The supplementary material for “Sparse quantile regression via ℓ_0-penalty”

Toshio Honda and Wei-Ying Wu

The proofs of Propositions 1-3 and Lemma 1 are given in this supplementary material.

S.1 Appendix(to be the supplementary material)

We prove Propositions 1-3 and Lemma 1 here.

The proof of Proposition 1 is given here.

Proof of Proposition 1:

The proof is similar to those of Lemma 1 of Fan et al. (2014) and Lemma 4 of Honda et al. (2019). However, $\Gamma_s(M)$ is more general than in those lemmas due to $\mathcal{S}$ in $(\gamma_{\mathcal{S}}, \mathcal{S})$ and the proof is more complicated.

We omit $\mathcal{S}$ in $(\boldsymbol{\gamma}_{\mathcal{S}}, \mathcal{S}) \in \Gamma_s(M)$ for notational simplicity and note that $\boldsymbol{W}_{i\mathcal{S}}^T \gamma_{\mathcal{S}}^* = \boldsymbol{W}_{i\mathcal{S}_0}^T \gamma_{\mathcal{S}_0}^*$.

Due to the Lipschitz continuity of $\rho_\tau(u)$ and application of the concentration inequalities (Theorems 14.3 and 14.4 in Bühlmann and van de Geer (2011))), we have

$$\begin{aligned}
\mathrm{E}\{G_s(M)\} &\leq 2\mathrm{E}\Big[\sup_{\boldsymbol{\gamma}_{\mathcal{S}} \in \Gamma_s(M)} \Big| \frac{1}{n} \sum_{i=1}^n \xi_i \{\rho_\tau(Y_i - \boldsymbol{W}_{i\mathcal{S}}^T \boldsymbol{\gamma}_{\mathcal{S}}) - \rho_\tau(Y_i - \boldsymbol{W}_{i\mathcal{S}}^T \boldsymbol{\gamma}_{\mathcal{S}}^*)\} \Big| \Big] \\
&\leq 4\mathrm{E}\Big[\sup_{\boldsymbol{\gamma}_{\mathcal{S}} \in \Gamma_s(M)} \Big| \frac{1}{n} \sum_{i=1}^n \xi_i \boldsymbol{W}_{i\mathcal{S}}^T (\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*) \Big| \Big], \qquad \text{(S.1)}
\end{aligned}$$

where $\{\xi_j\}_{j=1}^n$ is a Rademacher sequence of and independent of $\{(Y_j, \boldsymbol{X}_j)\}_{j=1}^n$.

Notice that

$$\begin{aligned}
\Big| \frac{1}{n} \sum_{i=1}^n \xi_i \boldsymbol{W}_{i\mathcal{S}}^T (\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*) \Big| &\leq \max_{j \in \mathcal{S}} \Big\| \frac{1}{n} \sum_{i=1}^n \xi_i \boldsymbol{W}_{ij} \Big\| \times \sum_{j \in \mathcal{S}} \| \boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^* \| \qquad \text{(S.2)} \\
&\leq L^{1/2} \max_{j,l} \Big| \frac{1}{n} \sum_{i=1}^n \xi_i W_{ijl} \Big| \times \sqrt{s} M
\end{aligned}$$

and

$$\mathrm{E}\Big\{ \max_{j,l} \Big| \frac{1}{n} \sum_{i=1}^n \xi_i W_{ijl} \Big| \Big\} \leq C_1 \sqrt{\frac{2 \log p_n}{nL} C_{Sp}^2} \qquad \text{(S.3)}$$

for some universal constant C_1 by Lemmas 2.2.10 in van der Vaart and Wellner (1996) and the facts on our spline basis. Note that j in the second line in (S.2) and in (S.3) ranges from 1 to p, not restricted to $\mathcal{S}$. We applied Bernstein's inequality to $\sum_{i=1}^n \xi_i W_{ijl}$ like Lemmas 2.2.9 and 2.2.11 before we used Lemmas 2.2.10.

Thus by (S.1)-(S.3), we have

$$\mathrm{E}\{G_s(M)\} \leq 4\sqrt{sL} M \times C_1 \sqrt{\frac{2 \log p_n}{nL} C_{Sp}^2} \leq 4\sqrt{2} C_1 M \sqrt{\frac{s \log p_n C_{Sp}^2}{n}}. \qquad \text{(S.4)}$$

Next we apply Massart's inequality (Theorem 14.2 in Bühlmann and van de Geer (2011)) to evaluate the stochastic part $G_s(M) - \mathrm{E}\{G_s(M)\}$. Write C_2 for $4\sqrt{2} C_1$ in (S.4).

Before we apply Massart's inequality, notice that

$$|\boldsymbol{W}_{i\mathcal{S}}^T(\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*)|^2 \le \|\boldsymbol{W}_{i\mathcal{S}}\|^2 \|\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*\|^2 \le sC_{Sp}^2 M^2$$

and

$$\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{W}_{i\mathcal{S}}\|^2 M^2 \le sC_{Sp}^2 M^2.$$

Thus as in the proof of Lemma 1 of Fan et al. (2014), we have

$$\mathrm{P}\Big(G_s(M) \ge C_2 M\sqrt{\frac{sC_{Sp}^2 \log p_n}{n}} + t\Big) \le \exp\Big(-\frac{nt^2}{8sC_{Sp}^2 M^2}\Big).$$

We used (S.4) to evaluate $\mathrm{E}\{G_s(M)\}$ inside the probability of Massart's inequality.

Taking $t^2 = K_A^2 M^2 sC_{Sp}^2 \log p_n/n$, we obtain

$$\mathrm{P}\Big(G_s(M) \ge (C_2 + K_A) M\sqrt{\frac{sC_{Sp}^2 \log p_n}{n}}\Big) \le \exp\Big(-\frac{K_A^2 \log p_n}{8}\Big).$$

Hence the proof of the proposition is complete. □

The proof of Proposition 2 is given here.

Proof of Proposition 2:

We follow those of Theorem 1 of Fan et al. (2014) and Proposition 1 of Honda et al. (2019).

The following arguments uniformly apply to any $\mathcal{S}$ satisfying $|\mathcal{S}| = s$ and $\mathcal{S}_0 \subset \mathcal{S}$. Then $G_s(M)$ appears only in (S.12) and we apply Proposition 1 to that $G_s(M)$ there. Besides, we can choose suitable K_A and K_B as in (S.13) and (S.14) to ensure the uniformity in s satisfying $s_0 \le s \le s_0 + k_1$. This means that we are dealing with $\mathcal{S}$ uniformly satisfying the condition in the proposition.

Taking $M_s = K_B L\sqrt{sC_{Sp}^2 \log p_n/n}$, we evaluate the following expression on $\Gamma_s(M_s)$. Note that this M_s depends on $s = |\mathcal{S}|$.

$$\mathrm{E}\{R_n(\boldsymbol{\gamma}_{\mathcal{S}}) - R_n(\boldsymbol{\gamma}_{\mathcal{S}}^*)\} = \mathrm{E}\Big[\frac{1}{n}\sum_{i=1}^n \{\rho_\tau(\epsilon_i' - a_i) - \rho_\tau(\epsilon_i')\}\Big], \tag{S.5}$$

where $a_i = \boldsymbol{W}_{i\mathcal{S}}^T(\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*)$ and recall the symbols in (17). Besides, by Assumption A5(1),

$$|a_i| \le \|\boldsymbol{W}_{i\mathcal{S}}\| M_s \le \sqrt{s} C_{Sp} M_s \to 0. \tag{S.6}$$

If $a_i > 0$, we have from the definition of $\rho_\tau(\cdot)$ that

$$\rho_\tau(\epsilon_i' - a_i) - \rho_\tau(\epsilon_i') = \int_0^{a_i} I\{0 < \epsilon_i' \le s\} ds + a_i(I\{\epsilon_i' \le 0\} - \tau).$$

Then by Assumption A2, we obtain

$$\begin{aligned}
&\mathrm{E}_\epsilon\Big[\int_0^{a_i} I\{0<\epsilon_i'\le s\}ds + a_i(I\{\epsilon_i'\le 0\}-\tau)\Big]\\
&=\int_0^{a_i}(F_i(s-\delta_i)-F_i(-\delta_i))ds + a_i(\tau_i-\tau)\\
&=\frac{1}{2}f_i(-\delta_i)a_i^2+o(a_i^2)+O(a_i^2(\log n)^{-1/2})+O(|\tau-\tau_i|^2(\log n)^{1/2}).
\end{aligned}$$

uniformly in i. Note that $|\tau-\tau_i|^2\le C_1|\delta_i|^2$ for some positive C_1 and that we can deal with the case of $a_i<0$ in the same way. Note that this kind of approximation is not accurate enough for the proof of Theorem 2 in the second setup since $s_0\to\infty$.

To evaluate (S.5), we need to consider

$$\frac{1}{2n}\sum_{i=1}^n f_i(-\delta_i)a_i^2 + o\Big(n^{-1}\sum_{i=1}^n a_i^2\Big)+O\Big(n^{-1}(\log n)^{1/2}\sum_{i=1}^n\delta_i^2\Big). \quad \text{(S.7)}$$

As for the first term of (S.7), we have

$$\begin{aligned}
\mathrm{E}\Big\{\frac{1}{2n}\sum_{i=1}^n f_i(-\delta_i)a_i^2\Big\}&=\frac{1}{2}(\boldsymbol{\gamma}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}^*)^T\mathrm{E}\Big\{\frac{1}{n}\sum_{i=1}^n f_i(-\delta_i)\boldsymbol{W}_{i\mathcal{S}}\boldsymbol{W}_{i\mathcal{S}}^T\Big\}(\boldsymbol{\gamma}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}^*)\\
&\ge\frac{\kappa_m}{2L}\|\boldsymbol{\gamma}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}^*\|^2.
\end{aligned} \quad \text{(S.8)}$$

Recall Assumption A3 for the inequality in the second line here.

As for the third term of (S.7), we have

$$\frac{(\log n)^{1/2}}{n}\sum_{i=1}^n\delta_i^2\le C_1(\log n)^{1/2}L^{-4} \quad \text{(S.9)}$$

for some positive constant C_1 by Assumption A1.

By combining (S.7), (S.8), and (S.9), we obtain

$$\mathrm{E}\{R_n(\boldsymbol{\gamma}_{\mathcal{S}})-R_n(\boldsymbol{\gamma}_{\mathcal{S}}^*)\}\ge\frac{\kappa_m}{2L}(1+o(1))\|\boldsymbol{\gamma}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}^*\|^2+O((\log n)^{1/2}L^{-4}). \quad \text{(S.10)}$$

Note that the above inequality holds only locally, not globally in $\boldsymbol{\gamma}_{\mathcal{S}}$.

We define $\boldsymbol{\gamma}_{\mathcal{S}}^\alpha$ by

$$\boldsymbol{\gamma}_{\mathcal{S}}^\alpha=\alpha\widetilde{\boldsymbol{\gamma}}_{\mathcal{S}}+(1-\alpha)\boldsymbol{\gamma}_{\mathcal{S}}^* \quad \text{(S.11)}$$

for

$$0\le\alpha=\frac{M_s}{M_s+\|\widetilde{\boldsymbol{\gamma}}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}^*\|}\le 1.$$

Then for the same $\mathcal{S}$,

$$\boldsymbol{\gamma}_{\mathcal{S}}^\alpha\in\mathit{\Gamma}_s(M_s).$$

Noticing that due to the convexity of $R_n(\boldsymbol{\gamma}_{\mathcal{S}})$,

$$R_n(\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha}) \leq \alpha R_n(\widetilde{\boldsymbol{\gamma}}_{\mathcal{S}}) + (1-\alpha) R_n(\boldsymbol{\gamma}_{\mathcal{S}}^{*}) \leq R_n(\boldsymbol{\gamma}_{\mathcal{S}}^{*}),$$

and applying Proposition 1, we have with probability larger than or equal to $1-\exp(-K_A^2 \log p_n/8)$ that

$$\begin{aligned}
&\mathrm{E}[R_n(\boldsymbol{\gamma}_{\mathcal{S}}) - R_n(\boldsymbol{\gamma}_{\mathcal{S}}^{*})]_{\boldsymbol{\gamma}_{\mathcal{S}}=\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha}} \qquad \text{(S.12)}\\
&= \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\boldsymbol{\gamma}_{\mathcal{S}}^{*}) - \mathrm{E}\Big\{\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\boldsymbol{\gamma}_{\mathcal{S}}^{*})\Big\} - \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha}) + \mathrm{E}\Big[\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\boldsymbol{\gamma}_{\mathcal{S}})\Big]_{\boldsymbol{\gamma}_{\mathcal{S}}=\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha}}\\
&\qquad + R_n(\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha}) - R_n(\boldsymbol{\gamma}_{\mathcal{S}}^{*})\\
&\leq G_s(M_s) \leq (C_{Pr1}+K_A) M_s \sqrt{\frac{sC_{Sp}^2 \log p_n}{n}}.
\end{aligned}$$

By (S.10) and (S.12), we have

$$\begin{aligned}
\|\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha} - \boldsymbol{\gamma}_{\mathcal{S}}^{*}\|^2 &\leq \frac{3(C_{Pr1}+K_A)L}{\kappa_m}\Big\{M_s \sqrt{\frac{sC_{Sp}^2 \log p_n}{n}} + o\Big(\frac{L\log p_n}{n}\Big)\Big\}\\
&\leq \frac{4(C_{Pr1}+K_A)L}{\kappa_m} M_s \sqrt{\frac{sC_{Sp}^2 \log p_n}{n}}
\end{aligned}$$

with probability larger than or equal to $1-\exp(-K_A^2 \log p_n/8)$.

We should choose K_A satisfying

$$k_1 \exp\Big(-\frac{K_A^2 \log p_n}{8}\Big) = O(n^{-K_C}) \qquad \text{(S.13)}$$

and then choose K_B satisfying

$$\frac{16(C_{Pr1}+K_A)}{\kappa_m} < K_B. \qquad \text{(S.14)}$$

Note that (S.13) gives an upper bound of the sum of probability from $s=s_0$ to $s=s_0+k_1$.

Then we have with probability $1-O(n^{-K_C})$ that uniformly in $\mathcal{S}$,

$$\|\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha} - \boldsymbol{\gamma}_{\mathcal{S}}^{*}\|^2 \leq \frac{1}{4}M_s^2 \quad \text{and} \quad \|\boldsymbol{\gamma}_{\mathcal{S}}^{\alpha} - \boldsymbol{\gamma}_{\mathcal{S}}^{*}\| \leq \frac{1}{2}M_s. \qquad \text{(S.15)}$$

(S.11), (S.15), and simple algebra yield

$$\|\widetilde{\boldsymbol{\gamma}}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^{*}\| \leq M_s = K_B L \sqrt{sC_{Sp}^2 \log p_n/n}$$

with probability $1-(n^{-K_C})$. We write C_{Pr2} for $K_B C_{Sp}$.

Hence the proof of the proposition is complete. □

The proof of Proposition 3 is given here.

Proof of Proposition 3:

We give another expression of $D_i(\boldsymbol{\gamma}_{\mathcal{S}})$ as

$$D_i(\boldsymbol{\gamma}_{\mathcal{S}}) = \overline{D}_i(\boldsymbol{\gamma}_{\mathcal{S}}) - \mathrm{E}\{\overline{D}_i(\boldsymbol{\gamma}_{\mathcal{S}})\}, \tag{S.16}$$

where

$$\overline{D}_i(\boldsymbol{\gamma}_{\mathcal{S}}) = \rho_\tau(Y_i - \boldsymbol{W}_{i\mathcal{S}}^T\boldsymbol{\gamma}_{\mathcal{S}}) - \rho_\tau(Y_i - \boldsymbol{W}_{i\mathcal{S}}^T\boldsymbol{\gamma}_{\mathcal{S}}^*) + \boldsymbol{W}_{i\mathcal{S}}^T(\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*)(\tau - I\{\epsilon_i' \le 0\})$$

and notice that

$$\rho_\tau(\epsilon_i' - a_i) - \rho_\tau(\epsilon_i') = -a_i(\tau - I\{\epsilon_i' \le 0\}) - (\epsilon_i' - a_i)[I\{\epsilon_i' \le a_i\} - I\{\epsilon_i' \le 0\}], \tag{S.17}$$

where $a_i = \boldsymbol{W}_{i\mathcal{S}}^T(\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*)$. The second term of the RHS in (S.17) is Lipschitz continuous. The argument based on the Lipschitz continuity, the symmetrization theorem, and the contraction theorem does not yield the desired order and we have to reply on (S.20) below.

By using (S.17) and the facts on our spline basis, we have the following three facts uniformly in $\boldsymbol{\gamma}_{\mathcal{S}}$ and $\mathcal{S}$. C_1 and C_2 are some positive constants.

$$\max_{1\le i\le n} \|\boldsymbol{W}_{i\mathcal{S}}\| \le C_{Sp}\sqrt{|\mathcal{S}|} \tag{S.18}$$

$$\max_{1\le i\le n} |\overline{D}_i(\boldsymbol{\gamma}_{\mathcal{S}})| \le \max_{1\le i\le n} \|\boldsymbol{W}_{i\mathcal{S}}\| C_{Pr2} L(|\mathcal{S}|\log p_n/n)^{1/2} \tag{S.19}$$

$$\begin{aligned}
\mathrm{E}\Big[\frac{1}{n^2}\sum_{i=1}^n\{\overline{D}_i(\boldsymbol{\gamma}_{\mathcal{S}})\}^2\Big] &\le \mathrm{E}\Big\{\frac{C_1}{n^2}\sum_{i=1}^n |\boldsymbol{W}_{i\mathcal{S}}^T(\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*)|^3\Big\} \qquad \text{(S.20)}\\
&\le \frac{C_1}{n} C_{Sp}\sqrt{|\mathcal{S}|}\{C_{Pr2}L(|\mathcal{S}|\log p_n/n)^{1/2}\}^3 \lambda_{\max}(\Omega_{\mathcal{S}})\\
&\le \frac{C_2 C_{Sp} C_{Pr2}^{3/2}\lambda_{2M}}{n} s_0^2 L^2 (n^{-1}\log p_n)^{3/2}.
\end{aligned}$$

We used Assumptions A2 and A6 here. The second term of the RHS in (S.17) is crucial to the first line of (S.20). We may be able to improve (S.20) and replace s_0^2 with $s_0^{3/2}$ for linear models. Then $n^{1/6}$ in (19) is replaced with $n^{1/5}$. However, considering the upper bound in (S.20) and the necessity of uniformity in $\boldsymbol{\gamma}_{\mathcal{S}}$ and $\mathcal{S}$, further improvements seem to be impossible.

By using (S.18)-(S.20) and Bernstein's inequality, we obtain

$$\mathrm{P}\Big(\Big|n^{-1}\sum_{i=1}^n D_i(\boldsymbol{\gamma}_{\mathcal{S}})\Big| \ge \frac{L\log p_n}{n\log n}\Big) \le C_3 \exp\Big\{-\frac{C_4(n\log p_n)^{1/2}}{s_0^2(\log n)^2}\Big\} \tag{S.21}$$

for any fixed $\boldsymbol{\gamma}_{\mathcal{S}}$ and $\mathcal{S}$, where C_3 and C_4 are some positive constants which are independent of $\boldsymbol{\gamma}_{\mathcal{S}}$ and $\mathcal{S}$.

To verify the uniformity in $\boldsymbol{\gamma}_{\mathcal{S}}$, we appeal to the standard small-block argument based on the Lipschitz continuity and the number of blocks is $O(\exp(C_5 s_0 L\log n))$. To obtain the uniformity in $\mathcal{S}$, we have to calculate the number of $\mathcal{S}$ and it is $O(\exp(C_6 s_0 \log p_n))$.

This implies that we should have

$$(s_0 L \log n) \vee (s_0 \log p_n) = o\Big(\frac{(n \log p_n)^{1/2}}{s_0^2 (\log n)^2}\Big) \quad \text{and} \quad \frac{(n \log p_n)^{1/2}}{s_0^2 (\log n)^3} \to \infty \quad \text{(S.22)}$$

to establish the statement for Proposition 3. Actually they are assured by Assumption A5', especially the former in (S.22) by Assumption A5'(2). Hence the proof of the proposition is complete. □

The proof of Lemma 1 is given here.

Proof of Lemma 1:

With $a_i = (\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*)^T \boldsymbol{W}_{i\mathcal{S}}$, we have

$$\frac{1}{n} \sum_{i=1}^{n} |a_i(\tau_i - \tau)| \le \Big(n^{-1} \sum_{i=1}^{n} a_i^2\Big)^{1/2} \Big(n^{-1} \sum_{i=1}^{n} (\tau_i - \tau)^2\Big)^{1/2}. \quad \text{(S.23)}$$

Note that

$$\frac{1}{n} \sum_{i=1}^{n} a_i^2 \le \lambda_{\max}(\widehat{\Omega}_{\mathcal{S}}) \|\boldsymbol{\gamma}_{\mathcal{S}} - \boldsymbol{\gamma}_{\mathcal{S}}^*\|^2 = O_p\Big(\frac{s_0 L \log p_n}{n}\Big) \quad \text{(S.24)}$$

uniformly in $\mathcal{S}$, where $\widehat{\Omega}_{\mathcal{S}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{W}_{i\mathcal{S}} \boldsymbol{W}_{i\mathcal{S}}^T$ and we applied the standard argument based on Bernstein's inequality with Assumption A6 to get an upper bound of $\lambda_{\max}(\widehat{\Omega}_{\mathcal{S}})$.

By Assumption A2', we also have

$$\frac{1}{n} \sum_{i=1}^{n} (\tau_i - \tau)^2 = O_p(L^{-6}). \quad \text{(S.25)}$$

Then the desired result follows from (S.23)-(S.25) and Assumption A5'(2).
□

References

Fan, J., Fan, Y. and Barut, E. (2014). Adaptive robust variable selection. *The Annals of Statistics* 42, 324–351.

van der Vaart, A. D. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.