Online Appendix to The Proximal Bootstrap for Finite-Dimensional Regularized Estimators

Jessie Li

January 20, 2021

1 Examples of Regularized Estimators

1.1 LASSO

$$\hat{\beta}_n = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 + \frac{\lambda_n}{\sqrt{n}} \|\beta\|_1 \right\}$$

$$r'_{\beta_0}(h) = \sum_{j=1}^{p} \left(h_j sign(\beta_{0j}) 1(\beta_{0j} \neq 0) + |h_j| 1(\beta_{0j} = 0) \right)$$

$$\hat{\beta}_n^* = \arg\min_{\beta} \alpha_n \lambda_n \|\beta\|_1 + \alpha_n \sqrt{n} \left(\hat{l}_n^*(\bar{\beta}_n) - \hat{l}_n(\bar{\beta}_n) \right)' (\beta - \bar{\beta}_n) + \frac{1}{2} \|\beta - \bar{\beta}_n\|_{\bar{H}_n}^2$$

$$l(\beta_0) = -E[x_i (y_i - x_i'\beta)], \quad \hat{l}_n(\bar{\beta}_n) = -\frac{1}{n} \sum_{i=1}^{n} x_i (y_i - x_i'\bar{\beta}_n)$$

$$H_0 = E[x_i x_i'], \quad \bar{H}_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i'$$

Examples of $\hat{l}_n^*(\bar{\beta}_n)$ include the multinomial and wild bootstrap analogs of $\hat{l}_n(\bar{\beta}_n)$:

$$\hat{l}_n^*(\bar{\beta}_n) = -\frac{1}{n} \sum_{i=1}^{n} x_i^* \left(y_i^* - x_i^{*\prime} \bar{\beta}_n \right), \quad \hat{l}_n^*(\bar{\beta}_n) = -\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi}) x_i \left(y_i - x_i' \bar{\beta}_n \right)$$

where ξ_i are i.i.d. variables with variance 1 and finite 3rd moment and $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$.

If $\bar{H}_n = \frac{1}{c} I_d$, $\hat{\beta}_n^*$ has a closed form solution:

$$\begin{aligned} \hat{\beta}_n^* &= prox_{c\alpha_n\lambda_n\|\cdot\|_1} \left(\bar{\beta}_n - c\alpha_n\sqrt{n} \left(\hat{l}_n^*(\bar{\beta}_n) - \hat{l}_n(\bar{\beta}_n) \right) \right) \\ &= \left(\bar{\beta}_n - c\alpha_n\sqrt{n} \left(\hat{l}_n^*(\bar{\beta}_n) - \hat{l}_n(\bar{\beta}_n) \right) - c\alpha_n\lambda_n \right)^+ - \left(\bar{\beta}_n - c\alpha_n\sqrt{n} \left(\hat{l}_n^*(\bar{\beta}_n) - \hat{l}_n(\bar{\beta}_n) \right) + c\alpha_n\lambda_n \right)^- \end{aligned}$$

where $x^+ \equiv \max(x, 0)$ and $x^- \equiv -\min(x, 0)$.

1.2 ℓ_1-norm support vector regression

The ℓ_1-norm support vector regression (SVR) estimator of Zhu et al. (2004) is similar to the ℓ_1 penalized quantile regression estimator of Belloni and Chernozhukov (2011):

$$\hat{\beta}_n = \arg\min \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\rho_\tau \left(y_i - x_i'\beta\right) - \kappa\right)^+ + \frac{\lambda_n}{\sqrt{n}} \|\beta\|_1 \right\}$$

The objective uses a relaxed version of the check function:

$$\begin{aligned}
&\left(\rho_\tau \left(y_i - x_i'\beta\right) - \kappa\right)^+ \\
&= \left(\left\{(1-\tau) 1 \left(y_i - x_i'\beta \leqslant 0\right) + \tau 1 \left(y_i - x_i'\beta > 0\right)\right\} \left|y_i - x_i'\beta\right| - \kappa\right)^+ \\
&= \begin{cases} \left(-(1-\tau)\left(y_i - x_i'\beta\right) - \kappa\right) 1 \left(-(1-\tau)\left(y_i - x_i'\beta\right) - \kappa > 0\right) & , y_i - x_i'\beta \leqslant 0 \\ \left(\tau \left(y_i - x_i'\beta\right) - \kappa\right) 1 \left(\tau\left(y_i - x_i'\beta\right) - \kappa > 0\right) & , y_i - x_i'\beta > 0 \end{cases} \\
&= \begin{cases} \left(-(1-\tau)\left(y_i - x_i'\beta\right) - \kappa\right) 1 \left(y_i < x_i'\beta - \frac{\kappa}{1-\tau}\right) & , y_i - x_i'\beta \leqslant 0 \\ \left(\tau \left(y_i - x_i'\beta\right) - \kappa\right) 1 \left(y_i > x_i'\beta + \frac{\kappa}{\tau}\right) & , y_i - x_i'\beta > 0 \end{cases} \\
&= \left(\tau \left(y_i - x_i'\beta\right) - \kappa\right) 1 \left(y_i > x_i'\beta + \frac{\kappa}{\tau}\right) - \left((1-\tau)\left(y_i - x_i'\beta\right) + \kappa\right) 1 \left(y_i < x_i'\beta - \frac{\kappa}{1-\tau}\right)
\end{aligned}$$

The proximal bootstrap estimator is

$$\hat{\beta}_n^* = \arg\min_\beta \alpha_n \lambda_n \|\beta\|_1 + \alpha_n \sqrt{n} \left(\hat{l}_n^* \left(\bar{\beta}_n\right) - \hat{l}_n \left(\bar{\beta}_n\right)\right)' \left(\beta - \bar{\beta}_n\right) + \frac{1}{2} \left\|\beta - \bar{\beta}_n\right\|_{\bar{H}_n}^2$$

$\hat{l}_n \left(\bar{\beta}_n\right)$ is a consistent estimate of $l\left(\beta_0\right)$ using $\bar{\beta}_n$:

$$\begin{aligned}
l\left(\beta_0\right) &= -E\left[x_i \left(\tau 1 \left(y_i > x_i'\beta_0 + \frac{\kappa}{\tau}\right) - (1-\tau) 1 \left(y_i < x_i'\beta_0 - \frac{\kappa}{1-\tau}\right)\right)\right] \\
\hat{l}_n \left(\bar{\beta}_n\right) &= -\frac{1}{n} \sum_{i=1}^{n} x_i \left(\tau 1 \left(y_i > x_i'\bar{\beta}_n + \frac{\kappa}{\tau}\right) - (1-\tau) 1 \left(y_i < x_i'\bar{\beta}_n - \frac{\kappa}{1-\tau}\right)\right)
\end{aligned}$$

The population Hessian and its consistent estimate using $\bar{\beta}_n$ are given by

$$\begin{aligned}
H_0 &= E\left[x_i x_i' \left(\tau f_{y|x} \left(x_i'\beta_0 + \frac{\kappa}{\tau}\right) + (1-\tau) f_{y|x} \left(x_i'\beta_0 - \frac{\kappa}{1-\tau}\right)\right)\right] \\
\bar{H}_n &= \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \left(\tau \hat{f}_{y|x} \left(x_i'\bar{\beta}_n + \frac{\kappa}{\tau}\right) + (1-\tau) \hat{f}_{y|x} \left(x_i'\bar{\beta}_n - \frac{\kappa}{1-\tau}\right)\right)
\end{aligned}$$

An example of $\hat{f}_{y|x}(y)$ is $\frac{1}{n}\sum_{j=1}^{n} K_h(y)$, where $K_h(y) = \frac{1}{h} K(y/h)$ and $K(u)$ is a kernel function that is symmetric around 0 and integrates to 1.

1.3 Trace Regression via Nuclear Norm Regularization

$$\hat{\Theta}_n = \arg\min_{\Theta \in \mathbb{R}^{d_1 \times d_2}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - tr\left(\Theta' X_i\right)\right) + \lambda_n \|\Theta\|_* \right\}$$

where $\|\Theta\|_* = \sum_{j=1}^{d_1 \wedge d_2} \sigma_j(\Theta)$ is the nuclear norm of Θ, and $\sigma_j(\Theta)$ is the jth largest singular value of Θ.

$$\hat{\Theta}_n^* = \arg\min_{\Theta} \alpha_n \lambda_n \|\Theta\|_* + \alpha_n \sqrt{n} \left(\hat{l}_n^* \left(\bar{\Theta}_n \right) - \hat{l}_n \left(\bar{\Theta}_n \right) \right)' \left(\Theta - \bar{\Theta}_n \right) + \frac{1}{2} \left\| \Theta - \bar{\Theta}_n \right\|_{\bar{H}_n}^2$$

$$\hat{l}_n \left(\bar{\Theta}_n \right) = -\frac{1}{n} \sum_{i=1}^{n} X_i \left(y_i - tr \left(\bar{\Theta}_n' X_i \right) \right), \quad \bar{H}_n = \frac{1}{n} \sum_{i=1}^{n} X_i X_i'$$

$$r'_{\Theta_0}(h) = \sum_{j=1}^{d_1 \wedge d_2} \left(h_j 1 \left(\sigma_j(\Theta_0) \neq 0 \right) + |h_j| 1 \left(\sigma_j(\Theta_0) = 0 \right) \right)$$

In the case of $\bar{H}_n = \frac{1}{c} I_{d_1}$, the proximal bootstrap has a closed form:

$$\hat{\Theta}_n^* = prox_{c\alpha_n\lambda_n\|\cdot\|_*} \left(\bar{\Theta}_n - c\alpha_n \sqrt{n} \left(\hat{l}_n^* \left(\bar{\Theta}_n \right) - \hat{l}_n \left(\bar{\Theta}_n \right) \right) \right) = U \Sigma_{c\alpha_n\lambda_n} V^T$$

where $\Sigma_{c\alpha_n\lambda_n} = \text{diag}\left\{ \max\left(\Sigma_1 - c\alpha_n\lambda_n, 0 \right), \max\left(\Sigma_2 - c\alpha_n\lambda_n, 0 \right), ..., \max\left(\Sigma_{d_1 \wedge d_2} - c\alpha_n\lambda_n, 0 \right) \right\}$, and for $j = 1...d_1 \wedge d_2$, Σ_j are the singular values of $\bar{\Theta}_n - c\alpha_n \sqrt{n} \left(\hat{l}_n^* \left(\bar{\Theta}_n \right) - \hat{l}_n \left(\bar{\Theta}_n \right) \right)$.

2 Proof of Theorem 1

Assumption 1 implies that the conditions of part 2 of Corollary 3.2.3 of van der Vaart and Wellner (1996) are satisfied, and therefore $\hat{\beta}_n \xrightarrow{p} \beta_0 = \arg\min_{\beta \in \mathbb{R}^d} Q(\beta)$. To derive its asymptotic distribution, use the centered and scaled parameter $h = \sqrt{n}(\beta - \beta_0)$:

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}_n - \beta_0 \right) &= \arg\min_h \left\{ n\hat{Q}_n \left(\beta_0 + \frac{h}{\sqrt{n}} \right) - n\hat{Q}_n(\beta_0) + \lambda_n \sqrt{n} r \left(\beta_0 + \frac{h}{\sqrt{n}} \right) \right\} \\
&= \arg\min_h \left\{ h'\sqrt{n} \left(\hat{l}_n(\beta_0) - l(\beta_0) \right) + \frac{1}{2} h' H_0 h + \lambda_n \left(\frac{r\left(\beta_0 + \frac{h}{\sqrt{n}} \right) - r(\beta_0)}{1/\sqrt{n}} \right) + o_p(1) \right\} \\
&\rightsquigarrow \arg\min_h \left\{ \lambda_0 r'_{\beta_0}(h) + h' W_0 + \frac{1}{2} h' H_0 h \right\}
\end{aligned}$$

The second line is due to the uniform in h local quadratic expansion of $n\hat{Q}_n \left(\beta_0 + \frac{h}{\sqrt{n}} \right) - n\hat{Q}_n(\beta_0)$, which follows from assumption 2. The last line follows from the following arguments. Assumption 3 implies the Lindeberg Condition is satisfied and $\sqrt{n}(P_n - P) g(\cdot, \beta_0) \rightsquigarrow W_0$. Assumption 5 implies $\frac{r\left(\beta_0 + \frac{h}{\sqrt{n}} \right) - r(\beta_0)}{1/\sqrt{n}} \to r'_{\beta_0}(h)$ for each $h \in \mathbb{R}^d$ and that $r'_{\beta_0}(h)$ is a convex function of h. Since $h'\sqrt{n} \left(\hat{l}_n(\beta_0) - l(\beta_0) \right) + \frac{1}{2} h' H_0 h + \lambda_n \left(\frac{r\left(\beta_0 + \frac{h}{\sqrt{n}} \right) - r(\beta_0)}{1/\sqrt{n}} \right)$ is a convex function of h , pointwise convergence implies uniform convergence over compact sets $K \subset \mathbb{R}^d$ (Pollard (1991)). Therefore,

$$n\hat{Q}_n \left(\beta_0 + \frac{h}{\sqrt{n}} \right) - n\hat{Q}_n(\beta_0) + \lambda_n \sqrt{n} r \left(\beta_0 + \frac{h}{\sqrt{n}} \right) - \lambda_n \sqrt{n} r(\beta_0) \rightsquigarrow h' W_0 + \frac{1}{2} h' H_0 h + \lambda_0 r'_{\beta_0}(h)$$

as a process indexed by h in the space of bounded functions $\ell^{\infty}(K)$ for any compact $K \subset \mathbb{R}^d$. Convexity implies $\lambda_0 r'_{\beta_0}(h) + h'W_0 + \frac{1}{2}h'H_0h$ has a unique minimum, so by the argmin continuous mapping theorem (Theorem 3.2.2 in van der Vaart and Wellner (1996)), $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \rightsquigarrow \mathcal{J}$.

Now we show $\hat{\beta}_n^* \xrightarrow{p} \beta_0$. Since $\alpha_n \to 0$ and $\alpha_n\lambda_n \to 0$ imply $\alpha_n\lambda_n r(h+\beta_0) = o(1)$ and $\alpha_n\sqrt{n}\bar{H}_n\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right) = o_p^*(1)$,

$$\begin{aligned}
\hat{\beta}_n^* - \beta_0 &= \arg\min_h \left\{\alpha_n\lambda_n r(h+\beta_0) + \frac{1}{2}\left\|h + \beta_0 - \bar{\beta}_n + \alpha_n\sqrt{n}\bar{H}_n^{-1}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right)\right\|_{\bar{H}_n}^2\right\} \\
&= \arg\min_h \left\{\frac{1}{2}h'H_0h + h'H_0\left(\beta_0 - \bar{\beta}_n\right) + \frac{1}{2}\left\|\beta_0 - \bar{\beta}_n\right\|_{H_0}^2\right\} + o_p(1) \\
&= \bar{\beta}_n - \beta_0 + o_p(1) = o_p(1)
\end{aligned}$$

The second line follows from convexity of the proximal bootstrap objective function, which implies the difference between $\alpha_n\lambda_n r(h+\beta_0) + \frac{1}{2}\left\|h + \beta_0 - \bar{\beta}_n + \alpha_n\sqrt{n}\bar{H}_n^{-1}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right)\right\|_{\bar{H}_n}^2$ and $\frac{1}{2}\left\|h + \beta_0 - \bar{\beta}_n\right\|_{H_0}^2 = \frac{1}{2}h'H_0h + h'H_0\left(\beta_0 - \bar{\beta}_n\right) + \frac{1}{2}\left\|\beta_0 - \bar{\beta}_n\right\|_{H_0}^2$ converges uniformly in probability to zero over any compact subset of $\mathbb{R}^d$.

To derive $\hat{\beta}_n^*$'s asymptotic distribution, first note that because $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = O_p(1)$ and $\sqrt{n}\alpha_n \to \infty$,

$$\frac{\hat{\beta}_n^* - \hat{\beta}_n}{\alpha_n} = \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} - \frac{\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)}{\sqrt{n}\alpha_n} = \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} + o_p(1)$$

It therefore suffices to show that $\frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$. To do this, use the centered and scaled parameter $h = (\beta - \beta_0)/\alpha_n$:

$$\begin{aligned}
\frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} &= \arg\min_h \left\{\alpha_n\lambda_n r(\beta_0 + \alpha_n h) + \alpha_n\sqrt{n}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right)'\left(\beta_0 - \bar{\beta}_n + \alpha_n h\right) + \frac{1}{2}\left\|\beta_0 - \bar{\beta}_n + \alpha_n h\right\|_{\bar{H}_n}^2\right\} \\
&= \arg\min_h \left\{\lambda_n\left(\frac{r(\beta_0 + \alpha_n h) - r(\beta_0)}{\alpha_n}\right) + \sqrt{n}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right)'\left(\frac{\beta_0 - \bar{\beta}_n}{\alpha_n} + h\right) + \frac{1}{2}\left\|\frac{\beta_0 - \bar{\beta}_n}{\alpha_n} + h\right\|_{\bar{H}_n}^2\right\} \\
&= \arg\min_h \left\{\lambda_n\left(\frac{r(\beta_0 + \alpha_n h) - r(\beta_0)}{\alpha_n}\right) + h'\sqrt{n}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right) + \frac{1}{2}h'\bar{H}_n h + o_p^*(1)\right\} \\
&\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_h \left\{\lambda_0 r'_{\beta_0}(h) + h'W_0 + \frac{1}{2}h'H_0h\right\}
\end{aligned}$$

We have used $\frac{\beta_0 - \bar{\beta}_n}{\alpha_n} = \frac{\sqrt{n}\left(\beta_0 - \bar{\beta}_n\right)}{\sqrt{n}\alpha_n} = o_p(1)$, $\bar{H}_n \xrightarrow{p} H_0$, the assumption of directional differentiability of $r(\beta)$ at β_0, and the following arguments. Assumption 4(i) says $\mathcal{G}_R \equiv \{g(\cdot,\beta) - g(\cdot,\beta_0) : \|\beta - \beta_0\| \leqslant R\}$ is a Donsker class for some $R > 0$, and $P\left(g(\cdot,\beta) - g(\cdot,\beta_0)\right)^2 \to 0$ for $\beta \to \beta_0$. By Lemma 3.3.5 of van der Vaart and Wellner (1996), $\sqrt{n}(P_n - P)g(\cdot,\beta)$ is stochastically equicontinuous, which implies

$$\left\|\sqrt{n}(P_n - P)\left(g\left(\cdot,\bar{\beta}_n\right) - g(\cdot,\beta_0)\right)\right\| = o_p\left(1 + \sqrt{n}\left\|\bar{\beta}_n - \beta_0\right\|\right) = o_p(1)$$

Stochastic equicontinuity and the envelope integrability condition in assumption 4(ii) imply that the assumptions of Lemma 4.2 in Wellner and Zhan (1996) are satisfied. Therefore, $\sqrt{n}\left(P_n^* - P_n\right) g\left(\cdot, \beta\right)$ is bootstrap equicontinuous, which implies

$$\left\|\sqrt{n}\left(P_n^* - P_n\right)\left(g\left(\cdot, \bar{\beta}_n\right) - g\left(\cdot, \beta_0\right)\right)\right\| = o_p^*\left(1 + \sqrt{n}\left\|\bar{\beta}_n - \beta_0\right\|\right) = o_p^*(1)$$

Therefore, $h'\sqrt{n}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right) = h'\sqrt{n}\left(P_n^* - P_n\right) g\left(\cdot, \beta_0\right) + h'\sqrt{n}\left(P_n^* - P_n\right)\left(g\left(\cdot, \bar{\beta}_n\right) - g\left(\cdot, \beta_0\right)\right) + o_p^*\left(1\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} h'W_0$. By convexity, pointwise convergence implies uniform convergence over compact sets $K \subset \mathbb{R}^d$, so

$$\lambda_n\left(\frac{r\left(\beta_0 + \alpha_n h\right) - r\left(\beta_0\right)}{\alpha_n}\right) + h'\sqrt{n}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right) + \frac{1}{2}h'\bar{H}_n h \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \lambda_0 r'_{\beta_0}\left(h\right) + h'W_0 + \frac{1}{2}h'H_0 h$$

as a process indexed by h in the space of bounded functions $\ell^\infty\left(K\right)$ for any compact $K \subset \mathbb{R}^d$. $\frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$ follows from the bootstrap version of the argmin continuous mapping theorem (see Lemma 14.2 in Hong and Li (2020)). ■

Monte Carlo Simulation for Finite-dimensional Lasso

We consider the following data generating process:

$$y_i = x_i'\beta_0 + \epsilon_i, \quad \beta_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}', x_i \sim N\left(0, I_5 + 0.5\left(\iota\iota' - I_5\right)\right), \epsilon_i \sim N\left(0, 1\right)$$

We compute the Lasso estimator $\hat{\beta}_n = \arg\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^n\left(y_i - x_i'\beta\right)^2 + \frac{\lambda_n}{\sqrt{n}}\left\|\beta\right\|_1\right\}$ using the CVX modeling software in Matlab developed by Grant and Boyd (2009). The proximal bootstrap estimator $\hat{\beta}_n^* = \arg\min_{\beta} \alpha_n\lambda_n\left\|\beta\right\|_1 + \alpha_n\sqrt{n}\left(\hat{l}_n^*\left(\bar{\beta}_n\right) - \hat{l}_n\left(\bar{\beta}_n\right)\right)'\left(\beta - \bar{\beta}_n\right) + \frac{1}{2}\left\|\beta - \bar{\beta}_n\right\|_{\bar{H}_n}^2$, for $\bar{\beta}_n = \hat{\beta}_n$, $\bar{H}_n = \frac{1}{n}\sum_{i=1}^n x_i x_i'$, $\hat{l}_n\left(\bar{\beta}_n\right) = -\frac{1}{n}\sum_{i=1}^n x_i\left(y_i - x_i'\bar{\beta}_n\right)$, and $\hat{l}_n^*\left(\bar{\beta}_n\right) = -\frac{1}{n}\sum_{i=1}^n x_i^*\left(y_i^* - x_i^{*\prime}\bar{\beta}_n\right)$, is computed using the fminunc Matlab function so that we can run the code in parallel (the current version of CVX does not support parallel for loops). We also tried using the fmincon Matlab function, and the results were the same.

We consider five different sample sizes $n \in \{100, 500, 1000, 5000, 10000\}$, three different α_n's for each n: $\alpha_n \in \left\{n^{-1/3}, n^{-1/4}, n^{-1/6}\right\}$, and two choices of $\lambda_n \in \{0.1, 0.5\}$. We use 5000 bootstrap iterations and 2000 Monte Carlo simulations. Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals $\left[\hat{\beta}_n - \frac{c_{97.5}}{\sqrt{n}}, \hat{\beta}_n - \frac{c_{2.5}}{\sqrt{n}}\right]$, where c_τ is the τ-th percentile of $\frac{\hat{\beta}_n^* - \hat{\beta}_n}{\alpha_n}$, and average interval lengths are reported in tables 1-3. Although the proximal bootstrap undercovers for smaller sample sizes, it achieves coverage very close to 95% for sufficiently large n.

Table 1: Proximal Bootstrap Coverage Frequencies and Interval Lengths for $\alpha_n = n^{-1/3}$

	$\lambda_n = 0.1$					$\lambda_n = 0.5$				
n	100	500	1000	5000	10000	100	500	1000	5000	10000
	0.940	0.940	0.945	0.957	0.951	0.933	0.933	0.938	0.958	0.950
	(0.489)	(0.222)	(0.157)	(0.070)	(0.050)	(0.450)	(0.204)	(0.145)	(0.065)	(0.046)
	0.922	0.944	0.946	0.946	0.947	0.919	0.940	0.942	0.950	0.949
	(0.458)	(0.209)	(0.147)	(0.066)	(0.047)	(0.308)	(0.143)	(0.101)	(0.045)	(0.032)
	0.935	0.945	0.942	0.953	0.954	0.934	0.944	0.939	0.953	0.945
	(0.459)	(0.208)	(0.147)	(0.066)	(0.047)	(0.308)	(0.143)	(0.101)	(0.046)	(0.032)
	0.933	0.935	0.948	0.953	0.949	0.936	0.938	0.940	0.945	0.945
	(0.456)	(0.208)	(0.147)	(0.066)	(0.047)	(0.306)	(0.142)	(0.101)	(0.045)	(0.032)
	0.929	0.947	0.953	0.939	0.950	0.936	0.949	0.951	0.938	0.945
	(0.457)	(0.208)	(0.148)	(0.066)	(0.047)	(0.306)	(0.143)	(0.102)	(0.045)	(0.032)

Table 2: Proximal Bootstrap Coverage Frequencies and Interval Lengths for $\alpha_n = n^{-1/4}$

	$\lambda_n = 0.1$					$\lambda_n = 0.5$				
n	100	500	1000	5000	10000	100	500	1000	5000	10000
	0.930	0.940	0.945	0.957	0.952	0.888	0.935	0.940	0.958	0.952
	(0.485)	(0.222)	(0.157)	(0.070)	(0.050)	(0.425)	(0.204)	(0.145)	(0.065)	(0.046)
	0.921	0.944	0.946	0.946	0.948	0.921	0.942	0.943	0.950	0.950
	(0.458)	(0.209)	(0.147)	(0.066)	(0.047)	(0.306)	(0.143)	(0.101)	(0.045)	(0.032)
	0.936	0.945	0.943	0.953	0.954	0.934	0.945	0.939	0.953	0.946
	(0.458)	(0.208)	(0.147)	(0.066)	(0.047)	(0.307)	(0.143)	(0.101)	(0.045)	(0.032)
	0.933	0.935	0.948	0.954	0.950	0.933	0.939	0.940	0.947	0.944
	(0.455)	(0.208)	(0.147)	(0.066)	(0.047)	(0.304)	(0.142)	(0.101)	(0.045)	(0.032)
	0.929	0.947	0.953	0.939	0.950	0.938	0.950	0.952	0.938	0.945
	(0.456)	(0.208)	(0.147)	(0.066)	(0.047)	(0.304)	(0.143)	(0.101)	(0.045)	(0.032)

Table 3: Proximal Bootstrap Coverage Frequencies and Interval Lengths for $\alpha_n = n^{-1/6}$

			$\lambda_n = 0.1$					$\lambda_n = 0.5$		
n	100	500	1000	5000	10000	100	500	1000	5000	10000
	0.913	0.934	0.946	0.958	0.952	0.784	0.902	0.929	0.958	0.953
	(0.462)	(0.220)	(0.157)	(0.070)	(0.050)	(0.349)	(0.190)	(0.143)	(0.065)	(0.046)
	0.921	0.944	0.946	0.946	0.948	0.919	0.941	0.943	0.951	0.950
	(0.457)	(0.208)	(0.147)	(0.066)	(0.047)	(0.302)	(0.142)	(0.101)	(0.045)	(0.032)
	0.934	0.946	0.943	0.953	0.954	0.930	0.944	0.939	0.953	0.946
	(0.458)	(0.208)	(0.147)	(0.066)	(0.047)	(0.303)	(0.142)	(0.101)	(0.045)	(0.032)
	0.933	0.936	0.949	0.953	0.950	0.933	0.937	0.941	0.948	0.945
	(0.455)	(0.207)	(0.147)	(0.066)	(0.047)	(0.300)	(0.142)	(0.101)	(0.045)	(0.032)
	0.928	0.948	0.954	0.940	0.950	0.939	0.951	0.952	0.938	0.947
	(0.456)	(0.208)	(0.147)	(0.066)	(0.047)	(0.301)	(0.142)	(0.101)	(0.045)	(0.032)

We also compare the proximal bootstrap to the standard multinomial bootstrap estimator $\hat{\beta}_n^{**} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left(y_i^* - x_i^{*\prime} \beta \right)^2 + \frac{\lambda_n}{\sqrt{n}} \|\beta\|_1 \right\}$. Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals $\left[\hat{\beta}_n - \frac{d_{97.5}}{\sqrt{n}}, \hat{\beta}_n - \frac{d_{2.5}}{\sqrt{n}} \right]$, where d_τ is the τ-th percentile of $\sqrt{n} \left(\hat{\beta}_n^{**} - \hat{\beta}_n \right)$, and average interval lengths are reported in table 4. We use 5000 bootstrap iterations and 2000 Monte Carlo simulations. Interestingly, for the case of $\lambda_n = 0.1$, the standard bootstrap coverage frequencies are close to the nominal level. The surprisingly good coverage of the standard bootstrap under certain DGPs is also documented in section 6.2 of Chatterjee and Lahiri (2011). However, when we use $\lambda_n = 0.5$, the standard bootstrap undercovers for the nonzero parameter and overcovers for the zero parameters. Additionally, the standard bootstrap confidence intervals are on average wider than the proximal bootstrap confidence intervals.

Table 4: Standard Bootstrap Coverage Frequencies and Interval Lengths

			$\lambda_n = 0.1$					$\lambda_n = 0.5$		
n	100	500	1000	5000	10000	100	500	1000	5000	10000
	0.947	0.944	0.945	0.961	0.953	0.915	0.917	0.914	0.926	0.920
	(0.509)	(0.224)	(0.158)	(0.071)	(0.050)	(0.474)	(0.211)	(0.149)	(0.067)	(0.047)
	0.950	0.960	0.959	0.966	0.965	0.982	0.986	0.983	0.987	0.991
	(0.477)	(0.211)	(0.149)	(0.067)	(0.047)	(0.329)	(0.150)	(0.107)	(0.048)	(0.034)
	0.963	0.961	0.961	0.969	0.966	0.987	0.991	0.985	0.989	0.990
	(0.478)	(0.211)	(0.149)	(0.067)	(0.047)	(0.328)	(0.151)	(0.107)	(0.048)	(0.034)
	0.959	0.955	0.964	0.967	0.964	0.989	0.982	0.986	0.990	0.991
	(0.474)	(0.211)	(0.149)	(0.067)	(0.047)	(0.327)	(0.151)	(0.107)	(0.048)	(0.034)
	0.958	0.967	0.966	0.954	0.965	0.987	0.993	0.989	0.989	0.993
	(0.476)	(0.211)	(0.149)	(0.067)	(0.047)	(0.327)	(0.150)	(0.107)	(0.048)	(0.034)

References

Belloni, Alexandre and Victor Chernozhukov, "ℓ_1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 2011, *39* (1), 82–130. 2

Chatterjee, Arindam and Soumendra Nath Lahiri, "Bootstrapping lasso estimators," *Journal of the American Statistical Association*, 2011, *106* (494), 608–625. 7

Grant, Michael and Stephen Boyd, "CVX: Matlab software for disciplined convex programming," 2009. 5

Hong, Han and Jessie Li, "The numerical bootstrap," *The Annals of Statistics*, 2020, *48* (1), 397–412. 5

Pollard, David, "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, 1991, *7* (2), 186–199. 3

van der Vaart, AW and Jon A Wellner, *Weak Convergence and Empirical Processes*, Springer, 1996. 3, 4

Wellner, Jon A and Yihui Zhan, "Bootstrapping Z-estimators," *University of Washington Department of Statistics Technical Report*, 1996, *308.* 5

Zhu, Ji, Saharon Rosset, Robert Tibshirani, and Trevor J Hastie, "1-norm support vector machines," in "Advances in neural information processing systems" 2004, pp. 49–56. 2