Asymptotic Properties and Empirical Evaluation of the NPMLE in the Proportional Hazards Mixed-effects Model

Anthony Gamst, Michael Donohue, Ronghui Xu

University of California, San Diego

Supplementary Material

PROOF OF THEOREM 1.

First we prove $\hat{\Lambda}_n(\cdot)$ is bounded on $[0,\tau]$. We then invoke the compactness of the parameter space and Helly's selection theorem to conclude the existence of a convergent subsequence of $\{\theta_n\}$. Finally we show the limit of this subsequence must be θ_0.

First, we let

$$\bar{\Lambda}_n(t) = \sum_{ij} \frac{\delta_{ij}(1-Y_{ij}(t))}{\sum_{kl} Y_{kl}(X_{ij}) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{kl}} \mathrm{E}_\theta(e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k)},$$

$$a_i(t) = n_i^{-1} \sum_{j=1}^{n_i} \int_{u=0}^{t} \{dN_{ij}(u) - Y_{ij}(u) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij}} \mathrm{E}_\theta(e^{\mathbf{b}_i' \mathbf{W}_{ij}} | \mathbf{y}_i)\, d\Lambda_0(u)\},$$

$$f_n(u) = n^{-1} \sum_{i=1}^{n} n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}(u) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij}} \mathrm{E}_\theta(e^{\mathbf{b}_i' \mathbf{W}_{ij}} | \mathbf{y}_i).$$

We show $\sup_{t\in[0,\tau]} |\bar{\Lambda}_n(t) - \Lambda_0(t)| \to 0$ almost surely. Note that $\{a_i(t) : i = 1, 2, \dots\}$ is a mean zero independent sequence for fixed t, and by the strong law of large numbers (*SLLN*) $n^{-1}\sum_i a_i(t) \to 0$ almost surely. Also, by the boundedness assumption on $\mathbf{W}_{ij}$ and $\mathbf{Z}_{ij}$, $\mathrm{E}_\theta(e^{\mathbf{b}_i' \mathbf{W}_{ij}} | \mathbf{y}_i)$ and $e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij}}$ are both bounded. Again by *SLLN*, $f_n(u) \to E[Y_{ij}(u) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij}} \mathrm{E}_\theta(e^{\mathbf{b}_i' \mathbf{W}_{ij}} | \mathbf{y}_i)]$ almost surely.

Now consider

$$\sum_{ij}\int_{u=0}^{t}\left\{dN_{ij}(u)-Y_{ij}(u)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}}\,\mathrm{E}_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)\,d\bar{\Lambda}_n(u)\right\}$$

$$=\sum_{ij}\left\{\delta_{ij}(1-Y_{ij}(t))-\sum_{kl}\frac{Y_{ij}(X_{kl})e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}}\,\mathrm{E}_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)\delta_{kl}(1-Y_{kl}(t))}{\sum_{rs}Y_{rs}(X_{kl})e^{\boldsymbol{\beta}_0'\mathbf{Z}_{rs}}\,\mathrm{E}_\theta(e^{\mathbf{b}_r'\mathbf{W}_{rs}}|\mathbf{y}_r)}\right\}$$

$$=\sum_{ij}\delta_{ij}(1-Y_{ij}(t))-\sum_{kl}\left\{\frac{\sum_{ij}Y_{ij}(X_{kl})e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}}\,\mathrm{E}_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)\delta_{kl}(1-Y_{kl}(t))}{\sum_{rs}Y_{rs}(X_{kl})e^{\boldsymbol{\beta}_0'\mathbf{Z}_{rs}}\,\mathrm{E}_\theta(e^{\mathbf{b}_r'\mathbf{W}_{rs}}|\mathbf{y}_r)}\right\}$$

$$=0.$$

By the above, for fixed t, we have:

$$\begin{aligned}\int_{u=0}^{t} f_n(u)\,d(\bar{\Lambda}_n-\Lambda_0)(u) &= n^{-1}\sum_{ij}n_i^{-1}\int_{u=0}^{t}\{dN_{ij}(u)-Y_{ij}(u)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}}\,\mathrm{E}_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)\,d\Lambda_0(u)\}\\ &\quad - n^{-1}\sum_{ij}n_i^{-1}\int_{u=0}^{t}\{dN_{ij}(u)-Y_{ij}(u)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}}\,\mathrm{E}_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)\,d\bar{\Lambda}_n(u)\}\\ &= n^{-1}\sum_{i=1}^{n}a_i(t)\\ &\to 0 \text{ a.s.,}\end{aligned}$$

by *SLLN*.

Since $\lim f_n(u) = \mathrm{E}[Y_{ij}(u)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}}\,\mathrm{E}_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)]$ is bounded away from zero, there exists some $c_1(u) > 0$ such that eventually $f_n(u) \geq c_1(u)$ almost surely. Let $c_1 = \sup_{u\in[0,\tau]} c_1(u)$. For sufficiently large n, we can write

$$0 \leq c_1\int_{u=0}^{t} d(\bar{\Lambda}_n-\Lambda_0)(u) \leq \int_{u=0}^{t} f_n(u)\,d(\bar{\Lambda}_n-\Lambda_0)(u) \to 0 \text{ a.s.,}$$

and by the squeeze theorem, we have

$$\int_{u=0}^{t} d(\bar{\Lambda}_n-\Lambda_0)(u) \to 0 \text{ a.s..}$$

It follows that $\bar{\Lambda}_n(t) \to \Lambda_0(t)$ a.s. for all $t \in [0,\tau]$. Pointwise convergence of non-decreasing functions to a continuous limit implies local (on $[0,\tau]$ in particular) uniform continuity.

Since $\hat{\boldsymbol{\beta}}_n$, $\hat{\boldsymbol{\Sigma}}_n$, $\mathbf{Z}_{kl}$, and $\mathbf{W}_{kl}$ are in compact sets, there exists some finite, possibly negative c_2 such that

$$\hat{\boldsymbol{\beta}}_n'\mathbf{Z}_{kl} + \log \mathrm{E}_{\hat{\theta}_n}[e^{\mathbf{b}_k'\mathbf{W}_{kl}}|\mathbf{y}_k] \geq \boldsymbol{\beta}_0'\mathbf{Z}_{kl} + \log \mathrm{E}_{\theta_0}[e^{\mathbf{b}_k'\mathbf{W}_{kl}}|\mathbf{y}_k] + c_2.$$

Therefore

$$\begin{aligned}\hat{\Lambda}_n(\tau) &= \sum_{ij} \frac{\delta_{ij}(1 - Y_{ij}(\tau))}{\sum_{kl} Y_{kl}(X_{ij}) \exp\{\hat{\boldsymbol{\beta}}_n' \mathbf{Z}_{kl} + \log \mathrm{E}_{\hat{\theta}_n}[e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k]\}} \\ &\leq \sum_{ij} \frac{\delta_{ij}(1 - Y_{ij}(\tau))}{\sum_{kl} Y_{kl}(X_{ij}) \exp\{\boldsymbol{\beta}_0' \mathbf{Z}_{kl} + \log \mathrm{E}_{\theta_0}[e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k] + c_2\}} \\ &= e^{-c_2} \bar{\Lambda}_n(\tau) \to e^{-c_2} \Lambda_0(\tau).\end{aligned}$$

Now that we have established $\hat{\Lambda}$ has an upper bound almost surely, and $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ are in compact sets; we can apply Helly's selection theorem to infer the existence of a convergent subsequence which we now denote by $\hat{\theta}_n = (\hat{\Lambda}_n, \hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Sigma}}_n)$ with limit θ^*. Next we show $\theta^* = \theta_0$.

Since

$$\hat{\Lambda}_n(t) = \int_0^t \frac{\sum_{kl} Y_{kl}(u) \exp\{\boldsymbol{\beta}_0' \mathbf{Z}_{kl} + \log \mathrm{E}_{\theta_0}[e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k]\}}{\sum_{kl} Y_{kl}(u) \exp\{\hat{\boldsymbol{\beta}}_n' \mathbf{Z}_{kl} + \log \mathrm{E}_{\hat{\theta}_n}[e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k]\}} \, d\bar{\Lambda}_n(u) \to \Lambda^*(t), \qquad (12)$$

we see that Λ^* is absolutely continuous with respect to Λ_0. Furthermore, $\Lambda^*(t)$ is differentiable with respect to t and $d\hat{\Lambda}_n(t)/\, d\bar{\Lambda}_n(t)$ converges to $d\Lambda^*(t)/\, d\Lambda_0(t)$.

Note that the finite sample likelihood as expressed via (4) has no finite maximum, since λ is free to go to infinity at any X_{ij}. We restrict Λ to be right continuous with jumps at X_{ij}; and for cluster i, conditional on the random effect $\mathbf{b}_i$, we let the log-likelihood be (6), where $\Lambda\{t\}$ is the size of the jump in Λ at t. The likelihood of the observed data, $L_n(\theta)$, is still as defined in (4) and we let $l_n(\theta) = \log L_n(\theta)$. In place of Λ_0, which is continuous at X_{ij}, we use $\bar{\Lambda}_n$. In particular we have:

$$\begin{aligned}0 &\leq n^{-1}\{l_n(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Sigma}}_n, \hat{\Lambda}_n) - l_n(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \bar{\Lambda}_n)\} \\ &= n^{-1} \sum_{i=1}^n \log \left\{ \int_{\mathbf{b}} R_{1i}(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \mathbf{b}) \phi(\mathbf{b}, \hat{\boldsymbol{\Sigma}}_n) d\mathbf{b} \right\} \\ &\quad - n^{-1} \sum_{i=1}^n \log \left\{ \int_{\mathbf{b}} R_{1i}(\boldsymbol{\beta}_0, \bar{\Lambda}_n, \mathbf{b}) \phi(\mathbf{b}, \boldsymbol{\Sigma}_0) d\mathbf{b} \right\} \\ &\quad + n^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} \log(\hat{\Lambda}_n\{X_{ij}\}/\bar{\Lambda}_n\{X_{ij}\}) \\ &\to \mathrm{E} \log \left\{ \int_{\mathbf{b}} R_{1i}(\boldsymbol{\beta}^*, \Lambda^*, \mathbf{b}) \phi(\mathbf{b}, \boldsymbol{\Sigma}^*) d\mathbf{b} \prod_{j=1}^{n_i} \lambda^*(X_{ij})^{\delta_{ij}} \right. \\ &\quad \left. \times \Big(\int_{\mathbf{b}} R_{1i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) \phi(\mathbf{b}, \boldsymbol{\Sigma}_0) d\mathbf{b} \prod_{j=1}^{n_i} \lambda_0(X_{ij})^{\delta_{ij}} \Big)^{-1} \right\},\end{aligned}$$

where

$$R_{1i}(\boldsymbol{\beta}, \Lambda, \mathbf{b}) = \prod_{j=1}^{n_i} \exp[\delta_{ij}(\boldsymbol{\beta}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}) - \Lambda(X_{ij}) \exp(\boldsymbol{\beta}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})],$$

and ϕ is the multivariate normal distribution. The limit above is negative the Kullback-Leibler information, so almost surely

$$\int_{\mathbf{b}} R_{1i}(\boldsymbol{\beta}^*, \Lambda^*, \mathbf{b})\phi(\mathbf{b}, \boldsymbol{\Sigma}^*)d\mathbf{b} \prod_{j=1}^{n_i} \lambda^*(X_{ij})^{\delta_{ij}} = \int_{\mathbf{b}} R_{1i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\phi(\mathbf{b}, \boldsymbol{\Sigma}_0)d\mathbf{b} \prod_{j=1}^{n_i} \lambda_0(X_{ij})^{\delta_{ij}},$$

or

$$\int_{\mathbf{b}} \prod_{j=1}^{n_i} \lambda^*(X_{ij})^{\delta_{ij}} \exp[\delta_{ij}(\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}) - \Lambda^*(X_{ij}) \exp(\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\phi(\mathbf{b}, \boldsymbol{\Sigma}^*)d\mathbf{b}$$
$$= \int_{\mathbf{b}} \prod_{j=1}^{n_i} \lambda_0(X_{ij})^{\delta_{ij}} \exp[\delta_{ij}(\boldsymbol{\beta}_0'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}) - \Lambda_0(X_{ij}) \exp(\boldsymbol{\beta}_0'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\phi(\mathbf{b}, \boldsymbol{\Sigma}_0)d\mathbf{b}. \quad (13)$$

Now we adapt techniques used in the identifiability step of the proportional odds case (Zeng *et al.*, 2005) to conclude $\theta^* = \theta_0$. Note that $P[(X_{ij}, \delta_{ij}) = (x, \delta)] > 0$ for any (x, δ) in $[0, \tau] \times \{1\} \cup \{\tau\} \times \{0\}$. This allows us to manipulate (X_{ij}, δ_{ij}) within that set and maintain the almost sure equality (13). A particular manipulation, which we now demonstrate, will allow us to conclude $\theta^* = \theta_0$.

Fix some k in $1, \ldots, n_i$. For $j \leq k$, let $\delta_{ij} = 1, X_{ij} = 0$ in (13) and note that we assume $\Lambda^*(0) = \Lambda_0(0) = 0$. If $j > k$ and $\delta_{ij} = 0$, we replace X_{ij} with τ. Otherwise, if $j = k + 1, \ldots, n_i$ and $\delta_{ij} = 1$, we integrate X_{ij} from 0 to τ. We get:

$$\int_{\mathbf{b}} \prod_{j=1}^{k} \lambda^*(0) \exp[\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}]$$
$$\times \prod_{j=k+1}^{n_i} \left\{\exp[-\Lambda^*(\tau) \exp(\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\right\}^{1-\delta_{ij}}$$
$$\times \prod_{j=k+1}^{n_i} \left\{\int_{y=0}^{\tau} \lambda^*(y) \exp[\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij} - \Lambda^*(y) \exp(\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]dy\right\}^{\delta_{ij}} \phi(\mathbf{b}, \boldsymbol{\Sigma}^*)d\mathbf{b}$$
$$= \int_{\mathbf{b}} \prod_{j=1}^{k} \lambda_0(0) \exp[\boldsymbol{\beta}_0'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}]$$
$$\times \prod_{j=k+1}^{n_i} \{\exp[-\Lambda_0(\tau) \exp(\boldsymbol{\beta}_0{}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\}^{1-\delta_{ij}}$$
$$\times \prod_{j=k+1}^{n_i} \left\{\int_{y=0}^{\tau} \lambda_0(y) \exp[\boldsymbol{\beta}_0{}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij} - \Lambda_0(y) \exp(\boldsymbol{\beta}_0{}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]dy\right\}^{\delta_{ij}} \phi(\mathbf{b}, \boldsymbol{\Sigma}_0)d\mathbf{b}$$

or

$$\begin{aligned}
&\int_{\mathbf{b}} \prod_{j=1}^{k} \lambda^*(0) \exp[\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}] \\
&\qquad \times \prod_{j=k+1}^{n_i} \left\{\exp[-\Lambda^*(\tau)\exp(\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\right\}^{1-\delta_{ij}} \\
&\qquad \times \prod_{j=k+1}^{n_i} \left\{1 - \exp[-\Lambda^*(\tau)\exp(\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\right\}^{\delta_{ij}} \phi(\mathbf{b}, \boldsymbol{\Sigma}^*) d\mathbf{b} \\
&\quad = \int_{\mathbf{b}} \prod_{j=1}^{k} \lambda_0(0) \exp[\boldsymbol{\beta}_0'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}] \\
&\qquad \times \prod_{j=k+1}^{n_i} \{\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0{}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\}^{1-\delta_{ij}} \\
&\qquad \times \prod_{j=k+1}^{n_i} \{1 - \exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0{}'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij})]\}^{\delta_{ij}} \phi(\mathbf{b}, \boldsymbol{\Sigma}_0) d\mathbf{b}. \qquad (14)
\end{aligned}$$

For $j > k$, we can choose δ_{ij} to be 0 or 1. If we sum (14) over all possible combinations of δ_{ij}, the $j > k$ factors sum to one and we are left with:

$$\int_{\mathbf{b}} \prod_{j=1}^{k} \lambda^*(0) \exp[\boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}] \phi(\mathbf{b}, \boldsymbol{\Sigma}^*) d\mathbf{b} = \int_{\mathbf{b}} \prod_{j=1}^{k} \lambda_0(0) \exp[\boldsymbol{\beta}_0'\mathbf{Z}_{ij} + \mathbf{b}'\mathbf{W}_{ij}] \phi(\mathbf{b}, \boldsymbol{\Sigma}_0) d\mathbf{b}$$

and

$$\begin{aligned}
&\exp\left\{\sum_{j=1}^{k} \boldsymbol{\beta}^{*\prime}\mathbf{Z}_{ij} + \frac{(\sum_{j=1}^{k}\mathbf{W}_{ij})'\boldsymbol{\Sigma}^*(\sum_{j=1}^{k}\mathbf{W}_{ij})}{2}\right\} \lambda^*(0)^k \\
&\quad = \exp\left\{\sum_{j=1}^{k} \boldsymbol{\beta}_0{}'\mathbf{Z}_{ij} + \frac{(\sum_{j=1}^{k}\mathbf{W}_{ij})'\boldsymbol{\Sigma}_0(\sum_{j=1}^{k}\mathbf{W}_{ij})}{2}\right\} \lambda_0(0)^k
\end{aligned}$$

We assume $\lambda^*(0) > 0$. Since k is arbitrarily chosen, the index set of the above summations can be replaced by any subset of $\{1, \ldots, n_i\}$. In particular, if we choose an index set $\{j, j'\}$, $j \neq j'$, we get

$$\begin{aligned}
&\boldsymbol{\beta}^{*\prime}(\mathbf{Z}_{ij} + \mathbf{Z}_{ij'}) + \frac{(\mathbf{W}_{ij} + \mathbf{W}_{ij'})'\boldsymbol{\Sigma}^*(\mathbf{W}_{ij} + \mathbf{W}_{ij'})}{2} + 2\log\lambda^*(0) \\
&\quad = \boldsymbol{\beta}_0{}'(\mathbf{Z}_{ij} + \mathbf{Z}_{ij'}) + \frac{(\mathbf{W}_{ij} + \mathbf{W}_{ij'})'\boldsymbol{\Sigma}_0(\mathbf{W}_{ij} + \mathbf{W}_{ij'})}{2} + 2\log\lambda_0(0).
\end{aligned}$$

If we subtract from this equality the resulting equalities with the singleton index sets $\{j\}$ and $\{j'\}$, we have:

$$\mathbf{W}_{ij}'(\boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}_0)\mathbf{W}_{ij'} = 0,\ j \neq j' : j, j' = 1, \ldots, n_i.$$

We also clearly have

$$\boldsymbol{c}'[1, \mathbf{Z}_{ij}']' + \mathbf{W}_{ij}'(\boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}_0)\mathbf{W}_{ij} = 0,\ j = 1, \ldots, n_i$$

where $\boldsymbol{c} = 2[\log \lambda^*(0) - \log \lambda_0(0), \boldsymbol{\beta}^{*\prime} - \boldsymbol{\beta}_0']'$. Therefore under C7 $\boldsymbol{c} = \mathbf{0}$ and it follows $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_0$, $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$, and $\lambda^*(0) = \lambda_0(0)$.

To show $\Lambda^* = \Lambda_0$, we manipulate the terms of (13) again. Let $\delta_{i1} = 1$ and integrate X_{i1} from 0 to t. Also for $j = 2, \ldots, n_i$, if $\delta_{ij} = 0$, replace X_{ij} with τ and if $\delta_{ij} = 1$ integrate X_{ij} from 0 to τ. Summing the result over all possible values of δ_{ij}, $j > 1$, this time we get

$$\begin{aligned}&\int_{\mathbf{b}} 1 - \exp[-\Lambda^*(t)\exp(\boldsymbol{\beta}_0{}'Z_{i1} + \mathbf{b}'\mathbf{W}_{i1})]\phi(\mathbf{b}, \boldsymbol{\Sigma}_0)d\mathbf{b} \\ &\quad = \int_{\mathbf{b}} 1 - \exp[-\Lambda_0(t)\exp(\boldsymbol{\beta}_0{}'Z_{i1} + \mathbf{b}'\mathbf{W}_{i1})]\phi(\mathbf{b}, \boldsymbol{\Sigma}_0)d\mathbf{b}. \qquad (15)\end{aligned}$$

Because both sides of (15) are strictly monotone in $\Lambda^*(t)$ and $\Lambda_0(t)$, we have $\Lambda^*(t) = \Lambda_0(t)$. Since Λ_0 is non-decreasing and continuous, the pointwise convergence can be extended to uniform convergence on $[0, \tau]$.

PROOF OF THEOREM 2.

First let

$$\begin{aligned}\mathcal{H} =&\Big\{(\mathbf{h}_1, \mathbf{h}_2, h_3) : \mathbf{h}_1 \in \mathbf{R}^{d_1}, \mathbf{h}_2 \in \mathbf{R}^{d_2(d_2+1)/2}, \\ &\quad h_3(\cdot) \text{ is a function on } [0, \tau]; \|\mathbf{h}_1\|, \|\mathbf{h}_2\|, \|h_3\|_V \le 1\Big\}\end{aligned}$$

where $\|h_3\|_V$ denotes the total variation of $h_3(\cdot)$ in $[0, \tau]$. Let S_n be a sequence of maps from $\mathcal{U}$, a neighborhood of $(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0)$, into $l^\infty(\mathcal{H})$:

$$\begin{aligned}&S_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \Lambda)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\ &\quad \equiv n^{-1}\frac{d}{d\epsilon} l_n\left(\boldsymbol{\beta} + \epsilon\mathbf{h}_1, \boldsymbol{\Sigma} + \epsilon\mathbf{h}_2, \Lambda(t) + \epsilon\int_0^t h_3(s)\, d\Lambda(s)\right)\Big|_{\epsilon=0} \\ &\quad \equiv A_{n1}[\mathbf{h}_1] + A_{n2}[\mathbf{h}_2] + A_{n3}[h_3].\end{aligned}$$

Here we treat $\boldsymbol{\Sigma}$ as an extended column vector consisting of the upper triangle elements of the covariance matrix. The terms A_{np}, $p = 1, 2, 3$, are linear functionals on $\mathbf{R}^{d_1}, \mathbf{R}^{d_2(d_2+1)/2}$ and $BV[0, \tau]$ (the space of functions with finite total variation in $[0, \tau]$). Let $l_{\boldsymbol{\beta}}$, $l_{\boldsymbol{\Sigma}}$ and l_Λ be the score functions for $\boldsymbol{\beta}, \boldsymbol{\Sigma}$, and Λ (along $\int_0^t 1 + \epsilon h_3(s)\, d\Lambda(s)$) for a single cluster, then

$$A_{n1}[\mathbf{h}_1] = \mathcal{P}_n[\mathbf{h}_1' l_{\boldsymbol{\beta}}],\ A_{n2}[\mathbf{h}_2] = \mathcal{P}_n[\mathbf{h}_2' l_{\boldsymbol{\Sigma}}],\ \text{and } A_{n3}[h_3] = \mathcal{P}_n\left[l_\Lambda[h3]\right]$$

where $\mathcal{P}_n$ denotes the empirical measure based on n independent clusters. We now seek explicit expressions for A_{np}. Recall the log-likelihood

$$n^{-1} l_n(\theta) = n^{-1} \sum_{i=1}^{n} \log \left\{ \int_{\mathbf{b}} R_{1i}(\boldsymbol{\beta}, \Lambda, \mathbf{b}) \phi(\mathbf{b}, \boldsymbol{\Sigma}) d\mathbf{b} \right\} + n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \delta_{ij} \log \Lambda\{X_{ij}\}$$

where

$$R_{1i}(\boldsymbol{\beta}, \Lambda, \mathbf{b}) = \exp \left\{ \sum_{j=1}^{n_i} \delta_{ij} (\boldsymbol{\beta}' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}) - \Lambda(X_{ij}) \exp(\boldsymbol{\beta}' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}) \right\}.$$

Note that

$$\frac{\partial}{\partial \epsilon} R_{1i}(\boldsymbol{\beta} + \epsilon \mathbf{h}_1, \Lambda, \mathbf{b}) \big|_{\epsilon=0} = R_{1i}(\boldsymbol{\beta}, \Lambda, \mathbf{b}) \sum_{j=1}^{n_i} \mathbf{h}_1' \mathbf{Z}_{ij} \left(\delta_{ij} - \Lambda(X_{ij}) \exp(\boldsymbol{\beta}' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}) \right).$$

Furthermore let $\Lambda_\epsilon(t) = \int_0^t 1 + \epsilon h_3 \, d\Lambda$, then $\frac{\partial}{\partial \epsilon} \Lambda_\epsilon(t) = \int_0^t h_3(s) \, d\Lambda(s)$ and

$$\frac{\partial}{\partial \epsilon} R_{1i} \left(\boldsymbol{\beta}, \Lambda_\epsilon, \mathbf{b} \right) \big|_{\epsilon=0} = -R_{1i} \left(\boldsymbol{\beta}, \Lambda, \mathbf{b} \right) \sum_{j=1}^{n_i} \int_0^{X_{ij}} h_3(s) \, d\Lambda(s) \exp(\boldsymbol{\beta}' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}).$$

Also $\Lambda_\epsilon\{t\} = (1 + \epsilon h_3(t)) \Lambda\{t\}$, so

$$\frac{d}{d\epsilon} \log \Lambda_\epsilon\{t\} \big|_{\epsilon=0} = \frac{h_3(t) \Lambda\{t\}}{\Lambda_\epsilon\{t\}} \bigg|_{\epsilon=0} = h_3(t).$$

If we let H_2 denote the matrix corresponding to the extended vector $\mathbf{h}_2$, then

$$\begin{aligned} \frac{\partial}{\partial \epsilon} \phi(\mathbf{b}; \boldsymbol{\Sigma} + \epsilon \mathbf{h}_2) \big|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon} |\boldsymbol{\Sigma} + \epsilon \mathbf{h}_2|^{-1/2} e^{-\mathbf{b}' (\boldsymbol{\Sigma} + \epsilon \mathbf{h}_2)^{-1} \mathbf{b}/2} \big|_{\epsilon=0} \\ &= \left\{ \mathbf{b}' \boldsymbol{\Sigma}^{-1} H_2 \boldsymbol{\Sigma}^{-1} \mathbf{b}/2 - \text{trace}(\boldsymbol{\Sigma}^{-1} H_2)/2 \right\} e^{-\mathbf{b}' \boldsymbol{\Sigma}^{-1} \mathbf{b}/2}. \end{aligned}$$

Finally, we can explicitly write A_{np} as

$$
\begin{aligned}
A_{n1}[\mathbf{h}_1] =& n^{-1}\sum_{i=1}^{n}\left(\int_{\mathbf{b}}\sum_{j=1}^{n_i}\mathbf{h}_1'\mathbf{Z}_{ij}\left(\delta_{ij}-\Lambda(X_{ij})e^{\boldsymbol{\beta}'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right)\right.\\
&\left.\times R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}\right)\\
&\times\left(\int_{\mathbf{b}}R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}\right)^{-1}\\
A_{n2}[\mathbf{h}_2] =& n^{-1}\sum_{i=1}^{n}\left(\int_{\mathbf{b}}\left\{\mathbf{b}'\boldsymbol{\Sigma}^{-1}H_2\boldsymbol{\Sigma}^{-1}\mathbf{b}/2-\mathrm{trace}(\boldsymbol{\Sigma}^{-1}H_2)/2\right\}\right.\\
&\left.\times R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}\right)\\
&\times\left(\int_{\mathbf{b}}R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}\right)^{-1}\\
A_{n3}[h_3] =& n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\delta_{ij}h_3(X_{ij})-\int_0^{X_{ij}}h_3(s)\,d\Lambda(s)\\
&\times\int_{\mathbf{b}}e^{\boldsymbol{\beta}'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}\\
&\times\left(\int_{\mathbf{b}}R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}\right)^{-1}
\end{aligned}
$$

or

$$
\begin{aligned}
A_{n1}[\mathbf{h}_1] =& n^{-1}\sum_{i=1}^{n}\int_{\mathbf{b}}\sum_{j=1}^{n_i}\mathbf{h}_1'\mathbf{Z}_{ij}\left(\delta_{ij}-\Lambda(X_{ij})e^{\boldsymbol{\beta}'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right)d\mu_i(\mathbf{b})\\
A_{n2}[\mathbf{h}_2] =& n^{-1}\sum_{i=1}^{n}\int_{\mathbf{b}}\left\{\mathbf{b}'\boldsymbol{\Sigma}^{-1}H_2\boldsymbol{\Sigma}^{-1}\mathbf{b}/2-\mathrm{trace}(\boldsymbol{\Sigma}^{-1}H_2)/2\right\}d\mu_i(\mathbf{b})\\
A_{n3}[h_3] =& n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\delta_{ij}h_3(X_{ij})-\int_0^{X_{ij}}h_3(s)\,d\Lambda(s)\int_{\mathbf{b}}e^{\boldsymbol{\beta}'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}d\mu_i(\mathbf{b})
\end{aligned}
$$

where

$$d\mu_i(\mathbf{b})=\frac{R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}}{\int_{\mathbf{b}}R_{1i}(\boldsymbol{\beta},\Lambda,\mathbf{b})e^{-\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}/2}d\mathbf{b}}.$$

We define the limit map $S:(\boldsymbol{\beta},\boldsymbol{\Sigma},\Lambda)[\mathbf{h}_1,\mathbf{h}_2,h_3]\to l^{\infty}(\mathcal{H})$ as

$$S(\boldsymbol{\beta},\boldsymbol{\Sigma},\Lambda)[\mathbf{h}_1,\mathbf{h}_2,h_3]=A_1[\mathbf{h}_1]+A_2[\mathbf{h}_2]+A_3[h_3]$$

where the linear functionals A_p are obtained by replacing the empirical sum in A_{np} by the expectation. By construction, $S_n(\hat{\boldsymbol{\beta}}_n,\hat{\boldsymbol{\Sigma}}_n,\hat{\Lambda}_n)=0$ and $S(\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_0,\Lambda_0)=0$.

As in Murphy (1995), asymptotic normality follows by verifying four conditions on the score function: convergence to a tight Gaussian process, Fréchet differentiability, invertibility, and the approximation condition. First, $\sqrt{n}(S_n(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0) - S(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0))$ weakly converges to a tight Gaussian process on $l^\infty(\mathcal{H})$, because $\mathcal{H}$ is a Donsker class and the functionals A_{np} are bounded Lipschitz functionals with respect to $\mathcal{H}$. The approximation condition that

$$\begin{aligned}\sup_{(\mathbf{h}_1,\mathbf{h}_2,h_3)\in\mathcal{H}} &|(S_n - S)(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Sigma}}_n, \hat{\Lambda}_n) - (S_n - S)(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0)| \\ &= o_p\left(n^{-1/2} \vee \left\{\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_0\| + \sup_{t\in[0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)|\right\}\right)\end{aligned}$$

can be proved in a manner similar to Lemma 1 in the appendix of Murphy (1995). Fréchet differentiability holds by the smoothness of $S(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \Lambda)$. We consider the derivative, denoted $\dot{S}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0)$, to be a linear map, T, from the space

$$\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0, \Lambda - \Lambda_0) : (\boldsymbol{\beta}, \boldsymbol{\Sigma}, \Lambda) \text{ is in the neighborhood } \mathcal{U} \text{ of } (\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0)\}$$

to $l^\infty(\mathcal{H})$. Lastly, we need to show T is continuously invertible on its range. We write

$$\begin{aligned}T(\boldsymbol{\beta} - \boldsymbol{\beta}_0, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0, \Lambda - \Lambda_0) =&(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathcal{Q}_1(\mathbf{h}_1, \mathbf{h}_2, h_3) + (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0)'\mathcal{Q}_2(\mathbf{h}_1, \mathbf{h}_2, h_3) \\ &+ \int_0^\tau \mathcal{Q}_3(\mathbf{h}_1, \mathbf{h}_2, h_3)\, d(\Lambda - \Lambda_0)\end{aligned}$$

where the $\mathcal{Q}_i$ are the respective partial derivatives of S with respect to $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, and Λ. The $\mathcal{Q}_i$ are of the form

$$\begin{aligned}\mathcal{Q}_1(\mathbf{h}_1, \mathbf{h}_2, h_3) =&B_1\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix} + \int_0^\tau h_3(t)D_1(t)\, dt, \\ \mathcal{Q}_2(\mathbf{h}_1, \mathbf{h}_2, h_3) =&B_2\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix} + \int_0^\tau h_3(t)D_2(t)\, dt,\end{aligned}$$

and

$$\mathcal{Q}_3(\mathbf{h}_1, \mathbf{h}_2, h_3) = B_3\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix} + b_4h_3(t) + \int_0^\tau h_3(t)D_3(t)\, dt;$$

where B_1, B_2, and B_3 are constant matrices; $D_1(t)$, $D_2(t)$, $D_3(t)$ are continuously differentiable functions; and $b_4 > 0$; each of which depends on θ_0. The operator $\mathcal{Q} = (\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3)'$ is the sum of a continuously invertible operator and a compact operator from $\mathcal{H}$ to itself and to prove T is invertible it suffices to show the invertibility of the linear operator $\mathcal{Q}(\mathbf{h}_1, \mathbf{h}_2, h_3)$ (Rudin, 1973). Suppose $\mathcal{Q}(\mathbf{h}_1, \mathbf{h}_2, h_3) = \mathbf{0}$, then $T(\boldsymbol{\beta} - \boldsymbol{\beta}_0, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0, \Lambda - \Lambda_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] = \mathbf{0}$ for any $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \Lambda)$ in the neighborhood $\mathcal{U}$.

For a small constant ϵ, let

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \epsilon \mathbf{h}_1, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 + \epsilon \mathbf{h}_2,$$
$$\Lambda(t) = \Lambda_0(t) + \epsilon \int_0^t h_3(t)\, d\Lambda_0(t).$$

It follows, by the definition of T, that

$$\begin{aligned} 0 =& T(\boldsymbol{\beta} - \boldsymbol{\beta}_0, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0, \Lambda - \Lambda_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\ =& \epsilon \mathrm{E}\{(l_{\boldsymbol{\beta}_0}[\mathbf{h}_1] + l_{\boldsymbol{\Sigma}_0}[\mathbf{h}_2] + l_{\Lambda_0}[h_3])^2\}, \end{aligned}$$

so that $l_{\boldsymbol{\beta}_0}[\mathbf{h}_1] + l_{\boldsymbol{\Sigma}_0}[\mathbf{h}_2] + l_{\Lambda_0}[h_3] = 0$ almost surely. Expanding this expression we get

$$\begin{aligned} 0 =& \sum_{j=1}^{n_i} \int_{\mathbf{b}} \mathbf{h}_1' \mathbf{Z}_{ij} \left(\delta_{ij} - \Lambda_0(X_{ij}) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}} \right) R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0) \\ &+ \int_{\mathbf{b}} \left\{ \mathbf{b}' \boldsymbol{\Sigma}_0^{-1} H_2 \boldsymbol{\Sigma}_0^{-1} \mathbf{b}/2 - \mathrm{trace}(\boldsymbol{\Sigma}_0^{-1} H_2)/2 \right\} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0) \\ &+ \sum_{j=1}^{n_i} \int_{\mathbf{b}} \left(\delta_{ij} h_3(X_{ij}) - \int_0^{X_{ij}} h_3(s)\, d\Lambda_0(s) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}} \right) R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0) \end{aligned} \tag{16}$$

where

$$\begin{aligned} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) =& R_{1i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) \prod_{j=1}^{n_i} \{\lambda_0(X_{ij})\}^{\delta_{ij}} \\ =& \prod_{j=1}^{n_i} \exp[\delta_{ij}(\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}) - \Lambda_0(X_{ij}) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij})] \{\lambda_0(X_{ij})\}^{\delta_{ij}}. \end{aligned}$$

Similar to the identifiability step of the consistency proof, we show that (16) implies $\mathbf{h}_1 = \mathbf{0}$, $\mathbf{h}_2 = \mathbf{0}$, and $h_3 = 0$. Let $\mathbf{Z}_{ij}$ and $\mathbf{W}_{ij}$ be fixed. Then for fixed integer k in $1, \ldots, n_i$, we define measures $\mu_1, \ldots, \mu_{n_i}$ on the set $\{0, 1\} \times [0, \tau]$ as follows:

$$\mu_m(\{0\} \times A) = 0, \quad \mu_m(\{1\} \times A) = I(0 \in A), \quad m \le k,$$

and

$$\mu_m(\{0\} \times A) = I(\tau \in A), \quad \mu_m(\{1\} \times A) = \int I_A dx, \quad m > k,$$

where A is any Borel set in $[0, \tau]$. We integrate both sides of (16) with respect to $\{(\delta_{i1}, X_{i1}), \ldots, (\delta_{in_i}, X_{in_i})\}$ and the product measure $d\mu_1 \cdots d\mu_{n_i}$. That is, we let $\delta_{im} = 1$ and $X_{im} = 0$ for all $m \le k$. Where $m > k$, we choose $X_{im} = \tau$ if $\delta_{im} = 0$, integrate X_{im} from 0 to τ if $\delta_{im} = 1$, then sum over $\delta_{ij} \in \{0, 1\}$. Then we sum all of the equalities of (16) for all possible combinations of $\{\delta_{i1}, \ldots, \delta_{in_i}\} \in \{0, 1\}^{n_i - k}$.

We compute the integral of each term on the right side of (16) with respect to the product measure, $\prod_{m=1}^{n_i} \mu_m$, the sum of which must be 0. First note, for any $\mathbf{b}$,

$$
\begin{aligned}
&\int R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d\left(\prod_{m=1}^{n_i} \mu_m\right) \\
&= \prod_{m\le k} \{\lambda_0(0) e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\} \\
&\quad \times \sum_{\substack{\delta_{im}\in\{0,1\}\\ m>k}} \prod_{m>k} \left(\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im})]\right)^{1-\delta_{im}} \\
&\quad \times \left\{\int_{y=0}^{\tau} \exp[\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im} - \Lambda_0(y)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im})]\lambda_0(y)dy\right\}^{\delta_{im}} \\
&= \prod_{m\le k} \{\lambda_0(0) e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\} \\
&\quad \times \sum_{\substack{\delta_{im}\in\{0,1\}\\ m>k}} \prod_{m>k} \left(\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im})]\right)^{1-\delta_{im}} \\
&\quad \times \left(1-\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im})]\right)^{\delta_{im}} \\
&= \prod_{m\le k} \{\lambda_0(0) e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}.
\end{aligned}
$$

For the first term of (16), where $j \le k$, we have for any $\mathbf{b}$:

$$
\begin{aligned}
&\int \mathbf{h}_1'\mathbf{Z}_{ij}\left(\delta_{ij} - \Lambda_0(X_{ij}) e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right) R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d\left(\prod_{m=1}^{n_i} \mu_m\right) \\
&= \int \mathbf{h}_1'\mathbf{Z}_{ij} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d\left(\prod_{m=1}^{n_i} \mu_m\right) \\
&= \mathbf{h}_1'\mathbf{Z}_{ij} \prod_{m\le k} \{\lambda_0(0) e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}.
\end{aligned}
$$

When $j > k$,

$$
\begin{aligned}
&\int \mathbf{h}_1'\mathbf{Z}_{ij}\left(\delta_{ij} - \Lambda_0(X_{ij})e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right) R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d\left(\prod_{m=1}^{n_i}\mu_m\right)\\
=&\mathbf{h}_1'\mathbf{Z}_{ij}\prod_{m\le k}\{\lambda_0(0)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}\\
&\times \sum_{\substack{\delta_{im}\in\{0,1\}\\ m>k, m\neq j}}\prod_{\substack{m>k\\ m\neq j}}(\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im})])^{1-\delta_{im}}\\
&\times (1-\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im})])^{\delta_{im}}\\
&\times \sum_{\delta_{ij}\in\{0,1\}}(1-\delta_{ij})\left(-\Lambda_0(\tau)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right)\exp[-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij})]\\
&+\delta_{ij}\int_{y=0}^{\tau}\left(1-\Lambda_0(y)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+b'\mathbf{W}_{ij}}\right)\exp[\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}-\Lambda_0(y)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij})]\lambda_0(y)dy\\
=&\mathbf{h}_1'\mathbf{Z}_{ij}\prod_{m\le k}\{\lambda_0(0)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}\\
&\times \sum_{\delta_{ij}\in\{0,1\}}(1-\delta_{ij})\,(-\Lambda_0(\tau)\exp[\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij})])\\
&+\delta_{ij}\Lambda_0(\tau)\exp[\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}-\Lambda_0(\tau)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij})]\\
=&0
\end{aligned}
$$

So contributions from the first term of (16) reduce to:

$$
\begin{aligned}
&\int\sum_{j=1}^{n_i}\int_{\mathbf{b}}\mathbf{h}_1'\mathbf{Z}_{ij}\left(\delta_{ij}-\Lambda_0(X_{ij})e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right)R_{2i}(\boldsymbol{\beta}_0,\Lambda_0,\mathbf{b})\,d_{\mathbf{b}}N(\mathbf{0},\boldsymbol{\Sigma}_0)\,d\left(\prod_{m=1}^{n_i}\mu_m\right)\\
&=\sum_{j\le k}\mathbf{h}_1'\mathbf{Z}_{ij}\int_{\mathbf{b}}\prod_{m\le k}\{\lambda_0(0)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}\,d_{\mathbf{b}}N(\mathbf{0},\boldsymbol{\Sigma}_0). \qquad (17)
\end{aligned}
$$

Similarly, contributions from the second term of (16) reduce to:

$$
\begin{aligned}
&\int\int_{\mathbf{b}}\left\{\mathbf{b}'\boldsymbol{\Sigma}_0^{-1}H_2\boldsymbol{\Sigma}_0^{-1}\mathbf{b}/2-\mathrm{trace}(\boldsymbol{\Sigma}_0^{-1}H_2)/2\right\}R_{2i}(\boldsymbol{\beta}_0,\Lambda_0,\mathbf{b})\,d_{\mathbf{b}}N(\mathbf{0},\boldsymbol{\Sigma}_0)\,d\left(\prod_{m=1}^{n_i}\mu_m\right)\\
&=\int_{\mathbf{b}}\left\{\mathbf{b}'\boldsymbol{\Sigma}_0^{-1}H_2\boldsymbol{\Sigma}_0^{-1}\mathbf{b}/2-\mathrm{trace}(\boldsymbol{\Sigma}_0^{-1}H_2)/2\right\}\prod_{m\le k}\{\lambda_0(0)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}\,d_{\mathbf{b}}N(\mathbf{0},\boldsymbol{\Sigma}_0). \qquad (18)
\end{aligned}
$$

From the third term of (16), if $j \le k$ then

$$
\begin{aligned}
&\int\left(\delta_{ij}h_3(X_{ij})-\int_0^{X_{ij}}h_3(s)\,d\Lambda_0(s)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}+\mathbf{b}'\mathbf{W}_{ij}}\right)R_{2i}(\boldsymbol{\beta}_0,\Lambda_0,\mathbf{b})\,d\left(\prod_{m=1}^{n_i}\mu_m\right)\\
&=h_3(0)\prod_{m\le k}\{\lambda_0(0)e^{\boldsymbol{\beta}_0'\mathbf{Z}_{im}+\mathbf{b}'\mathbf{W}_{im}}\}; \qquad (19)
\end{aligned}
$$

if $j > k$, then

$$
\begin{aligned}
&\int \left(\delta_{ij} h_3(X_{ij}) - \int_0^{X_{ij}} h_3(s)\, d\Lambda_0(s) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}} \right) R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d\left(\prod_{m=1}^{n_i} \mu_m \right) \\
&\quad = \prod_{m \le k} \{ \lambda_0(0) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{im} + \mathbf{b}' \mathbf{W}_{im}} \} \\
&\quad \times \sum_{\delta_{ij} \in \{0,1\}} \Bigg\{ -(1-\delta_{ij}) \int_0^{\tau} h_3(s)\, d\Lambda_0(s) \\
&\quad \times \exp[\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij} - \Lambda_0(t) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij})] \\
&\quad + \delta_{ij} \int_{y=0}^{\tau} \left(h_3(y) - \int_{s=0}^{y} h_3(s)\, d\Lambda_0(s) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}} \right) \\
&\quad \times \exp[\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij} - \Lambda_0(t) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij})] \lambda_0(y) dy \Bigg\} \\
&\quad = \prod_{m \le k} \{ \lambda_0(0) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{im} + \mathbf{b}' \mathbf{W}_{im}} \} \\
&\quad \times \sum_{\delta_{ij} \in \{0,1\}} \Bigg\{ -(1-\delta_{ij}) \int_0^{\tau} h_3(s)\, d\Lambda_0(s) \\
&\quad \times \exp[\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij} - \Lambda_0(t) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij})] \\
&\quad + \delta_{ij} \int_{s=0}^{\tau} h_3(s)\, d\Lambda_0(s) \exp[\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij} - \Lambda_0(t) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij})] = 0.
\end{aligned}
$$

So contributions from the third term (16) reduce to:

$$
\begin{aligned}
&\int \sum_{j=1}^{n_i} \int_{\mathbf{b}} \left(\delta_{ij} h_3(X_{ij}) - \int_0^{X_{ij}} h_3(s)\, d\Lambda_0(s) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{ij} + \mathbf{b}' \mathbf{W}_{ij}} \right) R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b})\, d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0)\, d\left(\prod_{m=1}^{n_i} \mu_m \right) \\
&\quad = \sum_{j \le k} h_3(0) \int_{\mathbf{b}} \prod_{m \le k} \{ \lambda_0(0) e^{\boldsymbol{\beta}_0' \mathbf{Z}_{im} + \mathbf{b}' \mathbf{W}_{im}} \}\, d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0) \qquad (20)
\end{aligned}
$$

Combining (17), (18), and (20) and integrating over $\mathbf{b}$, we obtain

$$
\sum_{j=1}^{k} \mathbf{h}_1' \mathbf{Z}_{ij} + \frac{1}{2} \left(\sum_{j=1}^{k} \mathbf{W}_{ij} \right)' H_2 \left(\sum_{j=1}^{k} \mathbf{W}_{ij} \right) + k h_3(0) = 0.
$$

Since the choice of k is arbitrary, we conclude

$$
\sum_{j=k_1+1}^{k_2} \mathbf{h}_1' \mathbf{Z}_{ij} + \frac{1}{2} \left(\sum_{j=k_1+1}^{k_2} \mathbf{W}_{ij} \right)' H_2 \left(\sum_{j=k_1+1}^{k_2} \mathbf{W}_{ij} \right) + (k_2 - k_1) h_3(0) = 0.
$$

for any $1 \le k_1 < k_2 \le n_i$. Finally we have $\mathbf{W}_{ij}' H_2 \mathbf{W}_{ij'} = 0$ for $j \ne j'$ and $\mathbf{Z}_{ij}' \mathbf{h}_1 + \mathbf{W}_{ij}' H_2 \mathbf{W}_{ij}/2 + h_3(0) = 0$, so that applying Condition C7 yields $H_2 = \mathbf{0}$, $\mathbf{h}_1 = \mathbf{0}$, and $h_3(0) = 0$.

Setting $X_{ij} = 0$, $j = 2, \cdots, n_i$, and $\delta_{ij} = 1$, $j = 1, \ldots, n_i$ in (20) gives us

$$h_3(X_{i1}) = \frac{\int_0^{X_{i1}} h_3(s)\, d\Lambda_0(s) \int_{\mathbf{b}} e^{\boldsymbol{\beta}_0' Z_{i1} + \mathbf{b}'\mathbf{W}_{i1}} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0)}{\int_{\mathbf{b}} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0)}.$$

Therefore $g(y) \equiv \int_0^y h_3(t)\, d\Lambda_0(t)$ satisfies the homogeneous equation

$$\frac{g'(y)}{\lambda_0(y)} - g(y) \frac{\int_{\mathbf{b}} e^{\boldsymbol{\beta}_0' Z_{i1} + \mathbf{b}'\mathbf{W}_{i1}} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0)}{\int_{\mathbf{b}} R_{2i}(\boldsymbol{\beta}_0, \Lambda_0, \mathbf{b}) d_{\mathbf{b}} N(\mathbf{0}, \boldsymbol{\Sigma}_0)} = 0$$

with boundary condition $g(0) = 0$. We conclude $g(y) = 0$, $h_3(y) = 0$, $\mathcal{Q}$ is one-to-one, and $\dot{S}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \Lambda_0)$ is invertible. Asymptotic normality follows from Theorem 2 of Murphy (1995).

Asymptotic efficiency of $\hat{\theta}_n$ follows from Bickel, Klaasen, Ritov, and Wellner (1993), Chapter 5 by showing each component is asymptotically linear with efficient influence function in the tangent space of the model. The proof relies on the decomposition of $\dot{S}(\theta_0)$ in terms of the invertible $\mathcal{Q}$ operator.

PROOF OF THEOREM 3.

The proof is analogous to that in Parner (1998). The first step is to derive the linear approximation to the Fréchet derivative, $\dot{S}(\theta_0)(\hat{\theta} - \theta_0)[\mathbf{h}_1, \mathbf{h}_2, h_3]$, in terms of $\mathbf{J}_n$. Then we use Theorem 2 and the invertibility of $\mathbf{J}_n$ to conclude the result.