Penalized empirical likelihood inference for sparse additive hazards regression with a diverging number of covariates

Wang, Shanshan; Xiang, Liming

2016

Wang, S., & Xiang, L. Penalized empirical likelihood inference for sparse additive hazards regression with a diverging number of covariates. Statistics and Computing, in press.

https://hdl.handle.net/10356/83404

https://doi.org/10.1007/s11222-016-9690-x

Supplementary Material for Penalized Empirical Likelihood Inference for Sparse Additive Hazards Regression with a Diverging number of covariates

Shanshan Wang[1,2] **and Liming Xiang**[2*]

[1]*School of Economics and Management, Beihang University, Beijing, China*

[2]*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore*

Sketches of Proofs

Here we provide sketches of proofs to Theorems 1-4. The presented lemmas are needed to prove theoretical results in these Theorems.

Lemma 1. *Suppose that Assumptions 1-5 are satisfied and that* $\boldsymbol{\beta} \in E_n$*, where* $E_n = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq Ca_n\}$ *for some constant* $C > 0$ *and* $a_n = (p/n)^{1/2}$*. We have*

(i) $\max_{1\leq i\leq n} \|\boldsymbol{W}_{ni}(\boldsymbol{\beta})\| = o_p(p^{1/2}n^{1/q})$,

(ii) $\left\|n^{-1}\sum_{i=1}^n W_{ni}(\beta)\boldsymbol{W}_{ni}^T(\boldsymbol{\beta}) - \boldsymbol{\Sigma}\right\| = o_p(p^{-1})$,

(iii) $\|n^{-1}\sum_{i=1}^n W_{ni}(\beta)\| = O_p(a_n)$,

(iv) and for the true value $\boldsymbol{\beta}_0$*,* $n^{-1/2}\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \xrightarrow{d} N_q(0, \boldsymbol{G})$*, where* $\boldsymbol{A}_n$ *is a* $q\times p$ *matrix such that* $\boldsymbol{A}_n\boldsymbol{A}_n^T \to \boldsymbol{G}$ *with* $\boldsymbol{G}$ *is* $q \times q$ *positive definite.*

Proof. Let $\boldsymbol{W}_i(\boldsymbol{\beta}) = \int_0^\tau \{\boldsymbol{Z}_i(t) - \boldsymbol{e}(t)\}dM_i(\boldsymbol{\beta}, t)$ and $W_{ij}(\boldsymbol{\beta}_0)$ be its $j-$th component. Under Assumptions 4 and by Lemma 11.2 in Owen (2001), we have $\max_{1\leq i\leq n} |W_{ij}(\boldsymbol{\beta}_0)| = o_p(n^{1/q})$. Then

$$\begin{aligned} \max_{1\leq i\leq n} \|W_i(\beta_0)\| = \max_{1\leq i\leq n} \Big(\sum_{j=1}^p W_{ij}^2(\beta_0)\Big)^{1/2} &= o_p(n^{1/q}p^{1/2}), \qquad \text{(A.1)} \\ \left\|n^{-1}\sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0) - \Sigma\right\| &= o_p(p^{-1}). \end{aligned}$$

* Corresponding author. Tel.: +65 65137451 ; Fax: +65 65158213.

E-mail address: lmxiang@ntu.edu.sg .

For any $\varepsilon > 0$ and under Assumption 5, it follows that

$$\begin{aligned} P\Big\{\|n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i(\boldsymbol{\beta}_0)\boldsymbol{W}_i^T(\boldsymbol{\beta}_0)-\boldsymbol{\Sigma}\| \geq \frac{\varepsilon}{p}\Big\} &\leq \frac{p^2}{n^2\varepsilon^2}E\sum_{j,k}\Big[\sum_{i=1}^{n}\Big\{W_{ij}(\boldsymbol{\beta}_0)W_{ik}(\boldsymbol{\beta}_0)-\sigma_{jk}\Big\}\Big]^2 \\ &= O_p\Big(\frac{p^4}{n}\Big) = O_p\Big(p^{-1}\frac{p^5}{n}\Big) = o_p(1). \end{aligned}$$

Decompose $\boldsymbol{W}_{ni}(\boldsymbol{\beta})$ as

$$\begin{aligned} \boldsymbol{W}_{ni}(\boldsymbol{\beta}) &= \boldsymbol{W}_i(\boldsymbol{\beta}) + \int_0^\tau \Big\{\boldsymbol{Z}_i(t)-\boldsymbol{e}(t)\Big\}Y_i(t)d\Big\{\Lambda_0(t)-\hat{\Lambda}_0(\boldsymbol{\beta},t)\Big\} \\ &\quad + \int_0^\tau \Big\{\boldsymbol{e}(t)-\bar{\boldsymbol{Z}}(t)\Big\}d\hat{M}_i(\boldsymbol{\beta},t) \\ &\stackrel{\text{def}}{=} \boldsymbol{W}_i(\boldsymbol{\beta}) + \boldsymbol{R}_{i1}(\boldsymbol{\beta}) + \boldsymbol{R}_{i2}(\boldsymbol{\beta}). \end{aligned}$$

With the true value $\boldsymbol{\beta}_0$,

$$\begin{aligned} \boldsymbol{R}_{i1}(\boldsymbol{\beta}_0) &= \int_0^\tau \Big\{\boldsymbol{Z}_i(t)-\boldsymbol{e}(t)\Big\}Y_i(t)d\Big\{\Lambda_0(t)-\hat{\Lambda}_0(\boldsymbol{\beta}_0,t)\Big\} \\ &= -\int_0^\tau \frac{Y_i(t)\{\boldsymbol{Z}_i(t)-\boldsymbol{e}(t)\}}{\sum_{j=1}^n Y_j(t)}\sum_{j=1}^{n}dM_j(\boldsymbol{\beta}_0,t) \end{aligned}$$

is a mean zero martingale. By Assumption 1, it holds that $\inf_{t\in[0,\tau]}\sum_{j=1}^n Y_j(t) \stackrel{P}{\to} \infty$ as $n\to\infty$. This simply guarantees that the number of individuals at risk at each time point becomes lager. It is then easy to show that $\sup|\boldsymbol{R}_{i1}(\boldsymbol{\beta}_0)| \stackrel{P}{\to} 0$. Also, by martingale central limiting theorem, $n^{-1}\sum_{j=1}^n Y_j(t)$ approaches a function $y(t)$ as n becomes large with $y(t)$ satisfying $\int_{t\in[0,\tau]} y(t)dt > 0$. We obtain that each component of $R_{i1}(\beta_0)$ is $O_p(n^{-1/2})$, and thus $\|R_{i1}(\beta_0)\| = O_p(a_n)$.

In addition, note that $\boldsymbol{R}_{i2}(\boldsymbol{\beta_0}) = \int_0^\tau\{\boldsymbol{e}(t)-\bar{\boldsymbol{Z}}(t)\}d\hat{M}_i(\boldsymbol{\beta}_0,t)$, $\boldsymbol{e}(t) = s^{(1)}(t)/s^{(0)}(t)$ and $\bar{\boldsymbol{Z}}(t) = \hat{s}^{(1)}(t)/\hat{s}^{(0)}(t)$, where $\hat{s}^{(1)}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)$ and $\hat{s}^{(0)} = n^{-1}\sum_{i=1}^n Y_i(t)$ are estimates of $s^{(1)}(t)$ and $s^{(0)}(t)$, respectively. Thus,

$$\boldsymbol{R}_{i2}(\boldsymbol{\beta}_0) = \int_0^\tau \Big\{\frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{\hat{s}^{(1)}(t)}{\hat{s}^{(0)}(t)}\Big\}d\hat{M}_i(\boldsymbol{\beta}_0,t).$$

By Lemma 5.1 of Zhou (1992) and Assumption 1, it follows that $\sup_{t\in[0,\tau]}|\hat{s}^{(0)}(t)-s^{(0)}(t)| = o_p(1)$ and $\sup_{t\in[0,\tau]}|[\hat{s}^{(0)}(t)]^{-1}| = O_p(1)$. Under Assumption 2 and using the fact that

$$\frac{\hat{s}^{(1)}(t)}{\hat{s}^{(0)}(t)} - \frac{s^{(1)}(t)}{s^{(0)}(t)} = \frac{\hat{s}^{(1)}(t)-s^{(1)}(t)}{\hat{s}^{(0)}(t)} + \frac{s^{(1)}(t)\{\hat{s}^{(0)}(t)-s^{(0)}(t)\}}{s^{(0)}(t)\hat{s}^{(0)}(t)},$$

we have

$$\begin{aligned} \Big\|\frac{\hat{s}^{(1)}(t)}{\hat{s}^{(0)}(t)} - \frac{s^{(1)}(t)}{s^{(0)}(t)}\Big\| &\leq \sup_{t\in[0,\tau]}\|\hat{s}^{(1)}(t)-s^{(1)}(t)\|\sup_{t\in[0,\tau]}|\{\hat{s}^{(0)}(t)\}^{-1}| \\ &\quad + \sup_{t\in[0,\tau]}|\hat{s}^{(0)}(t)-s^{(0)}(t)|\sup_{t\in[0,\tau]}|\{\hat{s}^{(0)}(t)\}^{-1}|\|\boldsymbol{e}(t)\| \\ &= O_p(1) + o_p(1)\|\boldsymbol{e}(t)\|. \end{aligned}$$

Therefore, $\|\boldsymbol{R}_{i2}(\boldsymbol{\beta}_0)\| \leq O_p(1)|\int_0^\tau d\hat{M}_i(\boldsymbol{\beta}_0, t)| + o_p(1)\|\int_0^\tau \boldsymbol{e}(t)d\hat{M}_i(\boldsymbol{\beta}_0, t)\|$.

Given that $d\hat{M}_i(\boldsymbol{\beta}_0, t) = d\tilde{N}_i(t) - Y_i(t)\boldsymbol{\beta}_0^T\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}dt$ and Assumption 4, we obtain that $\max_{1\leq i\leq n}\|\boldsymbol{R}_{i2}(\boldsymbol{\beta}_0)\| = o_p(p^{1/2}n^{1/q})$. Thus, it follows that

$$\max_{1\leq i\leq n}\|\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\| \leq \max_{1\leq i\leq n}\|\boldsymbol{W}_i(\boldsymbol{\beta}_0)\| + \max_{1\leq i\leq n}\|\boldsymbol{R}_{i1}(\boldsymbol{\beta}_0)\| + \max_{1\leq i\leq n}\|\boldsymbol{R}_{i2}(\boldsymbol{\beta}_0)\| = o_p(p^{1/2}n^{1/q}).$$

Since

$$W_{ni}(\beta) = W_{ni}(\beta_0) + \int_0^\tau Y_i(t)\Big\{Z_i(t) - \bar{Z}(t)\Big\}^{\otimes 2} dt\Big(\beta_0 - \beta\Big), \tag{A.2}$$

thus,

$$\begin{aligned}\max_{1\leq i\leq n}\|\boldsymbol{W}_{ni}(\boldsymbol{\beta})\| &\leq \max_{1\leq i\leq n}\|\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\| + \max_{1\leq i\leq n}\Big\|\int_0^\tau Y_i(t)\Big\{Z_i(t) - \bar{Z}(t)\Big\}^{\otimes 2} dt\Big(\beta_0 - \beta\Big)\Big\| \\ &= o_p(p^{1/2}n^{1/q}) + O_p(a_n) = o_p(p^{1/2}n^{1/q}). \end{aligned}\tag{A.3}$$

The last equality holds since $a_n = o_p(p^{1/2}n^{1/q})$.

Next we show part (ii). For any vector $\boldsymbol{a} \in \mathbb{R}^p$, we have

$$\begin{aligned}&\boldsymbol{a}^T\Big\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\boldsymbol{W}_{ni}^T(\boldsymbol{\beta}_0) - \frac{1}{n}\sum_{i=1}^n \boldsymbol{W}_i(\boldsymbol{\beta}_0)\boldsymbol{W}_i^T(\boldsymbol{\beta}_0)\Big\}\boldsymbol{a} \\ &= \frac{1}{n}\sum_{i=1}^n\Big\{\boldsymbol{a}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) - \boldsymbol{a}^T\boldsymbol{W}_i(\boldsymbol{\beta}_0)\Big\}^2 + \frac{2}{n}\sum_{i=1}^n\Big\{\boldsymbol{a}^T\boldsymbol{W}_i(\boldsymbol{\beta}_0)\Big\}\Big\{\boldsymbol{a}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) - \boldsymbol{a}^T\boldsymbol{W}_i(\boldsymbol{\beta}_0)\Big\} \\ &= O_p(a_n),\end{aligned}$$

which implies that $\frac{1}{n}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\boldsymbol{W}_{ni}^T(\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{W}_i(\boldsymbol{\beta}_0)\boldsymbol{W}_i^T(\boldsymbol{\beta}_0) + O_p(a_n)$ and

$$\begin{aligned}&\Big\|n^{-1}\textstyle\sum_{i=1}^n W_{ni}(\beta_0)W_{ni}^T(\beta_0) - \Sigma\Big\| \\ &= \Big\|n^{-1}\textstyle\sum_{i=1}^n W_{ni}(\beta_0)W_{ni}^T(\beta_0) - n^{-1}\sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0) + n^{-1}\sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0) - \Sigma\Big\| \\ &\leq \Big[2\Big\{\Big\|n^{-1}\textstyle\sum_{i=1}^n W_{ni}(\beta_0)W_{ni}^T(\beta_0) - n^{-1}\sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0)\Big\|^2 \\ &\quad + \Big\|n^{-1}\textstyle\sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0) - \Sigma\Big\|^2\Big\}\Big]^{1/2} \\ &= \Big\{2\{O_p(a_n^2) + o_p(p^{-2})\}\Big\}^{1/2} = o_p(p^{-1}),\end{aligned}$$

where the last equation holds as $p^3/n \to 0$. Recall Eq. (A.2) and the fact that the eigenvalues of matrix $\boldsymbol{D}_n = n^{-1}\sum_{i=1}^n Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2}dt \to \boldsymbol{D}$ in probability. By Assumption 3, we have $\|n^{-1}\sum_{i=1}^n W_{ni}(\beta)W_{ni}^T(\beta) - \Sigma\| = o_p(p^{-1})$. The proof of part (ii) of Lemma 1 is completed.

For $\boldsymbol{\beta} \in E_n$, i.e., $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| \leq Ca_n$, note that $n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + \boldsymbol{D}_n(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$ and the inequality $\|\boldsymbol{a}+\boldsymbol{b}\|^2 \leq 2\{\|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2\}$. It holds that

$$\Big\|n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\| = O_p(a_n). \tag{A.4}$$

The proof of part (iii) is completed.

To show the results in part (iv), it suffices to prove that $n^{-1/2}\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^n \boldsymbol{W}_i(\boldsymbol{\beta}_0) \to N_q(0,\boldsymbol{G})$ in distribution. Let $\boldsymbol{Y}_{ni} = n^{-1/2}\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}_i(\boldsymbol{\beta}_0)$.

$$\begin{aligned} P\Big(\|\boldsymbol{Y}_{ni}\| \geq \epsilon\Big) &\leq \frac{E\|\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}_i(\boldsymbol{\beta}_0)\|^2}{n\epsilon^2} = \frac{tr(\boldsymbol{A}_n\boldsymbol{A}_n^T)}{n\epsilon^2} = O_p(n^{-1}), \\ E\|\boldsymbol{Y}_{ni}\|^4 &= \frac{1}{n^2}E\Big\{\boldsymbol{W}_i^T(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}_n^T\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}_i(\boldsymbol{\beta}_0)\Big\}^2 \\ &\leq \frac{1}{n^2}\gamma_{\max}^2\{\boldsymbol{A}_n^T\boldsymbol{A}_n\}\gamma_{\min}^{-2}\{\boldsymbol{\Sigma}\}E\|\boldsymbol{W}_i(\boldsymbol{\beta}_0)\|^4 = O_p\Big(\frac{p^2}{n^2}\Big), \end{aligned}$$

which implies that

$$\begin{aligned} \sum_{i=1}^n E\|\boldsymbol{Y}_{ni}\|^2 I(\|\boldsymbol{Y}_{ni}\| \geq \epsilon) &\leq n\Big\{E\|\boldsymbol{Y}_{ni}\|^4\Big\}^{1/2}\Big\{P(\|\boldsymbol{Y}_{ni}\| \geq \epsilon)\Big\}^{1/2} \\ &= O_p\Big(n\frac{p}{n}n^{-1/2}\Big) \\ &= O_p\Big(\frac{p}{\sqrt{n}}\Big) = o_p(1) \end{aligned}$$

and $\sum_{i=1}^n \mathrm{cov}(\boldsymbol{Y}_{ni}) = n\mathrm{cov}(\boldsymbol{Y}_{n1}) = \mathrm{cov}(\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}_i(\boldsymbol{\beta}_0)) = \boldsymbol{A}_n\boldsymbol{A}_n^T \to \boldsymbol{G}$. By the Lindeberg-Feller central limit theorem, we have $n^{-1/2}\boldsymbol{A}_n\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^n \boldsymbol{W}_i(\boldsymbol{\beta}_0) \xrightarrow{d} N_q(0,\boldsymbol{G})$. Part (iv) of Lemma 1 is thus proved.

Proof of Theorem 1.

Denote $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n\{\boldsymbol{W}_{ni}(\boldsymbol{\beta})\}^{\otimes 2}$, and write $\boldsymbol{\nu} = \rho\boldsymbol{\theta}$ with $\rho \geq 0$ and $\|\boldsymbol{\theta}\| = 1$. By an argument similar to Owen (2001), we can show that

$$\rho\Big\{\boldsymbol{\theta}^T\boldsymbol{\Sigma}_n(\boldsymbol{\beta})\boldsymbol{\theta} - \max_i\Big\|\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\|\Big\|n^{-1}\sum_{i=1}^n\boldsymbol{\theta}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\|\Big\} \leq \Big\|n^{-1}\sum_{i=1}^n\boldsymbol{\theta}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\|.$$

From Assumptions 2 and 5 and results in (A.3) and (A.4, we have $0 < b \leq \boldsymbol{\theta}^T\boldsymbol{\Sigma}_n(\boldsymbol{\beta})\boldsymbol{\theta} \leq B$. It follows that

$$\max_i\Big\|\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\|\Big\|n^{-1}\sum_{i=1}^n\boldsymbol{\theta}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\| = o_p(n^{1/q}p^{1/2})O_p(a_n) = o_p\Big(\frac{p}{n^{1/2-1/q}}\Big) = o_p(1),$$

for any $q \geq 10/3$. Therefore, $\|\boldsymbol{\nu}\| = \rho = O_p(a_n)$ and

$$\max_i|\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})| = O_p(a_n)o_p(n^{1/q}p^{1/2}) = o_p\Big(\frac{p}{n^{1/2-1/q}}\Big) = o_p(1).$$

For $\boldsymbol{\beta} \in E_n$, we obtain from equation (9) that

$$\begin{aligned} 0 &= n^{-1}\sum_{i=1}^n\frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})}{1+\nu^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})} = n^{-1}\sum_{i=1}^n\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\{1 - \boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}) + \frac{(\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}))^2}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}\Big\} \\ &= n^{-1}\sum_{i=1}^n\boldsymbol{W}_{ni}(\boldsymbol{\beta}) - \Big\{n^{-1}\sum_{i=1}^n\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}\boldsymbol{\nu} + n^{-1}\sum_{i=1}^n\frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})(\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}))^2}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}. \end{aligned}$$

Hence,

$$\boldsymbol{\nu} = \Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}) + \gamma_n, \tag{A.5}$$

where $\gamma_n = \{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\}^{-1} n^{-1}\sum_{i=1}^{n}\frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})(\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}))^2}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}$ and $\Big\|n^{-1}\sum_{i=1}^{n}\frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})(\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta}))^2}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}\Big\|$ is bounded by $\max_i\|\boldsymbol{W}_{ni}(\boldsymbol{\beta})\|\|\boldsymbol{\nu}\|^2 O_p(1) = o_p(p^{1/2}n^{1/q})O_p(p/n) = o_p(p^{3/2}/n^{1-1/q}) = o_p(n^{-1/2})$ because $\frac{p^{3/2}}{n^{1/2-1/q}} = o_p(1)$ as $q \geq 5$ and $p^5/n \to 0$. Then,

$$\boldsymbol{\nu} = \Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}) + \boldsymbol{\vartheta}, \tag{A.6}$$

with $\|\boldsymbol{\vartheta}\| = o_p(n^{-1/2})$. Let $\zeta_i(\boldsymbol{\beta}) = \boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})$. We have that $\max_i|\zeta_i(\boldsymbol{\beta})| = o_p(1)$ because of Taylor's expansion $\log\{1+\zeta_i(\boldsymbol{\beta})\} = \zeta_i(\boldsymbol{\beta}) - \frac{1}{2}\zeta_i^2(\boldsymbol{\beta}) + \varsigma_i$, where ς_i satisfies that $P(|\varsigma_i| \leq C|\zeta_i|^3, 1 \leq i \leq n) \to 1$ for some constant C as $n \to \infty$. Therefore, it follows from equation (A.6) that

$$\begin{aligned}
l(\boldsymbol{\beta}) &= 2\sum_{i=1}^{n}\log\{1+\zeta_i\} = 2\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\zeta_i^2 + \sum_{i=1}^{n}\varsigma_i \\
&= 2n\nu^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\} - n\boldsymbol{\nu}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}\boldsymbol{\nu} + \sum_{i=1}^{n}\varsigma_i \\
&= n\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}^{-1}\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\} \\
&\quad -n\boldsymbol{\vartheta}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}^{-1}\boldsymbol{\vartheta} + \sum_{i=1}^{n}\varsigma_i.
\end{aligned}$$

Here, $\|n\boldsymbol{\vartheta}^T\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\}^{-1}\boldsymbol{\vartheta}\| = n\|\boldsymbol{\vartheta}\|^2 O_p(1) = n o_p(n^{-1})O_p(1) = o_p(1)$, and

$$\begin{aligned}
\Big\|\sum_{i=1}^{n}\varsigma_i\Big\| &\leq \sum_{i=1}^{n}C\Big\|\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\|^3 \leq Cn\|\boldsymbol{\nu}\|\max_i\|\boldsymbol{W}_{ni}(\boldsymbol{\beta})\|\boldsymbol{\nu}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta})\boldsymbol{W}_{ni}^T(\boldsymbol{\beta})\Big\}\boldsymbol{\nu} \\
&= Cn\|\boldsymbol{\nu}\|^3 o_p(p^{1/2}n^{1/q})O_p(1) = o_p(p^2/n^{1/2-1/q}) = o_p(1),
\end{aligned}$$

since $q \geq 10$ and $p^5/n \to 0$. Thus, we have

$$l(\boldsymbol{\beta}) = n\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}^{-1}\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta})\Big\} + o_p(1).$$

Let ∂E_n denote the boundary of E_n. For any $\boldsymbol{\beta} \in \partial E_n$, we write that $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + Ca_n\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a unit vector. Then we have $l(\boldsymbol{\beta}) = T_0 + T_1 + T_2$, where

$$\begin{aligned}
T_0 &= n\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta}_0)\Big\}^{-1}\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\}, \\
T_1 &= n\Big\{n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2}dt(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\Big\}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta})\Big\}^{-1} \\
&\quad \Big\{n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2}dt(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\Big\}, \\
T_2 &= l(\boldsymbol{\beta}) - T_0 - T_1.
\end{aligned}$$

By Assumption 3, we have $T_1 = O_p(na_n^2)$ with $na_n^2 \to \infty$, $T_2/T_1 \to 0$ in probability as $n \to \infty$, and $l(\boldsymbol{\beta}_0) = T_0 + o_p(1)$. This implies that for any given constant C, $P\{l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0) > C\} \to 1$ as $n \to \infty$. Therefore, $l(\beta)$ has a minimum in E_n with probability tending to 1.

Proof of Theorem 2.

Based on the definition of empirical likelihood, the estimator $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\beta}}_E)$ satisfies $\boldsymbol{Q}_{jn}(\hat{\boldsymbol{\beta}}_E, \hat{\boldsymbol{\nu}}) = 0$ $(j = 1, 2)$, where $\boldsymbol{Q}_{1n}(\boldsymbol{\beta}, \boldsymbol{\nu}) = n^{-1} \sum_{i=1}^n \frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})}{1+\boldsymbol{\nu}^T \boldsymbol{W}_{ni}(\boldsymbol{\beta})}$, $\boldsymbol{Q}_{2n}(\boldsymbol{\beta}, \boldsymbol{\nu}) = n^{-1} \sum_{i=1}^n \frac{(\frac{\partial \boldsymbol{W}_{ni}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}})^T \boldsymbol{\nu}}{1+\boldsymbol{\nu}^T \boldsymbol{W}_{ni}(\boldsymbol{\beta})}$ and $\frac{\partial \boldsymbol{W}_{ni}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\int_0^\tau Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2} dt$. Note that $\|\hat{\boldsymbol{\nu}}\| = O_p(a_n)$ is stochastically small and $\|\hat{\beta}_E - \beta_0\| = O_p(a_n)$. Using the stochastic expansions of $\boldsymbol{Q}_{jn},\ j = 1, 2$, around the value $(\boldsymbol{\beta}_0, \mathbf{0})$, it follows that

$$\begin{pmatrix} -\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}) \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\Sigma}_n & -\boldsymbol{D}_n \\ -\boldsymbol{D}_n^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0 \end{pmatrix} + O_p(\epsilon_n), \tag{A.7}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_n &= n^{-1} \sum_{i=1}^n \boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta}_0) \quad \text{with} \quad \|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| = o_p(p^{-1}), \\
\boldsymbol{D}_n &= n^{-1} \sum_{i=1}^n \int_0^\tau Y_i(t) \Big\{ \boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t) \Big\}^{\otimes 2} dt \quad \text{with} \quad \|\boldsymbol{D}_n - \boldsymbol{D}\| = o_p(p^{-1}), \text{ and} \\
\epsilon_n &= p^{1/2} \Big\{ \|\hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0\|^2 + \|\hat{\boldsymbol{\nu}}\|^2 \Big\} = O_p\Big(\frac{p^{3/2}}{n}\Big) = o_p(n^{-1/2}).
\end{aligned}$$

The last equation holds since $p^3/n \to 0$. Note that $\|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| = O_p(p^{-1})$ and $\|\boldsymbol{D}_n - \boldsymbol{D}\| = O_p(p^{-1})$, thus $\|(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma})\hat{\nu}\| = O_p((np)^{-1/2}) = o_p(n^{-1/2})$ and $\|(\boldsymbol{D}_n - \boldsymbol{D})\hat{\boldsymbol{\nu}}\| = O_p((np)^{-1/2}) = o_p(n^{-1/2})$ because that $p \to \infty$. Therefore, equation (A.7) can be rewritten as

$$\begin{pmatrix} -n^{-1} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\Sigma} & -\boldsymbol{D} \\ -\boldsymbol{D} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0 \end{pmatrix} + o_p(n^{-1/2}).$$

After simple calculations, we obtain that $\hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0 = \boldsymbol{D}^{-1} n^{-1} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_p(n^{-1/2})$. Thus

$$\begin{aligned}
n^{1/2} \boldsymbol{A}_n \boldsymbol{\Psi}^{-1/2} \{\hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0\} &= n^{-1/2} \boldsymbol{A}_n \boldsymbol{\Psi}^{-1/2} \boldsymbol{D}^{-1} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_p(n^{-1/2}) \\
&= n^{-1/2} \tilde{\boldsymbol{A}}_n \boldsymbol{\Sigma}^{-1/2} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_p(1),
\end{aligned}$$

where $\tilde{\boldsymbol{A}}_n = \boldsymbol{A}_n \boldsymbol{\Psi}^{-1/2} \boldsymbol{D}^{-1} \boldsymbol{\Sigma}^{1/2}$. Note that $\tilde{\boldsymbol{A}}_n \tilde{\boldsymbol{A}}_n^T = \boldsymbol{A}_n \boldsymbol{\Psi}^{-1/2} \boldsymbol{D}^{-1} \boldsymbol{\Sigma} \boldsymbol{D}^{-1} \boldsymbol{\Psi}^{-1/2} \boldsymbol{A}_n^T = \boldsymbol{A}_n \boldsymbol{A}_n^T$. From part (iv) of Lemma 1, we obtain that $n^{-1/2} \tilde{\boldsymbol{A}}_n \boldsymbol{\Sigma}^{-1/2} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \xrightarrow{d} N_q(0, \boldsymbol{G})$ as $n \to \infty$, which implies that $n^{1/2} \boldsymbol{A}_n \boldsymbol{\Psi}^{-1/2} \{\hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0\} \xrightarrow{d} N_q(0, \boldsymbol{G})$ as $n \to \infty$. The proof of Theorem 2 is completed.

Lemma 2. *Suppose that conditions in Theorem 3 hold. With probability tending to 1, $l_p(\boldsymbol{\beta})$ has a minimum in E_n.*

Proof. Note that $p_\lambda(0)=0$ and for large enough n,

$$\begin{aligned} l_p(\boldsymbol{\beta})-l_p(\boldsymbol{\beta}_0) &= \frac{l(\boldsymbol{\beta})}{2}-\frac{l(\boldsymbol{\beta}_0)}{2}+n\sum_{j=1}^{n}\Big\{p_\lambda(|\beta_j|)-p_\lambda(|\beta_{0j}|)\Big\} \\ &\geq \frac{l(\boldsymbol{\beta})}{2}-\frac{l(\boldsymbol{\beta}_0)}{2}+n\sum_{j\in\mathcal{A}}\Big\{p_\lambda(|\beta_j|)-p_\lambda(|\beta_{0j}|)\Big\} \geq \frac{l(\boldsymbol{\beta})-l(\boldsymbol{\beta}_0)}{2} \end{aligned}$$

where the last inequality holds due to Condition 1 and the unbiased property of the SCAD penalty, i.e., $p_\lambda(\beta_{0j})=p_\lambda(\beta_j)$ for $j\in\mathcal{A}$ when n is large. Hence, with probability tending to 1, $l_p(\boldsymbol{\beta})>l_p(\boldsymbol{\beta}_0)$ for $\boldsymbol{\beta}\in\partial E_n$. This implies that $l_p(\boldsymbol{\beta})$ has a minimum in E_n with probability tending to 1. The proof of Lemma 2 is completed.

Lemma 3. *Suppose that conditions in Theorem 3 hold. The local minimizer $\hat{\boldsymbol{\beta}}$ of (11) satisfies $\hat{\boldsymbol{\beta}}_2=0$ with probability tending to 1.*

Proof. For $\boldsymbol{\beta}\in E_n$ and its j-th component β_j, we have

$$\begin{aligned} n^{-1}\frac{\partial l_p(\boldsymbol{\beta})}{\partial\beta_j} &= n^{-1}\sum_{i=1}^{n}\frac{-\boldsymbol{\nu}^T\Big\{\int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}(t)]^{\otimes 2}dt\Big\}_j}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}+p'_\lambda(|\beta_j|)\mathrm{sign}(\beta_j) \\ &\overset{\text{def}}{=} Q_j+P_j,\quad j=1,\cdots,p \end{aligned}$$

where $\Big\{\int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}(t)]^{\otimes 2}dt\Big\}_j$ denotes the j-th column of $p\times p$ matrix $\int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}(t)]^{\otimes 2}dt$. Note that $\|\boldsymbol{\nu}\|=O_p(a_n)$ and $\lambda(n/p)^{1/2}\to\infty$ from Condition 1. We have

$$\begin{aligned} \Big\|\boldsymbol{\nu}^T n^{-1}\sum_{i=1}^{n}\frac{\int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}(t)]^{\otimes 2}dt}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}\Big\| &\leq \|\boldsymbol{\nu}\|\gamma_{\max}\Big\{n^{-1}\sum_{i=1}^{n}\frac{\int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}(t)]^{\otimes 2}dt}{1+\boldsymbol{\nu}^T\boldsymbol{W}_{ni}(\boldsymbol{\beta})}\Big\} \\ &= O_p(a_n)O_p(1)o_p(\lambda), \end{aligned}$$

This implies that $P(\max_{j\in\mathcal{A}^C}|Q_j|>\lambda/2)\to 0$, where $\mathcal{A}^C$ is the complementary of set $\mathcal{A}$. In addition, $|\beta_j|\leq ca_n$ for $j\in\mathcal{A}^C$ and $\lambda/a_n\to\infty$. For large enough n and $j\in\mathcal{A}^C, p_\lambda(|\beta_j|)=\lambda$ and $P_j=\lambda\mathrm{sign}(\beta_j)$. It is easy to see that the sign of β_j determines the sign of $\frac{\partial l_p(\boldsymbol{\beta})}{\partial\beta_j}$ asymptotically for $j\in\mathcal{A}^C$. In other words, with probability tending to 1, $n^{-1}\frac{\partial l_p(\boldsymbol{\beta})}{\partial\beta_j}<0$ if $\beta_j\in(-ca_n,0)$ and $n^{-1}\frac{\partial l_p(\boldsymbol{\beta})}{\partial\beta_j}>0$ if $\beta_j\in(0,ca_n)$. Hence $\hat{\boldsymbol{\beta}}_2=\mathbf{0}$ with probability tending to 1 and the sparsity of $\hat{\boldsymbol{\beta}}$ follows.

Proof of Theorem 3.

The sparsity of estimator $\hat{\boldsymbol{\beta}}$ in part (i) of the theorem follows directly from Lemma 3. It suffices to prove the asymptotic normality of $\hat{\boldsymbol{\beta}}_1$. By Lemma 3 and the definition of the penalized empirical likelihood, $\hat{\boldsymbol{\beta}}$ is the constrained minimizer of $l_p(\boldsymbol{\beta})$ in (11) subject to $\boldsymbol{H}_2\boldsymbol{\beta} = 0$, where $\boldsymbol{H}_2$ is a $(p-d)\times p$ block matrix of the $p\times p$ identical matrix $\boldsymbol{I}_p$ such that $\boldsymbol{I}_p = (\boldsymbol{H}_1^T, \boldsymbol{H}_2^T)$ with $\boldsymbol{H}_1 \in \mathbf{R}^{d\times p}$. Using the Lagrange multiplier method, the estimator $\hat{\boldsymbol{\beta}}$ can be equivalently obtained through minimizing the following objective function

$$\tilde{l}_p(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{\kappa}) = n^{-1}\sum_{i=1}^{n} \log\left\{1 + \boldsymbol{\nu}^T \boldsymbol{W}_{ni}(\boldsymbol{\beta})\right\} + \sum_{j=1}^{p} p_\lambda(|\beta_j|) + \boldsymbol{\kappa}^T H_2 \boldsymbol{\beta}, \tag{A.8}$$

where $\boldsymbol{\kappa} \in \mathbf{R}^{p-d}$ is another vector of Lagrange multipliers. Define

$$\begin{aligned}
\boldsymbol{Q}_{1n}(\boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\kappa}) &= n^{-1}\sum_{i=1}^{n} \frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})}{1+\boldsymbol{\nu}^T \boldsymbol{W}_{ni}(\boldsymbol{\beta})}, \\
\boldsymbol{Q}_{2n}(\boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\kappa}) &= n^{-1}\sum_{i=1}^{n} \frac{\left(\frac{\partial \boldsymbol{W}_{ni}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{\nu}}{1+\boldsymbol{\nu}^T \boldsymbol{W}_{ni}(\boldsymbol{\beta})} + \boldsymbol{b}(\boldsymbol{\beta}) + \boldsymbol{H}_2^T \boldsymbol{\kappa}, \\
\boldsymbol{Q}_{3n}(\boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\kappa}) &= \boldsymbol{H}_2 \boldsymbol{\beta},
\end{aligned}$$

where $\frac{\partial \boldsymbol{W}_{ni}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\int_0^\tau Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2} dt$ and $\boldsymbol{b}(\boldsymbol{\beta}) = \left\{p'_\lambda(|\beta_1|)\text{sign}(\beta_1), \ldots, p'_\lambda(|\beta_d|)\text{sign}(\beta_d), \mathbf{0}^T\right\}^T$. Then $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\kappa}})$ are solutions to the equations $Q_{jn}(\beta, \nu, \kappa) = 0$ for $j = 1, 2, 3$. Note that $\|\hat{\boldsymbol{\nu}}\| = O_p(a_n)$ and $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(a_n)$. Similar to an argument of Qin and Lawless (1995), $Q_{2n}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\kappa}}) = \mathbf{0}$ implies that $\|\boldsymbol{\kappa}\| = O_p(a_n)$. We expand $\boldsymbol{Q}_{jn}$ $(j = 1, 2, 3)$ around value $(\boldsymbol{\beta}_0, \mathbf{0}, \mathbf{0})$, and obtain that

$$\begin{pmatrix} -\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}, \mathbf{0}) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\Sigma}_n & -\boldsymbol{D}_n & \mathbf{0} \\ -\boldsymbol{D}_n^T & \boldsymbol{\Omega}(\boldsymbol{\beta}_0) & \boldsymbol{H}_2^T \\ \mathbf{0} & \boldsymbol{H}_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\kappa}} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{b}|(\boldsymbol{\beta}_0) \\ \mathbf{0} \end{pmatrix} + O_p(\epsilon_n), \tag{A.9}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_n &= n^{-1}\sum_{i=1}^{n} \boldsymbol{W}_{ni}^{\otimes 2}(\boldsymbol{\beta}_0) \quad \text{with } \|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| = o_p(p^{-1}) \\
\boldsymbol{D}_n &= n^{-1}\sum_{i=1}^{n} \int_0^\tau Y_i(t)\left\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\right\}^{\otimes 2} dt \quad \text{with } \|\boldsymbol{D}_n - \boldsymbol{D}\| = o_p(p^{-1}) \\
\boldsymbol{\Omega}(\boldsymbol{\beta}_0) &= \text{diag}\left\{p''_\lambda(|\beta_{01}|), \ldots, p''_\lambda(|\beta_{0d}|), 0, \ldots, 0\right\} \\
\epsilon_n &= p^{1/2}\left\{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\hat{\boldsymbol{\nu}}\|^2 + \|\hat{\boldsymbol{\kappa}}\|^2\right\} = O_p\left(\frac{p^{3/2}}{n}\right) = o_p(n^{-1/2}).
\end{aligned}$$

The last equation holds since $\frac{p^3}{n} \to 0$. Given Condition 2, we have $\|\boldsymbol{\Omega}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\| = o_p(p^{-1/2})O_p(a_n) = o_p(n^{-1/2})$. Note that $\|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| = O_p(p^{-1})$ and $\|\boldsymbol{D}_n - \boldsymbol{D}\| = O_p(p^{-1})$. It follows from Condition

1 that $\|(\Sigma_n - \Sigma)\hat{\nu}\| = O_p((np)^{-1/2}) = o_p(n^{-1/2})$ and $\|(\boldsymbol{D}_n - \boldsymbol{D})\hat{\boldsymbol{\nu}}\| = O_p((np)^{-1/2}) = o_p(n^{-1/2})$ Therefore, equation (A.9) can be rewritten as

$$\begin{pmatrix} -n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\Sigma} & -\boldsymbol{D} & \boldsymbol{0} \\ -\boldsymbol{D} & \boldsymbol{0} & \boldsymbol{H}_2^T \\ \boldsymbol{0} & \boldsymbol{H}_2 & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\kappa}} \end{pmatrix} + o_p(n^{-1/2}).$$

Let $\boldsymbol{K}_{11} = -\boldsymbol{\Sigma}, \boldsymbol{K}_{12} = (-\boldsymbol{D}, \boldsymbol{0})$, $\boldsymbol{K}_{21} = \boldsymbol{K}_{12}^T$,

$$\boldsymbol{K}_{22} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{H}_2^T \\ \boldsymbol{H}_2 & \boldsymbol{0} \end{pmatrix} \text{ and } \boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{pmatrix}.$$

It follows that

$$\begin{pmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\kappa}} \end{pmatrix} = \boldsymbol{K}^{-1} \begin{pmatrix} -n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} + o_p(n^{-1/2}),$$

Applying matrix inverse by blocks, we obtain that

$$\boldsymbol{K}^{-1} = \begin{pmatrix} \boldsymbol{K}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{K}_{11}^{-1}\boldsymbol{K}_{12} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{K}_{22.1}^{-1} \begin{pmatrix} -\boldsymbol{K}_{21}\boldsymbol{K}_{11}^{-1} & I \end{pmatrix},$$

where $\boldsymbol{K}_{22.1} = \boldsymbol{K}_{22} - \boldsymbol{K}_{21}\boldsymbol{K}_{11}^{-1}\boldsymbol{K}_{12}$. Then,

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\kappa}} \end{pmatrix} = \boldsymbol{K}_{22.1}^{-1}\boldsymbol{K}_{21}\boldsymbol{K}_{11}^{-1} n^{-1} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_p(n^{-1/2}).$$

with

$$\boldsymbol{K}_{22.1}^{-1} = \begin{pmatrix} \boldsymbol{\Xi}^{-1} - \boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{\Xi}^{-1} & \boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1} \\ (\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{\Xi}^{-1} & -(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1} \end{pmatrix},$$

and $\boldsymbol{\Xi} = \boldsymbol{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}$. Consequently

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &= \Big\{\boldsymbol{\Xi}^{-1} - \boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\Big\}\boldsymbol{D}\boldsymbol{\Sigma}^{-1}n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_p(n^{-1/2}) \\ &= \Big\{\boldsymbol{I}_p - \boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\Big\}\boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{1/2}\Big\{\boldsymbol{\Sigma}^{-1/2}n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} + o_p(n^{-1/2}) \end{aligned}$$

and $\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10} = \Big\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\Big\}\boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{1/2}\Big\{\boldsymbol{\Sigma}^{-1/2}n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} + o_p(n^{-1/2})$. The asymptotic covariance of $\sqrt{n}\hat{\boldsymbol{\beta}}_1$ is then given by

$$\begin{aligned} \boldsymbol{I}_{\mathcal{A}} &= \Big\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}H_2\Big\}\boldsymbol{D}^{-1}\boldsymbol{\Sigma}\boldsymbol{D}^{-1}\Big\{\boldsymbol{H}_1^T - \boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_1^T\Big\} \\ &= \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_1^T - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_1^T. \end{aligned}$$

Denote $\boldsymbol{\Psi} = \boldsymbol{\Xi}^{-1}$. Note that

$$\boldsymbol{\Xi}^{-1} = \begin{pmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \end{pmatrix} \boldsymbol{\Xi}^{-1}(\boldsymbol{H}_1^T, \boldsymbol{H}_2^T) = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{pmatrix},$$

where submatrices $\boldsymbol{\Psi}_{ij} = \boldsymbol{H}_i \boldsymbol{\Xi}^{-1} \boldsymbol{H}_j^T$ for $i = 1, 2$ and $j = 1, 2$. Then, $I_{\mathcal{A}} = \boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}$. Let $\boldsymbol{Y}_{ni} = n^{-1/2}\boldsymbol{A}_n \boldsymbol{I}_{\mathcal{A}}^{-1/2}\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\}\boldsymbol{D}^{-1}\boldsymbol{W}_{ni}(\beta_0)$ with $\boldsymbol{A}_n$ being a $q \times d$ matrix with fixed q and $\boldsymbol{A}_n\boldsymbol{A}_n^T \to \boldsymbol{G}$. It is straightforward to verify that

$$\begin{aligned}
P(\|\boldsymbol{Y}_{ni}\| \geq \epsilon) &\leq \frac{E\Big\|\boldsymbol{A}_n \boldsymbol{I}_{\mathcal{A}}^{-1/2}\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\}\boldsymbol{D}^{-1}\boldsymbol{W}_{ni}(\beta_0)\Big\|^2}{n\epsilon^2} = O_p(n^{-1}) \\
E\|\boldsymbol{Y}_{ni}\|^4 &= \frac{1}{n^2}E\Big\{\boldsymbol{W}_{ni}^T(\beta_0)\boldsymbol{D}^{-1}\{\boldsymbol{H}_1^T - \boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_1^T\Big\}\boldsymbol{I}_{\mathcal{A}}^{-1/2}\boldsymbol{A}_n^T\boldsymbol{A}_n \\
&\quad \boldsymbol{I}_{\mathcal{A}}^{-1/2}\Big\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\}\boldsymbol{D}^{-1}\boldsymbol{W}_{ni}(\beta_0)\Big\}^2 \\
&\leq \frac{1}{n^2}\gamma_{\max}^2\{\boldsymbol{A}_n\boldsymbol{A}_n^T\}\gamma_{\min}^{-2}\{\boldsymbol{I}_{\mathcal{A}}\}E\Big\|\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\}\boldsymbol{D}^{-1}\boldsymbol{W}_{ni}(\beta_0)\Big\|^4 \\
&= O_p\Big(\frac{p^2}{n^2}\Big).
\end{aligned}$$

It follows that

$$\sum_{i=1}^n E\|\boldsymbol{Y}_{ni}\|^2 I(\|\boldsymbol{Y}_{ni}\| \geq \epsilon) \leq n\Big\{E\|\boldsymbol{Y}_{ni}\|^4\}^{1/2}\{P(\|\boldsymbol{Y}_{ni}\| \geq \epsilon)\Big\}^{1/2} = O_p\Big(\frac{p}{\sqrt{n}}\Big) = o_p(1).$$

By the Lindeberg-Feller Martingale central limit theorem, we have

$$n^{-1/2}\boldsymbol{A}_n\boldsymbol{I}_{\mathcal{A}}^{1/2}\Big\{\boldsymbol{H}_1 - \boldsymbol{H}_1\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\Big\}\boldsymbol{D}^{-1}\Big\{n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\beta_0)\Big\} \xrightarrow{d} N_q(0, \boldsymbol{G})$$

as $\boldsymbol{A}_n\boldsymbol{A}_n^T \to \boldsymbol{G}$. This implies the asymptotic normality of $\hat{\boldsymbol{\beta}}_1$, i.e., $n^{1/2}\boldsymbol{A}_n\boldsymbol{I}_{\mathcal{A}}^{-1/2}\{\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}\} \to N_q(0, \boldsymbol{G})$ in distribution. The proof of Theorem 3 is completed.

Proof of Theorem 4.

From the proof of theorem 3, we have

$$\hat{\boldsymbol{\beta}} - \beta_0 = \Big\{\boldsymbol{I}_p - \boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\Big\}\boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{1/2}\Big\{\boldsymbol{\Sigma}^{-1/2}n^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\beta_0)\Big\} + o_p(n^{-1/2})$$

and $\|(\boldsymbol{D}_n - \boldsymbol{D})(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})\| = o_p(p^{-1})O_p(a_n) = o_p(n^{-1/2})$. Then,

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\hat{\boldsymbol{\beta}}) &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \Big\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\Big\}\Big\{d\tilde{N}_i(t) - Y_i(t)\hat{\boldsymbol{\beta}}^T\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}dt\Big\} \\
&= n^{-1}\sum_{i=1}^{n}\int_0^\tau \Big\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\Big\}\Big\{d\tilde{N}_i(t) - Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^T dt(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0 + \hat{\boldsymbol{\beta}})\Big\} \\
&= n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + \boldsymbol{D}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) + (\boldsymbol{D}_n - \boldsymbol{D})(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) \\
&= \boldsymbol{\Sigma}\boldsymbol{D}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{D}^{-1}\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} + o_p(n^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
l(\hat{\boldsymbol{\beta}}) &= n\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\hat{\boldsymbol{\beta}})\Big\}^T\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}^{\otimes 2}(\hat{\boldsymbol{\beta}})\Big\}^{-1}\Big\{n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\hat{\boldsymbol{\beta}})\Big\} + o_p(1) \\
&= \Big\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\}^T\boldsymbol{D}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{D}^{-1}\Big\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} + o_p(1).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
2l_p(\hat{\boldsymbol{\beta}}) &= \Big\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\}^T\boldsymbol{D}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{D}^{-1}\Big\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} \\
&\quad +2n\sum_{j=1}^{p}p_\lambda(\hat{\beta}_j) + o_p(1).
\end{aligned}
$$

Let $\tilde{\boldsymbol{H}}_2 = \begin{pmatrix}\tilde{\boldsymbol{A}} \\ \boldsymbol{H}_2\end{pmatrix} \in \mathbf{R}^{(p-d+q)\times p}$ with $\tilde{\boldsymbol{A}} = (\boldsymbol{A}, \mathbf{0}) \in \mathbf{R}^{q\times p}$ and denote $\tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{A}\boldsymbol{\beta}_1=0} l_p(\boldsymbol{\beta})$. Similar to the proof of Theorem 3, we obtain that under $H_0 : \boldsymbol{A}\boldsymbol{\beta}_1 = 0$,

$$
\begin{aligned}
2l_p(\tilde{\boldsymbol{\beta}}) &= \Big\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\}^T\boldsymbol{D}^{-1}\tilde{\boldsymbol{H}}_2^T(\tilde{\boldsymbol{H}}_2\boldsymbol{\Xi}^{-1}\tilde{\boldsymbol{H}}_2^T)^{-1}\tilde{\boldsymbol{H}}_2\boldsymbol{D}^{-1}\Big\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} \\
&\quad +2n\sum_{j=1}^{p}p_\lambda(\tilde{\beta}_j) + o_p(1).
\end{aligned}
$$

By Condition 1 and the oracle properties of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$, we have $n\sum_{j=1}^{n}\{p_\lambda(\tilde{\beta}_j) - p_\lambda(\hat{\beta}_j)\} = 0$ with probability tending to 1. Therefore,

$$
PELR = 2l_p(\tilde{\boldsymbol{\beta}}) - 2l_p(\hat{\boldsymbol{\beta}}) = \Big\{n^{-1/2}\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\}^T(\tilde{\boldsymbol{P}} - \boldsymbol{P})\Big\{n^{-1/2}\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\Big\} + o_p(1),
$$

where $\tilde{\boldsymbol{P}} = \Sigma^{1/2}\boldsymbol{D}^{-1}\tilde{\boldsymbol{H}}_2^T(\tilde{\boldsymbol{H}}_2\boldsymbol{\Xi}^{-1}\tilde{\boldsymbol{H}}_2^T)^{-1}\tilde{\boldsymbol{H}}_2\boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{P} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{D}^{-1}\boldsymbol{H}_2^T(\boldsymbol{H}_2\boldsymbol{\Xi}^{-1}\boldsymbol{H}_2^T)^{-1}\boldsymbol{H}_2\boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{1/2}$. Note that $\boldsymbol{H}_2 = (\mathbf{0}, \boldsymbol{I}_{p-d})\tilde{\boldsymbol{H}}_2$. It is straightforward to verify that $\tilde{\boldsymbol{P}}$ and $\boldsymbol{P}$ are idempotent matrices of ranks $p - d + q$ and $p - d$, respectively. $\tilde{\boldsymbol{P}} - \boldsymbol{P}$ is also an idempotent matrix of rank q because $(\tilde{\boldsymbol{P}} - \boldsymbol{P})(\tilde{\boldsymbol{P}} - \boldsymbol{P}) = \tilde{\boldsymbol{P}}^2 - \tilde{\boldsymbol{P}}\boldsymbol{P} - \boldsymbol{P}\tilde{\boldsymbol{P}} + \boldsymbol{P}^2 = \tilde{\boldsymbol{P}} - \boldsymbol{P}$. Therefore, $\tilde{\boldsymbol{P}} - \boldsymbol{P}$ can be written as $\boldsymbol{A}_n^T \boldsymbol{A}_n$ with $\boldsymbol{A}_n \boldsymbol{A}_n^T = \boldsymbol{I}_q$. By the oracle property, we have that $n^{-1/2} \boldsymbol{A}_n \boldsymbol{\Sigma}^{-1/2} \sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \xrightarrow{d} N_q(0, \boldsymbol{I}_q)$. Hence, $PELR \xrightarrow{d} \chi_q^2$ under H_0. The proof of Theorem 4 is completed.