Seemingly Unrelated Regression with Measurement Error: Estimation via Markov chain Monte Carlo and Mean Field Variational Bayes Approximation

Supplementary material

Georges Bresson
Department of Economics
Université Paris II, France

Anoop Chaturvedi
Department of Statistics
University of Allahabad, India

Mohammad Arshad Rahman
Department of Economic Sciences
Indian Institute of Technology, Kanpur, India

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur, India

June 12, 2020

Contents

1 Appendix A - Gibbs sampling for SURME model.

The joint posterior distribution of the SURME model is given by,

$$\begin{aligned}
& p\left(\beta, \gamma, \Sigma_{\varepsilon}, Z, \omega, \sigma_Z^2, \sigma_u^2 | W, y\right) \\
& \propto \prod_{i=1}^{N} \left\{ p\left(y_i | \beta, \gamma, \Sigma_{\varepsilon}, X_i, Z_i\right) \times p\left(\widetilde{Z}_i | X_i, \omega, \sigma_Z^2\right) \times p\left(\widetilde{W}_i | \widetilde{Z}_i, \sigma_u^2\right) \right\} \\
& \qquad \times N_K\left(\beta_0, B_0\right) N_M\left(\gamma_0, G_0\right) IW_M\left(\nu_0, S_0\right) N_K\left(\omega_0, O_0\right) IG\left(\delta_1, \delta_2\right) IG\left(\delta_3, \delta_4\right) \\
& \propto \prod_{i=1}^{N} \Bigg\{ (2\pi)^{-M/2} \left|\Sigma_{\varepsilon}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(y_i - X_i\beta - Z_i\gamma\right)' \Sigma_{\varepsilon}^{-1} (y_i - X_i\beta - Z_i\gamma)\right] \qquad \text{(A.1)} \\
& \qquad \times (2\pi)^{-M/2} \left(\sigma_Z^2\right)^{-M/2} \exp\left[-\frac{1}{2\sigma_Z^2}\left(\widetilde{Z}_i - X_i\omega\right)' \left(\widetilde{Z}_i - X_i\omega\right)\right] \\
& \qquad \times (2\pi)^{-M/2} \left(\sigma_u^2\right)^{-M/2} \exp\left[-\frac{1}{2\sigma_u^2}\left(\widetilde{W}_i - \widetilde{Z}_i\right)' \left(\widetilde{W}_i - \widetilde{Z}_i\right)\right] \Bigg\} \\
& \qquad \times N_K\left(\beta_0, B_0\right) N_M\left(\gamma_0, G_0\right) W_M\left(\nu_0, S_0\right) N_K\left(\omega_0, O_0\right) IG\left(\delta_1, \delta_2\right) IG\left(\delta_3, \delta_4\right).
\end{aligned}$$

The conditional posteriors for $(\beta, \gamma, \Sigma_{\varepsilon}^{-1}, Z, \omega, \sigma_Z^2, \sigma_u^2)$ in the SUR model with measurement error can obtained from the joint posterior density given by equation (A.1). The derivations presented below follow the ordering as presented in Algorithm 1.

1.1 Conditional posterior density for $\beta \sim N_K(\overline{\beta}, B_1)$

The conditional posterior density for β given by $p(\beta|\gamma, \Sigma_{\varepsilon}, Z, W, y)$ is proportional to $p(\beta) \times p(y, W, Z|X, \Delta)$, where $\Delta = (\beta, \gamma, \Sigma_{\varepsilon}, \omega, \sigma_Z^2, \sigma_u^2)$, and its kernel can be written as,

$$\begin{aligned}
& p(\beta|\gamma, \Sigma_{\varepsilon}, Z, y) \\
& \propto \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{N} \left(y_i - X_i\beta - Z_i\gamma\right)' \Sigma_{\varepsilon}^{-1} \left(y_i - X_i\beta - Z_i\gamma\right) + \left(\beta - \beta_0\right)' B_0^{-1} \left(\beta - \beta_0\right)\right\}\right] \\
& \propto \exp\Bigg[-\frac{1}{2}\Bigg\{\sum_{i=1}^{N} y_i'\Sigma_{\varepsilon}^{-1}X_i\beta - \sum_{i=1}^{N} \beta' X_i'\Sigma_{\varepsilon}^{-1}y_i + \sum_{i=1}^{N} \beta' X_i'\Sigma_{\varepsilon}^{-1}X_i\beta + \sum_{i=1}^{N} \beta' X_i'\Sigma_{\varepsilon}^{-1}Z_i\gamma \\
& \qquad + \sum_{i=1}^{N} \gamma' Z_i'\Sigma_{\varepsilon}^{-1}X_i\beta + \beta' B_0^{-1}\beta - \beta_0' B_0^{-1}\beta - \beta' B_0^{-1}\beta_0\Bigg\}\Bigg] \\
& \propto \exp\Bigg[-\frac{1}{2}\Bigg\{\beta'\Bigg(\sum_{i=1}^{N} X_i'\Sigma_{\varepsilon}^{-1}X_i + B_0^{-1}\Bigg)\beta - \Bigg(\sum_{i=1}^{N} \left(y_i'\Sigma_{\varepsilon}^{-1}X_i - \gamma' Z_i'\Sigma_{\varepsilon}^{-1}X_i\right) \\
& \qquad + \beta_0' B_0^{-1}\Bigg)\beta - \beta'\Bigg(\sum_{i=1}^{N} \left(\beta' X_i'\Sigma_{\varepsilon}^{-1}y_i - X_i'\Sigma_{\varepsilon}^{-1}Z_i\gamma\right) + B_0^{-1}\beta_0\Bigg)\Bigg\}\Bigg] \\
& \propto \exp\left[-\frac{1}{2}\left\{\beta' B_1^{-1}\beta - \overline{\beta}' B_1^{-1}\beta - \beta' B_1^{-1}\overline{\beta}\right\}\right]
\end{aligned}$$

where the second proportionality opens the square and drops all terms not involving β, the third proportionality rearranges the terms and the last proportionality introduces two terms, B_1 and $\overline{\beta}$, which are defined as follows,

$$B_1^{-1} = \left[\sum_{i=1}^{N} X_i'\Sigma_{\varepsilon}^{-1}X_i + B_0^{-1}\right] \quad \text{and} \quad \overline{\beta} = B_1\left[\sum_{i=1}^{N} X_i'\Sigma_{\varepsilon}^{-1}\left(y_i - Z_i\gamma\right) + B_0^{-1}\beta_0\right].$$

Adding and substracting $\overline{\beta}' B_1^{-1} \overline{\beta}$ inside the curly braces, the square can be completed as,

$$
\begin{aligned}
& p(\beta | \gamma, \Sigma_\varepsilon, Z, y) \\
& \propto \exp \left[-\frac{1}{2} \left\{ \beta' B_1^{-1} \beta - \overline{\beta}' B_1^{-1} \beta - \beta' B_1^{-1} \overline{\beta} - \overline{\beta}' B_1^{-1} \overline{\beta} + \overline{\beta}' B_1^{-1} \overline{\beta} \right\} \right] \\
& \propto \exp \left[-\frac{1}{2} (\beta - \overline{\beta})' B_1^{-1} (\beta - \overline{\beta}) \right],
\end{aligned}
$$

where the last line follows by recognizing that $\overline{\beta}' B_1^{-1} \overline{\beta}$ does not involve β and can therefore be absorbed in the constant of proportionality. The result is the kernel of a Gaussian or normal density and hence, $\beta | \gamma, \Sigma_\varepsilon, Z, y \sim N_K \left(\overline{\beta}, B_1 \right)$.

1.2 Conditional posterior density for $\gamma \sim N_M(\overline{\gamma}, G_1)$

The conditional posterior density for γ can be derived in an identical manner as that of β. So, to obtain the conditional posterior $p(\gamma | \beta, \Sigma_\varepsilon, Z, y)$, we focus on terms involving γ in the joint posterior density (A.1). Hence, we have,

$$
\begin{aligned}
& p(\gamma | \beta, \Sigma_\varepsilon, Z, y) \\
& \propto \exp \Bigg[-\frac{1}{2} \Bigg\{ \sum_{i=1}^{N} (y_i - X_i \beta - Z_i \gamma)' \Sigma_\varepsilon^{-1} (y_i - X_i \beta - Z_i \gamma) + (\gamma - \gamma_0)' G_0^{-1} (\gamma - \gamma_0) \Bigg\} \Bigg] \\
& \propto \exp \Bigg[-\frac{1}{2} \Bigg\{ -\sum_{i=1}^{N} y_i' \Sigma_\varepsilon^{-1} Z_i \gamma + \sum_{i=1}^{N} \beta' X_i' \Sigma_\varepsilon^{-1} Z_i \gamma - \sum_{i=1}^{N} \gamma' Z_i' \Sigma_\varepsilon^{-1} y_i + \sum_{i=1}^{N} \gamma' Z_i' \Sigma_\varepsilon^{-1} X_i \beta \\
& \quad + \sum_{i=1}^{N} \gamma' Z_i' \Sigma_\varepsilon^{-1} Z_i \gamma + \gamma' G_0^{-1} \gamma - \gamma_0' G_0^{-1} \gamma - \gamma' G_0^{-1} \gamma_0 \Bigg\} \Bigg] \\
& \propto \exp \Bigg[-\frac{1}{2} \Bigg\{ \gamma' \Bigg(\sum_{i=1}^{N} Z_i' \Sigma_\varepsilon^{-1} Z_i + G_0^{-1} \Bigg) \gamma - \Bigg(\sum_{i=1}^{N} \Big(y_i' \Sigma_\varepsilon^{-1} Z_i - \beta' X_i' \Sigma_\varepsilon^{-1} Z_i \Big) + \gamma_0' G_0^{-1} \Bigg) \gamma \\
& \quad - \gamma' \Bigg(\sum_{i=1}^{N} \Big(Z_i' \Sigma_\varepsilon^{-1} y_i - Z_i' \Sigma_\varepsilon^{-1} X_i \beta \Big) + G_0^{-1} \gamma_0 \Bigg) \Bigg\} \Bigg] \\
& \propto \exp \Bigg[-\frac{1}{2} \Big\{ \gamma' G_1^{-1} \gamma - \overline{\gamma}' G_1^{-1} \gamma - \gamma' G_1^{-1} \overline{\gamma} \Big\} \Bigg],
\end{aligned}
$$

where the derivation follows the same line of thought as for β and the terms G_1 and $\overline{\gamma}$ are defined as,

$$
G_1^{-1} = \Bigg[\sum_{i=1}^{N} Z_i' \Sigma_\varepsilon^{-1} Z_i + G_0^{-1} \Bigg] \quad \text{and} \quad \overline{\gamma} = G_1 \Bigg[\sum_{i=1}^{N} Z_i' \Sigma_\varepsilon^{-1} (y_i - X_i \beta) + G_0^{-1} \gamma_0 \Bigg]
$$

As done earlier, we add and subtract $\overline{\gamma}' G_1^{-1} \overline{\gamma}$, to complete the square as,

$$
\begin{aligned}
& p(\gamma | \beta, \Sigma_\varepsilon, Z, y) \\
& \propto \exp \left[-\frac{1}{2} \left\{ \gamma' G_1^{-1} \gamma - \overline{\gamma}' G_1^{-1} \gamma - \gamma' G_1^{-1} \overline{\gamma} - \overline{\gamma}' G_1^{-1} \overline{\gamma} + \overline{\gamma}' G_1^{-1} \overline{\gamma} \right\} \right] \\
& \propto \exp \left[-\frac{1}{2} (\gamma - \overline{\gamma})' G_1^{-1} (\gamma - \overline{\gamma}) \right].
\end{aligned}
$$

The last expression is the kernel of a Gaussian or normal density and hence we obtain $\gamma | \beta, \Sigma_\varepsilon, Z, y \sim N_M (\overline{\gamma}, G_1)$.

1.3 Conditional posterior density for $\Sigma_\varepsilon^{-1} \sim W_M(\nu_1, S_1)$

The conditional posterior distribution of the inverse of covariance matrix Σ_ε^{-1} is also derived in a similar way — collect terms that involve Σ_ε^{-1} from the joint posterior distribution (A.1) and then identify the distribution. This is done as follows:

$$
\begin{aligned}
&p(\Sigma_\varepsilon^{-1}|\beta,\gamma,Z,y) \\
&\propto \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1}\Sigma_\varepsilon^{-1}\right) + \sum_{i=1}^{N}(y_i - X_i\beta - Z_i\gamma)'\,\Sigma_\varepsilon^{-1}\,(y_i - X_i\beta - Z_i\gamma)\right\}\right] \\
&\quad\times |\Sigma_\varepsilon|^{-\frac{\nu_0+M+N+1}{2}} \\
&\propto \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1}\Sigma_\varepsilon^{-1}\right) + \sum_{i=1}^{N}\mathrm{tr}\left\{(y_i - X_i\beta - Z_i\gamma)'\,\Sigma_\varepsilon^{-1}\,(y_i - X_i\beta - Z_i\gamma)\right\}\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1}\Sigma_\varepsilon^{-1}\right) + \sum_{i=1}^{N}\mathrm{tr}\left\{(y_i - X_i\beta - Z_i\gamma)\,(y_i - X_i\beta - Z_i\gamma)'\,\Sigma_\varepsilon^{-1}\right\}\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1} + \sum_{i=1}^{N}(y_i - X_i\beta - Z_i\gamma)\,(y_i - X_i\beta - Z_i\gamma)'\right)\Sigma_\varepsilon^{-1}\right\}\right]
\end{aligned}
$$

where in the second proportionality $\sum_{i=1}^{N}(y_i - X_i\beta - Z_i\gamma)'\,\Sigma_\varepsilon^{-1}\,(y_i - X_i\beta - Z_i\gamma)$ is a scalar and trace (tr) of a scalar is the scalar itself. The last expression can be simplified by using the following substitution,

$$
\begin{aligned}
\nu_1 &= \nu_0 + N \\
S_1^{-1} &= S_0^{-1} + \sum_{i=1}^{N}(y_i - X_i\beta - Z_i\gamma)\,(y_i - X_i\beta - Z_i\gamma)'.
\end{aligned}
$$

Bringing back the term $|\Sigma_\varepsilon|^{-\frac{\nu_0+M+N+1}{2}}$, the conditional posterior density of Σ_ε^{-1} simplifies to,

$$
p(\Sigma_\varepsilon^{-1}|\beta,\gamma,Z,y) \propto \exp\left[-\frac{1}{2}\mathrm{tr}\left(S_1^{-1}\Sigma_\varepsilon^{-1}\right)\right] \times |\Sigma_\varepsilon|^{-\frac{\nu_1+M+1}{2}}
$$

which we recognize as the kernel of a Wishart distribution. Hence, we have $\Sigma_\varepsilon^{-1}|\beta,\gamma,Z,y \sim W_M(\nu_1, S_1)$.

1.4 Conditional posterior density for $\tilde{Z}_i \sim N_M(M_{1,i}, M_2)$

The conditional posterior for Z is also derived from the joint posterior density (A.1) by collecting terms involving Z. In this case, we focus on $p(\widetilde{Z}_i|W,y,\Delta)$ since the distribution of $\widetilde{Z}$ is given by $p(\widetilde{Z}|W,y,\Delta) = \prod_{i=1}^{N} p(\widetilde{Z}_i|W,y,\Delta)$, where we recall that $\widetilde{Z}_i = (z_{1i},\cdots,z_{Mi})'$, or equivalently $Z_i = diag(\widetilde{Z}_i)$ and $\Delta = (\beta,\gamma,\Sigma_\varepsilon,\omega,\sigma_Z^2,\sigma_u^2)$. So, collecting terms involving $\widetilde{Z}_i$ in $p(\widetilde{Z}_i|W,y,\Delta)$, the kernel can be written as,

$$
\begin{aligned}
&p(\widetilde{Z}_i|W,y,\Delta) \\
&\propto \exp\left[-\frac{1}{2}\left\{(y_i - X_i\beta - Z_i\gamma)'\,\Sigma_\varepsilon^{-1}\,(y_i - X_i\beta - Z_i\gamma)\right.\right. \\
&\qquad \left.\left. + \frac{1}{\sigma_Z^2}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right) + \frac{1}{\sigma_u^2}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right\}\right]
\end{aligned} \tag{A.2}
$$

The derivations are involved, so we ignore the exponentiation and work with one expression at a time. Concentrating on the first term we have,

$$\begin{aligned}
&(y_i - X_i\beta - Z_i\gamma)' \Sigma_\varepsilon^{-1} (y_i - X_i\beta - Z_i\gamma) \\
&\propto \gamma' Z_i' \Sigma_\varepsilon^{-1} Z_i \gamma - \gamma' Z_i' \Sigma_\varepsilon^{-1} (y_i - X_i\beta) - (y_i - X_i\beta)' \Sigma_\varepsilon^{-1} Z_i \gamma \\
&\propto \widetilde{Z}_i' \Psi \widetilde{Z}_i - \widetilde{Z}_i' D_\gamma \Sigma_\varepsilon^{-1} (y_i - X_i\beta) - (y_i - X_i\beta)' \Sigma_\varepsilon^{-1} Z_i \gamma
\end{aligned} \tag{A.3}$$

where[1] $D_\gamma = diag(\gamma)$ converts the column vector γ to a diagonal matrix, and $\Psi = \Gamma \odot \Sigma_\varepsilon^{-1}$ with $\Gamma = \gamma\gamma'$ where $\odot$ is the dot product (or Hadamard product) of Γ and Σ_ε^{-1}. The second and the third terms in equation (A.2) lend themselves to easy manipulation,

$$\frac{1}{\sigma_Z^2}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right) = \frac{1}{\sigma_Z^2}\left[\widetilde{Z}_i'\widetilde{Z}_i - \widetilde{Z}_i' X_i\omega - \omega' X_i' \widetilde{Z}_i + \omega' X_i' X_i \omega\right], \tag{A.4}$$

$$\frac{1}{\sigma_u^2}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right) = \frac{1}{\sigma_Z^2}\left[\widetilde{W}_i'\widetilde{W}_i - \widetilde{W}_i'\widetilde{Z}_i - \widetilde{Z}_i'\widetilde{W}_i + \widetilde{Z}_i'\widetilde{Z}_i\right]. \tag{A.5}$$

Combining (A.3), (A.4) and (A.5), the resulting expression is as follows,

$$\begin{aligned}
&(y_i - X_i\beta - Z_i\gamma)' \Sigma_\varepsilon^{-1} (y_i - X_i\beta - Z_i\gamma) + \frac{1}{\sigma_Z^2}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right) \\
&+ \frac{1}{\sigma_u^2}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right) \\
&\propto \widetilde{Z}_i'\left[\Psi + \left(\frac{1}{\sigma_Z^2} + \frac{1}{\sigma_u^2}\right) I_M\right]\widetilde{Z}_i - \widetilde{Z}_i'\left[D_\gamma \Sigma_\varepsilon^{-1} (y_i - X_i\beta) + \frac{\widetilde{W}_i}{\sigma_u^2} + \frac{X_i\omega}{\sigma_Z^2}\right] \\
&- \left[(y_i - X_i\beta)' \Sigma_\varepsilon^{-1} D_\gamma + \frac{\widetilde{W}_i'}{\sigma_u^2} + \frac{\omega' X_i'}{\sigma_Z^2}\right]\widetilde{Z}_i \\
&\propto \widetilde{Z}_i' M_2^{-1} \widetilde{Z}_i - \widetilde{Z}_i' M_2^{-1} M_{1i} - M_{1i}' M_2^{-1} \widetilde{Z}_i + M_{1i}' M_2^{-1} M_{1i} - M_{1i}' M_2^{-1} M_{1i} \\
&\propto \left(\widetilde{Z}_i - M_{1i}\right)' M_2^{-1} \left(\widetilde{Z}_i - M_{1i}\right)
\end{aligned} \tag{A.6}$$

where the derivation follows a similar line of thought as that of β and the terms M_{1i} and M_2 are defined as,

$$M_2 = \left[\Psi + \left(\frac{1}{\sigma_Z^2} + \frac{1}{\sigma_u^2}\right) I_M\right]^{-1} \text{ and } M_{1i} = M_2\left[D_\gamma \Sigma_\varepsilon^{-1} (y_i - X_i\beta) + \frac{\widetilde{W}_i}{\sigma_u^2} + \frac{X_i\omega}{\sigma_Z^2}\right].$$

Substituting (A.6) in (A.2), we arrive at the following expression,

$$p(\widetilde{Z}_i | W, y, \Delta) \propto \exp\left[-\frac{1}{2}\left(\widetilde{Z}_i - M_{1i}\right)' M_2^{-1} \left(\widetilde{Z}_i - M_{1i}\right)\right]. \tag{A.7}$$

We recognize equation (A.7) as the kernel of a normal distribution, hence the conditional posterior distribution is $\widetilde{Z}_i | W, y, \Delta \sim N_M(M_{1,i}, M_2)$. Alternatively, we can write $\widetilde{Z} | W, y, \Delta \sim N_M(\iota_N \otimes M_{1,i}, I_N \otimes M_2)$, where ι_N is an $(N \times 1)$ vector of ones and I_N is an $(N \times N)$ identity matrix.

[1] With some manipulations it can be shown that,

$$\begin{aligned}
\gamma' Z_i' \Sigma_\varepsilon^{-1} Z_i \gamma &= \widetilde{Z}_i' \Psi \widetilde{Z}_i \\
\gamma' Z_i' \Sigma_\varepsilon^{-1} (y_i - X_i\beta) &= \widetilde{Z}_i' D_\gamma \Sigma_\varepsilon^{-1} (y_i - X_i\beta).
\end{aligned}$$

1.5 Conditional posterior density for $\omega \sim N_M(\omega_1, \Sigma_\omega)$

The derivation for the conditional posterior density of ω follows a similar line of thought as that of β and γ. Collecting terms involving ω from the joint posterior distribution (A.1)), the derivation proceeds as follows,

$$
\begin{aligned}
&p(\omega|Z,\sigma_Z^2)\\
&\propto \exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right) + (\omega-\omega_0)'\,O_0^{-1}\,(\omega-\omega_0)\right\}\right]\\
&\propto \exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}\left(\widetilde{Z}_i'\widetilde{Z}_i - \widetilde{Z}_i'X_i\omega - \omega'X_i'\widetilde{Z}_i + \omega'X_i'X_i\omega\right)\right.\right.\\
&\qquad\left.\left. + \,\omega'O_0^{-1}\omega - \omega'O_0^{-1}\omega_0 - \omega_0'O_0^{-1}\omega + \omega_0'O_0^{-1}\omega_0\right\}\right]\\
&\propto \exp\left[-\frac{1}{2}\left\{\omega'\left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}X_i'X_i + O_0^{-1}\right) - \left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}\widetilde{Z}_i'X_i + \omega_0'O_0^{-1}\right)\omega\right.\right.\\
&\qquad\left.\left. - \,\omega'\left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}X_i'\widetilde{Z}_i + O_0^{-1}\omega_0\right)\right\}\right]\\
&\propto \exp\left[-\frac{1}{2}\left\{\omega'\Sigma_\omega^{-1}\omega - \omega_1'\Sigma_\omega^{-1}\omega - \omega'\Sigma_\omega^{-1}\omega_1\right\}\right]
\end{aligned}
$$

where the terms Σ_ω^{-1} and ω_1 are defined as follows,

$$
\Sigma_\omega^{-1} = \left[\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}X_i'X_i + O_0^{-1}\right] \quad \text{and} \quad \omega_1 = \Sigma_\omega\left[\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}X_i'\widetilde{Z}_i + O_0^{-1}\omega_0\right].
$$

As done earlier, we add and subtract $\omega_1'\Sigma_\omega^{-1}\omega_1$, to complete the square as,

$$
\begin{aligned}
p(\omega|Z,\sigma_Z^2) &\propto \exp\left[-\frac{1}{2}\left\{\omega'\Sigma_\omega^{-1}\omega - \omega_1'\Sigma_\omega^{-1}\omega - \omega'\Sigma_\omega^{-1}\omega_1 + \omega_1'\Sigma_\omega^{-1}\omega_1 - \omega_1'\Sigma_\omega^{-1}\omega_1\right\}\right]\\
&\propto \exp\left[-\frac{1}{2}(\omega-\omega_1)'\,\Sigma_\omega^{-1}\,(\omega-\omega_1)\right],
\end{aligned}
$$

which we recognize as the kernel of a normal density, hence $\omega|Z,\sigma_Z^2 \sim N_M(\omega_1, \Sigma_\omega)$.

1.6 Conditional posterior density for $\sigma_Z^2 \sim IG(\delta_1^*, \delta_2^*)$

The conditional posterior density of σ_Z^2 can be derived from the joint posterior density (A.1) by collecting terms involving σ_Z^2 as below.

$$
\begin{aligned}
&p(\sigma_Z^2|Z,\omega)\\
&\propto \left(\sigma_Z^2\right)^{-(\delta_1+1)}\exp\left[-\frac{\delta_2}{\sigma_Z^2}\right]\prod_{i=1}^{N}\left\{\left(\sigma_Z^2\right)^{-M/2}\exp\left[-\frac{1}{2\sigma_Z^2}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right)\right]\right\}\\
&\propto \left(\sigma_Z^2\right)^{-(\delta_1+1+NM/2)}\exp\left[-\frac{1}{\sigma_Z^2}\left\{\delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right)\right\}\right]\\
&\propto \left(\sigma_Z^2\right)^{-(\delta_1^*+1)}\exp\left[-\frac{\delta_2^*}{\sigma_Z^2}\right],
\end{aligned}
$$

which we recognize as the kernel of an inverse gamma (IG) distribution,

$$\begin{aligned}\sigma_Z^2|Z,\omega \sim IG\left(\delta_1^*,\delta_2^*\right) \qquad &\text{where}\\ \delta_1^* = \delta_1 + NM/2 \qquad &\text{and}\\ \delta_2^* = \delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right).\end{aligned}$$

1.7 Conditional posterior density for $\sigma_u^2 \sim IG(\delta_3^*, \delta_4^*)$

Similar to σ_Z^2, we can show that,

$$\begin{aligned}p(\sigma_u^2|W,Z) &\propto \left(\sigma_u^2\right)^{-(\delta_3+1+NM/2)}\exp\left[-\frac{1}{\sigma_u^2}\left\{\delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right\}\right]\\ &\propto \left(\sigma_u^2\right)^{-\left(\delta_3^*+1\right)}\exp\left[-\frac{\delta_4^*}{\sigma_u^2}\right],\end{aligned}$$

which we recognize as the kernel of an inverse Gamma (IG) distribution,

$$\begin{aligned}\sigma_u^2|Z,W \sim IG\left(\delta_3^*,\delta_4^*\right) \qquad &\text{where}\\ \delta_3^* = \delta_3 + NM/2 \qquad &\text{and}\\ \delta_4^* = \delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right).\end{aligned}$$

2 Appendix B - MFVB for SURME model

To derive the optimal q-densities, we utilize some properties of the inverse gamma and Wishart distributions. Let us start with the inverse gamma distribution. If $p(\sigma_Z^2) \sim IG(\delta_1, \delta_2)$, the pdf is given by,

$$p(\sigma_Z^2) = \frac{\delta_2^{\delta_1}}{\Gamma(\delta_1)}(\sigma_Z^2)^{-\delta_1-1}\exp\left[-\frac{\delta_2}{\sigma_Z^2}\right],$$

and, $E\left[\sigma_Z^2\right] = \frac{\delta_2}{\delta_1-1}$, $E\left[\ln\sigma_Z^2\right] = \ln(\delta_2) - \psi(\delta_1)$, $E\left[\sigma_Z^{-2}\right] = \frac{\delta_1}{\delta_2}$. The notation $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ is the digamma function.

If $p(X) \sim W_p(n, V)$, with $n > p$, then the pdf of the Wishart distribution is given by,

$$p(X) = \frac{|X|^{\frac{n-p-1}{2}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})}|V|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\text{tr}\left[V^{-1}X\right]\right]$$

and $E[X] = nV$, $E[\ln|X|] = \ln|V| + p\ln 2 + \psi_p\left(\frac{n}{2}\right)$. The notation $\psi_p(z)$ denotes the multivariate digamma function and is the derivative of the log of the multivariate gamma function:

$$\psi_p(z) = \frac{\partial\ln\Gamma_p(z)}{\partial z} = \sum_{l=1}^{p}\psi\left(z + \frac{(1-l)}{2}\right).$$

Moreover, If $p(X) \sim W_p(n, V)$, then $p(Y = 1/X) \sim IW_p(n, V^{-1})$ and the pdf of Y is given by,

$$p(Y) = \frac{|Y|^{-\frac{n+p+1}{2}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})}|V|^{\frac{n}{2}}\exp\left[-\frac{1}{2}\text{tr}\left[VY^{-1}\right]\right].$$

Moving to the SURME model, the prior distribution on the parameters and latent variables are as follows:

$$\begin{aligned}
&p(\beta) \sim N_M(\beta_0, B_0), \qquad p(\gamma) \sim N_M(\gamma_0, G_0), \quad p\left(\Sigma_\varepsilon^{-1}\right) \sim W_M(\nu_0, S_0),\\
&p(\tilde{Z}_i) \sim N_M\left(X_i\omega, \sigma_z^2 I_M\right), \quad p(\omega) \sim N_M(\omega_0, O_0), \qquad p(\sigma_Z^2) \sim IG(\delta_1, \delta_2),\\
&p(\sigma_u^2) \sim IG(\delta_3, \delta_4).
\end{aligned}$$

The optimal q-densities for the SURME model are as follows:

$$\begin{aligned}
&q(\beta) \sim N_K\left(\mu_{q(\beta)}, \Sigma_{q(\beta)}\right), \qquad q(\gamma) \sim N_M\left(\mu_{q(\gamma)}, \Sigma_{q(\gamma)}\right), \quad q\left(\Sigma_\varepsilon^{-1}\right) \sim W_M\left(\nu_1, B_{q(\Sigma)}\right),\\
&q(\omega) \sim N_K\left(\mu_{q(\omega)}, \Sigma_{q(\omega)}\right), \quad q(\tilde{Z}_i) \sim N_M(\mu_{q(\tilde{Z}_i)}, \Sigma_{q(\tilde{Z}_i)}), \quad q\left(\sigma_Z^2\right) \sim IG\left(\delta_1^*, B_{q(\sigma_Z^2)}\right),\\
&q\left(\sigma_u^2\right) \sim IG\left(\delta_3^*, B_{q(\sigma_u^2)}\right),
\end{aligned}$$

where we have,

$$\begin{aligned}
E_{q(\theta)}[\beta] &= \mu_{q(\beta)}, & E_{q(\theta)}[\beta\beta'] &= \Sigma_{q(\beta)} + \mu_{q(\beta)}\mu'_{q(\beta)},\\
E_{q(\theta)}[\gamma] &= \mu_{q(\gamma)}, & E_{q(\theta)}[\gamma\gamma'] &= \Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)},\\
E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] &= \nu_1 B_{q(\Sigma)}, & E_{q(\theta)}\left[\ln\left|\Sigma_\varepsilon^{-1}\right|\right] &= -E_{q(\theta)}\left[\ln|\Sigma_\varepsilon|\right],\\
E_{q(\theta)}[\omega] &= \mu_{q(\omega)}, & E_{q(\theta)}[\omega\omega'] &= \Sigma_{q(\omega)} + \mu_{q(\omega)}\mu'_{q(\omega)},\\
E_{q(\theta)}\left[\tilde{Z}_i\right] &= diag(\mu_{q(\tilde{Z}_i)}), & E_{q(\theta)}\left[\tilde{Z}_i\tilde{Z}_i'\right] &= \Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)},\\
E_{q(\theta)}\left[\sigma_Z^{-2}\right] &= \delta_1^*/B_{q(\sigma_Z^2)}, & E_{q(\theta)}\left[\ln\sigma_Z^2\right] &= \ln B_{q(\sigma_Z^2)} - \psi(\delta_1^*),
\end{aligned}$$

$$E_{q(\theta)}\left[\sigma_u^{-2}\right]=\delta_3^*/B_{q(\sigma_u^2)}, \qquad E_{q(\theta)}\left[\ln\sigma_u^2\right]=\ln B_{q(\sigma_u^2)}-\psi\left(\delta_3^*\right),$$

and $E_{q(\theta)}[\ln|\Sigma_\varepsilon|] = -\ln\left|B_{q(\Sigma)}\right| - M\ln 2 - \psi_M\left(\frac{\nu_1}{2}\right)$. Moreover, in the following derivations we will interchangeably use the following notations: $Var_{q(\theta)}[\beta]=\Sigma_{q(\beta)}$, $Var_{q(\theta)}[\gamma]=\Sigma_{q(\gamma)}$, $Var_{q(\theta)}[\omega]=\Sigma_{q(\omega)}$, and $Var_{q(\theta)}[\tilde{Z}_i]=\Sigma_{q(\tilde{Z}_i)}$.

We next present two equalities which are used later to derive the expressions for the optimal density of Σ_ε^{-1},

$$\begin{aligned} E_{q(\theta)}\left[X_i\beta\beta' X_i'\right] &= X_i E_{q(\theta)}\left[\beta\beta'\right]X_i' = X_i\left(\Sigma_{q(\beta)}+\mu_{q(\beta)}\mu_{q(\beta)}'\right)X_i' \\ &= X_i\Sigma_{q(\beta)}X_i' + X_i\mu_{q(\beta)}\mu_{q(\beta)}'X_i' \end{aligned} \tag{B.1}$$

$$\begin{aligned} E_{q(\theta)}\left[Z_i\gamma\gamma' Z_i'\right] &= E_{q(\theta)}\left[\left(\tilde{Z}_i\tilde{Z}_i'\right)\odot(\gamma\gamma')\right] = E_{q(\theta)}\left[\tilde{Z}_i\tilde{Z}_i'\right]\odot E_{q(\theta)}\left[\gamma\gamma'\right] \\ &= \left(\Sigma_{q(\tilde{Z}_i)}+\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\right)\odot\left(\Sigma_{q(\gamma)}+\mu_{q(\gamma)}\mu_{q(\gamma)}'\right). \end{aligned} \tag{B.2}$$

2.1 Updates of $q(\beta)$, $\mu_{q(\beta)}$, and $\Sigma_{q(\beta)}$)

The Gibbs sampling presented in Algorithm 1 shows that the conditional posterior for $\beta \sim N_K\left(\overline{\beta}, B_1\right)$, where,

$$B_1^{-1}=\left[\sum_{i=1}^N X_i'\Sigma_\varepsilon^{-1}X_i + B_0^{-1}\right], \quad \overline{\beta}=B_1\left[\sum_{i=1}^N X_i'\Sigma_\varepsilon^{-1}y_i^* + B_0^{-1}\beta_0\right],$$

and $y_i^*=y_i-Z_i\gamma$. The kernel of the conditional posterior density for β can be written as,

$$\begin{aligned} p(\beta|\gamma,\Sigma_\varepsilon,Z,y) &= \exp\left[-\frac{1}{2}\left\{\beta' B_1^{-1}\beta - \overline{\beta}' B_1^{-1}\beta - \beta' B_1^{-1}\overline{\beta}\right\}\right]\times\kappa_1 \\ &= \exp\Bigg[-\frac{1}{2}\Bigg\{\beta'\Bigg(\sum_{i=1}^N X_i'\Sigma_\varepsilon^{-1}X_i + B_0^{-1}\Bigg)\beta - \Bigg(\sum_{i=1}^N\left(y_i'\Sigma_\varepsilon^{-1}X_i - \gamma' Z_i'\Sigma_\varepsilon^{-1}X_i\right) + \beta_0' B_0^{-1}\Bigg)\beta \\ &\quad - \beta'\Bigg(\sum_{i=1}^N\left(X_i'\Sigma_\varepsilon^{-1}y_i - X_i'\Sigma_\varepsilon^{-1}Z_i\gamma\right) + B_0^{-1}\beta_0\Bigg)\Bigg\}\Bigg]\times\kappa_1, \end{aligned}$$

where κ_1 denotes constants with respect to the function argument. Taking logarithms on both sides of the equation yields,

$$\begin{aligned} \log p(\beta|\gamma,\Sigma_\varepsilon,Z,y) &= -\frac{1}{2}\Bigg\{\beta'\Bigg(\sum_{i=1}^N X_i'\Sigma_\varepsilon^{-1}X_i + B_0^{-1}\Bigg)\beta - \Bigg(\sum_{i=1}^N\left(y_i'\Sigma_\varepsilon^{-1}X_i - \gamma' Z_i'\Sigma_\varepsilon^{-1}X_i\right) \\ &\quad + \beta_0' B_0^{-1}\Bigg)\beta - \beta'\Bigg(\sum_{i=1}^N\left(X_i'\Sigma_\varepsilon^{-1}y_i - X_i'\Sigma_\varepsilon^{-1}Z_i\gamma\right) + B_0^{-1}\beta_0\Bigg)\Bigg\} + \log(\kappa_1) \end{aligned}$$

Then, taking expectations with respect to all parameters except β gives,

$$\begin{aligned} \log q\left(\beta\right) =& E_{q(\theta)}\left\{\log p(\beta|\gamma,\Sigma_\varepsilon,Z,y)\right\} \\ =& -\frac{1}{2}\Bigg\{\beta'\Bigg(\sum_{i=1}^N X_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_i + B_0^{-1}\Bigg)\beta \end{aligned}$$

$$
-\Bigg(\sum_{i=1}^{N}\Big(y_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] X_i - E_{q(\theta)}\left[\gamma' Z_i' \Sigma_\varepsilon^{-1}\right] X_i\Big) + \beta_0' B_0^{-1}\Bigg)\beta
$$

$$
-\beta'\Bigg(\sum_{i=1}^{N}\Big(X_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] y_i - X_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1} Z_i \gamma\right]\Big) + B_0^{-1}\beta_0\Bigg)\Bigg\} + \log(\kappa_1)
$$

From the properties of Wishart distribution, we know $E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] = \nu_1 B_{q(\Sigma)}$. Moreover, $E_{q(\theta)}\left[\Sigma_\varepsilon^{-1} Z_i \gamma\right] = E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] E_{q(\theta)}\left[Z_i\right] E_{q(\theta)}\left[\gamma\right] = \nu_1 B_{q(\Sigma)} diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}$. Consequently, we have,

$$
\begin{aligned}
\log q(\beta) = -\frac{1}{2}\Bigg\{\beta'\Bigg(\sum_{i=1}^{N} X_i'\left(\nu_1 B_{q(\Sigma)}\right) X_i + B_0^{-1}\Bigg)\beta \\
-\Bigg(\sum_{i=1}^{N}\Big(y_i'\left(\nu_1 B_{q(\Sigma)}\right) X_i - \mu_{q(\gamma)}' diag\left(\mu_{q(\tilde{Z}_i)}\right)\left(\nu_1 B_{q(\Sigma)}\right) X_i\Big) + \beta_0' B_0^{-1}\Bigg)\beta \\
-\beta'\Bigg(\sum_{i=1}^{N}\Big(X_i'\left(\nu_1 B_{q(\Sigma)}\right) y_i - X_i'\left(\nu_1 B_{q(\Sigma)}\right) diag\left(\mu_{q(\tilde{Z}_i)}\right)\mu_{q(\gamma)}\Big) + B_0^{-1}\beta_0\Bigg)\Bigg\} + \log(\kappa_1)
\end{aligned}
$$

We next make use of the following notations in the expression for $\log q(\beta)$,

$$
\begin{aligned}
\Sigma_{q(\beta)}^{-1} &= \left[\sum_{i=1}^{N} X_i'\left(\nu_1 B_{q(\Sigma)}\right) X_i + B_0^{-1}\right], \\
\mu_{q(\beta)} &= \Sigma_{q(\beta)}\left[\sum_{i=1}^{N} X_i'\left(\nu_1 B_{q(\Sigma)}\right)\left(y_i - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right) + B_0^{-1}\beta_0\right].
\end{aligned}
$$

Then the optimal density can be written as,

$$
q(\beta) \propto \exp\Bigg[-\frac{1}{2}\left\{\beta'\Sigma_{q(\beta)}^{-1}\beta - \mu_{q(\beta)}'\Sigma_{q(\beta)}^{-1}\beta - \beta'\Sigma_{q(\beta)}^{-1}\mu_{q(\beta)}\right\}\Bigg].
$$

Adding and subtracting $\mu_{q(\beta)}'\Sigma_{q(\beta)}^{-1}\mu_{q(\beta)}$ inside the curly braces, the square can be completed as,

$$
\begin{aligned}
q(\beta) &\propto \exp\left[-\frac{1}{2}\left\{\beta'\Sigma_{q(\beta)}^{-1}\beta - \mu_{q(\beta)}'\Sigma_{q(\beta)}^{-1}\beta - \beta'\Sigma_{q(\beta)}^{-1}\mu_{q(\beta)} - \mu_{q(\beta)}'\Sigma_{q(\beta)}^{-1}\mu_{q(\beta)} + \mu_{q(\beta)}'\Sigma_{q(\beta)}^{-1}\mu_{q(\beta)}\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}(\beta - \mu_{q(\beta)})'\Sigma_{q(\beta)}^{-1}(\beta - \mu_{q(\beta)})\right].
\end{aligned}
$$

The result is the kernel of a Gaussian or normal density. Hence, the optimal q-density for β has the following form,

$$
\begin{aligned}
q(\beta) &= f_{N_K}\left(\mu_{q(\beta)}, \Sigma_{q(\beta)}\right), \qquad \text{where,} \\
\mu_{q(\beta)} &= \Sigma_{q(\beta)}\left[\sum_{i=1}^{N} X_i'\left(\nu_1 B_{q(\Sigma)}\right)\left(y_i - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right) + B_0^{-1}\beta_0\right], \\
\Sigma_{q(\beta)} &= \left[\sum_{i=1}^{N} X_i'\left(\nu_1 B_{q(\Sigma)}\right) X_i + B_0^{-1}\right]^{-1}.
\end{aligned}
\tag{B.3}
$$

2.2 Updates of $q(\gamma), \mu_{q(\gamma)}$, and $\Sigma_{q(\gamma)}$

According to the Gibbs sampling presented in Algorithm 1, the conditional posterior for $\gamma \sim N_M\left(\overline{\gamma}, G_1\right)$, where,

$$G_1^{-1} = \left[\sum_{i=1}^{N} Z_i'\Sigma_\varepsilon^{-1}Z_i + G_0^{-1}\right], \quad \overline{\gamma} = G_1\left[\sum_{i=1}^{N} Z_i'\Sigma_\varepsilon^{-1}\tilde{y}_i + G_0^{-1}\gamma_0\right],$$

and $\tilde{y}_i = y_i - X_i\beta$. The kernel of the conditional posterior density for γ can be written as,

$$\begin{aligned}
p(\gamma|\beta, \Sigma_\varepsilon, Z, y) &= \exp\left[-\frac{1}{2}\left\{\gamma' G_1^{-1}\gamma - \overline{\gamma}' G_1^{-1}\gamma - \gamma' G_1^{-1}\overline{\gamma}\right\}\right] \times \kappa_2 \\
&= \exp\Bigg[-\frac{1}{2}\Bigg\{\gamma'\Big(\sum_{i=1}^{N} Z_i'\Sigma_\varepsilon^{-1}Z_i + G_0^{-1}\Big)\gamma - \Big(\sum_{i=1}^{N}\Big(y_i'\Sigma_\varepsilon^{-1}Z_i - \beta' X_i'\Sigma_\varepsilon^{-1}Z_i\Big) + \gamma_0' G_0^{-1}\Big)\gamma \\
&\quad - \gamma'\Big(\sum_{i=1}^{N}\Big(Z_i'\Sigma_\varepsilon^{-1}y_i - Z_i'\Sigma_\varepsilon^{-1}X_i\beta\Big) + G_0^{-1}\gamma_0\Big)\Bigg\}\Bigg] \times \kappa_2,
\end{aligned}$$

where κ_2 are constants with respect to function argument. Taking logarithms on both sides yields,

$$\begin{aligned}
\log p(\gamma|\beta, \Sigma_\varepsilon, Z, y) = -\frac{1}{2}\Bigg\{&\gamma'\Big(\sum_{i=1}^{N} Z_i'\Sigma_\varepsilon^{-1}Z_i + G_0^{-1}\Big)\gamma - \Big(\sum_{i=1}^{N}\Big(y_i'\Sigma_\varepsilon^{-1}Z_i - \beta' X_i'\Sigma_\varepsilon^{-1}Z_i\Big) + \gamma_0' G_0^{-1}\Big)\gamma \\
&- \gamma'\Big(\sum_{i=1}^{N}\Big(Z_i'\Sigma_\varepsilon^{-1}y_i - Z_i'\Sigma_\varepsilon^{-1}X_i\beta\Big) + G_0^{-1}\gamma_0\Big)\Bigg\} + \log(\kappa_2)
\end{aligned}$$

Then taking expectations with respect to all parameters except γ gives,

$$\begin{aligned}
&\log q\left(\gamma\right) = E_{q(\theta)}\left\{\log p(\gamma|\beta, \Sigma_\varepsilon, Z, y)\right\} \\
&= -\frac{1}{2}\Bigg\{\gamma'\Big(\sum_{i=1}^{N} E_{q(\theta)}\left[Z_i'\Sigma_\varepsilon^{-1}Z_i\right] + G_0^{-1}\Big)\gamma - \Big(\sum_{i=1}^{N}\Big(y_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}Z_i\right] - E_{q(\theta)}\left[\beta'\right] X_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}Z_i\right]\Big) \\
&\quad + \gamma_0' G_0^{-1}\Big)\gamma - \gamma'\Big(\sum_{i=1}^{N}\Big(E_{q(\theta)}\left[Z_i'\Sigma_\varepsilon^{-1}\right] y_i - E_{q(\theta)}\left[Z_i'\Sigma_\varepsilon^{-1}\right] X_i E_{q(\theta)}\left[\beta\right]\Big) + G_0^{-1}\gamma_0\Big)\Bigg\} + \log(\kappa_2).
\end{aligned}$$

In the above expression for $\log q(\gamma)$, there are two terms involving expectations and their final expression are as follows:

$$\begin{aligned}
E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}Z_i\right] &= \left(\nu_1 B_{q(\Sigma)}\right) diag(\mu_{q(\tilde{Z}_i)}), \\
E_{q(\theta)}\left[Z_i'\Sigma_\varepsilon^{-1}Z_i\right] &= E_{q(\theta)}\left[\left(\tilde{Z}_i\tilde{Z}_i'\right) \odot \Sigma_\varepsilon^{-1}\right] \\
&= E_{q(\theta)}\left[\left(\tilde{Z}_i\tilde{Z}_i'\right)\right] \odot E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] \\
&= \left(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)}\right) \odot \left(\nu_1 B_{q(\Sigma)}\right).
\end{aligned}$$

Substituting the expressions for $E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}Z_i\right]$ and $E_{q(\theta)}\left[Z_i'\Sigma_\varepsilon^{-1}Z_i\right]$ in the equation for $\log q(\gamma)$ we have,

$$\log q\left(\gamma\right) = E_{q(\theta)}\left\{\log p(\gamma|\beta, \Sigma_\varepsilon, Z, y)\right\}$$

$$
\begin{aligned}
&= -\frac{1}{2}\Bigg\{\gamma'\Bigg(\sum_{i=1}^{N}\Big[\Big(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)}\Big) \odot \Big(\nu_1 B_{q(\Sigma)}\Big)\Big] + G_0^{-1}\Bigg)\gamma \\
&\quad - \Bigg(\sum_{i=1}^{N}\Big[y_i'\Big(\nu_1 B_{q(\Sigma)}\Big) diag(\mu_{q(\tilde{Z}_i)}) - \mu'_{q(\beta)}X_i'\Big(\nu_1 B_{q(\Sigma)}\Big) diag(\mu_{q(\tilde{Z}_i)})\Big] + \gamma_0' G_0^{-1}\Bigg)\gamma \\
&\quad - \gamma'\Bigg(\sum_{i=1}^{N}\Big[\Big(\nu_1 B_{q(\Sigma)}\Big) diag(\mu_{q(\tilde{Z}_i)})y_i - \Big(\nu_1 B_{q(\Sigma)}\Big) diag(\mu_{q(\tilde{Z}_i)})X_i\mu_{q(\beta)}\Big] + G_0^{-1}\gamma_0\Bigg)\Bigg\} + \log(\kappa_2).
\end{aligned}
$$

Next, we introduce two terms which are defined as follows,

$$
\begin{aligned}
\Sigma_{q(\gamma)}^{-1} &= \left[\sum_{i=1}^{N}\Big(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)}\Big) \odot \Big(\nu_1 B_{q(\Sigma)}\Big) + G_0^{-1}\right], \\
\mu_{q(\gamma)} &= \Sigma_{q(\gamma)}\left[\sum_{i=1}^{N} diag(\mu_{q(\tilde{Z}_i)})\Big(\nu_1 B_{q(\Sigma)}\Big)\Big(y_i - X_i\mu_{q(\beta)}\Big) + G_0^{-1}\gamma_0\right],
\end{aligned}
$$

and substitute them in the expression for $\log q(\gamma)$. The optimal density $q(\gamma)$ can then be expressed as,

$$
q\left(\gamma\right) \propto \exp\left[-\frac{1}{2}\left\{\gamma'\Sigma_{q(\gamma)}^{-1}\gamma - \mu'_{q(\gamma)}\Sigma_{q(\gamma)}^{-1}\gamma - \gamma'\Sigma_{q(\gamma)}^{-1}\mu_{q(\gamma)}\right\}\right].
$$

Adding and subtracting $\mu'_{q(\gamma)}\Sigma_{q(\gamma)}^{-1}\mu_{q(\gamma)}$ inside the curly braces, the square can be completed as,

$$
\begin{aligned}
q\left(\gamma\right) &\propto \exp\left[-\frac{1}{2}\left\{\gamma'\Sigma_{q(\gamma)}^{-1}\gamma - \mu'_{q(\gamma)}\Sigma_{q(\gamma)}^{-1}\gamma - \gamma'\Sigma_{q(\gamma)}^{-1}\mu_{q(\gamma)} - \mu'_{q(\gamma)}\Sigma_{q(\gamma)}^{-1}\mu_{q(\gamma)} + \mu'_{q(\gamma)}\Sigma_{q(\gamma)}^{-1}\mu_{q(\gamma)}\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}(\gamma - \mu_{q(\gamma)})'\Sigma_{q(\gamma)}^{-1}(\gamma - \mu_{q(\gamma)})\right].
\end{aligned}
$$

The result is the kernel of a Gaussian or normal density. Hence, the optimal q-density for γ is given by,

$$
\begin{aligned}
q\left(\gamma\right) &= f_{N_M}\Big(\mu_{q(\gamma)}, \Sigma_{q(\gamma)}\Big), \qquad \text{where,} \\
\mu_{q(\gamma)} &= \Sigma_{q(\gamma)}\left[\sum_{i=1}^{N} diag(\mu_{q(\tilde{Z}_i)})\Big(\nu_1 B_{q(\Sigma)}\Big)\Big(y_i - X_i\mu_{q(\beta)}\Big) + G_0^{-1}\gamma_0\right], \\
\Sigma_{q(\gamma)} &= \left[\sum_{i=1}^{N}\Big(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)}\Big) \odot \Big(\nu_1 B_{q(\Sigma)}\Big) + G_0^{-1}\right]^{-1}.
\end{aligned}
\tag{B.4}
$$

2.3 Updates of $q(\Sigma_{\varepsilon}^{-1}), \nu_1$, and $B_{q(\Sigma_{\varepsilon})}$

According to the Gibbs sampling, presented in Algorithm 1, we know that the conditional posterior for $\Sigma_{\varepsilon}^{-1} \sim W_M\left(\nu_1, S_1\right)$, where $\nu_1 = \nu_0 + N$, and

$$
S_1^{-1} = \left[S_0^{-1} + \sum_{i=1}^{N}\left(y_i - X_i\beta - Z_i\gamma\right)\left(y_i - X_i\beta - Z_i\gamma\right)'\right].
$$

Therefore, the kernel of the conditional posterior density for Σ_ε can be written as,[2]

$$p(\Sigma_\varepsilon|\beta,\gamma,Z,y) = \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1} + \sum_{i=1}^{N}\left(y_i - X_i\beta - Z_i\gamma\right)\left(y_i - X_i\beta - Z_i\gamma\right)'\right)\Sigma_\varepsilon^{-1}\right\}\right] \times \kappa_3,$$

where κ_3 are constants with respect to function argument. Taking logarithm of the above expression we have,

$$\log p(\Sigma_\varepsilon|\beta,\gamma,Z,y) = -\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1} + \sum_{i=1}^{N}\left(y_i - X_i\beta - Z_i\gamma\right)\left(y_i - X_i\beta - Z_i\gamma\right)'\right)\Sigma_\varepsilon^{-1}\right\}\Bigg] + \log(\kappa_3).$$

We now take expectation with respect to all parameters except Σ_ε, which yields,

$$\begin{aligned}
&\log q\left(\Sigma_\varepsilon\right) = E_{q(\theta)}\left\{\log p(\Sigma_\varepsilon|\beta,\gamma,Z,y)\right\}\\
&= -\frac{1}{2}\left\{\mathrm{tr}\left(S_0^{-1} + \sum_{i=1}^{N} E_{q(\theta)}\left[\left(y_i - X_i\beta - Z_i\gamma\right)\left(y_i - X_i\beta - Z_i\gamma\right)'\right)\right]\Sigma_\varepsilon^{-1}\right\}\Bigg] + \log(\kappa_3).
\end{aligned}$$

Focusing on the quadratic terms and taking expectation, we have,

$$\begin{aligned}
&E_{q(\theta)}\left[\left(y_i - X_i\beta - Z_i\gamma\right)\left(y_i - X_i\beta - Z_i\gamma\right)'\right]\\
&= y_iy_i' - y_i\mu_{q(\beta)}'X_i' - y_i\mu_{q(\gamma)}'diag(\mu_{q(\tilde{Z}_i)}) - X_i\mu_{q(\beta)}y_i' + E_{q(\theta)}\left[X_i\beta\beta'X_i'\right] + X_i\mu_{q(\beta)}\mu_{q(\gamma)}'\\
&\quad \times diag(\mu_{q(\tilde{Z}_i)}) - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}y_i' + diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\mu_{q(\beta)}'X_i' + E_{q(\theta)}\left[Z_i\gamma\gamma'Z_i'\right]\\
&= y_iy_i' - y_i\mu_{q(\beta)}'X_i' - y_i\mu_{q(\gamma)}'diag(\mu_{q(\tilde{Z}_i)}) - X_i\mu_{q(\beta)}y_i' + \left[X_i\Sigma_{q(\beta)}X_i' + X_i\mu_{q(\beta)}\mu_{q(\beta)}'X_i'\right]\\
&\quad + X_i\mu_{q(\beta)}\mu_{q(\gamma)}'diag(\mu_{q(\tilde{Z}_i)}) - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}y_i' + diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\mu_{q(\beta)}'X_i'\\
&\quad + \left(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\right) \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu_{q(\gamma)}'\right),
\end{aligned} \tag{B.5}$$

where in the last equation, we substitute the expression for $E_{q(\theta)}\left[X_i\beta\beta'X_i'\right]$ and $E_{q(\theta)}\left[Z_i\gamma\gamma'Z_i'\right]$ from equation (B.1) and equation (B.2), respectively.

We see that most of the terms in the last equation belong to the following quadratic expression,

$$\begin{aligned}
&\left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right)\left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right)'\\
&= y_iy_i' - y_i\mu_{q(\beta)}'X_i' - y_i\mu_{q(\gamma)}'diag(\mu_{q(\tilde{Z}_i)}) - X_i\mu_{q(\beta)}y_i' + X_i\mu_{q(\beta)}\mu_{q(\beta)}'X_i'\\
&\quad + X_i\mu_{q(\beta)}\mu_{q(\gamma)}'diag(\mu_{q(\tilde{Z}_i)}) - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}y_i'\\
&\quad + diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\mu_{q(\beta)}'X_i' + \left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\right) \odot \left(\mu_{q(\gamma)}\mu_{q(\gamma)}'\right),
\end{aligned} \tag{B.6}$$

where the last term utilizes the identity,

$$diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\mu_{q(\gamma)}'diag(\mu_{q(\tilde{Z}_i)}) = \left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\right) \odot \left(\mu_{q(\gamma)}\mu_{q(\gamma)}'\right).$$

Substituting equation (B.6) into equation (B.5) and rearranging terms, we get,

$$E_{q(\theta)}\left[\sum_{i=1}^{N}\left(y_i - X_i\beta - Z_i\gamma\right)\left(y_i - X_i\beta - Z_i\gamma\right)'\right]$$

[2]If $\Sigma_\varepsilon^{-1} \sim W_M\left(\nu_1, S_1\right)$, then $\Sigma_\varepsilon \sim IW_M\left(\nu_1, S_1^{-1}\right)$.

$$
\begin{aligned}
&= \sum_{i=1}^{N} \left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right) \left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right)' \\
&\quad + \sum_{i=1}^{N} X_i\Sigma_{q(\beta)}X_i' + \sum_{i=1}^{N} \left(\mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)} \odot \Sigma_{q(\gamma)}\right) \\
&\quad + \sum_{i=1}^{N} \Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right).
\end{aligned}
$$

We now make use of the following notation,

$$
\begin{aligned}
B^{-1}_{q(\Sigma)} = S_0^{-1} + \sum_{i=1}^{N} \Bigg[&\left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right) \\
&\times \left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right)' + X_i\Sigma_{q(\beta)}X_i' \\
&+ \left(\mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)}\right) \odot \Sigma_{q(\gamma)} + \Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right) \Bigg].
\end{aligned}
$$

Bringing back the term $|\Sigma_\varepsilon|^{-\frac{\nu_0+M+N+1}{2}}$, the optimal q-density for Σ_ε simplifies to,

$$
q\left(\Sigma_\varepsilon\right) \propto \exp\left[-\frac{1}{2}\text{tr}\left(B^{-1}_{q(\Sigma)}\Sigma_\varepsilon^{-1}\right)\right] \times |\Sigma_\varepsilon|^{-\frac{\nu_1+M+1}{2}},
$$

which we recognize as the kernel of an inverse Wishart distribution. Hence, the optimal q-density for Σ_ε^{-1} has the following form,

$$
\begin{aligned}
q\left(\Sigma_\varepsilon^{-1}\right) &= f_{W_M}\left(\nu_1, B_{q(\Sigma)}\right), \qquad \text{where,} \\
\nu_1 &= \nu_0 + N, \qquad \text{and} \\
B_{q(\Sigma)} &= \Bigg[S_0^{-1} + \sum_{i=1}^{N} \Bigg[\left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right) \\
&\quad \times \left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)}\right)' + X_i\Sigma_{q(\beta)}X_i' \\
&\quad + \left(\mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)}\right) \odot \Sigma_{q(\gamma)} + \Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right) \Bigg]\Bigg]^{-1}.
\end{aligned} \tag{B.7}
$$

2.4 Updates of $q(\sigma_Z^2)$, δ_1^*, and $B_{q(\sigma_Z^2)}$

We know from the Gibbs sampling, presented in Algorithm 1, that the conditional posterior for $\sigma_Z^2 \sim IG(\delta_1^*, \delta_2^*)$, where,

$$
\begin{aligned}
\delta_1^* &= \delta_1 + \frac{N.M}{2}, \quad \text{and} \\
\delta_2^* &= \delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left(\tilde{Z}_i - X_i\omega\right)'\left(\tilde{Z}_i - X_i\omega\right).
\end{aligned}
$$

The kernel of the conditional posterior density for σ_Z^2 can be written as,

$$
p(\sigma_Z^2|Z,\omega) = \exp\left[-\frac{1}{\sigma_Z^2}\left\{\delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right)\right\}\right] \times \kappa_4,
$$

where κ_4 are constants with respect to function arguments. Taking logarithms on both sides yields,

$$\log p(\sigma_Z^2|Z,\omega) = -\frac{1}{\sigma_Z^2}\left\{\delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right)\right\} + \log(\kappa_4).$$

Next, we take expectations with respect to all parameters except σ_Z^2, which yields,

$$\begin{aligned}
\log q\left(\sigma_Z^2\right) &= E_{q(\theta)}\left\{\log p(\sigma_Z^2|Z,\omega)\right\} \\
&= -\frac{1}{\sigma_Z^2}\left\{\delta_2 + \frac{1}{2}\sum_{i=1}^{N} E_{q(\theta)}\left[\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right)\right]\right\} + \log(\kappa_4) \\
&= -\frac{1}{\sigma_Z^2}\left\{B_{q(\sigma_Z^2)}\right\} + \log(\kappa_4),
\end{aligned}$$

where $B_{q(\sigma_Z^2)} = \delta_2 + \frac{1}{2}\sum_{i=1}^{N} E_{q(\theta)}\left[\left(\widetilde{Z}_i - X_i\omega\right)'\left(\widetilde{Z}_i - X_i\omega\right)\right]$. Taking expectation of the quadratic term in $B_{q(\sigma_Z^2)}$, we have,

$$\begin{aligned}
&E_{q(\theta)}\left[\left(\tilde{Z}_i - X_i\omega\right)'\left(\tilde{Z}_i - X_i\omega\right)\right] \\
&= E_{q(\theta)}\left[\tilde{Z}_i'\tilde{Z}_i\right] - E_{q(\theta)}\left[\tilde{Z}_i'X_i\omega\right] - E_{q(\theta)}\left[\omega'X_i'\tilde{Z}_i\right] + E_{q(\theta)}\left[\omega'X_i'X_i\omega\right] \\
&= \operatorname{tr}\left[E_{q(\theta)}\left[\tilde{Z}_i\tilde{Z}_i'\right]\right] - \mu'_{q\left(\tilde{Z}_i\right)}X_i\mu_{q(\omega)} - \mu'_{q(\omega)}X_i'\mu_{q\left(\tilde{Z}_i\right)} + \operatorname{tr}\left[E_{q(\theta)}\left[X_i\omega\omega'X_i'\right]\right] \\
&= \operatorname{tr}\left[\left\{\Sigma_{q\left(\tilde{Z}_i\right)} + \mu_{q\left(\tilde{Z}_i\right)}\mu'_{q\left(\tilde{Z}_i\right)}\right\}\right] - \mu'_{q\left(\tilde{Z}_i\right)}X_i\mu_{q(\omega)} - \mu'_{q(\omega)}X_i'\mu_{q\left(\tilde{Z}_i\right)} + \operatorname{tr}\left[X_i\Sigma_{q(\omega)}X_i'\right] \\
&= ||\mu_{q\left(\tilde{Z}_i\right)} - X_i\mu_{q(\omega)}||^2 + \operatorname{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)}\right].
\end{aligned}$$

Consequently, the expression for $B_{q(\sigma_Z^2)}$ simplifies to,

$$B_{q(\sigma_Z^2)} = \delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left\{\parallel \mu_{q\left(\tilde{Z}_i\right)} - X_i\mu_{q(\omega)} \parallel^2 + \operatorname{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)}\right]\right\}.$$

Therefore, the optimal q-density for σ_Z^2 can be expressed as,

$$q(\sigma_Z^2) \propto \left(\sigma_Z^2\right)^{-\left(\delta_1^*+1\right)} \exp\left[-\frac{B_{q(\sigma_Z^2)}}{\sigma_Z^2}\right],$$

which we recognize as the kernel of an inverse gamma (IG) distribution. The optimal q-density for σ_Z^2 is given by,

$$\begin{aligned}
q\left(\sigma_Z^2\right) &= f_{IG}\left(\delta_1^*, B_{q(\sigma_Z^2)}\right), \qquad \text{where,} \\
\delta_1^* &= \delta_1 + \frac{N.M}{2}, \\
B_{q(\sigma_Z^2)} &= \delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left\{||\mu_{q\left(\tilde{Z}_i\right)} - X_i\mu_{q(\omega)}||^2 + \operatorname{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)}\right]\right\}.
\end{aligned} \tag{B.8}$$

2.5 Updates of $q(\sigma_u^2), \delta_3^*$, and $B_{q(\sigma_u^2)}$

Once again, we know from the Gibbs sampling, presented in Algorithm 1, that the conditional posterior for $\sigma_u^2 \sim IG(\delta_3^*, \delta_4^*)$, where,

$$\delta_3^* = \delta_3 + \frac{N.M}{2}$$
$$\delta_4^* = \delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left(\tilde{W}_i - \tilde{Z}_i\right)'\left(\tilde{W}_i - \tilde{Z}_i\right).$$

The kernel of the conditional posterior density for σ_u^2 can be written as,

$$p(\sigma_u^2|W,Z) = \exp\left[-\frac{1}{\sigma_u^2}\left\{\delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right\}\right] \times \kappa_5,$$

where κ_5 denotes all constants with respect to function argument. Taking logarithms on both sides yields,

$$\log p(\sigma_u^2|W,Z) = -\frac{1}{\sigma_u^2}\left\{\delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right\} + \log(\kappa_5).$$

Next, taking expectations with respect to all parameters except σ_u^2 yields,

$$\begin{aligned}\log q\left(\sigma_u^2\right) &= E_{q(\theta)}\left\{\log p(\sigma_u^2|W,Z)\right\}\\ &= -\frac{1}{\sigma_u^2}\left\{\delta_4 + \frac{1}{2}\sum_{i=1}^{N}E_{q(\theta)}\left[\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right]\right\} + \log(\kappa_5)\\ &= -\frac{1}{\sigma_u^2}\left\{B_{q(\sigma_u^2)}\right\} + \log(\kappa_5),\end{aligned}$$

where $B_{q(\sigma_u^2)} = \delta_4 + \frac{1}{2}\sum_{i=1}^{N}E_{q(\theta)}\left[\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right]$. Taking expectation of the quadratic term in $B_{q(\sigma_u^2)}$, we have,

$$\begin{aligned}&E_{q(\theta)}\left[\left(\widetilde{W}_i - \widetilde{Z}_i\right)'\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right]\\ &= \widetilde{W}_i'\widetilde{W}_i - \widetilde{W}_i'E_{q(\theta)}\left[\widetilde{Z}_i\right] - E_{q(\theta)}\left[\widetilde{Z}_i'\right]\widetilde{W}_i + E_{q(\theta)}\left[\widetilde{Z}_i'\widetilde{Z}_i\right]\\ &= \widetilde{W}_i'\widetilde{W}_i - \widetilde{W}_i'\mu_{q(\widetilde{Z}_i)} - \mu'_{q(\widetilde{Z}_i)}\widetilde{W}_i + \mathrm{tr}\left[E_{q(\theta)}\left[\widetilde{Z}_i\widetilde{Z}_i'\right]\right]\\ &= \widetilde{W}_i'\widetilde{W}_i - \widetilde{W}_i'\mu_{q(\widetilde{Z}_i)} - \mu'_{q(\widetilde{Z}_i)}\widetilde{W}_i + \mathrm{tr}\left[\left\{\Sigma_{q(\widetilde{Z}_i)} + \mu_{q(\widetilde{Z}_i)}\mu'_{q(\widetilde{Z}_i)}\right\}\right]\\ &= \parallel \widetilde{W}_i - \mu_{q(\widetilde{Z}_i)} \parallel^2 + \mathrm{tr}\left[\Sigma_{q(\widetilde{Z}_i)}\right].\end{aligned}$$

Consequently, the expression for $B_{q(\sigma_u^2)}$ simplifies to,

$$B_{q(\sigma_u^2)} = \delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left\{\parallel \widetilde{W}_i - \mu_{q(\widetilde{Z}_i)} \parallel^2 + \mathrm{tr}\left[\Sigma_{q(\widetilde{Z}_i)}\right]\right\}.$$

Therefore, the optimal q-density for σ_u^2 can be written as,

$$q(\sigma_u^2) \propto \left(\sigma_u^2\right)^{-\left(\delta_3^*+1\right)} \exp\left[-\frac{B_{q(\sigma_u^2)}}{\sigma_u^2}\right],$$

which we recognize as the kernel of an inverse gamma (IG) distribution. Hence, the optimal q-density for σ_u^2 is given by,

$$\begin{aligned} q\left(\sigma_u^2\right) &= f_{IG}\left(\delta_3^*, B_{q(\sigma_u^2)}\right), \qquad \text{where,} \\ \delta_3^* &= \delta_3 + \frac{N.M}{2}, \\ B_{q(\sigma_u^2)} &= \delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left\{\| \widetilde{W}_i - \mu_{q\left(\widetilde{Z}_i\right)} \|^2 + \text{tr}\left[\Sigma_{q\left(\widetilde{Z}_i\right)}\right]\right\}. \end{aligned} \tag{B.9}$$

2.6 Updates of $q(\omega), \mu_{q(\omega)}$, and $\Sigma_{q(\omega)}$

According to the Gibbs sampling presented in Algorithm 1, the conditional posterior for $\omega \sim N_M\left(\omega_1, \Sigma_\omega\right)$ where,

$$\begin{aligned} \Sigma_\omega &= \left[\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i' X_i + O_0^{-1}\right]^{-1} \\ \omega_1 &= \Sigma_\omega\left[\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'\tilde{Z}_i + O_0^{-1}\omega_0\right]. \end{aligned}$$

The kernel of the conditional posterior density for ω can be written as,

$$\begin{aligned} p(\omega|Z,\sigma_Z^2) = \exp\Bigg[-\frac{1}{2}\Bigg\{&\omega'\left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right) - \left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}\widetilde{Z}_i'X_i + \omega_0'O_0^{-1}\right)\omega \\ &- \omega'\left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'\widetilde{Z}_i + O_0^{-1}\omega_0\right)\Bigg\}\Bigg] \times \kappa_6, \end{aligned}$$

where κ_6 are constants with respect to function argument. Taking logarithms on both sides yields,

$$\begin{aligned} \log p(\omega|Z,\sigma_Z^2) = -\frac{1}{2}\Bigg\{&\omega'\left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right)\omega - \left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N}\widetilde{Z}_i'X_i + \omega_0'O_0^{-1}\right)\omega \\ &- \omega'\left(\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'\widetilde{Z}_i + O_0^{-1}\omega_0\right)\Bigg\} + \log(\kappa_6). \end{aligned}$$

Next, we take expectation with respect to all parameters except ω, which results in the following expression,

$$\begin{aligned} \log q\left(\omega\right) =& E_{q(\theta)}\left\{\log p(\omega|Z,\sigma_Z^2)\right\} \\ =& -\frac{1}{2}\Bigg\{\omega'\left(E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right)\omega - \left(E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N} E_{q(\theta)}\left[\widetilde{Z}_i'\right]X_i + \omega_0'O_0^{-1}\right)\omega \\ &- \omega'\left(E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N} X_i'E_{q(\theta)}\left[\widetilde{Z}_i\right] + O_0^{-1}\omega_0\right)\Bigg\} + \log(\kappa_6) \end{aligned}$$

$$= -\frac{1}{2}\left\{\omega'\Sigma^{-1}_{q(\omega)}\omega - \mu'_{q(\omega)}\Sigma^{-1}_{q(\omega)}\omega - \omega'\Sigma^{-1}_{q(\omega)}\mu_{q(\omega)}\right\} + \log(\kappa_6),$$

where the last expression makes use of the following notations,

$$\Sigma^{-1}_{q(\omega)} = \left[E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right] = \left[\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right],$$

$$\begin{aligned}
\mu_{q(\omega)} &= \Sigma_{q(\omega)}\left[E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N} X_i'E_{q(\theta)}\left[\tilde{Z}_i\right] + O_0^{-1}\omega_0\right] \\
&= \Sigma_{q(\omega)}\left[E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N} X_i'\mu_{q(\tilde{Z}_i)} + O_0^{-1}\omega_0\right] \\
&= \Sigma_{q(\omega)}\left[\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\sum_{i=1}^{N} X_i'\mu_{q(\tilde{Z}_i)} + O_0^{-1}\omega_0\right].
\end{aligned}$$

Adding and subtracting $\mu'_{q(\omega)}\Sigma^{-1}_{q(\omega)}\mu_{q(\omega)}$ inside the curly braces, the square can be completed, and the optimal q-density can be expressed as follows,

$$\begin{aligned}
q(\omega) &\propto \exp\left[-\frac{1}{2}\left\{\omega'\Sigma^{-1}_{q(\omega)}\omega - \mu'_{q(\omega)}\Sigma^{-1}_{q(\omega)}\omega - \omega'\Sigma^{-1}_{q(\omega)}\mu_{q(\omega)} - \mu'_{q(\omega)}\Sigma^{-1}_{q(\omega)}\mu_{q(\omega)} + \mu'_{q(\omega)}\Sigma^{-1}_{q(\omega)}\mu_{q(\omega)}\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}(\omega - \mu_{q(\omega)})'\Sigma^{-1}_{q(\omega)}(\omega - \mu_{q(\omega)})\right],
\end{aligned}$$

where all terms not involving ω is observed in the constant of proportionality. The result is the kernel of a Gaussian or normal density and the optimal q-density for ω is given by the following,

$$\begin{aligned}
q(\omega) &= f_{N_K}\left(\mu_{q(\omega)}, \Sigma_{q(\omega)}\right), \qquad \text{where,} \\
\Sigma_{q(\omega)} &= \left[\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right]^{-1} \\
\mu_{q(\omega)} &= \Sigma_{q(\omega)}\left[\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\sum_{i=1}^{N} X_i'\mu_{q(\tilde{Z}_i)} + O_0^{-1}\omega_0\right].
\end{aligned} \tag{B.10}$$

2.7 Updates of $q(\tilde{Z}_i)$, $\mu_{q(\tilde{Z}_i)}$, and $\Sigma_{q(\tilde{Z}_i)}$

According to the Gibbs sampling presented in Algorithm 1, the conditional posterior for $\tilde{Z}_i \sim N_M(M_{1,i}, M_2)$, $\forall\, i = 1, \cdots, N$ where,

$$\begin{aligned}
M_2 &= \left[\Psi + \left(\frac{1}{\sigma_Z^2} + \frac{1}{\sigma_u^2}\right)I_M\right]^{-1} \text{ with } \Psi = \Gamma \odot \Sigma_\varepsilon^{-1},\ \Gamma = \gamma\gamma', \text{ and,} \\
M_{1,i} &= M_2\left[diag(\gamma)\Sigma_\varepsilon^{-1}(y_i - X_i\beta) + \frac{\tilde{W}_i}{\sigma_u^2} + \frac{X_i\omega}{\sigma_Z^2}\right],
\end{aligned}$$

where $\odot$ is the dot product (or Hadamard product). The kernel of the conditional posterior density for $\tilde{Z}_i$ can be written as,

$$p(\tilde{Z}_i|W, y, \Delta) = \exp\left[-\frac{1}{2}\left\{\tilde{Z}_i'M_2^{-1}\tilde{Z}_i - M_{1,i}'M_2^{-1}\tilde{Z}_i - \tilde{Z}_i'M_2^{-1}M_{1,i}\right\}\right] \times \kappa_7$$

where κ_7 are constants with respect to function arguments, and $\Delta = (\beta, \gamma, \Sigma_\varepsilon, \omega, \sigma_Z^2, \sigma_u^2)$. Taking logarithms on both sides yields,

$$\log p(\tilde{Z}_i|W, y, \Delta) = -\frac{1}{2}\left\{\tilde{Z}_i' M_2^{-1}\tilde{Z}_i - M_{1,i}' M_2^{-1}\tilde{Z}_i - \tilde{Z}_i' M_2^{-1} M_{1,i}\right\} + \log(\kappa_7)$$

Next, taking expectations with respect to all parameters except $\tilde{Z}_i$ we have,

$$\begin{aligned}
\log q\left(\tilde{Z}_i\right) &= E_{q(\theta)}\left\{\log p(\tilde{Z}_i|W, y, \Delta)\right\} \\
&= -\frac{1}{2}\Big\{\tilde{Z}_i' E_{q(\theta)}\left[M_2^{-1}\right]\tilde{Z}_i - E_{q(\theta)}\left[M_{1,i}'\right] E_{q(\theta)}\left[M_2^{-1}\right]\tilde{Z}_i \\
&- \tilde{Z}_i' E_{q(\theta)}\left[M_2^{-1}\right] E_{q(\theta)}\left[M_{1,i}\right]\Big\} + \log(\kappa_7) \\
&= -\frac{1}{2}\left\{\widetilde{Z}_i' \Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\widetilde{Z}_i - \mu'_{q\left(\widetilde{Z}_i\right)}\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\widetilde{Z}_i - \widetilde{Z}_i'\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\mu_{q\left(\widetilde{Z}_i\right)}\right\} + \log(\kappa_7),
\end{aligned}$$

where the last equation introduces two terms $\Sigma^{-1}_{q\left(\tilde{Z}_i\right)}$ and $\mu_{q\left(\tilde{Z}_i\right)}$, whose expression can be derived as shown below:

$$\begin{aligned}
\Sigma^{-1}_{q\left(\tilde{Z}_i\right)} &= E_{q(\theta)}\left[M_2^{-1}\right] = E_{q(\theta)}\left[\Psi + \left(\sigma_Z^{-2} + \sigma_u^{-2}\right) I_M\right] \\
&= E_{q(\theta)}\left[\left(\Gamma \odot \Sigma_\varepsilon^{-1}\right) + \left(\sigma_Z^{-2} + \sigma_u^{-2}\right) I_M\right] \\
&= E_{q(\theta)}\left[\left(\gamma\gamma' \odot \Sigma_\varepsilon^{-1}\right) + \left(\sigma_Z^{-2} + \sigma_u^{-2}\right) I_M\right] \\
&= E_{q(\theta)}\left[\gamma\gamma' \odot \Sigma_\varepsilon^{-1}\right] + \left(E_{q(\theta)}\left[\sigma_Z^{-2}\right] + E_{q(\theta)}\left[\sigma_u^{-2}\right]\right) I_M \\
&= E_{q(\theta)}\left[\gamma\gamma'\right] \odot E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} + \frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right) I_M \\
&= \left\{\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right\} \odot \left(\nu_1 B_{q(\Sigma)}\right) + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} + \frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right) I_M, \qquad \text{and,} \\
\mu_{q\left(\tilde{Z}_i\right)} &= E_{q(\theta)}\left[M_{1,i}\right] = \Sigma_{q\left(\tilde{Z}_i\right)} E_{q(\theta)}\left[diag(\gamma)\Sigma_\varepsilon^{-1}\left(y_i - X_i\beta\right) + \frac{\widetilde{W}_i}{\sigma_u^2} + \frac{X_i\omega}{\sigma_Z^2}\right] \\
&= \Sigma_{q\left(\tilde{Z}_i\right)}\left[E_{q(\theta)}\left[diag(\gamma)\Sigma_\varepsilon^{-1}\left(y_i - X_i\beta\right)\right] + E_{q(\theta)}\left[\sigma_u^{-2}\right]\widetilde{W}_i + E_{q(\theta)}\left[\sigma_Z^{-2}\right] X_i E_{q(\theta)}\left[\omega\right]\right] \\
&= \Sigma_{q\left(\tilde{Z}_i\right)}\Big[E_{q(\theta)}\left[diag(\gamma)\right] E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(y_i - X_i E_{q(\theta)}\left[\beta\right]\right) + E_{q(\theta)}\left[\sigma_u^{-2}\right]\widetilde{W}_i \\
&\quad + E_{q(\theta)}\left[\sigma_Z^{-2}\right] X_i E_{q(\theta)}\left[\omega\right]\Big] \\
&= \Sigma_{q\left(\tilde{Z}_i\right)}\left[diag(\mu_{q(\gamma)})\left(\nu_1 B_{q(\Sigma)}\right)\left(y_i - X_i\mu_{q(\beta)}\right) + \left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\widetilde{W}_i + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right) X_i\mu_{q(\omega)}\right].
\end{aligned}$$

Consequently, the optimal q-density for $\widetilde{Z}_i$ can be written as,

$$q\left(\widetilde{Z}_i\right) \propto \exp\left[-\frac{1}{2}\left\{\widetilde{Z}_i'\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\widetilde{Z}_i - \mu'_{q\left(\widetilde{Z}_i\right)}\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\widetilde{Z}_i - \widetilde{Z}_i'\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\mu_{q\left(\widetilde{Z}_i\right)}\right\}\right].$$

Adding and subtracting $\mu'_{q\left(\widetilde{Z}_i\right)}\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\mu_{q\left(\widetilde{Z}_i\right)}$ inside the curly braces, the square can be completed as follows,

$$
\begin{aligned}
q\left(\widetilde{Z}_i\right) &\propto \exp\left[-\frac{1}{2}\left\{\begin{array}{c} \widetilde{Z}_i'\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\widetilde{Z}_i - \mu'_{q\left(\widetilde{Z}_i\right)}\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\widetilde{Z}_i - \widetilde{Z}_i'\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\mu_{q\left(\widetilde{Z}_i\right)} \\ -\mu'_{q\left(\widetilde{Z}_i\right)}\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\mu_{q\left(\widetilde{Z}_i\right)} + \mu'_{q\left(\widetilde{Z}_i\right)}\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}\mu_{q\left(\widetilde{Z}_i\right)} \end{array}\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}(\widetilde{Z}_i - \mu_{q\left(\widetilde{Z}_i\right)})'\Sigma^{-1}_{q\left(\widetilde{Z}_i\right)}(\widetilde{Z}_i - \mu_{q\left(\widetilde{Z}_i\right)})\right].
\end{aligned}
$$

The result is the kernel of a Gaussian or normal density and the optimal q-density for $\widetilde{Z}_i$ is given by,

$$
\begin{aligned}
q\left(\tilde{Z}_i\right) &= f_{N_M}\left(\mu_{q\left(\tilde{Z}_i\right)}, \Sigma_{q\left(\tilde{Z}_i\right)}\right), \qquad \text{where,} \\
\Sigma_{q\left(\tilde{Z}_i\right)} &= \left[\left\{\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right\} \odot \left(\nu_1 B_{q(\Sigma)}\right) + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} + \frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right) I_M\right]^{-1} \\
\mu_{q\left(\tilde{Z}_i\right)} &= \Sigma_{q\left(\tilde{Z}_i\right)}\Bigg[diag(\mu_{q(\gamma)})\left(\nu_1 B_{q(\Sigma)}\right)\left(y_i - X_i\mu_{q(\beta)}\right) + \left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\widetilde{W}_i \\
&\quad + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right) X_i\mu_{q(\omega)}\Bigg].
\end{aligned} \tag{B.11}
$$

3 Appendix C - Derivation of the Evidence Lower Bound (ELBO)

The evidence lower bound (ELBO) is given by the equation below,

$$\ell = E_{q(\theta)}\left[\ln\left(\frac{p(\theta, y)}{q(\theta)}\right)\right] = T_1 + T_2 + T_3 \tag{C.1}$$

where,

$$T_1 = E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^{N} p\left(y_i|\beta,\gamma,\Sigma_\varepsilon^{-1}\right) p(\beta)\, p(\gamma)\, p\left(\Sigma_\varepsilon^{-1}\right)}{q(\beta)\, q(\gamma)\, q\left(\Sigma_\varepsilon^{-1}\right)}\right)\right] \tag{C.2}$$

$$T_2 = E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^{N} p\left(\tilde{Z}_i|\omega,\sigma_Z^2\right) p(\omega)\, p\left(\sigma_Z^2\right)}{q(\omega)\, q\left(\sigma_Z^2\right)}\right)\right] \tag{C.3}$$

$$T_3 = E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^{N} p\left(\widetilde{W}_i|\tilde{Z}_i,\sigma_u^2\right) p\left(\sigma_u^2\right)}{\prod_{i=1}^{N} q\left(\tilde{Z}_i\right) q\left(\sigma_u^2\right)}\right)\right]. \tag{C.4}$$

3.1 Derivation of T_1

On the terms that constitute T_1, we have the following information:

$$\begin{aligned}
p\left(y|\beta,\gamma,\Sigma_\varepsilon^{-1}\right) &= \prod_{i=1}^{N} p\left(y_i|\beta,\gamma,\Sigma_\varepsilon^{-1}\right)\\
p\left(y_i|\beta,\gamma,\Sigma_\varepsilon^{-1}\right) &\sim N_M\left(X_i\beta + Z_i\gamma, \Sigma_\varepsilon\right)\\
p(\beta) &\sim N_M\left(\beta_0, B_0\right)\\
p(\gamma) &\sim N_M\left(\gamma_0, G_0\right)\\
p(\Sigma_\varepsilon^{-1}) &\sim W_M\left(\nu_0, S_0\right)\\
q(\beta) &\sim N(\mu_{q(\beta)}, \Sigma_{q(\beta)})\\
q(\gamma) &\sim N(\mu_{q(\gamma)}, \Sigma_{q(\gamma)})\\
q(\Sigma_\varepsilon^{-1}) &\sim W(\nu_1, B_{q(\Sigma)}).
\end{aligned}$$

First, we focus on the terms in the numerator and begin with the expectation of the likelihood. The response variable y follows a normal distribution, so the resulting likelihood has the expression,

$$\begin{aligned}
&\ln p\left(y|\beta,\gamma,\Sigma_\varepsilon^{-1}\right)\\
&= \sum_{i=1}^{N}\left\{-\frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_\varepsilon| - \frac{1}{2}(y_i - X_i\beta - Z_i\gamma)'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma)\right\}\\
&= -\frac{MN}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma_\varepsilon| - \frac{1}{2}\sum_{i=1}^{N}(y_i - X_i\beta - Z_i\gamma)'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma).
\end{aligned}$$

Taking expectation over the likelihood, we have,

$$\begin{aligned}
&E_{q(\theta)}\left[\ln p\left(y|\beta,\gamma,\Sigma_\varepsilon^{-1}\right)\right]\\
&= -\frac{MN}{2}\ln(2\pi) - \frac{N}{2}E_{q(\theta)}\left[\ln|\Sigma_\varepsilon|\right] - \frac{1}{2}\sum_{i=1}^{N}E_{q(\theta)}\left[(y_i - X_i\beta - Z_i\gamma)'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma)\right]
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{MN}{2}\ln(2\pi) + \frac{N}{2}\left[\ln\left|B_{q(\Sigma)}\right| + M\ln 2 + \psi_M\left(\frac{\nu_1}{2}\right)\right] \\
&\quad - \frac{1}{2}\sum_{i=1}^{N} E_{q(\theta)}\left[(y_i - X_i\beta - Z_i\gamma)'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma)\right].
\end{aligned} \tag{C.5}
$$

where we have used the equality $E_{q(\theta)}[\ln|\Sigma_\varepsilon|] = -\ln\left|B_{q(\Sigma)}\right| - M\ln 2 - \psi_M\left(\frac{\nu_1}{2}\right)$ and note that $\psi(M)(\cdot)$ is the multivariate digamma function. Focusing on the last term in equation (C.5) that involves the expectation, we have,

$$
\begin{aligned}
&E_{q(\theta)}\left[(y_i - X_i\beta - Z_i\gamma)'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma)\right] \\
&= E_{q(\theta)}\left[(y_i - X_i\beta - (\tilde{Z}_i \odot \gamma))'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - (\tilde{Z}_i \odot \gamma))\right] \\
&= E_{q(\theta)}\left[y_i'\Sigma_\varepsilon^{-1}y_i\right] - E_{q(\theta)}\left[y_i'\Sigma_\varepsilon^{-1}X_i\beta\right] - E_{q(\theta)}\left[y_i'\Sigma_\varepsilon^{-1}\left(\tilde{Z}_i \odot \gamma\right)\right] - E_{q(\theta)}\left[\beta'X_i'\Sigma_\varepsilon^{-1}y_i\right] \\
&\quad + E_{q(\theta)}\left[\beta'X_i'\Sigma_\varepsilon^{-1}X_i\beta\right] + E_{q(\theta)}\left[\beta'X_i'\Sigma_\varepsilon^{-1}\left(\tilde{Z}_i \odot \gamma\right)\right] - E_{q(\theta)}\left[\left(\tilde{Z}_i \odot \gamma\right)'\Sigma_\varepsilon^{-1}y_i\right] \\
&\quad + E_{q(\theta)}\left[\left(\tilde{Z}_i \odot \gamma\right)'\Sigma_\varepsilon^{-1}X_i\beta\right] + E_{q(\theta)}\left[\left(\tilde{Z}_i \odot \gamma\right)'\Sigma_\varepsilon^{-1}\left(\tilde{Z}_i \odot \gamma\right)\right] \\
&= y_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]y_i - y_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_iE_{q(\theta)}[\beta] - y_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(E_{q(\theta)}\left[\tilde{Z}_i\right] \odot E_{q(\theta)}[\gamma]\right) \\
&\quad - E_{q(\theta)}[\beta']X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]y_i + \operatorname{tr}\left[E_{q(\theta)}\left[\beta\beta'X_i'\Sigma_\varepsilon^{-1}X_i\right]\right] \\
&\quad + E_{q(\theta)}[\beta']X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(E_{q(\theta)}\left[\tilde{Z}\right] \odot E_{q(\theta)}[\gamma]\right) - \left(E_{q(\theta)}[\gamma'] \odot E_{q(\theta)}[Z_i']\right)E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]y_i \\
&\quad + \left(E_{q(\theta)}[\gamma'] \odot E_{q(\theta)}[Z_i']\right)E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_iE_{q(\theta)}[\beta] + \operatorname{tr}\left[E_{q(\theta)}\left[Z_i\gamma\gamma'Z_i'\Sigma_\varepsilon^{-1}\right]\right].
\end{aligned} \tag{C.6}
$$

There are two components in equation (C.6) that involves the trace operator. We simplify these two terms below.

$$
\begin{aligned}
&\operatorname{tr}\left[E_{q(\theta)}\left[\beta\beta'X_i'\Sigma_\varepsilon^{-1}X_i\right]\right] \\
&= \operatorname{tr}\left[E_{q(\theta)}[\beta\beta']X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_i\right] \\
&= \operatorname{tr}\left[\Sigma_{q(\beta)}X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_i\right] + \operatorname{tr}\left[\mu_{q(\beta)}\mu_{q(\beta)}'X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_i\right] \\
&= \operatorname{tr}\left[\Sigma_{q(\beta)}X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_i\right] + \mu_{q(\beta)}'X_i'E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]X_i\mu_{q(\beta)}.
\end{aligned} \tag{C.7}
$$

$$
\begin{aligned}
&\operatorname{tr}\left[E_{q(\theta)}\left[Z_i\gamma\gamma'Z_i'\Sigma_\varepsilon^{-1}\right]\right] \\
&= \operatorname{tr}\left[E_{q(\theta)}\left[\left(\tilde{Z}_i\tilde{Z}_i' \odot \gamma\gamma'\right)\Sigma_\varepsilon^{-1}\right]\right] \\
&= \operatorname{tr}\left[\left(E_{q(\theta)}\left[\tilde{Z}_i\tilde{Z}_i'\right] \odot E_{q(\theta)}[\gamma\gamma']\right)E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] \\
&= \operatorname{tr}\left[\left(\left(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\right) \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu_{q(\gamma)}'\right)\right)E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] \\
&= \operatorname{tr}\left[\left(\Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu_{q(\gamma)}'\right)\right)E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] + \operatorname{tr}\Big[\left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}' \odot \Sigma_{q(\gamma)}\right) \\
&\quad \times E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\Big] + \operatorname{tr}\left[\left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}' \odot \mu_{q(\gamma)}\mu_{q(\gamma)}'\right)E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right].
\end{aligned} \tag{C.8}
$$

Using the property for Hadamard product,

$$
AA' \odot BB' = (A \odot B)(B' \odot A') \text{ and } A \odot (B + C) = A \odot B + A \odot C,
$$

the last term in equation (C.8) can be written as below,

$$\begin{aligned}
&\operatorname{tr}\left[\left(\mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\mu'_{q(\gamma)}\right) E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] \\
&= \operatorname{tr}\left[\left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right)\left(\mu'_{q(\gamma)} \odot \mu'_{q(\tilde{Z}_i)}\right) E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] \\
&= \operatorname{tr}\left[\left(\mu'_{q(\gamma)} \odot \mu'_{q(\tilde{Z}_i)}\right) E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right)\right] \\
&= \left(\mu'_{q(\gamma)} \odot \mu'_{q(\tilde{Z}_i)}\right) E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right) \\
&= \left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right)' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right). \qquad \text{(C.9)}
\end{aligned}$$

Substituting equations (C.7), (C.8) and (C.9) into equation (C.6), and collecting terms yields the following expression,

$$\begin{aligned}
&E_{q(\theta)}\left[(y_i - X_i\beta - Z_i\gamma)'\Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma)\right] \\
&= \left(y_i - X_i\mu_{q(\beta)} - \left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right)\right)' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\left(y_i - X_i\mu_{q(\beta)} - \left(\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)}\right)\right) \\
&+ \operatorname{tr}\left[\Sigma_{q(\beta)}X_i' E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right] X_i\right] + \operatorname{tr}\left[\left(\Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right)\right) E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] \\
&+ \operatorname{tr}\left[\left(\mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)} \odot \Sigma_{q(\gamma)}\right) E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right]. \qquad \text{(C.10)}
\end{aligned}$$

Substituting equation (C.10) into equation (C.5), we have the expected value of the log-likelihood as follows:

$$\begin{aligned}
E_{q(\theta)}\left[\ln p\left(y|\beta,\gamma,\Sigma_\varepsilon^{-1}\right)\right] &= -\frac{MN}{2}\ln(2\pi) + \frac{N}{2}\left\{\ln\left|B_{q(\Sigma)}\right| + M\ln 2 + \psi_M\left(\frac{\nu_1}{2}\right)\right\} \\
&- \frac{1}{2}\sum_{i=1}^{N}\left\{\begin{array}{c}
\left(y_i - X_i\mu_{q(\beta)} - (\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)})\right)'\left(\nu_1 B_{q(\Sigma)}\right) \\
\times\left(y_i - X_i\mu_{q(\beta)} - (\mu_{q(\tilde{Z}_i)} \odot \mu_{q(\gamma)})\right) + \operatorname{tr}\left[\Sigma_{q(\beta)}X_i'\left(\nu_1 B_{q(\Sigma)}\right)X_i\right] \\
+\operatorname{tr}\left[\left(\Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu'_{q(\gamma)}\right)\right)\left(\nu_1 B_{q(\Sigma)}\right)\right] \\
+\operatorname{tr}\left[\left(\mu_{q(\tilde{Z}_i)}\mu'_{q(\tilde{Z}_i)} \odot \Sigma_{q(\gamma)}\right)\left(\nu_1 B_{q(\Sigma)}\right)\right]
\end{array}\right\}. \qquad \text{(C.11)}
\end{aligned}$$

Next, we work on the terms involving the prior distributions. We know $\beta \sim N(\beta_0, B_0)$, so the resulting logarithm of the pdf is,

$$\ln p(\beta) = -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|B_0| - \frac{1}{2}(\beta - \beta_0)'B_0^{-1}(\beta - \beta_0).$$

Taking expectation of the logarithm of the pdf yields,

$$\begin{aligned}
&E_{q(\theta)}\left[\ln p(\beta)\right] = -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|B_0| - \frac{1}{2}E_{q(\theta)}\left[(\beta - \beta_0)'B_0^{-1}(\beta - \beta_0)\right] \\
&= -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|B_0| - \frac{1}{2}E_{q(\theta)}\left[\operatorname{tr}\left[\beta\beta'B_0^{-1}\right]\right] - \mu'_{q(\beta)}B_0^{-1}\beta_0 - \beta_0'B_0^{-1}\mu_{q(\beta)} + \beta_0'B_0^{-1}\beta_0 \\
&= -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|B_0| - \frac{1}{2}\left\{(\mu_{q(\beta)} - \beta_0)'B_0^{-1}(\mu_{q(\beta)} - \beta_0) + \operatorname{tr}\left[\Sigma_{q(\beta)}B_0^{-1}\right]\right\} \qquad \text{(C.12)}
\end{aligned}$$

Similar to β, $\gamma \sim N(\gamma_0, G_0)$. Hence, the expectation of the ln(pdf) can be obtained to be the following expression,

$$E_{q(\theta)}\left[\ln p\left(\gamma\right)\right]=-\frac{M}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left|G_0\right|-\frac{1}{2}\Big\{(\mu_{q(\gamma)}-\gamma_0)'G_0^{-1}(\mu_{q(\gamma)}-\gamma_0)$$
$$+\operatorname{tr}\left[\Sigma_{q(\gamma)}G_0^{-1}\right]\Big\}. \tag{C.13}$$

The prior distribution for Σ_ε^{-1} is a Wishart distribution with degrees of freedom ν_0 and scale matrix S_0, i.e., $\Sigma_\varepsilon^{-1}\sim W_M\left(\nu_0,S_0\right)$. Therefore, the pdf of Σ_ε^{-1} has the following expression,

$$p\left(\Sigma_\varepsilon^{-1}\right)=\frac{1}{2^{\frac{\nu_0 M}{2}}\Gamma_M(\frac{\nu_0}{2})}\left|S_0^{-1}\right|^{\frac{\nu_0}{2}}\left|\Sigma_\varepsilon^{-1}\right|^{\frac{(\nu_0-M-1)}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left[S_0^{-1}\Sigma_\varepsilon^{-1}\right]\right],$$

and logarithm of the pdf has the expression,

$$\ln p\left(\Sigma_\varepsilon^{-1}\right)=-\frac{\nu_0 M}{2}\ln\left(2\right)-\ln\Gamma_M\left(\frac{\nu_0}{2}\right)+\frac{\nu_0}{2}\ln\left|S_0^{-1}\right|+\frac{(\nu_0-M-1)}{2}\ln\left|\Sigma_\varepsilon^{-1}\right|-\frac{1}{2}\operatorname{tr}\left[S_0^{-1}\Sigma_\varepsilon^{-1}\right].$$

Taking expectation over $\ln p\left(\Sigma_\varepsilon^{-1}\right)$, we get the following,

$$\begin{aligned}
&E_{q(\theta)}\left[\ln p\left(\Sigma_\varepsilon^{-1}\right)\right]\\
&=-\frac{\nu_0 M}{2}\ln\left(2\right)-\ln\Gamma_M(\frac{\nu_0}{2})+\frac{\nu_0}{2}\ln\left|S_0^{-1}\right|-\frac{(\nu_0-M-1)}{2}E_{q(\theta)}\left[\ln\left|\Sigma_\varepsilon\right|\right]\\
&\quad-\frac{1}{2}\operatorname{tr}\left[S_0^{-1}E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right]\\
&=-\frac{\nu_0 M}{2}\ln\left(2\right)-\ln\Gamma_M\left(\frac{\nu_0}{2}\right)+\frac{\nu_0}{2}\ln\left|S_0^{-1}\right|\\
&\quad+\frac{(\nu_0-M-1)}{2}\left(\ln\left|B_{q(\Sigma)}\right|+M\ln\left(2\right)+\psi_M\left(\frac{\nu_1}{2}\right)\right)-\frac{1}{2}\operatorname{tr}\left[S_0^{-1}\nu_1 B_{q(\Sigma)}\right]\\
&=-\frac{M\left(M+1\right)}{2}\ln\left(2\right)-\ln\Gamma_M\left(\frac{\nu_0}{2}\right)+\frac{\nu_0}{2}\ln\left|S_0^{-1}\right|+\frac{(\nu_0-M-1)}{2}\ln\left|B_{q(\Sigma)}\right|\\
&\quad+\frac{(\nu_0-M-1)}{2}\psi_M\left(\frac{\nu_1}{2}\right)-\frac{1}{2}\operatorname{tr}\left[S_0^{-1}\nu_1 B_{q(\Sigma)}\right].
\end{aligned} \tag{C.14}$$

Finally, we look at the optimal densities which appears in the denominator of T_1. We know that $q(\beta)\sim N(\mu_{q(\beta)},\Sigma_{q(\beta)})$, so the logarithm of the optimal density is,

$$\ln q\left(\beta\right)=-\frac{K}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left|\Sigma_{q(\beta)}\right|-\frac{1}{2}(\beta-\mu_{q(\beta)})'\Sigma_{q(\beta)}^{-1}(\beta-\mu_{q(\beta)}).$$

Taking expectation of $\ln q(\beta)$ we obtain,

$$\begin{aligned}
E_{q(\theta)}\left[\ln q\left(\beta\right)\right]&=-\frac{K}{2}\ln\left(2\pi\right)-\frac{1}{2}\left[\ln\left|\Sigma_{q(\beta)}\right|\right]-\frac{1}{2}E_{q(\theta)}\left[(\beta-\mu_{q(\beta)})'\Sigma_{q(\beta)}^{-1}(\beta-\mu_{q(\beta)})\right]\\
&=-\frac{K}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left|\Sigma_{q(\beta)}\right|-\frac{K}{2},
\end{aligned} \tag{C.15}$$

where the last line makes use of the following equality,

$$\begin{aligned}
&E_{q(\theta)}\left[(\beta-\mu_{q(\beta)})'\Sigma_{q(\beta)}^{-1}(\beta-\mu_{q(\beta)})\right]\\
&=\left(E_{q(\theta)}\left[\beta\right]-\mu_{q(\beta)}\right)'\Sigma_{q(\beta)}^{-1}\left(E_{q(\theta)}\left[\beta\right]-\mu_{q(\beta)}\right)+\operatorname{tr}\left[\Sigma_{q(\beta)}^{-1}\Sigma_{q(\beta)}\right]\\
&=0+\operatorname{tr}\left[I_K\right]=K.
\end{aligned}$$

Similar to $q(\beta)$, we know $q(\gamma) \sim N(\mu_{q(\gamma)}, \Sigma_{q(\gamma)})$. Proceeding analogous to $q(\beta)$, we can obtain,

$$E_{q(\theta)}\left[\ln q(\gamma)\right] = -\frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\Sigma_{q(\gamma)}\right| - \frac{M}{2}. \tag{C.16}$$

Lastly, the optimal density $q(\Sigma_\varepsilon^{-1}) \sim W(\nu_1, B_{q(\Sigma)})$. Hence, the logarithm of the optimal density is the following:

$$\begin{aligned}\ln q\left(\Sigma_\varepsilon^{-1}\right) = &-\frac{\nu_1 M}{2}\ln(2) - \ln\Gamma_M\left(\frac{\nu_1}{2}\right) - \frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right| - \frac{(\nu_1 - M - 1)}{2}\ln\left|\Sigma_\varepsilon\right| \\ &- \frac{1}{2}\text{tr}\left[B_{q(\Sigma)}^{-1}\Sigma_\varepsilon^{-1}\right].\end{aligned}$$

Taking expectation of the above quantity, we can write,

$$\begin{aligned}&E_{q(\theta)}\left[\ln q\left(\Sigma_\varepsilon^{-1}\right)\right] \\ &= -\frac{\nu_1 M}{2}\ln(2) - \ln\Gamma_M(\frac{\nu_1}{2}) - \frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right| \\ &\quad + \frac{(\nu_1 - M - 1)}{2}\left(\ln\left|B_{q(\Sigma)}\right| + M\ln(2) + \psi_M\left(\frac{\nu_1}{2}\right)\right) - \frac{1}{2}\text{tr}\left[B_{q(\Sigma)}^{-1}E_{q(\theta)}\left[\Sigma_\varepsilon^{-1}\right]\right] \\ &= -\frac{\nu_1 M}{2}\ln(2) - \ln\Gamma_M\left(\frac{\nu_1}{2}\right) - \frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right| \\ &\quad + \frac{(\nu_1 - M - 1)}{2}\left(\ln\left|B_{q(\Sigma)}\right| + M\ln(2) + \psi_M\left(\frac{\nu_1}{2}\right)\right) - \frac{1}{2}\text{tr}\left[B_{q(\Sigma)}^{-1}\nu_1 B_{q(\Sigma)}\right] \\ &= -\frac{M(M+1)}{2}\ln(2) - \ln\Gamma_M\left(\frac{\nu_1}{2}\right) - \frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right| + \frac{(\nu_1 - M - 1)}{2}\ln\left|B_{q(\Sigma)}\right| \\ &\quad + \frac{(\nu_1 - M - 1)}{2}\psi_M\left(\frac{\nu_1}{2}\right) - \frac{\nu_1 M}{2}.\end{aligned} \tag{C.17}$$

We go back to the expression for T_1 and take logarithm of each terms as follows:

$$\begin{aligned}T_1 &= E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^N p\left(y_i|\beta,\gamma,\Sigma_\varepsilon^{-1}\right)p(\beta)\,p(\gamma)\,p\left(\Sigma_\varepsilon^{-1}\right)}{q(\beta)\,q(\gamma)\,q\left(\Sigma_\varepsilon^{-1}\right)}\right)\right] \\ &= E_{q(\theta)}\left[\ln p\left(y|\beta,\gamma,\Sigma_\varepsilon^{-1}\right)\right] + E_{q(\theta)}\left[\ln p(\beta)\right] + E_{q(\theta)}\left[\ln p(\gamma)\right] + E_{q(\theta)}\left[\ln p\left(\Sigma_\varepsilon^{-1}\right)\right] \\ &\quad - E_{q(\theta)}\left[\ln q(\beta)\right] - E_{q(\theta)}\left[\ln q(\gamma)\right] - E_{q(\theta)}\left[\ln q\left(\Sigma_\varepsilon^{-1}\right)\right].\end{aligned} \tag{C.18}$$

Substituting the expressions from equations (C.11), (C.12), (C.13), (C.14), (C.15), (C.16) and (C.17) into equation (C.18), we can write:

$$\begin{aligned}T_1 = &-\frac{MN}{2}\ln(2\pi) + \frac{N}{2}\ln\left|B_{q(\Sigma)}\right| + \frac{MN}{2}\ln(2) + \frac{N}{2}\psi_M\left(\frac{\nu_1}{2}\right) - \frac{K}{2}\ln(2\pi) \\ &- \frac{1}{2}\ln|B_0| - \frac{1}{2}\left\{(\mu_{q(\beta)} - \beta_0)'B_0^{-1}(\mu_{q(\beta)} - \beta_0) + \text{tr}\left[\Sigma_{q(\beta)}B_0^{-1}\right]\right\} - \frac{M}{2}\ln(2\pi) \\ &- \frac{1}{2}\ln|G_0| - \frac{1}{2}\left\{(\mu_{q(\gamma)} - \gamma_0)'G_0^{-1}(\mu_{q(\gamma)} - \gamma_0) + \text{tr}\left[\Sigma_{q(\gamma)}G_0^{-1}\right]\right\} - \frac{M(M+1)}{2}\ln(2) \\ &- \ln\Gamma_M\left(\frac{\nu_0}{2}\right) + \frac{\nu_0}{2}\ln\left|S_0^{-1}\right| + \frac{(\nu_0 - M - 1)}{2}\ln\left|B_{q(\Sigma)}\right| + \frac{(\nu_0 - M - 1)}{2}\psi_M\left(\frac{\nu_1}{2}\right) \\ &- \frac{1}{2}\text{tr}\left[S_0^{-1}\nu_1 B_{q(\Sigma)}\right] + \frac{K}{2}\ln(2\pi) + \frac{1}{2}\left[\ln\left|\Sigma_{q(\beta)}\right|\right] + \frac{K}{2} + \frac{M}{2}\ln(2\pi) + \frac{1}{2}\left[\ln\left|\Sigma_{q(\gamma)}\right|\right]\end{aligned}$$

$$+\frac{M}{2}+\frac{M(M+1)}{2}\ln(2)+\ln\Gamma_M\left(\frac{\nu_1}{2}\right)+\frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right|-\frac{(\nu_1-M-1)}{2}\ln\left|B_{q(\Sigma)}\right|$$
$$-\frac{(\nu_1-M-1)}{2}\psi_M\left(\frac{\nu_1}{2}\right)+\frac{\nu_1 M}{2}$$
$$-\frac{1}{2}\sum_{i=1}^{N}\left\{\begin{array}{c}\left(y_i-X_i\mu_{q(\beta)}-\left(\mu_{q(\tilde{Z}_i)}\odot\mu_{q(\gamma)}\right)\right)'\left(\nu_1 B_{q(\Sigma)}\right)\times\\ \left(y_i-X_i\mu_{q(\beta)}-\left(\mu_{q(\tilde{Z}_i)}\odot\mu_{q(\gamma)}\right)\right)+\mathrm{tr}\left[\Sigma_{q(\beta)}X_i'\left(\nu_1 B_{q(\Sigma)}\right)X_i\right]\\ +\mathrm{tr}\left[\left(\Sigma_{q(\tilde{Z}_i)}\odot\left(\Sigma_{q(\gamma)}+\mu_{q(\gamma)}\mu_{q(\gamma)}'\right)\right)\left(\nu_1 B_{q(\Sigma)}\right)\right]\\ +\mathrm{tr}\left[\left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\odot\Sigma_{q(\gamma)}\right)\left(\nu_1 B_{q(\Sigma)}\right)\right]\end{array}\right\}.$$

Canceling terms, we finally obtain the following expression:

$$\begin{aligned}T_1 = &-\frac{MN}{2}\ln(2\pi)+\frac{MN}{2}\ln(2)+\frac{K}{2}+\frac{M}{2}+\frac{\nu_1 M}{2}-\frac{1}{2}\ln|B_0|-\frac{1}{2}\ln|G_0|\\ &+\frac{\nu_0}{2}\ln\left|S_0^{-1}\right|+\frac{1}{2}\ln\left|\Sigma_{q(\beta)}\right|+\frac{1}{2}\ln\left|\Sigma_{q(\gamma)}\right|+\frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right|-\frac{1}{2}\mathrm{tr}\left[\Sigma_{q(\beta)}B_0^{-1}\right]\\ &-\frac{1}{2}\mathrm{tr}\left[\Sigma_{q(\gamma)}G_0^{-1}\right]-\frac{1}{2}\mathrm{tr}\left[S_0^{-1}\nu_1 B_{q(\Sigma)}\right]-\ln\Gamma_M\left(\frac{\nu_0}{2}\right)+\ln\Gamma_M(\frac{\nu_1}{2})\\ &-\frac{1}{2}\left\{(\mu_{q(\beta)}-\beta_0)'B_0^{-1}(\mu_{q(\beta)}-\beta_0)\right\}-\frac{1}{2}\left\{(\mu_{q(\gamma)}-\gamma_0)'G_0^{-1}(\mu_{q(\gamma)}-\gamma_0)\right\}\\ &-\frac{1}{2}\sum_{i=1}^{N}\left\{\begin{array}{c}\left(y_i-X_i\mu_{q(\beta)}-\left(\mu_{q(\tilde{Z}_i)}\odot\mu_{q(\gamma)}\right)\right)'\left(\nu_1 B_{q(\Sigma)}\right)\times\\ \left(y_i-X_i\mu_{q(\beta)}-\left(\mu_{q(\tilde{Z}_i)}\odot\mu_{q(\gamma)}\right)\right)+\mathrm{tr}\left[\Sigma_{q(\beta)}X_i'\left(\nu_1 B_{q(\Sigma)}\right)X_i\right]\\ +\mathrm{tr}\left[\left(\Sigma_{q(\tilde{Z}_i)}\odot\left(\Sigma_{q(\gamma)}+\mu_{q(\gamma)}\mu_{q(\gamma)}'\right)\right)\left(\nu_1 B_{q(\Sigma)}\right)\right]\\ +\mathrm{tr}\left[\left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}'\odot\Sigma_{q(\gamma)}\right)\left(\nu_1 B_{q(\Sigma)}\right)\right]\end{array}\right\}.\end{aligned} \tag{C.19}$$

3.2 Derivation of T_2

On terms that constitute,

$$T_2 = E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^{N}p\left(\tilde{Z}_i|\omega,\sigma_Z^2\right)p(\omega)\,p(\sigma_Z^2)}{q(\omega)\,q(\sigma_Z^2)}\right)\right]$$

we have the following information,

$$\begin{aligned}p(\tilde{Z}_i|\omega,\sigma_Z^2) &\sim N_M\left(X_i\omega,\sigma_z^2 I_M\right)\\ p(\omega) &\sim N(\omega_0,A_0)\\ p(\sigma_Z^2) &\sim IG(\delta_1,\delta_2)\\ q(\omega) &\sim N_M\left(\mu_{q(\omega)},\Sigma_{q(\omega)}\right)\\ q\left(\sigma_Z^2\right) &\sim IG\left(\delta_1^*,B_{q(\sigma_Z^2)}\right).\end{aligned}$$

We begin by working on the terms in the numerator of T_2. Taking logarithm on the density of $\tilde{Z}_i$ we get,

$$\ln p(\tilde{Z}_i|\omega,\sigma_Z^2) = -\frac{M}{2}\ln(2\pi)-\frac{M}{2}\ln\sigma_Z^2-\frac{1}{2\sigma_Z^2}(\tilde{Z}_i-X_i\omega)'(\tilde{Z}_i-X_i\omega).$$

Taking expectation of the sum of the above term yields,

$$\begin{aligned}
&E_{q(\theta)}\left[\sum_{i=1}^{N}\ln p(\tilde{Z}_i|\omega,\sigma_Z^2)\right]\\
&=-\frac{MN}{2}\left[\ln(2\pi)+E_{q(\theta)}\left[\ln\sigma_Z^2\right]\right]-\frac{1}{2}E_{q(\theta)}\left[\sigma_Z^{-2}\right]\sum_{i=1}^{N}E_{q(\theta)}\left[(\tilde{Z}_i-X_i\omega)'(\tilde{Z}_i-X_i\omega)\right]\\
&=-\frac{MN}{2}\left[\ln(2\pi)+B_{q(\sigma_Z^2)}-\psi\left(\delta_1^*\right)\right]-\frac{1}{2}\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\sum_{i=1}^{N}E_{q(\theta)}\left[(\tilde{Z}_i-X_i\omega)'(\tilde{Z}_i-X_i\omega)\right]\\
&=-\frac{MN}{2}\left[\ln(2\pi)+B_{q(\sigma_Z^2)}-\psi\left(\delta_1^*\right)\right]\\
&-\frac{1}{2}\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\sum_{i=1}^{N}\left\{\parallel\mu_{q(\tilde{Z}_i)}-X_i\mu_{q(\omega)}\parallel^2+\mathrm{tr}\left[\Sigma_{q(\tilde{Z}_i)}+X_i\Sigma_{q(\omega)}X_i'\right]\right\},
\end{aligned}\tag{C.20}$$

where in the second equality we have used $E_{q(\theta)}\left[\ln\sigma_Z^2\right]=B_{q(\sigma_Z^2)}-\psi\left(\delta_1^*\right)$, and the last expression follows from Section 2.4.

Next, we work on the terms involving the prior distributions. We know $\omega\sim N(\omega_0,A_0)$, so the logarithm of the pdf is,

$$\ln p(\omega)=-\frac{K}{2}\ln(2\pi)-\frac{1}{2}\ln|O_0|-\frac{1}{2}(\omega-\omega_0)'O_0^{-1}(\omega-\omega_0).$$

Taking expectation of the above term we get,

$$\begin{aligned}
&E_{q(\theta)}\left[\ln p(\omega)\right]=-\frac{K}{2}\ln(2\pi)-\frac{1}{2}\ln|O_0|-\frac{1}{2}E_{q(\theta)}\left[(\omega-\omega_0)'O_0^{-1}(\omega-\omega_0)\right]\\
&=-\frac{K}{2}\ln(2\pi)-\frac{1}{2}\ln|O_0|-\frac{1}{2}\left\{(\mu_{q(\omega)}-\omega_0)'O_0^{-1}(\mu_{q(\omega)}-\omega_0)+\mathrm{tr}\left[\Sigma_{q(\omega)}O_0^{-1}\right]\right\},
\end{aligned}\tag{C.21}$$

where the expectation of the quadratic term can be taken analogous to the derivation in equation (C.12).

The prior distribution for σ_Z^2 is an inverse gamma distribution, so $\sigma_Z^2\sim IG\left(\delta_1,\delta_2\right)$, whose pdf is given by,

$$p(\sigma_Z^2)=\frac{\delta_2^{\delta_1}}{\Gamma(\delta_1)}(\sigma_Z^2)^{-\delta_1-1}\exp\left[-\frac{\delta_2}{\sigma_Z^2}\right].$$

Taking expectation of the logarithm of the above pdf, we have,

$$\begin{aligned}
&E_{q(\theta)}\left[\ln p(\sigma_Z^2)\right]=\delta_1\ln\delta_2-\ln\Gamma(\delta_1)-\left(\delta_1+1\right)E_{q(\theta)}\left[\ln\sigma_Z^2\right]-\delta_2E_{q(\theta)}\left[\sigma_Z^{-2}\right]\\
&=\delta_1\ln\delta_2-\ln\Gamma(\delta_1)-\left(\delta_1+1\right)\left\{\ln B_{q(\sigma_Z^2)}-\psi\left(\delta_1^*\right)\right\}-\delta_2\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right),
\end{aligned}\tag{C.22}$$

where once again we utilize the relationship $E_{q(\theta)}\left[\ln\sigma_Z^2\right]=B_{q(\sigma_Z^2)}-\psi\left(\delta_1^*\right)$.

We next move to the optimal densities present in the denominator. We know $q(\omega)\sim N(\mu_{q(\omega)},\Sigma_{q(\omega)})$, so the logarithm of the optimal density is,

$$\ln q(\omega)=-\frac{K}{2}\ln(2\pi)-\frac{1}{2}\ln\left|\Sigma_{q(\omega)}\right|-\frac{1}{2}(\omega-\mu_{q(\omega)})'\Sigma_{q(\omega)}^{-1}(\omega-\mu_{q(\omega)}).$$

Taking expectation of $\ln q(\omega)$ we obtain,

$$\begin{aligned}
E_{q(\theta)}\left[\ln q(\omega)\right] &= -\frac{K}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\Sigma_{q(\omega)}\right| - \frac{1}{2}E_{q(\theta)}\left[(\omega-\mu_{q(\omega)})'\Sigma^{-1}_{q(\omega)}(\omega-\mu_{q(\omega)})\right] \\
&= -\frac{K}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\Sigma_{q(\omega)}\right| - \frac{K}{2},
\end{aligned} \tag{C.23}$$

where the last expression makes use of the following equality,

$$\begin{aligned}
&E_{q(\theta)}\left[(\omega-\mu_{q(\omega)})'\Sigma^{-1}_{q(\omega)}(\omega-\mu_{q(\omega)})\right] \\
&= (\mu_{q(\omega)}-\mu_{q(\omega)})'\Sigma^{-1}_{q(\omega)}(\mu_{q(\omega)}-\mu_{q(\omega)}) + \operatorname{tr}\left[\Sigma^{-1}_{q(\omega)}\Sigma_{q(\omega)}\right] \\
&= \operatorname{tr}\left[I_K\right] = K.
\end{aligned}$$

Lastly, we know $q(\sigma^2_Z) \sim IG(\delta^*_1, B_{q(\sigma^2_Z)})$. So, the logarithm of the optimal density is,

$$\ln q(\sigma^2_Z) = \delta^*_1 \ln B_{q(\sigma^2_Z)} - \ln\Gamma(\delta^*_1) - (\delta^*_1+1)\ln\sigma^2_Z - B_{q(\sigma^2_Z)}\sigma_Z^{-2}.$$

Taking expectation of $\ln q(\sigma^2_Z)$, we have,

$$\begin{aligned}
&E_{q(\theta)}\left[\ln q(\sigma^2_Z)\right] = \delta^*_1 \ln B_{q(\sigma^2_Z)} - \ln\Gamma(\delta^*_1) - (\delta^*_1+1)\,E_{q(\theta)}\left[\ln\sigma^2_Z\right] - B_{q(\sigma^2_Z)}E_{q(\theta)}\left[\sigma_Z^{-2}\right] \\
&= \delta^*_1 \ln B_{q(\sigma^2_Z)} - \ln\Gamma(\delta^*_1) - (\delta^*_1+1)\left\{\ln B_{q(\sigma^2_Z)} - \psi\left(\delta^*_1\right)\right\} - B_{q(\sigma^2_Z)}\left(\frac{\delta^*_1}{B_{q(\sigma^2_Z)}}\right) \\
&= -\delta^*_1 + \delta^*_1 \ln B_{q(\sigma^2_Z)} - \ln\Gamma(\delta^*_1) - (\delta^*_1+1)\ln B_{q(\sigma^2_Z)} + (\delta^*_1+1)\,\psi\left(\delta^*_1\right).
\end{aligned} \tag{C.24}$$

Substituting equations (C.20)-(C.24) in the expression for T_2, we have,

$$\begin{aligned}
T_2 &= E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^N p\left(\tilde{Z}_i|\omega,\sigma^2_Z\right)p\left(\omega\right)p\left(\sigma^2_Z\right)}{q\left(\omega\right)q\left(\sigma^2_Z\right)}\right)\right] \\
&= \sum_{i=1}^N E_{q(\theta)}\left[\ln p\left(\tilde{Z}_i|\omega,\sigma^2_Z\right)\right] + E_{q(\theta)}\left[\ln p\left(\omega\right)\right] + E_{q(\theta)}\left[\ln p\left(\sigma^2_Z\right)\right] \\
&\quad - E_{q(\theta)}\left[\ln q\left(\omega\right)\right] - E_{q(\theta)}\left[\ln q\left(\sigma^2_Z\right)\right] \\
&= -\frac{MN}{2}\ln\left(2\pi\right) - \frac{MN}{2}\ln B_{q(\sigma^2_Z)} + \frac{MN}{2}\psi\left(\delta^*_1\right) \\
&\quad - \frac{1}{2}\left(\frac{\delta^*_1}{B_{q(\sigma^2_Z)}}\right)\sum_{i=1}^N\left\{\ \|\ \mu_{q(\tilde{Z}_i)} - X_i\mu_{q(\omega)}\ \|^2 + \operatorname{tr}\left[\Sigma_{q(\tilde{Z}_i)} + X_i\Sigma_{q(\omega)}X_i'\right]\ \right\} \\
&\quad - \frac{K}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln|O_0| - \frac{1}{2}\left\{\ (\mu_{q(\omega)}-\omega_0)'O_0^{-1}(\mu_{q(\omega)}-\omega_0) + \operatorname{tr}\left[\Sigma_{q(\omega)}O_0^{-1}\right]\ \right\} \\
&\quad + \delta_1\ln\delta_2 - \ln\Gamma(\delta_1) - (\delta_1+1)\left\{\ln B_{q(\sigma^2_Z)} - \psi\left(\delta^*_1\right)\right\} - \delta_2\left(\frac{\delta^*_1}{B_{q(\sigma^2_Z)}}\right) \\
&\quad + \frac{K}{2}\ln\left(2\pi\right) + \frac{1}{2}\ln\left|\Sigma_{q(\omega)}\right| + \frac{K}{2} \\
&\quad + \delta^*_1 - \delta^*_1\ln B_{q(\sigma^2_Z)} + \ln\Gamma(\delta^*_1) + (\delta^*_1+1)\ln B_{q(\sigma^2_Z)} - (\delta^*_1+1)\,\psi\left(\delta^*_1\right).
\end{aligned}$$

After canceling terms and rearranging, we have,

$$
\begin{aligned}
T_2 = &\ \delta_1^* + \frac{K}{2} - \frac{MN}{2}\ln(2\pi) - \ln\Gamma(\delta_1) + \ln\Gamma(\delta_1^*) - \frac{1}{2}\ln|O_0| + \delta_1 \ln\delta_2 + \frac{1}{2}\ln\left|\Sigma_{q(\omega)}\right| \\
&- \delta_1^* \ln B_{q(\sigma_Z^2)} - \frac{1}{2\sigma_\mu^2}\left\{(\mu_{q(\omega)} - \omega_0)' O_0^{-1}(\mu_{q(\omega)} - \omega_0) + \mathrm{tr}\left[\Sigma_{q(\omega)} O_0^{-1}\right]\right\} \\
&- \frac{1}{2}\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\left[2\delta_2 + \sum_{i=1}^{N}\left\{ \parallel \mu_{q(\tilde{Z}_i)} - X_i \mu_{q(\omega)} \parallel^2 + \mathrm{tr}\left[\Sigma_{q(\tilde{Z}_i)} + X_i \Sigma_{q(\omega)} X_i'\right] \right\}\right].
\end{aligned}
\tag{C.25}
$$

3.3 Derivation of T_3

On terms that constitute,

$$
T_3 = E_{q(\theta)}\left[\ln\left(\frac{\prod_{i=1}^{N} p\left(\widetilde{W}_i|\tilde{Z}_i, \sigma_u^2\right) p\left(\sigma_u^2\right)}{\prod_{i=1}^{N} q\left(\tilde{Z}_i\right) q\left(\sigma_u^2\right)}\right)\right],
$$

we have the following information,

$$
\begin{aligned}
p\left(\widetilde{W}_i|\tilde{Z}_i, \sigma_u^2\right) &\sim N_M\left(\tilde{Z}_i, \sigma_u^2 I_M\right) \\
p(\sigma_u^2) &\sim IG(\delta_3, \delta_4) \\
q\left(\tilde{Z}_i\right) &\sim N_M\left(\mu_{q(\tilde{Z}_i)}, \Sigma_{q(\tilde{Z}_i)}\right) \\
q\left(\sigma_u^2\right) &\sim IG\left(\delta_3^*, B_{q(\sigma_u^2)}\right).
\end{aligned}
$$

We begin by working on the terms in the numerator of T_3. Taking logarithm of the density of $\tilde{W}_i$ we get,

$$
\ln p\left(\widetilde{W}_i|\tilde{Z}_i, \sigma_u^2\right) = -\frac{M}{2}\ln(2\pi) - \frac{M}{2}\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}(\widetilde{W}_i - \tilde{Z}_i)'(\widetilde{W}_i - \tilde{Z}_i).
$$

Taking expectation of the above expression we have,

$$
\begin{aligned}
&E_{q(\theta)}\left[\sum_{i=1}^{N}\ln p\left(\widetilde{W}_i|\tilde{Z}_i, \sigma_u^2\right)\right] \\
&= -\frac{MN}{2}\ln(2\pi) - \frac{MN}{2}E_{q(\theta)}\left[\ln\sigma_u^2\right] - \frac{1}{2\sigma_u^2}\sum_{i=1}^{N}E_{q(\theta)}\left[(\widetilde{W}_i - \tilde{Z}_i)'(\widetilde{W}_i - \tilde{Z}_i)\right] \\
&= -\frac{MN}{2}\ln(2\pi) - \frac{MN}{2}\left\{\ln B_{q(\sigma_u^2)} - \psi\left(\delta_3^*\right)\right\} - \frac{1}{2}\left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\sum_{i=1}^{N}E_{q(\theta)}\left[(\widetilde{W}_i - \tilde{Z}_i)'(\widetilde{W}_i - \tilde{Z}_i)\right] \\
&= -\frac{MN}{2}\left\{\ln(2\pi) + \ln B_{q(\sigma_u^2)} - \psi\left(\delta_3^*\right)\right\} - \frac{1}{2}\left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\sum_{i=1}^{N}\left\{\parallel \widetilde{W}_i - \mu_{q(\tilde{Z}_i)} \parallel^2 + \mathrm{tr}\left[\Sigma_{q(\tilde{Z}_i)}\right]\right\},
\end{aligned}
\tag{C.26}
$$

where we use the equality $E_{q(\theta)}\left[(\widetilde{W}_i - \tilde{Z}_i)'(\widetilde{W}_i - \tilde{Z}_i)\right] = \parallel \widetilde{W}_i - \mu_{q(\tilde{Z}_i)} \parallel^2 + \mathrm{tr}\left[\Sigma_{q(\tilde{Z}_i)}\right]$.

We know the prior for σ_u^2 is an inverse gamma distribution i.e., $\sigma_u^2 \sim IG(\delta_3, \delta_4)$, and the pdf for σ_u^2 is,

$$
p(\sigma_u^2) = \frac{\delta_4^{\delta_3}}{\Gamma(\delta_3)}(\sigma_u^2)^{-\delta_3 - 1}\exp\left[-\frac{\delta_4}{\sigma_u^2}\right].
$$

Taking logarithm of the above expression we have,

$$\ln p(\sigma_u^2) = \delta_3 \ln \delta_4 - \ln \Gamma(\delta_3) - (\delta_3 + 1) \ln \sigma_u^2 - \delta_4 \sigma_u^{-2},$$

and expectation of $\ln p(\sigma_u^2)$ yields,

$$\begin{aligned} E_{q(\theta)}\left[\ln p(\sigma_u^2)\right] &= \delta_3 \ln \delta_4 - \ln \Gamma(\delta_3) - (\delta_3 + 1)\, E_{q(\theta)}\left[\ln \sigma_u^2\right] - \delta_4 E_{q(\theta)}\left[\sigma_u^{-2}\right] \\ &= \delta_3 \ln \delta_4 - \ln \Gamma(\delta_3) - (\delta_3 + 1)\left[\ln B_{q(\sigma_u^2)} + \psi\left(\delta_3^*\right)\right] - \delta_4 \left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right). \end{aligned} \tag{C.27}$$

Next, we look at the optimal densities in the denominator of T_3. We know $q(\tilde{Z}_i) \sim N_M(\mu_{q(\tilde{Z}_i)}, \Sigma_{q(\tilde{Z}_i)})$, so the logarithm of the corresponding pdf is,

$$\ln q\left(\tilde{Z}_i\right) = -\frac{M}{2} \ln(2\pi) - \frac{1}{2} \ln\left|\Sigma_{q(\tilde{Z}_i)}\right| - \frac{1}{2}(\tilde{Z}_i - \mu_{q(\tilde{Z}_i)})' \Sigma^{-1}_{q(\tilde{Z}_i)} (\tilde{Z}_i - \mu_{q(\tilde{Z}_i)}).$$

Taking expectation of the sum of the above expression yields,

$$\begin{aligned} &E_{q(\theta)}\left[\sum_{i=1}^{N} \ln q\left(\tilde{Z}_i\right)\right] \\ &= -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln\left|\Sigma_{q(\tilde{Z}_i)}\right| - \frac{1}{2} \sum_{i=1}^{N} E_{q(\theta)}\left[(\tilde{Z}_i - \mu_{q(\tilde{Z}_i)})' \Sigma^{-1}_{q(\tilde{Z}_i)} (\tilde{Z}_i - \mu_{q(\tilde{Z}_i)})\right] \\ &= -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln\left|\Sigma_{q(\tilde{Z}_i)}\right| - (\mu_{q(\tilde{Z}_i)} - \frac{MN}{2}, \end{aligned} \tag{C.28}$$

where we use the following equality,

$$\begin{aligned} &E_{q(\theta)}\left[(\tilde{Z}_i - \mu_{q(\tilde{Z}_i)})' \Sigma^{-1}_{q(\tilde{Z}_i)} (\tilde{Z}_i - \mu_{q(\tilde{Z}_i)})\right] \\ &= (\mu_{q(\tilde{Z}_i)} - \mu_{q(\tilde{Z}_i)})' \Sigma^{-1}_{q(\tilde{Z}_i)} (\mu_{q(\tilde{Z}_i)} - \mu_{q(\tilde{Z}_i)}) + \operatorname{tr}\left[\Sigma^{-1}_{q(\tilde{Z}_i)} \Sigma_{q(\tilde{Z}_i)}\right] \\ &= \operatorname{tr}\left[I_M\right] = M. \end{aligned}$$

The optimal density $q(\sigma_u^2) \sim IG(\delta_3^*, B_{q(\sigma_u^2)})$ has the pdf,

$$q(\sigma_u^2) = \frac{B_{q(\sigma_u^2)}^{\delta_3^*}}{\Gamma(\delta_3^*)} (\sigma_u^2)^{-\delta_3^* - 1} \exp\left[-\frac{B_{q(\sigma_u^2)}}{\sigma_u^2}\right].$$

Taking expectation of the logarithm of the above pdf we get,

$$\begin{aligned} &E_{q(\theta)}\left[\ln q(\sigma_u^2)\right] = \delta_3^* \ln B_{q(\sigma_u^2)} - \ln \Gamma(\delta_3^*) - (\delta_3^* + 1)\, E_{q(\theta)}\left[\ln \sigma_u^2\right] - B_{q(\sigma_u^2)} E_{q(\theta)}\left[\sigma_u^{-2}\right] \\ &= \delta_3^* \ln B_{q(\sigma_u^2)} - \ln \Gamma(\delta_3^*) - (\delta_3^* + 1)\left\{\ln B_{q(\sigma_u^2)} - \psi\left(\delta_3^*\right)\right\} - B_{q(\sigma_u^2)} \left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right) \\ &= -\delta_3^* + \delta_3^* \ln B_{q(\sigma_u^2)} - \ln \Gamma(\delta_3^*) - (\delta_3^* + 1) \ln B_{q(\sigma_u^2)} + (\delta_3^* + 1)\, \psi\left(\delta_3^*\right). \end{aligned} \tag{C.29}$$

Substituting equations (C.26)-(C.29) into the expression for T_3 we have,

$$T_3 = \sum_{i=1}^{N} E_{q(\theta)}\left[\ln p\left(\widetilde{W}_i | \tilde{Z}_i, \sigma_u^2\right)\right] + E_{q(\theta)}\left[\ln p\left(\sigma_u^2\right)\right] - \sum_{i=1}^{N} E_{q(\theta)}\left[\ln q\left(\tilde{Z}_i\right)\right]$$

$$
\begin{aligned}
& - E_{q(\theta)}\left[\ln q\left(\sigma_u^2\right)\right] \\
= & -\frac{MN}{2}\ln(2\pi) - \frac{MN}{2}\ln B_{q(\sigma_u^2)} + \frac{MN}{2}\psi\left(\delta_3^*\right) \\
& - \frac{1}{2}\left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\sum_{i=1}^{N}\left\{\parallel \widetilde{W}_i - \mu_{q\left(\tilde{Z}_i\right)} \parallel^2 + \text{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)}\right]\right\} + \delta_3\ln\delta_4 - \ln\Gamma(\delta_3) \\
& - (\delta_3+1)\ln B_{q(\sigma_u^2)} + (\delta_3+1)\psi\left(\delta_3^*\right) - \delta_4\left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right) + \frac{MN}{2}\ln(2\pi) + \frac{N}{2}\ln\left|\Sigma_{q\left(\tilde{Z}_i\right)}\right| \\
& + \frac{MN}{2} + \delta_3^* - \delta_3^*\ln B_{q(\sigma_u^2)} + \ln\Gamma(\delta_3^*) + (\delta_3^*+1)\ln B_{q(\sigma_u^2)} - (\delta_3^*+1)\psi\left(\delta_3^*\right).
\end{aligned}
$$

After canceling terms and rearranging, we obtain,

$$
\begin{aligned}
T_3 = & \ \delta_3^* + \frac{MN}{2} + \delta_3\ln\delta_4 - \ln\Gamma(\delta_3) + \ln\Gamma(\delta_3^*) + \frac{N}{2}\ln\left|\Sigma_{q\left(\tilde{Z}_i\right)}\right| - \delta_3^*\ln B_{q(\sigma_u^2)} \\
& - \frac{1}{2}\left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\left[2\delta_4 + \sum_{i=1}^{N}\left\{\parallel \widetilde{W}_i - \mu_{q\left(\tilde{Z}_i\right)} \parallel^2 + \text{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)}\right]\right\}\right].
\end{aligned} \tag{C.30}
$$

By summing the three terms T_1, T_2, T_3 (from equations (C.19), (C.25) and (C.30), respectively), we obtain the evidence lower bound (ELBO):

$$
\begin{aligned}
\ell = & \ E_{q(\theta)}\left[\ln\left(\frac{p(\theta,y)}{q(\theta)}\right)\right] = T_1 + T_2 + T_3 \\
= & -MN\ln(2\pi) + \frac{MN}{2}\ln(2) + \frac{K}{2} + \frac{M}{2}\left[N+\nu_1+2\right] + \delta_1^* + \delta_3^* + \delta_1\ln\delta_2 \\
& + \delta_3\ln\delta_4 - \ln\Gamma(\delta_1) + \ln\Gamma(\delta_1^*) - \ln\Gamma(\delta_3) + \ln\Gamma(\delta_3^*) - \frac{1}{2}\ln|O_0| \\
& - \ln\Gamma_M\left(\frac{\nu_0}{2}\right) + \ln\Gamma_M\left(\frac{\nu_1}{2}\right) - \frac{1}{2}\ln|B_0| - \frac{1}{2}\ln|G_0| + \frac{\nu_0}{2}\ln|S_0| \\
& + \frac{1}{2}\ln\left|\Sigma_{q(\beta)}\right| + \frac{1}{2}\ln\left|\Sigma_{q(\gamma)}\right| + \frac{1}{2}\ln\left|\Sigma_{q(\omega)}\right| + \frac{\nu_1}{2}\ln\left|B_{q(\Sigma)}\right| + \frac{N}{2}\ln\left|\Sigma_{q\left(\tilde{Z}_i\right)}\right| \\
& - \delta_1^*\ln B_{q(\sigma_Z^2)} - \delta_3^*\ln B_{q(\sigma_u^2)} - \frac{1}{2}\text{tr}\left[\Sigma_{q(\beta)}B_0^{-1}\right] - \frac{1}{2}\text{tr}\left[\Sigma_{q(\gamma)}G_0^{-1}\right] - \frac{1}{2}\text{tr}\left[S_0\nu_1 B_{q(\Sigma)}\right] \\
& - \frac{1}{2}(\mu_{q(\beta)} - \beta_0)' B_0^{-1}(\mu_{q(\beta)} - \beta_0) - \frac{1}{2}(\mu_{q(\gamma)} - \gamma_0)' G_0^{-1}(\mu_{q(\gamma)} - \gamma_0) \\
& - \frac{1}{2}(\mu_{q(\omega)} - \omega_0)' O_0^{-1}(\mu_{q(\omega)} - \omega_0) - \frac{1}{2}\text{tr}\left[\Sigma_{q(\omega)}O_0^{-1}\right] \\
& - \frac{1}{2}\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}}\right)\left[2\delta_2 + \sum_{i=1}^{N}\left\{\ \parallel \mu_{q\left(\tilde{Z}_i\right)} - X_i\mu_{q(\omega)} \parallel^2 + \text{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)} + X_i\Sigma_{q(\omega)}X_i'\right]\ \right\}\right] \\
& - \frac{1}{2}\left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}}\right)\left[2\delta_4 + \sum_{i=1}^{N}\left\{\parallel \widetilde{W}_i - \mu_{q\left(\tilde{Z}_i\right)} \parallel^2 + \text{tr}\left[\Sigma_{q\left(\tilde{Z}_i\right)}\right]\right\}\right] \\
& - \frac{1}{2}\sum_{i=1}^{N}\left\{\begin{array}{c}
\left(y_i - X_i\mu_{q(\beta)} - \left(\mu_{q\left(\tilde{Z}_i\right)} \odot \mu_{q(\gamma)}\right)\right)'\left(\nu_1 B_{q(\Sigma)}\right) \\
\times\left(y_i - X_i\mu_{q(\beta)} - \left(\mu_{q\left(\tilde{Z}_i\right)} \odot \mu_{q(\gamma)}\right)\right) \\
+\text{tr}\left[\Sigma_{q(\beta)}X_i'\left(\nu_1 B_{q(\Sigma)}\right)X_i\right] \\
\text{tr}\left[\left(\Sigma_{q\left(\tilde{Z}_i\right)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu_{q(\gamma)}'\right)\right)\left(\nu_1 B_{q(\Sigma)}\right)\right] \\
+\text{tr}\left[\left(\mu_{q\left(\tilde{Z}_i\right)}\mu_{q\left(\tilde{Z}_i\right)}' \odot \Sigma_{q(\gamma)}\right)\left(\nu_1 B_{q(\Sigma)}\right)\right]
\end{array}\right\}
\end{aligned} \tag{C.31}
$$

4 Appendix D. Gibbs sampling algorithm for SUR model.

Without any prior information on the measurement error, the SUR model for M equations is the following:

$$y_i = X_i\beta + W_i\gamma + \varepsilon_i \ , \ \varepsilon_i \sim N\left(0, \Sigma_\varepsilon\right) \ , \ i = 1, .., N$$

where W_i is the covariate with measurement error. Let us stack the covariates $V_i = [X_i, W_i]$ and the vectors of parameters $\theta = (\beta', \gamma')'$, then

$$y_i = V_i\theta + \varepsilon_i \ , \ \varepsilon_i \sim N\left(0, \Sigma_\varepsilon\right) \ , \ i = 1, .., N$$

The regression coefficients are assumed to have a Gaussian prior, $\theta \sim N_{K+M}\left(\theta_0, D_0\right)$ and for the precision matrix Σ_ε^{-1}, we assume the same Wishart distribution $\Sigma_\varepsilon^{-1} \sim W_M\left(\nu_0, S_0\right)$ as previously. Following Greenberg (2012), the Gibbs sampling algorithm for the SUR model on cross-section is given in the Algorithm 1, where we use the notation $\theta_0 = 0_{K+M}$, $D_0 = I_{K+M}$, $\nu_0 = 10$, and $S_0 = I_M$.

Algorithm 1 (Gibbs sampling for SUR model)

---

1. Sample $\theta \sim N_{K+M}\left(\overline{\theta}, D_1\right)$
   where $D_1^{-1} = \left[\sum_{i=1}^{N} V_i'\Sigma_\varepsilon^{-1}V_i + D_0^{-1}\right]$
   and $\overline{\theta} = D_1\left[\sum_{i=1}^{N} V_i'\Sigma_\varepsilon^{-1}y_i + D_0^{-1}\theta_0\right]$, $K = \sum_{m=1}^{M} k_m$.
2. Sample $\Sigma_\varepsilon^{-1} \sim W_M\left(\nu_1, S_1\right)$, where $\nu_1 = \nu_0 + N$
   and $S_1^{-1} = \left[S_0^{-1} + \sum_{i=1}^{N}\left(y_i - V_i\theta\right)\left(y_i - V_i\theta\right)'\right]$.

---

5 Appendix E. Some Additional Quantities

In the paper, we compare different models using the deviance information criterion (DIC) introduced by Spiegelhalter et al. (2002). The DIC has certain advantages, but it also has some limitations and they have been carefully reviewed by Spiegelhalter et al. (2014). One such drawback is that DIC, as introduced by Spiegelhalter et al. (2002), may lead to misleading inferences while comparing latent variable models. This has been noted in several papers including Celeux et al. (2006), Li et al. (2012), and Chan and Grant (2016). According to Li et al. (2012), the DIC should not be utilized for comparing models involving data augmentation. This is because the complete data likelihood (i.e., likelihood involving the latent variables) is often non-regular and so standard asymptotic reasonings required for DIC are no longer applicable. Moreover, such a DIC is sensitive not only to distributional representations but also to transformation of latent variables. In such cases, the correct alternative is to compute the observed-data DIC based on observed-data likelihood or integrated likelihood (i.e., likelihood obtained by integrating out the latent variables). This has been discussed and well exhibited by Chan and Grant (2016). In our case, we follow the approach of Chan and Grant (2016) and compute an integrated likelihood for calculating the marginal likelihood used for the DIC (see below).

5.1 Integrated likelihood, DIC and integrated marginal likelihood.

We derive an integrated likelihood — instead of the complete-data or conditional likelihoods — for the SURME model to compute the DIC and the effective number of parameters p_D. So, we need to rewrite the complete-data likelihood of our model for evaluating the integrated likelihood by integrating out the latent variables Z. The initial specification is given by:

$$\left\{ \begin{array}{lcl} y_i & = & X_i\beta + Z_i\gamma + \varepsilon_i \\ \widetilde{W}_i & = & \widetilde{Z}_i + \widetilde{u}_i \\ \widetilde{Z}_i & \sim & N_M\left(X_i\omega, \Sigma_{\widetilde{Z}}\right) \quad \text{with } \Sigma_{\widetilde{Z}} = \sigma_Z^2 I_M \\ \widetilde{u}_i & \sim & N_M\left(0, \Sigma_{\widetilde{u}}\right) \quad \text{with } \Sigma_{\widetilde{u}} = \sigma_i^2 I_M \end{array} \right. \tag{E.1}$$

where y_i, X_i, W_i are $(M \times 1)$, $(M \times K)$ and $(M \times 1)$, respectively. The vectors $\widetilde{W}_i$, $\widetilde{Z}_i$ and $\widetilde{u}_i$ are all $(M \times 1)$. The likelihood of the SURME model is given by,

$$p\left(y|X, W, Z, \theta\right) = \prod_{i=1}^{N} \left\{ \begin{array}{c} (2\pi)^{-M/2} \left|\Sigma_\varepsilon\right|^{-1/2} \exp\left[-\frac{1}{2}\left\{ \begin{array}{c} (y_i - X_i\beta - Z_i\gamma)' \\ \times\Sigma_\varepsilon^{-1}\left(y_i - X_i\beta - Z_i\gamma\right) \end{array} \right\}\right] \\ \times (2\pi)^{-M/2} \left|\Sigma_{\widetilde{u}}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\widetilde{W}_i - \widetilde{Z}_i\right)' \Sigma_{\widetilde{u}}^{-1}\left(\widetilde{W}_i - \widetilde{Z}_i\right)\right] \\ \times (2\pi)^{-M/2} \left|\Sigma_{\widetilde{Z}}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\widetilde{Z}_i - X_i\omega\right)' \Sigma_{\widetilde{Z}}^{-1}\left(\widetilde{Z}_i - X_i\omega\right)\right] \end{array} \right\}$$

where θ represents the parameters $\left(\beta, \gamma, \omega, \Sigma_\varepsilon, \Sigma_{\widetilde{u}}, \Sigma_{\widetilde{Z}}\right)$ of the model. We rewrite equation (E.1) in terms of $\tilde{Z}_i$ as follows,

$$\begin{aligned} y_i &= X_i\beta + Z_i\gamma + \varepsilon_i = X_i\beta + \left(\widetilde{Z}_i \odot \gamma\right) + \varepsilon_i \\ &= X_i\beta + (X_i\omega) \odot \gamma + (\widetilde{\varepsilon}_{Zi} \odot \gamma) + \varepsilon_i. \end{aligned}$$

Taking expectation of Y_i we have,

$$\begin{aligned} E\left[y_i\right] &= E\left[X_i\beta + (X_i\omega) \odot \gamma + (\widetilde{\varepsilon}_{Zi} \odot \gamma) + \varepsilon_i\right] \\ &= X_i\beta + (X_i\omega) \odot \gamma = X_i\beta + diag\left(X_i\omega\right)\gamma. \end{aligned}$$

Subtracting the expected value from the response, we obtain,

$$y_i - E[y_i] = (\widetilde{\varepsilon}_{Zi} \odot \gamma) + \varepsilon_i.$$

The variance of y_i can be derived as follows,

$$\begin{aligned} Var[y_i] &= E\left[(y_i - E[y_i])(y_i - E[y_i])'\right] \\ &= E\left[((\widetilde{\varepsilon}_{Zi} \odot \gamma) + \varepsilon_i)((\widetilde{\varepsilon}_{Zi} \odot \gamma) + \varepsilon_i)'\right] \\ &= E\left[\widetilde{\varepsilon}_{Zi}\widetilde{\varepsilon}'_{Zi}\right] \odot \gamma\gamma' + E\left[\varepsilon_i\varepsilon_i'\right] \\ &= \Sigma_{\widetilde{Z}} \odot \gamma\gamma' + \Sigma_\varepsilon = \sigma_Z^2 D_{\gamma^2} + \Sigma_\varepsilon \end{aligned}$$

where D_{γ^2} is a $(M \times M)$ diagonal matrix,

$$D_{\gamma^2} = diag(\gamma^2) = \begin{pmatrix} \gamma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_M^2 \end{pmatrix}$$

wherein γ is a $(M \times 1)$ vector. Then,

$$\begin{aligned} y_i|X_i, \theta &\sim N_M\left(X_i\beta + diag(X_i\omega)\gamma, \Sigma_Y\right) \\ \text{with} \quad & \Sigma_Y = \sigma_Z^2.D_{\gamma^2} + \Sigma_\varepsilon \\ \text{and} \quad & \theta = \left\{\beta, \gamma, \omega, \sigma_Z^2, D_{\gamma^2}, \Sigma_\varepsilon\right\}. \end{aligned} \tag{E.2}$$

By integrating out the latent variables Z_i, the integrated likelihood is given by,

$$p(y|X_i, \theta) = \prod_{i=1}^{N} (2\pi)^{-M/2} |\Sigma_y|^{-1/2} \exp\left[\begin{array}{c} -\frac{1}{2}(y_i - X_i\beta - diag(X_i\omega)\gamma)' \Sigma_y^{-1} \\ \times (y_i - X_i\beta - diag(X_i\omega)\gamma) \end{array}\right]$$

Then, the log-integrated likelihood of the model is,

$$\begin{aligned} \ln p(y|X_i, \theta) = &-\frac{MN}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma_y| - \frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\beta - diag(X_i\omega)\gamma)' \Sigma_y^{-1} \\ &\times (y_i - X_i\beta - diag(X_i\omega)\gamma). \end{aligned} \tag{E.3}$$

We know $DIC = -4E_{\theta|y}[\ln p(y|\theta)] + 2\ln p\left(y|\widehat{\theta}\right)$. The first term $E_{\theta|y}[\ln p(y|\theta)]$ can be estimated by averaging $\ln p(y|\theta)$ over the posterior draws of θ and $\ln p\left(y|\widehat{\theta}\right)$ can be readily computed. The effective number of parameters is defined as $p_D = -2\big(E_{\theta|y}[\ln p(y|\theta)] - \ln p(y|\widehat{\theta})\big)$.

On Table G5 (see below), we present the DIC as well as the effective number of parameters p_D for both the SUR and the SURME models, where we have used a thinning of factor 1 and of factor 100. While we use DIC as for model comparison, it should be noted that model adequacy measure accounting for measurement error in a multi-equation setup is not available yet and thus is an open area of research. More importantly, the effective numbers of parameters p_D are strongly different between the two models: one is positive (for SURME), the other is negative (for SUR). Spiegelhalter et al. (2002) have shown that p_D is approximately equal to trace of the product of Fisher's information and the posterior covariance, which simplifies to the trace of the "hat" matrix for normal models. The effective number of parameters p_D is approximately equal to the actual number of independent parameters for approximately normal likelihoods and negligible prior information. However, it is possible that p_D is negative for models outside the log-concave densities. (Spiegelhalter et al., 2002) note that when p_D is negative, it is indicative of a poor fit between the model and the data or sometimes a substantial conflict between the prior and data. However, (Celeux et al., 2006) conclude that when $p_D < 0$, the DIC's are not adequate for evaluating the model fit and its complexity.

5.2 Conditional posterior distribution of γ, marginalized over Z_i

Both Z and γ are unknown, and drawing them conditional on each other can cause high autocorrelation in MCMC draws, which leads to high inefficiency factor. To reduce high values of inefficiency factor, a common trick is to sample them jointly i.e., sample γ from marginal distribution and then sample $Z|\gamma$ or *vice versa.* From (E.2), we can derive the conditional posterior distribution of the vector of coefficients γ, marginalized over Z_i. Since $\gamma \sim N_M(\gamma_0, G_0)$ and $y_i \sim N_M(X_i\beta + diag(X_i\omega)\gamma, \Sigma_y)$, it follows,

$$p(\gamma|y, X, \beta, \omega, \Sigma_y) \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(y_i - X_i\beta - diag(X_i\omega)\gamma)'\Sigma_y^{-1}\right.$$
$$\left.\times (y_i - X_i\beta - diag(X_i\omega)\gamma)\right] \times \exp\left[-\frac{1}{2}(\gamma-\gamma_0)' G_0^{-1}(\gamma-\gamma_0)\right].$$

Opening the quadratic terms and completing the square for γ yields a kernel of a normal density and hence,

$$\gamma|y, X, \beta, \omega, \Sigma_y \sim N_M(\overline{\gamma}, G_1)$$

$$\text{where } G_1^{-1} = \left[\sum_{i=1}^{N} diag(X_i\omega)\Sigma_y^{-1} diag(X_i\omega) + G_0^{-1}\right]$$
$$\text{and } \overline{\gamma} = G_1\left[\sum_{i=1}^{N} diag(X_i\omega)\Sigma_y^{-1}(y_i - X_i\beta) + G_0^{-1}\gamma_0\right]$$
$$\text{with } \Sigma_y = \sigma_Z^2.D_{\gamma^2} + \Sigma_\varepsilon.$$

The problem is that Σ_y depends on γ. The only way to make it is to consider that D_{γ^2} uses the estimated γ's at the previous draw. However, the several Monte Carlo tests we realized did not improve our initial results with standard Gibbs sampling.

6 Appendix F. Optimal thinning

Suppose we generate a Markov chain x_t for $t > 1$. Starting at some value x_0, suppose it costs 1 unit of computation to progress from x_{t-1} to x_t. Further, suppose our interest lies in calculating the expected value of $y_t = f(x_t)$, where f is some real-valued function. Let the cost for computing f be θ. While θ is typically less than 1, it is possible for θ to be equal to or greater than 1. So, the efficiency of thinning depends on the following 3 factors: the computing cost of y_t, transition cost of moving from x_{t-1} to x_t, and the auto-covariance of the quantity of interest y_t.

For a first order autoregressive model with an autocorrelation given by $\rho_k = \rho^{|k|}$ for $\rho \in (-1, 1)$ and $k \in \mathbb{N}$, Owen (2017) shows that the relative efficiency of thinning at factor k (relatively to factor 1) is given by,

$$\text{effar}(k, \theta, \rho) = \frac{1+\theta}{k+\theta}\frac{1+\rho}{1-\rho}\frac{1-\rho^k}{1+\rho^k},$$

and argmax_k effar(k, θ, ρ) for a range of correlations ρ and costs θ gives the value of the optimal thinning k computed via a search procedure[3]. It should be noted that in the limit efficiency gain approaches $\theta + 1$ as $\rho \to 1$, as shown in Table F1.

Table F1 - Optimal thinning (k) and asymptotic efficiency (effar) for some ρ and θ values.

	ρ	0.1	0.5	0.8	0.9	0.95	0.99	0.995	0.996	0.997	0.999	0.9999
$\theta = 0.5$	k	1	2	4	6	10	31	49	57	69	144	669
	effar	1	1.080	1.256	1.341	1.398	1.464	1.477	1.4805	1.484	1.492	1.498
$\theta = 1$	k	1	2	5	8	13	39	62	72	87	182	843
	effar	1	1.200	1.519	1.681	1.792	1.925	1.953	1.959	1.966	1.984	1.9964
$\theta = 2.4892$	k	1	3	6	11	18	53	84	97	118	246	1143
	effar	1	1.483	2.162	2.567	2.865	3.256	3.340	3.360	3.382	3.437	3.478
$\theta = 5$	k	2	4	8	13	22	66	106	123	149	310	1442
	effar	1.027	1.765	2.960	3.766	4.430	5.382	5.599	5.652	5.710	5.858	5.969

[3]An R code of this search procedure in given in Owen (2017).

7 Appendix G. Tables of results

Table G1 - Bayesian estimation of SUR model. The table presents the posterior mean (MEAN), relative error (RE), standard deviation (STD), inefficiency factor (IF), 95% highest posterior density interval (INF-HPDI, SUP-HPDI) and Geweke's convergence diagnostics (CD). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Draws=51,000, Burnin draws=1,000, thinning=1, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
MEAN	4.142	5.621	4.246	5.147	4.648	3.353	3.207	3.217	4.130	0.515	4.103
RE	0.381	0.124	0.062	0.354	0.549	-0.162	-0.198	-0.196	3.130	0.029	3.103
STD	0.348	0.218	0.106	0.346	0.232	0.111	0.104	0.105	0.337	0.239	0.334
IF	1.001	1.000	1.002	1.001	1.000	1.001	1.001	1.002	1.007	1.010	1.007
INF-HPDI	3.570	5.262	4.072	4.577	4.267	3.170	3.035	3.045	3.610	0.127	3.585
SUP-HPDI	4.715	5.979	4.420	5.717	5.028	3.536	3.379	3.389	4.712	0.913	4.681
CD	1.000	1.000	0.990	1.000	0.990	1.000	0.980	1.000	1.000	0.980	0.990

Table G2 - Bayesian estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, IF is inefficiency factor, and (INF-HPDI, SUP-HPDI) represents 95% HPDI, and CD represents Geweke's convergence diagnostics. $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Draws=51,000, Burnin draws=1,000, thinning=1, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.25	1	0.5	1
MEAN	2.896	4.859	3.986	3.716	3.694	2.999	4.083	4.103	0.974	0.251	1.026	0.504	1.015
RE	-0.035	-0.028	-0.004	-0.071	-0.028	-0.000	0.021	0.026	-0.026	0.006	0.026	0.009	0.015
STD	0.371	0.243	0.119	0.376	0.260	0.126	0.131	0.130	0.071	0.020	0.187	0.135	0.184
IF	4.594	4.344	4.028	5.342	5.843	5.295	8.623	10.518	4.208	2.294	3.161	2.677	3.146
INF-HPDI	2.279	4.454	3.789	3.090	3.261	2.789	3.872	3.893	0.862	0.220	0.753	0.298	0.747
SUP-HPDI	3.499	5.254	4.179	4.327	4.117	3.204	4.301	4.321	1.094	0.284	1.363	0.741	1.345
CD	0.980	0.960	0.940	0.940	0.940	0.960	0.990	0.940	0.950	0.900	0.910	0.900	0.930

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.466	0.773	0.306	1.489	1.068	0.443
RE	-0.022	0.030	0.021	-0.008	0.017	-0.016
STD	0.165	0.109	0.055	0.165	0.109	0.055
IF	1.420	1.446	1.498	1.427	1.454	1.503
INF-HPDI	1.195	0.594	0.216	1.217	0.889	0.353
SUP-HPDI	1.738	0.952	0.396	1.760	1.247	0.533
CD	0.980	0.970	0.950	0.950	0.990	0.930

Table G3 - Bayesian estimation of SURME model. Autocorrelations ($\rho(\tau)$). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Draws=$51,000$, Burnin draws=$1,000$, thinning=1, Replications=100.

$\rho(\tau)$	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
1	0.848	0.829	0.865	0.854	0.852	0.881	0.980	0.986	0.364	0.445	0.819	0.791	0.816
2	0.735	0.708	0.755	0.745	0.747	0.783	0.961	0.972	0.301	0.272	0.677	0.631	0.672
3	0.645	0.614	0.662	0.658	0.665	0.700	0.942	0.959	0.290	0.206	0.564	0.506	0.558
4	0.572	0.540	0.583	0.587	0.599	0.630	0.923	0.946	0.283	0.175	0.473	0.408	0.466
5	0.512	0.481	0.516	0.529	0.547	0.569	0.905	0.933	0.277	0.154	0.399	0.331	0.391
6	0.462	0.433	0.458	0.481	0.504	0.517	0.888	0.920	0.271	0.140	0.339	0.269	0.331
7	0.420	0.393	0.409	0.441	0.468	0.473	0.870	0.908	0.265	0.128	0.291	0.221	0.282
8	0.385	0.360	0.367	0.407	0.438	0.434	0.853	0.895	0.259	0.118	0.251	0.183	0.243
9	0.356	0.332	0.331	0.378	0.413	0.401	0.837	0.883	0.254	0.109	0.218	0.152	0.211
10	0.330	0.309	0.300	0.354	0.391	0.372	0.820	0.871	0.249	0.102	0.191	0.128	0.184

Table G4 - Bayesian estimation of SURME model. Autocorrelations ($\rho(\tau)$). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Draws=$51,000$, Burnin draws=$1,000$, thinning=100, Replications=100.

$\rho(\tau)$	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
1	0.039	0.036	0.028	0.067	0.087	0.062	0.153	0.270	0.046	0.013	0.012	0.004	0.017
2	0.009	0.003	-0.001	0.022	0.024	0.014	0.025	0.082	0.004	-0.001	0.004	-0.007	0.003
3	-0.006	0.002	-0.005	0.005	0.006	-0.002	0.001	0.023	0.001	-0.002	-0.007	0.000	-0.003
4	-0.004	0.002	0.002	0.006	-0.003	-0.001	0.003	0.007	-0.001	-0.001	0.004	0.004	-0.007
5	-0.001	-0.002	0.000	0.004	-0.006	-0.010	0.003	-0.002	0.002	0.001	-0.002	0.006	-0.001
6	-0.007	-0.004	0.001	-0.005	-0.005	-0.010	0.004	-0.002	-0.005	-0.003	0.004	-0.001	-0.006
7	-0.007	-0.001	-0.012	-0.006	-0.006	0.004	0.004	-0.006	-0.001	-0.001	-0.001	0.001	-0.006
8	-0.004	0.001	-0.004	-0.002	-0.005	0.006	0.002	-0.003	0.004	0.003	-0.007	-0.000	0.001
9	-0.000	-0.002	0.011	-0.002	0.003	0.002	-0.008	-0.002	-0.011	-0.001	-0.004	0.001	-0.000
10	-0.007	-0.009	0.006	-0.002	0.004	-0.002	-0.007	-0.003	-0.006	0.003	0.000	-0.010	-0.009

Table G5 - Deviance Information Criterion (DIC) and effective number of parameters (p_D). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Draws=$51,000$, Burnin draws=$1,000$, thinning=100, Replications=100.

	thinning=1		thinning=100	
	SUR	SURME	SUR	SURME
DIC	2558.322	3406.346	2558.292	3407.118
p_D	-4.273	7.300	-4.249	7.278

Table G6 - Frequentist estimates of SUR model. The table presents the true parameter values (TRUE), regression coefficients (COEF), relative error (RE), standard error (SE) and 95% confidence interval (INF, SUP). $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.8$, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
COEFF	4.157	5.589	4.236	5.127	4.611	3.339	3.221	3.242	1.195	0.501	1.187
RE	0.386	0.118	0.059	0.282	0.213	0.113	-0.195	-0.190	0.195	0.002	0.187
SE	0.348	0.189	0.079	0.347	0.243	0.105	0.206	0.206			
INF	3.473	5.218	4.081	4.446	4.135	3.133	2.818	2.839			
SUP	4.840	5.959	4.392	5.808	5.086	3.544	3.625	3.645			

Table G7 - Bayesian estimation of SUR model. The table presents the posterior mean (MEAN), relative error (RE), standard deviation (STD), inefficiency factor (IF), 95% highest posterior density interval (INF-HPDI, SUP-HPDI) and Geweke's convergence diagnostics (CD). $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.8$, Draws=$51,000$, Burnin draws=$1,000$, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
MEAN	4.153	5.585	4.236	5.122	4.606	3.337	3.224	3.245	1.183	0.493	1.176
RE	0.384	0.117	0.059	0.348	0.535	-0.166	-0.194	-0.189	0.183	-0.014	0.176
STD	0.346	0.188	0.079	0.344	0.242	0.104	0.204	0.205	0.096	0.074	0.096
IF	1.002	1.000	1.000	0.996	1.000	1.003	0.999	1.000	1.001	1.004	1.000
INF-HPDI	3.581	5.276	4.107	4.551	4.205	3.165	2.887	2.906	1.033	0.377	1.026
SUP-HPDI	4.719	5.893	4.365	5.686	4.999	3.508	3.557	3.581	1.348	0.617	1.340
CD	0.990	0.980	1.000	0.970	0.980	0.990	0.980	0.970	0.980	0.970	0.980

Table G8 - Bayesian estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, IF is inefficiency factor, and (INF-HPDI, SUP-HPDI) represents 95% HPDI, and CD represents Geweke's convergence diagnostics. $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.8$, Draws=$51,000$, Burnin draws=$1,000$, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	0.0625	0.0156	1	0.5	1
MEAN	2.119	4.521	3.820	2.976	3.092	2.704	4.600	4.671	0.055	0.022	0.921	0.496	0.910
RE	-0.294	-0.096	-0.045	-0.256	-0.186	-0.099	0.150	0.168	-0.123	0.433	-0.079	-0.008	-0.090
STD	0.487	0.265	0.111	0.477	0.341	0.148	0.311	0.302	0.005	0.004	0.100	0.069	0.099
IF	1.969	1.921	1.871	2.314	2.342	2.318	2.100	2.497	1.475	1.559	1.162	1.002	1.185
INF-HPDI	1.307	4.078	3.634	2.176	2.520	2.459	4.089	4.175	0.047	0.016	0.765	0.387	0.754
SUP-HPDI	2.906	4.950	3.997	3.746	3.642	2.945	5.115	5.169	0.064	0.028	1.090	0.613	1.079
CD	0.950	0.950	0.970	0.880	0.900	0.900	0.960	0.880	0.900	0.920	0.960	0.970	0.940

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.497	0.752	0.301	1.497	1.054	0.448
RE	-0.002	0.002	0.003	-0.002	0.003	-0.003
STD	0.042	0.028	0.014	0.043	0.028	0.014
IF	1.000	0.994	0.998	1.005	1.001	1.001
INF-HPDI	1.427	0.706	0.278	1.426	1.008	0.425
SUP-HPDI	1.567	0.797	0.324	1.567	1.099	0.472
CD	1.000	1.000	0.980	0.980	0.990	0.980

Table G9 - MFVB estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, and (INF-CRI, SUP-CRI) represents 95% credible interval. $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.8$, Replications=100, Cycles = 742.12, Maximum Elbo = -2300.493.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	0.0625	0.0156	1	0.5	1
MEAN	2.490	4.718	3.899	3.555	3.509	2.883	4.345	4.279	0.059	0.018	0.974	0.505	0.973
RE	-0.170	-0.056	-0.025	-0.111	-0.077	-0.039	0.086	0.070	-0.049	0.174	-0.026	0.011	-0.027
STD	0.140	0.097	0.042	0.140	0.097	0.042	0.275	0.226	0.003	0.001	0.074	0.059	0.074
INF-CRI	2.216	4.527	3.817	3.282	3.318	2.801	3.807	3.837	0.053	0.016	0.829	0.390	0.828
SUP-CRI	2.763	4.908	3.981	3.829	3.699	2.965	4.883	4.721	0.066	0.020	1.120	0.621	1.119

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.498	0.751	0.301	1.497	1.053	0.448
RE	-0.001	0.001	0.002	-0.002	0.003	-0.004
STD	0.037	0.024	0.012	0.037	0.024	0.012
INF-CRI	1.425	0.703	0.277	1.424	1.006	0.424
SUP-CRI	1.572	0.799	0.324	1.570	1.101	0.472

Table G10 - Frequentist estimates of SUR model. The table presents the true parameter values (TRUE), regression coefficients (COEF), relative error (RE), standard error (SE) and 95% confidence interval (INF, SUP). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.5714$, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
COEFF	5.502	6.315	4.522	6.508	5.627	3.766	2.293	2.298	7.796	0.544	7.793
RE	0.834	0.263	0.130	0.627	0.481	0.255	-0.427	-0.426	6.796	0.088	6.793
SE	0.466	0.294	0.144	0.465	0.309	0.150	0.122	0.122			
INF	4.588	5.738	4.239	5.597	5.021	3.472	2.053	2.058			
SUP	6.416	6.891	4.805	7.420	6.233	4.060	2.532	2.538			

Table G11 - Bayesian estimation of SUR model. The table presents the posterior mean (MEAN), relative error (RE), standard deviation (STD), inefficiency factor (IF), 95% highest posterior density interval (INF-HPDI, SUP-HPDI) and Geweke's convergence diagnostics (CD). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.5714$, Draws=$51,000$, Burnin draws=$1,000$, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
MEAN	5.498	6.311	4.523	6.503	5.624	3.767	2.294	2.299	7.723	0.536	7.720
RE	0.833	0.262	0.131	0.711	0.875	-0.058	-0.427	-0.425	6.723	0.071	6.720
STD	0.464	0.292	0.143	0.460	0.309	0.149	0.121	0.122	0.626	0.446	0.631
IF	1.002	1.008	1.001	0.987	1.001	1.000	0.999	1.000	1.002	1.002	1.000
INF-HPDI	4.734	5.828	4.288	5.742	5.116	3.520	2.094	2.098	6.746	-0.193	6.740
SUP-HPDI	6.253	6.790	4.758	7.254	6.126	4.010	2.491	2.499	8.800	1.267	8.797
CD	0.970	0.990	0.990	0.990	0.980	0.990	0.970	0.970	0.980	0.970	0.980

Table G12 - Bayesian estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, IF is inefficiency factor, and (INF-HPDI, SUP-HPDI) represents 95% HPDI, and CD represents Geweke's convergence diagnostics. $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.5714$, Draws=51,000, Burnin draws=1,000, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.75	1	0.5	1
MEAN	2.805	4.612	3.947	3.398	3.532	2.992	4.212	4.232	0.919	0.758	1.023	0.502	1.020
RE	-0.065	-0.078	-0.013	-0.151	-0.070	-0.003	0.053	0.058	-0.081	0.010	0.023	0.005	0.020
STD	0.536	0.381	0.191	0.552	0.400	0.199	0.194	0.192	0.086	0.048	0.200	0.152	0.199
IF	1.079	1.102	1.077	1.140	1.189	1.137	1.413	1.622	1.279	1.011	1.012	1.005	1.016
INF-HPDI	1.900	3.970	3.626	2.479	2.861	2.658	3.902	3.925	0.784	0.681	0.738	0.275	0.736
SUP-HPDI	3.674	5.225	4.253	4.286	4.175	3.311	4.539	4.556	1.064	0.840	1.384	0.770	1.380
CD	0.980	0.970	0.970	0.960	0.980	0.960	0.980	0.950	0.950	1.000	0.980	0.970	0.960

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.444	0.807	0.306	1.520	1.073	0.431
RE	-0.037	0.076	0.020	0.013	0.022	-0.042
STD	0.182	0.122	0.062	0.182	0.122	0.063
IF	1.005	1.001	1.002	1.000	1.008	1.001
INF-HPDI	1.142	0.605	0.203	1.218	0.872	0.328
SUP-HPDI	1.741	1.006	0.408	1.816	1.274	0.534
CD	0.980	1.000	0.990	0.980	0.980	0.980

Table G13 - MFVB estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, and (INF-CRI, SUP-CRI) represents 95% credible interval. $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.5714$, Replications=100, Cycles $= 210.580$, Maximum Elbo $= -3808.019$.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.75	1	0.5	1
MEAN	2.996	4.771	4.015	3.745	3.816	3.118	4.046	3.992	1.000	0.738	1.081	0.534	1.097
RE	-0.001	-0.046	0.004	-0.064	0.004	0.039	0.012	-0.002	0.000	-0.016	0.081	0.069	0.097
STD	0.148	0.102	0.045	0.149	0.103	0.045	0.065	0.055	0.058	0.043	0.082	0.065	0.084
INF-CRI	2.705	4.570	3.927	3.453	3.614	3.029	3.918	3.884	0.886	0.654	0.919	0.407	0.933
SUP-CRI	3.286	4.972	4.103	4.038	4.018	3.207	4.174	4.100	1.113	0.822	1.242	0.662	1.261

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.455	0.802	0.302	1.522	1.067	0.426
RE	-0.030	0.069	0.008	0.015	0.016	-0.053
STD	0.151	0.099	0.050	0.151	0.099	0.050
INF-CRI	1.159	0.607	0.205	1.226	0.873	0.329
SUP-CRI	1.750	0.996	0.400	1.818	1.261	0.524

Table G14 - Frequentist estimates of SUR model. The table presents the true parameter values (TRUE), regression coefficients (COEF), relative error (RE), standard error (SE) and 95% confidence interval (INF, SUP). $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.5714$, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
COEFF	5.533	6.281	4.514	6.499	5.577	3.752	2.301	2.324	1.424	0.500	1.413
RE	0.844	0.256	0.128	0.625	0.468	0.251	-0.425	-0.419	0.424	0.000	0.413
SE	0.344	0.189	0.081	0.343	0.238	0.104	0.196	0.196			
INF	4.858	5.910	4.354	5.827	5.111	3.548	1.917	1.940			
SUP	6.208	6.652	4.673	7.171	6.044	3.956	2.685	2.708			

Table G15 - Bayesian estimation of SUR model. The table presents the posterior mean (MEAN), relative error (RE), standard deviation (STD), inefficiency factor (IF), 95% highest posterior density interval (INF-HPDI, SUP-HPDI) and Geweke's convergence diagnostics (CD). $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.5714$, Draws=$51,000$, Burnin draws=$1,000$, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
MEAN	5.527	6.277	4.513	6.492	5.572	3.750	2.305	2.329	1.410	0.492	1.400
RE	0.842	0.255	0.128	0.708	0.857	-0.062	-0.424	-0.418	0.410	-0.015	0.400
STD	0.342	0.188	0.081	0.339	0.237	0.104	0.194	0.195	0.115	0.086	0.114
IF	1.001	1.001	0.999	0.994	0.999	1.002	0.998	1.000	1.001	1.004	1.000
INF-HPDI	4.963	5.967	4.381	5.929	5.179	3.579	1.984	2.006	1.232	0.356	1.222
SUP-HPDI	6.086	6.586	4.645	7.049	5.957	3.920	2.621	2.648	1.606	0.636	1.595
CD	0.990	0.980	1.000	0.970	0.980	0.990	0.980	0.970	0.980	0.970	0.980

Table G16 - Bayesian estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, IF is inefficiency factor, and (INF-HPDI, SUP-HPDI) represents 95% HPDI, and CD represents Geweke's convergence diagnostics. $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.5714$, Draws=51,000, Burnin draws=1,000, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	0.0625	0.0469	1	0.5	1
MEAN	2.048	4.466	3.801	2.866	3.021	2.679	4.657	4.738	0.053	0.054	0.930	0.488	0.913
RE	-0.317	-0.107	-0.050	-0.283	-0.205	-0.107	0.164	0.184	-0.154	0.154	-0.070	-0.023	-0.087
STD	0.545	0.305	0.130	0.531	0.382	0.168	0.348	0.333	0.006	0.005	0.112	0.077	0.110
IF	2.143	2.091	2.004	2.500	2.520	2.475	2.319	2.719	1.646	1.354	1.201	1.005	1.203
INF-HPDI	1.137	3.953	3.582	1.973	2.381	2.399	4.084	4.192	0.044	0.046	0.755	0.367	0.741
SUP-HPDI	2.930	4.954	4.009	3.722	3.635	2.950	5.233	5.288	0.063	0.063	1.121	0.619	1.101
CD	0.930	0.950	0.950	0.880	0.890	0.870	0.950	0.890	0.900	0.920	0.960	0.970	0.930

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.494	0.754	0.301	1.499	1.054	0.448
RE	-0.004	0.005	0.004	-0.001	0.004	-0.005
STD	0.050	0.032	0.016	0.050	0.033	0.016
IF	1.002	1.000	0.998	1.006	1.001	1.006
INF-HPDI	1.412	0.700	0.274	1.416	1.000	0.420
SUP-HPDI	1.576	0.807	0.328	1.580	1.107	0.474
CD	1.000	1.000	0.990	0.980	0.960	0.980

Table G17 - MFVB estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, and (INF-CRI, SUP-CRI) represents 95% credible interval. $N = 300$, $\sigma_Z^2 = 0.0625$, $R_Z = 0.5714$, Replications=100, Cycles = 863.32, Maximum Elbo = -2462.561.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	0.0625	0.0469	1	0.5	1
MEAN	2.502	4.715	3.901	3.563	3.526	2.896	4.340	4.264	0.059	0.049	0.981	0.501	0.984
RE	-0.166	-0.057	-0.025	-0.109	-0.072	-0.035	0.085	0.066	-0.055	0.042	-0.019	0.003	-0.016
STD	0.140	0.098	0.042	0.141	0.098	0.042	0.277	0.228	0.003	0.003	0.075	0.059	0.075
INF-CRI	2.227	4.524	3.818	3.287	3.335	2.813	3.797	3.817	0.052	0.043	0.834	0.385	0.837
SUP-CRI	2.777	4.906	3.984	3.838	3.718	2.979	4.882	4.710	0.066	0.054	1.127	0.617	1.131

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.497	0.752	0.301	1.500	1.053	0.447
RE	-0.002	0.003	0.002	0.000	0.003	-0.007
STD	0.037	0.024	0.012	0.037	0.024	0.012
INF-CRI	1.424	0.705	0.277	1.427	1.005	0.423
SUP-CRI	1.570	0.800	0.324	1.573	1.101	0.471

Table G18 - Deviance Information Criterion (DIC) and effective number of parameters (p_D) for the all the four simulation studies. Replications=100, Draws=$51,000$, Burnin draws=$1,000$, thinning=100.

		$R_Z = 0.8$		$R_Z = 0.5714$	
		SUR	SURME	SUR	SURME
$\sigma_Z^2 = 1$	DIC	2558.292	3407.118	2945.197	3405.877
	p_D	-4.249	7.278	-7.477	7.489
$\sigma_Z^2 = 0.0625$	DIC	1666.708	2132.055	1817.049	2140.954
	p_D	-39.772	21.274	-26.364	25.858

References

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006), "Deviance Information Criteria for Missing Data Models," *Bayesian Analysis*, 1, 651–674.

Chan, J. and Grant, A. D. (2016), "Fast Computation of the Deviance Information Criterion for Latent Variable Models," *Computational Statistics and Data Analysis*, 100, 847–859.

Greenberg, E. (2012), *Introduction to Bayesian Econometrics*, Cambridge University Press, New York.

Li, Y., Zeng, T., and Yu, J. (2012), "Robust Deviance Information Criterion for Latent Variable Models," *Working Paper, Singapore Management University*, pp. 1–45.

Owen, A. B. (2017), "Statistically Efficient Thinning of a Markov Chain Sampler," *Journal of Computational and Graphical Statistics*, 26, 738–744.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society – Series B*, 64, 583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2014), "The Deviance Information Criterion: 12 years on," *Journal of the Royal Statistical Society – Series B*, 76, 485–493.