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 \((\hat{\gamma}, \hat{\theta})\) solves (7) for \(\phi_n\), there exists a \(c_n\) such that \((\hat{\gamma}, \hat{\theta})\) also solves (6) for \(c_n\); and vice versa.

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

\[
L(\tilde{\gamma}, \tilde{\theta}) < L(\hat{\gamma}, \hat{\theta}) \quad \text{and} \quad \sum_{j=1}^{p_n} \|\tilde{\gamma}_j - \tilde{\theta}_j\|_2 \leq c_n.
\]
Then, by definition,

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

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

Next, we prove that, for any given \(c_n\), if \((\hat{\gamma}, \hat{\theta})\) is the solution to (6), we can always find a \(\phi_n\) such that \((\hat{\gamma}, \hat{\theta})\) also solves (7). Suppose \((\gamma, \theta) = \arg\min_{\gamma, \theta} L(\gamma, \theta)\) is the minimizer of the unconstrained problem. Let \(C_{\text{max}} = \sum_{j=1}^{p_n} \| \gamma_j^* - \theta_j^* \|_2\). Then, for any \(c_n \geq C_{\text{max}}\), \((\hat{\gamma}, \hat{\theta})\) is also the solution to (6). In this case, it’s easy to check that \((\hat{\gamma}, \hat{\theta})\) also solves (7) with \(\phi_n = 0\).

For \(c_n < C_{\text{max}}\), suppose the solution to (6) is given by \((\hat{\gamma}, \hat{\theta})\). Let \(C_{\phi_n} = \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_n} - \hat{\theta}_j^{\phi_n} \|_2\), where \((\hat{\gamma}^{\phi_n}, \hat{\theta}^{\phi_n})\) is the solution to (7) for \(\phi_n\). We prove that \(C_{\phi_n}\) is a decreasing function of \(\phi_n\). In fact, suppose \((\hat{\gamma}^{\phi_2}, \hat{\theta}^{\phi_2})\) and \((\hat{\gamma}^{\phi_1}, \hat{\theta}^{\phi_1})\) are solutions to (7) for \(\phi_1\) and \(\phi_2\) respectively and \(\phi_1 < \phi_2\). By definition,

\[
L(\hat{\gamma}^{\phi_2}, \hat{\theta}^{\phi_2}) + \phi_2 \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_2} - \hat{\theta}_j^{\phi_2} \|_2 \\
\leq L(\hat{\gamma}^{\phi_1}, \hat{\theta}^{\phi_1}) + \phi_2 \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_1} - \hat{\theta}_j^{\phi_1} \|_2 \\
= L(\hat{\gamma}^{\phi_1}, \hat{\theta}^{\phi_1}) + \phi_1 \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_1} - \hat{\theta}_j^{\phi_1} \|_2 + (\phi_2 - \phi_1) \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_1} - \hat{\theta}_j^{\phi_1} \|_2 \\
\leq L(\hat{\gamma}^{\phi_2}, \hat{\theta}^{\phi_2}) + \phi_1 \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_2} - \hat{\theta}_j^{\phi_2} \|_2 + (\phi_2 - \phi_1) \sum_{j=1}^{p_n} \| \hat{\gamma}_j^{\phi_1} - \hat{\theta}_j^{\phi_1} \|_2
\]

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

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

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

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

This contradicts with the assumption that \((\tilde{\gamma}, \tilde{\theta})\) is the global minimizer of (6).

\section*{S2 Proof of Lemma 1}

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

\[
U_{n,j}(\gamma^*) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) \{Z_{ij}(t_{iv}, t) - E_{nj}(\gamma^*, t)\} d\Lambda_i(t)
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) \{Z_{ij}(t_{iv}, t) - E_{nj}(\gamma^*, t)\} dM_i(t)
\]

\[= I_1 + I_2.\]
The upper bound of $I_1$ will be given in Lemma S1 in Section S4. For $I_2$, we have

\[
I_2 = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) \{ Z_{ij}(t_{iv}, t) - e_{nj}(\gamma^*, t) \} dM_i(t)
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) \{ E_{nj}(\gamma^*, t) - e_{nj}(\gamma^*, t) \} dM_i(t)
\]

\[:= J_{1n} - J_{2n}.\]

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

\[
P\left( |(nh_n^2)^{1/2}J_{1n}| > x \right) \leq 2 \exp(-Cx^2). \tag{S.1} \]

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

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

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

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

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

4
In fact, we have

\[
E_{n_j}(\gamma^*, t) - e_{n_j}(\gamma^*, t) = \frac{S_{n,j}^{(1)}(\gamma^*, t)}{S_{n}^{(0)}(\gamma^*, t)} - \frac{s_{n,j}^{(1)}(\gamma^*, t)}{s_{n}^{(0)}(\gamma^*, t)} = \frac{1}{S_{n}^{(0)}(\gamma^*, t)} \{S_{n,j}^{(1)}(\gamma^*, t) - s_{n,j}^{(1)}(\gamma^*, t)\} + \frac{s_{n,j}^{(1)}(\gamma^*, t)}{S_{n}^{(0)}(\gamma^*, t)s_{n}^{(0)}(\gamma^*, t)} \{S_{n}^{(0)}(\gamma^*, t) - s_{n}^{(0)}(\gamma^*, t)\}.
\]

Then, conditioning on \( A \), condition 8 implies (S.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

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

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

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

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 \( \langle (nh_n^2)^{1/2} J_{2n} \rangle \),

\[
\langle (nh_n^2)^{1/2} J_{2n}(t) \rangle = n^{-1} h_n^2 \int_0^t \{E_{ij}(t_{iv}, u) - e_{n_j}(\gamma^*, u)\}^2 d\langle \bar{M}(u) \rangle \leq h_n^2 \left[ \sup_{t \in [0, \tau]} \{E_{ij}(t_{iv}, u) - e_{n_j}(\gamma^*, u)\} \right]^2 \int_0^t S_{n}^{(0)}(\beta^*, u) d\Lambda_0(u) = O(1),
\]
where the last equality follows from (S.2), condition 1 and the fact that \( \sup_{t \in [0, r]} |S_n^{(0)}(\beta^*, t)| \lesssim h_n^{-1} \). Then, it follows from Lemma 2.1 of van de Geer (1995) that

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

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

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

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] \( \{j : \hat{\gamma}_j \neq 0\} = \{j : \gamma_j^* \neq 0\} \).

[2] \( \max_{j \in A} |\hat{\gamma}_{ji} - \gamma_j^*| \leq M\nu_n \sqrt{q_n} \).


By optimization theory (Boyd and Vandenberghe, 2004), any vector \( \gamma \) satisfies the fol-
lowing KKT conditions is a solution to (5):

\[ U_{n,j}(\gamma) = \nu_n \sqrt{q_n} \rho'(\|\gamma_j\|_2) \|\gamma_j\|_2^{-1} \gamma_j, \text{ if } \gamma_j \neq 0, \]  
(S.4)

\[ \|U_{n,j}(\gamma)\|_\infty < \nu_n \sqrt{q_n} \rho'(0^+), \text{ if } \gamma_j = 0, \]  
(S.5)

\[ \lambda_{\min}(I_{n,\hat{A}}(\gamma)) > \nu_n \kappa(\rho, \gamma), \]  
(S.6)

where \( \hat{A} := \{ j_l : \gamma_j \neq 0 \text{ and } 1 \leq l \leq q_n \} \).

We define the event \( A \) as

\[ A = \{ n_i \leq M_\varepsilon \} \cap \{ \|U_n(\gamma^*)\|_\infty \leq \nu_n \sqrt{q_n} \rho'(0^+)/2 \} \cap \left\{ \inf_{\gamma \in B_0} : \lambda_{\min}(I_{n,\hat{A}}(\gamma)) > C_{\min}/2 \right\} \]
\[ \cap \left\{ \sup_{\gamma \in B_0} \|I_{n,\hat{A}}(\gamma)I_{n,\hat{A}}(\gamma)^{-1}\|_\infty < \frac{1}{2}(1 - \zeta) \rho'(0^+) \rho'(d_n/2) \right\}. \]

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

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

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

[1] Let \( \mathcal{N} \) denote the hypercube \( \{ \gamma_A \in \mathcal{R}^{r_n q_n} : \|\gamma_A - \gamma_A^*\|_\infty \leq M \nu_n \sqrt{q_n} \} \), where \( M \) is a sufficiently large constant. We show that within \( \mathcal{N} \), there exists a solution \( \hat{\gamma}_A \) to equation (S.4). We define a function \( f : \mathcal{R}^{r_n q_n} \to \mathcal{R}^{r_n q_n} \) as

\[ f(\gamma_A) = \gamma_A + 2 I_{n,\hat{A}}(\gamma^*)^{-1} \{ U_{n,\hat{A}}(\gamma) - \nabla_{A} p_n(\gamma) \}, \]  
(S.7)
where $\gamma \in \mathbb{R}^{p \times q}$ such that $\gamma_{A^c} = 0$, $\nabla_{\mathcal{A}p_{\nu_n}}(\gamma) := \nu_n \sqrt{q_n} \rho'(\|\gamma_j\|_2)\|\gamma_j\|_2^{-1} \gamma_j$. By the Taylor expansion,

$$U_{n,\mathcal{A}}(\gamma) = U_{n,\mathcal{A}}(\gamma^*) - \frac{1}{2} I_{n,\mathcal{A}}(\tilde{\gamma})(\gamma_{\mathcal{A}} - \gamma_{\mathcal{A}}^*),$$

where $\tilde{\gamma}$ lies on the line segment connecting $\gamma$ and $\gamma^*$. Substituting it into (S.7) gives

$$f(\gamma_{\mathcal{A}}) - \gamma_{\mathcal{A}}^* = \{I_{r \times q} - I_{n,\mathcal{A}}(\gamma^*)^{-1} I_{n,\mathcal{A}}(\tilde{\gamma})\}(\gamma_{\mathcal{A}} - \gamma_{\mathcal{A}}^*)$$

$$+ 2 I_{n,\mathcal{A}}(\gamma^*)^{-1}\{U_{n,\mathcal{A}}(\gamma_{\mathcal{A}}^*) - \nabla_{\mathcal{A}p_{\nu_n}}(\gamma)\},$$

where $I_{r \times q}$ is a $r \times q$ identity matrix. Without loss of generality, we assume

$$\|I_{r \times q} - I_{n,\mathcal{A}}(\gamma^*)^{-1} I_{n,\mathcal{A}}(\tilde{\gamma})\|_\infty \leq \frac{1}{2}. \quad (S.8)$$

Moreover, since $d_n \geq 2\|\gamma_j - \gamma_j^*\|_\infty$, it follows that

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

Hence,

$$\|\gamma_j\|_2 \geq \|\gamma_j^*\|_2 - \|\gamma_j - \gamma_j^*\|_2 \geq d_n/2.$$ 

By the concavity assumption of $\rho(t)$, we have $\rho'(\|\gamma_j\|_2) \leq \rho'(d_n/2)$. Therefore,

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

\[ \|f(\gamma) - \gamma_A^*\|_\infty \leq \frac{1}{2}\|\gamma - \gamma_A^*\|_\infty + 2\|I_{n,AA}(\gamma^*)^{-1}\|_\infty \{\|U_{n,AA}(\gamma_A^*)\|_\infty + \|\nabla_A p_{\nu_n}(\gamma)\|_\infty\} \]

\[ \leq \frac{1}{2} M \nu_n \sqrt{q_n} + \frac{4}{C_{\text{min}}} \left\{ \frac{\rho'(0+)}{2} \nu_n \sqrt{q_n} + \nu_n \sqrt{q_n} \rho'(d_n/2) \right\} \]

\[ \leq \left( \frac{M}{2} + \frac{6 \rho'(0+)}{C_{\text{min}}} \right) \nu_n \sqrt{q_n} \]

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 \(\text{sign}(\gamma_A) = \text{sign}(\gamma_A^*)\) for any \(\gamma_A \in \mathcal{N}\). Therefore, \(f(\gamma_A)\) is a continuous function on the convex and compact set \(\mathcal{N}\). By Brouwer’s fixed point theorem, there exists a solution \(\hat{\gamma}_A \in \mathcal{N}\) to the problem \(f(\gamma_A) = \gamma_A\), which also solves (S.4).

[2] We expand \(\hat{\gamma}_A\) to be \(\hat{\gamma} \in \mathbb{R}^{p_nq_n}\) such that \(\hat{\gamma}_{Ac} = 0\). We further show that \(\hat{\gamma}\) satisfies (S.5). Again, by the Taylor expansion of \(U_{n,Ac}(\hat{\gamma})\) around \(\gamma^*\), we have

\[ U_{n,Ac}(\hat{\gamma}) = U_{n,Ac}(\gamma^*) - \frac{1}{2} I_{n,AcA}(\hat{\gamma})(\hat{\gamma}_A - \gamma_A^*), \]  

where \(\hat{\gamma}\) lies on the line segment connecting \(\hat{\gamma}\) and \(\gamma^*\). Since \(f(\hat{\gamma}_A) = 0\), it holds that

\[ \hat{\gamma}_A - \gamma_A^* = 2I_{n,AA}(\gamma^*)^{-1}\{U_{n,AA}(\hat{\gamma}) - \nabla_A p_{\nu_n}(\hat{\gamma})\} \]
Substituting it into (S.11) gives

\[
U_{n,A^c}(\hat{\gamma}) = U_{n,A^c}(\gamma^*) - I_{n,A^c}(\hat{\gamma})I_{n,A\Lambda}(\gamma^*)^{-1}\{U_{n,A}(\hat{\gamma}) - \nabla_{A\Lambda}\nu_0(\hat{\gamma})\} \\
= U_{n,A^c}(\hat{\gamma}) - I_{n,A^c}(\gamma^*)I_{n,A\Lambda}(\gamma^*)^{-1}\{U_{n,A}(\hat{\gamma}) - \nabla_{A\Lambda}\nu_0(\hat{\gamma})\} \\
+ \{I_{n,A^c}(\hat{\gamma}) - I_{n,A^c}(\gamma^*)\}I_{n,A\Lambda}(\gamma^*)^{-1}\{U_{n,A}(\hat{\gamma}) - \nabla_{A\Lambda}\nu_0(\hat{\gamma})\}.
\]

Therefore,

\[
\|U_{n,A^c}(\hat{\gamma})\|_{\infty} \leq \frac{1}{4(1-\zeta)}\|I_{n,A^c}(\gamma^*)I_{n,A\Lambda}(\gamma^*)^{-1}\|_{\infty}\{\|U_{n,A}(\hat{\gamma})\|_{\infty} + \|\nabla_{A\Lambda}\nu_0(\hat{\gamma})\|_{\infty}\} \\
+ \|U_{n,A^c}(\gamma^*)\|_{\infty} \\
< \frac{1}{2}\nu_n\sqrt{q_n}\rho'(0+) + \frac{\rho'(0+)}{4d_n/2}\nu_n\sqrt{q_n}\rho'(d_n/2) + \frac{\nu_n}{4d_n/2}\nu_n\sqrt{q_n}\rho'(d_n/2) \\
< \frac{1}{2}\nu_n\sqrt{q_n}\rho'(0+) + \frac{1}{2}\nu_n\sqrt{q_n}\rho'(0+) \\
= \nu_n\sqrt{q_n}\rho'(0+).
\]

Therefore, (S.5) holds.

Finally, as we have shown, \(\hat{\gamma} \in B_0\) and \(A = \hat{A}\). Then, by condition 11 and Lemma S5, conditioning on event \(A\), (S.6) 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 $C_1 \exp(-C_2 nh_n^6 x^2)$, it holds that

$$
\left| \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) \{ Z_{ij}(t_{iv}, t) - E_{nj}(\gamma^*, t) \} d\Lambda_i(t) \right|
\geq D \left[ \{ (r_n q_n a d_n^1/n)^{1/2} \} (1 + x) + h_n^2 + r_n q_n^{-\alpha} \right].
$$

Proof of Lemma S1. Let $\tilde{S}_n^{(l)}(\beta^*, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \{ Z_i(t, t) \}^{\otimes l} \exp[\{ \beta^*(t) \}^T X_i(t)]$, for $l = 0, 1, 2$, $\tilde{E}_n(\beta^*, t) = \tilde{S}_n^{(1)}(\beta^*, t)/\tilde{S}_n^{(0)}(\beta^*, t)$ and $\tilde{E}_{nj}(\gamma^*, t)$ be the $j$-th element of $\tilde{E}_n(\beta^*, t)$. Note that,

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

Then,

$$
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) \{ Z_{ij}(t_{iv}, t) - E_{nj}(\gamma^*, t) \} d\Lambda_i(t)
= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) Z_{ij}(t_{iv}, t) - \lambda_v(t) Z_{ij}(t, t) d\Lambda_i(t)
+ \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \sum_{v=1}^{n_i} K_{h_n}(t - t_{iv}) E_{nj}(\gamma^*, t) - \lambda_v(t) \tilde{E}_{nj}(\beta^*, t) d\Lambda_i(t)
:= I_1 + I_2.
$$

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 $z_{ij}(s, t) = \ldots$
E\{Z_{ij}(s, t)\}. We first bound E(V_i).

\[
E \left\{ \int_0^\tau \sum_{v=1}^n K_{hn}(t - t_{iv}) Z_{ij}(t_{iv}, t) d\Lambda_i(t) \right\} \\
= E \left[ \int_0^\tau \left\{ \int K_{hn}(t - s) z_{ij}(s, t) \lambda_v(s) ds \right\} d\Lambda_i(t) \right] \\
= E \left[ \int_0^\tau \left\{ \int K(u) z_{ij}(t + uh_n, t) \lambda_v(t + uh_n) du \right\} d\Lambda_i(t) \right] \\
= E \left\{ \int_0^\tau \left( \int K(u) [z_{ij}(t) \lambda_v(t) + \{z_{ij}(t) \lambda_v(t)\}' uh_n + \{z_{ij}(t) \lambda_v(t)\}''(uh_n)^2/2 + o(h_n^2)] du \right) d\Lambda_i(t) \right\} \\
= E \left\{ \int_0^\tau Z_{ij}(t) \lambda_v(t) d\Lambda_i(t) \right\} + ch_n^2 + o(h_n^2),
\]

(S.10)

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

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

(S.11)

For I_2, we have

\[
I_2 = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \sum_{v=1}^n K_{hn}(t - t_{iv}) - \lambda_v(t) \right\} E_{nj}(\gamma^*, t) d\Lambda_i(t) \\
+ \frac{1}{n} \sum_{i=1}^n \int_0^\tau \lambda_v(t) \left\{ E_{nj}(\gamma^*, t) - \tilde{E}_{nj}(\beta^*, t) \right\} d\Lambda_i(t) \\
:= J_1 + J_2.
\]

Similarly as (S.11), it can be shown that

\[
P\{|J_1| \geq Dh_n^2(1 + x)\} \leq 2 \exp(-Cnh_n^6x^2).
\]

(S.12)
Next, we bound $J_2$ by

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

Recall that $E_{n,j}(\gamma^*, t) = S_n^{(1)}(\gamma^*, t) / S_n^{(0)}(\gamma^*, t)$ and $\tilde{E}_{n,j}(\beta^*, t) = \tilde{S}_n^{(1)}(\beta^*, t) / \tilde{S}_n^{(0)}(\beta^*, t)$, where

$$S_n^{(l)}(\gamma^*, t) = n^{-1} \sum_{i=1}^{n} \sum_{v=1}^{n_i} K_n(t - t_{iv})Y_i(t)\{Z_i(t_{iv}, t)\}^\otimes l \exp\{(\gamma^*)^T Z_i(t_{iv}, t)\}$$

$$\tilde{S}_n^{(l)}(\beta^*, t) = n^{-1} \sum_{i=1}^{n} Y_i(t)\{Z_i(t, t)\}^\otimes l \exp\{[\beta^*(t)]^T X_i(t)\}.$$  

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

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

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

We have

$$E_{n,j}(\gamma^*, t) - \bar{E}_{n,j}(\beta^*, t) = E_{n,j}(\gamma^*, t) - \frac{s_n^{(l)}(\gamma^*, t)}{s_n^{(0)}(\gamma^*, t)} + \frac{\bar{s}_n^{(l)}(\beta^*, t)}{\bar{s}_n^{(0)}(\beta^*, t)} - \bar{E}_{n,j}(\beta^*, t)$$

$$+ \frac{s_n^{(1)}(\gamma^*, t)}{s_n^{(0)}(\gamma^*, t)} - \frac{\bar{s}_n^{(1)}(\beta^*, t)}{\bar{s}_n^{(0)}(\beta^*, t)}.$$  

For $L_1$, we have

$$L_1 = \frac{1}{\bar{s}_n^{(0)}(\gamma^*, t)} \{S_n^{(1)}(\gamma^*, t) - s_n^{(1)}(\gamma^*, t)\} - \frac{s_n^{(1)}(\gamma^*, t)}{S_n^{(0)}(\gamma^*, t)s_n^{(0)}(\gamma^*, t)} \{S_n^{(0)}(\gamma^*, t) - s_n^{(0)}(\gamma^*, t)\}.$$  

13
By Lemma S2, with probability no less than $1 - \exp(-Cr_nq_n c_n d_n^{1/2} h_n^2 x^2)$, we have

$$\sup_{t \in [0, \tau]} |L_1| \lesssim (r_n q_n c_n d_n^{1/2} / n)^{1/2} (1 + x).$$  \hfill (S.13)

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

$$\sup_{t \in [0, \tau]} |L_2| \lesssim (r_n / n)^{1/2} (1 + x).$$  \hfill (S.14)

For $L_3$, we have

$$L_3 = \frac{1}{\lambda_v(t) \tilde{s}^{(0)}(\gamma^*, t)} \{ s^{(1)}_{n,j}(\gamma^*, t) - \lambda_v(t) \tilde{s}^{(1)}_j(\gamma^*, t) \}$$

$$- \frac{s^{(1)}_{n,j}(\gamma^*, t)}{\lambda_v(t) \tilde{s}^{(0)}(\beta^*, t) s^{(0)}_n(\gamma^*, t)} \{ s^{(0)}_n(\gamma^*, t) - \lambda_v(t) \tilde{s}^{(0)}_j(\gamma^*, t) \}. $$  \hfill (S.15)

By the same calculation as in (S.10), we have

$$s^{(0)}_n(\gamma^*, t) - \lambda_v(t) \tilde{s}^{(0)}(\gamma^*, t) = O(h_n^2),$$  \hfill (S.16)

$$s^{(1)}_{n,j}(\gamma^*, t) - \lambda_v(t) \tilde{s}^{(1)}_j(\gamma^*, t) = O(h_n^2).$$  \hfill (S.17)
Moreover,

\[
|\tilde{s}^{(0)}(\gamma^*, t) - \tilde{s}^{(0)}(\beta^*, t)| \\
\leq |E \{ Y_i(t) \exp[(\gamma^*)^T Z_i(t, t) - \{\beta^*(t)\}^T X_i(t)]\}| \\
\lesssim E |(\gamma^*)^T Z_i(t, t) - \{\beta^*(t)\}^T X_i(t)| \\
= E \left| \sum_{j=1}^{r_n} \left\{ \beta^*_j(t) - (\gamma^*_j)^T \phi(t) \right\} X_{ij}(t) \right| \\
\lesssim \left| \sum_{j=1}^{r_n} \left\{ \beta^*_j(t) - (\gamma^*_j)^T \phi(t) \right\} \right| \lesssim r_n q_n^{-\alpha},
\]

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

\[
s_n^{(0)}(\gamma^*, t) - \lambda_n(t)\tilde{s}^{(0)}(\gamma^*, t) = O(h_n^2 + r_n q_n^{-\alpha}), \\
s_n^{(1)}(\gamma^*, t) - \lambda_n(t)\tilde{s}^{(1)}_j(\gamma^*, t) = O(h_n^2 + r_n q_n^{-\alpha}).
\]

Then, it follows from (S.15) that

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

Equations (S.13), (S.14) and (S.19) together imply that

\[
P \left( |J_2| \geq D \left\{ (r_n q_n c_n d_n^{1/2}/n)^{1/2} (1 + x) + h_n^2 + r_n q_n^{-\alpha} \right\} \right) \leq C_1 \exp(-C_2 r_n q_n c_n d_n^{1/2} h_n^2 x^2).
\]

(S.20)

Finally, the result follows from (S.11), (S.12) and (S.20).
Lemma S2. Under conditions 1 to 8, there exist positive constants $C$ and $D$ such that, for any $x > 0$,

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

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

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

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

Proof of Lemma S2. Let

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

We prove the upper bound for $W_n$. The other two cases can be shown similarly. First, we bound $E(W_n)$. Let $\mathcal{F} = \{ \sum_{v=1}^{n_1} K_{h_n}(t - t_{iv}) Y(t) \exp{\{\gamma^T Z(t_{iv}, t)\}} : \gamma \in \mathcal{B}_0, t \in [0, \tau] \}$. We calculate the bracketing number of the function class $\mathcal{F}$.

\[
\left| \sum_{v=1}^{n_1} K_{h_n}(t_1 - t_{iv}) Y(t_1) \exp{\{\gamma_1^T Z(t_{iv}, t_1)\}} - \sum_{v=1}^{n_1} K_{h_n}(t_2 - t_{iv}) Y(t_2) \exp{\{\gamma_2^T Z(t_{iv}, t_2)\}} \right|
\]

\[
\leq \sum_{v=1}^{n_1} K_{h_n}(t_1 - t_{iv}) |Y(t_1) \exp{\{\gamma_1^T Z(t_{iv}, t_1)\}} - Y(t_2) \exp{\{\gamma_2^T Z(t_{iv}, t_2)\}}|
\]

\[+ \sum_{v=1}^{n_1} |K_{h_n}(t_1 - t_{iv}) - K_{h_n}(t_2 - t_{iv})| |Y(t_2) \exp{\{\gamma_2^T Z(t_{iv}, t_2)\}}|
\]

\[:= I_1 + I_2.\]
For $I_1$, let $d_{1j} = \gamma_{1,j}^T \phi(t)$ and $d_{2j} = \gamma_{2,j}^T \phi(t)$, we have

$$|Y(t_1) \exp\{\gamma_{1}^T Z(t_{iv}, t_1)\} - Y(t_2) \exp\{\gamma_{2}^T Z(t_{iv}, t_2)\}|$$

$$\lesssim |\gamma_{1}^T Z(t_{iv}, t_1) - \gamma_{2}^T Z(t_{iv}, t_2)| + |Y(t_1) - Y(t_2)|$$

$$\leq |(\gamma_1 - \gamma_2)^T Z(t_{iv}, t_1)| + |\gamma_{2}^T \{Z(t_{iv}, t_1) - Z(t_{iv}, t_2)\}| + |Y(t_1) - Y(t_2)|$$

$$\leq \sum_{j=1}^{r_n} |d_{1j} - d_{2j}| X_j(t_{iv}) + |\gamma_{2}^T [X(t_{iv}) \otimes \{\phi(t_1) - \phi(t_2)\}]| + |Y(t_1) - Y(t_2)|$$

$$\lesssim r_n q_n \|\gamma_1 - \gamma_2\|_{\infty} + r_n q_n^2 |t_1 - t_2| + |Y(t_1) - Y(t_2)|$$

Since $K_{h_n}(t - t_{iv}) = O(h_n^{-1})$ and $n_i = O(1)$, we have $I_1 \lesssim r_n q_n^2 h_n^{-1} (\|\gamma_1 - \gamma_2\|_{\infty} + |t_1 - t_2|) + h_n^{-1} |Y(t_1) - Y(t_2)|$. For $I_2$, by conditions 2 and 4, we have $I_2 \lesssim h_n^{-2} |t_1 - t_2|$. Denote $\theta_1 = (\gamma_1, t_1)^T$ and $\theta_2 = (\gamma_2, t_2)^T$. Then, we have

$$I_1 + I_2 \lesssim c_n \{\|\theta_1 - \theta_2\|_2 + |Y(t_1) - Y(t_2)|\},$$

where $c_n = r_n q_n^2 h_n^{-1} \vee h_n^{-2}$. When $\|\theta_1 - \theta_2\|_2 \leq \epsilon^2 / c_n^2$,

$$|f_{\theta_1} - f_{\theta_2}| \leq \epsilon^2 / c_n + c_n |Y(t_1) - Y(t_2)|,$$

where $f_{\theta_j} := \sum_{i=1}^{n_i} K_{h_n}(t_j - t_{iv}) Y(t_j) \exp\{\gamma_{j}^T Z(t_{iv}, t_j)\}$. The $L_2(P)$-size of the above bracket is

$$2 \epsilon^2 / c_n + 2 c_n \{E|Y(t_1) - Y(t_2)|^2\}^{1/2} = 2 \epsilon^2 / c_n + 2 c_n \left\{ \int_{t_1}^{t_2} dE_{T}(t) \right\}^{1/2} \leq 2 \epsilon^2 / c_n + 2 \epsilon \lesssim \epsilon.$$
Then, to cover $\mathcal{F}$, we need as many brackets as we need balls of radius $\epsilon^2/(2c_n^2)$ to cover $\Theta$, where $\Theta = B_0 \otimes [0, \tau]$. Hence, the bracketing entropy of $\mathcal{F}$ (see Example 19.7 of Van der Vaart (2000)) is

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

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

$$P\{W_n \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2}(1 + x)\} \leq P\{W_n - E(W_n) \geq D(r_n q_n c_n d_n^{1/2}/n)^{1/2} x\} \leq \exp(-Cr_n q_n c_n d_n^{1/2} h_n^2 x^2).$$

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

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

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

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

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

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

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

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

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

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

$$|Y(t_1) \exp\{\{\beta^*(t_1)\}^T X(t_1)\} - Y(t_2) \exp\{\{\beta^*(t_2)\}^T X(t_2)\}|$$

$$\lesssim |Y(t_1) - Y(t_2)| + |\exp\{\{\beta^*(t_1)\}^T X(t_1)\} - \exp\{\{\beta^*(t_2)\}^T X(t_2)\}|$$

$$\lesssim |Y(t_1) - Y(t_2)| + \left|\{\beta^*(t_1)\}^T X(t_1) - \{\beta^*(t_2)\}^T X(t_2)\right|$$

$$\lesssim |Y(t_1) - Y(t_2)| + \sum_{j=1}^{r_n} |\beta^*_j(t_1) - \beta^*_j(t_2)| + \sum_{j=1}^{r_n} |X_j(t_1) - X_j(t_2)|.$$

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
which forms a grid of points $0 = t_0 < t_1 < \cdots < t_k = \tau$. The $L_2$-size of these brackets is $\epsilon$. By the continuity assumption of $\beta_j^*(t)$ in condition 6, to cover $\{\beta_j^*(t) : t \in [0, \tau]\}$, we need as many $\epsilon$-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

$$\log N_0(\epsilon, \mathcal{F}, L_2(p)) \lesssim r_n \log(\epsilon^{-1}).$$

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

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

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

$$P \left\{ W_n \geq D(r_n/n)^{1/2}(1 + x) \right\} \leq P \left\{ W_n - E[W_n] \geq D(r_n/n)^{1/2}x \right\} \leq \exp(-Cr_nx^2).$$

Lemma S4. Under conditions 1 to 8, there exist positive constants $C_1, C_2$ and $D$, such that for any $x > 0$,

$$P \left( \sup_{\gamma \in B_0} |I_{n,ij}(\gamma) - \Sigma_{ij}\gamma| \geq D\left( (r_nq_n/cn d_n^{1/2}/n)^{1/2} (1 + x) + h_n^2 \right) \right) \leq C_1 \exp(-C_2r_nq_n/cn d_n^{1/2}h_n^2x^2).$$
Proof of Lemma S4. Note that

\[ I_{n,ij}(\gamma, t) - \Sigma_{ij}(\gamma, t) \]
\[ = \int_0^\tau \left\{ S_{n,ij}^{(2)}(\gamma, t) - \lambda_v(t)\bar{s}_{ij}^{(2)}(\gamma, t) \right\} d\Lambda_0(t) \]
\[ - \int_0^\tau \left\{ \frac{S_{n,ij}^{(1)}(\gamma, t)S_{n,j}^{(1)}(\gamma, t)}{S_n^{(0)}(\gamma, t)} - \frac{\bar{s}_{i}^{(1)}(\gamma, t)\bar{s}_{j}^{(1)}(\gamma, t)}{\bar{s}^{(0)}(\gamma, t)} \lambda_v(t) \right\} d\Lambda_0(t) \]
\[ := J_1(\gamma) - \int_0^\tau j_{2,n}(\gamma, t) d\Lambda_0(t). \]

For the term \( J_1(\gamma) \), we have

\[ |J_1(\gamma)| \leq \sup_{t \in [0,\tau]} |S_{n,ij}^{(2)}(\gamma, t) - \lambda_v(t)\bar{s}_{ij}^{(2)}(\gamma, t)| \cdot \Lambda_0(\tau). \] (S.21)

Similar as in (S.16) and (S.17), we have

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

This together with Lemma S2 imply that

\[ P \left( \sup_{\gamma \in B_\delta, t \in [0,\tau]} |S_{n,ij}^{(2)}(\gamma, t) - \lambda_v(t)\bar{s}_{ij}^{(2)}(\gamma, t)| \geq D_2\{ (r_nq_nc_n d_n^{1/2}/n)^{1/2}(1 + x) + h_n^2 \} \right) \]
\[ \leq \exp(-C_1r_nq_nc_n d_n^{1/2} h_n^2 x^2). \] (S.22)

Then, by (S.21), we have

\[ P \left( \sup_{\gamma \in B_\delta} |J_1(\gamma)| \geq D_1\{ (r_nq_nc_n d_n^{1/2}/n)^{1/2}(1 + x) + h_n^2 \} \right) \leq \exp(-C_1r_nq_n c_n d_n^{1/2} h_n^2 x^2). \] (S.23)
For the second term, we write \( j_{2,n}(\gamma, t) \) as

\[
j_{2,n}(\gamma, t) = \frac{S^{(1)}_{n,j}(\gamma, t)}{S^{(0)}_{n}(\gamma, t)} \{ S^{(1)}_{n,j}(\gamma, t) - \lambda_v(t) \hat{s}^{(1)}_j(\gamma, t) \} + \frac{\lambda_v(t) \hat{s}^{(1)}_j(\gamma, t)}{S^{(0)}_{n}(\gamma, t)} \{ S^{(1)}_{n,i}(\gamma, t) - \lambda_v(t) \hat{s}^{(1)}_i(\gamma, t) \}
\]

\[
- \frac{\lambda_v(t) \hat{s}^{(1)}_i(\gamma, t) \hat{s}^{(1)}_j(\gamma, t)}{S^{(0)}_{n}(\gamma, t) \hat{s}^{(0)}(\gamma, t)} \{ S^{(0)}_{n}(\gamma, t) - \lambda_v(t) \hat{s}^{(0)}(\gamma, t) \}.
\]

Since \( \lambda_v(t) \), \( S^{(1)}_{n,i}(\gamma, t) \), \( \hat{s}^{(1)}_j(\gamma, t) \) are all bounded and \( S^{(0)}_{n}(\gamma, t) \) and \( \hat{s}^{(0)}(\gamma, t) \) are bounded away from zero, it follows that

\[
\sup_{\gamma \in B_0} |J_2(\gamma)| \lesssim \sup_{\gamma \in B_0, t \in [0, \tau]} |j_{2,n}(\gamma, t)|
\]

\[
\lesssim \sup_{\gamma \in B_0} |S^{(1)}_{n,j}(\gamma, t) - \lambda_v(t) \hat{s}^{(1)}_j(\gamma, t)| + \sup_{\gamma \in B_0, t \in [0, \tau]} |S^{(0)}_{n}(\gamma, t) - \lambda_v(t) \hat{s}^{(0)}(\gamma, t)|.
\]

Similar as (S.22), we have

\[
P \left( \sup_{\gamma \in B_0} |J_2(\gamma)| \geq D_2 \{ (r_n q_n c_n d_n^{1/2} / n)^{1/2} (1 + x) + h_n^2 \} \right) \leq \exp(-C_2 r_n q_n c_n d_n^{1/2} h_n^2 x^2). \quad (S.24)
\]

(S.23) and (S.24) together complete the proof. \( \square \)

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

\[
P \left\{ \inf_{\beta \in B_0} \lambda_{\min}(I_{n,AA}(\gamma)) \leq \frac{C_{\min}}{2} \right\} \leq C_1 r_n^2 q_n^2 \exp\{-C_2 n h_n^2\}, \quad (S.25)
\]
and

\[
P \left\{ \sup_{\gamma \in B_0} \| I_{n,A^c,A}(\gamma) I_{n,A\Lambda}(\gamma)^{-1} \|_{\infty} \geq \frac{1}{2} (1 - \zeta) \frac{\rho'(0^+)}{\rho'(d_n/2)} \right\} \leq C_3 p_n q_n \exp \left\{ -C_4 n h_n^2 (r_n q_n)^{-1} \right\}.
\]

Proof of Lemma S5. By Weyl's inequality,

\[
|\lambda_{\min}(I_{n,A\Lambda}(\gamma)) - \lambda_{\min}(\Sigma_{A\Lambda}(\gamma))| \leq \| I_{n,A\Lambda}(\gamma) - \Sigma_{A\Lambda}(\gamma) \|_2 \leq \| I_{n,A\Lambda}(\gamma) - \Sigma_{A\Lambda}(\gamma) \|_1.
\]

By condition 9,

\[
\inf_{\gamma \in B_0} \lambda_{\min}(\Sigma_{A\Lambda}(\gamma)) = 1/\{ \sup_{\gamma \in B_0} \lambda_{\max}(\Sigma_{A\Lambda}(\gamma)) \} \geq 1/(\sup_{\gamma \in B_0} \| \Sigma_{A\Lambda}(\gamma) \|_{\infty}) \geq 1/M. \quad (S.27)
\]

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

\[
P \left\{ \sup_{\gamma \in B_0} |\lambda_{\min}(I_{n,A\Lambda}(\gamma)) - \lambda_{\min}(\Sigma_{A\Lambda}(\gamma))| \geq \frac{C_{\min}}{2} \right\}
\]

\[
\leq P \left\{ \sup_{\gamma \in B_0} \| I_{n,A\Lambda}(\gamma) - \Sigma_{A\Lambda}(\gamma) \|_{\infty} \geq \frac{C_{\min}}{2} \right\}
\]

\[
\leq P \left\{ \sup_{\gamma \in B_0} \max_{i \in A} \sum_{j \in A} |I_{n,ij}(\gamma) - \Sigma_{ij}(\gamma)| \geq \frac{C_{\min}}{2} \right\} \leq r_n^2 q_n^2 P \left\{ \sup_{\gamma \in B_0} |I_{n,ij}(\gamma) - \Sigma_{ij}(\gamma)| \geq \frac{C_{\min}}{2} \right\}
\]

\[
\leq C_1 r_n^2 q_n^2 \exp \left\{ -C_2 (n h_n^2 \vee r_n q_n c_n d_n^{1/2} h_n^2) \right\} = C_1 r_n^2 q_n^2 \exp \left\{ -C_2 n h_n^2 \right\}.
\]

This result together with (S.27) imply (S.25).
To prove (S.26), observe that

\[
I_{n, A^c; A}(\gamma)I_{n, A; A}(\gamma)^{-1} - \Sigma_{A^c; A}(\gamma)\Sigma_{A; A}(\gamma)^{-1}
= \{I_{n, A^c; A}(\gamma) - \Sigma_{A^c; A}(\gamma)\}I_{n, A; A}(\gamma)^{-1}
- \Sigma_{A^c; A}(\gamma)\Sigma_{A; A}(\gamma)^{-1}\{I_{n, A; A}(\gamma) - \Sigma_{A; A}(\gamma)\}I_{n, A; A}(\gamma)^{-1}
:= J_1(\gamma) + J_2(\gamma).
\]

For \( J_1(\gamma) \), it follows from Lemma S4 that

\[
P\left\{ \sup_{\gamma \in B_0} \| I_{n, A^c; A}(\gamma) - \Sigma_{A^c; A}(\gamma) \|_\infty \geq \frac{(1 - \zeta)C_{\min}}{8\sqrt{r_nq_n}} \right\}
\leq P\left\{ \sup_{\gamma \in B_0} \max_{i \in A^c} \sum_{j \in A} |I_{n, ij}(\gamma) - \Sigma_{ij}(\gamma)| \geq \frac{(1 - \zeta)C_{\min}}{8\sqrt{r_nq_n}} \right\}
\leq \sum_{i \in A^c} P\left\{ \sup_{\gamma \in B_0} |I_{n, ij}(\gamma) - \Sigma_{ij}(\gamma)| \geq \frac{(1 - \zeta)C_{\min}}{8\sqrt{r_nq_n}} \right\}
\leq C_3(p_n - r_n)r_nq_n^2 \exp\{-C_4n^2h_n^2(r_nq_n)^{-1}\}.
\]

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

\[
P\left\{ \sup_{\gamma \in B_0} \| I_{n, A; A}(\gamma)^{-1} \|_\infty \geq \frac{2\sqrt{r_nq_n}}{C_{\min}} \right\}
\leq P\left\{ \inf_{\gamma \in B_0} \lambda_{\min}(I_{n, A; A}(\gamma)) \leq \frac{C_{\min}}{2} \right\}
\leq C_1r_n^2q_n^2 \exp\{-C_2n^2h_n^2\}.
\]

Therefore, by the union bound, (S.30) and (S.31) together imply that

\[
P\left\{ \sup_{\gamma \in B_0} |J_1(\gamma)| \geq \frac{1 - \zeta}{4} \right\}
\leq P\left\{ \sup_{\gamma \in B_0} |J_1(\gamma)| \geq \frac{1 - \zeta}{4} \right\}
\leq C_1r_n^2q_n^2 \exp\{-C_2n^2h_n^2\} + C_3(p_n - r_n)r_nq_n^2 \exp\{-C_4n^2h_n^2(r_nq_n)^{-1}\},
\]

24
since $\rho'(0+)/\rho'(d_n/2) \geq 1$ by the concavity assumption in condition 11.

Similar as (S.30), we have

$$P \left\{ \sup_{\gamma \in B_0} \| I_{n,A}(\gamma) - \Sigma_{AA}(\gamma) \|_\infty \geq \frac{C_{\min}}{8 \sqrt{r_n q_n}} \right\} \leq C_3 r_n^2 q_n^2 \exp \{-C_4 n h_n^2 (r_n q_n)^{-1} \}.$$  

This together with (S.31) imply that

$$P \left\{ \sup_{\gamma \in B_0} |J_2(\gamma)| \geq \frac{(1 - \zeta) \rho'(0+)}{4 \rho'(d_n/2)} \right\} \leq C_1 p_n r_n^2 q_n^2 \exp \{-C_2 n h_n^2 \} + C_3 r_n^2 q_n^2 \exp \{-C_4 n h_n^2 (r_n q_n)^{-1} \}.$$  

Finally, it follows from (S.29), (S.32) and (S.33) that

$$P \left\{ \| I_{n,A}(\gamma) - \Sigma_{AA}(\gamma) \| \geq \frac{(1 - \zeta) \rho'(0+)}{2 \rho'(d_n/2)} \right\} \leq C_3 p_n r_n^2 q_n^2 \exp \{-C_4 n h_n^2 (r_n q_n)^{-1} \}.$$  

This together with condition 10 complete the proof.  

\[ \square \]

**S5 Additional simulation results**

Figure S1 shows the running time of the proposed method with $\ell_0$-regularization penalty based on $\lambda$ with length of 10 and fixed $\alpha$ 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.

Table S1 summarizes the comparison results by using different kernel functions for both
Figure S1: Running time in seconds of the proposed $\ell_0$Net for various sample sizes and number of covariates
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 S1: Comparison of estimation and selection performance of the proposed DB-hazard using different kernel functions under various penalty functions.

<table>
<thead>
<tr>
<th>Setting (a)</th>
<th>Epanechnikov</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gLasso\textsuperscript{†}</td>
<td>gNet\textsuperscript{‡}</td>
</tr>
<tr>
<td>$n = 100, \ p_n = 1000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSE\textsuperscript{1}</td>
<td>8.34</td>
<td>6.25</td>
</tr>
<tr>
<td>TP\textsuperscript{2}</td>
<td>7.7</td>
<td>8.0</td>
</tr>
<tr>
<td>FP\textsuperscript{3}</td>
<td>33.2</td>
<td>127.2</td>
</tr>
<tr>
<td>$n = 200, \ p_n = 1000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>5.04</td>
<td>4.17</td>
</tr>
<tr>
<td>TP</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>FP</td>
<td>57.1</td>
<td>149.0</td>
</tr>
</tbody>
</table>

| 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 |

\textsuperscript{†}: group Lasso; \textsuperscript{‡}: group Lasso with a Laplacian penalty; \textsuperscript{*}: $\ell_0$-regularization penalty (10)

Table S2 summarizes the performance of bandwidth selection. It can be 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 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
Table S2: Performance of the bandwidth selection procedure for DB-hazard using different kernel functions.

<table>
<thead>
<tr>
<th>Setting (a)</th>
<th>Epanechnikov</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Selected</td>
<td>Best(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>0.056</td>
<td>0.085</td>
</tr>
<tr>
<td>SSE(^2)</td>
<td>4.57</td>
<td>3.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>0.059</td>
<td>0.086</td>
</tr>
<tr>
<td>SSE</td>
<td>2.83</td>
<td>2.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Setting (b)</th>
<th>Epanechnikov</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Selected</td>
<td>Best(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>0.055</td>
<td>0.113</td>
</tr>
<tr>
<td>SSE</td>
<td>12.59</td>
<td>11.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>0.061</td>
<td>0.104</td>
</tr>
<tr>
<td>SSE</td>
<td>8.06</td>
<td>6.90</td>
</tr>
</tbody>
</table>

\(^{[1]}\): defined as the bandwidth leading to the smallest SSE; \(^{[2]}\): sum of squared errors.

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.

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

<table>
<thead>
<tr>
<th>Setting (a)</th>
<th>Setting (b)</th>
<th>gLasso†</th>
<th>gNet‡</th>
<th>(\ell_0)Net∗</th>
<th>gLasso</th>
<th>gNet</th>
<th>(\ell_0)Net</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n = 100, (p_n = 1000), (q_n = 8)</td>
<td>n = 100, (p_n = 1000), (q_n = 10)</td>
<td>n = 100, (p_n = 1000), (q_n = 13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>7.8</td>
<td>8.0</td>
<td>8.0</td>
<td>2.0</td>
<td>3.0</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>98.9</td>
<td>327.9</td>
<td>14.9</td>
<td>48.1</td>
<td>115.6</td>
<td>34.5</td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>10.43</td>
<td>8.94</td>
<td>7.88</td>
<td>14.59</td>
<td>14.55</td>
<td>14.25</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>7.9</td>
<td>8.0</td>
<td>8.0</td>
<td>1.9</td>
<td>3.3</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>93.7</td>
<td>339.8</td>
<td>23.2</td>
<td>50.2</td>
<td>160.8</td>
<td>39.7</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>7.9</td>
<td>8.0</td>
<td>8.0</td>
<td>2.6</td>
<td>3.8</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>169.7</td>
<td>900.8</td>
<td>52.0</td>
<td>96.3</td>
<td>258.0</td>
<td>70.4</td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>6.61</td>
<td>5.43</td>
<td>4.12</td>
<td>12.07</td>
<td>11.49</td>
<td>9.91</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>5.0</td>
<td>6.7</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>149.4</td>
<td>416.6</td>
<td>6.6</td>
<td>125.9</td>
<td>329.9</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>7.31</td>
<td>6.05</td>
<td>5.05</td>
<td>12.63</td>
<td>12.33</td>
<td>11.04</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>5.1</td>
<td>6.3</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>149.9</td>
<td>458.3</td>
<td>10.4</td>
<td>103.7</td>
<td>311.2</td>
<td>22.8</td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>8.05</td>
<td>7.59</td>
<td>6.74</td>
<td>13.31</td>
<td>13.18</td>
<td>12.31</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>5.9</td>
<td>6.6</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>262.2</td>
<td>917.8</td>
<td>18.4</td>
<td>199.8</td>
<td>584.3</td>
<td>40.4</td>
<td></td>
</tr>
</tbody>
</table>

†: group Lasso; ‡: group Lasso with a Laplacian penalty; ∗: \(\ell_0\)-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.

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

(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 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 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.

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.
References


