LOCALLY SPARSE ESTIMATOR OF GENERALIZED VARYING COEFFICIENT MODEL FOR ASYNCHRONOUS LONGITUDINAL DATA

Rou Zhong[1,2], Chunming Zhang[3], Jingxiao Zhang[1,2,*]

[1] *Center for Applied Statistics, Renmin University of China*

[2] *School of Statistics, Renmin University of China*

[3] *Department of Statistics, University of Wisconsin-Madison*

Supplementary Material

In the Supplementary Material, we first provide the proofs of Theorems 1–3. Then, the pointwise asymptotic distributions of $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$ are studied. Finally, some additional simulation results are presented.

S1 Proof of Theorem 1

Proof of Theorem 1. The estimating equations are equivalent to

$$U_n(\boldsymbol{\gamma}) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i} K_h(T_{ij}-S_{ik})\widetilde{\mathbf{X}}_i^{\star}(S_{ik})\Big[Y_i(T_{ij}) - g\Big\{\widetilde{\mathbf{X}}_i^{\star}(S_{ik})^{\top}\boldsymbol{\gamma}\Big\}\Big] - \bar{N}\widetilde{P}_1(\boldsymbol{\gamma}) - \bar{N}\widetilde{P}_2(\boldsymbol{\gamma}) = \mathbf{0},$$

where $\bar{N} = n^{-1}\sum_{i=1}^{n} L_i M_i$, $\widetilde{P}_1(\boldsymbol{\gamma}) = \mathbf{V}_{\rho_0,\rho_1}\boldsymbol{\gamma}$ and $\widetilde{P}_2(\boldsymbol{\gamma}) = \frac{\partial \mathrm{PEN}_{\lambda(\gamma)}}{\partial \boldsymbol{\gamma}}$. By using counting process $N_i(t,s)$, we can rewrite the estimating equations as

$$\psi_n(\boldsymbol{\gamma}) = n^{-1}\sum_{i=1}^{n}\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\widetilde{\mathbf{X}}_i^\star(s)^\top\boldsymbol{\gamma}\}]dN_i(t,s) - \bar{N}\widetilde{P}_1(\boldsymbol{\gamma}) - \bar{N}\widetilde{P}_2(\boldsymbol{\gamma}) = \mathbf{0},$$

where $N_i(t,s) = \sum_{j=1}^{L_i}\sum_{k=1}^{M_i} I(T_{ij} < t, S_{ik} < s)$ and $I(\cdot)$ is the indicator function. Let $\alpha_n = M^{1/2}h^2 + n^{-1/2}M^{1/2}h^{-1/2} + \rho M^{-1/2} + M^{-r}$. We then want to show that $\forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma} : \boldsymbol{\gamma}_0 + \alpha_n w, \|w\|_2 = C_1\}, \forall \epsilon > 0$, we have

$$P\{\inf_{\|w\|_2 = C_1} \psi_n(\boldsymbol{\gamma})^\top\psi_n(\boldsymbol{\gamma}) > \psi_n(\boldsymbol{\gamma}_0)^\top\psi_n(\boldsymbol{\gamma}_0)\} \geq 1 - \epsilon, \tag{S1.1}$$

when constant C_1 is large enough. It implies that there exists a local minimizer $\widehat{\boldsymbol{\gamma}}$ in the ball $\{\boldsymbol{\gamma} : \boldsymbol{\gamma}_0 + \alpha_n w, \|w\|_2 \leq C_1\}$, with probability at least $1-\epsilon$. That means $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|_2 = O_p(\alpha_n)$.

Let

$$U_{ni}(\boldsymbol{\gamma}) = \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\widetilde{\mathbf{X}}_i^\star(s)^\top\boldsymbol{\gamma}\}]dN_i(t,s) - \bar{N}\widetilde{P}_1(\boldsymbol{\gamma}) - \bar{N}\widetilde{P}_2(\boldsymbol{\gamma}).$$

Then $\psi_n(\boldsymbol{\gamma}) = n^{-1}\sum_{i=1}^{n} U_{ni}(\boldsymbol{\gamma})$. For $U_{ni}(\boldsymbol{\gamma})$, we have

$$\begin{aligned} U_{ni}(\boldsymbol{\gamma}) = U_{ni}(\boldsymbol{\gamma}_0) - \\ &\Big[\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)g'\{\widetilde{\mathbf{X}}_i^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}_i^\star(s)^\top dN_i(t,s) + \bar{N}\frac{\partial\widetilde{P}_1(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}_0} + \bar{N}\cdot\frac{\partial\widetilde{P}_2(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}_0}\Big]\alpha_n w\{1+o(1)\} \\ &\triangleq U_{ni}(\boldsymbol{\gamma}_0) - U_{ni}^{(1)}(w). \end{aligned}$$

Therefore,

$$\psi_n(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\{U_{ni}(\boldsymbol{\gamma}_0) - U_{ni}^{(1)}(w)\} = \psi_n(\boldsymbol{\gamma}_0) - U_n^{(1)}(w),$$

where $U_n^{(1)}(w) = \frac{1}{n}\sum_{i=1}^{n} U_{ni}^{(1)}(w)$. Then we have

$$\psi_n(\boldsymbol{\gamma})^\top \psi_n(\boldsymbol{\gamma}) - \psi_n(\boldsymbol{\gamma}_0)^\top \psi_n(\boldsymbol{\gamma}_0) = U_n^{(1)}(w)^\top U_n^{(1)}(w) - 2\psi_n(\boldsymbol{\gamma}_0)^\top U_n^{(1)}(w) \triangleq S_1 - S_2.$$

Let

$$A_1 = \frac{1}{n}\sum_{i=1}^{n}\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s) g'\{\widetilde{\mathbf{X}}_i^\star(s)^\top \boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}_i^\star(s)^\top dN_i(t,s) + \bar{N}\frac{\partial \widetilde{P}_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}_0} + \bar{N}\cdot\frac{\partial \widetilde{P}_2(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}_0},$$

We have

$$S_1 = \|A_1\alpha_n w\|^2 \geq O(M)\lambda_{\min}(A_1^\top A_1)\alpha_n^2\|w\|_2^2 = O(M)\lambda_{\min}(A_1)^2\alpha_n^2\|w\|_2^2,$$

$$|S_2| \leq 2\|\psi_n(\boldsymbol{\gamma}_0)\|_2 S_1^{1/2} \leq 2\|\psi_n(\boldsymbol{\gamma}_0)\|_2 O(M^{1/2})\lambda_{\max}(A_1)\alpha_n\|w\|_2.$$

By Lemma 1 and Lemma 2, there exists constants $C_2 > 0, C_3 > 0$, such that

$$S_1 \geq C_2 M\alpha_n^2\|w\|_2^2$$

$$|S_2| \leq C_3 M^{1/2}\alpha_n^2\|w\|_2.$$

Then

$$S_1 - S_2 \geq C_2 M\alpha_n^2\|w\|_2^2 - C_3 M^{1/2}\alpha_n^2\|w\|_2.$$

Thus, when C_1 is large enough, we have $S_1 - S_2 > 0$. Then (S1.1) is obtained. So $\|\widehat{\boldsymbol{\gamma}}^{(0)} - \boldsymbol{\gamma}_0^{(0)}\|_2 = O_p(\alpha_n)$ and $\|\widehat{\boldsymbol{\gamma}}^{(1)} - \boldsymbol{\gamma}_0^{(1)}\|_2 = O_p(\alpha_n)$.

Since $\|\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B}-\beta_0\|_\infty = O(M^{-r})$ by Assumption 1 (Zhong et al., 2021), we have

$$
\begin{aligned}
\|\widehat{\beta}_0-\beta_0\|_\infty &\le \|\widehat{\beta}_0-\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B}\|_\infty + \|\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B}-\beta_0\|_\infty \\
&= \|(\widehat{\boldsymbol{\gamma}}^{(0)}-\boldsymbol{\gamma}_0^{(0)})^\top\mathbf{B}\|_\infty + \|\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B}-\beta_0\|_\infty \\
&\le \|\widehat{\boldsymbol{\gamma}}^{(0)}-\boldsymbol{\gamma}_0^{(0)}\|_\infty\big(\sum_{j=1}^{L} B_j\big) + \|\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B}-\beta_0\|_\infty \\
&= \|\widehat{\boldsymbol{\gamma}}^{(0)}-\boldsymbol{\gamma}_0^{(0)}\|_\infty + \|\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B}-\beta_0\|_\infty \\
&= O_p(\alpha_n) + O_p(M^{-r}) \\
&= O_p(\alpha_n).
\end{aligned}
$$

We can get $\|\widehat{\beta}_1-\beta_1\|_\infty = O_p(\alpha_n)$ in the same way. The proof is completed.

□

Lemma 1. *Suppose that the conditions of Theorem 1 are satisfied, there exists constants* $c_1>0$ *and* $c_2>0$, *such that* $c_1 \le \lambda_{\min}(A_1) \le \lambda_{\max}(A_1) \le c_2$.

Proof. Let

$$
B_1 = \frac{1}{n}\sum_{i=1}^{n}\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)g'\{\widetilde{\mathbf{X}}_i^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}_i^\star(s)^\top dN_i(t,s).
$$

Then

$$\begin{aligned} EB_1 &= \int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}^\star(s)^\top]\lambda(t,s)dtds \\ &= \int\int K(z)E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}^\star(s)^\top]\lambda(s+hz,s)dzds \\ &= \{1+O(h^2)\}\int E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}^\star(s)^\top]\lambda(s,s)ds. \end{aligned}$$

First, EB_1 is positive definite. In specific, if there exists a vector $\mathbf{a}$, such that

$$\mathbf{a}^\top\int E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}^\star(s)^\top]\lambda(s,s)ds\mathbf{a}=0.$$

Then $\mathbf{a}^\top\widetilde{\mathbf{X}}^\star(s)=0$ for any $s\in\mathcal{G}$ with probability 1, which means $\mathbf{a}_1^\top\mathbf{B}(s)+\mathbf{a}_2^\top\mathbf{B}(s)X(s)=0$, where $\mathbf{a}_1,\mathbf{a}_2$ are the first and second L elements of $\mathbf{a}$. By Assumption 6, we have $\mathbf{a}=\mathbf{0}$. That means all eigenvalues of EB_1 are positive.

Then, eigenvalues of EB_1 are finite. Specifically, $\forall\mathbf{b}\in\mathbb{R}^{2L}$ satisfying

$\|\mathbf{b}\|_2 = 1$, we have

$$\begin{aligned}
&\mathbf{b}^\top \int E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}\widetilde{\mathbf{X}}^\star(s)^\top]\lambda(s,s)ds\mathbf{b} \\
&= \int E[\mathbf{b}^\top \widetilde{\mathbf{X}}^\star(s)\widetilde{\mathbf{X}}^\star(s)^\top \mathbf{b} g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}]\lambda(s,s)ds \\
&\leq \int E[\widetilde{\mathbf{X}}^\star(s)^\top \widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}]\lambda(s,s)ds \\
&= \int_{\mathcal{T}} \mathbf{B}(s)^\top \mathbf{B}(s) E\big[g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\} + X^2(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}\big]\lambda(s,s)ds \\
&\leq \int_{\mathcal{T}} \mathbf{B}(s)^\top \mathbf{B}(s) E^{1/2}\big[g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}^2]\lambda(s,s)ds \\
&\quad + \int_{\mathcal{T}} \mathbf{B}(s)^\top \mathbf{B}(s) E^{1/2}\big[X^4(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}^2]\lambda(s,s)ds < \infty
\end{aligned}$$

The first inequality is derived by

$$\mathbf{b}^\top \widetilde{\mathbf{X}}^\star(s)\widetilde{\mathbf{X}}^\star(s)^\top \mathbf{b} = \Big\{\sum_{j=1}^{2L} b_j \widetilde{X}_j^\star(s)\Big\}^2 \leq \sum_{j=1}^{2L} b_j^2 \sum_{j=1}^{2L} \widetilde{X}_j^{\star 2}(s) = \widetilde{\mathbf{X}}^\star(s)^\top \widetilde{\mathbf{X}}^\star(s),$$

where $\widetilde{X}_j^\star(s)$ is the j-th element of $\widetilde{\mathbf{X}}^\star(s)$. The last inequality can be obtained by Assumption 2 and Assumption 5. Hence, eigenvalues of EB_1 are finite.

We have $\|A_1 - EB_1\|_1 \leq \|A_1 - B_1\|_1 + \|B_1 - EB_1\|_1$, where $\|\cdot\|_1$ is the

L_1 norm for matrix. For $\|A_1 - B_1\|_1$,

$$
\begin{aligned}
\|A_1 - B_1\|_1 =& \Big\|\bar{N}\frac{\partial \widetilde{P}_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}_0} + \bar{N}\cdot\frac{\partial \widetilde{P}_2(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}_0}\Big\|_1 \\
=&\Big\|\bar{N}\mathbf{V}_{\rho_0,\rho_1} + \frac{\bar{N}}{2}\cdot\frac{M+1}{T}\frac{\partial^2}{\partial \boldsymbol{\gamma}_0^2}\int p_\lambda(|\mathbf{B}^\top(t)\boldsymbol{\gamma}_0^{(1)}|)dt\Big\|_1 \\
\leq&\bar{N}\|\mathbf{V}_{\rho_0,\rho_1}\|_1 + \frac{\bar{N}}{2}\cdot\frac{M+1}{T}\Big\|\frac{\partial^2}{\partial \boldsymbol{\gamma}_0^2}\int p_\lambda(|\mathbf{B}^\top(t)\boldsymbol{\gamma}_0^{(1)}|)dt\Big\|_1 \\
=&\bar{N}\|\mathbf{V}_{\rho_0,\rho_1}\|_1 + \frac{\bar{N}}{2}\cdot\frac{M+1}{T}\Big\|\frac{\partial^2}{\partial \boldsymbol{\gamma}_0^{(1)2}}\int p_\lambda(|\mathbf{B}^\top(t)\boldsymbol{\gamma}_0^{(1)}|)dt\Big\|_1 \\
=&\bar{N}o(1) + \frac{\bar{N}}{2}\cdot\frac{M+1}{T}o(M^{-1}) = o_p(1). \qquad \text{(S1.2)}
\end{aligned}
$$

The third equality is derived by Assumption 3 according to Lin et al. (2017). Moreover,

$$
\bar{N} = \frac{1}{n}\sum_{i=1}^n L_i M_i = n^{-1}\sum_{i=1}^n\sum_{j=1}^{L_i}\sum_{k=1}^{M_i} 1 = \frac{1}{n}\sum_{i=1}^n\int\int dN_i(t,s),
$$

$$
E\Big|\int\int dN_i(t,s)\Big| = E\Big\{\int\int dN_i(t,s)\Big\} = \int\int \lambda(t,s)dtds < \infty.
$$

Then by Markov inequality, we have $\bar{N} = O_p(1)$, which is used in the derivation of the last equality of (S1.2). For $\|B_1 - EB_1\|_1$, let

$$
\eta_{j_1j_2} = \frac{1}{n}\sum_{i=1}^n\int\int K_h(t-s)\widetilde{X}^\star_{ij_1}(s)g'\{\widetilde{\mathbf{X}}^\star_i(s)^\top\boldsymbol{\gamma}_0\}\widetilde{X}^\star_{ij_2}(s)dN_i(t,s).
$$

Then $\|B_1 - EB_1\|_1 = \sum_{j_1=1}^{2L}\sum_{j_2=1}^{2L}|\eta_{j_1j_2} - E\eta_{j_1j_2}|$. Similar to the proof of

Theorem 1 in Cao et al. (2015), we have

$$\begin{aligned}
\text{var}(\eta_{j_1j_2}) &= \frac{1}{n}\text{var}\Big[\int\int K_h(t-s)\widetilde{X}^\star_{j_1}(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{X}^\star_{j_2}(s)dN(t,s)\Big] \\
&\le \frac{1}{n}E\Big[\int\int K_h(t-s)\widetilde{X}^\star_{j_1}(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\widetilde{X}^\star_{j_2}(s)dN(t,s)\Big]^2 \\
&= \frac{1}{nh}\int\int K^2(z)E[\widetilde{X}^{\star 2}_{j_1}(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}^2\widetilde{X}^{\star 2}_{j_2}(s)]\lambda(s+hz,s)dzds + O(n^{-1}M^{-1}) \\
&= \frac{1}{nh}\int K^2(z)dz\int E[\widetilde{X}^{\star 2}_{j_1}(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}^2\widetilde{X}^{\star 2}_{j_2}(s)]\lambda(s,s)ds + O(n^{-1}M^{-1}) \\
&= O(M^{-1}n^{-1}h^{-1}).
\end{aligned}$$

The above derivation is obtained by Assumption 5 and $\int \widetilde{B}^2_{j_1}(s)\widetilde{B}^2_{j_2}(s)\lambda(s,s)ds = O(M^{-1})$, where $\widetilde{B}_j(s)$ is the j-th element of $\widetilde{\mathbf{B}}(s) = (\mathbf{B}(s)^\top, \mathbf{B}(s)^\top)^\top$. Then

$$\|B_1 - EB_1\|_1 = O_p(M^{3/2}n^{-1/2}h^{-1/2}) = o_p(1). \tag{S1.3}$$

Thus, by (S1.2) and (S1.3), we have $\|A_1 - EB_1\|_1 = o_p(1)$. Since

$$|\lambda_{\min}(A_1) - \lambda_{\min}(EB_1)| \le \|A_1 - EB_1\|_1,$$

$$|\lambda_{\max}(A_1) - \lambda_{\max}(EB_1)| \le \|A_1 - EB_1\|_1,$$

eigenvalues of A_1 are bounded away from 0 and infinity as eigenvalues of EB_1. The proof is completed.

□

Lemma 2. *Suppose that the conditions of Theorem 1 are satisfied,* $\|\psi_n(\boldsymbol{\gamma}_0)\|_2 = O_p(\alpha_n)$.

Proof. Let

$$Q_n(\boldsymbol{\gamma}_0) = n^{-1}\sum_{i=1}^{n}\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^{\star}(s)[Y_i(t) - g\{\widetilde{\mathbf{X}}_i^{\star}(s)^{\top}\boldsymbol{\gamma}_0\}]dN_i(t,s).$$

Then

$$\|\psi_n(\boldsymbol{\gamma}_0)\|_2 \leq \|Q_n(\boldsymbol{\gamma}_0)\|_2 + \bar{N}\|\widetilde{P}_1(\boldsymbol{\gamma}_0)\|_2 + \bar{N}\|\widetilde{P}_2(\boldsymbol{\gamma}_0)\|_2. \tag{S1.4}$$

First, for $Q_n(\boldsymbol{\gamma}_0)$, we have

$$\begin{aligned} E\|Q_n(\boldsymbol{\gamma}_0)\|_2^2 &= E\{Q_n(\boldsymbol{\gamma}_0)^{\top}Q_n(\boldsymbol{\gamma}_0)\} = \mathrm{tr}[\mathrm{var}\{Q_n(\boldsymbol{\gamma}_0)\}] + E\{Q_n(\boldsymbol{\gamma}_0)\}^{\top}E\{Q_n(\boldsymbol{\gamma}_0)\} \\ &= \frac{1}{nh}\mathrm{tr}[\mathrm{var}\{h^{1/2}U_{n1}(\boldsymbol{\gamma}_0)\}] + E\{Q_n(\boldsymbol{\gamma}_0)\}^{\top}E\{Q_n(\boldsymbol{\gamma}_0)\}. \end{aligned} \tag{S1.5}$$

For $E\{Q_n(\boldsymbol{\gamma}_0)\}$, we have

$$\begin{aligned}
E\{Q_n(\boldsymbol{\gamma}_0)\} =& E\Big(\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)[Y(t)-g\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]dN(t,s)\Big)\\
=& E\Big[E\Big\{\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)Y(t)dN(t,s)\Big|\widetilde{\mathbf{X}}^\star\Big\}\Big]\\
&-\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]\lambda(t,s)dtds\\
=&\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\beta_0(t)+X(t)\beta_1(t)\}]\lambda(t,s)dtds\\
&-\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]\lambda(t,s)dtds\\
=&\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(t)^\top\boldsymbol{\gamma}_0+R_n^{(0)}(t)+X(t)R_n^{(1)}(t)\}]\lambda(t,s)dtds\\
&-\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]\lambda(t,s)dtds\\
=&\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(t)^\top\boldsymbol{\gamma}_0\}]\lambda(t,s)dtds\\
&-\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]\lambda(t,s)dtds\\
&+\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(t)^\top\boldsymbol{\gamma}_0\}\{R_n^{(0)}(t)+X(t)R_n^{(1)}(t)\}]\lambda(t,s)dtds\\
&+\int\int K_h(t-s)E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(t)^\top\boldsymbol{\gamma}_0\}o_p\{R_n^{(0)}(t)+X(t)R_n^{(1)}(t)\}]\lambda(t,s)dtds\\
&\triangleq I_1+I_2+I_3, \qquad \text{(S1.6)}
\end{aligned}$$

where $R_n^{(0)}(t)=\beta_0(t)-\mathbf{B}^\top(t)\boldsymbol{\gamma}_0^{(0)}$ and $R_n^{(1)}(t)=\beta_1(t)-\mathbf{B}^\top(t)\boldsymbol{\gamma}_0^{(1)}$.

Let $F_{\boldsymbol{\gamma}_0}(s,hz)=E[\widetilde{\mathbf{X}}^\star(s)g\{\widetilde{\mathbf{X}}^\star(s+hz)^\top\boldsymbol{\gamma}_0\}]$. Then by Taylor expansion,

$$\begin{aligned}
I_1 &= \int\int K(z)\{F_{\boldsymbol{\gamma}_0}(s,hz)-F_{\boldsymbol{\gamma}_0}(s,0)\}\lambda(s+hz,s)dzds\\
&= \mathbf{C}h^2+o(h^2),
\end{aligned}$$

where

$$\mathbf{C} = \int z^2 K(z)dz \int \Big\{ \frac{\partial F_{\gamma_0}(s,y)}{\partial y}\Big|_{y=0} \cdot \frac{\partial \lambda(x,s)}{\partial x}\Big|_{x=s} + \frac{1}{2}\frac{\partial^2 F_{\gamma_0}(s,y)}{\partial y^2}\Big|_{y=0} \cdot \lambda(s,s) \Big\} ds.$$

So we have

$$I_1^\top I_1 = O(Mh^4). \tag{S1.7}$$

Let

$$\widetilde{I}_2 = \int\int K_h(t-s) E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(t)^\top \boldsymbol{\gamma}_0\}\{1 + X(t)\}]\lambda(t,s)dtds.$$

Then we have $|I_2| \leq WM^{-r}|\widetilde{I}_2|$, where W is a constant. Further, by Taylor expansion,

$$\begin{aligned}\widetilde{I}_2 &= \int\int K(z) E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s+hz)^\top \boldsymbol{\gamma}_0\}\{1 + X(s+hz)\}]\lambda(s+hz,s)dzds \\ &= \int E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}\{1 + X(s)\}]\lambda(s,s)ds + O(h^2).\end{aligned}$$

Further,

$$\begin{aligned}&\int E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}\{1 + X(s)\}]\lambda(s,s)ds \\ =& \int E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}]\lambda(s,s)ds + \int E[\widetilde{\mathbf{X}}^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\} X(s)]\lambda(s,s)ds.\end{aligned}$$

According to Assumption 5, for $j = 1, \dots, L$, there exists a constant C_4 such that

$$\begin{aligned}&\Big|\int E[\widetilde{X}_j^\star(s) g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}]\lambda(s,s)ds\Big| = \Big|\int B_j(s) E[g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}]\lambda(s,s)ds\Big| \\ \leq& \int B_j(s) E^{1/2}[g'\{\widetilde{\mathbf{X}}^\star(s)^\top \boldsymbol{\gamma}_0\}^2]\lambda(s,s)ds \leq C_4 \int B_j(s)\lambda(s,s)ds,\end{aligned}$$

$$\Big|\int E[\widetilde{X}_j^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}X(s)]\lambda(s,s)ds\Big| = \Big|\int B_j(s)E[X(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]\lambda(s,s)ds\Big|$$

$$\leq \int B_j(s)E^{1/2}[X^2(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}^2]\lambda(s,s)ds \leq C_4\int B_j(s)\lambda(s,s)ds,$$

Similarly, for $j = L+1,\dots,2L$, there exists a constant C_5 such that

$$\Big|\int E[\widetilde{X}_j^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]\lambda(s,s)ds\Big| \leq C_5\int B_j(s)\lambda(s,s)ds,$$

$$\Big|\int E[\widetilde{X}_j^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}X(s)]\lambda(s,s)ds\Big| \leq C_5\int B_j(s)\lambda(s,s)ds,$$

Let $C_6 = 2\max(C_4, C_5)$, we have

$$\Big|\int E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\{1+X(s)\}]\lambda(s,s)ds\Big| \leq C_6\int \widetilde{\mathbf{B}}(s)\lambda(s,s)ds.$$

On the other hand, by Assumption 2, there exists a constant C_7 such that

$$\Big\|\int \widetilde{\mathbf{B}}(s)\lambda(s,s)ds\Big\|_2^2 = \sum_{j=1}^{2L}\Big\{\int \widetilde{B}_j(s)\lambda(s,s)ds\Big\}^2 \leq C_7\sum_{j=1}^{2L}\Big\{\int \widetilde{B}_j(s)ds\Big\}^2 \leq 2\cdot C_7\sum_{j=1}^{L}\|B_j\|_2^2 = O(1).$$

Hence,

$$\Big\|\int E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\{1+X(s)\}]\lambda(s,s)ds\Big\|_2^2 < \infty.$$

Further,

$$\widetilde{I}_2^\top\widetilde{I}_2 \leq 2\Big\|\int E[\widetilde{\mathbf{X}}^\star(s)g'\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}\{1+X(s)\}]\lambda(s,s)ds\Big\|_2^2 + O(Mh^4) < \infty.$$

Therefore, $I_2^\top I_2 = O(M^{-2r})$. Moreover, we have $I_3^\top I_3 = o(M^{-2r})$. Then by (S1.6) and (S1.7),

$$E\{Q_n(\boldsymbol{\gamma}_0)\}^\top E\{Q_n(\boldsymbol{\gamma}_0)\} = O(Mh^4 + M^{-2r}). \qquad \text{(S1.8)}$$

On the other hand,

$$\text{var}\{h^{1/2}U_{n1}(\boldsymbol{\gamma}_0)\} = \text{var}\Big(\int\int h^{1/2}K_h(t-s)\widetilde{\mathbf{X}}^\star(s)[Y(t) - g\{\widetilde{\mathbf{X}}^\star(s)^\top\boldsymbol{\gamma}_0\}]dN(t,s)\Big)$$

$$=hE\Big[\text{var}\Big\{\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)Y(t)dN(t,s)|X(s), s\in\mathcal{T}, N(t,s),(t,s)\in\mathcal{T}^2\Big\}\Big]$$

$$+ h\text{var}\Big(\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)[g\{\beta_0(t)+\beta_1(t)X(t)\} - g\{\widetilde{\mathbf{X}}^\star(s)^\top\gamma_0\}]dN(t,s)\Big)$$

$$\triangleq D_1 + D_2. \tag{S1.9}$$

According to the derivation of (19) and (20) in Cao et al. (2015), we have

$$D_1 = \int K^2(z)dz\int E\{\widetilde{\mathbf{X}}^\star(s)\widetilde{\mathbf{X}}^\star(s)^\top\}\sigma\{s,X(s)\}^2\lambda(s,s)ds + O(h). \tag{S1.10}$$

For D_2, by Taylor expansion, we have

$$D_2 = h\text{var}\Big(\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)\Big[g\{\widetilde{\mathbf{X}}^\star(t)^\top\gamma_0 + R_n^{(0)}(t) + X(t)R_n^{(1)}(t)\} - g\{\widetilde{\mathbf{X}}^\star(s)^\top\gamma_0\}\Big]dN(t,s)\Big)$$

$$= h\text{var}\Big(\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)\Big[g\{\widetilde{\mathbf{X}}^\star(t)^\top\gamma_0\} - g\{\widetilde{\mathbf{X}}^\star(s)^\top\gamma_0\}$$

$$+ g'\{\widetilde{\mathbf{X}}^\star(t)^\top\gamma_0\}\{R_n^{(0)}(t) + X(t)R_n^{(1)}(t)\} + o\{R_n^{(0)}(t) + X(t)R_n^{(1)}(t)\}\Big]dN(t,s)\Big)$$

$$\triangleq h\text{var}\Big\{\int\int K_h(t-s)\widetilde{\mathbf{X}}^\star(s)G(t,s)dN(t,s)\Big\}$$

$$= hE\Big\{\int\int\int\int K_h(t_1-s_1)K_h(t_2-s_2)\widetilde{\mathbf{X}}^\star(s_1)\widetilde{\mathbf{X}}^\star(s_2)^\top G(t_1,s_1)G(t_2,s_2)dN(t_1,s_1)dN(t_2,s_2)\Big\}$$

$$- h\Big[\int\int K_h(t-s)E\{\widetilde{\mathbf{X}}^\star(s)G(t,s)\}\lambda(t,s)dtds\Big]^2$$

$$\triangleq D_{21} - D_{22}.$$

For D_{21}, we have

$$
\begin{aligned}
D_{21} =& h\int_{t_1\neq t_2}\int_{s_1\neq s_2} K_h(t_1-s_1)K_h(t_2-s_2)E\{\widetilde{\mathbf{X}}^{\star}(s_1)\widetilde{\mathbf{X}}^{\star}(s_2)^{\top}G(t_1,s_1)G(t_2,s_2)\} \\
&\cdot f(t_1,s_1,t_2,s_2)\lambda(t_2,s_2)dt_1dt_2ds_1ds_2 \\
&+ h\int_{t_1}\int_{s_1\neq s_2} K_h(t_1-s_1)K_h(t_1-s_2)E\{\widetilde{\mathbf{X}}^{\star}(s_1)\widetilde{\mathbf{X}}^{\star}(s_2)^{\top}G(t_1,s_1)G(t_1,s_2)\} \\
&\cdot f(t_1,s_1,t_1,s_2)\lambda(t_1,s_2)dt_1ds_1ds_2 \\
&+ h\int_{t_1\neq t_2}\int_{s_1} K_h(t_1-s_1)K_h(t_2-s_1)E\{\widetilde{\mathbf{X}}^{\star}(s_1)\widetilde{\mathbf{X}}^{\star}(s_1)^{\top}G(t_1,s_1)G(t_2,s_1)\} \\
&\cdot f(t_1,s_1,t_2,s_1)\lambda(t_2,s_1)dt_1dt_2ds_1 \\
&+ h\int_{t_1}\int_{s_1} K_h(t_1-s_1)^2E\{\widetilde{\mathbf{X}}^{\star}(s_1)\widetilde{\mathbf{X}}^{\star}(s_1)^{\top}G(t_1,s_1)^2\}\lambda(t_1,s_1)dt_1ds_1
\end{aligned}
$$

Through Taylor expansion, we can get that the first three terms are of order $O(hM^{-2r}+h^3)$ and the last term is of order $O(M^{-2r}+h^2)$ element-wise. Moreover, $D_{22}=O(hM^{-2r}+h^3)$ by Taylor expansion. That means

$$D_2 = O(M^{-2r}+h^2). \tag{S1.11}$$

Similar to the proof of Lemma 1, under Assumption 5, we have that the eigenvalues of $\int E\{\widetilde{\mathbf{X}}^{\star}(s)\widetilde{\mathbf{X}}^{\star}(s)^{\top}\}\sigma\{s,X(s)\}^2\lambda(s,s)ds$ are bounded away from 0 and infinity. Thus, according to (S1.9)-(S1.11), we have $\mathrm{var}\{h^{1/2}U_{n1}(\boldsymbol{\gamma}_0)\} = O(1)$. Then

$$\frac{1}{nh}\mathrm{tr}[\mathrm{var}\{h^{1/2}U_{n1}(\boldsymbol{\gamma}_0)\}] = O(n^{-1}Mh^{-1}). \tag{S1.12}$$

By combining (S1.5), (S1.8) and (S1.12), we can get $E\|Q_n(\boldsymbol{\gamma}_0)\|_2^2 = O(Mh^4 + M^{-2r} + n^{-1}Mh^{-1})$. Therefore,

$$\|Q_n(\boldsymbol{\gamma}_0)\|_2 = O_p(M^{1/2}h^2 + M^{-r} + n^{-1/2}M^{1/2}h^{-1/2}). \tag{S1.13}$$

For $\bar{N}\|\widetilde{P}_1(\boldsymbol{\gamma}_0)\|_2$ and $\bar{N}\|\widetilde{P}_2(\boldsymbol{\gamma}_0)\|_2$, we have

$$\widetilde{P}_1(\boldsymbol{\gamma}_0)^\top \widetilde{P}_1(\boldsymbol{\gamma}_0) = \boldsymbol{\gamma}_0^\top \mathbf{V}_{\rho_0,\rho_1}^\top \mathbf{V}_{\rho_0,\rho_1}\boldsymbol{\gamma}_0 = O(\rho^2 M^{-1}), \tag{S1.14}$$
$$\widetilde{P}_2(\boldsymbol{\gamma}_0)^\top \widetilde{P}_2(\boldsymbol{\gamma}_0) = \left\|\frac{M+1}{2T}\frac{\partial}{\partial \boldsymbol{\gamma}_0^{(1)}}\int_{\mathcal{T}} p_\lambda(|\mathbf{B}(t)^\top \boldsymbol{\gamma}_0^{(1)}|)dt\right\|_2^2.$$

Refer to Lin et al. (2017), by Assumption 3,

$$\left|\frac{\partial}{\partial \gamma_{0j}}\int_{\mathcal{T}} p_\lambda(|\mathbf{B}(t)^\top \boldsymbol{\gamma}_0^{(1)}|)dt\right| = O(n^{-1/2}M^{-2}), j = L+1, \ldots, 2L.$$

Thus,

$$\widetilde{P}_2(\boldsymbol{\gamma}_0)^\top \widetilde{P}_2(\boldsymbol{\gamma}_0) = O(n^{-1}M^{-1}). \tag{S1.15}$$

As $\bar{N} = O_p(1)$, by (S1.14) and (S1.15), we have

$$\bar{N}\|\widetilde{P}_1(\boldsymbol{\gamma}_0)\|_2 = O_p(\rho M^{-1/2}), \tag{S1.16}$$

$$\bar{N}\|\widetilde{P}_2(\boldsymbol{\gamma}_0)\|_2 = O_p(n^{-1/2}M^{-1/2}). \tag{S1.17}$$

By combining (S1.4), (S1.13), (S1.16) and (S1.17), we have $\|\psi_n(\boldsymbol{\gamma}_0)\|_2 = O_p(M^{1/2}h^2 + M^{-r} + n^{-1/2}M^{1/2}h^{-1/2} + \rho M^{-1/2}) = O_p(\alpha_n)$. The proof is completed.

□

S2 Proof of Theorem 2

Define

$$\mathcal{T}_1 = \{t \in \mathcal{T} : |\beta_1(t)| > aC_8(\lambda + M^{-r})\},$$

$$\mathcal{T}_2 = \{t \in \mathcal{T} : \beta_1(t) = 0\},$$

$$\mathcal{T}_3 = \mathcal{T} - \mathcal{T}_1 - \mathcal{T}_2.$$

We further define $\mathcal{S}_l = \mathrm{SUPP}(B_l), l = 1, \dots, L$. Let $\mathcal{A}_j = \{l : \mathcal{S}_l \subset \mathcal{T}_j\}, j = 1, 2$, and $\mathcal{A}_3 = \{1, \dots, L\} - \mathcal{A}_1 - \mathcal{A}_2$.

Proof of Theorem 2. Let $U_n^{(l)}(\boldsymbol{\gamma})$ be the $(L + l)$-th element of $U_n(\boldsymbol{\gamma})$ and $Q_n^{(l)}(\boldsymbol{\gamma})$ be the $(L + l)$-th element of $Q_n(\boldsymbol{\gamma})$. For $l \in \mathcal{A}_2$,

$$\begin{aligned} U_n^{(l)}(\boldsymbol{\gamma}) =& n^{-1} \sum_{i=1}^{n} \int\int K_h(t-s)\widetilde{X}_{il}(s)[Y_i(t) - g\{\widetilde{\mathbf{X}}_i^{\star}(s)^\top \boldsymbol{\gamma}\}]dN_i(t,s) \\ & - \bar{N} \cdot (\rho_1 \mathbf{V}\boldsymbol{\gamma})_l - \frac{\bar{N}}{2} \cdot \frac{M+1}{T} \frac{\partial}{\partial \gamma_l^{(1)}} \int p_\lambda(|\mathbf{B}(t)^\top \boldsymbol{\gamma}^{(1)}|)dt \\ =& Q_n^{(l)}(\boldsymbol{\gamma}) - \bar{N}(\rho_1 \mathbf{V}\boldsymbol{\gamma})_l - \frac{\bar{N}}{2} \cdot \frac{M+1}{T} \int p'_\lambda(|\mathbf{B}(t)^\top \boldsymbol{\gamma}^{(1)}|)B_l(t)\mathrm{sgn}(\gamma_l^{(1)})dt. \end{aligned}$$

Then

$$\begin{aligned}
&\left|\lambda^{-1}U_n^{(l)}(\widehat{\boldsymbol{\gamma}})+\frac{\bar{N}}{2}\cdot\frac{M+1}{T}\mathrm{sgn}(\widehat{\gamma}_l^{(1)})\int\lambda^{-1}p'_\lambda(|\mathbf{B}(t)^\top\boldsymbol{\gamma}^{(1)}|)\Big|_{\boldsymbol{\gamma}^{(1)}=\widehat{\boldsymbol{\gamma}}^{(1)}}B_l(t)dt\right|\\
=&\lambda^{-1}\left|Q_n^{(l)}(\widehat{\boldsymbol{\gamma}})-\bar{N}\cdot(\rho_1\mathbf{V}\widehat{\boldsymbol{\gamma}})_l\right|\leq\lambda^{-1}\left|Q_n^{(l)}(\widehat{\boldsymbol{\gamma}})\right|+\lambda^{-1}\bar{N}\left|(\rho_1\mathbf{V}\widehat{\boldsymbol{\gamma}})_l\right|\\
=&\lambda^{-1}\left|Q_n^{(l)}(\boldsymbol{\gamma}_0)+\sum_{j=1}^{2L}\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial\gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}(\widehat{\gamma}_g-\gamma_{0g})\right|+\lambda^{-1}O_p(\rho M^{-1})\\
\leq&\lambda^{-1}\left|Q_n^{(l)}(\boldsymbol{\gamma}_0)\right|+\lambda^{-1}\sum_{j=1}^{2L}\left|\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial\gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}\right|\cdot|\widehat{\gamma}_g-\gamma_{0g}|+\lambda^{-1}O_p(\rho M^{-1}),
\end{aligned}\tag{S2.18}$$

where $\boldsymbol{\gamma}_0^\star$ lies between $\boldsymbol{\gamma}_0$ and $\widehat{\boldsymbol{\gamma}}$. According to the derivation of Lemma 2, we have $\mathrm{var}\{h^{1/2}U_{n1}(\boldsymbol{\gamma}_0)\}=O(1)$. Thus, $\mathrm{var}\{Q_n^{(l)}(\boldsymbol{\gamma}_0)\}=O(n^{-1}h^{-1})$. Then

$$\left|Q_n^{(l)}(\boldsymbol{\gamma}_0)-EQ_n^{(l)}(\boldsymbol{\gamma}_0)\right|=O_p(n^{-1/2}h^{-1/2}).\tag{S2.19}$$

Moreover, based on the computation of $E\{Q_n(\boldsymbol{\gamma}_0)\}$ in the proof of Lemma 2,

$$\left|EQ_n^{(l)}(\boldsymbol{\gamma}_0)\right|=O(h^2).\tag{S2.20}$$

By combining (S2.19) and (S2.20), we can get

$$\left|Q_n^{(l)}(\boldsymbol{\gamma}_0)\right|=O_p(n^{-1/2}h^{-1/2}+h^2)=O_p(n^{-1/2}h^{-1/2}).\tag{S2.21}$$

On the other hand,

$$\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial\gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}=-\frac{1}{n}\sum_{i=1}^n\int\int K_h(t-s)\widetilde{X}_{il}(s)g'\{\widetilde{\mathbf{X}}_i^\star(s)^\top\boldsymbol{\gamma}_0^\star\}\widetilde{X}_{ig}^\star(s)dN_i(t,s).$$

Similar to the computation of $\mathrm{var}(\eta_{j_1 j_2})$ in the proof of Lemma 1, we have $\mathrm{var}\Big\{\frac{\partial Q_n^{(l)}(\gamma)}{\partial \gamma_g}\Big|_{\gamma=\gamma_0^\star}\Big\} = O(M^{-1}n^{-1}h^{-1})$. Then

$$\left|\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial \gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star} - E\Big\{\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial \gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}\Big\}\right| = O_p(M^{-1/2}n^{-1/2}h^{-1/2}). \quad \text{(S2.22)}$$

Furthermore, by Taylor expansion and Assumption 5, we have

$$\left|E\Big\{\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial \gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}\Big\}\right| = \left|\int\int K_h(t-s)E[\widetilde{X}_{il}(s)g'\{\widetilde{\mathbf{X}}_i^\star(s)^\top\boldsymbol{\gamma}_0^\star\}\widetilde{X}_{ig}^\star(s)]\lambda(t,s)dtds\right| = O(M^{-1}). \quad \text{(S2.23)}$$

By combining (S2.22) and (S2.23), we have

$$\left|\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial \gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}\right| = O_p(M^{-1/2}n^{-1/2}h^{-1/2} + M^{-1}) = O_p(M^{-1}).$$

Since $|\widehat{\gamma}_g - \gamma_{0g}| = O_p(n^{-1/2}M^{1/2}h^{-1/2})$, we have

$$\sum_{j=1}^{2L}\left|\frac{\partial Q_n^{(l)}(\boldsymbol{\gamma})}{\partial \gamma_g}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0^\star}\right| \cdot |\widehat{\gamma}_g - \gamma_{0g}| = O_p(n^{-1/2}M^{1/2}h^{-1/2}). \quad \text{(S2.24)}$$

Then by (S2.18), (S2.21) and (S2.24), we have

$$\left|\lambda^{-1}U_n^{(l)}(\widehat{\boldsymbol{\gamma}}) + \frac{\bar{N}}{2}\cdot\frac{M+1}{T}\mathrm{sgn}(\widehat{\gamma}_l^{(1)})\int \lambda^{-1}p'_\lambda(|\mathbf{B}(t)^\top\boldsymbol{\gamma}^{(1)}|)\Big|_{\boldsymbol{\gamma}^{(1)}=\widehat{\boldsymbol{\gamma}}^{(1)}}B_l(t)dt\right|$$

$$=O_p(\lambda^{-1}n^{-1/2}h^{-1/2} + \lambda^{-1}n^{-1/2}M^{1/2}h^{-1/2} + \lambda^{-1}\rho M^{-1}) \to 0.$$

Therefore, $\lambda^{-1}U_n^{(l)}(\widehat{\boldsymbol{\gamma}})$ and $-\frac{\bar{N}}{2}\cdot\frac{M+1}{T}\mathrm{sgn}(\widehat{\gamma}_l^{(1)})\int \lambda^{-1}p'_\lambda(|\mathbf{B}(t)^\top\boldsymbol{\gamma}^{(1)}|)\Big|_{\boldsymbol{\gamma}^{(1)}=\widehat{\boldsymbol{\gamma}}^{(1)}}B_l(t)dt$ share the same sign. Since $U_n^{(l)}(\widehat{\boldsymbol{\gamma}}) = 0$ and $\liminf_{n\to\infty}\liminf_{x\to 0^+}\lambda^{-1}p'_\lambda(x) > 0$, we have $\widehat{\gamma}_l^{(1)} = 0$ in probability for all $l \in \mathcal{A}_2$.

Define $\widehat{\mathcal{A}}_2 = \{l \in \mathcal{A}_2 : \widehat{\gamma}_l^{(1)} = 0\}$. Then we have $\widehat{\mathcal{A}}_2 = \mathcal{A}_2$ in probability. Based on the compact support property of B-spline basis, $\bigcup_{l\in\mathcal{A}_2} \mathcal{S}_l$ converges to $\text{NULL}(\beta_1)$ as $M \to \infty$. Therefore,

$$\bigcup_{l\in\widehat{\mathcal{A}}_2} \mathcal{S}_l \to \text{NULL}(\beta_1) \tag{S2.25}$$

in probability. Moreover, by the definition, we have for any $\varepsilon > 0$, there exists sufficient large n and M, such that

$$\bigcup_{l\in\widehat{\mathcal{A}}_2} \mathcal{S}_l \subset \text{NULL}^{\varepsilon}(\widehat{\beta}_1), \tag{S2.26}$$

where $\text{NULL}^{\varepsilon}(\widehat{\beta}_1)$ is the ε-neighborhood of $\text{NULL}(\widehat{\beta}_1)$. Here the ε-neighborhood of a subset G of $\mathcal{T}$ is defined by $\{t \in \mathcal{T} : \inf_{u\in G} |u - t| < \varepsilon\}$. According to Theorem 1, we have $\|\widehat{\beta}_1 - \beta_1\|_\infty = O_p(n^{-1/2}M^{1/2}h^{-1/2} + M^{-r})$. Since $n^{-1/2}M^{1/2}h^{-1/2}$ is dominated by λ, we also have $\|\widehat{\beta}_1 - \beta_1\|_\infty = O_p(\lambda + M^{-r})$. So for $t \in \mathcal{T}_1$, there exists a constant $C_9 > 1$ such that $|\widehat{\beta}_1(t) - \beta_1(t)| \leq aC_9(\lambda + M^{-r})$ in probability. Let $C_8 = 2C_9$. As $|\beta_1(t)| > aC_8(\lambda + M^{-r})$, we have $|\widehat{\beta}_1(t)| \geq aC_9(\lambda + M^{-r}) > a\lambda$ in probability. Thus, we have $t \in \text{SUPP}(\widehat{\beta}_1)$ in probability. That means $\mathcal{T}_1 \subset \text{SUPP}(\widehat{\beta}_1)$ in probability. So as $n \to \infty$ and $M \to \infty$,

$$\text{NULL}(\widehat{\beta}_1) \subset \mathcal{T}_2 \cup \mathcal{T}_3 = \text{NULL}(\beta_1) \cup \mathcal{T}_3. \tag{S2.27}$$

Since $\mathcal{T}_3 \to \emptyset$ in probability and $\text{NULL}(\widehat{\beta}_1)$ is closed, we have $\text{NULL}(\widehat{\beta}_1) \to$

NULL(β_1) and SUPP($\widehat{\beta}_1$) $\to$ SUPP(β_1) in probability by (S2.25) - (S2.27). The proof is completed.

□

S3 Proof of Theorem 3

Proof of Theorem 3. By Taylor expansion, we have

$$Y_i(t) - g\{\widetilde{\mathbf{X}}_i^\star(s)^\top \widehat{\boldsymbol{\gamma}}\} = Y_i(t) - g\{\eta_i(s, \boldsymbol{\beta}_0)\} - g'\{\eta_i(s, \boldsymbol{\beta}_0)\}[\widetilde{\mathbf{X}}_i^\star(s)^\top(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) - e_i(s)]\{1 + o_p(1)\},$$

where $e_i(s) = R_n^{(0)}(s) + X_i(t)R_n^{(1)}(s)$. Then

$$\begin{aligned}
\psi_n(\widehat{\boldsymbol{\gamma}}) =& n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\widetilde{\mathbf{X}}_i^\star(s)^\top\widehat{\boldsymbol{\gamma}}\}]dN_i(t,s) - \bar{N}\widetilde{P}_1(\widehat{\boldsymbol{\gamma}}) - \bar{N}\widetilde{P}_2(\widehat{\boldsymbol{\gamma}}) \\
=& n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s) - \bar{N}\widetilde{P}_1(\boldsymbol{\gamma}_0) - \bar{N}\widetilde{P}_2(\boldsymbol{\gamma}_0) \\
& - \Big[n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)g'\{\eta_i(s,\boldsymbol{\beta}_0)\}\widetilde{\mathbf{X}}_i^\star(s)^\top dN_i(t,s)\Big](\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\{1 + o_p(1)\} \\
& + \Big[n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)g'\{\eta_i(s,\boldsymbol{\beta}_0)\}e_i(s)dN_i(t,s)\Big]\{1 + o_p(1)\} \\
& - \bar{N}\{\widetilde{P}_1(\widehat{\boldsymbol{\gamma}}) - \widetilde{P}_1(\boldsymbol{\gamma}_0)\} - \bar{N}\{\widetilde{P}_2(\widehat{\boldsymbol{\gamma}}) - \widetilde{P}_2(\boldsymbol{\gamma}_0)\}. \qquad \text{(S3.28)}
\end{aligned}$$

Moreover,

$$\bar{N}\{\widetilde{P}_1(\widehat{\boldsymbol{\gamma}}) - \widetilde{P}_1(\boldsymbol{\gamma}_0)\} = \bar{N}\mathbf{V}_{\rho_0,\rho_1}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0),$$

$$\begin{aligned}\bar{N}\{\widetilde{P}_2(\widehat{\boldsymbol{\gamma}}) - \widetilde{P}_2(\boldsymbol{\gamma}_0)\} &= \bar{N} \cdot \frac{M+1}{2T}\left\{\frac{\partial}{\partial \boldsymbol{\gamma}_0}\int_{\mathcal{T}} p_\lambda(|\mathbf{B}(t)^\top \boldsymbol{\gamma}_0^{(1)}|)dt - \frac{\partial}{\partial \widehat{\boldsymbol{\gamma}}}\int_{\mathcal{T}} p_\lambda(|\mathbf{B}(t)^\top \widehat{\boldsymbol{\gamma}}^{(1)}|)dt\right\} \\ &= \bar{N} \cdot \frac{M+1}{2T} \cdot \left\{\frac{\partial^2}{\partial \boldsymbol{\gamma}_0^2}\int_{\mathcal{T}} p_\lambda(|\mathbf{B}(t)^\top \boldsymbol{\gamma}_0^{(1)}|)dt\right\} \cdot (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\{1 + o_p(1)\}.\end{aligned}$$

Then, (S3.28) can be written as

$$\begin{aligned}\psi_n(\widehat{\boldsymbol{\gamma}}) =& n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s) \\ &- \{\Omega_n + o_p(1)\}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \boldsymbol{\gamma}_n,\end{aligned}$$

where

$$\Omega_n = n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s) g'\{\eta_i(s,\boldsymbol{\beta}_0)\}\widetilde{\mathbf{X}}_i^\star(s)^\top dN_i(t,s),$$

$$\boldsymbol{\gamma}_n = -\bar{N}\widetilde{P}_1(\boldsymbol{\gamma}_0) - \bar{N}\widetilde{P}_2(\boldsymbol{\gamma}_0) + n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s) g'\{\eta_i(s,\boldsymbol{\beta}_0)\}e_i(s)dN_i(t,s).$$

As $\psi_n(\widehat{\boldsymbol{\gamma}}) = 0$, we have

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 = \{\Omega_n + o_p(1)\}^{-1}\left(n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s) + \boldsymbol{\gamma}_n\right). \tag{S3.29}$$

According to the derivation of $I_2^\top I_2 = O(M^{-2r})$ in the proof of Lemma 2, we have

$$\left\|n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s) g'\{\eta_i(s,\boldsymbol{\beta}_0)\}e_i(s)dN_i(t,s)\right\|_2 = O_p(M^{-r}).$$

Through (S1.16) and (S1.17),

$$\bar{N}\|\widetilde{P}_1(\boldsymbol{\gamma}_0)\|_2 = O_p(\rho M^{-1/2}),$$

$$\bar{N}\|\widetilde{P}_2(\boldsymbol{\gamma}_0)\|_2 = O_p(n^{-1/2}M^{-1/2}).$$

Then, it follows that

$$nh(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^\top \Omega_n^2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + O_p(1) = n^{-1}\sum_{i,j}^n P_i^\top P_j, \tag{S3.30}$$

where

$$P_i = h^{1/2}\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s).$$

Since $O_p(1/\sqrt{2\mathrm{tr}(\Sigma_0^2)}) = o_p(1)$, we then want to show that

$$\frac{n^{-1}\sum_{i,j}^n P_i^\top P_j - \mathrm{tr}(\Sigma_0)}{\sqrt{2\mathrm{tr}(\Sigma_0^2)}} \xrightarrow{d} N(0,1). \tag{S3.31}$$

Here $\Sigma_0 = \mathrm{var}(P_i)$. Let $\Delta_0 = E(P_i)$. By using similar technique in the proof of Lemma 2, we have

$$\Delta_0^\top \Delta_0 = O(Mh^5),$$

$$\mathrm{tr}(\Sigma_0^l) = O(M), l = 1, 2, 4.$$

The proof of (S3.31) is analogous to the proof of Theorem 4 in Li et al. (2020), so we just briefly introduce the idea here. First, we have

$$\begin{aligned} & n^{-1}\sum_{i,j}^{n} P_i^\top P_j - n\Delta_0^\top \Delta_0 - \mathrm{tr}(\Sigma_0) \\ =& n^{-1}\sum_{i\neq j}^{n}(P_i-\Delta_0)^\top(P_j-\Delta_0) + \left\{n^{-1}\sum_{i=1}^{n}(P_i-\Delta_0)^\top(P_i-\Delta_0) - \mathrm{tr}(\Sigma_0)\right\} \\ &+ n^{-1}\sum_{i,j}^{n}(P_i^\top\Delta_0 + P_j^\top\Delta_0 - 2\Delta_0^\top\Delta_0) \triangleq Q_1 + Q_2 + Q_3. \end{aligned} \tag{S3.32}$$

Then through Corollary 3.1 of Hall and Heyde (1980), it can be shown that

$$\frac{Q_1}{\sigma_n} \xrightarrow{d} N(0,1),$$

where $\sigma_n = \sqrt{2\mathrm{tr}(\Sigma_0^2)}$. Furthermore, as

$$E(Q_2) = E(Q_3) = 0, \mathrm{var}(Q_2) \leq O(n^{-1}\mathrm{tr}^2(\Sigma_0)), \mathrm{var}(Q_3) = O(nh^5M^2),$$

we have

$$\frac{Q_2}{\sigma_n} \leq O_p(n^{-1/2}M^{1/2}) = o_p(1), \frac{Q_3}{\sigma_n} = O_p(n^{1/2}h^{5/2}M^{1/2}) = o_p(1).$$

Moreover,

$$\frac{n\Delta_0^\top\Delta_0}{\sigma_n} = O_p(nh^5M^{1/2}) = o_p(1).$$

Therefore, by (S3.32),

$$\frac{n^{-1}\sum_{i,j}^{n} P_i^\top P_j - \mathrm{tr}(\Sigma_0)}{\sqrt{2\mathrm{tr}(\Sigma_0^2)}} = \frac{Q_1}{\sigma_n} + o_p(1) \xrightarrow{d} N(0,1).$$

Hence, according to (S3.30), we have

$$\frac{nh(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)^\top\Omega_n^2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)-\mathrm{tr}(\Sigma_0)}{\sqrt{2\mathrm{tr}(\Sigma_0^2)}}\xrightarrow{d}N(0,1).$$

The proof is completed.

□

S4 Point-wise asymptotic distribution

Theorem 4. *Suppose that the conditions of Theorem 2 are satisfied, then for any point $t\in\mathcal{T}$, we have*

$$\sqrt{nh}\{\widehat{\beta}_0(t)-\beta_0(t)\}\xrightarrow{d}N(0,\sigma_0^2(t)),$$

$$\sqrt{nh}\{\widehat{\beta}_1(t)-\beta_1(t)\}\xrightarrow{d}N(0,\sigma_1^2(t)),$$

where $\sigma_0^2(t)=\lim_{n\to\infty}\widetilde{\boldsymbol{B}}_0(t)^\top\Omega_X^{-1}\Sigma_X\Omega_X^{-1}\widetilde{\boldsymbol{B}}_0(t)$, $\sigma_1^2(t)=\lim_{n\to\infty}\widetilde{\boldsymbol{B}}_1(t)^\top\Omega_X^{-1}\Sigma_X\Omega_X^{-1}\widetilde{\boldsymbol{B}}_1(t)$, $\widetilde{\boldsymbol{B}}_0(t)=(\boldsymbol{B}(t)^\top,\boldsymbol{0}^\top)^\top$, $\widetilde{\boldsymbol{B}}_1(t)=(\boldsymbol{0}^\top,\boldsymbol{B}(t)^\top)^\top$, *$\boldsymbol{0}$ is a zero-valued vector with length L, and*

$$\Omega_X=\int E\{\widetilde{\boldsymbol{X}}^\star(s)\widetilde{\boldsymbol{X}}^\star(s)^\top\}g'\{\eta_i(s,\boldsymbol{\beta}_0)\}\lambda(s,s)ds,$$

$$\Sigma_X=\int K^2(z)dz\int E\{\widetilde{\boldsymbol{X}}^\star(s)\widetilde{\boldsymbol{X}}^\star(s)^\top\}\sigma\{s,X(s)\}^2\lambda(s,s)ds.$$

Proof of Theorem 4. For any $t\in\mathcal{T}$, we have

$$\sqrt{nh}\{\widehat{\beta}_1(t)-\beta_1(t)\}=\sqrt{nh}\widetilde{\mathbf{B}}_1(t)^\top(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)+\sqrt{nh}\{\widetilde{\mathbf{B}}_1(t)^\top\boldsymbol{\gamma}_0-\beta_1(t)\}. \tag{S4.33}$$

First, by Assumption 1, we have

$$\sup_t \sqrt{nh}|\widetilde{\mathbf{B}}_1(t)^\top \boldsymbol{\gamma}_0 - \beta_1(t)| = O_p(n^{1/2}h^{1/2}M^{-r}) = o_p(1). \tag{S4.34}$$

On the other hand, by (S3.29), we have

$$\begin{aligned}
&\sqrt{nh}\widetilde{\mathbf{B}}_1(t)^\top(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \\
=&\sqrt{nh}\widetilde{\mathbf{B}}_1(t)^\top\{\Omega_n + o_p(1)\}^{-1}\left(n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s) + \boldsymbol{\gamma}_n\right) \\
=&F_1 + F_2,
\end{aligned} \tag{S4.35}$$

where

$$F_1 = \sqrt{nh}\widetilde{\mathbf{B}}_1(t)^\top\{\Omega_n + o_p(1)\}^{-1}\left(n^{-1}\sum_{i=1}^n \int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s)\right),$$

$$F_2 = \sqrt{nh}\widetilde{\mathbf{B}}_1(t)^\top\{\Omega_n + o_p(1)\}^{-1}\boldsymbol{\gamma}_n.$$

According to the proof of Theorem 3, it can be shown that

$$\|F_2\|_2 = o_p(1). \tag{S4.36}$$

For F_1, let

$$\phi_i = \sqrt{nh}n^{-1}\widetilde{\mathbf{B}}_1(t)^\top\{\Omega_n + o_p(1)\}^{-1}\int\int K_h(t-s)\widetilde{\mathbf{X}}_i^\star(s)[Y_i(t) - g\{\eta_i(s,\boldsymbol{\beta}_0)\}]dN_i(t,s).$$

Then $F_1 = \sum_{i=1}^n \phi_i$. Similar to the proof of Theorem 1 in Cao et al. (2015), we also have

$$\sum_{i=1}^n E\{|\phi_i - E\phi_i|^3\} = nO(n^{3/2}h^{3/2}n^{-3}h^{-2}) = O(n^{-1/2}h^{-1/2}),$$

which verifies the Lyapunov condition. Hence, we have

$$\sum_{i=1}^{n}(\phi_i - E\phi_i) \xrightarrow{d} N(0, \sigma_1^2(t)),$$

where $\sigma_1^2(t)$ can be obtained analogously to the computation of $\mathrm{var}\{h^{1/2}U_{n1}(\boldsymbol{\gamma}_0)\}$ in (S1.9). Moreover, we have $\| \sum_{i=1}^{n} E\phi_i\|_2^2 = o(1)$ by (S1.7). Therefore,

$$F_1 \xrightarrow{d} N(0, \sigma_1^2(t)). \tag{S4.37}$$

Then combining (S4.33)-(S4.36), we have

$$\sqrt{nh}\{\widehat{\beta}_1(t) - \beta_1(t)\} \xrightarrow{d} N(0, \sigma_1^2(t)).$$

The asymptotic normality of $\beta_0(t)$ can be derived in the same way. The proof is completed.

□

S5 Additional simulation studies

S5.1 The effect of L

In this section, we report the simulation results of LocKer and PLSE methods with the use of various values of L in Bernoulli and Poisson cases. The settings are the same as settings in Section 4.1, except that the observation times of response and covariate are set to be synchronous. Tables 1-2 provide the averaged ISE_0, ISE_1, TP and FN for Bernoulli cases. Here PLSE becomes invalid because it only adapts to regression model with Gaussian response. For identifying ability of the proposed LocKer method, it also performs the best when using $L = 13$ for the sparse setting, which is caused by the same reason as Gaussian cases. However, we find that large value of L does not improve the estimation here. We conjecture the reason is that large value of L can bring more parameters in the estimation, which is quite adverse for Bernoulli cases. Furthermore, Tables 3-4 present the simulation results for Poisson cases. It is shown that for Poisson cases, large value of L can bring helps to the estimation of our method in terms of both accuracy and identifying ability. But large value of L would complicate the estimation, which should also be taken into account.

Table 1: The averaged ISE_0, ISE_1, TP and FN across 100 runs for PLSE and LocKer using various values of L when $n = 200, m = 15$ in Bernoulli cases, with standard deviation in parentheses.

			ISE_0	ISE_1	TP	FN
$L = 10$	Nonsparse	PLSE	0.5297 (0.0156)	0.3267 (0.0218)	–	0 (0)
		LocKer	0.0242 (0.0098)	0.0315 (0.0207)	–	0 (0)
	Sparse	PLSE	0.5337 (0.0147)	0.4211 (0.0181)	0.2287 (0.2756)	0 (0)
		LocKer	0.0184 (0.0089)	0.0839 (0.0447)	0.6082 (0.2252)	0 (0)
$L = 13$	Nonsparse	PLSE	0.5340 (0.0192)	0.3225 (0.0289)	–	0 (0)
		LocKer	0.0189 (0.0092)	0.0328 (0.0226)	–	0 (0)
	Sparse	PLSE	0.5320 (0.0126)	0.4195 (0.0167)	0.3110 (0.2518)	0 (0)
		LocKer	0.0162 (0.0076)	0.0857 (0.0533)	0.8307 (0.1817)	0 (0)
$L = 15$	Nonsparse	PLSE	0.5302 (0.0200)	0.3280 (0.0283)	–	0 (0)
		LocKer	0.0194 (0.0080)	0.0319 (0.0217)	–	0 (0)
	Sparse	PLSE	0.5304 (0.0126)	0.4214 (0.0149)	0.2232 (0.2158)	0.0339 (0.0474)
		LocKer	0.0158 (0.0085)	0.0985 (0.0385)	0.8084 (0.1222)	0.0196 (0.0420)
$L = 20$	Nonsparse	PLSE	0.5238 (0.0190)	0.3408 (0.0299)	–	0.0140 (0.0289)
		LocKer	0.0163 (0.0083)	0.0314 (0.0220)	–	0 (0)
	Sparse	PLSE	0.5259 (0.0122)	0.4224 (0.0114)	0.2918 (0.2206)	0.0646 (0.0646)
		LocKer	0.0141 (0.0074)	0.0966 (0.0379)	0.7865 (0.1106)	0.0031 (0.0158)

Table 2: The averaged ISE_0, ISE_1, TP and FN across 100 runs for PLSE and LocKer using various values of L when $n = 200, m = 20$ in Bernoulli cases, with standard deviation in parentheses.

			ISE_0	ISE_1	TP	FN
$L = 10$	Nonsparse	PLSE	0.5332 (0.0174)	0.3178 (0.0238)	–	0 (0)
		LocKer	0.0148 (0.0069)	0.0225 (0.0165)	–	0 (0)
	Sparse	PLSE	0.5362 (0.0113)	0.4174 (0.0112)	0.1601 (0.1962)	0 (0)
		LocKer	0.0130 (0.0067)	0.0597 (0.0265)	0.6826 (0.1948)	0 (0)
$L = 13$	Nonsparse	PLSE	0.5305 (0.0184)	0.3243 (0.0270)	–	0 (0)
		LocKer	0.0135 (0.0057)	0.0228 (0.0155)	–	0 (0)
	Sparse	PLSE	0.5339 (0.0107)	0.4186 (0.0120)	0.2811 (0.2380)	0 (0)
		LocKer	0.0117 (0.0070)	0.0607 (0.0327)	0.8926 (0.1584)	0 (0)
$L = 15$	Nonsparse	PLSE	0.5332 (0.0170)	0.3222 (0.0266)	–	0 (0)
		LocKer	0.0133 (0.0064)	0.0237 (0.0147)	–	0 (0)
	Sparse	PLSE	0.5318 (0.0116)	0.4184 (0.0126)	0.3171 (0.2334)	0.0401 (0.0520)
		LocKer	0.0112 (0.0060)	0.0716 (0.0365)	0.8462 (0.0844)	0.0252 (0.0389)
$L = 20$	Nonsparse	PLSE	0.5252 (0.0202)	0.3338 (0.0284)	–	0.0105 (0.0254)
		LocKer	0.0115 (0.0054)	0.0228 (0.0149)	–	0 (0)
	Sparse	PLSE	0.5268 (0.0120)	0.4247 (0.0136)	0.3450 (0.2281)	0.0875 (0.0595)
		LocKer	0.0116 (0.0065)	0.0741 (0.0382)	0.8549 (0.1199)	0.0035 (0.0154)

Table 3: The averaged ISE_0, ISE_1, TP and FN across 100 runs for PLSE and LocKer using various values of L when $n = 200, m = 15$ in Poisson cases, with standard deviation in parentheses.

			ISE_0	ISE_1	TP	FN
$L = 10$	Nonsparse	PLSE	1.6899 (0.0574)	0.3516 (0.0719)	–	0 (0)
		LocKer	0.0090 (0.0037)	0.0134 (0.0040)	–	0 (0)
	Sparse	PLSE	1.4547 (0.0712)	0.0921 (0.0402)	0.2177 (0.2269)	0 (0)
		LocKer	0.0117 (0.0044)	0.0286 (0.0113)	0.6507 (0.1580)	0 (0)
$L = 13$	Nonsparse	PLSE	1.7498 (0.0713)	0.3881 (0.1143)	–	0 (0)
		LocKer	0.0069 (0.0036)	0.0116 (0.0040)	–	0 (0)
	Sparse	PLSE	1.5041 (0.0708)	0.0842 (0.0382)	0.2127 (0.2165)	0 (0)
		LocKer	0.0090 (0.0036)	0.0112 (0.0099)	0.9550 (0.1115)	0 (0)
$L = 15$	Nonsparse	PLSE	1.7761 (0.0642)	0.3724 (0.0838)	–	0.0017 (0.0117)
		LocKer	0.0068 (0.0035)	0.0111 (0.0038)	–	0 (0)
	Sparse	PLSE	1.5229 (0.0757)	0.0943 (0.0371)	0.2530 (0.2117)	0.0739 (0.0613)
		LocKer	0.0096 (0.0038)	0.0237 (0.0092)	0.8528 (0.0432)	0.0139 (0.0375)
$L = 20$	Nonsparse	PLSE	1.8080 (0.0809)	0.3674 (0.0803)	–	0.0169 (0.0356)
		LocKer	0.0061 (0.0037)	0.0101 (0.0044)	–	0 (0)
	Sparse	PLSE	1.5288 (0.0709)	0.1153 (0.0430)	0.2679 (0.1899)	0.1237 (0.0847)
		LocKer	0.0097 (0.0038)	0.0256 (0.0106)	0.8808 (0.0675)	0.0024 (0.0187)

Table 4: The averaged ISE_0, ISE_1, TP and FN across 100 runs for PLSE and LocKer using various values of L when $n = 200, m = 20$ in Poisson cases, with standard deviation in parentheses.

			ISE_0	ISE_1	TP	FN
$L = 10$	Nonsparse	PLSE	1.7484 (0.0582)	0.3319 (0.0678)	–	0 (0)
		LocKer	0.0059 (0.0023)	0.0095 (0.0026)	–	0 (0)
	Sparse	PLSE	1.5469 (0.0651)	0.0791 (0.0340)	0.1689 (0.2231)	0 (0)
		LocKer	0.0084 (0.0034)	0.0241 (0.0066)	0.6560 (0.1649)	0 (0)
$L = 13$	Nonsparse	PLSE	1.8154 (0.0661)	0.3414 (0.0652)	–	0 (0)
		LocKer	0.0046 (0.0022)	0.0086 (0.0028)	–	0 (0)
	Sparse	PLSE	1.5771 (0.0760)	0.0813 (0.0345)	0.2748 (0.2236)	0.0020 (0.0200)
		LocKer	0.0064 (0.0027)	0.0091 (0.0064)	0.9954 (0.0283)	0 (0)
$L = 15$	Nonsparse	PLSE	1.8283 (0.0694)	0.3395 (0.0529)	–	0 (0)
		LocKer	0.0046 (0.0022)	0.0084 (0.0030)	–	0 (0)
	Sparse	PLSE	1.5891 (0.0652)	0.0856 (0.0322)	0.2806 (0.2482)	0.0745 (0.0573)
		LocKer	0.0068 (0.0032)	0.0180 (0.0093)	0.8691 (0.0493)	0.0238 (0.0368)
$L = 20$	Nonsparse	PLSE	1.8344 (0.0665)	0.3262 (0.0581)	–	0.0105 (0.0224)
		LocKer	0.0044 (0.0024)	0.0076 (0.0030)	–	0 (0)
	Sparse	PLSE	1.5903 (0.0665)	0.0973 (0.0300)	0.2870 (0.2094)	0.1162 (0.0778)
		LocKer	0.0056 (0.0029)	0.0103 (0.0079)	0.9430 (0.0278)	0.0087 (0.0331)

S5.2 Asymptotic distribution

In this section, we explore the asymptotic distribution of $\widehat{\boldsymbol{\gamma}}$ by numerical study. We consider the sparse setting in Gaussian case with sample sizes being $100, 200, 300, 400$, respectively. For various sample size settings, we conduct 100 runs and compute $(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)^\top\Omega_n^2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)$ for each run. To reduce computational cost, we fix $L = 13$ in the estimation. Figure 1 shows the Q-Q plot of $(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)^\top\Omega_n^2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)$ for each sample size. We can find that $(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)^\top\Omega_n^2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)$ is getting closer to Gaussian distribution with the increase of sample size, which is consistent with the result in Theorem 3.

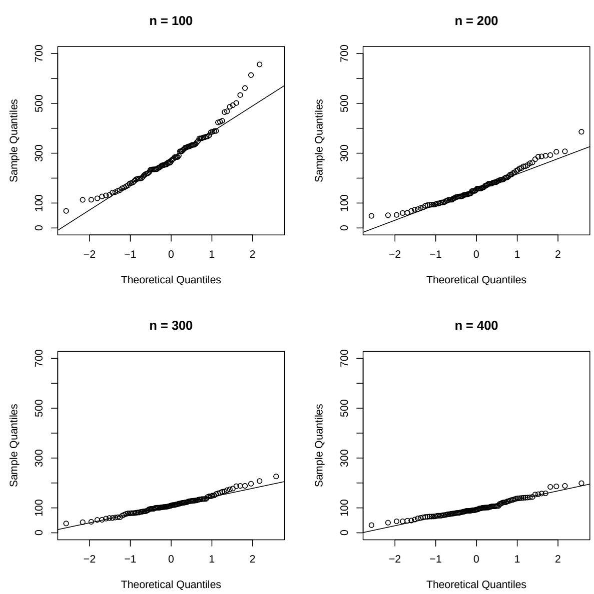

Figure 1: Q-Q plot of $(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)^\top\Omega_n^2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)$ for the sparse setting in Gaussian case with sample sizes being $100, 200, 300, 400$, respectively.

References

Cao, H., D. Zeng, and J. P. Fine (2015). Regression analysis of sparse asynchronous longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(4), 755–776.

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application.* Academic press.

Li, T., T. Li, Z. Zhu, and H. Zhu (2020). Regression analysis of asynchronous longitudinal functional and scalar data. *Journal of the American Statistical Association*, 1–15.

Lin, Z., J. Cao, L. Wang, and H. Wang (2017). Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics 26*(2), 306–318.

Zhong, R., S. Liu, H. Li, and J. Zhang (2021). Sparse logistic functional principal component analysis for binary data. *arXiv preprint arXiv:2109.08009*.