Supplement to "Estimating the Lasso's Effective Noise"

Johannes Lederer JOHANNES.LEDERER@RUB.DE
Department of Mathematics
Ruhr-University Bochum
44801 Bochum, Germany

Michael Vogt M.VOGT@UNI-ULM.DE
Institute of Statistics
Department of Mathematics and Economics
Ulm University
89081 Ulm, Germany

S.1. Technical Details

In what follows, we provide the technical details and proofs that are omitted in the paper.

Proof of Lemma 4

To show the result, we slightly generalize the proof of Lemma 5 in Chichignoud et al. (2016). Standard arguments from the lasso literature (Bühlmann and van de Geer, 2011) show that on the event $\mathcal{T}_\lambda$,

$$\|\hat{\beta}_{\lambda',S^\complement} - \beta^*_{S^\complement}\|_1 \leq \frac{2+\delta}{\delta}\|\hat{\beta}_{\lambda',S} - \beta^*_S\|_1,$$

that is, $\hat{\beta}_{\lambda'} - \beta^* \in \mathbb{C}_\delta(S)$ for every $\lambda' \geq (1+\delta)\lambda$. Under the ℓ_∞-restricted eigenvalue condition (13), we thus obtain that on $\mathcal{T}_\lambda$,

$$\phi\|\hat{\beta}_{\lambda'} - \beta^*\|_\infty \leq \frac{\|\boldsymbol{X}^\top\boldsymbol{X}(\hat{\beta}_{\lambda'} - \beta^*)\|_\infty}{n} \tag{S.1}$$

for every $\lambda' \geq (1+\delta)\lambda$. Moreover, since the lasso satisfies the zero-subgradient condition $2\boldsymbol{X}^\top(\boldsymbol{X}\hat{\beta}_{\lambda'} - Y)/n + \lambda'\hat{z} = 0$ with $\hat{z} \in \mathbb{R}^p$ belonging to the subdifferential of the function $f(\beta) = \|\beta\|_1$, it holds that

$$\frac{2\boldsymbol{X}^\top\boldsymbol{X}}{n}(\hat{\beta}_{\lambda'} - \beta^*) = -\lambda'\hat{z} + \frac{2\boldsymbol{X}^\top\varepsilon}{n}.$$

Taking the supremum norm on both sides of this equation and taking into account that $2\|\boldsymbol{X}^\top\varepsilon\|_\infty/n \leq \lambda$ on the event $\mathcal{T}_\lambda$, we obtain that

$$\frac{2\|\boldsymbol{X}^\top\boldsymbol{X}(\hat{\beta}_{\lambda'} - \beta^*)\|_\infty}{n} \leq \lambda' + \frac{2\|\boldsymbol{X}^\top\varepsilon\|_\infty}{n} \leq 2\lambda' \tag{S.2}$$

for every $\lambda' \geq (1+\delta)\lambda$ on $\mathcal{T}_\lambda$. The statement of Lemma 4 follows upon combining (S.1) and (S.2).

Proof of Lemma A.1

Let δ be a small positive constant with $0 < \delta < (\theta - 4)/\theta$ and $\theta > 4$ defined in (C3). Define $Z_{ijk} = X_{ij} X_{ik} \varepsilon_i^2$ along with $Z_{ijk} = Z_{ijk}^{\leq} + Z_{ijk}^{>}$, where

$$Z_{ijk}^{\leq} = Z_{ijk} \, \mathbf{1}\big(|\varepsilon_i| \leq n^{\frac{1-\delta}{4}}\big) \qquad \text{and} \qquad Z_{ijk}^{>} = Z_{ijk} \, \mathbf{1}\big(|\varepsilon_i| > n^{\frac{1-\delta}{4}}\big),$$

and write $\Delta \leq \Delta^{\leq} + \Delta^{>}$ with

$$\Delta^{\leq} = \max_{1 \leq j,k \leq p} \bigg| \frac{1}{n} \sum_{i=1}^n (Z_{ijk}^{\leq} - \mathbb{E} Z_{ijk}^{\leq}) \bigg|$$
$$\Delta^{>} = \max_{1 \leq j,k \leq p} \bigg| \frac{1}{n} \sum_{i=1}^n (Z_{ijk}^{>} - \mathbb{E} Z_{ijk}^{>}) \bigg|.$$

In what follows, we prove that

$$\mathbb{P}\big(\Delta^{\leq} > B\sqrt{\log(n \vee p)/n}\big) \leq C n^{-K} \tag{S.3}$$
$$\mathbb{P}\big(\Delta^{>} > B\sqrt{\log(n \vee p)/n}\big) \leq C n^{1 - (\frac{1-\delta}{4})\theta}, \tag{S.4}$$

where B, C and K are positive constants depending only on the parameters Θ, and K can be made as large as desired by choosing B and C large enough. Lemma A.1 is a direct consequence of the two statements (S.3) and (S.4).

We start with the proof of (S.3). A simple union bound yields that

$$\mathbb{P}\big(\Delta^{\leq} > B\sqrt{\log(n \vee p)/n}\big) \leq \sum_{j,k=1}^p P_{jk}^{\leq}, \tag{S.5}$$

where

$$P_{jk}^{\leq} = \mathbb{P}\bigg(\bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^n U_{ijk} \bigg| > B\sqrt{\log(n \vee p)} \bigg)$$

with $U_{ijk} = Z_{ijk}^{\leq} - \mathbb{E} Z_{ijk}^{\leq}$. Using Markov's inequality, $P_{jk}^{\leq}$ can be bounded by

$$\begin{aligned} P_{jk}^{\leq} &\leq \exp\big(- \mu B \sqrt{\log(n \vee p)}\big) \, \mathbb{E}\bigg[\exp\bigg(\mu \bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^n U_{ijk} \bigg| \bigg) \bigg] \\ &\leq \exp\big(- \mu B \sqrt{\log(n \vee p)}\big) \bigg\{ \mathbb{E}\bigg[\exp\bigg(\frac{\mu}{\sqrt{n}} \sum_{i=1}^n U_{ijk} \bigg) \bigg] \\ &\qquad\qquad + \mathbb{E}\bigg[\exp\bigg(-\frac{\mu}{\sqrt{n}} \sum_{i=1}^n U_{ijk} \bigg) \bigg] \bigg\} \end{aligned} \tag{S.6}$$

with an arbitrary constant $\mu > 0$. We now choose $\mu = \sqrt{\log(n \vee p)}/C_\mu$, where the constant $C_\mu > 0$ is picked so large that $\mu |U_{ijk}|/\sqrt{n} \leq 1/2$ for all n. With this choice of μ, we obtain that

$$\mathbb{E}\left[\exp\left(\pm\frac{\mu}{\sqrt{n}}\sum_{i=1}^{n}U_{ijk}\right)\right] = \prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\pm\frac{\mu}{\sqrt{n}}U_{ijk}\right)\right] \leq \prod_{i=1}^{n}\left(1+\frac{\mu^2}{n}\mathbb{E}[U_{ijk}^2]\right)$$
$$\leq \prod_{i=1}^{n}\exp\left(\frac{\mu^2}{n}\mathbb{E}[U_{ijk}^2]\right) \leq \exp(C_U\mu^2),$$

where the first inequality follows from the fact that $\exp(x) \leq 1+x+x^2$ for $|x| \leq 1/2$ and $C_U < \infty$ is an upper bound on $\mathbb{E}[U_{ijk}^2]$. Plugging this into (S.6) gives

$$P_{jk}^{\leq} \leq 2\exp\big(-\mu B\sqrt{\log(n\vee p)} + C_U\mu^2\big)$$
$$\leq 2\exp\left(-\left\{\frac{B}{C_\mu} - \frac{C_U}{C_\mu^2}\right\}\log(n\vee p)\right) = 2(n\vee p)^{\frac{C_U}{C_\mu^2}-\frac{B}{C_\mu}}.$$

Inserting this bound into (S.5), we finally obtain that

$$\mathbb{P}\big(\Delta^{\leq} > B\sqrt{\log(n\vee p)/n}\big) \leq 2p^2(n\vee p)^{\frac{C_U}{C_\mu^2}-\frac{B}{C_\mu}} \leq Cn^{-K},$$

where $K > 0$ can be chosen as large as desired by picking B sufficiently large. This completes the proof of (S.3).

We next turn to the proof of (S.4). It holds that

$$\mathbb{P}\big(\Delta^{>} > B\sqrt{\log(n\vee p)/n}\big) \leq P_1^{>} + P_2^{>},$$

where

$$\begin{aligned} P_1^{>} &:= \mathbb{P}\left(\max_{1\leq j,k\leq p}\left|\frac{1}{n}\sum_{i=1}^{n}Z_{ijk}^{>}\right| > \frac{B}{2}\sqrt{\frac{\log(n\vee p)}{n}}\right) \\ &\leq \mathbb{P}\big(|\varepsilon_i| > n^{\frac{1-\delta}{4}} \text{ for some } 1\leq i\leq n\big) \\ &\leq \sum_{i=1}^{n}\mathbb{P}\big(|\varepsilon_i| > n^{\frac{1-\delta}{4}}\big) \leq \sum_{i=1}^{n}\mathbb{E}\big[|\varepsilon_i|^\theta\big]/n^{(\frac{1-\delta}{4})\theta} \\ &\leq C_\theta n^{1-(\frac{1-\delta}{4})\theta} \end{aligned} \tag{S.7}$$

and

$$P_2^{>} := \mathbb{P}\left(\max_{1\leq j,k\leq p}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}Z_{ijk}^{>}\right| > \frac{B}{2}\sqrt{\frac{\log(n\vee p)}{n}}\right) = 0 \tag{S.8}$$

for sufficiently large n, since

$$\begin{aligned} \max_{1\leq j,k\leq p}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}Z_{ijk}^{>}\right| &\leq C_X^2\max_{1\leq i\leq n}\mathbb{E}\Big[\varepsilon_i^2\,\mathbf{1}(|\varepsilon_i| > n^{\frac{1-\delta}{4}})\Big] \\ &\leq C_X^2\max_{1\leq i\leq n}\mathbb{E}\Big[|\varepsilon_i|^\theta/n^{\frac{(\theta-2)(1-\delta)}{4}}\Big] \\ &\leq C_X^2 C_\theta n^{-\frac{(\theta-2)(1-\delta)}{4}} = o\left(\sqrt{\frac{\log(n\vee p)}{n}}\right). \end{aligned}$$

(S.4) follows upon combining (S.7) and (S.8).

Proof of Lemma A.2

Suppose we are on the event $\mathcal{S}_\gamma$ and let $\gamma' \geq \gamma$. In the case that $\beta^* = 0$, it holds that $\hat{\beta}_{2\gamma'/\sqrt{n}} = 0$ for all $\gamma' \geq \gamma$, implying that $R(\gamma', e) = 0$. Hence, Lemma A.2 trivially holds true if $\beta^* = 0$. We can thus restrict attention to the case that $\beta^* \neq 0$. Define $a_n = B(\log n)^2 \sqrt{\|\beta^*\|_1}$ with some $B > 0$ and write $e_i = e_i^{\leq} + e_i^{>}$ with

$$
\begin{aligned}
e_i^{\leq} &= e_i \, \mathbf{1}(|e_i| \leq \log n) - \mathbb{E}[e_i \, \mathbf{1}(|e_i| \leq \log n)] \\
e_i^{>} &= e_i \, \mathbf{1}(|e_i| > \log n) - \mathbb{E}[e_i \, \mathbf{1}(|e_i| > \log n)].
\end{aligned}
$$

With this notation, we get that

$$
\begin{aligned}
&\mathbb{P}_e\bigg(R(\gamma', e) > \frac{a_n \sqrt{\gamma'}}{n^{1/4}}\bigg) \\
&= \mathbb{P}_e\bigg(\max_{1 \leq j \leq p} \Big|\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij} X_i^\top (\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'}) e_i\Big| > \frac{a_n \sqrt{\gamma'}}{n^{1/4}}\bigg) \\
&\leq \sum_{j=1}^p \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij} X_i^\top (\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'}) e_i\Big| > \frac{a_n \sqrt{\gamma'}}{n^{1/4}}\bigg) \\
&\leq \sum_{j=1}^p \big\{P_{e,j}^{\leq} + P_{e,j}^{>}\big\},
\end{aligned}
\tag{S.9}
$$

where

$$
\begin{aligned}
P_{e,j}^{\leq} &= \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij} X_i^\top (\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'}) e_i^{\leq}\Big| > \frac{a_n \sqrt{\gamma'}}{2n^{1/4}}\bigg) \\
P_{e,j}^{>} &= \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij} X_i^\top (\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'}) e_i^{>}\Big| > \frac{a_n \sqrt{\gamma'}}{2n^{1/4}}\bigg).
\end{aligned}
$$

In what follows, we prove that for every $j \in \{1, \ldots, p\}$,

$$
P_{e,j}^{\leq} \leq C n^{-K} \qquad \text{and} \qquad P_{e,j}^{>} \leq C n^{-K}, \tag{S.10}
$$

where the constants C and K depend only on the parameters Θ, and K can be chosen as large as desired by picking C large enough. Plugging this into (S.9) immediately yields the statement of Lemma A.2.

We first show that $P_{e,j}^{\leq} \leq C n^{-K}$. To do so, we make use of the prediction bound (2) which implies that

$$
\frac{1}{n} \sum_{i=1}^n \big\{X_i^\top (\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})\big\}^2 \leq \frac{4\gamma' \|\beta^*\|_1}{\sqrt{n}} \tag{S.11}
$$

for any $\gamma' \geq \gamma$ on the event $\mathcal{S}_\gamma$. From this, it immediately follows that on $\mathcal{S}_\gamma$,

$$
\bigg|\frac{X_i^\top (\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})}{\sqrt{n}}\bigg| \leq \frac{2\sqrt{\gamma' \|\beta^*\|_1}}{n^{1/4}} \tag{S.12}
$$

for all i. Using Markov's inequality, $P_{e,j}^{\leq}$ can be bounded by

$$\begin{aligned}
P_{e,j}^{\leq} &= \mathbb{P}_e\bigg(\Big|\frac{1}{n^{1/4}}\sum_{i=1}^{n} X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})e_i^{\leq}\Big| > \frac{a_n\sqrt{\gamma'}}{2}\bigg) \\
&\leq \mathbb{E}_e \exp\bigg(\mu\Big|\frac{1}{n^{1/4}}\sum_{i=1}^{n} X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})e_i^{\leq}\Big|\bigg)\Big/\exp\bigg(\frac{\mu a_n\sqrt{\gamma'}}{2}\bigg) \\
&\leq \mathbb{E}_e \exp\bigg(\frac{\mu}{n^{1/4}}\sum_{i=1}^{n} X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})e_i^{\leq}\bigg)\Big/\exp\bigg(\frac{\mu a_n\sqrt{\gamma'}}{2}\bigg) \\
&\quad + \mathbb{E}_e \exp\bigg(-\frac{\mu}{n^{1/4}}\sum_{i=1}^{n} X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})e_i^{\leq}\bigg)\Big/\exp\bigg(\frac{\mu a_n\sqrt{\gamma'}}{2}\bigg)
\end{aligned} \tag{S.13}$$

with any $\mu > 0$. We make use of this bound with the particular choice $\mu = (4C_X\sqrt{\gamma'\|\beta^*\|_1}\log n)^{-1}$. Since $|\mu X_{ij}X_i^\top(\beta^* - \hat{\beta}_{2\gamma'/\sqrt{n}})e_i^{\leq}/n^{1/4}| \leq 1/2$ by condition (C2) and (S.12) and since $\exp(x) \leq 1 + x + x^2$ for any $|x| \leq 1/2$, we obtain that

$$\begin{aligned}
&\mathbb{E}_e \exp\bigg(\pm\frac{\mu}{n^{1/4}}\sum_{i=1}^{n} X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})e_i^{\leq}\bigg) \\
&= \prod_{i=1}^{n}\mathbb{E}_e \exp\bigg(\pm\mu X_{ij}\,\frac{X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})}{n^{1/4}}\,e_i^{\leq}\bigg) \\
&\leq \prod_{i=1}^{n}\bigg\{1 + \mu^2 X_{ij}^2\bigg(\frac{X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})}{n^{1/4}}\bigg)^2\mathbb{E}(e_i^{\leq})^2\bigg\} \\
&\leq \prod_{i=1}^{n}\exp\bigg(\mu^2 X_{ij}^2\bigg(\frac{X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})}{n^{1/4}}\bigg)^2\mathbb{E}(e_i^{\leq})^2\bigg) \\
&\leq \exp\bigg(\frac{c\mu^2}{\sqrt{n}}\sum_{i=1}^{n}\big\{X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})\big\}^2\bigg)
\end{aligned} \tag{S.14}$$

with a sufficiently large $c > 0$. Plugging (S.14) into (S.13) and using (S.11) along with the definition of μ, we arrive at

$$\begin{aligned}
P_{e,j}^{\leq} &\leq 2\exp\bigg(\frac{c\mu^2}{\sqrt{n}}\sum_{i=1}^{n}\big\{X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})\big\}^2 - \frac{\mu a_n\sqrt{\gamma'}}{2}\bigg) \\
&\leq 2\exp\bigg(4c\mu^2\gamma'\|\beta^*\|_1 - \frac{\mu a_n\sqrt{\gamma'}}{2}\bigg) \\
&\leq 2\exp\bigg(\frac{c}{4C_X^2(\log n)^2} - \frac{B\log n}{8C_X}\bigg) \leq Cn^{-K},
\end{aligned}$$

where K can be chosen as large as desired by picking C large enough.

We next verify that $P^{>}_{e,j} \le Cn^{-K}$. The term $P^{>}_{e,j}$ can be bounded by $P^{>}_{e,j} \le P^{>}_{e,j,1} + P^{>}_{e,j,2}$, where

$$P^{>}_{e,j,1} = \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})e_i\,\mathbf{1}(|e_i| > \log n)\Big| > \frac{a_n\sqrt{\gamma'}}{4n^{1/4}}\bigg)$$

$$P^{>}_{e,j,2} = \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{ij}X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})\mathbb{E}[e_i\,\mathbf{1}(|e_i| > \log n)]\Big| > \frac{a_n\sqrt{\gamma'}}{4n^{1/4}}\bigg).$$

Since the variables e_i are standard normal, it holds that

$$\begin{aligned} P^{>}_{e,j,1} &\le \mathbb{P}\big(|e_i| > \log n \text{ for some } 1 \le i \le n\big) \\ &\le \sum_{i=1}^n \mathbb{P}\big(|e_i| > \log n\big) \le \frac{2n}{\sqrt{2\pi}\log n}\exp\bigg(-\frac{(\log n)^2}{2}\bigg) \le Cn^{-K} \end{aligned} \tag{S.15}$$

for any $n > 1$, where $K > 0$ can be chosen as large as desired. Moreover, with the help of condition (C2) and (S.12), we get that

$$\begin{aligned} P^{>}_{e,j,2} &\le \mathbb{P}_e\bigg(\sum_{i=1}^n |X_{ij}|\bigg|\frac{X_i^\top(\beta^* - \hat{\beta}_{\frac{2}{\sqrt{n}}\gamma'})}{\sqrt{n}}\bigg|\mathbb{E}[|e_i|\,\mathbf{1}(|e_i| > \log n)] > \frac{a_n\sqrt{\gamma'}}{4n^{1/4}}\bigg) \\ &\le \mathbb{P}_e\bigg(C_X\frac{2\sqrt{\gamma'}\|\beta^*\|_1}{n^{1/4}}\sum_{i=1}^n \mathbb{E}[|e_i|\,\mathbf{1}(|e_i| > \log n)] > \frac{a_n\sqrt{\gamma'}}{4n^{1/4}}\bigg) \\ &\le \mathbb{P}_e\bigg(\sum_{i=1}^n \mathbb{E}[|e_i|\,\mathbf{1}(|e_i| > \log n)] > \frac{B(\log n)^2}{8C_X}\bigg) = 0 \end{aligned} \tag{S.16}$$

for n large enough, where the last equality follows from the fact that for any $c > 1$,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[|e_i|\,\mathbf{1}(|e_i| > \log n)] &\le \sum_{i=1}^n \mathbb{E}\bigg[\frac{|e_i|\exp(c|e_i|)}{\exp(c\log n)}\mathbf{1}(|e_i| > \log n)\bigg] \\ &\le \frac{n\mathbb{E}[|e_i|\exp(c|e_i|)]}{\exp(c\log n)} = o(1). \end{aligned}$$

Combining (S.15) and (S.16), we can conclude that $P^{>}_{e,j} \le Cn^{-K}$, where K can be picked as large as desired.

Proof of Lemma A.3

The proof is based on standard concentration and maximal inequalities. According to the Gaussian concentration inequality stated in Theorem 7.1 of Ledoux (2001) (see also Lemma 7 in Chernozhukov et al. (2015)), it holds that

$$\mathbb{P}\Big(\max_{1\le j\le p}\big|G_j/\sigma_j\big| \ge \mathbb{E}\Big[\max_{1\le j\le p}\big|G_j/\sigma_j\big|\Big] + \sqrt{2\log(n\vee p)}\Big) \le \frac{1}{n\vee p}, \tag{S.17}$$

where we use the notation $\sigma_j^2 = \mathbb{E}[G_j^2]$. Combining (S.17) with the maximal inequality $\mathbb{E}[\max_{1\le j\le p}|G_j/\sigma_j|] \le \sqrt{2\log(2p)}$ (see e.g. Proposition 1.1.3 in Talagrand (2003)) and multiplying each term inside the probability of (S.17) with $C_G = C_X C_\sigma$ yields

$$\mathbb{P}\Big(C_G \max_{1\le j\le p} \big|G_j/\sigma_j\big| \ge C_G\big[\sqrt{2\log(2p)} + \sqrt{2\log(n\vee p)}\big]\Big) \le \frac{1}{n\vee p}. \tag{S.18}$$

Since $\sigma_j \le C_G$ for any j, it holds that $C_G \max_{1\le j\le p}|G_j/\sigma_j| \ge \max_{1\le j\le p}|G_j|$. Plugging this into (S.18), we arrive at

$$\mathbb{P}\Big(\max_{1\le j\le p}|G_j| \ge C_G\big[\sqrt{2\log(2p)} + \sqrt{2\log(n\vee p)}\big]\Big) \le \frac{1}{n\vee p},$$

which implies that $\gamma_\alpha^G \le C_G[\sqrt{2\log(2p)} + \sqrt{2\log(n\vee p)}]$ for any $\alpha > 1/(n\vee p)$.

Proof of Lemma A.7

The proof is by contradiction. Suppose that $\mathbb{P}(\max_{1\le j\le p} V_j \le \gamma_\alpha^V) > 1-\alpha$, in particular, $\mathbb{P}(\max_{1\le j\le p} V_j \le \gamma_\alpha^V) = 1-\alpha+\eta$ with some $\eta > 0$. By Lemma A.4,

$$\sup_{t\in\mathbb{R}} \mathbb{P}\bigg(\Big|\max_{1\le j\le p} V_j - t\Big| \le \delta\bigg) \le b(\delta) := C\delta\sqrt{1\vee\log(p/\delta)}$$

for any $\delta > 0$, which implies that

$$\begin{aligned}
\mathbb{P}\Big(\max_{1\le j\le p} V_j \le \gamma_\alpha^V - \delta\Big) &= \mathbb{P}\Big(\max_{1\le j\le p} V_j \le \gamma_\alpha^V\Big) - \mathbb{P}\Big(\gamma_\alpha^V - \delta < \max_{1\le j\le p} V_j \le \gamma_\alpha^V\Big) \\
&\ge \mathbb{P}\Big(\max_{1\le j\le p} V_j \le \gamma_\alpha^V\Big) - \sup_{t\in\mathbb{R}} \mathbb{P}\bigg(\Big|\max_{1\le j\le p} V_j - t\Big| \le \delta\bigg) \\
&\ge 1-\alpha+\eta-b(\delta).
\end{aligned}$$

Since $b(\delta)\to 0$ as $\delta\to 0$, we can find a specific $\delta > 0$ with $b(\delta) < \eta$. For this specific δ, we get that $\mathbb{P}(\max_{1\le j\le p} V_j \le \gamma_\alpha^V - \delta) > 1-\alpha$, which contradicts the definition of the quantile γ_α^V according to which $\gamma_\alpha^V = \inf\{q : \mathbb{P}(\max_{1\le j\le p} V_j \le q) \ge 1-\alpha\}$.

Proof of Proposition A.11

We first have a closer look at the statistic $\Pi_B^* := \|(\mathcal{P}\boldsymbol{X}_B)^\top u\|_\infty/\sqrt{n}$. Without loss of generality, we let $A = \{1,\dots,p_A\}$ and $B = \{p_A+1,\dots,p_A+p_B\}$ with $p_A+p_B = p$, and we write $X_{i,A} = (X_{i1},\dots,X_{ip_A})^\top$ to shorten notation. Moreover, we define $\hat\psi_{jk} = n^{-1}\sum_{i=1}^n X_{ij}X_{ik}$ and set $\hat\psi_{j,A} = (\hat\psi_{j1},\dots,\hat\psi_{jp_A})^\top \in \mathbb{R}^{p_A}$ along with $\hat\Psi_A = (\hat\psi_{jk} : 1 \le j,k \le p_A) \in \mathbb{R}^{p_A\times p_A}$. Similarly, we let $\psi_{jk} = \mathbb{E}[X_{ij}X_{ik}]$, $\psi_{j,A} = (\psi_{j1},\dots,\psi_{jp_A})^\top$ and $\Psi_A = (\psi_{jk} : 1 \le j,k \le p_A)$. With this notation, the statistic $\Pi_B^* = \|(\mathcal{P}\boldsymbol{X}_B)^\top u\|_\infty/\sqrt{n} = \|(\mathcal{P}\boldsymbol{X}_B)^\top \mathcal{P}\varepsilon\|_\infty/\sqrt{n} = \|(\mathcal{P}\boldsymbol{X}_B)^\top \varepsilon\|_\infty/\sqrt{n}$ can be rewritten as $\Pi_B^* = \max_{j\in B}|W_{j,B}^*|$, where

$$W_{j,B}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat Z_{ij}\varepsilon_i \quad \text{with} \quad \hat Z_{ij} = X_{ij} - \hat\psi_{j,A}^\top \hat\Psi_A^{-1} X_{i,A},$$

and $W_B^* = (W_{j,B}^* : j \in B)$ is the vector with the elements $W_{j,B}^*$. In contrast to X_i, the random vectors $\hat Z_i = (\hat Z_{ij} : j \in B)$ are not independent across i in general. In order to deal with this complication, we introduce the auxiliary statistic $\Pi_B^{**} = \max_{j \in B} |W_{j,B}^{**}|$, where $W_B^{**} = (W_{j,B}^{**} : j \in B)$ and

$$W_{j,B}^{**} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij} \varepsilon_i \quad \text{with} \quad Z_{ij} = X_{ij} - \psi_{j,A}^{\top} \Psi_A^{-1} X_{i,A}.$$

The random vectors $Z_i = (Z_{ij} : j \in B)$ have the following properties: (i) Unlike $\hat{Z}_i$, they are independent across i. (ii) Since $|X_{ij}| \le C_X$ by (C2) and Ψ_A is positive definite by assumption, $|Z_{ij}| \le C_Z < \infty$ with a constant C_Z that depends only on the model parameters Θ'. (iii) Since Z_{ij} can be expressed as $Z_{ij} = X_{ij} - X_{i,A}^{\top} \vartheta^{(j)}$ with $\vartheta^{(j)}$ introduced before the formulation of Proposition 7, it holds that $\mathbb{E}[Z_{ij}^2] \ge c_Z^2 > 0$ with $c_Z^2 = c_\vartheta$. We denote the $(1-\alpha)$-quantile of Π_B^{**} by $\gamma_{\alpha,B}^{**}$. In the course of the proof, we will establish that $\gamma_{\alpha,B}^{**}$ is close to the quantile $\gamma_{\alpha,B}^{*}$ of the statistic Π_B^{*} in a suitable sense.

In addition to the above quantities, we introduce some auxiliary statistics that parallel those defined in the proof of Theorem 1. To start with, let $\hat{\Pi}_B(\gamma, e) = \max_{j \in B} |\hat{W}_{j,B}(\gamma, e)|$, where $\hat{W}_B(\gamma, e) = (\hat{W}_{j,B}(\gamma, e) : j \in B)$ with

$$\hat{W}_{j,B}(\gamma, e) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{Z}_{ij} \hat{u}_{\frac{2}{\sqrt{n}}\gamma, i} e_i,$$

and let $\hat{\pi}_{\alpha,B}(\gamma)$ be the $(1-\alpha)$-quantile of $\hat{\Pi}_B(\gamma, e)$ conditionally on $\boldsymbol{X}$ and ε. With this notation, the estimator $\hat{\gamma}_{\alpha,B}$ can be expressed as

$$\hat{\gamma}_{\alpha,B} = \inf \big\{ \gamma > 0 : \hat{\pi}_{\alpha,B}(\gamma') \le \gamma' \text{ for all } \gamma' \ge \gamma \big\}.$$

Moreover, let $\Pi_B^G = \max_{j \in B} |G_j|$, where $G_B = (G_j : j \in B)$ is a Gaussian random vector with $\mathbb{E}[G_B] = \mathbb{E}[W_B^{**}] = 0$ and $\mathbb{E}[G_B G_B^{\top}] = \mathbb{E}[W_B^{**} (W_B^{**})^{\top}]$, and let $\gamma_{\alpha,B}^G$ denote the $(1-\alpha)$-quantile of Π_B^G. Finally, define the statistic $\Pi_B(e) = \max_{j \in B} |W_{j,B}(e)|$, where $W_B(e) = (W_{j,B}(e) : j \in B)$ with

$$W_{j,B}(e) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij} \varepsilon_i e_i,$$

and let $\gamma_{\alpha,B}$ be the $(1-\alpha)$-quantile of $\Pi_B(e)$ conditionally on $\boldsymbol{X}$ and ε.

We next define some expressions which play a similar role as the quantity Δ in the proof of Theorem 1. In particular, we let $\Delta_1 = \|n^{-1} \sum_{i=1}^n X_{i,A} \varepsilon_i\|_2$ along with

$$\Delta_2 = \max_{j \in A} \left| \frac{1}{n} \sum_{i=1}^{n} \big\{ X_{ij}^2 \varepsilon_i^2 - \mathbb{E}[X_{ij}^2 \varepsilon_i^2] \big\} \right|$$

$$\Delta_3 = \max_{1 \le j,k \le p} \left| \frac{1}{n} \sum_{i=1}^{n} \big\{ X_{ij} X_{ik} - \mathbb{E}[X_{ij} X_{ik}] \big\} \right|$$

$$\Delta_4 = \max_{j,k \in B} \left| \frac{1}{n} \sum_{i=1}^{n} \big\{ Z_{ij} Z_{ik} \varepsilon_i^2 - \mathbb{E}[Z_{ij} Z_{ik} \varepsilon_i^2] \big\} \right|.$$

Applying Markov's inequality, we obtain that

$$\mathbb{P}\big(\Delta_1 > n^{-\frac{1}{2}+\rho}\big) \le Cn^{-2\rho} \tag{S.19}$$
$$\mathbb{P}\big(\Delta_2 > n^{-\frac{1}{2}+\rho}\big) \le Cn^{-2\rho}, \tag{S.20}$$

where we choose ρ to be a fixed constant with $\rho \in (0,1/2)$ and C depends only on Θ'. Moreover, noticing that $|Z_{ij}| \le C_Z < \infty$ and $\mathbb{E}[Z_{ij}^2] \ge c_Z^2 > 0$ under the conditions of Proposition 7, the same arguments as for Lemma A.1 yield the following: there exist positive constants C, D and K depending only on Θ' such that

$$\mathbb{P}\big(\Delta_3 > D\sqrt{\log(n \vee p)/n}\big) \le Cn^{-K} \tag{S.21}$$
$$\mathbb{P}\big(\Delta_4 > D\sqrt{\log(n \vee p)/n}\big) \le Cn^{-K}. \tag{S.22}$$

Taken together, (S.19)–(S.22) imply that the event

$$\mathcal{A}_n' := \Big\{(\Delta_1 \vee \Delta_2) \le n^{-\frac{1}{2}+\rho} \text{ and } (\Delta_3 \vee \Delta_4) \le D\sqrt{\log(n \vee p)/n}\Big\}$$

occurs with probability at least $1 - O(n^{-K} \vee n^{-2\rho})$.

With the above notation at hand, we now turn to the proof of Proposition A.11. In a first step, we show that the quantiles of the statistic Π_B^* are close to those of the auxiliary statistic Π_B^{**} in the following sense: there exist positive constants C and K depending only on Θ' such that

$$\begin{aligned} \gamma^*_{\alpha+\zeta_n,B} \le \gamma^{**}_{\alpha,B} \le \gamma^*_{\alpha-\zeta_n,B} \\ \gamma^{**}_{\alpha+\zeta_n,B} \le \gamma^{*}_{\alpha,B} \le \gamma^{**}_{\alpha-\zeta_n,B} \end{aligned} \tag{S.23}$$

for any $\alpha \in (\zeta_n, 1-\zeta_n)$ with $\zeta_n = Cn^{-K}$. The proof of (S.23) is postponed until the arguments for Proposition A.11 are complete. In the second step, we relate the quantiles $\gamma^{**}_{\alpha,B}$ of Π_B^{**} to the quantiles $\gamma_{\alpha,B}$ of $\Pi_B(e)$. Arguments completely analogous to those for Proposition A.10 yield the following: there exist positive constants C and K depending only on Θ' such that on the event $\mathcal{A}_n'$,

$$\begin{aligned} \gamma_{\alpha+\xi_n',B} \le \gamma^{**}_{\alpha,B} \le \gamma_{\alpha-\xi_n',B} \\ \gamma^{**}_{\alpha+\xi_n',B} \le \gamma_{\alpha,B} \le \gamma^{**}_{\alpha-\xi_n',B} \end{aligned} \tag{S.24}$$

for any $\alpha \in (\xi_n', 1-\xi_n')$ with $\xi_n' = Cn^{-K}$. In the third step, we relate the auxiliary statistic $\Pi_B(e)$ to the criterion function $\hat{\Pi}_B(\gamma, e)$, which underlies the estimator $\hat{\gamma}_{\alpha,B}$. Straightforward calculations show that

$$\hat{\Pi}_B(\gamma,e) \begin{cases} \le \Pi_B(e) + R_B(\gamma,e) \\ \ge \Pi_B(e) - R_B(\gamma,e), \end{cases} \tag{S.25}$$

where $R_B(\gamma, e) = R_{B,1}(\gamma, e) + R_{B,2}(e) + R_{B,3}(e)$ with

$$R_{B,1}(\gamma, e) = \max_{j \in B} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_{ij} \big\{ \mathcal{P} \boldsymbol{X}_B \big(\beta_B^* - \hat{\beta}_{B,\frac{2}{\sqrt{n}}\gamma} \big) \big\}_i \, e_i \right|$$

$$R_{B,2}(e) = \max_{j \in B} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{Z}_{ij} - Z_{ij}) \varepsilon_i e_i \right|$$

$$R_{B,3}(e) = \max_{j \in B} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_{ij} (\varepsilon_i - u_i) e_i \right|.$$

The terms $R_{B,1}(\gamma, e)$, $R_{B,2}(e)$ and $R_{B,3}(e)$ have the following properties: on the event $S'_\gamma \cap \mathcal{A}'_n$,

$$\mathbb{P}_e\left(R_{B,1}(\gamma', e) > \frac{D (\log n)^2 \sqrt{\|\beta_B^*\|_1 \gamma'}}{n^{1/4}} \right) \le C n^{-K} \tag{S.26}$$

for every $\gamma' \ge \gamma$, where the constants C, D and K depend only on Θ'. Moreover, on the event $\mathcal{A}'_n$,

$$\mathbb{P}_e\left(R_{B,2}(e) > \frac{D \log^{1/2}(n \vee p)}{n^{1/2-\rho}} \right) \le C n^{-2\rho} \tag{S.27}$$

$$\mathbb{P}_e\left(R_{B,3}(e) > \frac{D \log(n \vee p)}{n^{1/2-\rho}} \right) \le C n^{-K}, \tag{S.28}$$

where $\rho \in (0, 1/2)$ has been introduced in (S.19)–(S.20) and the constants C, D and K depend only on Θ'. The proofs of (S.26)–(S.28) are provided below. With (S.23)–(S.28) in place, we can now use the same arguments as in the proof of Theorem 1 (with minor adjustments) to obtain that $\gamma^*_{\alpha+\nu'_n,B} \le \hat{\gamma}_{\alpha,B} \le \gamma^*_{\alpha-\nu'_n,B}$ on the event $\mathcal{S}'_{\gamma^*_{\alpha+\nu'_n,B}} \cap \mathcal{A}'_n$.

Proof of (S.23) We prove that

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\Pi_B^{**} \le t) - \mathbb{P}(\Pi_B^G \le t) \right| \le C n^{-K} \tag{S.29}$$

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\Pi_B^{*} \le t) - \mathbb{P}(\Pi_B^G \le t) \right| \le C n^{-K}, \tag{S.30}$$

where C and K depend only on Θ'. Applying the same arguments as in the proof of Proposition A.9 to the statements (S.29) and (S.30) yields that

$$\begin{aligned} \gamma^{**}_{\alpha+Cn^{-K},B} &\le \gamma^G_{\alpha,B} \le \gamma^{**}_{\alpha-Cn^{-K},B} \\ \gamma^{G}_{\alpha+Cn^{-K},B} &\le \gamma^{**}_{\alpha,B} \le \gamma^{G}_{\alpha-Cn^{-K},B} \end{aligned} \quad \text{and} \quad \begin{aligned} \gamma^{*}_{\alpha+Cn^{-K},B} &\le \gamma^G_{\alpha,B} \le \gamma^{*}_{\alpha-Cn^{-K},B} \\ \gamma^{G}_{\alpha+Cn^{-K},B} &\le \gamma^{*}_{\alpha,B} \le \gamma^{G}_{\alpha-Cn^{-K},B}, \end{aligned}$$

from which (S.23) follows immediately.

It remains to prove (S.29) and (S.30). (S.29) is a direct consequence of Lemma A.6, since $0 < c_\sigma^2 c_Z^2 \le n^{-1} \sum_{i=1}^n \mathbb{E}[(Z_{ij}\varepsilon_i)^2] \le C_\sigma^2 C_Z^2 < \infty$ and $\max_{k=1,2} \{ n^{-1} \sum_{i=1}^n \mathbb{E}[|Z_{ij}\varepsilon_i|^{2+k}/C^k] \} + \mathbb{E}[\{\max_{j \in B} |Z_{ij}\varepsilon_i|/C\}^4] \le 4$ for C large enough, where we have used (C3) and the fact that $|Z_{ij}| \le C_Z < \infty$ and $\mathbb{E}[Z_{ij}^2] \ge c_Z^2 > 0$ under the conditions of Proposition 7. For the proof of (S.30), it suffices to show that

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\Pi_B^{*} \le t) - \mathbb{P}(\Pi_B^{**} \le t) \right| \le C n^{-K} \tag{S.31}$$

with C and K depending only on Θ', since by (S.29),

$$\begin{aligned}\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^G\le t)\right| &\le \sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^{**}\le t)\right| \\ &\quad+\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^G\le t)\right| \\ &\le \sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^{**}\le t)\right|+Cn^{-K}.\end{aligned}$$

To prove (S.31), we fix a constant $d\in(0,1/2)$ and let $c_n=Dn^d\sqrt{\log(n\vee p)/n}$, where D is a sufficiently large constant that depends only on Θ'. In the case that $\mathbb{P}(\Pi_B^*\le t)\ge\mathbb{P}(\Pi_B^{**}\le t)$, the difference $\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^{**}\le t)$ can be bounded as follows:

$$\begin{aligned}&\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^{**}\le t)\\ &=\mathbb{P}(\Pi_B^{**}\le t+\Pi_B^{**}-\Pi_B^*,|\Pi_B^{**}-\Pi_B^*|\le c_n)\\ &\quad+\mathbb{P}(\Pi_B^{**}\le t+\Pi_B^{**}-\Pi_B^*,|\Pi_B^{**}-\Pi_B^*|>c_n)-\mathbb{P}(\Pi_B^{**}\le t)\\ &\le\mathbb{P}(\Pi_B^{**}\le t+c_n)-\mathbb{P}(\Pi_B^{**}\le t)+\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n)\\ &\le\left|\mathbb{P}(\Pi_B^{**}\le t+c_n)-\mathbb{P}(\Pi_B^G\le t+c_n)\right|+\left|\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^G\le t)\right|\\ &\quad+\left|\mathbb{P}(\Pi_B^G\le t+c_n)-\mathbb{P}(\Pi_B^G\le t)\right|+\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n)\\ &\le\left|\mathbb{P}(\Pi_B^{**}\le t+c_n)-\mathbb{P}(\Pi_B^G\le t+c_n)\right|+\left|\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^G\le t)\right|\\ &\quad+\mathbb{P}(|\Pi_B^G-t|\le c_n)+\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n).\end{aligned}\tag{S.32}$$

For the case that $\mathbb{P}(\Pi_B^*\le t)<\mathbb{P}(\Pi_B^{**}\le t)$, we similarly get that

$$\begin{aligned}&\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^*\le t)\\ &\le\left|\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^G\le t)\right|+\left|\mathbb{P}(\Pi_B^{**}\le t-c_n)-\mathbb{P}(\Pi_B^G\le t-c_n)\right|\\ &\quad+\mathbb{P}(|\Pi_B^G-t|\le c_n)+\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n).\end{aligned}\tag{S.33}$$

(S.32) and (S.33) immediately yield that

$$\begin{aligned}\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^{**}\le t)\right| &\le 2\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^G\le t)\right|\\ &\quad+\sup_{t\in\mathbb{R}}\mathbb{P}(|\Pi_B^G-t|\le c_n)\\ &\quad+\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n).\end{aligned}$$

Since $\sup_{t\in\mathbb{R}}|\mathbb{P}(\Pi_B^{**}\le t)-\mathbb{P}(\Pi_B^G\le t)|\le Cn^{-K}$ by (S.29) and $\sup_{t\in\mathbb{R}}\mathbb{P}(|\Pi_B^G-t|\le c_n)\le Cn^{-K}$ by Lemma A.4, we further get that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\Pi_B^*\le t)-\mathbb{P}(\Pi_B^{**}\le t)\right|\le\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n)+Cn^{-K},$$

where C and K depend only on Θ'. To complete the proof of (S.31), we thus need to show that

$$\mathbb{P}(|\Pi_B^{**}-\Pi_B^*|>c_n)\le Cn^{-K}\tag{S.34}$$

with C and K depending only on Θ'. To do so, we bound the term $|\Pi_B^{**} - \Pi_B^*|$ by

$$\begin{aligned}
|\Pi_B^{**} - \Pi_B^*| &\le \max_{j \in B} \bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{Z}_{ij} - Z_{ij}) \varepsilon_i \bigg| \\
&= \max_{j \in B} \bigg| \{\psi_{j,A}^\top \Psi_A^{-1} - \hat{\psi}_{j,A}^\top \hat{\Psi}_A^{-1}\} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,A} \varepsilon_i \bigg| \\
&\le \Big\{ \max_{j \in B} \|\psi_{j,A} - \hat{\psi}_{j,A}\|_2 \|\Psi_A^{-1}\|_2 \\
&\qquad + \max_{j \in B} \|\hat{\psi}_{j,A}\|_2 \|\Psi_A^{-1} - \hat{\Psi}_A^{-1}\|_2 \Big\} \bigg\| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,A} \varepsilon_i \bigg\|_2. \qquad (S.35)
\end{aligned}$$

From (S.21), it immediately follows that

$$\mathbb{P}\Big(\max_{j \in B} \|\psi_{j,A} - \hat{\psi}_{j,A}\|_2 > D\sqrt{\log(n \vee p)/n} \Big) \le Cn^{-K} \qquad (S.36)$$

$$\mathbb{P}\Big(\|\Psi_A - \hat{\Psi}_A\|_2 > D\sqrt{\log(n \vee p)/n} \Big) \le Cn^{-K} \qquad (S.37)$$

with C, D and K depending only on Θ'. Moreover, it holds that

$$\mathbb{P}\Big(\|\Psi_A^{-1} - \hat{\Psi}_A^{-1}\|_2 > D\sqrt{\log(n \vee p)/n} \Big) \le Cn^{-K}, \qquad (S.38)$$

which is a consequence of (S.37) and the fact that

$$\|Q^{-1} - R^{-1}\|_2 \le \frac{\|R^{-1}\|_2^2 \|R - Q\|_2}{1 - \|R - Q\|_2 \|R^{-1}\|_2} \qquad (S.39)$$

for every pair of invertible matrices Q and R that are close enough such that $\|R - Q\|_2 \|R^{-1}\|_2 < 1$. Finally, a simple application of Markov's inequality yields that

$$\mathbb{P}\bigg(\bigg\| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,A} \varepsilon_i \bigg\|_2 > n^d \bigg) \le Cn^{-2d}, \qquad (S.40)$$

where C depends only on Θ'. The statement (S.34) follows upon applying the results (S.36)–(S.38) and (S.40) to the bound (S.35). ■

Proof of (S.26) To start with, we bound $R_{B,1}(\gamma, e)$ by

$$\begin{aligned}
R_{B,1}(\gamma, e) &\le \Big\{ 1 + \sqrt{p_A} \max_{j \in B} \|\hat{\psi}_{j,A}^\top \hat{\Psi}_A^{-1}\|_2 \Big\} \\
&\qquad \times \max_{1 \le j \le p} \bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij} \big\{ \mathcal{P} \boldsymbol{X}_B \big(\beta_B^* - \hat{\beta}_{B, \frac{2}{\sqrt{n}}\gamma} \big) \big\}_i e_i \bigg|. \qquad (S.41)
\end{aligned}$$

The same arguments as in the proof of Lemma A.2 yield that on the event S'_γ,

$$\begin{aligned}
\mathbb{P}_e \bigg(\max_{1 \le j \le p} \bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij} \big\{ \mathcal{P} \boldsymbol{X}_B \big(\beta_B^* - \hat{\beta}_{B, \frac{2}{\sqrt{n}}\gamma'} \big) \big\}_i e_i \bigg| \\
> \frac{D (\log n)^2 \sqrt{\|\beta_B^*\|_1 \gamma'}}{n^{1/4}} \bigg) \le Cn^{-K} \qquad (S.42)
\end{aligned}$$

for every $\gamma' \geq \gamma$, where C, D and K depend only on Θ'. Moreover, on the event $\mathcal{A}_n'$,

$$\max_{j \in B} \|\hat{\psi}_{j,A} - \psi_{j,A}\|_2 \leq C\sqrt{\log(n \vee p)/n} \tag{S.43}$$
$$\|\hat{\Psi}_A - \Psi_A\|_2 \leq C\sqrt{\log(n \vee p)/n} \tag{S.44}$$
$$\|\hat{\Psi}_A^{-1} - \Psi_A^{-1}\|_2 \leq C\sqrt{\log(n \vee p)/n}, \tag{S.45}$$

where C is a sufficiently large constant that depends only on Θ', and (S.45) is a simple consequence of (S.44) and (S.39). To complete the proof, we apply (S.42)–(S.45) to the bound (S.41), taking into account that $\|\Psi_A^{-1}\|_2 \leq C < \infty$ and $\max_{j \in B} \|\psi_{j,A}\|_2 \leq C < \infty$. ■

Proof of (S.27) We have the bound

$$R_{B,2}(e) \leq \max_{j \in B} \big\|\hat{\psi}_{j,A}^\top \hat{\Psi}_A^{-1} - \psi_{j,A}^\top \Psi_A^{-1}\big\|_2 \Big\|\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,A}\varepsilon_i e_i\Big\|_2. \tag{S.46}$$

On the event $\mathcal{A}_n'$,

$$\begin{aligned}\mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{ij}\varepsilon_i e_i\Big| > n^\rho\bigg) &\leq n^{-2\rho}\Big\{\frac{1}{n}\sum_{i=1}^n X_{ij}^2\varepsilon_i^2\Big\} \leq n^{-2\rho}\big\{\mathbb{E}[X_{ij}^2\varepsilon_i^2] + \Delta_2\big\} \\ &\leq n^{-2\rho}\big\{C_X^2 C_\sigma^2 + n^{-\frac{1}{2}+\rho}\big\}\end{aligned}$$

for every $j \in A$, which implies that $\mathbb{P}_e(\|n^{-1/2}\sum_{i=1}^n X_{i,A}\varepsilon_i e_i\|_2 > n^\rho) \leq Cn^{-2\rho}$ with C depending only on Θ'. To complete the proof, we apply this, (S.43)–(S.45) and the fact that $\|\Psi_A^{-1}\|_2 \leq C < \infty$ and $\max_{j \in B} \|\psi_{j,A}\|_2 \leq C < \infty$ to the bound (S.46). ■

Proof of (S.28) Let $d_n = D\log(n \vee p)/n^{1/2-\rho}$ and define

$$\begin{aligned} e_i^{\leq} &= e_i\,\mathbf{1}(|e_i| \leq \log n) - \mathbb{E}[e_i\,\mathbf{1}(|e_i| \leq \log n)] \\ e_i^{>} &= e_i\,\mathbf{1}(|e_i| > \log n) - \mathbb{E}[e_i\,\mathbf{1}(|e_i| > \log n)]. \end{aligned}$$

It holds that

$$\begin{aligned}\mathbb{P}_e\big(R_{B,3}(e) > d_n\big) &\leq \sum_{j \in B} \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i\Big| > d_n\bigg) \\ &\leq \sum_{j \in B}\big\{P_{e,j}^{\leq} + P_{e,j}^{>}\big\}, \end{aligned} \tag{S.47}$$

where

$$\begin{aligned} P_{e,j}^{\leq} &= \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}\Big| > \frac{d_n}{2}\bigg) \\ P_{e,j}^{>} &= \mathbb{P}_e\bigg(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{>}\Big| > \frac{d_n}{2}\bigg). \end{aligned}$$

We first analyze the term $P_{e,j}^{\leq}$. With the help of (S.43)–(S.45), we obtain that on the event $\mathcal{A}_n'$,

$$
\begin{aligned}
|\hat{Z}_{ij}| &= |X_{ij} - \hat{\psi}_{j,A}^\top \hat{\Psi}_A^{-1} X_{i,A}| \leq \{1 + \|\hat{\psi}_{j,A}\|_2 \|\hat{\Psi}_A^{-1}\|_2 \sqrt{p_A}\} C_X \leq C \\
|\varepsilon_i - u_i| &= \big|\{\boldsymbol{X}_A(\boldsymbol{X}_A^\top \boldsymbol{X}_A)^{-1} \boldsymbol{X}_A^\top \varepsilon\}_i\big| = \bigg|\boldsymbol{X}_{i,A}^\top \hat{\Psi}_A^{-1} \Big\{\frac{1}{n}\sum_{\ell=1}^n \boldsymbol{X}_{\ell,A}\varepsilon_\ell\Big\}\bigg| \\
&\leq \sqrt{p_A} C_X \|\hat{\Psi}_A^{-1}\|_2 \Big\|\frac{1}{n}\sum_{\ell=1}^n \boldsymbol{X}_{\ell,A}\varepsilon_\ell\Big\|_2 \leq \frac{C}{n^{1/2-\rho}},
\end{aligned}
$$

which implies that $|\hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}| \leq C \log n / n^{1/2-\rho}$. Using Markov's inequality, $P_{e,j}^{\leq}$ can be bounded by

$$
\begin{aligned}
P_{e,j}^{\leq} &\leq \mathbb{E}_e \exp\bigg(\mu\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}\Big|\bigg) \Big/ \exp\Big(\frac{\mu d_n}{2}\Big) \\
&\leq \bigg\{\mathbb{E}_e \exp\bigg(\frac{\mu}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}\bigg) \\
&\qquad + \mathbb{E}_e \exp\bigg(-\frac{\mu}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}\bigg)\bigg\} \Big/ \exp\Big(\frac{\mu d_n}{2}\Big),
\end{aligned} \tag{S.48}
$$

where we choose $\mu = c_\mu n^{1/2-\rho}$ with $c_\mu > 0$ so small that $\mu|\hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}|/\sqrt{n} \leq 1/2$. Since $\exp(x) \leq 1 + x + x^2$ for $|x| \leq 1/2$, we further get that

$$
\begin{aligned}
\mathbb{E}_e \exp\bigg(\pm\frac{\mu}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}\bigg) &= \prod_{i=1}^n \mathbb{E}_e \exp\bigg(\pm\frac{\mu \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{\leq}}{\sqrt{n}}\bigg) \\
&\leq \prod_{i=1}^n \bigg\{1 + \frac{\mu^2 \hat{Z}_{ij}^2(\varepsilon_i - u_i)^2 \mathbb{E}(e_i^{\leq})^2}{n}\bigg\} \\
&\leq \prod_{i=1}^n \exp\bigg(\frac{\mu^2 \hat{Z}_{ij}^2(\varepsilon_i - u_i)^2 \mathbb{E}(e_i^{\leq})^2}{n}\bigg) \\
&= \exp\bigg(\frac{\mu^2}{n}\sum_{i=1}^n \hat{Z}_{ij}^2(\varepsilon_i - u_i)^2 \mathbb{E}(e_i^{\leq})^2\bigg) \leq \exp(c)
\end{aligned}
$$

with a sufficiently large constant c that depends only on Θ'. Plugging this into (S.48) yields that

$$
P_{e,j}^{\leq} \leq 2\exp\Big(c - \frac{c_\mu D \log(n \vee p)}{2}\Big) \leq C n^{-K}, \tag{S.49}
$$

where K can be made as large as desired.

We next have a closer look at the term $P_{e,j}^{>}$. Since $\max_{j \in B}|\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)| = \|(\mathcal{P}\boldsymbol{X}_B)^\top(\varepsilon - \mathcal{P}\varepsilon)\|_\infty = 0$, it holds that

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i^{>} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_{ij}(\varepsilon_i - u_i)e_i\,\mathbf{1}(|e_i| > \log n),
$$

and thus, as already proven in (S.15),

$$P_{e,j}^{>} \leq \mathbb{P}_e\big(|e_i| > \log n \text{ for some } 1 \leq i \leq n\big) \leq Cn^{-K}, \tag{S.50}$$

where K can be made as large as desired. To complete the proof, we insert equations (S.49) and (S.50) into (S.47) and invoke condition (C4). ■

Proof of Equation (10)

Let $G_1, \ldots, G_p$ be independent normal random variables with $\mathbb{E}[G_j] = 0$ for all j and suppose w.l.o.g. that $\mathbb{E}[G_j^2] = 1$ for all j. By Lemma A.7,

$$\mathbb{P}\Big(\max_{1\leq j\leq p} |G_j| \leq \gamma_\alpha^G\Big) = 1 - \alpha. \tag{S.51}$$

Moreover, standard arguments from classic extreme value theory show that

$$\mathbb{P}\Big(\max_{1\leq j\leq p} |G_j| \leq \frac{x}{a_p} + b_p\Big) \to e^{-2e^{-x}}$$

as $p \to \infty$ with $a_p = \sqrt{2\log p}$ and $b_p = \sqrt{2\log p} - \{\log\log p + \log(4\pi)\}/\{2\sqrt{2\log p}\}$, which in particular implies that for any fixed $\delta > 0$,

$$\mathbb{P}\Big(\max_{1\leq j\leq p} |G_j| \leq \frac{x_{\alpha\pm\delta}}{a_p} + b_p\Big) \to 1 - \{\alpha \pm \delta\} \tag{S.52}$$

with $x_{\alpha\pm\delta} = -\log(-\log(1 - \{\alpha \pm \delta\})/2)$. From (S.51) and (S.52), it follows that for any null sequence of positive numbers η_p,

$$\frac{x_{\alpha+\delta}}{a_p} + b_p \leq \gamma_{\alpha+\eta_p}^G \leq \gamma_{\alpha-\eta_p}^G \leq \frac{x_{\alpha-\delta}}{a_p} + b_p$$

for p sufficiently large. We thus arrive at

$$|\gamma_{\alpha-\eta_p}^G - \gamma_{\alpha+\eta_p}^G| \leq \frac{x_{\alpha-\delta} - x_{\alpha+\delta}}{\sqrt{2\log p}} \leq \frac{C}{\sqrt{2\log p}}$$

with some sufficiently large constant C.

S.2. Robustness Checks

Choice of α for Tuning Parameter Calibration

Our estimates of the quantiles of the effective noise can be used for different tasks, with inference and tuning parameter calibration as two examples. In inference, the choice of α is determined by the significance level. In tuning parameter calibration, in contrast, α can be chosen freely. In what follows, we examine how our tuning parameter calibration is influenced by the choice of α. To do so, we repeat the simulation exercises from Section 5.2 (with $\kappa = 0.25$) for three different values of α, namely $\alpha = 0.01, 0.05, 0.1$. Choosing α in the range between 0.01 and 0.1 in practice is sensible for the following reasons: The constraint $\alpha \leq 0.1$ makes sure that the finite sample guarantees for tuning parameter calibration from Section 4.1 hold with reasonably high probability ($\approx 90\%$ or higher). The constraint $\alpha \geq 0.01$, on the other hand, ensures that the bias of the lasso does not get overly strong. We thus restrict attention to $\alpha \in [0.01, 0.1]$, which is also the range of typical significance levels in testing.

Figure S.1 reports the results for the Hamming loss. The grey-shaded area in each panel depicts the histogram of the Hamming distances $\Delta_H(\hat{\beta}, \beta^*)$ that are produced by our estimator $\hat{\beta}$ over the $N = 1000$ simulation runs when α is set to 0.05, the red line depicts the histogram for $\alpha = 0.01$, and the blue line the histogram for $\alpha = 0.1$. In addition, the histogram of the cross-validated estimator is shown as a dotted line. Notice that the grey-shaded histograms in Figure S.1 are the same as those in Figure 3a. Figures S.2–S.4 present the results for the ℓ_1-, ℓ_∞- and prediction loss in an analogous way. Inspecting the plots, we conclude that the precise choice of α only has a minor effect on tuning parameter calibration with our method.

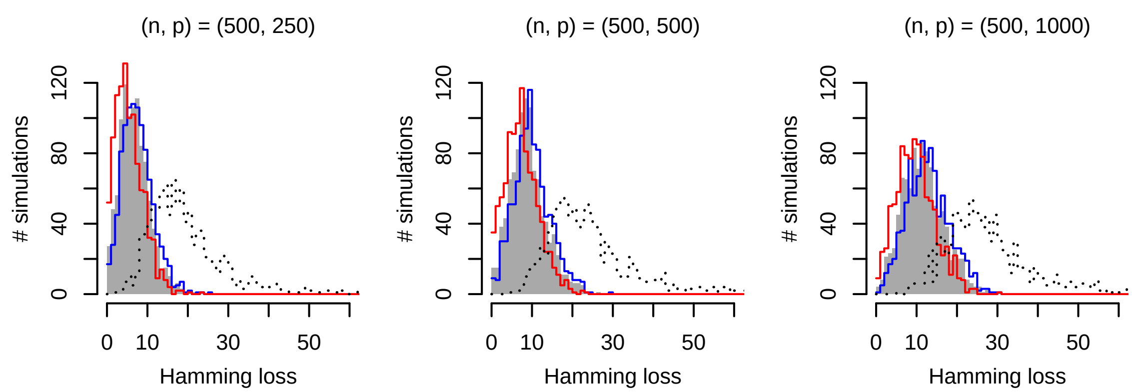


Figure S.1: Histograms of the Hamming loss for different values of α.

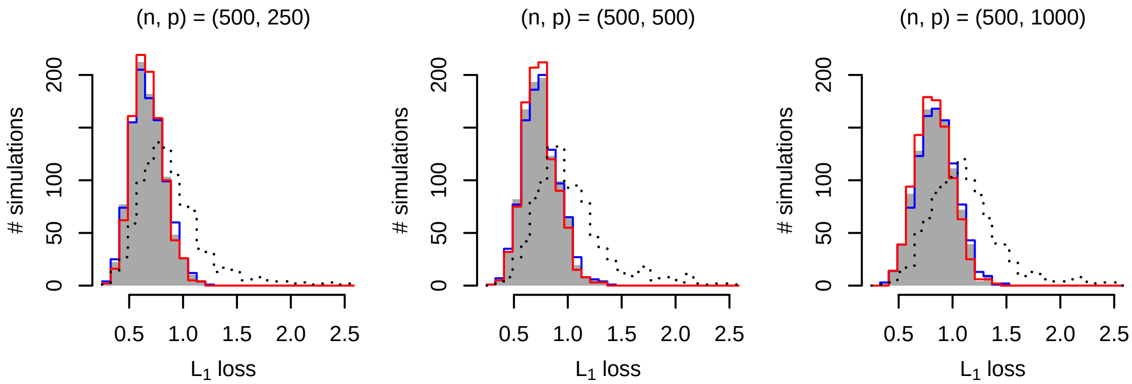


Figure S.2: Histograms of the ℓ_1-loss for different values of α.

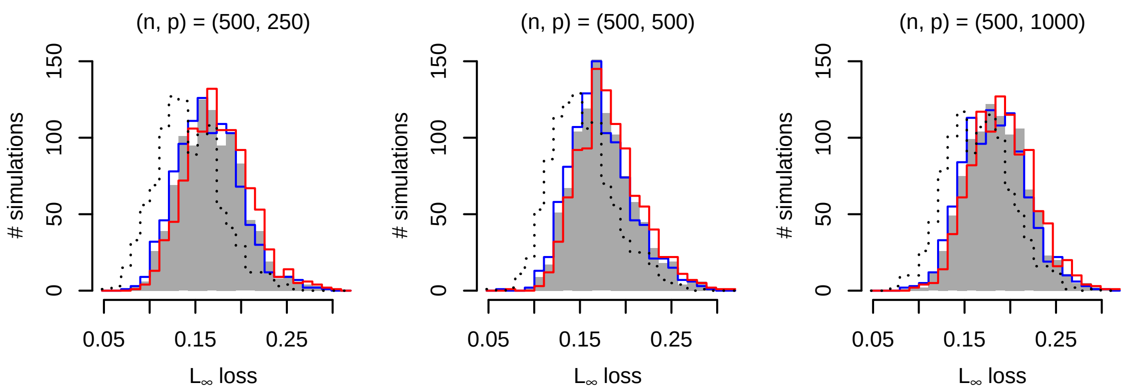


Figure S.3: Histograms of the ℓ_∞-loss for different values of α.

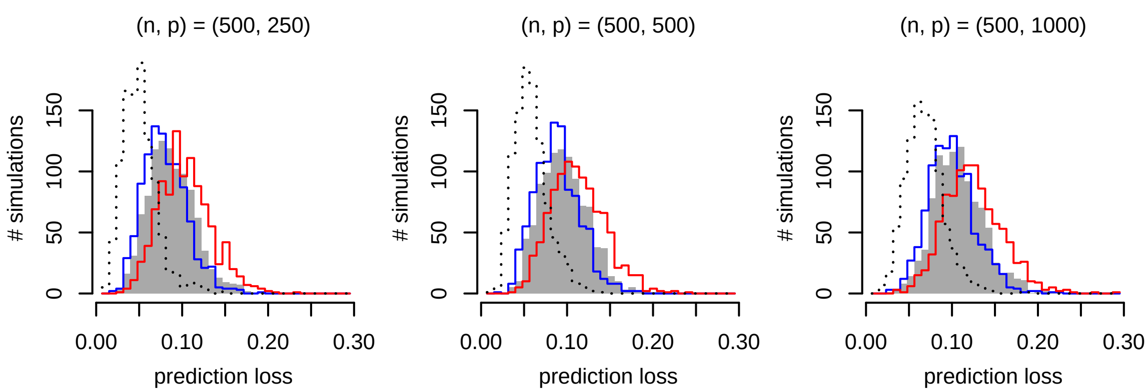


Figure S.4: Histograms of the prediction loss for different values of α.

Different Distributions of the Noise and the Design

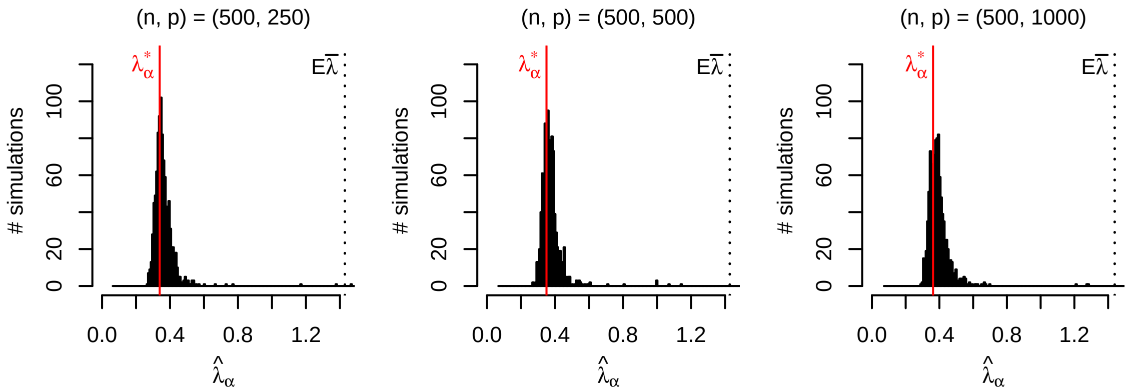
In this section, we investigate how our simulation results are influenced by the distribution of the noise ε_i and the design X_i. In order to do so, we repeat the simulation exercises from Sections 5.1–5.3 with non-normal noise variables ε_i and design vectors X_i. Specifically, we sample ε_i independently from a t-distribution with d degrees of freedom and variance normalized to 1. Moreover, X_i is drawn from a multivariate t-distribution with the same number of degrees of freedom, where the covariance matrix is the same as in Section 5 (in particular, it is given by $(1-\kappa)\boldsymbol{I} + \kappa\boldsymbol{E}$ with $\kappa = 0.25$). We consider three different choices of d, namely $d \in \{5, 10, 30\}$. For small d, the t-distribution differs substantially from the standard normal law, having much heavier tails. (Note in particular that $d = 5$ is the smallest integer for which the t-distribution has $\theta > 4$ moments as required by condition (C3).) As d increases, the t-distribution becomes less heavy-tailed and more akin to a standard normal law.

We start with the simulations from Section 5.1, which concern the approximation quality of our estimator $\hat{\lambda}_\alpha$. To see how the quality of $\hat{\lambda}_\alpha$ depends on the distribution of the noise and the design, we reproduce Figure 2 for the case of t-distributed errors and design vectors with $d \in \{5, 10, 30\}$. The results are reported in Figure S.5. As can be seen, the precision of our estimator diminishes somewhat as d gets smaller. Nevertheless, even for the case $d = 5$, we obtain quite precise results.

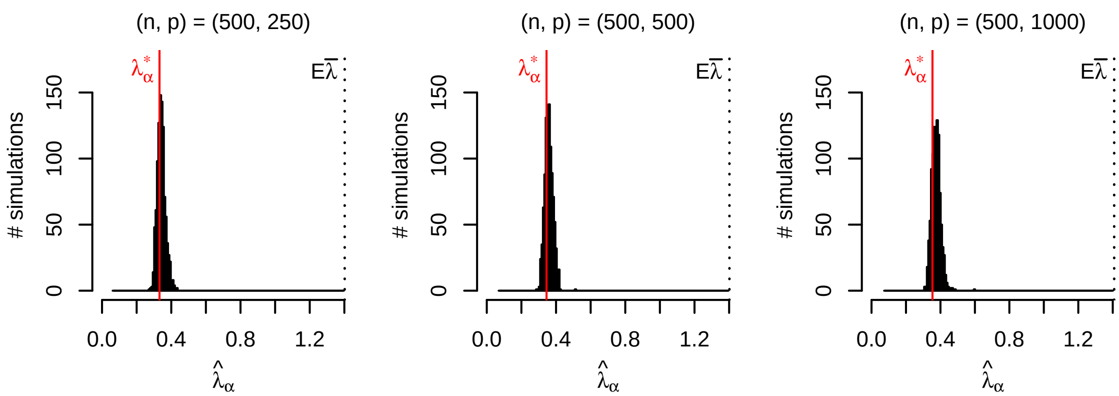
We now turn to the simulations on tuning parameter calibration from Section 5.2. We reproduce Figures 3a, 4, 5 and 6, which correspond to the four different losses under consideration, for the case of t-distributed noise terms and design vectors with $d \in \{5, 10, 30\}$. The results are presented in Figures S.6–S.9. The format is the same as in Figures 3a, 4, 5 and 6: the grey-shaded areas correspond to the histograms produced by our estimator, the black lines correspond to the histograms of the oracle method, and the dotted lines correspond to the histograms of the cross-validated lasso. In all of the considered cases, the histograms of our estimator are extremely close to those of the oracle. Moreover, the histograms are very similar to those of Figures 3a, 4, 5 and 6 for the Gaussian case.

We finally revisit the inference results from Section 5.3. As before, we repeat the simulations with t-distributed noise and design vectors for $d \in \{5, 10, 30\}$. The results are given in Tables S.1–S.3. For all considered values of d, the size of the test under the null is close to the target α. Moreover, the power of the test is comparable to that in the Gaussian case, even though it gets a bit lower for smaller d.

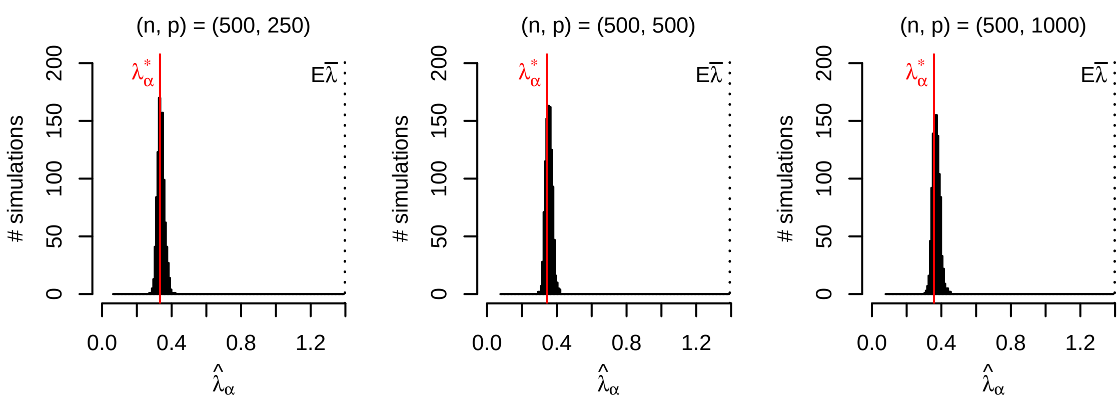
To summarize, the results demonstrate that our method does not require normally distributed noise and design, which supports our general theory in the main part of the paper.



(a) histograms for t-distributed noise and design with $d = 5$

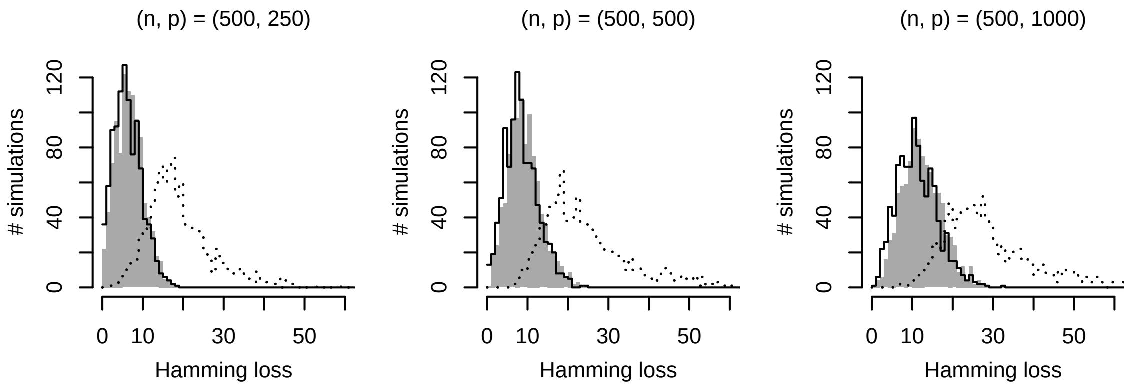


(b) histograms for t-distributed noise and design with $d = 10$

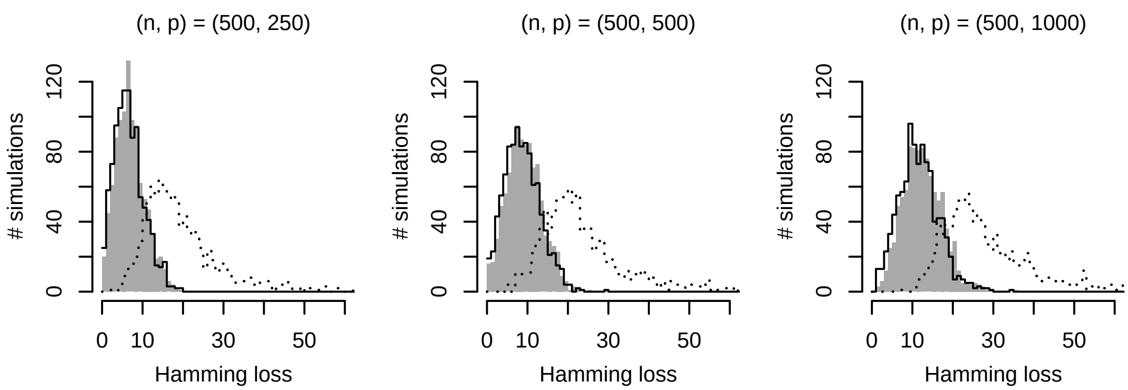


(c) histograms for t-distributed noise and design with $d = 30$

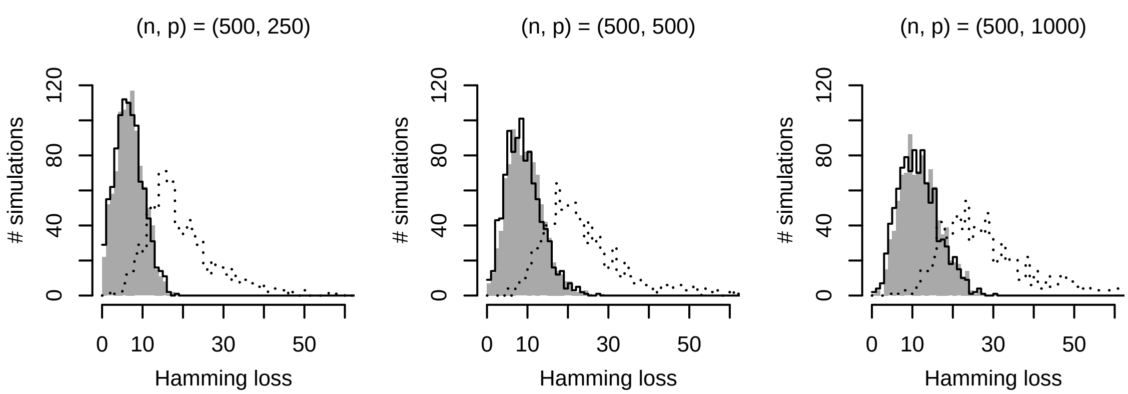
Figure S.5: Histograms of the estimates $\hat{\lambda}_\alpha$ for t-distributed noise variables and design vectors with $d \in \{5, 10, 30\}$.



(a) histograms for t-distributed noise and design with $d = 5$

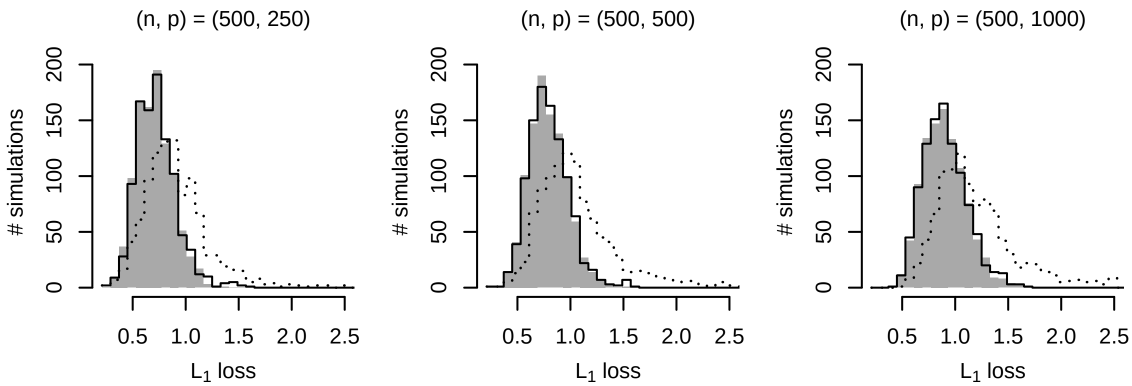


(b) histograms for t-distributed noise and design with $d = 10$

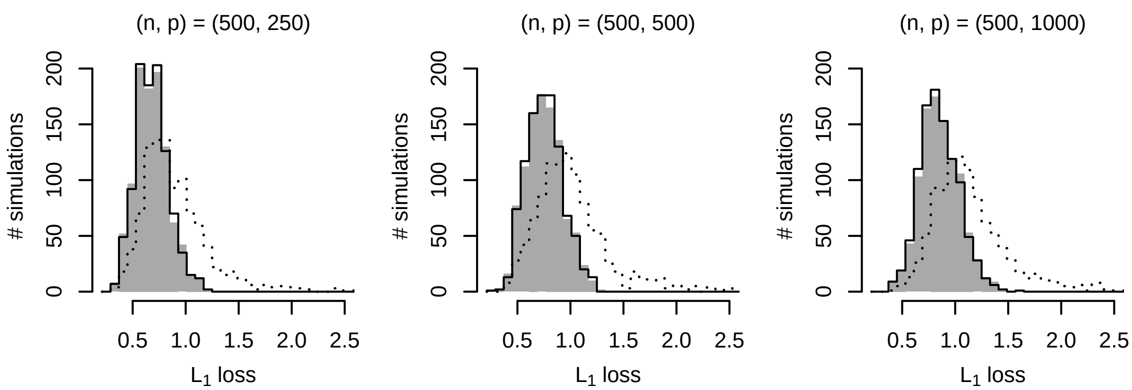


(c) histograms for t-distributed noise and design with $d = 30$

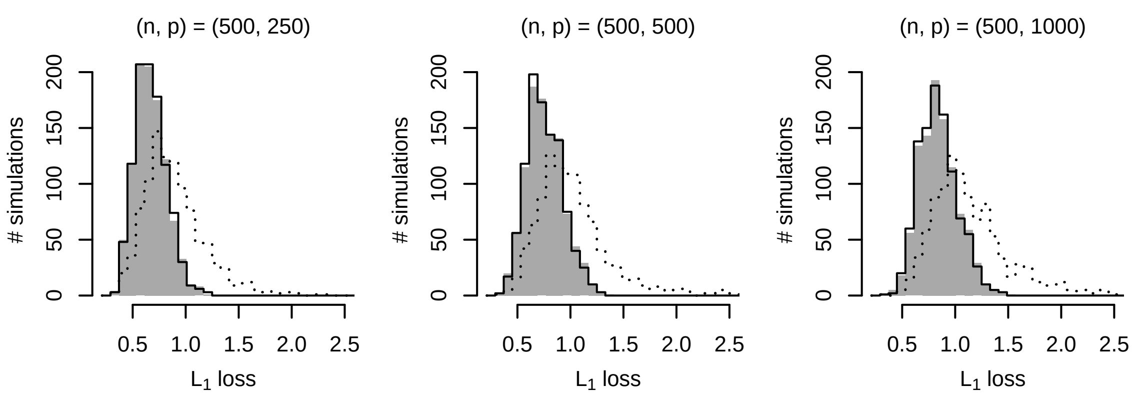
Figure S.6: Histograms of the Hamming loss for t-distributed noise variables and design vectors with $d \in \{5, 10, 30\}$.



(a) histograms for t-distributed noise and design with $d = 5$

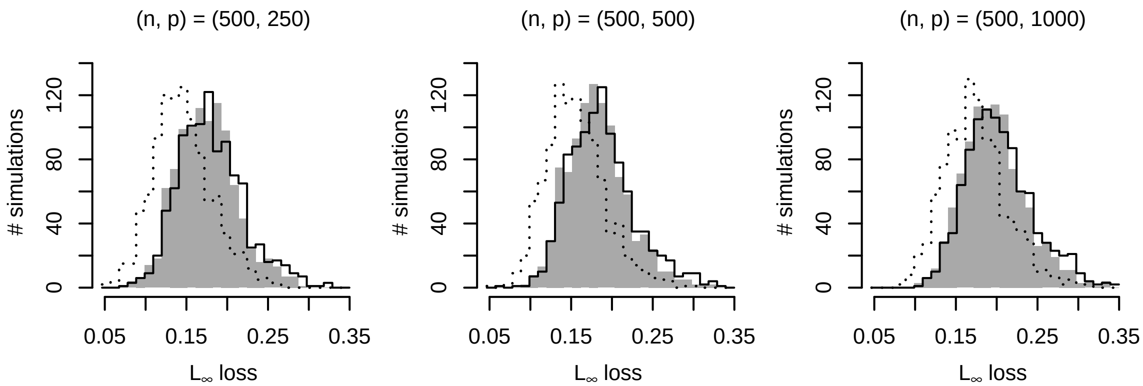


(b) histograms for t-distributed noise and design with $d = 10$

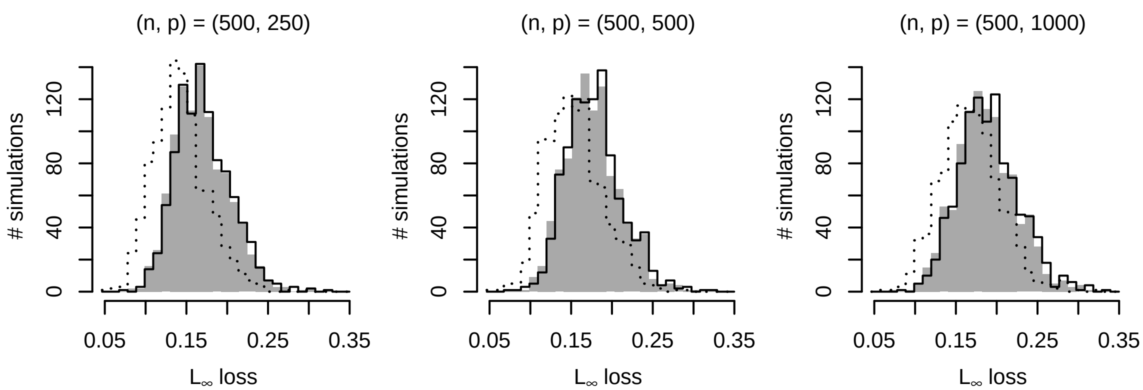


(c) histograms for t-distributed noise and design with $d = 30$

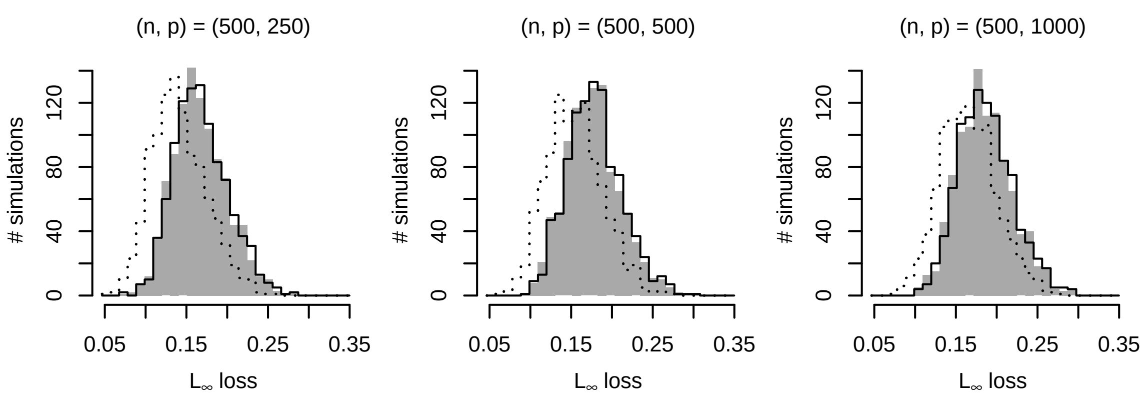
Figure S.7: Histograms of the ℓ_1-loss for t-distributed noise variables and design vectors with $d \in \{5, 10, 30\}$.



(a) histograms for t-distributed noise and design with $d = 5$

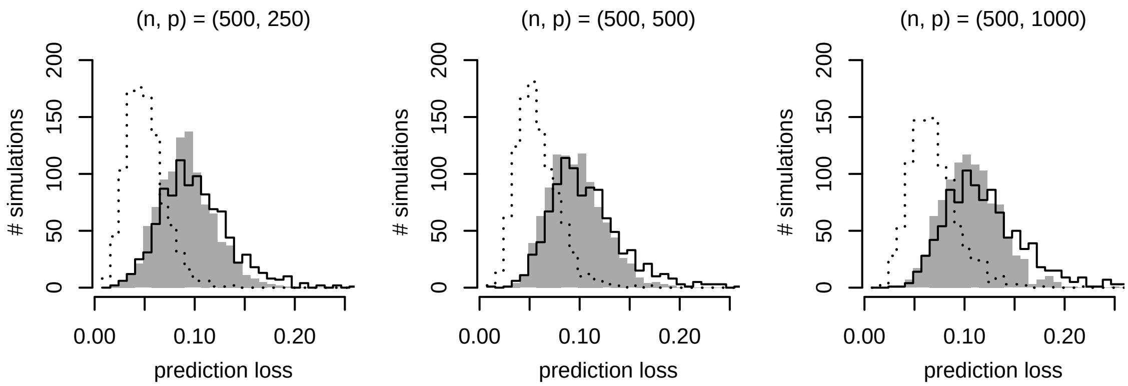


(b) histograms for t-distributed noise and design with $d = 10$

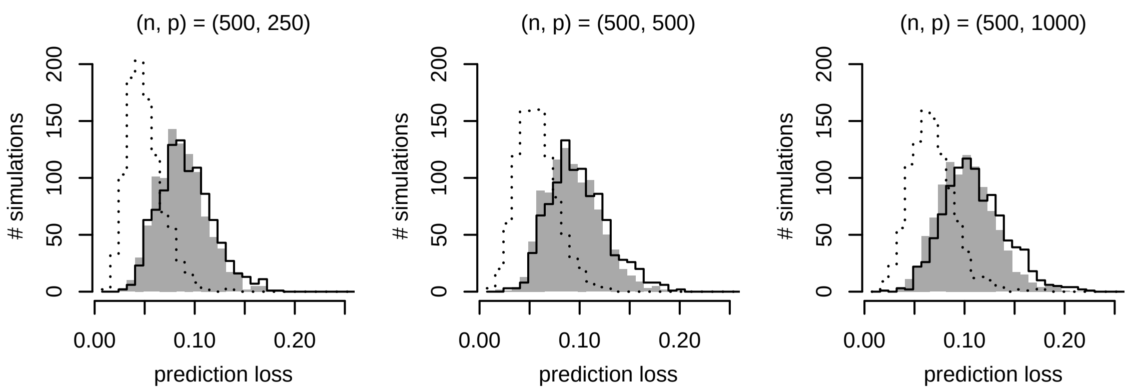


(c) histograms for t-distributed noise and design with $d = 30$

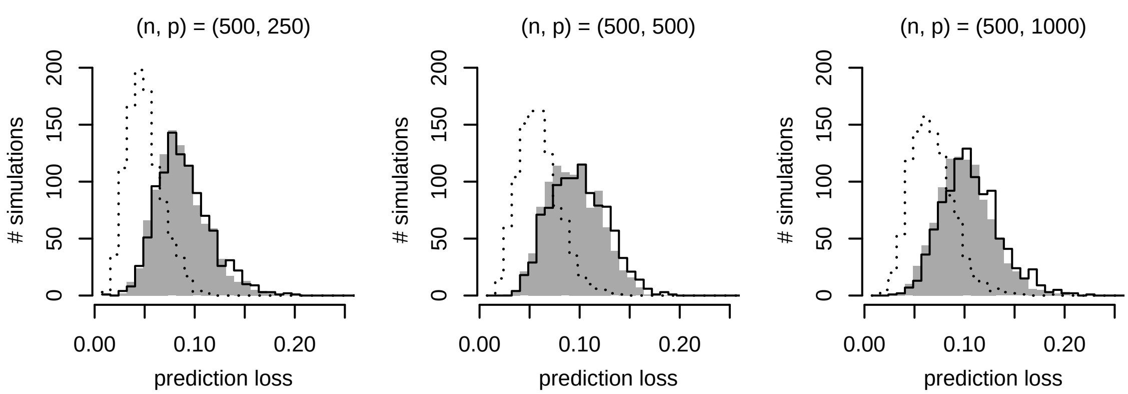
Figure S.8: Histograms of the ℓ_∞-loss for t-distributed noise variables and design vectors with $d \in \{5, 10, 30\}$.



(a) histograms for t-distributed noise and design with $d = 5$



(b) histograms for t-distributed noise and design with $d = 10$



(c) histograms for t-distributed noise and design with $d = 30$

Figure S.9: Histograms of the prediction loss for t-distributed noise variables and design vectors with $d \in \{5, 10, 30\}$.

(a) empirical size under $H_0 : \beta^* = 0$

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.011	0.033	0.080	0.009	0.058	0.105
$(n, p) = (500, 500)$	0.009	0.036	0.078	0.013	0.054	0.094
$(n, p) = (500, 1000)$	0.007	0.028	0.067	0.018	0.061	0.095

(b) empirical power under the alternative with SNR = 0.1

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.092	0.220	0.346	0.060	0.247	0.360
$(n, p) = (500, 500)$	0.100	0.247	0.401	0.107	0.298	0.419
$(n, p) = (500, 1000)$	0.085	0.223	0.365	0.139	0.309	0.397

(c) empirical power under the alternative with SNR = 0.2

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.471	0.741	0.856	0.602	0.836	0.917
$(n, p) = (500, 500)$	0.510	0.762	0.874	0.617	0.865	0.929
$(n, p) = (500, 1000)$	0.453	0.725	0.852	0.656	0.843	0.908

Table S.1: Empirical size under the null and power against different alternatives for t-distributed noise variables and design vectors with $d = 5$.

(a) empirical size under $H_0 : \beta^* = 0$

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.012	0.054	0.087	0.011	0.045	0.095
$(n, p) = (500, 500)$	0.012	0.047	0.100	0.004	0.053	0.102
$(n, p) = (500, 1000)$	0.005	0.033	0.080	0.005	0.041	0.084

(b) empirical power under the alternative with SNR = 0.1

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.111	0.260	0.403	0.109	0.279	0.399
$(n, p) = (500, 500)$	0.119	0.270	0.393	0.096	0.297	0.418
$(n, p) = (500, 1000)$	0.106	0.247	0.374	0.088	0.262	0.376

(c) empirical power under the alternative with SNR = 0.2

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.607	0.822	0.922	0.649	0.860	0.938
$(n, p) = (500, 500)$	0.578	0.790	0.891	0.592	0.832	0.910
$(n, p) = (500, 1000)$	0.556	0.806	0.895	0.567	0.851	0.909

Table S.2: Empirical size under the null and power against different alternatives for t-distributed noise variables and design vectors with $d = 10$.

(a) empirical size under $H_0 : \beta^* = 0$

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.012	0.057	0.111	0.007	0.056	0.105
$(n, p) = (500, 500)$	0.019	0.057	0.099	0.012	0.070	0.115
$(n, p) = (500, 1000)$	0.011	0.044	0.082	0.004	0.055	0.101

(b) empirical power under the alternative with SNR = 0.1

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.120	0.275	0.421	0.113	0.288	0.417
$(n, p) = (500, 500)$	0.130	0.278	0.391	0.129	0.335	0.437
$(n, p) = (500, 1000)$	0.140	0.287	0.407	0.119	0.327	0.446

(c) empirical power under the alternative with SNR = 0.2

	feasible test			oracle test		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$(n, p) = (500, 250)$	0.629	0.824	0.908	0.653	0.855	0.922
$(n, p) = (500, 500)$	0.605	0.823	0.905	0.660	0.892	0.935
$(n, p) = (500, 1000)$	0.619	0.831	0.906	0.620	0.854	0.935

Table S.3: Empirical size under the null and power against different alternatives for t-distributed noise variables and design vectors with $d = 30$.

Choice of L and M

When implementing our method, we need to choose the number of bootstrap iterations L as well as the grid size M for computing the lasso estimates. We have experimented with different choices of L and M and found that they have little effect on the simulation results. To illustrate this, we consider the same simulation setting as in Section 5.1 and produce $N = 1000$ estimates of $\hat{\lambda}_\alpha$ for different choices of (L, M). In addition to the choice $(L, M) = (100, 100)$ which is used in Section 5, we consider the choices $(L, M) = (200, 200)$ and $(L, M) = (300, 300)$. Figure S.10 reports the results. In each panel, the grey-shaded area is the histogram of the $N = 1000$ estimates of $\hat{\lambda}_\alpha$ for the choice $(L, M) = (100, 100)$, the blue line is the histogram for $(L, M) = (200, 200)$, and the red line is the histogram for $(L, M) = (300, 300)$. As one can see, the histograms are very similar across the different choices of (L, M), which suggests that the precise choice of L and M has little effect on our method.

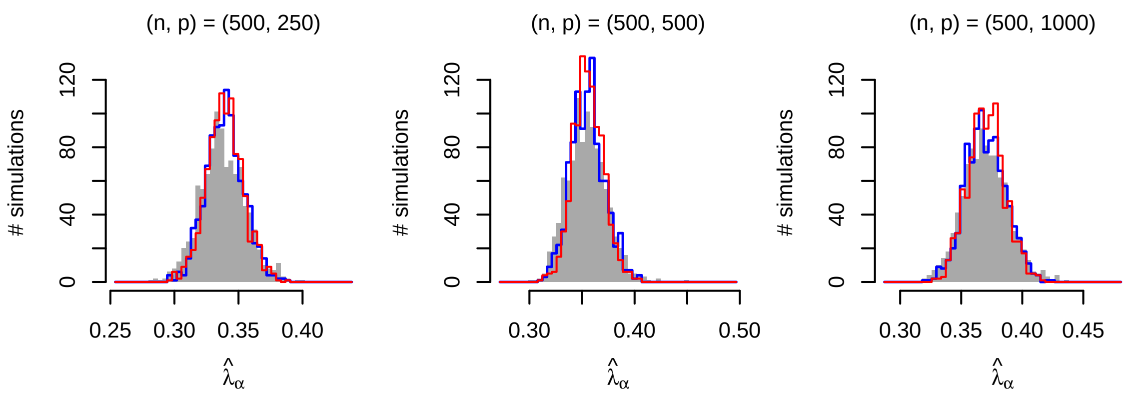


Figure S.10: Histograms of the estimates $\hat{\lambda}_\alpha$ for different choices of (L, M).