Statistica Sinica Preprint No: SS-2019-0304	
Title	CONDITIONAL MARGINAL TEST FOR HIGH DIMENSIONAL QUANTILE REGRESSION
Manuscript ID	SS-2019-0304
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0304
Complete List of Authors	Yanlin Tang, Yinfeng Wang, Huixia Judy Wang and Qing Pan
Corresponding Author	Yinfeng Wang
E-mail	dairy-2006@163.com
Notice: Accepted version subject to English editing.	

Statistica Sinica

CONDITIONAL MARGINAL TEST FOR HIGH DIMENSIONAL QUANTILE REGRESSION

Yanlin Tang[1], Yinfeng Wang[2], Huixia Judy Wang[3] and Qing Pan[3]

[1]*East China Normal University*

[2]*Shanghai Lixin University of Accounting and Finance*

and [3]*George Washington University*

Abstract: Analysis of tail quantiles of the response distribution is sometimes more important than the mean in biomarker studies. Inference in quantile regression is complicated when there exist a large number of candidate markers together with some pre-specified controlled covariates. In this paper, we develop a new and simple testing procedure to detect the effects of biomarkers in high-dimensional quantile regression in the presence of protected covariates. The test is based on the maximum-score-type statistic obtained from conditional marginal regression. We establish the asymptotic properties of the proposed test statistic under both null and alternative hypotheses, and further propose an alternative multiplier bootstrap method with theoretical justifications. We demonstrate through numerical studies that the proposed method provides adequate controls of the family-wise error rate with competitive power, and it can also be used as a stopping rule in the forward regression. The proposed method is applied to a motivating genome-wide association study

Corresponding author, email: dairy-2006@163.com.

to detect single nucleotide polymorphisms associated with low glomerular filtration rates in Type 1 diabetes patients.

Key words and phrases: Conditional marginal regression, Extreme value distribution, High dimensional, Maximal score statistic, Multiplier bootstrap.

1. Introduction

Genome-wide association study (GWAS) screens for associations between a large number of single-nucleotide polymorphisms (SNPs) and phenotypes such as disease symptoms and clinical index. It has been known that genes often do not function individually but tend to work together in a biological process; see for instance Zou et al. (2004), de Leeuw et al. (2016), and Sun et al. (2019). Therefore, it is important to identify gene sets, that is, classes of genes that jointly have an association with disease phenotypes. Inference in the context of gene set detection faces the challenges of both high-dimensionality and multiplicity, since the number of genes in a set can be much larger than the sample size and genes in different sets may overlap.

This paper is motivated by a GWAS from the Diabetes Complication and Control Trial (DCCT), where SNPs associated with the glomerular filtration rate (GFR) are searched through a genome-wide screening. The GFR is an important clinical index for the risks of nephropathy, one of the major microvascluar complications in diabetic patients. The study has three unique features. Firstly,

the mean level of GFR among participants is less clinically informative than the left tail quantiles, because the mean values are usually driven by the majority of participants without nephropathy, while the lower quantiles reflect the characteristics of the subset of participants that progressed to nephropathy. Secondly, the GFR data are skewed to the left even after log-transformation; see Figure 4.1 in Section 4. Thirdly, the data contains a large number of SNPs and some "protected" covariates, which are known to impact GFR levels, such as age, duration of diabetes and body mass index. Let Y be a scalar response variable, corresponding to GFR, and $(\mathbf{Z}^\top, \mathbf{X}^\top)^\top$ be a p_n-dimensional set of covariates, where n is the sample size, $\mathbf{Z}$ is a q-dimensional (q is fixed) conditioning set corresponding to the "protected" covariates, and $\mathbf{X}$ is the remaining d_n-dimensional covariates with $d_n = p_n - q$, corresponding to the SNPs. Our goal is to assess the association between SNPs and the lower tails of the GFR distribution to identify SNPs and gene pathways associated with patients at higher risk of kidney failure, after controlling the effect of the protected covariates.

In GWAS, the most commonly used approach is to test trait-SNP associations (conditioning on $\mathbf{Z}$) for one SNP at a time, followed by a multiple comparison adjustment, e.g. Bonferroni adjustment or false discovery rate (FDR) control. Bonferroni adjustment controls the family-wise error rate (FWER) well, however, such adjustment is usually conservative, which may result in low power

under the alternative. The FDR control works in a different way than FWER control, which is suitable for the case where there exist many important covariates. Other existing work in GWAS mainly focused on mean-regression-based tests. Without the inclusion of $\mathbf{Z}$, Zou et al. (2004) proposed a resampling procedure to assess the significance of genome-wide quantitative trait loci mapping for Drosophila backcross, and McKeague and Qian (2015) proposed an adaptive resampling test and applied it to the analysis of a glioblastoma cancer data. Guo and Chen (2016) proposed to test the overall significance of $\mathbf{X}$ conditional on $\mathbf{Z}$, based on a quadratic form of the score functions. Tang et al. (2018) proposed a hybrid test of maximum- and sum-squared-type statistics based on conditional marginal regressions, by regressing Y on $\mathbf{Z}$ and each X_j separately. Based on the sum of powered scores (Pan et al., 2014; Xu et al., 2016), Wu et al. (2019) proposed an adaptive test for generalized linear models, by assuming that the errors satisfy the sub-Gaussian condition and $p_n = o(n^2)$. None of these mean-based methods are suitable for analyzing the GFR data to meet the research goals.

As a valuable alternative to the mean regression, quantile regression provides a natural way to capture the impact of covariates on the tail of the response distribution. Quantile regression generally does not require any parametric distributional assumptions, and can accommodate skewed distributions and heteroscedasticity automatically. There exist various inference methods for quantile

regression, including Wald-type, quasi-likelihood-ratio and rank score tests, and resampling-based approaches; see related discussions in Koenker (2005, Chapter 3), Kocherginsky et al. (2005), Feng et al. (2011), and Wang et al. (2018b). Unfortunately, the existing tests are for low-dimensional covariates, and they either have low power for large p_n or are infeasible for cases with $p_n \geq n$. For inference with high-dimensional $\mathbf{X}$, one may first select a subset of predictors using some variable selection methods (Wu and Liu, 2009; Belloni and Chernozhukov, 2011; Wang et al., 2012; Sherwood and Wang, 2016), and then conduct hypothesis testing using conventional methods on the selected model. However, such practice ignores the uncertainty involved in the model selection step and thus often leads to inflated FWER (Leeb and Pötscher, 2003, 2005).

To detect significant predictors while accounting for uncertainties involved in the selection stage, Wang et al. (2018a) proposed a quantile marginal effect test, based on the maximum of the marginal t-statistics, and Wang et al. (2018b) considered wild residual bootstrap inference for penalized quantile regression, without the presence of $\mathbf{Z}$. However, their theories only work for fixed dimension, and the method in Wang et al. (2018a) uses a computationally intensive double bootstrap procedure for selecting the tuning parameter involved in the test calibration. Furthermore, in clinical studies, prognostic factors should be selected after accounting for the effects of some protected covariates with known

impacts on the outcome. With the inclusion of $\mathbf{Z}$, Park and He (2017) extended the rank score test for quantile regression with fixed dimensions to settings with diverging p_n; however, this method still requires $p_n < n$.

In this paper, we propose a conditional marginal score-type test for quantile regression in the ultra-high-dimensional setting, to detect the overall significance of $\mathbf{X}$ on the quantile of Y, in the presence of "protected" covariates $\mathbf{Z}$. More specifically, for $j = 1, \ldots, d_n$, we evaluate the additional effect of each X_j conditional on $\mathbf{Z}$, through rescaled conditional marginal rank scores, and define the final test statistic by the maximum of d_n squared score statistics. Different from the existing work, our method allows the dimension d_n to diverge with n and be much larger than n, for instance, $d_n = O\{\exp(n^{c_0})\}$ for some $c_0 > 0$. Under some regularity conditions, we establish the asymptotic properties of the proposed test statistic under both null and alternative hypotheses. To improve the finite-sample performance, we propose an alternative calibration method based on a multiplier bootstrap procedure and provide theoretical justifications. Numerical studies show that the proposed test provides adequate control of FWER with competitive power. We demonstrate that the proposed procedures are computationally efficient, taking much less time than those methods that require intensive resampling or double bootstrap (McKeague and Qian, 2015; Wang et al., 2018a; Tang et al., 2018). Different from the proposed maximum-type statistic,

another common approach to determine group-wise significance is the combination test, for instance, the Cauchy combination test (Liu and Xie, 2019, CCT), which combines the P-values obtained from individual test of each covariate into a single P-value to assess the group-wise significance. However, our simulation studies show that CCT tends to be conservative in high dimensions.

Besides the nice properties presented in the above paragraph, the proposed test can be used as a stopping rule in forward selection, where in each step the pre-selected set is treated as the conditioning set. Under the setting of high-dimensional covariates, penalization and variable screening methods are commonly used to select significant covariates. For example, Wu and Liu (2009), Belloni and Chernozhukov (2011), Peng and Wang (2015) and others, proposed penalized variable selection methods in quantile regression. Zhao and Li (2015) proposed a score-test-based variable screening method, while Li et al. (2015) and Ma et al. (2017) proposed screening methods based on the quantile partial correlation. The screening and penalized selection methods can only tell us whether one covariate is selected or not, while the proposed method can assess the significance of the covariate by providing a P-value that is more informative.

The rest of the paper is organized as follows. In Section 2, we describe the proposed conditional marginal score-type test, present the asymptotic properties under the null and local alternative hypotheses, and introduce the multiplier

bootstrap method. In Section 3, the finite sample performance of the proposed test is assessed through simulation studies. In Section 4, we apply the proposed method to the motivating GWAS data with GFR outcomes. In Section 5, we conclude the paper and give some discussions. Some additional simulation results and all technical proofs are provided in the online Supplementary Materials.

2. Conditional maximum-score test

2.1 Model settings

Let $\{(Y_i, \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}), i = 1, \ldots, n\}$ be independent and identical copies of the triplet $(Y, \mathbf{Z}, \mathbf{X})$. Let $Q_\tau(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot})$ be the conditional τ-th quantile of Y_i given $\{\mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}\}$. We assume the following linear quantile regression model:

$$Q_\tau(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}) = \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z},0}(\tau) + \mathbf{X}_{i\cdot}^\top \boldsymbol{\beta}_{\mathbf{X},0}(\tau),\ i = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau) = (\alpha_{1,0}(\tau), \ldots, \alpha_{q,0}(\tau))^\top$ and $\boldsymbol{\beta}_{\mathbf{X},0}(\tau) = (\beta_{1,0}(\tau), \ldots, \beta_{d_n,0}(\tau))^\top$ are the quantile-specific coefficient vectors of $\mathbf{Z}$ and $\mathbf{X}$, respectively. We are interested in testing the existence of association between $\mathbf{X}$ and the τ-th quantile of Y, after accounting for the effect of $\mathbf{Z}$, that is, testing

$$H_0 :\ \boldsymbol{\beta}_{\mathbf{X},0}(\tau) = \mathbf{0}_{d_n} \ \text{ versus } \ H_a :\ \boldsymbol{\beta}_{\mathbf{X},0}(\tau) \neq \mathbf{0}_{d_n}. \tag{2.2}$$

The testing of (2.2) can be viewed as a first step in GWAS to assess the overall significance of a gene set, and if H_0 is rejected, a second step can be conducted to identify important SNPs in the gene set.

2.2 Proposed test statistic

We define

$$\varepsilon_i(\tau) = Y_i - Q_\tau(Y_i|\mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}) = Y_i - \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z},0}(\tau) - \mathbf{X}_{i\cdot}^\top \boldsymbol{\beta}_{\mathbf{X},0}(\tau), \tag{2.3}$$

so that $Q_\tau\{\varepsilon_i(\tau)|\mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}\} = 0$. We let $\mathbb{X} = (\mathbf{X}_{1\cdot}, \ldots, \mathbf{X}_{n\cdot})^\top$, $\mathbb{Z} = (\mathbf{Z}_{1\cdot}, \ldots, \mathbf{Z}_{n\cdot})^\top$, $\mathbb{X}_{\cdot j} = (X_{1,j}, \ldots, X_{n,j})^\top$, $j = 1, \ldots, d_n$, and $\mathbf{f}_\tau = \text{diag}(f_{1,\tau}(0), \ldots, f_{n,\tau}(0))$, where $f_{i,\tau}(\cdot)$ is the density of $\varepsilon_i(\tau)|\{\mathbf{X}_{i\cdot}, \mathbf{Z}_{i\cdot}\}$. To detect the significance of $\mathbf{X}$ in the presence of $\mathbf{Z}$, we construct a score-type test statistic as follows.

First, we estimate the marginal effect of $\mathbf{Z}$ as

$$\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau) = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}),$$

where $\rho_\tau(t) = t\{\tau - I(t < 0)\}$ is the quantile check loss function. To evaluate the additional effect of each X_j conditional on $\mathbf{Z}$, we project X_j on $\mathbf{Z}$ with weights $\mathbf{f}_\tau$ to obtain

$$\mathbb{X}^*_{\cdot j,\tau} = \left\{\mathbf{I}_n - \mathbf{f}_\tau \mathbb{Z}(\mathbb{Z}^\top \mathbf{f}_\tau^2 \mathbb{Z})^{-1} \mathbb{Z}^\top \mathbf{f}_\tau\right\} \mathbb{X}_{\cdot j} \doteq (X^*_{1,j,\tau}, \ldots, X^*_{n,j,\tau})^\top, \tag{2.4}$$

so that the j-th component of $\mathbf{X}$ is orthogonal to $\mathbf{Z}$ in a weighted manner, i.e., $\mathbb{Z}^\top \mathbf{f}_\tau \mathbb{X}^*_{\cdot j,\tau} = \mathbf{0}, j = 1, \ldots, d_n$. We consider the weighted projection to account for the heteroscedasticity through $f_{i,\tau}(\cdot)$, to eliminate the first order difference; see the proof of Theorem 1 in the Supplementary Materials (equation (S.16), Section S3.2) for more details. Similar projections can also be found in the quantile literature, for instance, Park and He (2017).

Second, we define the rescaled conditional marginal score statistic as

$$S_{\tau,j}(\boldsymbol{\alpha}_{\mathbf{Z}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^*_{i,j,\tau} \psi_\tau(Y_i - \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z}}) / \{\tau(1-\tau)\|\mathbb{X}^*_{\cdot j,\tau}\|^2/n\}^{1/2},\ j = 1, \ldots, d_n,$$

where $\psi_\tau(t) = \tau - I(t < 0)$. The score statistic $S_{\tau,j}(\boldsymbol{\alpha}_{\mathbf{Z}})$ is the rescaled negative subgradient of $\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z}} - \beta_j X^*_{i,j,\tau})$ with respect to β_j evaluated at $\beta_j = 0$, which captures the association between the j-th component of $\mathbf{X}$ and the signs of the quantile residuals after accounting for the effect of $\mathbf{Z}$.

Finally, the proposed maximum-score test statistic is defined as

$$\begin{aligned} T_{n,1}(\tau) &= \max_{1 \le j \le d_n} S^2_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\} \\ &= \max_{1 \le j \le d_n} \Big[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^*_{i,j,\tau} \psi_\tau\{Y_i - \mathbf{Z}_{i\cdot}^\top \widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\}\Big]^2 / \{\tau(1-\tau)\|\mathbb{X}^*_{\cdot j,\tau}\|^2/n\}. \end{aligned} \quad (2.5)$$

In practice, $\mathbf{f}_\tau$ is unknown and has to be estimated and plugged in. We propose to estimate $\mathbf{f}_\tau$ by the quotient method (Siddiqui, 1960), i.e.,

$$\widehat{f}_{i,\tau}(0) = \frac{2h}{\widehat{Q}_{\tau+h}(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}) - \widehat{Q}_{\tau-h}(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot})}, \quad (2.6)$$

and $\widehat{\mathbf{f}}_\tau = \mathrm{diag}(\widehat{f}_{1,\tau}(0), \ldots, \widehat{f}_{n,\tau}(0))$, where $\widehat{Q}_\tau(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}) = (\mathbf{Z}_{i\cdot}^\top, \mathbf{X}_{i\cdot}^\top)\widehat{\boldsymbol{\theta}}(\tau)$, and $\widehat{\boldsymbol{\theta}}(\tau)$ is the L_1-penalized estimator of $\boldsymbol{\theta}_0(\tau) = (\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)^\top, \boldsymbol{\beta}_{\mathbf{X},0}(\tau)^\top)^\top$ (Belloni and Chernozhukov, 2011). The bandwidth h is specified by the "bandwidth.rq" function of the R package *quantreg*. By the proofs in Section S3 of the Supplementary Materials, the effect of the plug-in estimator $\widehat{\mathbf{f}}_\tau$ can be ignored asymptotically, thus we ignore the difference between $\mathbf{f}_\tau$ and $\widehat{\mathbf{f}}_\tau$ for the ease of presentation, but we need to be aware of the finite-sample difference.

The test statistic $T_{n,1}(\tau)$ can be simplified in the special homoscedastic case such that $f_{i,\tau}(\cdot) \equiv f_\tau(\cdot)$ for some $f_\tau(\cdot)$, that is, the errors $\varepsilon_i(\tau)$ have a common distribution that does not depend on covariates. In this case, $\mathbf{f}_\tau(\cdot)$ cancels out in the expression (2.4), and the test statistic $T_{n,1}(\tau)$ reduces to

$$T_{n,2}(\tau) = \max_{1\le j\le d_n} \widetilde{S}^2_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\},$$

where $\widetilde{S}_{\tau,j}(\boldsymbol{\alpha}_{\mathbf{Z}}) = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n} X^*_{i,j}\psi_\tau(Y_i - \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z}})/\{\tau(1-\tau)\|\mathbb{X}^*_{\cdot j}\|^2/n\}^{1/2}$,

with $\mathbb{X}^*_{\cdot j} = \{\mathbf{I} - \mathbb{Z}(\mathbb{Z}^\top\mathbb{Z})^{-1}\mathbb{Z}\}\mathbb{X}_{\cdot j} \doteq (X^*_{1,j},\ldots,X^*_{n,j})^\top$. Note that the score function $\widetilde{S}_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\}$ used to construct the test statistic $T_{n,2}(\tau)$ is the same as the sample quantile partial correlation between Y and X_j given $\mathbf{Z}$ as defined inMa et al. (2017). The test statistic $T_{n,2}(\tau)$ has a simpler form and it does not depend on the unknown density function. In the low-dimensional quantile regression setting, it was known that the score test assuming homoscedastic errors still performs competitively well when the homoscedasticity assumption is violated; see Wang and Fygenson (2009) and Park and He (2017). We shall show in Section 3 that the proposed test based on $T_{n,2}(\tau)$ is also robust against the violation of homoscedasticity in the high-dimensional setting.

2.3 Asymptotic properties under the null

In this section, we present the asymptotic properties of $T_{n,k}(\tau), k = 1, 2$ under the null hypothesis. We define the partial correlation matrix of $\mathbf{X}$ conditional

on $\mathbf{Z}$, weighted by the density matrix $\mathbf{f}_\tau$, as $\mathbf{R}_{\tau,\mathbf{X}|\mathbf{Z}} = \text{corr}(\mathbf{X}^*_{i,\tau}) = (r_{j,l})_{j,l=1}^{d_n}$, where $\mathbf{X}^*_{i,\tau} = (X^*_{i,1,\tau}, \ldots, X^*_{i,d_n,\tau})^\top$. Under the special case of homoscedastic errors, $\mathbf{R}_{\tau,\mathbf{X}|\mathbf{Z}} = \text{corr}(\mathbf{X} \mid \mathbf{Z})$. We assume the following conditions, where $C_k, k = 1, \ldots, 5$ are some positive constants.

A1. (i) The dimension of $\mathbf{Z}$, q is fixed; (ii) the dimension of $\mathbf{X}$, $\log(d_n) = o\{n^{1/4}/\log(n)^{3/4}\}$; (iii) $E(X_j) = 0$, and X_j is sub-Gaussian, i.e. $E[\exp\{C_1 X_j^2/\text{var}(X_j)\}] \le C_2$, $j = 1, \ldots, d_n$.

A2. For $\mathbf{R}_{\tau,\mathbf{X}|\mathbf{Z}} = (r_{j,l})_{j,l=1}^{d_n}$: (i) $C_3^{-1} \le \lambda_{\min}(\mathbf{R}_{\tau,\mathbf{X}|\mathbf{Z}}) \le \lambda_{\max}(\mathbf{R}_{\tau,\mathbf{X}|\mathbf{Z}}) \le C_3$; (ii) $\max_{1\le j<l\le d_n} |r_{j,l}| \le r_0 < 1$ for some constant $0 < r_0 < 1$; (iii) $\max_{1\le j\le d_n} \sum_{l=1}^{d_n} r_{j,l}^2 \le C_4$.

A3. The density function $f_{i,\tau}(\cdot)$ and its derivative $f'_{i,\tau}(\cdot)$ are continuous and bounded from above, and $f_{i,\tau}(0)$ is bounded away from zero, $i = 1, \ldots, n$, uniformly in n.

A4. Let h_n^* be some positive sequence satisfying $n^{1/5} h_n^* \ge C_5$. For $\nu \in [\tau - h_n^*, \tau + h_n^*]$, assume that $Q_\nu(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}) = (\mathbf{Z}_{i\cdot}^\top, \mathbf{X}_{i\cdot}^\top)\boldsymbol{\theta}_0(\nu)$, where $s_n = \max_{\nu\in[\tau-h_n^*,\tau+h_n^*]} \|\boldsymbol{\theta}_0(\nu)\|_0$ is bounded, and $Q_\nu(Y_i \mid \mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot})$ is smooth in ν and has bounded third derivative with respect to ν.

Condition A1 (i) requires the dimension of $\mathbf{Z}$ to be fixed, which is for technical convenience and also practically reasonable in the GWAS. We can relax

this condition by allowing q to diverge slowly, and a possible relaxation is that $h_n^{*-1}(q+s_n)\sqrt{\log(p_n \vee n)/n} \to 0$, which is required in Lemma S.1. Our Lemma S.2 is based on fixed q, and more careful investigation is needed for diverging q. Conditions A1 (ii) and (iii) are about the dimension and distribution of $\mathbf{X}$, which are standard in high-dimensional settings. Condition A3 is an assumption on the density function that is standard in quantile regression. Condition A4 ensures that $\mathbf{f}_\tau$ can be consistently estimated; see Lemma S.1 in the Supplementary Materials for more details. Now we discuss the condition A2. By the assumption that X_j is centralized, under H_0, we have

$$\begin{aligned} & \operatorname{corr}(S_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\}, S_{\tau,l}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\} \mid \mathbb{Z}, \mathbb{X}) \\ = & \operatorname{corr}\Big[\frac{\sum_{i=1}^n X^*_{i,j,\tau}\psi_\tau\{\varepsilon_i(\tau)\}}{\{\tau(1-\tau)\|\mathbb{X}^*_{\cdot j,\tau}\|^2\}^{1/2}}, \frac{\sum_{i'=1}^n X^*_{i',l,\tau}\psi_\tau\{\varepsilon_{i'}(\tau)\}}{\{\tau(1-\tau)\|\mathbb{X}^*_{\cdot l,\tau}\|^2\}^{1/2}} \mid \mathbb{Z}, \mathbb{X}\Big] \\ = & \frac{\sum_{i=1}^n X^*_{i,j,\tau}X^*_{i,l,\tau}}{\|\mathbb{X}^*_{\cdot j,\tau}\|\|\mathbb{X}^*_{\cdot l,\tau}\|} = r_{j,l} + O_p(n^{-1/2}). \end{aligned}$$

Let $\mathbf{S}_\tau\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\} = (S_{\tau,1}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\}, \ldots, S_{\tau,d_n}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\})^\top$, then $\operatorname{corr}[\mathbf{S}_\tau\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\} \mid \mathbb{Z}, \mathbb{X}] = \mathbf{R}_{\tau,\mathbf{X}|\mathbf{Z}} + O_p(n^{-1/2})$, where the convergence rate $O_p(n^{-1/2})$ is componentwise. That is, conditions A2 (i)-(iii) are essentially imposed on the score functions under the null hypothesis, which are analogous to conditions 1, 3 and that in Lemma 6 of Cai et al. (2014). Conditions A2 (i)-(ii) are mild, while A2 (iii) is needed to control the number of positively correlated covariates, which is a key condition in the proof of the asymptotic results.

Theorem 1 presents the asymptotic null distribution of $T_{n,1}(\tau)$.

Theorem 1. *Suppose that conditions A1-A4 hold. Then for any $x \in \mathbb{R}$, we have*

$$P[T_{n,1}(\tau) - 2\log(d_n) + \log\{\log(d_n)\} \leq x \mid H_0] \to \exp\{-\pi^{-1/2}\exp(-x/2)\},$$

as $n, d_n \to \infty$.

The proof of Theorem 1 consists of two parts, where the first part is to control $\max_{1\leq j\leq d_n} |S_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\} - S_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\}|$, and the second part is to derive the asymptotic distribution of $\max_{1\leq j\leq d_n} S^2_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\}$. The derivation of the first part is challenging because the asymptotic difference between $\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)$ and $\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)$ is reflected through the indicator function, and we overcome this challenge by applying the Hoeffding inequality and a chaining argument as in Lemma A.2 of Wang and He (2007). We prove the second part by using the fact that for each $j \in \{1, \ldots, d_n\}$, $\#\{l : |\text{corr}[S_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau), S_{\tau,l}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)]| > d_n^{-\gamma_0}\}$ is well controlled by A2 (iii) for some $\gamma_0 > 0$, which is similar to the proof of Theorem 6 in Cai et al. (2014).

By Theorem 1, we can reject the null hypothesis at the significance level γ if $T_{n,1}(\tau) > 2\log(d_n) - \log\{\log(d_n)\} + q_\gamma$, where $q_\gamma = -\log(\pi) - 2\log\{\log(1-\gamma)^{-1}\}$. Alternatively we can calculate the P-value as

$$1 - \exp\Big(-\pi^{-1/2}\exp\Big[-\frac{T_{n,1}(\tau) - 2\log(d_n) + \log\{\log(d_n)\}}{2}\Big]\Big).$$

For the homoscedastic case, we have the following corollary.

Corollary 1. *Assume that $f_{i,\tau}(\cdot) \equiv f_\tau(\cdot)$ for some $f_\tau(\cdot)$ across i, and conditions A1-A3 hold. Then for any $x \in \mathbb{R}$, we have*

$$P[T_{n,2}(\tau) - 2\log(d_n) + \log\{\log(d_n)\} \le x \mid H_0] \to \exp\{-\pi^{-1/2}\exp(-x/2)\},$$

as $n, d_n \to \infty$.

2.4 Asymptotic properties under the local alternative

In this section, we study the asymptotic properties of $T_{n,k}(\tau), k = 1, 2$ under the local alternative,

$$\begin{aligned} H_a: \ Q_\tau(Y_i|\mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}) &= \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z},0}(\tau) + \mathbf{X}_{i\cdot}^\top \boldsymbol{\beta}_{\mathbf{X},n}(\tau), \ \ i = 1, \ldots, n, \\ \boldsymbol{\beta}_{\mathbf{X},n}(\tau) &= \mathbf{b}_0(\tau)\sqrt{\log(d_n)/n}, \end{aligned} \tag{2.7}$$

where $\mathbf{b}_0(\tau) = (b_{1,0}(\tau), \ldots, b_{d_n,0}(\tau))^\top$. We assume that the number of nonzero components in $\mathbf{b}_0(\tau)$, denoted as $s_0(\tau)$, is fixed. Without loss of generality, we assume that the first $s_0(\tau)$ components of $\mathbf{b}_0(\tau)$ are nonzero.

To establish the asymptotic property of the test statistics under (2.7), we make an additional assumption, with a discussion in Section S2.

A5. Let $\omega^*_{j,l,\tau} = E\{f_{i,\tau}(0)X^*_{i,j,\tau}X^*_{i,l,\tau}\}/\{\tau(1-\tau)E(X^{*2}_{i,j,\tau})\}^{1/2}$. Assume that $\max_{1\le j\le d_n} |\sum_{l=1}^{s_0(\tau)} b_{l,0}(\tau)\omega^*_{j,l,\tau}| > \sqrt{2} + \epsilon$ for some positive constant ϵ.

Theorem 2. *Assume that conditions A1-A5 hold, and $s_0(\tau)$ is fixed. Under the local alternative (2.7), for any $\gamma > 0$, we have*

$$P[T_{n,1}(\tau) - 2\log(d_n) + \log\{\log(d_n)\} > q_\gamma \mid H_a] \to 1, \ \text{as } n, d_n \to \infty.$$

Since $\sqrt{\log(d_n)/n}$ is the optimal convergence rate that can be obtained in high-dimensional settings (Belloni and Chernozhukov, 2011), Theorem 2 indicates that the proposed test is asymptotically sharp.

2.5 Multiplier bootstrap

The asymptotic results in Theorem 1 and Corollary 1 provide a simple calibration method for the proposed maximum-score test statistic. Our preliminary results show that this asymptotic calibration performs well for large samples, but it tends to be conservative in finite samples. To achieve better finite sample performance, we propose an alternative calibration method based on a multiplier bootstrap procedure. The idea of multiplier bootstrap was also considered in other settings for low-dimensional data, e.g. He and Zhu (2003), Zhang et al. (2014), Horowitz (2019). We shall show that the proposed multiplier bootstrap method is computationally convenient and theoretically valid under the high-dimensional setting. Below we describe the procedure for the test statistic $T_{n,1}(\tau)$, which covers $T_{n,2}(\tau)$ as a special case.

Step 1. Let

$$T_{n,1}(\tau)^* = \max_{1\le j\le d_n} \Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i X^*_{i,j,\tau}\psi_\tau(e_i)\Big\}^2 / \Big\{\tau(1-\tau)\|\mathbb{X}^*_{\cdot j,\tau}\|^2/n\Big\},$$

where $\{e_i; i = 1,\dots,n\}$ is a random sample with the τ-th quantile zero, and $\{w_i; i = 1,\dots,n\}$ is a random sample independent of e_i with zero mean, unit variance and a finite third moment. We generate e_i from $N(-\Phi^{-1}(\tau), 1)$ and w_i

from a two-point distribution with $P(w = 1) = P(w = -1) = 1/2$.

Step 2. Repeat Step 1 M times to obtain bootstrap statistics $\{T_{n,1}(\tau)^{*1}, \ldots, T_{n,1}(\tau)^{*M}\}$, and calculate the P-value as $M^{-1}\sum_{b=1}^{M} I\{T_{n,1}(\tau)^{*b} > T_{n,1}(\tau)\}$.

Unlike the conventional bootstrap methods, the multiplier bootstrap does not require reanalyzing the data repeatedly and thus is computationally efficient. An intuitive justification is given by (S.17) in Section S3.2, where we show

$$\begin{aligned} & S_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\} = S_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\} + O_p\left\{n^{-1/4}(\log n)^{3/4}\right\} \\ = & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X^*_{i,j,\tau}\psi_\tau\{\varepsilon_i(\tau)\}/\{\tau(1-\tau)\|\mathbb{X}^*_{\cdot,j,\tau}\|^2/n\}^{1/2} + O_p\left\{n^{-1/4}(\log n)^{3/4}\right\} \end{aligned}$$

under the null hypothesis. Theorem 3 provides the theoretical justification for the multiplier bootstrap method in the high-dimensional setting. Similar results can also be obtained under conditions A1-A3 for the homoscedastic case.

Theorem 3. *Suppose that conditions A1-A4 hold. Then for any $x \in \mathbb{R}$, we have*

$$P^D\left[T_{n,1}(\tau)^* - 2\log(d_n) + \log\{\log(d_n)\} \le x | H_0\right] \to \exp\{-\pi^{-1/2}\exp(-x/2)\},$$

as $n, d_n \to \infty$, where the superscript D means conditional on the observed data $\{(Y_i, \mathbf{Z}_{i\cdot}, \mathbf{X}_{i,\cdot}), i = 1, \ldots, n\}$.

2.6 Forward selection via a sequential conditional test

The proposed conditional maximum-score test aims to assess the overall significance of $\mathbf{X}$. If the test leads to the rejection of H_0, indicating that at least

one component of $\mathbf{X}$ is associated with the τth quantile of Y after accounting for the effect of $\mathbf{Z}$, the next natural question is to identify those important variables. The proposed test can be used as a stopping rule in forward regression to discover significant components in $\mathbf{X}$. To account for multiple testing in the sequential procedure, we follow a similar two-stage selection as in Tang et al. (2018).

In the first stage, we initialize the forward regression by sequentially applying the proposed test. Specifically, we perform the conditional marginal test with $\mathbf{X}^{(0)} = \mathbf{X}$ and $\mathbf{Z}^{(0)} = \mathbf{Z}$. Let $\widehat{j}_1$ be the index of the predictor in $\mathbf{X}^{(0)}$ that gives the largest squared conditional marginal-score statistic and P_1 be the associated P-value. If $P_1 > \gamma$, the pre-specified significance level, we stop and declare that there is no significant X_j's. Otherwise, we move $X_{\widehat{j}_1}$ from $\mathbf{X}^{(0)}$ to $\mathbf{Z}^{(0)}$ and repeat the procedure until no more significant predictors are detected. Assume that the selected covariate set is $\mathbf{Z}^{(K)} = \{\mathbf{Z}, X_{\widehat{j}_1}, \ldots, X_{\widehat{j}_K}\}$, with associated P-values as $\{P_1, \ldots, P_K\}$. In the second stage, we perform multiple test adjustment. Suppose that $K \geq 1$. Define $K^* = 1$ if $P_1 > \gamma/K$, otherwise $K^* = \max_{1 \leq k \leq K}\{k : P_l \leq \gamma/(K-l+1), l = 1, \ldots, k\}$, and the finally selected covariate set is chosen as $\mathbf{Z}^{(K^*)} = \{\mathbf{Z}, X_{\widehat{j}_1}, \ldots, X_{\widehat{j}_{K^*}}\}$.

While it is challenging to establish a formal theoretical justification for the proposed two-stage method due to its sequential nature, our numerical studies in Section 3 show that the method performs well in terms of both false positives

and false negatives for modest and large samples.

3 Simulation study

3.1 Size and power study

We generate the simulation data from the following model,

$$Y_i = \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z},0} + \mathbf{X}_{i\cdot}^\top \sqrt{\log(d_n)/n}\mathbf{b}_0 + (1 + a_0 X_{i,1})\varepsilon_i,\ i = 1, \ldots, n,$$

where $\mathbf{Z}_{i\cdot} = (1, \widetilde{\mathbf{Z}}_{i\cdot}^\top)^\top$, a_0 is the parameter controlling the heterogeneity of the noises, $\boldsymbol{\alpha}_{\mathbf{Z},0} = \mathbf{1}_{q=6}$, $\mathbf{b}_0 = \mathbf{0}_{d_n=p_n-q}$ under H_0 and $\mathbf{b}_0 = \delta(1, 0.8, 0.6, 0.4, 0.2, 0_{d_n-5}^\top)^\top$ under H_a. We let $\delta \in (0, \delta_{\max,p_n}]$ for some pre-specified $\delta_{\max,p_n}$.

We consider three cases to examine the performance of the proposed test. In Case 1, $(\widetilde{\mathbf{Z}}_{i\cdot}^\top, \mathbf{X}_{i\cdot}^\top)^\top \sim N(\mathbf{0}, \mathbf{I}_{(p_n-1)\times(p_n-1)})$ and $\varepsilon_i \sim N(0,1)$ with $a_0 = 0$. In Case 2, $(\widetilde{\mathbf{Z}}_{i\cdot}^\top, \mathbf{X}_{i\cdot}^\top)^\top \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\sigma_{l,l'})_{l,l'=1,\ldots,p_n-1}$, $\sigma_{l,l'} = 0.5^{|l-l'|}$ and $\varepsilon_i \sim t_3$ with $a_0 = 0$. In Case 3, non-Gaussian regressors with heteroscedastic errors are considered. Specifically, we first generate $\mathbf{U}_{i\cdot} = (U_{i,1}, \ldots, U_{i,p_n-1})^\top \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the same as in case 2, and then let $\widetilde{Z}_{i,l} = 2\sqrt{3}\Phi(U_{i,l}) - \sqrt{3}$ for $l = 1, \ldots, 5$, and $X_{i,l-5} = 2\sqrt{3}\Phi(U_{i,l}) - \sqrt{3}$ for $l = 6, \ldots, p_n - 1$. Furthermore, we let $\varepsilon_i \sim t_3$ with $a_0 = 1/2$. Therefore, in this heteroscedastic case, the true quantile coefficient of $X_{i,1}$ is $\beta_{1,0}(\tau) = b_{0,1} + a_0 F_{t_3}^{-1}(\tau)$, which is nonzero and thus corresponds to the alternative model for all $\tau \neq 0.5$ even when $\mathbf{b}_0 = \mathbf{0}_{d_n}$. For all cases, we consider $p_n = 10, 50, 200, 1000$ and $n = 200, 800$, and set the nominal level as $\gamma = 0.05$, and the number of rep-

etitions in the multiplier bootstrap method as $M = 500$. We also consider a case to mimic the motivating GFR study in Section S1.2, and the main observations are similar to Cases 1-3.

The following tests are compared: (i) four variations of the proposed test, $T^E_{n,1}(\tau), T^B_{n,1}(\tau), T^E_{n,2}(\tau), T^B_{n,2}(\tau)$, where the superscript indicates using the asymptotic extreme value distribution (E) or the multiplier bootstrap procedure (B) to obtain the critical value; (ii) RS, the regularized rank score test in Park and He (2017), with $p_n < n$; (iii) QME, the quantile marginal effect test from Wang et al. (2018a) with the tuning parameter set as $\lambda_n = 3\sqrt{\tau(1-\tau)\log n}$; (iv) BON, Bonferroni adjustment method, where the individual P-values are based on $S_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\}$ and its asymptotic normality, $j = 1, \ldots, d_n$, i.e., the proposed conditional marginal rank score statistics for heteroscedastic cases; (v) CCT, Cauchy combination test in Liu and Xie (2019), where the individual P-values are the same as in BON; (vi) CAR, the conditional adaptive resampling test in Tang et al. (2018) for the mean model, with the tuning parameter set as $\lambda_n = \max\left[3(\log n)^{1/2}, \Phi^{-1}\left\{1-\gamma/(2d_n)\right\}\right]$; (vii) GC, the partial test in Guo and Chen (2016) for the mean model, which is based on a sum-squared-type U-statistic. The number of bootstraps is set as 500 for both QME and CAR. In Wang et al. (2018a) and Tang et al. (2018), the tuning parameter is selected by double bootstrap, which is computationally intensive, so we fix the parameter

at one value that performs relatively better to save the computation time. Table S.1 in the Supplementary Materials summarizes the average computing time for each method. Results show that the methods that do not require the estimation of $\mathbf{f}_\tau$, namely RS, $T_{n,2}(\tau)$ and GC, are computationally more efficient than those do, namely $T_{n,1}(\tau)$, BON and CCT. In addition, the resampling-bootstrap-based methods QME and CAR are computationally much more expensive than all the other methods, even if double bootstrap is not used for the tunings.

Table 3.1 summarizes the rejection percentages of different methods in Cases 1 and 3 with $\mathbf{b}_0 = \mathbf{0}$; empirical sizes from Case 2 are similar to those from Case 1, thus are moved to Table S.2 in the Supplementary Materials. In all scenarios but Case 3 with $\tau = 0.25$, the null hypothesis is true so the rejection rate corresponds to the empirical size; while in Case 3 with $\tau = 0.25$, $\beta_{1,0}(\tau) = \frac{1}{2}F_{t_3}^{-1}(0.25)$, thus the rejection rate corresponds to the power.

Under the null model, all four variations of the proposed test result in Type I errors close to the nominal level, but the tests based on the asymptotic critical values are slightly more conservative, especially for $n = 200$. Even though the test based on $T_{n,2}(\tau)$ assumes homoscedastic errors, the method still performs competitively well in the heteroscedastic Case 3 in terms of both Type I error and power, and it is computationally much simpler than the test based on $T_{n,1}(\tau)$. The RS performs well for small p_n, but it becomes quite conservative for larger

p_n and it is not applicable when $p_n \geq n$. The QME is sensitive to the choice of the tuning parameter; it gives deflated Type I error in most scenarios but inflated Type I errors at $\tau = 0.25$ for $n = 200$ and $p_n = 1000$. The BON and CCT control the Type I errors reasonably well, but in heteroscedastic Case 3 with $\tau = 0.25$, they are both more conservative than the proposed multiplier bootstrap method for detecting signals, especially for $n = 200$. Finally, the mean-based tests CAR and GC perform well in the homoscedastic cases, but they are not able to detect the signal at tail quantiles caused by the heteroscedasticity as seen in Case 3.

The limited performance of QME is probably caused by three reasons. First, the theory of QME only works for fixed-dimensional covariates. Second, QME is proposed for the marginal test. When adapting it to the conditional test, the method treats quantile residuals obtained from regressing Y on $\mathbf{Z}$ as the new response and then applies marginal test over $\mathbf{X}$. This may lead to inflated error rate if components in $\mathbf{Z}$ and $\mathbf{X}$ are highly correlated, which is often seen in the high-dimensional settings, due to the spurious correlation in the sample. Third, the tuning parameter λ_n is chosen using the same rule of thumb across simulations and thus is not data adaptive. Its performance may be improved by using double bootstrap to select a data-adaptive λ_n, but the computation is heavily intensive and not practical for large p_n.

To compare the power of different tests, we focus on $\tau = 0.5$ and $n =$

Table 3.1: Rejection percentages of different tests for Cases 1 & 3 with $\mathbf{b}_0 = \mathbf{0}$. All scenarios correspond to the null model, except $\tau = 0.25$ in Case 3.

			Case 1								Case 3							
			n=200				n=800				n=200				n=800			
location	method	p_n	10	50	200	1000	10	50	200	1000	10	50	200	1000	10	50	200	1000
$\tau = 0.25$	$T^E_{n,1}(\tau)$		3.1	3.5	4.6	3.1	2.4	4.6	5.3	4.0	48.9	25.5	15.5	8.2	99.7	99.0	97.6	94.1
	$T^B_{n,1}(\tau)$		5.2	5.1	5.4	4.3	4.7	5.6	6.5	5.5	60.3	30.2	17.6	10.2	99.9	99.2	98.0	94.4
	$T^E_{n,2}(\tau)$		2.7	3.6	4.8	3.4	2.5	4.6	5.4	4.1	57.2	33.6	20.9	11.5	100.0	99.7	99.0	96.9
	$T^B_{n,2}(\tau)$		5.5	5.0	6.1	4.9	4.6	5.8	6.6	5.3	66.1	38.6	23.7	14.5	100.0	99.7	99.1	97.1
	RS		4.7	2.7	/	/	4.0	3.7	3.4	/	56.8	9.6	/	/	99.9	87.1	27.5	/
	QME		2.0	2.7	7.5	14.7	3.4	3.1	3.7	5.9	21.9	4.0	2.9	5.5	99.0	92.1	81.4	66.2
	BON		4.4	4.1	6.2	4.1	4.3	5.3	4.9	4.6	57.1	28.8	17.0	9.1	99.9	99.0	97.9	94.3
	CCT		2.4	2.1	3.0	1.7	1.8	3.3	2.7	2.8	48.1	20.9	11.4	6.1	99.6	98.8	96.6	91.5
$\tau = 0.5$	$T^E_{n,1}(\tau)$		2.4	3.8	4.9	4.2	3.7	3.0	4.4	3.8	3.5	3.2	3.3	4.2	4.3	4.1	3.6	3.5
	$T^B_{n,1}(\tau)$		5.2	6.2	7.6	5.9	5.8	4.8	5.2	4.6	6.0	4.3	5.4	5.9	7.2	5.5	4.6	5.1
	$T^E_{n,2}(\tau)$		2.5	3.8	5.0	4.2	3.7	3.0	4.5	3.8	3.7	3.7	3.7	4.6	4.6	3.9	3.6	4.3
	$T^B_{n,2}(\tau)$		5.1	5.8	7.7	5.8	5.6	4.4	5.4	4.8	7.2	5.1	5.3	6.5	7.6	6.5	5.1	5.6
	RS		5.1	3.9	/	/	5.2	3.9	2.6	/	5.7	3.8	/	/	7.7	6.8	1.9	/
	QME		1.6	2.1	1.8	2.3	3.8	2.0	2.3	1.7	1.4	0.6	0.5	0.2	2.1	2.0	1.3	0.7
	BON		4.6	5.2	6.2	4.5	5.1	4.2	5.0	4.3	5.4	3.8	3.9	4.5	6.8	5.3	4.4	4.3
	CCT		2.1	2.4	2.2	2.1	3.1	2.1	2.9	1.9	3.3	2.2	2.4	2.5	4.6	3.3	2.6	2.0
mean	CAR		6.2	5.3	6.0	7.3	5.5	5.1	5.3	4.5	5.3	2.3	4.0	3.1	4.1	3.8	3.9	3.4
	GC		7.4	5.3	6.2	6.2	7.2	5.0	6.2	5.1	4.0	5.1	6.2	5.8	4.2	4.8	5.4	6.2

$T^E_{n,k}(\tau)$ and $T^B_{n,k}(\tau)$, $k = 1, 2$: four variations of the proposed test; RS: the rank score test in Park and He (2017); QME: the quantile marginal effect test in Wang et al. (2018a); BON, Bonferroni adjustment on d_n individual P-values; CCT, Cauchy combination test in Liu and Xie (2019); CAR: the conditional adaptive resampling test in Tang et al. (2018); GC: the sum-squared-type test in Guo and Chen (2016).

200. We let the number of nonzero coefficients be $s_0(\tau) = 5$ and set $\mathbf{b}_0 = \delta(1, 0.8, 0.6, 0.4, 0.2, 0_{d_n-5}^\top)^\top$, where δ varies from 0 to $\delta_{\max,p_n}$ with $\delta_{\max,p_n} = 6$ for $p_n = 10$ and $\delta_{\max,p_n} = 5$ for $p_n = 50, 200, 1000$. In the following analysis, we exclude QME because it is difficult to control the Type I error due to its sensitivity to the choice of the tuning parameter, and we also exclude BON and CCT since they were shown in Table 3.1 to be more conservative for detecting signals in heteroscedastic cases with small samples.

Figure 3.1 presents the power curves of different methods. Both CAR and GC are designed for detecting the mean effect. The CAR method gives higher power in Case 1 with homoscedastic normal errors, but the method is less powerful for models with heavy-tailed (Case 2) and heteroscedastic (Case 3) errors. The GC test is based on a sum-squared-type test statistic so it is less powerful to detect the sparse signal in all four cases, especially for large p_n. In addition, neither CAR nor GC can identify the signal at tails as shown in Table 3.1. The rank score test (RS) performs competitively well for $p_n = 10$ but it quickly loses for larger p_n and the method does not work for cases with $p_n \geq n$. The four variations of the proposed test perform similarly, giving either competitive or higher power than the other three methods. Among the four variations, the tests based on the multiplier bootstrap tend to be more powerful than their asymptotic counterparts, and the tests based on $T_{n,2}(\tau)$ assuming homoscedastic errors are

slightly more powerful than those based on $T_{n,1}(\tau)$.

3.2 Forward selection

In this section, we assess the performance of forward selection by using the proposed test as the stopping rule. Data are generated from the following model,

$$Y_i = \mathbf{Z}_{i\cdot}^\top \boldsymbol{\alpha}_{\mathbf{Z},0} + \mathbf{X}_{i\cdot}^\top \boldsymbol{\beta}_{\mathbf{X},0} + (1 + a_0 X_{i,1})\varepsilon_i,\ i = 1, \ldots, n,$$

where $\mathbf{Z}_{i\cdot} = (1, \widetilde{\mathbf{Z}}_{i\cdot}^\top)^\top$, $\boldsymbol{\alpha}_{\mathbf{Z},0} = \mathbf{1}_6$, $\boldsymbol{\beta}_{\mathbf{X},0} = (0, 1, 1, 0.8, 0.8, 0_{p_n-11}^\top)^\top$, with $n = 200$, $p_n = 200$ and 1000; $(\widetilde{\mathbf{Z}}_{i\cdot}^\top, \mathbf{X}_{i\cdot}^\top)^\top$ and ε_i are generated as in Cases 1 and 3 with $a_0 = 0$ for Case 1 and 0.5 for Case 3, and 1000 replicates are considered, with nominal level $\gamma = 0.05$. We compare the following forward selection procedures: (i) $T_{n,1}^B(\tau)$, sequential test based on $T_{n,1}^B(\tau)$; (ii) L_1, the L_1-penalized variable selection method in Belloni and Chernozhukov (2011), without penalizing the coefficients of $\mathbf{Z}$; (iii) QPCOR-L_1, the quantile partial correlation screening in Ma et al. (2017), and we use their algorithm 3 to reduce the dimension of $\mathbf{X}$ from d_n to $n/\log n$, then followed by the L_1-penalized method in Belloni and Chernozhukov (2011); (iv) CAR, sequential test based on CAR, with the same tuning parameter over replicates as in Section 3.1. For the sequential-test-based methods $T_{n,1}^B(\tau)$ and CAR, multiple test adjustments as in Section 2.6 are applied. For the quantile based methods, we focus on $\tau = 0.5$.

In evaluating the performance of different methods, we consider the percentages of replicates in which $X_j, j = 1, \ldots, 5$ are selected (PS), the average

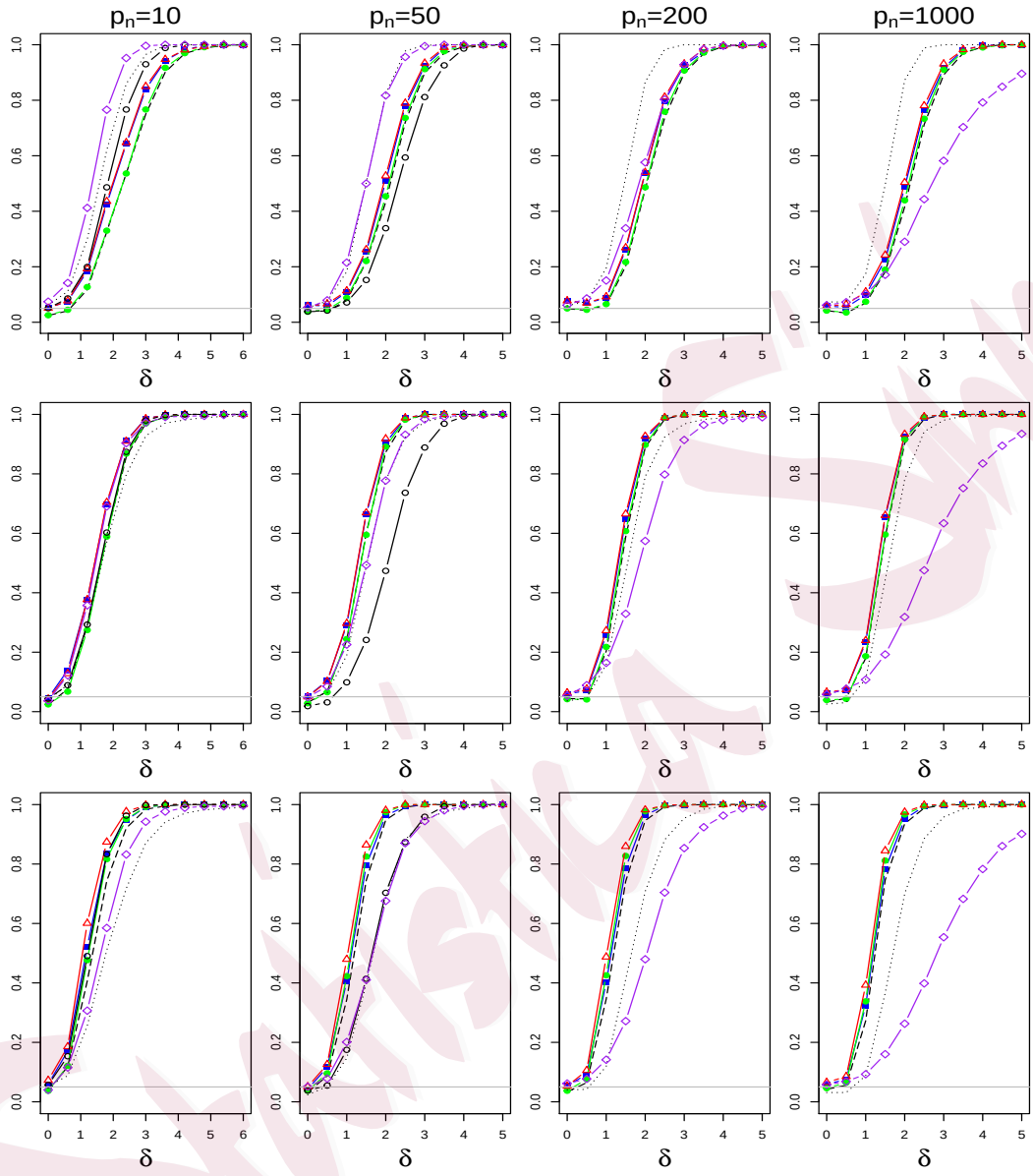

Figure 3.1: Power curves of different methods in Cases 1 (first row), 2 (second row) and 3 (third row), with $n = 200$ and $\tau = 0.5$: $T^E_{n,1}(\tau)$ (dashed), $T^B_{n,1}(\tau)$ (line with solid square), $T^E_{n,2}(\tau)$ (line with solid dots), $T^B_{n,2}(\tau)$ (line with triangle), RS (line with open circle); CAR (dotted); GC (line with diamond). The gray horizontal line stands for the nominal level of 0.05.

number of false positives (FP), percentages of replicates of under-fit (UF) in which at least one important X_j is not selected, and percentages of replicates in which the exact true model (TM) is selected. We find that: (i) the performance of $T^B_{n,1}(\tau)$ is competitive or better in all the scenarios; (ii) both L_1 and QPCOR-L_1 tend to over-fit the model (higher FP), and further steps may be applied to the selected model to refine the selection accuracy, but inherent uncertainty may be accumulated; (iii) CAR performs well when the noise is homoscedastic normal (Case 1), but the under-fit percentages (UF) can be high when the noises are heavy-tailed with heteroscedasticity (Case 3).

4. Analysis of the glomerular filtration rate

A single-nucleotide polymorphism (SNP) is a substitution of a single nucleotide that occurs at a specific position in the genome, some of which are linked to genes affecting specific phenotypes. In this section, we apply the proposed test and forward selection procedure to screen a large number of SNPs in a thorough search for mutations associated with phenotypes of interests, in the presence of some "protected" demographic covariates. Over a million SNPs are mapped in the GWAS of the Diabetes Control and Complications Trial (DCCT), a randomized clinical trial studying the effects of intensive monitoring of glucose levels on long-term microvascular complications, among Type 1 diabetes patients. The response variable of interest is the glomerular filtration rate (GFR,

Table 3.3: Forward selection results in Cases 1 and 3, $n = 200$.

Case	p_n	Method	PS					FP	UF	TM
			X_1	X_2	X_3	X_4	X_5			
1	200	$T^B_{n,1}(\tau)$	0.0	100.0	100.0	100.0	100.0	0.071	0.0	93.1
		L_1	0.0	100.0	100.0	99.6	99.9	0.085	0.5	91.5
		QPCOR-L_1	0.1	100.0	100.0	100.0	99.9	0.276	0.1	76.2
		CAR	0.0	100.0	100.0	100.0	100.0	0.058	0.0	94.4
	1000	$T^B_{n,1}(\tau)$	0.0	100.0	99.8	99.8	99.8	0.071	0.2	92.9
		L_1	0.0	99.9	100.0	97.5	98.0	0.092	4.5	87.0
		QPCOR-L_1	0.4	100.0	100.0	99.9	100.0	0.688	0.1	50.0
		CAR	0.0	100.0	100.0	100.0	100.0	0.071	0.0	93.3
3	200	$T^B_{n,1}(\tau)$	0.0	99.2	99.9	97.9	99.3	0.074	2.8	90.4
		L_1	0.6	100.0	100.0	100.0	100.0	0.112	0.0	89.9
		QPCOR-L_1	0.9	100.0	100.0	100.0	100.0	0.332	0.0	72.3
		CAR	0.1	96.4	95.1	83.2	88.4	0.045	30.0	66.3
	1000	$T^B_{n,1}(\tau)$	0.0	98.7	98.7	94.2	98.0	0.071	8.1	85.5
		L_1	0.3	100.0	100.0	100.0	100.0	0.085	0.0	91.8
		QPCOR-L_1	0.5	100.0	100.0	100.0	100.0	0.813	0.0	43.4
		CAR	0.1	93.2	92.8	75.7	83.1	0.033	43.9	54.6

$T^B_{n,1}(\tau)$: forward selection based on $T^B_{n,1}(\tau)$; L_1: the L_1-penalized variable selection method in Belloni and Chernozhukov (2011); QPCOR-L_1: the QPCOR in Ma et al. (2017); CAR: forward selection based on CAR in Tang et al. (2018). PS: percentages of being selected; FP: average number of false positives; UF: percentages of replicates in which at least one important X_j is not selected; TM: percentages of replicates in which the exact true model is selected.

measured in percentages), a well-used clinical index of overall kidney function. Although multiple GFR measurements were collected during follow-up, we are interested in the most severe status of nephropathy risks, which is usually measured by the most recent kidney functions, i.e. the GFR measurement at the last visit. The "protected" covariates includes gender, treatment, age (in

years, centered), duration of diabetes (in weeks, centered) and body mass index (BMI, centered), where the duration of diabetes measures the different stages of nephropathy development across the patients.

The GWAS of DCCT contains in total 1.18 million candidate SNPs, while the number of patients is only 1304, far less than the number of SNPs. One important statistical issue concerns assessing the overall significance of groups of SNPs, i.e. whether there exists SNPs in a set of genes that has effect on the disease, while controlling for the family-wise error rate. Most work in GWAS considered mean-based tests. However, in this study, the mean of GFR is less important clinically than the tail quantiles, because the mean values are usually driven by the majority of participants with normal kidney function, while the lower quantiles reflect the characteristics of the subset of participants with elevated risks of nephropathy. Furthermore, GFR values are skewed to the left even after logarithm transformation (Figure 4.1), thus quantile regression at several lower quantile levels could provide more clinically relevant information than the mean regression, and it also enables us to work on the original scale providing better interpretation to clinicians and patients. For these reasons, we would like to assess the significance of SNPs on lower quantiles of GFR, to identify SNPs and gene pathways associated with patients at higher risks of nephropathy, and we consider quantile levels $\tau = 0.1, 0.25$ and 0.5.

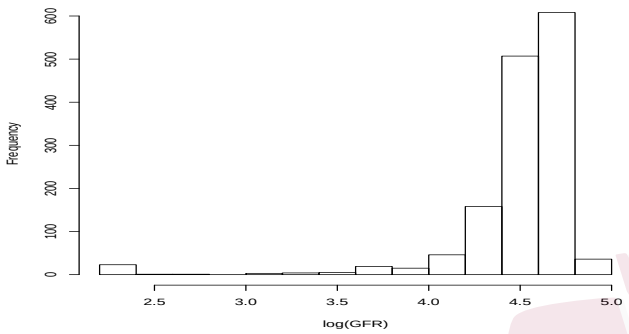

Figure 4.1: Log-transformed GFR% from the type 1 diabetes patients at the last visit in the DCCT study.

To apply the proposed method, we focus on a subset of SNPs, which belong to genes related to nephropathy in MSigDB Curated Gene Sets (`http://software.broadinstitute.org/gsea/msigdb/`), including 2908 SNPs after deleting those not satisfying the Hardy-Weinberg equilibrium (Crow, 1999). For further pre-processing, we (i) delete one female patient who has 98% of the SNPs missing; (ii) delete SNPs with any missing values; (iii) delete SNPs with minor allele frequency less than 5%; (iv) prune highly correlated SNP pairs, defined as correlation coefficients larger than 0.99. Finally, we have 1303 patients, consisting of 695 males and 608 females, and 981 SNPs. The SNPs are coded as -1, 0, 1, that is, the number of minor alleles minus 1. Previous work suggested that the risk factor mechanisms of nephropathy may be different in males and females (Silbiger and Neugarten, 2003). Therefore, we study male and female

participants separately, and the forward selection presented later shows that different sets of SNPs are identified for men and women, which further validates our stratified analysis by gender.

We first apply all the four variations of the proposed test for overall significance test at different quantiles. Conditional on $\mathbf{Z}$, all variations of the proposed test conclude that there exist significant SNPs, at $\tau = 0.1, 0.25, 0.5$, with P-values smaller than 0.0001. We also apply CAR in Tang et al. (2018) for overall significance test on the conditional mean. We consider the test with tuning parameter $\lambda_n = \max\left[a(\log n)^{1/2}, \Phi^{-1}\left\{1-\gamma/(2d_n)\right\}\right], a \in \{3,4,5,6,7\}$, and all the λ_n's lead to the same P-values, which are 0.572 and 0.324 for male and female groups, respectively, which indicates that no SNP is significantly associated with the conditional mean of the GFR.

Next we proceed to forward selection by applying $T^B_{n,1}(\tau)$ (sequentially), L_1 and QPCOR-L_1 to select the significant SNPs. To account for the randomness in the selection procedure, the covariate selection procedure is repeated in randomly selected subsets of size $0.8n$ in each gender group. No SNP is selected by either L_1 or QPCOR-L_1 in any random split, which is probably caused by weak signals and/or over penalization. Table 4.1 presents the frequencies of the SNPs which are selected at least 10 times by our method, over 50 random subsets.

For further verification, we regress Y on $\mathbf{Z}$ at the τ-th quantile, obtaining the

residuals under the null model. In general, if one SNP has effect on the response, residuals with different genotypes would have different distributions. We calculate the variance, mean and the τ-th quantile of the residuals in genotype "AA" and "Aa" for each SNP, and report the ranks of the differences among 981 SNPs in Table 4.1. We find that, most of the SNPs selected with high frequencies have high ranks in at least one of the three criteria, which provides further evidence for the impacts of the selected SNPs on the lower quantiles of GFR.

Furthermore, we searched PubMed for publications that studied the SNPs identified in our analysis as validation from external data for the functions of the reported SNPs. Specifically, rs9331949 and rs1044506 were found to be associated with dementia, epilepsy and Alzhermer's disease (Bennet et al., 2011; Du et al., 2016; Stage et al., 2016; Tan et al., 2016) and rs3830041 was found to be associated with HepB-related hepatocellular carcinoma (Yu et al., 2017). Further study of the functions of these two SNPs in dbSNP (`https://www.ncbi.nlm.nih.gov/snp`) shows that rs9331949 is involved in cell death while rs3830041 belongs to the NOTCH family which plays a role in vascular, renal and hepatic development.

Finally, we discuss the selected SNPs via gene pathways. The 11 selected SNPs in females belong to nine different genes (EML1, FAM53B, PPM1F, PTGIS, PTPRB, PTPRM, UGT2B7, UNC5B, ZCCHC24), which have significant

overlaps with three known gene sets:

BAELDE_DIABETIC_NEPHROPATHY_UP (P-value=6.48×10^{-22}),

GO_TRANSMEMBRANE_RECEPTOR_PROTEIN_PHOSPHATASE_ACTIVITY

(P-value=5.25×10^{-6}), and GO_PHOSPHOPROTEIN_PHOSPHATASE_ACTIVITY

(P-value=5.92×10^{-6}). The first set include genes up-regulated in glomeruli of kidneys from patients with diabetic nephropathy (type 2 diabetes mellitus), while the other two gene sets are related to catalysis which controls the state of phosphorylation of cell proteins and thereby provide an important mechanism for regulating cellular activity. The 15 selected SNPs in males belong to 10 different genes (ATP10B, CAPN3, CLU, FAM53B, NOTCH4, PTGIS, PTPRM, SLC6A7, TEK, ZCCHC24), which also overlaps significantly with BAELDE_DIABETIC_NEPHROPATHY_UP (P-value=1.38×10^{-24}). Furthermore, the markers selected in males also overlaps with gene set GO_CELLULAR_COMPONENT_MORPHOGENESIS (P-value=2.44×10^{-6}), which functions in the process of cellular structure generation and organization.

5. Discussion

The proposed method is based on a maximum-type test statistic, which is known to be powerful when the signals are sparse. In some studies, it is possible that the signals are weak and dense, that is, groups of markers jointly affect the phenotype, while the signal of each marker is faint. To adapt different types of

Table 4.1: Summary of SNPs selected in the random subsets.

τ	Gender	SNP	Frequency	r_{var}	r_{mean}	r_{Q_τ}
0.1	M	rs11742097_A	19	22	14	94
		rs6866731_G	12	26	15	89
	F	rs8091758_T	27	3	4	13
		rs16952201_T	13	1	95	21
0.25	M	rs2240785_C	26	3	6	5
		rs9331949_C	20	1	1	1
	F	rs2083564_A	18	902	822	741
		rs16952201_T	14	1	3	9
0.5	M	rs9331949_C	28	1	1	1
		rs2240785_C	26	5	9	5
		rs28364433_A	26	921	208	578
		rs28364475_T	26	443	5	401
		rs1982285_T	26	2	23	4
		rs3830041_T	24	386	8	158
		rs1044506_T	22	889	90	623
		rs11002952_A	21	730	955	720
		rs10050146_T	14	259	31	32
		rs7099298_A	12	18	254	10
		rs10129739_C	10	26	34	77
	F	rs11002951_T	40	249	4	87
		rs10999763_C	24	17	242	21
		rs10050146_T	22	65	11	860
		rs7099298_A	22	48	343	52
		rs2567136_T	20	792	375	60
		rs2241199_C	11	4	26	19
		rs11746151_C	10	697	360	220

Frequency: selected times among 50 random partitions; r_{var}, r_{mean}, r_{Q_τ}: rank of differences in residual variances, means and quantiles, in two genotypes "AA" and "Aa".

signals, we may consider a hybrid test statistic by taking a weighted average of the maximum- and sum-squared-type statistics as in Tang et al. (2018). However, the existing literature for sum-squared-type test requires either smoothed loss

functions (Guo and Chen, 2016), or limited dimensionality of markers (Park and He, 2017), or stronger conditions on the noises (Wu et al., 2019). Further investigation is needed in this direction for high-dimensional quantile regression with possibly heavy-tailed noises.

Supplementary Materials

The Supplementary Materials include some additional numerical results, discussion of condition A5, and the proofs of Theorems 1-3.

Acknowledgements

The authors would like to thank two anonymous reviewers, an associate editor and the editor for constructive comments and helpful suggestions. This work is partially supported by King Abdullah University of Science and Technology, Office of Sponsored Research under Award No: OSR-2015-CRG4-2582, the National Science Foundation (NSF) grant DMS-1712760, National Natural Science Foundation of China grants 11801355 and 11871376, Shanghai Pujiang Program 18PJ1409800, and the IR/D program from the NSF. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. All correspondence should be addressed to Dr. Yinfeng Wang, School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai,

201209, China. Email: dairy-2006@163.com.

References

Belloni, A. and Chernozhukov, V. (2011). l_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*, pp. 82–130.

Bennet, A., Reynolds, C., Eriksson, U., Hong, M., Blennow, K., Gatz, M., Alexeyenko, A., Pedersen, N. and Prince, J. (2011). Genetic association of sequence variants near AGER/NOTCH4 and dementia. *Journal of Alzheimer's Disease 24*, pp. 475–484.

Cai, T., Liu, W. and Xia, Y. (2014). Two-sample test of high-dimensional means under dependence. *Journal of the Royal Statistical Society, Series B 76*, pp. 349–372.

Crow, J. (1999). Hardy, Weinberg and language impediments. *Genetics 152*, pp. 821–825.

de Leeuw, C. A., Neale, B. M., Heskes, T. and Posthuma, D. (2016). The statistical properties of gene-set analysis. *Nature Reviews Genetics 17*, pp. 353–364.

Du, W., Tan, J., Chen, J. and Wang, L. (2016). Association between clusterin gene polymorphism rs11136000 and late-onset Alzheimer's disease susceptibility: A review and meta-analysis of case-control studies. *Experimental and Therapeutic Medicine 12*, pp. 2915–2927.

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

Guo, B. and Chen, S. (2016). Tests for high-dimensional generalized linear models. *Journal of the Royal Statistical Society, Series B 78*, pp. 1079–1102.

He, X. and Zhu, L. (2003). A lack-of-fit test for quantile regression. *Journal of American Statistical*

Association 98, pp. 1013–1022.

Horowitz, J. L. (2019). Bootstrap methods in econometrics. *Annual Review of Economics* to appear.

Kocherginsky, M., He, X. and Mu, Y. (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics 14*, pp. 41–55.

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

Leeb, H. and Pötscher, B. (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory 19*, pp. 100–142.

Leeb, H. and Pötscher, B. (2005). Model selection and inference: facts and fiction. *Econometric Theory 21*, pp. 21–59.

Li, G., Li, Y. and Tsai, C. L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association 110*, 246–261.

Liu, Y. and Xie, J. (2019). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, *115*, pp. 393–402.

Ma, S., Li, R. and Tsai, C. L. (2017). Variable screening via quantile partial correlation. *Journal of the American Statistical Association 112*, pp. 650–663.

McKeague, I. and Qian, M. (2015). An adaptive resampling test for detecting the presence of significant predictors. *Journal of American Statistical Association 110*, pp. 1422–1433.

Pan, W., Kim, J., Zhang, Y., Shen, X. and Wei, P. (2014). A powerful and adaptive association test for rare variants. *Genetics 197*, pp. 1081–1095.

REFERENCES

Park, S. and He, X. (2017). Hypothesis testing for regional quantiles. *Journal of Statistical Planning and Inference 191*, pp. 13–24.

Peng, B. and Wang, L. (2015). An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. *Journal of Computational and Graphical Statistics 24*, pp. 676–694.

Sherwood, B. and Wang, L. (2016). Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics 44*, pp. 288–317.

Siddiqui, M. (1960). Distribution of quantiles from a bivariate population. *Journal of Research of the National Bureau of Standards 64*, pp. 145–150.

Silbiger, S. and Neugarten, J. (2003). The role of gender in the progression of renal disease. *Advances in Renal Replacement Therapy 10*, pp. 3–14.

Stage, E., Duran, T., Risacher, S., Goukasian, N., Do, T., West, J., Wilhalme, H., Nho, K., Phillips, M., Elashoff, D., Saykin, A. and Apostolova, L. (2016). The effect of the top 20 Alzheimer disease risk genes on gray-matter density and FDG PET brain metabolism. *Alzheimers Dement 5*, pp. 53–66.

Sun, R., Hui, S., Bader, G. D., Lin, X. and Kraft, P. (2019). Powerful gene set analysis in GWAS with the generalized Berk-Jones statistic. *PLOS Genetics 15(3)*, e1007530.

Tan, L., Wang, H., Tan, M., Tan, C., Zhu, X., Miao, D., Yu, W., Jiang, T., Tan, L. and Yu, J. (2016). Effect of CLU genetic variants on cerebrospinal fluid and neuroimaging markers in healthy, mild cognitive impairment and Alzheimer's disease cohorts. *Scientific Reports 6*, 26027.

Tang, Y., Wang, H. and Barut, E. (2018). Testing the presence of significant covariates through conditional

marginal regression. *Biometrika 105*, pp. 57–71.

Wang, H. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics 37*, pp. 756–781.

Wang, H. and He, X. (2007). Detecting differential expressions in GeneChip microarray studies-a quantile approach. *Journal of American Statistical Association 102*, pp. 104–112.

Wang, H., McKeague, I. and Qian, M. (2018a). Testing for marginal effect in quantile regression. *Journal of the Royal Statistical Society, Series B 80*, pp. 433–452.

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

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

Wu, C., Xu, G. and Pan, W. (2019). An adaptive test on high-dimensional parameters in generalized linear models. *Statistica Sinica, 29*, pp. 2163–2186.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica 19*, pp. 801–817.

Xu, G., Lin, L., Wei, P. and Pan, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*, pp. 609–624.

Yu, T., Han, C., Zhu, G., Liao, X., Qin, W., Yang, C., Liu, Z., Su, H., Liu, X., Yu, L., Liu, Z., Lu, S., Chen, Z., Liang, Y., Huang, J., Qin, X., Gui, Y., Li, J. and Peng, T. (2017). Prognostic value of Notch receptors in postsurgical patients with hepatitis B virus-related hepatocellular carcinoma. *Cancer*

Medicine 6, pp. 1587–1600.

Zhang, L., Wang, H. and Zhu, Z. (2014). Testing for change points due to a covariate threshold in regression quantiles. *Statistica Sinica 24*, pp. 1859–1877.

Zhao, S. D. and Li, Y. (2014). Score test variable screening. *Biometrics 70*, pp. 862–871.

Zou, F., Fine, J., Hu, J. and Lin, D.Y. (2004). An efficient resampling method for assessing Genome-wide statistical significance in mapping quantitative trait loci. *Genetics 168*, pp. 2307–2316.

Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, School of Statistics, East China Normal University, Shanghai, 200062, China

E-mail: yltang@fem.ecnu.edu.cn

School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China

E-mail: dairy-2006@163.com

Department of Statistics, George Washington University, Washington D.C., 20052, USA

E-mail: judywang@gwu.edu

Department of Statistics, George Washington University, Washington D.C., 20052, USA

E-mail: qpan@gwu.edu