Web Appendices A and B provide the EM algorithm referenced in Sections 2.2 and the technical proofs for Theorem 1 and 2 referenced in Section 3, respectively. Supplementary proofs needed in Web Appendix B are given in Web Appendix C. Web Appendix D contains the tables referenced in Sections 4.3 and 5.

Web Appendix A: EM Algorithm

In Appendices A.1 and A.2, the expressions of the conditional expectation and the conditional score equations calculated in the E- and M-steps of the EM algorithm described in Section 2.2 are given for continuous longitudinal outcomes following a normal distribution and binary longitudinal outcomes with survival time, respectively.

A.1. EM algorithm – Continuous longitudinal data and survival time

(1) E-step : For continuous longitudinal outcomes following a normal distribution and survival time, we calculate the conditional expectation of $q(\boldsymbol{b}_i, \alpha_i)$ for subject i with $S_i = s$ given the observations and the current estimate $(\boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)})$ for some known function $q(\cdot)$. The conditional expectation denoted by $E\big[q(\boldsymbol{b}_i, \alpha_i)|\boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)}\big]$ can be expressed as the following:
Given the current estimate $(\boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)})$,

$$E\big[q(\boldsymbol{b}_i, \alpha_i)|\boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)}\big] = \frac{\sum_{\alpha_i=1}^K C_{\alpha_i} \int_{\boldsymbol{z}_{\alpha_i}} q\big(R(\boldsymbol{z}_{\alpha_i})\big)\kappa(\boldsymbol{z}_{\alpha_i}) \exp\{-\boldsymbol{z}_{\alpha_i}^T \boldsymbol{z}_{\alpha_i}\} d\boldsymbol{z}_{\alpha_i}}{\sum_{a=1}^K C_a \int_{\boldsymbol{z}_a} \kappa(\boldsymbol{z}_a) \exp\{-\boldsymbol{z}_a^T \boldsymbol{z}_a\} d\boldsymbol{z}_a}, \tag{B.01}$$

where

$$\begin{aligned}
R(\boldsymbol{z}_{\alpha_i}) &= \left(\frac{1}{\sigma_y^2}\sum_{j=1}^{n_i} \widetilde{\boldsymbol{X}}_{ij}^T \widetilde{\boldsymbol{X}}_{ij} + \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\right)^{-1} \Bigg[\sum_{j=1}^{n_i} \frac{1}{\sigma_y^2}\big(y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}\big)\widetilde{\boldsymbol{X}}_{ij}^T + \Delta_i\big(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi}^{(m)}\big) \\
&\qquad + \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\boldsymbol{\mu}_{\alpha_i}\Bigg] + \sqrt{2}\left[\frac{1}{\sigma_y^2}\sum_{j=1}^{n_i} \widetilde{\boldsymbol{X}}_{ij}^T \widetilde{\boldsymbol{X}}_{ij} + \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\right]^{-\frac{1}{2}} \boldsymbol{z}_{\alpha_i}, \\
\kappa(\boldsymbol{z}_{\alpha_i}) &= \exp\left\{-\sum_{s=1}^S I(S_i = s)\int_0^{V_i} e^{\widetilde{\boldsymbol{Z}}_i(u)\left(\boldsymbol{\psi}^{(m)} \circ R(\boldsymbol{z}_{\alpha_i})\right) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}^{(m)}} d\Lambda_s^{(m)}(u)\right\}, \\
C_{\alpha_i} &= \exp\Bigg\{\frac{1}{2}\left[\sum_{j=1}^{n_i} \frac{1}{\sigma_y^2}\big(y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}\big)\widetilde{\boldsymbol{X}}_{ij}^T + \Delta_i\big(\widetilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{\psi}^{(m)}\big) + \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\boldsymbol{\mu}_{\alpha_i}\right]^T \\
&\qquad \times \left(\frac{1}{\sigma_y^2}\sum_{j=1}^{n_i} \widetilde{\boldsymbol{X}}_{ij}^T \widetilde{\boldsymbol{X}}_{ij} + \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\right)^{-1} \times \Bigg[\sum_{j=1}^{n_i} \frac{1}{\sigma_y^2}\big(y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}\big)\widetilde{\boldsymbol{X}}_{ij}^T \\
&\qquad + \Delta_i\big(\widetilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{\psi}^{(m)}\big) + \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\boldsymbol{\mu}_{\alpha_i}\Bigg] \\
&\qquad - \frac{1}{2}\boldsymbol{\mu}_{\alpha_i}^T\big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1}\boldsymbol{\mu}_{\alpha_i} + \log w_{\alpha_i}\Bigg\}
\end{aligned}$$

is a constant, $\big(\boldsymbol{\Sigma}_b^{(m)}\big)^{\frac{1}{2}}$ is an unique non-negative square root of $\boldsymbol{\Sigma}_b^{(m)}$ (i.e. $\big(\boldsymbol{\Sigma}_b^{(m)}\big)^{\frac{1}{2}} \times \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{\frac{1}{2}} = \boldsymbol{\Sigma}_b^{(m)}$), and $\boldsymbol{z}_{\alpha_i}$ follows a multivariate Gaussian distribution with mean zero.
(2) M-step : Since normal distribution has a dispersion parameter $\boldsymbol{\phi}$ as σ_y^2, we estimate $\boldsymbol{\beta}^{(m+1)}$ and σ_y^2

in longitudinal process. $\boldsymbol{\beta}^{(m+1)}$ is the linear regression coefficients of regressing $\{\boldsymbol{Y}_i - \mathrm{E}[\widetilde{\boldsymbol{X}}_i \boldsymbol{b}_i | \boldsymbol{\theta}^{(m)}, \boldsymbol{\Lambda}^{(m)}],$ $i = 1, \ldots, n\}$ on $\{\boldsymbol{X}_i, i = 1, \ldots, n\}$, where $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}^T, ..., \boldsymbol{X}_{in_i}^T)^T$ and $\widetilde{\boldsymbol{X}}_i = (\widetilde{\boldsymbol{X}}_{i1}^T, ..., \widetilde{\boldsymbol{X}}_{in_i}^T)^T$.

$$(\sigma_y^2)^{(m+1)} = \frac{\sum_{i=1}^n \left[\boldsymbol{D}_i^T \boldsymbol{D}_i + \mathrm{E}[(\widetilde{\boldsymbol{X}}_i \boldsymbol{b}_i)^2 | \boldsymbol{\theta}^{(m)}, \boldsymbol{\Lambda}^{(m)}] - \left(\mathrm{E}[\widetilde{\boldsymbol{X}}_i \boldsymbol{b}_i | \boldsymbol{\theta}^{(m)}, \boldsymbol{\Lambda}^{(m)}]\right)^2\right]}{\sum_{i=1}^n n_i},$$

where $\boldsymbol{D}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}^{(m+1)} - \mathrm{E}[\widetilde{\boldsymbol{X}}_i \boldsymbol{b}_i | \boldsymbol{\theta}^{(m)}, \boldsymbol{\Lambda}^{(m)}]$. $\boldsymbol{\Sigma}_b^{(m+1)}$, $\boldsymbol{\mu}^{(m+1)}$, $\boldsymbol{w}^{(m+1)}$, $(\boldsymbol{\psi}^{(m+1)}, \boldsymbol{\gamma}^{(m+1)})$, and $\Lambda_s^{(m+1)}$ have the same expressions as in Section 2.2.

A.2. EM algorithm – Binary longitudinal data and survival time

(1) E-step : For binary longitudinal outcomes and survival time, given the current estimate $(\boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)})$, the conditional expectation denoted by $E\big[q(\boldsymbol{b}_i, \alpha_i) | \boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)}\big]$ can be expressed as in (B.01), where

$$\begin{aligned}
R(\boldsymbol{z}_{\alpha_i}) &= \boldsymbol{\Sigma}_b^{(m)} \left[\sum_{j=1}^{n_i} y_{ij} \widetilde{\boldsymbol{X}}_{ij}^T + \Delta_i \big(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi}^{(m)}\big)\right] + \boldsymbol{\mu}_{\alpha_i} + \sqrt{2} \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{\frac{1}{2}} \boldsymbol{z}_{\alpha_i}, \\
\kappa(\boldsymbol{z}_{\alpha_i}) &= \exp\Bigg\{ - \sum_{j=1}^{n_i} \log\left(1 + e^{\boldsymbol{X}_{ij} \boldsymbol{\beta}^{(m)} + \widetilde{\boldsymbol{X}}_{ij} R(\boldsymbol{z}_{\alpha_i})}\right) \\
&\quad - \sum_{s=1}^{S} I(S_i = s) \int_0^{V_i} e^{\widetilde{\boldsymbol{Z}}_i(u) \left(\boldsymbol{\psi}^{(m)} \circ R(\boldsymbol{z}_{\alpha_i})\right) + \boldsymbol{Z}_i(u) \boldsymbol{\gamma}^{(m)}} d\Lambda_s^{(m)}(u) \Bigg\}, \text{ and} \\
C_{\alpha_i} &= \exp\Bigg\{ \frac{1}{2} \left[\boldsymbol{\Sigma}_b^{(m)} \left(\sum_{j=1}^{n_i} y_{ij} \widetilde{\boldsymbol{X}}_{ij}^T + \Delta_i \big(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi}^{(m)}\big)\right) + \boldsymbol{\mu}_{\alpha_i}\right]^T \times \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1} \\
&\quad \times \left[\boldsymbol{\Sigma}_b^{(m)} \left(\sum_{j=1}^{n_i} y_{ij} \widetilde{\boldsymbol{X}}_{ij}^T + \Delta_i \big(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi}^{(m)}\big)\right) + \boldsymbol{\mu}_{\alpha_i}\right] \\
&\quad - \frac{1}{2} \boldsymbol{\mu}_{\alpha_i}^T \big(\boldsymbol{\Sigma}_b^{(m)}\big)^{-1} \boldsymbol{\mu}_{\alpha_i} + \log w_{\alpha_i} \Bigg\}
\end{aligned}$$

is a constant.
(2) M-step : Since the parameter $\boldsymbol{\phi}$ is set to 1 for logistic distribution, we estimate only $\boldsymbol{\beta}$ in the longitudinal process. $\boldsymbol{\beta}^{(m+1)}$ solves the conditional expectation of complete data log-likelihood score equation, using one-step Newton-Raphson iteration,

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(y_{ij} - \sum_{s=1}^{S} \mathrm{E}\left[\frac{\exp\{\boldsymbol{X}_{ij} \boldsymbol{\beta}^{(m+1)} + \widetilde{\boldsymbol{X}}_{ij} \boldsymbol{b}_i\}}{1 + \exp\{\boldsymbol{X}_{ij} \boldsymbol{\beta}^{(m+1)} + \widetilde{\boldsymbol{X}}_{ij} \boldsymbol{b}_i\}} \middle| \boldsymbol{\theta}^{(m)}, \Lambda_s^{(m)} \right] I(S_i = s) \right) \boldsymbol{X}_{ij}^T = \mathbf{0}.$$

$\boldsymbol{\Sigma}_b^{(m+1)}$, $\boldsymbol{\mu}^{(m+1)}$, $\boldsymbol{w}^{(m+1)}$, $\boldsymbol{\psi}^{(m+1)}$, $\boldsymbol{\gamma}^{(m+1)}$, and $\Lambda_s^{(m+1)}$ have the same expressions as in Section 2.2.

Web Appendix B: Proofs for Theorems

In Appendices B.1 and B.2, we present the detailed technical proofs for Theorem 1 and Theorem 2 given in Section 3, which are the asymptotic properties of the proposed estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Lambda}}(t))$ with $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^T, \widehat{\boldsymbol{\phi}}^T,$ $\mathrm{Vech}(\widehat{\boldsymbol{\Sigma}}_b)^T, \boldsymbol{\mu}^T, \boldsymbol{w}^T, \widehat{\boldsymbol{\psi}}^T, \widehat{\boldsymbol{\gamma}}^T)^T$ and $\widehat{\boldsymbol{\Lambda}}(t) = (\widehat{\Lambda}_1(t), \ldots, \widehat{\Lambda}_S(t))^T$. Meanwhile, the supplementary proofs needed to prove the asymptotic properties are provided in Appendices C.1 ~ C.3. From the full likelihood function of observed data $(\boldsymbol{Y}, \boldsymbol{V})$ for $(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ given in Section 2.2,

$$
\begin{aligned}
& L_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}) \\
&= \sum_{\boldsymbol{\alpha}} \int_{\boldsymbol{b}} L_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}, \boldsymbol{\alpha}) d\boldsymbol{b} \\
&= \prod_{i=1}^{n} \left(\sum_{\alpha_i=1}^{K} \int_{\boldsymbol{b}_1} \cdots \int_{\boldsymbol{b}_K} L_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}_i, V_i, \boldsymbol{b}_{\alpha_i}, \alpha_i) d\boldsymbol{b}_1 \cdots d\boldsymbol{b}_K \right) \\
&= \prod_{i=1}^{n} \left(\sum_{\alpha_i=1}^{K} \int_{\boldsymbol{b}_{\alpha_i}} L_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}_i, V_i, \boldsymbol{b}_{\alpha_i}, \alpha_i) d\boldsymbol{b}_{\alpha_i} \right) \\
&= \prod_{i=1}^{n} \left(\sum_{\alpha_i=1}^{K} \prod_{k=1}^{K} \left[\int_{\boldsymbol{b}_k} \left(\exp\left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_{ik}) - B(\boldsymbol{\beta}; \boldsymbol{b}_{ik})}{A(\phi_{ij})} + C(Y_{ij}; \phi_{ij}) \right] \right\} \right. \right. \right. \\
&\qquad \times \left(\prod_{s=1}^{S} \left[\lambda_s(V_i)^{\Delta_i} \exp\left\{ \Delta_i \left[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi} \circ \boldsymbol{b}_{ik}) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma} \right] \right. \right. \right. \\
&\qquad\qquad \left. \left. \left. - \int_0^{V_i} \exp\left\{ \widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_{ik}) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma} \right\} d\Lambda_s(u) \right\} \right]^{I(S_i=s)} \right) \\
&\qquad \left. \left. \left. \times (2\pi)^{-d_b/2} |\boldsymbol{\Sigma}_b|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{b}_{ik} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{b}_{ik} - \boldsymbol{\mu}_k) \right\} \times w_k \right) d\boldsymbol{b}_k \right]^{I(\alpha_i=k)} \right),
\end{aligned}
$$

we have the observed log-likelihood function,

$$
\begin{aligned}
& l_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}) = \log\{L_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V})\} \\
&= \sum_{i=1}^{n} \log \left(\sum_{\alpha_i=1}^{K} \prod_{k=1}^{K} \left[\int_{\boldsymbol{b}_k} \left(\exp\left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_{ik}) - B(\boldsymbol{\beta}; \boldsymbol{b}_{ik})}{A(\phi_{ij})} + C(Y_{ij}; \phi_{ij}) \right] \right\} \right. \right. \right. \\
&\qquad \times \left(\prod_{s=1}^{S} \left[\lambda_s(V_i)^{\Delta_i} \exp\left\{ \Delta_i \left[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi} \circ \boldsymbol{b}_{ik}) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma} \right] \right. \right. \right. \\
&\qquad\qquad \left. \left. \left. - \int_0^{V_i} \exp\left\{ \widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_{ik}) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma} \right\} d\Lambda_s(u) \right\} \right]^{I(S_i=s)} \right) \\
&\qquad \left. \left. \left. \times (2\pi)^{-d_b/2} |\boldsymbol{\Sigma}_b|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{b}_{ik} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{b}_{ik} - \boldsymbol{\mu}_k) \right\} \times w_k \right) d\boldsymbol{b}_k \right]^{I(\alpha_i=k)} \right).
\end{aligned}
$$

Then, we obtain the following modified object function by replacing $\lambda_s(V_i)$ with $\Lambda_s\{V_i\}$ in the above $l_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V})$ where $\Lambda_s\{V_i\}$ is the jump size of $\Lambda_s(t)$ at the observed time V_i with $\Delta_i = 1$,

$$
\begin{aligned}
& l_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}) \\
&= \sum_{i=1}^{n} \log \left(\sum_{\alpha_i=1}^{K} \prod_{k=1}^{K} \left[\int_{\boldsymbol{b}_k} \left(\exp\left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_{ik}) - B(\boldsymbol{\beta}; \boldsymbol{b}_{ik})}{A(\phi_{ij})} + C(Y_{ij}; \phi_{ij}) \right] \right\} \right. \right. \right. \\
&\qquad \times \left(\prod_{s=1}^{S} \left[\Lambda_s\{V_i\}^{\Delta_i} \exp\left\{ \Delta_i \left[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi} \circ \boldsymbol{b}_{ik}) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma} \right] \right. \right. \right. \\
&\qquad\qquad \left. \left. \left. - \int_0^{V_i} \exp\left\{ \widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_{ik}) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma} \right\} d\Lambda_s(u) \right\} \right]^{I(S_i=s)} \right) \\
&\qquad \left. \left. \left. \times (2\pi)^{-d_b/2} |\boldsymbol{\Sigma}_b|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{b}_{ik} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{b}_{ik} - \boldsymbol{\mu}_k) \right\} \times w_k \right) d\boldsymbol{b}_k \right]^{I(\alpha_i=k)} \right), \qquad \text{(B.02)}
\end{aligned}
$$

and $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Lambda}})$ maximizes $l_n(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ over the space $\{(\boldsymbol{\theta}, \boldsymbol{\Lambda}) : \boldsymbol{\theta} \in \Theta, \boldsymbol{\Lambda} \in \mathbb{W}_n \times \mathbb{W}_n \cdots \times \mathbb{W}_n\}$, where $\mathbb{W}_n$ consists of all the right-continuous step functions only; that is, $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_S)^T, s = 1, \ldots, S, \Lambda_s \in \mathbb{W}_n$. For the proofs of both Theorem 1 and Theorem 2, the modified object function is used in the place of the observed log-likelihood function.

B.1. Proof of Consistency – Theorem 1

Consistency can be proved by verifying the following three steps: First, we show the maximum likelihood estimate $(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Lambda}})$ exists. Second, we show that, with probability one, $\widehat{\Lambda}_s(\tau)$, $s=1,\ldots,S$, are bounded as $n\to\infty$. Third, if the second step is true, by Helly's selection theorem (p9 of [3]), we can choose a subsequence of $\widehat{\Lambda}_s$ such that $\widehat{\Lambda}_s$ weakly converges to some right-continuous monotone function Λ_s^* with probability one; that is, the measure given by $\mu_s([0,t])=\widehat{\Lambda}_s(t)$ for $t\in[0,\tau]$ weakly converges to the measure given by $\mu_s^*([0,t])=\Lambda_s^*(t)$. By choosing a sub-sequence, we can further assume $\widehat{\boldsymbol{\theta}}\to\boldsymbol{\theta}^*$. Thus, in this third step, we show $\boldsymbol{\theta}^*=\boldsymbol{\theta}_0$ and $\Lambda_s^*=\Lambda_{s0}$, $s=1,\ldots,S$.

Once the three steps are completed, we can conclude that, with probability one, $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\Lambda}}_s$ converges to $\boldsymbol{\Lambda}_{s0}$ in $[0,\tau]$, $s=1,\ldots,S$. However, since $\boldsymbol{\Lambda}_{s0}$ is continuous in $[0,\tau]$, the latter can be strengthened to uniform convergence; that is, $\sup_{t\in[0,\tau]}\|\widehat{\boldsymbol{\Lambda}}(t)-\boldsymbol{\Lambda}_0(t)\|\to 0$ almost surely. Then, the proof of Theorem 1 will be done.

In the first step, we will show the existence of the maximum likelihood estimate $(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Lambda}})$. Since $\boldsymbol{\theta}$ belongs to a compact set Θ by the assumption (A1), it is sufficient to show that $\Lambda_s\{V_i\}$, the jump size of Λ_s at V_i for which $\Delta_i=1$, is finite. Since, for each subject i with $\Delta_i=1$,

$$\begin{aligned}&\Lambda_s\{V_i\}\exp\left\{-\int_0^{V_i}\exp\left\{\widetilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi}\circ\boldsymbol{b})+\boldsymbol{Z}_i(t)\boldsymbol{\gamma}\right\}d\Lambda_s(t)\right\}\\ &\quad\le\ \exp\left\{-2\left(\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi}\circ\boldsymbol{b})+\boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}\right)\right\}\left(\Lambda_s\{V_i\}\right)^{-1},\end{aligned}$$

we have that, from (B.02),

$$\begin{aligned}l_n(\boldsymbol{\theta},\boldsymbol{\Lambda})\ \le\ &\sum_{i=1}^n\log\Bigg(\sum_{\alpha_i=1}^K\prod_{k=1}^K\Bigg[\int_{\boldsymbol{b}_k}\Bigg(\exp\Bigg\{\sum_{j=1}^{n_i}\Bigg[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_k)-B(\boldsymbol{\beta};\boldsymbol{b}_k)}{A(\phi_{ij})}+C(Y_{ij};\phi_{ij})\Bigg]\Bigg\}\\ &\times\Bigg(\prod_{s=1}^S\Bigg[\left(\Lambda_s\{V_i\}\right)^{-\Delta_i}\exp\Bigg\{-\Delta_i\Big[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi}\circ\boldsymbol{b}_k)+\boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}\Big]\Bigg\}\Bigg]^{I(S_i=s)}\Bigg)\\ &\times(2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b|^{-1/2}\exp\Bigg\{-\frac{1}{2}(\boldsymbol{b}_k-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_b^{-1}(\boldsymbol{b}_k-\boldsymbol{\mu}_k)\Bigg\}\times w_k\Bigg)d\boldsymbol{b}_k\Bigg]^{I(\alpha_i=k)}\Bigg).\end{aligned}$$

Thus, if $\Lambda_s\{V_i\}\to\infty$ for some i with $\Delta_i=1$, then $l_n(\boldsymbol{\theta},\boldsymbol{\Lambda})\to-\infty$, which is contradictory to that $l_n(\boldsymbol{\theta},\boldsymbol{\Lambda})$ is bounded. Therefore, we conclude that $\Lambda_s\{\cdot\}$, the jump size of Λ_s for stratum s, must be finite. By the conclusion and the assumption (A1), the maximum likelihood estimate $(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Lambda}})$ exists. ♯

In the second step, we will show that $\widehat{\Lambda}_s(\tau)$ is bounded as n goes to infinity with probability one. We define $\widehat{\zeta}_s=\log\widehat{\Lambda}_s(\tau)$ and rescale $\widehat{\Lambda}_s$ by the factor $e^{\widehat{\zeta}_s}$. Then, we let $\widetilde{\Lambda}_s$ denote the rescaled function; that is, $\widetilde{\Lambda}_s(t)=\widehat{\Lambda}_s(t)/\widehat{\Lambda}_s(\tau)=\widehat{\Lambda}_s(t)e^{-\widehat{\zeta}_s}$. thus, $\widetilde{\Lambda}_s(\tau)=1$. To prove this second step, it is sufficient to show $\widehat{\zeta}_s$ is bounded. After some algebra in (B.02), we obtain that, for any $\boldsymbol{\Lambda}\in\mathbb{W}\times\mathbb{W}\cdots\times\mathbb{W}$,

$$\begin{aligned}&n^{-1}l_n(\widehat{\boldsymbol{\theta}},\boldsymbol{\Lambda})\\ &=\frac{1}{n}\sum_{i=1}^n\Bigg[\sum_{j=1}^{n_i}\Bigg(\frac{Y_{ij}\boldsymbol{X}_{ij}\widehat{\boldsymbol{\beta}}}{A(\widehat{\phi}_{ij})}+C(Y_{ij};\widehat{\phi}_{ij})\Bigg)+\sum_{s=1}^S\Delta_iI(S_i=s)\log\Lambda_s\{V_i\}+\Delta_i\boldsymbol{Z}_i(V_i)\widehat{\boldsymbol{\gamma}}\\ &\quad-\frac{1}{2}\log\left\{(2\pi)^{d_b}|\widehat{\boldsymbol{\Sigma}}_b|\right\}-\frac{1}{2}\log|\widehat{\boldsymbol{\Sigma}}_b|+\log\Bigg[\sum_{\alpha=1}^K\Bigg[\widehat{w}_\alpha\exp\Bigg\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}-\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\Bigg\}\\ &\quad\times\int_{\boldsymbol{b}_{\alpha0}}\Bigg[\exp\Bigg\{-\frac{1}{2}\boldsymbol{b}_{\alpha0}^T\boldsymbol{b}_{\alpha0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha0})}{A(\widehat{\phi}_{ij})}-\sum_{s=1}^SI(S_i=s)\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha0},\widehat{\boldsymbol{\theta}})}d\Lambda_s(t)\Bigg\}\Bigg]d\boldsymbol{b}_{\alpha0}\Bigg]\Bigg]\Bigg],\end{aligned}$$

$$
\begin{aligned}
\text{where} \qquad \boldsymbol{M}_{i\alpha} &= \left[\left(\sum_{j=1}^{n_i}\frac{Y_{ij}\widetilde{\boldsymbol{X}}_{ij}}{A(\widehat{\phi}_{ij})} + \Delta_i\big(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\widehat{\boldsymbol{\psi}}^T\big)\right)\widehat{\boldsymbol{\Sigma}}_b^{1/2} + \widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1/2}\right]^T, \\
\boldsymbol{b}_{\alpha 0} &= \boldsymbol{\Sigma}_b^{-1/2}\boldsymbol{b}_\alpha - \boldsymbol{M}_{i\alpha}, \\
\text{and} \qquad Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}}) &= \big(\widetilde{\boldsymbol{Z}}_i(t)\circ\widehat{\psi}^T\big)\widehat{\boldsymbol{\Sigma}}_b^{1/2}\boldsymbol{b}_{\alpha 0} + \boldsymbol{Z}_i(t)\widehat{\boldsymbol{\gamma}} + \big(\widetilde{\boldsymbol{Z}}_i(t)\circ\widehat{\psi}^T\big)\widehat{\boldsymbol{\Sigma}}_b^{1/2}\boldsymbol{M}_{i\alpha}.
\end{aligned}
$$

Thus, since $0 \le n^{-1}l_n(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Lambda}}) - n^{-1}l_n(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\Lambda}})$ where $\widehat{\boldsymbol{\Lambda}} = e^{\widehat{\boldsymbol{\xi}}}\circ\widetilde{\boldsymbol{\Lambda}}$, it follows that

$$
\begin{aligned}
0 \le\ & \frac{1}{n}\sum_{i=1}^n\sum_{s=1}^S I(S_i=s)\Delta_i\left(\log e^{\widehat{\zeta}_s}\widetilde{\Lambda}_s\{V_i\} - \log\widetilde{\Lambda}_s\{V_i\}\right) \\
&+ \frac{1}{n}\sum_{i=1}^n\log\left[\sum_{\alpha=1}^K\left[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\right.\right. \\
&\quad\left.\left.\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0} - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \sum_{s=1}^S I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\boldsymbol{b}_{\alpha 0}\right]\right] \\
&- \frac{1}{n}\sum_{i=1}^n\log\left[\sum_{\alpha=1}^K\left[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\right.\right. \\
&\quad\left.\left.\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0} - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \sum_{s=1}^S I(S_i=s)\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}d\boldsymbol{b}_{\alpha 0}\right]\right]. \qquad \text{(B.03)}
\end{aligned}
$$

According to the assumption (A3), there exist some positive constants $C1$, $C2$ and $C3$ such that $|Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})| \le C_1\|\boldsymbol{b}_{\alpha 0}\| + C_2\|\boldsymbol{Y}_i\| + C_3$. By denoting $\boldsymbol{b}_{\alpha 0}$ as a vector of variables following a standard multivariate normal distribution, from concavity of the logarithm function, in the third term of (B.03),

$$
\begin{aligned}
&\log\left[\sum_{\alpha=1}^K\left[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\right.\right. \\
&\qquad\left.\left.\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0} - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \sum_{s=1}^S I(S_i=s)\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}d\boldsymbol{b}_{\alpha 0}\right]\right] \\
&= \frac{d_b}{2}\log(2\pi) + \log\left[\mathrm{E}_\alpha\left[\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\right.\right. \\
&\qquad\left.\left.\times\mathrm{E}_{\boldsymbol{b}_{\alpha 0}|\alpha}\left[\exp\left\{-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \sum_{s=1}^S I(S_i=s)\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\right]\right]\right] \\
&= \frac{d_b}{2}\log(2\pi) + \log\left[\mathrm{E}_{\alpha,\boldsymbol{b}_0}\left[\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\right.\right. \\
&\qquad\left.\left.\times\exp\left\{-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \sum_{s=1}^S I(S_i=s)\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\right]\right] \\
&\ge \frac{d_b}{2}\log(2\pi) + \log\left[\mathrm{E}_{\alpha,\boldsymbol{b}_0}\left[\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - e^{C_1\|\boldsymbol{b}_{\alpha 0}\|+C_2\|\boldsymbol{Y}_i\|+C_3}\right\}\right]\right] \\
&\ge \frac{d_b}{2}\log(2\pi) + \log\left[\mathrm{E}_{\alpha,\boldsymbol{b}_0}\left[\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - e^{C_1\|\boldsymbol{b}_{\alpha 0}\|+C_2\|\boldsymbol{Y}_i\|+C_3}\right]\right] \\
&= -e^{C_2\|\boldsymbol{Y}_i\|+C_4} - C_5,
\end{aligned}
$$

where C_4 and C_5 are positive constants. Then, since it is easily verified that $\mathrm{E}_{\alpha,\boldsymbol{b}_0}\left[\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - e^{C_1\|\boldsymbol{b}_{\alpha 0}\|+C_2\|\boldsymbol{Y}_i\|+C_3}\right] < \infty$, by the strong law of large numbers and the assumption (A5), the third term of (B.03)

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\log\Bigg[\sum_{\alpha=1}^{K}\Bigg[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}-\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\\
&\qquad\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}-\sum_{s=1}^{S}I(S_i=s)\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}d\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\
&\le\ \frac{1}{n}\sum_{i=1}^{n}(e^{C_2\|\boldsymbol{Y}_i\|+C_4}+C_5)\ \triangleq\ C_6
\end{aligned}
$$

can be bounded by some constant C_6 from above. Then (B.03) becomes

$$
\begin{aligned}
0\ \le\ &\frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{S}\Delta_i I(S_i=s)\widehat{\zeta}_s\\
&+\frac{1}{n}\sum_{i=1}^{n}\log\Bigg[\sum_{\alpha=1}^{K}\Bigg[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}-\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\\
&\quad\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}-\sum_{s=1}^{S}I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{V_i}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\
&+C_6\\
\le\ &\frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{S}\Delta_i I(S_i=s)\widehat{\zeta}_s\\
&+\frac{1}{n}\sum_{i=1}^{n}I(V_i=\tau)\log\Bigg[\sum_{\alpha=1}^{K}\Bigg[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}-\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\\
&\quad\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}-\sum_{s=1}^{S}I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{\tau}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\
&+\frac{1}{n}\sum_{i=1}^{n}I(V_i\neq\tau)\log\Bigg[\sum_{\alpha=1}^{K}\Bigg[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}-\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\\
&\qquad\qquad\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}\right\}\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\
&+C_6\\
\le\ &\frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{S}\Delta_i I(S_i=s)\widehat{\zeta}_s\\
&+\frac{1}{n}\sum_{i=1}^{n}I(V_i=\tau)\log\Bigg[\sum_{\alpha=1}^{K}\Bigg[\widehat{w}_\alpha\exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha}-\frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\\
&\quad\times\int_{\boldsymbol{b}_{\alpha 0}}\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}-\sum_{s=1}^{S}I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{\tau}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\
&+C_7, \qquad\qquad (\text{B.04})
\end{aligned}
$$

where C_7 is a constant. On the other hand, since, for any $\Gamma\ge 0$ and $x>0$, $\Gamma\log\big(1+x/\Gamma\big)\le\Gamma x/\Gamma=x$, we have that $e^{-x}\le\big(1+x/\Gamma\big)^{-\Gamma}$. Therefore, in the second term of (B.04),

$$
\begin{aligned}
&\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}-\sum_{s=1}^{S}I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{\tau}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}\\
&\le\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}\right\}\times\left\{1+\frac{\sum_{s=1}^{S}I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{\tau}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)}{\Gamma}\right\}^{-\Gamma}\\
&\le\Gamma^{\Gamma}\times\exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0}-\sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}};\boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})}\right\}\times\left\{\sum_{s=1}^{S}I(S_i=s)e^{\widehat{\zeta}_s}\int_0^{\tau}e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})}d\widetilde{\Lambda}_s(t)\right\}^{-\Gamma}. \qquad (\text{B.05})
\end{aligned}
$$

Since $Q_{1i\alpha}(t, \boldsymbol{b}_{\alpha 0}, \widehat{\boldsymbol{\theta}}) \geq -C_1\|\boldsymbol{b}_{\alpha 0}\| - C_2\|\boldsymbol{Y}_i\| - C_3$,

$$\begin{aligned}\int_0^\tau e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})} d\widetilde{\Lambda}_s(t) &\geq \int_0^\tau e^{-C_1\|\boldsymbol{b}_{\alpha 0}\|-C_2\|\boldsymbol{Y}_i\|-C_3} d\widetilde{\Lambda}_s(t)\\ &= e^{-C_1\|\boldsymbol{b}_{\alpha 0}\|-C_2\|\boldsymbol{Y}_i\|-C_3} \times \left\{\widetilde{\Lambda}_s(\tau) - \widetilde{\Lambda}_s(0)\right\}\\ &= e^{-C_1\|\boldsymbol{b}_{\alpha 0}\|-C_2\|\boldsymbol{Y}_i\|-C_3}.\end{aligned}$$

Thus, in (B.05),

$$\left\{\sum_{s=1}^S I(S_i = s) e^{\widehat{\zeta}_s} \int_0^\tau e^{Q_{1i\alpha}(t,\boldsymbol{b}_{\alpha 0},\widehat{\boldsymbol{\theta}})} d\widetilde{\Lambda}_s(t)\right\}^{-\Gamma} \leq \left\{\sum_{s=1}^S I(S_i = s) e^{\widehat{\zeta}_s}\right\}^{-\Gamma} e^{C_1\Gamma\|\boldsymbol{b}_{\alpha 0}\|+C_2\Gamma\|\boldsymbol{Y}_i\|+C_3\Gamma},$$

and

$$\begin{aligned}(B.05) \leq \Gamma^\Gamma \exp\Bigg\{ &-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0} - \sum_{j=1}^{n_i} \frac{B(\widehat{\boldsymbol{\beta}}; \boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \Gamma \log\left(\sum_{s=1}^S I(S_i = s) e^{\widehat{\zeta}_s}\right)\\ &+ C_1\Gamma\|\boldsymbol{b}_{\alpha 0}\| + C_2\Gamma\|\boldsymbol{Y}_i\| + C_3\Gamma\Bigg\}\end{aligned}$$

Therefore, (B.04) gives that

$$\begin{aligned}0 \leq\ & C_7 + \frac{1}{n}\sum_{i=1}^n \Delta_i\left(\sum_{s=1}^S I(S_i = s)\widehat{\zeta}_s\right) + \frac{1}{n}\sum_{i=1}^n I(V_i = \tau)\log\Bigg[\sum_{\alpha=1}^K \Bigg[\widehat{w}_\alpha \exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}\\ &\times \int_{\boldsymbol{b}_{\alpha 0}} \Gamma^\Gamma \exp\Bigg\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0} - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}}; \boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} - \Gamma\sum_{s=1}^S I(S_i = s)\widehat{\zeta}_s + C_1\Gamma\|\boldsymbol{b}_{\alpha 0}\|\\ &\qquad + C_2\Gamma\|\boldsymbol{Y}_i\| + C_3\Gamma\Bigg\} d\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\ =\ & C_7 + \frac{1}{n}\sum_{i=1}^n \Delta_i\left(\sum_{s=1}^S I(S_i = s)\widehat{\zeta}_s\right) + \frac{1}{n}\sum_{i=1}^n I(V_i = \tau)\log\Bigg[\Gamma^\Gamma \exp\left\{-\Gamma\sum_{s=1}^S I(S_i = s)\widehat{\zeta}_s\right\}\\ &\times \sum_{\alpha=1}^K \Bigg[\widehat{w}_\alpha \exp\left\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha\right\}(2\pi)^{d_b/2}(2\pi)^{-d_b/2}\\ &\times \int_{\boldsymbol{b}_{\alpha 0}} \exp\left\{-\frac{1}{2}\boldsymbol{b}_{\alpha 0}^T\boldsymbol{b}_{\alpha 0} - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}}; \boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} + C_1\Gamma\|\boldsymbol{b}_{\alpha 0}\| + C_2\Gamma\|\boldsymbol{Y}_i\| + C_3\Gamma\right\} d\boldsymbol{b}_{\alpha 0}\Bigg]\Bigg]\\ =\ & C_7 + \frac{1}{n}\sum_{i=1}^n \Delta_i\left(\sum_{s=1}^S I(S_i = s)\widehat{\zeta}_s\right) + \frac{1}{2}\log(2\pi) + \frac{1}{n}\sum_{i=1}^n I(V_i = \tau)\Bigg[\Gamma\log\Gamma - \Gamma\sum_{s=1}^S I(S_i = s)\widehat{\zeta}_s\\ &+ \log \mathrm{E}_{\alpha,\boldsymbol{b}_0}\Bigg[\exp\Bigg\{\frac{1}{2}\boldsymbol{M}_{i\alpha}^T\boldsymbol{M}_{i\alpha} - \frac{1}{2}\widehat{\boldsymbol{\mu}}_\alpha^T\widehat{\boldsymbol{\Sigma}}_b^{-1}\widehat{\boldsymbol{\mu}}_\alpha - \sum_{j=1}^{n_i}\frac{B(\widehat{\boldsymbol{\beta}}; \boldsymbol{b}_{\alpha 0})}{A(\widehat{\phi}_{ij})} + C_1\Gamma\|\boldsymbol{b}_{\alpha 0}\|\\ &\qquad + C_2\Gamma\|\boldsymbol{Y}_i\| + C_3\Gamma\Bigg\}\Bigg]\Bigg]\\ =\ & C_8 + \frac{1}{n}\sum_{i=1}^n \Delta_i\left(\sum_{s=1}^S \widehat{\zeta}_s\right) - \frac{\Gamma}{n}\sum_{i=1}^n I(V_i = \tau)\left(\sum_{s=1}^S \widehat{\zeta}_s\right) + C_9(\Gamma), \qquad \text{(B.06)}\end{aligned}$$

where C_8 is a constant and $C_9(\Gamma)$ is a deterministic function of Γ. For the s-th stratum, (B.06) is that

$$0 \leq C_8 + \sum_{i=1}^n \Delta_i I(S_i = s)\widehat{\zeta}_s - \frac{\Gamma}{n}\sum_{i=1}^n I(V_i = \tau) I(S_i = s)\widehat{\zeta}_s + C_9(\Gamma).$$

By the strong law of large numbers, $\sum_{i=1}^n I(V_i = \tau)I(S_i = s)/n \longrightarrow P(V_i = \tau, S_i = s) > 0$. Then, we can choose Γ large enough such that $\sum_{i=1}^n \Delta_i I(S_i = s)/n \leq (\Gamma/2n)\sum_{i=1}^n I(V_i = \tau)I(S_i = s)$. Thus, we obtain that

$$0 \leq C_8 + C_9(\Gamma) - \frac{\Gamma}{2n}\sum_{i=1}^n I(V_i = \tau)I(S_i = s)\widehat{\zeta}_s.$$

In other words,

$$\widehat{\zeta}_s \leq \frac{(C_8 + C_9(\Gamma))2n}{\Gamma \sum_{i=1}^n I(V_i = \tau) I(S_i = s)} \longrightarrow \frac{(C_8 + C_9(\Gamma))2}{\Gamma P(V_i = \tau, S_i = s)}.$$

If we denote $B_{s0} = \exp\left\{2\big(C_8 + C_9(\Gamma)\big)/\big(\Gamma P(V_i = \tau, S_i = s)\big)\right\}$, we conclude that $\widehat{\Lambda}_s(\tau) \leq B_{s0}, s = 1, \ldots, S$. Note that the above arguments hold for every sample in the probability space except a set with zero probability. Therefore, we have shown that, with probability one, $\widehat{\Lambda}_s(\tau)$ is bounded for any sample size n. ♯

In the third step, the goal of this step is to show that, if $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^*$ and $\widehat{\Lambda}_s$ weakly converges to Λ_s^* with probability one, then $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $\Lambda_s^* = \Lambda_{s0}$, $s = 1, \ldots, S$. We set some preliminaries as the followings: For convenience, we omit the index i for subject and use $\boldsymbol{O}$ to abbreviate the observed statistics $(\boldsymbol{Y}, \boldsymbol{X}, \widetilde{\boldsymbol{X}}, V, \Delta, n_N, s)$ and $\{\boldsymbol{Z}(t), \widetilde{\boldsymbol{Z}}(t), 0 \leq t \leq V\}$ for a subject. By dropping $(\lambda_s(V))^{\Delta}$ from the complete data likelihood function, we define that

$$\begin{aligned} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) &= \exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_i\boldsymbol{b}_\alpha) - B(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)} + C(Y_j; \phi_j)\right]\right\} \\ &\times \exp\left\{\Delta\big[\widetilde{\boldsymbol{Z}}(V)(\boldsymbol{\psi} \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(V)\boldsymbol{\gamma}\big] - \int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(t)\boldsymbol{\gamma}\right\} d\Lambda_s(t)\right\} \\ &\times (2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{b}_\alpha - \boldsymbol{\mu}_\alpha)^T \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{b}_\alpha - \boldsymbol{\mu}_\alpha)\right\} w_\alpha, \end{aligned}$$

and $$Q(v, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) = \frac{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}_\alpha, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) \exp\left\{\widetilde{\boldsymbol{Z}}(v)(\boldsymbol{\psi} \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(v)\boldsymbol{\gamma}\right\} d\boldsymbol{b}_\alpha}{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}_\alpha, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) d\boldsymbol{b}_\alpha}.$$

Furthermore, for any measurable function $f(\boldsymbol{O})$, we use operator notation to define $\mathbf{P}_n f = n^{-1}\sum_{i=1}^n f(\boldsymbol{O}_i)$ and $\mathbf{P} f = \int f d\mathbf{P} = \mathrm{E}[f(\boldsymbol{O})]$. Thus, $\mathbf{P}_n f$ is the empirical measure from n i.i.d observations and $\sqrt{n}(\mathbf{P}_n - \mathbf{P})$ is the empirical process based on these observations. We also define a class $\mathscr{F} = \{Q(v, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) : v \in [0, \tau], \boldsymbol{\theta} \in \Theta, \Lambda_s \in \mathbb{W}, \Lambda_s(0) = 0, \Lambda_s(\tau) \leq B_{s0}\}$, where B_{s0} is the constant given in the second step and $\mathbb{W}$ contains all nondecreasing functions in $[0, \tau]$. According to the result proved in Section C.1 $\mathscr{F}$ is P-Donsker.

Let m_s denote the number of subjects in stratum s; i.e. $n = \sum_{s=1}^S m_s$. V_s and Δ_s denote the observed time and censoring indicator for a subject belonging to stratum s, respectively. Thus, V_{sl} and Δ_{sl} are the l-th subject observed time and censoring indicator in stratum s.

Now we start the proof of the third step. Since $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Lambda}})$ maximizes the function $l_n(\boldsymbol{\theta}, \boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_S)^T$ and Λ_s, $s = 1, \ldots, S$, are any step functions with jumps only at V_i belonging to stratum s for which $\Delta_i = 1$, we differentiate $l_n(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ with respect to $\Lambda_s\{V_{sl}\}$ and obtain the following equation, satisfied by $\widehat{\Lambda}_s$,

$$\widehat{\Lambda}_s\{V_{sl}\} = \frac{\Delta_{sl}}{m_s \mathbf{P}_{m_s}\left\{I(V_s \geq v) Q(v, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)\right\}\big|_{v = V_{sl}}}$$

Imitating the above equation, we also can construct another function, denoted by $\bar{\boldsymbol{\Lambda}} = (\bar{\Lambda}_1, \ldots, \bar{\Lambda}_S)^T$ such that $\bar{\Lambda}_s$, $s = 1, \ldots, S$, are also step functions with jumps only at the observed V_{sl} and the jump size $\bar{\Lambda}_s\{V_{sl}\}$ is given by

$$\bar{\Lambda}_s\{V_{sl}\} = \frac{\Delta_{sl}}{m_s \mathbf{P}_{m_s}\left\{I(V_s \geq v) Q(v, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0})\right\}\big|_{v = V_{sl}}}.$$

Equivalently,

$$\bar{\Lambda}_s(t) = \frac{1}{m_s}\sum_{l=1}^{m_s} \frac{I(V_{sl} \leq t)\Delta_{sl}}{\mathbf{P}_{m_s}\left\{I(V_s \geq v) Q(v, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0})\right\}\big|_{v = V_{sl}}}.$$

Then, we claim $\bar{\Lambda}_s(t)$ uniformly converges to $\Lambda_{s0}(t)$ in $[0,\tau]$. To prove the claim, note that

$$
\begin{aligned}
&\sup_{t\in[0,\tau]}\left|\bar{\Lambda}_s(t) - \mathrm{E}\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right]\right| \\
&= \sup_{t\in[0,\tau]}\left|\frac{1}{m_s}\sum_{l=1}^{m_s}\frac{I(V_{sl}\le t)\Delta_{sl}}{\mathbf{P}_{m_s}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_{sl}}} - \mathbf{P}_{m_s}\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right]\right. \\
&\quad \left. + \ \mathbf{P}_{m_s}\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right] - \mathbf{P}\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right]\right| \\
&\le \sup_{t\in[0,\tau]}\left|\frac{1}{m_s}\sum_{l=1}^{m_s}I(V_{sl}\le t)\Delta_{sl}\left[\frac{1}{\mathbf{P}_{m_s}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}} - \frac{1}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}}\right]\bigg|_{v=V_{sl}}\right| \\
&\quad + \sup_{t\in[0,\tau]}\left|\left(\mathbf{P}_{m_s} - \mathbf{P}\right)\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right]\right| \\
&\le \sup_{t\in[0,\tau]}\left|\frac{1}{\mathbf{P}_{m_s}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}} - \frac{1}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}}\right| \\
&\quad + \sup_{t\in[0,\tau]}\left|\left(\mathbf{P}_{m_s} - \mathbf{P}\right)\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right]\right|. \qquad \text{(B.07)}
\end{aligned}
$$

In (B.07), the right hand side converges to 0 because the first and second terms on the right hand side converges to 0 in the following: First, according to Section C.1, $\{Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) : v \in [0,\tau]\}$ is a bounded and Glivenko-Cantelli class. $\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) : v \in [0,\tau]\}$ is also a Glivenko-Cantelli class because $\{I(V_s \ge v) : v \in [0,\tau]\}$ is a Glivenko-Cantelli class and the functional $(f,g) \to fg$ for any bounded two functions f and g is Lipschitz continuous. Then, we obtain that

$\sup_{t\in[0,\tau]}\left|\mathbf{P}_{m_s}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\} - \mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\right|$ converges to 0.

Besides, from Section C.1, $\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\} > \mathbf{P}\left\{I(V_s \ge v)\exp\{-C_{10} - C_{11}\|\boldsymbol{Y}\|\}\right\}$ for the two constants C_{10} and C_{11}, which means $\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}$ is bounded from below. Thus, the first term tends to 0. Second, since the class $\left\{I(V_s \le t)\Delta_s / \mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s} : t \in [0,\tau]\right\}$ is also a Glivenko-Cantelli class, the second term vanishes as m_s goes to infinity.

Therefore, we conclude that $\bar{\Lambda}_s(t)$ uniformly converges to

$$
\mathrm{E}\left[\frac{I(V_s \le t)\Delta_s}{\mathbf{P}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\big|_{v=V_s}}\right]. \qquad \text{(B.08)}
$$

We can easily verify that (B.08) is equal to $\Lambda_{s0}(t)$. Thus, the claim that $\bar{\Lambda}_s(t)$ uniformly converges to $\Lambda_{s0}(t)$ in $[0,\tau]$ has been proved.

From the construction of $\bar{\Lambda}_s(t)$, we obtain that

$$
\widehat{\Lambda}_s(t) = \int_0^t \frac{d\widehat{\Lambda}_s(v)}{d\bar{\Lambda}_s(v)}d\bar{\Lambda}_s(v) = \int_0^t \frac{\mathbf{P}_{m_s}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}}{\mathbf{P}_{m_s}\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\widehat{\boldsymbol{\theta}},\widehat{\Lambda}_s)\right\}}d\bar{\Lambda}_s(v). \qquad \text{(B.09)}
$$

$\widehat{\Lambda}_s(t)$ is absolutely continuous with respect to $\bar{\Lambda}_s(t)$. On the other hand, since both $\{I(V_s \ge v) : v \in [0,\tau]\}$ and $\mathscr{F}$ are Glivenko-Cantelli classes, $\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s) : v \in [0,\tau]\}$ is also a Glivenko-Cantelli class. Thus, we have

$$
\sup_{v\in[0,\tau]}\left|(\mathbf{P}_{m_s} - \mathbf{P})\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\widehat{\boldsymbol{\theta}},\widehat{\Lambda}_s)\right\}\right| + \sup_{v\in[0,\tau]}\left|(\mathbf{P}_{m_s} - \mathbf{P})\left\{I(V_s \ge v)Q(v,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}\right| \longrightarrow 0 \quad \text{a.s.}
$$

By the bounded convergence theorem and the fact that $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}^*$ and $\widehat{\Lambda}_s$ converges to Λ_s^*, for each v, $\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)\right\} \longrightarrow \mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)\right\}$; moreover, it is straightforward to check the derivative of $\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)\right\}$ with respect to v. Thus, by the Arzela-Ascoli theorem, uniformly in $[0, \tau]$,

$$\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)\right\} \longrightarrow \mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)\right\}.$$

Then, combining the above result and (B.09), it holds that, uniformly in $[0, \tau]$,

$$\frac{d\widehat{\Lambda}_s(v)}{d\bar{\Lambda}_s(v)} = \frac{\mathbf{P}_{m_s}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0})\right\}}{\mathbf{P}_{m_s}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)\right\}} \longrightarrow \frac{\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0})\right\}}{\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)\right\}}. \tag{B.010}$$

After taking limits on both sides of (B.09), we obtain that

$$\lim_{m_s \to \infty} \widehat{\Lambda}_s(t) = \int_0^t \frac{\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0})\right\}}{\mathbf{P}\left\{I(V_s \geq v)Q(v, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)\right\}} d\Lambda_{s0}(v), \tag{B.011}$$

Therefore, since $\Lambda_{s0}(t)$ is differentiable with respect to the Lebesque measure, so is $\Lambda_s^*(t)$; that is, (B.011) is equal to

$$\int_0^t \frac{d\Lambda_s^*(v)}{d\Lambda_{s0}(v)} d\Lambda_{s0}(v). \tag{B.012}$$

And we denote $\lambda_s^*(t)$ as the derivative of $\Lambda_s^*(t)$. Additionally, from (B.010) ~ (B.012), note that $\widehat{\Lambda}_s\{V_s\}/\bar{\Lambda}_s\{V_s\}$ uniformaly converges to $d\Lambda_s^*(V_s)/d\Lambda_{s0}(V_s) = \lambda_s^*(V_s)/\lambda_{s0}(V_s)$. Therefore, a second conclusion is that $\widehat{\Lambda}_s$ uniformly converges to Λ_s^* since Λ_s^* is continuous.

On the other hand,

$$\begin{aligned} & n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Lambda}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \bar{\boldsymbol{\Lambda}}) \\ & = \sum_{s=1}^{S} \left(\mathbf{P}_{m_s}\left[\Delta_s \log \frac{\widehat{\Lambda}_s\{V_s\}}{\bar{\Lambda}_s\{V_s\}} \right] + \mathbf{P}_{m_s}\left[\log \frac{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s) d\boldsymbol{b}_\alpha}{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \bar{\Lambda}_s) d\boldsymbol{b}_\alpha} \right] \right) \\ & \geq 0. \end{aligned} \tag{B.013}$$

Using the result of Section C.1 and similar arguments as above, we can verify that

$$\log \frac{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s) d\boldsymbol{b}_\alpha}{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \bar{\Lambda}_s) d\boldsymbol{b}_\alpha}$$

belongs to a Glivenko-Cantelli class and

$$\mathbf{P}\left[\log \frac{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s) d\boldsymbol{b}_\alpha}{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \bar{\Lambda}_s) d\boldsymbol{b}_\alpha} \right] \longrightarrow \mathbf{P}\left[\log \frac{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha}{\sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha} \right].$$

Since $\widehat{\Lambda}_s\{V_s\}/\bar{\Lambda}_s\{V_s\}$ uniformaly converges to $\lambda_s^*(V_s)/\lambda_{s0}(V_s)$, we obtain that, from (B.013),

$$\mathbf{P}\left[\log \left\{ \frac{\left(\lambda_s^*(V_s)\right)^{\Delta_s} \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha}{\left(\lambda_{s0}(V_s)\right)^{\Delta_s} \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha} \right\} \right] \geq 0.$$

Note that the left-hand side of the inequality is the negative Kullback-Leibler information. Then, the equality holds with probability one, and it immediately follows that

$$\left(\lambda_s^*(V_s)\right)^{\Delta_s} \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha = \left(\lambda_{s0}(V_s)\right)^{\Delta_s} \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha. \tag{B.014}$$

Our proof will be completed if we can show $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $\Lambda_s^* = \Lambda_{s0}$ from (B.014). Since (B.014) holds with probability one, (B.014) holds for any $(V_s, \Delta_s = 1)$ and the case $(V_s = \tau, \Delta_s = 0)$, but may not hold for $(V_s, \Delta_s = 0)$ when $V_s \in (0, \tau)$. However, we can show that (B.014) is also true for $(V_s, \Delta_s = 0)$ when $V_s \in (0, \tau)$. To do this, treating both sides of (B.014) as functions of V_s, we integrate these functions over an interval (V_s, τ) for $\Delta_s = 0$ as the following;

$$\int_{V_s}^{\tau} \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha = \int_{V_s}^{\tau} \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha$$

to obtain that

$$\begin{aligned} &\sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=\tau} - \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=V_s} \\ &\quad = \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=\tau} - \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=V_s}. \end{aligned}$$

After comparing this above equality with another following equality, which is given by (B.014) at $\Delta_s = 0$ and $V_s = \tau$,

$$\sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=\tau} = \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=\tau},$$

we obtain

$$\sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=V_s} = \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0, V_s=V_s},$$

and therefore

$$\sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0} = \sum_{\alpha} \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) d\boldsymbol{b}_\alpha \Big|_{\Delta_s=0}; \qquad \text{(B.015)}$$

that is, (B.014) also holds for any V_s and $\Delta_s = 0$.

Thus, to show $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $\Lambda_s^* = \Lambda_{s0}$, we let $\Delta_s = 0$ in (B.014). Also, we define $\widetilde{\boldsymbol{b}}_\alpha^* = \boldsymbol{b}_\alpha - \boldsymbol{\mu}_\alpha^*$ and $\widetilde{\boldsymbol{b}}_{\alpha 0} = \boldsymbol{b}_\alpha - \boldsymbol{\mu}_{\alpha 0}$, and therefore $\widetilde{\boldsymbol{b}}_\alpha^* \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_b^*)$ and $\widetilde{\boldsymbol{b}}_{\alpha 0} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{b0})$. This leads to $\widetilde{\boldsymbol{b}}^* \sim \sum_{\alpha=1}^K w_\alpha^* N(\boldsymbol{0}, \boldsymbol{\Sigma}_b^*)$ and $\widetilde{\boldsymbol{b}}_0 \sim \sum_{\alpha=1}^K w_{\alpha 0} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{b0})$ since $\widetilde{\boldsymbol{b}} \sim \sum_{\alpha=1}^K w_\alpha N(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_b)$. Then, (B.014) for $\Delta_s = 0$ can be expressed as following,

$$\begin{aligned} &\sum_{\alpha=1}^K w_\alpha^* \,\mathrm{E}_{\widetilde{\boldsymbol{b}}^*} \Bigg[\exp\Bigg\{ \sum_{j=1}^{n_N} \Bigg[\frac{Y_j(\boldsymbol{X}_j \boldsymbol{\beta}^* + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_\alpha^*)) - B(\boldsymbol{\beta}^*, \widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_\alpha^*)}{A(\phi_j^*)} + C(Y_j; \phi_j^*) \Bigg] \Bigg\} \\ &\qquad \times \exp\Bigg\{ -\int_0^V \exp\big\{ \widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}^* \circ (\widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_\alpha^*)) + \boldsymbol{Z}(t)\boldsymbol{\gamma}^* \big\} d\Lambda_s^*(t) \Bigg\} \Bigg] \\ &= \sum_{\alpha=1}^K w_{\alpha 0} \,\mathrm{E}_{\widetilde{\boldsymbol{b}}_0} \Bigg[\exp\Bigg\{ \sum_{j=1}^{n_N} \Bigg[\frac{Y_j(\boldsymbol{X}_j \boldsymbol{\beta}_0 + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{\alpha 0})) - B(\boldsymbol{\beta}_0, \widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{\alpha 0})}{A(\phi_{j0})} + C(Y_j; \phi_{j0}) \Bigg] \Bigg\} \\ &\qquad \times \exp\Bigg\{ -\int_0^V \exp\big\{ \widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0 \circ (\widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{\alpha 0})) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0 \big\} d\Lambda_{s0}(t) \Bigg\} \Bigg], \end{aligned} \qquad \text{(B.016)}$$

where $\widetilde{\boldsymbol{b}}^* \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_b^*)$ and $\widetilde{\boldsymbol{b}}_0 \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{b0})$. Furthermore, we assume $\boldsymbol{\mu}_{10} > \boldsymbol{\mu}_{20} > \cdots > \boldsymbol{\mu}_{K0}$ and

$\boldsymbol{\mu}_1^* > \boldsymbol{\mu}_2^* > \cdots > \boldsymbol{\mu}_K^*$ in (B.016), multiply $\exp\left\{\sum_{j=1}^{n_N} t_j Y_j \widetilde{X}_j\right\}$ to both sides and obtain

$$
\begin{aligned}
&\sum_{\alpha=1}^{K} w_\alpha^* \, \mathrm{E}_{\widetilde{\boldsymbol{b}}^*}\left[\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^* + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}}^* + t_j + \boldsymbol{\mu}_\alpha^*)) - B(\boldsymbol{\beta}^*, \widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_\alpha^*)}{A(\phi_j^*)} + C(Y_j;\phi_j^*)\right]\right\}\right.\\
&\qquad\left.\times \exp\left\{-\int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}^* \circ (\widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_\alpha^*)) + \boldsymbol{Z}(t)\boldsymbol{\gamma}^*\right\} d\Lambda_s^*(t)\right\}\right]\\
&= \sum_{\alpha=1}^{K} w_{\alpha 0} \, \mathrm{E}_{\widetilde{\boldsymbol{b}}_0}\left[\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}_0 + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}}_0 + t_j + \boldsymbol{\mu}_{\alpha 0})) - B(\boldsymbol{\beta}_0, \widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{\alpha 0})}{A(\phi_{j0})} + C(Y_j;\phi_{j0})\right]\right\}\right.\\
&\qquad\left.\times \exp\left\{-\int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0 \circ (\widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{\alpha 0})) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\} d\Lambda_{s0}(t)\right\}\right]. \qquad \text{(B.017)}
\end{aligned}
$$

Then, we use the fact that, for $\boldsymbol{\mu}_1 > \boldsymbol{\mu}_2$,

$$
\begin{aligned}
&\lim_{t_j \to \infty}\left\{ \mathrm{E}_{\widetilde{\boldsymbol{b}}}\left[\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}} + t_j + \boldsymbol{\mu}_1)) - B(\boldsymbol{\beta}, \widetilde{\boldsymbol{b}} + \boldsymbol{\mu}_1)}{A(\phi_j)} + C(Y_j;\phi_j)\right]\right\}\right.\right.\\
&\qquad\left.\times \exp\left\{-\int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi} \circ (\widetilde{\boldsymbol{b}} + \boldsymbol{\mu}_1)) + \boldsymbol{Z}(t)\boldsymbol{\gamma}\right\} d\Lambda_s(t)\right\}\right] \Big/ \\
&\qquad \mathrm{E}_{\widetilde{\boldsymbol{b}}}\left[\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}} + t_j + \boldsymbol{\mu}_2)) - B(\boldsymbol{\beta}, \widetilde{\boldsymbol{b}} + \boldsymbol{\mu}_2)}{A(\phi_j)} + C(Y_j;\phi_j)\right]\right\}\right.\\
&\qquad\left.\left.\times \exp\left\{-\int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi} \circ (\widetilde{\boldsymbol{b}} + \boldsymbol{\mu}_2)) + \boldsymbol{Z}(t)\boldsymbol{\gamma}\right\} d\Lambda_s(t)\right\}\right] \right\}\\
&= \infty. \qquad \text{(B.018)}
\end{aligned}
$$

Hence, when $t_j \to \infty$, the dominating term of the left hand side of (B.017) is

$$
\begin{aligned}
&w_1^* \, \mathrm{E}_{\widetilde{\boldsymbol{b}}^*}\left[\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^* + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}}^* + t_j + \boldsymbol{\mu}_1^*)) - B(\boldsymbol{\beta}^*, \widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_1^*)}{A(\phi_j^*)} + C(Y_j;\phi_j^*)\right]\right\}\right.\\
&\qquad\left.\times \exp\left\{-\int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}^* \circ (\widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_1^*)) + \boldsymbol{Z}(t)\boldsymbol{\gamma}^*\right\} d\Lambda_s^*(t)\right\}\right]. \qquad \text{(B.019)}
\end{aligned}
$$

Similarly, the dominating term of the right hand side of (B.017) is

$$
\begin{aligned}
&w_{10} \, \mathrm{E}_{\widetilde{\boldsymbol{b}}_0}\left[\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}_0 + \widetilde{\boldsymbol{X}}_j(\widetilde{\boldsymbol{b}}_0 + t_j + \boldsymbol{\mu}_{10})) - B(\boldsymbol{\beta}_0, \widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{10})}{A(\phi_{j0})} + C(Y_j;\phi_{j0})\right]\right\}\right.\\
&\qquad\left.\times \exp\left\{-\int_0^V \exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0 \circ (\widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_{10})) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\} d\Lambda_{s0}(t)\right\}\right]. \qquad \text{(B.020)}
\end{aligned}
$$

Thus, now we can compare (B.019) and (B.020) which are for one normal distribution.

First to show that $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$, $\boldsymbol{\phi}^* = \boldsymbol{\phi}_0$, $\boldsymbol{\Sigma}_b^* = \boldsymbol{\Sigma}_{b0}$, $\boldsymbol{\mu}_1^* = \boldsymbol{\mu}_{10}$, and $w_1^* = w_{10}$, from (B.019) and (B.020), we let $V_s = 0$ use $\boldsymbol{b}$ back in place of $\widetilde{\boldsymbol{b}}^* + \boldsymbol{\mu}_1^*$ and $\widetilde{\boldsymbol{b}}_0 + \boldsymbol{\mu}_0$ in (B.019) and (B.020), respectively. Then, with probability one, by some algebra, (B.019) becomes

$$\int_{\boldsymbol{b}} \exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^* + \widetilde{\boldsymbol{X}}_j(\boldsymbol{b}+t_j)) - B(\boldsymbol{\beta}^*, \boldsymbol{b}_\alpha + t_j)}{A(\phi_j^*)} + C(Y_j;\phi_j^*)\right]\right\}$$

$$\times(2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b^*|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{b}-\boldsymbol{\mu}_1^*)^T\boldsymbol{\Sigma}_b^{*-1}(\boldsymbol{b}-\boldsymbol{\mu}_1^*)\right\}d\boldsymbol{b}$$

$$= \exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^* + \widetilde{\boldsymbol{X}}_j t_j)}{A(\phi_j^*)} + C(Y_j;\phi_j^*)\right]\right\}(2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b^*|^{-1/2}$$

$$\times\int_{\boldsymbol{b}}\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j\boldsymbol{b}}{A(\phi_j^*)} - \sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)} - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{\Sigma}_b^{*-1}\boldsymbol{b} + \boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\boldsymbol{b} - \frac{1}{2}\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\boldsymbol{\mu}_1^*\right\}w_1^* d\boldsymbol{b}$$

$$= \exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^* + \widetilde{\boldsymbol{X}}_j t_j)}{A(\phi_j^*)} + C(Y_j;\phi_j^*)\right]\right\}(2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b^*|^{-1/2}$$

$$\times\int_{\boldsymbol{b}}\exp\left\{-\frac{1}{2}\left[\left(\Sigma_b^{*-1/2}\boldsymbol{b}\right)^T\left(\Sigma_b^{*-1/2}\boldsymbol{b}\right) - 2\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)} + \boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\boldsymbol{\Sigma}_b^{*-1/2}\boldsymbol{b}\right.\right.$$

$$
\begin{aligned}
&+\left[\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\right]\left[\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\right]^T\Bigg]\\
&+\frac{1}{2}\left[\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\right]\left[\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\right]^T\\
&-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)}-\frac{1}{2}\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\boldsymbol{\mu}_1^*\Bigg\}w_1^*d\boldsymbol{b}\\
&=\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^*+\widetilde{\boldsymbol{X}}_jt_j)}{A(\phi_j^*)}+C(Y_j;\phi_j^*)\right]\right\}(2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b^*|^{-1/2}w_1^*\\
&\quad\times\exp\left\{\frac{1}{2}\left[\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\right]\left[\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^{*1/2}\right]^T-\frac{1}{2}\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\boldsymbol{\mu}_1^*\right\}\\
&\quad\times\int_{\boldsymbol{b}}\exp\left\{-\frac{1}{2}\left[\Sigma_b^{*-1/2}\boldsymbol{b}-\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)^T\right]^T\left[\Sigma_b^{*-1/2}\boldsymbol{b}_1-\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)^T\right]\right\}\\
&\qquad\times\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)}\right\}d\boldsymbol{b}\\
&=\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^*+\widetilde{\boldsymbol{X}}_jt_j)}{A(\phi_j^*)}+C(Y_j;\phi_j^*)\right]\right\}\\
&\quad\times w_1^*\exp\left\{\frac{1}{2}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)\boldsymbol{\Sigma}_b^*\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}+\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)^T-\frac{1}{2}\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\boldsymbol{\mu}_1^*\right\}\\
&\quad\times\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\left(\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)}\right)\right\}\right]\\
&=\exp\left\{\sum_{j=1}^{n_N}\left[\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}^*+\widetilde{\boldsymbol{X}}_jt_j)}{A(\phi_j^*)}+C(Y_j;\phi_j^*)\right]\right\}\\
&\quad\times w_1^*\exp\left\{\frac{1}{2}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)\boldsymbol{\Sigma}_b^*\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)^T+\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)\boldsymbol{\Sigma}_b^*\left(\boldsymbol{\mu}_1^{*T}\boldsymbol{\Sigma}_b^{*-1}\right)^T\right\}\\
&\quad\times\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)}\right\}\right]\\
&=\exp\left\{\sum_{j=1}^{n_N}C(Y_j;\phi_j^*)\right\}\\
&\quad\times w_1^*\exp\left\{\frac{1}{2}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)\boldsymbol{\Sigma}_b^*\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)^T+\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_j^*)}\left(\boldsymbol{X}_j\boldsymbol{\beta}^*+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_1^*\right)\right\}\\
&\quad\times\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}_\alpha+t_j)}{A(\phi_j^*)}\right\}\right]\\
&=\exp\left\{\frac{1}{2}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)\boldsymbol{\Sigma}_b^*\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)^T+\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_j^*)}\left(\boldsymbol{X}_j\boldsymbol{\beta}^*+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_1^*\right)\right\}\\
&\quad\times\exp\left\{\sum_{j=1}^{n_N}C(Y_j;\phi_j^*)\right\}w_1^*\,\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)}\right\}\right]. \qquad \text{(B.021)}
\end{aligned}
$$

Likewise, (B.020) becomes

$$\exp\left\{\frac{1}{2}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)\boldsymbol{\Sigma}_{b0}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)^T+\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\left(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{10}\right)\right\}$$
$$\times\exp\left\{\sum_{j=1}^{n_N}C(Y_j;\phi_{j0})\right\}w_{10}\,\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}+t_j)}{A(\phi_{j0})}\right\}\right]. \qquad \text{(B.022)}$$

Then, to compare the coefficients of $\boldsymbol{Y}^T\boldsymbol{Y}$ and $\boldsymbol{Y}$ in the exponential part and the constant term out of the exponential part from (B.021) and (B.022), we have

$$\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)\boldsymbol{\Sigma}_b^*\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j^*)}\right)^T=\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)\boldsymbol{\Sigma}_{b0}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)^T, \qquad \text{(B.023)}$$

$$\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_j^*)}\left(\boldsymbol{X}_j\boldsymbol{\beta}^*+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_1^*\right)=\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\left(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{10}\right), \qquad \text{(B.024)}$$

and

$$\exp\left\{\sum_{j=1}^{n_N}C(Y_j;\phi_j^*)\right\}w_1^*\,\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}+t_j)}{A(\phi_j^*)}\right\}\right]$$
$$=\exp\left\{\sum_{j=1}^{n_N}C(Y_j;\phi_{j0})\right\}w_{10}\,\mathrm{E}_{\boldsymbol{b}|\alpha=1}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}+t_j)}{A(\phi_{j0})}\right\}\right]. \qquad \text{(B.025)}$$

Furthermore, by the assumption of the generalized linear mixed model with canonical link function for longitudinal outcome $Y(t)$ at time t, we have $\mu(t)=\mathrm{E}(Y(t)|\boldsymbol{b})=B'(\eta(t))$ and $v(t)=\mathrm{Var}(Y(t)|\boldsymbol{b})=B''(\eta(t))A(\phi(t))$, where $\boldsymbol{b}=\sum_{k=1}^K I(\alpha=k)\boldsymbol{b}_k$, $\eta(t)=g(\mu(t))=\boldsymbol{X}(t)\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}(t)\boldsymbol{b}$, $v(t)=v(\mu(t))A(\phi(t))$, $g(\cdot)$ and $v(\cdot)$ are known link and variance functions respectively, and $B'(\eta(t))$ and $B''(\eta(t))$ are the first and second derivatives of $B(\eta(t))$ with respect to the canonical parameter $\eta(t)$. Hence, we have

$$\mathrm{E}(Y_j|\boldsymbol{b},\alpha=1) = B'(\eta_j|\alpha=1)=B'(\boldsymbol{\beta}^*;\boldsymbol{b}|\alpha=1)=B'(\boldsymbol{\beta}_0;\boldsymbol{b}|\alpha=1) \qquad \text{(B.026)}$$

and
$$\mathrm{Var}(Y_j|\boldsymbol{b},\alpha=1) = B''(\eta_j|\alpha=1)A(\phi_j)=B''(\boldsymbol{\beta}^*;\boldsymbol{b}|\alpha=1)A(\phi_j^*) = B''(\boldsymbol{\beta}_0;\boldsymbol{b}|\alpha=1)A(\phi_{j0}). \qquad \text{(B.027)}$$

By the continuous mapping theorem and (B.026), we obtain $\boldsymbol{\beta}^*=\boldsymbol{\beta}_0$. Then, (B.027) becomes $B''(\boldsymbol{\beta}_0;\boldsymbol{b}|\alpha=1)A(\phi_j^*)=B''(\boldsymbol{\beta}_0;\boldsymbol{b}|\alpha=1)A(\phi_{j0})$. Hence, by assumption (A6), $A(\phi_j^*)=A(\phi_{j0})$, and, by the continuous mapping theorem, we obtain $\phi_j^*=\phi_{j0}$, $j=1,\ldots,n_N$. Therefore, (B.023) can be written as

$$\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)\boldsymbol{\Sigma}_b^*\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)^T=\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)\boldsymbol{\Sigma}_{b0}\left(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}\right)^T.$$

Then, by assumption (A6), we obtain $\boldsymbol{\Sigma}_b^*=\boldsymbol{\Sigma}_{b0}$. Since $\boldsymbol{\beta}^*=\boldsymbol{\beta}_0$ and $\phi_j^*=\phi_{j0}$, (B.024) can be written as

$$\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\left(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_1^*\right)=\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\left(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_jt_j+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{10}\right).$$

Also, by assumption (A6), we obtain $\boldsymbol{\mu}_1^*=\boldsymbol{\mu}_{10}$. In (B.025) for the constant terms, note that the random effect $\boldsymbol{b}$ given $\alpha=1$ on the left-hand side follows a multivariate normal distribution with mean $\boldsymbol{\Sigma}_b^*\big(\sum_{j=1}^{n_N}Y_j\widetilde{\boldsymbol{X}}_j/A(\phi_j^*)\big)^T+\boldsymbol{\mu}_1^*$ and covariance $\boldsymbol{\Sigma}_b^*$ and the random effect $\boldsymbol{b}$ given $\alpha=1$ on the right-hand side follows a multivariate normal distribution with mean $\boldsymbol{\Sigma}_{b0}\big(\sum_{j=1}^{n_N}Y_j\widetilde{\boldsymbol{X}}_j/A(\phi_{j0})\big)^T+\boldsymbol{\mu}_{10}$ and covariance $\boldsymbol{\Sigma}_{b0}$. (i) Because $\boldsymbol{\Sigma}_b^*=\boldsymbol{\Sigma}_{b0}$, $\boldsymbol{\mu}_1^*=\boldsymbol{\mu}_{10}$ and $\phi_j^*=\phi_{j0}$, $j=1,\ldots,n_N$, the random effects $\boldsymbol{b}$'s given $\alpha=1$ on both sides follow the same multivariate normal distribution. (ii) Besides, because $\boldsymbol{\beta}^*=\boldsymbol{\beta}_0$ and $\phi_j^*=\phi_{j0}$, $j=1,\ldots,n_N$, we have $\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b}|\alpha=1)}{A(\phi_j^*)}=\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}|\alpha=1)}{A(\phi_{j0})}$. By (i) and (ii), we obtain (iii) $\mathrm{E}_{\boldsymbol{b}|\alpha=1}\big[\exp\big\{\frac{B(\boldsymbol{\beta}^*;\boldsymbol{b})}{A(\phi_j^*)}\big\}\big] = \mathrm{E}_{\boldsymbol{b}|\alpha=1}\big[\exp\big\{\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b})}{A(\phi_{j0})}\big\}\big]$. Also, (iv) since $\phi_j^* = \phi_{j0}$, $j = 1,\ldots,n_N$, we have $\exp\big\{\sum_{j=1}^{n_N} C(Y_j;\phi_j^*)\big\} = \exp\big\{\sum_{j=1}^{n_N} C(Y_j;\phi_{j0})\big\}$. By (iii) and (iv), (B.025) can be written as

$$\exp\Bigg\{\sum_{j=1}^{n_N} C(Y_j;\phi_{j0})\Bigg\} w_1^* \,\mathrm{E}_{\boldsymbol{b}|\alpha=1}\Bigg[\exp\Bigg\{\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}+t_j)}{A(\phi_{j0})}\Bigg\}\Bigg]$$
$$= \exp\Bigg\{\sum_{j=1}^{n_N} C(Y_j;\phi_{j0})\Bigg\} w_{10}\,\mathrm{E}_{\boldsymbol{b}|\alpha=1}\Bigg[\exp\Bigg\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}+t_j)}{A(\phi_{j0})}\Bigg\}\Bigg].$$

Then, by assumption (A6), we obtain $w_1^* = w_{10}$.

For $\boldsymbol{\mu}_2^* = \boldsymbol{\mu}_{20}$ and $w_2^* = w_{20}$, we go back to (B.017). Since $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$, $\boldsymbol{\phi}^* = \boldsymbol{\phi}_0$, $\boldsymbol{\Sigma}_b^* = \boldsymbol{\Sigma}_{b0}$, $\boldsymbol{\mu}_1^* = \boldsymbol{\mu}_{10}$, and $w_1^* = w_{10}$, the most dominating terms of the left and right hand sides of (B.017) are canceled out. Then, the next dominating terms of the both sides of (B.017) are for $\alpha = 2$. Thus, by replacing $\boldsymbol{\mu}_1^*$, w_1^*, $\boldsymbol{\mu}_{10}$ and w_{10} with $\boldsymbol{\mu}_2^*$, w_2^*, $\boldsymbol{\mu}_{20}$ and w_{20} in (B.019) and (B.020) and comparing (B.019) and (B.020) for $\alpha = 2$ which are for one normal distribution, we can obtain $\boldsymbol{\mu}_2^* = \boldsymbol{\mu}_{20}$ and $w_2^* = w_{20}$ through the same argument done for $\alpha = 1$. By repeating this also for $\alpha = 3,\ldots,K$, we finally can obtain $\boldsymbol{\mu}_\alpha^* = \boldsymbol{\mu}_{\alpha 0}$, and $w_\alpha^* = w_{\alpha 0}$, $\alpha = 1,\ldots,K$,

Next, to show that $\boldsymbol{\psi}^* = \boldsymbol{\psi}_0$, $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_0$ and $\Lambda_s^* = \Lambda_{s0}$, we also let $\Delta_s = 0$ in (B.014) and, with $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$, $\boldsymbol{\phi}^* = \boldsymbol{\phi}_0$, $\boldsymbol{\Sigma}_b^* = \boldsymbol{\Sigma}_{b0}$, $\boldsymbol{\mu}_\alpha^* = \boldsymbol{\mu}_{\alpha 0}$, and $w_\alpha^* = w_{\alpha 0}$, $\alpha = 1,\ldots,K$, (B.014) can be expressed as, for $\alpha = 1,\ldots,K$,

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{b}|\alpha}\Bigg[\exp\Bigg\{-\int_0^{V_s}\exp\big\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}^*\circ\boldsymbol{b}) + \boldsymbol{Z}(t)\boldsymbol{\gamma}^*\big\}d\Lambda_s^*(t)\Bigg\}\Bigg]\\ &= \mathrm{E}_{\boldsymbol{b}|\alpha}\Bigg[\exp\Bigg\{-\int_0^{V_s}\exp\big\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0\big\}d\Lambda_{s0}(t)\Bigg\}\Bigg],\end{aligned} \tag{B.028}$$

where the random effects $\boldsymbol{b}$'s on both sides follow a multivariate normal distribution with mean $\boldsymbol{\Sigma}_{b0}\big(\sum_{j=1}^{n_N} Y_j\widetilde{\boldsymbol{X}}_j/A(\phi_{j0})\big)^T + \boldsymbol{\mu}_{\alpha 0}$ and covariance $\boldsymbol{\Sigma}_{b0}$.

For any fixed $\widetilde{\boldsymbol{X}} = \big(\widetilde{\boldsymbol{X}}_1^T,\ldots,\widetilde{\boldsymbol{X}}_{n_N}^T\big)^T$, treating $\widetilde{\boldsymbol{X}}^T\boldsymbol{Y}$ as a parameter in this normal family, $\boldsymbol{b} = \sum_{k=1}^K I(\alpha = k)\boldsymbol{b}_k$ is the complete statistic for $\widetilde{\boldsymbol{X}}^T\boldsymbol{Y}$. Therefore,

$$\exp\Bigg\{-\int_0^{V_s}\exp\big\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}^*\circ\boldsymbol{b}) + \boldsymbol{Z}(t)\boldsymbol{\gamma}^*\big\}d\Lambda_s^*(t)\Bigg\}$$
$$= \exp\Bigg\{-\int_0^{V_s}\exp\big\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0\big\}d\Lambda_{s0}(t)\Bigg\},$$

and equivalently we have

$$\exp\big\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}^*\circ\boldsymbol{b}) + \boldsymbol{Z}(t)\boldsymbol{\gamma}^*\big\}\lambda_s^*(t) = \exp\big\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0\big\}\lambda_{s0}(t).$$

By assumptions (A3) and (A6), $\boldsymbol{\psi}^* = \boldsymbol{\psi}_0$, $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_0$ and $\Lambda_s^* = \Lambda_{s0}$. ♯

Since all the three steps are completed, we can conclude that, with probability one, $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\Lambda}}$ converges to $\boldsymbol{\Lambda}_0$ in $[0,\tau]$. Moreover, as mentioned in the beginning of this proof for consistency, since $\boldsymbol{\Lambda}_0$ is continuous in $[0,\tau]$, the latter can be strengthened to uniform convergence; that is, $\sup_{t\in[0,\tau]}\|\widehat{\boldsymbol{\Lambda}}(t) - \boldsymbol{\Lambda}_0(t)\| \to 0$ almost surely. Therefore, Theorem 1 is proved.

B.2. Proof of Asymptotic Normality – Theorem 2

Asymptotic distribution for the proposed estimators can be shown if we can verify the conditions of Theorem 3.3.1 in van der Vaart and Wellner (1996). Then, it will be shown that the distribution is Gaussian. For completeness, we use Theorem 4 in Parner (1998) which restated the Theorem 3.3.1 of van der Vaart and Wellner (1996).

Theorem 4 (Parner 1998) *Let U_n and U be random maps and a fixed map, respectively, from ξ to a Banach space such that:*
(a) $\sqrt{n}(U_n - U)(\widehat{\xi}_n) - \sqrt{n}(U_n - U)(\xi_0) = o_P^*(1 + \sqrt{n}\|\widehat{\xi}_n - \xi_0\|)$.
(b) The sequence $\sqrt{n}(U_n - U)(\xi_0)$ converges in distribution to a tight random element Z.
(c) the function $\xi \to U(\xi)$ is Fréchet differentiable at ξ_0 with a continuously invertible derivative ∇U_{ξ_0} (on its range).
(d) U_{ξ_0} and $\widehat{\xi}_n$ satisfies $U_n(\widehat{\xi}_n) = o_P^*(n^{-1/2})$ and converges in outer probability to ξ_0.
Then $\sqrt{n}(\widehat{\xi}_n - \xi_0) \Rightarrow \nabla U_{\xi_0}^{-1} Z$.

We will prove the conditions (a)~(d). In our situation, the parameter $\xi_s = (\boldsymbol{\theta}, \boldsymbol{\Lambda}_s) \in \Xi = \{(\boldsymbol{\theta}, \boldsymbol{\Lambda}_s) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \sup_{t\in[0,\tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \leq \delta,\ s = 1, \ldots, S\}$ for a fixed small constant δ. We note that Ξ is a convex set. Define a set $\mathscr{H} = \{(\mathbf{h}_1, h_2) : \|\mathbf{h}_1\| \leq 1, \|h_2\|_V \leq 1\}$, where $\|h_2\|_V$ is the total variation of h_2 in $[0, \tau]$ defined as

$$\sup_{0=t_0\leq t_1\leq\cdots\leq t_l=\tau} \sum_{j=1}^{l} |h_2(t_j) - h_2(t_{j-1})|.$$

Furthermore, we define that, for stratum s,

$$\begin{aligned} U_{m_s}(\xi_s)(\mathbf{h}_1, h_2) &= \mathbf{P}_{m_s}\{l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2]\} \\ \text{and} \qquad U_s(\xi_s)(\mathbf{h}_1, h_2) &= \mathbf{P}\{l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2]\}, \end{aligned}$$

where $l_\theta(\boldsymbol{\theta}, \Lambda_s)$ is the first derivative of the log-likelihood function from one single subject belonging to stratum s, denoted by $l(\boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)$, with respect to $\boldsymbol{\theta}$, and $l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)$ is the derivative of $l(\boldsymbol{O}; \boldsymbol{\theta}, \Lambda_{s\varepsilon})$ at $\varepsilon = 0$, where $\Lambda_{s\varepsilon}(t) = \int_0^t (1 + \varepsilon h_2(u)) d\Lambda_s(u)$. Therefore, we can see that both U_{m_s} and U_s map from Ξ to $\ell^\infty(\mathscr{H})$ and $\sqrt{m_s}\{U_{m_s}(\xi_s) - U_s(\xi_s)\}$ is an empirical process in the space $\ell^\infty(\mathscr{H})$.

Denote $(\mathbf{h}_1^\beta, \mathbf{h}_1^\phi, \mathbf{h}_1^{\Sigma_b}, \mathbf{h}_1^\mu, \mathbf{h}_1^w, \mathbf{h}_1^\psi, \mathbf{h}_1^\gamma)$ as the corresponding components of $\mathbf{h}_1$ for the parameters $(\boldsymbol{\beta}, \boldsymbol{\phi}, \mathrm{Vec}(\boldsymbol{\Sigma}_b),\ \boldsymbol{\mu}, \boldsymbol{w}, \boldsymbol{\psi}, \boldsymbol{\gamma})$, respectively. From Section C.2, for any $(\mathbf{h}_1, h_2) \in \mathscr{H}$, the class

$$\begin{aligned} \mathscr{G} = \Big\{ & l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}_0, \Lambda_{s0})[h_2], \\ & \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \sup_{t\in[0,\tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \leq \delta,\ (\mathbf{h}_1, h_2) \in \mathscr{H} \Big\} \end{aligned}$$

is shown as P-Donsker (Section 2.1 of [4]), and it is also implied that

$$\sup_{(\mathbf{h}_1, h_2)\in\mathscr{H}} \mathbf{P}\left[l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}_0, \Lambda_{s0})[h_2]\right]^2 \longrightarrow 0$$

as $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \sup_{t\in[0,\tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \to 0$. Then we conclude the followings:

(a) follows from Lemma 3.3.5 (p311) of [4].

(b) holds as a result of Section C.2 and the convergence is defined in the metric space $\ell^\infty(\mathscr{H})$ by the Donsker theorem (Section 2.5 of [4]).

(d) is true because $(\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)$ maximizes $\mathbf{P}_{m_s} l(\boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)$, $(\boldsymbol{\theta}_0, \Lambda_{s0})$ maximizes $\mathbf{P}\, l(\boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)$, and $(\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)$ converges to $(\boldsymbol{\theta}_0, \Lambda_{s0})$ from Theorem 1.

Now, we need to verify the conditions in (c). Since the proof of the first half in (c), that the function $\xi \to U(\xi)$ is *Fréchet* differentiable at ξ_0, is given in Section C.3, we will only prove that the derivative ∇U_{ξ_0} is continuously invertible on its range $\ell^\infty(\mathscr{H})$. According to Section C.3, ∇U_{ξ_0} can be expressed as follows: for any $(\boldsymbol{\theta}_1, \Lambda_{s1})$ and $(\boldsymbol{\theta}_2, \Lambda_{s2})$ in Ξ,

$$\nabla U_{\xi_0}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2, \Lambda_{s1} - \Lambda_{s2})[\mathbf{h}_1, h_2] = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \Omega_1[\mathbf{h}_1, h_2] + \int_0^\tau \Omega_2[\mathbf{h}_1, h_2] d(\Lambda_{s1} - \Lambda_{s2})(t), \tag{B.029}$$

where both Ω_1 and Ω_2 are linear operators on $\mathscr{H}$, and $\Omega = (\Omega_1, \Omega_2)$ maps $\mathscr{H} \subset \boldsymbol{R}^d \times \mathrm{BV}[0, \tau]$ to $\boldsymbol{R}^d \times \mathrm{BV}[0, \tau]$, where $\mathrm{BV}[0, \tau]$ contains all the functions with finite total variation in $[0, \tau]$. The explicit expressions of Ω_1 and Ω_2 are given in Section C.3. From (B.029), we can treat $(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2, \Lambda_{s1} - \Lambda_{s2})$ as an element in $\ell^\infty(\mathscr{H})$ via the following definition:

$$(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2, \Lambda_{s1} - \Lambda_{s2})[\mathbf{h}_1, h_2] = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \mathbf{h}_1 + \int_0^\tau h_2(t) d(\Lambda_{s1} - \Lambda_{s2})(t), \quad \forall (\mathbf{h}_1, h_2) \in \boldsymbol{R}^d \times \mathrm{BV}[0,\tau].$$

Then ∇U_{ξ_0} can be expanded as a linear operator from $\ell^\infty(\mathscr{H})$ to itself. Therefore, if we can show that there exists some positive constant ε such that $\varepsilon\mathscr{H} \subset \Omega(\mathscr{H})$, then we will have that for any $(\delta\boldsymbol{\theta}, \delta\Lambda_s) \in \ell^\infty(\mathscr{H})$,

$$\begin{aligned} \|\nabla U_{\xi_0}(\delta\boldsymbol{\theta}, \delta\Lambda_s)\|_{\ell^\infty(\mathscr{H})} &= \sup_{(\mathbf{h}_1,h_2)\in\mathscr{H}} \left|\delta\boldsymbol{\theta}^T \Omega_1[\mathbf{h}_1, h_2] + \int_0^\tau \Omega_2[\mathbf{h}_1, h_2] d\delta\Lambda_s(t)\right| \\ &= \|(\delta\boldsymbol{\theta}, \delta\Lambda_s)\|_{\ell^\infty(\Omega(\mathscr{H}))} \ge \varepsilon \|(\delta\boldsymbol{\theta}, \delta\Lambda_s)\|_{\ell^\infty(\mathscr{H})}, \end{aligned}$$

and ∇U_{ξ_0} will be continuously invertible.

Note that to prove $\varepsilon\mathscr{H} \subset \Omega(\mathscr{H})$ for some ε is equivalent to showing that Ω is invertible. We also note from Section C.3, that Ω is the summation of an invertible operator and a compact operator. By Theorem 4.25 of [**?**], for the proof of the invertibility of Ω, it is sufficient to verify that Ω is one to one: if $\Omega[\mathbf{h}_1, h_2] = 0$, then, by choosing $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 = \varepsilon^* \mathbf{h}_1$ and $\Lambda_{s1} - \Lambda_{s2} = \varepsilon^* \int h_2 d\Lambda_{s0}$ in (B.029) for a small constant ε^*, we obtain

$$\nabla U_{\xi_0}(\mathbf{h}_1, \int h_2 d\Lambda_{s0})[\mathbf{h}_1, h_2] = \varepsilon^*(\mathbf{h}_1^T, h_2) \begin{pmatrix} \Omega_1[\mathbf{h}_1, h_2] \\ \Omega_2[\mathbf{h}_1, h_2] \end{pmatrix} = \varepsilon^*(\mathbf{h}_1^T, h_2)\Omega[\mathbf{h}_1, h_2] = 0.$$

By the definition of ∇U_{ξ_0}, we note that $\nabla U_{\xi_0}(\mathbf{h}_1, \int h_2 d\Lambda_{s0})[\mathbf{h}_1, h_2]$ is the negative information matrix in the submodel $(\boldsymbol{\theta}_0 + \varepsilon\mathbf{h}_1, \Lambda_{s0} + \varepsilon \int h_2 d\Lambda_{s0})$. Thus, the score function along this submodel should be zero with probability one; that is, $l_\theta(\boldsymbol{\theta}_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}_0, \Lambda_{s0})[h_2] = 0$; that is, with probability one, for the numerator of the score function

$$\begin{aligned} 0 = & \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) \times \Bigg[\sum_{j=1}^{n_N} \frac{1}{A(\phi_{j0})} \big(Y_j \boldsymbol{X}_j - B'(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)\big) \mathbf{h}_1^\beta \\ & + \sum_{j=1}^{n_N} \left\{ -\left(\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}_0 + \widetilde{\boldsymbol{X}}_j \boldsymbol{b}_\alpha) - B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})^2} \right) A'(\phi_{j0}) + C'(Y_j; \phi_{j0}) \right\} \mathrm{h}_1^{\phi_j} \\ & + \frac{1}{2}(\boldsymbol{b}_\alpha - \boldsymbol{\mu}_{\alpha 0})^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} (\boldsymbol{b}_\alpha - \boldsymbol{\mu}_{\alpha 0}) - \frac{1}{2} \operatorname{Tr}\left(\boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b\right) \\ & + \left(\boldsymbol{b}_\alpha - \frac{1}{2}\boldsymbol{\mu}_{\alpha 0}\right)^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}} \mathbf{h}_1^{w_{\alpha 0}} + \Delta_s \big\{ (\widetilde{\boldsymbol{Z}}(V_s) \circ \boldsymbol{b}_\alpha^T) \mathbf{h}_1^\psi + \boldsymbol{Z}(V_s) \mathbf{h}_1^\gamma \big\} \\ & - \int_0^{V_s} \exp\left\{ \widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0 \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0 \right\} \times \left\{ (\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{b}_\alpha^T) \mathbf{h}_1^\psi + \boldsymbol{Z}(t) \mathbf{h}_1^\gamma \right\} d\Lambda_{s0}(t) \Bigg] d\boldsymbol{b}_\alpha \\ & + \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) \times \Bigg[\Delta_s h_2(V_s) \\ & \qquad - \int_0^{V_s} h_2(t) \exp\left\{ \widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(t)\boldsymbol{\gamma}_0 \right\} d\Lambda_{s0}(t) \Bigg] d\boldsymbol{b}_\alpha, \end{aligned} \tag{B.030}$$

where $A'(\phi_{j0})$and $C'(Y_j; \phi_{j0})$ are the derivatives of $A(\phi_j)$and $C(Y_j; \phi_j)$ with respect to ϕ_j evaluated at ϕ_{j0} and $B'(\boldsymbol{\beta}_0; \boldsymbol{b})$ is the derivative of $B(\boldsymbol{\beta}; \boldsymbol{b})$ with respect to $\boldsymbol{\beta}$ evaluated at $\boldsymbol{\beta}_0$. Note that (B.030) holds with probability one, so it may not hold for any $V_s \in [0, \tau]$ when $\Delta_s = 0$. However, by the similar arguments done in Section B.1, if we integrate both sides from V_s to τ and substract the obtained equation from (B.030) at $\Delta_s = 0$ and $V_s = \tau$, it is easily shown that (B.030) also holds for any $V_s \in [0, \tau]$ when $\Delta_s = 0$. Hence, the proof of the invertibility of Ω will be completed if we can show $\mathbf{h}_1 = 0$ and $h_2(t) = 0$ from (B.030).

To show $\mathbf{h}_1 = 0$, particularly we let $\Delta_s = 0$ and $V_s = 0$ in (B.030) and obtain

$$\begin{aligned}
0 &= \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b},\alpha,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) \times \Bigg[\sum_{j=1}^{n_N} \frac{1}{A(\phi_{j0})}\big(Y_j \boldsymbol{X}_j - B'(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)\big)\mathbf{h}_1^\beta \\
&\quad + \sum_{j=1}^{n_N}\left\{ -\left(\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}_0 + \widetilde{\boldsymbol{X}}_j\boldsymbol{b}_\alpha) - B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})^2}\right)A'(\phi_{j0}) + C'(Y_j;\phi_{j0})\right\} \mathrm{h}_1^{\phi_j} \\
&\quad + \frac{1}{2}(\boldsymbol{b}_\alpha - \boldsymbol{\mu}_{\alpha 0})^T \boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\boldsymbol{\Sigma}_{b0}^{-1}(\boldsymbol{b}_\alpha - \boldsymbol{\mu}_{\alpha 0}) - \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\right) \\
&\quad + \left(\boldsymbol{b}_\alpha - \frac{1}{2}\boldsymbol{\mu}_{\alpha 0}\right)^T \boldsymbol{\Sigma}_{b0}^{-1}\mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}}\mathbf{h}_1^{w_\alpha}\Bigg] d\boldsymbol{b}_\alpha \\
&= \sum_\alpha \Bigg[\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha 0}\right\} w_{\alpha 0} \\
&\quad \times \Bigg[\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\}\right] \times \Bigg(\sum_{j=1}^{n_N}\frac{1}{A(\phi_{j0})} Y_j \boldsymbol{X}_j \mathbf{h}_1^\beta \\
&\qquad + \sum_{j=1}^{n_N}\left\{ -\frac{Y_j\boldsymbol{X}_j\boldsymbol{\beta}_0}{A(\phi_{j0})^2}A'(\phi_{j0}) + C'(Y_j;\phi_{j0})\right\} \mathrm{h}_1^{\phi_j} \\
&\qquad + \frac{1}{2}\boldsymbol{\mu}_{\alpha 0}^T\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{\mu}_{\alpha 0} - \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\right) - \frac{1}{2}\boldsymbol{\mu}_{\alpha 0}^T\boldsymbol{\Sigma}_{b0}^{-1}\mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}}\mathbf{h}_1^{w_\alpha}\Bigg) \\
&\quad + \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\Bigg[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\} \times \Bigg(-\sum_{j=1}^{n_N}\frac{B'(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\mathbf{h}_1^\beta \\
&\qquad + \sum_{j=1}^{n_N}\left\{ -\left(\frac{Y_j\widetilde{\boldsymbol{X}}_j\boldsymbol{b}_\alpha - B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})^2}\right)A'(\phi_{j0})\right\} \mathrm{h}_1^{\phi_j} \\
&\qquad + \frac{1}{2}\left(\boldsymbol{b}_\alpha^T\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{b}_\alpha - 2\boldsymbol{b}_\alpha^T\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{\mu}_\alpha\right) + \boldsymbol{b}_\alpha^T\boldsymbol{\Sigma}_{b0}^{-1}\mathbf{h}_1^{\mu_\alpha}\Bigg)\Bigg]\Bigg]\Bigg]. \qquad \text{(B.031)}
\end{aligned}$$

We first examine the coefficient for $\boldsymbol{Y}$ in (B.031).

$$\begin{aligned}
&\sum_{j=1}^{n_N}\Bigg\{ Y_j\left(\frac{1}{A(\phi_{j0})}\boldsymbol{X}_j\mathbf{h}_1^\beta - \frac{1}{A(\phi_{j0})^2}\boldsymbol{X}_j\boldsymbol{\beta}_0 A'(\phi_{j0})\ \mathrm{h}_1^{\phi_j}\right) \\
&\qquad \times \left[\sum_\alpha\left[\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha 0}\right\} w_{\alpha 0}\,\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\}\right]\right]\right]\Bigg\} \\
&- \sum_{j=1}^{n_N}\Bigg\{ \frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})^2}A'(\phi_{j0})\ \mathrm{h}_1^{\phi_j} \\
&\qquad \times \left[\sum_\alpha\left[\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha 0}\right\} w_{\alpha 0}\,\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\}\boldsymbol{b}_\alpha\right]\right]\right]\Bigg\} \\
&= \sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\Bigg\{ \boldsymbol{X}_j\left(\mathbf{h}_1^\beta - \frac{1}{A(\phi_{j0})}\boldsymbol{\beta}_0 A'(\phi_{j0})\ \mathrm{h}_1^{\phi_j}\right) \\
&\qquad \times \left[\sum_\alpha\left[\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha 0}\right\} w_{\alpha 0}\,\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\}\right]\right]\right] \\
&\quad - \frac{\widetilde{\boldsymbol{X}}_j}{A(\phi_{j0})}A'(\phi_{j0})\ \mathrm{h}_1^{\phi_j} \\
&\qquad \times \left[\sum_\alpha\left[\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha 0}\right\} w_{\alpha 0}\,\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\left[\exp\left\{-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\}\boldsymbol{b}_\alpha\right]\right]\right]\Bigg\} \\
&= 0
\end{aligned}$$

Then, for all $j = 1, \ldots, n_N$,

$$
\begin{aligned}
&\boldsymbol{X}_j \mathbf{h}_1^{\beta} \sum_{\alpha} \Bigg[\exp\Big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} \Big\} w_{\alpha 0} \, \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha} \Big[\exp\Big\{ - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \Big] \Bigg] \\
&- \frac{1}{A(\phi_{j0})} A'(\phi_{j0}) \, \mathrm{h}_1^{\phi_j} \Bigg(\boldsymbol{X}_j \boldsymbol{\beta}_0 \sum_{\alpha} \Bigg[\exp\Big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} \Big\} w_{\alpha 0} \, \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha} \Big[\exp\Big\{ - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \Big] \Bigg] \\
&\qquad + \widetilde{\boldsymbol{X}}_j \sum_{\alpha} \Bigg[\exp\Big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} \Big\} w_{\alpha 0} \, \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha} \Big[\exp\Big\{ - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \boldsymbol{b}_\alpha \Big] \Bigg] \Bigg) \\
&= 0.
\end{aligned}
$$

Based on assumption (A6), $\mathbf{h}_1^{\beta} = 0$ and $\mathrm{h}_1^{\phi_j} = 0$, $j = 1, \ldots, n_N$.

Then, we examine the constant terms without $\boldsymbol{Y}$ in (B.031). Since $\mathbf{h}_1^{\beta} = 0$ and $\mathrm{h}_1^{\phi_j} = 0$, $j = 1, \ldots, n_N$, (B.031) becomes

$$
\begin{aligned}
&\sum_{\alpha} \Bigg[\exp\Big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} \Big\} w_{\alpha 0} \Bigg[\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha} \Big[\exp\Big\{ - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \Big] \\
&\qquad \times \Big(\frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_{\alpha 0} - \frac{1}{2} \mathrm{Tr}\big(\boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \big) - \frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}} \mathbf{h}_1^{w_\alpha} \Big) \\
&\quad + \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha} \Big[\exp\Big\{ - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \times \Big(\frac{1}{2} \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{b}_\alpha - \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_\alpha + \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} \Big) \Big] \Bigg] \Bigg] \\
&= \mathrm{E}_{\alpha, \boldsymbol{b}} \Bigg[\exp\Big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} \Big\} \times \exp\Big\{ - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \times \Big(\frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_{\alpha 0} \\
&\qquad - \frac{1}{2} \mathrm{Tr}\big(\boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \big) - \frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}} \mathbf{h}_1^{w_\alpha} + \frac{1}{2} \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{b}_\alpha - \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_\alpha \\
&\qquad + \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} \Big) \Bigg] \\
&= 0,
\end{aligned}
$$

where $\boldsymbol{b}$ follows a multivariate normal distribution with mean $\boldsymbol{\Sigma}_{b0}\big(\sum_{j=1}^{n_N} (Y_j \widetilde{\boldsymbol{X}}_j / A(\phi_{j0})) \big) + \boldsymbol{\mu}_{\alpha 0}$ and covariance $\boldsymbol{\Sigma}_{b0}$. For any fixed $\widetilde{\boldsymbol{X}}$, treating $\boldsymbol{X}^T \boldsymbol{Y}$ as a parameter in this normal family, $\boldsymbol{b} = \sum_{k=1}^{K} I(\alpha = k) \boldsymbol{b}_k$ is the complete statistic for $\boldsymbol{X}^T \boldsymbol{Y}$, therefore,

$$
\begin{aligned}
&\exp\Big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \Big\} \times \Big(\frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_{\alpha 0} - \frac{1}{2} \mathrm{Tr}\big(\boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \big) \\
&\quad - \frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}} \mathbf{h}_1^{w_\alpha} + \frac{1}{2} \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{b}_\alpha - \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_\alpha + \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} \Big) \; = \; 0.
\end{aligned}
$$

Since $\exp\big\{ \sum_{j=1}^{n_N} \frac{Y_j}{A(\phi_{j0})} \widetilde{\boldsymbol{X}}_j \boldsymbol{\mu}_{\alpha 0} - \sum_{j=1}^{n_N} \frac{B(\boldsymbol{\beta}_0; \boldsymbol{b}_\alpha)}{A(\phi_{j0})} \big\} \neq 0$, by (A6), we have

$$
\begin{aligned}
&\frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_{\alpha 0} - \frac{1}{2} \mathrm{Tr}\big(\boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \big) + \Big(- \frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T + \boldsymbol{b}_\alpha^T \Big) \boldsymbol{\Sigma}_{b0}^{-1} \mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_{\alpha 0}} \mathbf{h}_1^{w_\alpha} \\
&\qquad + \frac{1}{2} \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{b}_\alpha - \boldsymbol{b}_\alpha^T \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_{b0}^{-1} \boldsymbol{\mu}_\alpha \; = \; 0.
\end{aligned}
$$

Then, by (A6), $-\frac{1}{2} \boldsymbol{\mu}_{\alpha 0}^T + \boldsymbol{b}_\alpha^T \neq 0$ and $\boldsymbol{\Sigma}_{b0}^{-1} \neq 0$ lead to $\mathbf{h}_1^{\mu_\alpha} = 0$, $\alpha = 1, \ldots, K$. Likewise, by (A6), $1/w_\alpha \neq 0$ and $\boldsymbol{\Sigma}_{b0}^{-1} \neq 0$ result in $\mathbf{h}_1^{w_\alpha} = 0$, $\alpha = 1, \ldots, K$, and $\boldsymbol{D}_b = \mathbf{0}$, respectively.

Next, we let $\Delta_s = 0$ in (B.030) and obtain

$$
\begin{aligned}
0 \;=\; & \sum_{\alpha}\int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b},\alpha,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\times\Bigg[\sum_{j=1}^{n_N}\frac{1}{A(\phi_{j0})}\big(Y_j\boldsymbol{X}_j - B'(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)\big)\mathbf{h}_1^{\beta} \\
& \quad +\sum_{j=1}^{n_N}\left\{-\left(\frac{Y_j(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_j\boldsymbol{b}_\alpha)-B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})^2}\right)A'(\phi_{j0})+C'(Y_j;\phi_{j0})\right\}\,\mathrm{h}_1^{\phi_j} \\
& +\frac{1}{2}(\boldsymbol{b}_\alpha-\boldsymbol{\mu}_{\alpha0})^T\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\boldsymbol{\Sigma}_{b0}^{-1}(\boldsymbol{b}_\alpha-\boldsymbol{\mu}_{\alpha0})-\frac{1}{2}\,\mathrm{Tr}\left(\boldsymbol{\Sigma}_{b0}^{-1}\boldsymbol{D}_b\right)+\left(\boldsymbol{b}_\alpha-\frac{1}{2}\boldsymbol{\mu}_{\alpha0}\right)^T\boldsymbol{\Sigma}_{b0}^{-1}\mathbf{h}_1^{\mu_\alpha}+\frac{1}{w_{\alpha0}}\mathbf{h}_1^{w_{\alpha0}} \\
& -\int_0^{V_s}\exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}_\alpha)+\boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\}\times\left\{(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{b}_\alpha^T)\mathbf{h}_1^{\psi}+\boldsymbol{Z}(t)\mathbf{h}_1^{\gamma}\right\}d\Lambda_{s0}(t)\Bigg]d\boldsymbol{b}_\alpha \\
& +\sum_{\alpha}\int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b},\alpha,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\times\left[-\int_0^{V_s}h_2(t)\exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}\circ\boldsymbol{b}_\alpha)+\boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\}d\Lambda_{s0}(t)\right]d\boldsymbol{b}_\alpha.
\end{aligned}
$$

Since $\mathbf{h}_1^{\beta}=0$, $\mathrm{h}_1^{\phi_j}$, $j=1,\ldots,n_N$, $\mathbf{h}_1^{\mu_\alpha}=0$, $\mathbf{h}_1^{w_{\alpha0}}=0$, $\alpha=1,\ldots,K$, and $\boldsymbol{D}_b=0$, the above expression can be written as

$$
\begin{aligned}
0 \;=\; & \mathrm{E}_{\alpha,\boldsymbol{b}}\Bigg[\exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\big(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha0}\big)-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\} \\
& \times\int_0^{V_s}\exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}_\alpha)+\boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\}\times\left[(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{b}_\alpha^T)\mathbf{h}_1^{\psi}+\boldsymbol{Z}(t)\mathbf{h}_1^{\gamma}+h_2(t)\right]d\Lambda_{s0}(t)\Bigg],
\end{aligned}
\tag{B.032}
$$

where $\boldsymbol{b}_\alpha$ follows a multivariate normal distribution with mean $\boldsymbol{\Sigma}_{b0}\big[\sum_{j=1}^{n_N}\big(Y_j\widetilde{\boldsymbol{Z}}_j/A(\phi_{j0})\big)\big]+\boldsymbol{\mu}_{\alpha0}$ and covariance $\boldsymbol{\Sigma}_{b0}$. Likewise, for any fixed $\widetilde{\boldsymbol{X}}$, treating $\boldsymbol{X}^T\boldsymbol{Y}$ as a parameter in this normal family, $\boldsymbol{b}_\alpha$ is the complete statistic for $\boldsymbol{X}^T\boldsymbol{Y}$, therefore,

$$
\begin{aligned}
& \exp\left\{\sum_{j=1}^{n_N}\frac{Y_j}{A(\phi_{j0})}\big(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha0}\big)-\sum_{j=1}^{n_N}\frac{B(\boldsymbol{\beta}_0;\boldsymbol{b}_\alpha)}{A(\phi_{j0})}\right\} \\
& \times\int_0^{V_s}\exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}_\alpha)+\boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\}\times\left[(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{b}_\alpha^T)\mathbf{h}_1^{\psi}+\boldsymbol{Z}(t)\mathbf{h}_1^{\gamma}+h_2(t)\right]d\Lambda_{s0}(t) \;=\; 0
\end{aligned}
$$

Since $\exp\left\{\sum_{j=1}^{n_N}\left[Y_j\big(\boldsymbol{X}_j\boldsymbol{\beta}_0+\widetilde{\boldsymbol{X}}_j\boldsymbol{\mu}_{\alpha0}\big)/A(\phi_{j0})\right]-\sum_{j=1}^{n_N}\left[B(\boldsymbol{\beta}_0;\boldsymbol{b})/A(\phi_{j0})\right]\right\}\neq 0$, equivalently

$$
\int_0^{V_s}\exp\left\{\widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi}_0\circ\boldsymbol{b}_\alpha)+\boldsymbol{Z}(t)\boldsymbol{\gamma}_0\right\}\times\left[(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{b}_\alpha^T)\mathbf{h}_1^{\psi}+\boldsymbol{Z}(t)\mathbf{h}_1^{\gamma}+h_2(t)\right]d\Lambda_{s0}(t)=0
$$

by assumption (A6). From assumption (A6), this immediately gives $\mathbf{h}_1^{\psi}=0$, $\mathbf{h}_1^{\gamma}=0$ and $h_2(t)=0$. Hence, the proof of condition (c) is completed.

Since the conditions (a)–(d) have been proved, Theorem 3.3.1 of [4] concludes that $\sqrt{m_s}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\widehat{\Lambda}_s-\Lambda_{s0})$ weakly converges to a tight random element in $\ell^\infty(\mathscr{H})$. Furthermore, we obtain

$$
\sqrt{m_s}\nabla U_{\xi_0}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\widehat{\Lambda}_s-\Lambda_{s0})[\mathbf{h}_1,h_2]=\sqrt{m_s}(\mathbf{P}_{m_s}-\mathbf{P})\left\{l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1+l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2]\right\}+o_P(1),
\tag{B.033}
$$

where $o_P(1)$ is a random variable which converges to zero in probability in $\ell^\infty(\mathscr{H})$. On the other hand, from (B.029), we have

$$
\sqrt{m_s}\nabla U_{\xi_0}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\widehat{\Lambda}_s-\Lambda_{s0})[\mathbf{h}_1,h_2]=\sqrt{m_s}\left\{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\Omega_1[\mathbf{h}_1,h_2]+\int_0^{\tau}\Omega_2[\mathbf{h}_1,h_2]d(\widehat{\Lambda}_s-\Lambda_{s0})(t)\right\}. \tag{B.034}
$$

By denoting $(\mathbf{h}_1^*,h_2^*)=\Omega^{-1}(\mathbf{h}_1,h_2)$, we have $(\mathbf{h}_1,h_2)=\Omega(\mathbf{h}_1^*,h_2^*)$, and replacing $(\mathbf{h}_1,h_2)$ with $(\mathbf{h}_1^*,h_2^*)$ in (B.033) and (B.034) leads to the followings, respectively.

$$
\sqrt{m_s}\nabla U_{\xi_0}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\widehat{\Lambda}_s-\Lambda_{s0})[\mathbf{h}_1^*,h_2^*]=\sqrt{m_s}(\mathbf{P}_{m_s}-\mathbf{P})\left\{l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1^*+l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2^*]\right\}+o_P(1),
$$

and

$$\sqrt{m_s}\nabla U_{\xi_0}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\widehat{\Lambda}_s-\Lambda_{s0})[\mathbf{h}_1^*,h_2^*] = \sqrt{m_s}\bigg\{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\Omega_1[\mathbf{h}_1^*,h_2^*] + \int_0^\tau \Omega_2[\mathbf{h}_1^*,h_2^*]d(\widehat{\Lambda}_s-\Lambda_{s0})(t)\bigg\}$$
$$= \sqrt{m_s}\bigg\{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\mathbf{h}_1 + \int_0^\tau h_2(t)d(\widehat{\Lambda}_s-\Lambda_{s0})(t)\bigg\}.$$

Thus, we obtain

$$\sqrt{m_s}\bigg\{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\mathbf{h}_1 + \int_0^\tau h_2(t)d(\widehat{\Lambda}_s-\Lambda_{s0})(t)\bigg\}$$
$$= \sqrt{m_s}(\mathbf{P}_{m_s}-\mathbf{P})\big\{l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1^* + l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2^*]\big\} + o_P(1). \quad \text{(B.035)}$$

Note that the first term on the right-hand side in (B.035) is $\sqrt{m_s}\{U_{m_s}(\boldsymbol{\theta}_0,\Lambda_{s0}) - U_s(\boldsymbol{\theta}_0,\Lambda_{s0})\}$, which is an empirical process in the space $\ell^\infty(\mathscr{H})$, and it is shown that $\mathscr{G}$ is P-Donsker in Section C.2. Therefore, $\sqrt{m_s}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\widehat{\Lambda}_s-\Lambda_{s0})$ weakly converges to a Gaussian process in $\ell^\infty(\mathscr{H})$.

In particular, if we choose $h_2 = 0$ in (B.035), then $\widehat{\boldsymbol{\theta}}^T\mathbf{h}_1$ is an asymptotic linear estimator for $\boldsymbol{\theta}_0^T\mathbf{h}_1$ with influence function being $l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1^* + l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2^*]$. Since this influence function is in the linear space spanned by the score functions for $\boldsymbol{\theta}_0$ and Λ_{s0}, Proposition 3.3.1 in Bickel *et al.* (1993) concludes that the influence function is the same as the efficient influence function for $\boldsymbol{\theta}_0^T\mathbf{h}_1$; that is $\widehat{\boldsymbol{\theta}}$ is an efficient estimator for $\boldsymbol{\theta}_0$. Therefore, Theorem 2 is proved.

Web Appendix C: Supplementary Proofs

The proofs for P-Donsker property of the classes $\mathscr{F}$ and $\mathscr{G}$ needed in Appendices B.1 and B.2 are presented in Appendices C.1 and C.2, respectively. In Appendix C.3, we prove *Fréchet* differentiability of $U(\xi)$ at ξ_0 and derive the derivative operator ∇U_{ξ_0} used in Appendix B.2.

C.1. Proof of P-Donsker Property of $\mathscr{F}$

We defined that a class $\mathscr{F} = \{Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s) : v \in [0,\tau], \boldsymbol{\theta} \in \Theta, \Lambda_s \in \mathscr{A}, s = 1,\ldots,S\}$, where $\mathscr{A} = \{\Lambda_s \in \mathbb{W}, \Lambda_s(0) = 0, \Lambda_s(\tau) \le B_{s0}, s = 1,\ldots,S\}$, B_{s0} is the constant given in the second step and $\mathbb{W}$ contains all nondecreasing functions in $[0,\tau]$. We can rewrite $Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)$ as

$$Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s) = Q_1(v,\boldsymbol{O};\boldsymbol{\theta})\frac{Q_2(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)}{Q_3(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)},$$

where

$$\begin{aligned}
Q_1(v,\boldsymbol{O};\boldsymbol{\theta}) &= \exp\bigg\{\boldsymbol{Z}(v)\boldsymbol{\gamma} + \bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\Sigma_b(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)^T + \frac{1}{2}R(v)\bigg\},\\
Q_2(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s) &= \sum_\alpha\int_{\boldsymbol{b}_\alpha}\exp\bigg\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha + \bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + (\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha\\
&\quad -\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)} - \int_0^{V_s}\exp\bigg\{(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha + \boldsymbol{Z}(t)\boldsymbol{\gamma}\\
&\quad +(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T + \boldsymbol{\mu}_\alpha\bigg]\\
&\quad +R(t)\bigg\}d\Lambda_s(t)\bigg\}w_\alpha d\boldsymbol{b}_\alpha,
\end{aligned}$$

$$
\begin{aligned}
Q_3(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s) \;=\; & \sum_{\alpha}\int_{\boldsymbol{b}_\alpha}\exp\bigg\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha+\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)} \\
& -\int_0^{V_s}\exp\bigg\{(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
& +(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg]\bigg\}d\Lambda_s(t)\bigg\}w_\alpha d\boldsymbol{b}_\alpha,
\end{aligned}
$$

$R(t) = (\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Sigma_b(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)^T$, $R(v)$ is $R(t)$ evaluated at $t=v$, $B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha) = B(\boldsymbol{\beta};g_1(\boldsymbol{b}_\alpha))$, $B_2(\boldsymbol{\beta};\boldsymbol{b}_\alpha) = B(\boldsymbol{\beta};g_2(\boldsymbol{b}_\alpha))$, $g_1(\boldsymbol{b}_\alpha) = \boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha+\boldsymbol{\Sigma}_b\big[\sum_{j=1}^{n_N}\big(Y_j\widetilde{\boldsymbol{X}}_j/A(\phi_j)\big)+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\big]^T+\boldsymbol{\mu}_\alpha$ and $g_2(\boldsymbol{b}_\alpha) = \boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha+\boldsymbol{\Sigma}_b\big[\sum_{j=1}^{n_N}\big(Y_j\widetilde{\boldsymbol{X}}_j/A(\phi_j)\big)+\Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\big]^T+\boldsymbol{\mu}_\alpha$.

Using assumption (A3), we can easily show that $Q_1(v,\boldsymbol{O};\boldsymbol{\theta})$ is continuously differentiable with respect to v and $\boldsymbol{\theta}$, and

$$
\|\nabla_{\boldsymbol{\theta}}Q_1(v,\boldsymbol{O};\boldsymbol{\theta})\|+\left|\frac{d}{dv}Q_1(v,\boldsymbol{O};\boldsymbol{\theta})\right| \;\le\; e^{k_1+k_2\|\boldsymbol{Y}\|}
$$

for some positive constants k_1 and k_2. Furthermore, it holds that

$$
\begin{aligned}
& \|\nabla_{\boldsymbol{\theta}}Q_2(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)\|+\left|\frac{d}{dv}Q_2(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)\right| \\
& \le \sum_{\alpha}\int_{\boldsymbol{b}_\alpha}\bigg[\exp\bigg\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha+\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\} \\
& \qquad \times e^{k_3\|\boldsymbol{b}_\alpha\|+k_4\|\boldsymbol{Y}\|+k_5(\alpha)}\times B_{s0}\times w_\alpha\bigg]d\boldsymbol{b}_\alpha \\
& \le e^{k_6+k_7\|\boldsymbol{Y}\|}
\end{aligned}
$$

and
$$
\|\nabla_{\boldsymbol{\theta}}Q_3(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)\|+\left|\frac{d}{dv}Q_3(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)\right| \;\le\; e^{k_8+k_9\|\boldsymbol{Y}\|}
$$

for some positive constants k_3,k_4,k_6,k_7,k_8, and k_9, and a deterministic function of α, $k(\alpha)$. Additionally, note that, for any $0<\Lambda<\infty$, $0<e^{-\Lambda}<1$ and $e^{-\Lambda}<\Lambda$ and thus $e^{-\Lambda_1}-e^{-\Lambda_2}<\Lambda_1-\Lambda_2$ for any Λ_1 and Λ_2 over $(0,\infty)$. Hence,

$$
\begin{aligned}
& |Q_2(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_{s1})-Q_2(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_{s2})| \\
& = \bigg|\sum_{\alpha}\int_{\boldsymbol{b}_\alpha}\exp\bigg\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha+\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)} \\
& \qquad \times\bigg[\exp\bigg\{-\int_0^v\exp\bigg\{(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha+\boldsymbol{Z}(t)\boldsymbol{\gamma}+(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T) \\
& \qquad\qquad \times\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg] \\
& \qquad +R(t)\bigg\}d\Lambda_{s1}(t)\bigg\}-\exp\bigg\{-\int_0^v\exp\bigg\{(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
& \qquad +(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg]+R(t)\bigg\}d\Lambda_{s2}(t)\bigg\}\bigg]w_\alpha d\boldsymbol{b}_\alpha\bigg| \\
& \le \bigg|\sum_{\alpha}\int_{\boldsymbol{b}_\alpha}\exp\bigg\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha+\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\} \\
& \times\bigg[\int_0^v\exp\bigg\{(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha+\boldsymbol{Z}(t)\boldsymbol{\gamma}+(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg]
\end{aligned}
$$

$$
\begin{aligned}
&\qquad +R(t)\Big\}d\big(\Lambda_{s1}-\Lambda_{s2}\big)(t)\Big]w_\alpha d\boldsymbol{b}_\alpha\Big|\\
&=\Big|\sum_\alpha\int_{\boldsymbol{b}_\alpha}\exp\Big\{\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\Big)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\Big\}\times(2\pi)^{d_b/2}\times(2\pi)^{-d_b/2}\\
&\quad\times\Big[\int_0^v\exp\Big\{-\frac{1}{2}\big[\boldsymbol{b}_\alpha-\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T\big]^T\big[\boldsymbol{b}_\alpha-\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T\big]\Big\}\\
&\quad\times\exp\Big\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma}\\
&\quad+(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big[\boldsymbol{\Sigma}_b\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big)^T+\boldsymbol{\mu}_\alpha\Big]+R(t)\Big\}d\big(\Lambda_{s1}-\Lambda_{s2}\big)(t)\Big]w_\alpha d\boldsymbol{b}_\alpha\Big|\\
&\le\Big|\sum_\alpha\Big[\exp\Big\{\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\Big)\boldsymbol{\mu}_\alpha\Big\}\times(2\pi)^{d_b/2}\\
&\quad\times\int_0^v\Big[\exp\Big\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma}\\
&\qquad\quad+(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big[\boldsymbol{\Sigma}_b\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big)^T+\boldsymbol{\mu}_\alpha\Big]+R(t)\Big\}\\
&\qquad\quad\times(2\pi)^{-d_b/2}\int_{\boldsymbol{b}_\alpha}\exp\Big\{-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\Big\}\\
&\times\exp\Big\{-\frac{1}{2}\big[\boldsymbol{b}_\alpha-\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T\big]^T\big[\boldsymbol{b}_\alpha-\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T\big]\Big\}d\boldsymbol{b}_\alpha\Big]d\big(\Lambda_{s1}-\Lambda_{s2}\big)(t)\Big]w_\alpha\Big|\\
&=\Big|\sum_\alpha\Big[\exp\Big\{\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\Big)\boldsymbol{\mu}_\alpha\Big\}\times(2\pi)^{d_b/2}\\
&\quad\times\int_0^v\Big[\exp\Big\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma}\\
&\qquad\qquad+(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big[\boldsymbol{\Sigma}_b\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big)^T+\boldsymbol{\mu}_\alpha\Big]+R(t)\Big\}\\
&\qquad\quad\times\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\Big[\exp\Big\{-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\Big\}\Big]\Big]d\big(\Lambda_{s1}-\Lambda_{s2}\big)(t)\Big]w_\alpha\Big|\\
&=\Big|\sum_\alpha\Big[\exp\Big\{\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\Big)\boldsymbol{\mu}_\alpha\Big\}\times(2\pi)^{d_b/2}\\
&\quad\times\Big[\big(\Lambda_{s1}(v)-\Lambda_{s2}(v)\big)\exp\Big\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma}\\
&\qquad\qquad+(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\Big[\boldsymbol{\Sigma}_b\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\Big)^T+\boldsymbol{\mu}_\alpha\Big]+R(v)\Big\}\\
&\qquad\quad\times\mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\Big[\exp\Big\{-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\Big\}\Big]\\
&\quad-\int_0^v\Big[\big(\Lambda_{s1}(t)-\Lambda_{s2}(t)\big)\frac{d}{dt}\Big[\exp\Big\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma}\\
&\qquad\qquad+(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big[\boldsymbol{\Sigma}_b\Big(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\Big)^T+\boldsymbol{\mu}_\alpha\Big]+R(t)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad \times \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\bigg[\exp\bigg\{-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\}\bigg]\bigg]\bigg]dt\bigg]\bigg]w_\alpha\bigg| \\
&\le \sum_\alpha w_\alpha\bigg[\exp\bigg\{\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha\bigg\}\times(2\pi)^{d_b/2} \\
&\quad \times\bigg[\big|\Lambda_{s1}(v)-\Lambda_{s2}(v)\big|\exp\bigg\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
&\qquad\qquad +(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg]+R(v)\bigg\} \\
&\qquad\qquad \times \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\bigg[\exp\bigg\{-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\}\bigg] \\
&\quad +\int_0^v\bigg[\big|\Lambda_{s1}(t)-\Lambda_{s2}(t)\big|\bigg|\frac{d}{dt}\bigg[\exp\bigg\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
&\qquad\qquad +(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg]+R(t)\bigg\} \\
&\qquad\qquad \times \mathrm{E}_{\boldsymbol{b}_\alpha|\alpha}\bigg[\exp\bigg\{-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\}\bigg]\bigg]\bigg|dt\bigg]\bigg] \\
&= \big|\Lambda_{s1}(v)-\Lambda_{s2}(v)\big|\exp\bigg\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
&\qquad\qquad +(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)^T+R(v)\bigg\}\times(2\pi)^{d_b/2} \\
&\qquad\qquad \times \mathrm{E}_{\alpha,\boldsymbol{b}}\bigg[\exp\bigg\{\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\}\bigg] \\
&\quad +\int_0^v\bigg[\big|\Lambda_{s1}(t)-\Lambda_{s2}(t)\big|\times(2\pi)^{d_b/2}\times\mathrm{E}_{\alpha,\boldsymbol{b}}\bigg[\exp\bigg\{\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha\bigg\} \\
&\qquad\quad \times\bigg|\frac{d}{dt}\bigg[\exp\bigg\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
&\quad +(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg[\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(t)\circ\boldsymbol{\psi}^T)\bigg)^T+\boldsymbol{\mu}_\alpha\bigg]+R(t)-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\}\bigg]\bigg|\bigg]\bigg]dt \\
&\le (2\pi)^{d_b/2}\exp\bigg\{\frac{1}{2}\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)\big((\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\big)^T+\boldsymbol{Z}(t)\boldsymbol{\gamma} \\
&\qquad\qquad +(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+\Delta(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)^T+R(v)\bigg\} \\
&\quad \times\mathrm{E}_{\alpha,\boldsymbol{b}}\bigg[\exp\bigg\{\bigg(\sum_{j=1}^{n_N}\frac{Y_j\widetilde{\boldsymbol{X}}_j}{A(\phi_j)}+(\Delta+1)(\widetilde{\boldsymbol{Z}}(v)\circ\boldsymbol{\psi}^T)\bigg)\boldsymbol{\mu}_\alpha-\sum_{j=1}^{n_N}\frac{B_1(\boldsymbol{\beta};\boldsymbol{b}_\alpha)}{A(\phi_j)}\bigg\}\bigg] \\
&\quad \times\bigg[\big|\Lambda_{s1}(v)-\Lambda_{s2}(v)\big|+\int_0^v\big|\Lambda_{s1}(t)-\Lambda_{s2}(t)\big|dt\bigg] \\
&\le e^{k_{10}+k_{11}\|\boldsymbol{Y}\|}\bigg[\big|\Lambda_{s1}(v)-\Lambda_{s2}(v)\big|+\int_0^\tau\big|\Lambda_{s1}(t)-\Lambda_{s2}(t)\big|dt\bigg],
\end{aligned}
$$

where k_{10} and k_{11} are positive constants. Similarly,

$$
|Q_3(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_{s1})-Q_3(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_{s2})|\le e^{k_{12}+k_{13}\|\boldsymbol{Y}\|}\bigg[\big|\Lambda_{s1}(v)-\Lambda_{s2}(v)\big|+\int_0^\tau\big|\Lambda_{s1}(t)-\Lambda_{s2}(t)\big|dt\bigg],
$$

where k_{12} and k_{13} are positive constants.

On the other hand, there exist positive constants $k_{14}, \ldots, k_{26}$ such that

$$
\begin{aligned}
&\left|Q_1(v, \boldsymbol{O}; \boldsymbol{\theta})\right| \\
&= \left|\exp\left\{\boldsymbol{Z}(v)\boldsymbol{\gamma} + \left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)\right)\Sigma_b(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)^T + \frac{1}{2}R(v)\right\}\right| \\
&\le e^{k_{14}+k_{15}\|\boldsymbol{Y}\|}, \\
&\left|Q_2(v, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)\right| \\
&= \left|\sum_\alpha \int_{\boldsymbol{b}_\alpha} \exp\left\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha + \left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + (\Delta+1)(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)\right)\boldsymbol{\mu}_\alpha - \sum_{j=1}^{n_N} \frac{B_1(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)}\right.\right. \\
&\qquad\qquad - \int_0^{V_s} \exp\left\{(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha + \boldsymbol{Z}(t)\boldsymbol{\gamma}\right. \\
&\qquad\qquad\qquad + (\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\left[\boldsymbol{\Sigma}_b\left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\right)^T + \boldsymbol{\mu}_\alpha\right] \\
&\qquad\qquad\qquad \left.\left.\left. + R(t)\right\}d\Lambda_s(t)\right\}w_\alpha d\boldsymbol{b}_\alpha\right| \\
&\le \left|\sum_\alpha \int_{\boldsymbol{b}_\alpha} \exp\left\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha + \left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + (\Delta+1)(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)\right)\boldsymbol{\mu}_\alpha - \sum_{j=1}^{n_N} \frac{B_1(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)}\right\}\right. \\
&\qquad \times\left[2\int_0^{v} \exp\left\{(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha + \boldsymbol{Z}(t)\boldsymbol{\gamma}\right.\right. \\
&\qquad\qquad\qquad + (\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\left[\boldsymbol{\Sigma}_b\left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\right)^T + \boldsymbol{\mu}_\alpha\right] \\
&\qquad\qquad\qquad \left.\left.\left. + R(t)\right\}d\Lambda_s(t)\right]w_\alpha d\boldsymbol{b}_\alpha\right| \\
&\le \left|\sum_\alpha \int_{\boldsymbol{b}_\alpha} \exp\left\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha + \left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + (\Delta+1)(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)\right)\boldsymbol{\mu}_\alpha - \sum_{j=1}^{n_N} \frac{B_1(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)}\right\}\right. \\
&\qquad \left. \times\ 2\ \times \exp\left\{k_{16}\|\boldsymbol{b}_\alpha)\| + k_{17}\|\boldsymbol{Y}\| + k_{18}\|\boldsymbol{\mu}_\alpha\| + k_{19}\right\} \times B_{s0} \times w_\alpha\ d\boldsymbol{b}_\alpha\right| \\
&\le e^{k_{19}+k_{20}\|\boldsymbol{Y}\|},
\end{aligned}
$$

and

$$
\begin{aligned}
&Q_3(v, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) \\
&= \sum_\alpha \int_{\boldsymbol{b}_\alpha} \exp\left\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha + \left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)\right)\boldsymbol{\mu}_\alpha - \sum_{j=1}^{n_N} \frac{B_1(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)}\right. \\
&\qquad - \int_0^{V_s} \exp\left\{(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\boldsymbol{\Sigma}_b^{1/2}\boldsymbol{b}_\alpha + \boldsymbol{Z}(t)\boldsymbol{\gamma}\right. \\
&\qquad\qquad \left.\left. + (\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\left[\boldsymbol{\Sigma}_b\left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{\psi}^T)\right)^T + \boldsymbol{\mu}_\alpha\right]\right\}d\Lambda_s(t)\right\}w_\alpha d\boldsymbol{b}_\alpha \\
&\ge \sum_\alpha \int_{\boldsymbol{b}_\alpha} \exp\left\{-\frac{1}{2}\boldsymbol{b}_\alpha^T\boldsymbol{b}_\alpha + \left(\sum_{j=1}^{n_N} \frac{Y_j \widetilde{\boldsymbol{X}}_j}{A(\phi_j)} + \Delta(\widetilde{\boldsymbol{Z}}(v) \circ \boldsymbol{\psi}^T)\right)\boldsymbol{\mu}_\alpha - \sum_{j=1}^{n_N} \frac{B_1(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)}\right. \\
&\qquad\qquad \left. - \exp\left\{k_{22}\|\boldsymbol{b}_\alpha\| + k_{23}\|\boldsymbol{Y}\| + k_{24}\|\boldsymbol{\mu}_\alpha\| + k_{25}\right\} \times B_{s0}\right\}w_\alpha d\boldsymbol{b}_\alpha \\
&\ge k_{26} > 0.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\|\nabla_{\boldsymbol{\theta}} Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)\| + \left|\frac{d}{dv}Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s)\right| \\
&= \left\|(\nabla_{\boldsymbol{\theta}} Q_1)\frac{Q_2}{Q_3} + Q_1\left(\nabla_{\boldsymbol{\theta}}\frac{Q_2}{Q_3}\right)\right\| + \left|\left(\frac{d}{dv}Q_1\right)\frac{Q_2}{Q_3} + Q_1\left(\frac{d}{dv}\left(\frac{Q_2}{Q_3}\right)\right)\right| \\
&= \left\|(\nabla_{\boldsymbol{\theta}} Q_1)\frac{Q_2}{Q_3} + Q_1\left[(\nabla_{\boldsymbol{\theta}} Q_2)\frac{1}{Q_3} + Q_2\frac{(-1)}{Q_3^2}(\nabla_{\boldsymbol{\theta}} Q_3)\right]\right\| \\
&\qquad + \left|\left(\frac{d}{dv}Q_1\right)\frac{Q_2}{Q_3} + Q_1\left[\left(\frac{d}{dv}Q_2\right)\frac{1}{Q_3} + Q_2\frac{(-1)}{Q_3^2}\left(\frac{d}{dv}Q_3\right)\right]\right| \\
&= \left\|(\nabla_{\boldsymbol{\theta}} Q_1)\frac{Q_2}{Q_3} + (\nabla_{\boldsymbol{\theta}} Q_2)\frac{Q_1}{Q_3} - (\nabla_{\boldsymbol{\theta}} Q_3)\frac{Q_1Q_2}{Q_3^2}\right\| \\
&\qquad + \left|\left(\frac{d}{dv}Q_1\right)\frac{Q_2}{Q_3} + \left(\frac{d}{dv}Q_2\right)\frac{Q_1}{Q_3} - \left(\frac{d}{dv}Q_3\right)\frac{Q_1Q_2}{Q_3^2}\right| \\
&\le \left(\|\nabla_{\boldsymbol{\theta}} Q_1\| + \left|\frac{d}{dv}Q_1\right|\right)\left|\frac{Q_2}{Q_3}\right| + \left(\|\nabla_{\boldsymbol{\theta}} Q_2\| + \left|\frac{d}{dv}Q_2\right|\right)\left|\frac{Q_1}{Q_3}\right| + \left(\|\nabla_{\boldsymbol{\theta}} Q_3\| + \left|\frac{d}{dv}Q_3\right|\right)\left|\frac{Q_1Q_2}{Q_3^2}\right| \\
&\le e^{k_{27}+k_{28}\|\boldsymbol{Y}\|},
\end{aligned}
$$

for some positive constants k_{27} and k_{28}. Therefore, by the mean-value theorem, we conclude that, for any $(v_1,\boldsymbol{\theta}_1,\Lambda_{s1})$ and $(v_2,\boldsymbol{\theta}_2,\Lambda_{s2})$ in $[0,\tau]\times\Theta\times\mathscr{A}$,

$$
\begin{aligned}
&\left|Q(v_1,\boldsymbol{O};\boldsymbol{\theta}_1,\Lambda_{s1}) - Q(v_2,\boldsymbol{O};\boldsymbol{\theta}_2,\Lambda_{s2})\right| \\
&\le e^{k_{27}+k_{28}\|\boldsymbol{Y}\|}\left[\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\| + \left|\Lambda_{s1}(V)-\Lambda_{s2}(V)\right| + \int_0^\tau \left|\Lambda_{s1}(t)-\Lambda_{s2}(t)\right|dt + |v_1-v_2|\right]
\end{aligned}
\tag{B.036}
$$

holds for some positive constants k_{27} and k_{28} and $0\le V\le\tau$ ($V=v_1$ or v_2).

Applying Theorem 2.7.5 (p159) in [4] to our situation, the entropy number for the class $\mathscr{A}$ satisfies $\log N_{[\cdot]}(\varepsilon,\mathscr{A},L_2(P))\le K/\varepsilon$, where K is a constant. Thus, we can find $\exp\{K/\varepsilon\}$ brackets, $\{[L_j,U_j]\}$, to cover the class $\mathscr{A}$ such that $\|U_j-L_j\|_{L_2(P)}\le\varepsilon$ for each pair of $[L_j,U_j]$. On the other hand, we can further find a partition of $[0,\tau]\times\Theta$, say $I_1\bigcup I_2\bigcup\cdots$, such that the number of partitions is of the order $(1/\varepsilon)^{d_\theta+1}$, and, for any $(v_1,\boldsymbol{\theta}_1)$ and $(v_2,\boldsymbol{\theta}_2)$ in the same partition, their Euclidean distance is less than ε. Therefore, the partition $\{I_1,I_2,\ldots\}\times\{[L_j,U_j]\}$ bracket covers $[0,\tau]\times\Theta\times\mathscr{A}$, and the total number of the partitions is of order $(1/\varepsilon)^{d_\theta+1}\exp\{1/\varepsilon\}$. Hence, from (B.036), for any I_l and $[L_j,U_j]$, the set of the functions $\{Q(v,\boldsymbol{O};\boldsymbol{\theta},\Lambda_s):(v,\boldsymbol{\theta})\in I_l,\Lambda_s\in\mathscr{A},\Lambda_s\in[L_j,U_j]\}$ can be bracket covered by

$$
\begin{aligned}
&\Bigg[Q(v_l,\boldsymbol{O};\boldsymbol{\theta}_l,\Lambda_{sl}) - e^{k_{27}+k_{28}\|\boldsymbol{Y}\|}\left\{\varepsilon + \left|U_j(V)-L_j(V)\right| + \int_0^\tau \left|U_j(t)-L_j(t)\right|dt\right\}, \\
&\qquad Q(v_l,\boldsymbol{O};\boldsymbol{\theta}_l,\Lambda_{sl}) + e^{k_{27}+k_{28}\|\boldsymbol{Y}\|}\left\{\varepsilon + \left|U_j(V)-L_j(V)\right| + \int_0^\tau \left|U_j(t)-L_j(t)\right|dt\right\}\Bigg],
\end{aligned}
\tag{B.037}
$$

where $(v_l,\boldsymbol{\theta}_l)$ is a fixed point in I_l and Λ_{sj} is a fixed function in $[L_j,U_j]$. Note that the $L_2(P)$ distance between these two functions in the above bracket (B.037) is less than $O(\varepsilon)$. Therefore, we have

$$
N_{[\cdot]}(\varepsilon,\mathscr{F},\|\cdot\|_{L_2(P)}\le O\left(\left(\frac{1}{\varepsilon}\right)^{d_\theta+1}e^{1/\varepsilon}\right).
$$

Furthermore, $\mathscr{F}$ has an $L_2(P)$-integrable covering function, which is equal to $O(e^{k_{27}+k_{28}\|\boldsymbol{Y}\|})$. From Theorem 2.5.6 (p130) in [4], $\mathscr{F}$ is P-Donsker.

Additionally, in the above derivation, we also note that all the functions in $\mathscr{F}$ are bounded from below by $e^{-k_{29}-k_{30}\|\boldsymbol{Y}\|}$ for some positive constants k_{29} and k_{30}. ♯

C.2. Proof of P-Donsker Property of $\mathscr{G}$

Recall that we defined the class

$$\mathscr{G} = \Big\{ l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}_0, \Lambda_{s0})[h_2], \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \sup_{t \in [0,\tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \le \delta,\ (\mathbf{h}_1, h_2) \in \mathscr{H} \Big\},$$

where $(\mathbf{h}_1^\beta, \mathbf{h}_1^\phi, \mathbf{h}_1^{\Sigma_b}, \mathbf{h}_1^\mu, \mathbf{h}_1^w, \mathbf{h}_1^\psi, \mathbf{h}_1^\gamma)$ denote the corresponding components of $\mathbf{h}_1$ for the parameters $(\boldsymbol{\beta}, \boldsymbol{\phi}, \mathrm{Vec}(\boldsymbol{\Sigma}_b),\ \boldsymbol{\mu}, \boldsymbol{w}, \boldsymbol{\psi}, \boldsymbol{\gamma})$, respectively. We can write that for $(\mathbf{h}_1, h_2) \in \mathscr{H}$,

$$\begin{aligned} & l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] \\ &= \left[\rho_1(\boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 - \int_0^{V_s} \rho_2(t, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 d\Lambda_s(t) \right] + \Delta h_2(V_s) \\ & \qquad - \int_0^{V_s} \rho_3(t, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) h_2(t) d\Lambda_s(t), \end{aligned}$$

where

$$\begin{aligned} & \rho_1(\boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 \\ &= \left\{ \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) d\boldsymbol{b}_\alpha \right\}^{-1} \times \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) \\ & \quad \times \Bigg[\sum_{j=1}^{n_N} \frac{1}{A(\phi_j)} \big(Y_j \boldsymbol{X}_j - B'(\boldsymbol{\beta}; \boldsymbol{b}_\alpha) \big) \mathbf{h}_1^\beta \\ & \qquad + \sum_{j=1}^{n_N} \left\{ - \left(\frac{Y_j(\boldsymbol{X}_j \boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_j \boldsymbol{b}_\alpha) - B(\boldsymbol{\beta}; \boldsymbol{b}_\alpha)}{A(\phi_j)^2} \right) A'(\phi_j) + C'(Y_j; \phi_j) \right\} \mathrm{h}_1^{\phi_j} \\ & \qquad + \frac{1}{2} (\boldsymbol{b}_\alpha - \boldsymbol{\mu}_\alpha)^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{D}_b \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{b}_\alpha - \boldsymbol{\mu}_\alpha) - \frac{1}{2} \mathrm{Tr}\left(\boldsymbol{\Sigma}_b^{-1} \boldsymbol{D}_b \right) \\ & \qquad + \left(\boldsymbol{b}_\alpha - \frac{1}{2} \boldsymbol{\mu}_\alpha \right)^T \boldsymbol{\Sigma}_b^{-1} \mathbf{h}_1^{\mu_\alpha} + \frac{1}{w_\alpha} \mathbf{h}_1^{w_\alpha} + \Delta_s \big\{ \big(\widetilde{\boldsymbol{Z}}(V_s) \circ \boldsymbol{b}_\alpha^T \big) \mathbf{h}_1^\psi + \boldsymbol{Z}(V_s) \mathbf{h}_1^\gamma \big\} \Bigg] d\boldsymbol{b}_\alpha, \\ & \rho_2(t, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 \\ &= \left\{ \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) d\boldsymbol{b}_\alpha \right\}^{-1} \\ & \quad \times \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) \times \exp\left\{ \widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(t) \boldsymbol{\gamma} \right\} \times \left[\big(\widetilde{\boldsymbol{Z}}(t) \circ \boldsymbol{b}_\alpha^T \big) \mathbf{h}_1^\psi + \boldsymbol{Z}(t) \mathbf{h}_1^\gamma \right] d\boldsymbol{b}_\alpha, \\ & \rho_3(t, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) \\ &= \left\{ \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) d\boldsymbol{b}_\alpha \right\}^{-1} \times \sum_\alpha \int_{\boldsymbol{b}_\alpha} G(\boldsymbol{b}, \alpha, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) \times \exp\left\{ \widetilde{\boldsymbol{Z}}(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_\alpha) + \boldsymbol{Z}(t) \boldsymbol{\gamma} \right\} d\boldsymbol{b}_\alpha, \end{aligned}$$

$B'(\boldsymbol{\beta}; \boldsymbol{b})$ is the derivative of $B(\boldsymbol{\beta}; \boldsymbol{b})$ with respect to $\boldsymbol{\beta}$, $A'(\phi_j)$ and $C'(Y_j; \phi_j)$ are the derivatives of $A(\phi_j)$ and $C(Y_j; \phi_j)$ with respect to ϕ_j respectively, and $\boldsymbol{D}_b$ is the symmetric matrix such that $\mathrm{Vec}(\boldsymbol{D}_b) = \mathbf{h}_1^b$.

For $l = 1, 2, 3$, we denote $\nabla_{\boldsymbol{\theta}} \rho_l$ and $\nabla_{\Lambda_s} \rho_l[\delta \Lambda_s]$ as the derivatives of ρ_l with respect to $\boldsymbol{\theta}$ and Λ_s along the path $\Lambda_s + \varepsilon \delta \Lambda_s$. Then, using the similar arguments done in Section C.1, it is verified that $\nabla_{\Lambda_s} \rho_l[\delta \Lambda_s] = \int_0^t \rho_{l+3}(u, \boldsymbol{O}; \boldsymbol{\theta}, \Lambda_s) d\delta \Lambda_s(u)$ and there exist two positive constants q_1 and q_2 such that

$$\sum_l \left\{ |\rho_l| + |\nabla_{\boldsymbol{\theta}} \rho_l| \right\} \le e^{q_1 + q_2 \|\boldsymbol{Y}\|}$$

By the mean value theorem, we have that, for any $(\boldsymbol{\theta}, \Lambda_s, \mathbf{h}_1, h_2)$ and $(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s, \widetilde{\mathbf{h}}_1, \widetilde{h}_2)$ in $\Xi \times \mathscr{H}$,

$$
\begin{aligned}
& l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] - l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T \widetilde{\mathbf{h}}_1 - l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)[\widetilde{h}_2] \\
&= l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] - l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T \mathbf{h}_1 - l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)[h_2] \\
&+ l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)[h_2] - l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T \widetilde{\mathbf{h}}_1 - l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)[\widetilde{h}_2] \\
&= \left[l_\theta(\boldsymbol{\theta}, \Lambda_s)^T - l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T\right]\mathbf{h}_1 + \left[l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s) - l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)\right][h_2] \\
&+ l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T(\mathbf{h}_1 - \widetilde{\mathbf{h}}_1) + l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)([h_2] - [\widetilde{h}_2]) \\
&= (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})^T \left[\frac{d}{d\boldsymbol{\theta}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \Lambda_s)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \Lambda_s=\Lambda_s^*}\right]\mathbf{h}_1 + \left[\frac{d}{d\Lambda_s} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \Lambda_s)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \Lambda_s=\Lambda_s^*}\right]^T [\Lambda_s - \widetilde{\Lambda}_s]\mathbf{h}_1 \\
&+ (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})^T \left[\frac{d}{d\boldsymbol{\theta}} l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \Lambda_s=\Lambda_s^*}\right][h_2] + \left[\frac{d}{d\Lambda_s} l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \Lambda_s=\Lambda_s^*}\right]^T [\Lambda_s - \widetilde{\Lambda}_s][h_2] \\
&+ l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T(\mathbf{h}_1 - \widetilde{\mathbf{h}}_1) + l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)([h_2] - [\widetilde{h}_2]) \\
&= (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})^T \nabla_{\boldsymbol{\theta}} \rho_1(\boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)\mathbf{h}_1 - (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})^T \int_0^{V_S} \nabla_{\boldsymbol{\theta}} \rho_2(t, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)^T d\Lambda_s^*(t)\mathbf{h}_1 \\
&+ \int_0^{V_s} \rho_4(t, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)^T \mathbf{h}_1 d(\Lambda_s - \widetilde{\Lambda}_s)(t) + \int_0^{V_s}\int_0^t \rho_5(u, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)^T d(\Lambda_s - \widetilde{\Lambda}_s)(u)\mathbf{h}_1 d\Lambda_s^*(t) \\
&- \int_0^{V_s} \rho_2(t, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)^T (\Lambda_s - \widetilde{\Lambda}_s)\mathbf{h}_1 dt - (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})^T \int_0^{V_S} \nabla_{\boldsymbol{\theta}} \rho_3(t, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) h_2(t) d\Lambda_s^*(t) \\
&+ \int_0^{V_s}\int_0^t \rho_6(u, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*) d(\Lambda_s - \widetilde{\Lambda}_s)(u) h_2(t) d\Lambda_s^*(t) - \int_0^{V_s} \rho_3(t, \boldsymbol{O}; \boldsymbol{\theta}^*, \Lambda_s^*)^T (\Lambda_s - \widetilde{\Lambda}_s)(t) h_2(t) dt \\
&+ \rho_1(\boldsymbol{O}; \widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T(\mathbf{h}_1 - \widetilde{\mathbf{h}}_1) - \int_0^{V_S} \rho_2(t, \boldsymbol{O}; \widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T(\mathbf{h}_1 - \widetilde{\mathbf{h}}_1) d\widetilde{\Lambda}_s(t) \\
&+ \Delta_s(h_2(V_s) - \widetilde{h}_2(V_s)) - \int_0^{V_S} \rho_3(t, \boldsymbol{O}; \widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)(h_2(V_s) - \widetilde{h}_2(V_s)) d\widetilde{\Lambda}_s(t),
\end{aligned} \tag{B.038}
$$

where $(\boldsymbol{\theta}^*, \Lambda_s^*)$ is equal to $\varepsilon^*(\boldsymbol{\theta}, \Lambda_s) + (1 - \varepsilon^*)(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)$ for some $\varepsilon^* \in [0, 1]$. Thus, we have that

$$
\begin{aligned}
& \left|l_\theta(\boldsymbol{\theta}, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta}, \Lambda_s)[h_2] - l_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)^T \widetilde{\mathbf{h}}_1 - l_{\Lambda_s}(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}_s)[\widetilde{h}_2]\right| \\
& \le e^{q_1 + q_2\|\boldsymbol{Y}\|}\bigg\{\|\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}\| + \|\mathbf{h}_1 - \widetilde{\mathbf{h}}_1\| + |\Lambda_s(V_s) - \widetilde{\Lambda}_s(V_s)| \\
& \qquad + \int_0^\tau |\Lambda_s(t) - \widetilde{\Lambda}_s(t)|[dt + d|h_2(t)| + d|\widetilde{h}_2(t)|] \\
& \qquad + |h_2(V_s) - \widetilde{h}_2(V_s)| + \int_0^\tau |h_2(V_s) - \widetilde{h}_2(V_s)|[d\Lambda_s(t) - d\widetilde{\Lambda}_s(t)]\bigg\},
\end{aligned} \tag{B.039}
$$

where $d|h_2(t)| = dh_2^+(t) + dh_2^-(t)$ and $d|\widetilde{h}_2(t)| = d\widetilde{h}_2^+(t) + d\widetilde{h}_2^-(t)$. As done in Section C.1, by applying Theorem 2.7.5 (p159) in [4], we note that for a set $\mathscr{H}_2 = \{h_2 : \|h_2\|_V \le B_1\}$, $\log N_{[\cdot]}\big(\varepsilon, \mathscr{H}_2, L_2(P)\big) \le K/\varepsilon$ for a constant B_1 and any probability measure P where K is a constant. Thus, we can find $\exp\{K/\varepsilon\}$ brackets, $\{[L_j, U_j]\}$, to cover the class $\mathscr{H}_2$ such that $\|U_j - L_j\|_{L_2(P)} \le \varepsilon$ for each pair of $[L_j, U_j]$. On the other hand, we can further find a partition of $\mathscr{H}_1 = \{\mathbf{h}_1 : \|\mathbf{h}_1\| \le 1\}$, say $I_1 \cup I_2 \cup \cdots$, such that the number of partitions is of the order $(1/\varepsilon)$, and, for any $\mathbf{h}_1$ and h_2 in the same partition, their Euclidean distance is less than ε. Therefore, the partition $\{I_1, I_2, \ldots\} \times \{[L_j, U_j]\}$ bracket covers $\mathscr{H}_1 \times \mathscr{H}_2$, and the total number of the partitions is of order $(1/\varepsilon)\exp\{1/\varepsilon\}$. Then, we obtain

$$
\log N_{[\cdot]}\big(\varepsilon, \mathscr{G}, L_2(P)\big) \le O\left(\frac{1}{\varepsilon} + \log \varepsilon\right).
$$

Moreover, $\mathscr{G}$ has an $L_2(P)$-integrable covering function, which is equal to $O(e^{q_1 + q_2\|\boldsymbol{Y}\|})$. Hence, from Theorem 2.5.6 (p130) in [4], $\mathscr{G}$ is P-Donsker.

Additionally, from (B.039), we can calculate that

$$
\begin{aligned}
&\left|l_\theta(\boldsymbol{\theta},\Lambda_s)^T\mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta},\Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1 - l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2]\right| \\
&\quad\le\ e^{q_1+q_2\|\boldsymbol{Y}\|}\left\{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + |\Lambda_s(V_s)-\Lambda_{s0}(V_s)| + \int_0^\tau |\Lambda_s(t)-\Lambda_{s0}(t)|dt\right\} \\
&\qquad + \left|\int_0^\tau \rho_3(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)h_2(t)d(\Lambda_s(t)-\Lambda_{s0}(t))\right|. \qquad \text{(B.040)}
\end{aligned}
$$

If $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\to 0$ and $\sup_{t\in[0,\tau]}|\Lambda_s(t)-\Lambda_{s0}(t)|\to 0$, the above expression converges to zero uniformly. Therefore,

$$
\sup_{(\mathbf{h}_1,h_2)\in\mathscr{H}} \mathbf{P}\left[l_\theta(\boldsymbol{\theta},\Lambda_s)^T\mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta},\Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1 - l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2]\right]^2 \longrightarrow 0. \quad \sharp
$$

C.3. Derivative Operator ∇U_{ξ_0}

From (B.038) in the previous Section C.2, we can obtain that

$$
\begin{aligned}
& l_\theta(\boldsymbol{\theta},\Lambda_s)^T\mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta},\Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1 - l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2] \\
&= \left[l_\theta(\boldsymbol{\theta},\Lambda_s)^T - l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\right]\mathbf{h}_1 + \left[l_{\Lambda_s}(\boldsymbol{\theta},\Lambda_s) - l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})\right][h_2] \\
&= (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\nabla_{\boldsymbol{\theta}}\rho_1(\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)\mathbf{h}_1 - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\int_0^{V_S}\nabla_{\boldsymbol{\theta}}\rho_2(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T\mathbf{h}_1 d\Lambda_s^*(t) \\
&\quad + \int_0^{V_s}\rho_4(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T\mathbf{h}_1 d(\Lambda_s-\Lambda_{s0})(t) \\
&\quad + \int_0^{V_s}\int_0^t \rho_5(u,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T d(\Lambda_s-\Lambda_{s0})(u)\mathbf{h}_1 d\Lambda_s^*(t) \\
&\quad - \int_0^{V_s}\rho_2(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T\mathbf{h}_1 d(\Lambda_s-\Lambda_{s0})(t) \\
&\quad - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\int_0^{V_S}\nabla_{\boldsymbol{\theta}}\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)h_2(t)d\Lambda_s^*(t) \\
&\quad + \int_0^{V_s}\int_0^t \rho_6(u,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T d(\Lambda_s-\Lambda_{s0})(u)h_2(t)d\Lambda_s^*(t) \\
&\quad - \int_0^{V_s}\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T h_2(t)d(\Lambda_s-\Lambda_{s0})(t) \\
&= (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\left\{\nabla_{\boldsymbol{\theta}}\rho_1(\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*) - \int_0^{V_S}\nabla_{\boldsymbol{\theta}}\rho_2(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)^T d\Lambda_s^*(t)\right\}\mathbf{h}_1 \\
&\quad + \mathbf{h}_1^T\left\{\int_0^\tau I(t\le V_s)\Big[\rho_4(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*) - \rho_2(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)\right. \\
&\qquad\qquad \left. + \rho_5(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)\int_t^{V_s}d\Lambda_s^*(u)\Big]d(\Lambda_s-\Lambda_{s0})(t)\right\} \\
&\quad - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\int_0^\tau I(t\le V_s)\nabla_{\boldsymbol{\theta}}\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)h_2(t)d\Lambda_s^*(t) \\
&\quad - \int_0^\tau\left\{-I(t\le V_s)\rho_6(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)\int_t^{V_s}h_2(u)d\Lambda_s^*(u)\right. \\
&\qquad\qquad \left. + I(t\le V_s)\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*)h_2(t)\right\}d(\Lambda_s-\Lambda_{s0})(t). \qquad \text{(B.041)}
\end{aligned}
$$

Then, we have that

$$
\nabla U_{\xi_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0,\Lambda_s-\Lambda_{s0})[\mathbf{h}_1,h_2]
$$

$$
\begin{aligned}
&= (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \mathbf{P}\left\{\nabla_{\boldsymbol{\theta}}\rho_1(\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) - \int_0^{V_S} \nabla_{\boldsymbol{\theta}}\rho_2(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})d\Lambda_{s0}(t)\right\}\mathbf{h}_1 \\
&\quad + \mathbf{h}_1^T\bigg\{\int_0^\tau \mathbf{P}\bigg[I(t\le V_s)\bigg(\rho_4(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) - \rho_2(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) \\
&\qquad\qquad + \rho_5(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\int_t^{V_s} d\Lambda_{s0}(u)\bigg)\bigg]d(\Lambda_s-\Lambda_{s0})(t)\bigg\} \\
&\quad - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\int_0^\tau \mathbf{P}\left\{I(t\le V_s)\nabla_{\boldsymbol{\theta}}\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}h_2(t)d\Lambda_{s0}(t) \\
&\quad - \int_0^\tau \mathbf{P}\bigg\{-I(t\le V_s)\rho_6(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\int_t^{V_s} h_2(u)d\Lambda_{s0}(u) \\
&\qquad\qquad + I(t\le V_s)\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})h_2(t)\bigg\}d(\Lambda_s-\Lambda_{s0})(t).
\end{aligned}
$$

By the similar algebra done in (B.040), we can verify that, for $j = 1,\ldots,6$,

$$
\sup_{t\in[0,\tau]} \|\rho_j(t,\boldsymbol{O};\boldsymbol{\theta}^*,\Lambda_s^*) - \rho_j(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\| \le e^{q_3+q_4\|\boldsymbol{Y}\|}\left\{\|\boldsymbol{\theta}^*-\boldsymbol{\theta}_0\| + \sup_{t\in[0,\tau]}|\Lambda_s^*-\Lambda_{s0}|\right\},
$$

which implies that the linear operator ∇U_{ξ_0} is bounded.
Then, we obtain

$$
\begin{aligned}
&\mathbf{P}\left[l_\theta(\boldsymbol{\theta},\Lambda_s)^T\mathbf{h}_1 + l_{\Lambda_s}(\boldsymbol{\theta},\Lambda_s)[h_2] - l_\theta(\boldsymbol{\theta}_0,\Lambda_{s0})^T\mathbf{h}_1 - l_{\Lambda_s}(\boldsymbol{\theta}_0,\Lambda_{s0})[h_2]\right] \\
&= \nabla U_{\xi_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0,\Lambda_s-\Lambda_{s0})[\mathbf{h}_1,h_2] + o\Big(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \sup_{t\in[0,\tau]}|\Lambda_s-\Lambda_{s0}|\Big)(\|\mathbf{h}_1\| + \|h_2\|_V).
\end{aligned}
$$

Therefore, U_ξ is *Fréchet* differentiable at ξ_0.

Additionally, from (B.041) and the above expression, we have

$$
\nabla U_{\xi_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0,\Lambda_s-\Lambda_{s0})[\mathbf{h}_1,h_2] = (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\Omega_1[\mathbf{h}_1,h_2] + \int_0^\tau \Omega_2[\mathbf{h}_1,h_2]d(\Lambda_s-\Lambda_{s0})(t),
$$

where

$$
\begin{aligned}
\Omega_1[\mathbf{h}_1,h_2] &= \mathbf{P}\left\{\nabla_{\boldsymbol{\theta}}\rho_1(\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) - \int_0^{V_S}\nabla_{\boldsymbol{\theta}}\rho_2(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})d\Lambda_{s0}(t)\right\}\mathbf{h}_1 \\
&\quad - \int_0^\tau \mathbf{P}\left\{I(t\le V_s)\nabla_{\boldsymbol{\theta}}\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}h_2(t)d\Lambda_{s0}(t)
\end{aligned}
$$

and

$$
\begin{aligned}
&\Omega_2[\mathbf{h}_1,h_2] \\
&= \mathbf{h}_1^T\mathbf{P}\left\{I(t\le V_s)\Big[\rho_4(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) - \rho_2(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0}) + \rho_5(u,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\int_t^{V_s} d\Lambda_{s0}(u)\Big]\right\} \\
&\quad + \mathbf{P}\left\{I(t\le V_s)\rho_6(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\int_t^{V_s} h_2(u)d\Lambda_{s0}(u)\right\} \\
&\quad - \mathbf{P}\left\{I(t\le V_s)\rho_3(t,\boldsymbol{O};\boldsymbol{\theta}_0,\Lambda_{s0})\right\}h_2(t).
\end{aligned}
$$

Thus, $\Omega = (\Omega_1,\Omega_2)$ is the bounded linear operator from $R^d \times BV[0,\tau]$ to itself. Furthermore, we note that $\Omega = \boldsymbol{H} + (\boldsymbol{M}_1,\boldsymbol{M}_2)$, where

$$
\begin{aligned}
\boldsymbol{H}(\mathbf{h}_1, h_2) &= \big(\mathbf{h}_1, -\mathbf{P}\left\{I(t \le V_s)\rho_3(t, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0})\right\}h_2(t)\big), \\
\boldsymbol{M}_1(\mathbf{h}_1, h_2) &= \Omega_1[\mathbf{h}_1, h_2] - \mathbf{h}_1, \\
\boldsymbol{M}_2(\mathbf{h}_1, h_2) &= \mathbf{h}_1^T \mathbf{P}\Bigg\{I(t \le V_s)\big[\nabla_{\boldsymbol{\theta}}\rho_4(t, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) - \rho_2(t, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) \\
&\qquad + \rho_5(t, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) \int_t^{V_s} d\Lambda_{s0}(u)\big]\Bigg\} \\
&\quad + \mathbf{P}\left\{I(t \le V_s)\rho_6(t, \boldsymbol{O}; \boldsymbol{\theta}_0, \Lambda_{s0}) \int_t^{V_s} h_2(u) d\Lambda_{s0}(u)\right\},
\end{aligned}
$$

and also note that $\boldsymbol{H}$ is obviously invertible. Since $\boldsymbol{M}_1$ maps into a finite-dimensional space, it is compact. The image of $\boldsymbol{M}_2$ is a continuously differentiable function in $[0, \tau]$. By the Arzela-Ascoli theorem (p41) in [4], $\boldsymbol{M}_2$ is a compact operator from $R^d \times BV[0, \tau]$ to $BV[0, \tau]$. Thus, we conclude that Ω is the summation of an invertible operator $\boldsymbol{H}$ and a compact operator $\boldsymbol{M} = (\boldsymbol{M}_1, \boldsymbol{M}_2)$. ♯

References

[1] Bickel PJ, Klaassen CAJ, Ritov Y, Wellner JA (1993) Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore

[2] Parner E (1998) Asymptotic Theory for the Correlated Gamma-frailty Model. Ann Stat 26:183–214

[3] van der Vaart AW (1998) Asymptotic Statistics. Cambridge University Press

[4] van der Vaart AW, Wellner JA (1996) Weak Convergence and Empirical Processes. New York: Springer–Verlag

Web Appendix D: Figures and Tables

Figures and tables referenced in Sections 4.3 and 5 are provided in Appendices D.1 and D.2, respectively.

D.1. Figures and tables referenced in Section 4.3

The results, density plots and relative bias plots obtained from additional simulations for the sensitivity to model-misspecification under a new true distribution, the mixture of non-central t_{20}(-2) and *Gamma*(7, 1/8), are presented in Table 1 and Figures 1 and 2 in this Appendix D.1 which are respectively corresponding to Table 3, Figures 1 and 2 in Section 4.3 for the original true distribution, the mixture of non-central t_{10}(-1) and *Gamma*(7, 1/8).

D.2. Tables referenced in Section 5

In the CHANCE data analysis, we additionally applied the rest four unselected distributions for random effects – one normal distribution without mixture and the mixtures of 2, 4 and 5 normal distributions – to the final simultaneous models derived under the mixture of 3 normal distributions and compared their results (provided in Tables 2–5, respectively, of this Appendix D.2) to those in Table 4 in Section 5. Most of the covariates in the final models yielded same conclusions under different distributions assumed for random effects except one covariate, "the number of 12 oz. beers consumed per week" whose inference is somewhat changed in comparison to the reference category of '30 or more': No beer consumption group, associated with higher odds of being satisfied under all mixtures, is not significant under one normal distribution although its p-value is at borderline around the significance level of 0.05; While the category of '15 to 29' is not associated with QoL satisfaction with a borderline p-value under the mixture of 3 normal distributions, it is likely to have higher odds of being satisfied under the mixture of 2 normal distributions; The category of '1 to 4', not associated with the risk of death under the mixture of 3 normal distributions, appears to significantly lower the risk of death under one normal distribution and has a borderline p-value under the mixture of 2 normal distributions. On the other hand, overall, the estimates for same variables are similar under all mixtures but slightly different from those under one normal distribution.

Table 6 is referenced in Paragraph 7 of Section 5 for additional simulations under the settings similar to the CHANCE data with the high censoring rate = 85% and the low average number of longitudinal observations per patient (n_i) = 1.93.

Table 1: Summary of simulation results of sensitivity for model-misspecification under the true distribution of random effects with a mixture of non-central t_{20}(-2) and *Gamma*(7, 1/8).

		1 Normal distribution				Mixture of 2 Normal distributions			
Par.	TRUE	Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
Longitudinal model									
β_1	1.0	.909	.116	.122	.899	.969	.074	.073	.920
β_2	- .5	- .654	.152	.211	.962	- .588	.128	.126	.885
β_3	- .2	- .201	.025	.026	.959	- .199	.024	.025	.955
σ_y^2	.5	.500	.010	.010	.946	.501	.010	.010	.951
Hazards model									
ψ	- .1	- .099	.024	.025	.959	- .098	.024	.025	.959
γ_1	- .1	- .091	.082	.084	.952	- .097	.083	.084	.952
γ_2	.1	.120	.150	.146	.936	.112	.149	.145	.937
$\Lambda(\ .9)$	.9	.903	.088	.089	.951	.904	.088	.089	.953
$\Lambda(1.4)$	1.4	1.403	.138	.140	.954	1.405	.138	.140	.954
$\Lambda(1.9)$	1.9	1.902	.209	.204	.946	1.903	.209	.204	.946

		Mixture of 3 Normal distributions				Mixture of 4 Normal distributions			
Par.	TRUE	Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
Longitudinal model									
β_1	1.0	.981	.059	.057	.934	.982	.058	.057	.936
β_2	- .5	- .563	.104	.099	.885	- .560	.104	.099	.886
β_3	- .2	- .199	.024	.025	.953	- .200	.024	.025	.958
σ_y^2	.5	.500	.010	.010	.947	.500	.010	.010	.945
Hazards model									
ψ	- .1	- .098	.024	.025	.959	- .099	.024	.025	.960
γ_1	- .1	- .098	.083	.083	.951	- .098	.083	.083	.952
γ_2	.1	.110	.149	.145	.936	.107	.150	.145	.937
$\Lambda(\ .9)$	.9	.904	.088	.089	.952	.904	.088	.089	.955
$\Lambda(1.4)$	1.4	1.405	.138	.140	.953	1.406	.139	.140	.950
$\Lambda(1.9)$	1.9	1.903	.209	.204	.946	1.906	.210	.204	.949

		Mixture of 5 Normal distributions			
Par.	TRUE	Est.	SSD	ESE	CP
Longitudinal model					
β_1	1.0	.987	.052	.050	.931
β_2	- .5	- .541	.091	.087	.912
β_3	- .2	- .200	.024	.025	.956
σ_y^2	.5	.500	.010	.010	.946
Hazards model					
ψ	- .1	- .100	.025	.025	.956
γ_1	- .1	- .098	.080	.083	.964
γ_2	.1	.107	.144	.145	.951
$\Lambda(\ .9)$	.9	.900	.086	.089	.960
$\Lambda(1.4)$	1.4	1.405	.136	.140	.962
$\Lambda(1.9)$	1.9	1.905	.203	.204	.957

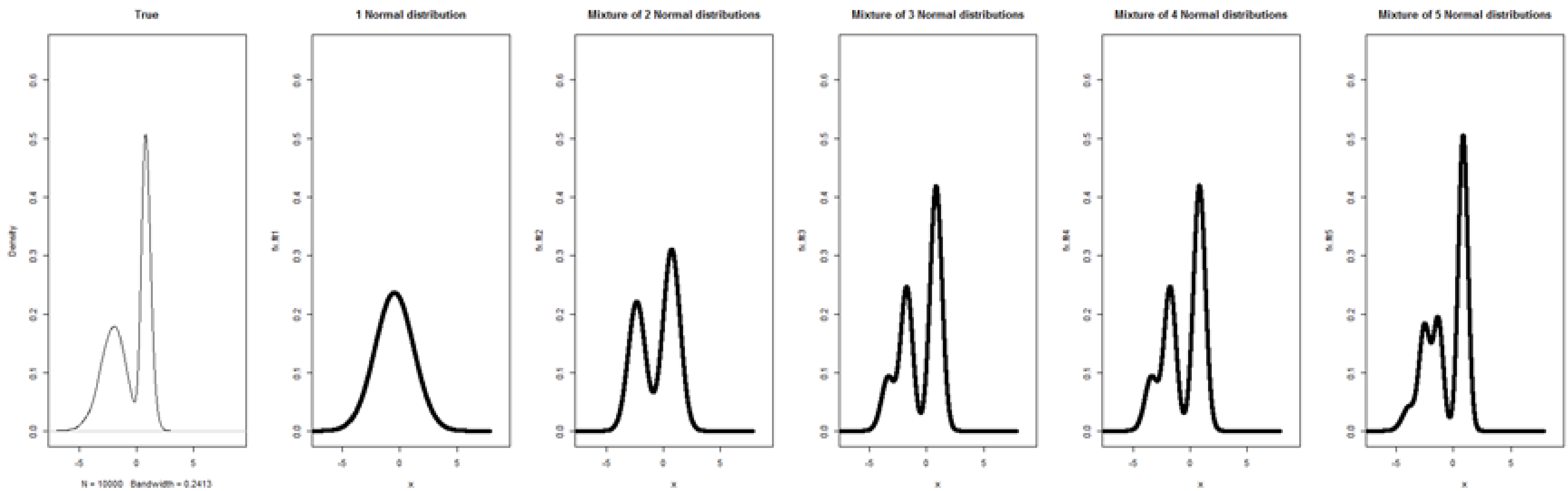

Figure 1: Density plots from simulation results of sensitivity for model-misspecification with true distribution as a mixture of non-central t_{20}(-2) and $Gamma(7, 1/8)$.

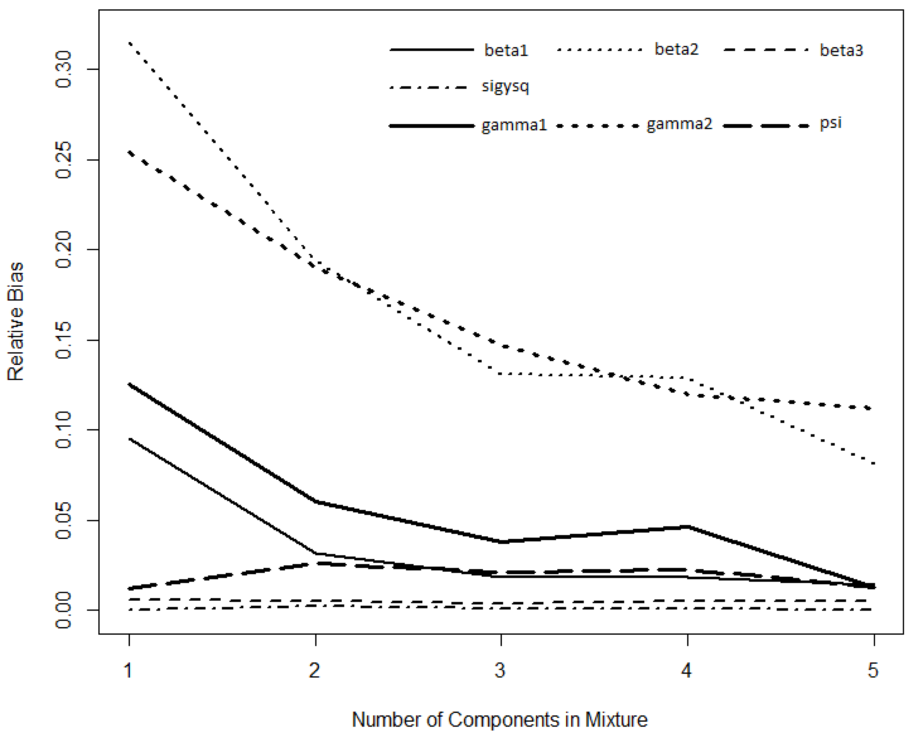

Figure 2: Relative bias plot of parameters in longitudinal and hazard models (thin and thick lines respectively) from simulation results of sensitivity for model-misspecification with true distribution as a mixture of non-central t_{20}(-2) and $Gamma$(7, 1/8).

Table 2: Results from the final model of simultaneous analysis for the Quality of Life and survival time of the CHANCE study, assuming 1 normal distribution for random effects

Parameter		Est.	ESE	P-value
< HNCS QoL longitudinal model >				
Race (ref= White)				
– African American	β_1	.948	.402	.018
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	β_2	.898	.472	.057
– less than 1	β_3	1.000	.598	.095
– 1 to 4	β_4	1.946	.523	<.001
– 5 to 14	β_5	1.644	.454	<.001
– 15 to 29	β_6	.903	.508	.075
Household income (ref= level1: 0–10K)				
– level2: 20–30K	β_7	- .263	.403	.515
– level3: 40–50K	β_8	.592	.446	.184
– level4: ≥ 60K	β_9	1.920	.487	<.001
Radiation therapy (ref= No)				
– Yes	β_{10}	-1.454	.423	.001
Tumor stage (ref= I)				
– II	β_{11}	- .590	.499	.238
– III	β_{12}	-1.762	.495	<.001
– IV	β_{13}	-1.679	.443	<.001
# of persons supported by household income	β_{14}	- .292	.146	.046
BMI	β_{15}	.054	.026	.037
Time at survey measurement (years)	β_{16}	.305	.092	.001
< Hazards model >				
Random effect coefficient	ψ	- .225	.094	.017
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	γ_1	- .708	.349	.043
– less than 1	γ_2	- .124	.397	.755
– 1 to 4	γ_3	- .766	.386	.048
– 5 to 14	γ_4	-1.030	.350	.003
– 15 to 29	γ_5	- .550	.371	.139
Household income (ref= level1: 0–10K)				
– level2: 20–30K	γ_6	- .227	.275	.409
– level3: 40–50K	γ_7	- .875	.342	.010
– level4: ≥ 60K	γ_8	-1.417	.374	<.001
Tumor stage (ref= I)				
– II	γ_9	- .245	.444	.581
– III	γ_{10}	.155	.402	.700
– IV	γ_{11}	.958	.307	.002
Total # of medical conditions reported	γ_{12}	.208	.095	.028

P-value for testing σ_b^2 being zero is based on a mixture of 0 and χ^2 distribution with 1 degree of freedom with equal mixing probabilities.

Table 3: Results from the final model of simultaneous analysis for the Quality of Life and survival time of the CHANCE study, assuming a mixture of 2 normal distributions for random effects

Parameter		Est.	ESE	P-value
< HNCS QoL longitudinal model >				
Race (ref= White)				
– African American	β_1	.910	.398	.022
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	β_2	.865	.441	.050
– less than 1	β_3	1.078	.600	.072
– 1 to 4	β_4	1.695	.549	.002
– 5 to 14	β_5	1.506	.451	.001
– 15 to 29	β_6	.964	.490	.049
Household income (ref= level1: 0–10K)				
– level2: 20–30K	β_7	- .343	.373	.359
– level3: 40–50K	β_8	.644	.449	.151
– level4: ≥ 60K	β_9	2.002	.500	<.001
Radiation therapy (ref= No)				
– Yes	β_{10}	-1.671	.534	.002
Tumor stage (ref= I)				
– II	β_{11}	- .701	.563	.213
– III	β_{12}	-2.035	.539	<.001
– IV	β_{13}	-1.862	.499	<.001
# of persons supported by household income	β_{14}	- .386	.145	.008
BMI	β_{15}	.054	.026	.039
Time at survey measurement (years)	β_{16}	.339	.094	<.001
< Hazards model >				
Random effect coefficient	ψ	- .207	.079	.008
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	γ_1	- .703	.348	.043
– less than 1	γ_2	- .148	.394	.707
– 1 to 4	γ_3	- .727	.385	.059
– 5 to 14	γ_4	- .998	.349	.004
– 15 to 29	γ_5	- .572	.370	.122
Household income (ref= level1: 0–10K)				
– level2: 20–30K	γ_6	- .206	.275	.455
– level3: 40–50K	γ_7	- .884	.342	.010
– level4: ≥ 60K	γ_8	-1.408	.373	<.001
Tumor stage (ref= I)				
– II	γ_9	- .250	.443	.574
– III	γ_{10}	.175	.402	.663
– IV	γ_{11}	.960	.307	.002
Total # of medical conditions reported	γ_{12}	.207	.095	.029

P-value for testing σ_b^2 being zero is based on a mixture of 0 and χ^2 distribution with 1 degree of freedom with equal mixing probabilities.

Table 4: Results from the final model of simultaneous analysis for the Quality of Life and survival time of the CHANCE study, assuming a mixture of 4 normal distributions for random effects

Parameter		Est.	ESE	P-value
< HNCS QoL longitudinal model >				
Race (ref= White)				
– African American	β_1	.891	.400	.026
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	β_2	.851	.430	.048
– less than 1	β_3	1.128	.614	.066
– 1 to 4	β_4	1.579	.577	.006
– 5 to 14	β_5	1.458	.431	.001
– 15 to 29	β_6	1.013	.535	.058
Household income (ref= level1: 0–10K)				
– level2: 20–30K	β_7	- .344	.358	.337
– level3: 40–50K	β_8	.630	.444	.156
– level4: ≥ 60K	β_9	1.960	.516	<.001
Radiation therapy (ref= No)				
– Yes	β_{10}	-1.716	.640	.007
Tumor stage (ref= I)				
– II	β_{11}	- .713	.557	.201
– III	β_{12}	-2.030	.548	<.001
– IV	β_{13}	-1.845	.522	<.001
# of persons supported by household income	β_{14}	- .396	.143	.006
BMI	β_{15}	.059	.027	.028
Time at survey measurement (years)	β_{16}	.353	.093	<.001
< Hazards model >				
Random effect coefficient	ψ	- .204	.078	.009
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	γ_1	- .703	.347	.043
– less than 1	γ_2	- .156	.393	.691
– 1 to 4	γ_3	- .711	.385	.065
– 5 to 14	γ_4	- .991	.348	.004
– 15 to 29	γ_5	- .579	.370	.117
Household income (ref= level1: 0–10K)				
– level2: 20–30K	γ_6	- .204	.274	.456
– level3: 40–50K	γ_7	- .881	.341	.010
– level4: ≥ 60K	γ_8	-1.398	.374	<.001
Tumor stage (ref= I)				
– II	γ_9	- .252	.443	.570
– III	γ_{10}	.168	.403	.677
– IV	γ_{11}	.951	.306	.002
Total # of medical conditions reported	γ_{12}	.207	.095	.030

P-value for testing σ_b^2 being zero is based on a mixture of 0 and χ^2 distribution with 1 degree of freedom with equal mixing probabilities.

Table 5: Results from the final model of simultaneous analysis for the Quality of Life and survival time of the CHANCE study, assuming a mixture of 5 normal distributions for random effects

Parameter		Est.	ESE	P-value
< HNCS QoL longitudinal model >				
Race (ref= White)				
– African American	β_1	.929	.379	.014
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	β_2	.895	.441	.043
– less than 1	β_3	1.116	.615	.069
– 1 to 4	β_4	1.654	.507	.001
– 5 to 14	β_5	1.489	.441	.001
– 15 to 29	β_6	.968	.530	.067
Household income (ref= level1: 0–10K)				
– level2: 20–30K	β_7	- .357	.349	.307
– level3: 40–50K	β_8	.643	.449	.152
– level4: ≥ 60K	β_9	2.004	.473	<.001
Radiation therapy (ref= No)				
– Yes	β_{10}	-1.651	.594	.005
Tumor stage (ref= I)				
– II	β_{11}	- .733	.515	.155
– III	β_{12}	-2.066	.550	<.001
– IV	β_{13}	-1.886	.528	<.001
# of persons supported by household income	β_{14}	- .390	.145	.007
BMI	β_{15}	.062	.027	.022
Time at survey measurement (years)	β_{16}	.356	.093	<.001
< Hazards model >				
Random effect coefficient	ψ	- .204	.077	.008
# of 12 oz. beers consumed per week (ref=30 or more)				
– None	γ_1	- .710	.348	.041
– less than 1	γ_2	- .153	.394	.697
– 1 to 4	γ_3	- .719	.384	.061
– 5 to 14	γ_4	- .996	.348	.004
– 15 to 29	γ_5	- .572	.370	.122
Household income (ref= level1: 0–10K)				
– level2: 20–30K	γ_6	- .202	.274	.460
– level3: 40–50K	γ_7	- .884	.341	.010
– level4: ≥ 60K	γ_8	-1.406	.373	<.001
Tumor stage (ref= I)				
– II	γ_9	- .249	.442	.573
– III	γ_{10}	.174	.403	.666
– IV	γ_{11}	.956	.307	.002
Total # of medical conditions reported	γ_{12}	.207	.095	.029

P-value for testing σ_b^2 being zero is based on a mixture of 0 and χ^2 distribution with 1 degree of freedom with equal mixing probabilities.

Table 6: Simulation results under the settings similar to the CHANCE data with 85% censoring and the average number of longitudinal observations per patient (n_i)=1.93 – mixture of 2 distributions, sample size of 400 and variance of random effect (σ_b^2)=0.5 considered.

Par.	True	Est.	SSD	ESD	CP
β_1	1.0	1.082	.516	.630	.972
β_2	- .5	- .525	.731	.847	.977
β_3	- .2	- .215	.254	.267	.964
μ_1	-3.0	-3.150	.490	.954	.995
μ_2	3.0	3.087	.427	.764	.995
w_1	.4	.400	.030	.050	.994
σ_b^2	.5	.577	.202	1.421	.998
ψ	- .1	- .099	.044	.049	.970
γ_1	- .1	- .103	.251	.242	.947
γ_2	.1	.091	.415	.419	.960
$\Lambda($ .104)	.104	.105	.031	.031	.954
$\Lambda($ .209)	.209	.210	.059	.059	.955
$\Lambda($ .313)	.313	.314	.087	.088	.953