Time-varying Hazards Model for Incorporating Irregularly Measured High-Dimensional Biomarkers

Xiang Li[1], Quefeng Li[2], Donglin Zeng[3],

Karen Marder[4], Jane Paulsen[5], and Yuanjia Wang[6]

[1,4,6] *Columbia University*

[2,3] *University of North Carolina, Chapel Hill*

[5] *University of Iowa*

Supplementary Material

The Supplementary Material includes proofs of Lemma 1 and Theorem 1, and additional information related to the simulation and real-data analysis.

S1 Proof of equivalence between (6) and (7)

We prove that if the global minimizers of (6) and (7) are unique, they are equivalent in the sense that if $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ solves (7) for ϕ_n, there exists a c_n such that $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ also solves (6) for c_n; and vice versa.

First, we prove that if $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ is the global minimizer of (7), it also solves (6) with $c_n = \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j - \hat{\boldsymbol{\theta}}_j\|_2$. Denote $L(\boldsymbol{\gamma}, \boldsymbol{\theta}) = -l_n(\boldsymbol{\gamma}) + p(\boldsymbol{\theta}; \nu_n)$. Suppose there exists $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}})$ different from $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ such that

$$L(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}}) < L(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}) \text{ and } \sum_{j=1}^{p_n} ||\tilde{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\theta}}_j||_2 \leq c_n.$$

Then, by definition,

$$L(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}}) + \phi_n \sqrt{q_n} \sum_{j=1}^{p_n} ||\tilde{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\theta}}_j||_2 < L(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}) + \phi_n \sqrt{q_n} \sum_{j=1}^{p_n} ||\hat{\boldsymbol{\gamma}}_j - \hat{\boldsymbol{\theta}}_j||_2,$$

which contradicts with the fact that $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ is the minimizer of (7).

Next, we prove that, for any given c_n, if $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}})$ is the solution to (6), we can always find a ϕ_n such that $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}})$ also solves (7). Suppose $(\check{\boldsymbol{\gamma}}, \check{\boldsymbol{\theta}}) = \arg\min_{\boldsymbol{\gamma}, \boldsymbol{\theta}} L(\boldsymbol{\gamma}, \boldsymbol{\theta})$ is the minimizer of the unconstrained problem. Let $C_{\max} = \sum_{j=1}^{p_n} \|\check{\boldsymbol{\gamma}}_j - \check{\boldsymbol{\theta}}_j\|_2$. Then, for any $c_n \geq C_{\max}$, $(\check{\boldsymbol{\gamma}}, \check{\boldsymbol{\theta}})$ is also the solution to (6). In this case, it's easy to check that $(\check{\boldsymbol{\gamma}}, \check{\boldsymbol{\theta}})$ also solves (7) with $\phi_n = 0$. For $c_n < C_{\max}$, suppose the solution to (6) is given by $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}})$. Let $C_{\phi_n} = \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_n} - \hat{\boldsymbol{\theta}}_j^{\phi_n}\|_2$, where $(\hat{\boldsymbol{\gamma}}^{\phi_n}, \hat{\boldsymbol{\theta}}^{\phi_n})$ is the solution to (7) for ϕ_n. We prove that C_{ϕ_n} is a decreasing function of ϕ_n. In fact, suppose $(\hat{\boldsymbol{\gamma}}^{\phi_1}, \hat{\boldsymbol{\theta}}^{\phi_1})$ and $(\hat{\boldsymbol{\gamma}}^{\phi_2}, \hat{\boldsymbol{\theta}}^{\phi_2})$ are solutions to (7) for ϕ_1 and ϕ_2 respectively and

$\phi_1 < \phi_2$. By definition,

$$
\begin{aligned}
& L(\hat{\boldsymbol{\gamma}}^{\phi_2}, \hat{\boldsymbol{\theta}}^{\phi_2}) + \phi_2 \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_2} - \hat{\boldsymbol{\theta}}_j^{\phi_2}\|_2 \\
\leq\ & L(\hat{\boldsymbol{\gamma}}^{\phi_1}, \hat{\boldsymbol{\theta}}^{\phi_1}) + \phi_2 \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_1} - \hat{\boldsymbol{\theta}}_j^{\phi_1}\|_2 \\
=\ & L(\hat{\boldsymbol{\gamma}}^{\phi_1}, \hat{\boldsymbol{\theta}}^{\phi_1}) + \phi_1 \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_1} - \hat{\boldsymbol{\theta}}_j^{\phi_1}\|_2 + (\phi_2 - \phi_1) \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_1} - \hat{\boldsymbol{\theta}}_j^{\phi_1}\|_2 \\
\leq\ & L(\hat{\boldsymbol{\gamma}}^{\phi_2}, \hat{\boldsymbol{\theta}}^{\phi_2}) + \phi_1 \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_2} - \hat{\boldsymbol{\theta}}_j^{\phi_2}\|_2 + (\phi_2 - \phi_1) \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_1} - \hat{\boldsymbol{\theta}}_j^{\phi_1}\|_2
\end{aligned}
$$

Therefore, $C_{\phi_2} = \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_2} - \hat{\boldsymbol{\theta}}_j^{\phi_2}\|_2 \leq \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j^{\phi_1} - \hat{\boldsymbol{\theta}}_j^{\phi_1}\|_2 = C_{\phi_1}$. Then, by the continuity of the objective function in (7) and the uniqueness of the global minimizer, for every $c_n < C_{\max}$, we can always find a ϕ_n such that $c_n = C_{\phi_n}$. We prove that $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}})$ solves (7) with such a ϕ_n. Otherwise, let $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ be the solution. Then,

$$
L(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}) + \phi_n \sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j - \hat{\boldsymbol{\theta}}_j\|_2 < L(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}}) + \phi_n \sum_{j=1}^{p_n} \|\tilde{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\theta}}_j\|_2.
$$

By definition, $\sum_{j=1}^{p_n} \|\hat{\boldsymbol{\gamma}}_j - \hat{\boldsymbol{\theta}}_j\|_2 = c_n$. Therefore,

$$
L(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}) < L(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}}) + \phi_n (\sum_{j=1}^{p_n} \|\tilde{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\theta}}_j\|_2 - c_n) \leq L(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}}).
$$

This contradicts with the assumption that $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\theta}})$ is the global minimizer of

(6).

S2 Proof of Lemma 1

As discussed in Remark 2, all following arguments are conditioned on the event $\{n_i \le M_\epsilon\}$, which has probability at least $1-\epsilon$ to hold. We have

$$
\begin{aligned}
U_{n,j}(\boldsymbol{\gamma}^*) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})\{Z_{ij}(t_{iv},t)-E_{nj}(\boldsymbol{\gamma}^*,t)\}d\Lambda_i(t)\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})\{Z_{ij}(t_{iv},t)-E_{nj}(\boldsymbol{\gamma}^*,t)\}dM_i(t)\\
&:= I_1 + I_2.
\end{aligned}
$$

The upper bound of I_1 will be given in Lemma S1 in Section S4. For I_2, we have

$$
\begin{aligned}
I_2 &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})\{Z_{ij}(t_{iv},t)-e_{nj}(\boldsymbol{\gamma}^*,t)\}dM_i(t)\\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})\{E_{nj}(\boldsymbol{\gamma}^*,t)-e_{nj}(\boldsymbol{\gamma}^*,t)\}dM_i(t)\\
&:= J_{1n} - J_{2n}.
\end{aligned}
$$

To bound J_{1n}, since nJ_{1n} is the sum of i.i.d random variables with mean zero, which are bounded by $O(h_n^{-1})$, it follows from the Hoeffding inequality that

$$P(|(nh_n^2)^{1/2}J_{1n}| > x) \leq 2\exp(-Cx^2). \tag{S2.1}$$

To bound J_{2n}, consider the event $A = A_1 \cap A_2$, where

$$A_1 := \left\{ \sup_{t\in[0,\tau]} |S_n^{(0)}(\boldsymbol{\gamma}^*, t) - s_n^{(0)}(\boldsymbol{\gamma}^*, t)| \leq D(r_n q_n c_n d_n^{1/2}/n)^{1/2} \right\},$$
$$A_2 := \left\{ \sup_{t\in[0,\tau]} |S_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t) - s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)| \leq D(r_n q_n c_n d_n^{1/2}/n)^{1/2} \right\}.$$

By Lemma S2 in Section S4, $P(A) \geq 1 - 2\exp(-Cr_n q_n c_n d_n^{1/2} h_n^2 x^2)$. Conditioning on A, we show that

$$\sup_{t\in[0,\tau]} |E_{nj}(\boldsymbol{\gamma}^*, t) - e_{nj}(\boldsymbol{\gamma}^*, t)| = o(1). \tag{S2.2}$$

In fact, we have

$$\begin{aligned} E_{nj}(\boldsymbol{\gamma}^*, t) - e_{nj}(\boldsymbol{\gamma}^*, t) &= \frac{S_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)}{S_n^{(0)}(\boldsymbol{\gamma}^*, t)} - \frac{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)}{s_n^{(0)}(\boldsymbol{\gamma}^*, t)} \\ = \frac{1}{S_n^{(0)}(\boldsymbol{\gamma}^*, t)} \{S_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t) - s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)\} & \\ + \frac{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)}{S_n^{(0)}(\boldsymbol{\gamma}^*, t) s_n^{(0)}(\boldsymbol{\gamma}^*, t)} \{S_n^{(0)}(\boldsymbol{\gamma}^*, t) - s_n^{(0)}(\boldsymbol{\gamma}^*, t)\}. & \end{aligned}$$

Then, conditioning on A, condition 8 implies (S2.2).

Let $\bar{M}(t) = \sum_{i=1}^{n} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) M_i(t)$. Since $M_i(t) = N_i(t) - \Lambda_i(t)$

is a martingale with compensator

$$\Lambda_i(t) = \int_0^t Y_i(u) \exp[\{\boldsymbol{\beta}^*(u)\}^T \boldsymbol{X}_i(u)]\lambda_0(u)du,$$

so $\bar{M}(t)$ is also a martingale. We have $|\Delta(\bar{M}(t))| = O(h_n^{-1})$. Next, we show that both $\Delta((nh_n^2)^{1/2}J_{2n}(t))$ and $\langle(nh_n^2)^{1/2}J_{2n}(t)\rangle$ are bounded. For $\Delta((nh_n^2)^{1/2}J_{2n}(t))$, we have

$$\begin{aligned}\Delta((nh_n^2)^{1/2}J_{2n}(t)) &\lesssim (nh_n^2)^{-1/2}\left(\sup_{t\in[0,\tau]}|E_{nj}(\boldsymbol{\gamma}^*,t) - e_{nj}(\boldsymbol{\gamma}^*,t)|\right) \lesssim (nh_n^2)^{-1/2} \\ &= O(1),\end{aligned}$$

where condition 7 and the fact that $|\Delta(\bar{M}(t))| = O(h_n^{-1})$ are used. Next, we calculate the predictable quadratic variation of $(nh_n^2)^{1/2}J_{2n}$, denoted by $\langle(nh_n^2)^{1/2}J_{2n}\rangle$,

$$\begin{aligned}\langle(nh_n^2)^{1/2}J_{2n}(t)\rangle &= n^{-1}h_n^2\int_0^t \{E_{ij}(t_{iv},u) - e_{nj}(\boldsymbol{\gamma}^*,u)\}^2 d\langle\bar{M}(u)\rangle \\ &\leq h_n^2\left[\sup_{t\in[0,\tau]}\{E_{ij}(t_{iv},u) - e_{nj}(\boldsymbol{\gamma}^*,u)\}\right]^2 \int_0^t S_n^{(0)}(\boldsymbol{\beta}^*,u)d\Lambda_0(u) \\ &= O(1),\end{aligned}$$

where the last equality follows from (S2.2), condition 1 and the fact that $\sup_{t\in[0,\tau]}|S_n^{(0)}(\boldsymbol{\beta}^*,t)| \lesssim h_n^{-1}$. Then, it follows from Lemma 2.1 of van de

Geer (1995) that

$$P\{|(nh_n^2)^{1/2}J_{2n}| > x|A\} \leq C_3\exp(-C_4x). \tag{S2.3}$$

(S2.1), (S2.3) and Lemma S2 in Section S4 together imply that

$$\begin{aligned} &P\{|I_2| \leq D(nh_n^2)^{-1/2}x\} \\ &\geq 1 - P\{|(nh_n^2)^{1/2}J_{1n}| > 0.5Dx\} - P\{|(nh_n^2)^{1/2}J_{2n}| > 0.5Dx|A\} - P(A^c) \\ &\geq 1 - C_1\exp(-C_2x^2) - C_3\exp(-C_4x). \end{aligned}$$

This result together with Lemma S1 prove the result after dropping high order terms.

S3 Proof of Theorem 1

We prove the following two results:

[1] $\{j : \hat{\boldsymbol{\gamma}}_j \neq 0\} = \{j : \boldsymbol{\gamma}_j^* \neq 0\}$.

[2] $\max_{j_l \in \mathcal{A}} |\hat{\gamma}_{j_l} - \gamma_{j_l}^*| \leq M\nu_n\sqrt{q_n}$.

Then, [1] implies $[a]$. [2] together with condition 6 imply $[b]$.

By optimization theory (Boyd and Vandenberghe, 2004), any vector $\boldsymbol{\gamma}$

satisfies the following KKT conditions is a solution to (5):

$$\boldsymbol{U}_{n,j}(\boldsymbol{\gamma}) = \nu_n \sqrt{q_n} \rho'(\|\boldsymbol{\gamma}_j\|_2) \|\boldsymbol{\gamma}_j\|_2^{-1} \boldsymbol{\gamma}_j, \text{ if } \boldsymbol{\gamma}_j \neq \mathbf{0}, \tag{S3.1}$$

$$\|\boldsymbol{U}_{n,j}(\boldsymbol{\gamma})\|_\infty < \nu_n \sqrt{q_n} \rho'(0+), \text{ if } \boldsymbol{\gamma}_j = \mathbf{0}, \tag{S3.2}$$

$$\lambda_{\min}(\boldsymbol{I}_{n,\hat{\mathcal{A}}\hat{\mathcal{A}}}(\boldsymbol{\gamma})) > \nu_n \kappa(\rho, \boldsymbol{\gamma}), \tag{S3.3}$$

where $\hat{\mathcal{A}} := \{j_l : \boldsymbol{\gamma}_j \neq \mathbf{0} \text{ and } 1 \leq l \leq q_n\}$.

We define the event A as

$$\begin{aligned} A = &\{n_i \leq M_\epsilon\} \cap \{\|\boldsymbol{U}_n(\boldsymbol{\gamma}^*)\|_\infty \leq \nu_n \sqrt{q_n} \rho'(0+)/2\} \\ &\cap \left\{ \inf_{\boldsymbol{\gamma} \in \mathcal{B}_0} : \lambda_{\min}(\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})) > C_{\min}/2 \right\} \\ &\cap \left\{ \sup_{\boldsymbol{\gamma} \in \mathcal{B}_0} \|\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma}) \boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\|_\infty < \frac{1}{2}(1-\zeta) \frac{\rho'(0+)}{\rho'(d_n/2)} \right\}. \end{aligned}$$

By Lemmas 1, S5 in Section S4, and the union bound,

$$\begin{aligned} P(A) \geq 1 - \epsilon &- C_1 p_n q_n \exp\{-C_2 n^2 h_n^8 (\nu_n \sqrt{q_n} - \pi_n)^2\} \\ &- C_3 p_n q_n \exp\{-C_4 (n h_n^2)^{1/2} (\nu_n \sqrt{q_n} - \pi_n)\} \\ &- C_5 p_n r_n q_n^2 \exp\{-C_6 n h_n^2 (r_n q_n)^{-1}\}. \end{aligned}$$

Next, we show that conditioning on event A, statements [1] and [2] hold.

[1] Let $\mathcal{N}$ denote the hypercube $\{\boldsymbol{\gamma}_{\mathcal{A}} \in \mathcal{R}^{r_n q_n} : \|\boldsymbol{\gamma}_{\mathcal{A}} - \boldsymbol{\gamma}^*_{\mathcal{A}}\|_\infty \leq M \nu_n \sqrt{q_n}\}$,

where M is a sufficiently large constant. We show that within $\mathcal{N}$, there exists a solution $\hat{\boldsymbol{\gamma}}_{\mathcal{A}}$ to equation (S3.1). We define a function $f : \mathcal{R}^{r_n q_n} \to \mathcal{R}^{r_n q_n}$ as

$$f(\boldsymbol{\gamma}_{\mathcal{A}}) = \boldsymbol{\gamma}_{\mathcal{A}} + 2\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\{\boldsymbol{U}_{n,\mathcal{A}}(\boldsymbol{\gamma}) - \nabla_{\mathcal{A}} p_{\nu_n}(\boldsymbol{\gamma})\}, \tag{S3.4}$$

where $\boldsymbol{\gamma} \in \mathcal{R}^{p_n q_n}$ such that $\boldsymbol{\gamma}_{\mathcal{A}^c} = \mathbf{0}$, $\nabla_{\mathcal{A}} p_{\nu_n}(\boldsymbol{\gamma}) := \nu_n \sqrt{q_n} \rho'(\|\boldsymbol{\gamma}_j\|_2)\|\boldsymbol{\gamma}_j\|_2^{-1}\boldsymbol{\gamma}_j$. By the Taylor expansion,

$$\boldsymbol{U}_{n,\mathcal{A}}(\boldsymbol{\gamma}) = \boldsymbol{U}_{n,\mathcal{A}}(\boldsymbol{\gamma}^*) - \frac{1}{2}\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\bar{\boldsymbol{\gamma}})(\boldsymbol{\gamma}_{\mathcal{A}} - \boldsymbol{\gamma}^*_{\mathcal{A}}),$$

where $\bar{\boldsymbol{\gamma}}$ lies on the line segment connecting $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^*$. Substituting it into (S3.4) gives

$$\begin{aligned} f(\boldsymbol{\gamma}_{\mathcal{A}}) - \boldsymbol{\gamma}^*_{\mathcal{A}} &= \{\mathcal{I}_{r_n q_n} - \boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\bar{\boldsymbol{\gamma}})\}(\boldsymbol{\gamma}_{\mathcal{A}} - \boldsymbol{\gamma}^*_{\mathcal{A}}) \\ &\quad + 2\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\{\boldsymbol{U}_{n,\mathcal{A}}(\boldsymbol{\gamma}^*_{\mathcal{A}}) - \nabla_{\mathcal{A}} p_{\nu_n}(\boldsymbol{\gamma})\}, \end{aligned}$$

where $\mathcal{I}_{r_n q_n}$ is a $r_n q_n \times r_n q_n$ identity matrix. Without loss of generality, we assume

$$\|\mathcal{I}_{r_n q_n} - \boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\bar{\boldsymbol{\gamma}})\|_\infty \leq 1/2. \tag{S3.5}$$

Moreover, since $d_n \geq 2\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*\|_\infty$, it follows that

$$\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*\|_2 \leq \sqrt{q_n}\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*\|_\infty \leq d_n/2.$$

Hence,

$$\|\boldsymbol{\gamma}_j\|_2 \geq \|\boldsymbol{\gamma}_j^*\|_2 - \|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*\|_2 \geq d_n/2.$$

By the concavity assumption of $\rho(t)$, we have $\rho'(\|\boldsymbol{\gamma}_j\|_2) \leq \rho'(d_n/2)$. Therefore,

$$\|\nabla_{\mathcal{A}} p_{\nu_n}(\boldsymbol{\gamma})\|_\infty \leq \nu_n \sqrt{q_n}\rho'(d_n/2).$$

Then, we obtain

$$\begin{aligned}
&\|f(\boldsymbol{\gamma}) - \boldsymbol{\gamma}_{\mathcal{A}}^*\|_\infty \\
&\leq 1/2\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_{\mathcal{A}}^*\|_\infty + 2\|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\|_\infty\{\|\boldsymbol{U}_{n,\mathcal{A}}(\boldsymbol{\gamma}_{\mathcal{A}}^*)\|_\infty + \|\nabla_{\mathcal{A}} p_{\nu_n}(\boldsymbol{\gamma})\|_\infty\} \\
&\leq \frac{1}{2}M\nu_n\sqrt{q_n} + \frac{4}{C_{\min}}\left\{\frac{\rho'(0+)}{2}\nu_n\sqrt{q_n} + \nu_n\sqrt{q_n}\rho'(d_n/2)\right\} \\
&\overset{(i)}{\leq} \left(\frac{M}{2} + \frac{6\rho'(0+)}{C_{\min}}\right)\nu_n\sqrt{q_n} \\
&\leq M\rho'(0+)\nu_n\sqrt{q_n},
\end{aligned}$$

where in (i), we use the fact that $\rho'(d_n/2) \leq \rho'(0+)$ due to the concavity assumption in condition 11.

Therefore, $f(\mathcal{N}) \subset \mathcal{N}$. It follows from the definition of d_n that $\text{sign}(\boldsymbol{\gamma}_{\mathcal{A}}) =$

sign($\boldsymbol{\gamma}^*_{\mathcal{A}}$) for any $\boldsymbol{\gamma}_{\mathcal{A}} \in \mathcal{N}$. Therefore, $f(\boldsymbol{\gamma}_{\mathcal{A}})$ is a continuous function on the convex and compact set $\mathcal{N}$. By Brouwer's fixed point theorem, there exists a solution $\hat{\boldsymbol{\gamma}}_{\mathcal{A}} \in \mathcal{N}$ to the problem $f(\boldsymbol{\gamma}_{\mathcal{A}}) = \boldsymbol{\gamma}_{\mathcal{A}}$, which also solves (S3.1).

[2] We expand $\hat{\boldsymbol{\gamma}}_{\mathcal{A}}$ to be $\hat{\boldsymbol{\gamma}} \in \mathcal{R}^{p_n q_n}$ such that $\hat{\boldsymbol{\gamma}}_{\mathcal{A}^c} = \mathbf{0}$. We further show that $\hat{\boldsymbol{\gamma}}$ satisfies (S3.2). Again, by the Taylor expansion of $\boldsymbol{U}_{n,\mathcal{A}^c}(\hat{\boldsymbol{\gamma}})$ around $\boldsymbol{\gamma}^*$, we have

$$\boldsymbol{U}_{n,\mathcal{A}^c}(\hat{\boldsymbol{\gamma}}) = \boldsymbol{U}_{n,\mathcal{A}^c}(\boldsymbol{\gamma}^*) - \frac{1}{2}\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\tilde{\boldsymbol{\gamma}})(\hat{\boldsymbol{\gamma}}_{\mathcal{A}} - \boldsymbol{\gamma}^*_{\mathcal{A}}), \tag{S3.6}$$

where $\tilde{\boldsymbol{\gamma}}$ lies on the line segment connecting $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}^*$. Since $f(\hat{\boldsymbol{\gamma}}_{\mathcal{A}}) = 0$, it holds that

$$\hat{\boldsymbol{\gamma}}_{\mathcal{A}} - \boldsymbol{\gamma}^*_{\mathcal{A}} = 2\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\{\boldsymbol{U}_{n,\mathcal{A}}(\hat{\boldsymbol{\gamma}}) - \nabla_{\mathcal{A}} p_{\nu_n}(\hat{\boldsymbol{\gamma}})\}.$$

Substituting it into (S4.2) gives

$$\begin{aligned}\boldsymbol{U}_{n,\mathcal{A}^c}(\hat{\boldsymbol{\gamma}}) &= \boldsymbol{U}_{n,\mathcal{A}^c}(\boldsymbol{\gamma}^*) - \boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\tilde{\boldsymbol{\gamma}})\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\{\boldsymbol{U}_{n,\mathcal{A}}(\hat{\boldsymbol{\gamma}}) - \nabla_{\mathcal{A}} p_{\nu_n}(\hat{\boldsymbol{\gamma}})\} \\ &= \boldsymbol{U}_{n,\mathcal{A}^c}(\hat{\boldsymbol{\gamma}}) - \boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma}^*)\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\{\boldsymbol{U}_{n,\mathcal{A}}(\hat{\boldsymbol{\gamma}}) - \nabla_{\mathcal{A}} p_{\nu_n}(\hat{\boldsymbol{\gamma}})\} \\ &\quad + \{\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\tilde{\boldsymbol{\gamma}}) - \boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma}^*)\}\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\{\boldsymbol{U}_{n,\mathcal{A}}(\hat{\boldsymbol{\gamma}}) - \nabla_{\mathcal{A}} p_{\nu_n}(\hat{\boldsymbol{\gamma}})\}.\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\|\boldsymbol{U}_{n,\mathcal{A}^c}(\hat{\boldsymbol{\gamma}})\|_\infty \\
\le\ &\frac{1}{4(1-\zeta)}\|\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma}^*)\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}^*)^{-1}\|_\infty\{\|\boldsymbol{U}_{n,\mathcal{A}}(\hat{\boldsymbol{\gamma}})\|_\infty + \|\nabla_{\mathcal{A}} p_{\nu_n}(\hat{\boldsymbol{\gamma}})\|_\infty\} \\
&+ \|\boldsymbol{U}_{n,\mathcal{A}^c}(\boldsymbol{\gamma}^*)\|_\infty \\
<\ &\frac{1}{2}\nu_n\sqrt{q_n}\rho'(0+) + \frac{\rho'(0+)}{4\rho'(d_n/2)}\{\nu_n\sqrt{q_n}\rho'(d_n/2) + \nu_n\sqrt{q_n}\rho'(d_n/2)\} \\
<\ &\frac{1}{2}\nu_n\sqrt{q_n}\rho'(0+) + \frac{1}{2}\nu_n\sqrt{q_n}\rho'(0+) \\
=\ &\nu_n\sqrt{q_n}\rho'(0+).
\end{aligned}
$$

Therefore, (S3.2) holds.

Finally, as we have shown, $\hat{\boldsymbol{\gamma}} \in \mathcal{B}_0$ and $\hat{\mathcal{A}} = \mathcal{A}$. Then, by condition 11 and Lemma S5, conditioning on event A, (S3.3) also holds.

S4 Additional lemmas and their proofs

Lemma S1. *Under conditions 1 to 8, there exist positive constants C_1, C_2 and D such that for any $x > 0$, with probability less than $C_1\exp(-C_2 n h_n^6 x^2)$,* *it holds that*

$$\left|\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\sum_{v=1}^{n_i}K_{h_n}(t-t_{iv})\{Z_{ij}(t_{iv},t)-E_{nj}(\boldsymbol{\gamma}^*,t)\}d\Lambda_i(t)\right|$$
$$\geq D\left[\{(r_nq_nc_nd_n^{1/2}/n)^{1/2}\}(1+x)+h_n^2+r_nq_n^{-\alpha}\right].$$

Proof of Lemma S1. Let

$$\tilde{S}_n^{(l)}(\boldsymbol{\beta}^*,t)=n^{-1}\sum_{i=1}^{n}Y_i(t)\{\boldsymbol{Z}_i(t,t)\}^{\otimes l}\exp[\{\boldsymbol{\beta}^*(t)\}^T\boldsymbol{X}_i(t)],$$

for $l=0,1,2$, $\tilde{\boldsymbol{E}}_n(\boldsymbol{\beta}^*,t)=\tilde{S}_n^{(1)}(\boldsymbol{\beta}^*,t)/\tilde{S}_n^{(0)}(\boldsymbol{\beta}^*,t)$ and $\tilde{E}_{nj}(\boldsymbol{\gamma}^*,t)$ be the j-th element of $\tilde{\boldsymbol{E}}_n(\boldsymbol{\beta}^*,t)$. Note that,

$$\sum_{i=1}^{n}\int_0^{\tau}\lambda_v(t)\{Z_{ij}(t,t)-\tilde{E}_{nj}(\boldsymbol{\beta}^*,t)\}d\Lambda_i(t)=0.$$

Then,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\sum_{v=1}^{n_i}K_{h_n}(t-t_{iv})\{Z_{ij}(t_{iv},t)-E_{nj}(\boldsymbol{\gamma}^*,t)\}d\Lambda_i(t)\\
=&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\sum_{v=1}^{n_i}K_{h_n}(t-t_{iv})Z_{ij}(t_{iv},t)-\lambda_v(t)Z_{ij}(t,t)d\Lambda_i(t)\\
&+\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\sum_{v=1}^{n_i}K_{h_n}(t-t_{iv})E_{nj}(\boldsymbol{\gamma}^*,t)-\lambda_v(t)\tilde{E}_{nj}(\boldsymbol{\beta}^*,t)d\Lambda_i(t)\\
:=&I_1+I_2.
\end{aligned}$$

For I_1, let $V_i := \int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})Z_{ij}(t_{iv},t) - \lambda_v(t)Z_{ij}(t,t)d\Lambda_i(t)$. Denote $z_{ij}(s,t) = \mathrm{E}\{Z_{ij}(s,t)\}$. We first bound $\mathrm{E}(V_i)$.

$$
\begin{aligned}
&\mathrm{E}\left\{\int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})Z_{ij}(t_{iv},t)d\Lambda_i(t)\right\} \\
&= \mathrm{E}\left[\int_0^\tau \left\{\int K_{h_n}(t-s)z_{ij}(s,t)\lambda_v(s)ds\right\} d\Lambda_i(t)\right] \\
&= \mathrm{E}\left[\int_0^\tau \left\{\int K(u)z_{ij}(t+uh_n,t)\lambda_v(t+uh_n)du\right\} d\Lambda_i(t)\right] \\
&= \mathrm{E}\left\{\int_0^\tau \left(\int K(u)\Big[z_{ij}(t)\lambda_v(t) + \{z_{ij}(t)\lambda_v(t)\}'uh_n\right.\right. \\
&\quad \left.\left. + \{z_{ij}(t)\lambda_v(t)\}''(uh_n)^2/2 + o(h_n^2)\Big]du\right) d\Lambda_i(t)\right\} \\
&= \mathrm{E}\left\{\int_0^\tau Z_{ij}(t)\lambda_v(t)d\Lambda_i(t)\right\} + ch_n^2 + o(h_n^2),
\end{aligned}
\tag{S4.1}
$$

where c is a constant. Hence, $\mathrm{E}[V_i] = O(h_n^2)$. Since $V_i = O(h_n^{-1})$, by the Hoeffding inequality,

$$
P\{|I_1| \geq Dh_n^2(1+x)\} \leq P\{|\bar{V} - \mathrm{E}(\bar{V})| \geq Dh_n^2 x\} \leq 2\exp(-Cnh_n^6x^2).
\tag{S4.2}
$$

For I_2, we have

$$
\begin{aligned}
I_2 &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})-\lambda_v(t)\right\}E_{nj}(\boldsymbol{\gamma}^*,t)d\Lambda_i(t)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\lambda_v(t)\left\{E_{nj}(\boldsymbol{\gamma}^*,t)-\tilde{E}_{nj}(\boldsymbol{\beta}^*,t)\right\}d\Lambda_i(t)\\
&:= J_1+J_2.
\end{aligned}
$$

Similarly as (S4.2), it can be shown that

$$
P\{|J_1|\geq Dh_n^2(1+x)\}\leq 2\exp(-Cnh_n^6x^2). \tag{S4.3}
$$

Next, we bound J_2 by

$$
|J_2|\lesssim \sup_{t\in[0,\tau]}|E_{nj}(\boldsymbol{\gamma}^*,t)-\tilde{E}_{nj}(\boldsymbol{\beta}^*,t)|.
$$

Recall that

$$
E_{nj}(\boldsymbol{\gamma}^*,t)=S_n^{(1)}(\boldsymbol{\gamma}^*,t)/S_n^{(0)}(\boldsymbol{\gamma}^*,t) \text{ and } \tilde{E}_{nj}(\boldsymbol{\beta}^*,t)=\tilde{S}_n^{(1)}(\boldsymbol{\beta}^*,t)/\tilde{S}_n^{(0)}(\boldsymbol{\beta}^*,t),
$$

where

$$
S_n^{(l)}(\boldsymbol{\gamma}^*, t) = n^{-1}\sum_{i=1}^{n}\sum_{v=1}^{n_i} K_{h_n}(t-t_{iv})Y_i(t)\{\boldsymbol{Z}_i(t_{iv}, t)\}^{\otimes l}\exp\{(\boldsymbol{\gamma}^*)^T\boldsymbol{Z}_i(t_{iv}, t)\}
$$
$$
\tilde{S}_n^{(l)}(\boldsymbol{\beta}^*, t) = n^{-1}\sum_{i=1}^{n} Y_i(t)\{\boldsymbol{Z}_i(t,t)\}^{\otimes l}\exp[\{\boldsymbol{\beta}^*(t)\}^T\boldsymbol{X}_i(t)].
$$

In addition, we define $\bar{E}(\boldsymbol{\gamma}^*, t) = \bar{S}_n^{(1)}(\boldsymbol{\gamma}^*, t)/\bar{S}_n^{(0)}(\boldsymbol{\gamma}^*, t)$, where

$$
\bar{S}_n^{(l)}(\boldsymbol{\gamma}^*, t) := n^{-1}\sum_{i=1}^{n} Y_i(t)\{\boldsymbol{Z}_i(t,t)\}^{\otimes l}\exp\{(\boldsymbol{\gamma}^*)^T\boldsymbol{Z}_i(t,t)\}.
$$

Let $s_n^{(l)}(\boldsymbol{\gamma}^*, t) = \mathrm{E}\{S_n^{(l)}(\boldsymbol{\gamma}^*, t)\}$, $\tilde{s}^{(l)}(\boldsymbol{\beta}^*, t) = \mathrm{E}\{\tilde{S}_n^{(l)}(\boldsymbol{\beta}^*, t)\}$ and $\bar{s}^{(l)}(\boldsymbol{\gamma}^*, t) = \mathrm{E}\{\bar{S}_n^{(l)}(\boldsymbol{\gamma}^*, t)\}$. We have

$$
\begin{aligned}
E_{nj}(\boldsymbol{\gamma}^*, t) - \tilde{E}_{nj}(\boldsymbol{\beta}^*, t) = & \underbrace{E_{nj}(\boldsymbol{\gamma}^*, t) - \frac{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)}{s_n^{(0)}(\boldsymbol{\gamma}^*, t)}}_{L_1} + \underbrace{\frac{\tilde{s}_j^{(1)}(\boldsymbol{\beta}^*, t)}{\tilde{s}^{(0)}(\boldsymbol{\beta}^*, t)} - \tilde{E}_{nj}(\boldsymbol{\beta}^*, t)}_{L_2} \\
& + \underbrace{\frac{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)}{s_n^{(0)}(\boldsymbol{\gamma}^*, t)} - \frac{\tilde{s}_j^{(1)}(\boldsymbol{\beta}^*, t)}{\tilde{s}^{(0)}(\boldsymbol{\beta}^*, t)}}_{L_3}.
\end{aligned}
$$

For L_1, we have

$$
\begin{aligned}
L_1 = & \frac{1}{S_n^{(0)}(\boldsymbol{\gamma}^*, t)}\{S_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t) - s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)\} \\
& - \frac{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t)}{S_n^{(0)}(\boldsymbol{\gamma}^*, t)s_n^{(0)}(\boldsymbol{\gamma}^*, t)}\{S_n^{(0)}(\boldsymbol{\gamma}^*, t) - s_n^{(0)}(\boldsymbol{\gamma}^*, t)\}.
\end{aligned}
$$

By Lemma S2, with probability no less than $1 - \exp(-Cr_n q_n c_n d_n^{1/2} h_n^2 x^2)$, we have

$$\sup_{t\in[0,\tau]} |L_1| \lesssim (r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x). \tag{S4.4}$$

Similarly, by Lemma S3, with probability no less than $1 - \exp(-Cr_n x^2)$, we have

$$\sup_{t\in[0,\tau]} |L_2| \lesssim (r_n/n)^{1/2}(1+x). \tag{S4.5}$$

For L_3, we have

$$\begin{aligned} L_3 = &\frac{1}{\lambda_v(t)\tilde{s}^{(0)}(\boldsymbol{\gamma}^*,t)}\{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*,t) - \lambda_v(t)\tilde{s}_j^{(1)}(\boldsymbol{\gamma}^*,t)\} \\ &- \frac{s_{n,j}^{(1)}(\boldsymbol{\gamma}^*,t)}{\lambda_v(t)\tilde{s}^{(0)}(\boldsymbol{\beta}^*,t)s_n^{(0)}(\boldsymbol{\gamma}^*,t)}\{s_n^{(0)}(\boldsymbol{\gamma}^*,t) - \lambda_v(t)\tilde{s}_j^{(0)}(\boldsymbol{\gamma}^*,t)\}. \end{aligned} \tag{S4.6}$$

By the same calculation as in (S4.1), we have

$$s_n^{(0)}(\boldsymbol{\gamma}^*,t) - \lambda_v(t)\bar{s}^{(0)}(\boldsymbol{\gamma}^*,t) = O(h_n^2), \tag{S4.7}$$

$$s_{n,j}^{(1)}(\boldsymbol{\gamma}^*,t) - \lambda_v(t)\bar{s}_j^{(1)}(\boldsymbol{\gamma}^*,t) = O(h_n^2). \tag{S4.8}$$

Moreover,

$$
\begin{aligned}
& |\bar{s}^{(0)}(\boldsymbol{\gamma}^*, t) - \tilde{s}^{(0)}(\boldsymbol{\beta}^*, t)| \\
\leq & \left| \mathrm{E} \left\{ Y_i(t) \exp[(\boldsymbol{\gamma}^*)^T \boldsymbol{Z}_i(t,t) - \{\boldsymbol{\beta}^*(t)\}^T \boldsymbol{X}_i(t)] \right\} \right| \\
\overset{(i)}{\lesssim} & \mathrm{E} \left| (\boldsymbol{\gamma}^*)^T \boldsymbol{Z}_i(t,t) - \{\boldsymbol{\beta}^*(t)\}^T \boldsymbol{X}_i(t) \right| \\
= & \mathrm{E} \left| \sum_{j=1}^{r_n} \left\{ \beta_j^*(t) - (\boldsymbol{\gamma}_j^*)^T \phi(t) \right\} X_{ij}(t) \right| \\
\lesssim & \left| \sum_{j=1}^{r_n} \left\{ \beta_j^*(t) - (\boldsymbol{\gamma}_j^*)^T \phi(t) \right\} \right| \overset{(ii)}{\lesssim} r_n q_n^{-\alpha},
\end{aligned}
\tag{S4.9}
$$

where (i) follows from condition 2 and (ii) follows from condition 6. Similarly, $|\bar{s}_j^{(1)}(\boldsymbol{\gamma}^*, t) - \tilde{s}_j^{(1)}(\boldsymbol{\beta}^*, t)| \lesssim r_n q_n^{-\alpha}$. Therefore,

$$
\begin{aligned}
s_n^{(0)}(\boldsymbol{\gamma}^*, t) - \lambda_v(t) \tilde{s}^{(0)}(\boldsymbol{\gamma}^*, t) &= O(h_n^2 + r_n q_n^{-\alpha}), \\
s_{n,j}^{(1)}(\boldsymbol{\gamma}^*, t) - \lambda_v(t) \tilde{s}_j^{(1)}(\boldsymbol{\gamma}^*, t) &= O(h_n^2 + r_n q_n^{-\alpha}).
\end{aligned}
$$

Then, it follows from (S4.6) that

$$
\sup_{t \in [0,\tau]} |L_3| \lesssim h_n^2 + r_n q_n^{-\alpha}. \tag{S4.10}
$$

Equations (S4.4), (S4.5) and (S4.10) together imply that

$$
\begin{aligned}
&P\left(|J_2| \geq D\{(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x) + h_n^2 + r_n q_n^{-\alpha}\}\right) \\
&\leq C_1 \exp(-C_2 r_n q_n c_n d_n^{1/2} h_n^2 x^2).
\end{aligned} \tag{S4.11}
$$

Finally, the result follows from (S4.2), (S4.3) and (S4.11). □

Lemma S2. *Under conditions 1 to 8, there exist positive constants C and D such that, for any $x > 0$,*

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |S_n^{(0)}(\boldsymbol{\gamma}, t) - s_n^{(0)}(\boldsymbol{\gamma}, t)| \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x)\right\} \\
&\leq \exp(-C r_n q_n c_n d_n^{1/2} h_n^2 x^2), \\
&P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |S_{n,j}^{(1)}(\boldsymbol{\gamma}, t) - s_{n,j}^{(1)}(\boldsymbol{\gamma}, t)| \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x)\right\} \\
&\leq \exp(-C r_n q_n c_n d_n^{1/2} h_n^2 x^2), \\
&P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |S_{n,ij}^{(2)}(\boldsymbol{\gamma}, t) - s_{n,ij}^{(2)}(\boldsymbol{\gamma}, t)| \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x)\right\} \\
&\leq \exp(-C r_n q_n c_n d_n^{1/2} h_n^2 x^2),
\end{aligned}
$$

where $c_n = r_n q_n^2 h_n^{-1} \vee h_n^{-2}$.

Proof of Lemma S2. Let

$$
W_n = \sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |S_n^{(0)}(\boldsymbol{\gamma}, t) - s_n^{(0)}(\boldsymbol{\gamma}, t)|.
$$

We prove the upper bound for W_n. The other two cases can be shown similarly. Let $\mathcal{F} = \{\sum_{v=1}^{n_i} K_{h_n}(t - t_{iv})Y(t)\exp\{\boldsymbol{\gamma}^T \boldsymbol{Z}(t_{iv}, t)\} : \boldsymbol{\gamma} \in \mathcal{B}_0, t \in [0,\tau]\}$. We calculate the bracketing number of the function class $\mathcal{F}$.

$$
\begin{aligned}
&\left|\sum_{v=1}^{n_i} K_{h_n}(t_1 - t_{iv})Y(t_1)\exp\{\boldsymbol{\gamma}_1^T \boldsymbol{Z}(t_{iv}, t_1)\}\right. \\
&\qquad \left. - \sum_{v=1}^{n_i} K_{h_n}(t_2 - t_{iv})Y(t_2)\exp\{\boldsymbol{\gamma}_2^T \boldsymbol{Z}(t_{iv}, t_2)\}\right| \\
&\le \sum_{v=1}^{n_i} K_{h_n}(t_1 - t_{iv})\left|Y(t_1)\exp\{\boldsymbol{\gamma}_1^T \boldsymbol{Z}(t_{iv}, t_1)\} - Y(t_2)\exp\{\boldsymbol{\gamma}_2^T \boldsymbol{Z}(t_{iv}, t_2)\}\right| \\
&+ \sum_{v=1}^{n_i} |K_{h_n}(t_1 - t_{iv}) - K_{h_n}(t_2 - t_{iv})|\,|Y(t_2)\exp\{\boldsymbol{\gamma}_2^T \boldsymbol{Z}(t_{iv}, t_2)\}| \\
&:= I_1 + I_2.
\end{aligned}
$$

For I_1, let $d_{1j} = \boldsymbol{\gamma}_{1,j}^T \boldsymbol{\phi}(t)$ and $d_{2j} = \boldsymbol{\gamma}_{2,j}^T \boldsymbol{\phi}(t)$, we have

$$
\begin{aligned}
&\left|Y(t_1)\exp\{\boldsymbol{\gamma}_1^T \boldsymbol{Z}(t_{iv}, t_1)\} - Y(t_2)\exp\{\boldsymbol{\gamma}_2^T \boldsymbol{Z}(t_{iv}, t_2)\}\right| \\
&\lesssim |\boldsymbol{\gamma}_1^T \boldsymbol{Z}(t_{iv}, t_1) - \boldsymbol{\gamma}_2^T \boldsymbol{Z}(t_{iv}, t_2)| + |Y(t_1) - Y(t_2)| \\
&\le |(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2)^T \boldsymbol{Z}(t_{iv}, t_1)| + |\boldsymbol{\gamma}_2^T\{\boldsymbol{Z}(t_{iv}, t_1) - \boldsymbol{Z}(t_{iv}, t_2)\}| + |Y(t_1) - Y(t_2)| \\
&\le \left|\sum_{j=1}^{r_n}(d_{1j} - d_{2j})X_j(t_{iv})\right| + |\boldsymbol{\gamma}_2^T[\boldsymbol{X}(t_{iv}) \otimes \{\boldsymbol{\phi}(t_1) - \boldsymbol{\phi}(t_2)\}]| + |Y(t_1) - Y(t_2)| \\
&\lesssim r_n q_n \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|_\infty + r_n q_n^2 |t_1 - t_2| + |Y(t_1) - Y(t_2)|
\end{aligned}
$$

Since $K_{h_n}(t - t_{iv}) = O(h_n^{-1})$ and $n_i = O(1)$, we have $I_1 \lesssim r_n q_n^2 h_n^{-1}(\|\boldsymbol{\gamma}_1 -$

$\boldsymbol{\gamma}_2\|_\infty + |t_1 - t_2|) + h_n^{-1}|Y(t_1) - Y(t_2)|$. For I_2, by conditions 2 and 4, we have $I_2 \lesssim h_n^{-2}|t_1 - t_2|$. Denote $\boldsymbol{\theta}_1 = (\boldsymbol{\gamma}_1, t_1)^T$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\gamma}_2, t_2)^T$. Then, we have

$$I_1 + I_2 \lesssim c_n\{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 + |Y(t_1) - Y(t_2)|\},$$

where $c_n = r_n q_n^2 h_n^{-1} \vee h_n^{-2}$. When $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 \leq \epsilon^2/c_n^2$,

$$|f_{\boldsymbol{\theta}_1} - f_{\boldsymbol{\theta}_2}| \leq \epsilon^2/c_n + c_n|Y(t_1) - Y(t_2)|,$$

where $f_{\boldsymbol{\theta}_j} := \sum_{v=1}^{n_i} K_{h_n}(t_j - t_{iv})Y(t_j)\exp\{\boldsymbol{\gamma}_j^T \boldsymbol{Z}(t_{iv}, t_j)\}$. The $L_2(P)$-size of the above bracket is

$$2\epsilon^2/c_n + 2c_n\{\mathrm{E}|Y(t_1) - Y(t_2)|^2\}^{1/2} = 2\epsilon^2/c_n + 2c_n\left\{\int_{t_1}^{t_2} dF_{\tilde{T}}(t)\right\}^{1/2}$$
$$\leq 2\epsilon^2/c_n + 2\epsilon \lesssim \epsilon.$$

Then, to cover $\mathcal{F}$, we need as many brackets as we need balls of radius $\epsilon^2/(2c_n^2)$ to cover $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta} = \mathcal{B}_0 \otimes [0, \tau]$. Hence, the bracketing entropy of $\mathcal{F}$ (see Example 19.7 of Van der Vaart (2000)) is

$$\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \lesssim r_n q_n \log(c_n^2 d_n/\epsilon^2).$$

The class $\mathcal{F}$ has an envelope function F with $||F||_{P,2} = O(h_n^{-1})$. Therefore, by the maximal inequality (Corollary 19.35 of Van der Vaart (2000)), we have

$$E(W_n) \lesssim n^{-1/2} \int_0^{||F||_{P,2}} \sqrt{r_n q_n \log(c_n^2 d_n/\epsilon^2)} d\epsilon \lesssim (r_n q_n c_n d_n^{1/2}/n)^{1/2}.$$

Then, by the functional Hoeffding inequality (Massart and Picard, 2007), for any $x > 0$, we have

$$\begin{aligned} & P\{W_n \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x)\} \\ \leq & P\{W_n - \mathrm{E}(W_n) \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2} x\} \\ \leq & \exp(-C r_n q_n c_n d_n^{1/2} h_n^2 x^2). \end{aligned}$$

□

Lemma S3. *Under conditions 1 to 8, there exist positive constants C and D such that for any $x > 0$,*

$$P\left\{\sup_{t\in[0,\tau]} |\tilde{S}_n^{(0)}(\boldsymbol{\beta}^*, t) - \tilde{s}^{(0)}(\boldsymbol{\beta}^*, t)| \geq D(r_n/n)^{1/2}(1+x)\right\} \leq \exp(-C r_n x^2).$$

$$P\left\{\sup_{t\in[0,\tau]} |\tilde{S}_{n,j}^{(1)}(\boldsymbol{\beta}^*, t) - \tilde{s}_j^{(1)}(\boldsymbol{\beta}^*, t)| \geq D(r_n/n)^{1/2}(1+x)\right\} \leq \exp(-C r_n x^2).$$

$$P\left\{\sup_{t\in[0,\tau]} |\tilde{S}_{n,ij}^{(2)}(\boldsymbol{\beta}^*, t) - \tilde{s}_{ij}^{(2)}(\boldsymbol{\beta}^*, t)| \geq D(r_n/n)^{1/2}(1+x)\right\} \leq \exp(-C r_n x^2).$$

$$P\left\{\sup_{t\in[0,\tau]}|\bar{S}_n^{(0)}(\boldsymbol{\gamma}^*,t)-\bar{s}^{(0)}(\boldsymbol{\gamma}^*,t)|\geq D(r_nq_n/n)^{1/2}(1+x)\right\}\leq \exp(-Cr_nq_nx^2).$$

$$P\left\{\sup_{t\in[0,\tau]}|\bar{S}_{n,j}^{(1)}(\boldsymbol{\gamma}^*,t)-\bar{s}_j^{(1)}(\boldsymbol{\gamma}^*,t)|\geq D(r_nq_n/n)^{1/2}(1+x)\right\}\leq \exp(-Cr_nq_nx^2).$$

$$P\left\{\sup_{t\in[0,\tau]}|\bar{S}_{n,ij}^{(2)}(\boldsymbol{\gamma}^*,t)-\bar{s}_{ij}^{(2)}(\boldsymbol{\gamma}^*,t)|\geq D(r_nq_n/n)^{1/2}(1+x)\right\}\leq \exp(-Cr_nq_nx^2).$$

Proof of Lemma S3. We prove the result for $\tilde{S}_n^{(0)}(\boldsymbol{\beta}^*,t)$. The other cases can be shown similarly. Let

$$W_n=\sup_{t\in[0,\tau]}|\tilde{S}_n^{(0)}(\boldsymbol{\beta}^*,t)-\tilde{s}^{(0)}(\boldsymbol{\beta}^*,t)|.$$

Denote $\mathcal{F}=\{Y(t)\exp[\{\boldsymbol{\beta}^*(t)\}^T\boldsymbol{X}(t)]:t\in[0,\tau]\}$. We calculate the bracketing number of the function class $\mathcal{F}$.

$$\begin{aligned}
&|Y(t_1)\exp[\{\boldsymbol{\beta}^*(t_1)\}^T\boldsymbol{X}(t_1)]-Y(t_2)\exp[\{\boldsymbol{\beta}^*(t_2)\}^T\boldsymbol{X}(t_2)]|\\
\lesssim\ &|Y(t_1)-Y(t_2)|+|\exp[\{\boldsymbol{\beta}^*(t_1)\}^T\boldsymbol{X}(t_1)]-\exp[\{\boldsymbol{\beta}^*(t_2)\}^T\boldsymbol{X}(t_2)]|\\
\lesssim\ &|Y(t_1)-Y(t_2)|+|\{\boldsymbol{\beta}^*(t_1)\}^T\boldsymbol{X}(t_1)-\{\boldsymbol{\beta}^*(t_2)\}^T\boldsymbol{X}(t_2)|\\
\lesssim\ &|Y(t_1)-Y(t_2)|+\sum_{j=1}^{r_n}|\beta_j^*(t_1)-\beta_j^*(t_2)|+\sum_{j=1}^{r_n}|X_j(t_1)-X_j(t_2)|.
\end{aligned}$$

We use brackets of the form $[I_{[t_i,\infty)},I_{[t_{i-1},\infty)}]$ with $F_{\tilde{T}}(t_i-)-F_{\tilde{T}}(t_{i-1}-)<\epsilon^2$ to cover $\{Y(t),t\in[0,\tau]\}$, which forms a grid of points $0=t_0<t_1<\cdots<t_k=\tau$. The L_2-size of these brackets is ϵ. By the continuity assumption

of $\beta_j^*(t)$ in condition 6, to cover $\{\beta_j^*(t) : t \in [0,\tau]\}$, we need as many ϵ-brackets as we need balls of radius $\epsilon/2$ to cover $[0,\tau]$. In addition, by continuity assumption in condition 5, to cover $\{X_j(t) : t \in [0,\tau]\}$, we also need as many brackets as we need balls of radius $\epsilon/2$ to cover $[0,\tau]$. Then, the bracketing entropy of $\mathcal{F}$ is given by

$$\log N_{[]}(\epsilon, \mathcal{F}, L_2(p)) \lesssim r_n \log(\epsilon^{-1}).$$

Moreover, the envelop function F of $\mathcal{F}$ has $||F||_{P,2} = O(1)$. Then, by the maximal inequality

$$\mathrm{E}(W_n) \lesssim n^{-1/2} \int_0^1 \sqrt{r_n \log(\epsilon^{-1})} d\epsilon = O\{(r_n/n)^{1/2}\}.$$

Then, it follows from the functional Hoeffding inequality that for any $x > 0$,

$$\begin{aligned} P\left\{W_n \geq D(r_n/n)^{1/2}(1+x)\right\} &\leq P\left\{W_n - \mathrm{E}[W_n] \geq D(r_n/n)^{1/2}x\right\} \\ &\leq \exp(-Cr_n x^2). \end{aligned}$$

$\square$

Lemma S4. *Under conditions 1 to 8, there exist positive constants C_1, C_2*

and D, such that for any $x > 0$,

$$P\left(\sup_{\gamma\in\mathcal{B}_0}|I_{n,ij}(\boldsymbol{\gamma})-\Sigma_{ij}(\boldsymbol{\gamma})|\geq D\{(r_nq_nc_nd_n^{1/2}/n)^{1/2}(1+x)+h_n^2\}\right)$$
$$\leq C_1\exp(-C_2r_nq_nc_nd_n^{1/2}h_n^2x^2).$$

Proof of Lemma S4. Note that

$$\begin{aligned}
&I_{n,ij}(\boldsymbol{\gamma},t)-\Sigma_{ij}(\boldsymbol{\gamma},t)\\
=&\int_0^\tau\{S_{n,ij}^{(2)}(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}_{ij}^{(2)}(\boldsymbol{\gamma},t)\}d\Lambda_0(t)\\
&-\int_0^\tau\left\{\frac{S_{n,i}^{(1)}(\boldsymbol{\gamma},t)S_{n,j}^{(1)}(\boldsymbol{\gamma},t)}{S_n^{(0)}(\boldsymbol{\gamma},t)}-\frac{\bar{s}_i^{(1)}(\boldsymbol{\gamma},t)\bar{s}_j^{(1)}(\boldsymbol{\gamma},t)}{\bar{s}^{(0)}(\boldsymbol{\gamma},t)}\lambda_v(t)\right\}d\Lambda_0(t)\\
:=&J_1(\boldsymbol{\gamma})-\int_0^\tau j_{2,n}(\boldsymbol{\gamma},t)d\Lambda_0(t).
\end{aligned}$$

For the term $J_1(\boldsymbol{\gamma})$, we have

$$|J_1(\boldsymbol{\gamma})|\leq\sup_{t\in[0,\tau]}|S_{n,ij}^{(2)}(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}_{ij}^{(2)}(\boldsymbol{\gamma},t)|\cdot\Lambda_0(\tau).\tag{S4.12}$$

Similar as in (S4.7) and (S4.8), we have

$$s_{n,ij}^{(2)}(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}_{ij}^{(2)}(\boldsymbol{\gamma},t)=O(h_n^2).$$

This together with Lemma S2 imply that

$$\begin{aligned} &P\Big(\sup_{\gamma\in\mathcal{B}_0,t\in[0,\tau]}|S^{(2)}_{n,ij}(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}^{(2)}_{ij}(\boldsymbol{\gamma},t)| \\ &\qquad \geq D_2\{(r_nq_nc_nd_n^{1/2}/n)^{1/2}(1+x)+h_n^2\}\Big) \\ &\leq \exp(-C_1r_nq_nc_nd_n^{1/2}h_n^2x^2). \end{aligned} \tag{S4.13}$$

Then, by (S4.12), we have

$$\begin{aligned} &P\left(\sup_{\gamma\in\mathcal{B}_0}|J_1(\boldsymbol{\gamma})|\geq D_1\{(r_nq_nc_nd_n^{1/2}/n)^{1/2}(1+x)+h_n^2\}\right) \\ &\leq \exp(-C_1r_nq_nc_nd_n^{1/2}h_n^2x^2). \end{aligned} \tag{S4.14}$$

For the second term, we write $j_{2,n}(\boldsymbol{\gamma},t)$ as

$$\begin{aligned} &j_{2,n}(\boldsymbol{\gamma},t) \\ =&\frac{S^{(1)}_{n,i}(\boldsymbol{\gamma},t)}{S^{(0)}_n(\boldsymbol{\gamma},t)}\{S^{(1)}_{n,j}(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}^{(1)}_j(\boldsymbol{\gamma},t)\} \\ &+\frac{\lambda_v(t)\bar{s}^{(1)}_j(\boldsymbol{\gamma},t)}{S^{(0)}_n(\boldsymbol{\gamma},t)}\{S^{(1)}_{n,i}(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}^{(1)}_i(\boldsymbol{\gamma},t)\} \\ &-\frac{\lambda_v(t)\bar{s}^{(1)}_i(\boldsymbol{\gamma},t)\bar{s}^{(1)}_j(\boldsymbol{\gamma},t)}{S^{(0)}_n(\boldsymbol{\gamma},t)\bar{s}^{(0)}(\boldsymbol{\gamma},t)}\{S^{(0)}_n(\boldsymbol{\gamma},t)-\lambda_v(t)\bar{s}^{(0)}(\boldsymbol{\gamma},t)\}. \end{aligned}$$

Since $\lambda_v(t)$, $S^{(1)}_{n,i}(\boldsymbol{\gamma},t)$, $\bar{s}^{(1)}_j(\boldsymbol{\gamma},t)$ are all bounded and $S^{(0)}_n(\boldsymbol{\gamma},t)$ and $\bar{s}^{(0)}(\boldsymbol{\gamma},t)$

are bounded away from zero, it follows that

$$\begin{aligned}
&\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0} |J_2(\boldsymbol{\gamma})| \\
\lesssim &\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |j_{2,n}(\boldsymbol{\gamma}, t)| \\
\lesssim &\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |S_{n,j}^{(1)}(\boldsymbol{\gamma}, t) - \lambda_v(t)\bar{s}_j^{(1)}(\boldsymbol{\gamma}, t)| \\
&+ \sup_{\boldsymbol{\gamma}\in\mathcal{B}_0, t\in[0,\tau]} |S_n^{(0)}(\boldsymbol{\gamma}, t) - \lambda_v(t)\bar{s}^{(0)}(\boldsymbol{\gamma}, t)|.
\end{aligned}$$

Similar as (S4.13), we have

$$\begin{aligned}
&P\left(\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0} |J_2(\boldsymbol{\gamma})| \geq D_2\{(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1+x) + h_n^2\}\right) \\
&\leq \exp(-C_2 r_n q_n c_n d_n^{1/2} h_n^2 x^2).
\end{aligned} \tag{S4.15}$$

(S4.14) and (S4.15) together complete the proof. □

Lemma S5. *Under conditions 1 to 11, there exist positive constants C_1, C_2, C_3, C_4 and $C_{\min}$ such that,*

$$P\left\{\inf_{\boldsymbol{\beta}\in\mathcal{B}_0} \lambda_{\min}(\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})) \leq \frac{C_{\min}}{2}\right\} \leq C_1 r_n^2 q_n^2 \exp\{-C_2 n h_n^2\}, \tag{S4.16}$$

and

$$P\left\{\sup_{\gamma\in\mathcal{B}_0}\|\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\|_\infty \geq \frac{1}{2}(1-\zeta)\frac{\rho'(0+)}{\rho'(d_n/2)}\right\}$$
$$\leq C_3 p_n r_n q_n^2 \exp\{-C_4 n h_n^2 (r_n q_n)^{-1}\}. \tag{S4.17}$$

Proof of Lemma S5. By Weyl's inequality,

$$\begin{aligned}|\lambda_{\min}(\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}))-\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}))| &\leq \|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})-\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})\|_2\\ &\leq \|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})-\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})\|_1.\end{aligned}$$

By condition 9,

$$\begin{aligned}\inf_{\gamma\in\mathcal{B}_0}\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})) &= 1/\{\sup_{\gamma\in\mathcal{B}_0}\lambda_{\max}(\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}))\} \geq 1/(\sup_{\gamma\in\mathcal{B}_0}\|\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})\|_\infty)\\ &\geq 1/M.\end{aligned} \tag{S4.18}$$

We denote $C_{\min} := 1/M$. Then, it follows from Lemma S4 that

$$
\begin{aligned}
& P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0} |\lambda_{\min}(\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})) - \lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}))| \geq \frac{C_{\min}}{2}\right\} \\
& \leq P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0} \|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})\|_\infty \geq \frac{C_{\min}}{2}\right\} \\
& \leq P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0} \max_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}} |I_{n,ij}(\boldsymbol{\gamma}) - \Sigma_{ij}(\boldsymbol{\gamma})| \geq \frac{C_{\min}}{2}\right\} \\
& \leq r_n^2 q_n^2 P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0} |I_{n,ij}(\boldsymbol{\gamma}) - \Sigma_{ij}(\boldsymbol{\gamma})| \geq \frac{C_{\min}}{2}\right\} \\
& \leq C_1 r_n^2 q_n^2 \exp\{-C_2(nh_n^2 \vee r_n q_n c_n d_n^{1/2} h_n^2)\} \\
& = C_1 r_n^2 q_n^2 \exp\{-C_2 n h_n^2\}.
\end{aligned}
\tag{S4.19}
$$

This result together with (S4.18) imply (S4.16).

To prove (S4.17), observe that

$$
\begin{aligned}
& \boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1} - \boldsymbol{\Sigma}_{\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1} \\
& = \{\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_{\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\}\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1} \\
& \quad - \boldsymbol{\Sigma}_{\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\{\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})\}\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1} \\
& := J_1(\boldsymbol{\gamma}) + J_2(\boldsymbol{\gamma}).
\end{aligned}
\tag{S4.20}
$$

For $J_1(\boldsymbol{\gamma})$, it follows from Lemma S4 that

$$\begin{aligned}
&P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}\|\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})-\boldsymbol{\Sigma}_{\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\|_\infty \geq \frac{(1-\zeta)C_{\min}}{8\sqrt{r_nq_n}}\right\} \\
&\leq P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}\max_{i\in\mathcal{A}^c}\sum_{j\in\mathcal{A}}|I_{n,ij}(\boldsymbol{\gamma})-\Sigma_{ij}(\boldsymbol{\gamma})| \geq \frac{(1-\zeta)C_{\min}}{8\sqrt{r_nq_n}}\right\} \\
&\leq \sum_{i\in\mathcal{A}^c j\in\mathcal{A}} P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}|I_{n,ij}(\boldsymbol{\gamma})-\Sigma_{ij}(\boldsymbol{\gamma})| \geq \frac{(1-\zeta)C_{\min}}{8\sqrt{r_nq_n}}\right\} \\
&\leq C_3(p_n-r_n)r_nq_n^2\exp\{-C_4nh_n^2(r_nq_n)^{-1}\}.
\end{aligned} \tag{S4.21}$$

By definition, $\|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\|_\infty \leq \sqrt{r_nq_n}\|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\|_2$. Then, we have

$$\begin{aligned}
P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}\|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\|_\infty \geq \frac{2\sqrt{r_nq_n}}{C_{\min}}\right\} &\leq P\left\{\inf_{\boldsymbol{\gamma}\in\mathcal{B}_0}\lambda_{\min}(\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})) \leq \frac{C_{\min}}{2}\right\} \\
&\leq C_1r_n^2q_n^2\exp\{-C_2nh_n^2\}.
\end{aligned} \tag{S4.22}$$

Therefore, by the union bound, (S4.21) and (S4.22) together imply that

$$\begin{aligned}
&P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}|J_1(\boldsymbol{\gamma})| \geq \frac{(1-\zeta)\rho'(0+)}{4\rho'(d_n/2)}\right\} \leq P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}|J_1(\boldsymbol{\gamma})| \geq \frac{1-\zeta}{4}\right\} \\
&\leq C_1r_n^2q_n^2\exp\{-C_2nh_n^2\} + C_3(p_n-r_n)r_nq_n^2\exp\{-C_4nh_n^2(r_nq_n)^{-1}\},
\end{aligned} \tag{S4.23}$$

since $\rho'(0+)/\rho'(d_n/2) \geq 1$ by the concavity assumption in condition 11.

Similar as (S4.21), we have

$$P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}\|\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})-\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})\|_\infty \geq \frac{C_{\min}}{8\sqrt{r_n q_n}}\right\} \leq C_3 r_n^2 q_n^2 \exp\{-C_4 n h_n^2 (r_n q_n)^{-1}\}.$$

This together with (S4.22) imply that

$$\begin{aligned} &P\left\{\sup_{\boldsymbol{\gamma}\in\mathcal{B}_0}|J_2(\boldsymbol{\gamma})| \geq \frac{(1-\zeta)\rho'(0+)}{4\rho'(d_n/2)}\right\} \\ &\leq C_1 r_n^2 q_n^2 \exp\{-C_2 n h_n^2\} + C_3 r_n^2 q_n^2 \exp\{-C_4 n h_n^2 (r_n q_n)^{-1}\}. \end{aligned} \tag{S4.24}$$

Finally, it follows from (S4.20), (S4.23) and (S4.24) that

$$\begin{aligned} &P\left\{\|\boldsymbol{I}_{n,\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\boldsymbol{I}_{n,\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}-\boldsymbol{\Sigma}_{\mathcal{A}^c\mathcal{A}}(\boldsymbol{\gamma})\boldsymbol{\Sigma}_{\mathcal{A}\mathcal{A}}(\boldsymbol{\gamma})^{-1}\|_\infty \geq \frac{(1-\zeta)\rho'(0+)}{2\rho'(d_n/2)}\right\} \\ &\leq C_3 p_n r_n q_n^2 \exp\{-C_4 n h_n^2 (r_n q_n)^{-1}\}. \end{aligned}$$

This together with condition 10 complete the proof. □

S5 Additional simulation results

Figure S1 shows the running time of the proposed method with ℓ_0-regularization penalty based on λ with length of 10 and fixed α and h. Overall, the computation time increased linearly with the number of covariates. When $p_n = 1000$ and $n = 200$, the running time is 634 seconds, with a total of

$p_n q_n = 5000$ parameters.

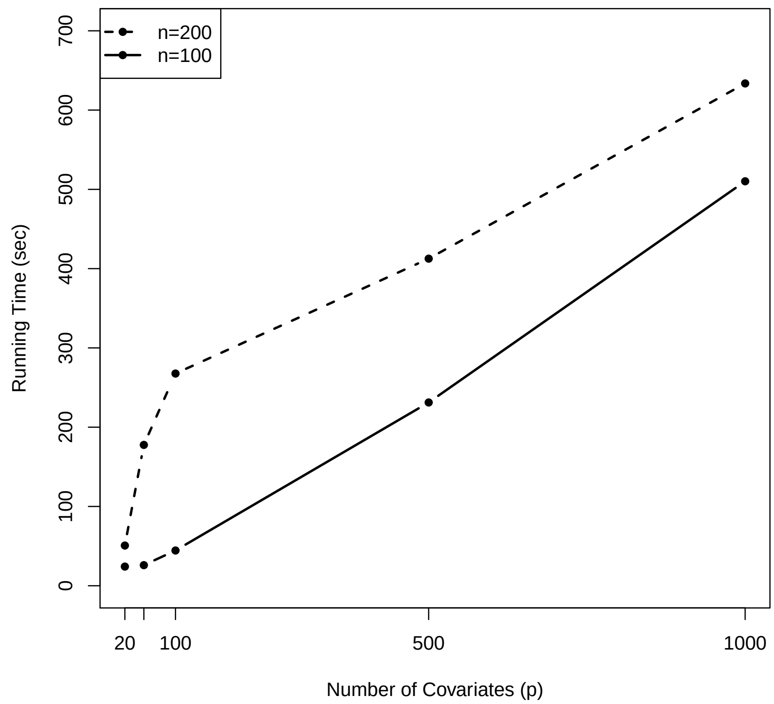

Figure S1: Running time in seconds of the proposed ℓ_0Net for various sample sizes and number of covariates

Table S1 summarizes the comparison results by using different kernel functions for both settings of $\beta(t)$. Epanechnikov and Gaussian kernels were considered. The simulation results are very similar between these two kernels. Both show our proposed approach has a smaller SSE, much better FP and comparable TP to either group LASSO or network regularization.

Table S2 summarizes the performance of bandwidth selection. It can be

Table S1: Comparison of estimation and selection performance of the proposed DB-hazard using different kernel functions under various penalty functions.

	Epanechnikov			Gaussian		
	gLasso†	gNet‡	ℓ_0Net*	gLasso	gNet	ℓ_0Net
Setting (a)						
			$n = 100,\ p_n = 1000$			
SSE[1]	8.34	6.25	4.57	8.23	6.13	4.26
TP[2]	7.7	8.0	8.0	7.7	8.0	8.0
FP[3]	33.2	127.2	1.6	38.4	125.5	1.7
			$n = 200,\ p_n = 1000$			
SSE	5.04	4.17	2.83	5.01	4.04	2.68
TP	8.0	8.0	8.0	8.0	8.0	8.0
FP	57.1	149.0	1.7	61.1	151.8	1.0
Setting (b)						
			$n = 100,\ p_n = 1000$			
SSE	14.14	13.91	12.59	14.00	13.78	12.42
TP	2.1	3.3	3.5	2.4	3.5	3.6
FP	14.8	38.9	5.2	16.8	44.0	5.0
			$n = 200,\ p_n = 1000$			
SSE	10.43	10.02	8.06	10.50	9.79	7.74
TP	5.9	7.1	7.4	5.9	7.3	7.4
FP	48.2	133.6	1.0	44.1	142.8	0.9

†: group Lasso; ‡: group Lasso with a Laplacian penalty; *: ℓ_0-regularization penalty (10)
[1]:sum of squared error; [2]:number of true positive; [3]:number of false positive.

seen from the table that the two kernel functions had similar performance. Our selected bandwidths by both kernel functions are very close to the "Best" bandwidth, indicating satisfactory performance of our data-driven procedure.

Table S2: Performance of the bandwidth selection procedure for DB-hazard using different kernel functions.

	Epanechnikov		Gaussian	
	Selected	Best[1]	Selected	Best
Setting (a)				
		$n = 100,\ p_n = 1000$		
Bandwidth	0.056	0.085	0.061	0.066
SSE[2]	4.57	3.89	4.26	3.91
		$n = 200,\ p_n = 1000$		
Bandwidth	0.059	0.086	0.065	0.077
SSE	2.83	2.19	2.68	2.29
Setting (b)				
		$n = 100,\ p_n = 1000$		
Bandwidth	0.055	0.113	0.057	0.110
SSE	12.59	11.31	12.42	11.36
		$n = 200,\ p_n = 1000$		
Bandwidth	0.061	0.104	0.062	0.085
SSE	8.06	6.90	7.74	6.91

[1]: defined as the bandwidth leading to the smallest SSE; [2]: sum of squared errors.

Table S3 summarizes the impact of various numbers of basis functions. Quadratic B-splines with 5, 7 and 10 interior knots, corresponding to $q_n = 8,\ 10,\ 13$, respectively, were considered. We observed an increase in SSE and the number of identified variables as the number of basis functions increased. Note that $\beta_j(t)$ is a linear combination of basis functions. To obtain $\beta_j(t) = 0$, all the elements in the coefficient vector $\gamma_j = (\gamma_{j1}, \cdots, \gamma_{jq_n})^T$

have to be zero. Thus, the trend is expected that it is more likely to obtain non-zero estimates with more basis functions. After increasing $n = 100$ to 200, the performance improved, which may suggest we need more sample sizes when describing a more complicated function $\beta_j(t)$ with more basis functions.

Table S3: Comparison of estimation and selection performance of the proposed DB-hazard using various numbers of knots under various penalty functions.

	Setting (a)			Setting (b)		
	gLasso†	gNet‡	ℓ_0Net*	gLasso	gNet	ℓ_0Net
			$n = 100,\ p_n = 1000,\ q_n = 8$			
SSE[1]	9.83	8.07	6.69	14.47	14.38	13.85
TP[2]	7.8	8.0	8.0	2.0	3.0	2.7
FP[3]	98.9	327.9	14.9	48.1	115.6	34.5
			$n = 100,\ p_n = 1000,\ q_n = 10$			
SSE	10.43	8.94	7.88	14.59	14.55	14.25
TP	7.9	8.0	8.0	1.9	3.3	2.4
FP	93.7	339.8	23.2	50.2	160.8	39.7
			$n = 100,\ p_n = 1000,\ q_n = 13$			
SSE	11.04	10.41	9.69	14.73	14.73	14.50
TP	7.9	8.0	8.0	2.6	3.8	2.8
FP	169.7	900.8	52.0	96.3	258.0	70.4
			$n = 200,\ p_n = 1000,\ q_n = 8$			
SSE	6.61	5.43	4.12	12.07	11.49	9.91
TP	8.0	8.0	8.0	5.0	6.7	6.9
FP	149.4	416.6	6.6	125.9	329.9	10.8
			$n = 200,\ p_n = 1000,\ q_n = 10$			
SSE	7.31	6.05	5.05	12.63	12.33	11.04
TP	8.0	8.0	8.0	5.1	6.3	6.5
FP	149.9	458.3	10.4	103.7	311.2	22.8
			$n = 200,\ p_n = 1000,\ q_n = 13$			
SSE	8.05	7.59	6.74	13.31	13.18	12.31
TP	8.0	8.0	8.0	5.9	6.6	6.2
FP	262.2	917.8	18.4	199.8	584.3	40.4

†: group Lasso; ‡: group Lasso with a Laplacian penalty; *: ℓ_0-regularization penalty (10)
[1]:sum of squared error; [2]:number of true positive; [3]:number of false positive.

Table S4: Estimates of time-dependent sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV) and area under curve (AUC) using our kernel smoothing method based on longitudinal data, the LVCF method and the model based on baseline data.

Year	SEN	SPE	PPV	NPV	AUC
			DB-hazard		
2	0.959	0.736	0.220	0.996	0.902
4	0.886	0.817	0.555	0.965	0.910
6	1.000	0.873	0.540	1.000	0.924
			LVCF		
2	0.499	0.873	0.234	0.957	0.708
4	0.658	0.832	0.502	0.904	0.736
6	0.900	0.651	0.278	0.978	0.735
			Baseline		
2	0.958	0.739	0.222	0.996	0.864
4	0.914	0.740	0.476	0.971	0.878
6	0.900	0.810	0.414	0.982	0.849

S6 Additional information for real data analysis

Table S4 summarizes the area under the ROC curve (AUC), time-dependent sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) at a given time where the threshold is obtained by optimizing Youden's index.

Figure S2 plots the number of subjects with available clinical measures (time-to-diagnosis outcome) and longitudinal imaging measurements at several follow up time (allowing a window of 6 month), which shows sparse measurements of imaging biomarkers at times (e.g., 18 month after baseline).

Figure S3 shows the heatmaps of the 136 features measured at the

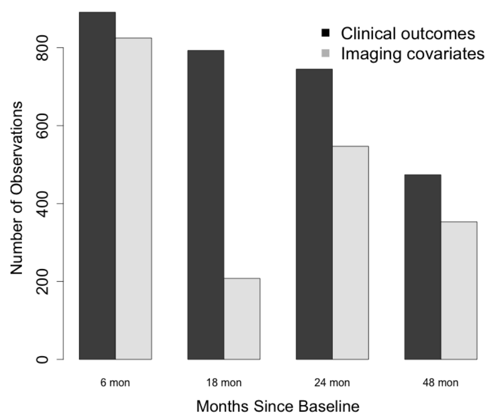

Figure S2: Number of subjects with clinical assessment of the time-to-diagnosis outcome and neuroimaging biomarker measures at several follow up time in PREDICT-HD study.

baseline and at the last visit for 142 subjects who were diagnosed with HD during the study (converters) and 390 subjects who remained free of HD diagnosis (non-converters).

Figure S4 shows the heatmaps of the selected features, where they are seen to better distinguish converters from non-converts than other non-selected noise features in Figure S3.

Figure S5 shows the estimated effect profiles of top 6 measures selected by DB-hazard.

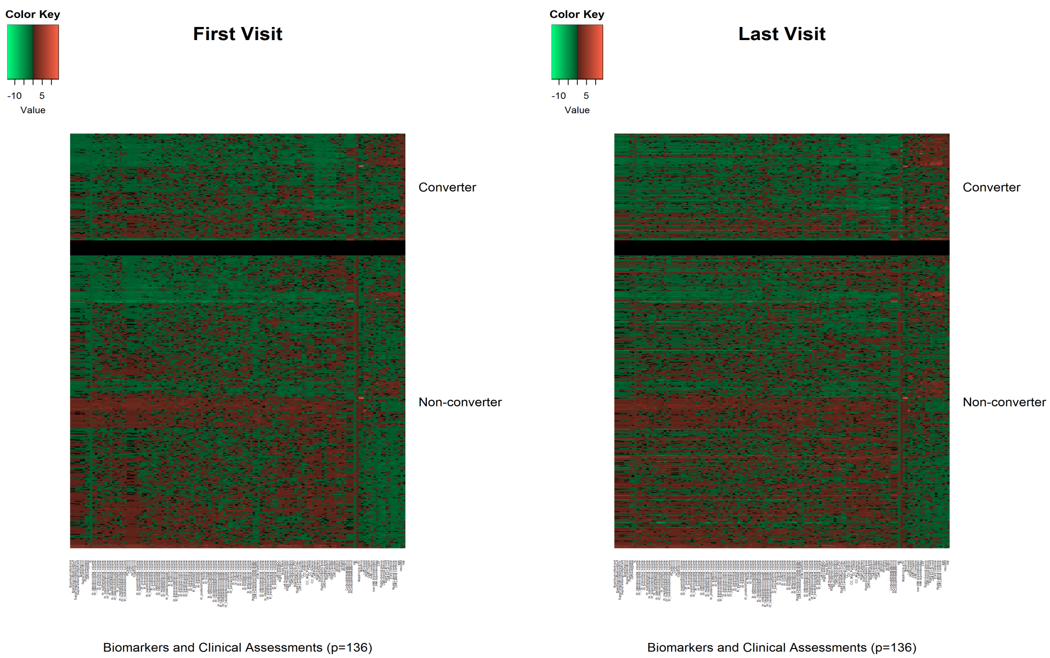

Figure S3: Heatmaps of all feature variables on subjects with at least two neuroimaging biomarker measures. “Converter”: Subjects who were diagnosed of HD during the follow up; ”Non-converter”: subjects who did not receive diagnosis during follow up.

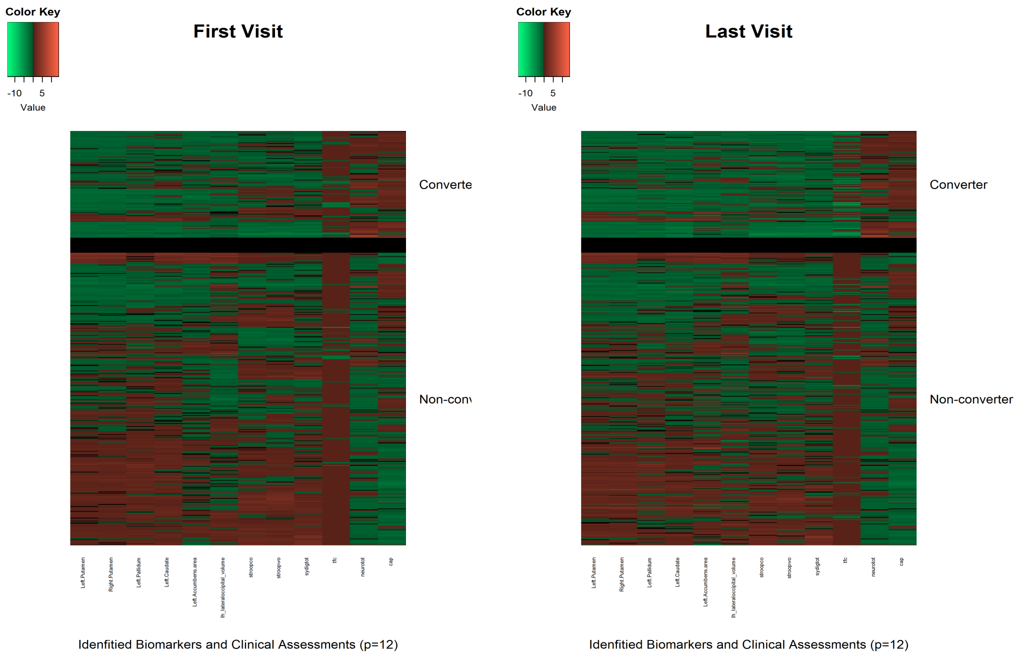

Figure S4: Heatmaps of feature variables selected by DB-hazard on subjects with at least two neuroimaging biomarker measures. “Converter”: Subjects who were diagnosed of HD during the follow up; ”Non-converter”: subjects who did not receive diagnosis during follow up.

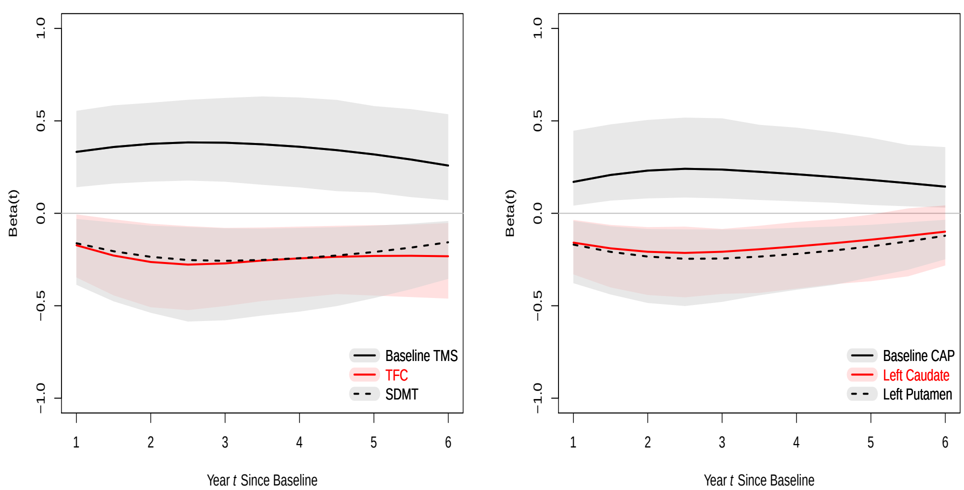

Figure S5: Estimated effects of six most informative markers identified by DB-hazard and their confidence intervals.

Bibliography

Boyd, S. and Vandenberghe, L. (2004) *Convex optimization*. Cambridge university press.

van de Geer, S. (1995) Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *The Annals of Statistics*, 1779–1801.

Massart, P. and Picard, J. (2007) *Concentration inequalities and model selection*, vol. 1896. Springer.

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