I N S T I T U T   D E   S T A T I S T I Q U E
B I O S T A T I S T I Q U E   E T
S C I E N C E S   A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2011/24

ON ASSESSING MODEL ADEQUACY IN LINEAR QUANTILE REGRESSION

NOH, H., EL GHOUCH, A. and I. VAN KEILEGOM

On Assessing Model Adequacy in Linear Quantile Regression

Hohsuk Noh *

Université catholique de Louvain

Anouar El Ghouch †

Université catholique de Louvain

Ingrid Van Keilegom ‡

Université catholique de Louvain

August 26, 2011

Abstract

In this paper we consider (possibly misspecified) linear quantile regression models, and study a measure for the quality-of-fit of these models, (a version of which has been) previously proposed by Koenker and Machado (1999). The measure is based on an adaptation to quantile regression of the famous coefficient of determination originally proposed for mean regression, and compares a 'reduced' model to a 'full' model, both of which can be misspecified. We propose an estimator of this measure, and prove its asymptotic distribution both in the non-degenerate and the degenerate case. The finite sample performance of the estimator is studied through a number of simulation experiments. The proposed measure is also applied to a data set on body fat measures.

Key Words: Coefficient of determination; full model; lack-of-fit; linear regression; model misspecification; quantile regression; reduced model.

*H. Noh acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy).

†A. El Ghouch acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

‡I. Van Keilegom acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

1 Introduction

Quantile regression has emerged as an attractive alternative to the classical mean regression approach based on the quadratic loss function. Since it was introduced by Koenker and Bassett (1978) as a robust (to outliers) and flexible (to error distribution) linear regression method, quantile regression has received considerable interest in both theoretical and applied statistics (see Koenker (2005) and references therein). As this method has widened its applications to many domains like economics, biology, ecology, and finance, the development of an effective assessment measure of goodness-of-fit for quantile regression models becomes very attractive for practitioners. In the context of fixed design median regression, Mckean and Sievers (1987) proposed a least absolute coefficient of determination that aims at measuring the prediction quality of covariates in a linear model. Later Koenker and Machado (1999) extended its definition to any quantile linear regression model. The proposed measure is motivated by the familiar R^2 coefficient of determination for linear mean regression, which is given by $1 - SSE/SST$, with SSE being the usual sum of squares of residuals and SST being the total sum of squares. In a similar way, for a fixed quantile level $q \in (0, 1)$, we can define a qth quantile based coefficient of determination by using the check loss function given by $\rho_q(u) = 2u(q - I(u < 0))$ instead of the squared loss. To be more precise, let Y be a one-dimensional dependent variable and $\mathbf{X} = (\mathbf{X}_0^\top, \mathbf{X}_1^\top)^\top$ be a random covariate vector of dimension $d_0 + d_1$, with $d_0, d_1 \geq 1$. The first (and only the first) element of $\mathbf{X}_0$ is 1. Define

$$\zeta(q) = 1 - \frac{\mathrm{E}[\rho_q(Y - \mathbf{X}^\top \boldsymbol{\beta}_q^*)]}{\mathrm{E}[\rho_q(Y - \mathbf{X}_0^\top \boldsymbol{\beta}_{0,q}^*)]}, \tag{1.1}$$

where $\boldsymbol{\beta}_{0,q}^*$ and $\boldsymbol{\beta}_q^*$ are pseudo true parameters in the sense that they are assumed to be the unique minimizers of $\mathrm{E}[\rho_q(Y - \mathbf{X}_0^\top \mathbf{b}_0)]$ and $\mathrm{E}[\rho_q(Y - \mathbf{X}^\top \mathbf{b})]$ with respect to $\mathbf{b}_0$ and $\mathbf{b}$, respectively. Equivalently, we can say that they are the best approximations to the true regression function that can be found within the two given families of linear models. None of the two linear models is supposed to be correct, they are both (possibly) subject to model misspecification.

In terms of the check loss distance, $\mathrm{E}[\rho_q(Y - \mathbf{X}^\top \boldsymbol{\beta}_q^*)]$ measures the amount of variation of Y that cannot be explained through a 'full' but possibly incorrect linear model in $\mathbf{X}$, and $\mathrm{E}[\rho_q(Y - \mathbf{X}_0^\top \boldsymbol{\beta}_{0,q}^*)]$ is the variation of Y that cannot be explained through the reduced linear model. So, $\zeta(q)$ is nothing but the missed fraction of variation in terms of the check loss when one uses the reduced qth quantile linear model instead of the full one. If $d_0 = 1$, then $\boldsymbol{\beta}_{0,q}^*$ becomes $\xi_q = \arg\min_b \mathrm{E}[\rho_q(Y - b)]$, which is the marginal qth quantile of Y, and then $\zeta(q)$ coincides with

$$R(q) = 1 - \frac{\mathrm{E}[\rho_q(Y - \mathbf{X}^\top \boldsymbol{\beta}_q^*)]}{\mathrm{E}[\rho_q(Y - \xi_q)]}. \tag{1.2}$$

This is the quantile analogue of the well known Pearson's correlation ratio $\eta^2 = 1 - \mathrm{E}(Y - \mathbf{X}^\top \boldsymbol{\beta}^*)^2 / \mathrm{E}(Y - \mathrm{E}(Y))^2$, i.e. the 'theoretical' R^2 for the linear mean regression model $\mathrm{E}(Y|\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}^*$.

Hereafter, for simplicity and when no confusion is possible, we will suppress the subscript q in all our notations, so we will write ρ, $\boldsymbol{\beta}^*$, $\boldsymbol{\beta}_0^*$ and ξ instead of ρ_q, $\boldsymbol{\beta}_q^*$, $\boldsymbol{\beta}_{0,q}^*$ and ξ_q. From now on, we will use the notation $\zeta(q)$ only when $d_0 > 1$. Like η^2, $R(q)$ lies in $[0, 1]$. $R(q) = 0$ corresponds to the case when $\mathbf{X}^\top \boldsymbol{\beta}^* = \xi$ with probability one, i.e. all components of $\boldsymbol{\beta}^*$ vanish except the first one, which coincides with ξ. In that case, no variability is captured by $\mathbf{X}$ via a linear qth quantile model. As for $\zeta(q)$, we observe that $0 \leq \zeta(q) \leq R(q)$. $\zeta(q) = 0$ is equivalent to saying that $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_0^*, \mathbf{0})$, i.e. no information is lost when considering only the restricted linear model, while, when $\mathrm{E}[\rho(Y - \mathbf{X}^\top \boldsymbol{\beta}^*)] > 0$, the equality $\zeta(q) = R(q)$ occurs if and only if $\mathbf{X}_0^\top \boldsymbol{\beta}_0^* = \xi$ with probability 1. This is the case when the reduced linear approximation fails to capture any variability in Y that could be captured if we use the 'full' linear model. To summarize, $R(q)$ quantifies the total gain of considering a linear model of $\mathbf{X}$ to fit the quantile of Y; the bigger the value of $R(q)$, the better is the qth quantile linear fit. In contrast to $R(q)$, $\zeta(q)$ quantifies the relative loss in the explained variation that can be attributed to the lack-of-fit of the reduced qth quantile linear model compared to the full one; the smaller the value of $\zeta(q)$, the better is the restricted qth quantile linear fit. This quantity can be used to compare two nested linear quantile models whether the corresponding true conditional quantile functions are linear or not. Unlike R^2 or any other measure based on the L^2 loss function, by varying q we get a more complete picture of the quality of different linear approximations both in the center and the tails. Also, when the underlying distribution is asymmetric or in the presence of outliers, $R(0.5)$ should be used as a robust alternative to R^2. Similarly, $\zeta(0.5)$ can be used to robustly reduce the dimensionality of $\mathbf{X}$ by keeping the significant components.

In this work our objective is to estimate and make inference about $\zeta(q)$ and $R(q)$ in the random design setting. In the literature, to the best of our knowledge, only the problem of estimating consistently $R(0.5)$ has been studied for fixed design median linear regression; see Mckean and Sievers (1987). In Section 2, we present asymptotic results about the proposed measures. In Section 3, we describe statistical inference based on them. We show an additional asymptotic result when $\zeta(q) = 0$ in Section 4. In Section 5, we present some Monte Carlo evidence of the developed theory, whereas the analysis of data on body fat measures is given in Section 6. All the theoretical proofs are deferred to the Appendix.

2 Main results

Consider i.i.d. random vectors $(Y_i, \mathbf{X}_i^\top)$, $i = 1, \dots, n$, where Y_i is a random scalar and $\mathbf{X}_i = (\mathbf{X}_{0,i}^\top, \mathbf{X}_{1,i}^\top)^\top$ is a random vector in $\mathbb{R}^d$, $d = d_0 + d_1$. The first element in $\mathbf{X}_{0,i}$ is one for all $i = 1, \dots, n$ and $(Y_i, \mathbf{X}_i^\top)$ is continuously distributed. Clearly, to consistently estimate $\zeta(q)$, see (1.1), we need consistent estimators of $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}_0^*$ under the misspecified setting. For that and in order to obtain an asymptotic linear representation of our estimator, see Theorem 2.2 below, we need to introduce some notations and make some assumptions.

Define

$$\hat{\boldsymbol{\beta}}_q \equiv \hat{\boldsymbol{\beta}} = \arg\min_{\mathbf{b} \in \mathcal{B}} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \mathbf{b}),$$

and $\varepsilon_q^* \equiv \varepsilon^* = Y - \mathbf{X}^\top \boldsymbol{\beta}^*$. Similarly, we define $\hat{\boldsymbol{\beta}}_{0,q} \equiv \hat{\boldsymbol{\beta}}_0$ and $\varepsilon_{0,q}^* \equiv \varepsilon_0^*$. Note that $\varepsilon^* = \varepsilon_0^*$ when $\zeta(q) = 0$.

Assumption A1: $\boldsymbol{\beta}^*$ ($\boldsymbol{\beta}_0^*$) is an interior point of $\mathcal{B}$ ($\mathcal{B}_0$), a compact subset of $\mathbb{R}^d$ ($\mathbb{R}^{d_0}$).

Assumption A2: $(\mathbf{X}, \varepsilon^*)$ satisfies the following:

A2.1 For all $\mathbf{x}$, $f_{\varepsilon^*|\mathbf{X}}(0|\mathbf{x}) > 0$, where $f_{\varepsilon^*|\mathbf{X}}(t|\mathbf{x})$ is the conditional density of ε^* given $\mathbf{X} = \mathbf{x}$.

A2.2 $f_{\varepsilon^*|\mathbf{X}}(t|\mathbf{x})$ is Lipschitz continuous, *i.e.* for all $\mathbf{x}$, $|f_{\varepsilon^*|\mathbf{X}}(t_1|\mathbf{x}) - f_{\varepsilon^*|\mathbf{X}}(t_2|\mathbf{x})| \le L|t_1 - t_2|$ for some $0 < L < \infty$, and there exists a constant $M > 0$ such that $f_{\varepsilon^*|\mathbf{X}}(t|\mathbf{x}) \le M$ for all t, x.

A2.3 $Q := \mathrm{E}[2 f_{\varepsilon^*|\mathbf{X}}(0|\mathbf{X}) \mathbf{X}\mathbf{X}^\top]$ and $V := \mathrm{E}[\varphi(\varepsilon^*)^2 \mathbf{X}\mathbf{X}^\top]$ are positive definite, where $\varphi_q(u) \equiv \varphi(u) = 2q - 2I(u < 0)$.

A2.4 The marginal density $f_{\varepsilon^*}(t)$ of ε^* is bounded in some neighborhood $\mathcal{E}$ of 0.

Assumption A3: $(\mathbf{X}_0, \varepsilon_0^*)$ satisfies Assumption A2, with $\mathbf{X}$ and ε^* replaced throughout by $\mathbf{X}_0$ and ε_0^*.

These conditions are needed to get Bahadur-type representations of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_0$ as given by the following lemma, whose proof can be found in Kim and White (2003). We use throughout the notation $\|\mathbf{X}\| = (\mathbf{X}^\top \mathbf{X})^{1/2}$ for the Euclidean norm of $\mathbf{X}$.

Lemma 2.1 *Suppose that A1 and A2 hold. If* $\mathrm{E}(\|\mathbf{X}\|^3) < \infty$, *then,*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* = \frac{1}{n} Q^{-1} \sum_{i=1}^{n} \mathbf{X}_i \varphi(\varepsilon_i^*) + o_p(n^{-1/2}).$$

Suppose that A1 and A3 hold. If $\mathrm{E}(\|\mathbf{X}_0\|^3) < \infty$, *then,*

$$\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^* = \frac{1}{n} Q_0{}^{-1} \sum_{i=1}^{n} \mathbf{X}_{0,i} \varphi(\varepsilon_{0,i}^*) + o_p(n^{-1/2}),$$

where $Q_0 = \mathrm{E}[2 f_{\varepsilon_0^* | \mathbf{X}_0}(0|\mathbf{X}_0)\mathbf{X}_0\mathbf{X}_0^\top]$.

Now that we have introduced all the necessary ingredients we move on to our main parameter of interest. An obvious estimator of $\zeta(q)$, is given by

$$\hat{\zeta}(q) = 1 - \frac{\sum_{i=1}^{n} \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}})}{\sum_{i=1}^{n} \rho(Y_i - \mathbf{X}_{0,i}^\top \hat{\boldsymbol{\beta}}_0)}. \tag{2.1}$$

The following theorem gives the asymptotic linear representation of our estimator.

Theorem 2.2 *Suppose that A1-A3 hold, and that there exists a* $M > 0$ *such that* $P(\|\mathbf{X}\| \leq M) = 1$. *Then,*

$$\sqrt{n}(\hat{\zeta}(q) - \zeta(q)) = \frac{1}{\sqrt{n}}(1 - \zeta(q)) \sum_{i=1}^{n} (e_i - u_i) + o_p(1),$$

where

$$e_i = \frac{\rho(\varepsilon_{0,i}^*) - \mathrm{E}\rho(\varepsilon_0^*)}{\mathrm{E}\rho(\varepsilon_0^*)} \quad \textit{and} \quad u_i = \frac{\rho(\varepsilon_i^*) - \mathrm{E}\rho(\varepsilon^*)}{\mathrm{E}\rho(\varepsilon^*)}.$$

The proof of this theorem is given in the Appendix. Theorem 2.2 implies the consistency of $\hat{\zeta}(q)$. It also implies that $n^{1/2}(\hat{\zeta}(q) - \zeta(q))$ is asymptotically normal with mean zero and variance $\sigma_q^2 \equiv \sigma^2 = (1 - \zeta(q))^2 \mathrm{Var}(e - u)$.

The boundedness of $\mathbf{X}$ assumed in Theorem 2.2 is stronger than the condition $\mathrm{E}(\|\mathbf{X}\|^3) < \infty$ in Kim and White (2003) and is used here to simplify the derivation of the stochastic expansion of both the denominator and numerator while avoiding the well known problems in the tails. We can relax this assumption by incorporating into the estimator a smooth weight function $w(\cdot)$, which reduces the inherent errors when the density $f_{\mathbf{X}}(\cdot)$ approaches zero. In fact, when $\mathbf{X}$ is not bounded, as a parameter of interest, it is more appropriate to consider

$$\zeta_w(q) = 1 - \mathrm{E}\left[\rho(Y - \mathbf{X}^\top \boldsymbol{\beta}_w^*) w(\mathbf{X})\right] \Big/ \mathrm{E}\left[\rho(Y - \mathbf{X}_0^\top \boldsymbol{\beta}_{w,0}^*) w(\mathbf{X})\right],$$

where $\boldsymbol{\beta}_w^* = \arg\min_{\mathbf{b} \in \mathcal{B}} \mathrm{E}\left[\rho(Y - \mathbf{X}^\top \mathbf{b}) w(\mathbf{X})\right]$ and $\boldsymbol{\beta}_{w,0}^* = \arg\min_{\mathbf{b}_0 \in \mathcal{B}_0} \mathrm{E}\left[\rho(Y - \mathbf{X}_0^\top \mathbf{b}_0) w(\mathbf{X})\right]$. Here, $w(\cdot)$ is a known bounded function, which has a compact support $\mathcal{D}$ contained in the support of the density of $\mathbf{X}$ so that only the set $\{x : w(x) \neq 0\}$ matters. For example, one may consider $w(\mathbf{x}) = I\{\|\mathbf{x}\| \leq M\}$, where M is any sufficiently large number. An estimator of $\zeta_w(q)$ can be defined as follows:

$$\hat{\zeta}_w(q) = 1 - \frac{\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}_w) w(\mathbf{X}_i)}{\sum_{i=1}^n \rho(Y_i - \mathbf{X}_{0,i}^\top \hat{\boldsymbol{\beta}}_{w,0}) w(\mathbf{X}_i)},$$

where $\hat{\boldsymbol{\beta}}_w = \arg\min_{\mathbf{b}\in\mathcal{B}} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \mathbf{b}) w(\mathbf{X}_i)$ and $\hat{\boldsymbol{\beta}}_{w,0} = \arg\min_{\mathbf{b}_0\in\mathcal{B}_0} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_{0,i}^\top \mathbf{b}_0) w(\mathbf{X}_i)$. $\zeta_w(q)$ has the same meaning as its unweighted version $\zeta(q)$ and using technical arguments very similar to those given in the Appendix, it can be shown that $\hat{\zeta}_w(q)$ has also the same asymptotic representation as $\hat{\zeta}(q)$, see Theorem 2.2, with $\varepsilon_w^* := (Y - \mathbf{X}^\top \boldsymbol{\beta}_w^*) w(\mathbf{X})$ and $\varepsilon_{w,0}^* := (Y - \mathbf{X}_0^\top \boldsymbol{\beta}_{w,0}^*) w(\mathbf{X})$, instead of ε^* and ε_0^*, respectively.

It is also clear that a similar approach can be used to estimate and make inference about $R(q)$; see (1.2). The following asymptotic result for $\hat{R}(q)$ is a simple corollary of Theorem 2.2, so its proof is omitted.

Theorem 2.3 *Suppose that A1-A2 hold. Suppose also that (i) there exists a* $M > 0$ *such that* $P(\|\mathbf{X}\| \leq M) = 1$*; and (ii) the distribution function of* Y*,* $F_Y(\cdot)$*, has bounded second derivative in a neighborhood of* ξ*, and* $f_Y(\xi) > 0$*, where* $f_Y(\cdot)$ *is the marginal density function of* Y*. Then,*

$$\sqrt{n}(\hat{R}(q) - R(q)) = \frac{1}{\sqrt{n}}(1 - R(q)) \sum_{i=1}^n (e_i - u_i) + o_p(1),$$

where

$$e_i = \frac{\rho(Y_i - \xi) - \mathrm{E}\rho(Y - \xi)}{\mathrm{E}\rho(Y - \xi)} \quad \text{and} \quad u_i = \frac{\rho(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}^*) - \mathrm{E}\rho(Y - \mathbf{X}^\top \boldsymbol{\beta}^*)}{\mathrm{E}\rho(Y - \mathbf{X}^\top \boldsymbol{\beta}^*)}.$$

3 Inference for $\zeta(q)$

To use the asymptotic normality results shown in the previous section, one needs a consistent estimator for the asymptotic variance $\sigma^2 = (1 - \zeta(q))^2 \mathrm{Var}(e - u)$. When $\zeta(q) = 0$ or 1, this variance vanishes and our result becomes $\hat{\zeta}(q) = \zeta(q) + o_p(n^{-1/2})$. Since $\zeta(q) = 0$ corresponds to the interesting case where the covariate $\mathbf{X}_1$ can be removed from the linear model with no loss of information, such a case will be investigated in more detail in Section 4. For now we consider the case when $\zeta(q) \in (0, 1)$. Put $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}$ and $\hat{\varepsilon}_{0,i} = Y_i - \mathbf{X}_{0,i}^\top \hat{\boldsymbol{\beta}}_0$, and let

$$\hat{e}_i = \frac{\rho(\hat{\varepsilon}_{0,i}) - n^{-1}\sum_i \rho(\hat{\varepsilon}_{0,i})}{n^{-1}\sum_i \rho(\hat{\varepsilon}_{0,i})} \quad \text{and} \quad \hat{u}_i = \frac{\rho(\hat{\varepsilon}_i) - n^{-1}\sum_i \rho(\hat{\varepsilon}_i)}{n^{-1}\sum_i \rho(\hat{\varepsilon}_i)}.$$

One can easily check that $\hat{e}_i = e_i + o_p(1)$ and $\hat{u}_i = u_i + o_p(1)$. So, as an estimator of σ^2, we propose $\hat{\sigma}^2 = \frac{1}{n}\sum_i (\hat{e}_i - \hat{u}_i - \overline{\hat{e} - \hat{u}})^2$, where $\overline{\hat{e} - \hat{u}} = n^{-1}\sum_i (\hat{e}_i - \hat{u}_i)$. An asymptotically valid confidence interval of $\zeta(q)$ of level $1-\alpha$ is given by $\hat{\zeta}(q) \pm z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}$, where z_α is the upper α quantile of the standard normal distribution.

Although this confidence interval gives us valuable information about $\zeta(q)$, it cannot be used to check the hypothesis

$$H_0 : \zeta(q) = 0 \quad \text{versus} \quad H_1 : \zeta(q) > 0. \tag{3.1}$$

The reason is that one needs the distribution of $\hat{\zeta}(q)$ under the null hypothesis. This will be discussed in Section 4. Instead, we can consider the following hypothesis

$$H_{0,\pi} : \zeta(q) \geq \pi \quad \text{versus} \quad H_{1,\pi} : \zeta(q) < \pi, \tag{3.2}$$

where $\pi \in (0, 1)$ is a small constant that can be considered by the analyst as a tolerable missed fraction of variation. In the literature, (3.2) is known as a neighborhood hypothesis or 'precise' hypothesis; see Hodges and Lehmann (1954). Note that (3.2) is designed to provide evidence in favor of the reduced model while it cannot be confirmed in the testing framework of (3.1) even if the p-value associated with (3.1) is large. For a detailed discussion about many aspects of neighborhood hypothesis testing we refer to Dette and Munk (2003). Testing (3.2) can be done directly using Theorem 3.2. In fact, since the limit of the upper confidence interval of $\zeta(q)$ is given by $\hat{\zeta}^+(q) = \hat{\zeta}(q) + z_\alpha\frac{\hat{\sigma}}{\sqrt{n}}$ and since, under $H_{0,\pi}$, $P(\zeta(q) \leq \hat{\zeta}^+(q)) \to 1-\alpha$, one should accept $H_{1,\pi}$ and so validate the reduced model whenever $\hat{\zeta}^+(q) < \pi$.

4 The case when $\zeta(q) = 0$

In this section we investigate the behavior of $\hat{\zeta}(q)$ when $\zeta(q) = 0$. To simplify the analysis we consider the homogeneous linear model given by

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon_i = \mathbf{X}_{0,i}^\top \boldsymbol{\beta}_0 + \mathbf{X}_{1,i}^\top \boldsymbol{\beta}_1 + \varepsilon_i, \tag{4.1}$$

where the errors ε_i are supposed to be independent of $\mathbf{X}_i = (\mathbf{X}_{0,i}^\top, \mathbf{X}_{1,i}^\top)^\top$ and the qth quantile of ε_i is 0. Note that (4.1) is equivalent to saying that $\mathbf{X}_i^\top \boldsymbol{\beta}$ is the true conditional quantile function of Y_i given $\mathbf{X}_i$, and $\boldsymbol{\beta}$ is now the true parameter that we assume to be an interior point of $\mathcal{B}$, a compact subset of $\mathbb{R}^d$, $d = d_0 + d_1$. These assumptions facilitate the technical development related to an uniform quadratic approximation, see Lemma 4.1 below, needed to derive the asymptotic distribution. Under model (4.1), $\zeta(q) = 0$ is equivalent to $\boldsymbol{\beta}_1 = \mathbf{0}$. In such a case $n^{-1}\sum_{i=1}^n \rho(Y_i - \mathbf{X}_{0,i}^\top \hat{\boldsymbol{\beta}}_0)$ and $n^{-1}\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}})$

have the same first order linear approximation, namely $n^{-1}\sum_{i=1}^n \rho(Y_i - \mathbf{X}_{0,i}^\top \boldsymbol{\beta}_0) + o_p(n^{-1/2})$. So to obtain the asymptotic distribution of $\hat{\zeta}(q)$ we need to inspect higher order terms in the asymptotic expansions. To this end, we introduce the following lemma (where f_ε denotes the density of ε).

Lemma 4.1 *Let $\mathbf{X}_{in} = \mathbf{X}_i/\sqrt{n}$. If A2 holds with ε instead of ε^* and if $f_\varepsilon(0) > 0$ and there exists a $M > 0$ such that $P(\|\mathbf{X}\| \le M) = 1$, then for any constant $C < \infty$,*

$$\sup_{\|\delta\| \le C} \left| \sum_{i=1}^n \Big(\rho(\varepsilon_i - \mathbf{X}_{in}^\top \delta) - \rho(\varepsilon_i) + \mathbf{X}_{in}^\top \delta \varphi(\varepsilon_i) \Big) - f_\varepsilon(0)\delta^\top \mathrm{E}(\mathbf{X}\mathbf{X}^\top)\delta \right| \to 0 \text{ in probability.}$$

as $n \to \infty$.

This is an adaptation of Lemma 2.2 in Rao and Zhao (1992) to the random design setting. Thanks to this lemma, we obtain the asymptotic distribution of $\hat{\zeta}(q)$ in the degenerate case. The proofs of both results can be found in the Appendix.

Theorem 4.2 *Under the conditions of Lemma 4.1, if $\zeta(q) = 0$, then,*

$$n\hat{\zeta}(q) \xrightarrow{d} \frac{q(1-q)}{f_\varepsilon(0)\lambda}\chi^2_{d_1},$$

where $\lambda = \mathrm{E}\rho(\varepsilon)$ and $\varepsilon = Y - \mathbf{X}^\top \boldsymbol{\beta} = Y - \mathbf{X}_0^\top \boldsymbol{\beta}_0$.

In other words $\hat{\zeta}(q)$ follows a scaled chi-square distribution with d_1 degrees of freedom. The scaling factor $q(1-q)/(f_\varepsilon(0)\lambda)$ can be estimated by plugging-in appropriate estimators of $f_\varepsilon(0)$ and λ; see for example Li and Zou (1994). Clearly one can use this result to construct an asymptotically valid test statistic for (3.1).

5 Numerical Studies

In this section we will present the results of four Monte Carlo experiments, to check whether the asymptotic theory developed in the previous sections provides good approximations for small samples. In all four examples, we use a truncated normal distribution $TN(\boldsymbol{\mu}, \Sigma, \mathbf{a}, \mathbf{b})$ for the distribution of the covariate vector, i.e. its density is given by

$$f(\mathbf{x}, \boldsymbol{\mu}, \Sigma, \mathbf{a}, \mathbf{b}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}}{\int_{\mathbf{a}}^{\mathbf{b}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} d\mathbf{x}}.$$

Example 1. Consider the model

$$Y_i = 0.5 + X_{1i} - 2X_{2i} + 3X_{3i} + \sigma\varepsilon_i, \quad i = 1, \dots, n, \tag{5.1}$$

where the $\mathbf{X}_i = (X_{1i}, X_{2i}, X_{3i})^\top$'s are i.i.d. and generated from a truncated multivariate normal distribution with

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}, \mathbf{a} = \begin{pmatrix} -5 \\ -5 \\ -5 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix},$$

the ε_i's are i.i.d. and generated from a standard normal, and ε_i and $\mathbf{X}_i$ are independent. The value of σ is given by 1, 3 or 5. Figure 1 shows the boxplot and quantile-quantile (Q-Q) plot of $\hat{R}(0.5)$ (with reference to the normal distribution) based on 1000 random samples of size 200 when $\sigma = 1$. The Q-Q plot in Figure 1 indicates that $\hat{R}(0.5)$ asymptotically follows a normal distribution. Moreover, the p-value of the Kolmogorov-Smirnov (KS) test for the distribution of $\hat{R}(0.5)$ is 0.193, which supports our asymptotic theory. Though not reported here, for other values of q we get similar findings. To validate our asymptotic theory in another point of view, we calculate the empirical coverage probabilities of confidence intervals for $R(0.5)$ and the results for different values of σ are reported in Table 1. The significance level (α) is 0.05. To construct confidence intervals, the asymptotic variance estimator given in Section 3 is used. From Table 1, we can clearly see that the empirical coverage probabilities are getting close to the nominal value as the sample size increases regardless of the value of σ.

Example 2. Consider the model

$$Y_i = \left(\frac{1}{2} + X_i + \frac{1}{4}X_i^2\right)\varepsilon_i, \quad i = 1, \dots, n, \tag{5.2}$$

where the X_i's are i.i.d. and have a truncated normal distribution with $\mu = 3$, $\sigma = 1$, $a = -2$ and $b = 8$, and the ε_i's are independent zero-mean normal random variables with variance 1/100. We assume again that ε_i and X_i are independent. Letting $Q_Y(q|X)$ denote the qth quantile of the distribution of Y given X, (5.2) may be expressed in the following way:

$$Q_Y(q|X) = \left(\frac{1}{2} + X + \frac{1}{4}X^2\right)\Phi^{-1}(q),$$

where $0 < q < 1$ and $\Phi^{-1}(q)$ is the standard Gaussian quantile function. Hence, by taking $\mathbf{X}_i = (X_i, X_i^2)^\top$, (5.2) can be regarded as a linear quantile regression model. Note that in this example the covariate vector $\mathbf{X}$ cannot account for any portion of the variation of Y in the mean regression framework, as the population version of R^2 equals zero. But for quantiles other than the median there is a clear benefit from the parametric form of the conditional quantile specification as can be seen in the $R(q)$-function in Figure 2. Only when $q = 0.5$, as in the case of mean regression, there is no benefit to consider $\mathbf{X}$ in the estimation of the conditional median of Y. As for the conditional median of Y, $R(0.5) = 0$ and $\hat{R}(0.5)$ is not asymptotically normal because its asymptotic variance is zero. Figure 3 shows the Q-Q plots and p-values of the KS test for the distribution of $\hat{R}(0.5)$ (upper panels, degenerate case) and $\hat{R}(0.2)$ (lower panels, non-degenerate case) with reference to the normal distribution based on 1000 random samples of different sample sizes ($n = 100, 500$ and 2000). From Figure 3 we observe that the distribution of $\hat{R}(0.2)$ converges to a normal distribution as the sample size increases, while that of $\hat{R}(0.5)$ stays severely skewed to the right regardless of how large the sample is. This observation coincides with our asymptotic theory.

Example 3. Consider the model

$$Y_i = 0.5 + X_{1i} - 2X_{2i} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{5.3}$$

where $\mathbf{X}_i = (X_{1i}, X_{2i}, X_{3i})$ and ε_i are the same as in Example 1, and where ε_i is independent of $\mathbf{X}_i$. Recall that when ε_i is standard normal,

$$f_\epsilon(0) = \frac{1}{\sqrt{2\pi}} \text{ and } \lambda = \mathrm{E}\rho(\varepsilon) = \frac{\sqrt{2}}{\sqrt{\pi}}.$$

Since $\beta_3 = 0$, we know from Theorem 4.2 that $n\hat{\zeta}(0.5)$ follows a scaled chi-square distribution $(\pi/4)\chi_1^2$. Figure 4 shows the Q-Q plot of $n\hat{\zeta}(0.5)$ with respect to $(\pi/4)\chi_1^2$. The $\hat{\zeta}(0.5)$'s are calculated using 500 data sets of size $n = 500$ and $n = 1000$. The quantile range is from 0.01 to 0.99 by steps of 0.01. From Figure 4 we see that the two distributions almost coincide when $n = 1000$ and differ only slightly for high quantiles when $n = 500$.

Example 4. Consider the true model

$$Y_i = 0.5 + X_{1i} - 2X_{2i} + 3X_{3i} + \nu(1 - \cos(\pi X_{1i}/2)) + \sigma\varepsilon_i, \quad i = 1, \ldots, n, \tag{5.4}$$

where the $\mathbf{X}_i$'s are i.i.d. and generated from a truncated multivariate normal distribution with

$$\boldsymbol{\mu} = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}, \mathbf{a} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

and the ε_i's are independent standard normal variables, that are independent of the $\mathbf{X}_i$'s. The value of σ is given by 0.4 or 0.8 and the value of ν is given by 0 or 2. Suppose that we only consider the linear quantile regression model for the analysis. When $\nu = 0$, there is no misspecification in the full model. In order to measure each covariate's importance in the linear model, we define

$$\hat{\zeta}_{X_1}(q) = 1 - \frac{\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}})}{\sum_{i=1}^n \rho(Y_i - \mathbf{X}_{0,i}^\top \hat{\boldsymbol{\beta}}_0)},$$

where $\mathbf{X}_{0,i} = (1, X_{2i}, X_{3i})$ and $\hat{\boldsymbol{\beta}}_0 = \arg\min_{\mathbf{b}_0} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_{0,i}^\top \mathbf{b}_0)$, and we define $\hat{\zeta}_{X_2}(q)$, $\hat{\zeta}_{X_3}(q)$, $\zeta_{X_1}(q)$, $\zeta_{X_2}(q)$ and $\zeta_{X_3}(q)$ in a similar manner. If $\zeta_{X_k}(q)$ is small, it means the the introduction of X_k into the model leads to little additional increase in explanatory power under the linearity restriction. Therefore, $\zeta_{X_k}(q)$ can be seen as the importance of X_k under the restriction of the linear model. Table 2 shows the values of $\zeta_{X_1}(q), \zeta_{X_2}(q)$ and $\zeta_{X_3}(q)$ for various settings. We observe that when no misspecification is present ($\nu = 0$), the relative importance of each covariate based on $\zeta_{X_k}(q)$ depends on the magnitude of its linear coefficient in the full model. But when the functional form of the covariate X_1 is misspecified ($\nu > 0$), the situation is different. Some portion of the nonlinear structure of X_1 can be explained by the linear component, but the amount depends on the structure of the nonlinear part. In our example, the nonlinear component is strictly increasing, and therefore that portion is quite large. Consequently, the relative importance of the covariates is not given by the magnitude of the linear coefficients, as is shown in Table 2. When $\nu = 3$, the proposed measure shows that X_1 becomes the most important covariate under the restriction of the linear model. Figure 5 shows the boxplots of $\hat{\zeta}_{X_k}(0.5)$, $k = 1, 2, 3$, when $\nu = 0$ (without misspecification) and $\nu = 3$ (with misspecification) based on 500 datasets of size 200. Clearly, we see that the values of the $\hat{\zeta}_{X_k}(q)$'s confirm our previous findings and intuition based on the $\hat{\zeta}_{X_k}(q)$'s.

6 Real Data Analysis

In this section, we illustrate the developed theory by using data on body fat measures (Penrose et al., 1985). The data set consists of the body fat percentage and various body circumference measurements for 252 men who are above age 20. We exclude one observation that appears to deviate remarkably from the other observations. Note that the body fat percentage can be determined by measuring the body average density, which requires the difference between the body weight measured in the air and during water submersion. Since this process can be cumbersome, it is sometimes convenient to build an equation to predict the percentage of body fat from body circumference measurements. This is also useful for individual health management.

For men two well known predictive factors (explanatory variables) for body fat percentage are abdominal circumference and body mass index. The body mass index is an individual's body weight (in kilogram) divided by the square of his or her height (in meter). Figure 6 shows scatter plots of abdomen circumference and body mass index versus body fat percentage, along with three conditional quantile lines ($q = 0.2, 0.5$ and 0.8). As is shown in Figure 6, there is a clear linear relationship between each of the predictive factors and body fat percentage. But we observe that for similar values of a predictive factor the body fat percentages can be fairly different. This indicates that if we want to manage body fat accurately based on either of the two predictive factors, we should have interest not only in the conditional mean of body fat percentage but also in the (high) conditional quantiles of it. This observation leads us to compare the goodness-of-fit of these two factors for predicting body fat in quantile regression. Table 3 shows the 95%-confidence intervals for $R(q)$ for each of the two factors and for different values of q. Across all the quantile levels considered, the goodness-of-fit of abdomen size is higher than that of body mass index. It implies that abdomen size is more helpful to predict the conditional quantile of body fat percentage as well as the mean conditional body fat. Recently, Coutinho et al. (2011) claimed that abdomen size is a more informative factor than body mass index for measuring the risk of coronary artery disease of an individual. Since the accumulation of excessive body fat is one of the crucial factors that cause coronary artery diseases, their recent report can be understood within the context of the results of our analysis. But from the fact that two confidence intervals of the goodness-of-fit measure overlap more for high conditional quantiles ($q = 0.8$ and 0.9), we can conclude that the difference between the goodness-of-fit of the two factors becomes a little bit less significant for high conditional quantiles than for the conditional median, if we take sampling variation into consideration. Our asymptotic theory enables us to consider sampling variation for comparison of goodness-of-fit and so helps us to perform more delicate analysis.

In this data set on body fat measures, we have possibly many other covariates which might be helpful for prediction of the quantile of body fat. But most of them have negligible effects. Recently, Capizzi et al. (2011) claimed that wrist circumference may identify children with overweight at risk for heart disease, the cause of which is mainly excess body fat. So we try to check whether wrist circumference is a factor worthy of consideration together with abdominal circumference and compare two nested linear models. One of them is the linear model containing only abdominal circumference, and the other linear model contains both covariates. In this case, the value of $\hat{\zeta}(q)$ across all q-levels is around 0.06, which suggests that wrist circumference does not need to be added to the model. Since Capizzi et al. (2011) focused in their study on children with overweight, their conclusions are not in disagreement with ours, since in our data set all men are above age 20. Additionally, we verify the explanatory power of wrist circumference. Table 4 shows the limit of the upper confidence interval for $R(q)$ at significance level $\alpha = 0.05$ for various values of q. Since all values are below 0.11, we see that wrist circumference may (in the best case) have a small effect on the quantile of body fat, but it is certainly less predictive than abdomen size or body mass index.

7 Conclusion

In this paper we have considered the problem of measuring the adequacy of a linear fit in quantile regression, when misspecification (both in the full and the reduced model) is possibly present. We have studied two coefficients, $\zeta(q)$ and $R(q)$, (a version of) which have been previously proposed by Koenker and Machado (1999), and which are both inspired by the so-called coefficient of determination, introduced long ago in the context of mean regression. The proposed estimators of both quantities have been studied from a theoretical point of view, as well as through simulations and via the analysis of data on body fat indicators.

It would be interesting to extend the results of this paper to nonlinear quantile regression. However, so far the asymptotic theory for parameter estimators under a misspecified nonlinear quantile model with random design, has not been developed yet, and so this needs to be considered first.

Another possible future research project consists of extending the present paper to the case where the response is subject to random right censoring (more complicated types of incomplete data can be considered in a later stage). So far, in the case of censored data, no quantile-analogue of the R^2-coefficient has been proposed in the literature. So, the first step will be to develop such a coefficient for censored data. Since censored data are often skewed (to the right), the development of a quantile-based coefficient for assessing model adequacy seems a very promising and useful project for practitioners.

Appendix

Theorem 2.2 can be easily proved using the fact that $\hat{a}\hat{b}^{-1} = ab^{-1} + \hat{b}^{-1}[\hat{a} - a - (\hat{b} - b)ab^{-1}]$ and the following expansions:

$$\frac{1}{n}\sum_{i=1}^{n} \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} \rho(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}^*) + o_p(n^{-1/2}), \tag{7.1}$$

$$\frac{1}{n}\sum_{i=1}^{n} \rho(Y_i - \mathbf{X}_{0,i}^\top \hat{\boldsymbol{\beta}}_0) = \frac{1}{n}\sum_{i=1}^{n} \rho(Y_i - \mathbf{X}_{0,i}^\top \boldsymbol{\beta}_0^*) + o_p(n^{-1/2}). \tag{7.2}$$

Since both (7.1) and (7.2) can be proven using the same arguments, we only provide the proof for (7.1), which is given in the following lemma.

Lemma 7.1 *Suppose that A1-A2 hold, and that there exists a* $M > 0$ *such that* $P(\|\mathbf{X}\| \leq M) = 1$. *Then, (7.1) holds true.*

Proof. Let $\hat{d}(\mathbf{X}_i) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top \mathbf{X}_i$ and consider the following decomposition of $\rho(\cdot)$:

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\rho(Y_i - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i) - \frac{1}{n}\sum_{i=1}^{n}\rho(Y_i - \boldsymbol{\beta}^{*\top}\mathbf{X}_i) \\ = & -\frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top \mathbf{X}_i\, \varphi(\varepsilon_i^*) \\ & -\frac{2}{n}\sum_{i=1}^{n}(\varepsilon_i^* - \hat{d}(\mathbf{X}_i)) \times \Big\{ I(\hat{d}(\mathbf{X}_i) > \varepsilon_i^* > 0) - I(\hat{d}(\mathbf{X}_i) < \varepsilon_i^* < 0) \Big\} \\ \equiv & \; A + B \quad (say). \end{aligned}$$

From Lemma 2.1, we have that

$$A = -(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \varphi(\varepsilon_i^*) = o_p(n^{-1/2}).$$

As for the term B, note that

$$\begin{aligned} |B| & \leq \frac{2}{n}\sum_{i=1}^{n}(|\hat{d}(\mathbf{X}_i)| + |\varepsilon_i^*|) I\Big\{|\varepsilon_i^*| < |\hat{d}(\mathbf{X}_i)|\Big\} \\ & \leq \frac{4}{n}\sum_{i=1}^{n}|\hat{d}(\mathbf{X}_i)|\; I\Big\{|\varepsilon_i^*| < |\hat{d}(\mathbf{X}_i)|\Big\} \\ & \leq 4\max_j |\hat{d}(\mathbf{X}_j)|\; \frac{1}{n}\sum_{i=1}^{n} I\Big\{|\varepsilon_i^*| < \max_j |\hat{d}(\mathbf{X}_j)|\Big\}. \end{aligned}$$

Using the Glivenko-Cantelli Theorem,

$$\begin{aligned} |B| & \leq 4\max_j |\hat{d}(\mathbf{X}_j)| \; \left\{ P\left(|\varepsilon^*| < \max_j |\hat{d}(\mathbf{X}_j)|\right) + o_p(1) \right\} \\ & = 4\max_j |\hat{d}(\mathbf{X}_j)| \left\{ F_{\varepsilon^*}\left(\max_j |\hat{d}(\mathbf{X}_j)|\right) - F_{\varepsilon^*}\left(-\max_j |\hat{d}(\mathbf{X}_j)|\right) \right\} + 4\max_j |\hat{d}(\mathbf{X}_j)| \times o_p(1) \\ & \leq 8 \sup_{e\in\mathcal{E}} f_{\varepsilon^*}(e) \left(\max_j |\hat{d}(\mathbf{X}_j)|\right)^2 + 4\max_j |\hat{d}(\mathbf{X}_j)| \times o_p(1), \end{aligned}$$

where F_{ε^*} is the distribution function of ε^*, and f_{ε^*} its density function. Finally, from the fact that the $\mathbf{X}_i$'s are bounded and using Lemma 2.1 we get that $\max_j |\hat{d}(\mathbf{X}_j)| = O_p(n^{-1/2})$, which proves that $B = o_p(n^{-1/2})$. □

Proof of Lemma 4.1. Let

$$f_n(\gamma) \quad = \quad \sum_{i=1}^{n} \Big(\rho(\varepsilon_i - \mathbf{X}_{in}^\top\gamma) - \rho(\varepsilon_i) + \mathbf{X}_{in}^\top\gamma\varphi(\varepsilon_i)\Big).$$

Recall from the proof of Lemma 3.3 in He and Shi (1994) that for any real number a, there exists a real r_a such that

$$\mathrm{E}_\varepsilon[\rho(\varepsilon + a) - \rho(\varepsilon)] = f_\varepsilon(0)a^2 + \frac{a^2}{2}r_a, \tag{7.3}$$

where $r_a \to 0$ as $|a| \to 0$. Hence, there exists $r_{i\gamma},\ i = 1, \ldots, n$, such that

$$\mathrm{E}f_n(\gamma) \quad = \quad \sum_{i=1}^{n}\mathrm{E}_{\mathbf{X}}\Big\{E_\varepsilon[\rho(\varepsilon_i - \mathbf{X}_{in}^\top\gamma) - \rho(\varepsilon_i)]\Big\} = \sum_{i=1}^{n}\mathrm{E}_{\mathbf{X}}\Big\{f_\varepsilon(0)(\mathbf{X}_{in}^\top\gamma)^2 + \frac{1}{2}(\mathbf{X}_{in}^\top\gamma)^2 r_{i\gamma}\Big\},$$

and, for any $\epsilon > 0$, there exists $\delta > 0$ such that $|\mathbf{X}_{in}^\top\gamma| \le \delta$ implies $|r_{i\gamma}| \le \epsilon$. Since $\|\mathbf{X}_{in}\| \to 0$ by the boundedness of $\mathbf{X}$, we have

$$\mathrm{E}f_n(\gamma) = \sum_{i=1}^{n}\mathrm{E}_{\mathbf{X}}\Big\{f_\varepsilon(0)(\mathbf{X}_{in}^\top\gamma)^2\Big\}(1 + o(1)) = f_\varepsilon(0)\gamma^\top E(\mathbf{X}\mathbf{X}^\top)\gamma(1 + o(1)).$$

By Schwarz's inequality,

$$\begin{aligned}
\mathrm{Var}f_n(\delta) \quad &\le \quad \sum_{i=1}^{n}\mathrm{E}_{\mathbf{X},\varepsilon}\Big\{\Big(\int_0^{-\mathbf{X}_{in}^\top\delta}[\varphi(\varepsilon_i + u) - \varphi(\varepsilon_i)]du\Big)^2\Big\} \\
&\le \quad \sum_{i=1}^{n}\mathrm{E}_{\mathbf{X}}\Big\{\|\mathbf{X}_{in}^\top\delta\|\Big|\int_0^{-\mathbf{X}_{in}^\top\delta}\mathrm{E}_\varepsilon[\varphi(\varepsilon_i + u) - \varphi(\varepsilon_i)]^2du\Big|\Big\} \\
&\le \quad \sum_{i=1}^{n}\mathrm{E}_{\mathbf{X}}\Big\{\delta^\top\mathbf{X}_{in}\mathbf{X}_{in}^\top\delta\Big\} \times o(1) \le \delta^\top\mathrm{E}(\mathbf{X}\mathbf{X}^\top)\delta \times o(1) \ \to \ 0, \ \textit{as } n \to \infty.
\end{aligned}$$

Since $f_n(\delta)$ is convex in δ, Lemma 4.1 follows from Lemma 2.1 in Rao and Zhao (1992). □

Proof of Theorem 4.2. For the proof, we adopt the following notations.

$$\begin{aligned}
\mathbf{\Sigma} \quad &= \quad \mathrm{E}(\mathbf{X}\mathbf{X}^\top), \quad \mathbf{T} = (\mathbf{I}_{d_0\times d_0}, \mathbf{O}_{d_0\times d_1})^\top, \quad \mathbf{P} = \mathbf{T}\{\mathbf{T}^\top\mathbf{\Sigma}\mathbf{T}\}^{-1}\mathbf{T}^\top, \\
\mathbf{S} \quad &= \quad (\mathbf{\Sigma}^{-1} - \mathbf{P})^{1/2} = (\mathbf{s}_1 \cdots \mathbf{s}_d)^\top,
\end{aligned}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{O}$ is a zero matrix and $\mathbf{s}_k$ is the kth row of $\mathbf{S}$. Let $\tilde{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0^\top, \mathbf{0}^\top)^\top$ be the minimizer of $\sum_{i=1}^{n}\rho(Y_i - \mathbf{X}_i^\top\boldsymbol{\beta})$ with the constraint $\boldsymbol{\beta}_1 = \mathbf{0}$. Following the steps in the proof of

Theorem 1 in Kim and White (2003), we have

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \frac{1}{2f_\varepsilon(0)} \frac{1}{n} \sum_{i=1}^n \mathbf{P}\mathbf{X}_i \varphi(\varepsilon_i) + o_p(n^{-1/2}), \tag{7.4}$$

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \frac{1}{2f_\varepsilon(0)} \frac{1}{n} \sum_{i=1}^n \boldsymbol{\Sigma}^{-1}\mathbf{X}_i \varphi(\varepsilon_i) + o_p(n^{-1/2}), \tag{7.5}$$

$$\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} = \frac{1}{2f_\varepsilon(0)} \frac{1}{n} \sum_{i=1}^n \{\boldsymbol{\Sigma}^{-1} - \mathbf{P}\}\mathbf{X}_i \varphi(\varepsilon_i) + o_p(n^{-1/2}), \tag{7.6}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^\top, \mathbf{0}^\top)^\top$. Observe that

$$\left(\boldsymbol{\Sigma}^{-1} - \mathbf{P}\right)^\top \boldsymbol{\Sigma} \left(\boldsymbol{\Sigma}^{-1} - \mathbf{P}\right) = \boldsymbol{\Sigma}^{-1} - \mathbf{P} \tag{7.7}$$

$$\left(\boldsymbol{\Sigma}^{-1} - \mathbf{P}\right)^\top \boldsymbol{\Sigma}\, \mathbf{P} = \mathbf{O} \tag{7.8}$$

Lemma 4.1 implies that

$$\begin{aligned}
&\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \tilde{\boldsymbol{\beta}}) - \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}) \\
= &\sum_{i=1}^n \rho(\varepsilon_i - \mathbf{X}_i^\top (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})) - \sum_{i=1}^n \rho(\varepsilon_i - \mathbf{X}_i^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\
= &\sum_{i=1}^n \mathbf{X}_i^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\varphi(\varepsilon_i) - \sum_{i=1}^n \mathbf{X}_i^\top (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\varphi(\varepsilon_i) \\
&+ n f_\varepsilon(0)(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) - n f_\varepsilon(0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1),
\end{aligned}$$

where $\varepsilon_i = Y_i - \mathbf{X}_i^\top \boldsymbol{\beta} = Y_i - \mathbf{X}_{0,i}^\top \boldsymbol{\beta}_0$. Since $(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_p(n^{-1})$ by (7.4),(7.6) and (7.8), we have

$$\begin{aligned}
&\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \tilde{\boldsymbol{\beta}}) - \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}) \\
= &\sum_{i=1}^n \mathbf{X}_i^\top (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\varphi(\varepsilon_i) - n f_\varepsilon(0)(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) + o_p(1)
\end{aligned} \tag{7.9}$$

Applying (7.6) and (7.7) to Equation (7.9), we obtain

$$\begin{aligned}
&\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \tilde{\boldsymbol{\beta}}) - \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}) \\
= &\frac{q(1-q)}{f_\varepsilon(0)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{S}\mathbf{X}_i \frac{\varphi(\varepsilon_i)}{\sqrt{4q(1-q)}} \right\|^2 + o_p(1)
\end{aligned} \tag{7.10}$$

Before deriving the asymptotic distribution of (7.10), we first show that $\text{rank}(\mathbf{S}) = d_1$. Since $\text{rank}(\mathbf{S}) = \text{rank}(\mathbf{S}^2)$, it is enough to show that $\text{rank}(\mathbf{S}^2) \leq d_1$ and $\text{rank}(\mathbf{S}^2) \geq d_1$. The first inequality follows by applying Sylvester's rank inequality in Kaw (2011) to (7.8). The second inequality holds because $\mathbf{e}_{d_0+1}, \cdots, \mathbf{e}_d$ are the eigenvectors of $\mathbf{\Sigma S}^2$ with common eigenvalue 1 and $\mathbf{\Sigma}^{-1}$ is invertible, where $\mathbf{e}_1, \dots, \mathbf{e}_d$ are the standard basis for $\mathbb{R}^d$.

Since rank($\mathbf{S}$) $= d_1$, we can choose a linearly independent subset of d_1 vectors from $\mathbf{s}_1, \dots, \mathbf{s}_d$. Applying the Gram-Schmidt process to such vectors with the inner product $< \mathbf{a}, \mathbf{b} >= \mathbf{a}^\top \mathbf{\Sigma b}$ leads us to obtain $\mathbf{u}_1, \dots, \mathbf{u}_{d_1}$ such that for $1 \le l, m \le d_1$,

$$\mathbf{u}_l^\top \mathbf{\Sigma u}_m = 0 \text{ for } l \neq m \quad \text{and} \quad \mathbf{u}_l^\top \mathbf{\Sigma u}_l = 1; \tag{7.11}$$

$$\mathbf{s}_k \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_{d_1}) \text{ for } k = 1, \dots, d, \tag{7.12}$$

where Span($\mathbf{u}_1, \dots, \mathbf{u}_{d_1}$) is the vector space spanned by $\mathbf{u}_1, \dots, \mathbf{u}_{d_1}$. Furthermore, it can be shown that for such $\mathbf{u}_1, \dots, \mathbf{u}_{d_1}$ there exists a $d \times d_1$ matrix $\mathbf{C}$ such that

$$\begin{pmatrix} \mathbf{s}_1^\top \\ \vdots \\ \mathbf{s}_d^\top \end{pmatrix} = \mathbf{S} = \mathbf{C} \begin{pmatrix} \mathbf{u}_1^\top \\ \vdots \\ \mathbf{u}_{d_1}^\top \end{pmatrix} \equiv \mathbf{CU} \quad \text{and} \quad \mathbf{C}^\top \mathbf{C} = \mathbf{I}_{d_1 \times d_1}. \tag{7.13}$$

From (7.13), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{SX}_i \frac{\varphi(\varepsilon_i)}{\sqrt{4q(1-q)}} = \mathbf{C} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{UX}_i \frac{\varphi(\varepsilon_i)}{\sqrt{4q(1-q)}} \equiv \mathbf{C} \rho_n$$

Using the fact that $\mathbf{X}_i$ and ε_i are independent and (7.11), it is evident that the distribution of ρ_n converges to a standard multivariate normal distribution of dimension $d_1 = d - d_0$. Since $\mathbf{C}^\top \mathbf{C} = \mathbf{I}_{d_1 \times d_1}$, we obtain that $\sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \tilde{\boldsymbol{\beta}}) - \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}})$ converges to $q(1-q)\chi^2_{d_1}/f_\varepsilon(0)$ as $n \to \infty$. Finally, by the law of large numbers and the decomposition

$$\frac{1}{n} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \tilde{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}) + o_p(n^{-1/2}),$$

we obtain the asymptotic distribution of $\hat{\zeta}(q)$. □

References

Capizzi, M., G. Leto, A. Petrone, S. Zampetti, R. E. Papa, M. Osimani, M. Spoletini, A. Lenzi, J. Osborn, M. Mastantuono, A. Vania, and R. Buzzetti (2011). Wrist circumference is a clinical marker of insulin resistance in overweight and obese children and adolescents. *Circulation (published online before print)*.

Coutinho, T., K. Goel, S. D. de Corrêa, C. Kragelund, A. M. Kanaya, M. Zeller, J. S. Park, L. Kober, C. Torp-Pedersen, Y. Cottin, L. Lorgis, S. H. Lee, Y. J. Kim, R. Thomas, V. L. Roger, V. K. Somers, and F. Lopez-Jimenez (2011). Central obesity and survival in subjects with coronary artery disease: a systematic review of the literature and collaborative analysis with individual subject data. *Journal of the American College of Cardiology 57*, 1877–1886.

Dette, H. and A. Munk (2003). Some methodological aspects of validation of models in nonparametric regression. *Statistica Neerlandica 57*, 207–244.

He, X. and P. Shi (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics 3*, 299–308.

Hodges, J. L. and E. L. Lehmann (1954). Testing the approximative validity of statistical hypotheses. *Journal of the Royal Statistical Society – Series B 16*, 261–268.

Kaw, A. (2011). *Introduction to Matrix Algebra*. Autar Kaw.

Kim, T.-W. and H. White (2003). Estimation, inference and specification testing for possibly misspecified quantile regression. *Advances In Econometrics 17*, 107–132.

Koenker, R. (2005). *Quantile Regression*, Volume 38 of *Econometric Society Monographs*. Cambridge: Cambridge University Press.

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

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

Li, Z. Y. and S. Z. Zou (1994). Nonparametric kernel estimation for error density. *Proceedings of 1994 Workshop on Information Theory and Statistics*, 93.

Mckean, J. W. and G. L. Sievers (1987). Coefficient of determination for least absolute deviation analysis. *Statistics & Probability Letters 5*, 49–54.

Penrose, K., A. Nelson, and A. Fisher (1985). Generalized body composition prediction equation for men using simple measurement techniques. *Medicine and Science in Sports and Exercise 17*, 189.

Rao, C. R. and L. C. Zhao (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. *Indian Journal of Statistics 54*, 323–331.

	$\sigma = 1$ ($R(0.5) = 0.657$)	$\sigma = 3$ ($R(0.5) = 0.262$)	$\sigma = 5$ ($R(0.5) = 0.124$)
$n = 50$	0.922	0.916	0.925
$n = 100$	0.929	0.936	0.939
$n = 200$	0.945	0.947	0.941
$n = 400$	0.946	0.952	0.944

Table 1: Empirical coverage probabilities of 95% confidence intervals for $R(0.5)$ from 1000 random samples of size $n = 50, 100, 200$ and 400.

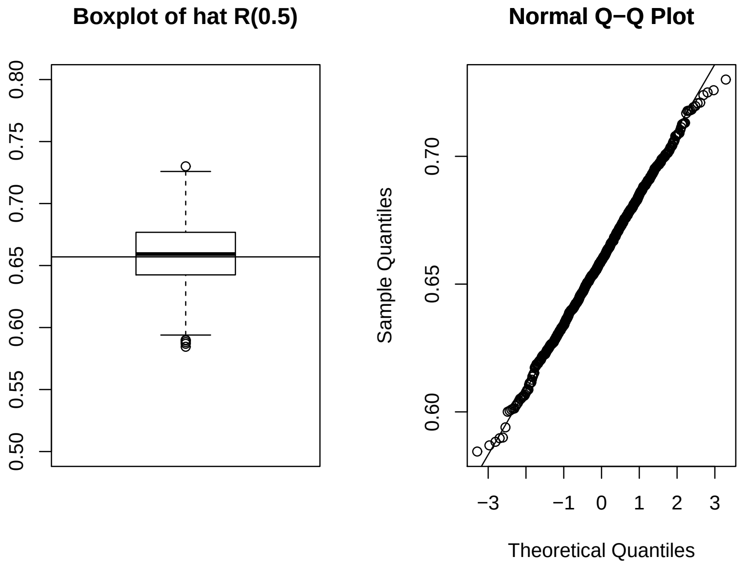

Figure 1: The boxplot and Q-Q plot of $\hat{R}(0.5)$.

			$\zeta_{X_1}(q)$	$\zeta_{X_2}(q)$	$\zeta_{X_3}(q)$
$\nu = 0$	$q = 0.5$	$\sigma = 0.4$	0.188	0.437	0.593
		$\sigma = 0.8$	0.057	0.186	0.324
	$q = 0.2$	$\sigma = 0.4$	0.184	0.420	0.572
		$\sigma = 0.8$	0.057	0.182	0.315
$\nu = 2$	$q = 0.5$	$\sigma = 0.4$	0.664	0.397	0.558
		$\sigma = 0.8$	0.441	0.179	0.316
	$q = 0.2$	$\sigma = 0.4$	0.609	0.386	0.541
		$\sigma = 0.8$	0.392	0.173	0.305

Table 2: The values of $\zeta_{X_1}(q), \zeta_{X_2}(q)$ and $\zeta_{X_3}(q)$ for various settings.

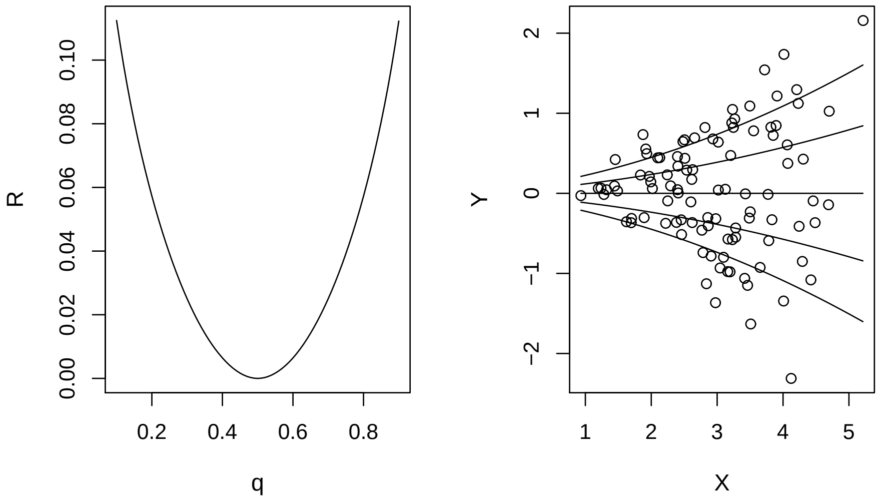

Figure 2: The $R(q)$ function and the scatterplot of one random sample of size 100 with the true conditional quantile functions for $q \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$.

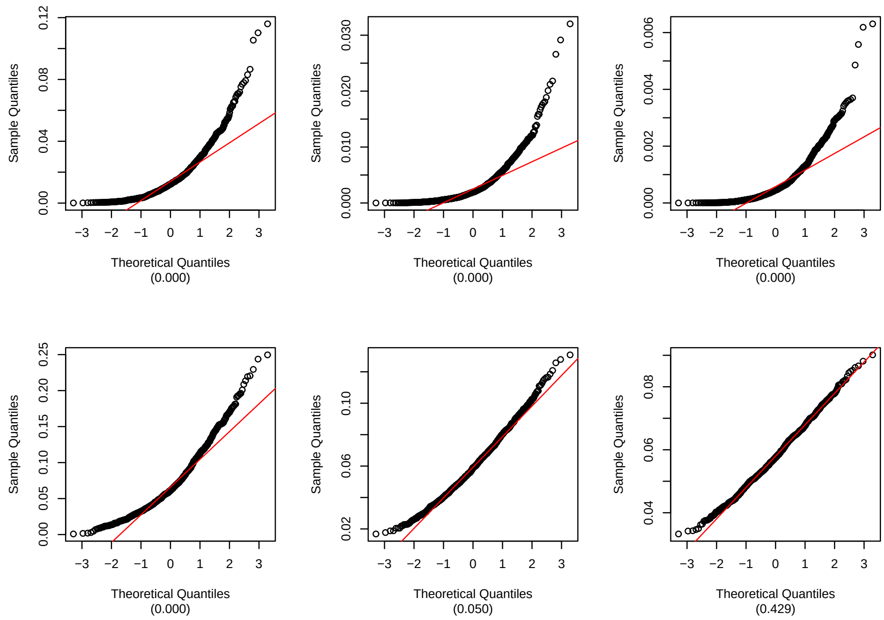

Figure 3: Q-Q plots and p-values of the KS test for the distribution of $\hat{R}(0.5)$ (upper panels, degenerate case) and $\hat{R}(0.2)$ (lower panels, non-degenerate case) based on 1000 random samples of different sample sizes ($n = 100, 500$ and 2000, from left to right). For each Q-Q plot the number between parentheses is the p-value of the corresponding KS-test.

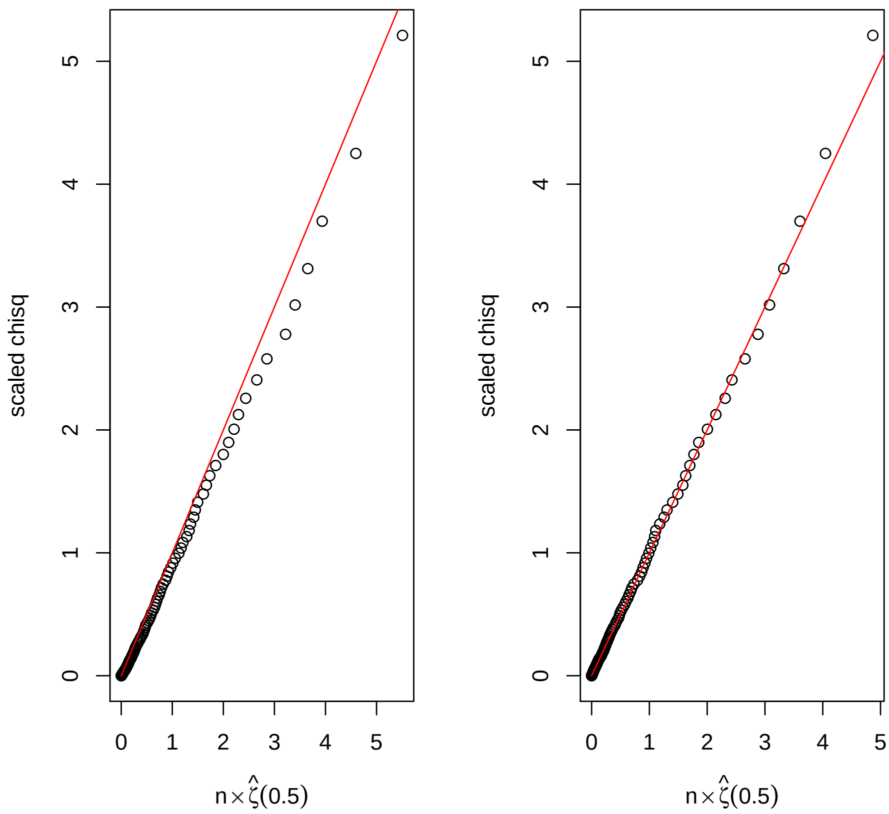

Figure 4: Q-Q plots of $n\hat{\zeta}(0.5)$ with respect to $(\pi/4)\chi^2_1$ when $n = 500$ (left panel) and $n = 1000$ (right panel). The solid line is the bisector.

	Abdomen circumference			Body mass index		
	lower limit	estimate	upper limit	lower limit	estimate	upper limit
$q = 0.5$	0.3779	0.4406	0.5033	0.2740	0.3436	0.4131
$q = 0.6$	0.3733	0.4368	0.5002	0.2692	0.3396	0.4101
$q = 0.7$	0.3665	0.4337	0.5008	0.2627	0.3363	0.4098
$q = 0.8$	0.3537	0.4236	0.4936	0.2595	0.3383	0.4170
$q = 0.9$	0.3670	0.4405	0.5139	0.2661	0.3539	0.4418

Table 3: 95%-confidence intervals for $R(q)$ for two factors and different values of q.

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
upper limit(q)	0.100	0.083	0.094	0.097	0.102	0.102	0.104	0.095	0.108

Table 4: The limit of the upper confidence interval for $R(q)$ at significance level $\alpha = 0.05$ for the model containing wrist circumference as only factor.

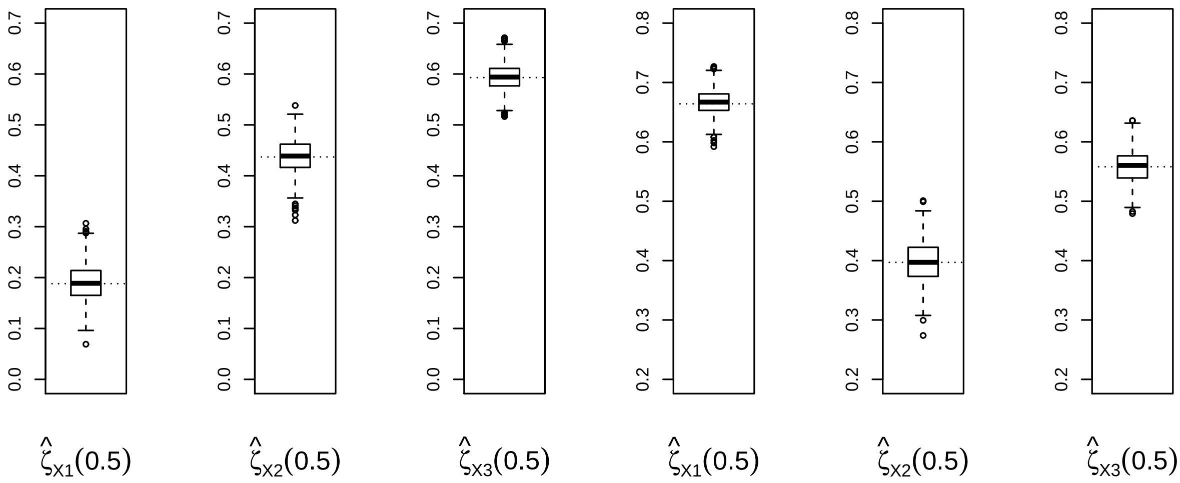

Figure 5: Boxplots of $\hat{\zeta}_{X_k}(0.5)$, $k = 1, 2, 3$, when $\nu = 0$ (left three panels) and $\nu = 3$ (right three panels) based on 500 datasets of size 200. The horizontal dotted line in each box plot represents the corresponding true value.

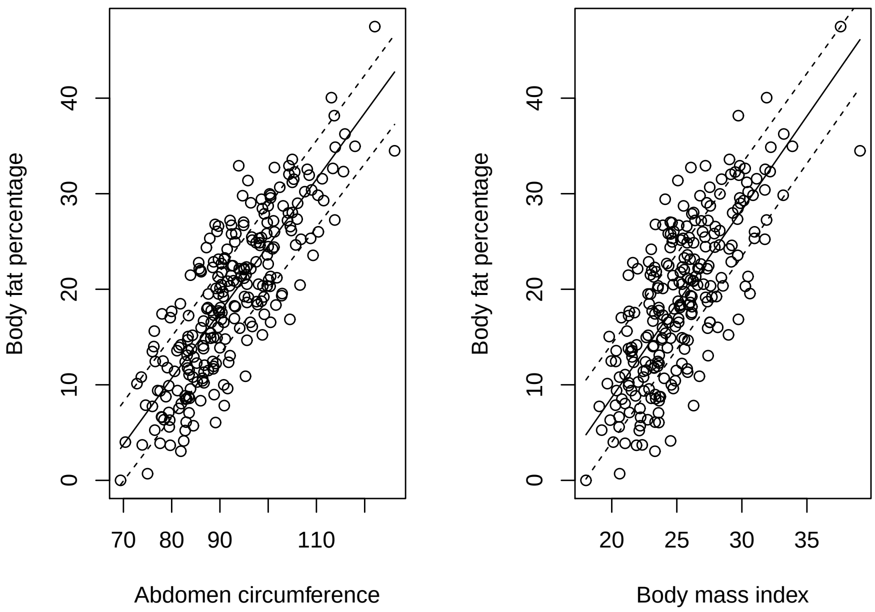

Figure 6: Scatter plots of abdomen circumference and body mass index versus body fat percentage, along with three conditional quantile lines ($q = 0.2$ – lower dashed, 0.5 – solid, and 0.8 – upper dashed).