SEMIPARAMETRIC PARTIAL LINEAR QUANTILE REGRESSION OF LONGITUDINAL DATA WITH TIME VARYING COEFFICIENTS AND INFORMATIVE OBSERVATION TIMES

Xuerong Chen[1], Jianguo Sun[2] and Lei Liu[3]

[1] *School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China*

[2] *Department of Statistics, University of Missouri, Columbia, USA*

[3] *Department of Preventive Medicine and Robert H. Lurie Cancer Center, Northwestern University, Chicago, USA*

Supplementary Material

Appendix I: Proof of the Asymptotic Properties.

In this appendix, we will first define some notation and then sketch the proofs for the three theorems given in the main text. Let $\mathbf{X}_1(t)_{n\times p_1} = (X_{11}(t), \cdots, X_{1n}(t))^T$, $\mathbf{Z}(t)_{n\times(p_2+p_3)} = (Z_1(t), \cdots, Z_n(t))^T$ and for a given t,

$$\mathbf{f} = \text{diag}[f_{\epsilon_1}(.\,|X_1(t), \mathcal{F}_{1t}), \cdots, f_{\epsilon_n}(.\,|X_n(t), \mathcal{F}_{nt})]\,,$$

where $f_{\epsilon_i}(.\,|X_i(t), \mathcal{F}_{it})$ denotes the conditional density function of $\epsilon_i(t)$. Define the varying coefficient functional space $\mathcal{S}$ as

$$\mathcal{S} \;=\; \{\, s(x,t) = x_1^T h(t) \equiv \sum_{j=1}^{p_1} x_{1j}(t) h_j(t) \;:\; \sum_{j=1}^{p_1} E\{x_j^2(t) h_j^2(t)\} < \infty,\, x_{1j}(t), h_j(t) \in \mathcal{H}_{rj} \,\}\,.$$

Furthermore, let $\zeta(X_{1i}, t) = X_{1i}^T(t) h(t) \in \mathcal{S}$ and define

$$\zeta_k^*(X_{1i}, t) = \arg\inf_{\zeta\in\mathcal{S}} \sum_{i=1}^n E\bigg\{[Z_{ik}(t) - \zeta(X_{1i}, t)] f_{\epsilon_i}(.\,|X_i(t), \mathcal{F}_{it})[Z_{ik}(t) - \zeta(X_{1i}, t)]\bigg\}$$

and $\omega_k(X_{1i}, t) = E(Z_{ik}(t)|X_{1i})$, where $Z_{ik}(t)$ denotes the kth coordinate of $Z_i(t)$, $k = 1, \cdots, p_2 + p_3$.

Note that

$$\sum_{i=1}^n E\bigg\{[Z_{ik}(t) - \zeta(X_{1i}, t)] f_{\epsilon_i}(.\,|X_i(t), \mathcal{F}_{it})[Z_{ik}(t) - \zeta(X_{1i}, t)]\bigg\}$$

$$= \sum_{i=1}^n E\bigg\{[Z_{ik}(t) - \omega_k(X_{1i}, t)] f_{\epsilon_i}(.\,|X_i(t), \mathcal{F}_{it})[Z_{ik}(t) - \omega_k(X_{1i}, t)]\bigg\}$$

$$+E\bigg\{[\omega_k(X_{1i},t)-\zeta(X_{1i},t)]f_{\epsilon_i}(.\,|X_i(t),\mathcal{F}_{it})[\omega_k(X_{1i},t)-\zeta(X_{1i},t)]\bigg\}.$$

It follows that the $\zeta_k^*(X_{1i},t)$'s are the projections of the $\omega_k(X_{1i},t)$'s onto the varying coefficient functional space $\mathcal{S}$. That is, $\zeta_k^*(X_{1i},t)\in\mathcal{S}$ and it is the function whose distance with $\omega_k(X_{1i},t)$ is shortest among all the functions in $\mathcal{S}$. Let $\zeta^*(X_{1i},t)=(\zeta_1^*(X_{1i},t),\cdots,\zeta_{(p_2+p_3)}^*(X_{1i},t))$. Then $Z_i^{**}(t)=Z_i(t)-\zeta^*(X_{1i},t)$ is orthogonal to the varying coefficient space $\mathcal{S}$ and hence the inner product $E\{Z_i^{**}(t)X_{1i}^T(t)\}$ between $Z_i^{**}(t)$ and $X_{1i}(t)$ is equal to 0. That is, $\zeta_k^*(X_{1i},t)$ have no effect on the asymptotic distribution of $\widehat{\theta}$.

The following notation is needed for the asymptotic variance of the proposed QR estimators and penalized estimators for parametric coefficients. Define

$$P_n^* \;=\; \frac{1}{n}\sum_{i=1}^n\bigg\{\int_0^{t_0}\bigg[\frac{\bar{s}^{(2)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)}-\frac{\bar{s}^{(1)}(t;\gamma_0)s^{(1)}(t;\gamma_0)}{(s^{(0)}(t;\gamma_0))^2}\bigg]dN_i(t)\bigg\},$$

$$\Omega_n=\frac{1}{n}\sum_{i=1}^n\bigg\{\int_0^{t_0}\bigg[\frac{s^{(2)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)}-\bar{x}^{\otimes 2}(t;\gamma_0)\bigg]dN_i(t)\bigg\},$$

where

$$s^{(0)}(t;\gamma)=E\big\{\xi_i(t)\exp(\gamma^TX_i(t))\big\}\,,\ \bar{s}^{(1)}(t;\gamma)=E\bigg\{\xi_i(t)Z_i^{**}(t)\exp(\gamma^TX_i(t))\bigg\},$$

$$\bar{z}^{**}(t;\gamma)=\frac{\bar{s}^{(1)}(t;\gamma)}{s^{(0)}(t;\gamma)}\,,\ M_i(t)=N_i(t)-\int_0^t\xi_i(s)\exp(\gamma^TX_i(s))d\Lambda_0(t)\,,$$

$$\bar{s}^{(2)}(t;\gamma)=E\bigg\{\xi_i(t)Z_i^{**}(t)X_i^T(t)\exp(\gamma^TX_i(t))\bigg\},$$

and

$$h_i^{**}(\alpha,\theta)=\int_0^{t_0}Z_i^{**}(t)\big[I(\epsilon_i(t)\le 0)-\tau\big]dN_i(t)\,.$$

Furthermore, we define some notation which are useful in the derivation of the asymptotic properties of the proposed estimation and model selection methods. Define

$$\begin{aligned}\Pi(t) &= \tilde{\mathbf{X}}_1(t)[\tilde{\mathbf{X}}_1^T(t)\mathbf{f}\tilde{\mathbf{X}}_1(t)]^{-1}\tilde{\mathbf{X}}_1^T(t)\mathbf{f},\quad \mathbf{Z}^*(t)=[I-\Pi(t)]\mathbf{Z}(t),\quad H_n^2(t)=\tilde{\mathbf{X}}^T(t)\mathbf{f}\tilde{\mathbf{X}}(t),\\ \mathbf{Z}^*(t) &= (Z_1^*(t),\cdots,Z_n^*(t))^T, X_{1i}^*(t)=H_n^{-1}(t)\tilde{X}_{1i}(t),\quad \varsigma=H_n(t)\vartheta+H_n^{-1}(t)\tilde{\mathbf{X}}_1^T(t)\mathbf{f}\mathbf{Z}(t)\theta,\end{aligned}$$

$$V_n=\frac{1}{n}\sum_{i=1}^nE\bigg\{h_i(\alpha_0,\theta_0)+\tau\int_0^{t_0}\Big[Z_i^*(t)-\bar{z}^*(t;\gamma_0)\Big]dM_i(t)-\tau P_n\Omega_n^{-1}\int_0^{t_0}\Big[X_i(t)-\bar{x}(t;\gamma_0)\Big]dM_i(t)\bigg\}^{\otimes 2},$$

$$P_n=\frac{1}{n}\sum_{i=1}^n\bigg\{\int_0^{t_0}\bigg[\frac{\tilde{s}^{(2)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)}-\frac{\tilde{s}^{(1)}(t;\gamma_0)s^{(1)}(t;\gamma_0)}{(s^{(0)}(t;\gamma_0))^2}\bigg]dN_i(t)\bigg\},h_i(\alpha,\theta)=\int_0^{t_0}Z_i^*(t)\big[I(\epsilon_i(t)\le 0)-\tau\big]dN_i(t),$$

$$\Omega_n=\frac{1}{n}\sum_{i=1}^n\bigg\{\int_0^{t_0}\bigg[\frac{s^{(2)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)}-\bar{x}^{\otimes 2}(t;\gamma_0)\bigg]dN_i(t)\bigg\},A_n=\frac{1}{n}\sum_{i=1}^nE\bigg\{\int_0^{t_0}Z_i^*(t)Z_i^{*T}(t)f_\epsilon(0|X_i(t),\mathcal{F}_{it})dN_i(t)\bigg\},$$

where $\bar{z}^*(t;\gamma)=\tilde{s}^{(1)}(t;\gamma)/s^{(0)}(t;\gamma)$,$\tilde{s}^{(k)}(t;\gamma)$ denotes the limit of $\tilde{S}^{(k)}(t;\gamma)$, $k=1,2$,

$$\tilde{S}^{(1)}(t;\gamma)=\frac{1}{n}\sum_{i=1}^n\xi_i(t)Z_i^*(t)\exp(\gamma^TX_i(t)),\tilde{S}^{(2)}(t;\gamma)=\frac{1}{n}\sum_{i=1}^n\xi_i(t)Z_i^*(t)X_i^T(t)\exp(\gamma^TX_i(t)).$$

Also let $\bar{A}$ denote the limit of $\bar{A}_n$.

Moreover, let $N_{[\,]}(\varepsilon, \mathcal{F}, \rho)$ and $N(\varepsilon, \mathcal{F}, \rho)$ be the bracketing number and covering number with respect to metric or semi-metric ρ of a function class $\mathcal{F}$. Let $\lesssim$ to indicate the function on its left-hand side is bounded by a positive constant times the function on its right-hand side.

Under condition (A_1) and by the Corollary 6.21 of Schumaker (1981), there exists a spline approximation $\alpha_{nj0}(t) = B_n^T(t)\vartheta_{j0}$ to $\alpha_{j0}(t)$ such that

$$\sup_{t\in[0,t_0]} \big|\alpha_{j0}(t) - \alpha_{nj0}(t)\big| = O(k_n^{-r}), j = 1, 2, \cdots, p_1, \tag{12}$$

where $\alpha_{nj0}(t) = B_n^T(t)\vartheta_{j0}$. Set $\alpha_{n0}(t) = (\alpha_{n10}(t), \cdots, \alpha_{np_10}(t))^T, \eta_{n0} = (\alpha_{n0}, \theta_0)$.

To give the proofs, we need some more notation. Let $\epsilon_{ni}(\varsigma, \theta, t) = Y_i(t) - \varsigma^T X_{1i}^*(t) - \theta^T Z_i^*(t)$ and $\varsigma = H_n(t)\vartheta_0 + H_n^{-1}(t)\tilde{\mathbf{X}}_1^T(t)\mathbf{f}\mathbf{Z}(t)\theta_0$, Define

$$\begin{aligned}
\psi_i(\alpha, \theta, \gamma, \Lambda_0) &= \int_0^{t_0} \Big\{\rho_\tau(\epsilon_i(t))dN_i(t) - \tau\Big[\epsilon_i(t)dN_i(t) - \epsilon_i(t)\xi_i(t)e^{\gamma^T X_i(t)}d\Lambda_0(t)\Big]\Big\},\\
m_i(\varsigma, \theta, \gamma, \Lambda_0) &= \int_0^{t_0} Z_i^*(t)\Big\{\Big[I(\epsilon_{ni}(\varsigma, \theta, t) \le 0) - \tau\Big]dN_i(t) + \tau\Big[dN_i(t) - \xi_i(t)e^{\gamma^T X_i(t)}d\Lambda_0(t)\Big]\Big\},\\
m_i^*(\alpha, \theta, \gamma, \Lambda_0) &= \int_0^{t_0} Z_i^{**}(t)\Big\{\Big[I(\epsilon_i(t) \le 0) - \tau\Big]dN_i(t) + \tau\Big[dN_i(t) - \xi_i(t)e^{\gamma^T X_i(t)}d\Lambda_0(t)\Big]\Big\},
\end{aligned}$$

$$\begin{aligned}
\Psi(\alpha, \theta, \gamma, \Lambda_0) &= E\Psi_n(\alpha, \theta, \gamma, \Lambda_0),\ \mathcal{M}_n^s(\varsigma, \theta, \gamma, \Lambda_0) = \frac{1}{n}\sum_{i=1}^n m_i(\varsigma, \theta, \gamma, \Lambda_0),\\
\mathcal{M}^s(\varsigma, \theta, \gamma, \Lambda_0) &= E\mathcal{M}_n^s(\varsigma, \theta, \gamma, \Lambda_0), \mathcal{M}_n^*(\alpha, \theta, \gamma, \Lambda_0) = \frac{1}{n}\sum_{i=1}^n m_i^*(\alpha, \theta, \gamma, \Lambda_0),
\end{aligned}$$

and

$$\begin{aligned}
I_1 &= \mathcal{M}_n^s(\varsigma, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) - \mathcal{M}^s(\varsigma, \theta, \gamma_0, \Lambda_0(t)) - \mathcal{M}_n^s(\varsigma_0, \theta_0, \widehat{\gamma}, \widehat{\Lambda}_0(t)) + \mathcal{M}^s(\varsigma_0, \theta_0, \gamma_0, \Lambda_0(t)),\\
I_2 &= \mathcal{M}_n^s(\varsigma, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) - \mathcal{M}_n^s(\varsigma, \theta, \gamma_0, \Lambda_0(t)) - \mathcal{M}_n^s(\varsigma_0, \theta_0, \widehat{\gamma}, \widehat{\Lambda}_0(t)) + \mathcal{M}_n^s(\varsigma_0, \theta_0, \gamma_0, \Lambda_0(t)),\\
I_3 &= \mathcal{M}_n^s(\varsigma, \theta, \gamma_0, \Lambda_0) - \mathcal{M}^s(\varsigma, \theta, \gamma_0, \Lambda_0(t)) - \mathcal{M}_n^s(\varsigma_0, \theta_0, \gamma_0, \Lambda_0(t)) + \mathcal{M}^s(\varsigma_0, \theta_0, \gamma_0, \Lambda_0(t)).
\end{aligned}$$

Now we are ready to describe the proofs.

Proof of Theorem 1. By Corollary 3.2.3 of Van der Vaart and Wellner (1996), we just only need to verity conditions of it hold, then we have $\rho(\widehat{\eta}_n - \eta_0) = o_p(1)$, where $\widehat{\eta}_n = (\widehat{\alpha}_n, \widehat{\theta})$ is the sieve estimator, and $\eta_0 = (\alpha_0, \theta_0)$ is the true parameter. Let's verify the first condition of it hold. For all $\varepsilon > 0$, we have

$$\begin{aligned}
&\inf_{\eta_n\in\Theta_n, \rho(\eta_n, \eta_{n0})\ge\varepsilon} \Psi(\alpha_n, \theta, \gamma_0, \Lambda_0) - \Psi(\alpha_0, \theta_0, \gamma_0, \Lambda_0)\\
=\ &\inf_{\eta_n\in\Theta_n, \rho(\eta_n, \eta_{n0})\ge\varepsilon} [\Psi(\alpha_n, \theta, \gamma_0, \Lambda_0) - \Psi(\alpha_{n0}, \theta_0, \gamma_0, \Lambda_0)] + [\Psi(\alpha_{n0}, \theta_0, \gamma_0, \Lambda_0) - \Psi(\alpha_0, \theta_0, \gamma_0, \Lambda_0)]\\
=:\ &J_1 + J_2.
\end{aligned}$$

Denote $\epsilon_{n0i}(t)) = Y_i(t) - \alpha_{n0}^T(t)X_{1i}(t) - \theta_0^T Z_i(t)$, then by (2), independence between censoring time C_i and $N_i^*(t), Y_i(t)$ after given $X_i(t), \mathcal{F}_{it}$, as well as the identity $\rho_\tau(x - y) - \rho_\tau(x) =$

$y\{I(x \le 0) - \tau\} + \int_0^y \{I(x \le z) - I(x \le 0)\}dz$, denote $\Delta_i(t) = (\eta_n - \eta_{n0})(X_{1i}^T(t), Z_i^T(t))^T$, we have

$$\begin{aligned} J_1 &= \inf_{\eta_n \in \Theta_n, \rho(\eta_n, \eta_{n0}) \ge \varepsilon} \frac{1}{n}\sum_{i=1}^n E\bigg\{\int_0^{t_0}\Big[\rho_\tau(\epsilon_{ni}(t)) - \rho_\tau(\epsilon_{n0i}(t))\Big]dN_i(t)\bigg\} \\ &= \inf_{\eta_n \in \Theta_n, \rho(\eta_n, \eta_{n0}) \ge \varepsilon} \int_0^{t_0} E\bigg\{\Big[(\eta_n - \eta_{n0})(X_{1i}^T(t), Z_i^T(t))^T\big(I(\epsilon_{n0i}(t) \le 0) - \tau\big)\Big] + \\ &\quad \int_0^{\Delta_i(t))}\Big[I(\epsilon_{n0i}(t) \le a) - I(\epsilon_{n0i}(t) \le 0)\Big]da\bigg\}dN_i(t) \\ &= \inf_{\eta_n \in \Theta_n, \rho(\eta_n, \eta_{n0}) \ge \varepsilon} \frac{1}{2}\frac{1}{n}\sum_{i=1}^n E\bigg\{\int_0^{t_0}(X_{1i}^T(t), Z_i^T(t))^{T\otimes 2} f_{\epsilon_i}(\Delta_i(t)|X_i(t), \mathcal{F}_{it})dN_i(t)\bigg\} \\ &\quad \times \rho(\eta_n, \eta_{n0})^2 \ge C_1\varepsilon^2 > 0. \end{aligned}$$

Similarly, let $\epsilon_{0i}(t) = Y_i(t) - \alpha_0^T(t)X_{1i}(t) - \theta_0^T Z_i(t)$, then

$$J_2 = \frac{1}{2}\frac{1}{n}\sum_{i=1}^n E\bigg\{\int_0^{t_0} X_{1i}^{\otimes 2}(t) f_{\epsilon_i}(0|X_i(t), \mathcal{F}_{it})dN_i(t)\bigg\}\|\alpha_{n0} - \alpha_0\|_{\mathcal{H}_r}^2.$$

Hence, $\inf_{\eta_n \in \Theta_n, \rho(\eta_n, \eta_{n0}) \ge \varepsilon} \Psi(\alpha_n, \theta, \gamma_0, \Lambda_0) - \Psi(\alpha_0, \theta_0, \gamma_0, \Lambda_0) > 0$. Thus, first condition of Corollary 3.2.3 of Van der Vaart and Wellner (1996) is satisfied. Let us verify the last condition $J_3 =: \lim_n \sup_{\eta_n \in \Theta_n} |\Psi_n(\alpha_n, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) - \Psi(\alpha, \theta, \gamma_0, \Lambda_0)| = o_p(1)$ also hold. Note that

$$\begin{aligned} J_3 &\le \lim_n \sup_{\eta_n \in \Theta_n} |\Psi_n(\alpha_n, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) - \Psi_n(\alpha_n, \theta, \gamma_0, \Lambda_0)| \\ &\quad + \lim_n \sup_{\eta_n \in \Theta_n} |\Psi_n(\alpha_n, \theta, \gamma_0, \Lambda_0) - \Psi(\alpha_n, \theta, \gamma_0, \Lambda_0)| \\ &\quad + \lim_n \sup_{\eta_n \in \Theta_n} |\Psi(\alpha_n, \theta, \gamma_0, \Lambda_0) - \Psi(\alpha, \theta, \gamma_0, \Lambda_0)| \\ &=: J_{31} + J_{32} + J_{33}, \end{aligned}$$

Hence, it's sufficient to prove that $J_{3k} = o_p(1), k = 1, 2, 3$. For J_{31}, we have

$$J_{31} = \lim_n \sup_{\eta_n \in \Theta_n} \tau\bigg|\frac{1}{n}\sum_{i=1}^n\int_0^{t_0}\epsilon_{ni}(t)\xi_i(t)e^{\widehat{\gamma}^T X_i(t)}d\widehat{\Lambda}_0(t) - \frac{1}{n}\sum_{i=1}^n\int_0^{t_0}\epsilon_{ni}(t)\xi_i(t)e^{\gamma_0^T X_i(t)}d\Lambda_0(t)\bigg|.$$

By Fatou lemma, Martingale Central Limit Theorem, the Continuous Mapping Theorem and $\|\widehat{\gamma} - \gamma_0\| = O_p(n^{-\frac{1}{2}}), \widehat{\Lambda}_0(t) - \Lambda_0(t) = O_p(n^{-\frac{1}{2}})$(see Lin et al. (2000)), it easy to obtain $J_{31} = O_p(n^{-\frac{1}{2}}) = o_p(1)$.

In order to prove $J_{32} = o_p(1)$, we just need to verify the class $\mathcal{E}_2 = \{\psi_i(\alpha_n, \theta, \gamma, \Lambda_0), \eta_n \in \Theta_n\}$ is Euclidean class for it's envelope function $\max\{\sup_{\eta_n \in \Theta_n, \psi_i \ne 0} \psi_i(\alpha_n, \theta, \gamma, \Lambda_0), 0\}$. By Euclidean properties of classes $\mathcal{E}_1 = \{\xi_i(t)\epsilon_{ni}(t)\exp(\gamma_0^T X_i(t)), \eta_n \in \Theta_n, t \in [0, t_0]\}, \mathcal{E}_{11} = \{\epsilon_i(\alpha_n, \theta), \alpha_n \in \mathcal{H}_r^n\}, \mathcal{E}_{12} = \{\epsilon_i(\alpha_n, \theta), \theta \in \mathcal{B}\}, \{I(\epsilon_{in}(t) \le 0)\}$, Lemma 5 in Sherman (1994) and Lemma 2.14 in Pakes and Pollard (1989), it is easy to see that classes $\mathcal{E}_2$ is Euclidean class. Hence, we have $J_{32} = o_p(1)$ immediately.

It is easy to see that $J_{33} = O(k_n^{-2r}) = o_p(1)$ because

$$
\begin{aligned}
J_{33} &= \lim_n \sup_{\eta_n \in \Theta_n} \frac{1}{n} \sum_{i=1}^n \left| E\left\{ \int_0^{t_0} \left[\rho_\tau(\epsilon_{ni}(t)) - \rho_\tau(\epsilon_i(t)) \right] dN_i(t) \right\} \right| \\
&= \lim_n \sup_{\eta_n \in \Theta_n} \frac{1}{2n} \sum_{i=1}^n E\left\{ \int_0^{t_0} X_{1i}^{\otimes 2}(t) f_{Y_i|X_i,\mathcal{F}_{it}}(\alpha^T(t) X_{1i}(t) + \theta^T Z_i(t) | X_i(t), \mathcal{F}_{it}) dN_i(t) \right\} \|\alpha_n - \alpha\|_{\mathcal{H}_r}^2.
\end{aligned}
$$

Thus, all conditions of Corollary 3.2.3 of Van der Vaart and Wellner (1996) are satisfied, so we have $\rho(\widehat{\eta}_n, \eta_0) = o_p(1)$.

In the following part, we verify the condition of Theorem 3.2.5 of Van der Vaart and Wellner (1996) to derive the convergence rate. Firstly, in the proof of consistency part, we have already shown that $\Psi(\alpha_n, \theta, \gamma_0, \Lambda_0) - \Psi(\alpha_0, \theta_0, \gamma_0, \Lambda_0) \geq C\rho^2(\eta_n, \eta_0)$. Consider the function classes

$$
\begin{aligned}
\mathcal{E}_{n,\delta} &= \{\psi_i(\alpha_n, \theta, \gamma_0, \Lambda_0) - \psi_i(\alpha_{n0}, \theta_0, \gamma_0, \Lambda_0), \rho(\eta, \eta_{n0}) \leq \delta\}, \\
\mathcal{E}_{n,\delta}^1 &= \{\psi_i(\alpha_n, \theta, \gamma_0, \Lambda_0) - \psi_i(\alpha_{n0}, \theta_0, \gamma_0, \Lambda_0), \|\alpha_n - \alpha_{n0}\|_{\mathcal{H}_r^n} \leq \delta\}, \\
\mathcal{E}_{n,\delta}^2 &= \{\psi_i(\alpha_n, \theta, \gamma_0, \Lambda_0) - \psi_i(\alpha_{n0}, \theta_0, \gamma_0, \Lambda_0), \|\theta - \theta_0\| \leq \delta\}, \\
\mathcal{E}_{31} &= \{I(\epsilon_i(\alpha_n, \theta)(t) \leq 0), \|\alpha_n - \alpha_{n0}\|_{\mathcal{H}_r^n} \leq \delta\}, \\
\mathcal{E}_3 &= \{\epsilon_i(\alpha_n, \theta)(t) I(\epsilon_i(\alpha_n, \theta)(t) \leq 0), \|\alpha_n - \alpha_{n0}\|_{\mathcal{H}_r^n} \leq \delta\}, \\
\mathcal{E}_{41} &= \{I(\epsilon_i(\alpha_n, \theta)(t) \leq 0), \|\theta - \theta_0\| \leq \delta\}, \\
\mathcal{E}_4 &= \{\epsilon_i(\alpha_n, \theta)(t) I(\epsilon_i(\alpha_n, \theta)(t) \leq 0), \|\theta - \theta_0\| \leq \delta\}.
\end{aligned}
$$

Similar to Zhang et al. (2010), we can obtain that $N_{[]}(\varepsilon, \mathcal{E}_{11}, \|.\|_{\mathcal{H}_r}) \lesssim (\delta/\varepsilon)^{p_{k_n}}$, $N_{[]}(\varepsilon, \mathcal{E}_{12}, \|.\|) \lesssim (\delta/\varepsilon)^{(p_2+p_3)}$, similarly, we could derive that $N_{[]}(\varepsilon, \mathcal{E}_{31}, \|.\|_{\mathcal{H}_r}) \lesssim (\delta/\varepsilon)^{p_{k_n}}$, $N_{[]}(\varepsilon, \mathcal{E}_{32}, \|.\|) \lesssim (\delta/\varepsilon)^{(p_2+p_3)}$, hence, it is easy to get that $N_{[]}(\varepsilon, \mathcal{E}_3, \|.\|_{\mathcal{H}_r}) \lesssim (\delta/\varepsilon)^{2p_{k_n}}$, $N_{[]}(\varepsilon, \mathcal{E}_4, \|.\|) \lesssim (\delta/\varepsilon)^{2(p_2+p_3)}$. By the Lemma 9.25 of Kosorok (2008) and the monotone property of $N(t), \Lambda_0(t)$ it's not difficult to get that $\log N_{[]}(\varepsilon, \mathcal{E}_{n,\delta}^1, \|.\|_{\mathcal{H}_r}) \lesssim p_{k_n} \log(\delta/\varepsilon)$, $\log N_{[]}(\varepsilon, \mathcal{E}_{n,\delta}^2, \|.\|) \lesssim (p_2 + p_3) \log(\delta/\varepsilon)$. And this lead to $\log N_{[]}(\varepsilon, \mathcal{E}_{n,\delta}, \rho) \lesssim [p_{k_n} + (p_2 + p_3)] \log(\delta/\varepsilon) \asymp p_{k_n} \log(\delta/\varepsilon)$, in which $\asymp$ means both sides of it have same order. Hence, the bracketing integral $J_{[]}(\delta_0, \mathcal{E}_{n,\delta}, \rho)$ of function class $\mathcal{E}_{n,\delta}$ satisfies

$$
J_{[]}(\delta_0, \mathcal{E}_{n,\delta}, \rho) = \int_0^\delta \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{E}_{n,\delta}, \rho)} d\varepsilon \lesssim \int_0^\delta \sqrt{1 + p_{k_n} \log(\delta/\varepsilon)} d\varepsilon \lesssim p_{k_n}^{1/2} \delta.
$$

Hence, by Lemma 3.4.2 of Van der Vaart and Wellner (1996), we have

$$
\begin{aligned}
& E\left(\sup_{\rho(\eta_n, \eta_{n0}) < \delta} |\sqrt{n}(\Psi_n - \Psi)(\alpha_n, \theta, \gamma_0, \Lambda_0) - \sqrt{n}(\Psi_n - \Psi)(\alpha_{n0}, \theta_0, \gamma_0, \Lambda_0)| \right) \\
\lesssim \;& J_{[]}(\delta, \mathcal{E}_{n,\delta}, \rho) \left(1 + \frac{J_{[]}(\delta, \mathcal{E}_{n,\delta}, \rho)}{\delta^2 \sqrt{n}} M \right) \lesssim p_{k_n}^{1/2} \delta \left(1 + \frac{p_{k_n}^{1/2} \delta}{\delta^2 \sqrt{n}} M \right) = O(p_{k_n}^{1/2} \delta).
\end{aligned}
$$

Moreover, denote $J_4 =: \Psi_n(\alpha_n, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) - \Psi_n(\alpha_n, \theta, \gamma_0, \Lambda_0)] - \sqrt{n}[\Psi_n(\alpha_{n0}, \theta_0, \widehat{\gamma}, \widehat{\Lambda}_0) - \Psi_n(\alpha_{n0}, \theta, \gamma_0, \Lambda_0)$, by using similar method which used to prove the result about J_{31}, we have

$$\begin{aligned} & E\Big(\sup_{\rho(\eta_n,\eta_{n0})<\delta} \big|\sqrt{n} J_4\big| \Big) \\ = \quad & E\Big(\sup_{\rho(\eta_n,\eta_{n0})<\delta} \Big| \frac{\tau}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{t_0} [\epsilon_{ni}(t) - \epsilon_{n0i}(t)] \xi_i(t) e^{\widehat{\gamma}^T X_i(t)} d\widehat{\Lambda}_0(t) \\ & - \int_0^{t_0} [\epsilon_{ni}(t) - \epsilon_{n0i}(t)] \xi_i(t) e^{\gamma_0^T X_i(t)} d\Lambda_0(t) \Big| \Big) = o(1). \end{aligned}$$

Hence it means that the key function $\phi_n(\delta)$ in Theorem 3.2.5 of Van der Vaart and Wellner (1996) is given by $\phi_n(\delta) = p_{k_n}^{1/2}\delta$, by Theorem 3.2.5 of Van der Vaart and Wellner (1996) and let $r_n = (n/k_n)^{1/2}$, we could get $\rho(\widehat{\eta}_n, \eta_{n0}) = (k_n/n)^{1/2}$. By (12), it is easy to get that $\rho(\widehat{\eta}_n, \eta_0) = (k_n/n)^{1/2} + k_n^{-r}$. When $k_n \asymp n^{1/(2r+1)}$, $\rho(\widehat{\eta}_n, \eta_0) = O_p(n^{-r/(2r+1)})$. This complete the proof of Theorem 1. □

For the proof of Theorem 2, we first give three lemmas needed.

Lemma 1. *Assume that* $(A_1) - (A_4)$ *and* $(A_6) - (A_9)$ *hold,* $k_n \to \infty, nk_n^{-(4r)} \to 0$ *as* $n \to \infty$, *then*

$$J_5 =: \mathcal{M}_n^s(\varsigma, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) - \mathcal{M}_n^*(\alpha, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) = o_p(n^{-1/2}).$$

By Lemma 1, it easy to see that the asymptotic distribution of $\widehat{\theta}$ will not change if $Z_i^*(t)$ are replaced by $Z_i^{**}(t)$ under the assumption of Theorem 2; the latter enjoy independence and are easier to handle mathematically. In order to simplify notation, we simply prove Theorem 2 by treating $Z_i^*(t), i = 1, \cdots, n$ are independent.

Lemma 2. *Assume that* $(A_1) - (A_4)$ *hold,* $k_n \to \infty, k_n^2/n \to 0, nk_n^{-(4r)} \to 0$ *as* $n \to \infty$ *and* $r \geq 1$, *then for all positive values* $\varepsilon_n = O(\sqrt{k_n/n})$,

$$\sup_{\rho(\eta_n,\eta_{n0})\leq\varepsilon_n} \{I_1\} = o_p(n^{-\frac{1}{2}}).$$

Lemma 3. *Assume that* $(A_1) - (A_4)$ *hold, then we have*

$$A_n^* - A_n = o_p(1), \;\; V_n^* - V_n = o_p(1) \;\; \text{and} \; P_n^* - P_n = o_p(1).$$

Proof of Theorem 2. Define

$$\begin{aligned} \Psi_n(\varsigma, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) \quad = \quad & \frac{1}{n} \sum_{i=1}^{n} \int_0^{t_0} \Big\{ \rho_\tau(\epsilon_{ni}(\varsigma, \theta, t)) dN_i(t) - \tau \Big[\epsilon_{ni}(\varsigma, \theta, t) dN_i(t) \\ & - \epsilon_{ni}(\varsigma, \theta, t) \xi_i(t) e^{\widehat{\gamma}^T X_i(t)} d\widehat{\Lambda}_0(t) \Big] \Big\}, \end{aligned}$$

it is easy to see that minimize $\Psi_n(\alpha_n, \theta, \widehat{\gamma}, \widehat{\Lambda}_0)$ respect to α_n, θ is equivalent to minimize $\Psi_n(\varsigma, \theta, \widehat{\gamma}, \widehat{\Lambda}_0)$ respect to ς, θ because $\varsigma^T X_{1i}^*(t) + \theta^T Z_i^*(t) = \alpha_n^T(t) X_{1i}(t) + \theta^T Z_i(t)$. Note that $\mathbf{Z}^*(t)$ is orthogonal to $\tilde{\mathbf{X}}_1(t)$ and hence also orthogonal to $\mathbf{X}_1^*(t)$, in order to get the asymptotic normality of $\widehat{\theta}$, it is sufficient to consider the estimating equation $\mathcal{M}_n^s(\varsigma,\theta,\widehat{\gamma},\widehat{\Lambda}_0)=o_p(n^{-1/2})$, in fact, it is correspond to the sub-gradient of $\Psi_n(\varsigma,\theta,\widehat{\gamma},\widehat{\Lambda}_0)$ respect to θ.

By the orthogonality of $\mathbf{Z}^*(t)$ and $\mathbf{X}_1^*(t)$, let $I_5=:\mathcal{M}^s(\varsigma,\theta,\gamma_0,\Lambda_0)-\mathcal{M}^s(\varsigma_0,\theta_0,\gamma_0,\Lambda_0)$, then for $\rho(\eta_n,\eta_{n0})=o_p(1)$, we also have $\|\theta-\theta_0\|=o_p(1)$, $\|\varsigma-\varsigma_0\|=o_p(1)$, then

$$\begin{aligned}
I_5 &= E\Big\{\int_0^{t_0} Z_i^*(t)\Big[\Big(F_{Y_i|X,\mathcal{F}_{it}}(\varsigma^T X_{1i}^*(t)+\theta^T Z_i^*(t))-F_{Y_i|X_i,\mathcal{F}_{it}}(\varsigma_0^T X_{1i}^*(t)+\theta_0^T Z_i^*(t))\Big)dN_i(t)\Big]\Big\}\\
&= \frac{1}{n}\sum_{i=1}^n E\Big\{\int_0^{t_0} Z_i^*(t)\Big[\Big(f_\epsilon(0|X_i(t),\mathcal{F}_{it})Z_i^{*T}(t)(\theta-\theta_0)\\
&\quad +f_\epsilon(0|X_i(t),\mathcal{F}_{it})X_{1i}^{*T}(t)(\varsigma-\varsigma_0)\Big)dN_i(t)\Big]\Big\}+O_p(\rho^2(\eta_n,\eta_{n0}))\\
&= \frac{1}{n}\sum_{i=1}^n E\Big\{\int_0^{t_0} Z_i^*(t)Z_i^{*T}(t)f_\epsilon(0|X_i(t),\mathcal{F}_{it})dN_i(t)\Big\}(\theta-\theta_0)+O_p(\rho^2(\eta_n,\eta_{n0}))\\
&= A_n(\theta-\theta_0)+O_p(\rho^2(\eta_n,\eta_{n0})).
\end{aligned}$$

Note that when $r\geq 1$, by the convergence rate which given by Theorem 1, we have $\rho^2(\eta_n,\eta_{n0})=o_p(n^{-1/2})$. By Lemma 2, above equation, when $r>1$, we have

$$\begin{aligned}
\mathcal{M}_n^s(\widehat{\varsigma},\widehat{\theta},\widehat{\gamma},\widehat{\Lambda}_0) &= \mathcal{M}_n^s(\varsigma_0,\theta_0,\widehat{\gamma},\widehat{\Lambda}_0)+\mathcal{M}^s(\widehat{\varsigma},\widehat{\theta},\gamma_0,\Lambda_0)-\mathcal{M}^s(\varsigma_0,\theta_0,\gamma_0,\Lambda_0)\\
&\quad +\mathcal{M}_n^s(\widehat{\varsigma},\widehat{\theta},\widehat{\gamma},\widehat{\Lambda}_0)-\mathcal{M}^s(\widehat{\varsigma},\widehat{\theta},\gamma_0,\Lambda_0(t))-\mathcal{M}_n^s(\varsigma_0,\theta_0,\widehat{\gamma},\widehat{\Lambda}_0(t))\\
&\quad +\mathcal{M}^s(\varsigma_0,\theta_0,\gamma_0,\Lambda_0(t))\\
&= \mathcal{M}_n^s(\varsigma_0,\theta_0,\widehat{\gamma},\widehat{\Lambda}_0)+A_n(\widehat{\theta}-\theta_0)+o_p(n^{-1/2}),
\end{aligned}$$

it is easy to get that

$$\theta-\theta_0=-A_n^{-1}\mathcal{M}_n^s(\varsigma_0,\theta_0,\widehat{\gamma},\widehat{\Lambda}_0)+o_p(n^{-1/2}).$$

On the other hand, it easy to get that $\frac{1}{n}\sum_{i=1}^n[h_i(\alpha_{n0},\theta_0)-h_i(\alpha_0,\theta_0)]=o_p(n^{-1/2})$. By (A_4), we have

$$\begin{aligned}
&\mathcal{M}_n^s(\varsigma_0,\theta_0,\widehat{\gamma},\widehat{\Lambda}_0)\\
&= \frac{1}{n}\sum_{i=1}^n h_i(\alpha_{n0},\theta_0)+\tau\frac{1}{n}\sum_{i=1}^n\int_0^{t_0}\Big[Z_i^*(t)-\frac{\tilde{S}^{(1)}(t;\widehat{\gamma})}{S^{(0)}(t;\widehat{\gamma})}\Big]dN_i(t)\\
&= \frac{1}{n}\sum_{i=1}^n h_i(\alpha_0,\theta_0)+\tau\frac{1}{n}\sum_{i=1}^n\int_0^{t_0}\Big\{Z_i^*(t)-\bar{Z}^*(t;\gamma_0)\Big\}dM_i(t)\\
&\quad -\tau P_n\Omega_n^{-1}\frac{1}{n}\sum_{i=1}^n\int_0^{t_0}\Big\{X_i(t)-\bar{X}(t;\gamma_0)\Big\}dM_i(t)+o_p(n^{-1/2}).
\end{aligned}$$

By functional central limit theorem, it is easy to get that

$$\sqrt{n}\mathcal{M}_n^s(\alpha_{n0},\theta_0,\widehat{\gamma},\widehat{\Lambda}_0)\xrightarrow{D}N(0,V),$$

Hence, we have $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{D} N(0, A^{-1}VA^{-1})$. This completes the proof of the theorem. □

For the proof of Theorem 3, we need more notation and two more lemmas.

Let $\bar{\vartheta}_0^{(1)}, \bar{\vartheta}_{*0}^{(1)}, \bar{\vartheta}_{*0}^{(2)}$ denote the true values of

$$\bar{\vartheta}^{(1)} = (\bar{\vartheta}_{11}, \cdots, \bar{\vartheta}_{p1})^T,\ \bar{\vartheta}_*^{(1)} = (\bar{\vartheta}_{1*}^T, \cdots, \bar{\vartheta}_{p_1*}^T)^T,\ \bar{\vartheta}_*^{(2)} = (\bar{\vartheta}_{p_1+1*}^T, \cdots, \bar{\vartheta}_{p*}^T)^T.$$

It is easy to see that $\bar{\vartheta}_{*0}^{(2)} = \mathbf{0}$, where $\mathbf{0}$ is a vector which element are 0 and has same dimension with $\bar{\vartheta}_*^{(2)}$. Denote $\bar{\epsilon}_{n0i}(t) = Y_i(t) - \bar{\vartheta}_0^T \mathcal{X}_i(t)$,

$$\begin{aligned} U_n(\mathbf{u}) &= \sum_{i=1}^n \int_0^{t_0} \Big\{ \rho_\tau\big(\bar{\epsilon}_{n0i}(t) - \mathbf{u}^T\mathcal{X}_i(t)/\sqrt{n/k_n}\big) dN_i(t) - \tau\big(\bar{\epsilon}_{n0i}(t) - \mathbf{u}^T\mathcal{X}_i(t)/\sqrt{n/k_n}\big) d\widehat{M}_i(t) \Big\} \\ &\quad - \sum_{i=1}^n \int_0^{t_0} \Big\{ \rho_\tau\big(\bar{\epsilon}_{n0i}(t)\big) dN_i(t) - \tau\big(\bar{\epsilon}_{n0i}(t)\big) d\widehat{M}_i(t) \Big\}. \end{aligned}$$

Lemma 4. *Assume that $(A_3)(\tilde{A}_5), (A_6)-(A_9)$ hold, then for any fixed $\mathbf{u}$, we have*

$$U_n(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T \bar{A}_n \mathbf{u} + \mathbf{u}^T G_n^s(\vartheta_0, \theta_0, \widehat{\gamma}, \widehat{\Lambda}_0) + o_p(1).$$

Lemma 5. *(Sparsity) Assume that $\sqrt{n/k_n}\lambda \to \infty$ as $\lambda \to \infty$, then for any $\bar{\vartheta}^{(1)}, \bar{\vartheta}_*^{(1)}$ satisfying $\|\bar{\vartheta}^{(1)} - \bar{\vartheta}_0^{(1)}\| = O_p(\sqrt{k_n/n}), \|\bar{\vartheta}_*^{(1)} - \bar{\vartheta}_{*0}^{(1)}\| = O_p(\sqrt{k_n/n})$ and any constant $\bar{c}$ we have*

$$\Psi_n^P\big((\bar{\vartheta}^{(1)T}, (\bar{\vartheta}_*^{(1)T}, \mathbf{0}^T)^T, \widehat{\gamma}, \widehat{\Lambda}_0\big) = \min_{\|\bar{\vartheta}_*^{(2)}\| \le \bar{c}(k_n/n)^{1/2}} \Psi_n^P\big((\bar{\vartheta}^{(1)T}, (\bar{\vartheta}_*^{(1)T}, \bar{\vartheta}_*^{(2)T}))^T, \widehat{\gamma}, \widehat{\Lambda}_0\big).$$

Proof of Theorem 3. We will first prove the conclusion (b) of Theorem 3.

(I) Proof of (b): Similar to Fan and Li (2001), to prove (b), it is enough to show that for any given $\delta > 0$, there exist a large constant C such that

$$P\Big\{ \inf_{\|\mathbf{u}\|=C} \Psi_n^P\big(\bar{\vartheta}_0 + \mathbf{u}/\sqrt{n/k_n}, \widehat{\gamma}, \widehat{\Lambda}_0\big) > \Psi_n^P\big(\bar{\vartheta}_0, \widehat{\gamma}, \widehat{\Lambda}_0\big) \Big\} \ge 1 - \delta, \tag{13}$$

which implies that with probability $1-\delta$ there exists a local minimum in the ball $\{\bar{\vartheta}_0^T + \mathbf{u}/\sqrt{n/k_n} : \|\mathbf{u}\| \le C\}$, namely there exists a local minimizer such that $\|\widehat{\bar{\vartheta}} - \widehat{\bar{\vartheta}}_0\| = O_p(\sqrt{k_n/n})$. It means that $\rho(\widehat{\eta}_n^P, \eta_0) = O_p((k_n/n)^{1/2} + k_n^{-r})$, which is exactly what we want to prove.

Note that, when n is enough large, $n\sum_{j=1}^{p_1}[p_\lambda(\|\bar{\vartheta}_{j*0} + u_{j*}/\sqrt{n/k_n}\|) - p_\lambda(\|\bar{\vartheta}_{j*0}\|)] = 0$ uniformly in any compact set because SCAD penalty is flat when coefficient lager than $a\lambda_n, \lambda_n \to 0$. By Lemma 4, we have

$$\begin{aligned} &\Psi_n^P\big(\bar{\vartheta}_0 + \mathbf{u}/\sqrt{n/k_n}, \widehat{\gamma}, \widehat{\Lambda}_0\big) - \Psi_n^P\big(\bar{\vartheta}_0, \widehat{\gamma}, \widehat{\Lambda}_0\big) \\ =\ & U_n(\mathbf{u}) + n\sum_{j=1}^p \Big[p_\lambda(\|\bar{\vartheta}_{j*0} + u_{j*}/\sqrt{n/k_n}\|) - p_\lambda(\|\bar{\vartheta}_{j*0}\|) \Big] \\ \ge\ & U_n(\mathbf{u}) + n\sum_{j=1}^{p_1} \Big[p_\lambda(\|\bar{\vartheta}_{j*0} + u_{j*}/\sqrt{n/k_n}\|) - p_\lambda(\|\bar{\vartheta}_{j*0}\|) \Big] \\ =\ & \frac{1}{2}\mathbf{u}^T \bar{A}_n \mathbf{u} + \mathbf{u}^T G_n^s(\vartheta_0, \theta_0, \widehat{\gamma}, \widehat{\Lambda}_0) + o_p(1). \end{aligned}$$

By Convexity Lemma(Lemma 2 of Wu and Liu (2009)), this equation hold uniformly on any compact set. It's easy to see that, when C is sufficiently large, $\Psi_n^P\big(\bar{\vartheta}_0^T + \mathbf{u}/\sqrt{n/k_n}, \widehat{\gamma}, \widehat{\Lambda}_0\big) - \Psi_n^P\big(\bar{\vartheta}_0^T, \widehat{\gamma}, \widehat{\Lambda}_0\big)$ is dominated by the quadratic term on the right hand side of above equation. By condition $(\tilde{A}_5)$, (13) holds. The proof of (b) is completed.

(II) Proof of (a): Similar to Fan and Li (2001), by Lemma 5, (a) holds.

(III) Proof of (c): By (a), it means that $\widehat{\bar{\vartheta}}_{j*} = 0, j = 1, \cdots, p_1$. Therefore, with probability approaching 1, we have

$$\Psi_n^P\big(\widehat{\bar{\vartheta}}, \widehat{\gamma}, \widehat{\Lambda}_0\big) = \Psi_n^P\big((\widehat{\bar{\vartheta}}^{(1)T}, (\widehat{\bar{\vartheta}}_*^{(1)T}, \mathbf{0}^T)^T, \widehat{\gamma}, \widehat{\Lambda}_0\big) = n\Psi_n(\widehat{\alpha}_n^P, \widehat{\theta}^P, \widehat{\gamma}, \widehat{\Lambda}_0) + n\sum_{j=1}^{p_1} p_\lambda(\|\widehat{\bar{\vartheta}}_{j*}\|)\,.$$

Note that the penalty term doesn't depend on $\widehat{\theta}^P$. Hence the asymptotic normality of $\widehat{\theta}^P$ is determined by $\Psi_n(\widehat{\alpha}_n^P, \widehat{\theta}^P, \widehat{\gamma}, \widehat{\Lambda}_0)$. Similar to the proof of Theorem 2, we could obtain the conclusion and this complete the proof of Theorem 3.

Appendix II: Proof of Weak Convergence of $\mathcal{L}(u, x_1, z)$.

Define $I_4 = L_n(\alpha_n, \theta) - L(\alpha_n, \theta) - L_n(\alpha_{n0}, \theta_0) + L(\alpha_{n0}, \theta_0)$,

$$\begin{aligned}
l_i(\alpha_n, \theta) &= \int_0^u I(X_{i1} \le x_1, Z_i \le z)[I(\epsilon_{ni}(t) \le 0) - \tau]\, dN_i(t), L_n(\alpha_n, \theta) = \frac{1}{n}\sum_{i=1}^n l_i(\alpha_n, \theta),\\
L(\alpha_n, \theta) &= EL_n(\alpha_n, \theta), \ddot{S}^{(1)}(t;\gamma) = \frac{1}{n}\sum_{j=1}^n \xi_i(t) I(X_{j1} \le x_1, Z_j \le z)\exp(\gamma^T X_j(t)),\\
\ddot{S}^{(2)}(t;\gamma) &= \frac{1}{n}\sum_{j=1}^n \xi_i(t) I(X_{j1} \le x_1, Z_j \le z) X_j(t)\exp(\gamma^T X_j(t)),\\
s_i(\alpha_{n0}, \theta) &= \int_0^{t_0} X_{i1}(t)[I(\epsilon_{ni}(t) \le 0) - \tau]\, dN_i(t), \dot{S}^{(1)}(t;\gamma) = \frac{1}{n}\sum_{j=1}^n \xi_i(t) X_{j1}(t)\exp(\gamma^T X_j(t)),
\end{aligned}$$

$$\dot{S}^{(2)}(t;\gamma) = \frac{1}{n}\sum_{j=1}^n \xi_i(t) X_{j1}(t) X_j^T(t)\exp(\gamma^T X_j(t)),$$

and

$$\begin{aligned}
\ddot{P}_n &= \frac{1}{n}\sum_{i=1}^n \bigg\{ \int_0^{t_0} \bigg[\frac{\dot{s}^{(2)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)} - \frac{\dot{s}^{(1)}(t;\gamma_0)s^{(1)}(t;\gamma_0)}{(s^{(0)}(t;\gamma_0))^2}\bigg] dN_i(t)\bigg\},\\
\tilde{P}_n &= \frac{1}{n}\sum_{i=1}^n \bigg\{ \int_0^{t_0} \bigg[\frac{\ddot{s}^{(2)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)} - \frac{\ddot{s}^{(1)}(t;\gamma_0)s^{(1)}(t;\gamma_0)}{(s^{(0)}(t;\gamma_0))^2}\bigg] dN_i(t)\bigg\},\\
B_{n1} &= \frac{1}{n}\sum_{i=1}^n E\bigg\{ \int_0^{t_0} I(X_{i1} \le x_1, Z_i \le z) X_{1i}^T(t) f_{\epsilon_i}(0|X_i(t), \mathcal{F}_{it}) dN_i(t)\bigg\},\\
B_{n2} &= \frac{1}{n}\sum_{i=1}^n E\bigg\{ \int_0^{t_0} I(X_{i1} \le x_1, Z_i \le z) Z_i^T(t) f_{\epsilon_i}(0|X_i(t), \mathcal{F}_{it}) dN_i(t)\bigg\},\\
B_{n3} &= \frac{1}{n}\sum_{i=1}^n E\bigg\{ \int_0^{t_0} X_{1i}^{\otimes 2}(t) f_{\epsilon_i}(0|X_i(t), \mathcal{F}_{it}) dN_i(t)\bigg\},
\end{aligned}$$

where $\dot{s}^{(k)}(t;\gamma), \ddot{s}^{(k)}(t;\gamma)$ are the limits of $\ddot{S}^{(k)}(t;\gamma)$ and $\ddot{S}^{(k)}(t;\gamma)$ respectively, for $k = 1, 2$. We also need the following lemma.

Lemma 6. *Assume that* $(A_1) - (A_4)$ *hold,* $k_n \to \infty, k_n^2/n \to 0, nk_n^{-(4r)} \to 0$ *as* $n \to \infty$ *and* $r \geq 1$*, then for all positive values* $\varepsilon_n = O(\sqrt{k_n/n})$*,*

$$\sup_{\rho(\eta_n,\eta_{n0})\leq\varepsilon_n} \{I_4\} = o_p(n^{-\frac{1}{2}}).$$

Proof of weak convergence of $\mathcal{L}(u, x_1, Z)$. By some calculation, Lemma 6 and Taylor expansion, we have

$$\begin{aligned}
& \mathcal{L}_\tau(u, x_1, z) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^u \left\{ I(X_{i1} \leq x_1, Z_i \leq z)[I(\widehat{\epsilon}_{ni}(t) \leq 0) - \tau] + \tau \left[I(X_{i1} \leq x_1, Z_i \leq z) - \frac{\ddot{S}^{(1)}(t;\widehat{\gamma})}{S^{(0)}(t;\widehat{\gamma})} \right] \right\} dN_i(t) \\
= & \sqrt{n}[L_n(\alpha_{n0}, \theta_0) + L(\alpha_n, \theta) - L(\alpha_{n0}, \theta_0)] + \tau \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^u \left[I(X_{i1} \leq x_1, Z_i \leq z) - \frac{\ddot{S}^{(1)}(t;\widehat{\gamma})}{S^{(0)}(t;\widehat{\gamma})} \right] \Bigg\} dN_i(t) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^n [l_i(\alpha_0, \theta_0) - B_{n1} B_{n3}^{-1} s_i(\alpha_0, \theta_0) - B_{n2} A_n^{-1} h_i(\alpha_0, \theta_0)] \\
& + \tau \frac{1}{\sqrt{n}} \sum_{i=1}^n \Bigg\{ \int_0^u \left[I(X_{i1} \leq x_1, Z_i \leq z) - \frac{\ddot{s}^{(1)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)} \right] dM_i(t) \\
& - B_{n1} B_{n3}^{-1} \int_0^{t_0} \left[X_{i1}(t) - \frac{\dot{s}^{(1)}(t;\gamma_0)}{s^{(0)}(t;\gamma_0)} \right] dM_i(t) - B_{n2} A_n^{-1} \int_0^{t_0} \left[Z_i^*(t) - \bar{z}^*(t;\gamma_0) \right] dM_i(t) \Bigg\} \\
& - \tau (\tilde{P}_n - B_{n1} B_{n3}^{-1} \ddot{P}_n - B_{n2} A_n^{-1} P_n) \Omega_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{t_0} \left\{ X_i(t) - \bar{X}(t;\gamma_0) \right\} dM_i(t) + o_p(1) \\
=: & \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{L}_{1i}(u) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{L}_{2i}(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{L}_{3i}(u) + o_p(1).
\end{aligned}$$

By the functional central limit theorem, $\mathcal{L}_\tau(u, x_1, z)$ converges in finite-dimensional distribution to a zero-mean Gaussian process. Since any function of bounded variation can be written as the difference of two increasing functions, the first terms of $\mathcal{L}_{1i}(u), \mathcal{L}_{2i}(u)$ and $\mathcal{L}_{3i}(u)$ are tight. And other remaining terms do not involve u. Hence, $\mathcal{L}_\tau(u, x_1, z)$ is tight and converges weakly to a zero-mean Gaussian process which could be approximated by the zero-mean Gaussian process $\tilde{\mathcal{L}}_\tau(u, x_1, z)$.

References

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348-1360.

Kosorok, M. R.(2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer Press, New York.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711-730.

Pakes, A. and Pollard, D.(1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

Schumaker, L. L. (1981). *Spline Functions: basic Theory.* Wiley, New York.

Sherman, R. P.(1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439-459.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wu Y. and Liu, Y. F. (2009). Variable selection in quantile regression. *Stat. Sinica* **19**, 801-817.

Zhang, Y., Hua, L. and Huang, J. (2010). A Spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scand. J. Statist.* **37**, 338-354.

School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

E-mail: (chenxr522@foxmail.com)

Department of Statistics, University of Missouri, Columbia, USA

E-mail: (sunj@missouri.edu)

Department of Preventive Medicine and Robert H. Lurie Cancer Center, Northwestern University, Chicago, USA

E-mail: (lei.liu@northwestern.edu)