Web-based Supplementary materials for "Model-free scoring system for risk prediction with application to hepatocellular carcinoma study"

BY WEINING SHEN, JING NING, YING YUAN, ANNA S. LOK, AND ZIDING FENG

1. REGULARITY CONDITIONS

We assume the following regularity conditions. These conditions are commonly used in the literature. Condition (C1) holds for a wide class of parametric functions including (4) and (5) with scaling constraints on the ψ_k's in the main paper.

(C1) The true parameter value $\boldsymbol{\beta_0}$ belongs to a known compact set $\mathcal{B}$. The utility functions $\ell_n(\boldsymbol{\beta})$ and $Q_n(\boldsymbol{\beta})$ have a unique minimizer in $\mathcal{B}$.

(C2) The kernel function $K(\cdot)$ is thrice-continuously differentiable and its r-th order derivatives $K^{(r)}$ are bounded for $r = 0, \ldots, 3$. We assume that the bandwidth $h_n \to 0$ and $nh_n^6 \to \infty$ as $n \to \infty$.

By examining the proof of Theorem 1, we notice that the asymptotic results are not restricted to the linear score system, but also work for nonlinear additive models such as $S(\boldsymbol{\beta}, \boldsymbol{X}) = \beta_1 X_1^2 + \beta_2 \log X_2$, given that the covariates are in a compact set and the partial derivatives of S with respect to $\boldsymbol{\beta}$ are continuously differentiable.

2. PROOF OF THEOREM 1

We first show that $\sup_{\boldsymbol{\beta} \in \mathcal{B}} |\ell_n^s(\boldsymbol{\beta}) - \ell_n(\boldsymbol{\beta})| \to 0$. Consider an ϵ-net of $\mathcal{B}$ and denote its covering number by $m_n(\epsilon)$. Then given the index i fixed, under Condition (C2), for any $\boldsymbol{\beta}^i, \ldots, \boldsymbol{\beta}^{m_n}$ from each partition of $\mathcal{B}$, by the approximation result in Schuster (1969),

$$P\left(\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{h_n^{-1}(S(\boldsymbol{X}_i;\boldsymbol{\beta})-S(\boldsymbol{X}_j;\boldsymbol{\beta}))}K(u)\,du - P\{S(\boldsymbol{X}_i;\boldsymbol{\beta}) > S(\boldsymbol{X};\boldsymbol{\beta})\}\right| > \xi_n\right)$$
$$\leq \sum_{k=1}^{m_n(\epsilon)} P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{h_n^{-1}(S(\boldsymbol{X}_i;\boldsymbol{\beta}^k)-S(\boldsymbol{X}_j;\boldsymbol{\beta}^k))}K(u)\,du - P\{S(\boldsymbol{X}_i;\boldsymbol{\beta}^k) > S(\boldsymbol{X};\boldsymbol{\beta}^k)\}\right| > \xi_n/2\right)$$
$$\leq C_1 m_n(\epsilon)\exp(-C_2 n h_n^2 \xi_n^2)$$

for some positive constants C_1, C_2 and any sequence $\xi_n = o(h_n)$. Choose ϵ such that $m_n(\epsilon)$ is a polynomial in n, then

$$\sum_{n=1}^{\infty} P\left(\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{h_n^{-1}(S(\boldsymbol{X}_i;\boldsymbol{\beta})-S(\boldsymbol{X}_j;\boldsymbol{\beta}))}K(u)\,du - P\{S(\boldsymbol{X}_i;\boldsymbol{\beta}) > S(\boldsymbol{X};\boldsymbol{\beta})\}\right| > \xi_n\right) < \infty$$

since $nh_n^2 \to \infty$ and $h_n \to 0$. By Borel-Cantelli lemma, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{h_n^{-1}(S(\boldsymbol{X}_i;\boldsymbol{\beta})-S(\boldsymbol{X}_j;\boldsymbol{\beta}))}K(u)\,du - P\{S(\boldsymbol{X}_i;\boldsymbol{\beta}) > S(\boldsymbol{X};\boldsymbol{\beta})\}\right| \to 0$$

almost surely. Similarly, for every i,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{\frac{1}{n}\delta_i\sum_{j=1}^{n}1\!\!1(Z_i > Z_j)\int_{-\infty}^{h_n^{-1}(S(\boldsymbol{X}_i;\boldsymbol{\beta})-S(\boldsymbol{X}_j;\boldsymbol{\beta}))}K(u)\,du}{\frac{1}{n}\sum_{j=1}^{n}1\!\!1(Z_i > Z_j)}\right.$$
$$\left. - \mathrm{E}\{\delta_i(S(\boldsymbol{X}_i;\boldsymbol{\beta}) > S(\boldsymbol{X};\boldsymbol{\beta}))|Z_i > Z\}\right| \to 0$$

almost surely. We obtain $\sup_{\boldsymbol{\beta}\in\mathcal{B}}|\ell_n^s(\boldsymbol{\beta}) - \ell_n(\boldsymbol{\beta})| \to 0$ almost surely. Also, by definition, $\boldsymbol{\beta}_0$ is the maximizer of $\ell_n(\boldsymbol{\beta})$. Since we assume an independent censoring condition, $\boldsymbol{\beta}_n$ is also the minimizer of $U(\boldsymbol{\beta})$. Hence by Theorem 5.7 of van der Vaart (2000), the consistency of $\boldsymbol{\beta}_n^s$ follows.

Define

$$\boldsymbol{\Sigma}_{\mathbf{1}}^i = -E\left(\frac{\partial w_n^i}{\partial\boldsymbol{\beta}};\boldsymbol{\beta}_0\right) \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{1}} = E(\delta_i\boldsymbol{\Sigma}_{\mathbf{1}}^i).$$

Following similar arguments in Zeng and Lin (2007), for every $\xi_n \to 0$

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\xi_n} |\sqrt{n}\{w_n(\boldsymbol{\beta}) - w_n(\boldsymbol{\beta}_0)\} + \Sigma_1\sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)| = o_p(\|\sqrt{n}\boldsymbol{\beta}-\boldsymbol{\beta}_0\|)(\boldsymbol{\beta}-\boldsymbol{\beta}_0).$$

By the arguments in Schuster (1969), the assumption that $nh_n^6 \to \infty$ and $K^{(r)}$ is uniformly bounded for $0 \le r \le 3$ together guarantee that $\nabla_{\boldsymbol{\beta}} w_n^i(\boldsymbol{\beta})$ and $\sqrt{n}\Delta_{\boldsymbol{\beta}} w_n^i(\boldsymbol{\beta})$ are well defined and uniformly bounded with repect to every i in a neighborhood around $\boldsymbol{\beta}_0$. In other words,

$$w_n^i(\boldsymbol{\beta}) - w_n^i(\boldsymbol{\beta}_0) = -\boldsymbol{\Sigma}_{\mathbf{1}}^i(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\Delta_{\boldsymbol{\beta}} w_n^i(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + o_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2).$$

In particular, for every $\xi_n \to 0$,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\xi_n} |\sqrt{n}\{w_n^i(\boldsymbol{\beta}) - w_n^i(\boldsymbol{\beta}_0)\} + \boldsymbol{\Sigma}_{\mathbf{1}}^i\sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)| = o_p(\sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)).$$

Therefore

$$\begin{aligned}
\sqrt{n}\{w_n(\boldsymbol{\beta}_n^s) - w_n(\boldsymbol{\beta}_0)\} &= \frac{1}{n}\sum_{i=1}^n \delta_i\sqrt{n}\left\{w_n^i(\boldsymbol{\beta}_n^s) - w_n^i(\boldsymbol{\beta}_0)\right\} \\
&= -\frac{1}{n}\sum_{i=1}^n \delta_i\boldsymbol{\Sigma}_{\mathbf{1}}^i\sqrt{n}(\boldsymbol{\beta}_n^s-\boldsymbol{\beta}_0) \\
&= -\boldsymbol{\Sigma}_{\mathbf{1}}\sqrt{n}(\boldsymbol{\beta}_n^s-\boldsymbol{\beta}_0).
\end{aligned}$$

Therefore the asymptotic normality follows since

$$\sqrt{n}w_n(\boldsymbol{\beta}_0) = \sqrt{n}\frac{1}{n}\sum_{i=1}^n \delta_i w_n^i(\boldsymbol{\beta}_0) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{2}}).$$

3. PROOF OF THEOREM 2

We follow the proofs of Zhang and Lu (2007). In order to show the root-n consistency of $\hat{\boldsymbol{\beta}}_n$, it suffices to show that, for any given $\epsilon > 0$, there exists a constant C such that

$$\Pr\left\{\inf_{\boldsymbol{\beta}=\boldsymbol{\beta}_{\mathbf{0}}+n^{-1/2}\boldsymbol{u},\|\boldsymbol{u}\|_2=C} Q_n(\boldsymbol{\beta}) > Q_n(\boldsymbol{\beta}_{\mathbf{0}})\right\} \ge 1-\epsilon,$$

where $\boldsymbol{u} = (u_1, \ldots, u_p)$ is a vector of dimension p. Note that

$$\begin{aligned}
&Q_n(\boldsymbol{\beta}_{\mathbf{0}} + n^{-1/2}\boldsymbol{u}) - Q_n(\boldsymbol{\beta}_{\mathbf{0}}) \\
&= \ell_n^s(\boldsymbol{\beta}_{\mathbf{0}} + n^{-1/2}\boldsymbol{u}) - \ell_n^s(\boldsymbol{\beta}_{\mathbf{0}}) - \lambda_n\sum_{j=2}^p\left(\frac{|\beta_{j0}+n^{-1/2}u_j|}{|\tilde{\beta}_j|} - \frac{|\beta_{j0}|}{|\tilde{\beta}_j|}\right) \\
&\le \ell_n^s(\boldsymbol{\beta}_{\mathbf{0}} + n^{-1/2}\boldsymbol{u}) - \ell_n^s(\boldsymbol{\beta}_{\mathbf{0}}) + \lambda_n\sum_{j=2}^d\frac{|n^{-1/2}u_j|}{|\tilde{\beta}_j|} \\
&= -\frac{1}{2n}\boldsymbol{u}^T\boldsymbol{\Sigma}_{\mathbf{2}}\boldsymbol{u} + \boldsymbol{u}^T O_p(1)\boldsymbol{u} + \lambda_n n^{-1/2}\sum_{j=2}^d\frac{|u_j|}{|\tilde{\beta}_j|} \\
&= -\frac{1}{2n}\boldsymbol{u}^T\{\boldsymbol{\Sigma}_{\mathbf{2}} + o_p(1)\}\boldsymbol{u} + \frac{1}{n}O_p(1)\sum_{j=2}^p|u_j| + \lambda_n n^{-1/2}\sum_{j=2}^d\frac{|u_j|}{|\tilde{\beta}_j|}.
\end{aligned} \tag{1}$$

Since $\tilde{\boldsymbol{\beta}}_n$ is root-n consistent, for $1 \le j \le d$,

$$\frac{1}{|\tilde{\beta}_j|} = \frac{1}{|\beta_{j0}|} - \frac{\text{sign}(\beta_{j0})}{\beta_{j0}^2}(\tilde{\beta}_j - \beta_{j0}) + o_p(|\tilde{\beta}_j - \beta_{j0}|) = \frac{1}{|\beta_{j0}|} + \frac{O_p(1)}{\sqrt{n}}.$$

Because $\sqrt{n}\lambda_n = O_p(1)$, it follows that

$$\lambda_n n^{-1/2} \sum_{j=2}^{d} \frac{|u_j|}{|\tilde{\beta}_j|} = \lambda_n n^{-1/2} \sum_{j=2}^{d} \left(\frac{|u_j|}{|\beta_{j0}|} + \frac{|u_j|}{\sqrt{n}} O_p(1) \right) \le C N^{-1} O_p(1).$$

Therefore, in (1), the first term is of the order $C^2 n^{-1}$, while the second and third terms are of the order Cn^{-1}, given sufficiently large C. Hence (1) is negative if C is chosen to be sufficiently large, and the root-n consistency follows.

Next, we show that $\hat{\boldsymbol{\beta}}_{2n} = \mathbf{0}$. Following the arguments by Zhang and Lu (2007), it suffices to show that, for any $\boldsymbol{\beta}_{\mathbf{1}}$ satisfying $\|\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{\beta}_{\mathbf{10}}\|_2 = O_p(n^{-1/2})$, $\partial Q_n(\boldsymbol{\beta})/\partial \beta_j$ and β_j have different signs for $\beta_j \in (-Cn^{-1/2}, Cn^{-1/2})$ for $j = (d+1), \ldots, p$ and some constant $C > 0$. Using the Taylor expansion,

$$\begin{aligned} \frac{\partial Q_n(\boldsymbol{\beta})}{\partial \beta_j} &= \frac{\partial \ell_n^s(\boldsymbol{\beta})}{\partial \beta_j} - \lambda_n \frac{\text{sign}(\beta_j)}{|\tilde{\beta}_j|} \\ &= O_p(n^{-1/2}) - \lambda_n \frac{\text{sign}(\beta_j)}{|\tilde{\beta}_j|} \\ &= n^{1/2} \left(O_p(1) - n\lambda_n \frac{\text{sign}(\beta_j)}{\sqrt{n}|\tilde{\beta}_j|} \right) \\ &= n^{1/2} \left(O_p(1) - n\lambda_n \frac{\text{sign}(\beta_j)}{O_p(1)} \right). \end{aligned}$$

Since $n\lambda_n \to \infty$, the sign of $\partial Q_n(\boldsymbol{\beta})/\partial \beta_j$ is always different from β_j. The result follows.

Last, we show the asymptotic normality of $\hat{\boldsymbol{\beta}}_{2n}$. Note that there exists a maximizer of $Q_n(\boldsymbol{\beta}_{\mathbf{1}}, \mathbf{0})$ (given $(d+1), \ldots, p$-th components are fixed 0), which we denote by $(\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0})$. Then

$$\begin{aligned} 0 &= \frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{\mathbf{1}}}\Big|_{\boldsymbol{\beta}^T = (\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0})} \\ &= S_{1n}(\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0}) - \nabla_{\boldsymbol{\beta}} S_{1n}(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_{10}) - \lambda_n \left(\frac{\text{sign}(\hat{\beta}_1)}{\tilde{\beta}_1}, \ldots, \frac{\text{sign}(\hat{\beta}_d)}{\tilde{\beta}_d} \right)^T \\ &= S_{1n}(\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0}) - \nabla_{\boldsymbol{\beta}} S_{1n}(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_{10}) - \lambda_n \left(\frac{\text{sign}(\beta_{10})}{\tilde{\beta}_1}, \ldots, \frac{\text{sign}(\beta_{d0})}{\tilde{\beta}_d} \right)^T, \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}^*$ is between $\boldsymbol{\beta}_{10}$ and $\hat{\boldsymbol{\beta}}_{1n}$. Note that $n\lambda_n \to 0$, we rewrite (2) as

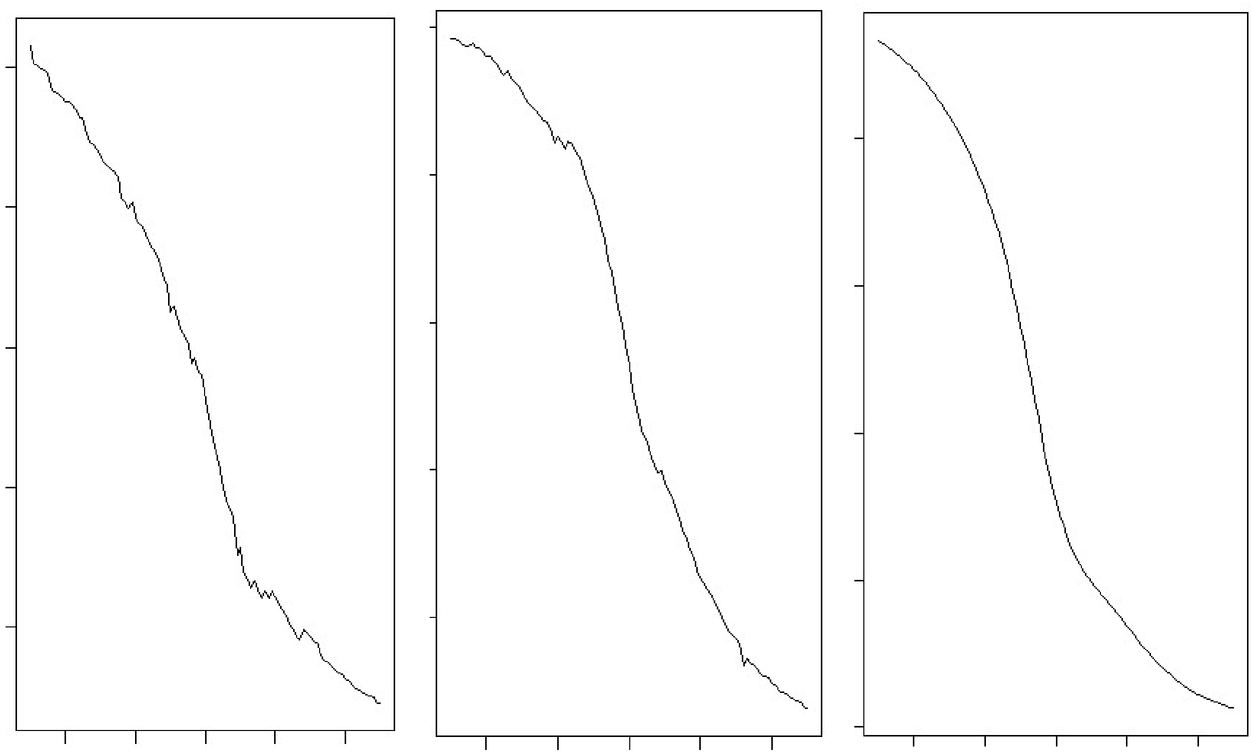
$$n^{1/2} S_{1n}(\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0}) = n^{1/2} \nabla_{\boldsymbol{\beta}} S_{1n}(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_{10}).$$

Similar to the proof of Theorem 1, we have $n^{1/2} S_{1n}(\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0}) \rightarrow_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{2,1}})$ and $\nabla_{\boldsymbol{\beta}} S_{1n}(\boldsymbol{\beta}^*) \rightarrow \boldsymbol{\Sigma}_{\mathbf{1,1}}$ as $n \rightarrow \infty$. The conclusion follows.

4. ADDITIONAL SIMULATION RESULTS

To demonstrate the kernel approximation (8) for the objective function (7) in the main paper, here we consider an example with two covariates X_1 and X_2 and their coefficients β_1 and β_2. We fix $\beta_1 = 1$ and plot the objective function with β_2 and their kernel approximations with bandwidths $h_n = n^{-1}$ and $h_n = n^{-1/3}$ in Figure 1. It can be seen that the recommended choice of bandwidth $h_n = n^{-1/3}$ works reasonably well.

Figure 1: Plots for the one-dimensional pseudo-likelihood function and its kernel approximation with bandwidth $h = n^{-1}$, and $h = n^{-1/3}$.



REFERENCES

Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *The Annals of Mathematical Statistics* **40,** 1187–1195.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press.

Zeng, D. and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association* **102,** 1387–1396.

Zhang, H. H. and Lu, W. (2007). Adaptive lasso for cox's proportional hazards model. *Biometrika* **94,** 691–703.