Supplementary Materials for "Globally Adaptive Longitudinal Quantile Regression with High Dimensional Compositional Covariates"

Huijuan Ma, Qi Zheng, Zhumin Zhang, Huichuan Lai and Limin Peng

East China Normal University, University of Louiswille,

University of Wisconsin-Madison, University of Wisconsin-Madison,

Emory University

S1 Notation and regularity conditions

Let $M_n(\tau, \boldsymbol{\delta}) = n^{1/2}\mathbb{E}_n \int_0^\infty \mathbf{V}_i(t)\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau) + \boldsymbol{\delta})\}dN_i(t)$, where $\psi_\tau(v) = \tau - I(v < 0)$, and define $D_i^I(\tau, \boldsymbol{\delta}) := I\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau) < 0\} - I\{Y_i(t) - \mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau) +\boldsymbol{\delta}) < 0\}$. Recall that $R_{r+s-1}(B) = \{\boldsymbol{\delta} \in R_{r+s-1} : \|\boldsymbol{\delta}\| \le \sqrt{(r+s)n^{-1}\log n}B\}$, for any $B > 0$.

We impose the following regularity conditions for our technical derivations:

(C1) (Condition on the number of longitudinal observations) $N_i(t)$ is independent of $\mathbf{V}_i(t)$ and $m_i = \int_0^\infty dN_i(t)$ is bounded by some finite constant M_0, $i = 1, \ldots, n$.

(C2) (Condition on the conditional density) Let $f_{t,\tau}(\cdot|\mathbf{v})$ denote the probability density

Corresponding author: lpeng@emory.edu (Limin Peng). The first two authors have equal contributions.

function of $Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)$ given $\mathbf{V}_i(t) = \mathbf{v}$. There exists some constants $\bar{f}$, $\underline{f} > 0$, such that $\underline{f} \leq \inf_{\tau\in\Delta,t,\mathbf{v}} f_{t,\tau}(0|\mathbf{v}) \leq \sup_{\tau\in\Delta,t,\mathbf{v}} f_{t,\tau}(\cdot|\mathbf{v}) < \bar{f}$. In addition, there exists a constant $C_f > 0$, such that $\sup_{\tau\in\Delta,t,\mathbf{v}} |f_{t,\tau}(u|\mathbf{v}) - f_{t,\tau}(0|\mathbf{v})| \leq C_f|u|$.

(C3) (Conditions on covariates) The covariates $\mathbf{V}_i(t)$ are bounded in the sense of $\|\mathbf{V}_i(t)\|_\infty < C_V$ for some finite constant C_V.

(C4) (Condition on the true regression coefficients $\boldsymbol{\gamma}_0(\tau)$) There exists a positive constant L, such that $\|\boldsymbol{\gamma}_0(\tau_1) - \boldsymbol{\gamma}_0(\tau_2)\|_\infty \leq L|\tau_1 - \tau_2|$, for all $\tau_1, \tau_2 \in \Delta$.

(C5) (Conditions on the identifiability of the true model) Let $R_{r+s-1} := \{\boldsymbol{\delta} = (\boldsymbol{\delta}_\mathbf{x}^\top, \boldsymbol{\delta}_\mathbf{z}^\top)^\top : \boldsymbol{\delta}_\mathbf{x} \in \mathbb{R}^r, \boldsymbol{\delta}_\mathbf{z} \in \mathbb{R}^p, \sum_{j=1}^s \boldsymbol{\delta}_\mathbf{z}^{(j)} = 0, \boldsymbol{\delta}_\mathbf{z}^{(l)} = 0, l = s+1, \ldots, p\}$. There exists some constants $\lambda_{\min}$ and $\lambda_{\max}$ such that $\lambda_{\min}\|\boldsymbol{\delta}_a\|^2 \leq \boldsymbol{\delta}_a^\top E[\int_0^\infty \mathbf{V}_{ia}(t)\mathbf{V}_{ia}(t)^\top dN_i(t)]\boldsymbol{\delta}_a \leq \lambda_{\max}\|\boldsymbol{\delta}_a\|^2$. Moreover,

$$q = \inf_{\boldsymbol{\delta}\in R_{r+s-1}, \boldsymbol{\delta}\neq 0} \frac{E[\int_0^\infty |\mathbf{V}_i(t)^\top \boldsymbol{\delta}|^2 dN_i(t)]^{3/2}}{E[\int_0^\infty |\mathbf{V}_i(t)^\top \boldsymbol{\delta}|^3 dN_i(t)]} > 0.$$

Condition (C1) is a mild condition, which assumes the number of longitudinal observations is bounded and the observation times are independent of the covariate process. Conditions (C2)–(C5) are common conditions in high dimensional quantile regression (for example, Belloni and Chernozhukov, 2011; Zheng et al., 2015). Condition (C2) imposes mild conditions on the conditional density. Condition (C3) assumes the boundedness of covariates in $\mathbf{X}_i(t)$ and $\mathbf{Z}_i(t)$. Condition (C4) imposes the true coefficient function are Lipschitz continuous. Condition (C5) implies that the eigenvalues of $E[\int_0^\infty \mathbf{V}_{ia}(t)\mathbf{V}_{ia}(t)^\top dN_i(t)]$ are bounded from below and above by $\lambda_{\min}$ and $\lambda_{\max}$, respectively. It is analogous to the sparse Riesz condition (Zhang and Huang, 2008)

S2 Technical lemmas and their proofs

In our theoretical studies, we utilize the following technical lemmas.

Lemma 1. *Suppose conditions (C1)–(C5) hold. If* $u < 3q\underline{f}\lambda_{\min}/(4C_f\lambda_{\max}^{3/2})$, *then*

$$\inf_{\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|=u} E[D_i(\boldsymbol{\delta},\tau)] \geq \underline{f}\lambda_{\min}u^2/8,$$

where $\underline{f}$ *and* C_f *are defined in condition (C2) and* $\lambda_{\min}, \lambda_{\max}$, *and* q *are defined in condition (C5).*

Lemma 2. *Under conditions (C1)–(C5), for any* n *satisfying* $\sqrt{r+s}\Big(24\sqrt{\log n} + 200$ $\sqrt{\log n + \log L/2} > 8M_0\lambda_{\max}$,

$$\begin{aligned} &\sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|\leq u} |\mathbb{G}_n\{D_i(\boldsymbol{\delta},\tau)\}| \\ \leq\ & M_0C_V\sqrt{r+s}u\Big(24\sqrt{\log n} + 200\sqrt{\log n + \log L/2 - \log u/2}\Big) \end{aligned}$$

holds with probability at least $1-16n^{-3}$, *where* C_V *and* L *are defined in conditions (C3) and (C4), respectively.*

Lemma 3. *Suppose conditions (C1)–(C5) hold. For any fixed* τ_0 *and any constant* $a > 8$,

$$\sup_{\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|\leq u} |\mathbb{G}_n\{D_i(\boldsymbol{\delta},\tau_0)\}| \leq 12aM_0\sqrt{\lambda_{\max}(r+s)}u$$

holds with probability at least $1-8a^{-1}$.

Lemma 4. *Suppose conditions (C1)–(C4) hold. For any given* $\boldsymbol{\xi} \in R_{r+s-1}$, $\|\boldsymbol{\xi}\| = 1$, *if* $(r+s)^3 \log^4 n = o(n)$, *then*

$$\sup_{\boldsymbol{\delta} \in R_{r+s-1}(B), \tau \in \Delta} |\boldsymbol{\xi}^\top [M_n(\tau, \boldsymbol{\delta}) - E\{M_n(\tau, \boldsymbol{\delta})\} - M_n(\tau, \mathbf{0}) + E\{M_n(\tau, \mathbf{0})\}]| = o_p(1).$$

Lemma 5. *Suppose conditions (C1), (C2) and (C5) hold. For any given* $\boldsymbol{\xi} \in R_{r+s-1}$, $\|\boldsymbol{\xi}\| = 1$, $\boldsymbol{\xi}^\top M_n(\tau, \mathbf{0})$ *converges weakly to a mean zero Gaussian process with the covariance function*

$$E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t), Y) h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t), Y)\} - E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t), Y)\} E\{h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t), Y)\},$$

where $h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t), Y) := \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t) \psi_\tau\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau)\} dN(t)$.

Lemma 6. *Suppose conditions (C1)–(C5) hold, and* $n/((r+s)^3 \log^2 \max\{n, r+p\}) \to \infty$. *For some constant* $C_2 > 0$, *we have*

$$\sup_{\tau \in \Delta, \boldsymbol{\delta} \in R_{r+s-1}(B)} \max_{j > r+s} n^{1/2} \boldsymbol{\eta}_j^\top [M_n(\tau, \boldsymbol{\delta}) - E\{M_n(\tau, \boldsymbol{\delta})\} - M_n(\tau, \mathbf{0}) + E\{M_n(\tau, \mathbf{0})\}]$$
$$\leq C_2 n^{1/4} (r+s)^{3/4} \log \max\{n, r+p\}$$

with probability at least $1 - 16\exp(-(r+s)\log\max\{n, r+p\}/2)$, *where* $\boldsymbol{\eta}_j$ *is a* $(r+p)$ *dimensional vector whose* j*th component is 1, all other components are 0.*

Lemma 7. *Under condition (C2), if* $E[\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)| dN_i(t)]$ *exists for some* $\tau_0 \in \Delta$, *then we have*

$$\tau^* \Big[E\Big[\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)| dN_i(t) \Big] - M_0 \underline{f}^{-1} \Big]$$

$$\leq \inf_{\tau\in\Delta}\sigma(\tau) \leq \sup_{\tau\in\Delta}\sigma(\tau) \leq E\Big[\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\Big] + M_0\underline{f}^{-1},$$

almost surely, where $\sigma(\tau) = \mathbb{E}_n\left(\int_0^\infty \rho_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)\right)$ *and* $\tau^* = \min\{\tau_L, 1-\tau_U\}$.

Lemma 8. *Suppose conditions (C1)–(C4), and (C5+) hold. For some* $r+1 \leq k \leq \kappa = o(n/\log(r+p))$, *some constant* $A > 2$,

$$\sup_{\substack{\boldsymbol{\delta}\in R_\ell,\|\boldsymbol{\delta}\|\leq u,\\ \ell\leq k,\tau\in\Delta}} \sum_{i=1}^n D_i(\boldsymbol{\delta},\tau) \geq \frac{\underline{f}\Lambda_{\min}}{8}nu^2 - A\sqrt{kn}u\Big(24\sqrt{\log\max\{n, r+p\}} + 200\sqrt{\log\max\{n,r+p\} + \log L/2 - \log u/2}\Big),$$

with probability no less than $1 - 8\exp(-(A^2-1)\log\max\{n, r+p\}) - 8\exp(-(A^2 - 3/2)\log\max\{n,r+p\} + \log L - \log u)$. *Consequently,*

$$\sup_{S_\Delta\subseteq S,|S|=k}\ \sup_{\tau\in\Delta}\|\hat{\boldsymbol{\gamma}}_S(\tau) - \boldsymbol{\gamma}_0(\tau)\| \leq 3584A\sqrt{k\log(\max\{r+p,n\})/n}/(\underline{f}\Lambda_{\min})$$

holds with probability at least $1 - 16\exp(-(A^2-2)\log\max\{n, r+p\})$.

Lemma 9. *Under the same conditions as in Theorem 3, for some constants* C_b *and* $A > 2$, *we have*

$$\inf_{S\in OF,|S|=k,\tau\in\Delta}\hat{\sigma}_S(\tau) - \hat{\sigma}_{S_\Delta}(\tau) \geq -8C_b(\underline{f}\Lambda_{\min})^{-1}(k-r-s)n^{-1}\log\max\{n, r+p\}$$

with probability at least $1 - 16\exp(-(A^2-2)\log\max\{n, r+p\}) - 2\exp(-3\log\max\{n, r+p\})$.

Lemma 10. *Under the same conditions as in Theorem 3, for some constant $A > 2$, we have*

$$\begin{aligned}&P\left(\inf_{S\in OF,|S|\le\kappa} GIC(S) > GIC(S_\Delta)\right)\\ \ge&1-16\exp(-(A^2-2)\log\max\{n,r+p\})-6/\max\{n,r+p\}-16/n^3.\end{aligned}$$

Lemma 11. *Under the same conditions as in Theorem 3, for some constant $A > 2$, we have*

$$\begin{aligned}&P\left(\inf_{S\in UF,|S|\le\kappa} GIC(S) > GIC(S_\Delta)\right)\\ \ge&\ 1-16\exp(-(A^2-2)\log\max\{n,r+p\})-16/n^3.\end{aligned}$$

We first state a fact, which will be often used in our proofs. Given any $\boldsymbol{\delta} \in R_{r+s-1}$ and $1 \le k \le s$,

$$\begin{aligned}\mathbf{V}_i^\top(t)\boldsymbol{\delta} = \mathbf{V}_{ia}(t)^\top\boldsymbol{\delta}_a &= \sum_{j=1}^r \delta_{\mathbf{x}j}X_{ij}(t) + \sum_{l=1}^s \delta_{\mathbf{z}l}Z_{il}(t)\\ &= \sum_{j=1}^r \delta_{\mathbf{x}j}X_{ij}(t) + \sum_{l=1}^s \delta_{\mathbf{z}l}(Z_{il}(t)-Z_{ik}(t)) + \sum_{l=1}^s \delta_{\mathbf{z}l}Z_{ik}(t)\\ &= \sum_{j=1}^r \delta_{\mathbf{x}j}X_{ij}(t) + \sum_{l=1}^s \delta_{\mathbf{z}l}(Z_{il}(t)-Z_{ik}(t)) = \mathbf{V}_{ia}^k(t)^\top\boldsymbol{\delta}_{a,\setminus k}.\end{aligned}$$

Proofs of lemmas

Proof of Lemma 1. We note that if $\boldsymbol{\delta} \in R_{r+s-1}$, $\boldsymbol{\delta} = (\boldsymbol{\delta}_a^\top, \mathbf{0}^\top)^\top$. We claim that there exists a $1 \le l \le s$, such that $\|\boldsymbol{\delta}_{a,\setminus l}\| \ge (1-1/s)^{1/2}u$. Suppose our claim does not hold, i.e., $\|\boldsymbol{\delta}_{a,\setminus l}\| < (1-1/s)^{1/2}u, \forall 1 \le l \le s$. Then $\delta_{\mathbf{z}l}^2 = u^2 - \|\boldsymbol{\delta}_{a,\setminus l}\|^2 > u^2/s, \forall 1 \le l \le s$, and consequently $u^2 < \sum_{j=1}^{s} \delta_{\mathbf{z}j}^2 \leq \|\boldsymbol{\delta}_a\|^2 \leq u^2$. Contradiction! Without loss of generality, we assume that $\|\boldsymbol{\delta}_{a,\setminus s}\| \geq (1-1/s)^{1/2}u$.

It follows from Knight and Fu (2000) that

$$\rho_\tau(x-y) - \rho_\tau(x) = y\left\{I(x<0)-\tau\right\} + \int_0^y \left\{I(x<u) - I(x<0)\right\} du$$

for any $x \neq 0$. Denote $F_{t,\tau}\{\cdot|\mathbf{V}_i(t)\}$ as the cumulative conditional distribution function of $Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)$ given $\mathbf{V}_i(t)$. The assumption that $N_i(t)$ is independent of $\mathbf{V}_i(t)$ and the fact $E[I\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau) < 0\}|\mathbf{V}_i(t)] = \tau$ together yield

$$E\left[D_i(\boldsymbol{\delta},\tau)\right] = E\left\{\int_0^\infty \left(\int_0^{\mathbf{V}_i(t)^\top\boldsymbol{\delta}} [F_{t,\tau}\{u|\mathbf{V}_i(t)\} - F_{t,\tau}\{0|\mathbf{V}_i(t)\}]du\right) dN_i(t)\right\}.$$

Hence,

$$\begin{aligned}
E\left[D_i(\boldsymbol{\delta},\tau)\right] &\geq E\Big[\int_0^\infty \int_0^{|\mathbf{V}_{ia}(t)^\top\boldsymbol{\delta}_a|} \big[f_{t,\tau}\{0|\mathbf{V}_i(t)\}u - C_f u^2\big] du\; dN_i(t)\Big] \\
&= E\Big\{\int_0^\infty \int_0^{|\mathbf{V}_{ia}^s(t)^\top\boldsymbol{\delta}_{a,\setminus s}|} \big[f_{t,\tau}\{0|\mathbf{V}_i(t)\}u - C_f u^2\big] du\; dN_i(t)\Big\} \\
&\geq \underline{f} E\Big\{\int_0^\infty |\mathbf{V}_{ia}^s(t)^\top\boldsymbol{\delta}_{a,\setminus s}|^2 dN_i(t)\Big\}/2 - C_f E\Big\{\int_0^\infty |\mathbf{V}_{ia}^s(t)^\top\boldsymbol{\delta}_{a,\setminus s}|^3 dN_i(t)\Big\}/3 \\
&\geq \frac{\underline{f}}{2} E\Big\{\int_0^\infty |\mathbf{V}_{ia}^s(t)^\top\boldsymbol{\delta}_{a,\setminus s}|^2 dN_i(t)\Big\} - \frac{C_f}{3q}\left[E\left\{\int_0^\infty |\mathbf{V}_{ia}^s(t)^\top\boldsymbol{\delta}_{a,\setminus s}|^2 dN_i(t)\right\}\right]^{\frac{3}{2}} \\
&\geq \frac{\underline{f}\lambda_{\min}}{2}\frac{s-1}{s}u^2 - \frac{C_f\lambda_{\max}^{3/2}}{3q}\left(\frac{s-1}{s}\right)^{3/2} u^3, \qquad \text{(S2.1)}
\end{aligned}$$

where the first inequality follows from the mean value theorem and condition (C1), the last two inequalities follow from condition (C5) and the claim we have verified at the beginning of the proof. When $u < 3q\underline{f}\lambda_{\min}/(4C_f\lambda_{\max}^{3/2})$, it follows from (S2.1) that $E[D_i(\boldsymbol{\delta},\tau)] \geq \underline{f}\lambda_{\min}u^2/8$. This completes the proof of Lemma 1. □

Proof of Lemma 2. Define $\mathcal{A}(u) := \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|\le u} |\mathbb{G}_n\{D_i(\boldsymbol{\delta},\tau)\}|$. We first find the upper bound for $\mathcal{A}(u)$. For any $\boldsymbol{\delta}\in R_{r+s-1}, \|\boldsymbol{\delta}\|\le u$,

$$\begin{aligned}
&\mathrm{Var}\{D_i(\boldsymbol{\delta},\tau)\} \le E[\{D_i(\boldsymbol{\delta},\tau)\}^2]\\
= \;& E\Bigg[\Big\{\int_0^\infty \Big(\rho_\tau\big[Y_i(t)-\mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta}\}\big]-\rho_\tau\big\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\big\}\Big)dN_i(t)\Big\}^2\Bigg]\\
\le \;& E\Bigg[\Big\{\int_0^\infty \Big(\rho_\tau\big[Y_i(t)-\mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta}\}\big]-\rho_\tau\big\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\big\}\Big)^2dN_i(t)\Big\}\\
&\quad\times\Big\{\int_0^\infty 1^2 dN_i(t)\Big\}\Bigg]\\
\le \;& M_0E\left[\int_0^\infty \{\mathbf{V}_i(t)^\top\boldsymbol{\delta}\}^2dN_i(t)\right] = M_0E\Big[\int_0^\infty \{\mathbf{V}_{ia}(t)^\top\boldsymbol{\delta}_a\}^2dN_i(t)\Big] \le M_0\lambda_{\max}u^2,
\end{aligned}$$

where the second inequality follows from Cauchy-Schwarz inequality and the last inequality follows from condition (C5). We use the symmetrization technique in empirical process to find the upper bound for $\mathcal{A}(u)$. Suppose $\{\varepsilon_i\}_{i=1}^n$ is a Rademacher sequence that is independent of $(Y_i(t),\mathbf{V}_i(t))$, where a Rademacher sequence refers to i.i.d. random variables taking values of ±1 with probability 1/2. Define $\mathcal{A}^0(u) := \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|\le u}|\mathbb{G}_n\{\varepsilon_iD_i(\boldsymbol{\delta},\tau)\}|$ as the symmetrized version of $\mathcal{A}(u)$. According to Lemma 2.3.7 in van der Vaart and Wellner (1996), we have $\forall M > 2M_0\sqrt{\lambda_{\max}}u$,

$$P\{\mathcal{A}(u)\ge M\} \le \frac{2P\{\mathcal{A}^0(u)\ge M/4\}}{1-4M_0\lambda_{\max}u^2/M^2}. \tag{S2.2}$$

Let $v_- := I(v<0)|v|$. As $\rho_\tau(v) = v\{\tau - I(v<0)\} = v\tau + v_-$, we have $D_i(\boldsymbol{\delta},\tau) = -\tau\int_0^\infty \mathbf{V}_i(t)^\top\boldsymbol{\delta}dN_i(t) + W_i(\tau,\boldsymbol{\delta})$, where $W_i(\tau,\boldsymbol{\delta}) := \int_0^\infty \big([Y_i(t)-\mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta}\}]_- -$

$\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\}_-\big) dN_i(t)$. Define

$$\mathcal{A}_1^0(u) := \sup_{\boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \leq u} \Big| \mathbb{G}_n \Big[\varepsilon_i \int_0^\infty \{\mathbf{V}_i(t)^\top \boldsymbol{\delta}\} dN_i(t) \Big] \Big|$$

and

$$\mathcal{A}_2^0(u) := \sup_{\tau \in \Delta, \boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \leq u} |\mathbb{G}_n \{\varepsilon_i W_i(\tau, \boldsymbol{\delta})\}|.$$

Then we have $\mathcal{A}^0(u) \leq \mathcal{A}_1^0(u) + \mathcal{A}_2^0(u)$.

We first evaluate $\mathcal{A}_1^0(t)$. For any $\varsigma > 0$,

$$\begin{aligned} & E[\exp\{\varsigma \mathcal{A}_1^0(t)\}] \leq E\Big(\exp \Big[\varsigma \sqrt{r+s} t \max_{j \leq r+s} \Big| \mathbb{G}_n \Big\{ \varepsilon_i \int_0^\infty V_{ij}(t) dN_i(t) \Big\} \Big| \Big] \Big) \\ \leq & \sum_{j \leq r+s} E\Big(\exp \Big[\varsigma \sqrt{r+s} u \Big| \mathbb{G}_n \Big\{ \varepsilon_i \int_0^\infty V_{ij}(t) dN_i(t) \Big\} \Big| \Big] \Big) \\ \leq & \; 2(r+s) E\Big(\exp \Big[\varsigma \sqrt{r+s} u \mathbb{G}_n \Big\{ \varepsilon_i \int_0^\infty V_{ij}(t) dN_i(t) \Big\} \Big] \Big) \\ = & \; 2(r+s) \prod_{i=1}^n E\Big(\exp \Big[\varsigma \sqrt{r+s} u \Big\{ \varepsilon_i \int_0^\infty V_{ij}(t) dN_i(t) \Big\} / \sqrt{n} \Big] \Big) \\ \leq & \; 2(r+s) \exp\{2\varsigma^2 M_0^2 C_V^2 (r+s) u^2\}, \end{aligned}$$

where the first two inequalities are elementary, the third inequality follows from the fact that $E(e^{|Z|}) \leq E(e^Z + e^{-Z}) = 2E(e^Z)$ for any symmetric random variable Z, the first equality follows from the definition of $\mathbb{G}_n$ and the fact that $\{\varepsilon_i \int_0^\infty V_{ij}(t) dN_i(t), i = 1, \cdots, n\}$ are independent, and the last inequality follows from the facts $E(e^{u\varepsilon_i}) \leq e^{u^2/2}$ and $\big| \int_0^\infty V_{ij}(t) dN_i(t) \big| \leq 2M_0 C_V$ by conditions (C1) and (C2). Then, for any $K_1 > 0$,

$$\begin{aligned} P\{\mathcal{A}_1^0(u) > K_1\} \;\; & \leq \;\; \min_{\varsigma \geq 0} e^{-\varsigma K_1} E\{e^{\varsigma \mathcal{A}_1^0(u)}\} \leq \min_{\varsigma \geq 0} e^{-\varsigma K_1} 2(r+s) \exp\{2\varsigma^2 M_0^2 C_V^2 (r+s) u^2\} \\ & \leq \;\; \min_{\varsigma \geq 0} 2m(r+s) \exp\big\{2\varsigma^2 M_0^2 C_V^2 (r+s) u^2 - \varsigma K_1\big\} \end{aligned}$$

$$\leq \quad 2(r+s)\exp[-K_1^2/\{8M_0^2C_V^2(r+s)u^2\}],$$

where the last inequality follows from choosing $\varsigma = K_1/\{4M_0C_V(r+s)u^2\}$. Let $K_1 = 6M_0C_V\sqrt{(r+s)\log n}t$. Then

$$P\left\{\mathcal{A}_1^0(u) > 6M_0C_V\sqrt{(r+s)\log n}u\right\} \leq 2\exp\{-4\log n + \log(r+s)\} \leq 2n^{-3}. \quad \text{(S2.3)}$$

Next, we evaluate $\mathcal{A}_2^0(u)$. Let $\Delta_m = \{\tau_1, \tau_2, \ldots, \tau_m\}$ be the ϵ-net of Δ, where $\epsilon = u/(\sqrt{r+s}L)$ and $m = 1/\epsilon$. Following similar arguments as (6) in Supplementary Materials of Zheng et al. (2015), we have

$$\mathbb{G}_n\{\varepsilon_i W_i(\tau, \boldsymbol{\delta})\} = \mathbb{G}_n[\varepsilon_i W_i\{\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}\}] - \mathbb{G}_n[\varepsilon_i W_i\{\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k)\}].$$

According to condition (C3), if $|\tau - \tau_k| \leq \epsilon$, $\|\boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k)\| \leq L\sqrt{r+s}\epsilon = u$. The foregoing two equations together imply

$$\mathcal{A}_2^0(u) \leq 2 \sup_{\tau_k \in \Delta_m, \boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \leq u} |\mathbb{G}_n\{\varepsilon_i W_i(\tau_k, 2\boldsymbol{\delta})\}| =: \mathcal{A}_3^0(u).$$

We note that $|W_i(\tau, \mathbf{v}_1) - W_i(\tau, \mathbf{v}_2)| = \big|\int_0^\infty \big([Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau) + \mathbf{v}_1\}]_- - [Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau) + \mathbf{v}_2\}]_-\big) dN_i(t)\big| \leq |\int_0^\infty \mathbf{V}_i(t)^\top(\mathbf{v}_1 - \mathbf{v}_2) dN_i(t)|$, for any $\mathbf{v}_1, \mathbf{v}_2 \in R_{r+s-1}$. Thus, $W_i(\tau, \mathbf{v})$ is a contraction (see page 117 of Ledoux and Talagrand (1991)). Then

$$\begin{aligned}
E[e^{\varsigma\mathcal{A}_3^0(u)}] &= E\Big(\exp\Big[2\varsigma \sup_{\tau_k \in \Delta_m, \boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \leq u} |\mathbb{G}_n\{\varepsilon_i W_i(\tau_k, 2\boldsymbol{\delta})\}|\Big]\Big) \\
&\leq E\Big(\sum_{k=1}^m \exp\Big[2\varsigma \sup_{\boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \leq u} |\mathbb{G}_n\{\varepsilon_i W_i(\tau_k, 2\boldsymbol{\delta})\}|\Big]\Big) \\
&\leq mE\Big(\exp\Big[4\varsigma \sup_{\boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \leq u} \Big|\mathbb{G}_n\Big\{2\varepsilon_i \int_0^\infty \mathbf{V}_{ia}(t)^\top \boldsymbol{\delta}_a dN_i(t)\Big\}\Big|\Big]\Big)
\end{aligned}$$

$$\begin{aligned}
&\leq \quad mE\Big(\exp\Big[8\varsigma\sqrt{r+s}u \max_{j\leq r+s}\Big|\mathbb{G}_n\Big\{\varepsilon_i\int_0^\infty V_{ij}(t)dN_i(t)\Big\}\Big|\Big]\Big)\\
&\leq \quad 2m(r+s)\exp\left\{128\varsigma^2 M_0^2 C_V^2(r+s)u^2\right\},
\end{aligned}$$

where the first inequality is elementary, the second inequality follows from Theorem 4.12 in Ledoux and Talagrand (1991) with the convex function $\exp(\cdot)$, and the last two inequalities follow from the same arguments used for $\mathcal{A}_1^0(u)$. Therefore, for any $K_2 > 0$,

$$\begin{aligned}
&P\left\{\mathcal{A}_2^0(u) > K_2\right\} \leq P\left\{\mathcal{A}_3^0(u) > K_2\right\} \leq \min_{\varsigma\geq 0} e^{-\varsigma K_2} E\{e^{\varsigma \mathcal{A}_3^0(u)}\}\\
\leq \quad & 2m(r+s)\exp[-K_2^2/\{512(r+s)u^2\}],
\end{aligned}$$

where the last equality holds at $\varsigma = K_2/\{256 M_0 C_V(r+s)u^2\}$. Since $m = 1/\epsilon = L\sqrt{r+s}/u$, then

$$\begin{aligned}
P\{\mathcal{A}_2^0(u) > K_2\} \quad &\leq \quad 2\exp\left\{-\frac{K_2^2}{512 M_0 C_V (r+s)u^2} + \log(r+s) + \log\left(\frac{L\sqrt{r+s}}{u}\right)\right\}\\
&\leq \quad 2\exp\left\{-\frac{K_2^2}{512 M_0 C_V (r+s)u^2} + \frac{3}{2}\log(r+s) + \log L - \log u\right\}.
\end{aligned}$$

If $K_2 = 50 M_0 C_V\sqrt{r+s}u\sqrt{\log n + \log L/2 - \log u/2}$, then

$$P\left\{\mathcal{A}_2^0(u) > 50 M_0 C_V\sqrt{r+s}u\sqrt{\log n + \log L/2 - \log u/2}\right\} \leq 2n^{-3}. \tag{S2.4}$$

Combining (S2.2)–(S2.4) together, we have

$$P\Big\{\mathcal{A}(u) > M_0 C_V\sqrt{r+s}u\Big(24\sqrt{\log n} + 200\sqrt{\log n + \log L/2 - \log u/2}\Big)\Big\} \leq 16n^{-3}.$$

This completes the proof of Lemma 2. □

Proof of Lemma 3. As compared with Lemma 2, we consider a fixed quantile τ_0 here. Define $\mathcal{A}(u)$, $\mathcal{A}^0(u)$, $\mathcal{A}_1^0(u)$, and $\mathcal{A}_2^0(u)$ the same way as in Lemma 2, which take the supreme only over $\boldsymbol{\delta} \in R_{r+s-1}, \|\boldsymbol{\delta}\| \le u$. Then

$$\begin{aligned}
E[\{\mathcal{A}_1^0(u)\}^2] &= n^{-1}E\Big\{\sup_{\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|\le u}\Big|\boldsymbol{\delta}_a^\top\sum_{i=1}^n\varepsilon_i\int_0^\infty \mathbf{V}_{ia}(t)dN_i(t)\Big|^2\Big\} \\
&\le n^{-1}E\Big\{\sup_{\boldsymbol{\delta}\in R_{r+s-1},\|\boldsymbol{\delta}\|\le u}\|\boldsymbol{\delta}_a\|^2\cdot\Big\|\sum_{i=1}^n\varepsilon_i\int_0^\infty \mathbf{V}_{ia}(t)dN_i(t)\Big\|^2\Big\} \\
&= n^{-1}u^2E\Big[\Big\{\sum_{i=1}^n\varepsilon_i\int_0^\infty \mathbf{V}_{ia}(t)^\top dN_i(t)\Big\}\Big\{\sum_{i=1}^n\varepsilon_i\int_0^\infty \mathbf{V}_{ia}(t)dN_i(t)\Big\}\Big] \\
&= u^2E\left[\Big\{\int_0^\infty \mathbf{V}_{ia}(t)^\top dN_i(t)\Big\}\Big\{\int_0^\infty \mathbf{V}_{ia}(t)dN_i(t)\Big\}\right]
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the third equality follows from that $\{\varepsilon_i\}_{i=1}^n$ are mutually independent and independent of the variables $\{(\mathbf{V}_i(t), N_i(t))\}_{i=1}^n$ and $E[\varepsilon_i] = 0$. As $E\left[\{\int_0^\infty \mathbf{V}_{ia}(t)^\top dN_i(t)\}\{\int_0^\infty \mathbf{V}_{ia}(t)dN_i(t)\}\right]$ is the trace of $E\left[\{\int_0^\infty \mathbf{V}_{ia}(t)dN_i(t)\}\{\int_0^\infty \mathbf{V}_{ia}(t)^\top dN_i(t)\}\right]$ and the trace of a square matrix equals the sum of the eigenvalues of the matrix, we have $E[\{\mathcal{A}_1^0(u)\}^2] \le M_0^2\lambda_{\max}(r+s)u^2$ by conditions (C1) and (C4). For any $a > 8$, $P(\mathcal{A}_1^0(u) > aM_0\sqrt{\lambda_{\max}(r+s)}u) \le a^{-1}$, by Markov inequality. Following similar arguments as in Lemma 2, $E[\{\mathcal{A}_2^0(u)\}^2] \le 4E[\{\mathcal{A}_1^0(u)\}^2] \le 4M_0^2\lambda_{\max}(r+s)u^2$, and $P\{\mathcal{A}_2^0(u) > 2aM_0\sqrt{\lambda_{\max}(r+s)}u\} \le a^{-1}$. Hence

$$P\{\mathcal{A}(u) > 12aM_0\sqrt{\lambda_{\max}(r+s)}u\} \le 8a^{-1}.$$

This completes the proof of Lemma 3. $\square$

Proof of Lemma 4: Let

$$\mathcal{A} := \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} |\boldsymbol{\xi}^\top[M_n(\tau,\boldsymbol{\delta}) - E\{M_n(\tau,\boldsymbol{\delta})\} - M_n(\tau,\mathbf{0}) + E\{M_n(\tau,\mathbf{0})\}]|$$

$$= \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} |\mathbb{G}_n\{U_i(\boldsymbol{\xi},\boldsymbol{\delta},\tau)\}|,$$

where $U_i(\boldsymbol{\xi},\boldsymbol{\delta},\tau) = \boldsymbol{\xi}^\top \int_0^\infty \mathbf{V}_i(t) D_i^I(\tau,\boldsymbol{\delta}) dN_i(t)$. We then obtain that

$$\mathrm{Var}[U_i(\boldsymbol{\xi},\boldsymbol{\delta},\tau)] \le E[\{U_i(\boldsymbol{\xi},\boldsymbol{\delta},\tau)\}^2] \le M_0 E\Big[\int_0^\infty \{\boldsymbol{\xi}^\top\mathbf{V}_i(t)\}^2 (D_i^I(\tau,\boldsymbol{\delta}))^2 dN_i(t)\Big]$$

$$\le \bar{f} M_0 E\Big[\int_0^\infty \{\boldsymbol{\xi}^\top\mathbf{V}_i(t)\}^2 |\mathbf{V}_i(t)^\top\boldsymbol{\delta}| dN_i(t)\Big]$$

$$\le \bar{f} M_0 \left\{E\Big\{\int_0^\infty |\boldsymbol{\xi}^\top\mathbf{V}_i(t)|^3 dN_i(t)\Big\}\right\}^{2/3} \left\{E\Big\{\int_0^\infty |\mathbf{V}_i(t)^\top\boldsymbol{\delta}|^3 dN_i(t)\Big\}\right\}^{1/3}$$

$$\le \bar{f} M_0 \left\{\frac{\left\{E\left[\int_0^\infty |\boldsymbol{\xi}^\top\mathbf{V}_i(t)|^2 dN_i(t)\right]\right\}^{3/2}}{q}\right\}^{2/3} \left\{\frac{\left\{E\left[\int_0^\infty |\mathbf{V}_i(t)^\top\boldsymbol{\delta}|^2 dN_i(t)\right]\right\}^{3/2}}{q}\right\}^{1/3}$$

$$\le C_B\sqrt{(r+s)n^{-1}\log n},$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from Law of iterated expectation and condition (C2), the third inequality follows from Holder's inequality, the last two inequalities follow from condition (C5), and $C_B = \bar{f} M_0 B \lambda_{\max}^{3/2} q^{-1}$. Then by Lemma 2.3.7 in van der Vaart and Wellner (1996), we have that for $\forall M$ such that $M^2 > 4C_B\sqrt{(r+s)n^{-1}\log n}$

$$P(\mathcal{A} \ge M) \le \frac{2P(\mathcal{A}^0 \ge M/4)}{1 - 4C_B\sqrt{(r+s)n^{-1}\log n}/M^2}, \tag{S2.5}$$

where $\mathcal{A}^0 := \sup_{\boldsymbol{\delta}\in R_{r+s-1}(B)} \sup_{\tau\in\Delta} |\mathbb{G}_n\{\varepsilon_i U_i(\boldsymbol{\xi},\boldsymbol{\delta},\tau)\}|$. Let $\Delta_m = \{\tau_1,\tau_2,\ldots,\tau_m\}$ be a

ϵ-net of Δ, where $\epsilon = B\sqrt{n^{-1}\log n}/L$ and $m = 1/\epsilon$. Then

$$\begin{aligned}
U_i(\boldsymbol{\xi}, \boldsymbol{\delta}, \tau) &= \boldsymbol{\xi}^\top \int_0^\infty \mathbf{V}_i(t) D_i^I(\tau, \boldsymbol{\delta}) dN_i(t) \\
&= -\boldsymbol{\xi}^\top \int_0^\infty \mathbf{V}_i(t) D_i^I(\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k)) dN_i(t) \\
&\quad + \boldsymbol{\xi}^\top \int_0^\infty \mathbf{V}_i(t) D_i^I(\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}) dN_i(t) \\
&= -\boldsymbol{\xi}^\top \mathbf{R}_i(\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k)) + \boldsymbol{\xi}^\top \mathbf{R}_i(\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}),
\end{aligned}$$

where $\mathbf{R}_i(\tau, \boldsymbol{\delta}) = \int_0^\infty \mathbf{V}_i(t) D_i^I(\tau, \boldsymbol{\delta}) dN_i(t)$. Hence

$$\begin{aligned}
\mathcal{A}^0 \le & \sup_{\tau\in\Delta, |\tau-\tau_k|\le\epsilon, \tau_k\in\Delta_m} |\mathbb{G}_n\{\varepsilon_i \boldsymbol{\xi}^\top \mathbf{R}_i(\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k))\}| \\
& + \sup_{\tau\in\Delta, |\tau-\tau_k|\le\epsilon, \tau_k\in\Delta_m, \boldsymbol{\delta}\in R_{r+s-1}(B)} |\mathbb{G}_n\{\varepsilon_i \boldsymbol{\xi}^\top \mathbf{R}_i(\tau_k, \boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta})\}|.
\end{aligned}$$

By condition (C3), if $|\tau - \tau_k| \le \epsilon$, $\|\boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_k)\| \le L\sqrt{r+s}\epsilon = B\sqrt{(r+s)n^{-1}\log n}$. Therefore,

$$\mathcal{A}^0 \le 2 \sup_{\tau_k\in\Delta_m, \boldsymbol{\delta}\in R_{r+s-1}(2B)} |\mathbb{G}_n\{\varepsilon_i \boldsymbol{\xi}^\top \mathbf{R}_i(\tau_k, \boldsymbol{\delta})\}| := 2\mathcal{B}^0.$$

Consider

$$T_{\boldsymbol{\xi},n,k}(B) := \sup_{\boldsymbol{\delta}\in R_{r+s-1}(2B)} |\mathbb{G}_n\{\varepsilon_i \boldsymbol{\xi}^\top \mathbf{R}_i(\tau_k, \boldsymbol{\delta})\}|.$$

Let $\{C(\boldsymbol{\delta}_l), l = 1, \ldots, N\}$ be cubes that cover the ball $R_{r+s-1}(2B)$, where $\boldsymbol{\delta}_l$ is the center of the cube $C(\boldsymbol{\delta}_l)$ with sides of length $B\sqrt{(r+s)n^{-5}\log n}$ so that the number of cubes $N = (4n^2)^{r+s}$, $\|\boldsymbol{\delta}_l\| \le 2B\sqrt{(r+s)n^{-1}\log n}$. For $\boldsymbol{\delta} \in C(\boldsymbol{\delta}_l)$, $\|\boldsymbol{\delta} - \boldsymbol{\delta}_l\| \le B(r+s)n^{-5/2}\log^{1/2} n := \zeta_n$. Let $T_{n,k}(\boldsymbol{\delta}) = n^{1/2}\mathbb{E}_n \varepsilon_i \int_0^\infty \mathbf{V}_i(t)\psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau_k) +$

$\boldsymbol{\delta})\}dN_i(t)$. Then

$$
\begin{aligned}
T_{\boldsymbol{\xi},n,k}(2B) \leq& \max_{1\leq l\leq N} |\boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\xi}^\top T_{n,k}(\mathbf{0})| + \max_{1\leq l\leq N} \sup_{\boldsymbol{\delta}\in C(\boldsymbol{\delta}_l)} |\boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}) - \boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}_l)| \\
\leq& \max_{1\leq l\leq N} |\boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\xi}^\top T_{n,k}(\mathbf{0})| + \max_{1\leq l\leq N} \Big| \\
& n^{1/2}\mathbb{E}_n \left[\int_0^\infty |\varepsilon_i \boldsymbol{\xi}^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} + \|\mathbf{V}_{ia}(t)\|\zeta_n\} dN_i(t)\right] \\
& - n^{1/2}E\Big[\int_0^\infty |\varepsilon_i \boldsymbol{\xi}^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} + \|\mathbf{V}_{ia}(t)\|\zeta_n\} dN_i(t)\Big] \\
& - n^{1/2}\mathbb{E}_n \left[\int_0^\infty |\varepsilon_i \boldsymbol{\xi}^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\}\} dN_i(t)\right] \\
& + n^{1/2}E\Big[\int_0^\infty |\varepsilon_i \boldsymbol{\xi}^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\}\}\Big]\Big| \\
& + \max_{1\leq l\leq N} n^{1/2}E\Big[\int_0^\infty |\varepsilon_i \boldsymbol{\xi}^\top \mathbf{V}_i(t)| \big\{\psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} + \|\mathbf{V}_{ia}(t)\|\zeta_n\} \\
& \quad - \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} - \|\mathbf{V}_{ia}(t)\|\zeta_n\}\big\} dN_i(t)\Big] =: \mathrm{IV}_1 + \mathrm{IV}_2 + \mathrm{IV}_3,
\end{aligned}
$$

where the second equality follows from the monotone property of ψ_{τ_k} and the triangle inequality. For IV_3,

$$
\begin{aligned}
\mathrm{IV}_3 \leq& 2\bar{f}\zeta_n\sqrt{n}E\Big[\int_0^\infty |\boldsymbol{\xi}^\top \mathbf{V}_i(t)| \cdot \|\mathbf{V}_{ia}(t)\| dN_i(t)\Big] \\
\leq& 2\bar{f}\zeta_n\sqrt{n}E\Big[\Big(\int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}_i(t)\mathbf{V}_i(t)^\top\boldsymbol{\xi} dN_i(t)\Big)^{1/2}\Big(\int_0^\infty \|\mathbf{V}_{ia}(t)\|^2 dN_i(t)\Big)^{1/2}\Big] \\
\leq& 2\bar{f}\zeta_n\sqrt{M_0 n(r+s)}C_V E\Big[\Big(\int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}_i(t)\mathbf{V}_i(t)^\top\boldsymbol{\xi} dN_i(t)\Big)^{1/2}\Big] \\
\leq& 2\bar{f}\sqrt{M_0}C_V\zeta_n\sqrt{n(r+s)}\left(E\Big[\int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}_i(t)\mathbf{V}_i(t)^\top\boldsymbol{\xi} dN_i(t)\Big]\right)^{1/2} \\
\leq& 2\bar{f}\sqrt{M_0\lambda_{\max}}\zeta_n\sqrt{n(r+s)} = 2\bar{f}\sqrt{M_0\lambda_{\max}}B(r+s)^{3/2}n^{-2}\log n,
\end{aligned}
$$

where the first inequality follows from the property of $\psi_\tau(u)$ and conditional expectation, the second inequality follows from the Cauchy-Schwartz inequality, and the third inequality follows from conditions (C2) and (C3), the fourth inequality is trivial, and the last inequality follows from condition (C5). Now, we consider IV_1. Since

$$\boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\xi}^\top T_{n,k}(\mathbf{0}) = n^{1/2}\mathbb{E}_n\varepsilon_i \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}_i(t) D_i^I(\tau_k, \boldsymbol{\delta}_l) dN_i(t),$$

then following the same arguments as in finding the upper bound of $\mathrm{Var}[U_i(\boldsymbol{\xi}, \boldsymbol{\delta}, \tau)]$, we have $\mathrm{Var}[\boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\xi}^\top T_{n,k}(\mathbf{0})] \le C_B\sqrt{(r+s)n^{-1}\log n}$.

Noting that $|\varepsilon_i \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}_i(t) D_i^I(\tau_k, \boldsymbol{\delta}_l) dN_i(t)|$ is bounded by $M_0 C_V\sqrt{r+s}$, we have

$$\begin{aligned}
&P\left(\left|\boldsymbol{\xi}^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\xi}^\top T_{n,k}(\mathbf{0})\right| > 4C_B^{1/2} n^{-1/4}(r+s)^{3/4}\log n\right)\\
&\le \exp\left(\frac{-8C_B n^{-1/2}(r+s)^{3/2} n \log^2 n}{C_B n\sqrt{(r+s)n^{-1}\log n} + 4C_B^{1/2} n^{-1/4}(r+s)^{3/4}\log n \times M_0 C_V\sqrt{(r+s)n}/3}\right)\\
&= 2\exp\left(\frac{-8(r+s)^{3/2}\log^2 n}{(r+s)^{1/2}\log^{1/2} n + 4M_0 C_V C_B^{-1/2} n^{-1/4}(r+s)^{5/4}\log n/3}\right)\\
&\le 2\exp(-4(r+s)\log n),
\end{aligned}$$

where the first inequality follows from Bernstein's inequality and the second inequality follows from the condition $(r+s)^3\log^4 n = o(n)$ and $4M_0C_VC_2C_B^{-1/2}n^{-1/4}(r+s)^{5/4}\log n/3 \le C_B(r+s)^{1/2}\log n$, when n is sufficiently large. Therefore, we have

$$\begin{aligned}
&P(\mathrm{IV}_1 \ge 4C_B^{1/2} n^{-1/4}(r+s)^{3/4}\log n)\\
&= P(\max_{1\le l\le N} |\boldsymbol{\xi}^\top\{T_{n,k}(\boldsymbol{\delta}_l) - T_{n,k}(\mathbf{0})\}| \ge 4C_B^{1/2} n^{-1/4}(r+s)^{3/4}\log n)\\
&= N\cdot P(|\boldsymbol{\xi}^\top\{T_{n,k}(\boldsymbol{\delta}_l) - T_{n,k}(\mathbf{0})\}| \ge 4C_B^{1/2} n^{-1/4}(r+s)^{3/4}\log n)\\
&\le 2N\exp\left[-4(r+s)\log n\right] \le 2\exp\left[-4(r+s)\log n + 3(r+s)\log n\right].
\end{aligned}$$

As the upper bound of IV_3 satisfies $2\bar{f}\sqrt{M_0\lambda_{\max}}B(r+s)^{3/2}n^{-2}\log n = o(4C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n)$. The results of IV_1 and IV_3 together yield

$$P\left(\mathrm{IV}_1 + \mathrm{IV}_3 > 8C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n\right) \leq 2\exp\left(-(r+s)\log n\right).$$

We can use similar arguments for IV_1 and IV_3 to find the bound for IV_2, and then get

$$P(T_{\boldsymbol{\xi},n,k}(2B) > 16C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n) \leq 4\exp\left(-(r+s)\log n\right).$$

Following the definition of $\mathcal{B}^0$,

$$\begin{aligned}
&P(\mathcal{B}^0 > 16C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n) \\
\leq &P(\max_{1\leq k\leq m} T_{\boldsymbol{\xi},n,k}(B) > 16C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n) \\
\leq &4m\exp\left(-(r+s)\log n\right) \leq 4\exp\left(-(r+s)\log n + \log\left(\frac{L\sqrt{n}}{B\log n}\right)\right) \to 0.
\end{aligned}$$

Consequently,

$$P\left(\mathcal{A} \geq 128C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n\right) \leq \frac{2P\left(\mathcal{A}_0 \geq 32C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n\right)}{1 - \frac{4C_B\sqrt{(r+s)n^{-1}\log n}}{\left(128C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n\right)^2}}$$

$$\leq 4P\left(2\mathcal{B}_0 \geq 32C_B^{1/2}n^{-1/4}(r+s)^{3/4}\log n\right) \to 0.$$

Since $n^{-1/4}(r+s)^{3/4}\log n = o(1)$, we have

$$\sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} |\boldsymbol{\xi}^\top[M_n(\tau,\boldsymbol{\delta}) - E\{M_n(\tau,\boldsymbol{\delta})\} - M_n(\tau,\mathbf{0}) + E\{M_n(\tau,\mathbf{0})\}]| = o_p(1).$$

This completes the proof of Lemma 4. $\square$

Proof of Lemma 5. Write

$$\boldsymbol{\xi}^\top M_n(\tau,\mathbf{0}) = n^{1/2}\mathbb{E}_n\left[\int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}_i(t)\psi_\tau\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)\right]$$
$$= \mathbb{G}_n\{h_{\tau,\boldsymbol{\xi},n}(\mathbf{V}_i(t),Y_i(t))\}.$$

We use Theorem 2.11.23 in van der Vaart and Wellner (1996) to prove the desired results. Consider the function class $\mathcal{H}_n := \left\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t),Y),\tau\in\Delta\right\}$, and its envelop function $H_n(\mathbf{V}(t)) = \int_0^\infty |\boldsymbol{\xi}^\top\mathbf{V}(t)|dN(t)$.

We first check the three conditions of H_n and $\mathcal{H}_n$ from (2.11.21) in van der Vaart and Wellner (1996). It is easy to see that $E[H_n^2]\le M_0\lambda_{\max}=O(1)$ and $E[H_n^2 I\{H_n\ge \eta\sqrt{n}\}]\to 0$, for any $\eta>0$. For $h_{n,\boldsymbol{\xi},\tau_1}-h_{n,\boldsymbol{\xi},\tau_2}$, since $\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)$ is the conditional τth quantile of $Y_i(t)|\mathbf{V}_i(t)$, then $\forall\tau_1>\tau_2,\tau_1,\tau_2\in\Delta$,

$$\tau_1-\tau_2 = P\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_1)<0\} - P\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_2)<0\}$$
$$= \int_{\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_2)}^{\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_1)} f_{Y_i(t)}\{u|\mathbf{V}_i(t)\}du \ge \underline{f}\mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_1)-\boldsymbol{\gamma}_0(\tau_2)\},$$

where $f_{Y_i(t)}\{\cdot|\mathbf{V}_i(t)\}$ is the conditional density of $Y_i(t)$ given $\mathbf{V}_i(t)$. Thus

$$|h_{n,\boldsymbol{\xi},\tau_1}-h_{n,\boldsymbol{\xi},\tau_2}| \le \int_0^\infty |\boldsymbol{\xi}^\top\mathbf{V}(t)|\Big\{|\tau_1-\tau_2| + |I\{Y(t)-\mathbf{V}(t)^\top\boldsymbol{\gamma}_0(\tau_1)<0\}$$
$$- I\{Y(t)-\mathbf{V}(t)^\top\boldsymbol{\gamma}_0(\tau_2)<0\}|\Big\}dN(t),$$

and for any $\eta_n\to 0$,

$$\sup_{|\tau_1-\tau_2|\le\eta_n} E[(h_{n,\boldsymbol{\xi},\tau_1}-h_{n,\boldsymbol{\xi},\tau_2})^2] \le M_0\lambda_{\max}\left\{2|\tau_1-\tau_2|^2 + 2\frac{|\tau_1-\tau_2|}{\underline{f}}\bar{f}\right\}$$

$$\leq M_0\lambda_{\max}\left\{2\eta_n^2 + 2\frac{\eta_n}{\underline{f}}\bar{f}\right\} \to 0.$$

Next we calculate the bracketing number of $\mathcal{H}_n$, where the definition of bracketing number can be found in Page 83 of van der Vaart and Wellner (1996). Define

$$\begin{aligned}
u_{n,\boldsymbol{\xi},\tau_1} =& \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t) I\{\boldsymbol{\xi}^\top \mathbf{V}(t) > 0\}[\tau_1 + \underline{f}\epsilon - I\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau_1) + \epsilon < 0\}]dN(t) \\
&+ \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t) I\{\boldsymbol{\xi}^\top \mathbf{V}(t) < 0\}[\tau_1 - \underline{f}\epsilon - I\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau_1) - \epsilon < 0\}]dN(t) \\
l_{n,\boldsymbol{\xi},\tau_1} =& \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t) I\{\boldsymbol{\xi}^\top \mathbf{V}(t) > 0\}[\tau_1 - \underline{f}\epsilon - I\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau_1) - \epsilon < 0\}]dN(t) \\
&+ \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t) I\{\boldsymbol{\xi}^\top \mathbf{V}(t) < 0\}[\tau_1 + \underline{f}\epsilon - I\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau_1) + \epsilon < 0\}]dN(t).
\end{aligned}$$

If $|\tau_1 - \tau_2| \leq \underline{f}\epsilon$, we have $l_{n,\boldsymbol{\xi},\tau_1} \leq h_{n,\boldsymbol{\xi},\tau_2} \leq u_{n,\boldsymbol{\xi},\tau_1}$. Furthermore, for $\epsilon < 1$,

$$\begin{aligned}
&\|u_{n,\boldsymbol{\xi},\tau_1} - l_{n,\boldsymbol{\xi},\tau_1}\|_2^2 = E[(u_{n,\boldsymbol{\xi},\tau_1} - l_{n,\boldsymbol{\xi},\tau_1})^2] \\
\leq& E\Bigg[\bigg\{\int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t)\Big\{2\underline{f}\epsilon + \Big|I\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau_1) - \epsilon < 0\} \\
&\qquad - I\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau_1) + \epsilon < 0\}\Big|\Big\}dN(t)\bigg\}^2\Bigg] \\
\leq& 8M_0\lambda_{\max}\underline{f}^2\epsilon^2 + 4M_0\lambda_{\max}\bar{f}\epsilon \leq 8(\bar{f} + \underline{f}^2)M_0\lambda_{\max}\epsilon.
\end{aligned}$$

Then the bracketing number $N_{[]}(\sqrt{8(\bar{f} + \underline{f}^2)M_0\lambda_{\max}\epsilon}, \mathcal{H}_n, P) = \max\{1, 1/(\underline{f}\epsilon)\}$, and thus $\log N_{[]}(\epsilon, \mathcal{H}_n, P) \leq \max\{0, \log\{8(\bar{f} + \underline{f}^2)M_0\lambda_{\max}/(\underline{f}\epsilon^2)\}\}$.

We can check that $\int_0^{\eta_n} \log N_{[]}(\epsilon, \mathcal{H}_n, P)d\epsilon \to 0$ for any $\eta_n \downarrow 0$. Following Theorem 2.11.23 in van der Vaart and Wellner (1996), $\{\mathbb{G}_n h_{n,\boldsymbol{\xi},\tau}, \tau \in \Delta\}$ is asymptotically tight and converges weakly to a mean zero process with covariance function

$$E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t),Y)h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t),Y)\} - E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t),Y)\}E\{h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t),Y)\}.$$

This completes the proof of Lemma 5. □

Proof of Lemma 6. We use similar arguments as in the proof of Lemma 6. Let

$$\begin{aligned}\mathcal{A}_j &:= \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} \boldsymbol{\eta}_j^\top[M_n(\tau,\boldsymbol{\delta}) - E\{M_n(\tau,\boldsymbol{\delta})\} - M_n(\tau,\mathbf{0}) + E\{M_n(\tau,\mathbf{0})\}] \\ &= \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} \Big|\mathbb{G}_n\Big\{\int_0^\infty V_{ij}(t)D_i^I(\tau,\boldsymbol{\delta})dN_i(t)\Big\}\Big|.\end{aligned}$$

Let $C_B' = 2M_0^{3/2}C_V^2\bar{f}\lambda_{\max}^{1/2}B$. For any $\boldsymbol{\delta}\in R_{r+s-1}(B)$, we have

$$\begin{aligned}&\mathrm{Var}\Big[\int_0^\infty V_{ij}(t)D_i^I(\tau,\boldsymbol{\delta})dN_i(t)\Big] \le 2M_0C_V^2E\Big[\int_0^\infty\{D_i^I(\tau,\boldsymbol{\delta})\}^2dN_i(t)\Big] \\ \le& 2M_0C_V^2\bar{f}E\Big[\int_0^\infty|\mathbf{V}_i(t)^\top\boldsymbol{\delta}|dN_i(t)\Big] \le 2M_0^{3/2}C_V^2\bar{f}E\Big\{\Big(\int_0^\infty|\mathbf{V}_i(t)^\top\boldsymbol{\delta}|^2dN_i(t)\Big)^{1/2}\Big\} \\ \le& 2M_0^{3/2}C_V^2\bar{f}\Big[E\Big\{\int_0^\infty|\mathbf{V}_i(t)^\top\boldsymbol{\delta}|^2dN_i(t)\Big\}\Big]^{1/2} \le 2M_0^{3/2}C_V^2\bar{f}\lambda_{\max}^{1/2}B\sqrt{(r+s)n^{-1}\log n} \\ =& C_B'\sqrt{(r+s)n^{-1}\log n}.\end{aligned}$$

where the first and third inequalities follow from Cauchy-Schwarz inequality, the second inequality follows from conditional expectation given $\mathbf{V}_i(t)$ along with condition (C2), the last equality follows from condition (C5). Let

$$\mathcal{A}_j^0 := \sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} \Big|\mathbb{G}_n\Big[\varepsilon_i\int_0^\infty V_{ij}(t)D_i^I(\tau,\boldsymbol{\delta})dN_i(t)\Big]\Big|.$$

Applying Lemma 2.3.7 in van der Vaart and Wellner (1996) yields that for $\forall M$ such that $M^2 > 4C'_B\sqrt{(r+s)n^{-1}\log n}$

$$P(\mathcal{A} \geq M) \leq \frac{2P(\mathcal{A}^0 \geq M/4)}{1 - 4C'_B\sqrt{(r+s)n^{-1}\log n}/M^2}, \tag{S2.6}$$

Let $\Delta_m, \mathbf{R}_i(\tau, \boldsymbol{\delta})$ be the same as defined in Lemma 4, we have

$$\mathcal{A}_j^0 \leq 2 \sup_{\tau_k \in \Delta_m, \boldsymbol{\delta} \in R_{r+s-1}(2B)} |G_n[\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{R}_i(\tau_k, \boldsymbol{\delta})]| := 2\mathcal{B}_j^0.$$

It is sufficient to consider

$$T_{j,n,k}(B) := \sup_{\boldsymbol{\delta} \in R_{r+s-1}(2B)} |G_n[\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{R}_i(\tau_k, \boldsymbol{\delta})]|.$$

Let $\{C(\boldsymbol{\delta}_l), l = 1, \ldots, N\}$ be cubes that cover the ball $R_{r+s-1}(2B)$, where $\boldsymbol{\delta}_l$ is the center of the cube $C(\boldsymbol{\delta}_l)$ with sides of length $B\sqrt{(r+s)\log n/n^5}$ so that the number of cubes $N = (4n^2)^{r+s}$, $\|\boldsymbol{\delta}_l\| \leq 2B\sqrt{(r+s)\log n/n}$. For $\boldsymbol{\delta} \in C(\boldsymbol{\delta}_l)$, $\|\boldsymbol{\delta} - \boldsymbol{\delta}_l\| \leq B(r+s)n^{-5/2}\log^{1/2} n := \zeta_n$.

Let $T_{n,k}(\boldsymbol{\delta}) = n^{-1/2}\sum_{i=1}^n \varepsilon_i \int_0^\infty \mathbf{V}_i(t)\psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta})\}dN_i(t)$. The monotone property of ψ_{τ_k} implies that

$$T_{\boldsymbol{\eta}_j,n,k}(B) \leq \max_{1\leq l\leq N} |\boldsymbol{\eta}_j^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\eta}_j^\top T_{n,k}(\mathbf{0})| + \max_{1\leq l\leq N} \sup_{\boldsymbol{\delta} \in C(\boldsymbol{\delta}_l)} |\boldsymbol{\eta}_j^\top T_{n,k}(\boldsymbol{\delta}) - \boldsymbol{\eta}_j^\top T_{n,k}(\boldsymbol{\delta}_l)|$$

$$\begin{aligned}
&\leq \max_{1\leq l\leq N} \left|\boldsymbol{\eta}_j^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\eta}_j^\top T_{n,k}(\mathbf{0})\right| + \max_{1\leq l\leq N} \sup_{\tau\in\Delta} \Bigg| \\
&\quad n^{1/2}\mathbb{E}_n\left[\int_0^\infty |\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} + \|\mathbf{V}_{ia}(t)\|\zeta_n\}dN_i(t)\right] \\
&\quad - n^{1/2}E\Big[\int_0^\infty |\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} + \|\mathbf{V}_{ia}(t)\|\zeta_n\}dN_i(t)\Big] \\
&\quad - n^{1/2}\mathbb{E}_n\left[\int_0^\infty |\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{V}_i(t)|\ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\}\}dN_i(t)\right]
\end{aligned}$$

$$
\begin{aligned}
&+ n^{-1/2} E\Big[\sum_{i=1}^{n} \int_0^\infty |\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{V}_i(t)| \ \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top \{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} dN_i(t)\Big]\Big| \\
\leq &\max_{1\leq l\leq N} \sup_{\tau\in\Delta} n^{-1/2} E\Big[\sum_{i=1}^{n} \int_0^\infty |\varepsilon_i \boldsymbol{\eta}_j^\top \mathbf{V}_i(t)| \ \big\{\psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top \{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} + \|\mathbf{V}_{ia}(t)\|\zeta_n\} \\
&- \psi_{\tau_k}\{Y_i(t) - \mathbf{V}_i(t)^\top \{\boldsymbol{\gamma}_0(\tau_k) + \boldsymbol{\delta}_l\} - \|\mathbf{V}_{ia}(t)\|\zeta_n\}\big\} dN_i(t)\Big] \\
=:&\mathrm{VI}_1 + \mathrm{VI}_2 + \mathrm{VI}_3.
\end{aligned}
$$

For VI_3, we have $\mathrm{VI}_3 \leq 4\bar{f} M_0 C_V^2 B(r+s) n^{-5/2} \log n$ from the proof of Lemma 4.

For VI_1, note that

$$
\boldsymbol{\eta}_j^\top T_{n,k}(\boldsymbol{\delta}_l) - \boldsymbol{\eta}_j^\top T_{n,k}(\mathbf{0}) = n^{-1/2} \sum_{i=1}^{n} \varepsilon_i \int_0^\infty \boldsymbol{\eta}_j^\top \mathbf{V}_i(t) D_i^I(\tau_k, \boldsymbol{\delta}_l) dN_i(t).
$$

Then

$$
\begin{aligned}
&P\left(\sqrt{n} |\boldsymbol{\eta}_j^\top \{T_{n,k}(\boldsymbol{\delta}_l) - T_{n,k}(\mathbf{0})\}| > 4 C_B'^{1/2} n^{1/4} (r+s)^{3/4} \log \max\{n, r+p\} \right) \\
\leq &2\exp\left(-\frac{1}{2} \frac{16 C_B' n^{1/2} (r+s)^{3/2} \log^2 \max\{n, r+p\}}{C_B' (r+s)^{1/2} n^{1/2} \log^{1/2} n + 8 C_V C_B'^{1/2} n^{1/4} (r+s)^{3/4} \log \max\{n, r+p\}/3} \right) \\
\leq &2\exp\left(-\frac{8(r+s) \log \max\{n, r+p\}}{\log^{-1/2} \max\{n, r+p\} + 8 C_V C_B'^{-1/2} n^{-1/4} (r+s)^{1/4}/3} \right) \\
\leq &2\exp\left(-4(r+s) \log \max\{n, r+p\} \right),
\end{aligned}
$$

where the first inequality follows from Bernstein's inequality, the second inequality is trivial, and the third inequality follows from the facts that $\log^{-1/2} \max\{n, r+p\} = o(1)$ and $n^{-1/4}(r+s)^{1/4} = o(1)$.

The rest arguments follow exactly as in the proof of Lemma 4 and we can find some constant C_2 such that Lemma 6 holds. □

Proof of Lemma 7: By the definition of τ^*, we have $\tau^*|u| \leq |\rho_\tau(u)| = |u\{\tau - I(u < 0)\}| \leq |u|$ when $\tau \in \Delta$. It then follows from the definition of $\sigma(\tau)$ that

$$\tau^*\mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)|dN_i(t)\right) \leq \sigma(\tau)$$
$$\leq \mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)|dN_i(t)\right).$$

It follows from triangle inequality, conditional expectation of $Y_i(t)$ given $\mathbf{V}_i(t)$, and condition (C2) that

$$\mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)|dN_i(t)\right)$$
$$\leq \mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\right) + \mathbb{E}_n\left(\int_0^\infty |\mathbf{V}_i(t)^\top (\boldsymbol{\gamma}_0(\tau) - \boldsymbol{\gamma}_0(\tau_0))|dN_i(t)\right)$$
$$\leq \mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\right) + \frac{|\tau - \tau_0|}{\underline{f}} M_0,$$

and

$$\mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)|dN_i(t)\right)$$
$$\geq \mathbb{E}_n\left(\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\right) - \frac{|\tau - \tau_0|}{\underline{f}} M_0.$$

According to the law of large numbers, $n^{-1}\sum_{i=1}^n \int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t) \rightarrow E[\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t)]$ almost surely. Combining the above results yields

$$\tau^*\left[E\Big[\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\Big] - M_0\underline{f}^{-1}\right]$$
$$\leq \inf_{\tau\in\Delta} \sigma(\tau) \leq \sup_{\tau\in\Delta} \sigma(\tau) \leq E\Big[\int_0^\infty |Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\Big] + M_0\underline{f}^{-1}$$

almost surely. This completes the proof of Lemma 7. $\square$

Proof of Lemma 8. Recall that $R_\ell = \{\boldsymbol{\delta} = (\boldsymbol{\delta}_{\mathbf{x}}^\top, \boldsymbol{\delta}_{\mathbf{z}}^\top)^\top : \boldsymbol{\delta}_{\mathbf{x}} \in \mathbb{R}^r, \sum_{j=1}^p \delta_{\mathbf{z}j} = 0, \|\boldsymbol{\delta}_z\|_0 \leq \ell - r\}$ with $\|\cdot\|_0$. For any $\ell < k$, $\boldsymbol{\delta} \in R_\ell$, write

$$\sum_{i=1}^n D_i(\boldsymbol{\delta}, \tau) = nE\left[D_i(\boldsymbol{\delta}, \tau)\right] + \sqrt{n}G_n\Big[D_i(\boldsymbol{\delta}, \tau)\Big].$$

Using arguments as in the proof of Lemma 1, we have

$$\frac{\underline{f}\Lambda_{\min}}{8}u^2 \leq \inf_{\substack{\boldsymbol{\delta}\in R_\ell, \|\boldsymbol{\delta}\|=u,\\ \ell\leq k, \tau\in\Delta}} E\Big[D_i(\boldsymbol{\delta}, \tau)\Big] \leq \sup_{\substack{\boldsymbol{\delta}\in R_\ell, \|\boldsymbol{\delta}\|\leq u,\\ \ell\leq k, \tau\in\Delta}} nE\Big[D_i(\boldsymbol{\delta}, \tau)\Big] \leq n\frac{2\bar{f}\Lambda_{\max} + \underline{f}\Lambda_{\min}}{4}u^2$$

when $u \leq 3q'\underline{f}\Lambda_{\min}/(4C_f\Lambda_{\max}^{3/2})$. Let

$$\mathcal{A}(u) \ := \sup_{\substack{\boldsymbol{\delta}\in R_\ell, \|\boldsymbol{\delta}\|\leq u,\\ \ell\leq k, \tau\in\Delta}} \Big|\mathbb{G}_n\Big[D_i(\boldsymbol{\delta}, \tau)\Big]\Big|.$$

Using arguments as in the proof of Lemma 2, we can show that

$$\begin{aligned} P\bigg(\mathcal{A}(u) \geq A\sqrt{k}u\Big(24\sqrt{\log(r+p)} + 200\sqrt{\log(r+p) + \log L/2 - \log u/2}\Big)\bigg) \\ \leq 8\exp(-(A^2-1)\log\max\{n, r+p\}) \\ + 8\exp(-(A^2-3/2)\log\max\{n, r+p\} + \log L - \log u), \end{aligned}$$

for some constant $A > 2$. Hence, we have with probability at least $1 - 8\exp(-(A^2 - 1)\log\max\{n, r+p\}) - 8\exp(-(A^2-3/2)\log\max\{n, r+p\} + \log L - \log u)$,

$$\sup_{\substack{\boldsymbol{\delta}\in R_\ell, \|\boldsymbol{\delta}\|\leq u,\\ \ell\leq k, \tau\in\Delta}} \sum_{i=1}^n D_i(\boldsymbol{\delta}, \tau) \geq \frac{\underline{f}\Lambda_{\min}}{8}nu^2 - A\sqrt{kn}u\Big(24\sqrt{\log\max\{n, r+p\}}$$

$$+ 200\sqrt{\log\max\{n, r+p\} + \log L/2 - \log u/2}\Big).$$

In particular, choose $u = 3584A\sqrt{k\log(\max\{r+p, n\})/n}/(\underline{f}\Lambda_{\min})$.

$$\sup_{\substack{\boldsymbol{\delta}\in R_\ell, \|\boldsymbol{\delta}\|=u,\\ \ell\le k, \tau\in\Delta}} \sum_{i=1}^n D_i(\boldsymbol{\delta}, \tau) \ge 0,$$

with probability at least $1 - 16\exp(-(A^2-2)\log\max\{n, r+p\})$. The above equation and the convex property of $\rho_\tau(\cdot)$ indicates that

$$\sup_{S_\Delta\subseteq S, |S|=k}\sup_{\tau\in\Delta}\|\hat{\boldsymbol{\gamma}}_S(\tau) - \boldsymbol{\gamma}_0(\tau)\| \le 3584A\sqrt{k\log(\max\{r+p, n\})/n}/(\underline{f}\Lambda_{\min}).$$

This completes the proof of Lemma 8. □

Proof of Lemma 9. Consider a model $S \in OF$ with $|S| = k \ge 1$. It follows from simple algebra that

$$n\{\hat{\sigma}_S(\tau) - \hat{\sigma}_{S_\Delta}(\tau)\} = D_i(\hat{\boldsymbol{\gamma}}_S(\tau) - \boldsymbol{\gamma}_0(\tau), \tau) - D_i(\hat{\boldsymbol{\gamma}}_{S_\Delta}(\tau) - \boldsymbol{\gamma}_0(\tau), \tau) \tag{S2.7}$$

According to Lemma 8, with probability at least $1-16\exp(-(A^2-2)\log\max\{n, r+p\})$,

$$\sup_{S_\Delta\subset S, |S|=k}\sup_{\tau\in\Delta}\|\hat{\boldsymbol{\gamma}}_S(\tau) - \boldsymbol{\gamma}_0(\tau)\| \le 3584A\sqrt{k\log(\max\{n, r+p\})/n}/(\underline{f}\Lambda_{\min}).$$

Define

$$H_{\tau,S} = \begin{pmatrix} E\{\int_0^\infty f_{t,\tau}(0|\mathbf{V}_i(t))\mathbf{V}_{iS}(t)\mathbf{V}_{iS}(t)^\top dN_i(t)\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

and

$$H_{\tau,S}^{-1} = \begin{pmatrix} [E\{\int_0^\infty f_{t,\tau}(0|\mathbf{V}_i(t))\mathbf{V}_{iS}(t)\mathbf{V}_{iS}(t)^\top dN_i(t)\}]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Following the lines as in the proof of Theorem A.2, with probability at least $1 - 16\exp(-(A^2-2)\log(r+p))$, we have

$$\begin{aligned} & D_i(\hat{\boldsymbol{\gamma}}_S(\tau) - \boldsymbol{\gamma}_0(\tau), \tau) \\ = & \frac{n}{2}\big[n^{-1/2}\{H_{\tau,S}\}^{-1}M_n(\tau,\mathbf{0})\big]^\top H_{\tau,S}\big[n^{-1/2}\{H_{\tau,S}\}^{-1}M_n(\tau,\mathbf{0})\big] \\ & -n^{1/2}\big[n^{-1/2}\{H_{\tau,S}\}^{-1}M_n(\tau,\mathbf{0})\big]^\top M_n(\tau,\mathbf{0}) + o_p(\sqrt{k\log\max\{n,r+p\}/n}) \\ = & -\frac{1}{2}M_n(\tau,\mathbf{0})^\top H_{\tau,S}^{-1}M_n(\tau,\mathbf{0}) + o_p(\sqrt{k\log\max\{n,r+p\}/n}), \end{aligned} \tag{S2.8}$$

and

$$\begin{aligned} & D_i(\hat{\boldsymbol{\gamma}}_{S_\Delta}(\tau) - \boldsymbol{\gamma}_0(\tau), \tau) \\ = & -\frac{1}{2}M_n(\tau,\mathbf{0})^\top H_\tau^{-1}M_n(\tau,\mathbf{0}) + o_p(\sqrt{k\log\max\{n,r+p\}/n}). \end{aligned} \tag{S2.9}$$

Combining (S2.7)–(S2.9) together yields

$$\begin{aligned} & n(\hat{\sigma}_S(\tau) - \hat{\sigma}_{S_\Delta}(\tau)) \\ = & -\frac{1}{2}M_n(\tau,\mathbf{0})^\top\{H_{\tau,S}^{-1} - H_\tau^{-1}\}M_n(\tau,\mathbf{0}) + o_p(\sqrt{k\log\max\{n,r+p\}/n}). \end{aligned} \tag{S2.10}$$

Let $D = S \setminus S_\Delta$, we have

$$H_{\tau,S}\{H_{\tau,S}^{-1} - H_\tau^{-1}\}H_{\tau,S} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$\begin{aligned}Q =&E\left\{\int_0^\infty f_{t,\tau}(0|\mathbf{V}_i(t))\mathbf{V}_{iD}(t)\mathbf{V}_{iD}(t)^\top dN_i(t)\right\}\\&-E\left\{\int_0^\infty f_{t,\tau}(0|\mathbf{V}_i(t))\mathbf{V}_{iD}(t)\mathbf{V}_{iS_\Delta}(t)^\top dN_i(t)\right\}\\&\times\left[E\left\{\int_0^\infty f_{t,\tau}(0|\mathbf{V}_i(t))\mathbf{V}_{iS_\Delta}(t)\mathbf{V}_{iS_\Delta}(t)^\top dN_i(t)\right\}\right]^{-1}\\&\times E\left\{\int_0^\infty f_{t,\tau}(0|\mathbf{V}_i(t))\mathbf{V}_{iS_\Delta}(t)\mathbf{V}_{iD}(t)^\top dN_i(t)\right\}.\end{aligned}$$

By simple linear algebra, we can show that the largest eigenvalue of Q is less than $\bar{f}\Lambda_{\max}$.

Now we consider $\boldsymbol{\eta}_j^\top H_{\tau,S}^{-1}M_n(\tau,\mathbf{0})$, where $\boldsymbol{\eta}_j$ is defined in Lemma 9. Conditional on $\mathbf{V}_i(\cdot)$, we note that $E[\boldsymbol{\eta}_j^\top H_{\tau,S}^{-1}\int_0^\infty \mathbf{V}_i(t)\psi_\tau\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)]=0$. Also by condition (C5+), $\left\|\boldsymbol{\eta}_j^\top H_{\tau,S}^{-1}\int_0^\infty \mathbf{V}_i(t)\psi_\tau\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)\right\|\le C_b$ for some constant C_b. By Bernstein's inequality,

$$\begin{aligned}&P\big(\sqrt{n}|\boldsymbol{\eta}_j^\top(H_{\tau,S})^{-1}M_n(\tau,\mathbf{0})|>4\sqrt{C_b}\sqrt{n\log\max\{n,r+p\}}\big)\\\le&2\exp\left(-\frac{1}{2}\frac{16C_bn\log\max\{n,r+p\}}{nC_b+4C_b^2\sqrt{n\log\max\{n,r+p\}}/3}\right)\le 2\exp\left(-4\log\max\{n,r+p\}\right).\end{aligned}$$

Thus, $P\big(|\boldsymbol{\eta}_j^\top H_{\tau,S}^{-1}M_n(\tau,\mathbf{0})|_\infty>4\sqrt{C_b}\sqrt{\log\max\{n,r+p\}}\big)\le 2\exp\left(-3\log\max\{n,r+p\}\right)$ and

$$\begin{aligned}&M_n(\tau,\mathbf{0})^\top\{H_{\tau,S}^{-1}-H_\tau^{-1}\}M_n(\tau,\mathbf{0})\\=&M_n(\tau,\mathbf{0})^\top H_{\tau,S}^{-1}H_{\tau,S}\{H_{\tau,S}^{-1}-H_\tau^{-1}\}H_{\tau,S}H_{\tau,S}^{-1}M_n(\tau,\mathbf{0})\\\le&\lambda_{\max}(Q)(\sqrt{k-r-s}4\sqrt{C_b}\sqrt{\log\max\{n,r+p\}})^2\end{aligned}$$

$$\leq 16C_b(k-r-s)\log\max\{n,r+p\}\bar{f}\Lambda_{\max},$$

with probability $1-2\exp(-3\log\max\{n,r+p\})$. This, coupled with (S2.10), yields the desired results. Hence we complete the proof of Lemma 9. □

Proof of Lemma 10. For any $S \in OF$, $|S| \leq \kappa$,

$$\begin{aligned}
&P\Big(\inf_{S\in OF,|S|\leq\kappa}\mathrm{GIC}(S) > \mathrm{GIC}(S_\Delta)\Big)\\
=&P\Big(\inf_{S\in OF,|S|\leq\kappa}\Big[\int_\Delta[\log(\hat{\sigma}_S(\tau)) - \log(\hat{\sigma}_{S_\Delta}(\tau))]d\tau + (|S|-|S_\Delta|)\phi_n\Big] > 0\Big)\\
=&P\Big(\inf_{S\in OF,|S|\leq\kappa}\Big[\int_\Delta \log\Big(1+\frac{\hat{\sigma}_S(\tau)-\hat{\sigma}_{S_\Delta}(\tau)}{\hat{\sigma}_{S_\Delta}(\tau)-\sigma(\tau)+\sigma(\tau)}\Big)d\tau + (|S|-|S_\Delta|)\phi_n\Big] > 0\Big)\\
\geq&P\Big(\inf_{S\in OF,|S|\leq\kappa}\Big[\int_\Delta \min\Big(\log 2, \frac{1}{2}\frac{\hat{\sigma}_S(\tau)-\hat{\sigma}_{S_\Delta}(\tau)}{\hat{\sigma}_{S_\Delta}(\tau)-\sigma(\tau)+\sigma(\tau)}\Big)d\tau + (|S|-|S_\Delta|)\phi_n\Big] > 0\Big).
\end{aligned}$$

By Lemma 9,

$$\inf_{S\in OF,|S|\leq\kappa,\tau\in\Delta}\hat{\sigma}_S(\tau)-\hat{\sigma}_{S_\Delta}(\tau) \geq -8C_b(\underline{f}\Lambda_{\min})^{-1}(\kappa-r-s)n^{-1}\log\max\{n,r+p\}$$

with probability at least $1-16\exp(-(A^2-2)\log\max\{n,r+p\})-2\exp(-3\log\max\{n,r+p\})$. Following lines as in the proof of Lemmas 1 and 2, we can show that

$$\sup_{\tau\in\Delta}|\hat{\sigma}_{S_\Delta}(\tau)-\sigma(\tau)| \leq C_3(r+s)\log n/n$$

with probability at least $1-16/n^3$, where $C_3>0$ is a constant. By Lemma 7,

$$\inf_{\tau\in\Delta}\sigma(\tau) \geq \tau^*\Big[E\Big[\int_0^\infty |Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\Big] - M_0\underline{f}^{-1}\Big]$$

for some $\tau_0 \in \Delta$, where $\tau^* = \min\{\tau_L, 1-\tau_U\}$. Since $\log(r+p)/n = o(\phi_n)$, combining the above results together yields

$$\begin{aligned}
&\inf_{S\in OF, |S|\le\kappa} \Big[\int_\Delta \min\Big(\log 2, \frac{1}{2}\frac{\hat\sigma_S(\tau)-\hat\sigma_{S_\Delta}(\tau)}{\hat\sigma_{S_\Delta}(\tau)-\sigma(\tau)+\sigma(\tau)} \Big) d\tau + (|S|-|S_\Delta|)\phi_n \Big] \\
&\ge \int_\Delta \min\Big(\log 2, \frac{-4C_b(n\underline{f}\Lambda_{\min})^{-1}(\kappa-r-s)\log\max\{n,r+p\}}{-C_3(r+s)\frac{\log n}{n} + \tau^*\big[E\{\int_0^\infty |Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\} - \frac{M_0}{\underline{f}}\big]} \Big) d\tau \\
&\quad + \inf_{S\in OF,|S|\le\kappa} \big[(|S|-|S_\Delta|)\phi_n\big] > 0
\end{aligned}$$

with probability $1-16\exp(-(A^2-2)\log\max\{n,r+p\}) - 2\exp(-3\log\max\{n,r+p\}) - 16/n^3$. This completes the proof of Lemma 10. □

Proof of Lemma 11: Given an underfitted model S with $|S|\le\kappa$, there exists an underfitted model S^+ with $|S^+|=\kappa$, such that $S\subseteq S^+$, and $S_\Delta - S = S_\Delta - S^+$. Thus,

$$\begin{aligned}
&\inf_{S\in UF,|S|\le\kappa} GIC(S) - GIC(S_\Delta) \\
&= \inf_{S\in UF,|S|\le\kappa} \Big[\int_\Delta [\log(\hat\sigma_S(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau + |S|\phi_n - |S_\Delta|\phi_n \Big] \\
&\ge \inf_{S^+\in UF,|S^+|=\kappa} \Big[\int_\Delta [\log(\hat\sigma_{S^+}(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau \Big] - |S_\Delta - S^+|\phi_n, \quad \text{(S2.11)}
\end{aligned}$$

where the last inequality follows from the fact that $\hat{\boldsymbol\beta}_S(\tau)$ is suboptimal for model S^+ but $\hat{\boldsymbol\beta}_{S^+}(\tau)$ is optimal or model S^+, and $|S|-|S_\Delta| \ge -|S_\Delta - S^+|$.

For $j\in S_\Delta$, we have $\int_\Delta |\beta_{0,j}(\tau)|d\tau \ge \xi_n$, and there exists $\Delta_j(\xi_n/2) := \{\tau\in\Delta : |\beta_{0,j}(\tau)| > \xi_n/2\}$, such that $\int_{\Delta_j(\xi_n/2)} |\beta_{0,j}(\tau)|d\tau > \xi_n/2$. By condition (C3), we have

$$\mathcal{L}(\Delta_j(\xi_n/2))\Big(\xi_n/2 + L\mathcal{L}(\Delta_j(\xi_n/2))\Big) \ge \int_{\Delta_j(\xi_n/2)} |\beta_{0,j}(\tau)|d\tau > \xi_n/2,$$

where $\mathcal{L}(\cdot)$ denote the Lebesgue measure. Thus, $\mathcal{L}(\Delta_j(\xi_n/2)) > \sqrt{\xi_n/(2L)} - \xi_n/(4L)$ for every $j \in S_\Delta$.

Since $S^+ \in UF$, $|S_\Delta - S^+| > 0$, then let $\Delta_{S^+}(\xi_n/2) := \bigcup \Delta_j(\xi_n/2), j \in S_\Delta - S^+$,

$$\begin{aligned}
&\int_\Delta [\log(\hat\sigma_{S^+}(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau - |S_\Delta - S^+|\phi_n \\
=& \int_{\Delta_{S^+}(\xi_n/2)} [\log(\hat\sigma_{S^+}(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau - |S_\Delta - S^+|\phi_n \\
&+ \int_{\Delta\setminus\Delta_{S^+}(\xi_n/2)} [\log(\hat\sigma_{S^+}(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau.
\end{aligned}$$

We first consider $\int_{\Delta_{S^+}(\xi_n/2)} [\log(\hat\sigma_{S^+}(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau - |S_\Delta - S^+|\phi_n$. By Lemma 8, with probability at least $1-8\exp\left(-(A^2-1)\log\max\{n, r+p\}\right) - 8\exp\left(-(A^2-3/2)\times \log\max\{n, r+p\} + \log L - \log(\xi_n/2)\right)$,

$$\begin{aligned}
&|\hat\sigma_{S^+}(\tau) - \sigma(\tau)| \\
\geq{}& m(\tau)\left[\frac{\underline{f}\Lambda_{\min}}{16}\xi_n^2 - A\sqrt{\kappa/n}\,\xi_n\Big(12\sqrt{\log\max\{n, r+p\}} \right. \\
&\left. \qquad + 100\sqrt{\log\max\{n, r+p\} + \log L/2 - \log \xi_n/4}\Big)\right] \\
\geq{}& C_4 m(\tau)\xi_n^2,
\end{aligned}$$

for $\tau \in \Delta_{S^+}(\xi_n/2)$, where $m(\tau) = |S_\tau - S^+|$ and C_4 is some constant. With probability at least $1-8\exp\left(-(A^2-1)\log\max\{n, r+p\}\right) - 8\exp\left(-(A^2-3/2)\log\max\{n, r+p\} + \log L - \log(\xi_n/2)\right)$,

$$\inf_{S^+\in UF, |S^+|=\kappa} \int_{\Delta_{S^+}(\xi_n/2)} [\log(\hat\sigma_{S^+}(\tau)) - \log(\hat\sigma_{S_\Delta}(\tau))]\, d\tau - |S_\Delta - S^+|\phi_n$$

$$\geq \int_{\Delta_{S^+}(\xi_n/2)} \min\left(\log 2, \frac{1}{2}\frac{(\hat{\sigma}_{S^+}(\tau)-\sigma(\tau))-(\hat{\sigma}_{S_\Delta}(\tau)-\sigma(\tau))}{\hat{\sigma}_{S_\Delta}(\tau)-\sigma(\tau)+\sigma(\tau)}\right)d\tau - |S_\Delta - S^+|\phi_n$$
$$\geq \frac{1}{2}\int_{\Delta_{S^+}(\xi_n/2)} \frac{C_4 m(\tau)\xi_n^2 - C_3(r+s)n^{-1}\log n}{C_3(r+s)n^{-1}\log n + E\{\int_0^\infty |Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\} + M_0\underline{f}^{-1}}d\tau$$
$$- |S_\Delta - S^+|\phi_n$$
$$\geq \frac{1}{4}\frac{|S_\Delta - S^+|C_4\xi_n^2}{E\{\int_0^\infty |Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\} + M_0\underline{f}^{-1}}\left[\frac{\sqrt{\xi_n}}{\sqrt{2L}} - \frac{\xi_n}{4L}\right] - |S_\Delta - S^+|\phi_n,$$

where the first two inequalities follow exactly as in Lemma 10, and the last inequality follows from $\mathcal{L}(\Delta_j(\xi_n/2)) > \sqrt{\xi_n/(2L)} - \xi_n/(4L)$ and $C_3(r+s)n^{-1}\log n = o(\xi_n^2)$.

With similar arguments and the condition $\kappa n^{-1}\log\max\{n, r+p\} = o(\xi_n^3)$, we can show

$$\inf_{S^+\in UF, |S^+|=\kappa}\int_{\Delta\setminus\Delta_{S^+}(\xi_n/2)}[\log(\hat{\sigma}_{S^+}(\tau)) - \log(\hat{\sigma}_{S_\Delta}(\tau))]\,d\tau = o(\xi_n^{5/2}).$$

Combining the above two results yields

$$\begin{aligned}
&\inf_{S^+\in UF, |S^+|=\kappa}\int_{\Delta}[\log(\hat{\sigma}_{S^+}(\tau)) - \log(\hat{\sigma}_{S_\Delta}(\tau))]\,d\tau - |S_\Delta - S^+|\phi_n \\
\geq\ & \frac{1}{4}\frac{|S_\Delta - S^+|C_4\xi_n^2}{E\{\int_0^\infty |Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau_0)|dN_i(t)\} + M_0\underline{f}^{-1}}\left[\frac{\sqrt{\xi_n}}{\sqrt{2L}} - \frac{\xi_n}{4L}\right] - |S_\Delta - S^+|\phi_n \\
& -o({\xi_n}^{5/2}) > 0. \qquad \text{(S2.12)}
\end{aligned}$$

As $\log(\xi_n) \geq \log(n^{-1}\log\max\{n, r+p\})/2$, with probability at least $1 - 16\exp(-(A^2 - 2)\log\max\{n, r+p\}) - 16/n^3$, we have

$$\inf_{S^+\in UF, |S^+|=\kappa}\left[\int_{\Delta}[\log(\hat{\sigma}_{S^+}(\tau)) - \log(\hat{\sigma}_{S_\Delta}(\tau))]\,d\tau\right] - |S_\Delta - S^+|\phi_n > 0.$$

This and $(S2.11)$ complete the proof of Lemma 11. □

S3 Additional Theorems and Proofs

In this section, we provide the statements and proofs of Theorem A.1, Corollary A.1, Theorem A.2 and Theorem A.3, which play a critical role in proving the results in Theorems 1–2.

Theorem A.1. *Under conditions (C1)–(C5), if* $r+s = o(n^{1/3})$, *and* $\sup_{\tau\in\Delta, j\in S_\tau} \lambda\omega_j(\tau) = O_p(\sqrt{n\log n})$, *then the oracle regularized estimator satisfies*

$$\sup_{\tau\in\Delta} \|\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\| = O_p(\sqrt{(r+s)\log n/n}).$$

Corollary A.1. *Suppose conditions (C1)–(C5) hold. Given a fixed* τ_0, *if* $r+s = o(n^{1/3})$, *and* $\max_{j\in S_{\tau_0}} \lambda w_j(\tau_0) = O_p(\sqrt{n})$, *then the oracle regularized estimator satisfies*

$$\|\hat{\boldsymbol{\gamma}}^o(\tau_0) - \boldsymbol{\gamma}_0(\tau_0)\| = O_p(\sqrt{(r+s)/n}).$$

Theorem A.2. *Under conditions (C1)–(C5), if* $\sup_{\tau\in\Delta, j\in S_\tau} \lambda\omega_j(\tau) = O_p(\sqrt{n\log n})$ *and* $(r+s)^3\log^4 n = o(n)$ *for any given* $\boldsymbol{\xi} \in R_{r+s-1}$ *and* $\|\boldsymbol{\xi}\| = 1$, *then we have the following results:*

(a) If $\sqrt{n/\{(r+s)\log n\}}\inf_{1\le j\le s, \tau\in\Delta} |\beta_{0j}(\tau)| \to \infty$, *then*

$$n^{1/2}\boldsymbol{\xi}^\top \left[\mathbf{H}_\tau \left\{\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\right\} + \frac{\lambda}{n}\boldsymbol{\varpi}(\tau)\right]$$

converges weakly to a mean zero Gaussian process with covariance

$$\boldsymbol{\Sigma}(\tau,\tau') = E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t),Y)h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t),Y)\} - E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t),Y)\}E\{h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t),Y)\},$$

where $h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t),Y)=\int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t)\psi_\tau\{Y(t)-\mathbf{V}(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN(t)$, $\psi_\tau(u)=\tau-I(u<0)$,

$$\mathbf{H}_\tau = \begin{pmatrix} E[\int_0^\infty f_{t,\tau}\{0|\mathbf{V}_i(t)\}\,\mathbf{V}_{ia}(t)\mathbf{V}_{ia}(t)^\top dN_i(t)] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$\boldsymbol{\varpi}(\tau)=\left(\mathbf{0}_r^\top,(\boldsymbol{\omega}(\tau)\circ sign(\boldsymbol{\beta}_0(\tau)))^\top,\mathbf{0}_{p-s}^\top\right)^\top$, $\circ$ *denotes the Hadamard product, and* $\boldsymbol{\omega}(\tau)=(\omega_1(\tau),\ldots,\omega_p(\tau))^\top$;

(b) If $\sup_{\tau\in\Delta} n^{-1/2}\{\sum_{j\in S_\tau}\lambda w_j^2(\tau)\}^{1/2}=o_p(1)$, *then* $n^{1/2}\boldsymbol{\xi}^\top\mathbf{H}_\tau\{\hat{\boldsymbol{\gamma}}^o(\tau)-\boldsymbol{\gamma}_0(\tau)\}$ *converges weakly to a mean zero Gaussian process with covariance* $\boldsymbol{\Sigma}(\tau,\tau')$.

Theorem A.3. *Suppose conditions (C1)–(C5) hold. Furthermore, we assume that* $n/((r+s)^3\log^2\max\{n,r+p\})\to\infty$ *and*

$$\sup_{j>r+s,\boldsymbol{\delta}\in R_{r+s-1}}\frac{E[\int_0^\infty\{V_{ij}(t)\mathbf{V}_i(t)^\top\boldsymbol{\delta}\}^2dN_i(t)]}{\|\boldsymbol{\delta}\|^2}=o\left(\frac{\log\max\{n,r+p\}}{(r+s)\log n}\right).$$

If $\sup_{j\in S_\Delta,\tau\in\Delta}\lambda w_j(\tau)=O_p(\sqrt{n\log n})$, $\lambda/(\sqrt{r+s}\log\max\{n,r+p\})\to\infty$, *and* $(\inf_{j>r+s,\tau\in\Delta}w_j(\tau))^{-1}\sqrt{n}/\sqrt{(r+s)\log\max\{n,r+p\}}=O_p(1)$, *then we have*

$$P\big(\sup_{\tau\in\Delta}|\hat{\boldsymbol{\gamma}}^o(\tau)-\hat{\boldsymbol{\gamma}}(\tau)|=\mathbf{0}\big)\to 1.$$

Proofs of Theorem A.1, Corollary A.1, Theorem A.2 and Theorem A.3

Proof of Theorem A.1: For any $\boldsymbol{\gamma}=(\boldsymbol{\alpha}^\top,\boldsymbol{\beta}^\top)^\top\in R_{r+s-1}$, let

$$Q(\boldsymbol{\gamma};\tau)=Q(\boldsymbol{\alpha},\boldsymbol{\beta};\tau):=\mathbb{E}_n\left(\int_0^\infty\rho_\tau\{Y_i(t)-\mathbf{V}_i(t)^\top\boldsymbol{\gamma}\}dN_i(t)\right)$$

and $L(\boldsymbol{\gamma};\tau):=Q(\boldsymbol{\gamma};\tau)+\lambda\sum_{j=1}^p w_j(\tau)|\beta_j|$, where $\boldsymbol{\gamma}$ in $L(\boldsymbol{\gamma};\tau)$ subjects to linear constraint. For any $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma} \in R_{r+s-1}$, $\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0(\tau)\| \leq u$, let $\boldsymbol{\delta} = \boldsymbol{\gamma} - \boldsymbol{\gamma}_0(\tau)$. Then,

$$
\begin{aligned}
L(\boldsymbol{\gamma}, \tau) - L(\boldsymbol{\gamma}_0(\tau), \tau) &= \sum_{i=1}^{n} D_i(\boldsymbol{\delta}, \tau) + \lambda \sum_{j=1}^{p} \omega_j(\tau)(|\beta_j| - |\beta_{0j}(\tau)|) \\
=& nE[D_i(\boldsymbol{\delta}, \tau)] + \sqrt{n}\mathbb{G}_n[D_i(\boldsymbol{\delta}, \tau)] + \lambda \sum_{j=1}^{p} \omega_j(\tau)(|\beta_j| - |\beta_{0j}(\tau)|) =: I_1 + I_2 + I_3.
\end{aligned}
$$

According to Lemma 1, we have if $u \leq 3q\underline{f}\lambda_{\min}/(4C_f\lambda_{\max}^{3/2})$,

$$
\inf_{\tau\in\Delta, \boldsymbol{\delta}\in R_{s+r-1}, \|\delta\|=u} I_1 \geq n\underline{f}\lambda_{\min}u^2/8.
$$

By Lemma 2, with probability at least $1 - 16n^{-3}$,

$$
\begin{aligned}
&\sup_{\tau\in\Delta, \boldsymbol{\delta}\in R_{s+r-1}, \|\delta\|\leq u} |I_2| \\
\leq& M_0 C_v \sqrt{r+s}u\Big(24\sqrt{\log n} + 200\sqrt{\log n + \log L/2 - \log u/2}\Big)\sqrt{n}.
\end{aligned}
$$

For I_3, since $\sup_{\tau\in\Delta, j\in S_\tau} \lambda w_j(\tau) = O_p(\sqrt{n\log n})$, then we obtain

$$
\begin{aligned}
I_3 \geq& \lambda \sum_{j=1}^{s} \omega_j(\tau)(|\beta_j| - |\beta_{0j}(\tau)|) \\
=& \lambda \sum_{j=1}^{s-1} \omega_j(\tau)(|\beta_j| - |\beta_{0j}(\tau)|) + \lambda\omega_s(\tau)\Big(\Big| - \sum_{j=1}^{s-1}\beta_j\Big| - \Big| - \sum_{j=1}^{s-1}\beta_{0j}(\tau)\Big|\Big) \\
\geq& -\max_{1\leq j\leq s-1} \lambda\{\omega_j(\tau) + \omega_s(\tau)\} \sum_{j=1}^{s-1}(|\beta_j - \beta_{0j}(\tau)|) \\
\geq& -\max_{1\leq j\leq s-1} \lambda\{\omega_j(\tau) + \omega_s(\tau)\}\sqrt{s-1}u = O_p(\sqrt{n(r+s)\log n}u).
\end{aligned}
$$

The results for I_1, I_2, and I_3 together imply that with probability at least $1 - 16n^{-3}$,

$$
\inf_{\tau\in\Delta, \boldsymbol{\gamma}\in R_{r+s-1}, \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0(\tau)\|=u} L(\boldsymbol{\gamma}, \tau) - L(\boldsymbol{\gamma}_0(\tau), \tau)
$$

$$\geq \frac{\underline{f}\lambda_{\min}}{8}C_1^2(r+s)\log n - O_p((r+s)\log n)C_1 \geq 0,$$

by choosing $u = C_1\sqrt{(r+s)\log n/n}$ for some sufficiently large C_1. Since $L(\boldsymbol{\gamma},\tau)$ is a convex function with respect to $\boldsymbol{\gamma}$ in R_{r+s-1}, then $\hat{\boldsymbol{\gamma}}^o(\tau)$ must locate within $\{\boldsymbol{\gamma} : \boldsymbol{\gamma} \in R_{r+s-1}, \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0(\tau)\| \leq C_1\sqrt{(r+s)\log n/n}\}$. This completes the proof of Theorem A.1. □

Proof of Corollary A.1: We can follow the same lines as the proof of Theorem A.1. We apply Lemma A.3 into I_2 rather than using Lemma A.2. The condition $\max_{j\in S_{\tau_0}} \lambda w_j(\tau_0) = O_p(\sqrt{n})$, coupled with Lemma 3, implies $\|\hat{\boldsymbol{\gamma}}^o(\tau_0) - \boldsymbol{\gamma}_0(\tau_0)\| = O_p(\sqrt{(r+s)/n})$. This completes the proof of Corollary A.1. □

Proof of Theorem A.2. We know from Theorem A.1 that for any $\epsilon > 0$, there exists a constant B such that

$$P\left(\|\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\| \leq B\sqrt{(r+s)\log n/n}\right) > 1-\epsilon.$$

Therefore, with probability at least $1-\epsilon$, minimizing the objective function $L(\boldsymbol{\gamma},\tau)$ over R_{r+s-1} is equivalent to minimizing $n^{-1/2}\{L(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta},\tau) - L(\boldsymbol{\gamma}_0(\tau),\tau)\}$ over $R_{r+s-1}(B)$.

$$\begin{aligned}
&n^{-1/2}\{L(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta},\tau) - L(\boldsymbol{\gamma}_0(\tau),\tau)\} \\
=&\mathbb{G}_n\Big\{D_i(\boldsymbol{\delta},\tau)\Big\} + n^{1/2}E\Big\{D_i(\boldsymbol{\delta},\tau)\Big\} + n^{-1/2}\lambda\sum_{j=1}^{s} w_j(\tau)\left(|\beta_{0j}(\tau)+\delta_{j+r}| - |\beta_{0j}(\tau)|\right)
\end{aligned}$$

$$
\begin{aligned}
=&\mathbb{G}_n\Big(\int_0^\infty \{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\}\big[\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) - \mathbf{V}_i(t)^\top \boldsymbol{\delta}\} \\
&\qquad - \psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\}\big] dN_i(t)\Big) \\
&- \boldsymbol{\delta}^\top [M_n(\tau, \boldsymbol{\delta}) - E\{M_n(\tau, \boldsymbol{\delta})\} - M_n(\tau, \mathbf{0}) + E\{M_n(\tau, \mathbf{0})\}] \\
&- \boldsymbol{\delta}^\top [M_n(\tau, \mathbf{0}) - E\{M_n(\tau, \mathbf{0})\}] \\
&+ n^{1/2} E\Big\{D_i(\boldsymbol{\delta}, \tau)\Big\} + n^{-1/2}\lambda \sum_{j=1}^{s} w_j(\tau)\left(|\beta_{0j}(\tau) + \delta_{j+r}| - |\beta_{0j}(\tau)|\right) \\
=:&\mathrm{II}_1 + \mathrm{II}_2 + \mathrm{II}_3 + \mathrm{II}_4 + \mathrm{II}_5.
\end{aligned}
$$

For II_1, note that

$$
\begin{aligned}
&\int_0^\infty \{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\}\Big[\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) - \mathbf{V}_i(t)^\top \boldsymbol{\delta}\} \\
&\qquad -\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\}\Big] dN_i(t) \\
= &\int_0^\infty \Big[- \{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\} I\{0 < Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) < \mathbf{V}_i(t)^\top \boldsymbol{\delta}\} \\
&\qquad + \{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\} I\{\mathbf{V}_i(t)^\top \boldsymbol{\delta} < Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) < 0\}\Big] dN_i(t).
\end{aligned}
$$

It is easy to see that

$$
\begin{aligned}
&\int_0^\infty \frac{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)}{\mathbf{V}_i(t)^\top \boldsymbol{\delta}}\Big[\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) - \mathbf{V}_i(t)^\top \boldsymbol{\delta}\} \\
&\qquad -\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\}\Big] dN_i(t)
\end{aligned}
$$

is a bounded function. Following similar arguments as in the proof of Lemma 4, we have

$$
\sup_{\boldsymbol{\delta} \in R_{r+s-1}(B), \tau \in \Delta} |\mathrm{II}_1| = o_p(1). \tag{S3.1}
$$

For II_2, it follows from Lemma 5 that,

$$\sup_{\boldsymbol{\delta}\in R_{r+s-1}(B),\tau\in\Delta} |\text{II}_2| = o_p(1). \tag{S3.2}$$

For II_3, since $E[M_n(\tau,\mathbf{0})] = 0$, then

$$\text{II}_3 = -\boldsymbol{\delta}^\top M_n(\tau,\mathbf{0}). \tag{S3.3}$$

For II_4, following similar arguments as in the proof of Lemma 1, we can obtain that

$$\begin{aligned} & \sup_{\boldsymbol{\delta}\in R_{r+s-1}(B),\tau\in\Delta} |\text{II}_4 - \frac{1}{2}n^{1/2}\boldsymbol{\delta}^\top \mathbf{H}_\tau \boldsymbol{\delta}| \\ = \;& O\Big\{(r+s)n^{-1/2}\log n \sup_{\boldsymbol{\delta}\in R_{r+s-1}(B)} \|\boldsymbol{\delta}_a\|\Big\} = o_p(1). \end{aligned} \tag{S3.4}$$

For II_5, we consider two cases:

(a) If $\sqrt{n/\{(r+s)\log n\}}\inf_{\tau\in\Delta,j\le s}|\beta_{0j}(\tau)| \to \infty$, then

$$\begin{aligned} & \text{II}_5 = n^{-1/2}\lambda\sum_{j=1}^{s}\omega_j(\tau)\left(|\beta_{0j}(\tau)+\delta_{j+r}| - |\beta_{0j}(\tau)|\right) \\ = \;& n^{-1/2}\lambda\sum_{j=1}^{s}\omega_j(\tau)\text{sgn}\{\beta_{0j}(\tau)\}\delta_{j+r} = n^{-1/2}\lambda\boldsymbol{\varpi}(\tau)^\top\boldsymbol{\delta}. \end{aligned} \tag{S3.5}$$

Combining the above equations (S3.1)–(S3.5) together, we have

$$\begin{aligned} & n^{-1/2}\{L(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta},\tau) - L(\boldsymbol{\gamma}_0(\tau),\tau)\} \\ = \;& \frac{1}{2}n^{1/2}\boldsymbol{\delta}^\top\mathbf{H}_\tau\boldsymbol{\delta} - \boldsymbol{\delta}^\top M_n(\tau,\mathbf{0}) + n^{-1/2}\lambda\boldsymbol{\varpi}(\tau)^\top\boldsymbol{\delta} + o_p(1) \end{aligned}$$

for any $\boldsymbol{\delta} \in R_{r+s-1}(B)$. Let $\hat{\boldsymbol{\delta}}$ denote the minimizer of $n^{-1/2}\{L(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta},\tau) - L(\boldsymbol{\gamma}_0(\tau),\tau)\}$ over $R_{r+s-1}(B)$, then

$$n^{1/2}\left(\mathbf{H}_\tau\hat{\boldsymbol{\delta}} + \frac{\lambda}{n}\boldsymbol{\varpi}(\tau)\right) = M_n(\tau,\mathbf{0}). \tag{S3.6}$$

For any given $\boldsymbol{\xi} \in R_{r+s-1}, \|\boldsymbol{\xi}\| = 1$, Lemma 5 implies that

$$n^{1/2}\boldsymbol{\xi}^\top \left[\mathbf{H}_\tau \left\{\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\right\} + \frac{\lambda}{n}\boldsymbol{\varpi}(\tau)\right]$$

converges weakly to a mean zero Gaussian process with covariance $\boldsymbol{\Sigma}(\tau, \tau')$.

(b) If $\sup_{\tau\in\Delta} n^{-1/2}(\sum_{j\in S_\tau} \lambda w_j^2(\tau))^{1/2} = o_p(1)$, then

$$\begin{aligned} &\sup_{\boldsymbol{\delta}\in R_{r+s-1}(B),\tau\in\Delta} \mathrm{II}_5 \\ =& \sup_{\boldsymbol{\delta}\in R_{r+s-1}(B),\tau\in\Delta} \left| n^{-1/2}\lambda \sum_{j=1}^{s} \omega_j(\tau)\left(|\beta_{0j}(\tau) + \delta_{j+r}| - |\beta_{0j}(\tau)|\right) \right| = o_p(1), \end{aligned} \quad \text{(S3.7)}$$

where the last equality follows from the Cauchy-Schwartz inequality. Consequently, it follows from (S3.1)–(S3.4) and (S3.7) that

$$n^{-1/2}\{L(\boldsymbol{\gamma}_0(\tau) + \boldsymbol{\delta}; \tau) - L(\boldsymbol{\gamma}_0(\tau); \tau)\} = \frac{1}{2}n^{1/2}\boldsymbol{\delta}^\top H_\tau \boldsymbol{\delta} - \boldsymbol{\delta}^\top M_n(\tau, \mathbf{0}) + o_p(1).$$

With similar arguments as in case (a), we can conclude that $n^{1/2}\boldsymbol{\xi}^\top\mathbf{H}_\tau \left\{\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\right\}$ converges weakly to a mean zero Gaussian process with covariance function $\boldsymbol{\Sigma}(\tau, \tau')$. This completes the proof of Theorem A.2. □

Proof of Theorem A.3. By KKT conditions, $\hat{\boldsymbol{\gamma}}^o(\tau)$ satisfies that

$$\begin{aligned} &\sum_{i=1}^{n}\int_0^\infty Z_{ij}(t)\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top\hat{\boldsymbol{\gamma}}^o(\tau)\}dN_i(t) - \varrho \\ &= \lambda w_j(\tau)u_j, \text{ where } u_j = \begin{cases} \operatorname{sign}\left(\hat{\beta}_j^o\right), & \text{if } \hat{\beta}_j^o \neq 0, \\ \in [-1, 1], & \text{if } \hat{\beta}_j^o = 0, \end{cases} \end{aligned}$$

for some ϱ and $1 \le j \le s$. If there exist $j_1, j_2 \in S_\Delta$ such that $\hat{\beta}^o_{j_1}(\tau), \hat{\beta}^o_{j_2}(\tau) \neq 0$, then we can find a_{j_1} and a_{j_2} such that $a_{j_1}+a_{j_2} = 1$ and $\sum_{i=1}^n \int_0^\infty (a_{j_1} Z_{ij_1}(t) + a_{j_2} Z_{ij_2}(t)) \psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \hat{\boldsymbol{\gamma}}^o(\tau)\} dN_i(t) = 0$. As a result,

$$\begin{aligned} |\varrho| &= \Big| \sum_{i=1}^n \int_0^\infty \Big(a_{j_1} Z_{ij_1}(t) + a_{j_2} Z_{ij_2}(t) \Big) \psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \hat{\boldsymbol{\gamma}}^o(\tau)\} dN_i(t) \\ &\qquad - a_{j_1} \lambda w_{j_1}(\tau) u_{j_1} - a_{j_2} \lambda w_{j_2}(\tau) u_{j_2} \Big| \\ &= |a_{j_1} \lambda w_{j_1}(\tau) u_{j_1} + a_{j_2} \lambda w_{j_2}(\tau) u_{j_2}| = O_p(n^{1/2} \log^{1/2} n), \end{aligned} \tag{S3.8}$$

where the last equality follows from that $\sup_{j \in S_\Delta, \tau \in \Delta} \lambda w_j(\tau) = O_p(n^{1/2} \log^{1/2} n)$.

To show $\hat{\boldsymbol{\gamma}}^o(\tau)$ is a global minimizer of $L(\boldsymbol{\gamma}; \tau)$ over $\{\boldsymbol{\gamma} : \boldsymbol{\gamma} \in R^{r+p}, \sum_{j=1}^p \gamma_{r+j} = 0\}$, it is sufficient to have the following KKT condition

$$\sup_{\tau \in \Delta} \Big\{ \Big| \sum_{i=1}^n \int_0^\infty V_{ij}(t) \psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \hat{\boldsymbol{\gamma}}^o(\tau)\} dN_i(t) \Big| + \varrho - \lambda w_j(\tau) \Big\} < 0$$

for $j > r+s$. By Theorem A.1 that for any $\epsilon > 0$, there exists a constant B such that

$$P\Big(\|\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\| \le B\sqrt{(r+s)\log n / n} \Big) > 1 - \epsilon.$$

Therefore, if we can show that

$$\begin{aligned} \sup_{\substack{\tau \in \Delta \\ j > r+s \\ \boldsymbol{\delta} \in R_{r+s-1}(B)}} & \Big| \sum_{i=1}^n \int_0^\infty V_{ij}(t) \psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) - \mathbf{V}_i(t)^\top \boldsymbol{\delta}\} dN_i(t) \Big| \\ & + \varrho - \lambda w_j(\tau) < 0, \end{aligned} \tag{S3.9}$$

then with probability approaching to 1, $\hat{\boldsymbol{\gamma}}^o(\tau)$ is also the global minimizer of $L(\boldsymbol{\gamma}; \tau)$ over $\{\boldsymbol{\gamma} : \boldsymbol{\gamma} \in R^{r+p}, \sum_{j=1}^p \gamma_{r+j} = 0\}$.

Let $\boldsymbol{\eta}_j$ be a $(r+p)$ dimensional vector whose jth component is 1, and all other components are 0, then $\sum_{i=1}^n \int_0^\infty V_{ij}(t)\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta})\}dN_i(t) = n^{1/2}\boldsymbol{\eta}_j^\top M_n(\tau,\boldsymbol{\delta})$ and

$$\begin{aligned} & n^{1/2}\boldsymbol{\eta}_j^\top M_n(\tau,\boldsymbol{\delta}) \\ = \; & n^{1/2}\boldsymbol{\eta}_j^\top[M_n(\tau,\boldsymbol{\delta}) - E\{M_n(\tau,\boldsymbol{\delta})\} - M_n(\tau,\mathbf{0}) + E\{M_n(\tau,\mathbf{0})\}] \\ & + n^{1/2}\boldsymbol{\eta}_j^\top[E\{M_n(\tau,\boldsymbol{\delta})\} - E\{M_n(\tau,\mathbf{0})\}] + n^{1/2}\boldsymbol{\eta}_j^\top M_n(\tau,\mathbf{0}) \\ := \; & \mathrm{IV}_1 + \mathrm{IV}_2 + \mathrm{IV}_3, \end{aligned} \tag{S3.10}$$

For IV_1, by Lemma 6,

$$\begin{aligned} & P\left(\sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} |\mathrm{IV}_1| > C_2 n^{1/4}(r+s)^{3/4}\log\max\{n,r+p\}\right) \\ \leq \; & 16\exp\left(-\frac{1}{2}(r+s)\log\max\{n,r+p\}\right). \end{aligned} \tag{S3.11}$$

For IV_2, $\forall\ \boldsymbol{\delta} \in R_{r+s-1}(B)$,

$$\begin{aligned} & n^{1/2}\boldsymbol{\eta}_j^\top[E\{M_n(\tau,\boldsymbol{\delta})\} - E\{M_n(\tau,\mathbf{0})\}] \\ = \; & nE\Big(\int_0^\infty V_{ij}(t)[I\{Y_i(t) - \mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta}) < 0\} \\ & \qquad - I\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau) < 0\}]dN_i(t)\Big) \\ \leq \; & \bar{f}nE\Big[\int_0^\infty |V_{ij}(t)\mathbf{V}_i(t)^\top\boldsymbol{\delta}|dN_i(t)\Big] \\ \leq \; & \bar{f}nM_0^{1/2}\left(E\Big[\int_0^\infty \left\{V_{ij}(t)\mathbf{V}_i(t)^\top\boldsymbol{\delta}\right\}^2 dN_i(t)\Big]\right)^{1/2} \\ = \; & \bar{f}nM_0^{1/2}B^2 o\left(\left[\frac{\log\max\{n,r+p\}}{(r+s)\log n}(r+s)n^{-1}\log n\right]^{1/2}\right) \end{aligned}$$

$$= o\left(n^{1/2}\log^{1/2}\max\{n, r+p\}\right), \tag{S3.12}$$

where the first inequality follows from the conditional expectation given $\mathbf{X}_i(t), \mathbf{Z}_i(t)$, along with condition (C2), the second inequality follows from the Cauchy-Schwarz inequality, and the second equality follows from the assumed condition. Hence

$$\sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)} |\mathrm{IV}_2| = o\left(n^{1/2}\log^{1/2}\max\{n, r+p\}\right).$$

Next, we consider IV_3. We note that

$$n^{1/2}\boldsymbol{\eta}_j^\top M_n(\tau, \mathbf{0}) = \sum_{i=1}^n \int_0^\infty V_{ij}(t)\{\tau - I\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau) < 0\}\}dN_i(t).$$

It follows from Lemma 2.3.7 in van der Vaart and Wellner (1996) that

$$P\Big(n^{-1/2}\sup_{\tau\in\Delta}\Big|\sum_{i=1}^n\int_0^\infty V_{ij}(t)\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)\Big| > M\Big)$$
$$\leq \frac{2P\left(n^{-1/2}\sup_{\tau\in\Delta}\left|\sum_{i=1}^n\int_0^\infty \varepsilon_i V_{ij}(t)\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)\right| > M/4\right)}{1 - 16M_0^2C_V^2/M^2}$$

for $M^2 > 16M_0^2C_V^2$, where $\{\varepsilon_i\}_{i=1}^n$ are independent and identically distributed random variables taking values of ±1 with probability $1/2$, and they are independent of $(Y_i(t), \mathbf{X}_i(t), \mathbf{Z}_i(t))$. On one hand,

$$P\Big(\tau\Big|\sum_{i=1}^n\int_0^\infty \varepsilon_i V_{ij}(t)dN_i(t)\Big| > 2M_0C_V\sqrt{n\log(r+p)}\Big)$$
$$\leq P\Big(\Big|\sum_{i=1}^n\int_0^\infty \varepsilon_i V_{ij}(t)dN_i(t)\Big| > 2M_0C_V\sqrt{n\log(r+p)}\Big)$$
$$\leq 2\exp\left(-\frac{1}{2}\frac{4nM_0^2C_V^2\log(r+p)}{\sum_{i=1}^n\{\int_0^\infty V_{ij}(t)dN_i(t)\}^2}\right) \leq 2\exp\left(-\frac{1}{2}\frac{4nM_0^2C_V^2\log(r+p)}{nC_V^2M_0^2}\right) \leq \frac{2}{(r+p)^2},$$

where the first inequality is elementary, the second inequality follows from Hoeffding inequality and the last inequality follows from condition (C2). On the other hand, we would like to bound $\sum_{i=1}^n \int_0^\infty \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) < 0\right\} dN_i(t)$. Let $\tau_L = \tau_1 < \tau_2 < \cdots < \tau_{m_n} = \tau_U$ be a τ-grid such that $\tau_{k+1} - \tau_k = \sqrt{\log(r+p)/n}$, then $m_n \leq (\tau_U - \tau_L)/(\sqrt{\log(r+p)/n})$.

First, we bound $\sum_{i=1}^n \int_0^\infty \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) < 0\right\} dN_i(t)$ for all $1 \leq k \leq m_n$. Given k, let $S_{ik} = \int_0^\infty \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) < 0\right\} dN_i(t)$, then $|S_{ik}| \leq M_0 C_V$ and $\mathrm{Var}(S_{ik}) \leq M_0^2 C_V^2$, since $\int_0^\infty dN_i(t) \leq M_0$ and $|V_{ij}(t)| \leq C_V$ by conditions (C1) and (C2). By Bernstein's inequality, we obtain that

$$
\begin{aligned}
&\Pr\left(\Big|\sum_{i=1}^n S_{ik}\Big| > 4M_0 C_V \sqrt{n \log \max\{n, r+p\}}\right) \\
&\leq 2\exp\left(-\frac{1}{2}\frac{-16 M_0^2 C_V^2 n \log \max\{n, r+p\}}{n M_0^2 C_V^2 + 4 M_0^2 C_V^2 n^{1/2} \log^{1/2} \max\{n, r+p\}/3}\right) = 2\exp(-4\log(r+p)),
\end{aligned}
$$

when n is sufficiently large such that $16 M_0^2 C_V^2 \sqrt{n \log \max\{n, r+p\}}/3 \leq n M_0^2 C_V^2$. Consequently,

$$
\begin{aligned}
& P\Big(\sup_{1\leq k \leq m_n} \Big|\sum_{i=1}^n S_{ik}\Big| > 4M_0 C_V \sqrt{n \log \max\{n, r+p\}}\Big) \\
\leq\ & 2 m_n \exp(-4 \log \max\{n, r+p\}) \\
\leq\ & 2\exp(-4\log\max\{n, r+p\} + \log n/2 - \log\log\max\{n, r+p\}/2) \\
\leq\ & 2\exp\left(-3\log\max\{n, r+p\}\right). \qquad \text{(S3.13)}
\end{aligned}
$$

Then we consider $\sum_{i=1}^n \int_0^\infty \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) < 0\right\} dN_i(t)$ for any $\tau_k < \tau <$

τ_{k+1}. It is easy to see that

$$\begin{aligned}
& \Big| \sum_{i=1}^{n} \int_{0}^{\infty} \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) < 0\right\} dN_i(t) \\
& - \sum_{i=1}^{n} \int_{0}^{\infty} \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) < 0\right\} dN_i(t) \Big| \\
\leq \; & \sum_{i=1}^{n} \int_{0}^{\infty} |\varepsilon_i V_{ij}(t)| \, I\left\{\mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) \leq Y_i(t) \leq \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\right\} dN_i(t) \\
\leq \; & \sqrt{n}\mathbb{G}_n \left(\int_{0}^{\infty} |\varepsilon_i V_{ij}(t)| \, I\left\{\mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) \leq Y_i(t) \leq \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_{k+1})\right\} dN_i(t) \right) \\
& + nE\Big[\int_{0}^{\infty} |\varepsilon_i V_{ij}(t)| \, I\left\{\mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) \leq Y_i(t) \leq \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_{k+1})\right\} dN_i(t) \Big] \\
\leq \; & \sqrt{n}\mathbb{G}_n \left(\int_{0}^{\infty} |\varepsilon_i V_{ij}(t)| \, I\left\{\mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) \leq Y_i(t) \leq \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_{k+1})\right\} dN_i(t) \right) \\
& + n M_0 C_V \sqrt{\log(r+p)/n}. \qquad \text{(S3.14)}
\end{aligned}$$

We note that $\mathrm{Var}\left(\int_0^\infty |\varepsilon_i V_{ij}(t)| \, I\left\{\mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) \leq Y_i(t) \leq \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_{k+1})\right\} dN_i(t) \right) \leq 4M_0^2 C_V^2 \sqrt{\log(r+p)/n}$, according to our choice of τ_k's. By Bernstein's inequality again,

$$\begin{aligned}
& P\Big(\sqrt{n}\mathbb{G}_n \left(\int_{0}^{\infty} |\varepsilon_i V_{ij}(t)| \, I\left\{\mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_k) \leq Y_i(t) \leq \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau_{k+1})\right\} dN_i(t) \right) \\
& \qquad > 4 M_0 C_V n^{1/4} \log^{3/4} \max\{n, r+p\} \Big) \\
\leq \; & 2 \exp\left(-\frac{1}{2} \frac{16 M_0^2 C_V^2 n^{1/2} \log^{3/2} \max\{n, r+p\}}{n M_0^2 C_V^2 \log^{1/2}(r+p) n^{-1/2} + 4 M_0^2 C_V^2 n^{1/4} \log^{3/4} \max\{n, r+p\}/3} \right) \\
= \; & 2 \exp\left(-\frac{8 \log \max\{n, r+p\}}{1 + 2n^{-1/4} \log^{1/4} \max\{n, r+p\}/3} \right) \\
\leq \; & 2 \exp\left(-4 \log \max\{n, r+p\} \right). \qquad \text{(S3.15)}
\end{aligned}$$

Therefore, it follows from (S3.13), (S3.14), (S3.15) that

$$P\Big(\sup_{\tau \in \Delta, \boldsymbol{\delta} \in R_{r+s-1}(B), j > r+s} \Big| \sum_{i=1}^{n} \int_{0}^{\infty} \varepsilon_i V_{ij}(t) I\left\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau) < 0\right\} dN_i(t) \Big|$$

$$> 9M_0C_Vn^{1/2}\log^{1/2}\max\{n,r+p\}\Big) \leq 4\exp\left(-3\log\max\{n,r+p\}\right). \quad \text{(S3.16)}$$

Let $\tilde{K} = (C_2+10)\sqrt{n\log\max\{n,r+p\}}$, combining (S3.10)–(S3.12) and (S3.16) and the condition that $n/((r+s)^3\log^2\max\{n,r+p\}) \to \infty$ together yields

$$\begin{aligned}
&P\Big(\max_{j>r+s}\sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)}\Big|\sum_{i=1}^n\int_0^\infty V_{ij}(t)\psi_\tau\{Y_i(t)-\mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta})\}dN_i(t)\Big| > \tilde{K}\Big)\\
\leq &pP\Big(\sup_{\tau\in\Delta,\boldsymbol{\delta}\in R_{r+s-1}(B)}\Big|\sum_{i=1}^n\int_0^\infty V_{ij}(t)\psi_\tau\{Y_i(t)-\mathbf{V}_i(t)^\top(\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta})\}dN_i(t)\Big| > \tilde{K}\Big)\\
\leq &p\frac{20}{(r+p)^2} \leq \frac{20}{r+p}
\end{aligned}$$

The condition $(\inf_{j>r+s-1,\tau\in\Delta} w_j(\tau))^{-1}\sqrt{n}/\sqrt{(r+s)\log\max\{n,r+p\}} = O_p(1)$, coupled with $\lambda/((r+s)\log\max\{n,r+p\}) \to \infty$ and (S3.8) implies that with probability approaching 1, $\lambda w_j(\tau) > \tilde{K} + \varrho$. This proves (S3.9) and hence completes the proof of Theorem A.3. □

S4 Proofs of Theorems 1–4 and Corollaries 1–2

In this section, we provide the proofs of Theorems 1–4 and Corollaries 1–2 based on the lemmas and theorems in Section S2–S3.

We first introduce some notation. Given a random sample $Z_1,\ldots,Z_n$, let $\mathbb{E}_n(f) = \mathbb{E}_n\{f(Z_i)\} := n^{-1}\sum_{i=1}^n f(Z_i)$ and $\mathbb{G}_n(f) = \mathbb{G}_n\{f(Z_i)\} := n^{1/2}\mathbb{E}_n[f(Z_i) - E\{f(Z_i)\}]$. Recall that $R_{r+s-1} := \{\boldsymbol{\delta} \in \mathbb{R}^{r+p} : \sum_{j=1}^s \delta_{r+j} = 0, \delta_{r+l} = 0, s < l \leq p\} = \{\boldsymbol{\delta} = (\boldsymbol{\delta}_{\mathbf{x}}^\top, \boldsymbol{\delta}_{\mathbf{z}}^\top)^\top : \boldsymbol{\delta}_{\mathbf{x}} \in \mathbb{R}^r, \boldsymbol{\delta}_{\mathbf{z}} \in \mathbb{R}^p, \sum_{j=1}^s \delta_{\mathbf{z}j} = 0, \delta_{\mathbf{z}l} = 0, s < l \leq p\}$. For any $B > 0$, define

$R_{r+s-1}(B) := \{\boldsymbol{\delta} \in R_{r+s-1} : \|\boldsymbol{\delta}\| \leq ((r+s-1)\log n/n)^{1/2}B\}$. For any $\boldsymbol{\delta} \in R_{r+s-1}$, let $\boldsymbol{\delta}_{\setminus l} := (\boldsymbol{\delta}_{\mathbf{x}}^\top, \boldsymbol{\delta}_{\mathbf{z},\setminus l}^\top)^\top$, $D_i(\boldsymbol{\delta}, \tau) := \int_0^\infty \rho_\tau\big[Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta}\}\big]dN_i(t) - \int_0^\infty \rho_\tau\big[Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\big]dN_i(t)$, and $M_n(\tau, \boldsymbol{\delta}) := n^{1/2}\mathbb{E}_n\big[\int_0^\infty \mathbf{V}_i(t)\psi_\tau[Y_i(t) - \mathbf{V}_i(t)^\top\{\boldsymbol{\gamma}_0(\tau)+\boldsymbol{\delta}\}]dN_i(t)\big]$, where $\psi_\tau(v) = \tau - I(v<0)$. We also define $\sigma(\tau) := n^{-1}\sum_{i=1}^n\int_0^\infty \rho_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top\boldsymbol{\gamma}_0(\tau)\}dN_i(t)$.

To prove Theorems 1–2 and Corollary 1, we first establish the asymptotic properties of the oracle regularized estimator, which is obtained by assuming that all relevant covariates are known in advance. Provided $S_\Delta = \{1, \ldots, s\}$, the oracle regularized estimator is given by $\hat{\boldsymbol{\gamma}}^o(\tau) = (\hat{\boldsymbol{\gamma}}_a^o(\tau)^\top, \mathbf{0}^\top)^\top$, where $\hat{\boldsymbol{\gamma}}_a^o(\tau) = (\hat{\boldsymbol{\alpha}}^o(\tau)^\top, \hat{\boldsymbol{\beta}}_a^o(\tau)^\top)^\top$ is the minimizer of

$$\sum_{i=1}^n\int_0^\infty \rho_\tau\{Y_i(t) - \mathbf{X}_i(t)^\top\boldsymbol{\alpha} - \mathbf{Z}_{ia}(t)^\top\boldsymbol{\beta}_a\}dN_i(t) + \lambda\sum_{j=1}^s w_j(\tau)|\beta_{aj}| \text{ s.t. } \sum_{j=1}^s \beta_{aj} = 0.$$

In Theorem A.1, we show that $\hat{\boldsymbol{\gamma}}^o(\tau)$ is uniformly consistent in $\tau \in \Delta$ with the convergence rate $\sqrt{(r+s)\log n/n}$. In Theorem A.2, we establish some weak convergence results for a linear combination of $\hat{\boldsymbol{\gamma}}^o(\tau)$. Specifically, we investigate the asymptotic behavior of $n^{1/2}\boldsymbol{\xi}^\top\{\hat{\boldsymbol{\gamma}}^o(\tau) - \boldsymbol{\gamma}_0(\tau)\}$ with the constraint that $\sum_{j=1}^s \hat{\beta}_j^o(\tau) = 0$. We show in Theorem A.3 that the oracle regularized estimator equals the proposed regularized estimator with probability tending to one.

Proof of Theorem 1. Theorem 1 follows immediately from Theorems A.1 and A.3.

□

Proof of Corollary 1. Corollary 1 follows immediately from Corollary A.1 and Theorem A.3. □

Proof of Theorem 2. The proof is completed by combining the results in Theorems A.2 and A.3. □

Proof of Theorem 3. The proof is completed by combining the results in Lemmas 10 and 11. □

Proof of Corollary 2. By definition,

$$\mathrm{GIC}(\lambda) = \int_{\Delta} \log \hat{\sigma}_{\lambda}(\tau) d\tau + |\hat{S}_{\lambda}|\phi_n \geq \int_{\tau\in\Delta} \log(\hat{\sigma}_{\hat{S}_{\lambda}}(\tau))d\tau + |\hat{S}_{\lambda}|\phi_n = \mathrm{GIC}(\hat{S}_{\lambda}).$$

By Theorem 3, for those λ's which lead to underfitted or overfitted model selection, we have

$$\inf_{S_{\lambda}\in UF\bigcup OF, |S_{\lambda}|\leq\kappa} \mathrm{GIC}(\lambda) \geq \inf_{S_{\lambda}\in UF\bigcup OF, |S_{\lambda}|\leq\kappa} \mathrm{GIC}(S_{\lambda}) > \mathrm{GIC}(S_{\Delta}), \tag{S4.1}$$

with probability approaching 1.

Consider an ancillary tuning parameter sequence $\tilde{\lambda}_n = \sqrt{(r+s)\log n}\log\max\{n, r+p\}$, it is easy to verify $\sup_{\tau\in\Delta, j\in S_{\tau}} \tilde{\lambda}_n\omega_j(\tau) = O_p(\sqrt{n\log n})$ and $\tilde{\lambda}_n/\sqrt{r+s}\log\max\{n, r+p\} \to \infty$, then by Theorem $A.3$, we have $Pr\left(\hat{\boldsymbol{\gamma}}_{\tilde{\lambda}_n}(\tau) - \boldsymbol{\gamma}_0(\tau) \in R_{r+s-1}(B)\right) \to 1$ and $Pr(S_{\tilde{\lambda}_n} = S_{\Delta}) \to 1$, where $\hat{\boldsymbol{\gamma}}_{\tilde{\lambda}_n}(\tau)$ is the minimizer of (2.3) with the tuning parameter $\tilde{\lambda}_n$.

Consequently, with probability approaching 1,

$$\begin{aligned}\mathrm{GIC}(\tilde{\lambda}_n) - GIC(S_{\tilde{\lambda}_n}) &= \mathrm{GIC}(\tilde{\lambda}_n) - \mathrm{GIC}(S_{\Delta}) = \int_{\Delta} \log(\hat{\sigma}_{\tilde{\lambda}_n}(\tau)) - \log(\hat{\sigma}_{S_{\Delta}}(\tau))d\tau \\ &= o\left(\inf_{S_{\lambda}\in UF\bigcup OF, |S_{\lambda}|\leq\kappa} \mathrm{GIC}(S_{\lambda}) - \mathrm{GIC}(S_{\Delta})\right), \end{aligned} \tag{S4.2}$$

where the last equation follows from the proofs of Lemmas 10 and 11 in the Supplementary Materials. By (S4.1), (S4.2), and the fact that $\hat{\lambda}_n$ is the minimizer of GIC(λ) together, with probability tending to 1,

$$\inf_{S_\lambda \in UF \bigcup OF, |S_\lambda| \le \kappa} \mathrm{GIC}(\lambda) > \mathrm{GIC}(\tilde{\lambda}_n) \ge \mathrm{GIC}(\hat{\lambda}_n).$$

Thus, $Pr(S_{\hat{\lambda}_n} = S_\Delta) \to 1$. This completes the proof of Corollary 2. □

Proof of Theorem ??: The proof of part (a) can be completed by modifying the arguments used in Theorems A.1, A.2 and A.3.

By Theorems A.2 and A.3, we obtain that $n^{1/2}\boldsymbol{\xi}^\top \mathbf{H}_\tau \{\hat{\boldsymbol{\gamma}}(\tau) - \boldsymbol{\gamma}_0(\tau)\} = \boldsymbol{\xi}^\top M_n(\tau, \mathbf{0}) + o_p(1)$. We decompose $\mathbf{Z}_i^l(t)$ into $(\mathbf{Z}_{ia}^l(t)^\top, \mathbf{Z}_{ib}^l(t)^\top)^\top$, where $\mathbf{Z}_{ia}^l(t)$ is $\mathbf{Z}_{ia}(t) - Z_{i,l}(t)\mathbf{1}_s$ with the lth component removed and $\mathbf{Z}_{ib}^l(t) = \mathbf{Z}_{ib}(t) - Z_{i,l}(t)\mathbf{1}_{p-s}$. Define $\mathbf{V}_i^l(t) = (\mathbf{V}_{ia}^l(t)^\top, \mathbf{V}_{ib}^l(t)^\top)^\top$, where $\mathbf{V}_{ia}^l(t) = (\mathbf{X}_i(t)^\top, \mathbf{Z}_{ia}^l(t)^\top)^\top$ and $\mathbf{V}_{ib}^l(t) = \mathbf{Z}_{ib}^l(t)$. By modifying the same arguments used in Theorems A.2 and A.3, we can show that

$$\begin{aligned} & n^{1/2}\mathbf{H}_\tau^l(\hat{\boldsymbol{\gamma}}_{\setminus l}(\tau) - \boldsymbol{\gamma}_{0\setminus l}(\tau)) \\ =& n^{-1/2}\sum_{i=1}^n \int_0^\infty \mathbf{V}_i^l(t)\psi_\tau\{Y_i(t) - \mathbf{V}_i(t)^\top \boldsymbol{\gamma}_0(\tau)\} dN_i(t) + o_p(1), \end{aligned}$$

where

$$\mathbf{H}_\tau^l = \begin{pmatrix} E[\int_0^\infty f_{t,\tau}\{0|\mathbf{V}_i(t)\}\mathbf{V}_{ia}^l(t)\mathbf{V}_{ia}^l(t)^\top dN_i(t)] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

Noting that for any given $\boldsymbol{\xi} \in R_{r+s-1}$ and $\|\boldsymbol{\xi}\| = 1$,

$$n^{1/2}\boldsymbol{\xi}^\top \mathbf{H}_\tau \{\hat{\boldsymbol{\gamma}}_l^u(\tau) - \boldsymbol{\gamma}_0(\tau)\} = n^{1/2}\boldsymbol{\xi}_{\setminus l}^\top \mathbf{H}_\tau^l(\hat{\boldsymbol{\gamma}}_{\setminus l}(\tau) - \boldsymbol{\gamma}_{0\setminus l}(\tau))$$

$$= n^{-1/2}\boldsymbol{\xi}_{\backslash l}^{\top}\sum_{i=1}^{n}\int_{0}^{\infty}\mathbf{V}_i^l(t)\psi_\tau\{Y_i(t)-\mathbf{V}_i(t)^{\top}\boldsymbol{\gamma}_0(\tau)\}dN_i(t)+o_p(1)=\boldsymbol{\xi}^{\top}M_n(\tau,\mathbf{0})+o_p(1)$$

$$= n^{1/2}\boldsymbol{\xi}^{\top}\mathbf{H}_\tau\left\{\hat{\boldsymbol{\gamma}}(\tau)-\boldsymbol{\gamma}_0(\tau)\right\}+o_p(1),$$

where the third equality follows from the fact that for any $\boldsymbol{\xi}\in R_{r+s-1}$ and $1\le l\le s$, $\boldsymbol{\xi}^{\top}\mathbf{V}_i(t)=\boldsymbol{\xi}_{\backslash l}^{\top}\mathbf{V}_i^l(t)$. This completes the proof of Theorem 4. □

S5 Additional Simulation Results

Tables S1–S3 present the simulation results for setup (I), (II), (III) described in Section 4.

We consider additional setups which are the same as setup (II) and setup (VI) except for different variance specifications for a_i and $\epsilon_i(t)$, and are denoted by setup (II') and setup (VI') respectively.

Setup (II'): Data are generated from a longitudinal linear model with independent homogeneous errors,

$$Y_i(t)=-X_{i1}+X_{i2}-t+\mathbf{Z}_i(t)^{\top}\mathbf{b}+a_i+\epsilon_i(t),$$

where $\mathbf{b}=(1,0.8,0.9,1,2,-1.5,-4.2,0,\ldots,0)^{\top}$, $a_i\sim N(0,1/4)$, $\epsilon_i(t)\sim N(0,1/4)$ or $N(0,1)$ for $t>0$, $\epsilon_i(t)$ and $\epsilon_i(t')$ are independent for $t>0$, $t'>0$ and $t\ne t'$, and a_i and $\epsilon_i(t)$ are independent for $t>0$.

Setup (IV'): Data are generated from a longitudinal linear model with dependent heterogeneous errors,

$$Y_i(t) = -X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \mathbf{b}_1 + (X_{i1} + \mathbf{Z}_i(t)^\top \mathbf{b}_2)(a_i + \epsilon_i(t)),$$

where $\mathbf{b}_1 = \mathbf{b} = (1, 0.8, 0.9, 1, 2, -1.5, -4.2, 0, \dots, 0)^\top$, $\mathbf{b}_2 = (0, 0.1, 0, 0.1, 0, 0, -0.2, 0, \dots, 0)^\top$, $a_i \sim N(0, 1/4)$, $\epsilon_i(t) \sim N(0, 1/4)$ or $N(0, 1)$ for $t > 0$, $\epsilon_i(t)$ and $\epsilon_i(t')$ are independent for $t > 0$, $t' > 0$ and $t \neq t'$, and a_i and $\epsilon_i(t)$ are independent for $t > 0$.

The simulation results for setups (II') and (IV') are summarized in Tables S4 and S5 respectively. We have generally consistent findings regarding the proposed method and the naive methods. That is, the proposed method still outperforms the SS(τ) and PS approaches. As expected, we observe more accurate variable selection (i.e. higher PCF) when $\epsilon_i(t)$ has a smaller variance.

S6 Additional Table from the Real Data Example

Table S6 presents basic summary statistics for gender, number of longitudinal records, and calprotectin levels in the FIRST dataset

References

Belloni, A. and V. Chernozhukov (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*, 82–130.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *The Annals of statistics 28*, 1356–1378.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete*, Volume 23. Springer, Berlin.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.

Zhang, C.-H. and J. Huang (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *The Annals of Statistics 36*(4), 1567–1594.

Zheng, Q., L. Peng, and X. He (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics 43*, 2225–2258.

Table S1: Simulation results under Setup (I) with independent homogeneous errors

	AEE_{ℓ_1}	AEE_{ℓ_2}	AEE_{ℓ_∞}	NCN	NIN	PUF (%)	PCF (%)	POF (%)	SUM
				Proposed					
AW_2	2.535	1.144	0.773	6.920	0.043	7.3	88.7	4.0	0.000
AW_3	2.611	1.178	0.797	6.830	0.017	14.7	84.0	1.3	0.000
SS(0.2)	2.723	1.235	0.838	6.190	0.037	47.3	49.3	3.3	0.000
SS(0.5)	2.015	0.918	0.624	6.600	0.007	31.0	68.3	0.7	0.000
SS(0.8)	2.851	1.289	0.870	6.077	0.020	53.3	45.0	1.7	0.000
PS	2.538	1.147	0.773	6.947	0.357	4.3	64.0	31.7	0.000
				ALasso (i)					
AW_2	2.666	1.158	0.777	6.923	1.020	6.7	1.3	92.0	0.000
AW_3	2.742	1.190	0.798	6.817	0.993	15.7	1.3	83.0	0.000
				ALasso (ii)					
AW_2	2.608	1.189	0.817	6.877	0.027	12.0	85.7	2.3	−2.144
AW_3	2.658	1.208	0.828	6.773	0.017	20.0	78.3	1.7	−2.491
				ALasso (iii)					
AW_2	2.879	1.294	0.881	6.877	0.027	12.0	85.7	2.3	0.000
AW_3	3.051	1.371	0.939	6.770	0.017	20.0	78.3	1.7	0.000
				ALasso (iv)					
AW_2	2.509	1.125	0.752	6.960	0.300	3.3	69.7	27.0	0.000
AW_3	2.544	1.141	0.765	6.933	0.313	5.3	65.7	29.0	0.000

Table S2: Simulation results under Setup (II) with dependent homogeneous errors

	AEE_{ℓ_1}	AEE_{ℓ_2}	AEE_{ℓ_∞}	NCN	NIN	PUF (%)	PCF (%)	POF (%)	SUM
				Proposed					
AW_2	2.500	1.126	0.754	6.930	0.037	6.0	90.7	3.3	0.000
AW_3	2.535	1.141	0.765	6.860	0.017	12.7	85.7	1.7	0.000
SS(0.2)	2.802	1.262	0.847	6.137	0.030	51.7	46.0	2.3	0.000
SS(0.5)	2.007	0.910	0.610	6.643	0.010	29.0	70.0	1.0	0.000
SS(0.8)	2.797	1.260	0.848	6.150	0.037	50.7	47.0	2.3	0.000
PS	2.489	1.123	0.753	6.950	0.337	4.7	69.0	26.3	0.000
				ALasso (i)					
AW_2	2.644	1.143	0.756	6.910	1.020	8.3	0.3	91.3	0.000
AW_3	2.694	1.165	0.773	6.837	1.003	14.3	0.0	85.7	0.000
				ALasso (ii)					
AW_2	2.501	1.136	0.771	6.900	0.033	9.3	87.7	3.0	−2.304
AW_3	2.553	1.160	0.790	6.817	0.010	15.7	83.3	1.0	−2.418
				ALasso (iii)					
AW_2	2.729	1.223	0.821	6.897	0.033	9.3	87.7	3.0	0.000
AW_3	2.902	1.302	0.885	6.813	0.010	15.7	83.3	1.0	0.000
				ALasso (iv)					
AW_2	2.412	1.077	0.715	6.973	0.373	2.0	66.0	32.0	0.000
AW_3	2.468	1.104	0.735	6.950	0.363	4.0	63.0	33.0	0.000

Table S3: Simulation results under Setup (III) with independent heterogeneous errors

	AEE_{ℓ_1}	AEE_{ℓ_2}	AEE_{ℓ_∞}	NCN	NIN	PUF (%)	PCF (%)	POF (%)	SUM
				Proposed					
AW_2	2.314	1.047	0.704	6.910	0.047	8.3	87.3	4.3	0.000
AW_3	2.373	1.074	0.723	6.840	0.013	14.3	84.7	1.0	0.000
SS(0.2)	2.992	1.341	0.889	5.973	0.023	61.0	36.7	2.3	0.000
SS(0.5)	1.905	0.861	0.567	6.687	0.013	24.0	74.7	1.3	0.000
SS(0.8)	2.751	1.254	0.855	6.180	0.023	43.7	54.0	2.3	0.000
PS	2.349	1.060	0.713	6.960	0.273	4.0	73.0	23.0	0.000
				ALasso (i)					
AW_2	2.489	1.080	0.709	6.910	1.023	8.3	2.0	89.7	0.000
AW_3	2.538	1.102	0.725	6.847	0.990	13.3	2.0	84.7	0.000
				ALasso (ii)					
AW_2	2.401	1.095	0.741	6.890	0.040	10.3	85.7	4.0	−2.678
AW_3	2.463	1.121	0.758	6.827	0.013	15.7	83.3	1.0	−2.693
				ALasso (iii)					
AW_2	2.652	1.194	0.806	6.890	0.040	10.3	85.7	4.0	0.000
AW_3	2.802	1.259	0.854	6.827	0.013	15.7	83.3	1.0	0.000
				ALasso (iv)					
AW_2	2.316	1.038	0.686	6.960	0.380	3.3	62.0	34.7	0.000
AW_3	2.409	1.080	0.718	6.940	0.283	4.7	69.0	26.3	0.000

Table S4: Simulation results for under Setup (II') with dependent homogeneous errors

	AEE_{ℓ_1}	AEE_{ℓ_2}	AEE_{ℓ_∞}	NCN	NIN	PUF (%)	PCF (%)	POF (%)	SUM
				$a_i \sim N(0, 1/4), \epsilon_i(t) \sim N(0, 1/4)$					
AW_2	1.133	0.522	0.356	7.000	0.007	0.0	99.3	0.7	0.000
AW_3	1.142	0.526	0.359	7.000	0.007	0.0	99.3	0.7	0.000
SS(0.2)	1.130	0.516	0.352	6.933	0.003	6.0	93.7	0.3	0.000
SS(0.5)	0.919	0.415	0.276	6.983	0.007	1.7	97.7	0.7	0.000
SS(0.8)	1.124	0.518	0.350	6.947	0.007	5.3	94.0	0.7	0.000
PS	1.048	0.479	0.323	7.000	0.103	0.0	90.3	9.7	0.000
				$a_i \sim N(0, 1/4), \epsilon_i(t) \sim N(0, 1)$					
AW_2	2.376	1.100	0.759	6.930	0.217	7.0	75.7	17.3	0.000
AW_3	2.435	1.131	0.784	6.903	0.163	9.3	77.0	13.7	0.000
SS(0.2)	2.898	1.340	0.922	6.050	0.030	66.0	33.3	0.7	0.000
SS(0.5)	2.162	0.997	0.682	6.560	0.030	37.3	61.0	1.7	0.000
SS(0.8)	2.872	1.326	0.910	6.060	0.037	63.3	34.0	2.7	0.000
PS	2.590	1.198	0.823	6.943	0.367	5.7	68.3	26.0	0.000

Table S5: Simulation results under Setup (IV') with dependent heterogeneous errors

	AEE_{ℓ_1}	AEE_{ℓ_2}	AEE_{ℓ_∞}	NCN	NIN	PUF (%)	PCF (%)	POF (%)	SUM
				$a_i \sim N(0, 1/4), \epsilon_i(t) \sim N(0, 1/4)$					
AW_2	0.804	0.364	0.241	7.000	0.003	0.0	99.7	0.3	0.000
AW_3	0.806	0.365	0.242	7.000	0.003	0.0	99.7	0.3	0.000
SS(0.2)	0.878	0.400	0.268	6.990	0.003	1.0	98.7	0.3	0.000
SS(0.5)	0.697	0.315	0.208	6.997	0.000	0.3	99.7	0.0	0.000
SS(0.8)	0.837	0.379	0.250	7.000	0.000	0.0	100.0	0.0	0.000
PS	0.697	0.316	0.210	7.000	0.057	0.0	95.3	4.7	0.000
				$a_i \sim N(0, 1/4), \epsilon_i(t) \sim N(0, 1)$					
AW_2	2.874	1.296	0.875	6.853	0.123	14.0	77.3	8.7	0.000
AW_3	2.976	1.341	0.906	6.797	0.087	19.3	75.7	5.0	0.000
SS(0.2)	3.694	1.645	1.091	5.457	0.010	79.7	20.3	0.0	0.000
SS(0.5)	2.572	1.165	0.784	6.317	0.007	49.7	49.7	0.7	0.000
SS(0.8)	3.440	1.561	1.059	5.707	0.010	68.0	31.7	0.3	0.000
PS	3.104	1.396	0.941	6.883	0.210	10.3	72.3	17.3	0.000

Table S6: Summary statistics for microbiota composition data

Gender		Boy		Girl
$n(\%)$		76 (56.3%)		59 (43.7%)
m	1	2	3	4
$n(\%)$	26 (19.26%)	42 (31.11%)	50 (37.04%)	17 (12.59%)
	Minimum	Median	Mean	Maximum
$Y(t)$	16	64.5	111.2	939