Web Appendix of "Conditional Inference Functions for Mixed-Effects Models with Unspecified Random-Effects Distribution"

A.1. Notation

We denote the estimate of the random effects as $\hat{\mathbf{b}}$, and let $Q(\boldsymbol{\beta}|\mathbf{b}_0)$ be the quadratic inference function defined in (4) conditional on the true random effects $\mathbf{b}_0$,

$$\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) = \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}|\mathbf{b}_0)\Big|_{\beta=\hat{\beta}_0},$$

and $\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_0|\hat{\mathbf{b}})$, $\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)$, and $\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}})$ can be defined similarly. In addition, let

$$\ddot{Q}_{\beta\beta}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}|\mathbf{b}_0)\Big|_{\beta=\hat{\beta}_0},$$

$$\ddot{Q}_{\beta\mathbf{b}}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \mathbf{b}} Q(\boldsymbol{\beta}|\mathbf{b}_0)\Big|_{\beta=\hat{\beta}_0,\ \mathbf{b}=\mathbf{b}_0},$$

and $\ddot{Q}_{\beta\beta}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)$ and $\ddot{Q}_{\beta\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)$ are defined similarly. Let $\mathbf{G}_N(\boldsymbol{\beta}|\mathbf{b}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i(\boldsymbol{\beta}|\mathbf{b}_i)$. We can define

$$\dot{\mathbf{G}}_{N,\beta}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) = \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{G}_{N,\beta}(\boldsymbol{\beta}|\mathbf{b}_0)\Big|_{\beta=\hat{\beta}_1},$$

$$\dot{\mathbf{G}}_{N,\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) = \frac{\partial}{\partial \mathbf{b}} \mathbf{G}_{N,b}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)\Big|_{\mathbf{b}=\mathbf{b}_0},$$

$$\text{and } \dot{\mathbf{G}}_{N,\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) = \frac{\partial}{\partial \mathbf{b}} \mathbf{G}_{N,b}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)\Big|_{\mathbf{b}=\mathbf{b}_0}.$$

The other second derivatives associated with the different parameters are defined in the same fashion. Let

$$\hat{\boldsymbol{\beta}}_0 = \arg\min Q(\boldsymbol{\beta}|\mathbf{b});\ \hat{\boldsymbol{\beta}}_1 = \arg\min Q(\boldsymbol{\beta}|\hat{\mathbf{b}}).$$

Both $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ are in S, that is,

$$\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) = 0, \quad \dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}) = 0.$$

Also let $\mathbf{A}_N(\boldsymbol{\beta}|\mathbf{b})$ be the weighting matrix such that

$$\mathbf{C}_N^{-1}(\boldsymbol{\beta}|\mathbf{b}) = \mathbf{A}_N(\boldsymbol{\beta}|\mathbf{b})'\mathbf{A}_N(\boldsymbol{\beta}|\mathbf{b})$$

and $Q(\boldsymbol{\beta}|\mathbf{b}) = |\mathbf{A}_N(\boldsymbol{\beta}|\mathbf{b})\mathbf{G}_N(\boldsymbol{\beta}|\mathbf{b})|^2$.

A.2. Regularity conditions and assumptions

Here we prove consistency and asymptotic normality for the fixed-effect estimator under the following assumptions.

1. Define n_i as the cluster size for subject i, let $n = \min(n_i)$, then $n_i = O_p(n)$ uniformly for $i = 1, \ldots, N$.
2. The parameter space S is compact.
3. Conditional on the true random effects $\mathbf{b}_0$, the parameter $\boldsymbol{\beta}$ is identifiable; that is, there is a unique $\boldsymbol{\beta}_0 \in S$ which satisfies $E\{\mathbf{g}(\boldsymbol{\beta}_0|\mathbf{b}_0)\} = 0$.
4. The derivative of the score function with respect to the random effects $\dot{\mathbf{g}}_{i,\mathbf{b}}(\hat{\boldsymbol{\beta}}|\mathbf{b}_0)$ is uniformly bounded in probability, i.e. $\dot{\mathbf{g}}_{i,\mathbf{b}}(\hat{\boldsymbol{\beta}}|\mathbf{b}_0) = O_p(1)$.
5. We require that $E[\mathbf{g}(\boldsymbol{\beta}|\mathbf{b})]$ be continuous and differentiable in both $\boldsymbol{\beta}$ and $\mathbf{b}$.
6. The expectation of $\mathbf{g}_i(\boldsymbol{\beta}_0|\hat{\mathbf{b}})$, the estimating functions conditional on the estimated random effects, converges to 0 in probability, i.e.

$$E[E\{\mathbf{g}_i(\boldsymbol{\beta}_0|\hat{\mathbf{b}})\}] \xrightarrow{p} 0 \quad \texttt{as } N \to \infty.$$

7. The weighting matrix $\mathbf{C}_N(\boldsymbol{\beta}|\mathbf{b})$ converges almost surely to a constant matrix $\mathbf{C}_0(\boldsymbol{\beta}|\mathbf{b})$, while $\mathbf{A}_N(\boldsymbol{\beta}|\mathbf{b})$ converges almost surely to a constant matrix $\mathbf{A}_0(\boldsymbol{\beta}|\mathbf{b})$ where $\mathbf{C}_0^{-1}(\boldsymbol{\beta}|\mathbf{b}) = \mathbf{A}_0(\boldsymbol{\beta}|\mathbf{b})\mathbf{A}_0(\boldsymbol{\beta}|\mathbf{b})'$.

A.3. Proofs of Lemmas and Theorem 1

Proof of Lemma 1. Define $B_N(r, \boldsymbol{\beta}_0) = \{\boldsymbol{\beta} \| \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < r/\sqrt{N}\}$ for a fixed constant r. Then by Taylor expansion, we have

$$\sup_{\boldsymbol{\beta} \in B_N(r,\boldsymbol{\beta}_0)} |\sqrt{N}\{\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\hat{\mathbf{b}}) - \dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})\}| = \sup_{\boldsymbol{\beta} \in B_N(r,\boldsymbol{\beta}_0)} |\sqrt{N}\ddot{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| + o_p(1).$$

Since $\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\hat{\mathbf{b}}) = \dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\hat{\mathbf{b}}) - \dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) + \dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})$, we have

$$\sup_{\boldsymbol{\beta} \in B_N(r,\boldsymbol{\beta}_0)} |\sqrt{N}\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\hat{b}) - \sqrt{N}\ddot{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - \sqrt{N}\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})| = o_p(1). \qquad \text{(A-1)}$$

Further, when $\boldsymbol{\beta}$ is on the boundary of $B_N(r, \boldsymbol{\beta}_0)$, i.e. $\boldsymbol{\beta} \in \{\boldsymbol{\beta} \| \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = r/\sqrt{N}\}$,

$$N(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\ddot{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{b})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = O(r^2) > 0$$

since $\ddot{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})$ is positive-definite and uniformly bounded.

In addition, by the weak law of large numbers and Condition 3, $\sqrt{N}\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0) = O_p(1)$, since

$$\sqrt{N}\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) = \sqrt{N}\dot{\mathbf{G}}_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})\mathbf{C}_N^{-1}(\hat{\mathbf{b}})\mathbf{G}_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) + o_p(1).$$

It can be concluded by Conditions 4 and 6 that

$$\sqrt{N}\{\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) - \dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\} = O_p(1).$$

Hence it follows from the above that

$$\sqrt{N}\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) = \sqrt{N}\{\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) - \dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\} + \sqrt{N}\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0) = O_p(1),$$

which leads to $N(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) = O_p(r)$. Therefore for any $\epsilon > 0$, there exists an M, such that when $r > M$,

$$P\{N(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\ddot{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + N(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) > 0\} > 1 - \epsilon \qquad \text{(A-2)}$$

for all $\boldsymbol{\beta}$ on the boundary of $B_N(r, \boldsymbol{\beta}_0)$. Therefore (A-2) certainly holds for all $\boldsymbol{\beta} \notin B_N(r, \boldsymbol{\beta}_0)$.

It follows from (A-2) and (A-1) that

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\hat{\mathbf{b}}) > 0 \qquad \text{(A-3)}$$

for $\boldsymbol{\beta} \notin B_N(r, \boldsymbol{\beta}_0)$ and some sufficiently large but finite r. Since the left-hand side of (A-3) is continuous for $\boldsymbol{\beta}$, by theorem (6.3.4) of Ortega and Rheinboldt (1973, p. 163), there must be a solution in $B_N(r, \boldsymbol{\beta}_0)$ satisfying

$$\dot{Q}_{\beta}(\boldsymbol{\beta}|\hat{\mathbf{b}}) = 0.$$

$\square$

Proof of Lemma 2. Since $\hat{\boldsymbol{\beta}}_0 = \texttt{arg}\min Q(\boldsymbol{\beta}|\mathbf{b}_0)$,

$$|Q(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)|^2 < |Q(\boldsymbol{\beta}_0|\mathbf{b}_0)|^2.$$

That is,

$$|\mathbf{A}_N(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)\frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_i(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)|^2 < |\mathbf{A}_N(\boldsymbol{\beta}_0|\mathbf{b}_0)\frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_i(\boldsymbol{\beta}_0|\mathbf{b}_0)|^2.$$

By the law of large numbers, we know that the right side of the above converges to 0 as $E[\mathbf{g}(\boldsymbol{\beta}_0|\mathbf{b}_0)] = 0$. Further, by Assumption 8, the uniform law of large numbers and the continuity mapping theorem, we can prove that

$$|\mathbf{A}_N(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)\frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_i(\hat{\boldsymbol{\beta}}_0) - \mathbf{A}_0(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)E[\mathbf{g}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)]| \rightarrow_{a.s.} 0.$$

It follows that

$$|A_0(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)E[\mathbf{g}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)]|^2 \rightarrow_{a.s.} 0.$$

Hence $\hat{\boldsymbol{\beta}}_0$ converges to $\boldsymbol{\beta}_0$ almost surely. $\square$

Proof of Lemma 3. Since $\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) = \dot{\mathbf{G}}_{N,\beta}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)'\mathbf{C}_N^{-1}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)\mathbf{G}_N(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0)$, by Taylor's Expansion,

$$\begin{aligned} 0 &= \dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) = \dot{Q}_{\beta}(\boldsymbol{\beta}_0|\mathbf{b}_0) + \ddot{Q}_{\beta\beta}(\tilde{\boldsymbol{\beta}}|\mathbf{b}_0)(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \\ &= \dot{\mathbf{G}}_{N,\beta}(\boldsymbol{\beta}_0|\mathbf{b}_0)'\mathbf{C}_N^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0)\mathbf{G}_N(\boldsymbol{\beta}_0|\mathbf{b}_0) + \ddot{Q}_{\beta\beta}(\tilde{\boldsymbol{\beta}}|\mathbf{b}_0)(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0), \end{aligned}$$

where $\tilde{\boldsymbol{\beta}}$ is between $\hat{\boldsymbol{\beta}}_0$ and $\boldsymbol{\beta}_0$. Then we have

$$\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0 = -\ddot{Q}_{\beta\beta}^{-1}(\tilde{\boldsymbol{\beta}}|\mathbf{b}_0)\dot{\mathbf{G}}_{N,\beta}(\boldsymbol{\beta}_0|\mathbf{b}_0)'\mathbf{C}_N^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0)\mathbf{G}_N(\boldsymbol{\beta}_0|\mathbf{b}_0).$$

Since $\hat{\boldsymbol{\beta}}_0 \rightarrow_{a.s.} \boldsymbol{\beta}_0$, it follows immediately that

$$\tilde{\boldsymbol{\beta}} \rightarrow_{a.s.} \boldsymbol{\beta}_0, \texttt{and } \dot{\mathbf{G}}_{N,\beta}(\tilde{\boldsymbol{\beta}}|\mathbf{b}_0) \rightarrow_{a.s.} \mathbf{d}_0.$$

By the Central Limit Theorem and Assumption 3, $\sqrt{N}\mathbf{G}_N(\boldsymbol{\beta}_0|\mathbf{b}_0) \xrightarrow{d} N(0,\boldsymbol{\Sigma})$ and $\mathbf{C}_N(\boldsymbol{\beta}_0|\mathbf{b}_0) \to_p \boldsymbol{\Sigma} = N\boldsymbol{\Sigma}_N$. Therefore $\sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)$ converges to a normal distribution of mean 0 with asymptotic covariance matrix

$$
\begin{aligned}
&\texttt{cov}(\sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)) \\
&= \ddot{Q}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\dot{\mathbf{G}}'_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\mathbf{C}_N^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0)\boldsymbol{\Sigma}\mathbf{C}_N^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0)\dot{\mathbf{G}}_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\ddot{Q}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0) \\
&\to (\mathbf{d}_0'\boldsymbol{\Sigma}^{-1}\mathbf{d}_0)^{-1}\mathbf{d}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{d}_0(\mathbf{d}_0'\boldsymbol{\Sigma}^{-1}\mathbf{d}_0)^{-1} = (\mathbf{d}_0\boldsymbol{\Sigma}\mathbf{d}_0)^{-1} = \boldsymbol{\Omega}_0. \qquad \text{(A-4)}
\end{aligned}
$$

This is because $\ddot{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0) = \dot{\mathbf{G}}_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)'\mathbf{C}_N^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0)\dot{\mathbf{G}}_{N,\beta}(\boldsymbol{\beta}_0|\mathbf{b}_0) + o_p(1)$. Hence it follows immediately that $\ddot{Q}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_0|\mathbf{b}_0) \to_{a.s.} \boldsymbol{\Omega}_0$. □

Proof of Theorem 1. Consistency of $\hat{\boldsymbol{\beta}}_1$ follows immediately from Lemma 1. By Lemma 3, $\sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)$ also converges to the normal distribution. Furthermore,

$$
\begin{aligned}
&\sqrt{N}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) = \sqrt{N}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0) + \sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \\
&= \sqrt{N}\ddot{Q}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}|\mathbf{b}_0)\dot{Q}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) - \sqrt{N}\ddot{Q}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\dot{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0) + o_p(1) \qquad \text{(A-5)} \\
&= \sqrt{N}\ddot{Q}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\dot{\mathbf{G}}_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\mathbf{C}_N^{-1}(\mathbf{b}_0)1/N\sum_{i=1}^{N}[\mathbf{g}_i(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) - \mathbf{g}_i(\boldsymbol{\beta}_0|\mathbf{b}_0)] + o_p(1).
\end{aligned}
$$

Define $\boldsymbol{\Sigma}^*$ as

$$
\boldsymbol{\Sigma}^* = \lim_{N\to\infty} E[N\{\mathbf{G}_N(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) - \mathbf{G}_N(\boldsymbol{\beta}_0|\mathbf{b}_0)\}\{\mathbf{G}_N(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) - \mathbf{G}_N(\boldsymbol{\beta}_0|\mathbf{b}_0)\}']. \qquad \text{(A-6)}
$$

Hence, the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)$ can be written as

$$
\boldsymbol{\Omega}_1 = (\mathbf{d}_0'\boldsymbol{\Sigma}^{-1}\mathbf{d}_0)^{-1}\mathbf{d}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}\mathbf{d}_0(\mathbf{d}_0'\boldsymbol{\Sigma}^{-1}\mathbf{d}_0)^{-1}. \qquad \text{(A-7)}
$$

When the estimate of the random effects is consistent, i.e. $\hat{\mathbf{b}} \to_p \mathbf{b}_0$ as $n \to \infty$, it can be shown that

$$
\sqrt{N}\dot{\mathbf{G}}_{N,\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}})(\hat{\mathbf{b}} - \mathbf{b}_0) = O_p(1)o_p(1) = o_p(1).
$$

Therefore,

$$
\begin{aligned}
\sqrt{N}\ddot{Q}_{\boldsymbol{\beta}\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}})(\hat{\mathbf{b}} - \mathbf{b}_0) &= \sqrt{N}\{\dot{\mathbf{G}}_{N,\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}})\}'\mathbf{C}_N^{-1}(\tilde{\mathbf{b}})\dot{\mathbf{G}}_{N,\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}})(\hat{\mathbf{b}} - \mathbf{b}_0) + o_p(1) \\
&= \sqrt{N}\{\dot{\mathbf{G}}_{N,\boldsymbol{\beta}}(\boldsymbol{\beta}_0|\mathbf{b}_0)\}'\mathbf{C}_N^{-1}(\mathbf{b}_0)\dot{\mathbf{G}}_{N,\mathbf{b}}(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}})(\hat{\mathbf{b}} - \mathbf{b}) + o_p(1) \\
&= o_p(1).
\end{aligned}
$$

Then by Taylor expansion, we have

$$\sqrt{N}\{\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}) - \dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)\} = \sqrt{N}\ddot{Q}_{\beta b}(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}})(\hat{\mathbf{b}} - \mathbf{b}_0) = o_p(1).$$

It follows immediately from $\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}) = 0$ that

$$\sqrt{N}\dot{Q}_{\beta}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) = \sqrt{N}\dot{\mathbf{G}}_{N,\beta}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)'\mathbf{C}_N^{-1}(\mathbf{b}_0)\mathbf{G}_N(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0) + o_p(1) = o_p(1). \tag{A-8}$$

Then by (A-5) and (A-8), we can conclude that

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) = \sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) + o_p(1).$$

Hence it follows from (A-4) that

$$\boldsymbol{\Omega}_1 = \ddot{Q}_{\beta\beta}^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0) + o_p(1),$$

which can be approximated by $\ddot{Q}_{\beta\beta}^{-1}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}})$ since $\ddot{Q}_{\beta\beta}^{-1}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}) \xrightarrow{p} \ddot{Q}_{\beta\beta}^{-1}(\boldsymbol{\beta}_0|\mathbf{b}_0)$. □

A.4. Conditions and proof of consistency of random-effect estimator

We estimate $b_{0,i}$ by solving

$$\mathbf{g}_i^*(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}_i) = \dot{\mu}_{i,b}(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_1|\hat{\mathbf{b}}_i)) = 0.$$

Therefore, by Taylor expansion we have

$$\hat{\mathbf{b}}_i - \mathbf{b}_{0i} = \{\dot{\mathbf{g}}_{i,\mathbf{b}_i}^*(\hat{\boldsymbol{\beta}}_1|\tilde{\mathbf{b}}_i)\}^{-1}\sum_{j=1}^{n_i}\dot{\mu}_{ij,b}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)(y_{ij} - \mu_{ij}(\hat{\boldsymbol{\beta}}_1|\mathbf{b}_0)).$$

Since $\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_0$, then

$$\hat{\mathbf{b}}_i - \mathbf{b}_{0i} \to \{\dot{\mathbf{g}}_{i,\mathbf{b}_i}^*(\boldsymbol{\beta}_0|\tilde{\mathbf{b}}_i)\}^{-1}\sum_{j=1}^{n_i}\dot{\mu}_{ij,b}(\boldsymbol{\beta}_0|\mathbf{b}_0)(y_{ij} - \mu_{ij}(\boldsymbol{\beta}_0|\mathbf{b}_0)).$$

Since $\{\dot{\mathbf{g}}_{i,\mathbf{b}_i}^*(\boldsymbol{\beta}_0|\tilde{\mathbf{b}}_i)\}^{-1}$ is bounded in probability, therefore if the law of large numbers holds for the sequence $\dot{\mu}_{i1,b}(\boldsymbol{\beta}_0|\mathbf{b}_0)\{y_{i1} - \mu_{i1}(\boldsymbol{\beta}_0|\mathbf{b}_0)\}, \ldots, \dot{\mu}_{in_i,b}(\boldsymbol{\beta}_0|\mathbf{b}_0)\{y_{in_i} - \mu_{in_i}(\boldsymbol{\beta}_0|\mathbf{b}_0)\}$, we can conclude that

$$\hat{\mathbf{b}}_i - \mathbf{b}_{0i} = O_p(n_i^{-1/2}).$$

That is, $\hat{\mathbf{b}}$ is a consistent estimator of $\mathbf{b}_0$. This is because $E\{\dot{\mu}_{ij,b}(\boldsymbol{\beta}_0|\mathbf{b}_0)(y_{ij}-\mu_{ij}(\boldsymbol{\beta}_0|\mathbf{b}_0))\}=0$.

Let $Z_{ij}=\dot{\mu}_{ij,b}(\boldsymbol{\beta}_0|\mathbf{b}_0)\{y_{ij}-\mu_{ij}(\boldsymbol{\beta}_0|\mathbf{b}_0)\}$. From Andrews (1988), if the sequence of random variables satisfies the L_1 mixingale conditions:

(a) $||E(Z_{ij}|Z_{i,j-m}||_1 \leq c_j\psi_m$, and

(b) $||Z_i - E(Z_{ij}|Z_{i,j+m})||_1 \leq c_j\psi_{m+1}$,

where $\{c_j : i \geq 1\}$ and $\{\psi_m : m \geq 0\}$ are some non-negative constants and $\psi_m \to 0$ as $m \to \infty$, and if $\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} c_j < \infty$ or $\{c_j\}$ can be given by $\{||Z_{ij}||_1\}$, we have the law of large numbers for the dependent sequence $\bar{Z}_i = 1/n_i\sum_{j=1}^{n_i} Z_{ij} \rightarrow_p 0$. Such conditions can be satisfied for sequences such as autoregressive, stationary Gaussian, or M-dependent and other sequences with decaying α mixing numbers.

References

Andrews, D.W.K. (1998), “Laws of large numbers for dependent non-identically distributed random variables,” *Econometric Theory*, 4, 458-467.

Ortega, J. M. and Rheinboldt, W. C. (1973), *Iterative solution of nonlinear equations in several variables*, Academic Press.