Supplemental material for adaptive inference after model selection

Eric B. Laber[1] and Susan A. Murphy[2]

[1]Department of Statistics, North Carolina State University, Raleigh, NC, 27695, U.S.A.

[2] Department of Statistics, University of Michigan, Ann Arbor, MI, 48019, U.S.A.

1 Appendix

1.1 Local inconsistency of PBS

Here we briefly discuss the PBS estimator and show that it need not be consistent under local alternatives. As in the main body of the article we consider the consistent tuning case. For simplicity we will consider the one-dimensional setting. In particular, we will assume that for each n that the data form a triangular array

$$y_{n,i} = x_i\beta_n^* + \epsilon_{n,i},$$

where $\beta_n^* = s_n/\sqrt{n}$ for some sequence of random variables $\{s_n\}$ converging almost surely to bounded random variable s. In addition, let $g_1, g_2, \ldots, g_n$ denote non-negative independent identically distributed random variables with unit mean and variance; we assume these random variables are jointly independent of the $\epsilon_{n,i}$'s for all n and i. We will also assume that $\frac{1}{n}\sum_{i=1}^n x_i \to 1$. The perturbation estimator is formed in two steps:

1. Compute the perturbed OLS estimator $\tilde{\beta}_n^0 \triangleq \arg\min_{\beta\in\mathbb{R}} \sum_{i=1}^n g_i(y_i - x_i\beta)^2 = \frac{\sum_{i=1}^n g_i x_i y_i}{\sum_{i=1}^n g_i x_i^2}$.

2. Compute the Adaptive LASSO perturbed estimator

$$\tilde{\beta}_n^{\lambda_n} \triangleq \arg\min_{\beta\in\mathbb{R}} \sum_{i=1}^n g_i(y_i - x_i\beta)^2 + \lambda_n|\beta|/|\tilde{\beta}_n^0| = \tilde{\beta}_n^0\left(1 - \frac{\lambda_n}{2|\tilde{\beta}_n^0|^2\sum_{i=1}^n g_i x_i^2}\right)_+ .$$

The PBS estimator uses the conditional distribution of $\sqrt{n}(\tilde{\beta}_n^{\lambda_n} - \hat{\beta}_n^{\lambda_n})$ to approximate the unconditional distribution of $\sqrt{n}(\hat{\beta}_n^{\lambda_n} - \beta_n^*)$. Some algebra (see Pötscher and Schneider, 2009, for details) shows that

$$\sqrt{n}(\hat{\beta}_n^{\lambda_n} - \beta_n^*) = (\mathbb{Z}_n + s_n)\left(1 - \frac{\lambda_n}{2(\mathbb{Z}_n + s_n)^2}\right)_+ - s_n \rightsquigarrow (\mathbb{Z}_\infty + s)\left(1 - \frac{\lambda_0}{2(\mathbb{Z}_\infty + s)^2}\right)_+ - s.$$

We now derive the limiting distribution of the PBS estimator $\sqrt{n}(\tilde{\beta}_n^{\lambda_n} - \hat{\beta}_n^{\lambda_n})$. First, we use a multiplier central limit theorem to argue that $\sqrt{n}(\tilde{\beta}_n^0 - \hat{\beta}_n^0)$, which we denote by $\tilde{\mathbb{Z}}_n$ converges conditionally in distribution to $\mathbb{Z}_\infty$. To see this note that

$$\sqrt{n}(\tilde{\beta}_n^0 - \hat{\beta}_n^0) = \frac{\sum_{g=1}^n g_i y_{n,i} x_i}{\sum_{i=1}^n g_i x_i^2} - \sqrt{n}\hat{\beta}_n^0 = \frac{\sum_{i=1}^n (g_i - 1) y_{n,i} x_i}{\sum_{i=1}^n g_i x_i^2} + \sqrt{n}\hat{\beta}_n^0 \frac{\sum_{i=1}^n (1 - g_i) x_i^2}{\sum_{i=1}^n g_i x_i^2}.$$

The last term on the right hand side of the above display equals $(\mathbb{Z}_n + s_n)\frac{\frac{1}{n}\sum_{i=1}^n (1-g_i)x_i^2}{\frac{1}{n}\sum_{i=1}^n g_i x_i^2}$ which converges conditionally to zero in probability (see Van Der Vaart and Wellner, 1996, Chapter 2.9 for details). Thus, using the conditional multiplier central limit theorem we have $\sqrt{n}(\tilde{\beta}_n^0 - \hat{\beta}_n^0) \rightsquigarrow \lim_{n\to\infty} \frac{\sum_{i=1}^n (g_i - 1) y_{n,i} x_i}{\sum_{i=1}^n g_i x_i} =_D \mathbb{Z}_\infty$.

Note that conditional on the observed data $\mathbb{Z}_n$ does not converge to a limit (it is a rather a non-convergent sequence of constants). We can write $\sqrt{n}(\tilde{\beta}_n^{\lambda_n} - \hat{\beta}_n^{\lambda_n})$ as

$$= \sqrt{n}\tilde{\beta}_n^0\left(1 - \frac{\lambda_n}{2(\sum_{i=1}^n g_i x_i^2)|\tilde{\beta}_n^0|^2}\right)_+ - \sqrt{n}\hat{\beta}_n^0\left(1 - \frac{\lambda_n}{2(\sum_{i=1}^n x_i^2)|\hat{\beta}_n^0|^2}\right)_+$$

$$= \tilde{\mathbb{Z}}_n\left(1 - \frac{\lambda_n}{2(n^{-1}\sum_{i=1} g_i x_i^2)|\tilde{\mathbb{Z}}_n + \mathbb{Z}_n + s_n|^2}\right)_+$$

$$-(\mathbb{Z}_n + s_n)\left[\left(1 - \frac{\lambda_n}{2(n^{-1}\sum_{i=1}^n x_i^2)|\mathbb{Z}_n + s_n|^2}\right)_+ - \left(1 - \frac{\lambda_n}{2(n^{-1}\sum_{i=1}^n g_i x_i^2)|\tilde{\mathbb{Z}} + \mathbb{Z}_n + s_n|^2}\right)_+\right],$$

which does not have a limit conditional on the data.

1.2 Proofs

Proof of Theorem 3.1. The existence and values of the roots r_1 and r_4 follow from applying the quadratic formula. Note that if $r_1 = \arg\sup_{s\in\mathbb{R}} g(s)$ and $r_4 = \arg\inf_{s\in\mathbb{R}} g(s)$, then because these are roots of $q(s) = \lambda_n^{(b)}$ it follows immediately that $g(r_1) = -r_1$ and $g(r_4) = -r_4$. The sequence of results we use to establish that $r_1 = \arg\sup_{s\in\mathbb{R}} g(s)$ and $r_4 = \arg\inf_{s\in\mathbb{R}}$ are as follows: (i) $g(s)$ is monotone increasing on $(-\infty, r_1] \cup [r_4, \infty)$; (ii) $g(s)$ is monotone decreasing on (r_1, r_4); and (iii) $\lim_{s\to-\infty} g(s) = \lim_{s\to\infty} g(s) = \mathbb{Z}_n^{(b)}$. Since $g(s)$ is continuous the foregoing three properties ensure that the maximum occurs at r_1 while the minimum occurs at r_4.

To see that $g(s)$ is monotone on $(-\infty, r_1]$ notice that because r_1 is the smallest root of $q(s) = \lambda_n^{(b)}$ and $q(s)$ tends to ∞ as $s \to -\infty$ we have $q(s) > \lambda_n^{(b)}$ on $(-\infty, r_1)$. Thus, $g(s) = \mathbb{Z}_n^{(b)} - \frac{\lambda_n^{(b)} \text{sgn}(\mathbb{Z}_n^{(b)}+s)}{2|\mathbb{W}_n^{(b)}+s|}$ on $(-\infty, r_1)$. Furthermore, because $\lambda_n^{(b)} > 0$ we have

$$r_1 < -\frac{1}{2}(\mathbb{W}_n^{(b)} + \mathbb{Z}_n^{(b)}) - \frac{1}{2}|\mathbb{W}_n^{(b)} - \mathbb{Z}_n^{(b)}| = \min(-\mathbb{Z}_n^{(b)}, -\mathbb{W}_n^{(b)}),$$

from whence it follows that $|\mathbb{W}_n^{(b)} + s|$ is monotone decreasing and $\text{sgn}(\mathbb{Z}_n^{(b)} + s) \equiv -1$ on $(-\infty, r_1]$. Subsequently $g(s) = \mathbb{Z}_n^{(b)} + \frac{\lambda_n^{(b)}}{2|\mathbb{W}_n^{(b)}+s|}$ is monotone increasing on $(-\infty, r_1)$. Analogous arguments show that $g(s) = \mathbb{Z}_n^{(b)} - \frac{\lambda_n^{(b)}}{2|\mathbb{W}_n^{(b)}+s|}$ is also monotone increasing on (r_4, ∞).

Note that for any $s, s' \in \mathbb{R}$ such that $s \le r_1$ and $s' \ge r_4$ we must have $g(s) > g(s')$. If r_1 and r_4 are the only real roots then it follows that $q(s) \le \lambda_n$ on $[r_1, r_4]$ and $g(s) = -s$ on this interval. On the other hand, if there are four real roots $r_1 < r_2 < r_3 < r_4$ (note that with probability one there are either two or four distinct real roots), then we must have $q(s) \le \lambda_n^{(b)}$ and hence $g(s) = -s$ on $[r_1, r_2] \cup [r_3, r_4]$. Thus, on (r_2, r_3) we must have $q(s) > \lambda_n^{(b)}$ and we can write $g(s)$ as

$$g(s) = (\mathbb{Z}_n^{(b)} + s)\left(1 - \frac{\lambda_n^{(b)}}{q(s)}\right)_+ - (\mathbb{Z}_n^{(b)} + s) + \mathbb{Z}_n^{(b)},$$

and since on (r_2, r_3) the first term in the above display is increasing at a slower rate than the second term is decreasing, it follows that $g(s)$ is monotone decreasing on (r_2, r_3). Combining the above pieces shows that (i) $g(s)$ is increasing on $(-\infty, r_1]$ and $[r_4, \infty)$ with $g(r_1) > g(s)$ for all $s \in (-\infty, r_1) \cup [r_4, \infty)$; (ii) $g(s)$ monotone decreasing between r_1 and r_4; and (iii) $\lim_{s\to-\infty} g(s) = \lim_{s\to+\infty} g(s) = \mathbb{Z}_n^{(b)}$ follows by inspection. This proves the result. □

Proof of Theorem 4.1. Part 1 of the theorem follows immediately from the requirement that $\beta^* \in \mathcal{S}_n$. The proof of part 2 of the theorem can be found in Zou (2006). Define

$$\tilde{\mathbb{V}}_n(t; s) = -2t^\intercal \mathbb{Z}_n + t^\intercal C_n t$$
$$+ \lambda_n \sum_{j=1}^{p} \left(\frac{|\sqrt{n}\beta_j^* + t_j| - |\sqrt{n}\beta_j^*|}{|\mathbb{W}_{n,j} + \sqrt{n}\beta_j^*|} 1_{\sqrt{n}|\hat{\beta}^0_{n,j}| > \hat{\sigma}_n \zeta_{n,j} \tau_{n,j}} + \frac{|s_j + t_j| - |s_j|}{|\mathbb{W}_{n,j} + s_j|} 1_{\sqrt{n}|\hat{\beta}^0_{n,j}| \le \hat{\sigma}_n \zeta_{n,j} \tau_{n,j}} \right), \quad (1)$$

and notice that $\tilde{\mathbb{V}}_n(t; \sqrt{n}s) = \mathbb{V}_n(t; s)$ for all $s \in \mathcal{S}_n$. Furthermore, for a fixed vector $\mathbf{c} \in \mathbb{R}^p$, it follows that

$$\sup_{s\in\mathcal{S}_n} \mathbf{c}^\intercal \arg\min_{t\in\mathbb{R}^p} \mathbb{V}_n(t; s) = \sup_{s\in\mathcal{S}_n} \mathbf{c}^\intercal \arg\min_{t\in\mathbb{R}^p} \tilde{\mathbb{V}}_n(t; \sqrt{n}s) = \sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal \arg\min_{t\in\mathbb{R}^p} \tilde{\mathbf{V}}_n(t; s).$$

Define

$$f(t, s, \gamma, \Sigma, \alpha) \triangleq -2t^\intercal \gamma + t^\intercal \Sigma t + \alpha \sum_{j=1}^{p} \frac{|s_j + t_j|}{|(\Sigma^{-1}\gamma)_j + s_j|} 1_{\beta_j^* = 0},$$

where $t, s, \gamma \in \mathbb{R}^p$, Σ is a $p \times p$ positive definite matrix, and $\alpha \in \mathbb{R}_+$. We view f as a function taking values on the positive extended real line $\mathbb{R} \cup \{+\infty\}$. The domain of f is thus given by

$$\operatorname{dom} f = \left\{ t, s, \gamma \in \mathbb{R}^p,\, \Sigma \in PD(p,p),\, \alpha \in \mathbb{R}_+ \ :\ \text{if } s_j = -(\Sigma^{-1}\gamma)_j \text{ then } \ t_j = (\Sigma_j^{-1}\gamma)_j \right\}.$$

An outline of the remainder of the proof is as follows: (i) we show that $g(s, \gamma, \Sigma, \alpha) \triangleq$

$\arg\min_{t\in\mathbb{R}^p} f(t,s,\gamma,\Sigma,\alpha)$ is bounded and continuous in all its arguments in a neighborhood of $\Sigma = C$ and $\alpha = \lambda_0$; (ii) we show that in studying the limiting behavior of the arg min of $f(t,s,\gamma,\Sigma,\alpha)$ it suffices to restrict attention to s and γ in compact subsets of $\mathbb{R}^p$; (iii) we next argue that $\arg\min_{t\in\mathbb{R}^p} \tilde{\mathbb{V}}_n(t;s) = g(s,\mathbb{Z}_n,C_n,\lambda_n) + r_1(s,\mathbb{Z}_n,C_n,\lambda_n)$, where $r_1(s,\gamma,\Sigma,\alpha)$ converges to zero in probability uniformly over s and γ in compact subsets of $\mathbb{R}^p$ and with Σ and α in neighborhoods of C and λ_n respectively; (iv) finally we use the equivalence between $\tilde{\mathbb{V}}_n$ and $\mathbb{V}_n$ established above combined with the continuous mapping theorem to conclude that

$$\sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal \arg\min_{t\in\mathbb{R}^p} \mathbb{V}_n(s;t) = \sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal g(s,\mathbb{Z}_n,C_n,\lambda_n) + r_1 \rightsquigarrow \sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal \arg\min_{t\in\mathbb{R}^p} \mathbb{V}_{\infty,\beta^*}(t;s).$$

Consider part (i). If γ is restricted to a bounded subset of $\mathbb{R}^p$, then we can restrict attention to t in a bounded set as follows. Note that $f(t,s,\gamma,\Sigma,\alpha) \geq -2t^\intercal\gamma + t^\intercal\Sigma t$ for all Σ, γ, and $\alpha > 0$. In addition, $f(\Sigma^{-1}\gamma,s,\gamma,\Sigma,\alpha) = -\gamma^\intercal\Sigma^{-1}\gamma + \alpha\#\{j : \beta_j^* = 0\} \leq -\gamma^\intercal\Sigma^{-1}\gamma + \alpha p$ (define $0/0 = 1$). Thus if Σ is restricted to live in a sufficiently small neighborhood of C so that it is assured to have eigenvalues bounded away from zero and ∞, α is restricted to a neighborhood of λ_0 bounded above zero and γ is restricted to a bounded subset of $\mathbb{R}^p$, then $g(s,\gamma,\Sigma,\alpha)$ is uniformly bounded. Next consider a j for which $\beta_j^* = 0$. Suppose $(s_m,\gamma_m,\Sigma_m,\alpha_m)$ is a sequence converging as $m \to \infty$ to (s,γ,Σ,α) where Σ is positive definite, α is finite, positive, s is finite and for which $s_j = -(\Sigma^{-1}\gamma)_j$. Then for $\delta > 0$ arbitrary, and sufficiently large m, $|(\Sigma_m^{-1}\gamma_m)_j + s_{m,j}| \leq \delta/(2p)$. If for any of these m, and any t we have $|t_j + s_{m,j}| > \delta$ then $\alpha_m|t_j + s_{m,j}|/|(\Sigma_m^{-1}\gamma_m)_j + s_{m,j}| > \alpha_m p \geq f(\Sigma_m^{-1}\gamma_m, s_m, \gamma_m, \Sigma_m, \alpha_m)$. Such a t is ruled out as a value for $g(s_m,\gamma_m,\Sigma_m,\alpha_m)$. Since δ is arbitrary, it follows that $[g(s_m,\gamma_m,\Sigma_m,\alpha_m)]_j$ converges to $(\Sigma^{-1}\gamma)_j$. We have continuity of g at points at which for one or more j, $\beta_j^* = 0$ and $s_j = -(\Sigma^{-1}\gamma)_j$.

The lemma immediately following this proof shows that g is continuous at the remaining points. The continuous mapping theorem shows $\sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal g(s, \mathbb{Z}_n, C_n, \lambda_n) \rightsquigarrow \sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal g(s, \mathbf{Z}_\infty, C, \lambda_0)$. Notice that $g(s, \mathbb{Z}_\infty, C, \lambda_0) = \arg\min_{t\in\mathbb{R}^p} \mathbb{V}_{\infty,\beta^*}(t; s)$.

We now argue that $\arg\min_{t\in\mathbb{R}^p} \tilde{V}_n(t; s) = g(s, \mathbb{Z}_n, C_n, \lambda_n) + r_1(s, \mathbb{Z}_n, C_n, \lambda_n)$ and that r_1 can be neglected in asymptotic arguments. Suppose that $\beta_j^* \neq 0$ then the event $\sqrt{n}|\hat{\beta}^0_{n,j}| \leq \hat{\sigma}_n\zeta_{n,j}\tau_{n,j}$ is contained in the event

$$|\beta_j^*| \leq \frac{\tau_{n,j}}{\sqrt{n}}\zeta_{n,j}\hat{\sigma}_n + |\hat{\beta}^0_{n,j} - \beta_j^*|,$$

where the right hand side is seen to be converging to zero in probability, hence $1_{\sqrt{n}|\hat{\beta}^0_n| \leq \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n} 1_{\beta_j^*\neq 0} = 0$ with probability tending to one. Writing $1_{\sqrt{n}|\hat{\beta}^0_n| > \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n} = 1 - 1_{\sqrt{n}|\hat{\beta}^0_{n,j}| \leq \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n}$ we see that $1_{\sqrt{n}|\hat{\beta}^0_n| \leq \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n} 1_{\beta_j^*\neq 0}$ equals $1_{\beta_j^*\neq 0}$ with probability tending to one. On the other hand, if $\beta_j^* = 0$ then the event $\sqrt{n}|\hat{\beta}^0_j| \leq \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n$ equals the event

$$|\sqrt{n}(\hat{\beta}^0_{n,j} - \beta_j^*)| \leq \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n,$$

where the right hand side of the above display tends to ∞ with probability tending to one (e.g., for any $M > 0$ it follows $P(\tau_{n,j}\zeta_{n,j}\hat{\sigma}_n > M) \to 1$) thus $1_{\sqrt{n}|\hat{\beta}^0_n| \leq \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n} 1_{\beta_j^* = 0}$ equals $1_{\beta_j^*=0}$ with probability tending to one. Thereby it also follows that $1_{\sqrt{n}|\hat{\beta}^0_{n,j}| > \tau_{n,j}\zeta_{n,j}\hat{\sigma}_n} 1_{\beta_j^*=0} = 0$ with probability tending to one. With probability tending to one it follows that $\tilde{\mathbb{V}}_n(t; s) = f(t, s, \mathbb{Z}_n, C_n, \lambda_n)$ and thus $r_1(s, t, \mathbb{Z}_n, C_n, \lambda_n) \triangleq g(s, t, \mathbb{Z}_n, C_n, \lambda_n) - \arg\min_{t\in\mathbb{R}^p} \tilde{\mathbb{V}}_n(t; s)$ converges to zero in probability uniformly over s and t where t is restricted to live in a compact subset of $\mathbb{R}^p$.

The proof is concluded by means of the equivalence between $\tilde{\mathbb{V}}_n(t; s)$ and $\mathbb{V}_n(t; s)$ as discussed in step (iv) above.

□

Lemma 1.1. *Using the notation and assumptions of the preceding proof,* $g(s, \gamma, \Sigma, \alpha)$ *is* *continuous at all points* $(s, \gamma, \Sigma, \alpha)$ *for which* Σ *is positive definite,* α *is positive and for all* j *for which* $\beta_j^* = 0$, $s_j \neq -(\Sigma^{-1}\gamma)_j$.

Proof. It will be convenient to work with $\tilde{f}(t, s, \gamma, \Sigma, \alpha) \triangleq f(t, s, \gamma, \Sigma, \alpha) + \gamma^\intercal \Sigma^{-1} \gamma$. Since the added term does not depend on t it follows that $g(s, \gamma, \Sigma, \alpha) = \arg\min_t \tilde{f}(t, s, \gamma, \Sigma, \alpha)$. Let $\epsilon \in [0, 1]$ be arbitrary, we first show that for any t, t'

$$f(\epsilon t + (1-\epsilon)t', s, \gamma, \Sigma, \alpha) \leq \epsilon f(t, s, \gamma, \Sigma, \alpha) + (1-\epsilon) f(t', s, \gamma, \Sigma, \alpha) - \epsilon(1-\epsilon) m ||t - t'||^2, \quad (2)$$

where m is a constant bounded below by the smallest eigenvalue of Σ. To see this, note that $\tilde{f}(t, s, \gamma, \Sigma, \alpha)$ can be written as

$$h(t, \gamma, \Sigma) + \alpha \sum_{j=1}^{p} \frac{|t_j + s_j|}{|(\Sigma^{-1}\gamma)_j + s_j|} 1_{\beta_j^* = 0}, \quad (3)$$

where $h(t, \gamma, \Sigma) = (t - \gamma^\intercal \Sigma^{-1})^\intercal \Sigma (t - \gamma^\intercal \Sigma^{-1})$. Hence, $h(\epsilon t + (1-\epsilon)t', \gamma, \Sigma)$ equals

$$\epsilon^2 h(t, \gamma, \Sigma) + (1-\epsilon)^2 h(t', \gamma, \Sigma) + (1-\epsilon)^2 h(t', \gamma, \Sigma) + 2\epsilon(1-\epsilon)(t - \gamma^\intercal \Sigma^{-1})^\intercal \Sigma (t' - \gamma^\intercal \Sigma^{-1}). \quad (4)$$

Furthermore, after some algebra we have

$$\begin{aligned} (t - \gamma^\intercal \Sigma^{-1})^\intercal \Sigma (t' - \gamma^\intercal \Sigma^{-1}) &= h(t', \gamma, \Sigma) + (t - t')^\intercal \Sigma (t' - \gamma^\intercal \Sigma^{-1}), \\ &= h(t, \gamma, \Sigma) + (t' - t)^\intercal \Sigma (t - \gamma^\intercal \Sigma^{-1}), \end{aligned}$$

adding down the right column of the above display yields,

$$2(t - \gamma^\intercal \Sigma^{-1})^\intercal \Sigma (t' - \gamma^\intercal \Sigma^{-1}) = h(t, \gamma, \Sigma) + h(t', \gamma, \Sigma) - (t - t')^\intercal \Sigma (t - t').$$

Thus, (4) is equal to

$$\epsilon h(t,\gamma,\Sigma)+(1-\epsilon)h(t',\gamma,\Sigma)-\epsilon(1-\epsilon)(t-t')^\intercal\Sigma(t-t'),$$

which is bounded above by

$$\epsilon h(t,\gamma,\Sigma)+(1-\epsilon)h(t',\gamma,\Sigma)-\epsilon(1-\epsilon)(\min\operatorname{eigen}(\Sigma))||t-t'||^2.$$

Since the second term in (3) is convex in t it follows that (2) is satisfied.

Now, let $t_{\min}$ denote $g(s,\gamma,\Sigma\alpha)$ and $t'_{\min}$ denote $g(s',\gamma',\Sigma',\alpha')$. Suppose first that $f(t'_{\min},s',\gamma',\Sigma',\alpha')\leq f(t_{\min},s,\gamma,\Sigma,\alpha)$. Then rearranging (2) we have

$$\epsilon(1-\epsilon)m||t_{\min}-t'_{\min}||^2\leq \epsilon f(t'_{\min},s,\gamma,\Sigma,\alpha)+(1-\epsilon)f(t_{\min},s,\gamma,\Sigma,\alpha)-f(\epsilon t'_{\min}+(1-\epsilon)t_{\min},s,\gamma,\Sigma,\alpha).$$

The right hand side of the above display is equal to

$$\begin{aligned}\epsilon\left[f(t'_{\min},s,\gamma,\Sigma,\alpha)-f(t'_{\min},s',\gamma',\Sigma',\alpha')\right]+\epsilon\left[f(t'_{\min},s',\gamma',\Sigma',\alpha')-f(t_{\min},s,\gamma,\Sigma,\alpha)\right]\\+\left[f(t_{\min},s,\gamma,\Sigma,\alpha)-f(\epsilon t'_{\min}+(1-\epsilon)t_{\min},s,\gamma,\Sigma,\alpha)\right],\end{aligned}$$

where the last two differences are nonpositive. Thus,

$$||t_{\min}-t'_{\min}||^2\leq\frac{1}{m(1-\epsilon)}\leq\left[f(t'_{\min},s,\gamma,\Sigma,\alpha)-f(t'_{\min},s',\gamma',\Sigma',\alpha')\right].$$

A parallel argument in the case that $f(t'_{\min},s',\gamma',\Sigma',\alpha')\geq f(t_{\min},s,\gamma,\Sigma,\alpha)$ shows that

$$||t_{\min}-t'_{\min}||^2\leq\frac{1}{m'(1-\epsilon)}\left[f(t_{\min},s',\gamma',\Sigma',\alpha')-f(t_{\min},s,\gamma,\Sigma,\alpha)\right],$$

where m' is larger than the minimum eigenvalue of Σ'. Thus, for Σ,Σ' in a sufficiently small neighborhood of C, the continuity of f implies the continuity of g. $\square$

Proof of Corollary 4.2. Suppose that $\beta_j^* \neq 0$ is fixed, then the event $\sqrt{n}|\hat\beta_{n,j}^0| > \tau_{n,j}\zeta_{n,j}\hat\sigma_n$ is contained in the event $\sqrt{n}|\beta_j^*| - \sqrt{n}|\hat\beta_{n,j}^0 - \beta_j^*| > \tau_{n,j}\zeta_{n,j}\hat\sigma_n$ which is equivalent to the event

$$|\beta_j^*| > \frac{\tau_{n,j}}{\sqrt{n}}\zeta_{n,j}\hat\sigma_n + |\hat\beta_{n,j} - \beta_j^*|,$$

which occurs eventually always with probability one if $\hat\sigma_n$ is strongly consistent or satisfies a log of the iterated logarithm (e.g., $\limsup_n \frac{\sqrt{n}(\hat\sigma_n - \sigma)}{\sqrt{\log\log n}}$ is bounded above almost surely). The usual plug-in estimator $\widehat\sigma_n^2 = n^{-1}\sum_{i=1}^n (y_i - \mathbf{x}_i^\intercal \widehat\beta_n^0)^2$ satisfies the foregoing conditions. $\square$

Proof of Theorem 4.3. To prove part 1 note that $\sqrt{n}(\hat\beta_n^{\lambda_n} - \beta_n^*)$ is the arg min over t of the localized process

$$\sum_{i=1}^n \left[(y_{n,i} - \mathbf{x}_i^\intercal(\beta_n^* + t/\sqrt{n}))^2 - (y_{n,i} - \mathbf{x}_i^\intercal\beta_n^*)^2\right] + \lambda_n \sum_{j=1}^p \frac{|\beta_{n,j}^* + t_j/\sqrt{n}| - |\beta_{n,j}^*|}{|\hat\beta_{n,j}^0|}. \qquad (5)$$

We rewrite the above localized process as

$$-\frac{2t^\intercal}{\sqrt{n}}\sum_{i=1}^n \epsilon_{n,i}\mathbf{x}_i + t^\intercal C_n t + \lambda_n \sum_{p=1}^j \frac{|\sqrt{n}\beta_j^* + s_{n,j} + t_j| - |\sqrt{n}\beta_j^* + s_{n,j}|}{|\sqrt{n}(\hat\beta_{n,j}^0 - \beta_{n,j}^*) + \sqrt{n}\beta_j^* + s_{n,j}|}.$$

Define $\tilde{\mathbb{Z}}_n \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_{n,i}\mathbf{x}_i$ and $\tilde{\mathbb{W}}_n \triangleq \sqrt{n}(\hat\beta_n^0 - \beta_n^*)$; using the regularity of the ordinary least squares estimator it follows that $\tilde{\mathbb{Z}}_n \rightsquigarrow \mathbb{Z}_\infty$ and $\tilde{\mathbb{W}}_n \rightsquigarrow C^{-1}\mathbb{Z}_\infty$. Thus, the foregoing localized process can be written as

$$-2t^\intercal\tilde{\mathbb{Z}}_n + t^\intercal C_n t + \lambda_n \sum_{j=1}^p \left[\frac{|s_{n,j} + t_j| - |s_{n,j}|}{|\tilde{\mathbb{W}}_{n,j} + s_{n,j}|}1_{\beta_j^*=0} + \frac{|\sqrt{n}\beta_j^* + s_{n,j} + t_j| - |\sqrt{n}\beta_j^* + s_{n,j}|}{|\sqrt{n}(\hat\beta_{n,j}^0 - \beta_{n,j}^*) + \sqrt{n}\beta_j^* + s_{n,j}|}1_{\beta_j^*\neq 0}\right],$$

which converges in distribution to

$$\mathbb{V}_{\infty,\beta^*}(t;s^*) = -2t^\intercal \mathbb{Z}_\infty + t^\intercal C t + \lambda_0 \sum_{j=1}^{p} \frac{|s_j^* + t_j| - |s_j^*|}{|\mathbb{W}_{\infty,j} + s_j^*|} 1_{\beta_j^*=0}.$$

The above convergence is uniform in a neighborhood of s^* (see the proof of Theorem 3.1) and all t in a compact set; furthermore the limiting process has a unique isolated minimizer over t at s^* and thus convergence of the arg min follows. Part 2 of this Theorem is established by the re-parametrization $s_j = \sqrt{n}\beta_j^* + s_{n,j}$ and applying the proof of Theorem 4.1. □

Proof of Theorem 4.4. The bootstrap analog $\mathbb{V}_n^{(b)}(t;s)$ of $\mathbb{V}_n(t;s)$ is given by

$$-2t^\intercal \mathbb{Z}_n^{(b)} + t^\intercal C_n t + \lambda_n^{(b)} \sum_{j=1}^{p} \frac{|\sqrt{n}s_j + t_j| - |\sqrt{n}s_j|}{|\mathbb{W}_{n,j}^{(b)} + \sqrt{n}s_j|}.$$

Notice that the only random components in the above expression are $\mathbb{Z}_n^{(b)}$, $\mathbb{W}_n^{(b)}$, and $\lambda_n^{(b)}$. Thus, provided that these quantities are (conditionally) asymptotically equal in distribution to their population analogs $\mathbb{Z}_n$, $\mathbb{W}_n$, and λ_n then the result follows immediately. However, this was proved by Chaterjee and Lahiri (2011) and so the desired result follows. □

Review of existing results for ordinary least squares.

Define $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^\intercal$ and let $\mathbb{X}_n$ denote the $n \times p$ design matrix with ith row $\mathbf{X}_i^\intercal$. Let $\mathbf{H}_n = \mathbb{X}_n(\mathbb{X}_n^\intercal \mathbb{X}_n)^{-1}\mathbb{X}_n^\intercal$ be the projection matrix onto the column space of $\mathbb{X}_n$. Then the residuals $\mathbf{e}_n = \mathbf{Y}_n - \mathbb{X}_n\widehat{\beta}_n^0 = (I - \mathbf{H}_n)\boldsymbol{\epsilon}_n$, where $\boldsymbol{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)^\intercal$. The estimated residual variance is $\widehat{\sigma}_n^2 = ||(I - \mathbf{H}_n)\boldsymbol{\epsilon}_n||^2/n = \boldsymbol{\epsilon}_n^\intercal(I - \mathbf{H}_n)\boldsymbol{\epsilon}_n/n$. Define $\widetilde{\boldsymbol{\epsilon}}_n = \boldsymbol{\epsilon}_n/\sigma$ so that the distribution of $\widetilde{\boldsymbol{\epsilon}}_n$ does not depend on either β^* or σ^2. Thus, neither the distribution of $\widehat{\sigma}_n^2/\sigma^2$ nor the distribution of the standardized residuals $\widehat{\mathbf{e}}_n = \mathbf{e}_n/\widehat{\sigma}_n = (I - \mathbf{H}_n)\widetilde{\boldsymbol{\epsilon}}_n/\sqrt{\widetilde{\boldsymbol{\epsilon}}_n^\intercal(I - \mathbf{H}_n)\widetilde{\boldsymbol{\epsilon}}_n n^{-1}}$ depends on either β^* or σ^2. Similarly, the distribution of $\sqrt{n}(\widehat{\beta}_n^0 - \beta^*)/\widehat{\sigma}_n = \sqrt{n}(\mathbb{X}_n^\intercal\mathbb{X}_n)^{-1}\mathbb{X}_n^\intercal\widetilde{\boldsymbol{\epsilon}}/\sqrt{\widetilde{\boldsymbol{\epsilon}}_n^\intercal(I - \mathbf{H}_n)\widetilde{\boldsymbol{\epsilon}}_n n^{-1}}$ does not depend on β^* or σ^2. We are assuming that both the predictors and outcome have been centered so that $\mathbf{1}_n^\intercal\widehat{\boldsymbol{e}}_n = 0$, where $\mathbf{1}_n$ is a vector of ones of length n.

The bootstrap distribution and the marginal distribution of the standardized residuals of the bootstrap estimator do not depend on either β^* or σ^2. To see this, let $M_n \in \{0,1\}^{n \times n}$ be a random $n \times n$ binary matrix whose rows are drawn uniformly with replacement from the canonical basis for $\mathbb{R}^n$. Then $M_n\widehat{\boldsymbol{e}}_n\widehat{\sigma}_n$ has the same bootstrap distribution as resampling the residuals $\boldsymbol{e}_n$ with replacement. Define $\mathbf{Y}_n^{(b)} = \mathbb{X}_n\widehat{\beta}_n^0 + M_n\widehat{\boldsymbol{e}}_n\widehat{\sigma}_n$ to be the bootstrap outcome and $\boldsymbol{\epsilon}_n^{(b)} = \mathbf{Y}_n^{(b)} - \mathbb{X}_n\widehat{\beta}_n^{0\,(b)} = (I - \mathbf{H}_n)\mathbf{Y}_n^{(b)} = (I - \mathbf{H}_n)M_n\widehat{\boldsymbol{e}}_n\widehat{\sigma}_n$ to be the bootstrap residuals. The bootstrap plug-in estimator of the residual variance is $\left(\widehat{\sigma}_n^{(b)}\right)^2 = ||(I - \mathbf{H}_n)M_n\widehat{\boldsymbol{e}}_n\widehat{\sigma}_n||^2/n = \widehat{\sigma}_n^2(M_n\widehat{\boldsymbol{e}}_n)^\intercal(I - \mathbf{H}_n)(M_n\widehat{\boldsymbol{e}}_n)/n$; thus, the distribution of $(\widehat{\sigma}_n^{(b)})^2/\widehat{\sigma}_n^2 = (M_n\widehat{\boldsymbol{e}}_n)^\intercal(I - \mathbf{H}_n)(M_n\widehat{\boldsymbol{e}}_n)/n$ does not depend on either β^* or σ^2. To see that the distribution of $\sqrt{n}(\widehat{\beta}_n^{0\,(b)} - \widehat{\beta}_n^0)/\widehat{\sigma}_n^{(b)}$ does not depend on β^* or σ write $\sqrt{n}(\widehat{\beta}_n^{0\,(b)} - \widehat{\beta}_n^0)/\widehat{\sigma}_n^{(b)}$ as

$$
\begin{aligned}
\frac{(\mathbb{X}_n^\intercal\mathbb{X}_n)^{-1}\mathbb{X}_n^\intercal(\mathbf{Y}_n^{(b)} - \mathbf{Y}_n)}{\widehat{\sigma}_n^{(b)}} &= (\mathbb{X}_n^\intercal\mathbb{X}_n)^{-1}\mathbb{X}_n^\intercal\left[\frac{M_n\widehat{\boldsymbol{e}}_n\widehat{\sigma}_n}{\widehat{\sigma}_n^{(b)}} - \frac{\widehat{\sigma}_n}{\widehat{\sigma}_n^{(b)}}\frac{(\mathbf{Y}_n - \mathbb{X}_n\widehat{\beta}_n^0)}{\widehat{\sigma}_n}\right] \\
&= (\mathbb{X}_n^\intercal\mathbb{X}_n)^{-1}\mathbb{X}_n^\intercal\left[\frac{M_n\widehat{\boldsymbol{e}}_n\widehat{\sigma}_n}{\widehat{\sigma}_n^{(b)}} - \frac{\widehat{\sigma}_n}{\widehat{\sigma}_n^{(b)}}\widehat{\boldsymbol{e}}_n\right] \\
&= (\mathbb{X}_n^\intercal\mathbb{X}_n)^{-1}\mathbb{X}_n^\intercal\left[M_n\widehat{\boldsymbol{e}}_n - \widehat{\boldsymbol{e}_n}\right]\frac{\widehat{\sigma}_n}{\widehat{\sigma}_n^{(b)}},
\end{aligned}
$$

and note that the (bootstrap and marginal) distribution of the last term does not depend on β^* or σ^2.

Proof of Lemma 4.6

Recall $\boldsymbol{\zeta}_n = \operatorname{diag}(C_n^{-1})$ and from (1)

$$\widetilde{\mathbb{V}}_n(t;s) = -2t^\intercal \mathbb{Z}_n + t^\intercal C_n t + \lambda_n \sum_{j=1}^{p} \frac{|\sqrt{n}\beta_j^* + t_j| - |\sqrt{n}\beta_j^*|}{|\mathbb{W}_{n,j} + \sqrt{n}\beta_j^*|} 1_{|\mathbb{W}_{n,j}+\sqrt{n}\beta_j^*|>\tau_{n,j}\zeta_{n,j}\widehat{\sigma}_n} + \lambda_n \sum_{j=1}^{p} \frac{|s_j + t_j| - |s_j|}{|\mathbb{W}_{n,j} + s_j|} 1_{|\mathbb{W}_j+\sqrt{n}\beta_j^*|\le\tau_{n,j}\zeta_{n,j}\widehat{\sigma}_n},$$

where we have replaced $\sqrt{n}\widehat{\beta}_{n,j}^0$ with $\mathbb{W}_{n,j} + \sqrt{n}\beta_j^*$. Because we are interested in uniform coverage of the proposed adaptive confidence interval we first make the dependence of $\widetilde{\mathbb{V}}_n(t;s)$ on β^* and σ explicit. Recall $\mathcal{U}(c) = \sup_{s\in\mathcal{S}_n} c^\intercal \arg\min_t \mathbb{V}_n(t;s) = \sup_s c^\intercal \arg\min_t \widetilde{\mathbb{V}}_n(t;s)$ so that (as shown in the discussion following (1))

$$\mathcal{U}(c)/\sigma = \sup_s c^\intercal \arg\min_t \widetilde{\mathbb{V}}_n(t;s)/\sigma = \left[\sup_s c^\intercal \arg\min_t \left\{\frac{\widetilde{\mathbb{V}}_n(\sigma\, t;\sigma\, s)}{\sigma^2}\right\}\right].$$

Define $\overline{\mathbb{V}}_n(t;s) = \widetilde{\mathbb{V}}_n(\sigma\, t, \sigma\, s)/\sigma^2$ then $\mathcal{U}(c)/\sigma = \sup_s c^\intercal \arg\min_t \overline{\mathbb{V}}_n(t;s)$. Furthermore,

$$\overline{\mathbb{V}}_n(t;s) = -2t^\intercal \widetilde{\mathbb{Z}}_n + t^\intercal C_n t + \frac{\lambda_n}{\sigma^2} \sum_{j=1}^{p} \frac{|\sqrt{n}\beta_j^*/\sigma + t_j| - |\sqrt{n}\beta_j^*/\sigma|}{|\widetilde{\mathbb{W}}_{n,j} + \sqrt{n}\beta_j^*/\sigma|} 1_{|\widetilde{W}_{n,j}+\sqrt{n}\beta_j^*/\sigma|>\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n} + \frac{\lambda_n}{\sigma^2} \sum_{j=1}^{p} \frac{|s_j + t_j| - |s_j|}{|\widetilde{\mathbb{W}}_{n,j} + s_j|} 1_{|\widetilde{W}_{n,j}+\sqrt{n}\beta_j^*/\sigma|\le\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n},$$

where $\widetilde{\mathbb{Z}}_n = \mathbb{Z}_n/\sigma$, $\widetilde{\mathbb{W}}_n = \mathbb{W}_n/\sigma$, and $\widetilde{\sigma}_n = \widehat{\sigma}_n/\sigma$; the review of least squares implies that the joint distribution of $\widetilde{\mathbb{Z}}_n$, $\widetilde{\mathbb{W}}_n$, and $\widetilde{\sigma}_n$ does not depend on β^* or σ. Recall $\lambda_n = r_n\widehat{\sigma}_n^2$ and by

(A5) r_n does not depend on β^* or σ^2. Thus, $\lambda_n/\sigma^2 = r_n\widetilde{\sigma}_n^2$. Write $\overline{\mathbb{V}}_n(t;s)$ as

$$- 2t^\intercal\widetilde{\mathbb{Z}}_n + t^\intercal C_n t + r_n\widetilde{\sigma}_n^2 \sum_{j=1}^p \frac{|s_j + t_j| - |s_j|}{|\widetilde{\mathbb{W}}_{n,j} + s_j|} 1_{|\widetilde{W}_{n,j}+\sqrt{n}\beta_j^*/\sigma|\le\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n} + r_n\widetilde{\sigma}_n^2 \sum_{j=1}^p \frac{|\sqrt{n}\beta_j^*/\sigma + t_j| - |\sqrt{n}\beta_j^*/\sigma|}{|\widetilde{\mathbb{W}}_{n,j} + \sqrt{n}\beta_j^*/\sigma|} 1_{|\widetilde{W}_{n,j}+\sqrt{n}\beta_j^*/\sigma|>\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}, \quad (6)$$

which depends on β^* and σ only through $\sqrt{n}\beta^*/\sigma$. Next we prove that $\overline{\mathbb{V}}_n(t;s)$ is strongly convex in t uniformly over s.

Lemma 1.2. *Assume (A1)-(A3), (A5). For each $t, u, s \in \mathbb{R}^p$*

$$\overline{\mathbb{V}}_n(u;s) - \overline{\mathbb{V}}_n(t;s) \ge g_{t,s}{}^\intercal(u-t) + \frac{m}{2}||u-t||^2,$$

where $g_{t,s}$ is any subgradient of $\overline{\mathbb{V}}_n(t;s)$, taken with respect to t, and m does not depend on t, u, or s.

Proof. To make the notation more compact, let $\omega_j^\beta = 1_{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|>\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}/|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|$ and $\omega_j^s = 1_{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|\le\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}/|\widetilde{\mathbb{W}}_{n,j} + s_j|$, then

$$\overline{\mathbb{V}}_n(t,s) = -2t^\intercal\widetilde{\mathbb{Z}}_n + t^\intercal C_n t + r_n\widetilde{\sigma}_n^2 \sum_{j=1}^p \omega_j^s \left(|t_j + s_j| - |s_j|\right) + r_n\widetilde{\sigma}_n^2 \sum_{j=1}^p \omega_j^\beta \left(|t_j + \sqrt{n}\beta_j^*/\sigma| - |\sqrt{n}\beta_j^*/\sigma|\right).$$

Any subgradient of $\overline{\mathbb{V}}_n(t;s)$ taken with respect to t has the form

$$g_{t,s} = -2\widetilde{\mathbb{Z}}_n + 2C_n t + \left\{r_n\widetilde{\sigma}_n^2\omega_j^s\left(1_{s_j>-t_j} - 1_{s_j<-t_j} + \rho_j 1_{s_j=-t_j}\right)\right\}_{j=1}^p + \left\{r_n\widetilde{\sigma}_n^2\omega_j^\beta\left(1_{\sqrt{n}\beta_j^*/\sigma>-t_j} - 1_{\sqrt{n}\beta_j^*/\sigma<-t_j} + \rho_j' 1_{\sqrt{n}\beta_j^*/\sigma=-t_j}\right)\right\}_{j=1}^p,$$

where $\rho_1, \ldots, \rho_p, \rho'_1, \ldots, \rho'_p \in [-1,\, 1]$. Now $\overline{\mathbb{V}}_n(u; s) - \overline{\mathbb{V}}_n(t; s)$ is equal to

$$- 2u^\intercal \widetilde{\mathbb{Z}}_n + u^\intercal C_n u + 2t^\intercal \widetilde{\mathbb{Z}}_n - t^\intercal C_n t + r_n \widetilde{\sigma}_n^2 \sum_{j=1}^{p} \omega_j^s \left(|u_j + s_j| - |t_j + s_j|\right) + r_n \widetilde{\sigma}_n^2 \sum_{j=1}^{p} \omega_j^\beta \left(|u_j + \sqrt{n}\beta_j^*/\sigma| - |t_j + \sqrt{n}\beta_j^*/\sigma|\right).$$

Because $-2t^\intercal \widetilde{\mathbb{Z}}_n + t^\intercal C_n t$ is strongly convex as a function of t it follows that

$$-2u^\intercal \widetilde{\mathbb{Z}}_n + u^\intercal C_n u + 2t^\intercal \widetilde{\mathbb{Z}}_n - t^\intercal C_n t \geq -2\widetilde{\mathbb{Z}}_n^\intercal (u - t) + 2t^\intercal C_n (u - t) + \frac{m}{2} ||u - t||^2$$

where $m = 2 \min \mathrm{eigen}\, C_n$. To prove that $\overline{\mathbb{V}}_n(u; s) - \overline{\mathbb{V}}_n(t; s)$ is larger than $g_{t,s}^\intercal (u - t) + (m/2)||u - t||^2$ we need only show $|u_j + s_j| - |t_j + s_j| \geq (u_j - t_j)(1_{s_j > -t_j} - 1_{s_j < -t_j} + \rho_j 1_{s_j = t_j})$ for any $s_j,\, t_j,\, u_j \in \mathbb{R}$ and $\rho_j \in [-1, 1]$ (note $\omega_j^\beta,\ \omega_j^s,\ \widetilde{\lambda}_n \widetilde{\sigma}_n^2$ are all nonnegative). We consider three cases: (i) $s_j > -t_j$ which implies $(u_j - t_j) = (u_j + s_j) - (s_j + t_j) \leq |u_j + s_j| - (s_j + t_j) = |u_j + s_j| - |t_j + s_j|$; (ii) $s_j < -t_j$ which implies $-(u_j - t_j) = -(u_j + s_j - (t_j + s_j)) = -(u_j + s_j) - |t_j + s_j| \leq |u_j + s_j| - |t_j + s_j|$; and (iii) $s_j = -t_j$ which implies $\rho_j(u_j - t_j) = \rho_j(u_j + s_j - (t_j + s_j)) = \rho(u_j + s_j) \leq |u_j + s_j|$. Because this result holds for any s_j it holds for $s_j = \sqrt{n}\beta_j^*/\sigma$. □

Recall that $\mathcal{U}(c)/\sigma = \sup_s c^\intercal \arg\min_t \overline{\mathbb{V}}_n(t; s)$ where $\overline{\mathbb{V}}_n(t; s)$ is defined in (6). We will show that $\arg\min_{t \in \mathbb{R}^p} \overline{\mathbb{V}}_n(t; s) \in \mathcal{B}_n$, where $\mathcal{B}_n = \left\{ t \in \mathbb{R}^p \,:\, ||t|| \leq \sqrt{4 r_n \widetilde{\sigma}_n^2 p/m} + ||\widetilde{\mathbb{W}}_n|| \right\}$. To see this we use the notation from the proof of Lemma 1.2. From Lemma 1.2 and strong convexity $\overline{\mathbb{V}}_n(\widetilde{\mathbb{W}}_n; s) - \overline{\mathbb{V}}_n(t_{\min,s}; s) \geq (m/2)||\widetilde{\mathbb{W}}_n - t_{\min,\, s}||^2$, where $t_{\min,\, s}$ minimizes $\overline{\mathbb{V}}_n(t; s)$ in t and we have chosen $g_{t_{\min,s}, s} = 0$. Furthermore because $\widetilde{\mathbb{W}}_n$ minimizes $-2t^\intercal \widetilde{\mathbb{Z}}_n + t^\intercal C_n t$ (recall that $\widetilde{\mathbb{Z}}_n = C_n \widetilde{\mathbb{W}}_n$) and using the fact that ω_j^β, ω_j^s are nonnegative, we have that $\overline{\mathbb{V}}_n(\widetilde{\mathbb{W}}_n; s) - \overline{\mathbb{V}}_n(t_{\min,s}; s) \leq 2 p r_n \widetilde{\sigma}_n^2$. Therefore, $||t_{\min,\, s}|| \leq ||t_{\min,\, s} - \widetilde{\mathbb{W}}_n|| + ||\widetilde{\mathbb{W}}_n|| \leq \sqrt{4 r_n \widetilde{\sigma}_n^2 p/m} + ||\widetilde{\mathbb{W}}_n||$ which does not depend on s, β^*, or σ^*.

We now show that the last term in (6) converges uniformly to zero in probability.

Lemma 1.3. *Assume (A1)-(A3), (A5), then*

$$\sup_{t\in\mathcal{B}_n,\,\beta^*\in\mathbb{R}^p,\,\sigma>0}\left|\sum_{j=1}^{p}\frac{|\sqrt{n}\beta_j^*/\sigma+t_j|-|\sqrt{n}\beta_j^*/\sigma|}{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|}\right|1_{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|>\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}=o_p(1).$$

Proof. From the triangle inequality and the preceding lemma,

$$\begin{aligned}\left|\sum_{j=1}^{p}\frac{|\sqrt{n}\beta_j^*/\sigma+t_j|-|\sqrt{n}\beta_j^*/\sigma|}{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|}\right|1_{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|>\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n} &\le \frac{||t||\sqrt{p}}{\min_j\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}\\ &\le \frac{\sqrt{p}}{\min_j\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}\left[\sqrt{\frac{2r_n\widetilde{\sigma}_n^2p}{m}}+||\widetilde{\mathbb{W}}_n||\right],\end{aligned}$$

the last expression in the above display does not depend on β^*, σ, or t and is $o_P(1)$ because $\min_j\tau_{n,j}\zeta_{n,j}$ converges to ∞ as $n\to\infty$. □

Define

$$\overline{\mathbb{V}}_n^*(t;s)=\overline{\mathbb{V}}_n(t;s)-r_n\widetilde{\sigma}_n^2\sum_{j=1}^{p}\frac{|\sqrt{n}\beta_j^*/\sigma+t_j|-|\sqrt{n}\beta_j^*/\sigma|}{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|}\Bigg|1_{|\widetilde{\mathbb{W}}_{n,j}+\sqrt{n}\beta_j^*/\sigma|>\tau_{n,j}\zeta_{n,j}\widetilde{\sigma}_n}.$$

From Lemma (1.3) for any $\epsilon>0$, if follows that $\sup_{\beta^*\in\mathbb{R}^p,\,\sigma>0}P_{\beta^*,\,\sigma}(\sup_{s\in\mathbb{R}^p,\,t\in\mathcal{B}_n}|\overline{\mathbb{V}}_n^*(t;s)-\overline{\mathbb{V}}_n(t;s)|>\epsilon)\to 0$ as $n\to\infty$. Furthermore, using the same arguments as given above, $\arg\min_{t\in\mathbb{R}^p}\overline{\mathbb{V}}_n^*(t;s)=\arg\min_{t\in\mathcal{B}_n}\overline{\mathbb{V}}_n^*(t;s)$. As before, let $t_{\min,\,s}=\arg\min_{t\in\mathcal{B}_n}\overline{\mathbb{V}}_n(t;s)$, and define $t^*_{\min,\,s}=\arg\min_{t\in\mathcal{B}_n}\overline{\mathbb{V}}_n^*(t;s)$.

Lemma 1.4. *Assume (A1)-(A3), (A5) and let $\epsilon>0$ be arbitrary. Then,* $\sup_{\beta^*\in\mathbb{R}^p,\,\sigma>0}P_{\beta^*,\sigma}(\sup_{s\in\mathbb{R}^p}|t_{\min,\,s}-t^*_{\min,\,s}|>\epsilon)\to 0$ *as* $n\to\infty$.

Proof. Because $t_{\min,s}$ is the minimizer of $\overline{\mathbb{V}}_n(t;s)$

$$\begin{aligned}
0 &\leq \overline{\mathbb{V}}_n(t^*_{\min,s};s) - \overline{\mathbb{V}}_n(t_{\min,s};s) \\
&= \overline{\mathbb{V}}_n(t^*_{\min,s};s) - \overline{\mathbb{V}}^*_n(t^*_{\min,s};s) + \overline{\mathbb{V}}^*_n(t^*_{\min,s};s) - \overline{\mathbb{V}}^*_n(t_{\min,s};s) + \overline{\mathbb{V}}^*_n(t_{\min,s};s) - \overline{\mathbb{V}}_n(t_{\min,s};s) \\
&\leq \overline{\mathbb{V}}_n(t^*_{\min,s};s) - \overline{\mathbb{V}}^*_n(t^*_{\min,s};s) + \overline{\mathbb{V}}^*_n(t_{\min,s};s) - \overline{\mathbb{V}}_n(t_{\min,s};s) \\
&\leq \sup_{s\in\mathbb{R}^p}\sup_{t\in\mathcal{B}_n} 2\left|\overline{\mathbb{V}}_n(t;s) - \overline{\mathbb{V}}^*_n(t;s)\right|.
\end{aligned}$$

Using strong convexity of $\overline{\mathbb{V}}_n(t;s)$, $P_{\beta^*,\sigma}(\sup_{s\in\mathbb{R}^p}||t_{\min,s}-t^*_{\min,s}||^2 > \epsilon) \leq P_{\beta^*,\sigma}((2/m)\left|\overline{\mathbb{V}}_n(t^*_{\min,s};s) - \overline{\mathbb{V}}_n(t_{\min,s};s)\right| > \epsilon)$. The above arguments imply that the latter is bounded above by $\sup_{\beta^*,\sigma} P_{\beta^*,\sigma}(|\sup_{s\in\mathbb{R}^p}\sup_{t\in\mathcal{B}_n} 2\left|\overline{\mathbb{V}}_n(t;s) - \overline{\mathbb{V}}^*_n(t;s)\right| > \epsilon)$ which converges to zero as $n\to\infty$ by Lemma (1.3). □

Let $\delta > 0$ be arbitrary. We will now use the foregoing results to prove that under (A0)-(A3) and (A5)

$$\sup_{\beta^*\in\mathbb{R}^p,\sigma>0} P_{\beta^*,\sigma}\left\{\inf_{a\in\mathbb{R}}\left[P_B\left(\frac{\mathcal{U}^{(b)}(\mathbf{c})}{\widehat{\sigma}^{(b)}_n} \geq a\right) - P_{\beta^*,\sigma}\left(\frac{\mathbf{c}^\intercal\sqrt{n}(\widehat{\beta}^{\lambda_n}_n - \beta^*)}{\sigma} \geq a\right)\right] \leq -\delta\right\} = o(1);$$

this statement is equivalent to the condition stated in the first part of Lemma 4.6.

Proof. We use a proof by contradiction. Suppose there exists $\delta,\epsilon > 0$ and subsequence $\beta^*_{n_k},\sigma_{n_k}$ indexed by k where $\{n_k\}_{k=1}^\infty$ is a strictly increasing sequence of integers so that

$$P_{\beta^*_{n_k},\sigma_{n_k}}\left\{\inf_{a\in\mathbb{R}}\left[P_B\left(\frac{\mathcal{U}^{(b)}(\mathbf{c})}{\widehat{\sigma}^{(b)}_{n_k}} \geq a\right) - P_{\beta^*_{n_k},\sigma_{n_k}}\left(\frac{\mathbf{c}^\intercal\sqrt{n_k}(\widehat{\beta}^{\lambda_{n_k}}_{n_k} - \beta^*_{n_k})}{\sigma_{n_k}} \geq a\right)\right] \leq -\delta\right\} \geq \epsilon \quad (7)$$

for all k. Define $\gamma_{n_k} = \sqrt{n_k}\beta^*_{n_k}/\sigma_{n_k}$; there exists a further subsequence, say n_{k_m}, so that for each $j = 1,\ldots,p$ $|\gamma_{n_{k_m},j}|$ is either bounded or diverges to $+\infty$. To simplify notation, hereafter we will use the subscript n_m as shorthand for n_{k_m} and we write $q_{n_m} = o_P(1)$ to mean for any $\eta > 0$, $\lim_{m\to\infty} P_{\beta^*_{n_m},\sigma_{n_m}}(q_{n_m} > \eta) = 0$. Define $I \triangleq \{j \in \{1,\ldots,p\} : |\gamma_{m,j}| \to \infty \text{ as } m\to\infty\}$.

Recall that $\sqrt{n_m}(\widehat{\beta}_{n_m}^{\lambda_{n_m}} - \beta_{n_m}^*)/\sigma_{n_m} = \arg\min_t \overline{\mathbb{V}}_{n_m}^*(t;\gamma_{n_m}) + o_P(1)$ and that $\sqrt{n_m}(\widehat{\beta}_{n_m}^{\lambda_{n_m}} - \beta_{n_m}^*)/\sigma_{n_m} \in \mathcal{B}_{n_m}$. Write $\overline{\mathbb{V}}_{n_m}^*(t;\gamma_{n_m})$ as

$$
\begin{aligned}
-2t^\intercal \widetilde{\mathbb{Z}}_{n_m} + t^\intercal C_{n_m} t + r_{n_m}\widetilde{\sigma}_{n_m}^2 \sum_{j\notin I} \frac{|\gamma_{n_m,j} + t_j| - |\gamma_{n_m,j}|}{|\widetilde{\mathbb{W}}_{n_m,j} + \gamma_{n_m,j}|} 1_{|\widetilde{\mathbb{W}}_{n_m,j}+\gamma_{n_m,j}| \le \tau_{n_m,j}\zeta_{n_m,j}\widetilde{\sigma}_{n_m}} \\
+ r_{n_m}\widetilde{\sigma}_{n_m}^2 \sum_{j\in I} \frac{|\gamma_{n_m,j} + t_j| - |\gamma_{n_m,j}|}{|\widetilde{\mathbb{W}}_{n_m,j} + \gamma_{n_m,j}|} 1_{|\widetilde{\mathbb{W}}_{n_m,j}+\gamma_{n_m,j}| \le \tau_{n_m,j}\zeta_{n_m,j}\widetilde{\sigma}_{n_m}}. \qquad (8)
\end{aligned}
$$

Because $|\gamma_{n_m,j}| \to \infty$ for $j \in I$ an essentially identical argument to the proof of Lemma (1.3) shows that the last term in (8) converges to zero in probability uniformly over $t \in \mathcal{B}_{n_m}$, that is

$$
\sup_{t\in\mathcal{B}_{n_m}} \left| r_{n_m}\widetilde{\sigma}_{n_m}^2 \sum_{j\in I} \frac{|\gamma_{n_m,j} + t_j| - |\gamma_{n_m,j}|}{|\widetilde{\mathbb{W}}_{n_m,j} + \gamma_{n_m,j}|} 1_{|\widetilde{\mathbb{W}}_{n_m,j}+\gamma_{n_m,j}| \le \tau_{n_m,j}\zeta_{n_m,j}\widetilde{\sigma}_{n_m}} \right| = o_P(1).
$$

Similarly, it can be shown that

$$
\sup_{t\in\mathcal{B}_{n_m}} \left| r_{n_m}\widetilde{\sigma}_{n_m}^2 \sum_{j\notin I} \frac{|\gamma_{n_m,j} + t_j| - |\gamma_{n_m,j}|}{|\widetilde{\mathbb{W}}_{n_m,j} + \gamma_{n_m,j}|} 1_{|\widetilde{\mathbb{W}}_{n_m,j}+\gamma_{n_m,j}| > \tau_{n_m,j}\zeta_{n_m,j}\widetilde{\sigma}_{n_m}} \right| = o_P(1)
$$

and thus $\mathbf{c}^\intercal \sqrt{n_m}\left(\widehat{\beta}_{n_m}^{\lambda_{n_m}} - \beta_{n_m}^*\right) = \arg\min_{t\in\mathbb{R}^p} \check{\mathbb{V}}_{n_m}(t;\gamma_{n_m}) + o_P(1)$, where

$$
\check{\mathbb{V}}_{n_m}(t;\gamma_{m_n}) \triangleq -2t^\intercal \widetilde{\mathbb{Z}}_{n_m} + t^\intercal C_{n_m} t + r_{n_m}\widetilde{\sigma}_{n_m}^2 \sum_{j\notin I} \frac{|\gamma_{n_m,j} + t_j| - |\gamma_{n_m,j}|}{|\widetilde{\mathbb{W}}_{n_m,j} + \gamma_{n_m,j}|}.
$$

Define $\ell(v,\Sigma,\alpha) \triangleq \arg\min_{\mu\in\mathbb{R}^p} \mu^\intercal \Sigma \mu + \alpha \sum_{j\notin I} |\mu_j + v_j|/|v_j|$ then some algebra shows that

$$
\mathbf{c}^\intercal \sqrt{n_m}\left(\widehat{\beta}_{n_m}^{\lambda_{n_m}} - \beta_{n_m}^*\right) = \mathbf{c}^\intercal \ell(\gamma_{n_m} + \widetilde{\mathbb{W}}_{n_m}, C_{n_m}, r_{n_m}\widetilde{\sigma}_{n_m}^2) + \mathbf{c}^\intercal \widetilde{\mathbb{W}}_{n_m} + o_P(1). \qquad (9)
$$

To obtain the desired result we will express $\mathcal{U}^{(b)}(\mathbf{c})$ in terms of the function ℓ and $\widetilde{\mathbb{W}}_{n_m}^{(b)}$. For clarity, we write $\mathcal{U}_{n_m}^{(b)}(\mathbf{c})$ to emphasize that we are considering a subsequence. Hereafter, without loss of generality we assume $t \in \mathcal{B}_{n_m}^{(b)}$. Using the results for the bootstrap and linear regression and extensions of the arguments given above, it follows that $\mathcal{U}_{n_m}^{(b)}(\mathbf{c}) = \sup_s \mathbf{c}^\intercal \arg\min_t \overline{\mathbb{V}}_{n_m}^{(b)*}(t;s)$, where

$$\overline{\mathbb{V}}_{n_m}^{(b)*}(t;s) = -2t^\intercal \widetilde{\mathbb{Z}}_{n_m}^{(b)} + t^\intercal C_{n_m} t + r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2 \sum_{j=1}^{p} \frac{|s_j+t_j|-|s_j|}{|\widetilde{\mathbb{W}}_{n_m,j}^{(b)}+s_j|} 1_{|\widetilde{W}_{n_m,j}^{(b)}+\gamma_{n_m,j}^{(b)}|\le \tau_{n_m,j}\zeta_{n_m,j}\widetilde{\sigma}_{n_m}^{(b)}} + o_{P_B}(1),$$

where $\gamma_{n_m,j}^{(b)} = \sqrt{n_m}\widehat{\beta}_{n_m}/\widehat{\sigma}_{n_m}$ and we write $q_{n_m}^{(b)} = o_{P_B}(1)$ to mean for any $\eta > 0$, $P_B(q_{n_m}^{(b)} > \eta) = o_P(1)$. Define $\mathcal{T}_n(I) = \{s \in \mathbb{R}^p : s_j = \gamma_{n,j} \text{ if } j \in I\}$. Then $\mathcal{U}_{n_m}^{(b)}(\mathbf{c}) \ge \sup_{s\in\mathcal{T}_{n_m}(I)} \mathbf{c}^\intercal \arg\min_t \overline{\mathbb{V}}_{n_m}^{(b)*}(t;s)$, but for any $s \in \mathcal{T}_{n_m}(I)$

$$\begin{aligned}\overline{\mathbb{V}}_{n_m}^{(b)*}(t;s) &= -2t^\intercal \widetilde{\mathbb{Z}}_{n_m}^{(b)} + t^\intercal C_{n_m} t + r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2 \sum_{j\notin I} \frac{|s_j+t_j|-|s_j|}{|\widetilde{\mathbb{W}}_{n_m,j}^{(b)}+s_j|} 1_{|\widetilde{W}_{n_m,j}^{(b)}+\gamma_{n_m,j}^{(b)}|\le \tau_{n_m,j}\zeta_{n_m,j}\widetilde{\sigma}_{n_m}^{(b)}} + o_{P_B}(1) \\ &= -2t^\intercal \widetilde{\mathbb{Z}}_{n_m}^{(b)} + t^\intercal C_{n_m} t + r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2 \sum_{j\notin I} \frac{|s_j+t_j|-|s_j|}{|\widetilde{\mathbb{W}}_{n_m,j}^{(b)}+s_j|} + o_{P_B}(1),\end{aligned}$$

where we have used $\gamma_{n_m,j}^{(b)} = (\gamma_{n_m,j}^{(b)} - \gamma_{n_m,j}) + \gamma_{n_m,j}$, which for $j \notin I$ is bounded in P_B probability uniformly over m. Thus if $s \in \mathcal{T}_{n_m}(I)$ then $\mathbf{c}^\intercal \arg\min_t \overline{\mathbb{V}}_{n_m}^{(b)*}(t;s) = \mathbf{c}^\intercal \ell(s + \widetilde{\mathbb{W}}_{n_m}^{(b)}, C_{n_m}, r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2) + \mathbf{c}^\intercal \widetilde{\mathbb{W}}_n^{(b)} + o_{P_B}(1)$, where we have used the strong convexity of $\overline{\mathbb{V}}_{n_m}^{(b)}(t;s)$ to ensure closeness of the arg mins. Furthermore, because ℓ only depends on components of its first argument that do not belong to I, $\sup_{s\in\mathcal{T}_{n_m}(I)} \mathbf{c}^\intercal \arg\min_{t\in\mathbb{R}^p} \overline{\mathbb{V}}_{n_m}^{(b)*}(t;s) = \sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal \ell(s, C_{n_m}, r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2)$ $\mathbf{c}^\intercal \widetilde{\mathbb{W}}_n^{(b)} + o_{P_B}(1)$. Thus,

$$\begin{aligned} P_B\left(\frac{\mathcal{U}_{n_m}(\mathbf{c})}{\widehat{\sigma}_{n_m}^{(b)}} \ge a\right) &\ge P_B\left(\sup_{s\in\mathcal{T}_{n_m}(I)} \mathbf{c}^\intercal \arg\min_t \overline{\mathbb{V}}_{n_m}^{(b)*}(t;s)\right) + o_P(1) \\ &= P_B\left(\sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal \ell(s, C_{n_m}, r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2) + \mathbf{c}^\intercal \widetilde{\mathbb{W}}_{n_m}^{(b)} \ge a\right) + o_P(1). \end{aligned}$$

Furthermore, using continuity of $f(\Sigma, \alpha) \triangleq \sup_{s\in\mathbb{R}^p} \mathbf{c}^\intercal \ell(s, \Sigma, \alpha)$ (see the proof of Lemma 1.1), and standard consistency results for the residual bootstrap (e.g., Freedman et al., 1981), for any $\eta > 0$

$$P\left(\sup_{a}\left|P_B\left(\sup_{s\in\mathbb{R}^p}\mathbf{c}^\intercal\ell(s, C_{n_m}, r_{n_m}(\widetilde{\sigma}_{n_m}^{(b)})^2) + \mathbf{c}^\intercal\widetilde{\mathbb{W}}_{n_m}^{(b)} \geq a\right)\right.\right.$$
$$\left.\left. - P\left(\sup_{s\in\mathbb{R}^p}\mathbf{c}^\intercal\ell(s, C_{n_m}, r_{n_m}\widetilde{\sigma}_{n_m}^2) + \mathbf{c}^\intercal\widetilde{\mathbb{W}}_{n_m} \geq a\right)\right| > \eta\right) = o(1).$$

Finally, since

$$P\left(\sup_{s\in\mathbb{R}^p}\mathbf{c}^\intercal\ell(s, C_{n_m}, r_{n_m}\widetilde{\sigma}_{n_m}^2) + \mathbf{c}^\intercal\widetilde{\mathbb{W}}_{n_m} \geq a\right) \geq P\left(\mathbf{c}^\intercal\ell(\widetilde{\mathbb{W}}_{n_m} + \gamma_{n_m}, C_{n_m}, r_{n_m}\widetilde{\sigma}_{n_m}^2) + \mathbf{c}^\intercal\widetilde{\mathbb{W}}_{n_m} \geq a\right),$$

it must be that (7) cannot hold. Thus, we have a contradiction. □

1.3 Additional simulation results

1.3.1 Double bootstrap algorithm for choosing τ_n

Let $\mathbf{c} \in \mathbb{R}^p$ be fixed. For a data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ let $\zeta_{1-\alpha}(\mathcal{D}, t)$ denote the $(1-\alpha)\times 100\%$ ACI confidence interval based on $\mathcal{D}$ with pretest threshold $\tau_{n,j} = t$ for all j. Let $a_{\min}$ and $a_{\max}$ denote the lower and upper endpoints of the grid of scale values and a_{step} denote a step size; in our simulations we used $a_{\min} = 0.1$, $a_{\max} = 4.0$ and $a_{\text{step}} = 0.1$. Let B denote the number of outer bootstrap iterations and W denote the number of inner bootstrap iterations (see below), in our simulations we chose $B = W = 1000$. The double bootstrap algorithm for choosing τ_n is as follows:

Algorithm 1: Double bootstrap procedure for tuning τ_n.

initialize $\varrho = \mathbf{c}^\intercal \hat{\beta}_n^{\lambda_n}$, $a = a_{\min}$

while $a \leq a_{\max}$ **do**

  for $b = 1, 2, \ldots, B$ **do**

    Draw a random sample of size n from $\mathbb{P}_n$, say $\mathcal{D}^{(b)}$

    Compute $\zeta_{1-\alpha}(\mathcal{D}^{(b)}, a\sqrt{\log\log n})$ using W resamples drawn from $\mathcal{D}^{(b)}$.

  end

  Compute $\hat{p} \triangleq \frac{1}{B}\sum_{b=1}^{B} 1_{\varrho \in \zeta_{1-\alpha}(\mathcal{D}^{(b)}, a\sqrt{\log\log n})}$.

  if $\hat{p} \geq 1 - \alpha - 2\sqrt{\frac{\alpha(1-\alpha)}{B}}$ **then**

    break.

  end

  else

    Set $a = a + a_{\text{step}}$.

  end

end

It is possible that the above algorithm will terminate by exhausting the grid, however, in our simulations with $a_{\max} = 4.0$ this never occurred. Should this happen, one can increase the value of $a_{\max}$.

References

Chatterjee, A. and Lahiri, S. "Asymptotic properties of the residual bootstrap for Lasso estimators." *Proceedings of the American Mathematical Society*, 138(12):4497–4509 (2010).

Freedman, D. A. et al. "Bootstrapping regression models." *The Annals of Statistics*, 9(6):1218–1228 (1981).

Hu, C. and Dong, Y. "Estimating the predictive quality of dose–response after model selection." *Statistics in medicine*, 26(16):3114–3139 (2007).

Jin, Z., Ying, Z., and Wei, L. "A simple resampling method by perturbing the minimand." *Biometrika*, 88(2):381–390 (2001).

Kabaila, P. "The effect of model selection on confidence regions and prediction regions." *Econometric Theory*, 11:537–537 (1995).

Pötscher, B. "Effects of model selection on inference." *Econometric Theory*, 7(2):163–185 (1991).

Pötscher, B. and Schneider, U. "On the distribution of the adaptive LASSO estimator." *Journal of Statistical Planning and Inference*, 139(8):2775–2790 (2009).

—. "Confidence sets based on penalized maximum likelihood estimators in Gaussian regression." *Electronic Journal of Statistics*, 4:334–360 (2010).

Shen, X., Huang, H., and Ye, J. "Inference after model selection." *Journal of the American Statistical Association*, 99(467):751–762 (2004).

Van Der Vaart, A. and Wellner, J. *Weak convergence and empirical processes*. Springer Verlag (1996).

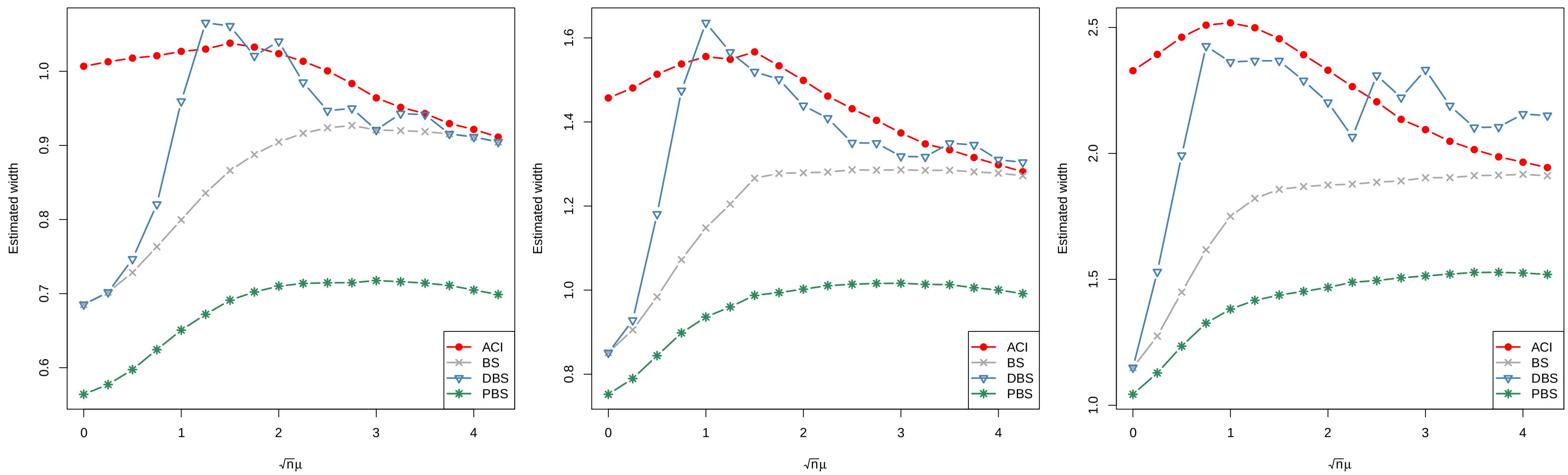

Figure 1: **Left:** Width of the ACI, BS, DBS, and PBS for dimension $p = 5$ and $s = 2$ large effects. **Center:** Width of the ACI, BS, DBS, and PBS for dimension $p = 10$ and $s = 4$ large effects. **Right:** Width of the ACI, BS, DBS, and PBS for dimension $p = 20$ and $s = 8$ large effects.

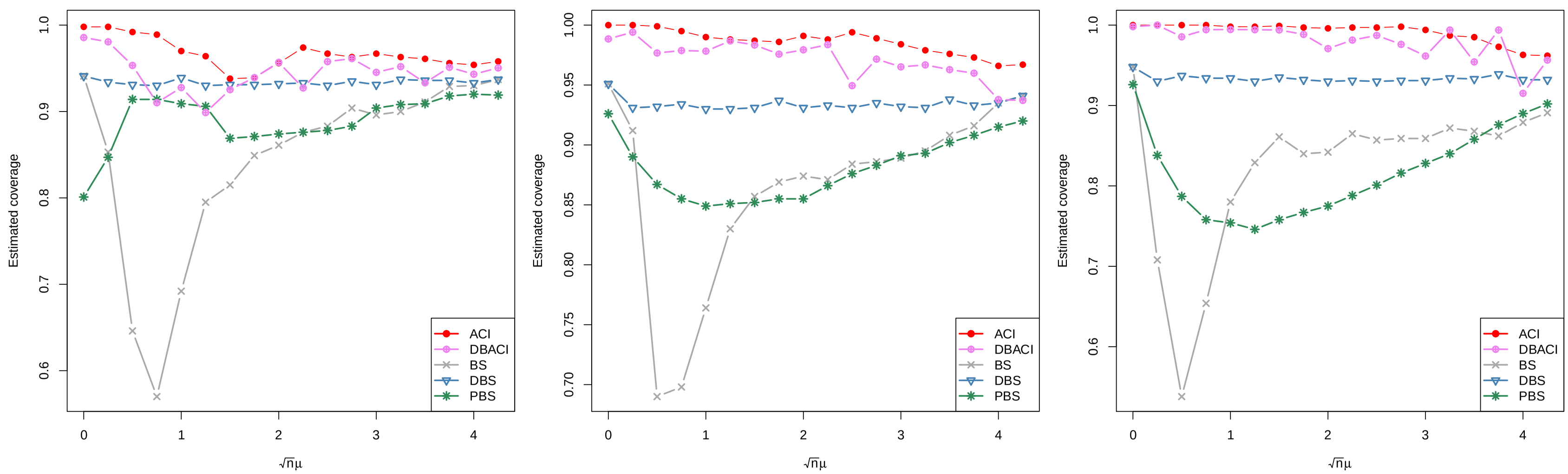

Figure 2: Estimated coverage probabilities with $\tau_{n,j} = \sqrt{\log n}$ for all j. **Left:** Estimated coverage of the ACI, BS, DBS, and PBS for dimension $p = 5$ and $s = 1$ large effects. **Center:** Estimated coverage of the ACI, BS, DBS, and PBS for dimension $p = 10$ and $s = 2$ large effects. **Right:** Estimated coverage of the ACI and BS for dimension $p = 20$ and $s = 4$ large effects. Target coverage is 95%.

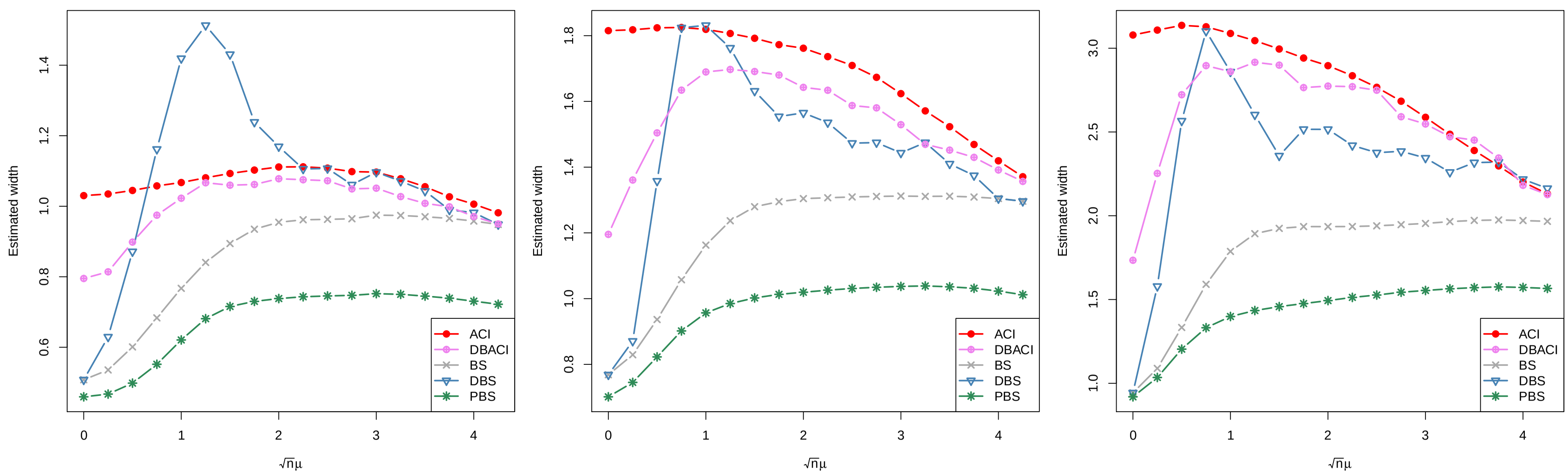

Figure 3: Estimated widths with $\tau_{n,j} = \sqrt{\log n}$ for all j. **Left:** Width of the ACI, BS, DBS, and PBS for dimension $p = 5$ and $s = 1$ large effects. **Center:** Width of the ACI, BS, DBS, and PBS for dimension $p = 10$ and $s = 2$ large effects. **Right:** Width of the ACI, BS, DBS, and PBS for dimension $p = 20$ and $s = 4$ large effects.

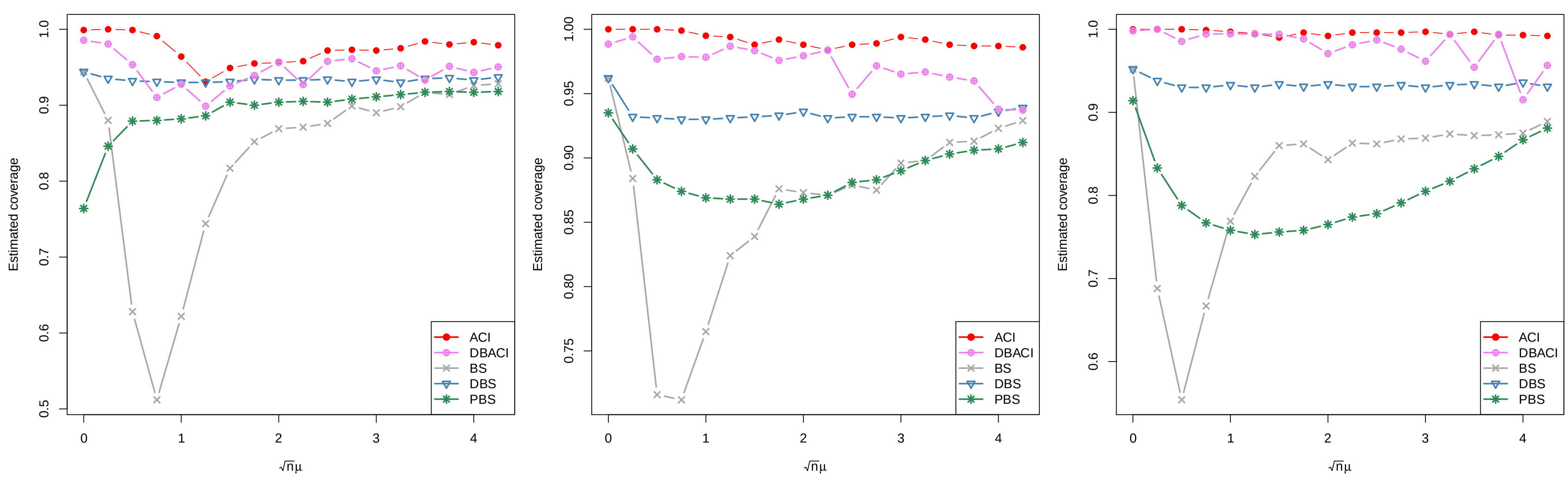

Figure 4: Estimated coverage probabilities with $\tau_{n,j} = n^{1/4}$ for all j. **Left:** Estimated coverage of the ACI, BS, DBS, and PBS for dimension $p = 5$ and $s = 1$ large effects. **Center:** Estimated coverage of the ACI, BS, DBS, and PBS for dimension $p = 10$ and $s = 2$ large effects. **Right:** Estimated coverage of the ACI and BS for dimension $p = 20$ and $s = 4$ large effects. Target coverage is 95%.

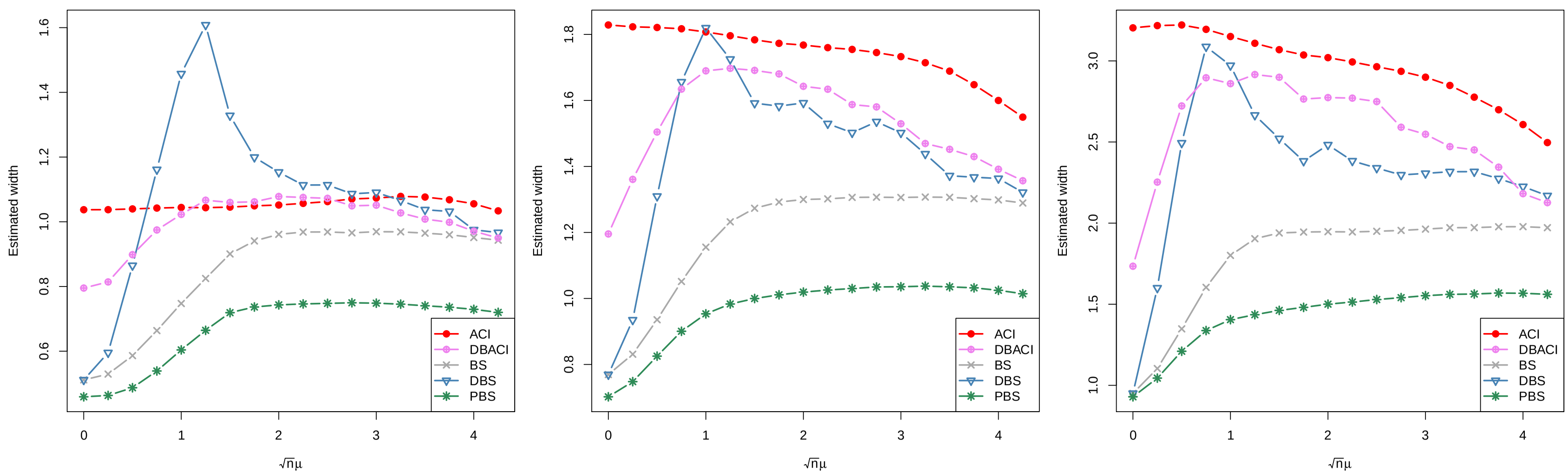

Figure 5: Estimated widths with $\tau_{n,j} = n^{1/4}$ for all j. **Left:** Width of the ACI, BS, DBS, and PBS for dimension $p = 5$ and $s = 1$ large effects. **Center:** Width of the ACI, BS, DBS, and PBS for dimension $p = 10$ and $s = 2$ large effects. **Right:** Width of the ACI, BS, DBS, and PBS for dimension $p = 20$ and $s = 4$ large effects.