Statistica Sinica: Supplement

MEASUREMENT ERROR IN LASSO: IMPACT AND LIKELIHOOD BIAS CORRECTION

Øystein Sørensen, Arnoldo Frigessi and Magne Thoresen

University of Oslo

Supplementary Material

This note contains proofs of theorems and propositions stated in the paper.

S1 Regularity conditions

We assume fixed true covariates which satisfy

$$(1/n)\mathbf{X}'\mathbf{X} = \mathbf{C}_{xx} \rightarrow \mathbf{\Sigma}_{xx}, \text{ as } n \rightarrow \infty, \tag{S1.1}$$
$$(1/n) \max_{1 \le i \le n} (\mathbf{x}_i'\mathbf{x}_i) \rightarrow 0, \text{ as } n \rightarrow \infty, \tag{S1.2}$$

where $\mathbf{\Sigma}_{xx}$ is a positive definite matrix.

The random measurement errors are assumed normally distributed with mean zero and covariance $\mathbf{\Sigma}_{uu}$. It follows (Anderson (2003, Th. 3.4.4)) that the limiting distribution of $n^{1/2}(\mathbf{C}_{uu} - \mathbf{\Sigma}_{uu})$ is normal with mean $\mathbf{0}$ and covariances $(\mathbf{\Sigma}_{uu})_{ik}(\mathbf{\Sigma}_{uu})_{jl} + (\mathbf{\Sigma}_{uu})_{il}(\mathbf{\Sigma}_{uu})_{jk}$, where $(\mathbf{\Sigma}_{uu})_{ik}$ is the (i,k)th element of $\mathbf{\Sigma}_{uu}$ and $i,j,k,l \in \{1,\ldots,p\}$. Now,

$$\mathbf{C}_{uu} \rightarrow \mathbf{\Sigma}_{uu}, \text{ as } n \rightarrow \infty, \tag{S1.3}$$
$$(1/n) \max_{1 \le i \le n} (\mathbf{u}_i'\mathbf{u}_i) \rightarrow 0, \text{ as } n \rightarrow \infty, \tag{S1.4}$$

hold with probability 1. It follows from (S1.1)-(S1.4) that with probability 1, $\mathbf{C}_{ww} \rightarrow \mathbf{\Sigma}_{ww}$ and $(1/n) \max_{1 \le i \le n} (\mathbf{w}_i'\mathbf{w}_i) \rightarrow 0$, as $n \rightarrow \infty$. The limiting distribution of $n^{1/2}(\mathbf{C}_{ww} - \mathbf{\Sigma}_{ww})$ has mean zero and finite covariances. Regularity conditions like these have also been assumed by, e.g., Knight and Fu (2000) and Zhao and Yu (2006).

S2 Karush-Kuhn-Tucker Conditions

We introduce the new coefficient $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$, which yields the naive lasso on the form

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left(-\frac{2}{n}\boldsymbol{\epsilon}'\mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\gamma}'\mathbf{C}_{ww}\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'\mathbf{C}_{wu}\boldsymbol{\beta}^0 + \lambda \left\|\boldsymbol{\gamma} + \boldsymbol{\beta}^0\right\|_1 \right), \tag{S2.1}$$

where we have removed all terms which are constant in $\boldsymbol{\gamma}$. Taking derivatives, we arrive at the Karush-Kuhn-Tucker (KKT) conditions for the naive Lasso.

Lemma 1. *$\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$ is a solution to (S2.1) if and only if $-(2/n)\boldsymbol{\epsilon}'\mathbf{W} + 2\mathbf{C}_{ww}\hat{\boldsymbol{\gamma}} + 2\mathbf{C}_{wu} = -\lambda\hat{\boldsymbol{\tau}}$, where $\hat{\boldsymbol{\tau}} \in \mathbb{R}^p$ satisfies $\|\hat{\boldsymbol{\tau}}\|_\infty \leq 1$ and $\hat{\tau}_j = sign\left(\hat{\beta}_j\right)$ for j such that $\hat{\beta}_j \neq 0$.*

The same change of variables for the corrected lasso yields

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}\,:\,\|\boldsymbol{\gamma}+\boldsymbol{\beta}^0\|_1 \leq R} \Big\{ -\frac{2}{n}\boldsymbol{\epsilon}'\mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\gamma}'\left(\mathbf{C}_{ww} - \boldsymbol{\Sigma}_{uu}\right)\boldsymbol{\gamma} \tag{S2.2}$$
$$+ 2\boldsymbol{\gamma}'\left(\mathbf{C}_{wu} - \boldsymbol{\Sigma}_{uu}\right)\boldsymbol{\beta}^0 + \lambda\left\|\boldsymbol{\gamma} + \boldsymbol{\beta}^0\right\|_1 \Big\}.$$

Due to the additional constraint $\left\|\boldsymbol{\gamma} + \boldsymbol{\beta}^0\right\|_1 \leq R$ added because of non-convexity, the KKT conditions can only characterize critical points in the interior of this domain. A critical point on the boundary may not have a zero subgradient. Under the assumptions of Loh and Wainwright (2012), for sufficiently large n, all local optima lie in a small ℓ_1-ball around $\boldsymbol{\beta}^0$. We assume that R is chosen large enough such that $\left\|\boldsymbol{\gamma} + \boldsymbol{\beta}^0\right\|_1 < R$ for all these optima, while R is small enough to avoid the trivial solutions for which one or more component of $\hat{\boldsymbol{\gamma}}$ is $\pm\infty$.

Lemma 2. *Assume $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$ is a critical point of (S2.2). If $\left\|\hat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^0\right\|_1 < R$, then $-(2/n)\boldsymbol{\epsilon}'\mathbf{W} + 2\left(\mathbf{C}_{ww} - \boldsymbol{\Sigma}_{uu}\right)\hat{\boldsymbol{\gamma}} + 2\left(\mathbf{C}_{wu} - \boldsymbol{\Sigma}_{uu}\right) = -\lambda\hat{\boldsymbol{\tau}}$, where $\hat{\boldsymbol{\tau}} \in \mathbb{R}^p$ is as defined in Lemma 1.*

S3 Proof of Proposition 1

By definition (Bühlmann and van de Geer (2011)),

$$(1/n)\left\|\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\beta}}\right\|_2^2 + \lambda\left\|\hat{\boldsymbol{\beta}}\right\|_1 \leq (1/n)\left\|\mathbf{y} - \mathbf{W}\boldsymbol{\beta}^0\right\|_2^2 + \lambda\left\|\boldsymbol{\beta}^0\right\|_1,$$

and after reorganizing terms,

$$(1/n)\left\|\mathbf{W}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)\right\|_2^2 + \lambda\left\|\hat{\boldsymbol{\beta}}\right\|_1 \leq (2/n)\left(\boldsymbol{\epsilon} - \mathbf{U}\boldsymbol{\beta}^0\right)'\mathbf{W}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right) + \lambda\left\|\boldsymbol{\beta}^0\right\|_1. \tag{S3.1}$$

Under (S3.1),

$$(2/n)\left(\boldsymbol{\epsilon} - \mathbf{U}\boldsymbol{\beta}^0\right)'\mathbf{W}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right) \leq (2/n)\left\|\left(\boldsymbol{\epsilon} - \mathbf{U}\boldsymbol{\beta}^0\right)\mathbf{W}\right\|_\infty\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right\|_1 \leq \lambda_0\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right\|_1,$$

which inserted into (S3.1) yields

$$(1/n)\left\|\mathbf{W}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right)\right\|_2^2+\lambda\left\|\hat{\boldsymbol{\beta}}\right\|_1\leq\lambda_0\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|_1+\lambda\left\|\boldsymbol{\beta}^0\right\|_1.$$

Now use,

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}}\right\|_1&\geq\|\boldsymbol{\beta}^0_{S_0}\|_1-\left\|\hat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}^0_{S_0}\right\|_1+\left\|\hat{\boldsymbol{\beta}}_{S_0^c}\right\|_1,\\ \left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|_1&=\left\|\hat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}^0_{S_0}\right\|_1+\left\|\hat{\boldsymbol{\beta}}_{S_0^c}\right\|_1,\end{aligned}\tag{S3.2}$$

and $\lambda\geq 2\lambda_0$, to obtain

$$(2/n)\left\|\mathbf{W}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right)\right\|_2^2+\lambda\left\|\hat{\boldsymbol{\beta}}_{S_0^c}\right\|_1\leq 3\lambda\left\|\hat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}^0_{S_0}\right\|_1.\tag{S3.3}$$

Inequality (S3.3) shows that $\|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1\leq 3\|\hat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}^0_{S_0}\|_1$. That is, the vector $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0$ is among the vectors to which the compatibility condition applies, for the index set S_0. Next, use (S3.2) again in (S3.3) to obtain

$$(2/n)\left\|\mathbf{W}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right)\right\|_2^2+\lambda\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|_1\leq 4\lambda\left\|\hat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}^0_{S_0}\right\|_1.\tag{S3.4}$$

Under the compatibility condition on S_0,

$$\left\|\hat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}^0_{S_0}\right\|_1\leq s_0^{1/2}\phi_0^{-1}n^{-1/2}\left\|\mathbf{W}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right)\right\|_2.$$

Using this and the inequality $4uv\leq 4u^2+v^2$ in (S3.4), we arrive at

$$\left\|\mathbf{W}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right)\right\|_2^2+\lambda\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0_{S_0}\right\|_1\leq 4\lambda^2 s_0/\phi_0^2.$$

S4 Proof of Proposition 2

This proof goes along the lines of the proof of Theorem 1 in Knight and Fu (2000), but with the addition of measurement error. We start with the naive Lasso after reparametrization, and let

$$\mathcal{L}_n(\boldsymbol{\gamma})=-\frac{2}{\sqrt{n}}\boldsymbol{\gamma}'\frac{\mathbf{W}}{\sqrt{n}}\boldsymbol{\epsilon}+\boldsymbol{\gamma}'\mathbf{C}_{ww}\boldsymbol{\gamma}+2\boldsymbol{\gamma}'\mathbf{C}_{wu}\boldsymbol{\beta}^0+\lambda\|\boldsymbol{\gamma}+\boldsymbol{\beta}^0\|_1.\tag{S4.5}$$

Note that $(2/\sqrt{n})(\mathbf{W}/\sqrt{n})\boldsymbol{\epsilon}\xrightarrow{d}\mathcal{N}(\mathbf{0},(4/n)\sigma^2\boldsymbol{\Sigma}_{ww})$; the first term in (S4.5) converges in distribution to a normally distributed quantity whose variance goes to zero as $1/n$, which is equivalent to convergence in probability to zero. Combining this result with the assumption that $\lambda\to 0$ as $n\to\infty$, yields $\mathcal{L}_n(\boldsymbol{\gamma})\xrightarrow{p}\mathcal{L}(\boldsymbol{\gamma})=\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{ww}\boldsymbol{\gamma}+2\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{uu}\boldsymbol{\beta}^0$. Since $\mathcal{L}_n(\boldsymbol{\gamma})$ is convex, it follows that $\underset{\boldsymbol{\gamma}}{\operatorname{argmin}}\{\mathcal{L}_n(\boldsymbol{\gamma})\}\xrightarrow{p}\underset{\boldsymbol{\gamma}}{\operatorname{argmin}}\{\mathcal{L}(\boldsymbol{\gamma})\}$ (Knight and Fu (2000)). The minimum of $\mathcal{L}(\boldsymbol{\gamma})$ is easily found, and accordingly, $\hat{\boldsymbol{\gamma}}\xrightarrow{p}-\boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{uu}\boldsymbol{\beta}^0$. The result follows immediately.

S5 Proof of Theorem 1

We follow the structure of the proof by Zhao and Yu (2006), who proved the corresponding result in the absence of measurement error. Consider the naive lasso, and note that

$$\left\{\text{sign}\left(\boldsymbol{\beta}_{S_0}^0\right)\hat{\gamma}_{S_0} > -\left|\boldsymbol{\beta}_{S_0}^0\right|\right\} \;\Rightarrow\; \left\{\text{sign}\left(\hat{\boldsymbol{\beta}}_{S_0}\right) = \text{sign}\left(\boldsymbol{\beta}_{S_0}^0\right)\right\}$$

and $\hat{\boldsymbol{\gamma}}_{S_0^c} = \mathbf{0} \;\Rightarrow\; \hat{\boldsymbol{\beta}}_{S_0^c} = \mathbf{0}$. Thus, by the KKT conditions for the naive lasso (Lemma 1), if a solution $\hat{\boldsymbol{\gamma}}$ exists, and

$$-\frac{\mathbf{W}_{S_0}'}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n}\mathbf{C}_{ww}\left(S_0, S_0\right)\hat{\boldsymbol{\gamma}}_{S_0} + \sqrt{n}C_{wu}\left(S_0, S_0\right)\boldsymbol{\beta}_{S_0}^0 = -\frac{\lambda\sqrt{n}}{2}\text{sign}\left(\boldsymbol{\beta}_{S_0}^0\right), \tag{S5.1}$$

$$\left|\hat{\boldsymbol{\gamma}}_{S_0}\right| < \left|\boldsymbol{\beta}_{S_0}^0\right|, \tag{S5.2}$$

$$\left|-\frac{W_{S_0^c}'}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n}\mathbf{C}_{ww}\left(S_0^c, S_0\right)\hat{\boldsymbol{\gamma}}_{S_0} + \sqrt{n}\mathbf{C}_{wu}\left(S_0^c, S_0\right)\boldsymbol{\beta}_{S_0}^0\right| \leq \frac{\lambda\sqrt{n}}{2}\mathbf{1}, \tag{S5.3}$$

then $\text{sign}(\hat{\boldsymbol{\beta}}_{S_0}) = \text{sign}(\boldsymbol{\beta}_{S_0}^0)$ and $\text{sign}(\hat{\boldsymbol{\beta}}_{S_0^c}) = \mathbf{0}$.

Event A implies the existence of $|\hat{\boldsymbol{\gamma}}_{S_0}| < |\boldsymbol{\beta}_{S_0}^0|$ such that

$$\left|\mathbf{Z}_1'\boldsymbol{\epsilon} - \mathbf{Z}_2\boldsymbol{\beta}_{S_0}^0\right| = \sqrt{n}\left(\left|\hat{\boldsymbol{\gamma}}_{S_0}\right| - \frac{\lambda}{2}\left|\mathbf{C}_{ww}\left(S_0, S_0\right)^{-1}\text{sign}\left(\boldsymbol{\beta}_{S_0}^0\right)\right|\right).$$

Buth then there must also exist $|\hat{\boldsymbol{\gamma}}_{S_0}| < |\boldsymbol{\beta}_{S_0}^0|$ such that

$$\mathbf{Z}_1'\boldsymbol{\epsilon} - \mathbf{Z}_2\boldsymbol{\beta}_{S_0}^0 = \sqrt{n}\left(\hat{\boldsymbol{\gamma}}_{S_0} - \frac{\lambda}{2}\mathbf{C}_{ww}\left(S_0, S_0\right)^{-1}\text{sign}\left(\boldsymbol{\beta}_{S_0}^0\right)\right),$$

which essentially means choosing the appropriate signs of the elements of $\hat{\boldsymbol{\gamma}}_{S_0}$. Multiplying through by $\mathbf{C}_{ww}(S_0, S_0)$ and reorganizing terms, we get (S5.1). Thus, A ensures that (S5.1) and (S5.2) are satisfied. Next, adding and subtracting $\sqrt{n}\mathbf{C}_{ww}(S_0^c, S_0)\hat{\boldsymbol{\gamma}}_{S_0}$ to the left-hand side of event B and then using the triangle inequaltity, yields

$$\begin{aligned} &\left|-\frac{\mathbf{W}_{S_0^c}}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n} + \mathbf{C}_{ww}(S_0^c, S_0)\hat{\boldsymbol{\gamma}}_{S_0} + \sqrt{n}\mathbf{C}_{wu}(S_0^c, S_0)\boldsymbol{\beta}_{S_0}^0\right| - \\ &\left| - \mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\frac{\mathbf{W}_{S_0}'}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n}\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}_{S_0}^0 \right. \\ &\left. + \sqrt{n}\mathbf{C}_{ww}(S_0^c, S_0)\hat{\boldsymbol{\gamma}}_{S_0}\right| \leq \frac{\lambda\sqrt{n}}{2}\left(1 - \theta\right)\mathbf{1}. \end{aligned}$$

The second term on the left-hand side of this expression is the left-hand side of (S5.1) multiplied by $\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}$. It can thus be replaced by the right-hand side of (S5.1) multiplied by this factor. This yields

$$\left|-\frac{\mathbf{W}_{S_0^c}}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n} + \mathbf{C}_{ww}(S_0^c, S_0)\hat{\boldsymbol{\gamma}}_{S_0} + \sqrt{n}\mathbf{C}_{wu}(S_0^c, S_0)\boldsymbol{\beta}_{S_0}^0\right| -$$

$$\left| \frac{\lambda\sqrt{n}}{2} \mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\text{sign}\left(\boldsymbol{\beta}^0_{S_0}\right)\right| \leq \frac{\lambda\sqrt{n}}{2}\left(1-\theta\right)\mathbf{1},$$

which implies, due to the IC-ME,

$$\left| -\frac{\mathbf{W}_{S_0^c}}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n}\mathbf{C}_{ww}(S_0^c, S_0)\hat{\boldsymbol{\gamma}}_{S_0} + \sqrt{n}\mathbf{C}_{wu}(S_0^c, S_0)\boldsymbol{\beta}^0_{S_0}\right| \leq \frac{\lambda\sqrt{n}}{2}\mathbf{1}.$$

This is indeed (S5.3). Altogether, A implies (S5.1) and (S5.2), while $B|A$ implies (S5.3).

For the asymptotic result, define the vectors

$$\begin{aligned}
\mathbf{z} &= \mathbf{C}_{ww}(S_0, S_0)^{-1}\frac{\mathbf{W}'_{S_0}}{\sqrt{n}}\boldsymbol{\epsilon},\\
\mathbf{a} &= \left|\boldsymbol{\beta}^0_{S_0}\right| - \left|\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0}\right|,\\
\mathbf{b} &= \mathbf{C}_{ww}(S_0, S_0)^{-1}\text{sign}\left(\boldsymbol{\beta}^0_{S_0}\right),\\
\boldsymbol{\zeta} &= \left(\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\frac{\mathbf{W}'_{S_0}}{\sqrt{n}} - \frac{\mathbf{W}'_{S_0^c}}{\sqrt{n}}\right)\boldsymbol{\epsilon},\\
\mathbf{f} &= \left(\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0) - \mathbf{C}_{wu}(S_0^c, S_0)\right)\boldsymbol{\beta}^0_{S_0}.
\end{aligned}$$

We have

$$\begin{aligned}
&1 - P(A \cap B) \leq P(A^c) + P(B^c) \leq \\
&\quad \sum_{j=1}^{s_0} P\left(|z_j| \geq \sqrt{n}\left(a_j - \frac{\lambda}{2}b_j\right)\right) + \sum_{j=1}^{p-s_0} P\left(\left|\zeta_j - \sqrt{n}f_j\right| \geq \frac{\lambda\sqrt{n}}{2}\left(1-\theta\right)\right).
\end{aligned}$$

It is clear that $\mathbf{z} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{C}_{ww}(S_0, S_0)^{-1}\right)$, as $n \to \infty$. Hence, there exists a finite constant k such that $E(z_j)^2 < k^2$ for $j = 1, \dots, s_0$. Next, we have by assumption

$$\begin{aligned}
\mathbf{a} &\to \left|\boldsymbol{\beta}^0_{S_0}\right| - \left|\boldsymbol{\Sigma}_{ww}(S_0, S_0)^{-1}\boldsymbol{\Sigma}_{uu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0}\right|, \text{ as } n \to \infty,\\
\mathbf{b} &\to \boldsymbol{\Sigma}_{ww}(S_0, S_0)^{-1}\text{sign}\left(\boldsymbol{\beta}^0_{S_0}\right), \text{ as } n \to \infty.
\end{aligned}$$

Now using the assumption $\lambda = o(1)$, we get

$$\begin{aligned}
P\left(A^c\right) &\leq \sum_{j=1}^{s_0}\left(1 - P\left(\frac{|z_j|}{k} < \frac{\sqrt{n}}{2k}a_j\left(1+o(1)\right)\right)\right)\\
&\leq \left(1+o(1)\right)\sum_{j=1}^{s_0}\left(1 - \Phi\left(\frac{\sqrt{n}}{2s}a_j\left(1+o(1)\right)\right)\right)\\
&= o\left(\exp(-n^c)\right),
\end{aligned}$$

where we used the bound for the Gaussian tail probability

$$1 - \Phi(t) < t^{-1}\exp\left(-(1/2)t^2\right). \tag{S5.4}$$

Next, we note that

$$\zeta \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2\left(\boldsymbol{\Sigma}_{ww}(S_0^c, S_0^c) - \boldsymbol{\Sigma}_{ww}(S_0^c, S_0)\boldsymbol{\Sigma}_{ww}(S_0, S_0)^{-1}\boldsymbol{\Sigma}_{ww}(S_0, S_0^c)\right)\right), \text{ as } n \to \infty.$$

Next, we consider $\mathbf{f}$, and note that the limiting distribution of $\sqrt{n}\mathbf{C}_{wu} = \sqrt{n}(\mathbf{C}_{uu}+\mathbf{C}_{xu})$ as $n \to \infty$ is normal with mean $\sqrt{n}\boldsymbol{\Sigma}_{wu} = \sqrt{n}\boldsymbol{\Sigma}_{uu}$ and finite variances (Anderson (2003, Th. 3.4.4)). In addition, $\mathbf{C}_{ww} \to \boldsymbol{\Sigma}_{ww}$ as $n \to \infty$. Thus, applying Slutsky's theorem to the product of the matrices, the limiting distribution of

$$\sqrt{n}\left(\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0) - \mathbf{C}_{wu}(S_0^c, S_0)\right)$$

as $n \to \infty$ has mean

$$\sqrt{n}\left(\boldsymbol{\Sigma}_{ww}(S_0^c, S_0)\boldsymbol{\Sigma}_{ww}(S_0, S_0)^{-1}\boldsymbol{\Sigma}_{wu}(S_0, S_0) - \boldsymbol{\Sigma}_{wu}(S_0^c, S_0)\right) = \mathbf{0}$$

and finite variances. The latter term equals zero by the MEC. Now

$$\sqrt{n}\mathbf{f} = \sqrt{n}\left(\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0) - \mathbf{C}_{wu}(S_0^c, S_0)\right)\boldsymbol{\beta}_{S_0}^0$$

is a vector in $\mathbb{R}^{p-s_0}$ whose elements are linear combinations of variables whose limiting distributions as $n \to \infty$ are normal with mean zero and finite variances. Accordingly, the limiting distribution of $\sqrt{n}\mathbf{f}$ as $n \to \infty$ is normal with mean zero and finite variances.

So again there exists a finite constant k such that $E(\zeta_j - \sqrt{n}f_j)^2 < k^2$ for $j = 1, \dots, (p - s_0)$. Thus, when $\lambda n^{(1-c)/2} \to \infty$ for $c \in [0, 1)$, we have

$$\begin{aligned}
P(B^c) &\leq \sum_{j=1}^{p-s_0}\left(1 - P\left(\frac{|\zeta_j - \sqrt{n}f_j|}{k} < \frac{1}{k}\frac{\lambda\sqrt{n}}{2}(1-\theta)\right)\right) \\
&\leq (1 + o(1))\sum_{j=1}^{p-s_0}\left(1 - \Phi\left(\frac{1}{k}\frac{\lambda\sqrt{n}}{2}(1-\theta)\right)\right) \\
&= o\left(\exp(-n^c)\right).
\end{aligned}$$

It follows that $P(A \cap B) = 1 - o(\exp(-n^c))$.

S6 Proof of Proposition 3

We consider now the Lasso with $\boldsymbol{\epsilon} = \mathbf{0}$. In this case, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^0$, and the Lasso becomes $\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left\{\|\mathbf{W}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta}^0\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right\}$. We follow the proof of Bühlmann and van de Geer (2011, Th. 7.1), but also take measurement error into account.

Part 1

The KKT conditions take the form

$$2\mathbf{C}_{ww}(S_0, S_0)(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0) + 2\mathbf{C}_{ww}(S_0, S_0^c)\hat{\boldsymbol{\beta}}_{S_0^c} + 2\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}_{S_0}^0 = -\lambda\hat{\boldsymbol{\tau}}_{S_0} \quad \text{(S6.1)}$$

$$2\mathbf{C}_{ww}(S_0^c, S_0)(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}^0_{S_0}) + 2\mathbf{C}_{ww}(S_0^c, S_0^c)\hat{\boldsymbol{\beta}}_{S_0^c} + 2\mathbf{C}_{wu}(S_0^c, S_0)\boldsymbol{\beta}^0_{S_0} = -\lambda\hat{\boldsymbol{\tau}}_{S_0^c}, \quad \text{(S6.2)}$$

where $\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\tau}}'_{S_0}, \hat{\boldsymbol{\tau}}'_{S_0^c})'$ has the properties $\|\hat{\boldsymbol{\tau}}\|_\infty \leq 1$ and $\hat{\tau}_j = \text{sign}(\hat{\beta}_j)$ if $\beta_j \neq 0$. We multiply (S6.1) by $\hat{\boldsymbol{\beta}}'_{S_0^c}\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}$ and (S6.2) by $\hat{\boldsymbol{\beta}}'_{S_0^c}$, and then subtract the first from the second, to get

$$\begin{aligned}
&2\hat{\boldsymbol{\beta}}'_{S_0^c}\left(\mathbf{C}_{ww}(S_0^c, S_0^c) - \mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{ww}(S_0, S_0^c)\right)\hat{\boldsymbol{\beta}}_{S_0^c} + \\
&\quad 2\hat{\boldsymbol{\beta}}'_{S_0^c}\left(\mathbf{C}_{wu}(S_0^c, S_0) - \mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\right)\boldsymbol{\beta}^0_{S_0} \\
&\quad = \lambda\left(\hat{\boldsymbol{\beta}}'_{S_0^c}\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\hat{\boldsymbol{\tau}}_{S_0} - \hat{\boldsymbol{\beta}}'_{S_0^c}\hat{\boldsymbol{\tau}}_{S_0^c}\right) \qquad \text{(S6.3)}
\end{aligned}$$

The matrix term within the parantheses in the leftmost term is positive semidefinite, since it is the Schur complement of the positive semidefinite matrix $\mathbf{C}_{ww}$, in which the part $\mathbf{C}_{ww}(S_0, S_0)$ is positive definite, since $s_0 < n$. Next, the term within the parantheses on the right-hand side is

$$\begin{aligned}
&\hat{\boldsymbol{\beta}}'_{S_0}\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\hat{\boldsymbol{\tau}}_{S_0} - \|\hat{\boldsymbol{\beta}}_{S_0}\|_1 \leq \\
&\left(\left\|\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\hat{\boldsymbol{\tau}}_{S_0}\right\|_\infty - 1\right)\left\|\hat{\boldsymbol{\beta}}_{S_0^c}\right\|_1 \leq 0.
\end{aligned}$$

The last inequality follows from the IC-ME, and is strict whenever $\|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \neq 0$. Finally, the second term on the left-hand side of (S6.3) is zero by assumption. Thus, if $\|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \neq 0$, the left-hand side of (S6.3) must be negative, which is a contradiction. We thus conclude that $\hat{\boldsymbol{\beta}}_{S_0^c} = \mathbf{0}$, and the KKT conditions (S6.1) and (S6.2) reduce to

$$2\mathbf{C}_{ww}(S_0, S_0)\left(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}^0_{S_0}\right) + 2\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0} = -\lambda\hat{\boldsymbol{\tau}}_{S_0} \qquad \text{(S6.4)}$$

$$2\mathbf{C}_{ww}(S_0^c, S_0)\left(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}^0_{S_0}\right) + 2\mathbf{C}_{wu}(S_0^c, S_0)\boldsymbol{\beta}^0_{S_0} = -\lambda\hat{\boldsymbol{\tau}}_{S_0^c}, \qquad \text{(S6.5)}$$

From (S6.4) we get

$$\begin{aligned}
&\left|\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}^0_{S_0}\right| = \left|\frac{\lambda}{2}\mathbf{C}_{S_0}(S_0, S_0)^{-1}\hat{\boldsymbol{\tau}}_{S_0} + \mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0}\right| \qquad \text{(S6.6)} \\
&\leq \left(\frac{\lambda}{2}\sup_{\|\boldsymbol{\tau}\|_\infty \leq 1}\left\|\mathbf{C}_{ww}(S_0, S_0)^{-1}\boldsymbol{\tau}_{S_0}\right\|_\infty\right)\mathbf{1} + \left|\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0}\right|.
\end{aligned}$$

Now, if $j \in S_0^{\text{det}}$ and $\hat{\beta}_j = 0$, then

$$\left|\hat{\beta}_j - \beta_j^0\right| = \left|\beta_j^0\right| > \frac{\lambda}{2} + \left(\sup_{\|\boldsymbol{\tau}_\infty\| \leq 1}\left\|\mathbf{C}_{ww}(S_0, S_0)^{-1}\boldsymbol{\tau}_{S_0}\right\|_\infty\right) + |v_j|,$$

where $\mathbf{v} = (v_1, \ldots, v_p)' = \mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0}$, contradicting (S6.6). Thus, $\hat{\beta}_j \neq 0$ for $j \in S_0^{\text{det}}$.

Part 2

We start by assuming $\text{sign}(\hat{\boldsymbol{\beta}}) = \text{sign}(\boldsymbol{\beta}^0)$. Thus, the KKT conditions are (S6.4) and (S6.5). From (S6.4) we get

$$\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}^0_{S_0} = -\frac{\lambda}{2}\mathbf{C}_{ww}(S_0, S_0)^{-1}\hat{\boldsymbol{\tau}}_{S_0} - \mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0)\boldsymbol{\beta}^0_{S_0}.$$

Inserting this into (S6.5) yields

$$\begin{aligned}&\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\hat{\boldsymbol{\tau}}_{S_0} + \\ &\frac{2}{\lambda}\left(\mathbf{C}_{ww}(S_0^c, S_0)\mathbf{C}_{ww}(S_0, S_0)^{-1}\mathbf{C}_{wu}(S_0, S_0) - \mathbf{C}_{wu}(S_0^c, S_0)\right)\boldsymbol{\beta}^0_{S_0} = \hat{\boldsymbol{\tau}}_{S_0^c},\end{aligned}$$

and the necessary condition stated in Proposition 3 follows by definition.

S7 Proof of Theorem 2

Starting from the KKT conditions of Lemma 2, we will redo the steps of the proof of Theorem 1, but with the insertion of extra terms representing the correction for measurement error. The corrected lasso is not in general convex, and our analysis will thus concern *any* critical point $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$ in the interior of the feasible set $\{\boldsymbol{\gamma} : \|\boldsymbol{\gamma} + \boldsymbol{\beta}^0\| < R\}$.

If $\hat{\boldsymbol{\gamma}}$ exists, and

$$\begin{aligned}&-\frac{\mathbf{W}'_{S_0}}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n}\left(\mathbf{C}_{ww}\left(S_0, S_0\right) - \boldsymbol{\Sigma}_{uu}\left(S_0, S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0} + \\ &\quad\sqrt{n}\left(C_{wu}\left(S_0, S_0\right) - \boldsymbol{\Sigma}_{uu}\left(S_0, S_0\right)\right)\boldsymbol{\beta}^0_{S_0} = -\frac{\lambda\sqrt{n}}{2}\text{sign}\left(\boldsymbol{\beta}^0_{S_0}\right),\end{aligned} \tag{S7.1}$$

$$\left|\hat{\boldsymbol{\gamma}}_{S_0}\right| < \left|\boldsymbol{\beta}^0_{S_0}\right|, \tag{S7.2}$$

$$\begin{aligned}&\Bigg| -\frac{W'_{S_0^c}}{\sqrt{n}}\boldsymbol{\epsilon} + \sqrt{n}\left(\mathbf{C}_{ww}\left(S_0^c, S_0\right) - \boldsymbol{\Sigma}_{uu}\left(S_0^c, S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0} + \\ &\quad\sqrt{n}\left(\mathbf{C}_{wu}\left(S_0^c, S_0\right) - \boldsymbol{\Sigma}_{uu}\left(S_0^c, S_0\right)\right)\boldsymbol{\beta}^0_{S_0}\Bigg| \leq \frac{\lambda\sqrt{n}}{2}\mathbf{1},\end{aligned} \tag{S7.3}$$

then $\text{sign}(\hat{\boldsymbol{\beta}}_{S_0}) = \text{sign}(\boldsymbol{\beta}^0_{S_0})$ and $\text{sign}(\hat{\boldsymbol{\beta}}_{S_0}) = \mathbf{0}$.

Event A in Theorem 2 implies the existence of $|\hat{\boldsymbol{\gamma}}_{S_0}| < |\boldsymbol{\beta}^0_{S_0}|$ such that

$$\left|\mathbf{Z}_6 - \mathbf{Z}_7\boldsymbol{\beta}_{S_0}\right| = \sqrt{n}\left(\left|\hat{\boldsymbol{\gamma}}_{S_0}\right| - \frac{\lambda}{2}\left|\left(\mathbf{C}_{ww}\left(S_0, S_0\right) - \boldsymbol{\Sigma}_{uu}\left(S_0, S_0\right)\right)^{-1}\text{sign}\left(\boldsymbol{\beta}^0_{S_0}\right)\right|\right).$$

But then there must exist $|\hat{\boldsymbol{\gamma}}_{S_0}| < |\boldsymbol{\beta}^0_{S_0}|$ such that

$$\mathbf{Z}_6 - \mathbf{Z}_7\boldsymbol{\beta}_{S_0} = \sqrt{n}\left(\hat{\boldsymbol{\gamma}}_{S_0} - \frac{\lambda}{2}\left(\mathbf{C}_{ww}\left(S_0, S_0\right) - \boldsymbol{\Sigma}_{uu}\left(S_0, S_0\right)\right)^{-1}\text{sign}\left(\boldsymbol{\beta}^0_{S_0}\right)\right).$$

Multiplying through by $\mathbf{C}_{ww}(S_0,S_0)-\mathbf{\Sigma}_{uu}(S_0,S_0)$ and reorganizing terms, we get (S7.1). Thus, A ensures that (S7.1) and (S7.2) are satisfied. Next, adding and subtracting $\sqrt{n}\left(\mathbf{C}_{ww}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0}$ to the left-hand side of event B and the using the triangle inequality, yields

$$\left|-\frac{\mathbf{W}_{S_0^c}'}{\sqrt{n}}\boldsymbol{\epsilon}+\sqrt{n}\left(\left(\mathbf{C}_{wu}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0}+\left(\mathbf{C}_{wu}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\boldsymbol{\beta}_{S_0}^0\right)\right|$$
$$\left|-\left(\mathbf{C}_{ww}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\left(\mathbf{C}_{ww}\left(S_0,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0,S_0\right)\right)^{-1}\frac{\mathbf{W}_{S_0}'}{\sqrt{n}}+\right.$$
$$\sqrt{n}\left(\mathbf{C}_{ww}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\left(\mathbf{S}_{ww}\left(S_0,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0,S_0\right)\right)^{-1}$$
$$\left.\left(\mathbf{C}_{wu}\left(S_0,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0,S_0\right)\right)\boldsymbol{\beta}_{S_0}^0+\sqrt{n}\left(\mathbf{S}_{ww}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0}\right|$$
$$\leq\frac{\lambda\sqrt{n}}{2}\left(1-\theta\right)\mathbf{1}.$$

The second term on the left-hand side of this expression is the left-hand side of (S7.1) multiplied by $\left(\mathbf{C}_{ww}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\left(\mathbf{C}_{ww}\left(S_0,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0,S_0\right)\right)^{-1}$. It can thus be replaced by the right-hand side of (S7.1) multiplied by this factor. This yields

$$\left|-\frac{\mathbf{W}_{S_0^c}'}{\sqrt{n}}\boldsymbol{\epsilon}+\sqrt{n}\left(\left(\mathbf{C}_{wu}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0}+\left(\mathbf{C}_{wu}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\boldsymbol{\beta}_{S_0}^0\right)\right|$$
$$\left|\frac{\lambda\sqrt{n}}{2}\left(\mathbf{C}_{ww}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\left(\mathbf{C}_{ww}\left(S_0,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0,S_0\right)\right)^{-1}\operatorname{sign}\left(\boldsymbol{\beta}_{S_0}^0\right)\right|$$
$$\leq\frac{\lambda\sqrt{n}}{2}\left(1-\theta\right)\mathbf{1},$$

which implies, due to the IC-CL,

$$\left|-\frac{\mathbf{W}_{S_0^c}'}{\sqrt{n}}\boldsymbol{\epsilon}+\sqrt{n}\left(\mathbf{C}_{wu}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\hat{\boldsymbol{\gamma}}_{S_0}\right.$$
$$\left.+\sqrt{n}\left(\mathbf{C}_{wu}\left(S_0^c,S_0\right)-\mathbf{\Sigma}_{uu}\left(S_0^c,S_0\right)\right)\boldsymbol{\beta}_{S_0}^0\right|\leq\frac{\lambda\sqrt{n}}{2}\mathbf{1}.$$

This is indeed (S7.3). Altogether, A implies (S7.1) and (S7.2), while $B|A$ implies (S7.3).

For the asymptotic result, define the vectors

$$\mathbf{z}=\left(\mathbf{C}_{ww}(S_0,S_0)-\mathbf{\Sigma}_{uu}\left(S_0,S_0\right)\right)^{-1}\frac{\mathbf{W}_{S_0}'}{\sqrt{n}}\boldsymbol{\epsilon},$$
$$\mathbf{a}=\left|\boldsymbol{\beta}_{S_0}^0\right|-\left|\left(\mathbf{C}_{ww}(S_0,S_0)-\mathbf{\Sigma}_{uu}(S_0,S_0)\right)^{-1}\left(\mathbf{C}_{wu}(S_0,S_0)-\mathbf{\Sigma}_{uu}(S_0,S_0)\right)\boldsymbol{\beta}_{S_0}^0\right|,$$
$$\mathbf{b}=\left(\mathbf{C}_{ww}(S_0,S_0)-\mathbf{\Sigma}_{uu}(S_0,S_0)\right)^{-1}\operatorname{sign}\left(\boldsymbol{\beta}_{S_0}^0\right),$$
$$\boldsymbol{\zeta}=\left(\left(\mathbf{C}_{ww}(S_0^c,S_0)-\mathbf{\Sigma}_{uu}(S_0^c,S_0)\right)\left(\mathbf{C}_{ww}(S_0,S_0)-\mathbf{\Sigma}_{uu}(S_0,S_0)\right)^{-1}\frac{\mathbf{W}_{S_0}'}{\sqrt{n}}-\frac{\mathbf{W}_{S_0^c}'}{\sqrt{n}}\right)\boldsymbol{\epsilon},$$

$$\mathbf{f} = \Big(\left(\mathbf{C}_{ww}(S_0^c, S_0) - \boldsymbol{\Sigma}_{uu}(S_0^c, S_0)\right) \left(\mathbf{C}_{ww}(S_0, S_0) - \boldsymbol{\Sigma}_{uu}(S_0, S_0)\right)^{-1} \left(\mathbf{C}_{wu}(S_0, S_0) - \boldsymbol{\Sigma}_{uu}(S_0, S_0)\right) - \left(\mathbf{C}_{wu}(S_0^c, S_0) - \boldsymbol{\Sigma}_{uu}(S_0^c, S_0)\right) \Big) \boldsymbol{\beta}_{S_0}^0.$$

We have

$$1 - P(A \cap B) \leq P(A^c) + P(B^c) \leq \sum_{j=1}^{s_0} P\left(|z_j| \geq \sqrt{n}\left(a_j - \frac{\lambda}{2} b_j\right)\right) + \sum_{j=1}^{p-s_0} P\left(\left|\zeta_j - \sqrt{n} f_j\right| \geq \frac{\lambda\sqrt{n}}{2}(1-\theta)\right).$$

It is clear that

$$\mathbf{z} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_{xx}(S_0, S_0)^{-1} \boldsymbol{\Sigma}_{ww}(S_0, S_0) \boldsymbol{\Sigma}_{xx}(S_0, S_0)^{-1}\right), \text{ as } n \to \infty.$$

Hence, there exists a finite constant k such that $E(z_j)^2 < k^2$ for $j = 1, \ldots, s_0$. Next, we have by assumption $\mathbf{a} \to \left|\boldsymbol{\beta}_{S_0}^0\right|$, as $n \to \infty$. Now using the assumption $\lambda = o(1)$, we get

$$\begin{aligned} P\left(A^c\right) &\leq \sum_{j=1}^{s_0} \left(1 - P\left(\frac{|z_j|}{k} < \frac{\sqrt{n}}{2k} a_j \left(1 + o(1)\right)\right)\right) \\ &\leq (1 + o(1)) \sum_{j=1}^{s_0} \left(1 - \Phi\left(\frac{\sqrt{n}}{2s} a_j \left(1 + o(1)\right)\right)\right) \\ &= o\left(\exp(-n^c)\right), \end{aligned}$$

where we used the bound (S5.4). Next, we note that

$$\zeta \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \left(\boldsymbol{\Sigma}_{xx}(S_0^c, S_0^c) - \boldsymbol{\Sigma}_{xx}(S_0^c, S_0) \boldsymbol{\Sigma}_{xx}(S_0, S_0)^{-1} \boldsymbol{\Sigma}_{xx}(S_0, S_0^c)\right)\right), \text{ as } n \to \infty.$$

Next, we consider $\mathbf{f}$, and note that the limiting distribution of $\sqrt{n}\left(\mathbf{C}_{wu} - \boldsymbol{\Sigma}_{uu}\right)$, as $n \to \infty$, is normal with mean $\mathbf{0}$ and finite variances (Anderson (2003, Th. 3.4.4)). In addition, $\mathbf{C}_{ww} - \boldsymbol{\Sigma}_{uu} \to \boldsymbol{\Sigma}_{xx}$, as $n \to \infty$. Thus, applying Slutsky's theorem to the product of the matrices, the limiting distribution of

$$\sqrt{n}\Big(\left(\mathbf{C}_{ww}(S_0^c, S_0) - \boldsymbol{\Sigma}_{uu}(S_0^c, S_0)\right) \left(\mathbf{C}_{ww}(S_0, S_0) - \boldsymbol{\Sigma}_{uu}(S_0, S_0)\right)^{-1} \left(\mathbf{C}_{wu}(S_0, S_0) - \boldsymbol{\Sigma}_{uu}(S_0, S_0)\right) - \left(\mathbf{C}_{wu}(S_0^c, S_0) - \boldsymbol{\Sigma}_{uu}(S_0^c, S_0)\right) \Big), \text{ as } n \to \infty,$$

is normal with mean $\mathbf{0}$ and finite variances. Now,

$$\sqrt{n}\mathbf{f} = \sqrt{n}\Big(\left(\mathbf{C}_{ww}(S_0^c, S_0) - \boldsymbol{\Sigma}_{uu}(S_0^c, S_0)\right) \left(\mathbf{C}_{ww}(S_0, S_0) - \boldsymbol{\Sigma}_{uu}(S_0, S_0)\right)^{-1} \left(\mathbf{C}_{wu}(S_0, S_0) - \boldsymbol{\Sigma}_{uu}(S_0, S_0)\right) - \left(\mathbf{C}_{wu}(S_0^c, S_0) - \boldsymbol{\Sigma}_{uu}(S_0^c, S_0)\right) \Big) \boldsymbol{\beta}_{S_0}^0$$

is a vector in $\mathbb{R}^{p-s_0}$ whose elements are linear combinations of variables whose limiting distributions as $n \to \infty$ are normal with mean zero and finite variances. Accordingly, the limiting distribution of $\sqrt{n}\mathbf{f}$ as $n \to \infty$ is normal with mean zero and finite variances.

So again there exists a finite constant k such that $E(\zeta_j - \sqrt{n} f_j)^2 < k^2$ for $j = 1, \ldots, (p - s_0)$. Thus, when $\lambda n^{(1-c)/2} \to \infty$ for $c \in [0, 1)$, we have

$$\begin{aligned} P(B^c) &\leq \sum_{j=1}^{p-s_0} \left(1 - P\left(\frac{|\zeta_j - \sqrt{n} f_j|}{k} < \frac{1}{k}\frac{\lambda\sqrt{n}}{2}(1-\theta)\right)\right) \\ &\leq (1 + o(1)) \sum_{j=1}^{p-s_0} \left(1 - \Phi\left(\frac{1}{k}\frac{\lambda\sqrt{n}}{2}(1-\theta)\right)\right) \\ &= o\left(\exp(-n^c)\right). \end{aligned}$$

It follows that $P(A \cap B) = 1 - o(\exp(-n^c))$.

Additional References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, Third Edition.* John Wiley and Sons, Hoboken.