Journal of Machine Learning Research 24 (2023) 1-49 Submitted 10/22; Revised 4/23; Published 7/23

Confidence Intervals and Hypothesis Testing for High-dimensional Quantile Regression: Convolution Smoothing and Debiasing

Yibo Yan yyb820nb@sina.com
School of Statistics
East China Normal University
Shanghai, 200062, China

Xiaozhou Wang* xzwang@sfs.ecnu.edu.cn
School of Statistics
Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE
East China Normal University
Shanghai, 200062, China

Riquan Zhang† zhangriquan@163.com
School of Statistics and Information
Shanghai University of International Business and Economics
Shanghai, 201620, China

Editor: Zhihua Zhang

Abstract

ℓ_1-penalized quantile regression (ℓ_1-QR) is a useful tool for modeling the relationship between input and output variables when detecting heterogeneous effects in the high-dimensional setting. Hypothesis tests can then be formulated based on the debiased ℓ_1-QR estimator that reduces the bias induced by Lasso penalty. However, the non-smoothness of the quantile loss brings great challenges to the computation, especially when the data dimension is high. Recently, the convolution-type smoothed quantile regression (SQR) model has been proposed to overcome such shortcoming, and people developed theory of estimation and variable selection therein. In this work, we combine the debiased method with SQR model and come up with the debiased ℓ_1-SQR estimator, based on which we then establish confidence intervals and hypothesis testing in the high-dimensional setup. Theoretically, we provide the non-asymptotic Bahadur representation for our proposed estimator and also the Berry-Esseen bound, which implies the empirical coverage rates for the studentized confidence intervals. Furthermore, we build up the theory of hypothesis testing on both a single variable and a group of variables. Finally, we exhibit extensive numerical experiments on both simulated and real data to demonstrate the good performance of our method.

Keywords: High-dimensional quantile regression, convolution-based smoothing, debiased method, hypothesis testing, non-asymptotic statistics

$*$. Corresponding Author
†. Corresponding Author

1. Introduction

Due to the development of modern technology, massive complex datasets are gradually becoming the main objects of research today, which is profoundly affecting statistics community. In addition to the explosive growth in scale and dimension of data, the heterogeneity of intrinsic data structure and the presence of outliers hinder the use of some classical methods. Take linear model as an example, ordinary least squares (OLS) estimator is statistically efficient under Gaussian and other light-tailed noises. However, it may fail to be consistent when dealing with heavy-tailed errors and it is also highly susceptible to outliers. To overcome the shortcomings of OLS estimator, Koenker and Bassett (1978) first proposed quantile regression (QR) in his famous seminal work. Compared to OLS regression, quantile regression is robust against the outliers and allows capturing the feature of the entire conditional distribution function. A comprehensive and systematic overview of quantile regression can be referred to Koenker et al. (2017).

Consider a linear model $y = \boldsymbol{x}^\top \boldsymbol{\beta}^* + \varepsilon$ in the high-dimensional setup, where the number of features p can be much greater than the sample size n. Under the assumption that the true parameter $\boldsymbol{\beta}^*$ is s-sparse, a widely-used way to perform parameter estimation and variable selection is applying penalized methods to minimize the empirical risk function. To achieve sparse estimate, Tibshirani (1996) proposed Lasso method that using ℓ_1-penalty to shrink OLS estimator. Over the last three decades, there are numerous articles focusing on penalized regression to extract inherent low-dimensional features from high-dimensional parameter space, and one can refer to Bühlmann and Van de Geer (2011), Hastie et al. (2019), Wainwright (2019), Fan et al. (2020) for an extensive and in-depth review of high-dimensional statistics. Beyond estimation and variable selection, another topic that appeals to statisticians is high-dimensional statistical inference. Due to the non-negligible bias induced by ℓ_1-regularization, many different concave penalties have been proposed to eliminate the shrinkage bias, such as SCAD designed by Fan and Li (2001) and MCP designed by Zhang (2010). Other works about non-convex penalized approach can be found in Fan and Lv (2011), Wang et al. (2013) and Fan et al. (2014). However, these methods rely on oracle properties, which is to say one can only make valid inference on the active set. Under specific minimum signal strength condition (Zhao and Yu, 2006; Wainwright, 2009), asymptotic normality for the selected variables can be verified, while other variables have no theoretical guarantees.

Another stream of high-dimensional inferential methods is termed debiasing technique originated from Zhang and Zhang (2014), in which they implemented low-dimensional projection (LDP) to correct the bias caused by Lasso estimator and then used the debiased Lasso estimates to construct confidence intervals. One prominent advantage of debiasing method is that it enables statistical inference uniformly on all parameters, since it does not need to impose the minimum signal strength condition. Van de Geer et al. (2014) reformulated the KKT condition of ℓ_1-penalized least squares regression and obtained the debiased (or desparsified in their work) Lasso estimator from an alternative way. The asymptotic normality they established depends on the accuracy of an approximate inverse of the sample covariance matrix, where the authors applied nodewise regression (see Meinshausen and Bühlmann, 2006) to achieve. Furthermore, Javanmard and Montanari (2014a,b) constructed the sample precision matrix (inverse of sample covariance matrix) via a series of convex programs, and then employed such matrix to design debiased estimator and applied it to the establishment of confidence intervals and hypothesis testing.

In the context of high-dimensional quantile regression (HDQR), Belloni and Chernozhukov (2011) first investigated ℓ_1-penalized QR model and developed a set of results on model selection and parameter estimation. Later, Wang et al. (2012) studied the methodology and theory of nonconvex penalized QR model via SCAD penalty. In Zheng et al. (2015), they used adaptive ℓ_1-penalty to acquire consistent shrinkage of regression quantile estimates across a continuous range of quantiles levels. Belloni et al. (2015, 2019) further considered the inference task of high-dimensional instrumental variable quantile regression (IVQR) model and employed Neyman's orthogonal score method to realize the goal. Latest results on the estimation and inference of high-dimensional censored QR model can be referred to Zheng et al. (2018) and Fei et al. (2023). By contrast, there is little work on hypothesis testing within this HDQR framework. In Zhao et al. (2014), they introduced a robust testing procedure by means of combining debiasing technique with composite quantile regression (CQR) model that proposed in Zou and Yuan (2008). Shortly afterwards Bradic and Kolar (2017) extended previous work on debiased method to develop uniform testing results for a range of quantiles. Based on their defined high-dimensional rank scores, they provided a distribution-free estimator of the sparsity function and adapted it for inference involving the QR process. Recently, Cheng et al. (2022) engaged researched on treatment effects in QR with high-dimensional confounding covariates. After applying the regularized projection score method, the authors showed the consistency and asymptotic normality of their proposed estimator of the treatment effects. Current testing methods on HDQR will encounter great difficulties in operation, where the non-smoothness of the quantile loss further impairs the efficiency of practical algorithms as dimension p increases.

In this paper, we are committed to designing a more computationally efficient testing procedure for HDQR framework. Generally speaking, QR problem can be transformed into a linear program (LP), and solving it requires significant computing resources whenever n and p are large. To speed up resolving QR problem, Horowitz (1998) first proposed to smooth the quantile loss with a kernel function and the new estimator can be obtained via first-order optimization algorithms. In Horowitz (1998), the author verified the asymptotic equivalence between the traditional QR estimator and the new proposed one. Subsequently, this smoothing method has been extended to various QR-related subjects, and we refer the reader to Whang (2006), Wu et al. (2015), Galvao and Kato (2016), Kaplan and Sun (2017), Chen et al. (2019) and de Castro et al. (2019) for more details. It should be pointed out that Horowitz's smoothing method gains smoothness at the cost of convexity, which cannot guarantee the solution is a global minima. This may lead to more complex situations in high-dimensional settings. An impressive work from Fernandes et al. (2021) recently presented a convolution-type smoothed quantile regression (SQR) model, of which the corresponding loss inherits convexity and is smooth with second derivative. For the fixed p setting, the researchers investigated the asymptotic properties of the SQR estimator. He et al. (2023) considered the SQR model under the "increasing dimension" regime that allows dimension p to grow with the sample size n while $p < n$, and provided an in-depth statistical analysis from the non-asymptotic viewpoint. Moreover, Tan et al. (2022b) extended this convolution smoothing technique to the HDQR framework. In their work, they proposed a multi-step procedure to iteratively solve a sequence of weighted ℓ_1-penalized SQR loss minimization. They showed that their iterative estimator achieves the optimal rate of convergence, and obtained the oracle rate under the minimum signal strength condition. As a special case, they also derived properties of ℓ_1-SQR estimate, the solution of minimizing SQR loss with Lasso penalty. For CQR model, Yan et al. (2023) and Moon and Zhou (2022) employed the similar smoothing method to overcome the computational challenges therein and also achieved abundant theoretical results in the increasing dimension regime and high-dimensional setting respectively. Other relevant works about convolution-type SQR model can be seen in Jiang and Yu (2021), Tan et al. (2022a), Man et al. (2022), Sang et al. (2022) and Zhang and Zhu (2022).

Motivated by Tan et al. (2022b), in this work we propose the debiased ℓ_1-SQR estimator, and use it to establish confidence intervals and hypothesis testing in HDQR framework. Different from previous debiased estimators built by classical QR estimates, our proposed estimator is generated from ℓ_1-SQR estimator that can be efficiently obtained via proximal gradient descent (PGD) and ADMM-based algorithms. According to He et al. (2023) and Tan et al. (2022b), the bias induced by smoothing loss can be well controlled after reasonably tunning the bandwidth parameter, which is insensitive to estimated results. The key to eliminate the shrinkage bias of ℓ_1-SQR estimator is to approximate the inverse of the Hessian matrix of SQR model, which is constructed in the way similar to that of CLIME estimator (Cai et al., 2011). Unlike preceding works of Zhao et al. (2014) and Bradic and Kolar (2017), our debiased ℓ_1-SQR estimator does not need to estimate the sparsity function separately since it is already included in the inverse approximation. After imposing some assumptions on the population covariance matrix of covariate $\boldsymbol{x}$ and its smoothing type, we obtain the non-asymptotic error bounds of the approximate inverse matrix in ∞- and L_1-norm at first. On the basis of these error bounds, we further provide the Bahadur representation and the Berry-Esseen bound of our debiased ℓ_1-SQR estimator, which refines the results of Tan et al. (2022b) that only hold for oracle estimator and does not depend on the minimum signal strength condition. The Berry-Esseen bound immediately implies the empirical coverage rates for our proposed confidence intervals. Moreover, we consider the hypothesis testing problems for both a single variable and a group of variables. For testing an individual null $H_{0,j} : \beta_j^* = 0$, we take a minimax way to acquire uniformly good performance of tests over a family of sparse vectors. Specifically, we control the upper bound of Type I errors and the lower bound of statistical powers. To test a group of hypotheses $\{H_{0,j} : \beta_j^* = 0\}_{j\in\mathcal{G}}$, the Bonferroni procedure is applied within this work to control the familywise error rate (FWER), which will be discussed in the subsequent sections.

The main contributions of this work are as follows:

1. We design the debiased ℓ_1-SQR estimator, and develop inference and hypothesis testing for high-dimensional quantile regression. Different from previous works of Zhao et al. (2014) and Bradic and Kolar (2017), we relax the independence condition of $\boldsymbol{x}$ and ε and thus the heteroscedasticity is allowed in this work. Besides, the dimension p is allowed to be much larger than the sample size n, and the sparsity s can slowly increase with n satisfying $s \ll n$. We provide theoretical results in a non-asymptotic type while simultaneously allowing both the sparsity s and the ambient dimension p to depend on the sample size n.

2. Compared to the existing works of Zhao et al. (2014) and Bradic and Kolar (2017), our method is computationally efficient and easy to realize. On one end, our proposed debiased estimator is established upon the ℓ_1-SQR estimator, which minimizes a smooth convex programming that can be solved by PGD, ADMM and other efficient algorithms. On the other end, there is no need to estimate sparsity function additionally in the construction of our estimator, whereas Zhao et al. (2014) estimated it via Koenker's quotient estimator and Bradic and Kolar (2017) introduced a high-dimensional regression rank scores method. Beyond that, our simulation results also show that the debiased ℓ_1-SQR estimator has better performance for large p setup.

3. We design procedures for testing a single variable and a group of variables. To the best of our knowledge, theoretical results for testing a single variable under HDQR in the minimax way is first established in this paper. Besides, we control the FWER for a group of variables to cope with the simultaneous testing. As far as we know, this is the first time that the FWER method is introduced into HDQR testing problems.

The rest of the paper is organized as follows. Section 2 presents the background of sparse quantile regression and the debiasing method, followed by the convolution-type SQR with ℓ_1-regularization. We combine the debiasing method with SQR model and propose the debiased ℓ_1-SQR estimator in Section 2.3. In Section 3, we provide a comprehensive and in-depth analysis of the debiased ℓ_1-SQR estimator from a non-asymptotic viewpoint. Specifically, properties of the approximated inverse matrix are obtained in Section 3.1. The Bahadur representation for our proposed estimator is provided in Section 3.2. In Section 3.3, we construct confidence intervals and verify its non-asymptotic empirical rate via the Berry-Esseen bound. Section 3.4 introduces the theoretical results on hypothesis testing, and the content of simultaneous testing can be found in Section 3.5. Extensive numerical studies of our proposed method on simulated data are exhibited in Section 4, and a real data study is provided in Section 5. Conclusions and future works are discussed in Section 6. The proofs of all theoretical results are relegated in the appendix.

Notation: In this work, we use $\mathbb{R}^p$ to denote the the p-dimensional Euclidean space. For every $\boldsymbol{u} = (u_1, \dots, u_p)^\top \in \mathbb{R}^p$, define $\|\boldsymbol{u}\|_0 = \sum_{j=1}^p \mathbb{I}\{u_j \neq 0\}$, $\|\boldsymbol{u}\|_1 = \sum_{j=1}^p |u_j|$, $\|\boldsymbol{u}\|_2 = \sqrt{\sum_{j=1}^p u_j^2}$ and $\|\boldsymbol{u}\|_\infty = \max_{1\le j\le p} |u_j|$. We use bold capital letters to represent matrices throughout this paper. For any $p \times q$ matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p\times q}$, we define the elementwise ℓ_∞-norm $\|\mathbf{A}\|_\infty = \max_{1\le i\le p, 1\le j\le q} = |a_{ij}|$, the elementwise ℓ_1-norm $\|\mathbf{A}\|_1 = \sum_{i=1}^p \sum_{j=1}^q |a_{ij}|$, and the matrix ℓ_1-norm $\|\mathbf{A}\|_{L_1} = \max_{1\le i\le p} \sum_{j=1}^q |a_{ij}|$. Here we denote $\mathbf{I}$ as the identity matrix. For some $r \geq 0$, the unit sphere and the ℓ_1-norm ball in $\mathbb{R}^p$ are defined as $\mathbb{S}^{p-1} = \{\boldsymbol{u} \in \mathbb{R}^p : \|\boldsymbol{u}\|_2 = 1\}$ and $\mathbb{B}_1(r) = \{\boldsymbol{u} \in \mathbb{R}^p : \|\boldsymbol{u}\|_1 \leq r\}$ respectively. Moreover, for two sequences of non-negative numbers $\{x_n\}_{n\geq 1}$ and $\{y_n\}_{n\geq 1}$, $x_n \lesssim y_n$ means that there exists some constant $C > 0$ independent of n such that $x_n \leq C y_n$; $x_n \gtrsim y_n$ is equivalent to $y_n \lesssim x_n$; $x_n \asymp y_n$ is equivalent to $x_n \lesssim y_n$ and $y_n \lesssim x_n$. For two positive definite matrixes $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \succ \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive definite matrix; and $\mathbf{A} \prec \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is a positive definite matrix.

2. Background and problem setup

In this section, we first give a brief introduction to sparse quantile regression and the debiasing technique. Then we present the recent results about the convolution-type smoothed quantile regression. At the end of this part, we extend the debiasing technique to the high-dimensional smoothed quantile regression and propose the debiased ℓ_1-SQR estimator.

2.1 Sparse quantile regression: ℓ_1 penalty and debiasing

Consider a scalar response variable $y \in \mathbb{R}$ and a p-dimensional covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ with $x_1 \equiv 1$ across the paper. Let $F_{y|\boldsymbol{x}}(\cdot)$ be the conditional probability distribution function of y given $\boldsymbol{x}$. For some $0 < \tau < 1$, we model the τ-th conditional quantile of y given $\boldsymbol{x}$ as $F^{-1}_{y|\boldsymbol{x}}(\tau|\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta}^*$, where $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_p^*)^\top \in \mathbb{R}^p$ is the true parameter. Generate n random samples from $(y, \boldsymbol{x})$ and denote them as $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$. The preceding model assumption can be equivalently written as a linear model

$$y_i = \boldsymbol{x_i}^\top\boldsymbol{\beta}^* + \varepsilon_i \text{ and } \mathbb{P}\left(\varepsilon_i \le 0|\boldsymbol{x_i}\right) = \tau. \tag{1}$$

Throughout the paper, we consider the high-dimensional scenario that the number of features p can be much greater than the sample size n. More details about the relationship between p and n will be discussed in Section 3.

Based on the dataset $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$, the ℓ_1-penalized QR estimator $\widehat{\boldsymbol{\beta}}$ is the global minima to the following optimization problem

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \underbrace{\frac{1}{n}\sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})}_{=:\widehat{Q}(\boldsymbol{\beta})} + \lambda\|\boldsymbol{\beta}\|_1 \right\}, \tag{2}$$

where $\rho_\tau(u) = u(\tau - \mathbb{I}\{u < 0\})$ is the asymmetric absolute deviation function, namely the check loss function, with $\mathbb{I}\{\cdot\}$ representing the indicator function. Due to the shrinkage property of Lasso penalty, there are extensive works focused on estimation, prediction and variable selection consistency about ℓ_1-QR model including Wang et al. (2007), Belloni and Chernozhukov (2011), Wang (2013) and Zheng et al. (2015).

For the high-dimensional QR inference problems, the non-negligible bias induced by ℓ_1 penalty hinders from using $\widehat{\boldsymbol{\beta}}$ to achieve valid results. To overcome this obstacle, Belloni et al. (2015) and Belloni et al. (2019) proposed a three-stage refitting procedure based on Neyman's orthogonal score function to construct studentized confidence intervals. In contrast, Bradic and Kolar (2017) developed the debiasing technique proposed in Zhang and Zhang (2014) and Van de Geer et al. (2014) for Lasso estimator, and then employed the debiased ℓ_1-QR estimator to implement uniform hypothesis testing. Specifically, given ℓ_1-penalized QR estimator $\widehat{\boldsymbol{\beta}}$, a debiased estimator $\widetilde{\boldsymbol{\beta}}$ can be constructed as

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + \frac{1}{n}\widetilde{\mathbf{D}}\sum_{i=1}^n \boldsymbol{x}_i\Psi_\tau(y_i - \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}), \tag{3}$$

where $\Psi_\tau(u) = \tau - \mathbb{I}(u < 0)$ is the subderivative of $\rho_\tau(u)$ and $\widetilde{\mathbf{D}}$ is an estimator of the inverse of the matrix $\mathbb{E}[f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\boldsymbol{x}^\top]$ with $f_{\varepsilon|\boldsymbol{x}}$ being the conditional density function. Under the condition that ε and $\boldsymbol{x}$ are independent, $\widetilde{\mathbf{D}}$ can be estimated as $\widetilde{\mathbf{D}} = \widehat{\zeta}\widehat{\mathbf{D}}$ with $\widehat{\zeta}$ being an estimator of the sparsity function $1/f_{\varepsilon}(0)$ and $\widehat{\mathbf{D}}$ being an estimator of the inverse of the covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^{\top})$, such results can be found in Zhao et al. (2014) and Bradic and Kolar (2017). When $p > n$, $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^{\top}$ is not of full rank. To approximate $\boldsymbol{\Sigma}^{-1}$ in this scenario, one can turn to using sparse precision matrix estimators, such as celebrated CLIME estimator (Cai et al., 2011, 2016). After establishing the Bahadur representation of the debiased estimator (3), one can easily verify its asymptotic normality and immediately construct the confidence intervals.

Without the independence of ε and $\boldsymbol{x}$, $\widehat{\zeta}\widehat{\mathbf{D}}$ is no more an appropriate estimator of the inverse of $\mathbb{E}[f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\boldsymbol{x}^{\top}]$. Given some kernel function $K(\cdot)$, matrix $\mathbb{E}[f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\boldsymbol{x}^{\top}]$ can be estimated by Powell's kernel-type estimator (Powell et al., 1991) $\frac{1}{nh}\sum_{i=1}^{n} K((y_i - \boldsymbol{x}_i^{\top}\widehat{\boldsymbol{\beta}})/h)\boldsymbol{x}_i\boldsymbol{x}_i^{\top}$, in which $h > 0$ is the bandwidth parameter. Thus, $\widetilde{\mathbf{D}}$ can be alternatively defined as the approximate inverse of the matrix $\frac{1}{nh}\sum_{i=1}^{n} K((y_i - \boldsymbol{x}_i^{\top}\widehat{\boldsymbol{\beta}})/h)\boldsymbol{x}_i\boldsymbol{x}_i^{\top}$, similar manipulations can be seen in Lian and Fan (2018) and Zhao et al. (2019). In those two papers, the authors focused on proposing consistent debiased estimates for support vector machine (SVM) and QR, respectively, in high-dimensional distributed settings. Nevertheless, to the best of our knowledge, the asymptotic normality of this kind of kernel-based debiased estimators still remains unknown. Actually, $\frac{1}{nh}\sum_{i=1}^{n} K((y_i - \boldsymbol{x}_i^{\top}\boldsymbol{\beta})/h)\boldsymbol{x}_i\boldsymbol{x}_i^{\top}$ is exactly the Hessian matrix of the SQR empirical risk function evaluated at $\boldsymbol{\beta}$. This motivates us to design a novel debiased QR estimator by incorporating the ideas of convolution-type smoothing method. Since the inverse of $\mathbb{E}[f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\boldsymbol{x}^{\top}]$ is estimated as a whole, our newly proposed debiased estimator does not rely on the homoscedastic assumption that $\boldsymbol{x}$ and ε are independent of each other. Besides, we also certify the asymptotic normality of our proposed estimator in the following analysis.

2.2 Sparse quantile regression: a smoothing approach

Recall that $F_{\varepsilon|\boldsymbol{x}}(\cdot)$ is denoted as the conditional distribution of ε given $\boldsymbol{x}$. Then the population quantile risk function can be defined as

$$Q(\boldsymbol{\beta}) = \mathbb{E}\left\{\int_{-\infty}^{\infty} \rho_{\tau}(u - \boldsymbol{x}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}^{*}))\,\mathrm{d}F_{\varepsilon|\boldsymbol{x}}(u)\right\}.$$

It is not hard to check that $Q(\boldsymbol{\beta})$ is twice differentiable and strongly convex in a neighborhood of $\boldsymbol{\beta}^{*}$ whenever the conditional distribution $F_{\varepsilon|\boldsymbol{x}}(\cdot)$ is smooth enough. For every $\boldsymbol{\beta} \in \mathbb{R}^{p}$, let $\widehat{F}(\cdot;\boldsymbol{\beta})$ be the empirical cumulative distribution function (CDF) of the residuals $\{\mathrm{r}_i(\boldsymbol{\beta}) := y_i - \boldsymbol{x}_i^{\top}\boldsymbol{\beta}\}_{i=1}^{n}$, namely $\widehat{F}(u;\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\mathrm{r}_i(\boldsymbol{\beta}) \le u\}$ for any $u \in \mathbb{R}$. After replacing $F_{\varepsilon|\boldsymbol{x}}(\cdot)$ with $\widehat{F}(u;\boldsymbol{\beta})$, the empirical risk for quantile loss, $\widehat{Q}(\cdot)$ in (2), can be equivalently expressed as

$$\widehat{Q}(\boldsymbol{\beta}) = \int_{-\infty}^{\infty} \rho_{\tau}(u)\,\mathrm{d}\widehat{F}(u;\boldsymbol{\beta}). \tag{4}$$

Note that $\widehat{Q}(\cdot)$ is non-smooth since it is actually a finite sum of check loss functions evaluated at various points. Instead of using empirical CDF $\hat{F}(\cdot;\boldsymbol{\beta})$, Fernandes et al. (2021) introduced a kernel-type CDF and derived a new empirical risk through integrating over the probability measure induced by it. Given the residuals $\mathrm{r}_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}$ and a bandwidth parameter h, the kernel-type CDF $\widehat{F}_h(\cdot;\boldsymbol{\beta})$ is established as

$$\widehat{F}_h(u;\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{u} K_h\big(t - \mathrm{r}_i(\boldsymbol{\beta})\big)\,\mathrm{d}t,$$

where $K(\cdot)$ is a symmetric and non-negative kernel function and $K_h(u)$ is defined as $K_h(u) = K(u/h)/h$. Thus, the corresponding empirical risk function, which is denoted by $\widehat{Q}_h$, can be obtained after integrating with respect to $\widehat{F}_h(u;\boldsymbol{\beta})$

$$\widehat{Q}_h(\boldsymbol{\beta}) := \int_{-\infty}^{\infty} \rho_\tau(u)\,\mathrm{d}\widehat{F}_h(u;\boldsymbol{\beta}) = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\rho_\tau(u) K\left(\frac{u + \boldsymbol{x}_i^\top\boldsymbol{\beta} - y_i}{h}\right)\mathrm{d}u. \tag{5}$$

By means of the convolution operator "$*$", such empirical risk can also be written as

$$\widehat{Q}_h(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\ell_{h,\tau}\left(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right) \text{ with } \ell_{h,\tau}(u) = \left(\rho_\tau * K_h\right)(u) = \int_{-\infty}^{\infty}\rho_\tau(v)K_h(v-u)\mathrm{d}v, \tag{6}$$

where $\ell_{h,\tau}$ is called as the SQR loss throughout the paper. Denote the integrated kernel function $\bar{K}:\mathbb{R}\to[0,1]$ as $\bar{K}(u) = \int_{-\infty}^{u} K(t)\,\mathrm{d}t$, and correspondingly $\bar{K}_h(u) = \int_{-\infty}^{u} K_h(t)\,\mathrm{d}t = \bar{K}(u/h)$. Then the SQR empirical risk function $\widehat{Q}_h(\boldsymbol{\beta})$ is twice continuously differentiable with gradient $\nabla\widehat{Q}_h(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{\bar{K}_h(-\mathrm{r}_i(\boldsymbol{\beta})) - \tau\}\boldsymbol{x}_i$ and Hessian matrix $\nabla^2\widehat{Q}_h(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} K_h(-\mathrm{r}_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i^\top$.

In Tan et al. (2022b), the authors investigated the high-dimensional sparse QR estimation problem with SQR loss function. Based on the dataset $\{(y_i,\boldsymbol{x}_i)\}_{i=1}^n$, the ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ is then defined as

$$\widehat{\boldsymbol{\beta}}_h = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}}\{\widehat{Q}_h(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1\} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\frac{1}{n}\sum_{i=1}^{n}\ell_{h,\tau}\left(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right) + \lambda\|\boldsymbol{\beta}\|_1\right\}. \tag{7}$$

As pointed out by Tan et al. (2022b), the bias of $\widehat{\boldsymbol{\beta}}_h$ is composed of two parts: the shrinkage bias induced by ℓ_1 penalty and the smoothing bias caused by SQR loss. After selecting a proper bandwidth, $\widehat{\boldsymbol{\beta}}_h$ reaches the same order convergence rate as the ℓ_1-QR estimator $\widehat{\boldsymbol{\beta}}$. Hence the ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ is capable of substituting $\widehat{\boldsymbol{\beta}}$ as the key role of debiasing.

2.3 Debiasing the ℓ_1-SQR estimator

We debias the ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ along the way of Van de Geer et al. (2014). The main idea is to invert the Karush-Kuhn-Tucker (KKT) conditions of (7). Since $\widehat{\boldsymbol{\beta}}_h$ is the minimizer of (7), then it satisfies the KKT conditions:

$$\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i) - \tau]\boldsymbol{x}_i + \lambda\boldsymbol{g} = \boldsymbol{0}, \tag{8}$$

where $\boldsymbol{g} = (g_1,\ldots,g_p)^\top$ is a subderivative of $\|\cdot\|_1$ at $\widehat{\boldsymbol{\beta}}_h$ satisfying $g_j = \mathrm{sign}(\widehat{\boldsymbol{\beta}}_{h;j})$ if $\widehat{\boldsymbol{\beta}}_{h;j} \neq 0$ and otherwise $g_j \in [-1,1]$. Here $\widehat{\boldsymbol{\beta}}_{h;j}$ denotes the j-th coordinate of $\widehat{\boldsymbol{\beta}}_h$.

For sufficiently large n, $\widehat{\boldsymbol{\beta}}_h$ is close to $\boldsymbol{\beta}^*$, and informally the following arguments hold:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i) - \tau]\boldsymbol{x}_i \\
\approx\ &\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\boldsymbol{\beta}^* - y_i) - \tau]\boldsymbol{x}_i + \mathbb{E}\{[\bar{K}_h(\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}}_h - y) - \tau]\boldsymbol{x}\} - \mathbb{E}\{[\bar{K}_h(\boldsymbol{x}^\top\boldsymbol{\beta}^* - y) - \tau]\boldsymbol{x}\} \\
\approx\ &\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\boldsymbol{\beta}^* - y_i) - \tau]\boldsymbol{x}_i + \nabla^2 Q_h(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*) \\
:=&\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\boldsymbol{\beta}^* - y_i) - \tau]\boldsymbol{x}_i + \mathbf{J}_h(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*),
\end{aligned} \tag{9}
$$

where $Q_h(\boldsymbol{\beta}) = \mathbb{E}[\widehat{Q}_h(\boldsymbol{\beta})]$ and $\mathbf{J}_h = \nabla^2 Q_h(\boldsymbol{\beta}^*)$, heuristically the first "$\approx$" is based on the theory of empirical process, and the second "$\approx$" is deduced from Taylor's expansion. Rearranging (9) leads to

$$
\widehat{\boldsymbol{\beta}}_h \approx \boldsymbol{\beta}^* + \mathbf{J}_h^{-1}\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i) - \tau]\boldsymbol{x}_i - \mathbf{J}_h^{-1}\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\boldsymbol{\beta}^* - y_i) - \tau]\boldsymbol{x}_i.
$$

According to the KKT conditions (8), we know that term $\mathbf{J}_h^{-1}\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i) - \tau]\boldsymbol{x}_i = -\lambda\mathbf{J}_h^{-1}\boldsymbol{g}$ is essentially the shrinkage bias that needs to be removed from the original ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$. On the other hand, He et al. (2023) has shown that $\mathbb{E}\{\mathbf{J}_h^{-1}\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\boldsymbol{\beta}^* - y_i) - \tau]\boldsymbol{x}_i\} = O(h^2)$, which suggests that the term $\mathbf{J}_h^{-1}\frac{1}{n}\sum_{i=1}^{n}[\bar{K}_h(\boldsymbol{x}_i^\top\boldsymbol{\beta}^* - y_i) - \tau]\boldsymbol{x}_i$ is asymptotic negligible whenever h is small. Therefore, we are motivated to define the debiased ℓ_1-SQR estimator as

$$
\widetilde{\boldsymbol{\beta}}_h = \widehat{\boldsymbol{\beta}}_h + \widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^{n}[\tau - \bar{K}_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i)]\boldsymbol{x}_i, \tag{10}
$$

in which $\widehat{\mathbf{W}}$ is an estimator of the matrix $\mathbf{J}_h^{-1}$.

Under the additional condition that the kernel function $K(\cdot)$ is Lipschitz continuous, He et al. (2023) verified that $\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) = \frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top$ is a consistent estimator of $\mathbf{J}_h = \mathbb{E}[K_h(\boldsymbol{x}^\top\boldsymbol{\beta}^* - y)\boldsymbol{x}\boldsymbol{x}^\top]$. As a consequence, we define $\widehat{\mathbf{W}}$ as the approximate inverse of $\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)$ by employing a similar method proposed by Cai et al. (2011). Concretely, $\widehat{\mathbf{W}}$ is the solution to the following optimization problem:

$$
\begin{aligned}
\widehat{\mathbf{W}} = \ &\underset{\mathbf{W}\in\mathbb{R}^{p\times p}}{\operatorname{argmin}} \|\mathbf{W}\|_1 \\
&\text{s.t. } \|\mathbf{W}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}\|_\infty \le \gamma_n,
\end{aligned} \tag{11}
$$

where γ_n is a predetermined tuning parameter. In general, $\widehat{\mathbf{W}}$ is not symmetric since there is no symmetry constraint on $\mathbf{W}$ in (11). To enforce the symmetry, $\widehat{\mathbf{W}} = (\widehat{b}_{i,j})_{1\le i,j\le p}$ only needs to be further operated as follows. Write $\widehat{\mathbf{W}}_1 = (\widehat{b}_{i,j}^1)_{1\le i,j\le p}$, which is defined as

$$
\widehat{b}_{i,j}^1 = \widehat{b}_{j,i}^1 = \widehat{b}_{i,j}\mathbb{I}\{|\widehat{b}_{i,j}| \le |\widehat{b}_{j,i}|\} + \widehat{b}_{j,i}\mathbb{I}\{|\widehat{b}_{i,j}| > |\widehat{b}_{j,i}|\}. \tag{12}
$$

Algorithm 1 Debiased estimator for $\boldsymbol{\beta}^*$ in high-dimensional smoothed quantile regression

Input: Data $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$, quantile index $\tau \in (0,1)$, kernel function $K(\cdot)$, smoothing bandwidth h, tuning parameters λ and γ_n.

1: Let ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ be a solution of the optimization problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^n \ell_{h,\tau}\left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right) + \lambda\|\boldsymbol{\beta}\|_1 \right\}.$$

2: Compute

$$\nabla^2 \widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) = \frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_h - y_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top.$$

3: Let $\widehat{\mathbf{W}}$ be a solution of the convex program:

$$\min_{\mathbf{W}\in\mathbb{R}^{p\times p}} \|\mathbf{W}\|_1 \quad \text{s.t.} \|\mathbf{W}\nabla^2 \widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}\|_\infty \le \gamma_n.$$

4: Define the debiased ℓ_1-SQR estimator as follows:

$$\widetilde{\boldsymbol{\beta}}_h = \widehat{\boldsymbol{\beta}}_h + \widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n [\tau - \bar{K}_h(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_h - y_i)]\boldsymbol{x}_i.$$

Output: The debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$.

Apparently, $\widehat{\mathbf{W}}_1$ is a symmetric matrix. In this work, without loss of generality we assume $\widehat{\mathbf{W}}$ is symmetric and use it in the rest of the paper. Note that there is a fine distinction between (11) and optimization problem (1) in Cai et al. (2011), in which the constraint is imposed on $\|\nabla^2 \widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)\mathbf{W} - \mathbf{I}\|_\infty$. Here we write in this way for convenience.

In Algorithm 1, we summarize the details about how to construct the debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$. In comparison to the previous debiased ℓ_1-QR estimators that originated from $\widehat{\boldsymbol{\beta}}$, our proposed $\widetilde{\boldsymbol{\beta}}_h$ is computationally efficient and easier to obtain. On the one hand, $\widetilde{\boldsymbol{\beta}}_h$ is established upon the ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$, which achieves the same statistical accuracy as $\widehat{\boldsymbol{\beta}}$ while greatly improving computational efficiency due to the smoothness of the SQR loss. Fast and efficient algorithms, like widely-used PGD, ADMM and coordinate descent, can be applied to solve the optimization problem (Tan et al., 2022b). A more comprehensive discussion can be found in Man et al. (2022). In that work, the authors introduced a major variant of the local adaptive majorize-minimization (LAMM) algorithm (Fan et al., 2018) for fitting penalized convolution smoothed quantile regression with many different regularization terms. On the other hand, there is no need to estimate the sparsity function separately in the construction of $\widetilde{\boldsymbol{\beta}}_h$, since it is already included when approximating the inverse of Hessian. Apart from this, $\widetilde{\boldsymbol{\beta}}_h$ works even if the independence between $\boldsymbol{x}$ and ε is violated. In the subsequent analysis, we first establish the non-asymptotic Bahadur representation of $\widetilde{\boldsymbol{\beta}}_h$. Based on this result, we then investigate the high-dimensional inference problems for quantile regression model. For arbitrary linear functional of $\boldsymbol{\beta}^*$, we construct the studentized confidence interval and verify its empirical coverage probability. Testing procedures for both a single variable and a group of variables are explored and provided with theoretical guarantees.

3. Statistical analysis

In this section, we provide a complete and in-depth analysis of the debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$. According to its construction, we first study the properties of $\widehat{\mathbf{W}}$ and explore the upper bound of γ_n. Then we establish the Bahadur representation of $\widetilde{\boldsymbol{\beta}}_h$ from a non-asymptotic viewpoint, which can further be used to validate the asymptotic normality of $\widetilde{\boldsymbol{\beta}}_h$. Based on these theoretical results, we construct the confidence intervals for arbitrary linear projection of $\boldsymbol{\beta}^*$, from which a pointwise inference for β_j^*, $j \in \{1, \ldots, p\}$ and the subgroup hypothesis testing for $\{\beta_j^* : j \in \mathcal{G} \subset \{1, \ldots, p\}\}$ can be directly achieved.

3.1 General CLIME estimator $\widehat{\mathbf{W}}$

Before proceeding, we impose some regularity assumptions at first.

Assumption 1 *$K(\cdot)$ is a kernel function, which is to say $K(\cdot)$ is non-negative, symmetric, and satisfies $\int_{-\infty}^{\infty} K(u)\,\mathrm{d}u = 1$. In this paper, $K(u)$ is assumed to be continuously twice differentiable with uniformly bounded zero-order, first-order and second-order derivatives, i.e., $\kappa_u = \sup_{u\in\mathbb{R}} K(u) < \infty$, $\kappa_l = \min_{u\in[-1,1]} K(u) > 0$, $\kappa_u' = \sup_{u\in\mathbb{R}} |K'(u)| < \infty$, and $\kappa_u'' = \sup_{u\in\mathbb{R}} |K''(u)| < \infty$. Moreover, denote $\kappa_k = \int_{-\infty}^{\infty} |u|^k K(u)\,\mathrm{d}u$ for $k \geq 1$.*

Assumption 2 *There exists some constant $l_0 > 0$ for the conditional density function $f_{\varepsilon|\boldsymbol{x}}(\cdot)$ such that $|f_{\varepsilon|\boldsymbol{x}}(u_1) - f_{\varepsilon|\boldsymbol{x}}(u_2)| \leq l_0 |u_1 - u_2|$ for all $u_1, u_2 \in \mathbb{R}$ almost surely (over $\boldsymbol{x}$). Moreover, suppose there exists two constants $\bar{f} \geq \underline{f} > 0$ such that $f_{\varepsilon|\boldsymbol{x}}(0) \geq \underline{f}$ almost surely for all $\boldsymbol{x}$ and $\sup_{u\in\mathbb{R}} f_{\varepsilon|\boldsymbol{x}}(u) \leq \bar{f}$.*

Assumption 3 *The covariate $\boldsymbol{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ is sub-Gaussian with $\mathbb{E}(\boldsymbol{x}) = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top) \succ \mathbf{0}$. That is, there exists $\sigma > 0$ (without loss of generality assume $\sigma \geq 1$) such that $\mathbb{P}\{|\langle \boldsymbol{u}, \boldsymbol{w} \rangle| \geq \sigma t\} \leq 2e^{-\frac{t^2}{2}}$ for all $\boldsymbol{u} \in \mathbb{S}^{p-1}$ and $t \geq 0$, where $\boldsymbol{w} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}$. Furthermore, the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$ satisfies $0 < \Lambda_{\min} \leq \Lambda_{\min}(\boldsymbol{\Sigma}) \leq 1 \leq \Lambda_{\max}(\boldsymbol{\Sigma}) \leq \Lambda_{\max} < \infty$ and the precision matrix $\boldsymbol{\Sigma}^{-1}$ satisfies $\|\boldsymbol{\Sigma}^{-1}\|_{L_1} \leq M$ for some $\Lambda_{\min}$, $\Lambda_{\max}$ and M. Besides, denote $\alpha_{\boldsymbol{x}}^2 = \max_{1\leq j\leq p} \mathbb{E}(x_j^2)$, $m_k = \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}(|\langle \boldsymbol{u}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}\rangle|^k), k = 1, 2, \ldots,$ in which $m_4 < \infty$, and without loss of generality assume $m_2 = 1$ in this work.*

Assumption 4 *$\boldsymbol{\beta}^*$ is s-sparse, which means that the cardinality of its support set $\mathcal{S} = \{j : \beta_j^* \neq 0\}$ satisfies $|\mathcal{S}| \leq s \ll n$.*

Assumption 1 is the standard condition for kernel function with some additional constraints of its derivatives. The requirement $\kappa_l = \min_{u\in[-1,1]} K(u) > 0$ is for theoretical simplicity and can be converted to $\min_{u\in[-c,c]} K(u) > 0$ for some $c \in (0,1)$ via transformation. Common kernel functions, such as Gaussian kernel, Epanechnikov kernel and logistic kernel, and their rescaled versions satisfy this condition. Assumption 2 imposes regularity conditions on the conditional density function. Such conditions are basic and standard in the research of quantile regression. Assumption 3 is about the covariate $\boldsymbol{x}$, which is supposed to be from a sub-Gaussian random vector family with zero mean and positive definite covariance matrix. Similar to Cai et al. (2011) and Bradic and Kolar (2017), the L_1-norm of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is assumed to be bounded, and this helps us to establish the convergence rate associated with $\widehat{\mathbf{W}}$. Besides, the kurtosis of arbitrary linear projection $\langle \boldsymbol{u}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}\rangle$ is required to be uniformly bounded. Assumption 4 is the hard sparsity condition of $\boldsymbol{\beta}^*$, which can be relaxed to a weaker version that allows some coefficients with small elements (Belloni et al., 2019). Under these assumptions, we could directly use the existing properties of $\widehat{\boldsymbol{\beta}}_h$ to establish theoretical results of $\widetilde{\boldsymbol{\beta}}_h$.

For completeness, we restate the non-asymptotic error bounds of ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ verified in Tan et al. (2022b) with a slightly different condition on $\boldsymbol{x}$.

Theorem 1 (Theorem 4.1 in Tan et al. (2022b)) *Suppose Assumptions 1-4 hold, and the bandwidth h falls into*

$$\max\left(\frac{\alpha_{\boldsymbol{x}}}{\underline{f}}\sqrt{\frac{s\log p}{n}}, \frac{\alpha_{\boldsymbol{x}}^2 \bar{f}}{\underline{f}^2}\frac{s\log p}{n}\right) \lesssim h \leq \min\left\{\underline{f}/(2l_0), (s^{1/2}\lambda)^{1/2}\right\}.$$

Then, the ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ with $\lambda \asymp \sigma\alpha_{\boldsymbol{x}}\sqrt{\tau(1-\tau)\log p/n}$ satisfies the bounds

$$\|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*\|_2 \leq C_1 \underline{f}^{-1} s^{1/2}\lambda \quad \textit{and} \quad \|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*\|_1 \leq C_2 \underline{f}^{-1} s\lambda \tag{13}$$

with probability at least $1-\frac{1}{p}$, where the constants $C_1, C_2 > 0$ depend only on $(l_0, \sigma, \Lambda_{\min}, \kappa_l, \kappa_2)$.

From Theorem 1 we find that ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ reaches the near-minimax rate of convergence whenever h is selected in the specified interval. On the basis of this result, we are devoted to studying the optimization problem (11) and quantifying γ_n. To achieve this target, we need to introduce another assumption on the inverse of the population Hessian $\mathbf{J}_h = \nabla^2 Q_h(\boldsymbol{\beta}^*)$.

Assumption 5 *There exists some $M' > 0$ such that $\|\mathbf{J}_h^{-1}\|_{L_1} \leq M'$. Moreover, $\mathbf{J}_h^{-1} := (\widetilde{\boldsymbol{b}}_1, \ldots, \widetilde{\boldsymbol{b}}_p)^\top = (\widetilde{b}_{i,j})_{1\leq i,j\leq p}$ is sparse row-wise, i.e., $\max_{1\leq i\leq p}\sum_{j=1}^p |\widetilde{b}_{i,j}|^q \leq c_{n,p}$ for $0 \leq q < 1$, where $c_{n,p}$ is positive and bounded away from 0 and allowed to increase as n and p grow.*

This assumption requires $\mathbf{J}_h^{-1}$ to be sparse both in the sense of L_1-norm and matrix row space. Similar conditions were considered in the literature on precision matrix estimation and more general inverse Hessian matrix estimation, one can turn to Van de Geer et al. (2014), Belloni et al. (2016), Cai et al. (2016) and Ning and Liu (2017) for more information. In fact, the assumption holds pointwise for h. This is because our proposed debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$ is constructed on the basis of ℓ_1-SQR estimator $\widehat{\boldsymbol{\beta}}_h$ with some certain bandwidth h. As He et al. (2023) and Tan et al. (2022b) mentioned, convolution-type smoothing method is not sensitive to h. When h satisfies certain order conditions, the errors of ℓ_1-SQR estimators with different h are almost the same. Hence, we actually select one specific value of h in constructing debiased ℓ_1-SQR estimator and then building up theories thereafter. Back to Assumption 5, here we essentially impose the row-wise sparse condition for the matrix $\mathbf{J}_h^{-1}$ with some certain h. Which is to say, the statement holds for specific $h = h(n, p)$. Once (n, p) is given, h is determined. This also fits our non-asymptotic framework.

Next, we turn to consider the accuracy bound of $\widehat{\mathbf{W}}$ as an approximation of the inverse of $\nabla^2 \widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)$. Since $\widehat{\mathbf{W}}$ can be treated as a general case of the CLIME estimator that Cai et al. (2011) proposed for estimating precision matrix, here we extend their method to investigate the properties of $\widehat{\mathbf{W}}$ and summarize these results in Theorem 2.

Theorem 2 *Suppose Assumptions 1-5 hold and the bandwidth h satisfies conditions in Theorem 1, with probability at least $1 - \frac{4}{p}$, we have the following results:*

$$\|\widehat{\mathbf{W}}\|_{L_1} \leq \|\mathbf{J}_h^{-1}\|_{L_1}, \quad \|\widehat{\mathbf{W}}\nabla^2 \widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}\|_\infty \lesssim \gamma_n \quad \textit{and} \quad \|\widehat{\mathbf{W}}\nabla^2 Q_h(\boldsymbol{\beta}^*) - \mathbf{I}\|_\infty \lesssim \gamma_n, \tag{14}$$

where $\gamma_n = \sqrt{\frac{\log p}{nh}} + \frac{\log p}{nh} + \frac{s^2 \log p (\log(p \vee n))^2}{nh^3} + s\big(\sqrt{\frac{\log p \log(p\vee n)}{nh^2}} + \sqrt{\frac{(\log p)^2 \log(p \vee n)}{n^2 h^3}} + \sqrt{\frac{(\log p)^3 \log(p\vee n)}{n^3 h^4}}\big).$

As Cai et al. (2011) mentioned that, the convex optimization (11) can be decomposed into p vector minimization problems. Denote $\boldsymbol{e}_j$ as the unit vector in $\mathbb{R}^p$ with its j-th coordinate being 1 and others being 0. For all $1 \leq j \leq p$, it has been proved that solving (11) is equivalent to solving the following p optimization problems:

$$\begin{aligned} \widehat{\boldsymbol{b}}_j = \ & \underset{\boldsymbol{b} \in \mathbb{R}^p}{\operatorname{argmin}} \|\boldsymbol{b}\|_1 \\ & \text{s.t. } \|\boldsymbol{b}^\top \nabla^2 \widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \boldsymbol{e}_j^\top\|_\infty \leq \gamma_n, \end{aligned} \tag{15}$$

and $\widehat{\mathbf{W}} = (\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_p)^\top = (\widehat{b}_{i,j})_{1 \leq i,j \leq p}$. Theorem 2 reveals that $\mathbf{J}_h^{-1} = (\widetilde{b}_{i,j})_{1\leq i,j\leq p}$ is feasible for (11), which means $\widetilde{\boldsymbol{b}}_j$ is feasible for (15). Due to the optimality of $\widehat{\boldsymbol{b}}_j$, we know that $\|\widehat{\boldsymbol{b}}_j\|_1 \leq \|\widetilde{\boldsymbol{b}}_j\|_1$ holds for $1 \leq j \leq p$. Hence, motivated by Lemma 7.1 of Cai et al. (2016), the error bound of $\|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_{L_1}$ can be obtained and will be applied to the further analysis.

At the end of this part, we provide the upper bound of $\|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_{L_1}$, which is crucial for establishing theoretical results afterwards.

Theorem 3 *Suppose Assumptions 1-5 hold and the bandwidth h satisfies conditions in Theorem 1, then we have*

$$\|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_{L_1} \leq 8c_{n,p}(\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1})^{1-q} \asymp \gamma_n^{1-q}. \tag{16}$$

3.2 Bahadur representation

In this part, we investigate the Bahadur representation of $\widetilde{\boldsymbol{\beta}}_h$. Different from the setup in He et al. (2023), the error rate of the Bahadur remainder term in this work is provided by allowing both p and s to increase with n. Theorems 1, 2 and 3 pave way for the establishment of this result.

Denote $\widehat{\boldsymbol{\delta}}_h = \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*$ and $\widehat{\varepsilon}_i = \mathrm{r}_i(\widehat{\boldsymbol{\beta}}_h) = y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_h$, and consider ℓ_1-ball $\mathbb{B}_1(r) = \{\boldsymbol{a} \in \mathbb{R}^p : \|\boldsymbol{a}\|_1 \le r\}$. For arbitrary $\boldsymbol{\alpha} \in \mathbb{B}_1(r)$, it is easy to see that

$$
\begin{aligned}
\sqrt{n}\boldsymbol{\alpha}^\top(\widetilde{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*) &= \sqrt{n}\boldsymbol{\alpha}^\top(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*) + \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n[\tau - \bar{K}_h(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_h - y_i)]\boldsymbol{x}_i \\
&= \sqrt{n}\boldsymbol{\alpha}^\top\widehat{\boldsymbol{\delta}}_h + \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i + \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n[\bar{K}_h(-\varepsilon_i) - \bar{K}_h(-\widehat{\varepsilon}_i)]\boldsymbol{x}_i \\
&= \sqrt{n}\boldsymbol{\alpha}^\top\widehat{\boldsymbol{\delta}}_h + \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i - \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n K_h(\zeta_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top\widehat{\boldsymbol{\delta}}_h \\
&= \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i + \boldsymbol{\alpha}^\top(\widehat{\mathbf{W}} - \mathbf{J}_h^{-1})\frac{1}{\sqrt{n}}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i \\
&\quad - \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n[K_h(\zeta_i) - K_h(-\widehat{\varepsilon}_i)]\boldsymbol{x}_i\boldsymbol{x}_i^\top\widehat{\boldsymbol{\delta}}_h - \sqrt{n}\boldsymbol{\alpha}^\top(\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I})\widehat{\boldsymbol{\delta}}_h \\
&:= \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i + (\Gamma_1 + \Gamma_2 + \Gamma_3),
\end{aligned} \tag{17}
$$

where ζ_i is the intermediate value from Taylor's expansion between $-\varepsilon_i$ and $-\widehat{\varepsilon}_i$, $\Gamma_1 = \boldsymbol{\alpha}^\top(\widehat{\mathbf{W}} - \mathbf{J}_h^{-1})\frac{1}{\sqrt{n}}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i$, $\Gamma_2 = \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\sum_{i=1}^n[K_h(-\widehat{\varepsilon}_i) - K_h(\zeta_i)]\boldsymbol{x}_i\boldsymbol{x}_i^\top\widehat{\boldsymbol{\delta}}_h$ and $\Gamma_3 = \sqrt{n}\boldsymbol{\alpha}^\top(\mathbf{I} - \widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h))\widehat{\boldsymbol{\delta}}_h$. Based on the error bounds of $\widetilde{\boldsymbol{\beta}}_h$ and $\widehat{\mathbf{W}}$, the residual terms Γ_1, Γ_2, and Γ_3 can be well controlled. The following theorem provides the non-asymptotic Bahadur representation for the linear projection of our proposed debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$. In the following, all of the results are established under the case of $q = 0$, which is one of the most common matrix sparsity assumptions that many works considered. Similar conditions can also be found in Liu and Wang (2017), Cheng et al. (2022), Li et al. (2022) and Tran and Yu (2022) on different topics.

Theorem 4 *Suppose Assumptions 1-5 hold, the bandwidth h is required to meet the condition $(\frac{s\log p}{n})^{1/2} \lesssim h \lesssim (\frac{s\log p}{n})^{1/4}$, and the sparsity s satisfies $s \lesssim (\log(p \vee n))^{1/2}$. Then for any $\boldsymbol{\alpha} \in \mathbb{B}_1(r)$, the debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$ satisfies*

$$
|\sqrt{n}\boldsymbol{\alpha}^\top(\widetilde{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*) - \frac{1}{\sqrt{n}}\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\sum_{i=1}^n[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i| \lesssim \frac{rc_{n,p}s^2(\log p)^{3/2}(\log(p \vee n))^{5/2}}{nh^3} \tag{18}
$$

with probability at least $1 - \frac{5}{p}$.

Remark 1 A necessary condition for the Bahadur residual term to be of order $o(1)$ is that $nh^3 \to \infty$ as $n \to \infty$. This suggests a less rigorous bound for the bandwidth h with respect to n, i.e., $n^{-1/3} \lesssim h \lesssim n^{-1/4}$.

Note that the main term of the Bahadur representation in (18) is the same as that of the SQR estimator in He et al. (2023). In fact, He et al. (2023) investigated the SQR model under the "increasing dimension" regime, in which p is allowed to grow up with n while $p < n$. In their work, the researchers did not impose the sparse condition on $\boldsymbol{\beta}^*$. All the theoretical results, such as the Bahadur representation, therein were established based on the SQR estimators without using penalty. While in this work, we focus on the high-dimensional setting, where p can be much larger than n. The ℓ_1-SQR estimator proposed by Tan et al. (2022b) is available for variable selection in this setting, but not directly for inference. The ℓ_1-penalty makes it sparse while introducing a shrinkage bias. On the basis of this penalized estimator, we propose the debiased ℓ_1-SQR estimator and build up a series of theories for it. In other words, the Bahadur representation and other results established in our work extend the corresponding results of the SQR estimator to the high-dimensional sparse scenarios, where the SQR estimator fails to work. Our proposed debiased ℓ_1-SQR estimator is designed for statistical inference in the high-dimensional setup. It provides valid inference results for arbitrary parameter, regardless of its signal strength.

3.3 Confidence intervals

In this section, we construct the confidence intervals and state the inference results for high-dimensional quantile regression based on our proposed debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$. For significance level $\varrho \in (0,1)$, let $z_{1-\varrho}$ be the $1-\varrho$ standard normal percentile point. We focus on the confidence intervals for $\boldsymbol{\alpha}^\top\boldsymbol{\beta}^*$, from which we can make inference for both a single variable and a subgroup of variables.

According to Theorem 4, an asymptotic $(1-\varrho)100\%$ confidence interval for $\boldsymbol{\alpha}^\top\boldsymbol{\beta}^*$ is given by

$$\widehat{\Delta}_n = \big[\boldsymbol{\alpha}^\top\widetilde{\boldsymbol{\beta}}_h - \frac{z_{1-\frac{\varrho}{2}}}{\sqrt{n}}\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}, \boldsymbol{\alpha}^\top\widetilde{\boldsymbol{\beta}}_h + \frac{z_{1-\frac{\varrho}{2}}}{\sqrt{n}}\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}\big], \quad (19)$$

where $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix previously defined. The next theorem provides the theoretical coverage probability of $\widehat{\Delta}_n$, which is essentially the Berry-Esseen bound for $\boldsymbol{\alpha}^\top\widetilde{\boldsymbol{\beta}}_h$.

Theorem 5 *Suppose Assumptions 1-5 hold, the bandwidth h is required to meet the condition $(\frac{s\log p}{n})^{1/2} \lesssim h \lesssim (\frac{s\log p}{n})^{1/4}$, and the sparsity s satisfies $s \lesssim (\log(p\vee n))^{1/2}$. Then for all $\boldsymbol{\alpha} \in \mathbb{B}_1(r)$, we have*

$$\begin{aligned}&\sup_{\boldsymbol{\beta}^*:\|\boldsymbol{\beta}^*\|_0\le s}\ \sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)} \mathbb{P}\big(\boldsymbol{\alpha}^\top\boldsymbol{\beta}^* \in \widehat{\Delta}_n\big)\\ &= 1-\varrho + O\Big(\frac{rc_{n,p}s^2\log p(\log(p\vee n))^2(r+\sqrt{\log p\log(p\vee n)})}{nh^3} + rh^2\sqrt{n\log(p\vee n)}\Big). \quad (20)\end{aligned}$$

Theorem 5 indicates that for appropriate choice of bandwidth $h = h(n,p) \to 0$, the coverage probabilities of $\widehat{\Delta}_n$ are close to $1-\varrho$ as $n, p \to \infty$ subject to some conditions. Whenever $c_{n,p}$ and r are considered fixed, the optimal tradeoff between the two error terms in (20) implicates the selection $h \asymp \frac{s^{2/5}(\log p)^{3/10}(\log(p\vee n))^{2/5}}{n^{3/10}}$. After taking h to be selected from the best tradeoff, the error of coverage probability is of order $\frac{s^{4/5}(\log p)^{3/5}(\log(p\vee n))^{13/10}}{n^{1/10}}$, which further implicates that the sparsity s and ambient dimension p obey the growth condition $s^8(\log p)^6(\log(p \vee n))^{13} \lesssim n$. Specially, for sparse loading vector $\boldsymbol{\alpha}$, (20) sheds light on the construction of the confidence interval for a single variable.

Remark 2 The confidence interval for one coordinate β_j^* is obtained by setting $\boldsymbol{\alpha} = \boldsymbol{e}_j$. For all $u \in \mathbb{R}$ and any $j \in \{1, \ldots, p\}$, we have

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}(\widetilde{\beta}_{h;j} - \beta_j^*)}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}} \le u \right) - \Phi(u) \right| = o_{\mathbb{P}}(1), \tag{21}$$

where $\widetilde{\beta}_{h;j}$ is the j-th element of $\widetilde{\boldsymbol{\beta}}_h$, $[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}$ is the (j,j)-th entry of matrix $\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}$ and $\Phi(\cdot)$ denotes the cumulative distribution function of standard normal random variable $N(0,1)$.

Remark 3 Whenever the sparsity s does not increase with n and p, our theoretical results are valid for the high-dimensional setup where $\log p = O(n^c)$ for some $c \in (0,1)$.

3.4 Hypothesis testing

Based on the result about confidence intervals for the single variable in (21), now we turn to considering the hypothesis testing problems $H_{0,j} : \beta_j^* = 0$ versus $H_{1,j} : \beta_j^* \neq 0$ for any $j \in \{1, \ldots, p\}$ and assign p-values for these tests. The p-value P_j for the null hypothesis $H_{0,j}$ is defined as

$$P_j = 2\left(1 - \Phi\left(\frac{\sqrt{n}|\widetilde{\beta}_{h;j}|}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}} \right)\right). \tag{22}$$

According to the procedure of hypothesis testing, we introduce the following decision rule $\widehat{T}_j$ based on P_j:

$$\widehat{T}_j = \begin{cases} 1 & \text{if } P_j \le \varrho \quad (\text{ reject } H_{0,j}), \\ 0 & \text{otherwise} \quad (\text{accept } H_{0,j}), \end{cases} \tag{23}$$

in which ϱ is the predetermined Type I error rate. In general, the quality of the test $\widehat{T}_j$ is measured by its significance level ϱ_j (probability of Type I error) and statistical power $1 - \pi_j$ (π_j is the probability of Type II error).

For $\beta_j^* \neq 0$ with sufficiently small absolute value, the null hypothesis $H_{0,j}$ is indistinguishable from $H_{1,j}$. Hence, in this work we enforce that $|\beta_j^*| > \mu$ whenever $\beta_j^* \neq 0$. Analogous to Javanmard and Montanari (2014a) and Javanmard and Montanari (2014b), we adopt the minimax viewpoint and control the performance of the test over all s-sparse vectors. Specifically, given a family of decision rules $T_j = T_{j,\mathbf{X}}(\boldsymbol{y}) : \mathbb{R}^n \to \{0,1\}$ with $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{y} \in \mathbb{R}^n$, for $\mu > 0$, we define

$$\varrho_j(T) = \sup_{\boldsymbol{\beta}^*} \left\{ \mathbb{P}_{\boldsymbol{\beta}^*}(T_j = 1) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \le s, \beta_j^* = 0 \right\} \tag{24}$$

and

$$\pi_j(T;\mu) = \sup_{\boldsymbol{\beta}^*} \left\{ \mathbb{P}_{\boldsymbol{\beta}^*}(T_j = 0) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \le s, |\beta_j^*| \ge \mu \right\}, \tag{25}$$

where $\mathbb{P}_{\boldsymbol{\beta}^*}$ is the probability measure induced by $(\mathbf{X}, \boldsymbol{y})$ with respect to the fixed parameter $\boldsymbol{\beta}^*$. Actually speaking, for any s-sparse parameter with $\beta_j^* = 0$, the probability of false rejection of the null is upper bounded by $\varrho_j(T)$. Moreover, if $\boldsymbol{\beta}^*$ is s-sparse with $|\beta_j^*| > \mu$, then $\pi_j(T;\mu)$ is the upper bound of the rate of erroneous detection. Next, we establish the bounds on $\varrho_j(\widehat{T}_j)$ and $\pi_j(\widehat{T}_j;\mu)$ and conclude these results in the following theorem.

Theorem 6 *Suppose Assumptions 1-5 hold, the bandwidth h is required to meet the condition $(\frac{s\log p}{n})^{1/2} \lesssim h \lesssim (\frac{s\log p}{n})^{1/4}$, and the sparsity s satisfies $s \lesssim (\log(p \vee n))^{1/2}$. Consider a sequence of design matrices $\mathbf{X} \in \mathbb{R}^{n\times p}$. For any $j \in \{1,\dots,p\}$ and $\varrho \in [0,1]$, we have the following bounds on the statistical significance and power for testing $H_{0,j} : \beta_j^* = 0$ against the alternative $|\beta_j^*| \geq \mu$:*

$$\lim_{n\to\infty} \varrho_j(\widehat{T}_j) \leq \varrho, \tag{26}$$

and

$$\liminf_{n\to\infty} \frac{1-\pi_j(\widehat{T}_j;\mu)}{1-\pi_j(\mu)} \geq 1, \qquad 1-\pi_j(\mu) = G\bigg(\varrho, \frac{\sqrt{n\mathbb{E}[K_h(\varepsilon)]}\mu}{\sqrt{\tau(1-\tau)}[\mathbf{J}_h^{-1}]_{j,j}^{1/2}}\bigg), \tag{27}$$

where, for $\varrho \in [0,1]$ and $u \in (0,\infty)$, the function $G(\varrho,u)$ is defined as follows:

$$G(\varrho,u) = 2 - \Phi(z_{1-\frac{\varrho}{2}} + u) - \Phi(z_{1-\frac{\varrho}{2}} - u).$$

Theorem 6 verifies that Type I error $\varrho_j(\widehat{T}_j)$ is uniformly upper bounded by the given significance level ϱ, and the statistical power $1-\pi_j(\widehat{T}_j;\mu)$ is at least $1-\pi_j(\mu)$. As mentioned in Javanmard and Montanari (2014a), the function $G(\varrho,u)$ is continuous and monotonically increasing with respect to u for fixed ϱ. It can be checked that $G(\varrho,0) = \varrho$, which is the trivial case obtained by randomly rejecting $H_{0,j}$ with probability ϱ. Given a target statistical power $1-\pi \in (\varrho,1)$, μ is required to satisfy $\mu \geq \frac{c_\pi[\mathbf{J}_h^{-1}]_{j,j}^{1/2}}{\sqrt{n}}$ for some constant c_π associated with π. Moreover, the larger μ is, the higher the power $1-\pi_j(\widehat{T}_j;\mu)$ achieves.

3.5 Simultaneous testing

In this part, we extend the method to test a group of variables $\{\beta_j^* : j \in \mathcal{G} \subset \{1,\dots,p\}\}$. Consider the simplest case that the cardinality of $\mathcal{G}$ is fixed (or increases with a slow rate) as $n,p \to \infty$. To test the group of hypotheses $\{H_{0,j} : \beta_j^* = 0\}_{j\in\mathcal{G}}$, here we aim at controlling the following familywise error rate (FWER)

$$\mathrm{FWER}(T,n) = \sup_{\boldsymbol{\beta}^*\in\mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \leq s} \mathbb{P}\big\{\exists j \in \mathcal{G} : \beta_j^* = 0, T_j = 1\big\}, \tag{28}$$

where $T = \{T_j\}_{j\in\mathcal{G}}$ is denoted as the family of tests.

After applying the Bonferroni procedure, we introduce the decision rule $\widehat{T}^{\mathcal{G}} = \{\widehat{T}_j^{\mathcal{G}}\}_{j\in\mathcal{G}}$ as

$$\widehat{T}_j^{\mathcal{G}} = \begin{cases} 1 & \text{if } P_j \leq \varrho/|\mathcal{G}| \quad (\text{ reject } H_{0,j})\,, \\ 0 & \text{otherwise} \qquad\;\; (\text{accept } H_{0,j})\,, \end{cases} \tag{29}$$

where P_j is given as per (22). Then we obtain the familywise error control and conclude such result in the following theorem.

Theorem 7 *Suppose Assumptions 1-5 hold, the bandwidth h is required to meet the condition $(\frac{s\log p}{n})^{1/2} \lesssim h \lesssim (\frac{s\log p}{n})^{1/4}$, and the sparsity s satisfies $s \lesssim (\log(p \vee n))^{1/2}$. Consider a sequence of design matrices $\mathbf{X} \in \mathbb{R}^{n\times p}$. For testing the group of hypotheses $\{H_{0,j} : \beta_j^* = 0\}_{j\in\mathcal{G}}$ with fixed cardinality $|\mathcal{G}|$, the familywise error rate of the test $\widehat{T}^{\mathcal{G}}$ is upper bounded as*

$$\limsup_{n\to\infty} \mathrm{FWER}(\widehat{T}^{\mathcal{G}}, n) \le \varrho. \tag{30}$$

4. Simulation experiments

In this section, we provide numerical simulations to illustrate finite sample properties of our proposed debiased ℓ_1-SQR estimator $\widetilde{\boldsymbol{\beta}}_h$. Under several different settings, we present the distribution of our test statistic under null hypothesis and the empirical coverage rates of the confidence intervals in (19). Moreover, we also display the empirical distribution function of the p-value (22) under the null, which is very similar to the uniform distribution, and the power function of the test as true variable β_j^* varies.

The data are generated from a linear regression model

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^* + \varepsilon_i, i = 1, \dots, n,$$

where the covariates $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,p})^\top$ are sampled from a multivariate normal distribution $N(\mathbf{0}, \mathbf{\Sigma})$ with Toeplitz covariance matrix $(\mathbf{\Sigma})_{j,k} = \rho^{|j-k|}$ for all $1 \le j, k \le p$. The support set of $\boldsymbol{\beta}^*$ is $\mathrm{supp}(\boldsymbol{\beta}^*) = \{1, 2, \dots, 10\}$ and the non-zero coefficients are $\beta_j^* = 1-(j-1)/18$ for $j \in \mathrm{supp}(\boldsymbol{\beta}^*)$. As for the random noise ε_i, it follows one of the following three distributions: (1) standard normal distribution $N(0, 1)$; (2) t-distribution with 1.5 degrees of freedom $t(1.5)$; (3) standard Cauchy distribution $Cauchy(0, 1)$. In our simulation experiments, we consider two settings: moderate one with $(n, p) = (500, 500)$ and high-dimensional one with $(n, p) = (500, 1000)$. The Gaussian kernel is applied with the bandwidth parameter h being $\max\{\sqrt{\tau(1-\tau)}(\log p)^{1/2}/n^{3/10}, 0.05\}$. As mentioned in Tan et al. (2022b), the estimated results are insensitive to the choice of bandwidth as long as it is in a proper range. Besides, tunning parameters γ_n and λ are selected by cross-validation. In this work, the quantile level τ is either 0.4 or 0.7, and the Toeplitz matrix coefficient ρ is fixed at 0.1 and 0.5. Empirical coverage rates of constructed confidence intervals are reported over 1000 independent replications.

Let $\boldsymbol{z} = (z_1, \dots, z_p)^\top$ denote the vector with $z_j = \frac{\sqrt{n}(\widetilde{\beta}_{h;j} - \beta_j^*)}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}$. In Figures 1, 2 and 3, we present the distribution of z_j ($j \in \{1, 10, 20\}$) under null hypothesis under various settings. From these histograms we find that z_j is very close to the standard normal random variable regardless of the type of noise, even if the stochastic error is from an extremely heavy tailed family. Our test statistics perform uniformly well for both large and small signal strengths. Additionally, Q-Q plots in Figures 4, 5 and 6 show the relationship between the sample quantiles of z_j and the quantiles of the standard normal distribution for one realization. The scattered points are close to the line y = x, which also confirms the asymptotic normality of $\boldsymbol{z}$.

Table 1 summarizes the empirical coverage rates of the 95% confidence intervals induced by $\widetilde{\boldsymbol{\beta}}_h$ under various settings. Overall, the coverage rates match relatively well to the preassigned level, as expected from our theoretical results. For most setups, our proposed

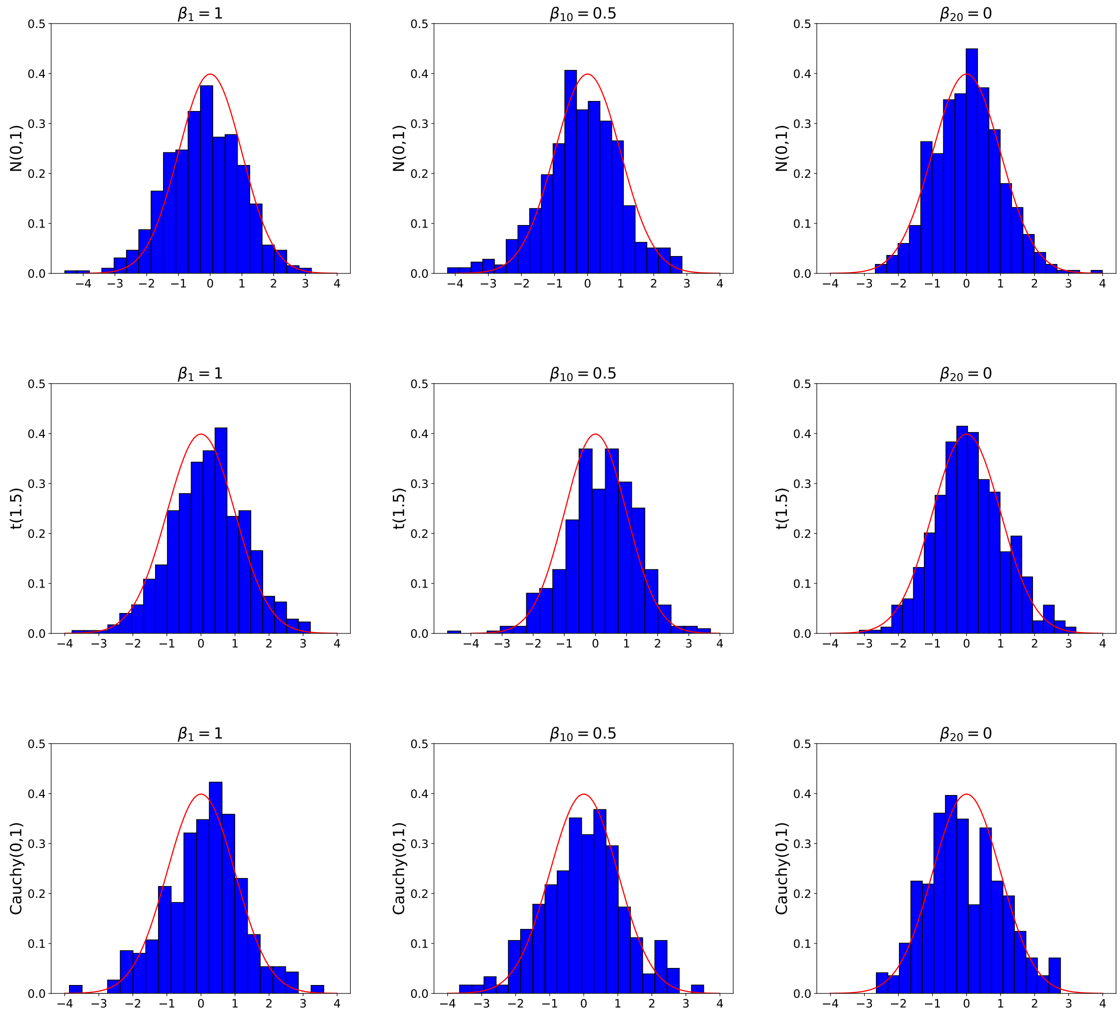

Figure 1: Histograms of $(\tau(1-\tau)[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j})^{-1/2} \cdot \sqrt{n}(\widetilde{\beta}_{h;j} - \beta_j^*)$ under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p,\tau,\rho) = (500, 500, 0.7, 0.1)$. The red curve depicts the probability density function of standard normal random variable.

debiased ℓ_1-SQR estimator creates confidence intervals with approximately 95% coverage. The parameters of medium signal strength (β_{10} in table) can be covered by such confidence intervals with somewhat lower probability around 90% in some scenarios, which is consistent with the descriptions in many previous articles about variable selection. Moreover, it can be found that the empirical coverage rates for $(n,p) = (500, 1000)$ are larger than those for $(n,p) = (500, 500)$. This manifests that our proposed testing procedure has great potential in high-dimensional settings.

Figures 7, 8 and 9 exhibit the empirical cumulative distribution functions (CDF) of the p-values of z_j restricted to the variables outside the support. We observe that the p-values

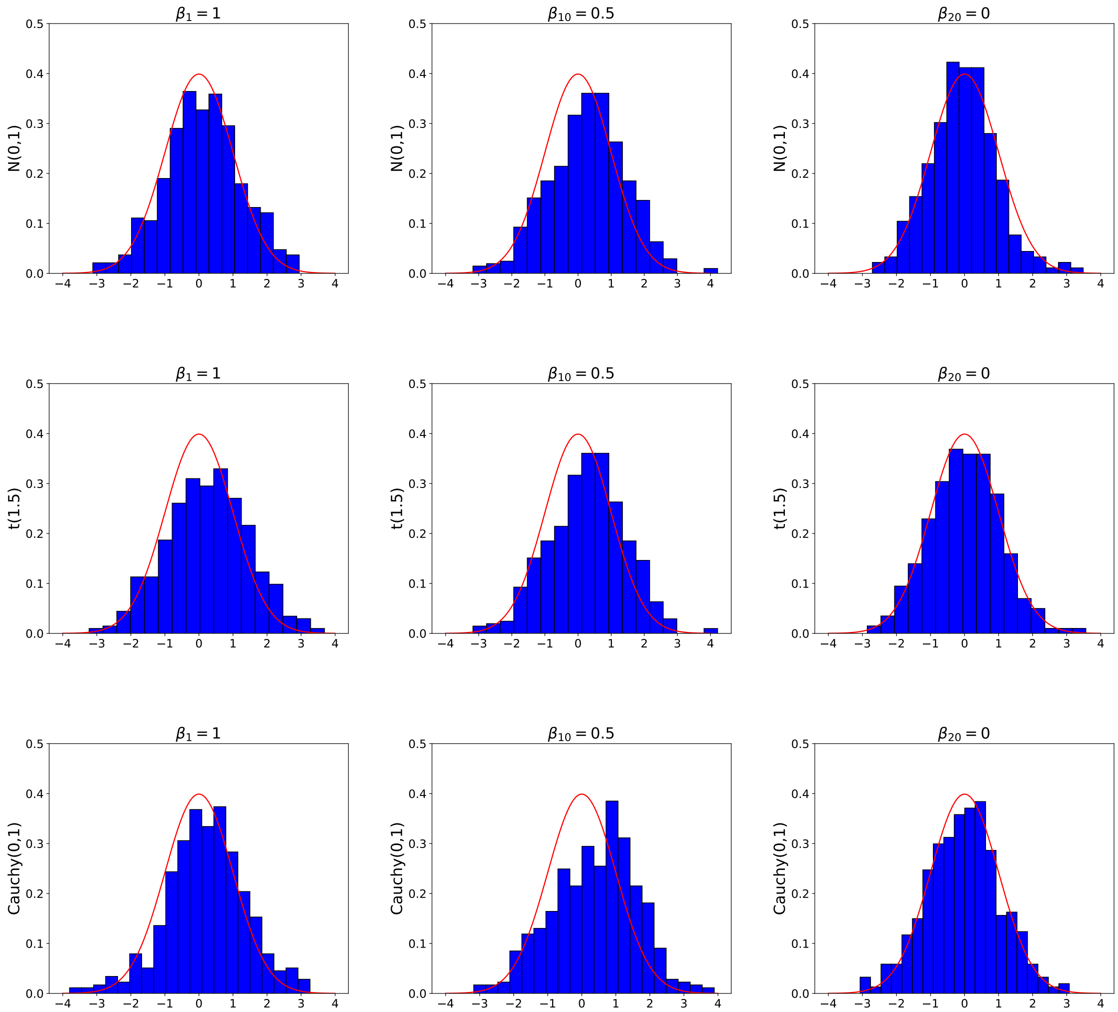

Figure 2: Histograms of $(\tau(1-\tau)[\widehat{\mathbf{W}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{W}}]_{j,j})^{-1/2} \cdot \sqrt{n}(\widetilde{\beta}_{h;j} - \beta_j^*)$ under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p,\tau,\rho) = (500, 500, 0.7, 0.5)$. The red curve depicts the probability density function of standard normal random variable.

for these entries are uniformly distributed as theoretically suggested. Compared to $N(0,1)$ noise setup, the empirical CDF of p-value is disturbed by the heavy-tailed noises to some extent.

Furthermore, we present a power curve for testing the null hypothesis $H_0 : \beta_1 = 0$. For different quantile level τ and Toeplitz matrix coefficient ρ, the associated results are included in Figure 10. As can be seen from the figure, the test reaches power in a small range of origin for all of the settings. Compared to the curve under $N(0,1)$ noise, the power curves under heavy-tailed noises are relatively backward.

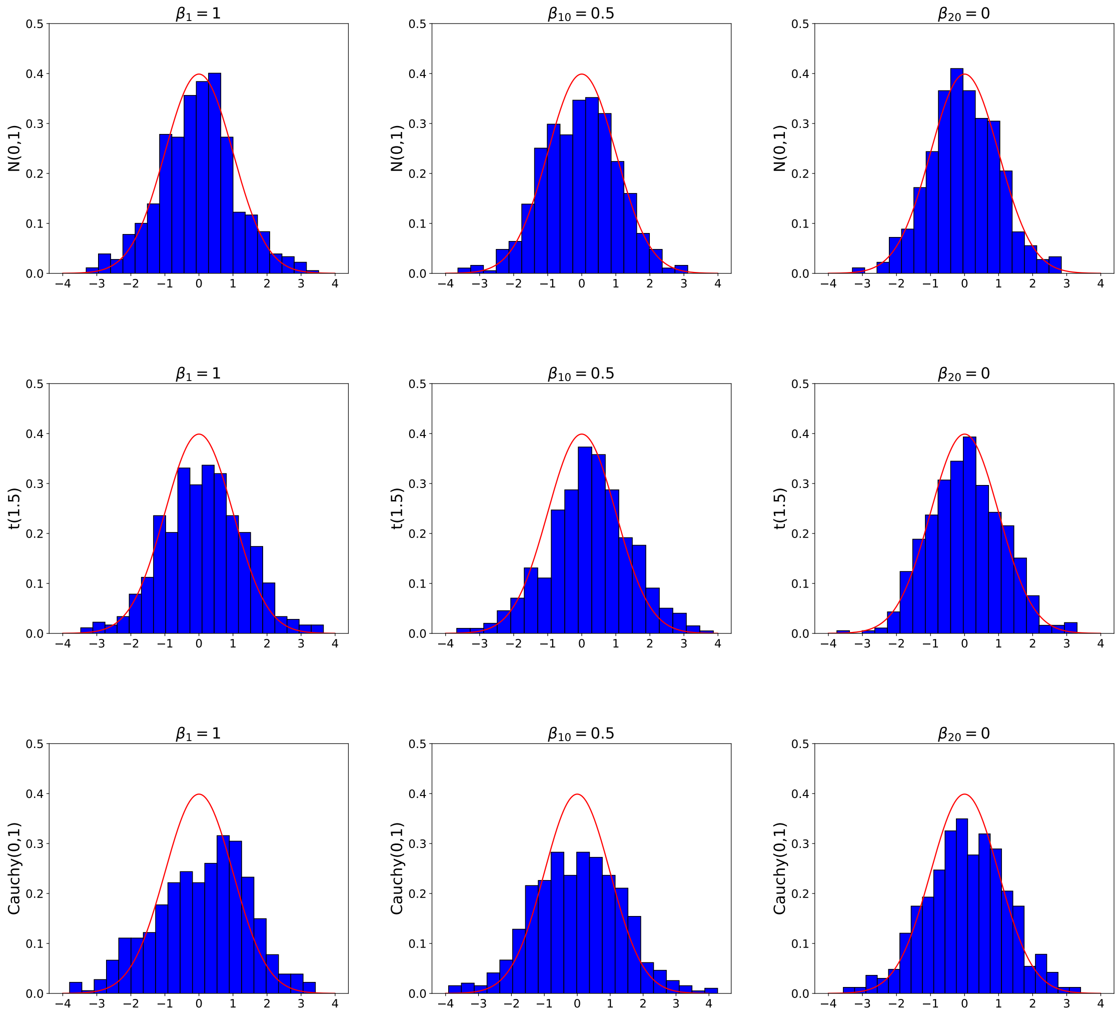

Figure 3: Histograms of $(\tau(1-\tau)[\widehat{\mathbf{W}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{W}}]_{j,j})^{-1/2} \cdot \sqrt{n}(\widetilde{\beta}_{h;j} - \beta_j^*)$ under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n, p, \tau, \rho) = (500, 1000, 0.7, 0.1)$. The red curve depicts the probability density function of standard normal random variable.

5. Real data

For an example of real data, we consider the famous genomic dataset about riboflavin (vitamin B_2) production rate, which is first made publicly available by Bühlmann et al. (2014). This dataset contains $n = 71$ samples and $p = 4088$ covariates that measure the logarithm of the expression level of 4088 genes. For each sample, there is a single real-valued response variable that represents the logarithm of the riboflavin production rate.

Many previous works modeled the riboflavin production rate as a linear regression and applied Lasso method to select important genes. Both of Bühlmann et al. (2014) and Javan-

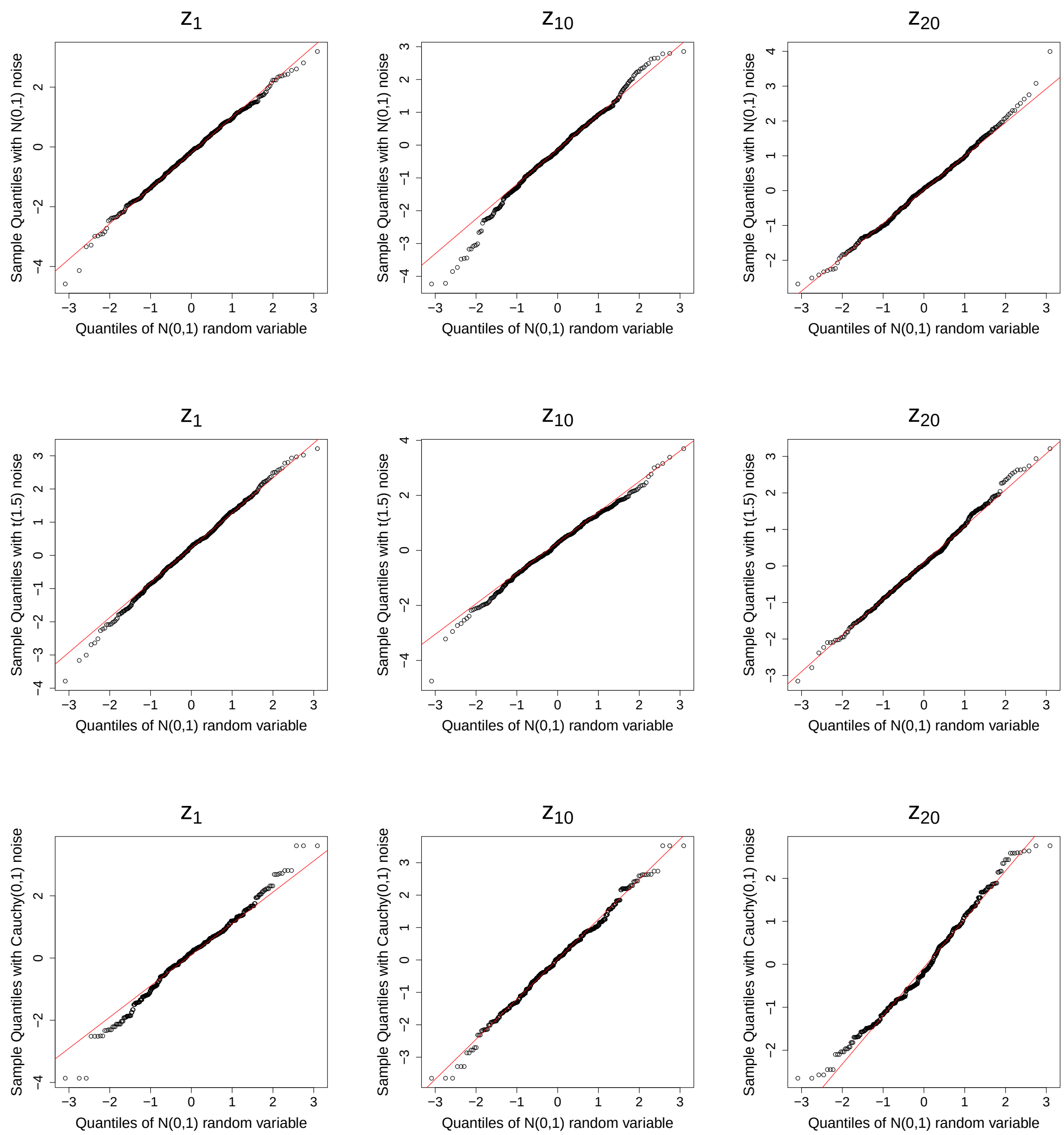

Figure 4: Q-Q plots of $\boldsymbol{z}$ for one realization under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p,\tau,\rho) = (500, 500, 0.7, 0.1)$.

mard and Montanari (2014a) picked out 30 genes corresponding to the nonzero parameters of the Lasso estimators. Then different methods have been implemented to further refine the search for crucial genes after variable selection. Among these articles, researchers only found two significant genes YXLD-at and YXLE-at to the best of our knowledge.

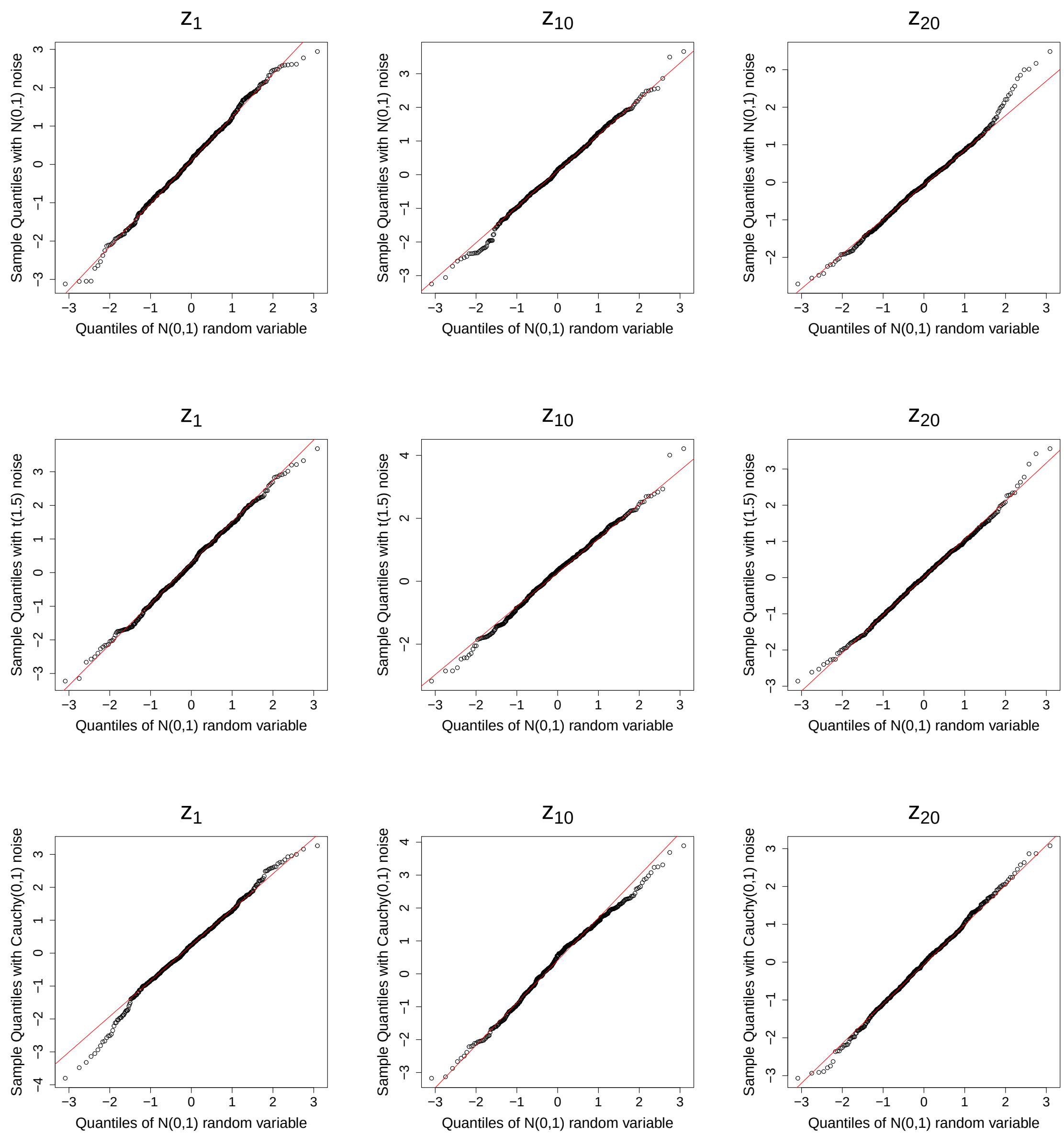

Figure 5: Q-Q plots of $\boldsymbol{z}$ for one realization under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p,\tau,\rho) = (500, 500, 0.7, 0.5)$.

In our experiment, we model the riboflavin production rate as a quantile regression and use ℓ_1-SQR to make variable selection at first. The bandwidth h is chosen to be 0.3 and the tunning parameter λ is selected by cross-validation. Implementing such procedures also picks out 30 genes, which are not exactly the same as those selected by Bühlmann et al. (2014) and Javanmard and Montanari (2014a). After that, we compute p-values for

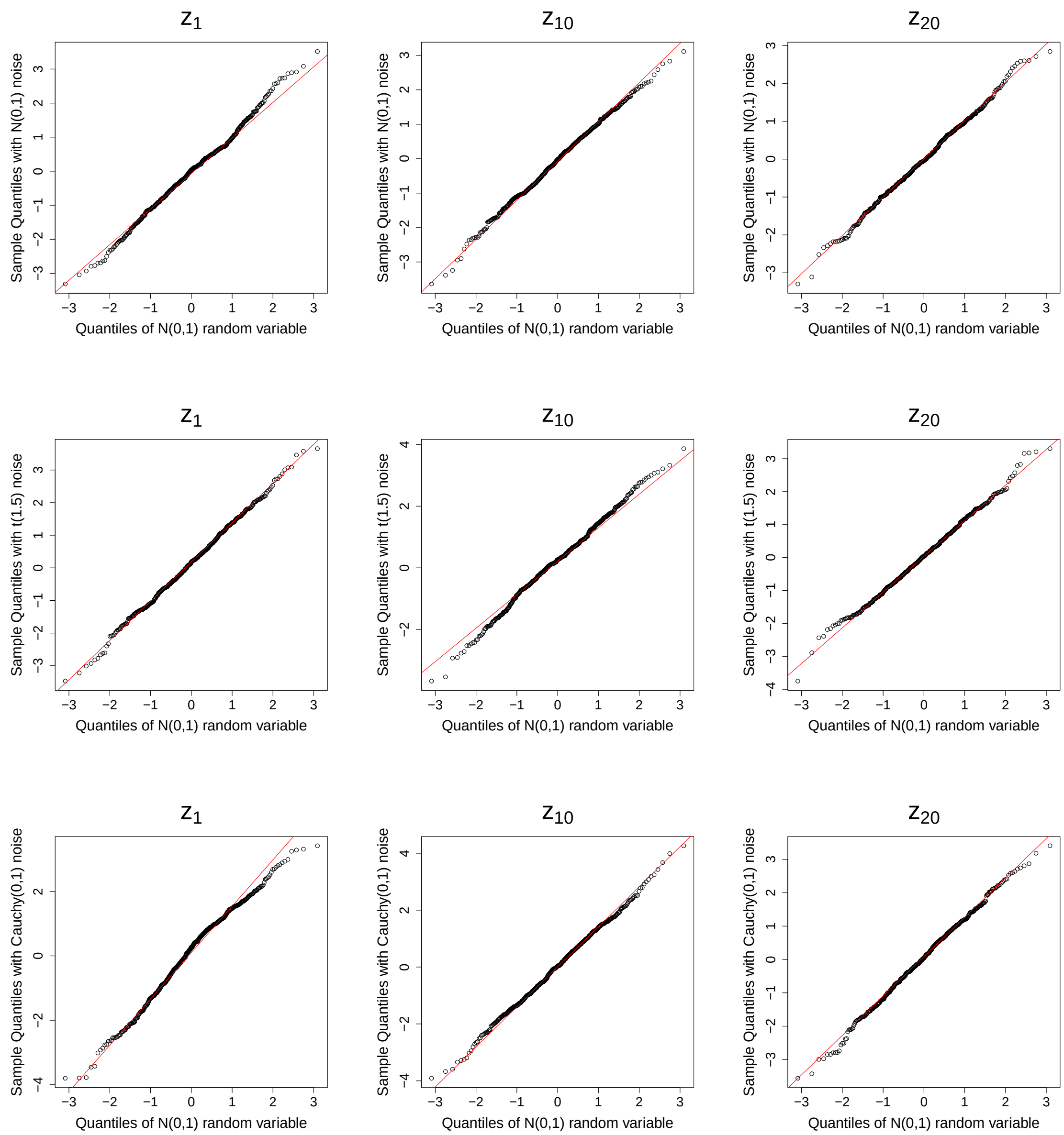

Figure 6: Q-Q plots of $\boldsymbol{z}$ for one realization under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n, p, \tau, \rho) = (500, 1000, 0.7, 0.1)$.

different genes based on formula (22), in which the tunning parameter γ_n is set to be 0.05, and adjust FWER to 5%. The smallest 5 p-values for genes are presented in Table 2. At the quantile levels $\tau \in \{0.1, 0.2, 0.3\}$, we all observe that our method successfully selects a new gene YOAB-at, which is one of the 30 genes that Bühlmann et al. (2014) and Javanmard and Montanari (2014a) found via Lasso but failed to pick out in the end. Along the route

Table 1: Empirical coverage rates with sample size $n = 500$

p	τ	ρ	distribution	β_1	β_{10}	β_{20}	β_{100}	β_{200}
			$N(0,1)$	0.9250	0.9260	0.9530	0.9560	0.9490
500	0.7	0.1	$t(1.5)$	0.9170	0.9220	0.9440	0.9470	0.9460
			$Cauchy(0,1)$	0.9110	0.9080	0.9380	0.9350	0.9360
			$N(0,1)$	0.9190	0.9240	0.9510	0.9480	0.9470
500	0.7	0.5	$t(1.5)$	0.9060	0.9020	0.9380	0.9430	0.9350
			$Cauchy(0,1)$	0.9030	0.8970	0.9320	0.9360	0.9290
			$N(0,1)$	0.9350	0.9240	0.9680	0.9670	0.9720
500	0.4	0.1	$t(1.5)$	0.9160	0.9160	0.9620	0.9560	0.9590
			$Cauchy(0,1)$	0.9120	0.9130	0.9690	0.9640	0.9610
			$N(0,1)$	0.9780	0.9670	0.9820	0.9740	0.9860
1000	0.7	0.1	$t(1.5)$	0.9650	0.9520	0.9610	0.9660	0.9560
			$Cauchy(0,1)$	0.9570	0.9480	0.9580	0.9530	0.9650

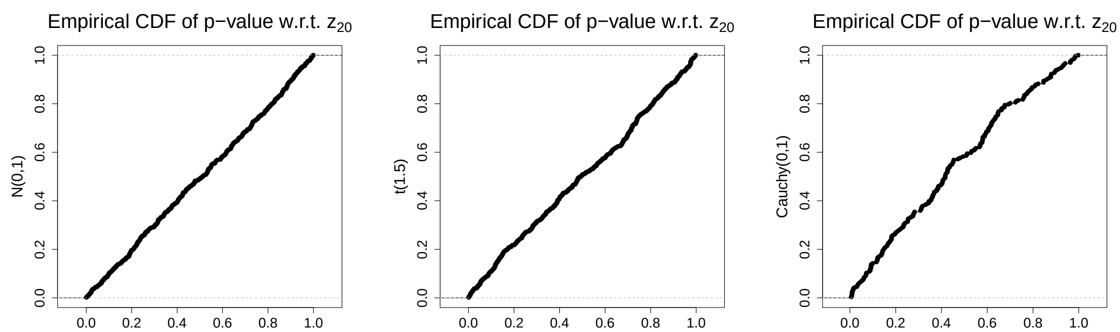

Figure 7: Empirical CDF of p-values (restricted to the entries out of the support set) for one realization under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n, p, \tau, \rho) = (500, 500, 0.7, 0.1)$.

of Bayes factor, Garcia-Donato and Steel (2021) also claimed that the gene YOAB-at has a strong effect on the production of riboflavin. This suggests that our debiased ℓ_1-SQR method provides a valid perspective for high-dimensional data analysis. The corresponding codes are provided in the supplementary materials.

6. Discussion

In this work, we propose the debiased ℓ_1-SQR estimator and use it to construct confidence intervals and implement hypothesis testing under the high-dimensional quantile regression

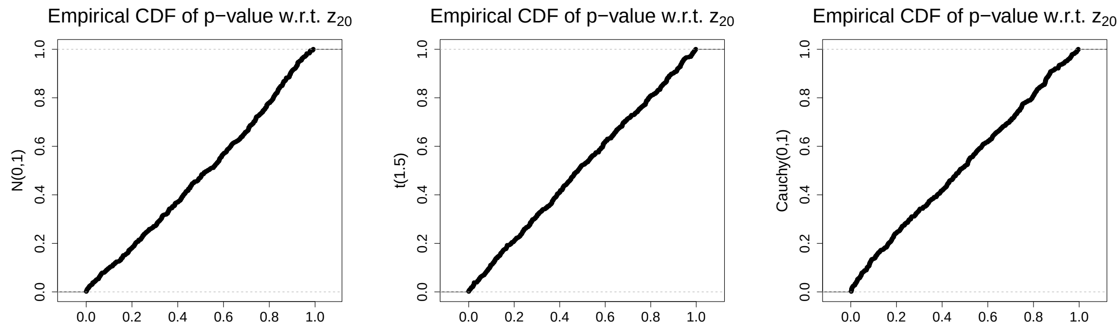

Figure 8: Empirical CDF of p-values (restricted to the entries out of the support set) for one realization under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p,\tau,\rho) = (500,500,0.7,0.5)$.

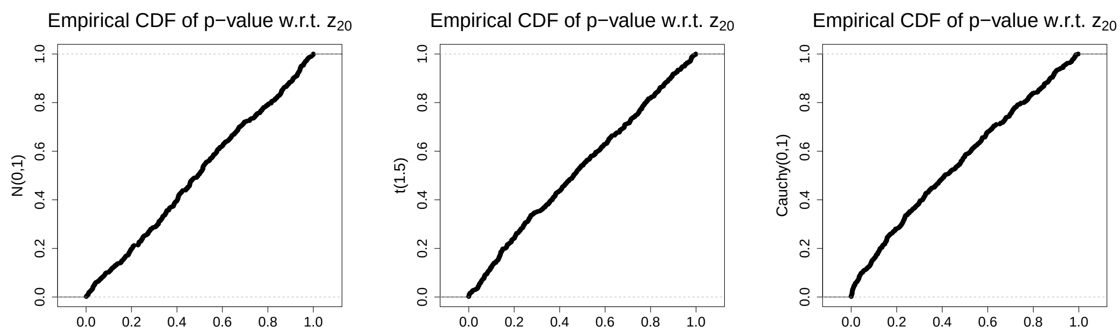

Figure 9: Empirical CDF of p-values (restricted to the entries out of the support set) for one realization under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p,\tau,\rho) = (500,1000,0.7,0.1)$.

framework. By taking advantage of convolution-type smoothing method, ℓ_1-SQR estimate, the vital part of our proposed estimator, can be efficiently obtained via coordinate descent and ADMM-based algorithms. Follow the path of Van de Geer et al. (2014), we then debias the ℓ_1-SQR estimate with the approximate inverse of SQR Hessian matrix. Theoretically, we provide the non-asymptotic Bahadur representation for our debiased ℓ_1-SQR estimator and also the Berry-Esseen bound, which immediately yields the empirical coverage rate for the studentized confidence intervals. Besides, procedures for testing a single variable are designed, and in a minimax perspective we give upper bound of Type I errors and lower bound of statistical powers over a family of sparse vectors. Furthermore, the familywise error rate is guaranteed to be well controlled for the simultaneous testing.

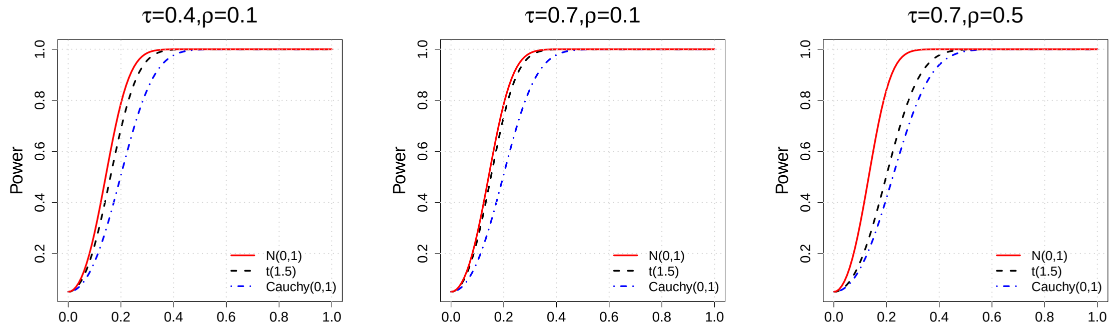

Figure 10: Power curves of testing the null $H_0 : \beta_{20} = 0$ under $N(0,1)$, $t(1.5)$ and $Cauchy(0,1)$ noises with $(n,p) = (500,500)$.

Table 2: The p-values for genes

	YOAB-at	LYSC-at	YEZB-at	YURQ-at	PRIA-at
$\tau = 0.1$	4.3921×10^{-4}	5.6768×10^{-2}	1.1334×10^{-1}	1.9568×10^{-1}	2.4639×10^{-1}
	YOAB-at	YEZB-at	LYSC-at	YYDA-at	YURQ-at
$\tau = 0.2$	6.5066×10^{-4}	2.5527×10^{-1}	3.3260×10^{-1}	3.5554×10^{-1}	3.6481×10^{-1}
	YOAB-at	YEBC-at	YHAI-at	LYSC-at	YYDA-at
$\tau = 0.3$	2.7525×10^{-4}	2.0100×10^{-1}	3.2014×10^{-1}	3.7809×10^{-1}	4.0173×10^{-1}

There are several open questions to be addressed. Nowadays individual data is protected by privacy policies and cannot be freely shared across different units. Traditional integrative analysis is no longer an efficient method and even highly challenging in the ultra high dimensional setting. To the best of our knowledge, latest papers are designing ℓ_1-penalized methods to work on this problem. It is of great interest to extend our debiased ℓ_1-SQR method to this field. On the one hand, we would like to design effective methods to test the heterogeneity based on our testing statistic. On the other hand, developing distributed estimator that can accommodate heterogeneity also merits further research. Besides, we implement simultaneous testing for a group of variables by controlling the familywise error rate in this work. It still remains unknown for the large-scale multiple testing of high-dimensional quantile regression. In the future, we are committed to exploring FDR and other methods to tackle this problem.

Acknowledgments

Xiaozhou Wang's research is supported by the National Natural Science Foundation of China (12101240), the Chenguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (20CG29), and Shanghai Sailing Program (21YF1410500). Riquan Zhang's research is supported by the National Natural Science Foundation of China (11971171, 11831008, 12171310), and the Basic Research Project of Shanghai Science and Technology Commission (22JC1400800). We thank the action editor and two anonymous referees for reading the paper carefully and providing valuable comments, which have improved the manuscript significantly.

Appendix A. Proof of Main Results

In this appendix, we present the proofs of the results in the paper.

Proof of Theorem 1. Following the proof of Theorem 4.1 in Tan et al. (2022b), we first bound the k-th ($k \geq 3$) absolute moments of all the one-dimensional linear projection of $\boldsymbol{w} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}$. Under Assumption 3, we know that $\mathbb{P}\{|\langle \boldsymbol{u}, \boldsymbol{w}\rangle| \geq \sigma t\} \leq 2e^{-\frac{t^2}{2}}$ for all $\boldsymbol{u} \in \mathbb{S}^{p-1}$ and $t \geq 0$. Then for any $k \geq 3$, we have

$$
\begin{aligned}
&\mathbb{E}|\langle \boldsymbol{u}, \boldsymbol{w}\rangle|^k \\
\leq\ & (\mathbb{E}|\langle \boldsymbol{u}, \boldsymbol{w}\rangle|^{2k})^{\frac{1}{2}} \\
=\ & \left(\sigma^{2k} \cdot 2k \int_0^\infty t^{2k-1}\mathbb{P}(|\langle \boldsymbol{u}, \boldsymbol{w}\rangle| \geq \sigma t)\,\mathrm{d}t\right)^{\frac{1}{2}} \\
\leq\ & \left(\sigma^{2k} \cdot 4k \int_0^\infty t^{2k-1}e^{-\frac{t^2}{2}}\,\mathrm{d}t\right)^{\frac{1}{2}} = \sigma^k(2^{k+1}k!)^{\frac{1}{2}} \leq 2\sigma^k k!, \qquad \text{(A.1)}
\end{aligned}
$$

where the first inequality follows from the Lyapunov's inequality and the last one is deduced from the fact $(2^{k+1}k!)^{\frac{1}{2}} = 2 \cdot (2^{k-1}k!)^{\frac{1}{2}} \leq 2 \cdot k!$. With this result, the rest of the proof is similar to that of Theorem 4.1 in Tan et al. (2022b). ■

Proof of Theorem 2. First, we need to find one feasible solution to the optimization problem (11). In fact, $\mathbf{J}_h^{-1}$ is feasible, i.e.,

$$
\|\mathbf{J}_h^{-1}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}\|_\infty \leq \gamma_n. \qquad \text{(A.2)}
$$

Once we have verified this result, the feasibility of $\widehat{\mathbf{W}}$ immediately implies $\|\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}\|_\infty \leq \gamma_n$. According to Cai et al. (2011), solving $\widehat{\mathbf{W}}$ can be equivalently transformed into solving p vector minimization problem in (15). Based on this fact, $\widehat{\mathbf{W}}$ can then be established row by row. Moreover, every row vector of $\widehat{\mathbf{W}}$ is also optimal according to the vector minimization problem. Hence, we also have $\|\widehat{\mathbf{W}}\|_{L_1} = \max\{\|\widehat{\boldsymbol{b}}_1\|_1, \ldots, \|\widehat{\boldsymbol{b}}_p\|_1\} = \|\widehat{\boldsymbol{b}}_k\|_1 \leq \|\widetilde{\boldsymbol{b}}_k\|_1 \leq \max\{\|\widetilde{\boldsymbol{b}}_1\|_1, \ldots, \|\widetilde{\boldsymbol{b}}_p\|_1\} = \|\mathbf{J}_h^{-1}\|_{L_1}$ for some k.

Now we are going to verify inequality (A.2). By the definition of the matrix norm, we know that

$$
\|\mathbf{J}_h^{-1}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}\|_\infty = \|\mathbf{J}_h^{-1}(\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{J}_h)\|_\infty \leq \|\mathbf{J}_h^{-1}\|_{L_1}\|\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \nabla^2 Q_h(\boldsymbol{\beta}^*)\|_\infty,
$$

from which the term $\|\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \nabla^2 Q_h(\boldsymbol{\beta}^*)\|_\infty$ can further be written as

$$
\begin{aligned}
\|\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \nabla^2 Q_h(\boldsymbol{\beta}^*)\|_\infty &\leq \|\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \nabla^2\widehat{Q}_h(\boldsymbol{\beta}^*)\|_\infty + \|\nabla^2\widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla^2 Q_h(\boldsymbol{\beta}^*)\|_\infty \\
&:= U_1 + U_2.
\end{aligned}
$$

Next, we control U_1 and U_2 separately. For U_2, note that

$$\begin{aligned}U_2 = \|\nabla^2 \widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla^2 Q_h(\boldsymbol{\beta}^*)\|_\infty &= \|\frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top - \mathbb{E}[K_h(-\varepsilon)\boldsymbol{x}\boldsymbol{x}^\top]\|_\infty \\ &\leq \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2 \cdot \|\frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i)\boldsymbol{w}_i\boldsymbol{w}_i^\top - \mathbb{E}[K_h(-\varepsilon)\boldsymbol{w}\boldsymbol{w}^\top]\|_\infty \\ &:= \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2 \cdot \|\frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\phi_i\boldsymbol{w}_i\boldsymbol{w}_i^\top\|_\infty,\end{aligned}$$

where $\boldsymbol{w}_i = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}_i$ and $\phi_i = K_h(-\varepsilon_i)$. It is easy to verify that $|\phi_i| \leq \frac{\kappa_u}{h}$ and

$$\mathbb{E}(\phi_i^2|\boldsymbol{x}_i) = \frac{1}{h^2}\int_{-\infty}^{\infty} K^2(-\frac{u}{h})f_{\varepsilon_i|\boldsymbol{x}_i}(u)\,\mathrm{d}u = \frac{1}{h}\int_{-\infty}^{\infty} K^2(v)f_{\varepsilon_i|\boldsymbol{x}_i}(-hv)\,\mathrm{d}v \leq \frac{\bar{f}\kappa_u}{h}.$$

Denote the j-th element of $\boldsymbol{w}_i$ as $w_{i,j}$. Then for arbitrary pair $(j,k) \in \{1,\dots,p\}\times\{1,\dots,p\}$, we bound the higher order moments of $\phi_i w_{i,j} w_{i,k}$ by

$$\mathbb{E}|\phi_i w_{i,j}w_{i,k}|^m = \mathbb{E}|\phi_i(\boldsymbol{e}_j^\top\boldsymbol{w}_i)(\boldsymbol{e}_k^\top\boldsymbol{w}_i)|^m \leq \frac{\bar{f}\kappa_u}{h}\cdot(\frac{\kappa_u}{h})^{m-2}\cdot\mathbb{E}|\boldsymbol{e}_j^\top\boldsymbol{w}_i|^m\cdot\mathbb{E}|\boldsymbol{e}_k^\top\boldsymbol{w}_i|^m.$$

By Lyapunov's inequality,

$$(\mathbb{E}|\boldsymbol{e}_j^\top\boldsymbol{w}_i|^m)^2 \leq \mathbb{E}|\boldsymbol{e}_j^\top\boldsymbol{w}_i|^{2m} \leq \sigma^{2m}\cdot 2m\int_0^\infty t^{2m-1}\mathbb{P}(|\langle\boldsymbol{e}_j,\boldsymbol{w}_i\rangle| \geq \sigma t)\,\mathrm{d}t \leq 2^{m+1}\sigma^{2m}m!.$$

Hence, the Bernstein's condition can be verified as

$$\mathbb{E}|\phi_i w_{i,j}w_{i,k}|^m \leq \frac{m!}{2}\cdot(4\sigma^2)^2\frac{\bar{f}\kappa_u}{h}\cdot(\frac{2\sigma^2\kappa_u}{h})^{m-2}$$

for $m \geq 2$. Applying Bernstein's inequality and taking union bound over all p^2 pairs, we obtain that for any $t > 0$, the following inequality holds with probability at least $1-2p^2e^{-t}$:

$$\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\phi_i\boldsymbol{w}_i\boldsymbol{w}_i^\top\|_\infty = \max_{1\leq j,k\leq p}|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\phi_i w_{i,j}w_{i,k}| \leq 2\sigma^2\big(2\sqrt{2\bar{f}\kappa_u\frac{t}{nh}} + \kappa_u\frac{t}{nh}\big). \tag{A.3}$$

Let $t = \log 2 + 3\log p$, then it follows that with probability at least $1-\frac{1}{p}$ we have

$$U_2 \leq \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2\cdot\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\phi_i\boldsymbol{w}_i\boldsymbol{w}_i^\top\|_\infty \leq 8\|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2\sigma^2\big(\sqrt{2\bar{f}\kappa_u\frac{\log p}{nh}} + \kappa_u\frac{\log p}{nh}\big). \tag{A.4}$$

To bound U_1, observe that

$$\begin{aligned}U_1 &= \|\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \nabla^2\widehat{Q}_h(\boldsymbol{\beta}^*)\|_\infty \\ &\leq \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2\cdot\|\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i)\boldsymbol{w}_i\boldsymbol{w}_i^\top - \frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i)\boldsymbol{w}_i\boldsymbol{w}_i^\top\|_\infty,\end{aligned}$$

where $\widehat{\varepsilon}_i = y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_h$. By the means of Taylor expansion, we have for arbitrary pair $(j,k) \in \{1,\ldots,p\} \times \{1,\ldots,p\}$,

$$
\begin{aligned}
& |\frac{1}{n}\sum_{i=1}^{n} K_h(-\widehat{\varepsilon}_i) w_{i,j} w_{i,k} - \frac{1}{n}\sum_{i=1}^{n} K_h(-\varepsilon_i) w_{i,j} w_{i,k}| \\
= & |\frac{1}{n}\sum_{i=1}^{n} \big[\frac{1}{h^2} K'\big(-\frac{\varepsilon_i}{h}\big)(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h) + \frac{1}{2h^3} K''(\eta_i)(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h)^2\big] w_{i,j} w_{i,k}| \\
\leq & |\frac{1}{nh^2}\sum_{i=1}^{n} K'\big(-\frac{\varepsilon_i}{h}\big) w_{i,j} w_{i,k} (\boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h)| + |\frac{1}{2nh^3}\sum_{i=1}^{n} K''(\eta_i) w_{i,j} w_{i,k} (\boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h)^2| \\
:= & J_1 + J_2,
\end{aligned} \tag{A.5}
$$

where $\widehat{\boldsymbol{\delta}}_h = \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*$ and η_i is an intermediate value between $-\widehat{\varepsilon}_i/h$ and $-\varepsilon_i/h$. Under Assumption 1, $|K''(u)| \leq \kappa_u''$ for any $u \in \mathbb{R}$. Then the union upper bound of J_2 is

$$
\begin{aligned}
\max_{1\leq j,k\leq p} J_2 = & \max_{1\leq j,k\leq p} |\frac{1}{2nh^3}\sum_{i=1}^{n} K''(\eta_i) w_{i,j} w_{i,k} (\boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h)^2| \\
& \leq \max_{1\leq j,k\leq p} \frac{\kappa_u''}{2h^3} |\frac{1}{n}\sum_{i=1}^{n} w_{i,j} w_{i,k} (\|\boldsymbol{x}_i\|_\infty \|\widehat{\boldsymbol{\delta}}_h\|_1)^2| \\
& \leq \frac{\kappa_u''}{2h^3} \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2 \|\widehat{\boldsymbol{\delta}}_h\|_1^2 \cdot \max_{1\leq i\leq n} \|\boldsymbol{w}_i\|_\infty^2 \cdot \max_{1\leq j,k\leq p} |\frac{1}{n}\sum_{i=1}^{n} w_{i,j} w_{i,k}| \\
& \leq \frac{\kappa_u''}{2h^3} \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2 \|\widehat{\boldsymbol{\delta}}_h\|_1^2 \cdot \max_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} |w_{i,j}|^4,
\end{aligned} \tag{A.6}
$$

where the first inequality is derived from the Hölder's inequality. The sub-Gaussian nature of $w_{i,j}$ leads to that for every $t > 0$,

$$
\mathbb{P}(\max_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} |w_{i,j}| \geq t) \leq \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbb{P}(|w_{i,j}| \geq t) \leq 2npe^{-\frac{t^2}{2\sigma^2}}, \tag{A.7}
$$

in which the first inequality follows from the union bound. Setting $t = \sigma\sqrt{2\log(2np^2)}$ and then we have

$$
\max_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} |w_{i,j}| \leq \sigma\sqrt{2\log(2np^2)} \tag{A.8}
$$

with probability at least $1 - \frac{1}{p}$. Inserting (A.8) and the ℓ_1-norm error bound (13) of $\widehat{\boldsymbol{\delta}}_h = \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*$ into (A.6) yields that the following union upper bound of J_2 holds with probability at least $1 - \frac{2}{p}$,

$$
\max_{1\leq j,k\leq p} J_2 \leq \frac{C_2^2 \kappa_u'' s^2 \lambda^2}{2\underline{f}^2 h^3} \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2 \sigma^4 (2\log(2np^2))^2 \asymp \frac{s^2 \log p (\log(p \vee n))^2}{nh^3}. \tag{A.9}
$$

At last, we derive the union upper bound of J_1. Similar to (A.6), it is easy to obtain that

$$\begin{aligned} J_1 &= \Big|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k} (\boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h)\Big| \\ &\le \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1} \cdot \|\widehat{\boldsymbol{\delta}}_h\|_1 \cdot \max_{1\le i\le n} \|\boldsymbol{w}_i\|_\infty \cdot \Big|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\Big|, \end{aligned} \tag{A.10}$$

from which the term $|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}|$ can be further controlled as

$$\begin{aligned} &\Big|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\Big| \\ \le &\Big|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k} - \mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\big]\Big| + \Big|\mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\big]\Big|. \end{aligned}$$

Given $\boldsymbol{x}_i$, the conditional mean $\mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h})|\boldsymbol{x}_i\big]$ satisfies

$$\begin{aligned} \Big|\mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h})|\boldsymbol{x}_i\big]\Big| &= \frac{1}{h^2}\Big|\int_{-\infty}^{\infty} K'(-\frac{u}{h}) f_{\varepsilon_i|\boldsymbol{x}_i}(u)\,\mathrm{d}u\Big| \\ &\le \frac{1}{h}\int_{-\infty}^{\infty} |K'(v)| f_{\varepsilon_i|\boldsymbol{x}_i}(-hv)\,\mathrm{d}v \le \frac{2\bar{f}C_K}{h}, \end{aligned}$$

where $C_K = \int_0^\infty |K'(v)|\,\mathrm{d}v < \infty$ is the total variation of $K(\cdot)$ on $[0,\infty)$. Hence we have

$$\begin{aligned} \Big|\mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\big]\Big| &\le \Big|\mathbb{E}\big[|\mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h})|\boldsymbol{x}_i\big]| w_{i,j} w_{i,k}\big]\Big| \\ &\le \frac{2\bar{f}C_K}{h}\big(\mathbb{E}(w_{i,j})^2 \cdot \mathbb{E}(w_{i,k})^2\big)^{\frac{1}{2}} \le \frac{8\bar{f}C_K\sigma^2}{h}, \end{aligned} \tag{A.11}$$

where the last inequality follows from $\mathbb{E}(w_{i,j})^2 \le 4\sigma^2$ that has been verified in (A.1). For the centered term $\big|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k} - \mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\big]\big|$, it can be bounded in a similar way to (A.3). Denote $\xi_i = \frac{1}{h^2}K'(-\frac{\varepsilon_i}{h})$, and then under Assumption 1 it can be verified that $|\xi_i| \le \frac{\kappa_u'}{h^2}$ and

$$\mathbb{E}[\xi_i^2|\boldsymbol{x}_i] = \frac{1}{h^4}\int_{-\infty}^{\infty} K'^2(-\frac{u}{h}) f_{\varepsilon_i|\boldsymbol{x}_i}(u)\,\mathrm{d}u = \frac{1}{h^3}\int_{-\infty}^{\infty} K'^2(v) f_{\varepsilon_i|\boldsymbol{x}_i}(-hv)\,\mathrm{d}v \le \frac{2\bar{f}\kappa_u' C_K}{h^3}.$$

Analogously, applying Bernstein's inequality and the union bound, we find that the following inequality holds with probability at least $1 - \frac{1}{p}$:

$$\begin{aligned} &\max_{1\le j,k\le p}\Big|\frac{1}{nh^2}\sum_{i=1}^n K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k} - \mathbb{E}\big[\frac{1}{h^2}K'(-\frac{\varepsilon_i}{h}) w_{i,j} w_{i,k}\big]\Big| \\ \le\ & 8\sigma^2\big(2\sqrt{\bar{f}\kappa_u' C_K \frac{\log p}{nh^3}} + \kappa_u' \frac{\log p}{nh^2}\big). \end{aligned} \tag{A.12}$$

Combining this bound with (13), (A.8), (A.10) and (A.11) yields that with probability at least $1-\frac{3}{p}$,

$$\begin{aligned}
&\max_{1\le j,k\le p} J_1\\
&\le \frac{8C_2\sigma^3 s\lambda\sqrt{2\log(2np^2)}}{\underline{f}}\|\mathbf{\Sigma}^{\frac{1}{2}}\|_{L_1}\big(\frac{\bar{f}C_K}{h}+2\sqrt{\bar{f}\kappa_u' C_K\frac{\log p}{nh^3}}+\kappa_u'\frac{\log p}{nh^2}\big)\\
&\asymp s\big(\sqrt{\frac{\log p\log(p\vee n)}{nh^2}}+\sqrt{\frac{(\log p)^2\log(p\vee n)}{n^2h^3}}+\sqrt{\frac{(\log p)^3\log(p\vee n)}{n^3h^4}}\big). \qquad \text{(A.13)}
\end{aligned}$$

Consequently, substituting (A.9) and (A.13) into (A.5) leads to that the following upper bound of U_1

$$U_1\lesssim \frac{s^2\log p(\log(p\vee n))^2}{nh^3}+s\big(\sqrt{\frac{\log p\log(p\vee n)}{nh^2}}+\sqrt{\frac{(\log p)^2\log(p\vee n)}{n^2h^3}}+\sqrt{\frac{(\log p)^3\log(p\vee n)}{n^3h^4}}\big) \qquad \text{(A.14)}$$

holds with probability at least $1-\frac{3}{p}$.

Finally, with probability at least $1-\frac{4}{p}$ we obtain

$$\|\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)-\nabla^2 Q_h(\boldsymbol{\beta}^*)\|_\infty\lesssim\gamma_n,$$

where $\gamma_n=\sqrt{\frac{\log p}{nh}}+\frac{\log p}{nh}+\frac{s^2\log p(\log(p\vee n))^2}{nh^3}+s\big(\sqrt{\frac{\log p\log(p\vee n)}{nh^2}}+\sqrt{\frac{(\log p)^2\log(p\vee n)}{n^2h^3}}+\sqrt{\frac{(\log p)^3\log(p\vee n)}{n^3h^4}}\big)$. In addition to this, we also have

$$\|\widehat{\mathbf{W}}\nabla^2 Q_h(\boldsymbol{\beta}^*)-\mathbf{I}\|_\infty\le\|\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)-\mathbf{I}\|_\infty+\|\widehat{\mathbf{W}}(\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)-\nabla^2 Q_h(\boldsymbol{\beta}^*))\|_\infty\lesssim\gamma_n. \qquad \text{(A.15)}$$

■

Proof of Theorem 3. Let $\boldsymbol{\theta}_i=\widehat{\boldsymbol{b}}_i-\widetilde{\boldsymbol{b}}_i$, $\boldsymbol{\theta}_i^u=(\widehat{b}_{i,1}\mathbb{I}\{|\widehat{b}_{i,1}|\ge 2\gamma_n\|\mathbf{J}_h^{-1}\|_{L_1}\},\ldots,\widehat{b}_{i,p}\mathbb{I}\{|\widehat{b}_{i,p}|\ge 2\gamma_n\|\mathbf{J}_h^{-1}\|_{L_1}\})^\top-\widetilde{\boldsymbol{b}}_i$ and $\boldsymbol{\theta}_i^l=\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^u$. Then we have

$$\|\widetilde{\boldsymbol{b}}_i\|_1-\|\boldsymbol{\theta}_i^u\|_1+\|\boldsymbol{\theta}_i^l\|_1\le\|\widetilde{\boldsymbol{b}}_i+\boldsymbol{\theta}_i^u\|_1+\|\boldsymbol{\theta}_i^l\|_1=\|\widehat{\boldsymbol{b}}_i\|_1\le\|\widetilde{\boldsymbol{b}}_i\|_1,$$

where the last inequality is according to the optimality of $\widehat{\boldsymbol{b}}_i$. Immediately we obtain $\|\boldsymbol{\theta}_i^l\|_1\le\|\boldsymbol{\theta}_i^u\|_1$, and also $\|\boldsymbol{\theta}_i\|_1\le\|\boldsymbol{\theta}_i^u\|_1+\|\boldsymbol{\theta}_i^l\|_1\le 2\|\boldsymbol{\theta}_i^u\|_1$. Based on the results obtained in Theorem 2, the following arguments hold

$$\max_{1\le i,j\le p}|\widehat{b}_{i,j}-\widetilde{b}_{i,j}|=\|\widehat{\mathbf{W}}-\mathbf{J}_h^{-1}\|_\infty\le\|\widehat{\mathbf{W}}\mathbf{J}_h-\mathbf{I}\|_\infty\cdot\|\mathbf{J}_h^{-1}\|_{L_1}\le\gamma_n\|\mathbf{J}_h^{-1}\|_{L_1}.$$

To find the upper bound of $\|\boldsymbol{\theta}_i^u\|_1$, we note that

$$\begin{aligned}
\|\boldsymbol{\theta}_i^u\|_1 \leq & \sum_{j=1}^p |\widehat{b}_{i,j} - \widetilde{b}_{i,j}| \mathbb{I}\{|\widehat{b}_{i,j}| \geq 2\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\} + \sum_{j=1}^p |\widetilde{b}_{i,j}| \mathbb{I}\{|\widehat{b}_{i,j}| < 2\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\} \\
\leq & \sum_{j=1}^p \|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_\infty \mathbb{I}\{|\widehat{b}_{i,j}| \geq 2\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}, |\widehat{b}_{i,j} - \widetilde{b}_{i,j}| \leq \gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\} \\
& + \sum_{j=1}^p |\widetilde{b}_{i,j}| \mathbb{I}\{|\widehat{b}_{i,j}| < 2\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}, |\widehat{b}_{i,j} - \widetilde{b}_{i,j}| \leq \gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\} \\
\leq & \sum_{j=1}^p \|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_\infty \mathbb{I}\{|\widetilde{b}_{i,j}| \geq \gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\} + \sum_{j=1}^p |\widetilde{b}_{i,j}| \mathbb{I}\{|\widetilde{b}_{i,j}| < 3\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\} \\
\leq & \sum_{j=1}^p (\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1})^{1-q} |\widetilde{b}_{i,j}|^q + \sum_{j=1}^p (3\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1})^{1-q} |\widetilde{b}_{i,j}|^q \\
\leq & 4c_{n,p} (\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1})^{1-q},
\end{aligned}$$

where we introduced the certain event $\{|\widehat{b}_{i,j} - \widetilde{b}_{i,j}| \leq \gamma_n \|\mathbf{J}_h^{-1}\|_{L_1}\}$ in deriving the second inequality, and the fourth inequality is based on Assumption 5. Therefore, $\|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_{L_1} = \max_{1 \leq i \leq p} \|\boldsymbol{\theta}_i\|_1 \leq 8c_{n,p} (\gamma_n \|\mathbf{J}_h^{-1}\|_{L_1})^{1-q} \asymp \gamma_n^{1-q}$. ∎

Proof of Theorem 4. First, we derive the union upper bound of $\Gamma_1 = \boldsymbol{\alpha}^\top (\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^n [\tau - \bar{K}_h(-\varepsilon_i)] \boldsymbol{x}_i$. Since $\boldsymbol{\alpha} \in \mathbb{B}_1(r)$, then we have

$$\begin{aligned}
& \sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} |\Gamma_1| = \sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} |\boldsymbol{\alpha}^\top (\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^n [\tau - \bar{K}_h(-\varepsilon_i)] \boldsymbol{x}_i| \\
\leq & \sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} \|(\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}) \boldsymbol{\alpha}\|_1 \cdot \|\frac{1}{\sqrt{n}} \sum_{i=1}^n [\tau - \bar{K}_h(-\varepsilon_i)] \boldsymbol{x}_i\|_\infty \\
\leq & r \sup_{\|\boldsymbol{u}\|_1 \leq 1} \|(\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}) \boldsymbol{u}\|_1 \cdot |\frac{1}{\sqrt{n}} \sum_{i=1}^n [\tau - \bar{K}_h(-\varepsilon_i)]| \cdot \max_{1 \leq i \leq n} \|\boldsymbol{x}_i\|_\infty \\
\leq & r \|\widehat{\mathbf{W}} - \mathbf{J}_h^{-1}\|_{L_1} \cdot \big[|\frac{1}{\sqrt{n}} \sum_{i=1}^n [\tau - \mathbb{E}\bar{K}_h(-\varepsilon)]| + |\frac{1}{\sqrt{n}} \sum_{i=1}^n [\bar{K}_h(-\varepsilon_i) - \mathbb{E}\bar{K}_h(-\varepsilon)]|\big] \cdot \max_{1 \leq i \leq n} \|\boldsymbol{x}_i\|_\infty,
\end{aligned} \tag{A.16}$$

where the first inequality follows from the Hölder's inequality, and the second inequality is based on the definition of $\boldsymbol{\alpha}$ and the property of $\|\cdot\|_\infty$-norm. Applying Hoeffding's inequality for the term $|\frac{1}{n} \sum_{i=1}^n [\bar{K}_h(-\varepsilon_i) - \mathbb{E}\bar{K}_h(-\varepsilon)]|$ implies that for any $t \geq 0$,

$$\mathbb{P}\big(|\frac{1}{n} \sum_{i=1}^n [\bar{K}_h(-\varepsilon_i) - \mathbb{E}\bar{K}_h(-\varepsilon)]| \geq t\big) \leq 2e^{-2nt^2}.$$

Let $t = \sqrt{\frac{\log(2p)}{2n}}$, then we obtain

$$|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\bar{K}_h(-\varepsilon_i) - \mathbb{E}\bar{K}_h(-\varepsilon)]| \leq \sqrt{\frac{\log(2p)}{2}} \tag{A.17}$$

with probability at least $1 - \frac{1}{p}$. On the other hand, it can be verified that

$$\begin{aligned}\mathbb{E}[\bar{K}_h(-\varepsilon)|\boldsymbol{x}] &= \int_{-\infty}^{\infty} \bar{K}_h(-\frac{u}{h}) \, \mathrm{d}F_{\varepsilon|\boldsymbol{x}}(u) = -\frac{1}{h} \int_{-\infty}^{\infty} K(-\frac{u}{h}) F_{\varepsilon|\boldsymbol{x}}(u) \, \mathrm{d}u \\ &= \int_{-\infty}^{\infty} K(v) F_{\varepsilon|\boldsymbol{x}}(-hv) \, \mathrm{d}v = \tau + \int_{-\infty}^{\infty} K(v) \int_{0}^{-hv} \{f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)\} \mathrm{d}t \; \mathrm{d}v,\end{aligned} \tag{A.18}$$

from which it follows that $|\mathbb{E}[\bar{K}_h(-\varepsilon)|\boldsymbol{x}] - \tau| \leq \frac{1}{2} l_0 \kappa_2 h^2$. Consequently,

$$|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\tau - \mathbb{E}\bar{K}_h(-\varepsilon)]| \leq \sqrt{n}\mathbb{E}|\mathbb{E}[\bar{K}_h(-\varepsilon)|\boldsymbol{x}] - \tau| \leq \frac{1}{2} l_0 \kappa_2 \sqrt{n} h^2. \tag{A.19}$$

Combining (A.16) with (A.8), (A.15), (A.18) and (A.19) leads to the following union upper bound of Γ_1 holds

$$\begin{aligned}\sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} |\Gamma_1| &\lesssim r c_{n,p} \gamma_n \sqrt{2 \log(2np^2)} \big(\frac{1}{2} l_0 \kappa_2 \sqrt{n} h^2 + \sqrt{\frac{\log(2p)}{2}} \big) \\ &\asymp \frac{r c_{n,p} s^2 (\log p)^{3/2} (\log(p \vee n))^{5/2}}{n h^3}\end{aligned} \tag{A.20}$$

with probability at least $1 - \frac{5}{p}$.

Next, we turn to consider the term $\Gamma_2 = \frac{1}{\sqrt{n}} \boldsymbol{\alpha}^\top \widehat{\mathbf{W}} \sum_{i=1}^{n} [K_h(-\widehat{\varepsilon}_i) - K_h(\zeta_i)] \boldsymbol{x}_i \boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h$. By a similar argument, we get that

$$\begin{aligned}\sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} |\Gamma_2| &= \sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} |\frac{1}{\sqrt{n}} \boldsymbol{\alpha}^\top \widehat{\mathbf{W}} \sum_{i=1}^{n} [K_h(-\widehat{\varepsilon}_i) - K_h(\zeta_i)] \boldsymbol{x}_i \boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h| \\ &\leq \sup_{\boldsymbol{\alpha} \in \mathbb{B}_1(r)} \|\widehat{\mathbf{W}} \boldsymbol{\alpha}\|_1 \cdot \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [K_h(-\widehat{\varepsilon}_i) - K_h(\zeta_i)] \boldsymbol{x}_i \boldsymbol{x}_i^\top \widehat{\boldsymbol{\delta}}_h\|_\infty \\ &\leq r \|\widehat{\mathbf{W}}\|_{L_1} \cdot \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [K_h(-\widehat{\varepsilon}_i) - K_h(\zeta_i)] \boldsymbol{x}_i \boldsymbol{x}_i^\top\|_\infty \cdot \|\widehat{\boldsymbol{\delta}}_h\|_1,\end{aligned} \tag{A.21}$$

in which the last inequality is derived from the fact that $\|\mathbf{A}\boldsymbol{u}\|_\infty \leq \|\mathbf{A}\|_\infty \|\boldsymbol{u}\|_1$ holds for any matrix $\mathbf{A}$ and vector $\boldsymbol{u}$. So we only need to upper bound the term $\|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [K_h(-\widehat{\varepsilon}_i) -$

$K_h(\zeta_i)]\boldsymbol{x}_i\boldsymbol{x}_i^\top\|_\infty$. Note that

$$
\begin{aligned}
&\|\frac{1}{\sqrt{n}}\sum_{i=1}^n[K_h(-\widehat{\varepsilon}_i)-K_h(\zeta_i)]\boldsymbol{x}_i\boldsymbol{x}_i^\top\|_\infty \le \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2\cdot\|\frac{1}{\sqrt{n}}\sum_{i=1}^n[K_h(-\widehat{\varepsilon}_i)-K_h(\zeta_i)]\boldsymbol{w}_i\boldsymbol{w}_i^\top\|_\infty\\
&\le \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2\cdot\Bigg\{\max_{1\le j,k\le p}|\frac{1}{\sqrt{n}}\sum_{i=1}^n[K_h(-\widehat{\varepsilon}_i)-K_h(-\varepsilon_i)]w_{i,j}w_{i,k}|\\
&\qquad\qquad + \max_{1\le j,k\le p}|\frac{1}{\sqrt{n}}\sum_{i=1}^n[K_h(\zeta_i)-K_h(-\varepsilon_i)]w_{i,j}w_{i,k}|\Bigg\},
\end{aligned}
\tag{A.22}
$$

and $|\zeta_i-(-\varepsilon_i)|\le|\varepsilon_i-\widehat{\varepsilon}_i|=|\boldsymbol{x}_i^\top\widehat{\boldsymbol{\delta}}_h|$. Therefore, applying (A.5), (A.9) and (A.13) implies

$$
\begin{aligned}
&\|\frac{1}{\sqrt{n}}\sum_{i=1}^n[K_h(-\widehat{\varepsilon}_i)-K_h(\zeta_i)]\boldsymbol{x}_i\boldsymbol{x}_i^\top\|_\infty \lesssim 2\sqrt{n}\|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2(\max_{1\le j,k\le p}J_1+\max_{1\le j,k\le p}J_2)\\
&\lesssim\frac{s^2\log p(\log(p\vee n))^2}{n^{1/2}h^3}+s\Big(\sqrt{\frac{\log p\log(p\vee n)}{h^2}}+\sqrt{\frac{(\log p)^2\log(p\vee n)}{nh^3}}+\sqrt{\frac{(\log p)^3\log(p\vee n)}{n^2h^4}}\Big).
\end{aligned}
\tag{A.23}
$$

This bound, together with (13) and (A.8), leads to that with probability at least $1-\frac{4}{p}$ we have

$$
\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}|\Gamma_2|\lesssim\frac{rs^2(\log p)^{3/2}(\log(p\vee n))^2}{nh^3}+rs\sqrt{\frac{(\log p)^2\log(p\vee n)}{nh^2}}\Big(1+\sqrt{\frac{\log p}{nh}}+\frac{\log p}{nh}\Big).
\tag{A.24}
$$

It remains to bound $\Gamma_3=\sqrt{n}\boldsymbol{\alpha}^\top(\mathbf{I}-\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h))\widehat{\boldsymbol{\delta}}_h$. Observe that

$$
\begin{aligned}
\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}|\Gamma_3| &= \sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}|\sqrt{n}\boldsymbol{\alpha}^\top(\mathbf{I}-\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h))\widehat{\boldsymbol{\delta}}_h|\\
&\le\sqrt{n}\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}\|\boldsymbol{\alpha}\|_1\cdot\|(\mathbf{I}-\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h))\widehat{\boldsymbol{\delta}}_h\|_\infty\\
&\le r\sqrt{n}\|\mathbf{I}-\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h)\|_\infty\cdot\|\widehat{\boldsymbol{\delta}}_h\|_1.
\end{aligned}
\tag{A.25}
$$

This, combined with (13), (14) and the condition $s\lesssim(\log(p\vee n))^{1/2}$, yields that

$$
\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}|\Gamma_3|\lesssim\gamma_n rs\sqrt{\log p}\lesssim\frac{rc_{n,p}s^2(\log p)^{3/2}(\log(p\vee n))^{5/2}}{nh^3}.
\tag{A.26}
$$

To sum up, the residual terms are upper bounded with probability at least $1-\frac{5}{p}$, i.e.,

$$
\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}|\Gamma_1+\Gamma_2+\Gamma_3|\lesssim\frac{rc_{n,p}s^2(\log p)^{3/2}(\log(p\vee n))^{5/2}}{nh^3}.
$$

■

Proof of Theorem 5. For arbitrary $\boldsymbol{\alpha} \in \mathbb{B}_1(r)$, from (17) we immediately obtain that

$$\begin{aligned}
&\frac{\sqrt{n}\boldsymbol{\alpha}^\top(\widetilde{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*)}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}} \\
&= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}[\tau - \bar{K}_h(-\varepsilon_i)]\boldsymbol{x}_i}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}} + \frac{\Gamma_1 + \Gamma_2 + \Gamma_3}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}\right) \cdot \sqrt{\frac{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}} \\
&= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\{\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)] - \bar{K}_h(-\varepsilon_i)\}\boldsymbol{x}_i}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}\right. \\
&\qquad \left. + \frac{\Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}\right) \cdot \sqrt{\frac{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}}, \qquad (A.27)
\end{aligned}$$

where $\Gamma_4 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\{\tau - \mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)]\}\boldsymbol{x}_i}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}$ and $\mathbb{E}_{\varepsilon_i}$ denotes the expectation with respect to the probability measure generated by the random noise ε_i.

First, we turn to explore the upper bound of Γ_4. Using a similar argument as in the proof of Theorem 4 yields

$$\begin{aligned}
|\Gamma_4| &\le \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{|\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\{\tau - \mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)]\}\boldsymbol{x}_i|}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}} \\
&\le \frac{l_0\kappa_2 h^2}{2\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^n |\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\boldsymbol{x}_i| \\
&\le \frac{\sqrt{n}l_0\kappa_2 h^2\|\boldsymbol{\alpha}\|_1\|\mathbf{J}_h^{-1}\|_{L_1}}{2\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}} \cdot \max_{1\le i\le n}\|\boldsymbol{x}_i\|_\infty, \qquad (A.28)
\end{aligned}$$

in which the second inequality is based on the fact $|\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)] - \tau| \le \frac{1}{2}l_0\kappa_2 h^2$ that has been verified beforehand and the last one follows from the Hölder's inequality. To control Γ_4, we still need to investigate the lower bound of $\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}$. After introducing the notation $\widetilde{f}(0) = \mathbb{E}[K_h(-\varepsilon)]$, we note that

$$\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha} - \frac{1}{\widetilde{f}(0)}\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\boldsymbol{\alpha} = \boldsymbol{\alpha}^\top\big[\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)}\mathbf{I}\big]\mathbf{J}_h^{-1}\boldsymbol{\alpha}.$$

Hence,

$$
\begin{aligned}
\boldsymbol{\alpha}^\top \mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} \mathbf{J}_h^{-1} \boldsymbol{\alpha} &\geq \frac{1}{\widetilde{f}(0)} \boldsymbol{\alpha}^\top \mathbf{J}_h^{-1} \boldsymbol{\alpha} - \big| \boldsymbol{\alpha}^\top \big[\mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)} \mathbf{I} \big] \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big| \\
&\geq \frac{1}{\widetilde{f}(0)} \|\boldsymbol{\alpha}\|_2^2 \Lambda_{\max}^{-1}(\mathbf{J}_h) - \big| \boldsymbol{\alpha}^\top \big[\mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)} \mathbf{I} \big] \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big| \\
&\geq \frac{1}{2\widetilde{f}(0) \cdot (\bar{f} + l_0 \kappa_1 h)} \Lambda_{\max}^{-1}(\boldsymbol{\Sigma}) - \big| \boldsymbol{\alpha}^\top \big[\mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)} \mathbf{I} \big] \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big| \\
&\geq \frac{1}{2(\bar{f} + l_0 \kappa_1 h)^2} \Lambda_{\max}^{-1}(\boldsymbol{\Sigma}) - \big| \boldsymbol{\alpha}^\top \big[\mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)} \mathbf{I} \big] \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big|,
\end{aligned} \tag{A.29}
$$

in which the last two inequalities follow from the fact $\widetilde{f}(0) \leq \mathbb{E}|\mathbb{E}[K_h(-\varepsilon)|\boldsymbol{x}]| \leq \bar{f} + l_0\kappa_1 h$ and $\mathbf{J}_h \prec 2(\bar{f} + l_0\kappa_1 h)\boldsymbol{\Sigma}$. The first result can be established by noting that

$$
\begin{aligned}
\mathbb{E}[K_h(-\varepsilon)|\boldsymbol{x}] &= \frac{1}{h} \int_{-\infty}^{\infty} K\big(-\frac{u}{h}\big) f_{\varepsilon|\boldsymbol{x}}(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} K(v) f_{\varepsilon|\boldsymbol{x}}(-hv)\,\mathrm{d}v \\
&= f_{\varepsilon|\boldsymbol{x}}(0) + \int_{-\infty}^{\infty} K(v) \big\{ f_{\varepsilon|\boldsymbol{x}}(-hv) - f_{\varepsilon|\boldsymbol{x}}(0) \big\}\,\mathrm{d}v \\
&\leq \bar{f} + \int_{-\infty}^{\infty} K(v) \big| f_{\varepsilon|\boldsymbol{x}}(-hv) - f_{\varepsilon|\boldsymbol{x}}(0) \big| \mathrm{d}v \\
&\leq \bar{f} + l_0 \kappa_1 h,
\end{aligned}
$$

where we use the conditions $f_{\varepsilon|\boldsymbol{x}}(0) \leq \bar{f}$ and $|f_{\varepsilon|\boldsymbol{x}}(u_1) - f_{\varepsilon|\boldsymbol{x}}(u_2)| \leq l_0|u_1 - u_2|$ in Assumption 2. And the second one is then obtained in the way

$$
\mathbf{J}_h = \mathbb{E}[K_h(-\varepsilon)\boldsymbol{x}\boldsymbol{x}^\top] \prec \mathbb{E}[2(\bar{f} + l_0\kappa_1 h)\boldsymbol{x}\boldsymbol{x}^\top] = 2(\bar{f} + l_0\kappa_1 h)\boldsymbol{\Sigma}.
$$

For the term $\big| \boldsymbol{\alpha}^\top \big[\mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)} \mathbf{I} \big] \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big|$, it can be bounded as

$$
\begin{aligned}
&\big| \boldsymbol{\alpha}^\top \big[\mathbf{J}_h^{-1} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widetilde{f}(0)} \mathbf{I} \big] \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big| \\
\leq& \big| \boldsymbol{\alpha}^\top \big(\mathbf{J}_h^{-1} - \widehat{\mathbf{W}} \big) \widehat{\boldsymbol{\Sigma}} \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big| + \big| \boldsymbol{\alpha}^\top \big(\widehat{\mathbf{W}} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widehat{f}(0)} \mathbf{I} \big) \mathbf{J}_h^{-1} \boldsymbol{\alpha} \big| + \big| \frac{1}{\widehat{f}(0)} - \frac{1}{\widetilde{f}(0)} \big| \cdot (\boldsymbol{\alpha}^\top \mathbf{J}_h^{-1} \boldsymbol{\alpha}) \\
\leq& \|\boldsymbol{\alpha}\|_1 \|(\mathbf{J}_h^{-1} - \widehat{\mathbf{W}}) \widehat{\boldsymbol{\Sigma}} \mathbf{J}_h^{-1} \boldsymbol{\alpha}\|_\infty + \|(\widehat{\mathbf{W}} \widehat{\boldsymbol{\Sigma}} - \frac{1}{\widehat{f}(0)} \mathbf{I}) \mathbf{J}_h^{-1} \boldsymbol{\alpha}\|_\infty \|\boldsymbol{\alpha}\|_1 \\
&+ \big| \frac{1}{\widehat{f}(0)} - \frac{1}{\widetilde{f}(0)} \big| \cdot \|\boldsymbol{\alpha}\|_2^2 \Lambda_{\min}^{-1}(\mathbf{J}_h) \\
\leq& \big(\|\mathbf{J}_h^{-1} - \widehat{\mathbf{W}}\|_{L_1} \|\widehat{\boldsymbol{\Sigma}}\|_\infty + \big| \frac{1}{\widehat{f}(0)} \big| \cdot \|\widehat{f}(0) \widehat{\mathbf{W}} \widehat{\boldsymbol{\Sigma}} - \mathbf{I}\|_\infty \big) \|\mathbf{J}_h^{-1}\|_{L_1} \|\boldsymbol{\alpha}\|_1^2 \\
&+ \big| \frac{1}{\widehat{f}(0)} - \frac{1}{\widetilde{f}(0)} \big| \cdot \|\boldsymbol{\alpha}\|_1^2 \Lambda_{\min}^{-1}(\mathbf{J}_h),
\end{aligned} \tag{A.30}
$$

where the second inequality follows from the Hölder's inequality and $\widehat{f}(0)$ is defined as $\widehat{f}(0) = \frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i)$. Consider the term $\widehat{f}(0)\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}$, it is easy to obtain that

$$\begin{aligned}
\widehat{f}(0)\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I} &= \widehat{\mathbf{W}}[\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i)\widehat{\boldsymbol{\Sigma}}] - \mathbf{I} \\
&= \widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top + \widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i)(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{x}_i\boldsymbol{x}_i^\top) - \mathbf{I} \\
&= [\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}] + \widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n [K_h(-\widehat{\varepsilon}_i) - K_h(-\varepsilon_i)](\widehat{\boldsymbol{\Sigma}} - \boldsymbol{x}_i\boldsymbol{x}_i^\top) \\
&\quad + \widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i)(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{x}_i\boldsymbol{x}_i^\top) \\
&:= [\widehat{\mathbf{W}}\nabla^2\widehat{Q}_h(\widehat{\boldsymbol{\beta}}_h) - \mathbf{I}] + \mathbf{R}_1 + \mathbf{R}_2. \qquad (A.31)
\end{aligned}$$

For $\mathbf{R}_1$, we have

$$\begin{aligned}
\|\mathbf{R}_1\|_\infty &= \|\widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n [K_h(-\widehat{\varepsilon}_i) - K_h(-\varepsilon_i)](\widehat{\boldsymbol{\Sigma}} - \boldsymbol{x}_i\boldsymbol{x}_i^\top)\|_\infty \\
&\le \|\widehat{\mathbf{W}}\|_{L_1} \cdot \big(\max_{1\le j,k\le p} \big|\frac{1}{n}\sum_{i=1}^n [K_h(-\widehat{\varepsilon}_i) - K_h(-\varepsilon_i)] w_{i,j}w_{i,k}\big| \\
&\qquad + \big|\frac{1}{n}\sum_{i=1}^n [K_h(-\widehat{\varepsilon}_i) - K_h(-\varepsilon_i)]\big| \|\widehat{\boldsymbol{\Sigma}}\|_\infty \big) \\
&\le \|\widehat{\mathbf{W}}\|_{L_1} \cdot \big(\max_{1\le j,k\le p} \big|\frac{1}{n}\sum_{i=1}^n [K_h(-\widehat{\varepsilon}_i) - K_h(-\varepsilon_i)] w_{i,j}w_{i,k}\big| \\
&\qquad + \big|\frac{1}{n}\sum_{i=1}^n [K_h(-\widehat{\varepsilon}_i) - K_h(-\varepsilon_i)]\big| \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2 \|\frac{1}{n}\sum_{i=1}^n \boldsymbol{w}_i\boldsymbol{w}_i^\top\|_\infty \big). \qquad (A.32)
\end{aligned}$$

By Bernstein's inequality and union bound, the following inequality holds

$$\max_{1\le j,k\le p} \big|\frac{1}{n}\sum_{i=1}^n w_{i,j}w_{i,k} - \mathbb{E}(w_{i,j}w_{i,k})\big| \le 8\sigma^2\big(\sqrt{\frac{2\log p}{n}} + \frac{\log p}{n}\big) \qquad (A.33)$$

with probability at least $1-\frac{1}{p}$. This result, along with (A.5) and the fact $|\mathbb{E}(w_{i,j}w_{i,k})| \le 4\sigma^2$, yields the upper bound

$$\|\mathbf{R}_1\|_\infty \lesssim \frac{s^2\log p(\log(p\vee n))^2}{nh^3} + s\sqrt{\frac{\log p\log(p\vee n)}{nh^2}}\big(1 + \sqrt{\frac{\log p}{nh}} + \sqrt{\frac{\log p}{n}} + \frac{\log p}{nh} + \frac{\log p}{n}\big). \qquad (A.34)$$

Turning to $\mathbf{R}_2$, note that

$$
\begin{aligned}
&\|\mathbf{R}_2\|_\infty = \|\widehat{\mathbf{W}}\frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i)(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{x}_i\boldsymbol{x}_i^\top)\|_\infty \\
&\le \|\widehat{\mathbf{W}}\|_{L_1}\cdot\big(\|\frac{1}{n}\sum_{i=1}^n\big[K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]\big]\widehat{\boldsymbol{\Sigma}}\|_\infty + \|\frac{1}{n}\sum_{i=1}^n\big[K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]\big]\boldsymbol{x}_i\boldsymbol{x}_i^\top\|_\infty\big) \\
&\le \|\widehat{\mathbf{W}}\|_{L_1}\cdot\big(\|\widehat{\boldsymbol{\Sigma}}\|_\infty\big|\frac{1}{n}\sum_{i=1}^n\big[K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]\big]\big| \\
&\qquad\qquad + \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{L_1}^2\big|\frac{1}{n}\sum_{i=1}^n\big[K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]\big]\big|\max_{\substack{1\le i\le n\\1\le j\le p}}|w_{i,j}|^2\big).
\end{aligned}
\tag{A.35}
$$

Since $|K_h(\cdot)| \le \kappa_u/h$, then the term $|\frac{1}{n}\sum_{i=1}^n\big[K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]\big]|$ can be controlled via Hoeffding's inequality as

$$
|\frac{1}{n}\sum_{i=1}^n[K_h(-\varepsilon_i) - \mathbb{E}K_h(-\varepsilon)]| \le \sqrt{\frac{\kappa_u^2\log(2p)}{2nh^2}} \tag{A.36}
$$

with probability at least $1-\frac{1}{p}$. This leads to the following upper bound

$$
\|\mathbf{R}_2\|_\infty \lesssim \sqrt{\frac{\log p}{nh^2}}\big(\log(p\vee n) + \sqrt{\frac{\log p}{n}} + \frac{\log p}{n}\big). \tag{A.37}
$$

Combining this bound with (14) and (A.34) implies

$$
\|\widehat{f}(0)\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}\|_\infty \lesssim \frac{s^2\log p(\log(p\vee n))^2}{nh^3} + \sqrt{\frac{\log p\log(p\vee n)}{nh^2}}\big(s + \sqrt{\log(p\vee n)}\big). \tag{A.38}
$$

Moreover, with overwhelming probability we have

$$
\begin{aligned}
|\widehat{f}(0)| &= |\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i)| \\
&= |\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i) - \frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i) + \frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)] + \widetilde{f}(0)| \\
&\ge \widetilde{f}(0) - |\frac{1}{n}\sum_{i=1}^n K_h(-\widehat{\varepsilon}_i) - \frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i)| - |\frac{1}{n}\sum_{i=1}^n K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]| \\
&\ge \underline{f} - l_0\kappa_1 h - C_3 s\sqrt{\frac{\log p\log(p\vee n)}{nh^2}} > 0,
\end{aligned}
\tag{A.39}
$$

where C_3 is some positive constant and the last inequality is based on the fact $\widetilde{f}(0) = \mathbb{E}[\mathbb{E}[K_h(-\varepsilon)|\boldsymbol{x}]] \geq \underline{f} - l_0\kappa_1 h$. On the other hand,

$$
\begin{aligned}
|\widehat{f}(0) - \widetilde{f}(0)| &= |\frac{1}{n}\sum_{i=1}^{n} K_h(-\widehat{\varepsilon}_i) - \mathbb{E}[K_h(-\varepsilon)]| \\
&\leq |\frac{1}{n}\sum_{i=1}^{n} K_h(-\widehat{\varepsilon}_i) - \frac{1}{n}\sum_{i=1}^{n} K_h(-\varepsilon_i)| + |\frac{1}{n}\sum_{i=1}^{n} K_h(-\varepsilon_i) - \mathbb{E}[K_h(-\varepsilon)]| \\
&\leq C_3 s\sqrt{\frac{\log p \log(p\vee n)}{nh^2}}. \qquad \text{(A.40)}
\end{aligned}
$$

Therefore, applying this bound with (16), (A.29), (A.30), (A.38), and (A.39), and then we find that

$$
\boldsymbol{\alpha}^\top \mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha} \geq \frac{1}{2(\bar{f} + l_0\kappa_1 h)^2}\Lambda_{\max}^{-1}(\boldsymbol{\Sigma}) - C_4\|\mathbf{J}_h^{-1}\|_{L_1}^2\|\boldsymbol{\alpha}\|_1^2 c_{n,p}\gamma_n > 0 \qquad \text{(A.41)}
$$

for some constant $C_4 > 0$. Eventually, the union upper bound of Γ_4 can be written as

$$
\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)} |\Gamma_4| \lesssim rh^2\sqrt{n\log(p\vee n)}. \qquad \text{(A.42)}
$$

Next, we investigate the limit properties of the zero-mean partial sum

$$
S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\{\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)] - \bar{K}_h(-\varepsilon_i)\}\boldsymbol{x}_i}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}} := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i.
$$

Given $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^n$, by the Berry-Esseen inequality (Tyurin (2012)), we have

$$
\sup_{u\in\mathbb{R}}\big|\mathbb{P}\big\{S_n \leq \operatorname{Var}(S_n)^{\frac{1}{2}}u|\mathcal{X}\big\} - \Phi(u)\big| \leq \frac{\sum_{i=1}^n \mathbb{E}_{\varepsilon_i}|Z_i|^3}{2(\sum_{i=1}^n \mathbb{E}_{\varepsilon_i} Z_i^2)^{\frac{3}{2}}}, \qquad \text{(A.43)}
$$

where $\mathbb{E}_{\varepsilon_i}$ denotes the expectation with respect to the probability measure generated by the random noise ε_i. Especially,

$$
\mathbb{E}_{\varepsilon_i}Z_i^2 = \frac{\mathbb{E}_{\varepsilon_i}\{\bar{K}_h(-\varepsilon_i) - \mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)]\}^2}{\tau(1-\tau)} = \frac{\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i) - \tau]^2 - \{\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)] - \tau\}^2}{\tau(1-\tau)}.
$$

According to (A.18), it can be seen that $|\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)] - \tau| \leq \frac{1}{2}l_0\kappa_2 h^2$. By a change of variable and integration by parts, we note that

$$
\begin{aligned}
&\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)]^2 = 2\int_{-\infty}^{\infty}\bar{K}(u)K(u)F_{\varepsilon_i|\boldsymbol{x}_i}(-hu)\,\mathrm{d}u \\
=&\tau - 2hf_{\varepsilon_i|\boldsymbol{x}_i}(0)\int_{-\infty}^{\infty} v\bar{K}(v)K(v)\,\mathrm{d}v + 2\int_{-\infty}^{\infty}\bar{K}(v)K(v)\int_0^{-hv}\{f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)\}\mathrm{d}t\ \mathrm{d}v,
\end{aligned}
$$

from which we immediately get that

$$
\tau(1-\tau) - C_5 h - (1+\tau)l_0\kappa_2 h^2 \leq \mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i) - \tau]^2 \leq \tau(1-\tau) + (1+\tau)l_0\kappa_2 h^2 \qquad \text{(A.44)}
$$

with $C_5 = 2\bar{f}\int_{-\infty}^{\infty} v\bar{K}(v)K(v)\,\mathrm{d}v = 2\bar{f}\int_0^{\infty}\bar{K}(v)(1-\bar{K}(v))\,\mathrm{d}v > 0$. Thus, we obtain $\mathbb{E}_{\varepsilon_i}Z_i^2 \le \frac{\tau(1-\tau)+(1+\tau)l_0\kappa_2 h^2}{\tau(1-\tau)}$ and $\mathbb{E}_{\varepsilon_i}Z_i^2 \ge \frac{\tau(1-\tau)-C_5 h-(1+\tau)l_0\kappa_2 h^2-\frac{1}{2}l_0\kappa_2 h^2}{\tau(1-\tau)}$. In other words, $\mathbb{E}_{\varepsilon_i}Z_i^2 = 1+O(h)$. For the 3-rd moment, it can be verified that

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}_{\varepsilon_i}|Z_i|^3\\
\le&\max_{1\le i\le n}|\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\boldsymbol{x}_i|\cdot\frac{\sum_{i=1}^n\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\mathbf{J}_h^{-1}\boldsymbol{\alpha}}{\big(\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}\big)^{\frac{3}{2}}}\cdot\mathbb{E}_{\varepsilon_i}\big|\bar{K}_h(-\varepsilon_i)-\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)]\big|^3\\
\le&2n\cdot\max_{1\le i\le n}|\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\boldsymbol{x}_i|\cdot\big(\tau(1-\tau)\big)^{-\frac{3}{2}}\big(\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}\big)^{-\frac{1}{2}}\cdot\mathbb{E}_{\varepsilon_i}\big|\bar{K}_h(-\varepsilon_i)-\mathbb{E}_{\varepsilon_i}[\bar{K}_h(-\varepsilon_i)]\big|^2\\
\le&2n\|\boldsymbol{\alpha}\|_1\|\mathbf{J}_h^{-1}\|_{L_1}\cdot\max_{1\le i\le n}\|\boldsymbol{x}_i\|_\infty\cdot\big(\tau(1-\tau)\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}\big)^{-\frac{1}{2}}\cdot\big(1+O(h^2)\big).
\end{aligned}\tag{A.45}$$

Consequently, for any $u\in\mathbb{R}$ and $\boldsymbol{\alpha}\in\mathbb{B}_1(r)$, we have

$$\begin{aligned}
&\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\mathbb{P}(S_n\le u)-\Phi(u)\big| = \sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\mathbb{E}\big[\mathbb{P}\{S_n\le u|\mathcal{X}\}-\Phi(u)\big]\big|\\
\le&\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\mathbb{E}\big|\mathbb{P}\{S_n\le\mathrm{Var}(S_n)^{\frac{1}{2}}u|\mathcal{X}\}-\Phi(\frac{u}{\mathrm{Var}(S_n)^{\frac{1}{2}}}|\mathcal{X})\big|+\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\mathbb{E}\big|\Phi(\frac{u}{\mathrm{Var}(S_n)^{\frac{1}{2}}}|\mathcal{X})-\Phi(u)\big|\\
\le&\mathbb{E}\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\mathbb{P}\{S_n\le\mathrm{Var}(S_n)^{\frac{1}{2}}u|\mathcal{X}\}-\Phi(\frac{u}{\mathrm{Var}(S_n)^{\frac{1}{2}}}|\mathcal{X})\big|+\mathbb{E}\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\Phi(\frac{u}{\mathrm{Var}(S_n)^{\frac{1}{2}}}|\mathcal{X})-\Phi(u)\big|,
\end{aligned}\tag{A.46}$$

where the second inequality is derived via the properties of the conditional expectation. Putting together (A.41), (A.43), (A.44) and (A.45) leads to that

$$\begin{aligned}
&\mathbb{E}\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\mathbb{P}\{S_n\le\mathrm{Var}(S_n)^{\frac{1}{2}}u|\mathcal{X}\}-\Phi(\frac{u}{\mathrm{Var}(S_n)^{\frac{1}{2}}}|\mathcal{X})\big|\\
\le&C_6\sup_{\boldsymbol{\alpha}\in\mathbb{B}_1(r)}\|\boldsymbol{\alpha}\|_1\cdot\frac{\mathbb{E}[\max_{1\le i\le n}\|\boldsymbol{w}_i\|_\infty]}{\sqrt{n}}\lesssim\frac{r\log(p\vee n)}{\sqrt{n}},
\end{aligned}\tag{A.47}$$

in which C_6 is some positive constant and the last step follows from the maximal inequality of the sub-Gaussian random variable, i.e., $\mathbb{E}[\max_{1\le i\le n}\|\boldsymbol{w}_i\|_\infty] = \mathbb{E}[\max_{\substack{1\le i\le n\\1\le j\le p}}|w_{i,j}|] \le \sigma\sqrt{2\log(2np)}$. Moreover, we have already verified that $|\mathrm{Var}(S_n)-1| = O(h)$. Then by an application of Lemma A.7 in the supplement of Spokoiny and Zhilova (2015), for sufficiently small h, we have

$$\mathbb{E}\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\Phi(\frac{u}{\mathrm{Var}(S_n)^{\frac{1}{2}}}|\mathcal{X})-\Phi(u)\big|\lesssim h.\tag{A.48}$$

And hence,

$$\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\big|\mathbb{P}(S_n\le u)-\Phi(u)\big|\lesssim\frac{r\log(p\vee n)}{\sqrt{n}}+h.\tag{A.49}$$

To apply Lemma D.3 in the supplement of Barber and Kolar (2018), we still need to find the upper bound of $\Big|\sqrt{\frac{\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}-1\Big|$. After simple calculation, we have

$$\Big|\sqrt{\frac{\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}-1\Big| = \frac{\Big|\frac{\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}-1\Big|}{\sqrt{\frac{\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}}+1} \leq \Big|\frac{\boldsymbol{\alpha}^\top(\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}-\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1})\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{\alpha}}\Big|. \tag{A.50}$$

For the difference of those two quadratic forms, it is easy to obtain that

$$\begin{aligned}
&\big|\boldsymbol{\alpha}^\top(\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}-\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1})\boldsymbol{\alpha}\big| \\
&\leq \|\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}-\mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\|_\infty\cdot\|\boldsymbol{\alpha}\|_1^2 \\
&\leq \big[\|(\mathbf{J}_h^{-1}-\widehat{\mathbf{W}})\widehat{\boldsymbol{\Sigma}}(\mathbf{J}_h^{-1}-\widehat{\mathbf{W}})\|_\infty+\|(\mathbf{J}_h^{-1}-\widehat{\mathbf{W}})\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\|_\infty+\|\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}(\mathbf{J}_h^{-1}-\widehat{\mathbf{W}})\|_\infty\big]\cdot\|\boldsymbol{\alpha}\|_1^2 \\
&\leq r^2\|\mathbf{J}_h^{-1}-\widehat{\mathbf{W}}\|_{L_1}\|\widehat{\boldsymbol{\Sigma}}\|_\infty\big(\|\mathbf{J}_h^{-1}-\widehat{\mathbf{W}}\|_{L_1}+2\|\widehat{\mathbf{W}}\|_\infty\big) \\
&\lesssim \frac{r^2c_{n,p}s^2\log p(\log(p\vee n))^2}{nh^3}.
\end{aligned}\tag{A.51}$$

Finally, putting all pieces together and applying Lemma D.3 aforementioned, we conclude that with probability approaching 1 the following bound holds

$$\begin{aligned}
&\sup_{\substack{u\in\mathbb{R}\\ \boldsymbol{\alpha}\in\mathbb{B}_1(r)}}\left|\mathbb{P}\left(\frac{\sqrt{n}\boldsymbol{\alpha}^\top(\widetilde{\boldsymbol{\beta}}_h-\boldsymbol{\beta}^*)}{\sqrt{\tau(1-\tau)\boldsymbol{\alpha}^\top\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{\alpha}}}\leq u\right)-\Phi(u)\right| \\
&\lesssim \frac{rc_{n,p}s^2\log p(\log(p\vee n))^2(r+\sqrt{\log p\log(p\vee n)})}{nh^3}+rh^2\sqrt{n\log(p\vee n)}.
\end{aligned}\tag{A.52}$$

Since the right hand side does not depend on $\boldsymbol{\beta}^*$, then the coverage probability (20) follows. ■

Proof of Theorem 6. At the beginning, we show that $\varrho_j(\widehat{T}_j)$ is uniformly upper bounded by ϱ. Let $V_j=\frac{\sqrt{n}(\widetilde{\beta}_{h;j}-\beta_j^*)}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}$. By the definition of $\varrho_j(\widehat{T}_j)$, we have

$$\begin{aligned}
\lim_{n\to\infty}\varrho_j(\widehat{T}_j) &= \lim_{n\to\infty}\sup_{\boldsymbol{\beta}^*}\big\{\mathbb{P}_{\boldsymbol{\beta}^*}(P_j\leq\varrho):\boldsymbol{\beta}^*\in\mathbb{R}^p,\|\boldsymbol{\beta}^*\|_0\leq s,\beta_j^*=0\big\} \\
&= \lim_{n\to\infty}\sup_{\boldsymbol{\beta}^*}\Big\{\mathbb{P}_{\boldsymbol{\beta}^*}\big(z_{1-\frac{\varrho}{2}}\leq\frac{\sqrt{n}|\widetilde{\beta}_{h;j}|}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}\big):\boldsymbol{\beta}^*\in\mathbb{R}^p,\|\boldsymbol{\beta}^*\|_0\leq s,\beta_j^*=0\Big\} \\
&= \lim_{n\to\infty}\sup_{\boldsymbol{\beta}^*}\big\{\mathbb{P}_{\boldsymbol{\beta}^*}(z_{1-\frac{\varrho}{2}}\leq|V_j|):\boldsymbol{\beta}^*\in\mathbb{R}^p,\|\boldsymbol{\beta}^*\|_0\leq s\big\}\leq\varrho,
\end{aligned}$$

in which the last inequality is derived from (21).

Now we turn to establishing the lower bound for the power of test $\widehat{T}_j$. According to (A.30) and (A.51), setting $\boldsymbol{\alpha}=\boldsymbol{e}_j$ therein and implementing a similar analysis thereafter yields that for arbitrary $\delta > 0$,

$$
\begin{aligned}
&\mathbb{P}\big(\boldsymbol{e}_j^\top \widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{e}_j - \frac{1}{\widetilde{f}(0)}\boldsymbol{e}_j^\top \mathbf{J}_h^{-1}\boldsymbol{e}_j \geq \delta\big)\\
&\leq \mathbb{P}\big(\boldsymbol{e}_j^\top \widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{e}_j - \boldsymbol{e}_j^\top \mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{e}_j \geq \frac{\delta}{2}\big) + \mathbb{P}\big(\boldsymbol{e}_j^\top \mathbf{J}_h^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{J}_h^{-1}\boldsymbol{e}_j - \frac{1}{\widetilde{f}(0)}\boldsymbol{e}_j^\top \mathbf{J}_h^{-1}\boldsymbol{e}_j \geq \frac{\delta}{2}\big)\\
&\leq \frac{C_7}{p^2},
\end{aligned}
\tag{A.53}
$$

where $\widetilde{f}(0) = \mathbb{E}[K_h(\varepsilon)]$, $C_7 > 0$ is some positive constant and the last inequality follows from a series of exponential inequalities. Therefore, by Borel-Cantelli's lemma we obtain that the following inequality holds almost surely

$$
\limsup_{n\to\infty} \big([\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j} - \frac{1}{\widetilde{f}(0)}[\mathbf{J}_h^{-1}]_{j,j}\big) = \limsup_{n\to\infty} \big[\boldsymbol{e}_j^\top \widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}\boldsymbol{e}_j - \frac{1}{\widetilde{f}(0)}\boldsymbol{e}_j^\top \mathbf{J}_h^{-1}\boldsymbol{e}_j\big] \leq 0. \tag{A.54}
$$

According to the definition of $\pi_j(\widehat{T}_j;\mu)$, we have

$$
\begin{aligned}
&\liminf_{n\to\infty} \frac{1-\pi_j(\widehat{T}_j;\mu)}{1-\pi_j(\mu)}\\
=\ &\liminf_{n\to\infty} \frac{1}{1-\pi_j(\mu)} \inf_{\boldsymbol{\beta}^*} \big\{\mathbb{P}_{\boldsymbol{\beta}^*}(P_j \leq \varrho) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \leq s, |\beta_j^*| \geq \mu\big\}\\
=\ &\liminf_{n\to\infty} \frac{1}{1-\pi_j(\mu)} \inf_{\boldsymbol{\beta}^*} \big\{\mathbb{P}_{\boldsymbol{\beta}^*}(z_{1-\frac{\varrho}{2}} \leq \frac{\sqrt{n}|\widetilde{\beta}_{h;j}|}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \leq s, |\beta_j^*| \geq \mu\big\}\\
=\ &\liminf_{n\to\infty} \frac{1}{1-\pi_j(\mu)} \inf_{\boldsymbol{\beta}^*} \big\{\mathbb{P}_{\boldsymbol{\beta}^*}\big(z_{1-\frac{\varrho}{2}} \leq \big|V_j + \frac{\sqrt{n}\beta_j^*}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}\big|\big) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \leq s, |\beta_j^*| \geq \mu\big\}\\
\geq\ &\liminf_{n\to\infty} \frac{1}{1-\pi_j(\mu)} \inf_{\boldsymbol{\beta}^*} \big\{\mathbb{P}_{\boldsymbol{\beta}^*}\big(z_{1-\frac{\varrho}{2}} \leq \big|V_j + \frac{\sqrt{n\widetilde{f}(0)}\mu}{\sqrt{\tau(1-\tau)}[\mathbf{J}_h^{-1}]_{j,j}^{1/2}}\big|\big) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \leq s\big\}\\
=\ &\liminf_{n\to\infty} \frac{1}{1-\pi_j(\mu)} G\Bigg(\varrho, \frac{\sqrt{n\mathbb{E}[K_h(\varepsilon)]}\mu}{\sqrt{\tau(1-\tau)}[\mathbf{J}_h^{-1}]_{j,j}^{1/2}}\Bigg) = 1,
\end{aligned}
\tag{A.55}
$$

where the inequality follows from (A.54) and the condition that $|\beta_j^*| \geq \mu$. ■

Proof of Theorem 7. Based on the definition of FWER, we have

$$
\begin{aligned}
&\limsup_{n\to\infty}\ \mathrm{FWER}(\widehat{T}^{\mathcal{G}}, n)\\
=&\limsup_{n\to\infty} \sup_{\boldsymbol{\beta}^*\in\mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0\leq s} \mathbb{P}_{\boldsymbol{\beta}^*}\big\{\exists j \in \mathcal{G},\ \text{s.t.}\ z_{1-\frac{\varrho}{2|\mathcal{G}|}} \leq \frac{\sqrt{n}|\widetilde{\beta}_{h;j}|}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}, \beta_j^* = 0\big\}\\
\leq&\limsup_{n\to\infty} \sum_{j\in\mathcal{G}} \sup_{\boldsymbol{\beta}^*} \big\{\mathbb{P}_{\boldsymbol{\beta}^*}(z_{1-\frac{\varrho}{2|\mathcal{G}|}} \leq \frac{\sqrt{n}|\widetilde{\beta}_{h;j}|}{\sqrt{\tau(1-\tau)}[\widehat{\mathbf{W}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{W}}]_{j,j}^{1/2}}) : \boldsymbol{\beta}^* \in \mathbb{R}^p, \|\boldsymbol{\beta}^*\|_0 \leq s, \beta_j^* = 0\big\}\\
\leq&\varrho + |\mathcal{G}| \cdot \limsup_{n\to\infty} C_8\big\{\frac{c_{n,p}s^2\log p(\log(p\vee n))^2\sqrt{\log p\log(p\vee n)}}{nh^3} + h^2\sqrt{n\log(p\vee n)}\big\} = \varrho,
\end{aligned}
$$

where the first inequality follows from the Bonferroni's inequality and the second inequality is via (A.52) with some positive constant $C_8 > 0$. ■

References

Rina Foygel Barber and Mladen Kolar. Rocket: Robust confidence intervals via Kendall's tau for transelliptical graphical models. *The Annals of Statistics*, 46(6B):3422–3450, 2018.

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

Alexandre Belloni, Victor Chernozhukov, and Kengo Kato. Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika*, 102 (1):77–94, 2015.

Alexandre Belloni, Victor Chernozhukov, and Ying Wei. Post-selection inference for generalized linear models with many controls. *Journal of Business and Economic Statistics*, 34(4):606–619, 2016.

Alexandre Belloni, Victor Chernozhukov, and Kengo Kato. Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Journal of the American Statistical Association*, 114(526):749–758, 2019.

Jelena Bradic and Mladen Kolar. Uniform inference for high-dimensional quantile regression: Linear functionals and regression rank scores. *arXiv preprint arXiv:1702.06209*, 2017.

Peter Bühlmann and Sara Van de Geer. *Statistics for High-dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media, 2011.

Peter Bühlmann, Markus Kalisch, and Lukas Meier. High-dimensional statistics with a view toward applications in biology. *Annual Review of Statistics and Its Application*, 1 (1):255–278, 2014.

T Tony Cai, Weidong Liu, and Xi Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494): 594–607, 2011.

T Tony Cai, Weidong Liu, and Harrison H Zhou. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics*, 44(2): 455–488, 2016.

Xi Chen, Weidong Liu, and Yichen Zhang. Quantile regression under memory constraint. *The Annals of Statistics*, 47(6):3244–3273, 2019.

Chao Cheng, Xingdong Feng, Jian Huang, and Xu Liu. Regularized projection score estimation of treatment effects in high-dimensional quantile regression. *Statistica Sinica*, 32 (1):23–41, 2022.

Luciano de Castro, Antonio F Galvao, David M Kaplan, and Xin Liu. Smoothed GMM for quantile models. *Journal of Econometrics*, 213(1):121–144, 2019.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

Jianqing Fan and Jinchi Lv. Nonconcave penalized likelihood with NP-dimensionality. *IEEE Transactions on Information Theory*, 57(8):5467–5484, 2011.

Jianqing Fan, Lingzhou Xue, and Hui Zou. Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics*, 42(3):819, 2014.

Jianqing Fan, Han Liu, Qiang Sun, and Tong Zhang. I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *The Annals of Statistics*, 46(2):814, 2018.

Jianqing Fan, Runze Li, Cun-Hui Zhang, and Hui Zou. *Statistical Foundations of Data Science*. Chapman and Hall/CRC, 2020.

Zhe Fei, Qi Zheng, Hyokyoung G Hong, and Yi Li. Inference for high-dimensional censored quantile regression. *Journal of the American Statistical Association*, 118(542):898–912, 2023.

Marcelo Fernandes, Emmanuel Guerre, and Eduardo Horta. Smoothing quantile regressions. *Journal of Business and Economic Statistics*, 39(1):338–357, 2021.

Antonio F Galvao and Kengo Kato. Smoothed quantile regression for panel data. *Journal of Econometrics*, 193(1):92–112, 2016.

Gonzalo Garcia-Donato and Mark FJ Steel. Bayes factors based on g-priors for variable selection. In *Handbook of Bayesian Variable Selection*, pages 351–370. Chapman and Hall/CRC, 2021.

Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC, 2019.

Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2):367–388, 2023.

Joel L Horowitz. Bootstrap methods for median regression models. *Econometrica*, 66(6): 1327–1351, 1998.

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

Adel Javanmard and Andrea Montanari. Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory*, 60(10):6522–6554, 2014b.

Rong Jiang and Keming Yu. Smoothing quantile regression for a distributed system. *Neurocomputing*, 466:311–326, 2021.

David M Kaplan and Yixiao Sun. Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory*, 33(1):105–157, 2017.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

Roger Koenker, Victor Chernozhukov, Xuming He, and Limin Peng. *Handbook of Quantile Regression*. CRC press, 2017.

Sai Li, T Tony Cai, and Hongzhe Li. Transfer learning in large-scale Gaussian graphical models with false discovery rate control. *Journal of the American Statistical Association*, in press, 2022.

Heng Lian and Zengyan Fan. Divide-and-conquer for debiased l_1-norm support vector machine in ultra-high dimensions. *Journal of Machine Learning Research*, 18(182):1–26, 2018.

Han Liu and Lie Wang. TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electronic Journal of Statistics*, 11(1):241–294, 2017.

Rebeka Man, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. A unified algorithm for penalized convolution smoothed quantile regression. *arXiv preprint arXiv:2205.02432*, 2022.

Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

Haeseong Moon and Wen-Xin Zhou. High-dimensional composite quantile regression: Optimal statistical guarantees and fast algorithms. *arXiv preprint arXiv:2208.09817*, 2022.

Yang Ning and Han Liu. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1):158–195, 2017.

James L Powell et al. Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics*, pages 357–384. Cambridge University Press, 1991.

Peijun Sang, Zuofeng Shang, and Pang Du. Statistical inference for functional linear quantile regression. *arXiv preprint arXiv:2202.11747*, 2022.

Vladimir Spokoiny and Mayya Zhilova. Bootstrap confidence sets under model misspecification. *The Annals of Statistics*, 43(6):2653–2675, 2015.

Kean Ming Tan, Heather Battey, and Wen-Xin Zhou. Communication-constrained distributed quantile regression with optimal statistical guarantees. *Journal of Machine Learning Research*, 23:1–61, 2022a.

Kean Ming Tan, Lan Wang, and Wen-Xin Zhou. High-dimensional quantile regression: Convolution smoothing and concave regularization. *Journal of the Royal Statistical Society: Series B*, 84(1):205–223, 2022b.

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

Chau Tran and Guo Yu. A completely tuning-free and robust approach to sparse precision matrix estimation. In *International Conference on Machine Learning*, pages 21733–21750. PMLR, 2022.

I. S. Tyurin. A refinement of the remainder in the Lyapunov theorem. *Theory of Probability and Its Applications*, 56(4):693–696, 2012.

Sara Van de Geer, Peter Bühlmann, Ya'acov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.

Martin J Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

Hansheng Wang, Guodong Li, and Guohua Jiang. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business and Economic Statistics*, 25(3):347–355, 2007.

Lan Wang, Yichao Wu, and Runze Li. Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107(497):214–222, 2012.

Lan Wang, Yongdai Kim, and Runze Li. Calibrating non-convex penalized regression in ultra-high dimension. *The Annals of Statistics*, 41(5):2505, 2013.

Lie Wang. The L_1 penalized LAD estimator for high dimensional linear regression. *Journal of Multivariate Analysis*, 120:135–151, 2013.

Yoon-Jae Whang. Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory*, 22(2):173–205, 2006.

Yuanshan Wu, Yanyuan Ma, and Guosheng Yin. Smoothed and corrected score approach to censored quantile regression with measurement errors. *Journal of the American Statistical Association*, 110(512):1670–1683, 2015.

Yibo Yan, Xiaozhou Wang, and Riquan Zhang. Composite smoothed quantile regression. *Stat*, 12(1):e542, 2023.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B*, 76 (1):217–242, 2014.

Yijiao Zhang and Zhongyi Zhu. Transfer learning for high-dimensional quantile regression via convolution smoothing. *arXiv preprint arXiv:2212.00428*, 2022.

Peng Zhao and Bin Yu. On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7:2541–2563, 2006.

Tianqi Zhao, Mladen Kolar, and Han Liu. A general framework for robust testing and confidence regions in high-dimensional quantile regression. *arXiv preprint arXiv:1412.8724*, 2014.

Weihua Zhao, Fode Zhang, and Heng Lian. Debiasing and distributed estimation for high-dimensional quantile regression. *IEEE Transactions on Neural Networks and Learning Systems*, 31(7):2569–2577, 2019.

Qi Zheng, Limin Peng, and Xuming He. Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics*, 43(5):2225, 2015.

Qi Zheng, Limin Peng, and Xuming He. High dimensional censored quantile regression. *The Annals of Statistics*, 46(1):308, 2018.

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