ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2019)
81, *Part* 3, *pp.* 629–648

Lack-of-fit tests for quantile regression models

Chen Dong,

Shanghai University of Finance and Economics, People's Republic of China

Guodong Li

University of Hong Kong, People's Republic of China

and Xingdong Feng

Shanghai University of Finance and Economics, People's Republic of China

[Received June 2016. Final revision March 2019]

Summary. The paper novelly transforms lack-of-fit tests for parametric quantile regression models into checking the equality of two conditional distributions of covariates. Accordingly, by applying some successful two-sample test statistics in the literature, two tests are constructed to check the lack of fit for low and high dimensional quantile regression models. The low dimensional test works well when the number of covariates is moderate, whereas the high dimensional test can maintain the power when the number of covariates exceeds the sample size. The null distribution of the high dimensional test has an explicit form, and the *p*-values or critical values can then be calculated directly. The finite sample performance of the tests proposed is examined by simulation studies, and their usefulness is further illustrated by two real examples.

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

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 a 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 at the τth quantile of Y conditionally 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 it is flexible to use a non-parametric approach to fit it. However, this method usually has a poor performance even when p is moderate, and it is also well known to lack interpretation (Koenker, 2005; Fan and Gijbels, 1996). The parametric method, therefore, is still routinely used in quantile regression as well as in other scenarios, and specifically a parametric form will be assumed for 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 $\boldsymbol{\beta}$. Accordingly, it is an important task in the literature to perform a lack-of-fit

Address for correspondence: Xingdong Feng, School of Statistics and Management and Institute of Data Science and Statistics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, People's Republic of China.
E-mail: feng.xingdong@mail.sufe.edu.cn

 1369–7412/19/81629

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 cumulative sum 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 cumulative sum processes. These tests are all non-parametric, and they can detect the departures at all directions when the sample size tends to ∞. As a cost, the number of covariates p is limited to a small value, say 1 or 2, in real applications. Conde-Amboage *et al.* (2015) suggested projecting the covariates $\mathbf{X}$ into a random variable first, and then applying He and Zhu's (2003) 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

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

where $I(\cdot)$ is the indicator function. The aforementioned lack-of-fit tests are all based on condition (1), wheras they do not pay attention to, or do not need, another condition that the random variable $I(\varepsilon < 0)$ takes only two possible values. Consider the distribution functions of $\mathbf{X}$ conditionally on $I(\varepsilon < 0)$ and $I(\varepsilon > 0)$. We can show that equation (1) holds if and only if these two conditional distributions are equal (see lemma 1 in Section 2 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 in 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 attention 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.

$$Q_\tau(Y|\mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}^{*\mathrm{T}}\boldsymbol{\beta} \qquad \mathbf{X}^* = (1, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}, \tag{2}$$

and almost all studies in this area concentrate on the variable selection. See He *et al.* (2013), Belloni and Chernozhukov (2011), Zheng *et al.* (2015), Ma *et al.* (2017) and references therein. Shah and Bühlmann (2018) first introduced the concept of lack of fit, or goodness of fit, for high dimensional linear mean models, which can be adopted for quantile regression models. When the number of covariates p is larger than that of observations, it usually reaches the exact fit of the data, leaving no room for a discussion of lack of fit. However, 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 $\mathbf{X}^{*\mathrm{T}}\boldsymbol{\beta}$ to $m_\tau(\mathbf{X}, \boldsymbol{\beta})$, a sparse non-linear model may be more suitable than a sparse linear model. The lack of fit in this paper also refers to the case where some important covariates have been missed in the process of searching for the good sparse approximation of $\mathbf{X}^{*\mathrm{T}}\boldsymbol{\beta}$. To construct a lack-of-fit test for the high dimensional linear quantile regression model (2), although we may first consider the residual prediction method in Shah and Bühlmann (2018), it heavily depends on ordinary least squares estimation and cannot be extended to the quantile regression model. By taking advantage of the relationship between lack-of-fit tests and two-sample problems, Section 4 constructs a test by applying two high dimensional two-sample test statistics in Cai *et al.* (2013, 2014). 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 the separate on-line supplementary file, and all data sets and codes that are used in the paper can be downloaded from `https://github.com/DurandalK/qrLOFT` and are available also from

`https://rss.onlinelibrary.wiley.com/hub/journal/14679867/series-b-datasets`

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

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

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

where $m_\tau(\cdot,\cdot)$ is a known function, $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ 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 $g(\mathbf{X}) = E\{I(\varepsilon < 0)|\mathbf{X}\}$. To check whether the parametric form of model (3) is correctly specified, we can summarize the hypothesis below,

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

Denote the observed data by $\{(Y_i, \mathbf{X}_i^{\mathrm{T}})^{\mathrm{T}}, i = 1, \ldots, n\}$, which are independent and identically distributed random vectors, where $\mathbf{X}_i = (X_{1i}, \ldots, X_{pi})^{\mathrm{T}}$, and n is the number of observations. Denote $\mathcal{S} = \{1 \leqslant i \leqslant n : \varepsilon_i < 0\}$ and $\mathcal{S}^{\mathrm{c}} = \{1 \leqslant i \leqslant n : \varepsilon_i \geqslant 0\}$, where $\varepsilon_i = Y_i - m_\tau(\mathbf{X}_i, \boldsymbol{\beta}_0)$. We then can separate the observed covariates $\{\mathbf{X}_i, 1 \leqslant i \leqslant n\}$ into two samples, $\{\mathbf{X}_i, i \in \mathcal{S}\}$ and $\{\mathbf{X}_i, i \in \mathcal{S}^{\mathrm{c}}\}$, and they have the distributions $F_{\mathcal{S}}(\mathbf{x}) = P(\mathbf{X} < \mathbf{x}|\varepsilon < 0)$ and $F_{\mathcal{S}^{\mathrm{c}}}(\mathbf{x}) = P(\mathbf{X} < \mathbf{x}|\varepsilon \geqslant 0)$ respectively.

Lemma 1. It holds that

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

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

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

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

The true parameter vector $\boldsymbol{\beta}_0$ is unknown, but we may estimate it by

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

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

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

$$\hat{F}_{\hat{\mathcal{S}}}(\mathbf{x}) = \frac{1}{n\tau} \sum_{i \in \hat{\mathcal{S}}} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leqslant \mathbf{x})$$

and

$$\hat{F}_{\hat{\mathcal{S}}^c}(\mathbf{x}) = \frac{1}{n(1-\tau)} \sum_{i \in \hat{\mathcal{S}}^c} \omega(\mathbf{X}_i) I(\mathbf{X}_i \leqslant \mathbf{x})$$

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

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

where

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

is just the weighted cumulative sum process of residuals in He and Zhu (2003), and their lack-of-fit test statistic is defined as the largest eigenvalue of $n^{-1}\sum_{i=1}^{n} R_n(\mathbf{X}_i) R_n^{\mathrm{T}}(\mathbf{X}_i)$. Therefore, we obtain He and Zhu's (2003) test statistic.

For the two-sample problem (4), there are a huge number of tests for the equality of two distributions in the literature, and we can always find a suitable test according to our experiences in covariates $\mathbf{X}_i$. For example, we may consider the Kolmogorov–Smirnov test. However, these non-parametric tests, and hence the resulting lack-of-fit tests, work well only for the case with a small number of covariates p, say 1 or 2, 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 it is used again in Section 4 for high dimensional data. For simplicity, we focus on a linear form of $m_\tau(\cdot, \cdot)$, i.e.

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

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

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

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 $\varepsilon_i = Y_i - \mathbf{X}_i^{*\mathrm{T}}\boldsymbol{\beta}_0$, and $\mathcal{S}$ and $\mathcal{S}^{\mathrm{c}}$ 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 $\hat{F}_{\mathcal{S},b}(\cdot)$, $\hat{F}_{\mathcal{S}^{\mathrm{c}},b}(\cdot)$ and $\hat{F}_b(\cdot)$ be the empirical distributions of $\{\mathbf{X}_i^{\mathrm{T}}b, i \in \mathcal{S}\}$, $\{\mathbf{X}_i^{\mathrm{T}}b, i \in \mathcal{S}^{\mathrm{c}}\}$ and $\{\mathbf{X}_i^{\mathrm{T}}b, 1 \leqslant i \leqslant n\}$ respectively. We then can verify that

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

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:

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

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 the Cramér–von Mises test (Baringhaus and Franz, 2004), and it is also easier to calculate the value of T_{1n}.

To estimate the parameter vector, we may consider

$$\hat{\boldsymbol{\beta}}_n = \arg\min \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i^{*\mathrm{T}}\boldsymbol{\beta}).$$

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

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

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

3.2. Asymptotic results

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

$$c_\tau = \frac{1}{2\tau(1-\tau)} E[\|\mathbf{X}_1 - \mathbf{X}_2\|\{f_{\varepsilon_1|\mathbf{X}_1}(0)\mathbf{X}_1^{*\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\mathbf{X}_1^* + f_{\varepsilon_2|\mathbf{X}_2}(0)\mathbf{X}_2^{*\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\mathbf{X}_2^*\}].$$

Theorem 1. Suppose that assumptions 1 and 2 in Appendix A hold. If the quantile regression model (5) is correctly specified, then $n\hat{T}_{1n} \to_{\mathrm{d}} c_\tau + \varrho_1$, where $\varrho_1 = 3\Sigma_{j=1}^{\infty}\lambda_j(\chi_{1j}^2 - 1)$, $\{\lambda_i\}$ are eigenvalues that are associated with the kernel κ_0, which is defined as in equation (11) in Appendix A, and $\{\chi_{1j}^2\}$ are independent χ^2 random variables with 1 degree of freedom.

For model (5), since the measurement units of covariates may vary in different scenarios, it is common in real applications to standardize some or even all covariates before performing the corresponding estimation, whereas the fitted conditional quantiles $\hat{Q}_\tau(Y_i|\mathbf{X}_i)$ are invariant. However, it may result in a different value of $\hat{T}_{1n}$, although the partition of $\hat{\mathcal{S}}$ and $\hat{\mathcal{S}}^{\mathrm{c}}$ is still unchanged, i.e. the proposed test $\hat{T}_{1n}$ 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, $\{(X_{ki} - \hat{\mu}_k)/\hat{\sigma}_{kk}^{1/2}, i = 1, \ldots, n\}$ for $1 \leqslant k \leqslant p$, where $\hat{\mu}_k = n^{-1}\Sigma_{i=1}^n X_{ki}$ and $\hat{\sigma}_{kk} = n^{-1}\Sigma_{i=1}^n (X_{ki} - \hat{\mu}_k)^2$. Note that $\hat{\mu}_k$s and $\hat{\sigma}_{kk}$s are all consistent and, by a method similar to the proof of theorem 1, we can readily derive the asymptotic distribution of the resulting test statistic under the null hypothesis.

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

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

where $h(\cdot)$ is a non-linear function satisfying $\min_{\mathbf{b}} \sup_{\mathbf{X}} \{h(\mathbf{X}) - \mathbf{X}^{\mathrm{T}}\mathbf{b}\}^2 > 0$ (He and Zhu, 2003). Let $c_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_1^{-1} E\{f_{\varepsilon|\mathbf{X}}(0)\mathbf{X}^* h(\mathbf{X})\}$. We then can obtain the Bahadur representation under the above local alternatives,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \boldsymbol{\Sigma}_1^{-1} n^{-1/2} \sum_{i=1}^{n} \psi_\tau(\varepsilon_i)\mathbf{X}_i^* + c_{\boldsymbol{\beta}} + o_p(1),$$

whereas it has the form

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \boldsymbol{\Sigma}_1^{-1} n^{-1/2} \sum_{i=1}^{n} \psi_\tau(\varepsilon_i)\mathbf{X}_i^* + o_p(1)$$

under the null hypothesis (5).

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

Theorem 2 shows that the test $\hat{T}_{1n}$ has non-trivial 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.

Step 1: generate independent and identically distributed random weights $\{w_i\}$ with the distribution function satisfying assumption 3 in Appendix A.
Step 2: generate the bootstrapped sample $\{Y_i^*\}$ with $Y_i^* = \mathbf{X}_i^{*\mathrm{T}}\hat{\boldsymbol{\beta}}_n + w_i|\hat{\varepsilon}_i|$, and calculate the bootstrapped estimator by

$$\hat{\boldsymbol{\beta}}_n^* = \arg\min \sum_{i=1}^{n} \rho_\tau(Y_i^* - \mathbf{X}_i^{*\mathrm{T}}\boldsymbol{\beta}).$$

Step 3: let $\hat{\varepsilon}_i^* = Y_i^* - (1, \mathbf{X}_i^{\mathrm{T}})\hat{\boldsymbol{\beta}}_n^*$, $\hat{\mathcal{S}}^* = \{1 \leqslant i \leqslant n : \hat{\varepsilon}_i^* < 0\}$ and $\hat{\mathcal{S}}^{*\mathrm{c}} = \{1 \leqslant i \leqslant n : \hat{\varepsilon}_i^* \geqslant 0\}$. Calculate the statistic $\hat{T}_{1n}^*(1)$ by replacing $\hat{\mathcal{S}}$ and $\hat{\mathcal{S}}^{\mathrm{c}}$ in the test statistic $\hat{T}_{1n}$ with $\hat{\mathcal{S}}^*$ and $\hat{\mathcal{S}}^{*\mathrm{c}}$ respectively.
Step 4: repeat steps 1–3 $B-1$ times. We then can use the empirical distribution of $\{\hat{T}_{1n}^*(1), \ldots, \hat{T}_{1n}^*(B)\}$ to approximate the distribution of test statistic $\hat{T}_{1n}$.

Theorem 3. Suppose that the conditions in theorem 1 hold. If assumption 3 in Appendix A is further satisfied, then

$$\sup_{x \in \mathbb{R}} |P^*(n\hat{T}_n^* \leqslant x) - P(n\hat{T}_n \leqslant x)| \to 0$$

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

Theorem 3 makes sure that the procedure proposed can be used to calculate critical values or p-values. For the random weights $\{w_i\}$, although there are many distribution functions satisfying assumption 3, they may lead to a similar result (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 in Section 2 to construct a lack-of-fit test to check whether there is a good sparse approximation of $\mathbf{X}^{*\mathrm{T}}\boldsymbol{\beta}$ to $Q_\tau(Y|\mathbf{X}) = m_\tau(\mathbf{X}, \boldsymbol{\beta})$ and/or whether some important covariates are missed by $\mathbf{X}^{*\mathrm{T}}\boldsymbol{\beta}$.

We first consider an l_1-penalized quantile regression estimator for model (5):

$$\tilde{\boldsymbol{\beta}}_n = \arg\min \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i^{*\mathrm{T}}\boldsymbol{\beta}) + \lambda\tau(1-\tau)\sum_{j=1}^{p} \tilde{\sigma}_j|\beta_j|,$$

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

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

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

We next consider a test statistic to check whether the distributions of two samples $\{\mathbf{X}_i, i \in \hat{\mathcal{S}}\}$ and $\{\mathbf{X}_i, i \in \hat{\mathcal{S}}^{\mathrm{c}}\}$ are equal, where $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i^{*\mathrm{T}}\hat{\boldsymbol{\beta}}_n$, and $\hat{\mathcal{S}}$ and $\hat{\mathcal{S}}^{\mathrm{c}}$ are defined as in Section 2. 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) and Li and Chen (2012) and references 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 $\mathbf{X} = (\mathbf{X}_{\mathcal{D}}^{\mathrm{T}}, \mathbf{X}_{\mathcal{D}^{\mathrm{c}}}^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{X}_{\mathcal{D}}$ and $\mathbf{X}_{\mathcal{D}^{\mathrm{c}}}$ consist of active and inactive covariates respectively. We can verify that the distributions of $\mathbf{X}_{\mathcal{D}^{\mathrm{c}}}$ conditionally on $I(\varepsilon < 0)$ and $I(\varepsilon \geqslant 0)$ are equal regardless of whether under the null or alternative hypothesis. Thus, for model (5) with a sparse structure, the corresponding two-sample problem also has a sparse structure. This section, therefore, 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 $\{\mathbf{X}_i, i \in \hat{\mathcal{S}}\}$ and $\{\mathbf{X}_i, i \in \hat{\mathcal{S}}^{\mathrm{c}}\}$ respectively.

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

$$\hat{\boldsymbol{\mu}}_{\hat{\mathcal{S}}} = \frac{1}{n\tau} \sum_{k\in\hat{\mathcal{S}}} \mathbf{X}_k$$

and

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

and the sample variances by

$$\hat{\mathbf{\Sigma}}_{\hat{\mathcal{S}}} = \frac{1}{n\tau} \sum_{k \in \hat{\mathcal{S}}} (\mathbf{X}_k - \hat{\boldsymbol{\mu}}_{\hat{\mathcal{S}}})(\mathbf{X}_k - \hat{\boldsymbol{\mu}}_{\hat{\mathcal{S}}})^{\mathrm{T}}$$

and

$$\hat{\mathbf{\Sigma}}_{\hat{\mathcal{S}}^{\mathrm{c}}} = \frac{1}{n(1-\tau)} \sum_{k \in \hat{\mathcal{S}}^{\mathrm{c}}} (\mathbf{X}_k - \hat{\boldsymbol{\mu}}_{\hat{\mathcal{S}}^{\mathrm{c}}})(\mathbf{X}_k - \hat{\boldsymbol{\mu}}_{\hat{\mathcal{S}}^{\mathrm{c}}})^{\mathrm{T}}.$$

Let

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

and

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

where $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{S}}} = (\hat{\mu}_1(\hat{\mathcal{S}}), \ldots, \hat{\mu}_p(\hat{\mathcal{S}}))^{\mathrm{T}}$, $\hat{\mu}_{\hat{\mathcal{S}}^{\mathrm{c}}} = (\hat{\mu}_1(\hat{\mathcal{S}}^{\mathrm{c}}), \ldots, \hat{\mu}_p(\hat{\mathcal{S}}^{\mathrm{c}}))^{\mathrm{T}}$, $\hat{\mathbf{\Sigma}}_{\hat{\mathcal{S}}} = (\hat{\sigma}_{ij}(\hat{\mathcal{S}}))_{p\times p}$ and $\hat{\mathbf{\Sigma}}_{\hat{\mathcal{S}}^{\mathrm{c}}} = (\hat{\sigma}_{ij}(\hat{\mathcal{S}}^{\mathrm{c}}))_{p\times p}$. The test statistic is given by

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

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

$$\lambda_{ij} = \frac{1}{n} \sum_{k \in \hat{\mathcal{S}}^{\mathrm{c}}} [\{X_{ik} - \hat{\mu}_i(\hat{\mathcal{S}}^{\mathrm{c}})\}\{X_{jk} - \hat{\mu}_i(\hat{\mathcal{S}}^{\mathrm{c}})\} - \hat{\sigma}_{ij}]^2 + \frac{1}{n} \sum_{k \in \hat{\mathcal{S}}} [\{X_{ik} - \hat{\mu}_i(\hat{\mathcal{S}})\}\{X_{jk} - \hat{\mu}_j(\hat{\mathcal{S}})\} - \hat{\sigma}_{ij}]^2,$$

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

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

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

$$\hat{T}_{2n} = \max[\hat{M}_{\boldsymbol{\mu}} - 2\log(p) + \log\{\log(p)\}, \hat{M}_{\mathbf{\Sigma}} - 4\log(p) + \log\{\log(p)\}].$$

4.2. Asymptotic results

Theorem 4. Suppose that assumptions 1 and 4–8 in the Appendix A hold. If $q^3\log^5(p \vee n) = o(n)$, then

$$P(\hat{T}_{2n} \leqslant u) \to \exp[-\{\pi^{-1/2} + (8\pi)^{-1/2}\}\exp(-u/2)], \tag{10}$$

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

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

In addition, let $\boldsymbol{\mu}_{\mathcal{S}} = E(\mathbf{X}|\varepsilon < 0)$, $\boldsymbol{\Sigma}_{\mathcal{S}} = \mathrm{var}(\mathbf{X}|\varepsilon < 0)$, $\boldsymbol{\mu}_{\mathcal{S}^c} = E(\mathbf{X}|\varepsilon \geqslant 0)$ and $\boldsymbol{\Sigma}_{\mathcal{S}^c} = \mathrm{var}(\mathbf{X}|\varepsilon \geqslant 0)$. The proposed test $\hat{T}_{2n}$ 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 $\hat{T}_{2n}$ 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 a 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, $\hat{T}_{1n}$ and $\hat{T}_{2n}$, for the cases with Gaussian and heavy-tailed covariates respectively. For comparison, we also conduct another two tests: the test in Conde-Amboage *et al.* (2015), hence denoted by CSG, and an oracle test, which refers to $\hat{T}_{2n}$ with the sparsity structure being known in advance.

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

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

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 $\hat{\varepsilon}_i + \hat{f}(0)^{-1}\mathbf{X}_i^{*\mathrm{T}}(\Sigma_{j=1}^n\mathbf{X}_j^*\mathbf{X}_j^{*\mathrm{T}})^{-1}\mathbf{X}_i^*\psi_\tau(\hat{\varepsilon}_i)$ in the bootstrapping procedure, where $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i^{*\mathrm{T}}\hat{\boldsymbol{\beta}}_n$, and $\hat{f}(\cdot)$ can be estimated from the residuals $\{\hat{\varepsilon}_i\}$ by the kernel method in Portnoy and Koenker (1989). In addition, 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 the Gaussian design, and the covariates $\mathbf{X}$ are generated from the multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma} = (2^{-|i-j|})_{p\times p}$. The data-generating process is

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

where $\mathbf{X}_i = (X_{1i}, \ldots, X_{pi})^{\mathrm{T}}$, $\{\varepsilon_i\}$ and $\{\mathbf{X}_i\}$ are two independent and identically distributed sequences and are independent of each other. We consider four distributions for the error term ε_i:

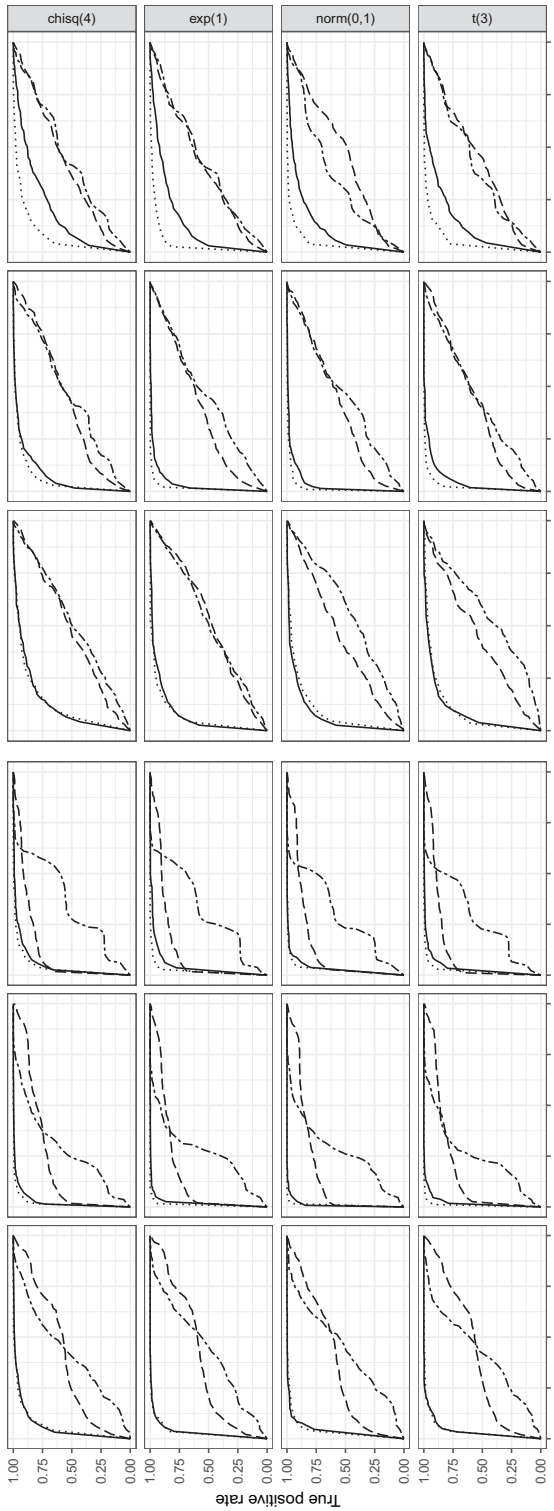

Fig. 1. ROC curves of $\hat{T}_{2n}$ (———), the oracle test (· · · · ·), $\hat{T}_{1n}$ (– – –) and CSG (· – · –) under the alternative model 1 with Gaussian covariates and (a)–(c) $(n,p)=(100,20)$ or (d)–(f) $(n,p)=(100,40)$: (a), (d) $\tau=0.25$; (b), (e) $\tau=0.5$; (c), (f) $\tau=0.75$

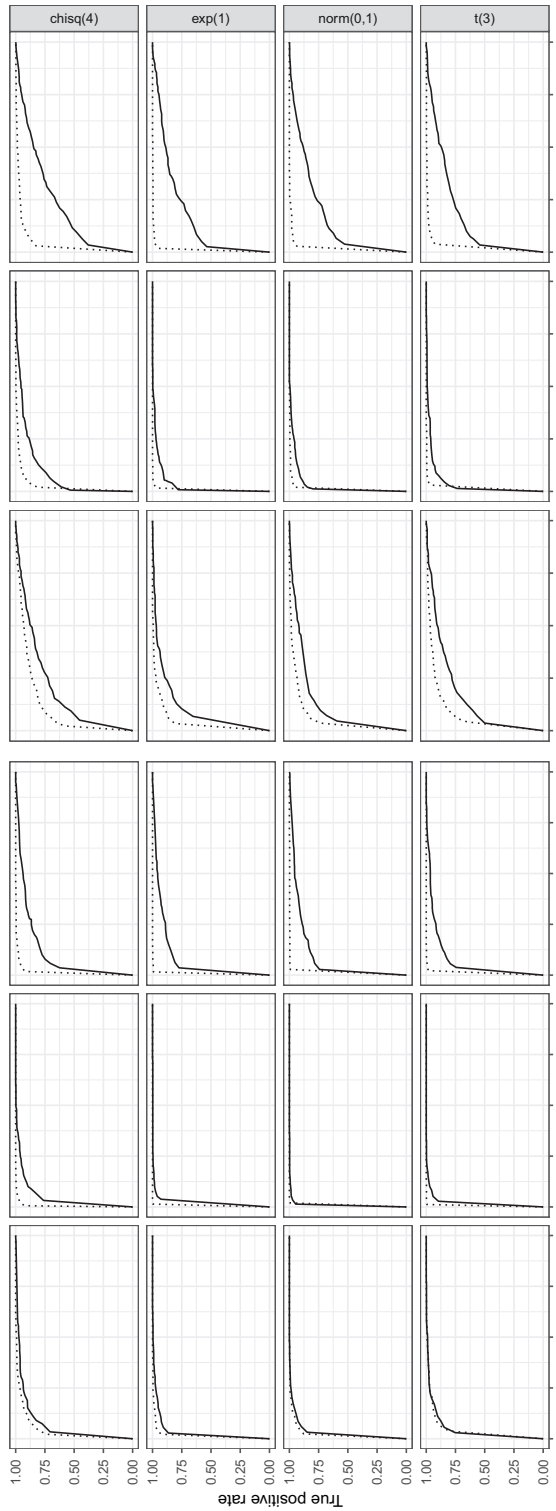

Fig. 2. ROC curves of $\hat{T}_{2n}$ (———) and the oracle test (· · · · ·) under the alternative model 1 with Gaussian covariates and (a)–(c) $(n, p) = (200, 400)$ or (d)–(f) $(n, p) = (200, 1000)$: (a), (d) $\tau = 0.25$; (b), (e) $\tau = 0.5$; (c), (f) $\tau = 0.75$

the standard normal distribution, the exponential distribution with rate 1, the χ^2-distribution with 4 degrees of freedom and the Student t-distribution with 3 degrees of freedom, which correspond to the symmetric, asymmetric, leptokurtic and platykurtic cases respectively. The coefficient vector is set to $\beta_i = 1$ for $1 \leqslant i \leqslant q$ and $\beta_j = 0$ for $q+1 \leqslant j \leqslant p$ with the cardinality of truly non-zero coefficients being $q = 5$. The function $\alpha(\cdot)$ is set to 0 for evaluating the size, and two alternatives are considered:

(a) $\alpha(\mathbf{X}_i) = 0.5(\Sigma_{1 \leqslant j \leqslant q} X_{ji})^2$ (model 1) and
(b) $\alpha(\mathbf{X}_i) = 4\exp\{-0.5(1 + \Sigma_{j=1}^{p} \beta_j X_{ji})\}$ (model 2).

To estimate the quantile regression model, we use the post-l_1-penalized method in Belloni and Chernozhukov (2011) along with the suggested tuning parameters in $\hat{T}_{2n}$. A quantile regression estimation is performed on all covariates in $\hat{T}_{1n}$ and CSG, but only on the truly active covariates, i.e. the first q covariates, in the oracle test. As a result, the tests $\hat{T}_{1n}$ 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, $\tau = 0.25, 0.5, 0.75$, and four combinations for sample size n and the number of covariates p: $(n, p) = (100, 20)$, $(100, 40)$, $(200, 400)$, $(200, 1000)$, where the first two combinations refer to the case with $n > p$, whereas the last two are for the case with $n < p$. The number of replications is set to 500, and we use $B = 500$ for the bootstrapping approximation in $\hat{T}_{1n}$ and CSG.

Fig. 1 gives the receiver operating characteristic (ROC) curves of all four tests under the alternative model 1 for the case of $n > p$, where the sample size is fixed at $n = 100$. The proposed test $\hat{T}_{2n}$ dominates the two low dimensional tests, and it is more obvious when the number of covariates increases from $p = 20$ to $p = 40$. Actually it is even as good as the oracle test, especially at the lower quantile levels. Moreover, the proposed low dimensional test $\hat{T}_{1n}$ outperforms CSG for most cases with a similar performance when p is larger. Fig. 2 presents the ROC curves of $\hat{T}_{2n}$ and the oracle test for the case with $n < p$, where the sample size is $n = 200$. It can be seen that $\hat{T}_{2n}$ is comparable with the oracle test even for $p = 1000$, especially at lower quantile levels. The ROC curves under the alternative model (2) 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 a heavy-tailed design is considered. Specifically, the covariates $\mathbf{X}$ are generated from the multivariate Student 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 are set to $\nu = 6$. All the other settings are the same as in the previous experiment.

Figs 3 and 4 give the ROC curves under the alternative model 1 for the cases of $n > p$ and $n < p$ respectively. Surprisingly the low dimensional test $\hat{T}_{1n}$ has an even better performance than the high dimensional test $\hat{T}_{2n}$ at higher quantile levels. This may be due to the compromise between two facts:

(a) $\hat{T}_{2n}$ is designed especially for the high dimensional data; however,
(b) it aims to check the equality of the first two moments rather than that of two conditional distributions as in $\hat{T}_{1n}$.

The other findings are similar to those in the first experiment.

The on-line supplementary file further reports the rejection rates of these tests at the level of significance of 10%, and we have the following findings.

(a) Generally speaking, all tests have acceptable sizes, whereas the two low dimensional tests $\hat{T}_{1n}$ and CSG can control the size better;

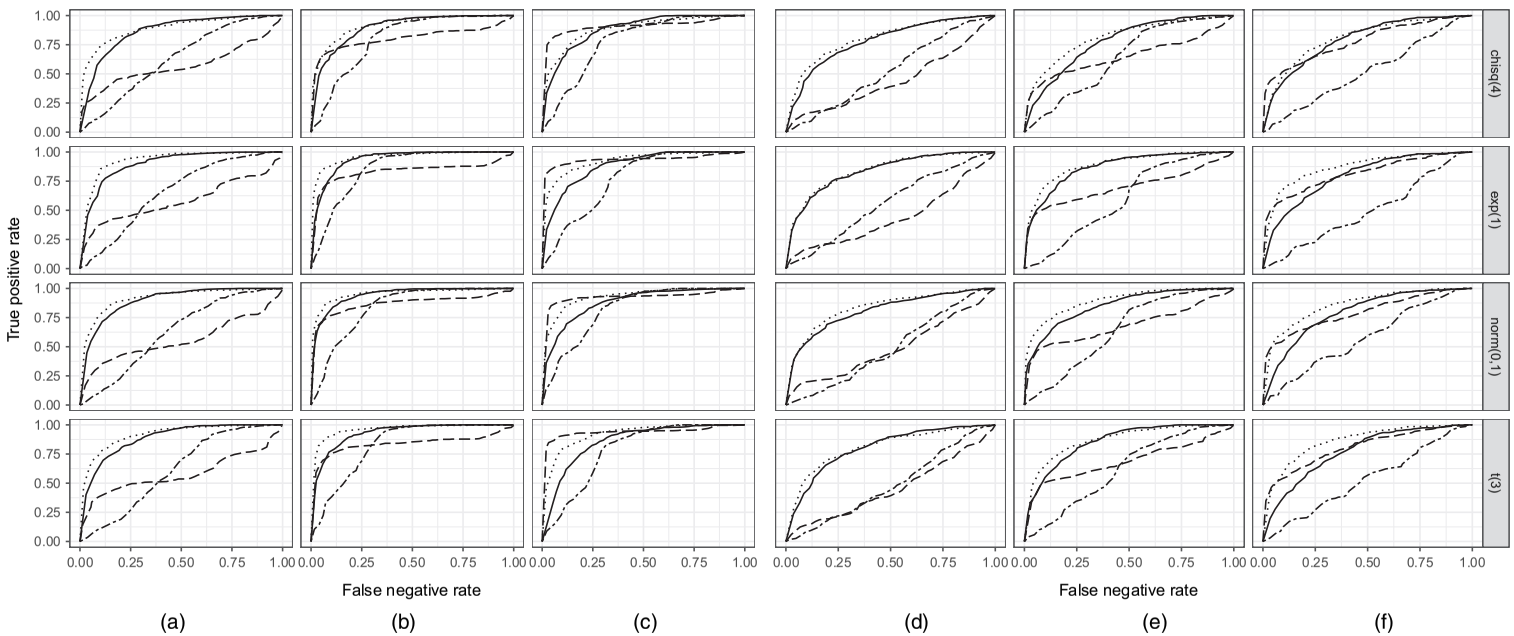

Fig. 3. ROC curves of $\hat{T}_{2n}$ (———), the oracle test (· · · · · · ·), $\hat{T}_{1n}$ (– – –) and CSG (· – · –) under the alternative model 1 with heavy-tailed covariates and (a)–(c) $(n, p) = (100, 20)$ or (d)–(f) $(n, p) = (100, 40)$: (a), (d) $\tau = 0.25$; (b), (e) $\tau = 0.5$; (c), (f) $\tau = 0.75$

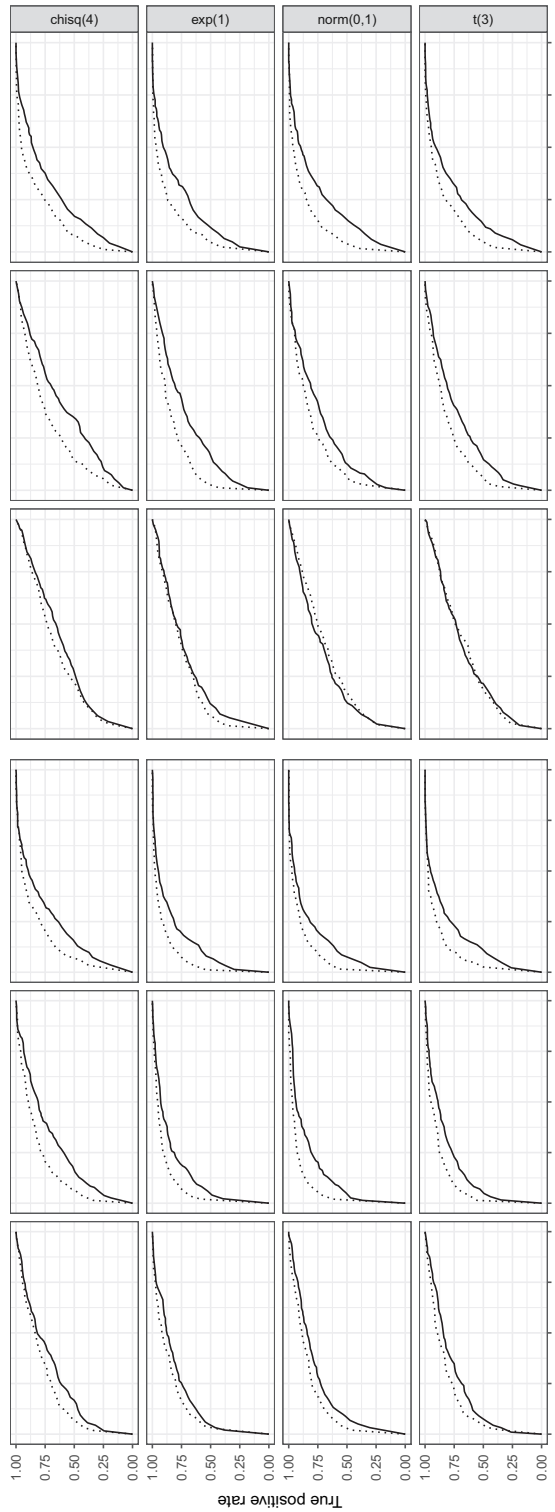

Fig. 4. ROC curves of $\hat{T}_{2n}$ (———) and the oracle test (· · · · ·) under the alternative model 1 with heavy-tailed covariates and (a)–(c) $(n, p) = (200, 400)$ or (d)–(f) $(n, p) = (200, 1000)$: (a), (d) $\tau = 0.25$; (b), (e) $\tau = 0.5$; (c), (f) $\tau = 0.75$

(b) For both tests $\hat{T}_{1n}$ and CSG, the powers drop dramatically as the number of covariates increases from $p=20$ to $p=40$, and actually they are not applicable when $p>n$.
(c) $\hat{T}_{2n}$ has a similar performance to the oracle test when the number of covariates is not too large, say $p=20$, and still has a comparable power, although it becomes less powerful, when p increases.
(d) Both $\hat{T}_{2n}$ and the oracle test have a relatively lower power for the heavy-tailed covariates than those for Gaussian covariates.

We may conclude that the proposed tests $\hat{T}_{1n}$ and $\hat{T}_{2n}$ 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 sales of a company can be affected by other factors. Note that the values of sales may vary in a very large range across different companies, and hence a quantile regression model may be more suitable here.

The data were 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 that 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 79 companies are included. The data set was downloaded from `http://lib.stat.cmu.edu/DASL/Datafiles/Companies.html`.

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

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

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

We applied our test $\hat{T}_{1n}$ to check the lack of fit for this model, and the bootstrapping procedure in Section 3 was employed to approximate the null distribution with $B=5000$ bootstrapped samples. The estimated p-values are 0.87, 4×10^{-4} and 0.0 at three quantile levels $\tau=0.25$, 0.5, 0.75 respectively. This implies that the linear regression model may fit data well for those companies with low sales, whereas contributions to sales from assets, profit and employee sizes may no longer linearly increase for companies with relatively high sales.

To explore further the relationship between the response and covariates at $\tau=0.25, 0.5, 0.75$, we first took a logarithmic transformation of the data, i.e. we let $\tilde{Y}_i=\log(Y_i)$ and $\tilde{X}_{ji}=\log(X_{ji})$ with $1\leqslant j\leqslant 4$, and then we fitted a linear quantile model:

$$Q_\tau(\tilde{Y}_i|\mathbf{X}_i^*)=\beta_0+\sum_{j=1}^{4}\beta_j\tilde{X}_{ji}, \qquad i=1,\ldots,79.$$

The estimated p-values of our test $\hat{T}_n$ are 0.99, 0.81 and 0.43 at quantile levels $\tau=0.25, 0.5$, 0.75 respectively, and this confirms the existence of non-linearity in the original model. We also performed the lack-of-fit test in He and Zhu (2003). However, all p-values are close to 1, and it fails to distinguish the above two models.

6.2. Gross domestic product growth rate data

This subsection attempts to analyse the data set in Barro and Lee (2013). The original data set contains statistics of economic development from 138 different countries, and they were collected quinquennially from 1950 to 2010 or averaged over a 5-year period between 1950 and 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, purchasing power parity deflators, political variables and trade policies. All economic features have been recorded in detail and more extensive information can be found at `http://www.barrolee.com`. A subset of the original data set is given in R package `hdm`, manufactured by Chernozhukov *et al.* (2016), and it consists of $n=90$ complete observations with $p=61$ variables. We shall use this subset data to demonstrate the usefulness of our proposed test $\hat{T}_{2n}$.

In the literature, many researchers have also studied the effect of lagged level of gross domestic product (GDP) *per capita* on the current GDP. For example, in the classical Solow–Swan–Ramsey growth model, there is a hypothesis of convergence in one country's economic development, which states 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. We chose 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 and free trade openness and the other 58 characteristics are set to covariates. We then considered the following quantile regression model:

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

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 were conducted accordingly.

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

As a matter of fact, variable selection has been an important issue in this study since the number of observations is comparable with the number of covariates. Belloni and Chernozhukov (2011) proposed the use of l_1-penalized quantile regression with an adaptive method in choosing

Table 1. p-values of the $\hat{T}_{2n}$- and CSG tests

Quantile level τ	$\hat{T}_{2n}$*-test p-value*	*CSG test p-value*
0.25	0.015	0.442
0.5	0.000	0.594
0.75	0.011	0.796

Table 2. *p*-values of the CSG and oracle tests

Quantile level τ	*Oracle test p-value*	*CSG test p-value*
0.25	0.058	0.502
0.5	0.346	0.570
0.75	0.007	0.786

penalty parameter λ to select the working model. According to the suggested relaxation of λ, several covariates were chosen: lagged GDP growth rate X_1, black market premium X_2, political instability X_3, a measure of tariff restriction X_4, infant mortality rate X_5, ratio of government 'consumption' net of defence and education X_6, exchange rate X_7, 'higher school complete' percentage in the females population X_8, 'secondary school complete' percentage in the males population X_9, females gross enrolment ratio for higher education X_{10}, percentage of non-education in the males population X_{11}, population proportion over 65 years old X_{12} and average years of secondary schooling in the males population X_{13}. We treated these covariates as the truly active covariates, and it then formed a low dimensional model:

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

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

The test proposed rejects the hypothesis of using a linear model to describe the latent relationship between current GDP growth rate and lagged GDP growth rate as well as other economic features at quantile levels $\tau = 0.25, 0.75$ at the level of significance 10%. It is consistent with the intuition that low or high GDP growth rates may be related to much more complicated social or political reasons, whereas a simple linear regression model may not be able to excavate enough information from the true underlying correspondence. Meanwhile, we can attempt to use a median linear regression model to provide some insights into interpreting the effects of a country's economic features on its GDP growth rate.

7. Conclusion and discussion

Our main contribution 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 in covariates such as the number of covariates, sample sizes and types of data (discrete or continuous covariates). As an illustration, by combining several successful two-sample tests in the literature, this paper has constructed two lack-of-fit tests, which are powerful for low dimensional and high dimensional data.

The tests that were proposed in this paper are for a fixed τ and can be easily extended to the case with finite quantile levels. Recently more research in quantile regression has been conducted on $\mathcal{I} \subset (0, 1)$. See Koenker and Machado (1999), Koenker and Xiao (2002), Angrist *et al*. (2006), Escanciano and Goh (2014) and 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 referees for their valuable comments that led to the substantial improvement in the quality of this paper. This work was partially supported by the National Natural Science Foundation of China (grants 11571218 and 11690012), the State Key Program in the Major Research Plan of the National Science Foundation of China (grant 91546202), the Program for Innovative Research Team of Shanghai University of Finance and Economics and the Hong Kong Research Grants Council (grants 17325416 and 17304617).

Appendix A: Technical conditions

A.1. Assumptions for low dimensional data

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

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

Assumption 3. The τth quantile of F_w is 0, $\int_0^\infty x^{-1} \mathrm{d}F_w(x) = -\int_{-\infty}^0 x^{-1} \mathrm{d}F_w(x) = 0.5$, $\int_{-\infty}^\infty |x| \mathrm{d}F_w(x) < \infty$ and there are two positive constants 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 $\hat{\boldsymbol{\beta}}_n$. Assumption 3 is just conditions (Q3)–(Q5) of Feng *et al.* (2011).

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

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

where

$$\zeta_i = \frac{1 - 2z_i}{2\tau(1-\tau)} - \frac{z_i}{2\tau^2} + \frac{1 - z_i}{2(1-\tau)^2}.$$

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

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

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

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

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

Assumption 7. The covariate $\mathbf{X}$ satisfies either of the following conditions.

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

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

(b) Polynomial-type tails: given some constants $\gamma_0, c_1 > 0$, $p \leqslant c_1 n^{\gamma_0}$ and for some constants $\varepsilon > 0$ and $K > 0$ such that

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

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

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

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

Assumption 8. There exist $\kappa \geqslant \frac{1}{3}$ 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 $\tilde{\boldsymbol{\beta}}_n$ and its debiased version $\hat{\boldsymbol{\beta}}_n$ (Bradic and Kolar, 2017). Assumption 5 consists of common assumptions in the high dimensional setting (Cai *et al.*, 2013). Assumption 6 further restricts the correlation matrix, whereas assumption 8 is used to control the fourth moment (Cai *et al.*, 2014). Assumption 7 specifies sub-Gaussian and polynomial-type distributions, and those families include many commonly used distributions, such as normal and Student t-distributions (Cai *et al.*, 2013, 2014).

References

Anderson, T. W. (1962) On the distribution of the two-sample Cramér-von Mises criterion. *Ann. Math. Statist.*, **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. *Statist. Sin.*, **6**, 311–329.

Baringhaus, L. and Franz, C. (2004) On a new multivariate two-sample test. *J. Multiv. Anal.*, **88**, 190–206.

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

Belloni, A. and Chernozhukov, V. (2011) l_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.*, **39**, 82–130.

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

Cai, T. T., Liu, W. and Xia, Y. (2013) Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Am. Statist. Ass.*, **108**, 265–277.

Cai, T. T., Liu, W. and Xia, Y. (2014) Two-sample test of high dimensional means under dependence. *J. R. Statist. Soc.* B, **76**, 349–372.

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

Chernozhukov, V., Hansen, C. and Spindler, M. (2016) hdm: high-dimensional metrics. *R J.*, **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. *Computnl Statist. Data Anal.*, **88**, 128–138.

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

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

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

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

He, X., Wang, L. and Hong, H. G. (2013) Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.*, **41**, 342–369.

He, X. and Zhu, L. (2003) A lack-of-fit test for quantile regression. *J. Am. Statist. Ass.*, **98**, 1013–1022.

Horowitz, J. and Spokoiny, V. G. (2002) An adaptive, rate-optimal test of linearity for median regression models. *J. Am. Statist. Ass.*, **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. *J. Am. Statist. Ass.*, **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-dimensional covariance matrices. *Ann. Statist.*, **40**, 908–940.
Ma, S., Li, R. and Tsai, C.-L. (2017) Variable screening via quantile partial correlation. *J. Am. Statist. Ass.*, **112**, 650–663.
Otsu, T. (2008) Conditional empirical likelihood estimation and inference for quantile regression models. *J. Econmetr.*, **142**, 508–538.
Portnoy, S. and Koenker, R. (1989) Adaptive estimation of linear models. *Ann. Statist.*, **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. *Computnl Statist. Data Anal.*, **51**, 6535–6542.
Shah, R. D. and Bühlmann, P. (2018) Goodness-of-fit tests for high dimensional linear models. *J. R. Statist. Soc.* B, **80**, 113–135.
Srivastava, M. and Yanagihara, H. (2010) Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Multiv. Anal.*, **101**, 1319–1329.
Stute, W. (1997) Nonparametric model checks for regression. *Ann. Statist.,* **25**, 613–641.
Székely, G. J. and Rizzo, M. L. (2005) A new test for multivariate normality. *J. Multiv. Anal.*, **93**, 58–80.
Whang, Y.-J. (2006) Smoothed empirical likelihood methods for quantile regression models. *Econmetr. Theory*, **22**, 173–205.
Zheng, J. X. (1998) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econmetr. Theory*, **14**, 123–138.
Zheng, Q., Peng, L. and He, X. (2015) Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Statist.*, **43**, 2225–2258.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for "Lack-of-fit tests for quantile regression models"'.