Statistica Sinica: Supplement

Inference for Low-Dimensional Covariates in a High-Dimensional Accelerated Failure Time Model

Hao Chai, Qingzhao Zhang, Jian Huang and Shuangge Ma

Yale University, Xiamen University, The Hong Kong Polytechnic University

Supplementary Material

This file contains proofs for the theoretical results described in the main text.

Define the norms of row sub-matrix as $\|\mathbf{D}\|_{*,S} = ||\mathbf{D}_S||_*$, where $* \in \{1,2\}$, $S \subset \{1,2,\ldots,p+q\}$, and $\mathbf{D}_S$ is the submatrix that consists of the rows of $\mathbf{D}$ indexed by S. Before proving the theoretical results, we present the following two useful lemmas. Lemma 1 shows that the empirical covariance matrix also enjoys the restricted eigenvalue condition when this matrix is close (in terms of maximum entry-wise distance) to a matrix which does satisfy the restricted eigenvalue condition. Lemma 2 gives an upper bound for the l_1-norm estimation accuracy of $\widetilde{\mathbf{B}} = (\widetilde{\mathbf{b}}_{,1}, \cdots, \widetilde{\mathbf{b}}_{,p})$. Following a similar idea, the upper bound of $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1$ is presented in Lemma 3. Note that these three lemmas rely on certain events, such as Ω_1 in Lemma 2 and Ω_2 in Lemma 3. Lemma 4 establishes certain tail probabilities of these events to ensure that the lemmas can be used to prove Theorem 1.

S1 Some useful lemmas

Lemma 1. *Denote* $\Gamma_n = \sum_{i=1}^n \omega_i \mathbf{u}_{i,} \mathbf{u}_{i,}^\top$ *with* $\mathbf{u}_{i,} = (\mathbf{x}_{i,}^\top, \mathbf{z}_{i,}^\top)^\top$. *Under Assumptions 1, 2, 4 and* $|A_0|\sqrt{\log q/n} = o(1)$, *if* $\max_{ij} |\Gamma_{n,ij} - \Gamma_{ij}| = O_p(\sqrt{\log q/n})$, *we have that* Γ_n *satisfies*

$$\inf_{\|\mathbf{a}\|_{1,A^c} \le 3\|\mathbf{a}\|_{1,A}} \frac{\mathbf{a}^\top \Gamma_n \mathbf{a}}{\|\mathbf{a}\|_{2,A}^2} > c_*/2 > 0$$

as $n \to \infty$, *where* A *and* c_* *are defined in Assumption 4.*

Proof. By applying Lemma 10.1 in van de Geer and Bühlmann (2009) and Lemma 6 in Kock and Callot (2015), we have

$$\inf_{\|\mathbf{a}\|_{1,A^c} \le 3\|\mathbf{a}\|_{1,A}} \frac{\mathbf{a}^\top \Gamma_n \mathbf{a}}{\|\mathbf{a}\|_{2,A}^2} \ge \kappa^2(|A|) - 16|A| \max_{ij} |\Gamma_{n,ij} - \Gamma_{ij}| > c_*/2.$$

The last inequality holds as $|A| \asymp |A_0|$ and $|A_0|\sqrt{\frac{\log q}{n}} = o(1)$ as $n \to \infty$. $\square$

Lemma 2. *Suppose that the event* $\Omega_1 = \{\|\mathbf{Z}^\top \mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\mathbf{b}_{,k})/n\|_\infty < \lambda_k/2,$ *for* $k = 1, \cdots, p\}$ *and the conditions in Lemma 1 hold. Then for each* $k \in \{1, 2, \cdots, p\}$,

$$||\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}||_{1,K_0^c} \le 3||\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}||_{1,K_0}, \tag{S1.1}$$

and

$$\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_1 \le 24\lambda_k |K_0|/c_*. \tag{S1.2}$$

Proof. By the definition of $\widetilde{\mathbf{B}}$, for each $k \in \{1, 2, \ldots, p\}$, we have

$$L_k(\widetilde{\mathbf{b}}_{,k}) - L_k(\mathbf{b}_{,k}) \leq \lambda_k(\|\mathbf{b}_{,k}\|_1 - \|\widetilde{\mathbf{b}}_{,k}\|_1).$$

The left-hand side is

$$\begin{aligned} LHS &= \frac{1}{2n}(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k})^\top \mathbf{Z}^\top \mathbf{W}\mathbf{Z}(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}) \\ &\quad - \frac{1}{n}(\mathbf{x}_{,k} - \ \mathbf{Z}\mathbf{b}_{,k})^\top \mathbf{W}\mathbf{Z}(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}). \end{aligned}$$

Note that $(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k})^\top \mathbf{Z}^\top \mathbf{W}\mathbf{Z}(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}) \geq 0$. If Ω_1 holds, we have

$$\begin{aligned} 0 &\leq \lambda_k(\|\mathbf{b}_{,k}\|_1 - \|\widetilde{\mathbf{b}}_{,k}\|_1) + \frac{1}{2}\lambda_k\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_1 \\ &\leq \lambda_k\|\mathbf{b}_{,k}\|_{1,K_0} - \lambda_k\|\widetilde{\mathbf{b}}_{,k}\|_{1,K_0} - \lambda_k\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_{1,K_0^c} + \frac{1}{2}\lambda_k\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_1 \\ &\leq \frac{3}{2}\lambda_k\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_{1,K_0} - \frac{1}{2}\lambda_k\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_{1,K_0^c}, \end{aligned}$$

which proves (S1.1). The second inequality holds because $b_{j,k} = 0$ for $j \in K_0^c$. The third inequality is due to the fact that $|x| - |y| \leq |x - y|$ for any $x, y \in \mathbb{R}$.

With the above discussions, we have

$$\frac{1}{n}(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k})^\top \mathbf{Z}^\top \mathbf{W}\mathbf{Z}(\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}) \leq 3\lambda_k\|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_{1,K_0}\ . \tag{S1.3}$$

Define two new $(p+q) \times 1$ vectors $\widetilde{\mathbf{a}}_{,k} = (0_{p\times 1}^\top, \widetilde{\mathbf{b}}_{,k}^\top)^\top$ and $\mathbf{a}_{,k} = (0_{p\times 1}^\top, \mathbf{b}_{,k}^\top)^\top$. Obviously, $||\widetilde{\mathbf{a}}_{,k} - \mathbf{a}_{,k}||_{1,A^c} \leq 3||\widetilde{\mathbf{a}}_{,k} - \mathbf{a}_{,k}||_{1,A}$. By applying the restricted eigenvalue result in Lemma 1,

$$\frac{1}{n}(\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k})^\top \mathbf{Z}^\top \mathbf{W}\mathbf{Z}(\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}) \geq \frac{1}{n}(\widetilde{\mathbf{a}}_{,k}-\mathbf{a}_{,k})^\top \Gamma_n(\widetilde{\mathbf{a}}_{,k}-\mathbf{a}_{,k}) > \frac{1}{2}c_*\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|^2_{2,K_0}. \quad \text{(S1.4)}$$

Combing (S1.3), (S1.4), and Jensen's inequality

$$\frac{1}{2}c_*\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|^2_{2,K_0} \leq 3\lambda_k\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|_{1,K_0} \leq 3\lambda_k\sqrt{|K_0|}\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|_{2,K_0}\ ,$$

we have that

$$\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|_{2,K_0} \leq 6\lambda_k\sqrt{|K_0|}/c_*\ .$$

Therefore, $\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|_1 \leq 4\|\widetilde{\mathbf{b}}_{,k}-\mathbf{b}_{,k}\|_{1,K_0} \leq 24\lambda_k|K_0|/c_*$. □

Lemma 3. *Suppose that the event* $\Omega_2 = \{\|(\mathbf{X},\mathbf{Z})^\top \mathbf{W}\epsilon/n\|_\infty < \lambda_0/2\}$ *and the conditions in Lemma 1 hold. Then we have*

$$\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_1 \leq 24\lambda_0|A_0|/c_*. \quad \text{(S1.5)}$$

Assume that the conditions in Lemma 2 hold. Then we have

$$\|\frac{1}{n}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\|_1 \leq 24p\lambda_0\lambda_{*,p}|A_0|/c_*, \quad \text{(S1.6)}$$

where $\lambda_{*,p} = \max\{\lambda_k, 1 \leq k \leq p\}$.

Proof. The proof of (S1.5) is similar to that of Lemma 2 and is omitted here. Below we prove (S1.6). For $k = 1, 2, \ldots, p$, the KKT conditions for the second objective function in (3) within the main text are

$$\begin{cases} \frac{1}{n}\mathbf{z}_{,j}^{\top}\mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\widetilde{\mathbf{b}}_{,k}) = \mathrm{sgn}(\widetilde{b}_{k,j})\lambda_k, & \text{if } \widetilde{b}_{k,j} \neq 0; \\ \frac{1}{n}|\mathbf{z}_{,j}^{\top}\mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\widetilde{\mathbf{b}}_{,k})| \leq \lambda_k, & \text{if } \widetilde{b}_{k,j} = 0. \end{cases}$$

Then, $\frac{1}{n}|\widetilde{\mathbf{x}}_{,k}^{\top}\mathbf{W}\mathbf{z}_{,j}| \leq \lambda_k$ for $j = 1, 2, \ldots, q$ and $k = 1, 2, \ldots, p$. Hence

$$|\frac{1}{n}\widetilde{\mathbf{x}}_{,k}^{\top}\mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)| \leq \lambda_k||\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||_1, \quad \text{for } k = 1, 2, \ldots, p.$$

Then, with the upper bound of $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1$, $\|\frac{1}{n}\widetilde{\mathbf{X}}^{\top}\mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|_1 \leq \sum_{k=1}^{p}\lambda_k\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq 24p\lambda_0\lambda_{*,p}|A_0|/c_*$. □

Lemma 4. *Under Assumptions 1–5 and $|A_0|\sqrt{\log q/n} = o(1)$, we have that for each $k \in \{1, 2, \cdots, p\}$, there exist positive constants $\hbar$ and ℓ,*

$$P\left(\|\mathbf{Z}^{\top}\mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\mathbf{b}_{,k})/n\|_{\infty} > t\right) \leq \hbar\exp(-\ell n t^2 + \log q), \quad \text{(S1.7)}$$

$$P\left(\|(\mathbf{X}, \mathbf{Z})^{\top}\mathbf{W}\boldsymbol{\epsilon}/n\|_{\infty} > t\right) \leq \hbar\exp(-\ell n t^2 + \log q), \quad \text{(S1.8)}$$

and $\max_{ij}|\Gamma_{n,ij} - \Gamma_{ij}| = O_p(\sqrt{\log q/n})$ *in Lemma 1 holds.*

Proof. Following Stute (1996), we have that for any given $j \in \{1, \cdots, q\}$ and $k \in \{1, \cdots, p\}$, the central limit theorem holds for $n^{-1/2}\mathbf{x}_{,k}^{\top}\mathbf{W}\boldsymbol{\epsilon}$, $n^{-1/2}\mathbf{z}_{,j}^{\top}\mathbf{W}\boldsymbol{\epsilon}$ and $n^{-1/2}\mathbf{z}_{,j}^{\top}\mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\mathbf{b}_{,k})$ under Assumptions 1–3 and 5. If the convergence properties for the above variables are uniform, then there exist positive constants $\hbar$ and ℓ such that $P(|\mathbf{z}_{,j}^{\top}\mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\mathbf{b}_{,k})/n| > t) \leq \hbar\exp(-\ell n t^2)$ and $P(|\mathbf{z}_{,j}^{\top}\mathbf{W}\boldsymbol{\epsilon}/n| > t) \leq \hbar\exp(-\ell n t^2)$ for any j and k. Therefore with the Bonferroni's inequality and a fixed p, we can obtain (S1.7) and (S1.8). The proof of $\max_{ij}|\Gamma_{n,ij} - \Gamma_{ij}| = O_p(\sqrt{\log q/n})$ follows a similar way. Below we employ the empirical process technique to establish the uniform central limit theorem.

Let Π be the space of parameter vectors for the family of X_k, Z_j, and ϵ. By Assumption 3, $X_k\epsilon$, $Z_j\epsilon$, and $Z_j(X_k - Z^\top \mathbf{b}_{,k})$ are sub-exponential random variables indexed by parameters in $\Pi \times \Pi$. Hence each of $X_k\epsilon$, $Z_j\epsilon$, and $Z_j(X_k - Z^\top \mathbf{b}_{,k})$ can be written as a function, indexed by $\boldsymbol{\pi} \in \Pi \times \Pi$, of some standard random variables in the sub-exponential family. It is clear that the Orlicz norm $\|\Psi\|_\psi < \infty$ for some sub-exponential random variable Ψ and $\psi(x) = \exp(x) - 1$ (van de Geer and Lederer, 2013). Define the semi-metric $\Delta(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$ in $\Pi \times \Pi$ using their corresponding sub-exponential random variables Ψ_{π_1} and Ψ_{π_2} as $\Delta(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2) = \|\Psi_{\pi_1} - \Psi_{\pi_2}\|_\psi$. By the previous arguments, the metric space $(\Pi \times \Pi, \Delta)$ is bounded. Following Lemma 19.15 of van der Vaart (1998), the covering number $N(\varepsilon, \Pi \times \Pi, L_2(Q))$ is bounded by a polynomial in $1/\varepsilon$ due to the finiteness of c_3 for all sub-exponential probability measure Q. Hence, the uniform entropy integral $J(1, \Pi \times \Pi, L_2)$ is finite (van der Vaart and Wellner, 2000). Following Bae and Kim (2003), we have that the uniform central limit theorem holds. The lemma follows from the above arguments. $\square$

S2 Proof of Theorem 1

First consider the events

$$\Omega_1 = \{\|\mathbf{Z}^\top \mathbf{W}(\mathbf{x}_{,k} - \ \mathbf{Z}\mathbf{b}_{,k})/n\|_\infty < \lambda_k/2, \text{ for } k = 1, \cdots, p\}$$

and

$$\Omega_2 = \{\|(\mathbf{X}, \mathbf{Z})^\top \mathbf{W}\boldsymbol{\epsilon}/n\|_\infty < \lambda_0/2\}.$$

Under condition $\min_k \lambda_k > M\sqrt{\log q/n}$ with a large enough M, we have $P(\|\mathbf{Z}^\top \mathbf{W}(\mathbf{x}_{,k} - \mathbf{Z}\mathbf{b}_{,k})/n\|_\infty > \lambda_k/2) \to 0$ as $n \to \infty$ by Lemma 4. Together with the condition that p is fixed and $P(\Omega_1) \leq 1$, we have $P(\Omega_1) \to 1$ as $n \to \infty$. Similarly, we can also obtain $P(\Omega_2) \to 1$ under the subgaussian condition and $\lambda_0 > M\sqrt{\log q/n}$. Based on the above discussions, we have

$$\lim_{n\to\infty} P(\Omega_1 \cup \Omega_2) = 1. \tag{S2.9}$$

By conditions $\sqrt{n}\lambda_{*,p}\lambda_0|A_0| \to 0$ and $\min_k \lambda_k > M\sqrt{\log q/n}$, we have $|A_0| \log q/\sqrt{n} \to 0$. Then obviously $|A_0|\sqrt{\log q/n} \to 0$ holds. Together with (S2.9) and Assumptions 1–4, the results in Lemma 2 and 3 can be used in the following proof.

From the definition of $\widetilde{\boldsymbol{\beta}}$ in (4) within the main text, we have

$$\begin{aligned}
\mathbf{0} &= \frac{1}{n}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}(\mathbf{y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}-\mathbf{Z}\widetilde{\boldsymbol{\theta}}). \\
\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\boldsymbol{\epsilon} &= \frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&\quad +\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0).
\end{aligned}$$

The left hand side

$$\begin{aligned}
LHS &= \frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\boldsymbol{\epsilon}+\frac{1}{\sqrt{n}}(\mathbf{B}_0-\widetilde{\mathbf{B}})^\top \mathbf{Z}^\top \mathbf{W}\boldsymbol{\epsilon} \\
&\overset{def}{=} A_n + B_n.
\end{aligned}$$

The right hand side

$$\begin{aligned}
RHS &= \frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
&\overset{def}{=} C_n + D_n.
\end{aligned}$$

Obviously, $A_n - C_n = -B_n + D_n$. Together with the results in Lemmas 2-3,

$$\begin{aligned}
& \|\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\boldsymbol{\epsilon}-\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 \\
= & \|-\frac{1}{\sqrt{n}}(\mathbf{B}_0-\widetilde{\mathbf{B}})^\top \mathbf{Z}^\top \mathbf{W}\boldsymbol{\epsilon}+\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\|_1 \\
\leq & \|-\frac{1}{\sqrt{n}}(\mathbf{B}_0-\widetilde{\mathbf{B}})^\top \mathbf{Z}^\top \mathbf{W}\boldsymbol{\epsilon}\|_1+\|\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\|_1 \\
\leq & \sqrt{n}\|\mathbf{B}_0-\widetilde{\mathbf{B}}\|_1 \cdot \|\mathbf{Z}^\top \mathbf{W}\boldsymbol{\epsilon}/n\|_\infty + \sqrt{n}\|(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)/n||_1 \\
\leq & 12p\sqrt{n}\lambda_{*,p}\lambda_0|K_0|/c_* + 24p\sqrt{n}\lambda_{*,p}\lambda_0|A_0|/c_* \\
\lesssim & \sqrt{n}\lambda_{*,p}\lambda_0|A_0|.
\end{aligned}$$

This converges to zero since $\sqrt{n}\lambda_{*,p}\lambda_0|A_0| \to 0$. As a result,

$$\|A_n - C_n\|_1 \xrightarrow{p} 0. \tag{S2.10}$$

Because $\|A_n\|_2 = \|\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\boldsymbol{\epsilon}\|_2$ is bounded in probability, $\|C_n\|_2 = \|\frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\widetilde{\mathbf{B}})^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_2$ is bounded in probability. Note that

$$C_n = \frac{1}{\sqrt{n}}(\mathbf{X}-\mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \frac{1}{\sqrt{n}}(\mathbf{B}_0-\widetilde{\mathbf{B}})^\top \mathbf{Z}^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \overset{def}{=} E_n + F_n.$$

Recall the definition of $\mathbf{B}_0$. Let $J_0^c = \{p+1, \cdots, p+q\}$. In fact,

$$\frac{1}{n}(\mathbf{X}-\mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\mathbf{X} = \Gamma_{J_0,J_0} - \Gamma^\top_{K_0^+,J_0}\Gamma^{-1}_{K_0^+,K_0^+}\Gamma_{K_0^+,J_0} + R_{n1},$$

where $R_{n1} = \mathbf{X}^\top \mathbf{W}\mathbf{X}/n - \Gamma_{J_0,J_0} + \mathbf{B}_0^\top(\Gamma_{J_0^c,J_0} - \mathbf{Z}^\top \mathbf{W}\mathbf{X}/n)$. From the sparsity of $\mathbf{B}_0$ and Lemma 1, we can obtain that $\|R_{n1}\|_\infty = O_p(|K_0|/\sqrt{n})$. Thus, for the term E_n, we have

$$\begin{aligned}
\|E_n\|_1 &\geq \|(\Gamma_{J_0,J_0} - \Gamma^\top_{K_0^+,J_0}\Gamma^{-1}_{K_0^+,K_0^+}\Gamma_{K_0^+,J_0})\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 - \|R_{n1}\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 \\
&\geq c_*\|\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 - O_p(|K_0|/\sqrt{n})\|\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1.
\end{aligned}$$

For F_n, by Lemma 2 and $\|\Gamma\|_\infty = O(1)$, we have

$$\begin{aligned}
\|F_n\|_1 &\leq \|(\mathbf{B}_0-\widetilde{\mathbf{B}})^\top \Gamma_{J_0^c,J_0}\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 \\
&\quad + \|(\mathbf{B}_0-\widetilde{\mathbf{B}})^\top(\mathbf{Z}^\top \mathbf{W}\mathbf{X}/n - \Gamma_{J_0^c,J_0})\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 \\
&\leq \{O(1) + O_p(\sqrt{\log q/n})\}\max_k \|\widetilde{\mathbf{b}}_{,k} - \mathbf{b}_{,k}\|_1\|\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1 \\
&\lesssim \lambda_{*,p}|K_0|\|\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\|_1.
\end{aligned}$$

As a result,

$$\begin{aligned}\|C_n\|_1 &\geq \|E_n\|_1 - \|F_n\|_1 \\ &\geq (c_* - O_p(|K_0|/\sqrt{n} + \lambda_{*,p}|K_0|) \cdot ||\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||_1.\end{aligned}$$

Obviously, $c_* - O_p(|K_0|/\sqrt{n} + \lambda_{*,p}|K_0| \xrightarrow{p} c_*$. Hence with a fixed p, $\|\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\|_2^2$ is bounded in probability. Then $\|F_n\|_1 \xrightarrow{p} 0$. Therefore, $\|C_n - E_n\|_1 \xrightarrow{p} 0$. Since $\|A_n - C_n\|_1 \xrightarrow{p} 0$ in (S2.10), we have $\|A_n - E_n\|_1 \xrightarrow{p} 0$. That is,

$$\|\frac{1}{\sqrt{n}}(\mathbf{X} - \mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\boldsymbol{\epsilon} - \frac{1}{\sqrt{n}}(\mathbf{X} - \mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\|_1 \xrightarrow{p} 0. \quad \text{(S2.11)}$$

Applying Theorem 3.1 of Stute (1996), we can obtain

$$\frac{1}{\sqrt{n}}(\mathbf{X} - \mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\boldsymbol{\epsilon} \xrightarrow{D} N(0, \boldsymbol{\Sigma}_1)$$

under Assumption 5. Under similar conditions, by Corollary 1.8 of Stute (1993), $\frac{1}{\sqrt{n}}(\mathbf{X} - \mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\mathbf{X} \xrightarrow{p} \boldsymbol{\Sigma}_0$. Hence, using the result in (S2.11) and Slutsky Lemma, we have

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}),$$

which concludes the proof. □

Bibliography

Bae, J. and Kim, S. (2003). The uniform central limit theorem for the kaplan-meier integral process. *Bulletin of the Australian Mathematical Society* **67**, 467–480.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* **186**, 325 – 344.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis* **45**, 89–103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics* **23**, 461–471.

van de Geer, S. and Bühlmann, P. (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics* **3**, 1360–1392.

van de Geer, S. and Lederer, J. (2013). The bernstein–orlicz norm and deviation inequalities. *Probability Theory and Related Fields* **157**, 225–250.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge university press.

van der Vaart, A. W. and Wellner, J. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.