Online Supplementary Materials for "Statistical Inference for Cox Proportional Hazards Models with a Diverging Number of Covariates"

Lu Xia[a], Bin Nan[b], and Yi Li[c]

[a]Department of Biostatistics, University of Washington, Seattle, WA; [b]Department of Statistics, University of Californina, Irvine, CA; [c]Department of Biostatistics, University of Michigan, Ann Arbor, MI

We provide detailed proofs for the lemmas presented in the Appendix of the article, as well as patient characteristics of the Boston Lung Cancer Study Cohort data analyzed in Section 5.

1 Technical proofs for the lemmas

Lemma A1 characterizes the difference between $\widehat{\eta}_n(t;\beta^0)$ and $\eta_0(t;\beta^0)$, which is needed to prove the asymptotic distribution for the leading term $\sqrt{n}c^T\Theta_{\beta^0}\dot{\ell}_n(\beta^0)$ as well as to establish the convergence rate for $\widehat{\Sigma}-\Sigma_{\beta^0}$.

Lemma A1. *Under Assumptions 1–3, we have*

$$\sup_{t\in[0,\tau]}|\widehat{\mu}_0(t;\beta^0)-\mu_0(t;\beta^0)|=\mathcal{O}_P(\sqrt{\log(p)/n}),$$
$$\sup_{t\in[0,\tau]}\|\widehat{\mu}_1(t;\beta^0)-\mu_1(t;\beta^0)\|_\infty=\mathcal{O}_P(\sqrt{\log(p)/n}),$$
$$\sup_{t\in[0,\tau]}\|\widehat{\eta}_n(t;\beta^0)-\eta_0(t;\beta^0)\|_\infty=\mathcal{O}_P(\sqrt{\log(p)/n}).$$

Proof of Lemma A1. The first two statements in the conclusion are similar to those in Kong and Nan (2014), but with differing setups. Consider a class of functions of $y\geq 0$ and $x\in\mathbb{R}^p$ indexed by t, $\mathcal{F}_0=\{1(y\geq t)\exp(x^T\beta^0):t\in[0,\tau]\}$. For any $0<\epsilon<1$, consider the cumulative distribution function for Y and take an positive integer $m<2/\epsilon$

and a sequence of points $0 = t_0 < t_1 < \cdots < t_{m-1} < t_m = \infty$ such that $P(t_i < Y \leq t_{i+1}) < \epsilon,\ i = 0, 1, \ldots, m-1$. For each $i = 1, \cdots, m$, define the bracket $[L_i, U_i]$, where $L_i(x, y) = 1(y \geq t_i)\exp(x^T\beta^0)$ and $U_i(x, y) = 1(y > t_{i-1})\exp(x^T\beta^0)$. We have $L_i(x, y) \leq 1(y \geq t)\exp(x^T\beta^0) \leq U_i(x, y)$ for $t_{i-1} < t \leq t_i$, and

$$[E\{U_i(X,Y) - L_i(X,Y)\}^2]^{1/2} = [E\{1(t_{i-1} < Y < t_i)\exp(2X^T\beta^0)\}]^{1/2} \leq e^{K_1}\sqrt{\epsilon},$$
$$E|U_i(X,Y) - L_i(X,Y)| = E\{1(t_{i-1} < Y < t_i)\exp(X^T\beta^0)\} \leq e^{K_1}\epsilon.$$

Then the bracketing numbers van der Vaart (1998) satisfy

$$N_{[]}(e^{K_1}\sqrt{\epsilon}, \mathcal{F}_0, L_2(P)) \leq \frac{2}{\epsilon}, \quad N_{[]}(e^{K_1}\epsilon, \mathcal{F}_0, L_1(P)) \leq \frac{2}{\epsilon},$$

or equivalently,

$$N_{[]}(\epsilon, \mathcal{F}_0, L_2(P)) \leq \frac{2e^{2K_1}}{\epsilon^2}, \quad N_{[]}(\epsilon, \mathcal{F}_0, L_1(P)) \leq \frac{2e^{K_1}}{\epsilon} < \infty.$$

By the Glivenko-Cantelli Theorem and the Donsker Theorem (van der Vaart, 1998), the class of $\mathcal{F}_0$ is P-Glivenko-Cantelli and P-Donsker. So $\sup_{t\in[0,\tau]} |\widehat{\mu}_0(t;\beta^0) - \mu_0(t;\beta^0)| \overset{a.s.}{\to} 0$, and moreover, by Theorem 2.14.9 of van der Vaart and Wellner (1996) with $V = 2$,

$$P\left(\sqrt{n}\sup_{t\in[0,\tau]} |\widehat{\mu}_0(t;\beta^0) - \mu_0(t;\beta^0)| > s\right) \leq De^{-s^2},$$

for every $s > 0$ and a constant $D > 0$ that only depends on K_1. Setting $s = \sqrt{2\log(p)}$ implies that

$$\sup_{t\in[0,\tau]} |\widehat{\mu}_0(t;\beta^0) - \mu_0(t;\beta^0)| = \mathcal{O}_P(\sqrt{\log(p)/n}).$$

For the second statement, we consider the classes of functions of $(x, y) = (x_1, \cdots, x_p, y)$ indexed by t,

$$\mathcal{F}_1^k = \{1(y \geq t)e^{x^T\beta^0}x_k : t \in [0,\tau]\},\ k = 1, \cdots, p.$$

Since $|e^{x^T\beta^0}x_k| \leq Ke^{K_1}$, similarly we have

$$N_{[]}(\epsilon, \mathcal{F}_1^k, L_2(P)) \leq \left(\frac{\sqrt{2}e^{K_1}K}{\epsilon}\right)^2.$$

By Theorem 2.14.9 of van der Vaart and Wellner (1996) with $V = 2$, we have

$$P\left(\sqrt{n}\sup_{t\in[0,\tau]}|\widehat{\mu}_{1k}(t;\beta^0) - \mu_{1k}(t;\beta^0)| > s\right) \le D' s^2 e^{-2s^2} \le D' e^{-1} e^{-s^2}$$

for every $s > 0$, where D' is a constant that only depends on K and K_1, and $\widehat{\mu}_{1k}$ and μ_{1k} are the kth components of $\widehat{\mu}_1$ and μ_1, respectively. Thus,

$$\begin{aligned}
&P\left(\sqrt{n}\sup_{t\in[0,\tau]}\|\widehat{\mu}_1(t;\beta^0) - \mu_1(t;\beta^0)\|_\infty > s\right) \\
&\le P\left(\bigcup_{k=1}^{p}\left\{\sqrt{n}\sup_{t\in[0,\tau]}|\widehat{\mu}_{1k}(t;\beta^0) - \mu_{1k}(t;\beta^0)| > s\right\}\right) \\
&\le pD'e^{-s^2}.
\end{aligned}$$

For example, taking $s = \sqrt{2\log(p)}$ would complete the proof for $\sup_{t\in[0,\tau]}\|\widehat{\mu}_1(t;\beta^0) - \mu_1(t;\beta^0)\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n})$.

Finally, we rewrite

$$\begin{aligned}
\widehat{\eta}_n(t;\beta^0) - \eta_0(t;\beta^0) &= \frac{\widehat{\mu}_1(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)} - \frac{\mu_1(t;\beta^0)}{\mu_0(t;\beta^0)} \\
&= \frac{\widehat{\mu}_1(t;\beta^0)}{\mu_0(t;\beta^0)} - \frac{\mu_1(t;\beta^0)}{\mu_0(t;\beta^0)} + \frac{\widehat{\mu}_1(t;\beta^0)}{\mu_0(t;\beta^0)}\left(\frac{\mu_0(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)} - 1\right).
\end{aligned}$$

By Assumptions 1–3, $\mu_0(t;\beta^0) \ge e^{-K_1}\pi_0 > 0$ and $\sup_{t\in[0,\tau]}\|\widehat{\mu}_1(t;\beta^0)\|_\infty = \mathcal{O}_P(1)$. Also, since

$$\inf_{t\in[0,\tau]}\widehat{\mu}_0(t;\beta^0) \ge \mu_0(t;\beta^0) - |\widehat{\mu}_0(t;\beta^0) - \mu_0(t;\beta^0)| \ge e^{-K_1}\pi_0 - \sup_{t\in[0,\tau]}|\widehat{\mu}_0(t;\beta^0) - \mu_0(t;\beta^0)| > e^{-K_1}\frac{\pi_0}{2}$$

almost surely, we have

$$\begin{aligned}
&\sup_{t\in[0,\tau]}\left\|\frac{\widehat{\mu}_1(t;\beta^0)}{\mu_0(t;\beta^0)}\left(\frac{\mu_0(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)} - 1\right)\right\|_\infty \\
&\le \sup_{t\in[0,\tau]}\left\|\frac{\widehat{\mu}_1(t;\beta^0)}{\mu_0(t;\beta^0)}\right\|_\infty \cdot \sup_{t\in[0,\tau]}\left|\frac{\mu_0(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)} - 1\right| \\
&\le \mathcal{O}_P(1)\sup_{t\in[0,\tau]}\left|\mu_0(t;\beta^0) - \widehat{\mu}_0(t;\beta^0)\right| = \mathcal{O}_P(\sqrt{\log(p)/n}).
\end{aligned}$$

Therefore,

$$\begin{aligned}\sup_{t\in[0,\tau]}\|\widehat{\eta}_n(t;\beta^0)-\eta_0(t;\beta^0)\|_\infty &\le \sup_{t\in[0,\tau]}\left\|\frac{1}{\mu_0(t;\beta^0)}\left(\widehat{\mu}_1(t;\beta^0)-\mu_1(t;\beta^0)\right)\right\|_\infty\\ &\quad+\sup_{t\in[0,\tau]}\left\|\frac{\widehat{\mu}_1(t;\beta^0)}{\mu_0(t;\beta^0)}\left(\frac{\mu_0(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)}-1\right)\right\|_\infty\\ &=\mathcal{O}_P(\sqrt{\log(p)/n}).\end{aligned}$$

□

Lemma A2 establishes the asymptotic distribution for the leading term $-c^T\Theta_{\beta^0}\dot{\ell}_n(\beta^0)$ in the decomposition of $c^T(\widehat{b}-\beta^0)$.

Lemma A2. *Assume $p^2\log(p)/n\to 0$. Under Assumptions 1–5, for any $c\in\mathbb{R}^p$ such that $\|c\|_2=1$ and $\|c\|_1\le a_*$ with some absolute constant $a_*>0$,*

$$\frac{\sqrt{n}c^T\Theta_{\beta^0}\dot{\ell}_n(\beta^0)}{\sqrt{c^T\Theta_{\beta^0}c}}\xrightarrow{\mathcal{D}}N(0,1).$$

Proof of Lemma A2. Using notation of martingales, we rewrite

$$\begin{aligned}\frac{-\sqrt{n}c^T\Theta_{\beta^0}\dot{\ell}_n(\beta^0)}{\sqrt{c^T\Theta_{\beta^0}c}} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_i-\frac{\widehat{\mu}_1(Y_i;\beta^0)}{\widehat{\mu}_0(Y_i;\beta^0)}\right\}\delta_i\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_i-\frac{\widehat{\mu}_1(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)}\right\}dN_i(t)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_i-\frac{\widehat{\mu}_1(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)}\right\}dM_i(t).\end{aligned}$$

Let $Q_i(t)=\frac{1}{\sqrt{n}}\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_i-\frac{\widehat{\mu}_1(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)}\right\}$, $i=1,\dots,n$, which are predictable with respect to the filtration $\mathcal{F}$. Then

$$\frac{-\sqrt{n}c^T\Theta_{\beta^0}\dot{\ell}_n(\beta^0)}{\sqrt{c^T\Theta_{\beta^0}c}}=\sum_{i=1}^n\int_0^\tau Q_i(t)dM_i(t). \tag{S1}$$

For any $t \in [0,\tau]$, let $U(t) = \sum_{i=1}^n \int_0^t Q_i(u) dM_i(u)$, whose predictable variation process is

$$\begin{aligned}
\langle U\rangle(t) &= \sum_{i=1}^n \int_0^t Q_i(u)^2 1(Y_i \geq u) e^{X_i^T \beta^0} dH_0(u) \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0} c}\left\{X_i - \frac{\widehat{\mu}_1(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)}\right\}^{\otimes 2} \Theta_{\beta^0} c 1(Y_i \geq u) e^{X_i^T\beta^0} dH_0(u) \\
&= \frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0} c}\left[\int_0^t \left\{\widehat{\mu}_2(u;\beta^0) - \frac{\widehat{\mu}_1(u;\beta^0)\widehat{\mu}_1(u;\beta^0)^T}{\widehat{\mu}_0(u;\beta^0)}\right\} dH_0(u)\right] \Theta_{\beta^0} c
\end{aligned}$$

Similar to the proof in Lemma A1, we can show that $\sup_{t\in[0,\tau]} \|\widehat{\mu}_2(t;\beta^0) - \mu_2(t;\beta^0)\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n})$, and thus

$$\begin{aligned}
\left\|\int_0^t \{\mu_2(u;\beta^0) - \widehat{\mu}_2(u;\beta^0)\} h_0(u) du\right\|_\infty &\leq \sup_{u\in[0,\tau]} \|\widehat{\mu}_2(u;\beta^0) - \mu_2(u;\beta^0)\|_\infty \int_0^\tau h_0(u) du \\
&= \mathcal{O}_P(\sqrt{\log(p)/n}). \qquad \text{(S2)}
\end{aligned}$$

Since

$$\frac{\widehat{\mu}_1\widehat{\mu}_1^T}{\widehat{\mu}_0} - \frac{\mu_1\mu_1^T}{\mu_0} = \frac{\widehat{\mu}_1\widehat{\mu}_1^T}{\widehat{\mu}_0\mu_0}(\mu_0 - \widehat{\mu}_0) + \frac{1}{\mu_0}[(\widehat{\mu}_1 - \mu_1)\widehat{\mu}_1^T + \mu_1(\widehat{\mu}_1 - \mu_1)^T],$$

by Assumption 1 and Lemma A1,

$$\left\|\int_0^t \left\{\frac{\widehat{\mu}_1(u;\beta^0)\widehat{\mu}_1^T(u;\beta^0)}{\widehat{\mu}_0(u;\beta^0)} - \frac{\mu_1(u;\beta^0)\mu_1^T(u;\beta^0)}{\mu_0(u;\beta^0)}\right\} h_0(u) du\right\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n}). \qquad \text{(S3)}$$

Combining (S2) and (S3), we have that, uniformly for all $t \in [0,\tau]$,

$$\begin{aligned}
&\left\|\int_0^t \left\{\widehat{\mu}_2(u;\beta^0) - \frac{\widehat{\mu}_1(u;\beta^0)\widehat{\mu}_1(u;\beta^0)^T}{\widehat{\mu}_0(u;\beta^0)}\right\} dH_0(u) - \right. \\
&\left. \int_0^t \left\{\mu_2(u;\beta^0) - \frac{\mu_1(u;\beta^0)\mu_1(u;\beta^0)^T}{\mu_0(u;\beta^0)}\right\} dH_0(u)\right\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n}).
\end{aligned}$$

Then

$$\begin{aligned}
&\left|\langle U\rangle(t) - \frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0} c}\left[\int_0^t \left\{\mu_2(u;\beta^0) - \frac{\mu_1(u;\beta^0)\mu_1(u;\beta^0)^T}{\mu_0(u;\beta^0)}\right\} dH_0(u)\right] \Theta_{\beta^0} c\right| \\
\leq& \zeta_{\min}^{-1}(\|c\|_1 \|\Theta_{\beta^0}\|_{1,1})^2 \mathcal{O}_P(\sqrt{\log(p)/n}) \\
\leq& \zeta_{\min}^{-1} a_*^2 p \zeta_{\max}^2 \mathcal{O}_P(\sqrt{\log(p)/n}) \to_P 0
\end{aligned}$$

if $p^2\log(p)/n \to 0$. By Assumption 4, $\langle U\rangle(t) - v(t;c) \to_P 0$.

Now we check the Lindeberg condition. For any $\epsilon > 0$, define the truncated process

$$U_\epsilon(t) = \sum_{i=1}^n \int_0^t Q_i(u)1\{|Q_i(u)| > \epsilon\}dM_i(u),$$

with a predictable variation process:

$$\begin{aligned}\langle U_\epsilon\rangle(t) &= \sum_{i=1}^n \int_0^t Q_i^2(u)1\{|Q_i(u)| > \epsilon\}1(Y_i \geq u)e^{X_i^T\beta^0}h_0(u)du \\ &= \sum_{i=1}^n \int_0^t Q_i^2(u)1\{|\sqrt{n}Q_i(u)| > \sqrt{n}\epsilon\}1(Y_i \geq u)e^{X_i^T\beta^0}h_0(u)du.\end{aligned}$$

Let $Q_{\max} = \sup_{t\in[0,\tau]}\max_{1\leq i\leq n}|\sqrt{n}Q_i(t)|$, then $1\{|\sqrt{n}Q_i(u)| > \sqrt{n}\epsilon\} \leq 1\{Q_{\max} > \sqrt{n}\epsilon\}$. By Assumption 1,

$$\sup_{t\in[0,\tau]}\max_{1\leq i\leq n}\left|\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_i - \frac{\widehat{\mu}_1(t;\beta^0)}{\widehat{\mu}_0(t;\beta^0)}\right\}\right| \leq \zeta_{\min}^{-1/2}\|c\|_1\|\Theta_{\beta^0}\|_{1,1}2K = \mathcal{O}(\sqrt{p}),$$

and $Q_{\max} = \mathcal{O}(\sqrt{p})$. When $p/n \to 0$, $1\{Q_{\max} > \sqrt{n}\epsilon\} = 0$ almost surely. Thus $\langle U_\epsilon\rangle(t) \to_P 0$. Finally, by the martingale central limit theorem, the asymptotic normality follows. □

Lemma A3 provides the theoretical properties of the lasso estimator in the Cox model. This is a direct result from Theorem 1 in Kong and Nan (2014), and thus the proof is omitted.

Lemma A3. *Under Assumptions 1–5, for the lasso estimator $\widehat{\beta}$, we have*

$$\|\widehat{\beta} - \beta^0\|_1 = \mathcal{O}_P(s_0\lambda_n), \quad \frac{1}{n}\sum_{i=1}^n |X_i^T(\widehat{\beta} - \beta^0)|^2 = \mathcal{O}_P(s_0\lambda_n^2),$$

where $s_0 = |\{j : \beta_j^0 \neq 0, j = 1, \cdots, p\}|$ is the true model size.

Lemma A4. *Under Assumptions 1–5, if $\lambda_n \asymp \sqrt{\log(p)/n}$, with probability going to 1, we have $\|\Theta_{\beta^0}\widehat{\Sigma} - I_p\|_\infty \leq \gamma_n$, for $\gamma_n \asymp \|\Theta_{\beta^0}\|_{1,1}s_0\lambda_n$.*

Lemma A4 shows that, unlike linear models with the tuning parameter in the constraint taking the order of $\sqrt{\log(p)/n}$, the Cox model requires a potentially larger γ_n for the feasibility of Θ_{β^0} that depends on $\|\Theta_{\beta^0}\|_{1,1}$, as the information matrix involves the regression coefficients.

Proof of Lemma A4. Write $A_n = \dfrac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{X_i - \eta_0(t;\beta^0)\right\}^{\otimes 2} dN_i(t) - \Sigma_{\beta^0}$.

$$
\begin{aligned}
\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty \leq & \left\| \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left[\left\{X_i - \widehat{\eta}_n(t;\widehat{\beta})\right\}^{\otimes 2} - \left\{X_i - \eta_0(t;\beta^0)\right\}^{\otimes 2}\right] dN_i(t)\right\|_\infty \\
& + \left\| \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{X_i - \eta_0(t;\beta^0)\right\}^{\otimes 2} dN_i(t) - \Sigma_{\beta^0}\right\|_\infty \\
\leq & \left\| \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{X_i - \widehat{\eta}_n(t;\widehat{\beta})\right\}\left\{\widehat{\eta}_n(t;\widehat{\beta}) - \eta_0(t;\beta^0)\right\}^T dN_i(t)\right\|_\infty \\
& + \left\| \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widehat{\eta}_n(t;\widehat{\beta}) - \eta_0(t;\beta^0)\right\}\left\{X_i - \eta_0(t;\beta^0)\right\}^T dN_i(t)\right\|_\infty + \|A_n\|_\infty .
\end{aligned}
$$

Note that for all $t \in [0,\tau]$, $\|X_i - \widehat{\eta}_n(t;\hat{\beta})\|_\infty \leq 2K$ and $\|X_i - \eta_0(t;\beta^0)\|_\infty \leq 2K$. Then

$$
\begin{aligned}
\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty \leq & \frac{4K}{n}\sum_{i=1}^{n}\int_0^{\tau}\|\widehat{\eta}_n(t;\hat{\beta}) - \eta_0(t;\beta^0)\|_\infty dN_i(t) + \|A_n\|_\infty \\
\leq & \frac{4K}{n}\sum_{i=1}^{n}\int_0^{\tau}\|\widehat{\eta}_n(t;\hat{\beta}) - \widehat{\eta}_n(t;\beta^0)\|_\infty dN_i(t) \\
& + \frac{4K}{n}\sum_{i=1}^{n}\int_0^{\tau}\|\widehat{\eta}_n(t;\beta^0) - \eta_0(t;\beta^0)\|_\infty dN_i(t) + \|A_n\|_\infty . \qquad \text{(S4)}
\end{aligned}
$$

By the mean value theorem, for the jth component in $\widehat{\eta}_n$ (denoted by $\widehat{\eta}_{nj}$), there exists some $\bar{\beta}^{(j)}$ lying inside the segments connecting $\widehat{\beta}$ and β^0 such that

$$
\widehat{\eta}_{nj}(t;\widehat{\beta}) = \widehat{\eta}_{nj}(t;\beta^0) + \left[\left.\frac{\partial \widehat{\eta}_{nj}(t;\beta)}{\partial \beta}\right|_{\beta=\bar{\beta}^{(j)}}\right]^T (\widehat{\beta} - \beta^0).
$$

Consider β in a neighborhood of β^0, i.e. when $\|\beta - \beta^0\|_1 \leq \delta'$ for some $\delta' > 0$, $e^{X_i^T\beta} \leq e^{|X_i^T\beta|} \leq e^{|X_i^T\beta^0| + K\delta'} \leq e^{K_1 + K\delta'}$, and $e^{X_i^T\beta} \geq e^{-|X_i^T\beta|} \geq e^{-K_1 - K\delta'}$. Since $\{1(Y \geq t) : t \in [0,\tau]\}$ is P-Glivenko-Cantelli, $\sup_{t\in[0,\tau]} |\frac{1}{n}\sum_{i=1}^{n} 1(Y \geq t) - P(Y \geq t)| \overset{a.s.}{\to} 0$, and then uniformly for $t \in [0,\tau]$ and $\beta \in \{\beta : \|\beta - \beta^0\|_1 \le \delta'\}$,

$$\widehat{\mu}_0(t;\beta) \ge \frac{1}{n}\sum_{i=1}^n 1(Y_i \ge t)e^{-K_1-K\delta'} \overset{a.s.}{\to} P(Y \ge t)e^{-K_1-K\delta'} \ge \frac{\pi_0}{2}e^{-K_1-K\delta'}.$$

In this case, uniformly for $t \in [0,\tau]$ and $\beta \in \{\beta : \|\beta - \beta^0\|_1 \le \delta'\}$,

$$\begin{aligned}\left\|\frac{\partial\widehat{\eta}_n(t;\beta)}{\partial\beta^T}\right\|_\infty &= \left\|\frac{\widehat{\mu}_2(t;\beta)\widehat{\mu}_0(t;\beta) - \widehat{\mu}_1(t;\beta)\widehat{\mu}_1(t;\beta)^T}{\widehat{\mu}_0^2(t;\beta)}\right\|_\infty \\ &\le_{a.s.} \left(\frac{\pi_0}{2}e^{-K_1-K\delta'}\right)^{-2}\left\{e^{K_1+K\delta'}K^2 \cdot e^{K_1+K\delta'} + e^{2(K_1+K\delta')}K^2\right\} \\ &= \frac{8}{\pi_0^2}e^{4(K_1+K\delta')}K^2 < \infty,\end{aligned}$$

i.e. $\left\|\frac{\partial\widehat{\eta}_n(t;\beta)}{\partial\beta^T}\right\|_\infty$ is uniformly bounded almost surely. When $s_0\lambda_n \to 0$, we have $\|\widehat{\eta}_n(t;\widehat{\beta}) - \widehat{\eta}_n(t;\beta^0)\|_\infty \le \mathcal{O}_P(\|\widehat{\beta}-\beta^0\|_1) = \mathcal{O}_P(s_0\lambda_n)$ and the first term in (S4) is $\frac{4K}{n}\sum_{i=1}^n\int_0^\tau \|\widehat{\eta}_n(t;\widehat{\beta}) - \widehat{\eta}_n(t;\beta^0)\|_\infty dN_i(t) = \mathcal{O}_P(s_0\lambda_n)$.

For the second term in (S4), we use an argument from Lemma A1 that $\sup_{t\in[0,\tau]}\|\widehat{\eta}_n(t;\beta^0) - \eta_0(t;\beta^0)\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n})$ and then have

$$\begin{aligned}&\frac{4K}{n}\sum_{i=1}^n\int_0^\tau \|\widehat{\eta}_n(t;\beta^0) - \eta_0(t;\beta^0)\|_\infty dN_i(t) \\ \le&\frac{4K}{n}\sum_{i=1}^n\int_0^\tau \sup_{t\in[0,\tau]}\|\widehat{\eta}_n(t;\beta^0) - \eta_0(t;\beta^0)\|_\infty dN_i(t) \\ =&\mathcal{O}_P(\sqrt{\log(p)/n}).\end{aligned}$$

For the last term A_n, by Hoeffding's concentration inequality, we have for every $t > 0$ and $j,k = 1,\cdots,p$,

$$P\left(|A_n(j,k)| \ge t\right) \le 2\exp\{-nt^2/C'\},$$

where C' is a constant only depending on K^4. Since A_n is a symmetric matrix,

$$\begin{aligned}
P\left(\|A_n\|_\infty \geq t\right) &= P\left(\bigcup_{1\leq j\leq p, j\leq k\leq p} |A_n(j,k)| \geq t\right) \\
&\leq \sum_{j=1}^{p}\sum_{k=j}^{p} P\left(|A_n(j,k)| \geq t\right) \\
&\leq p(p+1)\exp\{-nt^2/C'\}.
\end{aligned}$$

So $\|A_n\|_\infty = \mathcal{O}_P\left(\sqrt{\log(p)/n}\right)$. Combining the three terms in (S4), we have $\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty \leq \mathcal{O}_P(s_0\lambda_n + \sqrt{\log(p)/n})$. Finally, we conclude that

$$\begin{aligned}
\|\Theta_{\beta^0}\widehat{\Sigma} - I_p\|_\infty &\leq \|\Theta_{\beta^0}\|_{1,1}\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty \\
&= \mathcal{O}_P\left(\|\Theta_{\beta^0}\|_{1,1}s_0\lambda_n + \|\Theta_{\beta^0}\|_{1,1}\sqrt{\log(p)/n}\right).
\end{aligned}$$

□

Lemma A5. *Assume* $\limsup_{n\to\infty} p\gamma_n \leq 1 - \epsilon'$ *for some* $\epsilon' \in (0,1)$*. Then, under the assumptions in Lemma A4,* $\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty = \mathcal{O}_P(\gamma_n\|\Theta_{\beta^0}\|_{1,1})$.

Proof of Lemma A5. Note that $\widehat{\Theta} - \Theta_{\beta^0} = \widehat{\Theta}(I_p - \widehat{\Sigma}\Theta_{\beta^0}) + (\widehat{\Theta}\widehat{\Sigma} - I_p)\Theta_{\beta^0}$, then on the event $\{\|\widehat{\Sigma}\Theta_{\beta^0} - I_p\|_\infty \leq \gamma_n\}$, we have

$$\begin{aligned}
\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty &\leq \|\widehat{\Theta}\|_{\infty,\infty}\|I_p - \widehat{\Sigma}\Theta_{\beta^0}\|_\infty + \|\widehat{\Theta}\widehat{\Sigma} - I_p\|_\infty\|\Theta_{\beta^0}\|_{1,1} \\
&\leq \gamma_n\|\widehat{\Theta}\|_{\infty,\infty} + \gamma_n\|\Theta_{\beta^0}\|_{1,1}.
\end{aligned}$$

Since $\|\widehat{\Theta}\|_{\infty,\infty} \leq \|\widehat{\Theta} - \Theta_{\beta^0}\|_{\infty,\infty} + \|\Theta_{\beta^0}\|_{\infty,\infty} \leq p\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty + \|\Theta_{\beta^0}\|_{1,1}$, we can obtain

$$\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty \leq \gamma_n\left(p\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty + \|\Theta_{\beta^0}\|_{1,1}\right) + \gamma_n\|\Theta_{\beta^0}\|_{1,1}.$$

When $\limsup_{n\to\infty}\gamma_n p \leq 1 - \epsilon' < 1$, then for n large enough,

$$\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty \leq 2\gamma_n\|\Theta_{\beta^0}\|_{1,1}/(1 - \gamma_n p) \asymp \gamma_n\|\Theta_{\beta^0}\|_{1,1}.$$

Therefore, by Lemma A4, $\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty = \mathcal{O}_P(\gamma_n\|\Theta_{\beta^0}\|_{1,1})$. □

Lemma A6. *Under Assumptions 1–3 and 5, for each $t > 0$,*

$$P\left(\|\dot{\ell}_n(\beta^0)\|_\infty > t\right) \le 2pe^{-nt^2/(8K^2)}.$$

Proof of Lemma A6. Noting that $\|X_i - \widehat{\eta}_n(t;\beta^0)\|_\infty \le 2K$ uniformly for all i, then Lemma A6 is a direct result of Lemma 3.3(ii) in Huang et al. (2013). $\square$

2 Further discussion on Assumption 4

Assumption 4 in the main text is a technical condition that ensures the convergence (in probability) of the variation process for the application of the martingale central limit theorem. Unlike the traditional application of the martingale central limit theorem to Cox models, the dimension of $\widetilde{\Sigma}_{\beta^0}(t)$ is $p \times p$, where p, in our setting, is diverging with n. Also, the deterministic scalar $c^T\Theta_{\beta^0}$ depends on n through the dimensions of c, Θ_{β^0}. To address the issue, we have formulated our Assumption 4 by designing a limiting function $v(t; c)$, which can be viewed as the "standardized" information number up to time point t.

We also examine conditions related to the martingale central limit theorem or other central limit theorems. For example, similar to van de Geer et al. (2014) and Fang et al. (2017), we may utilize the Lindeberg-Feller central limit theorem by making an assumption like $\|\Theta_{\beta^0}X_i\|_\infty = \mathcal{O}(1)$ uniformly for all $i = 1, \ldots, n$, which is analogous to $\|X_{\beta^0,-j}\gamma^0_{\beta^0,j}\|_\infty = \mathcal{O}(1)$ given in Theorem 3.3(iv) of van de Geer et al. (2014) or $\sup_{t\in[0,\tau]}\max_{i\in[n]}|X^T_{i,2:d}(t)w^*| = \mathcal{O}(1)$ given in Assumption 4 of Fang et al. (2017). However, the assumption that $\|\Theta_{\beta^0}X_i\|_\infty = \mathcal{O}(1)$ uniformly for all $i = 1, \ldots, n$ is fairly strong. In addition, such assumptions are effectively a sparse Θ_{β^0} assumption when p diverges with n and covariates X_i are uniformly bounded. A sparse Θ_{β^0} assumption is in fact what we try to avoid.

3 Additional Simulation Results

We conduct simulations to evaluate the performance of simultaneous inference, using the criterion of the empirical false discover proportions (FDP) defined as

$$\text{FDP} = \frac{\#\text{false discoveries}}{\max(1, \#\text{discoveries})},$$

where a discovery occurs if the adjusted Benjamini-Hochberg p-value for testing $\beta_j^0 = 0$ is less than a pre-specified level, i.e., 0.05 and 0.1 in this case. The simulation settings are the same as those used for Figure 2 and Figure 3 in Section 4 of the main text.

Figures S1 and S2 are related to independent and AR(1) ($\rho = 0.5$) covariance matrix for covariates. All the three debiased lasso methods maintain FDP well-controlled below the nominal levels and tend to be more conservative when p is large, especially for QP with an AR(1) covariance matrix.

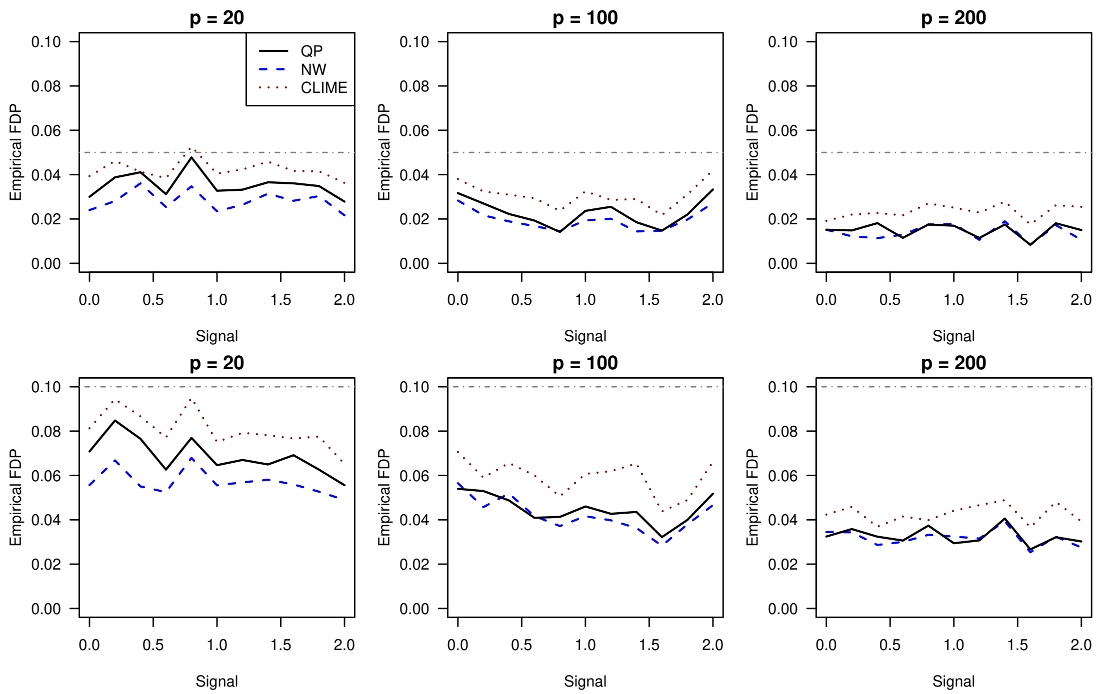

Figure S1: Empirical FDP with $n = 500$ and independent covariance matrix, as β_1^0 varies from 0 to 2. The top and bottom panels correspond to nominal false discovery rate levels of 0.05 and 0.1, respectively.

4 Boston Lung Cancer Study Cohort data

Table S1 shows the patient characteristics for the subset of the Boston Lung Cancer Study Cohort data analyzed in Section 5.

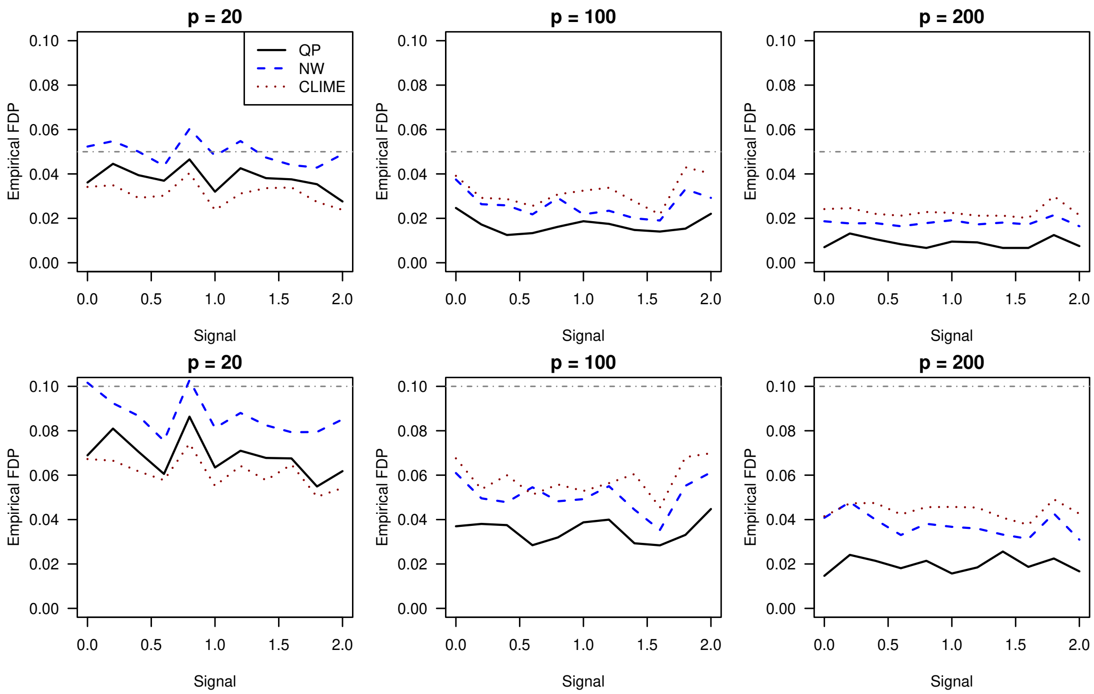

Figure S2: Empirical FDP with $n = 500$ and AR(1) covariance matrix ($\rho = 0.5$), as β_1^0 varies from 0 to 2. The top and bottom panels correspond to nominal false discovery rate levels of 0.05 and 0.1, respectively.

Table S1: Characteristics of $n = 561$ patients in the Boston Lung Cancer Study for survival analysis

Variable	Category / Unit	Count (%) / Mean (SD)
Age	Years old	60.0 (10.9)
Race	Caucasian	528 (94.1%)
	Others	33 (5.9%)
Education	No high school	79 (14.1%)
	High school	141 (25.1%)
	At least 1-2 years of college	341 (60.8%)
Gender	Male	215 (38.3%)
	Female	346 (61.7%)
Smoker	Current or recently quit	508 (90.6%)
	Never	53 (9.4%)
Histology	Adenocarcinoma	360 (64.2%)
	Squamous cell carcinoma	115 (20.5%)
	Large cell carcinoma	45 (8.0%)
	Unspecified	41 (7.3%)
Stage[a]	Early	243 (43.3%)
	Late	318 (56.7%)
Surgery	No	177 (31.6%)
	Yes	361 (64.3%)
Chemotherapy	No	300 (53.5%)
	Yes	238 (42.4%)
Radiation	No	332 (59.2%)
	Yes	206 (36.7%)
Treatment record	Missing[b]	23 (4.1%)

[a] Stages I and II classified as early stage, and stages III and IV as late stage.
[b] No treatment information on surgery, chemotherapy or radiation available for these patients.

References

Fang, E. X., Ning, Y., and Liu, H. (2017). Testing and confidence intervals for high dimensional proportional hazards models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1415–1437.

Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C.-H. (2013). Oracle inequalities for the lasso in the Cox model. *Annals of Statistics*, 41(3):1142–1165.

Kong, S. and Nan, B. (2014). Non-asymptotic oracle inequalities for the high-dimensional Cox regression via lasso. *Statistica Sinica*, 24(1):25–42.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

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