Appendix of "Maximum Likelihood Algorithm for Spatial Generalized Linear Mixed Models without Numerical Evaluations of Intractable Integrals"

Tonglin Zhang *

A MLE in SLMMs

The maximization step in our method relies on the ML algorithm for the SLMM given by

$$\boldsymbol{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{19}$$

where $\boldsymbol{z} \in \mathbb{R}^n$ is a response vector, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is a design matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a parameter vector for fixed effects, $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$ is an n-dimensional vector for random effects, $\boldsymbol{\omega} \in \mathbb{R}^q$ is a parameter vector for variance components in the random effects, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}^{-1})$ is an n-dimensional vector for random errors, and $\mathbf{W}$ is a known diagonal matrix for weights. We assume that $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are independent. By integrating $\boldsymbol{\gamma}$ out in the h-likelihood function of (19), we obtain the log-likelihood function of the model as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\omega}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log[\det(\mathbf{V}_{\boldsymbol{\omega}})] - \frac{1}{2}(\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}_{\boldsymbol{\omega}}^{-1}(\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta}) \tag{20}$$

where $\mathbf{V}_{\boldsymbol{\omega}} = \boldsymbol{\Sigma}_{\boldsymbol{\omega}} + \mathbf{W}^{-1}$.

We propose a profile likelihood approach to calculate the MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$. Given $\boldsymbol{\omega}$, the conditional MLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} = (\mathbf{X}^\top \mathbf{V}_{\boldsymbol{\omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}_{\boldsymbol{\omega}}^{-1}\mathbf{z}. \tag{21}$$

Put this into (20). We obtain the profile log-likelihood function of the model as

$$\ell_p(\boldsymbol{\omega}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log[\det(\mathbf{V}_{\boldsymbol{\omega}})] - \frac{1}{2}\boldsymbol{z}^\top \mathbf{M}_{\boldsymbol{\omega}}\boldsymbol{z}, \tag{22}$$

*Department of Statistics, Purdue University, 250 North University Street,West Lafayette, IN 47907-2066, Email: tlzhang@purdue.edu

where $\mathbf{M}_{\boldsymbol{\omega}} = \mathbf{V}_{\boldsymbol{\omega}}^{-1} - \mathbf{V}_{\boldsymbol{\omega}}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}_{\boldsymbol{\omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}_{\boldsymbol{\omega}}^{-1}$.

We devise a Newton-Raphson algorithm to compute the maximizer of $\ell_p(\boldsymbol{\omega})$. To implement the Newton-Raphson algorithm, we need to calculate the first-order and the second-order partial derivatives of $\ell_p(\boldsymbol{\omega})$ with respect to the components of $\boldsymbol{\omega}$. The derivation is straightforward. We only display the results below.

The first-order partial derivative of $\ell_p(\boldsymbol{\omega})$ is

$$\frac{\partial \ell_p(\boldsymbol{\omega})}{\partial \omega_j} = -\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_{\boldsymbol{\omega}}^{-1}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_j}\right) + \frac{1}{2}\boldsymbol{z}^\top\mathbf{M}_{\boldsymbol{\omega}}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_j}\mathbf{M}_{\boldsymbol{\omega}}\boldsymbol{z}, \tag{23}$$

where ω_j is the jth component of $\boldsymbol{\omega}$. The second-order partial derivative of $\ell_p(\boldsymbol{\omega})$ is

$$\begin{aligned}\frac{\partial^2 \ell_p(\boldsymbol{\omega})}{\partial \omega_{j_1}\partial \omega_{j_2}} =& \frac{1}{2}\mathrm{tr}\left(\mathbf{V}_{\boldsymbol{\omega}}^{-1}\frac{\partial^2 \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_1}\partial \omega_{j_2}}\right) - \frac{1}{2}\mathrm{tr}\left(\mathbf{V}_{\boldsymbol{\omega}}^{-1}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_1}}\mathbf{V}_{\boldsymbol{\omega}}^{-1}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_2}}\right) + \frac{1}{2}\boldsymbol{z}^\top\mathbf{M}_{\boldsymbol{\omega}}\frac{\partial^2 \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_1}\partial \omega_{j_2}}\mathbf{M}_{\boldsymbol{\omega}}\boldsymbol{z} \\ &- \frac{1}{2}\boldsymbol{z}^\top\mathbf{M}_{\boldsymbol{\omega}}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_1}}\mathbf{M}_{\boldsymbol{\omega}}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_2}}\mathbf{M}_{\boldsymbol{\omega}}\boldsymbol{z} - \frac{1}{2}\boldsymbol{z}^\top\mathbf{M}_{\boldsymbol{\omega}}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_2}}\mathbf{M}_{\boldsymbol{\omega}}\frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_1}}\mathbf{M}_{\boldsymbol{\omega}}\boldsymbol{z}.\end{aligned} \tag{24}$$

The MLE of $\boldsymbol{\omega}$, denoted by $\hat{\boldsymbol{\omega}}$, can be efficiently derived by the Newton-Raphson algorithm. After $\hat{\boldsymbol{\omega}}$ is derived, the MLE of $\boldsymbol{\beta}$ can be quickly derived by $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\omega}}}$. The computation of the first and the second order derivatives does not need any numerical evaluations of HDIIs because only partial derivatives of $\mathbf{V}_{\boldsymbol{\omega}}$ are needed. Because q is usually small, the indirect usage by the profile likelihood approach for the MLEs is more efficient than the direct usage by the likelihood approach. The two approaches provide identical results in parametric models (Murphy and van der Vaart, 2000). Therefore, we recommend using the profile likelihood approach in the computation of the MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$.

B Prediction of Random Effects

By (19), we obtain the joint distribution of $\boldsymbol{z}$ and $\boldsymbol{\gamma}$ as

$$\begin{pmatrix} \boldsymbol{z} \\ \gamma \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{\boldsymbol{\omega}} & \boldsymbol{\Sigma}_{\boldsymbol{\omega}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\omega}} & \boldsymbol{\Sigma}_{\boldsymbol{\omega}} \end{pmatrix}\right]. \tag{25}$$

If $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ are known, then we predict $\boldsymbol{\gamma}$ by

$$\hat{\gamma}_{\boldsymbol{\beta}\boldsymbol{\omega}} = \mathrm{E}(\boldsymbol{\gamma}|\mathbf{z}) = \boldsymbol{\Sigma}_{\boldsymbol{\omega}}\mathbf{V}_{\boldsymbol{\omega}}^{-1}(\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta}). \tag{26}$$

If $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ are unknown, then we replace them by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\omega}}$, respectively, in (26), leading to a similar formulation for $\hat{\gamma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\omega}}}$. It can predict $\boldsymbol{\gamma}$ under (19) with unknown $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$.

C Fisher Information

The Fisher information is the expected value of the negative Hessian matrix of $\ell(\boldsymbol{\beta}, \boldsymbol{\omega})$ divided by n. It can be straightforwardly derived by (20) under (19). The results are

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\omega}) = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\omega}} \\ \mathbf{I}_{\boldsymbol{\omega}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\omega}\boldsymbol{\omega}} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \mathbf{X}^\top \mathbf{V}_{\boldsymbol{\omega}}^{-1} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathrm{tr} \left(\mathbf{V}_{\boldsymbol{\omega}}^{-1} \frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_1}} \mathbf{V}_{\boldsymbol{\omega}}^{-1} \frac{\partial \mathbf{V}_{\boldsymbol{\omega}}}{\partial \omega_{j_2}} \right) \end{pmatrix}, \tag{27}$$

implying that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}^{-1})$ and $\sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_{\boldsymbol{\omega}_0\boldsymbol{\omega}_0}^{-1})$ as $n \to \infty$, where $\boldsymbol{\beta}_0$ and $\boldsymbol{\omega}_0$ are true parameter vectors.

D Proofs

Proof of Theorem 1. By $\ell_h(\boldsymbol{\beta}, \boldsymbol{\omega}) = \log[L_h(\boldsymbol{\beta}, \boldsymbol{\omega})]$ with $L_h(\boldsymbol{\beta}, \boldsymbol{\omega})$, we obtain the h-loglikelihood function of the SGLMM as

$$\ell_h(\boldsymbol{\beta}, \boldsymbol{\omega}) = \boldsymbol{y} * \boldsymbol{\theta} - b(\boldsymbol{\theta}) + c(\boldsymbol{y}) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log[\det(\boldsymbol{\Sigma}_{\boldsymbol{\omega}})] - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\gamma}.$$

Based on the working SLMM

$$\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{28}$$

where $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\omega)$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\gamma}}})$, $\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}} = (z_{1\tilde{\boldsymbol{\beta}}\boldsymbol{\gamma}}, \ldots, z_{n\tilde{\boldsymbol{\beta}}\boldsymbol{\gamma}})^\top$, $z_{i\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}} = \boldsymbol{x}_i^\top \tilde{\boldsymbol{\beta}} + \tilde{\gamma}_i$, $\mathbf{W}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}} = \mathrm{diag}(w_{1\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}, \ldots, w_{n\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}})$, $w_{i\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}^{-1} = (\partial\tilde{\eta}_i/\partial\tilde{\mu}_i)^2 b''(\tilde{\theta}_i)$, $\tilde{\mu}_i = g^{-1}(\tilde{\eta}_i)$, and $\tilde{\theta}_i = \tilde{\eta}_i = \boldsymbol{x}_i^\top \tilde{\boldsymbol{\beta}} + \tilde{\gamma}_i$, we obtain the working h-likelihood function as

$$\begin{aligned} L_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta}, \boldsymbol{\omega}) =& (2\pi)^{-n}[\det(\mathbf{W}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}})]^{\frac{1}{2}}[\det(\boldsymbol{\Sigma}_{\boldsymbol{\omega}})]^{-\frac{1}{2}} \\ & \times \exp\left\{-\frac{1}{2}(\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}^\top - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma})^\top \mathbf{W}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma}) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\gamma}\right\}. \end{aligned}$$

Taking the logarithm of the above, we obtain the working h-loglikelihood function as

$$\begin{aligned} \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta}, \boldsymbol{\omega}) =& -n\log(2\pi) + \frac{1}{2}\log[\det(\mathbf{W}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}})] - \frac{1}{2}\log[\det(\boldsymbol{\Sigma}_{\boldsymbol{\omega}})]^{-\frac{1}{2}} \\ & - \frac{1}{2}(\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}^\top - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma})^\top \mathbf{W}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma}) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\gamma}. \end{aligned}$$

It is enough to prove that the maximizers of $\ell_h(\boldsymbol{\beta}, \boldsymbol{\omega})$ and $\ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta}, \boldsymbol{\omega})$ are identical when $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$, because $(\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\omega}}_{h,\boldsymbol{\gamma}}^\top)^\top = \mathrm{argmin}_{\boldsymbol{\beta},\boldsymbol{\omega}}\, \ell_h(\boldsymbol{\beta}, \boldsymbol{\omega})$ and $(\hat{\boldsymbol{\beta}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}^\top, \hat{\boldsymbol{\omega}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}^\top)^\top = \mathrm{argmin}_{\boldsymbol{\beta},\boldsymbol{\omega}}\, \tilde{\ell}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta}, \boldsymbol{\omega})$.

We first compute the score functions of $\ell_h(\boldsymbol{\beta},\boldsymbol{\omega})$ and $\ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})$. The partial derivative of $\ell_h(\boldsymbol{\beta},\boldsymbol{\omega})$ with respect to β_j for every $j \in \{1,\cdots,p\}$ is

$$\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_j} = \tilde{\boldsymbol{x}}_j^\top(\boldsymbol{y}-\boldsymbol{\mu}), \tag{29}$$

where $\tilde{\boldsymbol{x}}_j$ is the jth column of $\mathbf{X}$. The partial derivative of $\ell_h(\boldsymbol{\beta},\boldsymbol{\omega})$ with respect to ω_j for every $j \in \{1,\cdots,q\}$ is

$$\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \omega_j} = -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\omega}}}{\partial \omega_j}\right) + \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\omega}}}{\partial \omega_j}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\gamma}. \tag{30}$$

The score function of $\ell_h(\boldsymbol{\beta},\boldsymbol{\omega})$ is

$$\dot{\ell}_h(\boldsymbol{\beta},\boldsymbol{\omega}) = \left(\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_1},\cdots,\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_p},\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \omega_1},\cdots,\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \omega_q}\right)^\top, \tag{31}$$

indicating that

$$\dot{\ell}_h(\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}},\hat{\boldsymbol{\omega}}_{h,\boldsymbol{\gamma}}) = \mathbf{0}. \tag{32}$$

The partial derivatives of $\ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})$ with respect to β_j for all $j \in \{1,\cdots,p\}$ is

$$\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_j} = \tilde{\boldsymbol{x}}_j^\top\mathbf{W}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{z}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\gamma}). \tag{33}$$

The partial derivatives of $\ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})$ with respect to ω_j for all $j \in \{1,\cdots,q\}$ is

$$\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \omega_j} = -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\omega}}}{\partial \omega_j}\right) + \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\omega}}}{\partial \omega_j}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\gamma}. \tag{34}$$

The score function of $\ell_h(\boldsymbol{\beta},\boldsymbol{\omega})$ is

$$\dot{\ell}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega}) = \left(\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_1},\cdots,\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_p},\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \omega_1},\cdots,\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \omega_q}\right)^\top, \tag{35}$$

indicating that

$$\dot{\ell}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\beta}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}},\hat{\boldsymbol{\omega}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}) = \mathbf{0} \tag{36}$$

for any $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$.

We next compare (29)–(36) for the relationship between $\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}$, and that between $\hat{\boldsymbol{\omega}}_{h,\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\omega}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}$ under (28) when $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$. We quickly obtain $\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})/\partial \omega_j = \partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})/\partial \omega_j$ by (30) and (34). Therefore, we only need to compare (29) and (33). By $\eta_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}+\gamma_i$ and $\partial\mu_i/\partial\eta_i = \partial\mu_i/\partial\theta_i = b''(\theta_i) = \mathrm{V}(y_i)$, we obtain $w_{i,\boldsymbol{\beta}\boldsymbol{\gamma}}^{-1} = (\partial\eta_i/\partial\mu_i)^2 b''(\theta_i) = \partial\eta_i/\partial\mu_i$.

Thus, (29) is equivalent to

$$\begin{aligned}
\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_j} &= \sum_{i=1}^{n} x_{ij}(y_i-\mu_i) \\
&= \sum_{i=1}^{n} x_{ij}(\frac{\partial \eta_i}{\partial \mu_i})^{-1}[(y_i-\mu_i)\frac{\partial \eta_i}{\partial \mu_i}] \\
&= \sum_{i=1}^{n} x_{ij}(\frac{\partial \eta_i}{\partial \mu_i})^{-1}[\eta_i+(y_i-\mu_i)\frac{\partial \eta_i}{\partial \mu_i}-\boldsymbol{x}_i^\top\boldsymbol{\beta}-\gamma_i] \\
&= \sum_{i=1}^{n} x_{ij}w_{i,\boldsymbol{\beta\gamma}}(z_{i,\boldsymbol{\beta\gamma}}-\boldsymbol{x}_i^\top\boldsymbol{\beta}-\gamma_i).
\end{aligned} \tag{37}$$

We obtain

$$\frac{\partial \ell_h(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_j}\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}},\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}_{h,\boldsymbol{\gamma}}} = \sum_{i=1}^{n} x_{ij}w_{i,\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}\boldsymbol{\gamma}}(z_{i,\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}\boldsymbol{\gamma}}-\hat{\boldsymbol{x}}_i^\top\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}-\gamma_i). \tag{38}$$

We equivalently express (33) as

$$\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij}w_{i,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(z_{i,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}-\boldsymbol{x}_i^\top\boldsymbol{\beta}-\gamma_i) \tag{39}$$

and obtain

$$\frac{\partial \ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})}{\partial \beta_j}\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}},\boldsymbol{\gamma}=\hat{\boldsymbol{\omega}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}} = \sum_{i=1}^{n} x_{ij}w_{i,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(z_{i,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}-\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}-\gamma_i). \tag{40}$$

We then compare (39) and (40) for the solutions of the score functions, specified by (32) and (35), respectively. We find that the solutions are identical when $\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_{h,\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}}=\boldsymbol{\gamma}$. We then draw the conclusion. ◇

Proof of Theorem 2. The objective function given by the E-step of the EM algorithm (Little and Rubin, 2002, P. 168) is $\int \log f_{h,\boldsymbol{\beta\omega}}(\boldsymbol{y},\boldsymbol{\gamma})f_{c,\boldsymbol{\beta\omega}}(\boldsymbol{\gamma}|\boldsymbol{y})d\boldsymbol{\gamma}$. Note that $f_{c,\boldsymbol{\beta\omega}}(\boldsymbol{\gamma}|\boldsymbol{y})$ is a PDF. The value of the integral is between the minimum and the maximum of $\log f_{h,\boldsymbol{\beta\omega}}(\boldsymbol{y},\boldsymbol{\gamma})$ as a function of $\boldsymbol{\gamma}$. Because $\log f_{h,\boldsymbol{\beta\omega}}(\boldsymbol{y},\boldsymbol{\gamma})$ is continuous in $\boldsymbol{\gamma}$, there exists $\tilde{\boldsymbol{\gamma}}$ satisfying the condition, which means existence. The M-step is carried out by the h-likelihood by replacing $\boldsymbol{\beta}$ with $\tilde{\boldsymbol{\gamma}}$. By Theorem 1, the solutions are identical to those given by (11). For normal responses, the EM algorithm provides identical solutions to those given by the likelihood method after treating missing values (Little and Rubin, 2002, P. 172). Thus, the final solutions are identical to those given by (13). ◇

Proof of Theorem 3. Both $\ell_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})$ and $\ell_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}(\boldsymbol{\beta},\boldsymbol{\omega})$ are the true log-likelihood functions. They satisfy the Shannon-Kolmogorov Information inequality (Ferguson, 1996, P. 113). Because $\boldsymbol{y}$ follows an exponential family distribution and the prior distribution for $\boldsymbol{\gamma}$ is normal, all the regularity conditions for consistency and asymptotic normality (e.g., the conditions given by Chapter

17 in Ferguson (1996) or Section 5.2 in van der Vaart (1998)) are satisfied. Thus, $\hat{\boldsymbol{\beta}}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}$ and $\hat{\boldsymbol{\beta}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}$ are $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\omega}}_{\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}$ and $\hat{\boldsymbol{\omega}}_{h,\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\gamma}}}$ are $\sqrt{n}$-consistent estimators of $\boldsymbol{\omega}$ under (10), implying the conclusion. $\diamondsuit$

Proof of Theorem 4. Because $\boldsymbol{y}$ follows an exponential family distribution, the usual regularity conditions for consistency of the MLEs given by Theorem 17 of Ferguson (1996) are satisfied. To apply the method in the proof of Theorem 17 of Ferguson (1996), we need the Lebesgue Dominate Theorem and an enhanced version of the strong low of large numbers. They are assumed by (ii) and (iii), respectively. Then, we can apply the method in the proof of Theorem 17 of Ferguson (1996), leading to $\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}_h^{(t)} \xrightarrow{P} 0$ and $\boldsymbol{\omega}^{(t)} - \boldsymbol{\omega}_h^{(t)} \xrightarrow{P} 0$ for any fixed t as $n \to \infty$. We derive the final conclusion by adopting the same method for the asymptotic normality of MLEs used in the proof of Theorem 18 of Ferguson (1996). $\diamondsuit$

Proof of Corollary 1. Asymptotic normality is implied by the Lyapunov condition with consistency. Note that $\hat{\boldsymbol{\beta}}_{PM}$ and $\hat{\boldsymbol{\omega}}_{PM}$ are the MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ with unobserved $\boldsymbol{\gamma}$, and $\hat{\boldsymbol{\omega}}_h$ and $\hat{\boldsymbol{\omega}}_h$ are those with observed $\boldsymbol{\gamma}$. We can use the formulation for the relationship between conditional and unconditional variance-covariance, i.e., $\mathrm{V}(\hat{\boldsymbol{\beta}}_{PM}) = \mathrm{E}[\mathrm{V}(\hat{\boldsymbol{\beta}}_{PM}|\boldsymbol{\gamma})] + \mathrm{V}[\mathrm{E}(\hat{\boldsymbol{\beta}}_{PM}|\boldsymbol{\gamma})]$ and $\mathrm{V}(\hat{\boldsymbol{\omega}}_{PM}) = \mathrm{E}[\mathrm{V}(\hat{\boldsymbol{\omega}}_{PM}|\boldsymbol{\gamma})] + \mathrm{V}[\mathrm{E}(\hat{\boldsymbol{\omega}}_{PM}|\boldsymbol{\gamma})]$. Then, we obtain the Fisher Information given by Appendix C. $\diamondsuit$

Proof of Corollary 2. By the Cramér-Rao Lower Bound Theorem (e.g., Ferguson (1996) Page 129), the variance-covariance matrix provided by Corollary 1 is identical to the variance-covariance matrix provided by the Fisher Information because it is optimized according to the connection between the conditional and unconditional variance-covariance matrices. Note that none of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\omega}}$, $\hat{\boldsymbol{\beta}}_{PM}$, and $\hat{\boldsymbol{\beta}}_{PM}$ depend on $\boldsymbol{\gamma}$. Using the standard Talyor expansion for the derivation of the asymptotics of the MLE with Corollary 1 and the properties of the Linderberg-Feller condition, we have

$$\begin{aligned} 0 =& \nabla \ell_{\hat{\boldsymbol{\beta}}_{PM}\hat{\boldsymbol{\omega}}_{PM}}(\hat{\boldsymbol{\beta}}_{PM}, \hat{\boldsymbol{\omega}}_{PM}) \\ =& \nabla \ell_{\boldsymbol{\beta}_0\boldsymbol{\omega}_0}(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0) + \nabla^2 \ell_{\boldsymbol{\beta}_0\boldsymbol{\omega}_0}(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0) \left[\begin{pmatrix} \hat{\boldsymbol{\beta}}_{PM} \\ \hat{\omega}_{PM} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\omega}_0 \end{pmatrix} \right] + o_p(\sqrt{n}) \end{aligned}$$

and

$$\begin{aligned} 0 =& \nabla \ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\omega}}) \\ =& \nabla \ell_{\boldsymbol{\beta}_0\boldsymbol{\omega}_0}(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0) + \nabla^2 \ell_{\boldsymbol{\beta}_0\boldsymbol{\omega}_0}(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0) \left[\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\omega} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\omega}_0 \end{pmatrix} \right] + o_p(\sqrt{n},) \end{aligned}$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\omega}_0$ are the true parameter vectors. Take the difference of the above, we have

$$\left\{\frac{1}{n}\nabla^2 \ell_{\boldsymbol{\beta}_0\boldsymbol{\omega}_0}(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0)\right\}\left\{\sqrt{n}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}}_{PM} \\ \hat{\omega}_{PM} \end{pmatrix} - \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\omega} \end{pmatrix}\right]\right\} + o_p(1) = 0.$$

Note that $-n^{-1}\nabla^2 \ell_{\boldsymbol{\beta}_0\boldsymbol{\omega}_0}(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0)$ is the Fisher Information matrix. It is positive definite. We obtain

$$\sqrt{n}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}}_{PM} \\ \hat{\omega}_{PM} \end{pmatrix} - \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\omega} \end{pmatrix}\right] \xrightarrow{P} 0$$

as $n \to \infty$, leading to the conclusion. $\diamondsuit$

References

Ferguson, T.S. (1996). *A Course in Large Sample Theory*, Chapman & Hall/CRC Press, Boca Raton, Florida.

Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, Wiley, New Jersey, USA.

Murphy, A. and van der Vaart, A.W. (2000). On profile likelihood. *Journal of the American Statistical Association*, **95**, 449-465.

van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK.