Model-assisted estimation in high-dimensional settings for survey data

Supplementary material

Mehdi DAGDOUG[(1)], Camelia GOGA[(1)] and David HAZIZA[(2)]

(1) Université de Bourgogne Franche-Comté,

Laboratoire de Mathématiques de Besançon, Besançon, FRANCE

(2) University of Ottawa, Departement of Mathematics and Statistics,

Ottawa, CANADA

Asymptotic assumptions

(H1) We assume that there exists a positive constant C_1 such that $N_v^{-1} \sum_{i \in U_v} y_i^2 < C_1$.

(H2) We assume that $\lim_{v \to \infty} \frac{n_v}{N_v} = \pi \in (0,1)$.

(H3) There exists a positive constant c such that $\min_{i \in U_v} \pi_i \geqslant c > 0$; also, we assume that $\limsup_{v \to \infty} n_v \max_{i \neq \ell \in U_v} |\pi_{i\ell} - \pi_i \pi_\ell| < \infty$.

(H4) We assume that there exists a positive constant C_2 such that, for all $i \in U_v$, $||\mathbf{x}_i||_2^2 \leq C_2 p_v$, where $|| \cdot ||_2$ denotes the usual Euclidean norm.

(H5) We assume that $||\widehat{\boldsymbol{\beta}}||_1 = \mathcal{O}_{\mathrm{p}}(p_v)$, where $\widehat{\boldsymbol{\beta}}$ is the least square estimator given in (8) in the main article and $|| \cdot ||_1$ denotes the L^1 norm.

Proof of Result 3.1

Result 3.1. *Assume (H1)-(H5). Consider a sequence of GREG estimators $\{\widehat{t}_{greg}\}_{v \in \mathbb{N}}$ of t_y. Then,*

$$\frac{1}{N_v}(\widehat{t}_{\mathrm{greg}} - t_y) = \mathcal{O}_{\mathrm{p}}\left(\sqrt{\frac{p_v^3}{n_v}}\right).$$

If the numbers of auxiliary variables $\{p_v\}_{v\in\mathbb{N}}$ *and the sample sizes* $\{n_v\}_{v\in\mathbb{N}}$ *satisfy* $p_v^3/n_v = o(1)$*, then* $N_v^{-1}(\widehat{t}_{greg} - t_y) = o_{\mathrm{p}}(1)$.

Proof. We adapt the proof of Robinson and Särndal (1983) to a high-dimensional setting. Let I_i be the sample membership indicator for unit i such that $I_i = 1$ if $i \in S$ and $I_i = 0$, otherwise. Let $\alpha_i := I_i/\pi_i - 1$ for all $i \in U_v$. We consider the following decomposition:

$$\frac{1}{N_v}\left(\widehat{t}_{\mathrm{greg}} - t_y\right) = \frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i - \sum_{j=1}^{p_v} b_j\widehat{\beta}_j, \tag{1}$$

where $b_j = \dfrac{1}{N_v}\sum_{i\in U_v}\alpha_i x_{ij}$ for $j = 1, 2, ..., p_v$. Now, the first term does not depend on the auxiliary information and we have (Robinson and Särndal, 1983; Breidt and Opsomer, 2000):

$$\mathbb{E}_{\mathrm{p}}\left(\frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i\right)^2 = \frac{1}{N_v^2}\sum_{i\in U} y_i^2\cdot\mathbb{E}_{\mathrm{p}}(\alpha_i^2) + \frac{1}{N_v^2}\sum_{i\in U_v}\sum_{\ell\in U_v,\ell\neq i} y_i y_\ell\cdot\mathbb{E}_{\mathrm{p}}(\alpha_i\alpha_\ell). \tag{2}$$

We have $\mathbb{E}_{\mathrm{p}}(\alpha_i^2) = (1-\pi_i)/\pi_i \leqslant 1/c$ and for $i \neq \ell$, $\mathbb{E}_{\mathrm{p}}(\alpha_i\alpha_\ell) = (\pi_{i\ell} - \pi_i\pi_\ell)/\pi_i\pi_\ell \leqslant \max_{i,\ell\in U_v,i\neq\ell}|\pi_{i\ell} - \pi_i\pi_\ell|/c^2$ by Assumption (H3). It follows from (H1), (H2) and (H3) that

$$\begin{aligned}\mathbb{E}_{\mathrm{p}}\left(\frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i\right)^2 &\leqslant \frac{1}{cN_v^2}\sum_{i\in U_v} y_i^2 + \frac{n_v\max_{i,\ell\in U_v,i\neq\ell}|\pi_{i\ell}-\pi_i\pi_\ell|}{c^2 n_v N_v^2}\sum_{i\in U_v}\sum_{\ell\in U_v,\ell\neq i}|y_i y_\ell| \\ &\leqslant \left(\frac{1}{cN_v} + \frac{n_v\max_{i,\ell\in U_v,i\neq\ell}|\pi_{i\ell}-\pi_i\pi_\ell|}{c^2 n_v}\right)\frac{1}{N_v}\sum_{i\in U_v} y_i^2 = \mathcal{O}\left(\frac{1}{n_v}\right)\end{aligned} \tag{3}$$

and so,

$$\left|\frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i\right| = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{n_v}}\right). \tag{4}$$

Now, consider the second term from the right-side of (1):

$$\left|\sum_{j=1}^{p_v}\widehat{\beta}_j b_j\right| \leqslant \sqrt{\left(\sum_{j=1}^{p_v}\widehat{\beta}_j^2\right)\left(\sum_{j=1}^{p_v} b_j^2\right)} = ||\widehat{\boldsymbol{\beta}}||_2\sqrt{\sum_{j=1}^{p_v} b_j^2} \leqslant ||\widehat{\boldsymbol{\beta}}||_1\sqrt{\sum_{j=1}^{p_v} b_j^2}. \tag{5}$$

By Assumption (H5), we have that $||\widehat{\boldsymbol{\beta}}||_1 = \mathcal{O}_{\mathrm{p}}(p_v)$. Furthermore,

$$\sqrt{\sum_{j=1}^{p_v} b_j^2} = \frac{1}{N_v}\left|\left|\sum_{i\in U_v}\alpha_i\mathbf{x}_i\right|\right|_2$$

and

$$\begin{aligned}\frac{1}{N_v^2}\mathbb{E}_{\mathrm{p}}\left|\left|\sum_{i\in U_v}\alpha_i\mathbf{x}_i\right|\right|_2^2 &= \frac{1}{N_v^2}\sum_{i\in U_v}||\mathbf{x}_i||_2^2\mathbb{E}_{\mathrm{p}}(\alpha_i^2)+\frac{1}{N_v^2}\sum_{i\in U_v}\sum_{\ell\neq i\in U_v}\mathbf{x}_i^\top\mathbf{x}_\ell\mathbb{E}_{\mathrm{p}}(\alpha_i\alpha_\ell)\\ &\leqslant \frac{1}{cN_v^2}\sum_{i\in U_v}||\mathbf{x}_i||_2^2+\frac{n_v\max_{i,\ell\in U_v,i\neq\ell}|\pi_{i\ell}-\pi_i\pi_\ell|}{c^2n_vN_v^2}\sum_{i\in U_v}\sum_{\ell\neq i\in U_v}|\mathbf{x}_i^\top\mathbf{x}_\ell|\\ &\leqslant\left(\frac{1}{cN_v}+\frac{n_v\max_{i,\ell\in U_v,i\neq\ell}|\pi_{i\ell}-\pi_i\pi_\ell|}{c^2n_v}\right)\frac{1}{N_v}\sum_{i\in U_v}||\mathbf{x}_i||_2^2\\ &=\mathcal{O}\left(\frac{p_v}{n_v}\right), \qquad (6)\end{aligned}$$

by Assumptions (H2)-(H4). It follows that

$$\sqrt{\sum_{j=1}^{p_v}b_j^2}=\mathcal{O}_{\mathrm{p}}\left(\sqrt{\frac{p_v}{n_v}}\right). \qquad (7)$$

The result follows by using (1), (4), (5), (7) and Assumption (H5):

$$\frac{1}{N_v}\left|\widehat{t}_{\mathrm{greg}}-t_y\right|\leqslant\left|\frac{1}{N_v}\sum_{i\in U_v}\alpha_iy_i\right|+\left|\sum_{j=1}^{p_v}\widehat{\beta}_jb_j\right|=\mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{n_v}}\right)+\mathcal{O}_{\mathrm{p}}\left(\sqrt{\frac{p_v^3}{n_v}}\right)=\mathcal{O}_{\mathrm{p}}\left(\sqrt{\frac{p_v^3}{n_v}}\right).$$

■

Proof of Result 3.2

Result 3.2. *Assume (H1)-(H5). Consider a sequence of penalized model-assisted estimators $\{\widehat{t}_{\mathrm{pen}}\}_{v\in\mathbb{N}}$ of t_y obtained by either ridge, lasso or elastic-net. Then,*

$$\frac{1}{N_v}(\widehat{t}_{\mathrm{pen}}-t_y)=\mathcal{O}_{\mathrm{p}}\left(\sqrt{\frac{p_v^3}{n_v}}\right).$$

If the numbers of auxiliary variables $\{p_v\}_{v\in\mathbb{N}}$ and the sample sizes $\{n_v\}_{v\in\mathbb{N}}$ satisfy $p_v^3/n_v = o(1)$, then $N_v^{-1}(\widehat{t}_{pen}-t_y)=o_{\mathrm{p}}(1)$.

Proof. From the proof of result (3.1), we only need to show that $||\hat{\boldsymbol{\beta}}_{\mathrm{pen}}||_2=\mathcal{O}_{\mathrm{p}}(p_v)$ or $||\hat{\boldsymbol{\beta}}_{\mathrm{pen}}||_1=\mathcal{O}_{\mathrm{p}}(p_v)$, where $\hat{\boldsymbol{\beta}}_{\mathrm{pen}}$ is one of the penalized regression coefficient: ridge, lasso and elastic-net.

Consider first the ridge regression coefficient, $\hat{\boldsymbol{\beta}}_{\mathrm{ridge}}$. The ridge regression estimator has the advantage of having an explicit expression. We will show that $||\hat{\boldsymbol{\beta}}_{\text{ridge}}||_2 < ||\hat{\boldsymbol{\beta}}||_2$ for $\lambda > 0$. Let denote $\hat{T}_\lambda = \boldsymbol{X}_{S_v}^\top \boldsymbol{\Pi}_{S_v}^{-1} \boldsymbol{X}_{S_v} + \lambda \boldsymbol{I}_{p_v} = \sum_{i \in S_v} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\pi_i} + \lambda \boldsymbol{I}_{p_v}$ sample counterpart of $T_\lambda = \boldsymbol{X}_{U_v}^\top \boldsymbol{X}_{U_v} + \lambda \boldsymbol{I}_{p_v} = \sum_{i \in U_v} \mathbf{x}_i \mathbf{x}_i^\top + \lambda \boldsymbol{I}_{p_v}$. Moreover, let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_{p_v}$ be the eigenvalues of $\sum_{i \in S_v} \mathbf{x}_i \mathbf{x}_i^\top / \pi_i$ in decreasing order and $\hat{\mathbf{v}}_j$ the orthonormal corresponding eigenvectors, $j = 1, \ldots, p_v$. Then, the eigenvalues of the matrix $\hat{T}_\lambda$ are $\hat{\lambda}_1 + \lambda \geq \hat{\lambda}_2 + \lambda \geq \ldots \geq \hat{\lambda}_{p_v} + \lambda \geq \lambda > 0$ with the same eigenvectors $\hat{\mathbf{v}}_j, j = 1, \ldots, p_v$. Using the same arguments as those used in Hoerl and Kennard (1970), we obtain $\hat{\boldsymbol{\beta}}_{\text{ridge}} = \sum_{j=1}^{p_v} (\hat{\lambda}_j + \lambda)^{-1} \hat{\mathbf{v}}_j \hat{\mathbf{v}}_j^\top \boldsymbol{X}_{S_v}^\top \boldsymbol{\Pi}_{S_v}^{-1} \mathbf{y}_{S_v}$ and $\hat{\boldsymbol{\beta}} = \sum_{j=1}^{p_v} (\hat{\lambda}_j)^{-1} \hat{\mathbf{v}}_j \hat{\mathbf{v}}_j^\top \boldsymbol{X}_{S_v}^\top \boldsymbol{\Pi}_{S_v}^{-1} \mathbf{y}_{S_v}$. Let denote by $c_j = \hat{\mathbf{v}}_j^\top \boldsymbol{X}_{S_v}^\top \boldsymbol{\Pi}_{S_v}^{-1} \mathbf{y}_{S_v} \in \mathbf{R}$, then

$$||\hat{\boldsymbol{\beta}}_{\text{ridge}}||_2^2 = \sum_{j=1}^{p_v} \frac{c_j^2}{(\hat{\lambda}_j + \lambda)^2} < ||\hat{\boldsymbol{\beta}}||_2^2 = \sum_{j=1}^{p_v} \frac{c_j^2}{(\hat{\lambda}_j)^2} \quad \text{for} \quad \lambda > 0.$$

It follows that $||\hat{\boldsymbol{\beta}}_{\text{ridge}}||_2 < ||\hat{\boldsymbol{\beta}}||_2 \leq ||\hat{\boldsymbol{\beta}}||_1 = \mathcal{O}_{\text{p}}(p_v)$ and we get $||\hat{\boldsymbol{\beta}}_{\text{ridge}}||_2 = \mathcal{O}_{\text{p}}(p_v)$.

We now consider the lasso regression estimator, $\hat{\boldsymbol{\beta}}_{\text{lasso}}$, which minimizes the design-based version of the optimization problem given in (13) in the main article:

$$\hat{\boldsymbol{\beta}}_{\text{lasso}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i \in S_v} \frac{1}{\pi_i} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \lambda ||\boldsymbol{\beta}||_1.$$

The lasso-estimator $\hat{\boldsymbol{\beta}}_{\text{lasso}}$ may be also obtained as the solution of a constrained optimization problem:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i \in S_v} \frac{1}{\pi_i} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$$

under the constraint

$$||\boldsymbol{\beta}||_1 \leq C,$$

for some small enough constant $C > 0$. If the ordinary least-square estimator $\widehat{\boldsymbol{\beta}}$ satisfies the constraint, namely if $||\widehat{\boldsymbol{\beta}}||_1 \leq C$, then the solution of the constrained optimization problem is $\widehat{\boldsymbol{\beta}}_{\text{lasso}} = \widehat{\boldsymbol{\beta}}$; otherwise, if $||\widehat{\boldsymbol{\beta}}||_1 > C$, then the solution $\widehat{\boldsymbol{\beta}}_{\text{lasso}}$ will be different from the least-square estimator $\widehat{\boldsymbol{\beta}}$ and $||\widehat{\boldsymbol{\beta}}_{\text{lasso}}||_1 \leq C < ||\widehat{\boldsymbol{\beta}}||_1$. So, in both cases, we have $||\widehat{\boldsymbol{\beta}}_{\text{lasso}}||_1 \leq ||\widehat{\boldsymbol{\beta}}||_1 = \mathcal{O}_{\text{p}}(p_v)$.

Finally, consider the elastic-net regression estimator, $\hat{\boldsymbol{\beta}}_{\text{en}}$. Consider the following objective functions:

$$\mathcal{L}_{ols}(\boldsymbol{\beta}) = \sum_{i \in S_v} \frac{1}{\pi_i}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$$
$$\mathcal{L}_{en}(\boldsymbol{\beta}) = \sum_{i \in S_v} \frac{1}{\pi_i}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \lambda_1 ||\boldsymbol{\beta}||_1 + \lambda_2 ||\boldsymbol{\beta}||_2^2 = \mathcal{L}_{ols}(\boldsymbol{\beta}) + \lambda_1 ||\boldsymbol{\beta}||_1 + \lambda_2 ||\boldsymbol{\beta}||_2^2,$$

where $\lambda_1 = \lambda\alpha$ and $\lambda_2 = \lambda(1-\alpha)$ with $\lambda > 0$ and $\alpha \in (0,1)$. The cases $\alpha = 0$ and $\alpha = 1$ lead, respectively, to the ridge and lasso regression estimators which have been discussed above. The ordinary least squares estimator $\widehat{\boldsymbol{\beta}}$ verifies $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathcal{L}_{ols}(\boldsymbol{\beta})$ and the elastic-net estimator verifies $\widehat{\boldsymbol{\beta}}_{en} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathcal{L}_{en}(\boldsymbol{\beta})$. Since $\widehat{\boldsymbol{\beta}}$ minimizes $\mathcal{L}_{ols}(\boldsymbol{\beta})$, we have $\mathcal{L}_{ols}(\widehat{\boldsymbol{\beta}}) \leqslant \mathcal{L}_{ols}(\widehat{\boldsymbol{\beta}}_{en})$. Similarly, we have $\mathcal{L}_{en}(\widehat{\boldsymbol{\beta}}_{en}) \leqslant \mathcal{L}_{en}(\widehat{\boldsymbol{\beta}}_{ols})$. Therefore, the following inequalities hold:

$$\begin{aligned} \mathcal{L}_{ols}(\widehat{\boldsymbol{\beta}}) + \lambda_1 ||\widehat{\boldsymbol{\beta}}_{en}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}_{en}||_2^2 &\leqslant \mathcal{L}_{ols}(\widehat{\boldsymbol{\beta}}_{en}) + \lambda_1 ||\widehat{\boldsymbol{\beta}}_{en}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}_{en}||_2^2 = \mathcal{L}_{en}(\widehat{\boldsymbol{\beta}}_{en}) \\ &\leqslant \mathcal{L}_{ols}(\widehat{\boldsymbol{\beta}}) + \lambda_1 ||\widehat{\boldsymbol{\beta}}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}||_2^2 = \mathcal{L}_{en}(\widehat{\boldsymbol{\beta}}_{ols}), \end{aligned}$$

which implies

$$\lambda_1 ||\widehat{\boldsymbol{\beta}}_{en}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}_{en}||_2^2 \leqslant \lambda_1 ||\widehat{\boldsymbol{\beta}}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}||_2^2. \tag{8}$$

Furthermore, since $\lambda_1 > 0$, we can write

$$\lambda_2 ||\widehat{\boldsymbol{\beta}}_{en}||_2^2 \leqslant \lambda_1 ||\widehat{\boldsymbol{\beta}}_{en}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}_{en}||_2^2. \tag{9}$$

Using (8), (9) and the fact that $||\widehat{\boldsymbol{\beta}}||_2 \leqslant ||\widehat{\boldsymbol{\beta}}||_1$, we obtain

$$\lambda_2 ||\widehat{\boldsymbol{\beta}}_{en}||_2^2 \leqslant \lambda_1 ||\widehat{\boldsymbol{\beta}}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}||_2^2 \leqslant \lambda_1 ||\widehat{\boldsymbol{\beta}}||_1 + \lambda_2 ||\widehat{\boldsymbol{\beta}}||_1^2$$

which implies

$$||\widehat{\boldsymbol{\beta}}_{en}||_2^2 \leqslant \frac{\alpha}{1-\alpha} ||\widehat{\boldsymbol{\beta}}||_1 + ||\widehat{\boldsymbol{\beta}}||_1^2 = \mathcal{O}_{\mathrm{p}}(p_v^2)$$

and so, $||\widehat{\boldsymbol{\beta}}_{en}||_2 = \mathcal{O}_{\mathrm{p}}(p_v)$.

■

Proof of Result 3.3

Result 3.3. *Assume (H1)-(H4). Also, assume that there exists a positive constant* $\tilde{C}$ *such that* $\lambda_{max}(\boldsymbol{X}_{U_v}^\top \boldsymbol{X}_{U_v}) \leqslant \tilde{C} N_v$*, where* $\lambda_{max}(\boldsymbol{X}_{U_v}^\top \boldsymbol{X}_{U_v})$ *is the largest eigenvalue of* $\boldsymbol{X}_{U_v}^\top \boldsymbol{X}_{U_v}$*. Assume also that* $N_v/\lambda_v = \mathcal{O}(1)$*.*

1. *Then, there exists a positive constant* C *such that* $\mathbb{E}_p\left[||\widehat{\boldsymbol{\beta}}_{\text{ridge}}||_2^2\right] \leqslant C$ *and*

$$\frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{\text{ridge}} - t_y\right| = \mathcal{O}\left(\sqrt{\frac{p_v}{n_v}}\right).$$

If the numbers of auxiliary variables $\{p_v\}_{v\in\mathbb{N}}$ *and the sample sizes* $\{n_v\}_{v\in\mathbb{N}}$ *satisfy* $p_v/n_v = o(1)$*, then* $N_v^{-1}\mathbb{E}_p|\widehat{t}_{\text{ridge}} - t_y| = o(1)$*.*

2. $\mathbb{E}_{\text{p}}(||\widehat{\boldsymbol{\beta}}_{\text{ridge}} - \tilde{\boldsymbol{\beta}}_{\text{ridge}}||_2^2) = \mathcal{O}(p_v/n_v)$*. Thus, if* $p_v/n_v = o(1)$*, then* $\mathbb{E}_{\text{p}}(||\widehat{\boldsymbol{\beta}}_{\text{ridge}} - \tilde{\boldsymbol{\beta}}_{\text{ridge}}||_2^2) = o(1)$*.*

3. *We have the following asymptotic equivalence:*

$$\frac{1}{N_v}\left(\widehat{t}_{\text{ridge}} - t_y\right) = \frac{1}{N_v}\left(\widehat{t}_{\text{diff},\lambda} - t_y\right) + \mathcal{O}_{\text{p}}\left(\frac{p_v}{n_v}\right),$$

where

$$\widehat{t}_{\text{diff},\lambda} = \sum_{i\in S_v} y_i/\pi_i - \left(\sum_{i\in S_v}\mathbf{x}_i/\pi_i - \sum_{i\in U_v}\mathbf{x}_i\right)^\top \tilde{\boldsymbol{\beta}}_{\text{ridge}}$$

and

$$\frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{\text{ridge}} - t_y\right| = \mathcal{O}\left(\frac{1}{\sqrt{n_v}}\right) + \mathcal{O}\left(\frac{p_v}{n_v}\right).$$

If $p_v = \mathcal{O}(n_v^a)$ *with* $0 \leq a < 1/2$*, then*

$$\frac{1}{N_v}\left(\widehat{t}_{\text{ridge}} - t_y\right) = \frac{1}{N_v}\left(\widehat{t}_{\text{diff},\lambda} - t_y\right) + o_{\text{p}}(1)$$

and

$$\frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{\text{ridge}} - t_y\right| = \mathcal{O}\left(\frac{1}{\sqrt{n_v}}\right).$$

Proof. 1. As in the proof of result (3.2), we consider the eigenvalues of the matrix $\hat{T}_\lambda$ in decreasing order: $\hat{\lambda}_1 + \lambda \geq \hat{\lambda}_2 + \lambda \geq \ldots \geq \hat{\lambda}_{p_v} + \lambda \geq \lambda > 0$. The matrix $\hat{T}_\lambda$ is always invertible and the eigenvalues of $\hat{T}_\lambda^{-1}$ are $0 < (\hat{\lambda}_1 + \lambda)^{-1} \leq (\hat{\lambda}_2 + \lambda)^{-1} \leq \ldots \leq (\hat{\lambda}_{p_v} + \lambda)^{-1} \leq \lambda^{-1}$. We then obtain

$$||\hat{T}_\lambda^{-1}||_2 \leqslant \lambda^{-1}, \tag{10}$$

where $||\cdot||_2$ is the spectral norm matrix defined for a squared $p \times p$ matrix $\mathbf{A}$ as $||\mathbf{A}||_2 = \sup_{\mathbf{x}\in\mathbf{R}^p, ||\mathbf{x}||_2 \neq 0} ||\mathbf{A}\mathbf{x}||_2/||\mathbf{x}||_2$. For a symmetric and positive definite matrix $\mathbf{A}$, we have that $||\mathbf{A}||_2 = \lambda_{max}(\mathbf{A})$, where $\lambda_{max}(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$. Now, we can write

$$\begin{aligned} \frac{1}{N_v^2}\left|\left|\sum_{i\in S_v}\frac{\mathbf{x}_i y_i}{\pi_i}\right|\right|_2^2 &= \frac{1}{N_v^2}\sum_{i\in U_v}\sum_{\ell\in U_v}\mathbf{x}_i^\top \mathbf{x}_\ell \frac{y_i I_i}{\pi_i}\frac{y_\ell I_\ell}{\pi_\ell} = \frac{1}{N_v^2}\mathcal{Y}^\top \boldsymbol{X}_{U_v}\boldsymbol{X}_{U_v}^\top \mathcal{Y} \\ &\leqslant \frac{1}{N_v}||\mathcal{Y}||_2^2 \frac{1}{N_v}||\boldsymbol{X}_{U_v}\boldsymbol{X}_{U_v}^\top||_2, \end{aligned}$$

where $\mathcal{Y}^\top = \left(\frac{y_i I_i}{\pi_i}\right)_{i\in U_v}$. The symmetric and positive semi-definite $N_v \times N_v$ matrix $\boldsymbol{X}_{U_v}\boldsymbol{X}_{U_v}^\top$ has the same non-null eigenvalues as those of the positive definite $p_v \times p_v$ matrix $\boldsymbol{X}_{U_v}^\top \boldsymbol{X}_{U_v}$,. Therefore,

$$\frac{1}{N_v}||\boldsymbol{X}_{U_v}\boldsymbol{X}_{U_v}^\top||_2 = \frac{1}{N_v}\lambda_{max}(\boldsymbol{X}_{U_v}^\top \boldsymbol{X}_{U_v}) \leqslant \tilde{C}.$$

Using Assumptions (H1) and (H3), we have

$$\frac{1}{N_v^2}\left|\left|\sum_{i\in S_v}\frac{\mathbf{x}_i y_i}{\pi_i}\right|\right|_2^2 \leqslant \frac{\tilde{C}}{N_v}||\mathcal{Y}||_2^2 = \frac{\tilde{C}}{N_v}\sum_{i\in U_v}\frac{y_i^2 I_i}{\pi_i^2} \leqslant \frac{\tilde{C}}{c^2 N_v}\sum_{i\in U_v} y_i^2 = \mathcal{O}(1).$$

Finally, using also the fact that $N_v/\lambda = O(1)$, we have

$$||\widehat{\boldsymbol{\beta}}_{\text{ridge}}||_2^2 \leqslant ||\hat{T}_\lambda^{-1}||_2^2 \left|\left|\sum_{i\in S_v}\frac{\mathbf{x}_i y_i}{\pi_i}\right|\right|_2^2 \leqslant N_v^2\lambda^{-2}\left|\left|\frac{1}{N_v^2}\sum_{i\in S_v}\frac{\mathbf{x}_i y_i}{\pi_i}\right|\right|_2^2 = \mathcal{O}(1).$$

It follows that

$$\mathbb{E}_{\text{p}}\left[||\widehat{\boldsymbol{\beta}}_{\text{ridge}}||_2^2\right] = \mathcal{O}(1). \tag{11}$$

To obtain the L^1 design-consistency of the ridge model-assisted estimator, we write as in the proof of Result 3.1:

$$\begin{aligned}\frac{1}{N_v}\left(\widehat{t}_{\text{ridge}}-t_y\right) &= \frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i-\sum_{j=1}^{p_v} b_j\widehat{\beta}_{j,\text{ridge}}\\ &= \frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i-\frac{1}{N_v}\left(\sum_{i\in U_v}\alpha_i\mathbf{x}_i\right)^\top\widehat{\boldsymbol{\beta}}_{\text{ridge}}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\text{p}}\left|\frac{1}{N_v}\left(\widehat{t}_{\text{ridge}}-t_y\right)\right| &\leqslant \mathbb{E}_{\text{p}}\left|\frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i\right|+\sqrt{\mathbb{E}_{\text{p}}\left(\frac{1}{N_v^2}\left|\left|\sum_{i\in U_v}\alpha_i\mathbf{x}_i\right|\right|_2^2\right)\mathbb{E}_{\text{p}}||\widehat{\boldsymbol{\beta}}_{\text{ridge}}||_2^2}\\ &= \mathcal{O}\left(\sqrt{\frac{1}{n_v}}\right)+\mathcal{O}\left(\sqrt{\frac{p_v}{n_v}}\right)=\mathcal{O}\left(\sqrt{\frac{p_v}{n_v}}\right)\end{aligned}$$

by (3), (6), (11).

2. We can write

$$\widehat{\boldsymbol{\beta}}_{\text{ridge}}-\widetilde{\boldsymbol{\beta}}_{\text{ridge}}=\widehat{T}_\lambda^{-1}\left(\sum_{i\in S_v}\frac{E_{i\lambda}}{\pi_i}-\sum_{i\in U_v}E_{i\lambda}\right), \tag{12}$$

where $E_{i\lambda}=\mathbf{x}_i(y_i-\mathbf{x}_i^\top\tilde{\boldsymbol{\beta}}_{\text{ridge}})$ with $\sum_{i\in U_v}E_{i\lambda}=\lambda\boldsymbol{I}_{p_v}\tilde{\boldsymbol{\beta}}_{\text{ridge}}$. Using the same arguments as those used in the proof of Result 3.1, we get

$$\frac{1}{N_v^2}\mathbb{E}_{\text{p}}\left|\left|\sum_{i\in S_v}\frac{E_{i\lambda}}{\pi_i}-\sum_{i\in U_v}E_{i\lambda}\right|\right|_2^2\leqslant\left(\frac{1}{cN_v}+\frac{n_v\max_{i,\ell\in U_v,i\neq\ell}|\pi_{i\ell}-\pi_i\pi_\ell|}{c^2n_v}\right)\frac{1}{N_v}\sum_{i\in U_v}||E_{i\lambda}||_2^2. \tag{13}$$

Furthermore,

$$\frac{1}{N_v}\sum_{i\in U_v}||E_{i\lambda}||_2^2\leqslant\frac{2C_2p_v}{N_v}\left(\sum_{i\in U_v}y_i^2+\sum_{i\in U_v}(\mathbf{x}_i^\top\tilde{\boldsymbol{\beta}}_{\text{ridge}})^2\right)=\mathcal{O}(p_v) \tag{14}$$

by Assumptions (H1) and (H4) and the fact that

$$\frac{1}{N_v}\sum_{i\in U_v}(\mathbf{x}_i^\top\tilde{\boldsymbol{\beta}}_{\text{ridge}})^2=\tilde{\boldsymbol{\beta}}_{\text{ridge}}^\top\left(\frac{1}{N_v}\sum_{i\in U_v}\mathbf{x}_i\mathbf{x}_i^\top\right)\tilde{\boldsymbol{\beta}}_{\text{ridge}}\leqslant||\tilde{\boldsymbol{\beta}}_{\text{ridge}}||_2^2\frac{1}{N_v}||\boldsymbol{X}_{U_v}^\top\boldsymbol{X}_{U_v}||_2=\mathcal{O}(1). \tag{15}$$

To obtain the above inequality, we have also used the fact that $||\tilde{\boldsymbol{\beta}}_{\text{ridge}}||_2=\mathcal{O}(1)$

which can be proved by using the same arguments as the ones used for showing that $||\widehat{\boldsymbol{\beta}}_{\text{ridge}}||_2 = \mathcal{O}(1)$ in point (1). Expressions (13) and (14) lead to

$$\frac{1}{N_v^2}\mathbb{E}_{\text{p}}\left|\left|\sum_{i\in S_v}\frac{E_{i\lambda}}{\pi_i} - \sum_{i\in U_v} E_{i\lambda}\right|\right|_2^2 = \mathcal{O}\left(\frac{p_v}{n_v}\right). \tag{16}$$

The result follows from (12), (16) and the fact that $||N_v\hat{T}_\lambda^{-1}||_2 = \mathcal{O}(1)$:

$$\mathbb{E}_{\text{p}}||\widehat{\boldsymbol{\beta}}_{\text{ridge}} - \tilde{\boldsymbol{\beta}}_{\text{ridge}}||_2^2 = \mathcal{O}\left(\frac{p_v}{n_v}\right). \tag{17}$$

3. We use the following decomposition:

$$\frac{1}{N_v}\left(\widehat{t}_{\text{ridge}} - t_y\right) = \frac{1}{N_v}\left(\widehat{t}_{\text{diff},\lambda} - t_y\right) - \frac{1}{N_v}\left(\sum_{i\in S_v}\frac{\mathbf{x}_i}{\pi_i} - \sum_{i\in U_v}\mathbf{x}_i\right)^\top\left(\widehat{\boldsymbol{\beta}}_{\text{ridge}} - \tilde{\boldsymbol{\beta}}_{\text{ridge}}\right),$$

and

$$\begin{aligned}\frac{1}{N_v}\left(\widehat{t}_{\text{diff},\lambda} - t_y\right) &= \frac{1}{N_v}\left(\sum_{i\in S_v}\frac{y_i}{\pi_i} - \sum_{i\in U_v} y_i\right) - \frac{1}{N_v}\left(\sum_{i\in S_v}\frac{\mathbf{x}_i}{\pi_i} - \sum_{i\in U_v}\mathbf{x}_i\right)^\top \tilde{\boldsymbol{\beta}}_{\text{ridge}} \\ &= \frac{1}{N_v}\sum_{i\in U_v}\alpha_i y_i - \frac{1}{N_v}\sum_{i\in U_v}\alpha_i\mathbf{x}_i^\top\tilde{\boldsymbol{\beta}}_{\text{ridge}},\end{aligned}$$

where $\alpha_i = I_i/\pi_i - 1, i \in U_v$. From (3), we have that $N_v^{-2}\mathbb{E}_{\text{p}}(\sum_{i\in U_v}\alpha_i y_i)^2 = \mathcal{O}(n_v^{-1})$ and we can get $N_v^{-2}\mathbb{E}_{\text{p}}\left(\sum_{i\in U_v}\alpha_i\mathbf{x}_i^\top\tilde{\boldsymbol{\beta}}_{\text{ridge}}\right)^2 = \mathcal{O}(n_v^{-1})$ by using similar arguments as those used in the proof of Result 3.1 and (15). We obtain

$$\frac{1}{N_v^2}\mathbb{E}_{\text{p}}\left(\widehat{t}_{\text{diff},\lambda} - t_y\right)^2 = \mathcal{O}\left(\frac{1}{n_v}\right).$$

The result follows since

$$\begin{aligned}\frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{\text{ridge}} - t_y\right| &\leqslant \frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{\text{diff},\lambda} - t_y\right| + \sqrt{\frac{1}{N_v^2}\mathbb{E}_p\left|\left|\sum_{i\in S_v}\frac{\mathbf{x}_i}{\pi_i} - \sum_{i\in U_v}\mathbf{x}_i\right|\right|_2^2\mathbb{E}_p\left|\left|\widehat{\boldsymbol{\beta}}_{\text{ridge}} - \tilde{\boldsymbol{\beta}}_{\text{ridge}}\right|\right|_2^2} \\ &= \mathcal{O}\left(\frac{1}{\sqrt{n_v}}\right) + \mathcal{O}\left(\frac{p_v}{n_v}\right)\end{aligned}$$

by using (6) and (17).

∎

Proof of Proposition 3.1

Proposition 3.1. *Suppose assumptions (H1)-(H3) and that the sampling design and the X-variables are such that the columns of $\mathbf{\Pi}_{S_v}^{-1/2}\mathbf{X}_{S_v}$ are orthogonal. Suppose also that there exist positive quantities C_3 and C_4 such that $max_{j=1,\ldots,p_v} N_v^{-1}\sum_{i\in U_v} x_{ij}^4 \leq C_3 < \infty$ and $min_{j=1,\ldots,p_v} N_v^{-1}\sum_{i\in U_v} x_{ij}^2 \geq C_4 > 0$. Then, $N_v^{-1}(\widehat{t}_{\text{greg}} - t_y) = \mathcal{O}_{\text{p}}(\sqrt{p_v/n_v})$ and $N_v^{-1}(\widehat{t}_{\text{pen}} - t_y) = \mathcal{O}_{\text{p}}(\sqrt{p_v/n_v})$, where $\widehat{t}_{\text{pen}}$ denotes either the lasso or the elastic-net estimator.*

Proof. From the proof of Result 3.1 (more specifically, Equations 5 and 6), we need to show that $\sum_{i\in U_v}||\mathbf{x}_i||_2^2/N_v = \mathcal{O}(p_v)$ and that $||\widehat{\boldsymbol{\beta}}||_2 = \mathcal{O}_{\text{p}}(1)$. The same result holds for $\widehat{\boldsymbol{\beta}}_{\text{lasso}}$ and $\widehat{\boldsymbol{\beta}}_{\text{en}}$. We have $\sum_{i\in U_v}||\mathbf{x}_i||_2^2/N_v = \sum_{j=1}^{p_v}\sum_{i\in U_v} x_{ij}^2/N_v \leq p_v\sqrt{C_3} = \mathcal{O}(p_v)$ under the assumption of uniformly bounded forth moment of $X_j, j = 1,\ldots,p_v$.

We first show that, under the assumed orthogonality condition, $||\widehat{\boldsymbol{\beta}}_{\text{lasso}}||_2 \leq ||\widehat{\boldsymbol{\beta}}||_2$, $||\widehat{\boldsymbol{\beta}}_{\text{en}}||_2 \leq ||\widehat{\boldsymbol{\beta}}||_2$ and also $||\widehat{\boldsymbol{\beta}}||_2 = \mathcal{O}_{\text{p}}(1)$.

Consider again the objective function $\mathcal{L}_{ols}(\boldsymbol{\beta})$ as in the proof of Result 3.2. We can write

$$\mathcal{L}_{ols}(\boldsymbol{\beta}) \;=\; \sum_{i\in S_v}\frac{1}{\pi_i}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2 = \sum_{i\in S_v}(\tilde{y}_i - \tilde{\mathbf{x}}_i^\top\boldsymbol{\beta})^2 \tag{18}$$

where $\tilde{y}_i = y_i/\sqrt{\pi_i}$ and $\tilde{\mathbf{x}}_i = (\tilde{x}_{ij})_{j=1}^{p_v} = \mathbf{x}_i/\sqrt{\pi_i}$ for all $i \in S_v$. Let $\tilde{\mathbf{X}}_{S_v} = \mathbf{\Pi}_{S_v}^{-1/2}\mathbf{X}_{S_v} = (\tilde{\mathbf{x}}_i^\top)_{i\in S_v} = (\tilde{\mathbf{X}}_1,\ldots,\tilde{\mathbf{X}}_{p_v})$. The columns of $\tilde{\mathbf{X}}_{S_v}$, denoted by $\tilde{\mathbf{X}}_j, j = 1,\ldots,p_v$ are assumed to be orthogonal. This means that $\tilde{\mathbf{X}}_j^\top\tilde{\mathbf{X}}_k = 0$ for $j \neq k$. The ordinary least-square estimator $\widehat{\boldsymbol{\beta}}$ is given by

$$\widehat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{X}}_{S_v})^{-1}\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{y}}_{S_v}.$$

Under the orthogonality condition, $\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{X}}_{S_v}$ is a diagonal matrix with diagonal elements given by $||\tilde{\mathbf{X}}_j||_2^2 = \sum_{i\in S_v}\tilde{x}_{ij}^2 = \sum_{i\in S_v}\frac{x_{ij}^2}{\pi_i}$, which corresponds to the Horvitz-Thompson estimator of $\sum_{i\in U_v} x_{ij}^2$. Therefore, $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_j)_{j\in S_v}$ and the j-th coordinate is given by $\hat{\beta}_j = (\sum_{i\in S_v}\tilde{x}_{ij}^2)^{-1}\sum_{i\in S_v}\tilde{x}_{ij}\tilde{y}_i$.

The lasso estimator $\hat{\boldsymbol{\beta}}_{\text{lasso}} = (\hat{\beta}_{j,\text{lasso}})_{j=1}^{p_v}$ as well as the elastic-net estimator $\hat{\boldsymbol{\beta}}_{\text{en}} = (\hat{\beta}_{j,\text{en}})_{j=1}^{p_v}$ are obtained by using the cyclic soft-thresholding algorithm (Hastie et al., 2011):

$$\hat{\beta}_{j,\text{lasso}} = \frac{\mathcal{S}_\lambda(\sum_{i\in S_v} r_{ij}\tilde{x}_{ij})}{\sum_{i\in S_v}\tilde{x}_{ij}^2}$$

and

$$\hat{\beta}_{j,\text{en}} = \frac{\mathcal{S}_{\lambda\alpha}(\sum_{i\in S_v} r_{ij}\tilde{x}_{ij})}{\sum_{i=1}^{n_v}\tilde{x}_{ij}^2 + \lambda(1-\alpha)},$$

where $r_{ij} = \tilde{y}_i - \sum_{k\neq j}\tilde{x}_{ik}\hat{\beta}_k$ and $\mathcal{S}_\lambda(z) = sign(z)(|z|-\lambda)_+$ is the soft-thresholding function with $(|z|-\lambda)_+ = |z|-\lambda$ if $|z|\geq\lambda$, and zero otherwise. If the columns of $\tilde{\mathbf{X}}_{S_v}$ are orthogonal, then $\sum_{i\in S_v} r_{ij}\tilde{x}_{ij} = \sum_{i\in S_v}\tilde{x}_{ij}\tilde{y}_i$ and $\hat{\beta}_{j,\text{lasso}}$ is the soft-threshold estimator of the least-square estimator $\widehat{\beta}_j$:

$$\hat{\beta}_{j,\text{lasso}} = \frac{\mathcal{S}_\lambda(\sum_{i\in S_v}\tilde{x}_{ij}\tilde{y}_i)}{\sum_{i\in S_v}\tilde{x}_{ij}^2}.$$

The elastic-net estimator is given by

$$\hat{\beta}_{j,\text{en}} = \frac{\mathcal{S}_{\lambda\alpha}(\sum_{i\in S_v}\tilde{x}_{ij}\tilde{y}_i)}{\sum_{i\in S_v}\tilde{x}_{ij}^2 + \lambda(1-\alpha)}.$$

It follows that

$$|\hat{\beta}_{j,\text{lasso}}| = \frac{|(|\sum_{i\in S_v}\tilde{x}_{ij}\tilde{y}_i| - \lambda)_+|}{\sum_{i\in S_v}\tilde{x}_{ij}^2} \leq \frac{|\sum_{i\in S_v}\tilde{x}_{ij}\tilde{y}_i|}{\sum_{i\in S_v}\tilde{x}_{ij}^2} = |\hat{\beta}_j|, \quad j = 1,\ldots,p_v$$

and $||\widehat{\boldsymbol{\beta}}_{\text{lasso}}||_2 \leq ||\widehat{\boldsymbol{\beta}}||_2$. Similarly, $||\widehat{\boldsymbol{\beta}}_{\text{en}}||_2 \leq ||\widehat{\boldsymbol{\beta}}||_2$.

We now show that $||\widehat{\boldsymbol{\beta}}||_2 = \mathcal{O}_{\text{p}}(1)$. We have

$$||\widehat{\boldsymbol{\beta}}||_2 \leq ||N_v(\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{X}}_{S_v})^{-1}||_2 \left|\left|\frac{1}{N_v}\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{y}}_{S_v}\right|\right|_2.$$

The matrix $\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{X}}_{S_v}$ is diagonal with diagonal elements equal to $\sum_{i\in S_v}\frac{x_{ij}^2}{\pi_i}$. Then,

$$||N_v(\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{X}}_{S_v})^{-1}||_2 = \max_{j=1,\ldots,p_v}\left(\frac{1}{N_v^{-1}\sum_{i\in S_v}\frac{x_{ij}^2}{\pi_i}}\right)$$

and for all $j = 1,\ldots,p_v$:

$$\frac{1}{N_v^{-1}\sum_{i\in S_v}\frac{x_{ij}^2}{\pi_i}} = \frac{1}{N_v^{-1}\sum_{i\in U_v}x_{ij}^2} + \mathcal{O}_{\text{p}}\left(\frac{1}{\sqrt{n_v}}\right) = \mathcal{O}_{\text{p}}(1)$$

by using (H2), (H3) and the assumption of uniformly bounded fourth moment of $X_j, j = 1,\ldots,p_v$. We have also used the fact that $1/(N_v^{-1}\sum_{i\in U_v}x_{ij}^2) \leq$

$1/(\min_{j=1,\ldots,p_v} N_v^{-1}\sum_{i\in U_v} x_{ij}^2) \leq 1/C_4 = \mathcal{O}(1)$ for all $j = 1,\ldots,p_v$. Then,

$$||N_v(\tilde{\mathbf{X}}_{S_v}^\top \tilde{\mathbf{X}}_{S_v})^{-1}||_2 = \mathcal{O}_{\mathrm{p}}(1). \tag{19}$$

Now,

$$\left|\left|\frac{1}{N_v}\tilde{\mathbf{X}}_{S_v}^\top \tilde{\mathbf{y}}_{S_v}\right|\right|_2^2 \leq \frac{1}{N_v}||\tilde{\mathbf{y}}_{S_v}||_2^2 \left|\left|\frac{1}{N_v}\tilde{\mathbf{X}}_{S_v}\tilde{\mathbf{X}}_{S_v}^\top\right|\right|_2.$$

We have

$$\left|\left|\frac{1}{N_v}\tilde{\mathbf{X}}_{S_v}\tilde{\mathbf{X}}_{S_v}^\top\right|\right|_2 = \left|\left|\frac{1}{N_v}\tilde{\mathbf{X}}_{S_v}^\top\tilde{\mathbf{X}}_{S_v}\right|\right|_2 = \max_{j=1,\ldots,p_v}\left(\frac{1}{N_v}\sum_{i\in S_v}\frac{x_{ij}^2}{\pi_i}\right) \leq \max_{j=1,\ldots,p_v}\left(\frac{1}{N_v}\sum_{i\in U_v}x_{ij}^2\right) \leq \sqrt{C_3}$$

and

$$\frac{1}{N_v}||\tilde{\mathbf{y}}_{S_v}||_2^2 = \frac{1}{N_v}\sum_{i\in S_v}\frac{y_i^2}{\pi_i^2} \leq \frac{1}{c^2 N_v}\sum_{i\in U_v} y_i^2 \leq \frac{C_1}{c^2}$$

by Assumption (H1). So, $||\frac{1}{N_v}\tilde{\mathbf{X}}_{S_v}\tilde{\mathbf{y}}_{S_v}||_2 = \mathcal{O}(1)$ and combined with (19), we obtain $||\widehat{\boldsymbol{\beta}}||_2 = \mathcal{O}_{\mathrm{p}}(1)$.

■

References

Breidt, F.-J. and Opsomer, J.-D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28:1023–1053.

Hastie, T., Tibshirani, R., and Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.

Isaki, C.-T. and Fuller, W.-A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77:49–61.

Robinson, P. M. and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā Series B*, 45:240–248.