The Canadian Journal of Statistics
Vol. xx, No. yy, 20??, Pages 1
La revue canadienne de statistique

Supplementary Materials to "Longitudinal Data Analysis Using the Conditional Empirical Likelihood Method"

Peisong Han[1]*, Peter X.-K. Song[2] and Lu Wang[2]

[1]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada*
[2]*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA*

1. TECHNICAL ASSUMPTIONS AND PROOFS

This section contains technical assumptions and proofs. These assumptions and proofs rely heavily on the results in Kitamura, Tripathi, & Ahn (2004). To better demonstrate the main steps of the proofs, some tedious theoretical details are omitted, which can be found in Kitamura, Tripathi, & Ahn (2004). Hereafter, let $\mathcal{B}$ denote the domain of $\boldsymbol{\beta}$, and let $\mathcal{B}_0 \subseteq \mathcal{B}$ be some closed ball around $\boldsymbol{\beta}_0$. For a matrix $\boldsymbol{A}$ with elements $\boldsymbol{A}_{ij}$, define $\| \boldsymbol{A} \| = \sqrt{\sum_{i,j} \boldsymbol{A}_{ij}^2}$.

Assumptions:

(i) *There exists $\omega_0 \in \Omega$, such that for any $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, we can find $\mathcal{X}_{\boldsymbol{\beta}}$, so that $P(\boldsymbol{x} \in \mathcal{X}_{\boldsymbol{\beta}}) > 0$ and $E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{X}_i = \boldsymbol{x}\} \neq \boldsymbol{0}$ for every $\boldsymbol{x} \in \mathcal{X}_{\boldsymbol{\beta}}$ and $i \in S^{\omega_0}$.*
(ii) *For any $1 \leq i \leq N$, $E\left\{\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\boldsymbol{g}_i(\boldsymbol{\beta})\|^m\right\} < \infty$ for some $m \geq 8$.*
(iii) *$\mathcal{B}$ is compact.*
(iv) *$\mu(\cdot)$ is continuously differentiable on $\mathcal{B}_0$.*
(v) *For any $\|\boldsymbol{\xi}\| = 1$, $0 < \inf_{\boldsymbol{X}_i,\boldsymbol{\beta}\in\mathcal{B}_0} \boldsymbol{\xi}^\top \boldsymbol{V}_i(\boldsymbol{\beta})\boldsymbol{\xi} \leq \sup_{\boldsymbol{X}_i,\boldsymbol{\beta}\in\mathcal{B}_0} \boldsymbol{\xi}^\top \boldsymbol{V}_i(\boldsymbol{\beta})\boldsymbol{\xi} < \infty$.*
(vi) *The range of $\boldsymbol{X}^c$ is compact.*
(vii) *$b_N \to 0$, $N^{1-2\nu-2/\delta} b_N^{2q} \to \infty$ and $N^{1-2\nu} b_N^{5q/2} \to \infty$ as $N \to \infty$, where $\nu \in (0, 1/2)$, $\delta \geq 8$ and $q = \max_i q_i$.*
(viii) *$\hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}) \in \left\{\boldsymbol{\lambda}_i \in \mathcal{R}^{n_i} : \|\boldsymbol{\lambda}_i\| \leq c|S_i|^{-1/m}\right\}$ for some $c > 0$ for $\boldsymbol{\beta} \in \mathcal{B}_0$.*
(ix) *$E\left\{\sup_{\boldsymbol{\beta}\in\mathcal{B}_0} \|\boldsymbol{G}_i(\boldsymbol{\beta})\|^2\right\} < \infty$.*

Remark: *Assumption (i) guarantees the identifiability of $\boldsymbol{\beta}_0$. Assumption (v) guarantees that the variance-covariance matrix restricted on each stratum is invertible. The restrictions on b_N in Assumption (vii) follow that in Smith (2007). Assumption (viii) is similar to Assumption 3.6 in Kitamura, Tripathi, & Ahn (2004). Since Lemma D.2 in Kitamura, Tripathi, & Ahn (2004) shows that $\max_{1\leq i\leq N} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0} \| \boldsymbol{g}_i(\boldsymbol{\beta}) \| = o_p(N^{1/m})$, Assumption (viii) ensures that*

$$\max_{1\leq i\leq N,\, j\in S_i} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0} |\hat{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta})\boldsymbol{g}_j(\boldsymbol{\beta})| = o_p(1), \tag{1}$$

which will be used in the proofs of Theorem 2 and Theorem 3. Actually, from (6) and (7) below and Assumption (vii), Assumption (viii) is true in a small neighborhood of $\boldsymbol{\beta}_0$. Here we explicitly make Assumption (viii) so that (1) can be readily used in the proofs. Assumption (ix) is used to guarantee the uniform weak law of large numbers in the proof of Theorem 3.

* *Author to whom correspondence may be addressed.*
E-mail: peisonghan@uwaterloo.ca

Proof of Theorem 1. Let

$$L_N(\boldsymbol{\beta}) \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{j\in S_i} -w_{ij}\log\left\{1+\hat{\boldsymbol{\lambda}}_i^{\top}(\boldsymbol{\beta})\boldsymbol{g}_j(\boldsymbol{\beta})\right\}\right] \tag{2}$$

$$= \sum_{\omega\in\Omega}\frac{|S^\omega|}{N}\frac{1}{|S^\omega|}\left[\sum_{i,j\in S^\omega} -w_{ij}\log\left\{1+\hat{\boldsymbol{\lambda}}_i^{\top}(\boldsymbol{\beta})\boldsymbol{g}_j(\boldsymbol{\beta})\right\}\right] \overset{\text{def}}{=} \sum_{\omega\in\Omega}\frac{|S^\omega|}{N}L_N^\omega(\boldsymbol{\beta}),$$

then $\hat{\boldsymbol{\beta}}_{CEL} = \arg\max_{\boldsymbol{\beta}\in\mathcal{B}} L_N(\boldsymbol{\beta})$. Let

$$\boldsymbol{u}_i(\boldsymbol{\beta}) = \frac{E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}}{1+\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|}$$

and $\boldsymbol{g}^b(\boldsymbol{\beta}) = \boldsymbol{g}(\boldsymbol{\beta})I\{(\boldsymbol{Y},\boldsymbol{X})\in\mathcal{D}_N\}$, where $\mathcal{D}_N = \{(\boldsymbol{Y},\boldsymbol{X}) : \sup_{\boldsymbol{\beta}\in\mathcal{B}}\|\boldsymbol{g}(\boldsymbol{\beta})\| \le cN^{1/m}\}$ and $0 < c < \min\{|S^\omega|^{1/m}/N^{1/m} : \omega\in\Omega\}$. For any $\omega\in\Omega$ and $\boldsymbol{\beta}\in\mathcal{B}$, we have

$$\begin{aligned}
|S^\omega|^{1/m}L_N^\omega(\boldsymbol{\beta}) &\le \frac{1}{|S^\omega|^{1-1/m}}\left[\sum_{i,j\in S^\omega} -w_{ij}\log\left\{1+|S^\omega|^{-1/m}\boldsymbol{u}_i^{\top}(\boldsymbol{\beta})\boldsymbol{g}_j^b(\boldsymbol{\beta})\right\}\right]\\
&= \frac{1}{|S^\omega|}\left[\sum_{i\in S^\omega} -\boldsymbol{u}_i^{\top}(\boldsymbol{\beta})E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\right] + o_p(1)\\
&= -E\left[\boldsymbol{u}_i^{\top}(\boldsymbol{\beta})E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\right] + o_p(1)\\
&= -E\left[\frac{\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|^2}{1+\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|}\right] + o_p(1),
\end{aligned}$$

where the inequality follows from the fact that (Equation (7) in the main paper)

$$\hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\lambda}_i\in\mathcal{R}^{n_i}}\left[-\sum_{j\in S_i} w_{ij}\log\left\{1+\boldsymbol{\lambda}_i^{\top}\boldsymbol{g}_j(\boldsymbol{\beta})\right\}\right],$$

the first equality is Lemma B.8 in Kitamura, Tripathi, & Ahn (2004), and the second equality follows from the uniform law of large numbers. Note that the introduction of $\boldsymbol{g}^b(\boldsymbol{\beta})$ above guarantees that the arguments inside the log functions are all positive. The uniform law of large numbers is applicable because, first, $\mathcal{B}$ is compact; second, $-\boldsymbol{u}^{\top}(\boldsymbol{\beta})E\{\boldsymbol{g}(\boldsymbol{\beta}) \mid \boldsymbol{x}\}$ is continuous in $\boldsymbol{\beta}$; and third, $E\left[\sup_{\boldsymbol{\beta}\in\mathcal{B}}|\boldsymbol{u}^{\top}(\boldsymbol{\beta})E\{\boldsymbol{g}(\boldsymbol{\beta}) \mid \boldsymbol{x}\}|\right] < \infty$ from Assumption (ii).

Based on the above results and the compactness of $\mathcal{B}$, for any $\omega\in\Omega$, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left(\frac{|S^\omega|}{N}\right)^{1-1/m}|S^\omega|^{1/m}L_N^\omega(\boldsymbol{\beta}) < o_p(1).$$

In particular, for any $\epsilon > 0$, let $B(\boldsymbol{\beta}_0,\epsilon)$ be an open ball centered at $\boldsymbol{\beta}_0$ with radius ϵ, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}\backslash B(\boldsymbol{\beta}_0,\epsilon)}\left(\frac{|S^\omega|}{N}\right)^{1-1/m}|S^\omega|^{1/m}L_N^\omega(\boldsymbol{\beta}) < o_p(1). \tag{3}$$

In addition, for stratum ω_0 as in Assumption (i), we have

$$
\begin{aligned}
|S^{\omega_0}|^{1/m} L_N^{\omega_0}(\boldsymbol{\beta}) &\leq -E\left[\frac{\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|^2}{1+\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|}\right] + o_p(1) \\
&\leq -E\left[I(\boldsymbol{x}_i \in \mathcal{X}_{\boldsymbol{\beta}})\frac{\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|^2}{1+\|E\{\boldsymbol{g}_i(\boldsymbol{\beta}) \mid \boldsymbol{x}_i\}\|}\right] + o_p(1).
\end{aligned}
$$

By Assumption (i), the last quantity above is strictly negative at each $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Therefore, there exists $H(\epsilon) > 0$, which is a constant depending on ϵ, such that

$$
\sup_{\boldsymbol{\beta} \in \mathcal{B} \setminus B(\boldsymbol{\beta}_0, \epsilon)} \left(\frac{|S^{\omega_0}|}{N}\right)^{1-1/m} |S^{\omega_0}|^{1/m} L_N^{\omega_0}(\boldsymbol{\beta}) < -r_{\omega_0}^{1-1/m} H(\epsilon) + o_p(1). \tag{4}
$$

The facts (3) and (4) lead to that

$$
\begin{aligned}
\sup_{\boldsymbol{\beta} \in \mathcal{B} \setminus B(\boldsymbol{\beta}_0, \epsilon)} N^{1/m} L_N(\boldsymbol{\beta}) \leq& \sum_{\omega \neq \omega_0} \sup_{\boldsymbol{\beta} \in \mathcal{B} \setminus B(\boldsymbol{\beta}_0, \epsilon)} \left(\frac{|S^{\omega}|}{N}\right)^{1-1/m} |S^{\omega}|^{1/m} L_N^{\omega}(\boldsymbol{\beta}) \\
&+ \sup_{\boldsymbol{\beta} \in \mathcal{B} \setminus B(\boldsymbol{\beta}_0, \epsilon)} \left(\frac{|S^{\omega_0}|}{N}\right)^{1-1/m} |S^{\omega_0}|^{1/m} L_N^{\omega_0}(\boldsymbol{\beta}) \\
&< -r_{\omega_0}^{1-1/m} H(\epsilon) + o_p(1).
\end{aligned} \tag{5}
$$

On the other hand, for any $\omega \in \Omega$ and $i \in S^{\omega}$, Lemma B.1 in Kitamura, Tripathi, & Ahn (2004) proves that

$$
\hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0) = \left\{\sum_{j \in S^{\omega}} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0)^{\top}\right\}^{-1} \left\{\sum_{j \in S^{\omega}} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0)\right\} \{1 + o_p(1)\}
$$

and Lemma B.7 in Kitamura, Tripathi, & Ahn (2004) proves that

$$
\left\{\sum_{j \in S^{\omega}} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0)^{\top}\right\}^{-1} = \boldsymbol{V}_i(\boldsymbol{\beta}_0)^{-1} \{1 + o_p(1)\} = O_p(1).
$$

Thus, we have

$$
\hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0) = O_p(1) \left\{\sum_{j \in S^{\omega}} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0)\right\} \{1 + o_p(1)\}. \tag{6}
$$

In addition, Lemma C.1 in Tripathi & Kitamura (2003) proves that

$$
\max_{i \in S^{\omega}} \| \sum_{j \in S^{\omega}} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0) \| = O_p\left(\sqrt{\frac{\log |S^{\omega}|}{|S^{\omega}| b_N^{q_{\omega}}}}\right), \tag{7}
$$

where q_ω is the dimension of the continuous components of $\boldsymbol{X}$ for the subjects in stratum ω. Now using the fact that $\log(1+x) \le x$ and (6) and (7), we have

$$\begin{aligned}
|S^\omega|^{1/m} L_N^\omega(\boldsymbol{\beta}_0) &\ge -|S^\omega|^{1/m} \frac{1}{|S^\omega|} \sum_{i \in S^\omega} \left[\hat{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}_0) \left\{ \sum_{j \in S^\omega} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0) \right\} \right] \\
&= -|S^\omega|^{1/m} O_p(1) \left\{ O_p\left(\sqrt{\frac{\log|S^\omega|}{|S^\omega| b_N^{q_\omega}}} \right) \right\}^2 \{1 + o_p(1)\} \\
&= O_p\left(\frac{\log|S^\omega|}{|S^\omega|^{1-1/m} b_N^{q_\omega}}\right) \\
&= o_p(1),
\end{aligned}$$

where the last equality comes from Assumption (vii). This fact implies that

$$N^{1/m} L_N(\boldsymbol{\beta}_0) = \sum_{\omega \in \Omega} \left(\frac{|S^\omega|}{N} \right)^{1-1/m} |S^\omega|^{1/m} L_N^\omega(\boldsymbol{\beta}_0) \ge o_p(1). \tag{8}$$

Since $\hat{\boldsymbol{\beta}}_{CEL}$ maximizes $L_N(\boldsymbol{\beta})$, based on the facts (5) and (8), $\hat{\boldsymbol{\beta}}_{CEL}$ must lie inside $B(\boldsymbol{\beta}_0, \epsilon)$. The consistency of $\hat{\boldsymbol{\beta}}_{CEL}$ then follows from the arbitrariness of ϵ. ■

Proof of Theorem 2. Since $\hat{\boldsymbol{\beta}}_{CEL}$ maximizes $L_N(\boldsymbol{\beta})$ in (2), we must have $\partial L_N(\hat{\boldsymbol{\beta}}_{CEL})/\partial \boldsymbol{\beta} = \mathbf{0}$. From Taylor expansion around $\boldsymbol{\beta}_0$, for some $\tilde{\boldsymbol{\beta}}$ between $\hat{\boldsymbol{\beta}}_{CEL}$ and $\boldsymbol{\beta}_0$, we have

$$\begin{aligned}
\mathbf{0} &= \sqrt{N} \frac{\partial L_N(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} + \frac{\partial^2 L_N(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \sqrt{N}(\hat{\boldsymbol{\beta}}_{CEL} - \boldsymbol{\beta}_0) \\
&= \sqrt{N} \sum_{\omega \in \Omega} \frac{|S^\omega|}{N} \frac{\partial L_N^\omega(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} + \sum_{\omega \in \Omega} \frac{|S^\omega|}{N} \frac{\partial^2 L_N^\omega(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \sqrt{N}(\hat{\boldsymbol{\beta}}_{CEL} - \boldsymbol{\beta}_0).
\end{aligned}$$

Therefore,

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{CEL} - \boldsymbol{\beta}_0) = -\left\{ \sum_{\omega \in \Omega} \frac{|S^\omega|}{N} \frac{\partial^2 L_N^\omega(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right\}^{-1} \left\{ \sum_{\omega \in \Omega} \frac{\sqrt{|S^\omega|}}{\sqrt{N}} \sqrt{|S^\omega|} \frac{\partial L_N^\omega(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right\}. \tag{9}$$

Now we have

$$\begin{aligned}
&\sqrt{|S^\omega|} \frac{\partial L_N^\omega(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\
&= -\frac{1}{\sqrt{|S^\omega|}} \sum_{i,j \in S^\omega} \frac{w_{ij} \boldsymbol{G}_j(\boldsymbol{\beta}_0)^\top \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0)}{1 + \hat{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0)} - \frac{1}{\sqrt{|S^\omega|}} \sum_{i \in S^\omega} \left\{ \frac{\partial \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} \sum_{j \in S^\omega} \frac{w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0)}{1 + \hat{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0)} \right\} \\
&= -\frac{1}{\sqrt{|S^\omega|}} \sum_{i \in S^\omega} \left\{ \sum_{j \in S^\omega} p_{ij}(\boldsymbol{\beta}_0) \boldsymbol{G}_j(\boldsymbol{\beta}_0) \right\}^\top \left\{ \sum_{j \in S^\omega} p_{ij}(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0)^\top \right\}^{-1} \left\{ \sum_{j \in S^\omega} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0) \right\} \\
&\xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{J}_\omega(\boldsymbol{\beta}_0)),
\end{aligned} \tag{10}$$

where the first equality follows from direct calculation, the second equality follows from

$$p_{ij}(\boldsymbol{\beta}_0) = \frac{w_{ij}}{1 + \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0)^\top \boldsymbol{g}_j(\boldsymbol{\beta}_0)},$$

$$\sum_{j \in S_i} \frac{w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0)}{1 + \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0)^\top \boldsymbol{g}_j(\boldsymbol{\beta}_0)} = \mathbf{0},$$

and

$$\hat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}_0) = \left\{ \sum_{j \in S_i} p_{ij}(\boldsymbol{\beta}_0) \boldsymbol{g}_j(\boldsymbol{\beta}_0) \boldsymbol{g}_j^\top(\boldsymbol{\beta}_0) \right\}^{-1} \left\{ \sum_{j \in S_i} w_{ij} \boldsymbol{g}_j(\boldsymbol{\beta}_0) \right\},$$

which are Equations (5), (6) and (10) in the main paper, respectively. The convergence in distribution above is proved by Lemma B.2 in Kitamura, Tripathi, & Ahn (2004).

On the other hand, direct calculation leads to

$$\begin{aligned} \frac{\partial^2 L_N^\omega(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{1}{|S^\omega|} \sum_{i,j \in S^\omega} \frac{w_{ij} \frac{\partial \{\hat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\beta}})^\top \boldsymbol{g}_j(\tilde{\boldsymbol{\beta}})\}}{\partial \boldsymbol{\beta}} \hat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\beta}})^\top \boldsymbol{G}_j(\tilde{\boldsymbol{\beta}})}{\{1 + \hat{\boldsymbol{\lambda}}_i^\top(\tilde{\boldsymbol{\beta}}) \boldsymbol{g}_j(\tilde{\boldsymbol{\beta}})\}^2} - \frac{1}{|S^\omega|} \sum_{i,j \in S^\omega} \frac{w_{ij} \frac{\partial \hat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \boldsymbol{G}_j(\tilde{\boldsymbol{\beta}})^\top}{1 + \hat{\boldsymbol{\lambda}}_i^\top(\tilde{\boldsymbol{\beta}}) \boldsymbol{g}_j(\tilde{\boldsymbol{\beta}})} \\ &\quad - \frac{1}{|S^\omega|} \sum_{i,j \in S^\omega} \frac{w_{ij}}{1 + \hat{\boldsymbol{\lambda}}_i^\top(\tilde{\boldsymbol{\beta}}) \boldsymbol{g}_j(\tilde{\boldsymbol{\beta}})} \left\{ \sum_{l=1}^{q_\omega} \frac{\partial g_j^{(l)}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \hat{\lambda}_i^{(l)}(\tilde{\boldsymbol{\beta}}) \right\} \\ &\stackrel{\text{def}}{=} \boldsymbol{T}_1(\tilde{\boldsymbol{\beta}}) - \boldsymbol{T}_2(\tilde{\boldsymbol{\beta}}) - \boldsymbol{T}_3(\tilde{\boldsymbol{\beta}}). \end{aligned}$$

Lemmas C.2, C.3 and C.4 in Kitamura, Tripathi, & Ahn (2004) prove that $\boldsymbol{T}_1(\tilde{\boldsymbol{\beta}}) = o_p(1)$, $\boldsymbol{T}_2(\tilde{\boldsymbol{\beta}}) = \boldsymbol{J}_\omega(\boldsymbol{\beta}_0) + o_p(1)$ and $\boldsymbol{T}_3(\tilde{\boldsymbol{\beta}}) = o_p(1)$, respectively. Thus, we have

$$-\frac{\partial^2 L_N^\omega(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \boldsymbol{J}_\omega(\boldsymbol{\beta}_0) + o_p(1). \tag{11}$$

Combining the facts (9), (10) and (11) we get the desired result. ■

Proof of Theorem 3. From (1), $p_{ij}(\boldsymbol{\beta}) = w_{ij}\{1 + o_p(1)\}$ and the $o_p(1)$ term is independent of i, j and $\boldsymbol{\beta} \in \mathcal{B}_0$. This result, together with the consistency of $\hat{\boldsymbol{\beta}}_{CEL}$, implies that for any $1 \le i \le N$,

$$\sum_{j \in S_i} p_{ij}(\hat{\boldsymbol{\beta}}_{CEL}) \boldsymbol{g}_j(\hat{\boldsymbol{\beta}}_{CEL}) \boldsymbol{g}_j(\hat{\boldsymbol{\beta}}_{CEL})^\top = \boldsymbol{V}_i(\hat{\boldsymbol{\beta}}_{CEL}) + o_p(1).$$

Then by Assumption (v) we have

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N \boldsymbol{G}_i(\hat{\boldsymbol{\beta}}_{CEL})^\top \left\{ \sum_{j \in S_i} p_{ij}(\hat{\boldsymbol{\beta}}_{CEL}) \boldsymbol{g}_j(\hat{\boldsymbol{\beta}}_{CEL}) \boldsymbol{g}_j(\hat{\boldsymbol{\beta}}_{CEL})^\top \right\}^{-1} \boldsymbol{G}_i(\hat{\boldsymbol{\beta}}_{CEL}) \\ &= \sum_{\omega \in \Omega} \frac{|S^\omega|}{N} \frac{1}{|S^\omega|} \sum_{i \in S^\omega} \boldsymbol{G}_i(\hat{\boldsymbol{\beta}}_{CEL})^\top \left\{ \boldsymbol{V}_i(\hat{\boldsymbol{\beta}}_{CEL})^{-1} + o_p(1) \right\} \boldsymbol{G}_i(\hat{\boldsymbol{\beta}}_{CEL}). \end{aligned}$$

Therefore Theorem 3 follows from Assumption (ix) and the uniform weak law of large numbers. ■

2. EXTRA SIMULATION RESULTS

This section contains some extra numerical results. The simulation settings considered here are built around the same basic simulation model as that used in Section 5, but with different features. More specifically, we consider three covariates: a continuous baseline covariate b_i, a binary treatment indicator d_i, and the visit time t. Assuming that there are five designed follow-up visits, we generate the longitudinal outcomes Y_{it} from the following model:

$$Y_{it} = \beta_1 + \beta_2 b_i + \beta_3 d_i + \beta_4 t + \beta_5 d_i \times t + \epsilon_{it}, \quad t = 1, \cdots, 5, \quad i = 1, \cdots, N,$$

where $(\beta_1, \ldots, \beta_5)^\top = (0.5, 0.5, 1.0, 0.3, 0.3)^\top$, $b_i \sim \mathcal{N}(0, 3^2)$ and is truncated between -7 and 7, $d_i \sim$ Bernoulli(0.5), $d_i \times t$ is the treatment and time interaction, and $(\epsilon_{i1}, \ldots, \epsilon_{i5})^\top \sim \mathcal{N}_5(\mathbf{0}, \boldsymbol{V}_i^{1/2} \boldsymbol{R} \boldsymbol{V}_i^{1/2})$. We consider two settings with balanced data. In the first setting, $\boldsymbol{V}_i$ is a 5×5 diagonal matrix with the t-th diagonal element $\sigma_{it}^2 = \exp(0.4 + 0.6|b_i| + 0.4t)$, and $\boldsymbol{R}$ takes the first-order autoregressive structure with correlation coefficient $\rho = 0.5$. In the second setting, $\boldsymbol{V}_i$ is a 5×5 diagonal matrix with the t-th diagonal element $\sigma_{it}^2 = (0.4 + b_i)^2$, and $\boldsymbol{R}$ takes the compound symmetry structure with correlation coefficient $\rho = 0.8$. For both settings, the bandwidth is selected by cross-validation. We use $N = 300$ and summarize the results in Table 1 and Table 2 based on 500 replications. The observations from these two tables agree with those in Section 5.

TABLE 1: Comparison of different estimators in the first setting ($N = 300$). cel: CEL; gee: GEE; $gel1$, $gel2$, $gel3$: GEL with σ_{it}^2 being modeled by $(\alpha_1 + \alpha_2 t)^2$, $\exp(\alpha_1 + \alpha_2 t)$ and the truth, respectively; ar: first-order autoregressive; cs: compound symmetry. Three summary statistics are calculated: bias, empirical standard error (number in ( )), and mean square error (number in [ ]). For the CEL estimator, the numbers in { } and ⟨ ⟩ are the means of standard errors calculated based on (11) and (12) in the main paper, respectively.

	cel	$gee.ar$	$gee.cs$	$gel1.ar$	$gel1.cs$	$gel2.ar$	$gel2.cs$	$gel3.ar$	$gel3.cs$
	0.0186	0.0119	0.0085	0.0150	0.0108	0.0146	0.0100	0.0127	0.0119
	(0.3319)	(0.4597)	(0.4891)	(0.4185)	(0.4302)	(0.4185)	(0.4334)	(0.2645)	(0.2775)
$\boldsymbol{\beta}_1$	[0.1105]	[0.2114]	[0.2393]	[0.1754]	[0.1852]	[0.1753]	[0.1879]	[0.0701]	[0.0771]
	{0.3537}	-	-	-	-	-	-	-	-
	⟨0.3640⟩	-	-	-	-	-	-	-	-
	-0.0005	-0.0038	-0.0046	-0.0024	-0.0024	-0.0026	-0.0026	0.0008	0.0011
	(0.0946)	(0.1415)	(0.1416)	(0.1177)	(0.1209)	(0.1178)	(0.1223)	(0.0815)	(0.0850)
$\boldsymbol{\beta}_2$	[0.0090]	[0.0200]	[0.0201]	[0.0139]	[0.0146]	[0.0139]	[0.0150]	[0.0066]	[0.0072]
	{0.0653}	-	-	-	-	-	-	-	-
	⟨0.0664⟩	-	-	-	-	-	-	-	-
	-0.0093	0.0107	0.0207	0.0012	0.0137	0.0029	0.0155	-0.0083	-0.0068
	(0.4808)	(0.6489)	(0.6859)	(0.5948)	(0.6069)	(0.5946)	(0.6108)	(0.3771)	(0.3882)
$\boldsymbol{\beta}_3$	[0.2313]	[0.4212]	[0.4709]	[0.3538]	[0.3685]	[0.3536]	[0.3733]	[0.1423]	[0.1508]
	{0.5006}	-	-	-	-	-	-	-	-
	⟨0.5144⟩	-	-	-	-	-	-	-	-
	-0.0062	-0.0036	-0.0028	-0.0050	-0.0037	-0.0050	-0.0035	-0.0016	-0.0014
	(0.1332)	(0.1856)	(0.1872)	(0.1661)	(0.1695)	(0.1656)	(0.1694)	(0.1069)	(0.1095)
$\boldsymbol{\beta}_4$	[0.0178]	[0.0344]	[0.0351]	[0.0276]	[0.0287]	[0.0275]	[0.0287]	[0.0114]	[0.0120]
	{0.1422}	-	-	-	-	-	-	-	-
	⟨0.1464⟩	-	-	-	-	-	-	-	-
	0.0013	-0.0079	-0.0105	-0.0041	-0.0079	-0.0045	-0.0082	-0.0034	-0.0045
	(0.1901)	(0.2586)	(0.2599)	(0.2325)	(0.2370)	(0.2325)	(0.2374)	(0.1507)	(0.1538)
$\boldsymbol{\beta}_5$	[0.0361]	[0.0669]	[0.0676]	[0.0541]	[0.0562]	[0.0541]	[0.0564]	[0.0227]	[0.0237]
	{0.2015}	-	-	-	-	-	-	-	-
	⟨0.2073⟩	-	-	-	-	-	-	-	-

TABLE 2: Comparison of different estimators in the second setting ($N = 300$). cel: CEL; gee: GEE; $gel1$, $gel2$, $gel3$: GEL with σ_{it}^2 being modeled by $(\alpha_1 + \alpha_2 t)^2$, $\exp(\alpha_1 + \alpha_2 b_i)$ and the truth, respectively; ar: first-order autoregressive; cs: compound symmetry. Three summary statistics are calculated: bias, empirical standard error (number in ( )), and mean square error (number in [ ]). For the CEL estimator, the numbers in { } and ⟨ ⟩ are the means of standard errors calculated based on (11) and (12) in the main paper, respectively.

	cel	$gee.ar$	$gee.cs$	$gel1.ar$	$gel1.cs$	$gel2.ar$	$gel2.cs$	$gel3.ar$	$gel3.cs$
	0.0048	-0.0009	0.0016	-0.0006	0.0016	0.0012	0.0034	0.0007	0.0017
	(0.1809)	(0.2323)	(0.2256)	(0.2305)	(0.2255)	(0.2274)	(0.2227)	(0.0894)	(0.0868)
$\boldsymbol{\beta}_1$	[0.0327]	[0.0540]	[0.0509]	[0.0531]	[0.0509]	[0.0517]	[0.0496]	[0.0080]	[0.0075]
	{0.1958}	-	-	-	-	-	-	-	-
	⟨0.2017⟩	-	-	-	-	-	-	-	-
	-0.0022	-0.0046	-0.0035	-0.0043	-0.0035	-0.0032	-0.0024	-0.0049	-0.0043
	(0.0711)	(0.0829)	(0.0811)	(0.0822)	(0.0813)	(0.0809)	(0.0798)	(0.0588)	(0.0578)
$\boldsymbol{\beta}_2$	[0.0051]	[0.0069]	[0.0066]	[0.0068]	[0.0066]	[0.0065]	[0.0064]	[0.0035]	[0.0034]
	{0.0505}	-	-	-	-	-	-	-	-
	⟨0.0520⟩	-	-	-	-	-	-	-	-
	-0.0097	-0.0100	-0.0089	-0.0097	-0.0090	-0.0126	-0.0123	-0.0083	-0.0094
	(0.2560)	(0.3288)	(0.3206)	(0.3266)	(0.3206)	(0.3219)	(0.3168)	(0.1275)	(0.1236)
$\boldsymbol{\beta}_3$	[0.0656]	[0.1082]	[0.1029]	[0.1068]	[0.1029]	[0.1038]	[0.1005]	[0.0163]	[0.0154]
	{0.2764}	-	-	-	-	-	-	-	-
	⟨0.2850⟩	-	-	-	-	-	-	-	-
	-0.0015	-0.0008	-0.0018	-0.0009	-0.0018	-0.0010	-0.0019	-0.0003	-0.0006
	(0.0253)	(0.0349)	(0.0318)	(0.0345)	(0.0318)	(0.0342)	(0.0313)	(0.0135)	(0.0121)
$\boldsymbol{\beta}_4$	[0.0006]	[0.0012]	[0.0010]	[0.0012]	[0.0010]	[0.0012]	[0.0010]	[0.0002]	[0.0001]
	{0.0273}	-	-	-	-	-	-	-	-
	⟨0.0281⟩	-	-	-	-	-	-	-	-
	0.0007	0.0003	0.0014	0.0005	0.0014	0.0004	0.0017	-0.0002	0.0001
	(0.0354)	(0.0491)	(0.0446)	(0.0485)	(0.0446)	(0.0484)	(0.0441)	(0.0187)	(0.0167)
$\boldsymbol{\beta}_5$	[0.0013]	[0.0024]	[0.0020]	[0.0024]	[0.0020]	[0.0023]	[0.0020]	[0.0004]	[0.0003]
	{0.0388}	-	-	-	-	-	-	-	-
	⟨0.0399⟩	-	-	-	-	-	-	-	-