JOINT STRUCTURE SELECTION AND ESTIMATION IN THE TIME-VARYING COEFFICIENT COX MODEL

Wei Xiao, Wenbin Lu and Hao Helen Zhang

North Carolina State University and University of Arizona

Supplementary Material

This note contains the proofs of Theorem 1, 2 and 3 in the main paper. Theorem 1 establishes the root-n consistency of $\widehat{\boldsymbol{\lambda}}^{(1)}$. Theorem 2 establishes the sparsity property of $\widehat{\boldsymbol{\lambda}}^{(0)}$. Theorem 3, which combines both Theorems 1 and 2, is our main theorem. It establishes the selection consistency of the KGNG estimator and its asymptotic normalities for both nonzero constant and time-varying regression coefficient.

Before we give the details of proofs, we first introduce some additional notation. With simple matrix multiplication, we have $\boldsymbol{\lambda}_0^{(1)} = P_1\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_0^{(0)} = P_2\boldsymbol{\lambda}_0$, where

$$P_1 = \begin{pmatrix} \mathbf{0}_{p_2\times p_1} & \boldsymbol{I}_{p_2} & \mathbf{0}_{p_2\times p_3} & \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_2} & \mathbf{0}_{p_2\times p_3} \\ \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_2} & \boldsymbol{I}_{p_3} & \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_2} & \mathbf{0}_{p_3\times p_3} \\ \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_2} & \mathbf{0}_{p_3\times p_3} & \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_2} & \boldsymbol{I}_{p_3} \end{pmatrix}_{(p_2+2p_3)\times 2p}; \quad \text{(S0.1)}$$

$$P_2 = \begin{pmatrix} \boldsymbol{I}_{p_1} & \mathbf{0}_{p_1\times p_2} & \mathbf{0}_{p_1\times p_3} & \mathbf{0}_{p_1\times p_1} & \mathbf{0}_{p_1\times p_2} & \mathbf{0}_{p_1\times p_3} \\ \mathbf{0}_{p_1\times p_1} & \mathbf{0}_{p_1\times p_2} & \mathbf{0}_{p_1\times p_3} & \boldsymbol{I}_{p_1} & \mathbf{0}_{p_1\times p_2} & \mathbf{0}_{p_1\times p_3} \\ \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_2} & \mathbf{0}_{p_2\times p_3} & \mathbf{0}_{p_2\times p_1} & \boldsymbol{I}_{p_2} & \mathbf{0}_{p_2\times p_3} \end{pmatrix}_{(2p_1+p_2)\times 2p}. \quad \text{(S0.2)}$$

Let $\overline{\boldsymbol{\lambda}}_0 = (\boldsymbol{\lambda}_0^{(1)^\intercal}, \boldsymbol{\lambda}_0^{(0)^\intercal})^\intercal$. We also have $\boldsymbol{\lambda}_{01} = P_3\overline{\boldsymbol{\lambda}}_0 = P_{31}\boldsymbol{\lambda}^{(1)} + P_{32}\boldsymbol{\lambda}^{(0)}$ and $\boldsymbol{\lambda}_{02} = P_4\overline{\boldsymbol{\lambda}}_0 = P_{41}\boldsymbol{\lambda}^{(1)} + P_{42}\boldsymbol{\lambda}^{(0)}$, where

$$P_3 = (P_{31}|P_{32}) = \left(\begin{array}{ccc|ccc} \mathbf{0}_{p_1\times p_2} & \mathbf{0}_{p_1\times p_3} & \mathbf{0}_{p_1\times p_3} & \boldsymbol{I}_{p_1} & \mathbf{0}_{p_1\times p_1} & \mathbf{0}_{p_1\times p_2} \\ \boldsymbol{I}_{p_2} & \mathbf{0}_{p_2\times p_3} & \mathbf{0}_{p_2\times p_3} & \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_2} \\ \mathbf{0}_{p_3\times p_3} & \boldsymbol{I}_{p_3} & \mathbf{0}_{p_3\times p_3} & \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_2} \end{array}\right)_{p\times 2p}; \quad \text{(S0.3)}$$

$$P_4 = (P_{41}|P_{42}) = \left(\begin{array}{ccc|ccc} \mathbf{0}_{p_1\times p_2} & \mathbf{0}_{p_1\times p_3} & \mathbf{0}_{p_1\times p_3} & \mathbf{0}_{p_1\times p_1} & \boldsymbol{I}_{p_1} & \mathbf{0}_{p_1\times p_2} \\ \mathbf{0}_{p_2\times p_2} & \mathbf{0}_{p_2\times p_3} & \mathbf{0}_{p_2\times p_3} & \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_1} & \boldsymbol{I}_{p_2} \\ \mathbf{0}_{p_3\times p_2} & \mathbf{0}_{p_3\times p_3} & \boldsymbol{I}_{p_3} & \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_1} & \mathbf{0}_{p_3\times p_2} \end{array}\right)_{p\times 2p}. \quad \text{(S0.4)}$$

Let $L_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot))$ and $Q_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot))$ denote the partial likelihood and penalized partial likelihood in equation (2.3) respectively. We then have

$$L_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}(\cdot)) = \sum_{i=1}^{n} \int_0^{\tau} \left[a(s)^{\intercal} \boldsymbol{Z}_i - \log \left(\sum_{j=1}^{n} Y_j(s) e^{a(s)^{\intercal} \boldsymbol{Z}_j} \right) \right] dN_i(s);$$
$$Q_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}(\cdot)) = - L_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}(\cdot)) + \theta_1 \|\boldsymbol{\lambda}_1\|_1 + \theta_2 \|\boldsymbol{\lambda}_2\|_1,$$

where

$$a(s) \equiv a(s, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot)) \equiv a(s, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot)) = \boldsymbol{\lambda}_1 \circ \boldsymbol{m} + \boldsymbol{\lambda}_2 \circ \boldsymbol{\beta}^*(s). \quad \text{(S0.5)}$$

Let $U(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot))$ and $H(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot))$ denote the vector and matrix of the first and second order partial derivatives of $L_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}(\cdot))$ with respect to $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^{\intercal}, \boldsymbol{\lambda}_2^{\intercal})^{\intercal}$. Thus

$$U(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}(\cdot)) = \sum_{i=1}^{n} \int_0^{\tau} A(s) \left(\boldsymbol{Z}_i - \boldsymbol{E}(a(s), s) \right) dN_i(s),$$
$$H(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \boldsymbol{m}, \boldsymbol{\beta}(\cdot)) = - \sum_{i=1}^{n} \int_0^{\tau} A(s) \left[\frac{S^{(2)}(a(s), s)}{S^{(0)}(a(s), s)} - \left(\frac{S^{(1)}(a(s), s)}{S^{(0)}(a(s), s)} \right)^{\otimes 2} \right] A(s)^{\intercal} dN_i(s),$$

where

$$A(s) \equiv A(s; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot)) = \begin{pmatrix} \operatorname{diag}(\boldsymbol{m}) \\ \operatorname{diag}(\boldsymbol{\beta}^*(s)) \end{pmatrix}_{2p \times p}. \quad \text{(S0.6)}$$

Rearranging the order of $\boldsymbol{\lambda}$, we rewrite $L_{2n}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2; \cdot)$ as $L_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$. Denote the first and second order partial derivatives of $L_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$ with respect to $\boldsymbol{\lambda}^{(1)}$ as $U_1(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$ and $H_{11}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$. Denote the first and second order partial derivatives of $L_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$ with respect to $\boldsymbol{\lambda}^{(0)}$ as $U_2(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$ and $H_{22}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \cdot)$. In the following proofs, we use $\|\cdot\|_1$ and $\|\cdot\|$ to denote ℓ_1 and ℓ_2 norms, respectively.

S1 Proof of Theorem 1

Define $\widetilde{a}(s) \equiv a(s, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot))$, $\widetilde{A}(s) \equiv A(s; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot))$ and

$$A_0(s) \equiv A(s; \boldsymbol{m}_0, \boldsymbol{\beta}_0^*(\cdot)), \quad \text{(S1.7)}$$

where $a(s, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot))$ and $A(s; \boldsymbol{m}, \boldsymbol{\beta}^*(\cdot))$ are defined in (S0.5) and (S0.6), respectively. The second order partial derivative of $Q_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot))$ with respect to $\boldsymbol{\lambda}^{(1)}$ is

$$\begin{aligned}
&\frac{\partial^2 Q_{2n}(\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))}{\partial\boldsymbol{\lambda}^{(1)}\partial\boldsymbol{\lambda}^{(1)\top}} = -H_{11}(\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))\\
=&\sum_{i=1}^n\int_0^\tau P_1\widetilde{A}(s)\left[\frac{S^{(2)}(\widetilde{a}(s),s)}{S^{(0)}(\widetilde{a}(s),s)}-\left(\frac{S^{(1)}(\widetilde{a}(s),s)}{S^{(0)}(\widetilde{a}(s),s)}\right)^{\otimes 2}\right]\left(P_1\widetilde{A}(s)\right)^\top dN_i(s)\\
=&\sum_{i=1}^n\int_0^\tau P_1\widetilde{A}(s)\left[\frac{\sum_{i<j}Y_i(s)Y_j(s)e^{\widetilde{a}(s)^\top\boldsymbol{Z}_i}e^{\widetilde{a}(s)^\top\boldsymbol{Z}_j}\left(\boldsymbol{Z}_i-\boldsymbol{Z}_j\right)^{\otimes 2}}{\left(S^{(0)}(\widetilde{a}(s),s)\right)^2}\right]\left(P_1\widetilde{A}(s)\right)^\top dN_i(s),
\end{aligned}$$

which is positive definite for large n and P_1 is defined in (S0.1). Thus, for any fixed $\boldsymbol{\lambda}^{(0)}$, there exists a unique minimizer $\widehat{\boldsymbol{\lambda}}^{(1)}$ of $Q_{2n}(\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))$ when n is large.

Consider the C-ball $B_n(C)=\{\boldsymbol{\lambda}^{(1)}:\boldsymbol{\lambda}^{(1)}=\boldsymbol{\lambda}_0^{(1)}+n^{-1/2}\boldsymbol{u},\|\boldsymbol{u}\|\le C\}$, $C>0$, and denote its boundary by $\partial B_n(C)$. To prove $\|\widehat{\boldsymbol{\lambda}}^{(1)}-\boldsymbol{\lambda}_0^{(1)}\|=O_p(n^{-1/2})$, it is sufficient to show that, for any given $\epsilon>0$, there exists a large constant C so that

$$\operatorname{pr}\left\{\sup_{\boldsymbol{\lambda}^{(1)}\in\partial B_n(C)} Q_{2n}(\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))<Q_{2n}(\boldsymbol{\lambda}_0^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))\right\}\le 1-\epsilon. \quad \text{(S1.8)}$$

Then we have

$$\begin{aligned}
&D_{2n}(u)=Q_{2n}(\boldsymbol{\lambda}_0^{(1)}+n^{-1/2}\boldsymbol{u},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))-Q_{2n}(\boldsymbol{\lambda}_0^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))\\
\le&-U_1^\top(\boldsymbol{\lambda}_0^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))\frac{\boldsymbol{u}}{\sqrt{n}}+\frac{1}{n}\boldsymbol{u}^\top\left[-H_{11}(\boldsymbol{\lambda}^{(1)*},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))\right]\boldsymbol{u}\\
&+\frac{\max(\theta_1,\theta_2)}{\sqrt{n}}\|\boldsymbol{u}\|_1, \qquad \text{(S1.9)}
\end{aligned}$$

where $\boldsymbol{\lambda}^{(1)*}$ lies in between $\boldsymbol{\lambda}_0^{(1)}$ and $\boldsymbol{\lambda}_0^{(1)}+n^{-1/2}\boldsymbol{u}$ and $\boldsymbol{\lambda}^{(1)*}\xrightarrow{p}\boldsymbol{\lambda}_0^{(1)}$. Here

$$\frac{1}{\sqrt{n}}U_1(\boldsymbol{\lambda}_0^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot))=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau P_1\widetilde{A}_s\left[\boldsymbol{Z}_i-\boldsymbol{E}\left(\bar{a}(s),s\right)\right]dN_i(s), \qquad \text{(S1.10)}$$

where $\bar{a}(s)\equiv a\left(s,\boldsymbol{\lambda}_0^{(1)},\boldsymbol{\lambda}^{(0)};\widetilde{\boldsymbol{m}},\widetilde{\boldsymbol{\beta}}^*(\cdot)\right)$. Define $B_1=\operatorname{diag}\left(P_3\left(\boldsymbol{\lambda}_0^{(1)\top},\boldsymbol{\lambda}^{(0)\top}\right)^\top\right)$ and $B_2=\operatorname{diag}\left(P_4\left(\boldsymbol{\lambda}_0^{(1)\top},\boldsymbol{\lambda}^{(0)\top}\right)^\top\right)$, where P_3 and P_4 are defined in (S0.3) and (S0.4), respectively. The first-order Taylor expansion of (S1.10) around m_0 and $\boldsymbol{\beta}_0^*(s)$ yields

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} U_1(\boldsymbol{\lambda}_0^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)) \\
=&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau P_1 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s\right)\right] dN_i(s) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau P_1 \widetilde{A}(s) V\left(a^*(s), s\right) B_1 \left(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0\right) dN_i(s) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau P_1 \widetilde{A}(s) V\left(a^*(s), s\right) B_2 \left(\widetilde{\boldsymbol{\beta}}^*(s) - \boldsymbol{\beta}_0^*(s)\right) dN_i(s),
\end{aligned}
$$

where $a^*(s) = a\left(s, \boldsymbol{\lambda}_0^{(1)}, \boldsymbol{\lambda}^{(0)}; \overline{\boldsymbol{m}}, \overline{\boldsymbol{\beta}}^*(\cdot)\right)$, $\overline{\boldsymbol{m}}$ lies between $\boldsymbol{m}_0$ and $\widetilde{\boldsymbol{m}}$, $\overline{\boldsymbol{\beta}}^*(s)$ lies between $\boldsymbol{\beta}_0^*(s)$ and $\widetilde{\boldsymbol{\beta}}^*(s)$, and $\boldsymbol{E}(\cdot,\cdot)$, $V(\cdot,\cdot)$ are defined in Section 3.1. Furthermore, we can prove (a)

$$
\sup_{t\in[0,\tau], \boldsymbol{\beta}\in\mathcal{B}} \|S^{(r)}(\boldsymbol{\beta}, t) - s^{(r)}(\boldsymbol{\beta}, t)\| = O_p(n^{-1/2})
$$

where $\mathcal{B}$ is a compact set of $\mathbb{R}^p$ that includes a neighborhood of $\boldsymbol{\beta}_0(t)$ for $t \in [0,\tau]$; (b) $\widetilde{\boldsymbol{\beta}}(t) \xrightarrow{p} \boldsymbol{\beta}_0(t)$ uniformly in the sense that $\sup_{t\in[0,\tau]} \|\widetilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| = o_p(1)$; (c) both $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(0)}$ are bounded. (a) can be justified using the central limit theorem for Banach space (Ledoux and Talagrand, 1991). (b) is proved in Appendix A of Tian et al. (2005). (c) can be justified with the following observations: When $\theta_1 = \theta_2 = 0$ and sample size n goes to infinity, approximately the minimizer of (3.2) are $\boldsymbol{\lambda}_1 = \mathbf{1}$ and $\boldsymbol{\lambda}_2 = \mathbf{1}$. Thus, for all $\theta_1 \geq 0$ and $\theta_2 \geq 0$, $\|\boldsymbol{\lambda}_1\|_1 + \|\boldsymbol{\lambda}_2\|_1 < 2p + 1$ as $n \to \infty$, and $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(0)}$ are bounded. Based on (b) and (c), we can prove that $a^*(s) \xrightarrow{p} \boldsymbol{\beta}_0(s)$ uniformly. Thus, given (a), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} U_1(\boldsymbol{\lambda}_0^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)) \\
\xrightarrow{p}\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau P_1 A_0(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s\right)\right] dN_i(s) \\
&- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau P_1 A_0(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right] dN_i(s) B_1 \sqrt{n} \left(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0\right) \\
&- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau P_1 A_0(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right] B_2 \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^*(s) - \boldsymbol{\beta}_0^*(s)\right) dN_i(s) \\
\triangleq\ & R_1 - R_2 - R_3,
\end{aligned}
$$

where $Q_0(s)$, $Q_1(s)$ and $Q_2(s)$ are defined in Section 3.1.

Next we want to show $R_1 - R_2 - R_3 = O_p(1)$. Using the fact that $Q_2(s)/Q_0(s) - (Q_1(s)/Q_0(s))^{\otimes 2} = O_p(1)$, $A_0(s) = O_p(1)$ and $\sqrt{n}\,(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0) = O_p(1)$, it is easy to verify that $R_2 = O_p(1)$. Further, similarly as Lemma 2 we can show

$$\begin{aligned} R_3 \approx & \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(h_n)^{-1/2}P_1A_0(s)\left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right]B_2\Sigma^{-1}(s) \\ & (nh_n)^{-1/2}\sum_{j=1}^{n}\int_0^{\tau}\left(\boldsymbol{Z}_j(u) - E\left(\boldsymbol{\beta}_0(s), u\right)\right)K\left(\frac{u-s}{h_n}\right)dM_j(u)dN_i(s) \\ \xrightarrow{p} & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^{\tau}\left\{\int_0^{\tau}P_1A_0(s)\left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right]\right. \\ & \left. B_2\Sigma^{-1}(s)\left(\boldsymbol{Z}_j(u) - E\left(\boldsymbol{\beta}_0(s), u\right)\right)\frac{1}{h_n}K\left(\frac{u-s}{h_n}\right)Q_0(s)ds\right\}dM_j(u) \\ = & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^{\tau}P_1A_0(u)\Sigma(u)B_2\Sigma^{-1}(u)\left(\boldsymbol{Z}_j(u) - E\left(\boldsymbol{\beta}_0(u), u\right)\right)dM_j(u), \end{aligned}$$

where the "$\xrightarrow{p}$" sign can be proved using the strong approximation argument (Yandell, 1983). By the martingale central limit theorem (Andersen and Gill, 1982), $R_3 = O_p(1)$. Then, we only need to prove $R_1 = O_p(1)$. Since

$$R_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}P_1A_0(s)\left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s\right)\right]dM_i(s) + o_p(1),$$

it follows from the martingale central limit theorem (Andersen and Gill, 1982) that $R_1 = O_p(1)$. Thus,

$$\frac{1}{\sqrt{n}}U_1(\boldsymbol{\lambda}_0^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)) \xrightarrow{p} R_1 - R_2 - R_3 = O_p(1).$$

Furthermore, we know

$$\begin{aligned} & -\frac{1}{n}H_{11}(\boldsymbol{\lambda}^{(1)*}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}P_1\widetilde{A}(s)V(\ddot{a}(s), s)\left(P_1\widetilde{A}(s)\right)^{\intercal}dN_i(s) \\ \xrightarrow{p} & \int_0^{\tau}P_1A_0(s)\Sigma(s)\left(P_1A_0(s)\right)^{\intercal}ds \triangleq I_{11}, \end{aligned}$$

where $\ddot{a}(s) = a(s, \boldsymbol{\lambda}^{(1)*}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot))$. From the definition of P_1 (S0.1) and $A_0(s)$ (S1.7), we can show that I_{11} is positive definite and $I_{11} = O_p(1)$. Therefore, given $\max(\theta_1, \theta_2)/\sqrt{n}$ is bounded, if we choose a sufficient large C, the second term in (S1.9) is of order C^2. The first and third terms are of order C, which are dominated by the second term. Therefore (S1.8) holds and it completes the proof.

S2 Proof of Theorem 2

For any $\boldsymbol{\lambda}^{(1)}$ satisfy $\|\boldsymbol{\lambda}^{(1)} - \boldsymbol{\lambda}_0^{(1)}\| = O_p(n^{-1/2})$, we have

$$
\begin{aligned}
& Q_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot)) - Q_{2n}(\boldsymbol{\lambda}^{(1)}, \mathbf{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot)) \\
\geq & - U_2^{\intercal}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot))\boldsymbol{\lambda}^{(0)} + \min(\theta_1, \theta_2) \|\boldsymbol{\lambda}^{(0)}\|_1 \\
\geq & \left[- \sup_k \left| U_{2k}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot)) \right| + \min(\theta_1, \theta_2) \right] \|\boldsymbol{\lambda}^{(0)}\|_1, \qquad \text{(S2.1)}
\end{aligned}
$$

where $U_{2k}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot))$ is the kth component of the vector $U_2(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot))$ and $\boldsymbol{\lambda}^{(0)*}$ lies in between $\boldsymbol{\lambda}^{(0)}$ and $\mathbf{0}$. Define $B_3 = \operatorname{diag}\left(P_3\left(\boldsymbol{\lambda}^{(1)^\intercal}, \boldsymbol{\lambda}^{(0)*^\intercal}\right)^\intercal\right)$ and $B_4 = \operatorname{diag}\left(P_4\left(\boldsymbol{\lambda}^{(1)^\intercal}, \boldsymbol{\lambda}^{(0)*^\intercal}\right)^\intercal\right)$. The first order Taylor expansion of $U_2(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot))$ around m_0 and $\boldsymbol{\beta}_0^*(s)$ yields

$$
\begin{aligned}
& \frac{\sqrt{nh_n}}{\sqrt{n}} U_2(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)) \triangleq \frac{\sqrt{nh_n}}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_2 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}(a(s), s)\right] dN_i(s) \\
= & \frac{\sqrt{nh_n}}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_2 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}(a^*(s), s)\right] dN_i(s) \\
& - \frac{\sqrt{nh_n}}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_2 \widetilde{A}(s) V(\overline{a}(s), s) B_3 \left(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0\right) dN_i(s) \\
& - \frac{\sqrt{nh_n}}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_1 \widetilde{A}(s) V(\overline{a}(s), s) B_4 \left(\widetilde{\boldsymbol{\beta}}^*(s) - \boldsymbol{\beta}_0^*(s)\right) dN_i(s) \\
\xrightarrow{p} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \sqrt{nh_n} P_2 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}(a^*(s), s)\right] dN_i(s) \\
& - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \sqrt{nh_n} P_2 \widetilde{A}(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right] dN_i(s) B_3 \sqrt{n} \left(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0\right) \\
& - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \sqrt{nh_n} P_2 \widetilde{A}(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right] B_4 \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^*(s) - \boldsymbol{\beta}_0^*(s)\right) dN_i(s) \\
\triangleq & R_4 - R_5 - R_6,
\end{aligned}
$$

where $a(s) = a\left(s, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)\right)$, $a^*(s) = a\left(s, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \boldsymbol{m}_0, \boldsymbol{\beta}_0^*(\cdot)\right)$, $\overline{a}(s) = a\left(s, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \overline{\boldsymbol{m}}, \overline{\boldsymbol{\beta}}^*(\cdot)\right)$ and $\boldsymbol{E}(\cdot,\cdot)$, $Q_0(\cdot)$, $Q_1(\cdot)$, $Q_2(\cdot)$ are defined in Section 3.1. Here $\overline{\boldsymbol{m}}$ lies in between $\boldsymbol{m}_0$ and $\widetilde{\boldsymbol{m}}$, and $\overline{\boldsymbol{\beta}}^*(s)$ lies in between $\boldsymbol{\beta}_0^*(s)$ and $\widetilde{\boldsymbol{\beta}}^*(s)$. Using the fact that $\sqrt{nh_n}P_2\widetilde{A}(s) = O_p(1)$, $Q_2(s)/Q_0(s) - (Q_1(s)/Q_0(s))^{\otimes 2} = O_p(1)$ and Lemma 1 and 2, it is easy to verify $R_5 = O_p(1)$ and $R_6 = O_p(1)$. For R_4, by the first order Taylor expansion of $\boldsymbol{\lambda}^{(1)}$ around $\boldsymbol{\lambda}_0^{(1)}$, we have

$$
\begin{aligned}
R_4 \equiv & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \sqrt{nh_n} P_2 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}\left(a^*(s), s\right)\right] dN_i(s) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \sqrt{nh_n} P_2 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s\right)\right] dN_i(s) \\
& - \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \sqrt{nh_n} P_2 \widetilde{A}(s) V(\ddot{a}(s), s) B_5(s) dN_i(s) \sqrt{n} \left(\lambda^{(1)} - \lambda_0^{(1)}\right) \\
= & O_p(1) - O_p(1) = O_p(1),
\end{aligned}
$$

where $\ddot{a}(s) = a(s, \boldsymbol{\lambda}^{(1)*}, \boldsymbol{\lambda}^{(0)*}; \boldsymbol{m}_0, \boldsymbol{\beta}_0^*(\cdot))$, $B_5(s) = \mathrm{diag}(\boldsymbol{m}_0)P_{31} + \mathrm{diag}(\boldsymbol{\beta}_0^*(s))P_{41}$, $\boldsymbol{\lambda}^{(1)*}$ lies in between $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}_0^{(1)}$, and P_{31}, P_{41} are defined in (S0.3) and (S0.4), respectively. Thus $\sqrt{h_n}U_2(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)*}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(\cdot)) = O_p(1)$. From (S2.1),

$$
\begin{aligned}
& Q_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot)) - Q_{2n}(\boldsymbol{\lambda}^{(1)}, 0; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}(\cdot)) \\
\geq & \left[-O_p(h_n^{-1/2}) + \min\left(\theta_1, \theta_2\right)\right] \|\boldsymbol{\lambda}^{(0)}\|_1.
\end{aligned}
$$

If $h_n^{1/2} \min(\theta_1, \theta_2) \to \infty$, then with probability one $Q_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(0)}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(s)) - Q_{2n}(\boldsymbol{\lambda}^{(1)}, \boldsymbol{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^*(s)) > 0$ for all $\boldsymbol{\lambda}^{(0)} \neq 0$. It completes the proof.

S3 Proof of Theorem 3

(a) From Theorm 1, $\|\widehat{\boldsymbol{\lambda}}^{(1)} - \boldsymbol{\lambda}_0^{(1)}\| = O_p(n^{-1/2})$ where $\boldsymbol{\lambda}_0^{(1)} = \boldsymbol{1}$. Thus $P(\widehat{\boldsymbol{\lambda}}^{(1)}[k] \neq 0) \to 1$ for all k, where $\widehat{\boldsymbol{\lambda}}^{(1)}[k]$ is the k-th component of $\widehat{\boldsymbol{\lambda}}^{(1)}$. This in couple with $P(\widehat{\boldsymbol{\lambda}}^{(0)} = \boldsymbol{0}) \to 1$ from Theorem 2 proves $\widehat{\mathrm{I}}_O = \mathrm{I}_O$, $\widehat{\mathrm{I}}_C = \mathrm{I}_C$ and $\widehat{\mathrm{I}}_{NC} = \mathrm{I}_{NC}$ hold with probability tending to one. It completes the proof for part (a).

(b) From part (a), we know with probability tending to one, $\widehat{\boldsymbol{\beta}}_C(t) \equiv \widehat{\boldsymbol{\beta}}_C$. Thus,

$$
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_C\right) = \sqrt{n}\left(\widetilde{\boldsymbol{m}}_C \circ \widehat{\boldsymbol{\lambda}}_1^C - \boldsymbol{m}_C\right) = \widetilde{\boldsymbol{m}}_C \circ \sqrt{n}\left(\widehat{\boldsymbol{\lambda}}_1^C - \boldsymbol{\lambda}_{01}^C\right) + \sqrt{n}\left(\widetilde{\boldsymbol{m}}_C - \boldsymbol{m}_C\right),
$$

where $\widetilde{\boldsymbol{m}}_C$, $\widehat{\boldsymbol{\lambda}}_1^C$, $\boldsymbol{m}_C$ are sub-vectors of $\widetilde{\boldsymbol{m}}$, $\widehat{\boldsymbol{\lambda}}_1$, $\boldsymbol{m}_0$ with indexes in I_C. Since $\widehat{\boldsymbol{\lambda}}_1^C$ and $\widetilde{\boldsymbol{m}}_C$ are root-n consistent estimator for $\boldsymbol{\lambda}_{01}^C$ and $\boldsymbol{m}_C$ respectively. We have $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_C\right) = O_p(1)$. It completes the proof for part (b).

(c) From the proofs of part (a), and Theorem 1 and 2, it is easy to show that $\widehat{\boldsymbol{\lambda}}^{(1)}$ is a root-n consistent maximizer of $Q_{2n}(\boldsymbol{\lambda}^{(1)}, \mathbf{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^{*}(\cdot))$. Thus

$$\frac{\partial Q_{2n}(\boldsymbol{\lambda}^{(1)}, \mathbf{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^{*}(\cdot))}{\partial \boldsymbol{\lambda}^{(1)}}\bigg|_{\boldsymbol{\lambda}^{(1)}=\widehat{\boldsymbol{\lambda}}^{(1)}} = \mathbf{0}.$$

This is equivalent to

$$-\frac{1}{\sqrt{n}}U_1(\widehat{\boldsymbol{\lambda}}^{(1)}, \mathbf{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^{*}(\cdot)) + \frac{\overline{\boldsymbol{\theta}}}{\sqrt{n}} = \mathbf{0}, \tag{S3.1}$$

where $\overline{\boldsymbol{\theta}}$ is a vector of length $(p_2 + 2p_3)$ consisting with elements θ_1 and θ_2.

The first order Taylor expansion of $n^{-1/2}U_1(\widehat{\boldsymbol{\lambda}}^{(1)}, \mathbf{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^{*}(\cdot))$ around $\boldsymbol{m}_0$ and $\boldsymbol{\beta}_0^*(\cdot)$ yields

$$\begin{aligned}
&\frac{1}{\sqrt{n}}U_1(\widehat{\boldsymbol{\lambda}}^{(1)}, \mathbf{0}; \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\beta}}^{*}(\cdot)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau} P_1\widetilde{A}(s)\left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\widetilde{a}(s), s\right)\right]dN_i(s)\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau} P_1\widetilde{A}(s)V\left(a^*(s), s\right)\widetilde{B}_6\left(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0\right)dN_i(s)\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau} P_1\widetilde{A}(s)V\left(a^*(s), s\right)\widetilde{B}_7\left(\widetilde{\boldsymbol{\beta}}^{*}(s) - \boldsymbol{\beta}_0^*(s)\right)dN_i(s)\\
&\triangleq R_7 - R_8 - R_9,
\end{aligned} \tag{S3.2}$$

where $\widetilde{a}(s) = a(s, \widehat{\boldsymbol{\lambda}}^{(1)}, \mathbf{0}; \boldsymbol{m}_0, \boldsymbol{\beta}_0^*(\cdot))$, $a^*(s) = a(s, \widehat{\boldsymbol{\lambda}}^{(1)}, \mathbf{0}; \overline{\boldsymbol{m}}, \overline{\boldsymbol{\beta}}^{*}(\cdot))$, $\widetilde{B}_6 = \text{diag}\left(P_3\left(\left(\widehat{\boldsymbol{\lambda}}^{(1)}\right)^{\intercal}, \mathbf{0}^{\intercal}\right)^{\intercal}\right)$ and $\widetilde{B}_7 = \text{diag}\left(P_4\left(\left(\widehat{\boldsymbol{\lambda}}^{(1)}\right)^{\intercal}, \mathbf{0}^{\intercal}\right)^{\intercal}\right)$. Using similar arguments for the proofs of Theorem 1 and 2, we can show

$$\begin{aligned}
R_8 &\xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} P_1A_0(s)\left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)}\right)^{\otimes 2}\right]dN_i(s)B_6\sqrt{n}\left(\widetilde{\boldsymbol{m}} - \boldsymbol{m}_0\right)\\
&\xrightarrow{p} \frac{1}{\tau}\int_0^{\tau} P_1A_0(s)\Sigma(s)dsB_6\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\Sigma^{-1}(u)\left(\boldsymbol{Z}_i(u) - \boldsymbol{E}\left(\boldsymbol{\beta}_0(u), u\right)\right)dM_i(u)\right],
\end{aligned} \tag{S3.3}$$

and

$$\begin{aligned}
R_9 \xrightarrow{p} & \frac{1}{nh_n^{1/2}} \sum_{i=1}^n \int_0^\tau P_1 A_0(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)} \right)^{\otimes 2} \right] B_7 \sqrt{nh_n} \left(\widetilde{\boldsymbol{\beta}}^*(s) - \boldsymbol{\beta}_0^*(s) \right) dN_i(s) \\
\approx & \frac{1}{nh_n^{1/2}} \sum_{i=1}^n \int_0^\tau P_1 A_0(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)} \right)^{\otimes 2} \right] B_7 \\
& \Sigma^{-1}(s)(nh_n)^{-1/2} \sum_{j=1}^n \int_0^\tau \left(\boldsymbol{Z}_j(u) - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), u \right) \right) K\left(\frac{u-s}{h_n} \right) dM_j(u) dN_i(s) \\
= & \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \int_0^\tau P_1 A_0(s) \left[\frac{Q_2(s)}{Q_0(s)} - \left(\frac{Q_1(s)}{Q_0(s)} \right)^{\otimes 2} \right] B_7 \Sigma^{-1}(s) \right. \\
& \left. \left(\boldsymbol{Z}_j(u) - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), u \right) \right) \frac{1}{h_n} K\left(\frac{u-s}{h_n} \right) dN_i(s) \right\} dM_j(u) \\
\xrightarrow{p} & \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^\tau P_1 A_0(u) \Sigma(u) B_7 \Sigma^{-1}(u) \left(\boldsymbol{Z}_j(u) - \boldsymbol{E}\left(\boldsymbol{\beta}_0(u), u \right) \right) dM_j(u), \qquad \text{(S3.4)}
\end{aligned}$$

where

$$B_6 = \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times (p_2+p_3)} \\ \mathbf{0}_{(p_2+p_3) \times p_1} & \boldsymbol{I}_{p_2+p_3} \end{pmatrix}; B_7 = \begin{pmatrix} \mathbf{0}_{(p_1+p_2) \times (p_1+p_2)} & \mathbf{0}_{(p_1+p_2) \times p_3} \\ \mathbf{0}_{p_3 \times (p_1+p_2)} & \boldsymbol{I}_{p_3} \end{pmatrix}.$$

Here B_6 and B_7 are the limits of $\widetilde{B}_6$ and $\widetilde{B}_7$, respectively. The first order Taylor expansion of R_7 around $\boldsymbol{\lambda}_0^{(1)}$ yields

$$\begin{aligned}
R_7 = & \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_1 \widetilde{A}(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s \right) \right] dN_i(s) \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_1 \widetilde{A}(s) V\left(\ddot{a}(s), s \right) B_8(s) dN_i(s) \sqrt{n} (\widehat{\boldsymbol{\lambda}}^{(1)} - \boldsymbol{\lambda}_0^{(1)}) \\
\xrightarrow{p} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau P_1 A_0(s) \left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s \right) \right] dM_i(s) - F \sqrt{n} (\widehat{\boldsymbol{\lambda}}^{(1)} - \boldsymbol{\lambda}_0^{(1)}), \qquad \text{(S3.5)}
\end{aligned}$$

where

$$B_8(s) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \operatorname{diag}(\boldsymbol{m}_C) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \operatorname{diag}(\boldsymbol{m}_{NC}) & \operatorname{diag}(\boldsymbol{\beta}_{NC}^*(s)) \end{pmatrix}$$

and

$$F = \int_0^\tau P_1 A_0(s) \Sigma(s) B_8(s) ds.$$

Here $\ddot{a}(s) = a(s, \boldsymbol{\lambda}^{(1)*}, \mathbf{0}; \boldsymbol{m}_0, \boldsymbol{\beta}_0^*(\cdot))$ and $\boldsymbol{\lambda}^{(1)*}$ lies in between $\boldsymbol{\lambda}_0^{(1)}$ and $\widehat{\boldsymbol{\lambda}}^{(1)}$.

Under the assumption $\max(\theta_1, \theta_2)/\sqrt{n} \to 0$, combining (S3.1), (S3.2), (S3.3), (S3.4), (S3.5), we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}^{(1)} - \boldsymbol{\lambda}_0^{(1)}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau} F^{-1}G(u)\left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(s), s\right)\right]dM_i(s) + o_p(1), \quad \text{(S3.6)}$$

where

$$G(u) = P_1A_0(u) - \frac{1}{\tau}\left(\int_0^{\tau} P_1A_0(s)\Sigma(s)ds\right)B_6\Sigma^{-1}(u) - P_1A_0(u)\Sigma(u)B_7\Sigma^{-1}(u).$$

Because

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_C\right) = \widetilde{\boldsymbol{m}}_C \circ \sqrt{n}\left(\widehat{\boldsymbol{\lambda}}_1^C - \boldsymbol{\lambda}_{01}^C\right) + n^{1/2}\left(\widetilde{\boldsymbol{m}}_C - \boldsymbol{m}_C\right)\\
=&\mathrm{diag}\left(\boldsymbol{m}_C\right)B_9\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}^{(1)} - \boldsymbol{\lambda}_0^{(1)}\right)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\frac{1}{\tau}B_{10}\Sigma^{-1}(u)\left[\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(u), u\right)\right]dM_i(u) + o_p(1), \qquad \text{(S3.7)}
\end{aligned}$$

where $B_9 = (\boldsymbol{I}_{p_2}|\boldsymbol{0}_{p_2\times p_3}|\boldsymbol{0}_{p_2\times p_3})$ and $B_{10} = (\boldsymbol{0}_{p_2\times p_1}|\boldsymbol{I}_{p_2}|\boldsymbol{0}_{p_2\times p_3})$. Combining (S3.6) and (S3.7), we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_C\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left[D(u) + \frac{1}{\tau}B_{10}\Sigma^{-1}(u)\right]\left(\boldsymbol{Z}_i - \boldsymbol{E}\left(\boldsymbol{\beta}_0(u), u\right)\right)dM_i(u) + o_p(1),$$

where

$$D(u) = \mathrm{diag}\left(\boldsymbol{m}_C\right)B_9F^{-1}G(u). \qquad \text{(S3.8)}$$

Finally by the martingale central limit theorem (Andersen and Gill, 1982), $n^{1/2}\left(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_C\right)$ converges weakly to a normal distribution with mean 0 and variance Σ_m^F, where

$$\Sigma_m^F = \int_0^{\tau}\left(D(u) + \frac{1}{\tau}B_{10}\Sigma^{-1}(u)\right)\Sigma(u)\left(D(u) + \frac{1}{\tau}B_{10}\Sigma^{-1}(u)\right)^{\intercal}du.$$

(d) Because

$$\begin{aligned}
&(nh_n)^{1/2}\left(\widehat{\boldsymbol{\beta}}_{NC}(t) - \boldsymbol{\beta}_{NC}(t)\right) = (nh_n)^{1/2}\left(\widetilde{\boldsymbol{m}}_{NC} \circ \widehat{\boldsymbol{\lambda}}_1^{NC} + \widetilde{\boldsymbol{\beta}}_{NC}^* \circ \widehat{\boldsymbol{\lambda}}_2^{NC} - \boldsymbol{\beta}_{NC}(t)\right)\\
=&(nh_n)^{1/2}\left[\widetilde{\boldsymbol{m}}_{NC} \circ \left(\widehat{\boldsymbol{\lambda}}_1^{NC} - \boldsymbol{\lambda}_{01}^{NC}\right) + \widetilde{\boldsymbol{\beta}}_{NC}^*(t) \circ \left(\widehat{\boldsymbol{\lambda}}_2^{NC} - \boldsymbol{\lambda}_{02}^{NC}\right) + \left(\widetilde{\boldsymbol{\beta}}_{NC}(t) - \boldsymbol{\beta}_{NC}(t)\right)\right]\\
=&(nh_n)^{1/2}\left(\widetilde{\boldsymbol{\beta}}_{NC}(t) - \boldsymbol{\beta}_{NC}(t)\right) + o_p(1)\\
\xrightarrow{d}& N\left\{0,\ \{\Sigma^{-1}(t)\}_{NC,NC}\int_{-1}^{1}K^2(s)ds\right\},
\end{aligned}$$

where $\{\Sigma^{-1}(t)\}_{NC,NC}$ is the submatrix of $\Sigma^{-1}(t)$ corresponding to I_{NC}. It completes the proof for part (d).

Additional References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist. 10*(4), 1100–1120.

Ledoux, M. and Talagrand, M. (1991). Probability in Banach Spaces: Isoperimetry and Processes. New York: Springer-Verlag Press.

Tian, L., Zucker, D., and Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *J. Amer. Statist. Assoc. 100*(469), 172–183.

Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Statist. 11*(4), 1119–1135.