ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)

Testing for marginal linear effects in quantile regression

Huixia Judy Wang

George Washington University, Washington DC, USA

and Ian W. McKeague and Min Qian

Columbia University, New York, USA

[Received April 2016. Final revision September 2017]

Summary. The paper develops a new marginal testing procedure to detect significant predictors that are associated with the conditional quantiles of a scalar response. The idea is to fit the marginal quantile regression on each predictor one at a time, and then to base the test on the t-statistics that are associated with the most predictive predictors. A resampling method is devised to calibrate this test statistic, which has non-regular limiting behaviour due to the selection of the most predictive variables. Asymptotic validity of the procedure is established in a general quantile regression setting in which the marginal quantile regression models can be misspecified. Even though a fixed dimension is assumed to derive the asymptotic results, the test proposed is applicable and computationally feasible for large dimensional predictors. The method is more flexible than existing marginal screening test methods based on mean regression and has the added advantage of being robust against outliers in the response. The approach is illustrated by using an application to a human immunodeficiency virus drug resistance data set.

Keywords: Bootstrap calibration; Inference; Marginal regression; Non-standard asymptotics; Quantile regression

1. Introduction

Consider a scalar response Y and a p-dimensional predictor $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$. We are interested in testing the presence of significant predictors that affect the conditional quantile of Y at a given quantile level τ or across multiple quantiles. To answer this question, we develop a new inference procedure based on marginal linear quantile regression. For theoretical development, throughout we assume that p is finite, though the test is applicable to cases with a large number of predictors.

Quantile regression has attracted increasing attention in recent years, mainly due to the following attractive features:

(a) robustness against outliers in the response, especially in the case of median regression;
(b) the ability to capture heterogeneity in the set of important predictors at different quantile levels of the response distribution caused by, for instance, heteroscedastic variance.

To test for significant predictors at the τth conditional quantile of Y, one natural approach is to carry out a hypothesis test comparing the null model of no predictors and the full model

Address for correspondence: Huixia Judy Wang, Department of Statistics, 739 Rome Hall, George Washington University, Washington DC 20052, USA.
E-mail: judywang@email.gwu.edu

 1369–7412/17/80000

consisting of all p predictors, by using either Wald-type, quasi-likelihood-ratio tests or bootstrap methods; see related discussions in Koenker (2005), chapter 3, Kocherginsky *et al.* (2005) and Feng *et al.* (2011). Unfortunately, these omnibus-type tests are based on fitting a model with all p predictors, which quickly becomes prohibitive and leads to less powerful tests for large p.

Recently McKeague and Qian (2015) proposed an adaptive resampling test, which provides valid post-selection inference for detecting significant predictors based on marginal linear regression. As in correlation learning or sure independence screening, the idea of marginal regression is to regress the response on each predictor one at a time (Fan and Lv, 2008; Genovese *et al.*, 2012). The adaptive resampling test successfully controls the familywise error rate by accounting for the variation that is caused by variable selection.

Partly inspired by this work, here we propose a new method for testing the existence of marginal effects in quantile regression. The new approach allows the most active predictors to vary at different quantile levels of the response distribution, so it is more flexible than mean regression and has the added advantage of being robust against outliers in the response. The test proposed is based on a maximum-type t-statistic associated with the selected most informative predictor at the quantile level of interest. Valid statistical inference must take into account the uncertainty that is involved in the selection step; see related discussion in Leeb and Pötscher (2006) and Belloni *et al.* (2014). We show that the limiting distribution of the proposed post-selected statistic changes abruptly in the proximity of the null hypothesis of no effect. To adapt to the non-regular asymptotic behaviour of the test statistic that is caused by variable selection, we develop a modified bootstrap procedure using a 'pre-test' through thresholding. To the best of our knowledge, the test is the first inference tool developed for quantile regression that scales in a computationally practical way with dimension. Unlike the omnibus-type tests that require fitting a model with all p predictors, the method proposed is based on fitting p linear quantile regression models with a single predictor, so its computational cost grows only linearly with p. Even though we assume a fixed dimension to derive the asymptotic results, there is numerical evidence that the test continues to work in cases with $p > n$, but a thorough investigation for the high dimensional case remains open.

This paper makes a novel contribution that is distinct from McKeague and Qian (2015). First of all, the test in McKeague and Qian (2015) is based on selecting the predictor that is maximally correlated with the response Y. However, correlation is not useful in the quantile regression set-up. Instead, we propose a new selection rule, which selects the most informative predictor at the quantile of interest as the predictor that minimizes an empirical asymmetric L_1-loss function. Secondly, our proposed test uses a scale invariant t-statistic, whereas that in McKeague and Qian (2015) is based on the maximum slope estimator. In quantile regression, the scale of the slope estimator depends not only on covariates but also on the quantile level. Therefore, a scale invariant statistic is more desirable for multiple-quantile analysis, since prestandardization of covariates is not sufficient to make the type of statistic as in McKeague and Qian (2015) scale free. Thirdly, unlike least squares regression, quantile regression allows the analysis over a set of quantile levels to capture the population heterogeneity. The flexibility of such globally concerned quantile regression has been discussed in Zheng *et al.* (2015). With the established convergence results across quantiles, our developed method can be used to detect the significance of predictors not only at a single quantile level but also at multiple quantiles jointly.

For quantile regression, developing asymptotic theory to calibrate the test statistic is enormously more challenging because of three main obstacles. Firstly, unlike in mean regression, the quantile estimator $\hat{\theta}_n(\tau)$ of the slope parameter for the predictor selected has no explicit form. Secondly, the asymmetric L_1-loss function is not differentiable everywhere. Lastly, since each marginal quantile regression model is (possibly) misspecified, there is a prediction bias which is

caused by omitting the correct predictors, and the bias takes a complicated form in the quantile regression setting.

To overcome these challenges and to study the non-regular limiting behaviour of $\hat{\theta}_n(\tau)$, we consider a general quantile regression model indexed by a (unidentifiable) local parameter that represents uncertainty at the $\sqrt{n}$-scale close to the null hypothesis. We establish the asymptotic properties of the t-statistic $T_n(\tau) = n^{1/2}\hat{\theta}_n(\tau)/\hat{\sigma}_n(\tau)$ and obtain a quadratic expansion for the non-differentiable loss function by using empirical process tools. In addition, we assess the bias of the marginal quantile regression estimator due to model misspecification under the local model by adapting the results in Angrist *et al.* (2006). On the basis of the asymptotic theory developed, we devise a non-parametric bootstrap procedure that adapts to the non-regular asymptotic behaviour of $T_n(\tau)$ by using a pre-test that involves thresholding. We establish the bootstrap consistency of this procedure under the general local quantile regression model.

The current paper is closely related to the post-selection inference literature. In particular, one can view the determination of the most informative predictor as a selection step, and the inference on the slope that is associated with the selected predictor in marginal quantile regression as the inference after selection. Lee and Taylor (2014) developed a post-selection inference method for marginal screening in linear regression, which selects the top k predictors that are most correlated with Y. Our proposed test solves a similar problem for the special case of $k = 1$ in a quantile regression set-up. In contrast, the method in Lee and Taylor (2014) relies on a strong model assumption that the regression errors are normally distributed with constant variance, and thus it cannot be applied to quantile regression. Among very few related works in the quantile regression literature, Belloni *et al.* (2014, 2015) proposed post-selection inference methods for quantile and median regression models respectively. However, inference in Belloni *et al.* (2014, 2015) is focused on the slope parameter of a single prespecified predictor (such as a treatment indicator), whereas the remaining predictors are selected through L_1-penalization. Thus their approach does not apply to a marginal screening-type test in which no predictor is singled out *a priori*.

Recently, Zhang *et al.* (2017) proposed a test based on the martingale difference divergence to detect the dependence of the conditional quantile of the response on covariates in a model-free setting. The current paper differs from Zhang *et al.* (2017) from several perspectives. When the interest is on the dependence at a single quantile level τ, our proposed method does not require any model assumptions on the quantiles near τ, whereas the test calibration in Zhang *et al.* (2017) requires a stronger local quantile independence assumption. In addition, our proposed method can be used to conduct joint tests across multiple quantiles, and to identify the most predictive variables, whereas that in Zhang *et al.* (2017) can only be applied to assess the overall dependence at a given single quantile level.

The rest of the paper is organized as follows. In Section 2, we formulate the problem, establish the non-regular local asymptotic distribution of $T_n(\tau)$ and develop the proposed test procedure. We assess the performance of the approach through a simulation study in Section 3 and apply it to a human immunodeficiency virus (HIV) drug susceptibility data set in Section 4. Some concluding remarks are made in Section 5. Proofs are collected in Appendix A and the on-line supplementary material. The R program developed and the HIV data are available from `http://www.columbia.edu/~im2131/ps/index.html`.

2. Method proposed

2.1. Marginal quantile regression

Suppose that $\{(y_i, \mathbf{x}_i), i = 1, \ldots, n\}$ is a random sample of $(Y, \mathbf{X})$, where $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^{\mathrm{T}}$. Let

$\mathcal{T} \subset (0,1)$ be a set with L prespecified quantile levels, where L is finite. We assume the following linear quantile regression model:

$$Q_\tau(Y|\mathbf{X}) = \alpha_0(\tau) + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau), \qquad \tau \in \mathcal{T}, \tag{1}$$

where $\alpha_0(\tau)$ and $\boldsymbol{\beta}_0(\tau) \in \mathbb{R}^p$ are the unknown quantile coefficients, $Q_\tau(Y|\mathbf{X}) = \inf\{y : F_Y(y|\mathbf{X}) \geqslant \tau\}$ is the τth conditional quantile of Y given $\mathbf{X}$, and $F_Y(\cdot|\mathbf{X})$ is the distribution function of Y given $\mathbf{X}$. Model (1) assumes only that the conditional quantile of Y is linear in $\mathbf{X}$ at quantiles of interest and thus is broader than the linear regression model in McKeague and Qian (2015). Unlike least squares regression, quantile regression analysis enables us to study at multiple quantiles. We aim to develop a formal test of whether any component of $\mathbf{X}$ has an effect on either a given quantile or at multiple quantiles of Y. Throughout we assume that all predictors are standardized before the data analysis. For notational simplicity we shall first present the proposed test at a single quantile level $\tau \in \mathcal{T}$ and discuss testing across multiple quantiles in Section 2.5.

Our proposed test is based on fitting the working marginal quantile regression models by regressing Y on X_k, $k = 1, \ldots, p$, for each k separately, i.e., for each k, the working marginal quantile regression solves the population minimization problem to obtain

$$(\alpha_k(\tau), \theta_k(\tau)) = \arg\min_{\alpha,\,\theta} E[\rho_\tau(Y - \alpha - \theta X_k) - \rho_\tau(Y)], \tag{2}$$

where $\rho_\tau(u) = \{\tau - I(u \leqslant 0)\}u$ is the quantile loss function (Koenker, 2005), and here we (implicitly) assume integrability. The coefficient $\theta_k(\tau)$ approximates the linear effect of X_k on the τth quantile of Y, and we refer to it as the *quantile marginal linear effect* of X_k hereafter. However, it is worth noting that the marginal quantile regression models are in general misspecified, and consequently $\alpha_k(\tau) + \theta_k(\tau)X_k$ may differ from $Q_\tau(Y|X_k)$, the conditional quantile of Y given a single predictor X_k, and $\theta_k(\tau)$ may differ from the derivative of $Q_\tau(Y|\mathbf{X})$ with respect to X_k. Since the seminal work of Fan and Lv (2008), marginal regression has been used for feature screening in various models. For conditional quantile screening, He *et al.* (2013) proposed a screening method based on marginal non-parametric quantile regression, and Shao and Zhang (2014) developed a model-free approach based on martingale difference correlation. Differently from these works, our focus is not on screening but on testing the existence of overall covariate effects through marginal regression, and the test proposed can be used as a first step before variable selection.

We define the index of the most informative predictor at the τth quantile as

$$k_0(\tau) = \arg\min_{k=1,\ldots,p} E[\rho_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\} - \rho_\tau(Y)] \tag{3}$$

and denote the corresponding slope parameter by $\theta_0(\tau) = \theta_{k_0(\tau)}(\tau)$. We focus on testing $H_0 : \theta_0(\tau) = 0$ *versus* $H_{\mathrm{a}} : \theta_0(\tau) \neq 0$. Note that the rejection of H_0 implies that at least one of the p predictors has an effect on the τth conditional quantile of Y.

On the basis of the random sample, we define

$$\begin{aligned} \hat{k}_n(\tau) &= \arg\min_{k=1,\ldots,p} \mathbb{P}_n[\rho_\tau\{Y - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau)X_k\}] \\ &\doteq \arg\min_{k=1,\ldots,p} n^{-1}\sum_{i=1}^{n} \rho_\tau\{y_i - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau)x_{i,k}\}, \end{aligned} \tag{4}$$

where $(\hat{\alpha}_k(\tau), \hat{\theta}_k(\tau)) = \arg\min_{\alpha,\theta}\mathbb{P}_n\{\rho_\tau(Y - \alpha - \theta X_k)\}$ are the sample quantile coefficient estimators that are obtained by regressing Y on X_k. The marginal quantile regression estimator of $\theta_0(\tau)$ is then given by $\hat{\theta}_n(\tau) = \hat{\theta}_{\hat{k}_n(\tau)}$.

For any given $k = 1, \ldots, p$, by theorem 3 of Angrist *et al.* (2006), the asymptotic covariance of $(\hat{\alpha}_k(\tau), \hat{\theta}_k(\tau))$ is $\Sigma_k(\tau) \doteq \mathbf{J}_k(\tau)^{-1} E[[\psi_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}]^2 \tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^{\mathrm{T}}] \mathbf{J}_k(\tau)^{-1}$, where $\mathbf{J}_k(\tau) = E[f_Y\{\alpha_k(\tau) + \theta_k(\tau)X_k | X_k\}\tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^{\mathrm{T}}]$, $\psi_\tau(u) = \tau - I(u < 0)$ and $\tilde{\mathbf{X}}_k = (1, X_k)^{\mathrm{T}}$. We can estimate $\mathbf{J}_k(\tau)$ by the kernel-based estimator that was suggested by Powell (1991),

$$\hat{\mathbf{J}}_k(\tau) = (nh_n)^{-1} \sum_{i=1}^{n} \mathcal{K}[\{y_i - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau)x_{i,k}\}/h_n]\tilde{\mathbf{x}}_{i,k}\tilde{\mathbf{x}}_{i,k}^{\mathrm{T}}, \tag{5}$$

where $\mathcal{K}(\cdot)$ is a kernel function, and h_n is the positive bandwidth satisfying $h_n \to 0$ and $h_n\sqrt{n} \to \infty$. By lemmas 1–3 in Appendix A and theorem 3 of Powell (1991), it is easy to show that $\hat{\mathbf{J}}_k(\tau) - \mathbf{J}_k(\tau) = o_p(1)$ uniformly in τ. Alternatively we can also estimate $\mathbf{J}_k(\tau)$ by using the difference quotient method in Hendricks and Koenker (1991). Consequently, we can estimate $\Sigma_k(\tau)$ consistently by $\hat{\Sigma}_k(\tau) = \hat{\mathbf{J}}_k^{-1}(\tau)\mathbb{P}_n([\tau - I\{Y < \hat{\alpha}_k(\tau) + \hat{\theta}_k(\tau)X_k\}])^2 \tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^{\mathrm{T}} \hat{\mathbf{J}}_k^{-1}(\tau)$. For any $k = 1, \ldots, p$, denote the lower right diagonal element of $\hat{\Sigma}_k(\tau)$ as $\hat{\sigma}_k^2(\tau)$.

Our proposed test statistic is defined as

$$T_n(\tau) = \frac{n^{1/2}\,\hat{\theta}_n(\tau)}{\hat{\sigma}_n(\tau)}, \tag{6}$$

where $\hat{\sigma}_n(\tau) = \hat{\sigma}_{\hat{k}_n(\tau)}(\tau)$. The test statistic $T_n(\tau)$ is a maximum-type statistic, which is a natural choice for testing H_0. However, when none of the components of $\mathbf{X}$ have an effect on the τth conditional quantile of Y, $k_0(\tau)$ is unidentifiable—it can be any of the p indices. In addition, the distribution of $n^{1/2}\{\hat{\theta}_n(\tau) - \theta_0(\tau)\}/\hat{\sigma}_n(\tau)$ does not converge uniformly with respect to $\theta_0(\tau)$ in the neighbourhood of $\theta_0(\tau) = 0$, so the normal limiting distribution that holds away from $\theta_0(\tau) = 0$ cannot be used to construct rejection regions. To construct a suitable test procedure for H_0, it is important to study the asymptotic behaviour of $T_n(\tau)$ under local alternatives.

2.2. Local model

In the local model, we replace the slope parameter $\boldsymbol{\beta}_0(\tau)$ in model (1) by

$$\boldsymbol{\beta}_n(\tau) = \boldsymbol{\beta}_0(\tau) + n^{-1/2}\,\mathbf{b}_0(\tau), \tag{7}$$

where $\mathbf{b}_0(\tau) \in \mathbb{R}^p$ is the local parameter. When $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, the quantile effect of $\mathbf{X}$ is $n^{-1/2}\mathbf{X}^{\mathrm{T}}\mathbf{b}_0(\tau)$ and it vanishes asymptotically. Let $\epsilon(\tau) = Y - \alpha_0 - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_n(\tau)$ denote the quantile regression residuals. It is clear that $Q_\tau\{\epsilon(\tau)|\mathbf{X}\} = 0$. Throughout, we assume that the distributions of $\mathbf{X}$ and $\epsilon(\tau)$ are fixed and only the distribution of Y depends on n, but we suppress n in the notation of Y for notational simplicity.

Under the local model, we define

$$\bar{k}_n(\tau, \mathbf{b}_0) = \arg\min_{k=1,\ldots,p} \min_{\alpha,\theta} E[\rho_\tau(Y - \alpha - X_k\theta) - \rho_\tau(Y)] \tag{8}$$

and rewrite $\theta_0(\tau)$ as $\theta_n(\tau) \doteq \theta_{\bar{k}_n(\tau,\mathbf{b}_0)}(\tau)$. When $\mathbf{b}_0(\tau) = \mathbf{0}$, $\bar{k}_n(\tau, \mathbf{0})$ coincides with $k_0(\tau)$ defined in equation (3) under the global model. If $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$ and $k_0(\tau)$ is unique, then $\bar{k}_n(\tau, \mathbf{b}_0) \to k_0(\tau)$ and $\theta_n(\tau)$ is asymptotically bounded away from zero, representing a non-local alternative case. In contrast if $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, then $\theta_n(\tau)$ is in the neighbourhood of zero representing a local alternative case. However, if $\boldsymbol{\beta}_0(\tau) = \mathbf{b}_0(\tau) = \mathbf{0}$, then $\bar{k}_n(\tau, \mathbf{b}_0)$ is not well defined and $\theta_n(\tau) = 0$, representing the null case i.e., under the local model (7), even though $\bar{k}_n(\tau, \mathbf{b}_0)$ is unidentifiable at $\boldsymbol{\beta}_0(\tau) = \mathbf{b}_0(\tau) = \mathbf{0}$, it is still 'weakly identifiable' when $\mathbf{b}_0(\tau) \neq \mathbf{0}$. By analysing the local model, we can study the (non-regular) asymptotic behaviour of the test statistic $T_n(\tau)$.

2.3. Limiting behaviour under the local model

Under the local model, in which the response Y depends on n, $(\alpha_k(\tau), \theta_k(\tau))$ defined in equation (2) also depends on n, but throughout the paper we suppress this dependence to simplify notation. We make the following assumptions.

Assumption 1. The conditional quantile regression model (1) with slope parameters defined in equation (7) holds, where $\alpha_0(\tau)$, $\boldsymbol{\beta}_0(\tau)$ and $\mathbf{b}_0(\tau)$ are in the interior of compact sets.

Assumption 2. Under the local model, for each $n \geqslant 1$, $k = 1, \ldots, p$ and any $\tau \in \mathcal{T}$, there are unique $(\alpha_k(\tau), \theta_k(\tau)) = \arg\min_{(\alpha,\theta)} E[\rho_\tau(Y - \alpha - \theta X_k) - \rho_\tau(Y)]$.

Assumption 3. The covariate $\mathbf{X}$ has a compact support.

Assumption 4. The conditional density $f_Y(\cdot|\mathbf{X})$ is bounded and uniformly continuous with a bounded first derivative, over the support of $\mathbf{X}$.

Assumption 5. Under the local model, for $k = 1, \ldots, p$, the matrices

$$\mathbf{J}_k\{\tau, \boldsymbol{\beta}_0(\tau)\} = \lim_{n\to\infty} E[f_{\epsilon(\tau)}\{e_k(\tau)|\mathbf{X}\}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^{\mathrm{T}}] \doteq \begin{pmatrix} \pi_k\{\tau, \boldsymbol{\beta}_0(\tau)\} & \mu_k\{\tau, \boldsymbol{\beta}_0(\tau)\} \\ \mu_k\{\tau, \boldsymbol{\beta}_0(\tau)\} & \mu_{kk}\{\tau, \boldsymbol{\beta}_0(\tau)\} \end{pmatrix}$$

and $\lim_{n\to\infty} E[[\psi_\tau\{\epsilon(\tau) - e_k(\tau)\}]^2\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^{\mathrm{T}}]$ exist and are positive definite for all $\tau \in \mathcal{T}$, where $e_k(\tau) = \alpha_k(\tau) + X_k\theta_k(\tau) - \alpha_0(\tau) - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_n(\tau)$ is the prediction bias due to marginal regression of Y on X_k at the quantile level τ.

Assumption 2 is an identifiability condition that is needed to ensure that the population quantile coefficient vector $(\alpha_k(\tau), \theta_k(\tau))$ that is obtained by regressing Y on X_k is unique. Assumptions 3–5 concern $\mathbf{X}$ and $\epsilon(\tau)$, and are needed to apply a result of Angrist *et al.* (2006), theorem 3, to obtain the asymptotic properties of the marginal regression estimator $(\hat{\alpha}_k(\tau), \hat{\theta}_k(\tau))$ under misspecification caused by omitting the correct predictors. Assumption 3 is needed to obtain the approximate representation of $(\alpha_k(\tau), \theta_k(\tau))$ when $\boldsymbol{\beta}_0(\tau) = 0$. This condition can be relaxed by assuming some boundedness condition on the higher moments of $\mathbf{X}$, but this would greatly complicate the proofs. In theory, the boundedness in the support of $\mathbf{X}$ is naturally needed to avoid quantile crossing if multiple linear quantile regression functions are assumed. When $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, $(\alpha_k(\tau), \theta_k(\tau)) \to (\alpha_0(\tau), 0)$ by the proof of lemma 2, so the first limiting matrix in assumption 5 is $\mathbf{J}_k(\tau, \mathbf{0}) = E[f_{\epsilon(\tau)}(0|\mathbf{X})\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^{\mathrm{T}}]$ with $\pi_k(\tau, \mathbf{0}) \equiv \pi(\tau, \mathbf{0}) = E[f_{\epsilon(\tau)}(0|\mathbf{X})]$, $\mu_k(\tau, \mathbf{0}) = E[f_{\epsilon(\tau)}(0|\mathbf{X})X_k]$ and $\mu_{kk}(\tau, \mathbf{0}) = E[f_{\epsilon(\tau)}(0|\mathbf{X})X_k^2]$, and the second limit is $\tau(1-\tau)E[\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^{\mathrm{T}}]$.

Theorem 1 gives the asymptotic representation of $n^{1/2}\{\hat{\theta}_n(\tau) - \theta_n(\tau)\}/\hat{\sigma}_n(\tau)$ when $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$ and $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$ separately. The asymptotic representation goes through a phase transition at $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, with a different form for $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$ (in which case $\bar{k}_n(\tau, \mathbf{b}_0)$ is identifiable). We first fix some notation. Let $\mathbb{X} = (\mathbf{1}_p^{\mathrm{T}}, X_1, \ldots, X_p)^{\mathrm{T}}$ and $\mathbb{X}_j$ be its jth element for $j = 1, \ldots, 2p$. In addition, for $k = 1, \ldots, p$, let $V_k(\tau, \boldsymbol{\beta}_0) = |\mathbf{J}_k(\tau, \boldsymbol{\beta}_0)|$, $\mathbf{C}_k(\tau) = E[f_{\epsilon(\tau)}(0|\mathbf{X})]E[f_{\epsilon(\tau)}(0|\mathbf{X})X_k\mathbf{X}] - E[f_{\epsilon(\tau)}(0|\mathbf{X})X_k]E[f_{\epsilon(\tau)}(0|\mathbf{X})\mathbf{X}]$ and $\mathbf{B}_k(\tau) = E[f_{\epsilon(\tau)}(0|\mathbf{X})\mathbf{X}\tilde{\mathbf{X}}_k^{\mathrm{T}}]$. Let $\mathbf{M}\{\tau, \boldsymbol{\beta}_0(\tau)\} \doteq (M_1\{\tau, \boldsymbol{\beta}_0(\tau)\}, \ldots, M_{2p}\{\tau, \boldsymbol{\beta}_0(\tau)\})^{\mathrm{T}}$ be a Gaussian process with mean 0 and covariance function $\Omega(\tau, \tau')$, whose (j, j')th element is $\Omega_{j,j'}(\tau, \tau') = \lim_{n\to\infty} E[\mathbb{X}_j\mathbb{X}_{j'}[\tau - I\{\epsilon(\tau) < e_j(\tau)\}][\tau' - I\{\epsilon(\tau') < e_{j'}(\tau')\}]]$, where $e_j(\tau) = e_{j-p}(\tau)$ for $j = p+1, \ldots, 2p$.

Theorem 1. Suppose that assumptions 1–5 hold, and, for each $\tau \in \mathcal{T}$, $\bar{k}_n(\tau, \mathbf{0}) \equiv k_0(\tau)$ is unique when $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$, and $\bar{k}_n\{\tau, \mathbf{b}_0(\tau)\} \to \kappa_\tau\{\mathbf{b}_0(\tau)\} \in \{1, \ldots, p\}$ when $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$ and $\mathbf{b}_0(\tau) \neq 0$. Then, jointly over $\tau \in \mathcal{T}$, we have

$$\frac{n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}}{\hat{\sigma}_n(\tau)}$$
$$\xrightarrow{\mathrm{d}} \begin{cases} \dfrac{M_{p+k_0(\tau)}\{\tau,\boldsymbol{\beta}_0(\tau)\}\pi_{k_0(\tau)}\{\tau,\boldsymbol{\beta}_0(\tau)\}-M_{k_0(\tau)}\{\tau,\boldsymbol{\beta}_0(\tau)\}\mu_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}}{V_{k_0(\tau)}\{\tau,\boldsymbol{\beta}_0(\tau)\}\sigma_{k_0(\tau)}(\tau)} & \text{if } \boldsymbol{\beta}_0(\tau)\neq\mathbf{0}, \\ \dfrac{M_{p+K(\tau)}(\tau,\mathbf{0})\pi(\tau,\mathbf{0})-M_{K(\tau)}(\tau,\mathbf{0})\mu_{K(\tau)}(\tau,\mathbf{0})}{V_{K(\tau)}(\tau,\mathbf{0})\sigma_{K(\tau)}(\tau)} & \\ \quad +\left(\dfrac{\mathbf{C}_{K(\tau)}(\tau)}{V_{K(\tau)}(\tau,\mathbf{0})}-\dfrac{\mathbf{C}_{\kappa_\tau\{\mathbf{b}_0(\tau)\}}(\tau)}{V_{\kappa_\tau\{\mathbf{b}_0(\tau)\}}(\tau,\mathbf{0})}\right)^{\mathrm{T}}\dfrac{\mathbf{b}_0(\tau)}{\sigma_{K(\tau)}(\tau)} & \text{if } \boldsymbol{\beta}_0(\tau)=\mathbf{0}, \end{cases}$$

where

$$K(\tau)=\underset{k=1,\ldots,p}{\arg\max}(\mathbf{M}_k(\tau)+\mathbf{B}_k(\tau)^{\mathrm{T}}\mathbf{b}_0(\tau))^{\mathrm{T}}\mathbf{J}_k(\tau,\mathbf{0})^{-1}(\mathbf{M}_k(\tau)+\mathbf{B}_k(\tau)^{\mathrm{T}}\mathbf{b}_0(\tau)) \tag{9}$$

with $\mathbf{M}_k(\tau)=(M_k(\tau,\mathbf{0}),M_{k+p}(\tau,\mathbf{0}))^{\mathrm{T}}$.

In theorem 1, the convergence condition on $\bar{k}_n\{\tau,\mathbf{b}_0(\tau)\}$ is a weak continuity condition on the joint distribution of $\epsilon(\tau)$ and $\mathbf{X}$. Because of the prediction bias $e_k(\tau)$ from marginal regression, $\Omega_{j,j}(\tau,\tau')$ does not have an explicit expression. However, when $\boldsymbol{\beta}_0(\tau)=\boldsymbol{\beta}_0(\tau')=\mathbf{0}$, $e_k(\tau)$ and $e_k(\tau')$ go to 0 under the local model for all $k=1,\ldots,p$, and consequently $\Omega_{j,j'}(\tau,\tau')=\{\min(\tau,\tau')-\tau\tau'\}E[\mathbb{X}_j\mathbb{X}_{j'}]$.

We refer to the special case in which $f_{\epsilon(\tau)}(\cdot|\mathbf{X})=f_{\epsilon(\tau)}(\cdot)$ is the same across all values of $\mathbf{X}$ as *homoscedastic*, i.e. the distribution of the regression error $\epsilon(\tau)$ does not depend on the covariates. In this case, $V_k(\tau,\mathbf{0})=f^2_{\epsilon(\tau)}(0)\,\mathrm{var}(X_k)$, $\mathbf{C}_k(\tau)=f^2_{\epsilon(\tau)}(0)\,\mathrm{cov}(X_k,\mathbf{X})$ and the limiting distribution in theorem 1 when $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$ can be simplified as in the following corollary.

Corollary 1. Under the assumptions of theorem 1, for the homoscedastic case, we have that, jointly over all $\tau\in\mathcal{T}$ such that $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$,

$$\frac{n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}}{\hat{\sigma}_n(\tau)}\xrightarrow{\mathrm{d}}\frac{M_{p+K(\tau)}(\tau)-M_{K(\tau)}(\tau)E[X_K]}{f_{\epsilon(\tau)}(0)\,\mathrm{var}(X_{K(\tau)})\sigma_{K(\tau)}(\tau)}$$
$$+\left(\frac{\mathrm{cov}(X_K,\mathbf{X})}{\mathrm{var}(X_K)}-\frac{\mathrm{cov}(X_{\kappa_\tau\{\mathbf{b}_0(\tau)\}},\mathbf{X})}{\mathrm{var}(X_{\kappa_\tau\{\mathbf{b}_0(\tau)\}})}\right)^{\mathrm{T}}\frac{\mathbf{b}_0(\tau)}{\sigma_{K(\tau)}(\tau)},$$

where $K(\tau)=\arg\max_{k=1,\ldots,p}\{M_{p+k}(\tau)-M_k(\tau)\,E[X_k]+f_{\epsilon(\tau)}(0)\mathbf{b}_0^{\mathrm{T}}(\tau)\,\mathrm{cov}(X_k,\mathbf{X})\}^2/\mathrm{var}(X_k)$ with $M_k(\tau)=M_k(\tau,\mathbf{0})$.

2.4. Adaptive bootstrap for testing at a single quantile

By theorem 1, we can obtain the asymptotic critical values for the test statistic $T_n(\tau)$ by simulating its asymptotic representation under the null hypothesis (with $\boldsymbol{\beta}_0(\tau)=\mathbf{b}_0(\tau)=\mathbf{0}$). This approach, however, requires estimating the weighted covariance matrix of $\mathbf{X}$ with weights to accommodate heteroscedasticity and thus does not perform well in finite samples with large p. For practical purposes, we propose to adopt the idea in McKeague and Qian (2015) and to develop an adaptive bootstrap procedure.

Denote $\{(y_i^*,\mathbf{x}_i^*),i=1,\ldots,n\}$ as the bootstrap sample that is obtained by sampling the observed data with replacement. The conventional bootstrap version of $R_n(\tau)\doteq n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}/\hat{\sigma}_n(\tau)$ is $R_n^*(\tau)\doteq n^{1/2}\{\hat{\theta}_n^*(\tau)-\hat{\theta}_n(\tau)\}/\hat{\sigma}_{n*}(\tau)$, where $\hat{\theta}_n^*(\tau)$ is the bootstrap version of $\hat{\theta}_n(\tau)$, and $\hat{\sigma}_{n*}(\tau)=\hat{\sigma}_{\hat{k}_n^*(\tau)}(\tau)$ with $\hat{k}_n^*(\tau)$ as the bootstrap counterpart of $\hat{k}_n(\tau)$.

However, since $\hat{\theta}_n(\tau)$ is a non-regular estimator, i.e. its distribution does not converge uniformly in an $n^{-1/2}$-neighbourhood of the null, the conventional bootstrap would fail to estimate the null distribution. To account for the discontinuity in the limiting distribution of $R_n(\tau)$, we propose to compare $|T_n(\tau)|$ with some threshold $\lambda_n(\tau) > 0$ to capture the different behaviours of $R_n(\tau)$ in two scenarios: $\beta_0(\tau) \neq \mathbf{0}$ (away from the null) and $\beta_0(\tau) = \mathbf{0}$ (in the $n^{-1/2}$-neighbourhood of the null). We consider the following modified bootstrap version of $R_n(\tau)$:

$$R_n^*\{\tau, \mathbf{b}_0(\tau)\} = \begin{cases} \dfrac{n^{1/2}\{\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau)\}}{\hat{\sigma}_{n*}(\tau)} & \text{if } |T_n(\tau)| > \lambda_n(\tau) \text{ or } |T_n^*(\tau)| > \lambda_n(\tau), \\ \mathbb{V}_n^*\{\tau, \mathbf{b}_0(\tau)\} & \text{otherwise,} \end{cases} \tag{10}$$

where $T_n^*(\tau) = n^{1/2}\hat{\theta}_n^*(\tau)/\hat{\sigma}_{n*}(\tau)$ and $\mathbb{V}_n^*(\tau, \mathbf{b})$ is the bootstrap version of $\mathbb{V}_n(\tau, \mathbf{b})$, which is a process in $\mathbf{b} \in \mathbb{R}^p$ defined in expression (11) that denotes the asymptotic representation of $R_n(\tau)$ in the local model. The definition of $\mathbb{V}_n^*\{\tau, \mathbf{b}(\tau)\}$ can be found in expression (13). For computational convenience, we use the sample estimator $\hat{\sigma}_{n*}(\tau)$ in the bootstrap to avoid recalculating $\hat{\mathbf{J}}_k(\tau)$ for each bootstrap sample and each k, but one can also use the bootstrap variance estimator to achieve potential second-order accuracy (Efron and Tibshirani, 1993; Hall and Kang, 2001).

The idea of the modified bootstrap is as follows. If $\lambda_n(\tau) = o(n^{1/2})$ and $\lambda_n(\tau) \to \infty$, theorem 1 indicates that $P\{|T_n| > \lambda_n(\tau)\} \to I(\beta_0 \neq \mathbf{0})$; thus $R_n(\tau)$ can be bootstrapped consistently by the naive bootstrap. The challenge lies in bootstrapping $R_n(\tau)$ when $\beta_0(\tau) = \mathbf{0}$ since $\hat{k}_n(\tau)$ does not converge to $k_0(\tau)$ in this case. From lemma 4 and expression (S.15) in the on-line supplementary material, we know that, when $\beta_0(\tau) = \mathbf{0}$, the estimated asymptotic representation of $R_n(\tau)$ is $\mathcal{V}_n\{\tau, \mathbf{b}_0(\tau), \hat{k}_n(\tau)\}$, where

$$\mathcal{V}_n(\tau, \mathbf{b}, k) = \frac{(-\hat{\mu}_k(\tau), \hat{\pi}(\tau))\mathbb{G}_n[\tilde{\mathbf{X}}_k \psi_\tau\{\epsilon(\tau)\}]}{\hat{V}_k(\tau)\sigma_k(\tau)} + \left(\frac{\hat{\mathbf{C}}_k(\tau)}{\hat{V}_k(\tau)} - \frac{\hat{\mathbf{C}}_{\bar{k}_n(\tau,\mathbf{b})}}{\hat{V}_{\bar{k}_n(\tau,\mathbf{b})}(\tau)}\right)^{\mathrm{T}} \frac{\mathbf{b}}{\sigma_k(\tau)},$$

$\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - P_n)$ with P_n being the distribution of $(Y, \mathbf{X})$, $\hat{V}_k(\tau) = |\hat{\mathbf{J}}_k(\tau)|$, $(\hat{\pi}(\tau), \hat{\mu}_k(\tau))$ are the elements in the first row of $\hat{\mathbf{J}}_k$, and $\hat{\mathbf{C}}_k(\tau)$ is the kernel-based estimator of $\mathbf{C}_k(\tau)$.

Therefore, we define the $\mathbb{V}_n(\tau, \mathbf{b})$ process as

$$\begin{aligned} \mathbb{V}_n(\tau, \mathbf{b}) &= \mathcal{V}_n\{\tau, \mathbf{b}, \mathbb{K}_n(\mathbf{b})\}, \\ \mathbb{K}_n(\tau, \mathbf{b}) &= \arg\max_{k=1,\ldots,p} (\mathbb{G}_n[\tilde{\mathbf{X}}_k \psi_\tau\{\epsilon(\tau)\}] + \hat{\mathbf{B}}_k^{\mathrm{T}}(\tau)\mathbf{b})^{\mathrm{T}} \hat{\mathbf{J}}_k^{-1}(\tau)(\mathbb{G}_n[\tilde{\mathbf{X}}_k \psi_\tau\{\epsilon(\tau)\}] + \hat{\mathbf{B}}_k^{\mathrm{T}}(\tau)\mathbf{b}), \end{aligned} \tag{11}$$

where $\hat{\mathbf{B}}_k(\tau)$ is the kernel-based estimator of $\mathbf{B}_k(\tau)$.

The following theorem 2 shows that the adaptive bootstrap version $R_n^*\{\tau, \mathbf{b}_0(\tau)\}$ provides a consistent estimator of the distribution of $R_n(\tau)$ jointly over $\tau \in \mathcal{T}$.

Theorem 2. Suppose that the assumptions in theorem 1 hold, and $\lambda_n(\tau) \to \infty$ and $\lambda_n(\tau) = o(n^{1/2})$ for all $\tau \in \mathcal{T}$. Then $R_n^*\{\tau, \mathbf{b}_0(\tau)\}$ converges to the limiting distribution of $R_n(\tau)$ jointly over $\tau \in \mathcal{T}$ conditionally (on the observed data) in probability.

For the calibration of the test statistic, we need to compute only $R_n^*(\tau) \doteq R_n^*(\tau, \mathbf{0})$, the bootstrap statistic with $\mathbf{b} = \mathbf{0}$. Let $R_{n(j)}^*(\tau)$, $j = 1, \ldots, m$, be the bootstrap statistics from m bootstrap repetitions. For a level γ test, we can calculate the lower and upper critical values of $T_n(\tau)$ by the $(\gamma/2)$th and $(1-\gamma/2)$th sample quantiles of $\{R_{n(j)}^*(\tau), j = 1, \ldots, m\}$. If $T_n(\tau)$ is smaller than the lower or larger than the upper critical values, we shall reject the null hypothesis and declare that there is at least one predictor that has a significant effect on the τth quantile of Y. Alternatively,

we can also compute the p-value as $2\min[m^{-1}\Sigma_{j=1}^{m} I\{R^*_{n(j)}(\tau) > T_n(\tau)\}, m^{-1}\Sigma_{j=1}^{m} I\{R^*_{n(j)}(\tau) < T_n(\tau)\}]$.

2.5. Testing across multiple quantiles

One attractive feature of quantile regression is that it enables us to assess the relationship between Y and $\mathbf{X}$ at multiple quantiles. We now discuss testing across multiple quantiles through marginal quantile regression. Let $\mathcal{T} = \{\tau_1, \ldots, \tau_L\}$ be a set of quantile levels of interest. To test whether any component of $\mathbf{X}$ has an effect on either of the L-quantiles, we can consider testing $\tilde{H}_0 : \theta_0(\tau_1) = \ldots = \theta_0(\tau_L) = 0$ *versus* $\tilde{H}_a$: at least one of the $\theta_0(\tau_l)$s is non-zero, where $\theta_0(\tau_l)$ is the slope of the most informative predictor at the τ_lth quantile as defined in equation (3).

To pool information across quantiles, we propose to consider the sum-type test statistic $S_n = \Sigma_{l=1}^{L} T_n^2(\tau_l)$, where $T_n(\tau)$ is defined in equation (6). By the joint convergence result in theorem 2 over $\tau \in \mathcal{T}$, we can extend the adaptive bootstrap for the test calibration of S_n. Specifically, define the modified bootstrap statistic as $S_n^* = \Sigma_{l=1}^{L} R_n^{*2}(\tau_l, \mathbf{0})$. The p-value of the multiple-quantile test can then be calculated by the proportion of modified bootstrap statistics that are larger than S_n.

One may also consider the maximum-type test statistic $\max_{1 \leqslant l \leqslant L} |T_n(\tau_l)|$ to combine information across quantiles. We choose the sum-type test statistic as it has good power against dense alternatives, and in quantile regression $\theta_0(\tau)$ is more likely to be non-zero in an interval of τ and thus dense when the alternative is true. The joint test for $\tilde{H}_0$ is an omnibus test and may not be informative in applications where the linear effect exists at some but not all quantile levels. In such cases, we may apply the test at several individual quantile levels after the joint test rejects $\tilde{H}_0$ to obtain a more comprehensive picture.

2.6. Selection of the tuning parameter $\lambda_n(\tau)$

The tuning parameter $\lambda_n(\tau)$ is involved in the pre-test with criterion $|T_n(\tau)| > \lambda_n(\tau)$ to determine whether a conventional bootstrap $R_n^*(\tau)$ or a bootstrap of the asymptotic representation $\mathbb{V}_n(\tau, \mathbf{b})$ should be used. In the extreme case with $\lambda_n(\tau) = 0$, the bootstrap proposed reduces to the naive bootstrap, which gives an inflated type I error rate. When $\lambda_n(\tau) = o(n^{1/2})$ and $\lambda_n(\tau) \to \infty$, satisfying the conditions in theorem 2, the pre-test has asymptotically negligible type I error as $\lim_{n\to\infty} P\{|T_n(\tau)| > \lambda_n(\tau) | \theta_n(\tau) = 0\} = 0$, so only the bootstrap of $\mathbb{V}_n(\tau, \mathbf{0})$ is used under the null hypothesis that $\theta_n(\tau) = 0$. In finite samples, however, if $\lambda_n(\tau)$ is overly large, the test will be too conservative. Our empirical investigation shows that, for large samples, $\lambda_n(\tau) = c\sqrt{\{\tau(1-\tau)\log(n)\}}$ with $c \in [4, 10]$ provides a good choice, where the rate chosen is proportional to the expected value of the maximum of n sub-Gaussian variables with mean 0. In finite samples, we propose to use a double-bootstrap procedure to choose the constant c. For illustration, we describe the procedure for testing at a single quantile τ as follows.

For a bootstrap sample $\{y_i^*, \mathbf{x}_i^*, i = 1, \ldots, n\}$, obtain the first-level bootstrap statistics $\hat{\theta}_n^*(\tau)$ and $\hat{\sigma}_{n*}(\tau)$, and m double-bootstrap statistics $R_{(1)}^{**}, \ldots, R_{(m)}^{**}$, the analogies of $R_n^*(\tau, \mathbf{0})$ as defined in expression (10), associated with a candidate c-value. For each c in the grid, we calculate the rejection rate as the proportion of bootstrap samples for which $n^{1/2}\{\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau)\}/\hat{\sigma}_{n*}(\tau)$ exceeds the double-bootstrap critical values, determined by the lower bound of the $(\gamma/2)$th and the upper bound of the $(1-\gamma/2)$th sample quantiles of $\{R_{(1)}^{**}, \ldots, R_{(m)}^{**}\}$. The double bootstrap then chooses the c-value that gives the rejection rate that is closest to the nominal significance level γ. The double bootstrap could be time consuming, so in practice we recommend using a double bootstrap with $m = 100$ for smaller samples of $n \leqslant 500$ and choosing quite a conservative $c = 5$ for larger samples.

3. Simulation study

3.1. Under linear regression models

We first assess the performance of the proposed test in linear regression models by considering four cases. For cases 1–3, data are generated from the model $Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon$, where $\epsilon \sim N(0,1)$ in cases 1 and 2 and $\epsilon \sim t_2$ in case 3. Case 1 corresponds to the null hypothesis of no active predictors with $\boldsymbol{\beta} = \mathbf{0}$. In case 2, we let $\boldsymbol{\beta} = (\frac{1}{3}, 0, \ldots, 0)^{\mathrm{T}}$ so there is a unique active predictor. In case 3, we let $\beta_1 = \ldots = \beta_5 = 0.25$, $\beta_6 = \ldots = \beta_{10} = -0.15$ and $\beta_j = 0$ for $j = 10, \ldots, p$, so there are 10 active predictors and k_0 is not unique. For case 4, data are generated from the model $Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + (1 + 0.45X_1)\epsilon$, where $\boldsymbol{\beta} = \mathbf{0}$ and $\epsilon \sim N(0,1)$. Case 4 represents a heteroscedatic case, where the predictor X_1 is active at $\tau \neq 0.5$ but inactive at $\tau = 0.5$. In all cases, the covariate vector $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ is from the multivariate normal distribution with mean 0, variance 1 and an exchangeability correlation of 0.5, truncated at -2 and 2. For each case, a random sample of size $n = 200$ of $(Y, X_1, \ldots, X_p)$ is generated, and the simulation is repeated 500 times. Even though the asymptotic results assume a finite p, we shall demonstrate that the test works empirically also for cases with $p > n$ through the analysis for five dimensions, $p = 10$, 100, 200, 400, 1000.

We apply the proposed quantile marginal effect test (QMET) at a single quantile level $\tau = 0.5$ and $\tau = 0.75$, separately, and at three quantiles $\tau = \{0.25, 0.5, 0.75\}$ jointly to detect the predictor effects on the conditional quantiles of Y. For comparison, we also include three competing methods:

(a) analysis of variance, AOV;
(b) Bonferroni, BONF;
(c) the centred percentile bootstrap CPB for single-quantile testing.

Each test is calibrated at the nominal level $\gamma = 0.05$. AOV is a rank-score-type method for testing the null hypothesis that no predictors have significant effects against the alternative that at least one predictor has a significant effect, and this method is feasible only in cases where $p < n$. The AOV method is implemented in the function '`anova.rq`' of the R package `quantreg`. BONF is a method of multiple comparison with Bonferroni adjustments, where a rank score test is carried out in the marginal quantile regression of Y on each X_j, $j = 1, \ldots, p$, and the null hypothesis is rejected if the minimum of p resulting p-values is smaller than γ. The rank-score-based tests are chosen for AOV and BONF since such tests were shown to be more stable than the Wald-type tests in the quantile regression literature (Kocherginsky *et al.*, 2005). CPB is the centred percentile bootstrap method based on the conventional bootstrap statistic $R_n^*(\tau)$. For both CPB and the QMET, 200 bootstrap samples are used. For the QMET method, we let the threshold $\lambda_n(\tau) = c\sqrt{\{\tau(1-\tau)\log(n)\}}$ and choose $c \in (0, 6)$ by a double bootstrap with 100 double-bootstrap samples.

Table 1 summarizes the rejection rates of the various methods against p. The rejection rates for case 1 with $\tau = 0.5$, 0.75 and $\tau = \{0.25, 0.5, 0.75\}$, and for case 4 with $\tau = 0.5$ correspond to type I errors where the null hypothesis is true, whereas they represent power in the other scenarios. In all three null scenarios, the CPB method gives high type I errors, confirming that the conventional bootstrap procedure fails to control the familywise error rate as it does not account for the uncertainty that is involved in the variable selection process. The AOV test is not feasible in cases with $p \geqslant n$, and it is highly conservative with power close to 0 for $p \geqslant 100$. The Bonferroni correction method has similar performance to that of the QMET for $p = 10$, but the QMET method is in general more powerful for larger p. Compared with the other methods, the QMET method with a double bootstrap maintains the level reasonably well and

Table 1. Percentages of rejections from various methods in cases 1–4 with $n = 200$†

Case	*p*	*Results for* $\tau = 0.5$				*Results for* $\tau = 0.75$				*Results for multivariate*
		CPB	*AOV*	*BONF*	*QMET*	*CPB*	*AOV*	*BONF*	*QMET*	*QMET*
1	10	27.8	6.6	5.2	3.8	24.8	6.2	3.2	5.6	2.4
	100	35.4	0.0	2.4	3.0	37.2	0.0	2.4	4.8	2.8
	200	39.4		2.2	4.6	34.6		1.8	5.6	2.6
	400	42.2		2.0	4.4	42.4		2.0	8.0	2.4
	1000	48.8		1.6	4.6	42.0		1.2	4.8	1.6
2	10	86.5	61.0	79.6	75.0	78.8	50.6	73.0	64.8	83.8
	100	85.7	0.0	67.2	67.0	75.4	0.2	55.8	54.4	70.2
	200	83.2		59.2	61.5	77.2		47.6	59.2	65.6
	400	84.8		58.8	62.1	74.8		43.0	60.0	70.1
	1000	85.7		50.6	64.6	75.6		38.2	61.2	65.4
3	10	94.0	94.2	91.6	85.4	75.0	70.2	70.6	62.0	84.2
	100	88.2	0.0	75.6	71.8	71.0	0.4	48.8	47.6	72.5
	200	89.6		74.2	72.0	71.2		52.2	50.2	71.3
	400	84.4		67.4	71.6	69.4		44.8	49.4	68.8
	1000	87.4		64.0	74.4	70.2		39.0	52.6	70.8
4	10	24.0	7.2	6.6	6.0	90.8	47.8	70.8	80.6	77.4
	100	26.2	0.0	4.4	3.2	86.0	0.0	53.4	61.4	46.8
	200	25.0		3.0	4.0	86.4		44.0	58.4	43.4
	400	30.6		2.2	5.8	85.8		41.4	57.0	30.7
	1000	28.6		2.2	6.6	87.4		38.2	58.2	26.8

†Case 1 with $\tau = 0.5$, 0.75 and case 4 with $\tau = 0.5$ correspond to the null model, and the others correspond to the alternative model. The last column is the proposed test across three quantiles 0.25, 0.5 and 0.75.

Table 2. Percentages of rejections of the QMET and Bonferroni methods at $\tau = 0.5$ in cases 5 and 6

Method	*Results for case 5*					*Results for case 6*				
	$p = 10$	$p = 100$	$p = 200$	$p = 400$	$p = 1000$	$p = 10$	$p = 100$	$p = 200$	$p = 400$	$p = 1000$
QMET	3.6	1.6	2.1	2.3	4.1	65.8	50.0	49.0	53.1	60.5
BONF	4.0	3.2	6.0	2.6	3.2	65.0	46.0	38.6	40.2	34.2

it provides relatively high power across all scenarios that were considered. Under alternative models with homoscedastic errors, the QMET across quantiles tends to be more powerful than the single-quantile test especially at the tail quantiles. However, in heteroscedastic models (case 4) where the signal has different magnitudes at different quantiles, it may be more advantageous to apply the test at a single quantile level with a stronger signal than the omnibus test across quantiles.

3.2. Under misspecified non-linear regression models

We consider two additional cases to assess the performance of the proposed test under the misspecification of linear models and for covariates with a different correlation structure. The data are generated from the model $Y = X_1^2/3 + bX_1 + \epsilon$, where $\epsilon \sim N(0, 1)$, and the covariates $(X_1, \ldots, X_p)$ are generated in the same way as in cases 1–4 but with an auto-regressive AR(1) correlation structure of parameter 0.5. We let $b = 0$ in case 5, and $b = \frac{1}{3}$ in case 6. Table 2

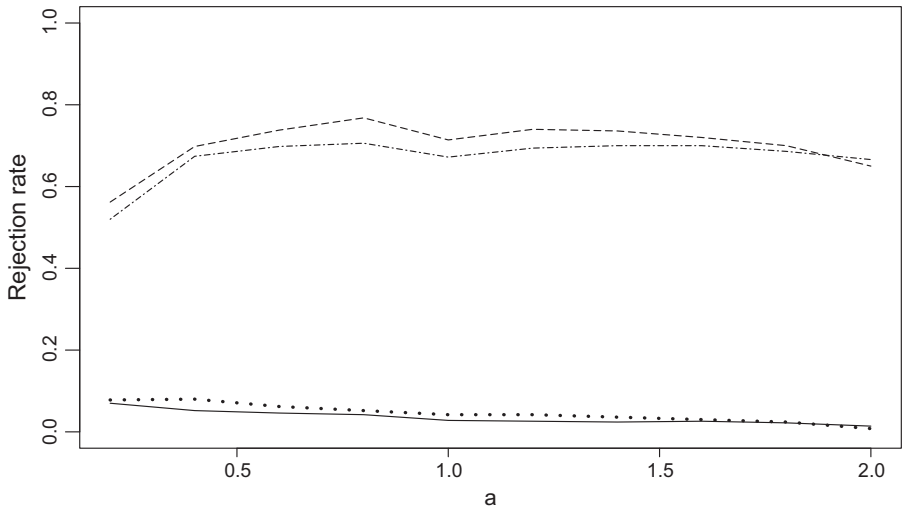

Fig. 1. Rejection rates of QMET at $\tau = 0.5$ in cases 1–4 against the constant a involved in the bandwidth parameter h_n: ———, case 1, ·-·-·-, case 2; - - - - - -, case 3; · · · · · ·, case 4

summarizes the rejection rates of the QMET and BONF methods at $\tau = 0.5$ in cases 5 and 6. Note that the test proposed is only for detecting linear covariate effects. When there is no linear relationship between $\mathbf{X}$ and Y at any quantiles (case 5), the QMET has difficulty identifying the non-linear relationship. However, if there is some linear trend in a misspecified model such as in case 6, the QMET can still identify the covariate effects with higher power than the Bonferroni method. The idea of the QMET method may be extended to detect non-linear covariate effects by marginally regressing Y on some polynomial or basis functions of each covariate separately.

3.3. Sensitivity against the bandwidth h_n

The calculation of $\hat{\sigma}_k(\tau)$ involves a bandwidth parameter h_n. In our implementation, we follow the suggestion in Hall and Sheather (1988) and choose $h_n = an^{-1/3}\{\Phi^{-1}(1-\gamma/2)\}^{2/3} \times (1.5\phi^2\{\Phi^{-1}(\tau)\}/[2\{\Phi^{-1}(\tau)\}^2+1])^{2/3}$ with $a = 1$, where Φ is the distribution function of $N(0,1)$. To assess the sensitivity of the test proposed against h_n, we plot the rejection rates of QMET at $\tau = 0.5$ against a in cases 1–4 in Fig. 1. Results suggest that the performance of the test is quite stable for $a \in [0.5, 1.5]$.

4. Application to the study of human immunodeficiency virus drug resistance

We illustrate the method proposed by analysing an HIV drug susceptibility data set from the HIV drug resistance database (`http://hivdb.stanford.edu`), which is a public resource for the study of sequence variation and mutations in the molecular targets of HIV drugs (Rhee *et al.*, 2003). After a patient starts antiretroviral therapy, the infecting HIV can form new mutations. Some mutations may not respond to existing drugs, which is a characteristic known as drug resistance or reduced drug susceptibility, meaning that the drugs become less effective at preventing the virus from multiplying. Researchers have estimated that, in an untreated HIV-infected subject, every possible single point mutation occurs between 10^4 and 10^5 times per day (Coffin, 1995). Drug resistance has become a major obstacle to the success of HIV therapy. Therefore, understanding the effect of mutations on drug resistance is an important research topic.

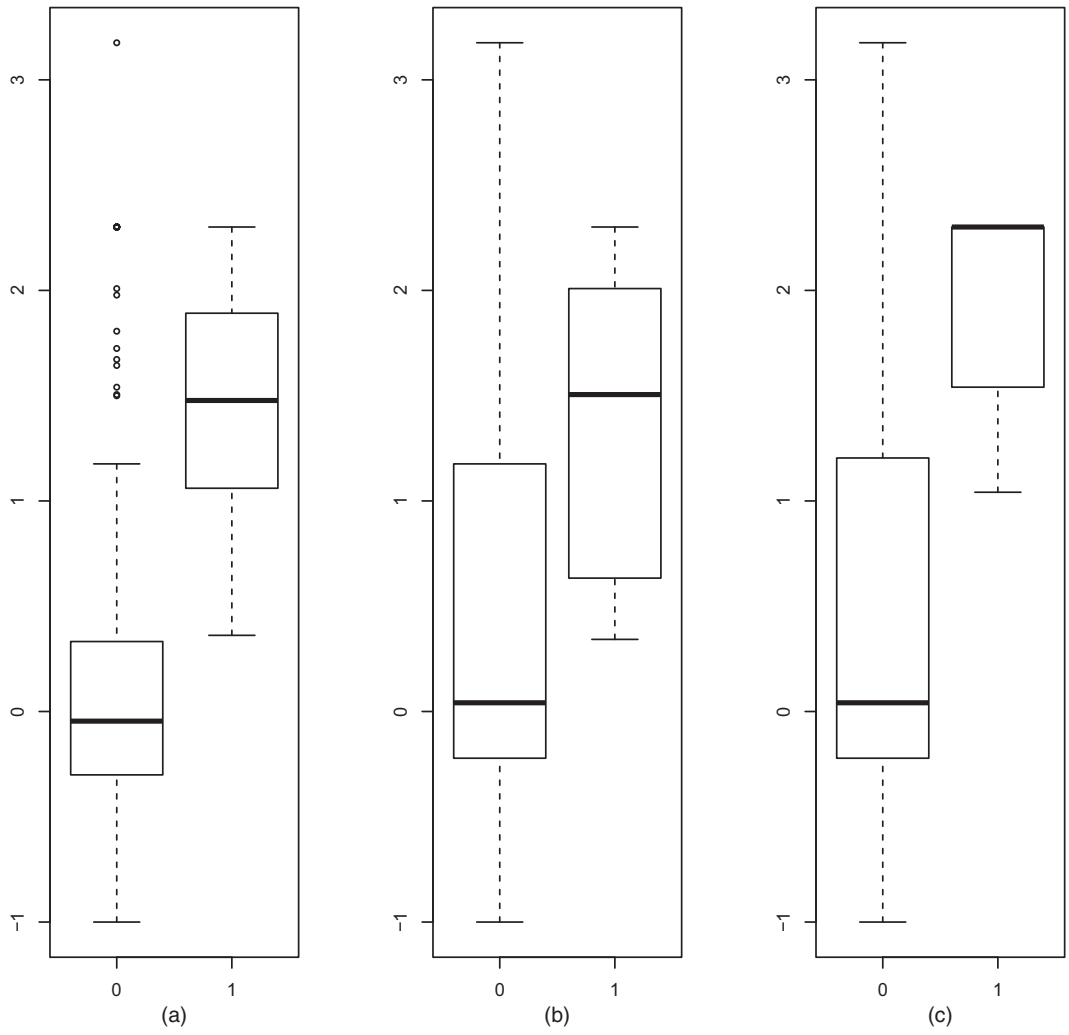

Fig. 2. Boxplots of common logarithmic susceptibility of a training data set *versus* three predictors (a) 103N, (b) 190A and (c) 230L that are selected by the forward QMET method at $\tau = 0.75$

We analyse the susceptibility data for the drug efavirenz. After excluding rare mutations, the data set includes 1472 HIV isolates and 197 locations of mutations. The susceptibility of an HIV sample is defined as the fold decrease in susceptibility of a single virus isolate compared with the susceptibility of a wild-type control isolate, i.e. the virus that has never been challenged by drugs. We focus on predicting common logarithmic susceptibility, denoted by Y, to efavirenz based on X_k, $k = 1, \ldots, p = 197$, indicating the presence of a mutation of interest in the kth viral sequence position. The susceptibility data are highly non-normal, even after log-transformation (see Fig. 2), so quantile regression provides a valuable way of analysing these data. In addition, analysis at the upper quantiles of susceptibility is of particular interest, being associated with stronger drug resistance. In this analysis, we consider two quantile levels: $\tau = 0.5$ and $\tau = 0.75$.

The test proposed can be used as a stopping rule in forward regression to select multiple significant predictors. Specifically, suppose that the test detects a significant predictor in the

first step. We can then use the residuals $Y - \hat{\theta}_n(\tau)X_{\hat{k}_n(\tau)}$ as a new outcome variable and carry out the marginal quantile regression over the remaining predictors. Repeat the procedure and record the p-values in the sequential tests as $p_1, p_2, \ldots$. The procedure will stop at the mth step, where $m = \inf\{j : p_j > \gamma\}$ and γ is the nominal level of significance. To account for the sequential testing that is involved in the forward regression, we further employ a multiple-test adjustment in the style of Holm (1979) and finally choose the covariates that are identified in the first $\tilde{m}$ steps, where $\tilde{m} = 1$ if $p_1 > \gamma/(m-1)$, and otherwise $\tilde{m} = \max_{1 \leqslant j \leqslant m-1}\{j : p_l \leqslant \gamma/(m-l)$ for all $l = 1, \ldots, j\}$.

To assess the performance of the method in settings with $n < p$, we randomly split the data into a training set of size $n = 190$ and a testing set of size 1282. For each split, we carry out 20 steps of forward quantile marginal effect testing (the QMET) and standard forward selection quantile regression, FWD, using the training data, and use the model that is selected in each step to predict the τth quantile of log-susceptibility of the testing data.

Figs 3(a) and 3(b) plot the training set p-values (plus and minus the median absolute deviation MAD) for the newly entered predictor at each step across 50 random splits at $\tau = 0.5$ and $\tau = 0.75$, and Figs 3(c) and 3(d) plot the corresponding prediction errors (median $\pm$ MAD) in the test sets. At a quantile level τ, the quantile prediction error is defined as the average of quantile loss $\rho_\tau(Y - \hat{Y})$, where Y is the true response and $\hat{Y}$ is the predicted value for subjects from the testing data. At the 0.05 level of significance, the QMET method selects about one predictor at the median and three at level $\tau = 0.75$, whereas the FWD method selects more than 20 predictors. The prediction error plots suggest that, with the QMET method, the improvement in prediction error becomes negligible after the first three predictors enter the model. In contrast, the prediction accuracy of the FWD method improves much more slowly. Generally speaking, at both quantile levels, FWD enters about 15 predictors to achieve the same prediction accuracy as the model selected by the QMET with just three predictors.

The obvious difference in the prediction accuracy suggests that two methods enter different predictors at each stage, which is not surprising given their different selection criteria. In each step of the forward QMET method, the QMET procedure is applied by treating residuals from the previous stage as new outcomes and identifies the predictor that gives the smallest quantile loss in marginal regression. In contrast, the standard FWD method identifies the predictor that gives the smallest Wald-type p-value conditionally on the predictors that have entered before the current step.

To demonstrate the value of regression at different quantiles, we look into one example training set, for which the forward QMET method at $\tau = 0.75$ at the level of significance of 0.05 selects three predictors, 103N, 190A and 230L, sequentially, whereas the method at the median selects only the first predictor. The binary predictors 103N, 190A and 230L indicate the presence of substitution of amino acid asparagine, alanine and leucine at positions 103, 190 and 230 respectively. Fig. 2 shows the boxplots of the common logarithmic susceptibility of the isolates in the training set with and without the three mutations. The boxplots suggest that isolates with these three mutations are associated with higher drug resistance at both the median and $\tau = 0.75$. After accounting for the effects of the first mutation, the effects of 190A and 230L become insignificant at the median but remain significant at $\tau = 0.75$, i.e. for the isolates that are more drug resistant; Table 3.

5. Discussion

We have developed a new procedure for detecting marginal effects in quantile regression. Our simulation study suggests that the test is effective and has stable performance, providing adequate control of the familywise error rate along with competitive power, even for cases with large p

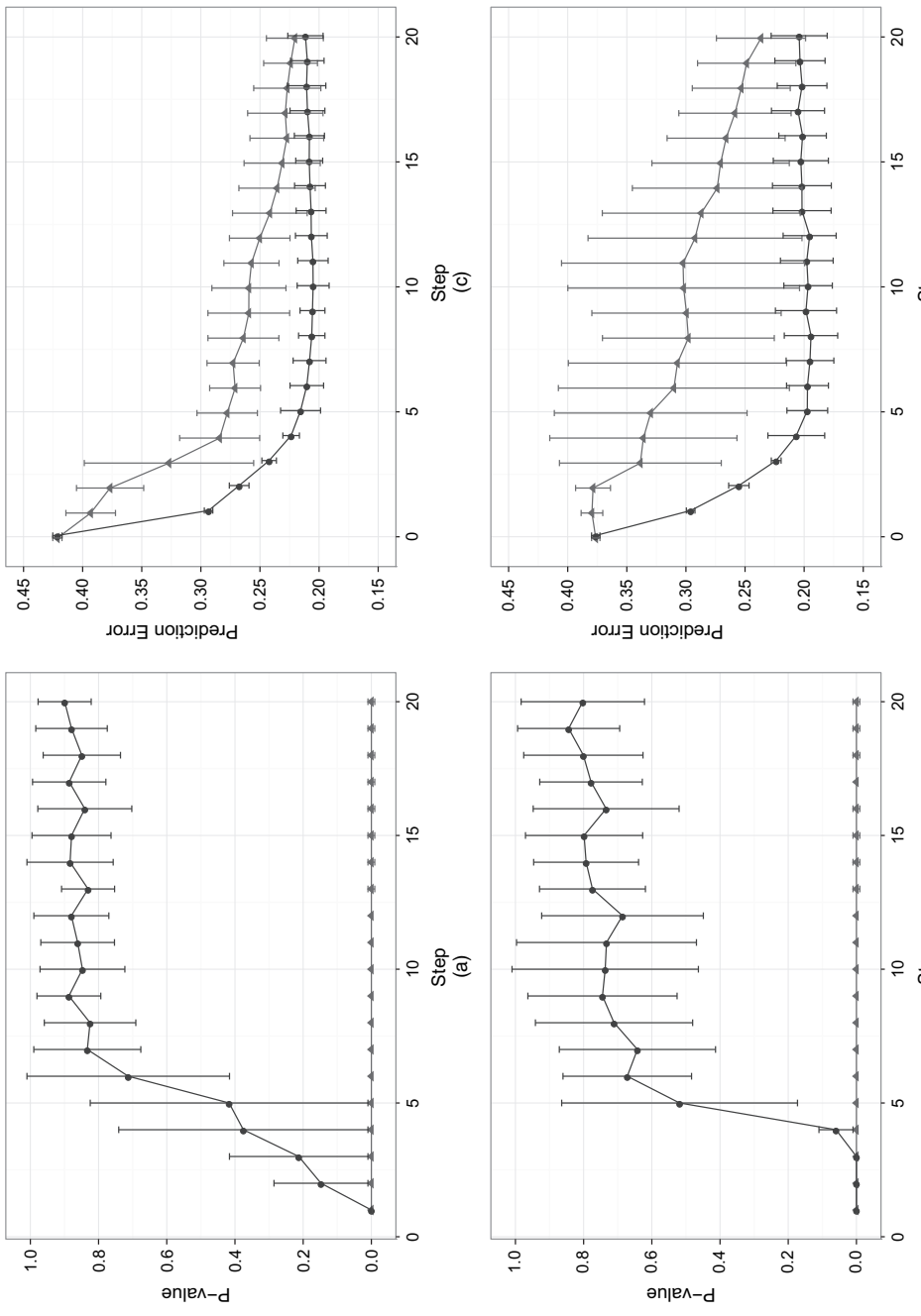

Fig. 3. (a), (b) Training set *p*-values (median ± MAD) for the newly entered predictor at each step for the standard forward quantile regression FWD (▲) and the proposed forward QMET method (●) across 50 random splits and (c), (d) corresponding prediction errors (median ± MAD) in the testing sets: (a), (c) $\tau = 0.5$; (b), (d) $\tau = 0.75$

Table 3. *p*-values from the forward QMET procedure for predictors 103N, 190A and 230L, and the corresponding estimated coefficients, standard errors and Wald-type *p*-values by regressing the log-susceptibility on all three predictors jointly at $\tau = 0.5$ and $\tau = 0.75$ in the example training set

τ	*Variable*	*Forward QMET p-value*	*Results from multiple quantile regression*		
			Coefficient	*Standard error*	*Wald p-value*
0.5	103N	0.0000	1.5593	0.1122	0.0000
	190A	0.8525	1.6021	0.5290	0.0028
	230L	0.7175	1.6372	0.8625	0.0592
0.75	103N	0.0000	1.7249	0.1827	0.0000
	190A	0.0125	1.8947	0.1900	0.0000
	230L	0.0000	2.1871	0.0928	0.0000

(although the asymptotic theory that we used to calibrate the test assumes fixed p). The theoretical study for diverging p is an interesting and challenging topic that deserves further research.

An alternative approach to the QMET beyond those which we have considered would be to use higher criticism, which is a way of synthesizing a collection of p-values from multiple tests that was originally proposed by Tukey (1989) and later systematically developed by Donoho and Jin (2004, 2015). However, as reported in McKeague and Qian (2015) for the mean regression version of our proposed test, higher criticism is typically anticonservative unless the predictors are close to being uncorrelated, which would be a highly restrictive assumption in most applications.

Our proposed test statistic can be viewed as a maximum-type test statistic across p covariates. Similarly to the discussion as in Chatterjee and Lahiri (2015) for mean regression, we may also consider an alternative statistic based on the sum of squared t-statistics, $\Sigma_{k=1}^{p} \hat{\theta}_k^2(\tau)/\hat{\sigma}_k^2(\tau)$. Our empirical studies show that the maximum-type test is more powerful for detecting sparse signals, and the sum-type test has more power for dense alternatives; see the results in the on-line supplementary material. This observation agrees with the findings in mean regression (Cai *et al.*, 2014; Gregory *et al.*, 2015; Chen and Qin, 2010; Fan *et al.*, 2015). Therefore, we recommend the proposed test for scenarios with sparse signals. How to combine two types of test statistics to accommodate signals of unknown sparsity levels is an interesting future research topic.

The test proposed can be used as a first step in applications to assess the overall significance of covariates on quantiles of Y. If the null hypothesis is rejected, one can use existing model selection procedures to see how many and which variables should be included next. Alternatively, one may use the test as a stopping rule in forward regression to select multiple significant predictors in a sequential manner, and our empirical study has shown some promising evidence. There is some recent work for sequential forward regression on error rate control and the probabilistic bounds for the number of selected covariates (Fithian *et al.*, 2015; Li and Barber, 2017; G'Sell *et al.*, 2016; Kozbur, 2015; Tibshirani *et al.*, 2016). However, because of the dependence between the new outcomes and the non-additive property of quantile functions, it would be challenging and also beyond the scope of the paper to study the theoretical properties of the forward regression procedure.

6. Supplementary material

The supplementary material that is available on line contains the proofs of lemmas 1–4, corollary 1 and lemmas 5 and 6, and some additional simulation results.

Acknowledgements

The research is partly supported by National Science Foundation grants DMS-1149355, DMS-1307838 and DMS-1712760, National Institutes of Health grants 2R01GM095722 and R21MH108999 and grant OSR-2015-CRG4-2582 from King Abdullah University of Science and Technology. The authors thank the Joint Editor, Associate Editor and three referees for their constructive comments and suggestions.

Appendix A

A.1. Proof of theorem 1

The proof of theorem 1 follows immediately from the following lemmas 1 and 4, which establish the joint asymptotic representations of $n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}/\hat{\sigma}_n(\tau)$ over $\tau\in\mathcal{T}$ for $\boldsymbol{\beta}_0(\tau)\neq 0$ and $\boldsymbol{\beta}_0(\tau)=0$, separately. Lemma 2 builds an approximate connection between $(\alpha_k(\tau),\theta_k(\tau))$ and $(\alpha_0(\tau),\boldsymbol{\beta}_n(\tau))$ under the local model with $\boldsymbol{\beta}_0(\tau)=0$, which together with lemma 3 are needed for proving lemma 4. The proofs for lemmas 1–4 are provided in the on-line supplementary material.

We first fix some notation. For any vector $\mathbf{v}\in\mathbb{R}^p$, let v_k denote its kth element, and $\mathbf{v}_{(-k)}$ denotes its subset excluding the kth element, for any $k=1,\ldots,p$. For notational simplicity, we write $k_n\{\tau,\mathbf{b}_0(\tau)\}$ as $k_n(\tau)$, and omit the argument τ in various expressions such as $M_k(\tau,\cdot)$, $\pi_k(\tau,\cdot)$, $V_k(\tau,\cdot)$, $\mathbf{J}_k(\tau,\cdot)$ etc. when needed.

Lemma 1. Suppose that assumptions 1–5 hold. For all τs in $\mathcal{T}$ for which $\beta_0(\tau)\neq 0$ and $k_0(\tau)$ is unique, we have $\hat{k}_n(\tau)\to k_0(\tau)$ almost surely and

$$\frac{n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}}{\hat{\sigma}_n(\tau)}\xrightarrow{\mathrm{d}}\frac{M_{p+k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\pi_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}-M_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\mu_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}}{V_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\sigma_{k_0(\tau)}(\tau)}.$$

Lemma 2. If assumptions 1–5 hold, we have

$$\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}=\begin{pmatrix}\alpha_0(\tau)\\ \beta_{n,k}(\tau)\end{pmatrix}+\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{A}_k^{\mathrm{T}}(\tau)\boldsymbol{\beta}_{n,(-k)}(\tau)+o(n^{-1/2}) \tag{12}$$

uniformly over τ for which $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$, where $\mathbf{A}_k(\tau)=(E[f_{\epsilon(\tau)}(0|\mathbf{X})\mathbf{X}_{(-k)}],E[f_{\epsilon(\tau)}(0|\mathbf{X})X_k\mathbf{X}_{(-k)}])$.

Remark 1. For the homoscedastic case where $f_{\epsilon(\tau)}(\cdot|\mathbf{X})$ is common across $\mathbf{X}$, by lemma 2, we have that uniformly in $\tau\in\mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$,

$$\theta_k(\tau)=\beta_{n,k}(\tau)+\frac{\mathrm{cov}(X_k,\mathbf{X}_{(-k)}^{\mathrm{T}})\boldsymbol{\beta}_{(-k)}(\tau)}{\mathrm{var}(X_k)}+o(n^{-1/2})=\frac{\mathrm{cov}(X_k,\mathbf{X}^{\mathrm{T}})\boldsymbol{\beta}_n(\tau)}{\mathrm{var}(X_k)}+o(n^{-1/2}).$$

Lemma 3. If assumptions 1–5 hold, we have

$$\left(n^{1/2}\begin{pmatrix}\hat{\alpha}_k(\tau)-\alpha_0(\tau)\\ \hat{\theta}_k(\tau)-\beta_{n,k}(\tau)\end{pmatrix}\right)_{k=1}^{p}\xrightarrow{\mathrm{d}}\left(\mathbf{J}_k^{-1}(\tau,\mathbf{0})\left\{\begin{pmatrix}M_k(\tau,\mathbf{0})\\ M_{p+k}(\tau,\mathbf{0})\end{pmatrix}+\mathbf{A}_k^{\mathrm{T}}(\tau)\mathbf{b}_{0,(-k)(\tau)}\right\}\right)_{k=1}^{p}$$

uniformly in $\tau\in\mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$.

Lemma 4. Under assumptions 1–5, we have

$$\begin{aligned}\frac{n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}}{\hat{\sigma}_n(\tau)}\xrightarrow{\mathrm{d}}&\frac{M_{p+K(\tau)}(\tau,\mathbf{0})\pi(\tau,\mathbf{0})-M_{K(\tau)}(\tau,\mathbf{0})\mu_{K(\tau)}(\tau,\mathbf{0})}{V_{K(\tau)}(\tau,\mathbf{0})\sigma_{K(\tau)}(\tau)}\\&+\left(\frac{\mathbf{C}_{K(\tau)}(\tau)}{V_{K(\tau)}(\tau,\mathbf{0})}-\frac{\mathbf{C}_{\kappa_\tau\{\mathbf{b}_0(\tau)\}}(\tau)}{V_{\kappa_\tau\{\mathbf{b}_0(\tau)\}}(\tau,\mathbf{0})}\right)^{\mathrm{T}}\frac{\mathbf{b}_0(\tau)}{\sigma_{K(\tau)}(\tau)}\end{aligned}$$

uniformly over $\tau\in\mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$, where

$$K(\tau)=\underset{k=1,\ldots,p}{\arg\max}\,(\mathbf{M}_k(\tau)+\mathbf{B}_k^{\mathrm{T}}(\tau)\mathbf{b}_0(\tau))^{\mathrm{T}}\mathbf{J}_k^{-1}(\tau,\mathbf{0})(\mathbf{M}_k(\tau)+\mathbf{B}_k^{\mathrm{T}}(\tau)\mathbf{b}_0(\tau))$$

with $\mathbf{M}_k(\tau)=(M_k(\tau,\mathbf{0}),M_{k+p}(\tau,\mathbf{0}))^{\mathrm{T}}$.

A.2. Proof of corollary 1

The proof of corollary 1 is provided in the on-line supplementary material.

A.3. Proof of theorem 2

Let $\mathbb{P}_n^*$ denote bootstrap average, and $\mathbb{G}_n^* = (\mathbb{P}_n^* - \mathbb{P}_n)\sqrt{n}$. Define $\hat{\epsilon}_n(\tau) = Y - \hat{\alpha}_n(\tau) - \hat{\theta}_n(\tau)X_{\hat{k}_n(\tau)}$. The bootstrapped process $\mathbb{V}_n^*\{\tau, \mathbf{b}(\tau)\}$ is defined as

$$\mathbb{V}_n^*\{\tau, \mathbf{b}(\tau)\} = \frac{(-\hat{\mu}_{\mathbb{K}_n^*(\tau,\mathbf{b})}(\tau), \hat{\pi}(\tau))\mathbb{G}_n^*[\tilde{\mathbf{X}}_{\mathbb{K}_n^*(\tau,\mathbf{b})}\psi_\tau\{\hat{\epsilon}_n(\tau)\}]}{\hat{V}_{\mathbb{K}_n^*(\tau,\mathbf{b})}(\tau)\hat{\sigma}_{\mathbb{K}_n^*(\tau,\mathbf{b})}(\tau)} + \left(\frac{\hat{\mathbf{C}}_{\mathbb{K}_n^*(\tau,\mathbf{b})}(\tau)}{\hat{V}_{\mathbb{K}_n^*(\tau,\mathbf{b})}(\tau)} - \frac{\hat{\mathbf{C}}_{\mathbb{K}_n(\tau,\mathbf{b})}}{\hat{V}_{\mathbb{K}_n(\tau,\mathbf{b})}(\tau)}\right)^{\mathrm{T}} \frac{\mathbf{b}}{\hat{\sigma}_{\mathbb{K}_n^*(\tau,\mathbf{b})}(\tau)}, \tag{13}$$

where $\mathbb{K}_n(\tau, \mathbf{b}) = \arg\min \mathbb{L}_{n,k}(\tau, \mathbf{b})$, $\mathbb{K}_n^*(\tau, \mathbf{b}) = \arg\max_k \mathbb{U}_{n,k}^*(\tau, \mathbf{b})$,

$$\mathbb{L}_{n,k}(\tau, \mathbf{b}) = \min_{\alpha,\theta} \mathbb{P}_n(\rho_\tau[\epsilon(\tau) + \alpha_0(\tau) + \mathbf{X}^{\mathrm{T}}\{\boldsymbol{\beta}_0(\tau) + n^{-1/2}\mathbf{b}\} - \alpha - \theta X_k]),$$

and

$$\mathbb{U}_{n,k}^*(\tau, \mathbf{b}) = (\mathbb{G}_n^*[\tilde{\mathbf{X}}_k\psi_\tau\{\hat{\epsilon}_n(\tau)\}] + \hat{\mathbf{B}}_k^{\mathrm{T}}(\tau)\mathbf{b})^{\mathrm{T}}\hat{\mathbf{J}}_k^{-1}(\tau)(\mathbb{G}_n^*[\tilde{\mathbf{X}}_k\psi_\tau\{\hat{\epsilon}_n(\tau)\}] + \hat{\mathbf{B}}_k^{\mathrm{T}}(\tau)\mathbf{b}).$$

Let E^M denote expectation conditional on the data, and let P^M be the corresponding probability measure. We shall show that $I\{|T_n^*(\tau)| > \lambda_n(\tau) \text{ or } |T_n(\tau)| > \lambda_n(\tau)\} \to^{P^M} I\{\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}\}$ and $I\{|T_n^*(\tau)| \leqslant \lambda_n(\tau)\}I\{|T_n(\tau)| \leqslant \lambda_n(\tau)\} \to^{P^M} I\{\boldsymbol{\beta}_0(\tau) = \mathbf{0}\}$ for all $\tau \in \mathcal{T}$ conditionally (on the data) in probability. This together with lemmas 5 and 6 below implies the result.

For $k = 1, \ldots, p$, the bootstrapped marginal regression coefficients satisfy

$$(\hat{\alpha}_k^*(\tau), \hat{\theta}_k^*(\tau)) = \arg\min_{\alpha,\theta} \mathbb{P}_n^*\{\rho_\tau(Y - \alpha - \theta X_k)\}.$$

By the first-order condition, we have $\mathbb{P}_n^*[\psi_\tau\{Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k\}\tilde{\mathbf{X}}_k] = 0$ for $\tau \in \mathcal{T}$. Similarly, by the definition of $(\alpha_k(\tau), \theta_k(\tau))$ in equation (2), $E[\psi_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}\tilde{\mathbf{X}}_k] = 0$. Under the assumptions that are listed in Section 2.3, it can be verified that, for any $\eta > 0$, the class of functions $\{\psi_\tau(Y - \alpha - \theta X_k)\tilde{\mathbf{X}}_k : \tau \in \mathcal{T}, (\alpha, \theta) \in \mathbb{R}^2, \sup_{\tau\in\mathcal{T}} \|(\alpha, \theta) - (\alpha_k(\tau), \theta_k(\tau))\| \leqslant \eta\}$ is a P-Donsker class, and $E\|[\psi_\tau(Y - \alpha_n - \theta_n X_k) - \psi_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}]\tilde{\mathbf{X}}_k\|^2 \to 0$ for any sequence (α_n, θ_n) such that $|(\alpha_n, \theta_n) - (\alpha_k(\tau), \theta_k(\tau))| \to 0$ for all $\tau \in \mathcal{T}$.

Using similar arguments to those in the proof of theorem 10.16 of Kosorok (2008), we have

$$n^{1/2}\begin{pmatrix} \hat{\alpha}_k^*(\tau) - \alpha_k(\tau) \\ \hat{\theta}_k^*(\tau) - \theta_k(\tau) \end{pmatrix} = \mathbf{J}_k^{-1}\{\tau, \boldsymbol{\beta}_0(\tau)\}n^{1/2}\mathbb{P}_n^*[\psi_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}\tilde{\mathbf{X}}_k] + o_p(1). \tag{14}$$

This, together with expression (S.2) in the on-line supplementary material, implies that

$$n^{1/2}\begin{pmatrix} \hat{\alpha}_k^*(\tau) - \hat{\alpha}_k(\tau) \\ \hat{\theta}_k^*(\tau) - \hat{\theta}_k(\tau) \end{pmatrix} = \mathbf{J}_k^{-1}\{\tau, \boldsymbol{\beta}_0(\tau)\}\mathbb{G}_n^*[\psi_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}\tilde{\mathbf{X}}_k] + o_p(1) \tag{15}$$

for all $\tau \in \mathcal{T}$ for $k = 1, \ldots, p$.

When $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, note that

$$|T_n^*(\tau)| = \left|\frac{n^{1/2}\hat{\theta}_n^*(\tau)}{\hat{\sigma}_{n*}(\tau)}\right| \leqslant \max_{k=1,\ldots,p}\left|\frac{n^{1/2}\{\hat{\theta}_k^*(\tau) - \hat{\theta}_k(\tau)\}}{\hat{\sigma}_k(\tau)} + \frac{n^{1/2}\hat{\theta}_k(\tau)}{\hat{\sigma}_k(\tau)}\right|$$

where $\hat{\sigma}_k^2(\tau)$ is the lower right-hand diagonal element of $\hat{\Sigma}_k(\tau)$, the consistent estimator of $\Sigma_k(\tau)$, which is assumed in assumption 5 to be bounded away from zero for all $\tau \in \mathcal{T}$ and k. This, together with equation (15), bootstrap consistency of the sample mean, lemma 3 and the condition that $\lambda_n(\tau) \to \infty$, implies that $|T_n^*(\tau)|/\lambda_n(\tau) = o_{P^M}(1)$ conditionally in probability.

When $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$, it is easy to verify that $|\theta_n(\tau)| \to |\theta_{0,k_0}(\tau)| > 0$ under the condition that k_0 is unique, where $(\alpha_{0,k}(\tau), \theta_{0,k}(\tau)) = \arg\min_{\alpha,\theta} E[\rho_\tau\{\epsilon(\tau) + \alpha_0(\tau) + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau) - \alpha - \theta X_k\}]$. Thus

$$\begin{aligned} P^M\{|T_n^*(\tau)| \leqslant \lambda_n(\tau)\} &= P^M\{n^{1/2}|\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau) + \hat{\theta}_n(\tau) - \theta_n(\tau) + \theta_n(\tau) - \theta_{0,k_0(\tau)}(\tau) + \theta_{0,k_0(\tau}(\tau)| \\ &\quad \leqslant \lambda_n(\tau)\hat{\sigma}_{n*}(\tau)\} \\ &\quad \leqslant P^M\{|\theta_{0,k_0(\tau)}(\tau)| \leqslant n^{-1/2}\lambda_n(\tau) \max_{k=1,\ldots,p} \hat{\sigma}_k(\tau) + |\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau)| \\ &\qquad + |\hat{\theta}_n(\tau) - \theta_n(\tau)| + |\theta_n(\tau) - \theta_{0,k_0(\tau)}(\tau)|\} \end{aligned}$$

tends to 0 in probability, where the convergence follows from lemma 1, lemma 5 below, and $\lambda_n(\tau) = o(\sqrt{n})$.

Therefore, for all $\tau \in \mathcal{T}$, we have that

$$\begin{aligned} E^M|I\{|T_n^*(\tau)| \leqslant \lambda_n(\tau)\} - I\{\boldsymbol{\beta}_0(\tau) = \mathbf{0}\}| &= E^M|I\{|T_n^*(\tau)| > \lambda_n(\tau)\} - I\{\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}\}| \\ &= P^M\{|T_n^*(\tau)| > \lambda_n(\tau), \boldsymbol{\beta}_0(\tau) = \mathbf{0}(\tau)\} \\ &\quad + P^M\{|T_n^*(\tau)| \leqslant \lambda_n(\tau), \boldsymbol{\beta}_0(\tau) \neq \mathbf{0}\} \\ &= P^M\{|T_n^*(\tau)| > \lambda_n(\tau)|\boldsymbol{\beta}_0(\tau) = \mathbf{0}\}I\{\boldsymbol{\beta}_0(\tau) = \mathbf{0}\} \\ &\quad + P^M\{|T_n^*(\tau)| \leqslant \lambda_n(\tau)|\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}\}I\{\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}\} \end{aligned}$$

tends to 0 in probability. This implies that $I\{|T_n^*(\tau)| > \lambda_n(\tau)\} \to^{P^M} I\{\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}\}$ and $I\{|T_n^*(\tau)| \leqslant \lambda_n(\tau)\} \to^{P^M} I\{\boldsymbol{\beta}_0(\tau) = \mathbf{0}\}$ conditionally in probability. By lemmas 1 and 3, it is easy to verify that $I\{|T_n(\tau)| \leqslant \lambda_n(\tau)\} \to^{\mathrm{P}} I\{\boldsymbol{\beta}_0(\tau) = \mathbf{0}\}$ for all $\tau \in \mathcal{T}$. The result of theorem 2 follows from Slutsky's lemma.

Lemma 5. Suppose that the assumptions in theorem 1 hold. Then $\hat{k}_n^*(\tau) \to^{P^M} k_0(\tau)$ conditionally (on the data) almost surely and

$$\frac{n^{1/2}\{\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau)\}}{\hat{\sigma}_{n*}(\tau)} \xrightarrow{\mathrm{d}} \frac{M_{p+k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\pi_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\} - M_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\mu_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}}{V_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\sigma_{k_0(\tau)}(\tau)}$$

for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$, conditionally (on the data) in probability, where $M_j\{\boldsymbol{\beta}_0(\tau)\} = M_j\{\tau, \boldsymbol{\beta}_0(\tau)\}$.

The proof is provided in the on-line supplementary material.

Lemma 6. Suppose that all the assumptions in theorem 1 hold. Then $\mathbb{V}_n^*\{\tau, \mathbf{b}_0(\tau)\}$ converges to the same limiting distribution as $\{\hat{\theta}_n(\tau) - \theta_n(\tau)\}\sqrt{n}/\hat{\sigma}_n(\tau)$ for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, conditionally (on the data) in probability.

The proof is provided in the on-line supplementary material.

References

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

Belloni, A., Chernozhukov, V. and Kato, K. (2014) Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Technical Report*. Duke University, Durham.

Belloni, A., Chernozhukov, V. and Kato, K. (2015) Uniform post selection inference for LAD regression and other Z-estimation problems. *Biometrika*, **102**, 77–94.

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.

Chatterjee, A. and Lahiri, S. N. (2015) Comment on "An adaptive resampling test for detecting the presence of significant predictors" by I. McKeague and M. Qian. *J. Am. Statist. Ass.*, **110**, 1434–1438.

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

Coffin, J. M. (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, **267**, 483–489.

Donoho, D. and Jin, J. (2004) Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, **32**, 962–994.

Donoho, D. and Jin, J. (2015) Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.*, **30**, 1–25.

Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. London: Chapman and Hall.
Fan, J., Liao, Y. and Yao, J. (2015) Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, **83**, 1497–1541.
Fan, J. and Lv, J. (2008) Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Statist. Soc.* B, **70**, 849–911.
Feng, X., He, X. and Hu, J. (2011) Wild bootstrap for quantile regression. *Biometrika*, **98**, 995–999.
Fithian, W., Taylor, J., Tibshirani, R. and Tibshirani, R. (2015) Selective sequential model selection. *Preprint arXiv:1512.02565*. University of California at Berkeley, Berkeley.
Genovese, C., Jin, J., Wasserman, L. and Yao, Z. (2012) A comparison of the lasso and marginal regression. *J. Mach. Learn. Res.*, **13**, 2107–2143.
Gregory, K. B., Carroll, R. J., Baladandayuthapani, V. and Lahiri, S. N. (2015) A two-sample test for equality of means in high dimension. *J. Am. Statist. Ass.*, **110**, 837–849.
G'Sell, M. G., Wager, S., Chouldechova, A. and Tibshirani, R. (2016) Sequential selection procedures and false discovery rate control. *J. R. Statist. Soc.* B, **78**, 423–444.
Hall, P. and Kang, K. H. (2001) Boostrapping nonparametric density estimators with empirically chosen bandwidths. *Ann. Statist.*, **29**, 1443–1468.
Hall, P. and Sheather, S. J. (1988) On the distribution of a Studentized quantile. *J. R. Statist. Soc.* B, **50**, 381–391.
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.
Hendricks, W. and Koenker, R. (1991) Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Am. Statist. Ass.*, **87**, 58–68.
Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, **6**, 65–70.
Kocherginsky, M., He, X. and Mu, Y. (2005) Practical confidence intervals for regression quantiles. *J. Computnl Graph. Statist.*, **14**, 41–55.
Koenker, R. (2005) *Quantile Regression*. Cambridge: Cambridge University Press.
Kosorok, M. K. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.
Kozbur, D. (2015) Testing-based forward model selection. *Preprint arXiv:1512.0266v2*. Eidgenössiche Technische Hochschule Zürich, Zürich.
Lee, J. D. and Taylor, J. E. (2014) Exact post model selection inference for marginal screening. *Preprint arXiv:1402.5596*. Stanford University, Stanford.
Leeb, H. and Pötscher, B. M. (2006) Can one estimate the conditional distribution of post-model-selection estimators? *Ann. Statist.*, **34**, 2554–2591.
Li, A. and Barber, R. F. (2017) Accumulation tests for FDR control in ordered hypothesis testing. *J. Am. Statist. Ass.*, **112**, 837–849.
McKeague, I. W. and Qian, M. (2015) An adaptive resampling test for detecting the presence of significant predictors. *J. Am. Statist. Ass.*, **110**, 1422–1433.
Powell, J. L. (1991) Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics* (eds W. Barnett, J. Powell and G. Tauchen). Cambridge: Cambridge University Press.
Rhee, S. Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J. and Shafer, R. W. (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.*, **31**, 298–303.
Shao, X. and Zhang, J. (2014) Martingale difference correlation and its use in high-dimensional variable screening. *J. Am. Statist. Ass.*, **109**, 1302–1318.
Tibshirani, R. J., Taylor, J., Lockhart, R. and Tibshirani, R. (2016) Exact post-selection inference for sequential regression procedures. *J. Am. Statist. Ass.*, **111**, 600–620.
Tukey, J. W. (1989) Higher criticism for individual significances in several tables or parts of tables. *Working Paper*. Princeton University, Princeton.
Zhang, X., Yao, S. and Shao, X. (2017) Conditional mean and quantile dependence testing in high dimension. *Ann. Statist.*, to be published.
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 "Testing for marginal linear effects in quantile regression"'.