Statistica Sinica (2013): Supplement

NETWORK-REGULARIZED HIGH-DIMENSIONAL COX REGRESSION FOR ANALYSIS OF GENOMIC DATA

Hokeun Sun[1], Wei Lin[2], Rui Feng[2] and Hongzhe Li[2]

[1]*Columbia University and* [2]*University of Pennsylvania*

Supplementary Material

S1. Additional Simulation Results

Standard errors for Tables 1 and 2 are given in Tables S1 and S2, respectively. We conducted an additional simulation study to check if the proposed methods are sensitive to the way in which the weights w_{ij} are generated. With the same settings as in Models 5 and 6, we generated the weights w_{ij} by sample correlation coefficients between two gene expressions. For illustrative purposes, we compare the performance of the adaptive Laplacian net method with unweighted and weighted networks based on 30 simulation replicates. The simulation results are summarized in Table S3, suggesting that the variable selection and estimation performance of the adaptive Laplacian net method is insensitive to the choice of the network between the unweighted and weighted versions.

S2. Proofs of Lemmas

Proof of Lemma 1. Let $Q(\boldsymbol{\beta})$ be the objective function in (2.6). We first consider the $|\widehat{A}|$-dimensional subspace $\mathcal{B}_s = \{\boldsymbol{\beta} \in \mathbb{R}^p : \boldsymbol{\beta}_{\widehat{A}^c} = \mathbf{0}\}$. The condition that $\mathcal{I}^*_{\widehat{A}\widehat{A}}(\widehat{\boldsymbol{\beta}}, \lambda_2)$ is positive definite implies that $Q(\boldsymbol{\beta})$ is strictly convex in a neighborhood of $\widehat{\boldsymbol{\beta}}$ in $\mathcal{B}_s$. Then the zero-gradient condition (A.1) implies that $\widehat{\boldsymbol{\beta}}$ is a strict minimizer of $Q(\boldsymbol{\beta})$ in the subspace $\mathcal{B}_s$.

It remains to show that, for any $\boldsymbol{\beta}_1 \in \mathbb{R}^p \setminus \mathcal{B}_s$, we have $Q(\boldsymbol{\beta}_1) < Q(\widehat{\boldsymbol{\beta}})$. Let $\boldsymbol{\beta}_2$ be the projection of $\boldsymbol{\beta}_1$ onto to the subspace $\mathcal{B}_s$; then $Q(\boldsymbol{\beta}_2) \le Q(\widehat{\boldsymbol{\beta}})$. It suffices to show that $Q(\boldsymbol{\beta}_1) < Q(\boldsymbol{\beta}_2)$. An application of the mean value theorem gives

$$Q(\boldsymbol{\beta}_1) - Q(\boldsymbol{\beta}_2) = \sum_{j \in \widehat{A}^c:\, \beta_{1j} \neq 0} \frac{\partial Q(\bar{\boldsymbol{\beta}})}{\partial \beta_j} \beta_{1j}$$

Table S1. Standard errors for Table 1.

Method	Sensitivity	Specificity	MCC	# of genes	# of FPs	MSE
			Model 1			
Lnet	0.010	0.001	0.009	0.83	0.54	0.001
AdaLnet	0.011	0.001	0.010	0.83	0.56	0.001
Lasso	0.013	0.002	0.004	2.92	2.40	0.001
Enet	0.012	0.001	0.010	1.05	0.74	0.001
GL_γ	0.009	0.001	0.009	0.69	0.47	0.001
			Model 2			
Lnet	0.011	0.001	0.009	0.86	0.61	0.001
AdaLnet	0.015	0.001	0.011	0.95	0.63	0.001
Lasso	0.011	0.002	0.006	2.34	1.98	0.001
Enet	0.015	0.001	0.010	1.52	1.07	0.001
GL_γ	0.015	0.001	0.014	1.04	0.74	0.001
			Model 3			
Lnet	0.014	0.002	0.006	2.21	1.67	0.001
AdaLnet	0.010	0.001	0.008	0.90	0.65	0.001
Lasso	0.005	0.001	0.005	0.86	0.78	0.001
Enet	0.014	0.001	0.011	1.82	1.48	0.001
GL_γ	0.010	0.001	0.008	0.52	0.20	0.001
			Model 4			
Lnet	0.010	0.001	0.009	1.05	0.84	0.001
AdaLnet	0.015	0.001	0.011	0.96	0.57	0.001
Lasso	0.009	0.001	0.007	1.15	0.92	0.001
Enet	0.016	0.001	0.009	1.99	1.46	0.001
GL_γ	0.009	0.001	0.007	1.15	0.92	0.001

$$= \sum_{j \in \widehat{A}^c:\ \beta_{1j} \neq 0} \{U_j(\bar{\boldsymbol{\beta}}) - \lambda_1 \operatorname{sgn}(\bar{\beta}_j) - \lambda_2 \widetilde{\mathbf{L}}_{j,\cdot} \bar{\boldsymbol{\beta}}\} \beta_{1j}, \tag{S.1}$$

where $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)^T$ lies between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, β_{1j} is the jth component of $\boldsymbol{\beta}_1$, and $\widetilde{\mathbf{L}}_{j,\cdot}$ is the jth row of $\widetilde{\mathbf{L}}$. Condition (A.2) and the fact that $\operatorname{sgn}(\bar{\boldsymbol{\beta}}) = \operatorname{sgn}(\boldsymbol{\beta}_1)$ entail that each term in (S.1) is negative. Hence, $Q(\boldsymbol{\beta}_1) < Q(\boldsymbol{\beta}_2)$ and the proof is complete.

The proofs of Lemmas 2 and 3 involve modern empirical process theory. For the reader's convenience, we collect some empirical process notation here. The unfamiliar reader is referred to Chapter 19 of van der Vaart (1998) for a short introduction and van der Vaart and Wellner (1996) or Kosorok (2008) for a detailed treatment. For a measurable function f, denote by $\mathbb{P}_n f$ and Pf the expectations of f under the empirical measure $\mathbb{P}_n$ and the probability measure P, respectively. Let $\|\cdot\|_{P,r}$ denote the usual $L_r(P)$-norm. The "size" of a

Table S2. Standard errors for Table 2.

Method	Sensitivity	Specificity	MCC	# of genes	# of FPs	MSE
			Model 5			
Lnet	0.014	0.001	0.007	0.78	1.51	0.001
AdaLnet	0.011	0.001	0.008	0.86	0.60	0.001
Lasso	0.006	0.001	0.005	1.10	0.96	0.001
Enet	0.015	0.002	0.009	2.14	1.65	0.001
GL_γ	0.011	0.001	0.009	0.50	0.20	0.001
			Model 6			
Lnet	0.009	0.001	0.009	0.84	0.71	0.001
AdaLnet	0.014	0.001	0.011	0.97	0.52	0.001
Lasso	0.009	0.001	0.007	1.17	0.98	0.001
Enet	0.017	0.001	0.010	1.91	1.39	0.001
GL_γ	0.014	0.001	0.011	0.70	0.31	0.001
			Model 7			
Lnet	0.013	0.002	0.008	2.43	2.04	0.001
AdaLnet	0.011	0.001	0.008	0.73	0.49	0.001
Lasso	0.007	0.001	0.006	1.31	1.19	0.001
Enet	0.014	0.001	0.009	1.82	1.47	0.001
GL_γ	0.012	0.001	0.008	1.00	0.76	0.001
			Model 8			
Lnet	0.011	0.001	0.010	1.07	0.88	0.001
AdaLnet	0.014	0.001	0.009	0.90	0.55	0.001
Lasso	0.009	0.001	0.007	1.30	1.14	0.001
Enet	0.015	0.001	0.009	1.57	1.19	0.001
GL_γ	0.015	0.000	0.010	0.74	0.38	0.001

Table S3. Simulation results for Models 5 and 6 with unweighted and weighted networks. Sensitivity, specificity, MCC, number of selected genes, number of false positives (FPs), and mean squared error (MSE) were averaged over 30 replicates, with standard errors given in parentheses. AdaLnet: adaptive Laplacian net with an unweighted network; wAdaLnet: adaptive Laplacian net with a weighted network.

Method	Sensitivity	Specificity	MCC	# of genes	# of FPs	MSE
			Model 5			
AdaLnet	0.569	0.996	0.688	29.37	4.33	0.078
	(0.016)	(0.001)	(0.012)	(0.94)	(0.61)	(0.001)
wAdaLnet	0.561	0.996	0.683	28.93	4.27	0.078
	(0.016)	(0.001)	(0.013)	(0.89)	(0.59)	(0.001)
			Model 6			
AdaLnet	0.604	0.995	0.701	32.23	6.60	0.078
	(0.011)	(0.001)	(0.011)	(0.96)	(0.40)	(0.001)
wAdaLnet	0.607	0.995	0.704	32.33	6.60	0.078
	(0.011)	(0.001)	(0.011)	(0.95)	(0.40)	(0.001)

class $\mathcal{F}$ of functions is measured by the *bracketing number* $N_{[]}(\varepsilon, \mathcal{F}, L_r(P))$, the minimum number of ε-brackets in $L_r(P)$ needed to cover $\mathcal{F}$, and the *covering number* $N(\varepsilon, \mathcal{F}, L_2(Q))$, the minimum number of $L_2(Q)$-balls of radius ε needed to cover $\mathcal{F}$. The logarithms of the bracketing number and covering number are called *entropy with bracketing* and *entropy*, respectively. The *bracketing integral* and *uniform entropy integral* are defined as

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon$$

and

$$J(\delta, \mathcal{F}, L_2) = \int_0^\delta \sqrt{\log \sup_Q N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, d\varepsilon,$$

respectively, where F is an envelope function of $\mathcal{F}$, i.e., $|f| \le F$ for all $f \in \mathcal{F}$, and the supremum is taken over all probability measures Q with $\|F\|_{Q,r} > 0$. To save notation, we will use "$\lesssim$" to denote "less than or equal to up to a constant."

The following lemma will be useful in the proofs of Lemmas 2 and 3.

Lemma 4 (Concentration of $\mathbf{S}^{(k)}(\cdot,\cdot)$, $k = 0, 1, 2$)**.** *Under Conditions* (C1) *and* (C2)*, there exist constants $C, K > 0$ such that*

$$P\left(\sup_{\boldsymbol{\beta} \in \mathcal{B}_0,\, t \in [0,\tau]} |S^{(0)}(\boldsymbol{\beta}, t) - s^{(0)}(\boldsymbol{\beta}, t)| \ge C\sqrt{s/n}(1+x)\right) \le \exp(-Ksx^2), \quad \text{(S.2)}$$

$$P\left(\sup_{\boldsymbol{\beta} \in \mathcal{B}_0,\, t \in [0,\tau]} |S_j^{(1)}(\boldsymbol{\beta}, t) - s_j^{(1)}(\boldsymbol{\beta}, t)| \ge C\sqrt{s/n}(1+x)\right) \le \exp(-Ksx^2), \quad \text{(S.3)}$$

and

$$P\left(\sup_{\boldsymbol{\beta} \in \mathcal{B}_0,\, t \in [0,\tau]} |S_{ij}^{(2)}(\boldsymbol{\beta}, t) - s_{ij}^{(2)}(\boldsymbol{\beta}, t)| \ge C\sqrt{s/n}(1+x)\right) \le \exp(-Ksx^2), \quad \text{(S.4)}$$

for all $x > 0$ and $i, j = 1, \ldots, p$, where $S_j^{(1)}(\cdot,\cdot)$ is the jth component of $\mathbf{S}^{(1)}(\cdot,\cdot)$ and $S_{ij}^{(2)}(\cdot,\cdot)$ is the (i,j)th entry of $\mathbf{S}^{(2)}(\cdot,\cdot)$.

Proof. We show (S.3) only; the other two inequalities follow similarly. Denote $W_j = \sup_{\boldsymbol{\beta} \in \mathcal{B}_0,\, t \in [0,\tau]} |S_j^{(1)}(\boldsymbol{\beta}, t) - s_j^{(1)}(\boldsymbol{\beta}, t)|$. We first control the expectation EW_j by bounding the bracketing number of the class of functions $\mathcal{S}_j \equiv \{Y(t) X_j \exp(\boldsymbol{\beta}^T \mathbf{X})\colon \boldsymbol{\beta} \in \mathcal{B}_0, t \in [0,\tau]\}$. By Condition (C2), for any $\boldsymbol{\beta} \in \mathcal{B}_0$, $|\boldsymbol{\beta}^T \mathbf{X}| \le M \|\boldsymbol{\beta}_A\|_\infty \le M(\|\boldsymbol{\beta}_0\|_\infty + d) < \infty$. Then we have, for any $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{B}_0$,

$$|\exp(\boldsymbol{\beta}_1^T \mathbf{X}) - \exp(\boldsymbol{\beta}_2^T \mathbf{X})| \le C|\boldsymbol{\beta}_1^T \mathbf{X} - \boldsymbol{\beta}_2^T \mathbf{X}| \le CM\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_\infty.$$

Hence, we need as many ε-brackets to cover the class of functions $\mathcal{C} \equiv \{\exp(\boldsymbol{\beta}^T\mathbf{X})\colon \boldsymbol{\beta} \in \mathcal{B}_0\}$ as we need hypercubes of edge length $\varepsilon/(CM)$ to cover $\mathcal{B}_0$, implying that the bracketing entropy of $\mathcal{C}$ is at most of order $s\log(1/\varepsilon)$. Also, one can easily show that the bracketing entropy of the class of functions $\{Y(t)\colon t \in [0,\tau]\}$ is at most of order $\log(1/\varepsilon)$ (van der Vaart (1998), Example 19.6). Thus, the bracketing entropy of $\mathcal{S}_j$ satisfies

$$\log N_{[]}(\varepsilon, \mathcal{S}_j, L_2(P)) \lesssim s\log(1/\varepsilon) + \log(1/\varepsilon) \lesssim s\log(1/\varepsilon).$$

An application of the maximal inequality in Corollary 19.35 of van der Vaart (1998) yields

$$EW_j \lesssim n^{-1/2} J_{[]}(\|F\|_{P,2}, \mathcal{S}_j, L_2(P)) \lesssim n^{-1/2} \int_0^{\|F\|_{P,2}} \sqrt{s\log(1/\varepsilon)}\, d\varepsilon \lesssim \sqrt{s/n},$$

where F is a bounded envelope function. We then apply the functional Hoeffding inequality (Massart (2007)) to conclude that

$$P\big(W_j \geq C\sqrt{s/n}(1+x)\big) \leq P(W_j \geq EW_j + C\sqrt{s/n}x) \leq \exp(-Ksx^2),$$

which completes the proof.

Proof of Lemma 2. We first write

$$\begin{aligned} U_j(\boldsymbol{\beta}_0) &= \mathbb{P}_n \int_0^\tau \{X_j - \bar{X}_j(\boldsymbol{\beta}_0, t)\}\, dN(t) = \mathbb{P}_n \int_0^\tau \{X_j - \bar{X}_j(\boldsymbol{\beta}_0, t)\}\, dM(t) \\ &= \mathbb{P}_n \int_0^\tau X_j\, dM(t) - \mathbb{P}_n \int_0^\tau \bar{X}_j(\boldsymbol{\beta}_0, t)\, dM(t) \equiv T_1 - T_2, \end{aligned}$$

where $M(t) = N(t) - \int_0^t Y(s)\lambda_0(s)\exp(\boldsymbol{\beta}_0^T\mathbf{X})\, ds$ is the counting process martingale and $\bar{X}_j(\cdot,\cdot)$ is the jth component of $\overline{\mathbf{X}}(\cdot,\cdot)$. Note that term T_1 is an independent sum of mean-zero, bounded random variables, and an application of Hoeffding's inequality (Hoeffding (1963)) gives $P(|T_1| \geq n^{-1/2}x) \leq 2\exp(-Kx^2)$.

Next consider term T_2. From Lemma 4, we have

$$P(\sup_{t\in[0,\tau]} |S^{(0)}(\boldsymbol{\beta}_0, t) - s^{(0)}(\boldsymbol{\beta}_0, t)| \geq \delta) \leq \exp(-Kn)$$

and

$$P(\sup_{t\in[0,\tau]} |S_j^{(1)}(\boldsymbol{\beta}_0, t) - s_j^{(1)}(\boldsymbol{\beta}_0, t)| \geq \delta) \leq \exp(-Kn),$$

for some constant $\delta > 0$ and $j = 1, \ldots, p$. From now on, we condition on the event that $\sup_{t\in[0,\tau]} |S^{(0)}(\boldsymbol{\beta}_0, t) - s^{(0)}(\boldsymbol{\beta}_0, t)| \leq \delta$ and $\sup_{t\in[0,\tau]} |S_j^{(1)}(\boldsymbol{\beta}_0, t) - s_j^{(1)}(\boldsymbol{\beta}_0, t)| \leq$

δ for $j = 1, \ldots, p$, and bound T_2. Write

$$\bar{X}_j(\boldsymbol{\beta}_0, t) - e_j(\boldsymbol{\beta}_0, t) = \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, t)}\{S_j^{(1)}(\boldsymbol{\beta}_0, t) - s_j^{(1)}(\boldsymbol{\beta}_0, t)\} - \frac{s_j^{(1)}(\boldsymbol{\beta}_0, t)}{S^{(0)}(\boldsymbol{\beta}_0, t)s^{(0)}(\boldsymbol{\beta}_0, t)}\{S^{(0)}(\boldsymbol{\beta}_0, t) - s^{(0)}(\boldsymbol{\beta}_0, t)\},$$

where $e_j(\cdot, \cdot)$ is the jth component of $\mathbf{e}(\cdot, \cdot)$. Since $S^{(0)}(\boldsymbol{\beta}_0, \cdot)$ and $s^{(0)}(\boldsymbol{\beta}_0, \cdot)$ are bounded away from zero on $[0, \tau]$, the above representation implies that $\sup_{t \in [0,\tau]} |\bar{X}_j(\boldsymbol{\beta}_0, t) - e_j(\boldsymbol{\beta}_0, t)| \leq \delta'$ for some constant $\delta' > 0$. Note also that $\bar{X}_j(\boldsymbol{\beta}_0, \cdot)$ is of uniformly bounded variation.

Let $\mathcal{F}_j$ be the class of functions $f\colon [0, \tau] \to \mathbb{R}$ of uniformly bounded variation and such that $\sup_{t \in [0,\tau]} |f(t) - e_j(t)| \leq \delta'$. By constructing hypercubes centered at piecewise constant functions on a grid, one can show that the entropy of $\mathcal{F}_j$ satisfies

$$\log N(\varepsilon, \mathcal{F}_j, \|\cdot\|_\infty) \lesssim (1/\varepsilon)\log(1/\varepsilon). \tag{S.5}$$

Furthermore, let $\mathcal{M}_j$ be the class of functions $\{\int_0^\tau f(t)\, dM(t)\colon f \in \mathcal{F}_j\}$ and denote $V_j = \sup_{g \in \mathcal{M}_j} |(\mathbb{P}_n - P)g| = \sup_{g \in \mathcal{M}_j} |\mathbb{P}_n g|$. Note that, for any $f_1, f_2 \in \mathcal{F}_j$,

$$\left| \int_0^\tau f_1(t)\, dM(t) - \int_0^\tau f_2(t)\, dM(t) \right| \leq \sup_{u \in [0,\tau]} |f_1(u) - f_2(u)| \int_0^\tau |dM(t)|.$$

This, in view of Theorem 2.7.11 of van der Vaart and Wellner (1996), implies that

$$N_{[\,]}(2\varepsilon\|F\|_{P,2}, \mathcal{M}_j, L_2(P)) \leq N(\varepsilon, \mathcal{F}_j, \|\cdot\|_\infty), \tag{S.6}$$

where $F = \int_0^\tau |dM(t)|$ is bounded. An application of the maximal inequality in Corollary 19.35 of van der Vaart (1998), along with (S.5) and (S.6), yields

$$\begin{aligned} EV_j &\lesssim n^{-1/2} J_{[\,]}(\|G\|_{P,2}, \mathcal{M}_j, L_2(P)) \\ &\lesssim n^{-1/2} \int_0^{\|G\|_{P,2}} \sqrt{(1/\varepsilon)\log(1/\varepsilon)}\, d\varepsilon \lesssim n^{-1/2}, \end{aligned}$$

where G is a bounded envelope function.

We now apply the functional Hoeffding inequality to obtain

$$P\big(|T_2| \geq Cn^{-1/2}(1 + x)\big) \leq P(|T_2| \geq EV_j + Cn^{-1/2}x) \leq \exp(-Kx^2).$$

Combining the bounds for T_1 and T_2 gives the desired inequality.

Proof of Lemma 3. We first write

$$\begin{aligned}\mathcal{I}_{ij}(\boldsymbol{\beta}) - \sigma_{ij}(\boldsymbol{\beta}) &= \int_0^\tau \{S_{ij}^{(2)}(\boldsymbol{\beta}, t) - s_{ij}^{(2)}(\boldsymbol{\beta}, t)\}\, dt \\ &\quad + \int_0^\tau \left\{ \frac{S_i^{(1)}(\boldsymbol{\beta}, t) S_j^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \frac{s_i^{(1)}(\boldsymbol{\beta}, t) s_j^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} \right\} dt \\ &\equiv T_1(\boldsymbol{\beta}) + T_2(\boldsymbol{\beta}).\end{aligned}$$

It follows from (S.4) in Lemma 4 that

$$P\left(\sup_{\boldsymbol{\beta} \in \mathcal{B}_0} |T_1(\boldsymbol{\beta})| \geq C\sqrt{s/n}(1+x) \right) \leq \exp(-Ksx^2).$$

To bound term $T_2(\boldsymbol{\beta})$, write

$$\begin{aligned}&\frac{S_i^{(1)}(\boldsymbol{\beta}, t) S_j^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \frac{s_i^{(1)}(\boldsymbol{\beta}, t) s_j^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} \\ &\quad = \frac{S_j^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \{S_i^{(1)}(\boldsymbol{\beta}, t) - s_i^{(1)}(\boldsymbol{\beta}, t)\} + \frac{s_i^{(\boldsymbol{\beta}, t)}}{S^{(0)}(\boldsymbol{\beta}, t)} \{S_j^{(1)}(\boldsymbol{\beta}, t) - s_j^{(1)}(\boldsymbol{\beta}, t)\} \\ &\quad\quad - \frac{s_i^{(1)}(\boldsymbol{\beta}, t) s_j^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t) s^{(0)}(\boldsymbol{\beta}, t)} \{S^{(0)}(\boldsymbol{\beta}, t) - s^{(0)}(\boldsymbol{\beta}, t)\}.\end{aligned}$$

As in the proof of Lemma 2, it is sufficient to condition on the event that $\sup_{\boldsymbol{\beta} \in \mathcal{B}_0,\, t \in [0,\tau]} |S^{(0)}(\boldsymbol{\beta}, t) - s^{(0)}(\boldsymbol{\beta}, t)| \leq \delta$ and $\sup_{\boldsymbol{\beta} \in \mathcal{B}_0,\, t \in [0,\tau]} |S_j^{(1)}(\boldsymbol{\beta}, t) - s_j^{(1)}(\boldsymbol{\beta}, t)| \leq \delta$ for some constant $\delta > 0$ and $j = 1, \ldots, p$. Then (S.2) and (S.3) in Lemma 4 imply that

$$P\left(\sup_{\boldsymbol{\beta} \in \mathcal{B}_0} |T_2(\boldsymbol{\beta})| \geq C\sqrt{s/n}(1+x) \right) \leq D\exp(-Ksx^2).$$

Putting the bounds for $T_1(\boldsymbol{\beta})$ and $T_2(\boldsymbol{\beta})$ together completes the proof.

References

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Massart, P. (2007). *Concentration Inequalities and Model Selection.* Springer, Berlin.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press, New York.

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