Statistica Sinica Preprint No: SS-2019-0247	
Title	Regularized projection score estimation of treatment effects in high-dimensional quantile regression
Manuscript ID	SS-2019-0247
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0247
Complete List of Authors	Chao Cheng, Xingdong Feng, Jian Huang and Xu Liu
Corresponding Author	Xu Liu
E-mail	liu.xu@mail.shufe.edu.cn
Notice: Accepted version subject to English editing.	

Regularized projection score estimation of treatment effects in high-dimensional quantile regression

Chao Cheng[a], Xingdong Feng[a], Jian Huang[b], Xu Liu[a] *

[a]School of Statistics and Management,

Shanghai University of Finance and Economics, Shanghai 200433, China

[b]Department of Statistics and Actuarial Science,

University of Iowa, Iowa City, Iowa 52246, USA

Abstract

We propose a regularized projection score method for estimating treatment effects in quantile regression in the presence of high-dimensional confounding covariates. We show that the proposed estimator of the treatment effects is consistent and asymptotically normal, with a root-n rate of convergence. We also provide an efficient algorithm for the proposed estimator. This algorithm can be easily implemented

*Corresponding author. Tel.: +86 21 6590 1156. E-mail address: *liu.xu@sufe.edu.cn*.

using existing software. Furthermore, we propose and validate a refitted wild bootstrapping approach for variance estimation. This enables us to construct confidence intervals for treatment effects in high-dimensional settings. Simulation studies are carried out to evaluate the finite sample performance of the proposed estimator. A GDP growth rate dataset is used to demonstrate the applications of the method.

Key Words: Efficiency score; High dimension; Quantile regression; Wild bootstrap.

1 Introduction

Quantile regression (Koenker and Bassett 1978) is an important tool for analyzing the relationship between a response variable and a set of covariates. It has a wide range of applications in the analysis of non-Gaussian data, which arises frequently in applied economic research. Unlike least squares regression, which models the conditional mean of a response given the covariates, quantile regression focuses on the conditional quantiles. Thus, it is able to provide a description of the conditional distribution of the response given the covariates. There is an extensive literature on the theoretical properties and computational algorithms for quantile regression when the number of regressors is fixed or increases at a lower rate than the sample size; see, for example, Koenker (2005) and the references therein. In this paper, we estimate low-dimensional treatment effects in the presence of a high-dimensional nuisance parameter vector.

There is now a substantial body of work on penalized methods for variable selection in high-dimensional models. Several important penalty functions have been introduced, including least absolute shrinkage and the selection operator (Lasso) or the ℓ_1 penalty

(Tibshirani 1996), the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li 2001), and the minimax concave penalty (MCP) (Zhang 2010). A common feature of these penalties is that they are capable of producing exact zero solutions, which automatically leads to variable selection. The penalized methods also have many attractive theoretical properties concerning selection, estimation, and prediction in the sparse setting ($p \gg n$), including the asymptotic oracle property under certain conditions. However, these methods provide no computable error assessment of the selection results in finite sample situations. The literature on this topic has grown too vast to be adequately summarized here, so we refer to the book by Bühlmann and van de Geer (2011), and the references therein for the results on convex selection and the papers by Fan and Li (2001); Zhang (2010) and Zhang and Zhang (2012) for concave selection.

Recently, many authors have studied the problem of statistical inference for low-dimensional parameters in high-dimensional regression models. Zhang and Zhang (2014) proposed a semiparametric efficient score approach for constructing confidence intervals of low-dimensional coefficients in high-dimensional linear models. van de Geer et al. (2014) considered the same problem using an approach that inverts the optimization conditions for the Lasso solution, extending the work of Zhang and Zhang (2014) to generalized linear models and problems with convex loss functions. Javanmard and Montanari (2014) considered the problem of hypothesis testing in high-dimensional regression using a method similar to that of Zhang and Zhang (2014). Fang et al. (2016) studied hypothesis testing and confidence intervals in high-dimensional proportional hazards models. Neykov et al. (2018) proposed a unified theory of confidence regions and testing for high-dimensional estimating equations. Ning and Liu (2017) proposed a decorrelated score approach for hypothesis tests and confidence regions in sparse high-dimensional models. Zhu and Bradic (2018) proposed an approach to test linear hypotheses without assumptions on model sparsity or the loading vector representing the hypothesis in high-dimensional linear models. For more related works using the regularized score method, refer to Belloni et al. (2013); Dezeure et al. (2015); Lockhart et al. (2014); Meinshausen (2014); Meinshausen et al. (2009); Ning and Liu (2017); Stucky and van de Geer (2018); Yang (2017).

Belloni et al. (2012) proposed a two-stage selection procedure with post-double selection to estimate a single treatment effect parameter in a high-dimensional linear model. Tibshirani et al. (2016) considered the statistical inference for forward stepwise and least angle regression in high-dimensional models after selection. Recently, various researchers have considered post-selection in the presence of high-dimensional parameters, including Berk et al. (2013, 2009); Lee et al. (2016); Lee and Taylor (2014); Rgamer and Greven (2018); Tibshirani et al. (2016).

Belloni and Chernozhukov (2011) studied the ℓ_1-penalized quantile regression under the high-dimensional setting and established a near-oracle property of the estimator. Wang et al. (2012) showed that the oracle property still holds when SCAD and MCP penalties are used. Zhao et al. (2014) provided a globally penalized framework for high-dimensional quantile regression models by employing adaptive ℓ_1 penalties; this approach could achieve consistent shrinkage of regression quantile estimates across a continuous range of quantile levels. Belloni et al. (2018) considered the robust inference of regression coefficients of high-dimensional quantile regression models via an optimal instrument, which was a residual from a density-weighted projection of the regressor of interest on other regressors. Zheng et al. (2015) proposed a robust and uniformly honest inference in high-dimensional quantile regression using a debiased composite quantile estimator.

Inspired by the work of Zhang and Zhang (2014) and Ning and Liu (2017), we consider the estimation of a pre-conceived low-dimensional parameter based on a projected score approach and study its statistical inference under linear quantile regression models. In particular, our proposed approach is similar to the decorrelated score method of Ning and Liu (2017). In essence, these approaches extend the efficient score method for dealing with infinite-dimensional nuisance parameters in semiparametric models (Bickel et al. 1998) to the high-dimensional settings. However, the decorrelated score method assumes a smooth loss function with second derivatives, which is not satisfied in the context of quantile regression.

The rest of the paper is organized as follows. Section 2 describes the estimation method based on regularized projection scores. The asymptotic properties of estimates of pre-conceived parameters are obtained in Section 3. We then propose a resampling approach based on cross-validation and confirm its validity in Section 4. An efficient computation algorithm is given in Section 5. On the basis of this algorithm, an one-step estimator is proposed in Section 6. Numerical studies are used to assess the finite-sample performance of the proposed method in Section 7. All proofs are given in the Appendix. An R package implementing the proposed method is available at `https://github.com/xliusufe/pqr`.

2 Regularized Projection Score Estimation

Suppose we have observations $\{(y_i, x_i, z_i), i = 1, \dots, n\}$ that are independent and identically distributed as (y, x, z), where $y \in \mathbb{R}$ is a response variable, $x \in \mathbb{R}^d$ is a d-dimensional vector containing covariates of main interest, and $z \in \mathbb{R}^q$ is a q-dimensional covariate with possibly confounding variables. Consider the linear quantile regression model

$$Q_\tau(y_i|x_i, z_i) = x_i'\beta_0 + z_i'\eta_0, \tag{1}$$

where $Q_\tau(\cdot|x_i, z_i)$ refers to the conditional τth quantile given the covariate (x_i, z_i). Here for notional simplicity, we assume that an intercept term is included in β_0. We would like to estimate the effect of the covariate vector x, represented by β_0, on the response variable, while taking into account the effect of the covariate z, represented by η_0. We are interested in the case where d is small (fixed), but q is large and may be far larger than the sample size n.

In the standard linear quantile regression, the parameters of model (1) are estimated by minimizing

$$M_n(\beta, \eta) = n^{-1} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta - z_i'\eta)$$

with respect to β and η, where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$. This approach works well in low-dimensional cases where both d and q are fixed and smaller than n. However, in the case where $q \gg n$, it no longer works owing to the singularity of the design matrix. There has been much work on penalized methods for estimating the parameter vector (β_0, η_0).

An important method is the Lasso estimator (Tibshirani 1996)

$$(\hat{\beta}_{lasso}, \hat{\eta}_{lasso}) = \operatorname*{argmin}_{\beta,\eta} M_n(\beta, \eta) + \lambda(\|\beta\|_1 + \|\eta\|_1).$$

This provides a point estimate of (β_0, η_0), denoted by $(\hat{\beta}, \hat{\eta})$. Owing to the shrinkage effect of the ℓ_1 penalty, $\hat{\beta}_{lasso}$ does not converge at the usual root-n rate, and its asymptotic distributional property is unknown. The penalized estimate $\hat{\beta}_{lasso}$ cannot be directly used for making statistical inferences about β_0, the main parameter of interest.

To reduce the shrinkage effect of penalization of the estimation of β_0, we consider the semi-penalized estimator

$$(\tilde{\beta}, \tilde{\eta}) = \operatorname*{argmin}_{\beta,\eta} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta - z_i'\eta) + \lambda_1 \|\eta\|_1. \tag{2}$$

Note that here β is not penalized. Intuitively, the estimator $\tilde{\beta}$ should be less biased than $\hat{\beta}_{lasso}$, as it is not subject to penalization. However, because x_i and z_i are correlated, the bias in $\tilde{\eta}$ will still lead to bias in $\tilde{\beta}$. This can be observed more clearly by considering the score equations corresponding to (2):

$$\frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\eta) x_i = 0, \tag{3}$$

$$\frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\eta) z_i = \lambda_1 \partial(\|\eta\|_1), \tag{4}$$

where $\psi_\tau(u) = \tau - I(u < 0)$ is the directional derivative of $\rho_\tau(u)$, and $\partial(\|\eta\|_1) = (\partial(|\eta_1|), \cdots, \partial(|\eta_q|))'$. Here, $\partial(|\eta_j|)$ is the subdifferential of $|\eta_j|$, that is, $\partial(|\eta_j|) = 1$ if $\eta_j > 0$, $\partial(|\eta_j|) = -1$ if $\eta_j < 0$, and $\partial(|\eta_j|) \in [-1, 1]$ if $\eta_j = 0$. The estimator $(\tilde{\beta}', \tilde{\eta}')'$

approximately satisfies (3) and (4). Therefore, $\tilde{\beta}$ is a solution to

$$\frac{1}{n}\sum_{i=1}^{n}\psi_\tau(y_i - x_i'\beta - z_i'\tilde{\eta})x_i = 0.$$

However, owing to the bias in the estimator $\tilde{\eta}$ and the correlation between x_i and z_i, the estimator $\tilde{\beta}$ does not have a root-n rate of convergence.

To obtain an estimator of β_0 with a root-n rate of convergence and an asymptotically normal distribution, we propose a regularized projection score approach. To describe this approach, we first consider the projection score function for β based on the loss function ρ_τ at the population level. The projection score is defined as the residual of the projection of the score function $\psi_\tau(y - x'\beta - z'\eta)x$ for β onto the closure of the linear span of the score function $\psi_\tau(y - x'\beta - z'\eta)z$ for the nuisance parameter η in the Hilbert space $L_2(P)$, where P is the distribution of (y, x, z) under model (1). That is, we need to find a matrix $H_0 \in \mathbb{R}^{d\times q}$ that minimizes

$$\mathrm{E}\|\psi_\tau(y - x'\beta_0 - z'\eta_0)x - \psi_\tau(y - x'\beta_0 - z'\eta_0)Hz\|^2 = \mathrm{E}\{\psi_\tau^2(\varepsilon)\|x - Hz\|^2\} \tag{5}$$

with respect to $H \in \mathbb{R}^{d\times q}$, where $\varepsilon = y - x'\beta_0 - z'\eta_0$. Here $\|\cdot\|$ denotes the Euclidean norm. Then the projection score function for β in the direction H_0 is

$$\psi_\tau(y - x'\beta - z'\eta)x - \psi_\tau(y - x'\beta - z'\eta)H_0 z = \psi_\tau(y - x'\beta - z'\eta)(x - H_0 z). \tag{6}$$

In general, (5) is a weighted least squares function. Under the quantile regression model (1), it can be simplified considerably. By the law of iterated expectations, we have

$$\begin{aligned}\mathrm{E}\{\psi_\tau^2(\varepsilon)\|x - Hz\|^2\} =& \mathrm{E}\{\mathrm{E}[\psi_\tau^2(\varepsilon)|x,z]\|x - Hz\|^2\}\\ =& \tau(1-\tau)\mathrm{E}\|x - Hz\|^2,\end{aligned} \tag{7}$$

where the last equation follows from (1). Thus, minimizing (5) is equivalent to minimizing (7). As τ is independent of H, we have

$$H_0 = \operatorname*{argmin}_{H\in\mathbb{R}^{d\times q}} \mathrm{E}\|x - Hz\|^2.$$

This is a least squares problem and can be solved explicitly. In particular, H_0 satisfies the normal equation $\mathrm{E}\{(x - Hz)z'\} = 0$, which yields

$$H_0 = \mathrm{E}(xz')\{\mathrm{E}(zz')\}^{-1}.$$

However, the sample version of $\mathrm{E}(zz')$, which is given by $n^{-1}\sum_{i=1}^{n} z_i z_i'$, is not invertible if $q > n$. Therefore, we cannot estimate H_0 by simply using the sample versions of $\mathrm{E}(xz')$ and $\mathrm{E}(zz')$. We need to regularize the projection calculation. We can consider either the standard Lasso or the group Lasso for the multi-response linear regression (Obozinski et al. 2011; Wang et al. 2013) estimation of the matrix H_0. For any $H \in \mathbb{R}^{d\times q}$, denote its jth column by h_j. We estimate H_0 by

$$\tilde{H} = \operatorname*{argmin}_{H\in\mathbb{R}^{d\times q}} \frac{1}{2n}\sum_{i=1}^{n}\|x_i - Hz_i\|^2 + \lambda_2 \sum_{j=1}^{d}\sum_{k=1}^{q}|h_{jk}| \tag{8}$$

or

$$\tilde{H} = \operatorname*{argmin}_{H\in\mathbb{R}^{d\times q}} \frac{1}{2n}\sum_{i=1}^{n}\|x_i - Hz_i\|^2 + \lambda_2 \sum_{j=1}^{q}\|h_j\|. \tag{9}$$

It is worth pointing out that Zhang and Zhang (2014) and van de Geer et al. (2014) use the standard Lasso to calculate the approximate projection.

By the KKT conditions, we obtain

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{H}z_i)z_{ij}\right\| \le \lambda_2, 1 \le j \le q.$$

This implies that the vectors z_i and $x_i - Hz_i$ are nearly orthogonal for a small λ_2. Furthermore, Lemma 1 of the Supplementary Material states that we need a sparsity assumption on H_0 in the sense that $\lambda_2 \sum_{j=1}^{q} \|h_{0j}\|$ is small, where h_{0j} is the jth column of H_0. The orthogonality property is important in establishing the theoretical properties of the proposed estimator described below.

We are now ready to describe the proposed regularized projection score estimator. Define the score function in the direction H as

$$\Psi_n(\beta, \eta)[H] \equiv \frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\eta)(x_i - Hz_i). \tag{10}$$

As the parameter η is unknown, we replace it by the initial estimator $\tilde{\eta}$ given in (2). We also estimate H by $\tilde{H}$. We then define the regularized projection score function for β as

$$\tilde{\Psi}_n(\beta) \equiv \Psi_n(\beta, \tilde{\eta})[\tilde{H}] = \frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\tilde{\eta})(x_i - \tilde{H}z_i). \tag{11}$$

Thus, we estimate the parameter β_0 based on the following estimating equation:

$$\tilde{\Psi}_n(\beta) = 0. \tag{12}$$

Owing to the nonsmoothness of ψ_τ, $\tilde{\Psi}_n$ may not have an exact zero root. In that case, we only need to solve (12) within $o_p(n^{-1/2})$ precision. In Section 5, we will consider a series of minimization problems that corresponds to solving (12) in an iterative way.

We summarize the proposed regularized projection score approach in two steps:

(S1) estimate the vector η_0 and the matrix H_0 by solving (2) and (9), respectively;

(S2) estimate the parameter vector β_0 by solving the estimation equation (12).

3 Asymptotic Properties

In this section, we establish the asymptotic results for $\hat{\beta}$, where $\hat{\beta}$ is a solution of (12). The asymptotic results of the Lasso estimate $\tilde{\eta}$ and the block Lasso estimate $\tilde{H}$ have been given by Belloni and Chernozhukov (2011), Obozinski et al. (2011), and Wang et al. (2013). To simplify the presentation, we summarize their regularity conditions below; moreover, we need to make some additional assumptions.

(A1) z follows $N(0, \Sigma_z)$, and the covariance Σ satisfies $(x', z')'$ $0 < c_\Lambda < \Lambda_{\min}(\Sigma) < \Lambda_{\max}(\Sigma) < C_\Lambda < \infty$. $\|\beta_0\| + \|\eta_0\| + \max_{1\le j\le q} \|h_{0j}\| \le C_0$, where C_0 is a constant and h_{0j} is the jth column of H_0.

(A2) The coefficient η_0 is sparse with $s = o(n)$, and $\lambda_1 = O(\sqrt{\log(q)/n})$, where $S = \{j : \eta_{0j} \neq 0, j = 1, \cdots, q\}$ and $s = |S|$.

(A3) If the estimated coefficient matrix $\tilde{H}$ is obtained from (8), H_0 is sparse with $s_{h,k} \le s_h = o(1)$ for $1 \le k \le d$, where $S_{h,k} = \{j : h_{0kj} \neq 0,\ j = 1, \cdots, q\}$ and $s_{h,k} = |S_{h,k}|$. If $\tilde{H}$ is obtained from (9), H_0 is sparse with $s_h = o(1)$, where $S_h = \{j : h_{0j} \neq 0,\ j = 1, \cdots, q\}$, $s_h = |S_h|$. $s_h^2 \vee s^2 = o\left(\sqrt{n}/\log(q)\right)$, and $\lambda_2 = O\left(\sqrt{\log(q)/n}\right)$. There exists a constant $c_0 \in (0, 1]$ such that $\|\Sigma^{-1}_{S_h S_h}\|_\infty \le c_0$, where $\Sigma_{I_1 I_2}$ is the submatrix of Σ with row and column index sets I_1 and I_2, respectively.

(A4) $|f(u|x, z) - f(0|x, z)| \le C|u|^{1/2}$ for some constant C uniformly on (x, z) in a neighborhood of zero. $f(0|x, z)$ is uniformly bounded from above by $f_{\max} < \infty$ and from below by $f_{\min} > 0$ for all (x, z), where $f(\cdot|x, z)$ is the density function of

$\varepsilon = y - x'\beta_0 - z'\eta_0$.

(A5) $\max_{1\le j\le q} \mathrm{E}\{\|(x - H_0 z)z_j\|\} = O(1)$, $\max_{1\le j\le d} \mathrm{E}\{\|(x - H_0 z)x_j\|\} = O(1)$, and $\{E[\|z\|_\infty]^2\}^{1/2} \le \zeta_n$ with $(s\vee s_h)^{3/2}\zeta_n\lambda_2 = o(n^{1/2})$ and $\tau_n(s\vee s_h)\log(\zeta_n s_h\lambda_2\tau_n^{-1/2}) = o(1)$, where $\tau_n = (s\vee s_h)(\lambda_1\vee\lambda_2)$. For any w_i between $x_i'(\hat\beta - \beta_0) + z_i'(\tilde\eta - \eta_0)$ and 0, and for any $H \in \mathcal{U}_H$,

$$\max_{1\le j\le q}\left\| n^{-1}\sum_{i=1}^n f(w_i|x_i, z_i)(x_i - Hz_i)z_{ij}\right\| = o_p(s^{-1}\{\log(q)\}^{-1/2}),$$

where $\mathcal{U}_H = \{H \in \mathbb{R}^{d\times q} : n^{-1/2}\sum_{i=1}^n \|(H - H_0)z_i\| = O_p(s\log(q)/n)\}$.

(A6) $\mathrm{E}\,\{f(0|x, z)(x - H_0 z)x'\}$ is an invertible matrix.

Assumption (A1) imposes an eigenvalue restriction on the design matrix. Assumption (A2) is the mutual incoherence and self-incoherence condition that bounds the difference of the estimator $\tilde H$ and the true matrix H_0 and the difference of the estimator $\tilde\eta$ and the true parameter η_0. Under Assumptions (A1) and (A2), the conditions of Belloni and Chernozhukov (2011), Obozinski et al. (2011), and Wang et al. (2013) are satisfied. Assumption (A3) limits the increasing rate of the covariate dimension relative to the sample size to ensure that the Bahadur representation of the estimator $\hat\beta$ holds. Assumption (A4) is used to obtain $\hat\beta$, which is widely used in the quantile regression literature. Assumption (A5) imposes the orthogonality of $x - \tilde H z$ and z, where $x - \tilde H z$ is the projection of x to the space of z. As $\mathrm{E}\,\{(x - H_0 z)z_j\} = 0$ from the definition of H_0, Assumption (A5) holds if $(x - H_0 z)z_j$ is weakly correlated with $f(0|x, z)$, the conditional density around 0. Thus, it is weaker than the assumption of independence between (x, z) and ϵ, which is imposed by

Zhao et al. (2014) and Bradic and Kolar (2017). Similar conditions are used in Theorem 3.1 of van de Geer et al. (2014) when generalized linear models are considered.

Theorem 1 *Under model (1), if Assumptions (A1)–(A4) hold,*

$$\hat{\beta} \xrightarrow{p} \beta_0.$$

Theorem 2 *Under model (1), if Assumptions (A1)–(A6) hold,*

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, Q^{-1}DQ^{'-1}),$$

where $Q = E\{f(0|x,z)(x - H_0 z)x'\}$ *and* $D = \tau(1-\tau)E\{(x - H_0 z)(x - H_0 z)'\}$.

Theorem 2 establishes that the proposed estimator is asymptotically normal. However, under the high-dimensional setting, it is challenging to estimate the asymptotic covariance matrix $Q^{-1}DQ^{'-1}$, in which the density of the error term is involved. In the following section, we propose a resampling method that avoids the estimation of the error density at zero.

4 Refitted Wild Bootstrap

Adopting the ideas of the refitted cross-validation of Fan et al. (2011) and the wild bootstrap of Feng et al. (2011), we propose a refitted wild bootstrap method to estimate the asymptotic variance-covariance matrix of $\hat{\beta}$. This resampling method accounts for heterogeneous errors and can bypass the estimation of different densities of errors at zero. Unlike the method of Wang et al. (2018), which only considered a fixed number of covariates, the proposed refitted wild bootstrap method can deal with high-dimensional confounding covariates with divergent dimension q.

We randomly split the original dataset into two even parts and carry out the refitted wild bootstrapping using the following steps.

(B1) Estimate parameters using the method described in Section 2 and the first part of the dataset, and denote the estimates as $\tilde{\eta}_1$.

(B2) Use the second part of the dataset to estimate parameters using the regular quantile regression method based on the nonzero coefficient set determined by the vector $\tilde{\eta}_1$, and denote the estimate as $(\hat{\beta}_2', \tilde{\eta}_2')$, where the vector $\tilde{\eta}_2$ includes those zero coefficients determined in Step (B1) for notation consistency.

(B3) Independently generate weights ζ_i satisfying the following conditions:

(B3.1) there are two positive constants c_1 and c_2 satisfying $\sup\{\zeta \in \mathbb{G} : \zeta \leq 0\} = -c_1$ and $\inf\{\zeta \in \mathbb{G} : \zeta \geq 0\} = c_2$, where $\mathbb{G}$ is the support of ζ;

(B3.2) the distribution G of ζ satisfies $\int_0^{+\infty} \zeta^{-1} g(\omega) d\zeta = -\int_{-\infty}^{0} \zeta^{-1} g(\zeta) d\zeta = 1/2$ and $E_\zeta[|\zeta|] < \infty$, where $g(\zeta)$ is the density of ζ and the expectation E_ζ is taken under G;

(B3.3) the τth quantile of the weight ζ is zero.

(B4) Use the second part of the dataset to obtain the bootstrapped samples as $y_i^* = \hat{\beta}_2' x_i + \tilde{\eta}_2' z_i + \zeta_i |\hat{r}_i|$, where $\hat{r}_i = y_i - \hat{\beta}_2' x_i - \tilde{\eta}_2' z_i$.

(B5) Use the bootstrapped samples to estimate parameters with the method of Section 2, and denote the estimate of β_0 by $\hat{\beta}^*$.

(B6) Repeat (B2)–(B5) B times, and denote the sample variance of B copies of $\hat{\beta}^*$ as $\widehat{V}_2$.

Similarly, we use the second part of the dataset to determine those variables with nonzero coefficients, and the first part to estimate the variance-covariance matrix with the approach described in (B1)–(B6). Denote the estimated matrix as $\widehat{V}_1$. We use $(\widehat{V}_1+\widehat{V}_2)/2$ to estimate the variance of $\hat{\beta}$ and repeat the above procedure a certain number of times to reduce the randomness effects of splitting data.

The growth rate of the dimension of β in condition (A3) is too fast to ensure the validity of the refitted wild bootstrap of (B1)–(B6). We need to further limit the rate to be

(A3′) $s\log(q)/n^{1/3} \to 0$.

Let P^* denote the probability under the resampling procedure given in (B1)–(B6).

Theorem 3 *Under Assumptions (A1)–(A2), A(4)–A(6), and (A3′), using the resampling approach described in steps (B1)–(B6), we have*

$$\sup_{x\in\mathbb{R}}\left|P^*\big(n^{1/2}(\hat{\beta}^*-\hat{\beta})\le x\big)-P\big(n^{1/2}(\hat{\beta}-\beta_0)\le x\big)\right| \xrightarrow{p} 0.$$

Theorem 3 provides a theoretical justification for using the refitted wild bootstrap to estimate the asymptotic variance-covariance matrix. This makes it possible to conduct statistical inferences without estimating the error densities. In the following section, we describe a computational algorithm for solving the estimating equation (12).

5 Computation

As pointed out in Section 2, a question is how to solve

$$\tilde{\Psi}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\tilde{\eta})(x_i - \tilde{H}z_i) = 0. \tag{13}$$

Let $\tilde{y}_i = y_i - z_i'\tilde{\eta}$ and $\widetilde{x}_i = x_i - \tilde{H}z_i$. Write

$$\sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\tilde{\eta})(x_i - \tilde{H}z_i) = \sum_{i=1}^{n} \psi_\tau\{\tilde{y}_i - (\tilde{H}z_i)'\beta - \widetilde{x}_i'\beta\}\widetilde{x}_i.$$

Let β^k be the value at the kth iteration, $k = 0, 1, 2, \ldots$. We take the Lasso estimator by solving (2) as the initial estimator β^0, and use the following iterative steps.

Step 1: Calculate

$$\tilde{y}_i^k = \tilde{y}_i - (\tilde{H}z_i)'\beta^k.$$

Step 2: Solve

$$\beta^{k+1} = \operatorname*{argmin}_{\beta} \sum_{i=1}^{n} \rho_\tau(\tilde{y}_i^k - \widetilde{x}_i'\beta).$$

Step 3: Set $k \leftarrow k + 1$; go to Step 1 until certain convergence criteria are satisfied.

Note that Step 2 is an optimization problem based on a low-dimensional quantile regression, so it can be solved using existing software. Refer to Koenker (2005) for details on its computation.

6 One-Step Estimator

The procedure given in Section 5 inspired the following one-step estimation approach.

First, we obtain an initial estimator of β by solving (2). Recall that the projected score function is

$$\tilde{\Psi}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \psi_\tau(y_i - x_i'\beta - z_i'\tilde{\eta})(x_i - \tilde{H}z_i),$$

where $\tilde{H}$ is obtained by solving (9). We consider a modified projected score function

$$\tilde{\Psi}_n^*(\beta) = \frac{1}{n}\sum_{i=1}^{n} \psi\{y_i - (x_i - \tilde{H}z_i)'\beta - (\tilde{H}z_i)'\tilde{\beta} - z_i'\tilde{\eta}\}(x_i - \tilde{H}z_i).$$

Let $\tilde{y}_i = y_i - (\tilde{H}z_i)'\tilde{\beta} - z_i'\tilde{\eta}$. Then, solving $\tilde{\Psi}^*(\beta) = 0$ is equivalent to solving

$$\hat{\beta}_{one} = \underset{\beta}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\big(\tilde{y}_i - (x_i - \tilde{H}z_i)'\beta\big).$$

Clearly, $\hat{\beta}_{one}$ can be considered a one-step update from the initial estimator $\tilde{\beta}$.

We replace Assumption (A6) with the following assumption:

(A6′) $\mathrm{E}\{f(0|x,z)(x - H_0z)(x - H_0z)'\}$ is an invertible matrix.

We then have the following result.

Theorem 4 *Under model (1), if Assumptions (A1)–(A5) and (A6′) hold, then*

$$n^{1/2}(\hat{\beta}_{one} - \beta_0) \xrightarrow{L} N(0, \widetilde{Q}^{-1}D\widetilde{Q}^{-1}),$$

where $\widetilde{Q} = E\{f(0|x,z)(x - H_0z)(x - H_0z)'\}$ *and* D *is defined as in Theorem 2.*

Note that $\widetilde{Q}$ is different from the Q in Theorem 2, owing to the modification of the score function. Also, the refitted wild bootstrap method of Section 4 can be similarly used to estimate the asymptotic covariance matrix $\widetilde{Q}^{-1}D\widetilde{Q}^{-1}$. The computation of this estimator is efficient because no iterations of (Step 1)–(Step 2) are needed.

7 Numerical Studies

7.1 A Simulation Study

We investigate the finite-sample performance of the estimation method of Section 2 with the variance-covariance matrix estimated by the refitted wild bootstrap method described in Section 4. Two sample sizes, $n = 50$ and $n = 100$, are used, and two quantile levels, $\tau = 0.5$ and $\tau = 0.75$, are considered.

We simulate data from the model

$$y_i = \mu + \sum_{j=1}^{3} x_{ij}\beta_j + \sum_{k=1}^{199} z_{ik}\eta_k + e_i, \quad i = 1, \cdots, n,$$

where all the covariate variables and the model error e_i are independently generated from the standard normal distribution. We consider a sparsity structure with coefficients given as

$$(\mu, \beta_1, \beta_2, \beta_3, \eta_1, \eta_2, \eta_3, \cdots, \eta_{199}) = (3, 3, 3, 3, 3, 3, 3, 0, \cdots, 0).$$

We use the method of Huang et al. (2012) to solve (9), with the Bayesian information criterion for the choice of penalties, and the method of Belloni and Chernozhukov (2011) to solve (2) at confidence levels 0.7 and 0.8, corresponding to sample sizes $n = 50$ and $n = 100$, respectively. For the bootstrap procedure, we repeat 1000 times to estimate the covariance matrix, where the random weights follow the discrete distribution

$$P(W = w) = \begin{cases} 1 - \tau, & w = 2(1 - \tau) \\ \tau, & w = -2\tau \end{cases},$$

for $0 < \tau < 1$. The R packages *quantreg* and *grpreg* are used to solve (2) and (9), respectively. We generate 1,000 Monte Carlo samples to compare the performances of the proposed method and the oracle method where the sparsity structure is assumed to be known.

We report the biases of the proposed and the oracle estimators, and the relative efficiency, which is the ratio of the mean square error, of the oracle estimator and of the proposed one. We also estimate the coverage probabilities of the proposed method at the 95% confidence level. As demonstrated in Table 1, the bootstrap leads to overall conservative interval estimates, especially when the quantile level $\tau = 0.75$. When the sample size is as small as 50, the relative efficiencies vary from 70% to 82%; these efficiencies can be improved to be in the range of 82% to 92% when the sample size is doubled. It is obvious from the results shown in Table 1 that the proposed method usually leads to estimates with smaller biases. The relatively smaller biases of the proposed method are probably due to the projection procedure used in our estimation.

7.2 Case Study of GDP Growth Rate

In this section, we analyze the national growth rate of GDP, using data collected by Barro and Lee (2013). Their results indicate that educational attainment serves as a proxy for the stock of human capital in a broad group of countries, as well as the economic development. This dataset includes 138 countries and eight broad categories comprising national income, education, population/fertility, government expenditure, PPP deflators, political variables, and trade policy and others. A detailed description can be found at http://www.barrolee.com/. Data are presented either quinquennially for the years 1950–2010 or as averages of five-year sub-periods over 1950–2010.

There is a subset of data including 90 complete observations (by country) with 61

Table 1: Estimated coverage probability (CP) at 95% confidence level, and the estimated relative efficiencies (RE) and biases (Bias) of the proposed estimator (EC) and the oracle estimator (Oracle).

$n = 50$	Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
	β_1	-9.608	-0.971	0.811	95.9
$\tau = 0.5$	β_2	0.945	1.993	0.701	95.8
	β_3	-3.486	-8.541	0.813	96.2
	β_1	-2.744	-6.880	0.697	99.0
$\tau = 0.75$	β_2	2.891	-0.413	0.617	97.8
	β_3	-4.957	-12.802	0.690	98.7
$n = 100$	Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
	β_1	-2.245	-1.684	0.992	95.6
$\tau = 0.5$	β_2	-2.913	0.455	0.919	96.5
	β_3	-7.060	-6.316	0.948	96.2
	β_1	-5.663	-0.863	0.854	97.2
$\tau = 0.75$	β_2	1.615	1.790	0.927	97.7
	β_3	-8.156	-2.809	0.938	97.8

covariates, which can be downloaded in the R package *hdm* (Chernozhukov et al. 2016). There are 41 observations out of 90 from 1965; the rest are from 1975. In this example, we only consider the 49 observations from 1975. We choose the national GDP growth rates per capita to be the response y_i, and denote the 61 scaled covariates by $\widetilde{x}_i = (\widetilde{x}_{i1}, \cdots, \widetilde{x}_{ip})'$,

$i = 1, \cdots, n$, where $n = 49$ and $p = 61$. We first take the logarithm or cubic root transformation such that each predictor's empirical distribution is more normally distributed.

There is a large body of literature on the relationship between economic development and government consumption expenditure, such as Landau (1986), Barro (1990), Barro (1991), Barro (1989), Devarajan et al. (1996), d'Agostino et al. (2016), and Dissou et al. (2016), which demonstrates their association. Owing to the correlation between government consumption expenditure and other variables characterizing population/fertility, political instability, the economic system, and so on, we need to reduce their influence by the proposed regularized projection procedure.

The following two variables are important to understand the influence of government consumption expenditure of a country on its economic growth rate: *Ratio of real government "consumption" expenditure to real GDP* (govsh41, denoted by $\widetilde{x}_{i1}$) and *Ratio of real government "consumption" expenditure net of spending on defense and on education to real GDP* (gvxdxe41, denoted by $\widetilde{x}_{i2}$). We use these two variables as treatments, denoted by $x_i = (\widetilde{x}_{i1}, \widetilde{x}_{i2})'$, and the remaining ones as confounders, denoted by z_i, $i = 1, \cdots, n$. Then we consider the linear quantile regression model (1) on these treatments and confounders:

$$Q_\tau(y_i|x_i, z_i) = \beta_0 + \sum_{j=1}^{2} x_{ij}\beta_j + \sum_{k=1}^{59} z_{ik}\eta_k, \ i = 1, \cdots, 49.$$

We report the estimated coefficients and the corresponding p-values in Table 2.

Barro (1989, 1990, 1991) found that both variables, *govsh41* and *gvxdxe41*, were negatively associated with the GDP growth rate. However, our results indicate that it may be a good strategy to promote GDP growth by increasing the total government consumption expenditure for those slowly growing economies. At the same time, countries with relatively slow GDP growth rates should limit government expenditure on defense and education to ensure economic growth.

Table 2: List of p-values of two variables for GDP growth rate. The numbers in the parentheses are the estimated coefficient at corresponding quantile level. *govsh41: *Ratio of real government "consumption" expenditure to real GDP* and *gvxdxe41: *Ratio of real government "consumption" expenditure net of spending on defense and on education to real GDP.*

Variable Name	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
govsh41*	0.0122 (0.8215)	0.2722 (0.1971)	0.9356 (-0.00498)
gvxdxe41*	0.0043 (-0.6523)	0.0026 (-0.3530)	0.7259 (-0.2403)

8 Conclusion

In this work, we used regularized projection scores to estimate low-dimensional preconceived parameters in high-dimensional quantile regression models. The asymptotic results we obtained facilitate classical statistical inference in high-dimensional scenarios, which has been largely overlooked in the quantile regression literature. Also, we proposed a refitted wild bootstrapping approach to bypass the estimation of the variance-covariance matrix of the estimator, which involves the probability densities of the errors. To the best of our knowledge, this is the first demonstration of wild bootstrapping in high-dimensional circumstances in the quantile regression literature.

The proposed method can be easily implemented because its computation is carried out based on existing algorithms, which can be accomplished using R packages. In practice, we advocate the one-step estimator owing to its computational efficiency in high-dimensional settings, especially when the resampling approach is needed.

Acknowledgements

The authors are grateful to the anonymous referees, the associate editor and the editor for their helpful comments, which have led to a significant improvement in the paper. Feng's research was supported by National Natural Science Foundation of China (11971292 and 11690012), Program for Innovative Research Team of SUFE, and Key Laboratory for Applied Statistics of MOE (130028906), Northeast Normal University. Huang's work is supported in part by the NSF grant DMS-1916199. Liu's research was supported by National Natural Science Foundation of China (11771267).

Supplementary Material

The proofs of Theorems 1–4 and related technical details can be found in the Supplementary Material.

References

Barro, R. J. (1989), "A cross-country study of growth, saving, and government," in *National Saving and Economic Performance*, National Bureau of Economic Research, pp. 271–304.

— (1990), "Government spending in a simple model of endogeneous growth," *Journal of Political Economy*, 98, 103–125.

— (1991), “Economic growth in a cross section of countries,” *The Quarterly Journal of Economics*, 106, 407–443.

Barro, R. J. and Lee, J. W. (2013), “A new data set of educational attainment in the world, 1950–2010,” *Journal of Development Economics*, 104, 184–198.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012), “Sparse models and methods for optimal instruments with an application to eminent domain,” *Econometrica*, 80, 2369–2429.

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

Belloni, A., Chernozhukov, V., and Kato, K. (2018), “Valid post-selection inference in high-dimensional approximately sparse quantile regression models,” *Journal of the American Statistical Association*, 114, 749–758.

Belloni, A., Chernozhukov, V., and Wei, Y. (2013), “Honest confidence regions for a regression parameter in logistic regression with a large number of controls,” cemmap working paper CWP67/13, Centre for Microdata Methods and Practice, London.

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013), “Valid post-selection inference,” *The Annals of Statistics*, 41, 802–837.

Berk, R., Brown, L., and Zhao, L. (2009), “Statistical inference after model selection,” *Journal of Quantitative Criminology*, 26, 217–236.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, Springer-Verlag New York, Inc.

Bradic, J. and Kolar, M. (2017), "Uniform inference for high-dimensional quantile regression: linear functionals and regression rank scores," *arXiv preprint arXiv:1702.06209*.

Bühlmann, P. and van de Geer, S. (2011), *Statistics for High-Dimensional Data*, Springer Berlin Heidelberg.

Chernozhukov, V., Hansen, C., and Spindler, M. (2016), "hdm: high-dimensional metrics," *arXiv preprint arXiv:1608.00354*.

d'Agostino, G., Dunne, J. P., and Pieroni, L. (2016), "Government spending, corruption and economic growth," *World Development*, 84, 190–205.

Devarajan, S., Swaroop, V., and fu Zou, H. (1996), "The composition of public expenditure and economic growth," *Journal of Monetary Economics*, 37, 313–344.

Dezeure, R., Bhlmann, P., Meier, L., and Meinshausen, N. (2015), "High-dimensional inference: confidence intervals, p-values and R-software hdi," *Statistical Science*, 30, 533–558.

Dissou, Y., Didic, S., and Yakautsava, T. (2016), "Government spending on education, human capital accumulation, and growth," *Economic Modelling*, 58, 9–21.

Fan, J., Guo, S., and Hao, N. (2011), "Variance estimation using refitted cross-validation in ultrahigh dimensional regression," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 74, 37–65.

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

Fang, E. X., Ning, Y., and Liu, H. (2016), "Testing and confidence intervals for high dimensional proportional hazards models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 1415–1437.

Feng, X., He, X., and Hu, J. (2011), "Wild bootstrap for quantile regression," *Biometrika*, 98, 995–999.

Huang, J., Breheny, P., and Ma, S. (2012), "A selective review of group selection in high dimension models," *Statistical Science*, 27, 481–499.

Javanmard, A. and Montanari, A. (2014), "Confidence intervals and hypothesis testing for high-dimensional regression," *Journal of Machine Learning Research*, 15, 2869–2909.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.

Koenker, R. and Bassett, G. (1978), "Regression quantiles," *Econometrica*, 46, 33–50.

Landau, D. (1986), "Government and economic growth in the less developed countries: an empirical study for 1960-1980," *Economic Development and Cultural Change*, 35, 35–75.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016), “Exact post-selection inference, with application to the Lasso,” *The Annals of Statistics*, 44, 907–927.

Lee, J. D. and Taylor, J. E. (2014), “Exact post model selection inference for marginal screening,” in *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., pp. 136–144.

Lockhart, R., Taylor, J., Tibshirani, R. J., and Tibshirani, R. (2014), “A significance test for the Lasso,” *The Annals of Statistics*, 42, 413–468.

Meinshausen, N. (2014), “Group bound: confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 923–945.

Meinshausen, N., Meier, L., and Bhlmann, P. (2009), “p-Values for high-dimensional regression,” *Journal of the American Statistical Association*, 104, 1671–1681.

Neykov, M., Ning, Y., Liu, J. S., and Liu, H. (2018), “A unified theory of confidence regions and testing for high dimensional estimating equations,” *Statistical Science*, 33, 427–443.

Ning, Y. and Liu, H. (2017), “A general theory of hypothesis tests and confidence regions for sparse high dimensional models,” *The Annals of Statistics*, 45, 158–195.

Obozinski, G., Wainwright, M. J., and Jordan, M. I. (2011), “Support union recovery in high-dimensional multivariate regression,” *The Annals of Statistics*, 39, 1–47.

Rgamer, D. and Greven, S. (2018), "Selective inference after likelihood- or test-based model selection in linear models," *Statistics & Probability Letters*, 140, 7–12.

Stucky, B. and van de Geer, S. (2018), "Asymptotic confidence regions for high-dimensional structured sparsity," *IEEE Transactions on Signal Processing*, 66, 2178–2190.

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

Tibshirani, R. J., Taylor, J., Lockhart, R., and Tibshirani, R. (2016), "Exact post-selection inference for sequential regression procedures," *Journal of the American Statistical Association*, 111, 600–620.

van de Geer, S., Bhlmann, P., Ritov, Y., and Dezeure, R. (2014), "On asymptotically optimal confidence regions and tests for high-dimensional models," *The Annals of Statistics*, 42, 1166–1202.

Wang, L., Keilegom, I. V., and Maidman, A. (2018), "Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors," *Biometrika*, 105, 859–872.

Wang, L., Wu, Y., and Li, R. (2012), "Quantile regression for analyzing heterogeneity in ultra-high dimension," *Journal of the American Statistical Association*, 107, 214–222.

Wang, W., Liang, Y., and Xing, E. (2013), "Block regularized lasso for multivariate multi-response linear regression," in *Proceedings of the Sixteenth International Conference on*

Artificial Intelligence and Statistics, PMLR, vol. 31 of *Proceedings of Machine Learning Research*, pp. 608–617.

Yang, Y. (2017), "Statistical inference for high dimensional regression via constrained Lasso," *arXiv preprint arXiv:1704.05098*.

Zhang, C.-H. (2010), "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, 38, 894–942.

Zhang, C.-H. and Zhang, S. S. (2014), "Confidence intervals for low dimensional parameters in high dimensional linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76, 217–242.

Zhang, C.-H. and Zhang, T. (2012), "A general theory of concave regularization for high-dimensional sparse estimation problems," *Statistical Science*, 27, 576–593.

Zhao, T., Kolar, M., and Liu, H. (2014), "A general framework for robust testing and confidence regions in high-dimensional quantile regression," *arXiv preprint arXiv:1412.8724*.

Zheng, Q., Peng, L., and He, X. (2015), "Globally adaptive quantile regression with ultra-high dimensional data," *The Annals of Statistics*, 43, 2225–2258.

Zhu, Y. and Bradic, J. (2018), "Linear hypothesis testing in dense high-dimensional linear models," *Journal of the American Statistical Association*, 113, 1583–1600.