Biostatistics (2015), **0**, 0, *pp.* 1–11
doi:10.1093/biostatistics/Sinnott˙Cai˙Biostatistics˙Supplementary˙Revision

Inference for Survival Prediction Under the Regularized Cox Model: Supplementary Materials

JENNIFER A. SINNOTT*, TIANXI CAI

Department of Biostatistics, Harvard University, Boston, MA 02115, USA

jas953@mail.harvard.edu

1. Web Appendix A: Oracle properties of $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\theta}}^*$, $\widehat{\boldsymbol{\theta}}_V$, and $\widehat{\boldsymbol{\theta}}_V^*$

Oracle properties of the adaptive elastic net estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}^*$ are established in Minnier *and others* (2011) for an arbitrary objective function which is the sum of iid contributions from the observations. Thus, oracle properties for these estimators in the Cox model follow once we show that the LPL and the perturbed LPL are asymptotically equivalent to an objective function which is the sum of iid contributions. We demonstrate this for the perturbed LPL, since the result then holds for the standard LPL by replacing each weight $\mathcal{V}_i$ by weight 1. Throughout, for simplicity of presentation, we assume no unpenalized covariates.

To demonstrate that $\widehat{\ell}_0^*(\boldsymbol{\theta})$ is asymptotically equivalent to an objective function that is the sum of iid contributions, we may follow similar arguments as given in Cai and Zheng (2013), by

*To whom correspondence should be addressed.

decomposing:

$$\widehat{\ell}_0^*(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\mathcal{V}_i\int[\boldsymbol{\theta}^\intercal\mathbf{W}_i - \log\{n\Pi^{(0)}(\boldsymbol{\theta},t)\}]dN_i(t)$$
$$-n^{-1}\sum_{i=1}^{n}\mathcal{V}_i\int[\log\{n\widehat{\Pi}^{(0)*}(\boldsymbol{\theta},t)\} - \log\{n\Pi^{(0)}(\boldsymbol{\theta},t)\}]dN_i(t),$$

where $\Pi^{(k)}(\boldsymbol{\theta},t) = E\{\mathbf{I}(X \geqslant t)e^{\boldsymbol{\theta}^\intercal\mathbf{W}}\mathbf{W}^{\otimes k}\}$ is the limit of $\widehat{\Pi}^{(k)*}(\boldsymbol{\theta},t) = n^{-1}\sum_{i=1}^{n}\mathcal{V}_i I(X_i \geqslant s)\mathbf{W}_i^{\otimes k}e^{\boldsymbol{\theta}^\intercal\mathbf{W}_i}$, and for any vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$ and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^\intercal$. We then use a Taylor series approximation to rewrite the second term as

$$\int\left\{\frac{\widehat{\Pi}^{(0)*}(\boldsymbol{\theta},t) - \Pi^{(0)}(\boldsymbol{\theta},t)}{\Pi^{(0)}(\boldsymbol{\theta},t)}\right\}d\left\{n^{-1}\sum_{i=1}^{n}\mathcal{V}_iN_i(t)\right\} + O_{\mathbb{P}^*}(n^{-1})$$
$$= n^{-1}\sum_{j=1}^{n}\mathcal{V}_j\int\left\{\frac{\mathbf{I}(X_j \geqslant t)e^{\boldsymbol{\theta}^\intercal\mathbf{W}_j} - \Pi^{(0)}(\boldsymbol{\theta},t)}{\Pi^{(0)}(\boldsymbol{\theta},t)}\right\}dA(t) + O_{\mathbb{P}^*}(n^{-1}),$$

where $\mathbb{P}^*$ is the probability measure generated by both the observed data and $\mathcal{V}$ and $A(t) = E\{N_j(t)\}$. Thus, $\widehat{\ell}_0^*(\boldsymbol{\theta})$ is asymptotically equivalent to the iid sum

$$n^{-1}\sum_{i=1}^{n}\mathcal{V}_i\left(\int[\boldsymbol{\theta}^\intercal\mathbf{W}_i - \log\{n\Pi^{(0)}(\boldsymbol{\theta},t)\}]dN_i(t) - \int\frac{\mathbf{I}(X_i \geqslant t)e^{\boldsymbol{\theta}^\intercal\mathbf{W}_i} - \Pi^{(0)}(\boldsymbol{\theta},t)}{\Pi^{(0)}(\boldsymbol{\theta},t)}dA(t)\right)$$

Finally, note that although Minnier *and others* (2011) do not explicitly address the aENET penalty, their proof under the adaptive lasso penalty generalizes to this case.

Next, we use the same approach to establish oracle properties for the ensemble-voting-based estimators $\widehat{\boldsymbol{\theta}}_V$ and $\widehat{\boldsymbol{\theta}}_V^*$. First, let $\widehat{I}_j$ be the vector indicating whether the perturbations vote the j^{th} variable into the model. That is, if we write $\widehat{q}_j = B^{-1}\sum_{b=1}^{B}\mathbf{I}(\widehat{\theta}_j^{*(b)} = 0)$ then $\widehat{I}_j = \mathbf{I}(\widehat{q}_j \leqslant \mathfrak{p}_j)$, where $\mathfrak{p}_j = \min\{\max(0.05, \bar{\mathfrak{p}}_j), 0.95\}$ is the threshold parameter as defined in section 2.5. Note that although $\mathfrak{p}_j$ is selected in a data-driven manner, the truncation ensures that $\mathfrak{p}_j \in [0.05, 0.95]$. Since the oracle properties of $\widehat{\boldsymbol{\theta}}^*$ ensure that $\widehat{q}_j \to 1$ for $j \in \mathcal{A}^c$ and $\widehat{q}_j \to 0$ for $j \in \mathcal{A}$, we have $P(\widehat{q}_j > \mathfrak{p}_j\ \forall j \notin \mathcal{A}) \geqslant P(\widehat{q}_j > 0.95\ \forall j \notin \mathcal{A}) \to 1$ and $P(\widehat{q}_j > \mathfrak{p}_j) \leqslant P(\widehat{q}_j > 0.05) \to 0$ for any $j \in \mathcal{A}$. It follows that $P(\widehat{\mathbf{I}} = \mathbf{I}_0) \to 1$, where $\widehat{\mathbf{I}} = (\widehat{I}_1, ..., \widehat{I}_p)^\intercal$, $\mathbf{I}_0 = (I_{01}, ..., I_{0p})^\intercal$ and $I_{0j} = I(j \in \mathcal{A}^c)$. The oracle properties of $\widehat{\boldsymbol{\theta}}_V$ follows from similar arguments as the oracle properties of $\widehat{\boldsymbol{\theta}}$ since the two estimators only differ by their initial estimators both of which are root-n consistent.

To establish the properties for

$$\widehat{\boldsymbol{\theta}}_V^* = (1+\lambda_2^*) \operatorname*{argmin}_{\boldsymbol{\theta}} \{-\widehat{\ell}_{0V}^*(\boldsymbol{\theta}) + \lambda_2^* \sum_{j=1}^{p_Z} \theta_j^2 + \lambda_1^* \sum_{j=1}^{p_Z} \widehat{w}_{Vj}^* |\theta_j|\},$$

where

$$\widehat{\ell}_{0V}^*(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \Delta_i \left[(\boldsymbol{\theta} \odot \widehat{\mathbf{I}})^{\intercal} \mathbf{W}_i - \log\{\widehat{\Pi}^{(0)*}(\boldsymbol{\theta} \odot \widehat{\mathbf{I}}, X_i)\}\right]$$

$\odot$ represents element-wise product, $\widehat{w}_{Vj}^* = |\widetilde{\theta}_{VRj}^*|^{-1}$, and

$$\widetilde{\boldsymbol{\theta}}_{VR}^* = (1+\lambda_2^*) \operatorname*{argmin}_{\boldsymbol{\theta}} \{-\widehat{\ell}_{0V}^*(\boldsymbol{\theta}) + \lambda_2^* \sum_{j=1}^{p_Z} \theta_j^2\}.$$

Note that in each minimization, the penalization term leads us to set the j^{th} coefficient to be 0 whenever $\widehat{I}_j = 0$. To establish that $\widehat{\boldsymbol{\theta}}_V^*$ also has oracle properties, we show again that $\widehat{\ell}_{0V}^*(\boldsymbol{\theta})$ may be approximated by a sum of iid terms up to $O_{\mathbb{P}^*}(n^{-1})$. To see this, we expand:

$$\widehat{\ell}_{0V}^*(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \int \left[(\boldsymbol{\theta} \odot \mathbf{I}_0)^{\intercal} \mathbf{W}_i - \log\{n\Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)\}\right] dN_i(s) \tag{1.1}$$

$$+ \{\boldsymbol{\theta} \odot (\widehat{\mathbf{I}} - \mathbf{I}_0)\}^{\intercal} n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \int \mathbf{W}_i dN_i(s) \tag{1.2}$$

$$+ n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \int [\log\{\widehat{\Pi}^{(0)*}(\boldsymbol{\theta} \odot \widehat{\mathbf{I}}, s)\} - \log\{\widehat{\Pi}^{(0)*}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)\}] dN_i(s) \tag{1.3}$$

$$- n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \int [\log\{\widehat{\Pi}^{(0)*}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)\} - \log\{\Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)\}] dN_i(s) \tag{1.4}$$

Looking at specific terms, we see $P\{(1.2) + (1.3) = 0\} \geqslant P(\widehat{\mathbf{I}} - \mathbf{I}_0 = 0) \to 1$ and thus (1.2) and (1.3) take value 0 with probability approaching 1. Using the same arguments as given above for the approximation of $\widehat{\ell}_0^*(\boldsymbol{\theta})$, (1.4) can be expressed as

$$-n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \int \left[\frac{I(X_i \geqslant s) e^{(\boldsymbol{\theta} \odot \mathbf{I}_0)^{\intercal} \mathbf{W}_i} - \Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)}{\Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)}\right] dA(s) + o_{\mathbb{P}^*}(n^{-1})$$

Thus, $\ell_{0V}^*(\boldsymbol{\theta})$ can be written as a sum of iid up to $O_{\mathbb{P}^*}(n^{-1})$:

$$n^{-1} \sum_{i=1}^{n} \mathcal{V}_i \int \Bigg(\left[(\boldsymbol{\theta} \odot \mathbf{I}_0)^{\intercal} \mathbf{W}_i - \log\{n\Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)\}\right] dN_i(s) - \frac{I(X_i \geqslant s) e^{(\boldsymbol{\theta} \odot \mathbf{I}_0)^{\intercal} \mathbf{W}_i} - \Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)}{\Pi^{(0)}(\boldsymbol{\theta} \odot \mathbf{I}_0, s)} dA(s) \Bigg) + O_{\mathbb{P}^*}(n^{-1}).$$

The oracle properties of $\widehat{\boldsymbol{\theta}}_V^*$ then follows from Minnier *and others* (2011).

2. Web Appendix B: Weak Convergence of $\sqrt{n}\{\widehat{S}(t_0; w_{\text{NEW}}) - S(t_0; w_{\text{NEW}})\}$

It suffices to show $\sqrt{n}\left\{\widehat{\Lambda}_0(t_0;\widehat{\boldsymbol{\theta}}) - \Lambda_0(t_0;\boldsymbol{\theta}_0)\right\}$ converges weakly to a Gaussian process in $t_0 \in [t_1, t_2]$. We may expand:

$$\begin{aligned}\sqrt{n}\left\{\widehat{\Lambda}_0(t_0;\widehat{\boldsymbol{\theta}}) - \Lambda_0(t_0;\boldsymbol{\theta}_0)\right\} &= \sqrt{n}\left\{\widehat{\Lambda}_0(t_0;\widehat{\boldsymbol{\theta}}) - \widehat{\Lambda}_0(t_0;\boldsymbol{\theta}_0)\right\} + \sqrt{n}\left\{\widehat{\Lambda}_0(t_0;\boldsymbol{\theta}_0) - \Lambda_0(t_0;\boldsymbol{\theta}_0)\right\} \\ &= \widehat{\mathbf{H}}(\boldsymbol{\theta}_0, t_0)^\intercal \cdot \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \sqrt{n}\left\{\widehat{\Lambda}_0(t_0;\boldsymbol{\theta}_0) - \Lambda_0(t_0;\boldsymbol{\theta}_0)\right\} + o_{\mathbb{P}}(1)\end{aligned}$$

where $\widehat{\mathbf{H}}(\boldsymbol{\theta}, t) = -\int_0^t \widehat{\Pi}^{(1)}(\boldsymbol{\theta}, s)\widehat{\Pi}^{(0)}(\boldsymbol{\theta}, s)^{-2} d\bar{N}(s)$ and $\widehat{\Pi}^{(k)}(\boldsymbol{\theta}; s) = n^{-1}\sum_{i=1}^n I(X_i \geqslant s)\mathbf{W}_i^{\otimes k} e^{\boldsymbol{\theta}^\intercal \mathbf{W}_i}$.

From the oracle properties of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_V$, we have $|n^{\frac{1}{2}}\widehat{\boldsymbol{\theta}}_{\mathcal{A}^c}| + |n^{\frac{1}{2}}\widehat{\boldsymbol{\theta}}_{V\mathcal{A}^c}| \to 0$ and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A}0}) = \sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A}0}) = n^{-\frac{1}{2}}\mathbb{A}_{\mathcal{A},\mathcal{A}}(\boldsymbol{\theta}_0)^{-1}\sum_{i=1}^n \mathbf{U}_{\mathcal{A}i}(\boldsymbol{\theta}_0) + o_{\mathbb{P}}(1),$$

where $\mathbb{A}_{\mathcal{A},\mathcal{A}}$ represents the sub-matrix of $\mathbb{A}$ whose rows and columns correspond to $\mathcal{A}$,

$$\begin{aligned}\mathbf{U}_i(\boldsymbol{\theta}_0) &= \int \left(\mathbf{W}_i - \frac{\Pi^{(1)}(\boldsymbol{\theta}, s)}{\Pi^{(0)}(\boldsymbol{\theta}, s)}\right) dN_i(s) - \int_0^t I(X_i \geqslant s) e^{\boldsymbol{\theta}^\intercal \mathbf{W}_i} \frac{\Pi^{(0)}(\boldsymbol{\theta}, t)\mathbf{W}_i - \Pi^{(1)}(\boldsymbol{\theta}, t)}{\Pi^{(0)}(\boldsymbol{\theta}, t)^2}, \\ \text{and} \quad \mathbb{A}(\boldsymbol{\theta}) &= -\int \frac{\Pi^{(2)}(\boldsymbol{\theta}, t)\Pi^{(0)}(\boldsymbol{\theta}, t) - \Pi^{(1)}(\boldsymbol{\theta}, t)\Pi^{(1)}(\boldsymbol{\theta}, t)^\intercal}{\Pi^{(0)}(\boldsymbol{\theta}, t)^2} dA(t).\end{aligned}$$

From a uniform law of large numbers (Pollard, 1990), we have the in probability convergence of $\widehat{\mathbf{H}}(\boldsymbol{\theta}, t) \to \mathbf{H}(\boldsymbol{\theta}, t) = -\int_0^t \Pi^{(1)}(\boldsymbol{\theta}, s)\Pi^{(0)}(\boldsymbol{\theta}, s)^{-2} dA(s)$ uniformly in t and $\boldsymbol{\theta}$. It follows that

$$\widehat{\Lambda}_0(t_0;\widehat{\boldsymbol{\theta}}) - \Lambda_0(t_0;\boldsymbol{\theta}_0) = \mathbf{H}_{\mathcal{A}}(\boldsymbol{\theta}_0, t_0)^\intercal \mathbb{A}_{\mathcal{A},\mathcal{A}}(\boldsymbol{\theta}_0)^{-1} n^{-1}\sum_{i=1}^n \mathbf{U}_{\mathcal{A}i}(\boldsymbol{\theta}_0) + \widehat{\Lambda}_0(t_0;\boldsymbol{\theta}_0) - \Lambda_0(t_0;\boldsymbol{\theta}_0) + o_{\mathbb{P}}(n^{-\frac{1}{2}}).$$

From Andersen and Gill (1982)

$$\sqrt{n}\left\{\widehat{\Lambda}_0(t_0;\boldsymbol{\theta}_0) - \Lambda_0(t_0;\boldsymbol{\theta}_0)\right\} = \int_0^{t_0} \frac{dM_i(s)}{\Pi^{(0)}(\boldsymbol{\theta}_0, s)} + o_{\mathbb{P}}(1),$$

where $M_i(t) = N_i(t) - \int_0^t I(X_i \geqslant s) e^{\boldsymbol{\theta}_0^\intercal \mathbf{W}_i} d\Lambda_0(s)$. Therefore, $\sqrt{n}\{\widehat{\Lambda}_0(t_0;\widehat{\boldsymbol{\theta}}) - \Lambda_0(t)\}$ is asymptotically equivalent to $n^{-\frac{1}{2}}\sum_{i=1}^n \mathcal{U}_i(t)$, which converges weakly to a mean zero Gaussian process by a functional central limit theorem (Pollard, 1990), where

$$\mathcal{U}_i(t) = \mathbf{H}_{\mathcal{A}}(\boldsymbol{\theta}_0, t)^\intercal \mathbb{A}_{\mathcal{A},\mathcal{A}}(\boldsymbol{\theta}_0)^{-1}\mathbf{U}_{\mathcal{A}i}(\boldsymbol{\theta}_0) + \int_0^t \frac{dM_i(s)}{\Pi^{(0)}(\boldsymbol{\theta}_0, s)}$$

Similar arguments given above and those as given in Lin *and others* (1994) can be used to show that $\sqrt{n}\{\widehat{\Lambda}_0^*(t_0;\widehat{\boldsymbol{\theta}}^*) - \widehat{\Lambda}_0^*(t_0;\widehat{\boldsymbol{\theta}})\} \mid \mathcal{O}$ converges to the same limiting distribution as that of $\sqrt{n}\{\widehat{\Lambda}_0(t_0;\widehat{\boldsymbol{\theta}}) - \Lambda_0(t)\}$.

3. Web Appendix C: Additional Simulation Results for Smaller Nonzero Signals

Here, we present simulation results for the setting where $h(\mathbf{z})$ includes nontrivial signals of smaller size. Specifically, we generated the genomic covariates $\mathbf{Z}$ from a multivariate normal distribution with mean 0 and compound symmetry structure with variance 1 and correlation ρ. We considered settings with $p_Z = 10, 20$, and 30 covariates, and correlations $\rho = 0$ and 0.5. The underlying signal was linear involving only the first five covariates; the structure of this signal was $h(\mathbf{z}) = 1 \cdot z_1 + 0.8 \cdot z_2 + 0.6 \cdot z_3 + 0.4 \cdot z_4 + 0.2 \cdot z_5$. For each setting, we generated survival times under the Cox model $\lambda(t_0;\mathbf{z}) = \lambda_0(t)\exp\{h(\mathbf{z})\}$, where $\lambda_0(t)$ is the hazard function from a Weibull($\lambda = 1, k = 3$). The censoring was generated from a uniform distribution with range chosen to produce approximately 50% censoring. We considered small and moderate sample sizes ($n = 200$ and $n = 500$). Here, we focus primarily on the two smallest signals $\theta_{04} = 0.4$ and $\theta_{05} = 0.2$. We consider the same patients as before: one is the "baseline" individual with $\mathbf{W} = \mathbf{W}^{(0)} = \mathbf{0}$; the second has $\mathbf{W} = \mathbf{W}^{(1)} = (0, 0, 2, 2, 2, 0, \ldots, 0)^\intercal$; and the third is $\mathbf{W} = \mathbf{W}^{(2)} = (-0.5, \ldots, -0.5)^\intercal$. Bootstrap and perturbation resampling methods use $B = 2000$ resamples. Results presented are based on 2000 simulations.

In Figure A1, we present the probabilities of inclusion of Z_4 and Z_5 in the model. We can see that these probabilities are still near 1 when $\theta_{04} = 0.4$ but range between 0.6 and 1 for $\theta_{05} = 0.2$, reflecting that the fifth variable is frequently excluded from the model even though $0.2 \neq 1$. The voting procedure tends to include θ_{05} slightly more frequently than the initial fit.

In Figure A2, we present the biases of the aENET estimator and voting based estimator for the zero-coefficients of $\mathbf{Z}$ as well as the two weak signals $\theta_{04} = 0.4$ and $\theta_{05} = 0.2$, and the empirical standard errors and estimated standard errors, and the empirical coverage levels of the 95% confidence intervals for the three methods across simulation settings. The voting method still most accurately estimates the standard errors when the true $\theta_{0j} = 0$. For the moderate signal $\theta_{04} = 0.4$, the performance patterns for the different methods are similar to those presented in the main text for the other setting. For the weakest signal $\theta_{05} = 0.2$, when $n = 200$, the asymptotic and voting-based confidence intervals substantially undercover, while the bootstrap overcovers; when $n = 500$, the coverage levels of all three methods are more similar, but still tend to undercover, especially with nontrivial correlation. The undercoverage we see here is because the signal is small and the aENET method does not always successfully include the fifth covariate and tends to produce substantial bias. These simulation results are consistent with theoretical results about the difficulty regularization methods have of selecting and providing valid confidence intervals for weak signals of similar magnitude of order $n^{-\frac{1}{2}}$ (Pötscher and Schneider, 2009; Wainwright, 2009). As we increase n to 1000, coverage improves (results not shown) – for example, when $n = 1000$, $\rho = 0.5$, and $p = 30$, the asymptotic confidence interval coverage is 89%; the bootstrap coverage is 92%; and the voting-based coverage is 93%.

The confidence interval and band performances (Figures A3 and A4) are quite similar to those in main text for the individuals $\mathbf{W}^{(0)}, \mathbf{W}^{(1)}$, and $\mathbf{W}^{(2)}$. The individual $\mathbf{W}^{(1)}$ would be most impacted by our difficulty capturing the small signal θ_{05}, and indeed the confidence interval coverage for the asymptotic-based method does drop substantially. The bootstrap intervals remain wide in this setting, and tend to overcover. By comparison, the confidence interval coverage for the voting-based method demonstrates performance very similar to the coverage presented in the main text, due to its improved ability to capture the small signal in its point estimation, as well as its improved standard error estimation for all coefficients. This further demonstrates the advantage of our proposed interval estimation procedures over existing methods based on asymptotic inference or bootstrap.

Fig. A1. Probabilities of inclusion of the j^{th} covariate, for $j = 4$ and $j = 5$

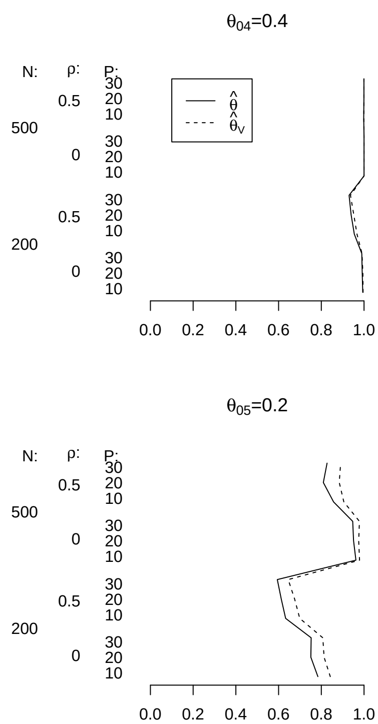


Fig. A2. Comparison of the standard errors, bias, and 95% confidence interval coverage of $\widehat{\theta}_j$, for true model parameters $\boldsymbol{\theta}_0 = (1, 0.8, 0.6, 0.4, 0.2, \dots, 0)$. Shown are values when $\theta_{0j} = 0$ (with absolute bias displayed), as well as $\theta_{04} = 0.4$ and $\theta_{05} = 0.2$. Bias and empirical standard errors are compared for the base aENET fit $(\widehat{\boldsymbol{\theta}})$ and the aENET fit after the voting procedure $(\widehat{\boldsymbol{\theta}}_V)$; the variability for the base aENET fit may be estimated using either the bootstrap or the asymptotic method, while the variability for the voting procedure is estimated using the resampled coefficient estimators after voting.

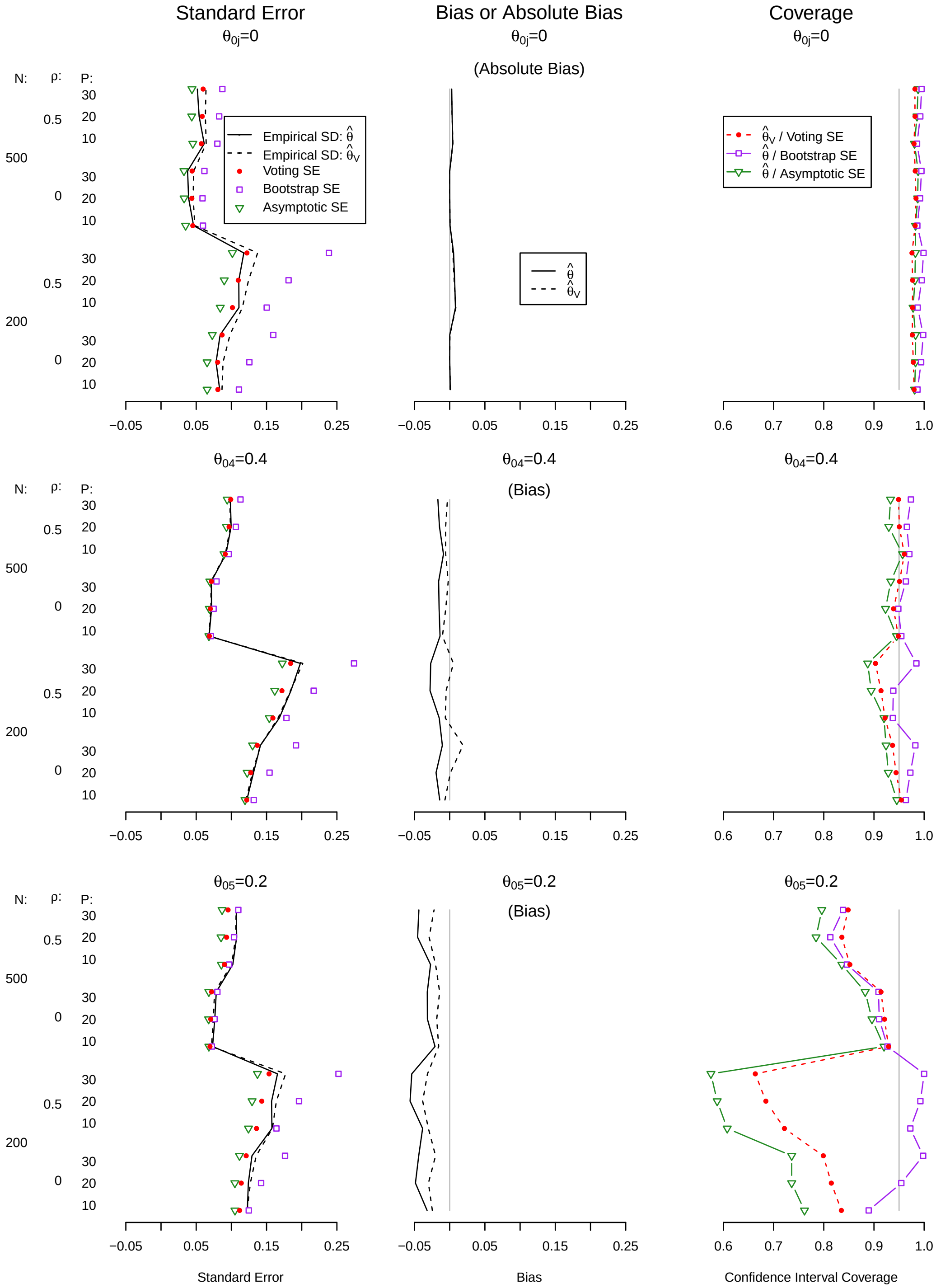


Fig. A3. Under the model with $\boldsymbol{\theta}_0 = (1, 0.8, 0.6, 0.4, 0.2, \ldots, 0)$, confidence interval coverage for t_0-year survival, and width, for three covariate levels: $W^{(0)}$, with all covariates 0; $W^{(1)} = (0, 0, 2, 2, 2, 0, \ldots, 0)$; and $W^{(2)} = (-.5, \ldots, -.5)$.

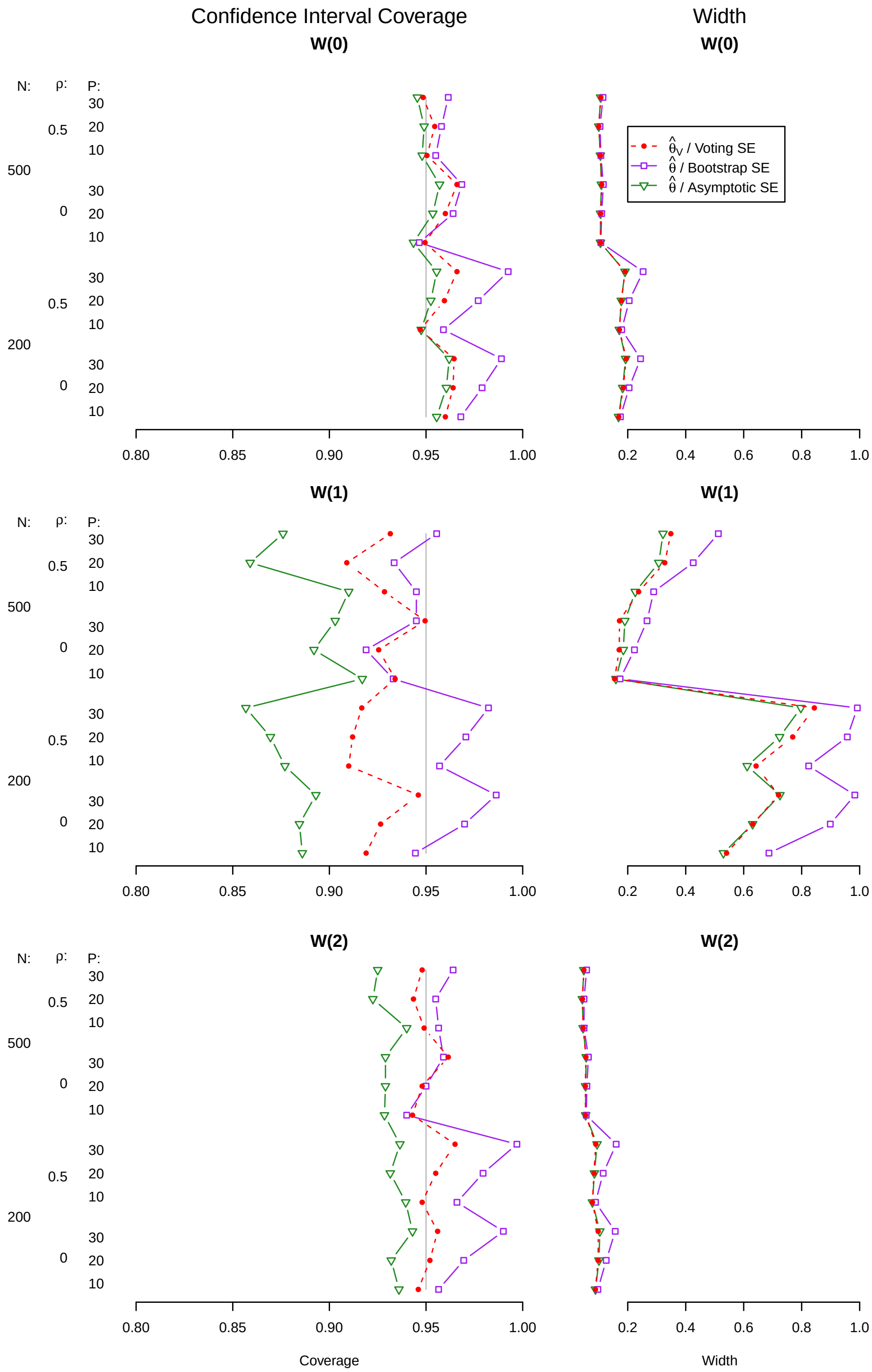
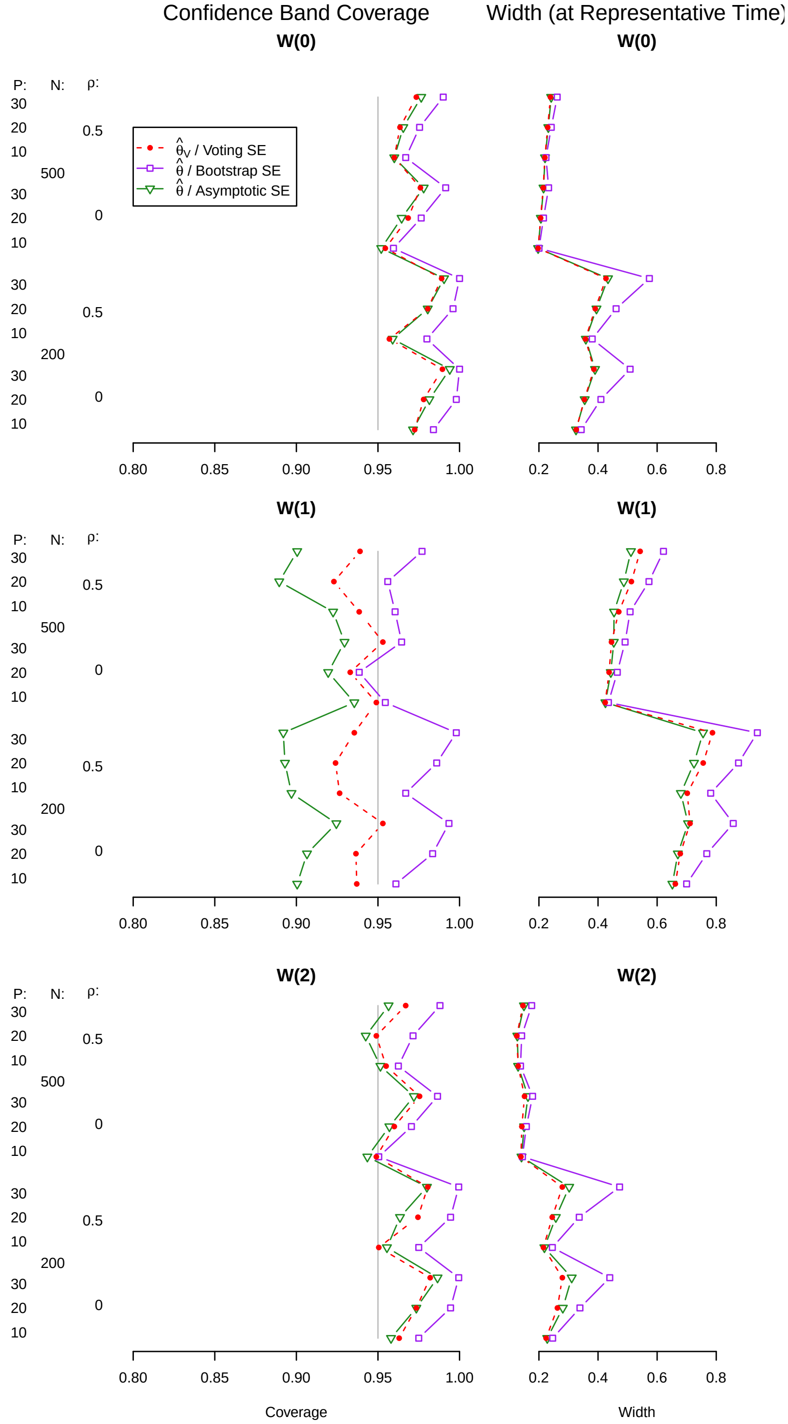


Fig. A4. Under the model with $\boldsymbol{\theta}_0 = (1, 0.8, 0.6, 0.4, 0.2, \ldots, 0)$, simultaneous confidence interval coverage for $W^{(0)}$, with all covariates 0; $W^{(1)} = (0, 0, 2, 2, 2, 0, \ldots, 0)$; and $W^{(2)} = (-.5, \ldots, -.5)$. Also shown are simultaneous confidence widths at representative times



References

Andersen, Per Kragh and Gill, Richard D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics* **10**(4), 1100–1120.

Cai, Tianxi and Zheng, Yingye. (2013). Resampling procedures for making inference under nested case–control studies. *Journal of the American Statistical Association* **108**(504), 1532–1544.

Lin, DY, Fleming, TR and Wei, LJ. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**(1), 73–81.

Minnier, Jessica, Tian, Lu and Cai, Tianxi. (2011). A perturbation method for inference on regularized regression estimates. *J. Am. Stat. Assoc.* **106**(496), 1371–1382.

Pollard, D. (1990). Empirical processes: theory and applications. In: *NSF-CBMS regional conference series in probability and statistics*. JSTOR.

Pötscher, Benedikt M and Schneider, Ulrike. (2009). On the distribution of the adaptive lasso estimator. *Journal of Statistical Planning and Inference* **139**(8), 2775–2790.

Wainwright, Martin J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (lasso). *Information Theory, IEEE Transactions on* **55**(5), 2183–2202.

[]