Lack-of-fit tests for quantile regression models

Chen Dong, Guodong Li and Xingdong Feng *

December 19, 2018

SUMMARY

This paper novelly transforms lack-of-fit tests for parametric quantile regression models into checking the equality of two conditional distributions of covariates. We then can borrow these successful test statistics from the rich literature of two-sample problems, and this gives us much flexibility in constructing a suitable lack-of-fit test according to our experiences on covariates. This finding is first demonstrated for the low dimensional data by using a practical two-sample test, which has a sound power for the case with a moderate dimension. We then apply it to the high dimensional data, and a lack-of-fit test for linear quantile regression models is thus constructed via combining two-sample test statistics in the literature. The asymptotic distribution of the test statistic under the null hypothesis has an explicit form, and we hence can calculate the critical values or p-values directly. The usefulness of these tests are illustrated by simulation experiments, and the real analysis gives further support.

Keywords: High dimensional data; Hypothesis test; Lack-of-fit; Quantile regression; Two-sample test.

*Xingdong Feng (Email: feng.xingdong@mail.sufe.edu.cn) is corresponding author and Professor, School of Statistics and Management & Institute of Data Science and Statistics, Shanghai University of Finance and Economics (SUFE), and Shanghai Key Laboratory of Financial Information Technology (SUFE), 777 Guoding Road, Shanghai 200433, China. Chen Dong (Email: dongchen@live.sufe.edu.cn) is Ph.D student, School of Statistics and Management, SUFE. Guodong Li (Email: gdli@hku.hk) is Associate Professor, Department of Statistics and Actual Sciences, University of Hong Kong, Hong Kong, China.

1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression has become an effective alternative to mean regression in many fields such as finance, economics, and geology; see Koenker (2005) for its literature review. For a response Y and covariates $\mathbf{X}$, instead of the conditional mean $E(Y \mid \mathbf{X})$ in mean regression, quantile regression aims to the τth quantile of Y conditional on $\mathbf{X}$,

$$Q_\tau(Y \mid \mathbf{X}) = m_\tau(\mathbf{X}),$$

where $0 < \tau < 1$, and random vector $\mathbf{X}$ consists of p covariates with a fixed p. The function $m_\tau(\cdot)$ is unknown and depends on τ, and it is flexible to use a nonparametric approach to estimate it. However, this method usually has a bad performance even when p is moderate, and it is also well known to be lack of interpretation (Koenker, 2005; Fan and Gijbels, 1996). As a result, the parametric method is still routinely used in quantile regression as well as in other scenarios, and specifically a parametric form will be assumed to the function of $m_\tau(\cdot)$, i.e. $m_\tau(\mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta})$ is known up to a parameter vector $\boldsymbol{\beta}$. In the meanwhile, it is an important task to perform a lack-of-fit test to check whether the parametric form is misspecified. Zheng (1998) first considered a kernel-based test for a general parametric quantile regression model. He and Zhu (2003) extended the approach in Stute (1997), and proposed a test based on a weighted cusum process of the residuals; see also Horowitz and Spokoiny (2002), Whang (2006), Otsu (2008), Escanciano and Velasco (2010) and Escanciano and Goh (2014) for more lack-of-fit tests based on cusum processes. The above tests are all nonparametric, and they can detect the departures at all directions when the sample size tends to infinity. As a cost, the number of covariates p is limited to a small value, say one or two, in real applications. Conde-Amboage, Sanchez-Sellero, and Gonzalez-Manteiga (2015) suggested to project the covariates $\mathbf{X}$ into a random variable first, and then applied He and Zhu's method to form a lack-of-fit test. It works well for a larger $p < n$.

Denote by $\boldsymbol{\beta}_0$ the true parameter vector, and let $\varepsilon = Y - m_\tau(\mathbf{X}, \boldsymbol{\beta}_0)$. It then holds that $P\{Q_\tau(Y|\mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta}_0)\} = 1$ if and only if

$$E\{I(\varepsilon < 0) \mid \mathbf{X}\} = \tau \quad \text{with probability one,} \tag{1}$$

where $I(\cdot)$ is the indicator function. The aforementioned lack-of-fit tests are all based on the fact at (1), while they do not pay attention to, or do not need, another fact that the random variable $I(\varepsilon < 0)$ only takes two possible values. Consider the distribution functions of $\mathbf{X}$ conditional on $I(\varepsilon < 0)$ and $I(\varepsilon > 0)$, respectively. We can show that equation (1) holds if and only if these two conditional distributions are equal; see Lemma 1 for details. This makes it possible to check whether the parametric form $m_\tau(\mathbf{X}, \boldsymbol{\beta})$ is correctly specified via solving a two-sample problem. For example, it will lead to He and Zhu's (2003) lack-of-fit test if the Cramér-von Mises test is applied to check the equality of the two conditional distributions of $\mathbf{X}$; see Section 2 for details. There is a rich literature of two-sample tests, and we can always find a suitable test statistic in this literature to form the corresponding lack-of-fit test according to our experiences on covariates. To demonstrate the idea here, we first consider the two-sample test statistic in Baringhaus and Franz (2004), which has a sound power even for the case with a moderate dimension, in Section 3.

Quantile regression has recently attracted more and more attentions in the literature of high dimensional data, where the number of covariates p may greatly exceed that of observations, the linear model is usually assumed, i.e.

$$Q_\tau(Y \mid \mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}^{*\top}\boldsymbol{\beta} \quad \text{with} \quad \mathbf{X}^* = (1, \mathbf{X}^\top)^\top, \tag{2}$$

and almost all researches in this area concentrate on the variable selection; see Belloni and Chernozhukov (2011), He et al. (2013), Zheng et al. (2015), Ma et al. (2017) and references therein. Shah and Buhlmann (2018) first introduced the concept of lack-of-fit, or goodness-of-fit, for high dimensional linear mean models, and it can be adopted to quantile regression models. When the number of covariates p is larger than that of observations, it usually reaches the exact fit of the data, and this leaves no room for the discussion of lack-of-fit. While the situation is different if model (2) is a sparse model. For a certain data generating process, if there is no good sparse approximation of $\mathbf{X}^{*\top}\boldsymbol{\beta}$ to $m_\tau(\mathbf{X}, \boldsymbol{\beta})$, then a sparse nonlinear model may be more suitable than a sparse linear one. Moreover, the lack-of-fit in this paper also refers to the case that some important covariates are missed in searching for the good sparse approximation of $\mathbf{X}^{*\top}\boldsymbol{\beta}$.

To construct a lack-of-fit test for the high dimensional linear quantile regression model at (2), we may first consider the residual prediction method in Shah and Buhlmann (2018), however, it heavily depends on the ordinary least squares estimation, and cannot be extended to the quantile regression model. Section 4 alternatively gets use of the relationship between lack-of-fit tests and two-sample problems, and then introduces a test by applying two high dimensional two-sample tests in Cai et al. (2013) and Cai et al. (2014) together. More importantly, the asymptotic distribution of the test statistic under the null hypothesis has an explicit form, and we can calculate the critical values or p-values directly.

The proofs of all lemmas and theorems are given in a separated supplementary file, and all datasets and codes used in the paper can by downloaded at the website `https://bb9.sufe.edu.cn/webapps/cmsmain/webui/users/2011000070/LoF/CodesAndDatasets.zip`.

2 Relationship between lack-of-fit tests and two-sample problems

Suppose that the τth quantile of Y conditional on $\mathbf{X}$ has a parametric form of

$$Q_\tau(Y \mid \mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta}), \tag{3}$$

where $m_\tau(\cdot,\cdot)$ is a known function, $\mathbf{X} = (X_1, \ldots, X_p)^\top$ consists of p covariates, and $\boldsymbol{\beta}$ is the parameter vector. Denote by $\boldsymbol{\beta}_0$ the true parameter vector. Let $\varepsilon = Y - m_\tau(\mathbf{X}, \boldsymbol{\beta}_0)$, and $g(\mathbf{X}) = E\{I(\varepsilon < 0) \mid \mathbf{X}\}$. To check whether the parametric form of model (3) is correctly specified, we can summarize the hypotheses below,

$$H_0 : P\{g(\mathbf{X}) = \tau\} = 1 \quad vs \quad H_1 : P\{g(\mathbf{X}) = \tau\} < 1.$$

Denote the observed data by $\{(Y_i, \mathbf{X}_i^\top)^\top, i = 1, ..., n\}$, which are independent and identically distributed ($i.i.d.$) random vectors, where $\mathbf{X}_i = (X_{1i}, \cdots, X_{pi})^\top$, and n is the number of observations. Denote $\mathcal{S} = \{1 \le i \le n : \varepsilon_i < 0\}$ and $\mathcal{S}^c = \{1 \le i \le n : \varepsilon_i \ge 0\}$, where $\varepsilon_i = Y_i - m_\tau(\mathbf{X}_i, \boldsymbol{\beta}_0)$. We then can separate the observed covariates $\{\mathbf{X}_i, 1 \le i \le n\}$

into two samples, $\{\mathbf{X}_i, i \in \mathcal{S}\}$ and $\{\mathbf{X}_i, i \in \mathcal{S}^c\}$, and they have the distributions of $F_{\mathcal{S}}(\mathbf{x}) = P(\mathbf{X} < \mathbf{x} \mid \varepsilon < 0)$ and $F_{\mathcal{S}^c}(\mathbf{x}) = P(\mathbf{X} < \mathbf{x} \mid \varepsilon \geq 0)$, respectively.

Lemma 1. *It holds that*

$$F_{\mathcal{S}}(\mathbf{x}) - F_{\mathcal{S}^c}(\mathbf{x}) = \frac{1}{\tau(1-\tau)} \int_{-\infty}^{\mathbf{x}} \{g(\mathbf{s}) - \tau\} dF_{\mathbf{X}}(\mathbf{s}),$$

where $F_{\mathbf{X}}(\cdot)$ is the distribution function of $\mathbf{X}_i$.

It is implied by the above lemma that $P\{g(\mathbf{X}_i) = \tau\} = 1$ if and only if $F_{\mathcal{S}}(\cdot) = F_{\mathcal{S}^c}(\cdot)$. As a result, in order to check whether model (3) is correctly specified, we can equivalently test for the hypotheses,

$$H_0 : F_{\mathcal{S}}(\cdot) = F_{\mathcal{S}^c}(\cdot) \quad vs \quad H_1 : F_{\mathcal{S}}(\cdot) \neq F_{\mathcal{S}^c}(\cdot). \tag{4}$$

Note that the true parameter vector $\boldsymbol{\beta}_0$ is unknown, and we may estimate it by

$$\widehat{\boldsymbol{\beta}}_n = \text{argmin} \sum_{i=1}^{n} \rho_\tau\{Y_i - m_\tau(\mathbf{X}_i, \boldsymbol{\beta})\},$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$; see Koenker (2005) for references. Let $\widehat{\varepsilon}_i = Y_i - m_\tau(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n)$, $\widehat{\mathcal{S}} = \{1 \leq i \leq n : \widehat{\varepsilon}_i < 0\}$, and $\widehat{\mathcal{S}}^c = \{1 \leq i \leq n : \widehat{\varepsilon}_i \geq 0\}$. We next consider the Cramér-von Mises test (Anderson, 1962) to check the equality of the distributions of samples $\{\mathbf{X}_i, i \in \widehat{\mathcal{S}}\}$ and $\{\mathbf{X}_i, i \in \widehat{\mathcal{S}}^c\}$.

Let $\kappa_n = \sum_{i \in \mathcal{S}} 1$ and $\widehat{\kappa}_n = \sum_{i \in \widehat{\mathcal{S}}} 1$ be the number of elements in the sets $\mathcal{S}$ and $\widehat{\mathcal{S}}$, respectively. When the function $m_\tau(\cdot, \cdot)$ at model (3) has a linear form, it holds that $\kappa_n = n\tau + o_p(n)$ and $\widehat{\kappa}_n = n\tau + o_p(n)$; see Theorem 2.2 of Koenker (2005) for details. The weighted empirical distributions of $F_{\mathcal{S}}(\cdot)$ and $F_{\mathcal{S}^c}(\cdot)$ then have the forms of

$$\widehat{F}_{\widehat{\mathcal{S}}}(\mathbf{x}) = \frac{1}{n\tau} \sum_{i \in \widehat{\mathcal{S}}} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leq \mathbf{x}) \text{ and } \widehat{F}_{\widehat{\mathcal{S}}^c}(\mathbf{x}) = \frac{1}{n(1-\tau)} \sum_{i \in \widehat{\mathcal{S}}^c} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leq \mathbf{x}),$$

respectively, where $\omega(\cdot)$ is the weight function. Let $\psi_\tau(u) = \tau - I(u < 0)$, and we can verify that

$$\begin{aligned}
\widehat{F}_{\widehat{\mathcal{S}}}(\mathbf{x}) - \widehat{F}_{\widehat{\mathcal{S}}^c}(\mathbf{x}) &= \frac{1}{n\tau} \sum_{i=1}^{n} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leq \mathbf{x}) I\{Y_i - m_\tau(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n) < 0\} \\
&\quad - \frac{1}{n(1-\tau)} \sum_{i=1}^{n} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leq \mathbf{x}) I\{Y_i - m_\tau(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n) \geq 0\} \\
&= -\frac{1}{\tau(1-\tau)\sqrt{n}} \cdot R_n(\mathbf{x}),
\end{aligned}$$

where

$$R_n(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leq \mathbf{x}) \psi_\tau\{Y_i - m_\tau(\mathbf{X}_i, \widehat{\beta}_n)\}$$

is just the weighted cusum process of residuals in He and Zhu (2003), and their lack-of-fit test statistic is defined as the largest eigenvalue of $n^{-1}\sum_{i=1}^{n} R_n(\mathbf{X}_i) R_n^\top(\mathbf{X}_i)$. As a result, we reach He and Zhu's (2003) test.

For the two-sample problem at (4), there are a huge number of tests for the equality of two distributions in the literature, and we can always find a suitable one according to our experiences on covariates $\mathbf{X}_i$. For example, we may also consider the Kolmogorov-Smirnov test. However, these nonparametric tests, and hence the resulting lack-of-fit ones, work well only for the case with a small number of covariates p, say one or two, in real applications.

The idea here is first demonstrated in the next section to form a practical lack-of-fit test for low dimensional data, and is then used again in Section 4 for high dimensional data. For simplicity, we focus on a linear form of $m_\tau(\cdot, \cdot)$, i.e.

$$Q_\tau(Y_i \mid \mathbf{X}_i) = m_\tau(\mathbf{X}_i, \boldsymbol{\beta}) = \mathbf{X}_i^{*\top} \boldsymbol{\beta}, \tag{5}$$

where $\mathbf{X}_i^* = (1, \mathbf{X}_i^\top)^\top$, $\boldsymbol{\beta}$ is the $(p+1)$-dimensional vector, and $\boldsymbol{\beta}_0$ is its true value. All results in this paper can be readily extended to the other parametric forms.

3 Lack-of-fit test for low dimensional data

3.1 Test statistic

Consider two samples $\{\mathbf{U}_i\}$ and $\{\mathbf{V}_i\}$ with distribution functions $F_{\mathbf{U}}(\cdot)$ and $F_{\mathbf{V}}(\cdot)$, respectively. It holds that

$$E\big(\|\mathbf{U}_1 - \mathbf{V}_1\|\big) - 0.5E\big(\|\mathbf{U}_1 - \mathbf{U}_2\|\big) - 0.5E\big(\|\mathbf{V}_1 - \mathbf{V}_2\|\big) \geq 0, \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm, and the equality holds if and only if $F_{\mathbf{U}}(\cdot) = F_{\mathbf{V}}(\cdot)$. This leads to a test statistic for the equality of $F_{\mathbf{U}}(\cdot)$ and $F_{\mathbf{V}}(\cdot)$ in Baringhaus and Franz (2004), which has a reasonable power even for a moderate dimension of random vectors $\mathbf{U}_i$ and $\mathbf{V}_i$; see also Székely and Rizzo (2005) for testing multivariate normality.

By applying Baringhaus and Franz's (2004) test to hypotheses (4), we have the test statistic,

$$T_{1n} = \frac{1}{n^2\tau(1-\tau)} \sum_{i\in\mathcal{S}, j\in\mathcal{S}^c} \|\mathbf{X}_i - \mathbf{X}_j\| - \frac{0.5}{n^2\tau^2} \sum_{i,j\in\mathcal{S}} \|\mathbf{X}_i - \mathbf{X}_j\| - \frac{0.5}{n^2(1-\tau)^2} \sum_{i,j\in\mathcal{S}^c} \|\mathbf{X}_i - \mathbf{X}_j\|, \tag{7}$$

where $\varepsilon_i = Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\beta}_0$, and $\mathcal{S}$ and $\mathcal{S}^c$ are defined as in the previous section. Denote the unit sphere in $\mathbb{R}^p$ by $\mathbb{S}^{p-1} = \{b \in \mathbb{R}^p : \|b\| = 1\}$, and let $\widehat{F}_{\mathcal{S},b}(\cdot)$, $\widehat{F}_{\mathcal{S}^c,b}(\cdot)$ and $\widehat{F}_b(\cdot)$ be the empirical distributions of $\{\mathbf{X}_i^\top b, i \in \mathcal{S}\}$, $\{\mathbf{X}_i^\top b, i \in \mathcal{S}^c\}$ and $\{\mathbf{X}_i^\top b, 1 \le i \le n\}$, respectively. We then can verify that

$$T_{1n} = \gamma_p \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{+\infty} \left\{ \widehat{F}_{\mathcal{S},b}(x) - \widehat{F}_{\mathcal{S}^c,b}(x) \right\}^2 dx d\mu(b) + o_p(n^{-1}),$$

where μ is the uniform distribution on $\mathbb{S}^{p-1}$ and γ_p is a constant depending on p only, and it actually is a Cramér-type statistic. It is of interest to define its Cramér-von Mises version,

$$\int_{\mathbb{S}^{p-1}} \int_{-\infty}^{+\infty} \left\{ \widehat{F}_{\mathcal{S},b}(x) - \widehat{F}_{\mathcal{S}^c,b}(x) \right\}^2 d\widehat{F}_b(x) d\mu(b),$$

which is equivalent to the test statistic in Conde-Amboage et al. (2015). This paper will focus on the Cramér-type statistic T_{1n} since the Cramér test is usually more powerful than Cramér-von Mises one (Baringhaus and Franz, 2004), and it is also easier to calculate the value of T_{1n}.

To estimate the parameter vector, we may consider

$$\widehat{\boldsymbol{\beta}}_n = \operatorname{argmin} \sum_{i=1}^n \rho_\tau\{Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\beta}\}.$$

Let $\widehat{\varepsilon}_i = Y_i - \mathbf{X}_i^{*\top}\widehat{\boldsymbol{\beta}}_n$, and $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{S}}^c$ are defined as in the previous section. Together with (7), we can define the lack-of-fit test statistic as

$$\widehat{T}_{1n} = \frac{1}{n^2\tau(1-\tau)} \sum_{i\in\widehat{\mathcal{S}}, j\in\widehat{\mathcal{S}}^c} \|\mathbf{X}_i - \mathbf{X}_j\| - \frac{0.5}{n^2\tau^2} \sum_{i,j\in\widehat{\mathcal{S}}} \|\mathbf{X}_i - \mathbf{X}_j\| - \frac{0.5}{n^2(1-\tau)^2} \sum_{i,j\in\widehat{\mathcal{S}}^c} \|\mathbf{X}_i - \mathbf{X}_j\|,$$

where $\mathcal{S}$ and $\mathcal{S}^c$ in T_{1n} are replaced by $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{S}}^c$, respectively.

3.2 Asymptotic results

Denote by $f_{\varepsilon|\mathbf{X}}(\cdot)$ the conditional density function of ε given the covariate $\mathbf{X}$. Let $\mathbf{\Sigma}_0 = E(\mathbf{X}^*\mathbf{X}^{*\top})$, $\mathbf{\Sigma}_1 = E\{f_{\varepsilon|\mathbf{X}}(0)\mathbf{X}^*\mathbf{X}^{*\top}\}$, and

$$c_\tau = \frac{1}{2\tau(1-\tau)} E\left(\|\mathbf{X}_1 - \mathbf{X}_2\|\{f_{\varepsilon_1|\mathbf{X}_1}(0)\mathbf{X}_1^{*\top}\mathbf{\Sigma}_1^{-1}\mathbf{X}_1^* + f_{\varepsilon_2|\mathbf{X}_2}(0)\mathbf{X}_2^{*\top}\mathbf{\Sigma}_1^{-1}\mathbf{X}_2^*\}\right).$$

Theorem 1. *Suppose that Assumptions 1 and 2 at the Appendix hold. If the quantile regression model at* (5) *is correctly specified, then* $n\widehat{T}_{1n} \to_d c_\tau + \varrho_1$, *where* $\varrho_1 = 3\sum_{j=1}^{\infty}\lambda_j(\chi^2_{1j} - 1)$, $\{\lambda_i\}$ *are eigenvalues associated with the kernel* κ_0, *which is defined as in* (11) *at the Appendix, and* $\{\chi^2_{1j}\}$ *are independent Chi-square random variables with one degree of freedom.*

For model (5), since the measurement units of covariates may vary at different scenarios, it is common in real applications to standardize all or even some covariates before performing the corresponding estimation, while the fitted conditional quanitles $\widehat{Q}_\tau(Y_i \mid \mathbf{X}_i)$ are invariant. However, it may result in a different value of $\widehat{T}_{1n}$, although the partition of $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{S}}^c$ is still unchanged, i.e. the proposed test $\widehat{T}_{1n}$ may be dominated by some covariates which have much larger variances than the others. As a result, we may standardize all covariates first, i.e. the test is performed on the scaled covariates, $\{(X_{ki} - \widehat{\mu}_k)/\widehat{\sigma}_{kk}^{1/2}, i = 1, \ldots, n\}$ for $1 \le k \le p$, where $\widehat{\mu}_k = n^{-1}\sum_{i=1}^n X_{ki}$ and $\widehat{\sigma}_{kk} = n^{-1}\sum_{i=1}^n (X_{ki} - \widehat{\mu}_k)^2$. Note that $\widehat{\mu}_k$'s and $\widehat{\sigma}_{kk}$'s are all consistent and, by a method similar to the proof of Theorem 1, we then can readily derive the null distribution of the resulting test statistic accordingly.

To evaluate the asymptotic power of $\widehat{T}_{1n}$, we consider the local alternatives

$$Q_\tau(Y_i \mid \mathbf{X}_i) = \mathbf{X}_i^{*\top}\boldsymbol{\beta} + n^{-1/2}h(\mathbf{X}_i), \tag{8}$$

where $h(\cdot)$ is a nonlinear function satisfying $\min_{\mathbf{b}}\sup_{\mathbf{X}}\{h(\mathbf{X}) - \mathbf{X}^\top\mathbf{b}\}^2 > 0$ (He and Zhu, 2003). Let $c_{\boldsymbol{\beta}} = \mathbf{\Sigma}_1^{-1}E\{f_{\varepsilon|\mathbf{X}}(0)\mathbf{X}^*h(\mathbf{X})\}$. We then can obtain the Bahadur representation under the above local alternatives, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \mathbf{\Sigma}_1^{-1}\cdot n^{-1/2}\sum_{i=1}^n\psi_\tau(\varepsilon_i)\mathbf{X}_i^* + c_{\boldsymbol{\beta}} + o_p(1)$, while it has the form of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \mathbf{\Sigma}_1^{-1}\cdot n^{-1/2}\sum_{i=1}^n\psi_\tau(\varepsilon_i)\mathbf{X}_i^* + o_p(1)$ under the null hypothesis of (5).

Theorem 2. *Under the local alternatives* (8)*, if Assumptions 1 and 2 at the Appendix hold, then* $n\widehat{T}_{1n} \to_d c_\tau + \varrho_1 + \varrho_2$*, where* ϱ_2 *is a Gaussian random variable with mean zero and variance* $\{\tau(1-\tau)\}^{-3}E\{E(\|\mathbf{X}_1 - \mathbf{X}_2\| \mid \mathbf{X}_1)f_{\varepsilon_1|\mathbf{X}_1}(0)[\mathbf{X}_1^{*\top}c_{\boldsymbol{\beta}} - h(\mathbf{X}_1)]\}^2$*, and* c_τ *and* ϱ_1 *are defined as in Theorem 1.*

The above theorem shows that the test $\widehat{T}_{1n}$ has nontrivial power under the local alternatives (8).

3.3 Bootstrapping approximation

The asymptotic distribution in Theorem 1 has a complicated form since it is usually difficult to derive those eigenvalues $\{\lambda_i\}$. By adopting the wild bootstrap method in Feng et al. (2011), we suggest the following procedure to approximate this distribution.

(S1) Generate *i.i.d.* random weights $\{w_i\}$ with the distribution function satisfying Assumption 3 at the Appendix.

(S2) Generate the bootstrapped sample $\{Y_i^*\}$ with $Y_i^* = \mathbf{X}_i^{*\top}\widehat{\boldsymbol{\beta}}_n + w_i|\widehat{\varepsilon}_i|$, and calculate the bootstrapped estimator by

$$\widehat{\boldsymbol{\beta}}_n^* = \text{argmin}\sum_{i=1}^{n}\rho_\tau(Y_i^* - \mathbf{X}_i^{*\top}\boldsymbol{\beta}).$$

(S3) Let $\widehat{\varepsilon}_i^* = Y_i^* - (1, \mathbf{X}_i^\top)\widehat{\boldsymbol{\beta}}_n^*$, $\widehat{\mathcal{S}}^* = \{1 \le i \le n : \widehat{\varepsilon}_i^* < 0\}$ and $\widehat{\mathcal{S}}^{*c} = \{1 \le i \le n : \widehat{\varepsilon}_i^* \ge 0\}$. Calculate the statistic $\widehat{T}_{1n}^*(1)$ by replacing $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{S}}^c$ in the test statistic $\widehat{T}_{1n}$ with $\widehat{\mathcal{S}}^*$ and $\widehat{\mathcal{S}}^{*c}$, respectively.

(S4) Repeat Steps (S1)–(S3) $B-1$ times. We then can use the empirical distribution of $\{\widehat{T}_{1n}^*(1), ..., \widehat{T}_{1n}^*(B)\}$ to approximate the distribution of test statistic $\widehat{T}_{1n}$.

Theorem 3. *Suppose that the conditions in Theorem 1 hold. If Assumption 3 at the Appendix is further satisfied, then*

$$\sup_{x\in\mathbb{R}}\left|P^*\left(n\widehat{T}_n^* \le x\right) - P\left(n\widehat{T}_n \le x\right)\right| \to 0$$

holds in probability as $n \to \infty$*, where* P^* *is the probability measure in the bootstrapped space.*

The above theorem makes sure that the proposed procedure can be used to calculate critical values or p-values. For the random weights $\{w_i\}$, there are many distribution functions satisfying Assumption 3, while they may lead to a similar result; see simulation results in Feng et al. (2011).

4 Lack-of-fit test for high dimensional data

4.1 Test statistic

Consider the high dimensional linear quantile regression model (5) with the number of covariates p being larger than the sample size n. A sparse structure is hence assumed, and this section applies the finding at Section 2 to construct a lack-of-fit test to check whether there exists a good sparse approximation of $\mathbf{X}^{*\top}\boldsymbol{\beta}$ to $Q_\tau(Y \mid \mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta})$ and/or whether some important covariates are missed by $\mathbf{X}^{*\top}\boldsymbol{\beta}$.

We first consider a ℓ_1 penalized quantile regression estimator for model (5),

$$\widetilde{\boldsymbol{\beta}}_n = \operatorname{argmin} \sum_{i=1}^{n} \rho_\tau\{Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\beta}\} + \lambda\tau(1-\tau)\sum_{j=1}^{p} \widetilde{\sigma}_j|\beta_j|,$$

where $\widetilde{\sigma}_j = n^{-1}\sum_{k=1}^{n} X_{jk}^2$; see Belloni and Chernozhukov (2011). Let $\mathcal{D} = \{1 \leq j \leq p : \beta_{0j} \neq 0\}$ and $\widehat{\mathcal{D}} = \{1 \leq j \leq p : \widetilde{\beta}_{jn} \neq 0\}$ be the set of truly active covariates and its estimated version, respectively, where $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, ..., \beta_{0p})^\top$ is the true parameter vector, and $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{0n}, \widetilde{\beta}_{1n}, \ldots, \widetilde{\beta}_{pn})^\top$. Denote by $q = \sum_{j\in\mathcal{D}} 1$ and $\widehat{q} = \sum_{j\in\widehat{\mathcal{D}}} 1$ the cardinalities of $\mathcal{D}$ and $\widehat{\mathcal{D}}$, respectively. Without loss of generality, we rearrange the p covariates such that $\widehat{\mathcal{D}} = \{0, 1, ..., \widehat{q}\}$. The probability structure of $\widetilde{\boldsymbol{\beta}}_n$ will be involved in constructing the test statistic, while it is well known to be biased. As a result, we further assume that $\mathbf{X}$ is independent of $\varepsilon = Y - \mathbf{X}^{*\top}\boldsymbol{\beta}_0$, and then consider a de-biased estimator,

$$\widehat{\boldsymbol{\beta}}_n = \widetilde{\boldsymbol{\beta}}_n + n^{-1}\widehat{f}^{-1}(0)\widehat{\boldsymbol{\Omega}}_0^q \sum_{k=1}^{n} \mathbf{X}_k^* \psi_\tau(\widetilde{\varepsilon}_k),$$

where $\widehat{f}(\cdot)$ is an estimated density function of ε, $\widehat{\boldsymbol{\Omega}}_0^q$ is a fitted precision matrix with the last $p - \widehat{q}$ rows replaced by zeros, and $\widetilde{\varepsilon}_k = Y_k - \mathbf{X}_k^{*\top}\widetilde{\boldsymbol{\beta}}_n$; see Bradic and Kolar (2017). Note that $\widehat{\beta}_{jn} = \widetilde{\beta}_{jn} = 0$ for $\widehat{q} + 1 \leq j \leq p$, where $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{0n}, \widehat{\beta}_{1n}, \ldots, \widehat{\beta}_{pn})^\top$.

We next consider a statistic to check whether the distributions of two samples $\{\mathbf{X}_i, i \in \widehat{\mathcal{S}}\}$ and $\{\mathbf{X}_i, i \in \widehat{\mathcal{S}}^c\}$ are equal, where $\widehat{\varepsilon}_i = Y_i - \mathbf{X}_i^{*\top}\widehat{\boldsymbol{\beta}}_n$, and $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{S}}^c$ are defined as in Section 2. Note that this is a high dimensional two-sample problem and, in the literature, the equality of moments, rather than distribution functions, has been checked; see Bai and Saranadasa (1996), Schott (2007), Chen and Qin (2010), Srivastava and Yanagihara (2010), Li and Chen (2012) and reference therein.

Cai et al. (2014) and Cai et al. (2013) proposed two-sample tests for the equality of the means and variances, respectively, and they are especially designed for the case with a sparse structure. Let $\mathbf{X} = (\mathbf{X}_{\mathcal{D}}^\top, \mathbf{X}_{\mathcal{D}^c}^\top)^\top$, where $\mathbf{X}_{\mathcal{D}}$ and $\mathbf{X}_{\mathcal{D}^c}$ consist of active and inactive covariates, respectively. We can verify that the distributions of $\mathbf{X}_{\mathcal{D}^c}$ conditional on $I(\varepsilon < 0)$ and $I(\varepsilon \geq 0)$ are equal regardless of the null or alternative hypothesis. Thus, for model (5) with a sparse structure, the corresponding two-sample problem also has a sparse structure. As a result, this section uses the test statistics in Cai et al. (2014) and Cai et al. (2013) to check for the equality of the means and variances of samples $\{\mathbf{X}_i, i \in \widehat{\mathcal{S}}\}$ and $\{\mathbf{X}_i, i \in \widehat{\mathcal{S}}^c\}$, respectively.

We first adopt the method in Cai et al. (2013) to check for the equality of variances matrices. Denote the sample means by

$$\widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}} = \frac{1}{n\tau}\sum_{k\in\widehat{\mathcal{S}}}\mathbf{X}_k \quad \text{and} \quad \widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}^c} = \frac{1}{n(1-\tau)}\sum_{k\in\widehat{\mathcal{S}}^c}\mathbf{X}_k,$$

and the sample variances by

$$\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathcal{S}}} = \frac{1}{n\tau}\sum_{k\in\widehat{\mathcal{S}}}(\mathbf{X}_k - \widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}})(\mathbf{X}_k - \widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}})^\top \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\widehat{\mathcal{S}}^c} = \frac{1}{n(1-\tau)}\sum_{k\in\widehat{\mathcal{S}}^c}(\mathbf{X}_k - \widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}^c})(\mathbf{X}_k - \widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}^c})^\top.$$

Let

$$\widehat{\gamma}_{ij}(\widehat{\mathcal{S}}) = \frac{1}{n\tau}\sum_{k\in\widehat{\mathcal{S}}}\Big[\{X_{ik} - \widehat{\mu}_i(\widehat{\mathcal{S}})\}\{X_{jk} - \widehat{\mu}_j(\widehat{\mathcal{S}})\} - \widehat{\sigma}_{ij}(\widehat{\mathcal{S}})\Big]^2$$

and

$$\widehat{\gamma}_{ij}(\widehat{\mathcal{S}}^c) = \frac{1}{n(1-\tau)}\sum_{k\in\widehat{\mathcal{S}}^c}\Big[\{X_{ik} - \widehat{\mu}_i(\widehat{\mathcal{S}}^c)\}\{X_{jk} - \widehat{\mu}_i(\widehat{\mathcal{S}}^c)\} - \widehat{\sigma}_{ij}(\widehat{\mathcal{S}}^c)\Big]^2,$$

where $\widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}} = \Big(\widehat{\mu}_1(\widehat{\mathcal{S}}), ..., \widehat{\mu}_p(\widehat{\mathcal{S}})\Big)^\top$, $\widehat{\mu}_{\widehat{\mathcal{S}}^c} = \Big(\widehat{\mu}_1(\widehat{\mathcal{S}}^c), ..., \widehat{\mu}_p(\widehat{\mathcal{S}}^c)\Big)^\top$, $\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathcal{S}}} = \Big(\widehat{\sigma}_{ij}(\widehat{\mathcal{S}})\Big)_{p\times p}$ and $\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathcal{S}}^c} = \Big(\widehat{\sigma}_{ij}(\widehat{\mathcal{S}}^c)\Big)_{p\times p}$. The test statistic can be given below:

$$\widehat{M}_{\boldsymbol{\Sigma}} = \max_{1\leq i\leq j\leq p}\frac{\{\widehat{\sigma}_{ij}(\widehat{\mathcal{S}}) - \widehat{\sigma}_{ij}(\widehat{\mathcal{S}}^c)\}^2}{(n\tau)^{-1}\widehat{\gamma}_{ij}(\widehat{\mathcal{S}}) + \{n(1-\tau)\}^{-1}\widehat{\gamma}_{ij}(\widehat{\mathcal{S}}^c)}.$$

The method in Cai et al. (2014) is then applied to check the equality of the means while the variance matrices are assumed to be equal. Denote the pooled sample covariance matrix by $\widehat{\mathbf{\Sigma}} = \tau\widehat{\mathbf{\Sigma}}_{\widehat{\mathcal{S}}}+(1-\tau)\widehat{\mathbf{\Sigma}}_{\widehat{\mathcal{S}}^c}$, and we then can calculate its adaptive thresholding estimator by $\widehat{\mathbf{\Sigma}}_{ATE} = (\widehat{\sigma}_{ij}I\{|\widehat{\sigma}_{ij}| \geq \delta\sqrt{\lambda_{ij}\log(p)/n}\})_{p\times p}$, where

$$\begin{aligned}\lambda_{ij} =& \frac{1}{n}\sum_{k\in\widehat{\mathcal{S}}^c}\left[\{X_{ik}-\widehat{\mu}_i(\widehat{\mathcal{S}}^c)\}\{X_{jk}-\widehat{\mu}_i(\widehat{\mathcal{S}}^c)\}-\widehat{\sigma}_{ij}\right]^2\\ &+\frac{1}{n}\sum_{k\in\widehat{\mathcal{S}}}\left[\{X_{ik}-\widehat{\mu}_i(\widehat{\mathcal{S}})\}\{X_{jk}-\widehat{\mu}_j(\widehat{\mathcal{S}})\}-\widehat{\sigma}_{ij}\right]^2,\end{aligned}$$

$\widehat{\mathbf{\Sigma}} = (\widehat{\sigma}_{ij})_{p\times p}$, δ is a tuning parameter which can be set to $\delta = 2$, or can be selected through cross-validation empirically. As a result, the precision matrix can be estimated by $\widehat{\mathbf{\Omega}} = \widehat{\mathbf{\Sigma}}^{-1}_{ATE}$, and then the test statistic is given below:

$$\widehat{M}_{\boldsymbol{\mu}} = \frac{\kappa_n(n-\kappa_n)}{n}\max_{1\leq i\leq p}\frac{\widehat{D}_i^2}{\widehat{b}_{ii}}, \tag{9}$$

where $(\widehat{D}_1,\cdots,\widehat{D}_p)^\top = \widehat{\mathbf{\Omega}}(\widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}}-\widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}^c})$, and $\widehat{b}_{ii}$ is the ith diagonal element of the matrix $\widehat{\mathbf{\Omega}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Omega}}$. By combining $\widehat{M}_{\boldsymbol{\mu}}$ and $\widehat{M}_{\mathbf{\Sigma}}$, we then can define the lack-of-fit test statistic below:

$$\widehat{T}_{2n} = \max\left\{\widehat{M}_{\boldsymbol{\mu}} - 2\log p + \log\log p, \widehat{M}_{\mathbf{\Sigma}} - 4\log p + \log\log p\right\}.$$

4.2 Asymptotic results

Theorem 4. *Suppose that Assumptions 1 and 4-8 at the Appendix hold. If* $q^3\log^5(p\vee n) = o(n)$*, then*

$$P\left(\widehat{T}_{2n}\leq u\right)\to\exp\left(-\left\{\pi^{-1/2}+(8\pi)^{-1/2}\right\}\exp(-u/2)\right), \tag{10}$$

as $\min(n,p)\to\infty$ *under the null hypothesis that model* (5) *is correctly specified.*

From the proof of Theorem 4, we have that $\sqrt{n}(\widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}}-\widehat{\boldsymbol{\mu}}_{\widehat{\mathcal{S}}^c}) = \sqrt{n}(\widehat{\boldsymbol{\mu}}_{\mathcal{S}}-\widehat{\boldsymbol{\mu}}_{\mathcal{S}^c}) - \mathbf{\Psi}_1\sqrt{n}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)+o_p(1)$ and $\sqrt{n}\{\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\widehat{\mathcal{S}}})-\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\widehat{\mathcal{S}}^c})\} = \sqrt{n}\{\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}})-\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}^c})\} - \mathbf{\Psi}_2\sqrt{n}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)+o_p(1)$, where $\mathbf{\Psi}_1 = f(0)E(\mathbf{X}_k\mathbf{X}_k^{*\top})$, $\mathbf{\Psi}_2 = f(0)E\{(\mathbf{X}_k\otimes\mathbf{X}_k)\mathbf{X}_k^{*\top}\}$, and the partitions of $\mathcal{S}$ and $\widehat{\mathcal{S}}$ are based on the true parameter vector $\boldsymbol{\beta}_0$ and the estimator $\widehat{\boldsymbol{\beta}}_n$, respectively. These two equations still hold when $\widehat{\boldsymbol{\beta}}_n$ is replaced by $\widetilde{\boldsymbol{\beta}}_n$. Under some local alternatives, we may expect that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)$ or $\sqrt{n}(\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)$ has a deterministic shift $c_{\boldsymbol{\beta}}$ as in the low dimensional case in the previous section. From Cai et al. (2013) and Cai et al. (2014), the test statistic $\widehat{T}_{2n}$ will have the power when $c_{\boldsymbol{\beta}} = O(\sqrt{\log(p)/n})$. However, it may be difficult to derive the asymptotic behavior of $\widehat{\boldsymbol{\beta}}_n$ or $\widetilde{\boldsymbol{\beta}}_n$ under alternative hypothesis, and we leave it for future research.

Moreover, let $\boldsymbol{\mu}_{\mathcal{S}} = E(\mathbf{X}|\varepsilon < 0)$, $\boldsymbol{\Sigma}_{\mathcal{S}} = \text{var}(\mathbf{X}|\varepsilon < 0)$, $\boldsymbol{\mu}_{\mathcal{S}^c} = E(\mathbf{X}|\varepsilon \geq 0)$ and $\boldsymbol{\Sigma}_{\mathcal{S}^c} = \text{var}(\mathbf{X}|\varepsilon \geq 0)$. The proposed test $\widehat{T}_{2n}$ is to check whether $\boldsymbol{\mu}_{\mathcal{S}} = \boldsymbol{\mu}_{\mathcal{S}^c}$ and $\boldsymbol{\Sigma}_{\mathcal{S}} = \boldsymbol{\Sigma}_{\mathcal{S}^c}$ rather than to check whether $F_{\mathcal{S}}(\cdot) = F_{\mathcal{S}^c}(\cdot)$ as in the previous two sections. As a result, the proposed test statistic $\widehat{T}_{2n}$ may have a lower power for some situations, and this may be the necessary cost when the number of covariates p is much larger.

When all covariates $\mathbf{X}$ are discretely distributed, i.e. the distribution function has finite number of parameters, we may figure out a more powerful lack-of-fit test by checking the equality of conditional distribution functions rather than their first two moments.

5 Simulation studies

This section conducts two simulation experiments to assess the finite-sample performance of the proposed tests, $\widehat{T}_{1n}$ and $\widehat{T}_{2n}$, for the cases with Gaussian and heavy-tailed covariates, respectively. For the sake of comparison, we also conduct another two tests: the one in Conde-Amboage, Sanchez-Sellero, and Gonzalez-Manteiga (2015), hence denoted by CSG, and an oracle test, which refers to $\widehat{T}_{2n}$ with the sparsity structure being known in advance.

For the test statistic $\widehat{T}_{1n}$, from the Bahadur representation of $\widehat{\boldsymbol{\beta}}_n$ at Section 3, if ε_i is further assumed to be independent of $\mathbf{X}_i$, then we have that

$$\widehat{\varepsilon}_i = \varepsilon_i - \{f(0)\}^{-1}\mathbf{X}_i^{*\top}\left(\sum_{j=1}^{n}\mathbf{X}_j^{*}\mathbf{X}_j^{*\top}\right)^{-1}\mathbf{X}_i^{*}\psi_\tau(\varepsilon_i) + o_p(n^{-1/2}),$$

where $f(\cdot) := f_{\varepsilon|\mathbf{X}}(\cdot)$ is the density function of ε_i. As in Feng et al. (2011), we use the corrected residuals $\widehat{\varepsilon}_i + \{\widehat{f}(0)\}^{-1}\mathbf{X}_i^{*\top}(\sum_{j=1}^{n}\mathbf{X}_j^{*}\mathbf{X}_j^{*\top})^{-1}\mathbf{X}_i^{*}\psi_\tau(\widehat{\varepsilon}_i)$ in the bootstrapping procedure, where $\widehat{\varepsilon}_i = Y_i - \mathbf{X}_i^{*\top}\widehat{\boldsymbol{\beta}}_n$, and $\widehat{f}(\cdot)$ can be estimated from the residuals $\{\widehat{\varepsilon}_i\}$ by the kernel method in Portnoy and Koenker (1989). Moreover, the following two-point mass distribution is employed for the random weights $\{w_i\}$:

$$w_i = \begin{cases} 2(1-\tau) & \text{with probability } 1-\tau \\ -2\tau & \text{with probability } \tau \end{cases}.$$

The first experiment is for the case with Gaussian design, and the covariates $\mathbf{X}$ are generated from the multivariate normal distribution with mean zeros and covariance matrix $\mathbf{\Sigma} = (2^{-|i-j|})_{p\times p}$. The data generating process is,

$$Y_i = 1 + \sum_{j=1}^{p} \beta_j X_{ji} + \alpha(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\mathbf{X}_i = (X_{1i}, ..., X_{pi})^\top$, $\{\varepsilon_i\}$ and $\{\mathbf{X}_i\}$ are two *i.i.d.* sequences, and are independent of each other. We consider four distributions for the error term ε_i: the standard normal distribution, the exponential distribution with rate one, the Chi-squared distribution with four degrees of freedom, and the Students-t distribution with three degrees of freedom, which correspond to the symmetric, asymmetric, leptokurtic and platykurtic cases, respectively. The coefficient vector is set to $\beta_i = 1$ for $1 \le i \le q$ and $\beta_j = 0$ for $q+1 \le j \le p$ with the cardinality of truly nonzero coefficients being $q = 5$. The function $\alpha(\cdot)$ is set to zero for evaluating the size, and two alternatives are considered:

(M1) $\alpha(\mathbf{X}_i) = 0.5\big(\sum_{1\le j\le q} X_{ji}\big)^2$; and

(M2) $\alpha(\mathbf{X}_i) = 4\exp\Big(-0.5\Big(1 + \sum_{j=1}^{p} \beta_j X_{ji}\Big)\Big)$.

To estimate the quantile regression model, we use the post-ℓ_1 penalized method in Belloni and Chernozhukov (2011) along with the suggesting tuning parameters in $\widehat{T}_{2n}$, while a quantile regression estimation is performed to all covariates in $\widehat{T}_{1n}$ and CSG, and only to the truly active covariates, i.e. the first q covariates, in the oracle test. As a result, the tests $\widehat{T}_{1n}$ and CSG are not applicable when $n < p$, and the oracle test will not be affected when p increases and n is fixed.

We consider three quantile levels, $\tau = 0.25$, 0.5 and 0.75, and four combinations for sample size n and the number of covariates p: $(n, p) = (100, 20)$, $(100, 40)$, $(200, 400)$ and $(200, 1000)$, where the first two combinations refer to the case with $n > p$, while the last two are for the case with $n < p$. The number of replications is set to 500, and we use $B = 500$ for bootstrapping approximation in $\widehat{T}_{2n}$ and CSG. The significance level is 10%.

Table 3 presents the rejection rates of all four tests for the case with $n > p$, where the sample size is fixed at $n = 100$. It can be seen that, when $p = 20$, the proposed test $\widehat{T}_{2n}$ has a similar performance to its oracle counterpart, and actually the partitions of datasets due to $I(\widehat{\varepsilon}_i < 0)$ are roughly the same in these two tests. When there are more covariates, i.e. $p = 40$, $\widehat{T}_{2n}$ becomes less powerful, but it is still comparable with the oracle test. For both tests $\widehat{T}_{1n}$ and CSG, the powers drop dramatically as the number of covariate increases from $p = 20$ to 40, and actually they are not applicable when $p > n$. Roughly speaking, all tests have acceptable sizes, while the two low dimensional tests, $\widehat{T}_{1n}$ and CSG, can control the size better. In the meanwhile, $\widehat{T}_{2n}$ is slightly more sensitive at the quantile levels $\tau = 0.25$ and 0.75, while it is more conservative at $\tau = 0.5$. To provide a more insightful comparison, we further draw the ROC curves of these tests in Figures 1 and 2, respectively for $p = 20$ and 40, where the alternative model (M1) is used. The proposed test $\widehat{T}_{2n}$ dominates the two low dimensional tests, and has a more obvious advantage when the number of covariates increases from $p = 20$ to 40. Actually it is even as good as the oracle test, especially at the lower quantile levels. Moreover, the proposed low dimensional test $\widehat{T}_{1n}$ outperforms CSG for most cases, and they have a similar performance when p is larger.

Table 4 gives the rejection rates of $\widehat{T}_{2n}$ and the oracle test for the case with $n < p$, where the sample size is $n = 200$. The proposed test $\widehat{T}_{2n}$ has a comparable power with its oracle counterpart when the number of covariates is as large as $p = 400$, and still provides a comparable power even for much larger p. Figures 3 and 4 presents their ROC curves under alternative model (M1), and $\widehat{T}_{2n}$ is almost as good as the oracle test even for $p = 1000$, especially at lower quantile levels. The ROC curves under alternative model (M2) are also calculated for both cases of $n > p$ and $n < p$, and similar findings can be observed.

The second experiment is to evaluate our tests for the case with non-Gaussian covariates, and we here consider a heavy-tailed designs. Specifically, the covariate $\mathbf{X}$ is generated from the multivariate Student's t distribution $t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the same as in the first experiment, and the degrees of freedom is set to $\nu = 6$. All the other settings are the same as in the previous experiment.

Table 5 presents the rejection rates of all four tests for the case with $n > p$. The two high dimensional tests have a relatively lower power than that in the first experiment, which is under expectation since $t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 6)$ is much more heavy-tailed than the Gaussian distribution. Moreover, the low dimensional test $\widehat{T}_{1n}$ surprisingly has a even better performance at higher quantile levels, and this is further confirmed by the observations on Figures 5 and 6, which gives the ROC curves of all four tests under the alternative model (M1) with $p = 20$ and 40, respectively. This may be due to the compromise between two facts: (1) $\widehat{T}_{2n}$ is designed especially for the high dimensional data; however, (2) it aims to checking the equality of first two moments rather than that of two conditional distributions as in $\widehat{T}_{1n}$. Table 6 lists the rejection rates of $\widehat{T}_{2n}$ and the oracle test for the case with $n < p$, while Figures 7 and 8 presents their ROC curves under the alternative model (M1) with $p = 400$ and 1000, respectively. The findings are similar to those in the first experiment.

In sum, the proposed tests $\widehat{T}_{1n}$ and $\widehat{T}_{2n}$ can provide a reliable lack-of-fit check for low dimensional and high dimensional data, respectively.

6 Empirical Analysis

6.1 Sales data

This subsection attempts to study how the sale of a company can be affected by other factors. Note that the values of sales may vary at a very large range across different companies, and hence a quantile regression model may be more suitable here.

The data is sampled from Forbes 500 companies. The variables include the amount of sales in millions (Y_i), the amount of assets in millions (X_{1i}), profits in millions (X_{2i}), the number of employees in thousands (X_{3i}), the type of market the company is associated with (X_{4i}), the market value of the company in millions (X_{5i}), and the cash flow in millions (X_{6i}). All values are for the year of 1986, and there are 79 companies included. The dataset is downloaded from the website http://lib.stat.cmu.edu/DASL/Datafiles/Companies.html.

The high correlations can be observed among profits (X_{2i}), market values (X_{5i}), and the cash flow (X_{6i}), and we then involve profits (X_{2i}) in the model only. To assess the linearity assumption on relationship between sales (Y_i) and four covariate variables at different quantiles, we consider the following model,

$$Q_\tau(Y_i \mid \mathbf{X}_i) = \beta_0 + \sum_{j=1}^{4} \beta_j X_{ji}, \quad i = 1, \cdots, 79,$$

where $\mathbf{X}_i = (X_{1i}, ..., X_{4i})^\top$.

We apply our test $\widehat{T}_{1n}$ to check the lack-of-fit for the above model, and the bootstrapping procedure at Section 3 is employed to approximate the null distribution with $B = 5,000$ bootstrapped samples. The estimated p-values are 0.87, 4×10^{-4} and 0.0 at three quantile levels $\tau = 0.25$, 0.5 and 0.75, respectively. This implies that the linear regression model may fit data well for those companies with low sales, while contributions to sales from asset, profit and employee sizes may no longer linearly increases for companies with relatively high sales.

To further explore the relationship between the response and covariates at $\tau = 0.25, 0.5$ and, 0.75. We first take logarithm transformation on the data, i.e. let $\widetilde{Y}_i = \log(Y_i)$ and $\widetilde{X}_{ji} = \log(X_{ji})$ with $1 \leq j \leq 4$, and then fit a linear quantile model,

$$Q_\tau(\widetilde{Y}_i \mid \mathbf{X}_i^*) = \beta_0 + \sum_{j=1}^{4} \beta_j \widetilde{X}_{ji}, \quad i = 1, \cdots, 79.$$

The estimated p-values of our test $\widehat{T}_n$ are 0.99, 0.81 and 0.43 at quantile levels $\tau = 0.25, 0.5$ and 0.75, respectively, and this confirms the existence of nonlinearity in the model. We also perform the lack-of-fit test in He and Zhu (2003). However, all p-values are close to one, and it fails to distinguish the above two models.

6.2 GDP growth rate data

This subsection attempts to analyze the dataset in Barro and Lee (2013). The original dataset contains the economic development's statistics from 138 different countries, and they were collected quinquennially from 1950 to 2010 or averaged of five-year period over 1950-2010. The profile of a country's economic growth can be depicted by using measurements such as national accounts of people's income, education status, population and fertility, government expenditures, PPP deflators, political variables and trade policies. All economic features have been recorded in details and more extensive information can be found on http://www.barrolee.com. A subset of the original dataset is given in R package *hdm*, manufactured by Chernozhukov, Hansen, and Spindler (2016), and it consists of $n = 90$ complete observations with $p = 61$ variables. We will use this subset data to demonstrate the usefulness of our proposed test $\widehat{T}_{2n}$.

In the literature, many researchers also care about the effect of lagged level of GDP per capita on the current one. For example, in the classical Solow-Swan-Ramsey growth model, there is a hypothesis of convergence in one country's economic development, and it is said that poorer countries should see faster economic growth than richer countries, i.e. the estimated coefficient of the lagged level of GDP should be negative. As a result, we choose the current GDP growth rates per capita as response (Y), and the lagged GDP growth rates per capita together with other economic features such as black market premium, free trade openness and the other 58 characteristics are set to covariates. We then consider the following quantile regression model,

$$Q_\tau(Y_i \mid \mathbf{X}_i) = \beta_0 + \sum_{j=1}^{61} \beta_j X_{ji}, \quad i = 1, \cdots, 90,$$

where $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{61i})$. Since some covariates are skewed and/or heavy-tailed, the logarithm and cube-root transformations are conducted accordingly.

The ℓ_1-penalized quantile regression with the same settings as in the previous section is used to fit the model, and the proposed test $\widehat{T}_{2n}$ is conducted to check the lack of fit at three quantile levels $\tau = 0.25$, 0.5 and 0.75. We also compute the test in Conde-Amboage et al. (2015), denoted by CSG, for the sake of comparison, and the p-values are summarized into Table 1. It can be seen that all p-values of $\widehat{T}_{2n}$ are smaller than 5%, and then the fitted model fails to provide a good fit to the data. Belloni and Chernozhukov (2011) also mentioned that ℓ_1-penalized quantile regression doesn't pick any features at first, and we have to shrink the penalty parameter such that some economic features can be selected accordingly. We may believe that there are some important covariates missed by the fitted model. However, the CSG fails to detect the problem, and it may be due to the fact that p is close to n here.

Table 1: p-values of $\widehat{T}_{2n}$ and CSG.

Quantile levels	$\widehat{T}_{2n}$	CSG
$\tau = 0.25$	0.015	0.442
$\tau = 0.5$	0.000	0.594
$\tau = 0.75$	0.011	0.796

As a matter of fact, variable selection has been an important issue in this study since the number of observations is comparable to that of covariates. Belloni and Chernozhukov (2011) proposed using ℓ_1-penalized quantile regression with an adaptive method in choosing penalty parameter λ to select the working model. According to the suggested relaxation of λ, several covariates are chosen: lagged GDP growth rate X_1, black market premium X_2, political instability X_3, measure of tariff restriction X_4, infant mortality rate X_5, ratio of government "consumption" net of defense and education X_6, exchange rate X_7, "higher school complete" percentage in female population X_8, "secondary school complete" percentage in male population X_9, female gross enrollment ratio for higher education X_{10}, percentage of non-education in male population X_{11}, population proportion over 65 X_{12}, average years of secondary schooling in male population X_{13}. We treat the above covariates to be the truly active ones, and it then forms a low dimensional model,

$$Q_\tau(Y_i \mid \mathbf{X}_i) = \beta_0 + \sum_{j=1}^{13} \beta_j X_{ji}, \quad i = 1, \cdots, 90,$$

where $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{13i})$. We conduct the CSG and the oracle tests again, and their p-values are listed in Table 2, where the oracle test refers to $\widehat{T}_{2n}$ with the sparse structure being know in advance; see also Section 5.

It can be seen that the hypothesis of using linear model to describe the latent relationship among current GDP growth rate and lagged GDP growth rate as well as other economic features are rejected by proposed test at quantile levels $\tau = 0.25, 0.75$ at the

Table 2: p-values of CSG and oracle tests.

Quantile levels	Oracle	CSG
$\tau = 0.25$	0.058	0.502
$\tau = 0.5$	0.346	0.570
$\tau = 0.75$	0.007	0.786

significance levle of 10%. It is consistent with the intuition that low or high GDP growth rates may be related to much complicated social or political reasons, while a simple linear regression model may not be able to excavate enough information from the true underlying correspondence. In the meanwhile, we can try to use a median linear regression model to provide some insights in interpreting a country's economic features' effects on its GDP growth rate.

7 Conclusion and discussion

The main contribution of this paper is to transform lack-of-fit tests for parametric quantile regression models into checking the equality of two conditional distributions. This makes it possible to construct a reliable test according to our experiences on covariates such as the number of covariates, sample sizes, types of data (discrete or continuous covariates) etc. As an illustration, this paper gives two lack-of-fit tests for low dimensional and high dimensional data, respectively.

The tests proposed in this paper is for a fixed τ, and can be easily extended to the case with finite quantile levels. Recently more and more researches in quantile regression have been developed on an interval belonging to $(0, 1)$ (Koenker and Machado, 1999; Koenker and Xiao, 2002; Angrist et al., 2006; Escanciano and Goh, 2014; Zheng et al., 2015). It is also interesting to extend the result in this paper to this scenario, and we leave it for possible future research.

Acknowledgements

We thank the joint editor, the associate editor and three anonymous referees for their valuable comments that led to the substantial improvement in the quality of this paper. This work was partially supported by National Natural Science Foundation of China (11571218, 11690012), the State Key Program in the Major Research Plan of National Science Foundation of China (91546202), Program for Innovative Research Team of SUFE, and the Hong Kong Research Grants Councils grants (HKU702613P, 17325416).

Appendix A: Technical conditions

A.1. Assumptions for low dimensional data

Assumption 1. *It holds that, uniformly for* $\mathbf{X}_i = \mathbf{X} \in \mathbb{R}^p$, $f_{\varepsilon|\mathbf{X}}(u) - f_{\varepsilon|\mathbf{X}}(0) = O(|u|^{1/2})$ *as* $u \to 0$, *and* $f_{\varepsilon|\mathbf{X}}(0)$ *and its derivative* $f'_{\varepsilon|\mathbf{X}}(0)$ *are bounded away from both zero and infinity.*

Assumption 2. $E(\|\mathbf{X}_i\|^3) < \infty$, *and matrices* $\mathbf{\Sigma}_0$ *and* $\mathbf{\Sigma}_1$ *are positive definite.*

Assumption 3. *The* τ*th quantile of* F_w *is zero,* $\int_0^{+\infty} x^{-1} dF_w(x) = -\int_{-\infty}^0 x^{-1} dF_w(x) = 0.5$, $\int_{-\infty}^{+\infty} |x| dF_w(x) < \infty$, *and there exist two positive constant* c_1 *and* c_2 *such that* $c_1 = -\sup_{x \in (-\infty, 0]} F_w(x)$ *and* $c_2 = \inf_{x \in [0, +\infty)} F_w(x)$, *where* $F_w(\cdot)$ *is the distribution function of* w_i.

Assumption 1 restricts the conditional density of the error term ε_i, and it is commonly used in the literature of quantile regression. Assumptions 1 and 2 are similar to Conditions A1 and A2 in section 4.2 of Koenker (2005), and they make sure the existence of Bahadur representation of $\widehat{\boldsymbol{\beta}}_n$. Assumption 3 is just Conditions (Q3)–(Q5) of Feng, He, and Hu (2011).

Let $z_i = I(\varepsilon_i < 0)$, $\mathbf{Z}_i = (\mathbf{X}_i^\top, z_i)^\top$, $\phi_\tau(z_k) = \tau - z_k$ and

$$\kappa(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k) = \|\mathbf{X}_i - \mathbf{X}_j\| \{\xi_{ij} + (\tau - z_k)[\zeta_i f_{\varepsilon_i|\mathbf{X}_i}(0) \mathbf{X}_i^{*\top} \Sigma_1^{-1} \mathbf{X}_k^* + \zeta_j f_{\varepsilon_j|\mathbf{X}_j}(0) \mathbf{X}_j^{*\top} \Sigma_1^{-1} \mathbf{X}_k^*]\}.$$

Denote

$$\kappa_0(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k) = \frac{1}{3!} \sum_p \kappa(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \mathbf{Z}_{i_3}), \tag{11}$$

where $\sum_p$ is the permutation of three distinct elements $\{i, j, k\}$. It is then the kernel of the U-statistic

$$U_{1n} = \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1} \sum_{1 \le i < j < k \le n} \kappa_0(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k),$$

which is used to derive the asymptotic distribution in Theorems 1 and 2.

A.2. Assumptions for high dimensional data

Assumption 4. *$\lambda = C_1\sqrt{\log(p)/n}$ for some $C_1 > 0$, $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p(\sqrt{q\log(p \vee n)/n})$, $\mathrm{card}(\widetilde{\boldsymbol{\beta}}_n) = O_p(q)$, $\max_{1\le i\le p}\sum_{i=1}^p |b_{ij}| \le C_2$ for $C_2 > 0$ with $\mathbf{\Omega} = (b_{ij})$, and the density of ε is three times continuously differentiable at origin with the derivative $f'''(0)$ is bounded by a constant.*

Assumption 5. *There exist $C_3 > 0$ and $0 < C_4 < 1$ such that $C_3^{-1} \le \lambda_{\min}(\mathbf{\Sigma}) \le \lambda_{\max}(\mathbf{\Sigma}) \le C_3$ and $\max_{1\le i<j\le p}|r_{i,j}| \le C_4 < 1$, where $\mathbf{\Sigma}$ and $\boldsymbol{R} = (r_{i,j})$ are the covariance and correlation matrices of covariate $\mathbf{X}$, respectively.*

Assumption 6. *Suppose that there exists a subset $\Upsilon \subset \{1, 2, \ldots, p\}$ with $\mathrm{card}(\Upsilon) = o(p)$ and a constant $\alpha_0 > 0$ such that, for all $\gamma > 0, \max_{1\le j\le p, j\notin\Upsilon} s_j(\alpha_0) = o(p^\gamma)$ with $s_j(\alpha_0) := \mathrm{card}\{i : |r_{ij}| \ge (\log p)^{-1-\alpha_0}\}$. Moreover, there exist some constant $r < 1$ and a sequence of number $\Lambda_{p,r}$ such that $\mathrm{card}(\mathbf{\Lambda}(r)) \le \Lambda_{p,r} = o(p)$.*

Assumption 7. *Suppose the covariates $\mathbf{X}$ satisfies either of the following conditions*

(i.) Sub-Gaussian-type tails: Given $\log p = o(n^{1/5})$, there exist some constants $\eta > 0$ and $K > 0$ such that

$$E\{\exp(\eta(X_{ik} - \mu_i)^2/\sigma_{ii})\} \le K, \quad 1 \le i \le p.$$

(ii.) Polynomial-type tails: Given some constants $\gamma_0, c_1 > 0, p \le c_1 n^{\gamma_0}$ and for some constants $\varepsilon > 0$ and $K > 0$ such that

$$E|(X_{ik} - \mu_i)/\sigma_{ii}^{1/2}|^{4\gamma_0+4+\varepsilon} \le K, \quad 1 \le i \le p.$$

Furthermore, we assume that for a constant $\tau > 0$,

$$\min_{1\le i\le j\le p} \frac{\gamma_{ij}}{\sigma_{ii}\sigma_{jj}} \le \tau$$

holds, where $\gamma_{ij} = \mathrm{var}\{(X_{ik} - \mu_i)(X_{jk} - \mu_j)\}$.

Assumption 8. *Suppose there exist $\kappa \ge 1/3$ such that for any $i, j, l, m \in \{1, 2, \ldots, p\}$,*

$$E\{(X_{ik} - \mu_i)(X_{jk} - \mu_j)(X_{mk} - \mu_m)(X_{lk} - \mu_l)\} = \kappa(\sigma_{ij}\sigma_{ml} + \sigma_{im}\sigma_{jl} + \sigma_{il}\sigma_{jm}).$$

Assumption 4 is needed for the ℓ_1 penalized estimator $\widetilde{\boldsymbol{\beta}}_n$ and its de-biased version $\widehat{\boldsymbol{\beta}}_n$; see Bradic and Kolar (2017). Assumption 5 consists of common assumptions in the high dimensional setting; see Cai et al. (2013). Assumption 6 further restricts the correlation matrix, while Assumption 8 is used to control the fourth moment; see Cai et al. (2014). Assumption 7 specifies sub-Gaussian and polynomial-type distributions, and those families include many commonly used distributions, such as the normal and Student's t distributions; see Cai et al. (2013) and Cai et al. (2014).

References

Anderson, T. W. (1962), "On the distribution of the two-sample Cramér-von Mises criterion," *The Annals of Mathematical Statistics*, 33, 1148–1159.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006), "Quantile regression under misspecification, with an application to the US wage structure," *Econometrica*, 74, 539–563.

Bai, Z. and Saranadasa, H. (1996), "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, 6, 311–329.

Baringhaus, L. and Franz, C. (2004), "On a new multivariate two-sample test," *Journal of Multivariate Analysis*, 88, 190–206.

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

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

Bradic, J. and Kolar, M. (2017), "Uniform inference for high-dimensional quantile regression: linear functionals and regression rank scores," *arXiv:1702.06209v1.*

Cai, T., Liu, W., and Xia, Y. (2013), "Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings," *Jounral of the American Statistical Association*, 108, 265–277.

— (2014), "Two-sample test of high dimension means under dependence," *Journal of the Royal Statistical Society, Series B*, 76, 349–372.

Chen, S. and Qin, Y. (2010), "A two-sample test for high-dimensional data with applications to gene-set testing," *The Annals of Statistics*, 38, 808–835.

Chernozhukov, V., Hansen, C., and Spindler, M. (2016), "hdm: High-Dimensional Metrics," *The R Journal*, 8, 185–199.

Conde-Amboage, M., Sanchez-Sellero, C., and Gonzalez-Manteiga, W. (2015), "A lack-of-fit test for quantile regression models with high-dimensional covariates," *Computational Statistics and Data Analysis*, 88, 128–138.

Escanciano, J. C. and Goh, S. C. (2014), "Specification analysis of linear quantile models," *Journal of Econometrics*, 178, 495–507.

Escanciano, J. C. and Velasco, C. (2010), "Specification tests of parametric dynamic conditional quantiles," *Journal of Econometrics*, 159, 209–221.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Boca Raton: Chapman & Hall/CRC.

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

He, X., Wang, L., and Hong, H. G. (2013), "Quantile-adapative model-free variable screeing for high-dimensional heterogeneous data," *The Annals of Statistics*, 41, 342–369.

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

Horowitz, J. and Spokoiny, V. G. (2002), "An adaptive, rate-optimal test of linearity for median regression models," *Journal of the American Statistical Association*, 97, 822–835.

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

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

Koenker, R. and Machado, J. (1999), "Goodness-of-fit and related inference processes for quantile regression," *Journal of the American Statistical Association*, 94, 1296–1310.

Koenker, R. and Xiao, Z. (2002), "Inference on the quantile regression process," *Econometrica*, 70, 1583–1612.

Li, J. and Chen, S. (2012), "Two sample tests for high dimension covariance matrices," *The Annals of Statistics*, 40, 908–940.

Ma, S., Li, R., and Tsai, C.-L. (2017), "Variable Screening via Quantile Partial Correlation," *Journal of the American Statistical Association*, 518, 650–663.

Otsu, T. (2008), "Conditional empirical likelihood estimation and inference for quantile regression models," *Journal of Econometrics*, 142, 508–538.

Portnoy, S. and Koenker, R. (1989), "Adaptive estimation of linear models," *The annals of Statistics*, 17, 362–381.

Schott, J. R. (2007), "A test for the equality of covariance matrices when the dimension is large relative to the sample sizes," *Computational Statistics and Data Analysis*, 51, 6535–6542.

Shah, R. D. and Buhlmann, P. (2018), "Goodness-of-fit tests for high dimensional linear models," *Journal of the Royal Statistical Society, Series B*, 80, 113–135.

Srivastava, M. and Yanagihara, H. (2010), "Testing the equality of several covariance matrices with fewer observations than the dimension," *Journal of Multivariate Analysis*, 101, 1319–1329.

Stute, W. (1997), "Nonparametric model checks for regression," *The Annals of Statistics*, 25, 613–641.

Székely, G. J. and Rizzo, M. L. (2005), "A new test for multivariate normality," *Journal of Multivariate Analysis*, 93, 58–80.

Whang, Y.-J. (2006), "Smoothed empirical likelihood methods for quantile regression models," *Econometric Theory*, 22, 173–205.

Zheng, J. X. (1998), "A consistent nonparametric test of parametric regression models under onditional quantile restrictions," *Econometric Theory*, 14, 123–138.

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

Table 3: Rejection rates of $\widehat{T}_{2n}$, the oracle test, $\widehat{T}_{1n}$ and CSG for the case with Gaussian covariates and $n > p$. The nominal significance level is 10%.

		Size				Power							
						(M1)				(M2)			
ε	τ	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{1n}$	CSG	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{1n}$	CSG	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{1n}$	CSG
						$(n, p) = (100, 20)$							
N(0,1)	0.25	0.106	0.124	0.092	0.100	0.956	0.948	0.404	0.106	0.906	0.906	0.316	0.072
	0.50	0.074	0.052	0.100	0.108	0.976	0.994	0.700	0.972	0.924	0.998	0.664	0.958
	0.75	0.102	0.120	0.112	0.088	0.958	0.978	0.762	0.982	0.942	0.986	0.776	0.930
Exp(1)	0.25	0.116	0.138	0.114	0.068	0.960	0.972	0.406	0.080	0.914	0.958	0.388	0.088
	0.50	0.086	0.064	0.100	0.102	0.980	0.986	0.712	0.988	0.966	0.994	0.702	0.960
	0.75	0.104	0.118	0.120	0.078	0.916	0.978	0.776	0.990	0.962	0.986	0.788	0.982
χ_4^2	0.25	0.112	0.132	0.108	0.092	0.874	0.882	0.376	0.120	0.764	0.870	0.298	0.110
	0.50	0.070	0.072	0.114	0.108	0.910	0.956	0.658	0.920	0.688	0.920	0.598	0.836
	0.75	0.102	0.114	0.086	0.090	0.870	0.914	0.768	0.958	0.844	0.876	0.704	0.926
t_3	0.25	0.114	0.150	0.084	0.086	0.912	0.952	0.376	0.102	0.790	0.864	0.326	0.070
	0.50	0.076	0.068	0.100	0.080	0.946	0.990	0.694	0.966	0.876	0.970	0.674	0.960
	0.75	0.114	0.112	0.094	0.072	0.940	0.978	0.784	0.984	0.930	0.968	0.734	0.972
						$(n, p) = (100, 40)$							
N(0,1)	0.25	0.122	0.168	0.086	0.106	0.830	0.836	0.162	0.062	0.744	0.830	0.150	0.098
	0.50	0.096	0.074	0.102	0.094	0.890	0.950	0.310	0.104	0.776	0.940	0.296	0.118
	0.75	0.114	0.134	0.076	0.076	0.716	0.930	0.212	0.118	0.858	0.936	0.200	0.116
Exp(1)	0.25	0.122	0.164	0.122	0.086	0.828	0.898	0.146	0.096	0.806	0.874	0.162	0.086
	0.50	0.060	0.082	0.106	0.088	0.866	0.974	0.332	0.104	0.850	0.938	0.284	0.076
	0.75	0.132	0.136	0.088	0.122	0.716	0.926	0.208	0.102	0.858	0.914	0.194	0.136
χ_4^2	0.25	0.122	0.172	0.086	0.096	0.700	0.772	0.130	0.082	0.614	0.698	0.126	0.098
	0.50	0.072	0.102	0.086	0.082	0.720	0.860	0.256	0.112	0.504	0.762	0.254	0.096
	0.75	0.132	0.148	0.094	0.116	0.606	0.848	0.234	0.104	0.746	0.782	0.208	0.124
t_3	0.25	0.110	0.136	0.076	0.098	0.788	0.806	0.138	0.098	0.680	0.740	0.148	0.066
	0.50	0.080	0.092	0.092	0.092	0.838	0.952	0.288	0.096	0.758	0.928	0.280	0.094
	0.75	0.142	0.146	0.104	0.084	0.696	0.906	0.236	0.120	0.878	0.930	0.204	0.104

Table 4: Rejection rates of $\widehat{T}_{2n}$ and the oracle test for the case with Gaussian covariates and $n < p$. The nominal significance level is 10%.

		Size		Power			
				(M1)		(M2)	
ε	τ	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{2n}$	Oracle
				$(n,p)=(200,400)$			
N(0,1)	0.25	0.130	0.112	0.946	0.956	0.864	0.928
	0.50	0.084	0.068	0.986	1.000	0.936	0.998
	0.75	0.138	0.092	0.834	0.996	0.994	0.998
Exp(1)	0.25	0.128	0.110	0.938	0.976	0.954	0.968
	0.50	0.124	0.058	0.984	1.000	0.942	0.996
	0.75	0.150	0.092	0.862	0.998	0.984	0.996
χ^2_4	0.25	0.130	0.098	0.874	0.894	0.686	0.796
	0.50	0.120	0.060	0.906	0.986	0.632	0.958
	0.75	0.138	0.084	0.796	0.974	0.888	0.942
t_3	0.25	0.146	0.098	0.920	0.922	0.768	0.858
	0.50	0.126	0.062	0.978	0.998	0.872	0.992
	0.75	0.112	0.090	0.844	0.998	0.948	0.990
				$(n,p)=(200,1000)$			
N(0,1)	0.25	0.144	0.132	0.790	0.892	0.610	0.816
	0.50	0.068	0.096	0.906	0.990	0.792	0.992
	0.75	0.128	0.102	0.676	0.976	0.948	0.980
Exp(1)	0.25	0.192	0.134	0.836	0.914	0.776	0.890
	0.50	0.054	0.082	0.900	0.998	0.810	0.980
	0.75	0.154	0.086	0.672	0.986	0.940	0.980
χ^2_4	0.25	0.154	0.080	0.668	0.736	0.462	0.610
	0.50	0.090	0.072	0.742	0.934	0.356	0.834
	0.75	0.142	0.114	0.538	0.926	0.786	0.864
t_3	0.25	0.148	0.104	0.686	0.826	0.540	0.730
	0.50	0.074	0.074	0.892	0.992	0.804	0.968
	0.75	0.130	0.126	0.676	0.978	0.930	0.970

Table 5: Rejection rates of $\widehat{T}_{2n}$, the oracle test, $\widehat{T}_{1n}$ and CSG for the case with heavy-tailed covariates and $n > p$. The nominal significance level is 10%.

		Size				Power							
						(M1)				(M2)			
ε	τ	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{1n}$	CSG	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{1n}$	CSG	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{1n}$	CSG
						$(n,p) = (100, 20)$							
N(0,1)	0.25	0.110	0.100	0.112	0.062	0.716	0.774	0.358	0.028	0.618	0.716	0.358	0.038
	0.50	0.056	0.060	0.094	0.130	0.712	0.840	0.766	0.858	0.624	0.792	0.730	0.804
	0.75	0.092	0.086	0.102	0.090	0.624	0.772	0.890	0.924	0.668	0.762	0.886	0.910
Exp(1)	0.25	0.108	0.094	0.092	0.072	0.744	0.812	0.372	0.024	0.748	0.782	0.370	0.032
	0.50	0.072	0.038	0.126	0.090	0.704	0.812	0.768	0.862	0.656	0.782	0.752	0.814
	0.75	0.100	0.078	0.110	0.074	0.636	0.762	0.894	0.940	0.688	0.726	0.852	0.886
χ_4^2	0.25	0.128	0.076	0.116	0.062	0.664	0.674	0.356	0.022	0.496	0.642	0.326	0.026
	0.50	0.072	0.070	0.086	0.092	0.600	0.690	0.688	0.734	0.354	0.592	0.632	0.644
	0.75	0.100	0.090	0.094	0.076	0.636	0.642	0.844	0.884	0.502	0.606	0.856	0.850
t_3	0.25	0.114	0.090	0.076	0.086	0.696	0.736	0.366	0.030	0.618	0.644	0.294	0.026
	0.50	0.070	0.036	0.094	0.096	0.702	0.818	0.740	0.846	0.578	0.712	0.722	0.756
	0.75	0.118	0.080	0.092	0.074	0.618	0.746	0.904	0.920	0.654	0.682	0.868	0.882
						$(n,p) = (100, 40)$							
N(0,1)	0.25	0.096	0.120	0.066	0.054	0.534	0.598	0.176	0.040	0.450	0.532	0.130	0.056
	0.50	0.074	0.078	0.108	0.084	0.532	0.670	0.510	0.088	0.430	0.616	0.482	0.092
	0.75	0.138	0.106	0.088	0.072	0.498	0.624	0.540	0.138	0.574	0.624	0.502	0.128
Exp(1)	0.25	0.116	0.108	0.118	0.068	0.606	0.604	0.178	0.036	0.544	0.626	0.158	0.060
	0.50	0.078	0.082	0.088	0.100	0.538	0.670	0.506	0.096	0.432	0.626	0.462	0.054
	0.75	0.120	0.080	0.086	0.082	0.480	0.660	0.576	0.134	0.564	0.674	0.476	0.136
χ_4^2	0.25	0.108	0.102	0.098	0.062	0.476	0.546	0.162	0.040	0.356	0.480	0.126	0.048
	0.50	0.108	0.092	0.102	0.078	0.394	0.502	0.448	0.066	0.190	0.410	0.440	0.096
	0.75	0.122	0.104	0.082	0.068	0.480	0.494	0.494	0.104	0.382	0.480	0.458	0.136
t_3	0.25	0.126	0.114	0.120	0.074	0.504	0.576	0.148	0.048	0.382	0.492	0.162	0.064
	0.50	0.088	0.088	0.090	0.088	0.514	0.630	0.500	0.078	0.348	0.558	0.446	0.090
	0.75	0.142	0.112	0.076	0.080	0.442	0.636	0.524	0.124	0.518	0.610	0.508	0.148

Table 6: Rejection rates of $\widehat{T}_{2n}$ and the oracle test for the case with heavy-tailed covariates and $n < p$. The nominal significance level is 10%.

		Size		Power			
				(M1)		(M2)	
ε	τ	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{2n}$	Oracle	$\widehat{T}_{2n}$	Oracle
				$(n,p) = (200, 400)$			
N(0,1)	0.25	0.098	0.104	0.534	0.666	0.406	0.542
	0.50	0.086	0.078	0.574	0.798	0.432	0.670
	0.75	0.104	0.102	0.584	0.762	0.676	0.772
Exp(1)	0.25	0.084	0.100	0.578	0.684	0.592	0.658
	0.50	0.094	0.074	0.620	0.762	0.476	0.748
	0.75	0.080	0.096	0.538	0.780	0.696	0.742
χ_4^2	0.25	0.094	0.090	0.448	0.542	0.330	0.426
	0.50	0.084	0.086	0.362	0.550	0.188	0.414
	0.75	0.100	0.092	0.390	0.596	0.470	0.556
t_3	0.25	0.078	0.078	0.508	0.592	0.396	0.450
	0.50	0.082	0.080	0.582	0.732	0.382	0.644
	0.75	0.116	0.094	0.512	0.724	0.644	0.726
				$(n,p) = (200, 1000)$			
N(0,1)	0.25	0.110	0.110	0.456	0.422	0.316	0.396
	0.50	0.082	0.108	0.388	0.572	0.252	0.500
	0.75	0.128	0.096	0.440	0.652	0.572	0.642
Exp(1)	0.25	0.146	0.078	0.528	0.538	0.484	0.500
	0.50	0.080	0.068	0.376	0.614	0.286	0.540
	0.75	0.110	0.116	0.486	0.690	0.600	0.656
χ_4^2	0.25	0.120	0.092	0.394	0.364	0.230	0.292
	0.50	0.084	0.106	0.208	0.404	0.142	0.302
	0.75	0.114	0.088	0.352	0.550	0.436	0.482
t_3	0.25	0.098	0.116	0.352	0.398	0.284	0.302
	0.50	0.070	0.100	0.350	0.572	0.180	0.454
	0.75	0.110	0.130	0.440	0.666	0.580	0.640

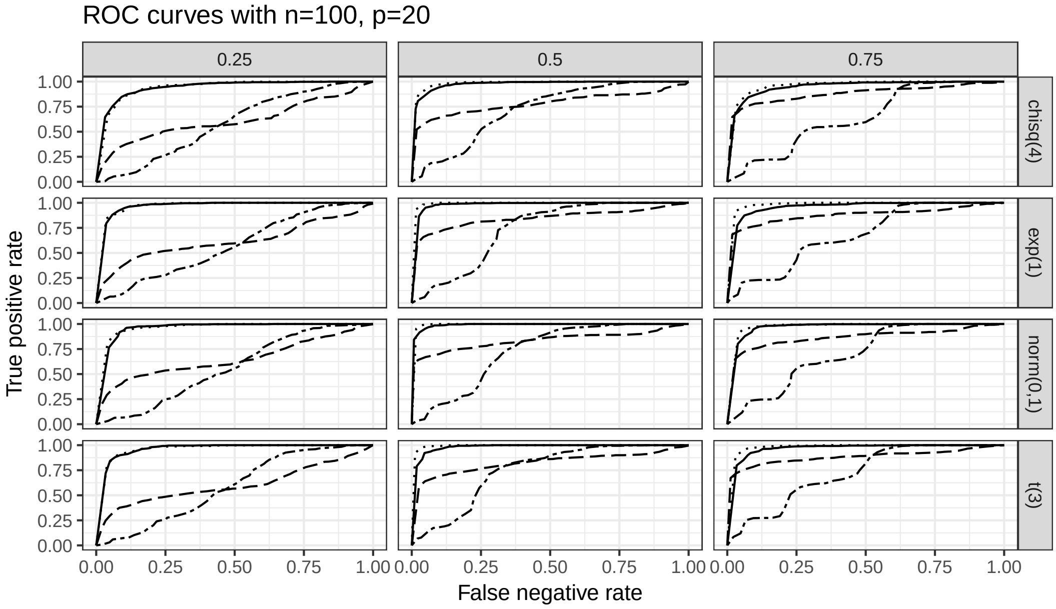

Figure 1: ROC curves of $\widehat{T}_{2n}$ (solid lines), the oracle test (dotted lines), $\widehat{T}_{1n}$ (dashed lines) and CSG (dot-dashed lines) under the alternative model (M1) with Gaussian covariates and $(n, p) = (100, 20)$.

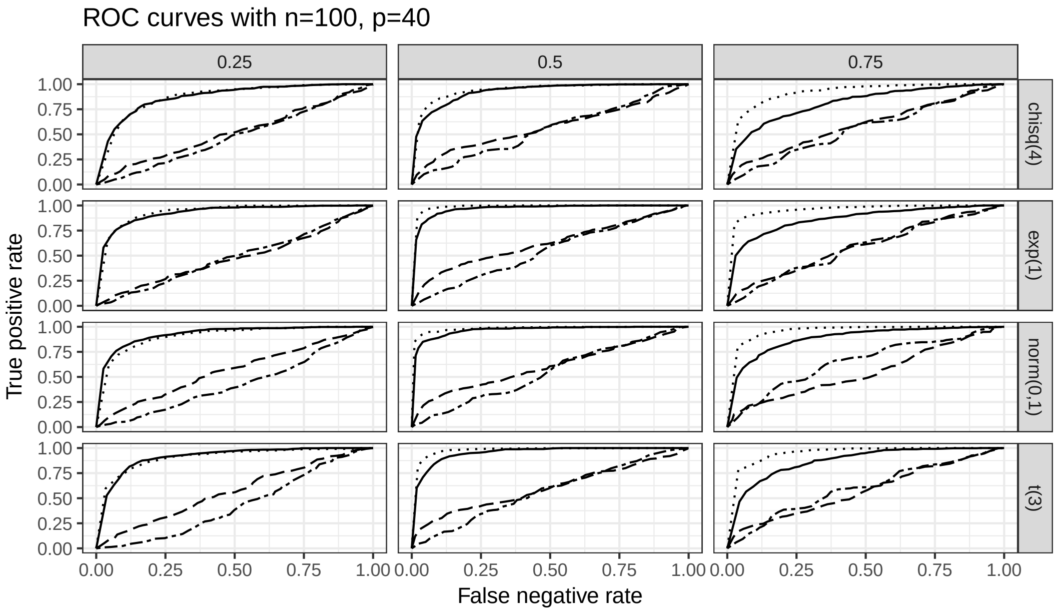

Figure 2: ROC curves of $\widehat{T}_{2n}$ (solid lines), the oracle test (dotted lines), $\widehat{T}_{1n}$ (dashed lines) and CSG (dot-dashed lines) under the alternative model (M1) with Gaussian covariates and $(n, p) = (100, 40)$.

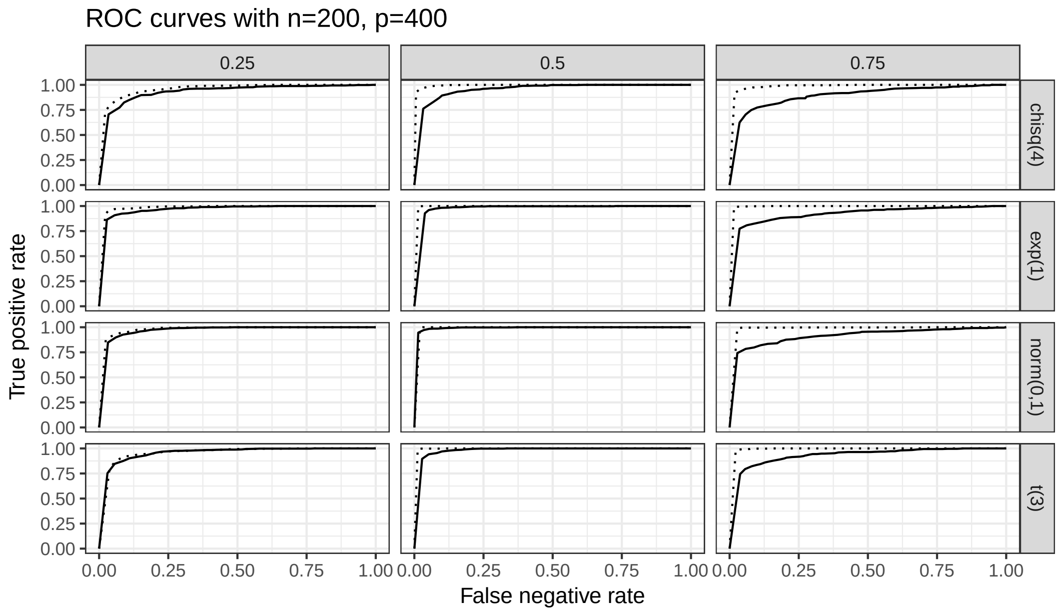

Figure 3: ROC curves of $\widehat{T}_{2n}$ (solid lines) and the oracle test (dotted lines) under the alternative model (M1) with Gaussian covariates and $(n, p) = (200, 400)$.

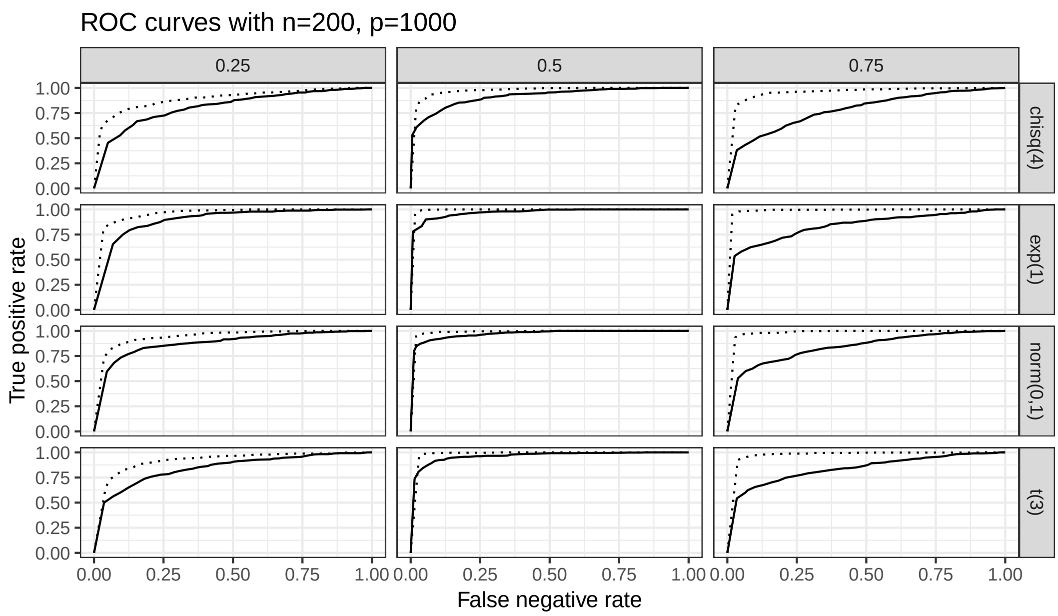

Figure 4: ROC curves of $\widehat{T}_{2n}$ (solid lines) and the oracle test (dotted lines) under the alternative model (M1) with Gaussian covariates and $(n, p) = (200, 1000)$.

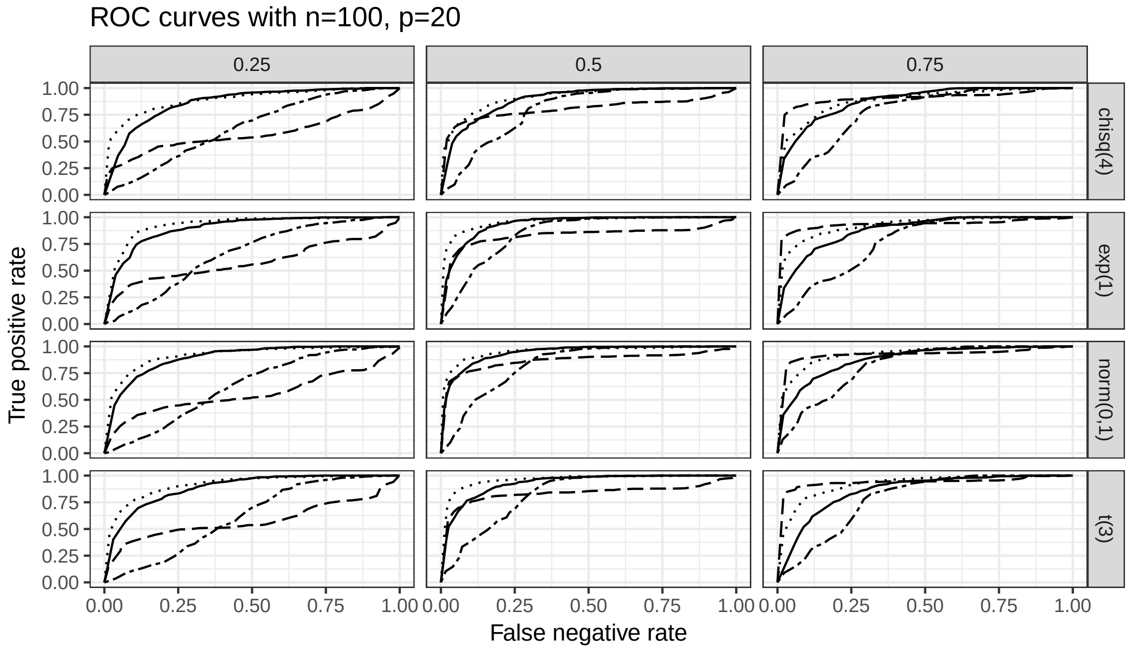

Figure 5: ROC curves of $\widehat{T}_{2n}$ (solid lines), the oracle test (dotted lines), $\widehat{T}_{1n}$ (dashed lines) and CSG (dot-dashed lines) under alternative model (M1) with heavy-tailed covariates and $(n, p) = (100, 20)$.

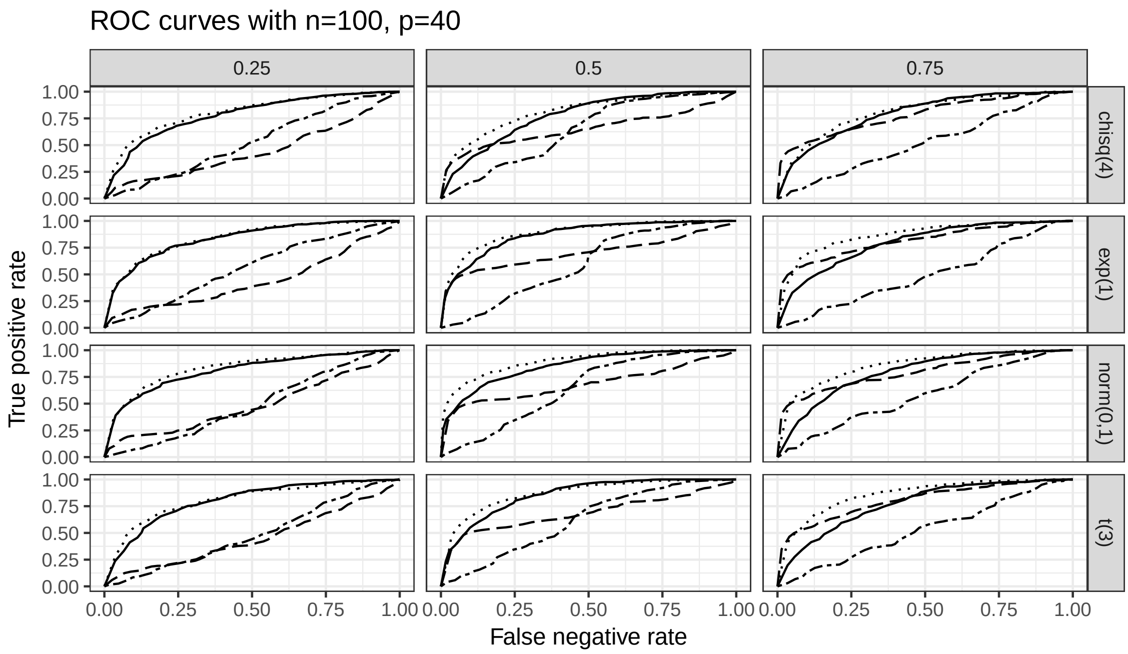

Figure 6: ROC curves of $\widehat{T}_{2n}$ (solid lines), the oracle test (dotted lines), $\widehat{T}_{1n}$ (dashed lines) and CSG (dot-dashed lines) under alternative model (M1) with heavy-tailed covariates and $(n, p) = (100, 40)$.

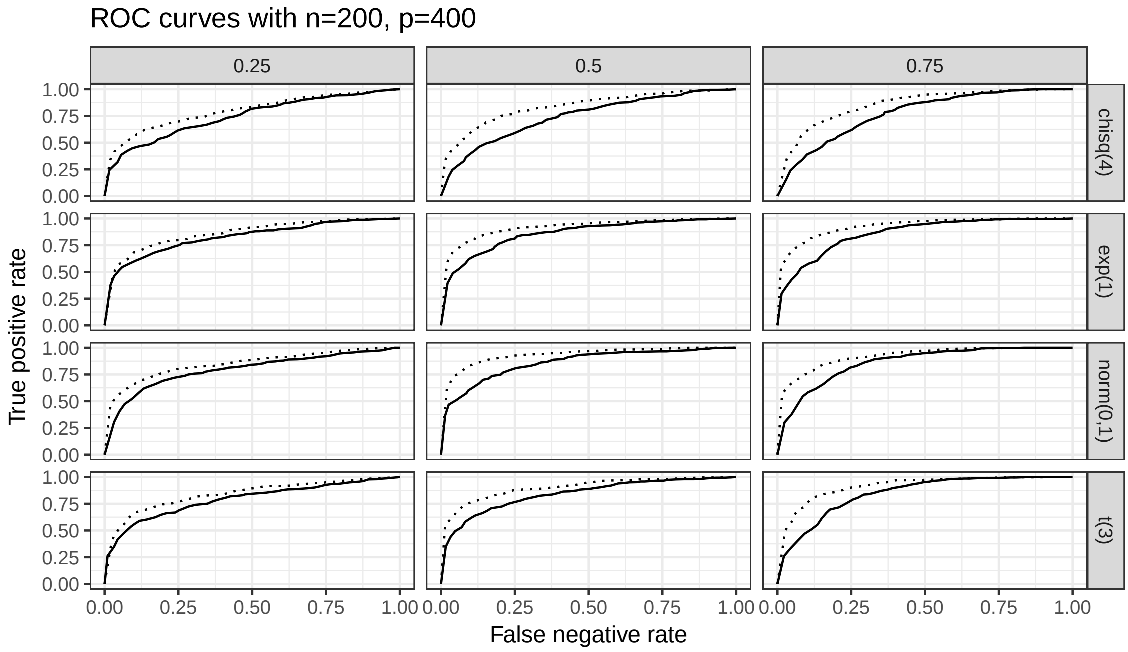

Figure 7: ROC curves of $\widehat{T}_{2n}$ (solid lines) and the oracle test (dotted lines) under the alternative model (M1) with heavy-tailed covariate and $(n, p) = (200, 400)$.

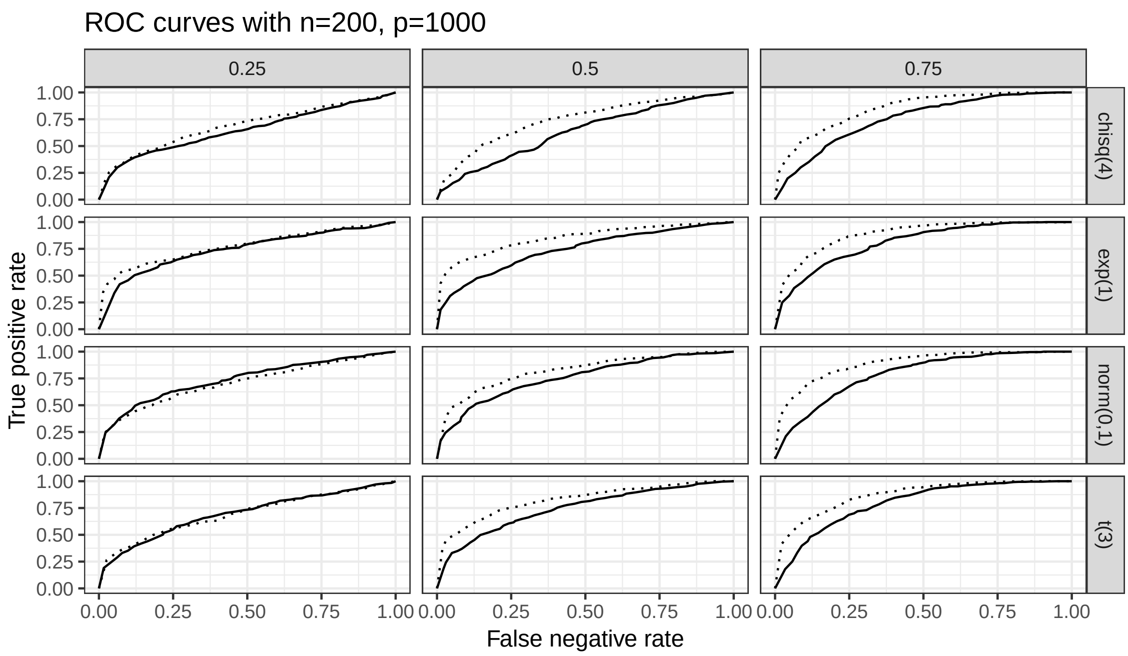

Figure 8: ROC curves of $\widehat{T}_{2n}$ (solid lines) and the oracle test (dotted lines) under the alternative model (M1) with heavy-tailed covariate and $(n, p) = (200, 1000)$.