Statistica Sinica Preprint No: SS-2025-0044	
Title	High-dimensional Extreme Quantile Regression
Manuscript ID	SS-2025-0044
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0044
Complete List of Authors	Yiwei Tang, Huixia Judy Wang and Deyuan Li
Corresponding Authors	Deyuan Li
E-mails	deyuanli@fudan.edu.cn

High-Dimensional Extreme Quantile Regression

Yiwei Tang*, Huixia Judy Wang**, Deyuan Li*

Fudan University, Rice University***

Abstract: The estimation of conditional quantiles at extreme tails is of great interest in numerous applications. Various methods that integrate regression analysis with an extrapolation strategy derived from extreme value theory have been proposed to estimate extreme conditional quantiles in scenarios with a fixed number of covariates. However, these methods become less effective in high-dimensional settings, where the number of covariates grows with the sample size. In this article, we develop new estimation methods tailored for extreme conditional quantiles with high-dimensional covariates. We establish the asymptotic properties of the proposed estimators and demonstrate their superior performance through simulation studies, particularly in scenarios of growing dimension and high dimension where existing methods may fail. Furthermore, the analysis of auto insurance data validates the efficacy of our methods in estimating extreme conditional insurance claims and selecting important variables.

Key words and phrases: Extrapolation; Extreme value; High-dimensional data; Regression analysis.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

1. Introduction

Quantile regression has become a widely recognized and useful alternative to classical least-squares regression for analyzing heterogeneous data. Since its introduction by Koenker and Bassett (1978), quantile regression has gradually been extended to a wide variety of data analytic settings; for a comprehensive review, see Koenker (2005); Koenker et al. (2017).

While traditional quantile regression allows exploration of a wide range of conditional quantiles for $\tau \in [\tau_l, \tau_u]$ with $\tau_l, \tau_u \in (0, 1)$, there is often interest in the extreme tails where τ is close to 0 or 1. Without loss of generality, we focus the discussion on τ close to 1. The inherent challenge in estimating tail quantiles lies in the fact that the number of observations in the tails, that is, above the τth quantile, is often small. In various contexts, methods have been developed to study extreme quantile regression by leveraging extreme value theory. Chernozhukov (2005), Wang, Li and He (2012), and Wang and Li (2013) studied extreme conditional quantiles under the linear quantile regression framework, and Li and Wang (2019) focused on the linear quantile autoregressive model. Daouia et al. (2013) and Gardes and Stupfler (2019) considered nonparametric smoothing methods. Wang and Tsai (2009) and Youngman (2019) applied generalized Pareto models to analyze the exceedance at a high threshold. Velthoen et al. (2023) and

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

Gnecco et al. (2024) explored tree-based methods such as gradient boosting and random forest.

With advancements in data collection techniques such as genomics, economics, finance, and imaging studies, the dimension of covariates p is becoming larger and can grow with the sample size n. Existing methodology and theory in quantile regression for high-dimensional covariates have primarily focused on a central quantile level or compact quantile sets in $(0, 1)$. In situations where p grows with n, various studies have investigated asymptotic behaviors of quantile regression estimators, e.g., Welsh (1989), He and Shao (2000), Belloni et al. (2019), Pan and Zhou (2021), and He et al. (2023). Assuming the sparsity in regression coefficients, researchers have explored the ℓ_1-penalized quantile regression estimator and its generalizations (Belloni and Chernozhukov, 2011; Fan et al., 2014), concave penalties (Wang, Wu and Li, 2012), and smoothed quantile regression with concave penalties (Tan et al., 2022). Additionally, the debiased estimator of high-dimensional quantile regression has been studied for inference at central quantiles (Zhao et al., 2014; Bradic and Kolar, 2017).

In this article, we address the challenge of estimating extreme conditional quantiles of $Y \in \mathbb{R}$ given a set of predictors $\mathbf{X} \in \mathbb{R}^p$ in a high-dimensional context. The challenge lies in the scarcity of efficient samples

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

at the tail and the sparsity of samples in high-dimensional space. As noted by Gnecco et al. (2024), a lack of samples exceeding the corresponding conditional τth quantile leads to empirical estimators with large bias and variance. The high dimensionality of the predictor space introduces additional bias, as noise covariates can obscure true signals.

Current research on extreme quantiles primarily focuses on estimation within fixed-dimensional settings and is either not applicable or less competitive in high-dimensional settings. Although tree-based methods (Velthoen et al., 2023; Gnecco et al., 2024) demonstrate effectiveness in relatively large and fixed covariate dimensions, addressing scenarios where the covariate dimension grows with the sample size, comparable to the sample size itself, remains inadequately explored. Recently, Sasaki et al. (2024) proposed a high-dimensional tail index regression model, assuming that the dimension grows with n and is comparable to $n(1-\tau)$. However, this approach imposes constraints such as a Pareto tail and specific link functions between the extreme value index and covariates.

In this article, we propose novel estimators for extreme quantiles in high dimensions by integrating concepts from extreme value theory with regularized estimators. Building upon a linear conditional quantile model, we employ regularized quantile regression to estimate intermediate con-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

ditional quantiles and extrapolate them to extreme quantiles. An intermediate quantile level τ_n approaches one at a moderate rate such that $n(1-\tau_n) \to \infty$, while an extreme quantile level approaches one more rapidly, i.e., $n(1-\tau_n) \to C$ for some constant $C \geq 0$.

In this article, we propose novel estimators for extreme quantiles in high-dimensional settings by integrating extreme value theory with regularized estimation. Based on a linear conditional quantile model, we first estimate intermediate conditional quantiles using regularized quantile regression and then extrapolate these estimates to extreme quantiles. An intermediate quantile level τ_n approaches one at a moderate rate such that $n(1-\tau_n) \to \infty$, while an extreme quantile level approaches one more rapidly, i.e., $n(1-\tau_n) \to C$ for some constant $C \geq 0$.

The theoretical analysis of extreme quantiles in high dimensions is challenging. The limiting distribution of a quantile estimator could be intractable in high dimensions, even at central quantile levels. Existing work on central quantiles within $(0,1)$ often requires a common assumption, namely $f_Y(Q_Y(\tau|\mathbf{X})|\mathbf{X}) > C > 0$, where $f_Y(\cdot|\mathbf{X})$ is the conditional density of Y given $\mathbf{X}$, and $Q_Y(\tau|\mathbf{X})$ is the τth conditional quantile of Y; see for instance Belloni and Chernozhukov (2011), and Wang, Wu and Li (2012). However, this assumption is violated in the tails as τ approaches 1. It is even

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

more challenging to achieve a uniform result of the regularized estimator over the entire tail, which is crucial for constructing an effective extreme value index estimator. This paper addresses the gap and establishes the uniform error rate for quantile regression estimator in high dimensions and at intermediate quantiles. This rate includes terms analogous to those in high-dimensional quantile regression at a central quantile level, augmented by an inflationary component attributed to escalating quantile levels or diminishing effective sample sizes. In addition, we propose a refined Hill approach for estimating the extreme value index, which is based on a fixed number of intermediate high-dimensional quantile estimates. This framework enables us to analyze the uniform behavior of regularized quantile regression in the tail region and assess the rate at which $f_Y(Q_Y(\tau|\mathbf{X})|\mathbf{X})$ tends to zero. The error rates of the proposed refined Hill and extreme quantile estimators offer theoretical insights into the choice of intermediate quantile levels in high-dimensional contexts.

The rest of the article is organized as follows. In Section 2, we present the proposed estimators for the extreme value index and extreme conditional quantiles, and derive theoretical results for the proposed estimators using techniques from both extreme value theory and high-dimensional statistics theory. We assess the finite sample performance of the proposed

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

methods through a simulation study in Section 3 and the analysis of auto insurance data in Section 4. Technical details and additional information for Sections 3 and 4 are provided in the online Supplementary Material.

2. Extreme Quantile Estimation in High Dimensions

Suppose we observe a random sample $\{(\mathbf{X}_i, Y_i), i = 1, \ldots, n\}$ of the random vector $(\mathbf{X}, Y)$, where Y_i is the univariate response variable and $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T$ is the p-dimensional centralized design vector. This article considers the high dimensional case: $p := p(n) \to \infty$ as $n \to \infty$. Let $Q_Y(\tau|\mathbf{X}) = \inf\{y : F_Y(y|\mathbf{X}) \geq \tau\}$ denote the τth conditional quantile of Y given $\mathbf{X}$, where $F_Y(\cdot|\mathbf{X})$ is the cumulative distribution function (CDF) of Y given $\mathbf{X}$. Denote $\mathcal{X} \subset \mathbb{R}^p$ as the support of $\mathbf{X}$.

Throughout the article, we assume that $F_Y(\cdot|\mathbf{X})$ is in the maximum domain of attraction of an extreme value distribution $G_\gamma(\cdot)$ with the extreme value index (EVI) $\gamma > 0$, denoted by $F_Y(\cdot|\mathbf{X}) \in D(G_\gamma)$. Here $F_Y(\cdot|\mathbf{X}) \in D(G_\gamma)$ means, for a given random sample $Y_1, \ldots, Y_n$ from $F_Y(\cdot|\mathbf{X})$, there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that, $P\left((\max_{1\leq i\leq n} Y_i - b_n)/a_n \leq y |\, \mathbf{X}\right) \to G_\gamma(y) = \exp\{-(1+\gamma y)^{-1/\gamma}\}$, as $n \to \infty$, for $1 + \gamma y \geq 0$. In this paper, we assume $\gamma > 0$, which means that $Y|\mathbf{X}$ has heavy-tailed distributions as commonly seen in many applications, such as stock market returns, insur-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

ance claims, earthquake magnitudes, river flows during floods.

The main objective of this paper is to estimate the conditional quantile $Q_Y(\tau_n|\mathbf{X})$ when $\tau_n \to 1$ and $p \to \infty$ as $n \to \infty$. We address two distinct regimes for the quantile level τ_n as it approaches 1: intermediate order quantile levels such that $n(1-\tau_n) \to \infty$, and extreme order quantile levels with $n(1-\tau_n) \to C$, where $C \geq 0$ is some constant.

We focus on the following tail linear quantile regression model:

$$Q_Y(\tau|\mathbf{X}) = \beta_0(\tau) + X_1\beta_1(\tau) + \cdots + X_p\beta_p(\tau) =: \mathbf{Z}^T\boldsymbol{\beta}(\tau), \tag{2.1}$$

for all $\tau \in [\tau_{ln}, 1)$, where $\mathbf{Z} = (1, \mathbf{X}^T)^T$, $\tau_{ln} \to 1$ as $n \to \infty$, and the quantile slope coefficients $\boldsymbol{\beta}(\tau)$ may vary across $\tau \in [\tau_{ln}, 1)$. The linear quantile model in (2.1) is specifically assumed at the upper tail. Similar tail model specifications have also been considered in Wang, Li and He (2012), Li and Wang (2019), Xu, Hou and Li (2022). In high-dimensional settings, we extend this assumption and additionally impose a sparsity condition to ensure model identifiability. We assume there are s tail-relevant variables, that is, $s := \#\{j = 1, 2, \ldots, p : \beta_j(\tau) \neq 0, \exists \tau \in [\tau_{ln}, 1)\}$, and $s = o(n)$. A tail-relevant variable may influence the tail quantiles of $Y|\mathbf{X}$, even if it does not affect the central quantiles. Without loss of

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

generality, we assume that the first s slope coefficients are nonzero, i.e., $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \beta_1(\tau), \ldots, \beta_s(\tau), \mathbf{0}_{p-s})^T$.

We propose a three-step procedure to estimate the extreme conditional quantile. In the first step, we obtain ℓ_1-penalized quantile estimators at a sequence of intermediate quantile levels. While the Hill estimator (Hill, 1975) is a standard tool for estimating the EVI in heavy-tailed settings, it is based on log-excess averages and requires analyzing the tail quantile process (de Haan and Ferreira, 2006), which is computationally and theoretically challenging in high dimensions. To address these challenges, we develop a refined Hill estimator in the second step, which relies on quantile estimates at a fixed number of intermediate quantile levels that approach to one at the same rate. In the third step, we develop an extrapolation estimator for the extreme conditional quantile $Q_Y(\tau_n|\mathbf{X})$ by leveraging extreme value theory and the previous results.

Before presenting the proposed estimators and their theoretical properties, we first introduce the notations used throughout the paper. For a sequence of random variables $E_1, E_2, \ldots, E_n$, we denote its order statistics as $E_{(1)} \leq E_{(2)} \leq \ldots \leq E_{(n)}$. Given a random sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, let $\mathbb{G}_n(f) = \mathbb{G}_n\{f(\mathbf{Z}_i)\} := n^{-1/2}\sum_{i=1}^n [f(\mathbf{Z}_i) - \mathrm{E}\{f(\mathbf{Z}_i)\}]$ and $\mathbb{E}_n f = \mathbb{E}_n f(\mathbf{Z}_i) := n^{-1} \sum_{i=1}^n f(\mathbf{Z}_i)$. We denote the ℓ_2, ℓ_1, ℓ_∞ and ℓ_0 norms by $\|\cdot\|$, $\|\cdot\|_1$, $\|\cdot\|_\infty$

and $\|\cdot\|_0$, respectively. Let $\|\boldsymbol{\beta}\|_{1,n} = \sum_{j=1}^{p} \widehat{\sigma}_j |\beta_j|$ denote the weighted ℓ_1-norm with $\widehat{\sigma}_j^2 := \mathbb{E}_n(X_{ij}^2)$. Given a vector $\boldsymbol{\delta} \in \mathbb{R}^p$, and a set of indices $T \subset \{1, \ldots, p\}$, we denote by $\boldsymbol{\delta}_T$ the vector where $\delta_{Tj} = \delta_j$ if $j \in T, \delta_{Tj} = 0$ if $j \notin T$. We use $a \lesssim b$ to denote $a = O(b)$, meaning $a \leq cb$ for some constant $c > 0$ that does not depend on n; and $a \asymp b$ to denote $a = O(b)$ and $b = O(a)$. We use $a \lesssim_P b$ to denote $a = O_P(b)$. Additionally, we use $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For notational simplicity, let $F_i(\cdot) = F_Y(\cdot|\mathbf{X}_i)$ be the conditional distribution function of Y given $\mathbf{X}_i$, and denote $f_i(\cdot) = f_Y(\cdot|\mathbf{X}_i)$.

2.1 ℓ_1-penalized Intermediate Quantile Estimator

Define $\mathcal{T}_n := \{c(1-\tau_{0n}) : c \in [c_1, c_2]\}$, where τ_{0n} is an intermediate quantile level such that $\tau_{0n} \to 1$ and $n(1-\tau_{0n}) \to \infty$, and $0 < c_1 < c_2 < \infty$ are constants. For any τ such that $1 - \tau \in \mathcal{T}_n$, we define the ℓ_1-penalized quantile estimator of $\boldsymbol{\beta}(\tau)$ as

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{(\beta_0, \boldsymbol{\beta}^T)^T \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \beta_0 - \mathbf{X}_i^T \boldsymbol{\beta}) + \frac{\lambda\sqrt{\tau(1-\tau)}}{n} \|\boldsymbol{\beta}\|_{1,n}, \quad (2.2)$$

where $\lambda > 0$ is the penalization parameter.

For high dimensional settings, the choice of λ is crucial for achieving estimation consistency. As suggested by Bickel et al. (2009), λ should exceed a suitably rescaled (sub)gradient of the sample criterion function evaluated at the true parameter value. While the asymptotic scale of λ has been derived for central quantiles (Belloni and Chernozhukov, 2011; Zheng et al., 2013), this scale differs for intermediate quantiles as τ approaches one. Specifically, a subgradient of the penalized quantile regression objective function at $\boldsymbol{\beta}(\tau)$ is given by $\mathbb{E}_n[\mathbf{Z}_i\{\tau - I(Y_i \leq \mathbf{Z}_i^T\boldsymbol{\beta}(\tau))\}] + \lambda\sqrt{\tau(1-\tau)}n^{-1}(0, \mathbf{g}^T)^T$, where $\mathbf{Z} = (1, \mathbf{X}^T)^T$ and $\mathbf{g} = (g_1, \ldots, g_p)^T$ is a subgradient of $\|\boldsymbol{\beta}\|_{1,n}$ with $|g_j| = \widehat{\sigma}_j$ for $j = 1, 2, \ldots, p$. The infinity norm of the first term is of the same order for both central and intermediate quantiles; however, the second term varies in order due to the factor $\sqrt{\tau(1-\tau)}$. Lemma 1 in the Supplementary Material provides an asymptotic uniform lower bound of λ over $\tau \in \mathcal{T}_n$, and a practical choice for λ is discussed in Section 2.4.

A major challenge in establishing the theoretical properties of $\hat{\beta}(\tau)$ is that the condition $f_Y(Q_Y(\tau|\mathbf{X})|\mathbf{X}) > C > 0$ fails at intermediate quantiles, complicating the derivation of a quadratic lower bound for the Taylor expansion of the expected quantile loss function. Consequently, the general framework of Negahban et al. (2012) cannot be applied to obtain the convergence rate through the restricted strong convexity property of the empirical loss function. However, leveraging condition C5 and extreme value theory, we assess the rate at which the conditional quantile and density converge, determining how $f_Y(Q_Y(\tau|\mathbf{X})|\mathbf{X})$ approaches zero as τ goes to one. This allows us to establish an asymptotic quadratic lower bound for the standardized expected intermediate quantile loss (see Lemma 4 in the Supplementary Material).

To establish the theoretical properties of $\widehat{\boldsymbol{\beta}}(\tau)$ over $\tau \in \mathcal{T}_n$, we impose Conditions C1-C6. Full details are provided in the Supplementary Material. Below, we highlight the key condition C5.

Condition C5. There exists an auxiliary line $\mathbf{Z} \to \mathbf{Z}^T\boldsymbol{\theta}_r$ with $0 < r < 1$ and a bounded vector $\boldsymbol{\theta}_r$ such that for $U = Y - \mathbf{Z}^T\boldsymbol{\theta}_r$ and some heavy-tailed distribution function $F_0(\cdot)$ with density $f_0(\cdot)$, the following hold for some positive continuous functions $K(\cdot), K_1(\cdot), K_2(\cdot)$ on $\mathcal{Z} = (1, \mathcal{X})$.

(i) There exists positive sequences d_n, d_{1n} such that $K(\mathbf{Z}) \asymp d_n, K_1(\mathbf{Z}) \asymp d_{1n}$ hold for all $\mathbf{Z} \in \mathcal{Z}$.

(ii) For any sequence $t_n \to \infty$, if $d_n\{1 - F_0(t_n)\} \to 0$ and $d_n f_0(t_n) \to 0$ as $n \to \infty$, then

$$\left|\frac{1 - F_U(t_n|\mathbf{Z})}{K(\mathbf{Z})\{1 - F_0(t_n)\}} - 1\right| = \{1 - F_0(t_n)\}^{\delta} K_1(\mathbf{Z})\{1 + o(1)\}, \text{ and}$$
$$\left|\frac{f_U(t_n|\mathbf{Z})}{K(\mathbf{Z}) f_0(t_n)} - 1\right| = \{f_0(t_n)\}^{\delta} K_2(\mathbf{Z})\{1 + o(1)\},$$

holds uniformly for $\mathbf{Z} \in \mathcal{Z}$, where F_U and f_U are the conditional distribution function and density function of U, and $\delta > 0$ is a constant.

(iii) $U_0(t) := F_0^{-1}(1-1/t)$ satisfies the second-order condition $A_1(t)^{-1}$ $\{U_0(tz)/U_0(t) - z^{\gamma}\} \to z^{\gamma}(z^{\rho}-1)/\rho$, as $t \to \infty$ with $\gamma > 0, \rho < 0$, and $A_1(t) = \gamma d t^{\rho}$ with $d \neq 0$. Additionally, $f_0(t)$ is regularly varying at infinity with index $-1/\gamma - 1$.

Remark 1. *Condition C5 presents a novel framework tailored for high-dimensional settings, where $K(\mathbf{Z})$, $K_1(\mathbf{Z})$, and $K_2(\mathbf{Z})$ may be unbounded and affect the tail behavior of $Y|\mathbf{X}$. Unlike fixed-dimensional cases (Wang, Li and He, 2012; Chernozhukov, 2005), this condition addresses the combined complexities of high dimensionality and tail behavior. The parameter $\boldsymbol{\theta}_r$ is used to remove the high-dimensional location component, yielding $U = Y - \mathbf{Z}^T\boldsymbol{\theta}_r$, whose conditional tail distribution $1 - F_U(\cdot|\mathbf{Z})$ is asymptotically equivalent to the univariate heavy-tailed distribution $1 - F_0(\cdot)$. The conditions $d_n\{1 - F_0(t_n)\} \to 0$ and $d_n f_0(t_n) \to 0$ as $n \to \infty$ ensure that the tail effect of $1 - F_0(t_n)$ outweighs the diverging scale effect of $K(\mathbf{Z})$ in high dimensions. This requirement allows us to apply extreme value theory to infer the tail behavior of Y, and is mild, as it is automatically satisfied when the number of important variables s is finite (see Example 1). Additionally, Lemma 3 in the Supplementary Material demonstrates that under*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

Condition C5, a second-order condition of $F_Y(t|\mathbf{X})$ still holds, similar to the fixed-dimensional case.

Define $s_r := \|\boldsymbol{\theta}_r\|_0$. The following Theorem 1 establishes a uniform bound of $\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)$ for $\tau \in \mathcal{T}_n$.

Theorem 1. *Assume Conditions C1-C6 hold, along with $1-\tau_{0n} > \sqrt{s\log(p)/n}$, $d_n^{-1}d_{1n}^{1/\delta}(1-\tau_{0n}) \to 0$, $s_r d_n^{-\gamma}(1-\tau_{0n})^{\gamma} \to 0$, and $\lambda \asymp \sqrt{n\log p}/\sqrt{1-\tau_{0n}}$. Then, for any $\epsilon \in (0,1)$ and some positive constant C, there exist a positive integer N such that, for all $n > N$,*

$$\sup_{1-\tau\in\mathcal{T}_n} \|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\| \leq C d_n^{\gamma}(1-\tau_{0n})^{-\gamma-1}\sqrt{\frac{s\log(p\vee n)}{n}},$$

with probability at least $1-4\epsilon-5p^{-4}$.

Theorem 1 establishes bounds similar in spirit to those found in Chernozhukov (2005) but with an additional term, $d_n^{\gamma}(1-\tau_{0n})^{-\gamma-1}$, which inflates the upper bound. This term reflects the decay of the conditional density in the tails. Specifically, as shown in Remark S.2 in the Supplementary Material, under certain conditions, as $n \to \infty$, $f_Y(Q_Y(\tau|\mathbf{X})|\mathbf{X}) \sim d_n^{-\gamma}(1-\tau_{0n})^{\gamma+1}$, which describes the rate at which the conditional density approaches zero at intermediate quantiles. Theorem 1 poses both upper and lower bounds conditions on $1-\tau_{0n}$. The condition $1-\tau_{0n} > \sqrt{s\log(p)/n}$ pre-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

vents τ_{0n} from approaching one too quickly, aligning with the requirement for restricted strong convexity of the objective function (see Lemma 4 in the Supplementary Material). Restricted strong convexity refers to maintaining strong convexity within a restricted set (Negahban et al., 2012). In the high-dimensional context, this restricted set is defined such that the error vector lies within it with high probability, given a suitably chosen penalty and certain conditions. Intuitively, restricted strong convexity holds only if the prediction error of the penalized estimator, $\mathbf{Z}^T(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau))$, grows slower than the quantile of $Y|\mathbf{X}$. The conditions $s_r d_n^{-\gamma}(1-\tau_{0n})^{\gamma} \to 0$ and $d_n^{-1} d_{1n}^{1/\delta}(1-\tau_{0n}) \to 0$ ensure that the tail effect dominates the high-dimensional effect. In Section 2.3, we provide an example illustrating the feasibility of these conditions.

2.2 Refined Hill Estimator

Let $k = \lfloor n(1-\tau_{0n}) \rfloor$. For simplicity, we omit the rounding notation, as it does not impact the theoretical discussion that follows. Define quantile levels $\tau_{0n} = \tau_1 < \tau_2 < \cdots < \tau_J \in (0,1)$, where $\tau_j = 1 - l_j k/n, l_j = s^{j-1}, s \in (0,1)$. For each $j = 1, 2, \ldots, J$, we estimate $\boldsymbol{\beta}(\tau_j)$ by the ℓ_1-penalized quantile regression estimator in (2.2). Suppose that we are interested in estimating the extreme conditional quantile of Y given $\mathbf{X} =$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

$\mathbf{x} \in \mathcal{X}$. Let $\mathbf{z} = (1, \mathbf{x}^T)^T$, and we can define the refined Hill estimator as $\hat{\gamma} = \left(\sum_{j=1}^{J} \phi(l_j) \log(1/l_j)\right)^{-1} \sum_{j=1}^{J} \phi(l_j) \log\left(\mathbf{z}^T \widehat{\boldsymbol{\beta}}(\tau_j)/\mathbf{z}^T \widehat{\boldsymbol{\beta}}(\tau_1)\right)$, where $\phi(\cdot)$ is some positive measurable function.

The refined Hill estimator is a weighted variant of the Hill-type estimators found in the literature but with the following key differences. First, $\hat{\gamma}$ utilizes a fixed number of log excesses at intermediate quantile levels with the same rate, whereas the Hill estimator typically relies on quantiles approaching one at varying rates, as seen in (Wang, Li and He, 2012). Second, $\hat{\gamma}$ allows for flexible weighting through ϕ, while the Hill estimator commonly applies equal weights to upper quantiles (Hill, 1975; Wang, Li and He, 2012; Daouia et al., 2023). We will discuss the choice of the weight function ϕ, and the tuning parameters J and s in Section 2.4. Similar weighting schemes have been explored by other researchers to generalize Hill and Pickands estimators (Drees, 1995; Daouia et al., 2013; He et al., 2022) and for estimating the Weibull tail coefficient (Gardes and Girard, 2016; He et al., 2020). These methods were developed in fixed-dimensional settings where the theoretical guarantees of the refined estimators could be established using the tractable asymptotic normality of intermediate quantile estimates. While this strategy is not feasible in high-dimensional settings, we can leverage the uniform convergence results from Theorem 1

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

to establish the convergence rate of the refined Hill estimator.

Theorem 2. *Assume that* $\mathbf{x}$ *belongs to a compact set in* $\mathcal{X} \subset \mathbb{R}^p$. *Assume the conditions of Theorem 1, and that* $d_n^{\rho}(n/k)^{-\rho} \vee d_{1n}d_n^{-\delta}(n/k)^{\delta} = o\left((ns/k)\sqrt{\log(p \vee n)/n}\right)$ *and* $s_r = o\left((d_n^{\gamma+\rho}(n/k)^{\gamma+\rho}) \vee \left(d_n^{\gamma-\delta}d_{1n}(n/k)^{\gamma-\delta}\right)\right)$. *Then we have* $\hat{\gamma} - \gamma = O_p\left(ns/k\sqrt{\log(p \vee n)/n}\right)$.

Remark 2. *Theorem 2 suggests that if* $(ns/k)\sqrt{\log(p \vee n)/n} \to 0$, *then* $\hat{\gamma} \xrightarrow{p} \gamma$. *In Theorem 2, the conditions on* s_r *are posed to account for the error in the second-order term in the tail expansion of* $F_Y(\cdot|\mathbf{X})$. *These conditions can be simplified for specific models, as discussed in Section 2.3. Note that the convergence rate of* $\hat{\gamma}$ *is* $(ns/k)\sqrt{\log(p \vee n)/n}$, *which is slower than the typical* $k^{-1/2}$ *rate achieved in the fixed-dimensional case (Wang and Tsai, 2009; Daouia et al., 2013). The difference arises from the slower convergence of the intermediate quantile estimators in high-dimensional settings. Furthermore, to ensure consistency,* k *must satisfy* $k > n^{1/2}$, *which is stricter than the requirement in the fixed dimensional case (Wang, Li and He, 2012), where* $k > n^{\eta}$ *with* η *being a small positive constant. Finally, while the error bound of the* ℓ_1*-penalized intermediate quantile estimator depends on* γ, *the convergence rate of* $\hat{\gamma}$ *remains invariant across* γ.

2.3 Extreme Conditional Quantile Estimator

Let $\tau_n \to 1$ be an extreme quantile level such that $1-\tau_n = o(1-\tau_{0n})$. Suppose that we are interested in estimating the τ_nth conditional quantile of Y given $\mathbf{X} = \mathbf{x}$. Direct estimation using the ℓ_1-penalized quantile estimator in (2.2) is often inaccurate or unstable due to limited data in the extreme tail. However, we can leverage extreme value theory and extrapolate from intermediate to extreme quantiles by using the relationship between quantiles with levels approaching one at different rates.

Define $U_Y(t|\mathbf{X}) = \inf\{y : F_Y(y|\mathbf{X}) \geq 1 - 1/t\} = F_Y^{-1}(1-1/t|\mathbf{X})$ for $t \geq 1$, the $(1-1/t)$th quantile of $F_Y(\cdot|\mathbf{X})$. According to Corollary 1.2.10 in de Haan and Ferreira (2006), for a heavy-tailed distribution $F_Y(\cdot|\mathbf{X} = \mathbf{x}) \in D(G_\gamma)$, we have $U_Y(tz|\mathbf{x})/U_Y(t|\mathbf{x}) \to z^\gamma$, as $t \to \infty$. Motivated by this, for a given $\mathbf{x}$, with notation $\mathbf{z} = (1, \mathbf{x}^T)^T$, we estimate $Q_Y(\tau_n|\mathbf{x})$ by $\widehat{Q}_Y(\tau_n|\mathbf{x}) = \{(1-\tau_{0n})/(1-\tau_n)\}^{\hat{\gamma}}\,\mathbf{z}^T\widehat{\boldsymbol{\beta}}(\tau_{0n})$.

Theorem 3. *Assume the conditions of Theorem 2. Let $\tau_n \to 1$ be a quantile level such that $n(1-\tau_n) = o(k)$, then we have*

$$\frac{\widehat{Q}_Y\left(\tau_n|\mathbf{x}\right)}{Q_Y\left(\tau_n|\mathbf{x}\right)} - 1 = O_p\left(\log\left(\frac{k}{n(1-\tau_n)}\right)\frac{ns}{k}\sqrt{\frac{\log(p\vee n)}{n}}\right).$$

Additionally, if $\log[k/\{n(1-\tau_n)\}]\ (ns/k)\sqrt{\log(p\vee n)/n} \to 0$,

$\widehat{Q}_Y(\tau_n|\mathbf{x})/Q_Y(\tau_n|\mathbf{x}) \xrightarrow{p} 1$.

According to our estimation procedure, the error in estimating extreme quantiles primarily arises from the EVI estimator and the ℓ_1-penalized intermediate quantile estimator. Specifically, the error rate of the EVI estimator and the relative error rate of the intermediate quantile estimator are both given by $(ns/k)\sqrt{\log(p \vee n)/n}$. Since the EVI estimator contributes to the error rate at the exponent, the final error rate for the extreme quantile estimation is $\log[k/\{n(1-\tau_n)\}(ns/k)\sqrt{\log(p \vee n)/n}$. In the fixed dimension case, the relative error rate simplifies to $\log[k/\{n(1-\tau_n)\}]/\sqrt{k}$. Comparing this with our results, we observe that the inflated term arises from the EVI estimator, which has an error rate of $1/\sqrt{k}$ in the fixed-dimensional setting.

Remark 3. *We assume a first-order condition, $F_Y(\cdot|\mathbf{X}) \in D(G_\gamma)$, to derive our estimator's properties. However, establishing properties of $\widehat{Q}_Y(\tau_n|\mathbf{x})$ also requires a second-order condition. Condition C5 ensures this condition holds via an auxiliary linear transformation $U = Y - \mathbf{Z}^T\boldsymbol{\theta}_r$. Lemma 3 shows that under C5, $F_Y(\cdot|\mathbf{X})$ meets the second-order condition even for high-dimensional $\mathbf{X}$, maintaining a common EVI across $\mathbf{X} \in \mathcal{X}$. Thus, intermediate conditional quantiles at any $\mathbf{x} \in \mathcal{X}$ can be used to estimate the EVI consistently. Theorem 2 provides the convergence rate for $\hat{\gamma}(\mathbf{x})$, emphasizing its dependence on $\mathbf{x}$. Selecting an appropriate $\mathbf{x}$ is crucial,* *particularly in high-dimensional cases where data sparsity is prevalent. We recommend using the mean* $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$ *for better numerical stability and improved convergence. Specifically, this choice yields the rate* $\hat{\gamma}(\bar{\mathbf{x}}) - \gamma = O_p\left((n/k)\sqrt{s\log(p \vee n)/n}\right)$ *(details in Supplementary Material). Alternatively, a pooled estimator averaging* $\{\hat{\gamma}(\mathbf{X}_i)\}_{i=1}^{n}$ *could enhance efficiency but is computationally demanding and less practical for high-dimensional problems.*

The regularity conditions outlined in our paper encompass a broad range of conventional regression settings. To illustrate these conditions, we consider the location-scale shift model as a representative example.

Example 1. Consider the location-scale shift model, $Y = \alpha + \mathbf{X}^T\boldsymbol{\beta} + (1 + \mathbf{X}^T\widetilde{\boldsymbol{\sigma}})\varepsilon$, where $1 + \mathbf{X}^T\widetilde{\boldsymbol{\sigma}} > 0$ for $\mathbf{X} \in \mathcal{X}$, and $\varepsilon \sim F_0(\cdot)$ with F_0 satisfying the second order condition for some $\gamma > 0$ and $\rho < 0$. Obviously, $Q_Y(\tau|\mathbf{X}) = (\alpha + F_0^{-1}(\tau)) + \mathbf{X}^T(\boldsymbol{\beta} + \tilde{\boldsymbol{\sigma}}F_0^{-1}(\tau))$.

Suppose that the scale effect variable is finite, i.e., $\|\tilde{\sigma}\|_0 < \infty$. Under this model, Condition C5 holds with $K(\mathbf{Z}) = \left(1 + \mathbf{X}^T\widetilde{\boldsymbol{\sigma}}\right)^{1/\gamma}$, $K_1(\mathbf{Z}) = C\left\{\left(1 + \mathbf{X}^T\widetilde{\boldsymbol{\sigma}}\right)^{-\rho/\gamma} - 1\right\}$, where C is some constant and $\delta = -\rho$. The derivation and verification of the remaining conditions are provided in Remark S.3 of the Supplementary Material. Let $\|\boldsymbol{\beta}\|_0 = n^b$, where $0 \le b < 1$. Then $d_n \sim C$, $d_{1n} \sim C$, and $s_r = n^b$. We next discuss two special cases to simplify k's conditions and examine their feasibility.

(I). The non-zero components of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}(\tau)$ are at the same positions, i.e., $s = s_r = n^b$, where $0 < b < 1$.By Theorem 2 and 3, the condition on k is $s\sqrt{\log p}\sqrt{n} \lesssim k \lesssim \left\{(s^2 \log p)^{1/[2(1-\rho)]} n^{(1-2\rho)/[2(1-\rho)]}\right\} \wedge ns^{-1/(\gamma+\rho)}$. To ensure the existence of an appropriate k, the additional sparsity conditions are $s^{1-1/(2\rho)} \log(p)/n \to 0$ and $s^{1+2/(\gamma+\rho)} \log p/n \to 0$.

(II). Suppose that $\|\boldsymbol{\beta}\|_0$ is finite, that is, $b = 0$. In this case, the condition on k can be relaxed to $\sqrt{\log p}\sqrt{n} \lesssim k \lesssim (\log p)^{1/[2(1-\rho)]}\ n^{(1-2\rho)/[2(1-\rho)]}$, which holds for any $\rho < 0$ as long as $\log(p) \lesssim n$.

To summarize, mild conditions commonly assumed in the high-dimensional literature for s, p, and n are sufficient to ensure the existence of a suitable k. To choose an appropriate k, for example, assume $s = n^{1/4}$ and $p = n^2$, in which case we can choose $k = \lfloor cn^{1/4}(2n \log n)^{0.51} \rfloor$, where $c > 0$ is a constant.

2.4 Computational Issues

The proposed method relies on several tuning parameters/functions, including the penalization parameter λ, the integer k associated with the intermediate quantile level $\tau_{0n} = 1 - k/n$, the weight function $\phi(\cdot)$ and the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

number of intermediate quantile levels J in the refined Hill estimator. In this section, we will discuss practical methods for selecting these quantities.

Selection of λ. In high-dimension settings, the optimal choice of λ depends on many factors, including p, the design matrix and error distribution. Bickel et al. (2009) suggested that λ should exceed the supremum norm of a suitably rescaled (sub)gradient of the sample criterion function; Belloni and Chernozhukov (2011) proposed a similar approach at central quantiles. However, direct application at intermediate quantiles typically leads to over-shrinkage. Our theoretical study suggests that at intermediate quantiles, λ should scale as $C\sqrt{n\log p}/\sqrt{1-\tau_{0n}}$, but accurately estimating the constant C is challenging, as noted in Homrighausen and McDonald (2017); Chetverikov et al. (2021). In our implementation, we conduct 10-fold cross-validations to select λ for each $\tau_j, j = 1, 2, \ldots, J$. This involves splitting the data into ten folds and using each fold to evaluate the estimator based on the remaining folds. For each λ, we calculate the average quantile loss at τ_j across all folds and select the λ that minimizes this loss. Numerical studies demonstrate that cross-validation performs well across various scenarios, although developing efficient, data-driven alternatives remains an important future direction.

Selection of k. The intermediate quantile level, or equivalently k, is

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

key to balancing bias and variance: a small k increases variance, while a large k increases bias. A common approach is to select k at the first stable point in the plot of the EVI estimator versus k; see Neves et al. (2015) for a heuristic algorithm. However, this method is computationally intensive in high-dimensional cases, as the estimator needs to be calculated over a sequence of k values, and the EVI may exhibit local variations across $\mathbf{x}$. We propose a simpler rule: set $k = \lfloor c_0 n^{0.5+\delta_1} (\log p)^{0.5+\delta_2} \rfloor$, where $c_0 > 0$ is a constant, and δ_1 and δ_2 are small positive constants, guided by theoretical insights from Theorem 2 and Example 1. Sensitivity analyses (Supplementary Material) confirm estimator stability across a wide range of parameters ($c_0 \in [0.5, 2.3]$, $\delta_1 \in [0.005, 0.016]$, $\delta_2 \in [0.01, 0.08]$). We use $c_0 = 0.8$, $\delta_1 = 0.01$, and $\delta_2 = 0.05$ throughout our numerical studies.

Choice of weights. The refined Hill estimator assigns different weights to intermediate quantiles. Following He et al. (2022), we use weights $\phi(\ell_j) = \ell_j^a$ with spacing $\ell_j = s^{j-1}$, $a, s \in (0, 1)$. As shown, in the autoregression setup, the asymptotic variance of the refined Hill estimator is proportional to a term depending only on (J, s, a). Minimizing this term yields the optimal choice (J, s, a). Although the limiting distribution is intractable without debiasing in high-dimensional settings, numerical evidence shows these weights enhance efficiency compared to the unweighted Hill estimator.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

3. Simulation Study

We conduct a simulation study to assess our method for estimating extreme quantiles in high-dimensional settings. Data are generated as $Y_i = X_{i1} + X_{i2} + (1 + wX_{i1})\varepsilon_i, i = 1, 2, \ldots, n$, where $X_{ij} \sim U(0, 1)$ for $j = 1, 2, \ldots, p$ and ε_i are independent and identically distributed student t random variables, and w is a constant controlling the degree of heteroscedasticity. Under this model, the τth conditional quantile of Y is $Q_Y(\tau|\mathbf{X}_i) = \alpha(\tau) + \mathbf{X}_i^T\boldsymbol{\beta}(\tau)$, with $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^T$, $\alpha(\tau) = Q_\varepsilon(\tau)$, $\boldsymbol{\beta}(\tau) = (1 + wQ_\varepsilon(\tau), 1, 0, \ldots, 0)^T$, and $Q_\varepsilon(\tau)$ denoting the τth quantile of ε_i. We consider six cases: Cases 1, 3, and 5 use $\varepsilon_i \sim t(5)$, while Cases 2, 4, and 6 use $\varepsilon_i \sim t(2)$. Moreover, Cases 1–2 are homoscedastic ($w = 0$), Cases 3–4 use $w = 0.5$, and Cases 5–6 use $w = 0.9$. (Note that for $t(v)$, the extreme value index is $\gamma = 1/v$.) We examine two sample size and dimension settings: one with $p = \lfloor n^{0.5} \rfloor$ and $n = 1000, 5000$, and another with high-dimensional cases $(n, p) = (1000, 1000)$ and $(1000, 1200)$. Each scenario is replicated 500 times.

For each simulation, we estimate $Q_Y(\tau_n|\mathbf{x})$ at $\tau_n = 0.995, 0.999$ using three methods: extreme quantile regression (EQR) from Wang, Li and He (2012), high-dimensional quantile regression (HQR) from Belloni and Chernozhukov (2011), and the proposed high-dimensional extreme quantile regression (HEQR). The parameter k in EQR is set to $[4.5n^{1/3}]$, as

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

recommended by Wang, Li and He (2012).

Each simulation computes the integrated squared error (ISE) as ISE $= L^{-1}\sum_{l=1}^{L}\left\{\widehat{Q}_Y(\tau_n|\mathbf{X}_l^*)/Q_Y(\tau_n|\mathbf{X}_l^*)-1\right\}^2$, where $\mathbf{X}_1^*,\ldots,\mathbf{X}_L^*$ are $L=100$ random covariate points drawn from $\mathbf{X}$. The mean ISE (MISE) is the average ISE over 500 simulations, with its standard error estimated as the standard deviation of ISE divided by $\sqrt{T}$.

Table 1: The mean integrated squared error (MISE) and standard error (in parentheses) of different estimators for the extreme conditional quantile at $\tau_n = 0.995, 0.999$ for $\gamma = 0.2$ and $p = \lfloor n^{0.5} \rfloor$. All values are in percentages.

w	n	HEQR	EQR	HQR	HEQR	EQR	HQR
			$\tau_n = 0.995$			$\tau_n = 0.999$	
0	1000	2.9(0.05)	3.6(0.05)	5.7(0.13)	4.7(0.09)	5.2(0.09)	12.5(0.12)
	5000	1.6(0.02)	2.7(0.02)	3.7(0.04)	2.4(0.03)	4.2(0.05)	6.1(0.06)
0.5	1000	3.3(0.05)	4.4(0.06)	7.5(0.21)	4.8(0.10)	5.8(0.10)	14.6(0.14)
	5000	2.0(0.02)	3.4(0.03)	5.0(0.06)	2.4(0.03)	4.8(0.06)	6.9(0.08)
0.9	1000	3.7(0.05)	5.2(0.07)	9.0(0.29)	5.0(0.09)	6.5(0.12)	15.8(0.17)
	5000	2.3(0.02)	4.0(0.04)	6.0(0.08)	2.6(0.03)	5.1(0.06)	7.8(0.12)

HEQR: the proposed estimator; EQR: the method in Wang, Li and He (2012); HQR: the high-dimensional qunatile regression estimator in Belloni and Chernozhukov (2011).

We begin by examining the $p = \lfloor n^{0.5} \rfloor$ setting. Tables 1 and 2 summarize the MISE of three estimators for the extreme conditional quantile at $\tau_n = 0.995$, and 0.999. Results show that for all three methods, both MISE and its standard error increase as τ_n approaches 1, and as the distribution becomes more heavy-tailed or heteroscedastic. In all scenarios, the

Table 2: The mean integrated squared error (MISE) and standard error (in parentheses) of different estimators for the extreme conditional quantile at $\tau_n = 0.995, 0.999$ for $\gamma = 0.5$ and $p = \lfloor n^{0.5} \rfloor$. All values are in percentages.

w	n	HEQR	EQR	HQR	HEQR	EQR	HQR
			$\tau_n = 0.995$			$\tau_n = 0.999$	
0	1000	7.4(0.26)	10.7(0.56)	24.6(3.33)	15.0(0.68)	19.9(2.11)	69.7(16.05)
	5000	5.3(0.11)	7.7(0.12)	12.7(0.28)	7.2(0.23)	10.5(0.26)	28.5(2.78)
0.5	1000	8.5(0.31)	13.5(0.70)	32.3(3.70)	15.4(0.84)	23.8(2.63)	96.8(28.76)
	5000	6.3(0.14)	9.7(0.18)	17.1(0.42)	8.1(0.33)	11.9(0.39)	36.5(3.93)
0.9	1000	9.6(0.33)	16.1(0.80)	40.5(4.25)	16.0(0.86)	27.3(2.86)	119.9(38.54)
	5000	7.3(0.17)	11.5(0.23)	21.5(0.58)	9.4(0.39)	13.5(0.49)	46.9(5.11)

HEQR: the proposed estimator; EQR: the method in Wang, Li and He (2012); HQR: the high-dimensional qunatile regression estimator in Belloni and Chernozhukov (2011).

Table 3: The mean integrated squared error (MISE) and standard error (in parentheses) of different estimators for the extreme conditional quantile at $\tau_n = 0.995, 0.999$ for the setting $p \geq n$, with $(n, p) = (1000, 1000)$ and $(1000, 1200)$. All values are in percentages.

γ	w	p	HEQR	HQR	HEQR	HQR
			$\tau_n = 0.995$		$\tau_n = 0.999$	
0.2	0	1000	0.23(0.03)	0.19(0.02)	0.23(0.03)	0.43(0.03)
		1200	0.20(0.02)	0.16(0.02)	0.22(0.03)	0.30(0.03)
	0.5	1000	0.26(0.03)	0.23(0.02)	0.26(0.03)	0.50(0.03)
		1200	0.23(0.03)	0.20(0.02)	0.24(0.03)	0.35(0.03)
0.5	0	1000	0.32(0.04)	0.44(0.33)	0.41(0.06)	0.63(0.13)
		1200	0.41(0.05)	0.36(0.23)	0.51(0.05)	0.55(0.14)
	0.5	1000	0.34(0.06)	0.52(0.71)	0.39(0.07)	0.67(0.17)
		1200	0.34(0.05)	0.41(0.41)	0.44(0.08)	0.58(0.20)

HEQR: the proposed estimator; HQR: the high-dimensional qunatile regression estimator in Belloni and Chernozhukov (2011).

proposed method HEQR outperforms the others in terms of both MISE and standard error. Both EQR and HEQR rely on extrapolation using extreme value theory. In this setting, since p increases with n but remains smaller than n, ERQ method is still applicable, and both EQR and HEQR perform better than HQR. The advantages of HEQR over EQR come from its penalized estimation of $\boldsymbol{\beta}(\tau)$ and its use of the refined Hill estimator, compared to EQR's unweighted Hill estimator.

We next examine the setting where $p \geq n$. Table 3 summarizes the results for HEQR and HQR, with $(n, p) = (1000, 1000)$ and $(1000, 1200)$, as EQR is not applicable in these cases. While HQR performs reasonably well for small $\gamma = 0.2$ at $\tau_n = 0.995$, it is significantly less efficient than HEQR for all other scenarios, regarding both MISE and its standard error. Overall, HEQR achieves greater accuracy and robustness, especially for heavier-tailed distributions.

We perform a sensitivity analysis to assess the stability of the proposed method with respect to the choices of c_0, δ_1, and δ_2 in the rule of thumb $k = \lfloor c_0 n^{0.5+\delta_1} (\log p)^{0.5+\delta_2} \rfloor$. For brevity, we present results only for Case 1. The other cases exhibit similar behavior, and the results are provided in the Supplementary file. Figure 1 (a), (b) show the MISE of EQR, HQR and HEQR for the extreme conditional quantiles at $\tau_n = 0.995$ against

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

$c_0 \in [0.4, 3]$, with (δ_1, δ_2) fixed at $(0.01, 0.05)$ for $n = 1000$ and 5000. The shaded area represents the 95% pointwise confidence band for the MISE, constructed as the estimated MISE $\pm$ 1.96 times its standard error. The two horizontal lines represent the MISE of HQR and EQR, respectively, while the shaded area corresponds to the 95% pointwise confidence band of the MISE. For $c_0 \in [0.5, 2.3]$, the MISE of HEQR is consistently smaller than that of HQR and EQR, indicating that c_0 in this range is a suitable choice. Figures 1 (c), (d), and (e),(f) plot the MISE against $\delta_1 \in [0.005, 0.03]$ and $\delta_2 \in [0.02, 0.08]$, respectively, with the other two constants fixed at their recommended values. The horizontal line shows the MISE for the recommended values $c_0 = 0.8$, $\delta_1 = 0.01$, and $\delta_2 = 0.05$, while the shaded area indicates its 95% pointwise confidence band. The results indicate that the method is insensitive to variations in $\delta_1 \in [0.005, 0.016]$ and $\delta_2 \in [0.01, 0.08]$.

4. Analysis of Auto Insurance Claims

In this section, we analyze an auto insurance claims dataset to investigate the effects of various factors on the higher quantiles of claim amounts. The data is available on Kaggle at `https://www.kaggle.com/datasets/xiaomengsun/car-insurance-claim-data`. Insurance claims data are typically heavy-tailed and heterogeneous, reflecting the diverse nature of risks

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

they embody and the array of factors that can influence claim sizes and frequencies. These factors not only complicate the analysis but also provide a rich vein of insights. We consider the following quantile regression model, $Q_{Y_i}(\tau|\mathbf{X}) = \beta_0(\tau) + X_{i1}\beta_1(\tau) + X_{i2}\beta_2(\tau) + \cdots + X_{ip}\beta_p(\tau), \quad i = 1, \ldots, n = 8423$, where Y represents the auto claim amount (in USD), and $X_{i1}, X_{i2}, \ldots, X_{ip}$ are $p = 43$ covariates included after preprocessing. To account for nonlinearity, we include quadratic terms for "Age" and "Income," as well as quadratic and cubic terms for "Vehicle Value" in the model, with both "Income" and "Vehicle Value" on the log-transformed scale. Detailed descriptions of variables appear in Section S.4 of the Supplementary Materials. We focus on high quantiles ($\tau = 0.991, 0.995, 0.999$) to analyze the large claims.

We conduct a cross-validation study comparing HEQR, EQR, and HQR methods in predicting extreme conditional quantiles. The data are randomly split into training (20%, $n_1 = 1684$) and testing (80%, $n_2 = 6739$) sets. For HEQR, we set $k = \lfloor 0.8 n_1^{0.51} (\log p)^{0.55} \rfloor$ and estimate γ at a central data point, with continuous covariates set to their mean values and binary variables set to their modes. Since $I\{Y < Q_Y(\tau|\mathbf{X})\}$ has a mean of τ and a variance of $\tau(1-\tau)$, this motivates us to consider the following prediction error (PE) metric, $\text{PE} = \{n_2\tau(1-\tau)\}^{-1/2} \sum_{j\in\mathcal{I}} [\tau - I\{Y_j < \widehat{Q}_Y(\tau|\mathbf{X}_j)\}]$, where $\mathcal{I}$ is the index set of the testing set. This metric measures the deviation between the empirical marginal distribution at the estimated extreme quantiles and the target quantile level τ_n. By aggregating over the entire test set, this marginal evaluation provides a more stable and robust assessment under tail sparsity. Similar metrics have been adopted in prior work on extreme quantile estimation, such as Xu, Wang and Li (2022). We repeat the cross-validation 500 times, summarizing the median absolute PE and median absolute deviation in Table 4. Results indicate that HEQR consistently achieves lower prediction errors than EQR and HQR, demonstrating superior performance at all three quantile levels.

Table 4: The median absolute value (with median absolute deviation in parentheses) of prediction errors for different methods at $\tau_n = 0.991$, 0.995, and 0.999, based on cross-validation of the auto claims data.

	$\tau_n = 0.991$	$\tau_n = 0.995$	$\tau_n = 0.999$
HEQR	1.84(1.13)	1.83(1.19)	1.28(0.63)
EQR	7.82(0.00)	5.82(0.00)	2.60(0.00)
HQR	8.95(2.06)	13.96(3.02)	32.86(6.74)

HEQR: the proposed estimator; EQR: the method in Wang, Li and He (2012); HQR: the high-dimensional qunatile regression estimator in Belloni and Chernozhukov (2011).

Next, we apply HEQR to the full dataset to estimate extreme quantiles and identify important predictors. Out of 43 covariates, 34 are selected at the intermediate quantile $\tau_{0n} = 1 - k/n$, where $k = \lfloor 0.8n^{0.51}(\log p)^{0.55}\rfloor$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044

and $n = 8423$. Notably, rural residence emerges as a significant predictor, aligning with findings by Clemente et al. (2023), who observed that rural residents typically exhibit lower auto insurance claims due to decreased claim frequency and severity. Additionally, our results confirm a negative relationship between vehicle age and claims, consistent with previous findings (Clemente et al., 2023). Figure 2(a) shows the ℓ_1-penalized quantile coefficient estimates $\widehat{\boldsymbol{\beta}}_j(\tau)$ for selected covariates at upper quantiles. The results indicate that being a commercial vehicle positively affects auto claims, with the effect increasing at higher quantiles. In contrast, "Rural Population" and "Time in Force" (the duration the insurance policy has been in effect, abbreviated as TIF) have negative effects, which become stronger at higher quantiles. The variable "Bachelor's Degree" shows little impact on auto claims. These findings highlight the heterogeneity of covariate effects, particularly in the upper tail of the claim distribution.

We also include the full-data analysis results from HQR for comparison. Figures 2(b) and 2(c) present the extreme quantile estimates of auto claims at $\tau_n = 0.991$ and 0.999 as a function of X_{20}, log-transformed Vehicle Value, for two groups of individuals with different TIF from both HQR and HEQR, with other continuous covariates set to their mean values and categorical variables to their modes. The two groups correspond to individuals

doi:10.5705/ss.202025.0044

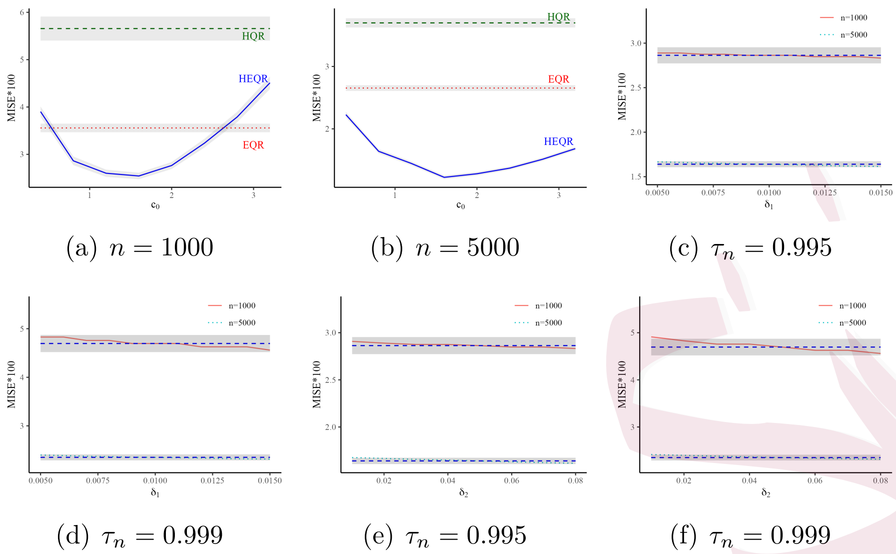
with TIF of 1 year and 7 years, representing the 0.25 and 0.75 quantiles, respectively. Results from HEQR suggest that, for both groups, higher vehicle values lead to larger claims, likely due to higher repair costs for more expensive vehicles, which aligns with common expectations and prior findings (Clemente et al., 2023). In contrast, HQR shows little variation in high quantiles of claims across vehicle values. Both methods suggest that shorter TIF is associated with higher claims at $\tau_n = 0.991$, likely due to less experienced or less stable policyholders. At $\tau_n = 0.999$, HEQR maintains clear differentiation between TIF groups, whereas HQR becomes unstable, indicating negligible group differences.

5. Discussion

Our work primarily addresses modeling heavy-tailed conditional distributions $Y \mid \mathbf{X}$. Extending the proposed methodology to include the special case $\gamma = 0$ represents an important direction for future research. In this scenario, the tail behavior of $Y \mid \mathbf{X}$ follows a Weibull-type distribution, characterized by additional parameters, including the Weibull tail coefficient (He et al., 2020; de Wet et al., 2016). Accommodating the case $\gamma = 0$ would entail substantial methodological and technical modifications.

Valid inference for extreme quantiles remains challenging, even in low

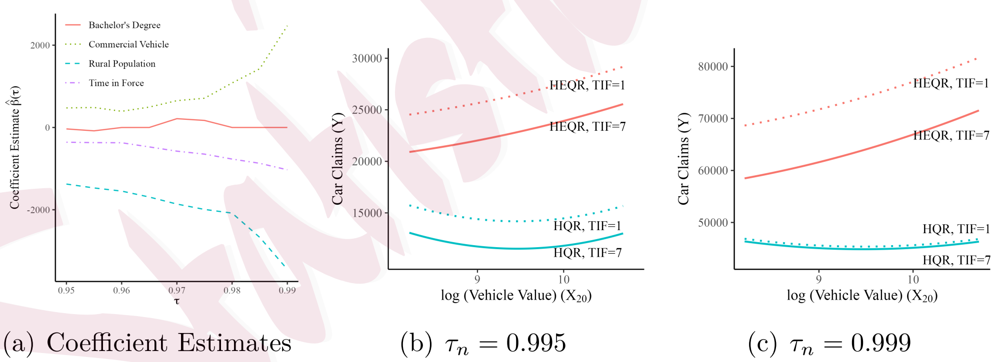
Statistica Sinica: Preprint
doi:10.5705/ss.202025.0044



(a) $n = 1000$ (b) $n = 5000$ (c) $\tau_n = 0.995$

(d) $\tau_n = 0.999$ (e) $\tau_n = 0.995$ (f) $\tau_n = 0.999$

Figure 1: The MISE of estimators for the extreme conditional quantile in Case 1: (a) and (b) against c_0 at $\tau_n = 0.995$ for $n = 1000$ and $n = 5000$, respectively; (c) and (d) against δ_1 at $\tau_n = 0.995$ and $\tau_n = 0.999$, respectively; (e) and (f) against δ_2 at $\tau_n = 0.995$ and $\tau_n = 0.999$, respectively.



(a) Coefficient Estimates (b) $\tau_n = 0.995$ (c) $\tau_n = 0.999$

Figure 2: (a) Estimated coefficients of selected covariates across quantile levels. (b) and (c) Estimated extreme quantiles of auto claims (\$) versus log-transformed Vehicle Value (X_{20}) from HEQR and HQR, for TIF $= 1$ and 7, at $\tau_n = 0.995, 0.999$. Other covariates are fixed at mean (continuous) or mode (categorical).

dimensions. One approach involves bias correction at intermediate quantiles using methods like the debiased Lasso (Javanmard and Montanari, 2014; van de Geer et al., 2014; Zhang and Zhang, 2014; Yan et al., 2023). However, estimating relevant nuisance parameters, such as the second-order parameter ρ and the function $K(\cdot)$, is practically difficult. Alternatively, bootstrap methods, which are effective at intermediate levels though less so at extreme quantiles, offer another promising path. For instance, Li and Wang (2019) suggest resampling blocks with geometrically distributed sizes. Given the proven success of bootstrap techniques in high-dimensional settings (Chatterjee and Lahiri, 2010, 2011; Wu and Wang, 2020), their adaptation to intermediate quantile inference warrants further study.

Supplementary Materials

The online Supplementary Material contains technical conditions, proofs, simulation results, and extra details on the auto insurance claims data—and a CSV file with the auto insurance claims dataset used in Section 4.

Acknowledgement

Deyuan Li's research was partially supported by the National Natural Science Foundation of China grants 11971115 and 12471279.

References

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

Belloni, A., Chernozhukov, V., Chetverikov, D. and Fernández-Val, I. (2019), 'Conditional quantile processes based on series or many regressors', *Journal of Econometrics* **213**(1), 4–29.

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009), 'Simultaneous analysis of lasso and dantzig selector', *Annals of Statistics* **37**(4), 1705–1732.

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

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

Chatterjee, A. and Lahiri, S. N. (2011), 'Bootstrapping lasso estimators', *Journal of the American Statistical Association* **106**(494), 608–625.

Chernozhukov, V. (2005), 'Extramal quantile regression', *Annals of Statistics* **33**(2), 806–839.

Chetverikov, D., Liao, Z. and Chernozhukov, V. (2021), 'On cross-validated lasso in high dimensions', *Annals of Statistics* **49**(3), 1300–1317.

Clemente, C., Guerreiro, G. R. and Bravo, J. M. (2023), 'Modelling motor insurance claim frequency and severity using gradient boosting', *Risks* **11**(9), 1–20.

Daouia, A., Gardes, L. and Girard, S. (2013), 'On kernel smoothing for extremal quantile regression', *Bernoulli* **19**(5B), 2557–2589.

Daouia, A., Stupfler, G. and Usseglio-Carleve, A. (2023), 'Inference for extremal regression with dependent heavy-tailed data', *Annals of Statistics* **51**(5), 2040–2066.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer Science & Business Media.

de Wet, T., Goegebeur, Y., Guillou, A. and Osmann, M. (2016), 'Kernel regression with Weibull-type tails', *Annals of the Institute of Statistical Mathematics* **68**, 1135–1162.

Drees, H. (1995), 'Refined pickands estimators of the extreme value index', *Annals of Statistics* **23**(6), 2059–2080.

Fan, J., Fan, Y. and Barut, E. (2014), 'Adaptive robust variable selection', *Annals of Statistics* **42**(1), 324–351.

Gardes, L. and Girard, S. (2016), 'On the estimation of the functional weibull tail-coefficient', *Journal of Multivariate Analysis* **146**, 29–45.

Gardes, L. and Stupfler, G. (2019), 'An integrated functional weissman estimator for conditional extreme quantiles', *REVSTAT-Statistical Journal* **17**(1), 109–144.

Gnecco, N., Terefe, E. M. and Engelke, S. (2024), 'Extremal random forests', *Journal of the American Statistical Association* **119**(548), 3059–3072.

He, F., Wang, H. J. and Tong, T. (2020), 'Extremal linear quantile regression with weibull-type tails', *Statistica Sinica* **30**(3), 1357–1377.

He, F., Wang, H. J. and Zhou, Y. (2022), 'Extremal quantile autoregression for heavy-tailed time series', *Computational Statistics & Data Analysis* **176**, 107563.

He, X., Pan, X., Tan, K. M. and Zhou, W.-X. (2023), 'Smoothed quantile regression with large-scale inference', *Journal of Econometrics* **232**(2), 367–388.

He, X. and Shao, Q.-M. (2000), 'On parameters of increasing dimensions', *Journal of Multivariate Analysis* **73**(1), 120–135.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *Annals of Statistics* **3**(5), 1163–1174.

Homrighausen, D. and McDonald, D. J. (2017), 'Risk consistency of cross-validation with lasso-type procedures', *Statistica Sinica* **49**(3), 1017–1036.

Javanmard, A. and Montanari, A. (2014), 'Confidence intervals and hypothesis testing for high-dimensional regression', *The Journal of Machine Learning Research* **15**(1), 2869–2909.

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

Koenker, R. and Bassett, J. G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Koenker, R., Chernozhukov, V., He, X. and Peng, L. (2017), *Handbook of quantile regression*, CRC press.

Li, D. and Wang, H. J. (2019), 'Extreme quantile estimation for autoregressive models', *Journal of Business & Economic Statistics* **37**(4), 661–670.

Negahban, S. N., Ravikumar, P., Wainwright, M. J. and Yu, B. (2012), 'A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers', *Statistical Science* **27**(4), 538–557.

Neves, M. M., Gomes, M. I., Figueiredo, F. and Prata Gomes, D. (2015), 'Modeling extreme events: sample fraction adaptive choice in parameter estimation', *Journal of Statistical Theory and Practice* **9**(1), 184–199.

Pan, X. and Zhou, W.-X. (2021), 'Multiplier bootstrap for quantile regression: non-asymptotic theory under random design', *Information and Inference* **10**(3), 813–861.

Sasaki, Y., Tao, J. and Wang, Y. (2024), 'High-dimensional tail index regression: with an application to text analyses of viral posts in social media', arXiv:2403.01318.

Tan, K. M., Wang, L. and Zhou, W.-X. (2022), 'High-dimensional quantile regression: Convolution smoothing and concave regularization', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84**(1), 205–233.

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014), 'On asymptotically optimal confidence regions and tests for high-dimensional models', *Annals of Statistics* **42**(3), 1166–1202.

Velthoen, J., Dombry, C., Cai, J.-J. and Engelke, S. (2023), 'Gradient boosting for extreme quantile regression', *Extremes* **26**(4), 639–667.

Wang, H. J. and Li, D. (2013), 'Estimation of extreme conditional quantiles through power transformation', *Journal of the American Statistical Association* **108**(503), 1062–1074.

Wang, H. J., Li, D. and He, X. (2012), 'Estimation of high conditional quantiles for heavy-tailed distributions', *Journal of the American Statistical Association* **107**(500), 1453–1464.

Wang, H. and Tsai, C.-L. (2009), 'Tail index regression', *Journal of the American Statistical Association* **104**(487), 1233–1240.

Wang, L., Wu, Y. and Li, R. (2012), 'Quantile regression for analyzing heterogeneity in ultra-high dimension', *Journal of the American Statistical Association* **107**(497), 214–222.

Welsh, A. (1989), 'On m-processes and m-estimation', *Annals of Statistics* **17**(1), 337–361.

Wu, Y. and Wang, L. (2020), 'A survey of tuning parameter selection for high-dimensional regression', *Annual review of statistics and its application* **7**(1), 209–226.

Xu, W., Hou, Y. and Li, D. (2022), 'Prediction of extremal expectile based on regression models with heteroscedastic extremes', *Journal of Business & Economic Statistics* **40**(2), 522–536.

Xu, W., Wang, H. J. and Li, D. (2022), 'Extreme quantile estimation based on the tail single-index model', *Statistica Sinica* **32**(2), 893–914.

Yan, Y., Wang, X. and Zhang, R. (2023), 'Confidence intervals and hypothesis testing for high-dimensional quantile regression: Convolution smoothing and debiasing', *Journal of Machine Learning Research* **24**(245), 1–49.

Youngman, B. D. (2019), 'Generalized additive models for exceedances of high thresholds with an application to return level estimation for us wind gusts', *Journal of the American Statistical Association* **114**(528), 1865–1879.

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**(1), 217–242.

Zhao, T., Kolar, M. and Liu, H. (2014), 'A general framework for robust testing and confidence regions in high-dimensional quantile regression', arXiv:1412.8724.

Zheng, Q., Gallagher, C. and Kulasekera, K. (2013), 'Adaptive penalized quantile regression for high dimensional data', *Journal of Statistical Planning and Inference* **143**(6), 1029–1038.

Department of Statistics and Data Science, School of Management, Fudan University, Fudan University, Shanghai, China

E-mail: yiweitang80@gmail.com

Department of Statistics, The George Washington University, Washington, DC, United States

E-mail: judywang@email.gwu.edu

Department of Statistics and Data Science, School of Management, Fudan University, Fudan University, Shanghai, China

E-mail: deyuanli@fudan.edu.cn