Statistica Sinica **32** (2022), 869-892
doi:https://doi.org/10.5705/ss.202019.0304

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

A 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 indices. It is 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. Inferences in the context of gene set detection face challenges in terms of both high-dimensionality and multiplicity, because the number of genes in a set can be much larger than the sample

Corresponding author: Yinfeng Wang, Interdisciplinary Research Institute of Data Science, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China. E-mail: dairy-2006@163.com.

size, and genes in different sets may overlap.

This study is motivated by a GWAS from the Diabetes Complication and Control Trial (DCCT), which searches for SNPs associated with the glomerular filtration rate (GFR) using genome-wide screening. The GFR is an important clinical index for the risks of nephropathy, one of the major microvascular complications in diabetic patients. The study has three unique features. First, the mean level of the GFR among participants is less clinically informative than the left tail quantiles. This is because the mean values are usually driven by the majority of participants without nephropathy, whereas the lower quantiles reflect the characteristics of the subset of participants that progressed to nephropathy. Second, the GFR data are skewed to the left, even after log-transformation; see Figure 2 in Section 4. Third, the data contain a large number of SNPs and some "protected" covariates that 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 the GFR, and let $(\mathbf{Z}^\top, \mathbf{X}^\top)^\top$ 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 set of the remaining d_n-dimensional covariates, with $d_n = p_n - q$, corresponding to the SNPs. Our goal is to assess the association between the SNPs and the lower tails of the GFR distribution in order to identify SNPs and gene pathways associated with patients at higher risk of kidney failure, after controlling the effect of the protected covariates.

In a 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, such as a Bonferroni adjustment or a false discovery rate (FDR) control. Although a Bonferroni adjustment controls the family-wise error rate (FWER) well, it is usually conservative, which may result in low power under the alternative. An FDR control works differently to an FWER control, which is suitable when there exist many important covariates. Other works based on a GWAS focus mainly on mean-regression-based tests. Without including $\mathbf{Z}$, Zou et al. (2004) proposed a resampling procedure to assess the significance of genome-wide quantitative trait loci mapping for Drosophila backcross. In addition, McKeague and Qian (2015) proposed an adaptive resampling test and applied it to glioblastoma cancer data. Guo and Chen (2016) proposed testing the overall significance of $\mathbf{X}$ conditional on $\mathbf{Z}$, based on a quadratic form of the score functions. Tang, Wang and Barut (2018) proposed a hybrid test of maximum- and sum-squared-type statistics based on conditional marginal regressions, where they regress 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, Xu and Pan (2019)

proposed an adaptive test for generalized linear models, in which they assume that the errors satisfy the subGaussian condition and $p_n = o(n^2)$. However, none of these mean-based methods are suitable for analyzing GFR data to meet the research goals.

As a valuable alternative to the mean regression, a quantile regression provides a natural way to capture the impact of covariates on the tail of the response distribution. A quantile regression does not, in general, require 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, as well as resampling-based approaches; see Koenker (2005, Chap. 3), Kochergin-sky, He and Mu (2005), Feng, He and Hu (2011), and Wang, Van Keilegom and Maidman (2018). Unfortunately, the existing tests apply only to low-dimensional covariates, and either have low power for large p_n or are infeasible for cases with $p_n \geq n$. For an inference with high-dimensional $\mathbf{X}$, one may first select a subset of predictors using a variable selection method (Wu and Liu (2009); Belloni and Chernozhukov (2011); Wang, Wu and Li (2012); Sherwood and Wang (2016)), and then conduct hypothesis testing on the selected model using conventional methods. However, this practice ignores the uncertainty involved in the model selection step and, thus, often leads to an inflated FWER (Leeb and Pötscher (2003, 2005)).

To detect significant predictors while accounting for the uncertainty involved in the selection stage, Wang, McKeague and Qian (2018) proposed a quantile marginal effect test based on the maximum of the marginal t-statistics. In addition, Wang, Van Keilegom and Maidman (2018) considered wild residual bootstrap inference for a penalized quantile regression, without the presence of $\mathbf{Z}$. However, their theories only work for a fixed dimension, and the method of Wang, McKeague and Qian (2018) uses a computationally intensive double bootstrap procedure to select the tuning parameter involved in the test calibration. Furthermore, in clinical studies, prognostic factors should be selected after accounting for the effects of protected covariates with known impacts on the outcome. By including $\mathbf{Z}$, Park and He (2017) extended the rank score test for quantile regressions with fixed dimensions to settings with diverging p_n; however, this method still requires $p_n < n$.

We propose a conditional marginal score-type test for a quantile regression in an ultrahigh-dimensional setting in order 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 $\mathbf{Z}$,

using rescaled conditional marginal rank scores, and define the final test statistic as the maximum of d_n squared score statistics. In contrast to existing works, 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 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 the 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 a double bootstrap (McKeague and Qian (2015); Wang, McKeague and Qian (2018); Tang, Wang and Barut (2018)). Combination tests are alternatives to the proposed maximum-type statistic for determining group-wise significance. For instance, the Cauchy combination test (Liu and Xie (2019, CCT)) combines the P-values obtained from the individual covariate tests into a single P-value in order to assess the group-wise significance. However, our simulation studies show that the CCT tends to be conservative in high dimensions.

In addition to the nice properties presented above, the proposed test can be used as a stopping rule in forward selection, where in each step, the preselected set is treated as the conditioning set. In settings with 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 have proposed penalized variable selection methods for quantile regressions. Zhao and Li (2014) proposed a score-test-based variable screening method, and Li, Li and Tsai (2015) and Ma, Li and Tsai (2017) proposed screening methods based on quantile partial correlations. Screening and penalized selection methods can only tell us whether a covariate is selected, whereas 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 using 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. Additional simulation results and all technical proofs are provided in the online Supplementary Material.

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 identically distributed (i.i.d.) 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 τth 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 an association between $\mathbf{X}$ and the τth quantile of Y, after accounting for the effect of $\mathbf{Z}$; that is, we test

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

The testing of (2.2) can be viewed as a first step in a 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

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

such that $Q_\tau\{\varepsilon_i(\tau)|\mathbf{Z}_{i\cdot}, \mathbf{X}_{i\cdot}\} = 0$. We let $\mathbb{X} = (\mathbf{X}_{1\cdot}, \ldots, \mathbf{X}_{n\cdot})^\top$, $\mathbb{Z} = (\mathbf{Z}_{1\cdot}, \ldots, \mathbf{Z}_{n\cdot})^\top$, $\mathbb{X}_{\cdot j} = (X_{1,j}, \ldots, X_{n,j})^\top$, for $j = 1, \ldots, d_n$, and $\mathbf{f}_\tau = \mathrm{diag}(f_{1,\tau}(0), \ldots, f_{n,\tau}(0))$, 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

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

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

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

such that the jth component of $\mathbf{X}$ is orthogonal to $\mathbf{Z}$ in a weighted manner;

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

Second, we define the rescaled conditional marginal score statistic as

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

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 jth 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

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

In practice, $\mathbf{f}_\tau$ is unknown and has to be estimated and substituted in. We propose estimating $\mathbf{f}_\tau$ using the quotient method (Siddiqui (1960)); that is,

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

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 Material, 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 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 the covariates. In this case, $\mathbf{f}_\tau(\cdot)$ cancels out in expression (2.4), and the test statistic $T_{n,1}(\tau)$ reduces to

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

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

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 in Ma, Li and Tsai (2017). The test statistic $T_{n,2}(\tau)$ has a simpler form, and does not depend on the unknown density function. In low-dimensional quantile regression settings, it is known that a score test that assumes homoscedastic errors still performs competitively well when the homoscedasticity assumption is violated; see Wang and Fygenson (2009) and Park and He (2017). We 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 $T_{n,k}(\tau)$, for $k = 1, 2$, 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, for $k = 1, \ldots, 5$, are some positive constants.

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

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, for $i = 1, \ldots, n$, 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 a bounded third derivative with respect to ν.

For technical convenience, condition A1 (i) requires the dimension of $\mathbf{Z}$ to be fixed, which is also practically reasonable in a GWAS. We can relax this con-

dition by allowing q to diverge slowly. 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 a fixed q. A more careful investigation is needed for diverging q. Conditions A1 (ii) and (iii) describe the dimension and distribution of $\mathbf{X}$, respectively, and are standard in high-dimensional settings. Condition A3 is an assumption on the density function that is standard in quantile regressions. Condition A4 ensures that $\mathbf{f}_\tau$ can be consistently estimated; see Lemma S.1 in the Supplementary Material for more details. Now, we discuss condition A2. By the assumption that X_j is centralized, under H_0, we have

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

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, and are analogous to conditions 1 and 3 and that in Lemma 6 of Cai, Liu and Xia (2014). Conditions A2 (i)–(ii) are mild, and 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*

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

as $n, d_n \to \infty$.

The proof of Theorem 1 consists of two parts, where the first part controls $\max_{1\le j\le d_n} |S_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\} - S_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\}|$, and the second part is used to derive the asymptotic distribution of $\max_{1\le j\le d_n} S_{\tau,j}^2\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)\}$. 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 in the indicator function. 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 $j \in \{1, \ldots, d_n\}$,

$\#\{l : |\mathrm{corr}[S_{\tau,j}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau), S_{\tau,l}\{\boldsymbol{\alpha}_{\mathbf{Z},0}(\tau)]| > d_n^{-\gamma_0}\}$ is well controlled by A2 (iii) for some $\gamma_0 > 0$, which is similar to the proof of Theorem 6 in Cai, Liu and Xia (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

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

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*

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

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,

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

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; see the 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*

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

Because $\sqrt{\log(d_n)/n}$ is the optimal convergence rate that can be obtained

in a high-dimensional setting (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 that 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. Multiplier bootstrap are also considered in other settings for low-dimensional data; see, for example, He and Zhu (2003), Zhang, Wang and Zhu (2014), and 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 includes $T_{n,2}(\tau)$ as a special case.

Step 1. Let

$$T_{n,1}(\tau)^* = \max_{1\le j\le d_n} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i X^*_{i,j,\tau} \psi_\tau(e_i) \right\}^2 \Big/ \Big\{ \tau(1-\tau) \|\mathbb{X}^*_{\cdot j,\tau}\|^2/n \Big\}, \quad (2.8)$$

where $\{e_i; i = 1, \ldots, n\}$ is a random sample with the τth quantile zero, and $\{w_i; i = 1, \ldots, n\}$ 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 the 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 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 that

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

under the null hypothesis. Theorem 3 provides a theoretical justification for the multiplier bootstrap method in a high-dimensional setting. Similar results can 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*

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

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 using 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 a forward regression to discover significant components in $\mathbf{X}$. To account for multiple testing in the sequential procedure, we follow a two-stage selection similar to that in Tang, Wang and Barut (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 let P_1 be the associated P-value. If $P_1 > \gamma$, the prespecified significance level, we stop and declare that there is no significant X_j. Otherwise, we move $X_{\widehat{j}_1}$ from $\mathbf{X}^{(0)}$ to $\mathbf{Z}^{(0)}$, and repeat the procedure until no further 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 adjustments. Suppose that $K \ge 1$. Define $K^* = 1$ if $P_1 > \gamma/K$, otherwise $K^* = \max_{1\le k\le K}\{k : P_l \le \gamma/(K - l + 1), l = 1, \ldots, k\}$. Finally, the selected covariate set is chosen as $\mathbf{Z}^{(K^*)} = \{\mathbf{Z}, X_{\widehat{j}_1}, \ldots, X_{\widehat{j}_{K^*}}\}$.

Note that it is challenging to establish a formal theoretical justification for the proposed two-stage method, owing to its sequential nature. However, 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:

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

where $\mathbf{Z}_{i\cdot} = (1, \widetilde{\mathbf{Z}}_{i\cdot}^\top)^\top$, a_0 is the parameter controlling the heterogeneity of the noise, $\boldsymbol{\alpha}_{\mathbf{Z},0} = \mathbf{1}_{q=6}$, $\mathbf{b}_0 = \mathbf{0}_{d_n = p_n - q}$ 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 prespecified $\delta_{\max,p_n}$.

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, nonGaussian 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. Then, we let $\widetilde{Z}_{i,l} = 2\sqrt{3}\Phi(U_{i,l}) - \sqrt{3}$, for $l = 1, \ldots, 5$, and let $X_{i,l-5} = 2\sqrt{3}\Phi(U_{i,l}) - \sqrt{3}$, 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 repetitions in the multiplier bootstrap method as $M = 500$. We also consider a case to mimic the motivating GFR study in Section S1.2; the main observations are similar to Cases 1–3.

The following tests are compared: (i) four variations of the proposed test, $T_{n,1}^E(\tau)$, $T_{n,1}^B(\tau)$, $T_{n,2}^E(\tau)$, $T_{n,2}^B(\tau)$, 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 of Park and He (2017), with $p_n < n$; (iii) QME, the quantile marginal effect test of Wang, McKeague and Qian (2018), with the tuning parameter set as $\lambda_n = 3\sqrt{\tau(1-\tau)\log n}$; (iv) BON, the Bonferroni adjustment method, where the individual P-values are based on $S_{\tau,j}\{\widehat{\boldsymbol{\alpha}}_{\mathbf{Z}}(\tau)\}$ and its asymptotic normality, for $j = 1, \ldots, d_n$, that is, the proposed conditional marginal rank score statistics for heteroscedastic cases; (v) CCT, the Cauchy combination test of Liu and Xie (2019), where the individual P-values are the same as in BON; (vi) CAR, the conditional adaptive resampling test of Tang, Wang and Barut (2018) for the mean model, and the tuning parameter is set as $\lambda_n = \max\left[3(\log n)^{1/2}, \Phi^{-1}\{1 - \gamma/(2d_n)\}\right]$; and (vii) GC, the partial test of 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, McKeague and Qian (2018) and Tang, Wang and Barut (2018), the tuning parameter is selected using a double bootstrap, which is computationally intensive; thus we fix the parameter to a value that performs

relatively well to reduce the computation time. Table S.1 in the Supplementary Material summarizes the average computing time for each method. The results show that the methods that do not require an estimation of $\mathbf{f}_\tau$, namely, RS, $T_{n,2}(\tau)$, and GC, are computationally more efficient than those that 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 the other methods, even if a double bootstrap is not used for the tuning parameters.

Table 1 summarizes the rejection percentages of the various methods in Cases 1 and 3 with $\mathbf{b}_0 = \mathbf{0}$. The empirical sizes from Case 2 are similar to those from Case 1, and thus are moved to Table S.2 in the Supplementary Material. In all scenarios except Case 3 with $\tau = 0.25$, the null hypothesis is true; thus, the rejection rate corresponds to the empirical size. In Case 3 with $\tau = 0.25$, $\beta_{1,0}(\tau) = (1/2)F_{t_3}^{-1}(0.25)$, and 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. However, 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 the type-I error and power, and is computationally much simpler than the test based on $T_{n,1}(\tau)$. The RS test performs well for small p_n, but becomes quite conservative for larger p_n and is not applicable when $p_n \geq n$. The QME test is sensitive to the choice of the tuning parameter; it gives a 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 tests control the type-I errors reasonably well. However, in the heteroscedastic Case 3 with $\tau = 0.25$, they are both more conservative than the proposed multiplier bootstrap method in terms of detecting signals, especially for $n = 200$. Finally, the mean-based tests CAR and GC perform well in the homoscedastic cases, but are not able to detect the signal at the tail quantiles caused by the heteroscedasticity, as seen in Case 3.

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

improved by using a double bootstrap to select a data-adaptive λ_n. However, 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 $n = 200$. 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 zero 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, owing to its sensitivity to the choice of the tuning parameter. We also exclude BON and CCT, because they are shown in Table 1 to be more conservative in terms of detecting signals in heteroscedastic cases with small samples.

Figure 1 presents the power curves of the different methods. Both CAR and GC are designed to detect 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, and so is less powerful in terms of detecting the sparse signal in all four cases, especially for large p_n. In addition, neither CAR nor GC can identify the signal at the tails, as shown in Table 1. The RS test performs competitively well for $p_n = 10$, but it quickly loses power 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, yielding 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 $T_{n,2}(\tau)$ that assume 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 $\mathbf{Z}_{i\cdot} = (1, \widetilde{\mathbf{Z}}_{i\cdot}^\top)^\top$, $\boldsymbol{\alpha}_{\mathbf{Z},0} = \mathbf{1}_6$, and $\boldsymbol{\beta}_{\mathbf{X},0} = (0, 1, 1, 0.8, 0.8, 0_{p_n-11}^\top)^\top$, with $n = 200$ and $p_n = 200$ and 1,000; $(\widetilde{\mathbf{Z}}_{i\cdot}^\top, \mathbf{X}_{i\cdot}^\top)^\top$ 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 1,000 replicates are considered with a nominal level $\gamma = 0.05$. We compare the following forward selection procedures: (i) $T_{n,1}^B(\tau)$, a sequential test based on $T_{n,1}^B(\tau)$; (ii) L_1, the L_1-penalized variable selection method of Belloni and Chernozhukov (2011), without penalizing the coefficients of $\mathbf{Z}$; (iii) QPCOR-L_1, the quantile partial correlation

Table 1. Rejection percentages of different tests for Cases 1 and 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	p_n / method	10	50	200	1,000	10	50	200	1,000	10	50	200	1,000	10	50	200	1,000
$\tau = 0.25$	$T^E_{n,1}(\tau)$	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 $T^B_{n,k}(\tau)$, for $k = 1, 2$: four variations of the proposed test; RS: the rank score test of Park and He (2017); QME: the quantile marginal effect test of Wang, McKeague and Qian (2018); BON, Bonferroni adjustment on d_n individual P-values; CCT, Cauchy combination test of Liu and Xie (2019); CAR: the conditional adaptive resampling test of Tang, Wang and Barut (2018); GC: the sum-squared-type test of Guo and Chen (2016).

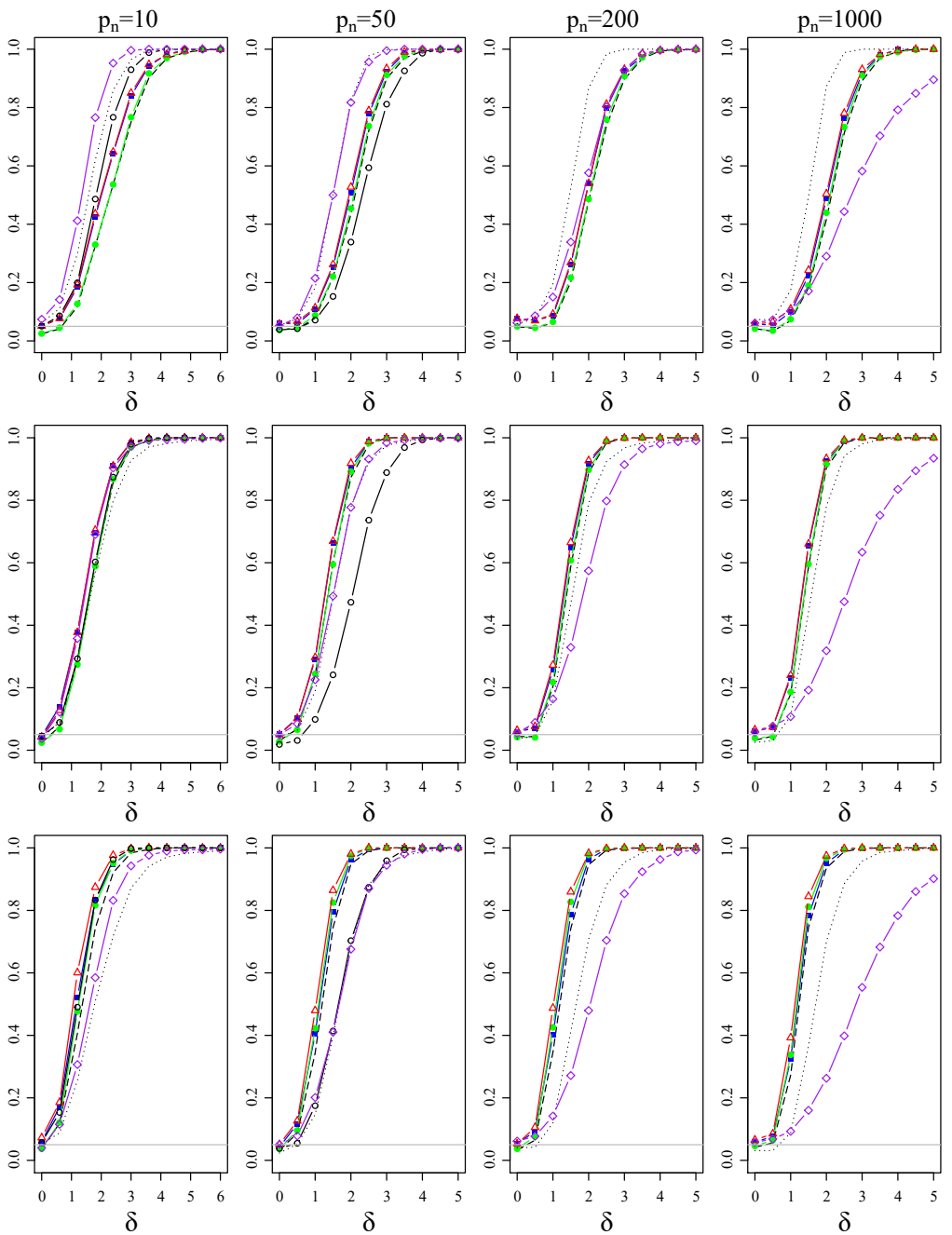

Figure 1. Power curves of the methods in cases 1 (first row), 2 (second row), and 3 (third row), with $n = 200$ and $\tau = 0.5$: $T^E_{n,1}(\tau)$ (dashed), $T^B_{n,1}(\tau)$ (line with solid square), $T^E_{n,2}(\tau)$ (line with solid dots), $T^B_{n,2}(\tau)$ (line with triangle), RS (line with open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

screening method of Ma, Li and Tsai (2017), where we use their algorithm 3 to reduce the dimension of $\mathbf{X}$ from d_n to $n/\log n$, followed by the L_1-penalized method of Belloni and Chernozhukov (2011); and (iv) CAR, a sequential test based on CAR, with the same tuning parameter over the replicates as in Section 3.1. For the sequential-test-based methods $T^B_{n,1}(\tau)$ and CAR, multiple test adjustments are applied; see Section 2.6. For the quantile based methods, we focus on $\tau = 0.5$.

To evaluate the performance of the methods, we consider the percentages of replicates in which X_j, for $j = 1, \ldots, 5$, are selected (PS), the average number of false positives (FP), the percentages of replicates of under-fit (UF) in which at least one important X_j is not selected, and the percentages of the replicates in which the exact true model (TM) is selected. We find the following: (i) the performance of $T^B_{n,1}(\tau)$ is competitive or better in all scenarios; (ii) L_1 and QPCOR-L_1 both tend to over-fit the model (higher FP); further steps may be applied to the selected model to refine the selection accuracy, but inherent uncertainty may accumulate; (iii) CAR performs well when the noise is homoscedastic normal (Case 1), but the under-fit percentages (UF) can be high when the noise is heavy-tailed with heteroscedasticity (Case 3).

4. Analysis of the Glomerular Filtration Rate

An SNP is a substitution of a single nucleotide that occurs at a specific position in the genome, and some 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 interest, in the presence of some "protected" demographic covariates. Over a million SNPs are mapped in the GWAS of the 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 GFR, measured in percentages, a popular 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 risk, which is usually measured using the most recent kidney functions, that is, the GFR measurement at the last visit. The "protected" covariates include 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 stage of nephropathy development in patients.

Table 2. 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
	1,000	$T^B_{n,1}(\tau)$	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
	1,000	$T^B_{n,1}(\tau)$	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 of Belloni and Chernozhukov (2011); QPCOR-L_1: the QPCOR of Ma, Li and Tsai (2017); CAR: forward selection based on CAR of Tang, Wang and Barut (2018). PS: percentage of being selected; FP: average number of false positives; UF: percentage of replicates in which at least one important X_j is not selected; TM: percentage of replicates in which the exact true model is selected.

The GWAS of the DCCT contains 1.18 million candidate SNPs for 1,304 patients, which is far less than the number of SNPs. An important statistical issue concerns assessing the overall significance of groups of SNPs, that is, whether SNPs exist in a set of genes that have an effect on the disease, while controlling for the family-wise error rate. Most works based on a GWAS consider mean-based tests. However, in this study, the mean of the 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, whereas the lower quantiles reflect the characteristics of participants with elevated risks of nephropathy. Furthermore, the GFR values are skewed to the left, even after a logarithm transformation (Figure 2). Thus, a quantile regression at several lower quantile levels could provide more clinically relevant information than that of the mean regression, and it enables us to work on the original scale, providing a better interpretation for clinicians and patients. For these reasons, we assess the significance of SNPs on lower quantiles of the GFR to identify SNPs and gene

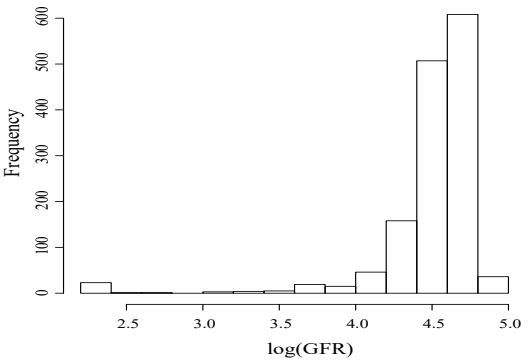

Figure 2. Log-transformed GFR% from the type 1 diabetes patients at the last visit in the DCCT study.

pathways associated with patients with a high risk of nephropathy; as such, we consider quantile levels $\tau = 0.05, 0.1, 0.25$ and 0.5.

To apply the proposed method, we focus on a subset of SNPs that belong to genes related to nephropathy in the MSigDB Curated Gene Sets (`http://software.broadinstitute.org/gsea/msigdb/`), including 2,908 SNPs after deleting those not satisfying the Hardy–Weinberg equilibrium (Crow (1999)). For further preprocessing, we (i) delete one female patient who has 98% of the SNPs missing, (ii) delete SNPs with any missing values, (iii) delete SNPs with a minor allele frequency of less than 5%, and (iv) prune highly correlated SNP pairs, defined as correlation coefficients larger than 0.99. Finally, we have 1,303 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 one. Previous works have 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. 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 four variations of the proposed test as the overall significance test at different quantiles. Conditional on $\mathbf{Z}$, all variations of the proposed test suggest that significant (at the level of 0.05) SNPs exist in the male group at $\tau = 0.05$, 0.1, but not for the female group or other quantiles. We also apply the CAR of Tang, Wang and Barut (2018) for the overall significance test on the conditional mean. We consider the test with tuning parameter $\lambda_n = \max\left[a(\log n)^{1/2}, \Phi^{-1}\left\{1 - \gamma/(2d_n)\right\}\right], a \in \{3, 4, 5, 6, 7\}$. All λ_n lead to the

Table 3. Summary of SNPs selected in the random subsets for males.

τ	SNP	Frequency	r_{var}	r_{mean}	r_{Q_τ}
0.05	rs9331949_C	11	17	3	1
	rs12411439_A	8	66	5	11
	rs6866731_G	7	18	38	49
	rs4714956_T	6	576	11	10
0.1	rs11742097_A	27	22	14	94
	rs463701_G	9	20	17	29
	rs6866731_G	6	26	15	89

same P-values, namely, 0.572 and 0.324 for the male and female groups, respectively, indicating 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 3 presents the frequencies of the SNPs selected at least 5 times by our method over 50 random subsets.

For further verification, we regress Y on $\mathbf{Z}$ at the τth quantile, obtaining the residuals under the null model. In general, if one SNP has an effect on the response, residuals with different genotypes have different distributions. We calculate the variance, mean, and τth quantile of the residuals in genotype "AA" and "Aa" for each SNP, and report the ranks of the differences between 981 SNPs in Table 3. We find that most of the SNPs selected with high frequencies have high ranks in at least one of the three criteria, providing further evidence of the effects of the selected SNPs on the lower quantiles of GFR.

Furthermore, we searched PubMed for publications that studied the same SNPs identified in our analysis as a source of validation from external data for the functions of the reported SNPs. Specifically, rs9331949 has been found to be associated with epilepsy, cognitive impairment, and Alzheimer's disease (Yu et al. (2013); Tan et al. (2016); Du et al. (2016); Xian et al (2017)). The six SNPs belong to five different genes (ATP10B, CLU, FAM53B, ADGRF5, SPG7) with rs11742097 and rs6866731 both locating within gene ATP10B. Therefore, we carried out gene set enrichment analysis (GSEA), searching for functional gene sets overlapping significantly with the selected SNPs.

The five identified genes all belong to gene set BAELDE DIABETIC NEPH-

ROPATHY UP (P-value=4.68×10^{-14}), which is a set of genes up-regulated in glomeruli of kidneys from patients with diabetic nephropathy (type 2 diabetes mellitus).

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, the signals may be weak and dense; that is, groups of markers may jointly affect the phenotype, while the signal of each marker is faint. To adapt different types of signals, we consider a hybrid test statistic by taking a weighted average of the maximum- and sum-squared-type statistics, tests as in Tang, Wang and Barut (2018). However, the existing literature on sum-squared-type tests requires smoothed loss functions (Guo and Chen (2016)), limited dimensionality of the markers (Park and He (2017)), or stronger conditions on the noise (Wu, Xu and Pan (2019)). Further investigation is needed in this direction for high-dimensional quantile regressions with possibly heavy-tailed noise.

Supplementary Material

The online Supplementary Material includes additional numerical results, a discussion of condition A5, and proofs of Theorems 1–3.

Acknowledgments

The authors thank two the anonymous reviewers, associate editor, and editor for their constructive comments and helpful suggestions. This work was partially supported by King Abdullah University of Science and Technology, Office of Sponsored Research under Award No. OSR-2015-CRG4-2582, 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. The opinions, findings, and conclusions and 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, Institute of Data Science and Interdisciplinary Studies, 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**, 82–130.

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 (Statistical Methodology)* **76**, 349–372.

Crow, J. (1999). Hardy, Weinberg and language impediments. *Genetics* **152**, 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**, 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**, 2915–2927.

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

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

He, X. and Zhu, L. (2003). A lack-of-fit test for quantile regression. *Journal of American Statistical Association* **98**, 1013–1022.

Horowitz, J. L. (2019). Bootstrap methods in econometrics. *Annual Review of Economics* **11**, 193–224.

Kocherginsky, M., He, X. and Mu, Y. (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics* **14**, 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**, 100–142.

Leeb, H. and Pötscher, B. (2005). Model selection and inference: facts and fiction. *Econometric Theory* **21**, 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**, 393–402.

Ma, S., Li, R. and Tsai, C. L. (2017). Variable screening via quantile partial correlation. *Journal of the American Statistical Association* **112**, 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**, 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**, 1081–1095.

Park, S. and He, X. (2017). Hypothesis testing for regional quantiles. *Journal of Statistical Planning and Inference* **191**, 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**, 676–694.

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

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

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

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**, e1007530.

Tan, L., Wang, H., Tan, M., Tan, C., Zhu, X., Miao, D. et al. (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**, 57–71.

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

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

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

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

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

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

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

Xian, W., Tao, H., Zhao, J., Fu, J., Zhong, W., Chen, Y. et al. (2017). Association between clusterin gene polymorphisms and epilepsy in a Han Chinese population. *Genet Test Mol Biomarkers* **21**, 692–697.

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

Yu, J., Ma, X., Wang, Y., Sun, L., Tan, L., Hu, N. et al. (2013). Genetic variation in clusterin gene and Alzheimer's disease risk in Han Chinese. *Neurobiology of Aging* **34**, 1921.e17–23.

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

Zhao, S. D. and Li, Y. (2014). Score test variable screening. *Biometrics* **70**, 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**, 2307–2316.

Yanlin Tang

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

Yinfeng Wang
Interdisciplinary Research Institute of Data Science, School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China.
E-mail: dairy-2006@163.com
Huixia Judy Wang
Department of Statistics, George Washington University, Washington D.C., 20052, USA.
E-mail: judywang@gwu.edu
Qing Pan
Department of Statistics, George Washington University, Washington D.C., 20052, USA.
E-mail: qpan@gwu.edu

(Received August 2019; accepted September 2020)