December 2022

Supporting Information for "Latent Class Proportional Hazards Regression with Heterogeneous Survival Data"

Teng Fei[1], John Hanfelt[2] and Limin Peng[2,*]

[1]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center,

633 3rd Avenue, New York, New York, 10017, U.S.A.

[2]Department of Biostatistics and Bioinformatics, Emory University,

1518 Clifton Road Northeast, Atlanta, Georgia 30322, U.S.A

[*]*email:* lpeng@emory.edu

Appendix A. Proof of asymptotic theories

Let $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\Lambda}$ be the maximum likelihood estimator corresponding to the observed data log-likelihood. Now define $N(t) = I(\tilde{T} \leqslant t, \Delta = 1)$, $\tilde{N}(t) = I(\tilde{T} \leqslant t, \Delta = 0)$ and let $\mathbb{P}_n$, P denote the empirical measure and probability measure, respectively. Then the log-likelihood ℓ satisfies $n^{-1}\ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda; \boldsymbol{O}) \equiv \ell_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda)$, where

$$\begin{aligned}
\ell_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda) &= n^{-1} \log L(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda; \boldsymbol{O}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \sum_{l=1}^{L} p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) \{\lambda(\tilde{T}_i) \exp(\boldsymbol{z}_{il}^T \boldsymbol{\gamma})\}^{\Delta_i} \exp\{-\Lambda(\tilde{T}_i) \exp(\boldsymbol{z}_{il}^T \boldsymbol{\gamma})\} \right\} \\
&= \frac{1}{n} \sum_{i=1}^{n} I(\Delta_i = 1) \log \left\{ \sum_{l=1}^{L} p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) \lambda(\tilde{T}_i) \exp(\boldsymbol{z}_{il}^T \boldsymbol{\gamma}) \exp\{-\Lambda(\tilde{T}_i) \exp(\boldsymbol{z}_{il}^T \boldsymbol{\gamma})\} \right\} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} I(\Delta_i = 0) \log \left\{ \sum_{l=1}^{L} p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) \exp\{-\Lambda(\tilde{T}_i) \exp(\boldsymbol{z}_{il}^T \boldsymbol{\gamma})\} \right\} \\
&= \mathbb{P}_n \int_0^{t^*} \left[\log \left\{ \sum_{l=1}^{L} p_l(\boldsymbol{x}; \boldsymbol{\alpha}) e^{\boldsymbol{z}_l^T \boldsymbol{\gamma}} \exp\left(- \int_0^t e^{\boldsymbol{z}_l^T \boldsymbol{\gamma}} d\Lambda(s) \right) \right\} + \log \Lambda\{t\} \right] dN(t) \\
&\quad + \mathbb{P}_n \int_0^{t^*} \log \left\{ \sum_{l=1}^{L} p_l(\boldsymbol{x}; \boldsymbol{\alpha}) \exp\left(- \int_0^t e^{\boldsymbol{z}_l^T \boldsymbol{\gamma}} d\Lambda(s) \right) \right\} d\tilde{N}(t).
\end{aligned}$$

Let $\mathcal{W}$ denote the space of functions on $[0, t^*]$ that are uniformly bounded by 1 and with total variation bounded by 1. Define $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^{(p+1)\times(L-1)} : ||\boldsymbol{u}|| \leqslant 1\}$ and $\mathcal{V} = \{\boldsymbol{v} \in \mathbb{R}^{q \times L + L - 1} : ||\boldsymbol{v}|| \leqslant 1\}$. Let $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{v} \in \mathcal{V}$, and $h \in \mathcal{W}$. Then $(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda)$ can be identified as elements in the space of bounded functions on $\mathcal{U} \times \mathcal{V} \times \mathcal{W}$, $\ell^\infty(\mathcal{U} \times \mathcal{V} \times \mathcal{W})$, by $\boldsymbol{u}^T \boldsymbol{\alpha} + \boldsymbol{v}^T \boldsymbol{\gamma} + \int_0^{t^*} h d\Lambda$. Similarly, $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0, \hat{\Lambda} - \Lambda_0)$ can also be identified in $\ell^\infty(\mathcal{U} \times \mathcal{V} \times \mathcal{W})$ by

$$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0, \hat{\Lambda} - \Lambda_0)[\boldsymbol{u}, \boldsymbol{v}, h] = \sqrt{n}\{\boldsymbol{u}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \boldsymbol{v}^T(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \int_0^{t^*} h d(\hat{\Lambda} - \Lambda_0)\}.$$

Appendix A.1 *Proof of Theorem 1*

Proof. Step 1. We show by contradiction that $\hat{\Lambda}(t^*) < \infty$. Condition (C1) indicates that for large n, there exists an observation with probability one such that $\tilde{T} = t^*$ and $\Delta = 0$. If $\hat{\Lambda}(t^*) = \infty$, then

$$\mathbb{P}_n \int_0^{t^*} \log \left\{ \sum_{l=1}^{L} p_l(\boldsymbol{x}; \boldsymbol{\alpha}) \exp\left(- \int_0^t e^{\boldsymbol{z}_l^T \boldsymbol{\gamma}} d\Lambda(s) \right) \right\} d\tilde{N}(t) = -\infty$$

and thus $\ell_n(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)=-\infty$. Therefore, it must satisfy $\hat{\Lambda}(t^*)<\infty$ to maximize ℓ_n.

Step 2. We show that $\limsup_n \hat{\Lambda}(t^*)<\infty$ by contradiction. By conditions (C2) and (C3) there exists a constant M such that $|\boldsymbol{z}_l^T\boldsymbol{\gamma}|\leqslant M$ for any $\boldsymbol{\gamma}$ and $\boldsymbol{z}_l$, and $p_l(\boldsymbol{x};\boldsymbol{\alpha})\in(0,1)$ for any $\boldsymbol{x}$ and $\boldsymbol{\alpha}$. Define $\bar{\Lambda}(t)=[\hat{\Lambda}(t)\wedge\tilde{M}]\vee\tilde{M}/2$, where $\tilde{M}=e^{-M}\{\log(\epsilon_0)\}^{-1}$ for a chosen $\epsilon_0\in(0,1)$.

By definition of MLE $\ell_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})\geqslant\ell_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\bar{\Lambda})$. Assuming that $\limsup_n\hat{\Lambda}(t^*)=\infty$, then by the following inequality

$$\log(\sum_{l=1}^{L}a_l)\leqslant\sum_{l=1}^{L}\log a_l+\log L,$$

we have

$$\begin{aligned}
\ell_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})\leqslant&\mathbb{P}_n\sum_{l=1}^{L}\int_0^{t*}\left\{\log p_l(\boldsymbol{x};\hat{\boldsymbol{\alpha}})-\int_0^t e^{\boldsymbol{z}_l^T\hat{\boldsymbol{\gamma}}}d\hat{\Lambda}(s)\right\}d\{N(t)+\tilde{N}(t)\}\\
&+\mathbb{P}_n\sum_{l=1}^{L}\boldsymbol{z}_l^T\hat{\boldsymbol{\gamma}}dN(t^*)+\mathbb{P}_n\sum_{l=1}^{L}\int_0^{t*}\log\hat{\Lambda}\{t\}dN(t)+\log L\mathbb{P}_n\{N(t^*)+\tilde{N}(t^*)\}\\
\leqslant&-\mathbb{P}_n\sum_{l=1}^{L}\int_0^{t*}\int_0^t e^{\boldsymbol{z}_l^T\hat{\boldsymbol{\gamma}}}d\hat{\Lambda}(s)d\{N(t)+\tilde{N}(t)\}\\
&+LM\mathbb{P}_n dN(t^*)+\mathbb{P}_n\sum_{l=1}^{L}\int_0^{t*}\log\hat{\Lambda}\{t\}dN(t)+\log L\mathbb{P}_n\{N(t^*)+\tilde{N}(t^*)\}\to-\infty.
\end{aligned}$$

On the other hand, by the following inequality

$$\log(\sum_{l=1}^{L}a_l)\geqslant\frac{1}{L}\sum_{l=1}^{L}\log a_l$$

we have

$$
\begin{aligned}
\ell_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \bar{\Lambda}) \geqslant & \frac{1}{L}\Bigg[\mathbb{P}_n \sum_{l=1}^{L} \int_0^{t*} \left\{ \log p_l(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}) - \int_0^t e^{\boldsymbol{z}_l^T \hat{\boldsymbol{\gamma}}} d\bar{\Lambda}(s) \right\} d\{N(t) + \tilde{N}(t)\} \\
& + \mathbb{P}_n \sum_{l=1}^{L} \boldsymbol{z}_l^T \hat{\boldsymbol{\gamma}} N(t^*) + \mathbb{P}_n \sum_{l=1}^{L} \int_0^{t*} \log \bar{\Lambda}\{t\} dN(t)\Bigg] \\
\geqslant & \frac{1}{L}\Bigg[\mathbb{P}_n \sum_{l=1}^{L} \left\{ \log p_l(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}) - \tilde{M} e^{M} \right\} \{N(t^*) + \tilde{N}(t^*)\} \\
& + \mathbb{P}_n \sum_{l=1}^{L} \boldsymbol{z}_l^T \hat{\boldsymbol{\gamma}} N(t^*) + \sum_{l=1}^{L} \log \frac{\tilde{M}}{2} \mathbb{P}_n N(t^*)\Bigg] \\
= & \frac{1}{L}\Bigg[\mathbb{P}_n \sum_{l=1}^{L} \left\{ \log p_l(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}) + \{\log(\epsilon_0)\}^{-1} \right\} \{N(t^*) + \tilde{N}(t^*)\} \\
& + \mathbb{P}_n \sum_{l=1}^{L} \boldsymbol{z}_l^T \hat{\boldsymbol{\gamma}} N(t^*) + \sum_{l=1}^{L} \{-M - \log(\log(\epsilon_0)) - \log 2\} \mathbb{P}_n N(t^*)\Bigg] > -\infty.
\end{aligned}
$$

The above contradiction shows that $\limsup_n \hat{\Lambda}(t^*) < \infty$. By Helly's selection theorem there exists a converging subsequence such that $\hat{\boldsymbol{\alpha}} \to \boldsymbol{\alpha}^*$, $\hat{\boldsymbol{\gamma}} \to \boldsymbol{\gamma}^*$ and $\hat{\Lambda} \to \Lambda^*$.

Step 3. We show that the limit of the subsequence mentioned in the end of step 2 are $\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0$ and Λ_0. Define $\Lambda^\epsilon(t) = \int_0^t \{1 + \epsilon h(s)\} d\Lambda(s)$, where $h(t) \in \mathcal{W}$, the space of functions on $[0, t^*]$ that are uniformly bounded by 1 and with total variation bounded by 1. Then we obtain the derivative of log-likelihood $\ell_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda^\epsilon)$ with respect to ϵ at 0, denoted by $\dot{\ell}_{n,\Lambda}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda)[h]$:

$$
\begin{aligned}
\dot{\ell}_{n,\Lambda}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda)[h] = & \mathbb{P}_n \int_0^{t^*} \left[\sum_{l=1}^{L} \tau_l(t; \boldsymbol{O}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda) \left\{ - \int_0^t h(s) e^{\boldsymbol{z}_l^T \boldsymbol{\gamma}} d\Lambda(s) \right\} + h(t) \right] dN(t) \\
& + \mathbb{P}_n \int_0^{t^*} \left[\sum_{l=1}^{L} \tilde{\tau}_l(t; \boldsymbol{O}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda) \left\{ - \int_0^t h(s) e^{\boldsymbol{z}_l^T \boldsymbol{\gamma}} d\Lambda(s) \right\} \right] d\tilde{N}(t),
\end{aligned}
\tag{A.1}
$$

where

$$
\begin{aligned}
\tau_{il}(t; \boldsymbol{O}_i, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda) = & \frac{p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) f_l(\tilde{T}_i = t, \Delta_i = 1 | \boldsymbol{x}_i; \boldsymbol{\gamma}, \Lambda)}{\sum_{d=1}^{L} p_d(\boldsymbol{x}_i; \boldsymbol{\alpha}) f_d(\tilde{T}_i = t, \Delta_i = 1 | \boldsymbol{x}_i; \boldsymbol{\gamma}, \Lambda)} \\
= & \frac{p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) \exp(\boldsymbol{z}_{il}^T \boldsymbol{\gamma}) \exp\{- \int_0^t e^{\boldsymbol{z}_{il}^T \boldsymbol{\gamma}} d\Lambda(s)\}}{\sum_{d=1}^{L} p_d(\boldsymbol{x}_i; \boldsymbol{\alpha}) \exp(\boldsymbol{z}_{id}^T \boldsymbol{\gamma}) \exp\{- \int_0^t e^{\boldsymbol{z}_{il}^T \boldsymbol{\gamma}} d\Lambda(s)\}}
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\tau}_{il}(t; \boldsymbol{O}_i, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda) = & \frac{p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) f_l(\tilde{T}_i = t, \Delta_i = 0 | \boldsymbol{x}_i; \boldsymbol{\gamma}, \Lambda)}{\sum_{d=1}^{L} p_d(\boldsymbol{x}_i; \boldsymbol{\alpha}) f_d(\tilde{T}_i = t, \Delta_i = 0 | \boldsymbol{x}_i; \boldsymbol{\gamma}, \Lambda)} \\
= & \frac{p_l(\boldsymbol{x}_i; \boldsymbol{\alpha}) \exp\{- \int_0^t e^{\boldsymbol{z}_{il}^T \boldsymbol{\gamma}} d\Lambda(s)\}}{\sum_{d=1}^{L} p_d(\boldsymbol{x}_i; \boldsymbol{\alpha}) \exp\{- \int_0^t e^{\boldsymbol{z}_{il}^T \boldsymbol{\gamma}} d\Lambda(s)\}}.
\end{aligned}
$$

By changing the order of integration we have

$$\dot{\ell}_{n,\Lambda}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)[h] = \mathbb{P}_n \int_0^{t^*} h(s)dN(s) - \mathbb{P}_n \sum_{l=1}^{L} \int_0^{t^*} h(s) e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}} \int_s^{t^*} \tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)dN(t)d\Lambda(s)$$

$$- \mathbb{P}_n \sum_{l=1}^{L} \int_0^{t^*} h(s) e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}} \int_s^{t^*} \tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)d\tilde{N}(t)d\Lambda(s).$$

By definition of the NPMLE, $\dot{\ell}_{n,\Lambda}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})[h] = 0$ for all $h \in \mathcal{W}$. By taking $h(\cdot) = I(\cdot \leqslant t)$, we have

$$\hat{\Lambda}(t) = \int_0^t \frac{\mathbb{P}_n dN(s)}{\phi_n(s;\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})},$$

where

$$\phi_n(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) = \mathbb{P}_n \sum_{l=1}^{L} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}} \left\{ \int_t^{t^*} \tau_l(s;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)dN(s) + \int_t^{t^*} \tilde{\tau}_l(s;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)d\tilde{N}(s) \right\}.$$

Under regularity condition (C3), $p_l(\boldsymbol{x}_i;\boldsymbol{\alpha})$ is Lipschitz in $\boldsymbol{\alpha}$ and $\boldsymbol{z}_{il}^T\boldsymbol{\gamma}$ is Lipschitz in $\boldsymbol{\gamma}$; thus $\{p_l(\boldsymbol{x}_i;\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \mathcal{A}\}$ and $\{\exp(\boldsymbol{z}_{il}^T\boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \Gamma\}$ are Donsker classes (van der Vaart and Wellner, 1996, Chapter 2.10). Noting that $\int_0^t e^{\boldsymbol{z}_{il}^T\boldsymbol{\gamma}} d\Lambda(s)\} = e^{\boldsymbol{z}_{il}^T\boldsymbol{\gamma}}\Lambda(t)$, we can view $\tau_l(t;\boldsymbol{O}_i,\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)$ and $\tilde{\tau}_l(t;\boldsymbol{O}_i,\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)$ as random quantities obtained from $p_l(\boldsymbol{x}_i;\boldsymbol{\alpha})$, $\boldsymbol{z}_{il}^T\boldsymbol{\gamma}$, and a deterministic function $\Lambda(t)$ through standard operations, under which the Donsker property preserves (van der Vaart and Wellner, 1996, Chapter 2.7). It then follows that $\{\tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) : t \in [0,t^*], l = 1,\ldots,L, \boldsymbol{\alpha} \in \mathcal{A}, \boldsymbol{\gamma} \in \Gamma, \Lambda \in \mathcal{B}\}$ and $\{\tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) : t \in [0,t^*], l = 1,\ldots,L, \boldsymbol{\alpha} \in \mathcal{A}, \boldsymbol{\gamma} \in \Gamma, \Lambda \in \mathcal{B}\}$ are Donsker classes.

Then by Glivenko-Cantelli theorem

$$\sup_{t\in[0,t^*],\boldsymbol{\alpha}\in\mathcal{A},\boldsymbol{\gamma}\in\Gamma,\Lambda\in\mathcal{B}} |\phi_n(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) - \phi^*(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)| \to 0,$$

where

$$\phi^*(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) = P \sum_{l=1}^{L} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}} \left\{ \int_t^{t^*} \tau_l(s;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)dN(s) + \int_t^{t^*} \tilde{\tau}_l(s;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)d\tilde{N}(s) \right\}.$$

Then by step 2 and the continuity of ϕ_n in $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and Λ, we have

$$\sup_{t\in[0,t^*]} |\phi_n(t;\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda}) - \phi^*(t;\boldsymbol{\alpha}^*,\boldsymbol{\gamma}^*,\Lambda^*)|$$

$$\leqslant \sup_{t\in[0,t^*]} |\phi_n(t;\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda}) - \phi_n(t;\boldsymbol{\alpha}^*,\boldsymbol{\gamma}^*,\Lambda^*)| + \sup_{t\in[0,t^*]} |\phi_n(t;\boldsymbol{\alpha}^*,\boldsymbol{\gamma}^*,\Lambda^*) - \phi^*(t;\boldsymbol{\alpha}^*,\boldsymbol{\gamma}^*,\Lambda^*)| \to 0.$$

In addition, we also have $\sup_{t\in[0,t^*]}|\mathbb{P}_n dN(t) - PdN(t)| \to 0$. Then we define

$$\tilde{\Lambda}(t) = \int_0^t \frac{\mathbb{P}_n dN(s)}{\phi_n(s;\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)}.$$

Then by previous derivations,

$$\tilde{\Lambda}(t) \to \int_0^t \frac{PdN(s)}{\phi^*(s;\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)} = \Lambda_0(t)$$

uniformly. Now define

$$\begin{aligned}\ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) =& P\int_0^{t^*}\left[\log\left\{\sum_{l=1}^L p_l(\boldsymbol{x};\boldsymbol{\alpha})e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}\exp\left(-\int_0^t e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)\right)\right\} + \log\Lambda\{t\}\right]dN(t)\\ &+ P\int_0^{t^*}\log\left\{\sum_{l=1}^L p_l(\boldsymbol{x};\boldsymbol{\alpha})\exp\left(-\int_0^t e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)\right)\right\}d\tilde{N}(t).\end{aligned}$$

Then by definition of NPMLE, $\ell_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda}) - \ell_n(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\tilde{\Lambda}) \geqslant 0$, thus $\lim_n\{\ell_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda}) - \ell_n(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\tilde{\Lambda})\} \geqslant 0$. However, we can also show that

$$\lim_n\{\ell_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda}) - \ell_n(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\tilde{\Lambda})\} = \ell(\boldsymbol{\alpha}^*,\boldsymbol{\gamma}^*,\Lambda^*) - \ell(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0) \leqslant 0.$$

Therefore, $\ell(\boldsymbol{\alpha}^*,\boldsymbol{\gamma}^*,\Lambda^*) = \ell(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)$ and by regularity conditions (C1)-(C3), $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}_0$, $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_0$ and $\Lambda^*(t) = \Lambda_0(t)$. Thus the consistency follows.

Appendix A.2 *Proof of theorem 2*

Proof. We use Theorem 19.26 from Van der Vaart (2000) to conduct the proof. In addition to the score function $\dot{\ell}_{n,\Lambda}$ for Λ derived in (A.1), we also derive the score functions for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ as following

$$\begin{aligned}\dot{\ell}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) =& \mathbb{P}_n\int_0^{t^*}\sum_{l=1}^L \tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha})dN(t)\\ &+ \mathbb{P}_n\int_0^{t^*}\sum_{l=1}^L \tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha})d\tilde{N}(t);\\ \dot{\ell}_{n,\boldsymbol{\gamma}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) =& \mathbb{P}_n\int_0^{t^*}\sum_{l=1}^L \tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\left\{\boldsymbol{z}_l - \int_0^t \boldsymbol{z}_l e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)\right\}dN(t)\\ &- \mathbb{P}_n\int_0^{t^*}\sum_{l=1}^L \tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\int_0^t \boldsymbol{z}_l e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)d\tilde{N}(t).\end{aligned}$$

Let

$$\begin{aligned}
\dot{\ell}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) &= \int_0^{t^*}\sum_{l=1}^{L}\tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha})dN(t)\\
&\quad+\int_0^{t^*}\sum_{l=1}^{L}\tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha})d\tilde{N}(t);\\
\dot{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda) &= \int_0^{t^*}\sum_{l=1}^{L}\tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\Big\{\boldsymbol{z}_l-\int_0^t\boldsymbol{z}_l e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)\Big\}dN(t)\\
&\quad-\int_0^{t^*}\sum_{l=1}^{L}\tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\int_0^t\boldsymbol{z}_l e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)d\tilde{N}(t);\\
\dot{\ell}_{\Lambda}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)[h] &= \int_0^{t^*}\Big[\sum_{l=1}^{L}\tau_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\Big\{-\int_0^t h(s)e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)\Big\}+h(t)\Big]dN(t)\\
&\quad+\int_0^{t^*}\Big[\sum_{l=1}^{L}\tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)\Big\{-\int_0^t h(s)e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}}d\Lambda(s)\Big\}\Big]d\tilde{N}(t).
\end{aligned}$$

Then there exists $\delta>0$ such that the class of functions $\{\dot{\ell}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda),\dot{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda),\dot{\ell}_{\Lambda}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)[h]:||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||+||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||+\sup_{t\in[0,t^*]}|\Lambda(t)-\Lambda_0(t)|<\delta, h\in\mathcal{W}\}$ is Donsker. Define $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$. Then by consistency of $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})$, the continuity of the score functions and the dominated convergence theorem,

$$\begin{aligned}
\sup_{\boldsymbol{u},\boldsymbol{v},h}\Big|&\mathbb{G}_n\{\boldsymbol{u}^T\dot{\ell}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})+\boldsymbol{v}^T\dot{\ell}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})+\dot{\ell}_{\Lambda}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}},\hat{\Lambda})[h]\}\\
&-\mathbb{G}_n\{\boldsymbol{u}^T\dot{\ell}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\boldsymbol{v}^T\dot{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\dot{\ell}_{\Lambda}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)[h]\}\Big|\to 0.
\end{aligned}$$

The next step is to show that the map $W:\ell^\infty(\mathcal{U},\mathcal{V},\mathcal{W})\to\ell^\infty(\mathcal{U},\mathcal{V},\mathcal{W})$ defined by

$$W(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)[\boldsymbol{u},\boldsymbol{v},h]=P\{\boldsymbol{u}^T\dot{\ell}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)+\boldsymbol{v}^T\dot{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)+\dot{\ell}_{\Lambda}(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)[h]\}$$

is Fréchet-differentiable at $(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)$ with a derivative $V(\boldsymbol{u},\boldsymbol{v},h)$ that has a continuous inverse. By direct calculation, we can show that

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}&W(\boldsymbol{\alpha}_0+\epsilon\tilde{\boldsymbol{u}},\boldsymbol{\gamma}_0+\epsilon\tilde{\boldsymbol{v}},\Lambda_0+\epsilon\int\tilde{h}d\Lambda_0)[\boldsymbol{u},\boldsymbol{v},h]\\
&=\tilde{\boldsymbol{u}}^T\boldsymbol{B}_{\boldsymbol{\alpha}}[\boldsymbol{u},\boldsymbol{v},h]+\tilde{\boldsymbol{v}}^T\boldsymbol{B}_{\boldsymbol{\gamma}}[\boldsymbol{u},\boldsymbol{v},h]+\int_0^{t^*}B_\Lambda[\boldsymbol{u},\boldsymbol{v},h]\tilde{h}(s)d\Lambda_0(s),
\end{aligned}$$

where the operator $\boldsymbol{B}[\boldsymbol{u},\boldsymbol{v},h] \equiv (\boldsymbol{B}_{\boldsymbol{\alpha}},\boldsymbol{B}_{\boldsymbol{\gamma}},B_\Lambda)[\boldsymbol{u},\boldsymbol{v},h]$ can be rewritten as

$$
\begin{aligned}
&-\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \phi^*(t;\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)h(t) \end{pmatrix} \\
&+\begin{pmatrix} \boldsymbol{u}^T\varphi_1(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\boldsymbol{v}^T\vartheta_1(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\int\nu_1(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)h(t)d\Lambda_0(t)+\boldsymbol{u} \\ \boldsymbol{u}^T\varphi_2(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\boldsymbol{v}^T\vartheta_2(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\int\nu_2(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)h(t)d\Lambda_0(t)+\boldsymbol{v} \\ \boldsymbol{u}^T\varphi_3(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\boldsymbol{v}^T\vartheta_3(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\int\nu_3(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)h(t)d\Lambda_0(t) \end{pmatrix}.
\end{aligned}
\tag{A.2}
$$

Detailed calculations for $\boldsymbol{B}_{\boldsymbol{\alpha}},\boldsymbol{B}_{\boldsymbol{\gamma}}$ and B_Λ can be found in the next subsection. We need to show that the operator $\boldsymbol{B}$ is invertible on its range.

By definition of $\phi^*(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)$, it is clear that $\phi^*(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)>0$ for any choice of $(\boldsymbol{\alpha},\boldsymbol{\gamma},\Lambda)$. Thus, the first term in (A.2) is an invertible operator. In addition, by conditions (C2) and (C3) the second term is a compact operator. Then we can show $\boldsymbol{B}$ is invertible by showing $\boldsymbol{B}$ is one-to-one. That is, if $\boldsymbol{B}(\boldsymbol{u},\boldsymbol{v},h)=\boldsymbol{0}$ then $(\boldsymbol{u},\boldsymbol{v},h)=\boldsymbol{0}$. Now assuming that $\boldsymbol{B}(\boldsymbol{u},\boldsymbol{v},h)=\boldsymbol{0}$ for some $(\boldsymbol{u},\boldsymbol{v},h)\in\mathcal{U}\times\mathcal{V}\times\mathcal{W}$, it follows that

$$\left.\frac{\partial}{\partial\epsilon}\right|_{\epsilon=0} W(\boldsymbol{\alpha}_0+\epsilon\tilde{\boldsymbol{u}},\boldsymbol{\gamma}_0+\epsilon\tilde{\boldsymbol{v}},\Lambda_0+\epsilon\int\tilde{h}d\Lambda_0)[\boldsymbol{u},\boldsymbol{v},h]=0,$$

which further indicates that the score function across the path $(\boldsymbol{\alpha}_0+\epsilon\tilde{\boldsymbol{u}},\boldsymbol{\gamma}_0+\epsilon\tilde{\boldsymbol{v}},\Lambda_0+\epsilon\int\tilde{h}d\Lambda_0)$ is zero. That is, with probability one

$$\boldsymbol{u}^T\dot{\ell}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\boldsymbol{v}^T\dot{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)+\dot{\ell}_\Lambda(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)[h]=0.$$

By setting $dN(t)=1$, we have for arbitrary t

$$\sum_{l=1}^{L}\tau_l(t;\boldsymbol{O},\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)\left\{\boldsymbol{u}^T\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha}_0)+\{\boldsymbol{v}^T\boldsymbol{z}_l+h(t)\}-\left(\int_0^t\{\boldsymbol{v}^T\boldsymbol{z}_l+h(t)\}e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0}d\Lambda_0(s)\right)\right\}=0.$$

The above equation holds only when $\boldsymbol{u}=0$ and $\boldsymbol{v}^T\boldsymbol{z}_l+h(t)=0$. Since $\boldsymbol{v}^T\boldsymbol{z}_l$ is a constant and $h(\cdot)$ is an arbitrary function in $\mathcal{W}$, the only solution which satisfies $\boldsymbol{v}^T\boldsymbol{z}_l+h(t)=0$ for arbitrary t is $\boldsymbol{v}=\boldsymbol{0}$ and $h(\cdot)=0$. Thus, $\boldsymbol{B}$ is one-to-one and consequently invertible, such that the derivative of W is also invertible. Now let $(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}},\tilde{h})=\boldsymbol{B}^{-1}(\boldsymbol{u},\boldsymbol{v},h)$ for some $(\boldsymbol{u},\boldsymbol{v},h)\in\mathcal{U}\times\mathcal{V}\times\mathcal{W}$, then it follows by Theorem 19.26 from Van der Vaart (2000) that uniformly in $(\boldsymbol{u}, \boldsymbol{v}, h)$,

$$\sqrt{n}\{\boldsymbol{u}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \boldsymbol{v}^T(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \int_0^{t^*} hd(\hat{\Lambda} - \Lambda_0)\}$$
$$= -\mathbb{G}_n\{\tilde{\boldsymbol{u}}^T\dot{\ell}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0, \Lambda_0) + \tilde{\boldsymbol{v}}^T\dot{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0, \Lambda_0) + \dot{\ell}_{\Lambda}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0, \Lambda_0)[\tilde{h}]\} + o_p(1).$$

Thus, $\sqrt{n}\{\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0, \hat{\Lambda} - \Lambda_0\}$ is asymptotically Gaussian. It further indicates the asymptotically multivariate zero-mean normality of $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ and $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$, and the weak convergence of $\sqrt{n}\{\hat{\Lambda}(t) - \Lambda_0(t)\}$ to a univariate zero-mean Gaussian process on $[0, t^*]$. By similar semiparametric efficiency arguments (Bickel et al., 1993) as also used in (Zeng and Lin, 2006) and (Mao and Lin, 2017), the estimators $(\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\gamma}}^T)^T$ for the parametric component of the model are asymptotically semiparametric efficient.

Appendix A.3 *Analytical variance estimator*

By similar arguments as in Zeng and Lin (2006), a consistent variance estimator for $\sqrt{n}\{\boldsymbol{u}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \boldsymbol{v}^T(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \int_0^{t^*} hd(\hat{\Lambda} - \Lambda_0)\}$ can be constructed as

$$\hat{\boldsymbol{V}} = (\boldsymbol{u}^T, \boldsymbol{v}^T, \boldsymbol{H}^T)\hat{\mathcal{I}}_n^{-1}\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{H} \end{pmatrix},$$

where $n\hat{\mathcal{I}}_n$ is the empirical information matrix of the observed-data log-likelihood $\ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \Lambda; \boldsymbol{O})$, which treats $\Lambda(\cdot)$ as a piecewise constant function, and $\boldsymbol{H}$ is a vector of length m with jth component equal to $h(t_j)$. Then it is straightforward to obtain variance estimations for $(\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\gamma}}^T)^T$ and $\hat{\Lambda}(\cdot)$ with appropriate choices of $\boldsymbol{u}$, $\boldsymbol{v}$ and h. Since $\Lambda(t)$ is positive, the 95% confidence interval is constructed by log-transformation

$$\left(\hat{\Lambda}(t)\exp\left\{\frac{-1.96\hat{SE}\{\hat{\Lambda}(t)\}}{\hat{\Lambda}(t)}\right\}, \hat{\Lambda}(t)\exp\left\{\frac{1.96\hat{SE}\{\hat{\Lambda}(t)\}}{\hat{\Lambda}(t)}\right\}\right),$$

where $\hat{SE}\{\hat{\Lambda}(t)\}$ is the estimated standard error of $\hat{\Lambda}(t)$.

Appendix A.4 *Calculation related to the proof of theorem 2*

By direct calculation, we can show that

$$\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} W(\boldsymbol{\alpha}_0 + \epsilon\tilde{\boldsymbol{u}}, \boldsymbol{\gamma}_0 + \epsilon\tilde{\boldsymbol{v}}, \Lambda_0 + \epsilon\int \tilde{h} d\Lambda_0)[\boldsymbol{u}, \boldsymbol{v}, h]$$
$$= \tilde{\boldsymbol{u}}^T \boldsymbol{B}_{\boldsymbol{\alpha}}[\boldsymbol{u}, \boldsymbol{v}, h] + \tilde{\boldsymbol{v}}^T \boldsymbol{B}_{\boldsymbol{\gamma}}[\boldsymbol{u}, \boldsymbol{v}, h] + \int_0^s B_\Lambda[\boldsymbol{u}, \boldsymbol{v}, h]\tilde{h}(s) d\Lambda_0(s),$$

where the operator $\boldsymbol{B}[\boldsymbol{u}, \boldsymbol{v}, h] = (\boldsymbol{B}_{\boldsymbol{\alpha}}, \boldsymbol{B}_{\boldsymbol{\gamma}}, B_\Lambda)[\boldsymbol{u}, \boldsymbol{v}, h]$ satisfies

$$\begin{aligned}
\boldsymbol{B}_{\boldsymbol{\alpha}}[\boldsymbol{u}, \boldsymbol{v}, h] =& P\Bigg[\int_0^{t^*}\Bigg\{\sum_{l=1}^L \tau_{l0}(t)\frac{\partial^2}{\partial \boldsymbol{\alpha}^2}\log p_l(\boldsymbol{x}; \boldsymbol{\alpha}_0)\boldsymbol{u} \\
&\qquad + \boldsymbol{B}_{\boldsymbol{\alpha},l}(t)\bigg(\frac{\partial}{\partial \boldsymbol{\alpha}}\log p_l(\boldsymbol{x}; \boldsymbol{\alpha}_0)^T\boldsymbol{u} + \boldsymbol{z}_l^T\boldsymbol{v} - \int_0^t \boldsymbol{z}_l^T e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\boldsymbol{v}\bigg)\Bigg\} dN(t) \\
&+ \int_0^{t^*}\Bigg\{\sum_{l=1}^L \tilde{\tau}_{l0}(t)\frac{\partial^2}{\partial \boldsymbol{\alpha}^2}\log p_l(\boldsymbol{x}; \boldsymbol{\alpha}_0)\boldsymbol{u} \\
&\qquad + \tilde{\boldsymbol{B}}_{\boldsymbol{\alpha},l}(t)\bigg(\frac{\partial}{\partial \boldsymbol{\alpha}}\log p_l(\boldsymbol{x}; \boldsymbol{\alpha}_0)^T\boldsymbol{u} - \int_0^t \boldsymbol{z}_l^T e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\boldsymbol{v}\bigg)\Bigg\} d\tilde{N}(t) \\
&+ \sum_{l=1}^L \int_0^{t^*} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0}\bigg(-\int_s^{t^*}\boldsymbol{B}_{\boldsymbol{\alpha},l}(t)dN(t) - \int_s^{t^*}\tilde{\boldsymbol{B}}_{\boldsymbol{\alpha},l}(t)d\tilde{N}(t)\bigg)h(s)d\Lambda_0(s)\Bigg];
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{B}_{\boldsymbol{\gamma}}[\boldsymbol{u},\boldsymbol{v},h] =& P\Bigg[\int_0^{t^*}\Bigg\{\sum_{l=1}^L \tau_{l0}(t)\Big(-\int_0^t \boldsymbol{z}_l^{\otimes 2} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\Big)\boldsymbol{v} \\
&\qquad + \boldsymbol{B}_{\boldsymbol{\gamma},l}(t)\Big(\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha}_0)^T\boldsymbol{u} + \boldsymbol{z}_l^T\boldsymbol{v} - \int_0^t \boldsymbol{z}_l^T e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\boldsymbol{v}\Big)\Bigg\} dN(t) \\
&+ \int_0^{t^*}\Bigg\{\sum_{l=1}^L \tilde{\tau}_{l0}(t)\Big(-\int_0^t \boldsymbol{z}_l^{\otimes 2} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\Big)\boldsymbol{v} \\
&\qquad + \tilde{\boldsymbol{B}}_{\boldsymbol{\gamma},l}(t)\Big(\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha}_0)^T\boldsymbol{u} - \int_0^t \boldsymbol{z}_l^T e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\boldsymbol{v}\Big)\Bigg\} d\tilde{N}(t) \\
&- \sum_{l=1}^L \int_0^{t^*} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0}\Big(\int_s^{t^*}\{\boldsymbol{z}_l\tau_{l0}(t) + \boldsymbol{B}_{\boldsymbol{\gamma},l}(t)\} dN(t) \\
&\qquad + \int_s^{t^*}\{\boldsymbol{z}_l\tilde{\tau}_{l0}(t) + \tilde{\boldsymbol{B}}_{\boldsymbol{\gamma},l}(t)\} d\tilde{N}(t)\Big) h(s) d\Lambda_0(s)\Bigg];
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{B}_{\Lambda}[\boldsymbol{u},\boldsymbol{v},h] =& P\Bigg[\sum_{l=1}^L \int_0^{t^*}\Bigg\{\{B_{\Lambda,l}(s,1) + \tilde{B}_{\Lambda,l}(s,1)\}\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha}_0)^T\boldsymbol{u} \\
&+ B_{\Lambda,l}\{s, \boldsymbol{z}_l^T\boldsymbol{v} - \int_0^t \boldsymbol{z}_l^T\boldsymbol{v} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(\dot{t})\} + \tilde{B}_{\Lambda,l}\{s, -\int_0^t \boldsymbol{z}_l^T\boldsymbol{v} e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(\dot{t})\} \\
&+ B_{\Lambda,l}\{s, -\int_0^t h(\dot{t}) e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(\dot{t})\} + \tilde{B}_{\Lambda,l}\{s, -\int_0^t h(\dot{t}) e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(\dot{t})\} \\
&+ h(s) e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0}\Big(\int_s^{t^*} \tau_{l0}(t) dN(t) + \tilde{\tau}_{l0}(t) d\tilde{N}(t)\Big)\Bigg\}\tilde{h}(s) d\Lambda_0(s)\Bigg],
\end{aligned}
$$

where $\tau_{l0}(t) = \tau_l(t;\boldsymbol{O},\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)$, $\tilde{\tau}_{l0}(t) = \tilde{\tau}_l(t;\boldsymbol{O},\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0,\Lambda_0)$ and

$$
\begin{aligned}
\boldsymbol{B}_{\boldsymbol{\alpha},l}(t) &= \tau_{l0}(t)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha}_0) - \tau_{l0}(t)\sum_{d=1}^L \tau_{d0}(t)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_d(\boldsymbol{x};\boldsymbol{\alpha}_0); \\
\tilde{\boldsymbol{B}}_{\boldsymbol{\alpha},l}(t) &= \tilde{\tau}_{l0}(t)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_l(\boldsymbol{x};\boldsymbol{\alpha}_0) - \tilde{\tau}_{l0}(t)\sum_{d=1}^L \tilde{\tau}_{d0}(t)\frac{\partial}{\partial\boldsymbol{\alpha}}\log p_d(\boldsymbol{x};\boldsymbol{\alpha}_0); \\
\boldsymbol{B}_{\boldsymbol{\gamma},l}(t) &= \tau_{l0}(t)\Big\{\boldsymbol{z}_l - \int_0^t \boldsymbol{z}_l e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\Big\} - \tau_{l0}(t)\sum_{d=1}^L \tau_{d0}(t)\Big\{\boldsymbol{z}_d - \int_0^t \boldsymbol{z}_d e^{\boldsymbol{z}_d^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\Big\}; \\
\tilde{\boldsymbol{B}}_{\boldsymbol{\gamma},l}(t) &= \tilde{\tau}_{l0}(t)\Big\{-\int_0^t \boldsymbol{z}_l e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\Big\} - \tilde{\tau}_{l0}(t)\sum_{d=1}^L \tilde{\tau}_{d0}(t)\Big\{-\int_0^t \boldsymbol{z}_d e^{\boldsymbol{z}_d^T\boldsymbol{\gamma}_0} d\Lambda_0(s)\Big\}; \\
\boldsymbol{B}_{\Lambda,l}(s,g(\dot{t})) &= e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0}\Big\{\int_s^{t^*} g(\dot{t})\tau_{l0}(\dot{t}) dN(\dot{t}) - \sum_{d=1}^L \int_s^{t^*} g(\dot{t})\tau_{l0}(\dot{t})\tau_{d0}(\dot{t}) dN(\dot{t})\Big\}; \\
\tilde{\boldsymbol{B}}_{\Lambda,l}(s,g(\dot{t})) &= e^{\boldsymbol{z}_l^T\boldsymbol{\gamma}_0}\Big\{\int_s^{t^*} g(\dot{t})\tilde{\tau}_{l0}(\dot{t}) dN(\dot{t}) - \sum_{d=1}^L \int_s^{t^*} g(\dot{t})\tilde{\tau}_{l0}(\dot{t})\tilde{\tau}_{d0}(\dot{t}) dN(\dot{t})\Big\}.
\end{aligned}
$$

Appendix B. An additional simulation study

To assess the robustness when the proportionality assumption between class-specific baseline hazard functions is violated, we conducted an additional simulation study with two latent classes, where the class-specific cumulative baseline hazard functions for the first class and second class are $\Lambda_1(t) = 0.1(e^t - 1)$ and $\Lambda_2(t) = 0.1(e^{2t} - 1)$, correspondingly. The covariate effects $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ are the same as scenario (I). The new scenario is named scenario (VI).

Table S.8 displays the simulation results for scenario (VI) from 10000 simulations with sample size 1000 and perturbed initialization (convergence rate = 96.98%, median censoring = 12.8%, median entropy = 0.7708). As observed, estimated covariate effects $\hat{\boldsymbol{\alpha}}$ for the latent polytomous logistic regression model are slightly biased. In addition, estimated reference cumulative baseline hazard function $\hat{\Lambda}(t)$ is also biased due to the wrong model specification. However, estimated covariate effects $\hat{\zeta}$ for the class-specific Cox model are still unbiased with coverage probabilities close to 0.95. This result demonstrates the robustness of the proposed model when the proportionality assumption of baseline hazards is violated.

Figure S.6 compares the cross-validated Brier scores for the proposed model and the standard Cox model under scenario (VI). As observed, the proposed method still achieves smaller median Brier scores than the Cox model when the proportionality assumption is violated.

Appendix C. Additional tables and figures for simulation

Please find Supplementary Tables S.1 - S.4, and Supplementary Figures S.1 - S.5 at the end of this document.

Appendix D. Simulation results with non-informative initialization

Please find Supplementary Tables S.5 - S.7 at the end of this document.

References

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*, volume 4. Johns Hopkins University Press Baltimore.

Mao, L. and Lin, D. (2017). Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79,** 573–587.

van der Vaart, A. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93,** 627–640.

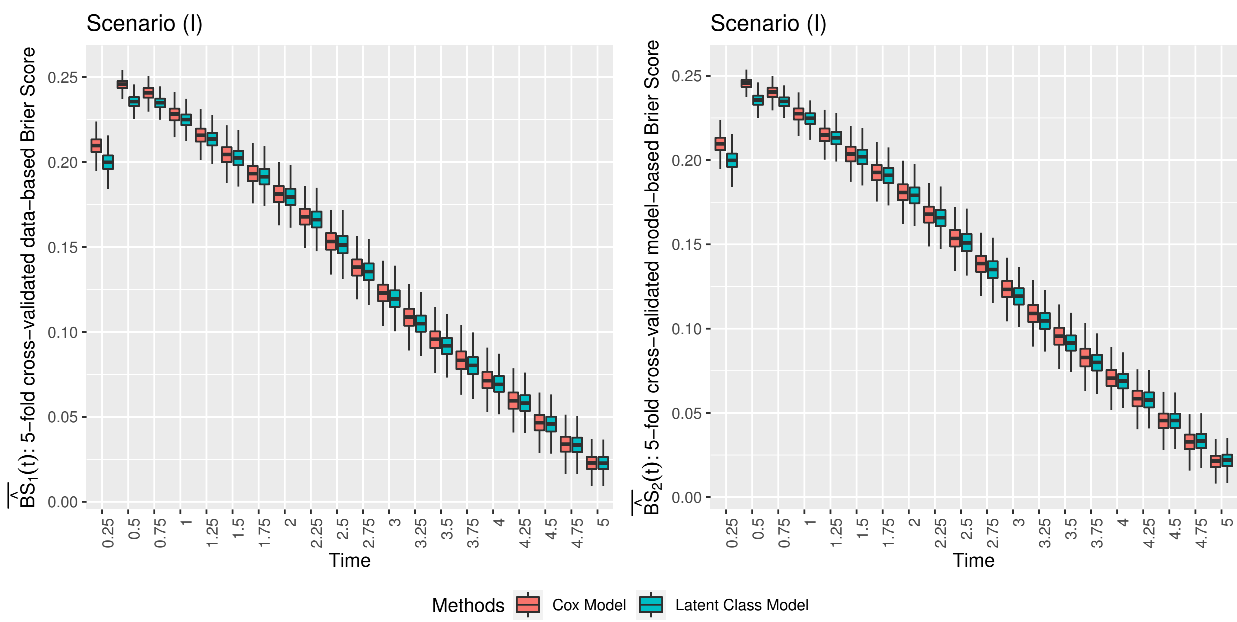

Figure S.1. Boxplots for average cross-validated Brier Score $\overline{\hat{\mathrm{BS}}}_1(t)$ and $\overline{\hat{\mathrm{BS}}}_2(t)$, $t \in (0, 5]$, from 1000 simulations under scenario (I) with sample size 1000, for the Cox model and the proposed latent class model with $L = 2$.

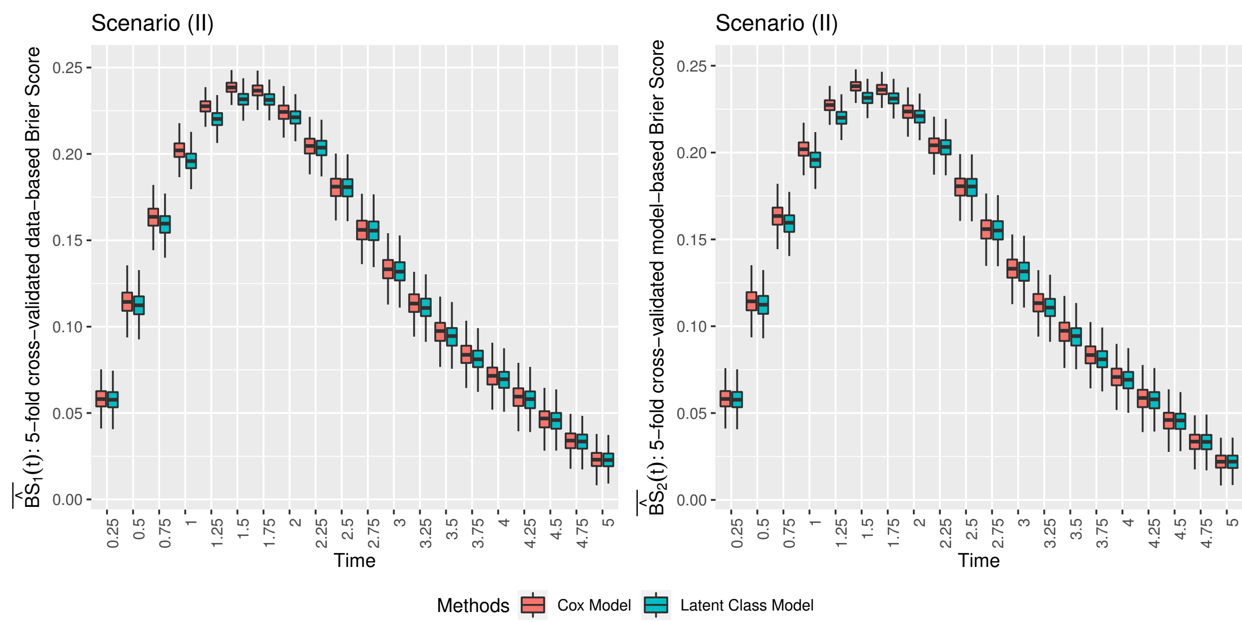

Figure S.2. Boxplots for average cross-validated Brier Score $\overline{\hat{\mathrm{BS}}_1}(t)$ and $\overline{\hat{\mathrm{BS}}_2}(t)$, $t \in (0, 5]$, from 1000 simulations under scenario (II) with sample size 1000, for the Cox model and the proposed latent class model with $L = 2$.

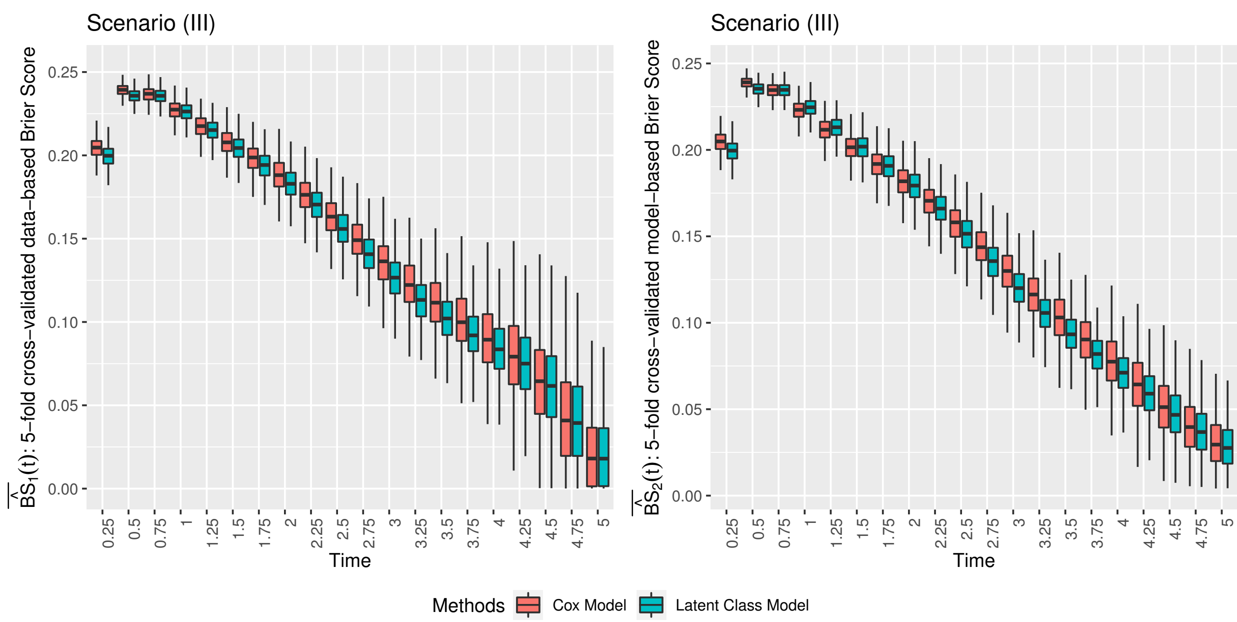

Figure S.3. Boxplots for average cross-validated Brier Score $\overline{\hat{\mathrm{BS}}_1}(t)$ and $\overline{\hat{\mathrm{BS}}_2}(t)$, $t \in (0, 5]$, from 1000 simulations under scenario (III) with sample size 1000, for the Cox model and the proposed latent class model with $L = 2$.

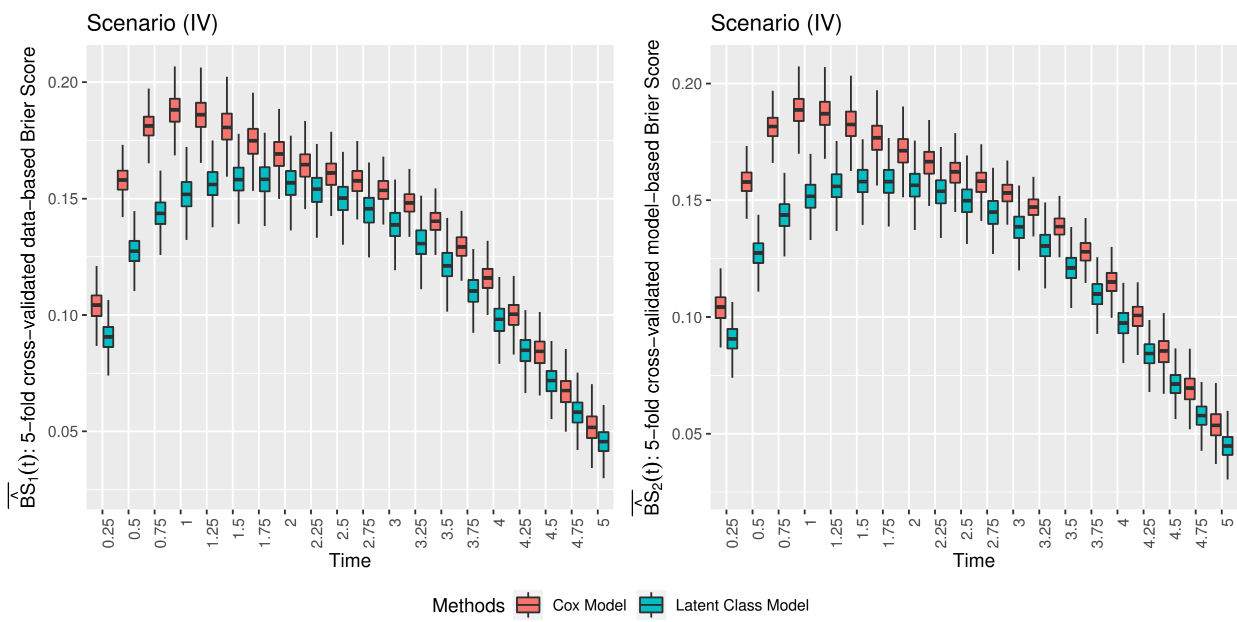

Figure S.4. Boxplots for average cross-validated Brier Score $\overline{\hat{\mathrm{BS}}}_1(t)$ and $\overline{\hat{\mathrm{BS}}}_2(t)$, $t \in (0, 5]$, from 1000 simulations under scenario (IV) with sample size 1000, for the Cox model and the proposed latent class model with $L = 3$.

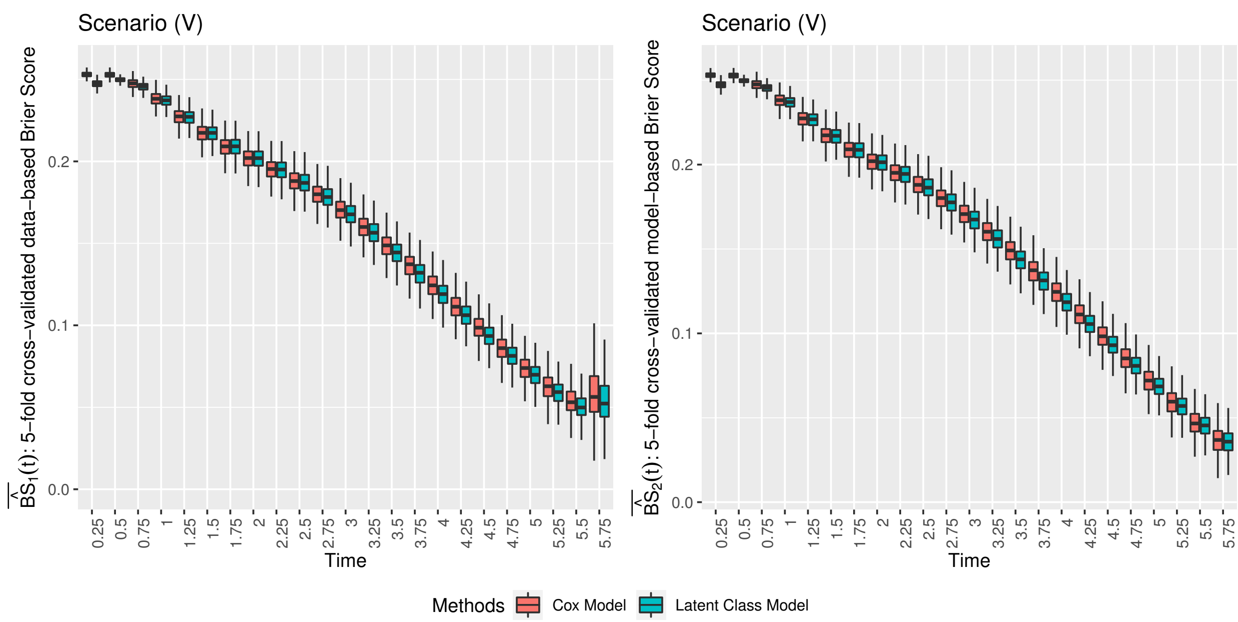

Figure S.5. Boxplots for average cross-validated Brier Score $\overline{\hat{\mathrm{BS}}_1}(t)$ and $\overline{\hat{\mathrm{BS}}_2}(t)$, $t \in (0, 5]$, from 1000 simulations under scenario (V) with sample size 1000, for the Cox model and the proposed latent class model with $L = 3$.

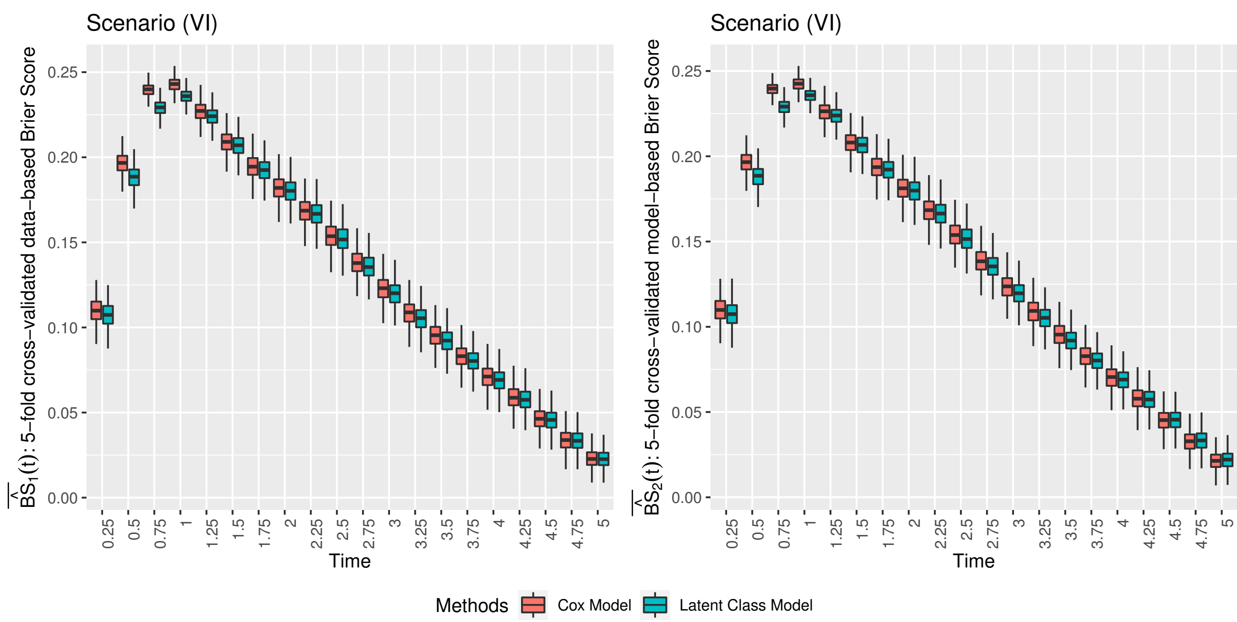

Figure S.6. Boxplots for average cross-validated Brier Score $\overline{\hat{\mathrm{BS}}_1}(t)$ and $\overline{\hat{\mathrm{BS}}_2}(t)$, $t \in (0, 5]$, from 1000 simulations under scenario (VI) with sample size 1000, for the Cox model and the proposed latent class model with $L = 2$.

Table S.1
Choices of parameters in the five simulation scenarios.

Simulation scenarios		Censoring parameter r	Parameters in model (2) $\boldsymbol{\alpha}$: $\boldsymbol{\alpha}_2$	$\boldsymbol{\alpha}_3$	Parameters in model (1) $\boldsymbol{\gamma}$: $\boldsymbol{\zeta}_1$	a_2	$\boldsymbol{\zeta}_2$	a_3	$\boldsymbol{\zeta}_{3,1}$
$L=2$	scenario (I)	0.1	(log(2),0,0)	NA	(-2,0)	2	(2,2)	NA	NA
	scenario (II)	0.1	(log(2),0,0)		(-2,0)	0	(2,2)		
	scenario (III)	0.6	(log(2),0,0)		(-2,0)	2	(2,2)		
	scenario (IV)	0.1	(2,-4,0)		(0,-3)	0.5	(0,6)		
$L=3$	scenario (V)	0.1	(0,-0.5,0)	(0,0,0.5)	(-2,-2)	2	(2,2)	4	(4,4)

Table S.2

Convergence rate, median standardized entropy index and median censoring rate out of 10000 simulations for the five simulation scenarios.

Simulation scenarios		Sample size	Convergence	Median entropy	Median censoring
$L = 2$	scenario (I)	1000	97.66%	0.7686	11%
	scenario (II)	1000	97.34%	0.4348	17%
	scenario (III)	1000	96.07%	0.6220	38%
	scenario (IV)	1000	97.40%	0.7771	19%
$L = 3$	scenario (V)	1000	98.65%	0.7602	15%
		2000	96.74%	0.7838	15%
		3000	94.06%	0.7785	15%

Table S.3

Simulation results for scenario (I)-(IV) out of 10000 simulations with sample size 1000 and perturbed initialization. Bias: mean bias; M.Bias: median bias; SE: standard deviation; SEE-P: median standard error estimate by profile likelihood variance estimation approach; CP-P: coverage probability by profile likelihood variance estimation approach; SEE-A: median standard error estimate by analytical variance estimation approach; CP-A: coverage probability by analytical variance estimation approach.

	Scenario (I)							Scenario (II)						
	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A
$\hat{\alpha}_{2,0}$	0.004	0.012	0.266	0.255	0.945	0.269	0.960	0.031	0.016	0.572	0.559	0.952	0.585	0.961
$\hat{\alpha}_{2,1}$	0.012	0.005	0.199	0.192	0.947	0.198	0.958	-0.014	-0.010	0.499	0.477	0.943	0.488	0.950
$\hat{\alpha}_{2,2}$	-0.017	-0.020	0.302	0.296	0.949	0.304	0.955	0.001	-0.007	0.522	0.500	0.945	0.517	0.956
$\hat{a}_2$	0.037	0.011	0.449	0.412	0.940	0.429	0.955	0.057	0.032	0.451	0.406	0.926	0.391	0.930
$\hat{\zeta}_{11}$	-0.033	-0.024	0.199	0.201	0.956	0.191	0.949	-0.070	-0.045	0.318	0.314	0.958	0.293	0.943
$\hat{\zeta}_{12}$	0.014	0.013	0.256	0.258	0.952	0.252	0.948	-0.010	-0.010	0.345	0.341	0.946	0.330	0.939
$\hat{\zeta}_{21}$	0.027	0.020	0.216	0.218	0.954	0.211	0.950	0.062	0.040	0.311	0.309	0.959	0.293	0.952
$\hat{\zeta}_{22}$	0.006	0.009	0.298	0.301	0.954	0.297	0.954	0.024	0.021	0.406	0.402	0.945	0.392	0.944
$\hat{\Lambda}(2)$	0.010	-0.005	0.166	NA	NA	0.165	0.953	0.006	-0.007	0.215	NA	NA	0.197	0.938
$\hat{\Lambda}(3)$	0.018	-0.010	0.351	NA	NA	0.344	0.951	0.041	0.003	0.545	NA	NA	0.504	0.949
$\hat{\Lambda}(4)$	0.180	0.028	1.164	NA	NA	1.047	0.945	0.419	0.158	1.596	NA	NA	1.348	0.955
	Scenario (III)							Scenario (IV)						
	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A
$\hat{\alpha}_{2,0}$	-0.003	0.011	0.377	0.340	0.922	0.372	0.946	0.016	0.006	0.484	0.505	0.964	0.520	0.970
$\hat{\alpha}_{2,1}$	0.037	0.025	0.259	0.234	0.935	0.245	0.949	-0.037	-0.023	0.316	0.303	0.947	0.309	0.953
$\hat{\alpha}_{2,2}$	-0.036	-0.037	0.410	0.380	0.936	0.400	0.953	0.011	0.010	0.665	0.696	0.968	0.717	0.971
$\hat{a}_2$	0.058	0.016	0.733	0.616	0.914	0.648	0.930	0.000	0.002	0.310	0.309	0.954	0.302	0.951
$\hat{\zeta}_{11}$	-0.105	-0.063	0.427	0.403	0.963	0.390	0.958	-0.011	-0.011	0.205	0.207	0.952	0.194	0.937
$\hat{\zeta}_{12}$	0.029	0.020	0.542	0.532	0.949	0.517	0.948	-0.027	-0.022	0.246	0.251	0.954	0.243	0.952
$\hat{\zeta}_{21}$	0.091	0.057	0.428	0.408	0.961	0.396	0.959	0.017	0.014	0.358	0.344	0.944	0.333	0.940
$\hat{\zeta}_{22}$	0.009	0.019	0.549	0.542	0.950	0.530	0.950	0.040	0.031	0.331	0.336	0.954	0.330	0.951
$\hat{\Lambda}(2)$	0.045	-0.003	0.302	NA	NA	0.278	0.932	0.025	0.001	0.188	NA	NA	0.179	0.950
$\hat{\Lambda}(3)$	0.106	-0.018	0.759	NA	NA	0.661	0.942	0.075	0.016	0.481	NA	NA	0.449	0.945
$\hat{\Lambda}(4)$	0.856	0.075	3.299	NA	NA	2.278	0.937	0.257	0.089	1.337	NA	NA	1.217	0.941

Table S.4

Simulation results for scenario (V) out of 10000 simulations with perturbed initialization and different choices of sample size. Bias: mean bias; M.Bias: median bias; SE: standard deviation; SEE-P: median standard error estimate by profile likelihood variance estimation approach; CP-P: coverage probability by profile likelihood variance estimation approach; SEE-A: median standard error estimate by analytical variance estimation approach; CP-A: coverage probability by analytical variance estimation approach.

	Scenario (V), $n = 1000$							Scenario (V), $n = 2000$							Scenario (V), $n = 3000$						
	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A	Bias	M.Bias	SE	SEE-P	CP-P	SEE-A	CP-A
$\hat{\alpha}_{2,0}$	-0.051	-0.034	0.677	0.591	0.882	0.614	0.903	-0.049	-0.040	0.470	0.437	0.908	0.443	0.905	-0.046	-0.038	0.377	0.360	0.928	0.369	0.912
$\hat{\alpha}_{2,1}$	0.039	0.033	0.441	0.371	0.910	0.386	0.924	0.033	0.024	0.291	0.268	0.928	0.265	0.925	0.027	0.022	0.230	0.219	0.938	0.213	0.931
$\hat{\alpha}_{2,2}$	0.044	0.039	0.581	0.515	0.908	0.548	0.936	0.042	0.034	0.389	0.371	0.929	0.383	0.943	0.039	0.036	0.311	0.305	0.940	0.315	0.945
$\hat{\alpha}_{3,0}$	0.003	-0.008	0.417	0.343	0.887	0.352	0.900	-0.006	-0.011	0.286	0.253	0.908	0.259	0.913	-0.012	-0.017	0.229	0.210	0.920	0.218	0.925
$\hat{\alpha}_{3,1}$	0.000	0.011	0.293	0.241	0.903	0.251	0.923	0.005	0.009	0.196	0.174	0.919	0.177	0.929	0.007	0.010	0.155	0.143	0.931	0.145	0.936
$\hat{\alpha}_{3,2}$	0.011	0.017	0.378	0.337	0.919	0.355	0.943	0.010	0.015	0.250	0.240	0.939	0.250	0.952	0.015	0.017	0.206	0.197	0.936	0.207	0.952
$\hat{a}_2$	-0.326	-0.256	1.160	0.787	0.791	0.667	0.718	-0.178	-0.122	0.828	0.628	0.872	0.548	0.793	-0.105	-0.074	0.631	0.534	0.932	0.502	0.851
$\hat{a}_3$	-0.226	-0.354	0.997	0.760	0.797	0.592	0.715	-0.137	-0.212	0.729	0.626	0.862	0.480	0.744	-0.093	-0.139	0.578	0.548	0.906	0.431	0.783
$\hat{\zeta}_{11}$	-0.056	-0.042	0.236	0.216	0.938	0.292	0.977	-0.027	-0.022	0.151	0.146	0.946	0.218	0.987	-0.015	-0.014	0.121	0.118	0.946	0.190	0.990
$\hat{\zeta}_{12}$	-0.072	-0.058	0.358	0.331	0.933	0.382	0.964	-0.032	-0.026	0.235	0.225	0.943	0.273	0.973	-0.021	-0.018	0.187	0.181	0.945	0.229	0.979
$\hat{\zeta}_{21}$	0.070	0.100	0.474	0.342	0.851	0.437	0.904	0.033	0.037	0.312	0.244	0.876	0.355	0.938	0.011	0.011	0.239	0.200	0.902	0.321	0.964
$\hat{\zeta}_{22}$	0.120	0.148	0.702	0.533	0.854	0.608	0.899	0.046	0.053	0.458	0.374	0.888	0.479	0.933	0.027	0.023	0.356	0.306	0.906	0.429	0.957
$\hat{\zeta}_{31}$	0.066	0.057	0.285	0.267	0.935	0.350	0.975	0.032	0.029	0.189	0.183	0.944	0.262	0.984	0.018	0.018	0.152	0.148	0.942	0.226	0.989
$\hat{\zeta}_{32}$	0.084	0.072	0.422	0.402	0.942	0.524	0.970	0.036	0.032	0.286	0.277	0.946	0.410	0.981	0.027	0.025	0.230	0.224	0.941	0.367	0.983
$\hat{\Lambda}(2)$	0.251	0.217	0.449	NA	NA	0.324	0.724	0.137	0.122	0.309	NA	NA	0.230	0.765	0.089	0.084	0.245	NA	NA	0.191	0.802
$\hat{\Lambda}(3)$	0.127	0.117	0.585	NA	NA	0.520	0.910	0.052	0.062	0.390	NA	NA	0.367	0.925	0.034	0.038	0.315	NA	NA	0.298	0.927
$\hat{\Lambda}(4)$	0.329	0.182	1.276	NA	NA	1.263	0.961	0.147	0.098	0.806	NA	NA	0.839	0.967	0.101	0.062	0.642	NA	NA	0.670	0.963

Table S.5

Convergence rate, median standardized entropy index and median censoring rate out of 10000 simulations for the five simulation scenarios with non-informative initialization.

Simulation scenarios		Sample size	Convergence	Median entropy	Median censoring
$L = 2$	scenario (I)	1000	97.66%	0.7667	11%
	scenario (II)	1000	97.38%	0.4348	17%
	scenario (III)	1000	96.06%	0.6228	38%
	scenario (IV)	1000	97.20%	0.7766	19%
$L = 3$	scenario (V)	1000	97.68%	0.7585	15%
		2000	98.14%	0.7660	15%
		3000	95.29%	0.7717	15%

Table S.6

Simulation results for the simulation scenarios (I) - (IV) out of 10000 simulations with sample size $n = 1000$ and non-informative initialization. M.Bias: median bias; SE: standard deviation; SEE: median standard error estimate; CP: coverage probability. Profile likelihood variance estimation approach was used for $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\gamma}}$. Analytical approach based on observed-data log-likelihood was used for $\hat{\Lambda}(t)$.

	Scenario (I)				Scenario (II)			
	M.Bias	SE	SEE	CP	M.Bias	SE	SEE	CP
$\hat{\alpha}_{2,0}$	0.014	0.267	0.255	0.941	0.003	0.580	0.560	0.947
$\hat{\alpha}_{2,1}$	0.007	0.201	0.192	0.943	0.000	0.503	0.477	0.945
$\hat{\alpha}_{2,2}$	-0.015	0.304	0.296	0.949	0.005	0.517	0.500	0.945
$\hat{a}_2$	0.011	0.446	0.412	0.941	0.023	0.441	0.406	0.930
$\hat{\zeta}_{11}$	-0.023	0.197	0.201	0.957	-0.045	0.319	0.313	0.956
$\hat{\zeta}_{12}$	0.009	0.256	0.258	0.953	-0.004	0.347	0.340	0.943
$\hat{\zeta}_{21}$	0.018	0.214	0.218	0.959	0.032	0.313	0.308	0.960
$\hat{\zeta}_{22}$	0.003	0.296	0.302	0.958	0.019	0.399	0.402	0.953
$\hat{\Lambda}(2)$	-0.004	0.163	0.165	0.956	-0.004	0.212	0.197	0.940
$\hat{\Lambda}(3)$	-0.008	0.352	0.344	0.951	0.010	0.540	0.503	0.952
$\hat{\Lambda}(4)$	0.034	1.146	1.051	0.944	0.151	1.606	1.345	0.950
	Scenario (III)				Scenario (IV)			
	M.Bias	SE	SEE	CP	M.Bias	SE	SEE	CP
$\hat{\alpha}_{2,0}$	0.018	0.381	0.340	0.924	0.018	0.488	0.504	0.960
$\hat{\alpha}_{2,1}$	0.021	0.258	0.234	0.936	-0.015	0.308	0.303	0.953
$\hat{\alpha}_{2,2}$	-0.035	0.417	0.380	0.936	-0.011	0.668	0.695	0.962
$\hat{a}_2$	0.010	0.729	0.618	0.918	0.003	0.311	0.309	0.952
$\hat{\zeta}_{11}$	-0.069	0.429	0.404	0.963	-0.013	0.204	0.206	0.951
$\hat{\zeta}_{12}$	0.010	0.542	0.534	0.950	-0.023	0.249	0.251	0.953
$\hat{\zeta}_{21}$	0.061	0.431	0.409	0.960	0.012	0.357	0.344	0.939
$\hat{\zeta}_{22}$	0.022	0.547	0.545	0.949	0.038	0.332	0.336	0.955
$\hat{\Lambda}(2)$	0.006	0.302	0.282	0.935	0.003	0.190	0.179	0.946
$\hat{\Lambda}(3)$	0.000	0.772	0.675	0.940	0.018	0.488	0.450	0.941
$\hat{\Lambda}(4)$	0.089	3.440	2.291	0.941	0.096	1.351	1.217	0.939

Table S.7

Simulation results for scenario (V) out of 10000 simulations with non-informative initialization and different choices of sample size. M.Bias: median bias; SE: standard deviation; SEE: median standard error estimate; CP: coverage probability. Profile likelihood variance estimation approach was used for $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\gamma}}$. Analytical approach based on observed-data log-likelihood was used for $\hat{\Lambda}(t)$.

	Scenario (V), $n = 1000$				Scenario (V), $n = 2000$				Scenario (V), $n = 3000$			
	M.Bias	SE	SEE	CP	M.Bias	SE	SEE	CP	M.Bias	SE	SEE	CP
$\hat{\alpha}_{2,0}$	0.053	0.725	0.623	0.890	0.032	0.476	0.458	0.935	-0.001	0.383	0.376	0.940
$\hat{\alpha}_{2,1}$	-0.011	0.501	0.403	0.909	-0.006	0.303	0.284	0.942	0.005	0.239	0.230	0.943
$\hat{\alpha}_{2,2}$	0.013	0.606	0.532	0.908	0.012	0.397	0.381	0.936	0.020	0.314	0.312	0.944
$\hat{\alpha}_{3,0}$	0.058	0.449	0.359	0.889	0.042	0.291	0.264	0.927	0.021	0.235	0.217	0.928
$\hat{\alpha}_{3,1}$	-0.027	0.334	0.255	0.902	-0.024	0.204	0.182	0.930	-0.014	0.162	0.148	0.931
$\hat{\alpha}_{3,2}$	-0.011	0.386	0.346	0.922	-0.018	0.257	0.246	0.940	-0.002	0.208	0.200	0.937
$\hat{a}_2$	-0.868	1.229	0.783	0.675	-0.515	0.884	0.653	0.789	-0.346	0.679	0.562	0.863
$\hat{a}_3$	-0.656	1.008	0.717	0.711	-0.445	0.696	0.611	0.799	-0.309	0.550	0.533	0.857
$\hat{\zeta}_{11}$	-0.063	0.256	0.229	0.941	-0.036	0.157	0.152	0.947	-0.024	0.122	0.120	0.952
$\hat{\zeta}_{12}$	-0.108	0.385	0.355	0.936	-0.058	0.240	0.235	0.947	-0.036	0.187	0.187	0.950
$\hat{\zeta}_{21}$	0.228	0.472	0.353	0.829	0.142	0.314	0.247	0.835	0.093	0.246	0.203	0.866
$\hat{\zeta}_{22}$	0.333	0.707	0.555	0.841	0.200	0.455	0.384	0.869	0.120	0.356	0.312	0.896
$\hat{\zeta}_{31}$	0.078	0.298	0.276	0.940	0.046	0.187	0.187	0.948	0.034	0.151	0.150	0.947
$\hat{\zeta}_{32}$	0.129	0.450	0.421	0.936	0.069	0.286	0.284	0.950	0.041	0.227	0.228	0.952
$\hat{\Lambda}(2)$	0.344	0.497	0.360	0.673	0.179	0.336	0.255	0.762	0.115	0.245	0.203	0.803
$\hat{\Lambda}(3)$	0.196	0.643	0.566	0.904	0.060	0.414	0.410	0.935	0.019	0.322	0.328	0.946
$\hat{\Lambda}(4)$	0.278	1.399	1.334	0.961	0.135	0.843	0.888	0.969	0.080	0.655	0.694	0.972

Table S.8

Simulation results for scenario (VI) out of 10000 simulations with sample size 1000 and perturbed initialization. M.Bias: median bias; SE: standard deviation; SEE: median standard error estimate; CP: coverage probability. Profile likelihood variance estimation approach was used for $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\gamma}}$. Analytical approach based on observed-data log-likelihood was used for $\hat{\Lambda}(t)$.

	M.Bias	SE	SEE	CP
$\hat{\alpha}_{2,0}$	0.204	0.276	0.251	0.841
$\hat{\alpha}_{2,1}$	-0.147	0.210	0.192	0.852
$\hat{\alpha}_{2,2}$	-0.043	0.309	0.298	0.946
$\hat{\zeta}_{11}$	-0.032	0.204	0.207	0.956
$\hat{\zeta}_{12}$	0.025	0.270	0.268	0.947
$\hat{\zeta}_{21}$	0.028	0.217	0.222	0.959
$\hat{\zeta}_{22}$	-0.070	0.311	0.308	0.948
$\hat{\Lambda}(2)$	-0.125	0.166	0.162	0.918
$\hat{\Lambda}(3)$	-0.145	0.354	0.340	0.935
$\hat{\Lambda}(4)$	-0.110	1.180	1.066	0.940