma^*,(2\mathbf{H}^*)^{-1}\right).$$

The conditional distribution of $\boldsymbol{\Sigma}^{-1}$ given $\boldsymbol{\xi}$,$\boldsymbol{\Lambda}$,$\mathbf{Y}^{\mathrm{o}}$,$\mathbf{Y}^{\mathrm{m}}$,τ,$\boldsymbol{\gamma}$, and $\mathbf{B}$ is as follows

$$\begin{aligned}
& p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{\xi},\boldsymbol{\Lambda},\mathbf{Y}^{\mathrm{o}},\mathbf{Y}^{\mathrm{m}},\tau,\boldsymbol{\gamma},\mathbf{B}) \\
\propto\ & |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(n+2\alpha-p-1)}\exp\left\{-\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_i^{-1})-\frac{1}{2}\sum_{j=1}^{n}\tau_j\Delta_j\right\} \\
=\ & |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(n+2\alpha-p-1)}\exp\{ \\
& -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\left(2\mathbf{B}+\sum_{j=1}^{n}\tau_j(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^{\top}\right)\right)\Bigg\} \\
=\ & |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(\alpha^*-p-1)}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{B}^*\right)\right\},
\end{aligned}$$

where $\alpha^* = n+2\alpha$ and $\mathbf{B}^* = 2\mathbf{B}+\sum_{j=1}^{n}\tau_j(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^{\top}$.

Thus,

$$\boldsymbol{\Sigma}^{-1}|(\boldsymbol{\xi},\boldsymbol{\Lambda},\mathbf{Y}^{\mathrm{o}},\mathbf{Y}^{\mathrm{m}},\tau,\boldsymbol{\gamma},\mathbf{B}) \sim \mathcal{W}_p\left(\alpha^*,(\mathbf{B}^*)^{-1}\right).$$

The conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{\Sigma}^{-1}$,$\boldsymbol{\Lambda}$,$\mathbf{Y}^{\mathrm{o}}$,$\mathbf{Y}^{\mathrm{m}}$,$\boldsymbol{\tau}$,$\boldsymbol{\gamma}$, and $\mathbf{B}$ is as follows

$$p(\boldsymbol{\xi}|\boldsymbol{\Sigma}^{-1},\boldsymbol{\Lambda},\mathbf{Y}^{\mathrm{o}},\mathbf{Y}^{\mathrm{m}},\boldsymbol{\tau},\boldsymbol{\gamma},\mathbf{B}) \ \propto\ \exp\left\{-\frac{1}{2}\left((\boldsymbol{\xi}-\mathbf{a})^{\top}\boldsymbol{\kappa}(\boldsymbol{\xi}-\mathbf{a})+\sum_{j=1}^{n}\tau_j\Delta_j\right)\right\},$$

where $(\boldsymbol{\xi}-\mathbf{a})^{\top}\boldsymbol{\kappa}(\boldsymbol{\xi}-\mathbf{a})+\sum_{j=1}^{n}\tau_j\Delta_j$ can be expressed as

$$\begin{aligned}
& \sum_{j=1}^{n}\tau_j\Delta_j+(\boldsymbol{\xi}-\mathbf{a})^{\top}\boldsymbol{\kappa}(\boldsymbol{\xi}-\mathbf{a}) \\
=\ & \sum_{j=1}^{n}\left(\tau_j(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\right)+(\boldsymbol{\xi}-\mathbf{a})^{\top}\boldsymbol{\kappa}(\boldsymbol{\xi}-\mathbf{a}) \\
=\ & \sum_{j=1}^{n}\tau_j\left(\mathbf{Y}_j^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{Y}_j+\boldsymbol{\xi}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}+\boldsymbol{\gamma}_j^{\top}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}\boldsymbol{\gamma}_j-2\mathbf{Y}_j^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}-2\boldsymbol{\gamma}_j^{\top}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\mathbf{Y}_j\right. \\
& \left.+2\boldsymbol{\gamma}_j^{\top}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\right)+\boldsymbol{\xi}^{\top}\boldsymbol{\kappa}\boldsymbol{\xi}-2\mathbf{a}^{\top}\boldsymbol{\kappa}\boldsymbol{\xi}+\mathbf{a}^{\top}\boldsymbol{\kappa}\mathbf{a} \\
\propto\ & \boldsymbol{\xi}^{\top}\left(\sum_{j=1}^{n}\tau_j\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\xi}+\boldsymbol{\xi}^{\top}\boldsymbol{\kappa}\boldsymbol{\xi}-2\sum_{j=1}^{n}\tau_j\mathbf{Y}_j^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}+2\sum_{j=1}^{n}\tau_j\boldsymbol{\gamma}_j^{\top}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi} \\
& -2\mathbf{a}^{\top}\boldsymbol{\kappa}\boldsymbol{\xi} \\
=\ & \boldsymbol{\xi}^{\top}\left(\sum_{j=1}^{n}\tau_j\boldsymbol{\Sigma}^{-1}+\boldsymbol{\kappa}\right)\boldsymbol{\xi}-2\left(\sum_{j=1}^{n}\tau_j\mathbf{Y}_j^{\top}\boldsymbol{\Sigma}^{-1}-\sum_{j=1}^{n}\tau_j\boldsymbol{\gamma}_j^{\top}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}+\mathbf{a}^{\top}\boldsymbol{\kappa}\right)\boldsymbol{\xi}
\end{aligned}$$

$$\begin{aligned}
&= \boldsymbol{\xi}^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} - 2\left(\sum_{j=1}^n \tau_j \mathbf{Y}_j^\top \boldsymbol{\Sigma}^{-1} - \sum_{j=1}^n \tau_j \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1} + \mathbf{a}^\top\boldsymbol{\kappa}\right)\boldsymbol{\kappa}^*(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} \\
&= \boldsymbol{\xi}^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} - 2\left(\boldsymbol{\kappa}^*\left(\sum_{j=1}^n \tau_j \mathbf{Y}_j^\top \boldsymbol{\Sigma}^{-1} - \sum_{j=1}^n \tau_j \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1} + \mathbf{a}^\top\boldsymbol{\kappa}\right)^\top\right)^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} \\
&= \boldsymbol{\xi}^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} - 2(\boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} \\
&= \boldsymbol{\xi}^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} - 2(\boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\xi} + (\boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\mu}^* - (\boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\mu}^* \\
&= (\boldsymbol{\xi} - \boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}(\boldsymbol{\xi} - \boldsymbol{\mu}^*) - (\boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}\boldsymbol{\mu}^* \\
&\propto (\boldsymbol{\xi} - \boldsymbol{\mu}^*)^\top(\boldsymbol{\kappa}^*)^{-1}(\boldsymbol{\xi} - \boldsymbol{\mu}^*),
\end{aligned}$$

with $\boldsymbol{\mu}^* = \boldsymbol{\kappa}^*\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^n \tau_j(\mathbf{Y}_j - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\right) + \boldsymbol{\kappa}\mathbf{a}\right]$ and $\boldsymbol{\kappa}^* = (\sum_{j=1}^n \tau_j \boldsymbol{\Sigma}^{-1} + \boldsymbol{\kappa})^{-1}$. Thus,

$$\boldsymbol{\xi}|(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Lambda}, \mathbf{Y}^{\mathrm{o}}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{B}) \sim N_p(\boldsymbol{\mu}^*, \boldsymbol{\kappa}^*).$$

The conditional distribution of $\boldsymbol{\lambda}$ given $\boldsymbol{\xi}$,$\boldsymbol{\Sigma}^{-1}$,$\mathbf{Y}^{\mathrm{o}}$,$\mathbf{Y}^{\mathrm{m}}$,$\boldsymbol{\tau}$,$\boldsymbol{\gamma}$, and $\mathbf{B}$ is as follows

$$p(\boldsymbol{\lambda}|\boldsymbol{\Sigma}^{-1}, \boldsymbol{\xi}, \boldsymbol{\Lambda}, \mathbf{Y}^{\mathrm{o}}, \mathbf{Y}^{\mathrm{m}}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{B}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda} + \sum_{j=1}^n \tau_j \Delta_j\right)\right\},$$

where $\boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda} + \sum_{j=1}^n \tau_j \Delta_j$ can be expressed as

$$\begin{aligned}
& \sum_{j=1}^n \tau_j \Delta_j + \boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda} \\
\propto\ & \sum_{j=1}^n \tau_j \left(\boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{\gamma}_j - 2\boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\xi})\right) + \boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda} \\
=\ & \sum_{j=1}^n \tau_j \left(\boldsymbol{\lambda}^\top \mathrm{Diag}(\boldsymbol{\gamma}_j)\boldsymbol{\Sigma}^{-1}\mathrm{Diag}(\boldsymbol{\gamma}_j)\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \mathrm{Diag}(\boldsymbol{\gamma}_{\mathrm{j}})\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{\mathrm{j}} - \boldsymbol{\xi})\right) + \boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda} \\
=\ & \boldsymbol{\lambda}^\top\left(\sum_{j=1}^n \tau_j \mathrm{Diag}(\boldsymbol{\gamma}_j)\boldsymbol{\Sigma}^{-1}\mathrm{Diag}(\boldsymbol{\gamma}_j)\right)\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \sum_{j=1}^n \tau_j \mathrm{Diag}(\boldsymbol{\gamma}_{\mathrm{j}})\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{\mathrm{j}} - \boldsymbol{\xi}) \\
& + \boldsymbol{\lambda}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\lambda} \\
=\ & \boldsymbol{\lambda}^\top\left(\boldsymbol{\Gamma}^{-1} + \sum_{j=1}^n \tau_j \mathrm{Diag}(\boldsymbol{\gamma}_j)\boldsymbol{\Sigma}^{-1}\mathrm{Diag}(\boldsymbol{\gamma}_j)\right)\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \sum_{j=1}^n \tau_j \mathrm{Diag}(\boldsymbol{\gamma}_{\mathrm{j}})\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{\mathrm{j}} - \boldsymbol{\xi})
\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{\lambda}^\top(\boldsymbol{\Gamma}^*)^{-1}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top\boldsymbol{\lambda}^* \\
&= \boldsymbol{\lambda}^\top(\boldsymbol{\Gamma}^*)^{-1}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top(\boldsymbol{\Gamma}^*)^{-1}\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^* + (\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*)^\top(\boldsymbol{\Gamma}^*)^{-1}(\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*) - (\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*)^\top(\boldsymbol{\Gamma}^*)^{-1}(\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*) \\
&= (\boldsymbol{\lambda} - \boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*)^\top(\boldsymbol{\Gamma}^*)^{-1}(\boldsymbol{\lambda} - \boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*) - (\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*)^\top(\boldsymbol{\Gamma}^*)^{-1}(\boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*) \\
&\propto (\boldsymbol{\lambda} - \boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*)^\top(\boldsymbol{\Gamma}^*)^{-1}(\boldsymbol{\lambda} - \boldsymbol{\Gamma}^*\boldsymbol{\lambda}^*) \\
&\propto (\boldsymbol{\lambda} - \boldsymbol{\delta}^*)^\top(\boldsymbol{\Gamma}^*)^{-1}(\boldsymbol{\lambda} - \boldsymbol{\delta}^*),
\end{aligned}
$$

with $\boldsymbol{\Gamma}^* = \left(\boldsymbol{\Gamma}^{-1} + \sum_{j=1}^n \tau_j \mathrm{Diag}(\boldsymbol{\gamma}_j)\boldsymbol{\Sigma}^{-1}\mathrm{Diag}(\boldsymbol{\gamma}_j)\right)^{-1} = \left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Sigma}^{-1} \odot \sum_{j=1}^n \tau_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^\top\right)^{-1}$ and $\boldsymbol{\delta}^* = \boldsymbol{\Gamma}^* \left(\sum_{j=1}^n \tau_j \mathrm{Diag}(\boldsymbol{\gamma}_\mathrm{j})\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_\mathrm{j} - \boldsymbol{\xi})\right) = \boldsymbol{\Gamma}^*\left(\boldsymbol{\Sigma}^{-1} \odot \sum_{j=1}^n \tau_j \boldsymbol{\gamma}_j (\mathbf{Y}_j - \boldsymbol{\xi})^\top\right)\mathbf{1}_p$.

Thus,

$$
\boldsymbol{\lambda}|(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\xi}, \boldsymbol{\Lambda}, \mathbf{Y}^\mathrm{o}, \mathbf{Y}^\mathrm{m}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{B}) \sim N_p(\boldsymbol{\delta}^*, \boldsymbol{\Gamma}^*).
$$

K. Proof of Theorem 3.1(c), (d), and (e)

Proof of Theorem 3.1(c):

We partitioned the random vector $\mathbf{Y}_j$ and its parameters $\boldsymbol{\xi}_i$, $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Lambda}_i$ as follows:

$$\mathbf{Y}_j = \begin{bmatrix} \mathbf{Y}_j^{\mathrm{o}} \\ \mathbf{Y}_j^{\mathrm{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\mathbf{Y}_j \\ \mathbf{M}_j\mathbf{Y}_j \end{bmatrix}, \quad \boldsymbol{\xi}_i = \begin{bmatrix} \boldsymbol{\xi}_{ij}^{\mathrm{o}} \\ \boldsymbol{\xi}_{ij}^{\mathrm{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\xi}_i \\ \mathbf{M}_j\boldsymbol{\xi}_i \end{bmatrix},$$

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_{ij}^{\mathrm{oo}} & \boldsymbol{\Sigma}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Sigma}_{ij}^{\mathrm{mo}} & \boldsymbol{\Sigma}_{ij}^{\mathrm{mm}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top & \mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{M}_j^\top \\ \mathbf{M}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top & \mathbf{M}_j\boldsymbol{\Sigma}_i\mathbf{M}_j^\top \end{bmatrix},$$

and

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_j\boldsymbol{\Lambda}_i\mathbf{O}_j^\top & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{M}_j\boldsymbol{\Lambda}_i\mathbf{M}_j^\top \end{bmatrix}.$$

By Lemma A.1

$$\boldsymbol{\Omega}_i = \boldsymbol{\Sigma}_i + \boldsymbol{\Lambda}_i^2 = \begin{bmatrix} \mathbf{O}_j(\boldsymbol{\Sigma}_i + \boldsymbol{\Lambda}_i^2)\mathbf{O}_j^\top & \mathbf{O}_j\boldsymbol{\Sigma}_i\mathbf{M}_j^\top \\ \mathbf{M}_j\boldsymbol{\Sigma}_i\mathbf{O}_j^\top & \mathbf{M}_j(\boldsymbol{\Sigma}_i + \boldsymbol{\Lambda}_i^2)\mathbf{M}_j^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}} & \boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Omega}_{ij}^{\mathrm{mo}} & \boldsymbol{\Omega}_{ij}^{\mathrm{mm}} \end{bmatrix}.$$

Note that $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Lambda}_i$ are symmetric matrix. Thus, $\boldsymbol{\Omega}_i$, $\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}$ and $\boldsymbol{\Omega}_{ij}^{\mathrm{mm}}$ are symmetric matrices and ${\boldsymbol{\Omega}_{ij}^{\mathrm{om}}}^\top = \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}$.

Let $\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1} = \mathbf{B}_i = \begin{bmatrix} \mathbf{B}_{i1} & \mathbf{B}_{i2} \end{bmatrix}$. Thus, we have

$$\begin{aligned}
\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1} &= \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}} & \boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Omega}_{ij}^{\mathrm{mo}} & \boldsymbol{\Omega}_{ij}^{\mathrm{mm}} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \begin{bmatrix} {\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1}(\boldsymbol{\Omega}_{ij}^{\mathrm{om}}{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}{\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1} + \mathbf{I}_{p_j^{\mathrm{o}}}) & -{\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1} \\ -{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}{\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1} & {\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}({\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}{\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1} + {\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1}) & -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}{\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1} \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}{\boldsymbol{\Omega}_{ij}^{\mathrm{oo}}}^{-1} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}{\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{B}_{i1} & \mathbf{B}_{i2} \end{bmatrix},
\end{aligned}$$

where $\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}} = \boldsymbol{\Omega}_{ij}^{\mathrm{mm}} - \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}$. Thus, we have

$$\begin{aligned}
\mathbf{B}_{i1} &= \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}(\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} + \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}) \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}(\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \mathbf{I}_{p_j^{\mathrm{o}}}) \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}
\end{aligned}$$

and

$$\mathbf{B}_{i2} = \begin{bmatrix} -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}} \\ \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}.$$

From the above calculation, we have the following results

$$\begin{aligned}
&\mathbf{B}_{i1} + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
= &\begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}(\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \mathbf{I}_{p_j^{\mathrm{o}}}) \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} + \begin{bmatrix} -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
= &\begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} - \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} + \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} \\
= &\begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} \\ \mathbf{0}_{p_j^{\mathrm{m}} \times p_j^{\mathrm{o}}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}},
\end{aligned}$$

$$\begin{aligned}
&\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Lambda}_i \\
= &\begin{bmatrix} \mathbf{B}_{i1} & \mathbf{B}_{i2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \\
= &\begin{bmatrix} \mathbf{B}_{i1}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{B}_{i2}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \\
= &\begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}} + \mathbf{I}_{p_j^{\mathrm{o}}})\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix},
\end{aligned} \tag{K.1}$$

and

$$
\begin{aligned}
&\mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top} \\
= &\begin{bmatrix} -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}(\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}})^{\top}\begin{bmatrix} -(\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}})^{\top} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}^{\top}} \end{bmatrix} \\
= &\begin{bmatrix} -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}} \\ \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\begin{bmatrix} -(\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}) & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix} \\
= &\begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & -\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \\ -\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \boldsymbol{\Lambda}_{ij}^{\mathrm{mm}}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{mm}} \end{bmatrix}. \quad \text{(K.2)}
\end{aligned}
$$

Since

$$\boldsymbol{\Delta}_i + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top} = \mathbf{I}_p - \boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Lambda}_i + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top}\,, \tag{K.3}$$

we substituted $(K.1)$, $(K.2)$ to $(K.3)$ to obtain

$$
\begin{aligned}
\boldsymbol{\Delta}_i + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top} &= \mathbf{I}_p - \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{m}}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}} & \mathbf{0}_{p_j^{\mathrm{o}}\times p_j^{\mathrm{m}}} \\ \mathbf{0}_{p_j^{\mathrm{m}}\times p_j^{\mathrm{o}}} & \mathbf{I}_{p_j^{\mathrm{m}}} \end{bmatrix}.
\end{aligned}
$$

Since

$$\mathbf{Y}_j \mid (\boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1) \;\sim\; N_p(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}_i/\tau_j)$$

and

$$\boldsymbol{\gamma}_j \mid (\tau_j, Z_{ij} = 1) \;\sim\; HN_p(\mathbf{0}, \mathbf{I}_p/\tau_j),$$

we have

$$f(\mathbf{Y}_j \mid \boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1) = \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}_i/\tau_j)$$

and

$$f(\boldsymbol{\gamma}_j \mid \tau_j, Z_{ij}=1) = 2^p \phi_p(\boldsymbol{\gamma}_j|\mathbf{0}, \mathbf{I}_p/\tau_j) I_{R_+^p}(\boldsymbol{\gamma}_j).$$

By Lemma B.1, we have

$$\begin{aligned} & (\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j) + \boldsymbol{\gamma}_j^\top \mathbf{I}_p^{-1}\boldsymbol{\gamma}_j \\ = \; & (\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i) + (\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i))^\top \boldsymbol{\Delta}_i^{-1}(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)), \end{aligned}$$

where $\boldsymbol{\Omega}_i = \boldsymbol{\Sigma}_i + \boldsymbol{\Lambda}_i^2$ and $\boldsymbol{\Delta}_i = (\mathbf{I}_p + \boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i)^{-1}$. Thus,

$$\begin{aligned} & \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}_i/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\mathbf{0}, \mathbf{I}_p/\tau_j) I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j) \\ = \; & \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i), \boldsymbol{\Delta}_i/\tau_j) I_{\mathbb{R}_+^p}(\boldsymbol{\gamma}_j). \end{aligned}$$

Thus,

$$\begin{aligned} f(\mathbf{Y}_j|Z_{ij}=1,\tau_j,\boldsymbol{\Theta}) &= \int_{\mathbb{R}_+^p} f(\mathbf{Y}_j \mid \boldsymbol{\gamma}_j, \tau_j, Z_{ij}=1, \boldsymbol{\Theta}) f(\boldsymbol{\gamma}_j \mid \tau_j, Z_{ij}=1, \boldsymbol{\Theta}) d\boldsymbol{\gamma}_j \\ &= \int_{\mathbb{R}_+^p} 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j, \boldsymbol{\Sigma}_i/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\mathbf{0}, \mathbf{I}_p/\tau_j) d\boldsymbol{\gamma}_j \\ &= \int_{\mathbb{R}_+^p} 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i/\tau_j) \cdot \phi_p(\boldsymbol{\gamma}_j|\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i), \boldsymbol{\Delta}_i/\tau_j) d\boldsymbol{\gamma}_j \\ &= 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i/\tau_j) \int_{\mathbb{R}_+^p} \phi_p(\boldsymbol{\gamma}_j|\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i), \boldsymbol{\Delta}_i/\tau_j) d\boldsymbol{\gamma}_j \\ &= 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i/\tau_j) \Phi_p(\sqrt{\tau_j}\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)|\boldsymbol{\Delta}_i). \end{aligned}$$

Thus,

$$\begin{aligned} & f(\mathbf{Y}_j^{\mathrm{o}}, \mathbf{Y}_j^{\mathrm{m}}|Z_{ij}=1,\tau_j,\boldsymbol{\Theta}) \\ = \; & 2^p \phi_p(\mathbf{Y}_j|\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i/\tau_j) \Phi_p(\sqrt{\tau_j}\boldsymbol{\Lambda}_i\boldsymbol{\Omega}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)|\boldsymbol{\Delta}_i) \\ = \; & 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j) \phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j) \Phi_p\Big(\sqrt{\tau_j}\begin{bmatrix}\mathbf{B}_{i1} & \mathbf{B}_{i2}\end{bmatrix}\begin{bmatrix}\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}} \\ \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}}\end{bmatrix}\Big|\boldsymbol{\Delta}_i\Big) \\ = \; & 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j) \phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j) \\ & \times \Phi_p\Big(\sqrt{\tau_j}\left(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}})\right) \mid \boldsymbol{\Delta}_i\Big), \end{aligned}$$

where condition mean $\boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}} = \boldsymbol{\xi}_{ij}^{\mathrm{m}} + \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})$ and conditional covariance $\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j = (\boldsymbol{\Omega}_{ij}^{\mathrm{mm}} - \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Omega}_{ij}^{\mathrm{om}})/\tau_j$.

The marginal density of $\mathbf{Y}_j^{\mathrm{o}}$ given $Z_{ij} = 1, \tau_j, \boldsymbol{\Theta}$ is as follows

$$\begin{aligned} f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\tau_j,\boldsymbol{\Theta}) &= \int f(\mathbf{Y}_j^{\mathrm{o}}, \mathbf{Y}_j^{\mathrm{m}}|Z_{ij}=1,\tau_j,\boldsymbol{\Theta})\, d\mathbf{Y}_j^{\mathrm{m}} \\ &= 2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j) \int \phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j) \\ &\quad \times \Phi_p\Big(\sqrt{\tau_j}\left(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}})\right) \mid \boldsymbol{\Delta}_i\Big) d\mathbf{Y}_j^{\mathrm{m}}. \end{aligned}$$

Let $\mathbf{z} = \mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}$, then $\mathbf{Y}_j^{\mathrm{m}} = \mathbf{z} + \boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}} = \mathbf{z} + \boldsymbol{\xi}_{ij}^{\mathrm{m}} + \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})$. Thus, we have

$$\phi_{p_j^{\mathrm{m}}}(\mathbf{Y}_j^{\mathrm{m}}|\boldsymbol{\xi}_{ij}^{\mathrm{m\cdot o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j) = \phi_{p_j^{\mathrm{m}}}(\mathbf{z}|\mathbf{0}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j)$$

and

$$\begin{aligned} &\Phi_p\Big(\sqrt{\tau_j}\left(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}(\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_{ij}^{\mathrm{m}})\right) \\ =\; &\Phi_p\Big(\sqrt{\tau_j}\Big(\mathbf{B}_{i1}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}(\mathbf{z} + \boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})\Big) \mid \boldsymbol{\Delta}_i\Big) \\ =\; &\Phi_p\Big(\sqrt{\tau_j}\Big((\mathbf{B}_{i1} + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}\mathbf{z}\Big) \mid \boldsymbol{\Delta}_i\Big). \end{aligned}$$

By Lemma 2.1 of Arellano-Valle and Genton (2005), we have

$$\begin{aligned} &f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\tau_j,\boldsymbol{\Theta}) \\ =\; &2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j) \\ &\times \int \phi_{p_j^{\mathrm{m}}}(\mathbf{z}|\mathbf{0}, \boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}/\tau_j)\Phi_p\Big(\sqrt{\tau_j}\Big((\mathbf{B}_{i1} + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}\mathbf{z}\Big)\Big|\boldsymbol{\Delta}_i\Big)\, d\mathbf{z} \\ =\; &2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j)\mathrm{E}_{\mathbf{z}}\Big\{\Phi_p\Big(\sqrt{\tau_j}\Big((\mathbf{B}_{i1} + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}}) + \mathbf{B}_{i2}\mathbf{Z}\Big)\Big|\boldsymbol{\Delta}_i\Big)\Big\} \\ =\; &2^p \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j)\Phi_p\Big(\sqrt{\tau_j}(\mathbf{B}_{i1} + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}})(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})\Big|\boldsymbol{\Delta}_i + \mathbf{B}_{i2}\boldsymbol{\Omega}_{ij}^{\mathrm{mm\cdot o}}\mathbf{B}_{i2}^{\top}\Big) \\ =\; &2^{p_j^{\mathrm{o}}} \phi_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}}, \boldsymbol{\Omega}_{ij}^{\mathrm{oo}}/\tau_j)\Phi_{p_j^{\mathrm{o}}}\Big(\sqrt{\tau_j}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_{ij}^{\mathrm{o}})\Big|\mathbf{I}_{p_j^{\mathrm{o}}} - \boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\Big). \end{aligned}$$

Since

$$\tau_j \mid (Z_{ij}=1) \quad \sim \quad \Gamma(\nu_i/2, \nu_i/2),$$

we have

$$\begin{aligned}
& f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\boldsymbol{\Theta}) \\
= & \int f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\tau_j,\boldsymbol{\Theta}) f(\tau_j|Z_{ij}=1) d\tau_j \\
= & \left(\frac{\pi}{2}\right)^{-\frac{p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}|^{-\frac{1}{2}} \\
& \times \int_0^\infty \tau_j^{\frac{\nu+p_j^{\mathrm{o}}}{2}-1} \exp\Big(-\tau_j\big[(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\nu_i\big]/2\Big) \\
& \times \Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}}-\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\big) d\tau_j \\
= & \left(\frac{\pi}{2}\right)^{-\frac{p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu_i+p_j^{\mathrm{o}}}{2})}{\Big\{\frac{1}{2}\big[(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\nu_i\big]\Big\}^{\frac{\nu_i+p_j^{\mathrm{o}}}{2}}} \\
& \times \int_0^\infty \Gamma\Big(\frac{\nu_i+p_j^{\mathrm{o}}}{2}, \frac{(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\nu_i}{2}\Big) \\
& \times \Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}}-\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\big) d\tau_j \\
= & \left(\frac{\pi}{2}\right)^{-\frac{p_j^{\mathrm{o}}}{2}} \left(\frac{\nu_i}{2}\right)^{-\frac{p_j^{\mathrm{o}}}{2}} \left(\frac{\nu_i}{2}\right)^{\frac{\nu_i+p_j^{\mathrm{o}}}{2}} |\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu_i+p_j^{\mathrm{o}}}{2})}{\Gamma(\frac{\nu_i}{2})} \\
& \times \left[\frac{(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\nu_i}{2}\right]^{-\frac{\nu_i+p_j^{\mathrm{o}}}{2}} \\
& \times \int_0^\infty \Gamma\Big(\frac{\nu_i+p_j^{\mathrm{o}}}{2}, \frac{(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top \boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\nu_i}{2}\Big) \\
& \times \Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}}-\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\big) d\tau_j
\end{aligned}$$

$$
\begin{aligned}
&= 2^{p_j^{\mathrm{o}}}(\pi\nu_i)^{-\frac{p_j^{\mathrm{o}}}{2}}|\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}|^{-\frac{1}{2}}\frac{\Gamma(\frac{\nu_i+p_j^{\mathrm{o}}}{2})}{\Gamma(\frac{\nu_i}{2})}\left[\frac{(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})}{\nu_i}+1\right]^{-\frac{\nu_i+p_j^{\mathrm{o}}}{2}} \\
&\quad\times\int_0^\infty\Gamma\Big(\frac{\nu_i+p_j^{\mathrm{o}}}{2},\frac{(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})^\top\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})+\nu_i}{2}\Big) \\
&\quad\times\Phi_{p_j^{\mathrm{o}}}\big(\sqrt{\tau_j}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\big|\mathbf{I}_{p_j^{\mathrm{o}}}-\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\big)d\tau_j \\
&= 2^{p_j^{\mathrm{o}}}\times t_{p_j^{\mathrm{o}}}(\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Omega}_{ij}^{\mathrm{oo}},\nu_i)T_{p_j^{\mathrm{o}}}\bigg(\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}(\mathbf{Y}_j^{\mathrm{o}}-\boldsymbol{\xi}_{ij}^{\mathrm{o}})\Big(\frac{\nu_i+p_j^{\mathrm{o}}}{U_{ij}^{\mathrm{o}}+\nu_i}\Big)^{\frac{1}{2}}\mid\boldsymbol{\Delta}_{ij}^{\mathrm{oo}};\nu_i+p_j^{\mathrm{o}}\bigg),
\end{aligned}
$$

where $U_{ij}^{\mathrm{o}}=(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\mathrm{oo}}(\mathbf{Y}_j-\boldsymbol{\xi}_i)$, $\boldsymbol{\Delta}_{ij}^{\mathrm{oo}}=\mathbf{I}_{p_j^{\mathrm{o}}}-\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}}$, and $\mathbf{C}_{ij}^{\mathrm{oo}}=\mathbf{O}_j^\top\boldsymbol{\Omega}_{ij}^{\mathrm{oo}^{-1}}\mathbf{O}_j$ with the last equality is by Lemma F.1.

Thus,

$$\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\boldsymbol{\Theta}\sim ST_{p_j^{\mathrm{o}}}\Big(\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}},\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}},\nu_i\Big).$$

Hence, it implies that

$$f(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\Theta}) = \sum_{i=1}^g f(\mathbf{Y}_j^{\mathrm{o}}|Z_{ij}=1,\boldsymbol{\Theta})p(Z_{ij}=1)=\sum_{i=1}^g w_i f_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\boldsymbol{\xi}_{ij}^{\mathrm{o}},\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}},\boldsymbol{\Lambda}_{ij}^{\mathrm{oo}},\nu_i).$$

Proof of Theorem 3.1(d):

The conditional density of $\mathbf{Y}_j^{\mathrm{o}}$, $\boldsymbol{\gamma}_j$, and τ_j given $Z_{ij}=1$ is as follows

$$
\begin{aligned}
&f(\mathbf{Y}_j^{\mathrm{o}},\boldsymbol{\gamma}_j,\tau_j|Z_{ij}=1) \\
&= f(\mathbf{Y}_j^{\mathrm{o}}\mid\boldsymbol{\gamma}_j,\tau_j,Z_{ij}=1)f(\boldsymbol{\gamma}_j\mid\tau_j,Z_{ij}=1)f(\tau_j|Z_{ij}=1) \\
&= (2\pi)^{-\frac{p_j^{\mathrm{o}}}{2}}\left|\frac{\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}}{\tau_j}\right|^{-\frac{1}{2}}\exp\Big(-[\mathbf{Y}_j^{\mathrm{o}}-\mathbf{O}_j(\boldsymbol{\xi}_i+\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)]^\top(\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}/\tau_j)^{-1}[\mathbf{Y}_j^{\mathrm{o}}-\mathbf{O}_j(\boldsymbol{\xi}_i+\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)]/2\Big) \\
&\quad\times 2^p(2\pi)^{-\frac{p}{2}}|\frac{\mathbf{I}_p}{\tau_j}|^{-\frac{1}{2}}\exp\Big(-\boldsymbol{\gamma}_j^\top(\frac{\mathbf{I}_p}{\tau_j})^{-1}\boldsymbol{\gamma}_j/2\Big)\frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})}\tau_j^{\frac{\nu_i}{2}-1}\exp(-\nu_i\tau_j/2) \\
&= 2^{\frac{p_j^{\mathrm{m}}}{2}}\pi^{-\frac{p+p_j^{\mathrm{o}}}{2}}\frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})}|\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}|^{-\frac{1}{2}}\tau_j^{(\frac{\nu_i+p+p_j^{\mathrm{o}}}{2})-1} \\
&\quad\times\exp\Big(-\tau_j[\big(\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)\big)^\top\boldsymbol{\Lambda}_i^\top\mathbf{S}_{ij}^{\mathrm{oo}}\boldsymbol{\Lambda}_i\big(\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)\big)+\boldsymbol{\gamma}_j^\top\mathbf{I}_p\boldsymbol{\gamma}_j+\nu_i]/2\Big).
\end{aligned}
$$

The conditional density of $\mathbf{Y}_j^{\text{o}}$ and $\boldsymbol{\gamma}_j$ given $Z_{ij}=1$ is as follows

$$
\begin{aligned}
& f(\mathbf{Y}_j^{\text{o}},\boldsymbol{\gamma}_j|Z_{ij}=1) \\
= & \int_0^{\infty} f(\mathbf{Y}_j^{\text{o}},\boldsymbol{\gamma}_j,\tau_j|Z_{ij}=1)d\tau_j \\
= & 2^{\frac{p_j^{\text{m}}}{2}}\pi^{-\frac{p+p_j^{\text{o}}}{2}}\frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})}|\boldsymbol{\Sigma}_{ij}^{\text{oo}}|^{-\frac{1}{2}}\int_0^{\infty}\tau_j^{(\frac{\nu+p+p_j^{\text{o}}}{2})-1} \\
& \times\exp\Big\{-\tau_j\Big([\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi}_i)]^{\top}\mathbf{S}_{ij}^{\text{oo}}[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi}_i)]+\boldsymbol{\gamma}_j^{\top}\boldsymbol{\gamma}_j+\nu_i\Big)/2\Big\}d\tau_j \\
= & \Gamma\left(\frac{\nu+p+p_j^{\text{o}}}{2}\right)\left(\frac{[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi}_i)]^{\top}\mathbf{S}_{ij}^{\text{oo}}[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi}_i)]+\boldsymbol{\gamma}_j^{\top}\boldsymbol{\gamma}_j+\nu_i}{2}\right)^{-\frac{\nu+p+p_j^{\text{o}}}{2}} \\
& \times 2^{\frac{p_j^{\text{m}}}{2}}\pi^{-\frac{p+p_j^{\text{o}}}{2}}\frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})}|\boldsymbol{\Sigma}_{ij}^{\text{oo}}|^{-\frac{1}{2}}.
\end{aligned}
$$

Let $\mathbf{x}=\boldsymbol{\gamma}_j$, $\mathbf{a}=\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)$, $\mathbf{b}=\mathbf{0}$, $\mathbf{Q}_1=\boldsymbol{\Lambda}_i\mathbf{S}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i$ and $\mathbf{Q}_2=\mathbf{I}_p$. By Lemma B.1, we have

$$
\begin{aligned}
& \big([\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)]^{\top}\boldsymbol{\Lambda}_i\mathbf{S}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i[\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)]+\boldsymbol{\gamma}_j^{\top}\boldsymbol{\gamma}_j\big) \\
& =\Big((\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*)^{\top}\boldsymbol{\Delta}_{ij}^{*^{-1}}(\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*)+U_{ij}^{\text{o}}\Big),
\end{aligned}
$$

where $\mathbf{q}_{ij}^*=\boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\text{oo}}(\mathbf{Y}_j-\boldsymbol{\xi}_i)$ and $\boldsymbol{\Delta}_{ij}^*=\mathbf{I}_p-\boldsymbol{\Lambda}_i\mathbf{C}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i$.

Thus,

$$
\begin{aligned}
& f(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\text{o}},Z_{ij}=1)\propto f(\mathbf{Y}_j^{\text{o}},\boldsymbol{\gamma}_j|Z_{ij}=1) \\
\propto & \big([\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi}_i)]^{\top}\mathbf{S}_{ij}^{\text{oo}}[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j-(\mathbf{Y}_j-\boldsymbol{\xi}_i)]+\boldsymbol{\gamma}_j^{\top}\boldsymbol{\gamma}_j+\nu_i\big)^{-\frac{\nu_i+p+p_j^{\text{o}}}{2}} \\
= & \big([\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)]^{\top}\boldsymbol{\Lambda}_i^{\top}\mathbf{S}_{ij}^{\text{oo}}\boldsymbol{\Lambda}_i[\boldsymbol{\gamma}_j-\boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i)]+\boldsymbol{\gamma}_j^{\top}\boldsymbol{\gamma}_j+\nu_i\big)^{-\frac{\nu_i+p+p_j^{\text{o}}}{2}} \\
= & \Big([\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*]^{\top}\boldsymbol{\Delta}_{ij}^{*^{-1}}[\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*]+(U_{ij}^{\text{o}}+\nu_i)\Big)^{-\frac{\nu_i+p+p_j^{\text{o}}}{2}} \\
\propto & \Big(\frac{[\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*]^{\top}\boldsymbol{\Delta}_{ij}^{*^{-1}}[\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*]}{U_{ij}^{\text{o}}+\nu_i}+1\Big)^{-\frac{\nu_i+p+p_j^{\text{o}}}{2}} \\
= & \Big(\frac{[\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*]^{\top}\boldsymbol{\Delta}_{ij}^{*^{-1}}[\boldsymbol{\gamma}_j-\mathbf{q}_{ij}^*]}{p_j^{\text{o}}+\nu_i}\frac{p_j^{\text{o}}+\nu}{U_{ij}^{\text{o}}+\nu}+1\Big)^{-\frac{\nu+p+p_j^{\text{o}}}{2}}.
\end{aligned}
$$

It implies $\boldsymbol{\gamma}_j|(\mathbf{Y}_j^{\mathrm{o}}, Z_{ij}=1) \sim Tt_p(\mathbf{q}_{ij}^*, \frac{U_{ij}^{\mathrm{o}}+\nu_i}{p_j^{\mathrm{o}}+\nu_i}\boldsymbol{\Delta}_{ij}^*, \nu_i + p_j^{\mathrm{o}}; \mathbb{R}_p^+)$.

Proof of Theorem 3.1(e):

The conditional density of τ_j given $\mathbf{Y}_j^{\mathrm{o}}$, $\boldsymbol{\gamma}_j$, and $Z_{ij}=1$ is as follows

$$f(\tau_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, Z_{ij}=1) = \frac{f(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j|Z_{ij}=1)}{f(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j|Z_{ij}=1)}$$

$$\begin{aligned} = \;& \frac{1}{\Gamma\left(\frac{\nu+p+p_j^{\mathrm{o}}}{2}\right)}\left(\frac{[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)]^\top \mathbf{S}_{ij}^{\mathrm{oo}}[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)] + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j + \nu_i}{2}\right)^{\frac{\nu_i+p+p_j^{\mathrm{o}}}{2}} \\ & \times \tau_j^{(\frac{\nu_i+p+p_j^{\mathrm{o}}}{2})-1}\exp\left\{-\tau_j\Big([\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)]^\top \mathbf{S}_{ij}^{\mathrm{oo}}[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)]\right. \\ & \left. + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j + \nu_i\Big)/2\right\}. \end{aligned}$$

It implies that

$$\begin{aligned} & \tau_j|(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, Z_{ij}=1) \\ \sim \;& \Gamma\left(\frac{\nu_i+p+p_j^{\mathrm{o}}}{2}, \frac{[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)]^\top \mathbf{S}_{ij}^{\mathrm{oo}}[\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)] + \boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j + \nu_i}{2}\right). \end{aligned}$$

By Lemma B.1, we have

$$\tau_j|(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, Z_{ij}=1) \sim \Gamma\left(\frac{\nu_i+p+p_j^{\mathrm{o}}}{2}, \frac{(\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*)^\top \boldsymbol{\Delta}_{ij}^{*^{-1}}(\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*) + U_{ij}^{\mathrm{o}} + \nu_i}{2}\right).$$

L. Proof of Corollary 3.1

The conditional density of $\mathbf{Y}_j^{\text{o}}$, $\boldsymbol{\gamma}_j$, and τ_j given $Z_{ij}=1$ is as follows

$$
\begin{aligned}
& f(\mathbf{Y}_j^{\text{o}}, \boldsymbol{\gamma}_j, \tau_j | Z_{ij}=1) \\
= \; & f(\mathbf{Y}_j^{\text{o}} \mid \boldsymbol{\gamma}_j, \tau_j, Z_{ij}=1) f(\boldsymbol{\gamma}_j \mid \tau_j, Z_{ij}=1) f(\tau_j | Z_{ij}=1) \\
= \; & (2\pi)^{-\frac{p_j^{\text{o}}}{2}} \left| \frac{\boldsymbol{\Sigma}_{ij}^{\text{oo}}}{\tau_j} \right|^{-\frac{1}{2}} \exp\Big(-[\mathbf{Y}_j^{\text{o}} - \mathbf{O}_j(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)]^\top (\boldsymbol{\Sigma}_{ij}^{\text{oo}}/\tau_j)^{-1} [\mathbf{Y}_j^{\text{o}} - \mathbf{O}_j(\boldsymbol{\xi}_i + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)]/2 \Big) \\
& \times 2^p (2\pi)^{-\frac{p}{2}} |\frac{\mathbf{I}_p}{\tau_j}|^{-\frac{1}{2}} \exp\Big(-\boldsymbol{\gamma}_j^\top (\frac{\mathbf{I}_p}{\tau_j})^{-1} \boldsymbol{\gamma}_j/2 \Big) \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} \tau_j^{\frac{\nu_i}{2}-1} \exp(-\nu_i\tau_j/2) \\
= \; & 2^{\frac{p_j^{\text{m}}}{2}} \pi^{-\frac{p+p_j^{\text{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Sigma}_{ij}^{\text{oo}}|^{-\frac{1}{2}} \tau_j^{(\frac{\nu_i+p+p_j^{\text{o}}}{2})-1} \exp\Big(-\tau_j\boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_j/2 \Big) \exp\Big(-\frac{\nu_i}{2}\tau_j \Big) \\
& \times \exp\Big(-\tau_j [\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)]^\top \mathbf{S}_{ij}^{\text{oo}} [\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j - (\mathbf{Y}_j - \boldsymbol{\xi}_i)]/2 \Big) \\
= \; & 2^{\frac{p_j^{\text{m}}}{2}} \pi^{-\frac{p+p_j^{\text{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Sigma}_{ij}^{\text{oo}}|^{-\frac{1}{2}} \tau_j^{(\frac{\nu_i+p+p_j^{\text{o}}}{2})-1} \\
& \times \exp\Big(-\tau_j [\big(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)\big)^\top \boldsymbol{\Lambda}_i^\top \mathbf{S}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i \big(\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)\big) + \boldsymbol{\gamma}_j^\top \mathbf{I}_p \boldsymbol{\gamma}_j + \nu_i]/2 \Big).
\end{aligned}
$$

Let $\mathbf{x} = \boldsymbol{\gamma}_j$, $\mathbf{a} = \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)$, $\mathbf{b} = \mathbf{0}$, $\mathbf{Q}_1 = \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i$ and $\mathbf{Q}_2 = \mathbf{I}_p$. By Lemma B.1, we have

$$
\begin{aligned}
& \big([\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)]^\top \boldsymbol{\Lambda}_i \mathbf{S}_{ij}^{\text{oo}} \boldsymbol{\Lambda}_i [\boldsymbol{\gamma}_j - \boldsymbol{\Lambda}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i)] + \boldsymbol{\gamma}_j^\top \boldsymbol{\gamma}_j \big) \\
& = \Big((\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*)^\top \boldsymbol{\Delta}_{ij}^{*^{-1}} (\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*) + U_{ij}^{\text{o}} \Big).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& f(\mathbf{Y}_j^{\text{o}}, \boldsymbol{\gamma}_j, \tau_j | Z_{ij}=1) \\
= \; & 2^{\frac{p_j^{\text{m}}}{2}} \pi^{-\frac{p+p_j^{\text{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Sigma}_{ij}^{\text{oo}}|^{-\frac{1}{2}} \tau_j^{(\frac{\nu_i+p+p_j^{\text{o}}}{2})-1} \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\text{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big]/2 \Big\} \\
& \times \exp\Big\{ -\tau_j \big(\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*\big)^\top \boldsymbol{\Delta}_{ij}^{*^{-1}} \big(\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*\big)/2 \Big\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& f(\mathbf{Y}_j^{\mathrm{o}}, \tau_j | Z_{ij}=1) \\
= & \int_{\mathbb{R}_p^+} f(\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j | Z_{ij}=1) d\boldsymbol{\gamma}_j \\
= & 2^{\frac{p_j^{\mathrm{m}}}{2}} \pi^{-\frac{p+p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}|^{-\frac{1}{2}} \tau_j^{(\frac{\nu_i+p+p_j^{\mathrm{o}}}{2})-1} \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big] / 2 \Big\} \\
& \times \int_{\mathbb{R}_p^+} \exp\Big\{ -\tau_j \big(\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*\big)^\top \boldsymbol{\Delta}_{ij}^{*^{-1}} \big(\boldsymbol{\gamma}_j - \mathbf{q}_{ij}^*\big)/2 \Big\} d\boldsymbol{\gamma}_j \\
= & 2^{\frac{p_j^{\mathrm{m}}}{2}} \pi^{-\frac{p+p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}|^{-\frac{1}{2}} \tau_j^{(\frac{\nu_i+p+p_j^{\mathrm{o}}}{2})-1} \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big] / 2 \Big\} \\
& \times (2\pi)^{\frac{p}{2}} |\frac{\boldsymbol{\Delta}_{ij}^*}{\tau_j}|^{\frac{1}{2}} \Phi_p \left(\mathbf{q}_{ij}^* | \boldsymbol{\Delta}_{ij}^* / \tau_j \right) \\
= & 2^{p_j^{\mathrm{m}}} (\frac{2}{\pi})^{\frac{p_j^{\mathrm{o}}}{2}} \frac{\frac{\nu_i}{2}^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} |\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}} \boldsymbol{\Delta}_{ij}^{*^{-1}}|^{-\frac{1}{2}} \tau_j^{\frac{\nu_i+p_j^{\mathrm{o}}}{2}-1} \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big] / 2 \Big\} \\
& \times \Phi_p \left(\mathbf{q}_{ij}^* \sqrt{\tau_j} | \boldsymbol{\Delta}_{ij}^* \right).
\end{aligned}
$$

The conditional density of τ_j given $\mathbf{Y}_j^{\mathrm{o}}$ and $Z_{ij}=1$ is as follows

$$
\begin{aligned}
& f(\tau_j \mid \mathbf{Y}_j^{\mathrm{o}}, Z_{ij}=1) \propto f(\mathbf{Y}_j^{\mathrm{o}}, \tau_j | Z_{ij}=1) \\
& \propto \tau_j^{\frac{\nu_i+p_j^{\mathrm{o}}}{2}-1} \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big] / 2 \Big\} \Phi_p \left(\mathbf{q}_{ij}^* \sqrt{\tau_j} | \boldsymbol{\Delta}_{ij}^* \right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
f(\tau_j \mid \mathbf{Y}_j^{\mathrm{o}}, Z_{ij}=1) = & C_1 \tau_j^{\frac{\nu_i+p_j^{\mathrm{o}}}{2}-1} \Phi_p \left(\mathbf{q}_{ij}^* \sqrt{\tau_j} | \boldsymbol{\Delta}_{ij}^* \right) \\
& \times \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big] / 2 \Big\}.
\end{aligned}
$$

Using the property of the probability density function, we have

$$
\int_0^\infty C_1 \tau_j^{\frac{\nu_i+p_j^{\mathrm{o}}}{2}-1} \exp\Big\{ -\tau_j \Big[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \Big] / 2 \Big\} \times \Phi_p \left(\mathbf{q}_{ij}^* \sqrt{\tau_j} | \boldsymbol{\Delta}_{ij}^* \right) d\tau_j = 1.
$$

It implies

$$C_1\Gamma(\frac{\nu_i+p_j^{\rm o}}{2})\left(\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu\Big]\right)^{-\frac{\nu_i+p_j^{\rm o}}{2}}$$
$$\times\int_0^\infty\Gamma\left(\frac{\nu_i+p_j^{\rm o}}{2},\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu\Big]\right)\Phi_p\left(\mathbf{q}_{ij}^*\sqrt{\tau_j}|\boldsymbol{\Delta}_{ij}^*\right)d\tau_j=1.$$

By Lemma F.1

$$C_1\Gamma(\frac{\nu_i+p_j^{\rm o}}{2})\left(\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu\Big]\right)^{-\frac{\nu_i+p_j^{\rm o}}{2}}$$
$$\times T_p\left(\mathbf{q}_{ij}^*\left(\frac{\nu_i+p_j^{\rm o}}{(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu}\right)^{\frac{1}{2}}|\boldsymbol{\Delta}_{ij}^*;\ \nu_i+p_j^{\rm o}\right)=1.$$

Therefore,

$$C_1=\frac{\left(\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i\Big]\right)^{\frac{\nu_i+p_j^{\rm o}}{2}}}{\Gamma(\frac{\nu_i+p_j^{\rm o}}{2})T_p\left(\mathbf{q}_{ij}^*\left(\frac{\nu_i+p_j^{\rm o}}{(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i}\right)^{\frac{1}{2}}|\boldsymbol{\Delta}_{ij}^*;\ \nu_i+p_j^{\rm o}\right)}.$$

By Lemma F.1

$$\begin{aligned}
&\mathrm{E}(\tau_j\mid\mathbf{Y}_j^{\rm o},Z_{ij}=1)\\
=\ &\int_0^\infty C_1\tau_j^{\frac{\nu_i+p_j^{\rm o}+2}{2}-1}\exp\left\{-\tau_j\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i\Big]/2\right\}\\
&\quad\times\Phi_p\left(\mathbf{q}_{ij}^*\sqrt{\tau_j}|\boldsymbol{\Delta}_{ij}^*\right)d\tau_j\\
=\ &C_1\times\Gamma(\frac{\nu_i+p_j^{\rm o}+2}{2})\left(\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i\Big]\right)^{-\frac{\nu_i+p_j^{\rm o}+2}{2}}\\
&\int_0^\infty\Gamma\left(\frac{\nu_i+p_j^{\rm o}+2}{2},\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i\Big]\right)\\
&\times\Phi_p\left(\mathbf{q}_{ij}^*\sqrt{\tau_j}|\boldsymbol{\Delta}_{ij}^*\right)d\tau_j\\
=\ &C_1\times\Gamma(\frac{\nu_i+p_j^{\rm o}+2}{2})\left(\frac{1}{2}\Big[(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i\Big]\right)^{-\frac{\nu_i+p_j^{\rm o}+2}{2}}\\
&\times T_p\left(\mathbf{q}_{ij}^*\left(\frac{\nu_i+p_j^{\rm o}+2}{(\mathbf{Y}_j-\boldsymbol{\xi}_i)^\top\mathbf{C}_{ij}^{\rm oo}(\mathbf{Y}_j-\boldsymbol{\xi}_i)+\nu_i}\right)^{\frac{1}{2}}|\boldsymbol{\Delta}_{ij}^*;\ \nu_i+p_j^{\rm o}+2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\nu_i + p_j^{\mathrm{o}}}{(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i} \\
&\quad \times \frac{T_p\left(\mathbf{q}_{ij}^*\left(\frac{\nu_i + p_j^{\mathrm{o}} + 2}{(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i}\right)^{\frac{1}{2}} |\boldsymbol{\Delta}_{ij}^*;\ \nu_i + p_j^{\mathrm{o}} + 2\right)}{T_p\left(\mathbf{q}_{ij}^*\left(\frac{\nu_i + p_j^{\mathrm{o}}}{(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i}\right)^{\frac{1}{2}} |\boldsymbol{\Delta}_{ij}^*;\ \nu_i + p_j^{\mathrm{o}}\right)} \\
&= \frac{\nu_i + p_j^{\mathrm{o}}}{U_{ij}^{\mathrm{o}} + \nu_i} \times \frac{T_p(\mathbf{q}_{ij}^* \sqrt{\frac{\nu_i + p_j^{\mathrm{o}} + 2}{\nu_i + U_{ij}^{\mathrm{o}}}} |\boldsymbol{\Delta}_{ij}^*; \nu_i + p_j^{\mathrm{o}} + 2)}{T_p(\mathbf{q}_{ij}^* \sqrt{\frac{\nu_i + p_j^{\mathrm{o}}}{\nu_i + U_{ij}^{\mathrm{o}}}} |\boldsymbol{\Delta}_{ij}^*; \nu_i + p_j^{\mathrm{o}})},
\end{aligned}
$$

where $U_{ij}^{\mathrm{o}} = (\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i)$ and $\mathbf{q}_{ij}^* = \boldsymbol{\Lambda}_i \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i)$.

Since $\int_0^\infty f(\tau_j \mid \mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1) d\tau_j = 1$, we have

$$
\begin{aligned}
&\frac{d}{d\nu_i} \int_0^\infty f(\tau_j \mid \mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1) d\tau_j = 0 \\
&\frac{d}{d\nu_i} \int_0^\infty C_1 \tau_j^{\frac{\nu_i + p_j^{\mathrm{o}}}{2} - 1} \exp\left\{ -\tau_j \left[(\mathbf{Y}_j - \boldsymbol{\xi}_i)^\top \mathbf{C}_{ij}^{\mathrm{oo}} (\mathbf{Y}_j - \boldsymbol{\xi}_i) + \nu_i \right] / 2 \right\} \\
&\times \Phi_p \left(\mathbf{q}_{ij}^* \sqrt{\tau_j} |\boldsymbol{\Delta}_{ij}^* \right) d\tau_j = 0,
\end{aligned}
$$

where $C_1 = \left(\frac{U_{ij}^{\mathrm{o}} + \nu_i}{2}\right)^{\frac{\nu_i + p_j^{\mathrm{o}}}{2}} \left(\Gamma(\frac{\nu_i + p_j^{\mathrm{o}}}{2}) T_p \left(\mathbf{q}_{ij}^* \sqrt{\frac{\nu_i + p_j^{\mathrm{o}}}{\nu_i + U_{ij}^{\mathrm{o}}}} |\boldsymbol{\Delta}_{ij}^*;\ \nu_i + p_j^{\mathrm{o}} \right) \right)^{-1}$.

By Leibnitz's rule, we have

$$
\begin{aligned}
&\log \left(\frac{U_{ij}^{\mathrm{o}} + \nu_i}{2} \right) + \left(\frac{\nu_i + p_j^{\mathrm{o}}}{U_{ij}^{\mathrm{o}} + \nu_i} \right) - \mathrm{DG} \left(\frac{\nu_i + p_j^{\mathrm{o}}}{2} \right) + \mathrm{E}(\log \tau_j | \mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1) \\
&- \mathrm{E}(\tau_j | \mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1) - \frac{1}{T_p(\mathbf{q}_{ij}^* \sqrt{\frac{\nu_i + p_j^{\mathrm{o}}}{\nu_i + U_{ij}^{\mathrm{o}}}} |\boldsymbol{\Delta}_{ij}^*;\ \nu_i + p_j^{\mathrm{o}})} \\
&\times \int_{-\infty}^{q_{ij1}^*} \cdots \int_{-\infty}^{q_{ijp}^*} g_{\nu_i}(\mathbf{x}_j) t_p(\mathbf{x}_j | \boldsymbol{\Delta}_{ij}^* \times \frac{U_{ij}^{\mathrm{o}} + \nu_i}{\nu_i + p_j^{\mathrm{o}}}; \nu_i + p_j^{\mathrm{o}}) d\mathbf{x}_j = 0,
\end{aligned}
$$

where $\mathbf{q}_{ij}^* = (q_{ij1}^*, ..., q_{ijp}^*)^\top$ and $g_{\nu_i}(\mathbf{x}_j) = \mathrm{DG}\left(\frac{\nu_i + p + p_j^{\mathrm{o}}}{2}\right) - \mathrm{DG}\left(\frac{\nu_i + p_j^{\mathrm{o}}}{2}\right) - \frac{p}{p_j^{\mathrm{o}} + \nu_i} + \frac{p(U_{ij}^{\mathrm{o}} - p_j^{\mathrm{o}})}{(\nu_i + p_j^{\mathrm{o}})(U_{ij}^{\mathrm{o}} + \nu_i)} - \log\left(1 + \frac{\mathbf{x}_j^\top {\boldsymbol{\Delta}_{ij}^*}^{-1} \mathbf{x}_j}{\nu_i + U_{ij}^{\mathrm{o}}}\right) + \frac{(\nu_i + p + p_j^{\mathrm{o}})(\mathbf{x}_j^\top {\boldsymbol{\Delta}_{ij}^*}^{-1} \mathbf{x}_j)}{(U_{ij}^{\mathrm{o}} + \nu_i)(U_{ij}^{\mathrm{o}} + \nu_i + \mathbf{x}_j^\top {\boldsymbol{\Delta}_{ij}^*}^{-1} \mathbf{x}_j)}$.

Let $\mathbf{A} = t_p(\mathbf{x}_j|\boldsymbol{\Delta}^*_{ij} \times \frac{U^{\circ}_{ij}+\nu_i}{\nu_i+p^{\circ}_j}; \nu_i + p^{\circ}_j)$. Thus,

$$\frac{\partial}{\partial \nu_i}\Big(T_p(\mathbf{q}^*_{ij} \times \Big(\frac{\nu_i + p^{\circ}_j}{U^{\circ}_{ij}+\nu_i}\Big)^{\frac{1}{2}}|\boldsymbol{\Delta}^*_{ij}; \nu_i + p^{\circ}_j)\Big)^{-1} = \frac{\partial}{\partial \nu_i}\Big(T_p(\mathbf{q}^*_{ij}|\boldsymbol{\Delta}^*_{ij} \times \frac{U^{\circ}_{ij}+\nu_i}{\nu_i+p^{\circ}_j}; \nu_i + p^{\circ}_j)\Big)^{-1}$$

$$= \frac{\partial}{\partial \nu_i}\Big(\int_{-\infty}^{q^*_{ij1}} \cdots \int_{-\infty}^{q^*_{ijp}} t_p(\mathbf{x}_j|\boldsymbol{\Delta}^*_{ij} \times \frac{U^{\circ}_{ij}+\nu_i}{\nu_i+p^{\circ}_j}; \nu_i + p^{\circ}_j)d\mathbf{x}_j\Big)^{-1}$$

$$= \frac{\partial}{\partial \nu_i}\Big(\int_{-\infty}^{q^*_{ij1}} \cdots \int_{-\infty}^{q^*_{ijp}} \frac{\Gamma(\frac{\nu_i+p+p^{\circ}_j}{2})|\boldsymbol{\Delta}^*_{ij}|^{-1/2}\Big(\frac{\nu_i+p^{\circ}_j}{U^{\circ}_{ij}+\nu_i}\Big)^{\frac{p}{2}}}{\Gamma(\frac{\nu_i+p^{\circ}_j}{2})(\pi(\nu_i+p^{\circ}_j))^{p/2}} \left[1 + \frac{\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j}{U^{\circ}_{ij}+\nu_i}\right]^{-\frac{\nu_i+p+p^{\circ}_j}{2}} d\mathbf{x}_j\Big)^{-1}$$

$$= -\frac{1}{T_p^2(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\circ}_j}{\nu_i+U^{\circ}_{ij}}}|\boldsymbol{\Delta}^*_{ij};\ \nu_i + p^{\circ}_j)}\Bigg\{\int_{-\infty}^{q^*_{ij1}} \cdots \int_{-\infty}^{q^*_{ijp}} \frac{\mathbf{A}}{2}\mathrm{DG}\left(\frac{\nu_i+p+p^{\circ}_j}{2}\right)$$

$$-\frac{\mathbf{A}}{2}\mathrm{DG}\left(\frac{\nu_i+p^{\circ}_j}{2}\right) - \frac{\mathbf{A}}{2}\frac{p}{p^{\circ}_j+\nu_i} + \frac{\mathbf{A}}{2}\frac{p(U^{\circ}_{ij}-p^{\circ}_j)}{(\nu_i+p^{\circ}_j)(U^{\circ}_{ij}+\nu_i)} - \frac{\mathbf{A}}{2}\log\left[1 + \frac{\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j}{U^{\circ}_{ij}+\nu_i}\right]$$

$$+\frac{\mathbf{A}}{2}\frac{(\nu_i+p+p^{\circ}_j)\left[\frac{\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j}{(\nu_i+U^{\circ}_{ij})^2}\right]}{1+\frac{\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j}{U^{\circ}_{ij}+\nu_i}}d\mathbf{x}_j\Bigg\}$$

$$= -\frac{1}{2T_p^2(\mathbf{q}^*_{ij}\sqrt{\frac{\nu_i+p^{\circ}_j}{\nu_i+U^{\circ}_{ij}}}|\boldsymbol{\Delta}^*_{ij};\ \nu_i + p^{\circ}_j))}\Bigg\{\int_{-\infty}^{q^*_{ij1}} \cdots \int_{-\infty}^{q^*_{ijp}} \Bigg(\mathrm{DG}\left(\frac{\nu_i+p+p^{\circ}_j}{2}\right)$$

$$-\mathrm{DG}\left(\frac{\nu_i+p^{\circ}_j}{2}\right) - \frac{p}{p^{\circ}_j+\nu_i} + \frac{p(U^{\circ}_{ij}-p^{\circ}_j)}{(\nu_i+p^{\circ}_j)(U^{\circ}_{ij}+\nu_i)} - \log\left[1 + \frac{\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j}{U^{\circ}_{ij}+\nu_i}\right]$$

$$+\frac{(\nu_i+p+p^{\circ}_j)(\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j)}{(\nu_i+U^{\circ}_{ij})(U^{\circ}_{ij}+\nu_i+\mathbf{x}_j^{\top}\boldsymbol{\Delta}^{*^{-1}}_{ij}\mathbf{x}_j)}\Bigg)$$

$$\times \mathbf{A}\ \ d\mathbf{x}_j\Bigg\}.$$

Hence,

$$
\begin{aligned}
&E(\log \tau_j|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij}=1)\\
&= \mathrm{DG}\left(\frac{\nu_i+p_j^{\mathrm{o}}}{2}\right) + \left(\frac{\nu_i+p_j^{\mathrm{o}}}{U_{ij}^{\mathrm{o}}+\nu_i}\right)\left(\frac{T_p(\mathbf{q}_{ij}^*\sqrt{\frac{\nu_i+p_j^{\mathrm{o}}+2}{\nu_i+U_{ij}^{\mathrm{o}}}}|\mathbf{\Delta}_{ij}^*;\nu_i+p_j^{\mathrm{o}}+2)}{T_p(\mathbf{q}_{ij}^*\sqrt{\frac{\nu_i+p_j^{\mathrm{o}}}{\nu_i+U_{ij}^{\mathrm{o}}}}|\mathbf{\Delta}_{ij}^*;\nu_i+p_j^{\mathrm{o}})} - 1\right)\\
&+\frac{1}{T_p(\mathbf{q}_{ij}^*\sqrt{\frac{\nu_i+p_j^{\mathrm{o}}}{\nu_i+U_{ij}^{\mathrm{o}}}}|\mathbf{\Delta}_{ij}^*;\nu_i+p_j^{\mathrm{o}})}\int_{-\infty}^{q_{ij1}^*}\cdots\int_{-\infty}^{q_{ijp}^*} g_{\nu_i}(\mathbf{x}_j)t_p(\mathbf{x}_j|\mathbf{\Delta}_{ij}^*\times\frac{U_{ij}^{\mathrm{o}}+\nu_i}{\nu_i+p_j^{\mathrm{o}}};\nu_i+p_j^{\mathrm{o}})d\mathbf{x}_j\\
&-\log\left(\frac{U_{ij}^{\mathrm{o}}+\nu_i}{2}\right).
\end{aligned}
$$

M. Proof of the $\mathbf{\Omega}_{ij}^{(k)}$

Let

$$\hat{Z}_{ij}^{(k)} = \mathrm{E}(Z_{ij}|\mathbf{Y}^{\mathrm{o}}, \hat{\mathbf{\Theta}}^{(k)}) = \frac{\hat{w}_i^{(k)} f_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)}, \hat{\mathbf{\Sigma}}_{ij}^{\mathrm{oo}(k)}, \hat{\mathbf{\Lambda}}_{ij}^{\mathrm{oo}(k)}, \hat{\nu}_i)}{\sum_{i=1}^{g} \hat{w}_i^{(k)} f_{p_j^{\mathrm{o}}}(\mathbf{Y}_j^{\mathrm{o}}|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{o}(k)}, \hat{\mathbf{\Sigma}}_{ij}^{\mathrm{oo}(k)}, \hat{\mathbf{\Lambda}}_{ij}^{\mathrm{oo}(k)}, \hat{\nu}_i)}$$

and

$$\begin{aligned}
\mathbf{\Omega}_{ij}^{(k)} &= \mathrm{E}(Z_{ij}\tau_j(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}^{\mathrm{o}}, \hat{\mathbf{\Theta}}^{(k)}) \\
&= \mathrm{E}(Z_{ij}|\mathbf{Y}^{\mathrm{o}}, \hat{\mathbf{\Theta}}^{(k)})\mathrm{E}(Z_{ij}\tau_j(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}^{\mathrm{o}}, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}) \\
&= \hat{Z}_{ij}^{(k)}\mathrm{E}(\tau_j(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}^{\mathrm{o}}, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}).
\end{aligned}$$

Since $\mathbf{Y}_j = \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{Y}_j^{\mathrm{m}}$ and $\mathbf{O}_j^\top\mathbf{O}_j(\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) = 0$, we have

$$\begin{aligned}
& \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}) \\
=\ & \mathrm{E}(\mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^\top\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi} - \mathbf{\Lambda}\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}) \\
=\ & \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j + \mathbf{M}_j^\top\mathrm{E}(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}) \\
=\ & \mathbf{O}_j^\top\mathbf{Y}_j^{\mathrm{o}} - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j + \mathbf{M}_j^\top\mathbf{M}_j(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\mathbf{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j + \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\mathbf{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j)) \\
=\ & \mathbf{O}_j^\top\mathbf{O}_j\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j + \mathbf{M}_j^\top\mathbf{M}_j(\hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\mathbf{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j)) \\
=\ & \mathbf{O}_j^\top\mathbf{O}_j\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j + (\mathbf{I}_p - \mathbf{O}_j^\top\mathbf{O}_j)\hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j \\
& +(\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\mathbf{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j) \\
=\ & \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j - (\boldsymbol{\xi}_i + \mathbf{\Lambda}_i\boldsymbol{\gamma}_j) + (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\mathbf{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j), \\
& \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}) \\
=\ & \mathrm{E}(\mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \mathbf{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)})|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}) \\
=\ & \mathrm{E}\Big(\hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j - (\boldsymbol{\xi}_i + \mathbf{\Lambda}_i\boldsymbol{\gamma}_j) + (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}) \\
& \times(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\mathbf{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j)|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij}=1, \hat{\mathbf{\Theta}}^{(k)}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\xi}}_i^{(k)} - \boldsymbol{\xi}_i \\
&\quad + \Big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\Big)\,\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
&= \hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i + (\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Cov}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \;& \mathrm{Cov}(\mathbf{O}_j^{\top}\mathbf{Y}_j^{\mathrm{o}} + \mathbf{M}_j^{\top}\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \;& \mathrm{Cov}(\mathbf{M}_j^{\top}\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \;& \mathrm{E}\big(\mathrm{Cov}(\mathbf{M}_j^{\top}\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
& + \mathrm{Cov}\big(\mathrm{E}(\mathbf{M}_j^{\top}\mathbf{Y}_j^{\mathrm{m}} - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \;& \mathrm{E}\big(\mathbf{M}_j^{\top}\mathrm{Cov}(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\mathbf{M}_j\big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
& + \mathrm{Cov}\big(\mathbf{M}_j^{\top}\mathrm{E}(\mathbf{Y}_j^{\mathrm{m}}|\mathbf{Y}_j^{\mathrm{o}}, \boldsymbol{\gamma}_j, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \;& \mathrm{E}\big(\frac{1}{\tau_j}\mathbf{M}_j^{\top}\mathbf{M}_j(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)}\mathbf{M}_j^{\top}\mathbf{M}_j\big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
& + \mathrm{Cov}\big(\mathbf{M}_j^{\top}\mathbf{M}_j(\hat{\boldsymbol{\xi}}_i^{(k)} + \hat{\boldsymbol{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j + \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i^{(k)} - \hat{\boldsymbol{\Lambda}}_i^{(k)}\boldsymbol{\gamma}_j)) \\
& - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\big) \\
= \;& \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} + \mathrm{Cov}\Big(\big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big)\boldsymbol{\gamma}_j\Big|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\Big) \\
= \;& \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} + \big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big)\mathrm{Cov}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
& \times\big((\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)} - \boldsymbol{\Lambda}_i\big)^{\top} \\
= \;& \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} + (\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i)\mathrm{Cov}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i)^{\top},
\end{aligned}
$$

where $\hat{\mathbf{A}}_{ij}^{(k)} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Lambda}}_i^{(k)}$ and $\hat{\mathbf{b}}_{ij}^{(k)} = \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\xi}}_i^{(k)}$.

Thus,

$$\begin{aligned}
& \mathrm{E}((\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\mathrm{E}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})^\top \\
& + \mathrm{Cov}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \left[\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i + \left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\right] \\
& \times \left[\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i + \left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\right]^\top + \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} \\
& + \big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)\mathrm{Cov}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top \\
= \; & \left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)^\top + \left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})^\top\left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)^\top \\
& + \left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)^\top + \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} \\
& + \big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})^\top\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top \\
& + \big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)\mathrm{Cov}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top \\
= \; & \left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)^\top + \left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})^\top\left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)^\top \\
& + \left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)\mathrm{E}(\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)^\top + \frac{1}{\tau_j}(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} \\
& + \big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)\mathrm{E}(\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top.
\end{aligned}$$

Thus,

$$\begin{aligned}
& \mathrm{E}(\tau_j(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top|\mathbf{Y}^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \mathrm{E}(\tau_j\mathrm{E}((\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)(\mathbf{Y}_j - \boldsymbol{\xi} - \boldsymbol{\Lambda}\boldsymbol{\gamma}_j)^\top|\mathbf{Y}_j^{\mathrm{o}}, \tau_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}) \\
= \; & \big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)\mathrm{E}(\tau_j\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\big(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\big)^\top \\
& + \left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)\mathrm{E}(\tau_j\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})^\top\left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)^\top \\
& + \left(\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i\right)\mathrm{E}(\tau_j\boldsymbol{\gamma}_j|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)^\top \\
& + (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}(k)})\hat{\boldsymbol{\Sigma}}_i^{(k)} + \mathrm{E}(\tau_j|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)})\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)\left(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i\right)^\top.
\end{aligned}$$

N. Proof of the CM-steps

Let $\hat{\mathbf{A}}_{ij}^{(k)} = (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\text{oo}(k)})\hat{\mathbf{\Lambda}}_i^{(k)}$ and $\hat{\mathbf{b}}_{ij}^{(k)} = \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\text{oo}(k)}\mathbf{Y}_j + (\mathbf{I}_p - \hat{\mathbf{\Sigma}}_i^{(k)}\hat{\mathbf{S}}_{ij}^{\text{oo}(k)})\hat{\boldsymbol{\xi}}_i^{(k)}$.

CM-step 1:

The mixing probabilities w_i's are subject to the constraint $\sum_{i=1}^g w_i = 1$. Define $L = \sum_{i=1}^g \sum_{j=1}^n \hat{Z}_{ij}^{(k)} \log(w_i) - \lambda(\sum_{i=1}^g w_i - 1)$, where λ is the Lagrange multiplier. Let $dL/dw_i = 0$ and $dL/d\lambda = 0$, we have $w_i = \sum_{j=1}^n \hat{Z}_{ij}^{(k)}/\lambda$ and $\sum_{i=1}^g w_i = 1$. It follows that $\lambda = n$ and the estimate of w_i at the $(k+1)$th iteration is

$$\hat{w}_i^{(k+1)} = \frac{\sum_{j=1}^n \hat{Z}_{ij}^{(k)}}{n}.$$

CM-step 2:

Taking the partial derivative of Q_2 with respect to $\boldsymbol{\xi}_i$ and setting it to zero yields

$$\begin{aligned}
\frac{\partial Q_2}{\partial \boldsymbol{\xi}_i} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\text{tr}\left(\mathbf{\Sigma}_i^{-1}\sum_{j=1}^n \mathbf{\Omega}_{ij}^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\sum_{j=1}^n \text{tr}\left(\mathbf{\Sigma}_i^{-1}\mathbf{\Omega}_{ij}^{(k)}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\sum_{j=1}^n \Big[\text{tr}\big(\mathbf{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}\hat{\tau}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top\big) \\
&\quad +\text{tr}\big(\mathbf{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top\big) \\
&\quad +\text{tr}\big(\mathbf{\Sigma}^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi})\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}(\hat{\mathbf{A}}_j^{(k)} - \mathbf{\Lambda})^\top\big)\Big] \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\xi}_i}\sum_{j=1}^n \Big[\hat{Z}_{ij}^{(k)}\hat{\tau}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top\mathbf{\Sigma}_i^{-1}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i) \\
&\quad +2\hat{Z}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i)^\top\mathbf{\Sigma}^{-1}(\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)}\Big] \\
&= -\frac{1}{2}\sum_{j=1}^n \Big[-2\hat{Z}_{ij}^{(k)}\mathbf{\Sigma}_i^{-1}\hat{\tau}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i) - 2\hat{Z}_{ij}^{(k)}\mathbf{\Sigma}_i^{-1}(\hat{\mathbf{A}}_{ij}^{(k)} - \mathbf{\Lambda}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)}\Big]
\end{aligned}$$

$$= \boldsymbol{\Sigma}_i^{-1} \sum_{j=1}^{n} \left[\hat{Z}_{ij}^{(k)} \hat{\tau}_{ij}^{(k)} (\hat{\mathbf{b}}_{ij}^{(k)} - \boldsymbol{\xi}_i) + \hat{Z}_{ij}^{(k)} (\hat{\mathbf{A}}_{ij}^{(k)} - \boldsymbol{\Lambda}_i) \hat{\boldsymbol{\eta}}_{ij}^{(k)} \right]$$

$$= 0.$$

With $\boldsymbol{\Lambda}_i$ fixed at $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, solving this equation gives

$$\hat{\boldsymbol{\xi}}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \hat{\tau}_{ij}^{(k)} \hat{\mathbf{b}}_{ij}^{(k)} - \sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\mathbf{S}}_{ij}^{\text{oo}(k)} \hat{\boldsymbol{\Lambda}}_i^{(k)} \hat{\boldsymbol{\eta}}_{ij}^{(k)}}{\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \hat{\tau}_{ij}^{(k)}}.$$

CM-step 3:

Taking the partial derivative of Q_2 with respect to $\boldsymbol{\Sigma}_i^{-1}$ and setting it to zero yields

$$\begin{aligned}
\frac{\partial Q_2}{\partial \boldsymbol{\Sigma}_i^{-1}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}_i^{-1}} \left[\frac{1}{2} \log |\boldsymbol{\Sigma}_i^{-1}| \left(\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \right) - \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \sum_{j=1}^{n} \boldsymbol{\Omega}_{ij}^{(k)} \right) \right] \\
&= \left[\frac{1}{2} (2\boldsymbol{\Sigma}_i - \text{Diag}(\boldsymbol{\Sigma}_i)) \left(\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)} \right) - \frac{1}{2} \left(2 \sum_{j=1}^{n} \boldsymbol{\Omega}_{ij}^{(k)} - \text{Diag} \left(\sum_{j=1}^{n} \boldsymbol{\Omega}_{ij}^{(k)} \right) \right) \right] \\
&= 0.
\end{aligned}$$

Fix $\boldsymbol{\xi}_i$ at $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i$ at $\hat{\boldsymbol{\Lambda}}_i^{(k)}$, solving this equation obtains

$$\hat{\boldsymbol{\Sigma}}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}}{\sum_{j=1}^{n} \hat{Z}_{ij}^{(k)}},$$

where $\hat{\boldsymbol{\Omega}}_{ij}^{(k+1/2)}$ is $\boldsymbol{\Omega}_{ij}^{(k)}$ in Eq. (3.6) with $\boldsymbol{\xi}_i$ replaced by $\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Lambda}_i$ replaced by $\hat{\boldsymbol{\Lambda}}_i^{(k)}$.

CM-step 4:

Here $\boldsymbol{\Lambda}_i$ assumed to be diagonal, say $\boldsymbol{\Lambda}_i = \text{Diag}(\boldsymbol{\lambda}_i)$, where $\boldsymbol{\lambda}_i$ is a p-dimensional vector. Taking the partial derivative of Q_2 with respect to $\boldsymbol{\lambda}_i$ and setting it to zero yields

$$
\begin{aligned}
&\frac{\partial Q_2}{\partial \boldsymbol{\lambda}_i} \\
= \;& -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\boldsymbol{\Omega}_{ij}^{(k)}\right) \\
= \;& -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\sum_{j=1}^{n}\mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Omega}_{ij}^{(k)}\right) \\
= \;& -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\sum_{j=1}^{n}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)\hat{\boldsymbol{\Psi}}_{ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)^\top\big) \\
& +\mathrm{tr}\big(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)}-\boldsymbol{\xi}_i)^\top\big)+\mathrm{tr}\big(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)^\top\big)\Big] \\
= \;& -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\lambda}_i}\sum_{j=1}^{n}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)\hat{\mathbf{s}}_{4ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)^\top\big) \\
& +2\mathrm{tr}\big(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)^\top\big)\Big] \\
= \;& \sum_{j=1}^{n}\Big[\mathrm{Diag}\left(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{A}}_{ij}^{(k)}-\boldsymbol{\Lambda}_i)\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)+\mathrm{Diag}\left(\boldsymbol{\Sigma}_i^{-1}\hat{Z}_{ij}^{(k)}(\hat{\mathbf{b}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}\right)\Big] \\
= \;& \mathrm{Diag}\left(\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\left((\hat{\mathbf{b}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}+\hat{\mathbf{A}}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)\right) \\
& -\mathrm{Diag}\left(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right) \\
= \;& \left(\boldsymbol{\Sigma}_i^{-1}\odot\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\left((\hat{\mathbf{b}}_{ij}^{(k)}-\boldsymbol{\xi}_i)\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}+\hat{\mathbf{A}}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^\top\right)\mathbf{1}_p \\
& -\left(\boldsymbol{\Sigma}_i^{-1}\odot\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)\boldsymbol{\lambda}_i.
\end{aligned}
$$

With $\boldsymbol{\xi}_i=\hat{\boldsymbol{\xi}}_i^{(k+1)}$ and $\boldsymbol{\Sigma}_i=\hat{\boldsymbol{\Sigma}}_i^{(k+1)}$, setting $\partial Q_2/\partial\boldsymbol{\lambda}_i=\mathbf{0}$ yields the estimate of $\boldsymbol{\lambda}_i$

$$
\begin{aligned}
\hat{\boldsymbol{\lambda}}_i^{(k+1)} = \;& \left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}}\odot\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^{-1} \\
& \times\left(\hat{\boldsymbol{\Sigma}}_i^{(k+1)^{-1}}\odot\sum_{j=1}^{n}\hat{Z}_{ij}^{(k)}\left((\hat{\mathbf{b}}_{ij}^{(k)}-\hat{\boldsymbol{\xi}}_i^{(k+1)})\hat{\boldsymbol{\eta}}_{ij}^{(k)\top}+\hat{\mathbf{A}}_{ij}^{(k)}\hat{\boldsymbol{\Psi}}_{ij}^{(k)}\right)^\top\right)\mathbf{1}_p.
\end{aligned}
$$

O. Proof of $\hat{\mathbf{s}}_j$ for the MSTMIX model

The single observation of the complete data log-likelihood, ignoring additive constant term, is

$$\begin{aligned}
&\ell_c(\boldsymbol{\Theta}|\mathbf{Y}_j, \mathbf{Z}_j, \boldsymbol{\gamma}_j, \tau_j) \\
= &\sum_{i=1}^{g} Z_{ij}\Big\{ \log(w_i) + \frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right) - \log\Gamma\Big(\frac{\nu_i}{2}\Big) - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \frac{\nu_i}{2}\log\tau_j \\
&-\frac{\tau_j}{2}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j) - \frac{\tau_j\nu_i}{2}\Big\}.
\end{aligned}$$

Let $\hat{\mathbf{D}}_{ij} = (\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}_{ij}^{\mathrm{oo}})$. The first derivatives of $\ell_c(\boldsymbol{\Theta}|\mathbf{Y}_j, \mathbf{Z}_j, \boldsymbol{\gamma}_j, \tau_j)$ with respect to w_i is

$$\frac{\partial\ell_{cj}}{\partial w_i} = \frac{Z_{ij}}{w_i} - \frac{Z_{gj}}{w_g}.$$

Thus,

$$\hat{u}^{\mathrm{o}}_{j,w_i} = \mathrm{E}\Big(\frac{Z_{ij}}{w_i} - \frac{Z_{gj}}{w_g}\Big|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}\Big) = \frac{\hat{Z}_{ij}}{\hat{w}_i} - \frac{\hat{Z}_{gj}}{\hat{w}_g}.$$

Furthermore, we can obtained

$$\mathrm{E}(\tau_j\mathbf{Y}_j|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij}=1, \hat{\boldsymbol{\Theta}}) = \hat{\tau}_{ij}\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij} + \hat{\tau}_{ij}\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}_{ij}^{\mathrm{oo}}\mathbf{Y}_j$$

from the law of iterative expectations. The first derivatives of $\ell_c(\boldsymbol{\Theta}|\mathbf{Y}_j, \mathbf{Z}_j, \boldsymbol{\gamma}_j, \tau_j)$ with respect to $\boldsymbol{\xi}_i$ is

$$\frac{\partial\ell_{cj}}{\partial\boldsymbol{\xi}_i} = Z_{ij}\tau_j\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j).$$

Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\mathrm{o}}_{j,\boldsymbol{\xi}_i} &= \mathrm{E}\Big(Z_{ij}\tau_j\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\Big|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}\Big) \\
&= \hat{Z}_{ij}\mathrm{E}\Big(\tau_j\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\Big|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij}=1, \hat{\boldsymbol{\Theta}}\Big) \\
&= \hat{Z}_{ij}\left(\hat{\tau}_{ij}\hat{\mathbf{S}}_{ij}^{\mathrm{oo}}(\mathbf{Y}_j - \hat{\boldsymbol{\xi}}_i) - \hat{\mathbf{S}}_{ij}^{\mathrm{oo}}\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\eta}}_{ij}\right).
\end{aligned}$$

The first derivatives of $\ell_c(\boldsymbol{\Theta}|\mathbf{Y}_j, \mathbf{Z}_j, \boldsymbol{\gamma}_j, \tau_j)$ with respect to $\boldsymbol{\Sigma}_i$ is

$$\begin{aligned}
\frac{\partial \ell_{cj}}{\partial \boldsymbol{\Sigma}_i} &= -\frac{Z_{ij}}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\Big\{ \log|\boldsymbol{\Sigma}_i| + \mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\tau_j\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top\right)\Big\} \\
&= -\frac{Z_{ij}}{2}\Big\{2\boldsymbol{\Sigma}_i^{-1} - \mathrm{Diag}(\boldsymbol{\Sigma}_i^{-1}) - 2\boldsymbol{\Sigma}_i^{-1}\tau_j\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}_i^{-1} \\
&\quad + \mathrm{Diag}\left(\boldsymbol{\Sigma}_i^{-1}\tau_j\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top\boldsymbol{\Sigma}_i^{-1}\right)\Big\} \\
&= \frac{1}{2}\left(2\mathbf{C}_{ij} - \mathrm{Diag}(\mathbf{C}_{ij})\right),
\end{aligned}$$

where $\mathbf{C}_{ij} = Z_{ij}\left(\boldsymbol{\Sigma}_i^{-1}\mathbf{R}_{ij}\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}\right)$ and $\mathbf{R}_{ij} = \tau_j\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)\left(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j\right)^\top$. Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\mathrm{o}}_{j,\boldsymbol{\sigma}_i} &= \mathrm{vech}\left\{\mathrm{E}\Big(\frac{1}{2}\left(2\mathbf{C}_{ij} - \mathrm{Diag}(\mathbf{C}_{ij})\right)\Big|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}\Big)\right\} \\
&= \mathrm{vech}\left\{\frac{1}{2}\mathrm{E}\Big(2\mathbf{C}_{ij} - \mathrm{Diag}(\mathbf{C}_{ij})\Big|\mathbf{Y}_j^{\mathrm{o}}, \hat{\boldsymbol{\Theta}}\Big)\right\} \\
&= \mathrm{vech}\left\{\frac{1}{2}\Big(2\hat{\mathbf{C}}_{ij} - \mathrm{Diag}(\hat{\mathbf{C}}_{ij})\Big)\right\},
\end{aligned}$$

where $\hat{\mathbf{C}}_{ij} = \hat{Z}_{ij}\left(\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\mathbf{R}}_{ij}\hat{\boldsymbol{\Sigma}}_i^{-1} - \hat{\boldsymbol{\Sigma}}_i^{-1}\right)$ with

$$\begin{aligned}
\hat{\mathbf{R}}_{ij} &= \mathrm{E}\Big(\mathbf{R}_{ij}\Big|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}\Big) \\
&= \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Sigma}}_i + (\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)\hat{\boldsymbol{\Psi}}_{ij}(\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)^\top \\
&\quad + (\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)\hat{\boldsymbol{\eta}}_{ij}(\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)^\top \\
&\quad + (\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)\hat{\boldsymbol{\eta}}_{ij}^\top(\hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{\Lambda}}_i)^\top \\
&\quad + \hat{\tau}_{ij}(\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)(\hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\mathbf{Y}_j + \hat{\mathbf{D}}_{ij}\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i)^\top.
\end{aligned}$$

Furthermore, we can obtained

$$\mathrm{E}(\tau_j\mathbf{Y}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\mathrm{o}}, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}) = \hat{\mathbf{D}}_{ij}(\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\eta}}_{ij}^\top + \hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij}) + \hat{\boldsymbol{\Sigma}}_i\hat{\mathbf{S}}^{\mathrm{oo}}_{ij}\mathbf{Y}_j\hat{\boldsymbol{\eta}}_{ij}^\top,$$

from the law of iterative expectations. The first derivatives of $\ell_c(\boldsymbol{\Theta}|\mathbf{Y}_j, \mathbf{Z}_j, \boldsymbol{\gamma}_j, \tau_j)$ with respect to $\boldsymbol{\lambda}_i$ is

$$\frac{\partial \ell_{cj}}{\partial \boldsymbol{\lambda}_i} = \mathrm{Diag}\left(Z_{ij}\tau_j\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j - \boldsymbol{\xi}_i - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\right).$$

Thus,

$$\begin{aligned}
\hat{\mathbf{u}}^{\mathrm{o}}_{j,\boldsymbol{\lambda}_i} &= \mathrm{E}\Big(\mathrm{Diag}\left(Z_{ij}\tau_j\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}_i-\boldsymbol{\Lambda}_i\gamma_j)\boldsymbol{\gamma}_j^\top\right)\Big|\mathbf{Y}_j^{\mathrm{o}},\hat{\boldsymbol{\Theta}}\Big)\\
&= \mathrm{Diag}\left(\mathrm{E}\Big(Z_{ij}\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\Big|\mathbf{Y}_j^{\mathrm{o}},\hat{\boldsymbol{\Theta}}\Big)\right)\\
&= \mathrm{Diag}\left(\hat{Z}_{ij}\mathrm{E}\Big(\tau_j\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j-\boldsymbol{\xi}-\boldsymbol{\Lambda}\boldsymbol{\gamma}_j)\boldsymbol{\gamma}_j^\top\Big|\mathbf{Y}_j^{\mathrm{o}},Z_{ij}=1,\hat{\boldsymbol{\Theta}}\Big)\right)\\
&= \mathrm{Diag}\left[\hat{Z}_{ij}\left(\hat{\boldsymbol{\Sigma}}_i^{-1}E(\tau_j\mathbf{Y}_j\boldsymbol{\gamma}_j^\top|\mathbf{Y}_j^{\mathrm{o}},Z_{ij}=1,\hat{\boldsymbol{\Theta}})-\hat{\boldsymbol{\Sigma}}_i^{-1}(\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\eta}}_{ij}^\top+\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij})\right)\right]\\
&= \mathrm{Diag}\left[\hat{Z}_{ij}\left(\hat{\mathbf{S}}_{ij}^{\mathrm{oo}}\left((\mathbf{Y}_j-\hat{\boldsymbol{\xi}}_i)\hat{\boldsymbol{\eta}}_{ij}^\top-\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Psi}}_{ij}\right)\right)\right].
\end{aligned}$$

The first derivatives of $\ell_c(\boldsymbol{\Theta}|\mathbf{Y}_j,\mathbf{Z}_j,\boldsymbol{\gamma}_j,\tau_j)$ with respect to ν_i is

$$\frac{\partial\ell_{cj}}{\partial\nu_i} = \frac{Z_{ij}}{2}\left\{\log\left(\frac{\nu_i}{2}\right)+1-\mathrm{DG}\left(\frac{\nu_i}{2}\right)+\log(\tau_j)-\tau_j\right\}.$$

Thus,

$$\begin{aligned}
\hat{u}^{\mathrm{o}}_{j,\nu_i} &= \mathrm{E}\Big(\frac{Z_{ij}}{2}\left\{\log\left(\frac{\nu_i}{2}\right)+1-\mathrm{DG}\left(\frac{\nu_i}{2}\right)+\log(\tau_j)-\tau_j\right\}\Big|\mathbf{Y}_j^{\mathrm{o}},\hat{\boldsymbol{\Theta}}\Big)\\
&= \frac{\hat{Z}_{ij}}{2}\left[\log\left(\frac{\hat{\nu}_i}{2}\right)+1-\mathrm{DG}\left(\frac{\hat{\nu}_i}{2}\right)+\hat{\kappa}_{ij}-\hat{\tau}_{ij}\right].
\end{aligned}$$

If the degrees of freedom are assumed to be equal, i.e., $\nu_1=\cdots=\nu_g=\nu$, then

$$\hat{u}^{\mathrm{o}}_{j,\nu} = \sum_{i=1}^{g}\frac{\hat{Z}_{ij}}{2}\left[\log\left(\frac{\hat{\nu}_i}{2}\right)+1-\mathrm{DG}\left(\frac{\hat{\nu}_i}{2}\right)+\hat{\kappa}_{ij}-\hat{\tau}_{ij}\right]=\sum_{i=1}^{g}\hat{u}^{\mathrm{o}}_{j,\nu_i}.$$

References

Anderson, T.W. 1957. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. Journal of the American Statistical Association 52: 200-203.

Anderson, T.W. 2003. An Introduction to Multivariate Statistical Analysis, third ed, John Wiley & Sons.

Azzalini, A. 2005. The skew-normal distribution and related multivariate families (with discussion). Scandinavian Journal of Statistics 32: 159-200.

Arellano-Valle, R.B. and Genton, M.G. 2005. On fundamental skew distributions. Journal of Multivariate Analysis 96: 93-116.

Arellano-Valle, R.B. and Azzalini, A. 2006. On the unification of families of skew-normal distributions. Scandinavian Journal of Statistics. 33: 561-574.

Azzalini, A. and Capitanio, A. 1999. Statistical applications of the multivariate skew-normal distribution, Journal of the Royal Statistical Society, Series B 65: 579-602.

Azzalini, A. and Capitanio, A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. Journal of the Royal Statistical Society, Series B 65: 367-389.

Azzalini, A. and Dalla Valle, A. 1996. The multivariate skew-normal distribution. Biometrika 83: 715-726.

Azzalini, A. and Genton, M.G. 2008. Robust Likelihood Methods Based on the Skew-t and Related Distributions. International Statistical Review 76: 106-129.

Basford, K.E. and McLachlan, G.J. 1985. Estimation of allocation rates in a cluster analysis text. Journal of the American Statistical Association 80: 286-293.

Bensmail, H., Celeux, G., Raftery, A.E, and Robert, C.P 1997. Inference in model-based cluster analysis. Statistics and Computing 7: 1V10.

Box, G.E.P. and Cox, D.R. 1964. An analysis of transformation. Journal of the Royal Statistical Society, Series A 26: 211-252.

Brooks, S.P. and Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7: 434-455.

Brooks, S.P., Giudici, P., and Robert, G.O. 2003. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions (with discussion). Journal of the Royal Statistical Society, Series B 65: 3-55.

Chan, K.S. and Ledolter, J. 1995. Monte Carlo EM estimation for time series models involving counts. Journal of the American Statistical Association 90: 242-252.

Chib, S. and Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49: 327-335.

Cook, R.D. and Weisberg, S. 1994. An Introduction to Regression Graphics. Wiley, New York.

Dellaportas, P. and Papageorgiou, I. 2006. Multivariate mixtures of normals with unknown number of components. Statistics and Computing 16: 57-68.

Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Series B 39: 1-38.

Diebolt, J. and Robert, C.P. 1994. Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society, Series B 56: 363-375.

Edwards, W.H., Lindman H.,and Savage, L.J. 1963. Bayesian statistical inference for psychological research. Psychological Review 70: 193-242.

Efron B. and Hinkley D.V. 1978. Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. Biometrika 65: 457-487.

Escobar, M.D. and West, M. 1995. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association. 90: 577-88.

Frühwirth-Schnatter S. 2006. Finite Mixture and Markov Switching Models. Springer, New York.

Gelfand, A.E. and Smith, A.F.M. 1990. Sampling based approaches to calculate marginal densities. Journal of the American Statistical Association 85: 398V409.

Ghahramani, Z. and Jordan, M.I. 1994. Supervised learning from incomplete data via an EM approach, in: Cowan, J.D., Tesarro, G., and Alspector, J.(Eds.),Advances in Neural Information Proccessing Systems, Vol. 6, Morgan Kaufmann, San Francisco, 1944, pp. 120-127.

Gneiting, T., Larson, K., Westrick, K., Genton, M.G., and Aldrich, E. 2006. Calibrated probabilistic forecasting at the Stateline wind energy center: the regime-switching space-time method. Journal of the American Statistical Association 101: 968-979.

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82: 711-732.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57: 97-109.

Hocking R.R. and Smith W.B. 1968. Estimation of parameters in the multivariate normal distribution with missing observations. Journal of the American Statistical Association 63: 159-173.

Jones, M.C. and Faddy, M.J. 2003. A skew extension of the t-distribution, with applications. Journal of the Royal Statistical Society, Series B 65: 159-174.

Kotz, S and Nadarajah, S. 2004. Multivariate t Distributions and Their Applications. Cambridge University Press.

Lange, K.L., Little R.J.A. and Taylor, J.M.G. 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84: 881-896.

Lin, T.I. 2009. Maximum likelihood estimation for multivariate skew normal mixture models, Journal of Multivariate Analysis 100: 257-265.

Lin, T.I., Lee, J.C. and Hsieh, W.J. 2007a. Robust mixture modeling using the skew t distribution. Statistics and Computing 17: 81-92.

Lin, T.I., Lee, J.C., and Ho, H.J. 2006. On fast supervised learning for normal mixture models with missing information. Pattern Recognition 39: 1177-1187.

Lin, T.I., Lee, J.C., and Ni, H.F. 2004. Bayesian Analysis of mixture modelling using the multivariate t distribution. Statistics and Computing. 14: 119-130.

Lin, T.I., Lee, J.C., and Yen, S.Y. 2007b. Finite mixture modelling using the skew normal distribution. Statistica Sinica 17: 909-927.

Lin, T.I., Ho, H.J., and Shen, P.S. 2008. Computationally efficient learning of multivariate t mixture models with missing information. Computational Statistics doi: 10.1007/s00180-008-0129-5 (In press).

Little, R.J.A. and Rubin, D.B. 2002. Statistical Analysis with Missing Data. 2rd ed. Wiley, New York.

Liu, C. 1995. Missing data imputation using the multivariate t distribution. Journal of Multivariate Analysis 53: 139-158.

Liu, C. 1999. Efficient ML estimation of the multivariate normal distribution from incomplete data. Journal of Multivariate Analysis 69: 206-217.

Liu, C.H., Rubin, D.B. and Wu, Y.N. 1998. Parameter Expansion to Accelerate EM: the PX-EM Algorithm. Biometrika 85: 755-770.

McLachlan, G.J. and Basford K.E. 1988. Mixture Models: Inference and Application to Clustering. Marcel Dekker, New York.

McLachlan, G.J. and Peel D. 2000. Finite Mixture Models. Wiely, New York.

Meilijson I. 1989. A fast improvement to the EM algorithm to its own terms. Journal of the Royal Statistical Society, Series B 51: 127-138.

Meng, X.L. and Rubin, D.B. 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80: 267-78.

Nadarajah, S. and Kotz, S. 2007. Programs in R for computing truncated t distributions. Quality and Reliability Engineering International 23: 273-278.

Pearson, K. 1894. Contributions to the theory of mathematical evolution, Phi. Trans. Roy. Soc. London A 185: 71-110.

Peel, D. and McLachlan, G.J. 2000. Robust Mixture modeling using the t distribution. Statistics and Computing 10: 339-348.

Richardson S. and Green P.J. 1997. On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society, Series B 59: 731-792.

Rubin, D.B. 1976. Inference and missing data, Biometrika 63: 581-592.

Rubin, D.B. 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley and Sons, New York.

Sahu, S.K., Dey, D.K. and Branco, M.D. 2003. A new class of multivariate skew distributions with application to Bayesian regression models. Canadian Journal of Statistics 31: 129-150.

SAS (2001), SAS/Stat User's Guide, Version 8.2, Cary, NC: SAS Institute Inc.

Schafer, J.L. 1997. Analysis of Incomplete Multivariate Data. Chapman and Hall, London.

Schimert, J., Schafer, J., Hesterberg, T., Fraley, C., and Clarkson, D. 2000. Analyzing Data with Missing Values in S-Plus, Seattle, WA: Data Analysis Products Division, Insightful Corporation.

Seidel, W., Mosler, K., and Alker, M. 2000. A cautionary note on likelihood ratio tests in mixture models. Annals of the Institute of Statistical Mathematics 52: 481-487.

Shoham, S. 2002. Robust clustering by deterministic agglomeration EM of mixtures of multivariate t-distributions. Pattern Recognition 35: 1127-1142.

Shoham, S., Fellows, M.R., and Normann, R.A. 2003. Robust, automatic spike sorting using mixtures of multivariate t-distributions. Journal of Neuroscience Methods 127: 111-122.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Linde, A.V.D. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64: 583-639.

Tanner, M.A. and Wong, W.H. 1987. The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association 82: 528-550.

Titterington D.M., Smith A.F.M. and Markov U.E. 1985. Statistical Analysis of Finite Mixture Distributions. Wiely, New York.

Wang, H.X., Zhang, Q.B., Luo, B., and Wei, S. 2004. Robust mixture modeling using multivariate t distribution with missing information. Pattern Recognition Letter 25: 701-710.

Wei, G.C.G. and Tanner, M.A. 1990. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. Journal of the American Statistical Association 85: 699-704.

Zacks S. 1971. The Theory of Statistical Inference. Wiley, New York.

Zhang, Z.H., Chan, K.L., Wu, Y.M., and Chen, C.B. 2004. Learning multivariate gaussian mixture model with the reversible jump MCMC algorithm. Statistics and Computing 14: 343-355.