Received: 25 March 2021 | Revised: 14 April 2021 | Accepted: 20 April 2021

DOI: 10.1002/sta4.386

ORIGINAL ARTICLE

Random projections for quantile ridge regression

Yan Zhou[1] | Jiang Liang[1] | Yaohua Hu[1] | Heng Lian[2]

[1]College of Mathematics and Statistics, Shenzhen University, Shenzhen, 518060, China

[2]Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

Correspondence
Heng Lian, Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong.
Email: henglian@cityu.edu.hk

Quantile regression estimate gives more complete information about the response distribution but is more costly to compute than mean regression. When the dimension is large, a ridge penalty is conventionally used to stabilize the estimate and achieve better bias-variance trade-off. We investigate a random projection approach to ease the computational burden and establish its statistical properties. Monte Carlo studies are carried out to illustrate the computational and statistical properties of the estimates.

KEYWORDS
dimension reduction, linear quantile regression, random projection, ridge regression

1 | INTRODUCTION

Quantile regression was proposed by Koenker and Bassett Jr (1978) and gives more complete information about the conditional response distribution than traditional mean regression. The popular implementation for standard linear quantile regression is based on converting the problem to a linear programming. In modern statistics, one is often concerned with problems containing a large number of predictors, in which case standard regression methodologies may become unstable or even infeasible. This stability issue can be rigorously quantified by a decomposition of the mean squared error into a bias term and a variance term, and in high-dimensional problems, the variance is dramatically inflated. Motivated by this, Hoerl and Kennard (1970) used ridge regression for mean regression with a quadratic penalty in which a parameter in the penalty can be tuned to achieve the optimal bias-variance trade-off. Using penalties to stabilize estimates has since become prevalent in the statistics and machine learning literature. Following the pioneering work of Fan and Li (2001) and Tibshirani (1996), an appropriately constructed penalty can also simultaneously serve the purpose of variable selection. In this paper, we only focus on the ridge penalty, whereas it is an interesting problem to investigate similar strategies for penalties that can perform variable selection.

For quantile regression, Yi and Huang (2017) recently proposed an efficient semi-smooth coordinate descent algorithm to compute the elastic-net estimator, with the ridge estimator being a special case. The complexity of the semi-smooth algorithm is $O(np)$ per iteration, which is still slow when both n and p are large. In this work, we consider the approximation of the linear quantile ridge regression estimator via random projection. Random projection is a classical technique for reducing storage and computational costs in various settings (Liu et al. 2019; Wang et al. 2013). Applications of random projection to vision problems have been popularly studied (Anand et al. 2012; Bingham & Mannila, 2001; Mu et al. 2011). Maillard (2012) considered random projection for standard nonpenalized linear regression and derived its excess risk bound, whereas Zhang et al. (2014) used a dual random projection approach for classification problems. The methodology we propose here is relatively straightforward, constructed by generating s random linear combinations of the original p predictors with $s < p$ and using these s linear combinations as the new predictors. Thus, the dimension of the quantile regression problem reduces from p to s. We establish some upper bound for the approximate estimator which suggests that the method works well when the covariate matrix is approximately low rank.

Related to this work, Zhang et al. (2021) considered random projection for nonparametric quantile regression when the sample size is large, whereas we consider the high-dimensional case in a parametric setting. Note that the technical aspects for nonparametric regression and high-dimensional regression are quite different. In particular, the current setting requires careful analysis of the spectrum of the covariance matrix of the predictors. Our work can thus be considered to be complementary to the work in the nonparametric setting.

2 | RANDOM PROJECTION METHOD

We consider the following optimization problem for quantile regression with a ridge penalty:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) + n\lambda \|\boldsymbol{\beta}\|^2,$$

where $\rho_\tau(x) = x(\tau - I\{x \le 0\})$ is the check loss function used for quantile regression, $(y_i, \mathbf{x}_i), i = 1, \ldots, n$ are independent copies of $(y, \mathbf{x})$ with p-dimensional predictor $\mathbf{x}$ and the response variable y, $\tau \in (0, 1)$ is the level of quantile under investigation and $\lambda > 0$ is a tuning parameter controlling the trade-off between bias and variance. In quantile regression, we assume that $\epsilon := y - \mathbf{x}^\top \boldsymbol{\beta}_0$ satisfies $P(\epsilon \le 0|\mathbf{x}) = \tau$ with $\boldsymbol{\beta}_0$ denoting the true parameter. Here, we are concerned with a problem with a large p that imposes computational efficiency constraints on finding the solution of the optimization problem. A complementary problem is concerned with very large n which will not be dealt with in this paper but bears some similarities with the problem studied below.

Given a $p \times s$ matrix $\mathbf{S}$ with $s \le p$, whose generation will be discussed later, we can find an approximate solution via

$$\tilde{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \mathbf{S}\boldsymbol{\alpha}) + n\lambda \|\mathbf{S}\boldsymbol{\alpha}\|^2$$

and set $\tilde{\boldsymbol{\beta}} = \mathbf{S}\tilde{\boldsymbol{\alpha}}$. When $s = p$ and $\mathbf{S}$ is invertible, we obviously have $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$, but the main interest is in the case where s is much smaller than p so that the optimization problem for $\boldsymbol{\alpha}$ is a quantile ridge regression problem with dimension s which can be solved faster. Our goal is to establish the statistical properties of $\tilde{\boldsymbol{\beta}}$ as an estimator of $\boldsymbol{\beta}_0$.

It is clear that if the $n \times p$ covariate matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ has rank r, then one can find $\mathbf{S}$ and $\boldsymbol{\alpha}$ such that $\mathbf{x}_i^\top \mathbf{S}\boldsymbol{\alpha} = \mathbf{x}_i^\top \boldsymbol{\beta}_0$ for all $i \in \{1, \ldots, n\}$, as long as $s \ge r$. Thus, the projection approach is expected to work well if $\boldsymbol{\Sigma} := E[\mathbf{x}\mathbf{x}^\top]$ is approximately low rank. Our established bound does not require any low-rank assumption, but the bound tends to be small when $\boldsymbol{\Sigma}$ is approximately low rank.

Let η, $\kappa \in (0, 1)$ be fixed numbers, and let $\mathbf{B}$ be any fixed matrix of a certain size such that the matrix products below are well defined. Let the singular value decomposition (SVD) of $\boldsymbol{\Sigma}$ be $\mathbf{U}\mathbf{D}\mathbf{U}^\top$ with diagonal entries of $\mathbf{D}$ arranged decreasingly and $\mathbf{U}_1$ the first r columns of $\mathbf{U}$ with $r \le s$. Our risk bound below depends on the choice of r. Ideally, we should choose r such that the risk bound is the smallest, but this optimal choice would depend on unknown quantities in a complicated way, and our bound is not claimed to be tight. Thus, we will not discuss the optimal choice of r to use in the proof. Note r is merely used as a parameter in our proof and is not a parameter to choose when computing the estimator. Let $\mathbf{S}$ be a random matrix of size $n \times s$. We assume $\mathbf{S}$ satisfies the following two properties with a probability that depends on s:

(i) $\|\mathbf{U}_1^\top \mathbf{S}\mathbf{S}^\top \mathbf{U}_1 - \mathbf{I}\|_{op} \le \eta$.
(ii) $\|\mathbf{U}_1^\top \mathbf{S}\mathbf{S}^\top \mathbf{B} - \mathbf{U}_1^\top \mathbf{B}\|_F^2 \le \kappa \|\mathbf{B}\|_F^2$.

In the above, $\|.\|_{op}$ and $\|.\|_F$ denote the operator norm and the Frobenius norm for matrices, respectively. In Wang et al. (2018), the first property is called Subspace Embedding Property (SEP), and the second is called Matrix Multiplication Property (MMP). It has been verified that, for various types of random matrices, SEP and MMP are satisfied. For example, tab. 5 in that paper shows that for a Gaussian random matrix (among many other possibilities shown there), if $s \ge C(r + \log(1/\delta_1))/\eta^2$, SEP holds with probability at least $1 - \delta_1$, and if $s \ge r/(\kappa\delta_2)$, MMP holds with probability at least $1 - \delta_2$.

We will establish an upper bound for the risk $E_{\mathbf{x}}\|\mathbf{x}^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\|^2 = \|\boldsymbol{\Sigma}^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\|^2$, where the subscript $\mathbf{x}$ in the expectation indicates that we are taking expectation over $\mathbf{x}$ which is independent of the data. To connect the check loss with the risk, we make the following assumptions.

Assumption (A). There exists a constant $C > 0$ such that $E[\rho_\tau(y - \mathbf{x}^\top \boldsymbol{\beta})] - E[\rho_\tau(y - \mathbf{x}^\top \boldsymbol{\beta}_0)] \ge CE(\mathbf{x}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0))^2$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$.
Assumption (B). The conditional density of $\epsilon := y - \mathbf{x}^\top \boldsymbol{\beta}_0$ given $\mathbf{x}$ (equivalently, the conditional density of y) is bounded by a constant uniformly over the support of $\mathbf{x}$.

Assumption (A) is basically the same as assumption (A) in Li et al. (2007), with the latter being slightly more general. It is also a special case of assumption 3 in Lv et al. (2018) with $q = 2$ in their assumption, and some sufficient condition for this assumption was also given there. Using the first four lines in (4) in the proof below, we can easily see that a sufficient condition for (A) is that the conditional density of y is bounded from below. Assumption (A) allows one to connect the loss to the estimation error and is thus required and prevalent in the literature of quantile regression (Belloni & Chernozhukov, 2011; Zou & Yuan, 2008).

Assumption (B) is mild and often used in quantile regression (He & Shi, 1994; Belloni & Chernozhukov, 2011). The following theorem is stated in a way that sacrifices clarity for generality. After the proof, a few remarks are presented to make the bound more explicit. In the statement as well as the proof of the theoretical results, C denotes a generic positive constant that can take different values at different places.

Theorem 1. *Assume assumptions (A) and (B) hold and* **S** *satisfies SEP and MMP stated previously. We also assume* $\|\mathbf{x}\| \le C$ *for some constant* C. *For any* $u > C/\sqrt{n}$ *with a fixed constant* C, *with* λ *set appropriately as in the proof, with probability at least* $1 - \exp\{-Cn(A_n(u)/u)^2\}$, *we have*

$$\|\boldsymbol{\Sigma}^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\|^2 \le CA_n(u)\sqrt{1+\kappa(p-r)}\|\boldsymbol{\beta}_0\| + C(A_n(u)/u)^2 + C(\sigma_{r+1} + \kappa s_{r+1})\|\boldsymbol{\beta}_0\|^2,$$

where $A_n(u) = \left(\sum_{j=1}^{p} \sigma_j u^2/(n(\sigma_j + u^2))\right)^{1/2}$, $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_p \ge 0$ *are the eigenvalues of* $\boldsymbol{\Sigma}$, *and* $s_{r+1} = \sum_{j=r+1}^{p} \sigma_j$.

In preparation for the proof of the theorem, we first state and prove several useful lemmas.

Lemma 1. For any $u > 0$,

$$E\left[\sup_{\beta:\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}\| \le u, \|\boldsymbol{\beta}\| \le 1} \frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{x}_i^\top \boldsymbol{\beta}\right] \le CA_n(u),$$

where $w_i, i = 1, \ldots, n$ are i.i.d. Rademacher variables.

Proof of Lemma 1. Assume the SVD $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ with $\mathbf{D}$ the diagonal matrix with entries $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_p \ge 0$. Using that $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} = \boldsymbol{\beta}^\top \mathbf{U}\mathbf{D}\mathbf{U}^\top \boldsymbol{\beta} \le u^2$ and $\boldsymbol{\beta}^\top \mathbf{U}\mathbf{U}^\top \boldsymbol{\beta} = \boldsymbol{\beta}^\top \boldsymbol{\beta} \le 1$, we have $\boldsymbol{\beta}^\top \mathbf{U}(\mathbf{D}/u^2 + \mathbf{I})\mathbf{U}^\top \boldsymbol{\beta} \le 2$. Thus, we can bound

$$\begin{aligned}
&E\left[\sup_{\beta:\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}\| \le u, \|\boldsymbol{\beta}\| \le 1} \frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{x}_i^\top \boldsymbol{\beta}\right] \\
&= E\left[\sup_{\beta:\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}\| \le u, \|\boldsymbol{\beta}\| \le 1} \frac{1}{n}\mathbf{w}^\top \mathbf{X}\boldsymbol{\beta}\right] \\
&\le E\left[\sup_{\beta:\boldsymbol{\beta}^\top \mathbf{U}(\mathbf{D}/u^2+\mathbf{I})\mathbf{U}^\top \boldsymbol{\beta} \le 2} \frac{1}{n}\mathbf{w}^\top \mathbf{X}\boldsymbol{\beta}\right],
\end{aligned}$$

where $\mathbf{w} = (w_1, \ldots, w_n)^\top$. Furthermore,

$$\begin{aligned}
&\left(E\left[\sup_{\beta:\boldsymbol{\beta}^\top \mathbf{U}(\mathbf{D}/u^2+\mathbf{I})\mathbf{U}^\top \boldsymbol{\beta} \le 2} \tfrac{1}{n}\mathbf{w}^\top \mathbf{X}\boldsymbol{\beta}\right]\right)^2 \\
&\le E\left[\left(\sup_{\beta:\boldsymbol{\beta}^\top \mathbf{U}(\mathbf{D}/u^2+\mathbf{I})\mathbf{U}^\top \boldsymbol{\beta} \le 2} \tfrac{1}{n}\mathbf{w}^\top \mathbf{X}\boldsymbol{\beta}\right)^2\right] \\
&= E\left[\left(\sup_{\beta:\boldsymbol{\beta}^\top \mathbf{U}(\mathbf{D}/u^2+\mathbf{I})\mathbf{U}^\top \boldsymbol{\beta} \le 2} \tfrac{1}{n}\mathbf{w}^\top \mathbf{X}\mathbf{U}(\mathbf{D}/u^2+\mathbf{I})^{-1/2}(\mathbf{D}/u^2+\mathbf{I})^{1/2}\mathbf{U}^\top \boldsymbol{\beta}\right)^2\right] \\
&\le E\left[\sup_{\beta:\boldsymbol{\beta}^\top \mathbf{U}(\mathbf{D}/u^2+\mathbf{I})\mathbf{U}^\top \boldsymbol{\beta} \le 2} \frac{1}{n^2}\|\mathbf{w}^\top \mathbf{X}\mathbf{U}(\mathbf{D}/u^2+\mathbf{I})^{-1/2}\|^2 \|(\mathbf{D}/u^2+\mathbf{I})^{1/2}\mathbf{U}^\top \boldsymbol{\beta}\|^2\right] \\
&\le \frac{C}{n^2}E\left[\mathbf{w}^\top \mathbf{X}\mathbf{U}(\mathbf{D}/u^2+\mathbf{I})^{-1}\mathbf{U}^\top \mathbf{X}^\top \mathbf{w}\right] \\
&\le \frac{C}{n^2}E\left[\mathrm{tr}\left(\mathbf{X}\mathbf{U}(\mathbf{D}/u^2+\mathbf{I})^{-1}\mathbf{U}^\top \mathbf{X}^\top\right)\right] \\
&\le \frac{C}{n}E\left[\mathrm{tr}\left((\mathbf{D}/u^2+\mathbf{I})^{-1}\mathbf{D}\right)\right] \\
&= \frac{C}{n}\sum_{j=1}^{p}\frac{\sigma_j u^2}{\sigma_j + u^2}.
\end{aligned}$$

□

Lemma 2. If $u > C/\sqrt{n}$, with probability at least $1-e^{-Cn(A_n(u)/u)^2}$,

$$\begin{aligned}&\Big|(1/n)\sum_i(\rho_\tau(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})-\rho_\tau(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_0))\\&\quad-E[\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta})-\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta}_0)]|\\&\quad\le C(A_n(u)/u)\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+CA_n(u)\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|,\forall\boldsymbol{\beta}\in\mathbb{R}^p.\end{aligned}$$

Proof of Lemma 2. By the standard symmetrization argument Pollard (1984), we have

$$\begin{aligned}&E\left[\sup_{\boldsymbol{\beta}}\left|(P-P_n)\frac{\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta})-\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\right]\\&\le CE\left[\sup_{\boldsymbol{\beta}}\left|\frac{(1/n)\sum_i w_i\mathbf{x}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\right],\end{aligned}\tag{1}$$

where $w_i, i=1,\ldots,n$ are i.i.d. Rademacher variables and the inequality follows from the contraction inequality for the Rademacher complexity (see, e.g., theorem 2.2 of Koltchinskii, 2011) since ρ_τ is Lipschitz continuous.

For the right-hand side of (1), we consider the set $\{\boldsymbol{\gamma}:\boldsymbol{\gamma}=(\boldsymbol{\beta}-\boldsymbol{\beta}_0)/(u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|),\boldsymbol{\beta}\in\mathbb{R}^p\}$. It is easy to see this class is actually contained in $\{\boldsymbol{\gamma}:\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\gamma}\|\le u,\|\boldsymbol{\gamma}\|\le 1\}$. Thus, by Lemma 1, we get

$$E\left[\sup_{\boldsymbol{\beta}}\left|\frac{(1/n)\sum_i w_i\mathbf{x}_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\right]\le CA_n(u).$$

Next, we use a concentration inequality to remove the expectation above. Since

$$\begin{aligned}&\left|\frac{\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta})-\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\\&\le C\left|\frac{\mathbf{x}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\\&\le C\|\mathbf{x}\|\le C,\end{aligned}$$

and

$$\begin{aligned}&Var\left(\frac{\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta})-\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right)\\&\le CVar\left(\frac{\mathbf{x}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right)\\&\le Cu^2,\end{aligned}$$

using the concentration inequality (see, e.g., the Adamczak bound on pages 24–25 of Koltchinskii, (2011)), we have

$$\begin{aligned}&\sup_{\boldsymbol{\beta}}\left|(P-P_n)\frac{\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta})-\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\\&\le CE\left[\sup_{\boldsymbol{\beta}}\left|(P-P_n)\frac{\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta})-\rho_\tau(y-\mathbf{x}^\top\boldsymbol{\beta}_0)}{u^{-1}\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|}\right|\right]\\&\quad+Cu\sqrt{t/n}+C\sqrt{t}/n,\end{aligned}$$

with probability at least $1-e^{-t}$. By setting $t=Cn(A_n(u)/u)^2$, we complete the proof of the lemma since $u\sqrt{t/n}\le CA_n(u)$ and $\sqrt{t}/n\le CA_n(u)$ using that $u\ge C/\sqrt{n}$. □

Write $\mathbf{U}=(\mathbf{U}_1,\mathbf{U}_2)$ where $\mathbf{U}_1$ is $p\times r$ and $\mathbf{U}_2$ is $p\times(p-r)$, and $\mathbf{D}=\begin{pmatrix}\mathbf{D}_1 & \\ & \mathbf{D}_2\end{pmatrix}$, where $\mathbf{D}_1$ contains the first r eigenvalues and $\mathbf{D}_2$ contains the rest. Define

$$\breve{\boldsymbol{\alpha}}=\mathbf{S}^\top\mathbf{U}_1(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\mathbf{U}_1^\top\boldsymbol{\beta}_0.$$

We have the following lemma.

Lemma 3. $E_{\mathbf{x}}\|\mathbf{x}^\top\boldsymbol{\beta}_0-\mathbf{x}^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2=O_p((\sigma_{r+1}+\kappa s_{r+1})\|\boldsymbol{\beta}_0\|^2)$ and $\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2=O_p((1+\kappa(p-r))\|\boldsymbol{\beta}_0\|^2)$.

Proof of Lemma 3. We have

$$\begin{aligned}
&\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}_0-\boldsymbol{\Sigma}^{1/2}\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2\\
=&\|\mathbf{U}\mathbf{D}^{1/2}\mathbf{U}^\top\boldsymbol{\beta}_0-\mathbf{U}\mathbf{D}^{1/2}\mathbf{U}^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2\\
=&\|\mathbf{D}^{1/2}\mathbf{U}^\top\boldsymbol{\beta}_0-\mathbf{D}^{1/2}\mathbf{U}^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2\\
=&\|\mathbf{D}_1^{1/2}\mathbf{U}_1^\top\boldsymbol{\beta}_0-\mathbf{D}_1^{1/2}\mathbf{U}_1^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2+\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\boldsymbol{\beta}_0-\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2.
\end{aligned}\tag{2}$$

For the first term in (2), we have

$$\|\mathbf{D}_1^{1/2}\mathbf{U}_1^\top\boldsymbol{\beta}_0-\mathbf{D}_1^{1/2}\mathbf{U}_1^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|=\|\mathbf{D}_1^{1/2}\mathbf{U}_1^\top\boldsymbol{\beta}_0-\mathbf{D}_1^{1/2}\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\mathbf{U}_1^\top\boldsymbol{\beta}_0\|=0.$$

For the second term in (2), we have

$$\begin{aligned}
&\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\boldsymbol{\beta}_0-\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|\\
\le&\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\boldsymbol{\beta}_0\|+\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\mathbf{S}\breve{\boldsymbol{\alpha}}\|\\
\le&C\sqrt{\sigma_{r+1}}\|\boldsymbol{\beta}_0\|+\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\mathbf{U}_1^\top\boldsymbol{\beta}_0\|\\
\le&C\sqrt{\sigma_{r+1}}\|\boldsymbol{\beta}_0\|+\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1\|_F\cdot\|(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\|_{op}\cdot\|\mathbf{U}_1^\top\boldsymbol{\beta}_0\|\\
\le&C\sqrt{\sigma_{r+1}}\|\boldsymbol{\beta}_0\|+C\sqrt{\kappa s_{r+1}}\|\boldsymbol{\beta}_0\|,
\end{aligned}$$

where the last step used the properties SEP and MMP for $\mathbf{S}$ and that $\|\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\|_F^2=\mathrm{tr}(\mathbf{D}_2^{1/2}\mathbf{U}_2^\top\mathbf{U}_2\mathbf{D}_2^{1/2})=\mathrm{tr}(\mathbf{D}_2)=s_{r+1}$. For the second bound in the lemma, we have

$$\begin{aligned}
&\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2\\
\le&\|\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\mathbf{U}_1^\top\boldsymbol{\beta}_0\|^2+\|\mathbf{U}_2^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\mathbf{U}_1^\top\boldsymbol{\beta}_0\|^2\\
\le&\|\boldsymbol{\beta}_0\|^2+C\kappa(p-r)\|\boldsymbol{\beta}_0\|^2.
\end{aligned}\tag{3}$$

□

Proof of Theorem 1. By the definition of $\tilde{\boldsymbol{\alpha}}$, we have

$$\frac{1}{n}\sum_i\rho_\tau(y_i-\mathbf{x}_i^\top\mathbf{S}\tilde{\boldsymbol{\alpha}})+\lambda\|\mathbf{S}\tilde{\boldsymbol{\alpha}}\|^2\le\frac{1}{n}\sum_i\rho_\tau(y_i-\mathbf{x}_i^\top\mathbf{S}\breve{\boldsymbol{\alpha}})+\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2.$$

Using Lemma 2, with probability at least $1-\exp\{-Cn(A_n(u)/u)^2\}$,

$$\begin{aligned}
&E[\rho_\tau(y-\mathbf{x}^\top\mathbf{S}\tilde{\boldsymbol{\alpha}})]-E[\rho_\tau(y-\mathbf{x}^\top\mathbf{S}\breve{\boldsymbol{\alpha}})]\\
&\le\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2-\lambda\|\mathbf{S}\tilde{\boldsymbol{\alpha}}\|^2+Cu^{-1}A_n(u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|+CA_n(u)\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|\\
&=-2\lambda\langle\mathbf{S}\breve{\boldsymbol{\alpha}},\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\rangle-\lambda\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|^2\\
&\quad+Cu^{-1}A_n(u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|+CA_n(u)\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|\\
&\le2\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|\cdot\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|-\lambda\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|^2\\
&\quad+Cu^{-1}A_n(u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|+CA_n(u)\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|\\
&\le2\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2+\frac{\lambda}{2}\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|^2-\lambda\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|^2\\
&\quad+Cu^{-1}A_n(u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|+CA_n(u)\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|\\
&\le2\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2-\frac{\lambda}{2}\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|^2+Cu^{-1}A_n(u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|\\
&\quad+C\left(\frac{CA_n^2(u)}{2\lambda}+\frac{\lambda}{2C}\|\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|^2\right)\\
&=2\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2+\frac{CA_n(u)^2}{\lambda}+Cu^{-1}A_n(u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|.
\end{aligned}$$

With the choice $\lambda \asymp A_n(u)/\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|$, we now have

$$E[\rho_\tau(y-\mathbf{x}^\top \mathbf{S}\tilde{\boldsymbol{\alpha}})]-E[\rho_\tau(y-\mathbf{x}^\top \mathbf{S}\breve{\boldsymbol{\alpha}})] \leq CA_n(u)\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|+C(A_n(u)/u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|.$$

Now, we bound $|E[\rho_\tau(y-\mathbf{x}^\top \mathbf{S}\breve{\boldsymbol{\alpha}})]-E[\rho_\tau(y-\mathbf{x}^\top \boldsymbol{\beta}_0)]|$. Using Knight's identity that $\rho_\tau(x-y)-\rho_\tau(x)=-y(\tau-I\{x\leq 0\})+\int_0^y(I\{x\leq t\}-I\{x\leq 0\})dt$, we have

$$\begin{aligned}
&|E[\rho_\tau(y-\mathbf{x}^\top \mathbf{S}\breve{\boldsymbol{\alpha}})]-E[\rho_\tau(y-\mathbf{x}^\top \boldsymbol{\beta}_0)]|\\
&=\left|-E[(\mathbf{x}^\top(\mathbf{S}\boldsymbol{\alpha}-\boldsymbol{\beta}_0)))(\tau-I\{y\leq \mathbf{x}^\top\boldsymbol{\beta}_0\})]+E\int_0^{\mathbf{x}^\top(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)}(F(t|\mathbf{x})-F(0|\mathbf{x}))dt\right|\\
&=\left|E\int_0^{\mathbf{x}^\top(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)}(F(t|\mathbf{x})-F(0|\mathbf{x}))dt\right|\\
&=\left|E\int_0^{\mathbf{x}^\top(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)}f(t^*|\mathbf{x})tdt\right|\\
&\leq C\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2,
\end{aligned} \tag{4}$$

where $F(.|x)$ is the conditional distribution of ϵ given x, $f(.|x)$ is the corresponding conditional density, t^* is a value between 0 and t that appears in the Taylor's expansion and the inequality uses Assumption (B). Using (4) and Assumption (A), we get that

$$\begin{aligned}
&\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\tilde{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2\\
&\quad\leq CA_n(u)\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|+C(A_n(u)/u)\|\boldsymbol{\Sigma}^{1/2}\mathbf{S}(\tilde{\boldsymbol{\alpha}}-\breve{\boldsymbol{\alpha}})\|+C\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2\\
&\quad\leq CA_n(u)\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|+(1/2)\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\tilde{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2+C\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2+C(A_n(u)/u)^2,
\end{aligned}$$

which implies

$$\begin{aligned}
&\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\tilde{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2\\
&\quad\leq CA_n(u)\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|+C(A_n(u)/u)^2+C\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2\\
&\quad\leq CA_n(u)\sqrt{1+\kappa(p-r)}\|\boldsymbol{\beta}_0\|+C(A_n(u)/u)^2+C(\sigma_{r+1}+\kappa s_{r+1})\|\boldsymbol{\beta}_0\|^2.
\end{aligned}$$

□

Remark 1. If $\sigma_r \geq C/n$, by choosing $u^2=\sigma_r$, we have $A_n^2(u)\asymp(r\sigma_r+s_{r+1})/n$, and thus,

$$\begin{aligned}
&\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\breve{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2\\
&\quad\leq C\sqrt{\frac{r\sigma_r+s_{r+1}}{n}(1+\kappa(p-r))}\|\boldsymbol{\beta}_0\|+C\frac{r+s_{r+1}/\sigma_r}{n}+C(\sigma_{r+1}+\kappa s_{r+1})\|\boldsymbol{\beta}_0\|^2.
\end{aligned}$$

If $\boldsymbol{\Sigma}$ has rank r, the bound can be further simplified due to $\sigma_{r+1}=s_{r+1}=0$.

Remark 2. If in addition to SEP and MMP, we impose the "bounded spectral norm property" that $\|\mathbf{S}\|_{op}^2 \leq Cp/s$ (Wang et al. 2018); we have the following bound in place of (3):

$$\begin{aligned}
&\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2\\
&\quad\leq \|\mathbf{S}\|_{op}^2\|\mathbf{S}^\top\mathbf{U}_1(\mathbf{U}_1^\top\mathbf{S}\mathbf{S}^\top\mathbf{U}_1)^{-1}\mathbf{U}_1^\top\boldsymbol{\beta}_0\|^2\\
&\quad\leq C\|\mathbf{S}\|_{op}^2\|\mathbf{U}_1^\top\boldsymbol{\beta}_0\|^2\\
&\quad\leq C(p/s)\|\boldsymbol{\beta}_0\|^2.
\end{aligned}$$

The factor $(1+\kappa(p-r))$ in the bound of Theorem 1 can thus be replaced by $C(p/s)$ giving a slightly different risk bound.

Remark 3. In the proof, $\lambda=A_u(u)/\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|$ is set to balance $2\lambda\|\mathbf{S}\breve{\boldsymbol{\alpha}}\|^2+CA_n(u)^2/\lambda$. This choice of λ merely serves theoretical purposes and is not feasible in practice since $\breve{\boldsymbol{\alpha}}$ is unknown. This is a reason we choose not to mention the choice of λ in the statement of the theorem. Furthermore, it is clear from the proof that we can derive a bound for any value of λ given by

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S}\tilde{\boldsymbol{\alpha}}-\boldsymbol{\beta}_0)\|^2$$
$$\leq C\lambda(1+\kappa(p-r))\|\boldsymbol{\beta}_0\|^2+CA_n(u)^2/\lambda+C(A_n(u)/u)^2+C(\sigma_{r+1}+\kappa s_{r+1})\|\boldsymbol{\beta}_0\|^2.$$

3 | NUMERICAL STUDIES

We perform some simulations to examine the performances of random projection in quantile ridge regression. The covariates are generated from $\mathbf{x}_i \sim N(0,\boldsymbol{\Sigma})$ with (i,j)-entry of $\boldsymbol{\Sigma}$ given by $\rho^{|i-j|}$ with $\rho\in\{0.1,0.4,0.7\}$. The components of $\boldsymbol{\beta}_0$ are generated from $N(0,0.2^2)$. The responses are generated independently from

$$y_i=\mathbf{x}_i^\top\boldsymbol{\beta}_0+0.3(1+\Phi(x_{i1}))(\epsilon_i-\Phi^{-1}(\tau)),$$

where $\epsilon_i \overset{i.i.d.}{\sim} N(0,1)$ and Φ is the standard normal cdf and Φ^{-1} is the corresponding quantile function so that the τth conditional quantile function is just $\mathbf{x}_i^\top\boldsymbol{\beta}_0$. We use $\tau=0.7$ for illustration. We set $n=p=2048$. In each case, 200 data sets are generated. We use a range of tuning parameters $\log(\lambda)\in\{-10,-9,\ldots,3,4\}$. For the projection matrix $\mathbf{S}$, we consider both Gaussian projection (entries of the matrix are i.i.d. Gaussian) and

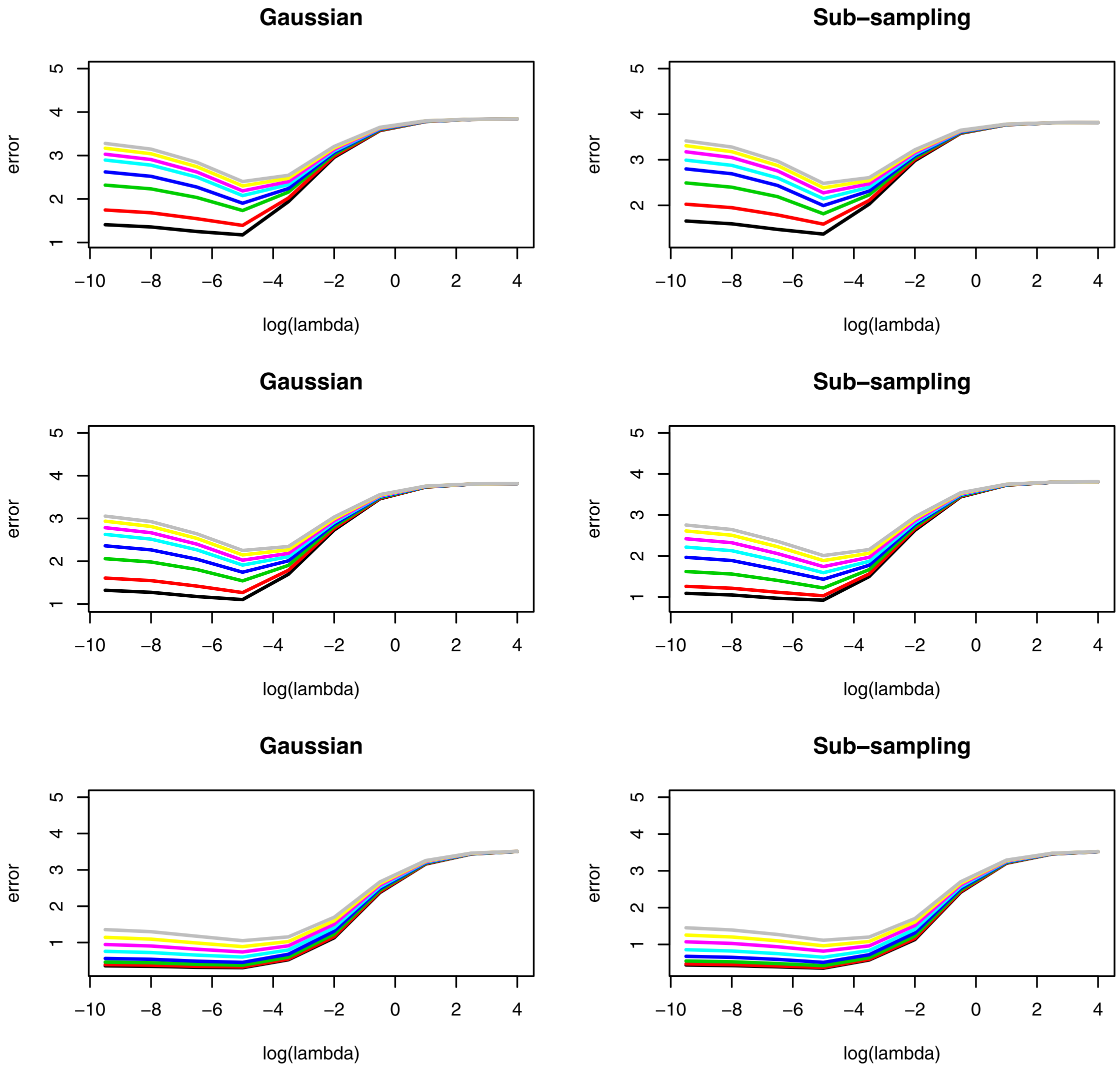

FIGURE 1 Errors in estimating $\boldsymbol{\beta}_0$ as s vary in $\{n/2,n/2^2,\ldots,n/2^7\}$. The black curve shows the results without using random projection, and other coloured curves show the errors for different values of s (with larger s corresponding to smaller errors). The three rows correspond to $\rho=0.1,0.4,0.7$, respectively

subsampling (here, $\mathbf{S}$ is constructed by randomly drawing s columns of the identity matrix). We compute the estimator with random projection for $s \in \{n/2, n/2^2, \ldots, n/2^7\}$. The performances are assessed by $\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\|$.

The model is fitted using the publicly available R package **hqreg** based on the semi-smooth coordinate descent method of Yi and Huang (2017), which used warm starts to compute the solution path efficiently. The results are presented in Figure 1. In this figure, the three rows correspond to $\rho = 0.1, 0.4, 0.7$ from top to bottom, whereas the two columns are for Gaussian projection and sub-sampling, respectively. Each curve in the figure shows how the error changes with λ, whereas different curves are for different s (the black curve at the bottom of each figure shows the error of the standard estimator without random projection). Naturally, the curve with larger error corresponds to smaller value of s. We see that as ρ increases, the performances typically become better.

4 | CONCLUSION

In this short manuscript, we considered quantile linear regression with a ridge penalty, which is suitable for high-dimensional models. We established some statistical property of the estimator when random projection is used.

As mentioned in the text, a closely related problem is to deal with the case when n is large, and a sketching method can be used in the case. Theoretical study of random sketching seems to bear some similarities with random projection and is worthy of investigation in the future.

As shown in the proof, the theoretical choice of λ depends on unknown quantities. In practice, one can of course use cross-validation to choose λ. The purpose of the current work is to show (by simulation) that random projection can be used to approximate the original problem for any choice of λ. Study of data-driven optimal choice of λ with theoretical guarantees seems hard, if not impossible.

ACKNOWLEDGEMENTS

The authors sincerely thank the editor, the associate editor and two anonymous reviewers for their insightful comments that improved the manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Heng Lian https://orcid.org/0000-0002-6008-6614

REFERENCES

Anand, A., Wilkinson, L., & Dang, T. N. (2012). Visual pattern discovery using random projections. In *IEEE Conference on Visual Analytics Science and Technology 2012, Vast 2012 - Proceedings* (pp. 43–52), Seattle, WA, USA.

Belloni, A., & Chernozhukov, V. (2011). l1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, *39*(1), 82–130.

Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: Applications to image and text data. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 245–250), San Francisco, USA.

Fan, J. Q., & Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

He, X., & Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, *3*, 299–308.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55–67.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, *1*(46), 33–50.

Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. New York: Springer.

Li, Y., Liu, Y., & Zhu, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, *102*, 255–268.

Liu, J., He, J., Zhang, W., Ma, T., Tang, Z., Niyoyita, J. P., & Gui, W. (2019). ANID-SEoKELM: Adaptive network intrusion detection based on selective ensemble of kernel ELMs with random features. *Knowledge-Based Systems*, *177*, 104–166.

Lv, S., Lin, H., Lian, H., & Huang, J. (2018). Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space. *Annals of Statistics*, *46*, 781–813.

Maillard, O. (2012). Linear regression with random projections. *Journal of Machine Learning Research*, *13*, 2735–2772.

Mu, Y., Dong, J., Yuan, X., & Yan, S. (2011). Accelerated low-rank visual recovery by random projection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2609–2616), Colorado Springs, CO, USA.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological*, *58*(1), 267–288.

Wang, S., Gittens, A., & Mahoney, M. W. (2018). Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging. *Journal of Machine Learning Research*, *18*, 1–50.

Wang, Z., Jie, W., Chen, S., & Gao, D. (2013). Random projection ensemble learning with multiple empirical kernels. *Knowledge-Based Systems*, *37*, 388–393.

Yi, C., & Huang, J. (2017). Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, *26*, 547–557.

Zhang, F., Li, R., & Lian, H. (2021). Approximate nonparametric quantile regression in reproducing kernel Hilbert spaces via random projection. *Information Sciences*, *547*, 244–254.

Zhang, L., Mahdavi, M., Jin, R., Yang, T., & Zhu, S. (2014). Random projections for classification: A recovery approach. *IEEE Transactions on Information Theory*, *60*(11), 7300–7316.

Zou, H., & Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, *36*(3), 1108–1126.

How to cite this article: Zhou, Y., Liang, J., Hu, Y., & Lian, H. (2021). Random projections for quantile ridge regression. *Stat*, *10*(1), e386. https://doi.org/10.1002/sta4.386